

**LOCALIZACIÓN DE FALLAS: RECONOCIMIENTO ESTADÍSTICO
DE PATRONES**

MARCO FIDEL SUÁREZ SÁNCHEZ
EDWIN SALAMANCA TORRES

ESCUELA DE INGENIERÍAS ELÉCTRICA, ELECTRÓNICA Y
TELECOMUNICACIONES
UNIVERSIDAD INDUSTRIAL DE SANTANDER
BUCARAMANGA

2006

**LOCALIZACIÓN DE FALLAS: RECONOCIMIENTO ESTADÍSTICO
DE PATRONES**

MARCO FIDEL SUÁREZ SÁNCHEZ
EDWIN SALAMANCA TORRES

Trabajo de grado para optar al título de Ingeniero Electrícista

Director
HERMANN RAÚL VARGAS TORRES
Doctor Ingeniero Electrícista

Codirector
JORGE ANDRÉS CORMANE
Magister(c) Ingeniero Electrícista

ESCUELA DE INGENIERÍAS ELÉCTRICA, ELECTRÓNICA Y
TELECOMUNICACIONES
UNIVERSIDAD INDUSTRIAL DE SANTANDER
BUCARAMANGA

2006

"La diferencia entre un escritor y alguien que no lo es. Es que al primero le cuesta mucho trabajo escribir."
Thomas Mann (1875-1955)

DEDICATORIA

*A mi madre y mi hermana gestadoras de todos mis logros.
A Maria Juliana gracias por su apoyo, paciencia y compañía.
Al mejor cluster de personas que pude encontrar (Roberto, Claudia, León, Silvia,
Christian, Toño, Erlin y Gladis).*

AGRADECIMIENTOS

Al PhD. Hermann Raúl Vargas, director del proyecto, por su orientación y asesoría las cuales hicieron posible la realización de este trabajo de grado.

Al Msc. Jorge Andrés Cormane, codirector del proyecto, por su constante colaboración y apoyo en el desarrollo de este proyecto.

Al Msc. Victor Barrera, por valoración que le dió a nuestro trabajo y por sus aportes para mejorarlo.

Al Los Ingenieros Alvaro Gómez Ruíz y Germán Morales España, por prestarnos su ayuda en momentos críticos.

A la Ingeniera Claudia Liliana Pacheco por su colaboración y aportes.

Ing. Marco Fidel Suárez Sánchez

DEDICATORIA

Este trabajo de grado está dedicado de una manera muy especial a las personas que me han acompañado durante todos estos años , mis padres. Que con su sacrificio y dedicación forjaron al hombre que soy.

Además a todos mis amigos y compañeros que creyeron en mi, y estuvieron firmes a mi lado.

AGRADECIMIENTOS

Doy gracias en primera medida a Dios todo poderoso por darme a unos padres excepcionales que me educaron de la mejor forma, a los profesores que con sus enseñanzas me han convertido en profesional, a mis amigos que siempre estuvieron allí para cambiar la rutina y volver más amena la vida universitaria.

*A Bienestar Universitario en especial al programa **PAMRA** por el apoyo recibido y la experiencia que me ofrecieron. A los patrocinadores de nuestro proyecto.*

Ing. Edwin Salamanca Torres

RESUMEN

TÍTULO:

LOCALIZACIÓN DE FALLAS: RECONOCIMIENTO ESTADÍSTICO DE PATRONES*

AUTORES:

MARCO FIDEL SUÁREZ SÁCHEZ

EDWIN SALAMANCA TORRES**

PALABRAS CLAVE:

Correlación, Componentes principales (ACP), conglomerados, localización de fallas, sistemas de distribución, Redes de secuencia.

DESCRIPCIÓN:

Mediante la aplicación de la estadística multivariada al campo de la energía eléctrica se puede dar solución al mejoramiento de los índices de calidad localizando oportunamente las fallas de cortocircuito en los sistemas de distribución. Esto se realizó mediante la simulación de fallas en un sistema prototipo con diferentes valores de resistencias de falla a fin de caracterizarlo. Las variables que intervienen en este análisis dependen de los medidores que existan en la cabecera del circuito, razón por la cual se plantean diferentes alternativas con el ánimo de dar solución a las diferentes empresas prestadoras del servicio buscando la mejor forma de adaptarse a la información histórica que estas posean sobre la ocurrencia de fallas. Para cada uno de los modelos propuestos se realiza un análisis de correlación seguido de un análisis de componentes principales para reducir la dimensión del problema y poderlo visualizar, lo cual facilita el descubrimiento de los patrones por medio de un análisis de conglomerados. Con este último se zonifica el circuito según las secciones con similitudes entre sí. Estas zonas claramente identificadas serán la base de la localización de una falla al capturar las señales de tensión y corriente ante un evento y procesarlas acorde con los análisis realizados anteriormente. Al ubicar este punto en el espacio de trabajo, se busca a cual grupo corresponde y de esta manera determinar la zona afectada. Al implementar esta metodología en un sistema real se tiene la limitante de que los datos de falla son escasos, motivo por el cual se recurre a las simulaciones tratando de acercarse a la realidad.

*Proyecto de Grado

**Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones.
Hermann Raúl Vargas Torres.

ABSTRACT

TITLE:

LOCALIZATION OF FAULTS: STATISTICAL RECOGNITION OF PATTERNS*

AUTHORS:

MARCO FIDEL SUÁREZ SÁCHEZ

EDWIN SALAMANCA TORRES**

KEY WORDS:

Correlation, Principal Components (ACP), Cluster, localization of faults, distribution systems, sequence Nets.

DESCRIPTION:

By the applications of the multivariate statistic into the field of electrical energy it is possible to give solution to the improvement of the quality indicators, by locating opportunely the faults of the short circuit in distribution systems. This was made by the simulation of faults in a prototype system with different values from faults resistances in order of characterizing it. The variables, that take part in this analysis, depend on the measurers that exist in the head of the circuit. Therefore, it has considered different alternatives with the spirit to give to solution to the different lending service companies, from the service, looking for the best form to adapt the historical information which these have on the occurrence of faults. For each one of the proposed models, it will realize an correlation analysis followed by an analysis of main components in order to reduce the problem dimension and to be able to visualize it, which facilitates the discovery of the patrons by the analysis of conglomerates. With this last one the circuit is zoned according to the sections with similarities to each other. These zones, identified clearly, will be the base of the location of a fault when capturing tension and current signals in front of an event and processing them agreed with the analyses previously made. When locating this point in the work space, it looks for the corresponding group in order to determine the affected zone. When implementing this methodology in a real system, it is limited by the fact that the fault data are few, reason by which it resorts to the simulations treating to approach the reality.

*Degree Project

**School of Electrical Engineering. Director PhD. Hermann Raúl Vargas Torres.

Índice general

| | |
|--|----------|
| 1. INTRODUCCIÓN | 1 |
| 1.1. Objetivos | 2 |
| 1.2. Estructura del documento | 2 |
| 2. ANÁLISIS ESTADÍSTICO MULTIVARIADO | 4 |
| 2.1. Introducción | 4 |
| 2.2. Definición | 4 |
| 2.3. Algunos conceptos básicos del análisis estadístico multivariado | 5 |
| 2.3.1. Los datos | 5 |
| 2.3.2. El valor teórico | 5 |
| 2.3.3. Escala de medida | 5 |
| 2.3.4. Vector de medias. \bar{x} | 6 |
| 2.3.5. La matriz de varianzas y covarianzas S | 6 |
| 2.3.6. La varianza total y la varianza media | 7 |
| 2.3.7. La matriz de correlación R | 7 |
| 2.3.8. La distancia | 8 |
| 2.3.8.1. Distancia euclídeana | 8 |
| 2.3.8.2. Distancia de mahalanobis | 9 |
| 2.3.8.3. Otras distancias | 9 |
| 2.3.9. Medidas de similaridad y de disimilitud. | 10 |
| 2.3.9.1. Las medidas de proximidad | 10 |
| 2.3.9.2. Las medidas de disimilitud | 11 |
| 2.4. Clasificación de las técnicas multivariadas. | 11 |
| 2.4.1. Métodos de dependencia | 12 |
| 2.4.2. Métodos de interdependencia | 12 |
| 2.4.3. Modelos estructurales | 12 |
| 2.5. Supuestos del análisis multivariado | 13 |

| | | |
|----------|--|----|
| 2.5.1. | Normalidad | 13 |
| 2.5.2. | Homocedasticidad | 13 |
| 2.5.3. | Linealidad | 14 |
| 2.6. | Análisis de componentes principales (ACP) | 14 |
| 2.6.1. | Ventajas del ACP | 15 |
| 2.6.2. | Cálculo de las componentes principales | 15 |
| 2.6.3. | Propiedades de las componentes principales (CP) | 16 |
| 2.6.4. | Selección del número de componentes principales | 17 |
| 2.6.5. | Interpretación de cada componente principal | 17 |
| 2.6.6. | Información de la última componente principal. | 17 |
| 2.6.7. | Ejemplo de aplicación de ACP | 18 |
| 2.7. | Reconocimiento de patrones o análisis de conglomerados | 21 |
| 2.7.1. | Objetivos del reconocimiento de patrones | 22 |
| 2.7.2. | Clasificación de los métodos de reconocimiento de patrones | 22 |
| 2.7.2.1. | Métodos jerárquicos | 22 |
| 2.7.2.2. | Métodos de partición | 25 |
| 2.7.3. | Impacto del escalamiento | 26 |
| 2.7.4. | Ejemplo de aplicación del reconocimiento de patrones. | 27 |
| 2.8. | Estado del Arte | 28 |

3. RECONOCIMIENTO ESTADÍSTICO DE PATRONES: LOCALIZACIÓN DE FALLAS 31

| | | |
|----------|--|----|
| 3.1. | Introducción | 31 |
| 3.2. | Planteamiento del problema | 31 |
| 3.2.1. | Antecedentes históricos de los métodos utilizados para la localización de fallas | 32 |
| 3.3. | Metodología propuesta | 32 |
| 3.3.1. | Selección del sistema de potencia de prueba | 34 |
| 3.3.2. | Simulación del sistema de potencia en condiciones de falla | 34 |
| 3.3.3. | Obtención de los descriptores de tipo determinista. | 35 |
| 3.3.3.1. | Modelo tensión-corriente | 36 |
| 3.3.3.2. | Modelo solo corriente | 37 |
| 3.3.3.3. | Modelo solo tensión | 38 |
| 3.3.4. | Cálculo de descriptores de tipo estadístico | 38 |
| 3.3.4.1. | Modelo tensión-corriente | 38 |
| 3.3.4.2. | Modelo corriente | 40 |
| 3.3.4.3. | Modelo de tensión | 42 |

| | | |
|-----------|---|-----------|
| 3.3.5. | Reconocimiento de patrones | 44 |
| 3.3.5.1. | Modelo tensión-corriente | 45 |
| 3.3.5.2. | Modelo corriente | 46 |
| 3.3.5.3. | Modelo tensión | 46 |
| 4. | PRUEBAS Y RESULTADOS | 53 |
| 4.1. | Introducción | 53 |
| 4.2. | Simulación del sistema de distribución de prueba en condiciones de falla | 53 |
| 4.3. | Zonificación del sistema de distribución de prueba | 54 |
| 4.4. | Validación de los modelos propuestos | 55 |
| 4.4.1. | Identificación del tipo de falla | 56 |
| 4.4.2. | Localización de la zona fallada con datos de entrenamiento | 57 |
| 4.4.3. | Localización de la zona fallada con datos de validación | 60 |
| 4.4.4. | Herramienta empleada para realizar la validación de los modelos y las zonas propuestas (<i>validación</i>) | 63 |
| 5. | CONCLUSIONES APORTES Y TRABAJOS FUTUROS | 67 |
| 5.1. | Conclusiones y observaciones | 67 |
| 5.2. | Aportes | 68 |
| 5.3. | Trabajos futuros | 68 |
| A. | Cálculo de componentes principales | 69 |
| B. | Características del sistema de distribución de prueba | 71 |
| C. | Fallas de cortocircuito | 76 |
| C.1. | Introducción | 76 |
| C.2. | Falla Simétrica | 77 |
| C.2.1. | Falla Trifásica (L-L-L) | 77 |
| C.3. | Fallas asimétricas | 79 |
| C.3.1. | Fallas monofásicas de línea a tierra L-T | 79 |
| C.3.2. | Fallas bifásicas o línea a línea L-L | 80 |
| C.3.3. | Fallas bifásicas tierra o doble línea tierra L-L-T | 80 |
| C.4. | Comportamiento de las redes de secuencia cuando la falla es vista desde el alimentador | 82 |
| D. | Componentes principales para cada modelo propuesto | 88 |
| D.1. | Modelo tensión-corriente | 88 |

| | |
|---|-----------|
| D.1.1. Descripción de las componentes: | 88 |
| D.2. Modelo de corriente | 89 |
| D.2.1. Descripción de las componentes: | 89 |
| D.3. Modelo de tensión | 89 |
| D.3.1. Descripción de las componentes: | 89 |
| E. PRtool Herramienta para el reconocimiento estadístico de patrones | 94 |
| E.1. Introducción | 94 |
| E.2. Como utilizar PRtool | 94 |
| E.2.1. Instalación | 94 |
| E.2.2. Utilizando PRtool | 96 |
| E.3. Help <i>PRtool</i> | 99 |
| E.4. Funciones creadas para <i>PRtool</i> | 100 |
| E.4.1. <i>guardar.m</i> | 100 |
| E.4.2. <i>plotcoref.m</i> | 100 |
| E.4.3. <i>grupos.m</i> | 101 |
| E.4.4. <i>reporte.m</i> | 101 |
| E.4.5. <i>barra.m</i> | 101 |
| E.4.6. <i>fasor2.m</i> | 101 |
| E.4.7. <i>abrir.m</i> | 101 |
| E.4.8. <i>dato.m</i> | 102 |
| E.4.9. <i>datom.m</i> | 102 |
| E.4.10. <i>iden.m</i> | 102 |
| E.4.11. <i>importar.m</i> | 102 |
| E.4.12. <i>pricom.m</i> | 102 |
| E.4.13. <i>locah.m</i> | 103 |
| E.5. Ejemplo <i>PRtool</i> | 103 |

Lista de figuras

| | |
|---|----|
| 2.1. (a)Gráfico de dispersión con relacion de homocedasticidad. (b)Gráfico de dispersión con relacion de heterocedasticidad. | 14 |
| 2.2. Gráfico de correlacion y distribución de las variables | 18 |
| 2.3. Porcentaje de información de cada CP | 20 |
| 2.4. Primer plano factorial ejemplo ACP | 20 |
| 2.5. Dendrograma utilizando distancia euclidea y enlace de Ward | 27 |
| 2.6. Reconocimiento de patrones sobre el primer plano factorial. | 28 |
| 3.1. Metodología propuesta. | 34 |
| 3.2. (a)Mínima tensión de fase. (b)Máximo valor de corriente de falla. | 37 |
| 3.3. Correlación modelo Tensión-Corriente. (a)L-T. (b)L-L. (c)L-L-T. (d)L-L-L. | 39 |
| 3.4. Porcentaje de información que contiene cada componente principal modelo Tensión-Corriente. (a)L-T. (b)L-L. (c)L-L-T. (d)L-L-L. | 40 |
| 3.5. Correlación modelo corriente. (a)L-T. (b)L-L. (c)L-L-T. (d)L-L-L. | 41 |
| 3.6. Porcentaje de información que contiene cada componente principal modelo corriente. (a)L-T. (b)L-L. (c)L-L-T. (d)L-L-L. | 42 |
| 3.7. Correlación modelo tensión. (a)L-T. (b)L-L. (c)L-L-T. (d)L-L-L. | 43 |
| 3.8. Porcentaje de información que contiene cada componente principal modelo tensión. (a)L-T. (b)L-L. (c)L-L-T. (d)L-L-L. | 44 |
| 3.9. Grupos generados para el modelo Tensión-Corriente. (a)L-T. (b)L-L. | 47 |
| 3.10. Grupos generados para el modelo Tensión-Corriente. (a)L-L-T. (b)L-L-L. | 48 |
| 3.11. Grupos generados para el modelo Corriente. (a)L-T. (b)L-L. | 49 |
| 3.12. Grupos generados para el modelo Corriente. (a)L-L-T. (b)L-L-L. | 50 |
| 3.13. Grupos generados para el modelo Tensión. (a)L-T. (b)L-L. | 51 |
| 3.14. Grupos generados para el modelo Tensión. (a)L-L-T. (b)L-L-L. | 52 |
| 4.1. Zonificación en el agrupamiento de la falla L-L del modelo Tensión-corriente | 55 |
| 4.2. Zonificación sistema de distribución de prueba | 56 |

| | |
|--|----|
| 4.3. Identificación del tipo de falla. | 58 |
| 4.4. Porcentaje de acierto de validación con datos de entrenamiento. (a)Datos de prueba. (b)Datos completos de entrenamiento. | 60 |
| 4.5. Porcentaje de acierto de localización de fallas con datos de validación. (a)Datos de prueba de validación. (b)Datos completos de validación. | 62 |
| 4.6. Comportamiento General. (a)Modelo Tensión-Corriente. (b)Modelo corriente. (c) Modelo tensión. | 65 |
| 4.7. Herramienta de Validación | 66 |
| | |
| B.1. Sistema de Potencia de Prueba | 71 |
| B.2. Fasores de tensión y secuencia cero en operación normal. | 74 |
| B.3. Fasores de corriente y secuencia cero en operación normal. | 74 |
| | |
| C.1. Falla trifásica | 77 |
| C.2. Equivalente simetrico falla trifásica | 77 |
| C.3. Señal de tensión bajo una falla trifásica | 78 |
| C.4. Señal de corriente bajo una falla trifásica | 78 |
| C.5. Fasores de tensión y secuencia cero bajo una falla trifásica | 79 |
| C.6. Fasores de corriente y secuencia cero bajo una falla trifásica | 79 |
| C.7. Falla monofásica | 80 |
| C.8. Equivalente simétrico falla monofásica | 80 |
| C.9. Señal de tensión bajo una falla monofásica | 81 |
| C.10.Señal de corriente bajo una falla monofásica | 81 |
| C.11.Fasores de tensión y secuencia cero bajo una falla monofásica | 82 |
| C.12.Fasores de corriente y secuencia cero bajo una falla monofásica | 82 |
| C.13.Falla bifásica | 82 |
| C.14.Equivalente simétrico falla bifásica | 83 |
| C.15.Señal de tensión bajo una falla bifásica | 83 |
| C.16.Señal de corriente bajo una falla bifásica | 84 |
| C.17.Fasores de tensión y secuencia cero bajo una falla bifásica | 84 |
| C.18.Fasores de corriente y secuencia cero bajo una falla bifásica | 85 |
| C.19.Falla bifásica Tierra | 85 |
| C.20.Equivalente simétrico falla bifásica tierra | 85 |
| C.21.Señal de tensión bajo una falla bifásica tierra | 86 |
| C.22.Señal de corriente bajo una falla bifásica tierra | 86 |
| C.23.Fasores de tensión y secuencia cero bajo una falla bifásica tierra. | 87 |

| | |
|---|-----|
| C.24. Fasores de corriente y secuencia cero bajo una falla bifásica tierra | 87 |
| E.1. Procedimiento para agregar carpeta de archivos al path de MatLab. | 95 |
| E.2. Procedimiento para buscar el directorio actual de MatLab. | 95 |
| E.3. Pantalla Inicial de PRtool | 96 |
| E.4. Interfaz <i>PRtool</i> | 97 |
| E.5. Diagrama de flujo de la sección agrupamiento. | 99 |
| E.6. Correlación bivariada, (a) Con todas las variables. (b) Manera de excluir variables. (c) Excluyendo las variables a y d | 103 |
| E.7. Explicaciones ACP, Base de datos ejemplo RPT.txt, (a) Con todas las variables. (b) Manera de excluir variables. (c) Excluyendo las variables a y d | 104 |
| E.8. Formas de visualización de componentes principales. | 104 |
| E.9. Agrupamientos en el primer plano factorial y cobertura aproximada de cada grupo . | 105 |
| E.10. Agrupamientos en hiperplano factorial y cobertura aproximada de cada grupo . . . | 105 |
| E.11. Agrupamientos en el primer plano factorial y cobertura aproximada de cada grupo por medio de repartición manual. (a) Ubicación de centros. (b) Grupos formados. (c) Cobertura aproximada de cada grupo. | 106 |
| E.12. Forma de almacenar las CP y los centros utilizados | 106 |

Lista de cuadros

| | |
|---|----|
| 2.1. Tabla de contingencia datos binarios | 11 |
| 2.2. Variables utilizadas en ejemplo de ACP | 18 |
| 2.4. Vectores propios asociados a valores propios en ejemplo de ACP | 19 |
| 3.1. Resistencias de falla empleadas en la simulación | 35 |
| 3.3. Variables utilizadas en el modelo Tensión-Corriente | 36 |
| 3.5. Variables utilizadas en el modelo de corriente | 37 |
| 3.7. Variables utilizadas en el modelo de tensión | 38 |
| 4.1. Clases de falla usadas en la simulación | 54 |
| 4.3. Resistencias de falla usadas en la simulación | 54 |
| 4.5. Zonificación sistema de distribución de prueba. | 55 |
| 4.6. Identificación de tipo de falla con datos de entrenamiento. | 56 |
| 4.7. Identificación de tipo de falla con datos de validación. | 57 |
| 4.8. Localización de la zona fallada con datos de prueba. | 58 |
| 4.9. Localización de la zona fallada con todos los datos de entrenamiento. | 59 |
| 4.10. Porcentaje de acierto en localización de la zona fallada con datos de entrenamiento con modelo corriente. | 59 |
| 4.11. Porcentaje de acierto en localización de la zona fallada con datos de entrenamiento con modelo tensión. | 59 |
| 4.12. Localización de la zona fallada con datos de prueba de validación. | 61 |
| 4.13. Localización de la zona fallada con todos los datos de entrenamiento. | 61 |
| 4.14. Porcentaje de acierto en localización de la zona fallada con datos de validación con modelo corriente. | 61 |
| 4.15. Porcentaje de acierto en localización de la zona fallada con datos de validación con modelo tensión. | 62 |
| B.1. Parámetros del generador. | 72 |

| | |
|--|-----------|
| B.2. Datos de cargas del sistema. | 72 |
| B.3. Parámetros de las líneas. | 73 |
| B.4. Factor de potencia. | 73 |
| C.1. Estadística de ocurrencia de fallas de acuerdo al tipo. | 76 |
| C.3. Estadística de ocurrencia de fallas de acuerdo al sitio. | 77 |
| D.1. Componentes principales en modelo Tensión-Corriente L-T | 89 |
| D.3. Componentes principales en modelo Tensión-Corriente L-L | 89 |
| D.5. Componentes principales en modelo Tensión-Corriente L-L-T | 90 |
| D.7. Componentes principales en modelo Tensión-Corriente L-L-L | 90 |
| D.9. Peso de cada CP modelo Tensión-Corriente. | 90 |
| D.11. Peso acumulado CP modelo Tensión-Corriente (en %). | 90 |
| D.13. Componentes principales en modelo corriente L-T | 91 |
| D.15. Componentes principales en modelo corriente L-L | 91 |
| D.17. Componentes principales en modelo corriente L-L-T | 91 |
| D.19. Componentes principales en modelo corriente L-L-L | 91 |
| D.21. Peso de cada CP modelo corriente. | 91 |
| D.23. Peso acumulado CP modelo corriente (en %). | 92 |
| D.25. Componentes principales en modelo tensión L-T | 92 |
| D.27. Componentes principales en modelo tensión L-L | 92 |
| D.29. Componentes principales en modelo tensión L-L-T | 92 |
| D.31. Componentes principales en modelo tensión L-L-L | 93 |
| D.33. Peso de cada CP modelo tensión. | 93 |
| D.35. Peso acumulado CP modelo tensión (en %) | 93 |

Capítulo 1

INTRODUCCIÓN

En la actualidad, la desregulación del mercado energético, ha llevado a que el sector eléctrico se interese por la calidad en la prestación del servicio, convirtiéndose en un parámetro importante a la hora de medir la competitividad de las empresas distribuidoras de energía eléctrica. Este esquema competitivo exige que las empresas distribuidoras implementen programas para mejorar su desempeño en todas las áreas. La nueva reglamentación enfocada a los sistemas de distribución, se basa en los principios relacionados con la eficiencia y la calidad del servicio. Como consecuencia de éste hecho las empresas distribuidoras de energía precisan tener amplio manejo de los temas concernientes a la calidad del servicio. Para cumplir con las exigencias técnicas y mínimo tiempo de interrupción, es necesario invertir en automatización de los elementos del sistema, por medio de hardware y software especializado para mejorar aspectos como restauración del servicio, reducción de pérdidas y monitoreo de la calidad, entre otros. La continuidad del suministro se ve directamente afectada por la ocurrencia de fallas en las redes de distribución y es así como la atención oportuna de éstas se convierte en un tema de gran importancia para las empresas distribuidoras.

Las fallas típicas que se presentan en los sistemas de distribución son:

Falla monofásica de línea a tierra.

Falla doble línea.

Falla doble línea a tierra.

Falla trifásica a tierra

A continuación, se presenta el proyecto de grado para obtener el título de ingeniero electricista titulado ***LOCALIZACIÓN DE FALLAS: RECONOCIMIENTO ESTADÍSTICO DE PATRONES***, el cual tiene como meta la localización de fallas en sistemas de distribución, a partir de la aplicación de técnicas estadísticas para el reconocimiento de patrones.

1.1. Objetivos

El objetivo general de este proyecto de grado es el de proponer una metodología para la identificación y la selección de patrones a partir de señales de tensión y corriente que se relacionen directamente con la localización de fallas en sistemas de distribución apoyados de un análisis estadístico multivariado.

Con la finalidad de que esta meta se cumpla se han trazado los siguientes objetivos específicos:

- Obtener señales de tensión y corriente a partir de la simulación de un sistema de distribución prototipo en condiciones de falla, utilizando ATP/EMTP y MatLab.
- Extraer descriptores que caractericen las señales de tensión y corriente en condiciones de falla, a partir del análisis exploratorio de los datos.
- Determinar sobre el sistema de distribución prototipo regiones características que representen al mismo, aplicando técnicas para el análisis de conglomerados.
- Establecer relaciones entre las regiones características y la localización de la falla.
- Implementar la metodología a través de una herramienta software en MatLab para la localización de las fallas basada en técnicas de reconocimiento de patrones.
- Elaborar un manual del usuario de la herramienta de software.

1.2. Estructura del documento

Este documento se dividirá en cinco partes en donde se mostrarán de forma breve los fundamentos que rigen la estadística multivariada y la forma correcta de utilizar la herramienta diseñada para solucionar el problema asociado con la localización de las fallas en las redes de distribución. En el capítulo 2 se dará la teoría relacionada con el análisis estadístico multivariado enfatizando en el análisis de componentes principales (ACP) y el análisis de conglomerados o reconocimiento de patrones (clusters), así como también el trabajo que se ha realizado a partir de la utilización de estas técnicas.

En el capítulo 3 se muestra la aplicación directa de las técnicas anteriormente mencionadas a la solución del problema de la localización de fallas en sistemas de distribución. Primero se muestra la caracterización del sistema de potencia prototipo a partir del cálculo de descriptores que detallan el comportamiento del sistema en condiciones de falla, posteriormente se aplican las técnicas de ACP y análisis de conglomerados (k-medias), por medio de la herramienta de software diseñada.

El capítulo 4 trata de las pruebas y resultados obtenidos. Las observaciones y conclusiones del trabajo de investigación se muestran en el capítulo 5.

Por último se agregan anexos A, B, C, D y E, los cuales son importantes para profundización de algunos temas.

Capítulo 2

ANÁLISIS ESTADÍSTICO MULTIVARIADO

2.1. Introducción

La filosofía del análisis estadístico multivariado¹ nace como respuesta a la necesidad de procesar gran cantidad de información contenida en el planteamiento de un problema y a la facilidad que brindan las nuevas tecnologías para procesar gran cantidad de datos complejos [R.T1, 2005]. Para los propósitos de cualquier disciplina aplicada, la mayor parte de los fenómenos se analizan y modelan de forma multivariada con el fin de entender y caracterizar su comportamiento. A menos que el problema sea tratado como un problema multivariado, éste es tratado superficialmente [A.H, 1999].

“Los métodos estadísticos multivariados hacen posible plantear preguntas específicas y precisas de considerable complejidad en marcos idóneos, lo que posibilita llevar a cabo investigaciones teóricamente significativas y evaluar los efectos de las variaciones paramétricas ocurridas de forma natural en el contexto en que normalmente ocurren”²

2.2. Definición

El análisis multivariado se refiere a todos los métodos estadísticos que analizan simultáneamente medidas de cada individuo u objeto sometido a investigación. Cualquier análisis simultáneo de dos o más variables³ puede ser considerado multivariado. El propósito del análisis multivariante es

¹Multivariado, multivariante o multivariable se refiere al mismo término.

²Según los estadísticos Hardyck y Petrinovich.[A.H, 1999]

³Se denomina variable a cualquier atributo o característica observable en los elementos de una muestra o de una población.

medir, explicar y predecir el grado de relación de los valores teóricos (combinaciones ponderadas de variables). Por lo tanto el carácter multivariante reside en los múltiples valores teóricos y no solo en el número de variables u observaciones [A.H, 1999].

2.3. Algunos conceptos básicos del análisis estadístico multivariado

2.3.1. Los datos

Las técnicas de análisis multivariado tratan con datos asociados a conjuntos de medidas sobre un número de individuos u objetos. El conjunto de individuos junto con sus variables se dispondrá en un arreglo matricial \mathbf{X} como se muestra en la ecuación (2.1), con n filas (observaciones) y p columnas (variables) [D.L, 2002].

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad (2.1)$$

Las variables pueden ser cuantitativas o cualitativas y estas a su vez continuas o de intervalos y binarias o generales, respectivamente. En éste documento solo se tendrán presentes las variables cuantitativas y continuas. Cada una de estas variables se denomina variable univariada y el conjunto de las p variables forman una variable multivariada.[D.L, 2002, P.D, 2002]

2.3.2. El valor teórico

Se refiere a una combinación lineal de variables con ponderaciones determinadas empíricamente. El investigador especifica las variables, mientras que las ponderaciones son objeto específico de determinación por parte de la técnica multivariada. El valor teórico capta el carácter multivariado del análisis. Por tanto el valor teórico es el punto central del análisis, sin olvidar la contribución de cada variable por separado [A.H, 1999].

2.3.3. Escala de medida

Un investigador no puede separar o identificar una variación a menos de que esta pueda ser medible. La medida es importante para representar con precisión el concepto interés y es crucial

en la selección del método estadístico multivariado. En este documento solo se usarán escalas de medida métricas⁴, debido a que estas proporcionan mayor precisión y además permite realizar operaciones matemáticas [A.H, 1999].

2.3.4. Vector de medias. \bar{x}

La medida de centralización más utilizada para describir los datos multivariados es el vector de medias, que es un vector de dimensión p , cuyas componentes son las medias de cada una de las p variables, la cual se calcula como se muestra en la ecuación (2.2). El vector de medias se organiza como se muestra en (2.3).

$$\bar{x}_p = \frac{1}{n} \sum_{i=1}^n x_i \quad (2.2)$$

$$\bar{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_p \end{bmatrix} \quad (2.3)$$

Las medidas de centralización escalares basadas en el orden de las observaciones no pueden generalizarse fácilmente al caso multivariado. Por ejemplo, se puede calcular el vector de medianas, pero éste punto no tiene necesariamente una situación como centro de los datos. Esta dificultad proviene de la falta de un orden natural de los datos multivariados [D.L, 2002, P.D, 2002].

2.3.5. La matriz de varianzas y covarianzas S

La varianza mide la variabilidad de las variables escalares respecto a la media. La relación lineal entre dos variables se mide por la covarianza. La covarianza entre las variables x_i y la variable x_j se calcula de la siguiente forma:

$$s_{ij} = \frac{1}{n} \sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j) \quad (2.4)$$

La matriz de varianzas y covarianzas S , ecuación (2.5), es una matriz cuadrada y simétrica cuya diagonal corresponde a las varianzas y los elementos por fuera de ésta corresponden a las covarianzas

⁴Valores cuantitativos o numéricos

entre las variables [P.D, 2002].

$$\mathbf{S} = \begin{bmatrix} s_{11}^2 & s_{12}^2 & \cdots & s_{1p}^2 \\ s_{21}^2 & s_{22}^2 & \cdots & s_{2p}^2 \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1}^2 & s_{p2}^2 & \cdots & s_{pp}^2 \end{bmatrix} \quad (2.5)$$

2.3.6. La varianza total y la varianza media

Para recopilar la variabilidad de un conjunto de variables, se calcula la traza de su matriz \mathbf{S} . Se define la varianza total de los datos por la ecuación (2.6):

$$T = tr(s) = \sum_{i=1}^n s_i^2 \quad (2.6)$$

Y la varianza media por la ecuación (2.7) :

$$\bar{s}^2 = \frac{1}{p} \sum_{i=1}^n s_i^2 \quad (2.7)$$

Estas medidas no tienen presente la estructura de dependencia entre las variables; es decir, si la dependencia entre las variables es muy alta, la variabilidad conjunta es pequeña [P.D, 2002].

2.3.7. La matriz de correlación \mathbf{R}

La interdependencia lineal entre las variables se mide con el coeficiente de correlación muestral r , el cual se debe analizar teniendo en cuenta su magnitud y signo así: valores cercanos a 1 o -1 indican que los puntos están a lo largo de una línea recta con pendiente positiva o negativa respectivamente, mientras que valores cercanos a cero indican que no hay relación lineal, sin descartar que exista otro tipo de relación [D.L, 2002, R.T1, 2005]. La correlación entre las variables x_i y la variable x_j se calcula como se muestra en la ecuación (2.8):

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}} \quad -1 \leq r_{ij} \leq 1 \quad (2.8)$$

En el caso en que las variables tengan una escala de medición distinta o en otras unidades y en particular grados de dispersión distintos, se puede incurrir en el error de descartar una variable importante en la estructura de variabilidad sólo porque su magnitud es menor que la de otra. Es decir las variables con mayor magnitud predominarán por completo en el análisis. En particular si se cambia las unidades, los resultados podrían cambiar completamente [D.L, 2002]. El arreglo

matricial de los coeficientes de correlación se muestra en (2.9).

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix} \quad (2.9)$$

Para el caso en que las variables sean de magnitud distinta, estas se pueden transformar para obtener nuevas variables que si se puedan comparar, lo cual se consigue tipificando las variables, es decir restándoles su media y dividiéndolas por su desviación típica (normalizar) [R.T1, 2005].

2.3.8. La distancia

El concepto de distancia entre puntos se utiliza alternativamente para estudiar la variabilidad de un grupo de observaciones [P.D, 2002]. Para el investigador puede ser importante determinar si dos o mas individuos están cerca o no, con el objetivo rechazar o no una hipótesis estadística de acuerdo con la discrepancia de la muestra. También se puede utilizar para determinar que tanto se ajusta un modelo de regresión a la tendencia de un grupo de puntos [D.L, 2002].

Las propiedades de la función distancia son [D.L, 2002, P.D, 2002]:

- $d(x, y) \geq 0$ para cualquier valor de x o y .
- $d(x, y) = 0$ si y solo si $x=y$.
- Debe existir simetría $d(x, y) = d(y, x)$
- Debe cumplir con la desigualdad triangular $d(x, y) \leq d(x, z) + d(z, y)$

Donde x , y y z son individuos de un grupo de datos que tienen coordenadas (x_1, \dots, x_p) , (y_1, \dots, y_p) y (z_1, \dots, z_p) respectivamente.

Las métricas más utilizadas en la estadística multivariada son la distancia Euclidiana y la distancia de Mahalanobis. Estas distancias se describen a continuación.

2.3.8.1. Distancia euclideana

Se define como sigue en la ecuación (2.10):

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.10)$$

El inconveniente de ésta distancia es la dependencia existente entre la métrica y las unidades de medida de las variables. De tal manera que cuando no existe una unidad fija natural, no es recomendable emplearla. Este problema se evita dividiendo cada variable por un término que elimine el efecto de escala, conduciendo a una familia de *métricas euclídeas ponderadas* [P.D, 2002].

