

TABLA 2.6: Datos Transformadores 2 devanados sistema Colombiano

			V1	V2	Z(1)	R(1)	Z(0)	R(0)	S
Nodo inicial	Nodo final	Grupo	kV	kV	%	%	%	%	MVA
ALAN220	ANCHG1	YNd1	230	13,8	12,8	0,32	12,8	0,32	115
ALAN220	ANCHG2	YNd1	230	13,8	12,8	0,32	12,8	0,32	115
ALAN220	ANCHG3	YNd1	230	13,8	12,8	0,32	12,8	0,32	115
ANCOR110	ANCO46.6	YNyn0	110	46,6	7,66	0,23	7,66	0,23	60
APAR110	APAR44	YNd1	110	44	7,16	0,25	7,16	0,25	40

- *Modelamiento transformadores 3 devanados.* Se cuenta con información de 130 transformadores tridevanados. Para cada elemento hay datos de tensión de cada devanado, potencia de cada devanado, impedancia y resistencia en secuencias desde un devanado a otro, tal como se muestra en las tablas 2.7 y 2.8.

TABLA 2.7: Datos transformadores 3 devanados

Devanado				V1	V2	V3	Sr12	Sr23	Sr31
Primario	Secundario	Terciario	Grupo	kV	kV	kV	MVA	MVA	MVA
ALTA220	ALTA115	ALTA13	YNyn0d5	220	115	13,8	90	90	90
ANCE5	ANCO110	ANCO46.6	YNyn0d5	220	110	46,6	180	180	180
ANCEA	ANCO110	ANCO14	YNyn0d5	220	110	46,6	180	180	180
BAC5	BAC115	BAC34.5	YNyn0d5	500	115	34,5	450	450	450
BAC5	BAC2	BAC.34.	YNyn0d5	500	230	34,5	450	450	450

TABLA 2.8: Datos transformadores 3 devanados (continuación)

Z12(1)	Z23(1)	Z31(1)	R12(1)	R23(1)	R31(1)	Z12(0)	Z23(0)	Z31(0)
%	%	%	%	%	%	%	%	%
9	19,2	28,8	0,24	0,52	0,77	9	19,2	28,8
8,8	26,8	36,8	0,2	0,71	0,97	8,8	26,8	36,8
9,31	27,45	39,18	0,21	0,62	0,88	9,31	27,45	39,18
12,29	7,39	19,56	0,29	0,17	0,46	12,29	7,39	19,56
11,53	16,64	31,06	0,25	0,35	0,66	11,53	16,64	31,06

Para los transformadores tridevanados es importante tener en cuenta que los valores de Z y R dados en las tablas anteriores, son valores medidos como se muestra en la figura 2.14.

De acuerdo con el modelo, R_{12} es la impedancia que hay entre los devanados primario y secundario. Teniendo los valores R_{12} , R_{23} y R_{13} , se calculan los valores de R_1 , R_2 y R_3 de la siguiente manera:

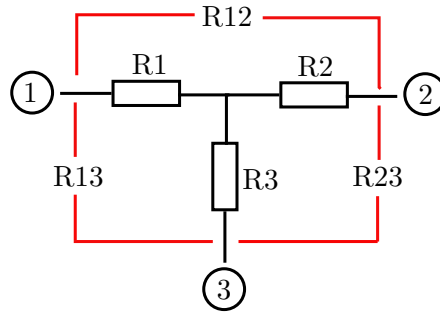


FIGURA 2.14: Datos de Resistencia entre los devanados. Transformador terciario.

$$\begin{aligned}
 R_1 &= \frac{R_{12} + R_{13} - R_{23}}{2} \\
 R_2 &= \frac{R_{12} + R_{23} - R_{13}}{2} \\
 R_3 &= \frac{R_{13} + R_{23} - R_{12}}{2}
 \end{aligned} \tag{2.33}$$

Se realiza el mismo procedimiento para hallar X_1 , X_2 y X_3 .

Posteriormente, con el ánimo de no generar un nodo ficticio en el centro del transformador tridevanado, se realiza una conversión estrella a delta teniendo como resultado lo siguiente:

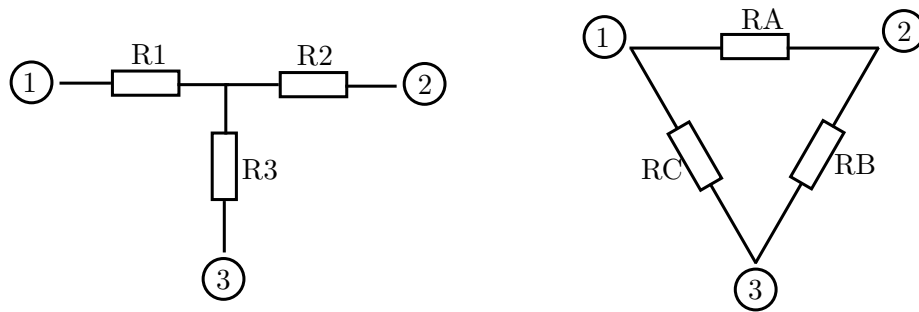


FIGURA 2.15: Modelo transformador tridevanado en estrella y delta

Donde:

$$\begin{aligned}
 R_A &= \frac{(R_1 * R_2) + (R_1 * R_3) + (R_2 * R_3)}{R_3} \\
 R_B &= \frac{(R_1 * R_2) + (R_1 * R_3) + (R_2 * R_3)}{R_1} \\
 R_C &= \frac{(R_1 * R_2) + (R_1 * R_3) + (R_2 * R_3)}{R_2}
 \end{aligned} \tag{2.34}$$

Para la secuencia cero, las conexiones entre elementos R_A , R_B y R_C del transformador se alteran de acuerdo con el grupo de conexión de los devanados.

- *Modelamiento Cargas.* En primer lugar, sólo algunos de los nodos del sistema tienen cargas conectadas y son identificados como PQcarga. La información que se tiene para el modelamiento es el valor de la potencia consumida por todas las cargas en un determinado tiempo, el factor de distribución de la potencia entre los nodos de carga y el factor de potencia de cada uno de los nodos.

TABLA 2.9: Datos de cargas del Sistema Colombiano

Nodo	Factor Distribución	Factor de Potencia	Tipo nodo
AGUABLANCA115	0,00760	0,91980	PQcarga
ALAN220	0,00000	1,00000	PQpaso
ALTAMIRA115	0,00340	0,94870	PQcarga
ALTAMIRA220	0,00000	1,00000	PQpaso
AMAGA1	0,00290	0,95310	PQcarga
AMAGA2	0,00000	1,00000	PQpaso
ANCONEPM220	0,00000	1,00000	PQpaso
ANCONISA220	0,00000	1,00000	PQpaso
ANCON_SUR110	0,00360	0,90000	PQcarga
ANCON_SUR46.6	0,00450	0,95960	PQcarga

El valor de potencia consumida por las cargas de todo el sistema es 8762MW, tomado de [UPME. 2006]. Al multiplicar este valor por los respectivos factores de distribución de cada nodo, se puede hallar el valor de potencia activa consumida por nodo. Posteriormente usando el valor de potencia activa hallado y el factor de potencia, se halla la potencia reactiva en cada nodo.

Con las potencias activa y reactiva calculadas se hallan los respectivos valores de R y XL o G y B para agregar al modelo general del sistema.

- *Modelamiento Generadores.* El sistema colombiano cuenta con 180 generadores conectados. Se cuenta con información para cada unidad de generación sobre potencia y tensión nominales, valores de puesta a tierra, impedancias transitorias y subtransitorias con las cuales se obtienen los valores de impedancia para el modelo de secuencias.

TABLA 2.10: Datos de Generadores del sistema Colombiano

Nodo	S nominal MVA	V nominal kV	R tierra Ohm	X tierra Ohm
A.ANCHICAYA_G1	126	13,8	26,66	0
A.ANCHICAYA_G2	126	13,8	26,66	0
A.ANCHICAYA_G3	126	13,8	26,66	0
AYURA_G1	19,1	6,9	0	0
B.ANCHICAYA_G1	25	6,9	28	0

TABLA 2.11: Datos de Generadores del sistema Colombiano (continuación)

x_d	x_d'	x_d''	x_q	x_q'	x_q''
%	%	%	%	%	%
107	25	18	63	63	24
107	25	18	63	63	24
107	25	18	63	63	24
123	32,1	18,7	75	75	23,8
143,96	32,14	24,37	91,06	91,06	29,19

Con los datos de las tablas 2.10 y 2.11, y usando las ecuaciones expuestas en secciones anteriores se obtienen los parámetros necesarios para completar el modelo de secuencias.

Debido a los protocolos de etiquetación, convenciones técnicas y generalizaciones en la información del modelo colombiano, los elementos del sistema tienen nombres distintos para identificar un mismo punto. Por lo tanto, en el trabajo de modelamiento se debe determinar la correcta conexión entre los elementos del sistema. Dichas conexiones fueron determinadas en gran parte en [Gallego L. 2008] y posteriormente fueron revisadas e implementadas en esta investigación. En seguida se asigna una identificación numérica para cada nodo, con el fin de identificarlos claramente dentro de las matrices de secuencia.

Posteriormente se procede a organizar la información como se indicó en la tabla 2.1. Es importante resaltar que para la secuencia (0) el modelo de los transformadores varía según la conexión de los devanados, por lo tanto, para organizar estas conexiones en tramos serie y shunt es necesario realizar un análisis exhaustivo de dicha información. Una vez organizada la información aplicamos el algoritmo (AGM) obteniendo las matrices de secuencia positiva, negativa y cero de (756x756) que caracterizan el sistema colombiano.

2.8.2. Construcción del unifilar y asignación de coordenadas

En esta sección se propone la construcción de un diagrama unifilar del sistema colombiano con información de posición (x,y) de cada una de las barras, con la intención de observar de manera gráfica:

- *Zonas Perturbadoras:* Zonas del sistema donde ocurren fallas que generan hundimientos en tensión en los nodos de interés.
- *Zonas Críticas:* Zonas donde ocurren hundimientos en tensión con alto nivel de severidad.

Esta construcción se realiza a través de la consulta de [UPME. 2006], ubicando cada uno de los elementos modelados en un diagrama general y etiquetando cada uno de los nodos. A falta de un software para realizar un diagrama de 756 nodos, este se implementó en el software Openoffice Draw, y para poder dibujar contornos de manera sistemática, el

diagrama se exportó, posteriormente, a texto plano en código SVG.²

En el diagrama se localizan las coordenadas de cada una de las barras del sistema, las cuales según el formato y el tamaño de la hoja, corresponden a un valor dado en pixeles. La visualización del diagrama es orientativa de las conexiones entre elementos y no es geográfica, por lo tanto, los resultados en secciones posteriores mostrarán las condiciones de calidad de potencia de las barras, pero dichos resultados no tienen significado de distancia geográfica entre dichas barras.

El modelo desarrollado del sistema colombiano será útil para hallar zonas en las cuales ocurren fallas que pueden afectar las tensiones de los nodos de Bogotá. También, dicho modelamiento servirá para posteriores proyectos de investigación a mayor escala. En la figura 2.16 se puede observar el diagrama unifilar del sistema colombiano, en el cual cada uno de los nodos se encuentra etiquetado con el número que lo representa en las matrices de secuencia.

²SVG es un formato para representar gráficas en texto plano.

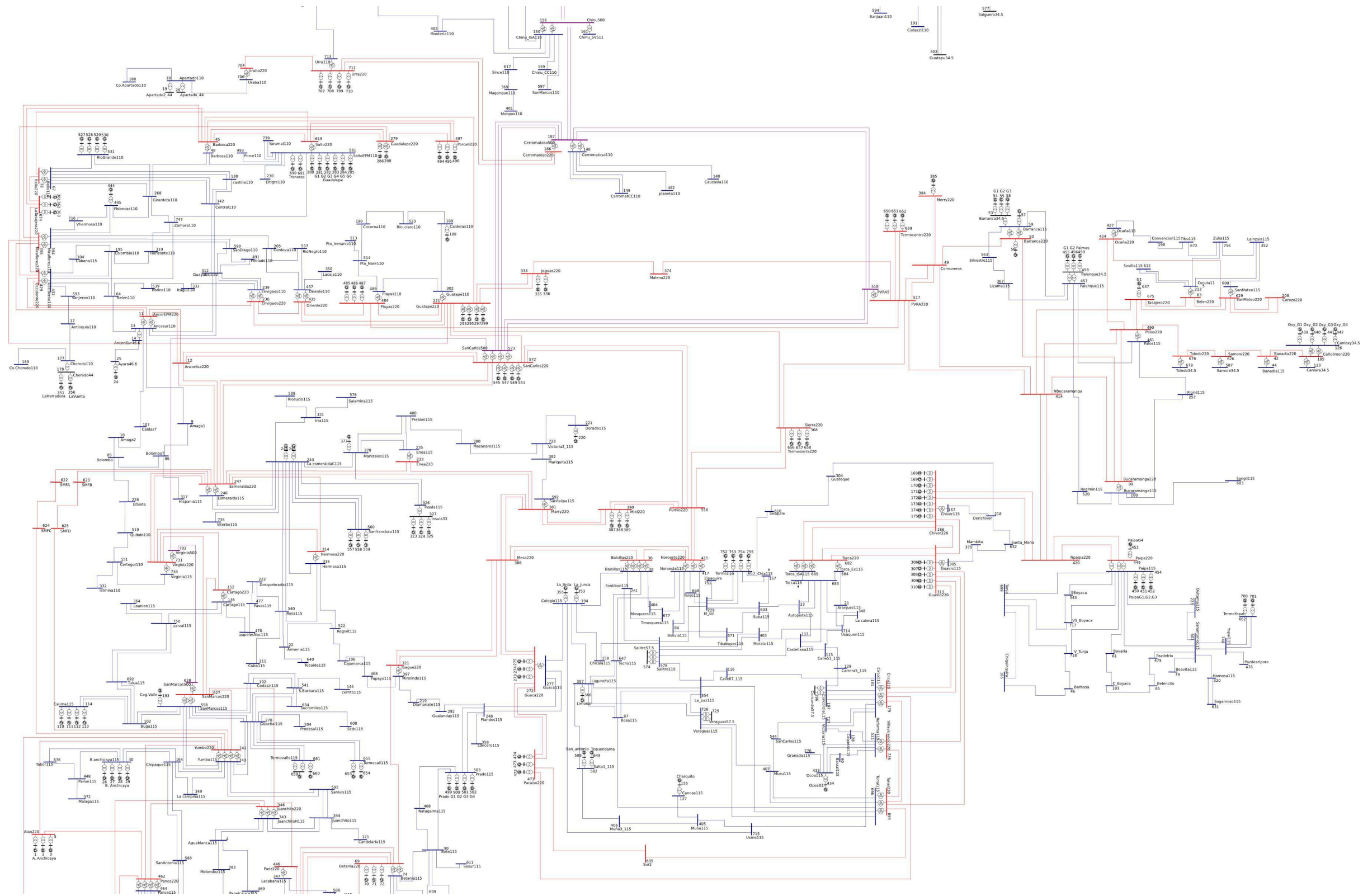


FIGURA 2.16: Diagrama unifilar del sistema Colombiano modelado (Parte).

Capítulo 3

Simulación Estocástica de fallas y análisis de hundimientos

En este capítulo se explican detalladamente los siguientes tres pasos de la metodología de análisis de hundimientos correspondientes a la identificación de zonas perturbadoras y zonas de interés, simulación de fallas en el sistema y por último el análisis estadístico de las simulaciones.

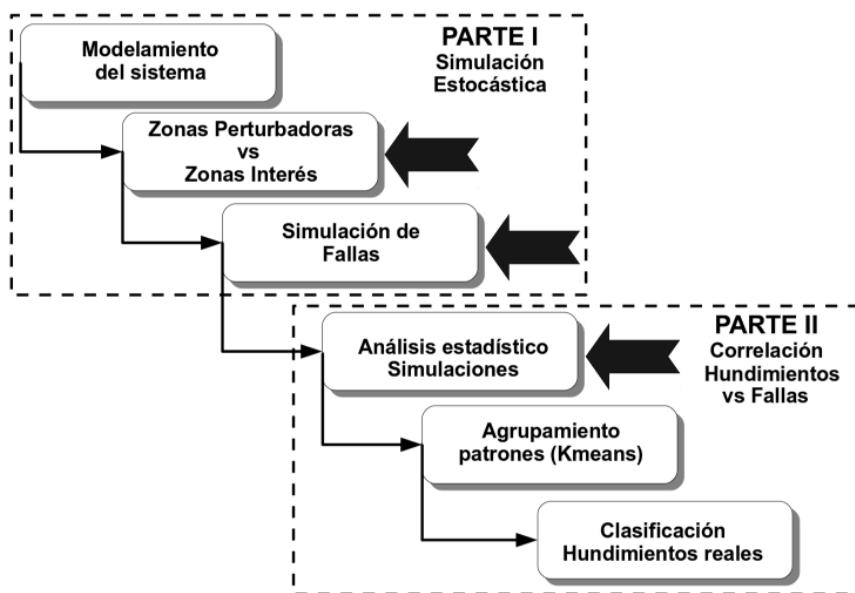


FIGURA 3.1: Zonas perturbadoras, simulación y análisis de hundimientos en la metodología propuesta

La evaluación del impacto de las fallas en los hundimientos de tensión observados en las barras del sistema puede realizarse por medio de dos metodologías, monitoreo de eventos

(fallas y hundimientos) o simulación estocástica de fallas.

En cuanto al monitoreo de eventos es necesario tener información de los hundimientos en tensión que ocurren en todas las barras objeto del análisis, e información de las fallas ocurridas en el sistema. En general el procesamiento de esta información puede presentar dificultades debido a los siguientes aspectos.

- El proceso para el registro de información de fallas es en gran parte manual, lo que permite impresiones por parte del operario al determinar el tipo de falla, la causa de dicha falla y la hora de ésta.¹ Esto conlleva a dificultades para correlacionar información, ya que el registro de hundimientos en tensión y el de fallas en el sistema deberían estar adecuadamente sincronizados.
- Existen equipos de protección y control que registran la ocurrencia de fallas, sin embargo gran parte de éstos no están integrados a un sistema de gestión información, por lo que dicha información en general se pierde o en el mejor de los casos se obtiene de manera manual con dificultad.
- Gran cantidad de fallas transitorias no se detectan o reportan ya que, debido a su rapidez, no ocasionaron ninguna maniobra de algún equipo, disparo o recierre de protecciones, apertura de circuitos, o daño de elementos en el sistema.

Por otro lado, la simulación estocástica se usa para determinar las características de los hundimientos en tensión de cada una de las subestaciones. En este procedimiento, las características de las fallas como lugar de ocurrencia, tipo de falla, frecuencia de ocurrencia, se modelan como distribuciones de probabilidad y posteriormente se simulan sobre el sistema [Romero M. 2008]. La simulación arroja como resultado las distribuciones de probabilidad de magnitud y frecuencia de los hundimientos en tensión en cada una de las barras del sistema.

La interpretación de los resultados de la simulación estocástica puede ser diversa. En caso de usar información real para modelar las características de las fallas (obviamente con las dificultades mencionadas en párrafos anteriores), el resultado será el valor esperado de la cantidad y las características de los hundimientos en tensión que ocurren realmente en las barras del sistema debido a fallas. Generalmente cuando se considera que las tendencias de fallas permanecen invariantes, los resultados se interpretan como una predicción de la cantidad y características de los hundimientos en tensión. Es importante resaltar que las tendencias halladas de hundimientos dependen fuertemente de la información estadística de las fallas, además, el grado de exactitud de la predicción va ligado a la cantidad de tiempo de registro de dichas fallas tal como lo indica la tabla 1 en la introducción de este documento.

