

Proyecto Final:

Eliminación de ruido de Señales de Ondas Gravitacionales Usando AI Generativa (Autoencodificador Variacional - VAE)

1. Introducción

Las ondas gravitacionales, ondulaciones en el espacio-tiempo causadas por objetos masivos como agujeros negros o estrellas de neutrones, se han convertido en un área fundamental de estudio en la física moderna. Estas ondas transportan información valiosa sobre eventos astrofísicos, pero su detección puede ser un desafío debido a la presencia de ruido en los datos. El ruido en las señales de ondas gravitacionales proviene de diversas fuentes, incluyendo el ruido instrumental, factores ambientales y señales de fondo.

La capacidad de eliminar el ruido de los datos de ondas gravitacionales es crucial para mejorar la precisión y sensibilidad de detectores como LIGO (Laser Interferometer Gravitational-Wave Observatory). Los métodos tradicionales para la eliminación de ruido son computacionalmente costosos y pueden no ser suficientes para tratar la complejidad inherente de las señales de ondas gravitacionales, ya que pueden no capturar adecuadamente las estructuras subyacentes y patrones de la señal.

El aprendizaje automático, en particular las técnicas de aprendizaje profundo, ha surgido como una solución prometedora para este problema. Entre estas técnicas, los Autoencodificadores Variacionales (VAEs) ofrecen una ventaja única debido a su capacidad para aprender representaciones latentes probabilísticas de los datos. Este proyecto tiene como objetivo aprovechar un modelo VAE para eliminar el ruido de las señales de ondas gravitacionales, mejorando su calidad y permitiendo un análisis más preciso de los eventos astrofísicos.

2. Objetivo

El objetivo principal de este proyecto es desarrollar un Autoencodificador Variacional (VAE) capaz de eliminar el ruido de las señales de ondas gravitacionales. Específicamente, el proyecto tiene como metas:

- Diseñar e implementar una arquitectura VAE adaptada para procesar datos de series temporales de ondas gravitacionales unidimensionales.
- Entrenar el modelo en un conjunto de datos que consiste en señales de ondas gravitacionales ruidosas, teniendo las señales limpias como objetivo de predicción.
- Evaluar el rendimiento del VAE comparando las salidas limpiadas con las señales limpias y cuantificando las mejoras en la calidad de la señal.

El objetivo final es crear un modelo que pueda eliminar de manera efectiva el ruido de los datos de ondas gravitacionales, contribuyendo así a la detección y análisis de eventos astrofísicos como las fusiones de agujeros negros.

3. Datos

Para entrenar al modelo, se creó un dataset de segmentos de ruido real a los cuales se les inyectaron señales obtenidas de simulaciones numericas de ondas producidas por la fusion de dos agujeros negros. La señal combinada funciona como entrada al modelo y la señal simulada como objetivo. Los datos fueron tratados de la siguiente manera.

3.1 Ruido Real

- Se descargaron 50 archivos de 4096 segundos de duración cada uno, con una tasa de muestreo de 4096 Hz. Los datos fueron obtenidos a través del API de Gravitational Wave Open Science Center.
- Los datos corresponden al detector LIGO Hanford (H1) ubicado en Washington.
- El rango de fechas va del 30 de noviembre de 2016 al 25 de agosto de 2017 (correspondientes a la corrida O2).
- Los periodos de tiempo fueron seleccionados aleatoriamente, con el objetivo de capturar la mayor variedad de condiciones de ruido posibles, dado que el ruido instrumental del LIGO presenta gran variabilidad temporal.
- Los datos fueron segmentados en intervalos de 8 segundos cada uno, resultando en un total de 25,261 señales (cada señal de 32,768 puntos).
- Se eliminaron:
 - Segmentos donde ocurrieron eventos gravitacionales reales (consultando catálogos públicos).
 - Segmentos etiquetados como de mala calidad (data quality vetoes).
- Para cada segmento de 8 segundos:
 - Se calculó su Densidad Espectral de Potencia (PSD) individualmente.
 - El PSD obtenido se utilizó para blanquear:
 - El propio segmento de ruido.
 - La señal simulada a inyectar.
- Filtrado posterior: luego del blanqueado, tanto ruido como señal simulada fueron filtrados usando un paso de banda entre 20 Hz y 500 Hz, ya que:
 - Por debajo de 20 Hz predomina el ruido sísmico.
 - Por encima de 500 Hz se incrementan efectos cuánticos y ruido instrumental.
- Segmentación y preparación final:
 - Cada segmento blanqueado se acortó de 8 segundos a 1 segundo (4096 puntos).
 - En la mitad de los segmentos (alrededor de 12,630):
 - Se inyectó una señal simulada, previamente escalada para simular distintas relaciones señal-ruido (SNR).

- En la otra mitad:
 - No se inyectó ninguna señal.
 - Se asignó una señal constante de ceros.
- Esto permite al modelo distinguir entre ruido puro y eventos reales, evitando un sesgo a detectar siempre una señal.

3.2 Señales Simuladas

- Las ondas gravitacionales simuladas fueron generadas utilizando el modelo numérico **SEOBNRv4_opt** a través de la librería **PyCBC**.
- Las simulaciones comienzan en **30 Hz** de frecuencia para cubrir el rango sensible del detector.
- El rango de **masas de los agujeros negros** utilizados para generar las señales fue de **10 a 60 masas solares**.
- Las señales fueron **normalizadas** para tener potencia unitaria antes de la inyección.
- Se aseguraron las siguientes condiciones:
 - La misma PSD usada para blanquear el ruido fue usada para blanquear la señal.
 - Cada señal inyectada mantiene una relación controlada de SNR tras blanqueo y filtrado.

3.3 Inyección de Señales

- Las señales simuladas fueron **inyectadas** directamente sobre el ruido real **blanqueado** y filtrado.
- El SNR fue controlado ajustando la amplitud de las señales inyectadas.
- El cálculo del SNR se realizó en **potencia**, no en decibeles.
- Posteriormente, los datos resultantes fueron usados para entrenar y validar el modelo.