2.3.8.2. Distancia de mahalanobis

Se emplea cuando las variables del estudio están en diferentes escalas de medición y se encuentran correladas. El valor que representa la distancia entre dos individuos no solo cambia de acuerdo con las unidades de medida empleadas sino por el grado de asociación que hay entre estas variables.

La distancia de Mahalanobis se emplea frecuentemente para medir la distancia entre una observación multivariada y el centro de la población de donde procede la observación.

$$d(x, y) = \sqrt{(x_i - y_i)^T \frac{1}{\sum (x_i - y_i)}} \quad (2.11)$$

El valor de $d(x, y)^2$ se considera como un residual⁵ multivariado [D.L, 2002, P.D, 2002].

2.3.8.3. Otras distancias

Finalmente se resumen otras distancias que se pueden emplear en el trabajo estadístico:

- **Distancia de minkowski:** Es la distancia existente entre un par de observaciones identificados como vectores fila $X_h = (X_{h1}, \dots, X_{hp})$ $X_i = (X_{i1}, \dots, X_{ip})$.

$$d_{hi} = \left(\sum_{j=1}^p |X_{hj} - X_{ij}|^r \right)^{\frac{1}{r}} \quad (2.12)$$

La distancia euclídea se obtiene cuando se hace $r=2$ en (2.12) [D.L, 2002].

- **Distancia Manhattan:** También denominada de la ciudad y esta dada por la ecuación (2.13):

$$d_{hi} = \sum_{j=1}^p |X_{hj} - X_{ij}| \quad (2.13)$$

Esta resulta de hacer $r=1$ en la distancia de Minkowski, ecuación (2.12). El calificativo de ciudad es porque la distancia entre dos puntos de ésta es igual al número de cuadras (calles

⁵Distancia entre la observación y el centro de gravedad de todos los datos.

o carreras) que se deben recorrer para ir de un punto a otro [D.L, 2002].

2.3.9. Medidas de similaridad y de disimilitud.

Estas medidas sirven como indicadores de que tanto son iguales o no, un par de objetos y que tan cercanos o distantes se encuentran. Se clasifican en medidas de proximidad y en medidas de similitud.

2.3.9.1. Las medidas de proximidad

También llamadas de similitud o semejanza. Miden el grado de semejanza entre dos objetos de forma que cuanto mayor es su valor, mayor es el grado de similaridad existente entre ellos y mayor es la probabilidad de clasificarlos en un mismo grupo. Esta clasificación se extiende si se tiene presente que las variables son cuantitativas, binarias o datos nominales [R.T2, 2005].

- **Medidas para variables cuantitativas:** Se usan cuando los valores que describen a la variables son continuos.

- *Coefficiente de congruencia:*

$$c_{ij} = \frac{\sum_{k=1}^p X_{ik}X_{jk}}{\sqrt{\sum_{k=1}^p X_{ik}^2} \sqrt{\sum_{k=1}^p X_{jk}^2}} \quad (2.14)$$

Este coeficiente representa el coseno del ángulo que forman los vectores de variables $(x_{i1}, \dots, x_{ip})'$ y $(x_{j1}, \dots, x_{jp})'$. Se consideran valores de c_{ij} mayores de 0.9 o 0.8 para afirmar que los datos son congruentes (muy similares).

- *Coefficiente de correlación:*

De este coeficiente se hace mención en el numeral 2.3.7 en la ecuación 2.8.

Es conveniente observar que tanto c_{ij} como r_{ij} toman valores comprendidos entre -1 y 1 tomando por lo tanto valores negativos. Dado que en algunos casos, los valores negativos cercanos a -1 pueden implicar fuerte semejanza entre los objetos clasificados, conviene en estas situaciones, utilizar como medida de semejanza sus valores absoluto.[R.T2, 2005]

- **Medidas para datos binarios:** Son apropiados cuando los datos están en escala nominal. Cada variable toma valores de 0 (de ausencia) y 1 (de presencia) de un atributo.

En este caso se contruyen para cada par de objetos r y s tablas de contingencia como se muestra en el cuadro 2.1.

Cuadro 2.1: Tabla de contingencia datos binarios

| Objeto s \ Objeto r | 0 | 1 |
|---------------------|---|---|
| 0 | a | b |
| 1 | c | d |

Donde:

a es el número de variables en las que los objetos r y s toman el valor de cero.

b es el número de variables en las que el objeto r es igual a 1 y s toma el valor de cero.

c es el número de variables en las que el objeto r es igual a 0 y s toma el valor de uno.

d es el número de variables en las que los objetos r y s toman el valor de uno.

$$p = a + b + c + d$$

Teniendo presente estas tablas de contingencia, las medidas de semejanza más utilizadas son:

Coefficiente de Jacard:

$$C_J = \frac{d}{b + c + d} \quad (2.15)$$

Coefficiente de acuerdo simple:

$$C_{AS} = \frac{a + d}{p} \quad (2.16)$$

Estas medidas toman valores entre 0 y 1, determinando la porción de acuerdo con los valores tomados de las p variables, existente entre los dos objetos [R.T2, 2005].

2.3.9.2. Las medidas de disimilitud

Conocidas también como de semejanza o distancia, miden la distancia entre dos objetos de forma que, cuanto mayor sea su valor, más diferentes son los objetos y menor es la probabilidad de que los métodos de clasificación los sitúen en el mismo grupo [R.T2, 2005].

De cada una de las normas utilizadas para medir distancia verlas en el numeral 2.3.8. Todas estas distancias no son invariantes a cambios de escala por lo que se sugiere estandarizar los datos si las unidades de medida de las variables no son comparables.

2.4. Clasificación de las técnicas multivariadas.

Los métodos utilizados en el análisis estadístico multivariado se pueden clasificar en tres categorías:

2.4.1. Métodos de dependencia

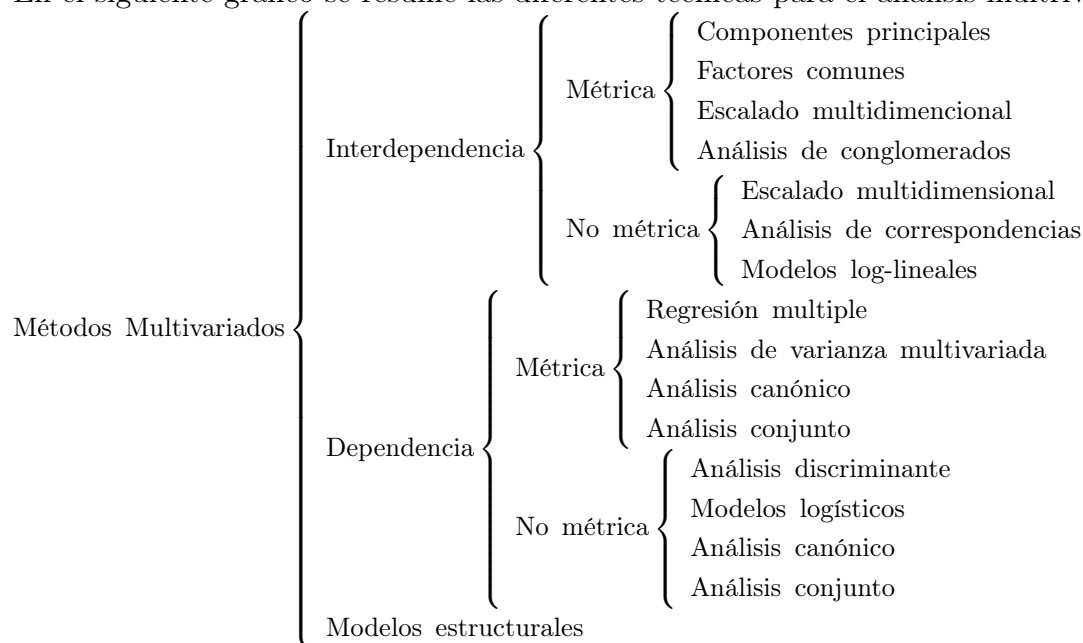
Un análisis de dependencia se puede definir como aquel en el que una variable o un conjunto de ellas se identifica como *variable dependiente* y que va a ser explicada por otras conocidas como *variables independientes*. Estos a su vez se pueden dividir en dos tipos según el número de variables dependientes y el tipo de escalas de medida empleadas [A.H, 1999].

2.4.2. Métodos de interdependencia

En contraste con los métodos de dependencia, un análisis de interdependencia es aquel en el que ninguna variable o ningún grupo de ellas se define como independiente o dependiente. Es decir, el procedimiento implica el análisis de todas las variables del conjunto simultáneamente [A.H, 1999]. Las técnicas de análisis de interdependencia buscan el *cómo* y el *por qué* se relacionan o asocian un conjunto de variables [D.L, 2002].

2.4.3. Modelos estructurales

En el siguiente gráfico se resume las diferentes técnicas para el análisis multivariado.



Aunque los modelos estructurales tienen aspectos tanto de dependencia como de interdependencia, no se etiqueta como una técnica perteneciente a estas. Tiene como objetivo el modelado que determina las relaciones entre las variables, a través de un sistema de ecuaciones lineales [D.L, 2002]. Se caracteriza por dos componentes básicos: El primero de tipo estructural, el cual relaciona las variables dependientes y las variables independientes (teoría antes que la experiencia) y

el segundo de tipo métrico, permite emplear varios indicadores para una única variable dependiente o independiente [A.H, 1999].

2.5. Supuestos del análisis multivariado

El análisis estadístico multivariado requiere de varios supuestos subyacentes a las técnicas estadísticas convencionales, esto con el fin de contrastar el valor teórico del modelo multivariado que actúa colectivamente sobre las variables a analizar y que debe cumplir con los mismos supuestos que las variables individuales [A.H, 1999]. A continuación se mencionan algunos de los supuestos que emplea el análisis multivariado.

2.5.1. Normalidad

Es el supuesto fundamental del análisis multivariante, y se refiere a la correspondencia del perfil de la distribución de los datos con la distribución normal. Si la variación respecto a ésta es lo suficientemente alta, todos los test estadísticos realizados no son válidos. Los valores de la mayoría de las medidas asociadas con objetos se aglomeran simétricamente en torno a un valor central específico. La mayor parte de estas medidas se ubican dentro de alguna distancia determinada respecto al valor central, a la izquierda o a la derecha, las demás se presentan de manera cada vez más escasa en tanto que la distancia al valor central es grande. La normalidad multivariada (combinación de dos o más variables) implica que las variables individuales son normales en sentido univariado y que sus combinaciones sean también normales. Por consiguiente si una variable es normal multivariada, es también normal univariada, sin ser necesariamente ésta suposición bidireccional. Por lo tanto, una situación en que todas las variables exhiben normalidad univariada ayudará a tener normalidad multivariada [A.H, 1999].

2.5.2. Homocedasticidad

La homocedasticidad es un supuesto primordial en las relaciones de dependencia entre variables, y se refiere a que las variables dependientes exhiben iguales niveles de variabilidad a lo largo del rango del predictor de las mismas. La homocedasticidad es deseable para que la varianza de la variable dependiente no se concentre solamente en un limitado rango de los valores independientes, y de ésta manera se puedan realizar predicciones [A.H, 1999].

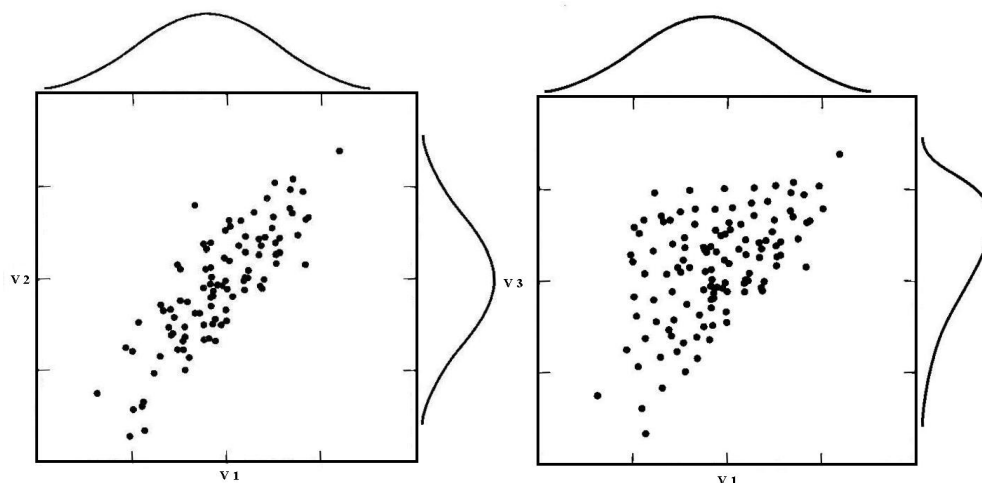


Figura 2.1: (a)Gráfico de dispersión con relacion de homocedasticidad. (b)Gráfico de dispersión con relacion de heterocedasticidad.

En el gráfico (a) de la Figura (2.1) se observa la dispersión de datos para dos variables ($V1$ y $V2$) con distribuciones normales, el cual exhibe la misma dispersión a lo largo de todos los valores de los datos (Homocedasticidad). Sin embargo, existen casos donde aparece una dispersión diferente entre dos variables ($V1$ y $V3$) provocada por la asimetría de una de éstas ($V3$) (Heterocedasticidad) como se ve en el gráfico (b) de la figura (2.1). El efecto de la heterocedasticidad está a menudo también relacionado con el tamaño de la muestra, especialmente cuando se examina la dispersión de la varianza entre grupos.

No se profundiza en éste supuesto debido a que solo se emplea métodos de interdependencia (Análisis de componentes principales y reconocimiento de patrones).

2.5.3. Linealidad

La linealidad es un supuesto implícito de todas las técnicas multivariadas, incluyendo el análisis de componentes principales. La medida de linealidad se haya mediante la correlación y los efectos no lineales no están representados en éste valor, por lo que es prudente revisar como impacta la correlación en conjunto con la gráfica de dispersión.

2.6. Análisis de componentes principales (ACP)

ACP es un método estadístico que suministra información sobre la interdependencia lineal de las variables. Esta interdependencia se mide con el coeficiente de correlación muestral r . El ACP

tiene como principal objetivo la estructuración de un conjunto de datos multivariados mediante la generación de nuevas variables que conservan la misma información y que además son incorrelacionadas entre si, donde éstas son combinaciones lineales de las primeras, y contienen la mayor parte de la variabilidad presente en el conjunto inicial [D.L, 2002]. Esta es una metodología de tipo determinista en la cual no es necesario hacer supuestos para indicar la adecuación de un modelo determinado para posteriormente validarlo. Por ejemplo el supuesto de normalidad, de asumir una distribución probabilística, por resultado.

2.6.1. Ventajas del ACP

Las ventajas de la aplicación del ACP se reflejan en los siguientes tópicos:

- Reducción de la dimensión del espacio donde están inscritos los datos.
- Transformación de las variables originales, en general correladas, en nuevas variables incorreladas, facilitando la interpretación de los datos [H.W, 2003, P.D, 2002].
- Determinación de relaciones insospechadas que vienen a enriquecer la interpretación. Muchas veces este análisis es un paso intermedio para establecer que metodología se va a utilizar para analizar de manera más profunda la información (Análisis de regresión o análisis de conglomerados) [R.T1, 2005].

Es importante recalcar que las nuevas variables se denominan componentes principales y los valores de las nuevas componentes se denominan puntajes de las componentes principales. La variabilidad total de los datos no se altera por transformaciones ortogonales de estos, además las nuevas variables tienen correlación igual a cero [D.L, 2002].

Generalmente las dos primeras componentes principales contienen la mayoría de la información, por lo tanto es común ubicar los datos en los dos nuevos ejes o componentes. A este plano se le conoce como plano factorial, y en el se pueden realizar algunas clasificaciones de los individuos, de las variables o de ambos y la detección de valores atípicos. Se recomienda el uso de ACP cuando la correlación entre variables es alta. Realizar un análisis de componentes principales sobre la matriz de correlación de las variables originales, es equivalente a realizarlo con la matriz de covarianzas de las variables tipificadas [R.T1, 2005].

2.6.2. Cálculo de las componentes principales

Como se explicó anteriormente y con el fin de evitar complicaciones generadas por interpretaciones erróneas debido al empleo de diferentes unidades, se sugiere trabajar con la matriz de

correlación de las variables originales \mathbf{R} . A continuación se muestra un algoritmo sencillo para la determinación de las componentes principales de un conjunto de variables que están fuertemente correlacionadas entre si.

1. Hallar la matriz de correlación de las variables que están describiendo un contexto.
2. Calcular los valores propios resolviendo la siguiente relación: $|(\mathbf{R} - \lambda\mathbf{I})| = 0$
3. Ordenar de manera descendiente los valores propios.
4. Los vectores propios asociados se calculan resolviendo: $(\mathbf{R} - \lambda\mathbf{I})\mathbf{X} = 0$

Donde:

- \mathbf{R} es la matriz de correlación de las variables originales.
- \mathbf{I} es la matriz identidad.
- λ son los valores propios hallados al resolver la ecuación.
- \mathbf{X} es el vector propio generado por cada valor propio hallado.

Cada componente principal es la combinación lineal de las variables originales cuyos coeficientes son los valores del vector propio asociado a cada valor propio hallado. El procedimiento asociado se detalla en el anexo A.

2.6.3. Propiedades de las componentes principales (CP)

Las componentes principales son nuevas variables con las siguientes propiedades [P.D, 2002]:

- Conservan la variabilidad inicial: La suma de las varianzas de componentes principales es igual a la suma de las varianzas de las variables originales, y la varianza generalizada de los componentes es igual a la original.
- La proporción de la variabilidad explicada por una componente es el cociente entre su varianza, el valor propio asociado al vector propio que lo define y la suma de los valores propios de la matriz.
- La correlación entre una CP y una variable X_i es proporcional al coeficiente de esa variable en la definición de la componente, y el coeficiente de proporcionalidad es el cociente entre la desviación típica de CP y la desviación típica de la variable.
- Las r componentes principales ($r < p$) proporcionan la predicción lineal óptima con r variables del conjunto de variables \mathbf{X} .

2.6.4. Selección del número de componentes principales

Es difícil encontrar un criterio estrictamente formal para determinar el número de CP a utilizar. El analista de los datos es quien define la cantidad de pérdida de variabilidad permitida.

Como se explicó anteriormente cada CP aporta una parte de variabilidad de los datos analizados, distribuyéndose en forma decreciente desde la primera hasta la última. La clave está en determinar que cantidad de variación explicada que se considera satisfactoria, y se selecciona el número de componentes que cumplen con éste requisito.

La mejor forma de explicar el procedimiento anterior, es utilizar un diagrama de barras que muestra que cantidad de información contiene cada componente principal.⁶

2.6.5. Interpretación de cada componente principal

“Un aspecto clave del ACP es la interpretación de los factores, ya que ésta no viene dada a priori, sino que será inducida tras observar la relación de los factores de las variables iniciales (habrá que estudiar tanto signo como magnitud de las correlaciones). Esto no siempre es fácil, y será de vital importancia en el conocimiento que el experto tenga sobre la materia de investigación.”

7

Cuando existe una alta correlación positiva entre todas las variables, el primer CP tiene todas las coordenadas del mismo signo y se pueden interpretar como un promedio ponderado de todas las variables. Se puede recurrir a aproximaciones de cada componente para facilitar la interpretación de estos.

2.6.6. Información de la última componente principal.

Generalmente las primeras componentes principales se consideran que poseen la mayor parte de la información de las variables originales, sin embargo las últimas componentes pueden poseer información importante [D.L, 2002, P.D, 2002]. Debido a que los autovalores sirven para mostrar las varianzas de las componentes principales, las últimas componentes principales tienen varianzas más pequeñas. Si la discrepancia de un componente es cero o íntima al cero, la componente representa una relación lineal entre las variables que es esencialmente constante, por lo tanto, si el último autovalor está cerca del cero, indica la presencia de una colinealidad que puede suministrar nueva información para el investigador [P.D, 2002].

⁶Dillon y Goldstein (1986, Págs. 47-50) describen técnicas gráficas para determinar el número de componentes a utilizar.

⁷“Manuel Terrádez “ Memorias de diplomado de análisis estadístico multivariado UIS 2005

2.6.7. Ejemplo de aplicación de ACP

A continuación se realiza un pequeño ejemplo de la aplicación de ACP para mostrar la reducción de la dimensionalidad de un grupo de variables que describen el comportamiento de un sistema de potencia en condiciones de falla.

Las variables utilizadas⁸ se muestran en el cuadro 2.2 :

Cuadro 2.2: Variables utilizadas en ejemplo de ACP

| Variable | Convención |
|---------------------------|------------|
| Profundidad (p.u) | PHT |
| Pendiente de bajada (p.u) | PBT |
| Pendiente de subida (p.u) | PST |
| Altura (p.u) | AHC |
| Pendiente de subida (p.u) | PSC |
| Pendiente de bajada (p.u) | PBC |

Como se mencionó en el numeral 2.6.2 lo primero que se debe hacer es revisar la matriz de correlaciones de las variables utilizadas, la cual se muestra en la Figura 2.2.

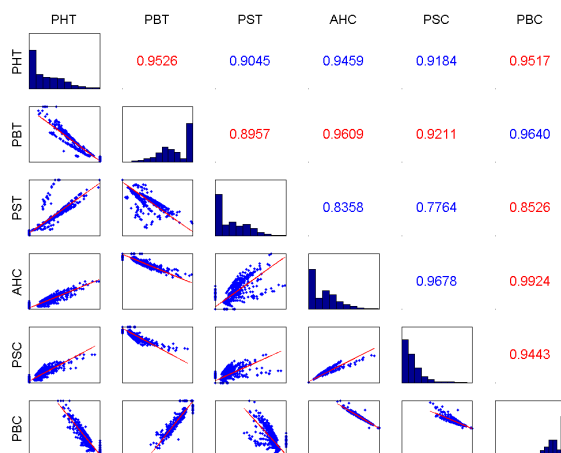


Figura 2.2: Gráfico de correlación y distribución de las variables

Se puede observar que las seis variables están fuertemente correlacionadas puesto que el coeficiente “ r ” de cada una de ellas es superior a 0,7. Esto hace que las variables sean linealmente dependientes entre ellas, además la visualización de toda la información no es posible debido a que

⁸Estas variables se utilizaron teniendo en cuenta el estudio mencionado en [M.J, 2003], y solo para mostrar la aplicación de ACP, no son determinantes para el objetivo principal de éste proyecto.

se deben manejar seis dimensiones. Luego el ACP es útil para reducir la dimensionalidad de los datos y para que las nuevas variables sean linealmente independientes.

Al desarrollar la relación $|(\mathbf{S} - \lambda\mathbf{I})| = 0$ se obtiene la siguiente ecuación característica⁹:

$$\lambda^6 - 6\lambda^5 + 2,2843\lambda^4 - 0,2393\lambda^3 + 0,0099\lambda^2 - 1,5505 * 10^{-4}\lambda + 4,496 * 10^7 = 0$$

La cual arroja los siguientes valores propios: $\lambda_1 = 5,5996$ $\lambda_2 = 0,2631$ $\lambda_3 = 0,05640$ $\lambda_4 = 0,04290$ $\lambda_5 = 0,03430$ $\lambda_6 = 0,00360$ y estos a su vez generan los vectores propios como se ve en el cuadro 2.4:

Cuadro 2.4: Vectores propios asociados a valores propios en ejemplo de ACP

| | Comp 1 | Comp 2 | Comp 3 | Comp 4 | Comp 5 | Comp 6 |
|------------|------------|-------------|-----------|------------|-----------|------------|
| PHT | 0,00022669 | -0,00023492 | 0,0070769 | -0,0003837 | 0,012283 | 0,9999 |
| PBT | -0,011414 | 0,013977 | -0,32426 | -0,94503 | -0,038033 | 0,0024055 |
| PST | 0,0098105 | -0,021046 | 0,94552 | -0,32411 | -0,01897 | -0,0065906 |
| AHC | 0,010893 | -0,0091234 | -0,00588 | 0,041985 | -0,99892 | 0,012325 |
| PSC | 0,8224 | 0,56888 | 0,0032486 | -0,0027799 | 0,0036348 | -0,0001215 |
| PBC | -0,56861 | 0,82198 | 0,027411 | 0,010161 | -0,013438 | 0,000297 |

Una vez halladas las componentes principales, se puede determinar con cuantas de ellas se tiene la mayor cantidad de información posible, empleando el diagrama de barras de la figura (2.3) donde se muestra que tanto explica cada CP.

⁹Se desarrolla la ecuación con la matriz de covarianza debido a que las variables empleadas están normalizadas al sistema po unidad (p.u)

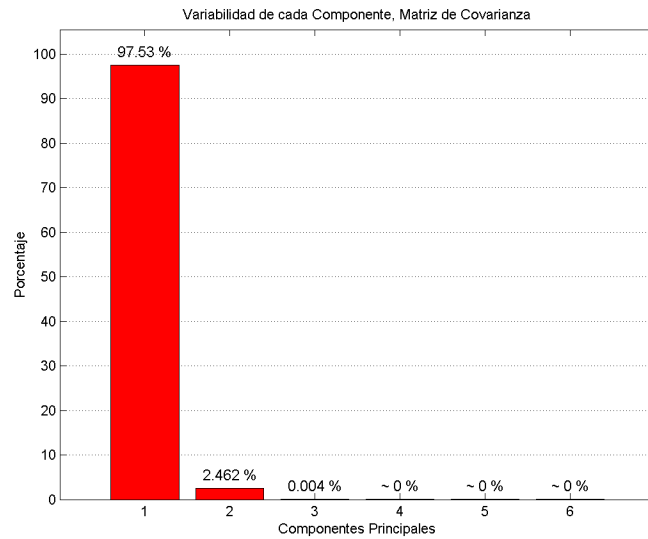


Figura 2.3: Porcentaje de información de cada CP

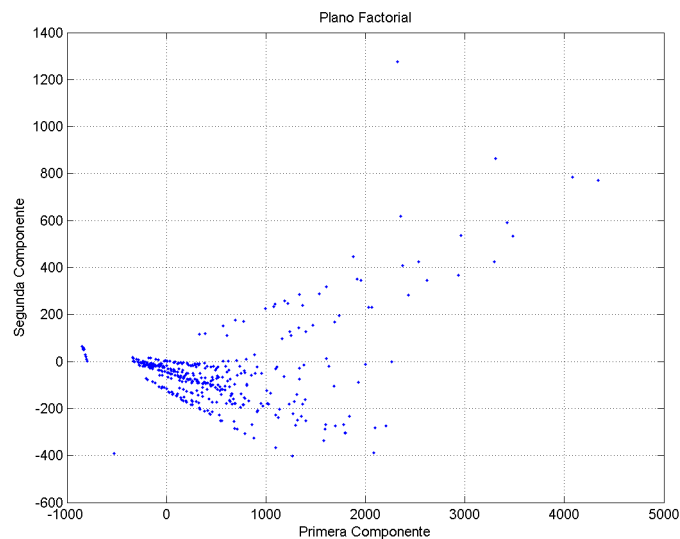


Figura 2.4: Primer plano factorial ejemplo ACP

Como se observa en la figura (2.3) el 99 % de la información se encuentra explicada en las dos primeras componentes o primer plano factorial, figura (2.4), reduciéndose la dimensionalidad del problema de seis dimensiones, a dos.

2.7. Reconocimiento de patrones o análisis de conglomerados

Actualmente se consideran las técnicas para el reconocimiento de patrones como disciplinas de la Inteligencia Artificial (IA). Para justificar esto se parte de dos principios:

- Adquisición y representación del conocimiento: Esta actividad se realiza a partir de la transferencia de información de una o más fuentes a un sistema informático de forma útil para que sea procesado. El reconocimiento de formas adquiere patrones de una clase conocida y se almacenan para establecer un patrón prototipo de cada clase.
- Aprendizaje: Implica la adaptación a cambios para poder llevar a cabo una tarea. En el reconocimiento de patrones es necesario llevar a cabo un algoritmo de aprendizaje para calcular el conjunto de patrones prototipo que caracterizan cada una de las clases a discriminar.

El análisis de conglomerados¹⁰ tiene como objeto congrega elementos en grupos en función de las similitudes entre ellos, de manera tal que los elementos correspondientes a un mismo grupo deben mostrar un alto grado de homogeneidad interna (dentro del conglomerado) y un alto grado de heterogeneidad externa (entre conglomerados), siendo evidente gráficamente la cercanía de los elementos de un grupo y la lejanía entre elementos de diferentes grupos.

Este análisis se conoce también como métodos de *clasificación automática no supervisada*¹¹, *reconocimiento de patrones sin supervisión*, taxonomía numérica o técnicas de *cluster*¹² [D.L, 2002, R.T2, 2005, P.D, 2002].

Partiendo de que el objetivo del reconocimiento de patrones es la comparación de objetos basados en algunas variables, éste análisis no está soportado sobre la base teórica que permita hacer inferencia y por ello suele calificarse como una técnica exploratoria. La solución no es única, las agrupaciones resultantes dependen de varios elementos en el procedimiento y de las variables utilizadas para su construcción [R.T2, 2005].

*“Conviene, sin embargo, estar siempre alerta ante el peligro de obtener, como resultado del análisis, no una clasificación de los datos sino una disección de los mismos, en distintos grupos que solo existen en la memoria del computador. El conocimiento que el analista tenga acerca del problema decidirá cuales de los grupos obtenidos son significativos y cuales no”*¹³

¹⁰“Son regiones continuas de un espacio que contiene una densidad relativamente alta de puntos, las cuales están separadas por regiones que contienen una densidad relativamente baja de puntos.” **Everitt 1980[D.L, 2002]**

¹¹No supervisados se aplica para distinguirlos del análisis discriminante

¹²La palabra conglomerado es la traducción más cercana al término ‘cluster’

¹³“Salvador Figueras” Memorias de diplomado de análisis estadístico multivariado UIS 2005

2.7.1. Objetivos del reconocimiento de patrones

El reconocimiento estadístico de patrones generalmente busca los siguientes propósitos:

- La identificación de una estructura natural de los objetos, es decir, el desarrollo de una tipología o clasificación de objetos.
- La búsqueda de esquemas conceptuales que expliquen el agrupamiento de algunos objetos.
- La formulación de hipótesis mediante la descripción y exploración de los grupos conformados.

Tres son los elementos requeridos para el reconocimiento de patrones, el primero, la medida que determine el grado de similaridad entre los objetos, el segundo, el procedimiento para la formación de grupos o de conglomerados y por último la técnica para determinar cuantos grupos se deben formar. [C.W, 2003, S.B, 2002].

2.7.2. Clasificación de los métodos de reconocimiento de patrones

Existe una gran variedad de algoritmos establecidos para la conformación de conglomerados, y cada uno de estos métodos tiene una perspectiva diferente, produciendo resultados generalmente distintos al ser aplicados sobre el mismo conjunto de datos. Es importante tener presente que al momento de aplicar algún método de reconocimiento de patrones debe existir concordancia entre la medida de similaridad utilizada, las variables que se consideran y la naturaleza de la tipología esperada [D.L, 2002].

2.7.2.1. Métodos jerárquicos

Los métodos jerárquicos parten de una matriz de distancias o similaridades entre los elementos de la muestra. Se construye una jerarquía basada en estas distancias, la cual posee la característica de que en cada paso del algoritmo solo un objeto cambia de grupo y los grupos están anidados en los pasos anteriores. Si un objeto ha sido asignado a un grupo este no cambia de grupo.

El proceso de agrupamiento se hace de una manera aglomerativa o por un proceso de división. En el tipo aglomerativo se comienza con tantos grupos como objetos existan, se recalculan las distancias entre los grupos existentes, para mezclarse los más cercanos sucesivamente hasta que el algoritmo termina con un solo grupo que contiene todos los elementos. Los métodos de división inician con todos los elementos dentro de un mismo conglomerado. En cada paso del algoritmo se divide en el grupo más heterogéneo. El algoritmo termina cuando existan tantos grupos como elementos disponibles [M.J, 2001, D.L, 2002, P.D, 2002].

Este trabajo se centra en el análisis de métodos aglomerativos ya que requieren menor esfuerzo de cálculo siendo los más utilizados. Si se quiere revisar los métodos divisivos se puede consultar los algoritmos de división de Seber (1984), Dillon-Golstein (1984, pag. 178) y Krzanowski-Marriot (1995 pags. 61-94) [D.L, 2002, P.D, 2002].

Los *Métodos aglomerativos* se caracterizan por poseer siempre la misma estructura, solo se diferencian en la forma de calcular las distancias entre grupos. Dicha estructura es:

1. Comenzar con tantas clases como elementos. Las distancias entre clases son las distancias entre elementos.
2. Seleccionar los dos elementos más próximos en la matrix de distancias y formar con ellos una clase.
3. Sustituir los dos elementos utilizados en 2, para definir la clase por un nuevo elemento que represente la clase construída. Las distancias entre este nuevo elemento y los anteriores se calculan dependiendo del tipo de enlace que se explicará ulteriormente.
4. Repetir 2 y 3 hasta que todos los elementos queden ubicados en un solo grupo [R.A, 2002, D.R, 1998, C.H, 2003, P.D, 2002].

Para determinar que grupos se integran se utiliza una función objetivo o criterio que, en el caso de los métodos aglomerativos recibe el nombre de enlace [D.L, 2002, R.T2, 2005, P.D, 2002]. Entre los que se destacan:

- **Enlace simple o “vecino más cercano”**: Mide la proximidad entre dos grupos calculando la distancia entre sus objetos más próximos o la similitud entre sus objetos más semejantes. Este criterio solo depende del orden de las distancias por lo cual será invariante ante transformaciones monótonas, ya que se obtendrá la misma jerarquía a pesar de que las distancias sean numéricamente distintas. La distancia entre el conglomerado \mathcal{A} y el conglomerado \mathcal{B} se define mediante la ecuación (2.17):

$$d_{\mathcal{A}\mathcal{B}} = \min \{d_{ij}\} \text{ donde } i \in \mathcal{A}, j \in \mathcal{B} \quad (2.17)$$

Así la distancia entre dos conglomerados es la menor distancia observada desde un punto de un conglomerado a un punto del otro conglomerado. Este criterio tiende a producir grupos alargados y pueden incluir elementos muy distintos en los extremos.

- **Enlace completo o “vecino más lejano”**: Mide la proximidad entre grupos calculando la distancia entre sus grupos más lejanos o la similitud entre sus objetos menos semejantes.

Este método es el opuesto lógico al del enlace simple. Para este procedimiento la distancia entre el conglomerado \mathcal{A} y el conglomerado \mathcal{B} está dada por la ecuación (2.18):

$$d_{AB} = \max \{d_{ij}\} \text{ donde } i \in \mathcal{A}, j \in \mathcal{B} \quad (2.18)$$

Este procedimiento es también invariante ante transformaciones monótonas. Tiende a producir grupos esféricos.

- **Enlace promedio:** Mide la proximidad entre grupos calculando la media de la distancia entre los objetos de ambos grupos o la media de las similitudes entre los objetos de ambos grupos. Como se ponderan los valores de las distancias, este criterio no es invariante ante transformaciones monótonas. La distancia entre dos conglomerados \mathcal{A} con un tamaño $n\mathcal{A}$ y \mathcal{B} con un tamaño $n\mathcal{B}$ se define como el promedio de las distancias entre todos los pares de objetos, uno del conglomerado \mathcal{A} y otro del conglomerado \mathcal{B} ; es decir:

$$d_{AB} = \frac{1}{n\mathcal{A}n\mathcal{B}} \sum_{i \in \mathcal{A}} \sum_{j \in \mathcal{B}} d_{ij} \quad (2.19)$$

- **Método del centroide:** Calcula la distancia existente entre el centroide de dos grupos, se aplica solo a variable continuas. Se toman como centros los vectores de las medias de las observaciones que pertenecen a un grupo. La distancia entre el grupo \mathcal{A} y el grupo \mathcal{B} se define así:

$$d_{AB}^2 = \sum_{k=1}^p (x_{\mathcal{A}k} - x_{\mathcal{B}k})^2 \quad (2.20)$$

donde $x_{\mathcal{A}k}$ y $x_{\mathcal{B}k}$ son las medias de la variable x_j en los grupos \mathcal{A} y \mathcal{B} respectivamente.