¹Por ejemplo, muchas de las fallas con causa incierta son atribuidas a descargas atmosféricas y la hora registrada es deducida o asignada por el operario que realiza el reporte.

En caso de no contar con información real o confiable, las características de las fallas se pueden modelar con distribuciones uniformes, de tal manera que en la simulación se realiza un barrido de todo tipo de fallas en todos los lugares posibles. Debido a que no se tiene información de la cantidad de fallas real en cada lugar de ocurrencia, con este análisis no es posible determinar la cantidad real de hundimientos que pueden ocurrir en las barras del sistema, sin embargo, sí permite determinar el valor esperado de la magnitud de los hundimientos en tensión que podría ocurrir debido a fallas. Con esto, se puede saber qué tan expuesta está una barra a experimentar hundimientos por fallas y cómo es el perfil de hundimientos que podrían tener las barras del sistema.

En las siguientes secciones se determinarán las barras en donde se evaluarán los hundimientos en tensión, y la zona perturbadora del sistema colombiano en la cual ocurren fallas que ocasionan hundimientos en dichas barras. Las características como lugar de ocurrencia y tipo de falla se modelarán con distribuciones uniformes, para posteriormente realizar una simulación estocástica generando fallas en modelo de componentes simétricas del sistema obtenido en el capítulo anterior. Esto permitirá, (i) identificar las zonas perturbadoras donde las fallas ocasionan hundimientos en las barras del sistema de Bogotá, (ii) determinar cuáles barras de 220kV y 115kV son más vulnerables a hundimientos en tensión y (iii) determinar las características de hundimientos generados por fallas. Posteriormente los resultados de las simulaciones se utilizarán en el capítulo siguiente para determinar el impacto real de las fallas usando información real de hundimientos en tensión.

3.1. Determinación de zonas perturbadoras

Una zona perturbadora es en la que ocurren fallas que generan hundimientos en tensión en alguna barra de interés. En esta investigación, el interés es conocer en qué zonas del sistema colombiano pueden ocurrir fallas que generen hundimientos en tensión en los nodos del sistema de 115kV y 220kV cercanos o dentro de la ciudad de Bogotá. Después de identificar las zonas perturbadoras, se realiza la simulación estocástica de fallas en dichas zonas y se evalúan los hundimientos generados en los nodos de interés.

Para hallar las zonas perturbadoras, a partir de las matrices de secuencia del sistema se construye otra matriz llamada matriz de hundimientos en tensión. Esta matriz entrega información de las tensiones de todos los nodos cuando ocurren fallas en cada uno de los nodos. Para cada tipo de falla es necesario construir la matriz de hundimientos con base en los cálculos de fallas explicados en el capítulo anterior. Posteriormente se filtra la matriz de hundimientos extrayendo sólo los nodos en los que ocurren fallas que generan hundimientos en los nodos de interés. En este procedimiento la condición de tensión prefalla $V=1$ es válida para todos los casos, ya que la detección de los hundimientos en tensión se hace con la referencia deslizante (ver página 9).

Para construir la matriz de hundimientos en tensión por fallas trifásicas suponemos de

manera ilustrativa que el nodo de interés es el nodo 8 en el sistema de prueba de 9 nodos utilizado anteriormente y descrito en la figura 3.2.

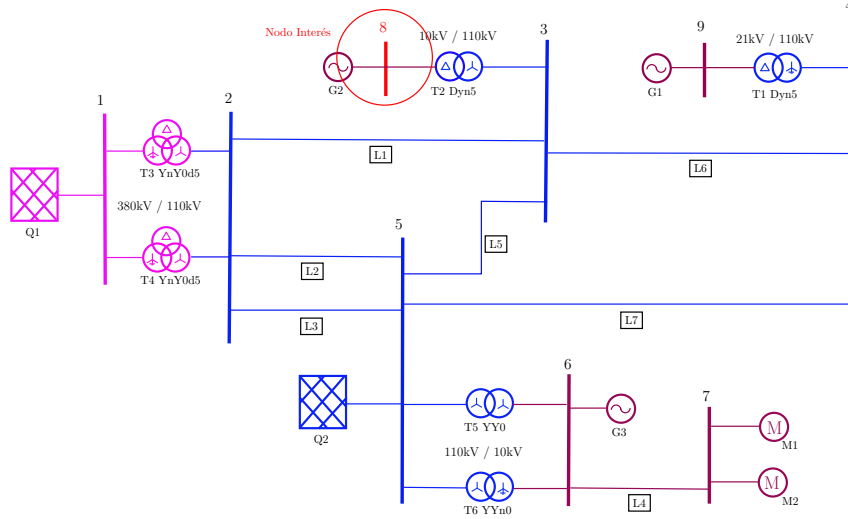


FIGURA 3.2: Nodo de interés para hundimientos en tensión

En primer lugar la tensión en cada uno de los nodos cuando ocurre una falla trifásica en k está dada por la ecuación:

$$V_i = 1 - \frac{Z_{ik}}{Z_{kk}} \quad (3.1)$$

Con base en la ecuación 3.1 se calcula la matriz de hundimientos como sigue:

$$V_{HT} = [1] - Z[DZ]^{-1} \quad (3.2)$$

En la ecuación 3.2 el término $[DZ]^{-1}$ es una matriz calculada a partir de la diagonal de la matriz de impedancia positiva, cuya dimensión es aumentada para ser multiplicada punto a punto con Z . Para calcular $[DZ]$ primero extraemos la diagonal de la matriz Z :

$$diag(Z) = \left\{ \begin{array}{c} 0,0037 \\ 0,0134 \\ 0,0433 \\ 0,0598 \\ 0,0138 \\ 0,0823 \\ 0,7550 \\ 0,0458 \\ 0,0350 \end{array} \right\} + \left\{ \begin{array}{c} 0,0374 \\ 0,1659 \\ 0,2752 \\ 0,3341 \\ 0,1590 \\ 1,6024 \\ 2,3363 \\ 0,7753 \\ 0,5703 \end{array} \right\} i \quad (3.3)$$

Posteriormente, es necesario transponer la diagonal y aumentar la dimensión de ésta, para ésto, hay que tener en cuenta que al transponer un vector complejo, la transpuesta también es la conjugada por lo que es necesario, conjugar nuevamente el resultado tal como se muestra en la ecuación 3.4.

$$DZ = meshgrid \frac{(abs(diag(Z)') .* abs(diag(Z)'))}{diag(Z)'} \quad (3.4)$$

El término *meshgrid* aumenta la dimensión del vector obteniendo una matriz cuadrada como se muestra a continuación.

$$DZ = \left\{ \begin{array}{cccccccc} 0,0037 & 0,0134 & 0,0433 & 0,0598 & 0,0138 & 0,0823 & 0,7550 & 0,0458 & 0,0350 \\ 0,0037 & 0,0134 & 0,0433 & 0,0598 & 0,0138 & 0,0823 & 0,7550 & 0,0458 & 0,0350 \\ 0,0037 & 0,0134 & 0,0433 & 0,0598 & 0,0138 & 0,0823 & 0,7550 & 0,0458 & 0,0350 \\ 0,0037 & 0,0134 & 0,0433 & 0,0598 & 0,0138 & 0,0823 & 0,7550 & 0,0458 & 0,0350 \\ 0,0037 & 0,0134 & 0,0433 & 0,0598 & 0,0138 & 0,0823 & 0,7550 & 0,0458 & 0,0350 \\ 0,0037 & 0,0134 & 0,0433 & 0,0598 & 0,0138 & 0,0823 & 0,7550 & 0,0458 & 0,0350 \\ 0,0037 & 0,0134 & 0,0433 & 0,0598 & 0,0138 & 0,0823 & 0,7550 & 0,0458 & 0,0350 \\ 0,0037 & 0,0134 & 0,0433 & 0,0598 & 0,0138 & 0,0823 & 0,7550 & 0,0458 & 0,0350 \end{array} \right\} + \quad (3.5)$$

$$\left\{ \begin{array}{cccccccc} 0,0374 & 0,1659 & 0,2752 & 0,3341 & 0,1590 & 1,6024 & 2,3363 & 0,7753 & 0,5703 \\ 0,0374 & 0,1659 & 0,2752 & 0,3341 & 0,1590 & 1,6024 & 2,3363 & 0,7753 & 0,5703 \\ 0,0374 & 0,1659 & 0,2752 & 0,3341 & 0,1590 & 1,6024 & 2,3363 & 0,7753 & 0,5703 \\ 0,0374 & 0,1659 & 0,2752 & 0,3341 & 0,1590 & 1,6024 & 2,3363 & 0,7753 & 0,5703 \\ 0,0374 & 0,1659 & 0,2752 & 0,3341 & 0,1590 & 1,6024 & 2,3363 & 0,7753 & 0,5703 \\ 0,0374 & 0,1659 & 0,2752 & 0,3341 & 0,1590 & 1,6024 & 2,3363 & 0,7753 & 0,5703 \\ 0,0374 & 0,1659 & 0,2752 & 0,3341 & 0,1590 & 1,6024 & 2,3363 & 0,7753 & 0,5703 \\ 0,0374 & 0,1659 & 0,2752 & 0,3341 & 0,1590 & 1,6024 & 2,3363 & 0,7753 & 0,5703 \end{array} \right\} i \quad (3.6)$$

Al aplicar la ecuación 3.2 obtenemos que para fallas trifásicas la matriz de hundimientos es la siguiente:

$$V_{HT} = \left\{ \begin{array}{cccccccc} 0,0000 & 0,8828 & \mathbf{0,9469} & 0,9600 & 0,9046 & 0,9925 & 0,9954 & 0,9891 & 0,9888 \\ 0,4807 & 0,0000 & \mathbf{0,5551} & 0,6656 & 0,1944 & 0,9363 & 0,9618 & 0,9074 & 0,9053 \\ 0,6059 & 0,2485 & \mathbf{0,0000} & 0,4085 & 0,1491 & 0,9329 & 0,9595 & 0,7962 & 0,8353 \\ 0,6367 & 0,3092 & \mathbf{0,2780} & 0,0000 & 0,1727 & 0,9346 & 0,9607 & 0,8516 & 0,7256 \\ 0,5948 & 0,2258 & \mathbf{0,5156} & 0,6145 & 0,0000 & 0,9215 & 0,9522 & 0,8997 & 0,8915 \\ 0,6799 & 0,3867 & \mathbf{0,6162} & 0,6945 & 0,2099 & 0,0000 & 0,4362 & 0,9207 & 0,9143 \\ 0,6918 & 0,4119 & \mathbf{0,6324} & 0,7077 & 0,2404 & 0,0471 & 0,0000 & 0,9237 & 0,9175 \\ \mathbf{0,7747} & \mathbf{0,5687} & \mathbf{0,4283} & 0,6610 & 0,5122 & 0,9616 & \mathbf{0,9768} & \mathbf{0,0000} & \mathbf{0,9059} \\ 0,8303 & 0,6755 & \mathbf{0,6619} & 0,5330 & 0,6120 & 0,9695 & 0,9816 & 0,9307 & 0,0000 \end{array} \right\} \quad (3.7)$$

En la matriz de la ecuación 3.7 cada columna se interpreta como el nodo donde ocurre la falla trifásica, y cada una de las filas son las tensiones de los nodos durante la ocurrencia de dicha falla. De esta manera, al ocurrir por ejemplo una falla trifásica en el nodo 3, las tensiones en todos los nodos serán las marcadas con letras en rojo. Por otro lado, para identificar cuáles fallas ocasionaron hundimientos en el nodo de interés (8), observamos la correspondiente fila (marcada en celda gris) y verificamos cuáles de los valores son hundimientos. De aquí podemos identificar que la ocurrencia de fallas en los nodos 1,2,3,4,5 y 8 generan hundimientos en tensión en el nodo 8.

Los nodos hallados hacen parte de la zona perturbadora para hundimientos en el nodo 8. El resultado se muestra gráficamente en la figura 3.3.

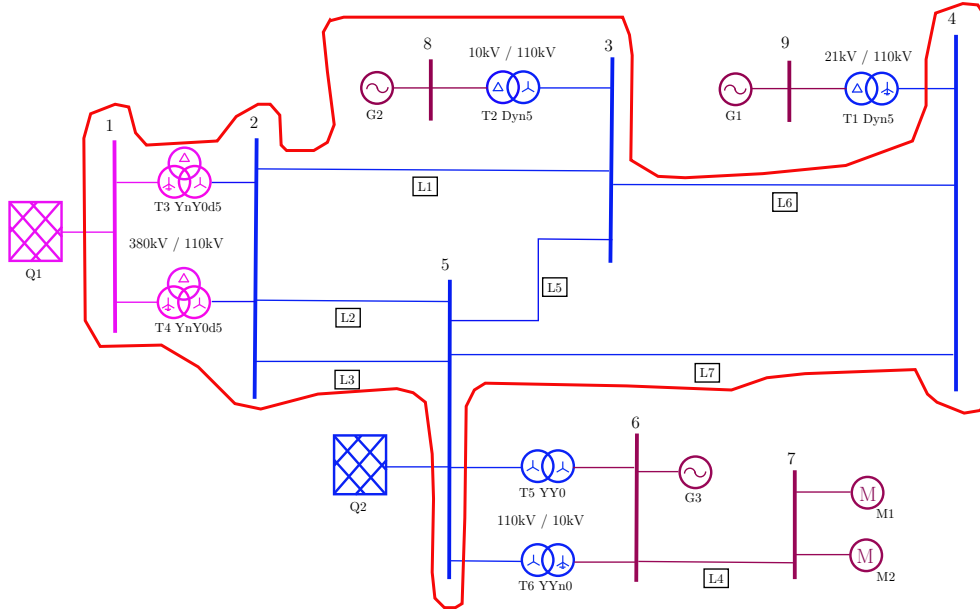


FIGURA 3.3: Zona perturbadora para el nodo 8

Para generar zonas perturbadoras en el sistema colombiano, se implementa el procedimiento anterior en el siguiente algoritmo:

1. Se calcula la matriz de hundimientos en tensión V_{HT} de acuerdo con la ecuación 3.2.
2. De la matriz V_{HT} se extraen los nodos de interés.
3. Se buscan las columnas donde los valores de tensión en los nodos de interés son menores a 0.9 pu.
4. Con los nodos encontrados se genera una grilla de datos para generar contornos.
5. Con la ubicación X,Y de las barras del sistema colombiano y los nodos encontrados, se genera el contorno de la zona perturbadora.

Debido a que los demás tipos de fallas son asimétricas (monofásica, bifásica y bifásica a tierra), la matriz de hundimientos en tensión se calcula a partir de las matrices de componentes simétricas. El resultado de esto son tres matrices de tensiones de secuencias (+, - y 0), de las cuales se obtienen las tensiones V_a , V_b y V_c de cada uno de los nodos. La matriz de hundimientos final para cada uno de los tipos de falla, está compuesta por el menor valor de tensión de las tres fases para cada uno de los nodos.

Para el caso de fallas monofásicas tenemos que la matriz $[DZ]$ y las matrices de secuencias se calculan según las ecuaciones 3.8 y 3.9.

$$DZ = meshgrid \frac{(abs(diag(Z + Z_{(-)} + Z_{(0)}))' . * abs(diag(Z + Z_{(-)} + Z_{(0)}))')}{diag(Z + Z_{(-)} + Z_{(0)})'} \quad (3.8)$$

$$\begin{aligned} V_{(+)} &= [1] - Z[DZ]^{-1} \\ V_{(-)} &= -Z_{(-)}[DZ]^{-1} \\ V_{(0)} &= -Z_{(0)}[DZ]^{-1} \end{aligned} \quad (3.9)$$

Posteriormente, se calculan las tensiones de las fases y se halla la matriz de hundimientos según las ecuaciones 3.10 y 3.11.

$$\begin{aligned} V_a &= V_{(0)} + V_{(+)} + V_{(-)} \\ V_b &= V_{(0)} + V_{(+)} * (1\angle - 120^\circ) + V_{(-)} * (1\angle 120^\circ) \\ V_c &= V_{(0)} + V_{(+)} * (1\angle 120^\circ) + V_{(-)} * (1\angle - 120^\circ) \end{aligned} \quad (3.10)$$

$$V_{HT} = \min(V_a, V_b, V_c) \quad (3.11)$$

Para el caso de fallas bifásicas la matriz $[DZ]$ y las matrices de secuencias son calculadas según las ecuaciones 3.12 y 3.13.

$$DZ = meshgrid \frac{(abs(diag(Z + Z_{(-)}))' . * abs(diag(Z + Z_{(-)}))')}{diag(Z + Z_{(-)})'} \quad (3.12)$$

$$\begin{aligned} V_{(+)} &= [1] - Z[DZ]^{-1} \\ V_{(-)} &= -Z_{(-)}[-DZ]^{-1} \\ V_{(0)} &= 0 \end{aligned} \quad (3.13)$$

Y finalmente para el caso de fallas bifásicas a tierra tenemos:

$$DZ = meshgrid \frac{(diag(Z) + ((diag(Z_{(-)}) . * diag(Z_{(0)})))}{(diag(Z_{(0)}) + diag(Z_{(-)}))} \quad (3.14)$$

$$\begin{aligned} V_{(+)} &= [1] - Z * [DZ]^{-1} \\ V_{(-)} &= -Z_{(-)}(-[DZ]^{-1}) * meshgrid \frac{diag(Z)}{(diag(Z_{(0)}) + diag(Z_{(-)}))} \\ V_{(0)} &= -Z_{(0)}(-[DZ]^{-1}) * meshgrid \frac{diag(Z_{(-)})}{(diag(Z_{(0)}) + diag(Z_{(-)}))} \end{aligned} \quad (3.15)$$

Para estos dos últimos casos de fallas la matriz de hundimientos se calcula de igual manera que el caso monofásico a través de las ecuaciones 3.10 y 3.11.

El procedimiento descrito, es implementado para cada tipo de falla sobre el sistema colombiano, con el fin de encontrar las zonas perturbadoras del sistema en las cuales ocurren fallas que afectan los nodos de interés para la ciudad de Bogotá. Los nodos de interés dentro del sistema colombiano para observar los hundimientos en tensión son los siguientes:

- 11 Nodos del anillo 220kV que alimentan al sistema de Bogotá.
- 50 Nodos del sistema de 115kV relacionado con el sistema de Bogotá.
- 3 Nodos de 57,5kV barras modeladas dentro de Bogotá.
- 21 Nodos de 13,8kV barras de generación.
- 1 Barra generación 6,9kV.

Los Nombres de estos nodos de interés y su correspondiente nivel de tensión pueden verse en la tabla 3.1 y las zonas para cada tipo de fallas se muestran en las figuras 3.4, 3.5 y 3.6.

