



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA

**Discriminación entre eventos
y ruido sísmico mediante un
modelo de clasificación**

TESIS

Que para obtener el título de

Ingeniero Geofísico

P R E S E N T A

Fernando Gamboa Rivero

DIRECTOR DE TESIS

Dr. Martín Cárdenas Soto



Ciudad Universitaria, Cd. Mx., 2019



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Dedicado a mis padres

Agradecimientos

A mi papá, por aportar la idea que me llevó a este trabajo, y por su apoyo y ayuda constante.

A mi mamá, por tenerme paciencia en lo que me llevó terminar este trabajo.

Al Dr. Martín Cárdenas, por proporcionarme los datos sobre los que se construyó este trabajo, y también por su paciencia.

A los demás integrantes del jurado: M.C. David Escobedo, Dr. Denis Legrand, Dra. Olivia Enríquez, e Ing. Thalía Reyes; por revisar el trabajo y contribuir con ideas para mejorarlo.

A todas las personas detrás de todos los *softwares* usados durante el desarrollo de este trabajo, sin los cuales no hubiera podido lidiar con los datos.

Al lector, para que se tome el tiempo de revisar al menos el Resumen, la Introducción y las Conclusiones.

Contenido

1. Resumen	1
2. Abstract	1
3. Introducción	2
4. Generalidades del modelo propuesto	2
4.1. Introducción a un modelo de clasificación	2
4.1.1. Planteamiento general de un problema y modelo de clasificación	2
4.1.2. El <i>score de corte</i>	3
4.1.3. La <i>matriz de confusión</i> y su relación con el <i>score de corte</i>	4
4.1.4. Conjuntos de datos de entrenamiento y prueba	5
4.1.5. El modelo de clasificación como una “caja negra” intercambiable	6
4.2. Propuesta de un modelo de clasificación basado en la distribución normal multivariada.	6
4.2.1. La distribución normal multivariada	6
4.2.2. Confección del modelo a partir de la distribución normal multivariada.....	7
4.2.3. Compensación numérica	8
5. Antecedentes	8
5.1. STA/LTA	11
5.2. Picado de ondas P y S por medio de redes neuronales	11
5.3. Picado de ondas P, S, y FZHW con una técnica particular	13
5.4. Una comparativa de técnicas basadas en la correlación cruzada, con una técnica proveniente del procesamiento de audio	14
5.5. Comentarios sobre los antecedentes	16
6. Datos disponibles	16
7. Metodología.....	18
7.1. Preprocesamiento	18
7.1.1. Preprocesamiento en SAC	18
7.1.2. Filtrado y transformación a coordenadas esféricas	19
7.1.3. Selección de <i>eventos</i> y <i>ruido</i>	20
7.1.4. Extracción de atributos de los sismogramas.....	21
7.1.5. Selección de atributos	22
7.2. Procesamiento	26
7.2.1. Conjuntos de entrenamiento y prueba	26
7.2.2. Entrenamiento del modelo.....	26

7.2.3. Evaluación del conjunto de datos.....	28
7.2.4. Métricas de rendimiento para el conjunto de entrenamiento	29
8. Resultados.....	31
8.1. Métricas de rendimiento para el conjunto de prueba.....	31
8.2. Elección de un score de corte.....	39
8.3. Figuras de los registros clasificados.....	39
9. Conclusiones	43
10. Anexos.....	46
10.1. Breviario sobre ondas sísmicas e instrumentación.....	46
10.2. ¿Qué es el ruido sísmico?.....	49
10.3. Métricas para la selección de atributos.....	51
10.3.1. Coeficiente de correlación entre dos variables continuas	51
10.3.2. Adaptación del coeficiente de correlación a una variable continua y otra nominal	52
10.3.3. Selección de variables por medio de teoría de la información.	52
10.4. Atributos extraídos de forma original.....	56
10.5. Selección preliminar de atributos.....	66
10.6. Sismogramas por trabajar	70
10.7. Sismogramas clasificados	93
Referencias.....	121

1. Resumen

Este documento presenta una técnica para discriminar ruido sísmico de eventos sísmicos, con base en un modelo de clasificación creado a partir del enfoque de la minería de datos. Bajo este enfoque se separa el modelo de clasificación particular de la evaluación de su rendimiento, y hasta cierto punto, de los registros sísmicos particulares que se quieren trabajar. El modelo de clasificación fue planteado para disminuir la cantidad de trabajo manual que debe hacer un sismólogo o intérprete al preprocesar registros sísmicos, de forma previa a una selección minuciosa; pero puede aplicarse como una técnica automatizada o asistir al intérprete al momento de realizar una selección pormenorizada de eventos o ruido sísmico.

Términos clave: evento sísmico, ruido, separación, clasificación, minería de datos.

2. Abstract

This document proposes a classification model for the discrimination of seismic noise from seismic events, with a background on data mining techniques. The procedures taken from this background allow for the separation of the classification model from the evaluation of its performance, and to some extent, from the seismic records to be classified. Although the classification model was designed to reduce the manual workload of a seismologist or interpreter while pre-processing seismic records before a thorough selection, it could also be used as a fully automated technique, or as an aid for the interpreter while performing an in-depth selection of seismic events or noise.

Keywords: seismic event, noise, classification, data mining.

3. Introducción

Dentro de la Sismología y áreas afines, como la Vulcanología, existen métodos que trabajan con eventos sísmicos, como temblores tectónicos o volcánicos; explosiones, de origen artificial o natural; y otros fenómenos controlados o conocidos. Otros métodos necesitan que se les proporcione sólo el ruido sísmico, sin la interferencia de los eventos antes mencionados. Todos estos fenómenos quedan registrados en la misma señal, o sismograma. Para obtener un contexto general de lo que es un sismograma, y la diferencia entre los eventos y el ruido sísmico, se recomienda revisar las secciones de Breviario sobre ondas sísmicas e instrumentación y ¿Qué es el ruido sísmico?, ambas en los Anexos.

Separar los eventos del ruido a partir del sismograma es necesario para respetar las suposiciones en las que se basan los métodos mencionados en el párrafo anterior. Sin embargo, bajo condiciones que no sean óptimas, esta separación se vuelve un problema difícil de resolver con las técnicas básicas automatizadas ya establecidas (por ejemplo, el umbral de amplitud, el *STA/LTA*, u otros métodos de reconocimiento de forma de onda).

La alternativa más inmediata a estas técnicas básicas es una separación manual de los eventos del ruido; pero esta actividad es tediosa, tardada, dependiente del criterio profesional del intérprete que trabaje los sismogramas, e imposible de implementar cuando se tienen muchos sismogramas por procesar. Por ello, se buscan técnicas automáticas de discriminación, que requieran la menor intervención posible y tengan un buen rendimiento, separando de la forma más pura posible a los eventos del ruido.

En el transcurso de este trabajo, se propone una técnica para lograr la separación comentada, que, si bien no se aparta por completo del trabajo manual, minimiza la intervención humana al tratar de replicar una clasificación proporcionada para el conjunto de sismogramas a trabajar. Además, la pureza de la separación puede controlarse de una forma sencilla; eligiendo, con base en unas métricas de rendimiento expuestas, un valor a partir del cual se declara a una señal como un evento sísmico.

4. Generalidades del modelo propuesto

4.1. Introducción a un modelo de clasificación

4.1.1. Planteamiento general de un problema y modelo de clasificación

La técnica propuesta en este trabajo es lo que en minería de datos se conoce como un *modelo de clasificación*. En términos generales, los modelos de clasificación toman unas variables de

entrada, conocidas como *atributos* o *variables de soporte*; y entregan una respuesta continua en un rango acotado, generalmente de cero a uno. A partir de esta respuesta, e introduciendo un umbral para su valor máximo, se puede obtener un valor nominal de salida para la combinación de atributos dados como entrada.

Los atributos son específicos al problema; por ejemplo, para el área de la Sismología serán características extraídas de un sismograma. En otras áreas, como en registros de pozo, podrían tomarse, entre otras, las características de permeabilidad y conductividad; o en ciencias atmosféricas las precipitaciones de lluvia en una zona.

Los problemas de clasificación más comunes son aquellos de clasificación binaria, donde el valor nominal que entrega el modelo puede tomar uno de dos valores (por ejemplo: *No/Sí*, *0/1*, *negativo/positivo*; o para los ejemplos de atributos mencionados en el párrafo anterior: *pozo no productor/pozo productor*, *no hay inundación aguas abajo/hay inundación aguas abajo*), y la problemática de este trabajo se puede plantear como uno de ellos. El resultado final de la técnica de clasificación presentada es una respuesta con valor de *ruido* o *evento*. Introduciendo lo expuesto en el párrafo anterior, y para la problemática de esta tesis, los sismogramas serán procesados de la siguiente manera: cada sismograma se partirá en ventanas, y a cada ventana se le extraerán ciertos atributos. La clasificación en *evento* o *ruido* se entregará para cada ventana; por lo que nos referiremos a ellas bajo el término genérico de *casos*. Los atributos de cada caso ingresan al modelo, el cual entrega un valor, o *score*, para cada una de las *clases*; por exclusión, un caso sólo puede tener *clase: evento*, o *ruido*. Esta situación se aprovecha para acotar el *score* que entrega el modelo para cada *clase* a un valor de cero a uno. Como consecuencia, la suma de los *scores* de las (dos) *clases* da uno. Al tratarse de un problema de clasificación binario, puede trabajarse únicamente con el *score* de una de las *clases*, por lo general la de interés, que puede ser *positivo*, *1*, o para este trabajo, *evento*.

4.1.2. El *score de corte*

Intuitivamente, podría asignarse el caso a la *clase* que cuenta con el *score* más alto, sea *ruido* o *evento*. Sin embargo, es más sofisticado introducir un umbral para cada *clase*, constante para todos los casos. Recordando que este problema puede manejarse sólo con el *score* de la *clase evento*, a dicho umbral se le denomina *score de corte* (para la *clase evento*). Todos los casos con *score* para la *clase evento* mayor o igual al *score de corte* se asignan a la *clase evento*. Con el *score de corte* se controla el rendimiento del modelo; el cual se cuantifica con las métricas expuestas en la sección de Resultados.

En estos momentos puede uno preguntarse, ¿cómo se sabe si el modelo está clasificando de manera correcta? Para saberlo, se vuelve necesario comparar los resultados del modelo con otra respuesta de referencia, considerada la “verdad absoluta”, y que es obtenida por cualquier otro método para el mismo problema. El otro método, generalmente, es una clasificación manual de los casos, o uno dependiente de información no disponible al momento de plantear el problema. Para este problema, donde no se dispone de más información, la “verdad absoluta” de la clasificación de cada caso se toma de una clasificación manual para cada ventana, en *evento sísmico* o *ruido*, elaborada por el sustentante. El valor de la respuesta de referencia es lo que se denomina *clase*. Por otra parte, a la respuesta del modelo se le denomina *clasificación*. Es muy importante tener presente esta diferencia.

4.1.3. La matriz de confusión y su relación con el score de corte

Una vez que se tiene una variable de *clase*, se puede comparar la clasificación del modelo con ella, para determinar si los casos fueron bien clasificados o no. Esta relación se expone en la forma de una *matriz de confusión*, expuesta en la Tabla 1, adaptada de (Davis & Goadrich, 2006). La *clase* es proporcionada de forma manual; mientras que la clasificación corresponde a la proporcionada por el modelo. Al tener dos valores tanto en la *clase* como en la clasificación del modelo, la matriz de confusión es de cuatro elementos.

Tabla 1: Matriz de confusión entre un modelo de clasificación y una clasificación de referencia.

	Clase: Positivo	Clase: Negativo
Clasificado por el modelo como Positivo	TP: Verdadero Positivo	FP: Falso Positivo
Clasificado por el modelo como Negativo	FN: Falso Negativo	TN: Verdadero Negativo

Considerando que la *clase* de interés es *evento*, podemos dar nombre y significado a los elementos de la matriz. Los casos que el modelo marcó como *eventos*, y que tienen un valor de *clase* de *evento* son los verdaderos positivos, o TP por las siglas en inglés para *True-Positive*. De forma similar, los casos que tanto el modelo como la clase marcaron como *ruido*, serán los verdaderos negativos, o TN (de *True-Negative*). Los otros dos elementos provienen de la confusión entre ambos valores: Los falsos positivos (FP, *False-Positive*) son aquellos que el modelo marcó como *evento*, pero la *clase* registra como *ruido*; y los falsos negativos (FN, *False-Negative*) los que el modelo dice que son *ruido*, pero tienen un valor de la *clase* de *evento*.

La clasificación de los casos en *evento* o *ruido* depende del *score de corte*. Por esta razón, la matriz de confusión y las métricas de rendimiento que se desprenden de ella son una función del

score de corte. De esta manera el *score de corte* controla el rendimiento del modelo, lo cual permite enfocar los resultados hacia la clasificación o discriminación de una *clase* en particular.

4.1.4. Conjuntos de datos de entrenamiento y prueba

Si se clasifican los datos manualmente para tener la “verdad absoluta” en la *clase* para cada caso, puede cuestionarse la necesidad de plantear un modelo de clasificación; ya que la clasificación manual ya es una clasificación de los datos. Sin embargo, si se presentan datos nuevos sin clasificar, podríamos aplicarles un modelo ya hecho y con un rendimiento garantizado, sin repetir el planteamiento y trabajo necesarios para proponer un modelo de clasificación. Para ello, tomamos todos los casos con *clase* que tenemos a nuestra disposición, y los separamos en dos conjuntos: el *conjunto de casos de entrenamiento*, y el *conjunto de prueba*. Los datos de entrenamiento se usan para entrenar al modelo de clasificación, al cual se le proporcionan junto con la *clase* correspondiente a cada caso.

El proceso de entrenamiento permite al clasificador identificar o reconocer alguna estructura o patrón en los atributos que le indique cuál es la *clase* correspondiente a cada caso. Por otra parte, el conjunto de casos de prueba es el que se usa para validar el rendimiento del modelo. Antes de clasificar los casos del conjunto de prueba, se les retira la *clase*, y sólo se entregan los atributos al modelo. El clasificador entrega el *score* para cada *clase*, que se compara con la clasificación original de los datos (la “verdad absoluta”). A partir de la comparación se obtienen las métricas de rendimiento, en función del *score de corte*. Al momento de evaluar un nuevo conjunto de datos sin clasificar, completamente desconocido por el modelo, podemos dar como garantía las métricas de rendimiento para el conjunto de prueba.

Otra razón para segmentar los datos en conjuntos de entrenamiento y de prueba es que los modelos pueden ofrecer un rendimiento excelente al clasificar el conjunto de entrenamiento, mientras que el rendimiento para el conjunto de prueba es terrible. A este fenómeno se le conoce como *sobreentrenamiento* del modelo, y resulta de que el modelo aprende a reconocer perfectamente los casos del conjunto de entrenamiento. Esto se da por tomar un conjunto con muy pocos casos, dar demasiadas iteraciones al algoritmo de entrenamiento del modelo, o formar un conjunto de entrenamiento con muy poca variedad en sus atributos o *clases*. Afortunadamente, este fenómeno es fácil de reconocer haciendo uso de uno o más conjuntos de prueba.

Para mitigar las causas del *sobreentrenamiento* hay varias estrategias, orientadas a modificar la forma de repartir los datos en ambos conjuntos, pero su discusión se sale del enfoque de este trabajo. Basta mencionar que la forma más sencilla consiste en tomar un porcentaje arbitrario de

los casos para formar el conjunto de entrenamiento, constituyendo el resto al conjunto de prueba; aunque con este procedimiento podríamos toparnos con los problemas mencionados.

4.1.5. El modelo de clasificación como una “caja negra” intercambiable

Hasta este momento se ha tratado al modelo de clasificación como una caja negra a la que entran los atributos de un caso, y sale un *score* para el caso; sin entrar en detalles de lo que sucede dentro. Esto se debe a que esa caja negra puede ser cualquier función, esquema, o artefacto que reciba los valores de los atributos, y entregue un valor del cero al uno para cada *clase*. Podemos elegir cualquiera de estos entes para formar un modelo de clasificación, cambiando la forma en la que se calcula la respuesta, pero conservando los mismos atributos y conjuntos de entrenamiento y prueba. Con ello, podemos encontrar al modelo que resulte más satisfactorio de entre los modelos que probemos. Sin embargo, la discusión de la búsqueda y elección del mejor modelo, o uno lo suficientemente bueno se sale del enfoque de este trabajo.

Al momento de elegir un modelo deben de considerarse factores más allá de su rendimiento, tales como el tipo de dato de las variables y su distribución estadística, la facilidad con la que se puede implementar, el consumo de recursos, o incluso la facilidad con la que se puede explicar.

4.2. Propuesta de un modelo de clasificación basado en la distribución normal multivariada.

Una vez que se consideraron los factores intrínsecos al problema, en particular la distribución de los atributos, se formó un modelo a partir de dos distribuciones normales multivariadas, con una para cada *clase* (*ruido* o *evento*). Por estos motivos, se introduce esta distribución a continuación. En la sección de Selección de atributos, dentro de la Metodología se detallan los atributos seleccionados para el modelo de clasificación; mientras que en la sección de Confección del modelo a partir de la distribución normal multivariada, más adelante, se exponen las razones por las que se eligió al modelo de clasificación presentado en este trabajo.

4.2.1. La distribución normal multivariada

Se dice que un grupo de p variables continuas x_i , acomodadas en un vector de variables x para un conjunto de n casos, tiene una distribución normal multivariada cuando su función de densidad de probabilidad es descrita por la expresión [1] (Peña, 2002, pág. 264).

$$f(x) = \frac{1}{\sqrt{|V|}^p \sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^T V^{-1}(x-\mu)} \quad [1]$$

En la expresión [1], $\boldsymbol{\mu}$ es el vector de valores medios y \mathbf{V} es la matriz de covarianza para las variables x_i . $|\mathbf{V}|$ es el determinante de \mathbf{V} , y \mathbf{V}^{-1} su matriz inversa. El vector $(\mathbf{x} - \boldsymbol{\mu})$ resulta de sustraer el vector de valores medios a un caso \mathbf{x} ; mientras que $(\mathbf{x} - \boldsymbol{\mu})'$ es el transpuesto de este vector.

La expresión [1] puede compararse con la función de densidad de una distribución normal univariante, y se observará que son esencialmente la misma función. A la expresión se le pueden distinguir dos partes: el valor constante que multiplica a la exponencial, y el exponente de la exponencial. El valor constante actúa como un factor de escala para acotar el valor máximo de la exponencial, mientras que el exponente toma un valor de cero cuando $\mathbf{x} = \boldsymbol{\mu}$, y, dado el signo negativo, disminuye según aumenta la diferencia entre \mathbf{x} y $\boldsymbol{\mu}$. Al tomar $(\mathbf{x} - \boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{x} - \boldsymbol{\mu})$ se condensa la diferencia en un escalar, conocido como *distancia de Mahalanobis*. De lo anterior se intuye que el valor máximo de $f(\mathbf{x})$ se obtiene cuando $\mathbf{x} = \boldsymbol{\mu}$; así como $f(\mathbf{x})$ toma valores asintóticos a cero cuando crece la diferencia $(\mathbf{x} - \boldsymbol{\mu})$.

4.2.2. Confección del modelo a partir de la distribución normal multivariada

Al momento de hacer un análisis exploratorio de los datos como parte del preprocesamiento (detalles en la sección Selección de atributos), se observó que algunas de las variables disponibles mostraban en sus histogramas una forma acampanada, semejante a una curva gaussiana. Además, al introducir la información de la *clase* como un color dentro del histograma, se notaba que los casos de la *clase ruido* contaban con una menor dispersión alrededor de la media del histograma; mientras que los casos de la *clase evento* podían alcanzar valores extremos, bastante alejados del centro de la campana, mucho más lejos que los de la *clase ruido*. En otras palabras, considerando a cada atributo dentro de una distribución normal unidimensional, los casos etiquetados como *ruido* tienen una menor desviación estándar que los de la *clase evento*. Por otra parte, el valor medio de ambas *clases* es muy similar, lo que resulta en que los valores cerca del valor medio correspondan en su mayoría a casos de *ruido*, mientras que aquellos alejados del valor medio son casi todos *eventos*.

Dada esta situación, se optó por ajustar una distribución normal multivariada, con la forma funcional de la expresión [1] a cada *clase*. Para formar un *score* a partir de los valores de $f(\mathbf{x})$ se aplicaron las normalizaciones expuestas en las expresiones [2] y [3].

$$s_0 = \frac{\pi_0 f_0(\mathbf{x})}{\pi_0 f_0(\mathbf{x}) + \pi_1 f_1(\mathbf{x})} \quad [2]$$

$$s_1 = \frac{\pi_1 f_1(\mathbf{x})}{\pi_0 f_0(\mathbf{x}) + \pi_1 f_1(\mathbf{x})} \quad [3]$$

En ellas, $f_0(\mathbf{x})$ corresponde a la densidad de probabilidad para la distribución ajustada a los casos de *ruido*, y $f_1(\mathbf{x})$ es el equivalente para los casos de *evento*. Los valores de π_0 y π_1 corresponden a la probabilidad a priori de que un caso sea de la *clase* correspondiente, y se obtienen contando los casos de cada *clase* y dividiendo entre n . Estos factores le dan una mejora marginal, pero cuantificable al rendimiento del modelo.

Por último, s_0 es el *score* para la *clase ruido* otorgado por el modelo al caso, y s_1 es el de la *clase evento*. Por inspección se observa que $s_0 + s_1 = 1$, por lo que podremos trabajar sólo con el *score* para la *clase evento*, s_1 , sabiendo que el de la otra *clase* es su complemento para uno.

4.2.3. Compensación numérica

Recordando que para ambas *clases* los valores de las medias son similares, puede notarse que en los casos que caigan muy lejos del valor medio de cualquiera de las distribuciones resultarán valores de $f(\mathbf{x})$ próximos a cero. Esto se vuelve un problema ya que el denominador de [2] y [3], $\pi_0 f_0(\mathbf{x}) + \pi_1 f_1(\mathbf{x})$, puede hacerse más pequeño que la magnitud más chica que puede representar la computadora donde se implementó el modelo. Para compensar esta situación, se asignó directamente a la *clase evento* los casos con un valor de $\pi_0 f_0(\mathbf{x}) + \pi_1 f_1(\mathbf{x}) = 0_{CPU}$, donde el 0_{CPU} es la magnitud más pequeña que puede representar la computadora. Empíricamente, se observó que $0_{CPU} < 10^{-60}$.

5. Antecedentes

Con la proliferación de instrumentos más económicos se volvió factible colocar más sismómetros para densificar redes, ampliar las ya existentes; y tomar cada vez más datos de lugares no considerados con anterioridad. Como consecuencia, se da una acumulación de datos que no han sido analizados, al agregarse más rápido de lo que es posible procesarlos. De no aprovecharse esos datos, el esfuerzo dedicado a su captura habrá sido en vano.

Una buena parte del trabajo de un sismólogo es extraer las señales de interés del sismograma, y para manejar la cantidad de datos presente en la actualidad es necesario automatizar este proceso. Además, la automatización acerca el trabajo del sismólogo a un proceso en tiempo real; cuyos resultados son de interés para los sistemas de alertamiento sísmico. Por otra parte, como menciona (Ross & Ben-Zion, 2014, pág. 368), las “inspecciones y correcciones de un analista pueden

refinar los resultados automáticos, pero no son factibles con los grandes conjuntos de datos de hoy en día, y son menos objetivas si se elaboran solas [sólo por el analista]”. En otras palabras, los métodos automáticos clásicos o sencillos, por ejemplo, *STA/LTA*, no tienen un rendimiento lo suficientemente bueno y es necesario un supervisor humano para obtener buenos resultados. Estas deficiencias no se limitan sólo a la Sismología; por ejemplo, en Vulcanología los eventos (de origen volcánico) pueden ser de baja frecuencia, con un comienzo emergente no ser evidente. Bajo esas condiciones, donde el inicio de la señal no es claro, los métodos más sencillos no funcionan.

Por lo general, este déficit se observa al trabajar datos poco óptimos tomados en condiciones desfavorables (más ruido o menos estaciones); para los cuales se requiere la asistencia de un intérprete humano, al cual no le alcanzará el tiempo para trabajar todo el conjunto de datos de forma manual. Es necesario complementar las carencias del intérprete humano y las de las técnicas automáticas para poder analizar los datos de forma eficiente.

En el transcurso de la siguiente sección se analiza una selección de cuatro documentos que abarcan avances y propuestas representativas para dar una solución a esta problemática. En general, todos los métodos presentados, excluyendo al *STA/LTA*, tienen la estructura presentada a continuación:

1. Retirar media de la señal
2. Aplicar filtros pasa-banda a la señal, excluyendo a la banda de frecuencias de mayor ruido, ubicada alrededor de algunos pocos Hertz (En la sección de los anexos: ¿Qué es el ruido sísmico?, se comenta al respecto de esta banda).
3. A cada banda se le aplican transformaciones y se le extraen o miden ciertas características.
4. Las características extraídas entran a un modelo de clasificación, u otra técnica para discriminar ruido de eventos.
5. El método se compara contra una técnica establecida, como el *STA/LTA*; o contra un intérprete humano. El intérprete humano se sigue considerando el mejor estándar, a falta de la “verdad absoluta”.

Sin embargo, es al momento de comparar el rendimiento entre dos métodos cuando surgen dificultades. Como expone (Gentili & Michelini, 2006, pág. 56) y se resume a continuación: ya que se usan diferentes conjuntos de datos en cada estudio o técnica presentada, una comparación directa no es posible. Además, no hay una forma estandarizada de probar el rendimiento de los métodos presentados. La cuestión se complica más cuando se consideran los siguientes factores:

- La discriminación entre ruido y eventos puede hacerse con énfasis en el instante de tiempo (por ejemplo, cuando llega una fase al instrumento; denominado *picking*, según *óp. cit.*), o con énfasis en una ventana de tiempo (denominado *labeling*, *ibíd.*). Comparar una de estas técnicas con otro método de referencia o intérprete exige diferentes metodologías, ya que en el *picking* es necesario dar la separación temporal entre el resultado del modelo y el de referencia; mientras que en el *labeling* el resultado de clasificar cada ventana sólo es correcto o incorrecto respecto a la referencia.
- No se expone a conciencia la manera en la que la proporción de señal a ruido afecta a los resultados del método; ni se menciona esta información para las señales usadas en el entrenamiento o prueba del modelo. Según *óp. cit.*, esto sobreestima el rendimiento de los métodos. Por si fuera poco, no hay una forma estandarizada de obtener esta proporción, si se llega a mencionar.
- Al momento de presentar el rendimiento del modelo, no se entregan elementos suficientes para construir la matriz de confusión, con lo cual no se puede formar un panorama completo del rendimiento del modelo. Para el caso de las técnicas de *picking*, no se distingue entre errores de clasificación entre las ondas P y S, con lo cual se sobreestima el rendimiento.
- El conjunto de prueba usado para medir el rendimiento puede ser muy pequeño, o no representativo de todas las señales analizadas, al tomarse de una o dos estaciones. Con ello, nuevamente se sobreestima el rendimiento del modelo.

La consecuencia de estas limitaciones es que por lo general la comparación entre dos propuestas de métodos de clasificación sólo puede hacerse de forma cualitativa (*óp. cit.*); al carecer de la información necesaria para una comparación exhaustiva, que incluso podría no ser aplicable.

A pesar de ello, al interior de una propuesta, donde varios de los factores mencionados son los mismos (mismos datos, mismo procesamiento, pero posiblemente diferentes métodos) sí es posible comparar los métodos cualitativamente, aunque sea contra un intérprete humano (por ejemplo, (Ross & Ben-Zion, 2014, pág. 372)).

A continuación, se comentan las técnicas en orden cronológico:

5.1. STA/LTA

El *STA/LTA*, detallado en (Trnkoczy, 1999) pero con orígenes anteriores, es generalmente considerado el estándar de referencia; y no sólo por su longevidad, ya que es un método comprensible y sencillo, que puede ofrecer un buen rendimiento en condiciones favorables.

En *óp. cit.*, pág. 2 se comenta que (al momento de su redacción) ya existían métodos más sofisticados, pero pocos de ellos se encontraban implementados de forma comercial al requerirse “un experto” para poder aprovecharlos al máximo y obtener un buen rendimiento con ellos. El funcionamiento del método es sencillo: se toma una señal sísmica, se filtra y convierte a valor absoluto. Posteriormente se extraen, muestra a muestra, dos ventanas: una larga y una corta.

La ventana corta “mide la amplitud <<instantánea>> de la señal sísmica” (*óp. cit.*, pág. 3); mientras la ventana larga considera el nivel promedio de la señal. Finalmente, se toma el promedio de cada ventana, y se obtiene la proporción de la ventana corta sobre la ventana larga. Este último paso da nombre al método: *Short-Term Average / Long-Term Average*.

Cuando esta proporción excede cierto umbral se declara un evento, y cuando decae por debajo de otro se declara su término. Hay muchos parámetros que se pueden ajustar dentro del método (consúltese *óp. cit.*), pero el funcionamiento se reduce al expuesto en el párrafo anterior.

5.2. Picado de ondas P y S por medio de redes neuronales

La siguiente propuesta representativa es el trabajo de (Gentili & Michelini, 2006). El método propuesto hace énfasis en *picking*, es decir, en encontrar el instante en el que arriban las ondas P y S a un instrumento; y para tal fin emplean redes neuronales.

Como mencionan (*óp. cit.*, p. 39), la bondad de las redes neuronales (y de los modelos de clasificación de minería de datos en general) es que “tienen la ventaja de adoptar esquemas de aprendizaje basados en los datos para encontrar la solución al problema. Por ello, no es necesario conocer a fondo la distribución estadística de las características [de los datos] para obtener una solución, aun cuando estas características sean redundantes o ruidosas”. Lo anterior implica que no es necesario buscar a tientas los mejores parámetros para el modelo, a diferencia del *STA/LTA* en (Trnkoczy, 1999).

La metodología de (Gentili & Michelini, 2006) parte de una señal de tres componentes sin media, a la cual se le aplican filtros en dos bandas: la primera de 2 a 50 Hertz, y la segunda de 2 a 8 Hz., para la detección de ondas P y S respectivamente. En palabras de los autores (*óp. cit.*, p. 40), el filtrado sirve para “retirar la señal ruidosa en frecuencias menos importantes”. Posteriormente cada banda se transforma a un sistema de coordenadas bidimensional: horizontal y vertical.

A partir de estas componentes se extraen características de varianza, sesgo, curtosis, sus derivadas temporales y una combinación de ellas; calculando las características muestra a muestra, tomando las muestras de dos ventanas de longitud fija, una para detectar ondas S y otra para las P. Estas características se dan como los atributos de entrada a la red neuronal.

En la técnica planteada en *óp. cit.* se ocupan dos redes neuronales, una para cada tipo de onda elástica (*óp. cit.*, pág. 44). Los autores resaltan que las redes obtenidas se componen de perceptrones sencillos, formando sólo una capa de neuronas. Por las razones expuestas en el apéndice de dicha propuesta (*óp. cit.*, pág. 60-61) los autores mencionan que el problema de clasificación que intentan resolver es linealmente separable; al menos en el espacio de los atributos que ellos obtuvieron de las señales sísmicas. Dado el comentario sobre la linealidad del problema, podría esperarse que los autores comentaran sobre una técnica de clasificación más sencilla que las redes neuronales. Sin embargo, los autores no hacen mayores comentarios al respecto.

En todo caso, a ambas redes neuronales entran los valores continuos de los atributos, calculados para varias muestras alrededor de un instante. Dentro de cada neurona se toman los valores proporcionados, y se calcula una combinación lineal de ellos. Sobre la combinación lineal se aplica una función logística, que da el valor de salida de la neurona (*óp. cit.*, p. 60). Estos valores de salida se proporcionan a otras dos neuronas, que calculan otra combinación lineal y le aplican la función logística. Finalmente, los valores de salida de las últimas dos neuronas se combinan (*óp. cit.*, pp. 44-45) para dar un *score* de salida. Dentro del mismo esquema se considera implícitamente la elección de un *score de corte* (*ibíd.*).

Tras obtener un *score*, los autores efectúan algunos pasos encaminados a mejorar el rendimiento del clasificador, pero dependen de tener señales de una red (de más de una estación), por lo cual no serán comentados aquí.

El rendimiento de la técnica propuesta por ellos se mide contrastando los resultados contra un picado manual de las fases en los sismogramas. Ya que el resultado del modelo es una clasificación, muestra a muestra, sobre el arribo de una fase P o S; los autores reportan sus resultados partiendo de la desviación con respecto al picado manual. Para marcar un picado de la red neuronal como un falso positivo, los autores requieren que difiera por una cantidad fija respecto al picado manual (*óp. cit.*, p. 48). Con ello, los autores presentan como métricas de rendimiento al *Precision* y al *Recall*, para proceder a analizar el rendimiento en función de la proporción de señal a ruido.

Para tener un punto de comparación, los autores repiten esta evaluación con un clasificador de *STA/LTA* (*óp. cit.*, p. 51-52); a partir de lo cual llegan a la conclusión de que la red neuronal tiene un mejor desempeño que el *STA/LTA* cuando la proporción de señal a ruido es baja. A ésta se le añaden otras conclusiones sobre la generalidad del método, que lo hace fácilmente aplicable a situaciones de análisis en tiempo real o monitoreo sísmico, entre otras (*óp. cit.*, p. 60).

5.3. Picado de ondas P, S, y FZHW con una técnica particular

El tercer documento analizado, (Ross & Ben-Zion, 2014); plantea un problema de clasificación entre arribos de ondas P, S, y un tipo especial de onda que se genera al refractarse una onda P a lo largo de una interfaz entre dos materiales con un alto contraste de velocidad. A éstas últimas se les denomina en inglés *Fault Zone Head Waves*. Dadas las propiedades de las *FZHW*, los métodos de picado automático convencionales las llegan a confundir con el primer arribo de las ondas P en las redes sísmicas ubicadas cerca de fallas con estas características (por ejemplo, el segmento de Parkfield de la falla de San Andrés (*óp. cit.*)). Esta confusión lleva a los autores a proponer un modelo de clasificación, con énfasis en *picking*, para los tres tipos de ondas.

Para detectar los tres tipos de onda, los autores parten de sismogramas de tres componentes, a los cuales se les sustrajo su media. Como es de esperarse, le sigue la aplicación de un filtro pasa-banda, que los autores tomaron de 0.5 a 30 Hz (*óp. cit.*, p. 369). Aprovechando las características de las ondas P, S, y *FZHW* dentro de la red de sismómetros, los autores calculan parámetros relativos a la polarización de las señales, y con ellos se modulan las tres componentes de los sismogramas, muestra a muestra (*óp. cit.*, p. 370). Según los autores, se espera que con esta modulación se mitigue la aportación de las ondas con proyección en más de una de las componentes, y que a partir de ello se puedan distinguir mejor los diferentes tipos de onda.

Una localización preliminar de los picados, para todos los tipos de onda, se realiza bajo un algoritmo de *STA/LTA*, pero la ubicación precisa del picado se refina al obtener el sesgo y curtosis de la señal, y sus derivadas. Los autores indican que estos valores responden mejor “a cambios abruptos en el carácter de la señal” (*óp. cit.*, p. 372), y a partir de ellos se puede discriminar una *FZHW*, de carácter emergente, de una onda P directa. Para tener una mayor robusticidad, los autores también consideran el tiempo mínimo que habría de existir entre los arribos de ambos tipos de ondas; lo cual se hace determinando intervalos temporales a partir de las cuales es factible observar un arribo después de otro.

Las pruebas del método se hacen tanto con sismogramas sintéticos, como con datos tomados de una red en las proximidades de la falla de San Andrés en Parkfield, California. Al tener énfasis en

el *picking*, los resultados se reportan en función de la diferencia en tiempo con respecto a un picado manual. Para finalizar, en la sección de discusión, los autores comentan sobre la posibilidad de ocupar el método en otras aplicaciones, con un mínimo de ajustes a sus parámetros (*óp. cit.*, p. 379), los cuales deben de ajustarse para mejorar el rendimiento, en términos de la cantidad de arribos detectados.

Resulta interesante que, según los autores, no sea necesario trabajar las tres componentes en conjunto, ya que al hacerlo no se obtienen resultados significativamente mejores (*óp. cit.*, p. 377). Sin embargo, puede argumentarse que al momento de calcular la matriz de covarianza para el análisis de polarización (*óp. cit.*, p. 370), se involucra la información de todas las componentes. Por último, los autores mencionan la posibilidad de asociar el modelo con otras fuentes de información, como la detección de un arribo a lo ancho de una red; para decidir sobre el resultado del método (*óp. cit.*, p. 379).

5.4. Una comparativa de técnicas basadas en la correlación cruzada, con una técnica proveniente del procesamiento de audio

Por último, se discute la propuesta de (Yoon, O'Reilly, Bergen, & Beroza, 2015). Para tener un contexto, los autores parten de la técnica de identificar sismos dentro de registros a través de la correlación cruzada, ya sea con una señal de referencia o con el mismo registro como referencia.

El método de los autores busca solventar las deficiencias de la búsqueda con correlación cruzada, debidas a la necesidad de tener una señal de referencia para detectar señales similares. Esta referencia puede ser desconocida, ya que en ocasiones no se sabe lo que se está buscando en los registros (*óp. cit.*, p. 2). Por otro lado, en el caso de una autocorrelación, la señal de referencia se toma de entre las mismas señales a procesar; pero hay que operar a todos los registros entre sí (*ibíd.*), lo cual no resulta práctico para procesar una cantidad masiva de registros.

Ambos métodos aprovechan que los registros procedentes de una fuente sísmica tienen formas similares y son cuasiinvariantes con el tiempo (*óp. cit.*, p. 1); y también son comparados con el *STA/LTA*, el cual tiene la ventaja de ser rápido y fácil de aplicar, pero falla al intentar encontrar señales en condiciones poco óptimas (*ibíd.*).

Para obtener un método que tenga las ventajas de los tres, los autores adaptan un algoritmo diseñado para comparar segmentos de audio entre sí e identificar a los más semejantes (*óp. cit.*, p. 2-3), el cual a su vez usa una técnica de *hashing* (las cuales ya son conocidas en bases de datos y programación desde antaño) para evitar comparar a todos los registros entre sí, y sólo efectuar la comparación entre unos pocos candidatos (*óp. cit.*, p. 2).

La técnica propuesta parte de un canal horizontal (de una de las tres componentes) de una estación, al cual se le aplica un filtrado pasa-banda de 4 a 10 Hertz, y se le diezma en el tiempo (*óp. cit.*, p. 3). Posteriormente, el método extrae ventanas empalmadas de los registros, y de cada ventana se obtiene un espectrograma. El espectrograma se proyecta sobre ondículas de Haar; y se seleccionan los coeficientes de las ondículas más representativas, es decir, que difieran más de los coeficientes del ruido. A los coeficientes seleccionados se les extrae el signo, y el resto se reemplaza con cero.

Al espectro de una ventana, transformado hasta este punto, es al que se le aplica una función *hash*, y al resultado de la función, denominado *hash* (de la ventana), se le acomoda en una *tabla de hash*, la cual organiza elementos similares en sus propios conjuntos o *hash buckets* (*óp. cit.*, p. 9-10). Para identificar los espectros, y por lo tanto las señales similares entre sí, sólo hace falta analizar los contenidos de cada *hash bucket*, y calcular una métrica de semejanza (*óp. cit.*, p. 10); sin tener que realizar la búsqueda exhaustiva de todos contra todos como en la autocorrelación. Al trabajar con ventanas, puede decirse que tanto el método propuesto, como las dos variantes de la correlación, hacen énfasis en el *labeling*.

Los pasos tomados por el método desde el preprocesamiento van encaminados a simplificar la representación de una señal sísmica, y culminan con el empleo de la técnica de *hashing* para facilitar la búsqueda de señales semejantes. Al simplificar las señales, y por el uso de una función *hash*, las señales marcadas como semejantes por el método no necesariamente lo son, y requieren de pasos adicionales para declararse como un evento detectado (*óp. cit.*, p. 6); entre los cuales los autores mencionan a los métodos de clasificación entre sismos y ruido (*óp. cit.*, p. 7).

Con lo anterior se infiere que una de las desventajas del método es que las comparaciones que realiza sólo son por pares, aunque este problema podría tratarse con otras técnicas de posprocesamiento (*ibíd.*). Otros inconvenientes provienen del diseño del método, ya que la propuesta particular de los autores sólo se limita a indicar pares de señales semejantes, sin llegar a un modelo de clasificación. Por ello, si hay ruido correlacionado entre dos de los registros, éstos se llegarán a identificar como señales semejantes, aún sin haber sido originados por un sismo (*ibíd.*).

Sin embargo, la gran ventaja de la propuesta es la rapidez (143 veces más rápido que la autocorrelación) con la que se ejecuta el método, aún sin optimizarse para el tiempo de ejecución (*óp. cit.*, p. 5). Al usar una *tabla de hash*, el método se escala de forma casi lineal en tiempo de ejecución con respecto a la cantidad de datos, lo cual lo hace práctico para procesar conjuntos masivos de registros (*óp. cit.*, p. 8). Esta misma rapidez lo haría propicio para aplicaciones de

detección en tiempo real (*óp. cit.*, p. 9). Los autores también mencionan la posibilidad de incorporar información de más canales y estaciones al método, para aumentar su robusticidad (*ibíd.*).

5.5. Comentarios sobre los antecedentes

Los documentos analizados en esta sección son de carácter representativo para el problema a tratar en este trabajo. Ninguna de las técnicas se hizo con un énfasis particular en la minería de datos; sin embargo, aspectos como la elección de un *score de corte*, la matriz de confusión, y las métricas de rendimiento entre otros, son mencionados en el desarrollo de los trabajos expuestos. También se nota que para dar solución al problema de clasificación se han tomado una gran variedad de enfoques, tanto en el planteamiento, como en la técnica de clasificación; abarcando operaciones tanto sencillas como complejas; con pocos y muchos parámetros de ajuste (longitud de ventanas, empalme, filtrado, etcétera).

A partir de lo anterior se intuye que no hay una técnica ideal para resolver todos los problemas de separación de ruido y eventos. De una forma similar, aunque no es mencionado explícitamente, si se desean aplicar las técnicas examinadas a un conjunto de datos distinto al que se usó en su desarrollo, será necesario volver a ajustar los parámetros operativos de la técnica; a lo que en minería de datos se le denomina *(re)entrenar el modelo*.

No debería de sorprendernos que las técnicas desarrolladas van adquiriendo complejidad con el paso del tiempo. Esto se debe al aumento de la disponibilidad (y disminución del costo) del poder de cómputo, gracias a la miniaturización de los componentes electrónicos y a la manufactura a gran escala. Como era de esperarse, este fenómeno también se encuentra detrás de la proliferación de datos en el área de la Sismología.

6. Datos disponibles

Los datos trabajados se tomaron a raíz del colapso de una cavidad en la Colonia Golondrinas, en la delegación Álvaro Obregón, Distrito Federal, México; a finales de 2006. La cavidad se localizó dentro de una zona urbanizada, por lo que al abrirse la cavidad colapsó una residencia. Esto motivó la toma de datos geofísicos por medio de varias técnicas, entre ellas la recopilación de registros sísmicos, con la finalidad de monitorear cualquier movimiento diferencial o actividad posterior al colapso en el área.

El instrumento empleado fue un Güralp 6TD de tres componentes (E, N, Z) que incluye sismómetro, digitalizador y registrador. La respuesta instrumental del sistema es desconocida, ya que dicho instrumento se encontraba en préstamo a la Facultad de Ingeniería de la Universidad

Nacional Autónoma de México. El instrumento se colocó en las proximidades de la cavidad, aproximadamente a veinte metros, en el punto WGS84 19.3782802° N, 99.2148679° W; el cual fue denominado “Estación Golondrinas” (GOLO).

Los registros capturados constan de poco más de 352 horas de registros, no contiguos, abarcando del 14 de octubre al 21 noviembre de 2006. La lista de estos registros, así como sus sismogramas se incluyen en la sección de Sismogramas por trabajar en los Anexos. Al ubicarse dentro de una urbanización, los registros se encuentran altamente contaminados con eventos transitorios de origen antropogénico; que suceden con mayor frecuencia entre las 06:00 y las 23:59 en hora local. El intervalo entre las 00:00 y las 06:00 cuenta con una menor amplitud en los registros y una menor cantidad de eventos transitorios.

Un ejemplo de estos fenómenos o eventos antropogénicos transitorios es la demolición de la residencia colapsada, la que se llevó a cabo mientras se tomaban los registros; como se observa en el registro *GOLO 323*, mostrado en la Figura 46 de la sección de Sismogramas por trabajar en los anexos. En este registro, el intervalo de los 55,000 a los 70,000 segundos muestra las percusiones por la demolición de la residencia colapsada; mientras que el intervalo de 25,000 a 30,000 segundos se ve afectado por un sismo de subducción al oeste del país (comunicación personal con el Dr. Martín Cárdenas Soto, 21 de enero de 2019).

Otros análisis de los registros previos a la elaboración de este trabajo no revelaron evidencia de algún colapso durante el periodo de registro; lo cual concuerda con los testimonios de los vecinos, quienes no reportaron movimientos diferenciales o sonidos originados en el subsuelo dentro del mismo periodo (*ibíd.*).

Por otra parte, los registros cuentan con algunas fallas instrumentales, durante las cuales al menos una de las componentes permanece en un valor constante o con una tendencia lineal pura, sin indicar algún otro tipo de movimiento. Estas fallas instrumentales se debieron a la interrupción de la alimentación al instrumento, porque alguien movió al instrumento, o por que empezó a derivar la masa del sismómetro.

En el transcurso de las actividades preliminares a este trabajo se observó que dichas fallas instrumentales eran fáciles de identificar, tanto de forma manual como empleando técnicas de minería de datos. Aún más, los modelos de clasificación se sesgaban a la identificación de las fallas instrumentales como *eventos*, indicando al resto del registro como *ruido*. Por estas razones, los segmentos con fallas instrumentales fueron retirados por completo de todos los registros, como primer paso del preprocesamiento. La depuración se realizó de forma manual como parte del

preprocesamiento, pero se comenta aquí ya que los segmentos retirados no llegaron a formar parte de ningún otro paso en el desarrollo del trabajo.

7. Metodología

7.1. Preprocesamiento

7.1.1. Preprocesamiento en SAC

Como se anticipa desde la sección anterior, el primer paso del preprocesamiento de los registros sísmicos fue el retiro de los segmentos con fallas instrumentales. Al estar los registros en el formato del programa Seismic Analysis Code (Goldstein & Snoke, 2005); se aprovecharon las facultades del programa para realizar todos los pasos del preprocesamiento. Estos pasos consistieron en las siguientes operaciones, una vez que se leyó cada registro al espacio de trabajo de SAC:

1. Sincronizar las tres componentes del sismograma a un comienzo común en el tiempo.
2. Delimitar el segmento de la señal que se va a conservar, usando las utilidades de visualización y picado propios de SAC.
3. Recortar el sismograma al segmento seleccionado.
4. Retirar valor medio del segmento seleccionado.
5. Retirar tendencia lineal del segmento seleccionado.
6. Escribir el segmento seleccionado a un nuevo archivo SAC.

En caso de que el registro tuviera una falla instrumental en alguna de las componentes (por ejemplo, *GOLO 315* mostrado en la Figura 34, en el tramo de 50,000 a 70,000 segundos); el segmento con la falla, en sus tres componentes, se excluía de la selección y de cualquier análisis posterior. Si el registro se encontraba interrumpido por la falla instrumental, el procedimiento mencionado se repetía para los segmentos a conservar a ambos lados de la falla instrumental. Las causas de estas fallas se han comentado en la sección de Datos disponibles.

Para facilitar el procesamiento posterior, particularmente en la clasificación manual de las ventanas en *ruido* o *evento*, se fragmentó a la mayoría de los registros en tramos de una longitud manejable. Todos los comandos de SAC fueron ejecutados de forma manual, para todos los sismogramas disponibles. Estos se muestran en sus versiones originales y preprocesadas en la sección de Sismogramas por trabajar, en los Anexos. En la misma sección también se da su fecha y hora de inicio, así como la duración de cada uno.

Ya que se desconoce la respuesta instrumental del sismómetro empleado, no fue posible aplicar la corrección correspondiente a los sismogramas. Por tal razón, los sismogramas se conservaron en las unidades arbitrarias al instrumento, denominadas *cuentas*. Los atributos calculados a partir de los sismogramas conservan unidades derivadas de las cuentas.

7.1.2. Filtrado y transformación a coordenadas esféricas

Cuando se plantearon modelos de clasificación preliminares, primero se experimentó con extraer las características de los sismogramas a partir de sus componentes cartesianas, sin filtrar. Sin embargo, se notó que los modelos tenían un mejor rendimiento al añadir dos pasos al preprocesamiento: filtrar el sismograma en varias bandas de frecuencia, y transformar las componentes filtradas a coordenadas esféricas.

El filtrado de las componentes cartesianas fue el primer paso en aplicarse. Este se llevó a cabo en Matlab (The MathWorks, Inc., 2012), con filtros de Butterworth de orden 2. Para obtener un filtrado sin desfase se usó el comando *filtfilt* del *Signal Processing Toolbox* de Matlab R2012a (*ibid.*); el cual aplica el filtro dos veces a la señal. La primera se aplica el filtro a la señal original, y una vez filtrada se invierte en el tiempo. La segunda se aplica a esta señal invertida, con lo cual se revierte el desfase causado por el primer filtrado. Para finalizar, se vuelve a invertir en tiempo la señal ya filtrada, restaurándose su orientación original. Como consecuencia de este procedimiento, se duplica el orden del filtro aplicado; por lo que el resultado es similar al de un filtro de orden 4, pero sin el desfase temporal asociado. Para poder cargar los archivos SAC a Matlab se usaron las funciones de la librería MatSAC (Peng, 2006).

Con el procedimiento de filtrado comentado se obtuvieron las siguientes bandas, incluyendo a la señal original. Las frecuencias de esquina de los filtros se eligieron de acuerdo con los rangos de frecuencia del ruido natural y antropogénico comentados en (Bonney-Claudet, Cotton, & Bard, 2006, pág. 209):

1. Sismograma sin filtrar.
2. Sismograma filtrado con un pasa bajas de frecuencia de esquina a 1 Hz.; para observar el ruido sísmico de origen natural.
3. Sismograma filtrado con un pasa altas de frecuencia de esquina a 5 Hz.; para observar el ruido sísmico de origen antropogénico.
4. Sismograma filtrado con un pasa bandas de frecuencias de esquina de 1 y 5 Hz.; para observar el ruido que puede provenir tanto de fuentes naturales como antropogénicas.

Partiendo de las cuatro bandas del paso anterior, a cada banda se le transformó de coordenadas cartesianas a esféricas, según la expresión [4]. Este paso también se implementó por medio de scripts de Matlab.

$$r = \sqrt{e^2 + n^2 + z^2}; \theta = \text{Atan}\left(\frac{\sqrt{e^2 + n^2}}{z}\right); \phi = \text{Atan2}(n, e) \quad [4]$$

En la expresión, (e, n, z) corresponde a las componentes cartesianas. En las componentes esféricas r corresponde al radio, con un rango de 0 a ∞ y unidades de cuentas (como las componentes del registro). θ corresponde a la colatitud, con unidades de radianes y un rango de $-\pi/2$ a $\pi/2$, con cero en el plano norte-este. ϕ corresponde al azimut, medido en radianes y con rango de $-\pi$ a π con cero en el eje $+\hat{e}$. El uso de la función Atan2 (presentada aquí con los argumentos en el orden solicitado por Matlab) permite salvar la discontinuidad entre los cuadrantes presente en la función tangente de un argumento.

7.1.3. Selección de *eventos* y *ruido*

El siguiente paso del preprocesamiento fue decidir cual segmento de la señal se marcaba como *ruido* y cuál como *evento*; para que el modelo de clasificación tuviera una variable de *clase* que aprender durante su entrenamiento. Por las razones que se comentaron en la sección de Generalidades del modelo propuesto y Antecedentes, a falta de una “verdad absoluta” se sigue considerando a un intérprete humano entrenado como la mejor clasificación posible.

Al plantear el problema como uno de clasificación binaria, un segmento del sismograma sólo puede ser de una de dos *clases* mutuamente exclusivas: o *ruido*, o *evento*. Por ello, se asume que la mayor parte de los sismogramas están formados por *ruido*, y sólo es necesario seleccionar a los *eventos* de entre el ruido. Para tal fin, se desarrolló una interfaz gráfica de Matlab que permite observar todas las bandas de todas las componentes esféricas. Una muestra de la interfaz gráfica se incluye en la Figura 1, incluida a continuación.

Dentro de la interfaz gráfica, al momento de seleccionar un segmento a indicar como *evento*, se marcan todas las muestras dentro de dicho segmento como pertenecientes a la *clase evento*. La *clase* asignada a cada muestra del sismograma se guarda en un vector de selección, con tantos elementos como muestras tenga el sismograma trabajado.

Para comprender mejor la diferencia entre eventos y ruido sísmico, así como la forma de distinguir entre ellos, se recomienda al lector revisar la sección de ¿Qué es el ruido sísmico?, en los Anexos.

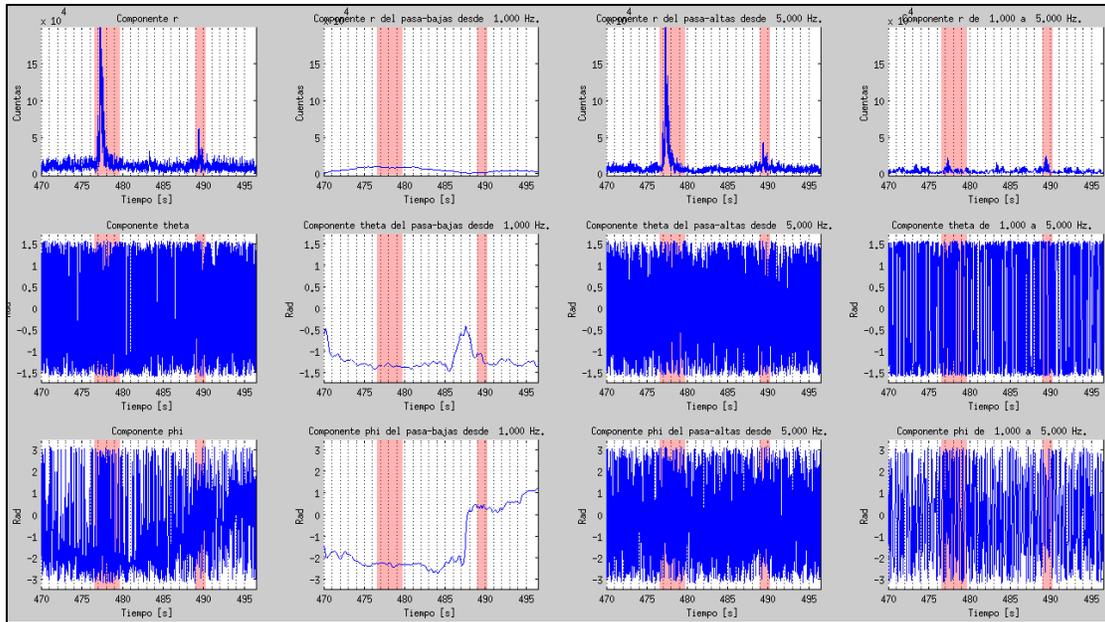


Figura 1: Interfaz gráfica para seleccionar muestras de eventos sísmicos, sobre las componentes de un registro en coordenadas esféricas. Las muestras seleccionadas se marcan en rojo. Cada banda de frecuencias ocupa una columna, y cada componente una fila.

7.1.4. Extracción de atributos de los sismogramas

Una vez que tenemos la *clase*, nos faltan los atributos para formar el conjunto de datos. Para formar los atributos se tomaron los sismogramas preprocesados de la manera expuesta en la sección de Filtrado y transformación a coordenadas esféricas; manteniendo las versiones tanto en coordenadas esféricas como en el sistema cartesiano original. A partir de estas señales se extrajeron ventanas de 256 muestras (2.56 segundos), con un empalme entre cada ventana y la siguiente de 128 muestras (1.28 segundos). Estos valores se escogieron para facilitar cualquier procesamiento posterior en el dominio de la frecuencia. La corta duración de las ventanas se escogió para tratar de aislar los eventos impulsivos dentro de la señal, tomando inspiración de (Trnkoczy, 1999, pág. 5).

Partiendo de las ventanas comentadas en el párrafo anterior se calcularon con otros códigos de Matlab varios atributos, partiendo de las siguientes propiedades del sismograma. Estos se calcularon para todas las bandas, incluyendo la señal sin filtrar:

- Elipticidades primaria, secundaria y terciaria.
- Rectilinearidad, parámetro global de polarización, y coeficientes de linealidad y planaridad.
- Orientación en coordenadas esféricas de los vectores característicos de la matriz de covarianza (en coordenadas cartesianas).

- Suma de cuadrados, promedio, pendiente de una recta ajustada por mínimos cuadrados, y desviación estándar muestral de las muestras dentro de la ventana; para todas las componentes esféricas
- Proporción entre un mismo atributo de una misma ventana, pero de diferentes bandas. Ésta se calculó para atributos selectos.
- Primera diferencia centrada (derivación numérica) de un mismo atributo. Calculada para atributos selectos.
- *Clase* de cada ventana: *ruido* o *evento*.

Con respecto al penúltimo último grupo de atributos, la primera diferencia centrada se calculó entre las dos ventanas contiguas (anterior y posterior) a una ventana particular, a la cual se atribuyó el valor de la primera diferencia centrada. Por ese motivo, no es posible calcular estos atributos para la primera y última ventana de cada sismograma. Por practicidad, se omitieron estas dos ventanas de cualquier análisis posterior. Al tratar con una cantidad considerable de ventanas, y gracias a la corta duración de ellas, no se pierde un segmento considerable de los sismogramas.

La clasificación de cada ventana se determinó observando las muestras del sismograma que pertenecían a la ventana, con la interfaz gráfica mostrada en la Figura 1. Si al menos una de las muestras dentro de la ventana se encontraba marcada como *evento*, se asignaba la ventana a dicha *clase*. Se llegaron a considerar esquemas más complicados para la asignación de la *clase* a la ventana, por ejemplo, a partir de la *clase* de la mayoría de las muestras, etcétera; pero fueron descartados al agregarle complejidad y más parámetros al método.

7.1.5. Selección de atributos

En total se obtuvieron 288 atributos diferentes de 967,526 ventanas, o casos; procedentes de los sismogramas trabajados. La totalidad de las variables extraídas se enuncia en la sección de Atributos extraídos de forma original, en los anexos. Por razones propias a la minería de datos y al sentido común, no es deseable ni necesario trabajar con todas estas variables. Por ello, el siguiente paso del preprocesamiento consistió en la selección de las variables más adecuadas para el modelo planteado. La razón por la cual se desaconseja trabajar con demasiadas variables es conocida como la *maldición de la dimensionalidad* (Hernández Orallo, Ramírez Quintana, & Ferri Ramírez, 2004, pág. 79); y se resume en que de haber demasiados atributos, aquellos patrones que pudieran indicar la *clase* de cada caso se diluyen en el espacio de los atributos hasta volverse imperceptibles por los modelos de clasificación.

La selección de variables se llevó a cabo en dos pasos, y sobre la totalidad de los casos. En el primero, se eligieron las 32 variables más relevantes, cuantificando su relevancia tanto por la correlación que tienen con la variable de *clase*, como por las métricas de cantidad de información expuestas en la sección de Métricas para la selección de atributos, en los anexos. Los atributos seleccionados abarcan el 50% de la relevancia normalizada de todas las variables, según se expone en la sección de Selección preliminar de atributos, también en los anexos. Para realizar este paso se empleó la interfaz gráfica del Weka Explorer (Frank, Hall, & Witten, 2016).

El segundo proceso de selección consistió en elegir a los atributos que, para cada *clase*, mostraran una distribución aproximadamente normal. Esto permite satisfacer el supuesto de que los atributos siguen una distribución normal multidimensional, con parámetros diferentes para cada *clase*. De este procedimiento, aplicado nuevamente sobre todos los casos; resultaron las 9 variables enunciadas en la Tabla 2, a continuación. Los histogramas de estos atributos, coloreados según la *clase*, se incluyen de forma posterior en la Figura 2, elaborada con Weka. El mismo programa asistió para realizar este paso. Es a partir de estas variables que se aplicó el procesamiento por minería de datos, como tal.

Tabla 2: Atributos seleccionados para el procesamiento, al estar dentro del 50% de los atributos más relevantes y tener una distribución cuasi normal.

Atributo	Descripción			
	Componente	Banda	Frecuencia de esquina	Atributo
<i>r_pbaj_01p00_sumc_pdc</i>	Radial	Pasa-bajas	1 Hz	Primera diferencia centrada, entre las ventanas contiguas (PDC); de la suma de los cuadrados de las muestras de la ventana.
<i>r_pbaj_01p00_prom_pdc</i>	Radial	Pasa-bajas	1 Hz	PDC del promedio de las muestras de la ventana.
<i>r_sin_filt_sumc_pdc</i>	Radial	(Sin filtrar)	(no aplica)	PDC de la suma de los cuadrados de las muestras de la ventana.
<i>r_pbaj_01p00_pend</i>	Radial	Pasa-bajas	1 Hz	Pendiente de una recta, ajustada por mínimos cuadrados a las muestras de la ventana.
<i>r_sin_filt_prom_pdc</i>	Radial	(Sin filtrar)	(no aplica)	PDC del promedio de las muestras de la ventana.
<i>r_pbaj_01p00_pend_pdc</i>	Radial	Pasa-bajas	1 Hz	PDC de la pendiente de una recta, ajustada por mínimos cuadrados a las muestras de la ventana.
<i>r_sin_filt_pend</i>	Radial	(Sin filtrar)	(no aplica)	Pendiente de una recta, ajustada por mínimos cuadrados a las muestras de la ventana.
<i>r_sin_filt_pend_pdc</i>	Radial	(Sin filtrar)	(no aplica)	PDC de la pendiente de una recta, ajustada por mínimos cuadrados a las muestras de la ventana.
<i>r_sin_filt_sd_pdc</i>	Radial	(Sin filtrar)	(no aplica)	PDC de la desviación estándar de las muestras de la ventana.

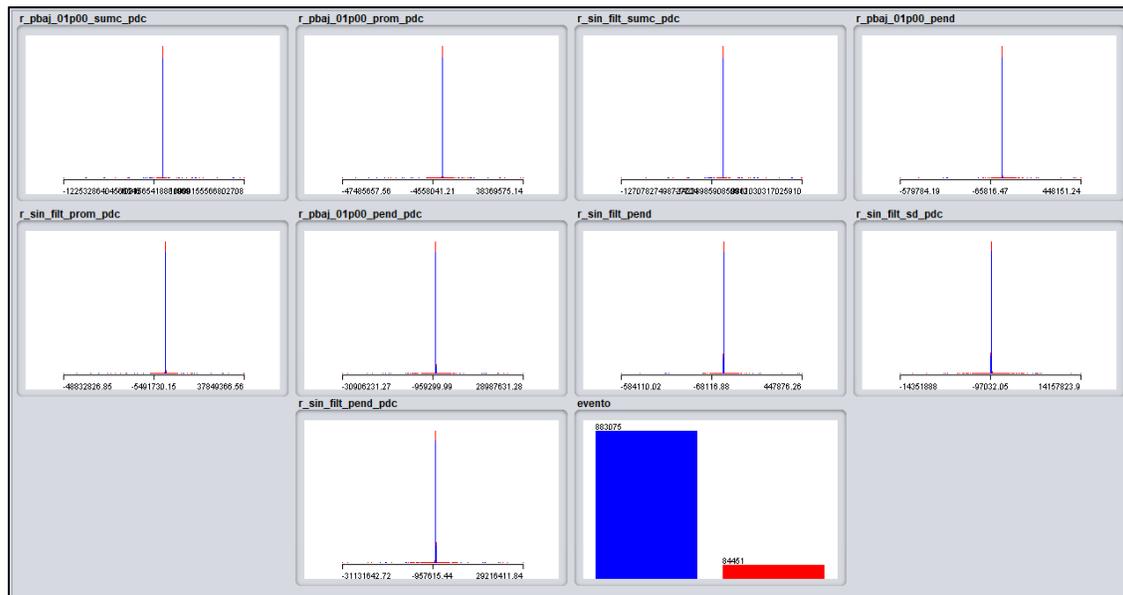


Figura 2: Histogramas de las variables seleccionadas para el procesamiento; y la distribución de los casos por clase (azul = ruido, rojo = evento). El nombre de cada atributo se incluye encima de cada gráfico. Figura elaborada con Weka

Explorer.

