

## **Estrategia de regresión basada en el método de los $k$ vecinos más cercanos para la estimación de la distancia de falla en sistemas radiales**

### **$k$ -NN based regression strategy used to estimate the fault distance in radial power systems**

*Germán Morales España, Juan Mora Flórez\*, Hermann Vargas Torres*

Grupo de Investigación en Sistema de Energía Eléctrica (GISEL) y en Calidad de Energía Eléctrica y Estabilidad (ICE3), Universidad Tecnológica de Pereira, Programa de Ingeniería Eléctrica, La Julita, Pereira, Colombia

(Recibido el 27 de agosto de 2007. Aceptado el 9 de mayo de 2008)

#### **Resumen**

En este artículo se presenta una estrategia de regresión para estimación de la distancia de falla en sistemas de potencia radiales, empleando la técnica de los  $k$ -Vecinos más cercanos ( $k$ -NN). Esta propuesta de localización de fallas utiliza las medidas de la componente fundamental de tensión y de corriente disponibles en la subestación, no depende del modelo del sistema de potencia y se adapta a las características particulares de los sistemas radiales. El localizador de fallas se prueba en un sistema de distribución, donde se obtienen errores promedio inferiores al 3%, a diferentes distancias y resistencias de falla.

----- **Palabras clave:** Localización de fallas,  $k$ -vecinos más cercanos ( $k$ -NN), sistemas radiales, regresión.

#### **Abstract**

A regression strategy based on  $k$  nearest neighbors ( $k$ -NN) to estimate the fault distance in radial power systems is proposed. This fault location approach uses measurements of the fundamental components of voltage and current measured at the power substation. In addition, the approach is not constrained by the power system modeling and it is easily adaptable to the special characteristics of radial systems. The proposed fault locator is tested in a power distribution system and the obtained mean error is lower than 3%, by considering all fault types, several faulted nodes and fault resistances.

----- **Keywords:** Faults location,  $k$  nearest neighbors ( $k$ -NN), radial systems, regression.

---

\* Autor de correspondencia: telefax: + 57 + 6 321 1757; correo electrónico: jjmora@utp.edu.co (J. Mora-Flórez)

## Introducción

La continuidad y por tanto la calidad del servicio de energía eléctrica en los sistemas de transmisión y distribución se ve afectada por muchos factores, y las fallas en estos es uno de los más críticos. Ante una falla, los procesos de identificación y de localización son fundamentales, pues esto posibilita un rápido reestablecimiento del servicio [1]. De otra parte, las empresas suministradoras del servicio eléctrico están interesadas en mantener unos índices de continuidad, especialmente motivadas por conservar la confianza de sus clientes, así como para cumplir las reglamentaciones impuestas por el ente regulador. Para el caso Colombiano, la comisión de regulación de energía y gas (CREG), ha reglamentado los índices asociados a la duración y frecuencia de las interrupciones mediante el establecimiento del DES (Duración equivalente de las interrupciones), y el FES (Frecuencia equivalente de las interrupciones), para cada tipo de circuito [2, 3].

En los últimos años, algunas técnicas relevantes aplicables a sistemas de radiales han sido planteadas para la localización de fallas [4-13]. Estas técnicas emplean el cálculo de la impedancia equivalente vista desde la subestación durante la falla. Utilizando la componente fundamental de la corriente y tensión en prefalla y falla medidas en la subestación, estiman la sección de línea fallada con la comparación de la impedancia obtenida a partir del modelo impuesto por el método y la impedancia equivalente calculada. Como inconvenientes de estos métodos está la alta dependencia del modelo del sistema y que ignoran las características especiales de los sistemas radiales como los de distribución [3].

En este artículo se propone el uso de la técnica de aprendizaje supervisado de regresión, conocida como los  $k$  vecinos más cercanos ( $k$ -NN). Esta técnica se aplica a la estimación de la distancia de falla, considerando las características fundamentales de los sistemas radiales, sin depender del modelo del sistema. Inicialmente se presentan los fundamentos básicos de la técnica del vecino más cercano aplicada a la regresión. Luego se

discute la aplicación de los  $k$ -NN a la localización de fallas. Seguidamente se muestran y analizan resultados de la metodología propuesta en un sistema prototipo y finalmente, se presentan las conclusiones.