TABLA 3.1: Barras de interés para detección zonas perturbadoras

Id_barra	Barra	Tensión kV	Id_barra	Barra	Tensión kV
21	aranjuez_115	115	404	mosquera115	115
23	autopista115	115	405	muña115	115
36	balsillas220	230	406	muña3_115	115
38	balsillas115	115	407	muzul115	115
60	barzal	115	415	noroeste220	230
84	bolivia115	115	417	noroeste115	115
87	bosa115	115	433	ocoa	115
115	calle51_115	115	434	ocoa_g1	13,8
116	calle67_115	115	471	paraiso220	230
127	canoas115	115	472	paraiso_g1	13,8
128	caqueza115	115	473	paraiso_g2	13,8
129	carrera5_115	115	474	paraiso_g3	13,8
137	castellana115	115	521	reforma115	115
155	charquitog1	13,8	544	s.carlos_eeb115	115
157	chia115	115	574	sal5_57.5	57,5
158	chicala115	115	578	salitre115	115
179	circo220	230	582	salto1_115	115
181	circo115	115	589	san_antoniog1	13,8
194	colegio115	115	633	suba115	115
196	con5_57.5	57,5	635	sur2	230
197	concordia115	115	647	techo115	115
229	el_sol115	115	648	tenjo115	115
261	fontibon115	115	649	tequendamag1	13,8
270	granada	115	663	termozipa115	115
272	guaca220	230	671	tibabuyes115	115
273	guaca_g1	13,8	677	tmosquera115	115
274	guaca_g2	13,8	682	torca220	230
275	guaca_g3	13,8	683	torca_eeb115	115
277	guaca115	115	684	torca_ex115	115
306	guavio_g1	13,8	685	torca_isa115	115
307	guavio_g2	13,8	694	tunal220	230
308	guavio_g3	13,8	696	tunal115	115
309	guavio_g4	13,8	714	usaquen115	115
310	guavio_g5	13,8	715	usme115	115
313	guavio220	230	725	ver5_57.5	57,5
348	la_calera115	115	726	veraguas115	115
353	la_juncag1	13,8	727	victoria1_115	115
354	la_paz115	115	736	villavivencio220	230
355	la_tintag1	13,8	751	zipaquira	115
357	laguneta115	115	752	zipaquira_g2	13,8
366	limonar_g1	6,9	753	zipaquira_g3	13,8
386	mesa220	230	754	zipaquira_g4	13,8
403	morato115	115	755	zipaquira_g5	13,8

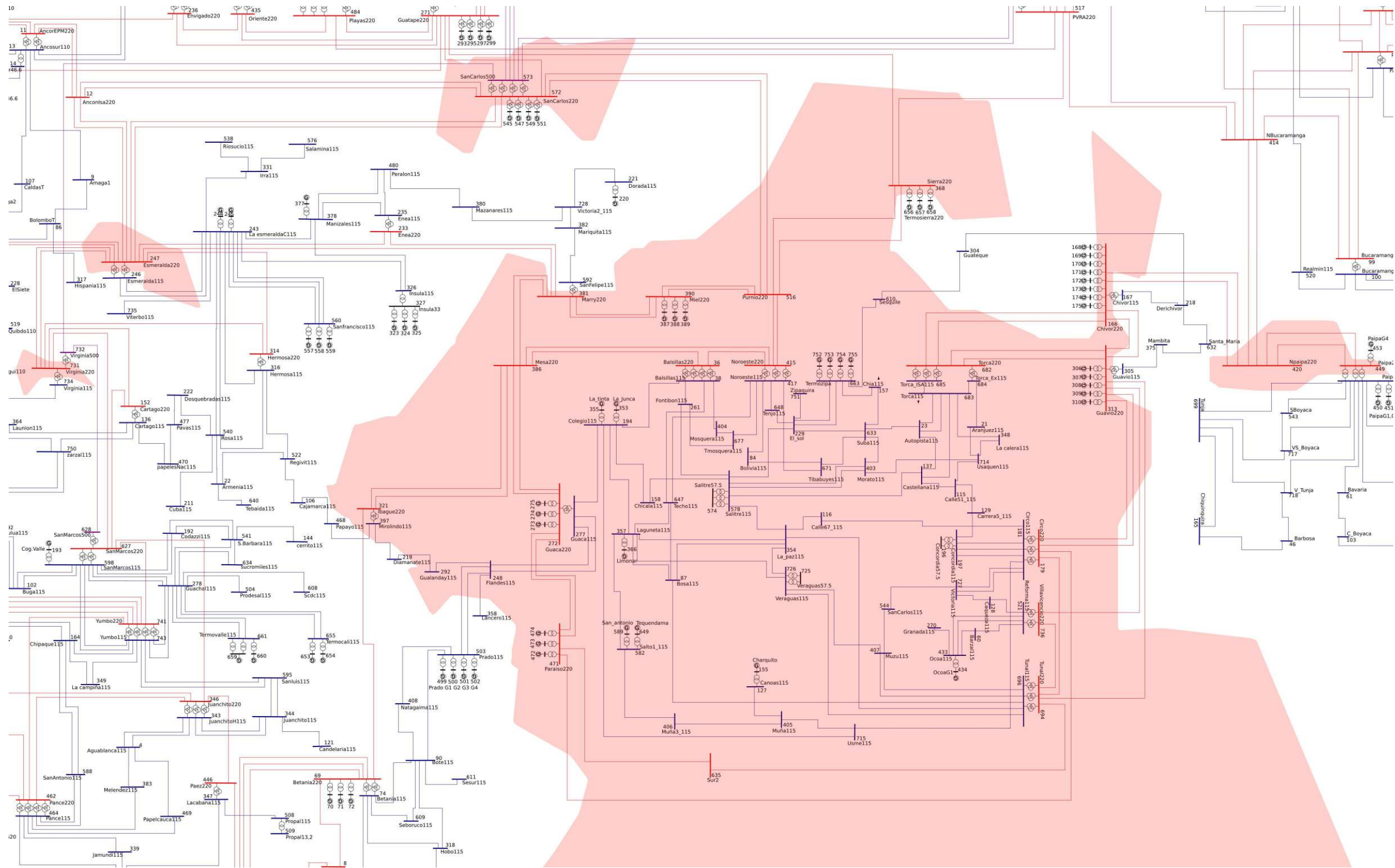


FIGURA 3.4: Zona perturbadora para fallas monofásicas a tierra.

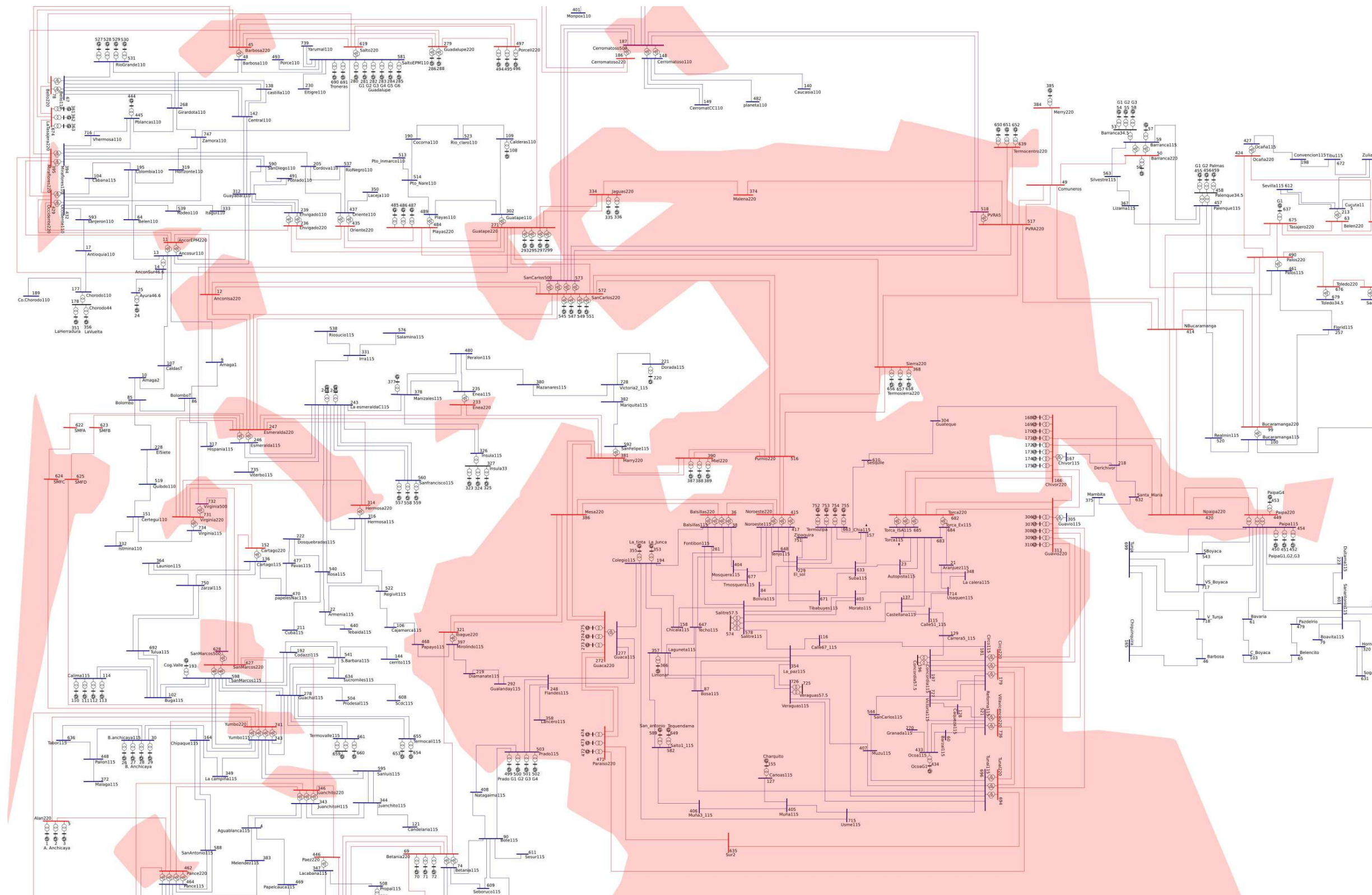


FIGURA 3.5: Zona perturbadora para fallas bifásicas y bifásicas a tierra.

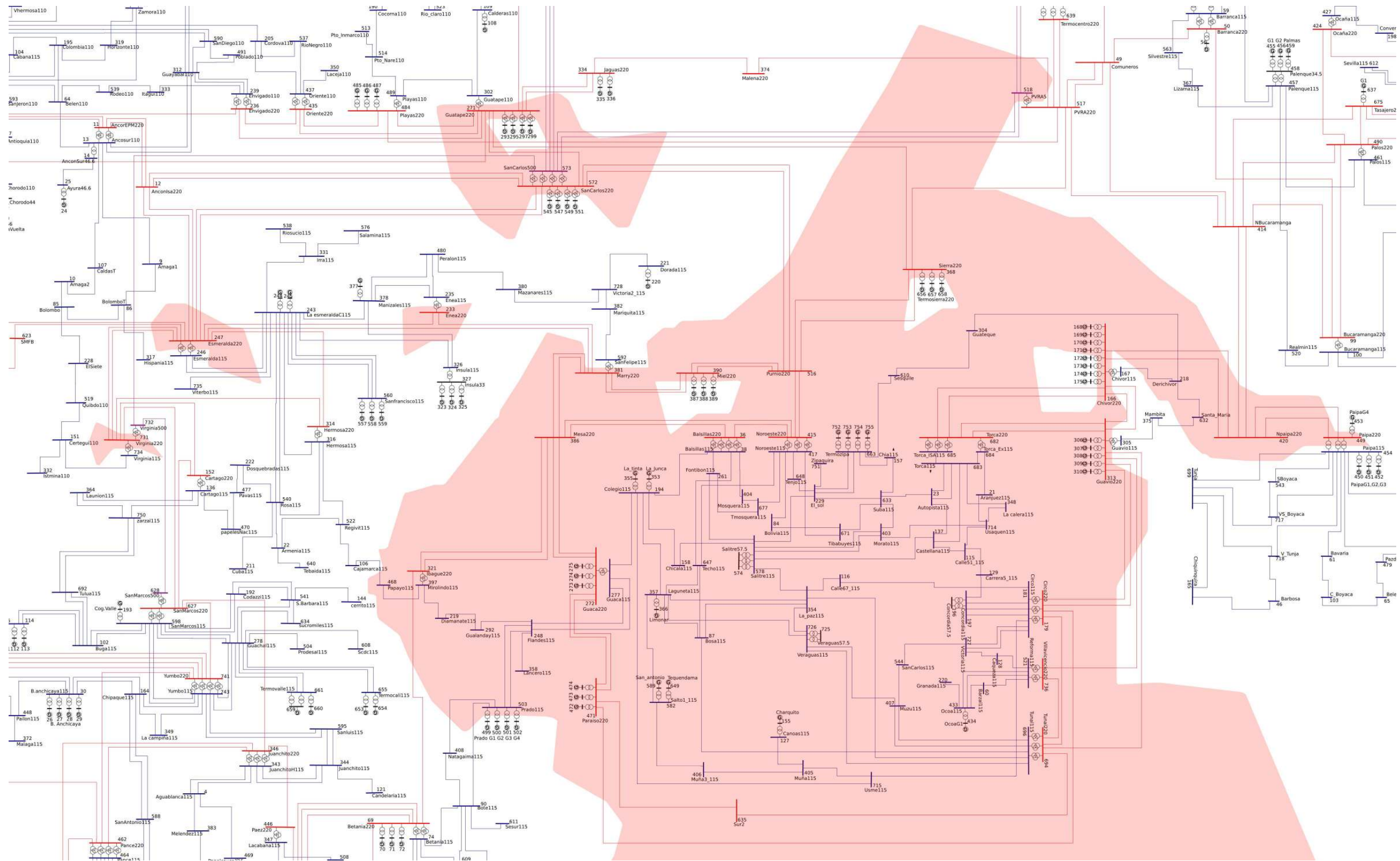


FIGURA 3.6: Zona perturbadora para fallas trifásicas.

3.2. Caracterización de variables aleatorias

Una vez determinadas las zonas perturbadoras para cada tipo de falla, se deben caracterizar los parámetros de ocurrencia de las fallas que se van a simular. Los parámetros establecidos para esta investigación son la ubicación y el tipo de falla. La ubicación de una falla es considerada como una variable aleatoria ya que no es posible saber de manera determinista en qué lugar del sistema ocurrirá una falla. De igual manera, cuando ocurre una falla no es posible saber si ésta será de tipo monofásico, bifásico o trifásico, por lo que el tipo de falla también es considerado una variable aleatoria. Estas variables aleatorias se modelan a partir de información estadística de fallas que ocurren en el sistema de transmisión y distribución. En las siguientes secciones se explica detalladamente como se realiza dicho modelamiento.

3.2.1. Ubicación de la falla

En las redes de distribución constantemente ocurren fallas que ocasionan la operación de las protecciones en los respectivos circuitos y generan hundimientos en tensión en varias barras del sistema, dichas fallas generalmente ocurren más en unos lugares que otros debido a:

- Condiciones atmosféricas distintas en cada zona.
- Niveles de exposición diferentes a la humedad, intemperie, viento y animales.
- Construcción topológica de los circuitos (aéreo o subterráneo).
- Usuarios perturbadores ubicados en ciertos lugares del sistema.

El modelamiento de la tendencia del lugar ocurrencia de las fallas es indispensable ya que esto afecta directamente las características de los hundimientos en tensión vistos en cada uno de los nodos del sistema.

En caso de tener información detallada de cada línea o circuito, el modelo de la ubicación de falla se puede obtener con la distribución de frecuencias normalizada de fallas. Un ejemplo del modelamiento se puede apreciar en [Romero M. 2006] en donde la información de fallas que ocurren en una muestra de 22 circuitos de media tensión en las redes de CODENSA ESP S.A entre 2002 y 2005, es analizada para caracterizar la ubicación de falla en dichos circuitos. Dicha información mostrada en la tabla 3.2, es normalizada dividiendo cada una de las cantidades de falla entre el total de fallas de todos los circuitos.

Al graficar la cantidad de fallas normalizada obtenemos la figura 3.9 donde podemos apreciar, por ejemplo, que es mucho más probable que ocurra una falla en el circuito 5 que en los demás circuitos. Con esta distribución de frecuencias o distribución de probabilidad se tiene el criterio para generar fallas en la simulación estocástica.

TABLA 3.2: Número de Fallas en circuitos de distribución de ejemplo.

ID Circuito	No. Fallas	Fallas normalizadas
1	6	0,01
2	1	0
3	1	0
4	1	0
5	248	0,43
6	4	0,01
7	2	0
8	3	0,01
9	6	0,01
10	1	0
11	3	0,01
12	19	0,03
13	1	0
14	3	0,01
15	3	0,01
16	154	0,27
17	1	0
18	4	0,01
19	98	0,17
20	5	0,01
21	3	0,01
22	8	0,01

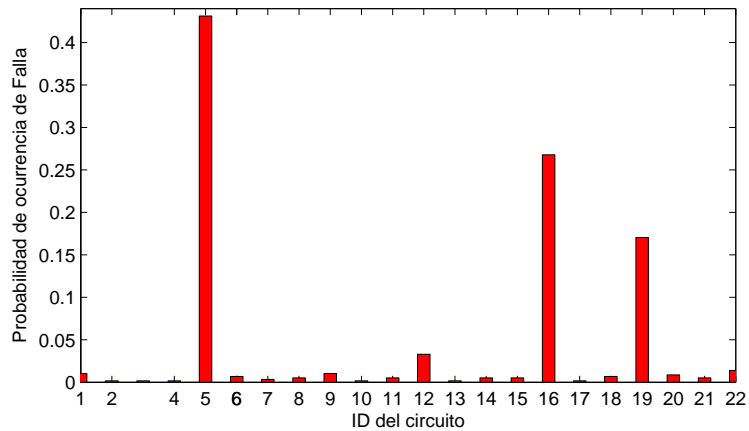


FIGURA 3.7: Distribución de probabilidad Falla en ejemplo de 22 circuitos.

De acuerdo con el nivel de detalle de la información de las fallas, se puede estimar la resolución de la ubicación de éstas. Por ejemplo, en la distribución de la figura 3.7, aunque sabemos en cuáles circuitos ocurren más fallas, no es posible determinar en qué parte del circuito están ocurriendo dichas fallas, por lo que, si no se cuenta con más información, se puede considerar que el circuito puede fallar en cualquier parte con igual probabilidad.

En otras ocasiones la información que se tiene de las fallas ocurridas es muy general, por ejemplo, en muchos casos se tienen para líneas de transmisión la tasa esperada de fallas por kilómetro al año, asumiendo que todas las líneas se comportan estadísticamente igual, se puede establecer las fallas de cada línea según su extensión. En otros casos se pueden tener estadísticas de fallas por niveles de tensión, por ejemplo se espera que en baja tensión ocurran más eventos porque hay usuarios perturbadores y las líneas están expuestas a muchas más eventualidades. En alta tensión se esperan pocas fallas, la mayoría debido a condiciones atmosféricas.

En casos de tener información estadística general es necesario evaluar dicha información con el fin de adjudicar a cada tramo de interés una probabilidad de falla. Un ejemplo de esto se explica en [Romero M. 2006], donde al no tener información detallada de fallas que ocurren en las redes eléctricas de la Universidad Nacional de Colombia (Bogotá), se tomó la información generalizada de tasa de fallas en circuitos para redes de distribución aéreas y subterráneas de Bogotá, mostrada en la tabla 3.3.

	Aérea	Subterránea
Numero fallas	386	36
Longitud (Km)	436,622	77,176
Fallas/Km	0,884	0,466

TABLA 3.3: Tasa de falla según la tipología de línea.

Con la información de la tabla 3.3, la información de la extensión de cada tramo modelado del circuito de la universidad y la topología, se hace la suposición que los tramos se comportan estadísticamente igual según lo indicado en la tabla 3.3, es decir, que los circuitos de la Universidad Nacional se comportan generalmente como lo haría cualquier circuito de la ciudad, por lo que para cada uno de los 33 tramos (aéreos y subterráneos) se puede calcular una tasa de falla cuyo resultado es mostrado en la figura 3.8.

