

Facultad de Ciencias
**Astronómicas
y Geofísicas**
UNIVERSIDAD NACIONAL DE LA PLATA


Tesis de Grado para obtener el título de Geofísico

Detección automática de eventos sísmicos marcianos

Agustín Reynaldo Gómez


Director: Prof. Dr. Danilo Rubén Velis

Firma y aclaración del Director o Codirector:



Danilo R. Velis

Firma y aclaración de uno de los Jurados:



Andrea Cecilia
Durán

La Plata
- 2022 -

A mis viejos, mi hermana y mis abuelas.

Agradecimientos

Si tuviese que ser completamente sincero con los agradecimientos, éstos serían más largos que el trabajo de Tesis. El motivo de esto es que tanta gente me ayudó a lo largo de mi trayecto académico, que se vuelve imposible agradecerle como corresponde. Sin embargo, me veo en la necesidad de aunque sea mencionar a aquellas personas que, al momento de escribir los agradecimientos, están presentes en mi vida.

Agradezco a Danilo, mi director, por todo el apoyo pedagógico y técnico que me dio en mi desarrollo de Tesis. De más está decir que no debió ser fácil ser mi director, y por eso lo que más agradezco es la paciencia que tuvo conmigo.

A Diego Mercerat, por cedernos los datos de Géozur, y por su buena predisposición a responder con lujo de detalle todas mis consultas. Sus aportes fueron fundamentales para este trabajo de Tesis.

A mis juradas Cecilia Durán y Verónica Martínez, por sus aportes de gran calidad para mejorar este trabajo. Su buena predisposición para comentar y corregir este trabajo fue esencial en la última etapa.

A mis compañeres y amigos de la facultad. Aunque quizás ellos no lo tengan tan presente, fue en gran parte por ellos que logré recibirme. Gracias por tantos mates, por tantas cursadas, por tantas jornadas de estudio, por bancarme cuando estaba mal y por celebrar cuando estaba bien. Aunque no lo digo tanto como me gustaría, lo dejo plasmado por escrito: los quiero.

A mis amigos de la vida. Su compañía alegró mi vida todos estos años. Espero algún día devolverles todo lo bueno que me dieron, y todo lo bien que me hicieron.

A la Agrupación Estudiantil Inti, por enseñarme lo que representa la militancia, y por hacerme entender que, más allá de lo que el futuro me depare, siempre seguiré siendo militante.

A mi mamá, a mi viejo, a mi hermana. Gracias infinitas, por todo. Esta tesis va dedicada a ustedes.

Finalmente, agradezco a todes les docentes, nodocentes, estudiantes y graduades que día a día trabajan para que la Educación Pública sea posible.

Gracias.

Índice general

Agradecimientos	1
Resumen	1
1. Introducción y marco teórico	3
1.1. Detección de eventos sísmológicos	3
1.2. Métodos de detección automática de eventos	5
1.2.1. Método de Allen (RAM)	6
1.2.2. Método de Earle-Shearer (ESM)	8
1.2.3. Método de Allen modificado (MAM)	9
2. Preprocesamiento de los datos	11
2.1. Características de los datos	12
2.2. Efecto del ruido en los datos	13
2.2.1. Definición de relación señal/ruido	13
2.2.2. Fuentes de ruido	14
2.3. Filtrado de los datos	14
3. Algoritmo de reducción a una componente	18
3.1. Características generales del método	19
3.2. Maximización de la energía	22
3.2.1. Energía en componentes rotadas	22
3.2.2. Método de <i>Gradient Descent</i>	26
3.2.3. Consideraciones prácticas	27
3.3. Aplicación a sismograma específico	28
3.3.1. Primera rotación	30
3.3.2. Segunda rotación	31

3.3.3. Observaciones de los resultados	32
3.3.4. Detección de eventos	34
3.4. Conclusiones	35
4. Resultados	39
4.1. Detección sin reducción a una componente	40
4.2. Detección con reducción a una componente	45
4.3. Comparación de resultados	46
4.4. Conclusiones	47
5. Conclusiones generales y discusión	50
Bibliografía	52

Resumen

La detección de eventos sísmológicos es muy importante en muchos ámbitos de la Sismología. Poder diferenciar, en un sismograma o en una serie de tiempo que describa el movimiento del suelo, cuándo una señal se asocia a una fase sísmológica y cuándo se asocia a ruido es fundamental para múltiples aplicaciones tanto en sismología global (para la caracterización de coordenadas del foco, hora de origen y magnitud de terremotos), como en aplicaciones sísmicas y microsísmicas (por ejemplo, para la determinación de correcciones estáticas a trazas sísmicas, o para el monitoreo de procesos de fracturación hidráulica).

Para la detección de los eventos, el método históricamente más eficaz es el manual, en el que la detección es realizada por una persona capacitada para reconocer las distintas fases. Sin embargo, la detección manual puede ser subjetiva, y resulta impráctica cuando se busca trabajar con grandes volúmenes de datos, o cuando los eventos que se busca detectar son de baja energía. Por ende, hoy en día también se utilizan algoritmos de detección automática o semiautomática que consideran a la traza como una serie de tiempo digital y aplican estrategias y técnicas nacidas del análisis de señales digitales, que tienen el objetivo de realzar los arribos sísmológicos por sobre todo tipo de ruido no deseado. La eficacia de estos algoritmos suele depender fuertemente de la relación señal/ruido. Cuando la relación señal/ruido es baja, los algoritmos reducen su capacidad de detección ya que los arribos sísmicos se ven enmascarados por la energía del ruido.

En el caso particular de la sismología global, la detección de eventos es crucial para estudiar el interior de la Tierra. Hoy en día, la sismología ha ampliado sus horizontes para estudiar otros cuerpos celestes en el sistema solar. Un ejemplo es el proyecto InSight (Interior exploration using Seismic Investigations, Geodesy and Heat Transport) de la NASA. Este proyecto consistió en la instalación de diversos instrumentos geofísicos en el planeta Marte, entre los cuales se encuentran dos sismómetros de tres componentes.

Para comprender mejor el comportamiento de los sismogramas que se obtendrían de Marte mediante los sismómetros allí instalados, se diseñaron datos sintéticos que modelan los datos reales registrados. Los datos sintéticos incluyen las fuentes de ruido que afectan a la señal, más las magnitudes esperadas por los eventos que puedan ocurrir en Marte. Luego, en el marco del proyecto *Preparing for InSight* (Clinton *et al.*, 2017) se convocó a diversos equipos de investigación a que realizaran una inversión de estos datos, a fin

de detectar los *martemotos* presentes en dicho set de datos, y obtener un modelo de velocidades del planeta. Uno de los grupos de trabajo que participó del proyecto es el Laboratorio de Géoazur, perteneciente a OCA (*Observatoire de la Cote d'Azur*), ubicado en Francia. Este grupo realizó la detección de arribos mediante dos métodos del tipo STA/LTA, con diferentes tamaños de ventana móvil y formas de filtrado, para luego realizar una detección manual.

Una de las características más importantes de este conjunto de datos sintéticos es que, debido a que las condiciones de observación no son las óptimas y a que los eventos sísmicos marcianos son de baja magnitud, su relación señal/ruido es considerablemente más baja que las que pueden llegar a obtenerse en la Tierra, por lo que resultan valiosos para el primer objetivo planteado.

El primer objetivo de este trabajo de Tesis es utilizar estos datos sintéticos de baja relación señal/ruido para estudiar la eficacia de tres métodos de detección automática comúnmente utilizados en la sismología global, para la detección de *martemotos* en estos datos: el método de Allen (Allen, 1978), el de Earle y Shearer (Earle y Shearer, 1994) y el método de Allen Modificado (Sabbione, 2010). La eficacia de los métodos se cuantifica en función de la cantidad de arribos coincidentes con las detecciones automáticas y manuales realizadas por Géoazur.

Los métodos de detección convencionales suelen tomar como dato de entrada un único sismograma. Esto lleva a adoptar una de dos estrategias. Una es utilizar aquella componente cuya relación señal/ruido sea mayor. Sin embargo, generalmente no es posible conocer *a priori* qué componente tiene la mejor relación señal/ruido. La otra alternativa es aplicar los algoritmos a las tres componentes en paralelo, lo que puede dar lugar a ambigüedades en caso de que una detección sea exitosa en una componente pero fracase en otra.

Por lo previamente mencionado, el segundo objetivo de este trabajo de Tesis consiste en desarrollar y aplicar a los datos sintéticos marcianos una estrategia para lidiar con el problema de la variabilidad componente a componente de la relación señal/ruido. La esencia del método es combinar las tres componentes, considerando hipótesis físicas y técnicas matemáticas de la sismología global, para maximizar la energía de la señal en una única componente. Esto trae dos ventajas: por un lado, permite trabajar con un único sismograma; por otro lado, puede mejorar la relación señal/ruido considerablemente, favoreciendo la efectividad de las detecciones. Una vez aplicada esta estrategia, se aplican los métodos de detección sobre el sismograma con la relación señal/ruido maximizada, y se comparan los resultados obtenidos por los mismos métodos en los datos antes de aplicar este algoritmo, para verificar si su capacidad de detección es mejorada con esta técnica. Esta metodología fue presentada por Gómez y Velis (2021).

Capítulo 1

Introducción y marco teórico

1.1. Detección de eventos sismológicos

La propagación de ondas sísmicas generadas por terremotos en el interior o superficie de la Tierra es un fenómeno altamente estudiado por la Geofísica. La información provista por los terremotos detectados por sismómetros estratégicamente dispuestos sobre la superficie terrestre ha sido fundamental para comprender la estructura del interior de nuestro planeta y es, hasta el momento, una de las formas más eficientes de comprender la dinámica y la estructura del interior de la Tierra. Hoy en día, la Sismología ha expandido sus horizontes y ha sido fundamental en el estudio del interior de otros cuerpos del sistema solar, como por ejemplo la Luna (Nunn *et al.*, 2020) o, más recientemente, el planeta Marte.

En general, para entender la propagación de las ondas sísmicas en el marco de la sismología global, se modela al interior del cuerpo como elástico. En un modelo elástico, las ondas que viajan por el interior del medio se denominan ondas internas, y pueden ser del tipo P o compresionales, o del tipo S o transversales. La incorporación de superficies de discontinuidad a los modelos es necesaria para describir las distintas capas que caracterizan la estructura del interior planetario. Las ondas pueden viajar sobre estas superficies de diferentes maneras. Cuando una onda está confinada a la superficie del planeta, se dice que es una onda superficial. A diferencia de las internas, las ondas superficiales tienen un comportamiento físico más complejo, por lo que requieren de tratamientos diferentes para su análisis, pero ambas aportan información crucial (Aki y Richards, 2002).