- **Método de Ward:** Con este método se busca la mínima variabilidad entre los grupos, tratándose de un problema de optimización. Ahora se parte de los elementos directamente, en lugar de utilizar la matriz de distancias. Este método se basa en la pérdida de información resultante al agrupar casos en grupos, medida por la suma total del cuadrado de las desviaciones de cada caso al centroide del grupo al cual pertenece. La suma de cuadrados de ward (SCW) se calcula mediante la ecuación (2.21).

$$SCW = \frac{1}{\frac{1}{n\mathcal{A}} + \frac{1}{n\mathcal{B}}} \|\bar{X}_{\mathcal{A}} - \bar{X}_{\mathcal{B}}\|^2 \quad (2.21)$$

Donde $\bar{X}_{\mathcal{A}}$ y $\bar{X}_{\mathcal{B}}$ son los centroides de los conglomerados \mathcal{A} y \mathcal{B} respectivamente. Se empieza con n grupos, un caso por grupo donde la SCW es igual a cero. En el segundo paso se

buscan los casos en los que se produzca el menor incremento en SCW dentro de las posibles combinaciones de a dos objetos. En la tercera etapa se toman los $(n-1)$ grupos conformados, se calcula la SCW y se juntan aquellos que muestren menor incremento. El proceso continua hasta obtener un grupo de n casos u objetos.

Es difícil dar reglas generales para justificar un criterio sobre otro, se recomienda analizar que criterio se acomoda mejor al tipo de datos que se va a manejar, a partir de la comparación de resultados [P.D, 2002].

Una manera de visualizar los grupos formados después de un reconocimiento de patrones de tipo jerárquico es el *dendrograma*¹⁴. La forma de contruir un dendrograma es la siguiente:

1. En la parte inferior del gráfico se disponen los n elementos iniciales.
2. Las uniones entre elementos se indican por tres líneas rectas, dos dirigidas a los elementos que se unen, y que son perpendiculares al eje de los elementos, y una paralela a este eje, que situa el nivel donde se unen.
3. El proceso se repite hasta que todos los elementos están conectados por líneas rectas.

Si se corta el dendrograma a un nivel de distancia dado, se obtiene una clasificación del número de grupos existentes a ese nivel y los elementos que los forman.

El dendrograma es útil en este estudio para ayudar a determinar el número de grupos óptimos y para hacer una idea de cómo quedarían conformados los grupos [G.S, 2002, H.J, 2000].

2.7.2.2. Métodos de partición

Difieren de los métodos jerárquicos en que se conoce *a priori* el número de grupos que se desean formar. Es un método muy utilizado cuando la cantidad de datos que se quieren clasificar es elevado y/o para refinar una clasificación realizada por un método jerárquico. Otra diferencia con las técnicas jerárquicas es que la ubicación de un objeto dentro de un grupo no es definitiva.

Esta técnica de clasificación ofrece las siguientes características:

1. Inician con una partición del conjunto de objetos en algún número específico de grupos. A cada uno de estos grupos se le calcula el centroide.
2. Ubican cada caso u objeto en el conglomerado cuyo centroide este más cercano a éste.
3. Calculan el nuevo centroide de los conglomerados. Estos no son actualizados hasta tanto no se comparen sus centroides con todos sus casos.

¹⁴Del griego *dendron* que significa árbol.

4. Continúan con los pasos 2 y 3 hasta que los casos resulten irremovibles.

▪ *Método de las K-medias*

Se parte de que se divide una muestra de n individuos con p variables en una partición inicial, para después modificar su configuración hasta obtener la fragmentación óptima en k grupos, en términos de una función objetivo. Este método es muy sensible a los valores iniciales dados, por lo que es conveniente utilizar los más adecuados. Una forma de realizarlo es mediante una clasificación obtenida por un algoritmo jerárquico [R.T2, 2005].

La implementación del algoritmo *K-medias* es la siguiente:

1. Seleccionar k centroides, donde k es el número de grupos deseado.
2. Asignar cada observación al grupo cuyo centro es el más cercano.
3. Recalcular los centros de cada grupo.
4. Iterar los pasos 2 y 3 hasta que se satisfaga un criterio de parada.

Para la implementación del algoritmo *k-medias* se puede utilizar una gama variada de medidas de distancia (Euclídea, Manhattan, etc) (ver numeral 2.3.8).

Como se mencionó anteriormente el método *k-medias* requiere tener definido el número de grupos que se quiere implementar. Para esto, el dendrograma sugiere un número de agrupaciones en cada paso, pero esto no define el número óptimo de grupos. Realmente determinar éste número óptimo es tarea del analista de los datos, ya que lo sugiere según la aplicación.

El reconocimiento de patrones no tiene pretensiones inferenciales hacia una población a partir de una muestra, se emplea esencialmente como una técnica exploratoria. Las soluciones no son únicas y además, siempre es posible conformar conglomerados [D.L, 2002].

2.7.3. Impacto del escalamiento

Cuando se realiza un proceso de reconocimiento de patrones es importante tener presente el efecto que tiene la escala de medida de cada una de las variables que se están utilizando, dicho de otra forma, el escalamiento dimensional de un vector característico puede impactar considerablemente los resultados del agrupamiento [MIT, 2003]. El escalamiento puede utilizarse para normalizar las dimensiones, de modo que una distancia métrica simple constituya un criterio razonable para la similitud.

2.7.4. Ejemplo de aplicación del reconocimiento de patrones.

Se presenta a continuación un ejemplo de la aplicación reconocimiento de patrones en un grupo de datos al cual se le ha aplicado un ACP. Como se mencionó en el numeral (2.7.3) es necesario realizar un escalamiento a los datos procesados con ACP para evitar inconvenientes con la escala de medida. Primero se realiza un análisis jerárquico como se muestra en la figura (2.5), esto con el fin de estimar que cantidad de grupos al momento de utilizar un método de partición.

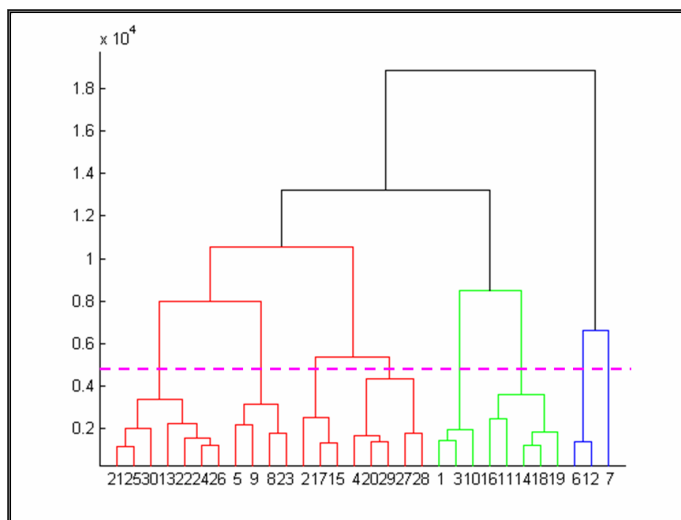


Figura 2.5: Dendrograma utilizando distancia euclídea y enlace de Ward

En la Figura (2.5) se observa que utilizando aproximadamente 8 grupos se tienen comportamientos definidos como se ve en la Figura (2.6).

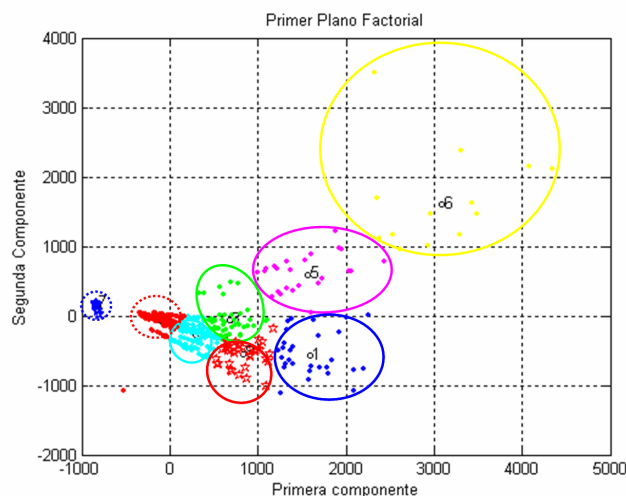


Figura 2.6: Reconocimiento de patrones sobre el primer plano factorial.

2.8. Estado del Arte

La minería de datos, es la exploración y el análisis, ya sea por medios parcial o totalmente automáticos, de grandes cantidades de datos con la meta de descubrir reglas y patrones significativos. Las técnicas para la minería de datos han sido aplicadas a una amplia variedad de campos, desde la aplicación de la ley y la justicia (policía), hasta la radioastronomía, medicina y procesos de control industrial.

El concepto de patrón es universal en el descubrimiento. Los patrones en datos biológicos contienen conocimiento, solo si se logra descubrir. La discriminación de patrones de señales han permitido la identificación personal de voz, escritura a mano alzada y escenas de imágenes. También incluye el reconocimiento de objetivos militares basados en radar, infrarojos y/o imágenes de video. Los patrones existen en sondeos electromagnéticos de alta frecuencia sobre cuerpos químicos, incluyendo el ADN.

Las posibilidades aumentan en geología, meteorología, personalidad, cultura, historia, electromagnetismo y otros datos; que vienen de imágenes microscópicas de celdas a imágenes macroscópicas de las regiones de la tierra obtenidas por sondeo satelital. El reconocimiento de patrones ha contribuido al descubrimiento de asociaciones y relaciones de causa y efecto. La asociación entre patrones y sus causas son el ladrillo desde el cual la ola del conocimiento científico está siendo construido [C.H, 2003].

Algunas de las aplicaciones de técnicas estadísticas multivariadas se muestran a continuación:

- *Remoción del ruido coloro Gaussiano de las imágenes digitales empleando análi-*

sis de componentes principales: En este trabajo se emplea el ACP para remover el ruido coloro gaussiano de las imágenes digitales. El espacio de la imagen entera fue dividido en dos subespacios complementarios, es decir, el subespacio de imagen más el ruido y subespacio del ruido. La estimación fue realizada quitando el subespacio del ruido. El método fue probado en un ancho rango de imágenes reales así como en las imágenes de un fractal sintetizadas. El algoritmo demuestra una supresión del ruido en una gran porción sin la destrucción de detalles anatómicos o funcionales importantes y deja una porción de ruido muy pobre [B.S].

- *Análisis de componentes principales para la generación de claves simétricas*: Este trabajo presenta un nuevo esquema de encriptación de biométricas basado en vectores de rasgo extraídos de un sistema de reconocimiento de cara. Este sistema usa el ACP para generar una clave confidencial simétrica, con la cual se encripta cualquier tipo de dato que contenga información, en una plantilla biométrica. Los datos son disimulados y sólo teniendo una biométrica individual similar se ofrece el vector que es capaz de regenerar la clave correcta. Este esquema se aplica a un sistema que emplea eigencaras para el reconocimiento donde la clase descubierta corregida de una imagen de la muestra puede garantizar la generación corregida de una clave simétrica. Debido a la eficacia del sistema, que es dependiente del algoritmo de reconocimiento de cara, las pruebas mostraron una proporción de 90.4 % de generación importante simétrica corregida, o encriptación satisfactoria del esquema de descripción, para 25 clases de la cara, con 5 imágenes cada uno [C.G, 2004].
- *Segmentación del color en imágenes*: Para la segmentación de colores en imágenes se utilizan algoritmos que generalmente usan la distancia euclidea como medida de similitud, lo que propone ésta nueva metodología es la utilización del ángulo del vector como medida de similitud alterna, Se usan los colores cercanos como medios para segmentar las imágenes. La introducción del ángulo del vector como una medida de similitud requirió la transformación del algoritmo k-medias. La mezcla del algoritmo de ACP se presentó como una manera alternada de acercar los colores (y posiblemente otros datos). Además, algunos de los desarrollos recientes en la matemática no lineal, transforma la idea de una mezcla de componentes principales y la extiende a una mezcla de curvas principales. Los variados colores de la imagen se segmentan empleando un nuevo método cuantitativo [W.S, 2004].
- *Comparación de los sistemas de reconocimiento de cara; análisis de componentes principales (ACP), wavelet de Gabor (GW) y la transformada de wavelet discreta(DWT)*: Las tres técnicas se llevan a cabo en MATLAB que programa el ambiente. La investigación se realiza con las imágenes faciales de la base de datos de FERET. Las imágenes se preprocesan para producir una imagen estandarizada que se emplea para la identificación.

ACP produce una base ortonormal para el espacio de la imagen que extrae los rasgos faciales dominantes y proporciona un reconocimiento excepcional. La técnica de GW es el modelado después de los experimentos biológicos y se emplea para filtrar rasgos de espacio-frecuencia de los puntos clave de la imagen de la cara. La DWT investiga el uso potencial en la extracción de los rasgos faciales y también se aplica a las versiones rotadas de la imagen, por esto aumenta la capacidad de la filtración direccional. Una medida de similitud de los rasgos extraídos de la cara proporciona un reconocimiento que es robusto contra las variaciones en la iluminación[M.M ,2003].

- ***Identificación coherente en los sistemas de potencia mediante el análisis de componentes principales:*** En este estudio se presenta una nueva técnica para hacer una identificación coherente entre los generadores de un gran sistema de potencia interconectado que emplea las medidas de velocidad del generador e información del ángulo de la tensión las barras. Este estudio se basa en la aplicación del ACP a medidas obtenidas de los estudios de simulación que representan ejemplos de eventos intra-área. Los resultados de la aplicación de ACP presenta separadamente la información de las velocidades del generador y los ángulos de las barras. La aproximación del ACP fue capaz de destacar grupos de generadores mostrando características comunes cuando se compara con técnicas de análisis convencionales. [A.K, 2004]

Capítulo 3

RECONOCIMIENTO ESTADÍSTICO DE PATRONES: LOCALIZACIÓN DE FALLAS

3.1. Introducción

Parte de la agenda de muchas empresas distribuidoras de energía eléctrica para fortalecerse en el mercado, consta de proporcionar una alternativa de solución de bajo costo a los problemas de la continuidad del suministro. Con la aplicación de técnicas estadísticas multivariadas a la resolución de problemas asociados a la localización de fallas de cortocircuito, se brindan alternativas orientadas a mejorar la planeación y operación de los sistemas de distribución.

3.2. Planteamiento del problema

Debido a las características técnicas, económicas y operativas de los sistemas de distribución, la implementación de los procedimientos para la localización de fallas se hacen poco eficientes, ya que han sido diseñados para sistemas de transmisión. La poca precisión en la localización de la falla incide en la duración de la atención de la misma reflejándose en la violación de los límites de los índices de calidad (DES¹ y FES²) establecidos por la CREG³ y en el pago de compensaciones por parte de las empresas distribuidoras y operadoras de red. Con la aplicación de las técnicas para el reconocimiento de patrones se ofrece una alternativa para la localización de fallas en sistemas

¹Duración de las interrupciones de energía eléctrica.

²Número de interrupciones de energía eléctrica.

³Comisión de Regulación de Energía y Gas

de distribución, identificando la zona probable de ubicación. Por tanto, es necesario observar el comportamiento de las características más importantes de los grupos encontrados.

3.2.1. Antecedentes históricos de los métodos utilizados para la localización de fallas

Se han desarrollado diversos métodos para la localización de fallas, uno de los métodos más usuales es el de la inspección visual de la línea. Éste método tiene el inconveniente de que se debe realizar a pie o en automóvil y apoyado por el uso de binoculares dependiendo del terreno, ocasionando demoras en la localización e inconvenientes debidos al estado del tiempo.

Los métodos existentes para la localización de fallas en líneas de transmisión se dividen en tres categorías:

- Fenómenos de Ondas Viajeras: Parte de conocer el comportamiento de ondas viajeras de tensión y corriente en condiciones de falla. Es un método preciso pero complejo.
- Componentes de Alta Frecuencia: Se obtienen componentes de alta frecuencia de señales de tensión y corriente de falla. El inconveniente de éste método es que es muy complejo y costoso debido a que requiere el empleo de filtros especiales para su medición.
- Frecuencia Fundamental: Emplea la frecuencia fundamental de señales de tensión y corriente en los terminales de una línea, y determina la impedancia vista desde estos terminales, y a su vez la compara con los parámetros de la línea para estimar una distancia desde el terminal de la línea hasta la falla.

Para localización de fallas en sistemas de distribución se adaptan los principios de los métodos utilizados en sistemas de transmisión teniendo en cuenta algunas consideraciones por las diferencias existentes entre los dos sistemas. Otro método utilizado consiste en colocar indicadores en cada sección de la línea provistos de enlaces de comunicación por radio, pero este método se hace poco viable a medida que el sistema de distribución se hace más extenso [D.R, 1998].

3.3. Metodología propuesta

Una forma de resolver el problema de la localización es conocer el comportamiento del sistema cuando se encuentra sometido a diferentes perturbaciones. 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 de corriente. 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 realiza la descripción de la metodología propuesta.

En la primera etapa, se selecciona un sistema de potencia prototipo (ver apéndice B) en el cual se pueda representar las características de los sistemas de distribución, esto es, topología radial, secciones de línea y cargas monofásicas y trifásicas, derivaciones y se supone que el equipo de medida está ubicado en la cabecera del circuito. Las fallas más comunes en sistemas de distribución son el resultado de cortocircuitos. En este trabajo se tuvieron presentes fallas del tipo monofásica a tierra, bifásica, bifásica a tierra, y trifásica, cada una con diferentes valores de resistencia de falla⁴. En la segunda etapa, se implementa el sistema prototipo en el software ATP/EMTP con el objetivo de realizar simulaciones, sometiénolo a las condiciones de falla planteadas anteriormente, 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 plantea la caracterización del sistema de dos formas diferentes. La primera, de tipo determinista basada en el cálculo de descriptores. La segunda, de tipo estadística la cual está basada en el análisis exploratorio de datos a partir de la aplicación de técnicas como el análisis de componentes principales - ACP, 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 homogeneidad entre los grupos, esto se logra con la aplicación de técnicas para el análisis de conglomerados. El conjunto de datos recolectados puede ser útil debido a que permite establecer de manera probable el lugar de ocurrencia de una falla luego que ésta sucede. En la figura (3.1) se presenta el diagrama de bloques de la metodología propuesta.

⁴7 valores de resistencia de falla para entrenamiento y 4 valores de resistencia para validación

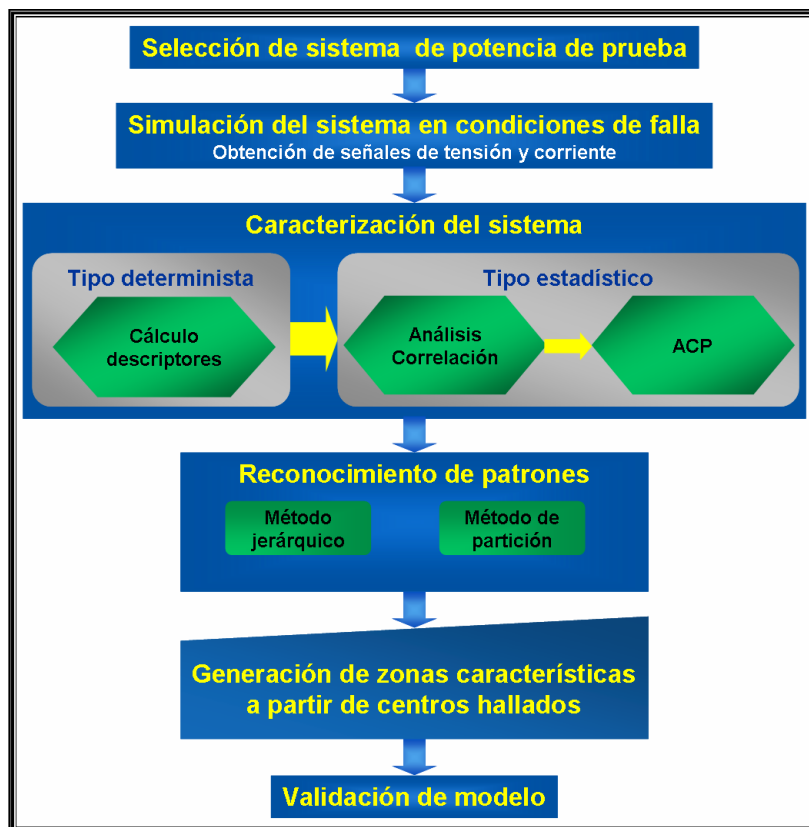


Figura 3.1: Metodología propuesta.

3.3.1. Selección del sistema de potencia de prueba

El sistema de potencia utilizado para realizar las simulaciones de cortocircuito es el especificado en el anexo B.

3.3.2. Simulación del sistema de potencia en condiciones de falla

Para poder extraer información de como se comporta el sistema de potencia bajo la ocurrencia de cualquier tipo de falla⁵, se utilizó el software ATP/EMTP, en donde se colocaron medidores de tensión y de corriente en la cabecera del circuito, es decir las mediciones se realizan en el alimentador. Se simuló cada tipo de falla en todas las barras del sistema utilizando once valores

⁵Ver Anexo C

de resistencia de falla, los cuales se muestran en el cuadro 3.1.

Cuadro 3.1: Resistencias de falla empleadas en la simulación

| Nombre | Valor [Ω] | Nombre | Valor [Ω] | Nombre | Valor [Ω] | Nombre | Valor [Ω] |
|-----------|--------------------|-----------|--------------------|-----------|--------------------|------------|--------------------|
| R1 | 0.05 | R4 | 15 | R7 | 30 | R10 | 45 |
| R2 | 5 | R5 | 20 | R8 | 35 | R11 | 50 |
| R3 | 10 | R6 | 25 | R9 | 40 | - | - |

Para la obtención de los descriptores de tipo determinista se emplearon los resultados extraídos de las simulaciones realizadas con las resistencias R1, R2, R4, R6, R8, R10 y R11. Los demás resultados de la simulación se emplearán para realizar la validación de la metodología propuesta.

3.3.3. Obtención de los descriptores de tipo determinista.

En el cálculo de los descriptores de tipo determinista se tiene presente que los instrumentos utilizados para realizar la adquisición de señales en los alimentadores, no siempre están en capacidad de almacenar todos los datos requeridos; es decir algunos instrumentos solo realizan mediciones de tensión y otros solo de corriente. Por esta razón, se presentan tres opciones para desarrollar la metodología.

- Datos de tensión y corriente⁶.
- Datos de solo corriente.
- Datos de solo tensión.

En cada una de las opciones propuestas se realizó un análisis por separado a cada uno de los tipos de falla de cortocircuito:

- Línea-Tierra (L-T).
- Línea-Línea (L-L).
- Línea-Línea-Tierra (L-L-T).
- Línea-Línea-Línea (L-L-L).

⁶A éste modelo se le denominará *Tensión-Corriente*.

La selección de descriptores del sistema se ha hecho bajo la premisa de que sean sencillos de obtener, bien sea directa o indirectamente y que el tiempo no sea un factor importante. También están calculados en estado estable de la falla y a una frecuencia única⁷.

3.3.3.1. Modelo tensión-corriente

Para éste modelo se tienen cuatro matrices por tipo de falla, y cada una contiene cuatro descriptores como se muestra en el cuadro (3.3):

Cuadro 3.3: Variables utilizadas en el modelo Tensión-Corriente

| Variable | Convención |
|--|------------|
| Mínimo valor de la tensión por fase (p.u) | MTF |
| Máximo valor de la corriente por fase. (p.u) | MCF |
| Potencia Activa consumida en la falla. (p.u) | PF |
| Potencia Reactiva consumida en la falla. (p.u) | QF |

- **MTF**: Indica el mínimo valor registrado entre las tres tensiones de fase durante el estado estable de la falla. (ver fig. 3.2 (a)). $MTF = \min(mTfa, mTfb, mTfc)$
- **MCF**: Es el máximo valor de corriente de cortocircuito en estado estable entre las tres fases (ver fig. 3.2 (b)). $MCF = \max(mcfa, mcfb, mcfc)$
- **PF**: Mide la exigencia que sufre el alimentador del sistema. Da una medida de la potencia activa que demanda el cortocircuito al sistema. Su cálculo se realizó de la siguiente forma: $PF = \frac{\text{real}(\mathbf{V} * \mathbf{I}^*)}{3}$ donde \mathbf{V} es un vector fila que contiene los fasores de tensión de fase en el estado estable de falla, así: $\mathbf{V} = [V_a, V_b, V_c]$ e \mathbf{I} es un vector columna donde se encuentran las corrientes de falla en estado estable $\mathbf{I} = \begin{bmatrix} I_a \\ I_b \\ I_c \end{bmatrix}$.
- **QF**: Se calcula de forma similar al anterior $QF = \frac{\text{Imag}(\mathbf{V} * \mathbf{I}^*)}{3}$ y complementa la exigencia que sufre el alimentador del sistema ante la ocurrencia de una falla.

⁷60 Hertz.

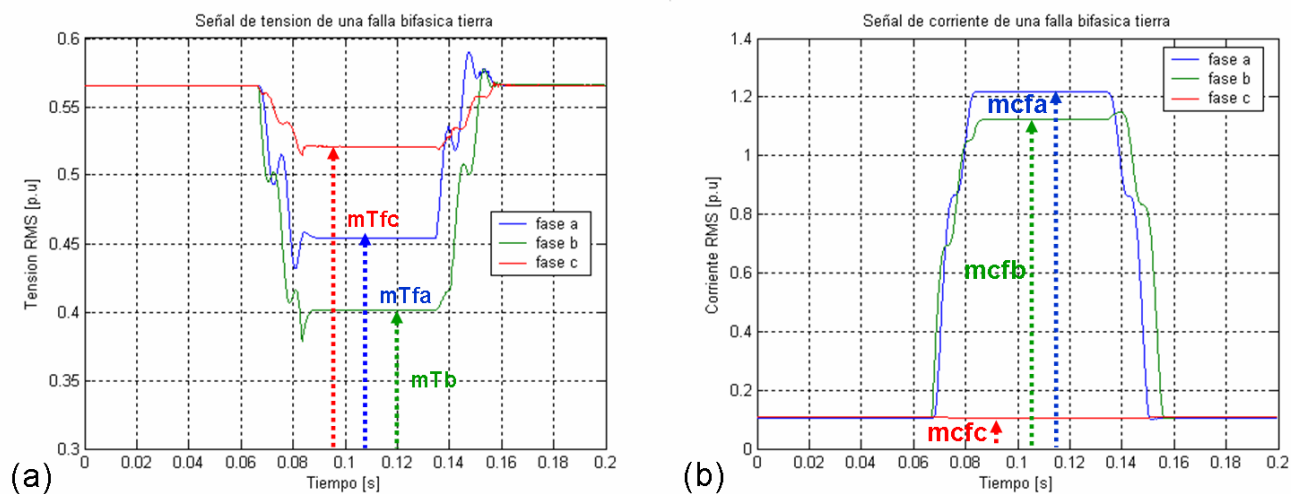


Figura 3.2: (a)Mínima tensión de fase. (b)Máximo valor de corriente de falla.

3.3.3.2. Modelo solo corriente

Igual que en el caso anterior en éste modelo se tienen cuatro matrices por tipo de falla, que contienen cuatro descriptores como se muestra en el cuadro 3.5:

Cuadro 3.5: Variables utilizadas en el modelo de corriente

| Variable | Convención |
|--|------------|
| Máximo valor de la corriente por fase. (p.u) | MCF |
| Promedio de alturas de las corrientes. (p.u) | PHC |
| Desviación estándar de las corrientes. (p.u) | DEC |
| Secuencia cero de la corriente. (p.u) | SCC |

- **MCF:** Es el mismo descriptor propuesto en el modelo Tensión-Corriente, y como se mencionó es el máximo valor de corriente de cortocircuito en estado estable entre las tres fases (ver fig. 3.2 (b)).
- **PHC:** El promedio de las corrientes indica el valor de corriente como si la falla hubiese sido trifásica, o en un equivalente monofásico.
- **DEC:** La desviación estandar muestra que tan lejos o cerca están las corrientes de PHC.
- **SCC:** La secuencia cero mide el desbalance del sistema ante un evento de falla.

3.3.3.3. Modelo solo tensión

El cuadro que muestra los descriptores que se utilizaron para éste modelo se exponen en el cuadro 3.7:

Cuadro 3.7: Variables utilizadas en el modelo de tensión

| Variable | Convención |
|---|------------|
| Mínima altura de Tensión. (p.u) | MTF |
| Promedio de alturas de las Tensiones. (p.u) | PDT |
| Desviación estándar de las Tensiones. (p.u) | DET |
| Secuencia cero de la Tensión. (p.u) | SCT |

- **MTF:** Mide el impacto del cortocircuito pero a la inversa del modelo de corriente, es decir, es más severo si éste valor es más pequeño.(ver fig. 3.2 (a))
- **PHT:** El promedio de las tensiones hace el mismo supuesto que el modelo de corriente.
- **DET:** La desviación estandar muestra que tan lejos o cerca está la suposición anterior.
- **SCT:** La secuencia cero de tensión mide el desbalance del sistema ante un evento de falla

3.3.4. Cálculo de descriptores de tipo estadístico

Después de haber calculado los descriptores de tipo determinista, es necesario realizar un análisis a partir de técnicas estadísticas multivariadas, esto para aprovechar las ventajas que se mencionaron en la sección 2.6.1. Este análisis se aplicará a los tres modelos propuestos como se muestra a continuación.

3.3.4.1. Modelo tensión-corriente

- **Análisis de correlación bivariada modelo tensión-corriente:** Para determinar la posibilidad de utilizar el ACP en los datos que se tienen, es necesario determinar la correlación entre las variables (descriptores) existentes. La correlación existente entre los descriptores, calculada por tipo de falla se muestra en la figura 3.3.

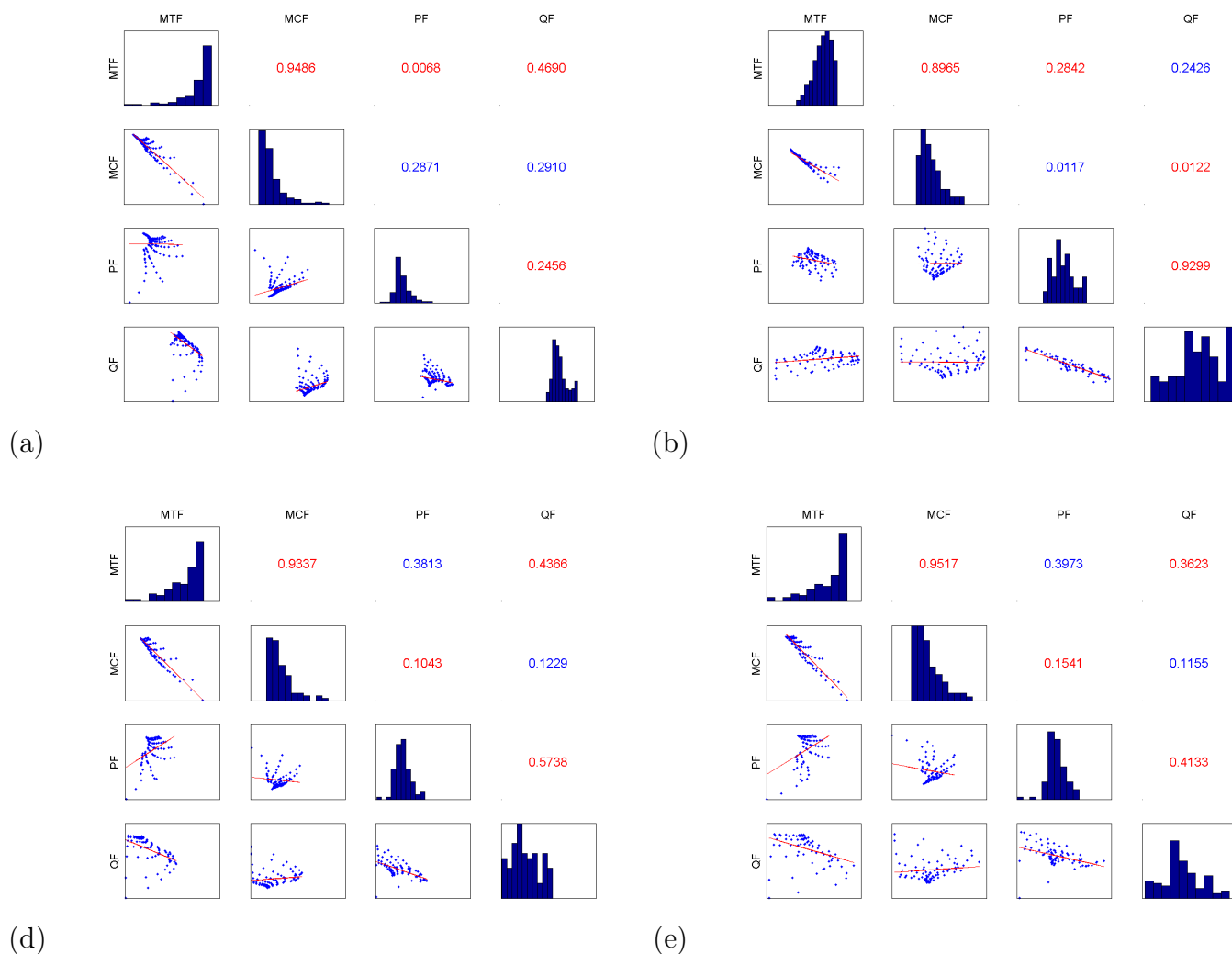


Figura 3.3: Correlación modelo Tensión-Corriente. (a)L-T. (b)L-L. (c)L-L-T. (d)L-L-L.

Se observa una alta interdependencia lineal entre las dos primeras variables en todos los tipos de falla, el resto de las variables tienen poca correlación pero no despreciable. La falla bifásica muestra alta correlación entre la potencia activa y la potencia reactiva. Este comportamiento justifica el uso de ACP para reducir la dimensionalidad de las variables.

- ACP modelo tensión-corriente:** El ACP se realiza sobre la matriz de covarianza \mathbf{S} de las variables originales obtenidas de las señales de tensión y de corriente, la razón es que los descriptores empleados están normalizados al sistema de medida por unidad [p.u]. El resultado de la aplicación de ACP a las variables del modelo Tensión-Corriente se observa en la figura 3.4.