4. Metodología

4.1 Arquitectura del VAE

Se diseñó un Autoencoder Variacional unidimensional adaptado a la estructura de las series temporales de 1 segundo.

Encoder:

- **3 capas convolucionales 1D** con funciones de activación **ReLU**.
- Compresión de la entrada a un espacio latente de **64 dimensiones**.

Espacio latente:

- Implementación de un **reparametrization trick** que permite muestrear a partir de la media (μ) y desviación estándar (\log_var) aprendidas.

Decoder:

- **3 capas convolucionales 1D** deconvolutivas (upsampling + convolución), también con activaciones **ReLU**.
- Activación final: **linear** para permitir la reconstrucción precisa de valores positivos y negativos.

4.2 Función de pérdida

La función de pérdida combinó dos términos:

- **Reconstrucción (MAE)**: error absoluto medio entre la señal denoised y la señal limpia inyectada.
- **KL Divergence**: regularizador del espacio latente para aproximarlos a una distribución gaussiana estándar.

4.3 Entrenamiento

- **75% del dataset** fue utilizado para entrenamiento y **25%** para **validación** (alrededor de **6315 señales** de validación).
- Se utilizó el optimizador **Adam** con una tasa de aprendizaje de **1e-4**.
- El modelo fue entrenado por **50 épocas**.
- Sin embargo, se observó que la **función de pérdida alcanzó su mínimo en la época 12**, después de lo cual se mantuvo constante.
- Por lo tanto, se **seleccionaron los parámetros** correspondientes a la **época 12** para evitar **overfitting**.
- Se utilizó **early stopping** para garantizar que no se siguiera entrenando innecesariamente después de converger.
- Se registraron métricas de desempeño tanto en el conjunto de entrenamiento como en el de validación.

5. Resultados

5.1 Evaluación sobre datos de validación

El modelo fue evaluado en el conjunto de validación, compuesto tanto por segmentos con señales inyectadas como por segmentos de ruido puro.

- Para señales inyectadas:
 - Se obtuvo una **correlación de Pearson promedio** de **0.88** entre las señales reconstruidas y las plantillas originales (antes de la inyección).
 - La reconstrucción conservó correctamente las fases y amplitudes relativas de las señales simuladas.

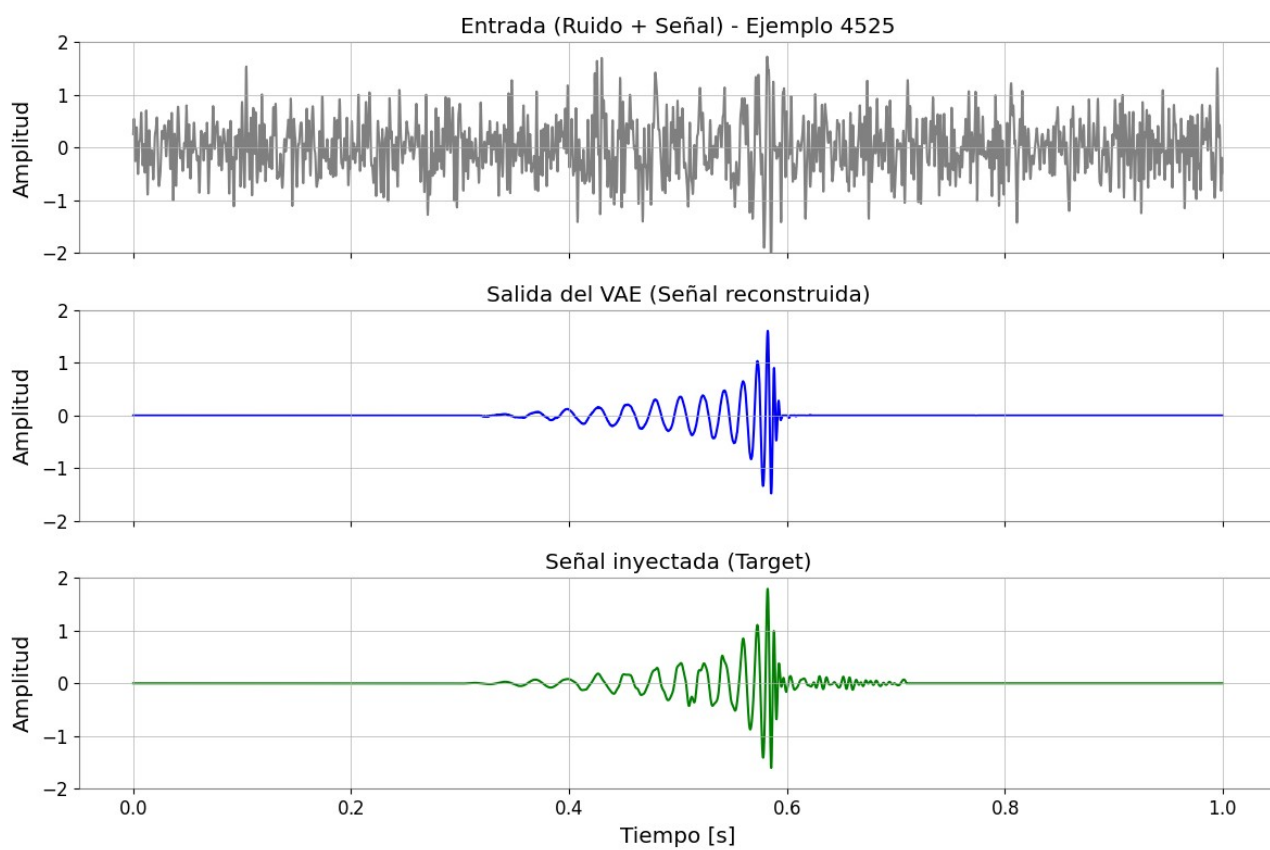
- Se observó una mejora significativa en la **relación señal-ruido** tras el proceso de denoising, logrando recuperar eventos que antes eran indetectables a simple vista.
- Para segmentos de ruido puro:
 - El modelo no introdujo falsas señales.
 - La salida para segmentos sin inyección correspondió a una señal casi nula, confirmando que el VAE no “alucina” señales donde no existen.