En el campo de la sismología global, las series de tiempo que describen el movimiento relativo de un punto de la Tierra se denominan *sismogramas* (ver Figura 1.1), y son registrados por instrumentos denominados *sismómetros*. Los sismómetros buscan representar el movimiento de un punto sobre la superficie terrestre respecto de algún sistema de referencia local. Dependiendo del tipo de sensor asociado al sismómetro, el sismogra-

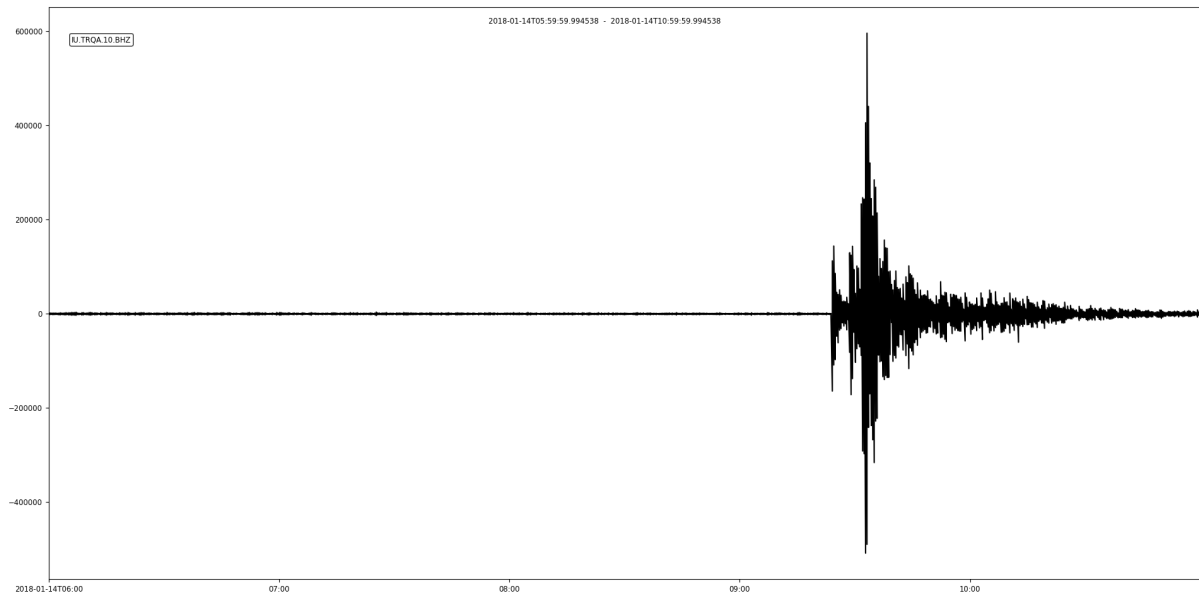


Figura 1.1: Ejemplo de sismograma de la componente Z correspondiente a un sismómetro de 3 componentes y de banda ancha en la estación sismológica de Tornquist TRQA. El evento en cuestión corresponde a la fecha 18 de enero de 2018.

ma resultante puede ser de *movimiento*, de *velocidad* o de *aceleración*. Cuando ocurre un terremoto, las ondas elásticas se propagan por el interior y la superficie del planeta, y esa energía elástica propagada se observa como una perturbación en los sismogramas.

Muchas fuentes pueden generar una respuesta en el sismómetro y pueden ser registradas; algunas naturales y otras artificiales. La señal resultante es la superposición de un gran conjunto de efectos, de los cuales en general solo los arribos sísmicos son de interés para la Sismología y de valor para la interpretación del interior planetario. Un *arribo sísmico* es una perturbación elástica que viaja desde el hipocentro de un terremoto hasta la estación de observación. Los arribos se observan en las series de tiempo, donde, provisto que no son enmascarados por ruido ni obviados, son detectados mediante diversos métodos.

En Sismología, la detección de eventos sismológicos es una de las actividades más importantes. Poder diferenciar, en un sismograma o en una serie de tiempo que describa el movimiento del suelo, cuándo una señal se asocia al arribo de una fase sismológica y cuándo se asocia a ruido es fundamental para múltiples aplicaciones tanto en sismología global (por ejemplo, para la caracterización de coordenadas del foco, hora de origen y magnitud de terremotos), como en aplicaciones sísmicas y microsísmicas (por ejemplo, para la determinación de correcciones estáticas a trazas sísmicas, o para el monitoreo de procesos de fracturación hidráulica).

1.2. Métodos de detección automática de eventos

Se considera que una de las maneras más seguras de detectar eventos sísmológicos es a través de una detección manual. Ésta consiste en la revisión cautelosa de un intérprete capacitado para reconocer los arribos de las distintas fases, y asociar a cada fase un tiempo de arribo. No obstante, este método es subjetivo, y además resulta muy dificultoso de aplicar cuando se trabaja con grandes volúmenes de datos, o cuando los eventos tienen una energía muy baja. Para intentar resolver estas problemáticas, en la actualidad se utilizan métodos de detección automática o semiautomática.

Actualmente, la digitalización y la masividad de las series de tiempo sísmológicas hacen de los métodos de detección automática una posibilidad y una necesidad, al punto tal de que distintos algoritmos de detección están incorporados en las herramientas digitales modernas para el tratamiento de datos sísmológicos.

Los métodos de detección automática de eventos sísmicos son algoritmos que consideran a las series de tiempo sísmológicas como señales digitales a las cuales se le calculan diversos *atributos sísmicos* (Küperkoch *et al.*, 2012). Estos atributos son funciones del tiempo, pero también de la energía y frecuencia de la señal. La idea detrás de ellos es realzar la componente de señal por sobre la del ruido, a fin de poder reconocer mejor los eventos para su posterior detección. En las secciones siguientes se detallará la teoría detrás de algunos de los métodos convencionales más utilizados en el marco de la sísmología global, que fueron seleccionados para su análisis en este trabajo.

En general, los métodos de detección automática se basan en la utilización de atributos sísmicos que dependen del tiempo y de la energía del sismograma. Sobre estos atributos, se consideran dos ventanas temporales móviles de distinta longitud y utilizando algún criterio se comparan los valores dentro de cada ventana; un evento sísmológico es considerado detectado si a partir de dicha comparación se satisfacen ciertas condiciones. En la bibliografía, el concepto de atributo sísmico también recibe el nombre de *función característica*; la función característica es específica del método en cuestión, y es construida de manera que se realcen ciertos parámetros de la señal por sobre los del ruido.

En la práctica, los métodos más utilizados son del tipo STA/LTA (*short term average/long term average*). Éstos utilizan una ventana temporal de larga duración temporal y una ventana de corta duración temporal. Dentro de esas ventanas se calculan los promedios temporales de la función característica y se los compara; si la comparación supera cierto umbral *THR*, entonces se clasifica el instante en el que eso ocurre como un arribo sísmico.

El principal problema con los métodos de detección automática es la dificultad para reconocer la eficacia de los mismos para detectar eventos telesísmicos. Los métodos pueden obtener un resultado falso de dos maneras: pueden no reconocer un arribo, o determinar un falso positivo, marcando el tiempo de un arribo inexistente. Naturalmente, la cantidad de

resultados falsos que se obtienen se correlaciona directamente con la relación señal/ruido: cuando ésta es baja, las fases se vuelven más difíciles de detectar. En ese sentido, uno de los objetivos de este trabajo es analizar la eficacia de algunos métodos del tipo STA/LTA cuando son aplicados a la detección de eventos telesísmicos en condiciones de baja relación señal/ruido.

En este trabajo, se seleccionaron tres métodos denominados *convencionales*. Los métodos convencionales se caracterizan por haber sido diseñados utilizando herramientas vinculadas al análisis de señales digitales (Sabbione, 2010). Estos métodos fueron elegidos de acuerdo a su desempeño para detectar eventos telesísmicos terrestres en otros trabajos (Sabbione *et al.*, 2011). A continuación, se describe la teoría de dichos métodos.

1.2.1. Método de Allen (RAM)

Este método fue diseñado por Allen (1978). En RAM, se construye una función característica CF que contiene información tanto de la energía como de la frecuencia de la señal en cada instante de tiempo. Si x_i es la serie de tiempo digital correspondiente a un sismograma, y N el número de muestras obtenidas,

$$CF_i = x_i^2 + C_i x_i'^2, \quad i = 1, 2, \dots, N, \quad (1.1)$$

donde $x_i' = x_i - x_{i-1}$ es la diferencia sucesiva hacia atrás de la serie de tiempo (una aproximación discreta de la derivada de la señal), y C_i es una función de pesos definida por:

$$C_i = \frac{\sum_{j=1}^i |x_j|}{\sum_{j=1}^i |x_j - x_{j-1}|}. \quad (1.2)$$

El primer término de la ecuación 1.1 contiene información de la energía de la señal (cuadrado de la amplitud), mientras que el segundo término contiene información de la frecuencia de la señal (cuadrado de la aproximación discreta de la derivada temporal). El coeficiente C_i , definido en la ecuación 1.2, modula el término de la frecuencia, de manera que ambos términos terminan teniendo el mismo orden de magnitud. Puede demostrarse que esta función característica CF se aproxima a la función envolvente de x_i (Allen, 1978).

RAM es un método del tipo STA/LTA, es decir calcula un *short-term average* (STA) y un *long-term average* (LTA) de la función característica mediante:

$$STA_i = STA_{i-1} + C_3(CF_i - STA_{i-1}), \quad (1.3)$$

$$LTA_i = LTA_{i-1} + C_4(CF_i - LTA_{i-1}). \quad (1.4)$$

Notar que las ecuaciones 1.3 y 1.4 representan procesos autorregresivos, y pueden interpretarse como filtros pasa-bajos aplicados a la función característica CF_i . La serie STA_i , como representa cortos períodos de tiempo, debería ser capaz de capturar frecuencias más altas que la asociadas a la serie de tiempo LTA_i . Las constantes C_3 y C_4

son parámetros arbitrarios (definidos por el intérprete) para modular el tamaño de las ventanas correspondientes a los promedios STA y LTA.

El algoritmo declara un evento en la i -ésima muestra de la serie de tiempo cuando el atributo r_i , definido como

$$r_i = \frac{STA_i}{LTA_i} \quad (1.5)$$

satisface que

$$r_i > THR, \quad (1.6)$$

donde THR es otro parámetro a definir por quien esté implementando el método. Este parámetro recibe el nombre de *umbral* (en inglés, *threshold*), y su valor dependerá sustancialmente del comportamiento del atributo r_i , el cual a su vez dependerá de cómo se modulen los promedios STA_i y LTA_i mediante los parámetros C_3 y C_4 .

La Figura 1.2 muestra una serie de tiempo sintética muy simple, con una ondícula de Ricker de fase mínima con ruido blanco incoherente de baja amplitud, a fin de ilustrar cómo funciona el método de Allen gráficamente. La selección del umbral es fundamental, puesto que un valor muy bajo puede derivar en detecciones falsas, y uno muy alto en que no se detecte ningún evento.

Uno de los problemas del algoritmo recién descrito es que no permite reconocer adecuadamente el inicio de un evento, ya que es posible que el atributo supere el umbral en un rango de muestras. Esto significaría que cada una de esas muestras representa un evento distinto. Para sortear este inconveniente, se define también un criterio de *comienzo* del evento, es decir, a partir de qué muestra comienza el evento. Para ello, se define que el evento comienza en la muestra i -ésima si se cumple que

$$r_i > THR \text{ y } r_{i-1} \leq THR.$$

Este criterio será utilizado nuevamente en el método de Earle-Shearer (Sección 1.2.2) y en una modificación al método de Allen, que se denomina método de Allen modificado (Sección 1.2.3).