Dentro de la Figura 2 se nota que todos los histogramas tienen un aspecto impulsivo, con la inmensa mayoría de los casos aglomerados alrededor de un valor. Los casos aglomerados se componen casi en su totalidad por casos azules, lo que indica que proceden de ventanas marcadas como *ruido*. En torno a las aglomeraciones, y con un rango mucho mayor, se observan a los casos rojos, procedentes de ventanas marcadas como *evento*.

A partir de la Figura 2 se desprende la idea de ajustar una distribución normal multivariada a cada *clase*: mientras más cerca esté un caso de la aglomeración, tendrá una mayor probabilidad de ser un caso de *clase ruido*. De una manera similar, mientras más se aleje un caso de la aglomeración, mayor será su probabilidad de ser *evento*. La probabilidad se cuantifica por medio de las distribuciones (expresión [1]) ajustadas a cada *clase*, y el *score* del caso para cada *clase* se obtiene de normalizar la probabilidad de dicha *clase* entre la suma de las probabilidades (expresiones [2] y [3]). Los factores de ponderación π_0 y π_1 (expresiones [2] y [3]) toman en cuenta que la *clase ruido* es más común que la *clase evento* (véase el último gráfico de la Figura 2, variable *evento*); por lo que, a priori, un caso tiene una mayor probabilidad de ser *ruido*.

Al observar los nombres de los atributos elegidos, puede observarse que, de los nueve, siete de ellos terminan con “_pdc”. Esto indica que son atributos obtenidos al calcular la primera diferencia centrada de un atributo, entre las ventanas anterior y posterior a la ventana a la que se le imputa el valor. Los otros dos atributos son una pendiente ajustada por mínimos cuadrados a los valores dentro de una ventana. En ambos casos, para cada atributo se presentan valores tanto positivos como negativos, con una preferencia hacia un valor medio, lo cual origina la distribución observada, que se asume como normal para el desarrollo del trabajo. Lo anterior sucede aun cuando el atributo de origen está acotado en su rango, como la suma de los cuadrados de las muestras en la ventana, en el atributo *r_pbaj_01p00_sumc_pdc*.

También, resulta importante que todos los atributos elegidos provienen de las bandas del filtrado pasa bajas, y de la señal sin filtrar. La implicación práctica de esto es que las frecuencias contenidas en las bandas del pasa altas y pasa bandas no resultan determinantes para el modelo, ya que sus atributos no cuentan con una distribución cuasi normal o son de poca relevancia comparados con los de las otras bandas. Además, puede argumentarse que al trabajar con un problema binario; al conocer aquello que es ruido a partir de las bajas frecuencias y de la señal sin filtrar, por eliminación se puede caracterizar a los eventos.

7.2. Procesamiento

7.2.1. Conjuntos de entrenamiento y prueba

Como se comentó hace algunas secciones, para evaluar el rendimiento del modelo es necesario partir los datos disponibles en un conjunto de entrenamiento y uno de prueba. Ya que se tienen elegidas las variables que participarán en el modelo, se eligió a las ventanas procedentes del sismograma *GOLO_320* ya preprocesado (9,416 ventanas, abarcando 3 horas 20 minutos y 54 segundos) para formar el conjunto de entrenamiento, y el resto de los sismogramas preprocesados (958,110 ventanas; abarcando 18 sismogramas, algunos en varios fragmentos; y 14 días 4 horas con 39 minutos, no contiguos en el tiempo) se mandaron al de prueba. Esta elección se debió a que *GOLO_320* fue el primer registro en ser preprocesado, al tener una duración convenientemente breve, lo que facilitó su manejo.

7.2.2. Entrenamiento del modelo

Retomando las generalidades del modelo, los parámetros en los que se basa una distribución normal multivariada son el vector de valores medios, μ ; y la matriz de covarianza, V . El modelo propuesto contempla ajustar dos de estas distribuciones, una a los datos de cada *clase*. Como se menciona en la sección de Confección del modelo a partir de la distribución normal multivariada, el *score* para cada *clase* se calcula como la proporción entre la “probabilidad” dada por la función de densidad de probabilidad, para la *clase* de interés; dividida entre la suma de las probabilidades que da cada función de densidad. A cada función de densidad se le multiplica por la proporción de casos en dicha *clase*.

Con todo lo anterior, debemos exponer siete parámetros para definir el modelo, incluyendo la cantidad de variables en las distribuciones. Estos valores se tomarán de los casos en el conjunto de entrenamiento, formado por las ventanas del sismograma *GOLO_320*, al considerarse éste lo suficientemente representativo del resto de los sismogramas. Los valores se representan por los siguientes símbolos, retomados de las expresiones [1], [2], y [3]: $\mu_0, \mu_1, V_0, V_1, \pi_0, \pi_1, p$. Los símbolos se presentan a continuación, junto con su valor, en la Tabla 3. El orden de los elementos en los vectores y matrices se indica en los encabezados de fila y columna; y los vectores de valores medios se muestran transpuestos por cuestiones de espacio.

Tabla 3: Parámetros de las dos distribuciones normales multivariadas que conforman el modelo. Se muestra el transpuesto del vector de valores medios por cuestiones de espacio.

Símbolo	Valor									
μ_0^T	$r_{_pbaj_01p00_sumc_pdc}$	$r_{_pbaj_01p00_prom_pdc}$	$r_{_sin_filt_sumc_pdc}$	$r_{_pbaj_01p00_pend}$	$r_{_sin_filt_prom_pdc}$	$r_{_pbaj_01p00_pend_pdc}$	$r_{_sin_filt_pend}$	$r_{_sin_filt_sd_pdc}$	$r_{_sin_filt_pend_pdc}$	
	-3.2050E+07	-1.3927E+01	3.1311E+08	-2.9830E-01	7.1905E+00	-2.0885E+01	2.2669E+00	5.3052E+01	3.7448E+02	
μ_1^T	$r_{_pbaj_01p00_sumc_pdc}$	$r_{_pbaj_01p00_prom_pdc}$	$r_{_sin_filt_sumc_pdc}$	$r_{_pbaj_01p00_pend}$	$r_{_sin_filt_prom_pdc}$	$r_{_pbaj_01p00_pend_pdc}$	$r_{_sin_filt_pend}$	$r_{_sin_filt_sd_pdc}$	$r_{_sin_filt_pend_pdc}$	
	-4.7039E+06	-1.0289E+02	-4.4578E+09	1.9623E+00	-3.0210E+02	3.1209E+02	-1.1095E+01	-6.2836E+02	-4.0488E+03	
V_0		$r_{_pbaj_01p00_sumc_pdc}$	$r_{_pbaj_01p00_prom_pdc}$	$r_{_sin_filt_sumc_pdc}$	$r_{_pbaj_01p00_pend}$	$r_{_sin_filt_prom_pdc}$	$r_{_pbaj_01p00_pend_pdc}$	$r_{_sin_filt_pend}$	$r_{_sin_filt_sd_pdc}$	$r_{_sin_filt_pend_pdc}$
	$r_{_pbaj_01p00_sumc_pdc}$	4.9667E+20	4.7664E+14	5.0184E+20	5.8782E+12	1.0681E+14	-4.1046E+12	1.3278E+12	4.4066E+13	-4.4976E+12
	$r_{_pbaj_01p00_prom_pdc}$	4.7664E+14	5.5363E+08	5.1007E+14	6.9658E+06	1.0790E+08	-2.8887E+05	1.3868E+06	4.7646E+07	-6.5265E+06
	$r_{_sin_filt_sumc_pdc}$	5.0184E+20	5.1007E+14	2.1757E+23	6.8967E+12	3.2541E+16	1.3438E+14	3.7362E+14	1.6140E+16	2.1320E+14
	$r_{_pbaj_01p00_pend}$	5.8782E+12	6.9658E+06	6.8967E+12	1.3167E+05	1.3595E+06	-1.5307E+04	2.1825E+04	6.3895E+05	-1.5827E+05
	$r_{_sin_filt_prom_pdc}$	1.0681E+14	1.0790E+08	3.2541E+16	1.3595E+06	5.1503E+09	2.3185E+07	5.8973E+07	2.3723E+09	1.3393E+07
	$r_{_pbaj_01p00_pend_pdc}$	-4.1046E+12	-2.8887E+05	1.3438E+14	-1.5307E+04	2.3185E+07	8.4404E+08	1.3393E+05	2.4251E+06	1.8799E+08
	$r_{_sin_filt_pend}$	1.3278E+12	1.3868E+06	3.7362E+14	2.1825E+04	5.8973E+07	1.3393E+05	1.1914E+06	2.7803E+07	1.1083E+05
	$r_{_sin_filt_sd_pdc}$	4.4066E+13	4.7646E+07	1.6140E+16	6.3895E+05	2.3723E+09	2.4251E+06	2.7803E+07	1.5526E+09	8.0516E+07
	$r_{_sin_filt_pend_pdc}$	-4.4976E+12	-6.5265E+06	2.1320E+14	-1.5827E+05	1.3393E+07	1.8799E+08	1.1083E+05	8.0516E+07	6.8809E+09
V_1		$r_{_pbaj_01p00_sumc_pdc}$	$r_{_pbaj_01p00_prom_pdc}$	$r_{_sin_filt_sumc_pdc}$	$r_{_pbaj_01p00_pend}$	$r_{_sin_filt_prom_pdc}$	$r_{_pbaj_01p00_pend_pdc}$	$r_{_sin_filt_pend}$	$r_{_sin_filt_sd_pdc}$	$r_{_sin_filt_pend_pdc}$
	$r_{_pbaj_01p00_sumc_pdc}$	1.3921E+27	1.9324E+19	1.3919E+27	1.5835E+17	1.9196E+19	-1.1511E+18	1.5771E+17	6.2590E+16	-1.1260E+18
	$r_{_pbaj_01p00_prom_pdc}$	1.9324E+19	3.1501E+11	1.9348E+19	2.5871E+09	3.0994E+11	9.0336E+06	2.5509E+09	2.4155E+09	9.7423E+08
	$r_{_sin_filt_sumc_pdc}$	1.3919E+27	1.9348E+19	1.8507E+27	1.5864E+17	2.6159E+19	-1.1495E+18	2.4182E+17	7.5918E+18	-1.3563E+18
	$r_{_pbaj_01p00_pend}$	1.5835E+17	2.5871E+09	1.5864E+17	2.1516E+07	2.5465E+09	1.2869E+05	2.1199E+07	2.0164E+07	3.6327E+06
	$r_{_sin_filt_prom_pdc}$	1.9196E+19	3.0994E+11	2.6159E+19	2.5465E+09	4.3626E+11	-8.6401E+08	4.0460E+09	1.3208E+11	-1.3026E+10
	$r_{_pbaj_01p00_pend_pdc}$	-1.1511E+18	9.0336E+06	-1.1495E+18	1.2869E+05	-8.6401E+08	4.1322E+10	-3.4575E+06	5.0886E+08	4.0871E+10
	$r_{_sin_filt_pend}$	1.5771E+17	2.5509E+09	2.4182E+17	2.1199E+07	4.0460E+09	-3.4575E+06	5.0502E+07	1.5795E+09	-1.3141E+06
	$r_{_sin_filt_sd_pdc}$	6.2590E+16	2.4155E+09	7.5918E+18	2.0164E+07	1.3208E+11	5.0886E+08	1.5795E+09	1.4628E+11	1.4774E+10
	$r_{_sin_filt_pend_pdc}$	-1.1260E+18	9.7423E+08	-1.3563E+18	3.6327E+06	-1.3026E+10	4.0871E+10	-1.3141E+06	1.4774E+10	2.1521E+11
π_0	0.9180									
π_1	0.0819									
p	9									

Antes de continuar, es preciso comentar que el cálculo de las matrices inversas de las matrices de covarianza, V_0^{-1} y V_1^{-1} , puede fallar con algunas técnicas de inversión dadas las magnitudes de los elementos que las componen. Como alternativa, se sugiere aprovechar las propiedades de la descomposición de Cholesky, y de ella obtener la matriz inversa. Este procedimiento fue el implementado para el desarrollo del trabajo.

7.2.3. Evaluación del conjunto de datos

Para evaluar cada caso, sin importar el conjunto del que se origine, se procede de la siguiente forma:

1. Obtener el factor de escala de la expresión [1], $(\sqrt{|V|} \sqrt{2\pi})^{-1}$, para cada *clase*; y calcular las proporciones a priori de las *clases* (π_0 y π_1). Estos valores no cambian con los casos, por lo que se les calcula sólo una vez.
2. Tomar sólo a los atributos seleccionados (Tabla 2), y con ellos formar un vector x en el mismo orden que las variables del vector de valores medios (mostrado en la Tabla 3).
3. Obtener el vector $(x - \mu)$ y su transpuesto; para los vectores de valores medios de cada *clase*: μ_0 y μ_1 .
4. Para cada *clase*, obtener la *distancia de Mahalanobis*: $(x - \mu)'V^{-1}(x - \mu)$.
5. Con la *distancia de Mahalanobis* y el factor de escala del inciso 1, evaluar la expresión [1] para cada *clase*, para obtener las densidades de probabilidad $f_0(x)$ y $f_1(x)$.
6. A partir de los valores obtenidos en el inciso anterior, y con las proporciones a priori de las *clases*, evaluar la expresión [3] para obtener el *score* para la *clase evento*, s_1 . Recordemos que los *scores* de ambas *clases* se complementan a uno, por lo que podemos trabajar con sólo uno de ellos.

Al momento de determinar el rendimiento del modelo, se toma el valor de s_1 para cada caso junto con el valor de la *clase*. Dos de las métricas de rendimiento más comunes, las curvas de *Receiver Operating Characteristic* (*ROC*, por sus siglas en inglés); y la curva de *Precision-Recall*, se forman variando el *score de corte* de 0 a 1 y comparando la clasificación que se otorga a los casos según este umbral contra la *clase*. Estas métricas de rendimiento se incluyen y explican en la sección de Resultados, más adelante.

7.2.4. Métricas de rendimiento para el conjunto de entrenamiento

Considerando lo expuesto al comentar la evaluación del conjunto de datos, a continuación, se exponen las métricas de rendimiento del modelo propuesto para el conjunto de entrenamiento. Es muy importante recordar que estas métricas de rendimiento resultarán optimistas comparadas con las del conjunto de prueba, ya que al modelo se le pide que evalúe los mismos datos con los que fue entrenado. La interpretación de las métricas de rendimiento se reserva para los valores correspondientes al conjunto de prueba (en la sección de Métricas de rendimiento para el conjunto de prueba), al ser los que dan sustento al rendimiento del modelo.

Para continuar, en la Figura 3, se exponen los elementos de la matriz de confusión en términos del *score de corte*, considerando a la *clase evento* como la verdadera, como se ha hecho en el resto del trabajo. A los elementos de ésta se les ha normalizado entre la cantidad de casos en el conjunto de entrenamiento ($n = 9416$). Analizando la figura, se observa que hay fluctuaciones mínimas de los elementos de la matriz con el *score de corte*, lo cual aparentemente indica que la clasificación es prácticamente independiente del *score de corte*. Sin embargo, veremos más adelante que este no es el caso. Lo único claro es que esta figura en particular no sirve para elegir al *score de corte*.

Tras la Figura 3, se introduce en la Figura 4 a la curva de ROC. En la siguiente, la Figura 5, se muestra la curva de *Precision-Recall*.

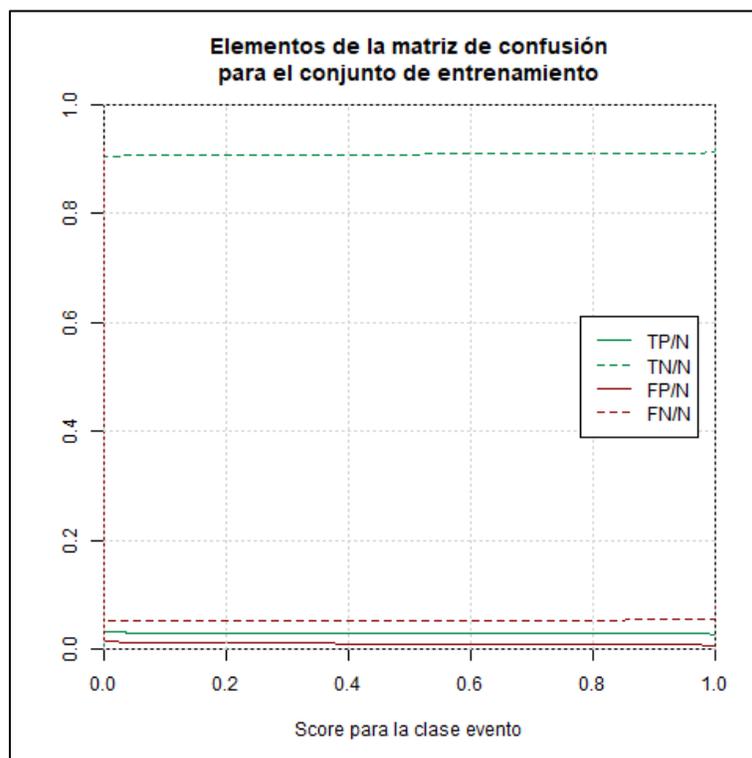


Figura 3: Elementos de la matriz de confusión, normalizados entre la cantidad de casos, para el conjunto de prueba.

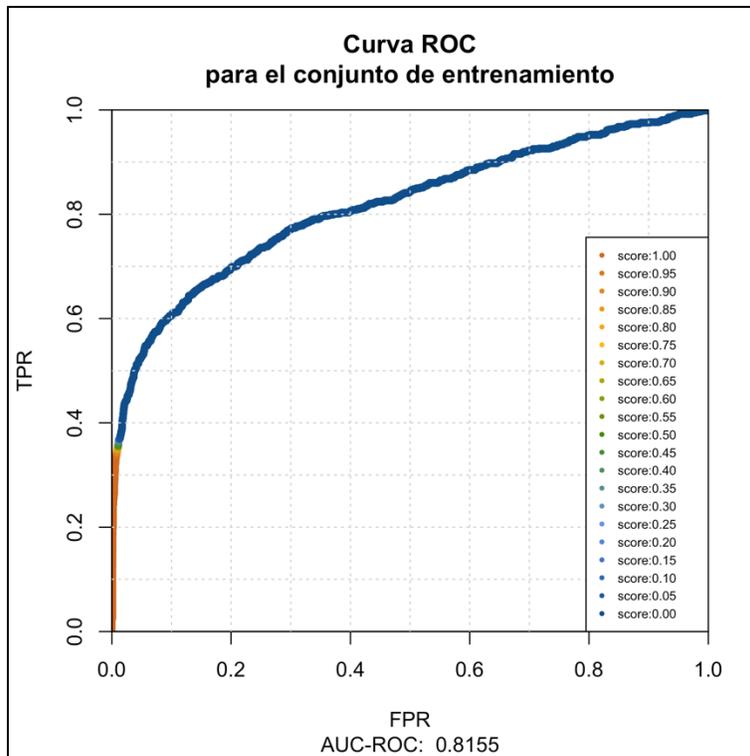


Figura 4: Curva de ROC para el conjunto de entrenamiento. El área bajo la curva se indica en la parte inferior. Para una explicación, véase la Figura 7 en la sección de Métricas de rendimiento para el conjunto de prueba.

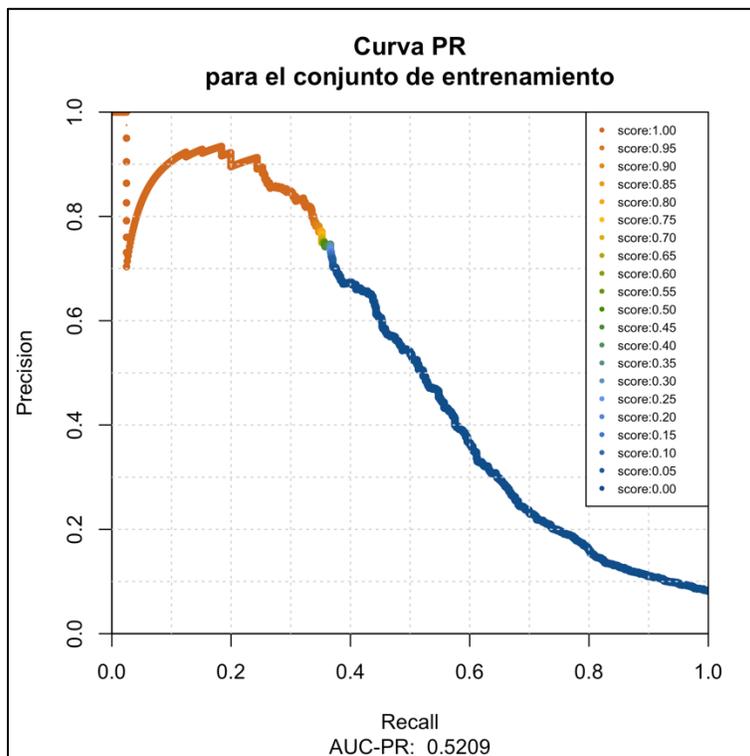


Figura 5: Curva de Precisión-Recall para el conjunto de entrenamiento. El área bajo la curva se indica en la parte inferior. Para una explicación, véase la Figura 8 en la sección de Métricas de rendimiento para el conjunto de prueba.

8. Resultados

8.1. Métricas de rendimiento para el conjunto de prueba

La metodología de trabajo expuesta muestra cómo se evalúa el conjunto de datos (sean de entrenamiento o prueba) por el modelo normal multivariado propuesto; y explica cómo se obtiene un valor del *score* s_1 para cada uno de los casos, o ventanas de los sismogramas. Las métricas por mostrar permiten elegir un *score de corte* a partir del rendimiento del modelo a dicho valor; y comparan la clasificación que hace el modelo con respecto a los diversos elementos de la matriz de confusión, según el *score de corte*.

En la Tabla 4 se presentan las métricas de rendimiento a utilizar, las cuales fueron tomadas de (Davis & Goadrich, 2006). Todas ellas se extraen de los elementos de la matriz de confusión, y al depender ésta del *score de corte*, también son función de él. La interpretación que se le da a estas métricas se define a continuación:

- El *True Positive Rate (TPR)*, también llamado *Recall*; es la proporción de casos bien clasificados como *evento*, entre la cantidad de casos con *clase evento*. En otras palabras, responde a la pregunta ¿cuántos de los casos con *clase evento* fueron bien identificados como *evento*?
- El *False Positive Rate (FPR)* es la proporción de casos mal clasificados como *evento*, entre la cantidad de casos con *clase ruido*, por lo que responde a la pregunta ¿cuántos de los casos con *clase ruido* fueron mal identificados como *evento*?
- La *Precision* es la proporción de casos bien clasificados como *evento*, respecto al total de casos clasificados como *evento*, por lo que responde a la pregunta: ¿cuántos de los casos de la *clase evento* fueron bien clasificados?

Tabla 4: Expresiones de las métricas de rendimiento derivadas de la matriz de confusión. Tomadas de (Davis & Goadrich, 2006)

Métrica	Expresión
<i>True Positive Rate (TPR), Recall</i>	$TP/(TP + FN)$
<i>False Positive Rate (FPR)</i>	$FP/(FP + TN)$
<i>Precision</i>	$TP/(TP + FP)$

Ninguna de las métricas anteriores es perfecta, ya que no toman en cuenta todos los elementos de la matriz de confusión. Por lo tanto, a partir de una sola de ellas no es posible reconstruir el rendimiento completo del modelo de clasificación. Por ejemplo, un clasificador que identifique todo como *positivo* puede tener una *Precision* de 0.99, pero debido a que el conjunto de

datos sólo tiene un caso *negativo* contra noventa y nueve *positivos*. Para compensar sus deficiencias, estas métricas se usan en pares. Dos combinaciones muy populares son las de *FPR* con *TPR*, y *Precision* con *Recall* (Davis & Goadrich, 2006).

Las combinaciones anteriores se exponen convencionalmente de forma gráfica; formando la comparación de *FPR* contra *TPR* la curva *ROC* (*Receiver Operating Characteristic*, en inglés); y la comparación de *Recall* contra *Precision* la curva de *Precision-Recall* (curva *PR*). Dos métricas más se desprenden de estas curvas, al integrar el área que hay entre la curva y el eje horizontal de la gráfica. Éstas son el área bajo la curva *ROC* (*Area Under Curve ROC*, *AUC-ROC* por sus siglas en inglés), y el área bajo la curva de *Precision-Recall* (*AUC-PR*). Ambas métricas de área bajo la curva tienen un rango de 0 a 1, siendo mejor los valores más altos; y al integrar las curvas resultantes para todos los *scores* de corte, condensan en un solo valor el rendimiento del modelo.

La Tabla 5 introduce a la matriz de confusión para el conjunto de prueba, acompañada de las métricas recién definidas. La misma matriz de confusión, pero en función del *score de corte* se introduce en la Figura 6. Se observa que, en el extremo superior de la tabla, correspondiente al extremo derecho de la figura; en los valores del *score de corte* próximos a uno, la *Precision* alcanza su valor máximo, al contrario del *TPR* y el *Recall*. Recordando las interpretaciones de estas métricas, se deduce que a *scores de corte* altos se llega a discriminar algunos cuantos de los casos de la *clase evento*, sin que se etiquete mal a los de la *clase ruido*. Sin embargo, la consecuencia es que muchos de los casos de *clase evento* son mal clasificados, y se les considera *ruido*.

Por ejemplo, a un *score de corte* de 0.98, el modelo ha clasificado a 74,067 casos (7.73% del total) como *evento*, mientras que a los 884,043 casos restantes (92.27%) los ha clasificado como *ruido*. Estos valores proceden de sumar las columnas de cantidad de casos a partir de la fila del *score de corte* hacia arriba, y de la fila inferior a la del *score de corte* hacia abajo, respectivamente. A este mismo *score* tenemos un *TPR* de 0.4871, un *FPR* de 0.0381, y un *Precision* de 0.5503. Esto significa que de los casos que clasificamos como *evento*, 55.03% sí lo son; mientras que para los casos con *clase ruido* estamos clasificando incorrectamente sólo a un 3.81% de ellos como si fueran *evento*. Sin embargo, al escoger un *score de corte* tan alto, sólo estamos clasificando bien al 48.71% de los casos con *clase evento*; mientras que el resto de ellos (51.29%) los clasificamos mal, como *ruido*.

En los extremos del *score de corte* se observa que todas estas medidas alcanzan sus valores máximos o mínimos. A un *score* de 1.00, la *Precision* alcanza su máximo de 0.5573. No es posible obtener un mejor rendimiento, bajo esta métrica, para el modelo propuesto. En el otro extremo de la Tabla 5, a un *score de corte* de 0.00; tanto el *TPR* como el *FPR* toman un valor máximo de 1.0000.

Esto se debe a que, al elegir dicho *score de corte*, se considera que todos los casos se clasifican como *evento*; y como consecuencia, clasificamos bien al 100% de los casos con *clase evento*, y clasificamos mal al 100% de aquellos de la *clase ruido*.

Tabla 5: Métricas de rendimiento en función del score de corte para la clase evento, para el conjunto de datos de prueba. Se han agrupado los casos según el score que les otorga el modelo, que fue redondeado a dos decimales.

Score de corte para la clase evento	Cantidad de casos con clase = ruido	Cantidad de casos con clase = evento	Casos clasificados como evento	Casos clasificados como ruido	TP: Verdaderos positivos	TN: Verdaderos negativos	FP: Falsos positivos	FN: Falsos negativos	Casos clasificados como evento / N	Casos clasificados como ruido / N	TP: Verdaderos positivos / N	TN: Verdaderos negativos / N	FP: Falsos positivos / N	FN: Falsos negativos / N	TPR = Recall	FPR	Precision
1.00	32,038	40,334	72,372	885,738	40,334	842,393	32,038	43,345	7.55%	92.45%	4.21%	87.92%	3.34%	4.52%	0.4820	0.0366	0.5573
0.99	855	287	73,514	884,596	40,621	841,538	32,893	43,058	7.67%	92.33%	4.24%	87.83%	3.43%	4.49%	0.4854	0.0376	0.5526
0.98	412	141	74,067	884,043	40,762	841,126	33,305	42,917	7.73%	92.27%	4.25%	87.79%	3.48%	4.48%	0.4871	0.0381	0.5503
0.97	290	95	74,452	883,658	40,857	840,836	33,595	42,822	7.77%	92.23%	4.26%	87.76%	3.51%	4.47%	0.4883	0.0384	0.5488
0.96	228	77	74,757	883,353	40,934	840,608	33,823	42,745	7.80%	92.20%	4.27%	87.74%	3.53%	4.46%	0.4892	0.0387	0.5476
0.95	198	52	75,007	883,103	40,986	840,410	34,021	42,693	7.83%	92.17%	4.28%	87.72%	3.55%	4.46%	0.4898	0.0389	0.5464
0.94	157	43	75,207	882,903	41,029	840,253	34,178	42,650	7.85%	92.15%	4.28%	87.70%	3.57%	4.45%	0.4903	0.0391	0.5455
0.93	161	41	75,409	882,701	41,070	840,092	34,339	42,609	7.87%	92.13%	4.29%	87.68%	3.58%	4.45%	0.4908	0.0393	0.5446
0.92	115	36	75,560	882,550	41,106	839,977	34,454	42,573	7.89%	92.11%	4.29%	87.67%	3.60%	4.44%	0.4912	0.0394	0.5440
0.91	101	43	75,704	882,406	41,149	839,876	34,555	42,530	7.90%	92.10%	4.29%	87.66%	3.61%	4.44%	0.4917	0.0395	0.5436
0.90	108	29	75,841	882,269	41,178	839,768	34,663	42,501	7.92%	92.08%	4.30%	87.65%	3.62%	4.44%	0.4921	0.0396	0.5430
0.89	99	33	75,973	882,137	41,211	839,669	34,762	42,468	7.93%	92.07%	4.30%	87.64%	3.63%	4.43%	0.4925	0.0398	0.5424
0.88	72	31	76,076	882,034	41,242	839,597	34,834	42,437	7.94%	92.06%	4.30%	87.63%	3.64%	4.43%	0.4929	0.0398	0.5421
0.87	67	20	76,163	881,947	41,262	839,530	34,901	42,417	7.95%	92.05%	4.31%	87.62%	3.64%	4.43%	0.4931	0.0399	0.5418
0.86	93	36	76,292	881,818	41,298	839,437	34,994	42,381	7.96%	92.04%	4.31%	87.61%	3.65%	4.42%	0.4935	0.0400	0.5413
0.85	78	17	76,387	881,723	41,315	839,359	35,072	42,364	7.97%	92.03%	4.31%	87.61%	3.66%	4.42%	0.4937	0.0401	0.5409
0.84	76	26	76,489	881,621	41,341	839,283	35,148	42,338	7.98%	92.02%	4.31%	87.60%	3.67%	4.42%	0.4940	0.0402	0.5405
0.83	65	14	76,568	881,542	41,355	839,218	35,213	42,324	7.99%	92.01%	4.32%	87.59%	3.68%	4.42%	0.4942	0.0403	0.5401
0.82	84	14	76,666	881,444	41,369	839,134	35,297	42,310	8.00%	92.00%	4.32%	87.58%	3.68%	4.42%	0.4944	0.0404	0.5396
0.81	69	16	76,751	881,359	41,385	839,065	35,366	42,294	8.01%	91.99%	4.32%	87.58%	3.69%	4.41%	0.4946	0.0404	0.5392
0.80	73	15	76,839	881,271	41,400	838,992	35,439	42,279	8.02%	91.98%	4.32%	87.57%	3.70%	4.41%	0.4947	0.0405	0.5388
0.79	55	18	76,912	881,198	41,418	838,937	35,494	42,261	8.03%	91.97%	4.32%	87.56%	3.70%	4.41%	0.4950	0.0406	0.5385
0.78	55	21	76,988	881,122	41,439	838,882	35,549	42,240	8.04%	91.96%	4.33%	87.56%	3.71%	4.41%	0.4952	0.0407	0.5383
0.77	58	25	77,071	881,039	41,464	838,824	35,607	42,215	8.04%	91.96%	4.33%	87.55%	3.72%	4.41%	0.4955	0.0407	0.5380
0.76	58	17	77,146	880,964	41,481	838,766	35,665	42,198	8.05%	91.95%	4.33%	87.54%	3.72%	4.40%	0.4957	0.0408	0.5377
0.75	55	15	77,216	880,894	41,496	838,711	35,720	42,183	8.06%	91.94%	4.33%	87.54%	3.73%	4.40%	0.4959	0.0408	0.5374
0.74	52	18	77,286	880,824	41,514	838,659	35,772	42,165	8.07%	91.93%	4.33%	87.53%	3.73%	4.40%	0.4961	0.0409	0.5371
0.73	51	15	77,352	880,758	41,529	838,608	35,823	42,150	8.07%	91.93%	4.33%	87.53%	3.74%	4.40%	0.4963	0.0410	0.5369
0.72	48	18	77,418	880,692	41,547	838,560	35,871	42,132	8.08%	91.92%	4.34%	87.52%	3.74%	4.40%	0.4965	0.0410	0.5367
0.71	62	21	77,501	880,609	41,568	838,498	35,933	42,111	8.09%	91.91%	4.34%	87.52%	3.75%	4.40%	0.4968	0.0411	0.5364
0.70	39	14	77,554	880,556	41,582	838,459	35,972	42,097	8.09%	91.91%	4.34%	87.51%	3.75%	4.39%	0.4969	0.0411	0.5362
0.69	51	14	77,619	880,491	41,596	838,408	36,023	42,083	8.10%	91.90%	4.34%	87.51%	3.76%	4.39%	0.4971	0.0412	0.5359
0.68	57	18	77,694	880,416	41,614	838,351	36,080	42,065	8.11%	91.89%	4.34%	87.50%	3.77%	4.39%	0.4973	0.0413	0.5356
0.67	45	9	77,748	880,362	41,623	838,306	36,125	42,056	8.11%	91.89%	4.34%	87.50%	3.77%	4.39%	0.4974	0.0413	0.5354
0.66	39	14	77,801	880,309	41,637	838,267	36,164	42,042	8.12%	91.88%	4.35%	87.49%	3.77%	4.39%	0.4976	0.0414	0.5352
0.65	47	15	77,863	880,247	41,652	838,220	36,211	42,027	8.13%	91.87%	4.35%	87.49%	3.78%	4.39%	0.4978	0.0414	0.5349
0.64	51	13	77,927	880,183	41,665	838,169	36,262	42,014	8.13%	91.87%	4.35%	87.48%	3.78%	4.39%	0.4979	0.0415	0.5347
0.63	46	15	77,988	880,122	41,680	838,123	36,308	41,999	8.14%	91.86%	4.35%	87.48%	3.79%	4.38%	0.4981	0.0415	0.5344
0.62	37	21	78,046	880,064	41,701	838,086	36,345	41,978	8.15%	91.85%	4.35%	87.47%	3.79%	4.38%	0.4983	0.0416	0.5343
0.61	38	13	78,097	880,013	41,714	838,048	36,383	41,965	8.15%	91.85%	4.35%	87.47%	3.80%	4.38%	0.4985	0.0416	0.5341
0.60	49	15	78,161	879,949	41,729	837,999	36,432	41,950	8.16%	91.84%	4.36%	87.46%	3.80%	4.38%	0.4987	0.0417	0.5339
0.59	48	14	78,223	879,887	41,743	837,951	36,480	41,936	8.16%	91.84%	4.36%	87.46%	3.81%	4.38%	0.4988	0.0417	0.5336
0.58	47	11	78,281	879,829	41,754	837,904	36,527	41,925	8.17%	91.83%	4.36%	87.45%	3.81%	4.38%	0.4990	0.0418	0.5334
0.57	42	12	78,335	879,775	41,766	837,862	36,569	41,913	8.18%	91.82%	4.36%	87.45%	3.82%	4.37%	0.4991	0.0418	0.5332
0.56	39	16	78,390	879,720	41,782	837,823	36,608	41,897	8.18%	91.82%	4.36%	87.45%	3.82%	4.37%	0.4993	0.0419	0.5330
0.55	31	14	78,435	879,675	41,796	837,792	36,639	41,883	8.19%	91.81%	4.36%	87.44%	3.82%	4.37%	0.4995	0.0419	0.5329
0.54	53	13	78,501	879,609	41,809	837,739	36,692	41,870	8.19%	91.81%	4.36%	87.44%	3.83%	4.37%	0.4996	0.0420	0.5326
0.53	40	7	78,548	879,562	41,816	837,699	36,732	41,863	8.20%	91.80%	4.36%	87.43%	3.83%	4.37%	0.4997	0.0420	0.5324
0.52	36	12	78,596	879,514	41,828	837,663	36,768	41,851	8.20%	91.80%	4.37%	87.43%	3.84%	4.37%	0.4999	0.0420	0.5322
0.51	54	14	78,664	879,446	41,842	837,609	36,822	41,837	8.21%	91.79%	4.37%	87.42%	3.84%	4.37%	0.5000	0.0421	0.5319
0.50	46	11	78,721	879,389	41,853	837,563	36,868	41,826	8.22%	91.78%	4.37%	87.42%	3.85%	4.37%	0.5002	0.0422	0.5317
0.49	44	12	78,782	879,328	41,870	837,519	36,912	41,809	8.22%	91.78%	4.37%	87.41%	3.85%	4.36%	0.5004	0.0422	0.5315
0.48	44	14	78,840	879,270	41,884	837,475	36,956	41,795	8.23%	91.77%	4.37%	87.41%	3.86%	4.36%	0.5005	0.0423	0.5313
0.47	40	18	78,898	879,212	41,902	837,435	36,996	41,777	8.23%	91.77%	4.37%	87.40%	3.86%	4.36%	0.5007	0.0423	0.5311
0.46	34	8	78,940	879,170	41,910	837,401	37,030	41,769	8.24%	91.76%	4.37%	87.40%	3.86%	4.36%	0.5008	0.0423	0.5309
0.45	47	11	78,998	879,112	41,921	837,354	37,077	41,758	8.25%	91.75%	4.38%	87.40%	3.87%	4.36%	0.5010	0.0424	0.5307
0.44	55	13	79,066	879,044	41,934	837,299	37,132	41,745	8.25%	91.75%	4.38%	87.39%	3.88%	4.36%	0.5011	0.0425	0.5304
0.43	46	16	79,128	878,982	41,950	837,253	37,178	41,729	8.26%	91.74%	4.38%	87.39%	3.88%	4.36%	0.5013	0.0425	0.5302

Score de corte para la clase evento	Cantidad de casos con clase = ruido	Cantidad de casos con clase = evento	Casos clasificados como evento	Casos clasificados como ruido	TP: Verdaderos positivos	TN: Verdaderos negativos	FP: Falsos positivos	FN: Falsos negativos	Casos clasificados como evento / N	Casos clasificados como ruido / N	TP: Verdaderos positivos / N	TN: Verdaderos negativos / N	FP: Falsos positivos / N	FN: Falsos negativos / N	TPR = Recall	FPR	Precision
0.42	43	19	79,190	878,920	41,969	837,210	37,221	41,710	8.27%	91.73%	4.38%	87.38%	3.88%	4.35%	0.5015	0.0426	0.5300
0.41	44	12	79,246	878,864	41,981	837,166	37,265	41,698	8.27%	91.73%	4.38%	87.38%	3.89%	4.35%	0.5017	0.0426	0.5298
0.40	43	18	79,307	878,803	41,999	837,123	37,308	41,680	8.28%	91.72%	4.38%	87.37%	3.89%	4.35%	0.5019	0.0427	0.5296
0.39	35	13	79,355	878,755	42,012	837,088	37,343	41,667	8.28%	91.72%	4.38%	87.37%	3.90%	4.35%	0.5021	0.0427	0.5294
0.38	47	15	79,417	878,693	42,027	837,041	37,390	41,652	8.29%	91.71%	4.39%	87.36%	3.90%	4.35%	0.5022	0.0428	0.5292
0.37	51	17	79,485	878,625	42,044	836,990	37,441	41,635	8.30%	91.70%	4.39%	87.36%	3.91%	4.35%	0.5024	0.0428	0.5290
0.36	51	14	79,550	878,560	42,058	836,939	37,492	41,621	8.30%	91.70%	4.39%	87.35%	3.91%	4.34%	0.5026	0.0429	0.5287
0.35	63	20	79,633	878,477	42,078	836,876	37,555	41,601	8.31%	91.69%	4.39%	87.35%	3.92%	4.34%	0.5029	0.0429	0.5284
0.34	50	17	79,700	878,410	42,095	836,826	37,605	41,584	8.32%	91.68%	4.39%	87.34%	3.92%	4.34%	0.5031	0.0430	0.5282
0.33	49	21	79,770	878,340	42,116	836,777	37,654	41,563	8.33%	91.67%	4.40%	87.34%	3.93%	4.34%	0.5033	0.0431	0.5280
0.32	60	13	79,843	878,267	42,129	836,717	37,714	41,550	8.33%	91.67%	4.40%	87.33%	3.94%	4.34%	0.5035	0.0431	0.5278
0.31	49	19	79,911	878,199	42,148	836,668	37,763	41,531	8.34%	91.66%	4.40%	87.32%	3.94%	4.33%	0.5037	0.0432	0.5274
0.30	71	9	79,991	878,119	42,157	836,597	37,834	41,522	8.35%	91.65%	4.40%	87.32%	3.95%	4.33%	0.5038	0.0433	0.5270
0.29	43	21	80,055	878,055	42,178	836,554	37,877	41,501	8.36%	91.64%	4.40%	87.31%	3.95%	4.33%	0.5040	0.0433	0.5269
0.28	59	16	80,130	877,980	42,194	836,495	37,936	41,485	8.36%	91.64%	4.40%	87.31%	3.96%	4.33%	0.5042	0.0434	0.5266
0.27	63	15	80,208	877,902	42,209	836,432	37,999	41,470	8.37%	91.63%	4.41%	87.30%	3.97%	4.33%	0.5044	0.0435	0.5262
0.26	75	16	80,299	877,811	42,225	836,357	38,074	41,454	8.38%	91.62%	4.41%	87.29%	3.97%	4.33%	0.5046	0.0435	0.5258
0.25	70	14	80,383	877,727	42,239	836,287	38,144	41,440	8.39%	91.61%	4.41%	87.29%	3.98%	4.33%	0.5048	0.0436	0.5255
0.24	72	15	80,470	877,640	42,254	836,215	38,216	41,425	8.40%	91.60%	4.41%	87.28%	3.99%	4.32%	0.5050	0.0437	0.5251
0.23	70	18	80,558	877,552	42,272	836,145	38,286	41,407	8.41%	91.59%	4.41%	87.27%	4.00%	4.32%	0.5052	0.0438	0.5247
0.22	56	20	80,634	877,476	42,292	836,089	38,342	41,387	8.42%	91.58%	4.41%	87.26%	4.00%	4.32%	0.5054	0.0438	0.5245
0.21	80	23	80,737	877,373	42,315	836,009	38,422	41,364	8.43%	91.57%	4.42%	87.26%	4.01%	4.32%	0.5057	0.0439	0.5241
0.20	93	21	80,851	877,259	42,336	835,916	38,515	41,343	8.44%	91.56%	4.42%	87.25%	4.02%	4.32%	0.5059	0.0440	0.5236
0.19	63	18	80,932	877,178	42,354	835,853	38,578	41,325	8.45%	91.55%	4.42%	87.24%	4.03%	4.31%	0.5061	0.0441	0.5233
0.18	89	25	81,046	877,064	42,379	835,764	38,667	41,300	8.46%	91.54%	4.42%	87.23%	4.04%	4.31%	0.5064	0.0442	0.5229
0.17	81	34	81,161	876,949	42,413	835,683	38,748	41,266	8.47%	91.53%	4.43%	87.22%	4.04%	4.31%	0.5069	0.0443	0.5226
0.16	95	26	81,282	876,828	42,439	835,588	38,843	41,240	8.48%	91.52%	4.43%	87.21%	4.05%	4.30%	0.5072	0.0444	0.5221
0.15	107	20	81,409	876,701	42,459	835,481	38,950	41,220	8.50%	91.50%	4.43%	87.20%	4.07%	4.30%	0.5074	0.0445	0.5216
0.14	110	29	81,548	876,562	42,488	835,371	39,060	41,191	8.51%	91.49%	4.43%	87.19%	4.08%	4.30%	0.5077	0.0447	0.5210
0.13	105	33	81,686	876,424	42,521	835,266	39,165	41,158	8.53%	91.47%	4.44%	87.18%	4.09%	4.30%	0.5081	0.0448	0.5205
0.12	135	37	81,858	876,252	42,558	835,131	39,300	41,121	8.54%	91.46%	4.44%	87.16%	4.10%	4.29%	0.5086	0.0449	0.5199
0.11	131	31	82,020	876,090	42,589	835,000	39,431	41,090	8.56%	91.44%	4.45%	87.15%	4.12%	4.29%	0.5090	0.0451	0.5193
0.10	134	57	82,211	875,899	42,646	834,866	39,565	41,033	8.58%	91.42%	4.45%	87.14%	4.13%	4.28%	0.5096	0.0452	0.5187
0.09	162	50	82,423	875,687	42,696	834,704	39,727	40,983	8.60%	91.40%	4.46%	87.12%	4.15%	4.28%	0.5102	0.0454	0.5180
0.08	191	62	82,676	875,434	42,758	834,513	39,918	40,921	8.63%	91.37%	4.46%	87.10%	4.17%	4.27%	0.5110	0.0457	0.5172
0.07	208	54	82,938	875,172	42,812	834,305	40,126	40,867	8.66%	91.34%	4.47%	87.08%	4.19%	4.27%	0.5116	0.0459	0.5162
0.06	221	70	83,229	874,881	42,882	834,084	40,347	40,797	8.69%	91.31%	4.48%	87.06%	4.21%	4.26%	0.5125	0.0461	0.5152
0.05	275	93	83,597	874,513	42,975	833,809	40,622	40,704	8.73%	91.27%	4.49%	87.03%	4.24%	4.25%	0.5136	0.0465	0.5141
0.04	345	114	84,056	874,054	43,089	833,464	40,967	40,590	8.77%	91.23%	4.50%	86.99%	4.28%	4.24%	0.5149	0.0468	0.5126
0.03	500	145	84,701	873,409	43,234	832,964	41,467	40,445	8.84%	91.16%	4.51%	86.94%	4.33%	4.22%	0.5167	0.0474	0.5104
0.02	718	193	85,612	872,498	43,427	832,246	42,185	40,252	8.94%	91.06%	4.53%	86.86%	4.40%	4.20%	0.5190	0.0482	0.5073
0.01	1,795	459	87,866	870,244	43,886	830,451	43,980	39,793	9.17%	90.83%	4.58%	86.68%	4.59%	4.15%	0.5245	0.0503	0.4995
0.00	830,451	39,793	958,110	0	83,679	0	874,431	0	100.00%	0.00%	8.73%	0.00%	91.27%	0.00%	1.0000	1.0000	0.0873
Casos por clase	874,431	83,679															
Total de casos en el conjunto: N	958,110																

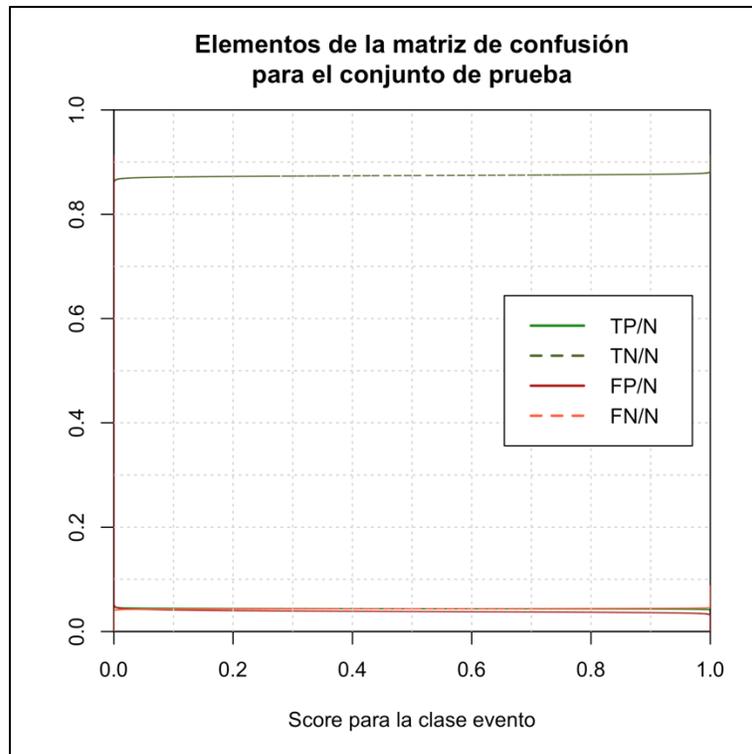


Figura 6: Elementos de la matriz de confusión para el conjunto de prueba, en función del score de corte.

La elección de un *score de corte* debe hacerse considerando el equilibrio entre estas métricas; las cuales se exponen de manera gráfica en la Figura 7, que presenta a la curva *ROC*; y en la Figura 8, que muestra a la curva de *PR*. En ambas, el *score de corte* de cada punto en la figura se indica con la escala de colores. En primera instancia, se observa que la transición entre *scores* de corte bajos (<0.1) a altos (>0.9) es muy breve, y se concentra en un espacio muy limitado tanto en el espacio de *ROC* como en el de *PR*. Esto se debe a la estructura particular de la matriz de confusión, y la manera en la que se distribuyen los casos según su *clase* y el *score* que les otorga el modelo.

Lo anterior indica que el modelo de clasificación implementado es robusto con respecto al *score de corte*, ya que, en términos de los elementos de la matriz de confusión, estos valores permanecen prácticamente constantes en los valores medios (de 0.1 a 0.9) del *score de corte*. Esta situación se observa en la Figura 6 y en la Tabla 5. En la Tabla 5, es notable que la cantidad de casos en cada *clase* (segunda y tercera columnas) cuentan con aglomeraciones en los extremos del *score de corte*. Estas aglomeraciones de varias decenas de miles de elementos son muy grandes a comparación de los valores en el resto de las filas. Por otra parte, al formar las curvas de *ROC* y *PR* se parte de una matriz de confusión calculada caso por caso, ordenando los casos según el *score* que les otorga el clasificador para trazarlos en la curva.

La implicación práctica de lo anterior es que, al momento de trazar las curvas de *ROC* y *PR*, primero se colocan los puntos de las aglomeraciones de casos en los extremos de la Tabla 5. Al ir recorriendo el *score de corte* caso por caso, cada uno de ellos altera en cierta medida los valores de *TPR*, *FPR*, y *Precision*, según su *clase* y clasificación; pero ya que su *score* puede ser muy similar al del caso anterior, tienen poco impacto en el *score de corte* de la Tabla 5. Por lo anterior, muchos casos se encuentran aglomerados en los valores extremos del *score de corte*. Estos casos aglomerados, con un *score* muy similar, son la causa de que la transición entre *scores de corte* altos y bajos sea tan breve en ambas curvas, aun cuando el trazo de la curva pasa por un amplio rango de valores.

En la curva de *PR* (Figura 8) se observa otro fenómeno, también relacionado con la manera en la que se traza la curva a partir de los casos ordenados por su *score*. En el extremo inferior del *Recall*, entre los valores de 0.0 a poco más de 0.4 hay varios picos en la forma de la curva, más pronunciados hacia el extremo inferior del *Recall*. Estas fluctuaciones se originan al recorrer los casos de uno en uno para trazar la curva. Los picos se dan cuando se recorren varios casos seguidos de la *clase ruido*, ubicados según su *score*, entre casos de la *clase evento*. Estos casos mezclados tienen un impacto en la *Precision*, disminuyendo su valor hasta que se recorren otros casos positivos, de la *clase evento*; los cuales se reflejan en los picos aserrados de la gráfica.

Es necesario aclarar que si se intenta reconstruir la Figura 7 o la Figura 8 a partir de la Tabla 5 habría que desglosar todos los casos aglomerados para poder recrear las figuras, de lo contrario sólo se obtendrán algunos puntos aislados en el espacio *ROC* o *PR*. La matriz de confusión fue condensada en la forma de la Tabla 5 para mantener una longitud aceptable.

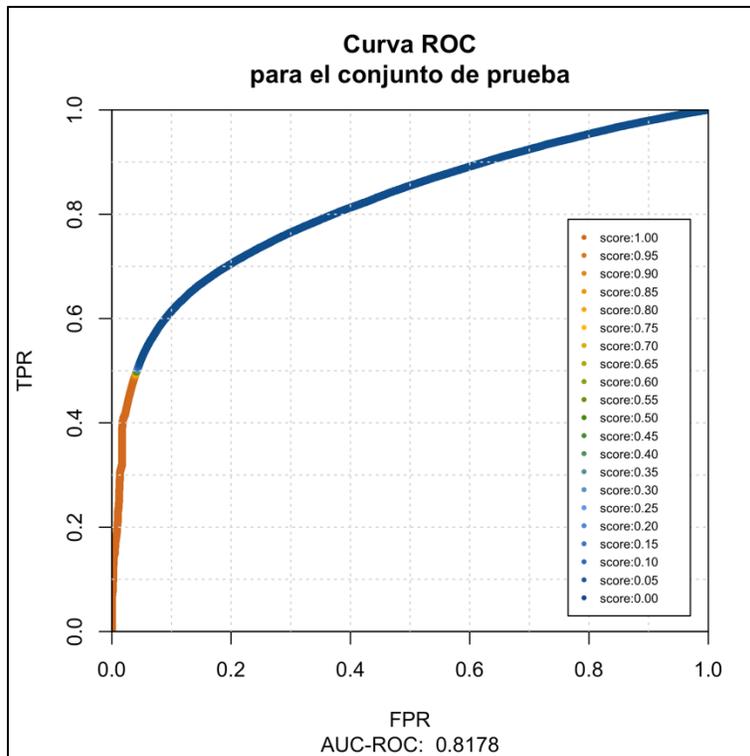


Figura 7: Curva de ROC para el conjunto de prueba, para el modelo clasificador propuesto. El score de corte de cada punto de la curva se indica con la escala de colores.

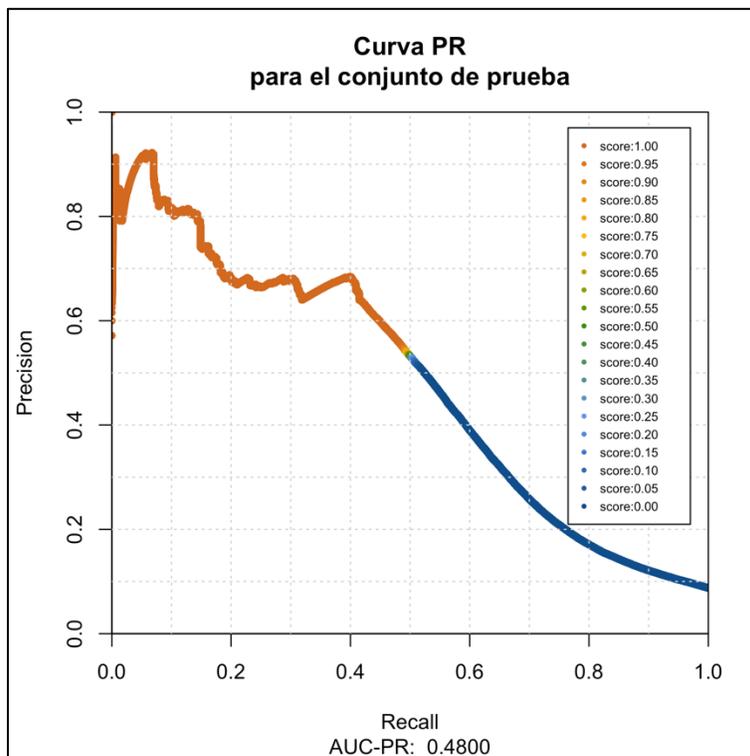


Figura 8: Curva de PR para el conjunto de prueba, para el modelo clasificador propuesto. El score de corte de cada punto de la curva se indica con la escala de colores.

En la Figura 7 y en la Figura 8 se incluyen las áreas bajo las curvas, las cuales se replican en la Tabla 6. Recordando que estas métricas toman valores en el rango de cero a uno, y que un clasificador perfecto tiene valores de *AUC* de 1.0, se nota que el rendimiento del clasificador propuesto, según la curva *ROC* es bastante bueno. Sin embargo, parece que el rendimiento según la curva *PR* deja algo que desear. Esta percepción se disipa cuando se interpretan las dimensiones de cada figura de manera conjunta.

Tabla 6: Áreas bajo las curvas *ROC* y *PR*, para el conjunto de entrenamiento.

<i>Curva</i>	<i>Área bajo la curva (estimación)</i>
<i>ROC</i>	0.8178
<i>PR</i>	0.4800

Por ejemplo, en el punto de la curva *ROC* donde $TPR=0.400$, $FPR=0.017$, el modelo clasifica bien al 40% de los casos de la *clase evento*, pero se equivoca al clasificar a 1.7% de los casos de la *clase ruido*. Desplazándonos sobre la curva, hacia arriba de la figura, aumentamos la cantidad de casos de *clase evento* que identifica correctamente el clasificador, a costa de irlos revolviendo con más y más casos de la *clase ruido*.

La curva de *PR* se emplea de una forma similar. Si analizamos la curva por su extremo derecho, donde $Recall=1.000$ y $Precision=0.0873$; encontramos que, en este punto, el clasificador está clasificando correctamente al 100% de los casos de *clase evento*, pero de los casos marcados como tal, sólo 8.73% son verdaderos positivos. Recordemos que este punto corresponde a un *score de corte* de 0, por lo que equivale a decir que todos los casos son de la *clase evento*. Si nos desplazamos sobre la curva hacia la izquierda, disminuye la cantidad de casos de la *clase evento* que clasificamos bien, pero reducimos la cantidad de casos de la *clase ruido* que se clasifican mal.

Supongamos que, antes de tener un modelo de clasificación, etiquetamos a todos los datos como si fueran de la *clase evento*. Esto se corresponde al punto analizado en el párrafo anterior, bajo las métricas de *Recall* y *Precision* mencionadas. Ahora, si introducimos el modelo de clasificación propuesto, podemos mejorar la *Precision* de nuestra clasificación eligiendo un punto sobre la curva de *PR*, llegando a multiplicar la métrica hasta por casi nueve veces si elegimos uno de los puntos ubicados cerca del extremo izquierdo de la curva.

En este momento deben de hacer consideraciones prácticas sobre los resultados. Hay que preguntarse si lo que se desea es conservar la mayor cantidad de *eventos* posible, sin revolverlos con tanto *ruido*; o si lo que nos interesa es aislar el *ruido*, con la menor cantidad de *eventos* mal

clasificados que sea posible. Con la ayuda de las curvas de *ROC* y *PR*, se puede elegir un *score de corte* que nos dé un rendimiento de acuerdo con nuestros intereses.

8.2. Elección de un score de corte

Considerando que el enfoque de este trabajo sólo se limita a separar ruido de eventos, sin un énfasis preferente por unos u otros, se elige un *score de corte* para la *clase evento* de 0.75017. En este punto tenemos el rendimiento presentado en la Tabla 7.

Tabla 7: Métricas de rendimiento para el score de corte elegido. A diferencia de la Tabla 5, no se han agrupado los casos según su score de corte redondeado a dos decimales.

Score de corte para la clase evento	Casos clasificados como evento	Casos clasificados como ruido	TP: Verdaderos positivos	TN: Verdaderos negativos	FP: Falsos positivos	FN: Falsos negativos	Casos clasificados como evento / N	Casos clasificados como ruido / N	TP: Verdaderos positivos / N	TN: Verdaderos negativos / N	FP: Falsos positivos / N	FN: Falsos negativos / N	TPR = Recall	FPR	Precisión
0.75017	77,181	880,929	41,487	838,737	35,694	42,192	8.06%	91.94%	4.33%	87.54%	3.73%	4.40%	0.4958	0.0408	0.5375

La Tabla 7 se extrajo directamente de la matriz de confusión desagregada y sus métricas de rendimiento para un caso con *score* de 0.75017; a diferencia de la Tabla 5 que se recalculó a partir de un agrupamiento de los casos según su *score de corte*. Por esta razón se omiten las columnas que indican la cantidad de casos por cada *clase* según el *score*. En la Tabla 7, todos los casos con *score* mayor o igual al indicado se consideran *evento* ($TP + FP = 77,181$ casos); mientras que aquellos con un *score* menor se clasifican como *ruido* ($TN + FN = 880,929$ casos). Aunque sólo clasificamos el 8.06% de las ventanas como *evento*, dentro de esta selección estamos clasificando bien al 49.58% de los casos con *clase evento* presentes en el conjunto de prueba. De forma similar, dentro de la selección sólo estamos clasificando (mal) como *evento* a 4.08% de los casos de *clase ruido* del conjunto de prueba; mientras que 53.75% de los casos en la selección sí tienen una *clase de evento*.

Los 77,181 casos, o ventanas, seleccionados como *positivos* a este *score de corte* se transforman en aproximadamente 98,793 segundos de registros; abarcando un día, 3 horas, y 26 minutos con 32 segundos de duración.

8.3. Figuras de los registros clasificados

Al sobreponer las ventanas clasificadas como *evento* sobre los sismogramas preprocesados se observa con claridad la acción clasificadora del modelo. Todas las figuras formadas por la

sobreposición de las ventanas a los registros se incluyen en la sección de Sismogramas clasificados, en los anexos; pero por su importancia a continuación se discute un par de ellas, ya que se consideran representativas del trabajo y sus resultados.

Primero, se presenta la Figura 9, que muestra al registro *GOLO 323-2*, el cual forma la segunda sección resultante del preprocesamiento del registro *GOLO 323*. Sobre el registro se han indicado, con color rojo, aquellas ventanas que tienen un *score* para la clase *evento* mayor o igual a 0.75017.

