

# APLICACIÓN DE MEZCLAS DE DISTRIBUCIONES A LA LOCALIZACIÓN DE FALLAS EN SISTEMAS DE DISTRIBUCIÓN DE ENERGÍA ELÉCTRICA

---

**JORGE ANDRÉS CORMANE ANGARITA**

*Ingeniero Electricista*

*Magíster en Ingeniería Eléctrica*

*Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones*

*Universidad Industrial de Santander*

*jcormane@uis.edu.co*

**HERMANN RAÚL VARGAS TORRES**

*Doctor Ingeniero Industrial*

*Profesor Asociado*

*Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones*

*Universidad Industrial de Santander*

*hrvargas@uis.edu.co*

**GABRIEL ORDÓÑEZ PLATA**

*Doctor Ingeniero Industrial*

*Profesor Titular*

*Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones*

*Universidad Industrial de Santander*

*gaby@uis.edu.co*

*Fecha Recepción: 19 de septiembre de 2005*

*Fecha Aceptación: 23 de mayo de 2006*

## RESUMEN

*Mejorar la confiabilidad de las redes de distribución requiere de una gran inversión, pero no todas las empresas están en capacidad de asumirlas. Por lo tanto, cualquier estrategia que permita mejorar la confiabilidad, se debe reflejar directamente en la disminución de la duración y frecuencia de las interrupciones. En este artículo se presenta una alternativa de solución a los problemas de continuidad del suministro asociados a la localización de fallas. Se propone una metodología de tipo estadístico basada en la mezclas de distribuciones. Con esta, se construye un modelo estadístico a partir de la extracción de patrones característicos de las señales registradas por los equipos de medida, junto con los parámetros y topología propios de la red durante el evento. La utilización de esta metodología tiene por objeto brindar una alternativa económica y de fácil implementación para el desarrollo de estrategias orientadas a mejorar la confiabilidad a partir de la disminución de los tiempos de atención y recuperación del sistema.*

**PALABRAS CLAVE:** *Localización de fallas, Calidad de Potencia, Índices de calidad, Modelo estadístico, Clasificación, Mezclas de distribuciones, Estadística multivariada.*

## ABSTRACT

*The enhancement of power distribution system reliability requires a great investment but not all the utilities are in a position to assume it. Therefore, any strategy that allows the improvement of reliability should be reflected directly in the decrease of the duration and frequency of interruptions. In this paper an alternative of solution to the problems of continuity associated to fault location is presented. A methodology of statistical nature is proposed using finite mixture. With this approach a statistical model is obtained from the extraction of characteristic patterns of the signals registered by measurement equipments, along with the parameters and own topology of the network during the event. The purpose of this methodology is to offer an economic alternative of easy implementation for the development of strategies oriented to improve the reliability from the decrease in the times of attention and recovery of the system.*

**KEYWORDS:** *Fault Location, Power Quality, Statistical Models, Classification, Mixture Model, Multivariate analysis.*

## INTRODUCCIÓN

Desde principios de los noventa se ha fomentado el interés en los temas concernientes a la calidad del servicio<sup>1</sup> por parte de algunos países. Esto se debe al ambiente de libre competencia del sector eléctrico, donde la calidad es un buen reflejo de la participación de las empresas en el mercado energético y uno de los aspectos de mayor interés frente a las exigencias del sector a nivel mundial. En Colombia, a partir del establecimiento del nuevo marco regulatorio del sector eléctrico, se pretende fortalecer la distribución y el mercado minorista desde la óptica de la calidad del producto<sup>2</sup> [1], [2]. La dependencia existente entre el ser humano y la energía eléctrica, ha determinado que cada vez y con más insistencia se requiera que el suministro se realice bajo criterios de seguridad, confiabilidad y calidad [3]. Tales acciones se han promovido por medio de tarifas fijas en los usuarios finales y a través de la aplicación de estándares para la prestación del servicio por parte de las empresas distribuidoras [4]. La continuidad en el suministro es uno de los aspectos más importantes para el cliente. Dicha importancia se origina por el impacto económico y social generado a partir de la ocurrencia de interrupciones [5], [6].

Aunque económicamente no es viable lograr un 100% de confiabilidad, las empresas se están esforzando en solucionar el problema de la continuidad mediante una adecuada planificación y operación del sistema [7]. En la actualidad, se estima que el 80% de las interrupciones son causadas por fallas en sistemas de distribución. Pensar en la implementación de algoritmos similares a los utilizados en los sistemas de transmisión no es viable en sistemas de distribución puesto que su funcionamiento se ve afectado por las características técnicas (topología, secciones no homogéneas, múltiples derivaciones, equipos de medida) y operativas (operación radial), comprometiendo la eficiencia del mismo [8].

Existe variedad de métodos para la localización de fallas en sistemas de distribución. Estos métodos se clasifican en tres categorías. La primera, métodos que detectan componentes de alta frecuencia de ondas viajeras. La segunda, métodos que utilizan el cálculo de la impedancia y la tercera, métodos de inspección visual, que implican el patrullaje y reconocimiento de la línea [8], [9]. La idea central de este trabajo es proponer una alternativa de solución a los problemas de continuidad del suministro asociados a la

localización de fallas, con el objetivo de disminuir el tiempo de restablecimiento de las condiciones de operación normal del sistema [10]. Esta situación permite mejorar la confiabilidad a partir del planteamiento de estrategias orientadas a la disminución de los tiempos de atención y recuperación del sistema, brindando una alternativa de bajo costo y fácil implementación a las empresas del sector.