Se puede observar en la figura 3.4, que para cada uno de los tipos de falla la mayor parte de

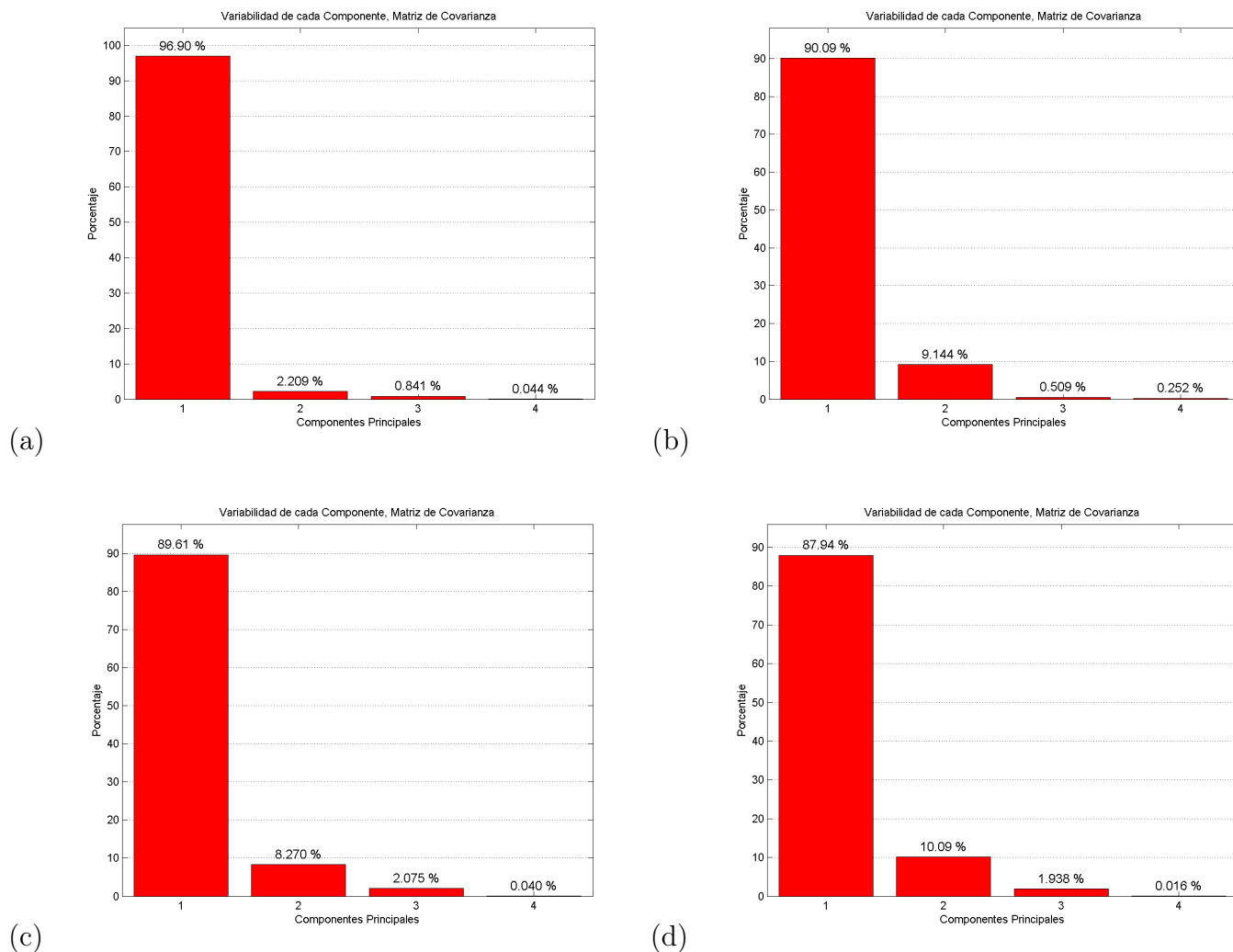


Figura 3.4: Porcentaje de información que contiene cada componente principal modelo Tensión-Corriente. (a)L-T. (b)L-L. (c)L-L-T. (d)L-L-L.

la información está contenida en las dos primeras componentes (entre el 98,03 % y 99,95 % acumulado) y la máxima pérdida de información que sufren los datos es de 2,12 % por lo que para cuestiones de visualización se puede trabajar con el primer plano factorial.

3.3.4.2. Modelo corriente

- Análisis de correlación modelo corriente:** De igual manera que en el modelo Tensión-Corriente, es importante determinar que tanta dependencia lineal ofrecen los descriptores propuestos en este modelo.

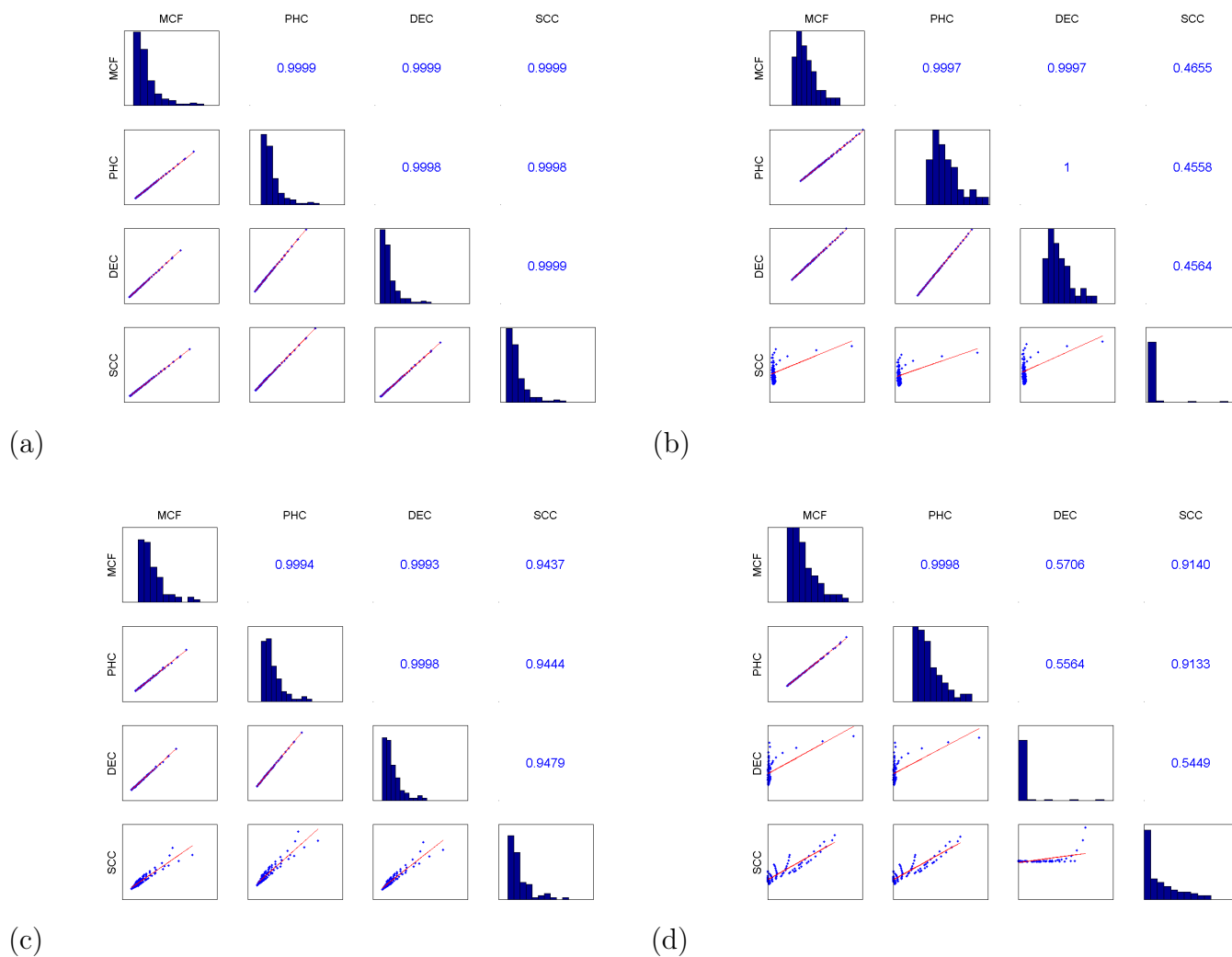


Figura 3.5: Correlación modelo corriente. (a)L-T. (b)L-L. (c)L-L-T. (d)L-L-L.

En la figura 3.5 se observa la evidente interdependencia entre las variables, esto debido a que al momento de calcular los descriptores, se utilizaron operaciones lineales (Promedio y desviación típica). La consecuencia de esto es que al momento de realizar el ACP se van a obtener resultados óptimos en lo referente a la disminución de la dimensionalidad.

- **ACP modelo corriente:** El uso de ACP está más que justificado debido a la alta interdependencia de los descriptores.

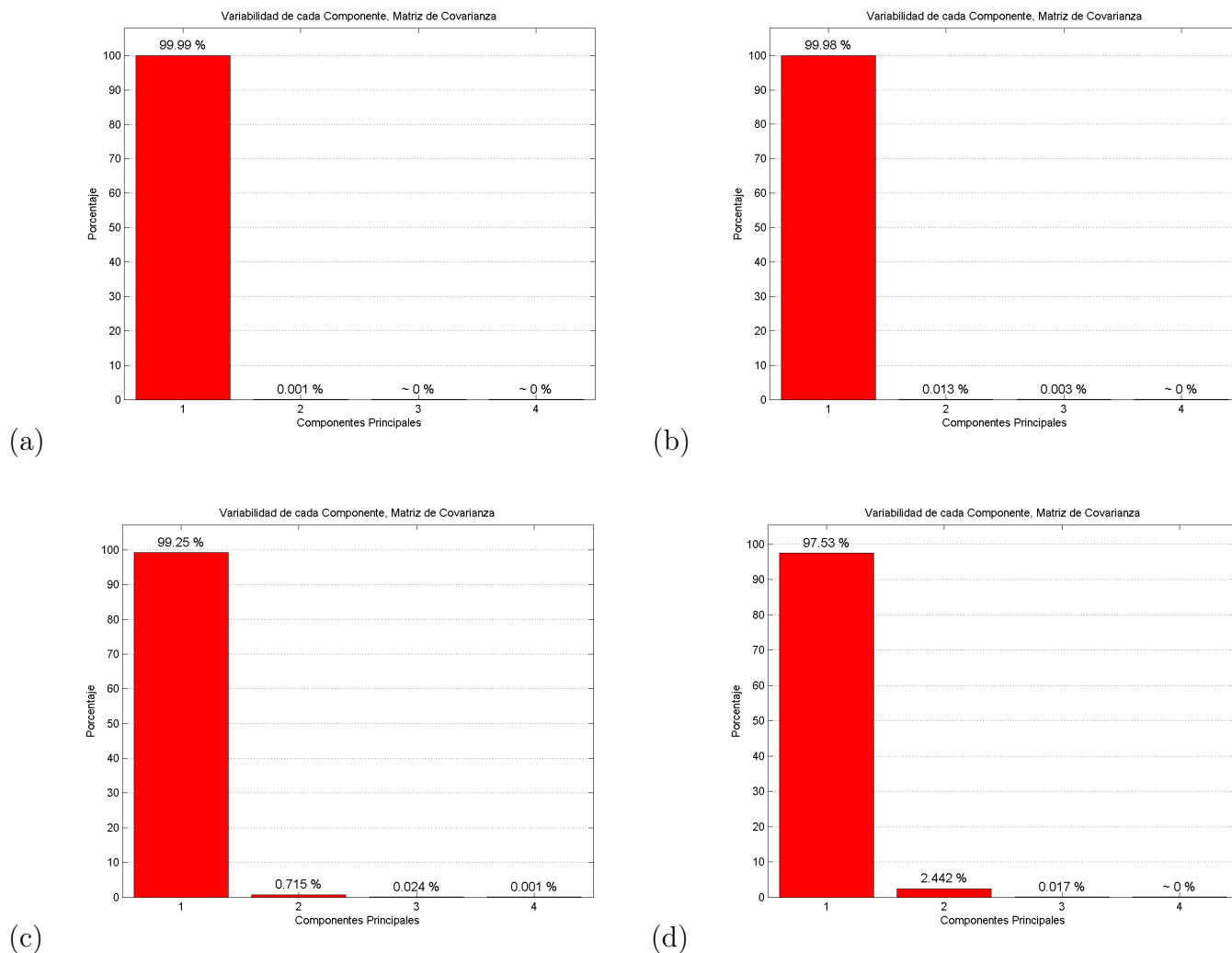


Figura 3.6: Porcentaje de información que contiene cada componente principal modelo corriente. (a)L-T. (b)L-L. (c)L-L-T. (d)L-L-L.

Como se mencionó al momento de analizar la correlación bivariada, el empleo de ACP para reducir dimensionalidad fué óptimo, pues como se observa en la figura (3.6), existe un 99,99 % de la información contenida en la primera componente y solo un 0,01 % de pérdida de información. Para fines de visualización se optó por trabajar con las dos primeras componentes principales.

3.3.4.3. Modelo de tensión

- **Análisis de correlación modelo tensión:** Es de esperarse que los descriptores estén altamente correlados como lo fueron en el modelo de corriente, esto como consecuencia de que las relaciones utilizadas para su cálculo fueron las mismas utilizadas en el modelo anterior.

Esto de ve reflejado en la figura 3.7.

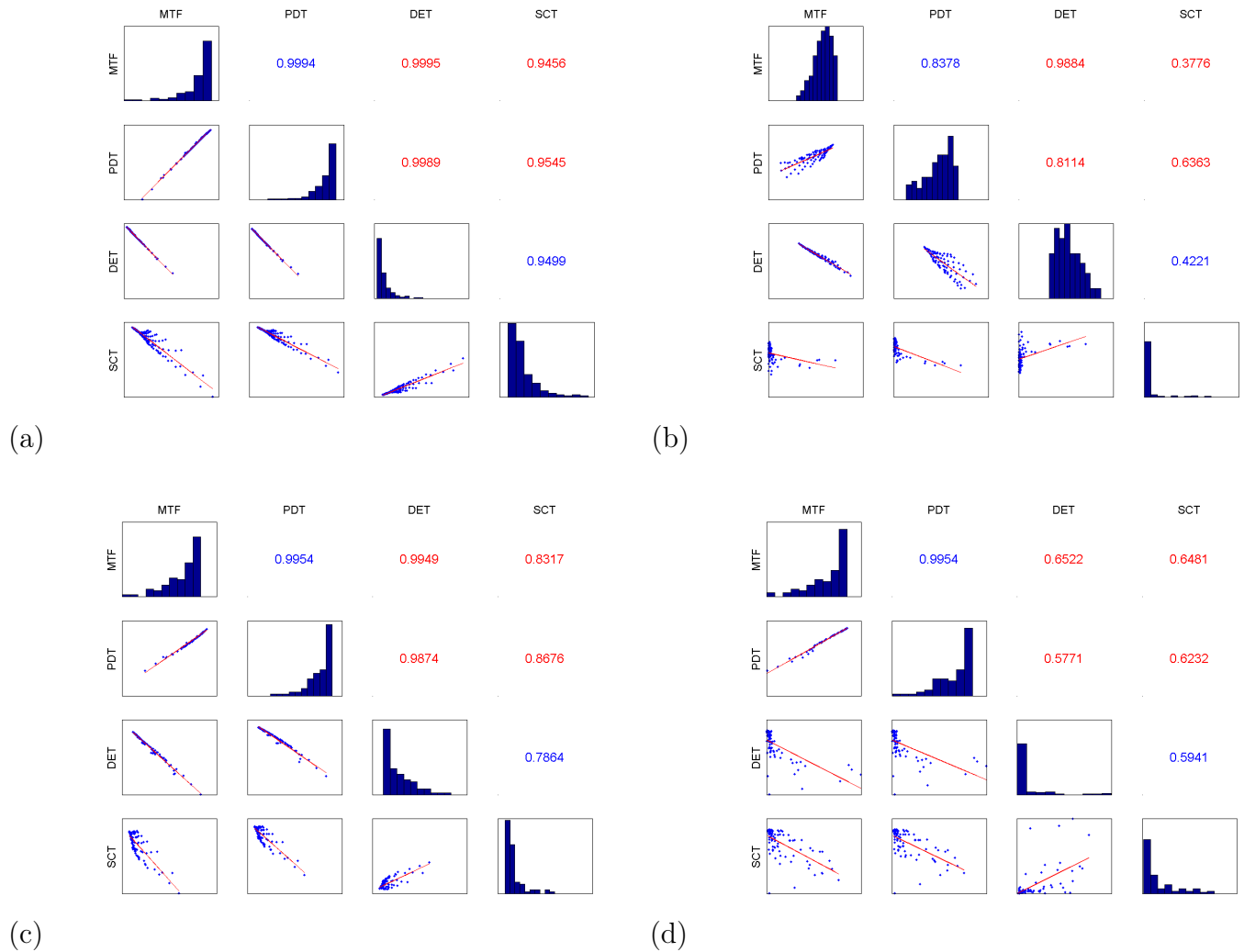


Figura 3.7: Correlación modelo tensión. (a)L-T. (b)L-L. (c)L-L-T. (d)L-L-L.

- **ACP modelo tensión:**La mayor parte de la variabilidad se explica en el primer plano factorial en todos los tipos de falla. En el peor de los casos la pérdida de información es 1,95% en la falla bifásica. Esto se puede observar en la figura 3.8.

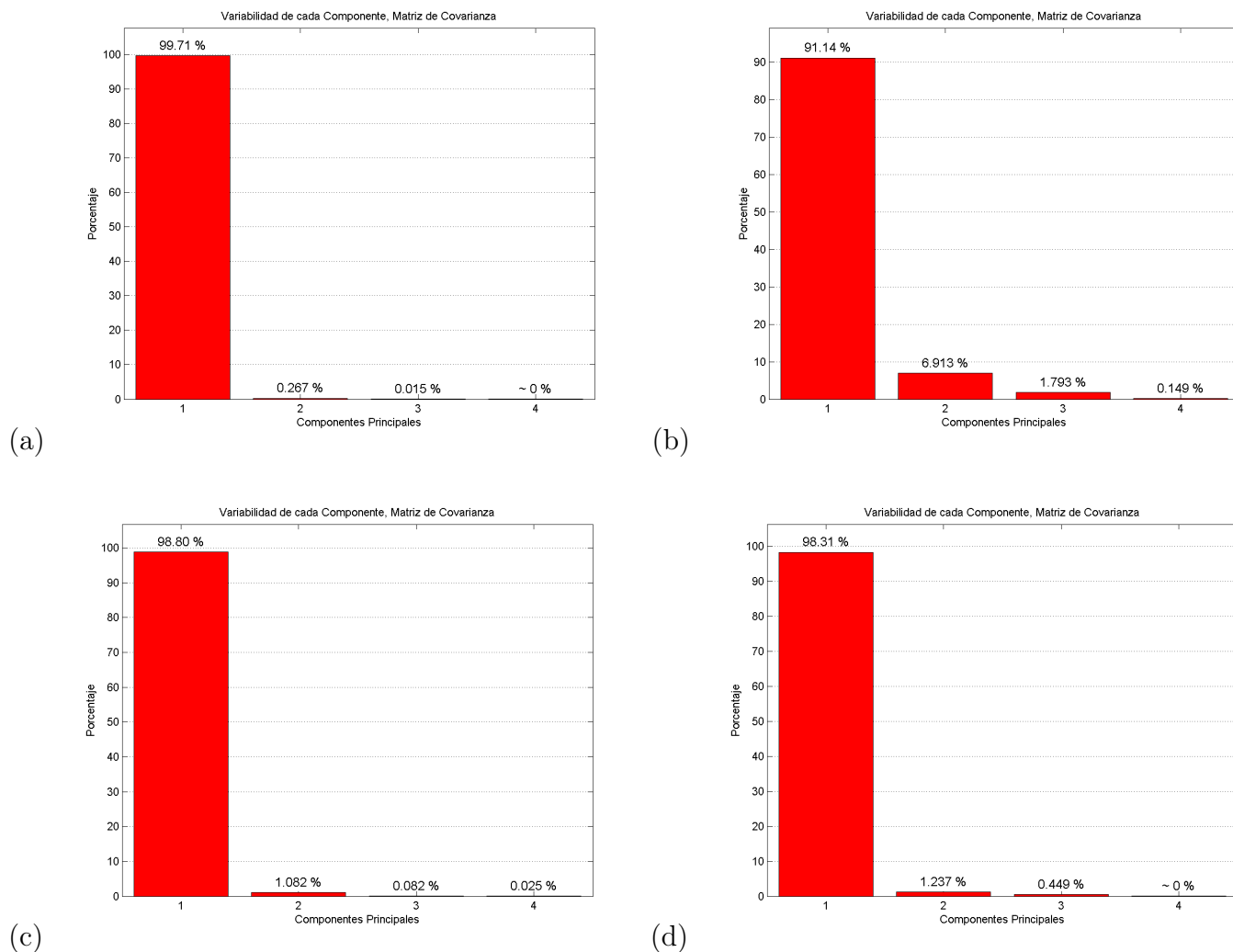


Figura 3.8: Porcentaje de información que contiene cada componente principal modelo tensión. (a)L-T. (b)L-L. (c)L-L-T. (d)L-L-L.

3.3.5. Reconocimiento de patrones

Con la construcción del modelo estadístico se pretende encontrar patrones que permitan la localización de 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 evento registrado, viene acompañado de información relevante a la hora de clasificar cada dato dentro de una clase establecida (zona dentro del sistema de distribución).

Para realizar el reconocimiento de patrones se utilizaron las dos primeras componentes halladas con ACP, ya que estas contienen un gran porcentaje de la información. Se utilizó un método de partición, seleccionando los centros según criterio personal, previamente realizado un proceso

automático que sugirió la formación de grupos.

Es importante aclarar que para poder realizar el agrupamiento fué necesario escalar la segunda componente, esto debido a que se requiere que la escala de los nuevos ejes tengan el mismo rango para poder hacer buen uso de las métricas. El factor de escala se calculó con la ecuación 3.1 , para luego ser aplicado a la ecuación 3.2.

$$f_{escala} = \frac{\frac{mx(score1)}{mx(score2)} + \frac{min(score1)}{min(score2)}}{2} \quad (3.1)$$

$$score2 = f_{escala} * score2 \quad (3.2)$$

donde:

1. $score1$, $score2$: Son las nuevas coordenadas de los datos transformados sobre las componentes principales.
2. $mx(score1)$: Es el máximo valor de los puntajes de la primera componente principal.
3. $mx(score2)$: Es el máximo valor de los puntajes de la segunda componente principal.
4. $min(score1)$: Es el mínimo valor de los puntajes de la primera componente principal.
5. $min(score2)$: Es el mínimo valor de los puntajes de la segunda componente principal.

Habiendo realizado el escalamiento respectivo se procedió a realizar los agrupamientos.

3.3.5.1. Modelo tensión-corriente

- **Falla L-T:** Para realizar el agrupamiento se tomó el primer plano factorial hallado en ACP, en el cual se observan los datos distribuidos formando 7 ramales, correspondientes a las resistencias de falla utilizadas para el entrenamiento del modelo. Los resultados obtenidos después de agrupar se muestran en la gráfica 3.9 (a), en donde se distinguen 49 grupos. La razón para que el número de grupos sea tan grande es que los datos en cada ramal se aglomeran en 6 zonas bien definidas y dos zonas más cuando hay fallas monofásicas, pero al ser 7 ramales se necesitan 56 grupos. El motivo por el cual no se usan los 56 grupos es que los datos que están más alejados del alimentador, están más cerca entre ellos indiferente de la resistencia de falla, lo que produce que se formen grupos que contienen diferentes resistencias.
- **Falla L-L:** Como en el caso anterior se revisa el plano factorial arrojado por el ACP, se obtienen los siguientes 28 grupos. Ver figura 3.9 (b)

- **Falla L-L-T:** Para éste caso se crearon 31 grupos, los que se muestran en la figura 3.10 (a).
- **Falla L-L-L:** Para la falla trifásica se crearon 38 grupos como se ve en la figura 3.10 (b).

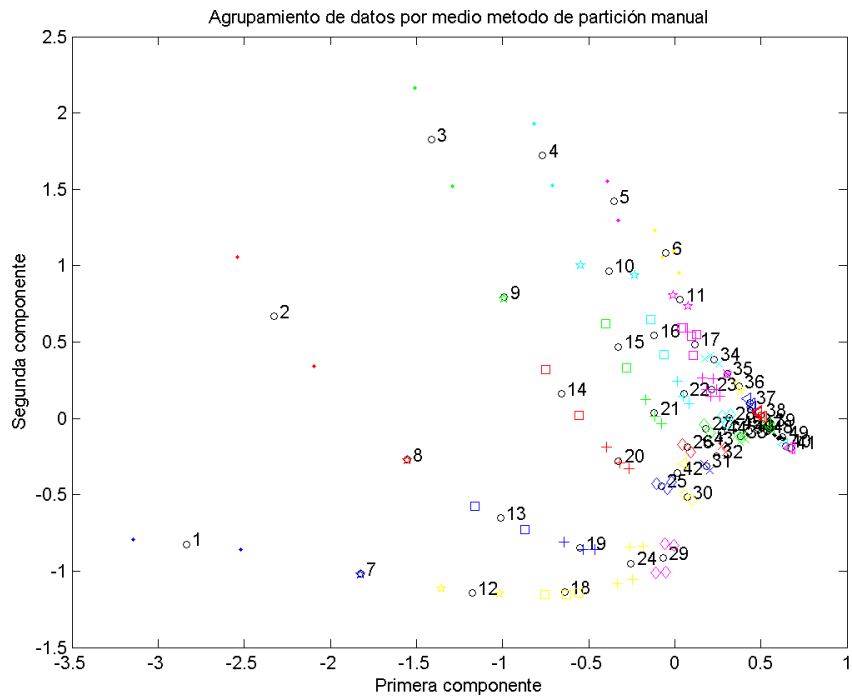
3.3.5.2. Modelo corriente

Al igual que en el modelo Tensión-Corriente, en el modelo de corriente también se realizó el escalamiento necesario antes de proceder con el análisis de conglomerados.

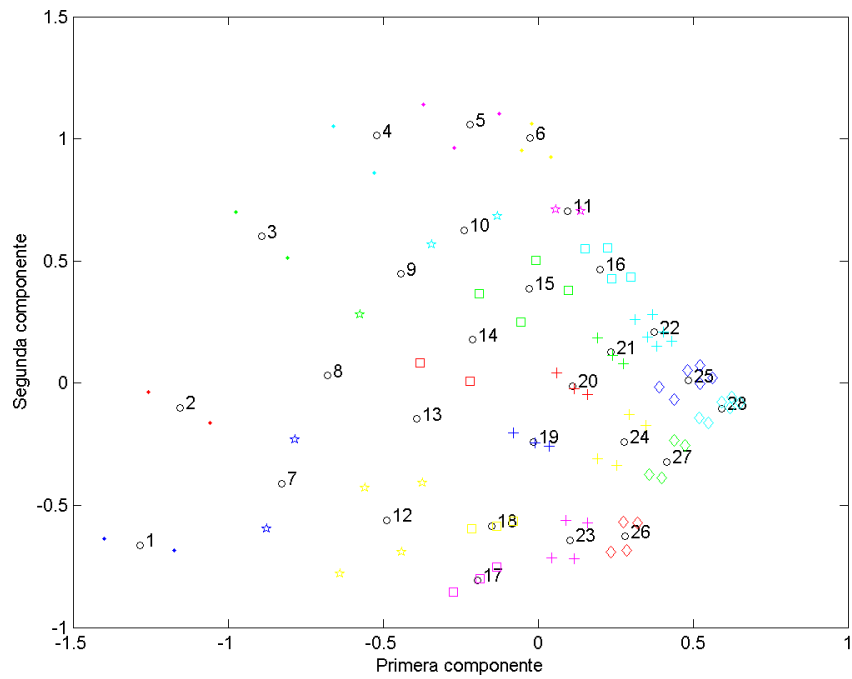
- **Falla L-T:** Se formaron 44 grupos sobre el primer plano factorial entregado por el ACP (Fig 3.11 (a)).
- **Falla L-L:** Para éste tipo de falla se formaron 32 grupos (Fig. 3.11 (b))
- **Falla L-L-T:** Se realizan 36 grupos como lo muestra la figura (3.12 (a)).
- **Falla L-L-L:** La grafica con los 36 grupos formados se muestra en la figura (3.12 (b)).

3.3.5.3. Modelo tensión

- **Falla L-T:** La figura (3.13 (a)) muestra la generación de 48 grupos.
- **Falla L-L:** Al desarrollar un método de aglomeración de partición manual sobre el plano factorial se consigue formar 31 grupos tal y como se ve en la figura (3.13 (b)).
- **Falla L-L-T:** Debido a la alta concentración de datos en la parte baja de la figura (3.14 (a)), el agrupamiento se puede hacer tedioso, sin embargo se crearon 31 grupos.
- **Falla L-L-L:** Presenta el mismo comportamiento que en los casos anteriores y se crearon 36 grupos como se observa en la figura (3.14 (b))

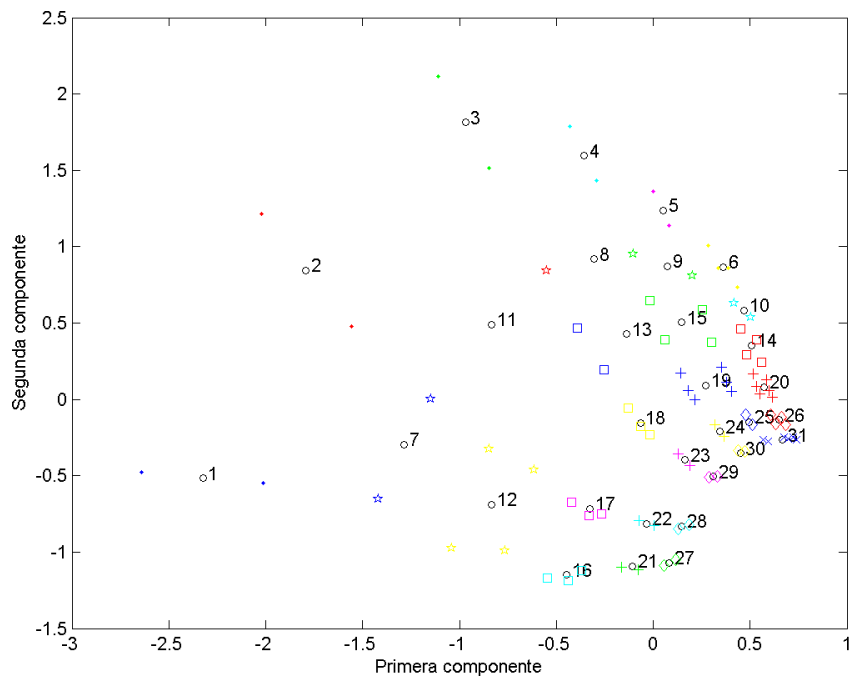


(a)

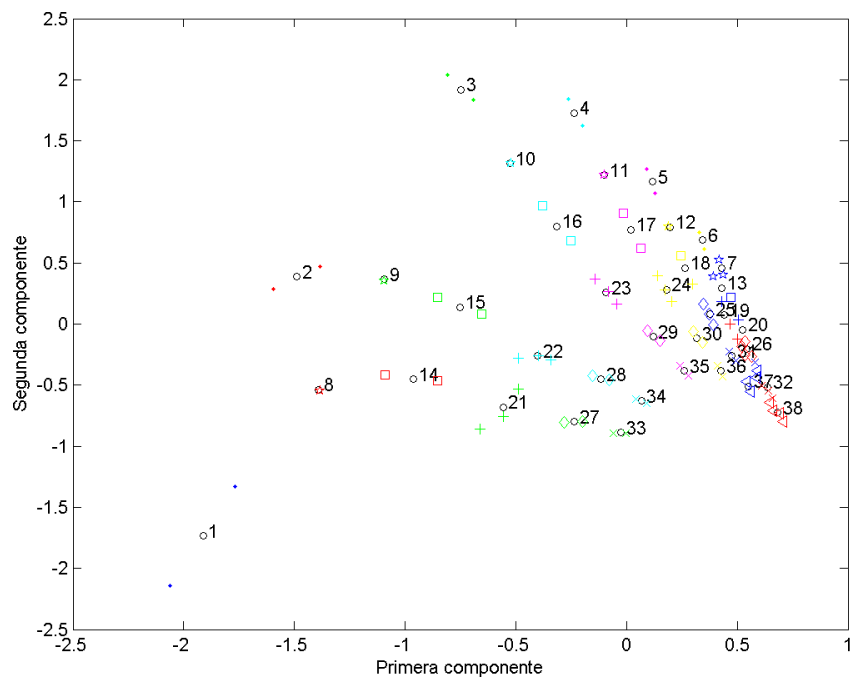


(b)

Figura 3.9: Grupos generados para el modelo Tensión-Corriente. (a)L-T. (b)L-L.

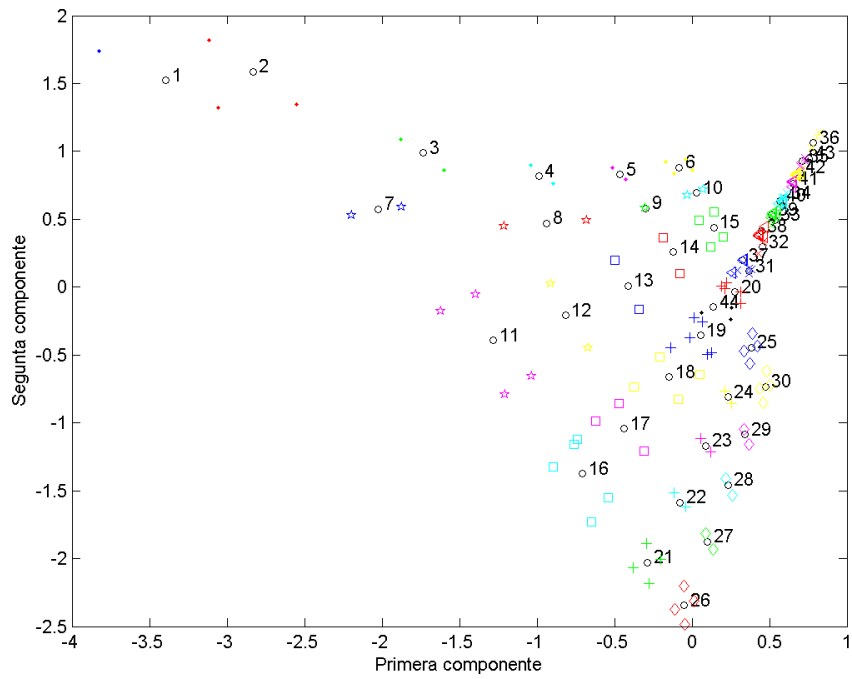


(a)

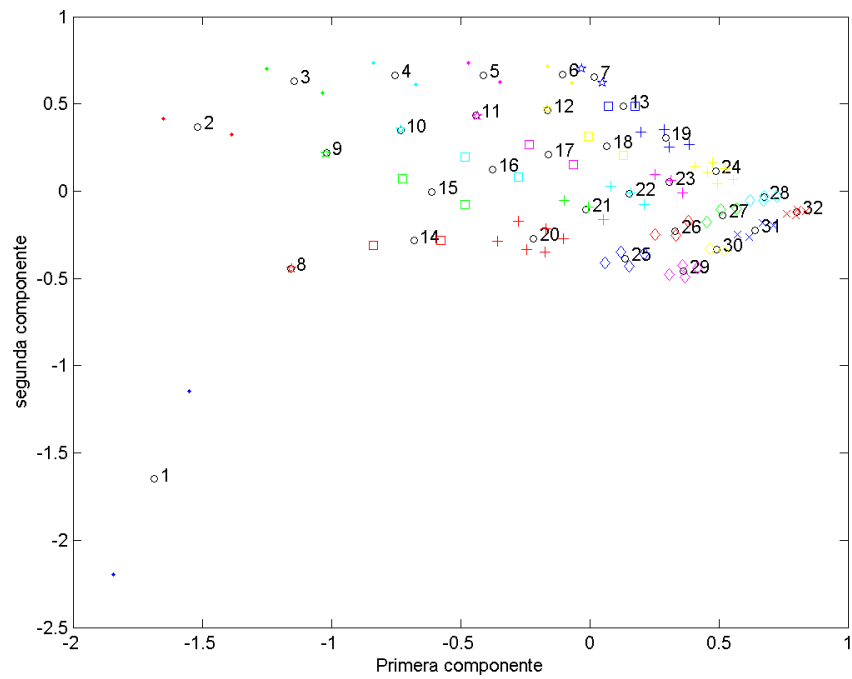


(b)

Figura 3.10: Grupos generados para el modelo Tensión-Corriente. (a)L-L-T. (b)L-L-L.