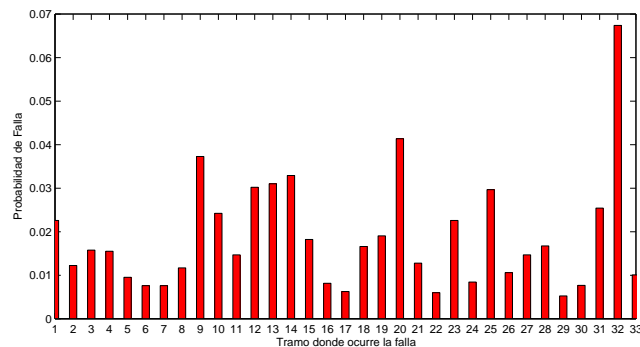


FIGURA 3.8: Probabilidad de Falla en tramos de la Universidad Nacional.

El resultado del modelamiento del lugar de ocurrencia de fallas es una distribución de probabilidad que indica dónde y qué tan frecuente pueden ocurrir fallas en algún sitio específico del sistema. Para modelar la ocurrencia de fallas en la zona perturbadora identificada en secciones anteriores, se necesita información estadística de fallas en líneas del sistema colombiano, que incluyen los niveles de tensión 115kV 220kV y 500kV, tal como se muestra en las gráficas de zonas perturbadoras. Dicha información no estuvo disponible durante la elaboración de este trabajo por lo que en este documento asumiremos una distribución de ocurrencia de falla uniforme, resaltando que la metodología propuesta se puede aplicar a cualquier distribución encontrada posteriormente.

3.2.2. Tipo de falla

Las fallas que ocurren en el sistema pueden ser de tipo monofásico, bifásico, bifásico a tierra o trifásico. El modelamiento del tipo de falla consiste en encontrar la distribución de probabilidad que indica qué tan posible es que la ocurrencia de una falla sea de uno u otro tipo. Para encontrar la distribución que caracteriza el tipo de falla se hace un análisis de acuerdo a la calidad y el detalle de la información de las fallas.

En un primer caso, es posible contar con registros de tipo de falla por parte de los operadores de red, éstos se pueden contabilizar de tal manera que se encuentre el número de eventos por cada tipo de falla y posteriormente normalizar cada uno de los valores de la misma manera que en la sección anterior. El resultado de esto será una distribución discreta de tres o cuatro posibles estados indicando la probabilidad de ocurrencia de cada uno de estos. En general en la literatura se encuentra que en los circuitos de distribución las fallas monofásicas son las más probables, les siguen bifásicas y posteriormente trifásicas.

Es importante tener en cuenta que la distribución del tipo de falla hallada representa el comportamiento de la variable para todos los circuitos modelados. Sin embargo, es posible realizar una representación más detallada del tipo de falla, haciendo que para cada línea se tenga una distribución de probabilidad distinta. En ese caso se pueden realizar simulaciones de montecarlo con variables dependientes (algoritmo Gibbs) explicado en [Romero M. 2006]. Esto implica un trabajo más elaborado por lo que es necesario verificar si el detalle y la cantidad de información, y las diferencias entre distribuciones ameritan la realización de un trabajo tan detallado.

En un segundo caso, la información del tipo de falla no se encuentra reportada directamente. En este caso los operadores de red tienen clasificados los eventos que ocurren en las redes con el fin de llevar estadísticas de operación y mantenimiento entre otros. Por ejemplo, en CODENSA ESP. S.A estos eventos son clasificados en 62 diferentes tipos de causas según [Duarte O. 2005], registrando información puntual del suceso físico pero no del tipo de falla que se generó.

Con la información anterior no es posible inferir de manera directa el tipo de falla ocurrido,

sin embargo, si podemos estimar de manera estocástica qué probabilidad de falla hay para cada una de las causas. Para realizar ésto, en [Romero M. 2006] se propone una clasificación de las causas de las fallas como se indica en la tabla 3.4.

Número	Causa
1	Vandalismo
2	Árboles
3	Falla Equipos
4	Accidentes
5	No Determinada
6	Viento
7	Otros
8	Falla Infraestructura
9	Descargas

TABLA 3.4: Clasificación de las causas de fallas

Una vez que los eventos son clasificados en cada tipo de causa, se debe estimar cuál es el tipo de falla para cada causa, sin embargo, una causa no ocasiona siempre el mismo tipo de fallas, por lo que la asignación del tipo de falla se puede lograr a través de estimación estocástica. Un ejemplo de ésto se puede ver en [Reyes D. 2005] en donde se estimó estocásticamente el tipo de falla para cada causa obteniendo la tabla 3.5.

Causa	Tipo de Falla		
	Monofásica	Bifásica	Trifásica
Descargas	0,79	0	0,21
Viento	0,08	0,9	0,02
Accidentes	0,33	0,33	0,33
Falla Infraestructura	0,85	0,1	0,05
Animales	0,45	0,45	0,1
Árboles	0,83	0,15	0,02
Vandalismo	0,33	0,33	0,33
Fallas Equipos	0,85	0,1	0,05

TABLA 3.5: Probabilidad de ocurrencia de una falla según la causa.

Con la tabla 3.5 se puede estimar de manera estocástica el tipo de falla para cada evento ocurrido en el sistema. De esta manera se tiene completa la información para hallar la tendencia general del tipo de falla en los circuitos evaluados. El resultado de este análisis es la distribución del tipo de falla tal como se muestra en la tabla 3.6 y en la figura 3.9 donde el tipo de falla para los 22 circuitos de baja tensión obtenido de [Romero M. 2006].

TABLA 3.6: Distribución del tipo de falla

TOTAL		
Tipo falla	Número de fallas	Probabilidad de ocurrencia
Monofásica	321	0,634
Bifásica	112	0,221
Trifásica	73	0,144
TOTAL	506	1,000

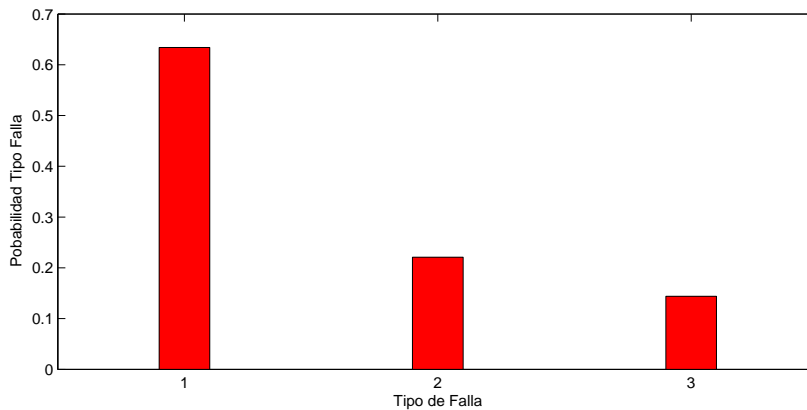


FIGURA 3.9: Probabilidad de tipo de Falla en 22 circuitos de baja tensión.

Otra posibilidad para realizar el modelamiento de las fallas es usar la causa como una variable aleatoria con lo cual se tendrían 3 distribuciones de probabilidad, el lugar de la falla, la causa de la falla y el tipo de falla, las cuales pueden simularse a través de algoritmos de montecarlo mencionados anteriormente. Finalmente, debido a la falta de información y a los problemas señalados al inicio de este capítulo, en este documento se modelará el tipo de falla en los circuitos de 220kV y 115kV con una distribución uniforme, sin embargo para la metodología planteada las interpretaciones de los resultados siguen siendo válidas.

3.3. Generación de fallas en zonas perturbadoras

Una vez se tienen las distribuciones de probabilidad que modelan las variables aleatorias del lugar de la falla y tipo de falla, se procede a generar los valores aleatorios de dichas distribuciones de tal manera que éstos sean estadísticamente similares a las distribuciones encontradas.

Para generar valores aleatorios de una variable cualquiera modelada por una distribución de probabilidad, se debe calcular el valor normalizado y el valor acumulado de la distribución. Como ejemplo podemos ver que para la distribución del tipo de falla mostrada en la figura 3.9 se tienen los valores normalizados y acumulados mostrados en la tabla 3.7.

TABLA 3.7: Valor acumulado del tipo de falla

TOTAL		
Tipo falla	Valor Normalizado	Valor Acumulado
Monofásica	0,634	0,634
Bifásica	0,221	0,855
Trifásica	0,144	1,000

Ahora con un software se generan números aleatorios entre 0 y 1 usando el valor acumulado de la tabla anterior para asignar el tipo de falla, tal como se muestra a continuación.

- Si el número aleatorio generado está entre 0 y 0,634 se asigna un valor de falla monofásica.
- Si el número aleatorio generado está entre 0,635 y 0,855 se asigna un valor de falla bifásica.
- Si el número aleatorio generado está entre 0,855 y 1 se asigna un valor de falla trifásica.

Al hacer un número de muestras suficientemente grande, se logrará que éstas sean estadísticamente similares a la distribución modelada. Se puede verificar si la cantidad de muestras es suficiente si las pruebas estadísticas explicadas posteriormente resultan satisfactorias. El resultado de la generación de números aleatorios para el tipo de falla es entonces un vector de valores de dicha variable aleatoria tal como se muestra en la tabla 3.8.

TABLA 3.8: Generación estocástica de valores de tipo de falla

Muestra	N. Aleatorio	Tipo de Falla
1	0,043	Monofásica
2	0,508	Monofásica
3	0,044	Monofásica
4	0,139	Monofásica
5	0,751	Bifásica
6	0,584	Monofásica
7	0,482	Monofásica
8	0,914	trifásica
9	0,261	Monofásica
10	0,785	Bifásica

El proceso anterior se aplica a todas las variables aleatorias con el fin de realizar la simulación estocástica de fallas. El algoritmo completo aplicado a las zonas perturbadoras del sistema colombiano se explica detalladamente a continuación.

1. Se genera una ubicación de falla a partir de la distribución uniforme de ubicación. Con esto se obtiene en cuál de las líneas de la zona perturbadora ocurrirá una falla.
2. Se genera un valor de Porcentaje de línea. Para estimar en qué tramo ocurrirá la falla, la línea se divide en 10 tramos y el número aleatorio indica en cuál de los tramos ocurrirá la falla.
3. Generación de tipo de falla. Por medio de la distribución de tipo de falla, se estima cuál es el tipo de falla que ocurrirá.
4. Ir la paso 1, o finalizar si se alcanzó el máximo de simulaciones.

El resultado de este proceso es la tabla 3.9 que son los datos de entrada para el simulador de fallas diseñado en secciones anteriores.

TABLA 3.9: Muestra de datos de entrada para simulación de fallas

Tramo	% tramo	Tipo falla
15	100	1
16	100	1
17	100	1
18	100	1
28	0	1
28	10	1
28	20	1
28	30	1
28	40	1
28	50	1
28	60	1

Para asegurar que las muestras obtenidas son estadísticamente idénticas a las distribuciones de entrada, se pueden aplicar criterios como la prueba sobre la media, varianza o ji-cuadrado, o aplicar otras pruebas más estrictas como la de kolmogorov y smirnov las cuales son explicadas detalladamente en [Romero M. 2006].

Con los datos anteriores se realiza la simulación de cada una de las fallas obteniendo como resultado una matriz de voltajes en cada uno de los nodos de interés. Esta matriz es procesada para extraer los valores de tensión que se encuentran en el rango de hundimientos en tensión y son analizados estadísticamente en la siguiente sección.

3.4. Desarrollo de algoritmos para generación de contornos iso-profundidad

Parte de los objetivos de este trabajo es desarrollar una propuesta de visualización para los indicadores calculados. Esta propuesta consiste en visualizar de manera gráfica la información de hundimientos en tensión en un mapa geográfico o unifilar por medio de contornos. Estos contornos ayudarán a identificar cuál es la distribución de zonas críticas en el sistema.

Con el fin de generar dichos contornos se realiza el siguiente procedimiento:

1. Cargar los índices de hundimientos para cada uno de los nodos. Para este caso, consiste en realizar una consulta desde MATLAB a una base de datos,² donde se encuentra toda la información de hundimientos simulados. A través de la consulta se puede calcular el correspondiente indicador que se quiere visualizar para cada uno de los nodos.
2. Cargar la información del mapa y coordenadas (x,y) de cada uno de los nodos. Es necesario tener un mapa geográfico o un plano unifilar en formato de SVG el cual permite agregar información de texto plano de forma directa. Además se carga la información de las coordenadas (x,y) de cada nodo dando su ubicación dentro del mapa. Estas coordenadas están definidas en pixeles y son precalculadas del mapa del sistema.
3. A cada dato de ubicación de nodo (x,y) se le asigna el indicador correspondiente. La tabla 3.10 muestra un ejemplo de la información del índice de hundimientos y las coordenadas de cada nodo.

TABLA 3.10: Ejemplo de información calculada para generación de contornos.

Nodo	Coordenada X	Coordenada Y	Índice (profundidad)
21	3824	4274	0,45
23	3670	4298	0,56
36	2952	4076	0,23
38	2919	4126	0,44
60	3117	4351	0,39

4. Se calcula el rango de los indicadores de todos los nodos y éste se divide en 10

²Debido a la gran cantidad de datos simulados, se implementó una base de datos en MYSQL con el fin de optimizar el manejo de la información y reducir tiempos de cálculos.

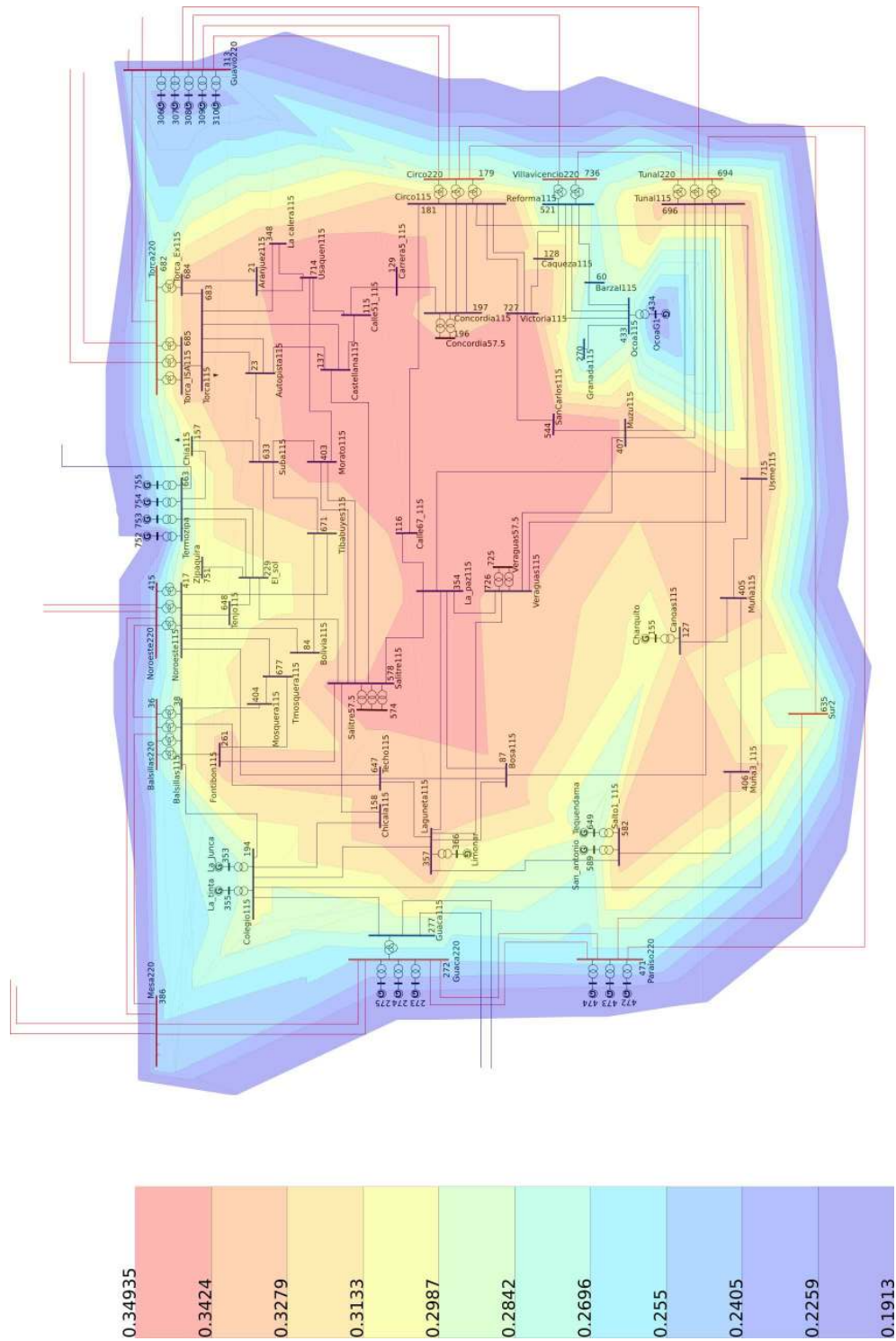


FIGURA 3.11: Contornos de profundidad promedio de hundimientos sobre diagrama unifilar.

La gráfica 3.11 muestra cuál es la profundidad promedio de los hundimientos que ocurren en cada uno de los nodos. La barra izquierda indica el color correspondiente al nivel de profundidad promedio de los hundimientos de tensión con lo cual podemos observar que los nodos en zonas color rojo tienen los hundimientos más críticos y los nodos en zonas azules son los menos perjudicados.

El algoritmo anterior se programa en matlab como una función de tal manera que se pueden generar mapas de manera fácil y sistemática. Posteriormente este algoritmo será modificado para generar contornos en un mapa geográfico usando herramientas de visualización como GoogleMaps³.

3.5. Análisis estadístico de hundimientos

Para observar cómo es la profundidad de los hundimientos debido a las fallas ocurridas en el sistema de 220kV y 115kV, se analizarán los hundimientos obtenidos en la simulación estocástica. Para este análisis, en primer lugar se deben cambiar los valores de tensión de los hundimientos de la simulación por valores de profundidad, lo cual según la figura 1.2 en el capítulo 1, corresponde a restar de 1pu el valor de tensión del hundimiento. Esto se hace con el ánimo de evitar confusiones en la interpretación de los indicadores calculados, logrando una correspondencia entre el valor del indicador y el nivel de gravedad, es decir, que entre más grande sea el valor del indicador éste es considerado más crítico. Para el actual caso, la interpretación implica que entre más grande es el valor de la profundidad, más crítico es el indicador.

Una vez que se calculan las profundidades, se excluyen los datos que no son considerados hundimientos en tensión (generalmente son datos de variación de tensión menor al 10 %). Posteriormente, los datos se agrupan por niveles de tensión, por nodos o por tipo de falla, para calcular los distintos indicadores estadísticos. En este análisis, los indicadores estadísticos propuestos son el Máximo, Percentil 95, Percentil 50 y promedio, los cuales son generalmente utilizados en evaluación de calidad de potencia y además permiten caracterizar de manera detallada el comportamiento de los hundimientos en tensión. Estos indicadores se calculan para los hundimientos de las correspondientes tres fases de cada nodo.

Para visualizar los indicadores mencionados se utilizarán las siguientes opciones:

- *Distribuciones de frecuencia.* Por medio de las distribuciones de frecuencia se puede visualizar la distribución de los datos, observando las posibles tendencias de la profundidad.

³GoogleMaps es una herramienta de Google la cual ofrece una plataforma para generar y modificar todo tipo de gráficos sobre mapas geo-referenciados

- *Tablas de índices.* A través de las tablas se pueden evaluar comparativamente los distintos indicadores calculados para cada grupo de datos.
- *Mapas de iso-profundidad.* Por medio de contornos de igual profundidad se pueden identificar qué zonas son críticas para cada tipo de fallas.
- *Top de nodos críticos y nodos robustos.* Estas tablas muestran cuales son los nodos que son más afectados por las fallas que ocurren en el sistema, así como los nodos que son más robustos.

