

# Estimación del Tamaño de la Perturbación en una Microrred durante Eventos de Frecuencia usando Analítica de Datos Multivariantes

Mejía, César<sup>1,\*</sup>; Rocha, Alex<sup>1</sup>; Cepeda, Jaime<sup>1,2</sup>

<sup>1</sup>Escuela Politécnica Nacional, Facultad de Ingeniería Eléctrica y Electrónica, Quito, Ecuador

<sup>2</sup>Operador Nacional de Electricidad CENACE, Gerencia Nacional de Desarrollo Técnico, Quito, Ecuador

**Resumen:** Las microrredes, definidas como conjuntos de cargas interconectadas y recursos energéticos distribuidos, operan dentro de límites eléctricos claramente definidos y pueden funcionar de manera conectada o aislada respecto a la red principal. Estos sistemas enfrentan desafíos críticos de estabilidad de frecuencia, especialmente tras perturbaciones significativas como la pérdida de generación eléctrica, lo que es particularmente problemático en microrredes aisladas o sin un Esquema de Alivio de Carga (EAC) eficaz. Este trabajo presenta una novedosa metodología para estimar el tamaño de las perturbaciones durante eventos de frecuencia, la cual se integra en un EAC adaptativo. La metodología se aplica en la microrred de Galápagos Santa Cruz – Baltra, utilizando modelos de regresión lineal múltiple. Además, se complementa con un análisis de contingencias basado en simulaciones de Monte Carlo, obteniendo datos críticos como la frecuencia mínima (*Nadir*), la tasa de cambio de frecuencia (*RoCoF*), el voltaje de establecimiento (*Vend*) y la frecuencia de establecimiento (*fend*). La evaluación realizada en siete escenarios diferentes demuestra la alta efectividad de la regresión lineal múltiple para predecir el tamaño de las perturbaciones, lo que subraya su potencial para mejorar la estabilidad y confiabilidad de microrredes.

**Palabras clave:** Monte Carlo, regresión lineal múltiple, microrred, estabilidad

## Estimating the Size of the Disturbance in Frequency Events using Multivariate Data Analytics

**Abstract:** Microgrids, defined as interconnected loads and distributed energy resources operating within clearly defined electrical boundaries, can function either connected to or isolated from the main grid. These systems face critical challenges related to frequency stability, especially following significant disturbances such as the loss of electrical generation, which is particularly problematic in isolated microgrids or those lacking an effective Load Shedding Scheme (LSS). This work presents an innovative methodology for estimating the size of disturbances during frequency events, which is integrated into an adaptive LSS. The methodology is applied to the Galápagos Santa Cruz – Baltra microgrid, using multiple linear regression models. Additionally, it is complemented by a contingency analysis based on Monte Carlo simulations, obtaining critical data such as minimum frequency (*Nadir*), rate of change of frequency (*RoCoF*), settling voltage (*Vend*), and settling frequency (*fend*). The evaluation conducted across seven different scenarios demonstrates the high effectiveness of multiple linear regression in predicting the size of disturbances, that highlights its potential to enhance the stability and reliability of microgrids.

**Keywords:** Monte Carlo, multiple linear regression, microgrid, stability

### 1. INTRODUCCIÓN

Las microrredes, según la definición del IEEE Std. 2030-7 (2018), son sistemas compuestos por cargas interconectadas y recursos energéticos distribuidos que operan dentro de límites eléctricos definidos. Estas pueden funcionar como entidades controlables, ya sea conectadas o desconectadas de la red principal. Esta flexibilidad operativa permite a las microrredes

enfrentar desafíos significativos de estabilidad, especialmente en términos de frecuencia, un aspecto crítico en la gestión de perturbaciones en un Sistema Eléctrico de Potencia (SEP).

La estabilidad de frecuencia es la capacidad de un SEP para mantener un equilibrio tras una perturbación significativa, como la pérdida de generación. Dicho evento puede provocar un desbalance entre la generación y la carga, lo que lleva a una

\*cesar.mejia@epn.edu.ec

Recibido: 19/09/2023

Aceptado: 21/01/2024

Publicado en línea el 28/02/2025

10.33333/rp.vol55n1.02

CC 4.0

disminución de la frecuencia y, eventualmente, al colapso del sistema. Durante los primeros 30 segundos posteriores a la perturbación, se activa la Regulación Primaria de Frecuencia (RPF), que incluye respuestas electromagnéticas, inerciales y del regulador de velocidad (Sun et al., 2010). Si la RPF resulta insuficiente, es necesario implementar un Esquema de Alivio de Carga (EAC) para salvaguardar la estabilidad del sistema (Wiest et al., 2018).

Los EAC pueden clasificarse en tres tipos principales: tradicional, semi-adaptativo, y adaptativo. Los enfoques tradicionales, como el propuesto por Mohd Zin et al. (2004) para el sistema Tenaga Nasional Berhad (TNC), utilizan pasos de frecuencia predefinidos para desconectar cargas. Sin embargo, este enfoque fijo resulta ineficaz en escenarios con alta variabilidad de condiciones operativas. Los EAC semi-adaptativos, como el de Jiang et al. (2010), ajustan la cantidad de carga a desconectar en función de la tasa de cambio de frecuencia (*RoCoF*), lo que mejora la respuesta del sistema.

Por otro lado, los EAC adaptativos, como el propuesto por Mohamad et al. (2017) el cual usa la técnica de Optimización por Enjambre de Partículas (PSO), calculan la potencia óptima a desconectar, aunque no siempre consideran la dependencia de *DP* con otras variables. En esta línea, Pérez et al. (2016) diseñaron un EAC adaptativo para evitar colapsos de voltaje, utilizando PSO combinado con un índice de estabilidad de voltaje (SVSI). A pesar de estos avances, aún son necesarias técnicas más sofisticadas para garantizar la estabilidad integral del sistema.

Escobar (2021) diseñó un EAC adaptativo basado en voltaje y frecuencia, implementado en una plataforma de simulación digital en tiempo real. Este esquema utiliza un regresor para determinar la carga a desconectar en caso de contingencia, considerando voltaje, frecuencia y *RoCoF*. Como mejora, se propone la incorporación de la variable *Nadir*, debido a su alta correlación con el tamaño de la perturbación (*DP*), lo que podría aumentar la precisión del modelo. Asimismo, se recomienda la aplicación de técnicas de despacho económico para optimizar la operación de generación térmica, lo que garantiza una distribución eficiente de la carga en el sistema.

En este trabajo, se propone una metodología para estimar los valores del tamaño de la perturbación (*DP*) generados por múltiples eventos de frecuencia, utilizando modelos regresores basados en la evolución dinámica del sistema. El documento está estructurado en tres secciones principales: en la Sección 2, se describe la metodología empleada, abarcando la modelación dinámica del sistema, la generación de datos mediante simulaciones Monte Carlo y el análisis post-contingencia. En la Sección 3, se presentan los resultados, que incluyen el análisis descriptivo de los datos, la estimación de los regresores y la evaluación de la capacidad predictiva de los modelos para determinar *DP*.

## 2. METODOLOGÍA

En la Figura 1, se presenta un esquema con la metodología propuesta para estimar el tamaño de la perturbación, estructurada en tres fases: a) modelación del sistema eléctrico

de prueba, b) preparación y estructura de datos de Monte Carlo, y c) análisis post contingencia. A continuación, se detallará el estado del arte en cada una de las fases de la metodología propuesta.

### 2.1 Sistema eléctrico de prueba

El sistema de prueba consiste en la microrred del sistema eléctrico aislado Santa Cruz – Baltra. La modelación dinámica se realiza en el programa computacional PowerFactory. La microrred se encuentra compuesta por unidades de generación convencional (unidades térmicas) y fuentes de energía renovable no convencional (parque eólico y plantas fotovoltaicas). Desde el sistema de distribución compuesto por varios alimentadores se provee de energía a centros educativos, hospitalares, sector doméstico, comercial, industrial y otros.