Dicha propuesta se apoya en las últimas modificaciones a las normas de calidad de la potencia eléctrica realizadas por la CREG<sup>3</sup>, respecto a la monitorización de la continuidad del suministro y calidad de la potencia suministrada [11]. Es de notar que los resultados mostrados en el estudio fueron obtenidos utilizando únicamente datos de simulación. Las etapas siguientes del estudio es la aplicación del modelo a datos reales (mediciones).

Este documento se encuentra estructurado de la siguiente manera, en la Metodología Propuesta para la localización de Fallas, se introduce la teoría de la técnica utilizada y sus requerimientos para la construcción del modelo. En el tema de los Algoritmos Utilizados, se presenta la metodología seguida durante la investigación, dejando claras las etapas planteadas. En el Ejemplo de Aplicación, ilustra la operación de los algoritmos utilizados. Además, se describe y presenta gráficamente los datos utilizados, adicionalmente se muestran los resultados del ejemplo de aplicación. Finalmente, se plantean las conclusiones del trabajo.

---

<sup>1</sup> La calidad del servicio prestado enmarca aspectos concernientes a la continuidad del suministro.

<sup>2</sup> La calidad del producto, de onda o de potencia suministrada, se refiere a las perturbaciones y variaciones del estado estacionario de tensión y corriente, regulación de frecuencia y tensión, armónicos, flicker y factor de potencia.

<sup>3</sup> Comisión Reguladora de Energía y Gas - CREG

## TEORÍA ACERCA DEL ANÁLISIS MULTIVARIABLE

El análisis multivariable puede plantearse en dos niveles, el primero, consiste en la extracción de la información que contienen los datos disponibles (Análisis Exploratorio de Datos - EDA) y el segundo, en el cual se pretende generar conocimiento sobre el problema y obtener conclusiones sobre la población de los datos, a partir de la construcción de un modelo estadístico que permita explicar la forma como se generan y predicen los datos (inferencia) [14]. Tomando como base la propuesta para el análisis de datos multivariantes mencionada anteriormente, se aplican tres técnicas para la construcción adecuada del modelo, la primera, Análisis de Componentes Principales – ACP, la segunda, k-medias y la tercera, Mezclas de distribuciones – MD. Con la aplicación de las dos primeras se realiza la exploración de los datos y la definición de las variables a utilizar, mientras que con la tercera se realiza la estimación de la función de probabilidad.

Como un paso adelante en el entendimiento de los datos multivariantes se tiene que la exploración visual es una herramienta poderosa para el análisis de la información, esto ayuda al entendimiento de los mismos aunque el aumento del volumen y la complejidad de la información sean evidentes [12]. La aplicación del ACP tiene como objetivo la estructuración de un conjunto de datos multivariable mediante la reducción del número de variables, a costa de una pérdida mínima de información [13]. La descomposición de los datos en *clusters* permite la conformación de grupos significativos de manera analítica, con el objetivo de clasificar los datos de una muestra según similitudes o afinidades [14]. Adicionalmente el uso de modelos relativamente simples para cada estructura local, ofrece gran facilidad de interpretación, análisis y simplificación computacional [15], [16]. Los algoritmos de agrupamiento se basan en la utilización de diferentes métricas para el cálculo de la distancia entre puntos. Las métricas se encuentran sujetas a las propiedades presentadas en (1). Donde  $d(A,B)$  es la distancia entre dos puntos dados, y pertenecientes a  $\mathbb{R}^p$

$$\begin{aligned} d(A,B) &\geq 0 \quad \forall (A,B) \\ d(A,B) &= 0 \Leftrightarrow B = A \\ d(A,B) &= d(B,A) \\ d(A,B) &\leq d(A,C) + d(C,B) \end{aligned} \quad (1)$$

Existe gran variedad de métricas asociadas con la cuantificación de la variabilidad de los datos, en este trabajo

se utilizan las métricas Euclidiana (2) y Mahalanobis (3) para el análisis de clusters por ser las que mejores resultados han arrojado en esta área [17]. Las Ecuaciones 2 y 3 están planteadas de forma matricial, donde  $\mathbf{x}$  es el vector de datos,  $\bar{\mathbf{x}}$  es el vector de medias,  $\mathbf{x}^T$  es la operación transpuesta de la matriz y  $\mathbf{V}$  es la matriz de covarianza.

$$dE_i = \left[ (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \right]^{1/2} \quad (2)$$

$$dM_i = \left[ (\mathbf{x} - \bar{\mathbf{x}})\mathbf{V}^{-1}(\mathbf{x} - \bar{\mathbf{x}})^T \right]^{1/2} \quad (3)$$

Con las métricas se evalúan dos aspectos importantes, el primero, concerniente a la cercanía entre puntos de un mismo grupo (homogeneidad interna), esto nos indica que tan compacto está un conjunto de datos. El segundo, relacionado con la distancia entre los elementos de los diferentes grupos (heterogeneidad externa), para garantizar que no se genere traslape entre grupos, por el hecho de estar relativamente cerca [17].

