

INTERPRETACIÓN DE SEÑALES PARA LA DETECCIÓN DE OBSTRUCCIONES  
EN INSPECCIONES INSTRUMENTADAS DE TUBERÍAS

JUAN CAMILO BOHORQUEZ REYES



UNIVERSIDAD AUTONOMA DE BUCARAMANGA – UNAB  
FACULTAD DE INGENIERIAS FISICOMECANICAS  
INGENIERIA MECATRONICA  
BUCARAMANGA  
2015

Universidad Autónoma De Bucaramanga  
Facultad de Ingenieras Fisicomencanicas  
Línea de Automatización y Control

Interpretación de señales para la detección de obstrucciones en inspecciones  
instrumentadas de tuberías

JUAN CAMILO BOHORQUEZ REYES  
U00061150

**Director:** PhD Carlos Julio Arizmendi

Bucaramanga  
2015

**Nota de Aceptación**

---

---

---

---

---

---

---

**Director**

---

**Evaluador**

**30 Enero 2015**

## Tabla de contenidos

1.	Objetivos .....	1
1.1.	General: .....	1
1.2.	Específicos: .....	1
2.	Introducción .....	2
3.	Antecedentes .....	4
4.	Diseño Metodológico .....	5
5.	Avances .....	6
5.1.	Pre-procesamiento de señales .....	7
5.2	Procesamiento de datos.....	14
5.2.1	Extracción y visualización de estadísticos.....	18
5.2.2	Correlaciones .....	18
6.	Clasificación de las oscilaciones periódicas.....	24
6.1	Red Neuronal.....	44
6.1.1	Resultados.....	48
6.1.2	Post-clasificación.....	52
6.2	Comparación de métodos .....	57
6.2.1	Interfaz grafica .....	70
6.2.2	Análisis de eventos en marcadores .....	72
6.2.2.1	Marcadores TA1, TA2 y GR1.....	73
6.2.2.1	Marcador TA3.....	74
6.2.2.2	Marcadores TA4, TA5, TA6, TA7, TA8 y GR3 .....	75
6.2.2.3	Marcador GR2 .....	76
6.3	Metodología Caliper .....	77
7.	Conclusiones.....	78
8.	Bibliografía .....	80

## 1. Objetivos

### 1.1. General:

Diseñar clasificadores basados en Redes Neuronales y/o Maquinas de Soporte Vectorial con el fin de detectar segmentos con variación de diámetro interno en líneas de transporte para la identificación de patrones y tendencias en datos de deformación mecánica.

### 1.2. Específicos:

- Realizar análisis estadísticos e identificar el problema con el fin de caracterizar los datos y comprender la información almacenada en las señales mecánicas tomadas por el sensor caliper del ITION.
- Desarrollar e implementar técnicas provenientes del procesamiento de las señales, basándose en la Transformada *Wavelet* y reducción de la dimensionalidad para encontrar las variables relevantes del sistema.
- Desarrollar algoritmos para realizar la clasificación de las variables relevantes del sistema, con el fin de detectar las diferentes clases de datos que representen características físicas de la tubería, por medio de métodos como Maquinas de Soporte Vectorial y/o Redes Neuronales.
- Identificar el método óptimo para la detección de obstrucciones en la tubería mediante la comparación de resultados obtenidos por medio de los métodos implementados.

## 2. Introducción

En un oleoducto o gasoducto se requieren kilómetros de tubos con el fin de transportar el material hasta el centro de distribución, debido a la gran distancia que se debe recorrer, existen complicaciones al producirse una falla como una obstrucción dentro de alguna sección de la tubería, donde el procedimiento para realizar la inspección de la zona que se vio comprometida puede retrasar la producción de todo el complejo. Si falla la estructura de la tubería y se produce una fuga antes de que se detecte a tiempo, debido a las pérdidas de material, se produciría un fuerte daño ambiental, afectando el ecosistema y las poblaciones cercanas.

En la industria de los hidrocarburos se implementa un método de inspección de tuberías, que detecta imperfecciones tales como fisuras, abolladuras, obstrucciones, etc. Este método es conocido como Pigging, que basa en una herramienta llamada *Piping Inspection Gauge* (PIG). Este instrumento posee múltiples sensores que detectan las imperfecciones ya mencionadas.

Para implementar este procedimiento, en la mayoría de los casos, se tiene que detener la actividad en la tubería que se va a inspeccionar y el análisis de los sensores puede llegar a demorar los resultados provocando consecuencias en la producción de la empresa y además reducir la calidad de los procedimientos al no solucionar los problemas rápidamente.

Para encontrar solución a estos problemas, en conjunto con la Corporación para la Investigación de la Corrosión (CIC) en su proyecto ITION, se propuso un dispositivo PIG que viaje por el tubo junto con el material, realizando una inspección dentro de este y por medio de sistemas de medición tipo caliper se toman datos del diámetro de la tubería, para luego realizar su interpretación utilizando técnicas provenientes del procesamiento de señales basado en la Transformada *Wavelet*. Estos datos serán adaptados a las entradas de dos clasificadores basados en Redes Neuronales.

Gracias a estos métodos inteligentes es posible detectar las imperfecciones de una forma más eficiente debido a que las señales serán interpretadas artificialmente, esto puede agilizar los procedimientos de mantenimiento y prevención de errores con el fin de evitar las consecuencias ya mencionadas anteriormente, además, dicho procedimiento se realiza sin detener el funcionamiento de la tubería y de esta forma no se ve afectada la producción mientras se realiza la inspección.

Se han realizado diferentes métodos de medición con el fin de detectar las imperfecciones en una tubería, métodos tales como ultrasonido, visión, MFL donde cada uno requiere un tratamiento de datos diferente, unos más complejos que otros. El artefacto PIG es una herramienta modular que puede integrar una variedad de sensores en ella y tomar los datos de cada uno al mismo tiempo. Este sistema se utiliza comercialmente aunque el valor para implementarlo puede ser bastante costoso.

La metodología propuesta soluciona los problemas más evidentes al utilizar este método en una línea de transporte activa, como son, evitar la pérdida de dinero a causa de detener la producción y la pérdida de calidad del servicio al prevenir desastres y mal manejo de los hidrocarburos.

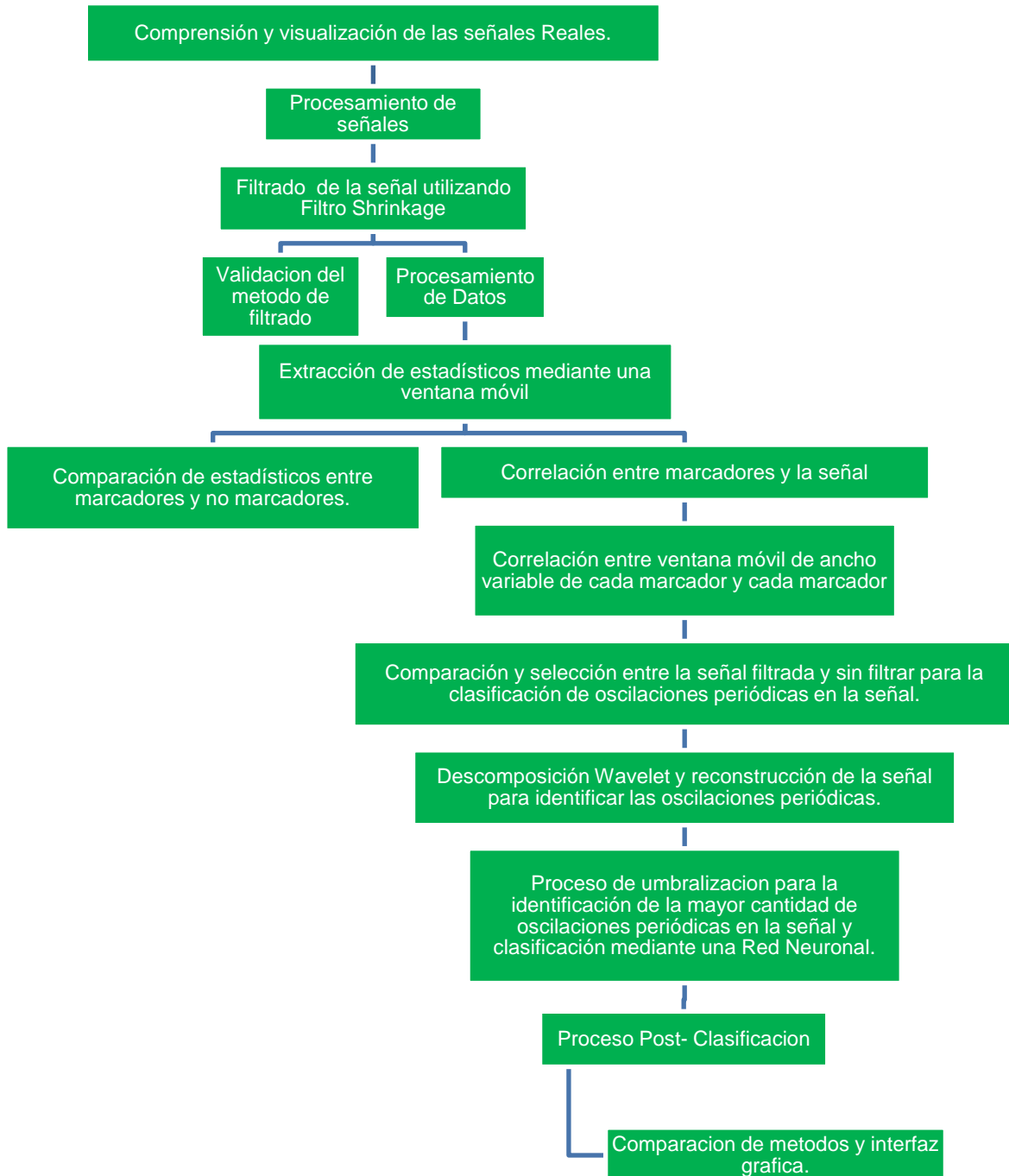
### **3. Antecedentes**

El sistema de inspección de tuberías que se habla en el presente proyecto, empezó a desarrollarse en La Corporación de Investigación para la Corrosión (CIC), quien realizó la investigación sobre los sistemas utilizados para hallar imperfecciones en tuberías, además de diseñar el prototipo de cada uno de los sistemas de medición y posteriormente la construcción de cada uno de ellos.

El CIC y la Universidad Autónoma de Bucaramanga (UNAB) realizaron un convenio con el fin de profundizar en el tema de tratamiento de las señales, donde estas fueron tomadas por cada uno de los instrumentos y además de implementar métodos de tratamiento y clasificación de datos con el fin de encontrar información pertinente para detectar las imperfecciones en las tuberías.



## 4. Diseño Metodológico



## 5. Avances

La corporación para la investigación de la corrosión realizó una toma de datos utilizando el dispositivo ITION en un tubo de 23 Km donde se tomaron un paquete de señales utilizando una tasa de muestreo de 300Hz.

Las señales que se tomaron corresponden a 8 sensores inerciales, 8 sensores de MFL, 2 señales obtenidas por medidores tipo caliper y una señal de odómetro.

La señal del odómetro almacena la distancia recorrida por el ITION en todo su recorrido. Debido a que el artefacto estuvo detenido por un tiempo al principio del trayecto, se eliminan los primeros datos de la señal para obtener el recorrido real del instrumento. En la Figura 1 se puede observar la señal del odómetro original y la tratada.

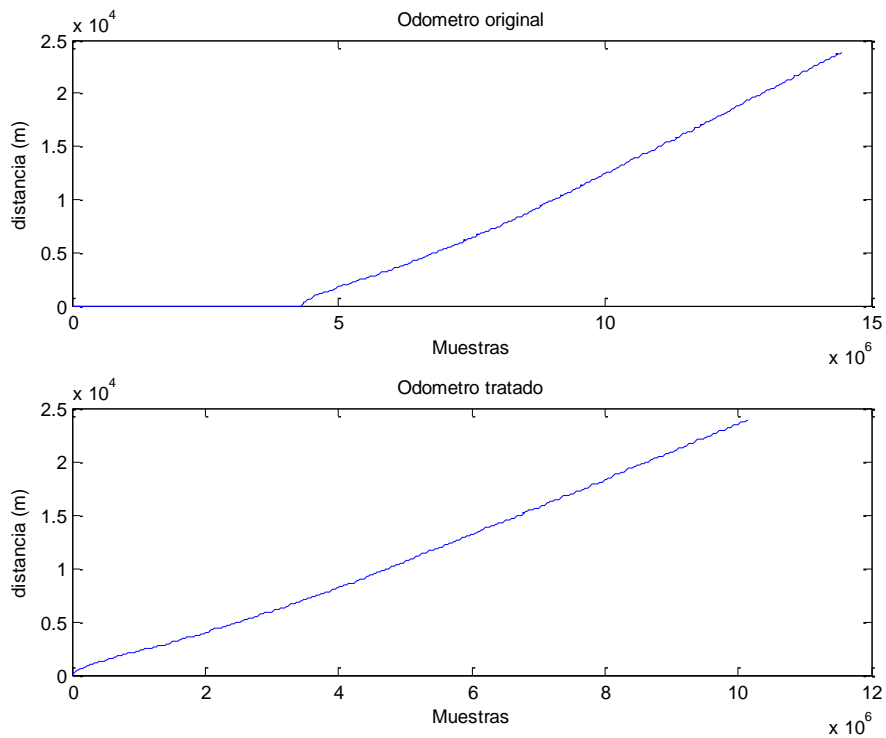


Figura 1. Señal de odómetro. Fuente: Autor.

A continuación se observan en las Figura 2 y Figura 3 las señales caliper con la que se implementara la metodología en este documento.

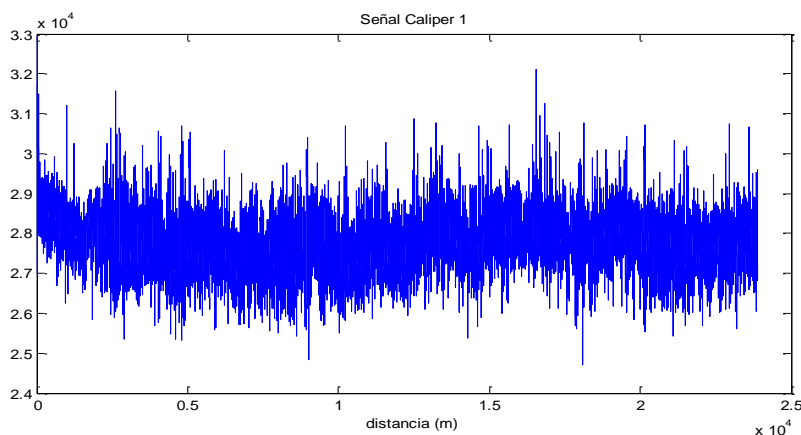


Figura 2. Señal Caliper 1. Fuente: Autor.

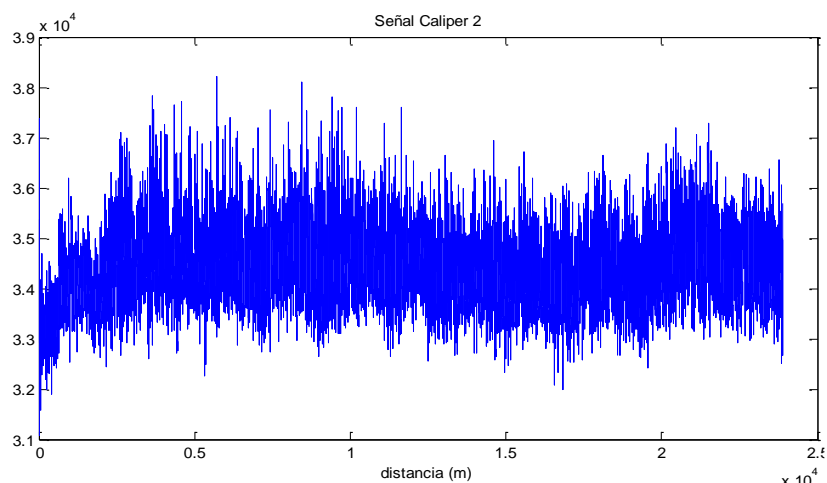


Figura 3. Señal Caliper 2. Fuente: Autor.

### 5.1. Pre-procesamiento de señales

En las señales en bruto proporcionadas por la Corporación para la Investigación de la Corrosión obtenidas tras la toma de datos realizada por el dispositivo ITION, poseen una cantidad de ruido considerable por ello se propone una metodología de pre-procesamiento con la cual encontrara la configuración optima de filtrado basado en un filtro *Shrinkage* [13].

La metodología propuesta para encontrar la configuración adecuada del filtro *Shrinkage* se basa en el criterio del menor MSE [10] entre la señal original y la filtrada. La señal se filtra utilizando la transformada *Wavelet*.

Para determinar el nivel de descomposición adecuado para el filtro se utiliza el criterio de la entropía de *Shannon*, midiendo el nivel de incertidumbre.

En la Tabla 1 se pueden observar el MSE y el nivel de descomposición utilizando 36 *Wavelets* madre diferentes con su respectivo nivel de descomposición para las dos señales Caliper.

<b>MSE y Niveles de descomposición para diferentes <i>Wavelet</i> madre para cada señal Caliper para Filtro <i>Shrinkage</i></b>				
	Caliper 1		Caliper 2	
<i>Wavelet</i> Madre	MSE	NIVEL	MSE	NIVEL
'haar'	2.30108926e-19	11	2.761492316e-19	12
'db2'	2.55827098e-17	10	3.128016126e-17	11
'db3'	9.07027191e-19	10	1.122966087e-18	11
'db4'	1.53828209e-18	9	1.974372648e-18	10
'db5'	4.61414166e-19	9	5.794310794e-19	10
'db6'	7.74151935e-19	9	1.015888823e-18	10
'db7'	4.01496409e-18	9	5.146467074e-18	10
'db8'	2.50850716e-16	9	2.205855395e-16	9
'db9'	2.43278303e-18	8	3.263592832e-18	9
'db10'	1.49949403e-21	10	2.704940356e-21	11
'bior1.3'	1.26273659e-21	9	2.405239452e-21	10
'bior1.5'	2.40145836e-21	10	4.695707161e-21	11
'bior2.2'	4.97317007e-22	9	8.955842241e-22	10
'bior2.4'	5.02085542e-22	9	9.596159682e-22	10
'bior2.6'	1.19353118e-21	9	1.909429016e-21	9
'bior2.8'	6.22392503e-21	11	1.165863055e-20	12
'bior3.1'	1.60247225e-21	10	3.082524609e-21	11
'bior3.3'	1.44811561e-21	9	2.818282722e-21	10
'bior3.5'	1.79249709e-21	9	3.512152747e-21	10
'bior3.7'	1.49883072e-21	8	3.015719023e-21	9
'bior3.9'	3.25885778e-19	9	4.254494892e-19	10
'bior4.4'	5.97231398e-19	9	7.690567139e-19	10
'bior5.5'	1.29381118e-20	9	1.150222649e-20	9
'bior6.8'	3.26198931e-19	10	4.136199866e-19	11
'coif1'	3.88042643e-17	9	5.068147064e-17	10
'coif2'	9.37537509e-20	9	9.056552404e-20	9
'coif3'	1.03335148e-16	8	1.353906339e-16	9
'coif4'	4.86415281e-12	8	6.366151660e-12	9
'coif5'	2.30108926e-19	11	2.761492316e-19	12

'sym2'	2.55827098e-17	10	3.128016126e-17	11
'sym3'	1.04122217e-19	10	1.394646181e-19	11
'sym4'	1.85987291e-20	9	3.235455317e-20	10
'sym5'	1.56759413e-19	9	2.111162205e-19	10
'sym6'	1.09766211e-19	9	1.493959082e-19	10
'sym7'	3.17553874e-22	9	5.677954804e-22	9
'sym9'	2.30108926e-19	11	2.761492316e-19	12

Tabla 1. MSE y Nivel de descomposición para cada *Wavelet* madre y para cada Señal Caliper para Filtro *Shrinkage*. Fuente: Autor.

Al obtener el mínimo MSE de los resultados anteriores se obtuvo que la mejor configuración de *Wavelet* madre y niveles de descomposición del filtro *Shrinkage*, utilizando la *Wavelet Symlets 9* a 9 niveles de descomposición.

Luego de obtener la *Wavelet* madre y el nivel de descomposición óptimo, se deben encontrar otros parámetros del filtro *Shrinkage* que son el *Universal Threshold*, *Threshold Scaling* e *Implementation of Thresold*. [14] utilizando el criterio del MSE para las dos señales Caliper.

*Universal Threshold* es un umbral que se calcula por medio de estimadores estadísticos. [14] Las funciones que se pueden escoger en *Matlab® Wavelet toolbox* para este parámetro son, *Sqtwolog*, *rigsure*, *heursere*, *minmax*.

*Threshold Scaling* es un parámetro que pondera los umbrales y se consideran de tres tipos, [13] *Sln*, *One*, *Mln*. Las ecuaciones para cada uno son Ec 1, Ec 2 y Ec 3 respectivamente.

$$\sigma = \frac{\text{Media}(1er \text{ Nivel de descomposicion})}{0.6745}$$

Ec 1. *Threshold Scaling Sln*. Fuente: [13]

$$\sigma = 1$$

Ec 2. *Threshold Scaling One*. Fuente: [13]

$$\sigma = \frac{\text{Media}(1er\ Nivel\ de\ descomposicion)}{0.6745}$$

Ec 2. *Threshold Scaling Min.* Fuente: [13]

En la metodología propuesta se obtuvo la configuración óptima de parámetro para lograr el mínimo MSE y se observa en la Tabla 2.

<b>Señal</b>	<b><i>Universal Threshold</i></b>	<b><i>Threshold Scaling</i></b>	<b><i>Implementation of the Threshold</i></b>
Caliper 1	Minmax	One	Hard
Caliper 2	Minmax	One	Hard

Tabla 2. Parámetros óptimos del Filtro *Shrinkage*

Al determinar todos los parámetros del filtro *Shrinkage* se debe calcular el ancho de ventana óptimo para trabajar las señales ya que el tamaño de las mismas es bastante grande como para realizar el filtro a la señal completa, debido a esto se realiza el filtrado por ventanas para cada señal.

La metodología para obtener el ancho de ventana óptimo se basa en el criterio de la relación señal a ruido (por siglas en inglés, SNR) [13]. Donde se probaron anchos de ventana empezando con 100.000 muestras hasta 2.000.000 de muestras con un aumento de 100.000 por cada iteración. Luego se compraron los SNR obtenidos. Los resultados obtenidos por esta prueba se pueden observar en la Figura 4 para la señal Caliper 1 y la Figura 5 para la señal Caliper 2.

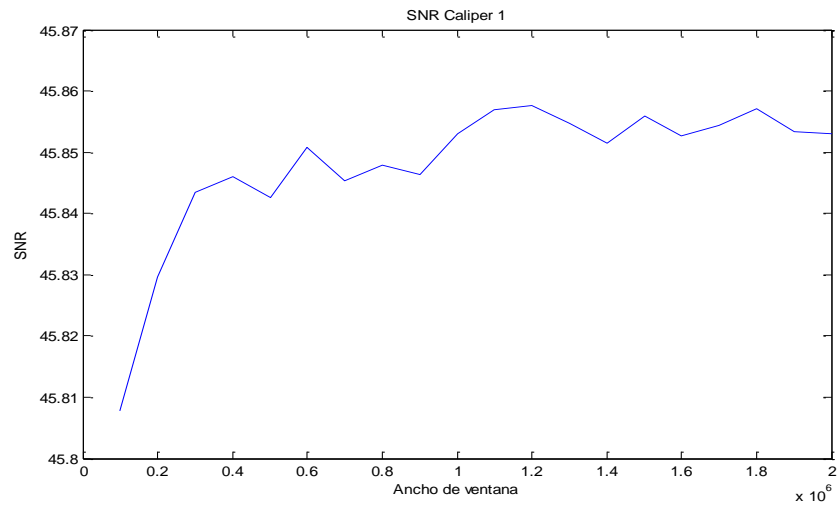


Figura 4. SNR vs Ancho de Ventana para la Señal Caliper 1. Fuente: Autor.

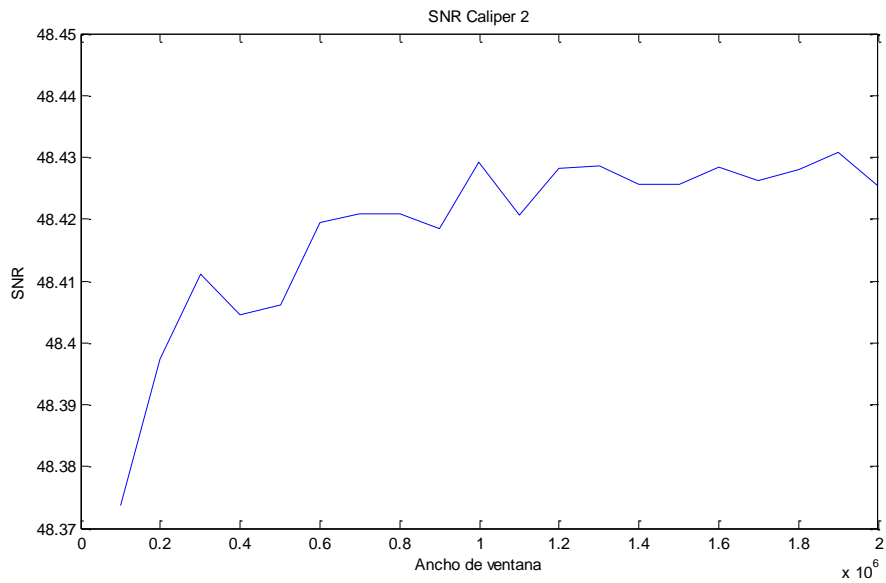


Figura 4. SNR vs Ancho de Ventana para la Señal Caliper 2. Fuente: Autor.

Como se puede observar en la Figura 4 y en la Figura 5, los SNR para todos los anchos de ventana varían muy poco, por esta razón el ancho de ventana que se utilice no genera un cambio mayor en el resultado del filtrado.

Debido a lo anterior se escogió una ventana de 1'048.576, con el fin de reducir el costo computacional y evitar errores matemáticos debido a que este valor es el numero en base 2 más cercano a 1'000.000.

Luego de realizar el proceso de filtrado utilizando el filtro *Shrinkage* configurado con los parámetros óptimos que se encontraron anteriormente, Se obtuvieron las señales filtradas que se observan en las Figuras 5 y Figura 6 donde además se comparan con las señales originales.

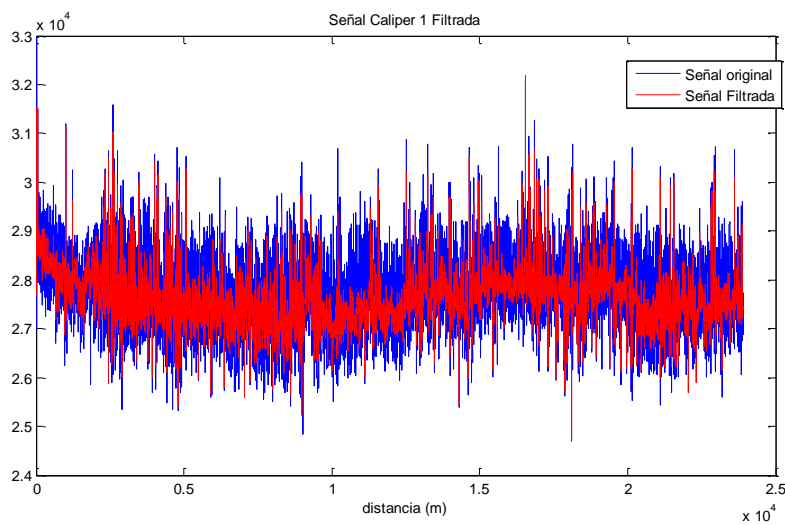


Figura 5. Señal original Vs Señal filtrada Caliper 1. Fuente: Autor

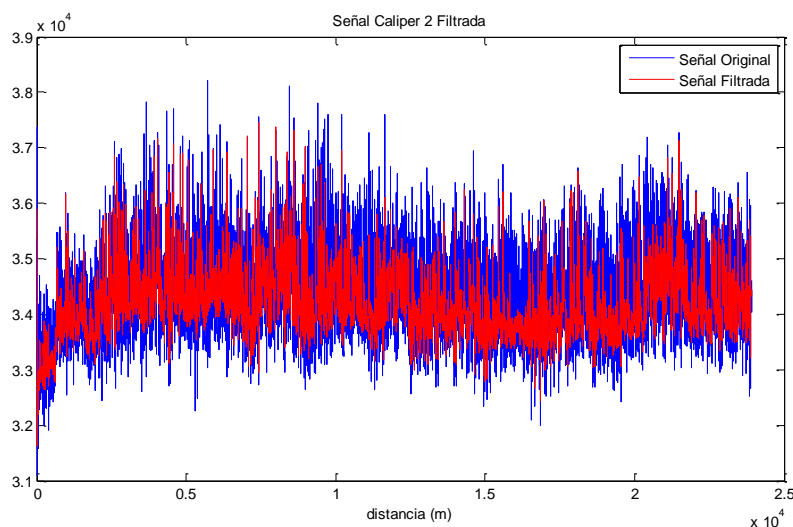


Figura 6. Señal original Vs Señal filtrada Caliper 2. Fuente: Autor



Luego del proceso de filtrado se deben identificar los marcadores que indican la presencia de un defecto en la abolladura. La corporación para la investigación de la corrosión proporciono una distancia que corresponde al centro de un rango de 14 metros donde se encuentra el defecto.

Cada uno de estos rangos que corresponden a un marcador, se dividen en dos grupos, para las señales Caliper. Estos grupos se identifican con los nombres TA y GR. En la tabla 3 se puede observar los marcadores proporcionados por la Corporación para la Investigación de la corrosión.

<b>Marcadores señales Caliper</b>	
<b>Nombre de la etiqueta</b>	<b>Rango de distancia del marcador (m)</b>
TA1	7.27 ± 7
TA2	8.25 ± 7
TA3	12582.44 ± 7
TA4	20265.49 ± 7
TA5	23856.97 ± 7
TA6	23857.32 ± 7
TA7	23857.66 ± 7
TA8	23858.01 ± 7
GR1	10.60 ± 7
GR2	30.01 ± 7
GR3	23855.97 ± 7

Tabla 3. Marcadores de las señales Caliper. Fuente: CIC

Al obtener los marcadores se pasó a realizar una inspección visual de las mismas con el fin de determinar una zona de dicho rango que se repita en cada marcador.

En el anexo 1 se observan cada uno de los marcadores filtrados y sin filtrar en las dos señales caliper.