Un problema presente en todos los métodos del tipo STA/LTA utilizados en este trabajo es el *efecto de borde* generado por el cociente STA/LTA. Éste consiste en un aumento drástico del STA/LTA al inicio de la serie de tiempo, el cual aumenta en amplitud para luego decaer y oscilar alrededor del valor 1. Este aumento inicial en ocasiones conduce a falsas detecciones, lo cual se desea evitar. Para mitigar este problema se utiliza una ventana de Hanning que atenúa los valores iniciales del STA/LTA. Esto implica agregar un parámetro extra T_{taper} , que representa la longitud de la ventana desde el inicio de la serie de tiempo que se desea atenuar. La atenuación consiste simplemente en multiplicar esta ventana inicial con la rama creciente de la ventana de Hanning.

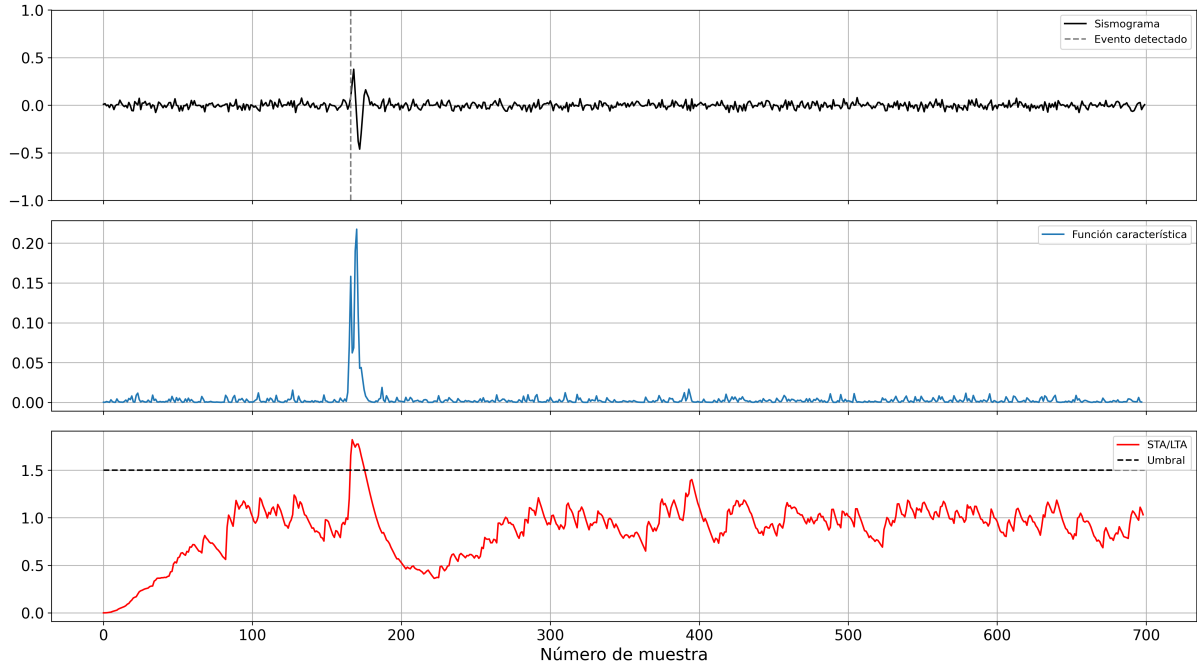


Figura 1.2: Ejemplo de detección mediante el método de Allen, a partir de una traza sintética construida utilizando una ondícula de Ricker equivalente de fase mínima.

1.2.2. Método de Earle-Shearer (ESM)

Este método es más moderno que RAM, siendo también del estilo STA/LTA (Earle y Shearer, 1994). En este caso, el método se basa en la envolvente de la señal, definida como

$$e_i = \sqrt{x_i^2 + \hat{x}_i^2}, \quad (1.7)$$

donde \hat{x}_i es la transformada de Hilbert de la señal.

Luego de calcular la envolvente de la señal, se calcula el cociente STA/LTA mediante los parámetros T_{STA} y T_{LTA} , siendo éstos la longitud de la ventana corta y larga respectivamente. Matemáticamente, los promedios STA y LTA se definen de la siguiente manera:

$$STA_i = \frac{1}{N_{STA}} \sum_{j=i}^{i+N_{STA}-1} e_j, \quad (1.8)$$

$$LTA_i = \frac{1}{N_{LTA}} \sum_{j=i-N_{LTA}}^{i-1} e_j, \quad (1.9)$$

donde N_{STA} y N_{LTA} representan la longitud en muestras de las ventanas de corto y largo período respectivamente. Luego se calcula el cociente

$$r_i = \frac{STA_i}{LTA_i}. \quad (1.10)$$

Este valor es algo inestable, por lo que el atributo es suavizado mediante una convolución con una ventana de Hanning de longitud temporal T_{Hann} . El atributo al cual se aplica el criterio de detección es el cociente STA/LTA filtrado, r_i^f .

La Figura 1.3 refleja el comportamiento de la envolvente de la señal y del atributo STA/LTA filtrado para el mismo evento sintético utilizado en el ejemplo anterior.

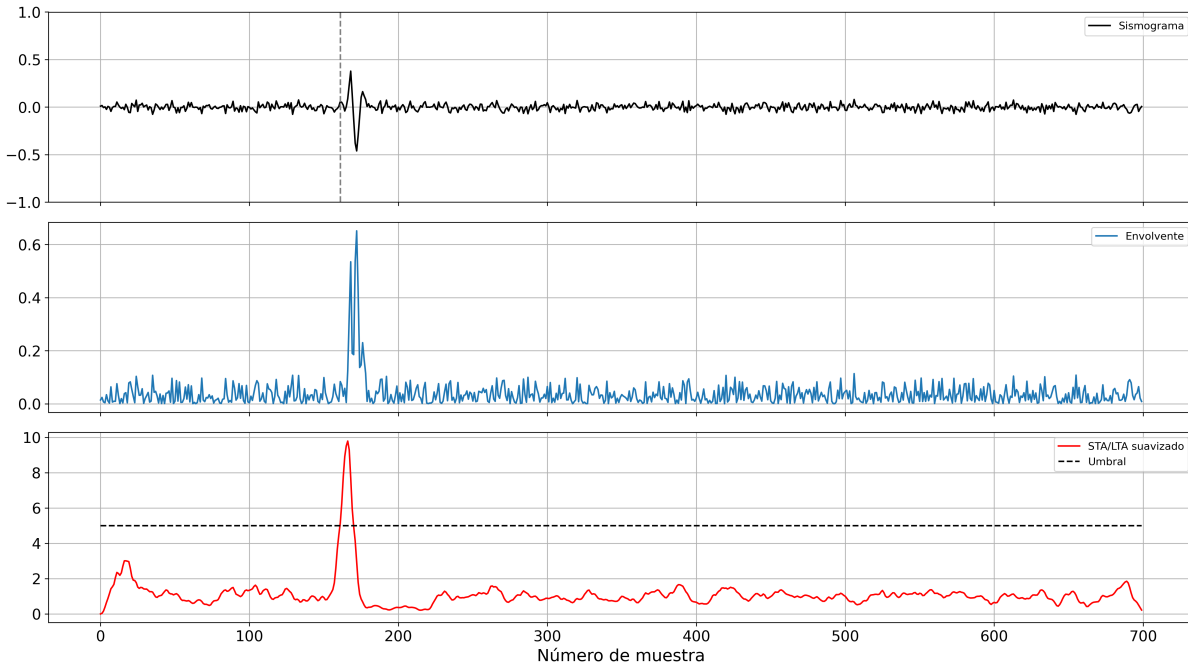


Figura 1.3: Ejemplo de detección mediante el método de Earle-Shearer, a partir de una traza sintética construida utilizando una ondícula de Ricker equivalente de fase mínima.

Al igual que con el método de Allen, se define un último parámetro THR como el umbral: si en la i -ésima muestra $r_i^f > THR$ y $r_{i-1}^f \leq THR$, se asigna a dicha muestra un arribo.

1.2.3. Método de Allen modificado (MAM)

Como su nombre indica, este método es una versión modificada del método presentado en la Sección 1.2.1, y cuya modificación fue propuesta en el trabajo realizado por Sabbione (2010). En MAM, la función característica utilizada es la misma que en el método original definida en 1.1. La modificación ocurre a la hora de calcular los promedios STA y LTA. En este algoritmo, al igual que en el método de Earle-Shearer, se calculan los promedios utilizando una ventana móvil de longitudes temporales T_{STA} y T_{LTA} . Ésto representa una diferencia sustancial respecto a cómo son calculados los promedios en el método de Allen tradicional (ver ecuaciones 1.3 y 1.4). Entonces, los promedios STA y LTA quedan definidos de la siguiente manera:

$$STA_i = \frac{1}{N_{STA}} \sum_{j=i}^{i+N_{STA}-1} CF_j, \quad (1.11)$$

$$LTA_i = \frac{1}{N_{LTA}} \sum_{j=i-N_{LTA}}^{i-1} CF_j, \quad (1.12)$$

De manera análoga a la estrategia utilizada en el método de Earle-Shearer, al atributo $r_i = STA_i/LTA_i$ se lo convoluciona con una ventana de Hanning a fin de suavizarlo, y luego se aplica el criterio ya previamente descrito: se asigna a la i -ésima muestra un arribo si el atributo r_i suavizado supera un determinado umbral THR , y $r_{i-1} < THR$.

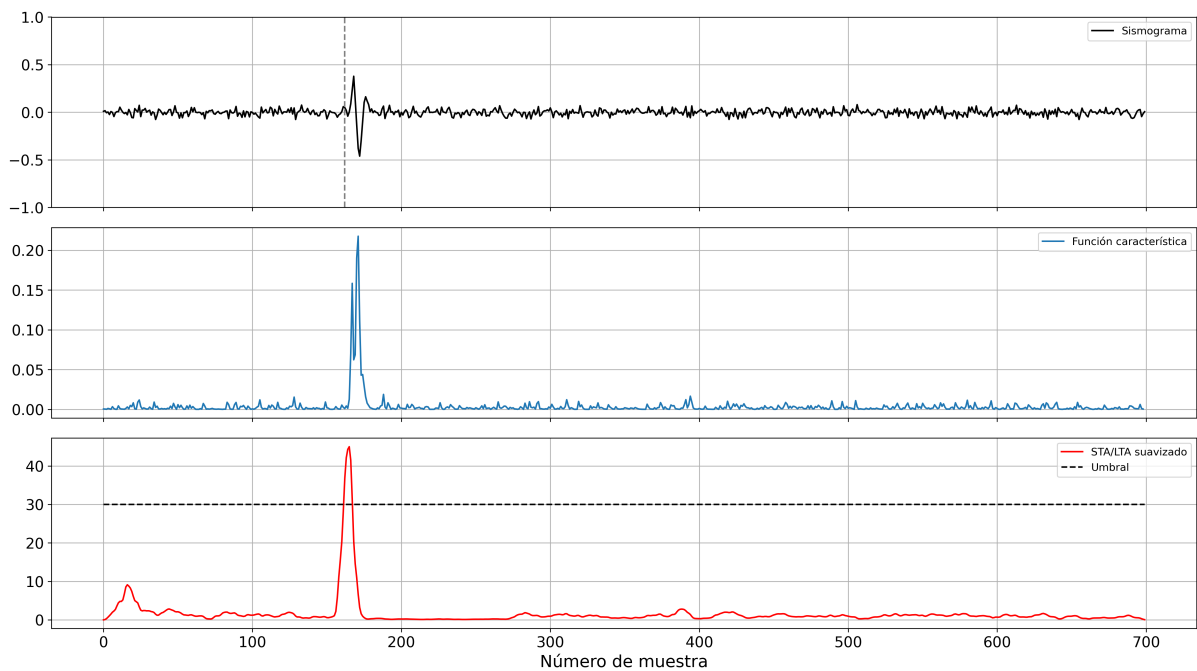


Figura 1.4: Ejemplo de detección mediante el método de Allen modificado, a partir de una traza sintética construida utilizando una ondícula de Ricker equivalente de fase mínima.