Las MD es una técnica de modelado estadístico con la cual se obtiene una estimación de la función densidad de probabilidad de los datos en una muestra aleatoria, la cual se representa como una suma ponderada finita de las componentes de densidad multivariantes. Esta técnica provee un sustento matemático para el modelado estadístico en una amplia variedad de fenómenos, reflejado en los últimos veinte años donde se han logrado avances considerables en el ajuste de modelos de mezclas [18]. Las MD se han aplicado con múltiples propósitos como: el modelado de la heterogeneidad de una población (Biología), manejo de datos faltantes, estimación de densidades de probabilidad y análisis de conglomerados (Estadística), reconocimiento de patrones (Imágenes) y obtención de consumos anómalos (Ingeniería Eléctrica). Dentro de las características principales de las MD se puede observar que no requieren un parámetro de suavizado de la función. Se puede establecer el número de términos de la mezcla y poseen un menor gasto computacional en comparación con otros métodos como la estimación de densidad Kernel y los histogramas [17]. La utilidad de la MD radica en el análisis y modelado de conglomerados (clusters) [18]. Al aplicar la técnica en el problema de localización de fallas, se utilizarán los conceptos sobre datos multivariantes. Las MD multivariantes se expresan como se muestra en (4).

$$f(\mathbf{x}) = \sum_{g=1}^G p_g f_g(\mathbf{x}, \boldsymbol{\theta}) \quad (4)$$

Las MD sobre una muestra poblacional  $\mathbf{X}$  compuesta de observaciones d-dimensionales, contiene  $G$  componentes

$f_g(\mathbf{x}, \theta)$  referentes a la densidad de distribución multivariada seleccionada. Cada componente de densidad  $f_g$  intenta describir el comportamiento de un grupo dentro de la población en el cual los datos relativos a dicho grupo poseen características similares, establecidas en el vector estimador  $\theta$  correspondiente a los parámetros de cada distribución. Un estimador es un parámetro que define el comportamiento de los datos dentro de un grupo, y por tanto, su forma. Por ejemplo, los estimadores para una distribución de probabilidad normal son: el vector de medias  $\hat{\mu}$  y la matriz de covarianza  $\hat{\Sigma}$ , que definen el punto central de la distribución y la forma como se concentran los datos alrededor de dichos puntos respectivamente. La cantidad  $p_g$ , llamada proporción de la mezcla o coeficiente de mezclado, brinda información sobre la importancia del grupo dentro de la mezcla. Las condiciones que deben cumplir los coeficientes de mezclado están dadas en (5).

$$\sum_{g=1}^G p_g = 1 \quad p_g > 0 \quad (5)$$

El propósito de la MD es identificar una cantidad desconocida de grupos en los cuales se aglomeran los datos de una muestra, es decir, la MD busca la homogeneidad dentro de una muestra inicialmente heterogénea. Para lograr este objetivo la MD utiliza el algoritmo Expectation-Maximization (EM) para la estimación de parámetros. El algoritmo EM aparece como una forma de aplicación del método de la Máxima Verosimilitud (MV) para la obtención de parámetros faltantes, permitiendo hallar de forma iterativa el vector estimador  $\theta$  de MV para cada distribución  $f_g(\mathbf{x}, \theta)$ , a partir del ingreso de valores iniciales para los parámetros de  $\hat{\theta}^{(0)}$  [19]. El algoritmo trabaja con la función soporte  $L(\mathbf{Y}, \mathbf{Z} | \theta)$  de la muestra, hallando la esperanza de las funciones de los datos ausentes  $\mathbf{Z}$  a partir del cálculo de la distribución de  $\mathbf{Z}$  con los valores iniciales  $\hat{\theta}^{(0)}$  y los datos observados  $\mathbf{Y}$ .

$$L(\theta | \mathbf{Y}) = E[\ell(\theta | \mathbf{Y}, \mathbf{Z})] \quad (6)$$

Al resultado de esta operación se le denomina paso E (Estimación) dada en (6). En el paso M (Maximización) se maximiza la función  $L(\theta | \mathbf{Y})$  respecto a  $\theta$  con el fin de conseguir los estimadores MV a partir de la sustitución de las observaciones faltantes por estimaciones de sus valores. Sea  $\hat{\theta}^{(i+1)}$  el estimador obtenido en el paso M, se retorna al paso E de manera iterativa hasta obtener la convergencia deseada [20]. El método MV escoge como estimador de los parámetros aquel valor que hace máxima la probabilidad de que el modelo a estimar genere la muestra observada [19], [21].

## METODOLOGÍA PROPUESTA PARA LA LOCALIZACIÓN DE FALLAS

Una forma de resolver el problema de la localización es conocer el comportamiento del sistema cuando se encuentra sometido a diferentes perturbaciones. Para lograr este propósito se propone realizar la caracterización de la respuesta del sistema en presencia de fallas, la cual, se ve reflejada en las señales de tensión y corriente. Las fallas más comunes en sistemas de distribución son el resultado de corto circuitos (L-T, L-L, L-L-T, L-L-L y L-L-L-T). En consecuencia, la obtención de las señales es un paso fundamental en el momento extraer las características que describan el comportamiento del sistema.

A continuación se efectuará una descripción por etapas de la metodología propuesta. En la primera etapa, se selecciona el sistema prototipo. En la segunda etapa, se implementa el sistema prototipo en el software ATP/EMTP con el objetivo de realizar simulaciones del sistema bajo condiciones de falla, conformando un banco de registros con las señales de tensión y corriente, obtenidas en cada simulación.

En la tercera etapa, se procede a realizar la adecuación de las señales de tensión y corriente, en la cual, se obtiene el valor eficaz de las señales en el tiempo y se convierten al sistema por unidad. En la cuarta etapa, se planteó la caracterización del sistema de dos formas: la primera, de tipo determinista basada en el cálculo de descriptores monofásicos de la señal de tensión [22]. La segunda, de tipo estadística la cual está basada en la aplicación de un análisis exploratorio de los datos, obteniendo finalmente descriptores característicos del comportamiento del sistema. En la quinta etapa, se procede a realizar el análisis de la información extractada en los descriptores, estableciendo reglas y condiciones para la conformación de zonas características y relaciones de heterogeneidad entre los grupos. Finalmente, en la sexta etapa, se conforma el modelo de mezclas y se clasifican los datos de falla, obteniendo la localización probable de la falla.