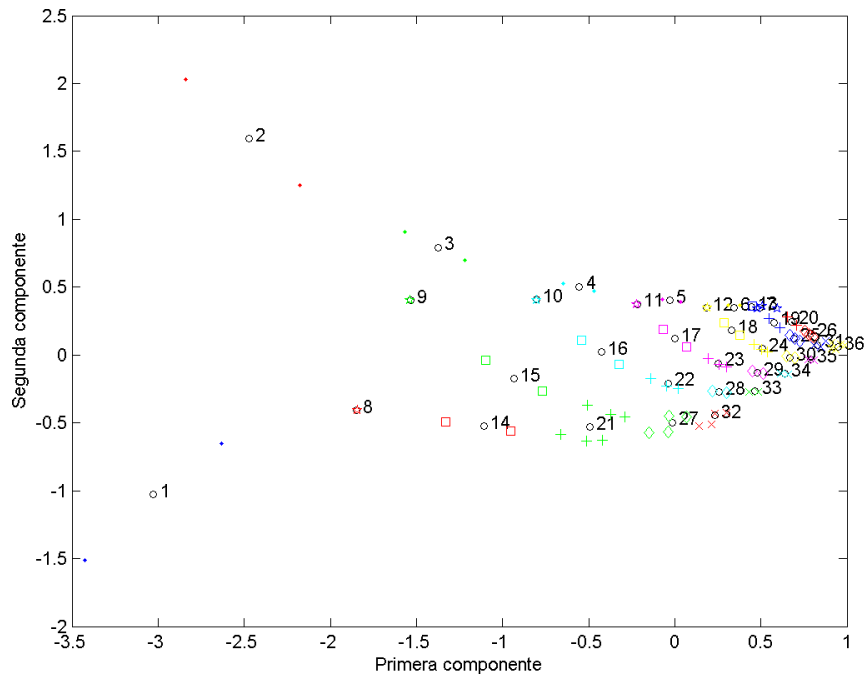


(a)

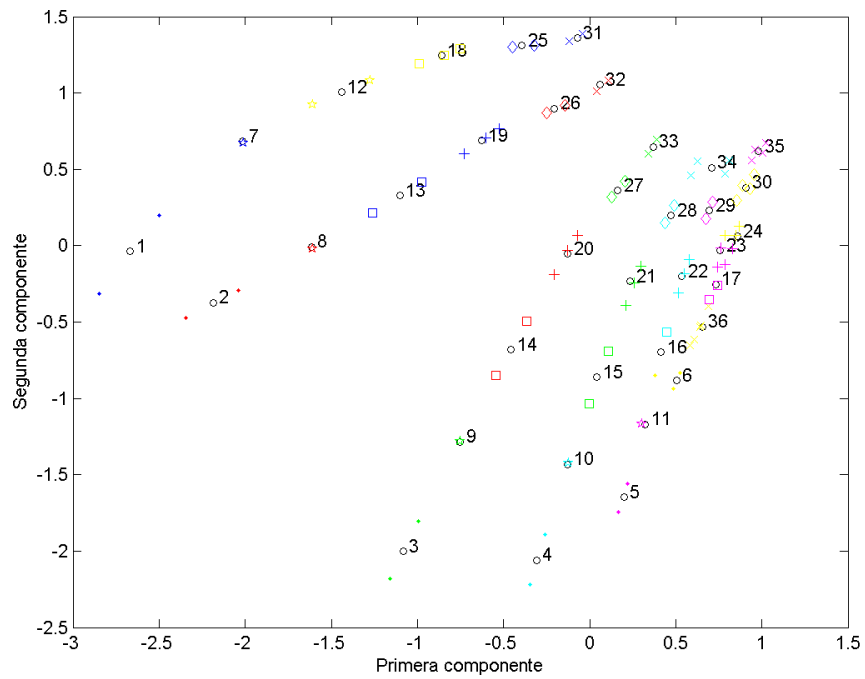


(b)

Figura 3.11: Grupos generados para el modelo Corriente. (a)L-T. (b)L-L.

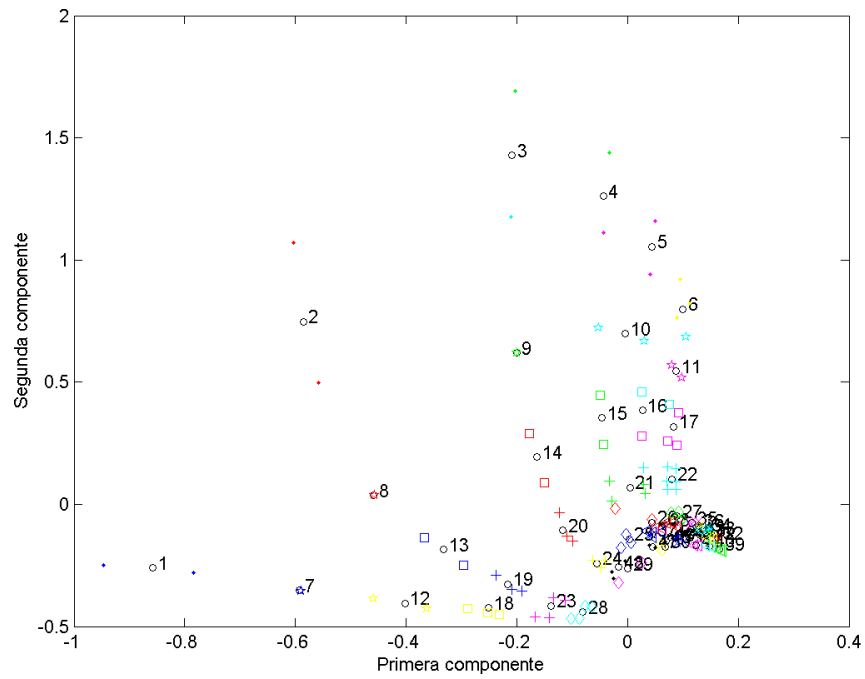


(a)

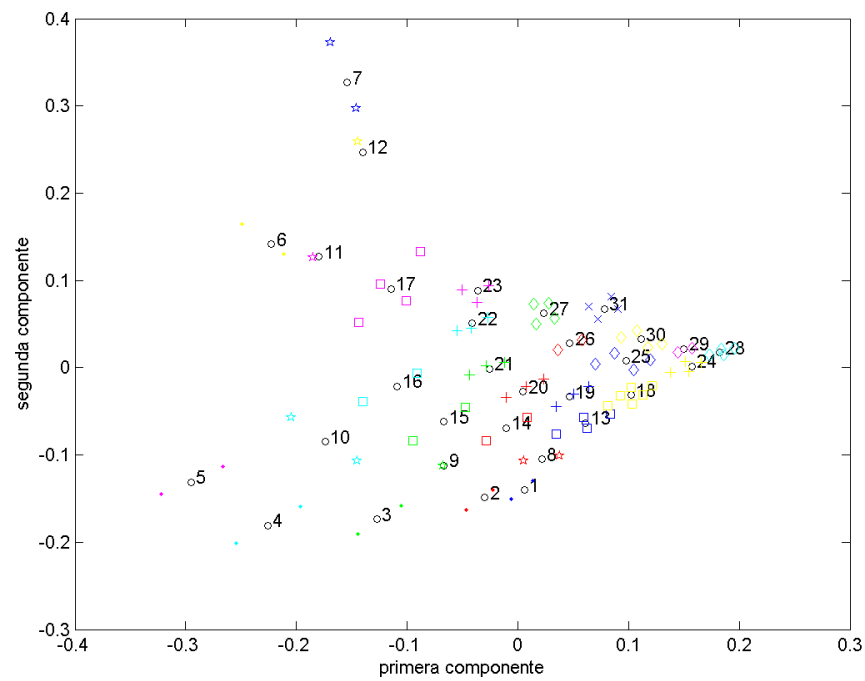


(b)

Figura 3.12: Grupos generados para el modelo Corriente. (a)L-L-T. (b)L-L-L.

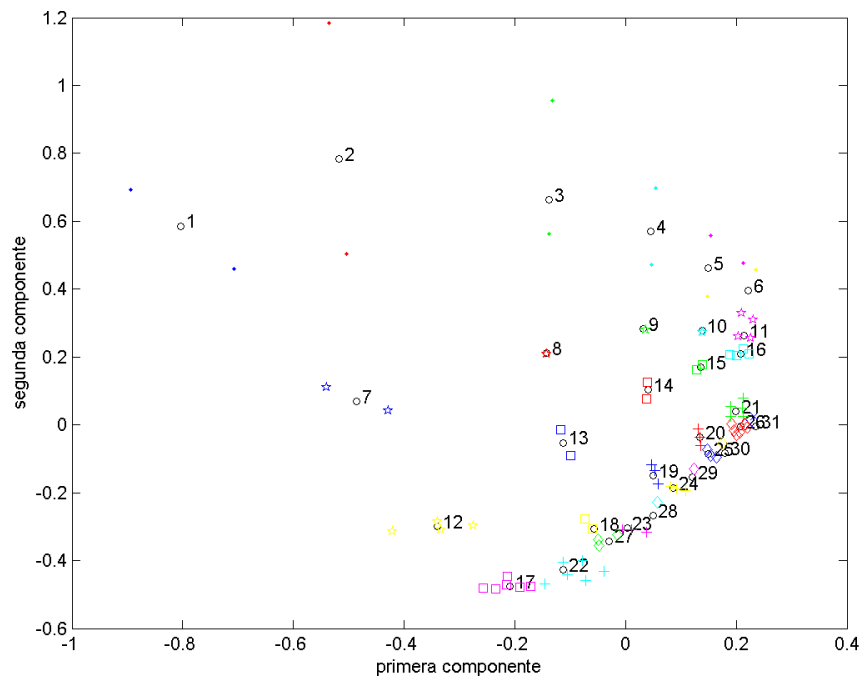


(a)

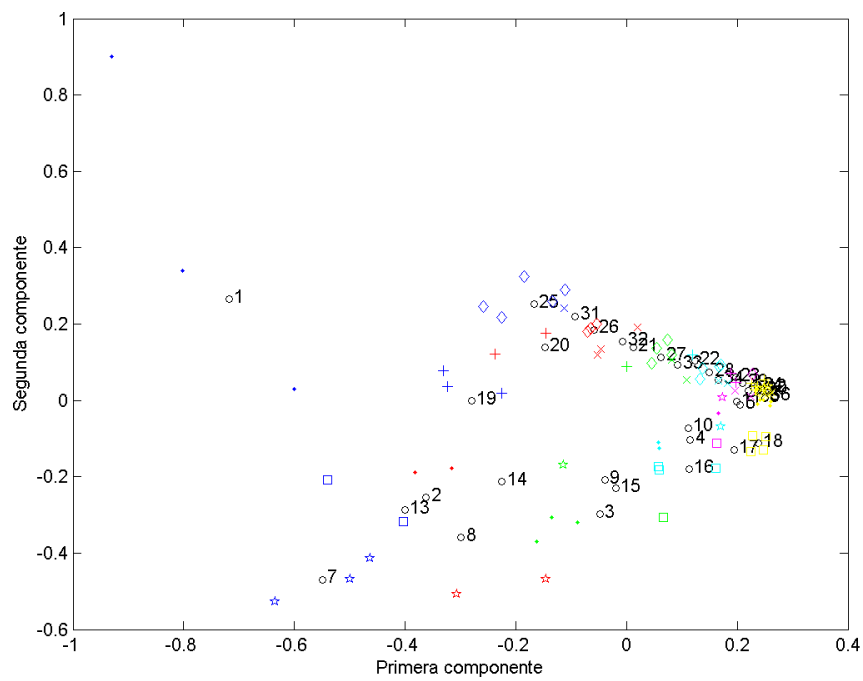


(b)

Figura 3.13: Grupos generados para el modelo Tensión. (a)L-T. (b)L-L.



(a)



(b)

Figura 3.14: Grupos generados para el modelo Tensión. (a)L-L-T. (b)L-L-L.

Capítulo 4

PRUEBAS Y RESULTADOS

4.1. Introducción

En el presente capítulo se muestran detalles de la simulación del sistema de distribución de prueba en condiciones de falla, las consideraciones que se tuvieron presentes para realizarla y la organización de los datos arrojados en ésta. También se expone la zonificación que se realizó al sistema de prueba, a partir de la aplicación de las técnicas para el reconocimiento de patrones de las que se habló en el capítulo 2. Para finalizar se muestran los resultados obtenidos después de realizar la validación a los modelos propuestos en el capítulo anterior, y el rendimiento que ofrece cada uno de ellos.

4.2. Simulación del sistema de distribución de prueba en condiciones de falla

Para realizar las simulaciones en ATP/EMTP, del sistema de distribución de prueba ante la ocurrencia de fallas de cortocircuito se tuvo presente dos consideraciones:

- Se simuló cada clase de falla (ver cuadro 4.1) por separado en cada una de las barras del sistema (desde la barra 1 hasta la barra 21).
- La simulación se realizó con 11 resistencias de falla (Ver cuadro 4.3).

La información arrojada en la simulación se almacenó en dos bases de datos. La primera corresponde a la base de datos de entrenamiento, la cual se compone de las simulaciones realizadas con los valores de resistencia R1, R2, R4, R6, R8, R10 y R11. La segunda corresponde a la base de datos de validación, la cual la conforman de las simulaciones realizadas con los valores restantes de

Cuadro 4.1: Clases de falla usadas en la simulación

| Tipo de falla | Fases falladas | Clase |
|-------------------------|---|-------|
| Falla monofásica (L-T) | Fase <i>a</i> | 01 |
| | Fase <i>b</i> | 02 |
| | Fase <i>c</i> | 03 |
| Falla bifásica (L-L) | Fase <i>a</i> , Fase <i>b</i> | 04 |
| | Fase <i>b</i> , Fase <i>c</i> | 05 |
| | Fase <i>c</i> , Fase <i>a</i> | 06 |
| Falla bifásica (L-L-T) | Fase <i>a</i> , Fase <i>b</i> | 07 |
| | Fase <i>b</i> , Fase <i>c</i> | 08 |
| | Fase <i>c</i> , Fase <i>a</i> | 09 |
| Falla trifásica (L-L-L) | Fase <i>a</i> , Fase <i>b</i> , Fase <i>c</i> | 10 |

Cuadro 4.3: Resistencias de falla usadas en la simulación

| Nombre | Valor [Ω] | Nombre | Valor [Ω] | Nombre | Valor [Ω] | Nombre | Valor [Ω] |
|-----------|--------------------|-----------|--------------------|-----------|--------------------|------------|--------------------|
| R1 | 0.05 | R4 | 15 | R7 | 30 | R10 | 45 |
| R2 | 5 | R5 | 20 | R8 | 35 | R11 | 50 |
| R3 | 10 | R6 | 25 | R9 | 40 | - | - |

resistencias (R3, R5, R7 y R9). Para el entrenamiento de la metodología propuesta no se emplearon todos los datos con el fin de no saturar con información que pudiera ser redundante, es decir solo se trabajó con las clases de falla 01 , 04, 07 y 10¹.

4.3. Zonificación del sistema de distribución de prueba

A partir del reconocimiento de patrones que se realizó en el capítulo 3, se crearon zonas en el sistema de potencia de prueba detallado en el anexo B. La selección de estas zonas se hizo a partir del comportamiento que presentaron las observaciones al momento de crear los grupos de cada modelo. Se proponen 8 zonas eléctricas; la razón de esto es que como se aprecia en la figura 4.1 hay 6 zonas claramente definidas, las 2 zonas restantes son las correspondientes a los ramales monofásicos. En la figura 4.2 y en el cuadro 4.5 se detalla la zonificación propuesta.

¹Estas clases de falla se llamarán datos de prueba en el resto del documento.

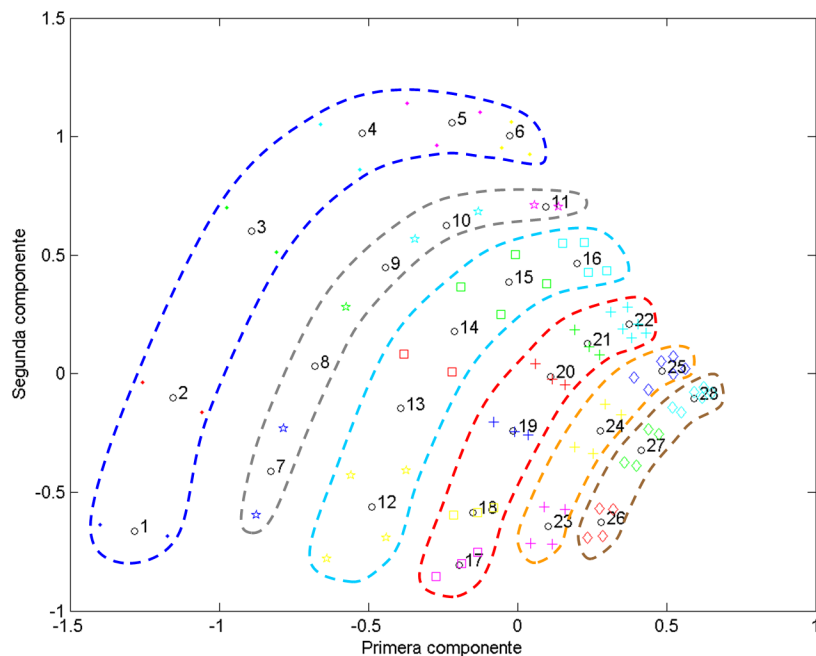


Figura 4.1: Zonificación en el agrupamiento de la falla L-L del modelo Tensión-corriente

Cuadro 4.5: Zonificación sistema de distribución de prueba.

| Zona | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|--------|-----|---|-----|--------|-----|-------|----------------|-------------|
| Barras | 1,2 | 3 | 4,5 | 6,7,12 | 8,9 | 10,11 | 13,14,15,16,17 | 18,19,20,21 |

4.4. Validación de los modelos propuestos

En ésta sección se muestra el resultado de las pruebas realizadas a los modelos propuestos. Primero se realizó la identificación del tipo de falla y la ubicación dentro de las zonas con los datos de entrenamiento en dos partes: la primera con los datos de prueba y la segunda con todos los datos de entrenamiento. Luego se procede a validar los modelos propuestos, empleando la herramienta software diseñada para éste fin, al igual que en el caso anterior se toman primero los datos de prueba y después todos los datos, esta vez de la base de validación. El porcentaje de acierto utilizado en la validación se calculó como se muestra en la ecuación (4.1).

$$P_{Acierto} = \frac{\#DCC}{\#DIC} * 100 \quad (4.1)$$

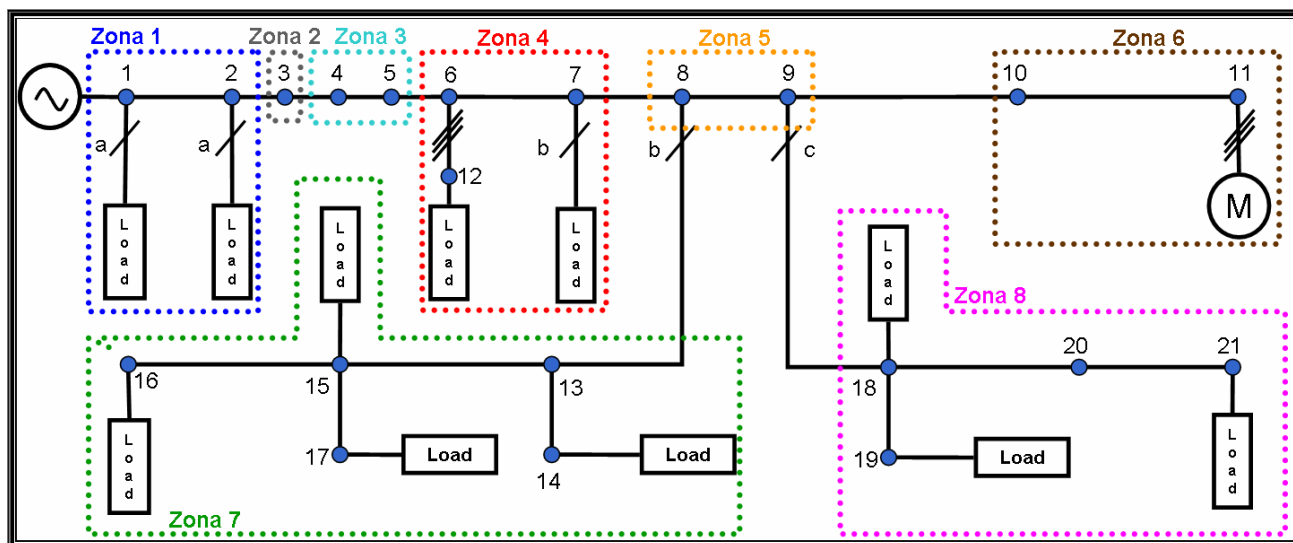


Figura 4.2: Zonificación sistema de distribución de prueba

Donde:

- #DCC: Corresponde al número de datos correctamente clasificados.
- #DIC: Corresponde al número total de datos a clasificar.

4.4.1. Identificación del tipo de falla

Se ha organizado una matriz para mostrar cuantos datos se clasificaron en la falla correcta y cuantos de ellos se clasificaron en un tipo de falla errada. Las filas indican el tipo de falla al cual pertenecen verdaderamente, mientras que las columnas son correspondientes a la clasificación. La diagonal corresponde a los valores #DCC (valor azul).

Cuadro 4.6: Identificación de tipo de falla con datos de entrenamiento.

| <i>Tipo de falla</i> | L-T | L-L | L-L-T | L-L-L |
|----------------------|-------|-------|-------|-------|
| L-T | 313 | 0 | 2 | 0 |
| L-L | 0 | 209 | 43 | 0 |
| L-L-T | 0 | 2 | 250 | 0 |
| L-L-L | 0 | 0 | 0 | 168 |
| #DIC | 315 | 252 | 252 | 168 |
| $P_{Acierto}$ en % | 99,36 | 82,93 | 99,20 | 100 |

Cuadro 4.7: Identificación de tipo de falla con datos de validación.

| <i>Tipo de falla</i> | L-T | L-L | L-L-T | L-L-L |
|----------------------|-----|-------|-------|-------|
| L-T | 180 | 0 | 0 | 0 |
| L-L | 0 | 120 | 24 | 0 |
| L-L-T | 0 | 10 | 134 | 0 |
| L-L-L | 0 | 0 | 0 | 96 |
| #DIC | 180 | 144 | 144 | 96 |
| $P_{Acierto}$ en % | 100 | 83,33 | 93,05 | 100 |

El cuadro (4.6) muestra el porcentaje de acierto al identificar el tipo de falla con los datos de entrenamiento, y el cuadro (4.7) muestra el porcentaje de acierto con los datos de validación. La figura (4.3) hace una comparación entre la identificación del tipo de falla para los dos tipos de datos. En ésta parte están todas las clases de fallas (01-11) para ambas bases de datos, se observa la efectividad en la localización, sin embargo para el caso de las fallas bifásicas es un poco menor, esto se debe a que los equivalente simétricos de fallas funcionan para el punto de falla y no visto desde otro punto, como en éste caso, desde el alimentador.

4.4.2. Localización de la zona fallada con datos de entrenamiento

Se realizó la localización de la falla con los datos de entrenamiento para cada uno de los modelos propuestos. La organización de la matriz de localización tiene la misma estructura que la de identificación de falla, solo que ahora se toman las 8 zonas propuestas y no los tipos de falla.

- **Modelo tensión-corriente:** De igual forma que en el proceso de identificación de falla se realiza la localización utilizando los datos de prueba como lo muestra el cuadro (4.8) y los datos completos de entrenamiento como lo muestra el cuadro (4.9).

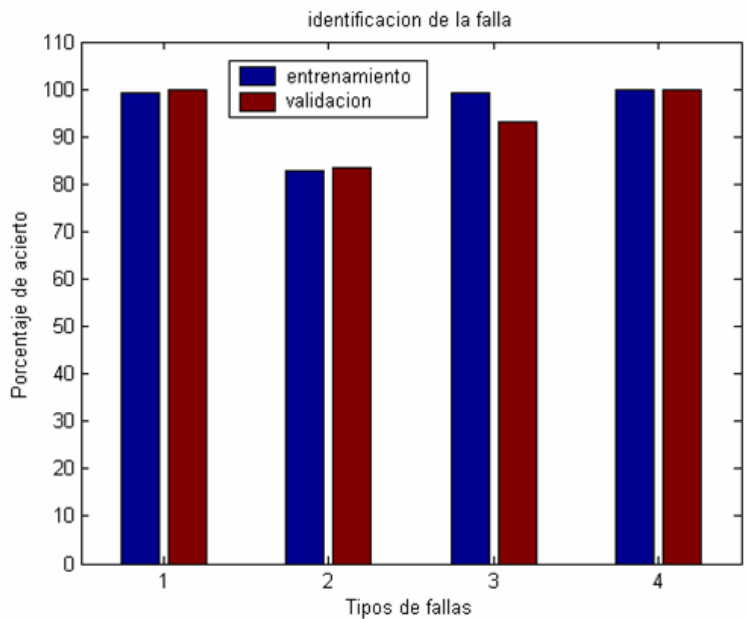


Figura 4.3: Identificación del tipo de falla.

Cuadro 4.8: Localización de la zona fallada con datos de prueba.

| Zona fallada | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|--------------------|-----|-------|-------|-------|-------|-----|-------|-------|
| 1 | 56 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 1 | 27 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 53 | 2 | 0 | 0 | 0 | 1 |
| 4 | 0 | 0 | 1 | 81 | 1 | 1 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 54 | 1 | 1 | 0 |
| 6 | 0 | 0 | 0 | 0 | 0 | 56 | 0 | 0 |
| 7 | 2 | 6 | 3 | 0 | 0 | 0 | 32 | 0 |
| 8 | 1 | 7 | 2 | 0 | 0 | 0 | 17 | 3 |
| #DIC | 56 | 28 | 56 | 84 | 56 | 56 | 35 | 28 |
| $P_{Acierto}$ en % | 100 | 96,43 | 94,64 | 96,43 | 96,43 | 100 | 91,43 | 10,71 |

Cuadro 4.9: Localización de la zona fallada con todos los datos de entrenamiento.

| Zona fallada | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|--------------------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1 | 152 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 5 | 71 | 1 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 9 | 135 | 7 | 0 | 0 | 0 | 3 |
| 4 | 0 | 0 | 9 | 219 | 1 | 2 | 0 | 0 |
| 5 | 0 | 0 | 0 | 10 | 135 | 2 | 7 | 0 |
| 6 | 0 | 0 | 0 | 2 | 9 | 143 | 0 | 0 |
| 7 | 0 | 0 | 0 | 0 | 1 | 2 | 32 | 0 |
| 8 | 0 | 0 | 0 | 0 | 0 | 8 | 17 | 3 |
| #DIC | 154 | 77 | 154 | 231 | 154 | 154 | 35 | 28 |
| $P_{Acierto}$ en % | 99,35 | 92,21 | 87,66 | 94,81 | 87,66 | 92,86 | 91,43 | 10,71 |

- **Modelo de corriente:** Aquí no se representa la matriz de localización con el fin de mostrar resultados específicos, la primera fila corresponde a los porcentajes de acierto con los datos de prueba y la segunda fila con todos los datos de entrenamiento (cuadro (4.10)).

Cuadro 4.10: Porcentaje de acierto en localización de la zona fallada con datos de entrenamiento con modelo corriente.

| Zona fallada | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|-------|-------|-------|-------|-------|-------|----|-------|
| $P_{Acierto}$ en % con datos de prueba. | 94,64 | 85,71 | 96,43 | 95,24 | 98,21 | 100 | 40 | 42,86 |
| $P_{Acierto}$ en % con datos de entren. | 87,59 | 61,04 | 75,33 | 76,62 | 74,03 | 81,17 | 40 | 42,86 |

- **Modelo de tensión:** Al igual que en el caso anterior solo se muestran resultados concretos (cuadro (4.11)).

Cuadro 4.11: Porcentaje de acierto en localización de la zona fallada con datos de entrenamiento con modelo tensión.

| Zona fallada | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|-------|-------|-------|-------|-------|-------|-------|-------|
| $P_{Acierto}$ en % con datos de prueba. | 80,36 | 85,71 | 89,29 | 78,57 | 78,57 | 62,5 | 57,14 | 82,14 |
| $P_{Acierto}$ en % con datos de entren. | 83,01 | 87,01 | 78,57 | 78,36 | 73,38 | 67,53 | 57,14 | 82,14 |

En la figura (4.4 (a)), se comparan para los mismos datos de prueba, los tres modelos planteados. El porcentaje de acierto es bueno para todas las zonas a excepción de la zona 7, el modelo Tensión-Corriente presenta un mejor comportamiento seguido por corriente y por último tensión. De forma similar se observa en la figura (4.4 (b)), donde no importa en que fase ocurrió la falla, se puede ver que el modelo de Tensión-Corriente es el que ofrece mejores resultados.

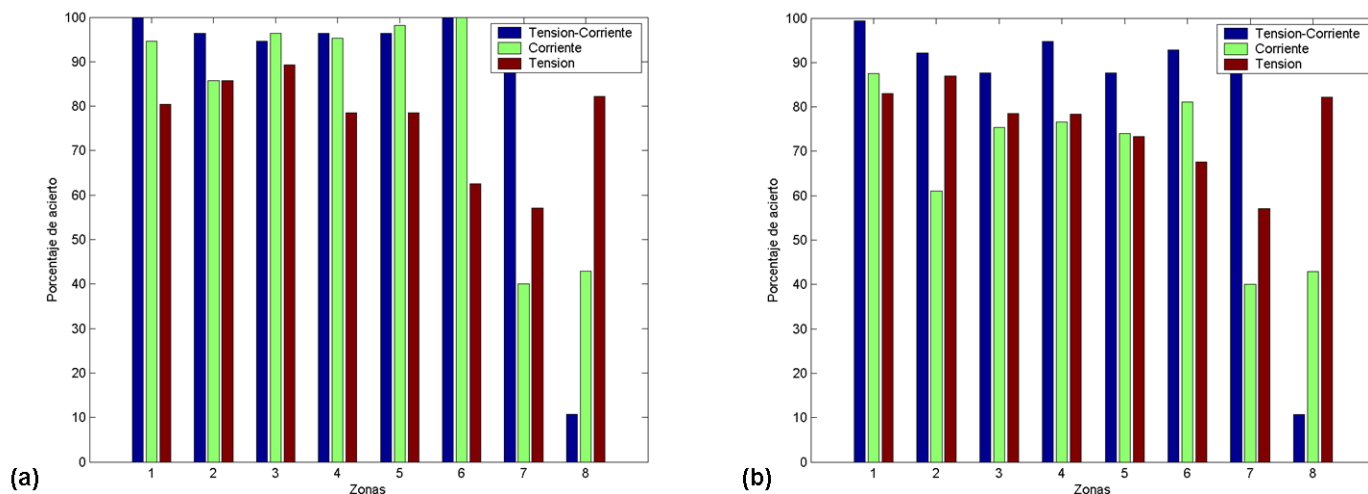


Figura 4.4: Porcentaje de acierto de validación con datos de entrenamiento. (a) Datos de prueba. (b) Datos completos de entrenamiento.

4.4.3. Localización de la zona fallada con datos de validación

Los resultados obtenidos después de realizar la localización de la falla utilizando los datos de validación, se organizan de igual manera que cuando se realizó con los datos de entrenamiento.

- **Modelo tensión-corriente:** El cuadro (4.12) muestra los resultados obtenidos utilizando los datos de prueba de validación y los resultados empleando los datos completos de validación se muestran en el cuadro (4.13).

Cuadro 4.12: Localización de la zona fallada con datos de prueba de validación.

| Zona fallada | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|--------------------|-------|-------|----|-------|-------|-------|----|------|
| 1 | 29 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 3 | 13 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 5 | 24 | 3 | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 3 | 45 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 26 | 2 | 4 | 0 |
| 6 | 0 | 0 | 0 | 0 | 1 | 30 | 1 | 0 |
| 7 | 0 | 0 | 0 | 0 | 1 | 0 | 19 | 0 |
| 8 | 0 | 0 | 0 | 0 | 0 | 5 | 9 | 2 |
| #DIC | 32 | 16 | 32 | 48 | 32 | 32 | 20 | 16 |
| $P_{Acierto}$ en % | 90,63 | 81,25 | 75 | 93,75 | 81,25 | 93,75 | 95 | 12,5 |

Cuadro 4.13: Localización de la zona fallada con todos los datos de entrenamiento.

| Zona fallada | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|--------------------|-------|-------|-------|-------|-------|------|----|------|
| 1 | 79 | 9 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 10 | 34 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 1 | 15 | 65 | 7 | 0 | 0 | 0 | 0 |
| 4 | 0 | 1 | 13 | 118 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 1 | 7 | 64 | 4 | 12 | 0 |
| 6 | 0 | 0 | 0 | 2 | 8 | 77 | 1 | 0 |
| 7 | 0 | 0 | 0 | 0 | 1 | 0 | 19 | 0 |
| 8 | 0 | 0 | 0 | 0 | 0 | 5 | 9 | 2 |
| #DIC | 88 | 44 | 88 | 132 | 88 | 88 | 20 | 16 |
| $P_{Acierto}$ en % | 89,77 | 77,27 | 73,86 | 89,39 | 72,73 | 87,5 | 95 | 12,5 |

- **Modelo de corriente:** La primera fila corresponde a los porcentajes de acierto con los datos de prueba de validación y la segunda fila con todos los datos de validación (cuadro (4.14)).

Cuadro 4.14: Porcentaje de acierto en localización de la zona fallada con datos de validación con modelo corriente.

| Zona fallada | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|--|-------|-------|-------|-------|-------|-------|----|-------|
| $P_{Acierto}$ en % datos de prueb. valida. | 75 | 75 | 90,63 | 93,75 | 75 | 78,13 | 60 | 56,25 |
| $P_{Acierto}$ en % datos de todos valida. | 69,32 | 52,27 | 63,67 | 69,7 | 59,09 | 65,91 | 60 | 56,25 |

- *Modelo de tensión:* Ver cuadro (4.11).

Cuadro 4.15: Porcentaje de acierto en localización de la zona fallada con datos de validación con modelo tensión.

| Zona fallada | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|--|-------|-------|-------|-------|-------|-------|----|------|
| $P_{Acierto}$ en % con datos de prueba valida. | 87,5 | 93,75 | 84,38 | 75 | 56,25 | 81,25 | 50 | 87,5 |
| $P_{Acierto}$ en % con todos datos de valida. | 85,23 | 84,1 | 75 | 76,52 | 61,36 | 76,14 | 50 | 87,5 |

La figura (4.5) compara la efectividad de los tres modelos, pero ésta vez con los datos de validación y se aprecian mejoras para algunas zonas pero se baja considerablemente el porcentaje de acierto en la zona 8 con el modelo tensión-corriente.

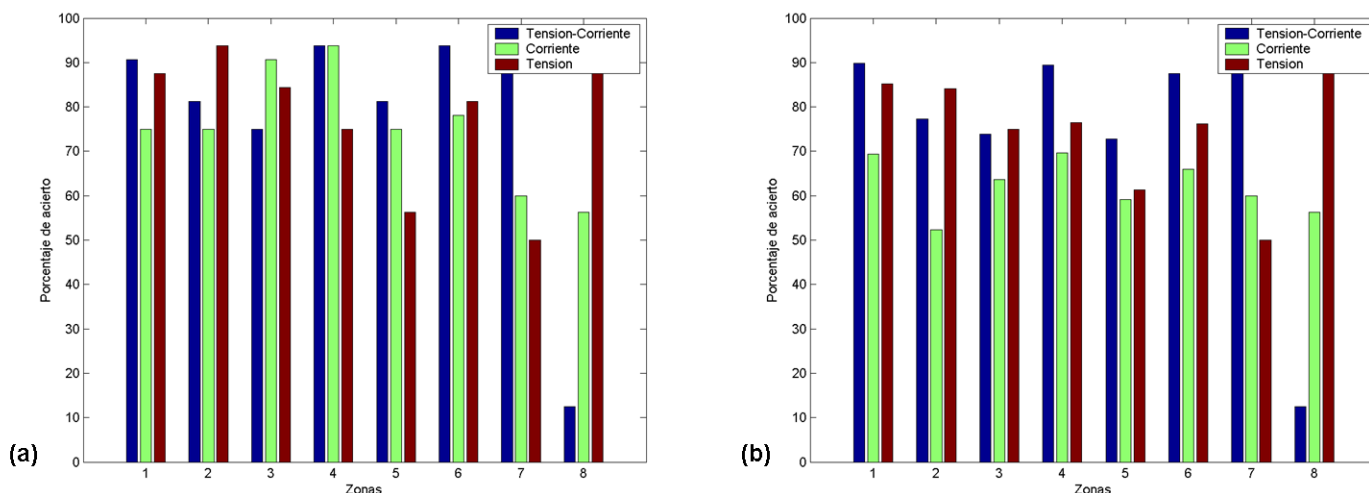


Figura 4.5: Porcentaje de acierto de localización de fallas con datos de validación. (a) Datos de prueba de validación. (b) Datos completos de validación.