## Método de regresión basado en $k$ -n

### Aspectos básicos del método $k$ -NN

Un método de aproximación simple no paramétrica es el basado en la regla del vecino más cercano, que consiste en estimar el valor de un dato desconocido a partir de las características del dato más próximo, según una medida de similitud o distancia [14]. Esta regla tiene propiedades estadísticas bien establecidas y facilidad de aplicación a sistemas reales [15]. El método del vecino más cercano se puede extender utilizando no uno, sino un conjunto de datos más cercanos para predecir el valor de los nuevos datos, en lo que se conoce como los  $k$ -vecinos más cercanos ( $k$ -NN o  $k$ -Nearest Neighbors). Al considerar más de un vecino, se brinda inmunidad ante ruido y se suaviza la curva de estimación [16].

El método de los  $k$ -vecinos más cercanos se adapta fácilmente a la regresión de funciones con valores continuos [16]. El algoritmo asume que todos los datos pertenecen a  $R^p$ , y mediante una medida de distancia en ese espacio se determinan los  $k$  datos más cercanos al nuevo dato  $x_q$  para aproximar una función  $f: R^p \rightarrow R$  a partir de los  $k$  valores ya seleccionados. Esta función corresponde al promedio de los  $k$  valores más cercanos; si se considera el promedio aritmético (todos los datos dentro del grupo tienen igual relevancia), la función aproximación tiene la forma que se presenta en (1).

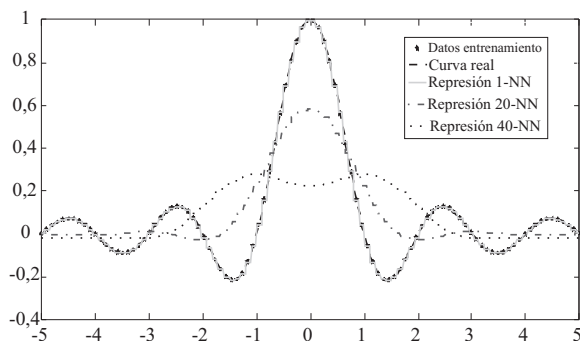
$$\hat{f}(x_q) \leftarrow \frac{1}{k} \sum_{i=1}^k f(x_i) \quad (1)$$

Todos los datos deben estar normalizados, para evitar que las características en el conjunto de entrada con valores más altos dominen el cálculo de la distancia [14].

### Ejemplo de análisis

Como análisis, se presenta la construcción de la curva de regresión de la función  $\text{sinc}(x)$ , con 100 datos linealmente espaciados en el intervalo  $[-5,5]$ . La medida de distancia utilizada es la euclídea clásica [16]. En la línea continua de forma escalonada de la figura 1 se muestra la curva de regresión utilizando un vecino mas cercano (1-NN).

Cuando se aplica el método de  $k$ -NN con promedio aritmético como se muestra en la curva a trazos de la figura 1 para  $k=20$ , se observa una curva más suavizada, sin embargo la aparente mejor regresión se descompensa con la tendencia lineal de la curva a medida que  $k$  aumenta, como se observa en la curva para  $k=40$  en la figura 1. Lo anterior se debe a que el valor calculado ante un nuevo dato corresponde al valor medio de los  $k$  datos más cercanos.



**Figura 1** Regresión de la función  $\text{sinc}(x)$  con el método del vecino más cercano, con  $k=1$ ,  $k=20$  y  $k=40$

Debido a que el método de  $k$ -NN se basa en la distancia, su desempeño se mejora al considerar un promedio ponderado, que da mayor importancia a los datos más cercanos al nuevo ejemplar, tal como se presenta en (2).

$$\hat{f}(x_q) \leftarrow \frac{\sum_{i=1}^k w_i \cdot f(x_i)}{\sum_{i=1}^k w_i} \quad (2)$$

$$w_i \equiv \text{KER}(d(x_q, x_i)) \quad (3)$$

El peso  $w_i$  que se muestra en (2) está definido en (3), donde  $\text{KER}$  es una función *Kernel* que determina la ponderación de cada punto basado en la distancia al punto de referencia dada por la función  $d$ . La función *Kernel* corresponde al peso que se le da a los datos para obtener un promedio ponderado, por lo tanto debe variar inversamente con la distancia para que los puntos más cercanos tengan mayor peso (mayor importancia).