## ALGORITMOS UTILIZADOS

Se presenta la descripción de los algoritmos utilizados en cada paso de la construcción del modelo. Un esquema tipo resumen de los algoritmos se muestra a continuación. El algoritmo k-medias para la determinación de clusters es:

1. Especificar el número de grupos  $u$  obtener una partición inicial de la muestra.
2. Determinar los centros de los grupos, puede ser de forma aleatoria o a priori.
3. Tomar cada observación y calcular la distancia a cada grupo (2).
4. Cada observación se ubica en el grupo cuya distancia sea mínima y se recalculan los centros.
5. Repetir los pasos 2, 3 y 4 hasta cuando no ocurran más cambios en los grupos.

La idea es que la suma de los cuadrados de la distancia de todos los puntos que forman un grupo respecto a su centro sea mínima. A partir del agrupamiento realizado se inicia la identificación de las características predominantes en cada grupo, con el objetivo de inferir y relacionar nuevos datos. Con la aplicación del algoritmo k-medias, se estiman los valores iniciales para los centros. El valor inicial de la matriz de covarianza se estima con la matriz identidad y el de los coeficientes de mezclado se calcula con la proporción de datos que poseen los grupos respecto a la muestra. Obtenidos los parámetros iniciales, se inicia la estimación de los parámetros con el algoritmo EM hasta alcanzar el valor de convergencia deseado. El resultado de este procedimiento son las estimaciones de los parámetros  $\hat{\mu}$ ,  $\hat{\Sigma}$  y  $P$  de cada grupo. Estos parámetros serán utilizados en (4) con el fin de clasificar otras observaciones. Los pasos del algoritmo EM se presentan a continuación.

1. Determinar el número de componentes de la mezcla (k-medias).
2. Determinar los valores iniciales de los parámetros de las componentes de densidad ( $\mu^{(0)}$ ,  $\Sigma^{(0)}$  y  $P^{(0)}$ ).
3. Calcular la probabilidad posterior para cada observación (Paso E), tal como se muestra en (7) y (8).

$$\hat{\tau}_{ij} = \frac{\hat{p}_i \phi(\mathbf{x}_j; \hat{\mu}_i, \hat{\Sigma}_i)}{\hat{f}(\mathbf{x}_j)} \quad (7)$$

$$\hat{f}(\mathbf{x}_j) = \sum_{k=1}^c \hat{p}_k \phi(\mathbf{x}_j; \hat{\mu}_k, \hat{\Sigma}_k) \quad (8)$$

Donde  $\hat{\tau}_{ij}$  representa la probabilidad posterior del punto  $\mathbf{x}_j$  correspondiente al i-esimo termino,  $\phi(\mathbf{x}_j; \hat{\mu}_i, \hat{\Sigma}_i)$  es la densidad normal multivariable y  $\hat{f}(\mathbf{x}_j)$  corresponde la MD estimada para el i-esimo termino evaluado en  $\mathbf{x}_j$ .

4. Actualizar  $\hat{\mu}$ ,  $\hat{\Sigma}$  y  $P$  de cada componente (Paso M), utilizando (9), (10) y (11). Donde  $\hat{p}_i$ ,  $\hat{\mu}_i$  y  $\hat{\Sigma}_i$  son las estimaciones actualizadas de los parámetros.

$$\hat{p}_i = \frac{1}{n} \sum_{j=1}^n \hat{\tau}_{ij} \quad (9)$$

$$\hat{\mu}_i = \frac{1}{n} \sum_{j=1}^n \frac{\hat{\tau}_{ij} \mathbf{x}_j}{\hat{p}_i} \quad (10)$$

$$\hat{\Sigma}_i = \frac{1}{n} \sum_{j=1}^n \frac{\hat{\tau}_{ij} (\mathbf{x}_j - \hat{\mu}_i)(\mathbf{x}_j - \hat{\mu}_i)^T}{\hat{p}_i} \quad (11)$$

5. Repetir los pasos 3 y 4 hasta obtener la convergencia deseada.

Posteriormente se realiza la organización de los grupos en clases, basadas en la probabilidad de aparición en cada

grupo, dada por el modelo de la mezcla.

$$\hat{f}_{FM}(\mathbf{x}) = \sum_{i=1}^c p_i \phi_i(\mathbf{x}; \hat{\mu}_i, \hat{\Sigma}_i) \quad (12)$$

### EJEMPLO DE APLICACIÓN

Con la construcción del modelo estadístico se pretende localizar la falla dentro del sistema según la respuesta del mismo durante la perturbación. Las señales de tensión y corrientes obtenidas son la base para hallar la solución. Cada falla registrada viene acompañada de información relevante a la hora de clasificar los datos dentro de una clase establecida en el modelo (zona dentro del sistema de distribución). La información utilizada en el estudio corresponde a magnitudes (no ángulos) de tensión y corriente, puesto que se quiere realizar una herramienta de bajo costo que no implique la utilización de equipos de última tecnología. Muchos de los equipos de medida que poseen las empresas actualmente únicamente manejan las magnitudes de las señales. El sistema prototipo escogido se muestra en **Figura 1**. [8], [23].