La Figura 1.4 sintetiza el comportamiento de la función característica y del atributo suavizado. El arribo se asocia a la primera muestra que supera el umbral. Notar que la función característica es la misma que en el método de Allen, pero la forma del atributo STA/LTA es completamente distinta.

Capítulo 2

Preprocesamiento de los datos

Como se ha mencionado en oportunidades previas, uno de los objetivos de este trabajo es reconocer cómo se comportan ciertos métodos de detección automática de eventos sísmicos en condiciones de relación señal/ruido baja. Para esto, es necesario contar con datos de actualidad sísmológicos que cuenten con una relación señal/ruido pobre. Un conjunto de datos que cumple con estas condiciones es el provisto por la misión InSight (Interior exploration using Seismic Investigations, Geodesy and Heat Transport), organizado por NASA desde el año 2017 hasta la actualidad (Banerdt *et al.*, 2013).

El proyecto InSight consiste en la instalación de varios instrumentos geofísicos para poder estudiar el interior del planeta Marte. Los instrumentos fueron instalados en la superficie del planeta por el *InSight Lander*, y pueden verse en la Figura 2.1. Estos instrumentos son:

- un conjunto de termómetros denominado HP³ para medir la temperatura y el flujo de calor saliente;
- un conjunto de antenas emisoras en ondas de radio y rayos X denominado RISE, con el objetivo de medir movimientos propios del planeta (como nutación y movimiento del polo);
- un conjunto de dos sismómetros de tres componentes con alta ganancia denominado SEIS.

El proyecto SEIS es el de principal interés para nuestro trabajo. Más específicamente, este trabajo se focaliza en el sismómetro de tres componentes, de banda de medio período, y sensor de velocidades allí instalado. Más información acerca de este proyecto enmarcado en la misión InSight puede encontrarse en Lognonné *et al.* (2019). Para entender mejor el comportamiento de los datos reales que se obtendrían del SEIS, se creó el proyecto *Preparing for InSight*, dentro del cual, por motivos que se expresarán en secciones

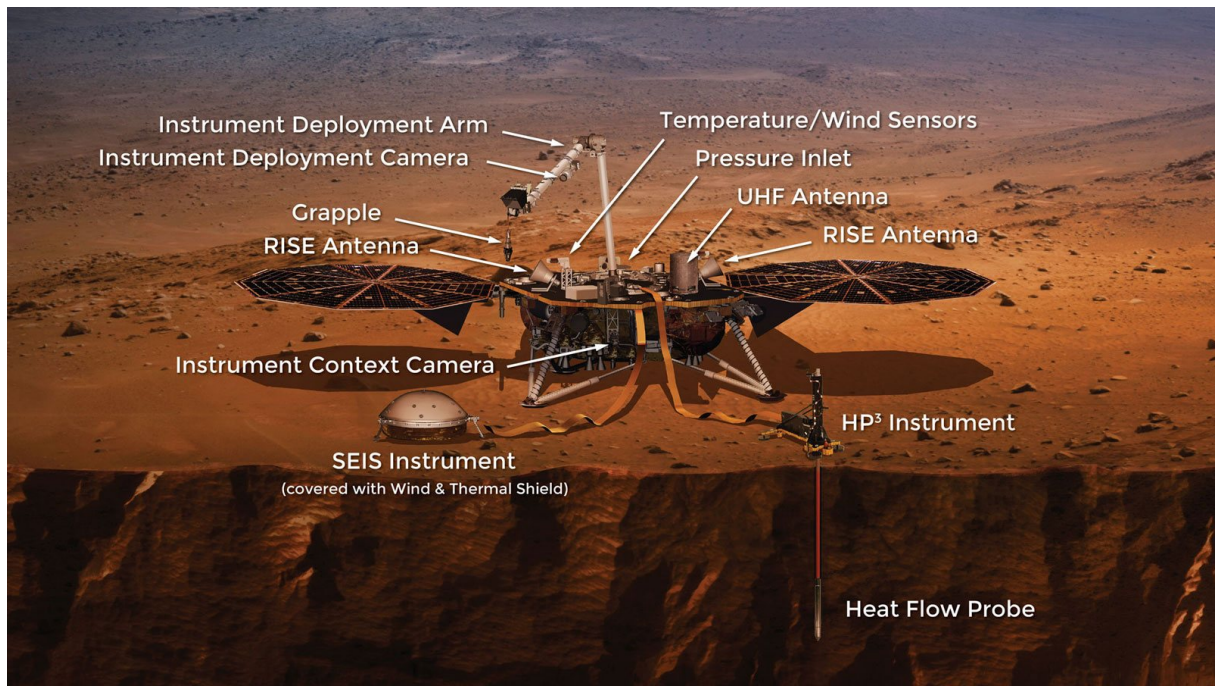


Figura 2.1: Imagen conceptual de los dispositivos instalados en Marte mediante la misión InSight. Créditos de la imagen: NASA.

posteriores, generó datos sintéticos que modelan las series de tiempo que registrarían los sensores SEIS; éstos son los datos que se utilizan en este trabajo.

2.1. Características de los datos

Los datos sintéticos fueron generados a partir de un modelo de velocidades 1D de Marte, y basándose en un catálogo que involucra tanto eventos tectónicos como impactos de meteoritos, y se realizaron en el marco del proyecto *Preparing for InSight* (Clinton *et al.*, 2017). Este proyecto consistió en otorgar a diversos participantes el set de datos y que, a partir de éstos, se pudiera determinar tanto el catálogo de eventos como el modelo de velocidades. La razón de esto es poder verificar de manera objetiva la eficacia de las metodologías utilizadas por distintos equipos dedicados a la sismología global cuando se trabaja con datos con las características de los sismogramas marcianos. En la actualidad, y desde el año 2018, se cuenta con datos reales que están disponibles en los catálogos sísmicos del proyecto SEIS. Sin embargo, se optó por utilizar los datos sintéticos ya que de esa forma se hace posible la comparación con los resultados obtenidos por otros grupos de investigación, como por ejemplo los de Géoazur, cuyos resultados serán descritos en secciones posteriores.

Tanto la sismicidad de Marte como las diversas fuentes de ruido esperadas en los registros fueron consideradas para la generación de los datos sintéticos. Con respecto a la sismicidad marciana, ésta es varias magnitudes menor a la observada en la Tierra, por el simple hecho de que Marte es un planeta tectónicamente inactivo. Para el caso de los

datos sintéticos, se consideró un catálogo de fuentes sísmicas, tanto tectónicas como asociadas a impactos de meteoritos, cuyas posiciones del foco, horas de origen y magnitudes momento (cuyos valores mínimo y máximo son $M_w = 2.5$ y $M_w = 5.0$ respectivamente) son aleatorias. Tanto el catálogo de impactos como el tectónico tiene una distribución temporal de un año terrestre, asignado arbitrariamente al año 2019.

Las formas de onda resultantes resultan de la resolución de la ecuación de onda considerando a Marte como un medio viscoelástico y con simetría axial. Los sismogramas resultan de la superposición de dichas formas de onda con las diversas formas de ruido esperables en los sismómetros instalados.

Los sismogramas sintéticos son presentados como datos provenientes de un sensor de velocidades triaxial Streckeisen STS-2, que registra en tres canales MHU, MHV, y MHV con la unidad en cuentas y con una frecuencia de Nyquist de 1.0 Hz. Si bien la deconvolución instrumental de los datos es posible, en principio en la banda de registración se asume que la función de transferencia es una constante (Clinton *et al.*, 2017). Por lo tanto, en esta Tesis la conversión a unidades de metros por segundo se realizó simplemente normalizando los datos para que éstos se encuentren en el rango de amplitudes $[-1,1]$.

2.2. Efecto del ruido en los datos

2.2.1. Definición de relación señal/ruido

En general, la relación señal/ruido de las series de tiempo marcianas es más baja que la obtenida en estaciones terrestres. Se entiende a una serie de tiempo como una componente de señal y otra componente de ruido, es decir, una serie de tiempo es la suma directa de una amplitud de señal deseada (el dato que se busca medir) y de ruido (aleatorio o coherente) que dificulta la observación directa del dato. Si bien existen muchas definiciones de relación señal/ruido, en este trabajo se adopta la siguiente:

$$SNR = \frac{E_s}{E_r}, \quad (2.1)$$

donde E_s es la energía de la señal, y E_r es la energía del ruido superpuesto a la señal. La energía de una serie de tiempo x_i se define como

$$E_x = \sum_i x_i^2. \quad (2.2)$$

La relación señal/ruido es función de múltiples aspectos que caracterizan a los terremotos, o *martemotos* para el caso de esta Tesis, como su magnitud, distancia epicentral, acimut, ángulo de incidencia, profundidad, la geología regional por la que el rayo viaja y la geología local en la cual la estación está colocada. También depende de la instalación de la estación de observación, entre otros factores.

2.2.2. Fuentes de ruido

Las fuentes de ruido que afectan los sismogramas marcianos son muy variados tanto en su contenido en frecuencia como en su energía, y pueden separarse en ruido ambiente y en ruido asociado a los sensores en sí mismos. En particular, el ruido ambiente se subdivide en aquél asociado a efectos que afectan al instrumento (cambios de temperatura, campo magnético, etcétera), y en aquél que provoca aceleraciones en la superficie (variaciones de presión y viento). Las fuentes que se tuvieron en consideración para el modelado del ruido son:

- Variaciones de presión atmosférica calculadas a partir de modelos globales meteorológicos;
- Variaciones en el campo de vientos, que producen tensiones sobre el domo protector que recubre los sismómetros y que resultan en señales con energía en la banda de registración;
- Variaciones de temperatura, que afectan la sensibilidad de las componentes de los sismómetros;
- Ruido asociado a variaciones magnéticas por el efecto del viento solar;
- Campos eléctricos generados por transporte de material en la superficie, que provoca efectos triboeléctricos;
- Ruido instrumental, que en la banda de 0.5 Hz - 1.0 Hz tiene un comportamiento browniano.

El ruido ambiente en particular tiene un efecto que resulta de gran relevancia en los datos. La razón de esto es que los instrumentos son instalados en superficie, sobre una capa superficial suave (Murdoch *et al.*, 2017). Esto presenta una importante diferencia con respecto a los datos obtenidos en sismómetros en la Tierra, donde las instalaciones sismológicas se dan sobre estructuras generalmente construidas en el subsuelo, con el objetivo de aislar el instrumental del ruido atmosférico y de la capa superficial. En los instrumentos SEIS, las fuentes de ruido ambiente de mayor amplitud media cuadrática son las asociadas a variaciones de temperatura y presión (Mimoun *et al.*, 2016).