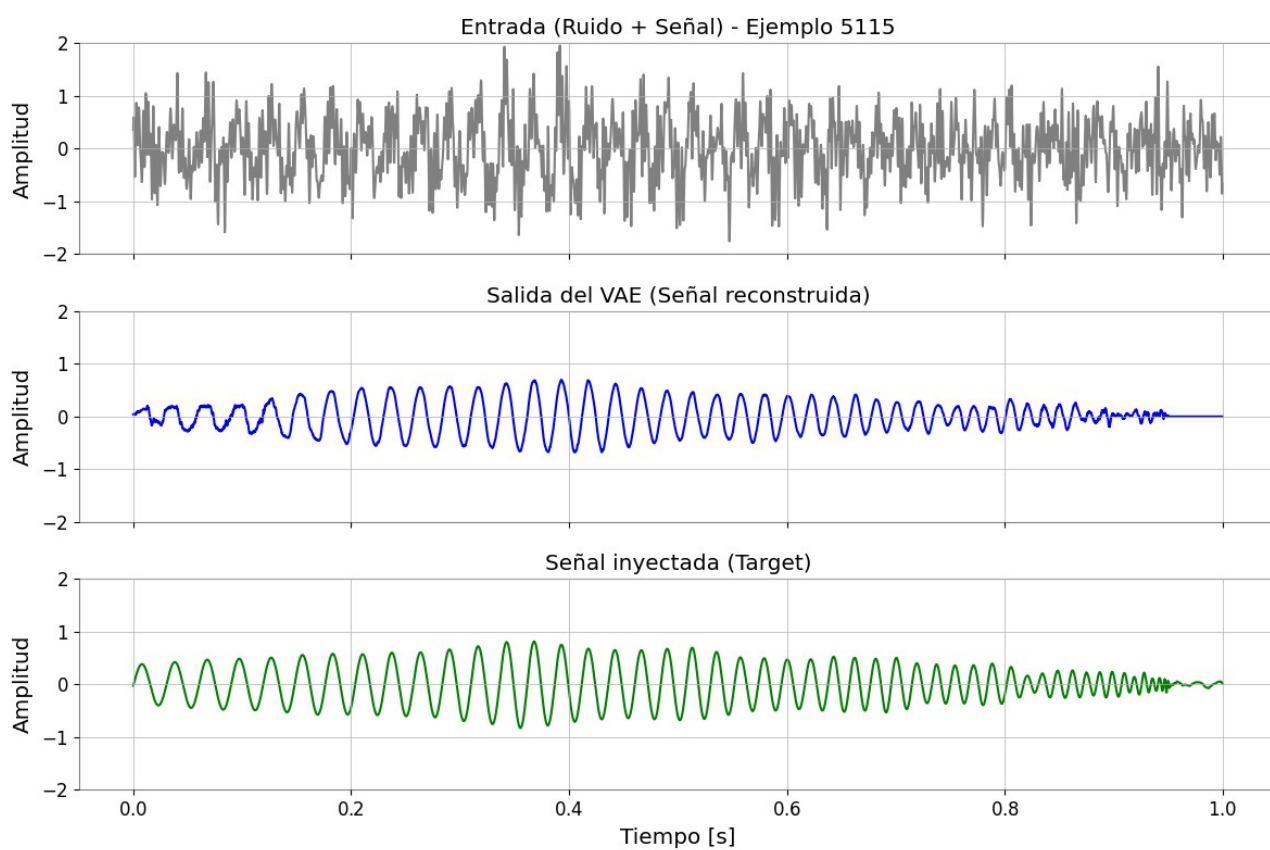
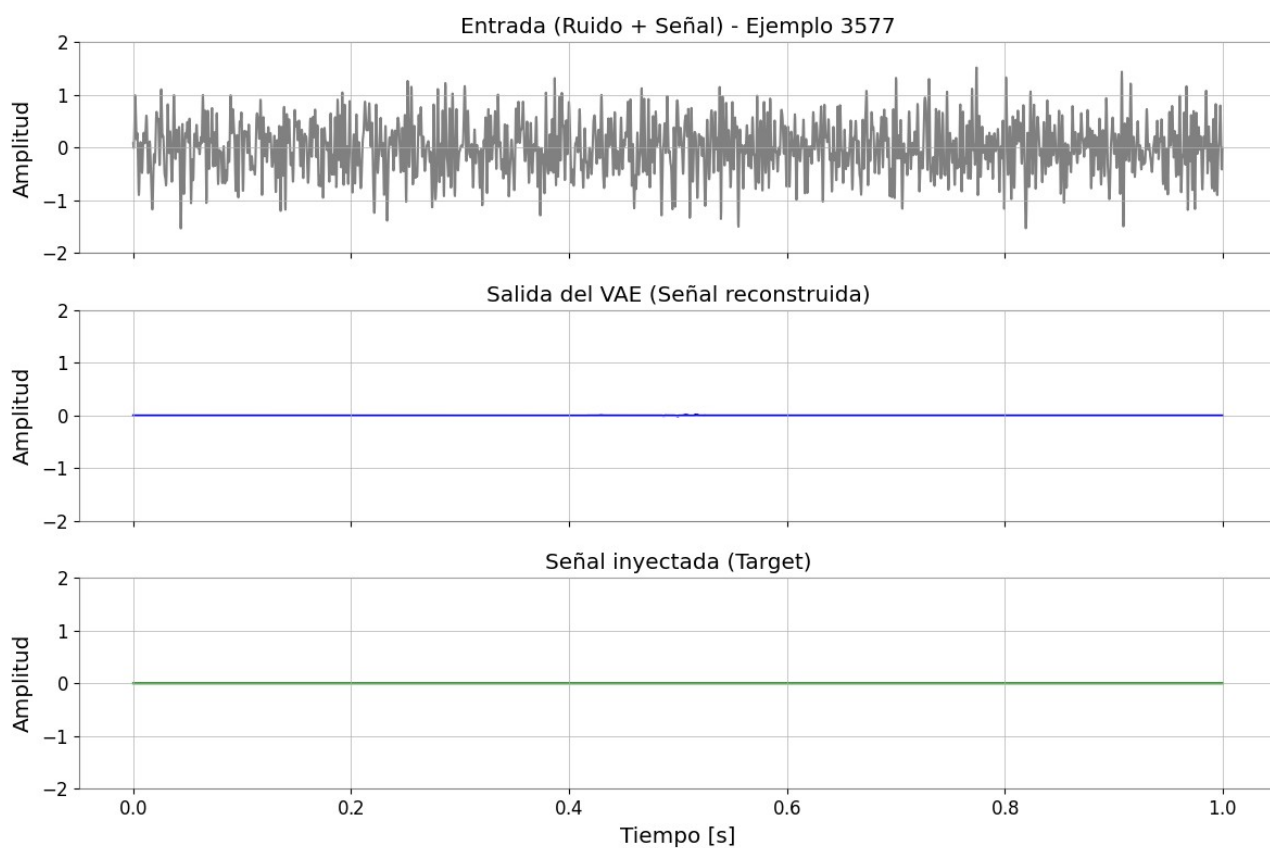
5.2 Análisis visual de ejemplos