Con la ponderación por distancia, se soluciona el problema de linealización que se presenta al considerar valores grandes de  $k$  en la predicción del nuevo valor, ya que los datos más lejanos tendrán poco efecto sobre el valor a predecir. Utilizando una ponderación igual al cuadrado del inverso de la distancia, se obtiene (4).

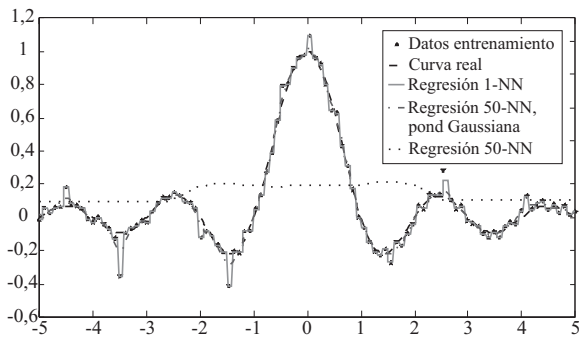
$$w_i \equiv d(x_q, x_i)^{-2} \quad (4)$$

Para el caso  $x_i = x_q$ , que resulta en un denominador cero para  $w$ , se asigna el valor de  $f(x_i)$  a  $\hat{f}(x_q)$ .

Sin embargo, pese a que se usa un *Kernel*, se tiene un sobre entrenamiento, ya que la curva siempre pasa por los valores de todos los datos, haciendo la curva vulnerable al ruido. Como solución a este problema se consideran funciones *Kernel* que a distancia cero la ponderación sea finita [14]. Un ejemplo de este tipo de ponderaciones y la más utilizada es el *Kernel* Gaussiano definido en (5).

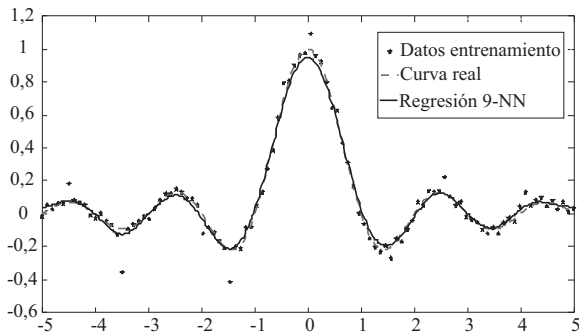
$$w_i \equiv e^{-d(x_q, x_i)^2} \quad (5)$$

La aplicación del *Kernel* Gaussiano para la regresión empleando 50-NN se presenta en la figura 2. Los datos tienen ruido de distribución normal con desviación estándar del 3% de la amplitud de la señal, y la característica de que cada 5 datos el ruido posee mayor intensidad (desviación estándar del 10%). La aplicación del *Kernel* Gaussiano mejora la curva de regresión, atenuando los cambios bruscos de la ponderación según (4). Aunque la curva de regresión usando *Kernel* Gaussiano tiene un alto error con los datos de entrenamiento, su comportamiento es fiel a la curva real, no se afecta por el ruido en los datos y conserva su tendencia.



**Figura 2** Regresión de la función  $\text{sinc}(x)$  con  $k=1$  (1-NN) y  $k=50$  (50-NN) sin ponderación y con ponderación, utilizando 100 datos con ruido

El valor  $k$  debe ser elegido de tal manera que la curva de regresión sea lo más similar posible a la original, lo cual se logra aplicando validación cruzada en la escogencia de  $k$  [17]. Con validación cruzada, se obtiene  $k=9$  y la curva resultante se muestra en la figura 3, donde se observa que esta curva es más similar a la curva real que la obtenida con  $k=50$  y mostrada en la figura 2.



**Figura 3** Regresión de la función  $\text{sinc}(x)$  con 9 vecinos más cercanos (9-NN) usando ponderación y 100 datos con ruido

### ***K-nn como localizador de fallas***

#### ***Método propuesto***

La relación entre las mediciones de tensión y corriente con la distancia de falla en un sistema radial es altamente compleja, debido principalmente a: a) la monitorización de tensión y co-

rriente únicamente en la cabecera del circuito; b) el efecto de la resistencia de falla; c) el desbalance del sistema debido a laterales monofásicos y bifásicos y carga desbalanceada; d) las redes no homogéneas, debido a que generalmente se presentan distintos calibres de conductor; y e) el desbalance en la impedancia de la línea debido a que éstas típicamente no se transponen, entre otros. En algunos sistemas radiales es posible que no se cumplan algunas de las restricciones mencionadas, pero se consideran debido a que son propias de los sistemas de distribución y estos son típicamente radiales.