En la siguiente sección se muestran los resultados de la simulación estocástica planteada. El análisis realizado está dividido en tres partes:

- Primero se analizarán los hundimientos ocasionados por todas las fallas en el sistema,
- luego, se analizarán los indicadores de hundimientos en tensión ocasionados por fallas discriminadas por el nivel de tensión donde ocurren.
- Por último se analizarán los indicadores de hundimientos en tensión discriminados por tipo de falla.

3.5.1. Caracterización de hundimientos por todo tipo de fallas

En un primer caso, se analizan todas las fallas simuladas en las redes de 220kV y 115kV. Los hundimientos en tensión de todos los nodos son caracterizados mediante la distribución de frecuencias de la figura 3.12.

De la figura 3.12 podemos observar que la profundidad de la mayoría de los hundimientos ocasionados por fallas en 115kV y 220kV está concentrada entre 0.1 y 0.5. Sólo algunos hundimientos tienen profundidad mayor a 0.5. Vemos que entre mayor sea la profundidad de los hundimientos, menor es su probabilidad de ocurrencia. Lo anterior se puede observar de manera más detallada en la tabla 3.11, en la cual, se calculan los indicadores estadísticos de máximo, percentil 95, 50 y promedio, discriminados por el nivel de tensión donde se observan los hundimientos en tensión.

TABLA 3.11: Hundimientos de tensión debido a fallas en 220kV y 115kV

Indicador	Profundidad de sags según Nivel tensión observado					
	Todos	220kV (11)	115kV(50)	57,5kV (3)	13,8kV (21)	6,9kV (1)
Maximo	1,000	1,000	1,000	1,000	0,900	0,896
Percentil 95	0,575	0,515	0,598	0,665	0,480	0,559
Percentil 50	0,248	0,232	0,284	0,259	0,199	0,249
Promedio	0,302	0,271	0,327	0,344	0,245	0,311

De la tabla 3.11 podemos observar:

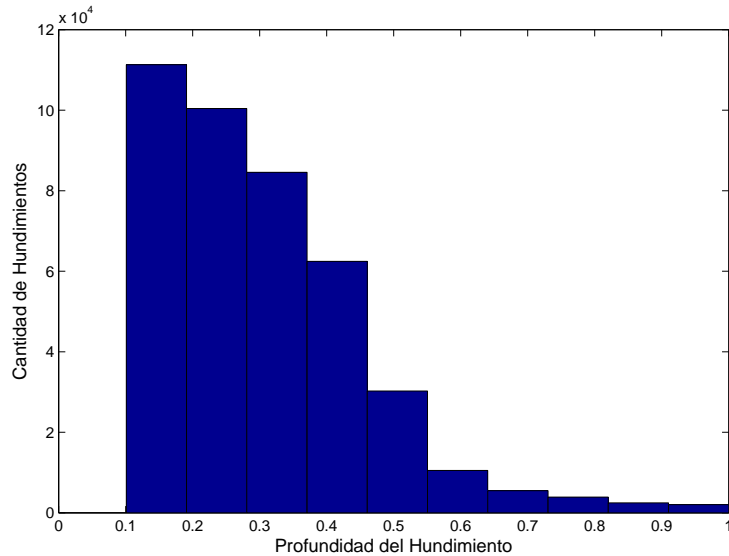


FIGURA 3.12: Profundidad Hundimientos por fallas en 220kV y 115kV

1. La mayoría (95 %) de los hundimientos observados en todos los nodos tienen una profundidad a lo sumo de 0,575. Este valor varía entre 0,480 y 0,665 si se observa en cada nivel de tensión .
2. De acuerdo con todos los indicadores, los nodos menos afectados son los de nivel de tensión 13,8kV que corresponden a los nodos de generación. El 95% de los hundimientos en este nivel de tensión tienen una profundidad de a lo sumo 0,480.
3. De acuerdo al indicador percentil 95, el nivel de tensión más afectado por las fallas es el de 57,5kV con una profundidad de 0,665. El indicador promedio indica lo mismo.

Es importante tener en cuenta la cantidad de nodos que existe en cada nivel de tensión la cual está indicada en paréntesis al lado del nivel de tensión en la tabla 3.11, con esto podemos verificar en cuántos nodos se observa la característica de los hundimientos de cada nivel de tensión.

Para observar la profundidad de los hundimientos en cada uno de los nodos de interés, se calcula el percentil 95 de las profundidades de los hundimientos en tensión en cada uno de los nodos, entonces, utilizando el algoritmo descrito en la sección anterior se genera la gráfica iso-profundidad 3.13.

En la gráfica 3.13, se observan los nodos de interés encerrados por contornos cuyos colores indican cuál es la profundidad del 95 % de los hundimientos para dicho nodo. En el mapa observamos que los nodos de generación que se encuentran unidos al anillo de 220kV son los menos afectados. Los nodos más afectados se encuentran en la zona central en los niveles de tensión 115kV y 57,5kV.

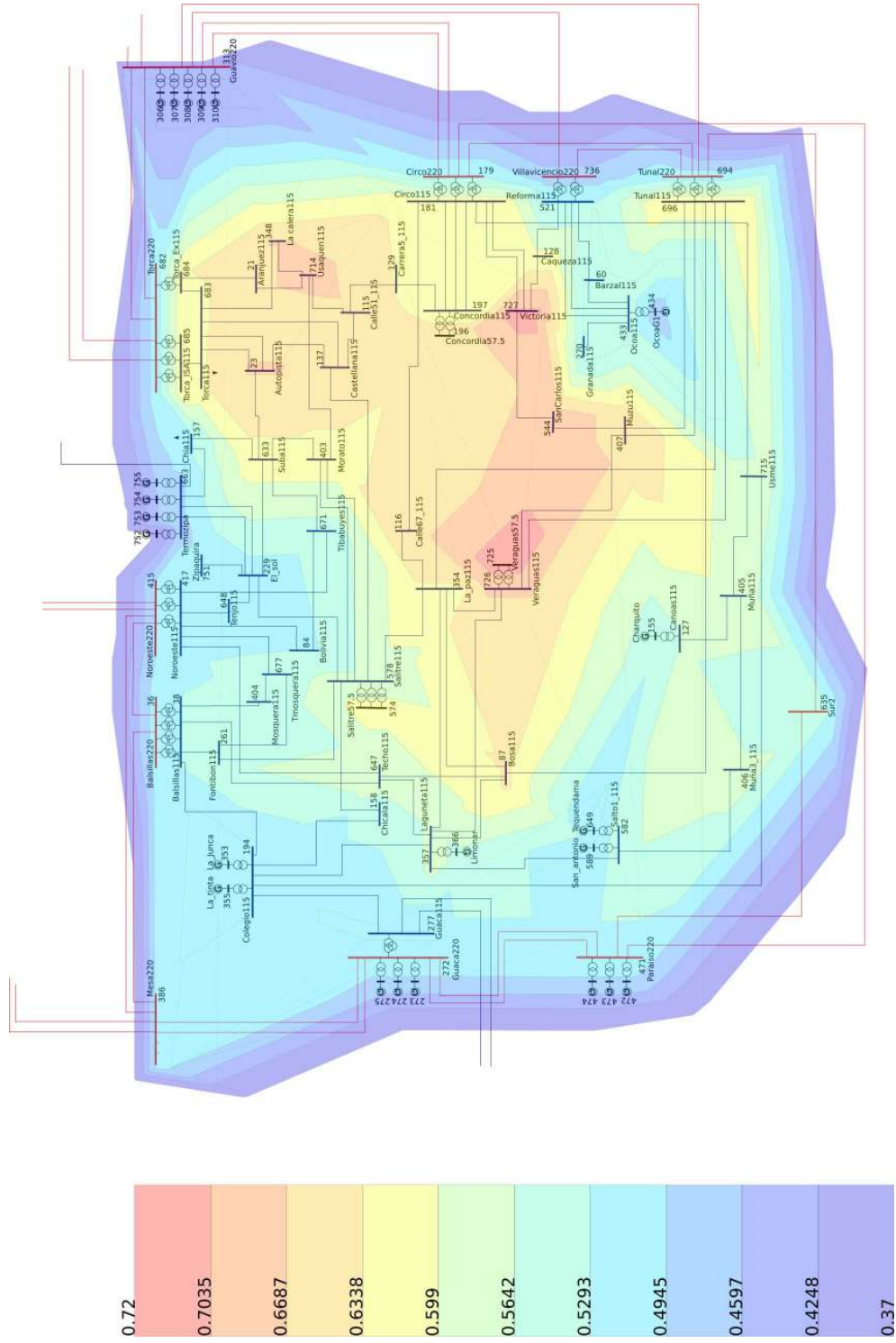


FIGURA 3.13: Mapa de hundimientos en tensión debido fallas en 220kV y 115kV

Con el fin de evaluar cómo es el impacto de las fallas con respecto a profundidad de los hundimientos en tensión generados en cada uno de los nodos de interés, se obtienen las tablas 3.12 y 3.13.

TABLA 3.12: Nodos más afectados por fallas en 220kV y 115kV

Nodo	Nombre	Profundidad
725	VER5 57.5	0,72
727	VICTORIA1 115	0,71
544	S.CARLOS	0,68
714	USAQUEN115	0,68
23	AUTOPISTA115	0,67
348	LA CALERA115	0,67
726	VERAGUAS115	0,67
116	CALLE67 115	0,66
137	CASTELLANA115	0,66
407	MUZU115	0,66

De la tabla 3.12 podemos ver que en general los nodos más afectados tienen profundidades de hundimientos en tensión entre 0,66 y 0,72, siendo el nodo de Veraguas 57,5kV el más afectado. También podemos observar que los nodos más afectados se encuentran concentrados en la parte central derecha del mapa del sistema de distribución, en la figura 3.13.

TABLA 3.13: Nodos menos afectados por fallas en 220kV y 115kV

Nodo	Nombre	Profundidad
752	ZIPAQUIRA_G2	0,39
306	GUAVIO G1	0,4
307	GUAVIO G2	0,4
308	GUAVIO G3	0,4
309	GUAVIO G4	0,4
310	GUAVIO G5	0,4
313	GUAVIO220	0,44
434	OCOA G1	0,44
353	LA JUNCAG1	0,46
355	LA TINTAG1	0,46

En la tabla 3.13 se observa que los nodos menos afectados son en general de 13,8kV y corresponden a los nodos de generación unidos al anillo de 220kV. Estos nodos están distribuidos en la parte exterior del anillo mostrado en la gráfica 3.13. El nodo menos afectado es una unidad de generación colgada al sistema de 220kV mediante la barra de Termozipa.

3.5.2. Caracterización de hundimientos por nivel de tensión de las fallas

A continuación, se observan de manera separada las distribuciones de frecuencia de los hundimientos en todos los nodos, de acuerdo al nivel de tensión donde ocurrieron las fallas. Es decir, se establece un análisis para los hundimientos en tensión debido a fallas en 220kV y otro para hundimientos en tensión debido a fallas en 115kV. El resultado se observa en la figura 3.14.

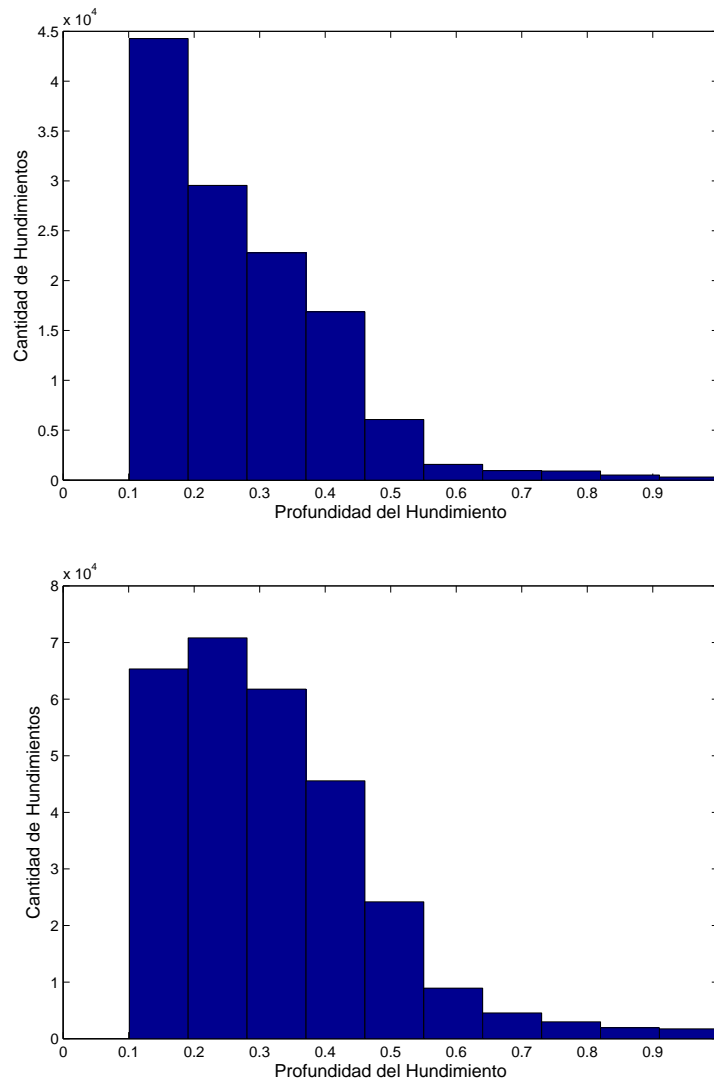


FIGURA 3.14: Profundidad Hundimientos por fallas en 220kV(arriba) y 115kV(abajo)

En la figura 3.14, podemos ver que la mayoría de los hundimientos en tensión causados por fallas en 220kV se encuentran entre 0,1 y 0,5 lo que no es muy distinto a los ocasionados por fallas en 115kV. Sin embargo, la distribución de la profundidad de los hundimientos

ocasionados por fallas en 115kV es más homogénea que la de hundimientos de 220kV.

TABLA 3.14: Hundimientos de tensión debido a fallas en 220kV

Indicador	Profundidad de sags según Nivel tensión observado					
	Todos	220kV (11)	115kV(50)	57,5kV (3)	13,8kV (21)	6,9kV (1)
Máximo	1,000	1,000	0,768	0,611	0,883	0,428
Percentil 95	0,506	0,698	0,465	0,450	0,644	0,393
Percentil 50	0,296	0,271	0,339	0,352	0,221	0,326
Promedio	0,271	0,304	0,268	0,268	0,256	0,233

TABLA 3.15: Hundimientos de tensión debido a fallas en 115kV

Indicador	Profundidad de sags según Nivel tensión observado					
	Todos	220kV (11)	115kV(50)	57,5kV (3)	13,8kV (21)	6,9kV (1)
Máximo	1,000	0,684	1,000	1,000	0,900	0,896
Percentil 95	0,600	0,424	0,653	0,736	0,449	0,580
Percentil 50	0,256	0,272	0,273	0,283	0,210	0,235
Promedio	0,316	0,256	0,352	0,375	0,240	0,341

Al observar las tablas 3.14 y 3.15 podemos ver que:

1. La mayoría (95 %) de los hundimientos en tensión ocasionados por fallas en 115kV tienen profundidad de 0,6 lo cual es más crítico que la profundidad de los hundimientos por fallas en 220kV que corresponde 0,506.
2. Las fallas que ocurren en 115kV afectan en mayor medida a las barras de 57,5kV con hundimientos de tensión a lo sumo de 0,736 de profundidad.
3. El nivel de tensión menos afectado por fallas que ocurren en 115kV es el de 220kV con profundidades de 0,424 (percentil95) seguido del de 13,8kV con profundidades de 4,49.
4. Según el indicador percentil 95, las fallas que ocurren en 220kV afectan en mayor medida a las barras de 220kV con profundidad máxima de 0,698.
5. El nivel de tensión menos afectado por fallas que ocurren en 220kV es el de 6,9kV seguido por 57,5kV con 0,450.
6. Los nodos de 220kV son más afectados por las fallas que ocurren en 220kV.
7. Los nodos de 115kV son más afectados por las fallas que ocurren en 115kV.
8. Los nodos de 57,5kV son más afectados por las fallas que ocurren en 115kV.
9. Los nodos de 13,8kV son más afectados por las fallas que ocurren en 220kV.
10. Los nodos de 6,9kV son más afectados por las fallas que ocurren en 115kV.

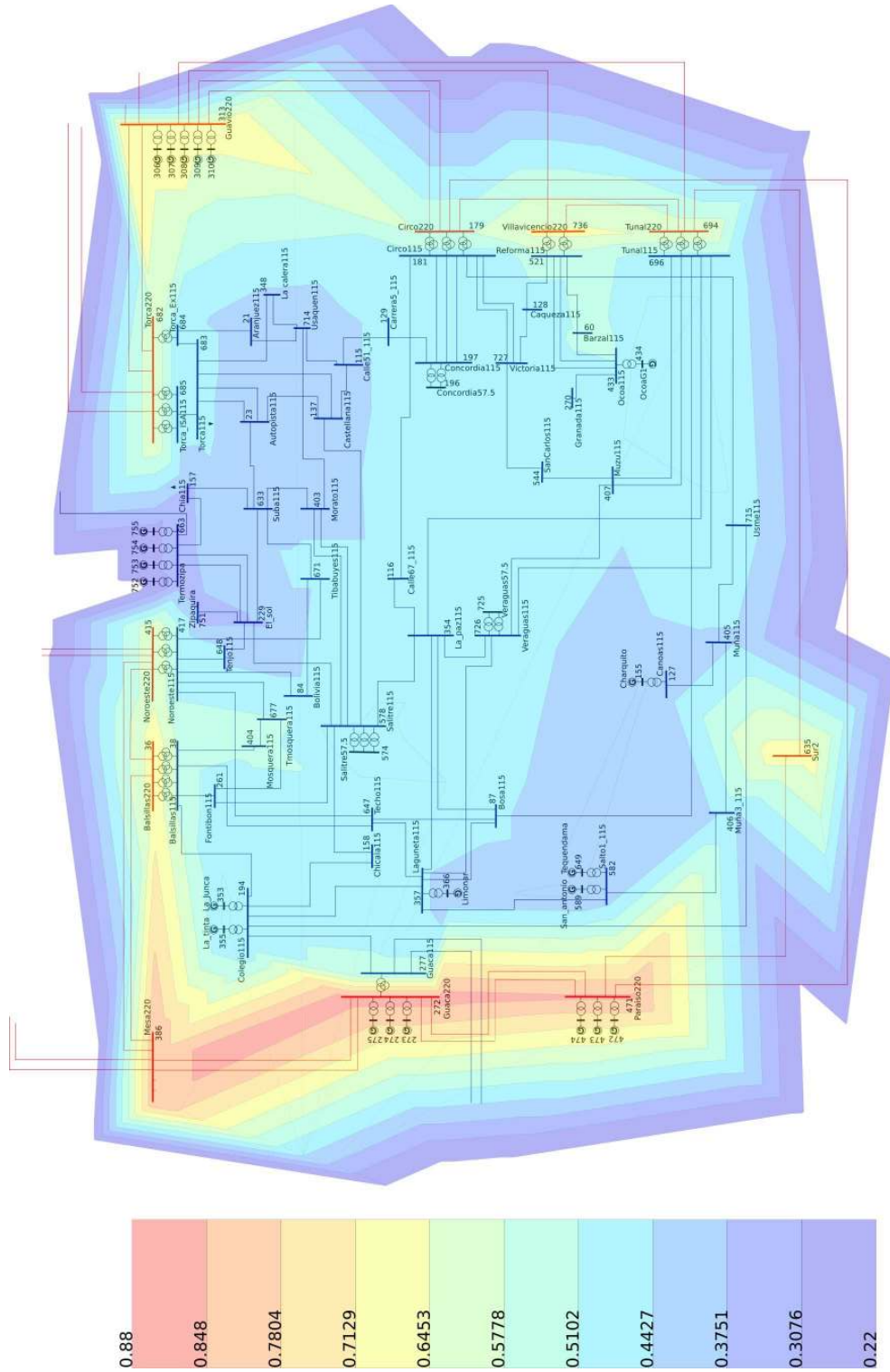


FIGURA 3.15: Mapa de hundimientos debido a fallas en 220kV

De acuerdo con la figura 3.15 Los nodos más afectados por las fallas que ocurren en 220kV son los nodos que se encuentran en el anillo de 220kV y los nodos generadores unidos a ese anillo. En general los nodos más afectados están en la parte izquierda del mapa del sistema, sin embargo, todos los nodos de 220kV son afectados.