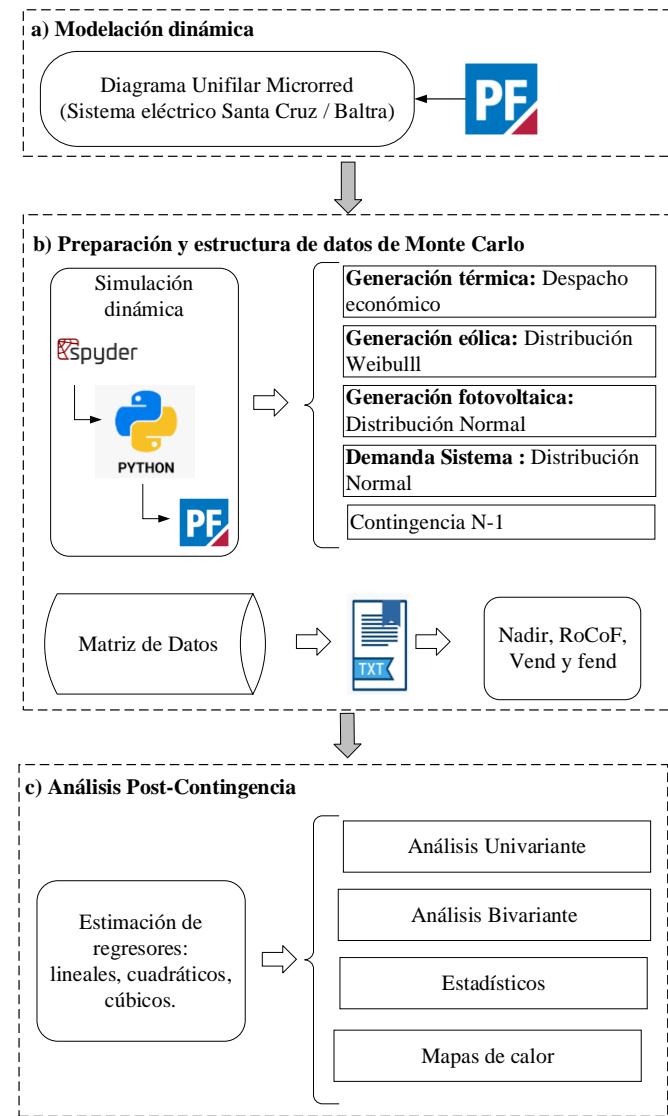


Figura 1. Esquema de la metodología

Se considera como generación convencional a la central ubicada en Santa Cruz que está formada por diez unidades de generación a diésel. Los generadores se encuentran subdivididos en tres grupos conectados a barras independientes. El grupo Hyundai 1 se encuentra integrado por dos generadores a diésel G08-HYU y G09-HYU de 1,702 MW

a 4,16 kV. El grupo Hyundai 2 se encuentra integrado por cuatro generadores a diésel G10-HYU, G11-HYU, G12-HYU, G13-HYU de 1,702 MW a 4,16 kV.

Finalmente, el grupo Caterpillar se encuentra formado por cuatro generadores a diésel G1-3512, G3-3512, G4-3512 y G6-3512 de 650 kW a 480 V. En la actualidad, los generadores Caterpillar no son usados como parte del despacho de generación.

Los controles de los generadores Hyundai se representaron mediante modelos y parámetros típicos. Los modelos de los reguladores consisten en: a) regulador de voltaje (modelo IEEE ESAC8B); y b) control de velocidad (modelo IEEE DEGOV1).

Las fuentes de energía renovable no convencional instaladas en el sistema eléctrico Santa Cruz - Baltra están formadas por: un parque eólico en la isla Baltra con tres aerogeneradores de 750 kW cada uno y una planta fotovoltaica de 200 kW, un banco de baterías para regulación de fluctuaciones viento/irradiación (baterías Ion-Litio) con capacidad de 500 kW y 400 kWh, un banco de baterías para almacenamiento de energía no usada (baterías de plomo ácido) con capacidad de 600 kW y 4 000 kWh; y finalmente, en la isla Santa Cruz se tiene una planta fotovoltaica de 1,5 MW.

Para la modelación de los generadores eólicos, se utiliza el modelo estándar de un aerogenerador tipo FRCWTG (en inglés *Fully Rated Converter Wind Turbine Generator*). La descripción detallada del modelo se encuentra en DIgSILENT GmbH (2022). El modelo se redimensiona para representar un aerogenerador de 750 kW con frecuencia nominal 60 Hz.

La modelación de la planta fotovoltaica de Santa Cruz de capacidad 1,5 MW utiliza el modelo estándar ElmPvsys, este modelo se encuentra disponible en PowerFactory. El modelo está basado en un generador estático. En general, el sistema fotovoltaico modela un conjunto de paneles fotovoltaicos conectados a la red a través de un único inversor. El detalle del modelo *ElmPvsys* se encuentra en la referencia DIgSILENT GmbH (2022).

## 2.2 Preparación y estructura de datos de Monte Carlo

Una vez completada la modelación dinámica del sistema de prueba, se implementó un algoritmo en Python para la generación de datos. Este algoritmo emplea el método de Monte Carlo para realizar análisis de contingencias N-1 mediante simulaciones de probabilidad múltiple.

Los pasos ejecutados por el algoritmo son los siguientes: i) cálculo de la demanda del sistema mediante distribución normal; ii) asignación de niveles de potencia en generadores convencionales (Hyundai) mediante despacho económico; iii) asignación de niveles de potencia en generadores convencionales (aerogeneradores y planta fotovoltaica) mediante distribución Weibull; iv) contingencia N-1 mediante la desconexión aleatoria de generación; y, v) simulación dinámica en el dominio del tiempo en PowerFactory.

### 2.2.1 Demanda del sistema

La demanda de potencia activa del sistema se determina mediante funciones de distribución normal o gaussiana, caracterizada por la Ecuación (1). La demanda de potencia reactiva se estimará basándose en el supuesto de que el factor de potencia mantiene un comportamiento casi invariable (Cepeda, 2013).

$$f(x, \mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1)$$

donde  $x$  es la variable,  $\mu$  es la media de la distribución,  $\sigma$  es la desviación estándar, y el cuadrado de la desviación estándar,  $\sigma^2$ , se denomina la varianza.

Con la Ecuación (1), se obtiene la matriz de la demanda aleatoria  $D_{2 \ 000 \times 1}$  formada por 2 000 observaciones. Para el cálculo de la demanda del sistema se utiliza una media  $\mu = 7,5 \text{ MW}$  y una desviación estándar  $\sigma = 0,5$ .

### 2.2.2 Asignación de potencia en unidades no convencionales

La función de distribución Weibull caracterizada por la Ecuación (2), ha sido adoptada para determinar la potencia de generación  $P_{eol}$  de los generadores eólicos (Mejía et al., 2024).

$$f(v) = \frac{k}{c} \left(\frac{v}{c}\right)^{k-1} e^{-\left(\frac{v_{prom}}{c}\right)^k}; k > 0, v > 0, c > 1 \quad (2)$$

Donde  $k$  y  $c$  constituyen los parámetros forma y escala, respectivamente, y, la variable  $v_{prom}$  representa la velocidad media o promedio del viento. El parámetro de forma  $k$  no tiene unidades y se puede obtener mediante la Ecuación (3).

$$K = \left(\frac{\sigma}{v_{prom}}\right)^{-1,086}; 1 \leq k \leq 10 \quad (3)$$

Donde  $\sigma$  es la desviación estándar obtenida a partir de la velocidad del viento. El factor de escala  $c$  tiene las mismas unidades de la velocidad del viento (m/s) y está dada por la Ecuación (4).

$$c = \frac{v_{prom}}{v_{prom} \left(1 + \frac{1}{k}\right)} \quad (4)$$

Para la planta fotovoltaica de 1,5 MW de Santa Cruz, se asume que, durante las horas de actividad de la planta, la producción se comporta como un proceso estocástico de acuerdo con una función de distribución Normal.