La figura (4.6) muestra el comportamiento que tiene cada modelo durante el proceso de validación. Esta figura es un resumen de todas las figuras anteriores. Vincula todos los datos, incluidos los datos de prueba y los datos completos de cada base de datos. Debido a que solo se empleó el 40% de las clases de fallas para el entrenamiento, se esperaría un porcentaje de acierto cercano al 40% con todos los datos de validación, pero como se puede apreciar la mayoría de veces se supera el 40% alcanzando a llegar al 99,35%, lo cual es satisfactorio comparado con el tamaño de los datos empleados. Exceptuando la zona 8 con el modelo de tensión-corriente, debido a la

confusión que se presenta con la zona 7. Esto sugiere la formación de una sola zona que comprenda los ramales monofásicos pero sería una zona muy grande y la idea es diferenciarlo debido a que están en diferentes fases.

4.4.4. Herramienta empleada para realizar la validación de los modelos y las zonas propuestas (*validación*)

La validación de los modelos y de las zonas propuestas se realizó mediante el empleo de una sencilla herramienta diseñada en Matlab 6.5.

Para ejecutar la herramienta es necesario seguir los siguientes pasos:

1. Copiar la carpeta *validación* en la siguiente ruta C:\MATLAB6p5.
2. Buscar en current Directory la carpeta *validación*.
3. Cumplido lo anterior, ya se tiene acceso a la herramienta digitando sobre la ventana de comandos de MatLab *validacion*².

La herramienta *validación*, emplea los autovectores y los centros de grupo calculados mediante el reconocimiento de patrones en cada uno de los tres modelos propuestos. La herramienta carga la base de datos que contiene las simulaciones destinadas para la validación³. Con ésta información reacomoda los datos en las zonas preestablecidas, localizando el probable sector de ocurrencia de la falla de cortocircuito.

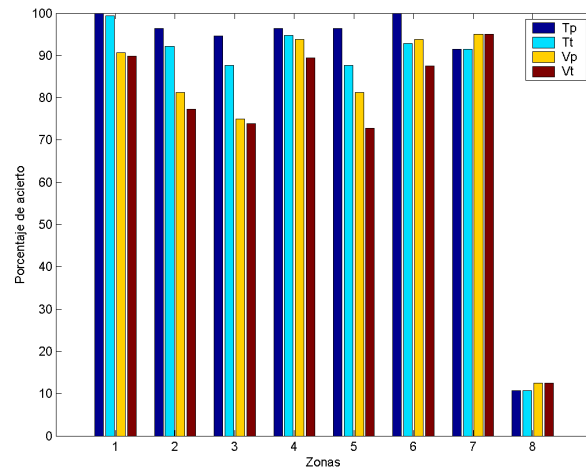
En la figura (4.7) se observan siete campos de la herramienta:

1. **Tipo de falla:** El usuario selecciona uno de los cuatro tipos de falla (LT-LL-LLT-LLL).
2. **Resistencia de falla:** Se selecciona uno de los cuatros valores de resistencia de falla empleados en la simulación destinada para realizar la validación (10 Ohm, 20 Ohm, 30 Ohm o 40 Ohm).
3. **Modelo:** Se escoge uno, dos o tres modelos con los que se quiera realizar el proceso de validación (tensión-corriente, corriente o tensión).
4. **Barra:** Se elige la barra del sistema de potencia de prueba en donde se quiere realizar la falla.

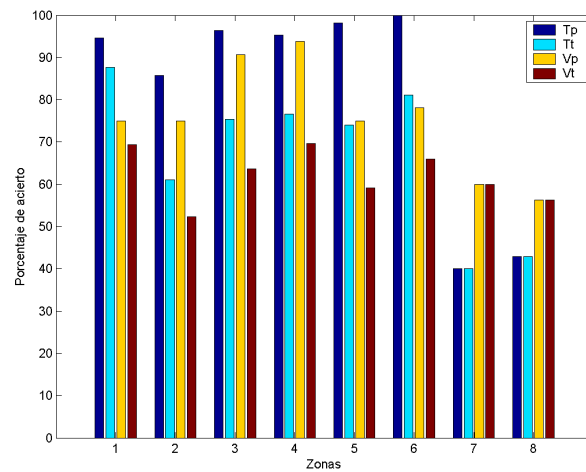
²No lleva tilde por tratarse de un comando

³Resistencias R3, R5, R7 y R9

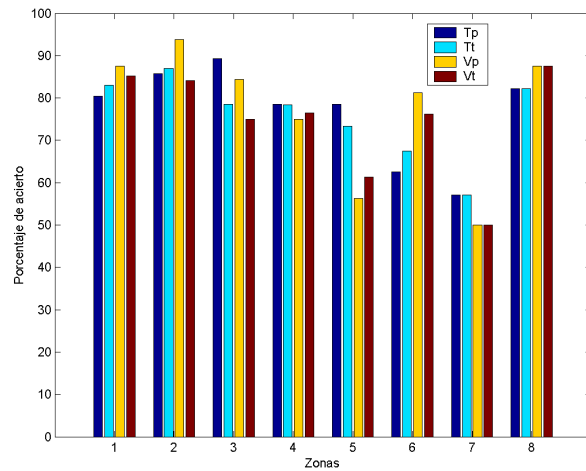
5. **Localizar:** Este botón ejecuta el proceso de validación.
6. **Zona:** Este campo muestra la zona real en donde ocurre el tipo de falla.
7. **Mensaje:** La localización realizada por cada uno de los modelos se detalla en éste campo.



(a)



(b)



(c)

Figura 4.6: Comportamiento General. (a)Modelo Tensión-Corriente. (b)Modelo corriente. (c) Modelo tensión.



Figura 4.7: Herramienta de Validación

Capítulo 5

CONCLUSIONES APORTES Y TRABAJOS FUTUROS

5.1. Conclusiones y observaciones

- El ACP redujo la dimensionalidad de los datos maximizando la varianza entre los nuevos descriptores.
- El ACP ayudó a la investigación pues proporcionó una interpretación de los datos, facilitando así el camino que llevó a la conformación de zonas donde se pudo identificar el comportamiento del sistema ante la ocurrencia de fallas, a partir de un análisis de conglomerados. Es decir el ACP es un paso fundamental en el desarrollo de la metodología propuesta para la localización de fallas.
- Un factor influyente para la localización de fallas, es la detección del tipo de falla, pues la propuesta trabajó cada uno de los tipos de falla de forma independiente. La falla bifásica es la que para los modelos propuestos presenta menor efectividad de identificación.
- La efectividad del método de reconocimiento de patrones que se utilizó depende entre otras cosas del conocimiento previo y profundo de los datos que van a ser manipulados.
- Tener un número de zonas considerable exige más exactitud en los centros de los grupos porque para un computador el valor 0,29601 es menor que 0,29602 y eso hace que un dato quede en un grupo que no es el suyo. La primera zona presenta más facilidad para su localización mientras que la octava zona no es fácilmente identificable, para algunos casos se confunde con la séptima, la dificultad está directamente relacionado con que la zona 7 y la zona 8 corresponden a los ramales monofásicos.

- El algoritmo utilizado para realizar los agrupamientos poseen variables iniciales de operación, las cuales son responsabilidad del analista. Sin una buena interpretación, la efectividad de éste bajaría notoriamente.
- Para la formación de los grupos fue necesario escalar la segunda componente principal, el factor de escala seleccionado lo escoge el analista de los datos, pues solo el sabe que es lo que busca agrupar. Se sugiere una forma neutral de escalamiento.
- Los textos consultados muestran que los datos que se analizan se conglomeran de forma en que se pueden identificar visualmente grupos bien definidos, sin embargo, se observó que no siempre se obtiene esta clase de resultados.

5.2. Aportes

- Se diseñó una herramienta software que no se limita solamente al análisis y solución del problema de la localización de fallas. Es decir, la herramienta diseñada puede ser utilizado para cualquier tipo de problema multivariado.
- La herramienta *PRtool* ofrece un confort visual al momento de realizar el análisis de datos multivariados, lo cual es muy importante para la toma de decisiones y las conclusiones se vayan a obtener sobre el análisis.
- La localización de fallas es posible con la utilización de técnicas estadísticas multivariadas de exploración de datos. La efectividad de ésta radica en la cantidad de datos que se utilicen para entrenar la herramienta.

5.3. Trabajos futuros

- Buscar descriptores que detallen más a fondo el comportamiento de un sistema de distribución ante la ocurrencia de fallas. Esto con el fin de aprovechar al máximo el análisis de componentes principales.
- La herramienta *PRtool* solo maneja variables cuantitativas, queda abierto el campo para incluir en éste análisis la caracterización por medio de variables cualitativas.
- La eliminación de datos atípicos de las bases de datos podría mejorar los resultados.

Anexo A

Cálculo de componentes principales

Supongamos que se dispone de los valores de p variables de n individuos de una población organizados en una matriz \mathbf{X} de dimension $n \times p$. La primer componente principal se define como la combinación lineal de las variables originales que tiene varianza máxima. Los valores de este primer componente en n individuos se representan por un vector \mathbf{z}_1 dado por:

$$\mathbf{z}_1 = \mathbf{X}\mathbf{a}_1 \quad (\text{A.1})$$

donde \mathbf{a}_1 es un vector columna de la forma $\mathbf{a}_1 = (a_{11}, \dots, a_{1p})^T$, con norma unitaria. Para ésta demostración se supone que previamente se ha restado a cada variable su media (variables tipificadas), de manera que las variables de la matriz \mathbf{X} tienen media cero y su matriz de varianzas y covarianzas está dada por $\mathbf{S} = \frac{1}{n}\mathbf{X}^T\mathbf{X}$, luego \mathbf{z}_1 tendrá media nula.

Su varianza será:

$$\frac{1}{n}\mathbf{z}_1^T\mathbf{z}_1 = \frac{1}{n}\mathbf{a}_1^T\mathbf{X}^T\mathbf{X}\mathbf{a}_1 = \mathbf{a}_1^T\mathbf{S}\mathbf{a}_1 \quad (\text{A.2})$$

Se puede maximizar la varianza sin límite aumentando el modulo del vector \mathbf{a}_1 . Para que la maximización de (A.2) tenga solución, se debe imponer una restricción al módulo del vector \mathbf{a}_1 , y sin pérdida de generalidad, se impondrá que $\mathbf{a}_1 \cdot \mathbf{a}_1 = 1$. Esta restricción se introduce mediante el multiplicador de Lagrange:

$$M = \mathbf{a}_1^T\mathbf{S}\mathbf{a}_1 - \lambda(\mathbf{a}_1 \cdot \mathbf{a}_1 - 1) \quad (\text{A.3})$$

y maximizando ésta expresión de la forma habitual derivando respecto a los componentes de \mathbf{a}_1 e igualando a cero. Entonces

$$\frac{\partial M}{\partial \mathbf{a}_1} = 2\mathbf{S}\mathbf{a}_1 - 2\lambda\mathbf{a}_1 = 0 \quad (\text{A.4})$$

cuya solución es:

$$\mathbf{S}\mathbf{a}_1 = \lambda\mathbf{a}_1 \quad (\text{A.5})$$

que implica que \mathbf{a}_1 es un vector propio de la matriz S , y λ su correspondiente valor propio. Para determinar qué valor propio de S es solución de (A.5), se premultiplica por \mathbf{a}_1^T ésta ecuación,

$$\mathbf{a}_1^T \mathbf{S} \mathbf{a}_1 = \lambda \mathbf{a}_1^T \mathbf{a}_1 = \lambda \quad (\text{A.6})$$

y se concluye por A.2, que λ es la varianza de \mathbf{z}_1 . Como ésta es la cantidad que se quiere maximizar, λ será el mayor valor propio de la matriz S . Su vector asociado, \mathbf{a}_1 , define los coeficientes de cada variable en el primer componente principal.

Después de tener la primera componente se procede a obtener el mejor plano de proyección de las variables X . Para esto se establece como función objetivo que la suma de las varianzas $\mathbf{z}_1 = \mathbf{X} \mathbf{a}_1$ y $\mathbf{z}_2 = \mathbf{X} \mathbf{a}_2$ sea máxima, donde \mathbf{a}_1 y \mathbf{a}_2 son vectores que definen el plano. La función objetivo será:

$$\phi = \mathbf{a}_1^T \mathbf{S} \mathbf{a}_1 + \mathbf{a}_2^T \mathbf{S} \mathbf{a}_2 - \lambda_1 (\mathbf{a}_1^T \cdot \mathbf{a}_1 - 1) - \lambda_2 (\mathbf{a}_2^T \cdot \mathbf{a}_2 - 1) \quad (\text{A.7})$$

que incorpora las restricciones de que las direcciones deben tener módulo unitario ($\mathbf{a}_i^T \mathbf{a}_i = 1$), $i=1,2$. Derivando e igualando a cero:

$$\frac{\partial \phi}{\partial \mathbf{a}_1} = 2\mathbf{S} \mathbf{a}_1 - 2\lambda_1 \mathbf{a}_1 = 0 \quad (\text{A.8})$$

$$\frac{\partial \phi}{\partial \mathbf{a}_2} = 2\mathbf{S} \mathbf{a}_2 - 2\lambda_2 \mathbf{a}_2 = 0 \quad (\text{A.9})$$

La solución de éste sistema es:

$$\mathbf{S} \mathbf{a}_1 = \lambda_1 \mathbf{a}_1 \quad (\text{A.10})$$

$$\mathbf{S} \mathbf{a}_2 = \lambda_2 \mathbf{a}_2 \quad (\text{A.11})$$

que indica que \mathbf{a}_1 y \mathbf{a}_2 deben ser vectores propios de S . Tomando los vectores de norma uno y sustituyendo en A.7, se tiene que, en el máximo, la función objetivo como se muestra en A.12.

$$\phi = \lambda_1 + \lambda_2 \quad (\text{A.12})$$

Es claro que λ_1 y λ_2 son los dos autovalores mayores de la matriz S y \mathbf{a}_1 y \mathbf{a}_2 sus correspondientes autovectores. Se observa que la covarianza entre \mathbf{z}_1 y \mathbf{z}_2 estarán incorreladas [P.D, 2002, O.P, 2000, H.W, 2003].

Anexo B

Características del sistema de distribución de prueba

El sistema de distribución utilizado para desarrollar la metodología planteada en el capítulo 3 se muestra en la figura (B.1). En éste modelo fueron realizadas las simulaciones de las fallas de cortocircuito en ATP/EMTP. El modelo fue obtenido de la tesis doctoral para localización de fallas en sistemas de distribución de Ratan Das [D.R, 1998].

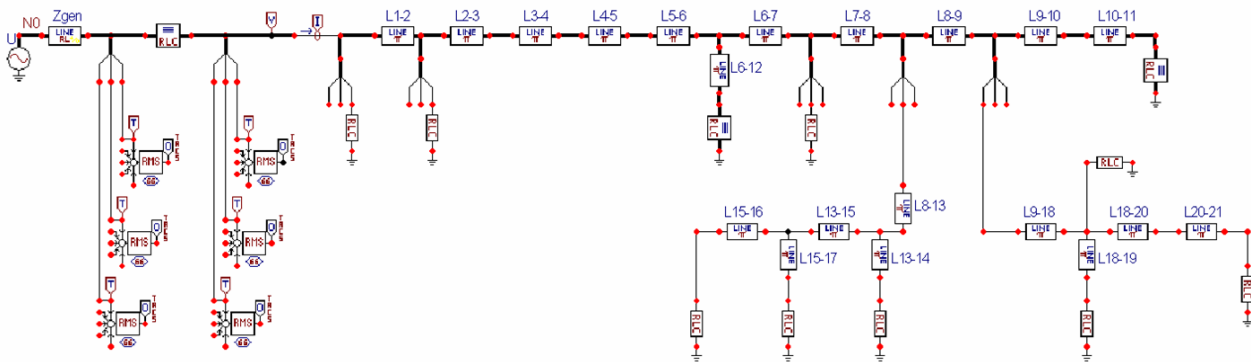


Figura B.1: Sistema de Potencia de Prueba

Los parámetros de la fuente ubicada en la cabecera del circuito se muestran en el cuadro (B.1). La información de cada una de las cargas ubicadas en cada una de las barras se muestran en el cuadro (B.2). La descripción de los parámetros de las líneas que interconectan las barras se muestran en el cuadro (B.3) y por último el factor de potencia de las cargas utilizadas en el cuadro (B.4).

Antes de caracterizar el comportamiento del sistema ante la ocurrencia de fallas es importante

Cuadro B.1: Parámetros del generador.

| Tensión Base (kV) | Potencia Base (MVA) | Impedancia de Secuencia Positiva y Negativa (p.u.) | Impedancia de Secuencia Cero (p.u.) |
|-------------------|---------------------|--|-------------------------------------|
| 25 | 10 | $0,68283 + j2,98139$ | $0,09496 + j1,39289$ |

Cuadro B.2: Datos de cargas del sistema.

| Nodo | Fase | Carga (kVA) | Composición (%) | | |
|------|----------|-------------|-----------------|-------------|-------|
| | | | Calentador | Iluminación | Motor |
| 1 | A | 15 | 99,8 | 0,1 | 0,1 |
| 2 | A | 15 | 99,8 | 0,1 | 0,1 |
| 7 | B | 15 | 99,8 | 0,1 | 0,1 |
| 11 | A, B, C. | 1000 | 0,1 | 0,1 | 99,8 |
| 12 | A, B, C. | 67,5 | 99,8 | 0,1 | 0,1 |
| 14 | B | 15 | 99,8 | 0,1 | 0,1 |
| 15 | B | 15 | 99,8 | 0,1 | 0,1 |
| 16 | B | 7,5 | 99,8 | 0,1 | 0,1 |
| 17 | B | 15 | 99,8 | 0,1 | 0,1 |
| 18 | C | 25 | 99,8 | 0,1 | 0,1 |
| 19 | C | 15 | 99,8 | 0,1 | 0,1 |
| 21 | C | 15 | 99,8 | 0,1 | 0,1 |

conocer como se comporta en condiciones normales de operación. En estado estable las magnitudes de las tensiones son aproximadamente $V_{\mathbf{a}, \mathbf{b}, \mathbf{c}} = \frac{25kV}{\sqrt{3}} = 14433,8[V]$ sin tener presente la caída interna.

Del diagrama fasorial (B.2) las magnitudes son: 0,9789, 0,9795, 0,9789 [p.u] y los ángulos de fase son: $-4,1941^\circ$, $-124,27^\circ$, $115,73^\circ$ y con referencia cero son 0° , $-120,07^\circ$, $119,92^\circ$, respectivamente. Pero dada la naturaleza del sistema, es decir, la diversidad de cargas conectadas existe un leve desbalance del sistema apreciándose en la magnitud de la tensión de secuencia cero que es cercana a los 0,00075 [p.u].

De forma análoga sucede con la corriente, cuyo valor en estado estable es: $I_a=0,1026$ [p.u], $I_b=0,1068$ [p.u], $I_c=0,1077$ [p.u]¹ y los ángulos de fase son $-32,501^\circ$, $-150,56^\circ$, 90° , respectivamente. Llevando a referencia cero son 0° , $-118,06^\circ$, $122,5^\circ$. También se aprecia que la corriente está en atraso con un factor de potencia $fp = \cos(\theta_v - \theta_i) = \cos(28,307^\circ) = 0,88$.

¹Con una corriente base de 230,94 [A]

Cuadro B.3: Parámetros de las líneas.

| Sección entre los nodos | Longitud de la sección (km) | Impedancia Serie (Ohms / km) | | Admitancia Paralelo (Ohms / km) | |
|-------------------------|-----------------------------|-------------------------------|--------------------|---------------------------------|---------------------|
| | | Secuencia Positiva y Negativa | Secuencia Cero | Secuencia Positiva y Negativa | Secuencia Cero |
| 1 - 2 | 2,414 | $0,3480 + j0,5166$ | $0,5254 + j1,704$ | $j3,74 \text{ E-6}$ | $j2,49 \text{ E-6}$ |
| 2 - 6 | 16,092 | $0,3480 + j0,5166$ | $0,5254 + j1,704$ | $j3,74 \text{ E-6}$ | $j2,49 \text{ E-6}$ |
| 6 - 7 | 4,023 | $0,3480 + j0,5166$ | $0,5254 + j1,704$ | $j3,74 \text{ E-6}$ | $j2,49 \text{ E-6}$ |
| 7 - 8 | 5,150 | $0,5519 + j0,5390$ | $0,7290 + j1,727$ | $j3,59 \text{ E-6}$ | $j2,39 \text{ E-6}$ |
| 8 - 9 | 2,414 | $0,5519 + j0,5390$ | $0,7290 + j1,727$ | $j3,59 \text{ E-6}$ | $j2,39 \text{ E-6}$ |
| 9 - 10 | 4,506 | $0,5519 + j0,5390$ | $0,7290 + j1,727$ | $j3,59 \text{ E-6}$ | $j2,39 \text{ E-6}$ |
| 10 - 11 | 2,414 | $0,3480 + j0,5166$ | $0,5254 + j1,704$ | $j3,74 \text{ E-6}$ | $j2,49 \text{ E-6}$ |
| 6 - 12 | 2,414 | $0,3480 + j0,5166$ | $0,5254 + j1,704$ | $j3,74 \text{ E-6}$ | $j2,49 \text{ E-6}$ |
| 8 - 13 | 2,414 | $7,3977 + j0,8998$ | $7,3977 + j0,8998$ | $j2,51 \text{ E-6}$ | $j2,51 \text{ E-6}$ |
| 13 - 14 | 2,414 | $7,3977 + j0,8998$ | $7,3977 + j0,8998$ | $j2,51 \text{ E-6}$ | $j2,51 \text{ E-6}$ |
| 13 - 15 | 2,414 | $7,3977 + j0,8998$ | $7,3977 + j0,8998$ | $j2,51 \text{ E-6}$ | $j2,51 \text{ E-6}$ |
| 15 - 16 | 2,414 | $7,3977 + j0,8998$ | $7,3977 + j0,8998$ | $j2,51 \text{ E-6}$ | $j2,51 \text{ E-6}$ |
| 15 - 17 | 2,414 | $7,3977 + j0,8998$ | $7,3977 + j0,8998$ | $j2,51 \text{ E-6}$ | $j2,51 \text{ E-6}$ |
| 9 - 18 | 2,414 | $7,3977 + j0,8998$ | $7,3977 + j0,8998$ | $j2,51 \text{ E-6}$ | $j2,51 \text{ E-6}$ |
| 18 - 19 | 2,414 | $7,3977 + j0,8998$ | $7,3977 + j0,8998$ | $j2,51 \text{ E-6}$ | $j2,51 \text{ E-6}$ |
| 18 - 20 | 3,219 | $7,3977 + j0,8998$ | $7,3977 + j0,8998$ | $j2,51 \text{ E-6}$ | $j2,51 \text{ E-6}$ |
| 20 - 21 | 3,219 | $7,3977 + j0,8998$ | $7,3977 + j0,8998$ | $j2,51 \text{ E-6}$ | $j2,51 \text{ E-6}$ |

Cuadro B.4: Factor de potencia.

| Tipo de carga | Factor de potencia |
|---------------|--------------------|
| Calentador | 1 |
| Iluminación | 0,85 atraso |
| Motor | 0,8 atraso |

La pregunta que surge con el resultado anterior es: ¿por qué siendo el sistema trifásico trifilar aparece una componente de secuencia cero? La respuesta a este interrogante es la siguiente: El hecho de tener cargas monofásicas permite que la corriente retorne por la tierra, es importante recordar que la tierra también se comporta como un conductor, a esto se le suma que la fuente también está conectada a tierra, y por último hay que tener presente los efectos capacitivos, aunque pequeños no se pueden despreciar.

Deben existir dos factores para que aparezca corriente de secuencia cero:

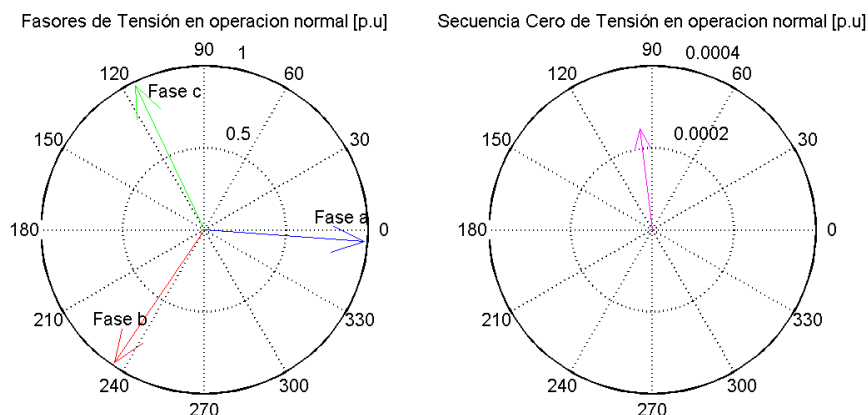


Figura B.2: Fasores de tensión y secuencia cero en operación normal.

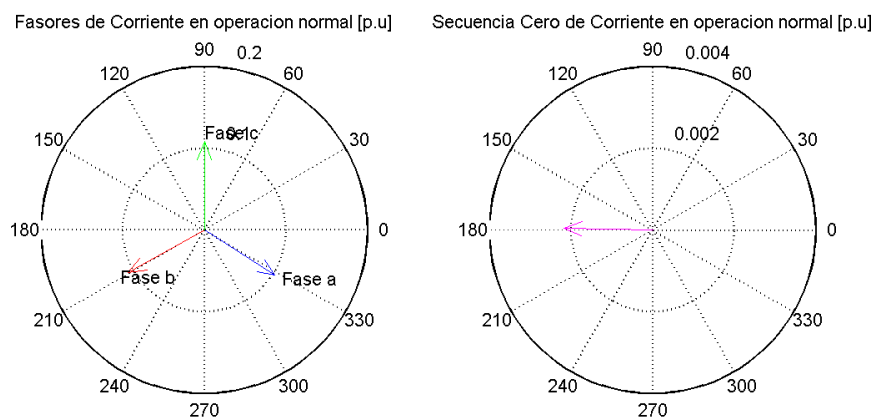


Figura B.3: Fasores de corriente y secuencia cero en operación normal.

1. Medio físico por el cual pueda circular, diferente a una fase, en este tópico aparece el neutro como la tierra.
2. El balance de cargas que presenta el sistema, debido a la diversidad de carga.

Para el caso que se está trabajando se cumplen estos dos factores, pues se cuentan con cargas monofásicas conectadas a tierra, a su vez el generador también se halla aterrizado. Otro elemento que no se puede ver pero que siempre está presente y que no se puede omitir el efecto capacitivo el cual provee una ruta de alta impedancia para las corrientes de secuencia cero al igual que las corrientes de fuga presentes en los aisladores. Se concluye que por más perfecto que sea el sistema de potencia (balanceado) existen elementos reales que no se pueden omitir para éste tipo de análisis. Los modelos ideales se aproximan a la realidad en una buena proporción con pequeños porcentajes

de error, y siempre van a existir pequeños detalles que no se tuvieron presentes al momento de diseñarlos, en algún momento cobran importancia para uno u otro tema de investigación.

Anexo C

Fallas de cortocircuito

C.1. Introducción

Una falla de cortocircuito es cualquier evento que interfiere con el flujo normal de corriente. La mayoría de las fallas en las líneas de transmisión de 115 kV, o mayores, son originadas por las descargas atmosféricas, que resultan en el flameo de los aisladores. La alta tensión entre un conductor y la estructura que lo sostiene, origina una ionización que facilita un camino a tierra para la carga inducida en la descarga [G.J, 1996].

La ocurrencia de fallas se ve directamente relacionada con el tipo de ésta. Aproximadamente el 85 % de las fallas que ocurren son de tipo monofásico (Línea a tierra), y las que menos ocurren son las que involucran las tres fases (trifásicas simétricas), aproximadamente el 2 % . La estadística de la ocurrencia de fallas según el tipo y el sitio se muestra en los cuadros (C.1) y (C.3) respectivamente [C.G, 1990]. A continuación se hace un recorrido por los tipos de falla de cortocircuito o fallas paralelo mostrando las características¹ que cada una presenta [A. A, 1999].

Cuadro C.1: Estadística de ocurrencia de fallas de acuerdo al tipo.

| Tipo de Falla | % Total |
|-----------------|-----------|
| Monofásica | 85 % |
| Bifásica | 8 |
| Bifásica Tierra | 5 |
| Trifásica | 2 o menos |

¹Estas características fueron obtenidas mediante la simulación de cada tipo de falla medidas desde el alimentador del sistema detallado en el apéndice B

Cuadro C.3: Estadística de ocurrencia de fallas de acuerdo al sitio.

| Sitio de la Falla | % Total |
|--|---------|
| Línea de Transmisión | 50 |
| Equipos de interrupción | 15 |
| Transformadores | 12 |
| Cables | 10 |
| Otros | 8 |
| Equipos de control | 3 |
| Transformadores de corriente y potencial | 2 |

C.2. Falla Simétrica

C.2.1. Falla Trifásica (L-L-L)

Este tipo de falla se presenta cuando las tres fases de una línea se interconectan entre sí como lo muestra la figura (C.1). Es la falla que presenta menor ocurrencia, pero en general es la más crítica. Su equivalente simétrico es como detalla la figura (C.2). Las señales de tensión y corriente bajo la ocurrencia de una falla trifásica toman características muy notables, como se puede apreciar en las gráficas (C.3), (C.4), (C.5) y (C.6).

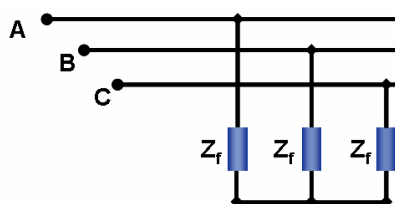


Figura C.1: Falla trifásica

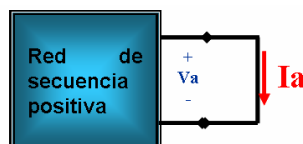


Figura C.2: Equivalente simétrico falla trifásica

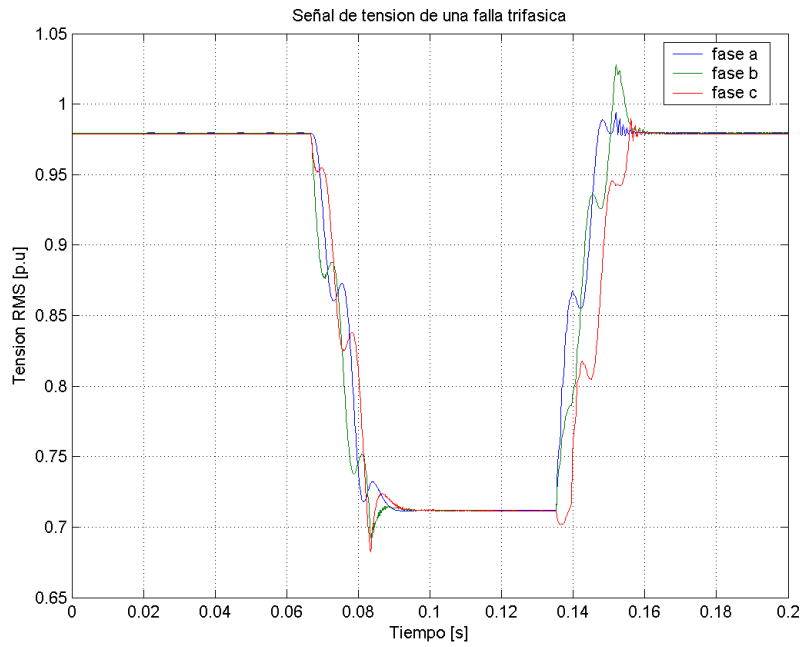


Figura C.3: Señal de tensión bajo una falla trifásica

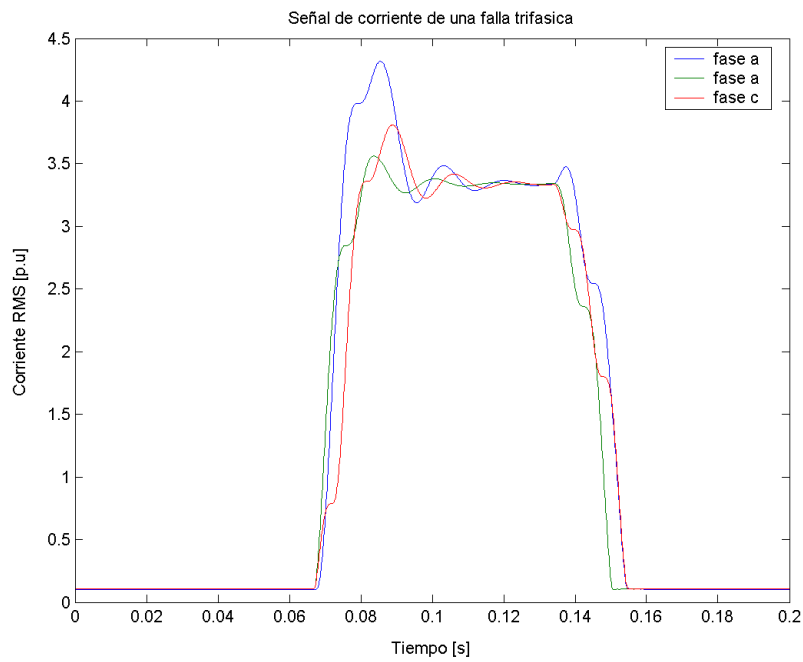


Figura C.4: Señal de corriente bajo una falla trifásica

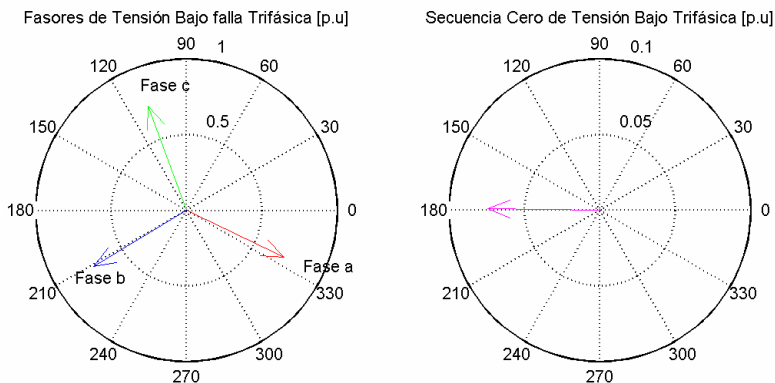


Figura C.5: Fasores de tensión y secuencia cero bajo una falla trifásica

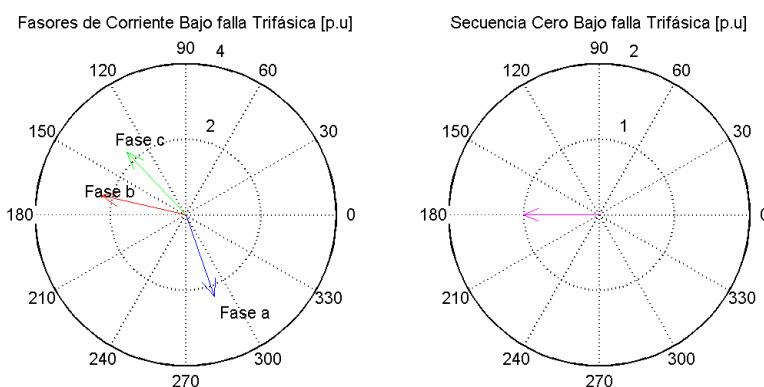


Figura C.6: Fasores de corriente y secuencia cero bajo una falla trifásica

C.3. Fallas asimétricas

La mayoría de las fallas que ocurren en los sistemas de potencia, son fallas asimétricas que consisten en cortocircuitos asimétricos, fallas asimétricas a través de impedancias o conductores abiertos.