2.3. Filtrado de los datos

En esta sección se describe todo el proceso de filtrado de los datos tras un análisis del espectro de frecuencias de los datos.

La Figura 2.2 es un ejemplo no representativo del comportamiento de los sismogramas cuando contienen un sismo. En general, las condiciones de los sismogramas son de baja

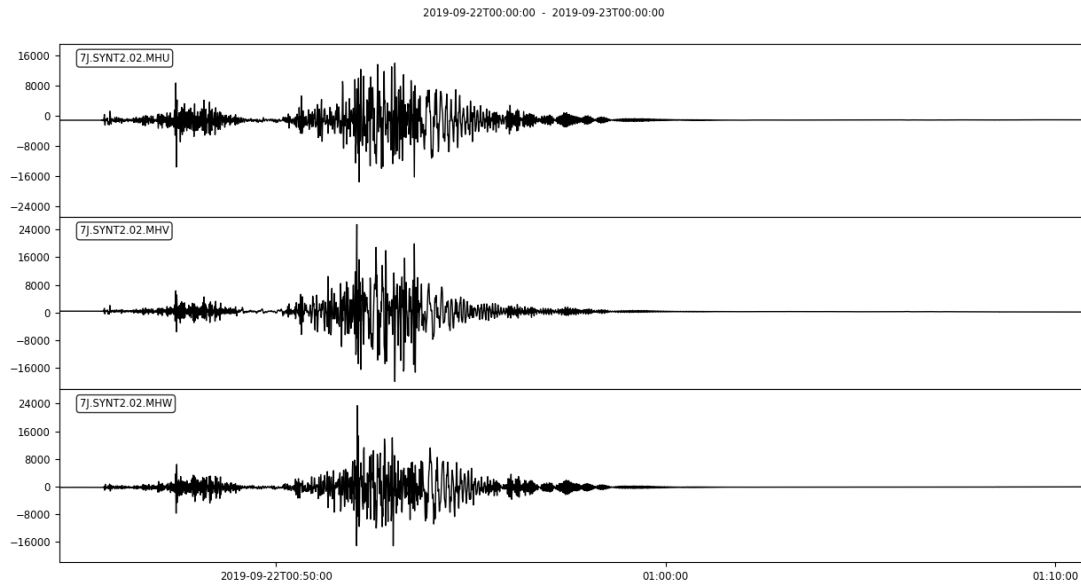


Figura 2.2: Ejemplo de eventos sísmicos sintéticos marcianos con buena relación señal/ruido, visibles sin preprocesamiento ni filtrado necesario.

relación señal/ruido, donde incluso ésta puede ser menor a la unidad. Por ello resulta necesario un filtrado de las señales para remover el ruido presente en los datos.

En general, el ruido ambiente tiene una componente de gran energía en los altos períodos, haciendo que la señal se vea enmascarada por estos ruidos. De manera análoga, el ruido ambiente y el ruido instrumental generan contenido de alta frecuencia, cercana a la frecuencia de Nyquist. La Figura 2.3 muestra un ejemplo sintético típico, donde en general no es posible reconocer la presencia de un evento sin un filtrado previo de las bajas frecuencias.

Se decidió aplicar a las señales un filtro pasa-altas del tipo Butterworth, de orden 4, de fase cero y con frecuencia de corte en 0.2 Hz. Se decidió no filtrar las altas frecuencias para evitar la pérdida de energía de alta frecuencia de la señal, lo que dificultaría la detección. Para mitigar efectos de pérdida de resolución espectral, es importante que la longitud de las señales sea lo más grande posible. En ese sentido, las señales fueron recortadas en porciones de 24 horas de longitud, para poder incorporar los efectos de ruido coherente de período diario.

La Figura 2.4 muestra los mismos datos de la Figura 2.3, luego de aplicado el filtro pasa-altas. En este caso, se observa la presencia de un potencial evento a las 15 horas, muy difícil de identificar en los datos sin filtrar.

Para el filtrado se aplicó la siguiente metodología:

1. Se removió la media y la tendencia lineal. En el espectro de frecuencias, esto equivale a remover la componente continua y los ruidos de muy baja frecuencia (mayor a un período diario);

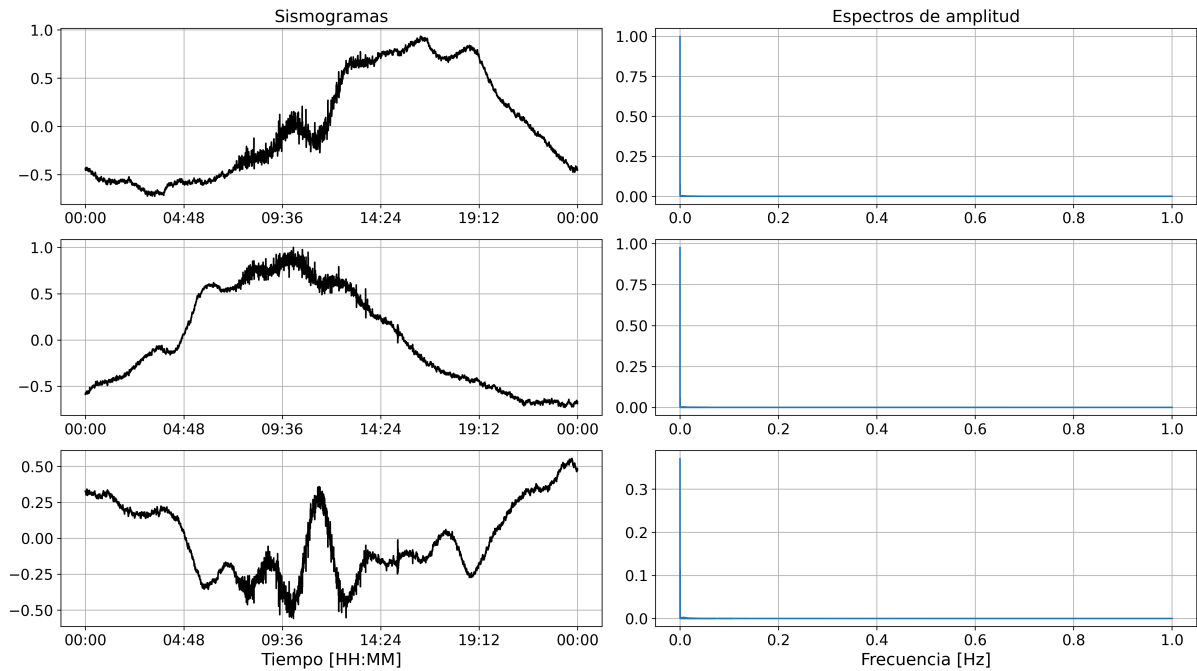


Figura 2.3: Ejemplo de los sismogramas con un día de longitud, con sus respectivos espectros de amplitud normalizados. De arriba a abajo, las componentes U , V y W . Notar la alta energía para bajas frecuencias.

2. Se aplicó una ventana de Hanning en los tiempos iniciales y finales para evitar efectos de *ripple* ocasionados por el truncamiento;
3. Se aplicó el filtro en frecuencia.

La elección de la frecuencia de corte fue hecha mediante un simple método de prueba y error, considerando eventos visibles y observando de qué manera la amplitud máxima de algunos eventos variaba en relación a la del ruido. Como se verá en secciones posteriores, la frecuencia de corte utilizada en el filtrado coincide con la utilizada por Géoazur.

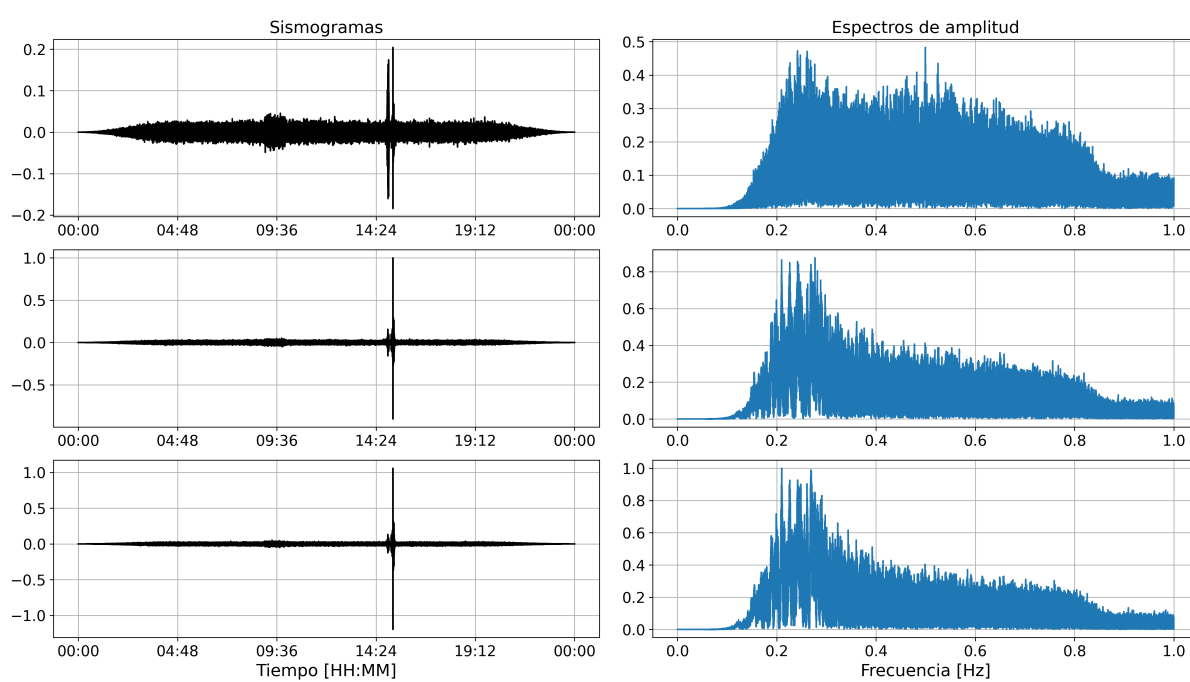


Figura 2.4: Sismogramas de la Figura 2.3 luego del filtrado. Notar la presencia de un posible evento alrededor de las 15 horas.

Capítulo 3

Algoritmo de reducción a una componente

El procesamiento de los datos, que consiste principalmente en la remoción de bajas frecuencias para eliminar gran parte del ruido coherente e incoherente presente en los sismogramas, es una etapa necesaria pero no siempre suficiente para optimizar la relación señal/ruido de los datos. Esto se debe en gran parte a dos motivos, descritos en mayor profundidad en la Sección 2.2:

1. las fuentes de ruido generan ruido incoherente de relativamente alta amplitud en todo el ancho de banda, haciendo que sea imposible que su atenuación sea adecuada sin perder resolución espectral de los eventos;
2. la magnitud de los eventos sísmicos marcianos es significativamente menor a la terrestre, haciendo que los efectos de atenuación tengan mucha mayor relevancia. Lo que es más, estos efectos generalmente no pueden ser removidos adecuadamente, principalmente debido a que, como solo se cuenta con una estación en todo el planeta, la estimación *a priori* de los hipocentros no puede realizarse, por lo cual la distancia al foco, esencial para remover efectos de atenuación, es generalmente una incógnita.