### 2.2.3 Asignación de potencia en unidades convencionales

El despacho económico es utilizado para determinar la potencia activa en los generadores convencionales (Hyundai). Debido a la existencia de generadores eólicos y fotovoltaicos, la potencia para el despacho económico en las unidades convencionales  $P_{UC}$ , se actualiza en cada paso de la simulación de Monte Carlo. La potencia  $P_{UC}$  se calcula mediante la diferencia entre la demanda total del sistema  $P_d$  con la potencia

de la central eólica  $P_{eol}$  y de la planta fotovoltaica  $P_{phv}$ , tal como se detalla en la Ecuación (5).

$$P_{UC} = P_d - (P_{eol} + P_{phv}) \quad (5)$$

Según Chowdhury y Rahman (1990), el despacho económico se define como el proceso de asignar niveles de potencia activa en las unidades de generación, de esta manera, la carga del sistema puede ser cubierta en su totalidad y de la forma más económica. El objetivo básico de la función de despacho económico es programar las salidas de las unidades de generación de combustible fósil en línea para satisfacer la carga del sistema al menor costo (Bakirtzis, 1994).

En un sistema de  $N$  generadores, el costo de combustible de cada unidad se representa como  $F_i$  y la potencia de salida como  $P_i$ . El costo total de combustible del sistema se representa como  $F_T$  que corresponde a la suma del costo de cada unidad individual.

$$\begin{aligned} F_T &= F_1 + F_2 + \dots + F_N \\ &\approx \sum_{i=1}^N F_i(P_i) \end{aligned} \quad (6)$$

Por tanto, la función objetivo a minimizar se muestra en la Ecuación (7), donde  $F$  es la tasa de costo de cada generador que entrega una potencia  $P_i$ .

$$f_o = \sum_{i=1}^N F(P_i) \quad (7)$$

La Ecuación (7) se puede expresar de la siguiente manera:

$$F(P_i) = a + bP_i + cP_i^2 \quad (8)$$

donde  $a, b, c$ , son los coeficientes de costo de combustible de cada generador. La restricción que se debe cumplir es que la suma de la potencia que entrega cada unidad de generación  $P_i$  debe ser igual a la potencia de demanda  $P_d$  que se está abasteciendo (Guo-Li Zhang et al., 2005). Para propósitos de este trabajo, no se considera las pérdidas del sistema, por tanto, el balance de potencia estará dado por la Ecuación (9) donde  $\xi$  es el error.

$$\sum_{i=1}^N P_i - P_d = \xi = 0 \quad (9)$$

Los límites de generación se encuentran dados en la Ecuación (10), donde  $P_i^{\min}$  y  $P_i^{\max}$  representan las potencias de salida mínimos y máximos de cada unidad de generación.

$$P_i^{\min} \leq P_i \leq P_i^{\max}. \quad (10)$$

Según Guo-Li Zhang et al. (2005), para resolver el problema de despacho económico de un sistema totalmente térmico y sin tener en cuenta las pérdidas, se recomienda utilizar el método de iteración Lambda. En este método, se asumen una tasa del costo incremental  $\lambda_i$  y se encuentra la potencia  $P_i$ . Con los dos primeros resultados, se extraña o interpola hasta disminuir el

ancho del intervalo donde se encuentra la solución (Muñoz & Quezada, 2015).

El costo incremental  $\lambda_i$  se caracteriza de acuerdo con la Ecuación (11) y corresponde a la derivada de la función de costo respecto a la potencia de salida de cada unidad de generación  $P_i$  como se muestra en la Ecuación (12).

$$\lambda = \frac{df_i}{P_i} \quad (11)$$

$$\lambda_i = b_i + 2c_i P_i \quad (12)$$

De la Ecuación (12), se obtiene la potencia de generación  $P_i$  mostrado en la Ecuación (13) y se calcula la variación de la potencia de cada unidad, como se expresa en la Ecuación (14).

$$P_i = \frac{\lambda_i - b_i}{2c_i} \quad (13)$$

$$\Delta P_i^{(1)} = P_D - \sum_{i=1}^N P_i^{(1)} = \xi \quad (14)$$

El costo incremental promedio  $\lambda_m$  se define en la Ecuación (15), donde  $\lambda_h$  es el costo incremental superior y  $\lambda_l$  el costo inferior.

$$\lambda_m = \frac{\lambda_h + \lambda_l}{2} \quad (15)$$

Con las expresiones anteriores, la potencia de generación  $P_i$  considerando el nuevo valor de  $\lambda_m$  se expresa como se muestra en la Ecuación (16).

$$P_i = \frac{\lambda_m - b_i}{2c_i} \quad (16)$$

A continuación, en la Ecuación (17) se muestra un arreglo para que la variación de potencia  $\Delta P_i$ , cumpla con una tolerancia, por ejemplo  $1,0 \times 10^{-4}$ .

$$\begin{aligned} P_i^{(1)} &= P_D - \sum_{i=1}^N P_i^{(1)} > \xi & \lambda_h = \lambda_m \\ P_i^{(1)} &= P_D - \sum_{i=1}^N P_i^{(1)} < \xi & \lambda_l = \lambda_m \end{aligned} \quad (17)$$

Con base en la teoría del despacho económico y al método de iteración Lambda, se desarrolla en Python un algoritmo para asignar niveles de generación en las unidades convencionales Hyundai.

#### 2.2.4 Contingencia N-1

Una vez definido el valor de la demanda del sistema y asignado los valores de potencia en los generadores convencionales y no convencionales, se procede a realizar la contingencia N-1. La contingencia que permite producir un desbalance entre la generación y carga es la salida aleatoria de generación, misma que está basada en una función de distribución normal.

#### 2.2.5 Simulación dinámica

En cada simulación de Monte Carlo se ejecuta una simulación dinámica en el programa PowerFactory durante un tiempo de 20 segundos. Para cada simulación, se obtienen los datos de *Nadir* (Hz), *RoCoF* (Hz/s), voltaje de establecimiento *Vend* (kV) y frecuencia de establecimiento *fend* (Hz). Este proceso se repite para un número *n* de simulaciones que varía dependiendo del número de iteraciones que se requiera, para este caso se consideraron 2 000 simulaciones. Los datos son almacenados en la matriz  $X_{2\,000 \times 4}$  que servirán para la estimación del regresor.

### 2.3 Análisis Post-Contingencia

Una vez obtenida la matriz de datos, se procede a su descripción mediante dos enfoques: a) análisis univariante y b) análisis multivariante. El análisis univariante se centra en la operación de cada vector de datos de manera individual, mientras que el análisis multivariante considera las relaciones geométricas dentro de la matriz de datos (Peña, 2002).

Para facilitar la comprensión, se presenta el esquema del *dataset* final, el cual distingue entre las variables características (*features*) y la variable objetivo (*target*). Las variables *features* incluyen *Nadir*, *RoCoF*, *Vend* y *fend*, todas de tipo numérico y continuo. Estas variables se emplean para predecir el tamaño de la perturbación (*DP*), la variable objetivo. Su naturaleza numérica y continua, tanto desde una perspectiva computacional como matemática, permite aplicar de manera efectiva técnicas de regresión lineal múltiple.