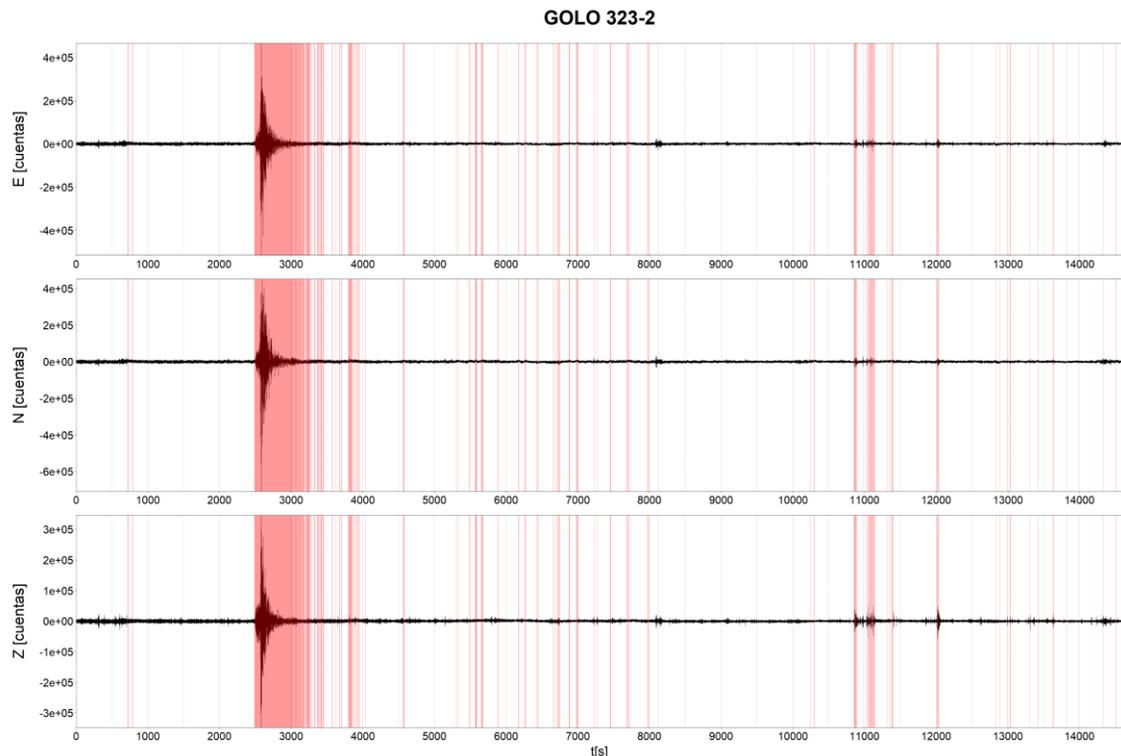


Figura 9: Registro *GOLO 323-2*, preprocesado. Sus ventanas con *score* para la clase evento mayor o igual a 0.75017 se han resaltado en rojo.

Como se puede ver a los 2,500 segundos del registro, el método de clasificación propuesto ha detectado un sismo, el cual da amplitudes de $\pm 4 \times 10^5$ a las tres componentes. Pasados los 3,000 segundos, se interrumpe la continuidad de las ventanas marcadas en la coda del sismo, hasta que se dejan de marcar ventanas entre los 3,500 y 4,000 segundos del registro. En este mismo intervalo hay un evento transitorio pasados los 3,750 segundos, el cual también se ha detectado. Para las ventanas resaltadas localizadas entre los 3,250 y 3,750 segundos, es complicado decidir hasta cuál instante se deben a la coda del sismo (ocurrido a los 2,500 segundos) o a otros eventos subsecuentes (réplicas); o incluso a falsos positivos.

Esta misma complicación es encontrada por el modelo de clasificación, al resaltar ventanas que aparentemente no muestran algún fenómeno en dicho intervalo y el resto del registro. Sin

embargo, hay que considerar que el modelo trabaja sobre los atributos indicados en la Tabla 2, y no sobre las componentes cartesianas de la Figura 9, por lo que el modelo puede revelar detalles que no son inmediatamente visibles.

De la misma forma, el que la selección manual de las ventanas en *evento* o *ruido* se haya realizado sobre el registro transformado a componentes esféricas podría tener una influencia al momento de entrenar al modelo, y con ello afectar al *score* de estas ventanas sin causa aparente. Esta situación también se da en otros de los registros trabajados, como *GOLO 297-2* (Figura 56), o *GOLO 296* (Figura 54); ambos expuestos en los anexos correspondientes.

Dentro de la Figura 9, el intervalo de 4,000 a 8,000 segundos tiene poca actividad, aunque en ciertos segmentos se resaltan varias ventanas sin alguna actividad evidente en el registro; lo cual se atribuye a las causas expuestas en el párrafo anterior. A poco más de 8,000 segundos, se observa en las tres componentes un fenómeno transitorio, que no es marcado por el modelo al *score de corte* elegido. Otros fenómenos de amplitud similar pueden observarse a partir de los 4,000 segundos, aunque no siempre tienen una proyección clara en las tres componentes.

Lo anterior indica que la amplitud de estos fenómenos no es la suficiente como para cambiar el valor de los atributos (respecto al valor de los atributos de las ventanas con ruido), y con ellos el *score* para la clase *evento* de las ventanas que les corresponden. En otras palabras, son muy pequeños como para que sean detectados al *score de corte* elegido. Aunque podría elegirse un *score de corte* para la *clase evento* más bajo en un intento de marcarlos, esto se haría a costa de marcar a más falsos positivos y de perder el refinamiento logrado con el método. Nuevamente, esta situación se puede observar en otros de los registros, como *GOLO 302-2* (Figura 66) o *GOLO 304-3* (Figura 71); también incluidos en los anexos correspondientes.

Para finalizar la discusión de la Figura 9, se comentan las ventanas marcadas entre los 10,750 a 11,250, y entre 12,000 y 12,250 segundos. Los fenómenos marcados entre estos intervalos tienen la característica de mostrarse en todas las componentes, y de alcanzar una mayor amplitud en la componente vertical (Z) respecto a los fenómenos entre 4,000 y 8,000 segundos. Por ello, estos fenómenos sí son marcados por el método, a diferencia de los del párrafo anterior. La situación expuesta en los dos últimos párrafos se ve en el resto de las figuras: aquellos fenómenos con una amplitud reducida pueden ser ignorados al *score de corte* elegido; pero intentar marcarlos disminuyendo el *score de corte* puede resultar contraproducente.

La segunda figura por discutir es la correspondiente al registro *GOLO 323-4*; el cual corresponde a la última sección del registro *GOLO 323*, una vez que fue preprocesado. Este se muestra como la Figura 10, a continuación.

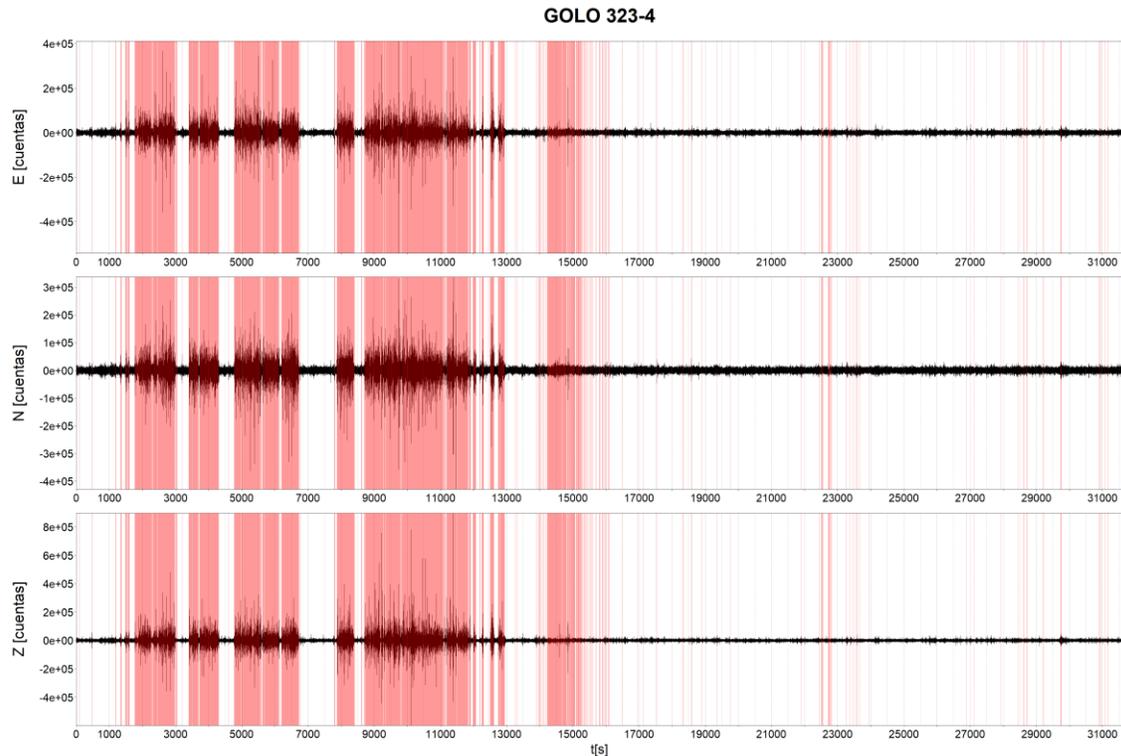


Figura 10: Registro *GOLO 323-4*, preprocesado. Sus ventanas con score para la clase evento mayor o igual a 0.75017 se han resaltado en rojo.

En la Figura 10, a partir de los 17,000 segundos el registro tiene un aspecto similar al de la Figura 9 una vez que ha pasado el sismo (entre los 4,000 y 10,000 segundos de aquella figura); por lo que no se comentará más al respecto. El resto del registro, de los 1,000 a los 16,000 segundos resulta mucho más interesante, en particular por que en ese intervalo se llevó a cabo la demolición de la casa colapsada mencionada en la sección de Datos disponibles (comunicación personal con el Dr. Martín Cárdenas Soto, 21 de enero de 2019). Las percusiones provocadas por las obras de demolición se observan como intervalos con una gran amplitud en todas las componentes, alcanzando amplitudes de hasta $+8 \times 10^5$ cuentas en la componente vertical. El modelo de clasificación al *score de corte* elegido no tiene problemas para identificar y delimitar estos fenómenos, como se observa en las ventanas marcadas en la Figura 10.

En conjunto con los resultados mostrados en la Figura 9, lo anterior representa una identificación exitosa de los eventos sísmicos en un registro, ya sean de fuente natural o artificial. Otros eventos se pueden observar en las figuras de los registros *GOLO 321-4* (Figura 88) y *GOLO*

322-3 (Figura 91), asemejándose a otro conjunto de percusiones por demolición y a un sismo, respectivamente.

En otros registros, como *GOLO 298-2* (Figura 58) o *GOLO 303-3* (Figura 67), entre otros; al *score de corte* elegido se marcan ciertos fenómenos de baja frecuencia como si fueran *eventos*. La causa más probable para estos fenómenos sea un movimiento físico del instrumento por personal no capacitado en su manejo.

Por último, hay que resaltar que el *score de corte* para la *clase evento* elegido, con un valor de 0.75017; es tan sólo una propuesta. Para otro enfoque particular de discriminación de eventos sobre ruido, o viceversa, será conveniente tomar un *score de corte* propio para obtener los resultados más apropiados para el problema.

9. Conclusiones

En el desarrollo de este documento se ha presentado un método de discriminación entre eventos y ruido sísmico, basado en las técnicas generales de minería de datos; acoplado a un preprocesamiento convencional para sismogramas triaxiales. Al apoyarse en este método se separa la tarea de decidir si un segmento de un sismograma es ruido o un evento sísmico; de la forma particular del modelo de clasificación. Lo anterior establece una cimentación sobre la cual se pueden probar otros modelos de clasificación.

A pesar de que el modelo de clasificación normal multivariado es sencillo, con tan sólo siete parámetros (nueve considerando la longitud de la ventana y el empalme ocupados durante el preprocesamiento, por lo que no son intrínsecos al modelo); se logra separar una cantidad significativa de casos con *clase ruido* de aquellos de *clase evento*. El modelo y el *score de corte* elegidos aumentan la densidad de casos de la *clase evento* en el conjunto de prueba, de 8.19% a 49.58%, tras aplicar el modelo de clasificación presentado; con un *score de corte* elegido como un compromiso para no enfocar los resultados hacia la separación de una *clase* en particular.

Es importante recordar que el *score de corte* nos permite orientar el rendimiento de la técnica hacia la depuración de *ruido* para mantener los *eventos*, o viceversa. Lo anterior amplía las posibles aplicaciones de la técnica, al poder obtener a partir de los datos originales un conjunto más reducido, pero más adecuado para diferentes análisis; sólo con tomar *scores* de corte diferentes a partir de la respuesta del modelo. No es posible maximizar todas las métricas de rendimiento a la vez, por lo que hay que aceptar que no se pueden obtener resultados perfectos, tan sólo unos que pueden llegar a ser significativamente mejores que los datos con los que contamos de origen.

Al analizar los resultados, se esperaba que el modelo propuesto alcanzara rendimientos más altos, medidos por cualquiera de las métricas de *TPR*, *FPR* o *Precision*. Estas cotas de rendimiento no se alcanzan al haber casos que, al momento de clasificar manualmente, no pudieron ser asignados con certeza a alguna de las dos *clases*. En otras palabras, les fue asignada de forma manual una *clase* sin tener la certidumbre necesaria. La consecuencia de esto es que hay rangos de los atributos en los que no es posible discriminar entre las *clases*; los cuales se observan en la Figura 2 como aquellas partes de las “campanas” que se empalman. Por esta razón, no hay modelo de clasificación que pueda separar fidedignamente a los casos cuyos atributos caen en estos valores.

Ya que se conoce que la responsabilidad de esta situación recae en quien realizó la clasificación manual, quien es el sustentante; se sugiere que en trabajos futuros haya un mayor control al realizar esta actividad, ya que de este paso dependen los resultados del modelo. Esto pone en evidencia que los resultados del método, así como las conclusiones que se deriven de él, son tan buenos como los datos de los que proceden. La alternativa de volver a clasificar manualmente catorce días de registros sísmicos no es una opción viable en este momento, dada la cantidad de horas-hombre necesarias para la tarea.

A pesar de ello, el método sigue mostrando su viabilidad y robusticidad bajo esta circunstancia. Sólo hace falta ver la matriz de confusión en la Figura 6 y sus valores casi constantes a través del *score de corte*; o las figuras de los sismogramas con la clasificación sobrepuesta (sección de Sismogramas clasificados, en los anexos). La simplicidad de la forma analítica del modelo, junto con la facilidad con la que se entrena; usando sólo la matriz de covarianza, el vector de valores medios y la proporción de casos por cada *clase*, junto con la cantidad de variables; permite entrenar el modelo aún bajo circunstancias desfavorables, como en la toma de los registros en un área urbana.

La separación de eventos de ruido presentada por esta técnica puede abarcar aplicaciones en un espectro amplio, con usos tan determinantes como la elección de los segmentos de un sismograma a almacenar permanentemente; como en formas más sutiles, por ejemplo, al depurar los segmentos de los registros que menos nos interesan como un paso previo a una inspección y una nueva clasificación manual o automática de los datos, pero esta vez en un conjunto mucho más acotado.

También, resulta interesante que sólo haya sido necesario conservar las bandas de bajas frecuencias y de la señal sin filtrar para poder separar las ventanas de cada *clase*, al menos bajo los atributos elegidos (Tabla 2). La interpretación que se le da a esto es que se puede conocer cuál es

una parte de la señal sin conocer la otra, por pura eliminación. Esta conjetura podría no ser válida para otros conjuntos de datos; y debe de verificarse con una mayor variedad de sismogramas.

Para finalizar, se propone como una línea de investigación a futuro el empleo de este modelo de clasificación para identificar dentro de un sismograma el segmento donde se declara un evento; con el fin de proporcionar este segmento a una nueva herramienta que realice una búsqueda intensiva de su comienzo y final dentro del sismograma. Incluso, la nueva herramienta podría ser el mismo modelo de clasificación, pero entrenado sobre un registro o tipo de evento característico; aunque no se descarta el empleo de modelos con otras formas funcionales o técnicas trabajando con atributos completamente diferentes. Un buen lugar para comenzar a trabajar bajo este esquema es en los segmentos de los sismogramas que aparentemente no muestran fenómeno o evento alguno, pero que de todas formas fueron marcados por el modelo de clasificación como si tuvieran algún evento. Por otro lado, también se manifiesta la necesidad de contar con una herramienta interactiva para la visualización de las ventanas seleccionadas en función del *score de corte*, para facilitar la elección del *score de corte* desde una perspectiva sismológica.

10. Anexos

10.1. Breviario sobre ondas sísmicas e instrumentación

Una onda sísmica es una perturbación mecánica que se propaga por la tierra. Estas perturbaciones pueden deformar de dos maneras básicas al medio; por lo que se puede distinguir entre al menos dos variedades de ondas (Stein & Wysession, 2003, pág. 57):

- Las ondas compresivas, P, o primarias: Estas ondas distorsionan al medio en la dirección en la que se propagan, estirándolo y compactándolo. Son las ondas de mayor velocidad, y por ello son las primeras que reciben desde su fuente; de ahí que se les diga primarias o P. Estas ondas se propagan tanto en materiales sólidos como fluidos; ya que el material se intenta oponer a la deformación que provocan.
- Las ondas de cizalla, corte, S, o secundarias: La distorsión causada por estas ondas es en dirección perpendicular a la de su propagación. Estas ondas causan una deformación como si las hojas de unas tijeras intentaran cortar el medio. Una parte de la onda intenta desplazar el medio hacia un sentido, y la otra en el sentido opuesto. Como los fluidos no se oponen a este tipo de deformación, estas ondas no se transmiten en ellos.

Además, se presentan fenómenos de interferencia entre las ondas ya mencionados y la superficie del medio, para nuestro caso la superficie terrestre (Stein & Wysession, 2003, pág. 86):

- Las ondas Love, descubiertas por A. E. H. Love; se dan como un fenómeno de interferencia entre la superficie de la tierra y ondas de corte con deformación en sentido paralelo a la superficie.
- Las ondas Rayleigh, descubiertas por J. W. Strutt, de título Lord Rayleigh; se dan con la interferencia entre la superficie, las ondas compresivas, y las ondas de corte con deformación perpendicular a la superficie.

Las dos ondas mencionadas con anterioridad son los fenómenos de interferencia más conocidos, pero hay otros que se dan en condiciones más particulares; como en el límite entre un medio líquido y sólido, o con una geometría especial de la superficie. Para dar una impresión más clara de los desplazamientos que causan estas ondas; se incluye la Figura 11.

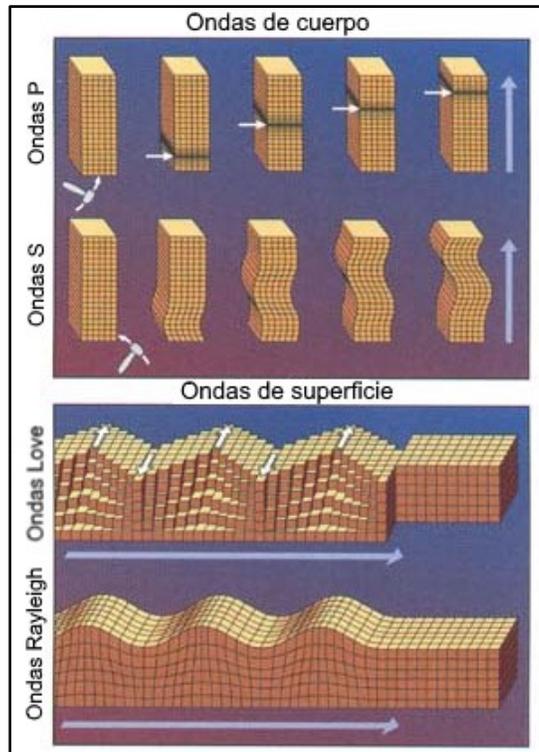


Figura 11: Desplazamientos causados por las ondas P, S, Love, y Rayleigh. Las ondas de cuerpo (P y S) se dan dentro del medio, y en este diagrama se propagan desde abajo hacia arriba. Las ondas de superficie se dan en la superficie del medio, y en el diagrama se propagan de izquierda a derecha. Adaptación de una imagen del dominio público (trabajo del United States Geological Survey), disponible en (Wikimedia Commons, 2004).

El objetivo de cualquier *sismómetro* es transformar el movimiento del instrumento, causado por las ondas sísmicas, en una señal que pueda ser analizada. Los instrumentos más convencionales basan su funcionamiento en el principio de inercia (Havskov & Alguacil, 2004, pág. 2), el cual explica que una masa tiende a permanecer en reposo o en movimiento rectilíneo en ausencia de fuerzas que actúen en ella. Esto se puede enunciar de manera parecida a la Primera Ley de Newton. En términos prácticos, los sismómetros cuentan con una masa suspendida, la cual tiende a permanecer en reposo mientras el instrumento, fijo al suelo, se mueve al paso de las ondas sísmicas.

La masa y el principio de inercia nos permiten observar el movimiento del suelo respecto a la masa, pero si no hubiera alguna forma de aplicar una fuerza contraria al movimiento de la masa, ésta se seguiría moviendo de forma indefinida. Por esto es necesario introducir una forma de amortiguamiento, para disipar el movimiento de la masa, y que vuelva a un estado de reposo. Dada esta necesidad, todos los sismómetros se pueden reducir a un sistema de masa-resorte amortiguado, como el presentado en la Figura 12. En los instrumentos modernos, este

amortiguamiento se da de forma electromagnética mediante bobinas e imanes incorporados a la masa y su montura.

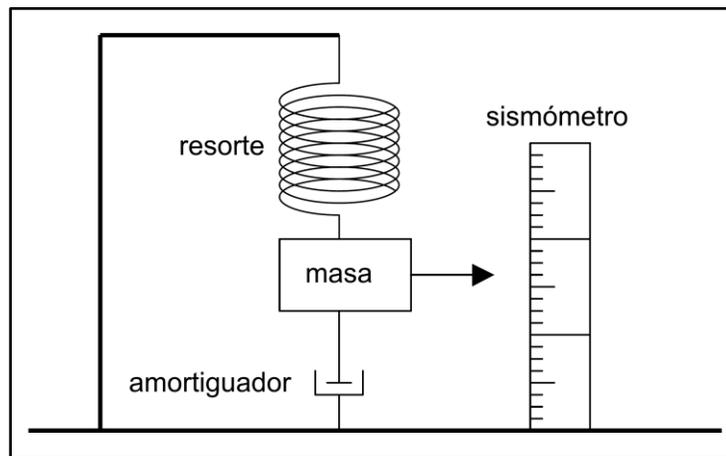


Figura 12: Sistema masa-resorte amortiguado. La masa se suspende por medio de un resorte, y un amortiguador disipa su movimiento. Cuando la montura de la masa se mueve, la masa tiende a permanecer estática. El movimiento entre la posición en reposo de la masa y la montura se puede leer con la aguja (flecha) y la escala a la derecha. Adaptada de (Havskov & Alguacil, 2004, pág. 12)

El funcionamiento básico de este sistema se describe en (Havskov & Alguacil, 2004, págs. 12-13) de la siguiente manera: Partiendo de un estado de reposo, si sucediera un movimiento del suelo de forma súbita la masa tenderá a permanecer estática, mientras el resto del instrumento se mueve con el suelo. La magnitud del desplazamiento de la masa respecto al suelo podrá ser medido directamente. Si ocurriera un movimiento sinusoidal rápido, es decir, de alta frecuencia; la masa permanecería estática, mientras el instrumento y el suelo se sacuden. Nuevamente, se podría medir el desplazamiento de manera directa como la distancia entre la masa y su montura. Sin embargo, si sucediera un movimiento lento, de baja frecuencia; la masa tendría tiempo de moverse con su montura y el suelo, por lo que el movimiento total ya no se podría medir directamente entre la masa y su montura. Además, la distancia entre masa y montura tenderá a ser la del sistema en reposo mientras más lento sea el movimiento. Entre estos dos extremos, puede imaginarse un movimiento sinusoidal que mueva al instrumento justo cuando la masa se encuentra al tope de su recorrido, “como empujar un columpio justo antes de que empiece a bajar” (óp. cit. pág. 13.). Esta es la frecuencia de resonancia del sistema, y de no ser por el amortiguamiento, la masa permanecería oscilando al cesar el movimiento.

De la situación planteada en el párrafo anterior, se extrae que a frecuencias bajas la respuesta del sistema, dada por el movimiento de la masa respecto al suelo; es proporcional a la *frecuencia* del desplazamiento de él. A frecuencias altas la amplitud de la respuesta es proporcional a su

desplazamiento. En términos del análisis armónico, a esto se le conoce como la función de transferencia del sistema, que relaciona las señales de entrada y salida del sistema; estando planteadas estas señales como una combinación de funciones armónicas.

Hasta ahora se ha hablado de cómo el sismómetro obtiene una señal de salida a partir del movimiento del suelo; pero antes de poder hacer un uso práctico de la señal hay que guardarla de alguna manera. El instrumento encargado de esto es el *registrador*. Antes de la proliferación de las computadoras y electrónica digital, el registro se hacía de manera analógica trazando físicamente la señal de salida en rollos de papel (blanco, ahumado o fotográfico). Con la introducción de la electrónica digital, el almacenamiento y procesamiento de los registros se vuelve más sencillo. Aun así, los sismómetros, como el resto de la naturaleza, funcionan de manera analógica. Ya sea mediante el desplazamiento mecánico directo de una pluma, aguja, o haz de luz; o con la corriente inducida en las bobinas de un sismómetro, la salida del sismómetro es un fenómeno físico, analógico; y requiere convertirse a un formato digital para almacenarse y trabajarse de esa manera. El instrumento responsable es el *digitalizador*, el cual también tiene su propia función de transferencia. Con la miniaturización de la electrónica, es posible combinar sismómetro, digitalizador, y registrador en un mismo aparato; y su fabricante puede proporcionar la respuesta instrumental de todo el sistema.

10.2. ¿Qué es el ruido sísmico?

Dentro de un contexto de análisis de señales, el ruido es todo aquello que no nos interesa y ofusca la señal, el objeto de nuestro interés. Por razones de desarrollo histórico, en sus comienzos la Sismología empezó por estudiar los sismos: las señales más impactantes que pueden registrarse de la tierra sólida. Por ello, se distinguen a los *eventos sísmicos*, de los cuales los sismos forman parte, del *ruido sísmico*. Mientras que los sismos se empezaron a estudiar de forma determinística, el ruido permaneció incomprensible y se mostraba como un obstáculo al análisis de los eventos; ya que oculta a aquellos de menor amplitud. Por ello, se le denominó ruido. Si asumimos que una señal sísmica, o *sismograma*, sólo se compone de ruido y eventos, los eventos deben de ser todo aquello que no sea ruido. Es preciso comentar que los instrumentos que registran estas señales también insertan cierto ruido dentro de la señal que registran, pero ahondar en esta situación se sale del enfoque de este trabajo; por lo que se asumirá a este componente del ruido, el *ruido instrumental*, como despreciable respecto a las demás componentes de la señal.

La distinción entre ruido y señal (o eventos, para este caso) depende de la disciplina en particular, ya que en la actualidad el ruido, y no los eventos, es la materia prima para varias técnicas

sismológicas con diversos objetivos. Como mencionan (Bonney-Claudet, Cotton, & Bard, 2006, pág. 209); clasificar al ruido y sus causas es complicado. Para ellos (*loc. cit.*) el ruido “es un término genérico para denominar a las vibraciones ambientales del suelo, provocadas por fuentes como las mareas, olas rompiendo en las costas, el viento y sus efectos en los objetos de la superficie, maquinaria industrial, vehículos, pasos humanos, etcétera.”. A partir de los trabajos de (Gutenberg, 1958), (Asten M. W., 1978) y (Asten & Henstridge, 1984) los mismos autores resumen las fuentes de acuerdo a su frecuencia y las presentan en una tabla (Bonney-Claudet, Cotton, & Bard, 2006, pág. 209), cuya adaptación se presenta como la Tabla 8.

Tabla 8: Resumen de fuentes de ruido ambiental de acuerdo con su frecuencia. Adaptada de (Bonney-Claudet, Cotton, & Bard, 2006, pág. 209); quienes revisan los trabajos de (Gutenberg, 1958), (Asten M. W., 1978) y (Asten & Henstridge, 1984) para su elaboración.

	(Gutenberg, 1958)	(Asten M. W., 1978) y (Asten & Henstridge, 1984)
Olas oceánicas impactando en las costas	0.05 – 0.1 Hz.	0.5 – 1.2 Hz.
Monzones y perturbaciones meteorológicas a gran escala	0.1 – 0.25 Hz.	0.16 – 0.5 Hz.
Ciclones sobre los océanos	0.3 – 1 Hz.	0.5 – 3 Hz.
Condiciones meteorológicas a escala local	1.4 – 5 Hz.	
Tremor volcánico	2 – 10 Hz.	
Actividad urbana	1 – 100 Hz.	1.4 – 30 Hz.

A partir de la información de la Tabla 8, (Bonney-Claudet, Cotton, & Bard, 2006, pág. 209) concluyen que en una primera aproximación, el ruido generado por la actividad humana, denominado *ruido antropogénico* o *cultural*, se presenta en frecuencias superiores a 5 Hz., y el causado por fuentes naturales a frecuencias inferiores a 1 Hz.; mientras que en la banda de 1 a 5 Hz. las fuentes son tanto naturales como artificiales. No es difícil concluir que en esta última banda la asignación de un ruido a una fuente natural o artificial puede resultar equívoca, aunque los autores (*loc. cit.*) dejan abierta, de forma implícita, la posibilidad de participar en el ruido de alta frecuencia a las fuentes naturales. Como evidencia de las causas del ruido a frecuencias mayores a 5 Hz., (*óp. cit., pág. 210*) comenta los resultados de algunos estudios donde se observa claramente una correlación entre la amplitud del ruido y la hora local del sitio de estudio, observándose cómo aumenta cuando se acerca el mediodía, y alcanza su mínima amplitud alrededor de la medianoche; con lo cual queda evidente su correlación con la actividad humana.

Considerando la dificultad para distinguir las fuentes del ruido entre naturales y antropogénicas, para el enfoque de este trabajo se considerará ruido a todo aquello que no parezca ser un evento; y a los eventos, como aquellos segmentos de una señal sísmica que tengan

características, comienzo impulsivo o emergente, periodicidad en el tiempo, amplitud sobresaliente; y en general un aspecto anómalo (respecto al ruido), que indiquen un origen posiblemente cultural o marcadamente sísmico. Esta forma de clasificar al ruido y los eventos es cualitativa y subjetiva; pero se sigue considerando, en varias ocasiones y frecuentemente de forma implícita, el estándar de referencia. Puede decirse que este estándar parte del *criterio*, el cual se basa en la *experiencia profesional*. Es un hecho que se sigue aplicando en muchas áreas de la Sismología, a falta de un método cuantitativo lo suficientemente versátil y robusto para reemplazar la flexibilidad de un intérprete humano entrenado. Como ejemplos se puede mencionar a (Peterson, 1993, pág. 11), quien al estudiar las características globales del ruido sísmico se encontró con esta necesidad de separar eventos del ruido, respecto a lo cual sólo comenta que “los sismogramas fueron visualmente revisados para buscar la presencia de eventos pequeños, brotes de ruido, u otros tipos de problemas instrumentales”; y a los trabajos citados en la sección de Antecedentes.

10.3. Métricas para la selección de atributos

No hay una regla general para seleccionar un atributo o variable sobre otro, así que la aplicación de las métricas presentadas a continuación depende del contexto del problema y el criterio de quien lo trabaja. Hay que mantener en proporción la cantidad de atributos y la de casos a trabajar para mitigar la maldición de la dimensionalidad (Hernández Orallo, Ramírez Quintana, & Ferri Ramírez, 2004, pág. 79). Las métricas expuestas a continuación fueron aplicadas para depurar la gran cantidad de atributos resultantes del preprocesamiento.

10.3.1. Coeficiente de correlación entre dos variables continuas

La forma más sencilla de descartar atributos formados por variables continuas es retirando aquellas que resulten redundantes respecto a otra, con un comportamiento aproximadamente lineal, de tendencia contraria o similar a la otra variable comparada. La intensidad de la relación lineal entre los vectores de dos variables continuas se puede cuantificar con el *coeficiente de correlación* de Pearson, mostrado en la expresión [5] (Meyer, 1973, pág. 148), donde \mathbf{X} y \mathbf{Y} son los vectores de las variables a comparar.

$$\rho_{X,Y} = \frac{E\{[\mathbf{X} - \bar{X}][\mathbf{Y} - \bar{Y}]\}}{\sqrt{V(\mathbf{X})V(\mathbf{Y})}} \quad [5]$$

En el numerador se encuentra la covarianza entre estas variables, y en el denominador el producto de la desviación estándar de cada una. Si el coeficiente entre dos atributos es, en valor absoluto, 1; hay una relación lineal entre ambas variables, con el signo dado por aquél del

coeficiente de correlación. Para fines prácticos, se escoge un umbral menor (en valor absoluto) para el coeficiente de correlación permitido, retirándose una variable de cada par que exceda el umbral. Al retirar una de las variables se rompe la dependencia lineal, por sí solo no tiene sentido retirar ambas. Sin embargo, el contexto del problema puede indicar las variables que se prefiere conservar, al ser las más fáciles de calcular, o con algún otro atractivo.

10.3.2. Adaptación del coeficiente de correlación a una variable continua y otra nominal

Al trabajar un problema de clasificación, es recomendable explorar la relación entre los atributos continuos; y la *clase*, que es una variable de tipo nominal. De acuerdo a (Frank, Hall, & Witten, 2016) y (University of Waikato, 2012), para obtener la correlación entre una variable continua X , y una variable nominal Y con k clases diferentes, ambas con n muestras, se procede de la siguiente manera:

1. Para cada *clase* j de Y , se forma una variable indicadora, o *dummy*, Y_j . Esta variable vale uno cuando el valor de un elemento de Y coincide con la *clase* a indicar, y cero cuando no.
2. Se cuenta cuantas muestras n_j hay con cada *clase* j de Y
3. Debe obtenerse el coeficiente de correlación (expresión [5]) entre la variable continua X y cada una de las Y_j ; al cual representaremos como ρ_{X,Y_j}
4. El coeficiente de correlación entre X y Y estará dado por el promedio de los ρ_{X,Y_j} , ponderado por la cantidad de casos en cada *clase* de Y , como muestra la expresión [6] (*óp. cit.*).

$$\rho_{X,Y} = \frac{\sum_{j=1}^k n_j \rho_{X,Y_j}}{\sum_{j=1}^k n_j} \quad [6]$$

Al ponderar los coeficientes de correlación con la cantidad de casos se considera que los valores de Y pueden estar distribuidos de una forma no uniforme entre las *clases*. Para fines prácticos, este cálculo se efectúa para todas las variables continuas, y se eligen aquellas con los mayores valores de $\rho_{X,Y}$. Nuevamente, puede haber consideraciones particulares al problema sobre las variables a escoger.

10.3.3. Selección de variables por medio de teoría de la información.

Otras métricas para la selección de variables se desprenden de considerar la información contenida en cada variable o atributo. El concepto más básico de la teoría de la información es la *cantidad de información* de un evento genérico (Abramson, 1981, pág. 25); el cual *no debe de*

confundirse con los *eventos sísmicos* de este documento. Si el evento tiene una probabilidad $P(E)$ de suceder, la cantidad de información $I(E)$ que se asocia al evento se muestra en la expresión [7]. Esta cantidad siempre es mayor o igual a cero, y según la base del logaritmo se definen sus unidades. Para el \log_2 las unidades son *bits* (*óp. cit.*, p. 26).

$$I(E) = \log\left(\frac{1}{P(E)}\right) \quad [7]$$

Otro concepto necesario es el de una *fente de información* de memoria nula, siendo ésta un ente S que emite *símbolos* con cierta probabilidad constante (*óp. cit.*, p. 27). La emisión de cierto símbolo s_i se corresponde a un evento, por lo que cada símbolo tiene asociada una cantidad de información. La cantidad media de información emitida cada vez que la fuente emite un símbolo es la *entropía de la fuente S* , mostrada en la expresión [8] (*ibíd.*). Esta cantidad también es mayor o igual a cero, y hereda las unidades de la cantidad de información (*óp. cit.*, p. 28-31).

$$H(S) = \sum_s P(s_i) \log\left(\frac{1}{P(s_i)}\right) \quad [8]$$

Interpretando las expresiones anteriores, se puede pensar en la cantidad de información como “la información necesaria para que la presencia de [el símbolo] s_i sea cierta” (*óp. cit.*, p. 29). Observando la expresión [7] se nota que los símbolos más probables tendrán una menor cantidad de información, mientras que los menos probables tendrán una mayor. Por ello, una fuente que emita símbolos con probabilidad uniforme maximizará su entropía, mientras que una fuente que siempre emita el mismo símbolo tendrá una entropía de cero. (*óp. cit.*, p. 31-32). De ahí se desprende la interpretación de $H(S)$ como la incertidumbre de la fuente. Si la fuente emite siempre el mismo símbolo, tendremos certeza sobre los símbolos que emite.

Para continuar, se introduce otro concepto, el de un *canal de información*. Un canal puede visualizarse como un objeto que toma un símbolo a_i de un *alfabeto A* y entrega otro símbolo b_j de un alfabeto B (*óp. cit.*, p. 112). Entre ambos símbolos existen probabilidades condicionadas $P(b_j|a_i)$ y $P(a_i|b_j)$, así como a cada conjunto de símbolos se le asocia una entropía $H(A)$ y $H(B)$ respectivamente, calculadas con la expresión [8]. Otra medida relevante es la probabilidad a posteriori de cualquier símbolo de A , dado un símbolo de salida b_j ; también con su entropía asociada, mostrada en la expresión [9] (*óp. cit.*, p. 118). La ausencia de índice en a se refiere a la suma sobre todo el alfabeto A (*ibíd.*).

$$H(A|b_j) = \sum_A P(a|b_j) \log\left(\frac{1}{P(a|b_j)}\right) \quad [9]$$

Si se expande esta expresión para abarcar a todos los símbolos de B , resulta el valor medio de las entropías a posteriori $H(A|B)$, mostrada en la expresión [10]; y que recibe el nombre de *equivocación del canal* (óp. cit., p. 123).

$$H(A|B) = \sum_B P(b)H(A|b) = \sum_{A,B} P(a,b) \log\left(\frac{1}{P(a|b)}\right) \quad [10]$$

Con todo lo anterior, se puede definir a la *información mutua* de A y B , como la media de la cantidad de información que da la observación de un símbolo de salida b , mostrada en la expresión [11] (óp. cit., p. 124). Entre las propiedades de la información mutua, se destaca que si los eventos, o símbolos, de A y B son independientes, la información mutua entre ambos alfabetos será de cero; y con alguna dependencia entre las variables será mayor a cero. Además, es una operación conmutativa (óp. cit., p. 126).

$$I(A; B) = H(A) - H(A|B) \quad [11]$$

La unión de $H(A)$ y $H(B)$, llamada *entropía afín* e interpretada como “la incertidumbre del suceso simultáneo (a_i, b_j) ” (óp. cit., p. 126-127) se expone en la expresión [12] (óp. cit., p. 127). La relación entre las cantidades anteriores se expone de forma visual mediante un diagrama de Venn, en la Figura 13 (loc. cit.).

$$H(A, B) = H(A) + H(B) - I(A; B) \quad [12]$$

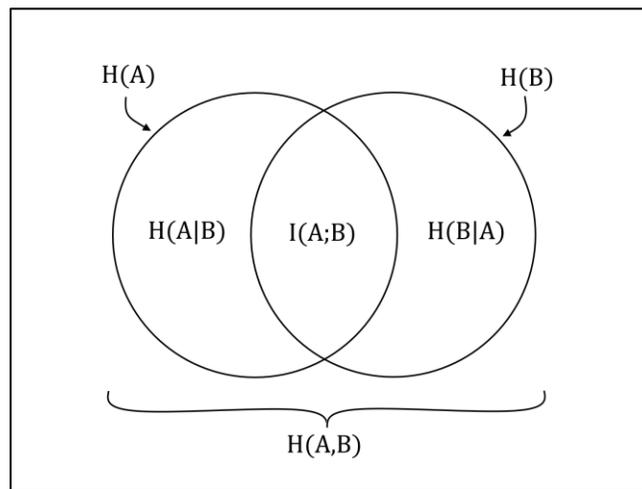


Figura 13: Relación entre las entropías e información mutua de un canal. Adaptado de (Abramson, 1981, pág. 127).

El uso práctico de estas magnitudes se da al tomar la relación entre un atributo y la *clase* (binaria, para los fines de este documento) como un canal; teniendo a la entrada un valor de la *clase* y a la salida el valor de uno de los atributos. Sin embargo, los atributos que se trabajan son variables continuas. De acuerdo con (Jones, 1979, pág. 136) se puede extender la noción de entropía cambiando la probabilidad por la densidad de probabilidad y las sumas por integrales, que deben de evaluarse de $-\infty$ a $+\infty$. Este último detalle implica que las integrales no siempre convergerán, así que las nociones sólo pueden aplicarse con cierta reserva.

Para fines prácticos, es más sencillo discretizar en intervalos al atributo continuo y manejarlo como si fuera una variable nominal, con lo que se pueden aplicar las métricas de los párrafos anteriores. A dichas métricas se les manejó de tres formas para determinar las variables más relevantes, como se describe a continuación.

La *información mutua* (expresión [11]) puede ser empleada como tal, definiendo a A como la *clase* y a B como una versión discretizada del atributo numérico; sin olvidar que esta cantidad es conmutativa. Retomando a los párrafos anteriores, si A y B se componen de eventos independientes, la información mutua tenderá a ser cero; por lo que nos interesa conservar a los atributos con la mayor información mutua respecto a la *clase*. Sin embargo, esta medida puede resultar en valores grandes para atributos con muchas categorías (intervalos diferentes, para las variables continuas que se han discretizado) (Witten, Frank, Hall, & Pal, 2017, págs. 110-111).

Otra medida, la *proporción de la ganancia de información* (*Information Gain Ratio*, en inglés), mostrada en la expresión [13] (óp. cit., p. 112); es la proporción de la información mutua entre la *clase* A y un atributo discretizado B , con respecto a la entropía del atributo (*ibíd.*). El incremento que pudiera llegar a tener la información mutua se normaliza o acota con respecto a la entropía del atributo, que será menor o igual a ésta (véase la Figura 13). Esta medida se ocupa para mitigar los inconvenientes de la información mutua. Nuevamente, nos interesan atributos que maximicen este valor. Como se podría sospechar, el defecto de esta medida es el favorecer a combinaciones de variables con alta información mutua, pero con una baja entropía del atributo (óp. cit., p. 112). A esta expresión también se le conoce como el *coeficiente de incertidumbre* de B (Press, Teukolsky, Vetterling, & Flannery, 1992, pág. 634). El coeficiente de incertidumbre de A se obtiene intercambiando las variables en [13] (*ibíd.*).

$$GR(A, B) = \frac{H(A) - H(A|B)}{H(B)} \quad [13]$$

$$GR(A, B) = \frac{I(A; B)}{H(B)}$$

La última medida basada en la teoría de la información que usará es la *incertidumbre simétrica*, incluida en la expresión [14] (Witten, Frank, Hall, & Pal, 2017, pág. 291). Esta es una media ponderada de ambos coeficientes de incertidumbre, donde se pondera cada coeficiente por la entropía de cada variable involucrada (Press, Teukolsky, Vetterling, & Flannery, 1992, pág. 634). Como en las cantidades anteriores, se busca elegir a los atributos que den los valores más altos.

$$U(A, B) = 2 \frac{H(A) + H(B) - H(A, B)}{H(A) + H(B)} \quad [14]$$

$$U(A, B) = 2 \frac{I(A; B)}{H(A) + H(B)}$$

10.4. Atributos extraídos de forma original

Para cada una de las ventanas extraídas a partir de los sismogramas preprocesados, se calcularon los siguientes atributos, presentados en la Tabla 9. Los atributos abarcan valores calculados a partir de las componentes tanto en coordenadas cartesianas como en esféricas. Más detalles sobre el procesamiento y justificación de estas variables se encuentran en la sección de Metodología.

Dentro de la Tabla 9, los atributos 1 a 3 son variables de rastreo, para identificar cada ventana dentro del sismograma, ya sea con un ordinal o por su punto medio en segundos o muestras desde el comienzo del sismograma. El cuarto atributo corresponde a la clasificación manual de cada ventana en *evento* (la variable de *clase* vale 1) o *ruido* (la *clase* vale 0). Los atributos 5 a 204 se repiten para cada una de las cuatro bandas disponibles al filtrar la señal: B1 corresponde al sismograma sin filtrar, B2 al sismograma filtrado con un pasa bajas de 1 Hz., B3 al de un pasa altas de 5Hz., y la B4 al de un pasa bandas entre 1 y 5 Hz.

Los primeros nueve atributos dentro de cada banda son valores usados en análisis de polarización, calculados según las expresiones de (Maercklin, 2010); y los nueve que les siguen sus primeras diferencias (con respecto al tiempo) centradas. Tras los atributos de polarización y sus primeras diferencias centradas, se encuentran atributos calculados a partir de las componentes esféricas de las muestras de cada ventana (atributos 35 a 54 en la banda 1), así como sus primeras derivadas (respecto al tiempo) centradas.

Los atributos del 205 al 264 representan proporciones calculadas entre los valores de un mismo atributo, pero de diferentes bandas, junto con sus primeras diferencias (respecto al tiempo) centradas. Las primeras diferencias centradas y las proporciones sólo fueron calculados para ciertos atributos, elegidos de forma empírica.

Tabla 9: Atributos originales calculados para cada ventana en los sismogramas preprocesados.

Atributo	Nombre del atributo	Banda del atributo o bandas entre las que se calcula la proporción	Atributo de origen, derivado, o comparado con proporción
1	<i>nvent</i>	(no aplica)	Número de ventana en el sismograma
2	<i>inst_t</i>	(no aplica)	Centro de la ventana, en segundos del sismograma
3	<i>inst_m</i>	(no aplica)	Centro de la ventana, en muestras del sismograma
4	<i>evento</i>	(no aplica)	Clase de la ventana. 1: <i>evento</i> . 0: <i>ruido</i> .
5	<i>pol_sin_filt_e_21</i>	B1: Sin filtrar	Elipticidad principal
6	<i>pol_sin_filt_e_31</i>	B1: Sin filtrar	Elipticidad subprincipal
7	<i>pol_sin_filt_e_32</i>	B1: Sin filtrar	Elipticidad transversal
8	<i>pol_sin_filt_rect_mk</i>	B1: Sin filtrar	Rectilinearidad (Montalbetti y Kanasewich)
9	<i>pol_sin_filt_rect_j</i>	B1: Sin filtrar	Rectilinearidad (Jurkevics)
10	<i>pol_sin_filt_gpp</i>	B1: Sin filtrar	Parámetro global de polarización
11	<i>pol_sin_filt_l_1</i>	B1: Sin filtrar	Coefficiente de linealidad
12	<i>pol_sin_filt_f_1</i>	B1: Sin filtrar	Coefficiente de planaridad
13	<i>pol_sin_filt_pl</i>	B1: Sin filtrar	Grado de planaridad
14	<i>pol_sin_filt_e_21_pdc</i>	B1: Sin filtrar	PDC de elipticidad principal
15	<i>pol_sin_filt_e_31_pdc</i>	B1: Sin filtrar	PDC de elipticidad subprincipal
16	<i>pol_sin_filt_e_32_pdc</i>	B1: Sin filtrar	PDC de elipticidad transversal
17	<i>pol_sin_filt_rect_mk_pdc</i>	B1: Sin filtrar	PDC de rectilinearidad (Montalbetti y Kanasewich)
18	<i>pol_sin_filt_rect_j_pdc</i>	B1: Sin filtrar	PDC de rectilinearidad (Jurkevics)
19	<i>pol_sin_filt_gpp_pdc</i>	B1: Sin filtrar	PDC de parámetro global de polarización
20	<i>pol_sin_filt_l_1_pdc</i>	B1: Sin filtrar	PDC de coeficiente de linealidad
21	<i>pol_sin_filt_f_1_pdc</i>	B1: Sin filtrar	PDC de coeficiente de planaridad
22	<i>pol_sin_filt_pl_pdc</i>	B1: Sin filtrar	PDC de grado de planaridad
23	<i>v1_sin_filt_theta</i>	B1: Sin filtrar	Colatitud del primer vector característico de la matriz de covarianza de las muestras de la ventana
24	<i>v1_sin_filt_phi</i>	B1: Sin filtrar	Azimut del primer vector característico de la matriz de covarianza de las muestras de la ventana
25	<i>v1_sin_filt_theta_pdc</i>	B1: Sin filtrar	PDC de la colatitud del primer vector característico de la matriz de covarianza de las muestras de la ventana
26	<i>v1_sin_filt_phi_pdc</i>	B1: Sin filtrar	PDC del azimut del primer vector característico de la matriz de covarianza de las muestras de la ventana
27	<i>v2_sin_filt_theta</i>	B1: Sin filtrar	Colatitud del segundo vector característico de la matriz de covarianza de las muestras de la ventana
28	<i>v2_sin_filt_phi</i>	B1: Sin filtrar	Azimut del segundo vector característico de la matriz de covarianza de las muestras de la ventana
29	<i>v2_sin_filt_theta_pdc</i>	B1: Sin filtrar	PDC de la colatitud del segundo vector característico de la matriz de covarianza de las muestras de la ventana

Atributo	Nombre del atributo	Banda del atributo o bandas entre las que se calcula la proporción	Atributo de origen, derivado, o comparado con proporción
30	<i>v2_sin_filt_phi_pdc</i>	B1: Sin filtrar	PDC del azimut del segundo vector característico de la matriz de covarianza de las muestras de la ventana
31	<i>v3_sin_filt_theta</i>	B1: Sin filtrar	Colatitud del tercer vector característico de la matriz de covarianza de las muestras de la ventana
32	<i>v3_sin_filt_phi</i>	B1: Sin filtrar	Azimut del tercer vector característico de la matriz de covarianza de las muestras de la ventana
33	<i>v3_sin_filt_theta_pdc</i>	B1: Sin filtrar	PDC de la colatitud del tercer vector característico de la matriz de covarianza de las muestras de la ventana
34	<i>v3_sin_filt_phi_pdc</i>	B1: Sin filtrar	PDC del azimut del tercer vector característico de la matriz de covarianza de las muestras de la ventana
35	<i>r_sin_filt_sumc</i>	B1: Sin filtrar	Suma de los cuadrados de las muestras en la ventana
36	<i>r_sin_filt_prom</i>	B1: Sin filtrar	Promedio de las muestras en la ventana
37	<i>r_sin_filt_pend</i>	B1: Sin filtrar	Pendiente de una recta ajustada por mínimos cuadrados a las muestras en la ventana
38	<i>r_sin_filt_sd</i>	B1: Sin filtrar	Desviación estándar de las muestras en la ventana
39	<i>r_sin_filt_sumc_pdc</i>	B1: Sin filtrar	PDC de suma de los cuadrados de las muestras en la ventana
40	<i>r_sin_filt_prom_pdc</i>	B1: Sin filtrar	PDC de promedio de las muestras en la ventana
41	<i>r_sin_filt_pend_pdc</i>	B1: Sin filtrar	PDC de pendiente de una recta ajustada por mínimos cuadrados a las muestras en la ventana
42	<i>r_sin_filt_sd_pdc</i>	B1: Sin filtrar	PDC de desviación estándar de las muestras en la ventana
43	<i>theta_sin_filt_prom</i>	B1: Sin filtrar	Promedio de las muestras en la ventana
44	<i>theta_sin_filt_pend</i>	B1: Sin filtrar	Pendiente de una recta ajustada por mínimos cuadrados a las muestras en la ventana
45	<i>theta_sin_filt_sd</i>	B1: Sin filtrar	Desviación estándar de las muestras en la ventana
46	<i>theta_sin_filt_prom_pdc</i>	B1: Sin filtrar	PDC de promedio de las muestras en la ventana
47	<i>theta_sin_filt_pend_pdc</i>	B1: Sin filtrar	PDC de pendiente de una recta ajustada por mínimos cuadrados a las muestras en la ventana
48	<i>theta_sin_filt_sd_pdc</i>	B1: Sin filtrar	PDC de desviación estándar de las muestras en la ventana
49	<i>phi_sin_filt_prom</i>	B1: Sin filtrar	Promedio de las muestras en la ventana
50	<i>phi_sin_filt_pend</i>	B1: Sin filtrar	Pendiente de una recta ajustada por mínimos cuadrados a las muestras en la ventana
51	<i>phi_sin_filt_sd</i>	B1: Sin filtrar	Desviación estándar de las muestras en la ventana
52	<i>phi_sin_filt_prom_pdc</i>	B1: Sin filtrar	PDC de promedio de las muestras en la ventana
53	<i>phi_sin_filt_pend_pdc</i>	B1: Sin filtrar	PDC de pendiente de una recta ajustada por mínimos cuadrados a las muestras en la ventana
54	<i>phi_sin_filt_sd_pdc</i>	B1: Sin filtrar	PDC de desviación estándar de las muestras en la ventana
55	<i>pol_pbaj_01p00_e_21</i>	B2: Pasa-bajas 1 Hz.	Elipticidad principal
56	<i>pol_pbaj_01p00_e_31</i>	B2: Pasa-bajas 1 Hz.	Elipticidad subprincipal
57	<i>pol_pbaj_01p00_e_32</i>	B2: Pasa-bajas 1 Hz.	Elipticidad transversal
58	<i>pol_pbaj_01p00_rect_mk</i>	B2: Pasa-bajas 1 Hz.	Rectilinearidad (Montalbetti y Kanasevich)
59	<i>pol_pbaj_01p00_rect_j</i>	B2: Pasa-bajas 1 Hz.	Rectilinearidad (Jurkevics)
60	<i>pol_pbaj_01p00_gpp</i>	B2: Pasa-bajas 1 Hz.	Parámetro global de polarización
61	<i>pol_pbaj_01p00_l_1</i>	B2: Pasa-bajas 1 Hz.	Coefficiente de linearidad

Atributo	Nombre del atributo	Banda del atributo o bandas entre las que se calcula la proporción	Atributo de origen, derivado, o comparado con proporción
62	<i>pol_pbaj_01p00_f_1</i>	B2: Pasa-bajas 1 Hz.	Coefficiente de planaridad
63	<i>pol_pbaj_01p00_pl</i>	B2: Pasa-bajas 1 Hz.	Grado de planaridad
64	<i>pol_pbaj_01p00_e_21_pdc</i>	B2: Pasa-bajas 1 Hz.	PDC de elipticidad principal
65	<i>pol_pbaj_01p00_e_31_pdc</i>	B2: Pasa-bajas 1 Hz.	PDC de elipticidad subprincipal
66	<i>pol_pbaj_01p00_e_32_pdc</i>	B2: Pasa-bajas 1 Hz.	PDC de elipticidad transversal
67	<i>pol_pbaj_01p00_rect_mk_pdc</i>	B2: Pasa-bajas 1 Hz.	PDC de rectilinearidad (Montalbetti y Kanasewich)
68	<i>pol_pbaj_01p00_rect_j_pdc</i>	B2: Pasa-bajas 1 Hz.	PDC de rectilinearidad (Jurkevics)
69	<i>pol_pbaj_01p00_gpp_pdc</i>	B2: Pasa-bajas 1 Hz.	PDC de parámetro global de polarización
70	<i>pol_pbaj_01p00_l_1_pdc</i>	B2: Pasa-bajas 1 Hz.	PDC de coeficiente de linearidad
71	<i>pol_pbaj_01p00_f_1_pdc</i>	B2: Pasa-bajas 1 Hz.	PDC de coeficiente de planaridad
72	<i>pol_pbaj_01p00_pl_pdc</i>	B2: Pasa-bajas 1 Hz.	PDC de grado de planaridad
73	<i>v1_pbaj_01p00_theta</i>	B2: Pasa-bajas 1 Hz.	Colatitud del primer vector característico de la matriz de covarianza de las muestras de la ventana
74	<i>v1_pbaj_01p00_phi</i>	B2: Pasa-bajas 1 Hz.	Azimut del primer vector característico de la matriz de covarianza de las muestras de la ventana
75	<i>v1_pbaj_01p00_theta_pdc</i>	B2: Pasa-bajas 1 Hz.	PDC de la colatitud del primer vector característico de la matriz de covarianza de las muestras de la ventana
76	<i>v1_pbaj_01p00_phi_pdc</i>	B2: Pasa-bajas 1 Hz.	PDC del azimut del primer vector característico de la matriz de covarianza de las muestras de la ventana
77	<i>v2_pbaj_01p00_theta</i>	B2: Pasa-bajas 1 Hz.	Colatitud del segundo vector característico de la matriz de covarianza de las muestras de la ventana
78	<i>v2_pbaj_01p00_phi</i>	B2: Pasa-bajas 1 Hz.	Azimut del segundo vector característico de la matriz de covarianza de las muestras de la ventana
79	<i>v2_pbaj_01p00_theta_pdc</i>	B2: Pasa-bajas 1 Hz.	PDC de la colatitud del segundo vector característico de la matriz de covarianza de las muestras de la ventana
80	<i>v2_pbaj_01p00_phi_pdc</i>	B2: Pasa-bajas 1 Hz.	PDC del azimut del segundo vector característico de la matriz de covarianza de las muestras de la ventana
81	<i>v3_pbaj_01p00_theta</i>	B2: Pasa-bajas 1 Hz.	Colatitud del tercer vector característico de la matriz de covarianza de las muestras de la ventana
82	<i>v3_pbaj_01p00_phi</i>	B2: Pasa-bajas 1 Hz.	Azimut del tercer vector característico de la matriz de covarianza de las muestras de la ventana
83	<i>v3_pbaj_01p00_theta_pdc</i>	B2: Pasa-bajas 1 Hz.	PDC de la colatitud del tercer vector característico de la matriz de covarianza de las muestras de la ventana
84	<i>v3_pbaj_01p00_phi_pdc</i>	B2: Pasa-bajas 1 Hz.	PDC del azimut del tercer vector característico de la matriz de covarianza de las muestras de la ventana
85	<i>r_pbaj_01p00_sumc</i>	B2: Pasa-bajas 1 Hz.	Suma de los cuadrados de las muestras en la ventana
86	<i>r_pbaj_01p00_prom</i>	B2: Pasa-bajas 1 Hz.	Promedio de las muestras en la ventana
87	<i>r_pbaj_01p00_pend</i>	B2: Pasa-bajas 1 Hz.	Pendiente de una recta ajustada por mínimos cuadrados a las muestras en la ventana
88	<i>r_pbaj_01p00_sd</i>	B2: Pasa-bajas 1 Hz.	Desviación estándar de las muestras en la ventana
89	<i>r_pbaj_01p00_sumc_pdc</i>	B2: Pasa-bajas 1 Hz.	PDC de suma de los cuadrados de las muestras en la ventana

Atributo	Nombre del atributo	Banda del atributo o bandas entre las que se calcula la proporción	Atributo de origen, derivado, o comparado con proporción
90	<i>r_pbaj_01p00_prom_pdc</i>	B2: Pasa-bajas 1 Hz.	PDC de promedio de las muestras en la ventana
91	<i>r_pbaj_01p00_pend_pdc</i>	B2: Pasa-bajas 1 Hz.	PDC de pendiente de una recta ajustada por mínimos cuadrados a las muestras en la ventana
92	<i>r_pbaj_01p00_sd_pdc</i>	B2: Pasa-bajas 1 Hz.	PDC de desviación estándar de las muestras en la ventana
93	<i>theta_pbaj_01p00_prom</i>	B2: Pasa-bajas 1 Hz.	Promedio de las muestras en la ventana
94	<i>theta_pbaj_01p00_pend</i>	B2: Pasa-bajas 1 Hz.	Pendiente de una recta ajustada por mínimos cuadrados a las muestras en la ventana
95	<i>theta_pbaj_01p00_sd</i>	B2: Pasa-bajas 1 Hz.	Desviación estándar de las muestras en la ventana
96	<i>theta_pbaj_01p00_prom_pdc</i>	B2: Pasa-bajas 1 Hz.	PDC de promedio de las muestras en la ventana
97	<i>theta_pbaj_01p00_pend_pdc</i>	B2: Pasa-bajas 1 Hz.	PDC de pendiente de una recta ajustada por mínimos cuadrados a las muestras en la ventana
98	<i>theta_pbaj_01p00_sd_pdc</i>	B2: Pasa-bajas 1 Hz.	PDC de desviación estándar de las muestras en la ventana
99	<i>phi_pbaj_01p00_prom</i>	B2: Pasa-bajas 1 Hz.	Promedio de las muestras en la ventana
100	<i>phi_pbaj_01p00_pend</i>	B2: Pasa-bajas 1 Hz.	Pendiente de una recta ajustada por mínimos cuadrados a las muestras en la ventana
101	<i>phi_pbaj_01p00_sd</i>	B2: Pasa-bajas 1 Hz.	Desviación estándar de las muestras en la ventana
102	<i>phi_pbaj_01p00_prom_pdc</i>	B2: Pasa-bajas 1 Hz.	PDC de promedio de las muestras en la ventana
103	<i>phi_pbaj_01p00_pend_pdc</i>	B2: Pasa-bajas 1 Hz.	PDC de pendiente de una recta ajustada por mínimos cuadrados a las muestras en la ventana
104	<i>phi_pbaj_01p00_sd_pdc</i>	B2: Pasa-bajas 1 Hz.	PDC de desviación estándar de las muestras en la ventana
105	<i>pol_palt_05p00_e_21</i>	B3: Pasa-altas 5 Hz.	Elipticidad principal
106	<i>pol_palt_05p00_e_31</i>	B3: Pasa-altas 5 Hz.	Elipticidad subprincipal
107	<i>pol_palt_05p00_e_32</i>	B3: Pasa-altas 5 Hz.	Elipticidad transversal
108	<i>pol_palt_05p00_rect_mk</i>	B3: Pasa-altas 5 Hz.	Rectilinearidad (Montalbetti y Kanasewich)
109	<i>pol_palt_05p00_rect_j</i>	B3: Pasa-altas 5 Hz.	Rectilinearidad (Jurkevics)
110	<i>pol_palt_05p00_gpp</i>	B3: Pasa-altas 5 Hz.	Parámetro global de polarización
111	<i>pol_palt_05p00_l_1</i>	B3: Pasa-altas 5 Hz.	Coefficiente de linearidad
112	<i>pol_palt_05p00_f_1</i>	B3: Pasa-altas 5 Hz.	Coefficiente de planaridad
113	<i>pol_palt_05p00_pl</i>	B3: Pasa-altas 5 Hz.	Grado de planaridad
114	<i>pol_palt_05p00_e_21_pdc</i>	B3: Pasa-altas 5 Hz.	PDC de elipticidad principal
115	<i>pol_palt_05p00_e_31_pdc</i>	B3: Pasa-altas 5 Hz.	PDC de elipticidad subprincipal
116	<i>pol_palt_05p00_e_32_pdc</i>	B3: Pasa-altas 5 Hz.	PDC de elipticidad transversal
117	<i>pol_palt_05p00_rect_mk_pdc</i>	B3: Pasa-altas 5 Hz.	PDC de rectilinearidad (Montalbetti y Kanasewich)