El método propuesto para la determinación de la distancia desde la subestación hasta el nodo de falla, está diseñado a partir de la técnica de regresión para construir una función ( $f$ ) no lineal de la forma presentada en (6).

$$y = f(V_{a,b,c}, I_{a,b,c}) \quad (6)$$

Donde  $y$  es la distancia a la falla,  $V_{a,b,c}$  son las magnitudes de las tensiones de las fases a, b y c en el estado estable de la falla,  $I_{a,b,c}$  son las magnitudes de las corrientes de las fases a, b y c en el estado estable de la falla, y  $f$  es la función que relaciona las medidas en la cabecera del circuito con la distancia a la falla. Se puede disminuir la complejidad de la función al conocer el tipo de falla, entonces se tiene una función de la forma (7).

$$y = f(V_{a,b,c}, I_{a,b,c}, \text{tipo de falla}) \quad (7)$$

La función propuesta en este artículo es la presentada en (7), debido a que determinar el tipo de falla utilizando redes de secuencia o con algoritmos de clasificación es un problema más sencillo [3, 9, 17]. A partir del tipo de falla, el localizador es más especializado y por tanto su desempeño es mejor.

#### ***Comparación del método propuesto con otros métodos algorítmicos***

En la tabla 1 se muestra una comparación del método propuesto con métodos comúnmente utilizados para localización de fallas en sistemas radiales. De la tabla 1 se tiene que PROP son las

características del método y LIB significa que se puede utilizar el modelo de línea o carga que se quiera. Cuando se utilizan simulaciones, la limitación del modelado la impone la herramienta de simulación de fallas.

Las características sombreadas de la tabla 1 indican los requerimientos mínimos que debe tener un método para poder aplicarlo a los sistemas radiales, en especial a los de distribución. En el

método [4] es el método más básico e ignora totalmente cualquier tipo de carga. Los métodos que cumplen con las características fundamentales son el método propuesto, el método de Aggarwal [8] y el método de Zhu [7]. Estos métodos a diferencia del método propuesto son altamente dependientes del modelo de las líneas y de las cargas, además que su implementación y costo computacional son más exigentes.

**Tabla 1** Comparación de los métodos para la localización de fallas

<b>Aspecto Analizado</b>	<b>PROP</b>	<b>[4]</b>	<b>[5]</b>	<b>[6]</b>	<b>[7]</b>	<b>[8]</b>	<b>[9]</b>	<b>[10]</b>	<b>[11]</b>	<b>[12]</b>	<b>[13]</b>
Componentes de fase	✓	✗	✗	✗	✓	✓	✗	✗	✗	✗	✓
Modelo de línea	LIB	Corta	Larga	Corta	Corta	Corta	Larga	Corta	Corta	Corta	Corta
Modelo de carga	LIB	✗	Z=f(V)	Z=cte	Z=f(V)	Z=cte	Z=f(V)	Z=cte	I=cte	Z=cte	Z=cte
No homogeneidad	✓	✗	✗	✓	✓	✓	✓	✗	✗	✓	✗
Sistemas desbalanceados	✓	✗	✗	✗	✓	✓	✗	✗	✗	✗	✓
Laterales	✓	✗	✗	✓	✓	✓	✓	✗	✗	✓	✗
Cargas intermedias	✓	✗	✓	✓	✓	✓	✓	✓	✗	✓	✓

### Metodología aplicada para la localización de fallas

La metodología desarrollada para obtener la distancia a la falla desde la subestación donde se miden las tensiones y corrientes, se presenta en esta sección.

#### Selección de los patrones de entrenamiento

Para construir la función (7), con el complemento de una herramienta de regresión, se realiza un aprendizaje supervisado a partir de ejemplos con medidas de todos las fallas localizados a distancia diferente y para varias resistencias de falla. Una de las ventajas de la metodología propuesta es que se puede simular según el modelo disponible, es decir se adapta la solución del problema al modelo y no se tiene que adaptar el modelo al método (como en la mayoría de los métodos de distancia).

En esta etapa se obtiene una base de datos con las señales de tensión y corriente vistas en la subestación durante distintos escenarios de falla. Los patrones de entrenamiento son las diferencias de las magnitudes de tensión y corriente medidas en el estado de prefalla y el estado estable de la falla.