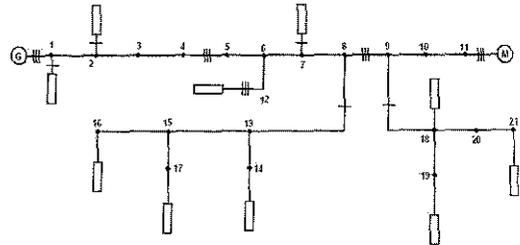


Figura 1. Sistema de distribución de prototipo

Para la discriminación de las zonas se tuvo presente los grupos visualmente apreciables en el análisis preliminar de datos. El propósito de asociar los grupos a las zonas, es establecer la correspondencia de la ubicación de la falla en el sistema con la clasificación de los datos dentro de los grupos. Para representar la información del sistema se utilizaron tres descriptores los cuales corresponden a la magnitud del hueco de tensión en cada fase cuando la falla ocurre [22]. Una vez los valores de los descriptores han sido calculados, se determina si estos deben ser transformados mediante logaritmos con el fin de mejorar las distribuciones de las observaciones sin afectar la variabilidad de los datos [19]. Para este caso, los datos de entrenamiento y validación son simulaciones de fallas realizadas en cada barra del sistema con diferentes resistencias de falla. Los valores de resistencia de falla utilizados en el estudio fueron tomados a partir del análisis realizado en [8] [23], [24]. Donde se toma un modelo de resistencia de falla que tiene presente los efectos de la resistencia del arco eléctrico y el de la trayectoria a tierra. En **Tabla 1** se presentan los valores de resistencia de falla utilizados para los procesos de entrenamiento (E) o validación (V).

Tabla 1. Valores de resistencia de falla

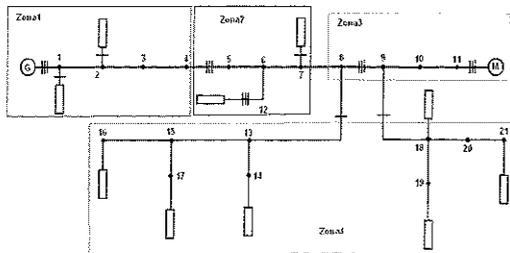
Resistencia de falla [ $\bullet$ ]	Proceso	Resistencia de falla [ $\bullet$ ]	Proceso
0.05	E	25	E
5	E	30	V
10	V	35	E
15	E	40	V
20	V	50	E

En **Tabla 2** se contabilizan el número de simulaciones realizadas para el entrenamiento (E) como para la validación (V), organizadas según el tipo de falla.

Tabla 2. Resultados de simulación realizada

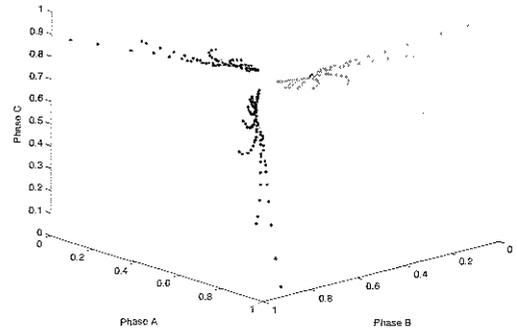
Tipo de Falla	Simulaciones por fase		Proceso		
	E	V	E	V	
Monofásica	132	187	176	315	180
Bifásica	132	132	132	252	144
Bifásica a tierra	132	132	132	252	144
Trifásica	132		84	48	
Trifásica a tierra	132		84	48	
<b>Total</b>	<b>1551</b>		<b>1323</b>	<b>228</b>	

Antes de iniciar la clasificación se realiza un ordenamiento manual de los datos de entrenamiento con el fin de identificar y establecer previamente zonas dentro del sistema de distribución. El objetivo es determinar la influencia que tiene la organización previa de los datos sobre los algoritmos a utilizar. A continuación en **Figura 2** se muestra cada una de las zonas del sistema obtenidas mediante la agrupación previa de los descriptores de tensión para fallas monofásicas.

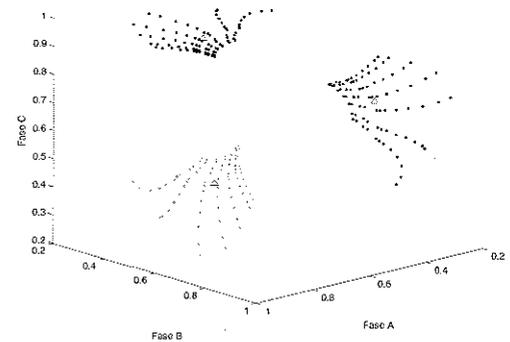


**Figura 2.** Determinación previa de las zonas a través del análisis visual de los datos.

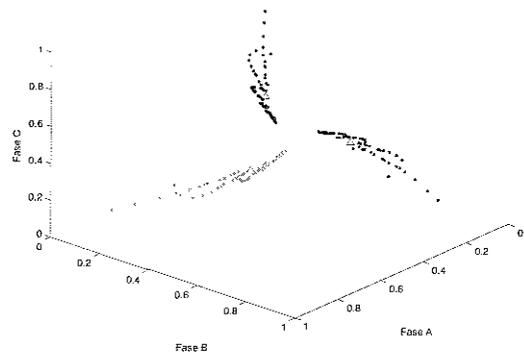
El modelo propuesto está estructurado por diferentes etapas de clasificación, la primera (EC1), determina la fase fallada, la segunda (EC2), determina el valor de resistencia de falla y la tercera (EC3), determina la zona del evento. La etapa EC1 aplica únicamente a las fallas monofásicas y bifásicas mientras que la EC2 y EC3 aplican a todas. En las **Figuras 3, 4 y 5** se muestran la distribución de los datos de entrenamiento correspondientes a tres tipos de falla, además se observan tres grupos claramente definidos, los cuales, corresponden a cada una de las fases.