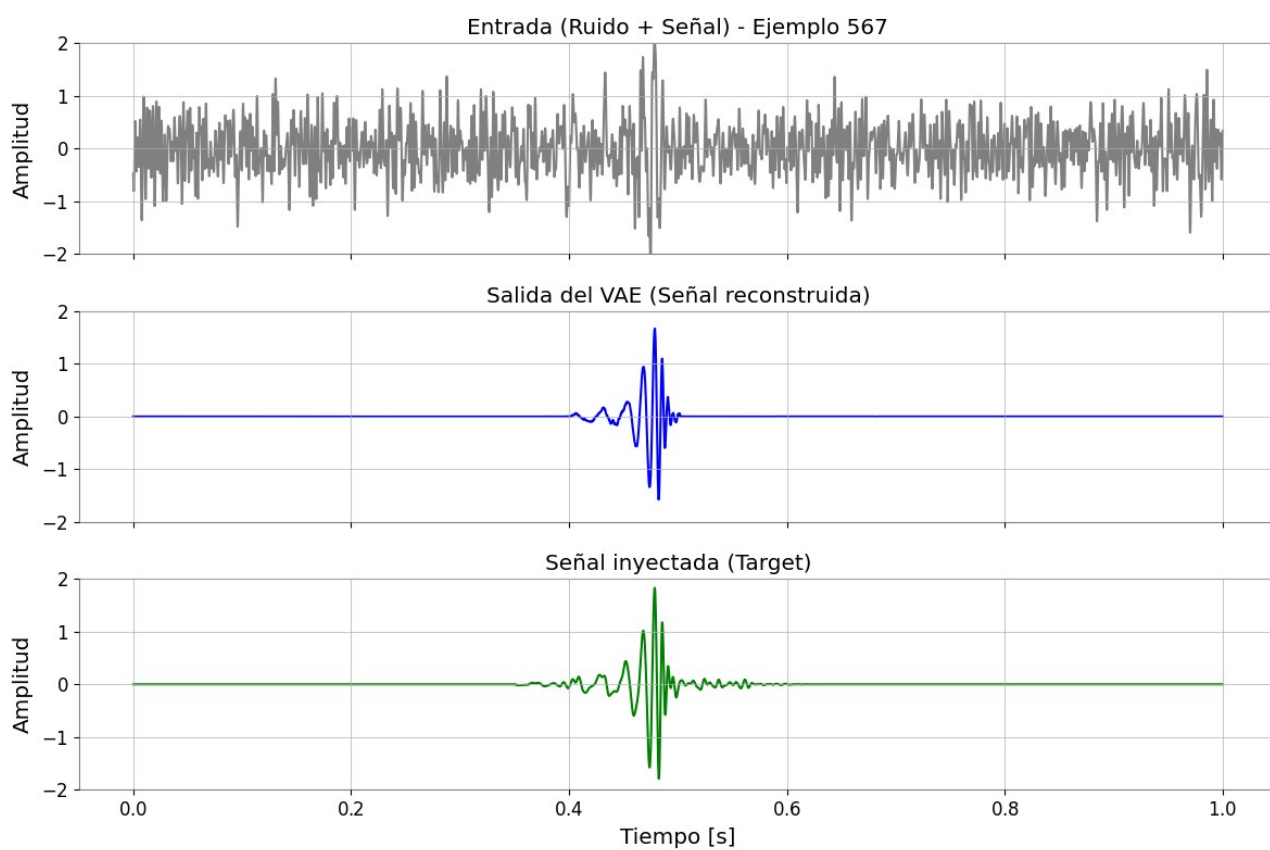
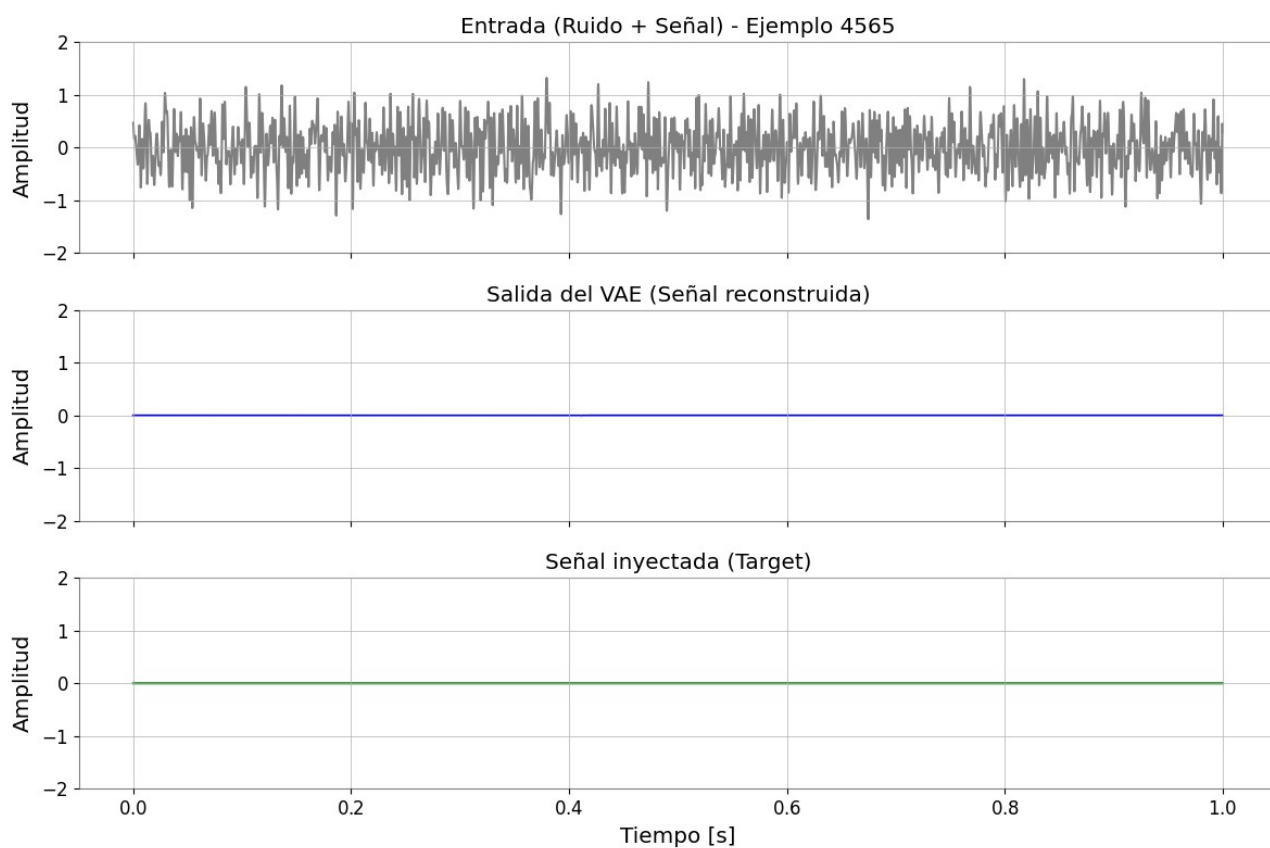
Se realizaron inspecciones visuales de segmentos individuales pertenecientes al set de validación:

- Antes del denoising, las señales inyectadas o reales eran **casi indistinguibles** del ruido.
- Después del denoising:
 - Las señales reconstruidas son **claramente visibles**.
 - La estructura de chirp característica de las fusiones de agujeros negros se recupera de forma precisa.

Se presentan a continuación algunos ejemplos:





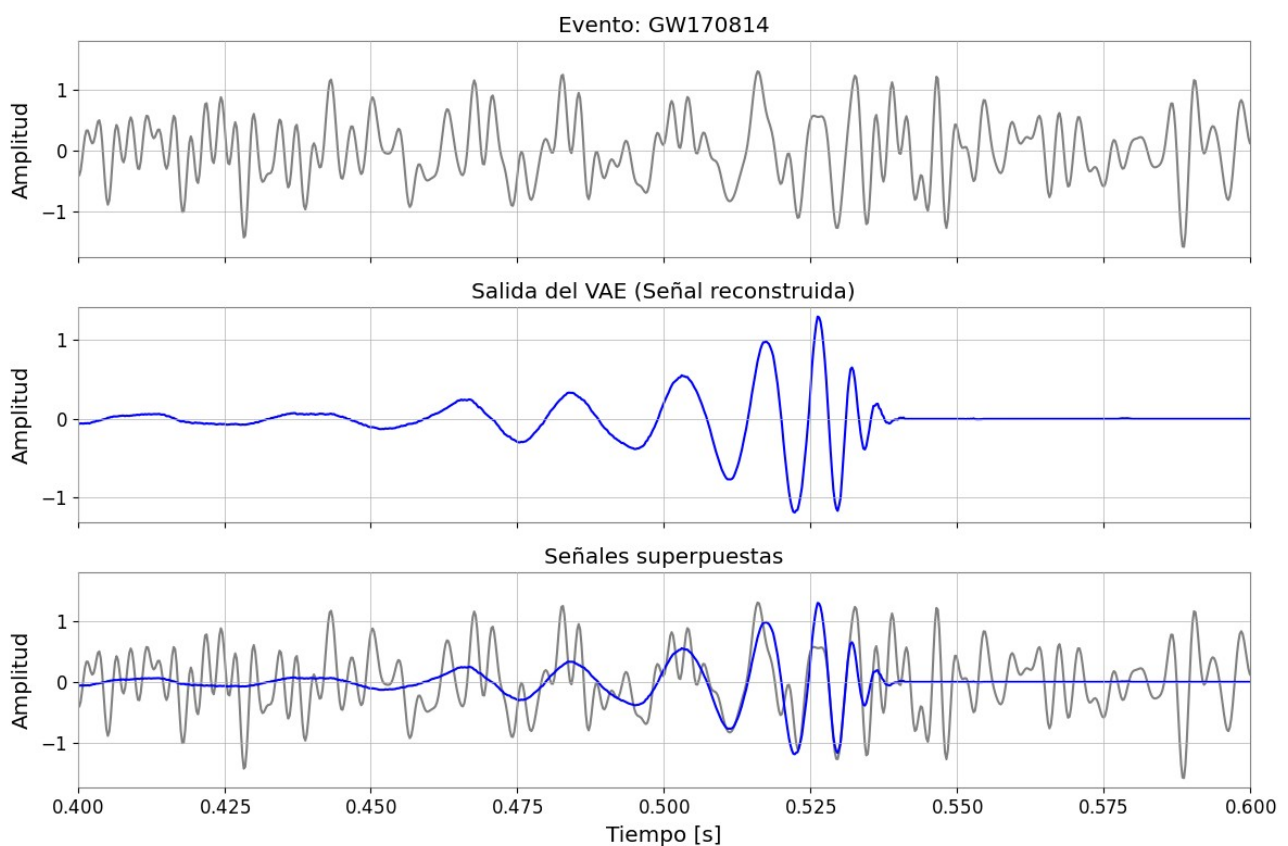


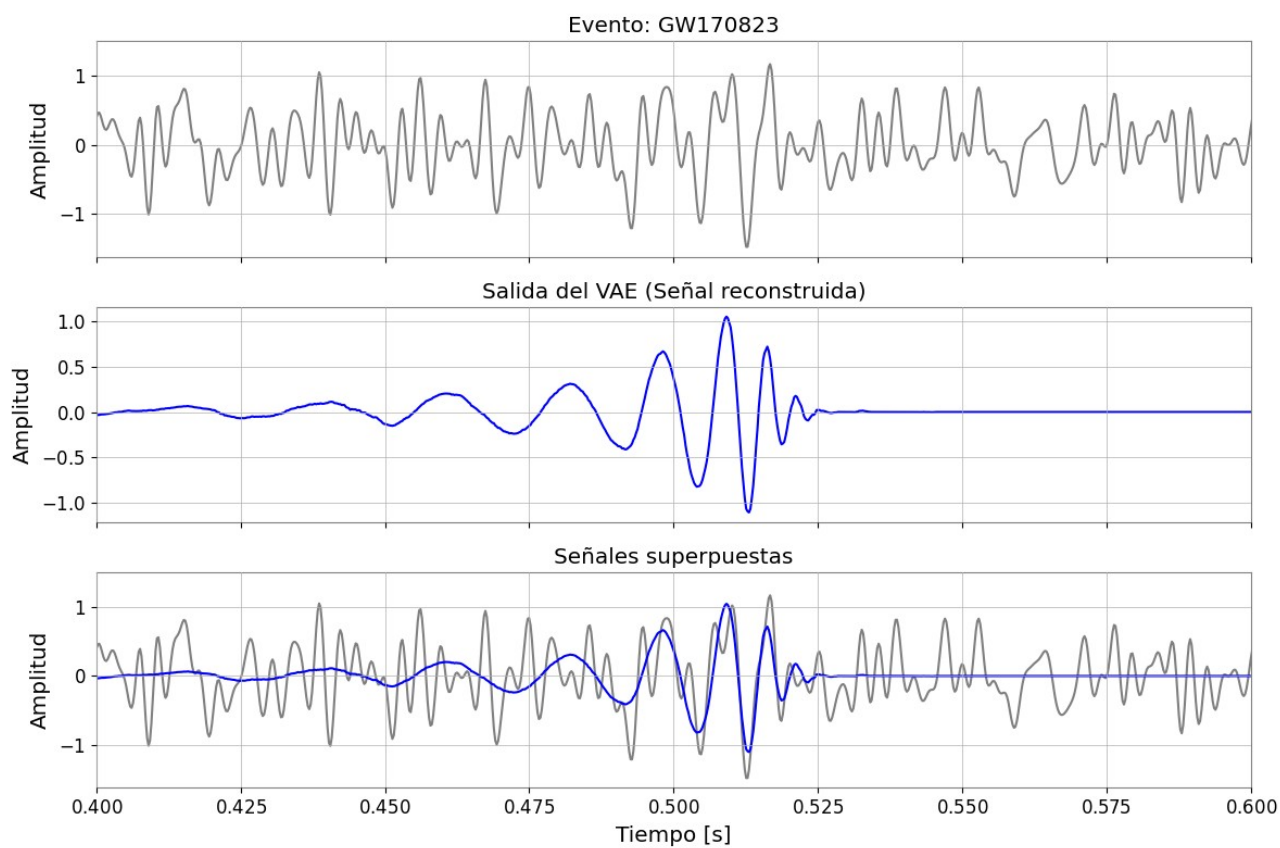
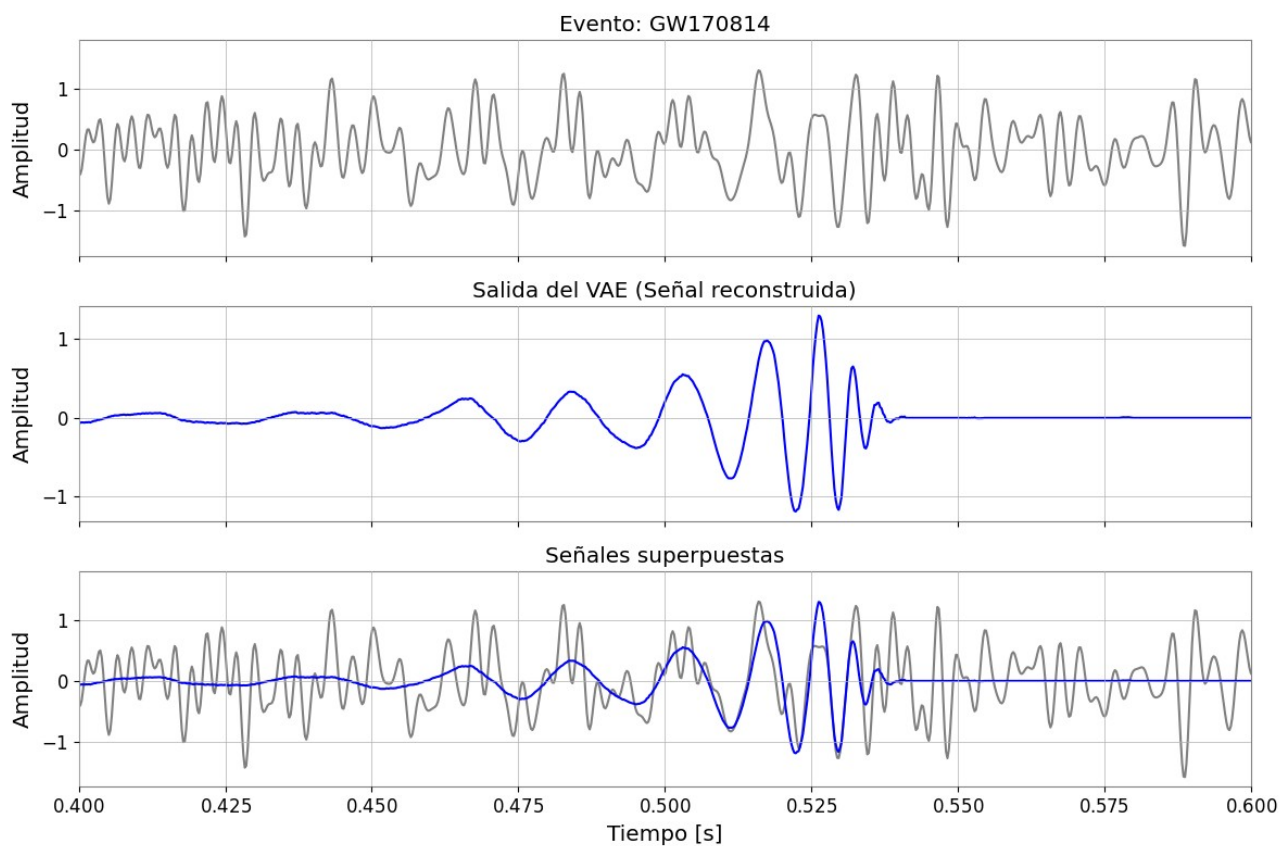
4.2 Evaluación sobre eventos reales de la corrida O2