La combinación de estos problemas provoca que la relación señal/ruido en las componentes del sensor sea lo suficientemente pequeña como para dificultar o impedir que los eventos sean detectados en alguna y menos aún en todas las componentes.

Por otro lado, cada componente del sismómetro tiene una relación señal/ruido que en principio depende de la orientación del mismo con respecto a la dirección de propagación y la dirección del evento. Generalmente no se cuenta con información de qué componente es conveniente utilizar para garantizar una óptima eficacia a la hora de aplicar los métodos de detección. Esto lleva, en principio, a aplicar los métodos a cada componente por separado, lo que dificulta y alarga el proceso de detección.

En ese sentido, se plantea la posibilidad de un algoritmo que precondicione los datos, de manera que la relación señal/ruido mejore y que solo sea necesario trabajar con una componente en vez de con tres, para independizarnos de la polarización y la posición del foco, y para reducir las incertidumbres que puedan generarse por la detección en cada componente. Es decir, se busca una forma de construir un sismograma que incorpore información de las tres componentes, de manera que la energía de la señal aumente por sobre la del ruido (Gómez y Velis, 2021). En resumen, el método propuesto tiene dos objetivos fundamentales:

- mejorar de manera significativa la relación señal/ruido de los eventos sísmicos, a fin de que éstos puedan ser detectados con mayor facilidad, y
- permitir trabajar con un único sismograma en vez de con un sismograma por componente.

3.1. Características generales del método

En un medio elástico las ondas internas se caracterizan, entre otras cosas, por estar linealmente polarizadas. Esto implica que la energía cinética de las partículas se distribuye en una única dirección. Para el caso de las ondas P, las amplitudes se dan en la dirección de propagación; para el caso de las ondas S, en la dirección perpendicular a la de propagación. En ese sentido, las ondas internas presentan una diferencia fundamental respecto de las ondas superficiales, que presentan una polarización elíptica (Kulháněk, 2012). Esto es fundamental porque permite, provisto que se conocen el acimut y la inclinación de un evento sísmico dado, diferenciar las fases internas (onda P, SV y SH) y las fases de onda superficial (ondas Rayleigh y Love). Un esquema que muestra el comportamiento de las ondas mientras viajan por el interior del planeta se observa en la Figura 3.1.

Los eventos sísmicos marcianos no son excepción. La Figura 3.2 muestra un ejemplo correspondiente a un evento marciano sintético, donde los *scatter plot* muestra la posición de la masa del sensor para distintos tiempos. En este ejemplo, la existencia de ondas linealmente polarizadas es clara, como lo indica la distribución de las amplitudes.

Una consecuencia importante de la polarización lineal de las ondas es que, para cada evento, existe una única dirección, determinada por un acimut y una inclinación respecto del sistema de referencia dado por la orientación del sismómetro, para la cual la energía de las ondas compresionales es máxima y la de las ondas transversales mínima.

En ese sentido, para un dado evento, se puede orientar las componentes (U, V, W) del sismómetro, rotándolo para diferentes acimuts e inclinaciones, hasta que la energía del mismo sea máxima en una componente y mínima en las demás.

En resumen, la metodología propuesta involucra, de manera general, las siguientes etapas:

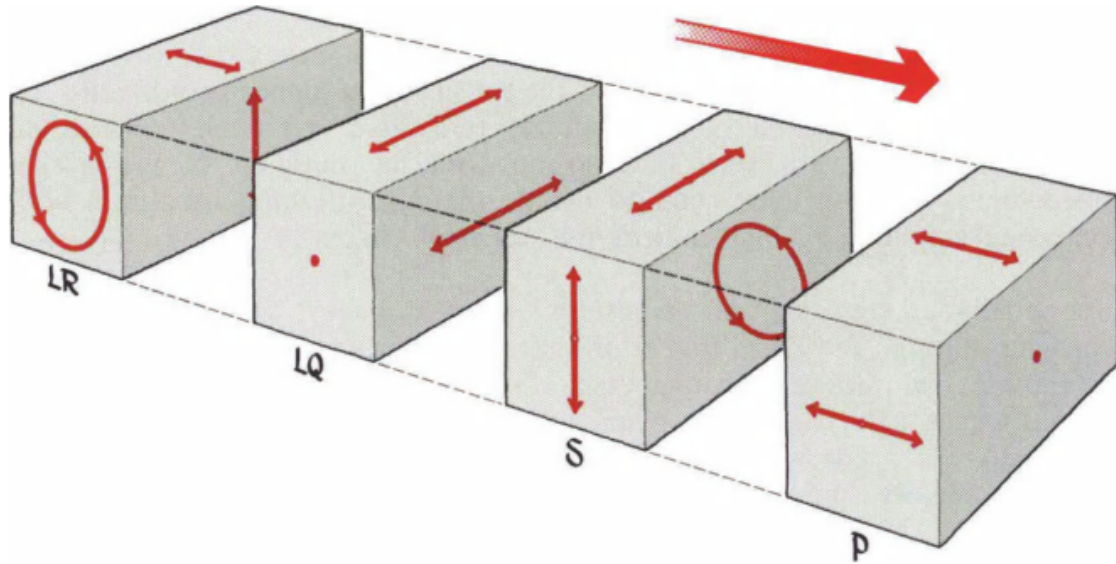


Figura 3.1: Esquema con el movimiento del suelo para las fases de onda interna y las fases de onda superficial. Imagen obtenida de Kulháněk (2012).

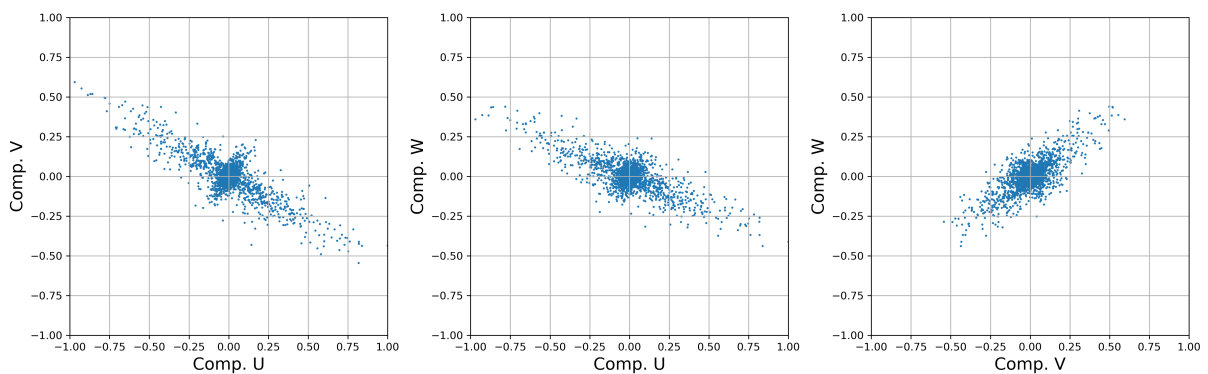


Figura 3.2: Ejemplo de evento sísmico marciano. Los *scatter plot* representan la trayectoria de la masa del sismómetro para distintos tiempos en presencia de un evento. Notar la polarización lineal, marcada por la tendencia lineal de la velocidad de la partícula.

1. Considerar un evento sísmico, asumiendo en principio que contiene fases linealmente polarizadas, de longitud δt en tiempo, o N_w en muestras.
2. Seleccionar arbitrariamente una componente y calcular en dicha componente la energía del evento.
3. A partir de las tres componentes del sismómetro, seleccionar un ángulo (en un rango de 0 a 2π) y una inclinación (en un rango de 0 a π) y rotar la componente seleccionada en la dirección determinada por los ángulos elegidos.
4. Con la componente rotada, recalcular la energía.
5. Repetir los pasos 3 y 4 para todos los ángulos y para todas las inclinaciones.
6. Seleccionar los ángulos para los cuales la energía del evento es máxima, que se denominan ángulos óptimos.
7. Rotar la componente seleccionada respecto del ángulo e inclinación óptimos.

Esta metodología garantiza que, finalizados los pasos previamente descritos, toda la energía calculada para ese evento esté contenida en la componente seleccionada. Repitiendo esto para todos los eventos, lo que se obtiene es un único sismograma, construido a partir de la información obtenida de las tres componentes, con la energía de los eventos maximizada.

En este punto es importante marcar dos cuestiones importantes:

- Este sismograma resultante es una construcción matemática basada en la hipótesis de polarización lineal, que altera la forma, amplitud, ángulo e inclinación de la fase sísmica, por lo que debe interpretarse como una herramienta auxiliar para los métodos de detección.
- Así planteado, este método no asegura su eficacia en presencia de ruido incoherente. De esto se hablará en secciones posteriores, donde también se explicará por qué este método (con modificaciones adecuadas) permite maximizar la energía de los eventos pero no la del ruido.

En este trabajo, la componente en la cual la energía se decidió maximizar es la componente W , o "vertical". En ese sentido, y con el objetivo de hacer más amena la explicación de los detalles del método, se hablará de la componente vertical cuando se haga referencia a la componente de energía máxima.

3.2. Maximización de la energía

En la práctica, el método de reducción a una componente se basa en el problema de maximizar la energía en función de dos ángulos, es decir es un problema de optimización. En general, se espera que para un dado evento, la dirección de propagación sea única, con lo que es esperable que la energía del evento como función del ácimut y de la inclinación contenga un único extremo máximo. Ésto implica que el problema de optimización serán de caracter local, lo que simplifica su resolución considerablemente.

Matemáticamente, la maximización de la energía puede descomponerse en dos rotaciones consecutivas (ver Figura 3.3 para una descripción gráfica de la descomposición en dos rotaciones):

1. Rotar respecto de la componente *vertical* W hasta que la energía de la componente U sea máxima o, de manera equivalente, que la energía de la componente V sea mínima. Esto garantiza que, luego de la primera rotación, toda la energía del evento esté contenida exclusivamente en el plano (U, W) ;
2. rotar respecto de la componente *horizontal* V (cuya energía fue minimizada en el paso anterior) hasta que la energía de la componente W sea máxima, o equivalentemente que la energía en U sea mínima.

Esto permite separar al problema único de optimización 2D en dos problemas de optimización 1D, minimizando la energía de la componente V en el primer paso, y la de la componente U en el segundo paso.