C.3.1. Fallas monofásicas de línea a tierra L-T

La falla monofásica de línea a tierra es el tipo más común de falla, se origina por las descargas atmosféricas o por los conductores al hacer contacto con las estructuras aterrizadas. Su equivalente simétrico se observa en la figura C.8. Se considera que la fase fallada es la fase *a*.

Como en el caso de la falla trifásica, las señales de tensión y corriente bajo la ocurrencia de una falla monofásica también toman características muy notables. La fase afectada en nivel de tensión disminuye notoriamente y las otras poco se ven afectadas, esto se aprecia en la gráficas (C.9) y

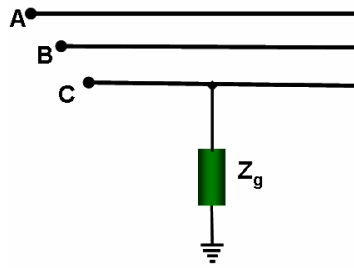


Figura C.7: Falla monofásica

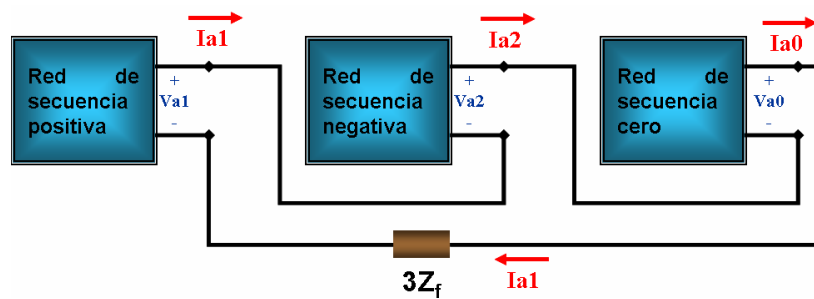


Figura C.8: Equivalente simétrico falla monofásica

(C.11). Del mismo modo se observa en las gráficas (C.10) y (C.12) que la corriente de fase fallada es tres veces la corriente de secuencia cero, a pesar que en el diagrama fasorial se graficaron las tres corrientes solo es apreciable la corriente de la fase afectada.

C.3.2. Fallas bifásicas o línea a línea L-L

Para representar una falla línea a línea a través de un impedancia Z_f se conectan los segmentos hipotéticos de las tres líneas en la falla, de la manera mostrada en la figura (C.13), sin perder generalidad la falla L-L se considera que está entre las fases a y b . Para determinar los valores de tensión y corriente se utiliza el equivalente mostrado en (C.14).

Las corrientes entre las fases involucradas son de igual magnitud y de sentidos opuestos y el valor de secuencia cero es muy pequeña pero no nula (fig. (C.18)), esto debido a la característica del sistema de prueba.

C.3.3. Fallas bifásicas tierra o doble línea tierra L-L-T

Para una falla L-L-T, los segmentos hipotéticos se conectan como lo muestra la gráfica (C.19), como en el caso anterior se considera que la falla ocurre entre las fases a y b .

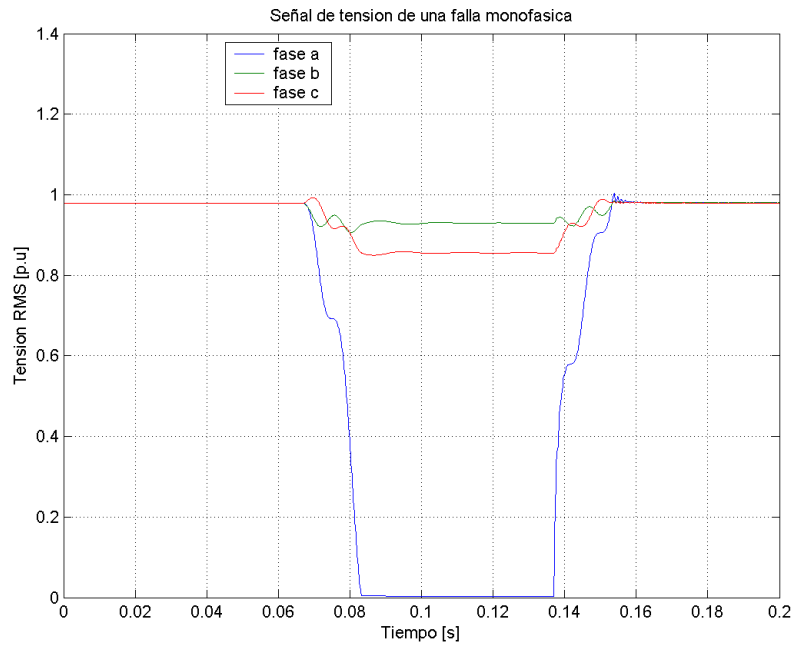


Figura C.9: Señal de tensión bajo una falla monofásica

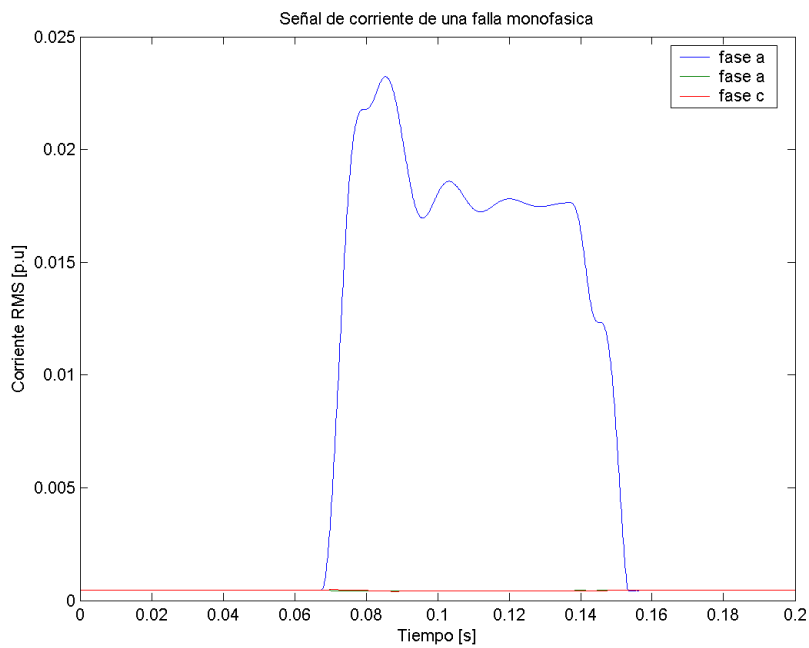


Figura C.10: Señal de corriente bajo una falla monofásica

El circuito de equivalente simétrico se muestra en (C.20). En este caso la corriente de secuencia cero para este tipo de falla es más apreciable debido al camino que ofrece Z_g .

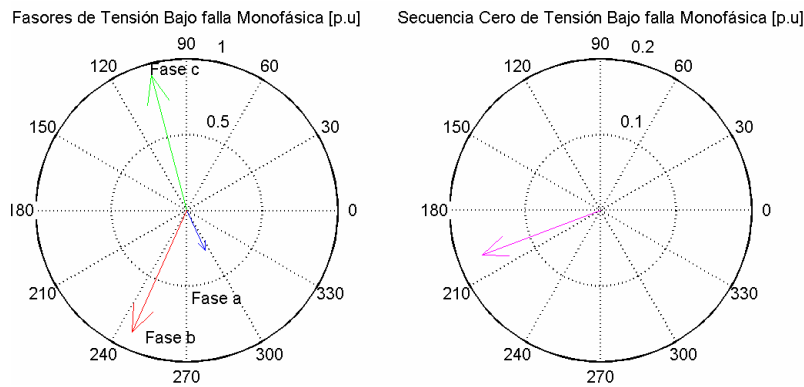


Figura C.11: Fasores de tensión y secuencia cero bajo una falla monofásica

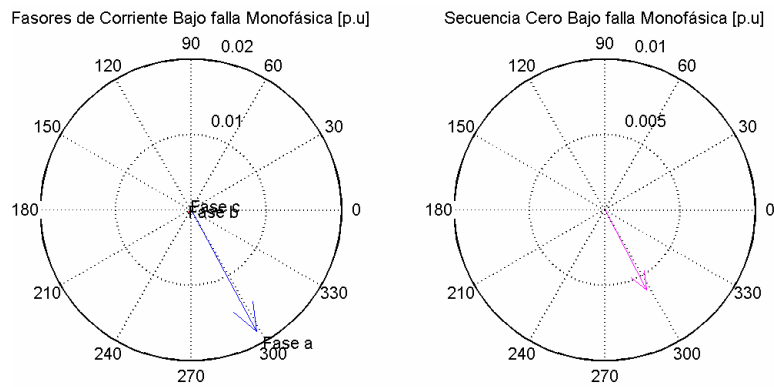


Figura C.12: Fasores de corriente y secuencia cero bajo una falla monofásica

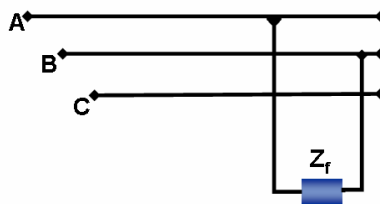


Figura C.13: Falla bifásica

C.4. Comportamiento de las redes de secuencia cuando la falla es vista desde el alimentador

Las redes de secuencia son transformaciones lineales de los fasores de fase (tensión y/o corriente) de elementos simétricos con el objetivo de facilitar el análisis de sistemas de potencia. Uno de los

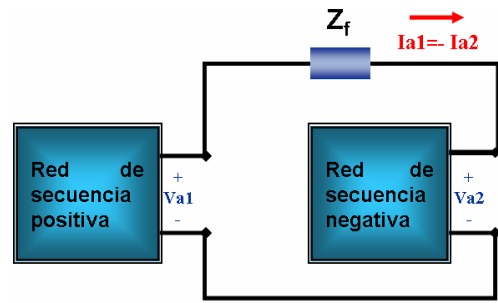


Figura C.14: Equivalente simétrico falla bifásica

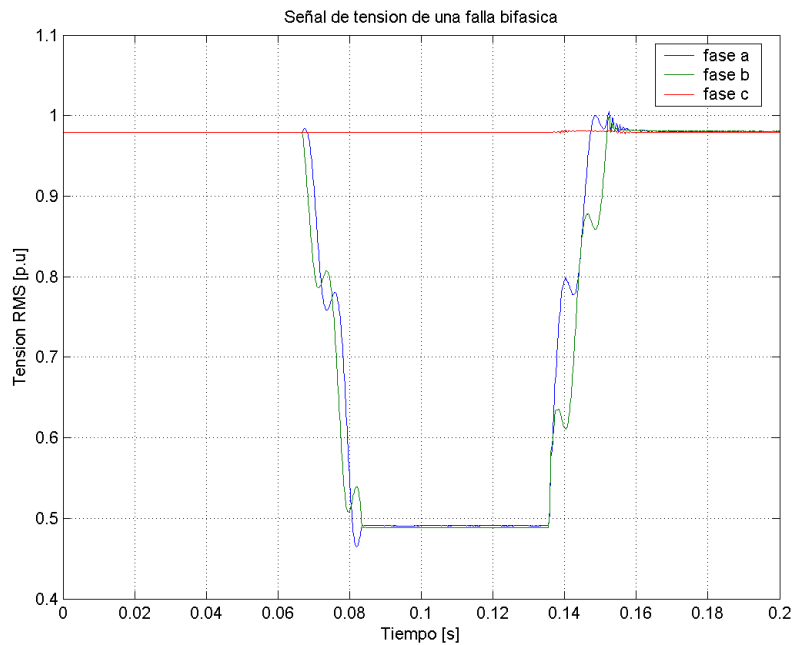


Figura C.15: Señal de tensión bajo una falla bifásica

campos donde más ayuda las redes de secuencia es en el análisis de fallas, las cuales son de índole asimétrica. Cuando se llevan los fasores a tres espacios simultáneos donde existe la simetría, generalmente es con respecto a la fase *a*.

Las redes de secuencia se aplican al punto donde ocurre la falla. Para cualquier otro punto no se cumple a cabalidad con lo propuestos por estas redes. Unas de las razones por las que no se cumple esto son:

1. A lo largo del circuito de potencia existen divisores de corriente, como lo son las líneas mismas y los efectos capacitivos.

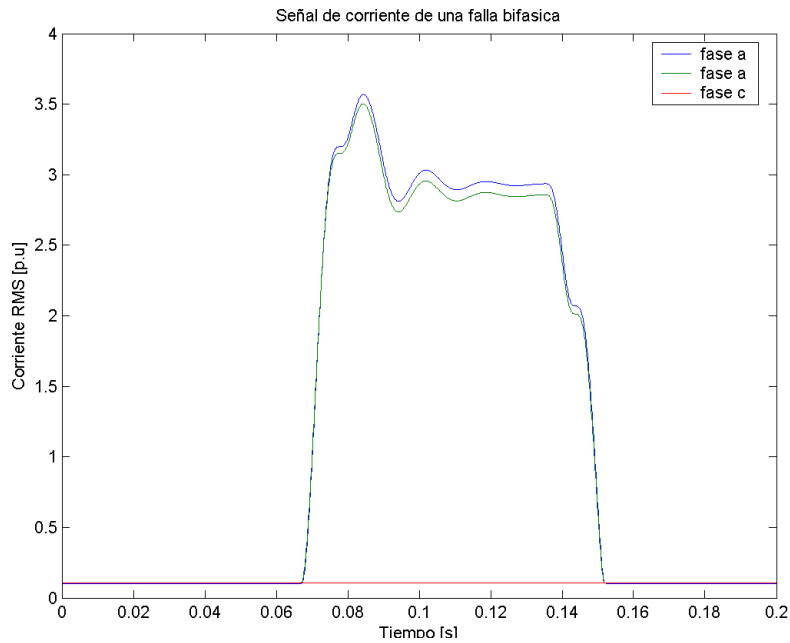


Figura C.16: Señal de corriente bajo una falla bifásica

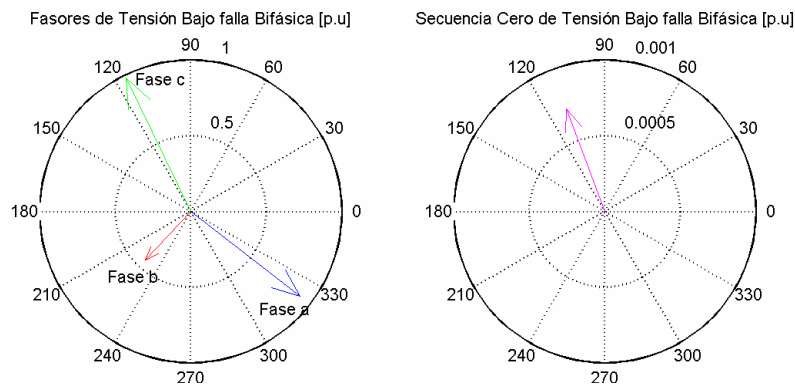


Figura C.17: Fasores de tensión y secuencia cero bajo una falla bifásica

2. La secuencia cero tiene una particularidad, y es que depende de la topología del circuito por su esquema de conexión como lo son los transformadores y generadores.
3. La secuencia positiva es la única que tiene fuente de alimentación es decir tiene elementos activos.

Es importante recordar que lo que se hace con las secuencias es un equivalente de Thevenin, al momento de desplegarlo en su totalidad, para aplicar las diferentes técnicas para su análisis con cada secuencia, empiezan aparecer cambios, en algunos casos considerables para cada barra en la

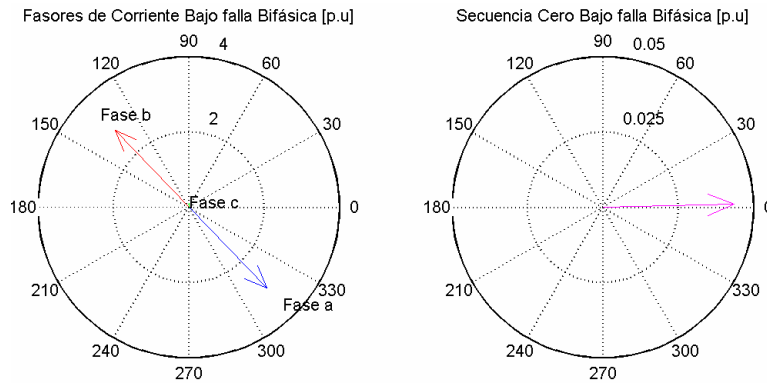


Figura C.18: Fasores de corriente y secuencia cero bajo una falla bifásica

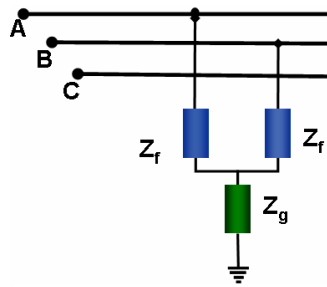


Figura C.19: Falla bifásica Tierra

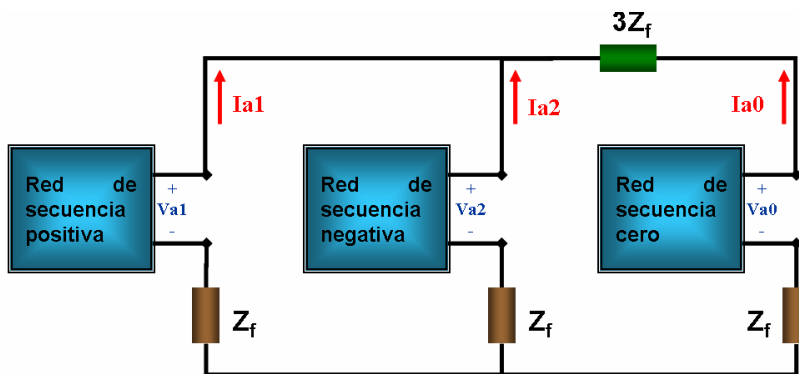


Figura C.20: Equivalente simétrico falla bifásica tierra

que se quiere conocer los valores de tensión y corriente.

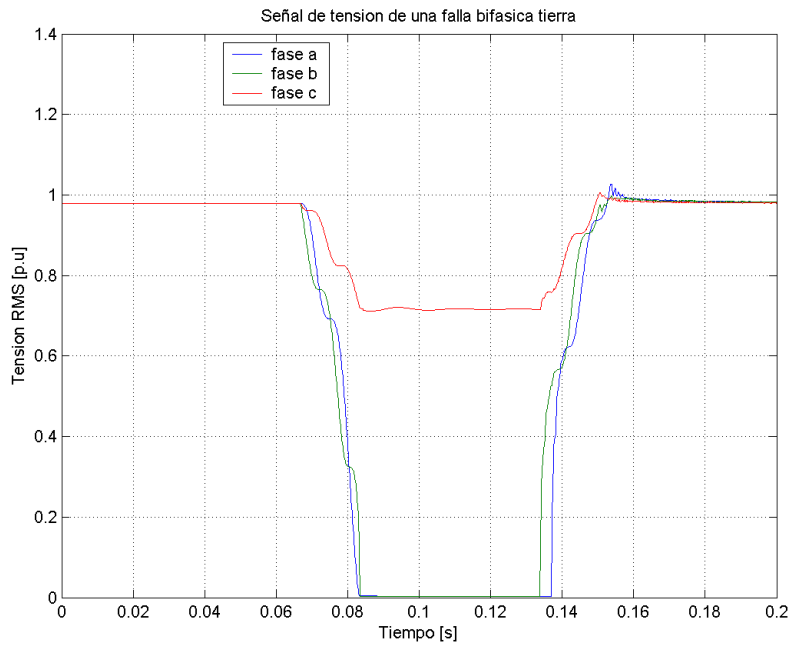


Figura C.21: Señal de tensión bajo una falla bifásica tierra

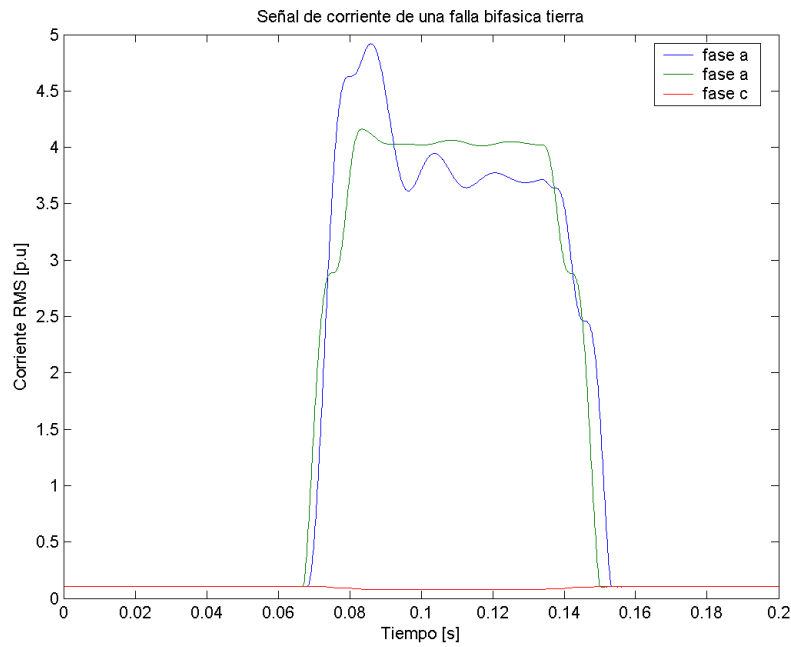


Figura C.22: Señal de corriente bajo una falla bifásica tierra

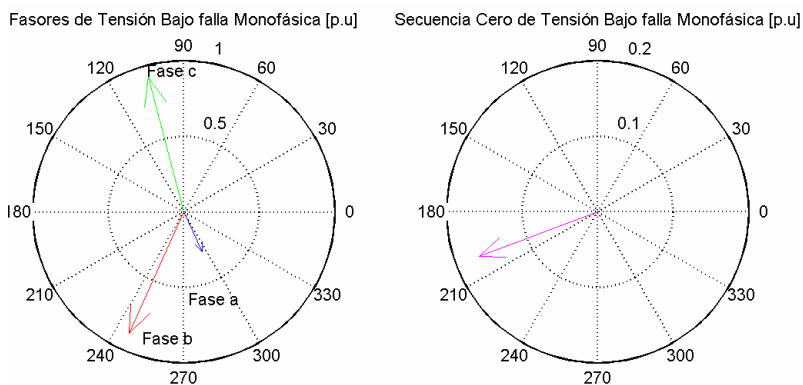


Figura C.23: Fasores de tensión y secuencia cero bajo una falla bifásica tierra.

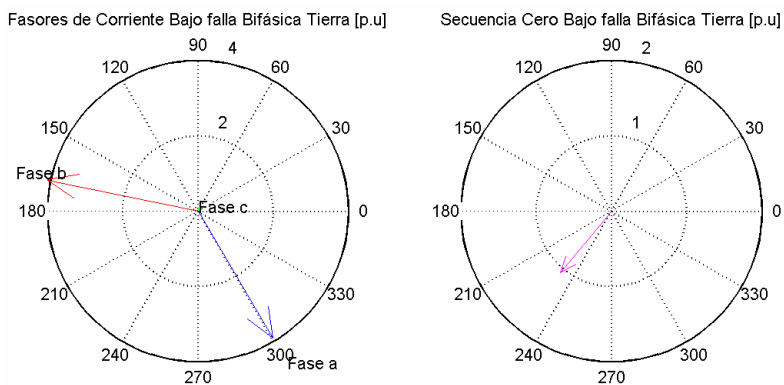


Figura C.24: Fasores de corriente y secuencia cero bajo una falla bifásica tierra

Anexo D

Componentes principales para cada modelo propuesto

El objetivo de éste apéndice es mostrar las componentes principales y el significado de las más representativas¹ de cada modelo propuesto para la localización de fallas.

D.1. Modelo tensión-corriente

Los valores en cada una de las componentes para cada tipo de falla son similares, al menos en las dos primeras, lo cual es muy bueno para el análisis porque indica que el modelo tiene una característica general en cada tipo de falla.

D.1.1. Descripción de las componentes:

- **Primera:** Como se observa en los cuadros (D.1), (D.3), (D.5) y (D.7), la mayor contribución está dada por la máxima corriente con un factor superior al 0.94 en contraposición, es decir que los datos están organizados desde aquellas fallas que no exigen gran corriente hasta las que exigen el sistema al menos en una fase para sobrecargarla. Esto se denomina “*impacto del cortocircuito*”.
- **Segunda:** Aquí las contribuciones provienen de las dos potencias (activa y reactiva) en proporciones similares, predominando la activa sobre la reactiva, lo cual formula la idea real de la verdadera exigencia del sistema, tanto por parte de los generadores como los primeros motores. Esta se denomina “*exigencia*”.

¹En todos los modelos se pudo explicar la mayor parte de la información en las dos primeras componentes.

Cuadro D.1: Componentes principales en modelo Tensión-Corriente L-T

| | Comp 1 | Comp 2 | Comp 3 | Comp 4 |
|------------|-----------|---------|----------|---------|
| MTF | 0,22342 | 0,44649 | -0,10882 | 0,85959 |
| MCF | -0,97347 | 0,1006 | 0,025302 | 0,20397 |
| PF | -0,03356 | 0,67765 | -0,60299 | -0,4196 |
| QF | -0,036342 | -0,5756 | -0,78989 | 0,20843 |

Cuadro D.3: Componentes principales en modelo Tensión-Corriente L-L

| | Comp 1 | Comp 2 | Comp 3 | Comp 4 |
|------------|------------|-----------|----------|----------|
| MTF | 0,18414 | -0,16806 | 0,8472 | 0,46915 |
| MCF | -0,98287 | -0,039163 | 0,1576 | 0,087146 |
| PF | -0,0059602 | 0,75738 | -0,19376 | 0,62354 |
| QF | 0,0048145 | -0,62976 | -0,4689 | 0,61928 |

Los pesos de cada componente y el peso acumulado se observa en los cuadros (D.9) y (D.11), respectivamente.

D.2. Modelo de corriente

El resultado del ACP indica que se puede trabajar con una única componente, pero al pensar en una dimensión los datos no se pueden apreciar por lo cual se trabajó con las dos primeras componentes, es decir en un plano. (Ver cuadros (D.13), (D.15), (D.17) y (D.19)).

D.2.1. Descripción de las componentes:

- **Primera:** Sencillamente es la “*Corriente de falla*”.
- **Segunda:** Se denomina, “*Corriente de fuga*”.

Los pesos de cada componente y el peso acumulado se observa en los cuadros (D.21) y (D.23), respectivamente.

D.3. Modelo de tensión

En este modelo el comportamiento es similar al anterior por lo cual su análisis no cambia significativamente. (Ver cuadros (D.25), (D.27), (D.29) y (D.31))

D.3.1. Descripción de las componentes:

- **Primera:** Se observa que es la tensión mínima y la media de las tensiones, las que lideran

Cuadro D.5: Componentes principales en modelo Tensión-Corriente L-L-T

| | Comp 1 | Comp 2 | Comp 3 | Comp 4 |
|------------|---------------|---------------|---------------|---------------|
| MTF | 0,24612 | 0,26416 | -0,10793 | -0,92628 |
| MCF | -0,96806 | 0,11416 | -0,031174 | -0,22103 |
| PF | 0,03359 | 0,74372 | 0,65169 | 0,14508 |
| QF | -0,034001 | -0,60338 | 0,75012 | -0,26851 |

Cuadro D.7: Componentes principales en modelo Tensión-Corriente L-L-L

| | Comp 1 | Comp 2 | Comp 3 | Comp 4 |
|------------|---------------|---------------|---------------|---------------|
| MTF | 0,31165 | 0,21531 | -0,25386 | -0,88999 |
| MCF | -0,94759 | 0,14216 | -0,08557 | -0,27302 |
| PF | 0,065884 | 0,93444 | 0,31105 | 0,16041 |
| QF | -0,024637 | -0,24547 | 0,91186 | -0,32811 |

Cuadro D.9: Peso de cada CP modelo Tensión-Corriente.

| | L-T | L-L | L-L-T | L-L-L |
|--------------|------------|------------|--------------|--------------|
| Comp1 | 0,96 | 0,90 | 0,89 | 0,87 |
| Comp2 | 0,02 | 0,09 | 0,08 | 0,10 |
| Comp3 | 0,00 | 0,00 | 0,02 | 0,01 |
| Comp4 | 0,00 | 0,00 | 0,00 | 0,00 |

Cuadro D.11: Peso acumulado CP modelo Tensión-Corriente (en %).

| | L-T | L-L | L-L-T | L-L-L |
|--------------|------------|------------|--------------|--------------|
| Comp1 | 96,90 | 90,09 | 89,61 | 87,95 |
| Comp2 | 99,11 | 99,23 | 97,88 | 98,05 |
| Comp3 | 99,95 | 99,74 | 99,96 | 99,98 |
| Comp4 | 100 | 100 | 100 | 100 |

Cuadro D.13: Componentes principales en modelo corriente L-T

| | Comp 1 | Comp 2 | Comp 3 | Comp 4 |
|------------|---------------|---------------|---------------|---------------|
| MCF | -0,80102 | -0,1328 | -0,20424 | 0,54683 |
| PHC | -0,26615 | -0,78787 | 0,087982 | -0,54835 |
| DEC | -0,46319 | 0,56918 | -0,24739 | -0,63268 |
| SCC | -0,27016 | 0,19406 | 0,94305 | 0,0036181 |

Cuadro D.15: Componentes principales en modelo corriente L-L

| | Comp 1 | Comp 2 | Comp 3 | Comp 4 |
|------------|---------------|---------------|---------------|---------------|
| MCF | -0,75621 | -0,57713 | 0,30777 | 0,018858 |
| PHC | -0,49448 | 0,5256 | -0,26847 | 0,6381 |
| DEC | -0,42852 | 0,41625 | -0,22518 | -0,76967 |
| SCC | -0,0039476 | -0,46627 | -0,88459 | 0,0088309 |

Cuadro D.17: Componentes principales en modelo corriente L-L-T

| | Comp 1 | Comp 2 | Comp 3 | Comp 4 |
|------------|---------------|---------------|---------------|---------------|
| MCF | -0,72849 | -0,21866 | 0,64886 | -0,021818 |
| PHC | -0,47802 | -0,12908 | -0,60127 | -0,62715 |
| DEC | -0,41943 | -0,059384 | -0,46478 | 0,77752 |
| SCC | -0,25472 | 0,9654 | 0,037981 | -0,04097 |

Cuadro D.19: Componentes principales en modelo corriente L-L-L

| | Comp 1 | Comp 2 | Comp 3 | Comp 4 |
|------------|---------------|---------------|---------------|---------------|
| MCF | -0,65249 | -0,27931 | -0,40404 | -0,57706 |
| PHC | -0,64487 | -0,28352 | 0,41382 | 0,57664 |
| DEC | -0,0076375 | 0,0030013 | -0,81575 | 0,57835 |
| SCC | -0,39793 | 0,91739 | 0,0075462 | 0,00062813 |

Cuadro D.21: Peso de cada CP modelo corriente.

| | L-T | L-L | L-L-T | L-L-L |
|--------------|------------|------------|--------------|--------------|
| Comp1 | 0,997 | 0,999 | 0,992 | 0,975 |
| Comp2 | 0,000 | 0,000 | 0,007 | 0,024 |
| Comp3 | 0,000 | 0,000 | 0,000 | 0,000 |
| Comp4 | 0,000 | 0,000 | 0,000 | 0,000 |

Cuadro D.23: Peso acumulado CP modelo corriente (en %).

| | L-T | L-L | L-L-T | L-L-L |
|--------------|------------|------------|--------------|--------------|
| Comp1 | 99,99 | 90,09 | 89,61 | 87,95 |
| Comp2 | 99,99 | 99,99 | 99,97 | 99,98 |
| Comp3 | 99,99 | 99,99 | 99,99 | 99,99 |
| Comp4 | 100 | 100 | 100 | 100 |

Cuadro D.25: Componentes principales en modelo tensión L-T

| | Comp 1 | Comp 2 | Comp 3 | Comp 4 |
|------------|---------------|---------------|---------------|---------------|
| MTF | 0,82863 | 0,22131 | 0,21418 | -0,46746 |
| PDT | 0,32926 | -0,097987 | 0,52579 | 0,77816 |
| DET | -0,42607 | 0,0061947 | 0,82321 | -0,37517 |
| SCT | -0,15303 | 0,97025 | -0,0010079 | 0,18761 |

Cuadro D.27: Componentes principales en modelo tensión L-L

| | Comp 1 | Comp 2 | Comp 3 | Comp 4 |
|------------|---------------|---------------|---------------|---------------|
| MTF | 0,79102 | 0,3618 | -0,0044816 | -0,49333 |
| PDT | 0,43489 | -0,73324 | -0,49629 | 0,16408 |
| DET | -0,41882 | -0,20494 | -0,33492 | -0,81879 |
| SCT | -0,09877 | 0,53802 | -0,80094 | 0,24348 |

Cuadro D.29: Componentes principales en modelo tensión L-L-T

| | Comp 1 | Comp 2 | Comp 3 | Comp 4 |
|------------|---------------|---------------|---------------|---------------|
| MTF | 0,7442 | 0,21527 | -0,6117 | -0,16015 |
| PDT | 0,56491 | -0,23748 | 0,69981 | -0,36708 |
| DET | -0,32373 | -0,32877 | -0,29004 | -0,83845 |
| SCT | -0,14913 | 0,88836 | 0,22797 | -0,36962 |

Cuadro D.31: Componentes principales en modelo tensión L-L-L

| | Comp 1 | Comp 2 | Comp 3 | Comp 4 |
|------------|---------------|---------------|---------------|---------------|
| MTF | 0,72937 | 0,82565 | 0,40089 | -0,54817 |
| PDT | 0,67654 | -0,2319 | -0,43403 | 0,54784 |
| DET | -0,045868 | -0,26952 | -0,72517 | -0,63197 |
| SCT | -0,090557 | -0,931 | 0,3536 | -0,0021221 |

Cuadro D.33: Peso de cada CP modelo tensión.

| | L-T | L-L | L-L-T | L-L-L |
|--------------|------------|------------|--------------|--------------|
| Comp1 | 0,985 | 0,853 | 0,971 | 0,981 |
| Comp2 | 0,012 | 0,103 | 0,026 | 0,014 |
| Comp3 | 0,001 | 0,042 | 0,001 | 0,004 |
| Comp4 | 0,000 | 0,000 | 0,000 | 0,000 |

Cuadro D.35: Peso acumulado CP modelo tensión (en %)

| | L-T | L-L | L-L-T | L-L-L |
|--------------|------------|------------|--------------|--------------|
| Comp1 | 98,52 | 85,34 | 97,18 | 98,15 |
| Comp2 | 99,81 | 95,66 | 99,84 | 99,59 |
| Comp3 | 99,98 | 99,92 | 99,98 | 99,99 |
| Comp4 | 100 | 100 | 100 | 100 |

Anexo E

PRtool Herramienta para el reconocimiento estadístico de patrones

E.1. Introducción

Para facilitar la implementación de los métodos estadísticos multivariados de interdependencia, a los datos que describen el comportamiento del sistema de potencia de prueba bajo condiciones de falla, se diseñó un herramienta sencilla programada en el paquete matemático MatLab 6.5.

La cualidad más grande que tiene ésta herramienta para el reconocimiento de patrones, es que no solo ésta diseñada para ser aplicada al problema de la localización de fallas, sino que puede ser utilizada para realizar un análisis de correlación bivariada, análisis de componentes principales y un análisis de conglomerados a cualquier tipo de datos que esten organizados en un arreglo matricial $n \times p$, como se indicó en 2.1.

E.2. Como utilizar PRtool

Para poder utilizar ésta herramienta estadística es necesario conocer algunos detalles que se explican a continuación:

E.2.1. Instalación

Es necesario tener instalado MatLab version 6.5. Cumpliendo con éste requisito se deben seguir las siguientes instrucciones:

1. Copiar la carpeta PRtool en la siguiente ruta C:\MATLAB6p5.

2. Agregar la carpeta PRtool al path de matlab como se indica a continuación: *File>Set Path>Add Folder...>Seleccionar PRtool>Archivos PRtool>Aceptar>save>close*. La figura (E.1) muestra el anterior procedimiento.

