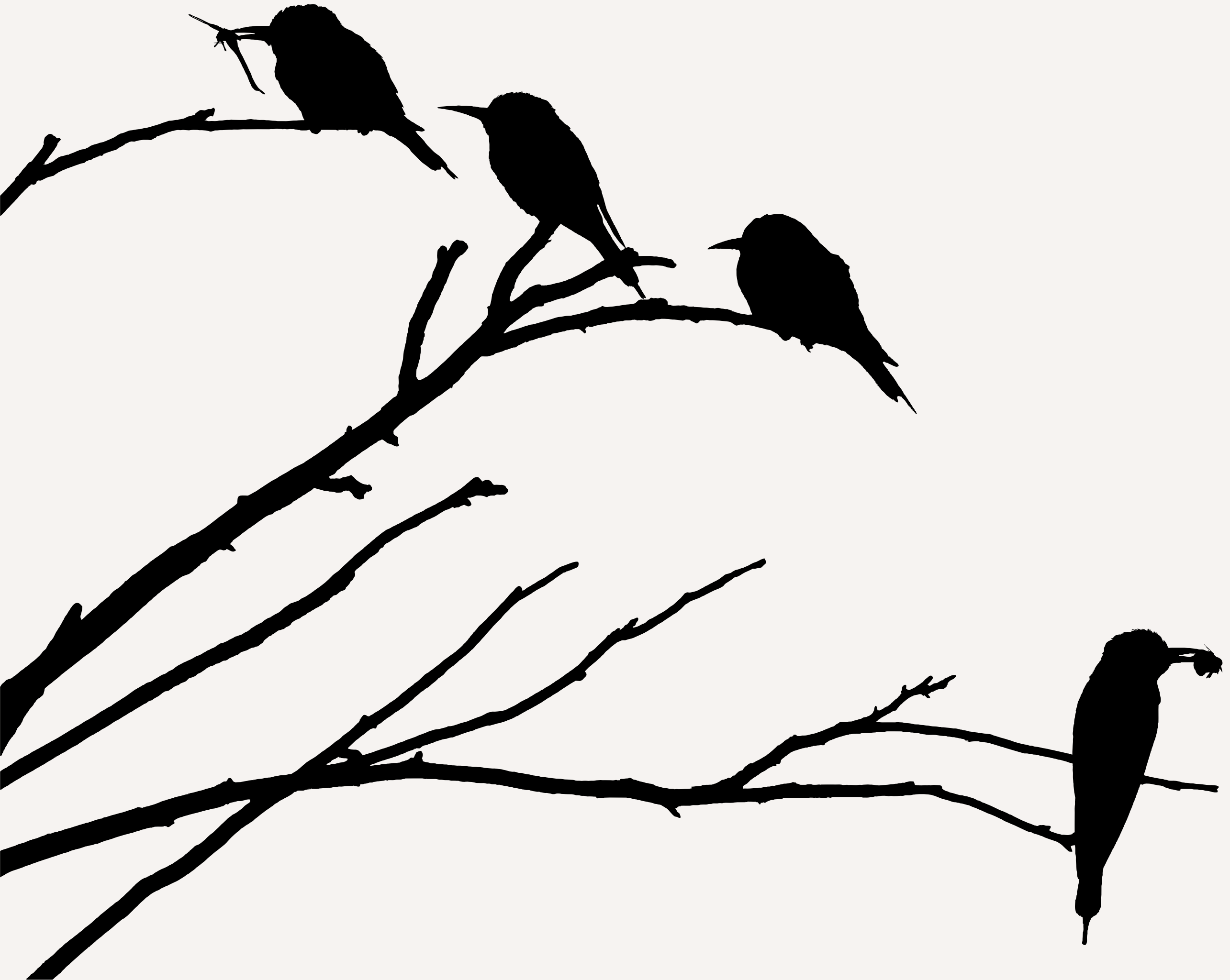
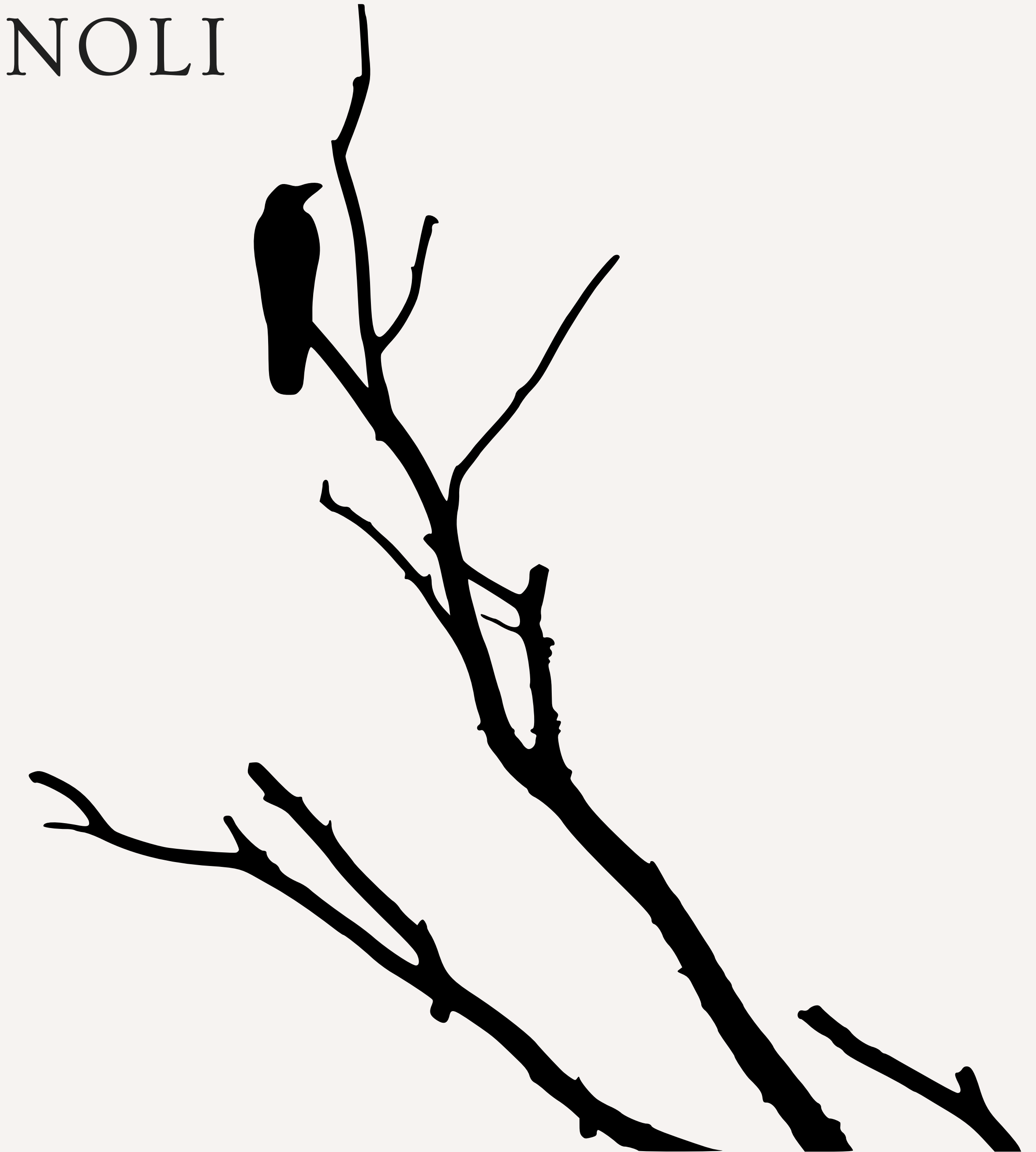
Localización de fuentes sonoras mediante computo de diferencias temporales de arribo

Laboratorio 6 y 7-2021-Departamento de Física ,FCEN,UBA

CONSTANZA FERRER - FELIPE CIGNOLI





Director: Dr. Gabriel Mindlin

Co-director: Ing. Roberto Bistel

Índice

Planteo del problema

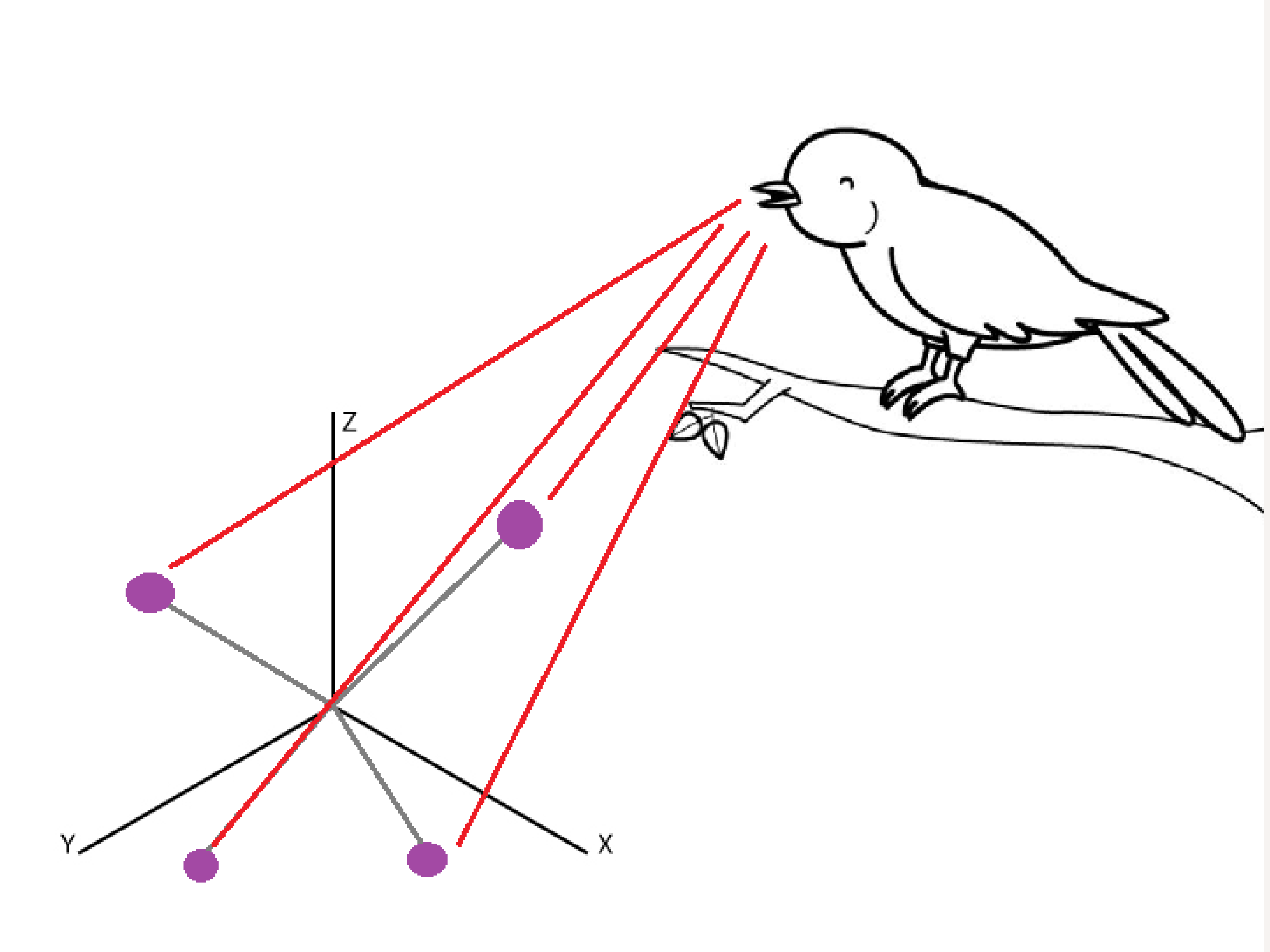
$$t_i = \frac{|X_i - p|}{c} + t_0$$

c Velocidad del sonido

t_0 Instante inicial

p Posición del ave

$$t_i - t_j = \frac{|X_i - p|}{c} - \frac{|X_j - p|}{c}$$



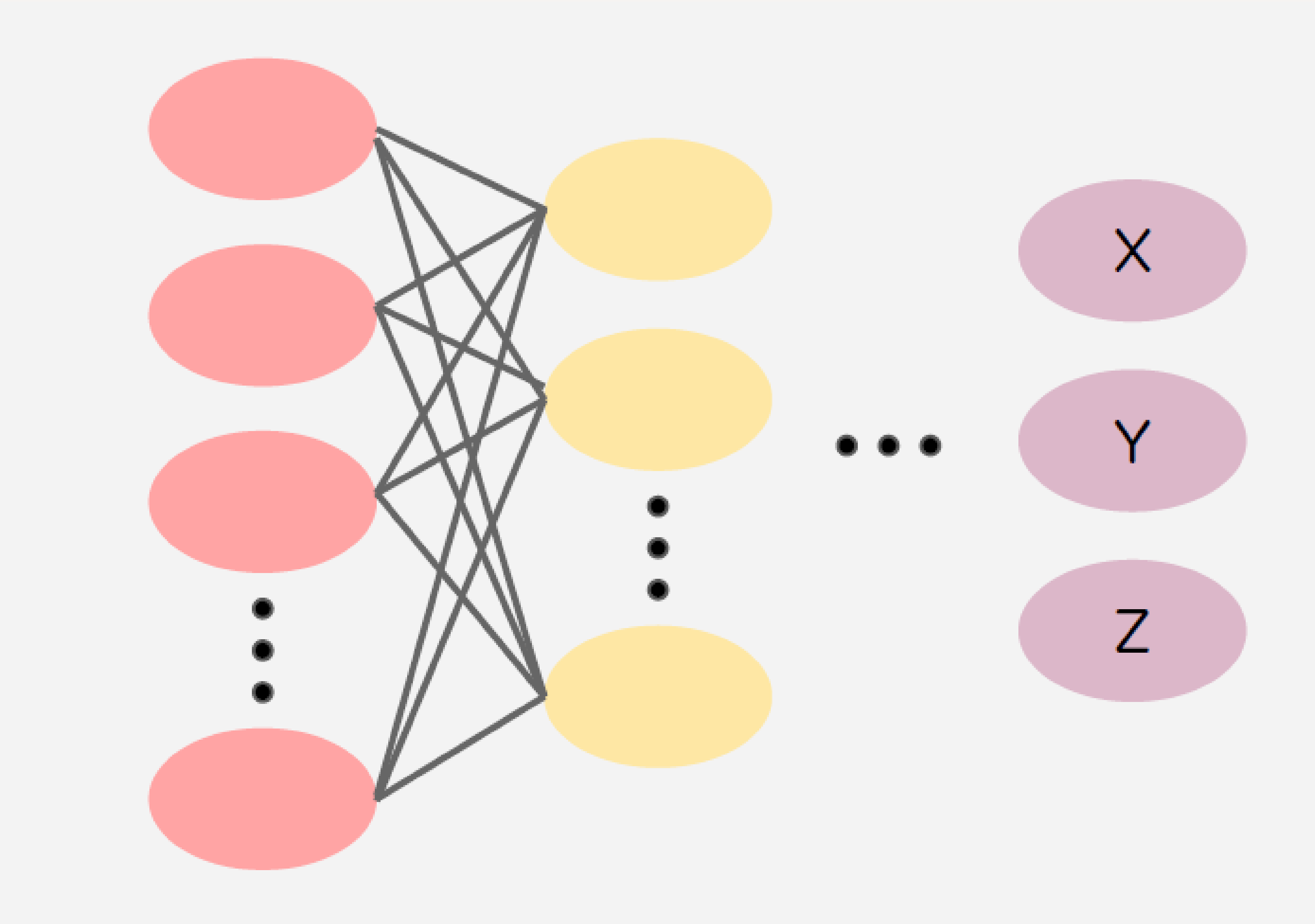
Objetivo General

Diseñar un algoritmo para realizar una regresión capaz de traducir 6 retrasos relativos temporales en posiciones de fuentes sonoras, captadas por un grabador ZOOM H6 con 4 canales.

¿Qué es una Red Neuronal?