## 5.2 Procesamiento de datos

Según lo anterior, las señales en bruto no proporcionan diferencias significativas entre los marcadores y la señal. Con el fin de encontrar alguna diferencia se propuso la obtención de estadísticos para cada una de las señales obtenidas por el dispositivo ITION, en primera instancia se compararan los boxplot de los marcadores con el resto de la señal para verificar que tan separables son con respecto a ella, esto aplica para las dos señales caliper.

Debido a que algunos marcadores del mismo tipo se solapan entre si se realizó el análisis tomando estos marcadores como uno solo.

En la Figura 7 se encuentra el conjunto de datos del marcador TA1 para las dos señales y el conjunto de datos de las muestras que no poseen defecto para las dos señales.

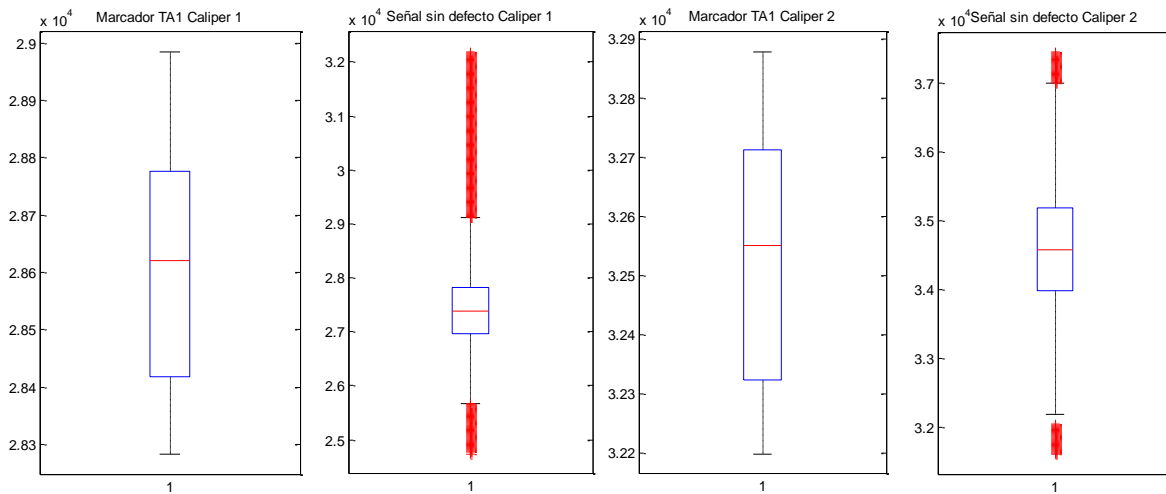


Figura 7. Boxplot Marcador TA1 y Boxplot de las muestras de la señal que no poseen abolladuras.

En la Figura 8 se encuentra el conjunto de datos del marcador TA2 para las dos señales y el conjunto de datos de las muestras que no poseen defecto para las dos señales.

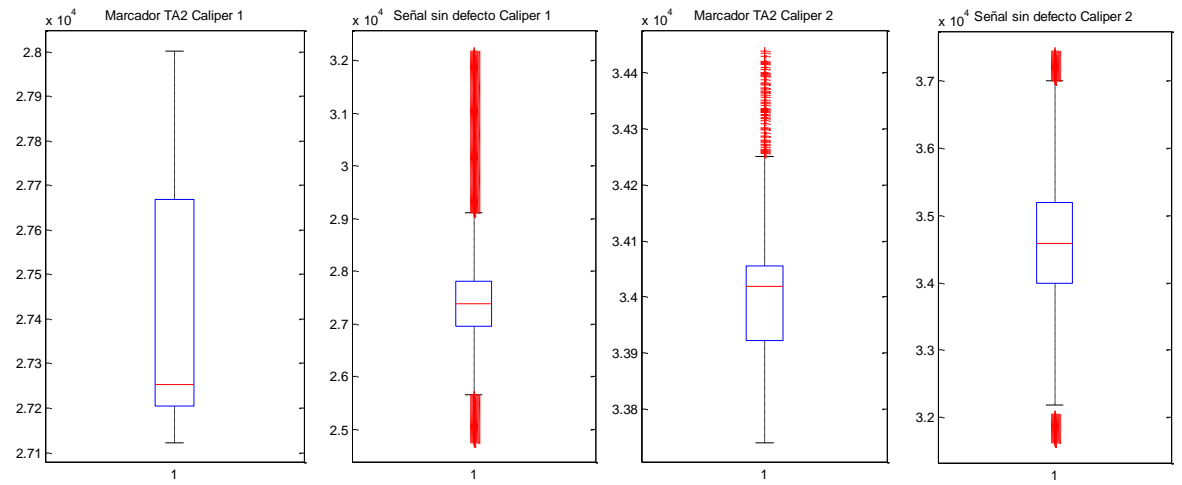


Figura 8. Boxplot Marcador TA2 y Boxplot de las muestras de la señal que no poseen abolladuras.

En la Figura 9 se encuentra el conjunto de datos del marcador TA3 para las dos señales y el conjunto de datos de las muestras que no poseen defecto para las dos señales.

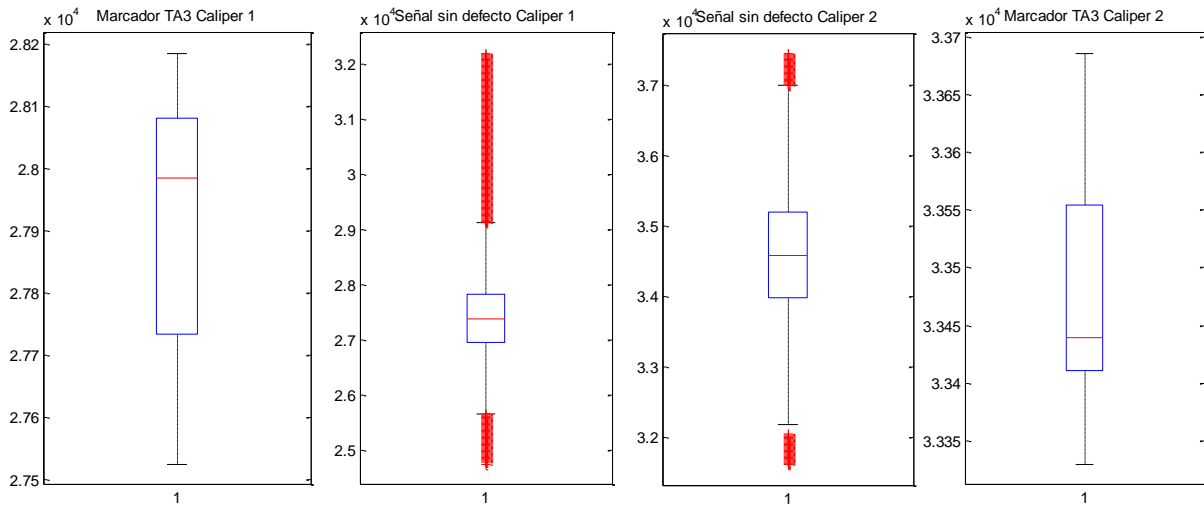


Figura 9. Boxplot Marcador TA3 y Boxplot de las muestras de la señal que no poseen abolladuras

En la Figura 10 se encuentra el conjunto de datos del marcador TA4 para las dos señales y el conjunto de datos de las muestras que no poseen defecto para las dos señales.

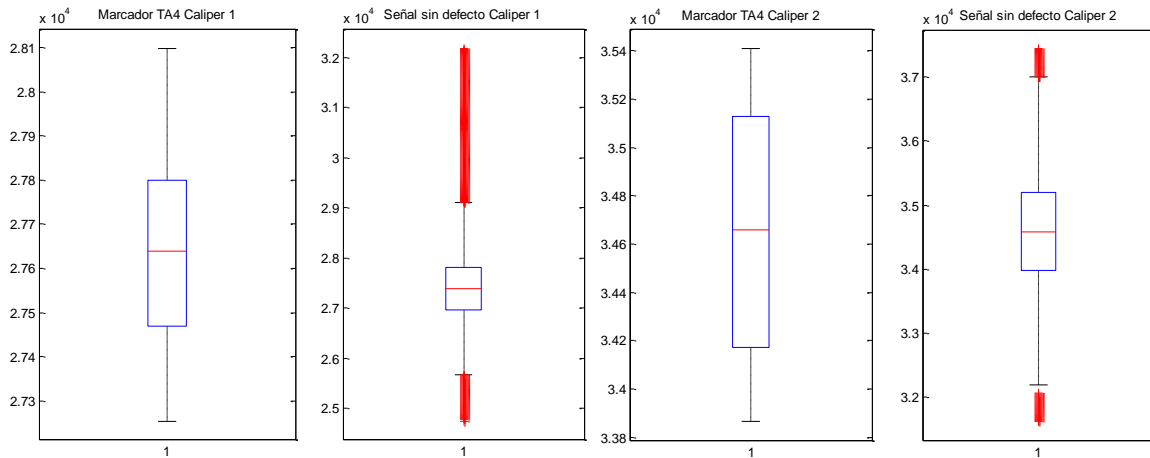


Figura 10. Boxplot Marcador TA4 y Boxplot de las muestras de la señal que no poseen abolladuras

En la Figura 11 se encuentra el conjunto de datos del marcador GR1 para las dos señales y el conjunto de datos de las muestras que no poseen defecto para las dos señales.

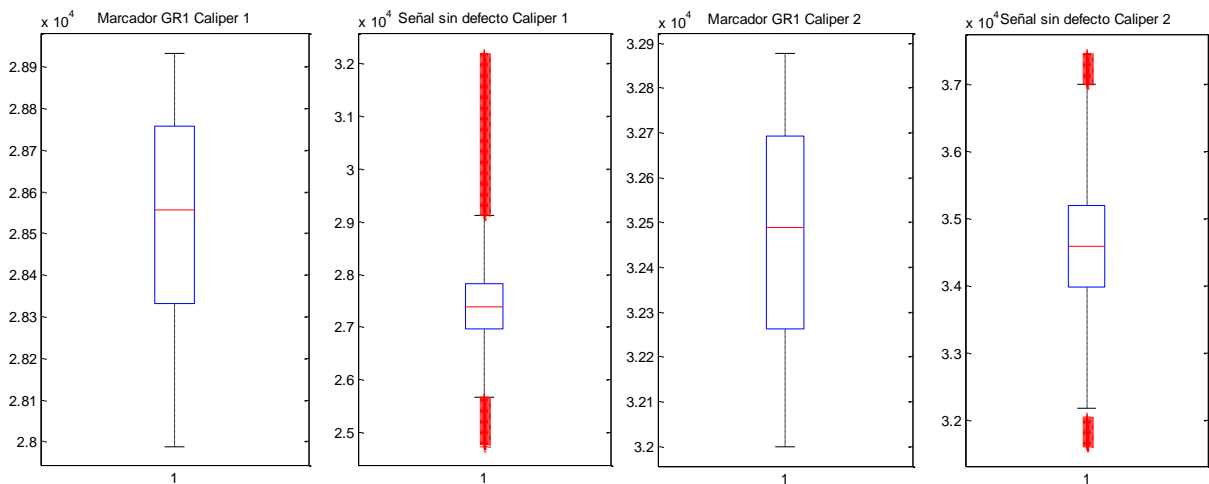


Figura 11. Boxplot Marcador GR1 y Boxplot de las muestras de la señal que no poseen abolladuras

En la Figura 12 se encuentra el conjunto de datos del marcador GR2 para las dos señales y el conjunto de datos de las muestras que no poseen defecto para las dos señales.

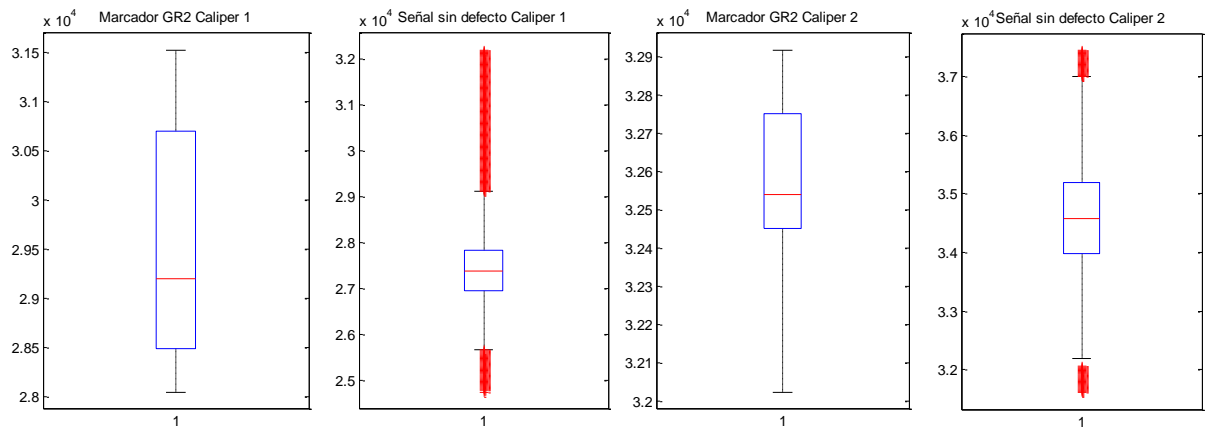


Figura 12. Boxplot Marcador GR2 y Boxplot de las muestras de la señal que no poseen abolladuras

En la Figura 13 se encuentra el conjunto de datos del marcador GR3 para las dos señales y el conjunto de datos de las muestras que no poseen defecto para las dos señales.

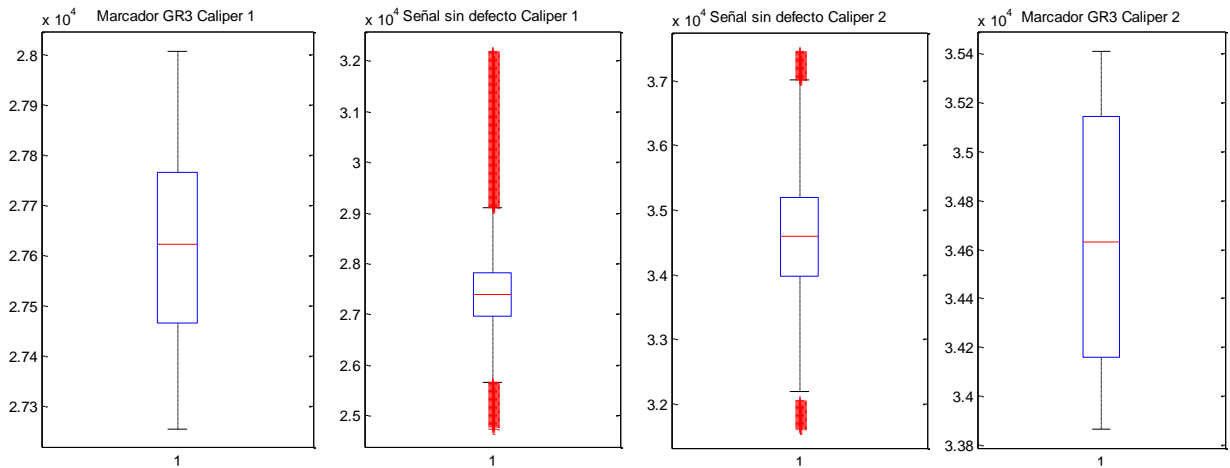


Figura 13. Boxplot Marcador GR3 y Boxplot de las muestras de la señal que no poseen abolladuras

### **5.2.1 Extracción y visualización de estadísticos**

Para obtener los estadísticos de las señales, se utilizó un método que consiste en una ventana deslizante de 14 metros de ancho que se desliza por toda la señal a un paso de una muestra. El ancho de esta ventana se escoge teniendo en cuenta los marcadores donde se encuentran los defectos, tienen este tamaño, así garantiza que se tomen estadísticos del marcador completo.

Se obtuvieron tres estadísticos. La media, la desviación estándar y el coeficiente de variabilidad. Este último se obtiene al realizar la división entre la desviación estándar y la media.

Los estadísticos de cada uno de los marcadores se pueden observar en el Anexo 2.

### **5.2.2 Correlaciones**

En esta sección se realiza un análisis de correlación utilizando el método anterior. Por medio de una ventana deslizante con un ancho de 14 metros y a un paso de una muestra. Se le calcula su correlación entre cada ventana y un marcador tipo TA y GR independientemente. Este procedimiento se realiza para las dos señales.

En las Figuras 14, 15, 16 y 17 se observan las muestras representativas que se escogieron de la señal filtrada para realizar el análisis de correlación utilizando la ventana deslizante para cada marcador y cada señal.

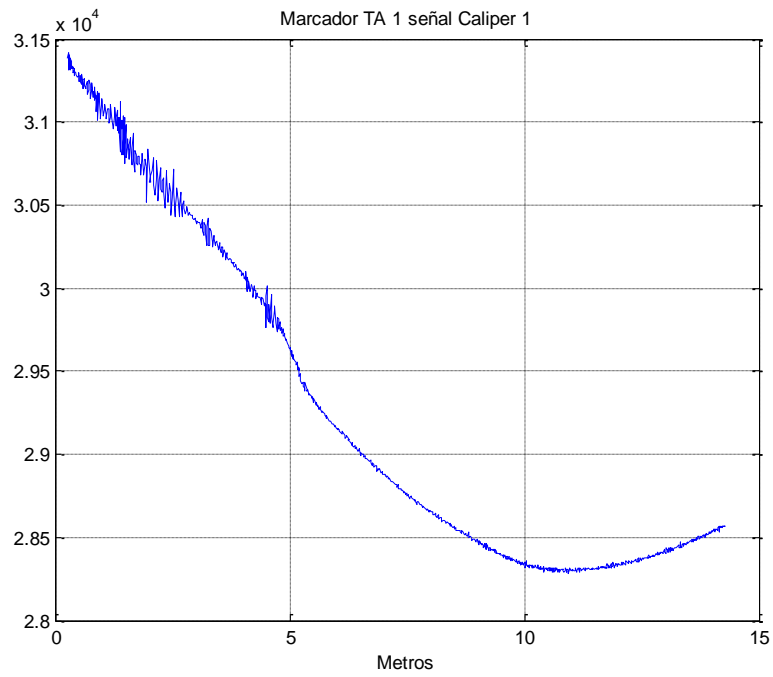


Figura 14. Muestra representativa del marcador TA para la señal Caliper 1.

Fuente: Autor.

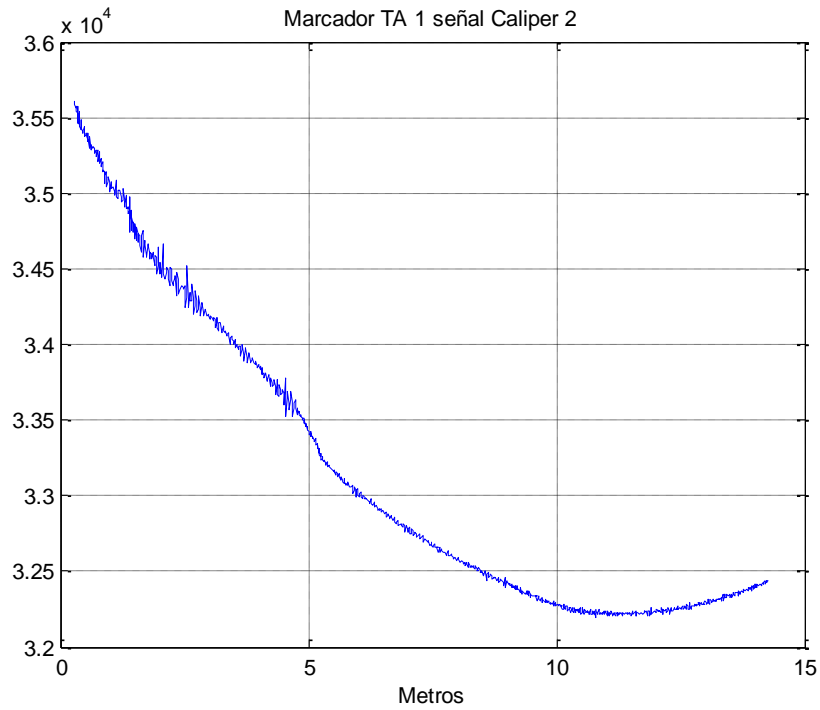


Figura 15. Muestra representativa del marcador TA para la señal Caliper 2.

Fuente: Autor.

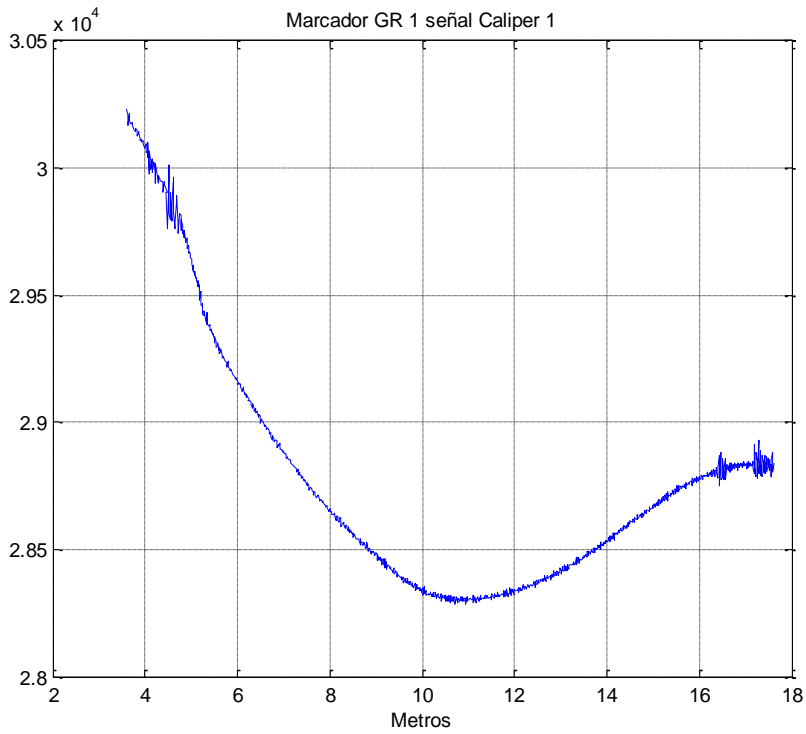


Figura 16. Muestra representativa del marcador GR para la señal Caliper 1.

Fuente: Autor.

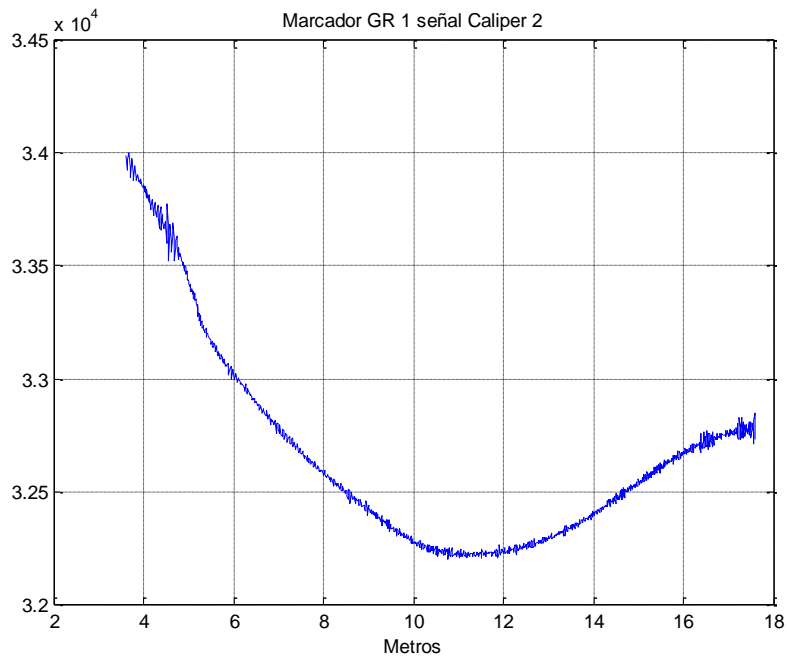


Figura 17. Muestra representativa del marcador GR para la señal Caliper 2.

Fuente: Autor.



A continuación de la Figura 18 hasta la Figura 21 se observan los resultados del análisis de correlación para las dos señales Caliper y para los dos tipos de marcadores.

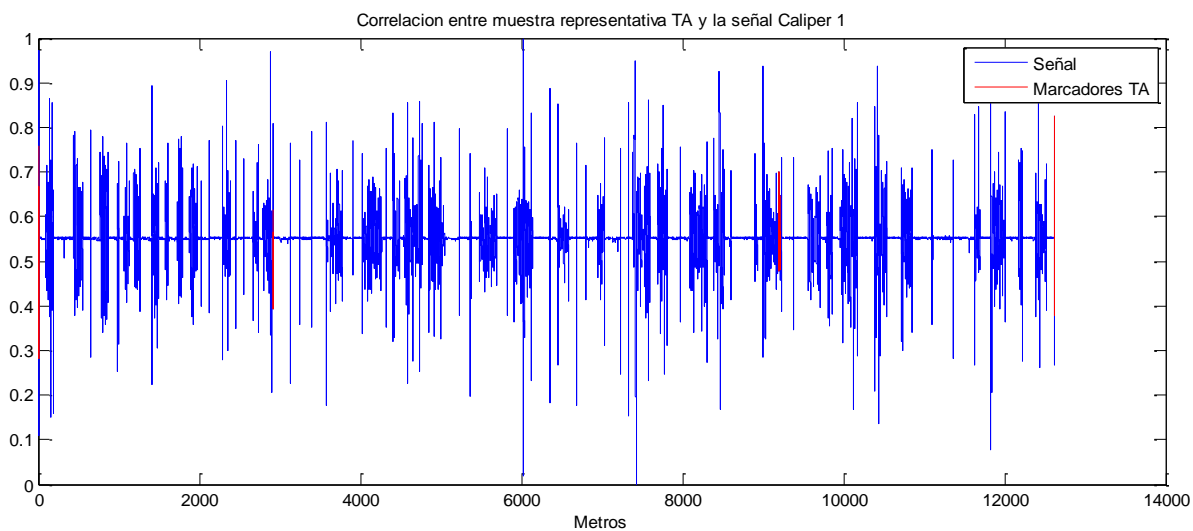
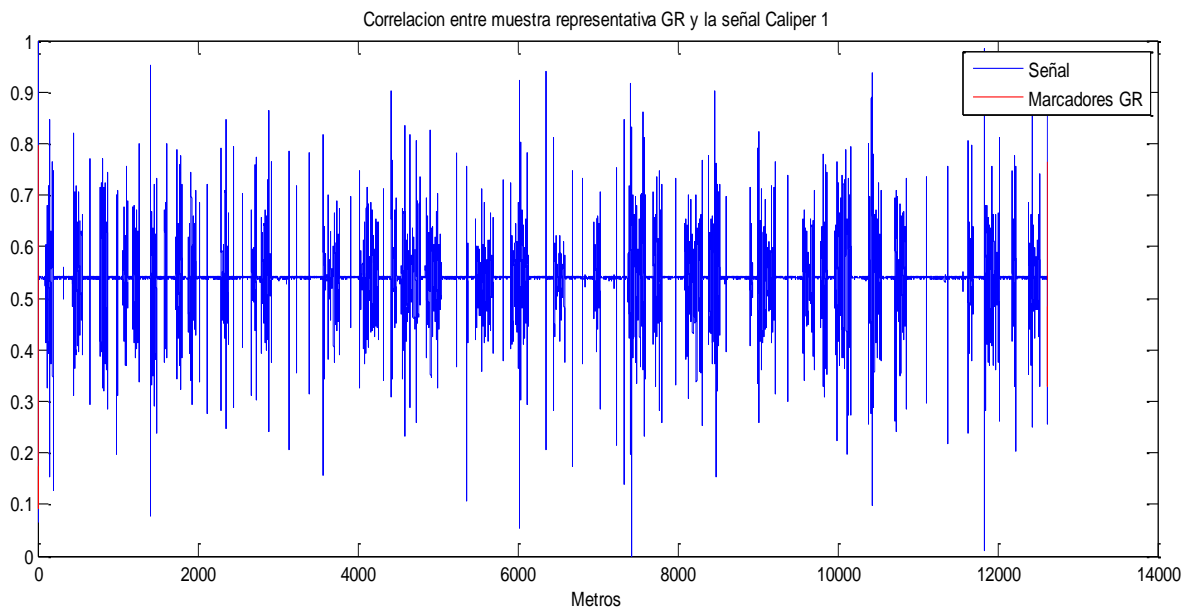


Figura 18. Correlación Señal Caliper 1 marcador TA. Fuente: Autor.



. Figura 19. Correlación Señal Caliper 1 marcador GR. Fuente: Autor.

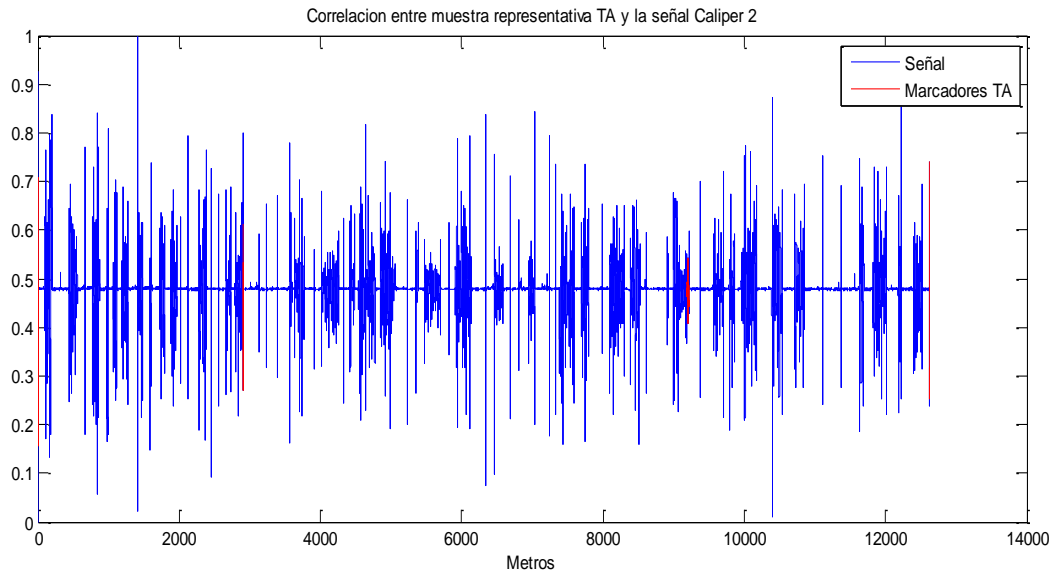


Figura 20. Correlación Señal Caliper 2 marcador TA. Fuente: Autor.