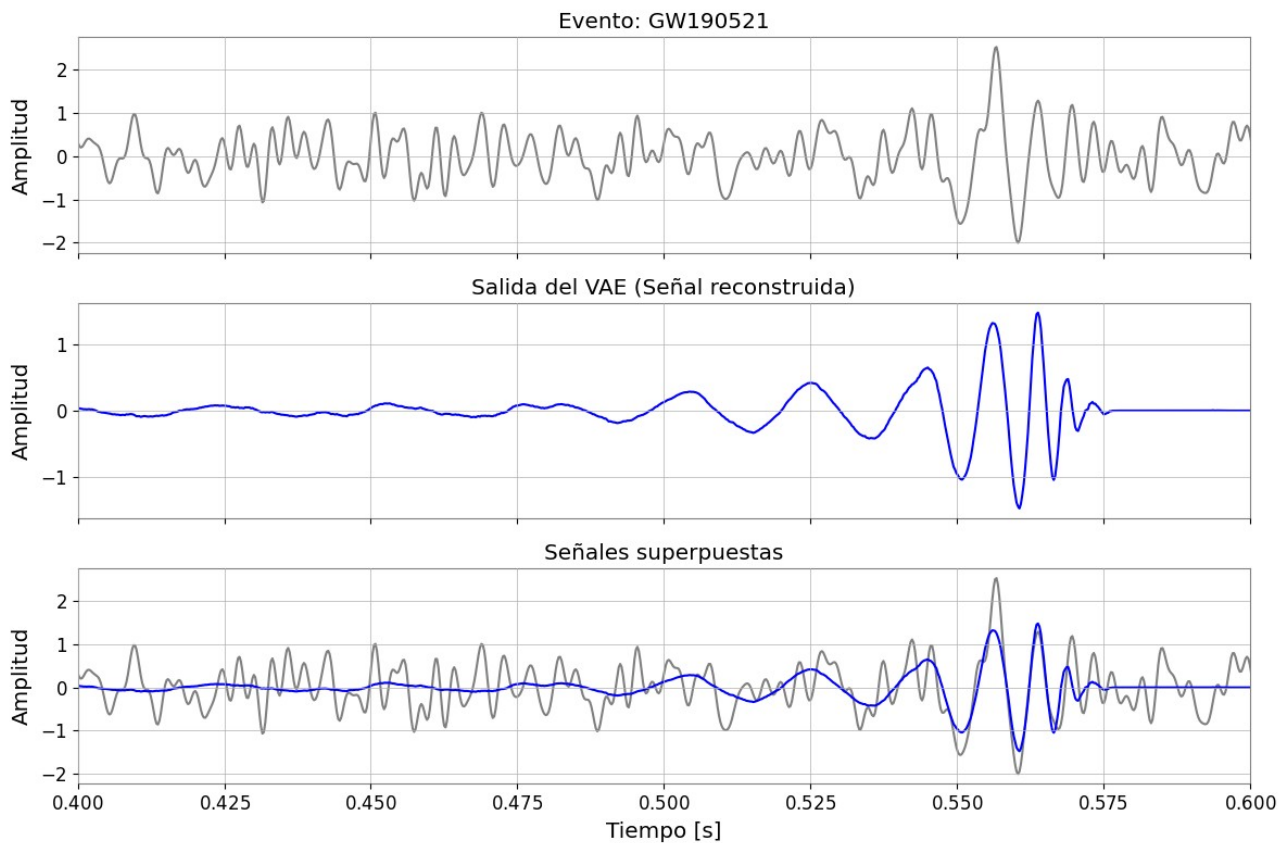
Se aplicó el modelo a datos reales que contienen eventos confirmados por la colaboración LIGO/Virgo durante la corrida O2:

- El modelo fue capaz de **resaltar y limpiar** los eventos reales de manera efectiva, recuperando la estructura de las señales gravitacionales esperadas.
- Las señales reconstruidas coincidieron tanto en **tiempo de llegada** como en **forma de onda** general con los templates publicados.
- Este resultado confirma que el modelo no solo funciona en datos simulados, sino que también generaliza a eventos verdaderos con ruido real complejo.