**Figura 3.** Distribución de los datos de entrenamiento falla monofásica (EC1)



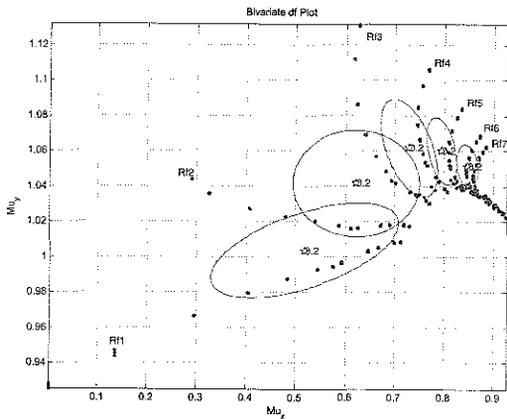
**Figura 4.** Distribución de los datos de entrenamiento falla bifásica (EC1)



**Figura 5.** Distribución de los datos de entrenamiento falla bifásica a tierra (EC1)

A partir de la definición a priori de los grupos y con la aplicación del algoritmo k-medias se estiman los centros de los grupos con los cuales se inicia el análisis. Los centros se someten al criterio del analista para su escogencia con el fin de obtener los estimados más convenientes. Posteriormente, se define la forma y la proporción final en que los grupos se encuentran dentro de la distribución, utilizando el algoritmo EM y los estimados iniciales.

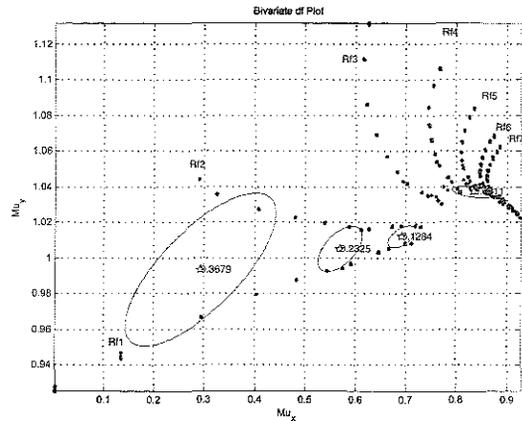
En esta etapa, se utiliza un modelo heterocedástico, el cual permite determinar libremente la forma de las matrices de covarianzas para cada distribución y por ende la forma de cada grupo. Adicionalmente, se utiliza el mismo valor inicial para los coeficientes de mezclado, con lo cual se supone igualmente probable la clasificación de los datos en cada uno de los grupos. En la etapa EC2, La información acerca del valor de resistencia de falla permite crear un posible escenario donde se conforman grupos que representan rangos de valores de resistencias de falla. En este caso se establecieron cinco grupos representativos, tal y como se muestra en la **Figura 6**. Lo anterior hace relación con la capacidad de predecir la causa de la falla. Esta clase de información puede ser relevante a la hora de determinar el tipo de procedimiento a seguir para solucionar el problema [25].



**Figura 6.** Distribución de los grupos en los datos de entrenamiento falla monofásica (EC2)

En la **Figura 6**, se muestra la forma y el tamaño final de los cinco grupos obtenidos (elipses). Los centros y los coeficientes de mezclado son representados por las estrellas. El valor de resistencia de falla crece, conforme se avanza hacia la derecha. La forma elíptica de cada grupo, está relacionada con los valores de la matriz de covarianza [14]. En la etapa EC3 los datos son agrupados, cada grupo representa una zona probable de la ocurrencia de la falla. Cada zona definida, está asociada a un número determinado de barras dentro del sistema de distribución, tal y como se observa en la **Figura 2**. Estas zonas son establecidas con ayuda de la información acerca del número de grupos obtenido en la etapa EC2. Por lo tanto, en la etapa EC3 se tienen " " grupos correspondientes a los grupos de resistencia de falla determinados en EC2 y estos a su vez contienen " " grupos correspondientes a cada una de las zonas del sistema. En el caso del sistema prototipo, existe una MD compuesta por cinco grupos (EC2) y cada grupo

está compuesto por cuatro subgrupos que representan las zonas del sistema. En la **Figura 7**, se ilustra la construcción de los grupos asociados a las zonas. Es de notar que los parámetros de cada grupo son mostrados en la grafica.



**Figura 7.** Distribución de las zonas asociadas al grupo 1 de la etapa EC2

El algoritmo EM tiene la libertad de operar los elementos de la matriz de covarianza en cada iteración hasta conseguir la mejor estimación. En este caso, si las matrices de covarianza permanecieran sin cambio en relación con su estimación inicial (matrices identidad) se obtendrían grupos de formas circulares [17]. Finalmente, se aplica el modelo de mezcla, con el fin de comparar las zonas supuestas inicialmente (**Figura 2**) con las determinadas por el algoritmo. En este caso de aplicación, la aproximación final de las zonas corresponde al resultado obtenido luego de la aplicación de los algoritmos. Se destaca, la influencia de una adecuada aproximación inicial de las zonas por parte del analista de los datos, permitiendo mejorar el ajuste del modelo de los datos durante su construcción.