TABLA 3.16: Nodos más afectados por fallas en 220kV

Nodo	Nombres	Profundidad
272	GUACA220	0,88
471	PARAISO220	0,86
386	MESA220	0,84
273	GUACA_G1	0,8
274	GUACA_G2	0,8
275	GUACA_G3	0,8
472	PARAISO_G1	0,74
473	PARAISO_G2	0,74
474	PARAISO_G3	0,74
313	GUAVIO220	0,7

Según la tabla 3.16, los nodos más afectados pertenecen a los niveles de tensión 220kV y 13,8kV, los cuales son el anillo de transmisión y los nodos de generación unidos a éste. El nodo más afectado es Guaca de 220kV con hundimientos en tensión de 0,88.

TABLA 3.17: Nodos menos afectados por fallas en 220kV

Nodo	Nombres	Profundidad
752	ZIPAQUIRA_G2	0,24
753	ZIPAQUIRA_G3	0,29
754	ZIPAQUIRA_G4	0,29
755	ZIPAQUIRA_G5	0,29
649	TEQUENDAMAG1	0,35
589	SAN_ANTONIOG1	0,36
663	TERMOZIPA115	0,36
155	CHARQUITOG1	0,37
157	CHIA115	0,37
229	EL_SOL115	0,37

Según la tabla 3.17, los nodos menos afectados pertenecen a los niveles de tensión 13,8kV, los cuales son los nodos de generación unidos al sistema de 115kV. El nodo menos afectado es el nodo de generación de Zipaquirá con hundimientos en tensión de 0,24.

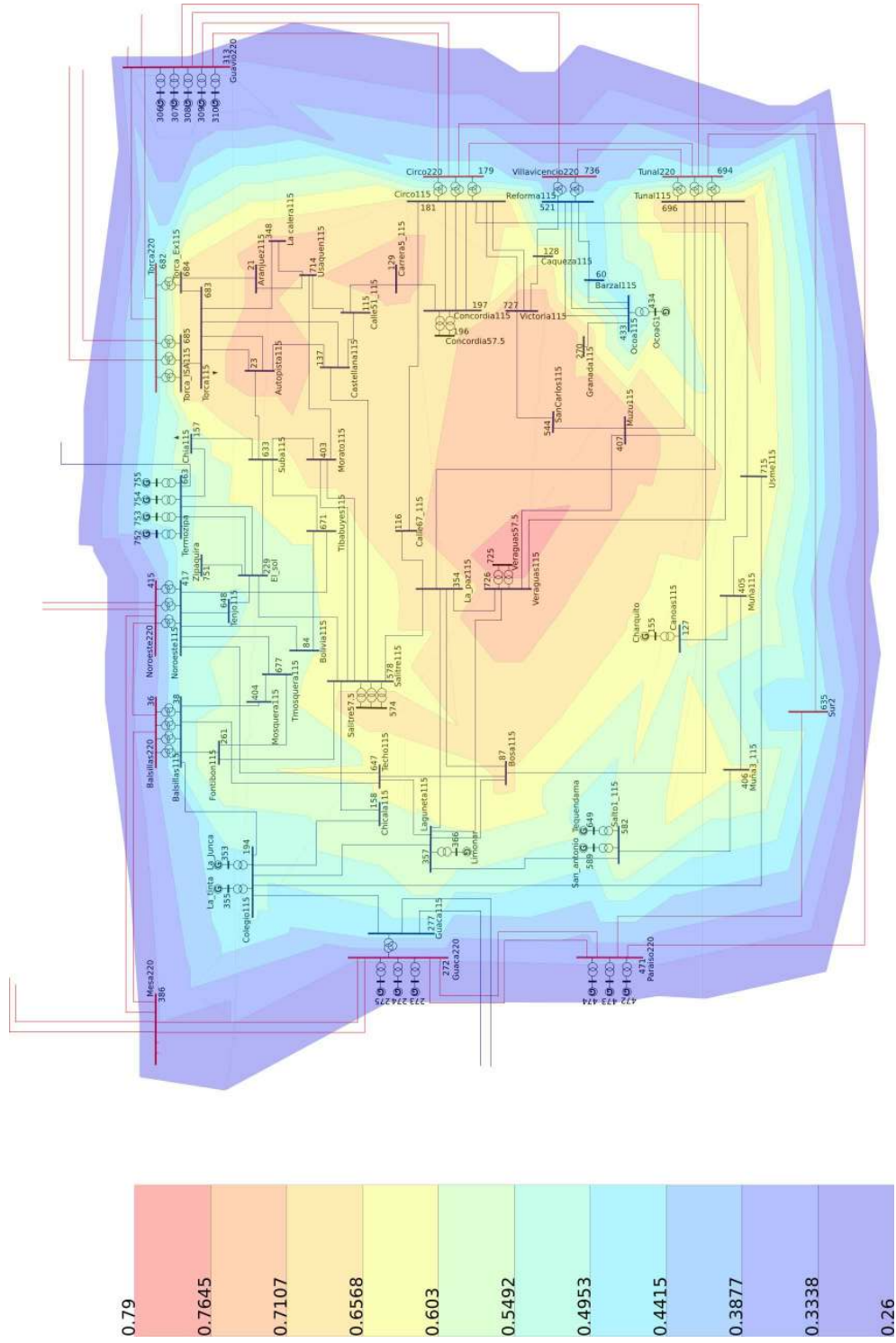


FIGURA 3.16: Mapa de hundimientos debido fallas en 115kV

En la figura 3.16 observamos que los nodos más afectados se encuentran en la zona central derecha, abarcando en mayor parte nodos de 115kV y un nodo de 57,5kV. Es importante notar que los nodos afectados se encuentran concentrados en una zona. Los nodos menos afectados son los del anillo de 220kV y los nodos generadores unidos a dicho anillo. Esto implica que las fallas que ocurren en 115kV afectan en menor medida a los nodos aguas arriba en 220kV.

TABLA 3.18: Nodos más afectados por fallas en 115kV

Nodo	Nombres	Profundidad
727	VICTORIA1.115	0,76
726	VERAGUAS115	0,74
725	VER5.57.5	0,79
714	USAQUEN115	0,71
544	S.CARLOS_EEB115	0,73
407	MUZU115	0,74
348	LA_CALERA115	0,73
129	CARRERA5.115	0,75
23	AUTOPISTA115	0,74
21	ARANJUEZ.115	0,72

Según la tabla 3.18, los nodos más afectados pertenecen a los niveles de tensión 115kV y 57,5kV. El nodo más afectado es Victoria1 de 115kV con hundimientos en tensión de 0,76.

TABLA 3.19: Nodos menos afectados por fallas en 115kV

Nodo	Nombres	Profundidad
306	GUAVIO_G1	0,28
307	GUAVIO_G2	0,28
308	GUAVIO_G3	0,28
309	GUAVIO_G4	0,28
310	GUAVIO_G5	0,28
273	GUACA.G1	0,29
274	GUACA.G2	0,29
275	GUACA.G3	0,29
472	PARAISO_G1	0,29
473	PARAISO_G2	0,29

Según la tabla 3.19 los nodos menos afectados por fallas que ocurren en 115kV son los nodos de generación unidos al anillo de 220kV. Los nodos más robustos pertenecen a Guavio 13,8kV con hundimientos en tensión de 0,28.

3.5.3. Caracterización de hundimientos por tipo de falla

Ahora el análisis se centrará en las diferencias de los hundimientos en tensión generados por distintos tipos de falla. Para esto se filtran los hundimientos en tensión según el tipo de falla que los generó, monofásicas, bifásicas, trifásicas y bifásicas a tierra. luego se calculan los indicadores para cada grupo de manera separada. A continuación se observan las distribuciones de frecuencia de la profundidad de los hundimientos en tensión obtenidos para cada tipo de falla.

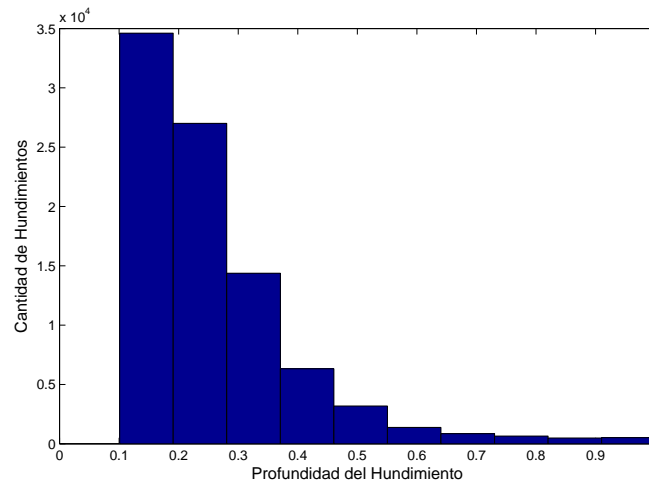


FIGURA 3.17: Profundidad Hundimientos por fallas monofásicas

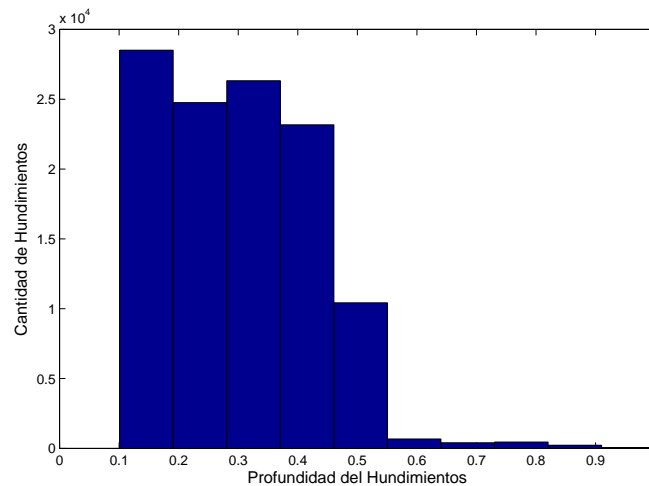


FIGURA 3.18: Profundidad Hundimientos por fallas bifásicas

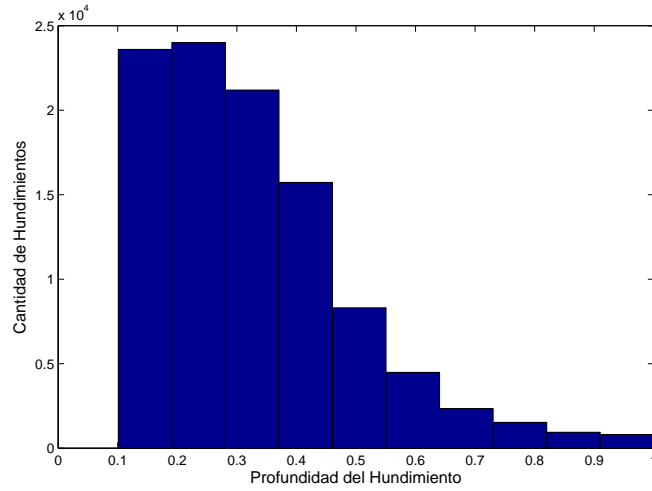


FIGURA 3.19: Profundidad Hundimientos por fallas trifásicas

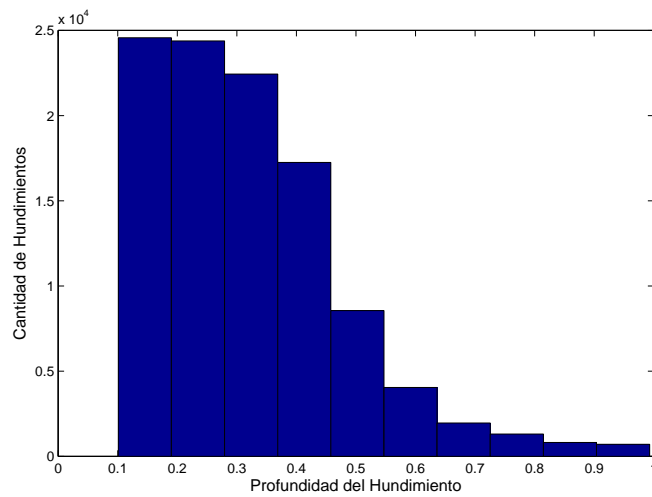


FIGURA 3.20: Profundidad Hundimientos por fallas bifásicas a tierra

De las gráficas 3.17, 3.18, 3.19 y 3.20 podemos ver que:

1. Las fallas monofásicas producen hundimientos en tensión con profundidades de entre 0,1 y 0,5. Entre mayor es la profundidad del hundimiento, menor es la ocurrencia de este.
2. Las fallas bifásicas producen hundimientos en tensión con profundidades de entre 0,1 y 0,5. La distribución de las profundidades es mucho más homogénea que los demás tipos de fallas.
3. Las fallas trifásicas producen hundimientos en tensión con profundidades de entre 0,1 y 0,9.
4. Las fallas bifásicas a tierra producen hundimientos en tensión con profundidades de entre 0,1 y 0,9. Sin embargo la distribución de las profundidades es distinta que la de las fallas trifásicas.

En la tabla 3.20 se observa de manera comparativa los indicadores para los hundimientos en tensión ocasionados por cada tipo de falla.

TABLA 3.20: Percentil 95 de hundimientos de tensión por tipo de falla

Percentil 95	Profundidad de sags según Nivel tensión observado					
	Todos	220kV (11)	115kV(50)	57,5kV (3)	13,8kV (21)	6,9kV (1)
todas	0,575	0,515	0,598	0,665	0,480	0,559
monofasicas	0,525	0,542	0,587	0,391	0,324	0,348
bifasicas	0,495	0,469	0,489	0,732	0,526	0,585
trifasicas	0,653	0,577	0,684	0,724	0,502	0,571
bif-tierra	0,621	0,563	0,642	0,707	0,490	0,558

Según la tabla 3.20 las fallas monofásicas afectan en mayor medida a los nodos de 115kV con profundidades de 0,587. Por otro lado, las fallas bifásicas, trifásicas y bifásicas a tierra afectan en mayor media a los nodos de 57,5kV con profundidades cercanas a 0,7.

Los nodos de niveles de tensión 220kV, 115kV y 57,5kV son afectados en mayor medida por las fallas trifásicas. Los nodos de 13,8kV y 6,9kV son afectados en mayor medida por fallas bifásicas.