Atributo	Nombre del atributo	Banda del atributo o bandas entre las que se calcula la proporción	Atributo de origen, derivado, o comparado con proporción
118	<i>pol_palt_05p00_rect_j_pdc</i>	B3: Pasa-altas 5 Hz.	PDC de rectilinearidad (Jurkevics)
119	<i>pol_palt_05p00_gpp_pdc</i>	B3: Pasa-altas 5 Hz.	PDC de parámetro global de polarización
120	<i>pol_palt_05p00_l_1_pdc</i>	B3: Pasa-altas 5 Hz.	PDC de coeficiente de linealidad
121	<i>pol_palt_05p00_f_1_pdc</i>	B3: Pasa-altas 5 Hz.	PDC de coeficiente de planaridad
122	<i>pol_palt_05p00_pl_pdc</i>	B3: Pasa-altas 5 Hz.	PDC de grado de planaridad
123	<i>v1_palt_05p00_theta</i>	B3: Pasa-altas 5 Hz.	Colatitud del primer vector característico de la matriz de covarianza de las muestras de la ventana
124	<i>v1_palt_05p00_phi</i>	B3: Pasa-altas 5 Hz.	Azimut del primer vector característico de la matriz de covarianza de las muestras de la ventana
125	<i>v1_palt_05p00_theta_pdc</i>	B3: Pasa-altas 5 Hz.	PDC de la colatitud del primer vector característico de la matriz de covarianza de las muestras de la ventana
126	<i>v1_palt_05p00_phi_pdc</i>	B3: Pasa-altas 5 Hz.	PDC del azimut del primer vector característico de la matriz de covarianza de las muestras de la ventana
127	<i>v2_palt_05p00_theta</i>	B3: Pasa-altas 5 Hz.	Colatitud del segundo vector característico de la matriz de covarianza de las muestras de la ventana
128	<i>v2_palt_05p00_phi</i>	B3: Pasa-altas 5 Hz.	Azimut del segundo vector característico de la matriz de covarianza de las muestras de la ventana
129	<i>v2_palt_05p00_theta_pdc</i>	B3: Pasa-altas 5 Hz.	PDC de la colatitud del segundo vector característico de la matriz de covarianza de las muestras de la ventana
130	<i>v2_palt_05p00_phi_pdc</i>	B3: Pasa-altas 5 Hz.	PDC del azimut del segundo vector característico de la matriz de covarianza de las muestras de la ventana
131	<i>v3_palt_05p00_theta</i>	B3: Pasa-altas 5 Hz.	Colatitud del tercer vector característico de la matriz de covarianza de las muestras de la ventana
132	<i>v3_palt_05p00_phi</i>	B3: Pasa-altas 5 Hz.	Azimut del tercer vector característico de la matriz de covarianza de las muestras de la ventana
133	<i>v3_palt_05p00_theta_pdc</i>	B3: Pasa-altas 5 Hz.	PDC de la colatitud del tercer vector característico de la matriz de covarianza de las muestras de la ventana
134	<i>v3_palt_05p00_phi_pdc</i>	B3: Pasa-altas 5 Hz.	PDC del azimut del tercer vector característico de la matriz de covarianza de las muestras de la ventana
135	<i>r_palt_05p00_sumc</i>	B3: Pasa-altas 5 Hz.	Suma de los cuadrados de las muestras en la ventana
136	<i>r_palt_05p00_prom</i>	B3: Pasa-altas 5 Hz.	Promedio de las muestras en la ventana
137	<i>r_palt_05p00_pend</i>	B3: Pasa-altas 5 Hz.	Pendiente de una recta ajustada por mínimos cuadrados a las muestras en la ventana
138	<i>r_palt_05p00_sd</i>	B3: Pasa-altas 5 Hz.	Desviación estándar de las muestras en la ventana
139	<i>r_palt_05p00_sumc_pdc</i>	B3: Pasa-altas 5 Hz.	PDC de suma de los cuadrados de las muestras en la ventana
140	<i>r_palt_05p00_prom_pdc</i>	B3: Pasa-altas 5 Hz.	PDC de promedio de las muestras en la ventana
141	<i>r_palt_05p00_pend_pdc</i>	B3: Pasa-altas 5 Hz.	PDC de pendiente de una recta ajustada por mínimos cuadrados a las muestras en la ventana
142	<i>r_palt_05p00_sd_pdc</i>	B3: Pasa-altas 5 Hz.	PDC de desviación estándar de las muestras en la ventana
143	<i>theta_palt_05p00_prom</i>	B3: Pasa-altas 5 Hz.	Promedio de las muestras en la ventana
144	<i>theta_palt_05p00_pend</i>	B3: Pasa-altas 5 Hz.	Pendiente de una recta ajustada por mínimos cuadrados a las muestras en la ventana
145	<i>theta_palt_05p00_sd</i>	B3: Pasa-altas 5 Hz.	Desviación estándar de las muestras en la ventana

Atributo	Nombre del atributo	Banda del atributo o bandas entre las que se calcula la proporción	Atributo de origen, derivado, o comparado con proporción
146	<i>theta_palt_05p00_prom_pdc</i>	B3: Pasa-altas 5 Hz.	PDC de promedio de las muestras en la ventana
147	<i>theta_palt_05p00_pend_pdc</i>	B3: Pasa-altas 5 Hz.	PDC de pendiente de una recta ajustada por mínimos cuadrados a las muestras en la ventana
148	<i>theta_palt_05p00_sd_pdc</i>	B3: Pasa-altas 5 Hz.	PDC de desviación estándar de las muestras en la ventana
149	<i>phi_palt_05p00_prom</i>	B3: Pasa-altas 5 Hz.	Promedio de las muestras en la ventana
150	<i>phi_palt_05p00_pend</i>	B3: Pasa-altas 5 Hz.	Pendiente de una recta ajustada por mínimos cuadrados a las muestras en la ventana
151	<i>phi_palt_05p00_sd</i>	B3: Pasa-altas 5 Hz.	Desviación estándar de las muestras en la ventana
152	<i>phi_palt_05p00_prom_pdc</i>	B3: Pasa-altas 5 Hz.	PDC de promedio de las muestras en la ventana
153	<i>phi_palt_05p00_pend_pdc</i>	B3: Pasa-altas 5 Hz.	PDC de pendiente de una recta ajustada por mínimos cuadrados a las muestras en la ventana
154	<i>phi_palt_05p00_sd_pdc</i>	B3: Pasa-altas 5 Hz.	PDC de desviación estándar de las muestras en la ventana
155	<i>pol_pban_01p00a05p00_e_21</i>	B4: Pasa-banda 1 a 5 Hz.	Elipticidad principal
156	<i>pol_pban_01p00a05p00_e_31</i>	B4: Pasa-banda 1 a 5 Hz.	Elipticidad subprincipal
157	<i>pol_pban_01p00a05p00_e_32</i>	B4: Pasa-banda 1 a 5 Hz.	Elipticidad transversal
158	<i>pol_pban_01p00a05p00_rect_mk</i>	B4: Pasa-banda 1 a 5 Hz.	Rectilinearidad (Montalbetti y Kanasevich)
159	<i>pol_pban_01p00a05p00_rect_j</i>	B4: Pasa-banda 1 a 5 Hz.	Rectilinearidad (Jurkevics)
160	<i>pol_pban_01p00a05p00_gpp</i>	B4: Pasa-banda 1 a 5 Hz.	Parámetro global de polarización
161	<i>pol_pban_01p00a05p00_l_1</i>	B4: Pasa-banda 1 a 5 Hz.	Coefficiente de linearidad
162	<i>pol_pban_01p00a05p00_f_1</i>	B4: Pasa-banda 1 a 5 Hz.	Coefficiente de planaridad
163	<i>pol_pban_01p00a05p00_pl</i>	B4: Pasa-banda 1 a 5 Hz.	Grado de planaridad
164	<i>pol_pban_01p00a05p00_e_21_pdc</i>	B4: Pasa-banda 1 a 5 Hz.	PDC de elipticidad principal
165	<i>pol_pban_01p00a05p00_e_31_pdc</i>	B4: Pasa-banda 1 a 5 Hz.	PDC de elipticidad subprincipal
166	<i>pol_pban_01p00a05p00_e_32_pdc</i>	B4: Pasa-banda 1 a 5 Hz.	PDC de elipticidad transversal
167	<i>pol_pban_01p00a05p00_rect_mk_pdc</i>	B4: Pasa-banda 1 a 5 Hz.	PDC de rectilinearidad (Montalbetti y Kanasevich)
168	<i>pol_pban_01p00a05p00_rect_j_pdc</i>	B4: Pasa-banda 1 a 5 Hz.	PDC de rectilinearidad (Jurkevics)
169	<i>pol_pban_01p00a05p00_gpp_pdc</i>	B4: Pasa-banda 1 a 5 Hz.	PDC de parámetro global de polarización
170	<i>pol_pban_01p00a05p00_l_1_pdc</i>	B4: Pasa-banda 1 a 5 Hz.	PDC de coeficiente de linearidad
171	<i>pol_pban_01p00a05p00_f_1_pdc</i>	B4: Pasa-banda 1 a 5 Hz.	PDC de coeficiente de planaridad
172	<i>pol_pban_01p00a05p00_pl_pdc</i>	B4: Pasa-banda 1 a 5 Hz.	PDC de grado de planaridad
173	<i>v1_pban_01p00a05p00_theta</i>	B4: Pasa-banda 1 a 5 Hz.	Colatitud del primer vector característico de la matriz de covarianza de las muestras de la ventana

Atributo	Nombre del atributo	Banda del atributo o bandas entre las que se calcula la proporción	Atributo de origen, derivado, o comparado con proporción
174	<i>v1_pban_01p00a05p00_phi</i>	B4: Pasa-banda 1 a 5 Hz.	Azimut del primer vector característico de la matriz de covarianza de las muestras de la ventana
175	<i>v1_pban_01p00a05p00_theta_pdc</i>	B4: Pasa-banda 1 a 5 Hz.	PDC de la colatitud del primer vector característico de la matriz de covarianza de las muestras de la ventana
176	<i>v1_pban_01p00a05p00_phi_pdc</i>	B4: Pasa-banda 1 a 5 Hz.	PDC del azimut del primer vector característico de la matriz de covarianza de las muestras de la ventana
177	<i>v2_pban_01p00a05p00_theta</i>	B4: Pasa-banda 1 a 5 Hz.	Colatitud del segundo vector característico de la matriz de covarianza de las muestras de la ventana
178	<i>v2_pban_01p00a05p00_phi</i>	B4: Pasa-banda 1 a 5 Hz.	Azimut del segundo vector característico de la matriz de covarianza de las muestras de la ventana
179	<i>v2_pban_01p00a05p00_theta_pdc</i>	B4: Pasa-banda 1 a 5 Hz.	PDC de la colatitud del segundo vector característico de la matriz de covarianza de las muestras de la ventana
180	<i>v2_pban_01p00a05p00_phi_pdc</i>	B4: Pasa-banda 1 a 5 Hz.	PDC del azimut del segundo vector característico de la matriz de covarianza de las muestras de la ventana
181	<i>v3_pban_01p00a05p00_theta</i>	B4: Pasa-banda 1 a 5 Hz.	Colatitud del tercer vector característico de la matriz de covarianza de las muestras de la ventana
182	<i>v3_pban_01p00a05p00_phi</i>	B4: Pasa-banda 1 a 5 Hz.	Azimut del tercer vector característico de la matriz de covarianza de las muestras de la ventana
183	<i>v3_pban_01p00a05p00_theta_pdc</i>	B4: Pasa-banda 1 a 5 Hz.	PDC de la colatitud del tercer vector característico de la matriz de covarianza de las muestras de la ventana
184	<i>v3_pban_01p00a05p00_phi_pdc</i>	B4: Pasa-banda 1 a 5 Hz.	PDC del azimut del tercer vector característico de la matriz de covarianza de las muestras de la ventana
185	<i>r_pban_01p00a05p00_sumc</i>	B4: Pasa-banda 1 a 5 Hz.	Suma de los cuadrados de las muestras en la ventana
186	<i>r_pban_01p00a05p00_prom</i>	B4: Pasa-banda 1 a 5 Hz.	Promedio de las muestras en la ventana
187	<i>r_pban_01p00a05p00_pend</i>	B4: Pasa-banda 1 a 5 Hz.	Pendiente de una recta ajustada por mínimos cuadrados a las muestras en la ventana
188	<i>r_pban_01p00a05p00_sd</i>	B4: Pasa-banda 1 a 5 Hz.	Desviación estándar de las muestras en la ventana
189	<i>r_pban_01p00a05p00_sumc_pdc</i>	B4: Pasa-banda 1 a 5 Hz.	PDC de suma de los cuadrados de las muestras en la ventana
190	<i>r_pban_01p00a05p00_prom_pdc</i>	B4: Pasa-banda 1 a 5 Hz.	PDC de promedio de las muestras en la ventana
191	<i>r_pban_01p00a05p00_pend_pdc</i>	B4: Pasa-banda 1 a 5 Hz.	PDC de pendiente de una recta ajustada por mínimos cuadrados a las muestras en la ventana
192	<i>r_pban_01p00a05p00_sd_pdc</i>	B4: Pasa-banda 1 a 5 Hz.	PDC de desviación estándar de las muestras en la ventana
193	<i>theta_pban_01p00a05p00_prom</i>	B4: Pasa-banda 1 a 5 Hz.	Promedio de las muestras en la ventana
194	<i>theta_pban_01p00a05p00_pend</i>	B4: Pasa-banda 1 a 5 Hz.	Pendiente de una recta ajustada por mínimos cuadrados a las muestras en la ventana
195	<i>theta_pban_01p00a05p00_sd</i>	B4: Pasa-banda 1 a 5 Hz.	Desviación estándar de las muestras en la ventana
196	<i>theta_pban_01p00a05p00_prom_pdc</i>	B4: Pasa-banda 1 a 5 Hz.	PDC de promedio de las muestras en la ventana
197	<i>theta_pban_01p00a05p00_pend_pdc</i>	B4: Pasa-banda 1 a 5 Hz.	PDC de pendiente de una recta ajustada por mínimos cuadrados a las muestras en la ventana
198	<i>theta_pban_01p00a05p00_sd_pdc</i>	B4: Pasa-banda 1 a 5 Hz.	PDC de desviación estándar de las muestras en la ventana
199	<i>phi_pban_01p00a05p00_prom</i>	B4: Pasa-banda 1 a 5 Hz.	Promedio de las muestras en la ventana
200	<i>phi_pban_01p00a05p00_pend</i>	B4: Pasa-banda 1 a 5 Hz.	Pendiente de una recta ajustada por mínimos cuadrados a las muestras en la ventana
201	<i>phi_pban_01p00a05p00_sd</i>	B4: Pasa-banda 1 a 5 Hz.	Desviación estándar de las muestras en la ventana

Atributo	Nombre del atributo	Banda del atributo o bandas entre las que se calcula la proporción	Atributo de origen, derivado, o comparado con proporción
202	<i>phi_pban_01p00a05p00_prom_pdc</i>	B4: Pasa-banda 1 a 5 Hz.	PDC de promedio de las muestras en la ventana
203	<i>phi_pban_01p00a05p00_pend_pdc</i>	B4: Pasa-banda 1 a 5 Hz.	PDC de pendiente de una recta ajustada por mínimos cuadrados a las muestras en la ventana
204	<i>phi_pban_01p00a05p00_sd_pdc</i>	B4: Pasa-banda 1 a 5 Hz.	PDC de desviación estándar de las muestras en la ventana
205	<i>pol_e_21_B2sB1</i>	B2 / B1	Elipticidad principal
211	<i>pol_e_21_B2sB1_pdc</i>	B2 / B1	PDC de elipticidad principal
206	<i>pol_e_21_B3sB1</i>	B3 / B1	Elipticidad principal
212	<i>pol_e_21_B3sB1_pdc</i>	B3 / B1	PDC de elipticidad principal
208	<i>pol_e_21_B3sB2</i>	B3 / B2	Elipticidad principal
214	<i>pol_e_21_B3sB2_pdc</i>	B3 / B2	PDC de elipticidad principal
207	<i>pol_e_21_B4sB1</i>	B4 / B1	Elipticidad principal
213	<i>pol_e_21_B4sB1_pdc</i>	B4 / B1	PDC de elipticidad principal
209	<i>pol_e_21_B4sB2</i>	B4 / B2	Elipticidad principal
215	<i>pol_e_21_B4sB2_pdc</i>	B4 / B2	PDC de elipticidad principal
210	<i>pol_e_21_B4sB3</i>	B4 / B3	Elipticidad principal
216	<i>pol_e_21_B4sB3_pdc</i>	B4 / B3	PDC de elipticidad principal
217	<i>pol_e_31_B2sB1</i>	B2 / B1	Elipticidad subprincipal
223	<i>pol_e_31_B2sB1_pdc</i>	B2 / B1	PDC de elipticidad subprincipal
218	<i>pol_e_31_B3sB1</i>	B3 / B1	Elipticidad subprincipal
224	<i>pol_e_31_B3sB1_pdc</i>	B3 / B1	PDC de elipticidad subprincipal
220	<i>pol_e_31_B3sB2</i>	B3 / B2	Elipticidad subprincipal
226	<i>pol_e_31_B3sB2_pdc</i>	B3 / B2	PDC de elipticidad subprincipal
219	<i>pol_e_31_B4sB1</i>	B4 / B1	Elipticidad subprincipal
225	<i>pol_e_31_B4sB1_pdc</i>	B4 / B1	PDC de elipticidad subprincipal
221	<i>pol_e_31_B4sB2</i>	B4 / B2	Elipticidad subprincipal
227	<i>pol_e_31_B4sB2_pdc</i>	B4 / B2	PDC de elipticidad subprincipal
222	<i>pol_e_31_B4sB3</i>	B4 / B3	Elipticidad subprincipal
228	<i>pol_e_31_B4sB3_pdc</i>	B4 / B3	PDC de elipticidad subprincipal
229	<i>pol_e_32_B2sB1</i>	B2 / B1	Elipticidad transversal
235	<i>pol_e_32_B2sB1_pdc</i>	B2 / B1	PDC de elipticidad transversal
230	<i>pol_e_32_B3sB1</i>	B3 / B1	Elipticidad transversal
236	<i>pol_e_32_B3sB1_pdc</i>	B3 / B1	PDC de elipticidad transversal
232	<i>pol_e_32_B3sB2</i>	B3 / B2	Elipticidad transversal
238	<i>pol_e_32_B3sB2_pdc</i>	B3 / B2	PDC de elipticidad transversal
231	<i>pol_e_32_B4sB1</i>	B4 / B1	Elipticidad transversal
237	<i>pol_e_32_B4sB1_pdc</i>	B4 / B1	PDC de elipticidad transversal
233	<i>pol_e_32_B4sB2</i>	B4 / B2	Elipticidad transversal
239	<i>pol_e_32_B4sB2_pdc</i>	B4 / B2	PDC de elipticidad transversal

Atributo	Nombre del atributo	Banda del atributo o bandas entre las que se calcula la proporción	Atributo de origen, derivado, o comparado con proporción
234	<i>pol_e_32_B4sB3</i>	B4 / B3	Elipticidad transversal
240	<i>pol_e_32_B4sB3_pdc</i>	B4 / B3	PDC de elipticidad transversal
241	<i>pol_gpp_B2sB1</i>	B2 / B1	Parámetro global de polarización
247	<i>pol_gpp_B2sB1_pdc</i>	B2 / B1	PDC de parámetro global de polarización
242	<i>pol_gpp_B3sB1</i>	B3 / B1	Parámetro global de polarización
248	<i>pol_gpp_B3sB1_pdc</i>	B3 / B1	PDC de parámetro global de polarización
244	<i>pol_gpp_B3sB2</i>	B3 / B2	Parámetro global de polarización
250	<i>pol_gpp_B3sB2_pdc</i>	B3 / B2	PDC de parámetro global de polarización
243	<i>pol_gpp_B4sB1</i>	B4 / B1	Parámetro global de polarización
249	<i>pol_gpp_B4sB1_pdc</i>	B4 / B1	PDC de parámetro global de polarización
245	<i>pol_gpp_B4sB2</i>	B4 / B2	Parámetro global de polarización
251	<i>pol_gpp_B4sB2_pdc</i>	B4 / B2	PDC de parámetro global de polarización
246	<i>pol_gpp_B4sB3</i>	B4 / B3	Parámetro global de polarización
252	<i>pol_gpp_B4sB3_pdc</i>	B4 / B3	PDC de parámetro global de polarización
265	<i>r_prom_B2sB1</i>	B2 / B1	Promedio de las muestras en la ventana
271	<i>r_prom_B2sB1_pdc</i>	B2 / B1	PDC de promedio de las muestras en la ventana
266	<i>r_prom_B3sB1</i>	B3 / B1	Promedio de las muestras en la ventana
272	<i>r_prom_B3sB1_pdc</i>	B3 / B1	PDC de promedio de las muestras en la ventana
268	<i>r_prom_B3sB2</i>	B3 / B2	Promedio de las muestras en la ventana
274	<i>r_prom_B3sB2_pdc</i>	B3 / B2	PDC de promedio de las muestras en la ventana
267	<i>r_prom_B4sB1</i>	B4 / B1	Promedio de las muestras en la ventana
273	<i>r_prom_B4sB1_pdc</i>	B4 / B1	PDC de promedio de las muestras en la ventana
269	<i>r_prom_B4sB2</i>	B4 / B2	Promedio de las muestras en la ventana
275	<i>r_prom_B4sB2_pdc</i>	B4 / B2	PDC de promedio de las muestras en la ventana
270	<i>r_prom_B4sB3</i>	B4 / B3	Promedio de las muestras en la ventana
276	<i>r_prom_B4sB3_pdc</i>	B4 / B3	PDC de promedio de las muestras en la ventana
277	<i>r_sd_B2sB1</i>	B2 / B1	Desviación estándar de las muestras en la ventana
283	<i>r_sd_B2sB1_pdc</i>	B2 / B1	PDC de desviación estándar de las muestras en la ventana
278	<i>r_sd_B3sB1</i>	B3 / B1	Desviación estándar de las muestras en la ventana
284	<i>r_sd_B3sB1_pdc</i>	B3 / B1	PDC de desviación estándar de las muestras en la ventana
280	<i>r_sd_B3sB2</i>	B3 / B2	Desviación estándar de las muestras en la ventana
286	<i>r_sd_B3sB2_pdc</i>	B3 / B2	PDC de desviación estándar de las muestras en la ventana
279	<i>r_sd_B4sB1</i>	B4 / B1	Desviación estándar de las muestras en la ventana
285	<i>r_sd_B4sB1_pdc</i>	B4 / B1	PDC de desviación estándar de las muestras en la ventana
281	<i>r_sd_B4sB2</i>	B4 / B2	Desviación estándar de las muestras en la ventana
287	<i>r_sd_B4sB2_pdc</i>	B4 / B2	PDC de desviación estándar de las muestras en la ventana
282	<i>r_sd_B4sB3</i>	B4 / B3	Desviación estándar de las muestras en la ventana
288	<i>r_sd_B4sB3_pdc</i>	B4 / B3	PDC de desviación estándar de las muestras en la ventana

Atributo	Nombre del atributo	Banda del atributo o bandas entre las que se calcula la proporción	Atributo de origen, derivado, o comparado con proporción
253	<i>r_sumc_B2sB1</i>	B2 / B1	Suma de los cuadrados de las muestras en la ventana
259	<i>r_sumc_B2sB1_pdc</i>	B2 / B1	PDC de suma de los cuadrados de las muestras en la ventana
254	<i>r_sumc_B3sB1</i>	B3 / B1	Suma de los cuadrados de las muestras en la ventana
260	<i>r_sumc_B3sB1_pdc</i>	B3 / B1	PDC de suma de los cuadrados de las muestras en la ventana
256	<i>r_sumc_B3sB2</i>	B3 / B2	Suma de los cuadrados de las muestras en la ventana
262	<i>r_sumc_B3sB2_pdc</i>	B3 / B2	PDC de suma de los cuadrados de las muestras en la ventana
255	<i>r_sumc_B4sB1</i>	B4 / B1	Suma de los cuadrados de las muestras en la ventana
261	<i>r_sumc_B4sB1_pdc</i>	B4 / B1	PDC de suma de los cuadrados de las muestras en la ventana
257	<i>r_sumc_B4sB2</i>	B4 / B2	Suma de los cuadrados de las muestras en la ventana
263	<i>r_sumc_B4sB2_pdc</i>	B4 / B2	PDC de suma de los cuadrados de las muestras en la ventana
258	<i>r_sumc_B4sB3</i>	B4 / B3	Suma de los cuadrados de las muestras en la ventana
264	<i>r_sumc_B4sB3_pdc</i>	B4 / B3	PDC de suma de los cuadrados de las muestras en la ventana

10.5. Selección preliminar de atributos

Los 260 atributos presentados en la Tabla 9 son demasiados como para manejarlos cómodamente. Además, mientras más atributos se agreguen más dispersos se harán los patrones que pudieran tener los casos, y establecer un modelo apropiado se convierte en la búsqueda de una aguja en un pajar. A esto último se le conoce como la *maldición de la dimensionalidad* (Hernández Orallo, Ramírez Quintana, & Ferri Ramírez, 2004, pág. 79). Por esta razón, es necesario reducir el número de atributos a trabajar.

Para ello, se ocuparon las métricas de cantidad de información, junto con el coeficiente de correlación entre variables continuas y una discreta expuestos en la sección de Métricas para la selección de atributos de los Anexos. Cada uno de estos valores mide en escalas distintas, pero comparten la característica de aumentar con la cantidad de información contenida en cada atributo o su correlación con la *clase*. Para poder comparar los resultados de cada medida, es preciso estandarizar dichos valores.

El procedimiento seguido para estandarizar los valores es el siguiente, y parte de que todas las medidas empleadas tienen un rango de valores mayor o igual a cero:

1. El valor de la métrica de cantidad de información (o correlación) i para el atributo j es $m_{i,j}$

2. Al sumar la cantidad de información (o correlación) de todos los atributos para cualquier métrica obtenemos un valor positivo: $\sum_j m_{i,j} = M_i$
3. Para cada métrica, se escala cada valor de $m_{i,j}$ entre M_i , para obtener la proporción de relevancia normalizada $p_{i,j}$ que da el atributo j a la métrica i : $p_{i,j} = m_{i,j}/M_i$. Este valor es adimensional.
4. Como consecuencia de lo anterior, la suma de $p_{i,j}$ para cada métrica i es de uno; y cada uno de estos valores tiene un rango entre cero y uno.
5. La suma de la relevancia normalizada de cada atributo j se obtiene sumando las $p_{i,j}$ para todas las métricas i consideradas (cuatro, en este trabajo): $P_j = \sum_i p_{i,j}$. La suma de P_j para todos los atributos es igual a la cantidad de métricas consideradas.

Una vez calculado el valor de P_j , es sencillo volver a normalizar este valor para que su suma valga uno. A partir de esto, se calculan porcentajes de relevancia para cada atributo, considerando todas las métricas en conjunto. Esta situación se muestra en la Tabla 10.

De los atributos presentados en la Tabla 10, se observa que al considerar a los primeros 32 se cubre el 50% de la relevancia normalizada acumulada. Para fines prácticos, el resto de los atributos resultan redundantes o irrelevantes para indicar la *clase* de cada caso. Es a partir de estos 32 atributos que se realiza la selección de aquellos con una distribución cuasi normal.

Tabla 10: Atributos originales calculados a partir de los sismogramas preprocesados; y sus métricas de cantidad de información y correlación, en bruto y normalizadas.

Orden según P normalizado	Atributo	Nombre del atributo	Correlación	Cantidad de información	Preparación de la ganancia de información	Entredumbre y simétrica	Correlación normalizada	Ganancia de información normalizado	Preparación de la ganancia de información normalizada	Entredumbre y simétrica normalizada	H'	H' normalizado	H' normalizado y acumulado
1	36	r_pbal_01p00_prom	0.24172	0.09610	0.03093	0.05180	0.03391	0.03429	0.02770	0.02941	0.12531	0.09133	0.03133
2	36	r_sin_filt_prom	0.28369	0.08663	0.01130	0.03856	0.03980	0.03030	0.02207	0.11256	0.02814	0.09947	0.05947
3	38	r_sin_filt_sd	0.33078	0.07066	0.01707	0.03094	0.04641	0.02470	0.01634	0.01771	0.10516	0.02629	0.00576
4	85	r_pbal_01p00_sunc	0.03077	0.09957	0.02737	0.04899	0.00432	0.02620	0.02804	0.09337	0.02334	0.10910	0.10910
5	89	r_pbal_01p00_sunc_pdc	0.00021	0.09354	0.02719	0.04897	0.00003	0.03263	0.02604	0.02769	0.08638	0.01159	0.13069
6	138	r_palt_05p00_sd	0.28863	0.04547	0.01278	0.02282	0.04050	0.01590	0.01223	0.01306	0.06168	0.02042	0.15112
7	186	r_pban_01p00a05p00_prom	0.27270	0.04458	0.01310	0.02328	0.03826	0.01558	0.01254	0.01333	0.07971	0.01993	0.17104
8	35	r_sin_filt_sunc	0.04141	0.08865	0.02116	0.03832	0.00581	0.03036	0.02025	0.02194	0.07836	0.01959	0.19063
9	188	r_pban_01p00a05p00_sd	0.26126	0.04291	0.01297	0.02238	0.03665	0.01500	0.01242	0.01315	0.07723	0.01931	0.20994
10	136	r_palt_05p00_prom	0.27079	0.04158	0.01175	0.02097	0.03799	0.01454	0.01125	0.01200	0.07578	0.01894	0.22888
11	135	r_palt_05p00_sunc	0.20292	0.04403	0.01232	0.02200	0.02847	0.01539	0.01179	0.01259	0.06625	0.01706	0.24593
12	90	r_pbal_01p00_prom_pdc	0.00014	0.06585	0.02247	0.03922	0.00002	0.02302	0.02151	0.02245	0.06700	0.01675	0.26270
13	185	r_pban_01p00a05p00_sunc	0.18086	0.04490	0.01357	0.02403	0.02218	0.01570	0.01299	0.01376	0.06462	0.01615	0.27885
14	267	r_prom_84s81	0.21590	0.03290	0.00679	0.01578	0.03029	0.01150	0.00842	0.00909	0.05924	0.01481	0.29366
15	51	phi_sin_filt_sd	0.16923	0.03312	0.01180	0.02048	0.02383	0.01158	0.01130	0.01172	0.05842	0.01461	0.30927
16	83	r_pbal_01p00_sd	0.11372	0.04607	0.01319	0.02351	0.01995	0.01610	0.01263	0.01345	0.05814	0.01454	0.32880
17	253	r_sunc_82s81	0.21308	0.02974	0.00857	0.01526	0.02990	0.01040	0.00821	0.00874	0.05724	0.01431	0.33711
18	265	r_prom_82s81	0.20114	0.02924	0.00617	0.01460	0.02822	0.01022	0.00783	0.00836	0.05462	0.01366	0.35077
19	39	r_sin_filt_sunc_pdc	0.00021	0.06550	0.01572	0.02852	0.00003	0.02290	0.01505	0.01633	0.05431	0.01358	0.36433
20	269	r_prom_84s82	0.13599	0.03493	0.01120	0.01970	0.01908	0.01221	0.01072	0.01127	0.05328	0.01332	0.37767
21	255	r_sunc_84s81	0.15568	0.03430	0.00954	0.01705	0.02184	0.01199	0.00913	0.00976	0.05273	0.01318	0.39083
22	101	phi_pbal_01p00_sd	0.08343	0.03694	0.01212	0.02125	0.01711	0.01291	0.01160	0.01217	0.04839	0.01210	0.40294
23	87	r_pbal_01p00_pend	0.00105	0.04946	0.01529	0.02702	0.00015	0.01729	0.01464	0.01546	0.04754	0.01189	0.41483
24	40	r_sin_filt_prom_pdc	0.00056	0.04895	0.01465	0.02599	0.00008	0.01711	0.01403	0.01487	0.04610	0.01152	0.42635
25	257	r_sunc_84s82	0.07545	0.03436	0.01054	0.01864	0.01059	0.01201	0.01009	0.01067	0.04336	0.01084	0.43719
26	195	theta_pban_01p00a05p00_sd	0.14412	0.01786	0.00626	0.01089	0.02022	0.00624	0.00599	0.00623	0.03869	0.00967	0.44687
27	91	r_pbal_01p00_pend_pdc	0.00011	0.03627	0.01239	0.02162	0.00002	0.01268	0.01186	0.01238	0.03693	0.00923	0.45610
28	37	r_sin_filt_pend	0.00086	0.03774	0.01136	0.02013	0.00012	0.01519	0.01087	0.01152	0.03571	0.00893	0.46503
29	42	r_sin_filt_sd_pdc	0.00048	0.03842	0.01113	0.01980	0.00007	0.01343	0.01065	0.01134	0.03549	0.00887	0.47390
30	221	pol_e_21_84s82	0.14916	0.01236	0.00511	0.00943	0.02983	0.00432	0.00489	0.00497	0.03510	0.00878	0.48265
31	268	r_prom_83s81	0.08118	0.02582	0.00705	0.01263	0.01139	0.00902	0.00875	0.00723	0.03440	0.00860	0.49127
32	41	r_sin_filt_pend_pdc	0.00079	0.03641	0.01097	0.01849	0.00011	0.01273	0.00993	0.01058	0.03335	0.00834	0.49961
33	266	r_prom_83s81	0.11650	0.01844	0.00534	0.00950	0.01635	0.00645	0.00511	0.00544	0.03334	0.00834	0.50795
34	92	r_pbal_01p00_sd_pdc	0.00011	0.02918	0.01171	0.01999	0.00002	0.01020	0.01121	0.01144	0.03287	0.00822	0.51616
35	181	v3_pban_01p00a05p00_theta	0.05910	0.01924	0.00910	0.01515	0.00829	0.00673	0.00872	0.00867	0.03241	0.00810	0.52426
36	139	r_palt_05p00_sunc_pdc	0.00006	0.03669	0.00924	0.01669	0.00001	0.01282	0.00885	0.00955	0.03124	0.00781	0.53207
37	95	theta_pbal_01p00_sd	0.03249	0.03069	0.00789	0.01421	0.00456	0.01073	0.00755	0.00814	0.03097	0.00774	0.53982
38	254	r_sunc_83s81	0.06472	0.02201	0.00716	0.01258	0.00908	0.00769	0.00886	0.00720	0.03083	0.00771	0.54753
39	189	r_pban_01p00a05p00_sunc_pdc	0.00003	0.03409	0.00900	0.01618	0.00000	0.01192	0.00862	0.00926	0.02981	0.00745	0.55498
40	280	r_sd_83s82	0.09242	0.00968	0.00690	0.01058	0.01297	0.00338	0.00661	0.00606	0.02901	0.00725	0.56223
41	209	pol_e_21_84s82	0.10811	0.00718	0.00564	0.00844	0.01517	0.00251	0.00540	0.00483	0.02791	0.00698	0.56921
42	142	r_palt_05p00_sd_pdc	0.00076	0.03080	0.00848	0.01517	0.00011	0.01077	0.00811	0.00868	0.02767	0.00692	0.57613
43	93	theta_pbal_01p00_prom	0.02691	0.02141	0.00835	0.01431	0.00378	0.00748	0.00799	0.00819	0.02745	0.00686	0.58299
44	163	pol_pban_01p00a05p00_pi	0.12116	0.00975	0.00353	0.00612	0.01700	0.00341	0.00338	0.00350	0.02729	0.00682	0.58981
45	256	r_sunc_83s82	0.01397	0.03627	0.00747	0.01132	0.00186	0.00918	0.00715	0.00762	0.02592	0.00648	0.59628
46	162	pol_pban_01p00a05p00_f_1	0.10407	0.00997	0.00377	0.00650	0.01480	0.00348	0.00361	0.00372	0.02542	0.00635	0.60264
47	156	pol_pban_01p00a05p00_e_31	0.00661	0.00661	0.00405	0.00675	0.01412	0.00301	0.00388	0.00623	0.02487	0.00622	0.60886
48	61	pol_pbal_01p00_l_1	0.07160	0.00872	0.00605	0.00933	0.01005	0.00305	0.00579	0.00534	0.02421	0.00606	0.61493
49	45	theta_sin_filt_sd	0.02387	0.01913	0.00730	0.01255	0.00335	0.00669	0.00718	0.00718	0.02421	0.00605	0.62097
50	49	phi_sin_filt_prom	0.02510	0.01841	0.00704	0.01211	0.00352	0.00643	0.00674	0.00693	0.02363	0.00591	0.62688
51	192	r_pban_01p00a05p00_sd_pdc	0.00016	0.02490	0.00734	0.01303	0.00002	0.00870	0.00702	0.00746	0.02321	0.00580	0.63268
52	190	r_pban_01p00a05p00_prom_pdc	0.00030	0.02431	0.00742	0.01312	0.00004	0.00850	0.00710	0.00751	0.02315	0.00579	0.63847
53	183	v3_pban_01p00a05p00_theta_pdc	0.00006	0.01374	0.00989	0.01513	0.00001	0.00480	0.00947	0.00866	0.02294	0.00573	0.64420
54	191	r_pban_01p00a05p00_pend_pdc	0.00038	0.02330	0.00747	0.01314	0.00005	0.00814	0.00715	0.00752	0.02287	0.00572	0.64992
55	140	r_palt_05p00_prom_pdc	0.00074	0.02520	0.00681	0.01221	0.00010	0.00699	0.00652	0.00699	0.02242	0.00561	0.65553
56	157	pol_pban_01p00a05p00_e_32	0.09862	0.00772	0.00303	0.00519	0.01384	0.00270	0.00290	0.00297	0.02241	0.00560	0.66113
57	141	r_palt_05p00_pend_pdc	0.00168	0.02401	0.00677	0.01209	0.00024	0.00692	0.00648	0.00692	0.02203	0.00551	0.66664
58	279	r_sd_84s81	0.00142	0.02400	0.00661	0.01183	0.00020	0.00639	0.00633	0.00677	0.02169	0.00542	0.67206
59	173	v1_pban_01p00a05p00_theta	0.01280	0.01566	0.00744	0.01210	0.00186	0.00947	0.00894	0.00869	0.02113	0.00538	0.67734
60	55	pol_pbal_01p00_e_21	0.04980	0.00754	0.00623	0.00921	0.00698	0.00364	0.00527	0.00527	0.02083	0.00521	0.68255
61	187	r_pban_01p00a05p00_pend	0.00045	0.02121	0.00644	0.01140	0.00006	0.00741	0.00617	0.00653	0.02017	0.00504	0.68759
62	137	r_palt_05p00_pend	0.00056	0.02173	0.00601	0.01075	0.00008	0.00760	0.00575	0.00575	0.01958	0.00490	0.69249
63	60	pol_pbal_01p00_gpp	0.04999	0.00826	0.00480	0.00769	0.00701	0.00289	0.00460	0.00440	0.01890	0.00473	0.69721
64	243	pol_gpp_84s81	0.01752	0.00748	0.00674	0.01064	0.00246	0.00621	0.00732	0.00609	0.01848	0.00462	0.70183
65	58	pol_pbal_01p00_rect_mk	0.03416	0.00756	0.00599	0.00895	0.00479	0.00264	0.00574	0.00512	0.01830	0.00457	0.70641
66	62	pol_pbal_01p00_f_1	0.06233	0.00729	0.00923	0.00544	0.00875	0.00310	0.00311	0.01750	0.00437	0.71078	0.71078
67	56	pol_pbal_01p00_e_31	0.05544	0.00633	0.00352	0.00597	0.00778	0.00291	0.00337	0.00342	0.01748	0.00437	0.71515
68	59	pol_pba_01p00_rect_j	0.03791	0.00625	0.00473	0.00780	0.00532	0.00288	0.00453	0.00435	0.01708	0.00427	0.71942
69	175	v1_pban_01p00a05p00_theta_pdc	0.00025	0.01542	0.00589	0.01012	0.00003	0.00539	0.00564	0.00579	0.01685	0.00421	0.72363
70	63	pol_pbal_01p00_pi	0.03948	0.00811	0.00419	0.00686	0.00554	0.00283	0.00401	0.00393	0.01631	0.00408	0.72771
71	208	pol_e_21_83s82	0.04308	0.00748	0.00390	0.00638	0.00604	0.00262	0.00373	0.00365	0.01604	0.00401	0.73172
72	263	r_sunc_84s82_pdc	0.00073	0.01720	0.00491	0.00875	0.00010	0.00501	0.00470	0.00501	0.01583	0.00396	0.73568
73	270	r_sd_83s81	0.00058	0.01379									

Orden según P normalizado	Atributo	Nombre del atributo	Correlación	Ganancia de información	Proporción de la ganancia de información	Incertidumbre simétrica	Correlación normalizada	Ganancia de información normalizada	Proporción de la ganancia de información normalizada	Incertidumbre simétrica normalizada	Pf	P normalizado	P normalizado y acumulado
101	231	pol_e_32_B4sB1	0.00039	0.00787	0.00452	0.00726	0.00005	0.00275	0.00433	0.00416	0.01129	0.00282	0.83033
102	258	r_sumc_B4sB3	0.00330	0.00806	0.00423	0.00691	0.00046	0.00282	0.00405	0.00396	0.01128	0.00282	0.83136
103	113	pol_palt_05p00_pl	0.04224	0.00301	0.00233	0.00350	0.00593	0.00105	0.00223	0.00200	0.01122	0.00280	0.83596
104	5	pol_sin_filt_e_21	0.02380	0.00411	0.00349	0.00512	0.00334	0.00144	0.00334	0.00293	0.01105	0.00276	0.83872
105	215	pol_e_21_B4sB2_pdc	0.00003	0.00699	0.00465	0.00724	0.00000	0.00244	0.00445	0.00415	0.01105	0.00276	0.84149
106	105	pol_palt_05p00_e_21	0.05407	0.00257	0.00127	0.00210	0.00759	0.00090	0.00122	0.00120	0.01090	0.00275	0.84421
107	277	r_sd_B2sB1	0.00217	0.01009	0.00353	0.00614	0.00030	0.00353	0.00338	0.00351	0.01072	0.00268	0.84689
108	19	pol_sin_filt_gpp_pdc	0.00033	0.00237	0.00603	0.00636	0.00005	0.00100	0.00577	0.00384	0.01046	0.00263	0.84953
109	106	pol_palt_05p00_rect_mk	0.05045	0.00257	0.00127	0.00210	0.00708	0.00090	0.00122	0.00120	0.01039	0.00260	0.85210
110	207	pol_e_21_B4sB1	0.00329	0.00623	0.00417	0.00649	0.00046	0.00213	0.00400	0.00371	0.01035	0.00259	0.85469
111	11	pol_sin_filt_l_1	0.03296	0.00418	0.00220	0.00359	0.00462	0.00146	0.00210	0.00205	0.01024	0.00256	0.85723
112	226	pol_e_31_B3sB2_pdc	0.00000	0.00711	0.00402	0.00647	0.00000	0.00248	0.00384	0.00370	0.01003	0.00251	0.85978
113	8	pol_sin_filt_rect_mk	0.01599	0.00411	0.00349	0.00512	0.00224	0.00144	0.00334	0.00293	0.00996	0.00249	0.86225
114	13	pol_sin_filt_pl	0.00943	0.00289	0.00448	0.00539	0.00132	0.00101	0.00429	0.00309	0.00972	0.00243	0.86468
115	217	pol_e_31_B2sB1	0.00048	0.00753	0.00366	0.00606	0.00007	0.00263	0.00350	0.00347	0.00967	0.00242	0.86710
116	145	theta_palt_05p00_sd	0.05136	0.00196	0.00078	0.00133	0.00721	0.00069	0.00074	0.00074	0.00939	0.00235	0.86944
117	272	r_prom_B3sB1_pdc	0.00075	0.00699	0.00361	0.00591	0.00011	0.00244	0.00345	0.00338	0.00939	0.00235	0.87179
118	94	theta_pbaj_01p00_pend	0.00010	0.00919	0.00311	0.00544	0.00001	0.00321	0.00298	0.00311	0.00932	0.00233	0.87412
119	12	pol_sin_filt_f_1	0.01690	0.00247	0.00363	0.00446	0.00237	0.00086	0.00348	0.00355	0.00927	0.00232	0.87644
120	10	pol_sin_filt_gpp	0.02593	0.00416	0.00218	0.00357	0.00367	0.00145	0.00209	0.00204	0.00923	0.00231	0.87874
121	234	pol_e_32_B4sB3	0.00030	0.00739	0.00336	0.00563	0.00004	0.00258	0.00322	0.00322	0.00906	0.00227	0.88101
122	73	v1_pbaj_01p00_theta	0.02492	0.00577	0.00177	0.00315	0.00350	0.00202	0.00169	0.00179	0.00899	0.00225	0.88328
123	285	r_sd_B4sB1_pdc	0.00000	0.00184	0.00363	0.00502	0.00000	0.00068	0.00338	0.00327	0.00889	0.00223	0.88549
124	270	r_prom_B4sB3	0.00320	0.00573	0.00247	0.00552	0.00045	0.00200	0.00333	0.00316	0.00889	0.00223	0.88772
125	99	phi_pbaj_01p00_prom	0.01783	0.00696	0.00302	0.00359	0.00250	0.00243	0.00193	0.00203	0.00889	0.00223	0.88996
126	9	pol_sin_filt_rect_l	0.01317	0.00413	0.00301	0.00459	0.00185	0.00144	0.00238	0.00263	0.00880	0.00220	0.89218
127	6	pol_sin_filt_e_31	0.01470	0.00335	0.00299	0.00433	0.00206	0.00117	0.00286	0.00248	0.00857	0.00214	0.89430
128	54	phi_sin_filt_sd_pdc	0.00139	0.00643	0.00317	0.00524	0.00020	0.00225	0.00304	0.00300	0.00848	0.00212	0.89642
129	161	pol_pban_01p00a05p00_l_1	0.00348	0.00275	0.00139	0.00238	0.00482	0.00096	0.00133	0.00133	0.00842	0.00211	0.89853
130	213	pol_e_21_B4sB1_pdc	0.00000	0.00562	0.00338	0.00538	0.00000	0.00197	0.00324	0.00308	0.00828	0.00207	0.90060
131	149	phi_palt_05p00_prom	0.03718	0.00202	0.00121	0.00193	0.00522	0.00071	0.00116	0.00110	0.00818	0.00205	0.90264
132	24	v1_sin_filt_phi	0.01760	0.00587	0.00181	0.00319	0.00247	0.00205	0.00173	0.00183	0.00808	0.00202	0.90466
133	97	theta_pbaj_01p00_pend_pdc	0.00033	0.00763	0.00269	0.00467	0.00005	0.00267	0.00257	0.00267	0.00796	0.00199	0.90665
134	286	r_sd_B3sB2_pdc	0.00049	0.00586	0.00297	0.00488	0.00007	0.00205	0.00284	0.00279	0.00775	0.00194	0.90859
135	242	pol_gpp_B3sB1	0.03227	0.00279	0.00113	0.00193	0.00453	0.00097	0.00109	0.00111	0.00769	0.00192	0.91051
136	172	pol_pban_01p00a05p00_pl_pdc	0.00022	0.00578	0.00296	0.00486	0.00003	0.00202	0.00283	0.00278	0.00766	0.00192	0.91243
137	18	pol_sin_filt_e_21_pdc	0.00026	0.00152	0.00389	0.00413	0.00004	0.00053	0.00469	0.00236	0.00763	0.00191	0.91434
138	218	pol_e_31_B3sB1	0.00352	0.00564	0.00270	0.00449	0.00049	0.00197	0.00259	0.00257	0.00762	0.00191	0.91624
139	98	theta_pbaj_01p00_sd_pdc	0.00040	0.00837	0.00223	0.00400	0.00006	0.00282	0.00213	0.00229	0.00740	0.00185	0.91809
140	102	phi_pbaj_01p00_prom_pdc	0.00086	0.00775	0.00227	0.00404	0.00012	0.00271	0.00218	0.00231	0.00732	0.00183	0.91992
141	275	r_prom_B4sB2_pdc	0.00087	0.00755	0.00211	0.00378	0.00012	0.00264	0.00202	0.00216	0.00695	0.00174	0.92166
142	31	v3_sin_filt_theta	0.01303	0.00475	0.00174	0.00301	0.00183	0.00166	0.00166	0.00172	0.00688	0.00172	0.92338
143	81	v3_pbaj_01p00_theta	0.00709	0.00561	0.00189	0.00331	0.00099	0.00196	0.00181	0.00190	0.00666	0.00167	0.92504
144	151	phi_palt_05p00_sd	0.03713	0.00101	0.00046	0.00078	0.00521	0.00035	0.00044	0.00044	0.00645	0.00161	0.92666
145	271	r_prom_B2sB1_pdc	0.00117	0.00599	0.00205	0.00357	0.00016	0.00209	0.00196	0.00205	0.00626	0.00157	0.92822
146	228	pol_e_31_B4sB3_pdc	0.00000	0.00481	0.00238	0.00392	0.00000	0.00168	0.00227	0.00225	0.00620	0.00155	0.92977
147	43	theta_sin_filt_prom	0.01885	0.00269	0.00135	0.00222	0.00284	0.00094	0.00129	0.00127	0.00615	0.00154	0.93131
148	75	v1_pbaj_01p00_theta_pdc	0.00027	0.00704	0.00179	0.00323	0.00004	0.00246	0.00172	0.00185	0.00606	0.00152	0.93282
149	206	pol_e_21_B3sB1	0.00329	0.00443	0.00210	0.00350	0.00046	0.00155	0.00201	0.00200	0.00603	0.00151	0.93433
150	260	r_sumc_B3sB1_pdc	0.00105	0.00535	0.00186	0.00325	0.00015	0.00187	0.00178	0.00185	0.00565	0.00141	0.93574
151	244	pol_gpp_B3sB2	0.02703	0.00118	0.00078	0.00121	0.00379	0.00041	0.00074	0.00069	0.00564	0.00141	0.93715
152	29	v2_sin_filt_phi	0.01170	0.00422	0.00126	0.00224	0.00164	0.00147	0.00121	0.00123	0.00561	0.00140	0.93856
153	159	pol_pban_01p00a05p00_rect_l	0.02385	0.00139	0.00079	0.00128	0.00383	0.00049	0.00076	0.00073	0.00560	0.00140	0.93996
154	284	r_sd_B3sB1_pdc	0.00002	0.00328	0.00243	0.00369	0.00000	0.00115	0.00233	0.00211	0.00559	0.00140	0.94136
155	232	pol_e_32_B3sB2	0.02949	0.00064	0.00063	0.00088	0.00414	0.00022	0.00060	0.00050	0.00546	0.00137	0.94272
156	124	v1_palt_05p00_phi	0.01681	0.00315	0.00101	0.00178	0.00236	0.00110	0.00097	0.00102	0.00545	0.00136	0.94408
157	274	r_prom_B3sB2_pdc	0.00109	0.00557	0.00169	0.00300	0.00015	0.00195	0.00162	0.00172	0.00543	0.00136	0.94544
158	107	pol_palt_05p00_e_32	0.01432	0.00073	0.00203	0.00186	0.00201	0.00026	0.00194	0.00106	0.00527	0.00132	0.94676
159	224	pol_e_31_B3sB1_pdc	0.00001	0.00379	0.00209	0.00338	0.00000	0.00133	0.00200	0.00194	0.00527	0.00132	0.94808
160	27	v2_sin_filt_theta	0.02567	0.00093	0.00067	0.00102	0.00360	0.00033	0.00064	0.00058	0.00515	0.00129	0.94936
161	165	pol_pban_01p00a05p00_e_31_pdc	0.00029	0.00323	0.00206	0.00324	0.00004	0.00113	0.00197	0.00185	0.00499	0.00125	0.95061
162	96	theta_pbaj_01p00_prom_pdc	0.00033	0.00484	0.00163	0.00285	0.00005	0.00169	0.00156	0.00163	0.00493	0.00123	0.95184
163	123	v1_palt_05p00_theta	0.01315	0.00263	0.00091	0.00159	0.00092	0.00087	0.00091	0.00091	0.00485	0.00114	0.95298
164	21	pol_sin_filt_f_1_pdc	0.00030	0.00234	0.00194	0.00286	0.00004	0.00082	0.00186	0.00164	0.00436	0.00109	0.95407
165	57	pol_pbaj_01p00_e_32	0.02108	0.00076	0.00057	0.00086	0.00296	0.00027	0.00054	0.00049	0.00426	0.00106	0.95514
166	33	v3_sin_filt_theta_pdc	0.00018	0.00518	0.00163	0.00268	0.00002	0.00111	0.00156	0.00153	0.00424	0.00106	0.95619
167	77	v2_pbaj_01p00_theta	0.01328	0.00207	0.00081	0.00140	0.00186	0.00072	0.00060	0.00078	0.00417	0.00104	0.95734
168	241	pol_gpp_B2sB1	0.01990	0.00112	0.00049	0.00083	0.00279	0.00039	0.00047	0.00047	0.00413	0.00103	0.95827
169	212	pol_e_21_B3sB1_pdc	0.00000	0.00340	0.00242	0.00341	0.00000	0.00119	0.00136	0.00138	0.00393	0.00098	0.95925
170	132	v3_palt_05p00_phi	0.00557	0.00359	0.00108	0.00192	0.00050	0.00126	0.00104	0.00110	0.00389	0.00097	0.96022
171	237	pol_e_32_B4sB1_pdc	0.00000	0.00297	0.00142	0.00236	0.00000	0.00104	0.00136	0.00135	0.00374	0.00094	0.96116
172	131	v3_palt_05p00_theta	0.01248	0.00179	0.00069	0.00118	0.00175	0.00062	0.00066	0.00071	0.00371	0.00093	0.96209
173	15	pol_sin_filt_e_31_pdc	0.00030	0.00173	0.00170	0.00239	0.00004	0.00060	0.00163	0.00137	0.00364	0.00091	0.

Orden según Pj normalizado	Atributo	Nombre del atributo	Correlación	Generación de información	Proporción de la ganancia de información	Incrédulumbre simétrica	Correlación normalizada	Generación de información normalizada	Proporción de la ganancia de información normalizada	Incrédulumbre simétrica normalizada	Pj	Pj normalizado	Pj normalizado y acumulado
201	174	v1_pban_01p00a05p00_phi	0.00187	0.00180	0.00064	0.00111	0.00026	0.00063	0.00061	0.00063	0.00214	0.00053	0.98230
202	276	r_prom_8483_pdc	0.00000	0.00064	0.00118	0.00132	0.00000	0.00023	0.00113	0.00076	0.00211	0.00053	0.98283
203	50	phi_sin_fit_pend	0.00106	0.00176	0.00069	0.00119	0.00015	0.00062	0.00066	0.00068	0.00211	0.00053	0.98335
204	169	pol_pban_01p00a05p00_gpp_pdc	0.00026	0.00115	0.00091	0.00136	0.00004	0.00040	0.00087	0.00078	0.00209	0.00052	0.98388
205	77	pol_sin_fit_e_32	0.00391	0.00079	0.00070	0.00102	0.00055	0.00028	0.00067	0.00058	0.00208	0.00052	0.98440
206	182	v3_pban_01p00a05p00_phi	0.00963	0.00041	0.00029	0.00045	0.00135	0.00014	0.00028	0.00026	0.00203	0.00051	0.98490
207	114	pol_palt_05p00_e_21_pdc	0.00082	0.00070	0.00099	0.00123	0.00012	0.00025	0.00099	0.00070	0.00200	0.00050	0.98540
208	133	v3_palt_05p00_theta_pdc	0.00002	0.00168	0.00073	0.00122	0.00000	0.00059	0.00070	0.00070	0.00198	0.00050	0.98590
209	123	r_sd_8381_pdc	0.00000	0.00135	0.00050	0.00127	0.00000	0.00047	0.00076	0.00073	0.00196	0.00049	0.98639
210	22	pol_sin_fit_pl_pdc	0.00025	0.00097	0.00086	0.00124	0.00004	0.00034	0.00082	0.00071	0.00191	0.00048	0.98687
211	170	pol_pban_01p00a05p00_l_1_pdc	0.00024	0.00045	0.00112	0.00108	0.00003	0.00016	0.00107	0.00062	0.00188	0.00047	0.98734
212	48	theta_sin_fit_sd_pdc	0.00004	0.00127	0.00076	0.00122	0.00001	0.00044	0.00073	0.00070	0.00188	0.00047	0.98781
213	288	r_sd_8483_pdc	0.00000	0.00063	0.00098	0.00118	0.00000	0.00022	0.00094	0.00067	0.00183	0.00046	0.98828
214	46	theta_sin_fit_prom_pdc	0.00008	0.00059	0.00095	0.00113	0.00001	0.00021	0.00091	0.00065	0.00178	0.00044	0.98871
215	72	pol_pban_01p00_phi_pdc	0.00000	0.00157	0.00063	0.00108	0.00000	0.00055	0.00060	0.00062	0.00177	0.00044	0.98915
216	116	pol_palt_05p00_e_32_pdc	0.00032	0.00009	0.00154	0.00037	0.00005	0.00003	0.00147	0.00021	0.00176	0.00044	0.98959
217	210	pol_e_21_8483	0.00021	0.00130	0.00067	0.00110	0.00003	0.00045	0.00064	0.00063	0.00176	0.00044	0.99003
218	117	pol_palt_05p00_rect_mk_pdc	0.00076	0.00035	0.00105	0.00090	0.00011	0.00012	0.00100	0.00052	0.00175	0.00044	0.99047
219	44	theta_sin_fit_pend	0.00098	0.00090	0.00081	0.00095	0.00014	0.00031	0.00059	0.00054	0.00158	0.00039	0.99088
220	127	v2_palt_05p00_theta	0.00443	0.00084	0.00032	0.00055	0.00062	0.00029	0.00031	0.00032	0.00154	0.00039	0.99125
221	47	theta_sin_fit_pend_pdc	0.00038	0.00098	0.00060	0.00095	0.00005	0.00034	0.00057	0.00054	0.00151	0.00038	0.99163
222	229	pol_e_32_8381	0.00039	0.00073	0.00066	0.00095	0.00006	0.00026	0.00048	0.00055	0.00149	0.00037	0.99200
223	125	v1_palt_05p00_theta_pdc	0.00016	0.00116	0.00051	0.00087	0.00002	0.00041	0.00049	0.00050	0.00142	0.00036	0.99235
224	198	phi_pban_01p00a05p00_prom	0.00772	0.00013	0.00017	0.00022	0.00108	0.00005	0.00016	0.00012	0.00141	0.00035	0.99271
225	250	pol_gpp_8381_pdc	0.00095	0.00025	0.00078	0.00087	0.00013	0.00009	0.00075	0.00038	0.00135	0.00034	0.99304
226	68	pol_pban_01p00_rect_j_pdc	0.00030	0.00095	0.00051	0.00083	0.00004	0.00033	0.00048	0.00048	0.00134	0.00033	0.99338
227	216	pol_e_21_8483_pdc	0.00000	0.00066	0.00062	0.00089	0.00000	0.00023	0.00060	0.00051	0.00134	0.00033	0.99371
228	67	pol_pban_01p00_rect_mk_pdc	0.00033	0.00101	0.00048	0.00080	0.00005	0.00035	0.00046	0.00046	0.00132	0.00033	0.99404
229	115	pol_palt_05p00_e_31_pdc	0.00092	0.00056	0.00056	0.00079	0.00013	0.00020	0.00054	0.00045	0.00131	0.00033	0.99437
230	236	pol_e_32_8381_pdc	0.00008	0.00075	0.00052	0.00081	0.00001	0.00026	0.00050	0.00046	0.00124	0.00031	0.99468
231	238	pol_e_32_8381_pdc	0.00005	0.00037	0.00067	0.00076	0.00001	0.00013	0.00064	0.00044	0.00122	0.00030	0.99499
232	32	v3_sin_fit_phi	0.00050	0.00091	0.00039	0.00085	0.00007	0.00032	0.00037	0.00037	0.00113	0.00028	0.99527
233	128	v2_palt_05p00_phi	0.00363	0.00047	0.00023	0.00038	0.00051	0.00016	0.00022	0.00011	0.00111	0.00028	0.99555
234	168	pol_pban_01p00a05p00_rect_j_pdc	0.00024	0.00047	0.00052	0.00071	0.00003	0.00016	0.00050	0.00041	0.00111	0.00028	0.99582
235	26	v1_sin_fit_phi_pdc	0.00009	0.00095	0.00038	0.00085	0.00001	0.00033	0.00036	0.00037	0.00108	0.00027	0.99609
236	118	pol_palt_05p00_rect_j_pdc	0.00088	0.00041	0.00044	0.00080	0.00012	0.00014	0.00042	0.00034	0.00103	0.00026	0.99635
237	122	pol_palt_05p00_phi_pdc	0.00080	0.00019	0.00055	0.00049	0.00011	0.00007	0.00052	0.00028	0.00098	0.00024	0.99659
238	30	v2_sin_fit_phi_pdc	0.00000	0.00093	0.00020	0.00053	0.00000	0.00033	0.00029	0.00030	0.00092	0.00023	0.99682
239	69	pol_pban_01p00_gpp_pdc	0.00032	0.00061	0.00032	0.00052	0.00005	0.00021	0.00030	0.00030	0.00086	0.00021	0.99704
240	264	r_sumc_8483_pdc	0.00000	0.00051	0.00053	0.00000	0.00000	0.00018	0.00033	0.00033	0.00081	0.00020	0.99724
241	129	v2_palt_05p00_theta_pdc	0.00011	0.00074	0.00026	0.00045	0.00002	0.00026	0.00025	0.00026	0.00078	0.00019	0.99744
242	184	v3_pban_01p00a05p00_phi_pdc	0.00011	0.00013	0.00050	0.00048	0.00002	0.00005	0.00048	0.00022	0.00076	0.00019	0.99763
243	78	v2_pban_01p00_phi	0.00048	0.00060	0.00025	0.00042	0.00007	0.00021	0.00024	0.00024	0.00075	0.00019	0.99781
244	249	pol_gpp_8481_pdc	0.00022	0.00052	0.00025	0.00042	0.00003	0.00018	0.00024	0.00024	0.00069	0.00017	0.99799
245	247	pol_gpp_8381_pdc	0.00025	0.00025	0.00024	0.00034	0.00003	0.00009	0.00023	0.00020	0.00055	0.00014	0.99812
246	180	v2_pban_01p00a05p00_phi_pdc	0.00022	0.00044	0.00016	0.00027	0.00003	0.00015	0.00015	0.00016	0.00049	0.00012	0.99825
247	176	v1_pban_01p00a05p00_phi_pdc	0.00001	0.00042	0.00017	0.00029	0.00000	0.00015	0.00017	0.00017	0.00048	0.00012	0.99837
248	126	v1_palt_05p00_phi_pdc	0.00019	0.00043	0.00014	0.00024	0.00003	0.00015	0.00015	0.00015	0.00045	0.00011	0.99848
249	34	v3_sin_fit_phi_pdc	0.00059	0.00013	0.00018	0.00023	0.00008	0.00005	0.00017	0.00013	0.00043	0.00011	0.99859
250	134	v3_palt_05p00_phi_pdc	0.00018	0.00024	0.00015	0.00024	0.00003	0.00003	0.00014	0.00013	0.00039	0.00010	0.99869
251	29	v2_sin_fit_theta_pdc	0.00017	0.00029	0.00013	0.00022	0.00002	0.00010	0.00013	0.00013	0.00038	0.00010	0.99878
252	223	pol_e_31_8381_pdc	0.00000	0.00027	0.00015	0.00024	0.00000	0.00010	0.00014	0.00014	0.00034	0.00009	0.99883
253	65	pol_pban_01p00_e_31_pdc	0.00010	0.00022	0.00014	0.00022	0.00001	0.00003	0.00013	0.00013	0.00035	0.00009	0.99896
254	211	pol_e_21_8381_pdc	0.00000	0.00023	0.00014	0.00022	0.00000	0.00008	0.00013	0.00013	0.00034	0.00008	0.99905
255	194	theta_pban_01p00a05p00_pend	0.00129	0.00006	0.00006	0.00008	0.00018	0.00002	0.00005	0.00005	0.00030	0.00008	0.99912
256	193	theta_pban_01p00a05p00_prom	0.00108	0.00008	0.00006	0.00009	0.00015	0.00003	0.00006	0.00005	0.00028	0.00007	0.99919
257	64	pol_pban_01p00_e_21_pdc	0.00036	0.00015	0.00010	0.00015	0.00005	0.00005	0.00009	0.00009	0.00028	0.00007	0.99927
258	76	v1_pban_01p00_phi_pdc	0.00003	0.00018	0.00011	0.00018	0.00000	0.00006	0.00011	0.00010	0.00028	0.00007	0.99934
259	204	phi_pban_01p00a05p00_sd_pdc	0.00013	0.00008	0.00014	0.00016	0.00002	0.00013	0.00009	0.00009	0.00027	0.00007	0.99940
260	130	v2_palt_05p00_phi_pdc	0.00011	0.00017	0.00010	0.00017	0.00002	0.00006	0.00010	0.00009	0.00027	0.00007	0.99947
261	144	theta_palt_05p00_pend	0.00172	0.00000	0.00000	0.00000	0.00024	0.00000	0.00000	0.00000	0.00024	0.00006	0.99953
262	248	pol_gpp_8381_pdc	0.00051	0.00007	0.00008	0.00010	0.00007	0.00002	0.00007	0.00006	0.00023	0.00006	0.99959
263	235	pol_e_32_8381_pdc	0.00000	0.00013	0.00007	0.00012	0.00000	0.00004	0.00007	0.00006	0.00018	0.00005	0.99963
264	80	v2_pban_01p00_phi_pdc	0.00013	0.00008	0.00008	0.00011	0.00002	0.00003	0.00007	0.00006	0.00018	0.00005	0.99968
265	196	theta_pban_01p00a05p00_prom_pdc	0.00003	0.00007	0.00008	0.00011	0.00000	0.00003	0.00008	0.00006	0.00017	0.00004	0.99972
266	84	v3_pban_01p00_phi_pdc	0.00008	0.00010	0.00006	0.00010	0.00001	0.00004	0.00006	0.00006	0.00016	0.00004	0.99976
267	71	pol_pban_01p00_e_1_pdc	0.00002	0.00010	0.00006	0.00010	0.00000	0.00003	0.00006	0.00006	0.00015	0.00004	0.99980
268	66	pol_pban_01p00_e_32_pdc	0.00024	0.00007	0.00005	0.00007	0.00003	0.00002	0.00005	0.00004	0.00014	0.00004	0.99983
269	251	pol_gpp_8481_pdc	0.00011	0.00007	0.00006	0.00008	0.00001	0.00002	0.00006	0.00005	0.00014	0.00004	0.99987
270	150	phi_palt_05p00_pend	0.00071	0.00000	0.00000	0.00000	0.00010	0.00000	0.00000	0.00000	0.00010	0.00003	0.99989
271	252	pol_gpp_8483_pdc	0.00066	0.00000	0.00000	0.00000	0.00009	0.00000	0.00000	0.00000	0.00009	0.00002	0.99992
272	152	phi_palt_05p00_prom_pdc	0.00052	0.00000	0.00000	0.00000	0.00007	0.00000	0.00000	0.00000	0.00007		

de Preprocesamiento en SAC. Como se comenta en la sección de Datos disponibles, varios de los sismogramas fueron segmentados en tramos de una longitud manejable. Por esta razón, de un sismograma pueden surgir hasta cuatro registros. Para fines de visualización, todos los tramos de un mismo sismograma se han unido en una sola figura, en la cual se muestran las componentes sin filtrar. De haber algún segmento con falla instrumental, la falla se ha sustituido por un valor constante de cero. La información básica de cada registro, después del preprocesamiento se incluye en la Tabla 12, mostrada después de la Tabla 11.