Establecido el modelo de los datos, se realizó la validación con datos desconocidos para el mismo. Las observaciones poseen datos con valores de resistencias de falla asociadas de 10, 20, 30 y 40  $\Omega$ . Para evaluar la eficiencia de la primera etapa de clasificación (EC1), se utiliza un conjunto de 468 observaciones. En la **Tabla 3** se muestra el número de datos pertenecientes a fallas en las tres fases. Como se observa en **Tabla 3**, el modelo clasificó correctamente la totalidad de las observaciones correctamente

**Tabla 3.** Resultados de validación Etapa EC1

Tipo de falla		Fase A	Fase B	Fase C
Monofásica	Datos	48	68	64
	Aciertos	48	68	64
Bifásica	Datos	48	48	48
	Aciertos	48	48	48
Bifásica a tierra	Datos	48	48	48
	Aciertos	48	48	48

En la **Tabla 4** se presentan los resultados obtenidos en la validación de los datos (564 observaciones) mediante el clasificador EC2.

**Tabla 4** Resultados de Validación Etapa EC2

Grupo		1	2	3	4	5
Monofásica	Datos	0	45	45	45	45
	Aciertos	0	31	30	37	42
Bifásica	Datos	0	36	36	36	36
	Aciertos	0	30	30	36	36
Bifásica a tierra	Datos	0	36	36	36	36
	Aciertos	0	26	30	36	36
Trifásica	Datos	0	24	24	24	24
	Aciertos	0	20	24	24	24

Los resultados presentados en la **Tabla 5**, reflejan un buen desempeño del clasificador EC3 de acuerdo al porcentaje de aciertos en la clasificación de los datos en las zonas definidas (564 observaciones).

**Tabla 5** Resultados de Validación Etapa EC3

Grupo		1	2	3	4
Monofásica	Datos	48	48	48	36
	Aciertos	48	48	48	19
Bifásica	Datos	48	48	48	0
	Aciertos	48	48	48	0
Bifásica a tierra	Datos	48	48	48	0
	Aciertos	48	48	48	0
Trifásica	Datos	32	32	32	0
	Aciertos	32	32	32	0

Los problemas de clasificación, se presentan en los puntos más alejados del alimentador. En estos casos, existe una mala asignación de la zona debido a un error de clasificación en la etapa previa (EC2). Otra posible causa es la localización de los datos en la intersección de dos o más grupos. En este caso, la zona es asignada es la que posee la mayor probabilidad a posteriori. Cada observación es evaluada respecto a cada uno de los grupos calculando la probabilidad de pertenecer a los mismos en todas las etapas. La observación se clasifica dentro del grupo en el cual presente la probabilidad más alta.

## CONCLUSIONES

Se presenta una novedosa alternativa de solución al problema de localización de fallas en los sistemas de distribución, la cual, está compuesta por una herramienta estadísticamente soportada, basada en la extracción y modelado de la información asociada a fallas. Esta

herramienta deja al descubierto las potenciales relaciones existentes entre los datos.

Este trabajo contribuye a satisfacer la necesidad actual de las empresas distribuidoras por mantener y mejorar la calidad del servicio, promoviendo la concepción de estrategias orientadas a la disminución de la duración y el número de las interrupciones. Respecto a la creación de estrategias, se está trabajando en una propuesta metodológica basada en el fraccionamiento del tiempo de indisponibilidad debido a una falla, en tiempos asociados a las acciones realizadas durante la atención de la misma y posterior recuperación del sistema.

Una ventaja en la construcción del modelo es la determinación de grupos de características bien definidos con los cuales se optimiza la clasificación de los datos, asegurando una buena aproximación por parte del modelo. Adicionalmente, la aproximación inicial de las zonas en el sistema resulta útil a la hora de estimar el número real y tamaño de las zonas dentro de sistema de distribución. El desempeño del modelo depende de la calidad en la obtención y procesamiento de la información (sistemas de información confiables), además de la caracterización detallada del sistema al cual se aplica.

Los requerimientos de software para la implementación de la herramienta propuesta son suministrados por paquetes estadísticos de uso comercial y de bajo costo, lo cual permite establecer la sencillez y facilidad en la implementación del modelo. Los requerimientos de hardware son básicos y económicos, ya que depende únicamente de un sistema de adquisición de datos que esté en capacidad de monitorear las señales de tensión y corriente en la cabecera del circuito.

El modelo propuesto contempla las implicaciones, técnicas, económicas y operativas existentes en los sistemas de distribución, destacando la mínima inversión en la implementación de la herramienta por parte de las empresas. Existen limitaciones potenciales para el modelo propuesto, como lo son, la selección de la cantidad de grupos a conformar, la proporción de los grupos en la muestra y los valores de inicio de los algoritmos utilizados. Para superar estas limitaciones se utilizaron criterios teóricos y heurísticos, siendo los últimos los de mayor influencia en la estructura del modelo ya que hacen parte del conocimiento a priori que posee el analista sobre la información.

Se está trabajando en la evaluación del desempeño del modelo con información obtenida a partir de sistemas de distribución reales. Además, se está explorando la opción de incluir mayor cantidad de descriptores que relacionen adecuadamente la información monitoreada con la ubicación de la falla. Otros trabajos van encaminados a la construcción de un modelo que mezcla las mejores características de un método algorítmico con el modelo estadístico.