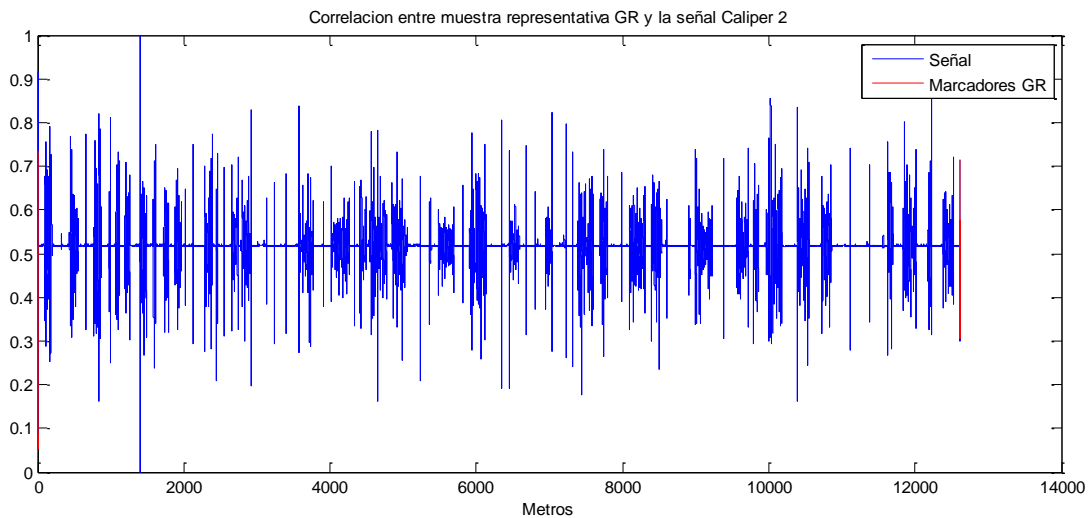


Figura 21. Correlación Señal Caliper 2 marcador GR. Fuente: Autor.

Para los resultados de correlación mostrados anteriormente se realizó un proceso de normalización basado en la Ec 4.

$$Correlacion Normalizada = \frac{Correlacion - \min(Correlacion)}{\max(Correlacion) - \min(Correlacion)}$$

Ec 4. Correlación Normalizada. Fuente: Autor.

En la etapa anterior se observó que en zonas donde no se encuentran defectos la correlación supera a las zonas donde están los marcadores. Por ello se puede afirmar que los marcadores no poseen un comportamiento similar entre si tomando el rango de 14 metros proporcionado.

Luego del análisis de correlación se propuso un método similar realizando una descomposición *Wavelet* utilizando un nivel de profundidad de 6 y la *Wavelet* madre Sym9 hallada optima anteriormente (ver 5.1 Pre-procesamiento de señales). Al obtener los niveles de descomposición se realizó el mismo análisis de correlación para cada uno, utilizando una ventana móvil comparándola con una muestra representativa de los marcadores para cada tipo y cada nivel de descomposición.

Los resultados del análisis para las dos señales caliper y para los marcadores TA y GR se observan en el Anexo 3.

Luego de esto se propuso un método de búsqueda dentro de los marcadores utilizando la correlación con el fin de encontrar un evento que no se repita dentro de cada marcador.

El procedimiento se realizó con cada uno de los marcadores y comparando una ventana deslizante cuyo ancho varía cada vez que se realiza el corrimiento por toda la señal, desde 1 cm hasta 30 cm, con otra ventana deslizante del mismo ancho que se desplaza por el mismo marcador y así determinar si existe un evento que suceda una sola vez en cada uno de los marcadores.

El resultado de este procedimiento se puede observar en el Anexo 4.

## 6. Clasificación de las oscilaciones periódicas

En el análisis anterior se encontró que en cada uno de los marcadores se encontraron patrones de alta frecuencia que se repiten a lo largo de las señales [21]. Se debe identificar cuáles de las muestras corresponden a esta señal periódica.

Para lo anterior se tomaron señales guía donde ya se identificó esta señal periódica utilizando un sensor MFL de la misma herramienta donde se obtuvieron las señales caliper [21].

Se compararon las señales filtradas caliper con la señal de referencia MFL para identificar como se representan la señal periódica en la señal caliper.

Se observó la misma sección en las dos señales (superior MFL de referencia, Inferior Caliper 1) y se observaron las similitudes que se muestran en la Figura 22.

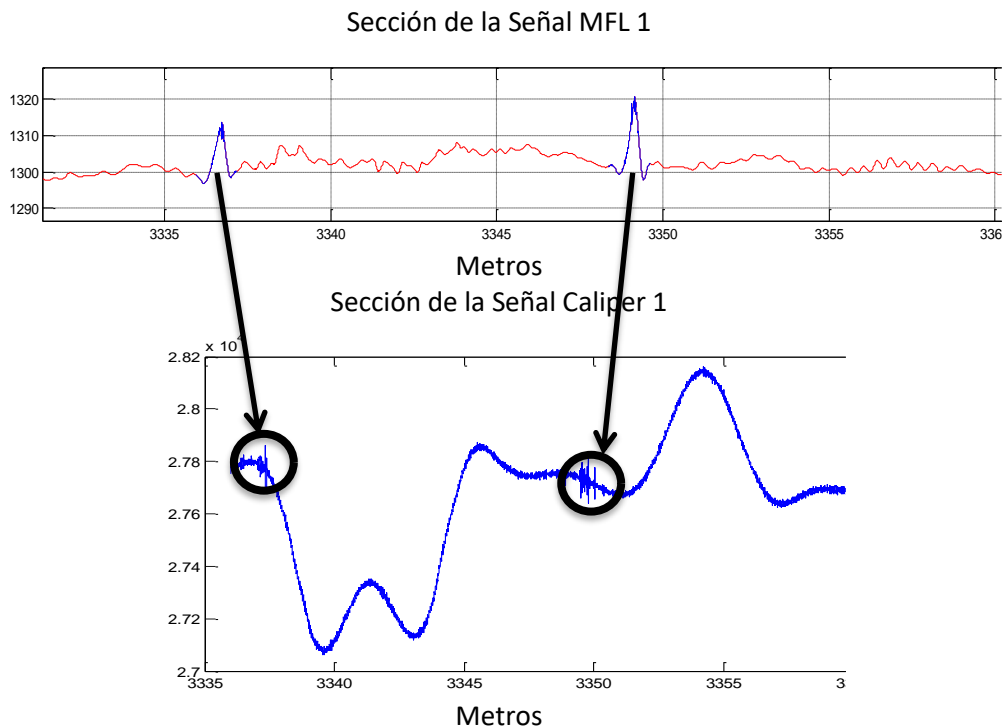


Figura 22. Comparación entre Señal MFL de referencia y Caliper. Fuente: [21], Autor.

Para validar la comparación anterior se selecciona otra parte de la señal y se observan las similitudes entre las oscilaciones periódicas de la señal MFL de referencia (superior) y una de las señales caliper (inferior). En la Figura 23 se muestran las similitudes encontradas.

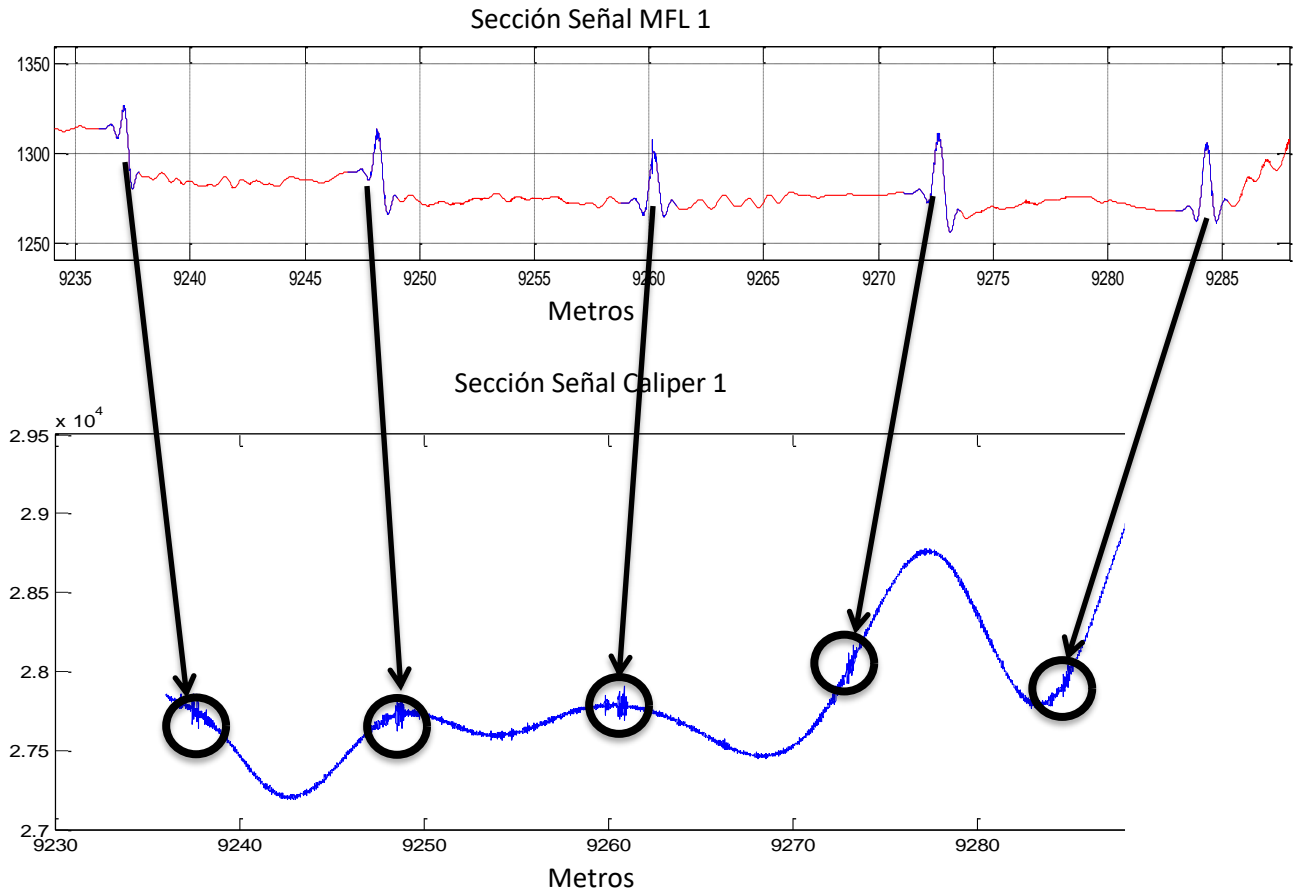


Figura 23. Comparación entre Señal MFL de referencia y Caliper en diferente sección. Fuente [21], Autor.

Se realiza una comparación entre los marcadores Caliper vistos desde la señal MFL de referencia y las señales Caliper.

En la Figura 24 se observa el marcador TA1 en las señales: MFL de referencia (superior) y la señal Caliper 1 (inferior). En esta Figura se observa donde se encuentra la oscilación periódica en dicho marcador.

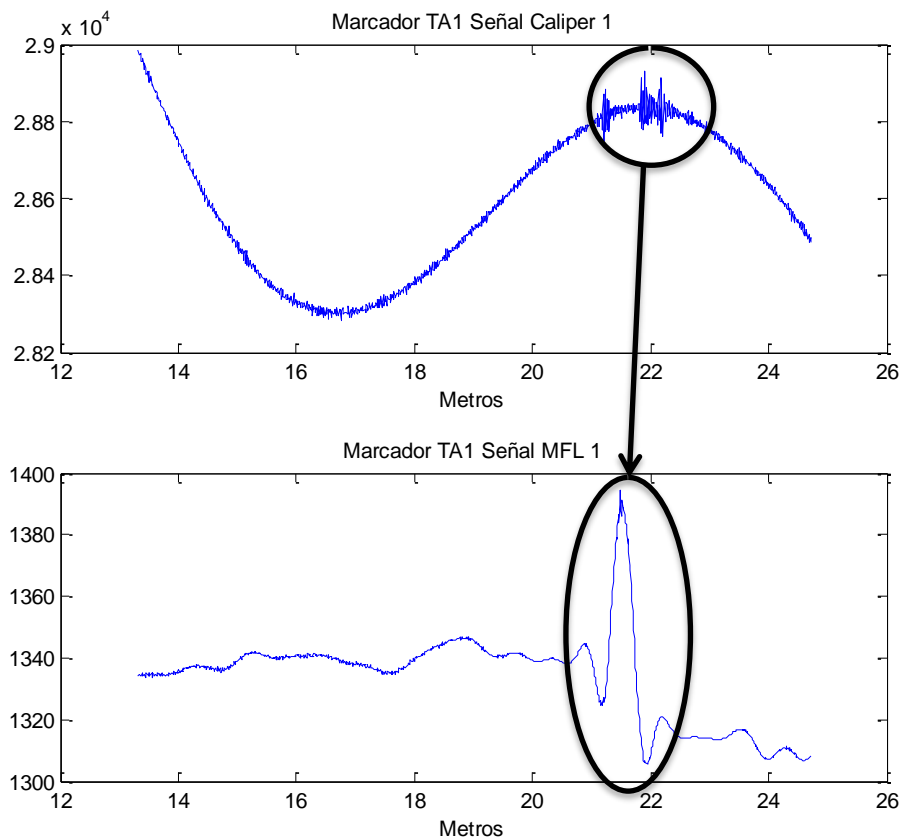


Figura 24. Marcador TA1 en la señal Caliper 1 y en la señal MFL de referencia.

Fuente. Autor.

Esta comparación se realizó con todos los marcadores hallando una soldadura similar en cada uno de ellos.

Luego se realiza este mismo análisis utilizando las señales sin filtrar es decir *Raw Data* para comparar los dos procedimientos y llegar a la conclusión de si la señal con filtro.

En la Figura 25 se observa la comparación entre la señal MFL de referencia (superior), y una sección de la señal Caliper sin filtrar (inferior), con el fin de encontrar las oscilaciones periódicas en la señal.

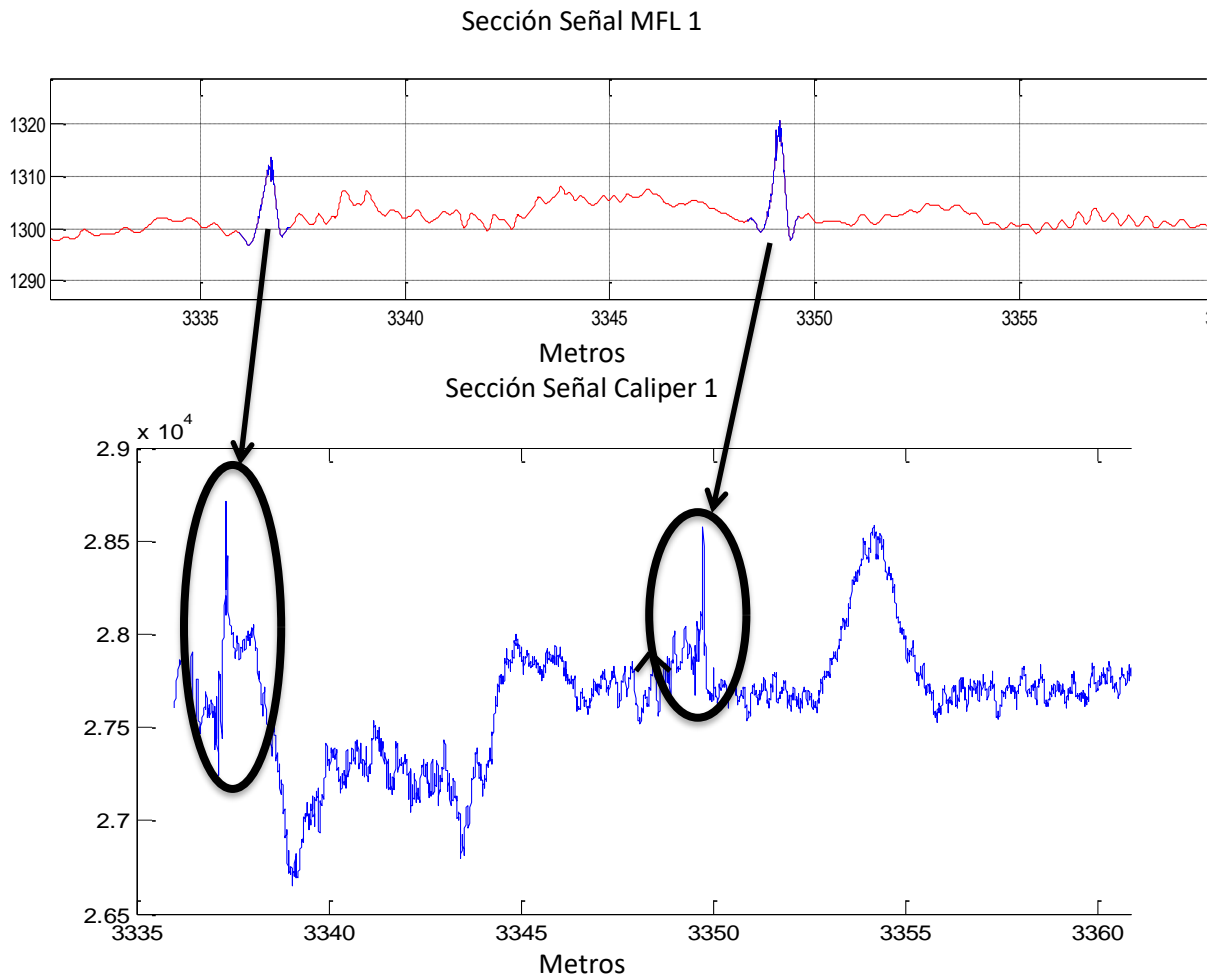


Figura 25. Comparación entre Señal MFL de referencia y la señal Caliper sin filtrar.

Autor: [21], Autor.

En la Figura 25 se observa que además de las oscilaciones periódicas, existen otras anomalías que pueden ser fenómenos que se estén buscando dentro de la señal.

Se realiza una validación con otra zona de la señal para confirmar que estos picos coinciden en toda la señal Caliper sin filtrar.

En la Figura 26 se observa la misma comparación anterior en otra zona de la señal. La señal MFL (superior), la señal Caliper sin filtrar (inferior).

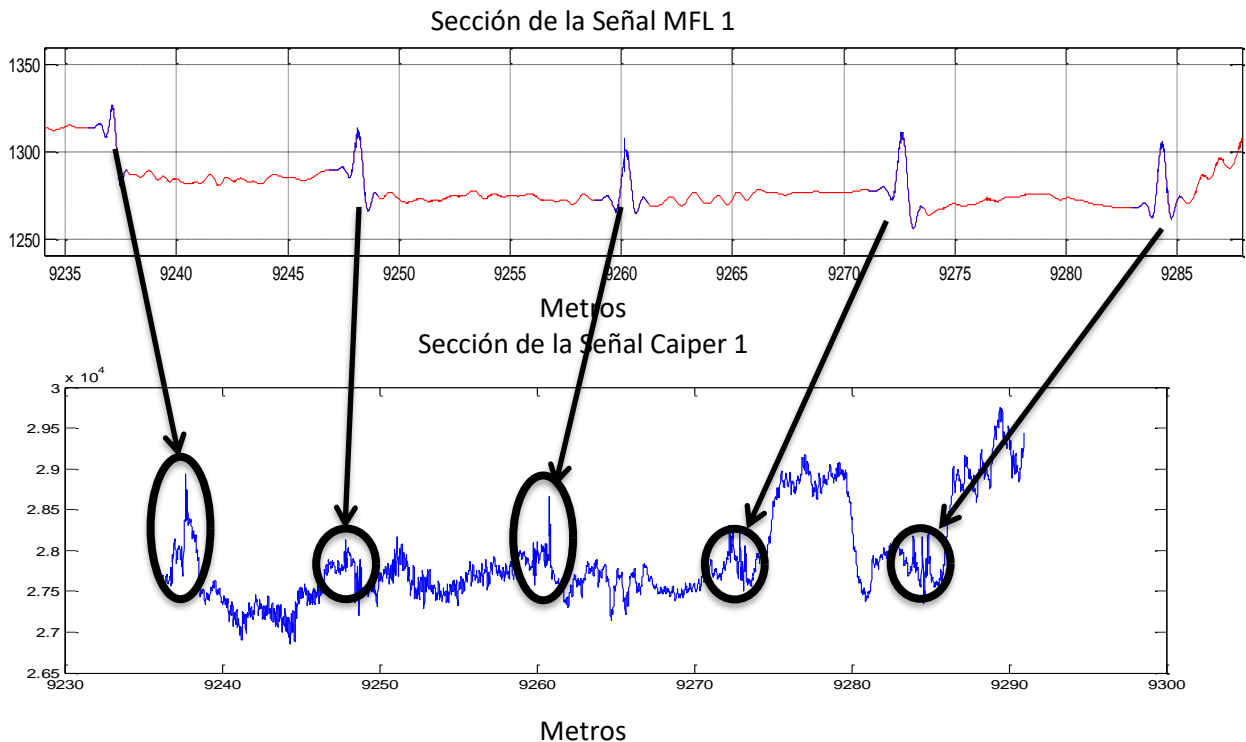


Figura 26. Comparación entre Señal MFL de referencia y Caliper sin filtrar en diferente sección. Fuente [21], Autor.

En otra sección también se encuentran estas oscilaciones y es posible que se solapen algunas de estas con el ruido de la señal al no filtrarse, de igual manera se elimina información que podría representar una abolladura.

Como se realizó anteriormente se compara entre los marcadores caliper, vistos desde la señal MFL de referencia y los mismos marcadores en la señal caliper.

En la Figura 27 se observa el marcador TA1 en las señales: MFL de referencia (Inferior) y la señal Caliper 1 sin filtrar (Superior). En esta Figura se observa donde se encuentra la oscilación en dicho marcador.



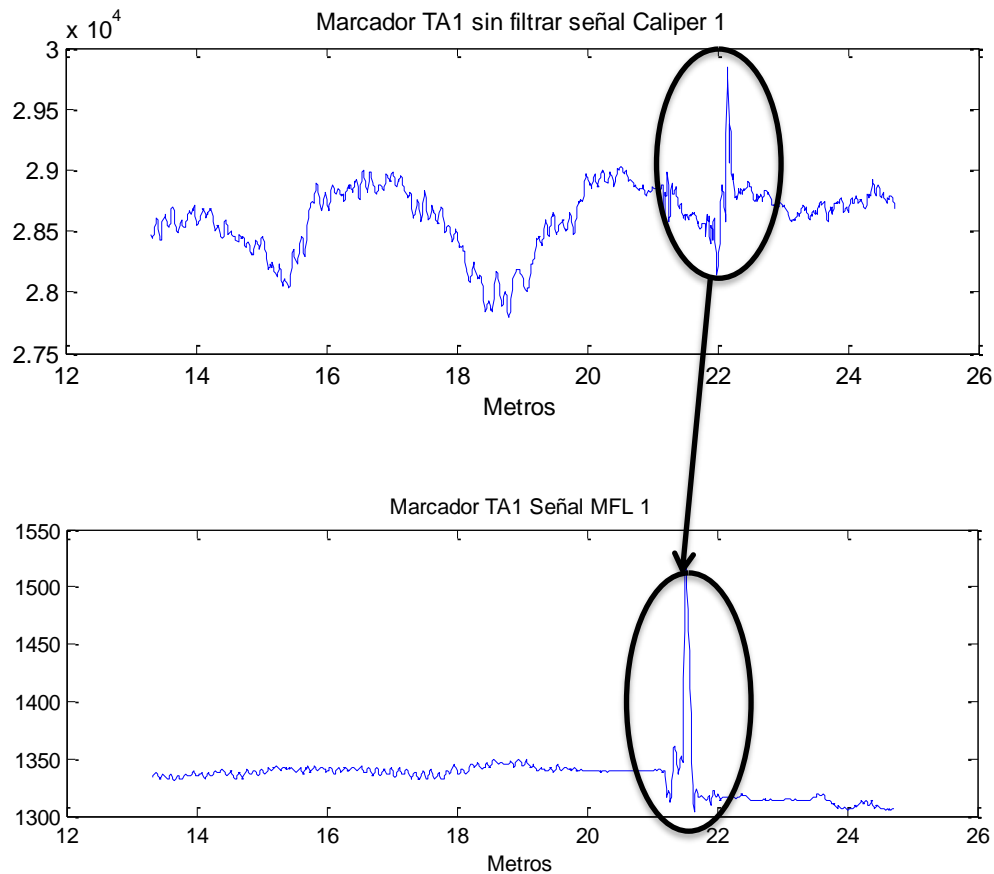


Figura 27. Marcador TA1 en la señal Caliper 1 sin filtrar y en la señal MFL de referencia. Fuente. Autor.

Esta comparación se realizó con todos los marcadores hallando una oscilación similar en cada uno de ellos.

En la Figura 27 se demuestra que los marcadores de la señal caliper sin filtrar también se pueden detectar estas oscilaciones, además de proporcionar mayor cantidad de información.

A continuación se realiza una descomposición *Wavelet* para los marcadores filtrados y sin filtrar con el fin de encontrar una señal reconstruida común que pueda identificar las oscilaciones más fácilmente en toda la señal.

Se utilizó la herramienta *Wavelet packet* de Matlab® para hallar el árbol de descomposición óptimo por medio de la función *Besttree* que se basa en la

entropía de Shannon, donde se observan todas las posibles derivaciones utilizando la transformada *Wavelet* de una señal como se observa en la Figura 28.

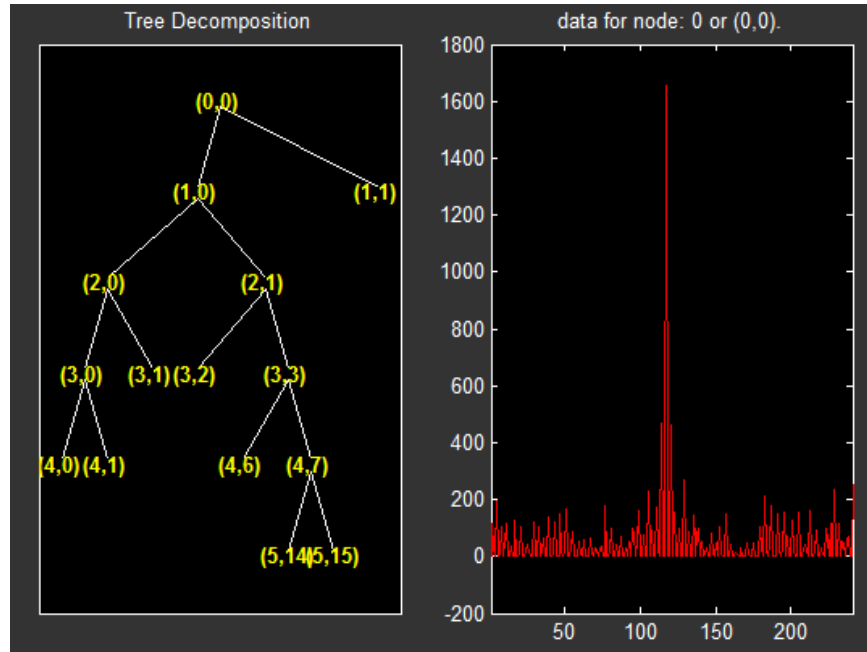


Figura 28. Árbol de descomposición de una señal. Fuente: Autor.

Los nodos que se observan en la Figura 50 son cada una de las posibles derivaciones de la señal mostrada en la misma figura. En el caso de dicha señal posee 15 derivaciones o nodos a una profundidad de 6 niveles.

Para los marcadores filtrados se obtuvo este árbol de descomposición para cada uno, a 9 niveles utilizando una *Wavelet* madre tipo *Haar* y para cada uno de estos se necesitó un nodo diferente para reconstruir la señal de manera que las oscilaciones se distingan del ruido de fondo. Debido a esto la señal filtrada necesitaría diferentes niveles de descomposición para diferentes partes de la señal lo que complica el procesamiento de la misma.

Los nodos que se mencionaron anteriormente con los cuales se debe reconstruir cada marcador para diferenciar las oscilaciones periódicas de la señal sin defecto se observan en la Tabla 4. Por ejemplo para el marcador TA1 de la señal Caliper 1

es necesario reconstruir el marcador con el nodo 13 del árbol de descomposición para diferenciar la oscilación del resto de la señal.

Marcador\Señal Caliper	Caliper 1	Caliper 2
TA1	Nodo 13	Nodo 13
TA2	Nodo 27	Nodo 39
TA3	Nodo 56	Nodo 19
TA4	Nodo 56	Nodo 20
GR1	Nodo 27	Nodo 13
GR2	Nodo 6	Nodo 10
GR3	Nodo 13	Nodo 10

Tabla 4. Nodos donde se observa la oscilación para cada uno de los marcadores de cada señal. Fuente: Autor.

Con el fin de validar el resultado de las señales filtradas se procede a reconstruir la señal utilizando el nodo más común de la Tabla 4, con el fin de obtener una señal que garantice la mayor cantidad de oscilaciones periódicas.

En la Figura 29 y en la Figura 30 se observa secciones de las señales reconstruidas de las dos señales caliper filtradas.