#### Definición de las etiquetas

Las distancias entre los nodos de falla y la subestación son seleccionadas como etiquetas para los *k*-NN. Las etiquetas son los valores de *y* en la función (7) que se desea construir.

#### Escalamiento de datos

Se realiza un escalamiento simple de las medidas de tensión y corriente entre [-1,1]. Este paso es muy importante para evitar dificultades numéricas durante los cálculos. Si las tensiones se encuentran en por unidad (p.u.), automáticamente



se encuentran escaladas. Hay distintos tipos de escalamiento, pero en esta aplicación simplemente se divide por el máximo a cada valor.

### Entrenamiento de los $k$ -NN

Esta etapa es una de las más importantes, debido a que un mal entrenamiento conlleva a resultados no deseables, y en el peor de los casos a un sobreentrenamiento [15]. Como los  $k$ -NN necesitan la definición a priori del número de vecinos  $k$ , este parámetro se selecciona en la etapa de entrenamiento utilizando validación cruzada [17].

La validación cruzada consiste en dividir el bloque de datos de entrenamiento en  $n$  partes iguales. Luego, para un cierto valor de hiperparámetros, se utilizan  $n-1$  de las  $n$  partes para entrenar la herramienta y la parte restante para hallar el error de validación. Este proceso se realiza  $n$  veces, lo que permite usar todas las muestras para hallar un error de validación o de desempeño. Por último, se promedian los  $n$  valores de error de validación encontrados para obtener un valor asociado al parámetro en uso. Éste procedimiento se repite desde  $k=1$  hasta  $k=\text{número total de datos}$ , y se escoge el parámetro  $k$  que tenga un menor error de validación cruzada. De esta manera, se puede afirmar que el parámetro escogido por validación cruzada es el que tendrá mejor comportamiento ante datos desconocidos.

### Prueba del localizador de fallas

Es necesario realizar una prueba final con datos desconocidos (datos que no fueron utilizados en la etapa de entrenamiento), con el fin de obtener la precisión final del localizador. La cantidad de los datos de prueba deben ser superiores al 20% de la base total de datos (entrenamiento y prueba), para tener una medida confiable del error [16]. Se define el error en proporción a la longitud total de la línea como se muestra en (8).

$$\%error = \frac{Long_{est} - Long_{real}}{Long_{total}} \times 100 \quad (8)$$

Donde  $Long_{est}$  es la distancia estimada por el método,  $Long_{real}$  es la distancia real al punto de fa-

lla y  $Long_{total}$  es la longitud total del ramal donde está ubicada la falla.

## Pruebas y resultados

### Sistema de pruebas

Los algoritmos desarrollados se prueban con el circuito de distribución SaskPower de 25 kV de Saskatchewan (Canadá), presentado en [9]. El diagrama unifilar del sistema se muestra en la figura 4. Este circuito ha sido utilizado para pruebas de distintos métodos de localización de fallas [3, 9, 17, 18].

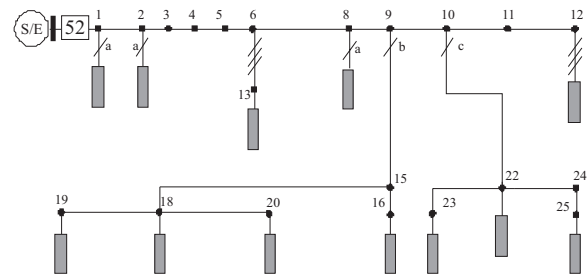


Figura 4 Circuito radial de distribución SaskPower de 25 kV de Saskatchewan (Canadá)

En el caso de múltiples ramales, la distancia debe ser estimada por cada ramal que hay en el sistema, por tanto se tendrán múltiples estimaciones de la falla. El problema de la múltiple estimación es una falencia de todos los métodos de distancia y algunas propuestas de solución con ayuda de herramientas de clasificación se presentan en [3, 17, 19].