Tabla 11: Sismogramas trabajados

Registro	Sismograma original	Sismograma preprocesado
<i>GOLO 287</i>	Figura 14	Figura 15
<i>GOLO 288</i>	Figura 16	Figura 17
<i>GOLO 296</i>	Figura 18	Figura 19
<i>GOLO 297</i>	Figura 20	Figura 21
<i>GOLO 298</i>	Figura 22	Figura 23
<i>GOLO 299</i>	Figura 24	Figura 25
<i>GOLO 301</i>	Figura 26	Figura 27
<i>GOLO 302</i>	Figura 28	Figura 29
<i>GOLO 303</i>	Figura 30	Figura 31
<i>GOLO 304</i>	Figura 32	Figura 33
<i>GOLO 315</i>	Figura 34	Figura 35
<i>GOLO 316</i>	Figura 36	Figura 37
<i>GOLO 317</i>	Figura 38	Figura 39
<i>GOLO 320</i>	Figura 40	Figura 41
<i>GOLO 321</i>	Figura 42	Figura 43
<i>GOLO 322</i>	Figura 44	Figura 45
<i>GOLO 323</i>	Figura 46	Figura 47
<i>GOLO 324</i>	Figura 48	Figura 49
<i>GOLO 325</i>	Figura 50	Figura 51

Tabla 12: Información sobre cada sismograma resultante del preprocesamiento. Para cada registro preprocesado se indica el registro original. Las columnas de año, día juliano, hora, minuto, segundo, y milisegundo se refieren al instante de tiempo (UTC) en el que se tomó la primera muestra del registro preprocesado.

Registro de origen	Registro preprocesado	Periodo de muestreo [s]	Duración [muestras]	Año	Día juliano	Hora	Minuto	Segundo	Milisegundo
GOLO 287	GOLO 287	0.01	8,640,000	2006	287	00	00	00	0000
GOLO 288	GOLO 288	0.01	4,680,000	2006	288	00	00	00	0000
GOLO 296	GOLO 296	0.01	1,938,038	2006	296	18	36	41	0660
GOLO 297	GOLO 297-1	0.01	3,992,678	2006	297	00	00	00	0000
GOLO 297	GOLO 297-2	0.01	4,647,323	2006	297	11	05	26	0770
GOLO 298	GOLO 298-1	0.01	3,991,764	2006	298	00	00	00	0000
GOLO 298	GOLO 298-2	0.01	2,847,716	2006	298	11	05	17	0629
GOLO 298	GOLO 298-3	0.01	1,467,866	2006	298	19	55	21	0336
GOLO 299	GOLO 299	0.01	673,647	2006	299	00	00	00	0000
GOLO 301	GOLO 301	0.01	1,533,952	2006	301	19	43	80	0480
GOLO 302	GOLO 302-1	0.01	1,079,434	2006	302	00	00	00	0000
GOLO 302	GOLO 302-2	0.01	3,089,057	2006	302	03	20	09	0430
GOLO 302	GOLO 302-3	0.01	4,350,001	2006	302	11	54	59	0988
GOLO 303	GOLO 303-1	0.01	3,566,250	2006	303	00	00	00	0000
GOLO 303	GOLO 303-2	0.01	3,168,151	2006	303	09	54	22	0488
GOLO 303	GOLO 303-3	0.01	273,780	2006	303	18	52	29	0992
GOLO 303	GOLO 303-4	0.01	1,563,839	2006	303	19	39	21	0609
GOLO 304	GOLO 304-1	0.01	1,194,376	2006	304	00	00	00	0000
GOLO 304	GOLO 304-2	0.01	1,214,688	2006	304	03	19	03	0750
GOLO 304	GOLO 304-3	0.01	1,165,938	2006	304	06	41	30	0619
GOLO 304	GOLO 304-4	0.01	1,105,001	2006	304	09	55	49	0988
GOLO 315	GOLO 315-1	0.01	2,637,813	2006	315	02	17	51	0870
GOLO 315	GOLO 315-2	0.01	2,285,414	2006	315	09	37	29	0990
GOLO 315	GOLO 315-3	0.01	1,074,645	2006	315	21	00	53	0547
GOLO 316	GOLO 316-1	0.01	1,605,000	2006	316	00	00	00	0000
GOLO 316	GOLO 316-2	0.01	2,400,001	2006	316	04	27	29	0989
GOLO 316	GOLO 316-3	0.01	2,250,001	2006	316	11	07	29	0988
GOLO 316	GOLO 316-4	0.01	2,385,001	2006	316	17	22	29	0988
GOLO 317	GOLO 317-1	0.01	1,530,000	2006	317	00	00	00	0000
GOLO 317	GOLO 317-2	0.01	2,752,501	2006	317	04	14	59	0989
GOLO 317	GOLO 317-3	0.01	2,002,501	2006	317	11	53	44	0988
GOLO 317	GOLO 317-4	0.01	2,355,001	2006	317	17	27	29	0988
GOLO 320	GOLO 320	0.01	1,205,685	2006	320	20	38	46	0470
GOLO 321	GOLO 321-1	0.01	1,991,250	2006	321	00	00	00	0000
GOLO 321	GOLO 321-2	0.01	1,653,751	2006	321	05	31	52	0490
GOLO 321	GOLO 321-3	0.01	1,940,626	2006	321	10	07	29	0988
GOLO 321	GOLO 321-4	0.01	3,054,376	2006	321	15	30	56	0238
GOLO 322	GOLO 322-1	0.01	1,993,800	2006	322	00	00	00	0000
GOLO 322	GOLO 322-2	0.01	1,977,718	2006	322	05	32	17	0990
GOLO 322	GOLO 322-3	0.01	2,454,275	2006	322	11	01	55	0160
GOLO 322	GOLO 322-4	0.01	2,214,210	2006	322	17	50	57	0898
GOLO 323	GOLO 323-1	0.01	2,272,500	2006	323	00	00	00	0000
GOLO 323	GOLO 323-2	0.01	1,468,126	2006	323	06	18	44	0990
GOLO 323	GOLO 323-3	0.01	1,721,251	2006	323	10	23	26	0238
GOLO 323	GOLO 323-4	0.01	3,178,126	2006	323	15	10	18	0738
GOLO 324	GOLO 324-1	0.01	2,581,875	2006	324	00	00	00	0000
GOLO 324	GOLO 324-2	0.01	2,205,001	2006	324	07	10	18	0740

<i>Registro de origen</i>	<i>Registro preprocesado</i>	<i>Periodo de muestreo [s]</i>	<i>Duración [muestras]</i>	<i>Año</i>	<i>Día juliano</i>	<i>Hora</i>	<i>Minuto</i>	<i>Segundo</i>	<i>Milisegundo</i>
GOLO 324	GOLO 324-3	0.01	1,878,752	2006	324	13	17	48	0738
GOLO 324	GOLO 324-4	0.01	1,974,375	2006	324	18	30	56	0250
GOLO 325	GOLO 325-1	0.01	1,882,500	2006	325	00	00	00	0000
GOLO 325	GOLO 325-2	0.01	1,852,501	2006	325	05	13	44	0990
GOLO 325	GOLO 325-3	0.01	2,542,501	2006	325	10	22	29	0988
GOLO 325	GOLO 325-4	0.01	2,362,501	2006	325	17	26	14	0988

A continuación, se presentan las figuras de los registros.

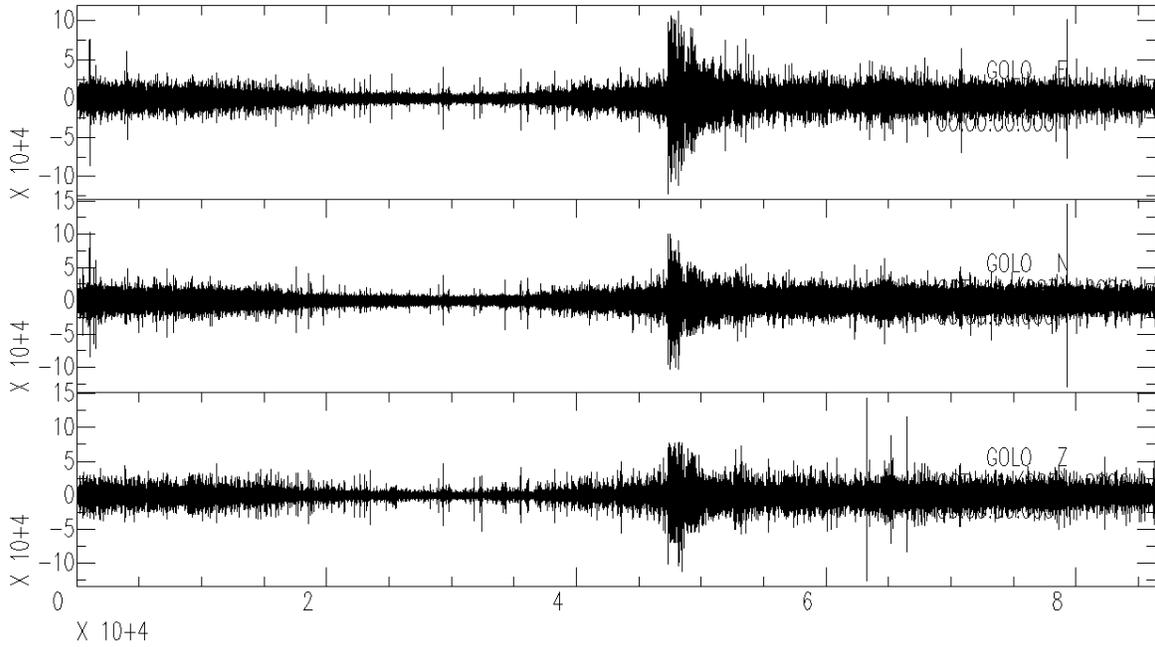


Figura 14: GLOL 287, original.

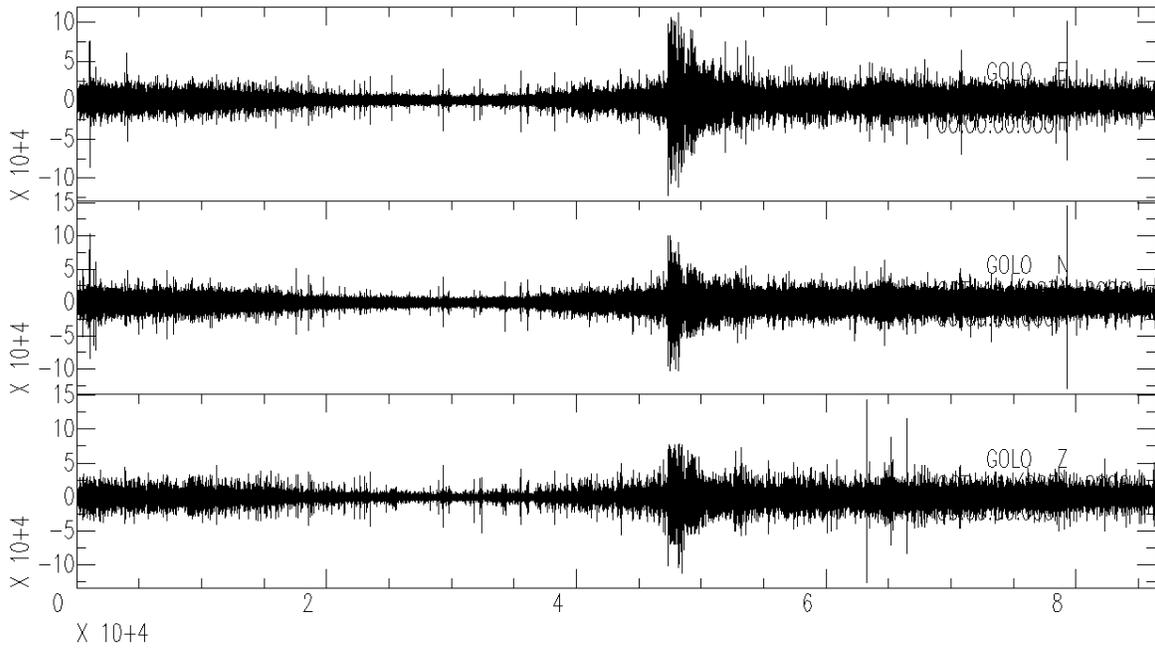


Figura 15: GLOL 287, preprocesado.

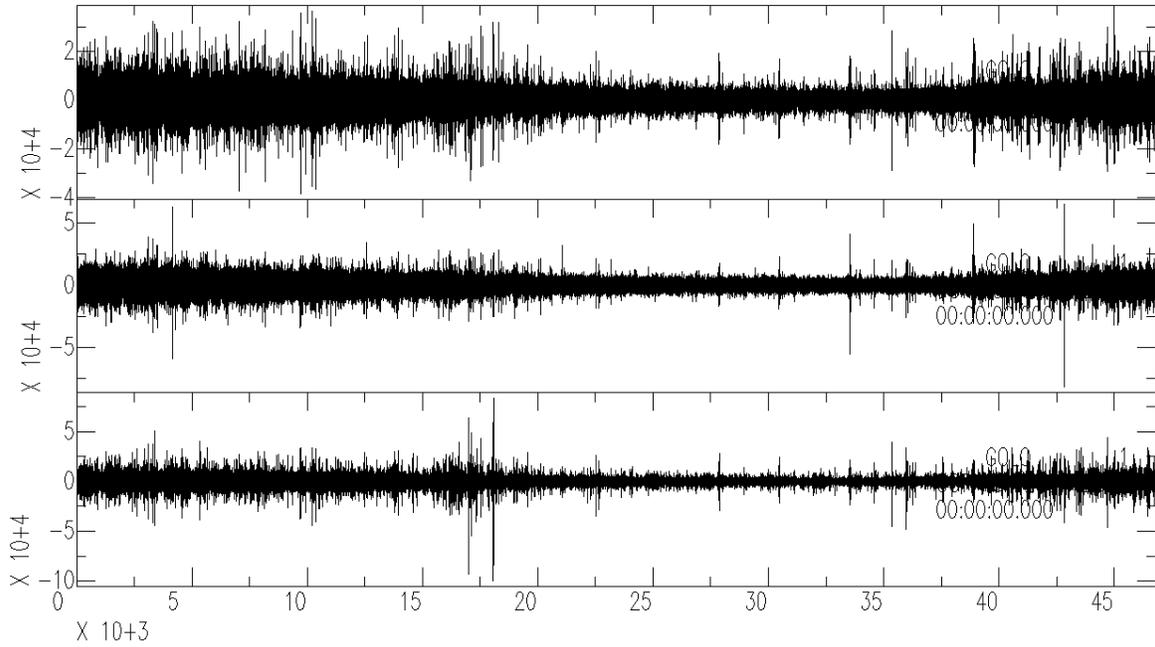


Figura 16: GOLO 288, original.

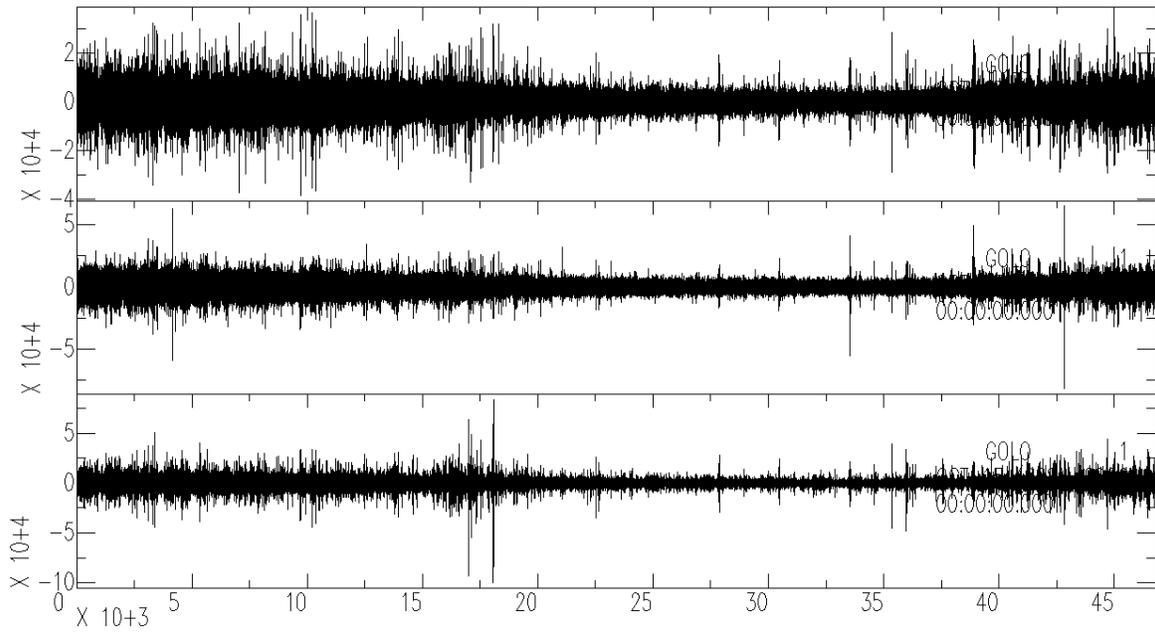


Figura 17: GOLO 288, preprocesado.

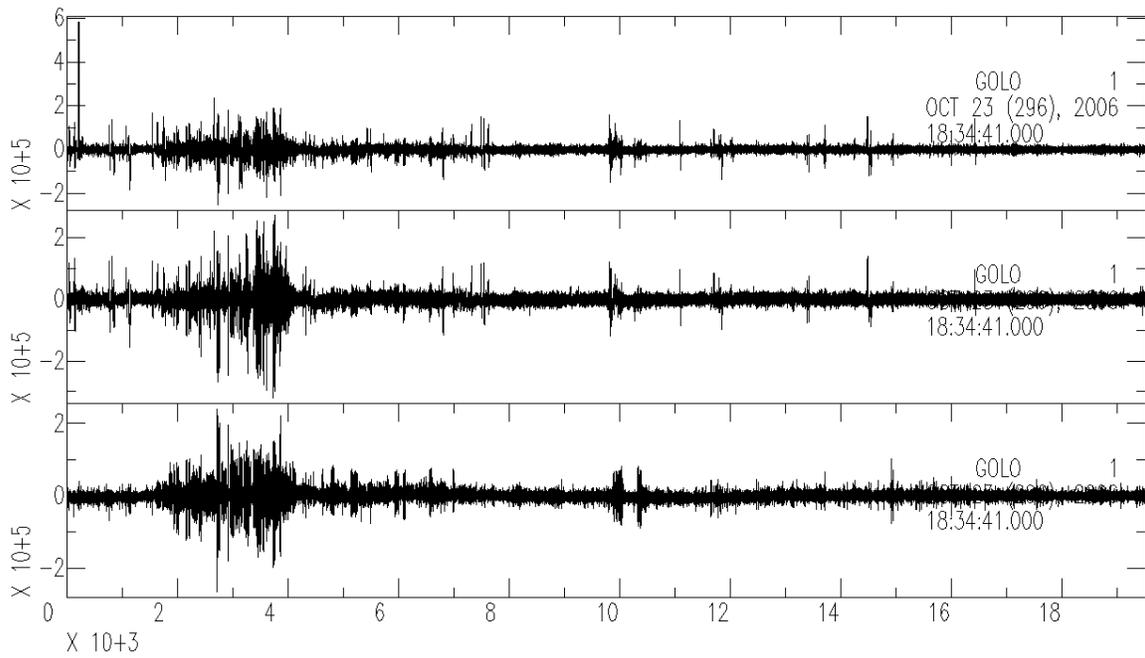


Figura 18: GULO 296, original.

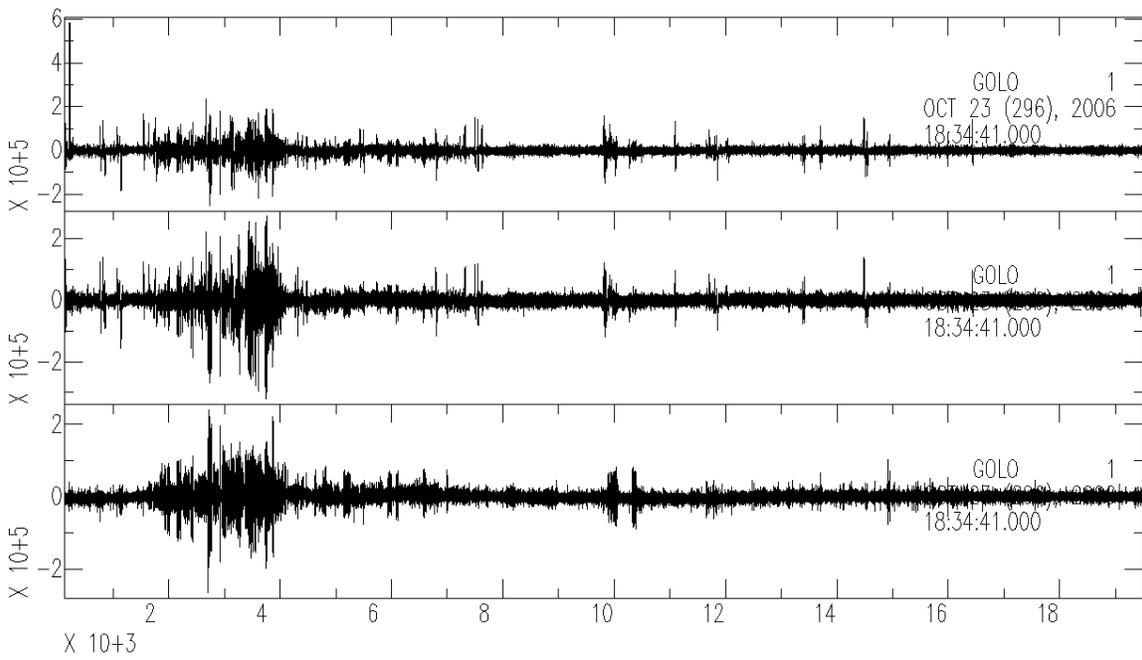


Figura 19: GULO 296, preprocesado.

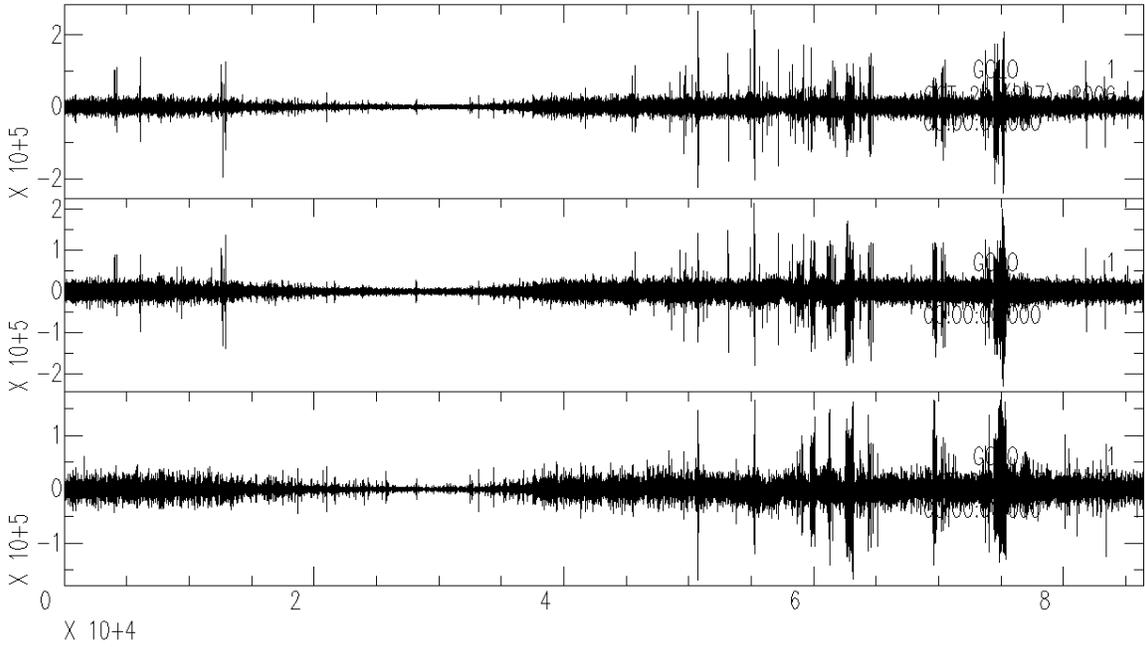


Figura 20: GOLO 297, original.

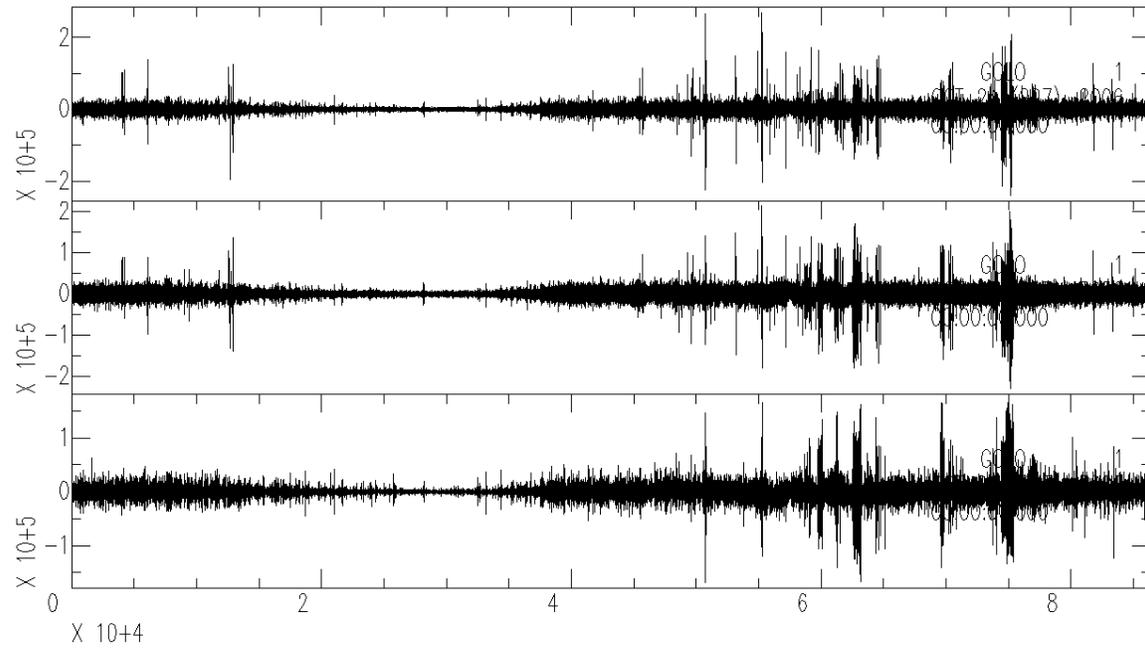


Figura 21: GOLO 297, preprocesado.

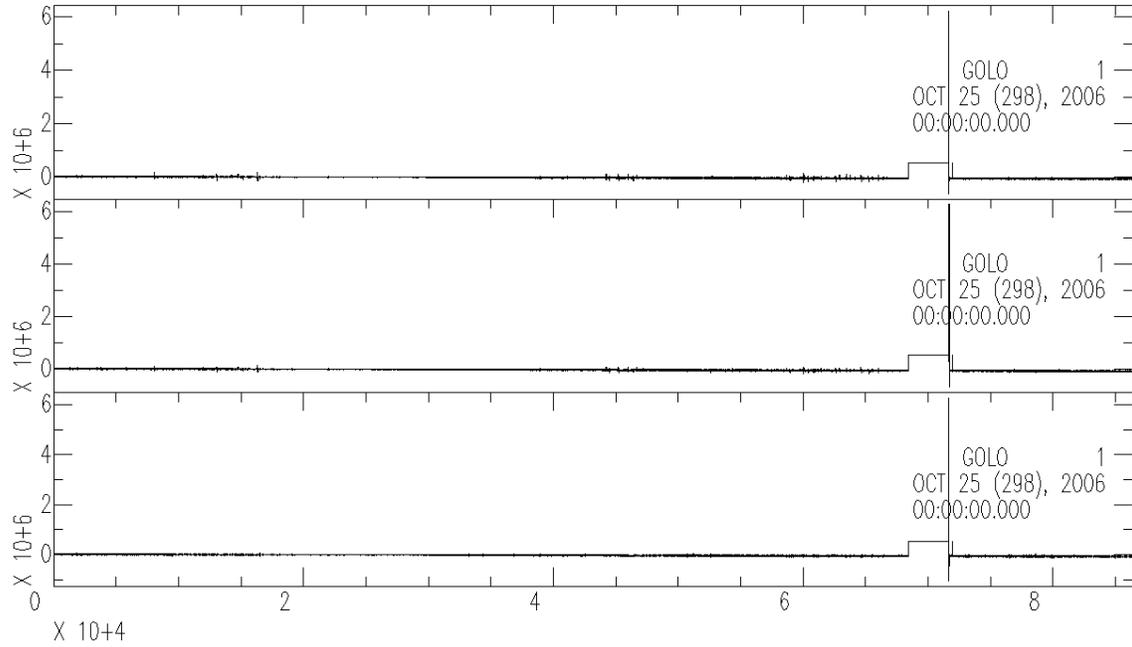


Figura 22: GOLO 298, original.

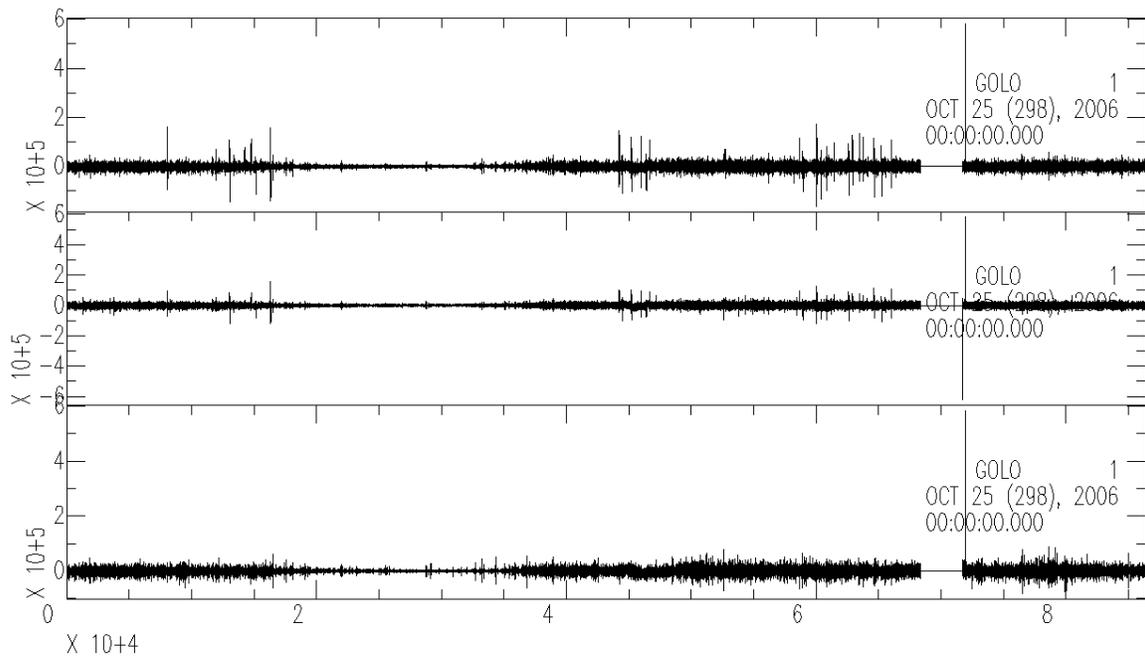


Figura 23: GOLO 298, preprocesado.

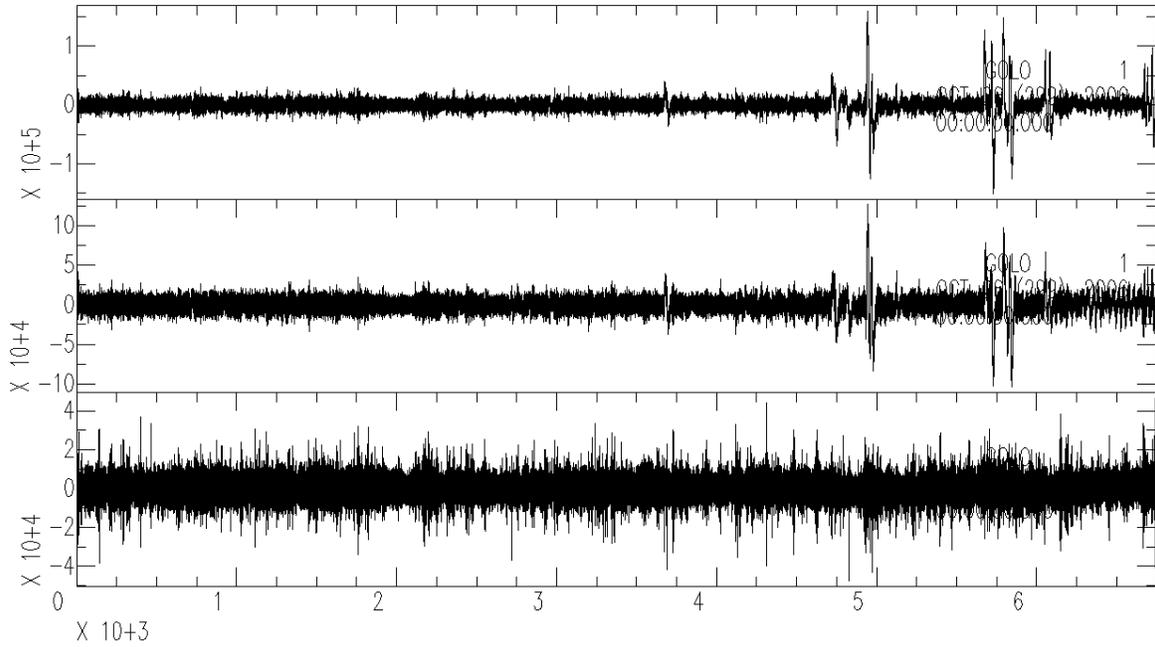


Figura 24: GOLO 299, original.

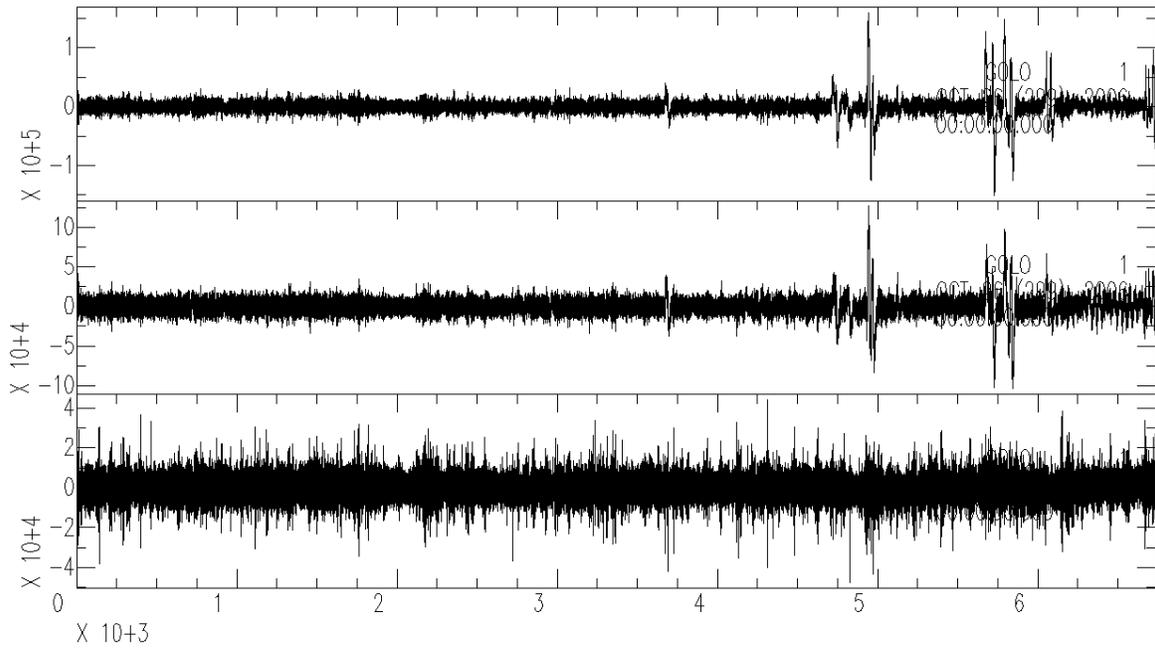


Figura 25: GOLO 299, preprocesado.

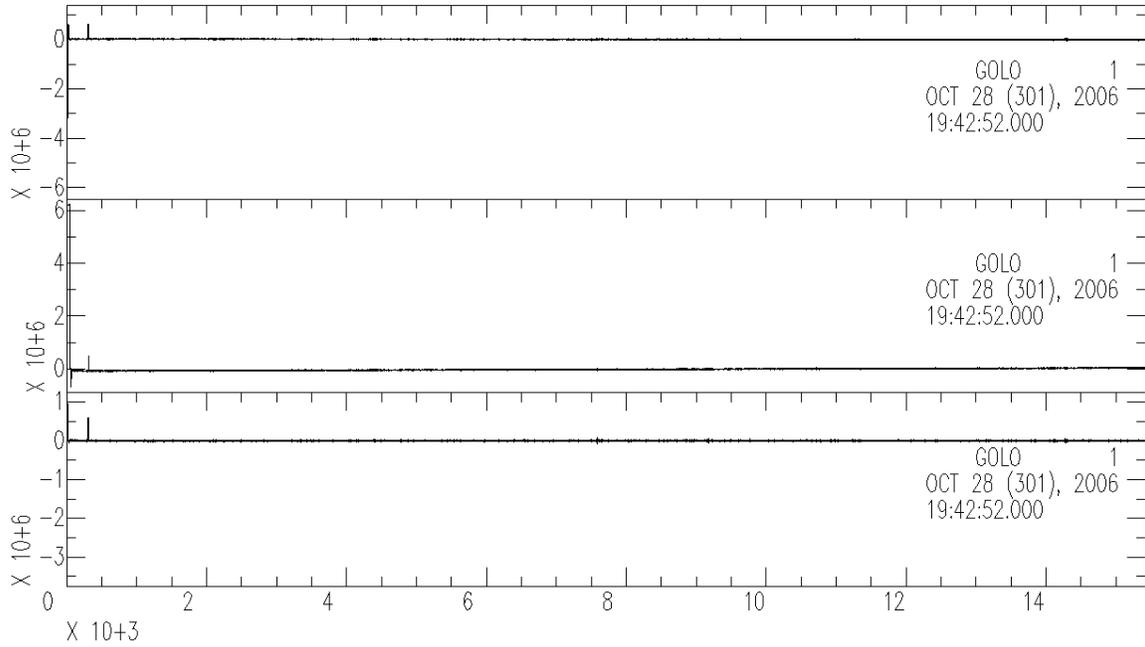


Figura 26: GOLO 301, original.

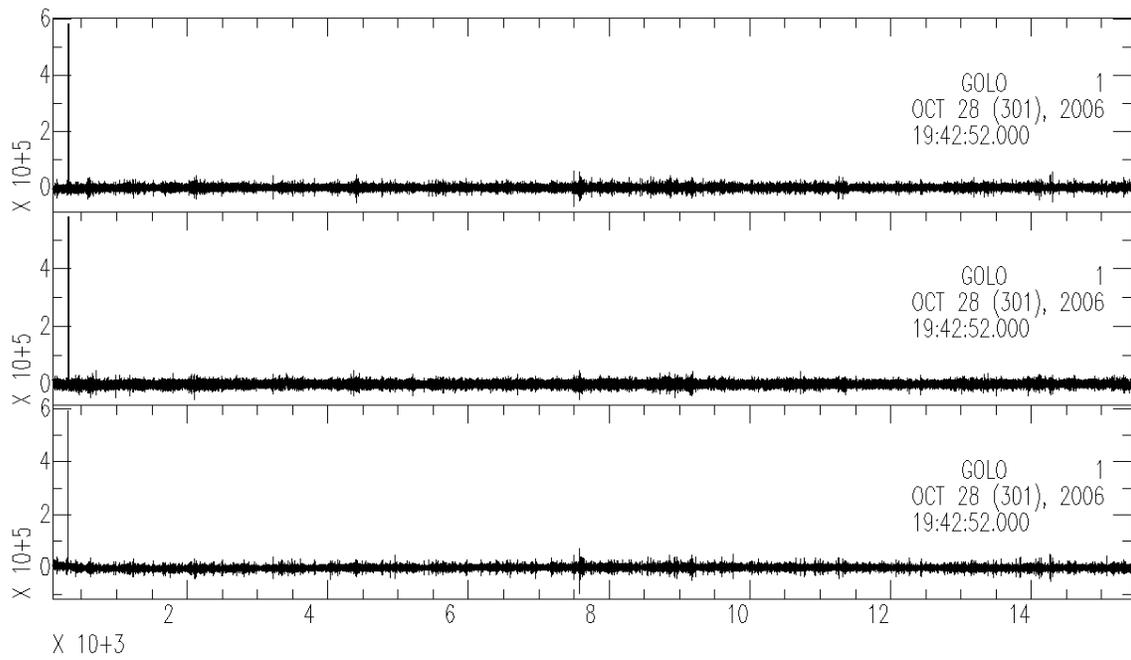


Figura 27: GOLO 301, preprocesado.

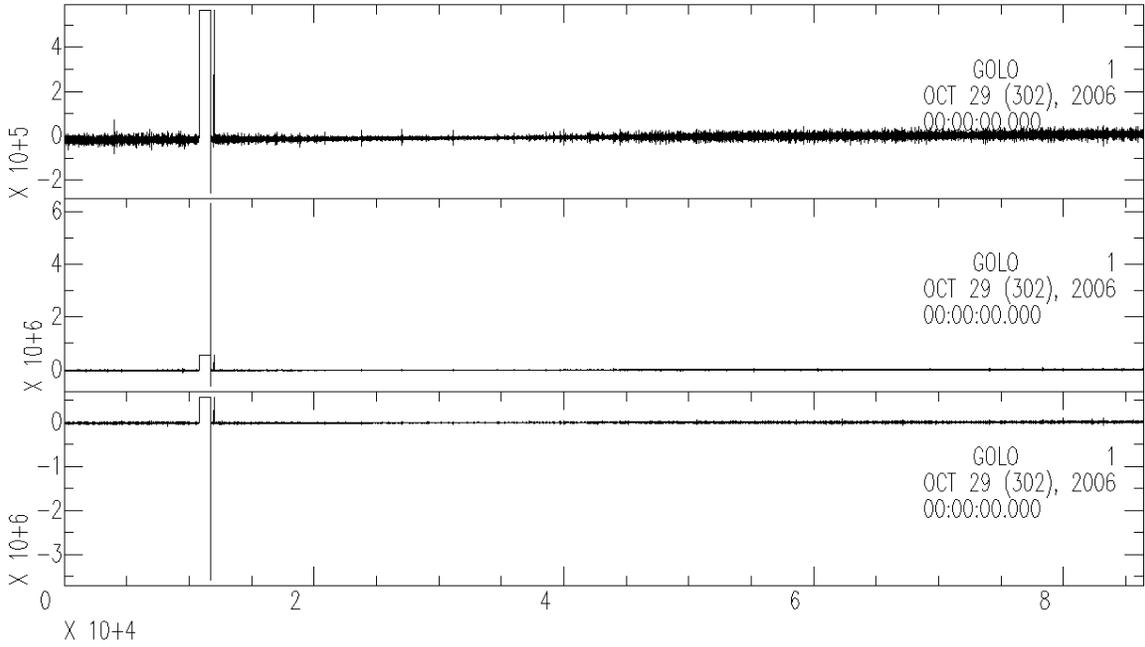


Figura 28: GOLO 302, original.

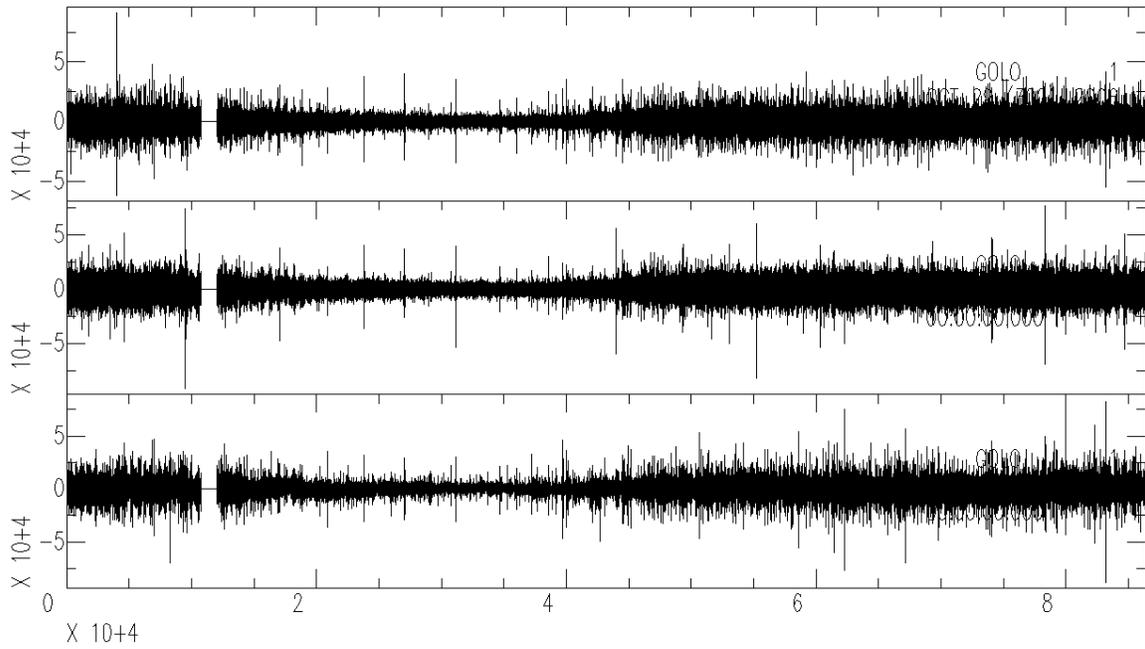


Figura 29: GOLO 302, preprocesado.

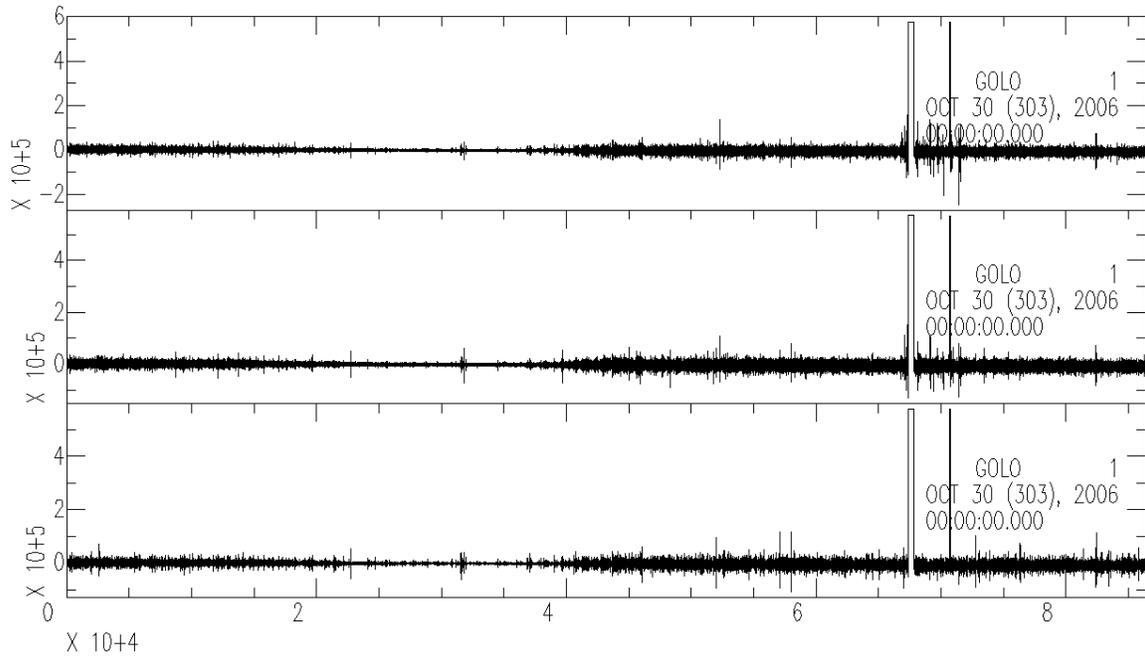


Figura 30: Golo 303, original.

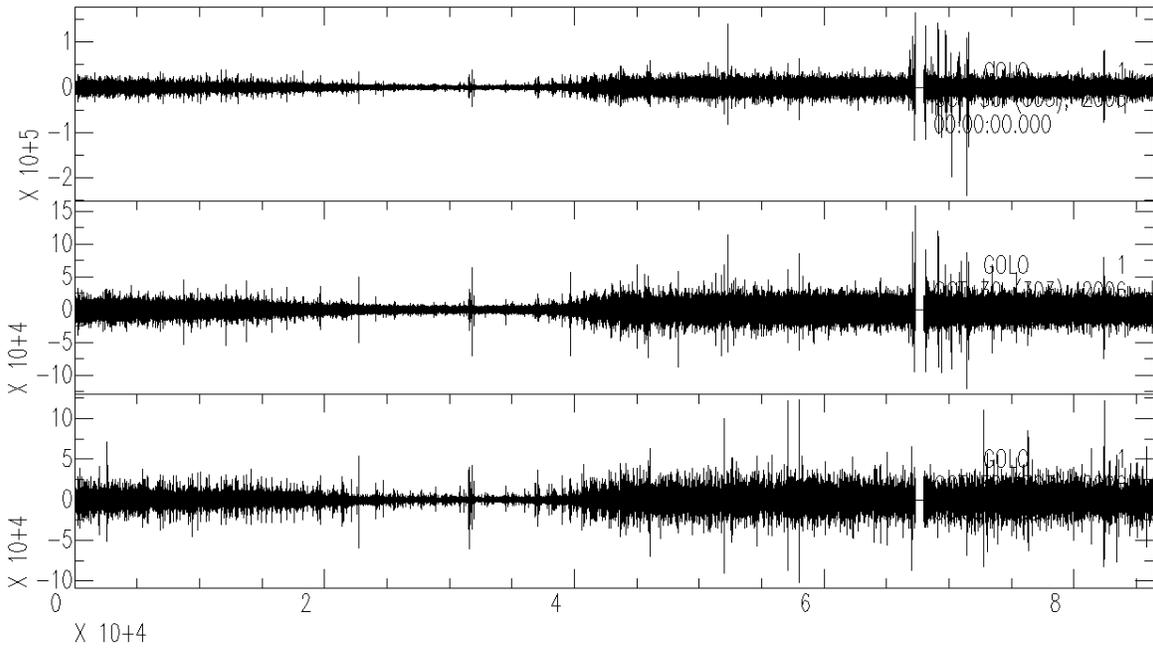


Figura 31: Golo 303, preprocesado.

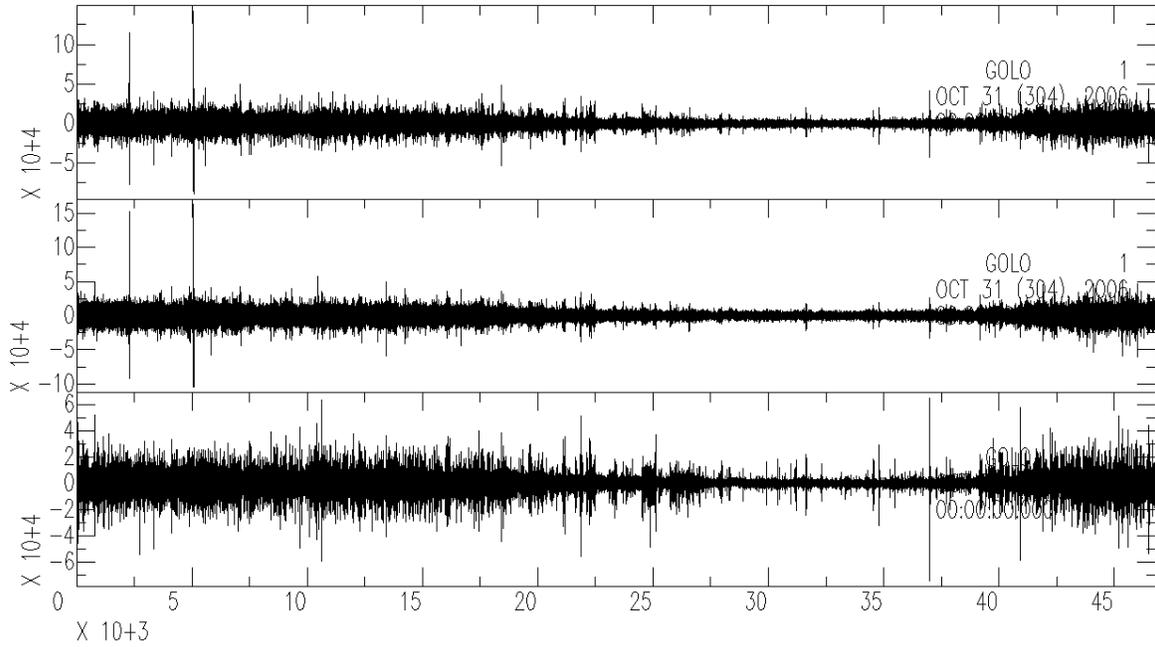


Figura 32: GOLO 304, original.

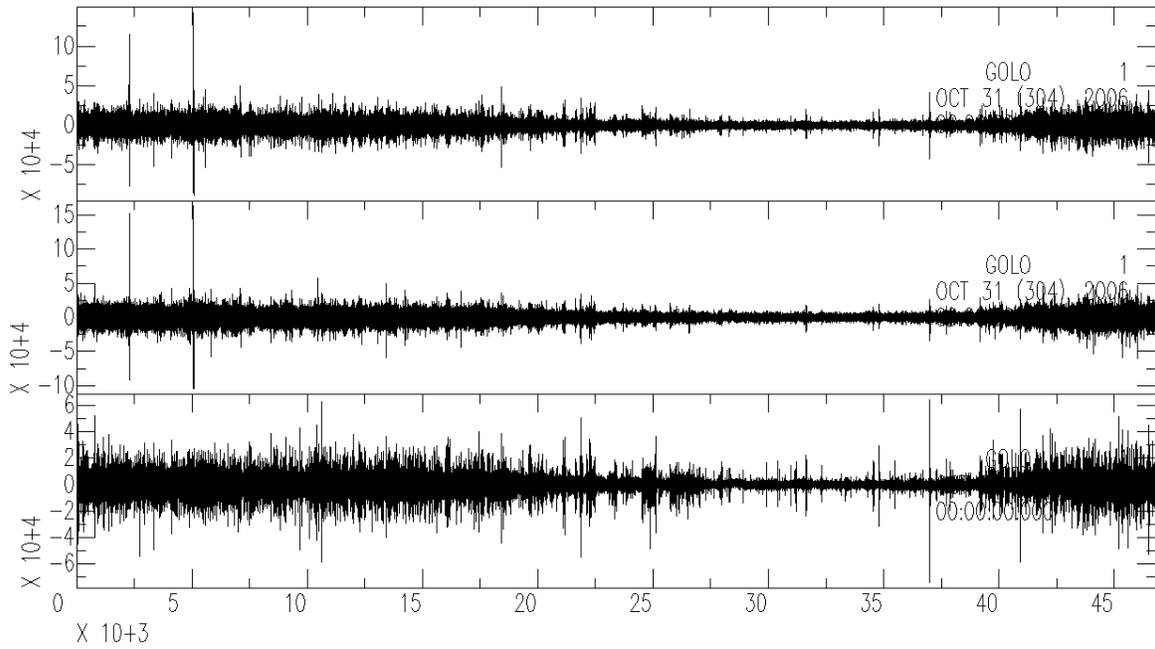


Figura 33: GOLO 304, preprocesado.

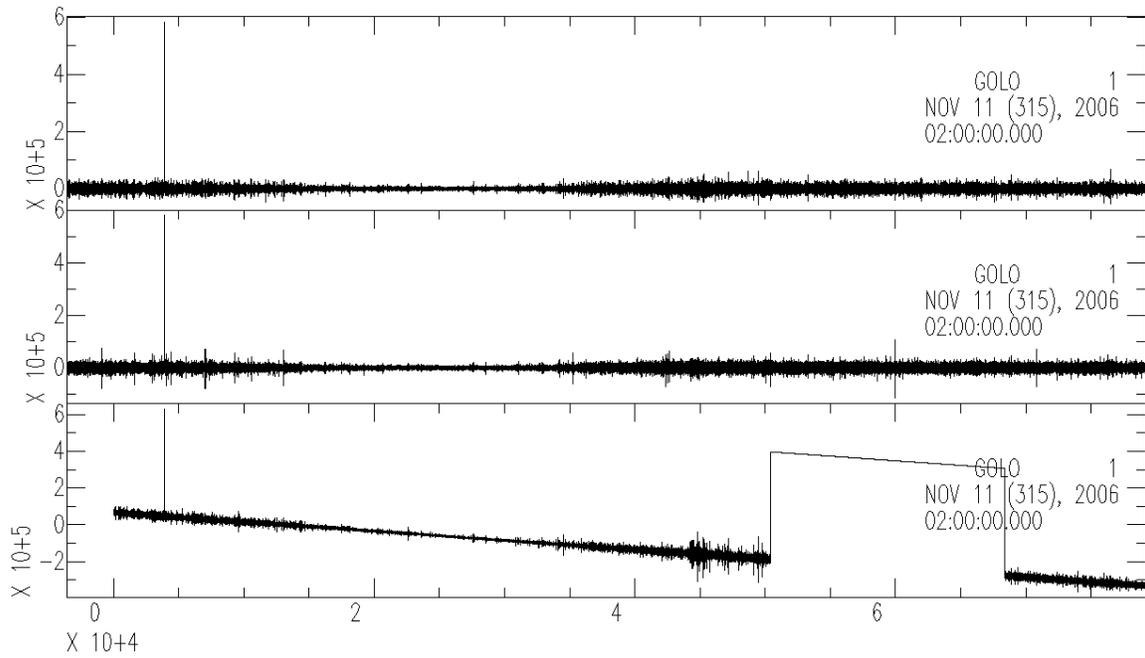


Figura 34: Golo 315, original.

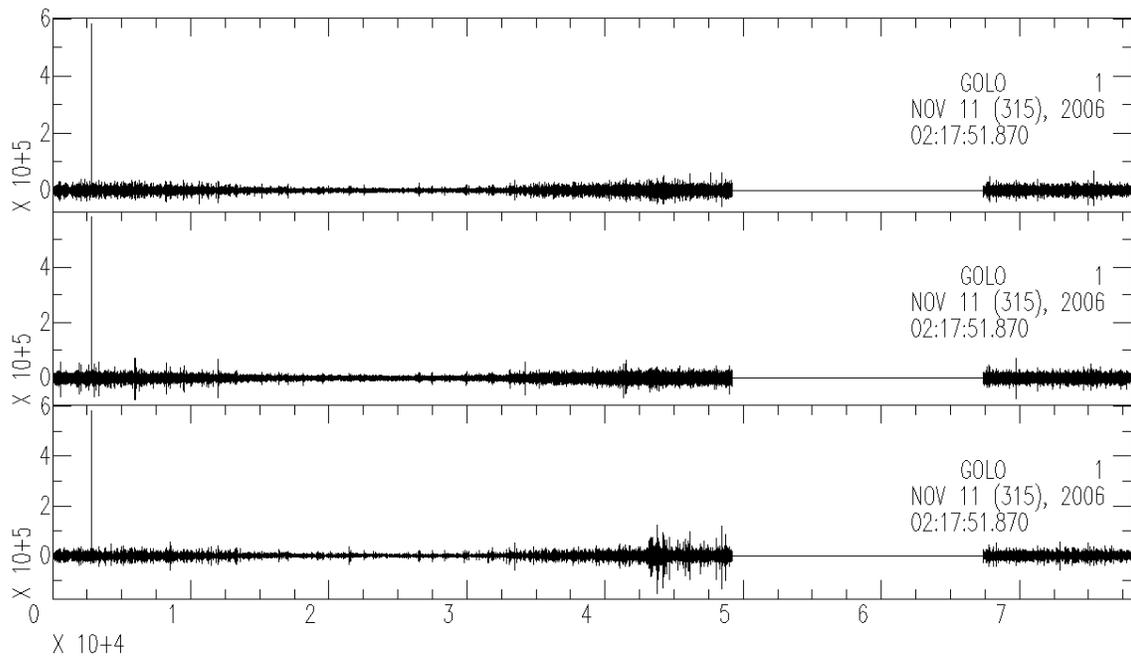


Figura 35: Golo 315, preprocesado.