FIGURA 3.21: Mapa de hundimientos debido fallas monofásicas



FIGURA 3.22: Mapa de hundimientos debido fallas bifásicas

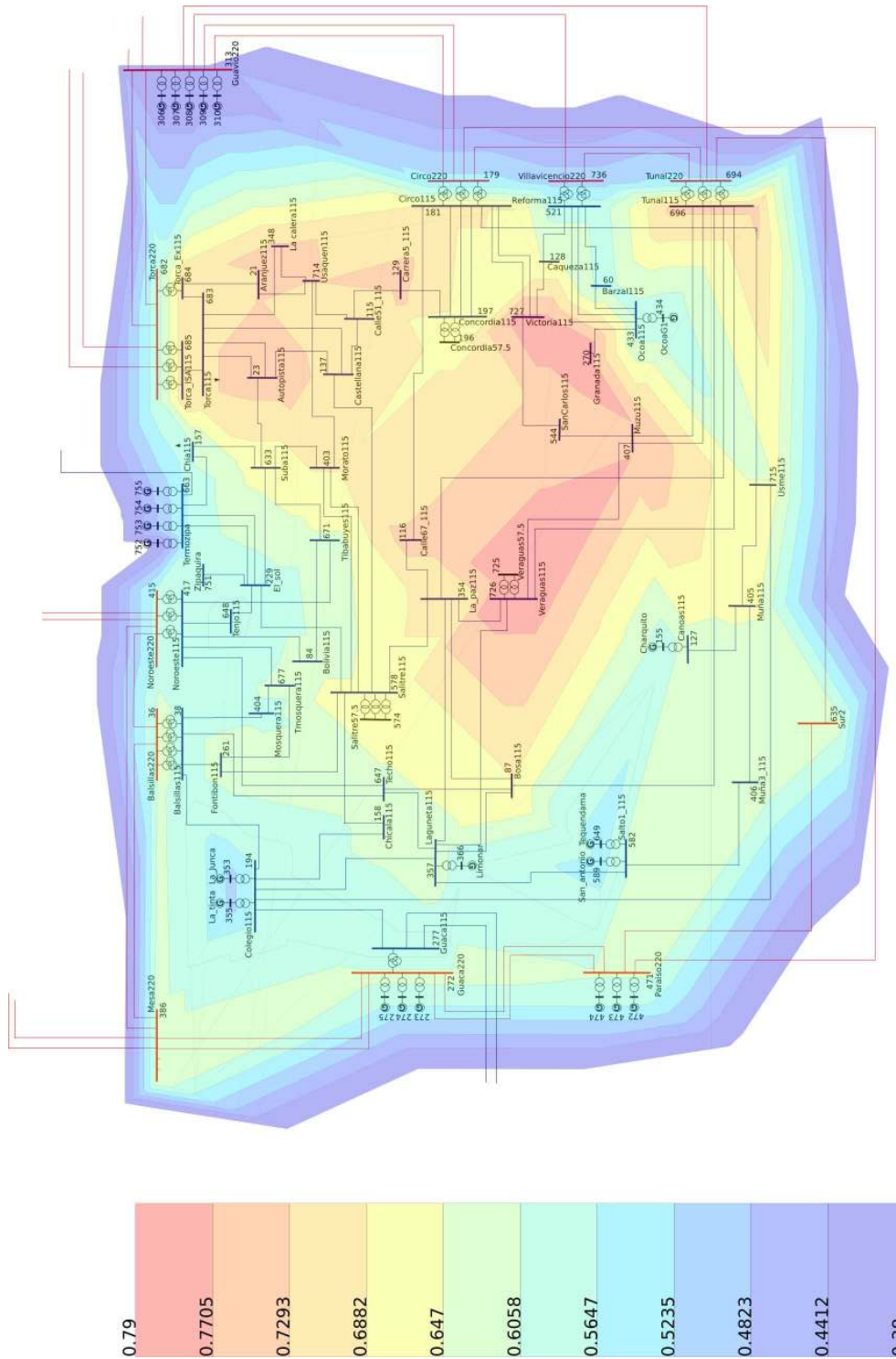


FIGURA 3.23: Mapa de hundimientos debido fallas trifásicas

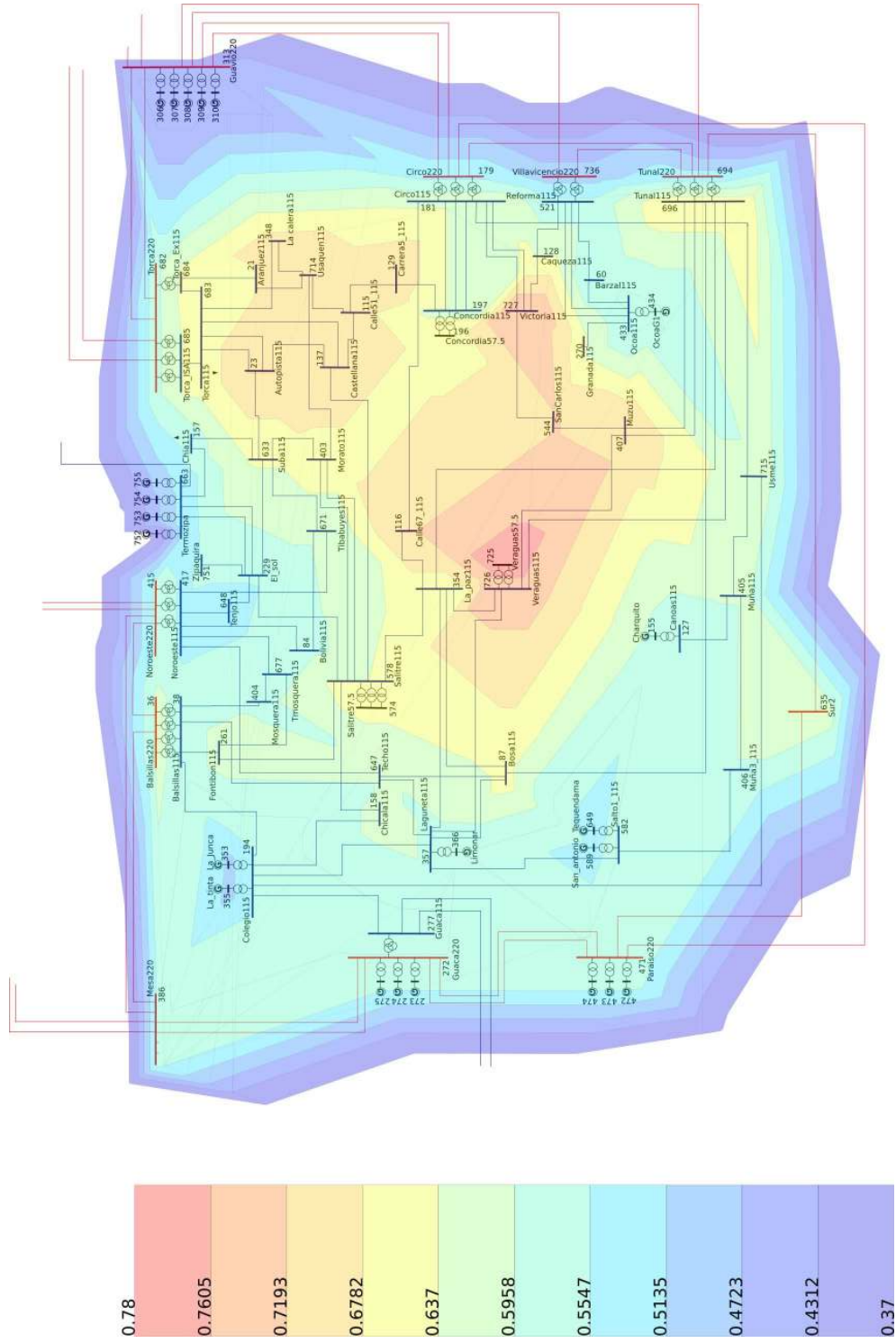


FIGURA 3.24: Mapa de hundimientos debido fallas bifásicas a tierra

De las gráficas 3.24, 3.23, 3.22 y 3.21 podemos observar que:

1. Las fallas monofásicas afectan los nodos de 220kV en la parte izquierda del sistema y los nodos de 115 en la parte central derecha. Los nodos más afectados tienen profundidades entre 0,68 y 0,76.
2. Las fallas bifásicas afectan en mayor parte a los nodos de 57,5kV con profundidades entre 0,65 y 0,8. El resto de los nodos tienen profundidades similares cercanas a 0,4.
3. Las fallas trifásicas afectan en mayor parte a los nodos de 115kV en la parte central derecha del mapa del sistema, con profundidades entre 0,72 y 0,79.
4. Las fallas bifásicas afectan en mayor parte a los nodos de 115kV y 57,5kV en la parte central derecha del mapa del sistema, con profundidades entre 0,71 y 0,89.

Finalmente, la tabla 3.21 muestra el indicador percentil 95 discriminado por lugar de ocurrencia de las fallas y por tipo de fallas.

TABLA 3.21: Indicadores de hundimientos por ubicación y tipo de falla

Percentil 95	Profundidad de sags según Nivel tensión observado					
	Todos	220kV (11)	115kV(50)	57,5kV (3)	13,8kV (21)	6,9kV (1)
Mono 220kV	0,470	0,801	0,406	0,348	0,474	0,321
Bi 220kV	0,484	0,506	0,467	0,466	0,719	0,413
Tri 220kV	0,553	0,827	0,471	0,445	0,714	0,387
Bit 220kV	0,547	0,786	0,481	0,435	0,696	0,378
Mono 115kV	0,553	0,368	0,659	0,409	0,289	0,355
Bi 115kV	0,498	0,411	0,495	0,799	0,497	0,612
Tri 115kV	0,679	0,445	0,735	0,793	0,462	0,598
Bit 115kV	0,643	0,445	0,702	0,776	0,450	0,585

De la anterior tabla podemos observar que:

1. Los nodos de 220kV son más afectados por fallas trifásicas en 220kV y menos afectados por fallas monofásicas en 115kV.
2. Los nodos de 115kV son más afectados por fallas trifásicas en 115kV y menos afectados por fallas monofásicas en 220kV.
3. Los nodos de 57,5kV son más afectados por fallas bifásicas en 115kV y menos afectados por fallas monofásicas en 220kV.
4. Los nodos de 13,8kV son más afectados por fallas bifásicas en 220kV y menos afectados por fallas monofásicas en 115kV.
5. Los nodos de 6,9kV son más afectados por fallas bifásicas en 115kV y menos afectados por fallas monofásicas en 220kV.

Capítulo 4

Correlación fallas vs hundimientos en tensión por medio de técnicas de agrupamiento (Clustering)

En este capítulo se explican detalladamente los últimos dos pasos de la metodología de análisis de hundimientos de tensión correspondientes a la identificación de patrones de hundimientos por medio de algoritmos de clustering (kmeans) y clasificación de hundimientos reales.

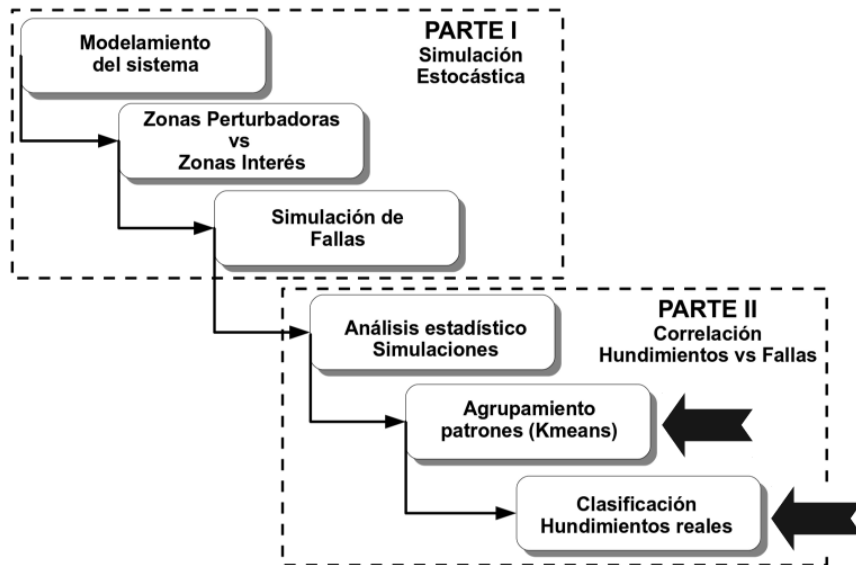


FIGURA 4.1: Identificación de patrones y clasificación de hundimientos en la metodología propuesta

Actualmente en Colombia, los operadores de red y el ente regulador CREG están

interesados en conocer las condiciones de calidad (entre ellas hundimientos en tensión) de las redes de distribución, con el fin de evaluar y posteriormente regular la emisión de dichas perturbaciones. En ese sentido, es importante para los operadores de red determinar cuáles son las causas de los hundimientos en tensión, con el ánimo de definir responsabilidades, determinar posibles acciones de mitigación, y mejorar los indicadores de evaluación reportados a la CREG.

Con respecto al análisis de causas de hundimientos, en el capítulo 3 después de realizar una simulación estocástica de fallas en el sistema colombiano, se obtuvieron las características de profundidad de los hundimientos en tensión debido a fallas en 220kV y 115kV. En este análisis se identificó cómo es la característica de los hundimientos generados por fallas, pero no se conoce en cuáles zonas ocurren las fallas que generan más hundimientos en tensión. En ese sentido es importante resaltar que los operadores de red actualmente registran los hundimientos ocurridos en cada una de las barras del sistema. Esta información es analizada dentro del proyecto de investigación *Innovación en Clusters de Calidad de Potencia*[CLUSTERS. 2010], del cual esta tesis hace parte, por lo que es posible tomar esa información para determinar en qué zonas ocurren las fallas que ocasionan dichos hundimientos.

En este capítulo se propone un algoritmo para correlacionar los hundimientos registrados por el operador de red con fallas simuladas en el capítulo anterior a través de algoritmos de agrupación de datos (clustering) con la intención de:

- Encontrar las zonas donde ocurren las fallas que ocasionan los hundimientos en tensión registrados por el operador de red CODENSA S.A. Esto se hace con el propósito de implementar estrategias de evaluación y mantenimiento en dichas zonas con el fin reducir la posibilidad de fallas.
- Identificar en cuáles zonas del sistema colombiano, que no hacen parte del operador de red CODENSA S.A , ocurren fallas que generan hundimientos en las barras de dicho operador de red. Ésto se hace con el fin de sentar bases para la definición de las responsabilidades de cada uno de los operadores con respecto a los hundimientos en tensión ocurridos.

En las próximas secciones se explican detalladamente las principales características de las técnicas de agrupamiento de datos analizando el funcionamiento del algoritmo kmeans mediante un ejemplo ilustrativo. Posteriormente se usa el algoritmo para clasificar los hundimientos generados en la simulación estocástica del capítulo anterior, mencionando las limitaciones de dimensionalidad del algoritmo kmeans y las opciones para determinar el número de grupos deseados. Debido a la relación causa-efecto entre las fallas simuladas y los hundimientos generados, los distintos grupos de hundimientos obtenidos serán una clasificación indirecta de las fallas, que hace referencia al lugar y tipo de falla, que ocasionan tales hundimientos. El agrupamiento de los datos simulados será usado como clasificador de los eventos reales, asignando cada evento real a un grupo de hundimientos simulados y por lo tanto, a un grupo de fallas en ciertas zonas que representan los lugares donde pudo

haber ocurrido la falla.

El resultado de lo anterior es la identificación de las zonas donde ocurren las fallas que posiblemente generan los hundimientos en tensión reportados por el operador de red CODENSA S.A.

4.1. Técnicas de agrupamiento de datos

El objetivo de las técnicas de agrupamiento (clustering) es identificar o clasificar datos en distintos grupos o categorías. Este agrupamiento se realiza de tal manera que los datos de un mismo grupo tienen características similares entre sí, y éstas son distintas a las características de los datos en otros grupos [Queipo. 2001, Ramos A. 2001].

Cada dato ' A_x ' es caracterizado por un conjunto de variables (dimensiones) que reflejan las cualidades fundamentales del mismo (x, y, \dots). Cada grupo es identificado por su centro (centroide), el cual se calcula como el promedio (o la mediana según el algoritmo de clustering) de los datos que pertenecen a dicho grupo. Para clasificar N datos en k grupos, es necesario ejecutar un algoritmo con dos objetivos fundamentales:

- Minimizar las desviaciones entre los datos que pertenecen a un mismo grupo.
- Maximizar las distancias entre los centroides de los grupos.

El resultado de esto es la identificación de distintos grupos (clusters) los cuales se emplean para clasificar cada uno de los datos tal como se muestra en la figura 4.2.

Existen varias técnicas para realizar agrupamiento de datos (kmeans, kmedoids, jerárquicos, etc.), cuyas diferencias se basan fundamentalmente en la forma de medir distancias entre los datos y los clusters, el criterio para definir la pertenencia de los datos a cada cluster y la forma de optimizar el agrupamiento [Berzal F. 2005]. Entre las técnicas de agrupamiento se encuentra el algoritmo kmeans, la idea de este algoritmo es realizar una clasificación de N datos en k grupos. Cada grupo es identificado por un centroide cuya posición se va optimizando por medio del cálculo de la media de los datos del respectivo grupo. El algoritmo de kmeans se desarrolla de manera detallada en los siguientes pasos:

1. En primer lugar se debe definir un número inicial de grupos o clusters k para la clasificación. Debido a la dificultad para determinar la cantidad correcta de clusters, generalmente se realiza un barrido para varios valores de k calculando algunos indicadores que estiman la correcta cantidad de clusters. Muchas veces la cantidad de clusters puede definirse dependiendo de la utilidad que se puede obtener de un agrupamiento específico.
2. Una vez se define la cantidad de clusters, se procede a ubicar k puntos (identificados como centroides) en el espacio muestral. Esta ubicación inicial de centroides puede realizarse de las siguientes maneras:

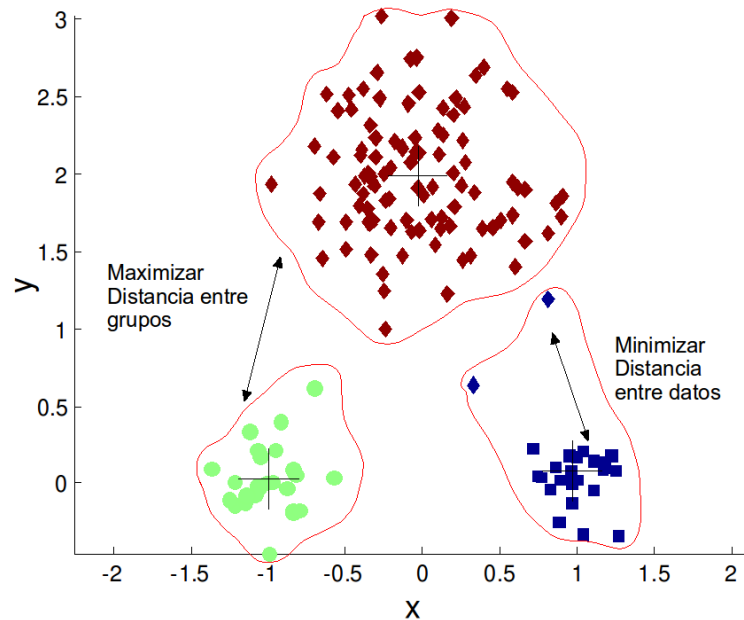


FIGURA 4.2: Agrupación de datos de dos dimensiones

- Se seleccionan k datos aleatorios entre todas las observaciones.
- Se seleccionan k posiciones de los datos más alejados entre sí.
- Se seleccionan k puntos aleatorios y uniformemente distribuidos del rango de las observaciones.
- Se selecciona una muestra preliminar del 10 % de los datos a la cual se le realiza un pre-agrupamiento inicializado con muestras aleatorias.
- Seleccionando un set de centroides pre-establecidos.

Esta ubicación inicial tiene un fuerte impacto en el resultado final del agrupamiento, por lo que es necesario inicializar muchas veces el algoritmo para evitar estancarse en máximos locales.

3. Posteriormente, se calculan las distancias entre los N datos y cada uno de los centroides. La manera de calcular estas distancias puede ser definida por el usuario, sin embargo lo más común es usar la distancia euclidiana.
4. Cada uno de los datos es asignado al grupo con el centroide más cercano. Luego se actualiza la posición de cada centroide calculando la media de los datos asignados al respectivo grupo.
5. Con la nueva ubicación de centroides se retorna al paso 4 hasta que las distancias entre datos y centroides sea mínima.

Un ejemplo ilustrativo se observa en la figura 4.3 donde se utiliza el algoritmo kmeans para clasificar los datos en tres grupos distintos.

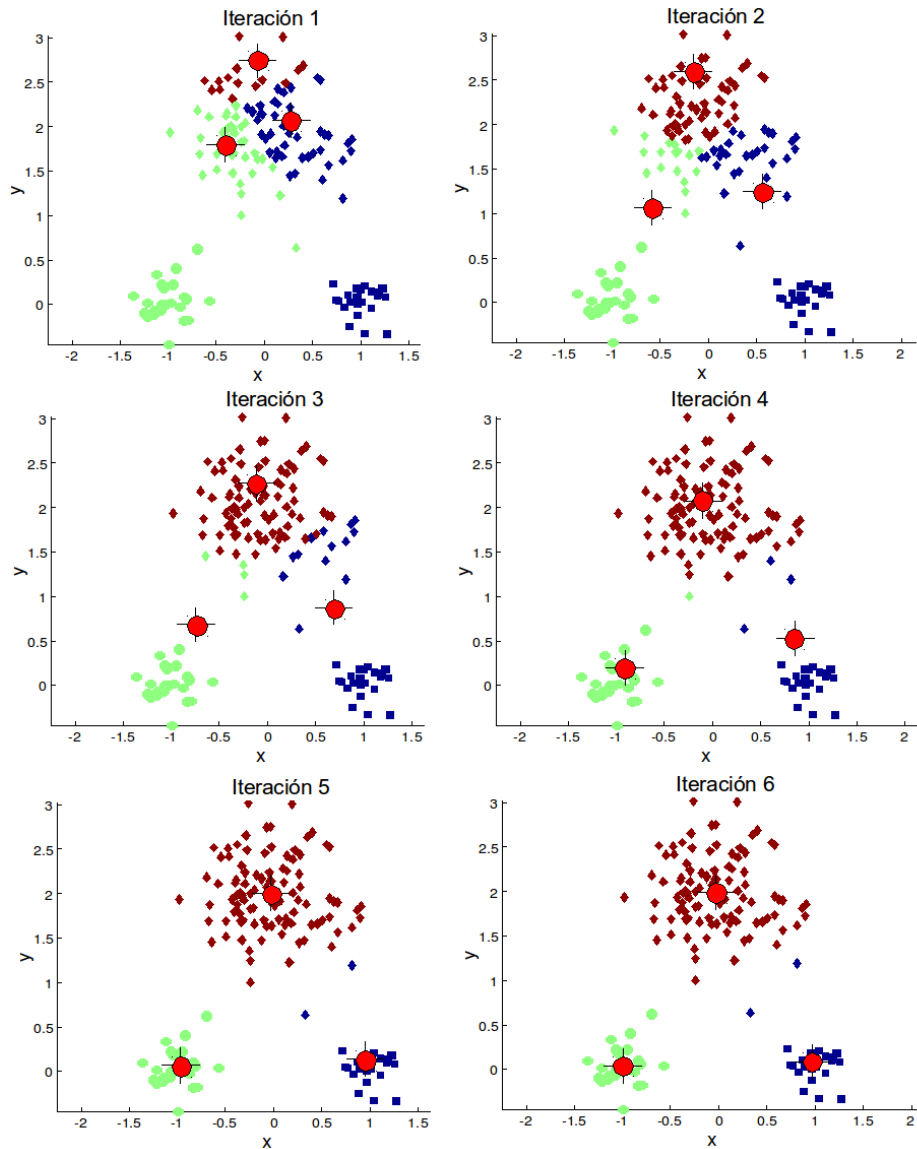


FIGURA 4.3: Proceso iterativo del algoritmo Kmeans

En una primera iteración 3 centroides (color rojo) son ubicados en unas posiciones aleatorias iniciales. Posteriormente se calculan las distancias entre los datos y los centroides y se asignan los datos a cada centroide. En esta iteración se definen los tres grupos de datos (Colores verde, rojo y azul) iniciales. Para cada grupo de datos se calcula la media y se actualiza el lugar de cada uno de los centroides, tal como se muestra en la iteración 2.