3.2.1. Energía en componentes rotadas

Se considera en primer lugar una rotación sobre el plano (U, V) de un ángulo ϕ . Este ángulo es análogo al ácimut. La componente $V(\phi)$ rotada puede calcularse para la i -ésima muestra como:

$$V_i(\phi) = U_i \sin \phi + V_i \cos \phi, \quad (3.1)$$

con lo que la energía de la componente V , expresada en términos del ácimut ϕ queda dada como:

$$E_V(\phi) = \sum_i (U_i \cos \phi + V_i \sin \phi)^2. \quad (3.2)$$

De manera análoga, si se considera una rotación de un ángulo θ (análogo a la inclinación) sobre el plano (U, W) la componente $U_i(\theta)$ puede expresarse como:

$$U_i(\theta) = U_i \cos \theta + W_i \sin \theta. \quad (3.3)$$

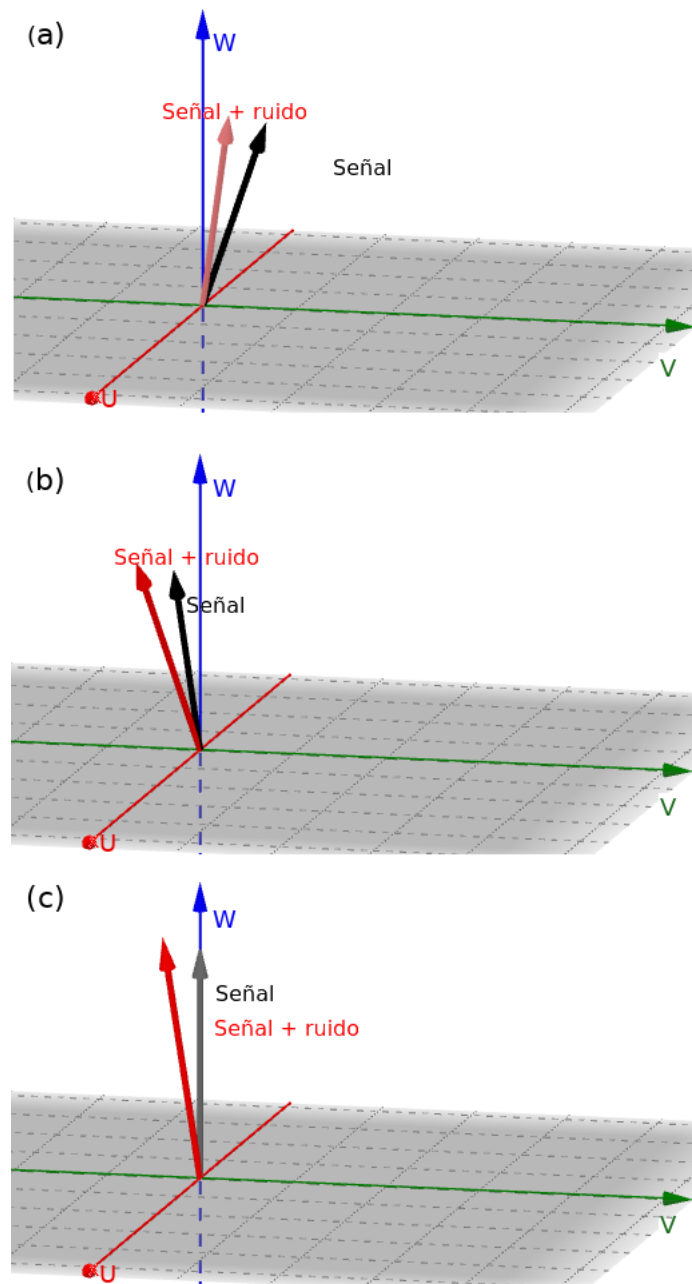


Figura 3.3: Esquema que resume la forma en la que se rota la señal en dos etapas. (a) Señal (negro) y la misma señal contaminada con ruido (rojo). (b) Luego de la primera rotación, la energía queda contenida en el plano (U, W) . La presencia de ruido afecta la rotación, habiendo un poco de energía todavía en la componente V . (c) Luego de la segunda rotación, la señal contaminada no está completamente proyectada sobre el eje W , por la presencia de ruido.

Luego, la energía de la componente U expresada como función de θ es:

$$E_U(\theta) = \sum_i (U_i \cos \theta + W_i \sin \theta)^2. \quad (3.4)$$

Las definiciones dadas en las ecuaciones 3.2 y 3.4 son las energías que se busca minimizar. En ese sentido, ambas pueden pensarse como funciones de costo.

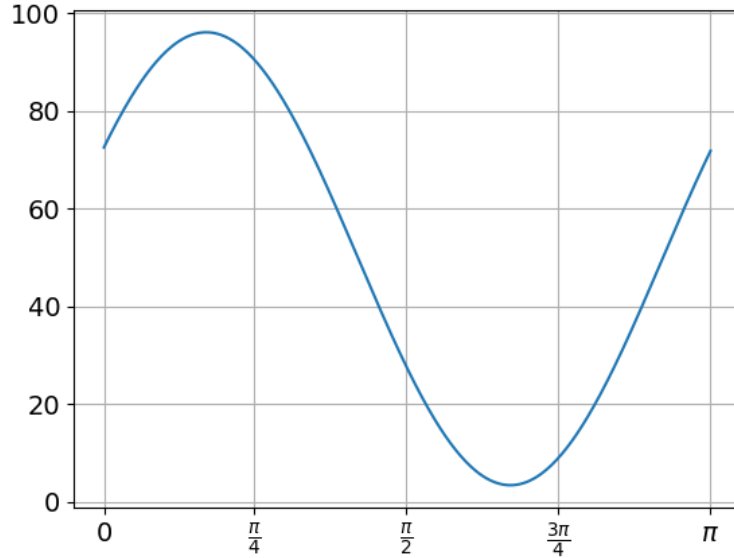


Figura 3.4: Ejemplo de energía para un evento. Para este ejemplo, se rotaron las componentes entre los ángulos 0 y π . La función tiene un máximo, que corresponde al ángulo óptimo, y un mínimo.

Extremos de la energía como función del ángulo de rotación

Una forma sencilla de minimizar la energía, teniendo las expresiones 3.2 y 3.4, es hallar los extremos a través de sus derivadas, las cuales pueden ser halladas analíticamente. La función de energía definida en la ecuación 3.2 puede reescribirse de la siguiente manera:

$$E_V(\phi) = \cos^2 \phi \sum_i U_i^2 + \sin^2 \phi \sum_i V_i^2 + 2 \sin \phi \cos \phi \sum_i U_i V_i. \quad (3.5)$$

Las sumatorias $\sum_i U_i^2 = E_U^{(0)}$ y $\sum_i V_i^2 = E_V^{(0)}$ son las energías de las componentes U y V respectivamente previo a la rotación. Por otro lado, teniendo en cuenta que $2 \sin \phi \cos \phi = \sin 2\phi$, la ecuación 3.5 se puede reescribir como:

$$E_V(\phi) = E_U^{(0)} \cos^2 \phi + E_V^{(0)} \sin^2 \phi + \sin 2\phi \sum_i U_i V_i. \quad (3.6)$$

Entendiendo que $E_U^{(0)}$, $E_V^{(0)}$ y $\sum_i U_i V_i$ son constantes respecto de ϕ (pues utilizan las muestras previas a la rotación), la derivada puede calcularse de manera sencilla como:

$$\begin{aligned} \frac{dE_V}{d\phi}(\phi) &= E_U^{(0)} 2 \cos \phi (-\sen \phi) + E_V^{(0)} 2 \sen \phi \cos \phi + 2 \cos 2\phi \sum_i U_i V_i \\ &= -E_U^{(0)} \sen 2\phi + E_V^{(0)} \sen 2\phi + 2 \cos 2\phi \sum_i U_i V_i. \end{aligned}$$

Finalmente,

$$\frac{dE_V}{d\phi}(\phi) = (E_V^{(0)} - E_U^{(0)}) \sen 2\phi + 2 \cos 2\phi \sum_i U_i V_i. \quad (3.7)$$

Luego, el ángulo óptimo ϕ^* que minimiza la energía cumple que

$$\frac{dE_V}{d\phi}(\phi^*) = 0, \quad (3.8)$$

lo que lleva a

$$\begin{aligned} (E_V^{(0)} - E_U^{(0)}) \sen 2\phi^* + 2 \cos 2\phi^* \sum_i U_i V_i &= 0, \\ (E_U^{(0)} - E_V^{(0)}) \sen 2\phi^* &= 2 \cos 2\phi^* \sum_i U_i V_i, \\ \frac{\tan 2\phi^*}{2} &= \frac{\sum_i U_i V_i}{E_U^{(0)} - E_V^{(0)}}, \end{aligned}$$

obteniendo finalmente que

$$\phi^* = \frac{1}{2} \tan^{-1} \left(\frac{2 \sum_i U_i V_i}{E_U^{(0)} - E_V^{(0)}} \right). \quad (3.9)$$

Análogamente, la variación de la energía en U (ver ecuación 3.4) respecto a θ está dada por:

$$\frac{dE_U}{d\theta}(\theta) = (E_W^{(1)} - E_U^{(1)}) \sen 2\theta + 2 \cos 2\theta \sum_i U_i^{(1)} W_i^{(1)}, \quad (3.10)$$

donde ahora el supraíndice (1) indica que las componentes utilizadas son las obtenidas luego de la primera rotación. De la ecuación 3.10, el ángulo óptimo θ^* para la segunda rotación será:

$$\theta^* = \frac{1}{2} \tan^{-1} \left(\frac{2 \sum_i U_i^{(1)} W_i^{(1)}}{E_U^{(1)} - E_W^{(1)}} \right). \quad (3.11)$$

Las expresiones 3.9 y 3.11 son las expresiones analíticas de los ángulos óptimos calculados según el criterio de mínima energía. No obstante, a pesar de contar con expresiones

cerradas para calcular ϕ^* y θ^* , en este trabajo se adoptó una estrategia de minimización numérica y no analítica, por dos motivos:

1. Las ecuaciones 3.9 y 3.11 tienen dos soluciones, una para el mínimo de la energía y otra, con un salto de $\pm\pi/2$, para el máximo de energía. Esto implica que se requiere un criterio adicional para el cálculo del valor mínimo, el cual se busca. Una opción es a través del criterio de la derivada segunda, el cual verifica si un extremo es mínimo o máximo de acuerdo al signo de la derivada evaluada en dicho extremo. Sin embargo, ésto requiere cálculos adicionales.
2. Se ha observado que el cálculo analítico de los ángulos óptimos provoca saltos de $\pm\pi$. Si bien matemáticamente esto no es un inconveniente, puesto que la energía no cambia para cambios de $\pm\pi$, sí impide una adecuada interpretación de los ángulos cuando se los observa gráficamente como función del tiempo, puesto que los ángulos presentan discontinuidades. En ese sentido, es preferible una alternativa que asegure la continuidad de los ángulos para observar su comportamiento constante en presencia de un evento linealmente polarizado.

Para evitar estos inconvenientes, y por simplicidad, en este trabajo se estimaron los ángulos óptimos mediante el método de optimización *Gradient Descent*, como se describe a continuación.

3.2.2. Método de *Gradient Descent*

El método de *Gradient Descent* es un algoritmo de optimización iterativo que permite hallar un mínimo local de una función diferenciable (Cauchy, 1847). La idea del mismo es considerar el gradiente de la función y tomar pasos discretos escalados por el opuesto de dicho gradiente hasta arribar a un mínimo local. La Figura 3.5 muestra un ejemplo concreto y conceptual del método de *Gradient Descent*. Entonces, si $f(\vec{x})$ es una función diferenciable, para hallar el mínimo local se realizan los siguientes pasos:

1. Se considera un valor inicial \vec{x}_0 y se evalúa tanto la función como su gradiente en dicho valor inicial.
2. Se considera un tamaño de paso ϵ y una cantidad de pasos N_{GD} .
3. Para $i = 0, 1, \dots, N_{GD}$, se calcula

$$\vec{x}_i = \vec{x}_{i-1} - \epsilon \nabla f(\vec{x}_{i-1}) \quad (3.12)$$

y se evalúa la función en x_i .