### Descripción de las pruebas

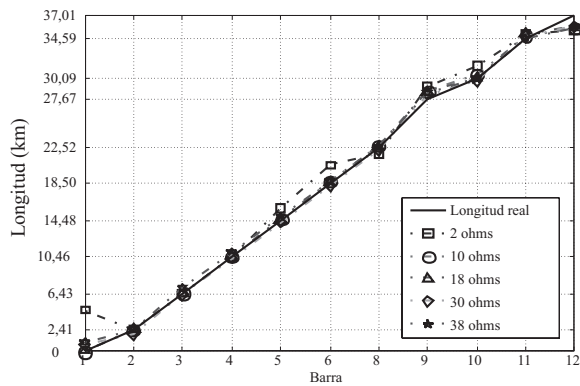
Para las pruebas se hacen fallas en cada nodo del circuito mostrado en la figura 4, con 21 resistencias de falla que varían desde  $0\Omega$  hasta  $40\Omega$  cada  $2\Omega$  [20]. Para el entrenamiento se utilizan los datos correspondientes a las fallas de todos los nodos y 11 resistencias de falla desde  $0\Omega$  hasta  $40\Omega$  cada  $4\Omega$ . Los datos restantes de resistencias de falla se utilizan para la etapa de prueba. De acuerdo a lo propuesto, se tienen 121 datos de

entrenamiento y 110 datos de prueba para cada tipo de falla (monofásica, bifásica y trifásica) en el ramal principal (se considera ramal principal entre las barras 1 y 12 debido a que es el único ramal trifásico, además es el que transporte la mayor cantidad de potencia). Se cuenta para cada tipo de falla el 52,38% de datos de entrenamiento y el 47,62% de datos de prueba.

### Resultados de las pruebas

Los resultados de la localización de falla en ramal principal (nodos 1 al 12), que tiene una longitud de 37,013 km, ante falla a-t se presentan en esta sección. Para fallas monofásicas en las otras fases el desempeño es similar.

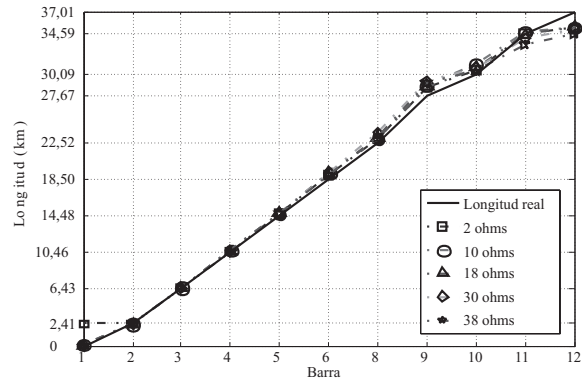
En la figura 5 se muestran las curvas con la distancia real y las estimadas con el método propuesto, para cinco valores de resistencia de falla, considerando únicamente los registros de tensión como variable de entrada del algoritmo de regresión. En las figuras 6 y 7 se presentan las curvas en caso de usar como entrada sólo la corriente y la corriente con la tensión, respectivamente.



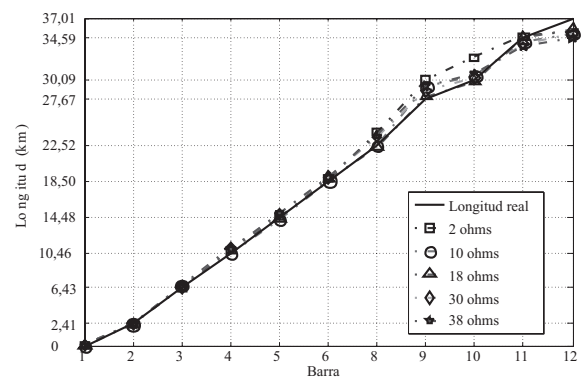
**Figura 5** Estimación de la distancia a la falla a-t en el alimentador principal, utilizando la tensión

A partir de las graficas 5, 6 y 7 y utilizando la ecuación (8) se calculan los errores en la estimación de la distancia, lo cual permite una mejor

visualización del desempeño del localizador, tal como se presenta en las figuras 8, 9 y 10. En general, se observa que los errores de localización en promedio son inferiores al 3% y el máximo es para el primer nodo en el caso de usar solo tensión, con un error aproximado del 12%.