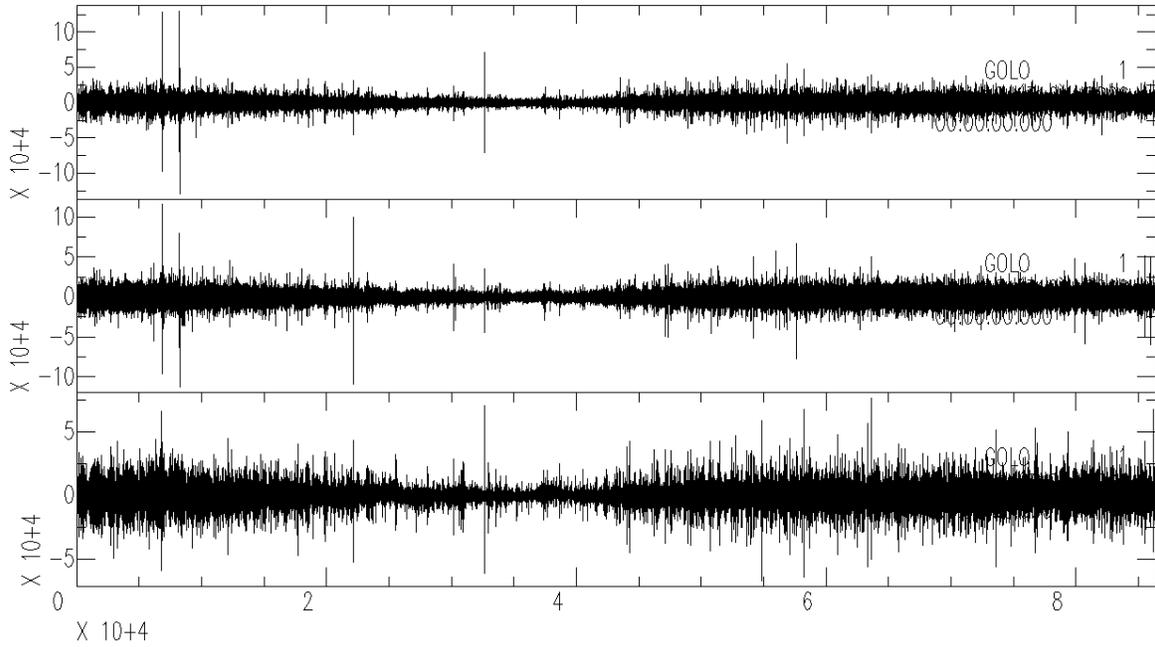


Figura 36: GOLA 316, original.

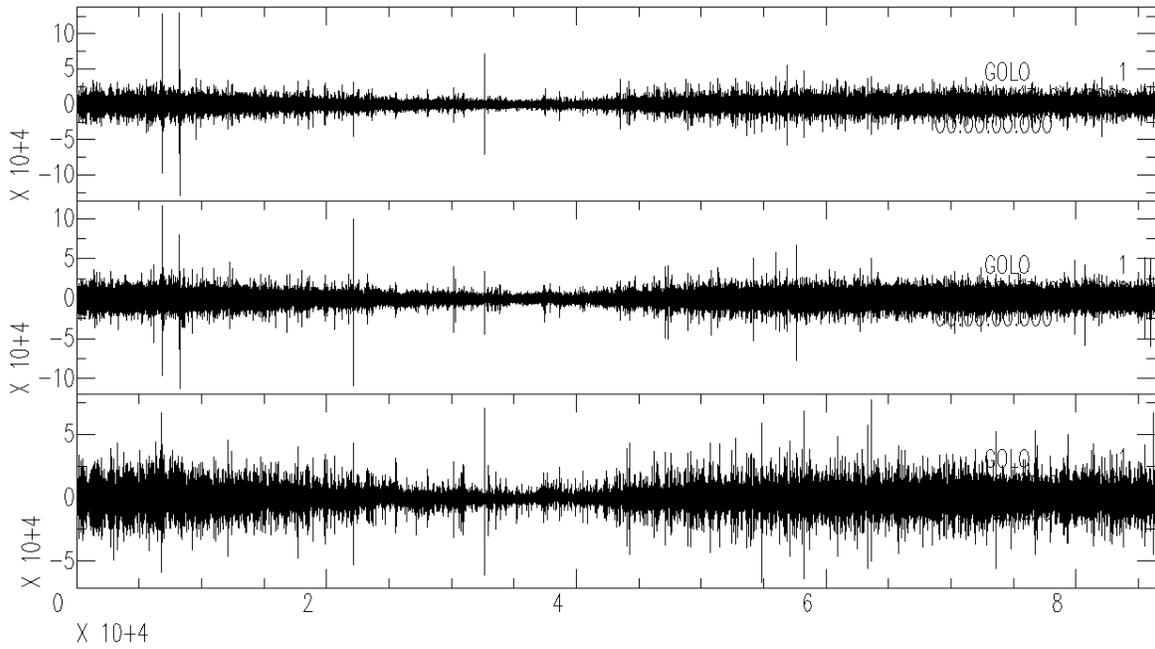


Figura 37: GOLA 316, preprocesado.

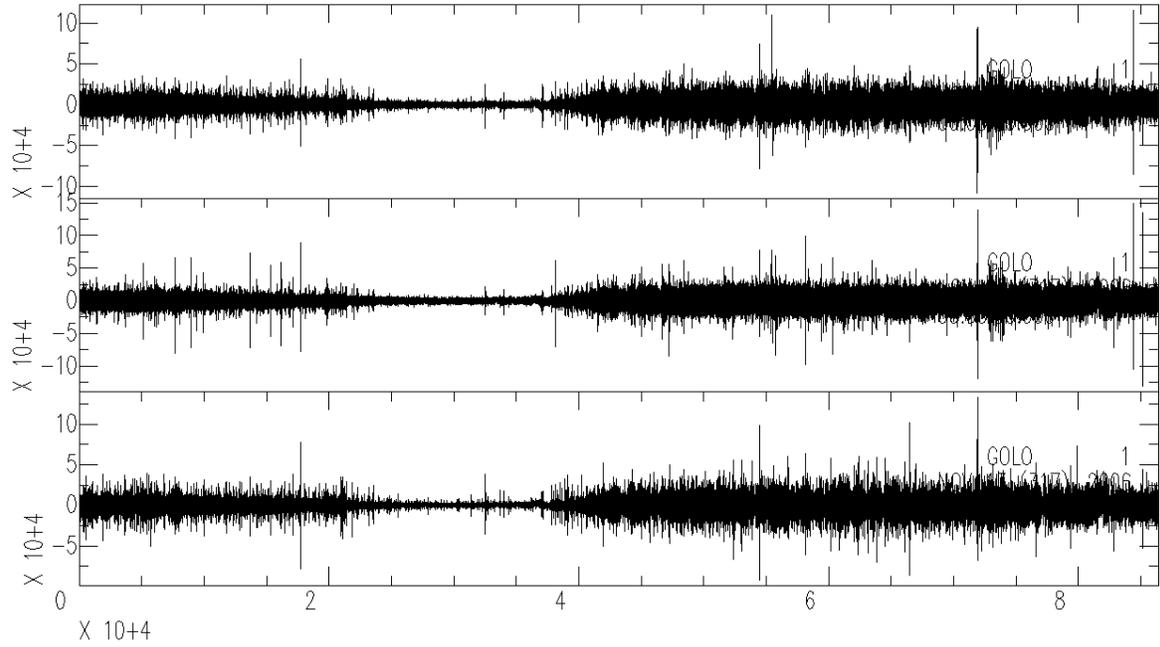


Figura 38: GOLO 317, original.

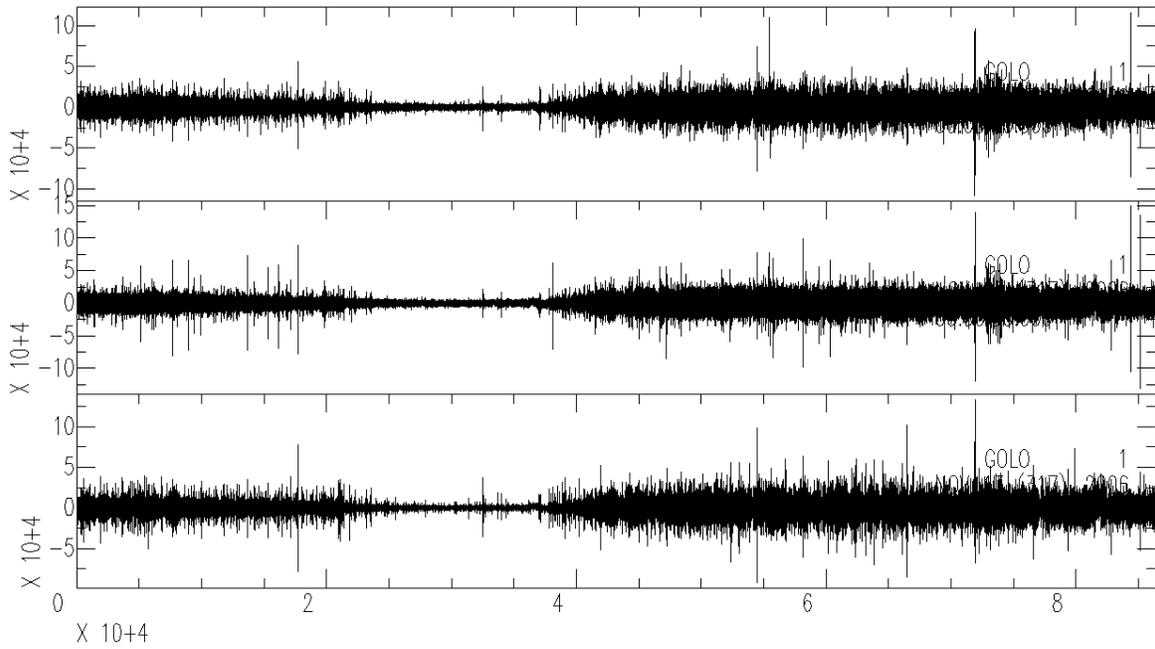


Figura 39: GOLO 317, preprocesado.

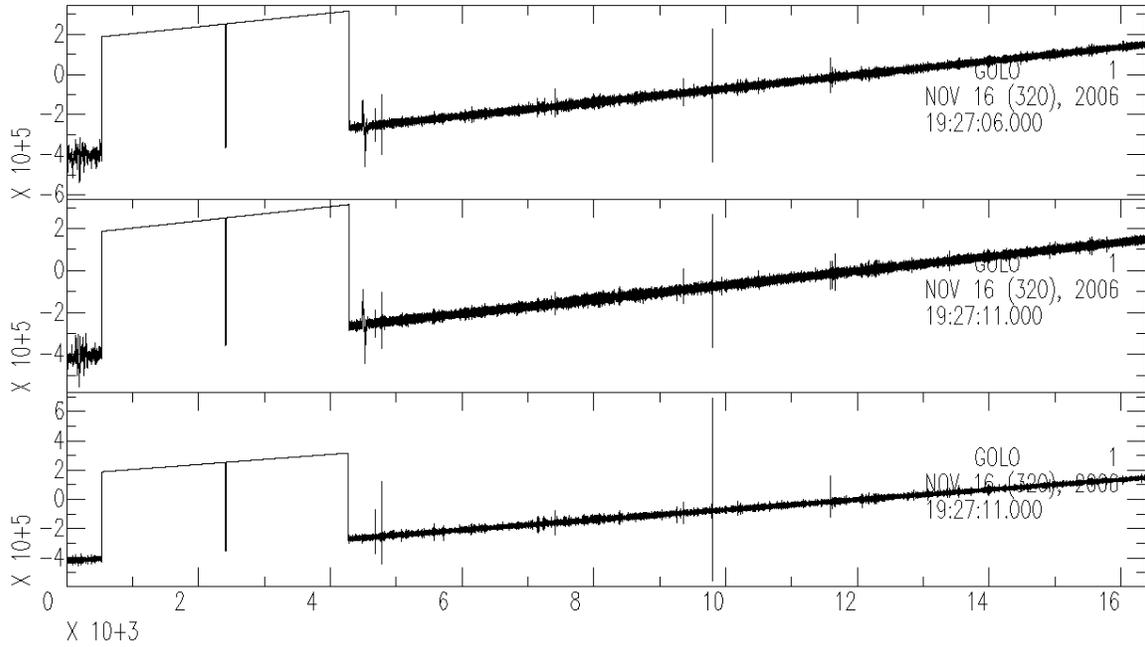


Figura 40: GOLA 320, original.

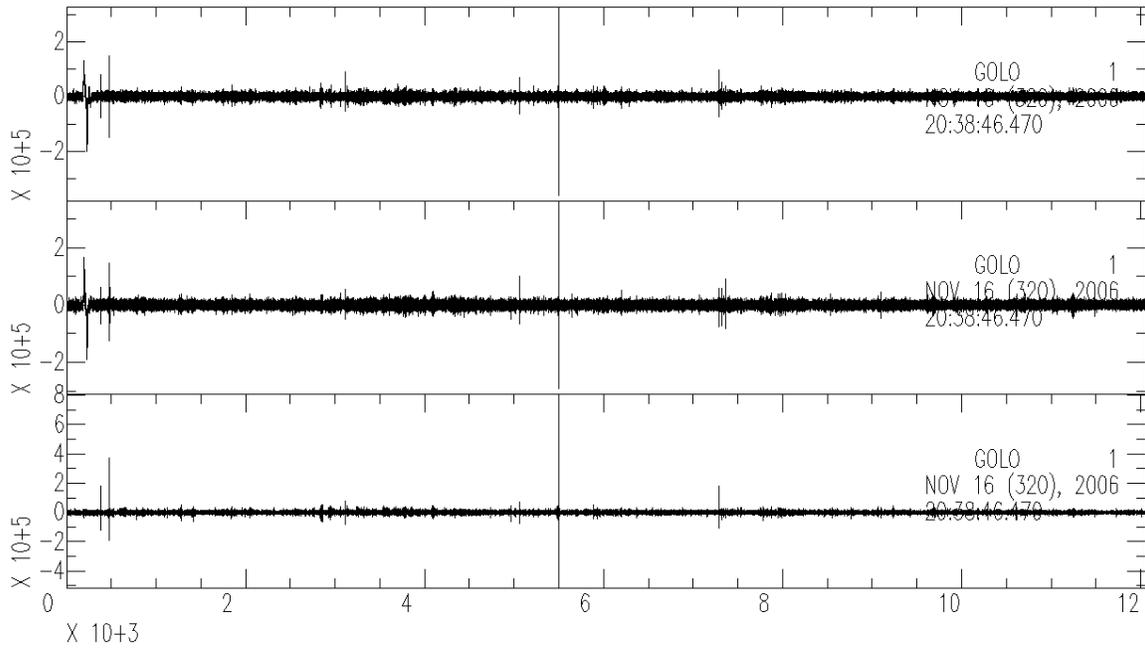


Figura 41: GOLA 320, preprocesado.

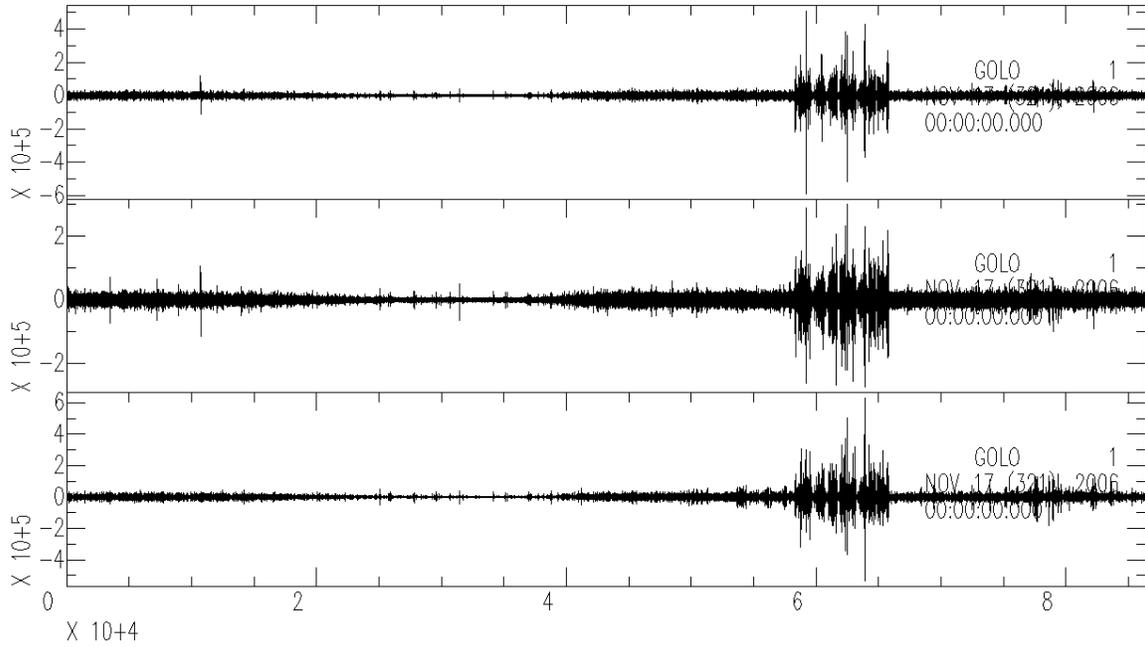


Figura 42: GOLA 321, original.

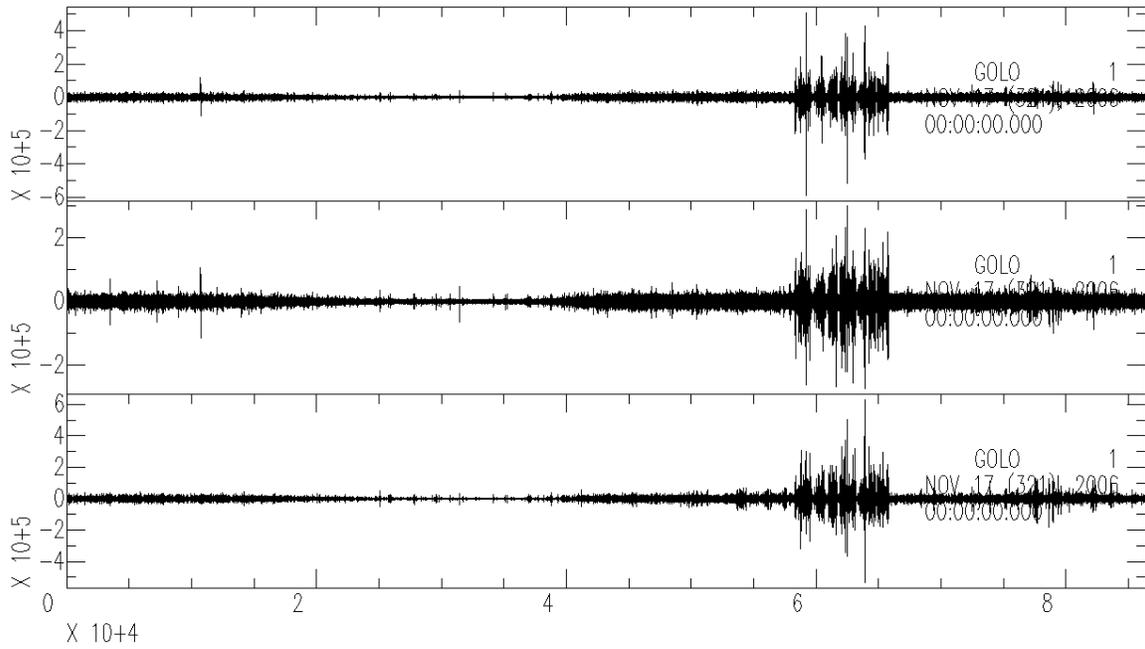


Figura 43: GOLA 321, preprocesado.

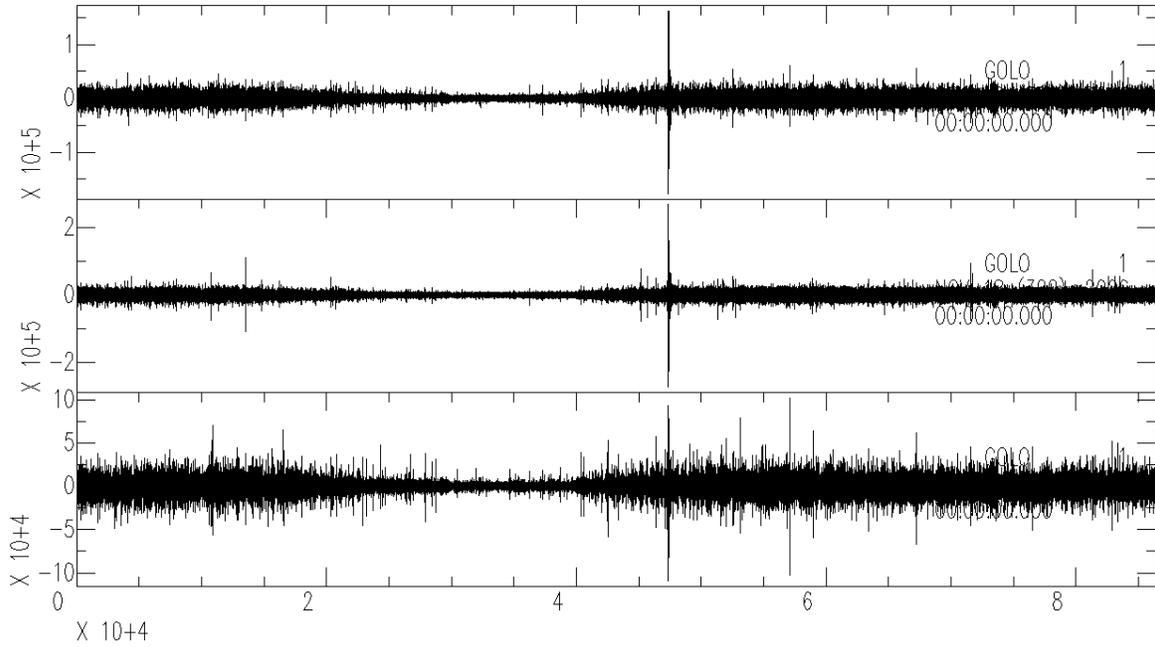


Figura 44: GOLA 322, original.

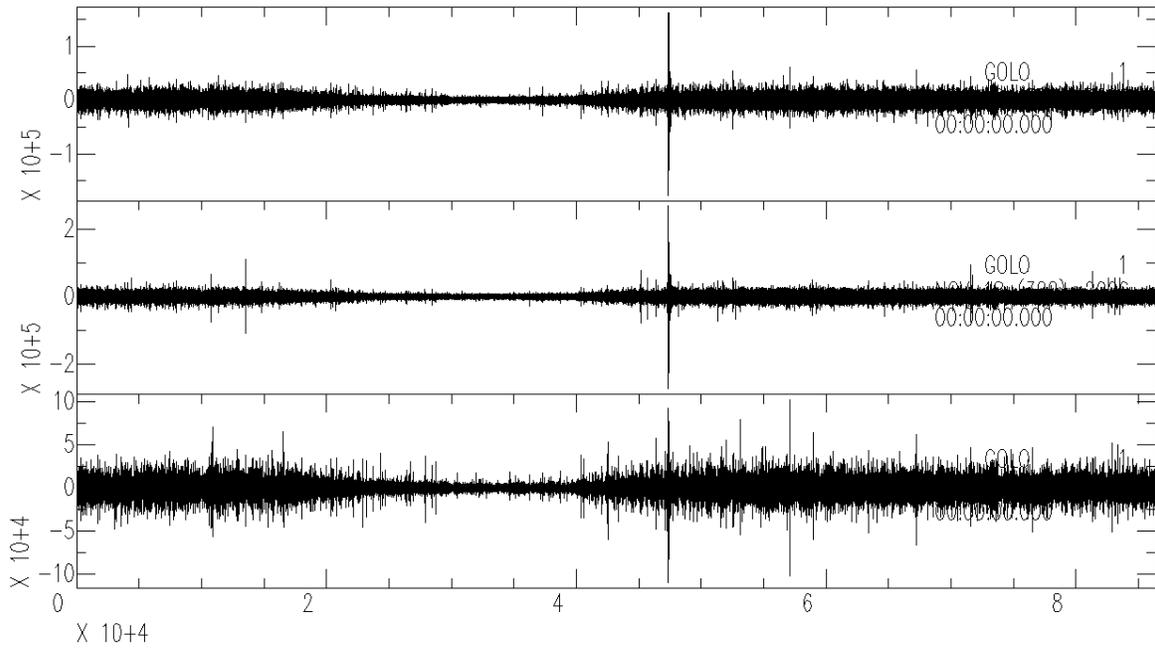


Figura 45: GOLA 322, preprocesado.

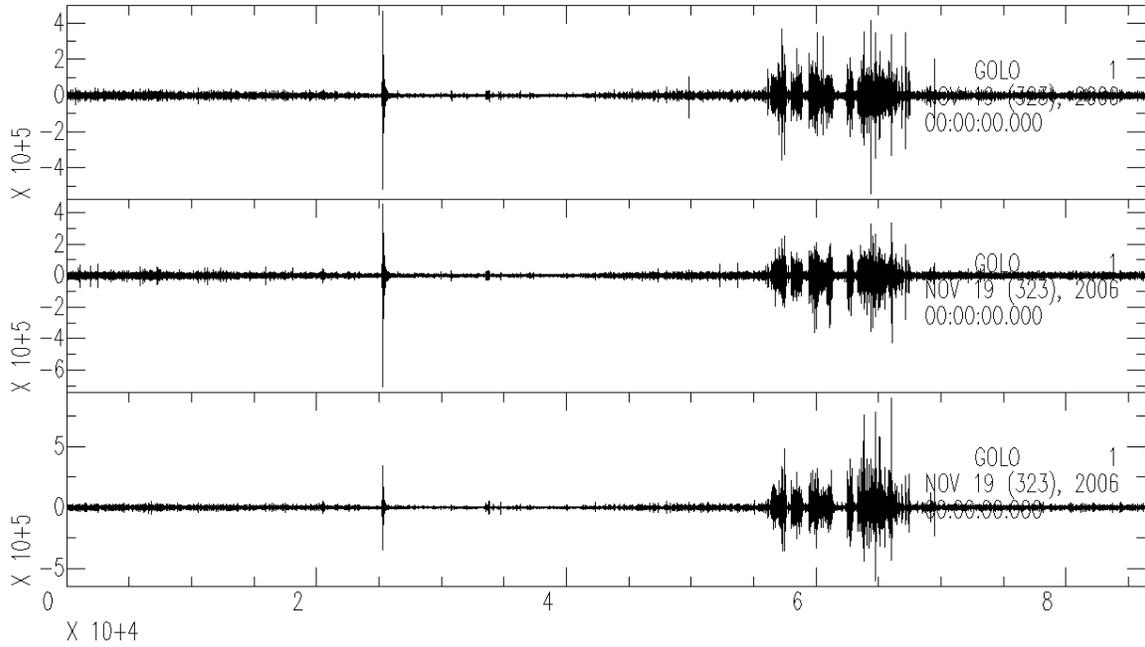


Figura 46: GOLA 323, original.

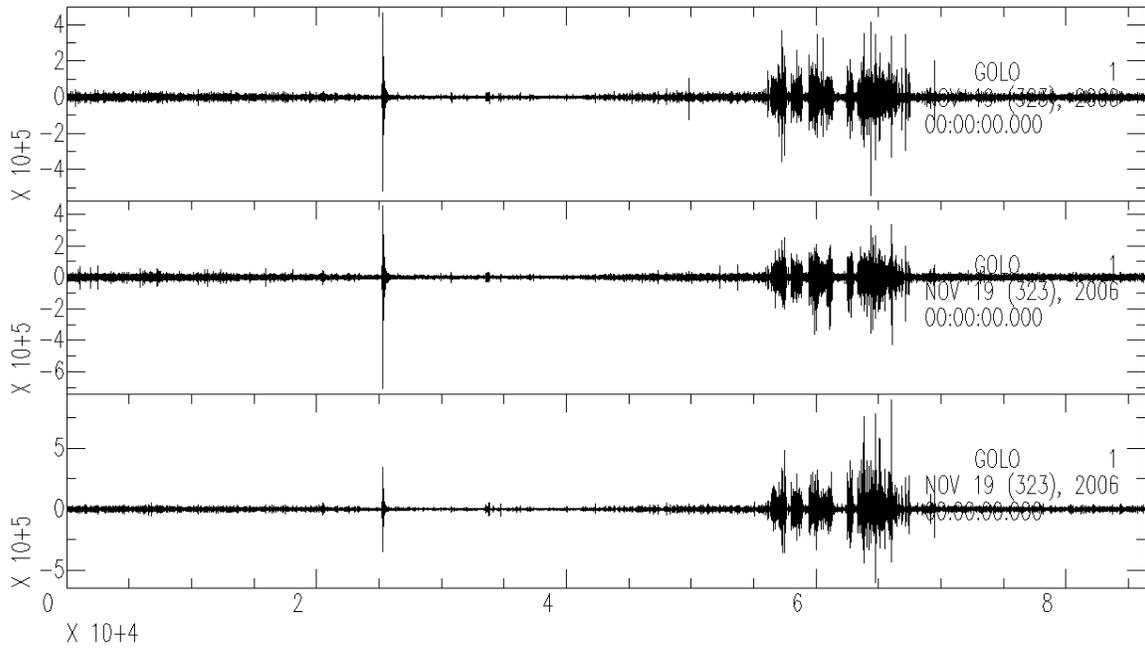


Figura 47: GOLA 323, preprocesado.

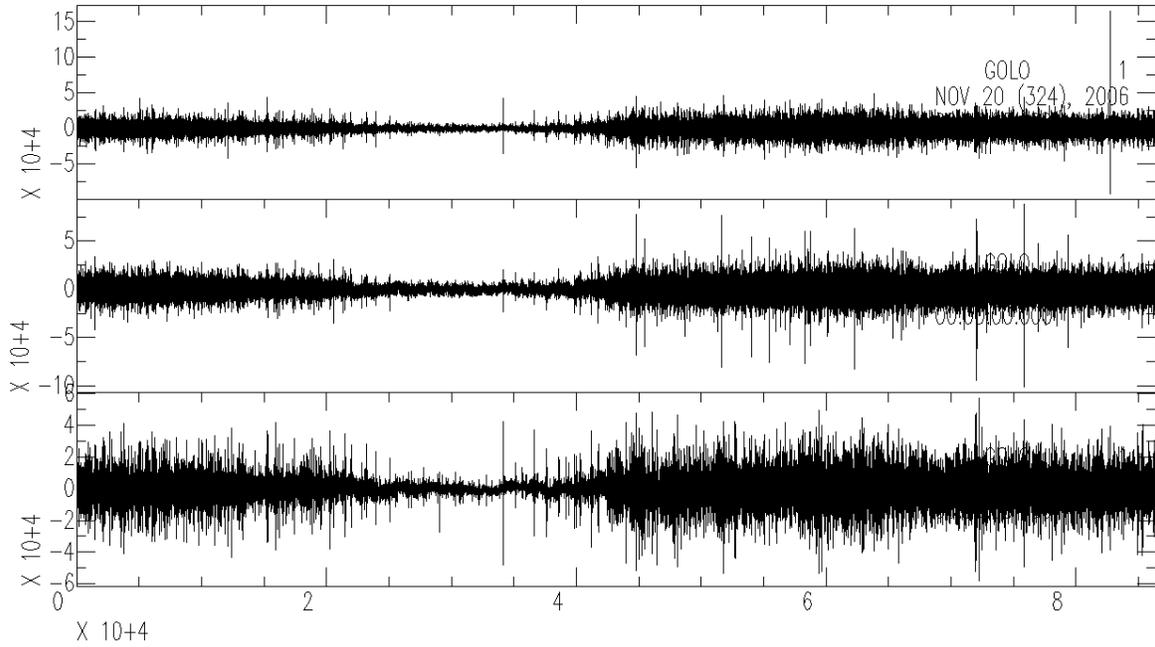


Figura 48: GOLA 324, original.

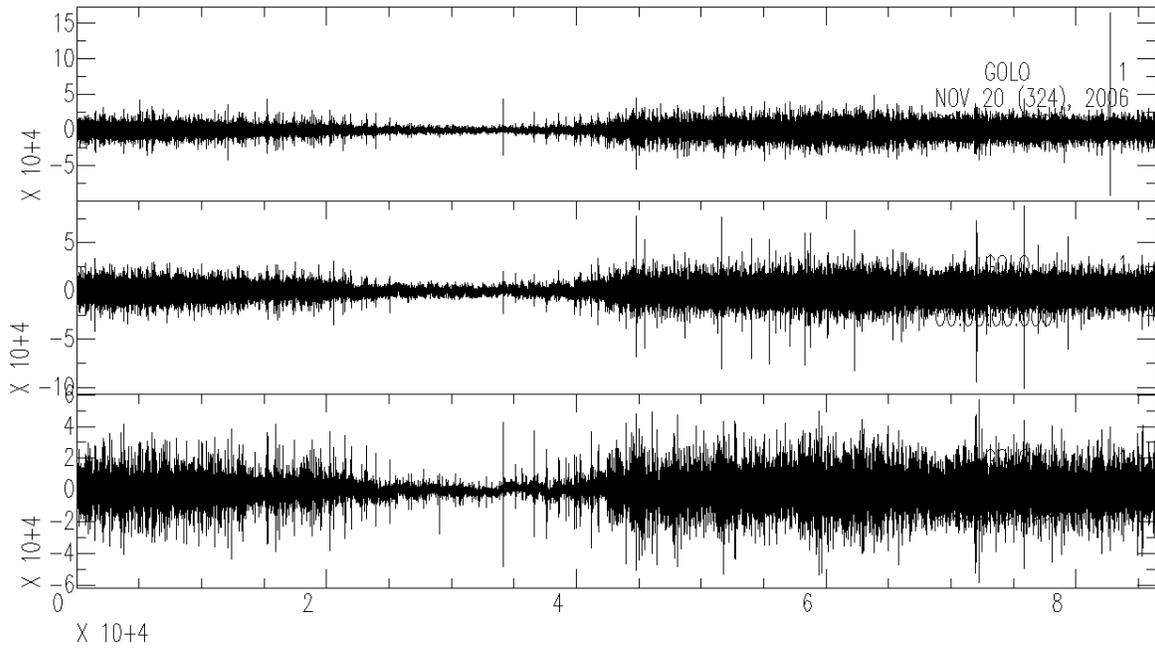


Figura 49: GOLA 324, preprocesado.

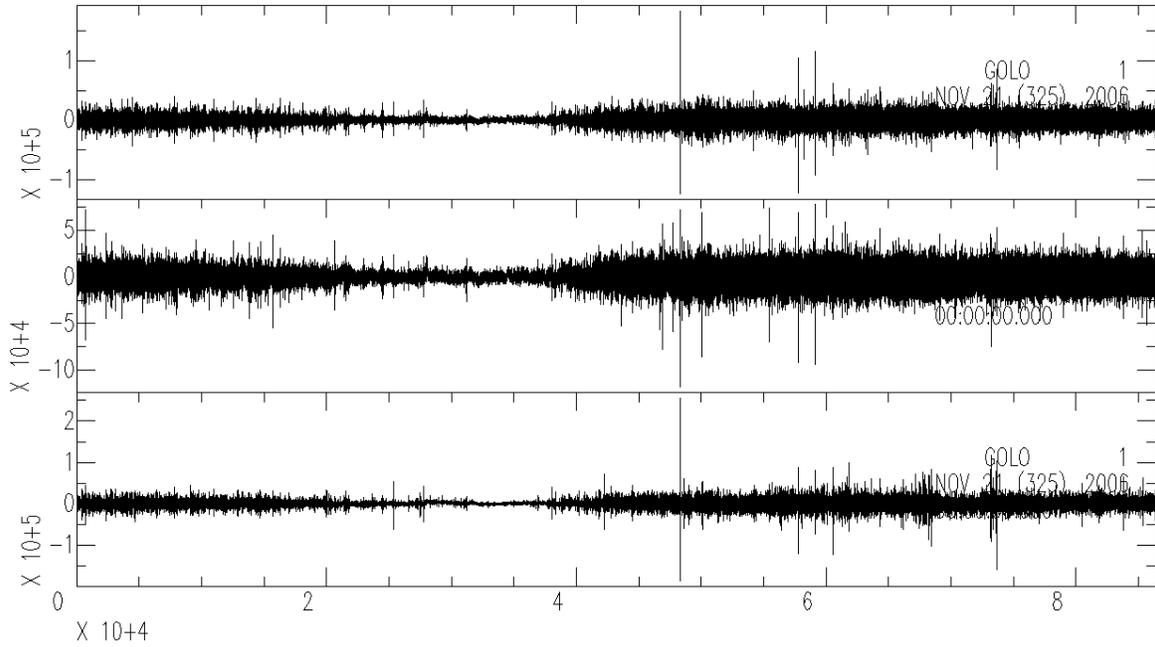


Figura 50: GOLO 325, original.

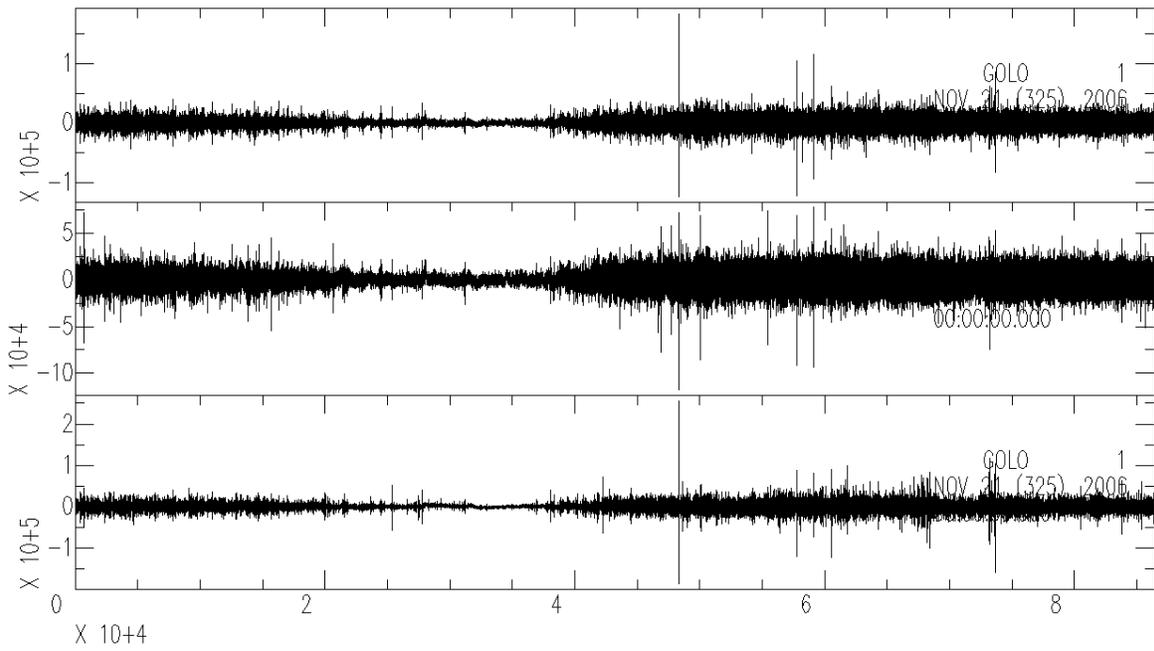


Figura 51: GOLO 325, preprocesado.

10.7. Sismogramas clasificados

Como complemento a la sección de Resultados, en las figuras de a continuación se muestran las ventanas marcadas como *evento*, a partir de un *score de corte* para dicha clase de 0.75017. Las ventanas marcadas se muestran en color rojo, y se han sobrepuesto a los sismogramas preprocesados, sin filtrar. En la Tabla 13, tras este párrafo, se hace referencia a todas las figuras, para ubicarlas fácilmente. Los detalles de cada uno de los registros preprocesados se incluyen en la Tabla 12, de la sección de Sismogramas por trabajar. El análisis general de estas figuras se expone en la sección de Figuras de los registros clasificados, en los Resultados.

Tabla 13: Figuras de los sismogramas preprocesados, con las ventanas de un score para la clase evento mayor o igual a 0.75017 resaltadas de color rojo.

Registro preprocesado	Figura	Registro preprocesado	Figura
GOLO 287	Figura 52	GOLO 317-1	Figura 80
GOLO 288	Figura 53	GOLO 317-2	Figura 81
GOLO 296	Figura 54	GOLO 317-3	Figura 82
GOLO 297-1	Figura 55	GOLO 317-4	Figura 83
GOLO 297-2	Figura 56	GOLO 320	Figura 84
GOLO 298-1	Figura 57	GOLO 321-1	Figura 85
GOLO 298-2	Figura 58	GOLO 321-2	Figura 86
GOLO 298-3	Figura 59	GOLO 321-3	Figura 87
GOLO 299	Figura 60	GOLO 321-4	Figura 88
GOLO 301	Figura 61	GOLO 322-1	Figura 89
GOLO 302-1	Figura 62	GOLO 322-2	Figura 90
GOLO 302-2	Figura 63	GOLO 322-3	Figura 91
GOLO 302-3	Figura 64	GOLO 322-4	Figura 92
GOLO 303-1	Figura 65	GOLO 323-1	Figura 93
GOLO 303-2	Figura 66	GOLO 323-2	Figura 94
GOLO 303-3	Figura 67	GOLO 323-3	Figura 95
GOLO 303-4	Figura 68	GOLO 323-4	Figura 96
GOLO 304-1	Figura 69	GOLO 324-1	Figura 97
GOLO 304-2	Figura 70	GOLO 324-2	Figura 98
GOLO 304-3	Figura 71	GOLO 324-3	Figura 99
GOLO 304-4	Figura 72	GOLO 324-4	Figura 100
GOLO 315-1	Figura 73	GOLO 325-1	Figura 101
GOLO 315-2	Figura 74	GOLO 325-2	Figura 102
GOLO 315-3	Figura 75	GOLO 325-3	Figura 103
GOLO 316-1	Figura 76	GOLO 325-4	Figura 104
GOLO 316-2	Figura 77		
GOLO 316-3	Figura 78		
GOLO 316-4	Figura 79		

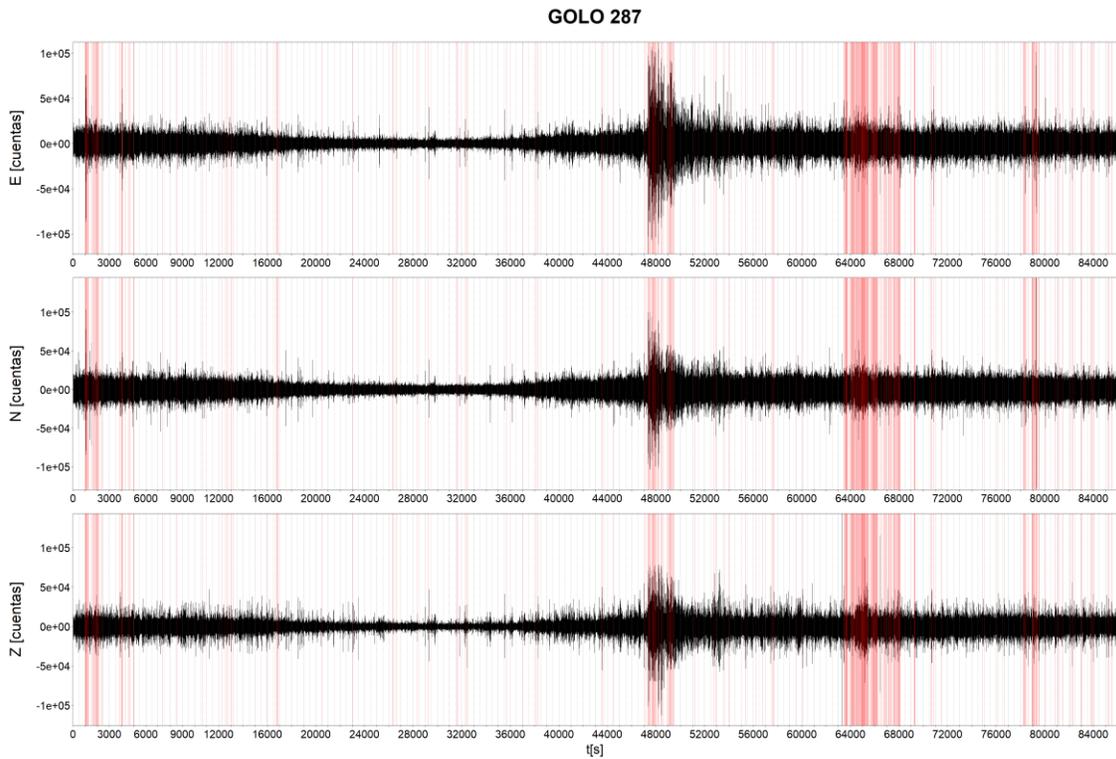


Figura 52: Sismograma de GOLO 287. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

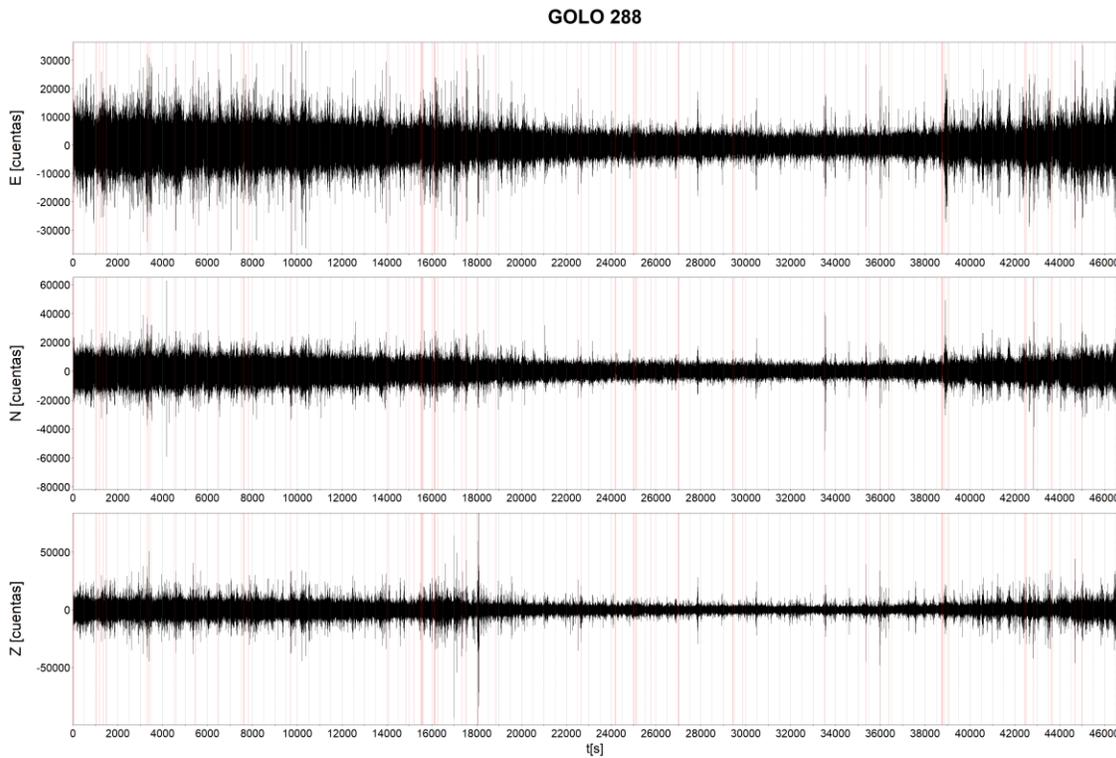


Figura 53: Sismograma de GOLO 288. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

GOLO 296

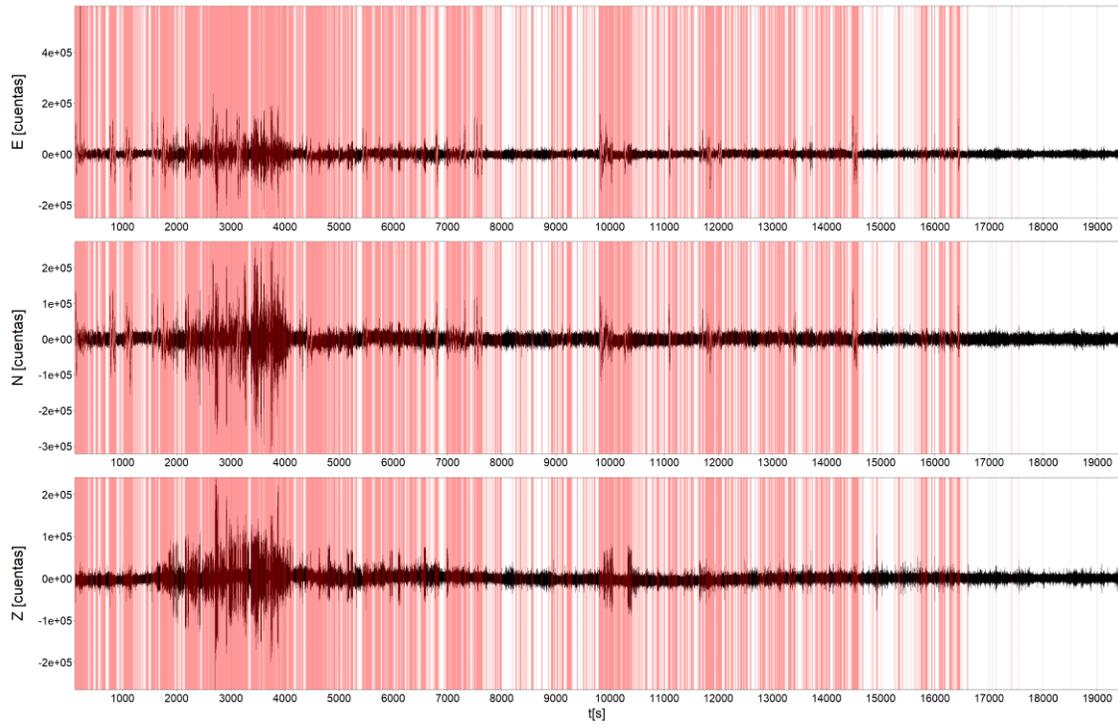


Figura 54: Sismograma de GOLO 296. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

GOLO 297-1

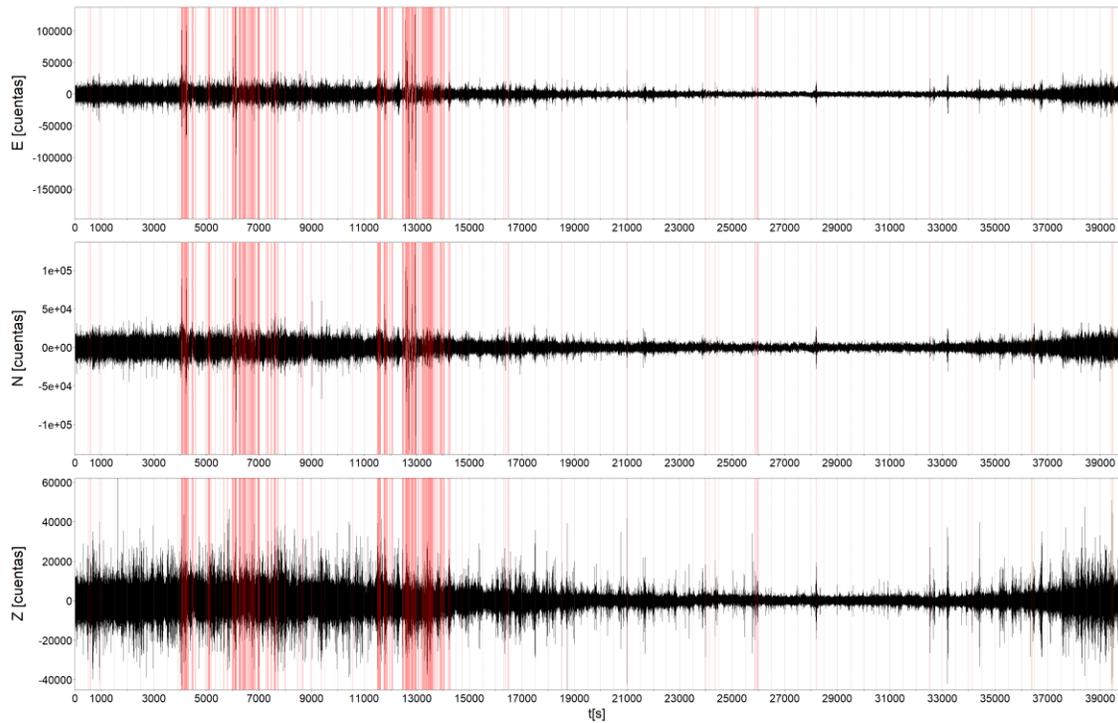


Figura 55: Sismograma de GOLO 297-1. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

GOLO 297-2

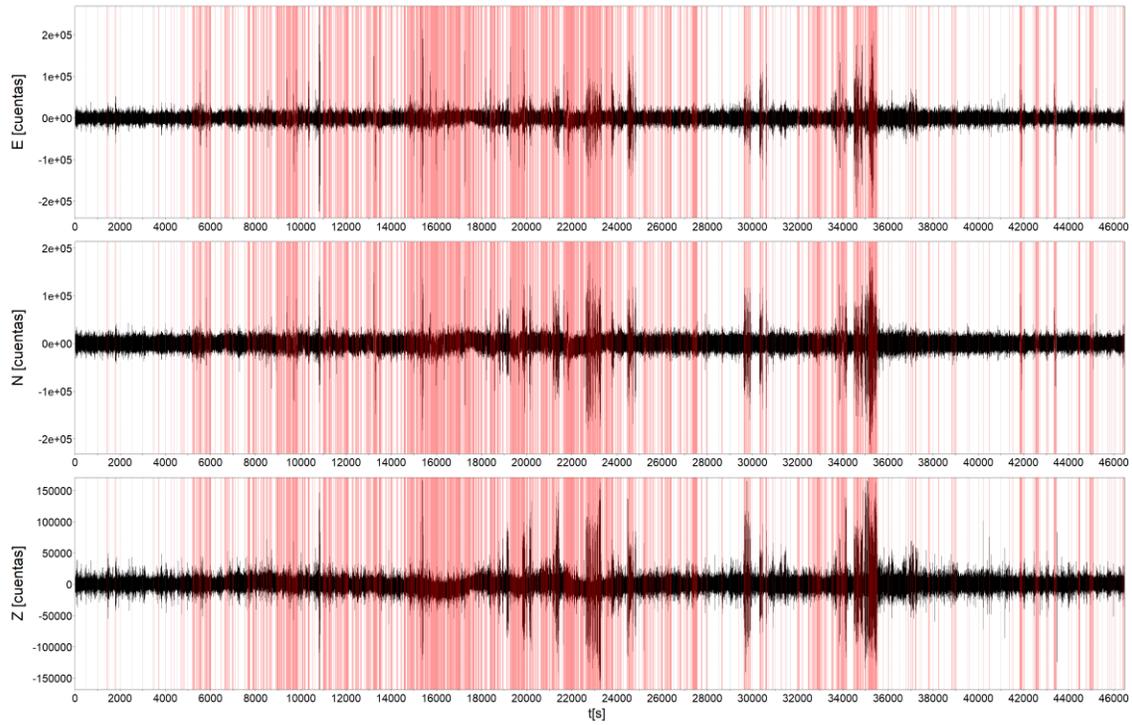


Figura 56: Sismograma de GOLO 297-2. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

GOLO 298-1

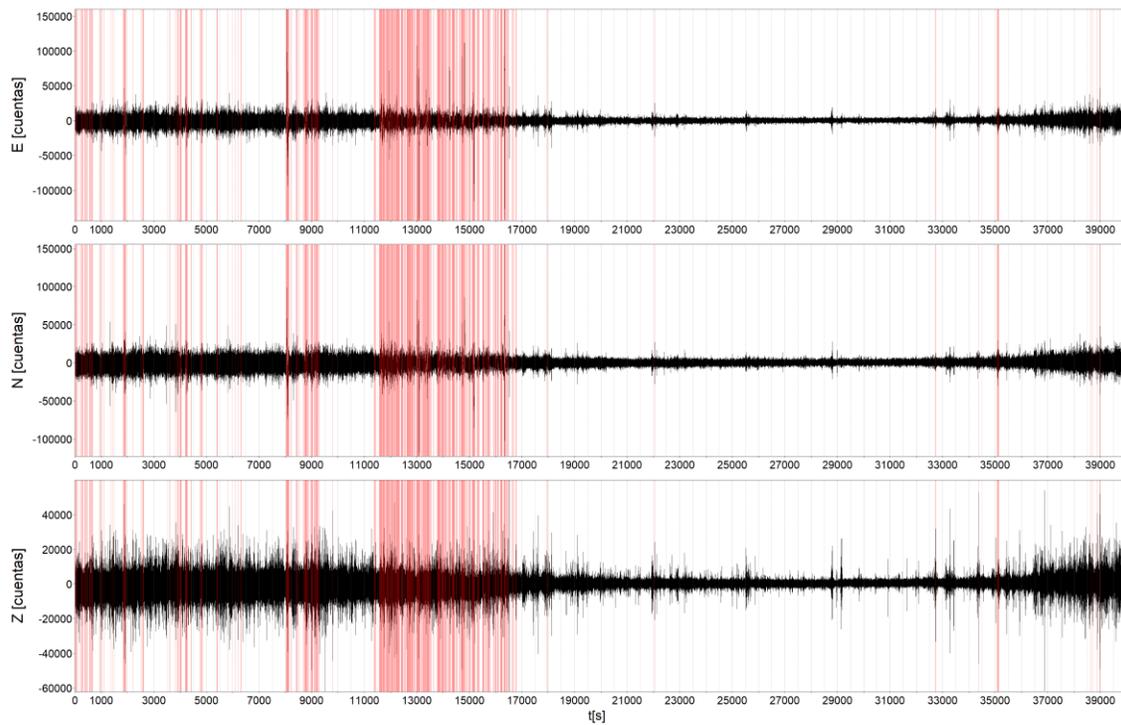


Figura 57: Sismograma de GOLO 298-1. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

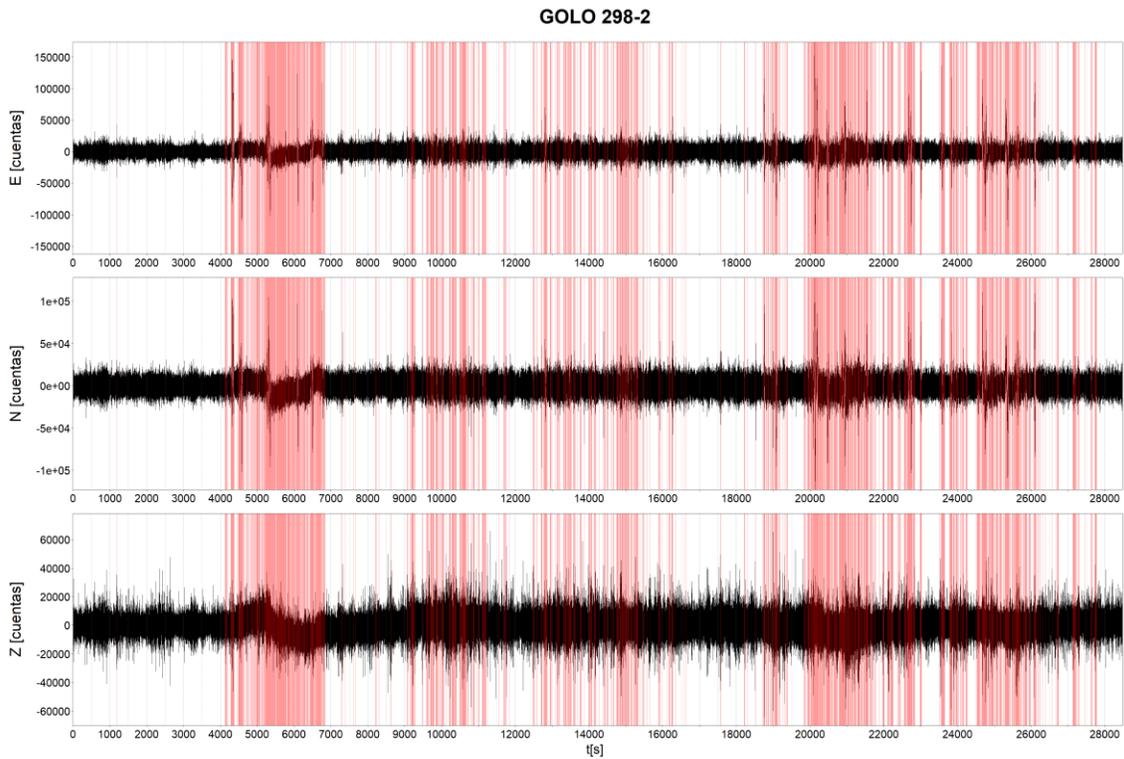


Figura 58: Sismograma de GOLO 298-2. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

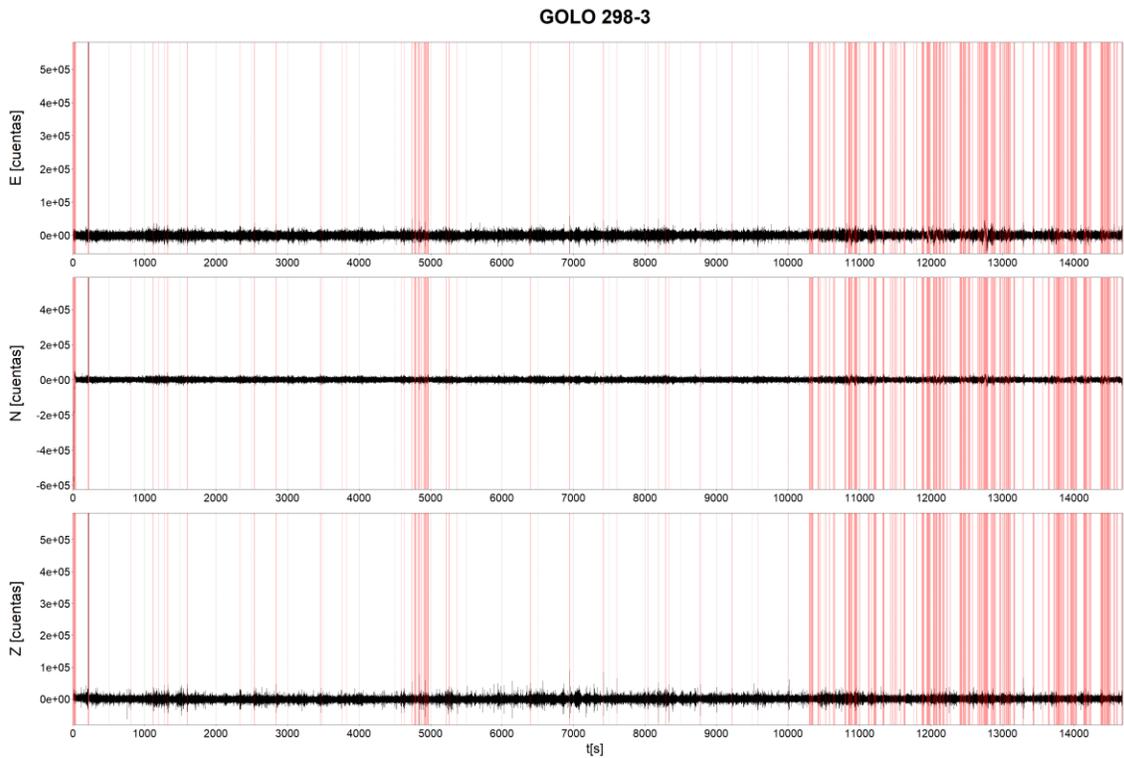


Figura 59: Sismograma de GOLO 298-3. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