La descripción detallada del esquema del *dataset* empleado se encuentra en la Tabla 1, donde se destaca cómo este enfoque estructurado mejora la capacidad predictiva y la robustez del análisis.

**Tabla 1.** Descripción del esquema del dataset

Variable	Tipo de variable	Descripción
<i>Nadir</i>	Feature	Frecuencia mínima (Hz)
<i>RoCoF</i>	Feature	Tasa de cambio de frecuencia (Hz/s)
<i>Vend</i>	Feature	Voltaje de establecimiento (kV)
<i>fend</i>	Feature	Frecuencia de establecimiento (Hz)
<i>DP</i>	Target	Tamaño de la perturbación (MW)

#### 2.3.1 Análisis univariante

El análisis univariante permite comprender el comportamiento individual de cada variable, examinándola de manera aislada y explorando posibles relaciones entre ellas (Peña, 2002). Este análisis utiliza como principales estadísticos: a) media; b) mediana; c) desviación estándar; d) varianza; e) coeficiente de asimetría; f) coeficiente de curtosis; y g) coeficiente de variación.

#### 2.4 Análisis multivariante

El análisis multivariante tiene como objetivo principal analizar la estructura de dependencias entre las variables. Estas dependencias se analizan de diferentes maneras: a) entre pares de variables; b) entre una variable y el conjunto de las demás; c) entre pares de variables controlando el efecto de otras; y d) considerando todas las variables en conjunto (Peña, 2002).

Este enfoque permite identificar patrones complejos y relaciones estructurales en los datos.

### 2.5 Regresión lineal

La implementación de un modelo supervisado de regresión lineal múltiple exige una división estructurada del conjunto de datos en dos subconjuntos: i) un conjunto de entrenamiento, que representa el 80 % de los datos disponibles (1 600 observaciones), y ii) un conjunto de prueba, que constituye el 20 % restante (400 observaciones).

El modelo de regresión lineal múltiple se centra en establecer una relación cuantitativa entre la variable dependiente, el tamaño de la perturbación (*DP*), y las variables independientes *Nadir*, *RoCoF*, *Vend* y *fend*. Estas variables, al ser numéricas y continuas, favorecen la aplicación eficiente del modelo, cuyo objetivo principal es minimizar los errores al cuadrado mediante un ajuste óptimo de los coeficientes de regresión.

La evaluación del modelo se realiza utilizando el coeficiente de determinación ajustado ( $R^2$ ), que mide la proporción de la variabilidad explicada por el modelo, y considera el número de variables independientes. Un  $R^2$  ajustado más alto indica un ajuste más sólido y efectivo.

En este trabajo, el análisis de las dependencias en los *casos a* y *b* se basa en este modelo, que proporciona una herramienta robusta para explorar las interacciones entre las variables. Adicionalmente, las simulaciones dinámicas se llevaron a cabo utilizando el software PowerFactory de DIgSILENT, versión 2023, lo que permitió evaluar el comportamiento del sistema en el dominio del tiempo.

## 3. RESULTADOS

### 3.1 Análisis de datos

Esta sección se enfoca exclusivamente en el análisis y exploración preliminar de los datos, previo a la implementación de los modelos de regresión.

#### 3.1.1 Análisis descriptivo univariante

El análisis descriptivo de datos tiene la finalidad de examinar los datos previo a la aplicación de cualquier técnica estadística. A partir de los principales estadísticos que se presentan en la Tabla 2, es posible realizar un análisis general de los datos, el cual se resume a continuación:

- Tamaño de muestra: Cada variable cuenta con 2 000 observaciones. Las medias y medianas de las variables están cercanas entre sí, lo que sugiere la ausencia de valores extremos significativos.
- En relación con el coeficiente de Kurtosis, generalmente se observa Coeficiente de Kurtosis; los valores de Kurtosis para todas las variables son menores a 3, lo que indica que las distribuciones tienden a ser más compactas alrededor de su centro. Esto implica una menor

probabilidad de valores extremos en comparación con una distribución normal. Sin embargo, es importante considerar que podrían existir anomalías o atípicos debido a características específicas de los datos o factores externos que afecten la variabilidad.

- c) Coeficiente de variación: Las variables *DP* y *RoCoF* presentan valores similares, lo que indica que ambas distribuciones de datos tienen una estabilidad relativa parecida en términos de su dispersión. Aunque estas variables tienen diferentes medias y desviaciones estándar, el coeficiente de variación (CV) sugiere que la variabilidad relativa en relación con sus medias es comparable.

**Tabla 2.** Estadísticos del análisis univariante

	<i>DP</i> (MW)	<i>Nadir</i> (Hz)	<i>RoCoF</i> (Hz/s)	<i>Vend</i> (kV)	<i>fend</i> (Hz)
Número de datos	2 000	2 000	2 000	2 000	2 000
Valor mínimo	0,030	58,600	-0,850	13,570	59,400
Valor máximo	1,360	59,970	-0,030	14,020	60,110
Media	0,699	59,314	-0,460	13,768	59,870
Mediana	0,760	59,330	-0,500	13,800	59,950
Desviación estándar	0,370	0,255	0,218	0,108	0,152
Varianza	0,137	0,065	0,048	0,012	0,023
Coeficiente de Asimetría	-0,072	-0,269	0,094	0,276	-1,141
Coeficiente de Kurtosis	-1,570	-0,398	-1,471	-0,918	0,166
Coeficiente de Variación	0,529	0,004	0,474	0,008	0,003

Para profundizar el entendimiento de los principales estadísticos, resulta interesante observar el comportamiento de las variables mediante histogramas y diagramas de cajas que se muestran en la Figura 2. Los datos con mayor frecuencia corresponden a las observaciones que se sitúan entre los valores mínimos y máximos de la caja.

Como se puede observar en la Figura 2, no se identifican datos atípicos significativos en las distribuciones de las variables *DP*, *Nadir*, *RoCoF*, *Vend* y *fend*. Los diagramas de cajas confirman que la mayoría de los datos se encuentran dentro del rango intercuartil, con pocas observaciones fuera de este rango, lo que sugiere que la presencia de *outliers* es mínima.

Las variables *DP*, *Nadir* y *RoCoF* presentan distribuciones bimodales evidentes en sus histogramas. Esto podría indicar la existencia de dos subgrupos bien definidos dentro de los datos, posiblemente asociados a diferentes fuentes de generación de energía, como plantas térmicas y fuentes renovables no convencionales, cada una con características operativas distintas. En las variables *Vend* y *fend*, la mayoría de los datos están concentrados alrededor de los valores centrales, con distribuciones más compactas y un comportamiento menos disperso en comparación con las variables anteriores.

