

Optimización en la Detección de Ondas Gravitacionales Integrando Análisis Topológico y CNN

07 de junio de 2024

Equipo 04

Ángel Azahel Ramírez Cabello - A01383328

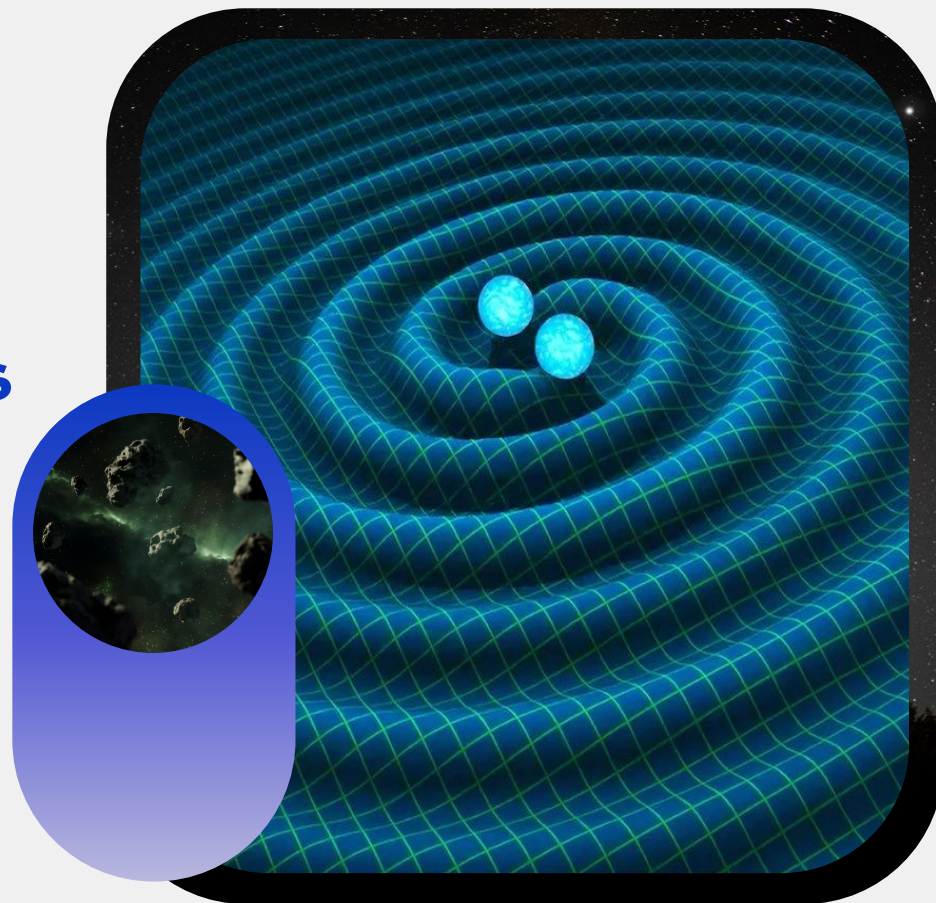
Annette Pamela Ruiz Abreu - A01423595

Avril Michelle Ruiz Martínez - A00833018

Franco Mendoza Muraira - A01383399

Jorge Raúl Rocha López - A01740816

Luis Angel López Chávez - A01571000



Contenido

01

Introducción

Descripción del problema y objetivos del trabajo

03

Resultados Finales

Modelos, sus métricas y la comparación entre estos

02

Metodología

Procedimientos, exploración e insights relevantes para el resultado final