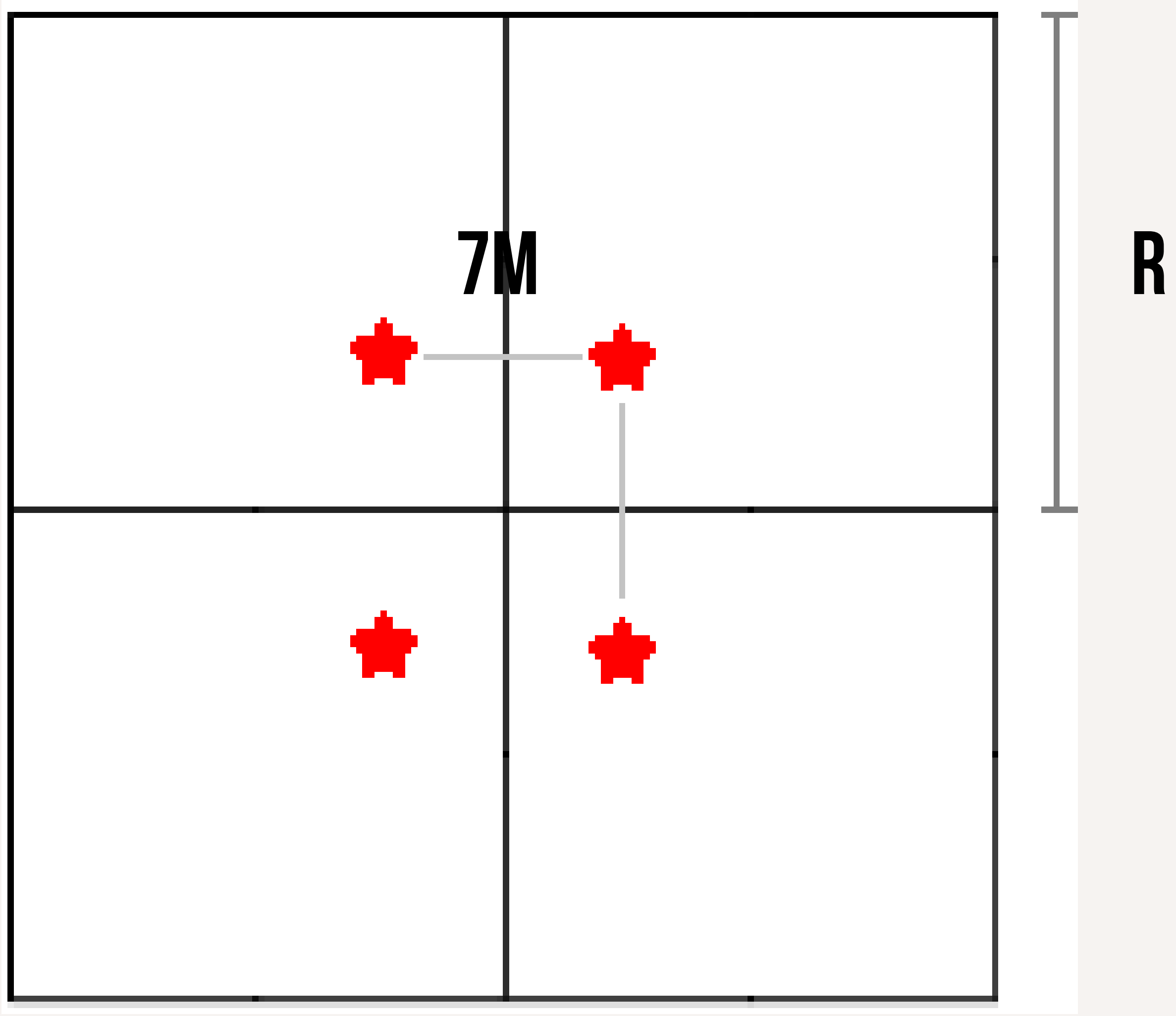
Modelo computacional que consiste en un conjunto de unidades (neuronas) conectadas entre sí para transmitirse señales. La información de entrada atraviesa la red (donde se somete a diversas operaciones) produciendo unos valores de salida.

Ingresamos los delays temporales



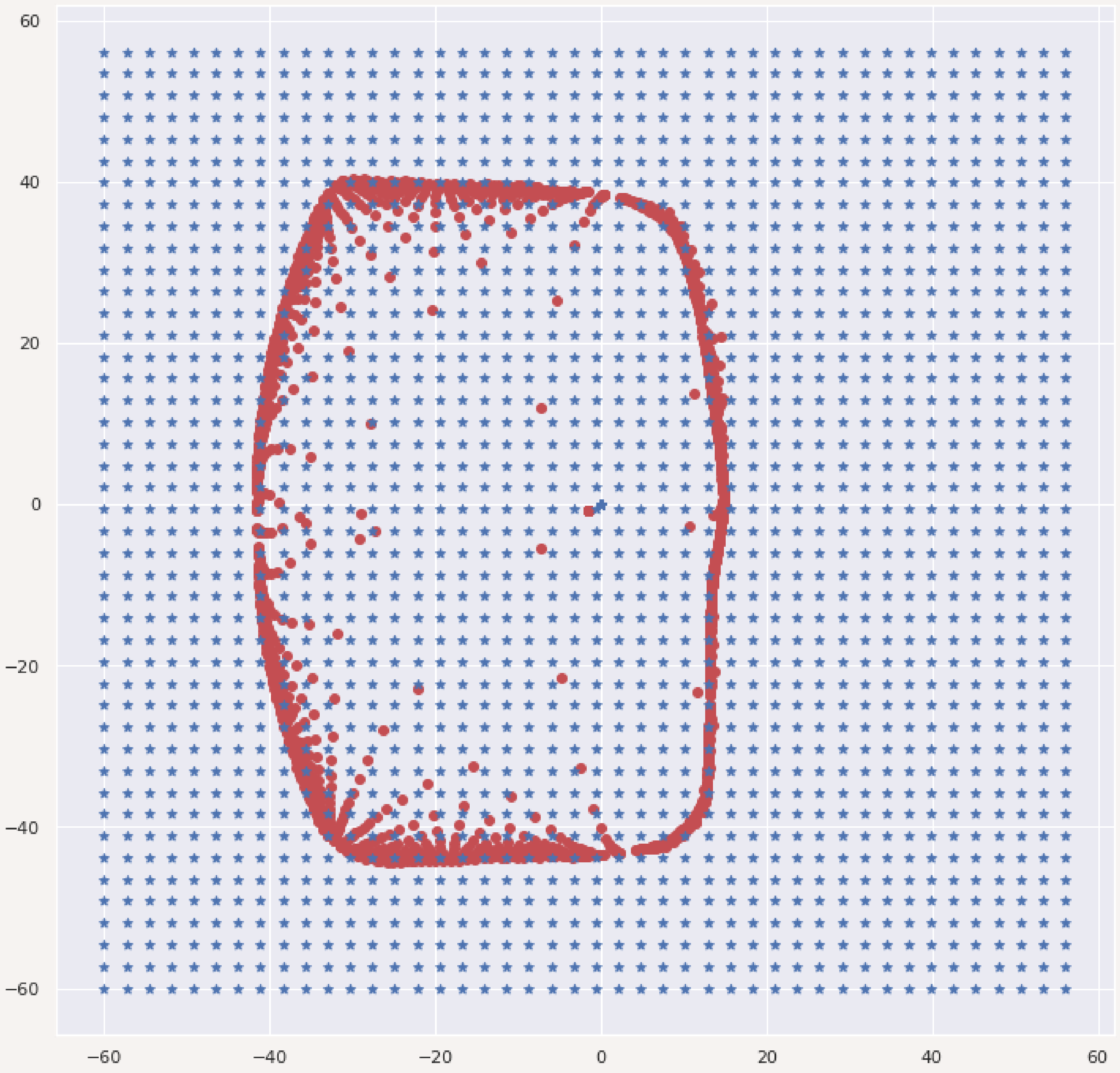
Buscamos recuperar las posiciones (x, y, z)

Punto de partida



Layers= 64-64

Épocas=100

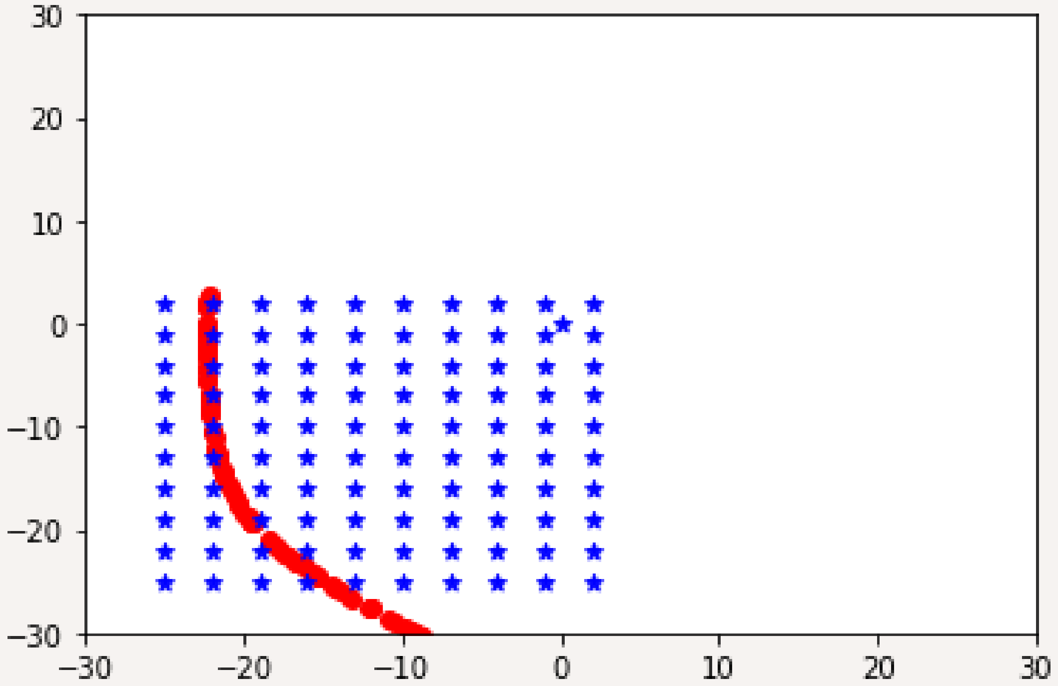


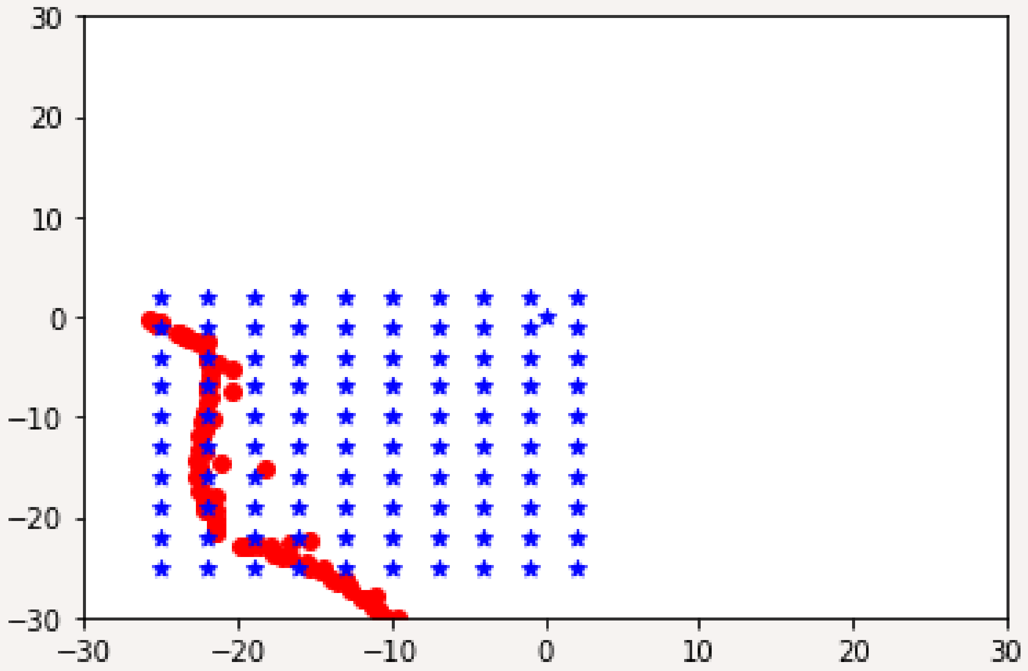
Nomenclatura para Redes

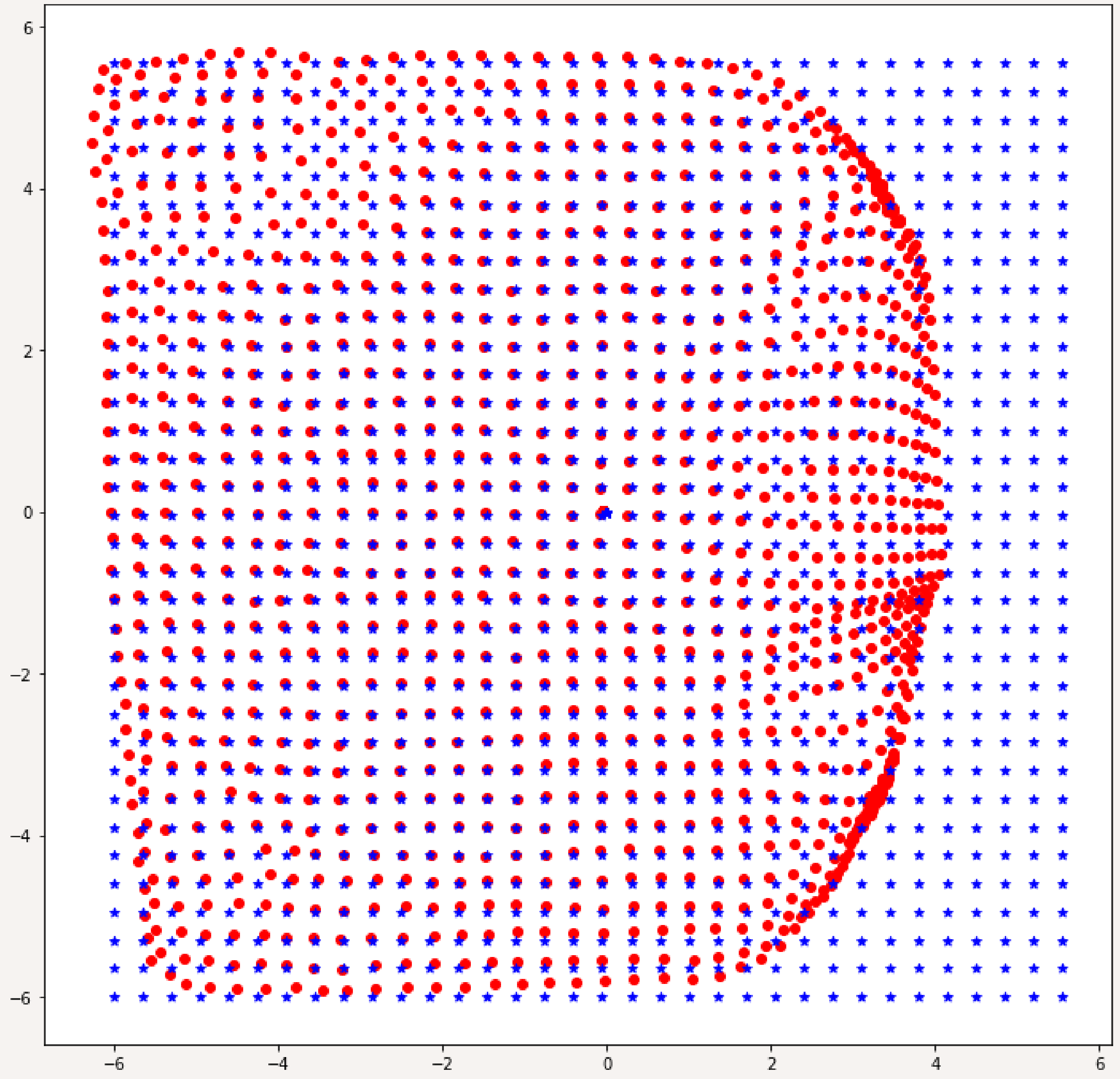
Épocas: épocas de entrenamiento del modelo

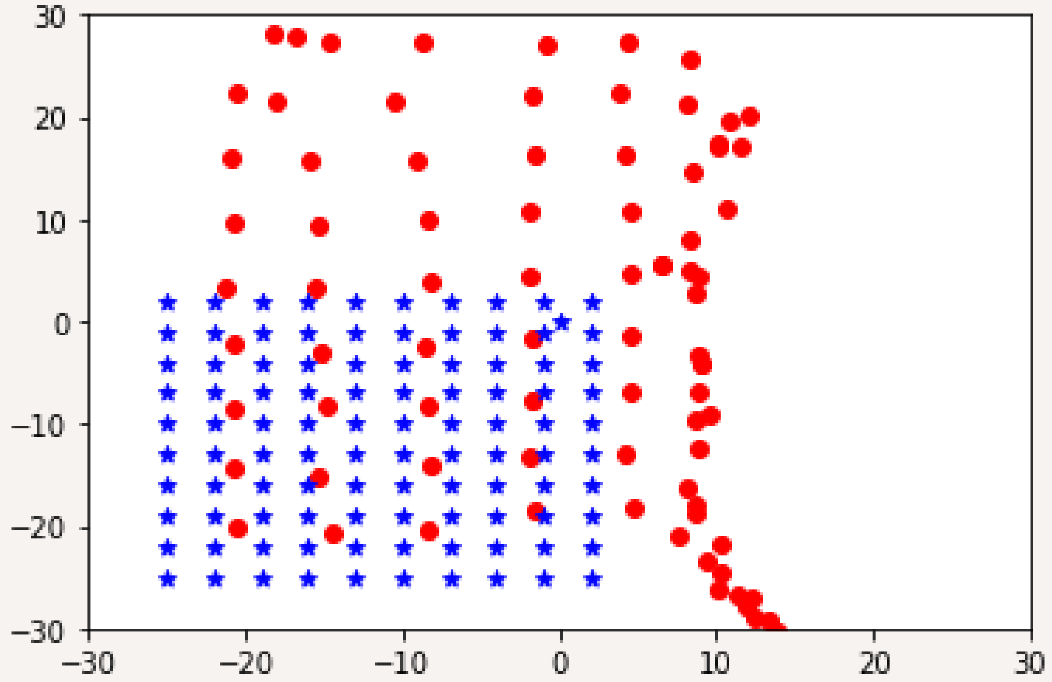
Layers: capas de la Red.