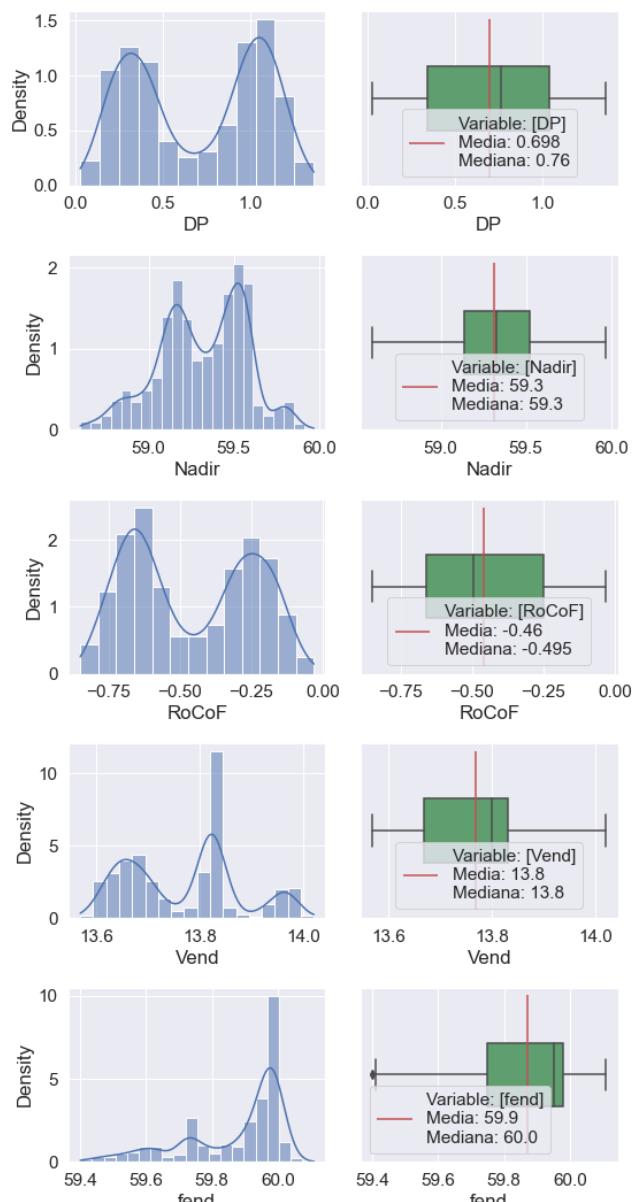
### 3.1.2 Análisis descriptivo bivariante

El análisis descriptivo bivariante se enfoca en explorar las relaciones entre pares de variables mediante herramientas gráficas y estadísticas. En la Figura 3 (a), se puede observar el diagrama de dispersión de las cinco variables en análisis,

donde se puede resaltar que la potencia de salida *DP* tiene un comportamiento lineal con las variables *Nadir* y *RoCoF*.

La Figura 3 (b) presenta la matriz de correlación, que permite identificar la dirección y la magnitud de las relaciones lineales. En este análisis, se observa una alta correlación negativa entre *DP* y las variables independientes *Nadir*, *RoCoF*, *Vend* y *fend*, lo que refuerza la idoneidad de un modelo de regresión lineal múltiple. Adicionalmente, la fuerte dependencia de *DP* con las variables predictoras sugiere que ninguna puede ser eliminada sin afectar significativamente el modelo.

Por último, la Figura 3 (c) muestra la matriz de varianzas y covarianzas, que complementa la matriz de correlación al destacar la variabilidad conjunta entre las variables. Este análisis confirma que las dependencias identificadas son consistentes en magnitud y dirección, reforzando la solidez del modelo propuesto.



**Figura 2.** Diagrama de cajas e histogramas de las variables *DP*, *Nadir*, *RoCoF*, *Vend* y *fend*

### 3.1.3 Regresión lineal múltiple

A continuación, se presenta la evaluación de los modelos predictivos, fundamentada en el análisis de los datos previamente descritos.

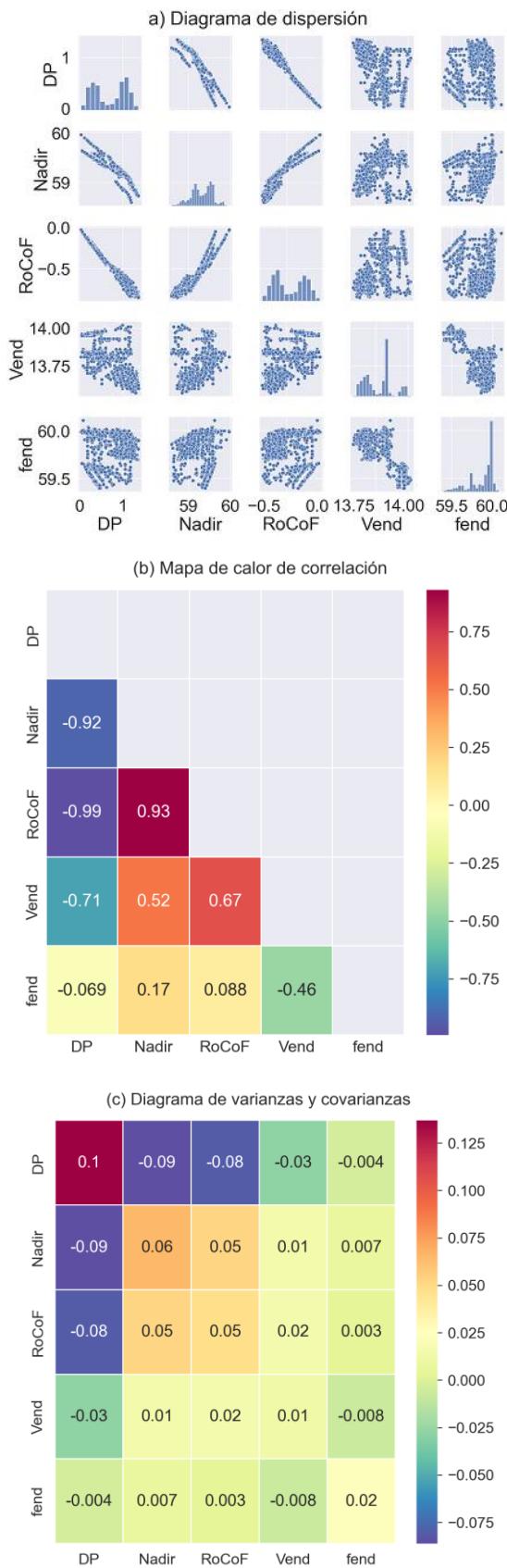


Figura 3. Diagramas de análisis bivariantes

### 3.1.4 Regresión lineal multivariante

El análisis de regresión lineal múltiple, destinado a determinar el tamaño de la perturbación ( $DP$ ), se llevó a cabo utilizando tres modelos distintos. Cada modelo consideró un conjunto diferente de variables independientes (*features*), según lo especificado en la Tabla 3.

- **Modelo 1 (caso a):** Incluye las variables *Nadir*, *RoCoF*, *Vend*, y *fend*. Este modelo mostró un  $R^2$  de 0,99226, e indica un alto grado de ajuste.
- **Modelo 2 (caso b):** Excluye la variable *Nadir*, utilizando solo *RoCoF*, *Vend*, y *fend*. Aunque  $R^2$  disminuye ligeramente a 0,99217, el modelo sigue siendo robusto.
- **Modelo 3 (caso c):** Considera solo las variables *Nadir* y *RoCoF*. El  $R^2$  es de 0,98596, lo que indica un menor ajuste en comparación con los modelos anteriores, aunque aún es considerable.

La comparación entre estos modelos muestra que el *caso a*, que incluye las cuatro variables, ofrece el mejor ajuste. Sin embargo, los modelos simplificados (*caso b* y *c*) también proporcionan resultados satisfactorios, aunque con una ligera disminución en la precisión.

Tabla 3. Resultados del entrenamiento de la regresión lineal múltiple –  
(a) *caso a*: dependencia de 4 variables, (b) *caso b*: 3 variables,  
(c) *caso c*: 2 variables

(a) Caso a		
Descripción	Parámetros	Valores
Coeficiente de correlación múltiple	R	$996,12 \times 10^{-3}$
Coeficiente de determinación	$R^2$	$992,26 \times 10^{-3}$
Coeficiente de determinación ajustado	$R^2$ ajustado	$992,24 \times 10^{-3}$
Error estimado estándar	$\varepsilon$	$32,713 \times 10^{-3}$
(b) Caso b		
Descripción	Parámetros	Valores
Coeficiente de correlación múltiple	R	$996,07 \times 10^{-3}$
Coeficiente de determinación	$R^2$	$992,17 \times 10^{-3}$
Coeficiente de determinación ajustado	$R^2$ ajustado	$992,15 \times 10^{-3}$
Error estimado estándar	$\varepsilon$	$32,78 \times 10^{-3}$
(c) Caso c		
Descripción	Parámetros	Valores
Coeficiente de correlación múltiple	R	$992,95 \times 10^{-3}$
Coeficiente de determinación	$R^2$	$985,96 \times 10^{-3}$
Coeficiente de determinación ajustado	$R^2$ ajustado	$985,94 \times 10^{-3}$
Error estimado estándar	$\varepsilon$	$43,88 \times 10^{-3}$

Además, en la Tabla 3 se comparan los resultados de ambos modelos en términos de la significancia de los coeficientes, la precisión de la predicción (medida por  $R^2$ ), y la importancia relativa de las variables independientes en cada caso.

Como complemento, se desarrollaron dos modelos de regresión lineal simple para evaluar la relación individual de las variables *Nadir* y *RoCoF* con la potencia de salida *DP*. El primer modelo, denominado *caso d*, estima *DP* en función de *Nadir*, como se muestra en la Figura 4(a). En este caso,  $R^2$  es 0,840, lo que indica que *Nadir* por sí sola no es un predictor fuerte de *DP*.

El segundo modelo, denominado *caso e*, estima *DP* en función de *RoCoF*, como se observa en la Figura 4(b). Aquí,  $R^2$  es

significativamente mayor, alcanzando 0,985, lo que demuestra que el *RoCoF* es un predictor mucho más robusto para *DP*. Aunque los modelos univariantes no alcanzan el nivel de precisión de los modelos multivariantes, el *caso e* destaca el papel clave de *RoCoF* como variable determinante en la predicción de perturbaciones, lo que supera notablemente a *Nadir* en términos de desempeño.

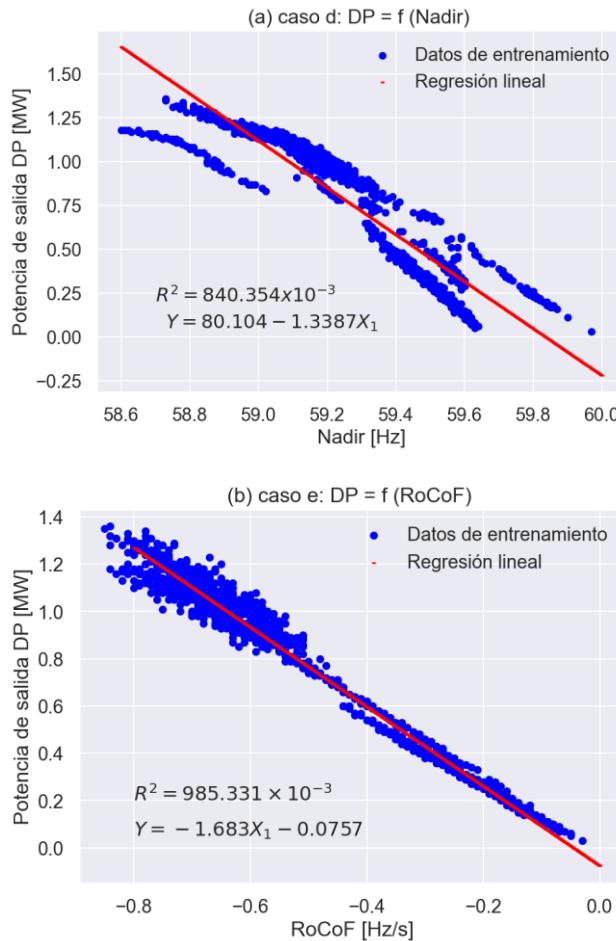


Figura 4. Resultados de regresión lineal simple

Con base en los resultados obtenidos en los diferentes casos analizados hasta aquí, se concluye que el modelo que presenta el mejor desempeño es el correspondiente al *caso a* (regresión lineal múltiple con cuatro variables independientes: *Nadir*, *RoCoF*, *Vend* y *fend*). El modelo del *caso a* permitirá estimar el tamaño de la perturbación de forma más exacta. Los resultados de los *casos d* y *e*, hasta aquí analizados, se incluyen en la Tabla 4.

### 3.2 Capacidad de predicción del modelo

Luego de identificar que el modelo de regresión múltiple del *caso a* (considerando las cuatro variables independientes: *Nadir*, *RoCoF*, *Vend* y *fend*) presentó los resultados más prometedores durante la fase de entrenamiento, se procedió a evaluar su capacidad predictiva en datos nuevos. Esto busca validar su generalización y evitar el sobreajuste. Para este fin, se realizaron pruebas con un conjunto independiente de 400 eventos, calculando el coeficiente de determinación  $R^2$  y las funciones de costo  $J = \sum \varepsilon_i^2$  y  $J_{med} = \sum \varepsilon_i^2/n$ , donde  $n$  es el

número de eventos de cada subconjunto:  $n_1 = 1\,600$  para entrenamiento,  $n_2 = 400$  para prueba.

Es importante destacar que un alto valor de  $R^2$  en el entrenamiento no garantiza un buen desempeño en datos nuevos. Esto ocurre porque los modelos pueden ajustarse a patrones específicos de los datos de entrenamiento que no se replican en el conjunto de prueba. Por ello, la evaluación debe enfocarse en el rendimiento en datos de prueba para asegurar su capacidad de generalización y aplicabilidad en escenarios reales.

Mediante las ecuaciones de función de coste, los resultados obtenidos para el entrenamiento son:  $R^2 = 992,2 \times 10^{-3}$ ,  $J = 1,706$  y  $J_{med} = 1,0 \times 10^{-3}$ ; mientras que los resultados para los datos de prueba son:  $R^2 = 991,2 \times 10^{-3}$ ,  $J = 0,462$  y  $J_{med} = 1,1 \times 10^{-3}$ . A partir de estos resultados, se aprecia que el coeficiente de determinación de la prueba mantiene un valor mayor a 0,99; esto refleja un buen modelo. El coeficiente de determinación de prueba disminuye en un 0,1 % respecto del entrenamiento. Esta leve disminución en la capacidad predictiva también se aprecia en un ligero incremento de la función de coste promedio.

En la Figura 5, se presenta la dispersión del conjunto de prueba, junto con las estimaciones logradas mediante el modelo de regresión lineal múltiple del *caso a*, que incluye las 4 variables independientes. El modelo alcanzó un  $R^2 = 991,294 \times 10^{-3}$ . En la Figura 5(a), se observa la dispersión de *DP* respecto a las 4 variables, graficando la frecuencia en el eje  $x$ , mientras que en la Figura 5(b), se muestra la dispersión de *DP* en función del *RoCoF*.

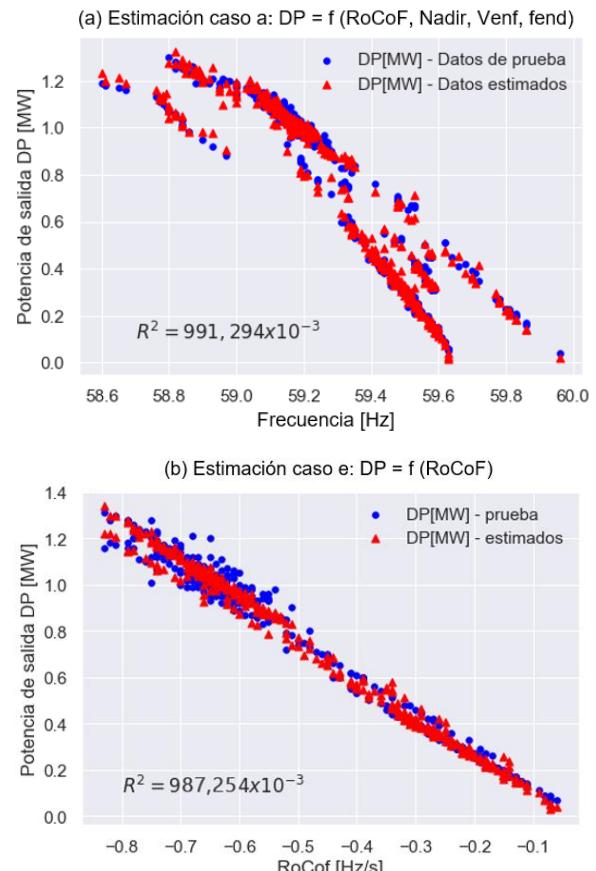


Figura 5. Resultados de estimación - casos de regresión lineal múltiple

Sobre la base de los resultados obtenidos con el modelo de regresión múltiple, se plantea evaluar un modelo adicional de regresión no lineal univariante (*caso f*), ilustrado en la Figura 6. Considerando que la variable independiente *RoCoF* presenta el mejor desempeño predictivo, se proponen dos modelos de  $DP = f(RoCoF)$ : a) *caso f1*, que utiliza un modelo de regresión cuadrática representado en la Figura 6(a); y (b) el *caso f2*, basado en un modelo de regresión cúbica mostrado en la Figura 6(b).

El modelo cuadrático tiene la forma:  $y = (-49,97 \times 10^{-3})x_1^2 - 1,729x_1 - 83,47 \times 10^{-3}$  con un coeficiente de determinación  $R^2 = 985,352 \times 10^{-3}$ . Como se observa en la Figura 6 (b), el modelo de regresión cúbica tiene un coeficiente  $R^2 = 987,792 \times 10^{-3}$  y se expresa mediante la ecuación:  $y = 2,415x_1^3 - 3,216x_1^2 - 0,431x_1 - 58,34 \times 10^{-3}$ . Como se puede revisar, el modelo cúbico presenta un coeficiente de determinación ligeramente mayor al modelo cuadrático, sin embargo, los dos modelos presentan menor desempeño al modelo de regresión lineal del *caso a*.

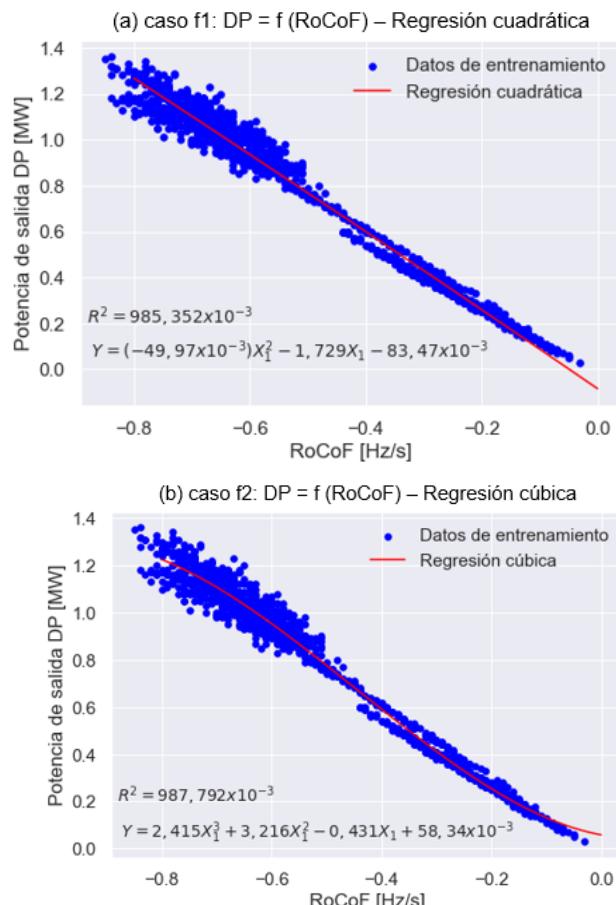


Figura 6. Resultados para regresiones no lineales

### 3.3 Discusión de resultados

Los modelos de análisis de datos multivariantes y las técnicas de aprendizaje automático han tenido relativamente poco desarrollo en el diseño de un esquema de alivio de carga (EAC) aplicado a microrredes. En este contexto, se ha desarrollado una metodología para estimar el tamaño de las perturbaciones (*DP*), el cual sirve como un parámetro crucial en el diseño de un EAC.

En la Tabla 4, se presenta un resumen de los casos de regresión que han sido objeto de estudio en esta investigación. Es evidente que el rendimiento más destacado se observa en el modelo de regresión lineal múltiple que emplea las cuatro variables independientes: *Nadir*, *RoCoF*, *Vend* y *fend*. Por consiguiente, este modelo se utilizaría para la estimación de *DP*, lo que, a su vez, facilitará la implementación de un EAC adaptativo.

Es relevante señalar que entre las cuatro variables que se han analizado, el *RoCoF* presenta la correlación más alta en relación con la magnitud de la perturbación (*DP*). En consecuencia, es posible diseñar un EAC adaptativo exclusivamente basado en la variable *RoCoF*, utilizando un modelo de regresión lineal cúbica.

Tabla 4. Resumen de casos de regresión analizados  $R^2$

Caso	Modelo	Coeficiente de determinación $R^2$
a	Regresión lineal con 4 variables ( <i>Nadir</i> , <i>RoCoF</i> , <i>Vend</i> y <i>fend</i> )	$992,26 \times 10^{-3}$
b	Regresión lineal con 3 variables ( <i>RoCoF</i> , <i>Vend</i> y <i>fend</i> )	$992,17 \times 10^{-3}$
c	Regresión lineal con 2 variables ( <i>Nadir</i> , <i>RoCoF</i> )	$985,96 \times 10^{-3}$
d	Regresión lineal con 1 variable ( <i>Nadir</i> )	$840,35 \times 10^{-3}$
e	Regresión lineal con 1 variable ( <i>RoCoF</i> )	$985,33 \times 10^{-3}$
f1	Regresión cuadrática con 1 variable ( <i>RoCoF</i> )	$985,35 \times 10^{-3}$
f2	Regresión cúbica con 1 variable ( <i>RoCoF</i> )	$987,79 \times 10^{-3}$

## 4. CONCLUSIONES

Las microrredes, especialmente las aisladas, enfrentan desafíos críticos en la estabilidad de la frecuencia debido a perturbaciones significativas, como la pérdida de generación eléctrica. Este problema es especialmente grave en microrredes que carecen de un Esquema de Alivio de Carga (EAC) efectivo. En este estudio, se ha estimado el tamaño de la perturbación (*DP*) en la microrred de Galápagos Santa Cruz – Baltra mediante el uso de análisis multivariante de datos, modelos de regresión, y técnicas de inteligencia computacional.

El análisis reveló que el modelo de regresión lineal múltiple, que incorpora cuatro variables independientes (*Nadir*, *RoCoF*, *Vend* y *fend*), es el más robusto para predecir la magnitud de *DP*. Este modelo no solo proporcionó una estimación precisa de *DP*, sino que también facilitó la implementación de un EAC adaptativo, crítico para mantener la estabilidad en la microrred. Entre las variables analizadas, el *RoCoF* demostró tener la correlación más fuerte con *DP*, destacando su importancia en la predicción de perturbaciones.

Por otro lado, los modelos de regresión polinómica, tanto cuadráticos como cúbicos, centrados en *RoCoF* como variable independiente, no mostraron mejoras en la fiabilidad predictiva en comparación con los modelos lineales. Esto sugiere que la complejidad adicional de los modelos polinómicos no necesariamente se traduce en un mejor desempeño para este tipo de problemas.

Para asegurar la transparencia y reproducibilidad de los resultados, el código y los modelos desarrollados en este estudio se han publicado en un repositorio público en [GitHub](#), facilitando el acceso a cualquier persona interesada en verificar los resultados. Los scripts creados documentan detalladamente todo el proceso de desarrollo y validación de los modelos, promoviendo su uso y análisis dentro de la comunidad científica.

Como recomendación para futuros estudios, se sugiere explorar la aplicación de redes neuronales para comparar los resultados obtenidos con los métodos de regresión tradicionales. Este enfoque podría aportar una perspectiva adicional y posiblemente revelar patrones o tendencias no detectadas por las técnicas convencionales. La implementación de redes neuronales podría ofrecer mejoras significativas en la capacidad predictiva y en la comprensión de las dinámicas complejas dentro de las microrredes.

## REFERENCIAS

- Bakirtzis, A. (1994). Genetic algorithm solution to the economic dispatch problem. *IEE Proceedings - Generation, Transmission and Distribution*, 141(4), 377. <https://doi.org/10.1049/ip-gtd:19941211>
- Cepeda, J. (2013). *Evaluación de la vulnerabilidad del sistema eléctrico de potencia en tiempo real usando tecnología de medición sincrofasorial* [Tesis doctoral]. Universidad Nacional de San Juan.
- Chowdhury, B. H., & Rahman, S. (1990). A review of recent advances in economic dispatch. *IEEE Transactions on Power Systems*, 5(4), 1248–1259. <https://doi.org/10.1109/59.99376>
- DIgSILENT GmbH. (2022a, enero 10). *Technical Reference Fully Rated Converter Wind Turbine Templates*. Power System Solutions.
- DIgSILENT GmbH. (2022b, enero 10). *Technical Reference PV System ElmPvsys, TypPvpanel*. Power System Solutions.
- Escobar, B. (2021). *Esquema adaptativo de seccionamiento de carga en función del voltaje y la frecuencia usando simulación digital en tiempo real* [Tesis de maestría]. Escuela Politécnica Nacional.
- Guo-Li Zhang, Hai-Yan Lu, Geng-Yin Li, & Guang-Quan Zhang. (2005). Dynamic economic load dispatch using hybrid genetic algorithm and the method of fuzzy number ranking. *2005 International Conference on Machine Learning and Cybernetics*, 2472-2477 Vol. 4. <https://doi.org/10.1109/ICMLC.2005.1527359>
- IEEE Std. 2030-7. (2018). IEEE Standard for the specification of microgrid controllers. *IEEE Std. 2030-7-2018*, 1–43. <https://doi.org/10.1109/IEEEESTD.2018.8340204>
- Jiang, H., Yan, G., Ji, H., Liu, L., & Shan, D. (2010). An Improved Under Frequency Load Shedding Scheme Based on Rate of Change of Frequency. *2010 International Conference on Electrical and Control Engineering*, 3292–3295. <https://doi.org/10.1109/ICECE.2010.803>
- Mejía, C. A., Gallardo, C., & Pesantez, P. (2024). Proposal of a Load Shedding Scheme for Microgrids Using IEC 61850 Standard and the Rate of Change of Frequency. *2024 IEEE Eighth Ecuador Technical Chapters Meeting (ETCM)*, 1–7. <https://doi.org/10.1109/ETCM63562.2024.10746010>
- Mohamad, H., Isa, A. I. M., Yasin, Z. M., Salim, N. A., & Mohd Rahim, N. N. A. (2017). Optimal load shedding technique for an islanding distribution system by using Particle Swarm Optimization. *2017 3rd International Conference on Power Generation Systems and Renewable Energy Technologies (PGSRET)*, 154–158. <https://doi.org/10.1109/PGSRET.2017.8251819>
- Mohd Zin, A. A., Mohd Hafiz, H., & Aziz, M. S. (2004). A review of under-frequency load shedding scheme on TNB system. *PECon 2004. Proceedings. National Power and Energy Conference, 2004.*, 170–174. <https://doi.org/10.1109/PECON.2004.1461637>
- Muñoz, L., & Quezada, P. (2015). *Flujo óptimo de potencia DC considerando restricciones por congestión y pérdidas en las líneas para resolución de problemas de despacho económico* [Universidad Politécnica Salesiana Sede Quito]. <https://dspace.ups.edu.ec/bitstream/123456789/11286/3/UPS-KT01199.pdf>
- Peña, D. (2002). *Análisis de datos multivariantes* (McGrawHill). McGrawHill. <https://b-ok.lat/book/1185829/d0e57e>
- Pérez, S., López, K., & Rodríguez, L. (2016). Optimal under voltage load shedding based on voltage stability index. *Ingeniería e Investigación*, 36(2), 43. <https://doi.org/10.15446/ing.investig.v36n2.53773>
- Sun, Y., Zhang, Z., Li, G., & Lin, J. (2010). Review on frequency control of power systems with wind power penetration. *2010 International Conference on Power System Technology*, 1–8. <https://doi.org/10.1109/POWERCON.2010.5666151>
- Wiest, P., Rudion, K., & Probst, A. (2018). Efficient integration of (n-1)-security into probabilistic network expansion planning. *International Journal of Electrical Power & Energy Systems*, 94, 151–159. <https://doi.org/10.1016/j.ijepes.2017.07.002>

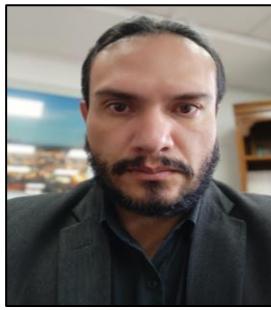
## BIOGRAFÍAS



**César Andrés, Mejía**, Ingeniero en Electrónica y Control (2010) e Ingeniero Eléctrico (2013) por la Escuela Politécnica Nacional, Quito, Ecuador. Posee una Maestría en Electricidad con mención en Redes Eléctricas Inteligentes (2024) por la misma institución y un MBA (2022) de la Universidad Andina Simón Bolívar. Su experiencia incluye estudios eléctricos para sistemas de potencia y microrredes, enfocados en estabilidad, confiabilidad y eficiencia energética. Es especialista en automatización de subestaciones eléctricas e integración de sistemas SCADA mediante DNP3 e IEC 61850, contribuyendo al diseño de soluciones innovadoras para redes eléctricas inteligentes y sostenibles.



**Alex Ricardo, Rocha**, Ingeniero en Electrónica y Control por la Escuela Politécnica Nacional, Quito, Ecuador (2019), y Magíster en Electricidad con mención en Redes Eléctricas Inteligentes de la misma institución. Actualmente, se desempeña como especialista en estudios de automatización y en la gestión de sistemas en laboratorios de metrología y calibración en magnitudes de potencia y energía. Sus áreas de investigación incluyen la automatización de procesos eléctricos, eficiencia energética, calidad de energía, transformación digital y minería de datos, enfocándose en el desarrollo de soluciones innovadoras para sistemas eléctricos y tecnologías avanzadas.



**Jaime Cristóbal, Cepeda**, Recibió el título de Ingeniero Eléctrico en la Escuela Politécnica Nacional (2005), Doctor en Ingeniería Eléctrica en la Universidad Nacional de San Juan, Argentina (2013), y Máster en Big Data en la Universidad Miguel de Cervantes, España (2021). Fue investigador en el Instituto de Energía Eléctrica, Argentina, y en el Instituto de Sistemas Eléctricos de Potencia, Alemania (2009-2013). Lideró áreas técnicas en CENACE (2014-2021) y fue Director Ejecutivo de la Agencia de Regulación y Control de Energía (2021-2022). Actualmente, es Profesor en la Escuela Politécnica Nacional. Ha publicado más de 120 artículos en temas como estabilidad de sistemas eléctricos y microrredes.