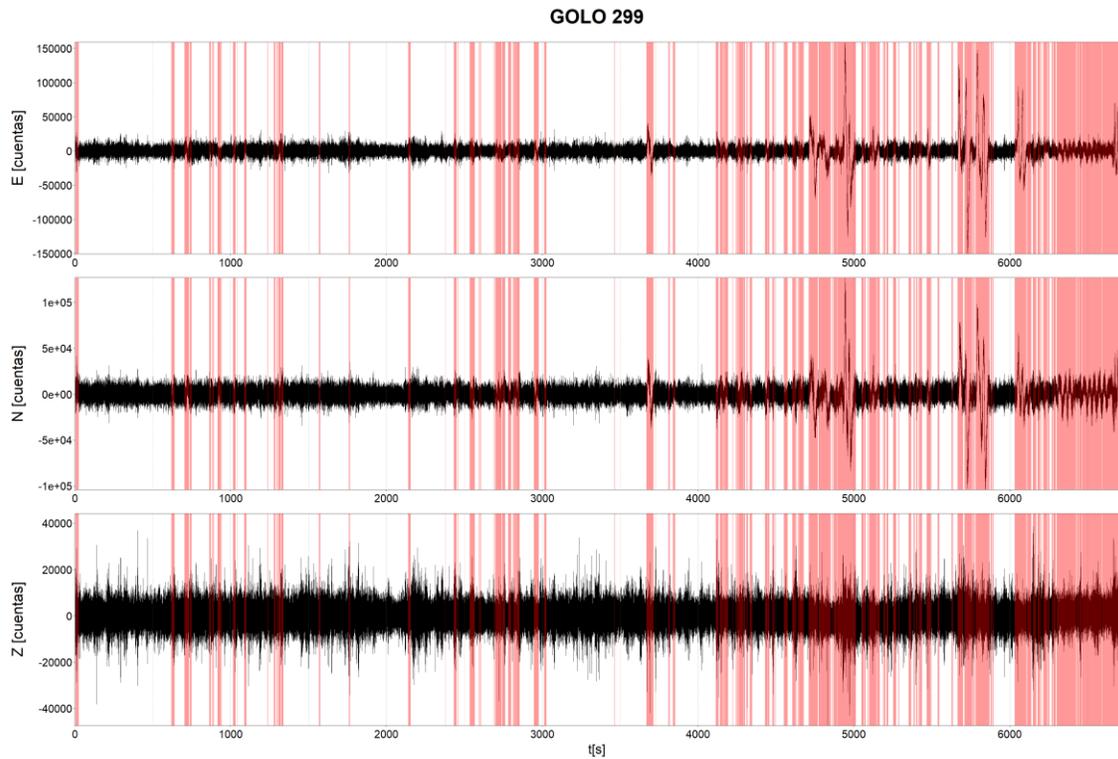


Figura 60: Sismograma de GOLO 299. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

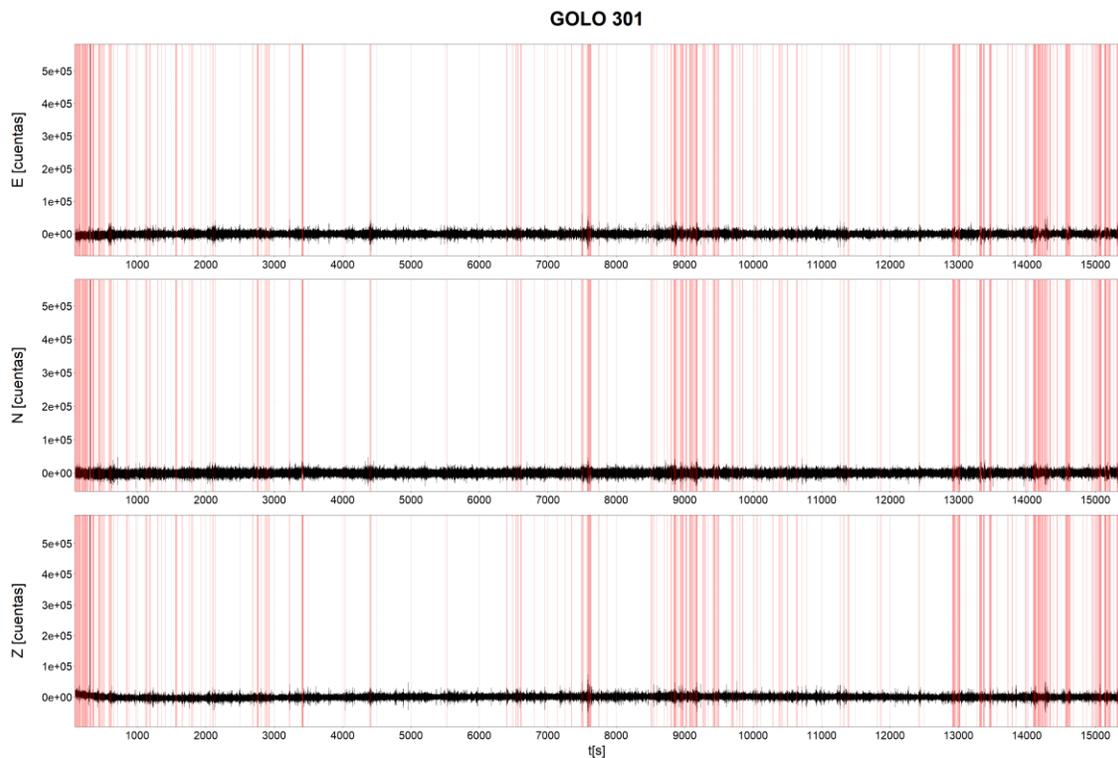


Figura 61: Sismograma de GOLO 301. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

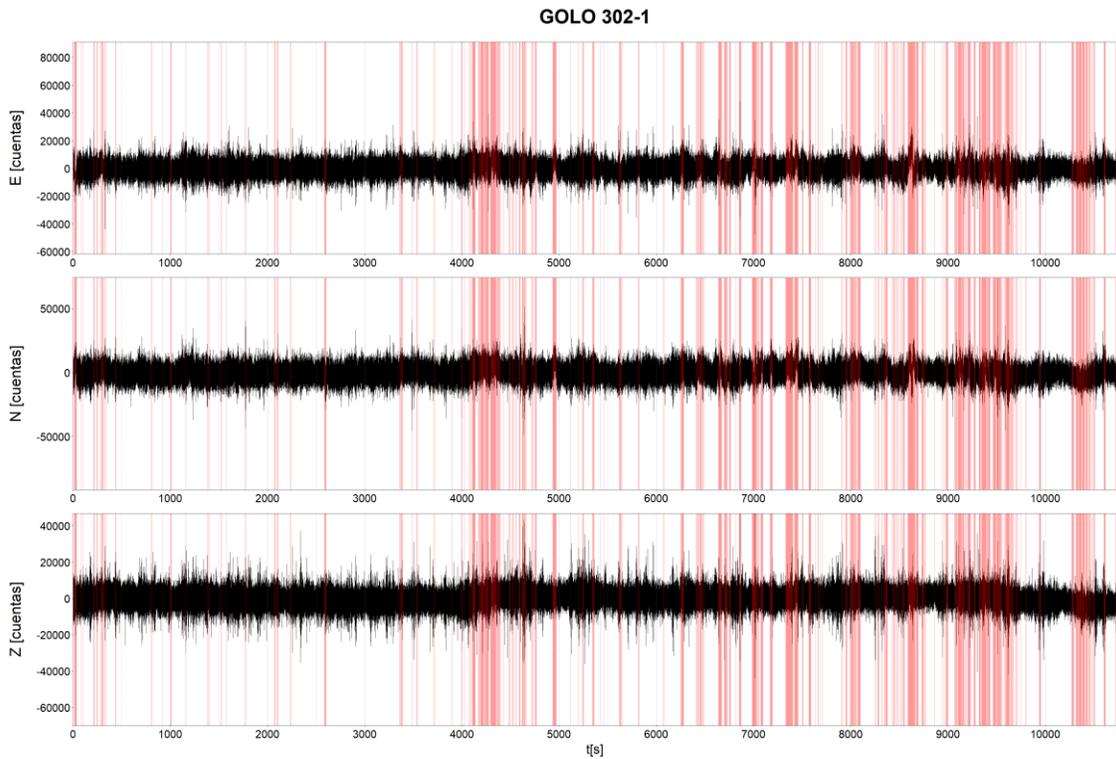


Figura 62: Sismograma de GOLO 302-1. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

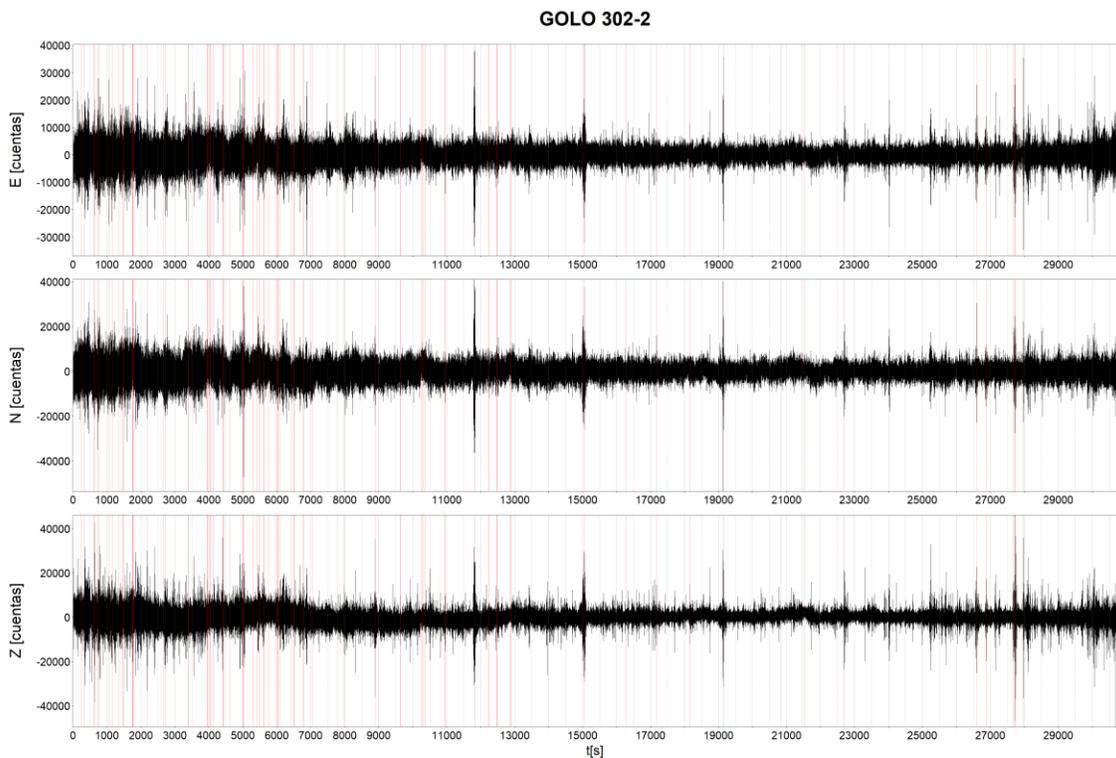


Figura 63: Sismograma de GOLO 302-2. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

GOLO 302-3

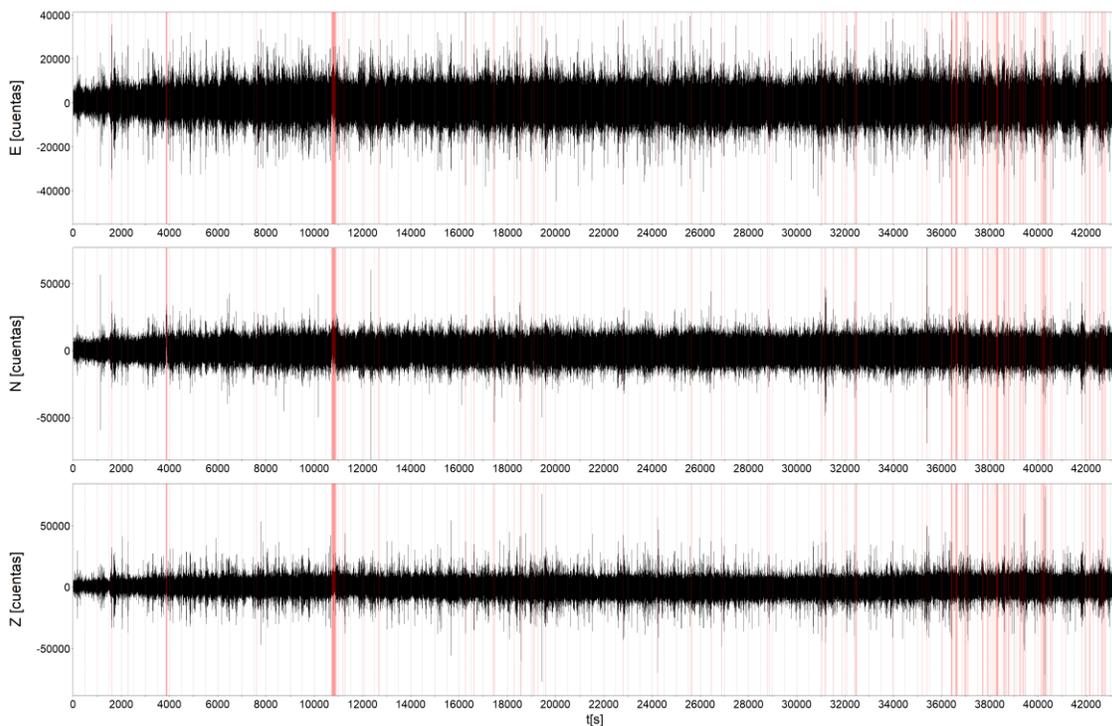


Figura 64: Sismograma de GOLO 302-3. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

GOLO 303-1

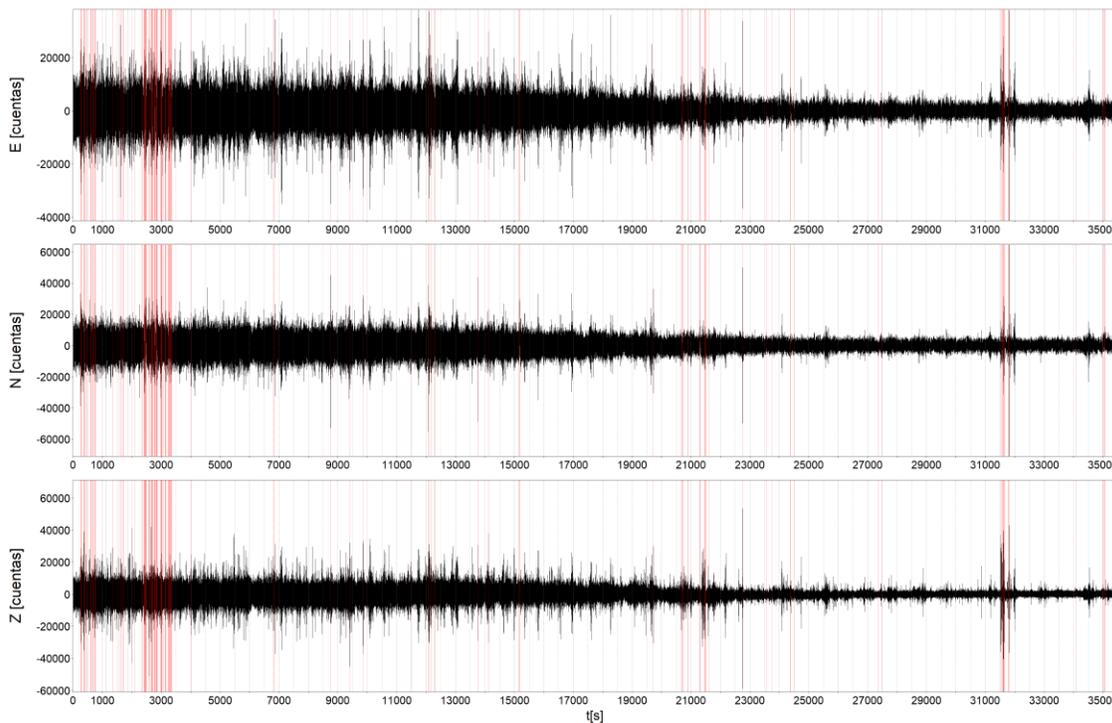


Figura 65: Sismograma de GOLO 303-1. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

GOLO 303-2

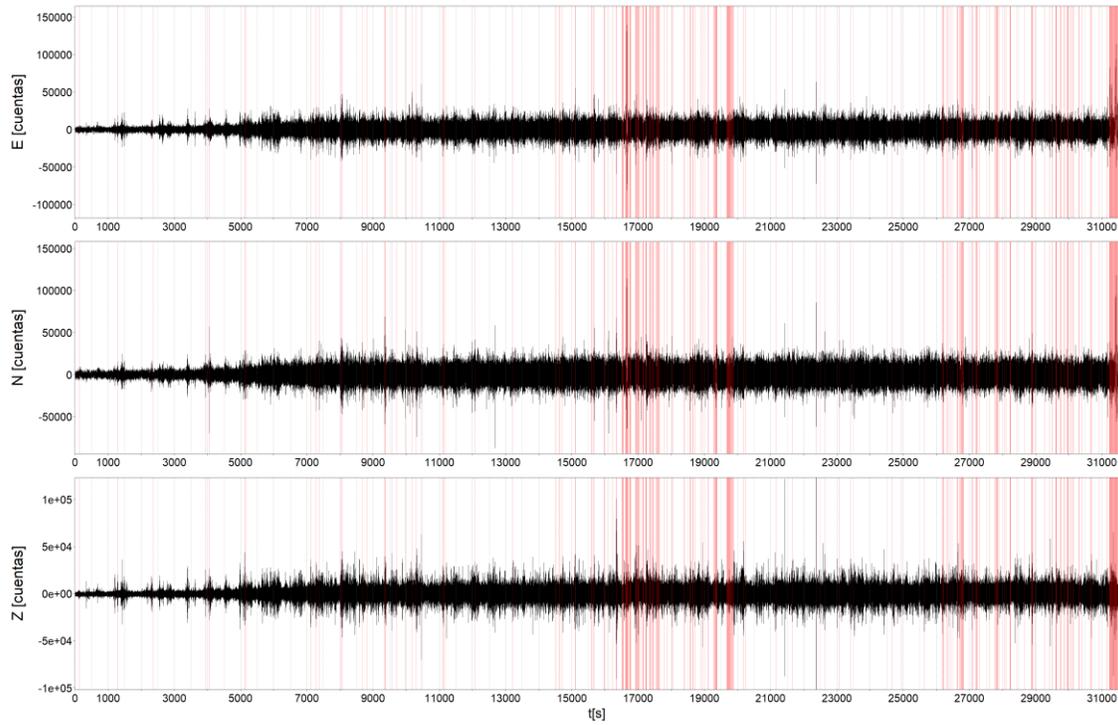


Figura 66: Sismograma de GOLO 303-2. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

GOLO 303-3

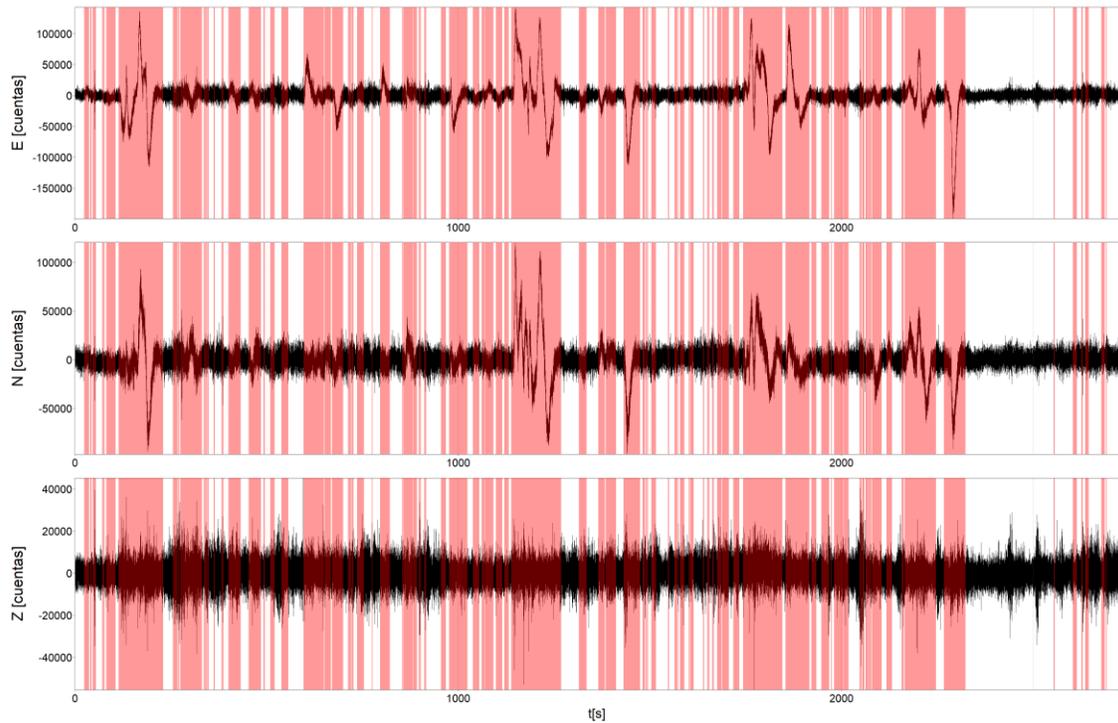


Figura 67: Sismograma de GOLO 303-3. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

GOLO 303-4

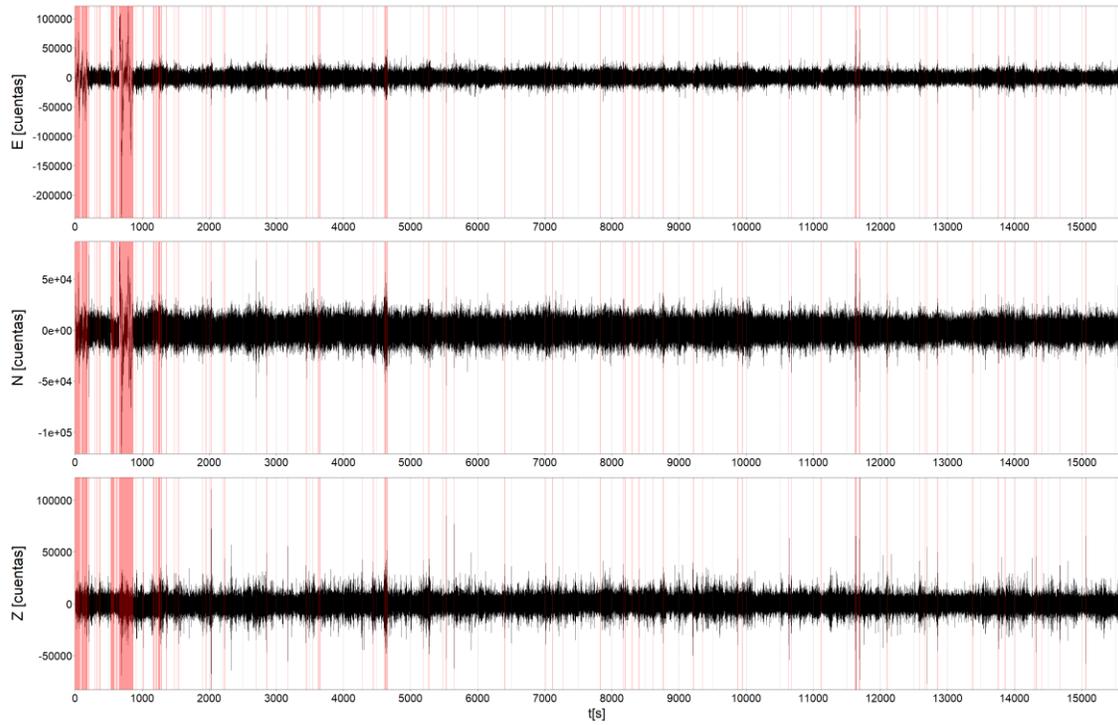


Figura 68: Sismograma de GOLO 303-4. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

GOLO 304-1

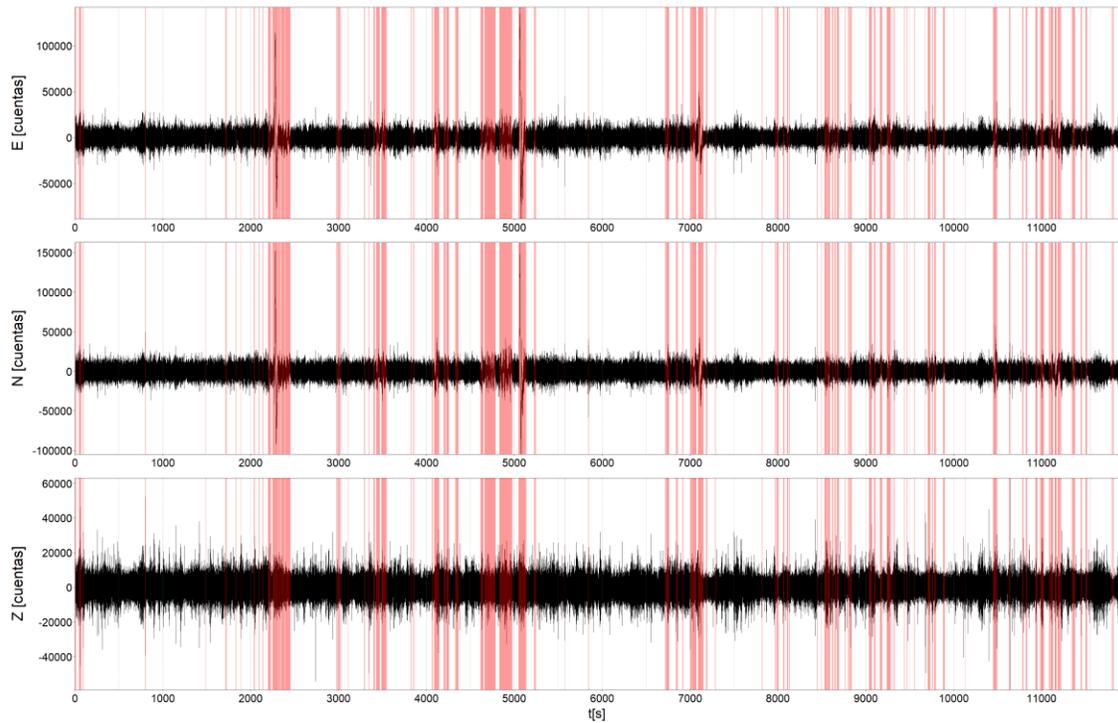


Figura 69: Sismograma de GOLO 304-1. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

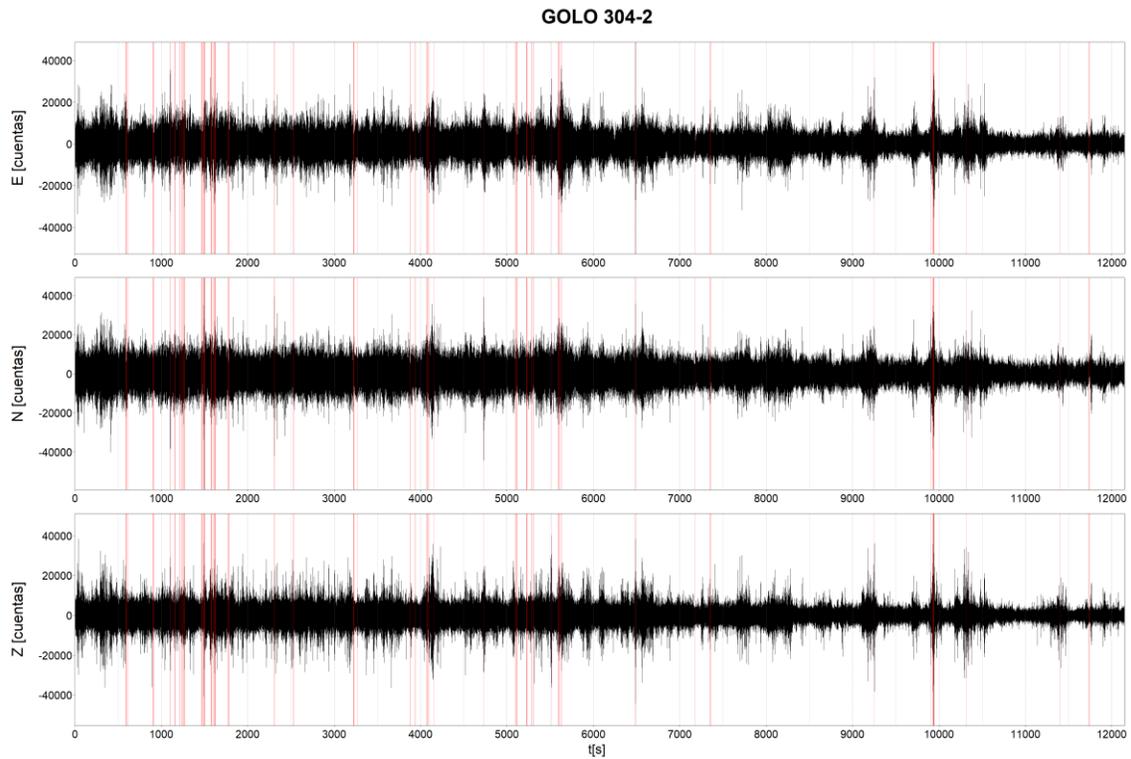


Figura 70: Sismograma de GOLO 304-2. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

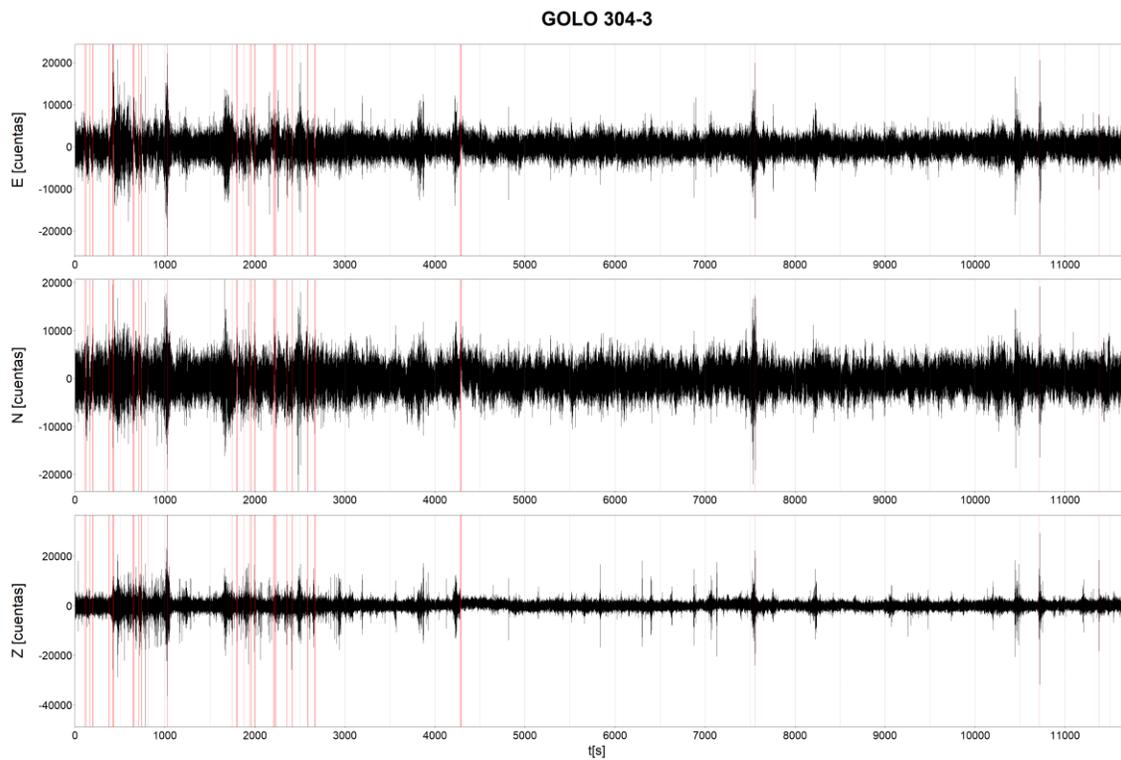


Figura 71: Sismograma de GOLO 304-3. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

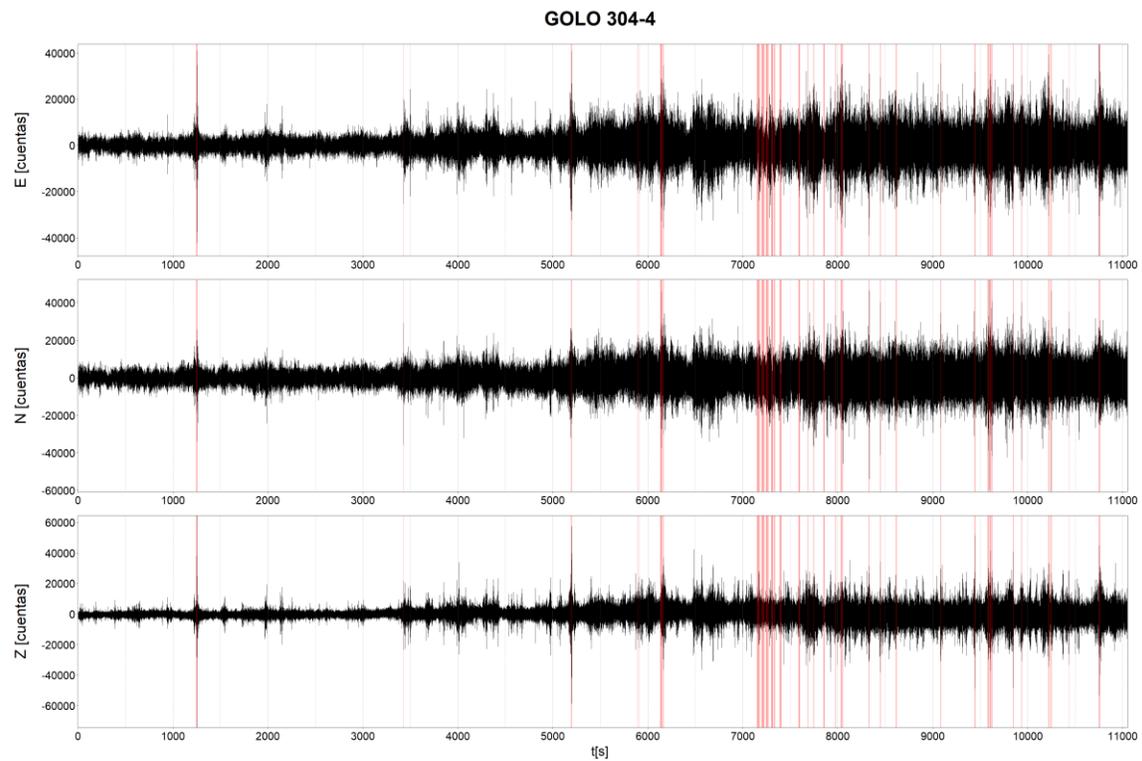


Figura 72: Sismograma de GOLO 304-4. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

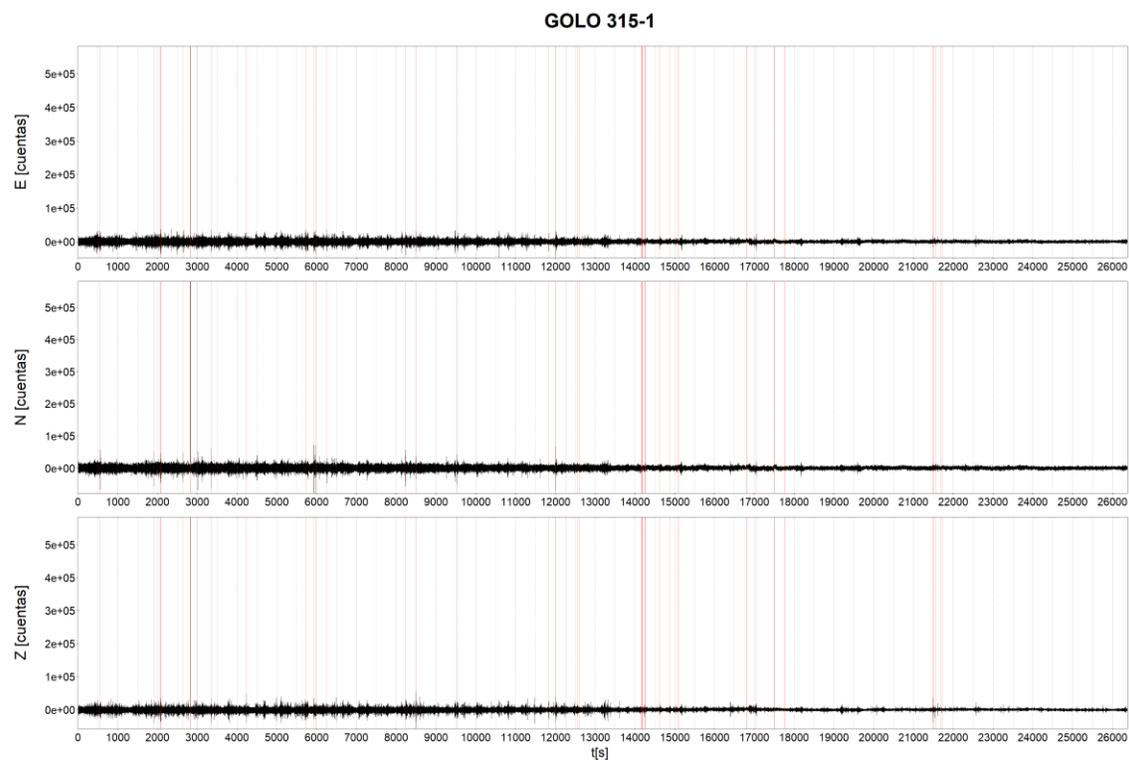


Figura 73: Sismograma de GOLO 315-1. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

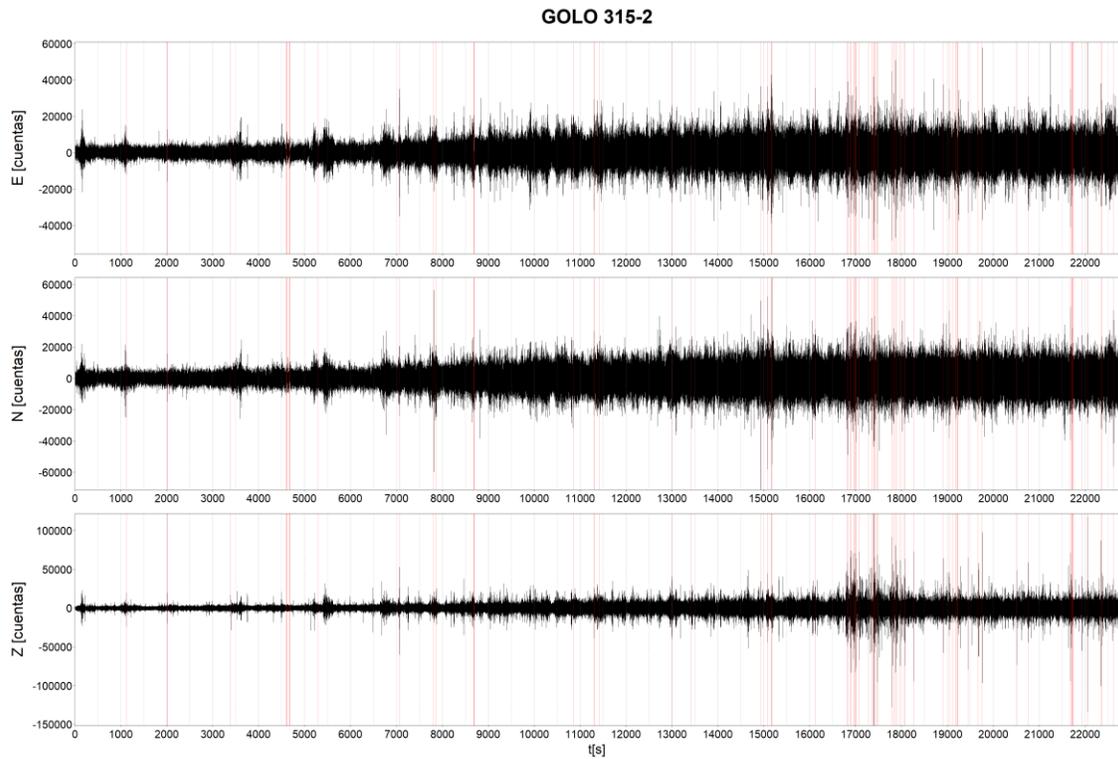


Figura 74: Sismograma de GOLO 315-2. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

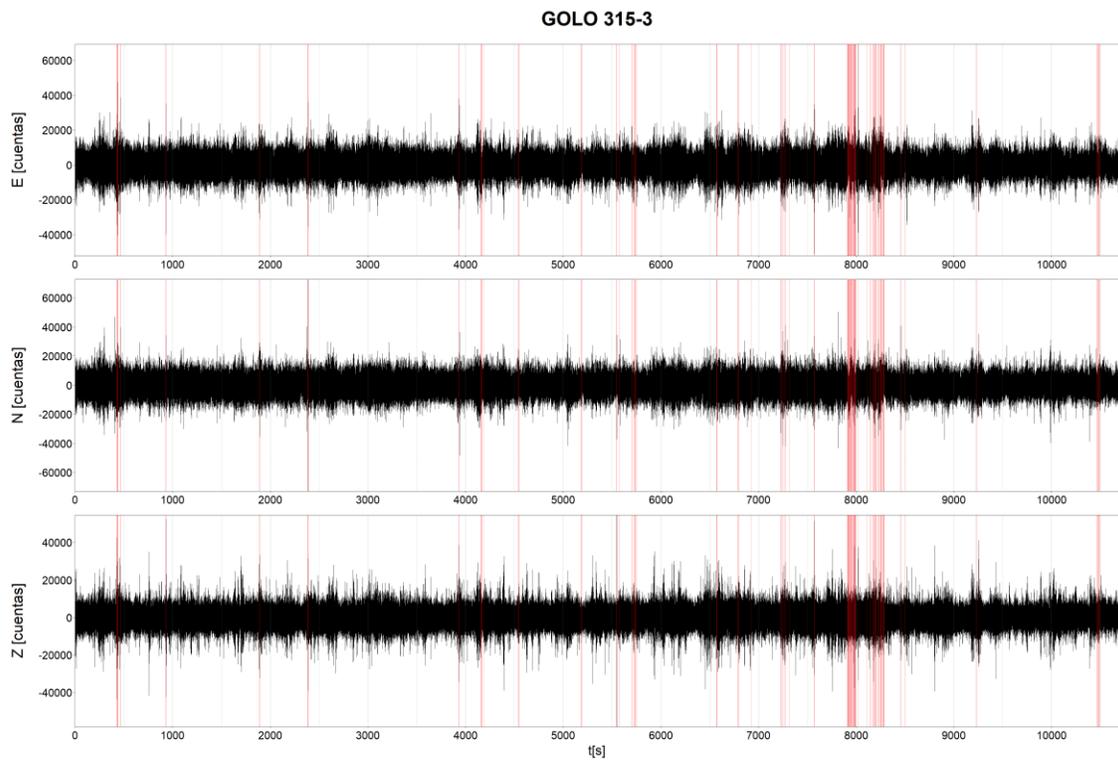


Figura 75: Sismograma de GOLO 315-3. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

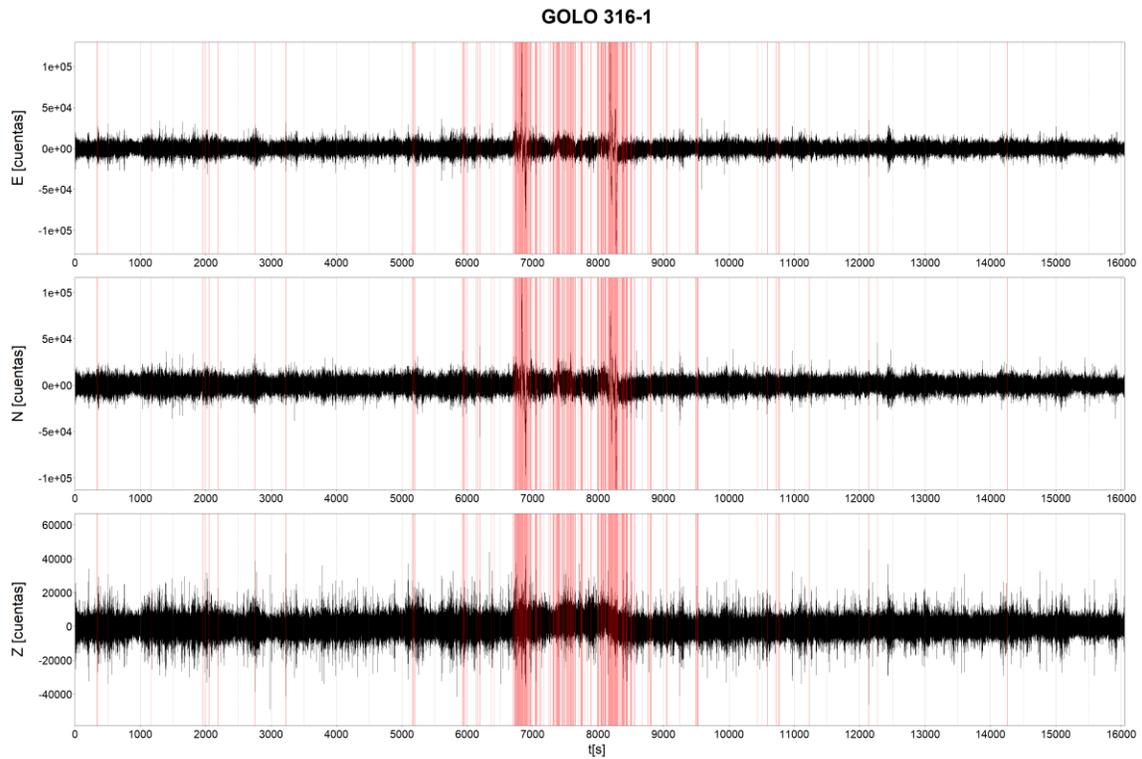


Figura 76: Sismograma de GOLO 316-1. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

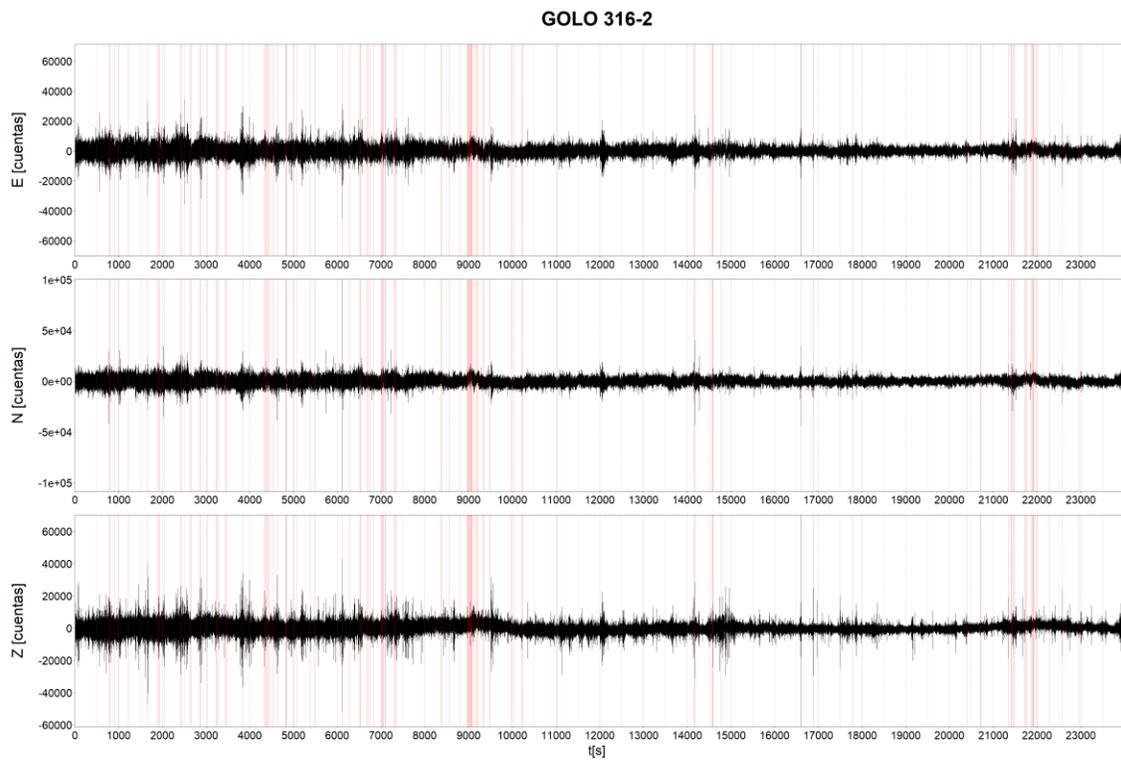


Figura 77: Sismograma de GOLO 316-2. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

GOLO 316-3

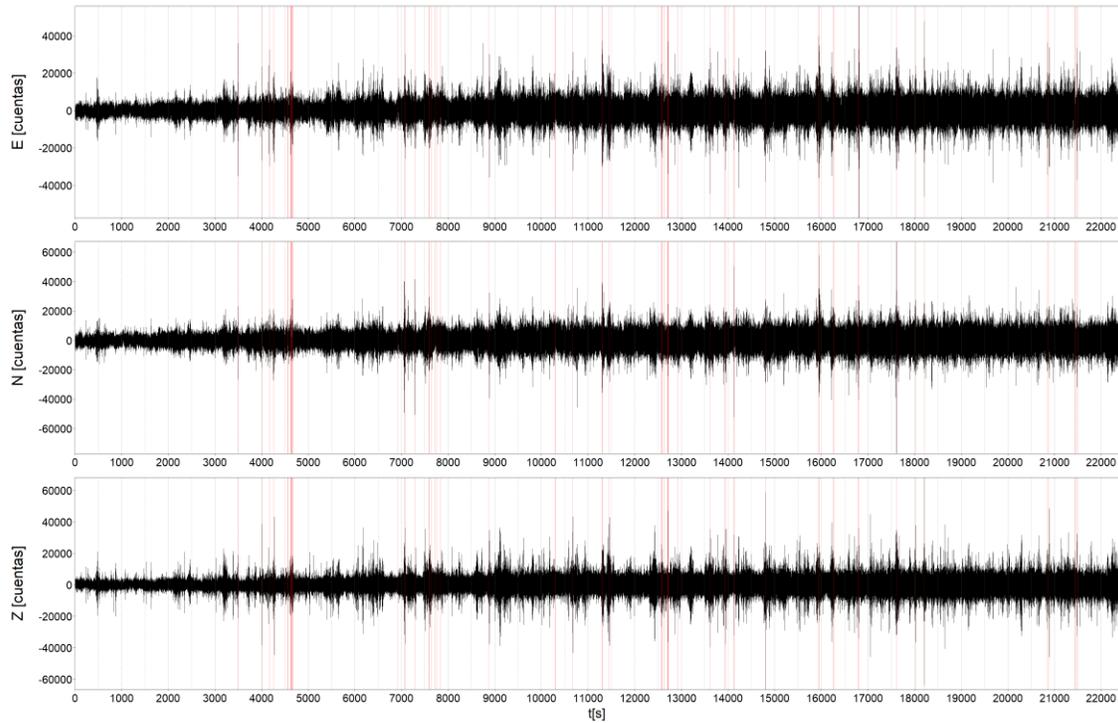


Figura 78: Sismograma de GOLO 316-3. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

GOLO 316-4

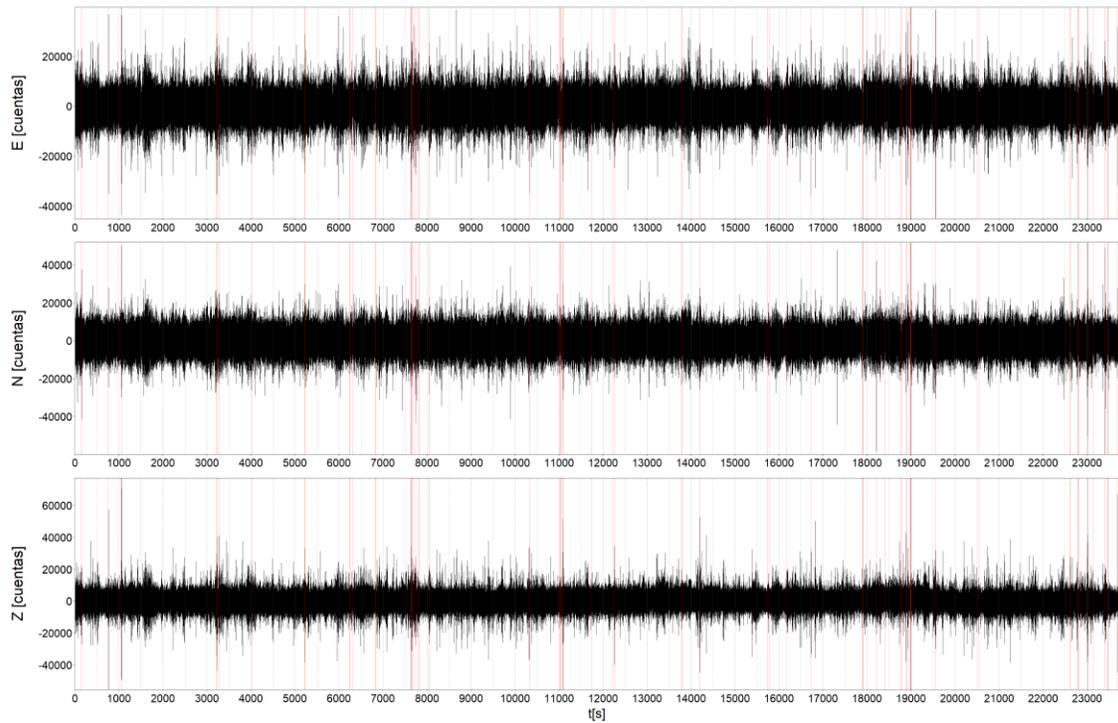


Figura 79: Sismograma de GOLO 316-4. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

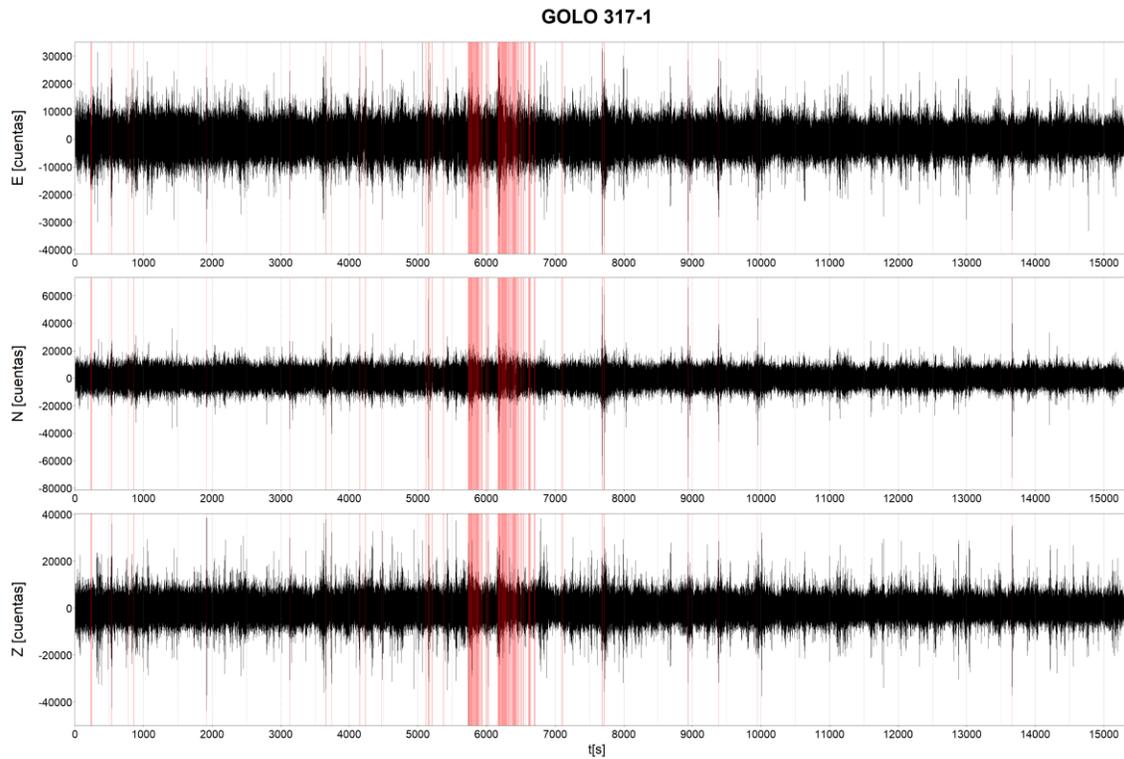


Figura 80: Sismograma de GOLO 317-1. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

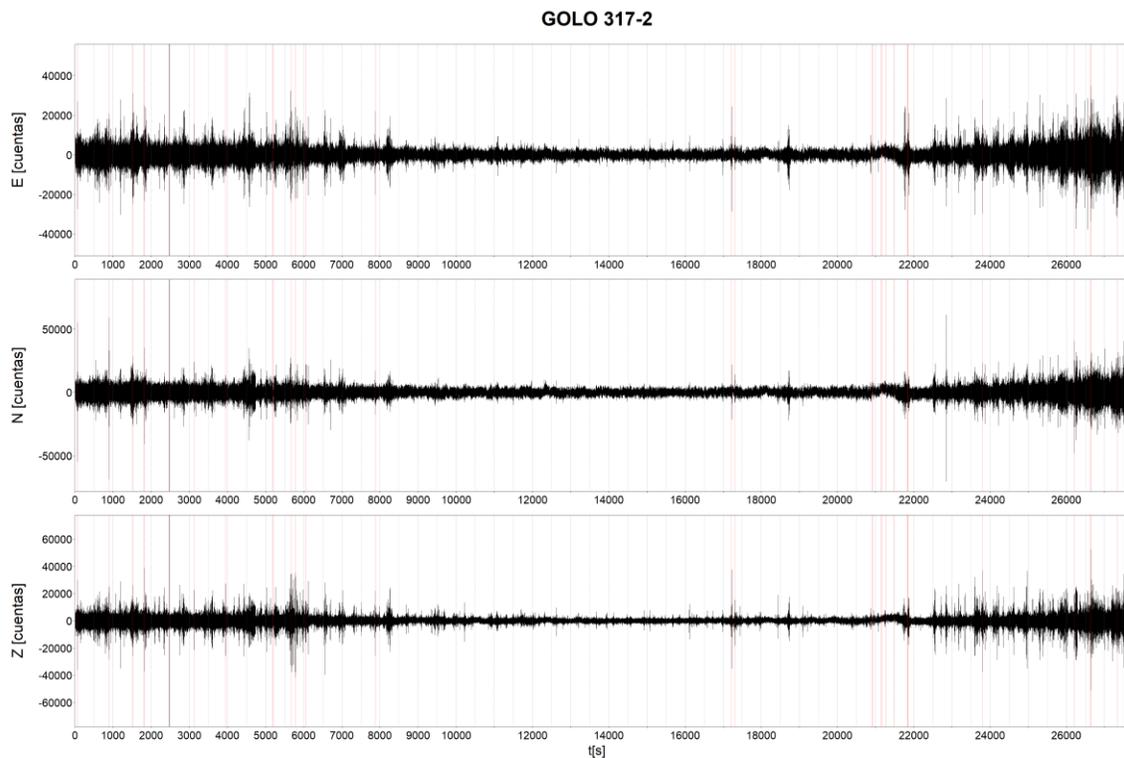


Figura 81: Sismograma de GOLO 317-2. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

GOLO 317-3

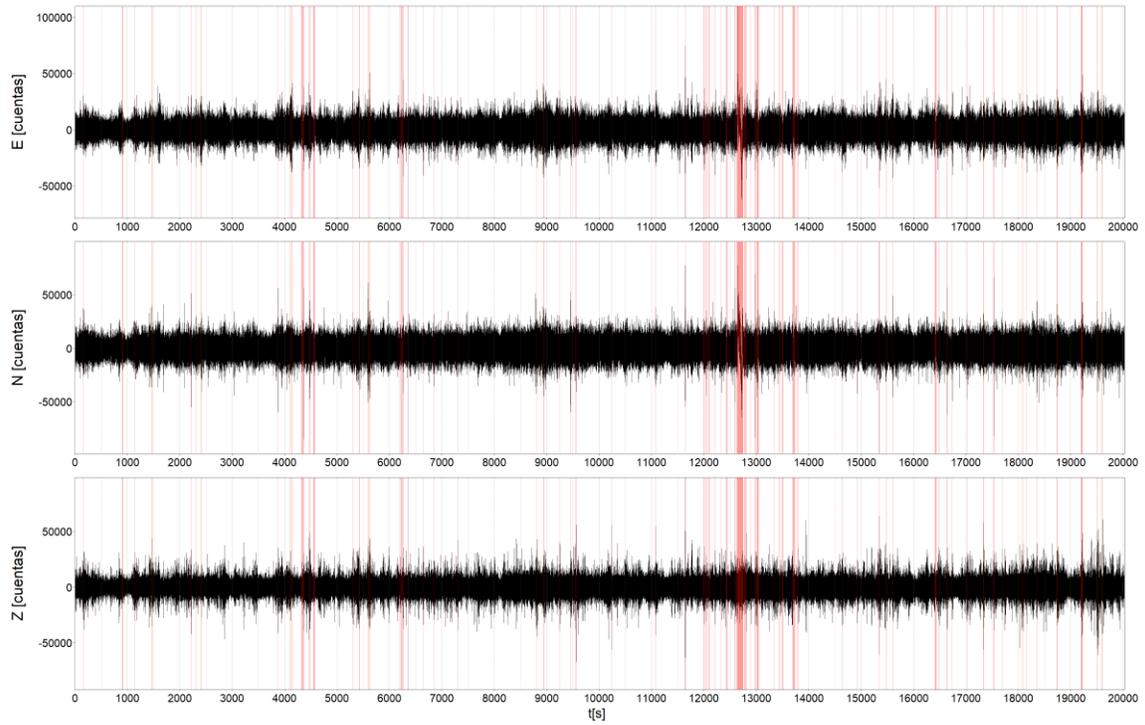


Figura 82: Sismograma de GOLO 317-3. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