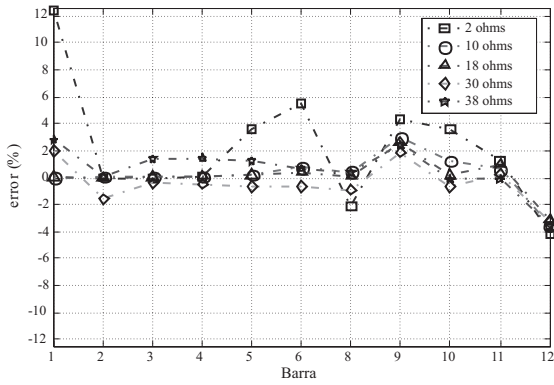


**Figura 6** Estimación de la distancia a la falla a-t en el alimentador principal, utilizando la corriente.

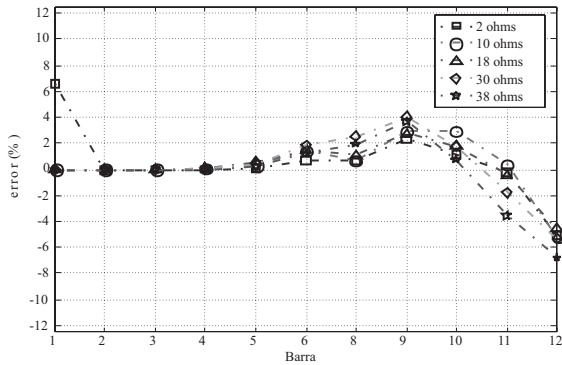


**Figuras 7** Estimación de la distancia a la falla a-t en el alimentador principal, utilizando la tensión y la corriente.

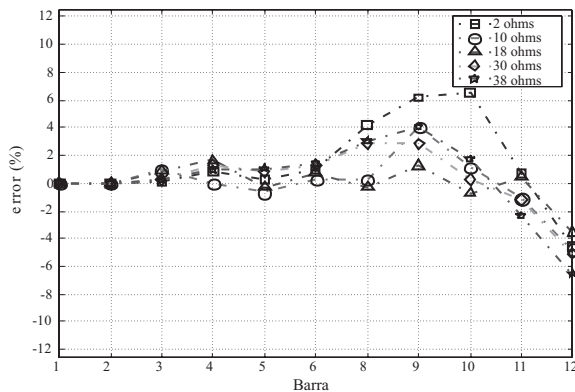
Como la técnica para ajustar el entrenamiento es la validación cruzada, el error obtenido con esta técnica y el error en la prueba presentado en las figuras 8, 9 y 10, se pueden tomar como índices para valorar el desempeño de la herramienta de regresión como localizador de fallas.



**Figura 8** Error en la estimación de la distancia a la falla a-t en el alimentador principal, utilizando la tensión



**Figura 9** Error en la estimación de la distancia a la falla a-t en el alimentador principal, utilizando la corriente



**Figura 10** Error en la estimación de la distancia a la falla a-t en el alimentador principal, utilizando la tensión y la corriente.

En la tabla 2 se presentan los promedios aritméticos del error de validación cruzada y el error de prueba del localizador obtenidos en este ejemplo de aplicación.

**Tabla 2** Promedios de los errores de validación cruzada (VC) y de prueba del localizador de fallas

Señales	% Error VC	% Error prueba
$V_{a,b,c}$	1,1584	1,4831
$I_{a,b,c}$	1,1092	1,4376
$V_{a,b,c}, I_{a,b,c}$	1,2457	1,5425
Promedio	1,1711	1,4877

Se observa que el comportamiento del error de validación cruzada tiene un comportamiento similar al error de prueba de las distintas señales. Según la tabla 2, la señal de mejor funcionamiento considerando el error de prueba y de validación cruzada es la corriente, de igual manera la señal de menor desempeño es la corriente y la tensión. Es importante resaltar el hecho que este método tiene un buen desempeño únicamente con señales de tensión, aspecto que lo hace muy útil, puesto que en muchas subestaciones sólo se cuenta con estas medidas.

## Conclusiones

En este artículo se presenta un método de localización de fallas en sistemas radiales empleando  $k$ -NN, que utiliza únicamente las medidas de tensión y corriente en la cabecera del circuito, obtenidas de registros de fallas ocurridas a lo largo del circuito de prueba y ante distintos escenarios de resistencia de falla. Para el entrenamiento del localizador de fallas se utilizaron aproximadamente el 53%, mientras que para la prueba se usaron el 47% del total de los registros. Los errores promedio son inferiores al 3% en la localización de fallas para el circuito analizado.