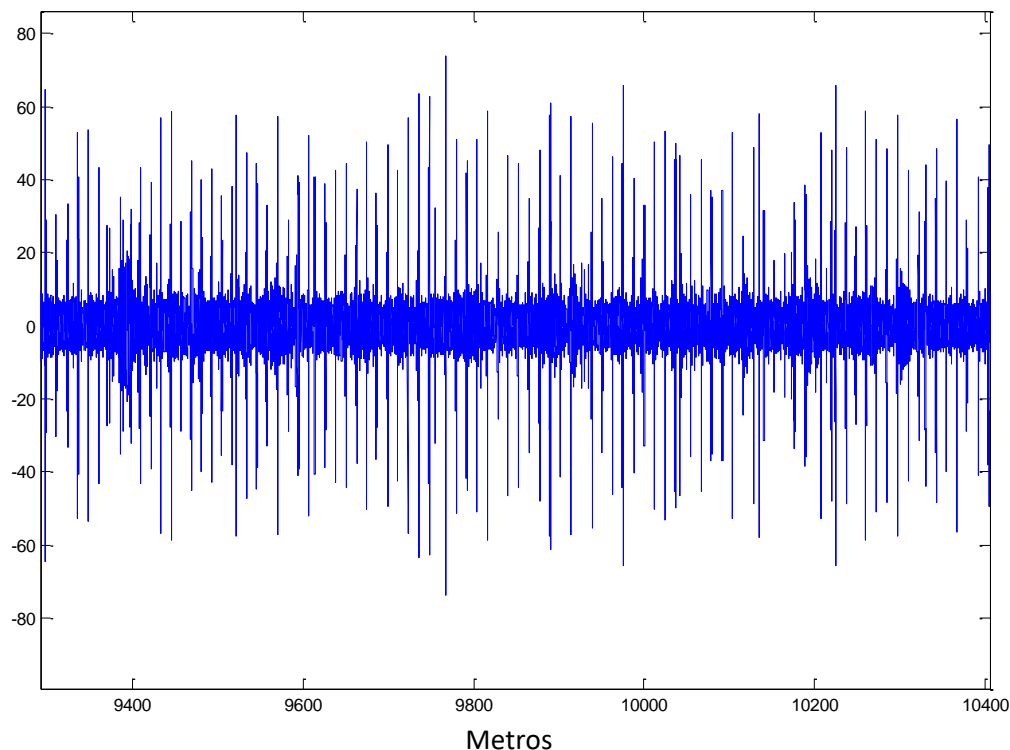


Figura 29. Señal reconstruida Caliper 1 Filtrada. Fuente: Autor.

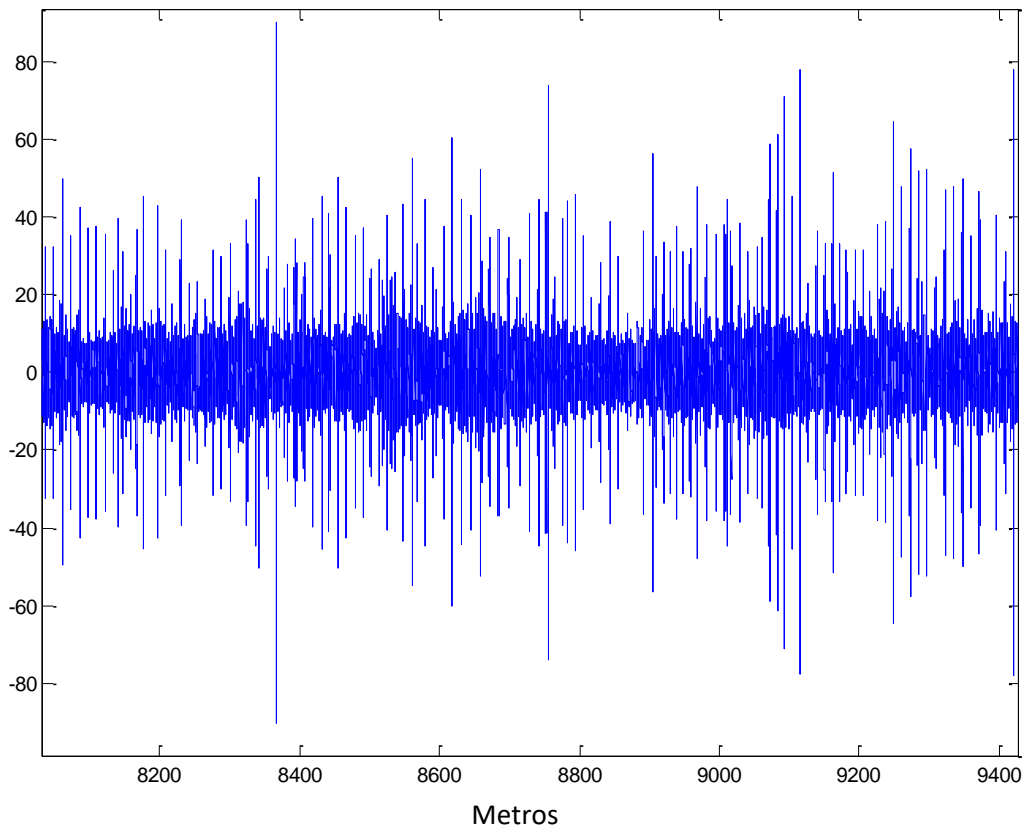


Figura 30. Señal Reconstruida Caliper 2 Filtrada. Fuente: Autor.

Cada pico que se observa en las figuras anteriores representa una oscilación periódica que se encuentra en las señales caliper. Estas oscilaciones se desean identificar y etiquetar.

Utilizando una de las señales reconstruidas se busca un umbral donde identifique la mayor cantidad de oscilaciones periódicas [21] (ver 8. Bibliografía). Para esto se separa el pico que representa la oscilación en cada marcador, se elimina los datos negativos debido a que la señal es simétrica en cero y su media correspondería a cero. Luego se calcula la media y la desviación estándar de cada una de las oscilaciones. En la Figura 31 se observa la oscilación de uno de los marcadores.

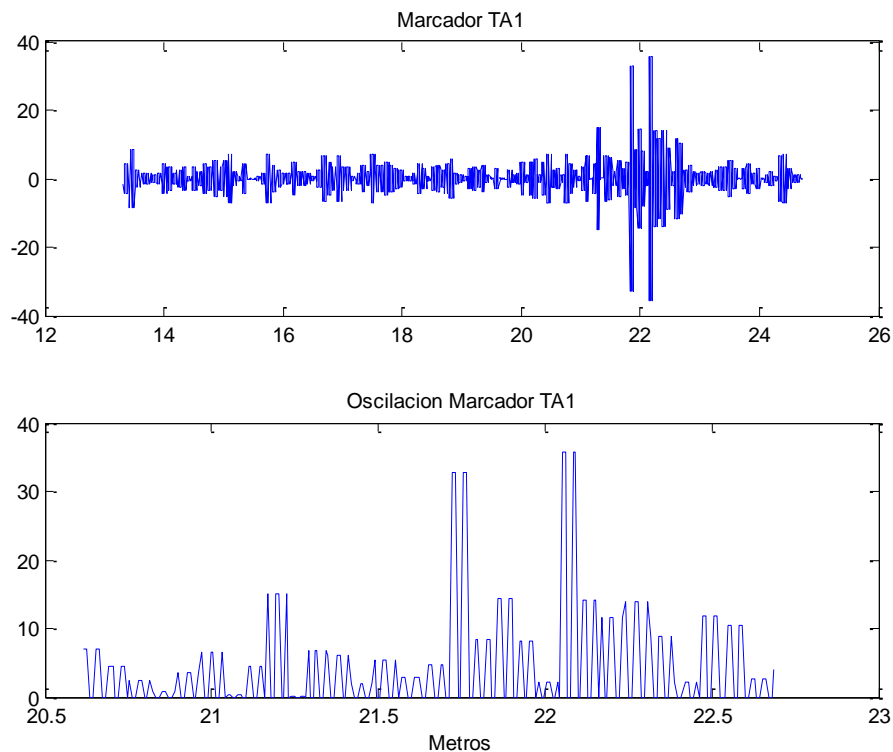


Figura 31. Marcador TA1 y oscilación señal Caliper 1 reconstruida y filtrada.

Fuente: Autor.

Los estadísticos obtenidos se observan en la Tabla 5.

Estadístico\Marcador	TA1	TA2	TA3	TA4	GR1	GR2	GR3
Media	4.1625	3.1093	3.7050	4.1017	4.1392	3.8136	4.1025
Desviación	7.1338	6.3927	7.7137	8.2024	7.1400	6.4824	8.2021
Media + Desviación	11.2965	9.5021	11.4188	12.3041	11.2791	10.296	12.304

Tabla 5. Estadísticos de las oscilaciones de cada marcador Caliper 1. Fuente:

Autor.

Al obtener estos estadísticos, se utiliza la suma entre ellos como umbral ya que esto permite obviar la magnitud del ruido que se presenta en la señal, además de asegurar que la mayoría de las oscilaciones periódicas de la señal sean detectadas por dicho umbral. Luego se procede a validar cuál de ellos detecta la mayor cantidad de oscilaciones en la señal.

Para validar el procedimiento se realiza un cómputo de los datos [21] mediante el cual se le otorga valores de 1 a cualquier muestra de la señal que este superior al umbral y 0 a los datos que sean inferiores. Luego se halla la distancia donde se encuentran, los flancos de subida en la señal de unos y ceros y la diferencia entre un flanco y otro.

Las oscilaciones se encuentran separadas entre 10 a 13 metros o entre 23 a 25 metros según la información inicial otorgada por la Corporación para la investigación de la corrosión. Se buscan los flancos que tengan estas distancias entre si y se intenta encontrar el mayor cantidad de oscilaciones.

Se utilizan los umbrales obtenidos por la suma de los estadísticos de cada uno de las oscilaciones de los marcadores y se escoge el que detecte la mayor cantidad de ellas. En la Tabla 6 se muestra las encontradas por cada uno de los umbrales para la señal Caliper 1 filtrada.

	TA1	TA2	TA3	TA4	GR1	GR2	GR3
Umbral	11.2965	9.5021	11.4188	12.3041	11.2791	10.296	12.304
Oscilaciones encontradas	1072	623	1106	1278	1068	841	1279

Tabla 6. Oscilaciones encontradas por el umbral obtenido de cada marcador en la señal reconstruida Caliper 1 filtrada. Fuente: Autor.

Se realizó el mismo procedimiento utilizando con la señal reconstruida Caliper 2 filtrada utilizando los mismos umbrales y se obtuvieron los siguientes resultados (Tabla 7).

	TA1	TA2	TA3	TA4	GR1	GR2	GR3
Umbral	11.2965	9.5021	11.4188	12.3041	11.2791	10.296	12.304
Oscilaciones encontradas	526	126	553	789	525	276	789

Tabla 7. Oscilaciones encontradas por el umbral obtenido de cada marcador en la señal reconstruida Caliper 2. Fuente: Autor.

En el procedimiento anterior se observa que utilizando la señal filtrada no se halla una gran cantidad de oscilaciones en la señal además, no funciona de la misma manera para las dos señales caliper esto se debe a que donde se encuentra la información de las oscilaciones periódicas son picos de alta frecuencia y al filtrar la señal estos picos atenúan y no se identifica la mayoría de las oscilaciones.

Se realiza el mismo procedimiento anterior utilizando las señales Caliper sin filtrar con el fin de encontrar una mayor cantidad de oscilaciones periódicas en la señal.

Como se realizó anteriormente se observa el árbol de descomposición para los marcadores sin filtrar de las dos señales con el fin de encontrar si era posible reconstruir la señal utilizando un nodo común para todos los marcadores.

Se realiza la prueba utilizando el nodo que más se repite en los marcadores filtrados (nodo 13) y observar si la señal caliper reconstruida con dicho marcador logra separar las oscilaciones periódicas del resto de la señal.

Se encontró por medio de inspección visual que la reconstrucción de la señal utilizando el nodo número 13 para las dos señales y todos los marcadores, esto garantiza que se pueda diferenciar la oscilación del resto del marcador debido a que los picos de la señal reconstruida que representan la oscilación poseen una mayor diferencia de amplitud entre el ruido y el valor más alto de este mismo comparado con la reconstrucción de la señal utilizando la señal filtrada.

En la Figura 32 y la Figura 33 se observan secciones de las dos señales reconstruidas utilizando las señales Caliper sin filtrar.

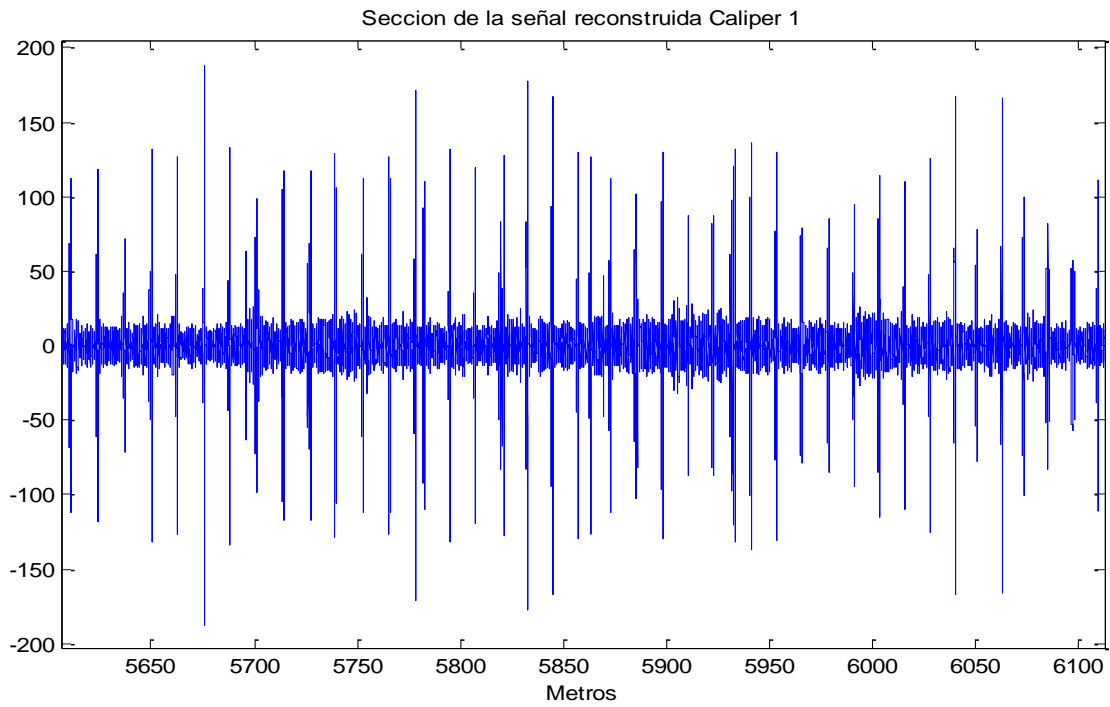


Figura 32. Señal reconstruida Caliper 1 sin filtrar. Fuente: Autor.

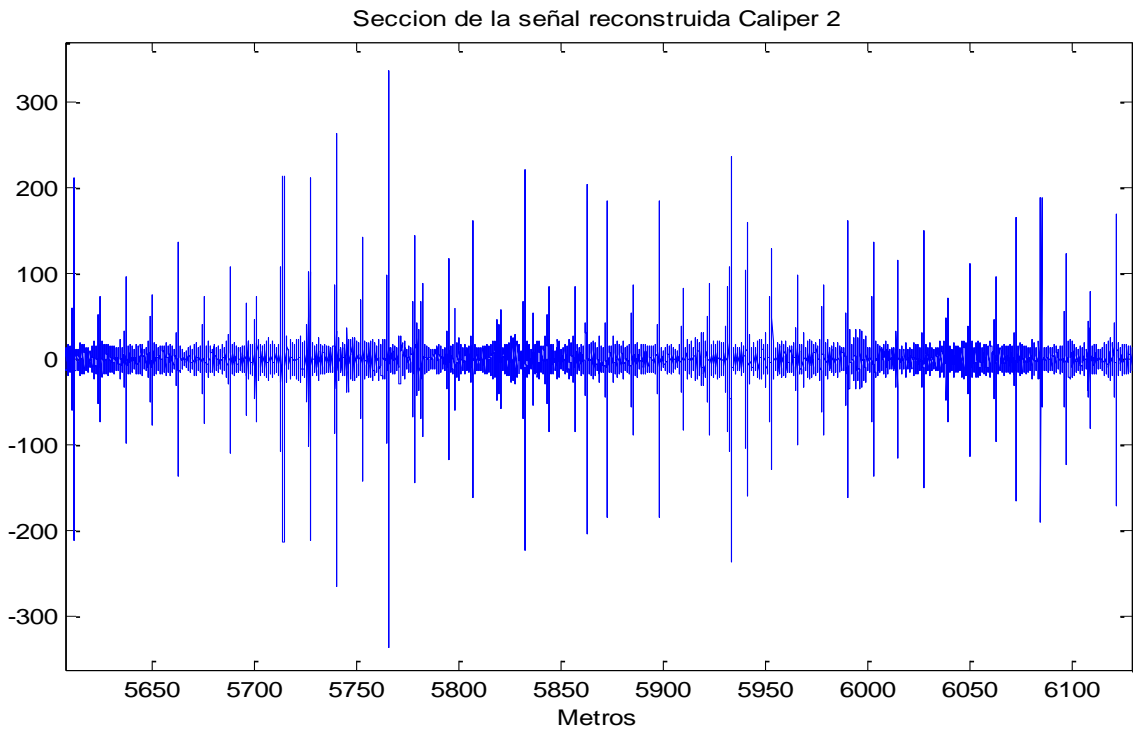


Figura 33. Señal reconstruida Caliper 2 sin filtrar. Fuente: Autor.



En las figuras anteriores representan las señales de alta frecuencia que forman parte de las señales caliper. Cada pico, como se observó anteriormente en las señales reconstruidas de las señales Caliper filtradas, representa una oscilación periódica que se encuentra a lo largo de la señal. Estos picos se desean identificar y etiquetar.

Utilizando una de las señales reconstruidas se busca un umbral donde identifique la mayor cantidad de oscilaciones periódicas [21]. Para esto se separa el pico que representa la oscilación en cada marcador, se elimina los datos negativos debido a que la señal es simétrica en cero y su media correspondería a cero. Luego se calcula la media y la desviación estándar de cada una de las oscilaciones. En la Figura 34 se observa la oscilación de uno de los marcadores.

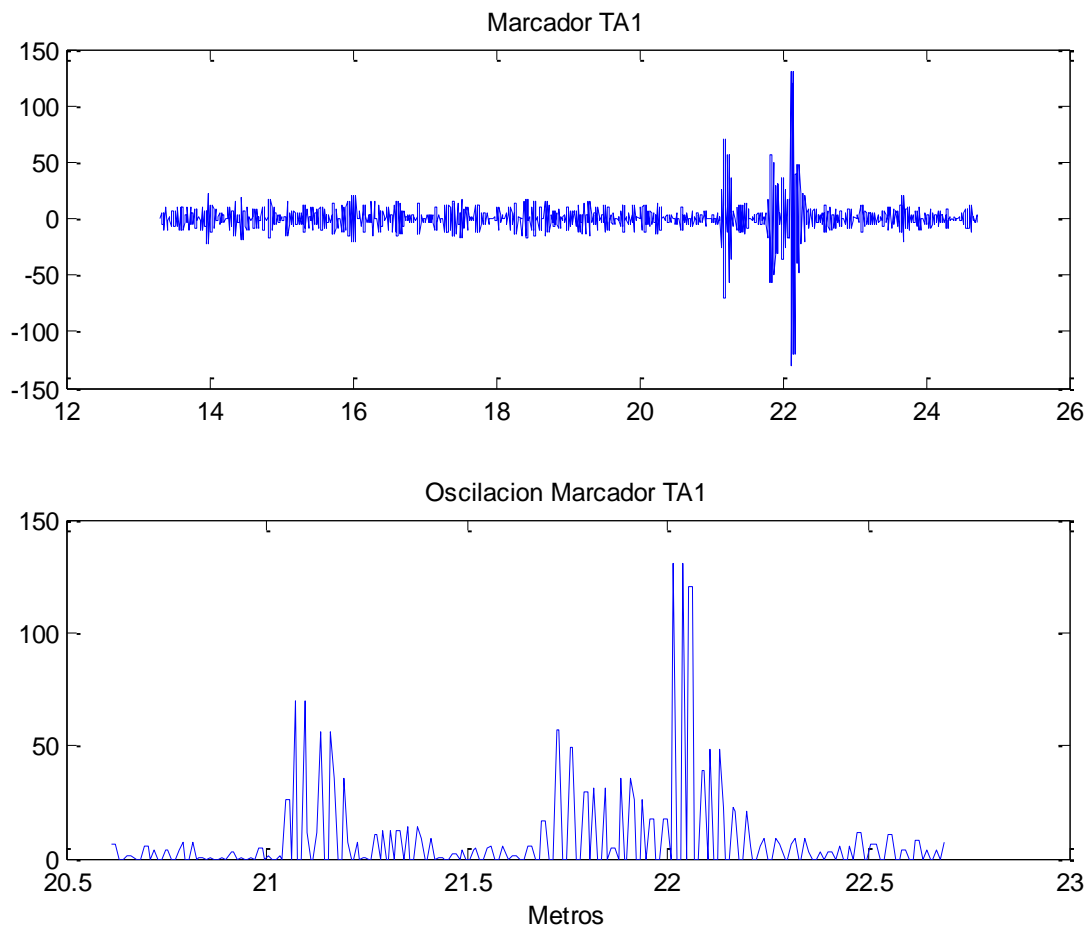


Figura 34. Marcador TA1 y oscilación señal Caliper 1. Fuente: Autor.

Los estadísticos obtenidos se observan en la Tabla 8.

Estadístico\Marcador	TA1	TA2	TA3	TA4	GR1	GR2	GR3
Media	9.0010	9.7541	6.0674	9.2313	8.9678	7.5601	9.1919
Desviación	20.1775	24.3046	12.7999	19.7756	20.1872	17.7863	19.7868
Media + Desviación	29.0785	34.0587	18.8673	29.0069	29.155	25.3464	28.987

Tabla 8. Estadísticos de las oscilaciones de cada marcador Caliper 1. Fuente: Autor.

Como se muestra anteriormente luego de obtener los estadísticos se utiliza la suma entre ellos como umbral ya que esto permite obviar la magnitud del ruido que se presenta en la señal, además de asegurar que la mayoría de las oscilaciones periódicas de la señal sean detectadas por dicho umbral. Luego se procede a validar cuál de ellos detecta la mayor cantidad de oscilaciones en la señal.

Para validar el procedimiento se realiza un cómputo de los datos [21] mediante el cual se le otorga valores de 1 a cualquier muestra de la señal que este superior al umbral y 0 a los datos que sean inferiores. Luego se halla la distancia donde se encuentran, los flancos de subida en la señal de unos y ceros y la diferencia entre un flanco y otro.

Las oscilaciones se encuentran separadas entre 10 a 13 metros o entre 23 a 25 metros según la información inicial otorgada por la Corporación para la investigación de la corrosión. Se buscan los flancos que tengan estas distancias entre si y se intenta encontrar el mayor cantidad de oscilaciones.

Se utilizan los umbrales obtenidos por la suma de los estadísticos de cada uno de las oscilaciones de los marcadores y se escoge el que detecte la mayor cantidad de ellas. En la Tabla 9 se muestra las encontradas por cada uno de los umbrales para la señal Caliper 1.

	TA1	TA2	TA3	TA4	GR1	GR2	GR3
Umbral	29.0785	34.0587	18.8673	29.0069	29.155	25.3464	28.987
Oscilaciones encontradas	1521	1588	835	1521	1521	1411	1513

Tabla 9. Oscilaciones encontradas por el umbral obtenido de cada marcador en la señal reconstruida Caliper 1. Fuente: Autor.

Se realizó el mismo procedimiento con la señal reconstruida Caliper 2 sin filtrar utilizando los mismos umbrales y se obtuvieron los siguientes resultados (Tabla 10).

	TA1	TA2	TA3	TA4	GR1	GR2	GR3
Umbral	29.0785	34.0587	18.8673	29.0069	29.155	25.3464	28.987
Oscilaciones encontradas	1432	1560	377	1432	1432	1254	1425

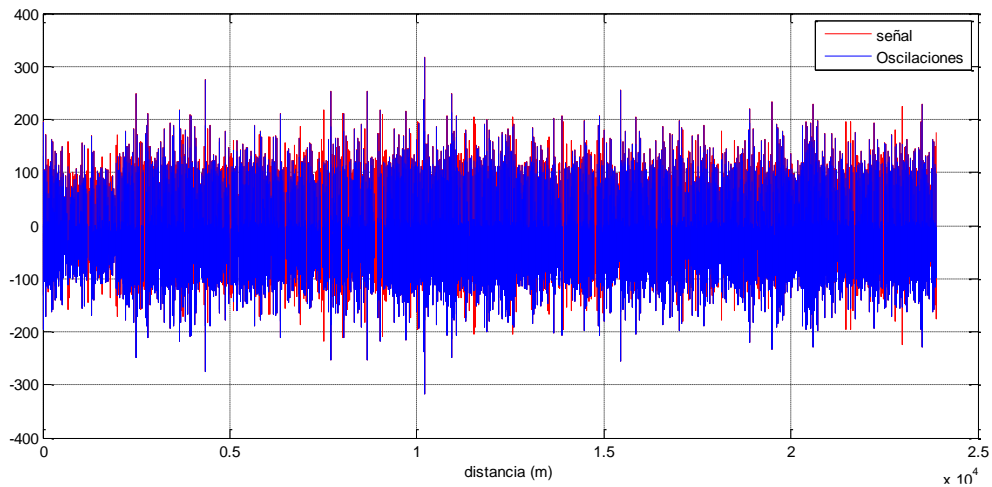
Tabla 10. Oscilaciones encontradas por el umbral obtenido de cada marcador en la señal reconstruida Caliper 2. Fuente: Autor.

El umbral obtenido por los estadísticos de la oscilación del marcador TA2 obtuvo los mejores resultados y los más similares entre las dos señales reconstruidas caliper.

En el procedimiento utilizando las señales reconstruidas obtenidas de las señales caliper sin filtrar, se llegó a encontrar un mayor número de oscilaciones en la señal, además el procedimiento funciona de la misma manera para las dos señales caliper obteniendo un buen resultado para las dos.

Utilizando las señales sin filtrar se obtiene un mejor resultado comprado con las señales filtradas.

En la Figura 35 se observa la señal reconstruida caliper 1 sin filtrar con las oscilaciones que se identificaron. El color azul representa las oscilaciones encontradas y el color rojo representa las partes de la señal que no se clasificaron como oscilaciones.



Señal 35. Señal reconstruida Caliper 1 y las oscilaciones identificadas. Fuente: Autor.

En la Figura 36 se observa una sección de la señal reconstruida caliper 1 sin filtrar donde se identifican a detalle algunas de las oscilaciones. El color azul representa las oscilaciones encontradas, el color rojo representa la señal que se detectó sin defecto y los picos que se señalan en la figura son las oscilaciones que no se detectaron en el proceso o falsos negativos.

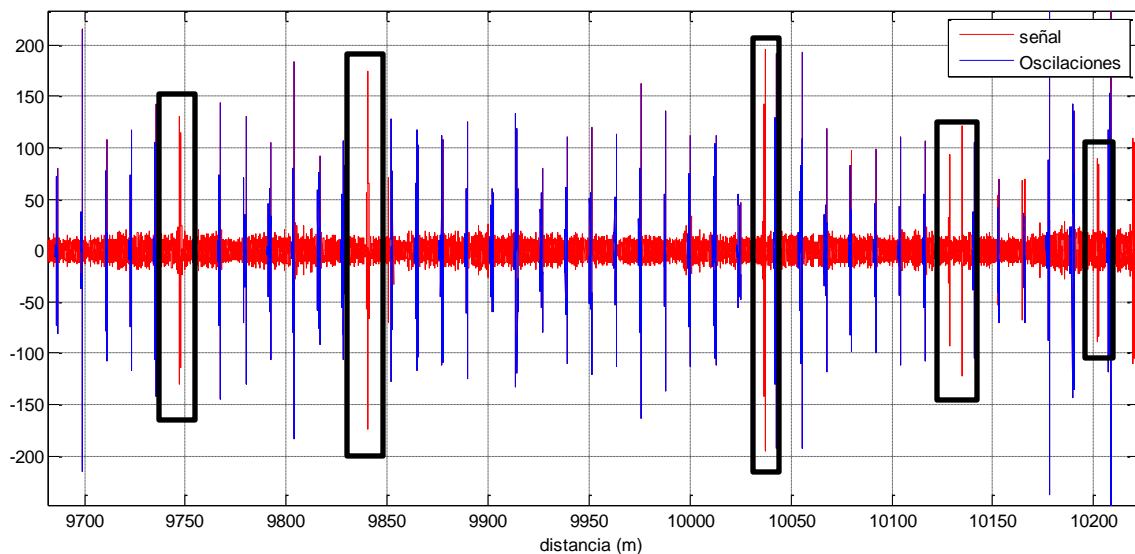


Figura 36. Sección de la señal reconstruida Caliper 1 y las oscilaciones identificadas. Fuente: Autor.

En la Figura 37 se observa la señal Caliper 1 sin filtrar con las oscilaciones identificadas. El color azul representa las oscilaciones encontradas, el color rojo representa la señal que no se consideran como defecto y los picos señalados son las oscilaciones que no se detectaron en el proceso o falsos positivos.

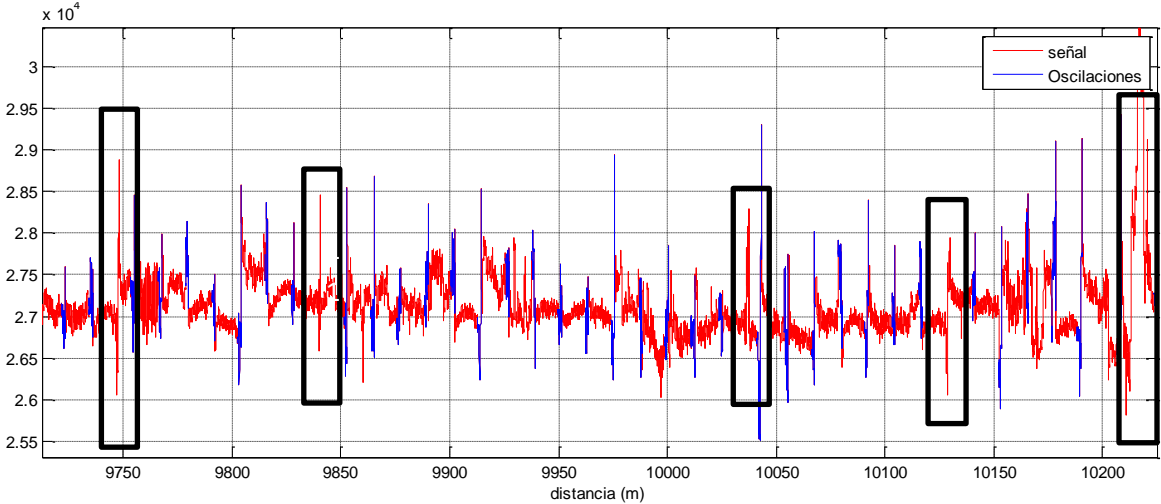
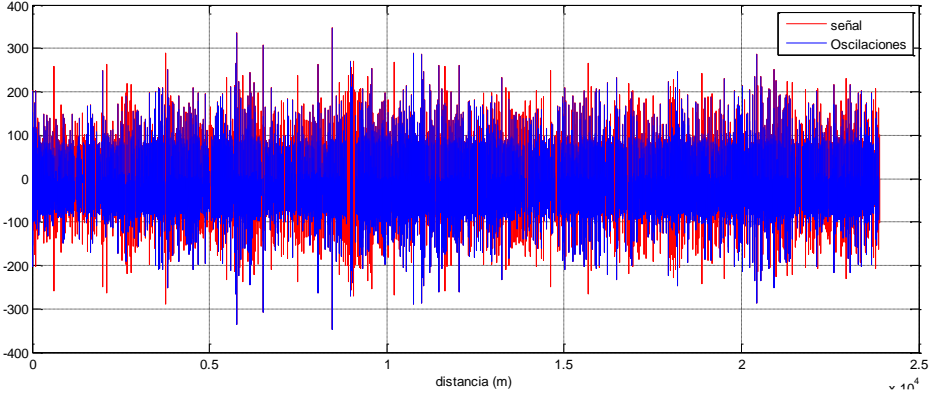


Figura 37, Sección de la señal Caliper 1 sin filtrar y las oscilaciones identificadas.