GOLO 317-4

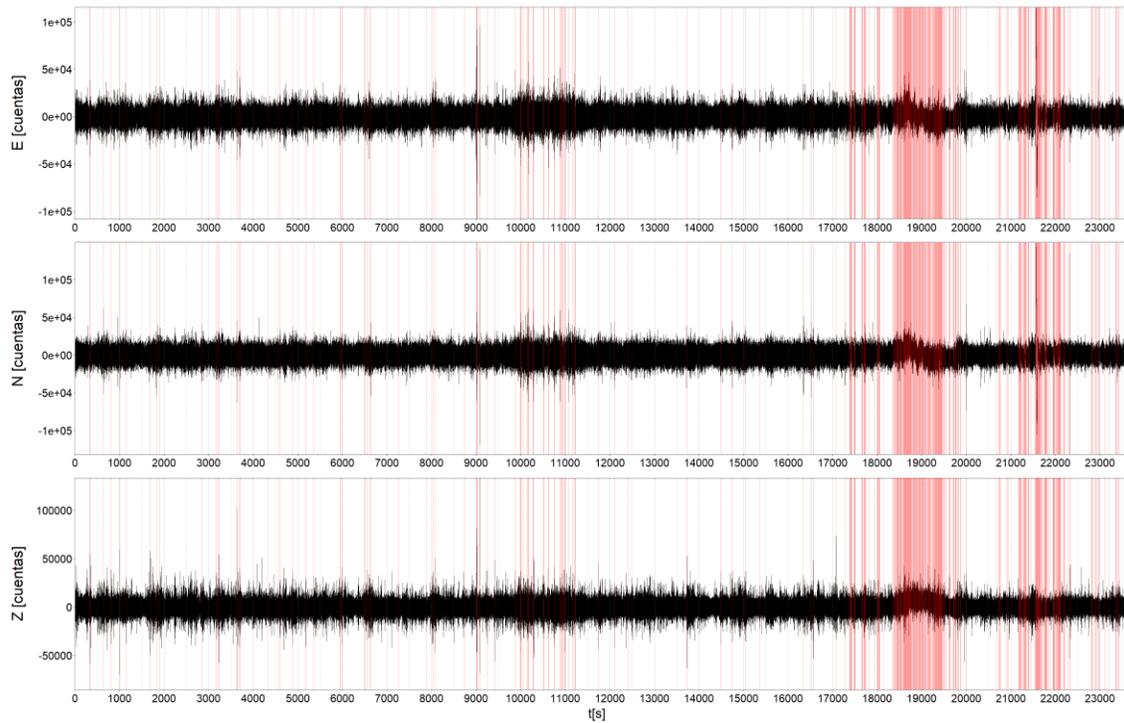


Figura 83: Sismograma de GOLO 317-4. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

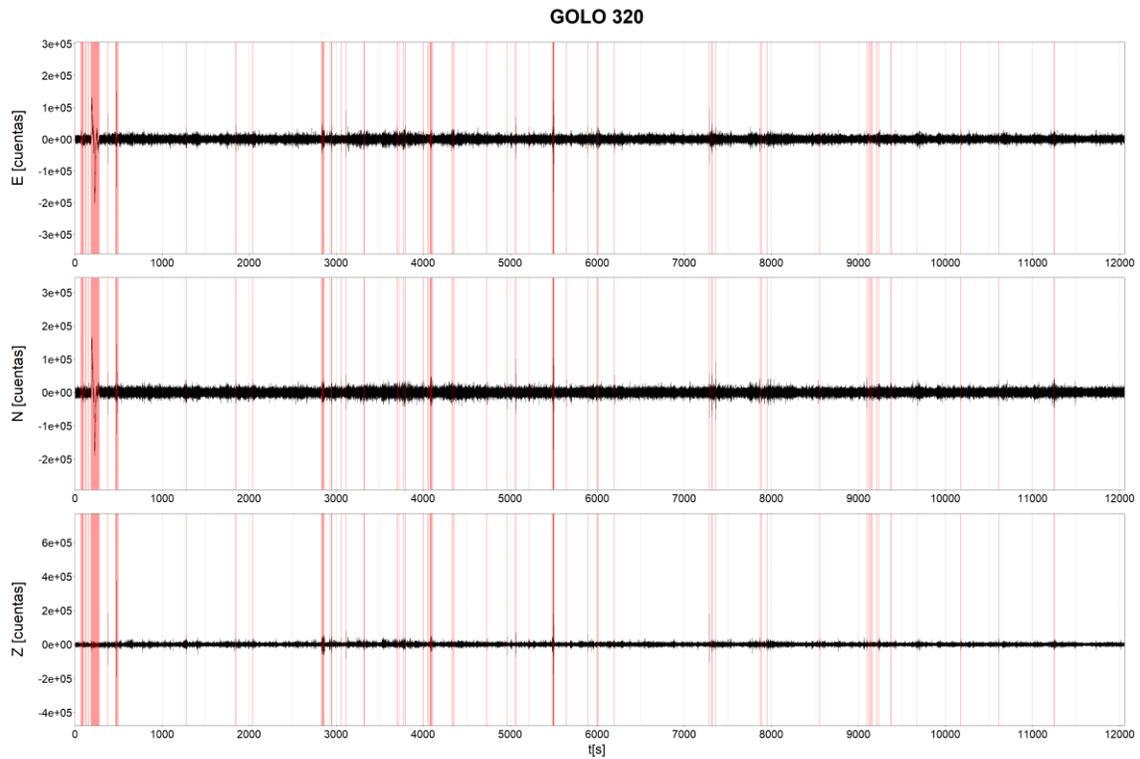


Figura 84: Sismograma de GOLO 320. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

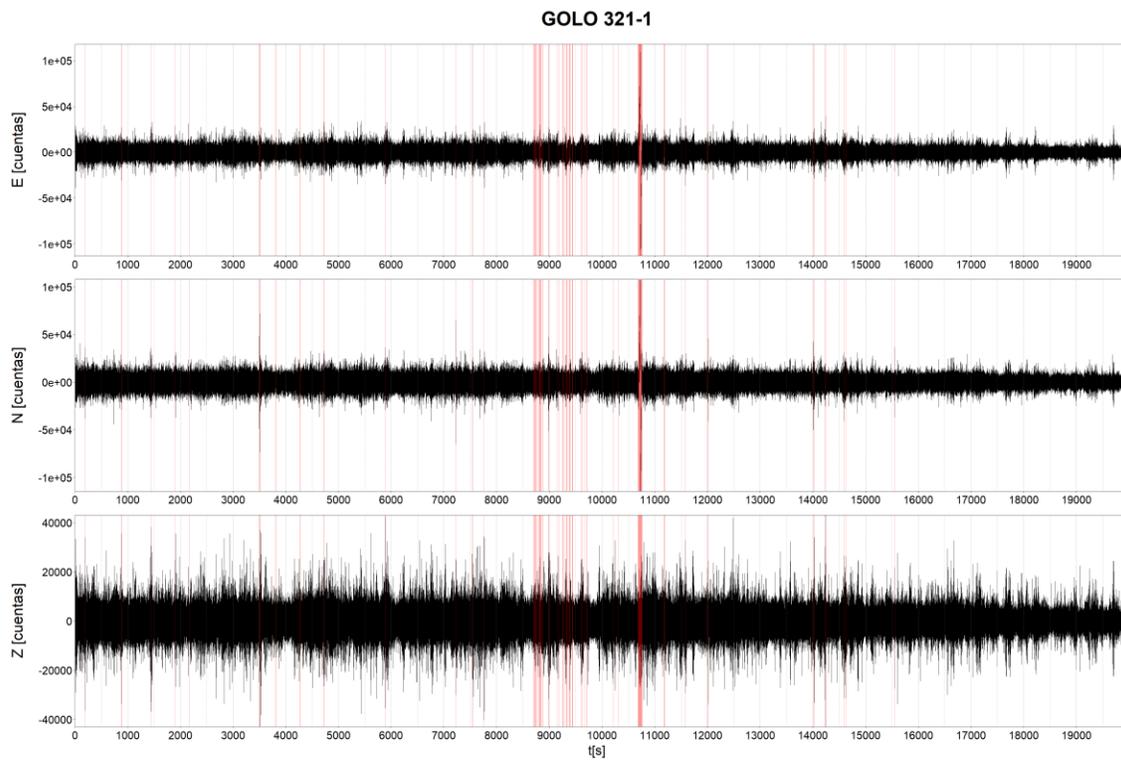


Figura 85: Sismograma de GOLO 321-1. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

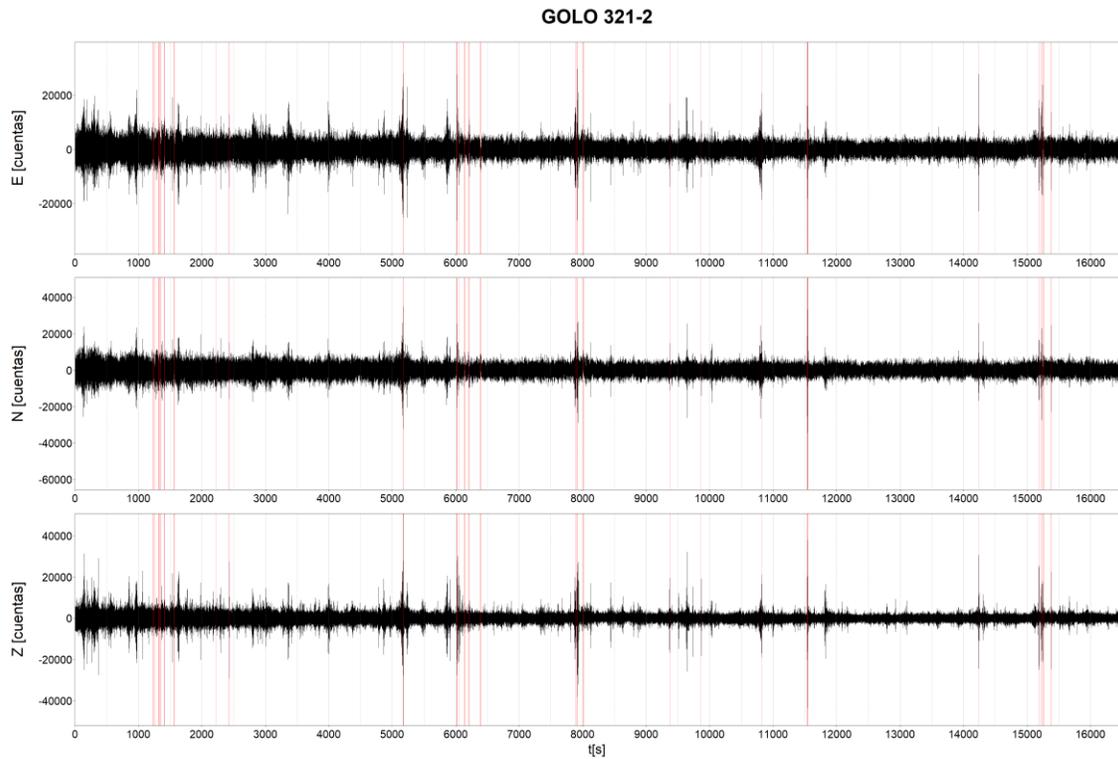


Figura 86: Sismograma de GOLO 321-2. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

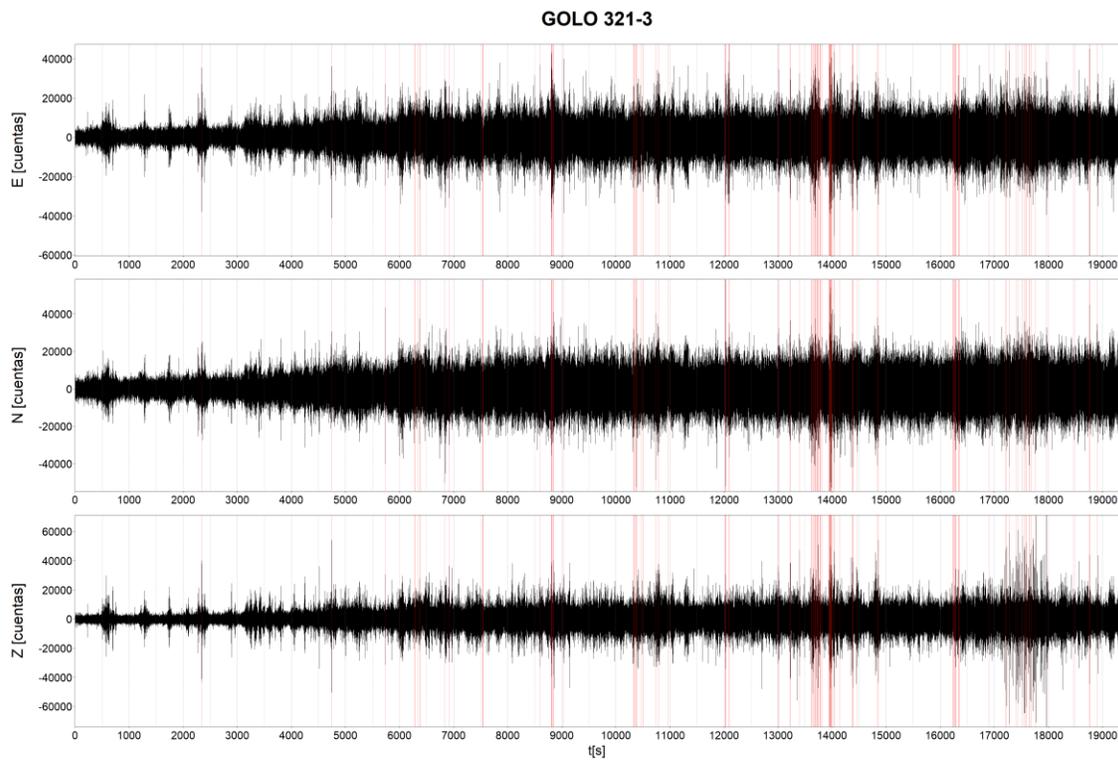


Figura 87: Sismograma de GOLO 321-3. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

GOLO 321-4

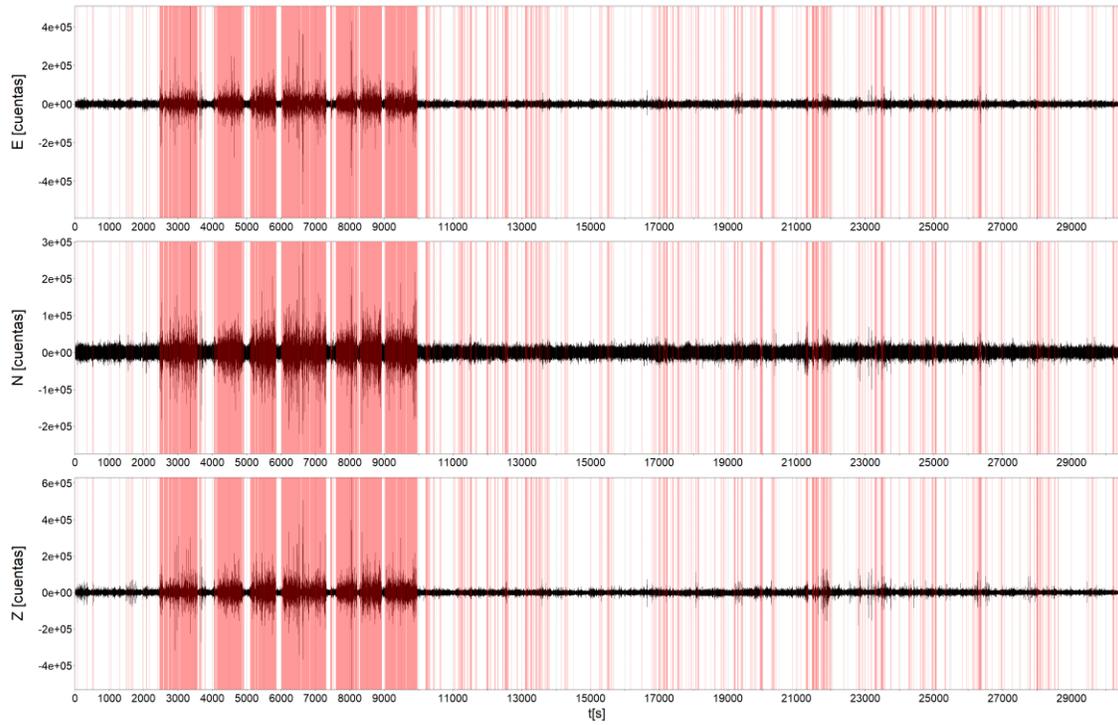


Figura 88: Sismograma de GOLO 321-4. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

GOLO 322-1

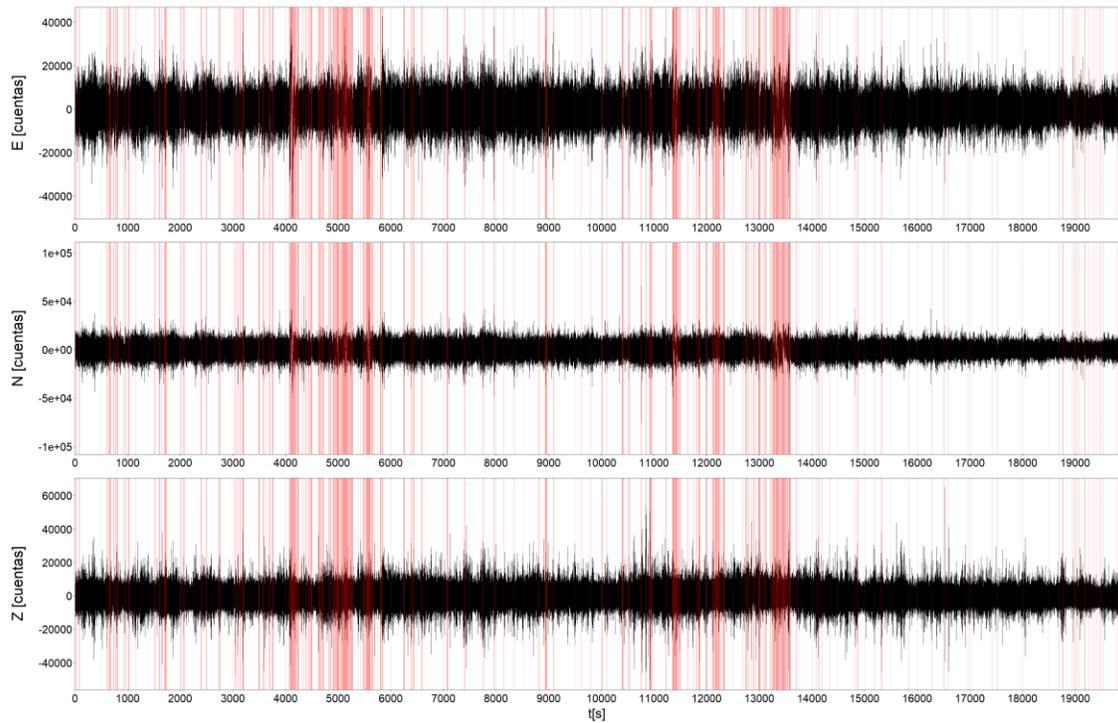


Figura 89: Sismograma de GOLO 322-1. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

GOLO 322-2

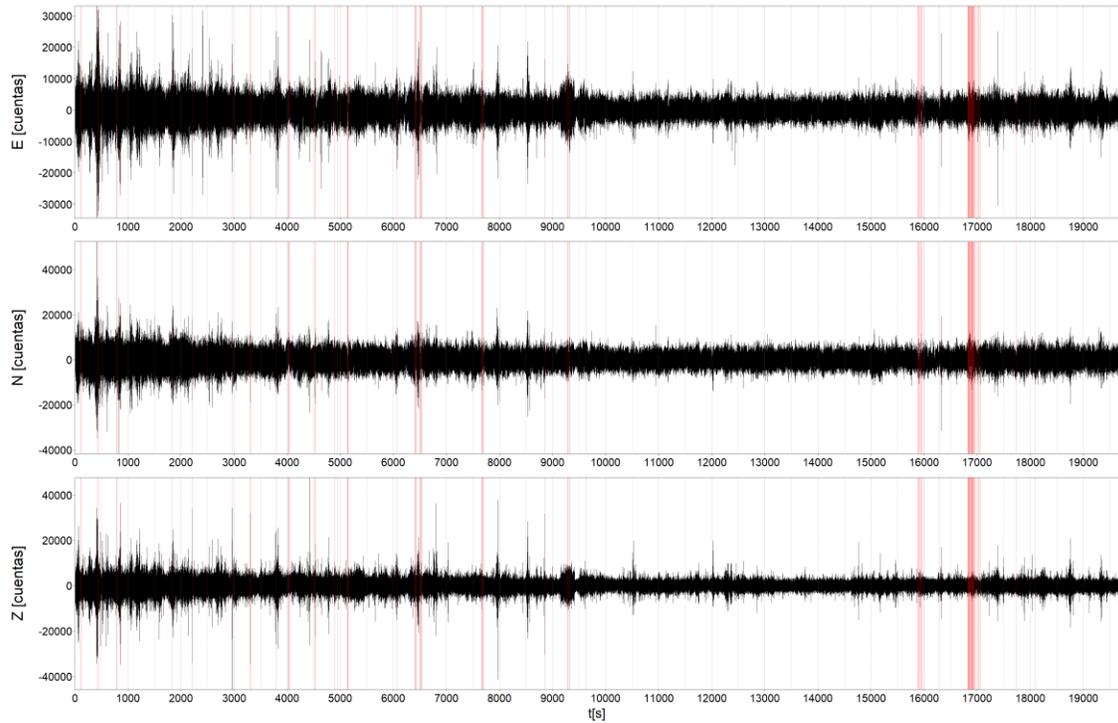


Figura 90: Sismograma de GOLO 322-2. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

GOLO 322-3

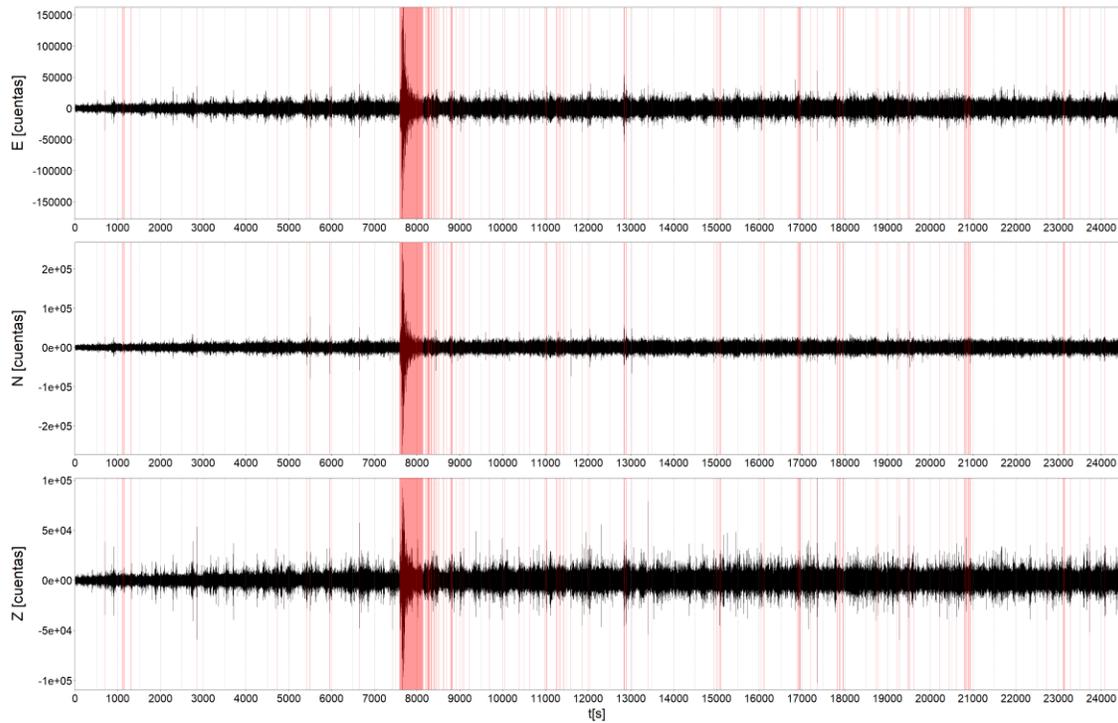


Figura 91: Sismograma de GOLO 322-3. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

GOLO 322-4

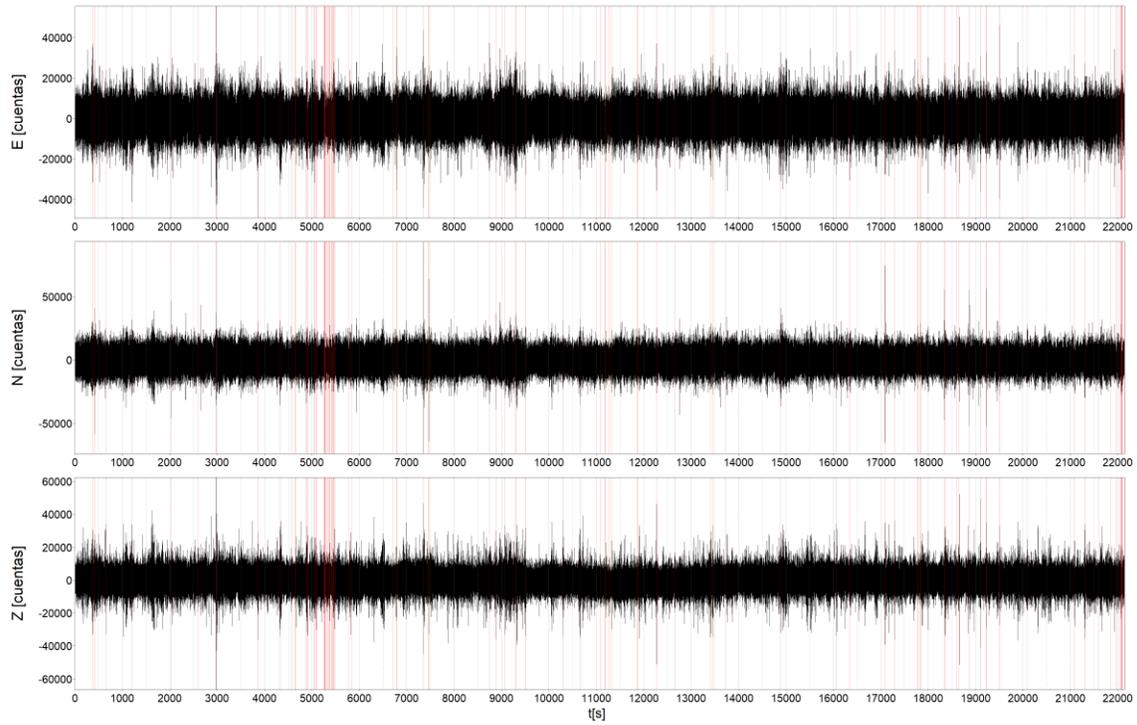


Figura 92: Sismograma de GOLO 322-4. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

GOLO 323-1

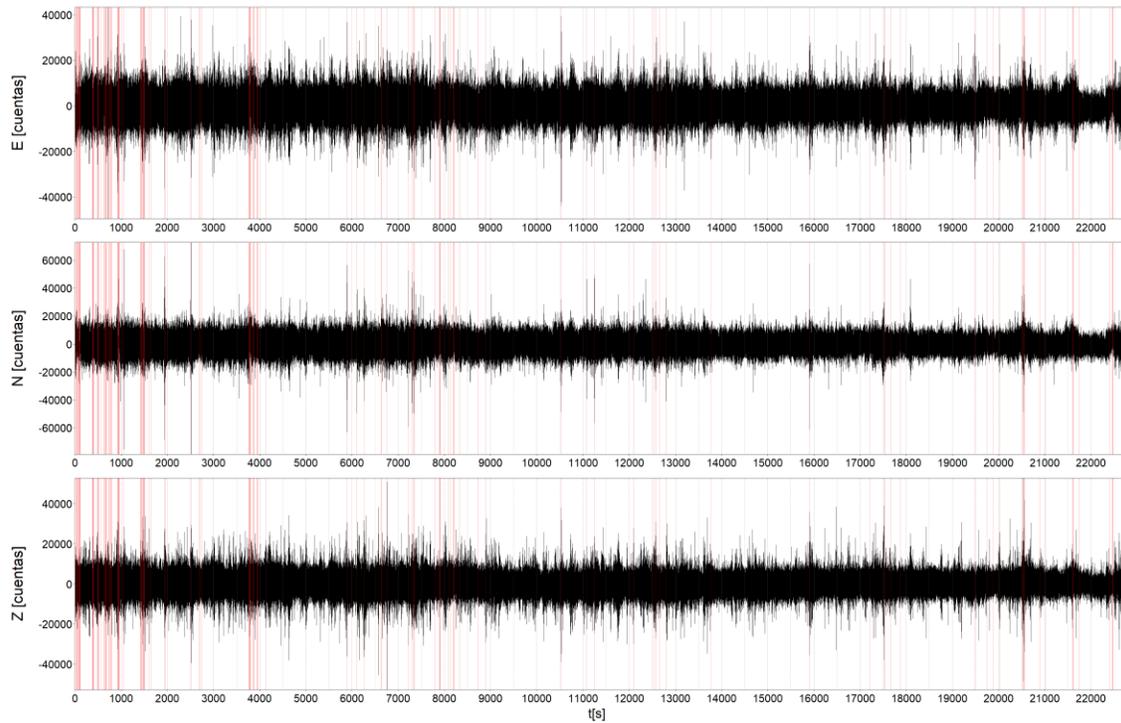


Figura 93: Sismograma de GOLO 323-1. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

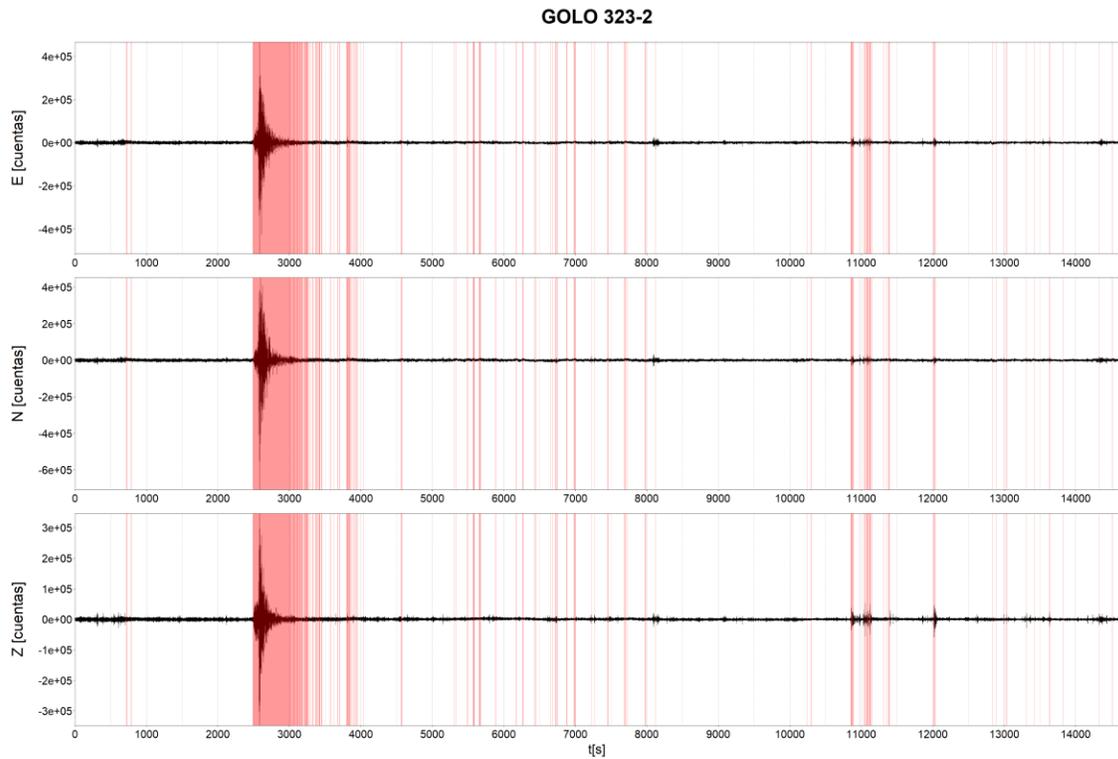


Figura 94: Sismograma de GOLO 323-2. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

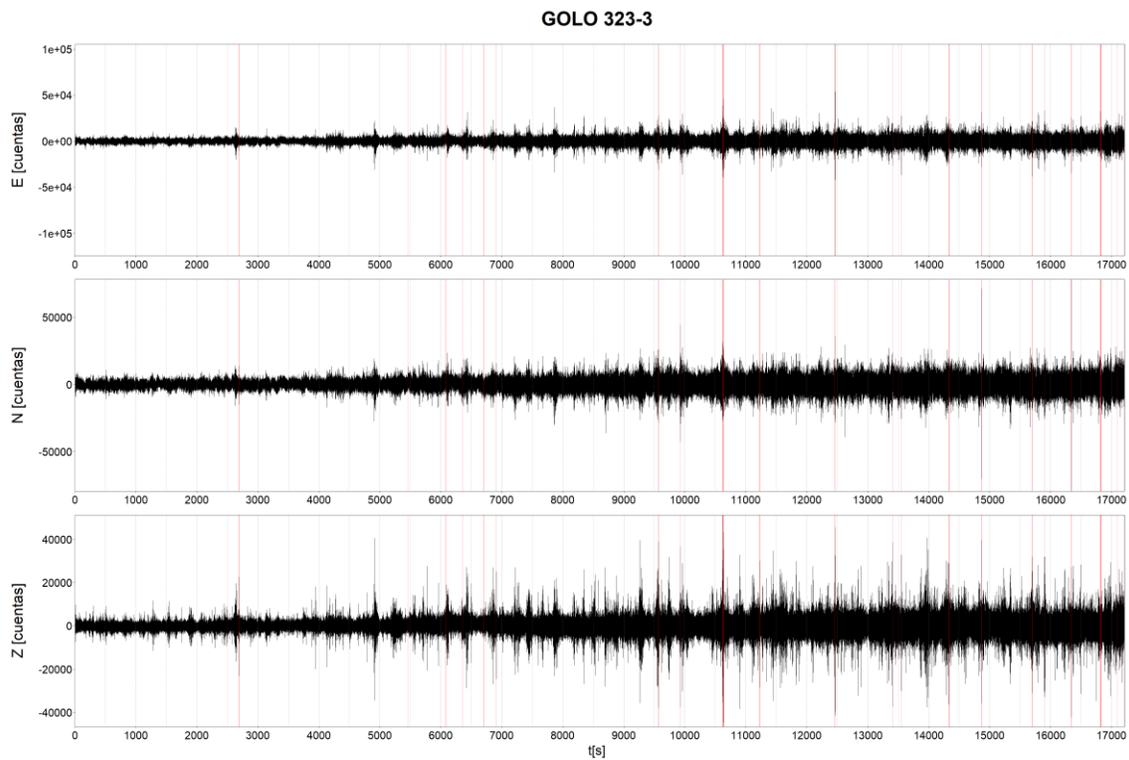


Figura 95: Sismograma de GOLO 323-3. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

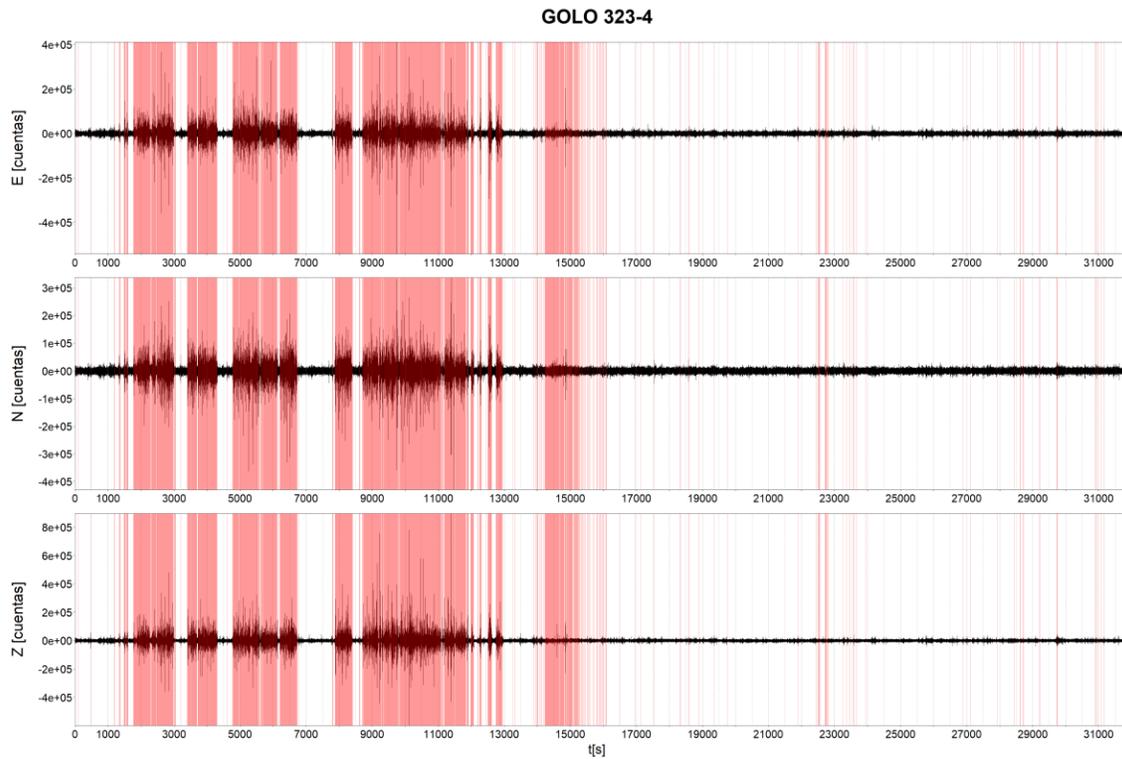


Figura 96: Sismograma de GOLO 323-4. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

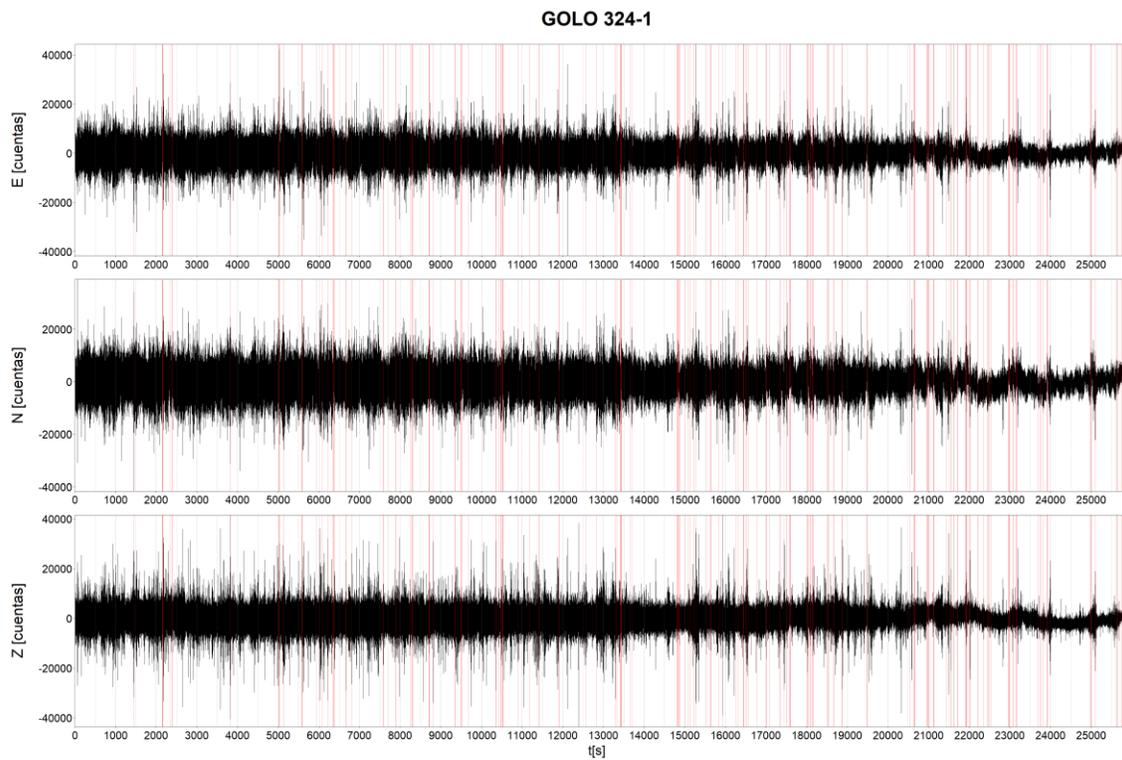


Figura 97: Sismograma de GOLO 324-1. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

GOLO 324-2

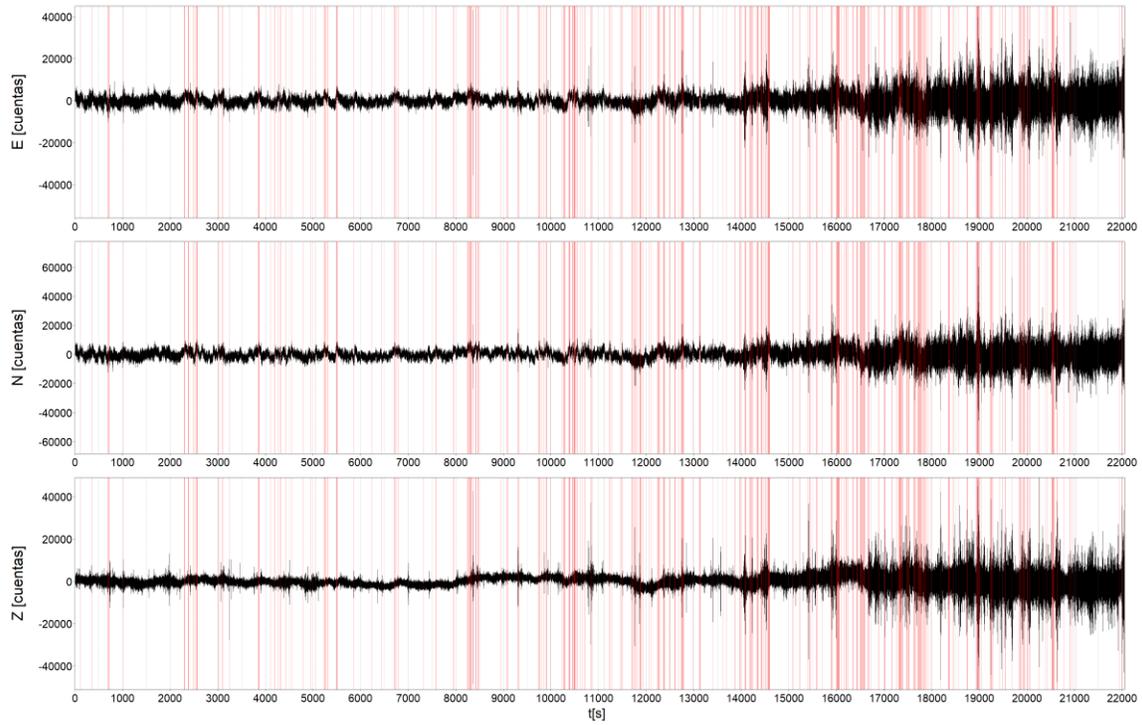


Figura 98: Sismograma de GOLO 324-2. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

GOLO 324-3

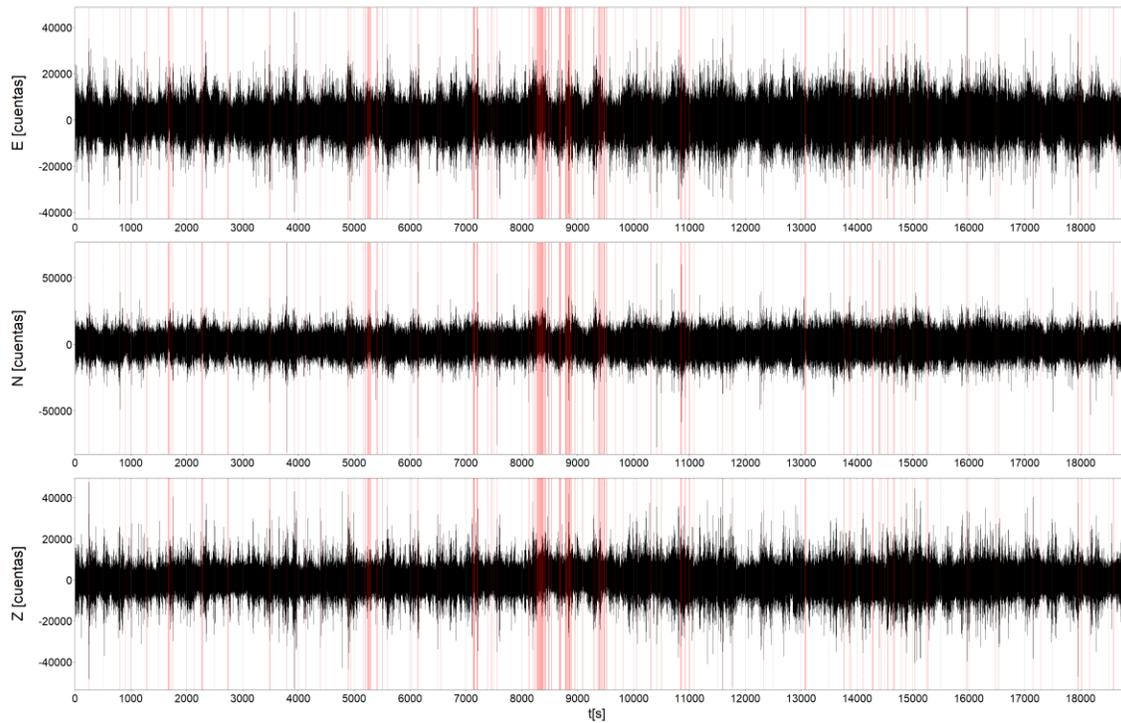


Figura 99: Sismograma de GOLO 324-3. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

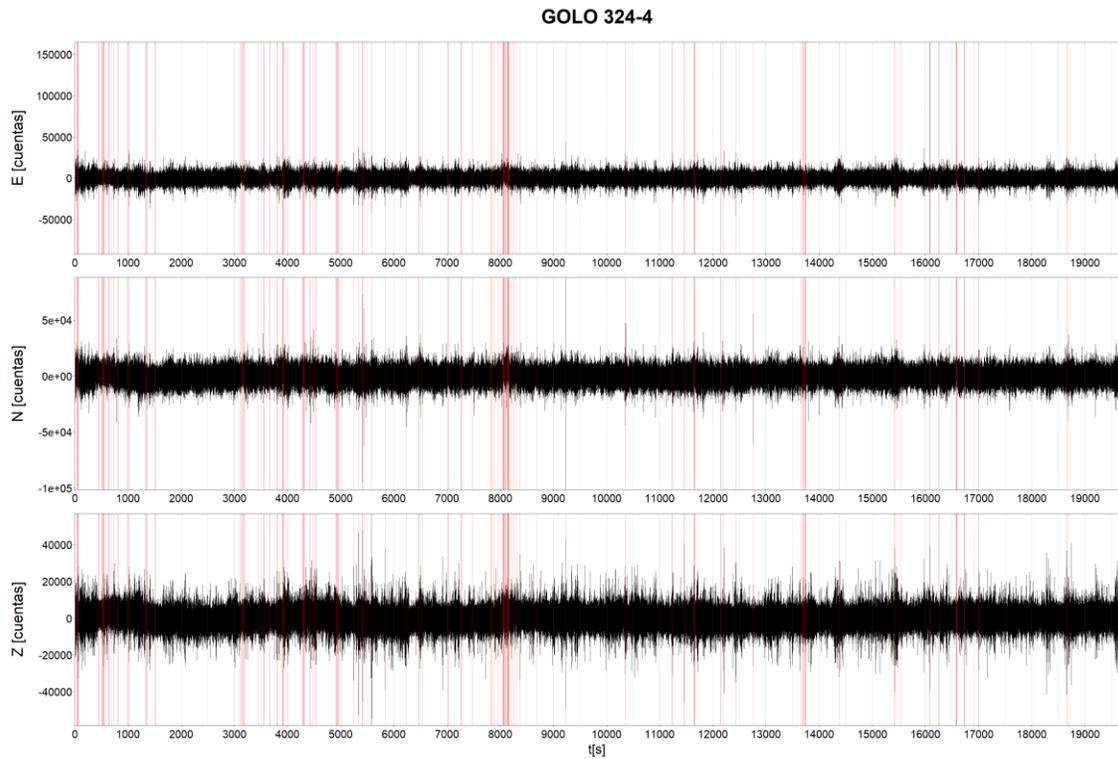


Figura 100: Sismograma de GOLO 324-4. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

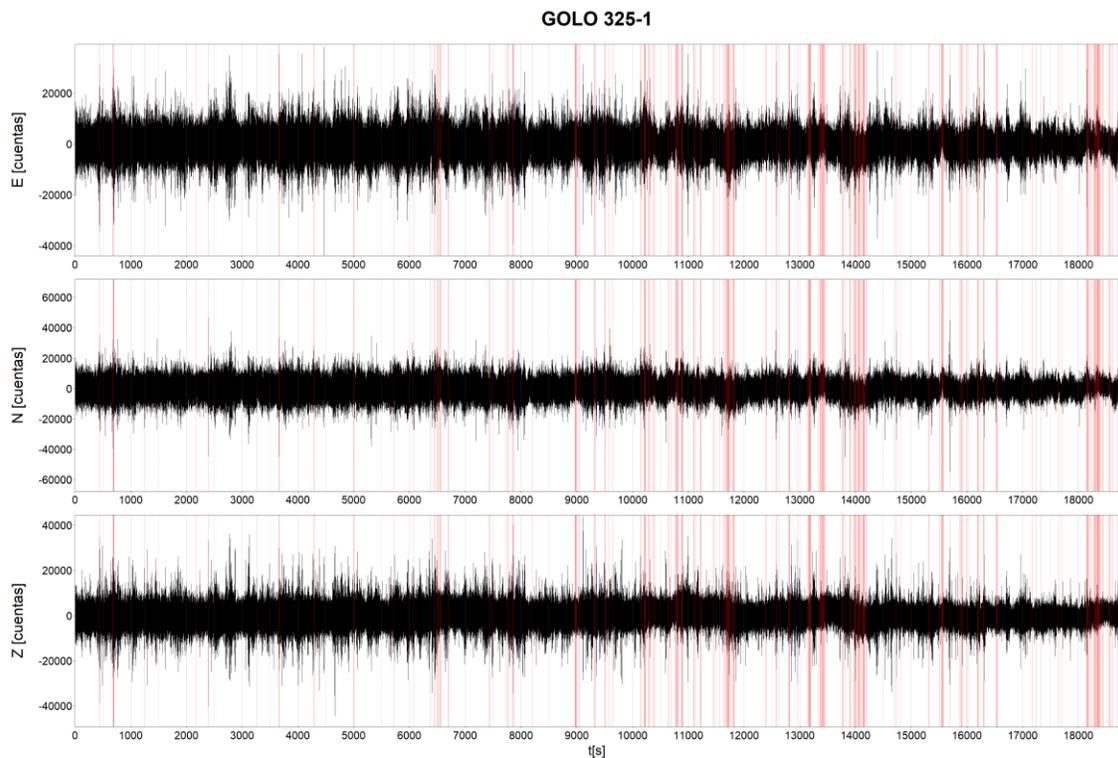


Figura 101: Sismograma de GOLO 325-1. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

GOLO 325-2

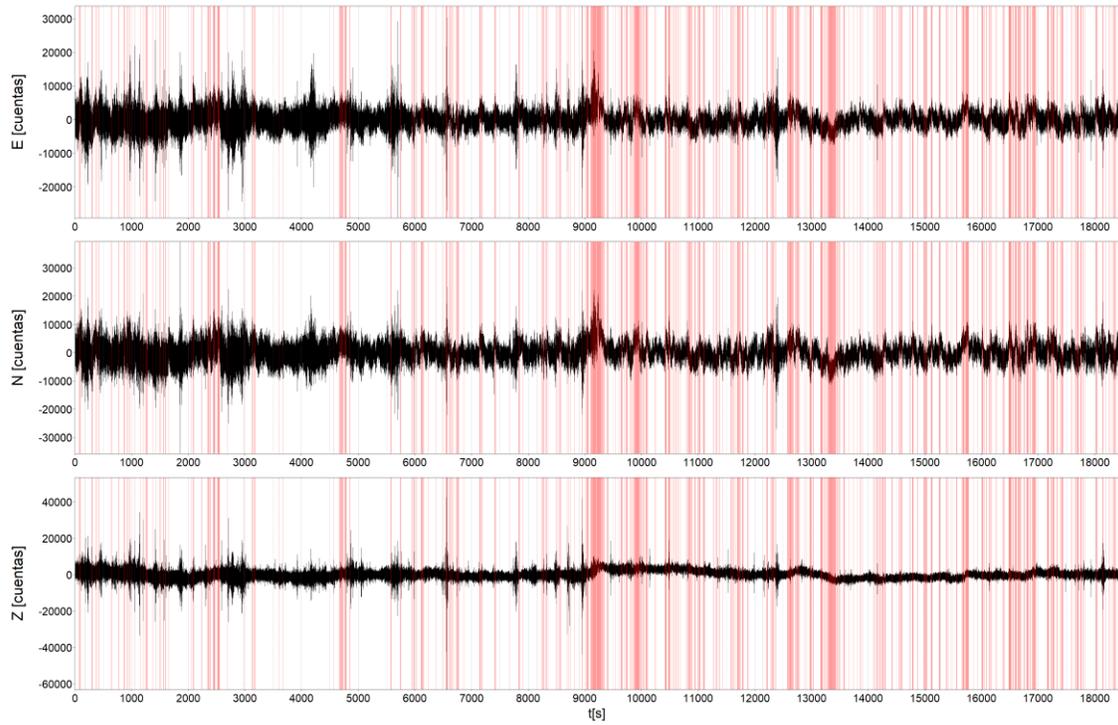


Figura 102: Sismograma de GOLO 325-2. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

GOLO 325-3

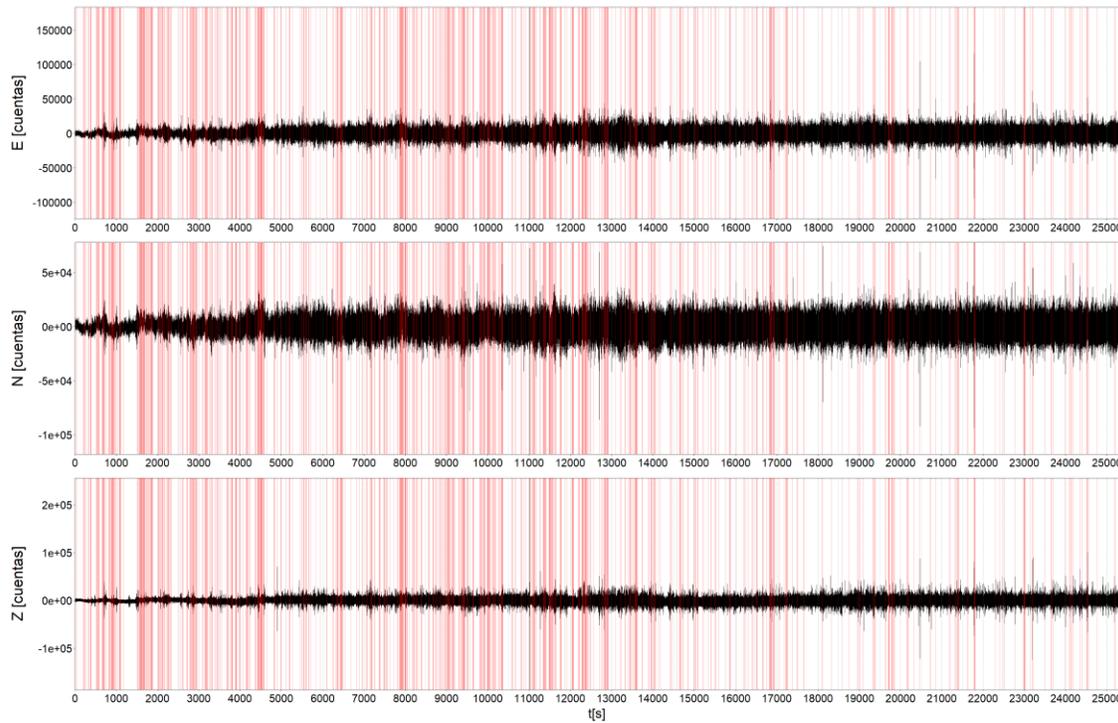


Figura 103: Sismograma de GOLO 325-3. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

GOLO 325-4

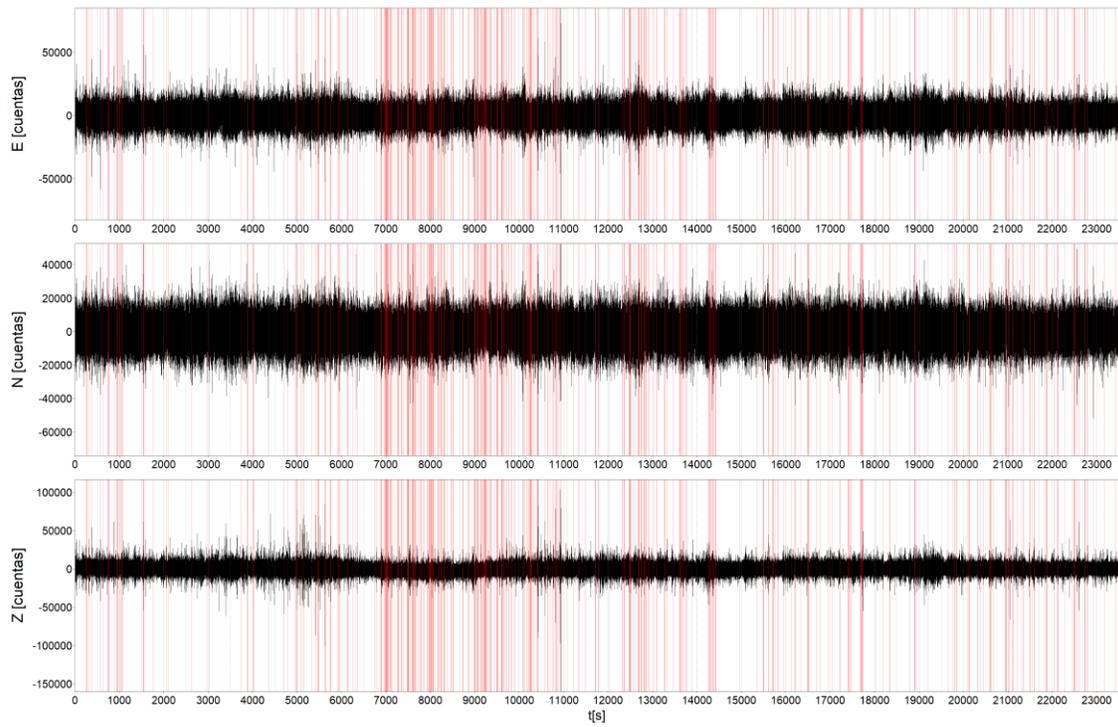


Figura 104: Sismograma de GOLO 325-4. Las ventanas clasificadas como evento al score de corte de 0.75017 se resaltan en color rojo.

Referencias

- Abramson, N. (1981). *Teoría de la información y codificación* (quinta ed.). (J. A. De Miguel Menoyo, Trad.) Madrid, España: Paraninfo.
- Asten, M. W. (1978). Geological Control on the Three-Component Spectra of Rayleigh-Wave Microseisms. *Bulletin of the Seismological Society of America*, 68(6), 1623-1636.
- Asten, M. W., & Henstridge, J. D. (1984). Array estimators and the use of microseisms for reconnaissance of sedimentary basins. *Geophysics*, 49(11), 1828-1837.
- Bonnefoy-Claudet, S., Cotton, F., & Bard, P.-Y. (2006). The nature of noise wavefield and its applications for site effects studies: A literature review. *Earth-Science Reviews*, 79(3-4), 205-227.
- Davis, J., & Goadrich, M. (2006). The Relationship Between Precision-Recall and ROC Curves. *Proceedings of the 23rd International Conference on Machine Learning* (págs. 233-240). Pittsburgh, Pennsylvania, USA; June 25 - 29, 2006: Association for Computing Machinery.
- Frank, E., Hall, M., & Witten, I. (2016). *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"* (Cuarta ed.). Morgan Kaufmann. Obtenido de <https://www.cs.waikato.ac.nz/ml/weka/book.html>
- Gentili, S., & Michelini, A. (2006). Automatic picking of P and S phases using a neural tree. *Journal of Seismology*(10), 39-63.
- Goldstein, P., & Snoke, A. (2005). SAC Availability for the IRIS Community. *Incorporated Institutions for Seismology Data Management Center Electronic Newsletter*.
- Gutenberg, B. (1958). Microseisms. En H. Landsberg, & J. Van Miegheem (Edits.), *Advances in Geophysics* (Vol. 5, págs. 53-92). Nueva York, Nueva York, EE. UU.: Academic Press. Recuperado el 10 de julio de 2017, de [https://doi.org/10.1016/S0065-2687\(08\)60075-8](https://doi.org/10.1016/S0065-2687(08)60075-8)
- Havskov, J., & Alguacil, G. (2004). *Instrumentation in Earthquake Seismology*. Dordrecht, Países Bajos: Springer.
- Hernández Orallo, J., Ramírez Quintana, M. J., & Ferri Ramírez, C. (2004). *Introducción a la Minería de Datos*. Madrid: Pearson Educación, S.A.
- Jones, D. S. (1979). *Elementary Information Theory* (Primera ed.). Oxford, Reino Unido: Oxford University Press.
- Maercklin, N. (2010). *Three-component processing and analysis tools for seismic data in SAC format*.
- Meyer, P. L. (1973). *Probabilidad y Aplicaciones Estadísticas*. México, D.F., México: Fondo Educativo Interamericano.
- Peng, Z. (5 de Noviembre de 2006). *Index of /people/zpeng/Teaching/SAC_Tutorial/MatSAC*. Recuperado el 5 de Diciembre de 2018, de http://geophysics.eas.gatech.edu/people/zpeng/Teaching/SAC_Tutorial/MatSAC

- Peña, D. (2002). *Análisis de datos multivariantes*. Madrid: McGraw-Hill Interamericana de España.
- Peterson, J. (1993). *Observations and modeling of seismic background noise*. Albuquerque, Nuevo Mexico: U.S. Department of Interior, Geological Survey.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical Recipes in C* (Segunda ed.). Cambridge University Press.
- Ross, Z. E., & Ben-Zion, Y. (2014). Automatic picking of direct P, S seismic phases and fault zone head waves. *Geophysical Journal International*(199), 368-381.
- Stein, S., & Wysession, M. (2003). *An Introduction to Seismology, Earthquakes, and Earth Structure*. Malden, Massachusetts, EE. UU.: Blackwell Publishing.
- The MathWorks, Inc. (2012). MATLAB y Signal Processing Toolbox Release 2012a. Natick, Massachusetts, Estados Unidos.
- Trnkoczy, A. (1999). Information Sheet 8.1: Understanding and parameter setting of STA/LTA trigger algorithm. En P. Bormann (Ed.), *New Manual of Seismological Observatory Practice (NMSOP-2)* (Vols. 2 - Annexes). Potsdam, Alemania: IASPEI, GFZ German Research Centre for Geosciences.
- University of Waikato. (2012). CorrelationAttributeEval.java. Hamilton, New Zealand.
- Wikimedia Commons. (junio de 2004). *File:Pswaves.jpg*. Recuperado el 10 de julio de 2017, de <https://commons.wikimedia.org/wiki/File:Pswaves.jpg>
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2017). *Data Mining: Practical Machine Learning Tools and Techniques* (Cuarta ed.). Morgan Kaufmann, an imprint of Elsevier.
- Yoon, C. E., O'Reilly, O., Bergen, K. J., & Beroza, G. C. (2015). Earthquake detection through computationally efficient similarity search. *Science Advances*, 1(11), 1-13.