El método propuesto, a pesar de estar fundamentado en una técnica sencilla como los  $k$ -NN, presenta un alto desempeño, comparable con los



métodos algorítmicos clásicos o basados en la reactancia aparente de falla vista desde la subestación, y con una ventaja adicional asociada a la flexibilidad en la escogencia del modelo del sistema de potencia. La metodología de localización de fallas aquí analizada es de bajo costo económico y de fácil implementación. Su aplicación permite reducir considerablemente el tiempo de restauración del sistema, aumentando así los índices de continuidad del servicio de energía eléctrica y reduciendo, por consiguiente, los pagos por compensación debida a la deficiente calidad del servicio.

### Referencias

1. IEEE Std 37.114. "IEEE Guide for Determining Fault Location on AC Transmission and Distribution Lines". *Power System Relaying Comité*. 2004.
2. Comisión de Regulación de Energía y Gas CREG (CREG 1998-2002). *Resoluciones CREG 070 de 1998, CREG 096 de 2000, CREG 084 de 2002, CREG 084*. 2002.
3. J. Mora, G. Carrillo, B. Barrera. "Fault Location in Power Distribution Systems Using a Learning Algorithm for Multivariable Data Analysis". *IEEE Transaction on Power Delivery*, Vol. 22. 2007. pp. 1715-1721.
4. A. Warrington, C. Van. *Protective relays. Their theory and practice*. Chapman and Hall, London. 1968. pp. 125-146.
5. K. Srinivasan, A. St-Jacques. "A new fault location algorithm for radial transmission lines with loads" *IEEE Transactions on Power Delivery*. Vol. 4. 1989. pp. 1676-1682.
6. A. Girgis, C. Fallon, D. Lubkeman. "A fault location technique for rural distribution feeders" *IEEE Transactions on Industry and Applications*. Vol. 26. 1993. pp. 1170-1175.
7. J. Zhu, D. Lubkeman, A. Girgis. "Automated fault location and diagnosis on electric power distribution feeders" *IEEE Transactions on Power Delivery*. Vol. 12. 1997. pp. 801-809.
8. R. Aggarwal, Y. Aslan, A. Johns. "New concept in fault location for overhead distribution systems using superimposed components" *IEE Proceedings. Generation, Transmission and Distribution*. Vol. 144. 1997. pp. 309-316.
9. R. Das. *Determining the locations of faults in distribution systems*. Ph.D. dissertation, University of Saskatchewan. Saskatoon, Canadá. 1998. pp. 16-73
10. D. Novosel, D. Hart, Y. Hu, J. Myllymaki, System for locating faults and estimating fault resistance in distribution networks with tapped loads 1998. US Patent number 5,839,093.
11. L. Yang. One terminal fault location system that corrects for fault resistance effects 1998. US Patent number 5,773,980.
12. M. Saha, E. Rosolowski, Method and device of fault location for distribution networks 2002. US Patent number 6,483,435.
13. M. Choi, S. Lee, D. Lee, B. Jin. "A new fault location algorithm using direct circuit analysis for distribution systems." *IEEE Transactions on Power Delivery*. Vol. 19. 2004. pp. 35-41.
14. T. Cover, P. Hart. "Nearest neighbor pattern classification". *IEEE Transactions on Information Theory*. Vol. 13. 1967. pp. 21-27
15. D. Aha, D. Kibler, M. Albert. "Instance-based learning algorithms". *Machine Learning*. 1991, N.º 6. pp. 37-66
16. F. Moreno. *Clasificadores eficaces basados en algoritmos rápidos de búsqueda del vecino más cercano*. Ph.D. dissertation, Universidad de Alicante. Departamento de lenguajes y sistemas informáticos. 2004. pp. 56-89
17. G. Morales, A. Gómez. *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, Universidad Industrial de Santander, Colombia. 2005. <http://tangara.uis.edu.co/biblioweb/pags/cat/popup/derautor.jsp?parametros=118738>. Consultada Mayo de 2007.
18. G. Morales, H. Vargas, J. Mora. "Impedance based method to fault location in power distribution, considering tapped loads and heavy unbalanced systems" *Proc. XII encuentro regional Iberoamericano del CI-GRÉ*. Foz de Iguazú. 2007. pp. 52-61
19. G. Morales, G. Carrillo, J. Mora. "Selección de descriptores de tensión para localización de fallas en redes de distribución de energía" *Revista Ingeniería*. Vol. 11. 2006. pp. 43-50.
20. J. B. Dagenhart. "The 40- Ground-Fault Phenomenon" *IEEE Transactions on Industry Applications*. Vol. 36. 2000. pp 30-32.