Fuente: Autor.

En la Figura 38 se observa la señal reconstruida caliper 2 sin filtrar con las oscilaciones que se identificaron. El color azul representa las oscilaciones encontradas y el color rojo representa las partes de la señal que no se clasificaron como oscilaciones.



Señal 38. Señal reconstruida Caliper 2 y las oscilaciones identificadas. Fuente:

Autor.

En la Figura 39 se observa una sección de la señal reconstruida caliper 2 sin filtrar donde se identifican a detalle algunas de las oscilaciones. El color azul representa las oscilaciones encontradas, el color rojo representa la señal que se detectó sin defecto y los picos que se señalan en la figura son las oscilaciones que no se detectaron en el proceso o falsos negativos.

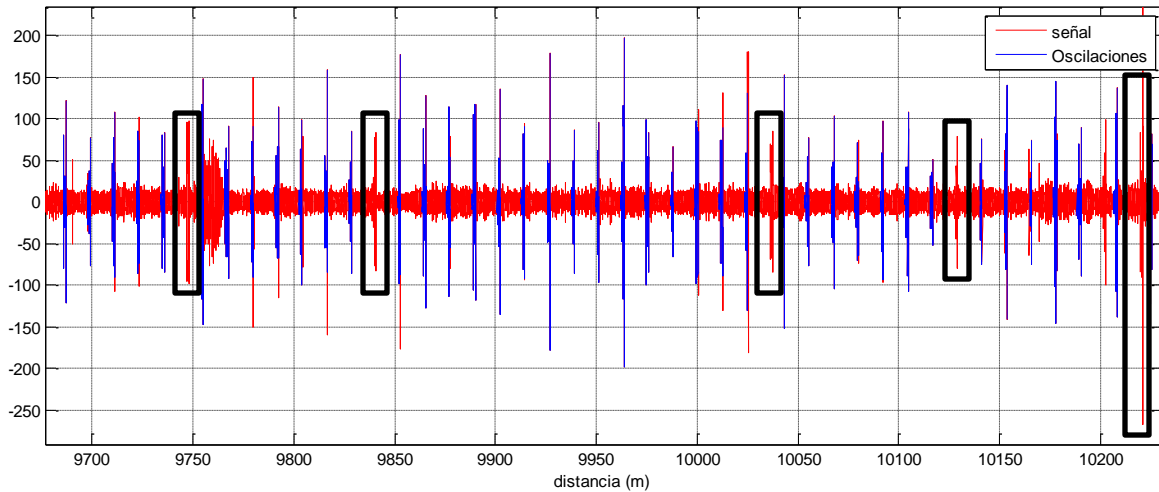


Figura 39. Sección de la señal reconstruida Caliper 2 y las oscilaciones identificadas. Fuente: Autor.

En la Figura 40 se observa la señal Caliper 2 sin filtrar con las oscilaciones identificadas. El color azul representa las oscilaciones encontradas, el color rojo representa la señal que se consideró sin defecto y los picos señalados son las oscilaciones que no se detectaron en el proceso o falsos negativos.

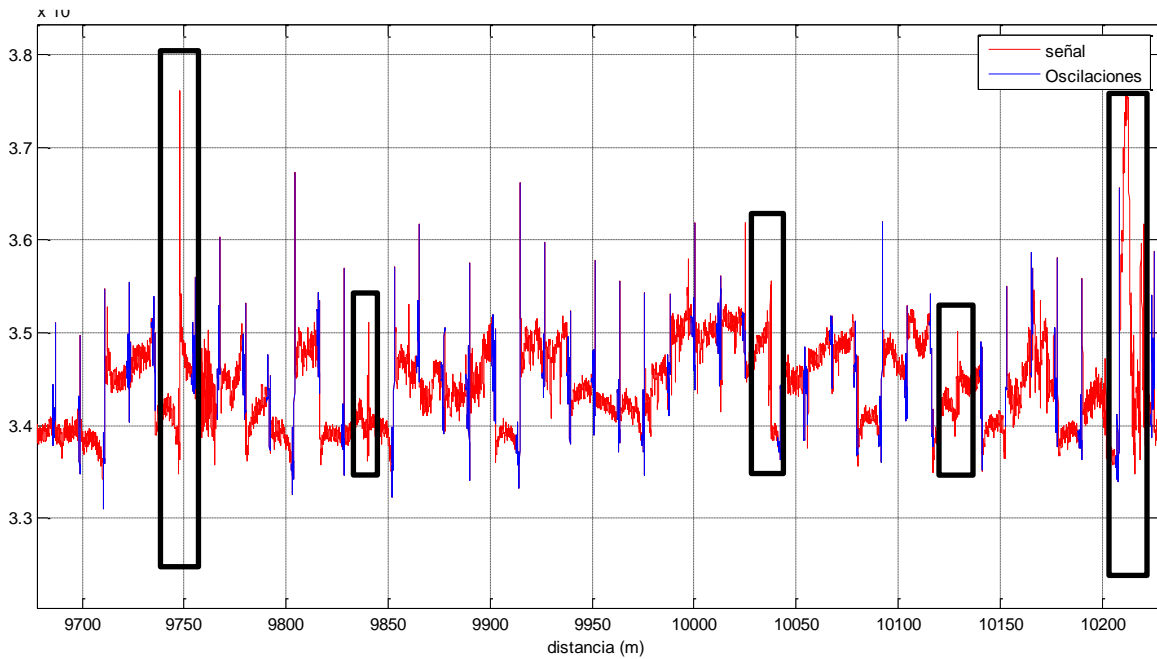


Figura 40. Sección de la señal Caliper 2 sin filtrar y las oscilaciones identificadas.

Fuente: Autor.

Como se observa en las Figuras anteriores algunas oscilaciones no fueron detectadas, esto se debe a que cada una de ellas no sigue con el mismo comportamiento de la mayoría de oscilaciones en la señal, bajando el número de picos clasificados utilizando un proceso de umbralización para su clasificación.

## 6.1 Red Neuronal

En el procedimiento anterior se identificaron una gran cantidad de oscilaciones pero no su totalidad. A continuación se propone un método de clasificación utilizando redes neuronales con el fin de encontrar una mayor cantidad de oscilaciones.

La red neuronal propuesta posee dos entradas que corresponden a las muestras o patrones de las dos señales reconstruidas utilizando el nodo 13 del árbol de descomposición obtenido por la *toolbox* de Matlab® *Wavelet Packet*.

Utilizando las señales reconstruidas y el vector de etiquetas que se encuentra utilizando el procedimiento clasificación por umbralización, se escogen tres secciones de la señal donde se hayan clasificado correctamente las oscilaciones con el fin de crear un grupo de datos etiquetados correctamente para realizar el procedimiento de entrenamiento, prueba y validación de la red neuronal. De esta manera la red neuronal aprenderá el comportamiento de las oscilaciones y así obteniendo un mejor resultado en su clasificación.

En las Figuras 41, Figura 42 y Figura 43 se observan las secciones de una de las señales que se escogieron para crear el grupo de datos para entrenar y validar la red neuronal.

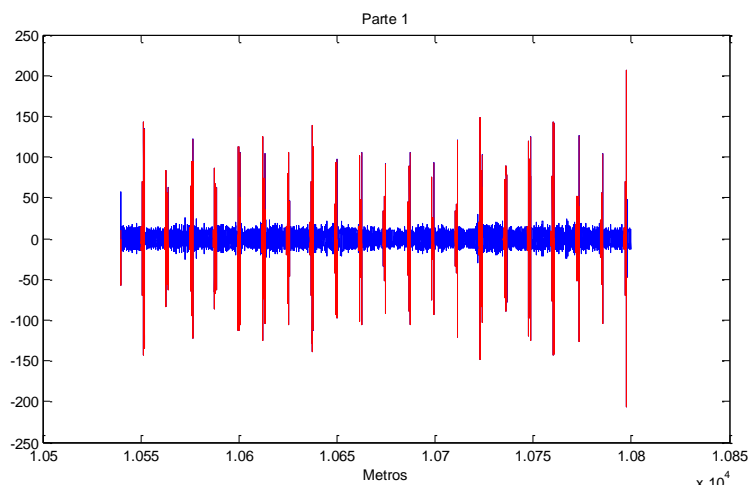


Figura 41. Parte 1 del conjunto de datos para clasificación. Fuente: Autor.



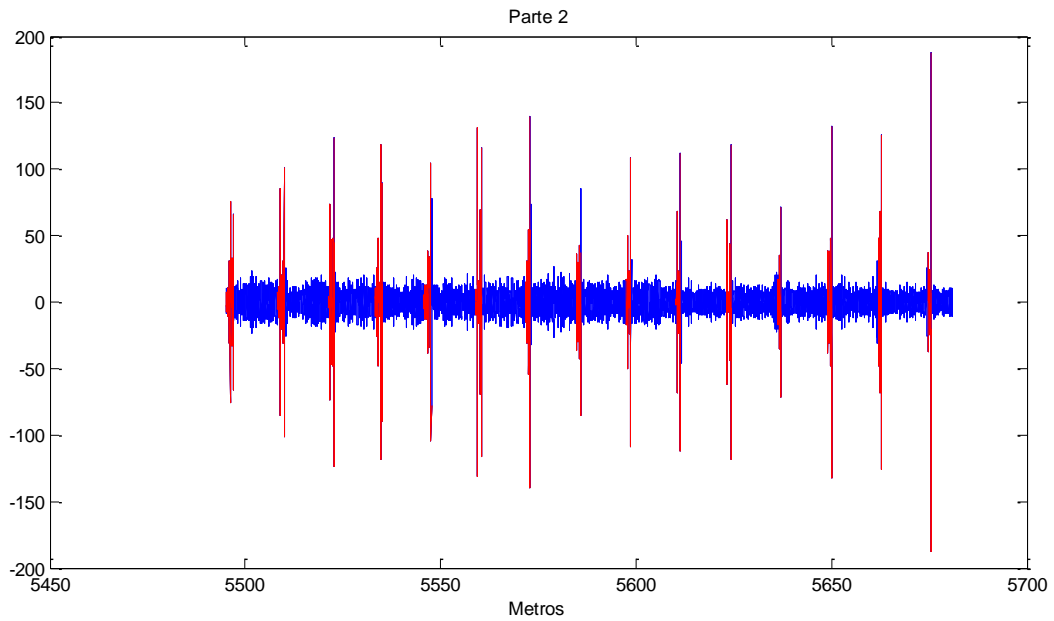


Figura 42. Parte 2 del conjunto de datos para clasificación. Fuente: Autor.

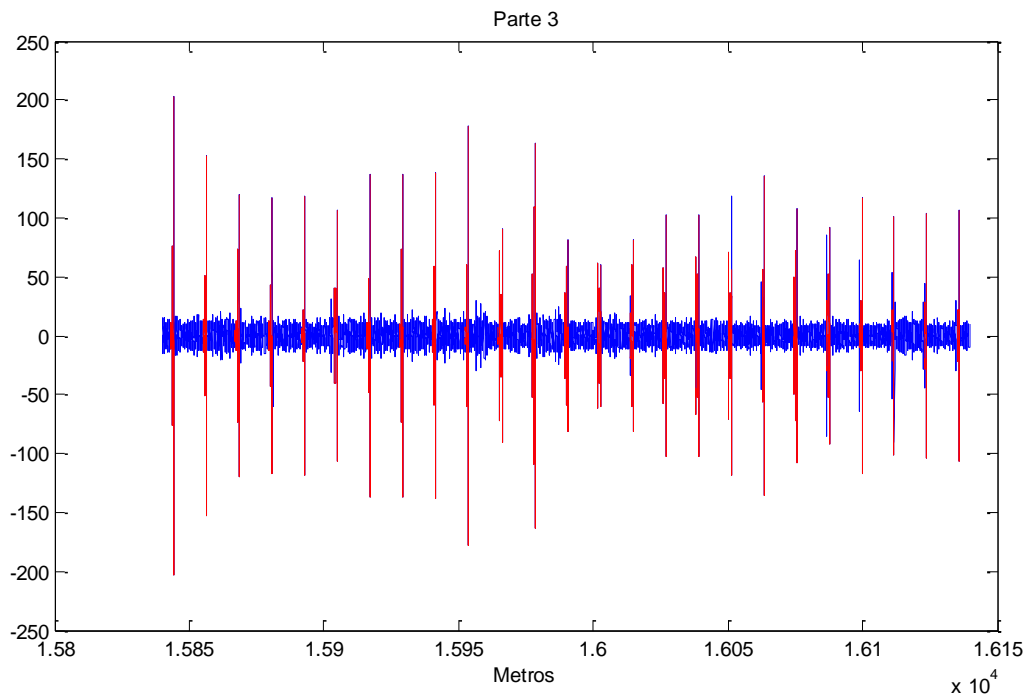


Figura 43. Parte 2 del conjunto de datos para clasificación. Fuente: Autor.

Los datos anteriores se concatenaron en un solo conjunto de datos y se tomaron las dos señales reconstruidas como características para ingresar al clasificador neuronal.

Para el diseño del clasificador se propuso una red con 2 capas ocultas, cada una con 5 y 10 neuronas respectivamente y una función tangente hiperbólica para cada una, en cuanto a la capa de salida se configuro una función lineal.

Para realizar el entrenamiento de la red se propuso los siguientes paramentos:

1. Iteraciones: 1000
2. Meta (mínimo MSE):  $1e-5$
3. Mínimo gradiente:  $1e-7$
4. Tasa de aprendizaje: 0.05
5. Decremento de la tasa de aprendizaje: 0.1
6. Incremento de la tasa de aprendizaje: 10
7. Máxima tasa de aprendizaje:  $1e10$
8. Parada por validación: 6

Se utilizó el algoritmo de entrenamiento Trainlm (*Levenberg-Marquart*) que utiliza un criterio de parada por validación, donde se detiene el entrenamiento si el rendimiento no mejora o se mantiene igual, un máximo número de iteraciones seguidas.

Se utilizó el método de Validación Cruzada con K iteraciones para asegurar que el modelo posea un nivel de generalización aceptable y así garantizar el porcentaje de clasificación.

En se realizó una aleatorización de los datos de las cuatro señales tomadas a diferentes velocidades para luego dividirlos en 5 grupos diferentes. Como se puede observar en la Figura 44,



Figura 44. Grupos de datos. Fuente: Autor.

Luego se determinaron los grupos de datos que se utilizaran para el entrenamiento, validación y prueba de la red neuronal. En este caso, las primeras tres cajas se utilizan para el entrenamiento, la caja número cuatro se utiliza para validación y por último la caja número cinco, se utiliza para probar la red y obtener el porcentaje de clasificación.

Cada vez que realiza un entrenamiento se debe reorganizar los datos de manera que todas las cajas realicen las tres funciones. Como se observa en la Figura 45.

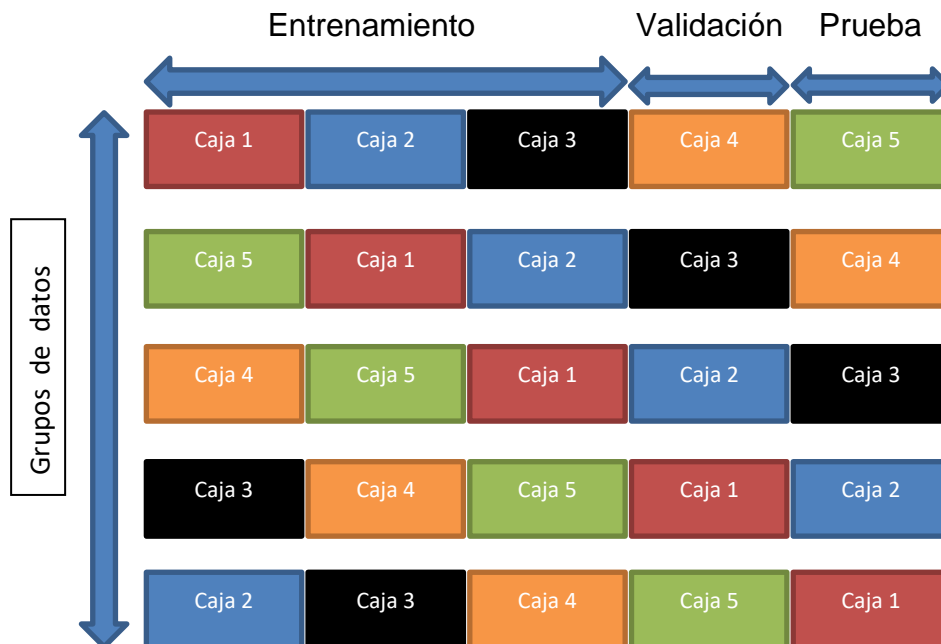


Figura 45. Procedimiento de validación cruzada por K iteraciones. Fuente: Autor

Luego de realizar este proceso, se debe volver a aleatorizar los datos para repetir de nuevo todo el procedimiento.

En este caso repitió este procedimiento 15 veces obteniendo 75 porcentajes de clasificación utilizando diferentes datos de prueba.

### 6.1.1 Resultados

Los porcentajes de clasificación que se obtuvieron de este análisis se pueden observar en la Tabla 11.

Iteraciones\Grupo de datos	1	2	3	4	5
1	86.2688	87.6730	89.8894	89.5327	88.0955
2	86.5376	87.6434	90.0195	89.4517	87.9428
3	86.7123	87.6730	90.0340	89.4810	87.9714
4	86.7623	87.7359	89.9315	89.5034	87.9879
5	86.3479	87.6558	90.1917	89.3786	88.0168
6	86.9391	87.7819	90.1629	89.7429	87.9096
7	86.6094	87.8578	90.2356	89.4888	87.9263
8	86.6486	87.7248	89.9042	89.5327	88.0498
9	86.6980	87.8062	89.8031	89.5181	87.7151
10	86.6633	87.7422	90.0049	89.6137	87.8648
11	86.5808	87.8111	90.2642	89.2330	87.9263
12	86.5852	87.9006	89.8894	89.6425	87.9263
13	86.6048	87.8749	90.0904	89.3925	87.8814
14	86.5808	87.7889	90.0620	89.5906	87.9879
15	86.6835	87.8331	89.8329	89.3279	87.9214

Tabla 11. Porcentaje de clasificación balanceado para cada grupo de datos en cada iteración.

Se obtuvo un porcentaje de clasificación calculado a partir de la ecuación de precisión equilibrada que se muestra a continuación.

$$\text{Precision equilibrada} = \frac{1}{2} * \frac{Cp}{Cp + Fp} + \frac{1}{2} * \frac{Cn}{Cn + Fn}$$

$Cp$  = Correctos Positivos  
 $Fp$  = Flacos Positivos  
 $Cn$  = Correctos Negativos  
 $Fn$  = Falsos Negativos

#### Ec.4. Precisión equilibrada (*Balanced Accuracy*)

Y de esta manera se obtuvo un porcentaje de clasificación total de:

$$\% \text{Clasificación total (Balanced Accuracy)} = 88.3680 \pm 0.1098\%$$

Al validar el clasificador neuronal se probó utilizando una de las señales reconstruidas y se obtuvo el resultado mostrado en las Figuras 46, Figura 47, Figura 48 y Figura 49.

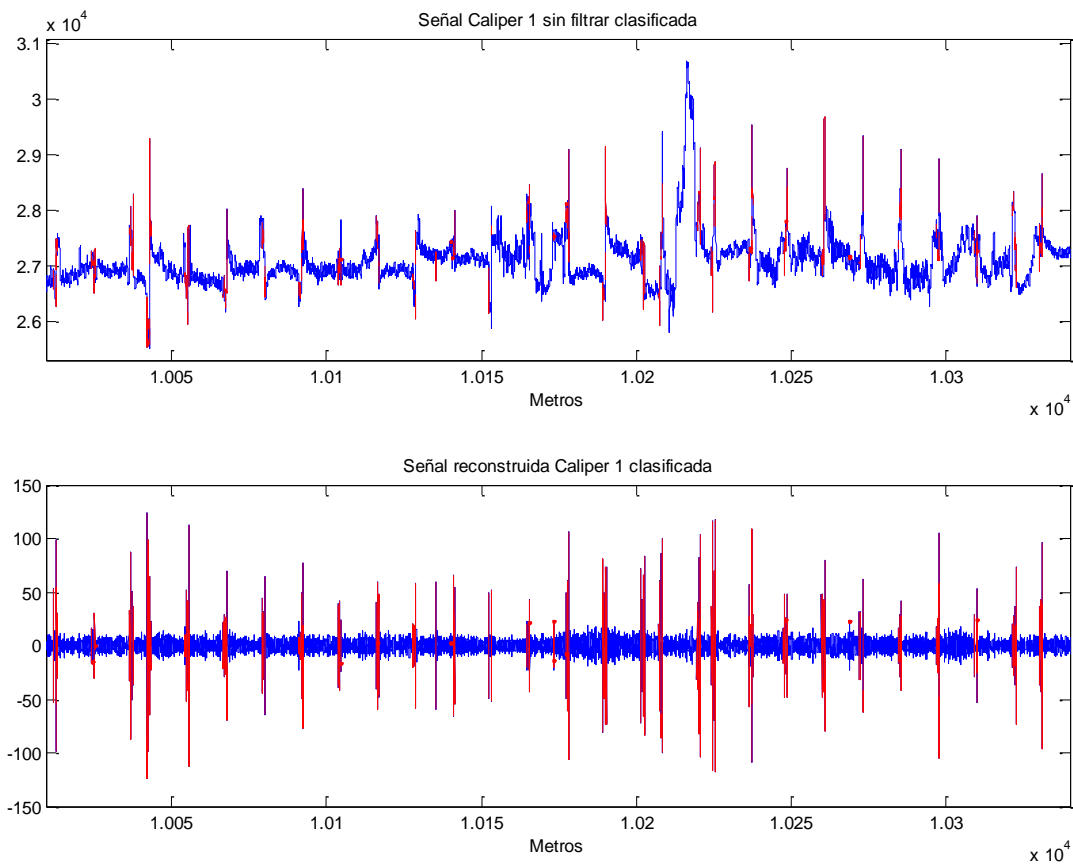


Figura 46. Señal reconstruida Caliper 1 clasificada. Fuente: Autor.

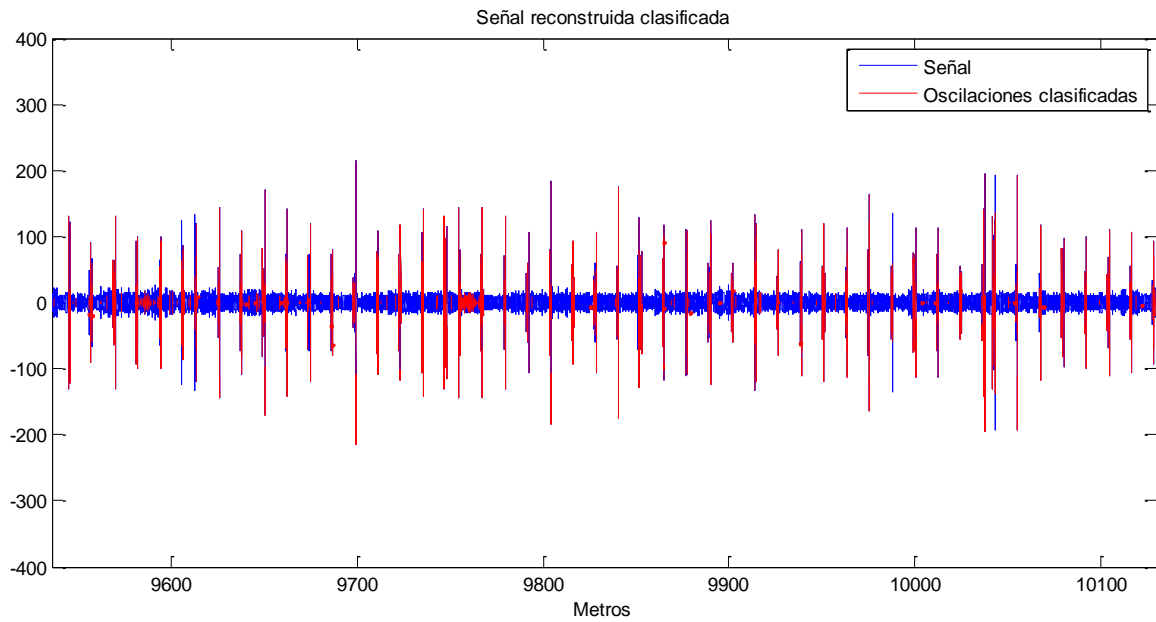


Figura 47. Sección de la señal reconstruida Caliper 1 clasificada. Fuente: Autor.

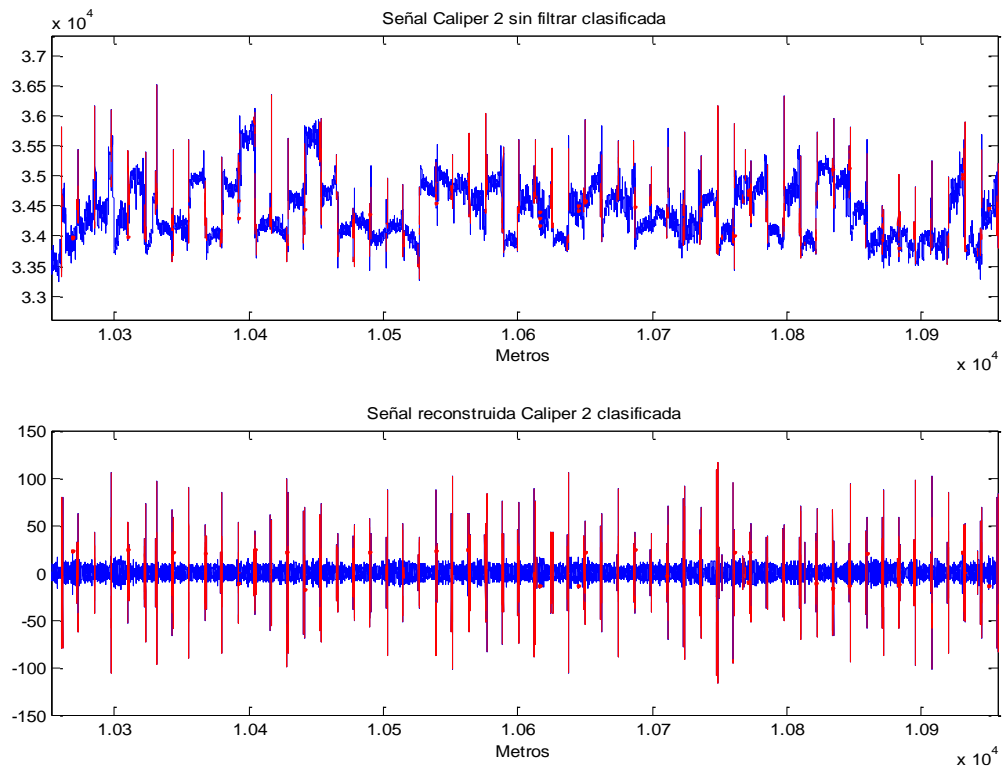


Figura 48. Señal reconstruida Caliper 2 clasificada. Fuente: Autor.

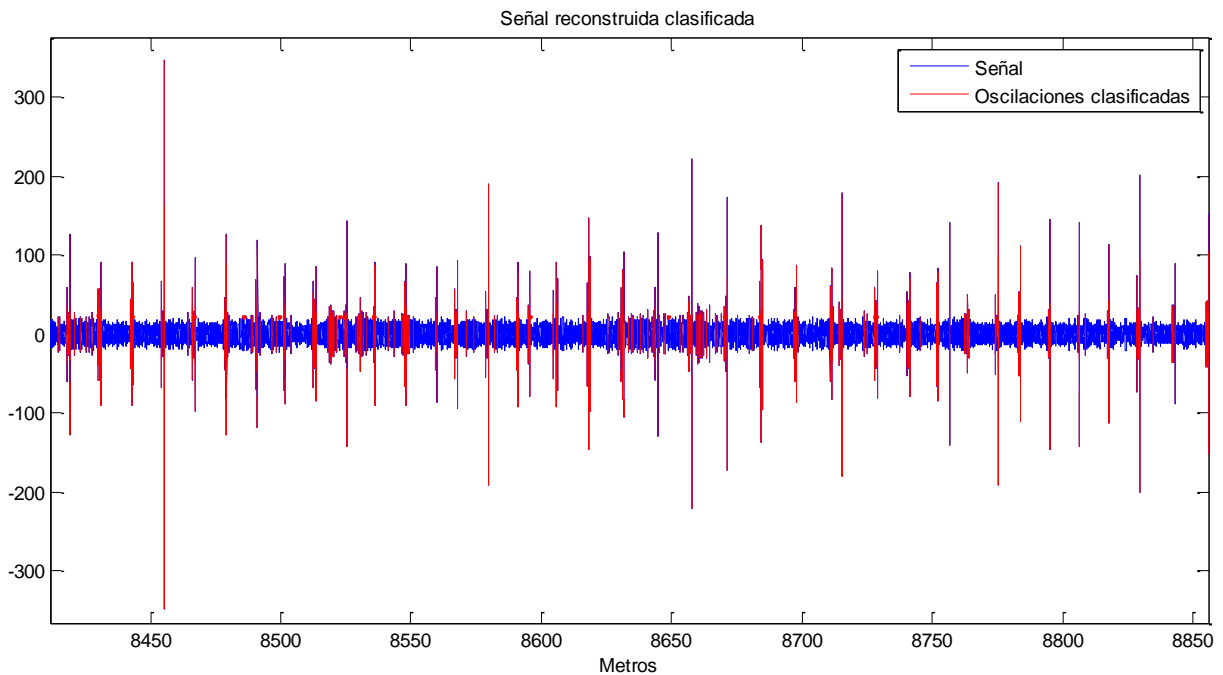


Figura 49. Sección de la señal reconstruida Caliper 2 clasificada. Fuente: Autor.

Debido a que la red neuronal se entrenó utilizando una porción pequeña de muestras de la señal reconstruida, al realizar una clasificación de la señal completa existen algunas oscilaciones que tienen un comportamiento diferente a la mayoría por esta razón existen errores en dicha clasificación.

En la Figura 50 se observa una oscilación de la señal donde algunas muestras de esta no se clasificaron correctamente, esto es debido al pico de gran amplitud que se encuentra en ella. La red neuronal no reconoce este tipo de datos debido a la poca cantidad de muestras en su entrenamiento comparado con la cantidad de las muestras de la señal completa.

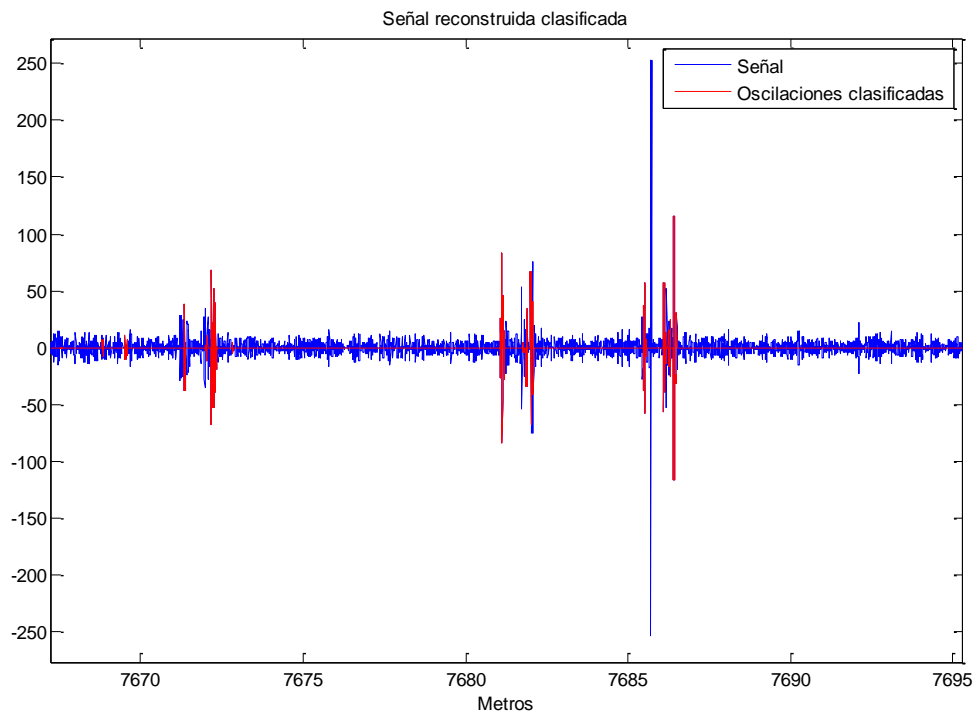


Figura 50. Falsos positivos clasificados. Fuente: Autor.

### 6.1.2 Post-clasificación

Luego de clasificar los datos utilizando una red, se proponen dos métodos de Post-clasificación con el fin de eliminar una gran cantidad de muestras mal clasificadas. El primer procedimiento consiste en recorrer la señal buscando flancos de subida, lugares donde se empieza a clasificar las muestras como oscilación, luego se cuentan las muestras clasificadas como oscilación justamente después del flanco. Se almacena la posición de cada una de estas muestras hasta que exista un flanco de bajada, esto se hace recorriendo toda el vector de etiquetas obtenidas a través de la red neuronal. Esto crea grupos de datos clasificados con un tamaño determinado, entre más pequeño sea el grupo, se puede considerar como un falso positivo, los grupos que tienen una mayor cantidad de muestras se consideran eventos significativos.

Se observó la cantidad de grupos que se obtuvieron anteriormente y la cantidad de muestras que posee cada uno, esto con el fin de encontrar la cantidad de



grupos clasificados que poseen la menor cantidad de muestras y escoger un umbral que solo discrimina estos considerados como falsos positivos.

En la Figura 51 se observa el histograma donde el eje x representa el ancho de grupos en muestras y en el eje y representa la cantidad de grupos que posee dicho ancho.

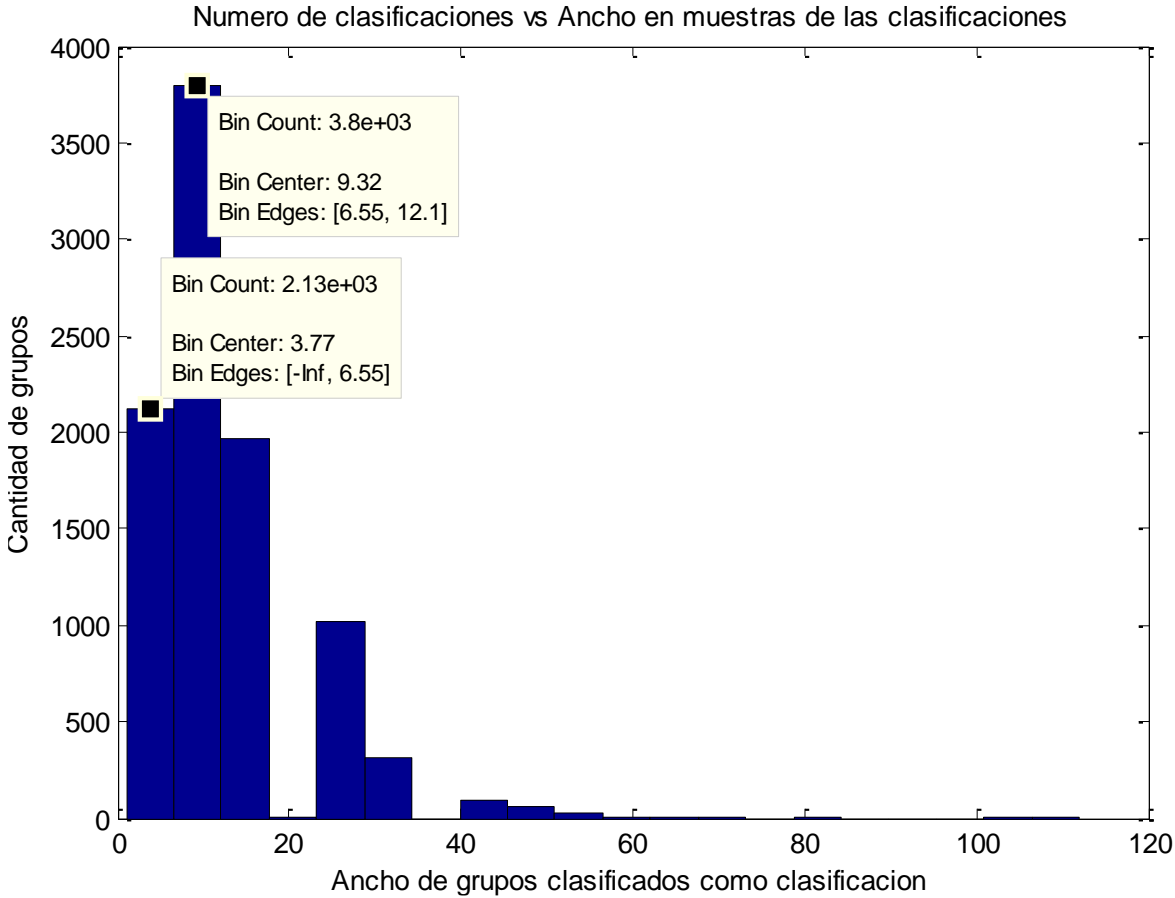


Figura 51. Numero de clasificación Vs Ancho en muestras de las clasificaciones.

Fuente: Autor.

Como se puede observar en la figura existe una gran cantidad de grupos de muestras clasificados como oscilación con un ancho menor a 10 muestras. Se propuso un umbral que cambiara la clasificación de grupos de datos que posean menos de este número de muestras con el fin de reducir la cantidad de falsos positivos en la señal. En la Figura 52 se observa una sección de la señal reconstruida clasificada, antes y después de este procediendo post-clasificación.

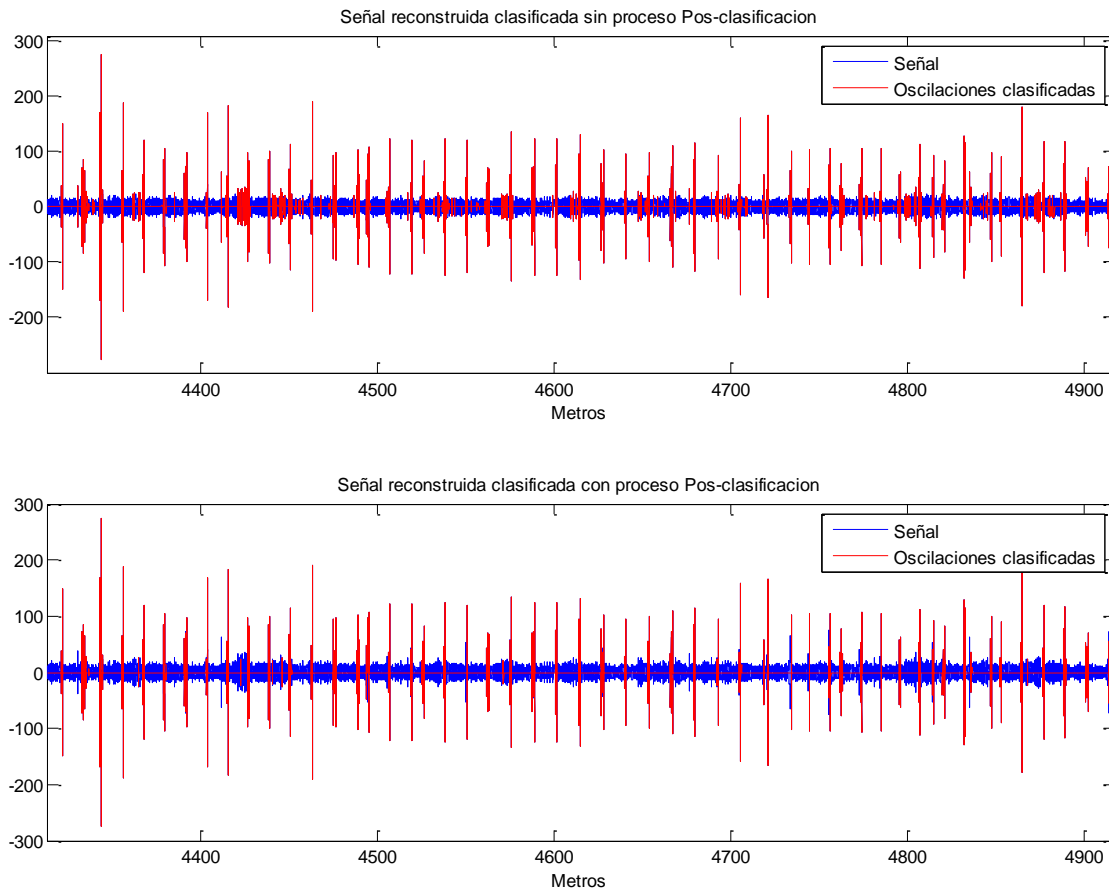


Figura 52. Señal clasificada antes y después del proceso post-clasificación.

Fuente: Autor.

Luego de reducir el número de falsos positivos en la señal, también se encontraron una gran cantidad de falsos negativos. Para que la clasificación sea más exacta se realiza otro procedimiento para eliminar una gran cantidad de muestras clasificadas como sin defecto estando en una oscilación.

El segundo procedimiento es muy similar al anterior, en primer lugar se detectan los flancos de bajada, es decir lugares donde la señal se empieza a clasificar como sin defecto luego de una muestra clasificada como oscilación. En segundo lugar se cuentan las muestras clasificadas como “sin defecto” justamente después del flanco además se almacena la posición de cada una, hasta que exista un flanco de subida o detección de oscilación. Esto crea grupos de datos que representan las muestras que se encuentran entre las oscilaciones detectadas,

donde no existe ningún evento para la red neuronal. En la Figura 54 se observa los grupos de datos obtenidos en este procedimiento marcado por las flechas.

Debido a la forma de onda de la oscilación en las señales caliper, existe un espacio en el centro de la oscilación que se puede confundir con el ruido de fondo, provocando que dichas muestras se clasifiquen incorrectamente. Para eliminar una gran cantidad de falsos negativos se observa un histograma donde se comparan la cantidad de grupos de datos clasificados como sin defecto con la cantidad de muestras que posee cada uno, esto se realiza para identificar el promedio de los tamaños en muestras de los grupos de datos que correspondieran a un falso negativo y cualquiera que posea menos muestras que esta cantidad se reclasifica como oscilación. En la Figura 53 se observa el histograma mencionado anteriormente.

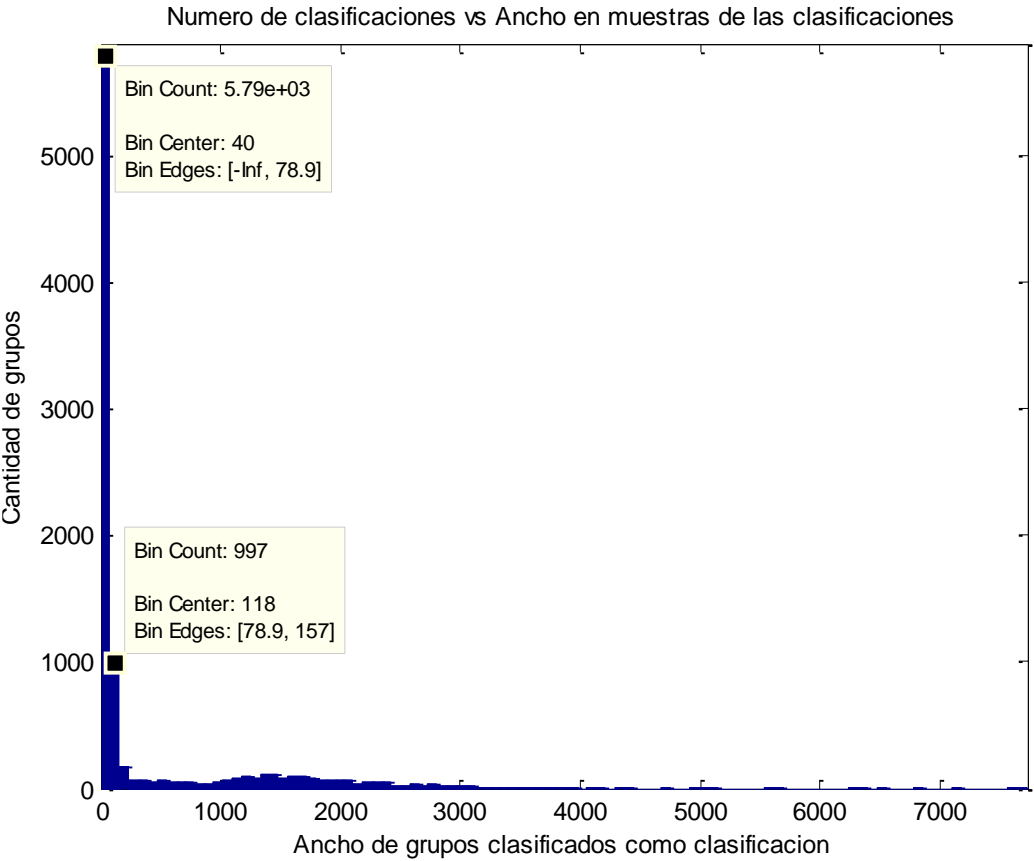


Figura 53. Numero de clasificaciones sin oscilación Vs Ancho de muestras de clasificaciones sin oscilación. Fuente: Autor.

Al observar la Figura anterior se observa que la mayor cantidad de grupos de datos poseen un tamaño menor a 120 muestras. Con el fin de no afectar de sobremanera la señal y eliminar la mayor cantidad de falsos negativos se toma el valor de muestras mencionado anteriormente considerando que el tamaño en muestras de un falso negativo es el evento que más se repite en la señal debido a que cada oscilación en esta tendría este problema de clasificación.

En la Figura 54 se observa una oscilación clasificada y un falso negativo en ella identificado con la flecha amarilla.

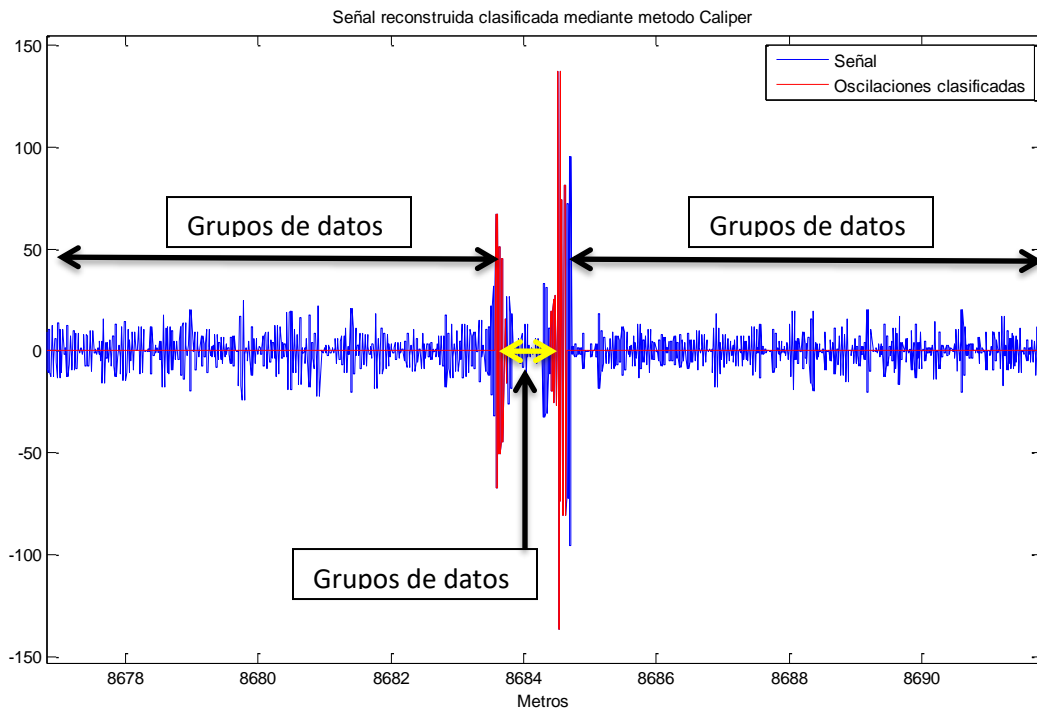


Figura 54. Falsos negativos en la señal reconstruida clasificada. Fuente: Autor.

En la Figura 55 se observa una oscilación antes del procedimiento de Post-clasificación y la misma oscilación luego del procedimiento.

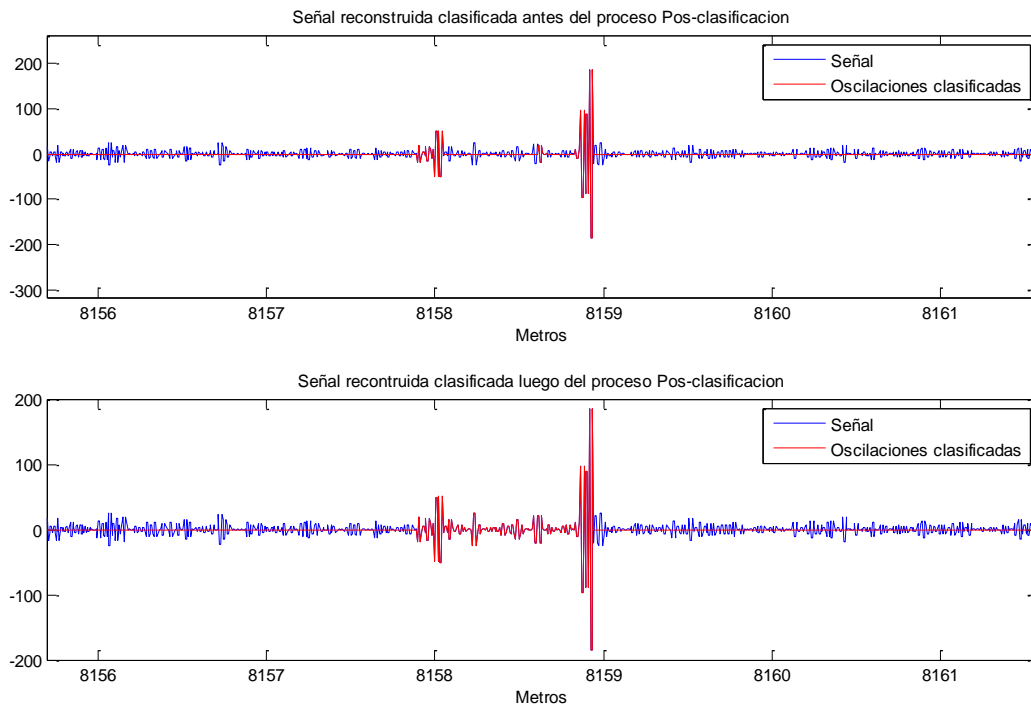


Figura 55. Oscilación antes y después del procedimiento Post-clasificación.

Fuente: Autor.

En la figura anterior se observa una de las oscilaciones que se clasificó correctamente luego del procedimiento. De igual manera existen una cantidad de oscilaciones que no se clasificaron correctamente no obstante mucho menor que la cantidad existente antes del proceso.

## 6.2 Comparación de métodos

En esta sección se realiza una comparación de dos metodologías para detectar oscilaciones dentro de las señales. La primera se explica más a fondo en el apéndice [21] (ver 8. Bibliografía), esta consiste en la clasificación de oscilaciones utilizando una señal de covarianza obtenida por ventaneo entre una muestra representativa de una oscilación y ventanas que se deslizan por una señal obtenida de un sensor MFL. La segunda metodología consiste en clasificar las oscilaciones utilizando la red neuronal mostrada en la Pág. 106 y un procedimiento

post-clasificación mostrado en la Pág. 114. Utilizando las señales obtenidas por un sensor tipo Caliper.

Se obtuvo el vector de datos clasificados entre las etiquetas 0 y 1 que corresponde a “sin defecto” y “defecto” respectivamente utilizando cada uno de los métodos para proceder a medir la distancia entre detección de oscilación, esto con el fin de identificar el evento que puede estar ocurriendo en la tubería para cada una de las oscilaciones.

Los eventos se clasifican utilizando la distancia que se existe entre ellos en la Tabla 12, se observan la clasificación de los eventos según la distancia.

<b>Evento</b>	<b>Intervalo de distancias entre eventos (Metros)</b>
Errores	0 – 1
Desconocidos	1 – 9
Soldaduras	9 - 13
Posibles Soldaduras	13 - $\infty$

Tabla 12. Clasificación de eventos clasificados según la diferencia de distancia entre ellos. Fuente: Autor.

En la Tabla anterior se muestran las distancias que hay entre un evento y otro, con su clasificación dependiendo de esta distancia. Los eventos que se consideran errores son problemas que persistieron, luego de la clasificación y los procesos post-clasificación nombrados anteriormente, como falsos negativos. Los eventos que se consideran desconocidos se reconocen ya que entre ellos existen distancias entre 1 y 7 metros, se les da esta denominación debido a que pueden ser cualquier evento físico dentro de la señal como reparaciones, válvulas, cambios de aceleración en la herramienta, etc. Los eventos nombrados como soldaduras son aquellos que poseen una distancia entre si entre 9 a 13 metros, debido a los análisis anteriores comparando la señal caliper y MFL se observó que estas distancias posiblemente correspondan a las soldaduras de la tubería. Por ultimo están los eventos de posibles soldaduras que son los que se encuentran a

más de 13 metros, se les llama de esta manera ya que una sección de tubería no puede tener más de esta distancia. Si existe una distancia entre eventos mayor esta es posible que no se haya detectado una oscilación en ese segmento

A continuación en las Figuras 56 y 57 se observan histogramas que comparan la distancia que corresponde a cada grupo de distancias y el tamaño en muestras de estos grupos. Estos histogramas corresponden a la clasificación utilizando el método MFL y el método Caliper respectivamente.

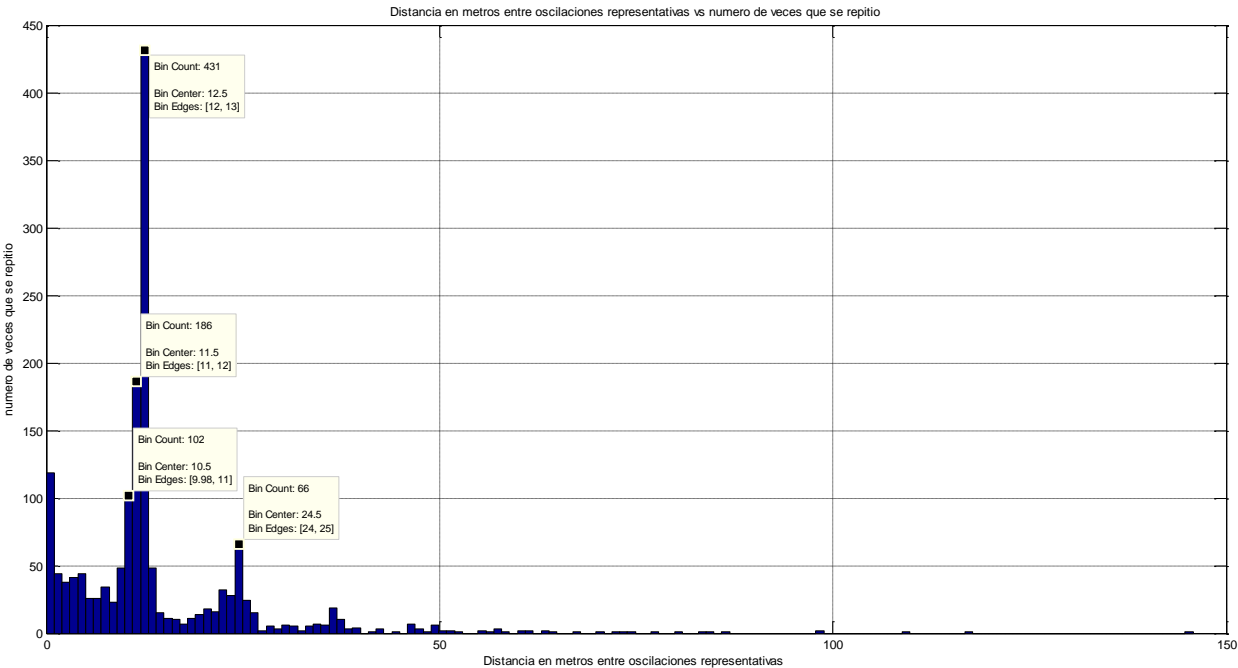


Figura 56. Distancia en metros entre oscilaciones respectivas y número de veces que se repite cada distancia utilizando el método MFL. Fuente: [21].

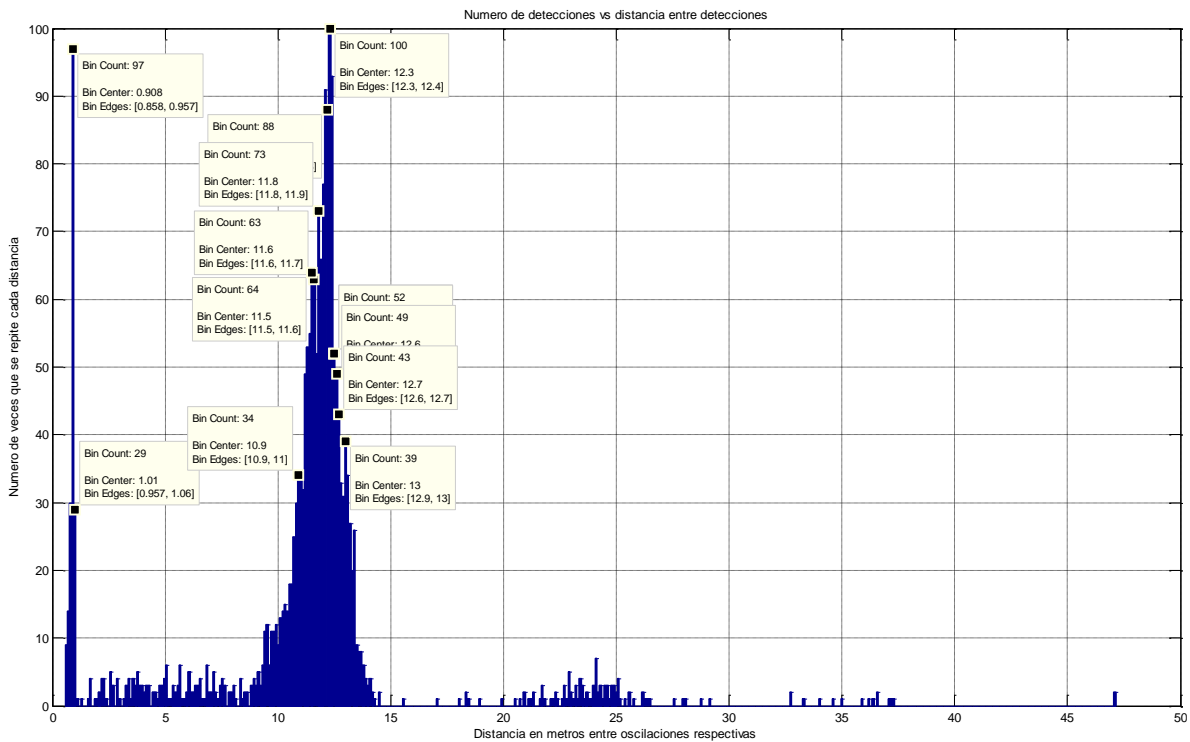


Figura 57. Distancia en metros entre oscilaciones respectivas y número de veces que se repite cada distancia utilizando el método Caliper. Fuente: Autor.

En las figuras anteriores en el eje x se encuentra la distancia que existe entre cada evento y en el eje y se encuentra la cantidad de estas distancias. Se puede observar un pico significativo que corresponde a distancias entre 9 a 13 metros considerados como soldaduras, este pico refleja la cantidad de eventos que están separados esta distancia, a la derecha de este pico se encuentran los eventos que están a más de 13 metros o lugares donde no se clasificaron oscilaciones, a la izquierda de este pico se encuentran dos tipos de eventos los desconocidos que son aquellos que se separan entre 1 a 9 metros y los errores que son aquellos que tienen una separación entre 0 y 1 metro. Implementando el método MFL se observa que existen una gran cantidad de oscilaciones no detectadas ya que se puede observar una distancia entre eventos de casi 150 metros lo que quiere decir que existe un segmento muy grande donde no se clasificaron las oscilaciones, de igual forma existe una menor cantidad de oscilaciones que se encuentran entre 9 a 13 metros comparado las detectadas en el método caliper.



Utilizando el método caliper se encontraron 1646 oscilaciones separadas de 9 a 13 metros que pueden considerarse como separaciones entre soldaduras. 138 distancias que superan los 13 metros es decir lugares donde no se clasificaron algunas soldaduras. 165 falsos negativos que persistieron luego del procesamiento post-clasificación y 196 eventos desconocidos que pueden ser falsos positivos o eventos como paso por válvulas, reparaciones de la tubería etc.

A continuación se observa algunos de los eventos que suceden dentro de las señales caliper que se encuentran separados por las distancias en que se mencionaron anteriormente, desde las más pequeñas hasta las mayores utilizando el método caliper.

En la Figura 58 se observa un evento considerado como error que persistió luego del procesamiento post-clasificación.

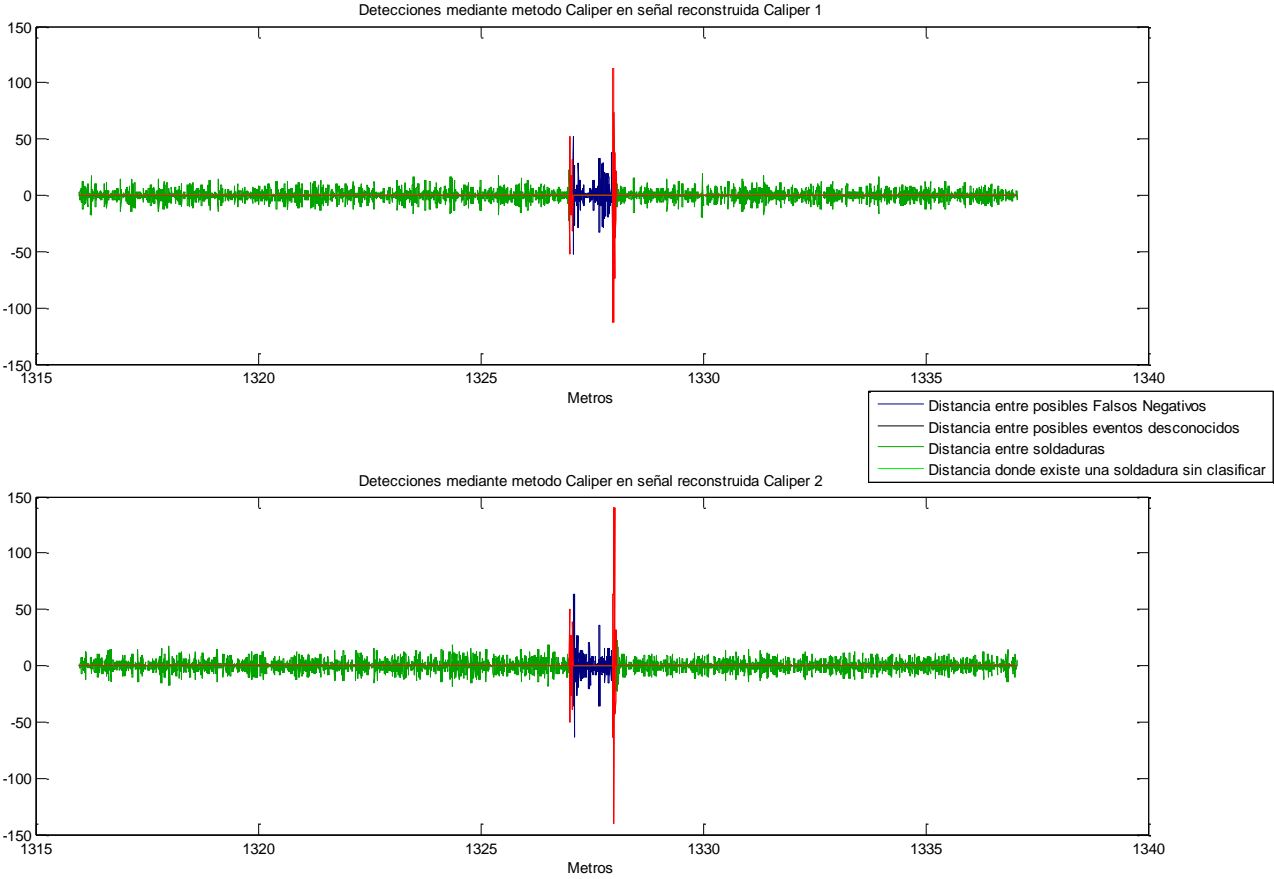


Figura 58. Falso positivo dentro de la señales caliper. Fuente: Autor.

En la Figura 59 se observa error que persistió luego del procesamiento post-clasificación.

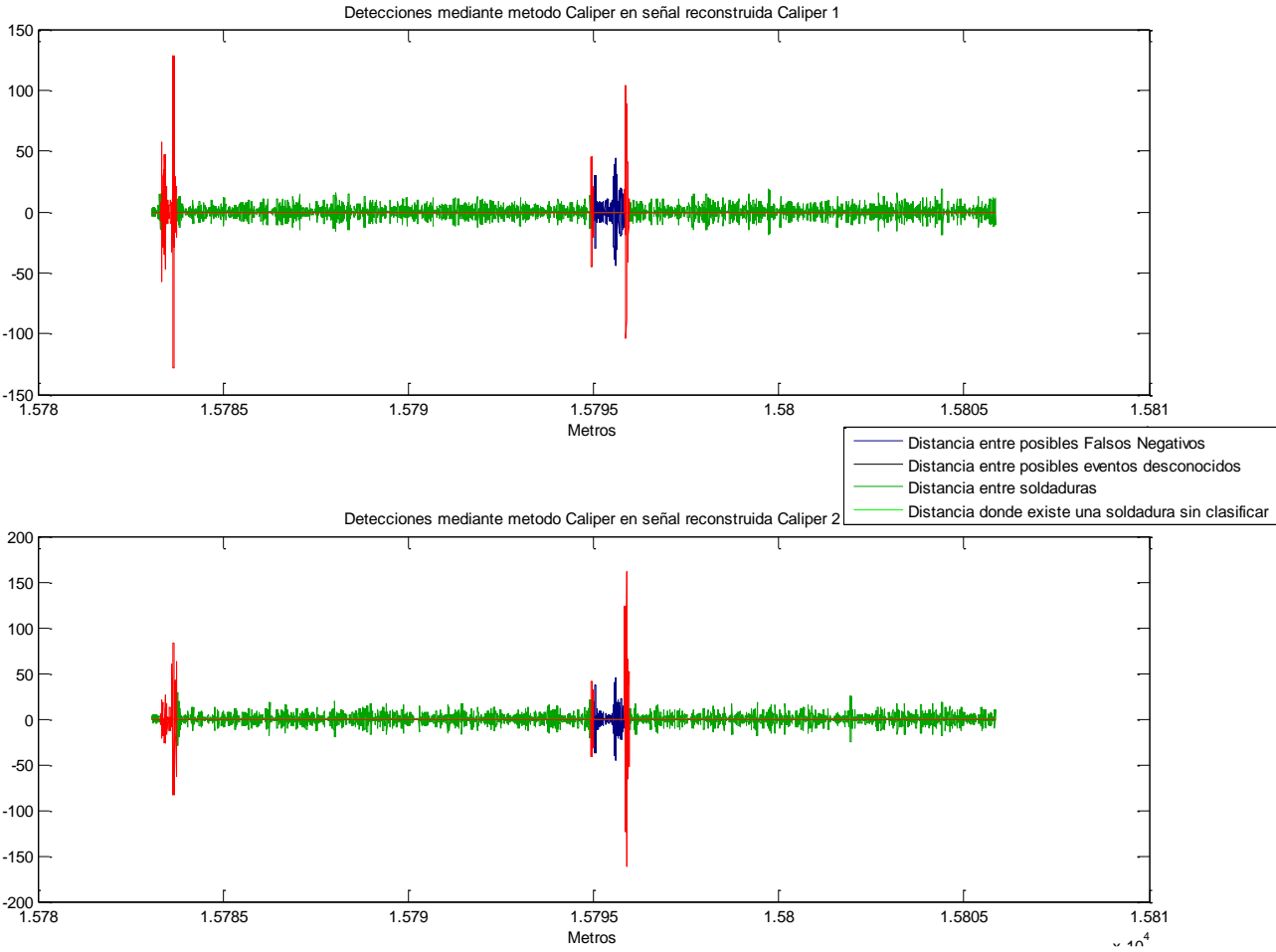


Figura 59. Falso positivo dentro de la señales caliper. Fuente: Autor.

En la Figura 60 se observa un evento desconocido que al observar la poca distancia entre las dos oscilaciones puede que se esté identificando en este lugar una reparación en la tubería.

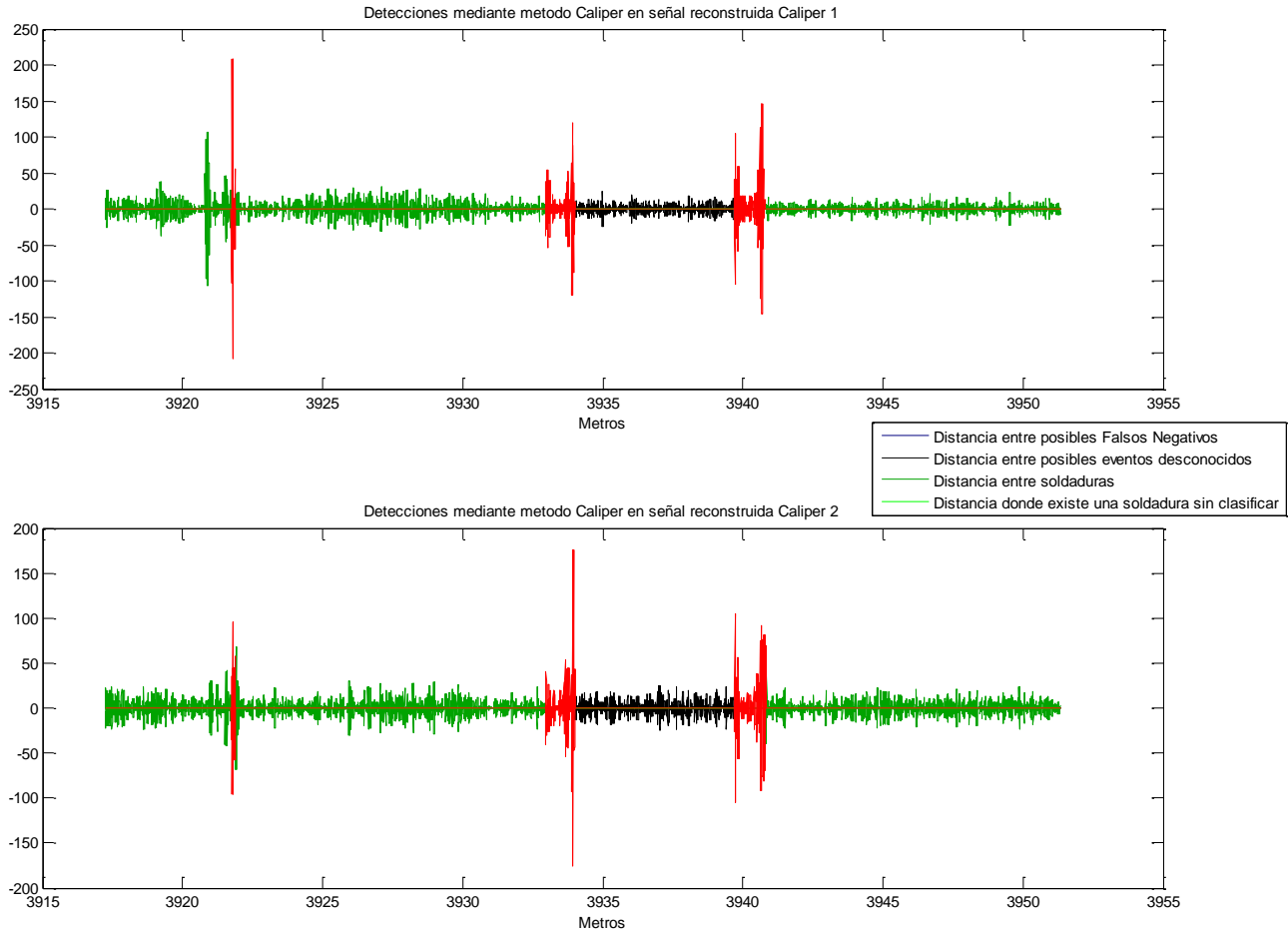


Figura 60. Evento desconocido dentro de la señal. Fuente: Autor.

En la Figura 61 se observa otro evento desconocido que al observar la poca distancia entre las dos oscilaciones puede que se esté identificando en este lugar una reparación en la tubería.

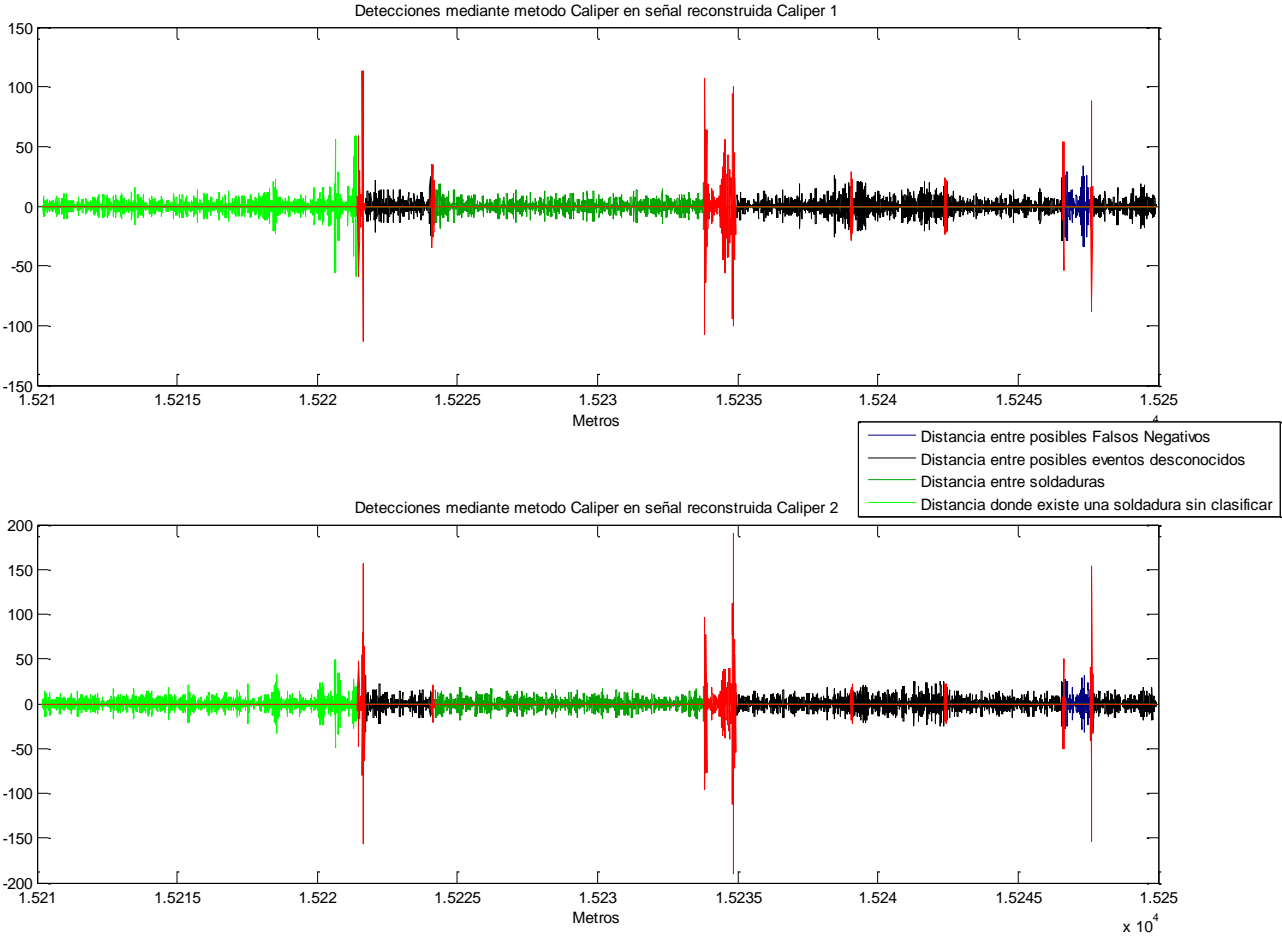


Figura 61. Errores de clasificación. Fuente: Autor.

En la Figura 62 se observan eventos considerados soldaduras poseen una distancia entre sí de 9 a 13 metros que podrían corresponder a las soldaduras de la tubería.

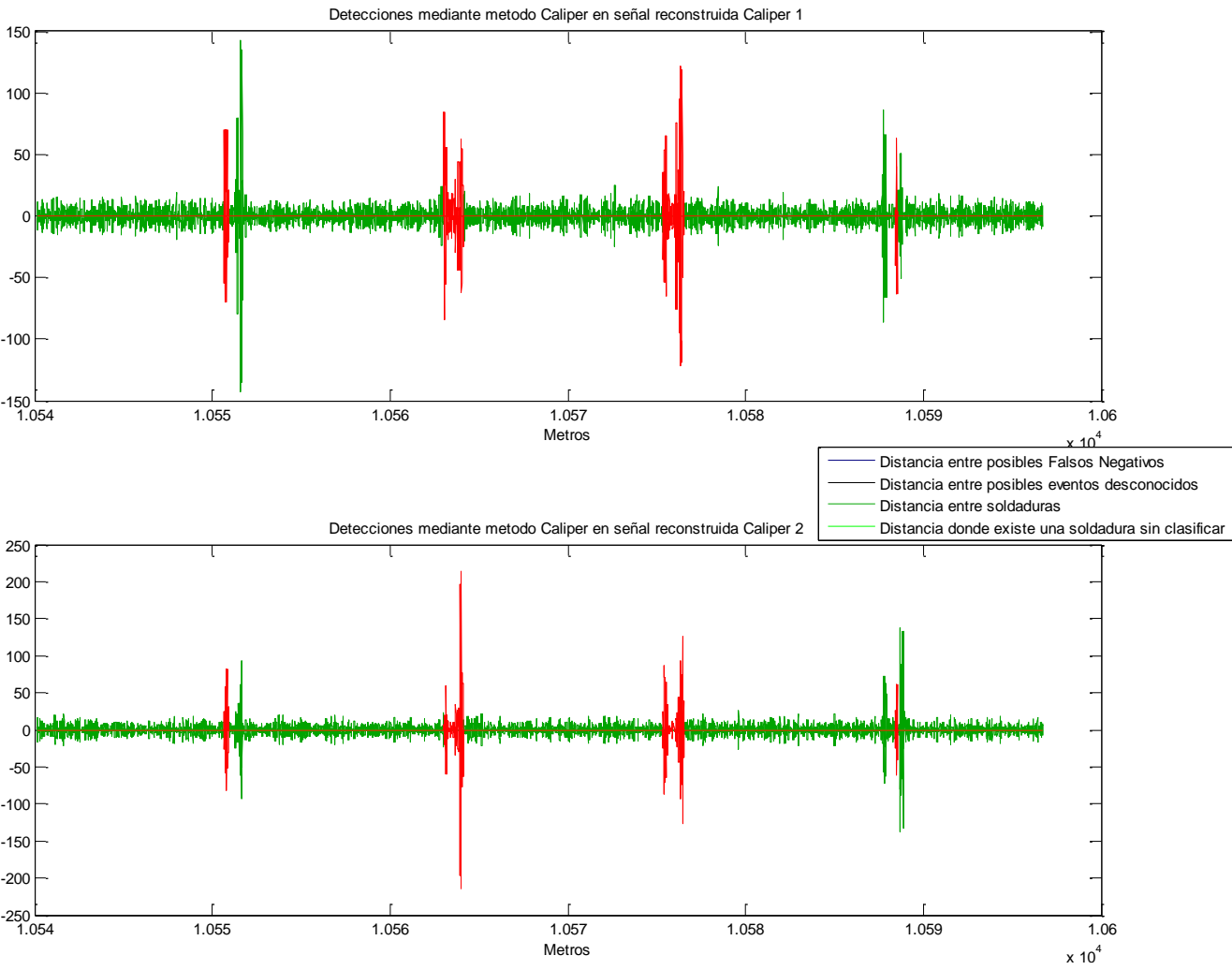


Figura 62. Oscilaciones con separaciones de 9 a 13 metros. Fuente: Autor.

En la figura 63 se observa otra sección de la señal donde se clasifica correctamente las oscilaciones que tienen una distancia entre sí de 9 a 13 metros que podrían corresponder a las soldaduras de la tubería.

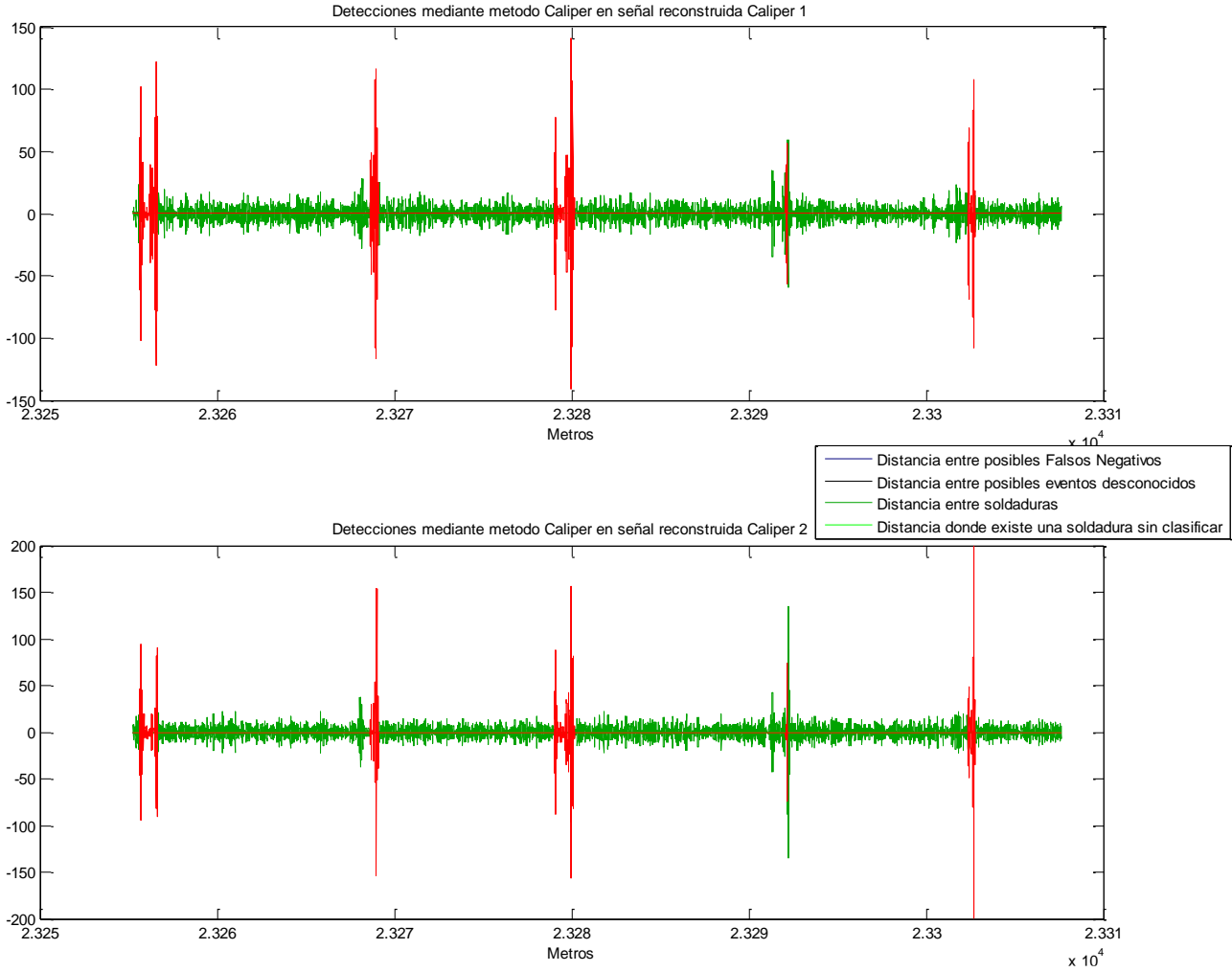


Figura 63. Oscilaciones con separaciones de 9 a 13 metros. Fuente: Autor.

En la Figura 64 se observan eventos considerados como posibles soldaduras debido a esto entre algunas de ellas existen distancias mayores a 13 metros.

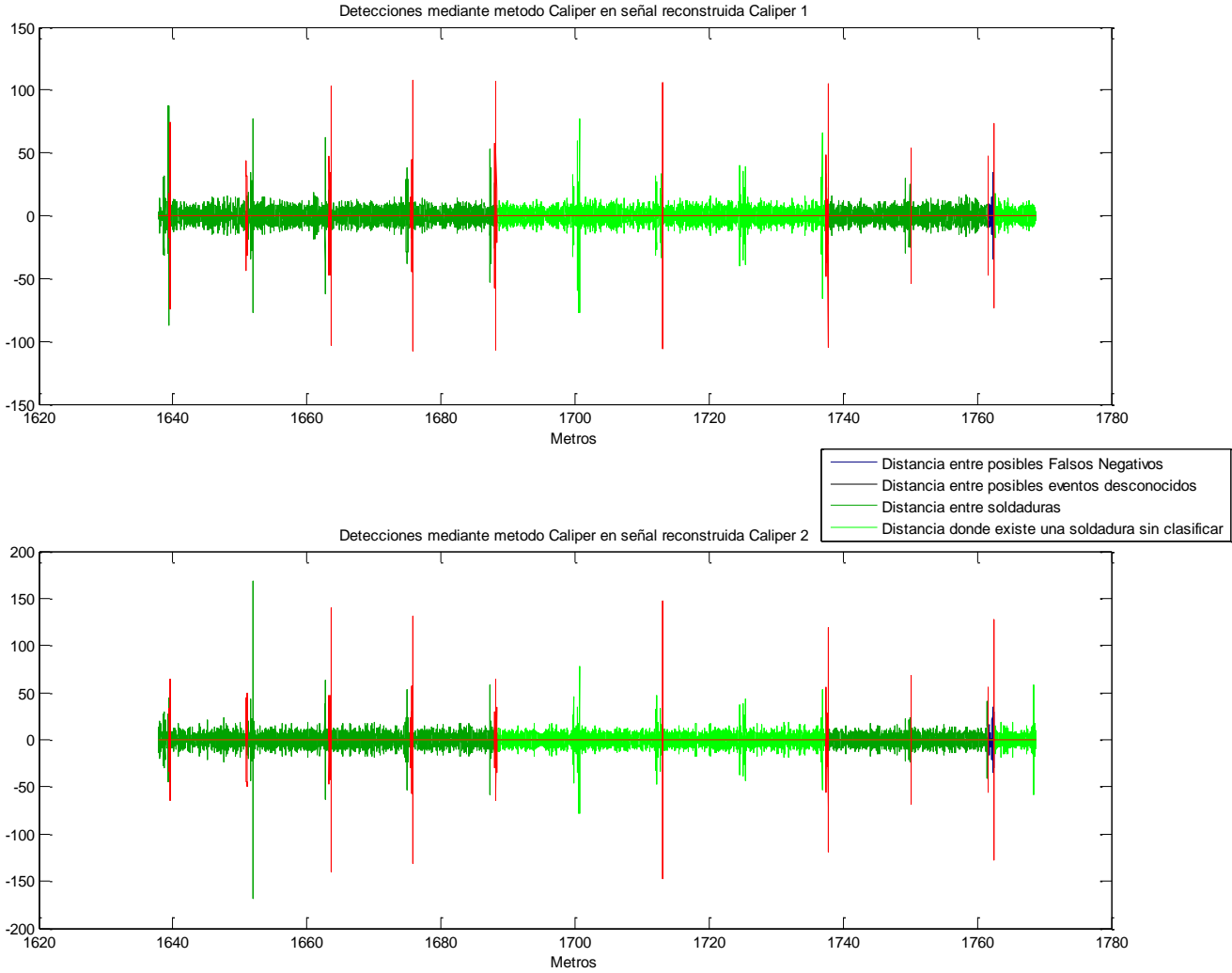


Figura 64. Oscilaciones sin clasificar. Fuente: Autor.



En la Figura 65 se observan otros eventos considerados como posibles soldaduras debido a esto entre algunas de ellas existen distancias mayores a 13 metros.