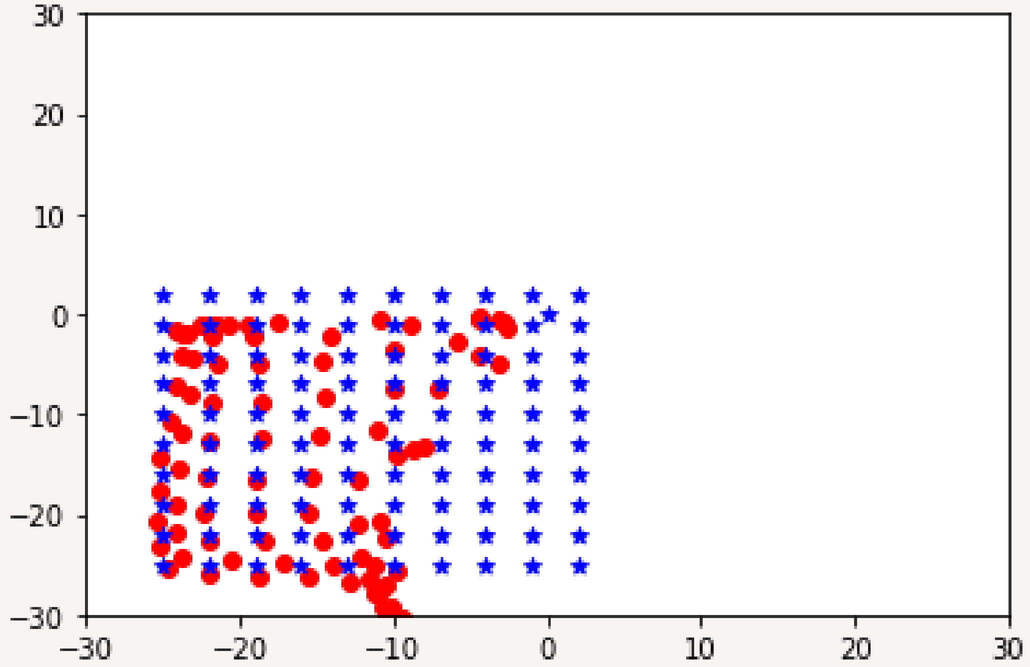
Comparación entre distintas arquitecturas de red







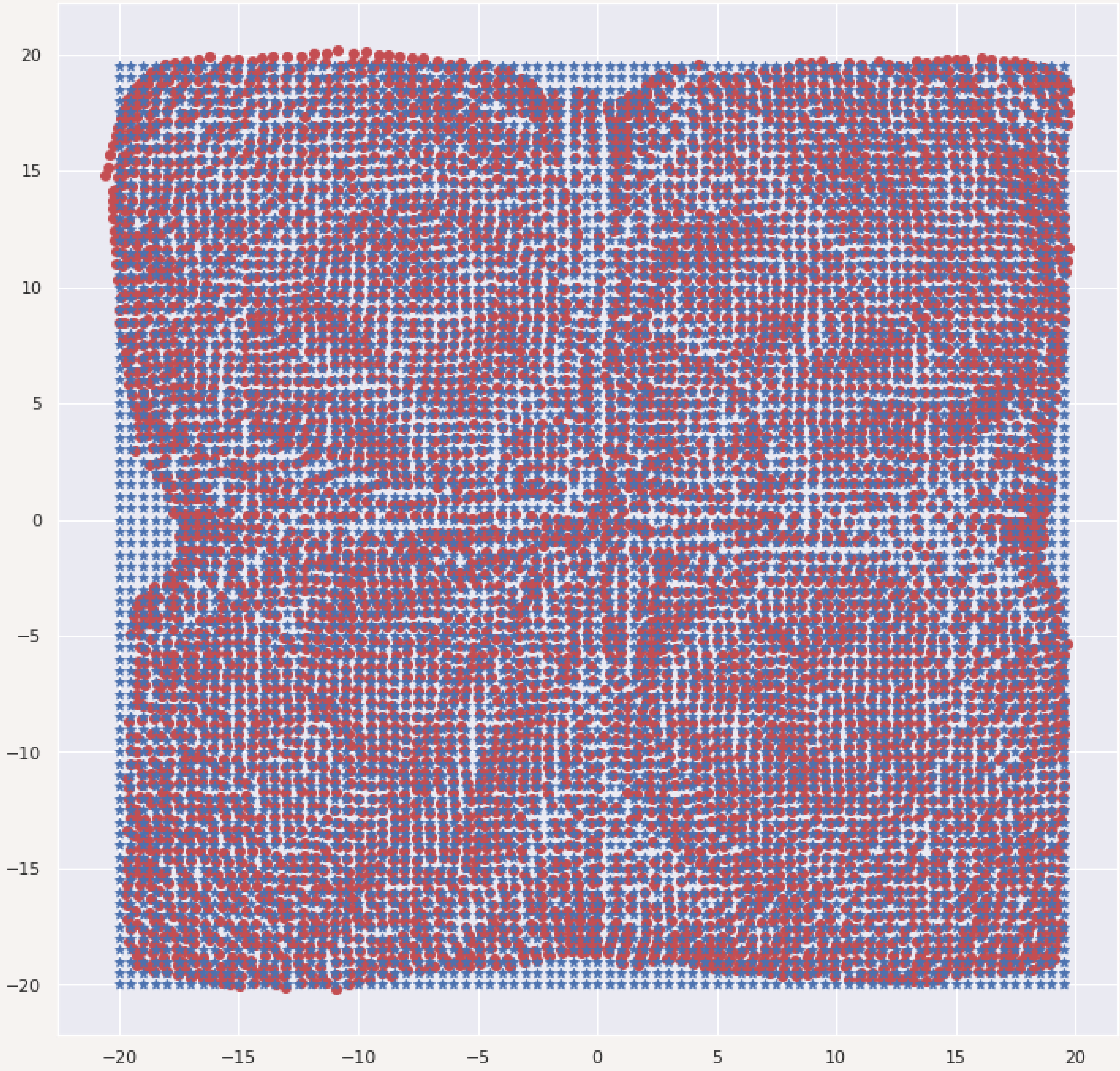




Layers= 64-128-128-64

Épocas=1000

Arquitectura seleccionada



Layers= 128-128-128-128

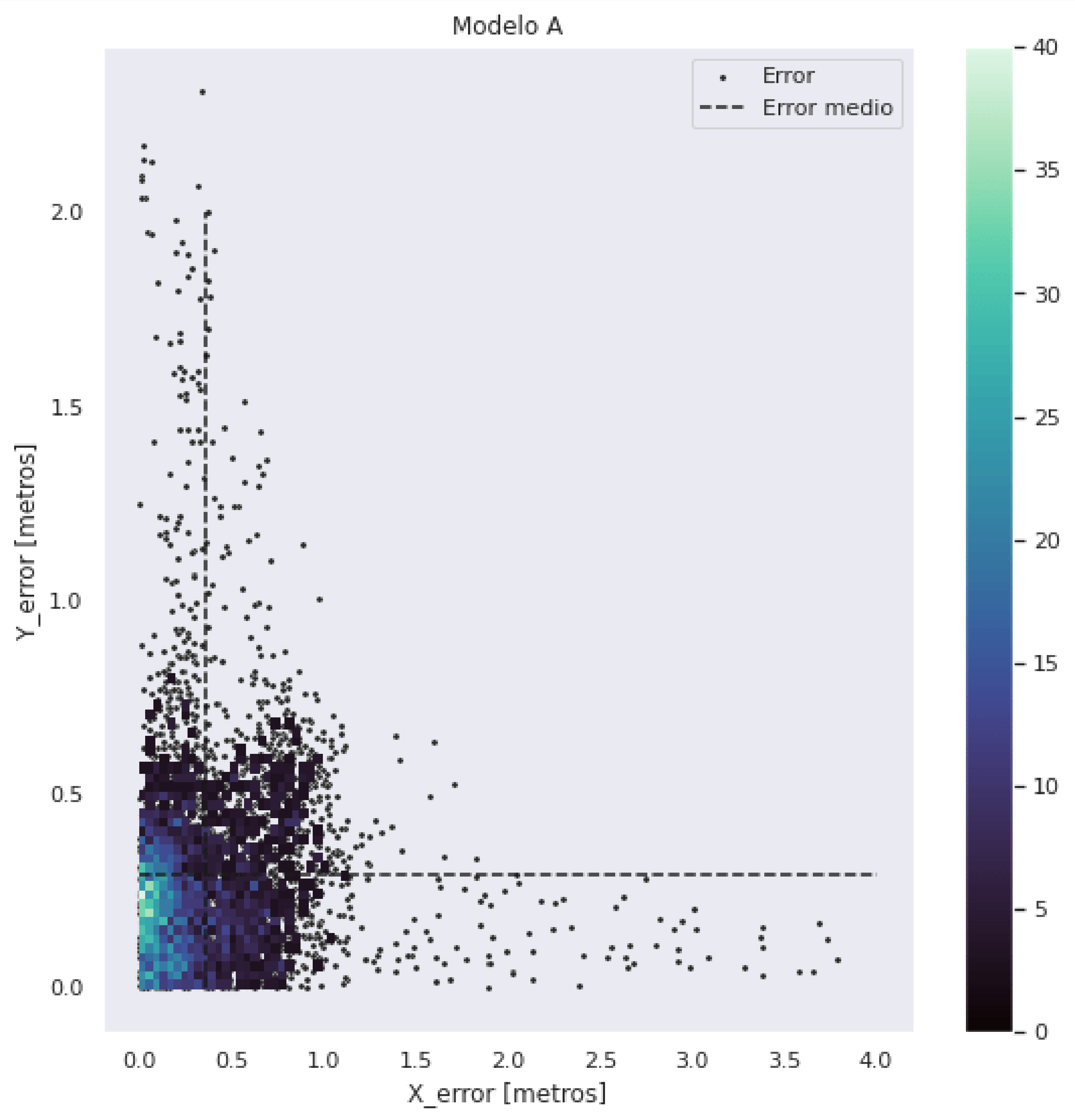
R= 20m Δ= 0.5m

k=3

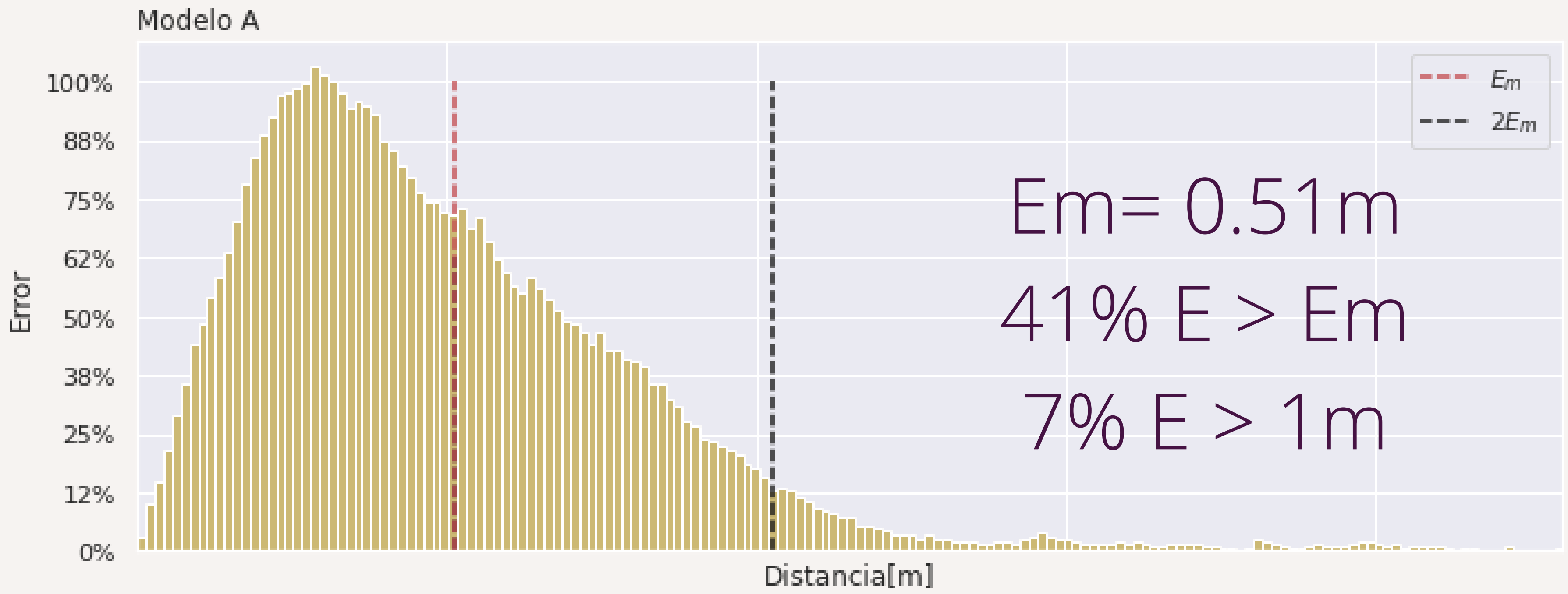
Epocas= 1200

Error = 0.52m

Análisis de Error



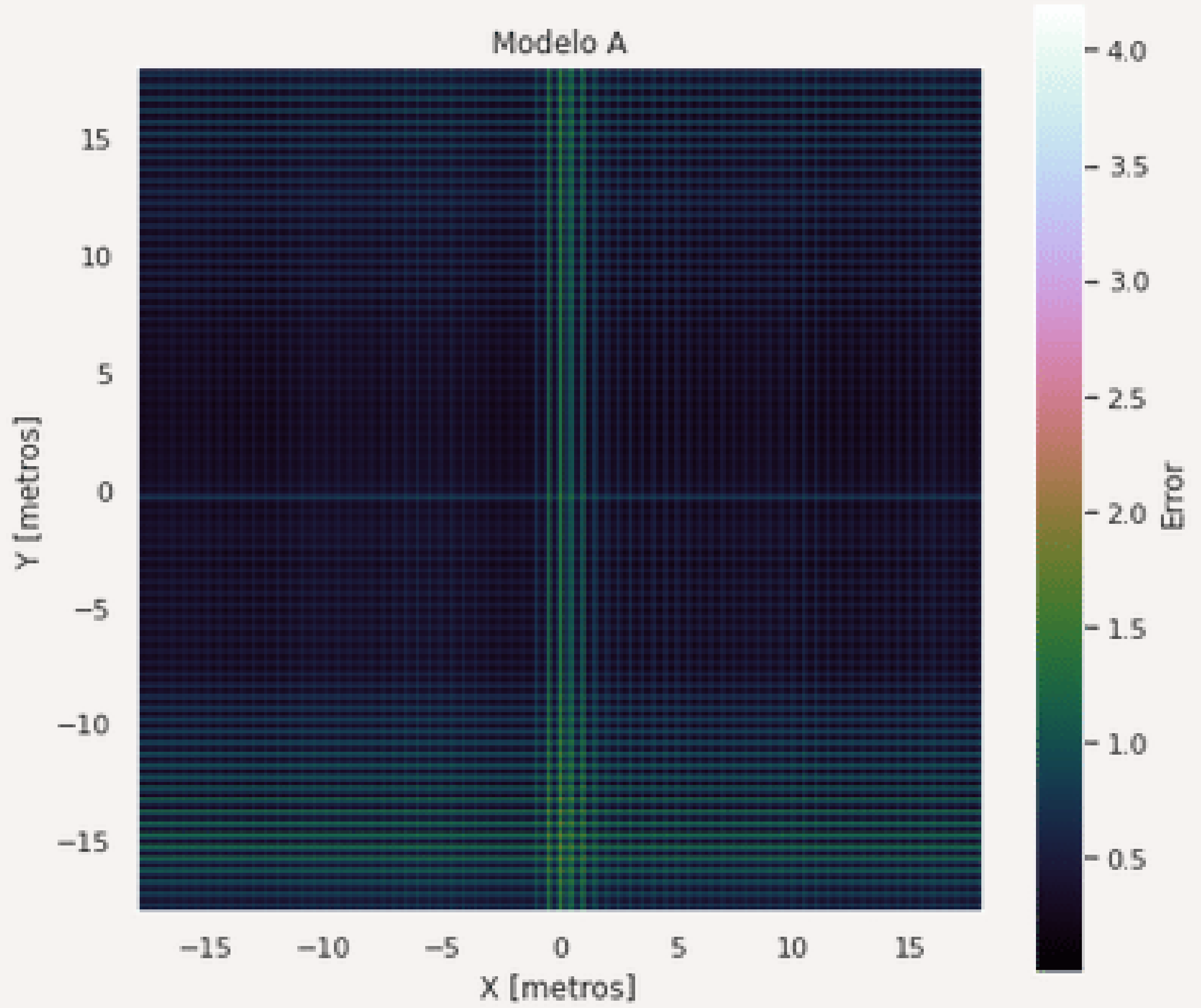
Histograma bidimensional de Error



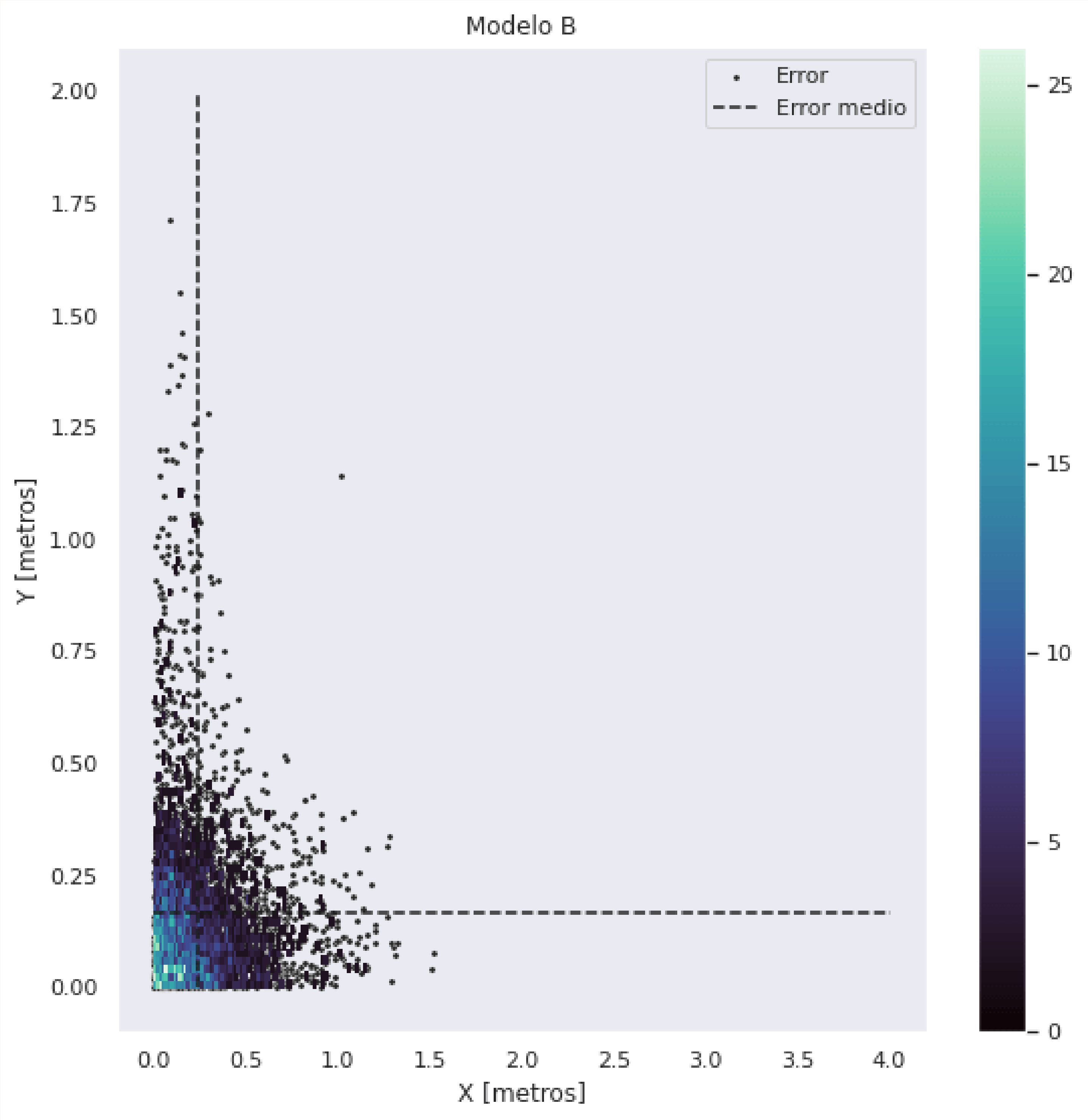
Em= 0.51m

41% E > Em

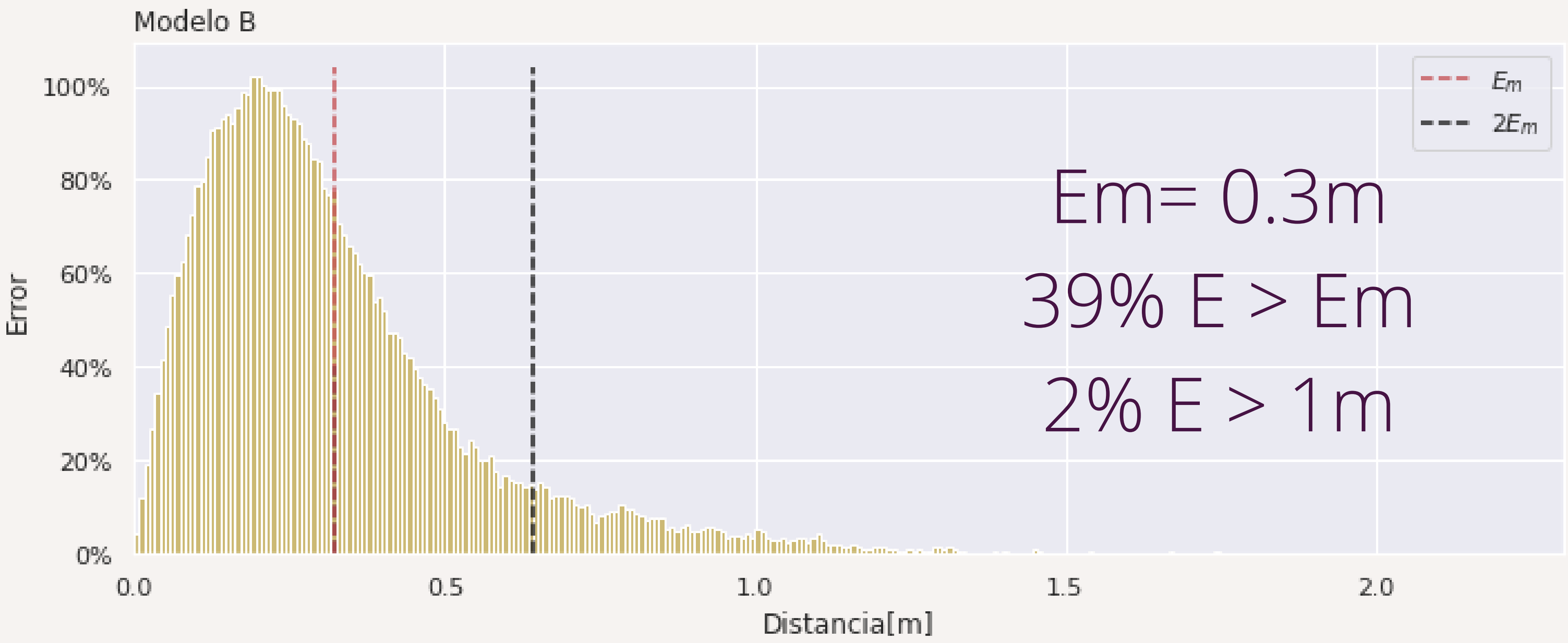
7% E > 1m

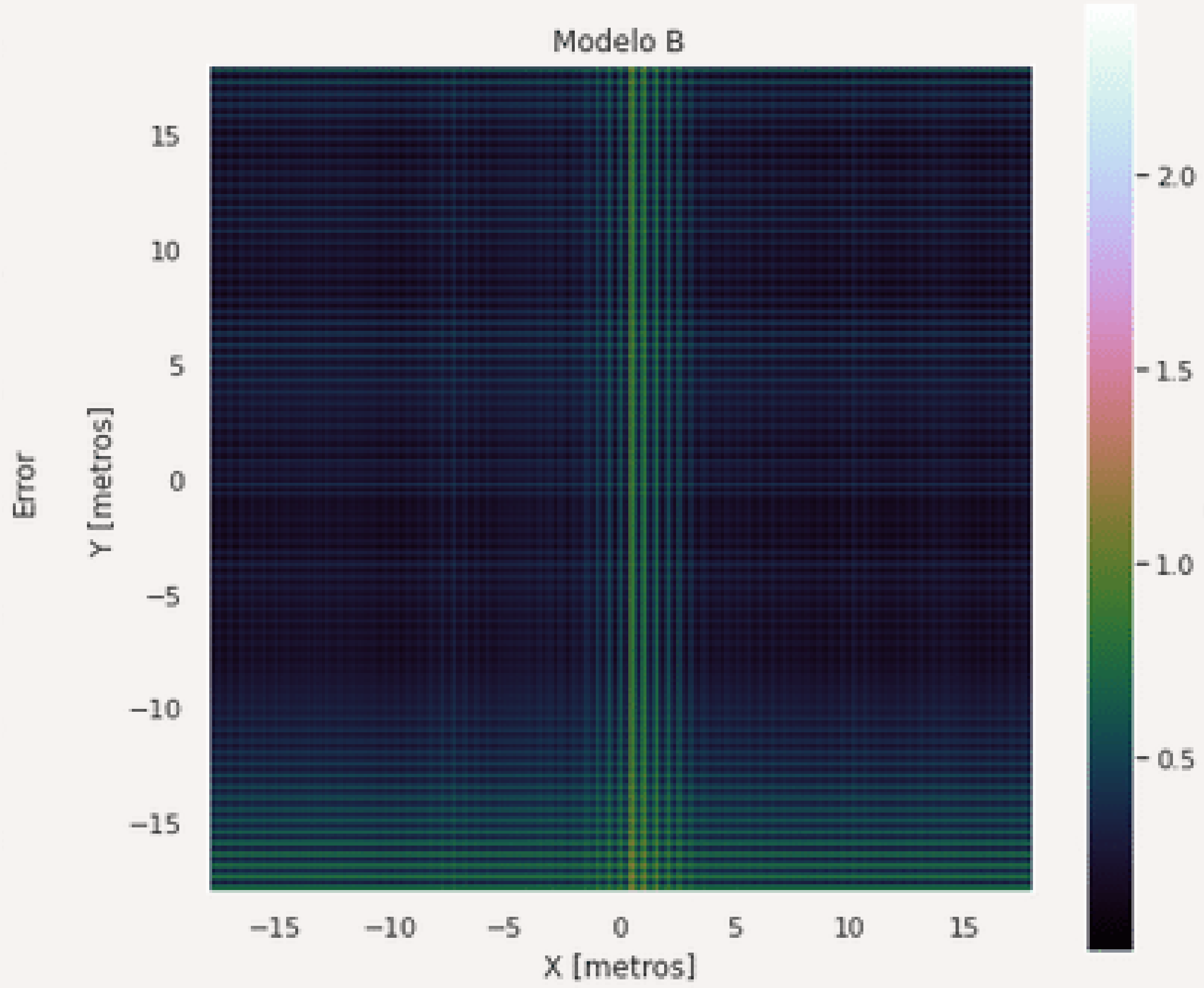


Análisis de Error

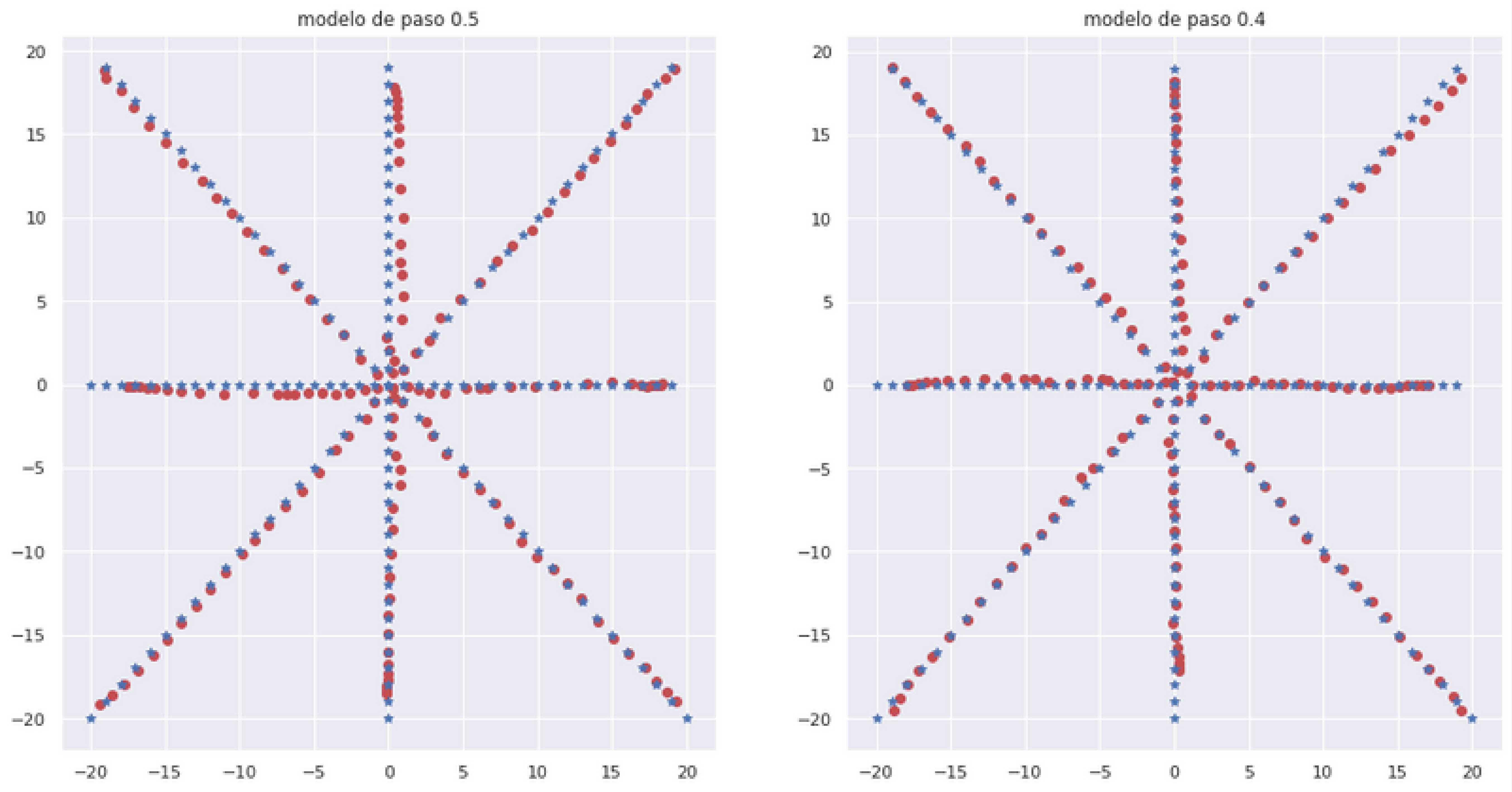


Histograma bidimensional de Error





¿Cómo predicen estos modelos?



R= 20m

k=3

ep=1200

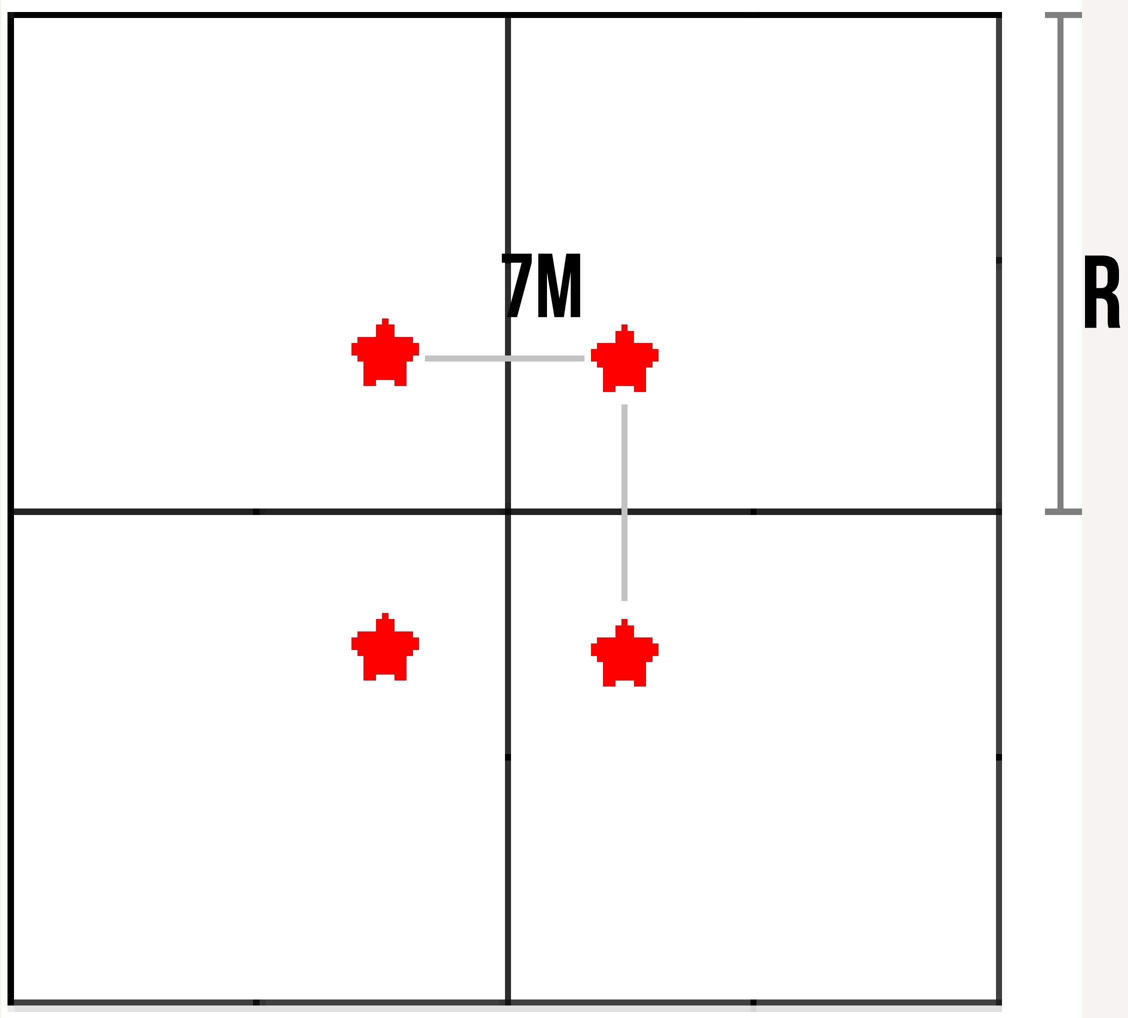
Δ= 0.5m

Modelo A

Δ= 0.4m

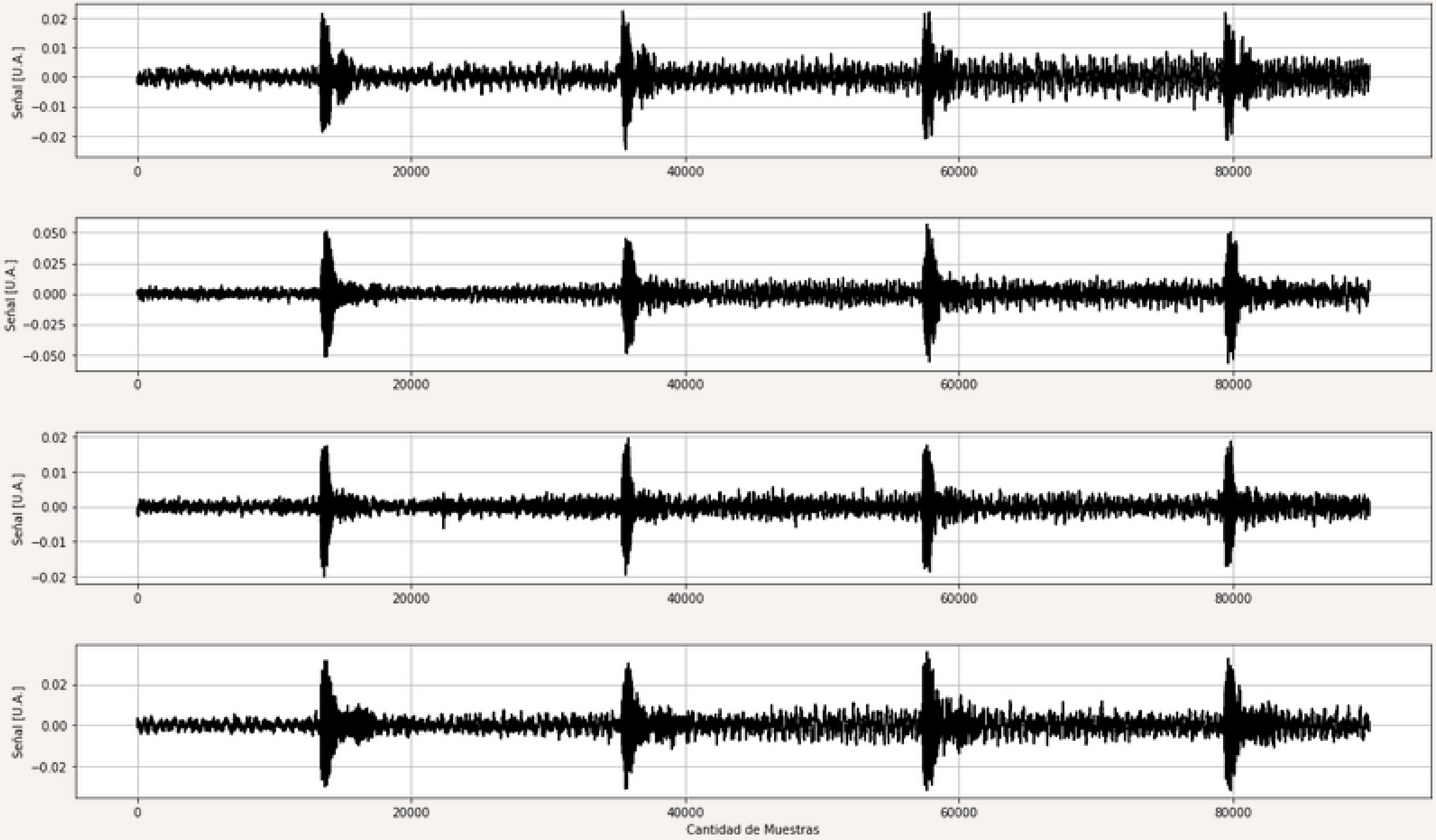
Modelo B


capturada en moto g⁶


7M
R

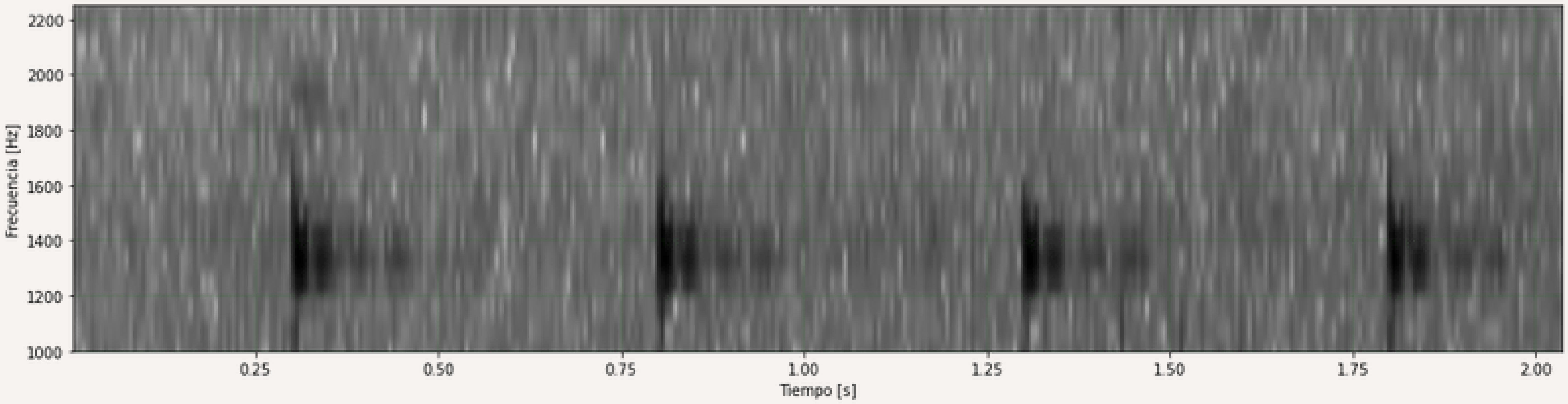
Procesamiento de Audio

Audios de metrónomo



$t = \ 4s$

Espectográma y frecuencias de corte

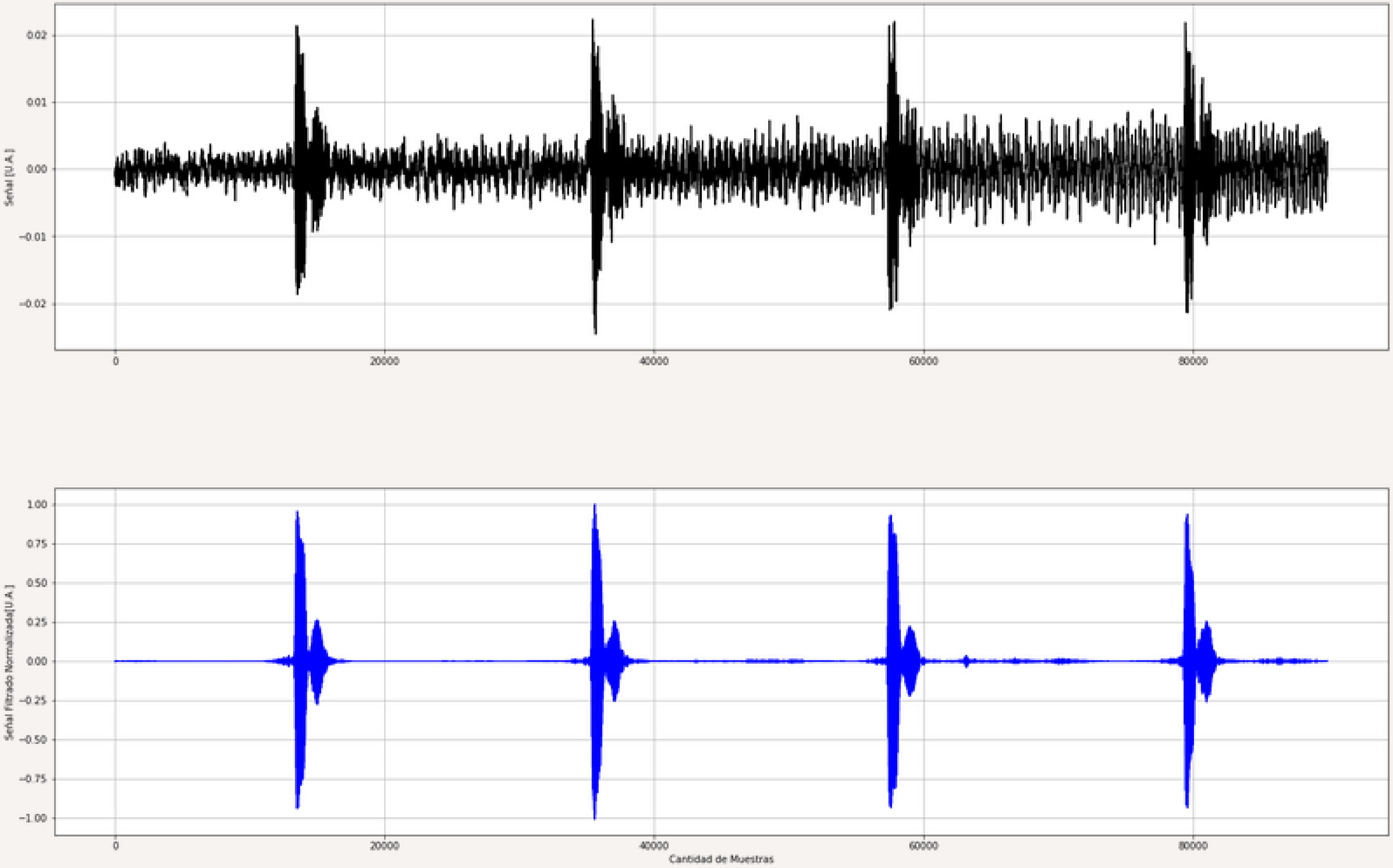


Espectográma correspondiente micrófono 1

Queremos aplicar un filtro pasa bajos, y pasa altos sobre la señal

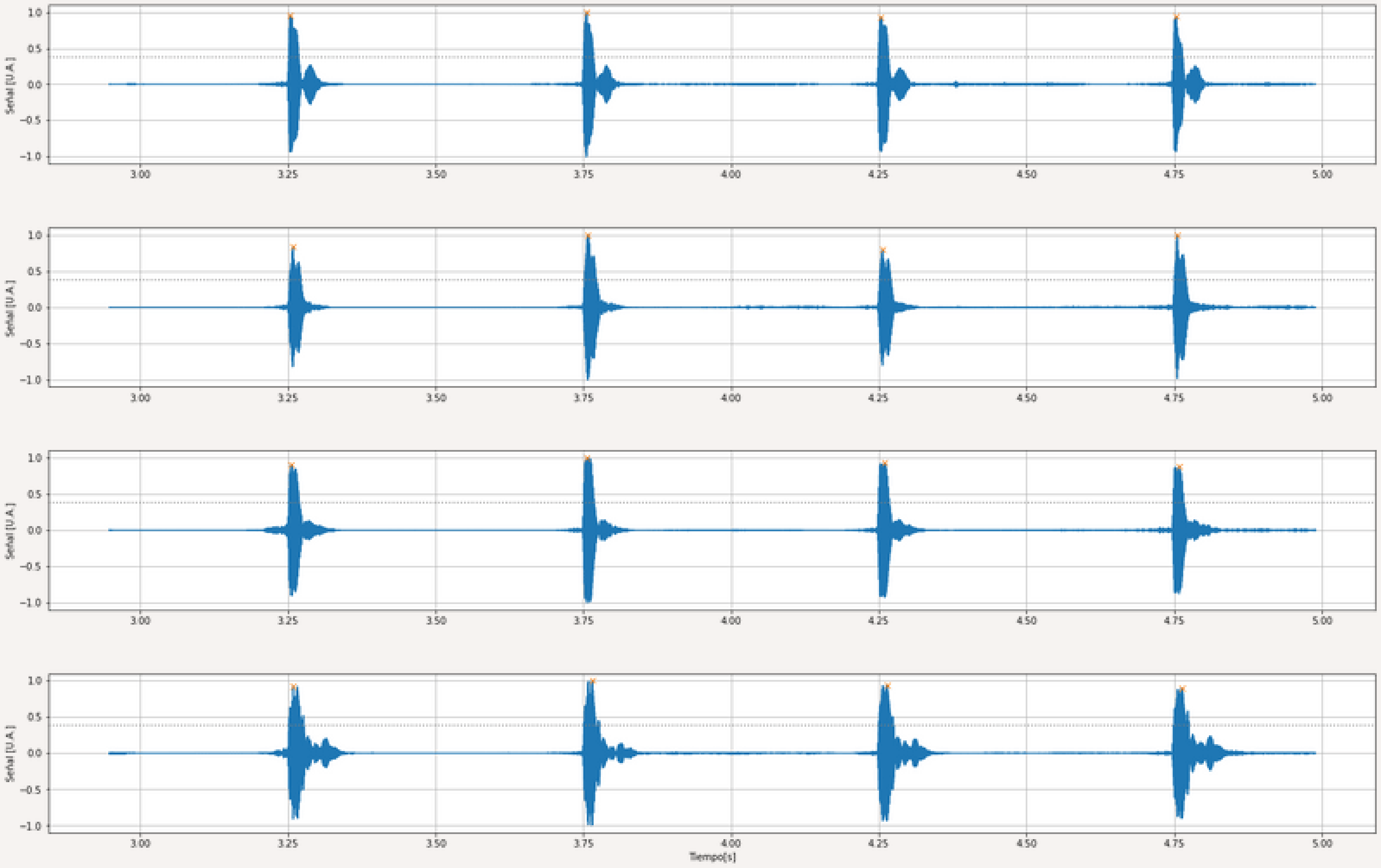
$$f_{PA} = 1000\ hz$$

$$f_{PB} = 2250\ hz$$



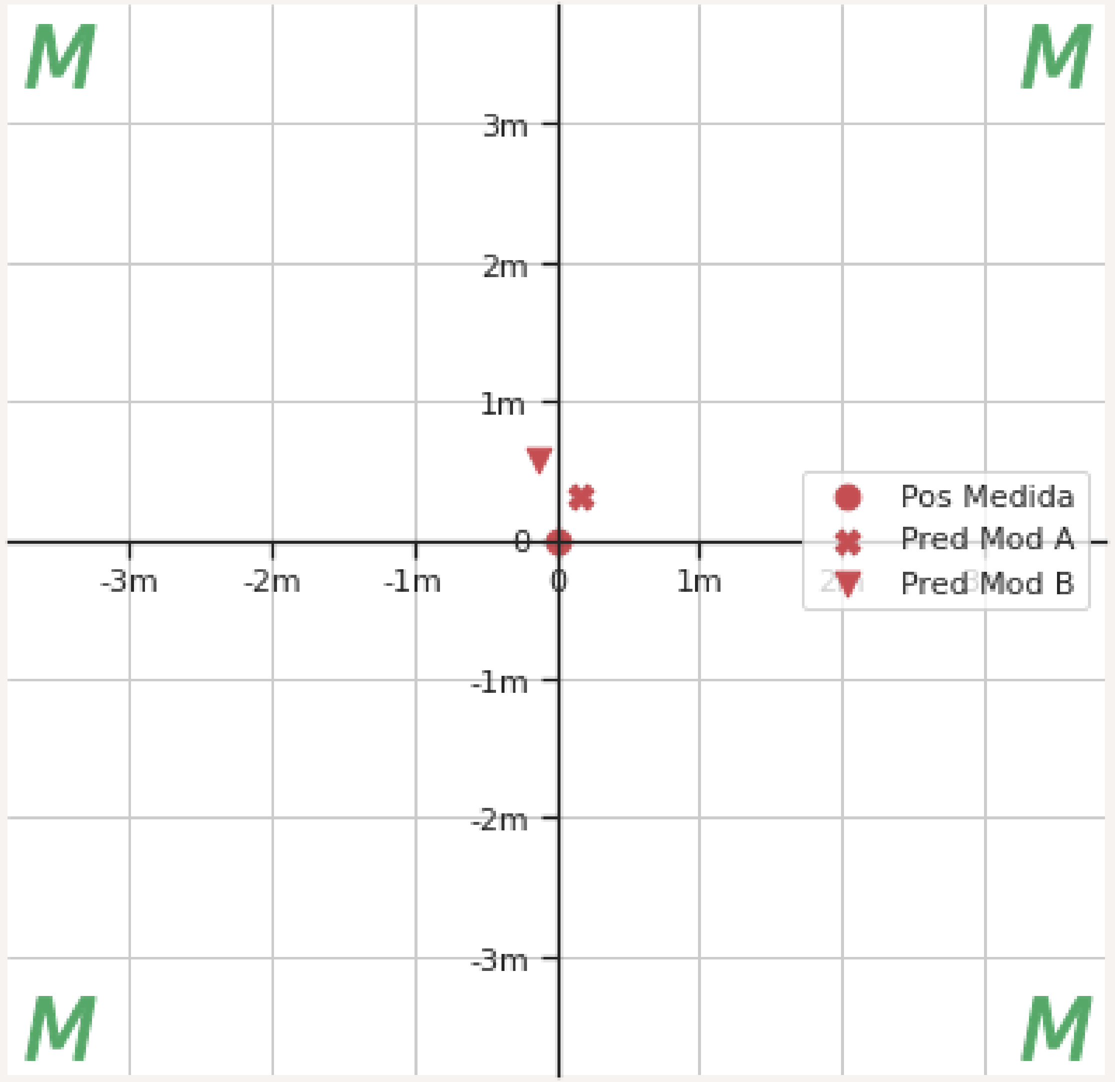
Señal [U.A.]
0.02
0.01
0.00
−0.01
−0.02
0
20000
40000
60000
80000
Señal Filtrado Normalizada[U.A.]
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
20000
40000
60000
80000
Cantidad de Muestras

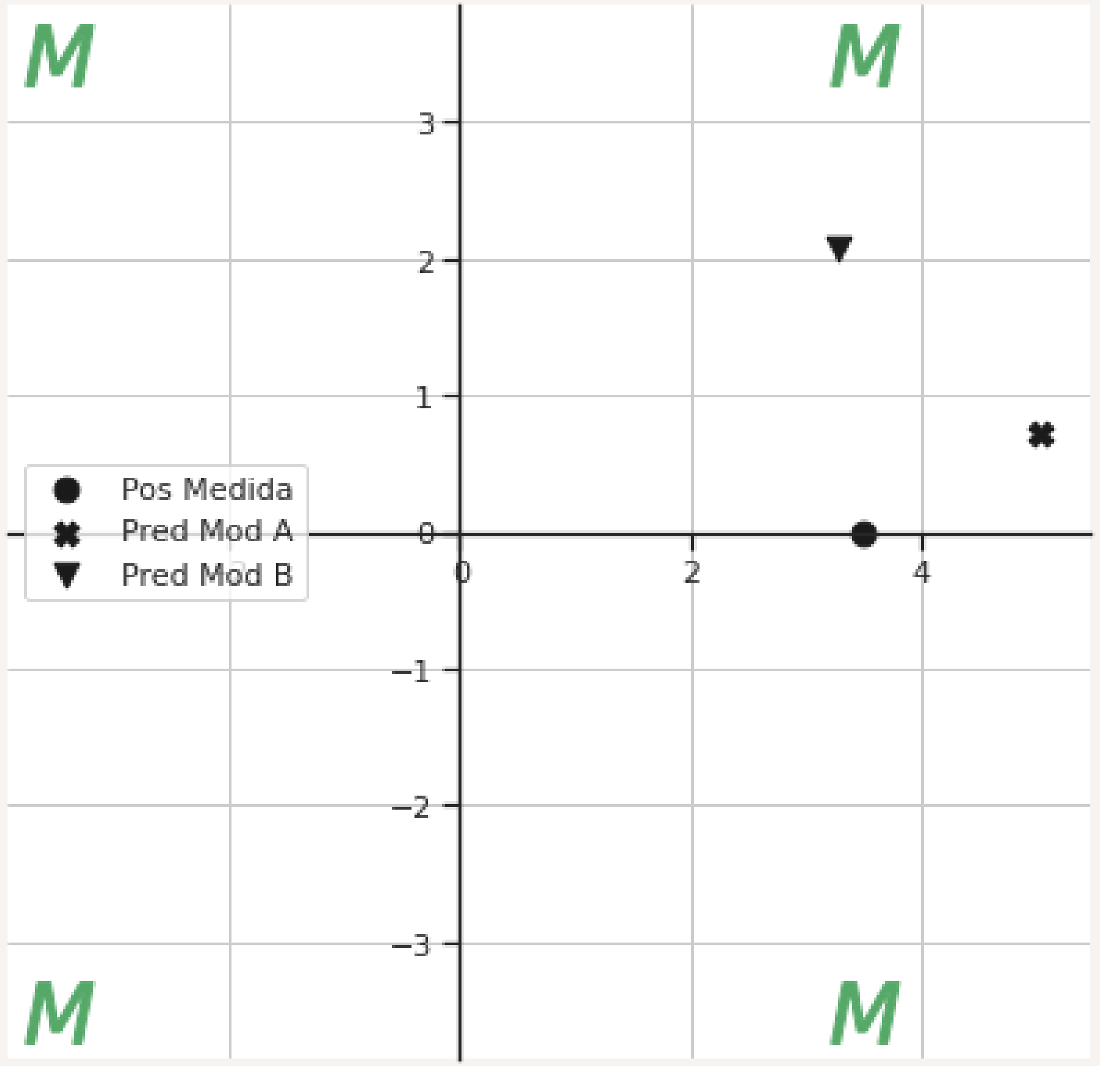
Picos en la señal filtrada

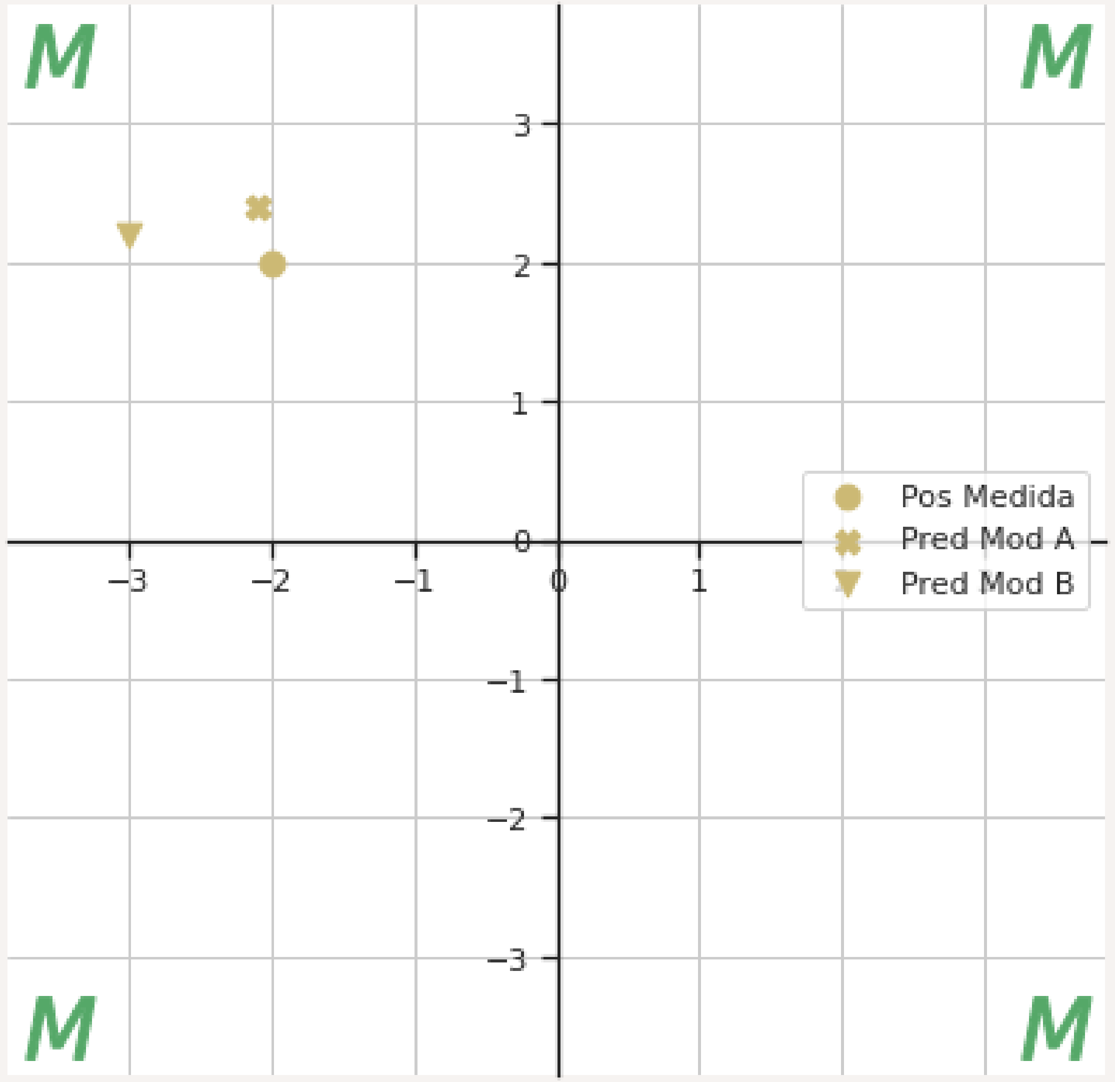


Sobre esta señal vamos a obtener los delays y aplicar nuestros modelos

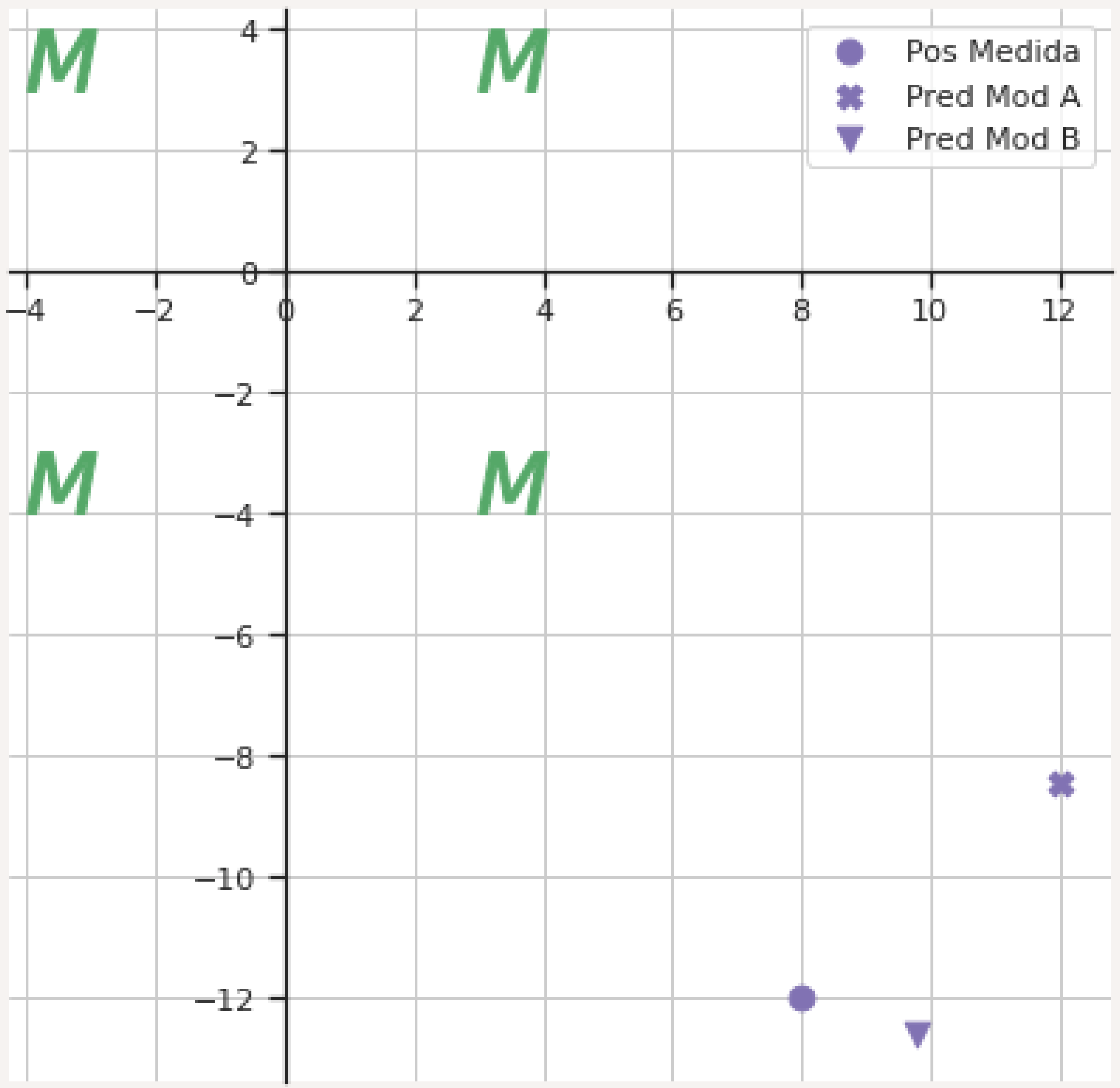
Algunos resultados

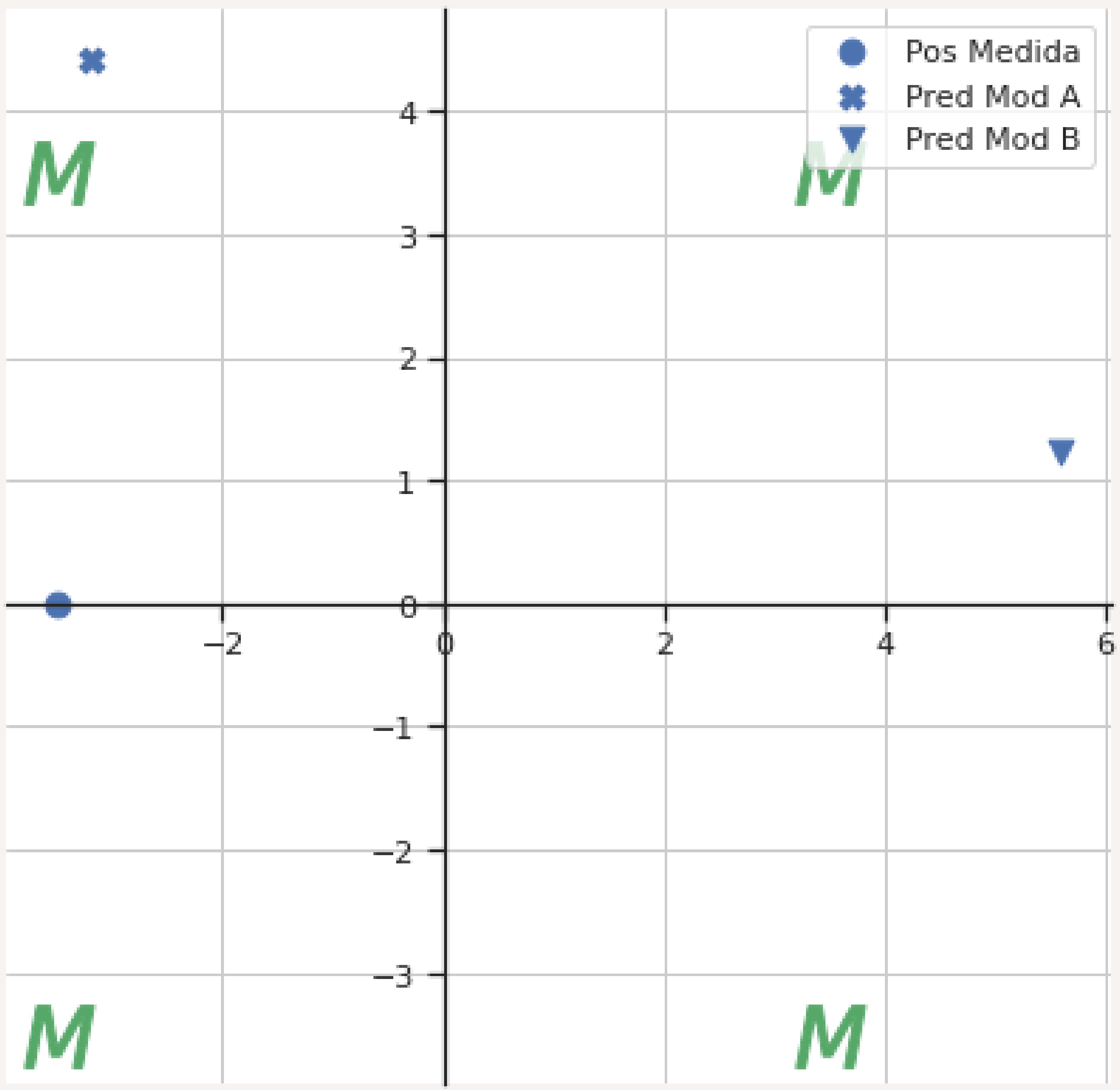




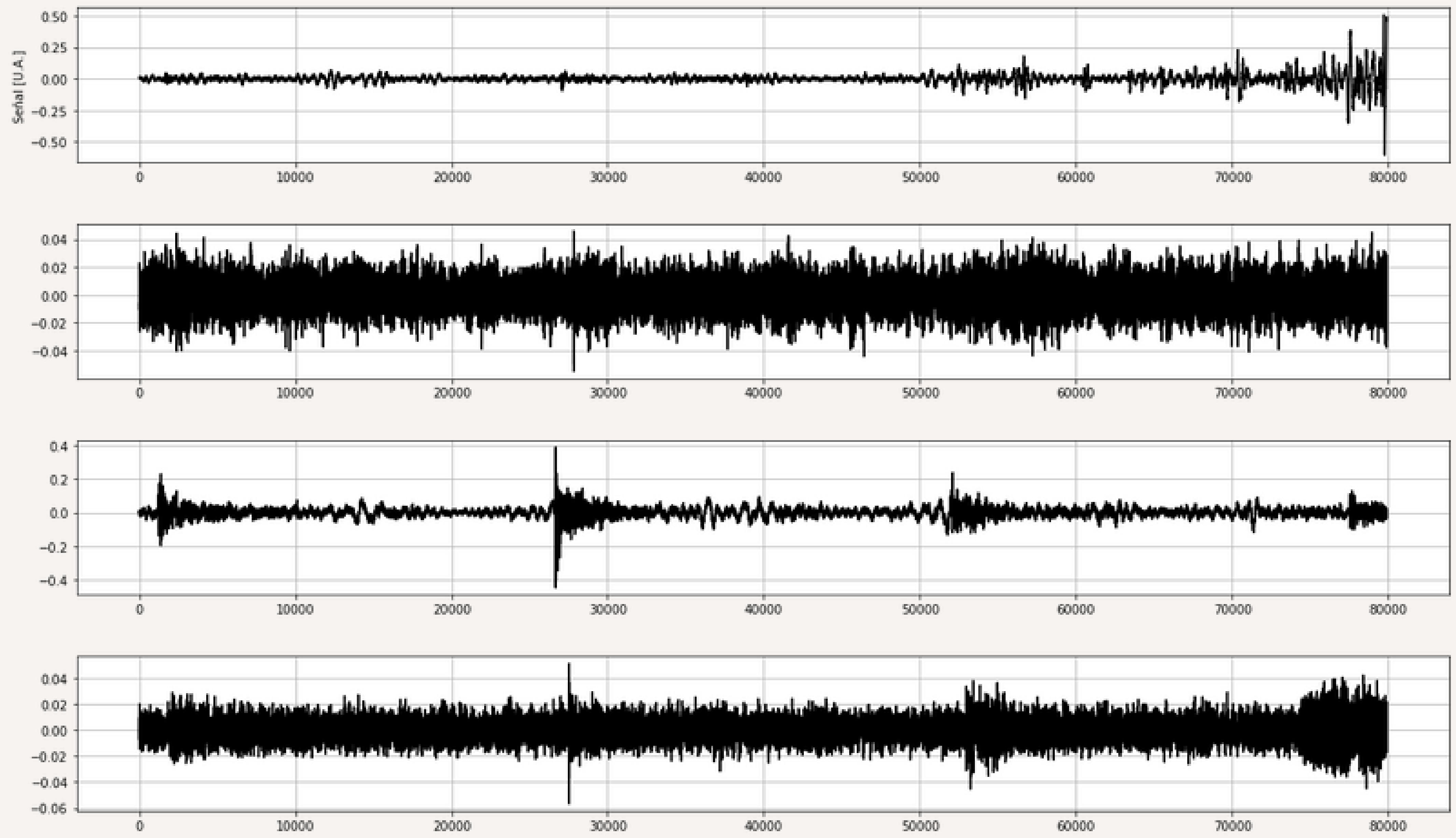


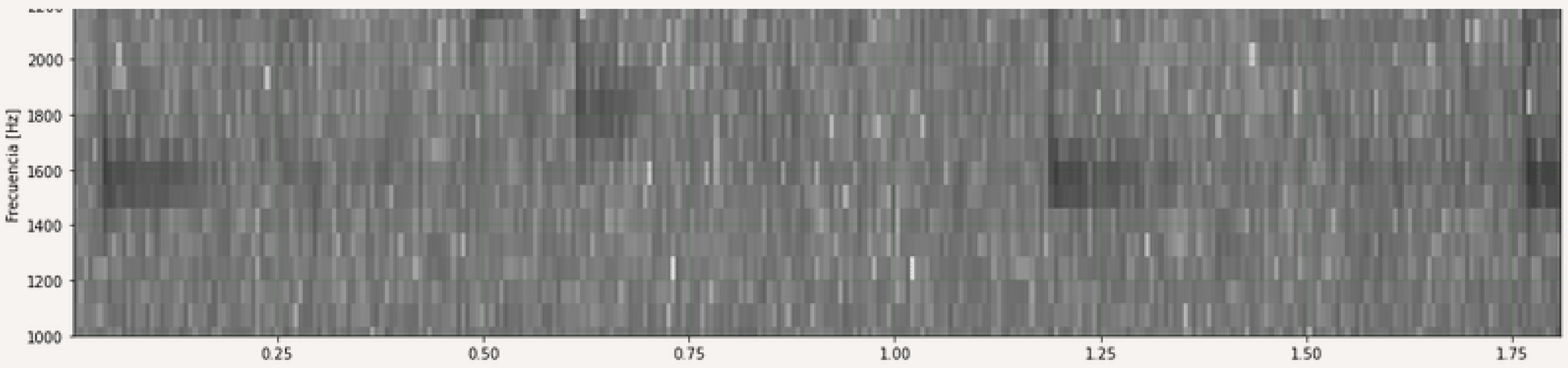
Algunos resultados





Retos

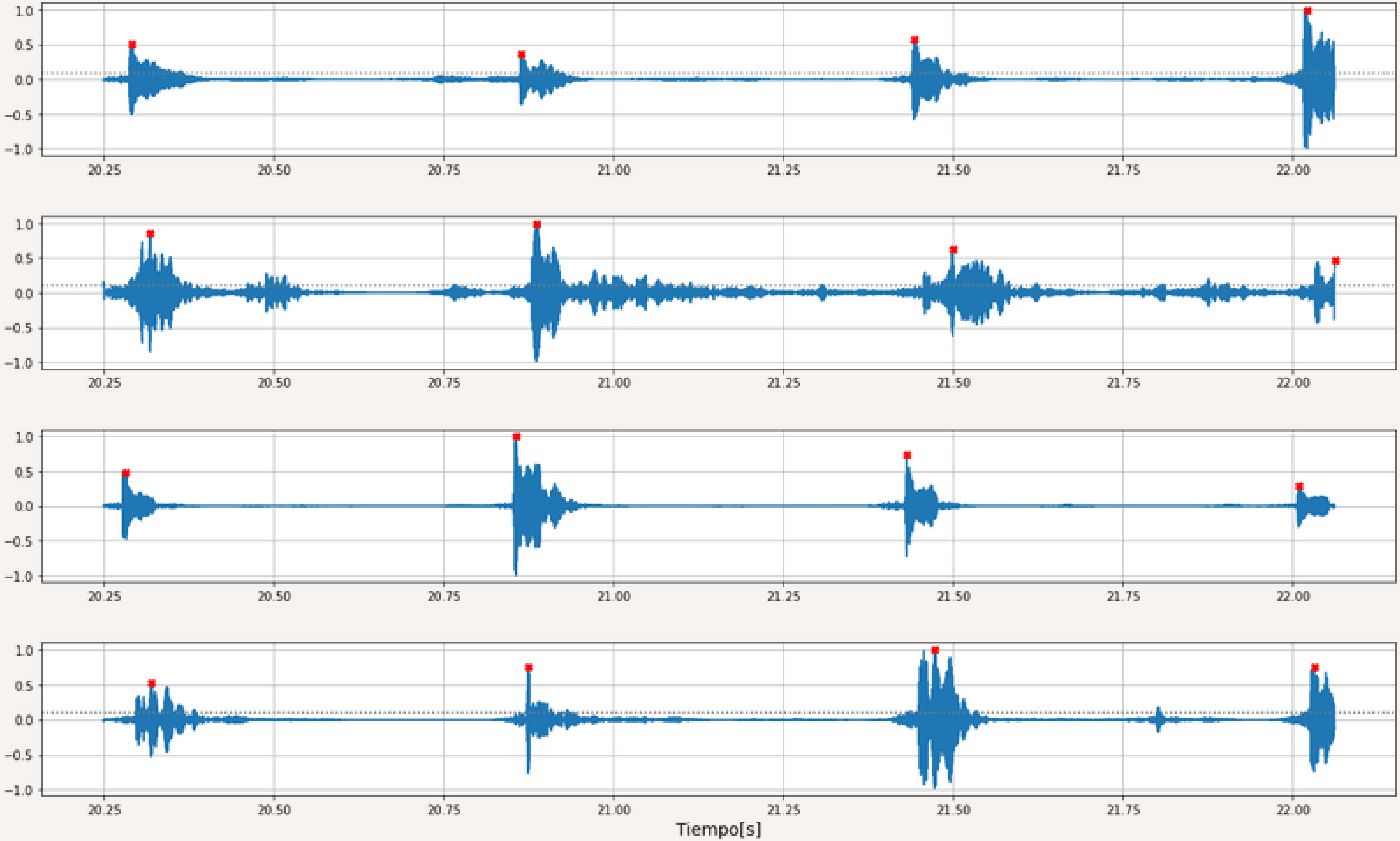




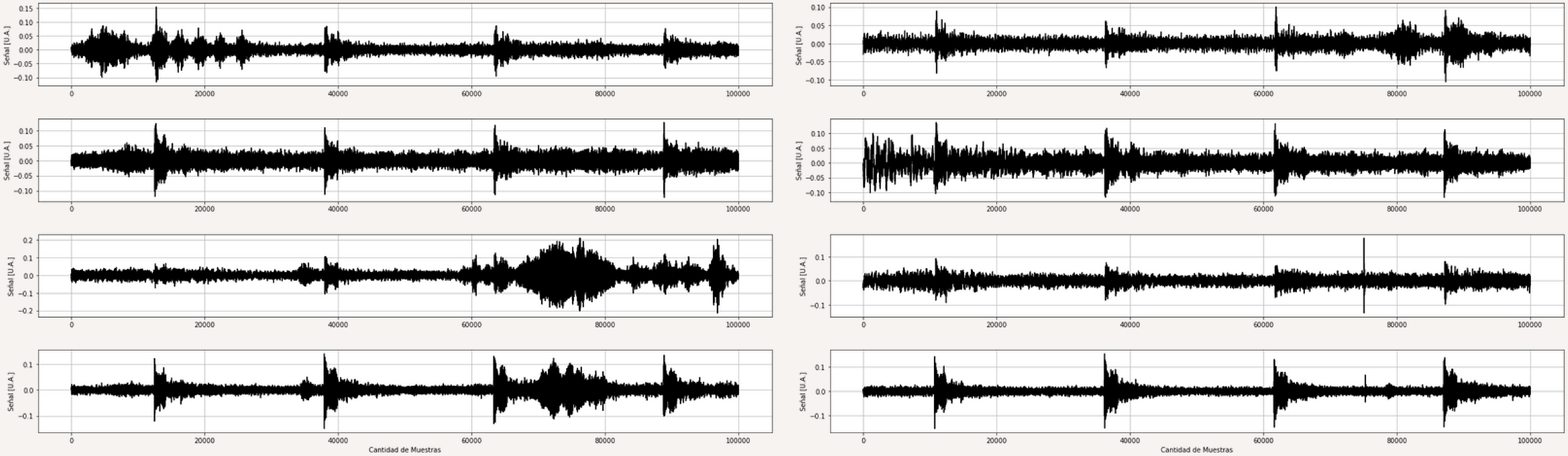
x=3.5m

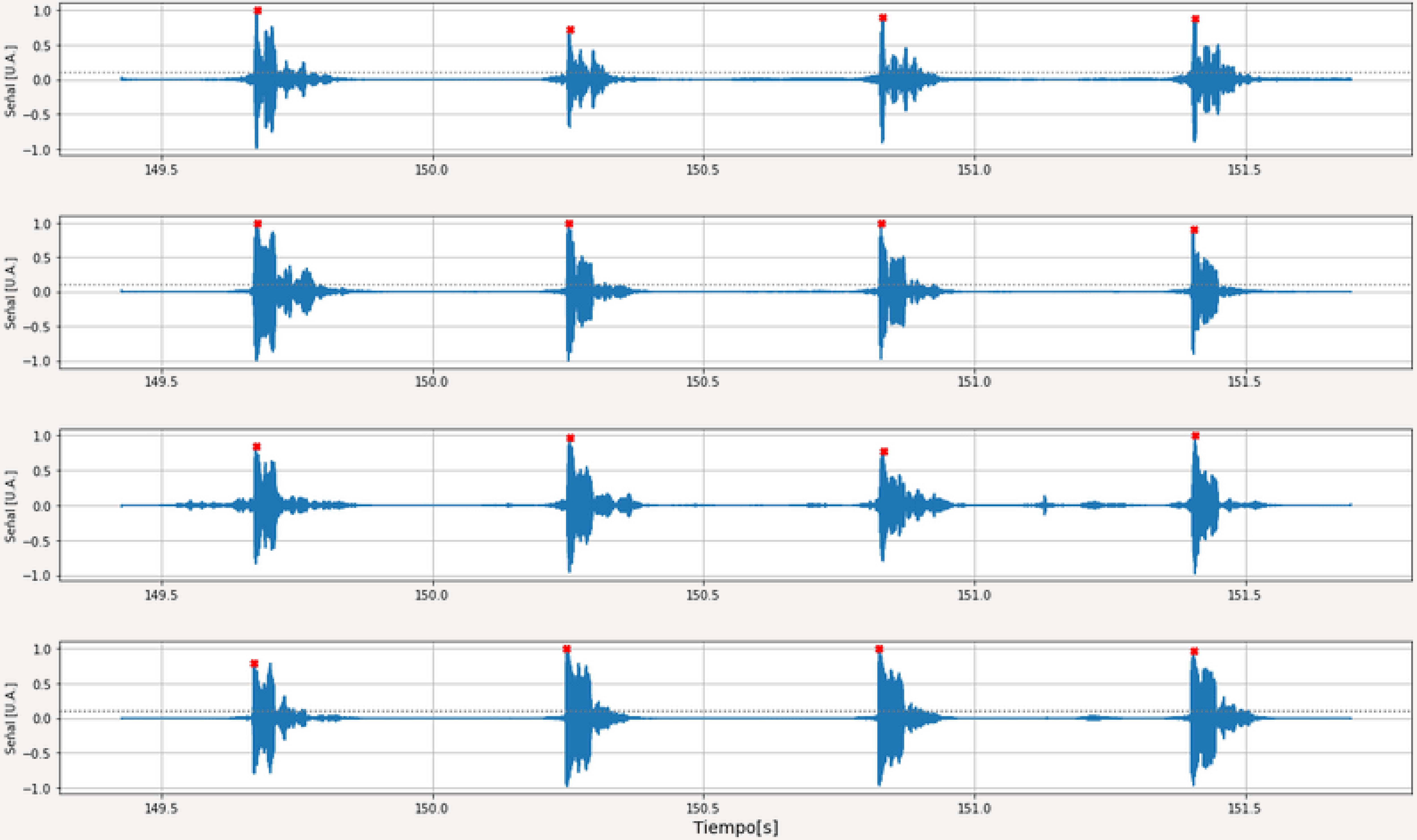
y=13.5m

Las señales n
siempre son
nítidas


1.0
0.5
0.0
−0.5
−1.0
20.25
20.50
20.75
21.00
21.25
21.50
21.75
22.00
Tiempo[s]

Elección de tramos de la señal de Audio




Señal [U.A.]
1.0
0.5
0.0
−0.5
−1.0
149.5
150.0
150.5
151.0
151.5
Tiempo[s]

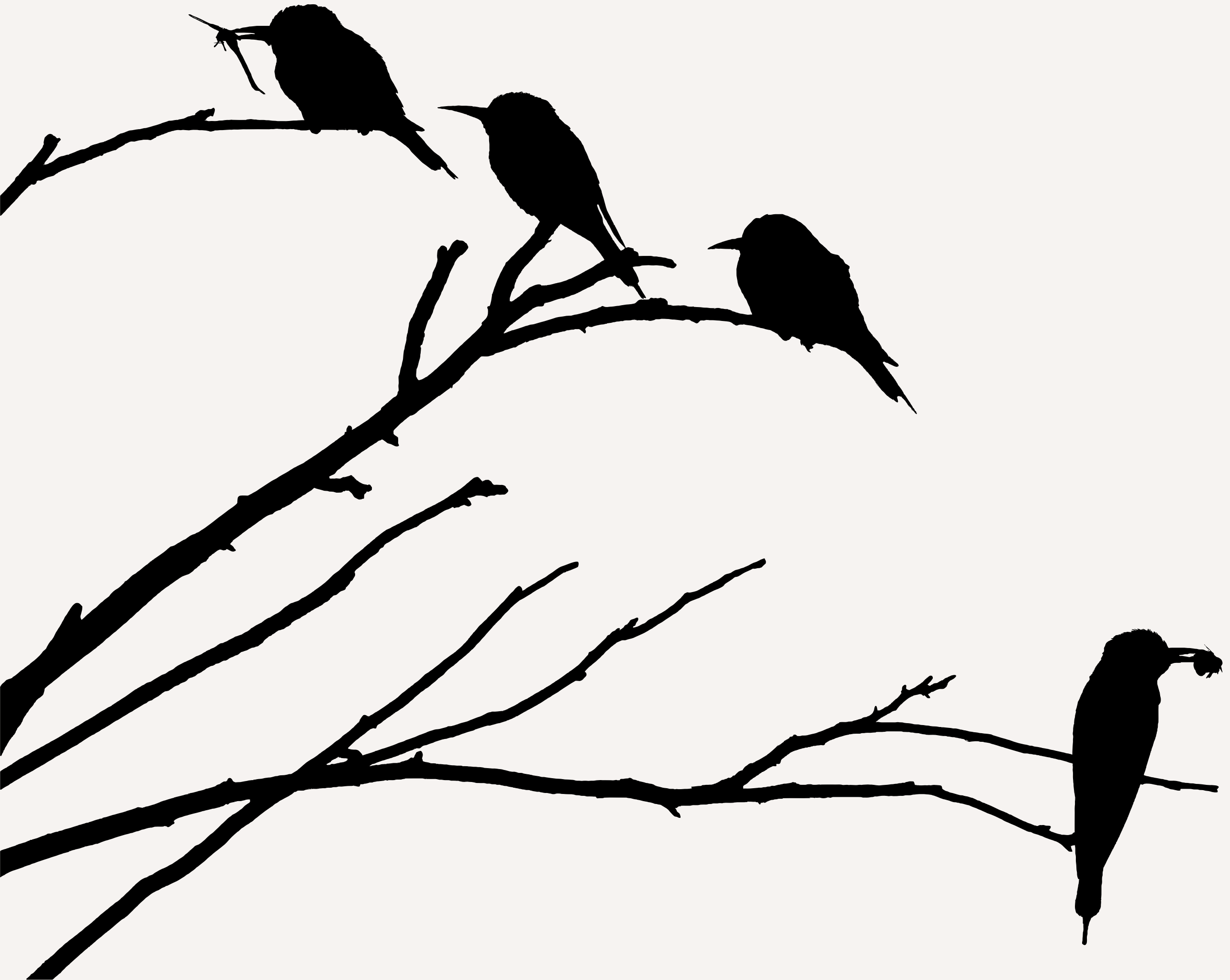
Conclusiones

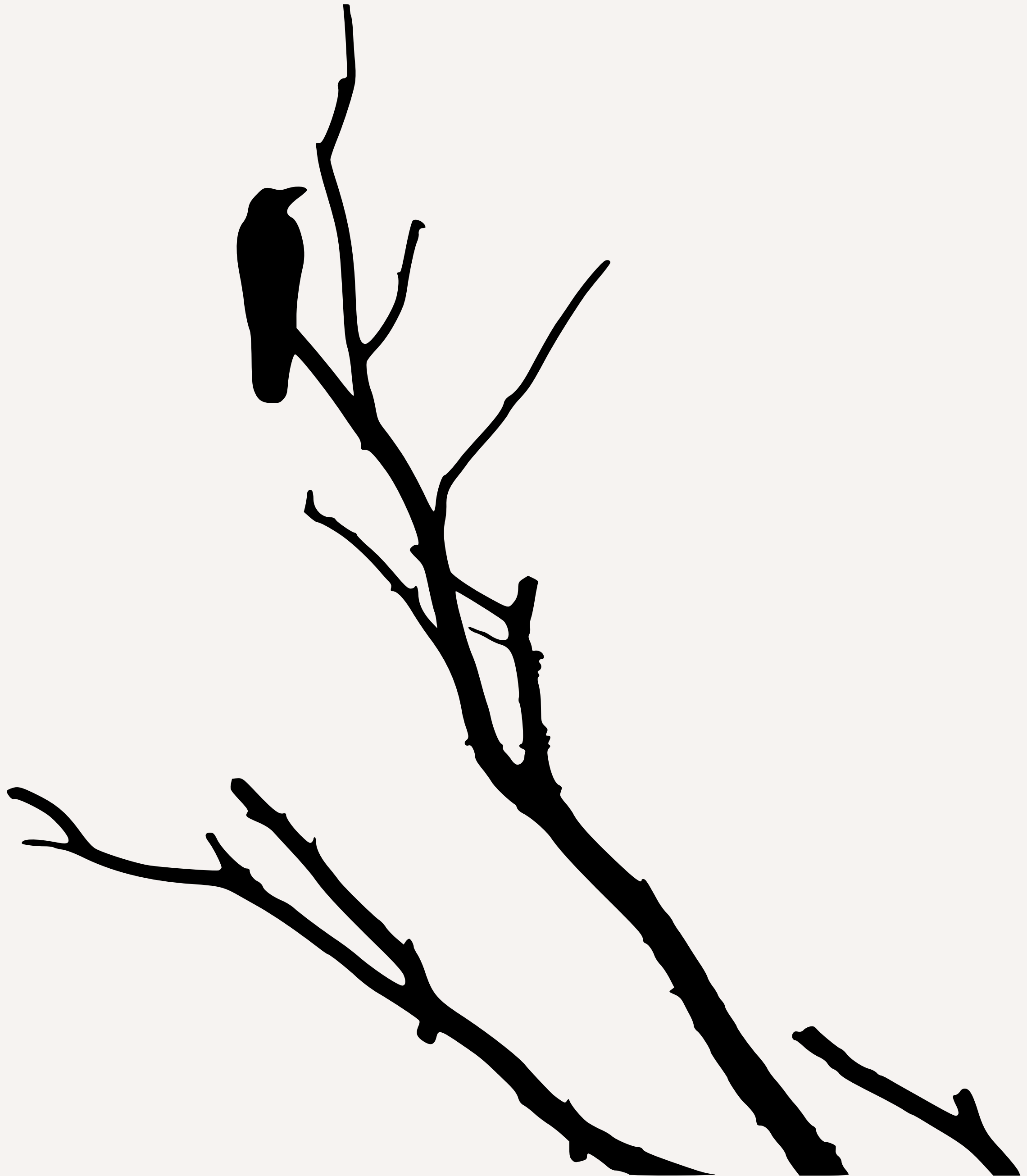
- Se diseñó una Red Neuronal capaz de traducir 6 delays temporales en posiciones y se caracterizó el error.

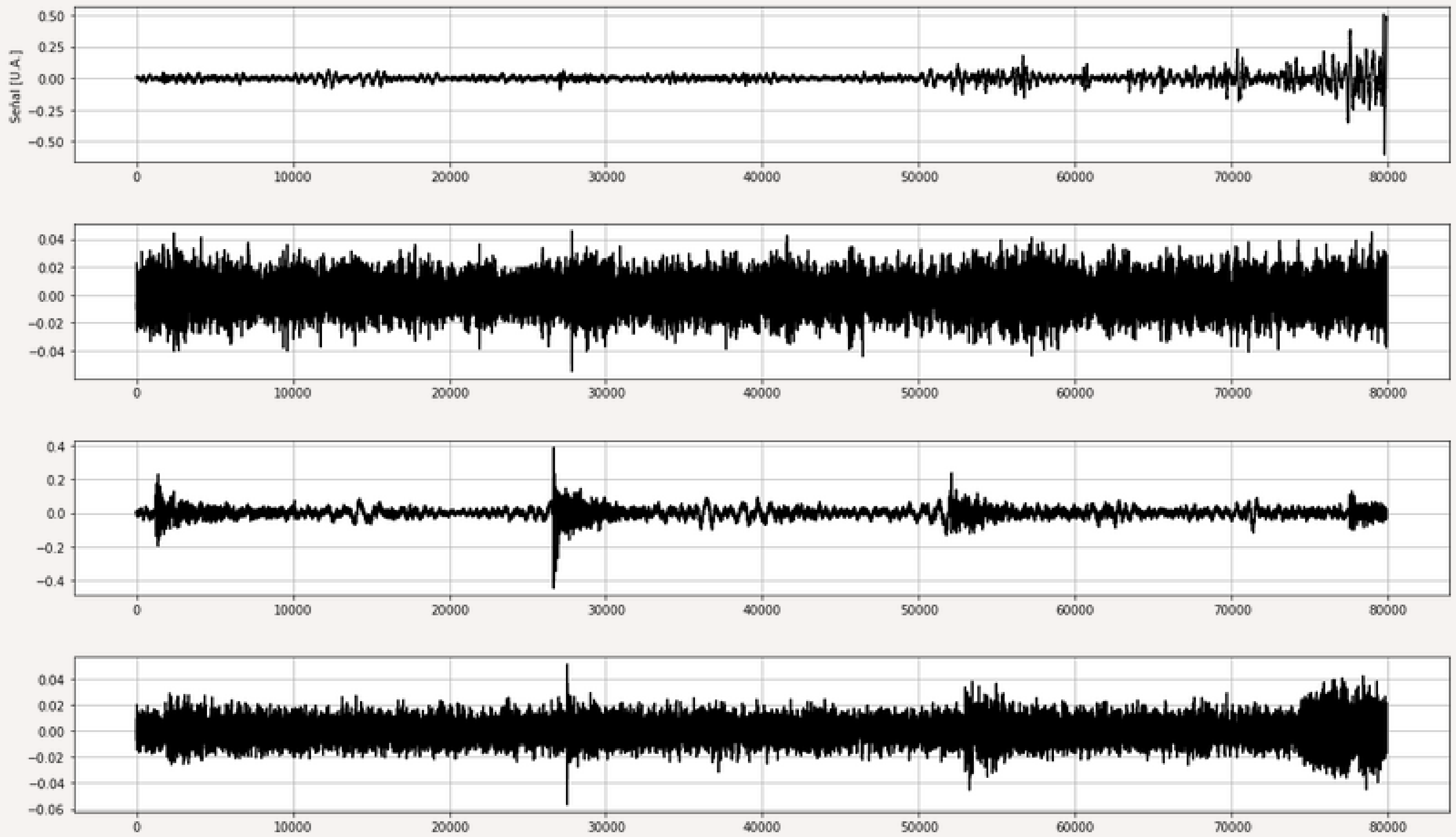
- Se desarrolló una rutina de procesamiento de datos capaz de determinar los delays temporales a través de grabaciones de audio.

- No fue posible automatizar el procesamiento de audio y la calidad original del mismo tiene un peso más elevado del esperado en la detección.

¡Gracias!






Señal [U.A.]
0.50
0.25
0.00
−0.25
−0.50
0
10000
20000
30000
40000
50000
60000
70000
80000
0.04
0.02
0.00
−0.02
−0.04
0.4
0.2
0.0
−0.2
−0.4
−0.06