El proceso se repite varias veces entre la iteración 3 y la iteración 6. En la gráfica se puede observar como los centroides se van desplazando hacia los grupos naturales de datos existentes. Finalmente en la iteración 6 se tienen 3 grupos de datos bien definidos.

Tal como se mencionó anteriormente, el algoritmo es muy sensible a los valores iniciales de los centroides, por lo que se puede quedar en un máximo local. Un ejemplo de ésto se ve en la figura 4.4 donde se muestra una primera iteración para el mismo grupo de datos un conjunto de centroides iniciales distintos.

En la figura 4.4 el algoritmo kmeans inicia con centroides distribuidos aleatoriamente en el rango de los datos. Al alcanzar la última iteración (6) vemos que los datos fueron agrupados de manera distinta a la de la figura 4.3. En este caso, el algoritmo alcanzó un máximo local en el cual los datos no se encuentran agrupados de manera óptima.

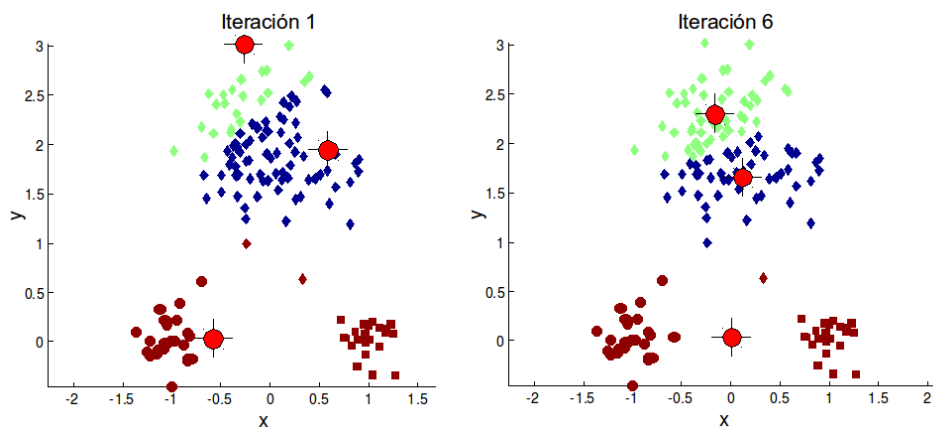


FIGURA 4.4: Datos agrupados con kmeans (Máximo local)

Para evitar que el algoritmo finalice en un máximo local se ejecuta el algoritmo kmeans varias veces aumentando de esta manera la probabilidad de alcanzar un máximo absoluto.

4.2. Agrupación de hundimientos en tensión

Tal como se mencionó al inicio de este capítulo, el objetivo de esta sección es clasificar los hundimientos simulados en el capítulo 3 usando el algoritmo kmeans. Debido a que la información se correlacionará posteriormente con hundimientos reales ocurridos en las barras de CODENSA S.A. ESP., se deben tener en cuenta los siguientes aspectos para el tratamiento de la información simulada.

1. En primer lugar, para realizar la correlación se debe tener las mismas variables en la simulación y en los datos reales, esto quiere decir, que se identificará en

cuáles barras de CODENSA existe información de hundimientos en tensión. Analizando la información del proyecto de investigación SIGCE (Sistema Integral de Gestión de Calidad [SIGCE. 2008]) y el proyecto CLUSTERS [CLUSTERS. 2010] se identificaron 48 barras de 115kV con información de hundimientos en tensión.

- Una vez que las barras de CODENSA son identificadas, se procede a extraer la información simulada de hundimientos en tensión de dichas barras organizando la información en una matriz de N filas x 144 columnas, que corresponden a las 3 fases de los 48 nodos.
- Debido a que en la realidad no existe concordancia entre el etiquetado de las fases(A,B,C) en la simulación con el etiquetado realizado por CODENSA¹, la información de cada una de las fases de cada barra se organiza según su severidad de la más crítica a la menor. De esta manera es posible hacer la correlación entre información simulada e información real.

Una vez organizada la información simulada podemos observar en la figura 4.5 dos matrices, una matriz con la información de las fallas que causan los hundimientos y otra de los hundimientos vistos en cada fase de 48 barras.

Fallas			Hundimientos de Tensión								
Tramo	%línea	Tipo Falla	V1a	V1b	V1c	V2a	V2b	V2c	V3a	V3b	V3c
17	100	1	0,612	0,821	1,000	0,059	0,795	1,000	0,000	0,723	1,000
28	30	1	0,831	1,000	1,000	0,824	1,000	1,000	0,715	0,846	1,000
259	90	4	0,780	0,078	0,000	0,780	0,000	0,000	0,754	0,754	0,754
138	10	3	1,000	0,033	0,000	1,000	1,000	0,569	0,807	1,000	1,000
CAUSA						EFECTO					

FIGURA 4.5: Organización de información para agrupamiento de datos

En la figura 4.5 observamos que la magnitud de los hundimientos en tensión está organizada como se mencionó anteriormente. Además se hace referencia a la relación causa-efecto entre las fallas en el sistema y los hundimientos observados en las barras.

La matriz de fallas se compone de tres variables, tramo de falla, porcentaje de línea y tipo de falla, con una cantidad de datos igual a 6355. (3 variables x 6355 datos). Por otro lado la matriz de hundimientos cuenta con 144 variables o dimensiones que son producto de 3

¹Es decir, en la instalación de equipos de medición de calidad se conectan los cables A, B y C a las fases del punto de medida, cuidando que queden en secuencia positiva, sin embargo las fases pueden quedar rotadas, es decir (B,C,A, o C,A,B)

fases por 48 nodos y 6355 datos (144 variables x 6355 datos).

El algoritmo kmeans será aplicado a la matriz de hundimientos en tensión identificando distintos grupos, los cuales harán referencia a datos de 144 dimensiones que tienen cierto grado de similitud entre si. Debido a la relación causa-efecto entre las fallas simuladas y los hundimientos en tensión generados, los distintos clusters de hundimientos obtenidos serán una clasificación indirecta de las fallas que hace referencia al lugar y tipo de falla. Es decir, que las fallas (monofásicas, bifásicas y trifásicas) que ocurren en distintos puntos del sistema y que generan perfiles de hundimientos en tensión similares, serán agrupadas en un mismo cluster. Por lo tanto, el cluster también se puede interpretar como la agrupación de zonas de fallas (y tipos de falla) con el mismo impacto.

En resumen se realizará una clasificación de las fallas (lugar y tipo), de acuerdo con su impacto en las barras (hundimientos en tensión).

Una vez se tiene organizada la matriz de hundimientos en tensión, se debe establecer el número de clusters en los cuales se clasificará dicha información. Para esto se propone ejecutar el algoritmo kmeans haciendo un barrido en el número de grupos (k), para posteriormente identificar el mejor valor para k. Sin embargo, existen problemas relacionados con la dimensión de la matriz simulada (6355 filas x 144 columnas) los cuales se enumeran a continuación.

En cuanto al algoritmo kmeans:

- Los algoritmos iterativos de clasificación de información como kmeans generalmente son eficientes para problemas de dimensión reducida pero se vuelven ineficientes en problemas de gran escala.
- Es necesario aumentar el número de simulaciones por lo tanto el algoritmo kmeans emplea grandes cantidades de tiempo para optimizar la agrupación.
- Debido a la alta dimensionalidad, las posibles ubicaciones iniciales aleatorias de los centroides aumenta exponencialmente, por lo que la probabilidad de quedarse en un máximo local también aumenta.

En cuanto a la matriz de hundimientos:

- Es posible que se encuentre información redundante en las 144 variables de la matriz de hundimientos.
- Es posible que exista información irrelevante que no ofrece capacidad de discriminación que ayude a definir los grupos.
- Es posible que la información se encuentre segmentada, es decir que una información interesante puede estar distribuida en varios atributos.

Teniendo en cuenta lo anterior, para realizar la agrupación de hundimientos en tensión es necesario reducir la dimensionalidad de la matriz de entrada. En la siguiente sección se explica detalladamente el análisis de componentes principales para este propósito.

4.3. Reducción de dimensionalidad por medio de análisis de componentes principales

Uno de los mayores inconvenientes en el análisis de datos cuando se tiene una gran cantidad de variables, es la dimensionalidad del problema. Sin embargo, si se logran describir de manera precisa los valores de p variables por un conjunto de r variables (donde $r < p$), se habrá reducido dicha dimensionalidad a costa de una pequeña pérdida de información [Meléndez J. 2007].

La idea anterior se emplea en la técnica conocida como *Análisis de Componentes Principales* (Hotelling 1933), la cual tiene como objetivo representar adecuadamente información expresada como n observaciones y p en variables, con un menor número r de variables construidas como combinaciones lineales de las originales. De esta forma, se pueden identificar las posibles variables más relevantes del problema, y transformar las variables originales (generalmente correlacionadas), en variables no correlacionadas, facilitando de esta forma la interpretación de las mismas.

4.3.1. Interpretación de las componentes principales

Sea una matriz \mathbf{X} de $[n] \times [p]$ (n datos y p variables). Se desea encontrar un subespacio de menor dimensión que p , tal que al proyectar sobre él cada una de las observaciones, éstas conserven su estructura con la menor distorsión posible. Un ejemplo ilustrativo se puede ver en la figura 4.6, en la cual se desea encontrar la componente fundamental (subespacio de dimensión uno) de un caso con 2 variables ($p = 2$).

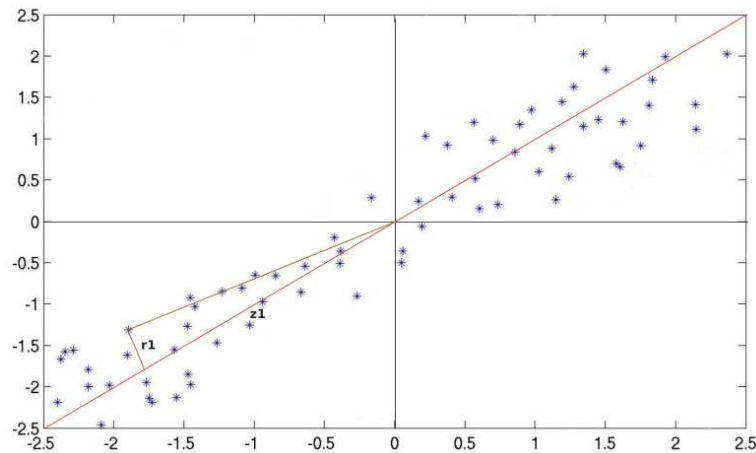


FIGURA 4.6: Ejemplo de componentes principales

La recta roja dibujada en la figura 4.6 puede transformar el problema de las dos variables en solo una, ya que ésta pasa cerca de todos los datos y a su vez la distancia entre éstos se

mantiene de manera aproximada cuando se proyectan dichos datos sobre la recta.

Para determinar dicha recta (que es la que mejor representa los datos) se debe cumplir la siguiente condición:

La sumatoria de las distancias entre los puntos originales y sus proyecciones sobre la recta debe ser lo mínimo posible.

Para explicar lo anterior, se toma un dato X_i y una dirección a_1 de norma uno, donde la proyección de X_i sobre a_1 es el escalar:

$$ProyX = \frac{\widehat{X}_i \cdot \widehat{a}_1}{|\widehat{a}_1|} = z_i \quad (4.1)$$

Denominando $z_i a_1$ como el vector que representa la proyección sobre la recta y r_i como la distancia entre X_i y la recta, el objetivo consiste en minimizar el cuadrado de la suma de las distancias r_i :

$$\text{Minimizar } \sum_{i=1}^n r_i^2 = \sum_{i=1}^n |\mathbf{x}_i - z_i \cdot \mathbf{a}_1|^2 \quad (4.2)$$

Se puede observar en la figura 4.6, que al proyectar cada punto sobre la recta se forma un triángulo rectángulo. Por el teorema de Pitágoras, se puede escribir:

$$X_i^2 = z_i^2 + r_i^2 \quad (4.3)$$

$$X_i' \cdot X_i = z_i^2 + r_i^2 \quad (4.4)$$

realizando la suma de todos los puntos $i=1 \dots n$

$$\sum_{i=1}^n X_i' \cdot X_i = \sum_{i=1}^n z_i^2 + \sum_{i=1}^n r_i^2 \quad (4.5)$$

Ya que el primer término de la ecuación 4.5 es constante, minimizar $\sum r_i^2$, equivale a maximizar $\sum z_i^2$, es decir la suma al cuadrado de las proyecciones. Esto a su vez equivale a maximizar la *varianza* [Peña D. 2002].

De acuerdo con lo anterior, podemos ver en la figura 4.6, que la mejor recta para representar los datos en una sola dimensión es aquella en la cual se maximizan las varianzas de los datos. Este criterio se puede extender a n dimensiones en el espacio, donde $n < p$.

El análisis para hallar las componentes principales de una matriz de datos a partir de las matrices de covarianza está implementado en Matlab mediante la función *princomp*. Esta función devuelve como resultado las componentes principales de la matriz de entrada y un vector que contiene los valores propios de la matriz de covarianza.

4.3.2. Principales componentes de los hundimientos en tensión simulados

Debido a que la matriz de hundimientos en tensión cuenta con 6355 observaciones de 144 variables, se realiza el análisis de componentes principales por medio de la función de Matlab *princomp*. El resultado de esto es una matriz de 144 componentes distintas organizadas de tal manera que las primeras componentes tienen la mayor varianza, es decir, las primeras componentes tienen la mayor información de la matriz de entrada.

Para saber cuántas componentes serán necesarias para tener el máximo de información, se calcula el porcentaje de varianza para las (n) primeras componentes principales mediante la siguiente ecuación:

$$PS_n = \frac{\sum_{i=1}^n S_i}{\sum_{i=1}^N S} * 100 \% \tag{4.6}$$

Donde:

PS es porcentaje de varianza para (n) componentes.

S_n es el valor propio (n) de la matriz de covarianza.

N es el número total de componentes.

n es el número de componentes.

El resultado de la ecuación 4.6 se muestra en la tabla 4.1 para las 6 primeras componentes principales.

TABLA 4.1: Porcentaje de varianza de las componentes principales

Número de componentes	Porcentaje de Varianza (%)
1	64,34
2	79,62
3	84,66
4	87,13
5	89,45
6	91,29

En la tabla 4.1 se observa que las primeras cinco componentes principales representan el 89.4% de la varianza de los datos originales. Finalmente se toman estas cinco componentes por lo que tenemos que la matriz inicial (6355 x 144) queda reducida a sus componentes principales (6355 x 5) sobre las cuales se les aplicará kmeans.

4.4. Implementación de kmeans e índices de agrupamiento

No hay un procedimiento específico para encontrar cuál es el número indicado de grupos en los cuales se puede clasificar la información [Davies D. 1979], sin embargo en la literatura se propone el uso de algunos indicadores que se expone a continuación.

1. *Cuadrado de la suma de las distancias entre las observaciones y los centroides*(R).

Este indicador muestra cómo varía la suma de las distancias entre los datos de cada uno de los clusters y sus centroides. A medida que la cantidad de clusters aumenta, la distancia entre los datos y los centroides disminuye. El objetivo es encontrar desde cual valor de (k) el indicador tiene poca variación. Es decir, el número apropiado de grupos esta relacionado con el codo de la curva tal como se muestra en la figura 4.7 donde tenemos $k=6$.

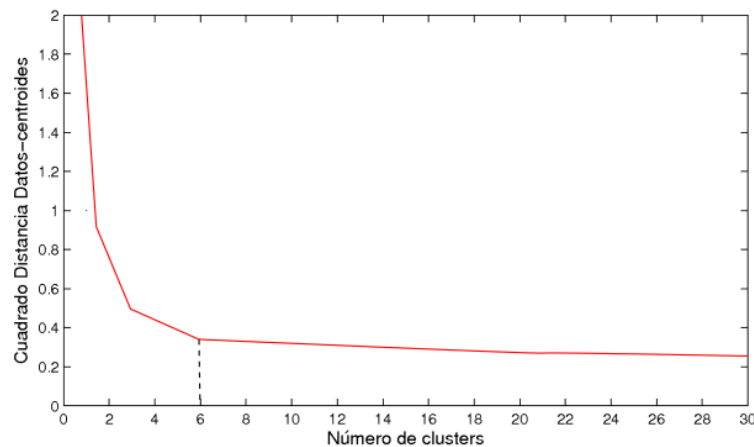


FIGURA 4.7: Índice cuadrado suma de las distancias datos-centroides

2. *Indicador Silueta*.

Este indicador asigna a cada dato un valor entre -1 y 1, que mide la similaridad de los datos en el mismo cluster comparándola con la similaridad de los datos de

otros clusters. Entre más cercano se encuentre el índice a 1, los datos están mejor agrupados. El indicador generalmente se resume como el promedio de los índices calculados para cada dato. En la gráfica 4.8 se muestra un ejemplo ilustrativo del indicador silueta para datos agrupados en 4 clusters.

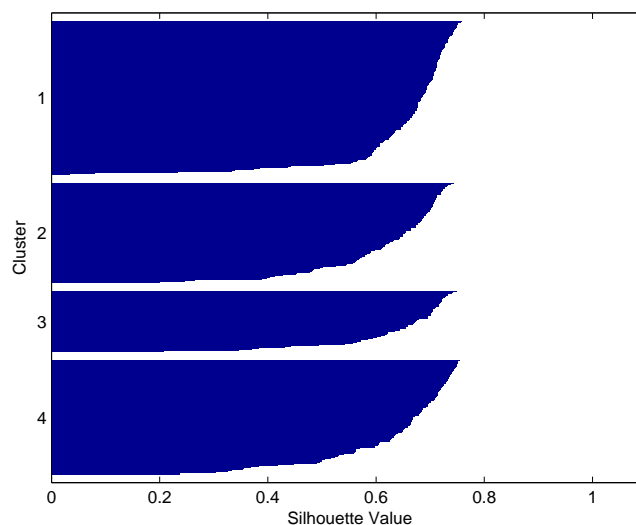


FIGURA 4.8: Indicador silueta para datos agrupados en 4 clusters

3. *Relación Inter-intra Cluster.*

En este indicador se calcula la proporción que hay entre el promedio de la distancia de los datos a los respectivos centroides y la mínima distancia que hay entre los centroides.

Para una agrupación adecuada, se tiene como objetivo maximizar el valor del indicador silueta; por otro lado el objetivo que se tiene para el indicador intra-inter clusters es minimizar su valor.

Teniendo las cinco componentes principales de la matriz de hundimientos en tensión simulados, se procede a ejecutar el algoritmo kmeans para diferentes cantidades de grupos (k) de entrada, calculando los indicadores explicados con el ánimo de hallar la cantidad apropiada de clusters para clasificar la información.

Los resultados de los indicadores para agrupaciones con valores de k entre 2 y 100 se muestran a continuación: