

Preprocesamiento de los datos de las Unidades de Medición Sincrofasorial (PMUs) utilizando la técnica limpieza de datos - Aplicación al Sistema Nacional Interconectado Ecuatoriano

Diana Guevara, Nelson Granda, Franklin Quilumba

Ingeniería Eléctrica, Escuela Politécnica Nacional

Resumen— En el presente artículo se aplican las técnicas de “Limpieza de datos” sobre la base de datos de las mediciones sincrofasoriales de frecuencia, tasa de frecuencia, fasores de secuencia positiva de las ondas sinusoidales de voltaje y corriente de las PMUs del Sistema Nacional Interconectado. En primera instancia, se realiza una descripción detallada del conjunto de tareas o procedimientos que comprenden esta técnica, las cuales hacen uso de diversas funciones o metodologías para detectar, diagnosticar, e imputar datos anómalos, con la finalidad de crear una base de datos fidedigna e influir en la calidad de los resultados de posteriores análisis o estudios. Por consiguiente, se implementa una aplicación para el desarrollo de la limpieza de datos de las diferentes mediciones sincrofasoriales en App Designer del software MATLAB. Dicha aplicación se encuentra dividida en secciones o etapas que permiten el desarrollo de esta técnica, y cada una de ellas ejecutan una rutina que permiten: el manejo apropiado de la base de datos, descripción de las características de la señal temporal seleccionada, tratamiento de datos vacíos, filtrado de ruido, detección e imputación de datos atípicos, con el objetivo de resolver las inconsistencias de una forma interactiva para el usuario. Finalmente, se obtienen los errores del conjunto de metodologías empleadas a lo largo del proceso de limpieza de datos, para así seleccionar y validar las más adecuadas dependiendo de la señal analizada.

Palabras Clave— base de datos, mediciones sincrofasoriales, PMU, datos anómalos, limpieza de datos, imputación de datos.

I. INTRODUCCIÓN

En la actualidad, para una operación segura y confiable de la red eléctrica se han desarrollado nuevas tecnologías, como son las Unidades de Medición Fasorial (PMUs), con la finalidad de mejorar el monitoreo, protección y control de los Sistemas Eléctricos de Potencia. La inclusión de las PMUs permite la estimación de mediciones instantáneas y precisas de frecuencia, variación de cambio de frecuencia, fasores de secuencia positiva de las ondas sinusoidales de voltaje y corriente de una red eléctrica en tiempo real, independientemente del estado operativo en el que esta se encuentre [1].

Estos medidores inteligentes PMUs, permiten la obtención y suministro de los datos fasoriales de las diferentes señales de acuerdo con una estampa de tiempo. El uso de estos datos es de

suma importancia, ya que facilita el diseño de técnicas que permiten evaluar la vulnerabilidad del sistema eléctrico de potencia, a través de análisis en tiempo real o después de ocurrida una contingencia.

Previo al uso de los datos fasoriales en cualquier análisis es necesario realizar su preprocesamiento utilizando técnicas de limpieza de datos, debido a que los datos pueden contener datos anómalos, lo que representa dificultades para futuros estudios. La limpieza de datos es un término usado de manera general para referirse a una variedad de tareas destinadas a mejorar la calidad de los datos. Estas tareas se consiguen mediante la unión de varias operaciones y en el presente proyecto se analizan algunas técnicas comunes de limpieza de datos para comprender mejor las operaciones subyacentes [2], [3].

El presente artículo se encuentra organizado de la siguiente manera: la sección 2 presenta un resumen de las principales características del análisis fasorial y sincrofasorial. La sección 3 describe la tecnología de medición sincrofasorial. La sección 4 menciona el preprocesamiento de datos en series temporales. La sección 5 resume las metodologías implementadas. La sección 6 explica la interfaz gráfica para la limpieza de datos de las mediciones sincrofasoriales. Finalmente, las secciones 6 y 7 muestran resultados y conclusiones, respectivamente.

II. ANÁLISIS FASORIAL Y SINCRUFASORIAL

A. Análisis fasorial

Un fasor es una representación vectorial constante de una función sinusoidal, asumiendo que la amplitud, frecuencia y fase permanecen constantes [4]. Usar fasores consiste en una forma de describir sintéticamente una señal sinusoidal en redes AC, asumiendo una frecuencia nominal constante.

La definición matemática clásica de fasor se basa en una señal de corriente alterna genérica $x(t)$ que se representa por:

$$x(t) = X_m \cos(\omega t + \phi) \quad (1)$$

Donde X_m es el valor pico de la señal, $\omega = 2\pi f$ es la frecuencia angular del sistema, f es la frecuencia instantánea y ϕ es el ángulo de fase de la señal. Empleando la identidad de

Euler, la Ecuación (1) puede ser expresada como:

$$\bar{X} = \left(\frac{X_m}{\sqrt{2}}\right) (\cos \phi + j \sin \phi) = X_r + jX_i \quad (2)$$

Dónde: X_r y X_i son las componentes rectangulares real e imaginaria de la representación fasorial, con su respectivo módulo $X_m/\sqrt{2}$ que corresponde al valor RMS (Root Mean Square) de la señal sinusoidal. La Fig. 1 muestra el instante de tiempo en el cual la longitud X_m va rotando a ω rad/s. A medida que la longitud X_m rota, su proyección instantánea a lo largo del eje horizontal es el valor instantáneo del fasor. Suponiendo que la longitud X_m está alineada con el eje horizontal en el tiempo "t = 0", entonces se asocia que el ángulo de fase es 0° . Por lo tanto, se puede expresar el fasor como $\bar{X}_1 = (X_m/\sqrt{2})\angle 0$. Ahora suponiendo que en el tiempo "t = 0", al rotar la longitud X_m esta se encuentra en el ángulo ϕ° respecto al eje horizontal, entonces se asocia que el ángulo de fase es ϕ° , expresando el fasor como $\bar{X}_2 = (X_m/\sqrt{2})\angle \phi$, de esta manera se muestra la representación fasorial de una onda sinusoidal dada por la Ecuación (2) [5].

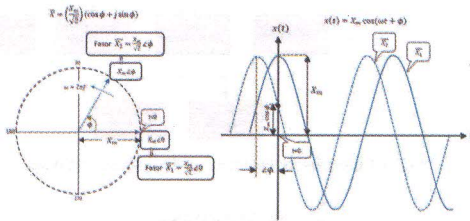


Fig. 1 Representación fasorial de una onda sinusoidal

B. Análisis sincrofasorial

Un sincrofasor se define como "un fasor calculado a partir de datos muestreados usando una señal de tiempo estándar como referencia para la medición" [6]. Es decir, la estimación del sincrofasor se basa en la misma idea del fasor, con diferencia que el ángulo de fase se calcula utilizando el Tiempo Universal Coordinado (UTC – Coordinated Universal Time) como referencia de tiempo, lo que permite tener una referencia única para todas las señales sinusoidales a medir en cualquier parte de la red eléctrica, ya que la difusión del tiempo depende únicamente de los sistemas de satélite [7].

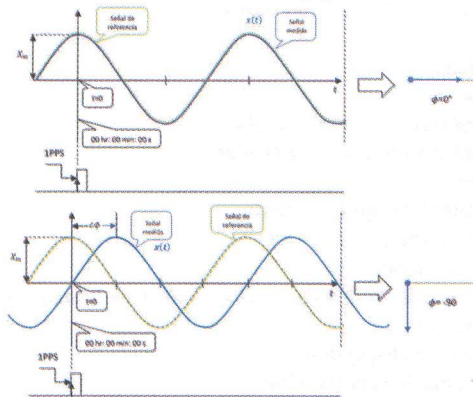


Fig. 2 Convención para representación de sincrofasores

La expresión de sincrofasor tiene la misma representación de la Ecuación (2) con la diferencia que el ángulo de fase ϕ es el desplazamiento entre la onda $x(t)$ con una forma de onda cosenoidal ficticia, la misma que tiene una frecuencia igual a la frecuencia nominal del sistema y es sincronizada al UTC. Dicha onda cosenoidal tiene su punto máximo en un instante de tiempo absoluto, que se toma como referencia "t = 0". Al comienzo de cada segundo (que se remonta a este instante "t = 0"), una señal conocida como "1 PPS" (1 pulso por segundo) se transmite desde el GPS (Sistema de Posicionamiento Global), entonces cuando el valor máximo de la función $x(t)$ coincide con la señal "1 PPS" el ángulo de fase ϕ tendrá un valor de 0° , y cuando el cruce por cero positivo de la función $x(t)$ coincide con la señal "1 PPS" tendrá un valor de -90° , tal como se muestra en la Fig. 2 [8].

III. TECNOLOGÍA DE MEDICIÓN SINCRIFASORIAL

A. Unidades de medición sincrofasorial (PMU)

Una unidad de medición sincrofasorial es un dispositivo electrónico moderno de medición y protección que permite calcular o estimar medidas de frecuencia, tasa de frecuencia, fasores de secuencia positiva de las ondas sinusoidales de voltaje y corriente de la red eléctrica con una tasa de muestreo de 60 muestras por segundo [9], [10].

B. Funcionamiento de una PMU

En una PMU se tiene en las entradas las ondas sinusoidales de corrientes y voltajes obtenidas de los devanados secundarios de los transformadores de corriente y de potencial. Las cuales se utilizan para llevar a cabo las mediciones de secuencia positiva a la salida de la PMU. Luego de lo cual, se filtra cada señal analógica mediante el uso de un filtro anti-aliasing para luego ser enviadas al convertidor analógico/digital. Para mantener la frecuencia constante de las señales se utiliza el reloj de muestreo, el cual se encuentra sincronizado en fase con la señal de un pulso por segundo "1 PPS" proporcionada por el receptor GPS a través del oscilador de enganche de fase. Las señales obtenidas del convertidor analógico/digital son enviadas al microprocesador junto con sus respectivas estampas de tiempo. El microprocesador mediante la Transformada Discreta de Fourier calcula los fasores de todas las señales de corriente y voltaje de secuencia positiva a través de N muestras en un período de tiempo específico. Finalmente, se transmite esta información mediante un puerto de comunicación hacia el concentrador de datos fasoriales, para el caso de CENACE el PDC es de tipo Super-PDC, en la Fig. 3 se puede observar el esquema de funcionamiento de una PMU [11], [12].

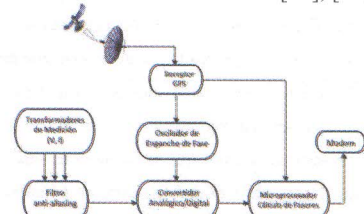


Fig. 3 Estructura básica de una PMU

C. Aplicaciones de las PMUs

El sistema de potencia está distribuido en un área geográfica muy grande normalmente la mayoría de los relés de protección hacen uso de la información disponible localmente, con tiempos de respuesta de varios milisegundos, suficiente para la protección local. Actualmente los sistemas SCADA permiten una visualización integral de los sistemas eléctricos de potencia y debido a las limitaciones en la frecuencia de muestreo proporcionan una visión relativamente estática. La introducción de las unidades de medición sincrofásorial permite la consideración del control basado en el valor medido de cantidades remotas, la información de diversos lugares puede estar disponible en cualquier parte del sistema mediante la integración de las PMUs y el sistema SCADA, permitiendo así una visión dinámica del sistema de potencia y a su vez una gran cantidad de aplicaciones novedosas que pueden proporcionar protección y control de un área amplia [12], [13], [14].

IV. PREPROCESAMIENTO DE DATOS EN SERIES TEMPORALES

Las bases de datos actuales son altamente susceptibles a datos atípicos, faltantes y al ruido lo que presenta importantes dificultades al ejecutar cualquier tipo de análisis. Existen varias técnicas de preprocesamiento de datos como: limpieza de datos, integración de datos, reducción de datos. En los últimos años, esta área se ha vuelto de gran importancia debido a que algoritmos para diferentes análisis requieren datos significativos y manejables para operar correctamente y proporcionar conocimiento útil [15].

A. Limpieza de datos en series temporales

La limpieza de datos consiste en detectar, diagnosticar, resolver inconsistencias y la imputación de datos anómalos. Si a los datos no se aplica la técnica de limpieza la confiabilidad de los resultados de cualquier análisis disminuye.

TABLA I.
TIPOS DE DATOS ANÓMALOS

TIPOS DE DATOS ANÓMALOS	
Tipo de datos	Característica
Datos faltantes	Los datos faltantes o también llamados datos NaN (Not a Number) se refiere a que no existen valores para una observación específica en un conjunto de datos univariantes o cuando no hay valores de datos para una variable particular de un conjunto de datos multivariantes. La falta de datos se debe principalmente a errores en la recopilación de datos.
Datos atípicos	Los datos atípicos son valores de datos que se desvía significativamente del resto del conjunto de datos, como si hubiera sido generado por un mecanismo diferente.
Datos Ruidosos	El ruido es una señal aleatoria que tiene diferentes frecuencias que van desde frecuencias cero (DC) hasta frecuencias infinitas que se suman a la señal original, lo que le da una densidad espectral de potencia constante, esto significa que la señal contiene todas las frecuencias y todas ellas muestran la misma potencia, por lo que el ruido se considera función de su dinámica de tiempo y frecuencia.

Específicamente, esta disertación se centra en la limpieza de datos de series temporales. El entrenamiento de un modelo en una serie de tiempo que contiene datos anómalos generalmente da como resultado parámetros erróneos y sesgados.

A continuación, se explicará los tipos de datos anómalos que serán tratados en la limpieza de datos y se exponen en la Tabla I. [15], [16], [17].

B. Detección e imputación de datos anómalos

La detección consiste en identificar diferentes tipos de datos anómalos, mientras que la imputación consiste en decidir sobre posibles correcciones para los valores anómalos encontrados [18]. Las técnicas o enfoques de detección e imputación de datos se eligen en función a los tipos de datos anómalos, las cuales pueden ser:

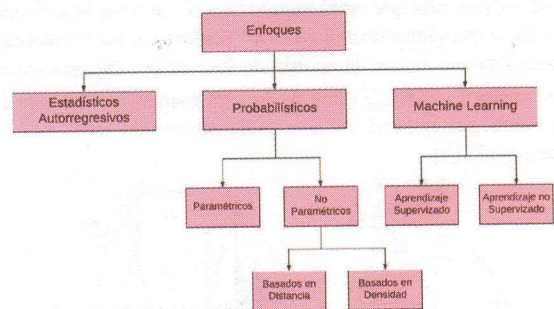


Fig. 4 Enfoques de las técnicas de limpieza de datos

V. METODOLOGÍA

Se presenta una descripción detallada de las metodologías implementadas en la aplicación, cada una de estas contienen diferentes enfoques para la evaluación, detección e imputación de datos anómalos.

A. Estadística descriptiva en series temporales

La estadística descriptiva permite recopilar, analizar, interpretar o explicar y representar los datos. Un modelo estadístico es un conjunto de funciones matemáticas que describen el comportamiento de los datos en términos de variables aleatorias y sus distribuciones de probabilidad asociadas [19]. A continuación, se enumera

- 1) *Medidas de Tendencia Central*
Media
Mediana
Moda
- 2) *Medidas de dispersión de datos*
Varianza y desviación estándar
Rango
Cuartiles y rango intercuartil
Máximo valor
Mínimo valor
- 3) *Gráficas de las descripciones estadísticas básicas de datos*
Histograma
Gráfico de Dispersión
Diagrama de caja (BoxPlot)
Gráfico Cuantil

B. Metodologías para la detección e imputación de los datos NaN

Este tipo de métodos analizan los datos faltantes de una serie temporal en un caso particular. La operación de decidir qué datos se usan para llenar estos espacios en blanco se denomina imputación de datos. Este término significa que se asigna datos al espacio en blanco en función de alguna heurística razonable. La selección de la técnica adecuada para manejar los valores perdidos depende de hacer la suposición correcta sobre el patrón de "faltantes" en el conjunto de datos [20]. A continuación, se enumera los algoritmos implementados.

- 1) *Interpolación lineal*
- 2) *Interpolación cúbica "Spline"*
- 3) *Interpolación cúbica "Akima"*
- 4) *Modelo Autorregresivo*
- 5) *Media Móvil*

C. Metodologías para filtrar ruido en series temporales

En las señales temporales existe ruido, el cual se constituye por señales indeseables que se introducen a lo largo del trayecto de transmisión, de manera que altera la señal deseada. Este ruido es generado por causas internas y externas al sistema y constituye uno de los principales factores que limitan un análisis de la señal. Un punto importante para una adecuada limpieza de datos es eliminar el ruido, el cual es una señal aleatoria generalmente de alta frecuencia, pero con un ancho de banda muy amplio, que se suma a la señal original. Primero se procede a la visualización en el dominio de la frecuencia haciendo uso de la Transformada de Fourier o la Transformada Wavelet para posteriormente filtrar el ruido de la señal [21]. A continuación, se enumera los algoritmos implementados.

- 1) *Transformada de Fourier*
- 2) *Transformada de Wavelet*

D. Metodologías para la detección e imputación de los datos atípicos

Las series temporales puede contener datos que no cumplan con el comportamiento general o modelo del conjunto de datos, estos valores son llamados atípicos. Los valores atípicos podrían causar problemas críticos debido a que una pequeña cantidad de este tipo de datos amenaza con perjudicar gravemente el contenido intrínseco de los datos al cambiar dramáticamente la tendencia de las estimaciones.

Se ha implementado algunos métodos de suavizado que permiten detectar e imputar los valores atípicos y aproximar los datos de la señal a una función estimada por patrones importantes de la serie temporal, modificando los valores atípicos y ajustando a una curva suave y fina en relación con la señal original. Otro de los métodos de detección son los basados en la agrupación, los cuales establecen que los objetos de datos normales pertenecen a agrupaciones grandes y densas, mientras que los valores atípicos pertenecen a agrupaciones pequeñas o dispersas, o no pertenecen a ninguna agrupación [22].

A continuación, se enumera de las técnicas de suavizado y de agrupamiento que detectan e imputan los valores atípicos en una serie temporal.

- 1) *Regresión lineal local "LOWESS"*
- 2) *Regresión cuadrática local "LOESS"*
- 3) *Savitzky-Golay "SGOLAY"*
- 4) *K-Means*

E. Criterios de evaluación de metodologías

Para comparar y medir el desempeño de los algoritmos presentados para la imputación de datos de las señales temporales se emplea dos indicadores: la raíz cuadrada del error cuadrático medio (RMSE) y el coeficiente de determinación (R2), los cuales miden la desviación entre la señal original y su versión reconstruida, de tal manera que esto aporte en la elección del algoritmo indicado para su aplicación.

VI. INTERFAZ GRÁFICA PARA LA LIMPIEZA DE DATOS DE LAS MEDICIONES SINCRFASORIALES

A. Diagrama de bloques de la aplicación desarrollada para la limpieza de datos de las mediciones sincrofásorales

La aplicación ha sido desarrollada con la finalidad de ofrecer al usuario facilidades al momento de interactuar con la ejecución del programa, de tal forma que mediante la interfaz gráfica creada le sea posible manejar cada una de las etapas en las que se efectuará la limpieza de datos. En la primera etapa "Selección de la señal", brinda la opción de escoger de una manera práctica una de las diferentes señales temporales registradas por las PMUs instaladas en el SNI. La etapa "Estadística Descriptiva" permite observar el comportamiento de los datos en términos de variables aleatorias y sus respectivas gráficas de descripciones estadísticas básicas del conjunto de datos. "Tratamiento de datos NaN" consiste en llenar los datos faltantes en función de algunos de los métodos específicos propuestos para esta etapa. "Filtrado de la señal" elimina el ruido. "Tratamiento de datos atípicos" esta etapa consiste en detectar e imputar la influencia de los valores atípicos. "Resultados de la señal preprocesada" permite obtener la señal temporal preprocesada y el cálculo de errores de la combinación de los métodos escogidos. En la Fig. 5 se puede observar la secuencia de las etapas a seguir para obtener la señal temporal preprocesada mediante la aplicación desarrollada.

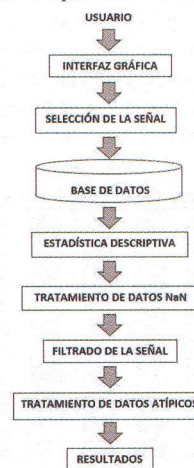


Fig. 5 Bosquejo del diagrama de bloques del funcionamiento general de la aplicación

B. Descripción de App Designer de MATLAB

App Designer integra dos tareas principales: diseñar los componentes visuales de la interfaz gráfica de usuario (GUI - Graphical User Interface) y programar el comportamiento de la aplicación, las cuales están estrechamente vinculadas por lo que si se realiza cambios en el diseño afecta inmediatamente a la programación de la aplicación. Incluye una versión totalmente integrada del editor MATLAB, fue introducido en la versión R2016a [23].

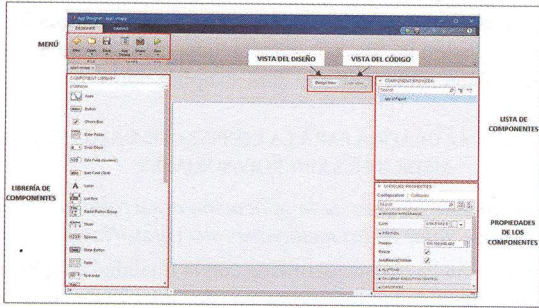


Fig. 6 Vista del diseño de la App Designer

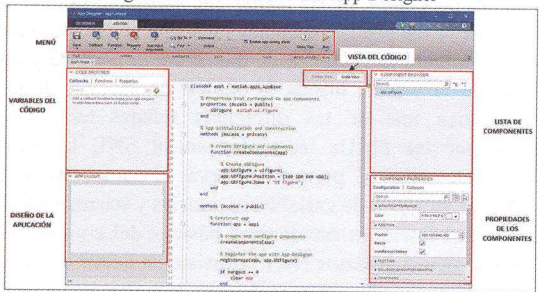


Fig. 7 Vista del código de la App Designer

C. Base de Datos

La base de datos proporcionada contiene información de 29 PMUs instaladas en el SIN, cada una de las PMUs obtienen las diferentes mediciones sincrofásorales de frecuencia, tasa de frecuencia, fasores de secuencia positiva de voltaje y corriente de su respectiva ubicación en la red eléctrica. Por lo que se obtiene una tabla característica correspondiente a cada PMU, la cual consta de 6 columnas y 47942 filas, y contienen los datos de las diferentes mediciones sincrofásorales a una tasa de muestreo de 60 muestras por segundo en un intervalo de tiempo de 800 [s]. A continuación, en la Fig. 8 se presenta un ejemplo de la PMU AGOYÁN-BAÑOS 1 con su correspondiente tabla.

UTC-Tiempo Universal	Coordenado	Conjunto de atributos
144797088 144797111	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797133	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797155	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797177	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797199	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797221	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797243	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797265	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797287	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797309	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797331	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797353	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797375	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797397	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797419	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797441	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797463	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797485	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797507	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797529	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797551	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797573	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797595	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797617	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797639	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797661	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797683	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797705	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797727	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797749	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797771	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797793	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797815	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797837	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797859	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797881	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797903	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797925	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797947	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144797969	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144798015	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144798061	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144798107	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144798153	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144798199	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144798245	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144798291	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144798337	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144798383	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144798429	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144798475	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144798521	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144798567	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144798613	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144798659	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144798705	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144798751	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144798797	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144798843	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144798889	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144798935	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144798981	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144799027	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144799073	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144799119	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144799165	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144799211	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144799257	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144799303	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144799349	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144799395	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144799441	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144799487	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144799533	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144799579	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144799625	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144799671	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144799717	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144799763	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144799809	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144799855	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144799901	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144799947	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
144797088 144800000	0.00000000	0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000

Fig. 8 Tabla de datos de la PMU AGOYÁN-BAÑOS 1

D. Interfaz gráfica implementada para la limpieza de datos

La aplicación consta de las siguientes etapas: selección de la señal, estadística descriptiva, tratamiento de datos NaN, filtrado de la señal, tratamiento de datos atípicos, resultados de la señal preprocesada. Estas permiten obtener como resultado final la limpieza de las diferentes mediciones sincrofásorales por cada una de las PMUs instaladas en el SIN.

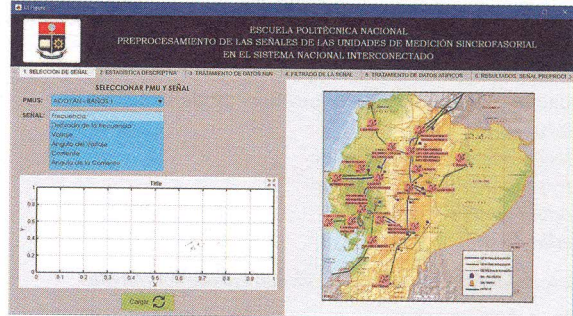


Fig. 9 Interfaz gráfica para la limpieza de mediciones sincrofásorales

VII. RESULTADOS

La aplicación para el preprocesamiento de señales mediante las técnicas de limpieza de datos presenta una gran cantidad de opciones en el uso y combinación de las diferentes metodologías expuestas en cada sección.

A. Caso 1: SEÑAL DE FRECUENCIA

1) Resultado de la sección 1: SELECCIONAR SEÑAL

Para el presente caso de estudio se escogió por señal a la frecuencia de la PMU AGOYÁN - BAÑOS 1

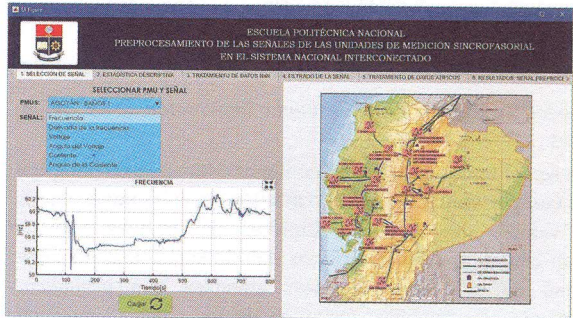


Fig.10 Sección 1: Selección de Señal

Se obtiene una mejor apreciación de la señal temporal de frecuencia, la cual a simple vista presenta discontinuidades, y un pico pronunciado al tiempo de 117 [s], estas observaciones serán tratadas en las secciones posteriores.

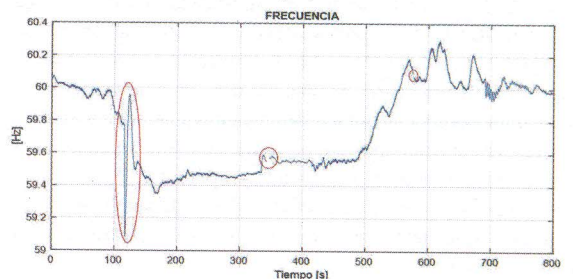


Fig. 11 Señal de la frecuencia de la PMU AGOYÁN - BAÑOS 1

2) Resultado de la sección 2: ESTADÍSTICA ESCRITIVA

Previo al uso de cualquier tipo de metodología sobre la señal de frecuencia, se procede a tener una imagen general de los datos. En la Tabla II se observa los resultados de las medidas de tendencia central y de dispersión, las cuales resumen y describen al conjunto de datos que representan la señal de frecuencia.

TABLA II. MEDIDAS DE TENDENCIA CENTRAL Y DE DISPERSIÓN

Medidas de tendencia central y de dispersión	Valor [Hz]
Media	59,77
Mediana	59,8
Moda	59,47
Varianza	0,07258
Desviación	0,2694
Máximo	60,3
Mínimo	59,08
Rango	1,214

A continuación, se presenta las gráficas de las descripciones estadísticas básicas de datos de la señal de frecuencia de la PMU AGOYÁN - BAÑOS I

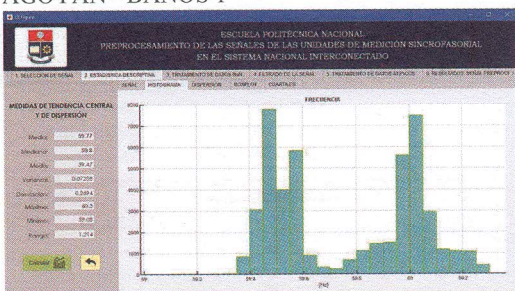


Fig. 12 Sección 2: Estadística Descriptiva - Histograma

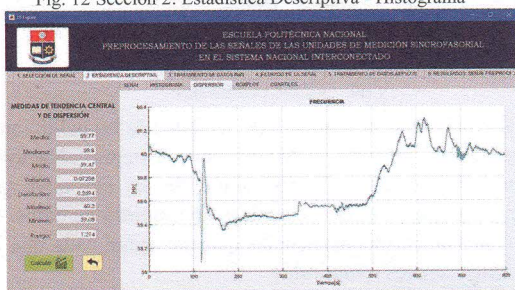


Fig. 13 Sección 2: Estadística Descriptiva – Diagrama de Dispersión

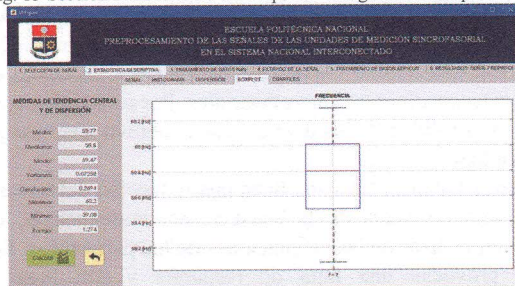


Fig. 14 Sección 2: Estadística Descriptiva – Gráfica BoxPlot



Fig. 15 Sección 2: Estadística Descriptiva – Gráfico Cuantil

3) Resultado de la sección 3: TRATAMIENTO DE DATOS NaN

El resultado de aplicar cada una de las diferentes metodologías para el tratamiento de datos NaN de la nueva señal de frecuencia se expone en la Fig. 16 y Fig.17.

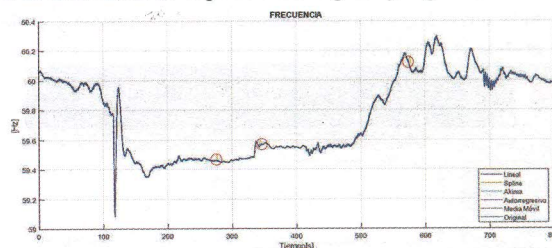


Fig. 16 Señal de la frecuencia con tratamiento de datos NaN

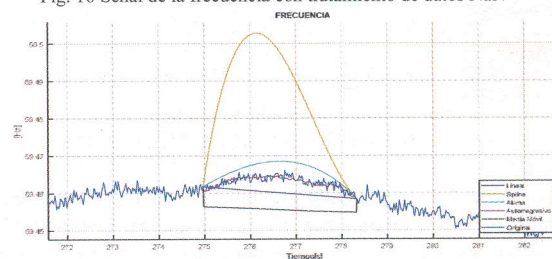


Fig. 17 Señal de la frecuencia con tratamiento de datos NaN

Como se puede apreciar el modelo autorregresivo presenta como resultado un valor RMSE muy bajo, por lo que en comparación a los demás modelos este resulta ser el óptimo para su presente aplicación.

TABLA III. CÁLCULO DE LOS ERRORES DE LOS MÉTODOS PARA EL TRATAMIENTO DE DATOS NaN EN LA SEÑAL DE LA FRECUENCIA

Metodología	RMSE
Lineal	0,0032
Spline	0,0259
Akima	0,0032
Autorregresivo	0,0008
Media Móvil	0,0075

4) Resultado de la sección 4: FILTRADO DE LA SEÑAL

En la Tabla IV se presenta el cálculo RMSE de los diferentes métodos empleados. Como se puede observar el análisis de Fourier y el filtro FIR presentan el valor más bajo de RMSE en comparación a los demás modelos, por esto resulta ser el modelo óptimo para su presente aplicación.

TABLA IV.
CÁLCULO DE LOS ERRORES DE LOS MÉTODOS PARA EL FILTRADO DE LA SEÑAL DE LA FRECUENCIA

Metodología	RMSE
Fourier y filtro Yulewalk	0,0018
Fourier y filtro Butterworth	0,0017
Fourier y filtro FIR	0,0014
Wavelet	0,0019

En la Fig. 18 se presenta los resultados del análisis de Fourier junto con el filtro FIR de orden 5. En la primera gráfica se obtiene la respuesta de la señal en el dominio de la frecuencia, en la segunda gráfica se observa la respuesta del filtro pasa bajo en el dominio de la frecuencia, y como se puede apreciar la diferencia de este filtro es que presenta una forma plana y lisa, y demora en decaer después de haber llegado a la frecuencia de corte, en la tercera gráfica se presenta la señal original y en la cuarta gráfica la señal filtrada.

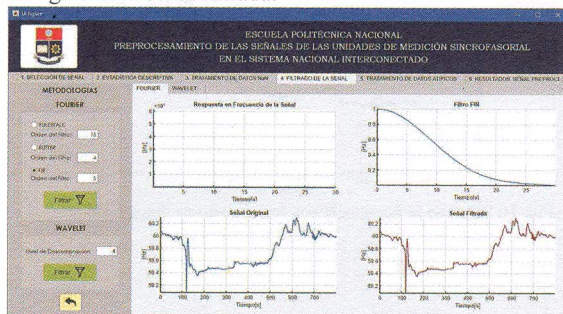


Fig. 18 Sección 4: Filtrado de la señal

5) Resultado de la sección 5: TRATAMIENTO DE DATOS ATÍPICOS

En la Tabla 4.4. se presenta el cálculo RMSE y R2 de los diferentes métodos empleados. Como se puede observar el método LOESS presentan el valor más bajo de RMSE y casi tiende a ser 1 su coeficiente de determinación, por lo que, en comparación a los demás modelos, este resulta ser el modelo óptimo para su presente aplicación.

TABLA V.
CÁLCULO DE LOS ERRORES DE LOS MÉTODOS PARA EL TRATAMIENTO DE DATOS ATÍPICOS EN LA SEÑAL DE FRECUENCIA

Metodología	RMSE	R2
LOWESS	0,02573	0,9909
LOESS	0,0189	0,9951
SGOLAY	0,0201	0,9944
K-MEANS	0,0397	0,9782

La Regresión cuadrática local “LOESS”, calcula y ajusta un polinomio cuadrático sobre la señal de frecuencia, mediante una ventana deslizante de 1030 puntos de datos, tras realizar varias pruebas empíricas se logró determinar este valor de ventana como el óptimo para el cálculo de este algoritmo. Se expone el resultado en la Fig. 19, la señal que se presenta en color rojo es la nueva función estimada.



Fig. 19 Sección 5: Tratamiento de datos atípicos

6) Resultado de la sección 6: SEÑAL PREPROCESADA

Después de todo el proceso ejecutado por las técnicas de limpieza de datos para tratar la señal de frecuencia de la PMU AGOYÁN - BAÑOS 1, en esta sección se presenta finalmente el resultado obtenido de la señal junto con la señal original escogida en la sección 1, como se puede observar en la Fig. 20. Adicionalmente, se incluye en esta sección los valores de la raíz cuadrada del error medio cuadrático y el coeficiente de determinación, los cuales fueron calculados entre la señal final y la original, para de esta manera poder comparar el rendimiento de cada método, y así seleccionar los más adecuados con la finalidad de no perder la forma de la señal original.



Fig. 20 Sección 6: Señal Preprocesada

VIII. CONCLUSIONES

Se realizó el preprocesamiento de los datos de las mediciones sincrofásorales de las PMUs del Sistema Nacional Interconectado, mediante la aplicación de los enfoques que comprenden el proceso de las técnicas de limpieza de datos, tal que, se resolvió las inconsistencias que presentan cada una de las diferentes señales temporales.

Los enfoques que comprenden las técnicas de limpieza de datos cumplieron una función específica a través de las diferentes metodologías propuestas, lo cual permitió el diagnóstico, detección e imputación de los datos anómalos de las diferentes mediciones sincrofásorales. Por consiguiente, se influyó en la calidad de los datos, y así se obtuvo una base de datos fidedigna.

Los datos faltantes NaN implican una reducción del tamaño del conjunto de datos de la señal y, en consecuencia, un aumento en el error de las estimaciones de los parámetros. Por lo que, provocan una falta de respuesta y dan lugar a estimaciones sesgadas, siendo necesario la imputación de estos

datos mediante mecanismos que permiten estimar valores de reemplazo, los cuales retienen las características generales del conjunto de datos.

Los métodos de imputación de los datos atípicos pueden ser altamente influenciados por observaciones extremas, y estos dependen específicamente de la longitud de la ventana. Al seleccionar una menor longitud de la ventana puede evitar la convergencia del algoritmo y ser más resistente a la presencia de valores atípicos, mientras que a mayor longitud de la ventana la curva ajustada puede ser muy suavizada y perder las características esenciales de la forma original de la señal.

REFERENCIAS

- [1] Aharon de la Torre, "Análisis técnico para la implementación de un sistema de monitoreo de área extendida (WAMS) en el Sistema Nacional Interconectado del Ecuador", Universidad Politécnica Salesiana, Septiembre del 2013.
- [2] R. P. Cody, "Cody's Data Cleaning Techniques", vol. 2, Editorial SAS, p. 272, 2008.
- [3] H. N. Akouemo, "Data Cleaning in the Energy Domain", Mayo, Marquette University, 2015.
- [4] Monti, C. Muscas, F. Ponci, "Phasor Measurement Units and Wide Area Monitoring Systems: From the Sensors to the System", Editorial ELSEVIER, pp. 2-100, 2016.
- [5] S. R. Bhide, "Digital power system protection", PHL, pp. 216-236, 2014.
- [6] G. Phadke, J. S. Thorp, "Power Electronics and Power Systems Synchronized Phasor Measurements and Their Applications Second Edition", vol. 2, Springer, pp. 84-220, 2017.
- [7] J. Momoh, "Smart Grid Fundamentals of Design and Analysis", John Wiley & Sons, pp. 20-215, 2012.
- [8] J. Ekanayake, K. Liyanage, J. Wu, A. Yokoyama, N. Jenkins, "Smart Grid : Technology and Applications.", Editorial Wiley, pp. 173-186, 2012.
- [9] L. Armas, "Estudio Técnico para la Determinación de la Ubicación Óptima de Unidades de Medición Fasorial (PMU) en el Sistema Nacional Interconectado S.N.I. Basado en Criterios de Observabilidad ante Contingencias", Escuela Politécnica Nacional, Quito, 2016.
- [10] M. Campos, M. Arias, "Ubicación Óptima de Unidades de Medición Fasorial aplicando Swarm Intelligence", IEEE, 2010.
- [11] N. Granda, "Esquema Adaptable de Separación Controlada en Islas para Sistemas Eléctricos", Universidad Nacional de San Juan, Argentina.
- [12] P. Mallet, J. A. Frazao, "PMU based situation awareness for smart distribution grids", Octubre, Grenoble Alpes University, 2015.
- [13] IEEE, "Dynamic Vulnerability Assessment and Intelligent Control for Sustainable Power Systems", Wiley & Sons, pp. 9-147, 2002.
- [14] K. Chandragupta, V. Ramkumar, "Phasor Measurement Units in Power System Networks", India, College of Technology, pp. 120-125, 2014.
- [15] J. Han, M. Kamber, J. Pei, "Data Mining Concepts and Techniques", vol. 3, Editorial
- [16] T. D. Johnson Theodore, "Exploratory Data Mining and Data Cleaning", vol. 39, Wiley & Sons, pp. 17-200, 2008
- [17] D. Swagatam, A. Ajith, K. Amit, "Metaheuristic Clustering", vol. 178, Springer, p. 249.
- [18] R. Nisbet, J. Elder, G. Miner, "Handbook of Statistical Analysis and Data Mining Applications", Editorial ELSEVIER, p. 823, 2007.
- [19] T. De Waal, J. Pannekoek, S. Scholtus, S. Netherlands, "Survey Methodology Statistical Data Editing and Imputation", Wiley & Sons, pp. 1-57, 2011.
- [20] J. Freund, J. Wilson, "Statistical Methods", vol. 2, Academic Press,

pp. 1-75, 1966.

- [21] D. Montgomery, C. Jennings, M. Kulahci, "Introduction to Time Series Analysis and Forecasting", Wiley & Sons, pp. 18-139, 2008.
- [22] D. S. Chi Fung, "Methods for the Estimation of Missing Values in Time Series", Cowan University, p. 332, 2006
- [23] R. Forest, B. Menke, "Environmental Data Analysis with MATLAB", vol. 2, Editorial ELSEVIER, pp. 1-20, 2012.
- [24] J. L. Gutierrez, A. I. Schlögl, "Comparación de métodos autorregresivos para la detección de artefactos en señales de EEG", University of Technology Graz, Austria, 2010.
- [25] W. Bowman, A. Azzalini, "Applied Smoothing Techniques for Data Analysis", Clarendon Press, pp. 75-91, 1997.



Diana Fernanda Guevara Estacio

nació en Quito el 10 de septiembre de 1992. Su estudio secundario lo realizó en el colegio 'Técnico Humanístico Experimental Quito', con la mención de Abanderada del Pabellón Nacional, graduándose con el título de bachiller en Físico

Matemático y mejor egresada en el año 2011, sus estudios superiores los realiza en la Escuela Politécnica Nacional en la carrera de Ingeniería Eléctrica, secretaria de la asociación de los estudiantes de eléctrica 2015-A, miembro de la IEEE sección PES en el año 2016. Ha trabajado como investigador en el laboratorio de tiempo real en el Operador Nacional de Electricidad. Sus áreas de interés son sistemas eléctricos de potencia y eficiencia energética (diana.guevara@epn.edu.ec).



Nelson Victoriano Granda Gutiérrez

obtuvo el título de Ingeniero Eléctrico en la Escuela Politécnica Nacional de Quito, en 2006. Se ha desempeñado como Ingeniero Eléctrico en el Centro Nacional de Control de Energía (CENACE) y posteriormente en el departamento de Optimización de la Generación Eléctrica (OGE) de

Petroamazonas EP. Actualmente se encuentra realizando su trabajo de investigación previo a la obtención de título de PhD en el Instituto de Energía Eléctrica de la Universidad Nacional de San Juan (IEE - UNSJ). Sus áreas de interés son análisis y control de sistemas de potencia en tiempo real. Actualmente trabaja en CELEC EP, Unidad de Negocio TRANSELECTRIC (nelson.granda @epn.edu.ec).



Franklin Lenin Quilumba Gudiño

obtuvo el título de Ingeniero Eléctrico en la Escuela Politécnica Nacional en Quito, Ecuador, en el 2008. Realizó sus estudios de posgrado en la Universidad de Texas Arlington, en Arlington, Estados Unidos de América, donde obtuvo el grado de Master of Science in Electrical Engineering y el título de

Doctor of Philosophy Ph.D. en el 2014 (quigufrale@gmail.com).