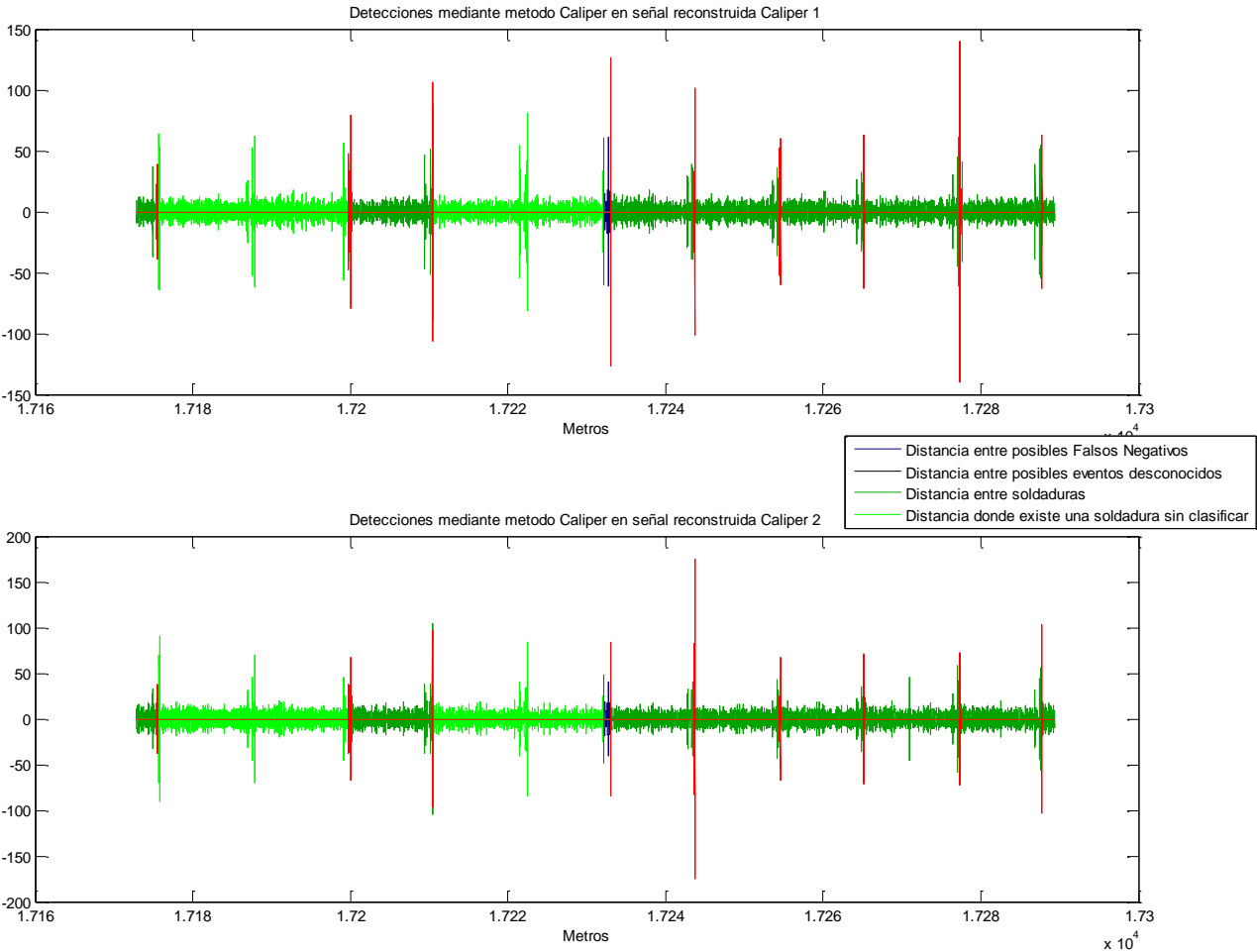


Figura 65. Soldaduras sin clasificar. Fuente: Autor.

## 6.2.1 Interfaz grafica

Utilizando la herramienta *GUIDE* para la creación de interfaces graficas de usuario o sus siglas en ingles GUI, se implementó una interfaz para realizar la comparación mencionada anteriormente. Dentro de esta se ejecutan los dos métodos secuencialmente y se obtienen las etiquetas respectivas para cada uno de ellos que representan las oscilaciones en la señal. Se grafica las señales y se comparan las muestras que se clasifican en cada método y las que se clasifican en común.

En la Figura 66 se observa la interfaz gráfica y la comparación de los métodos utilizando las señales MFL y Caliper.

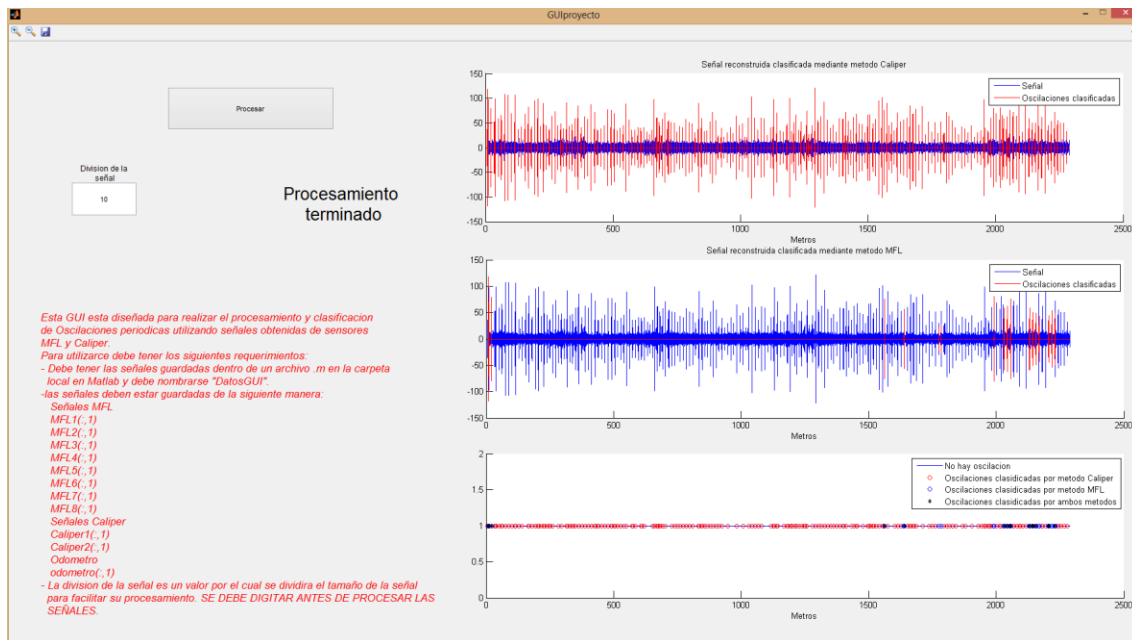


Figura 66. Interfaz Gráfica de comparación de métodos de clasificación de oscilaciones. Fuente: Autor.

En la Figura 66 se observa la comparación de los métodos Caliper y MFL mediante tres graficas que corresponden a la clasificación de oscilaciones utilizando el método caliper, la clasificación de oscilaciones utilizando el método MFL y la comparación de ambos métodos utilizando una gráfica donde se pueden

identificar las muestras clasificadas por cada método independientemente y las muestras que se clasificaron por ambos métodos.

En esta interfaz se realiza todo el procesamiento de la señal al pulsar el botón 'procesar' dentro de ella, también se puede dividir la señal en una sección más pequeña con el fin de acelerar su procesamiento ya que el tiempo que demora este, para todas las señales completas, es demasiado.

En la interfaz también se observa un texto explicativo para realizar la comparación y garantizar el correcto funcionamiento de la misma.

Luego de se implementó otra interfaz que tiene como función realizar todo el procesamiento del método caliper utilizando la red neuronal mencionada anteriormente (ver 6.1 Red Neuronal), realiza el procesamiento post-clasificación y grafica la señal con posibilidad de observar todos los eventos que se encontraron en ella, estos eventos se mencionaron anteriormente y de clasifican dependiendo de la distancia que existe entre una oscilación y otra (ver Pág. 58). En la Figura 67 se observa la interfaz mencionada anteriormente.

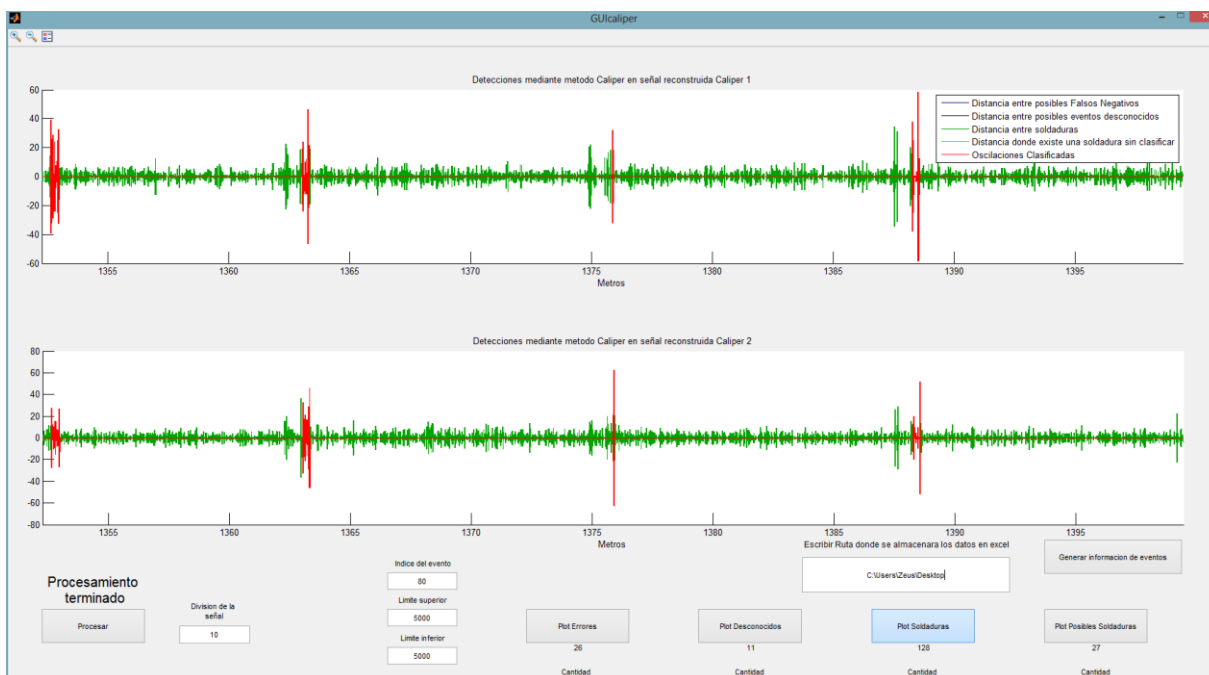


Figura 67. Interfaz de usuario método Caliper. Fuente: Autor.

En la esta interfaz igual que la anterior se realiza el procesamiento de los datos y se puede dividir la señal por un valor con el fin de disminuir el tiempo de procesamiento. También posee opciones para graficar cualquier evento que se haya encontrado en la señal, basado en las distancias que existen entre las oscilaciones detectadas, también se puede escoger la cantidad de muestras que se van a graficar con el fin de observar detalladamente o generalmente los eventos. En la interfaz se imprimen las dos señales Caliper reconstruidas y se grafican las respectivas etiquetas encontradas, de esta manera se pueden comparar los eventos que pasaron en una distancia específica en ambas señales, también posee la opción de generar las distancias donde se encuentran cada uno de los eventos de cada uno de los cuatro grupos de datos y el porcentaje que posee cada uno de estos con respecto a todos los eventos detectados en un archivo.xls.

### **6.2.2 Análisis de eventos en marcadores**

En esta sección observaremos los marcadores entregados por la CIC utilizando la interfaz gráfica mostrada anteriormente, con el fin de comprobar si en estos se pueden encontrar eventos que correspondan a algún defecto que se encuentre en este lugar además de las oscilaciones que se repiten en toda la señal.

Para esto se busca el lugar exacto en metros dentro de la interfaz gráfica y se observa si existe un evento desconocido en el intervalo donde se encuentra el marcador.

Debido a que algunos marcadores se encuentran solapados entre sí, en una gráfica se pueden analizar varios marcadores. Estos marcadores se observan en la Tabla 3 (ver pág. 13).

### 6.2.2.1 Marcadores TA1, TA2 y GR1

En la Figura 68 se observa en la interfaz gráfica el intervalo de distancia donde se encuentran los marcadores TA1, TA2 y GR1. Los eventos que se presentan en estos marcadores se encuentran marcados de color rojo en la Figura. Debido a que los marcadores se encuentran solapados no se conoce exactamente los eventos que corresponden a cada uno de ellos.



Figura 68. Marcadores TA1 y TA2 observados a través interfaz gráfica. Fuente: Autor.

### 6.2.2.1 Marcador TA3

En la Figura 69 se observa el lugar en metros donde se encuentra intervalo de 14 metros del marcador TA3, se observa que el clasificador no identifico un evento diferente a soldaduras en este marcador.

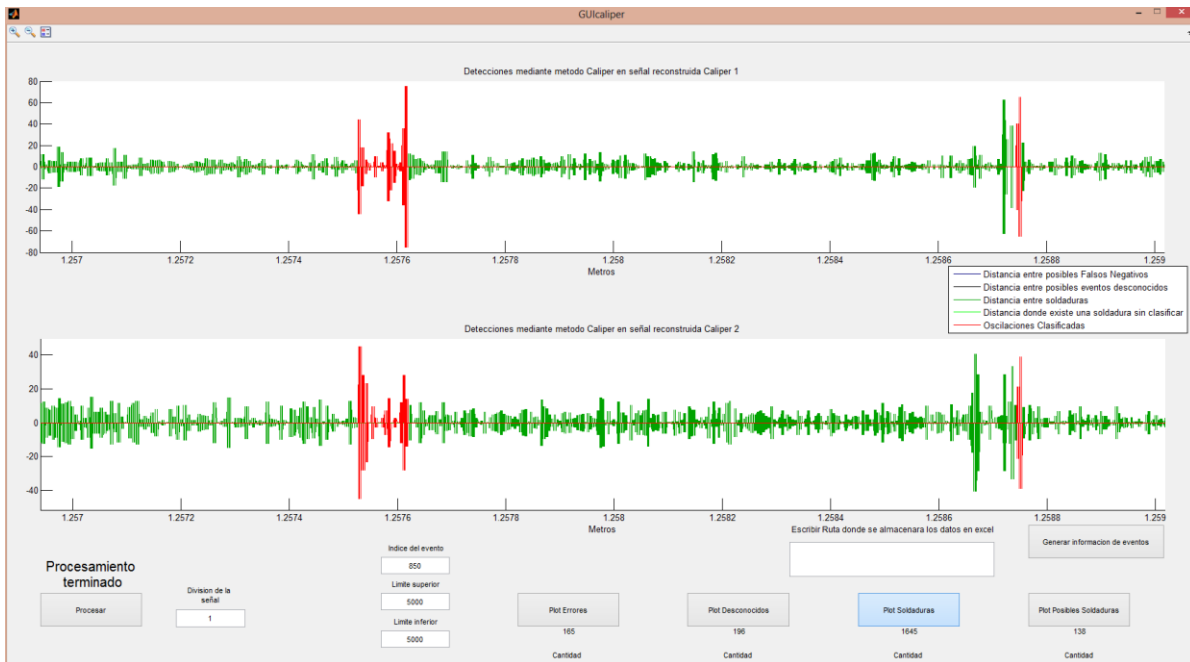


Figura 69. Marcador TA3 observados a través interfaz gráfica. Fuente: Autor.

### 6.2.2.2 Marcadores TA4, TA5, TA6, TA7, TA8 y GR3

En la Figura 70 se observan eventos desconocidos que podrían pertenecer a los marcadores mencionados en esta sección, también existe una oscilación bastante grande en medio de dos picos que poseen la misma forma de una oscilación de soldadura. Debido a que los marcadores se encuentran solapados, no se conoce exactamente que evento corresponde a cada uno de ellos.

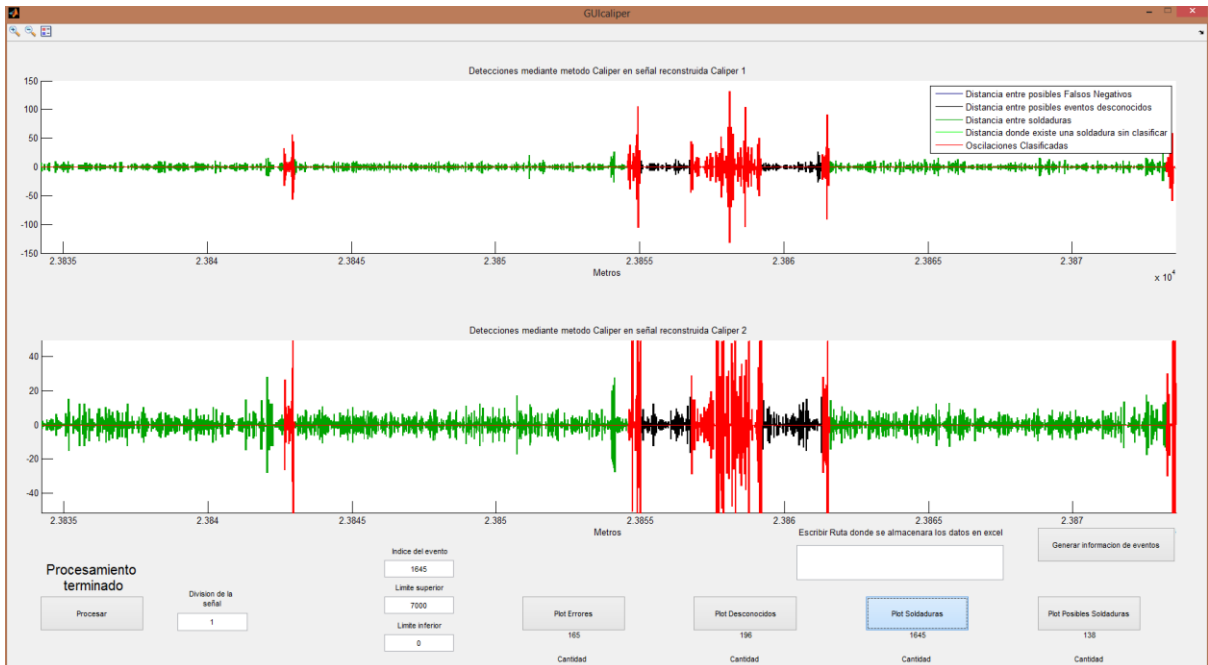


Figura 70. Marcadores TA4, TA5, TA6, TA7, TA8 y GR3 observados a través de la interfaz gráfica. Fuente: Autor.

### 6.2.2.3 Marcador GR2

En la Figura 71 se observa el intervalo del marcador GR2 donde se encuentra una soldadura y un evento que se considera desconocido que podría ser el que puede estar representando este marcador.

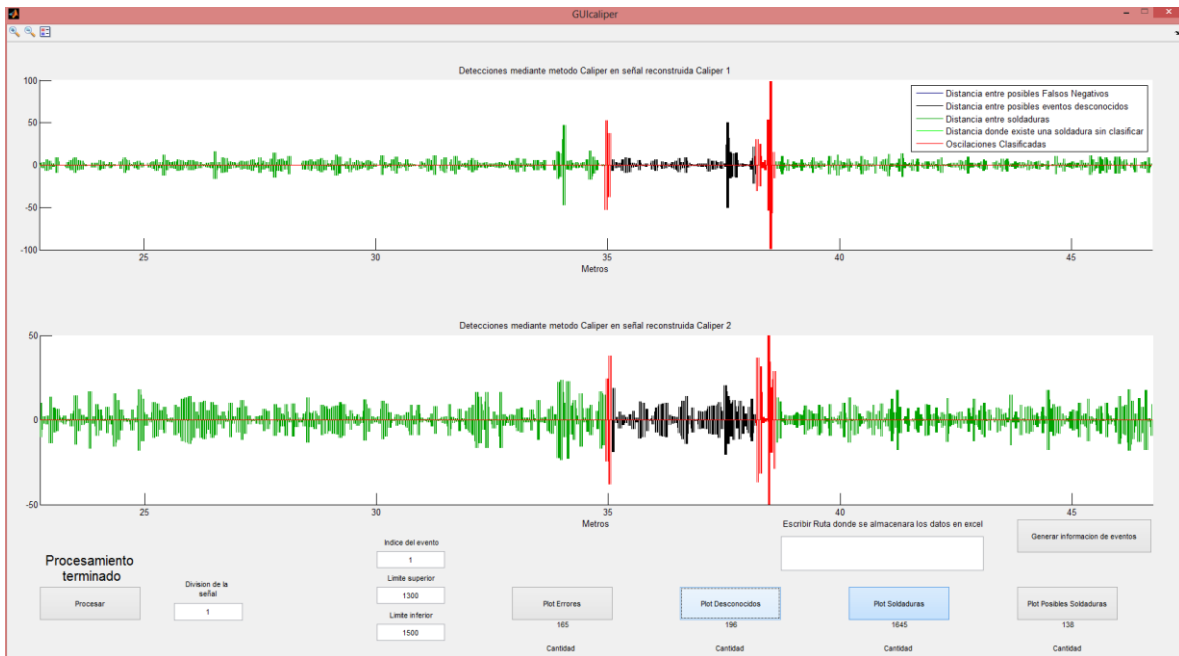
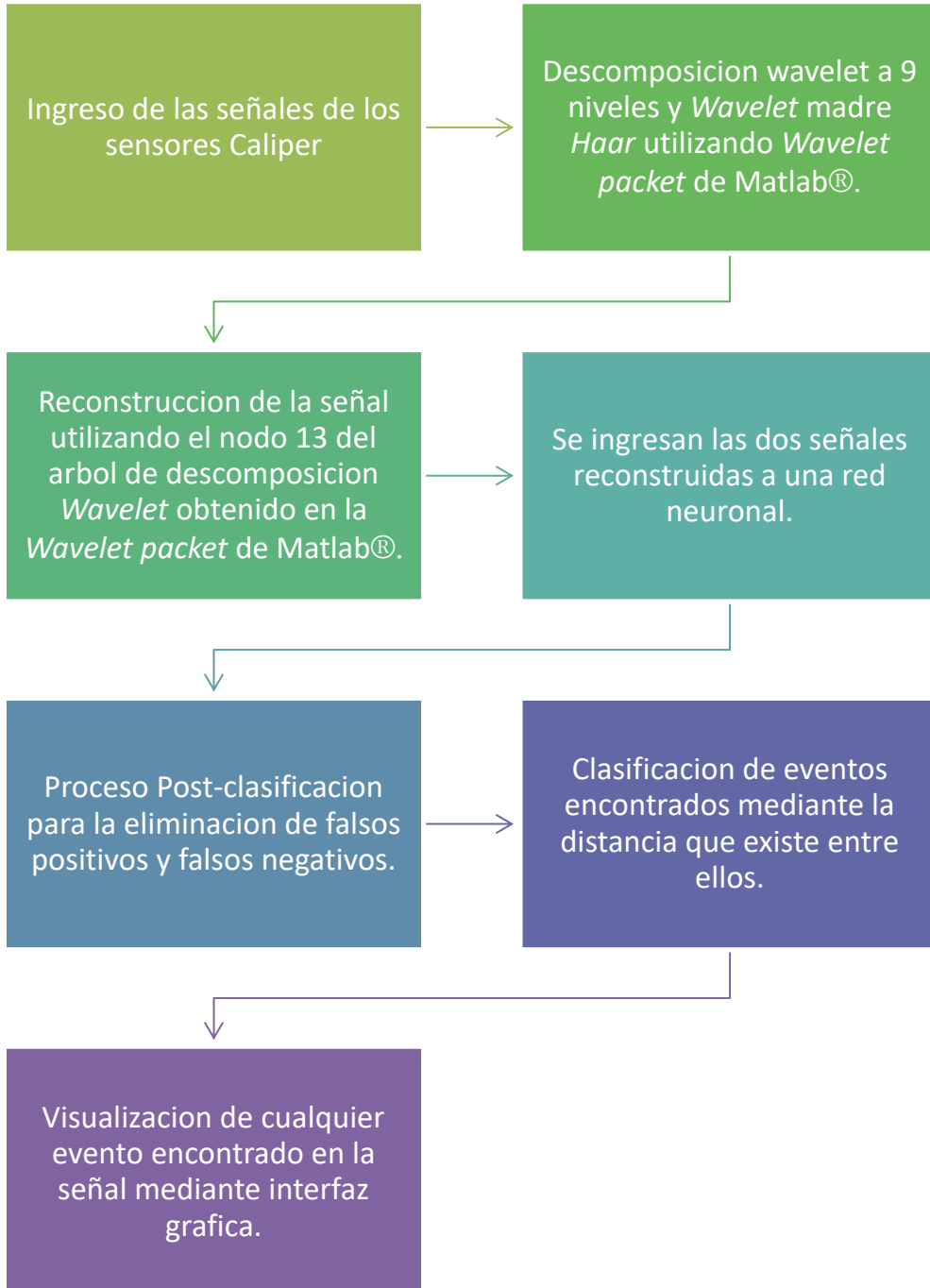


Figura 71. Marcador GR2 observados a través de la interfaz gráfica. Fuente: Autor.



### 6.3 Metodología Caliper

A continuación se muestra los pasos realizados para clasificar y visualizar los eventos encontrados en las señales del sensor caliper del dispositivo ITION. Los códigos que realizan los procedimientos de procesamiento de señales, clasificación de datos y post-procesamiento se pueden observar en el Anexo 5.



## 7. Conclusiones

- Debido a que la herramienta toma los datos a velocidad variable, se debe tener en cuenta la cantidad de muestras de cada una para realizar el procesamiento de las señales y la extracción de características de las mismas.
- En el proceso de clasificación y validación se debe aleatorizar muy bien los datos, ya que de esta manera se puede obtener elevados porcentajes de clasificación, al enseñarle a la red la mayoría de comportamientos que se puede encontrar en datos nunca antes vistos.
- Debido a que para realizar la validación del modelo se utilizan métodos iterativos, se debe tener en cuenta la cantidad de datos que se van a procesar, la cantidad de iteraciones, y el procedimiento a realizar, ya que, por estas razones, el tiempo de procesamiento puede aumentar de una manera significativa, además de necesitar mucha más potencia de cálculo.
- Al diseñar un filtro para una señal se debe tener en cuenta que a pesar de que esta tenga una gran cantidad de muestras, la información puede estar almacenada en frecuencias muy altas y es posible que al filtrar esta información se pierda.
- Al tener una señal con mucha cantidad de ruido la correlación de esta señal con un marcador de alta frecuencia no es muy efectivo ya que este puede ser muy similar al ruido y no aportaría información relevante.
- Si se necesita encontrar una anomalía en una señal filtrada que este en una frecuencia alta realizando una descomposición *Wavelet* es probable que no se pueda observar correctamente debido a que esta se puede mezclar con otros componentes de alta frecuencia y la señal necesitaría diferentes niveles de descomposición para diferentes partes de la señal dificultando el procedimiento.
- El procedimiento realizado para encontrar soldaduras en las señales caliper por medio de un umbral obtenido por los estadísticos de los marcadores no encuentra la totalidad de ellas debido a que las oscilaciones que

representan las soldaduras en los marcadores no son muy representativas en la señal por esta razón el número de soldaduras encontrada baja.

- Otra razón por la cual no se encuentran la totalidad de las soldaduras de la señal utilizando los sensores tipo caliper es por la falta de estos mismos. Se necesitan una cantidad mayor de ellos para detectar todas las soldaduras.
- Para realizar una un proceso de clasificación de obstrucciones o cambios de diámetro en la señal se deben identificar y etiquetar de manera exacta las muestras que correspondan a dicha anomalía ya que si una etiqueta cuenta con demasiado margen de error, la clasificación no se puede realizar correctamente y más en una señal con una cantidad muy grande de muestras.
- Para realizar una clasificación más eficiente utilizando las señales caliper sería necesario poseer datos de más sensores de este tipo.
- Se debe tener en cuenta que todas las oscilaciones periódicas no poseen una distancia de 10 a 12 metros entre sí, en el caso de algunas se encuentran mucho menos de esa distancia y utilizando el método de flancos de subida separado a dichas distancias se dejarían de clasificar una cantidad considerable de soldaduras.
- Debido a la forma de la oscilación en la señal caliper reconstruida, es probable que se generen falsos negativos dentro de la clasificación debido a que una sección de la oscilación se puede mezclar con el ruido de fondo de la señal y confundir a la red neuronal.

## 8. Bibliografía

- [1] H.R.Yoo, S.H.Cho, Y.W.Ryo, S.H.Park y G. S Park, Desing of a Permanent Magnet to Maximize the Sensing Signals in Geometry PIG Caliper System, International Conference on Electrical Engineering, (2002).
- [2] Inessa L Yablonskikh, Sarah J Dawson, Raja Venkatanarayanan. Assessment And Analysis Of Pipeline Buckles. PPL Pipeline Solutions, Cramlington, UK. (2007).
- [3] Dong Kyu Kim, Sung Ho Cho, Seoung Soo Park, Hui Ryong Yoo, Yong Woo Rho. Desing And Implementation of 30" Geometry PIG. R&D división, Korea Gas Corporation. (2003).
- [4] David Russel, Pigging in Pipeline Pre-commissioning, Weatherford & Specialty Services UK. (2005).
- [5] D. Isa, R. Rajkumar y KC Woo. Pipeline Defect Detection Usisng Support Vector Machines. School of Electrical and Electronic Engineering y University of Nottingham. (2007).
- [6] Tao Jin, Peiwen Que y Zhengsu Tao. Designing and Signal Processing of Intelligent Inspection Pig Applying Magnetic Flux Leakage Methods. Dept. of Information Measurement technology and Instrumental. (2004).
- [7] Tomas Arredondo Vidal. Introduccion a las Redes Neuronales {diapositivas}. Depto Electrónica Universidad Técnica Federico Santa Maria. (2012)
- [8] Gustavo A. Betancourt. Las Maquinas de Soporte Vectorial. Universidad Tecnologica de Pereira. (2005).
- [9] Ph.D Carlos Julio Arizmendi P. Transformada *Wavelet* {diapositivas}. Facultad de Ingenierías Fisicomecanicas de la Universidad Autónoma de Bucaramanga. (2013)
- [10] Donoho, David L. (1993). "Different Perspectives on Wavelets". American Mathematical Society. Estados Unidos. Páginas 168.

- [11] Jheyston Omar Serrano Luna. Sistema Caliper Para medir Deformaciones de la Geometría Interna en las Tuberías. CIC. (2014).
- [12] Breeding, Andy (2004). The Music Internet Untangled: Using Online Services to Expand Your Musical Horizons. Giant Path. p. 128.
- [13] Ph.D Carlos Julio Arizmendi, Signal Processing Techniques for Brain Tumour Diagnosis from Magnetic Resonance Spectroscopy Data. Departament de Llenguatges i Sistemes Informatics, Universitat Politècnica de Catalunya. (2011).
- [14] P.D Agoris, S. Meijer, E. Gulski, J.J. Smit. (2004). "Threshold selection for wavelet denoising of partial discharge data". Conference record of the 2004 IEEE International Symposium on Electrical Insulation. Páginas 62-65.
- [15] Frank Wilcoxon. Individual Comparisons by Ranking Methods. Biometrics. (1945).
- [16] Devijver, P. A., and J. Kittler, Pattern Recognition: A Statistical Approach, Prentice-Hall, Londres. (1982).
- [17] Jeff Schneider, The holdout method, The school of computer science. (1997).
- [18] Payam Refaeilzadeh, Lei Tang, Huan Lui. K-fold Cross-Validation, Arizona State University. (2008).
- [19] FH Joanneum. Cross-Validation Explained, Institute for Genomics and Bioinformatics. (2005-2006)
- [20] Emmendoerfer, Gustavo. "desenvolvimento e implementação de um sistema de medição inercial de trajetórias de dutos". Brasil. Páginas 159. (2007)
- [21] Freddy Reyes. "Interpretación de señales para la detección de referencias en inspecciones instrumentadas de tuberías". Facultad de Ingenierías Fisicomecánicas de la Universidad Autónoma de Bucaramanga. (2014).
- [22] MathWorks®. "MATLAB®. Creating Graphical User Interfaces". (2014).