04

Conclusiones

Posibles mejoras y resultados destacados

01

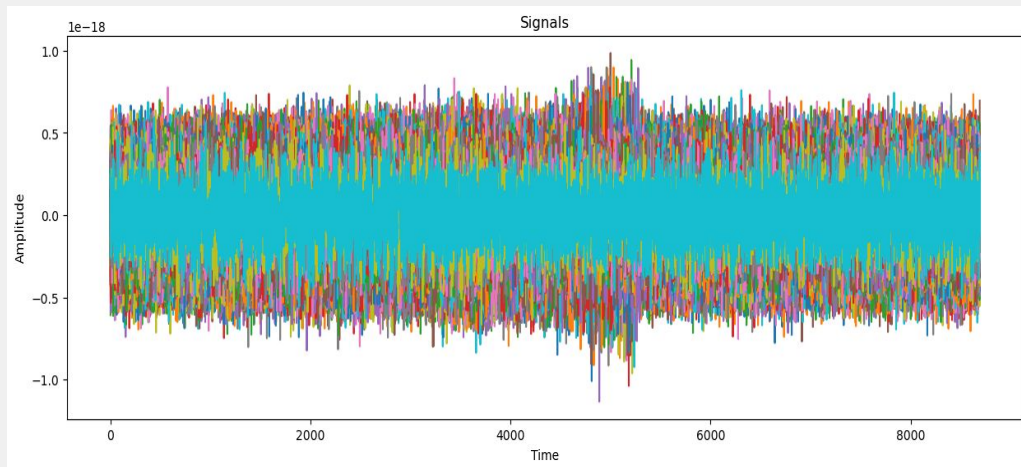
Introducción

Problema & Objetivo



Problema

¿Cómo separamos las ondas gravitacionales del ruido detectado?



Objetivos

Desarrollo de una CNN para la detección de ondas gravitacionales

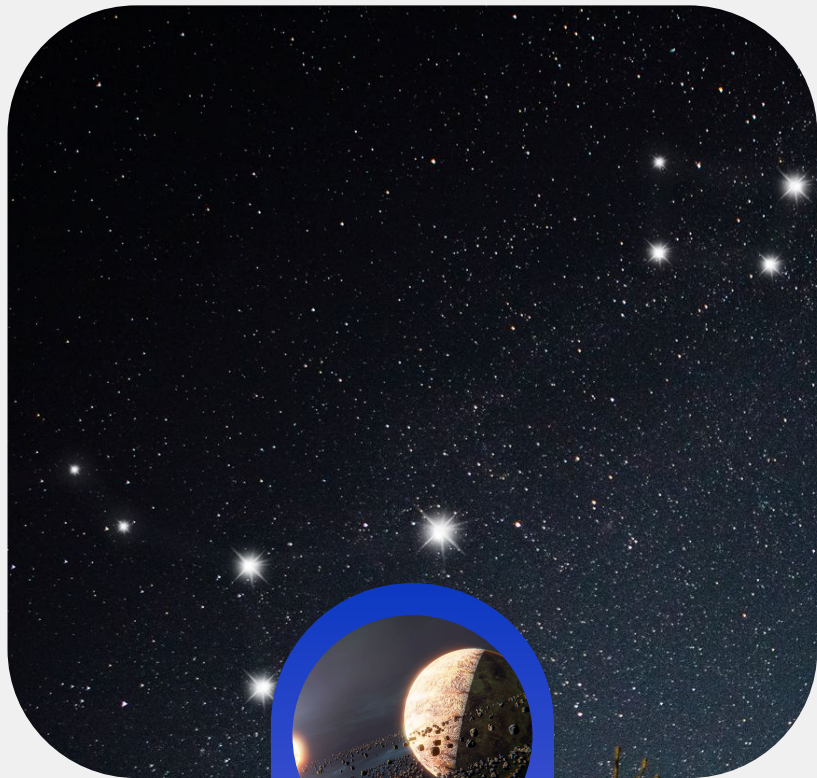
Reducción del tiempo de entrenamiento

Extracción de descriptores topológicos

Identificación y caracterización de la señal de "chirp"

Aplicación de métodos de interpretabilidad de modelos

Evaluación de otros modelos más '*tradicionales*' (XGBoost)

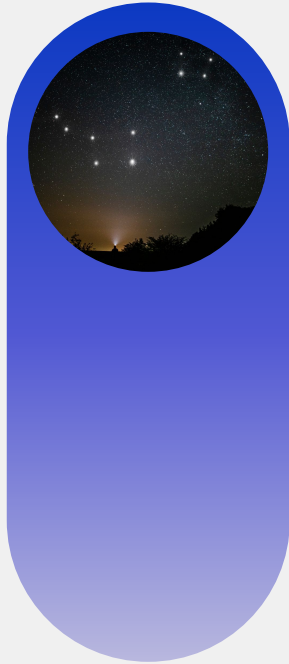


02

Metodología

Exploración, preparación e implementación

Datos



Giotto-tda

**Valores
de R**

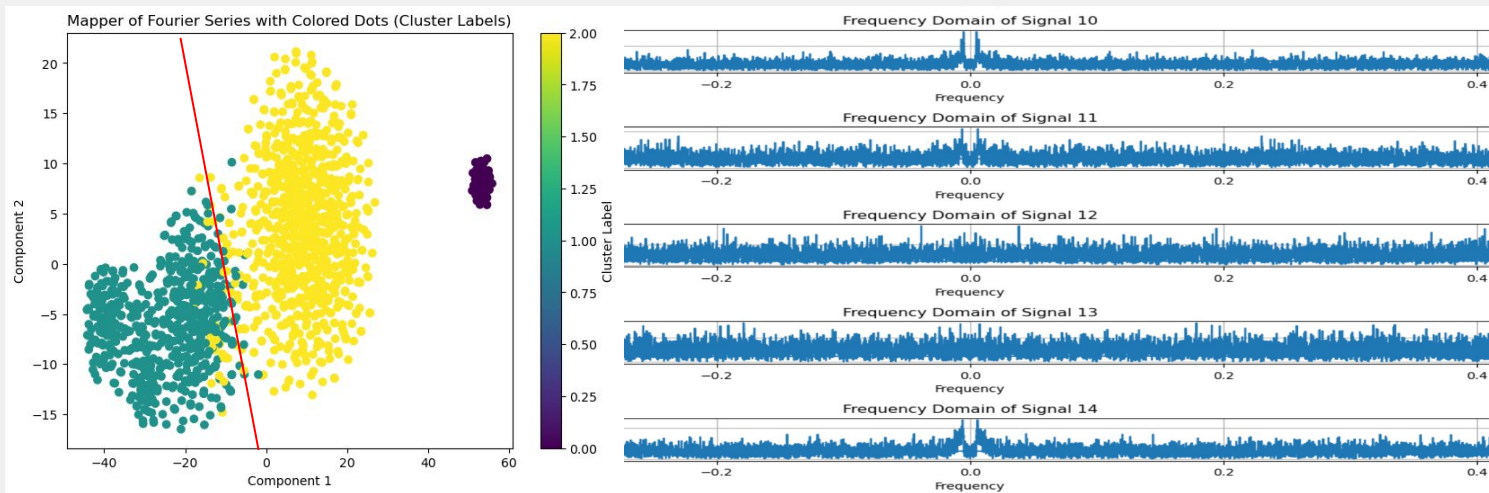
Para generar los datos, se utilizó el repositorio que se encuentra en la documentación de la librería de `giotto-tda` y se fue modificando el valor del parámetro `R` (0.075, 0.65) (que controla la relación del ruido-señal). Y se fue guardando las señales generadas y sus etiquetas.

Se probaron valores entre 0.075 a 0.65, con **1500 señales** cada uno. Se probaron valores de `R` = [0.065, 0.15, 0.2, 0.3, 0.4, 0.45, 0.5, 0.55, 0.6]

Exploración

Las series de Fourier de las señales logran dividir correctamente en 3 clusters las señales (transformando las magnitudes usando PCA en 50 componentes). Dos clústeres son de ondas y uno es de ruido.

*A partir de $R=0.15$ las series de Fourier dejan de ser suficientes para diferenciar ondas de ruido (accuracy < 60 %)

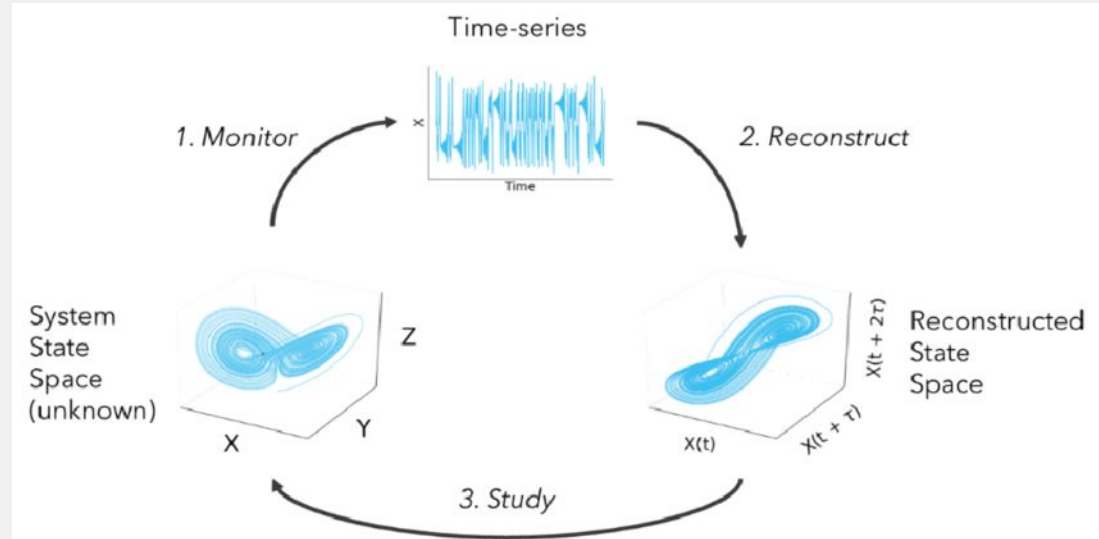


Accuracy: 93.20%

Takens Embedding

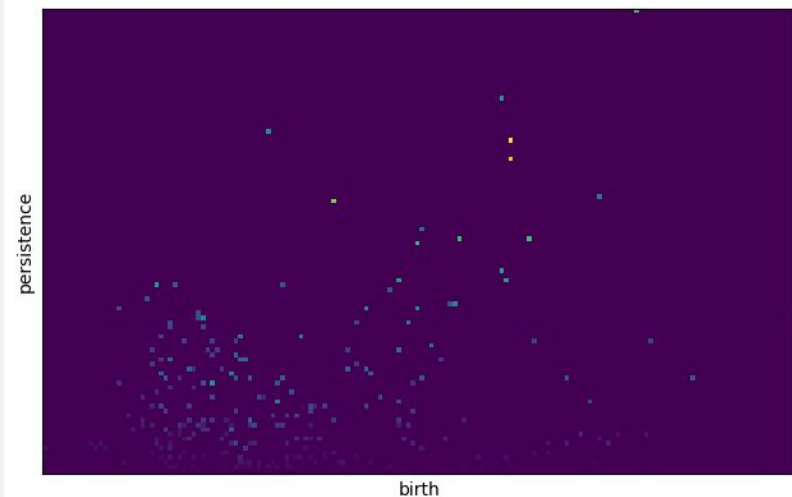
Envolver la serie de tiempo

A partir de pruebas en la serie de tiempo, se encontró que todas las señales tienen como parámetros óptimos (delay= 2, dimensiones =4)



Imágenes de Persistencia

A partir de la generación del diagrama de persistencia es posible generar un mapa con la distribución de densidad de la homología persistente en hoyos de dimensión 2.



03

Resultados



Modelos de Machine Learning

Con 50 componentes PCA de Fourier (accuracy)

Datos de entrenamiento

Modelo	R = 0.65	R = 0.5	R = 0.4	R = 0.3	R = 0.25	R = 0.2	R = 0.15	R = 0.065
RandomForest	0.97	0.96	0.96	0.84	0.7	0.6	0.54	0.48
SVM	0.95	0.91	0.95	0.85	0.77	0.69	0.56	0.5
XGBoost 1	0.97	0.96	0.95	0.84	0.72	0.62	0.55	0.5
Logistic Regression	0.95	0.87	0.94	0.85	0.74	0.66	0.55	0.5
Neural Network (MLP)	0.94	0.86	0.94	0.84	0.75	0.7	0.55	0.5

De todas las pruebas de modelos y datos de entrenamiento, se concluye que el SVM y el XGBoost entrenados con el dataset de $R = 0.4$ son la mejor opción para clasificar ondas solamente con el PCA de Fourier como input. (trained: 1200 observaciones, 80% de 1500)

*A partir de $R \leq 0.15$ las series de Fourier dejan de ser suficientes para diferenciar ondas de ruido (accuracy < 55 %)

Modelos de Machine Learning

Con 50 componentes PCA de Fourier (recall)

Datos de entrenamiento

Modelo	R = 0.65	R = 0.5	R = 0.4	R = 0.3	R = 0.25	R = 0.2	R = 0.15	R = 0.065
RandomForest	0.95	0.94	0.91	0.93	0.96	0.95	0.96	0.89
SVM	0.85	0.85	0.90	0.84	0.85	0.95	0.85	0.85
XGBoost 1	0.94	0.94	0.92	0.92	0.94	0.93	0.93	0.86
Logistic Regression	0.93	0.93	0.91	0.92	0.91	0.77	0.9	0.85
Neural Network (MLP)	0.91	0.9	0.93	0.9	0.9	0.89	0.85	0.8

Mejor Modelo con Fourier

Al analizar el recall, observamos que la mayoría de los modelos pueden identificar al menos el 85 % de las ondas, independientemente del nivel de ruido presente.

Predicción con diferentes SNR

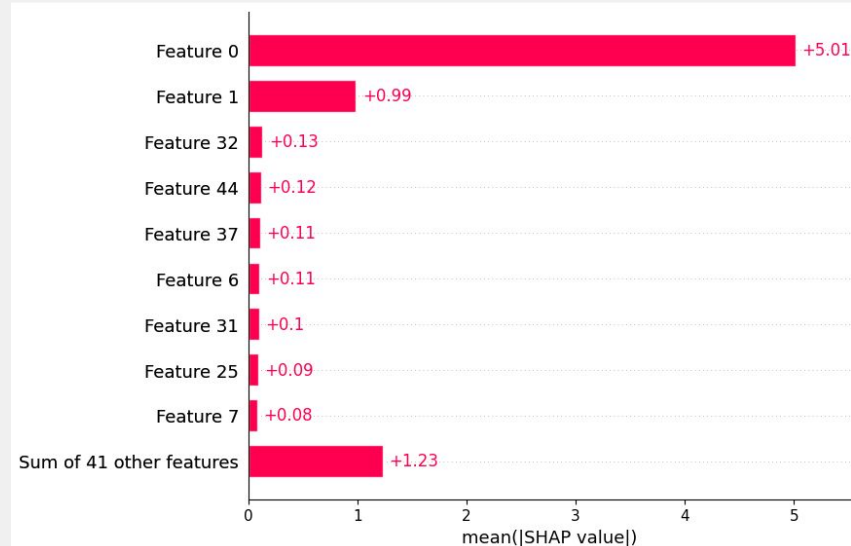
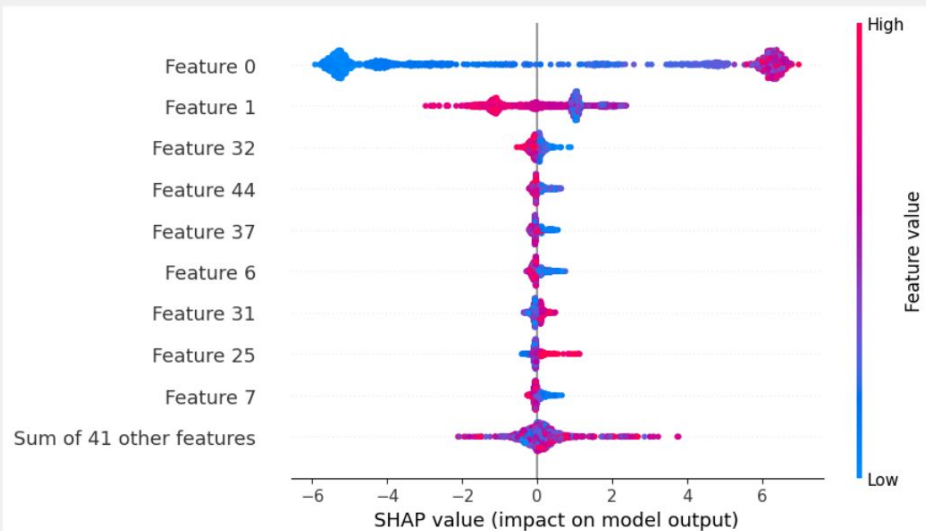


XGBoost 1

XGBoost	10 diferentes R de 0.15 a 0.65	500 diferentes R de 0.05 a 0.65	100 diferentes R de 0.05 a 0.25
Recall	0.81	0.77	0.58
Accuracy	0.85	0.82	0.60
Tiempo de entrenamiento	10s	12s	10s

Interpretabilidad de Modelo

XGBoost



Con 50 componentes PCA de Fourier

Versión 1 de la CNN

CNN 1	R = 0.65	R = 0.5	R = 0.25
Recall	0.80	0.7	0.5
Accuracy	0.85	0.73	0.49
Tiempo de entrenamiento	2m 16s	2m 03s	22s

Se entrenó con las imágenes de persistencia de las señales. (1200 observaciones)

Comparación de Modelos

CNN 1

Accuracy: 85%

Recall: 80%

Tiempo de Ejecución: 2m 16s

Este fue hecho usando imágenes de persistencia de las señales

XGBoost 2

Accuracy: 75%

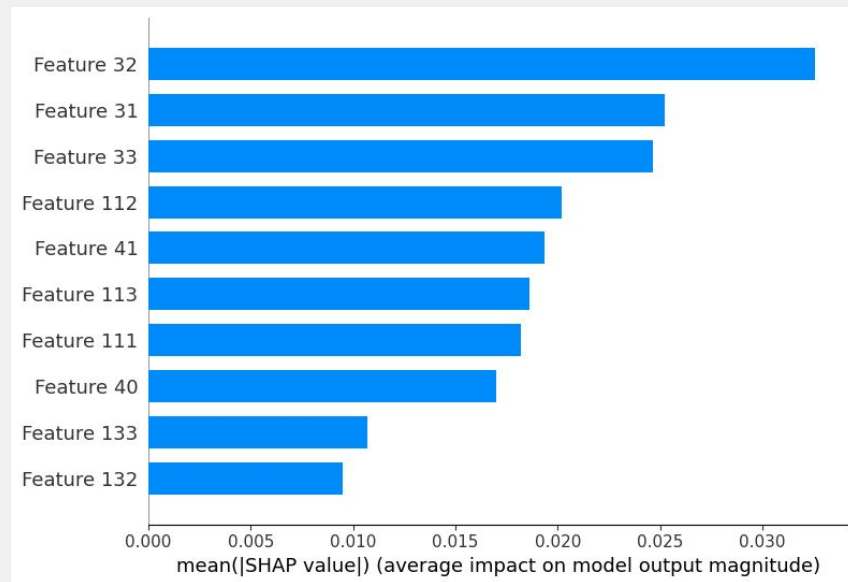
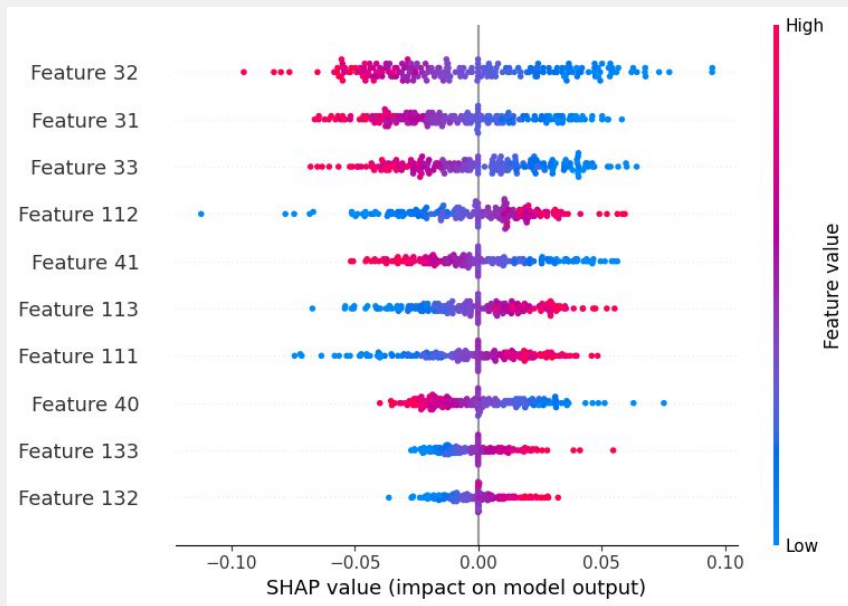
Recall: 72%

Tiempo de Ejecución: 2 s

Este fue hecho usando imágenes de persistencia de las señales

Interpretabilidad de Modelo

CNN



Con imágenes de persistencia

Versión 2 de la CNN

CNN 2	R = 0.65	R = 0.45	R = 0.25
Recall	0.90	0.8	0.72
Accuracy	0.92	0.81	0.69
Tiempo de entrenamiento	4m 12s	6m 03s	10m 12s

Se entrenó con las imágenes de persistencia de las series de Fourier. (1200 observaciones)



O4



Conclusiones

El análisis de las señales gravitacionales resulta un proceso muy complejo que requiere diversos campos de estudio para poder tomar decisiones a partir de las lecturas con ruido en el sistema, por lo que, el uso de la transformada discreta de Fourier junto con otros estudios matemáticos de las series de tiempo logra ser una metodología eficiente para identificar la naturaleza de una señal.

¡Muchas gracias!

¿Preguntas o Comentarios?