## BIBLIOGRAFÍA

- [1] Comisión de Regulación de Energía y Gas CREG, Resoluciones CREG 070 de 1998, CREG 096 de 2000, CREG 084 de 2002, CREG 084 de 2002.
- [2] J. Jaimes, J. Duarte, G. Ordóñez, G. Carrillo, P. Gaona, "La calidad de la energía eléctrica en el sector eléctrico colombiano: utopía o realidad", II Congreso Internacional de la Región Andina, Andescon IEEE, Colombia, 2004.
- [3] J. Cormane, V. Barrera, G. Ordóñez, "Predicción de corrientes armónicas en ventanas de carga residenciales mediante modelos neuronales artificiales". II Congreso Internacional de la Región Andina, Andescon IEEE, Colombia, 2004.
- [4] G. Carrillo, G. Ordóñez, "Calidad del Suministro de Energía Eléctrica (Power Quality)" ALURE Program - CREG Project. Tutorial Course, Colombia - 2002.
- [5] M. Bollen, Understanding power quality problems: voltages sags and interruptions, New York, IEEE Press, 2000.
- [6] A. Peñuela, J. Pacheco, G. Ordóñez, J. Duarte, "Propuesta metodología para la valoración de costos en empresas manufactureras, debidos a una inadecuada continuidad del suministro de la energía eléctrica", tesis de grado, Universidad Industrial de Santander, Bucaramanga., Colombia, 2003.
- [7] J. Jaimes, "Conceptualización técnica del servicio de reserva operativa", M.S. thesis, Universidad Industrial de Santander, Bucaramanga., Colombia, 2003.
- [8] R. Das, "Determining the Locations of Faults in Distribution Systems", Ph.D dissertation, Saskatchewan Univ., Canada, 1998.
- [9] J. Zhu, D. Lubkeman, A. Girgis, "Automated fault location and diagnosis on electric power distribution feeders", IEEE Trans. Power delivery, Vol. 12, pp. 801-809, Apr, 1997.
- [10] A. Girgis, C. Fallon, D. Lubkeman, "A fault location technique for rural distribution feeders", IEEE Trans. Industry applications, Vol. 29, pp. 1170-1175, Dec, 1993.
- [11] Comisión de Regulación de Energía y Gas CREG, Resolución CREG 025 de 2005.
- [12] J. Mora, J. Colomer, "Voltage Sag Characterization and Classification for Diagnosis in Electric Power Quality Domain", Department of Electronics, Computer Science and Automatic Control, Research report, Girona España, 2003.
- [13] D. Peña, Análisis de datos multivariantes. Madrid, McGraw-Hill, 2002.
- [14] Y. Wang, L. Luo, M. Freedman and S. Kung, "Probabilistic principal component subspaces: A hierarchical finite mixture model for data visualization", IEEE Trans. Neural Networks, Vol 11, pp. 625-636, May, 2000.
- [15] L. Díaz, Estadística multivariada: inferencia y métodos. Bogota D.C, Universidad Nacional de Colombia, 2002.
- [16] J. Hair, R. Anderson, R. Tatham, W. Black, Análisis multivariante. Madrid, Prentice Hall, 1999.
- [17] Y. Wang, S. Lin, H. Li, and S.Kung, "Data mapping by probabilistic modular networks and information theoretic criteria", IEEE Trans. Signal processing, Vol. 46, pp. 3378-3397, Dec, 1998.
- [18] M. Jordan and R. Jacobs, "Hierarchical mixture of experts and the EM algorithm", IEEE Trans. Neural Computat, Vol 6, pp. 181-214, 1994.
- [19] W. Martínez, and A Martínez Computational statistics Handbook whit MatLab. New York, Chapman & Hall, 2002.
- [20] E. Dalla, Jonson, Métodos multivariados aplicados al análisis de datos, México, Thompson, 2000.
- [21] G. Mclachlan and D. Peel, Finite mixture models, Canada, Wiley, 2000.
- [22] M. Figueiredo, "Unsupervised Learning of Finite Mixture Models", IEEE Trans. Pattern analysis and Machine intelligence, Vol. 24, Mar, 2002.
- [23] J.B. Dagenhart, "The 40-Ground-Fault Phenomenon" IEEE Trans. on Industry Applications, vol. 36, no. 1, pp 30-32, 2000
- [24] S. Lee, M. Choi, S. Kang, B. Jin, D. Lee, B. Ahn, N. Yoon, H. Kim, S. Wee. "An intelligent and efficient fault location and diagnosis écheme for radial distribution systems", IEEE Trans. Power Delivery, Vol. 19, Apr, pp. 524-532, 2004.
- [25] G. Carrillo, J. Román, J. Rivier, S. Ramírez "Metodología coste-beneficio aplicada a la instalación de equipos de localización y seccionamiento del tramo con avería en redes de distribución", Jornadas hispano-lusas de ingeniería eléctrica, Tomo 3, Barcelona, 1993.

**AUTORES**

**Jorge Andrés Cormane Angarita.** Ingeniero Electricista UIS, Magíster en Ingeniería Eléctrica UIS. Investigador GISEL. Áreas de Trabajo: Calidad del Suministro de Energía, Inteligencia artificial aplicada a Calidad del Suministro de Energía. Bucaramanga, Colombia.



**Gabriel Ordóñez Plata.** Ingeniero Electricista UIS, Doctor Ingeniero Industrial UPCO, Madrid, España. Profesor Titular UIS. Director del Grupo de Investigación en Sistemas de Energía Eléctrica GISEL-UIS. Áreas de trabajo: Calidad del Suministro de Energía, Procesamiento Digital de Señales Eléctricas. Bucaramanga, Colombia.



**Hermann Raúl Vargas Torres.** Ingeniero Electricista UIS, Magíster en Potencia Eléctrica UIS, Doctor Ingeniero Industrial UPCO, Madrid, España. Profesor Asociado UIS. Investigador del Grupo de Investigación en Sistemas de Energía Eléctrica GISEL-UIS. Áreas de trabajo: Calidad del Suministro de Energía, Estabilidad en Sistemas de Potencia, Mercados de Energía Eléctrica. Bucaramanga, Colombia.