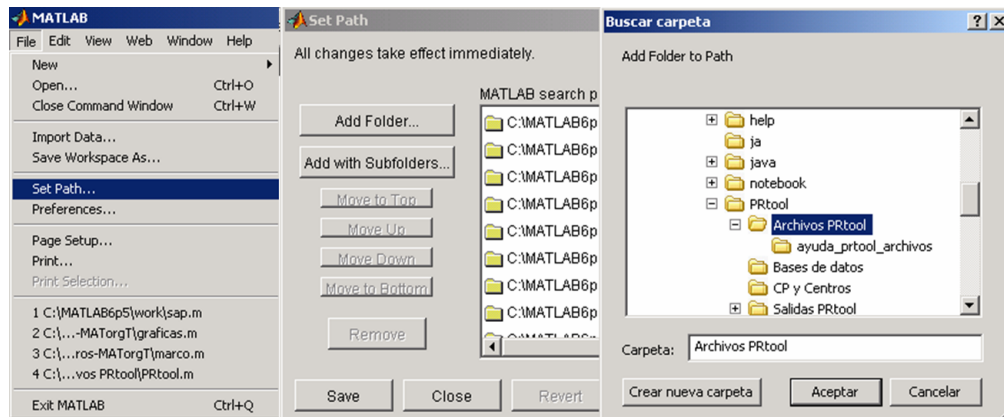


Figura E.1: Procedimiento para agregar carpeta de archivos al path de MatLab.

3. Buscar en current Directory la carpeta PRtool, como lo indica la figura (E.2).

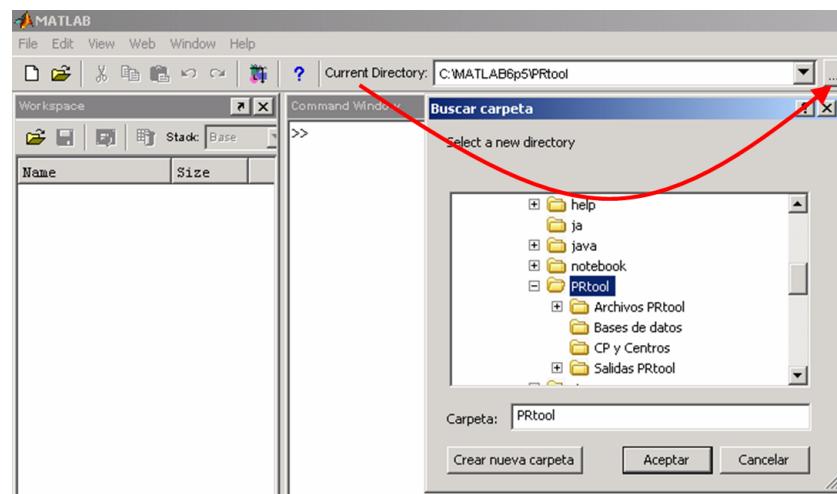


Figura E.2: Procedimiento para buscar el directorio actual de MatLab.

4. Digitando sobre la ventana de comandos de MatLab *star_PRtool*, se tiene acceso a la herramienta (ver figura (E.3)).



Figura E.3: Pantalla Inicial de PRtool

E.2.2. Utilizando PRtool

PRtool está diseñada para que se maneje de manera secuencial, es decir no se puede avanzar hasta que el proceso inmediatamente anterior no se haya realizado. A continuación se hace un recorrido por todas las funciones que ofrece esta herramienta para realizar reconocimiento de patrones. La figura (E.4) detalla cada una de las secciones de la herramienta *PRtool*.

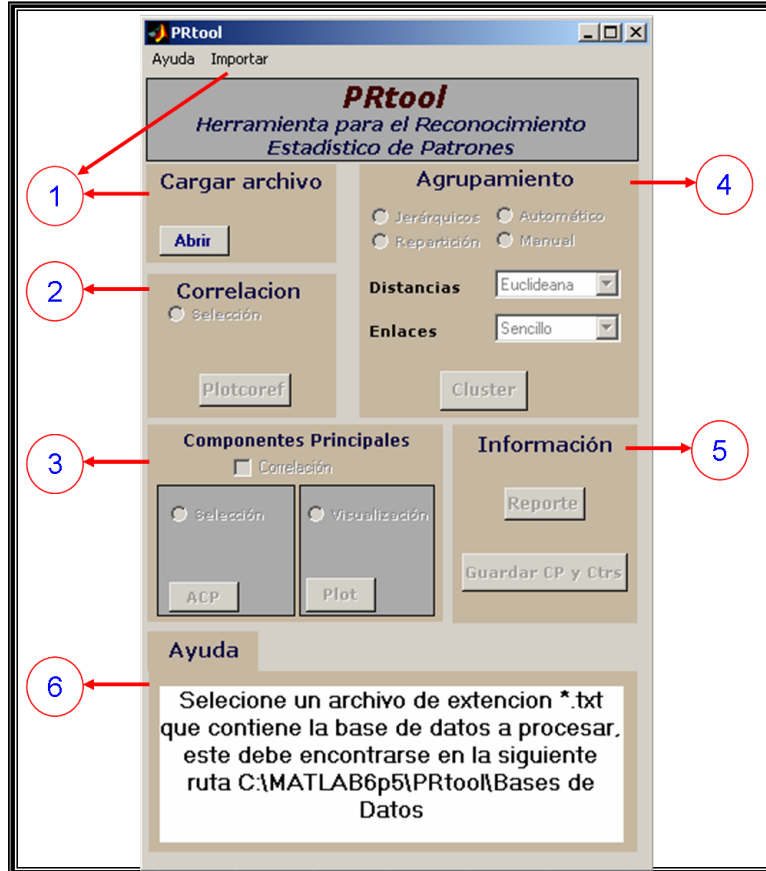


Figura E.4: Interfaz *PRtool*.

1. **Cargar Archivo:** Existen dos formas de cargar archivos a *PRtool*:

- a) Boton **Abrir**: Para cargar un archivo a *PRtool* por éste medio, se deben tener las siguientes consideraciones; primero, la base de datos debe encontrarse en un archivo con extension **.txt*¹, y segundo, se recomienda que el archivo se encuentre en el directorio *Base de datos* que se encuentra dentro de la carpeta *PRtool*.
- b) Boton **Importar**: Es posible trabajar con bases de datos que se encuentre en una extensión distinta a **.txt*. *PRtool* permite importar archivos con extensión *.xls* y *.mat*. Al igual que en el caso anterior el archivo debe encontrarse en el directorio *Base de datos* que se encuentra dentro de la carpeta *PRtool*.

2. **Correlación:** En esta sección de la interfaz se realiza un análisis de interdependencia de las variables contenidas en el archivo de datos. Soporta un máximo de doce variables, la razón

¹Formato de texto o block de notas.

es que con más de doce variables la visualización de las correlaciones se satura, impidiendo la comprensión de ésta. Cuando el número de variables supera al límite, *PRtool* permite la selección de las variables que se considere, sean las más representativas. Este procedimiento se realiza haciendo click sobre el marcador llamado selección, inmediatamente se activa un campo donde se introduce el número de las variables de interés, separadas por un espacio². Una vez seleccionadas las variables que entrarán en el análisis se procede a presionar el botón *plotcoref* , con el cual se presentará una gráfica matricial, donde se muestran: La distribución de cada variable, los valores del coeficiente de correlación bivariado³ r y los gráficos de dispersión variable a variable.

3. **Componentes Principales:** Para realizar un ACP, *PRtool* permite que el usuario defina si prefiere trabajar con la matriz de covarianza (predeterminado) ó con la matriz de correlación. Al igual que para el análisis de interdependencia, *PRtool*, permite seleccionar que variables entran a formar parte en análisis de componentes principales. Al presionar el botón *ACP*, se muestra un diagrama de barras que muestra la cantidad de información que contiene cada nueva variable o dicho de otra forma, cada componente principal. Seguidamente se activa el campo de visualización de las componentes principales, en el cual es posible seleccionar cuales se requieren visualizar. La visualización predeterminada es la de las tres primeras componentes principales.
4. **Agrupamiento:** Finalizado el ACP, se procede a realizar el análisis de conglomerados. Esta sección de *PRtool*, ofrece la opción de realizar un análisis jerárquico o un análisis de repartición. El diagrama de flujo de ésta sección se muestra en E.5.

²Este procedimiento también se realiza para ACP y para la Visualización de las CP.

³Azul: Corresponde a valores positivos. Rojo: Corresponde a valores negativos.

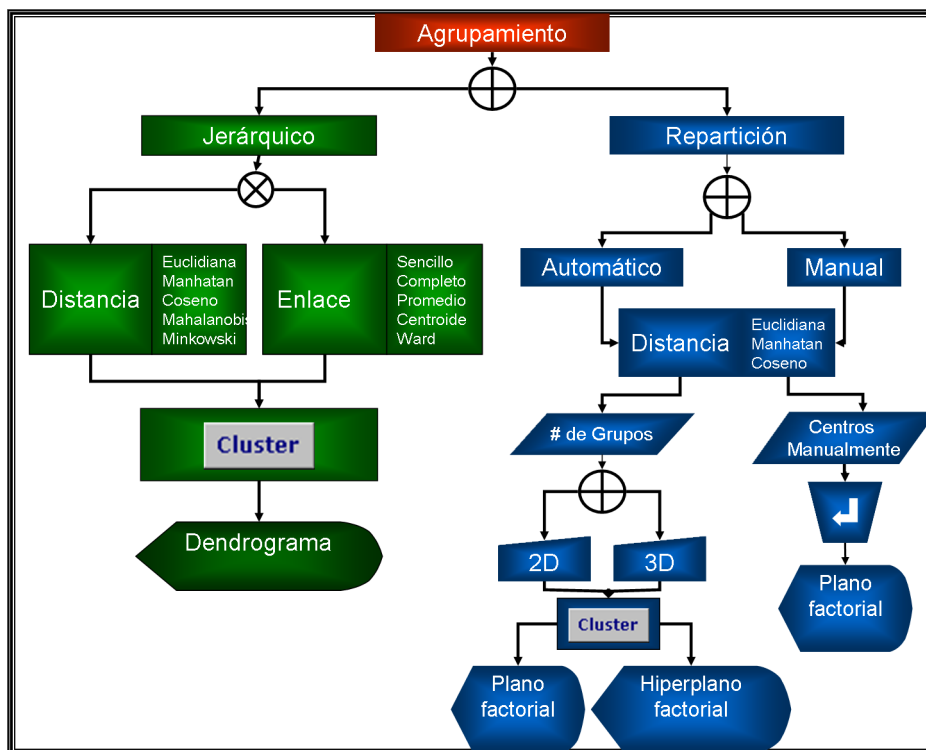


Figura E.5: Diagrama de flujo de la sección agrupamiento.

5. **Información:** Es la sección de la interfaz que se encarga de presentar un informe detallado del proceso de reconocimiento de patrones (oprimiendo el botón reporte), y además *PRtool*, permite guardar las componentes principales halladas, así como los centros de los grupos hallados, (Presionando el botón *Guardar PC y Ctrs*) en un archivo *.mat*.
6. **Ayuda:** El campo denominado ayuda presenta algunas instrucciones para el manejo de *PRtool*,

E.3. Help *PRtool*

Para conocer las funciones disponibles en la *PRtool*, digitar:

```

>>help PRtool
RECONOCIMIENTO ESTADÍSTICO DE PATRONES
% Herramienta para el reconocimiento estadístico multivariado de tipo exploratorio.
% Las funciones diseñadas para el desarrollo de esta herramienta son:
% ANÁLISIS DE CORRELACIÓN BIVARIADA
% plotcoref - Genera gráfico matricial que da informacion sobre la interdependencia de las variables.
% ver help plotcoref.
  
```

```

% ANÁLISIS DE COMPONENTES PRINCIPALES
% princomp - Genera las componentes principales (Incluido en Matlab) ver help princomp
% ANÁLISIS DE CONGLOMERADOS
% kmeans - Realiza agrupamientos por el metodo de la k-medias (Incluido en Matlab) ver help kmeans
% dendrogram - genera el dendrograma de un análisis de agrupamiento jerárquico. (Incluido Matlab)
ver help dendrogram
% grupos - Realiza agrupamientos con los centros que el usuario definio. Ver help grupos.
% INFORMACIÓN
% reporte - Genera un archivo txt en el cual se dan los resultados obtenidos del análisis exploratorio
ver help reporte.
% Autores:
%%%%%%%%%EDWIN SALAMANCA TORRES E3T UIS
%%%%%%%%%MARCO FIDEL SUAREZ SANCHEZ E3T UIS
%%%%%%%%%FEBRERO 2006

```

E.4. Funciones creadas para *PRtool*

A continuación se enumeran las funciones que se crearon para ser implementadas en *PRtool*, también se muestra la sintaxis y su utilidad.

E.4.1. *guardar.m*

Genera un archivo **.txt* en el cual se guardan matrices como base de datos, los cuales van a ser usados para un analisis exploratorio "Reconocimiento de Patrones".

SINTAXIS: `guardar(D,E,F)`, donde D es la matriz de datos $n \times p$, donde n es el número de individuos y p el número de variables. E es un arreglo (cell array) que contiene el nombre de las p variables. F es el nombre del archivo sin extensión.

ej: `guardar(D,{'V1','V2',...,'Vn'},'nuevabase')`

E.4.2. *plotcoref.m*

Esta función da información sobre la interdependencia de las variables, a través de una gráfica matricial que tiene las siguientes características: La triangular superior representa la correlación bivariada donde el color⁴ representa el signo de esta. La diagonal representa la distribución de cada variable. La triangular inferior representa el gráfico de dispersión de los datos.

SINTAXIS: `plotcoref(X)` o `plotcoref(X,E)`, donde X es la matriz de datos y E son las etiquetas.

⁴azul (+) y roja (-)

E.4.3. *grupos.m*

Esta función le permite graficar que datos corresponde a los centros definidos por el usuario utilizando una métrica definida por el mismo.

SINTAXIS: `grupos(c,D,d)`, donde c es la matriz de centros $k \times p$, D es la matriz de datos $n \times p$, d es la métrica a emplear entre las siguientes; '1' para Euclideana, '2' para Manhattan y '3' para Coseno. Por defecto es la norma Euclideana.

E.4.4. *reporte.m*

Genera un archivo **.txt* en el cual se dan los resultados obtenidos del análisis exploratorio.

SINTAXIS: `reporte(R,Ex,Pc,C)`, donde: R es la matriz de correlaciones. Ex son los porcentajes de varianza explicada (p.u) de cada componente principal y Pc son los pesos de cada componente. C son los centros de los grupos formados.

E.4.5. *barra.m*

Realiza un diagrama de barras donde se muestra el porcentaje de información que está contenida en cada componente principal

SINTAXIS: `barra(ex,'color')`, donde ex es un vector donde los valores están en p.u y 'color' corresponde a un color ya definido en MatLab ('c','r','b','k','m','y','g','w')

E.4.6. *fasor2.m*

Calcula los fasores de una señal de tiempo trifásica.

SINTAXIS: `F=fasor2(Ft,n,n1,s)`, donde Ft es la señal en el tiempo la cual debe ser una matriz $N \times 3$ (debe contener las tres fases), n es el número de muestras por ciclo, $n1$ es el ciclo en el cual ocurre la falla (función debe contener por lo menos $n1+4$ ciclos, S es el tipo de señal (1 para tensión y 2 para corriente), F es un vector complejo donde se halla en forma rectangular el fasor.

E.4.7. *abrir.m*

SINTAXIS: `[E,D]=abrir(lectura)`, donde *lectura* es una base de datos con extensión *.txt*, E es un cell-array donde se almacena las etiquetas de las variables, D es la matriz $n \times p$ que corresponde a la base de datos.

E.4.8. *dato.m*

Calcula los datos estadísticos necesarios para el modelo de corriente.

SINTAXIS: `d=dato(Ff)`, donde *Ff* es el fasor de corriente, `d=[máximo, media, desviación-estandar, corriente-sec(0)]`.

E.4.9. *datom.m*

Calcula los datos estadísticos necesarios para el modelo de tensión.

SINTAXIS: `d=datom(Ff)`, donde *Ff* es el fasor de tensión `d=[mínimo, media, desviación-estandar tensión-sec(0)]`.

E.4.10. *iden.m*

Identifica el tipo de falla de cortocircuito basandose en los fasores de corriente.

SINTAXIS: `id=iden(Fct)`, donde *id* son valores entre 0 y 4, y según su valor se catalogan las fallas: 0 la falla no se pudo identificar, 1 la falla es monofásica, 2 la falla es bifásica, 3 la falla es bifásica a tierra, 4 la falla es trifásica, *Fcf* corresponde a los fasores de corriente en estado de falla en p.u.

E.4.11. *importar.m*

Importa bases de datos con extensión *.xls* y *.mat*.

SINTAXIS: `[E,D]=importar(lectura)`, *lectura* es una base de datos con extension *.xls* ó *.mat*, con ciertas características para ser leida. *E* es un cell array donde se almacenan las etiquetas de las variables, *D* es la matriz $n \times p$ que corresponde a la base de datos. En los archivos *.xls* no se debe etiquetar los individuos, solo las variables. Para archivos de *.mat* las etiquetas deben estar guardadas en *E* (cell array) y en *D*, la matriz de datos.

E.4.12. *pricom.m*

Calcula las componentes principales a partir de la matriz de correlación de los datos.

SINTAXIS: `[pc,score,explained,Lamda]=pricom(D)`, donde *D* es la matriz de correlación de los datos $n \times p$, *pc* es una matriz con los coeficientes para la transformacion lineal, *score* son los valores de las componentes, *explained* son los porcentajes de varianza explicada por cada componente, *Lamda* son los valores propios de la Matriz de correlación.

E.4.13. `locah.m`

Localiza una de ocho zonas definidas para el sistema prototipo

SINTAXIS: `[z,c,y]=locah(d,id)`, donde d es el vector de variables, id es el identificador del tipo de falla, z es un valor que corresponde a una de las zonas definidas (1,2,3,4,5,6,7,8), c es una matriz que contiene los centros empleados para la localización, y es una vector en el espacio de componentes.

E.5. Ejemplo *PRtool*

A continuación se utilizará *PRtool* para realizar un reconocimiento de patrones a la base de datos llamada *ejemploPRtool.txt*, existente en la carpeta base de datos, ésta consta de 600 observaciones y de siete variables con una interdependencia lineal alta entre ellas.

- **Correlación:** La herramienta *PRtool* calcula el coeficiente de correlación, muestra la distribución de cada variable y hace un gráfico de la dispersión de observaciones, también permite la exclusión de variables según el criterio del analista de datos. La gráfica E.6 (a) muestra la correlación teniendo en cuenta todas las variables, la gráfica E.6 (b) muestra la manera de excluir variables del análisis y la gráfica E.6 (c) muestra el resultado del análisis excluyendo las variables a y d.

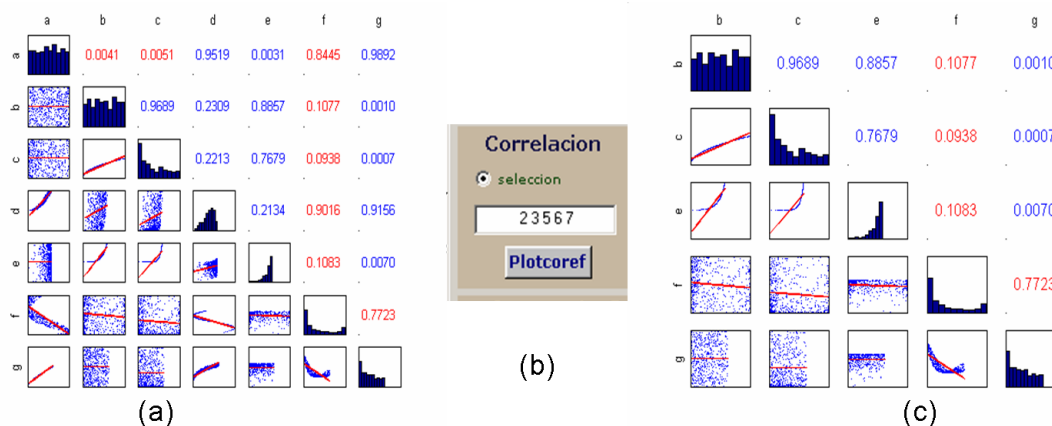


Figura E.6: Correlación bivariada, (a) Con todas las variables. (b) Manera de excluir variables. (c) Excluyendo las variables a y d

- **ACP:** *PRtool* también permite la exclusión de variables para realizar un ACP. La gráfica E.7 (a) muestra la correlación teniendo en cuenta todas las variables, la gráfica E.7 (b) muestra

la manera de excluir variables del análisis y la gráfica E.7 (c) muestra la el resultado del análisis excluyendo las variables a y d. Para fines de visualización se presenta la opción predeterminada de mostrar el hiperplano formado por las tres primeras componentes o de mostrar el plano factorial formado por cualquier par de componentes principales (ver fig E.8).

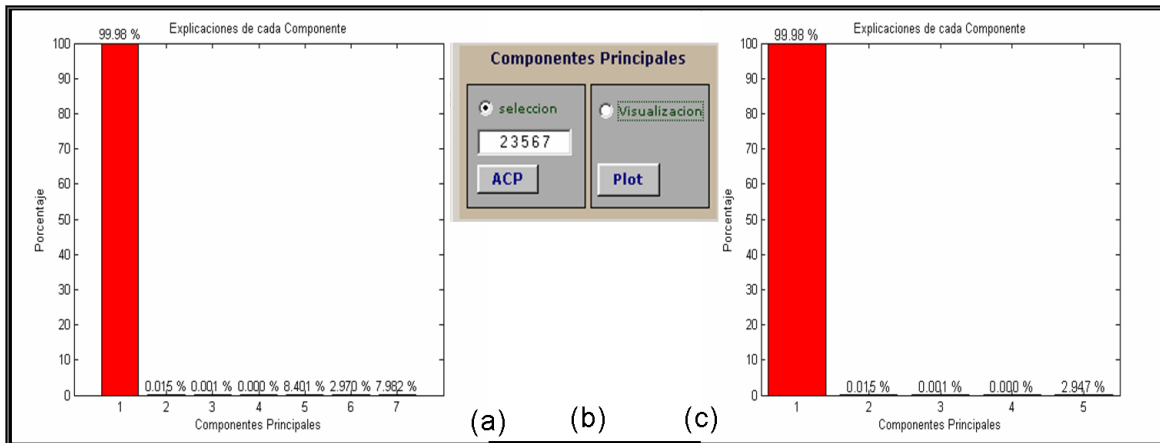


Figura E.7: Explicaciones ACP, Base de datos ejemplo RPT.txt, (a) Con todas las variables. (b) Manera de excluir variables. (c) Excluyendo las variables a y d

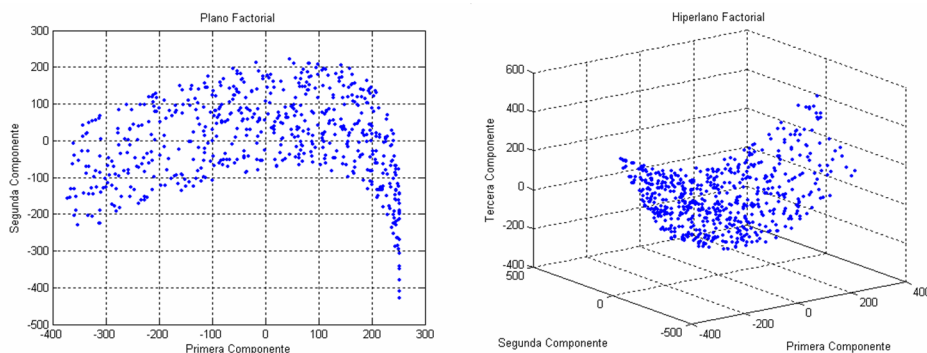


Figura E.8: Formas de visualización de componentes principales.

- Agrupamiento:** En este ejemplo no se realizará el agrupamiento jerárquico, ya que por la gran cantidad de observaciones no es recomendable, por lo tanto se opta por realizar un agrupamiento por repartición. *PRtool* permite elegir entre tres tipos de distancia para realizar un agrupamiento por repartición (Euclídea, Manhattan y coseno), también ofrece la opción de visualizar el agrupamiento en un plano factorial o en un hiperplano tridimensional, detallando la cobertura aproximada de cada grupo. En éste ejemplo se formaron 3 grupos, como se observa en E.9 y E.10.

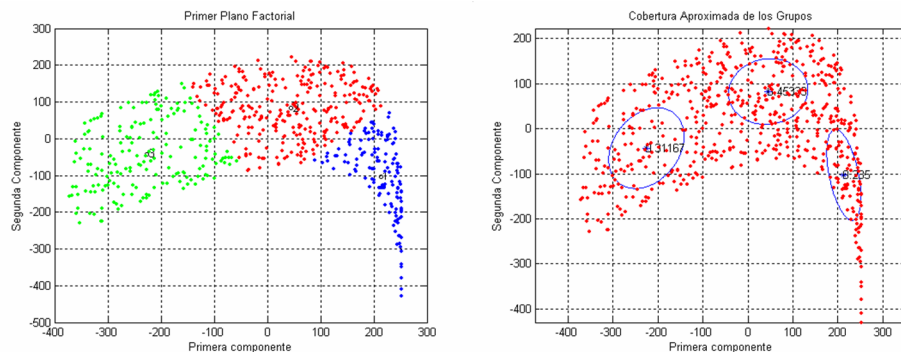


Figura E.9: Agrupamientos en el primer plano factorial y cobertura aproximada de cada grupo

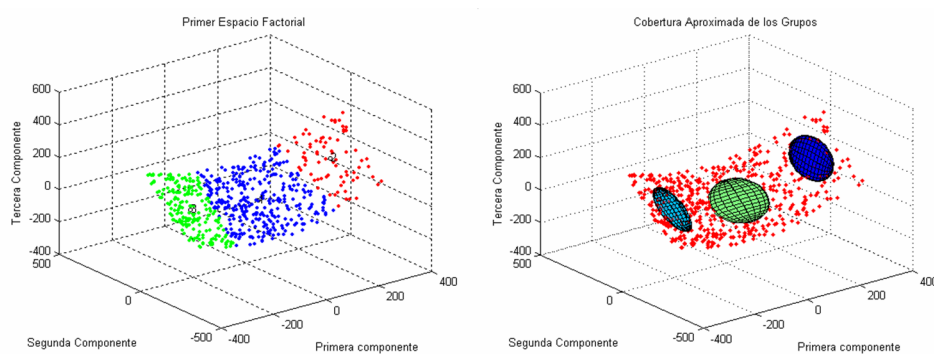


Figura E.10: Agrupamientos en hiperplano factorial y cobertura aproximada de cada grupo

Otra de las ventajas que tiene *PRtool* es que permite indicar los centros de los grupos manualmente, según el criterio del analista de los datos, tal y como muestra el gráfico E.11.

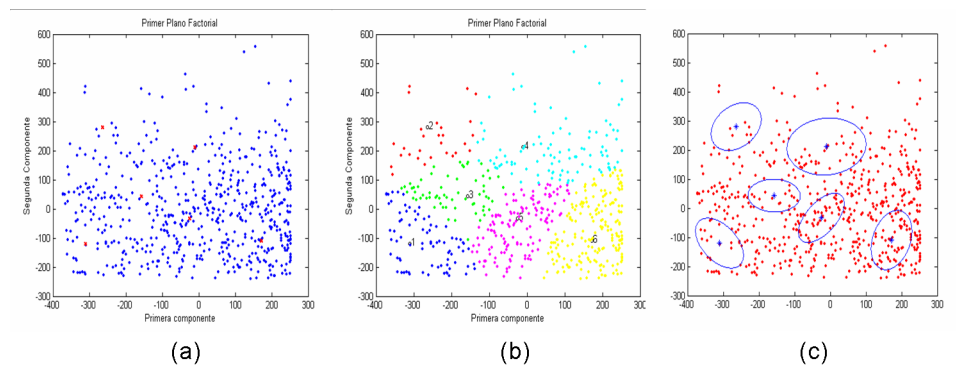


Figura E.11: Agrupamientos en el primer plano factorial y cobertura aproximada de cada grupo por medio de repartición manual. (a) Ubicación de centros. (b) Grupos formados. (c) Cobertura aproximada de cada grupo.

- **Información:** *PRtool* genera un informe del todo el proceso estadístico multivariado realizado por medio del botón llamado *reporte*, también permite almacenar en un archivo **.mat* las componentes principales calculadas y los centros utilizados para realizar los agrupamientos. La gráfica E.12 muestra el procedimiento para almacenar ésta información.

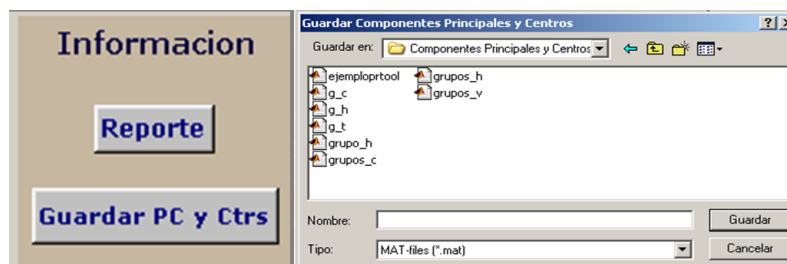


Figura E.12: Forma de almacenar las CP y los centros utilizados

Bibliografía

- [P.D, 2002] Peña, Daniel, *Análisis de Datos Multivariantes*, Mc Graw Hill.
- [D.L, 2002] Diaz M, Luis Guillermo, *Estadística Multivariada: Inferencia y Métodos*, Universidad Nacional.
- [A.H, 1999] Hair Anderson, Tatham Black, *Análisis Multivariante*, Prentice Hall.
- [R.T1, 2005] Rivera Flórez, Tulia Esther, *Componentes Principales*, Diplomado En Métodos Estadísticos Multivariados, Universidad Industrial de Santander. Septiembre 2005
- [O.P, 2000] Carvajal Olaya, Patricia, Soto Mejia, Jose y Trejos Carpintero, Álvaro Antonio, *Búsqueda de la Relación entre Áreas Icfes en Matemáticas, Físico, Lenguaje Y Rendimiento en Matemáticas I y Matemáticas II a través del Análisis de Componentes Principales*. Grupo De Investigación Estudio y Aplicación de Herramientas Estadísticas Modernas en la Solución de Problemas del Entorno, Universidad Tecnología de Pereira.
- [D.R, 1998] Duda, Richard O, *Pattern Classification*. Pag.597
- [R.A, 2002] Rencher, Alvin C, *Methods of Multivariate Analysis*, A John Wiley & Sons Inc. Publication. Brigham Young University Pagina 380.
- [H.W, 2003] Hardle, Wolfgang, *Applied Multivariate Statistical Analysis*, Version: 29th April 2003 pagina 235
- [H.D, 2001] Hand, David, Heikki Mannila and Padhraic Smyth, *Principles of Data Mining*, The MIT Press. Pagina 50
- [M.J, 2001] Marques de Sá, Joaquim P. *Pattern Recognition, Concepts, Methods and Applications*. Oporto University, Portugal. Pagina 53
- [J.D, 2000] Dalla E. Johnson, *Métodos multivariados Aplicados al Análisis de Datos*.

- [C.H, 2003] Cardenas Henry A, Brewer Carlos A. *KDD ENGINE Herramienta De Software Para El Reconocimiento Estadístico De Patrones Basado En Algoritmos De Agrupamiento*. Tesis de Grado UIS.
- [M.W, 2002] Wendy L. Martínez, Ángel R. Martínez. *Computational Statitisc HandBook with MatLab*. by Chapman & Hall/CRC
- [W.A, 2002] Webb, Andrew R, Malvern UK. *Statistical Pattern Recognition*. John Wiley & Sons Ltd,
- [F.K, 1999] Fukugana Keinosuke, *Introduction to Statistical Pattern Recognition*. Academic Press
- [R.T2, 2005] Rivera Flórez, Tulia Esther, *Análisis de conglomerados*, Diplomado en Métodos Estadísticos Multivariados, Universidad Industrial de Santander. Septiembre 2005
- [D.R, 1998] Ratan Das. *Determining the Locations of Faults in Distribution Systems*. Doctoral Thesis, University of Saskatchewan Saskatoon, Canada, 1998.
- [M.J, 2003] J. Mora and J. Colomer. *Voltage Sag Characterization and Classication for Diagnosis in Electric Power Quality Domain. Research Report*, Department of Electronics, Computer Science and Automatic Control, Group eXiT, University of Girona, España, 2003.
- [V.L, 2005] Libardo Villamizar y Carlos Quiñones. *Implementación del método de Ratan Das para la localización de fallas en sistemas de distribución de energía eléctrica*. Tesis de grado, Escuela de Ingenierías Eléctrica, Electrónica y Telecomunicaciones, Universidad Industrial de Santander, Colombia, 2005.
- [M.G, 2005] Germán Andrés Morales España y Alvaro Gómez Ruiz. *Estudio e implementación de una herramienta basada en máquinas de soporte vectorial aplicada a la localización de fallas en sistemas de distribución*. Tesis de grado, Escuela de Ingenierías Eléctrica Electrónica y Telecomunicaciones, Universidad Industrial de Santander Bucaramanga 2005
- [MIT, 2003] Reconocimiento automático del habla, Cuantización vectorial y agrupamiento, Clase n° 6 Año académico 2003, MIT.
- [G.J, 1996] John J.Grainger y William D. Stevenson. *Análisis de sistemas de potencia*. Mc Graw Hill. Cap 10 y 12.

- [C.G, 1990] Gilberto Carrillo Caicedo. *Fundamentos de Protecciones*. Bucaramanga, Abril de 1990.
- [A. A, 1999] Alvaro Acosta M. *Análisis de sistemas de potencia*. Facultad de ingeniería eléctrica. Universidad Tecnológica de Colombia. Capítulo 4.
- [B.S] Sourav Bhattacharya and P. Barat. Abstracts of Manuscripts in Review EDICS Number 2-LFLT *Linear Filtering and Enhancement*. www.ieee.org/xplore.
- [C.G, 2004] Gilmar Caiado Fleury Medeiros, Miguel G. Lizárraga, Luan Ling Lee, School of Electrical and Computer Engineering, State University of Campinas, Campinas, SP, Brazil. Este artículo aparece en Revista IEEE América Latina, Publication March 2004 Volume: 2, Issue: 1 ISSN: 1548-0992.
- [W.S, 2004] Slawo Wesolkowski, M.A.Sc. School of Engineering University of Guelph Guelph, ON email: kinsner@ee.umanitoba.ca www.ieee.org/xplore
- [M.M ,2003] Mike Meade, Shyamala C. Sivakumar, and William J. Phillips. *Comparative performance of principal component analysis, Gabor wavelets and discrete wavelet transforms for face recognition*. www.ieee.org/xplore.
- [A.K, 2004] Krishna K. Anaparthi, Student Member, IEEE, Balarko Chaudhuri, Student Member, IEEE, Nina F. Thornhill, and Bikash C. Pal, Senior Member, IEEE. *Coherency Identification in Power Systems through Principal Component Analysis*. www.ieee.org/xplore.
- [C.W, 2003] Wu Chou. *Pattern Recognition in Speech and Language Processing*. Avaya Labs Research. CRC PRESS. 2003
- [G.S, 2002] Sanjay Goil and Alok Choudhary, *High Performance Multidimensional Analysis and Data Mining*. Department of Electrical Computer Engineering Northwestern University. 2002
- [S.B, 2002] Sing-Tze Bow, *Pattern Recognition and Image Preprocessing*. Northern Illinois University. Marcel Dekker, Inc. 2002
- [H.J, 2000] Jiawei Han and Micheline Kamber, *Data Mining: Concepts and Techniques*, Simon Fraser University. Morgan Kaufmann Publishers. 2000.
- [T.K, 2005] Kurt Thearling, Ph.D. *An Introduction to Data Mining*. www.thearling.com <Graphics file: C:/lyx/bin/worklix/reconocimiento_1/graficas/capitulo4/validacion1.png>