Nota: No todos los detalles finos de las señales reales fueron recuperados, en parte debido a las limitaciones impuestas por el preprocesamiento (ver siguiente sección).







4.3 Análisis de la reconstrucción en frecuencia

Al analizar las señales reconstruidas en el dominio de Fourier, se observaron algunas limitaciones:

- Las **frecuencias por debajo de 20 Hz** y por encima de **500 Hz** no fueron recuperadas.
- Esto es **esperado** debido a dos factores principales:
 1. Se aplicó un **filtro de pase de banda de 20 Hz a 500 Hz** durante el preprocesamiento, eliminando información fuera de este rango.
 2. En las señales originales, las componentes de muy baja y muy alta frecuencia tienen **amplitudes comparables o inferiores** a las fluctuaciones típicas del ruido en esas bandas, haciendo extremadamente difícil su recuperación incluso teórica.

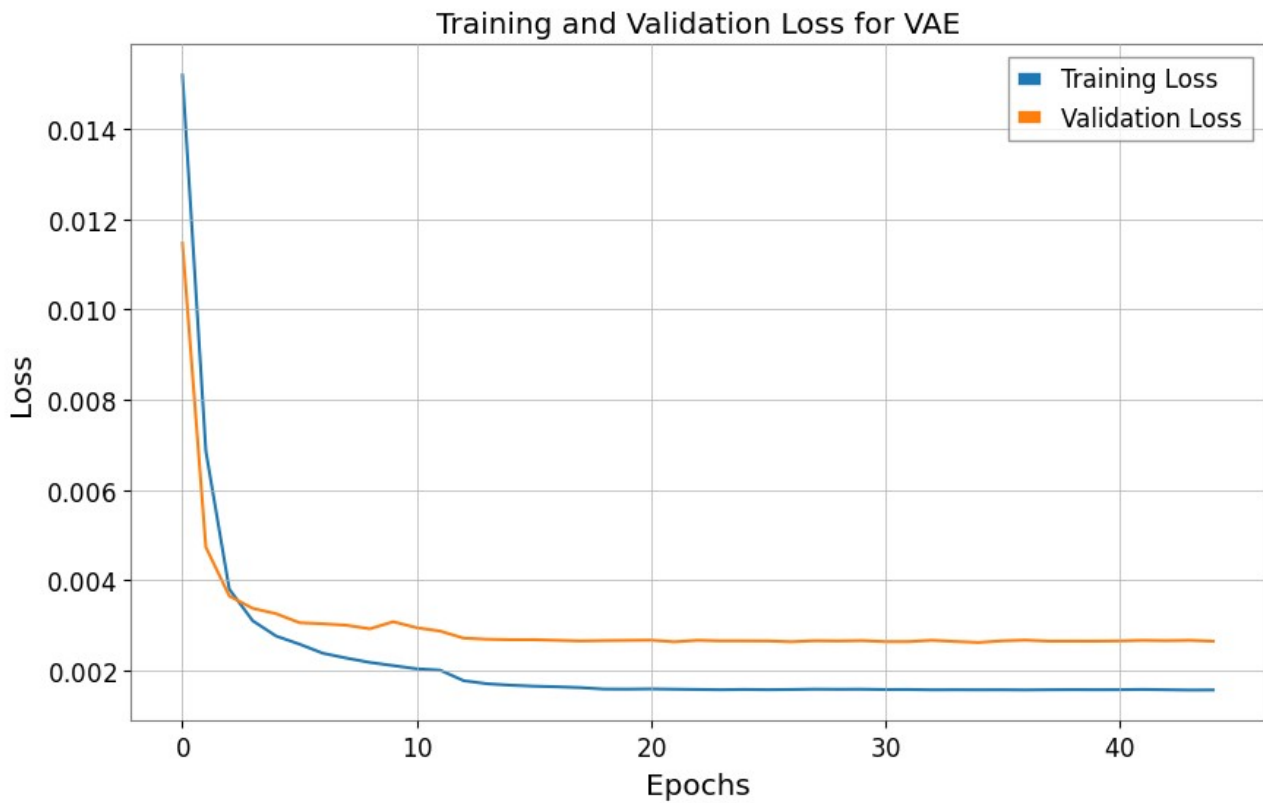
Así, el modelo se centra correctamente en reconstruir las partes de la señal donde la relación señal-ruido permite una recuperación fiable.

4.5 Evolución de la función de pérdida

El comportamiento de la función de pérdida durante el entrenamiento mostró:

- Una **disminución rápida** durante las primeras 10 épocas.
- Un **mínimo claro en la época 12**, después del cual la pérdida se estabilizó.
- No se observó un aumento en la pérdida de validación, indicando ausencia de overfitting hasta la detención.

La siguiente gráfica ilustra la evolución:



5. Conclusiones

La estrategia basada en Autoencoders Variacionales demuestra ser efectiva para limpiar señales gravitacionales inmersas en ruido altamente variable. La preselección cuidadosa de los datos, la normalización de las señales simuladas y la diferenciación explícita entre ruido puro y señales inyectadas resultaron ser factores clave en el éxito del modelo.

Este trabajo sugiere que técnicas de denoising basadas en machine learning pueden complementar métodos tradicionales como el filtrado adaptado en futuras búsquedas de ondas gravitacionales.