4. Finalizadas las iteraciones, el mínimo local puede ser aproximado como $f(x_{N_{GD}})$.

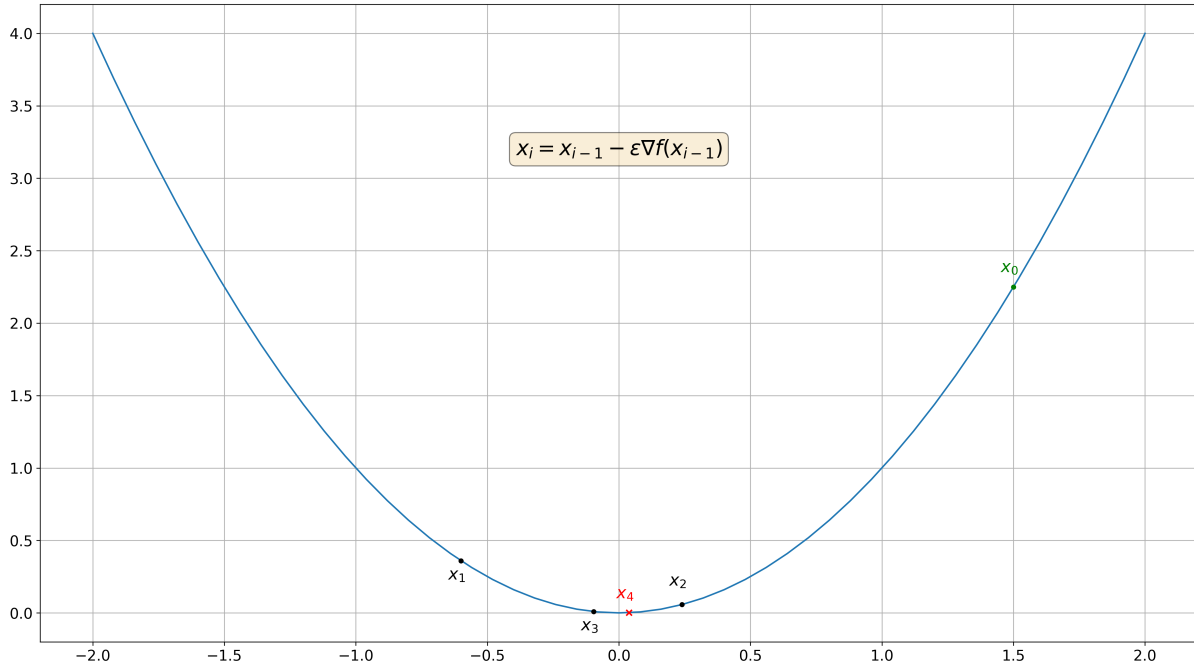


Figura 3.5: Ejemplo del método de *Gradient Descent* para hallar el mínimo de $f(x) = x^2$. En este caso, se usaron $N = 5$ iteraciones con un paso de $\epsilon = 0.7$, y como valor inicial se tomó $x_0 = 1.5$. Luego de solo 5 iteraciones, el resultado tiene un error del orden de $|x_5| \sim 10^{-2}$.

En principio, el método de *Gradient Descent* depende de tres parámetros: el punto inicial x_0 , el número de iteraciones N_{GD} y el tamaño de paso ϵ . Como se verá en secciones posteriores, el parámetro más problemático es el valor inicial, puesto que éste no es igual para cada muestra. Sin embargo, como veremos, existen alternativas que involucran un cálculo aproximado del mínimo, que luego es utilizado como x_0 de entrada al algoritmo.

3.2.3. Consideraciones prácticas

En la práctica, en general no es posible conocer *a priori* la presencia de un evento cuando se trabaja con grandes volúmenes de datos. Por lo tanto, este método trabaja sobre todos las componentes y sobre todos los tiempos mediante el uso de una ventana móvil que se mueve muestra a muestra. El algoritmo de optimización se aplica sobre toda la ventana móvil de longitud N_w y los ángulos óptimos son asignados a la muestra central de la ventana.

La longitud de esta ventana móvil debe ser representativa de un evento, pero no debe ser demasiado grande para no incorporar ruido incoherente que dificulte la determinación de los ángulos óptimos.

Dentro de la ventana móvil puede o no haber señal, y ésta puede o no estar linealmente polarizada. Si este no es el caso, el método no podrá concentrar toda la energía en una

sola componente. Sin embargo, una ventaja importante de este método es que, cuando la señal no está linealmente polarizada, la relación señal/ruido no empeora, como veremos en los ejemplos siguientes.

Por otro lado, el método requiere, para cada muestra dentro de la ventana móvil, un ángulo inicial a partir del cual aproximar al valor mínimo. Para esto, se hace un cálculo grueso de la energía del evento en una cantidad pequeña N_{gr} de ángulos equiespaciados entre 0 y π en la primera ventana móvil de la traza. El ángulo dentro de este conjunto pequeño para el cual la energía calculada es mínima es utilizado como valor inicial. Para las muestras siguientes, se toma como ángulo inicial el resultado del algoritmo aplicado a la ventana anterior.

Una ventaja de seleccionar el valor inicial de esta manera es que asegura la continuidad del ángulo calculado en función del tiempo. Utilizando otras maneras de seleccionar el ángulo inicial, se observa que los ángulos obtenidos presentan saltos de $\pm\pi$. Si bien matemáticamente esto no es necesariamente un problema (ya que los saltos de $\pm\pi$ no afectan la energía de la señal, solo su polaridad), sí es un inconveniente a la hora de interpretar gráficamente los ángulos como funciones del tiempo.

En la Tabla 3.1, se resumen todos los parámetros que son necesarios para la implementación del método.

Parámetro	Descripción
N_w	Longitud de la ventana móvil.
N_{gr}	Puntos para estimar valor inicial en la primera ventana.
ϵ	Paso en cada iteración. Factor de escala del gradiente.
N_{GD}	Cantidad de iteraciones del método <i>Gradient Descent</i> .

Tabla 3.1: Parámetros necesarios para la obtención del ángulo óptimo.

3.3. Aplicación a sismograma específico

Para demostrar la eficacia de la metodología propuesta, se procede a exponer el caso de varios *martemotos* sintéticos dispuestos en secuencia, correspondientes a la fecha 09 de septiembre de 2019 (ver Figura 3.6). Se observan 3 tramos de los sismogramas que podrían asociarse con eventos sísmicos (aunque *a priori* no es posible determinar si efectivamente esto es así): una secuencia de arribos de baja amplitud y duración entre las 17:49 y las 17:51 horas (azul), un tren de ondas de baja amplitud entre las 17:52 y las 17:58 (verde), y un último tren de ondas de alta amplitud y larga duración, entre las horas 18:01 y 18:09 (rojo). Todas las componentes muestran esos eventos, lo que indica que su energía está distribuida de manera desigual en las tres componentes.

La relación señal/ruido estimada componente a componente para la ventana de mayor amplitud tiene los siguientes valores:

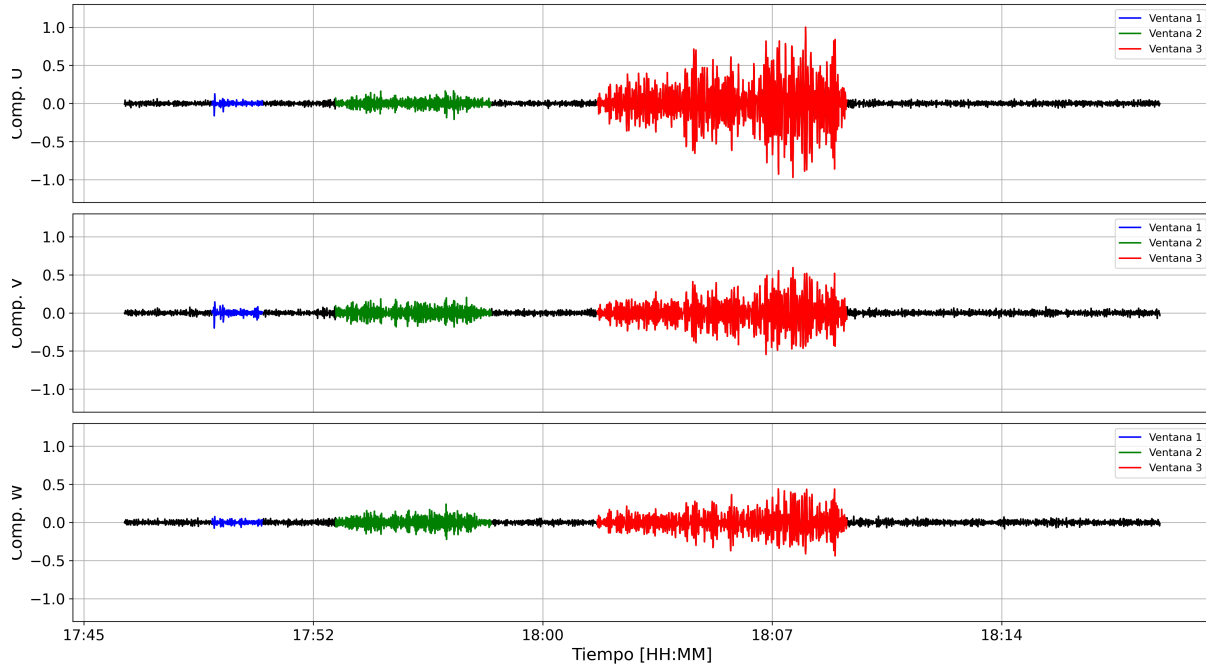


Figura 3.6: *Martemotos* seleccionados para el estudio de la eficacia del algoritmo. Se separa al sismograma en tres segmentos de interés. La energía de los eventos (sísmicos o no) está distribuida en las tres componentes.

$$SNR_U^{(0)} \approx 170, \quad (3.13)$$

$$SNR_V^{(0)} \approx 52, \quad (3.14)$$

$$SNR_W^{(0)} \approx 26. \quad (3.15)$$

Para la estimación de la relación señal/ruido se calculan por separado la energía del ruido y la del evento. Para la energía de la señal se tomó una ventana entre las horas 18:01 y 18:09, dando una longitud de 8 minutos, a la cual se le calculó la energía; esta ventana corresponde al sismo 3, que está marcado en la Figura 3.6. La ventana del ruido se seleccionó entre las horas 18:11 y 18:19, con la misma longitud que la ventana de señal.

Para la etapa de minimización de las energías, los parámetros utilizados están resumidos en la Tabla 3.2. Éstos fueron obtenidos subjetivamente, analizando el comportamiento de los ángulos óptimos y de la relación señal/ruido obtenida una vez finalizado el proceso.

El método es aplicado en dos etapas, una correspondiente a la primera rotación, y otra correspondiente a la segunda rotación. La primera rotación se realiza en el plano (U, V) , es decir con el eje W como eje de rotación, con el objetivo de que la energía en la componente V sea mínima. La segunda rotación se realiza de manera que el algoritmo concentre toda la energía de la señal en la componente W .

Parámetro	Valor
N_w	51
N_{gr}	5
ϵ	0.01
N_{GD}	50

Tabla 3.2: Valores seleccionados para la aplicación del método de minimización de la energía.

3.3.1. Primera rotación

La primera rotación implica la minimización de la energía de la componente V . La Figura 3.7 muestra el resultado de calcular los ángulos óptimos ϕ y rotar las componentes U y V alrededor del eje W .

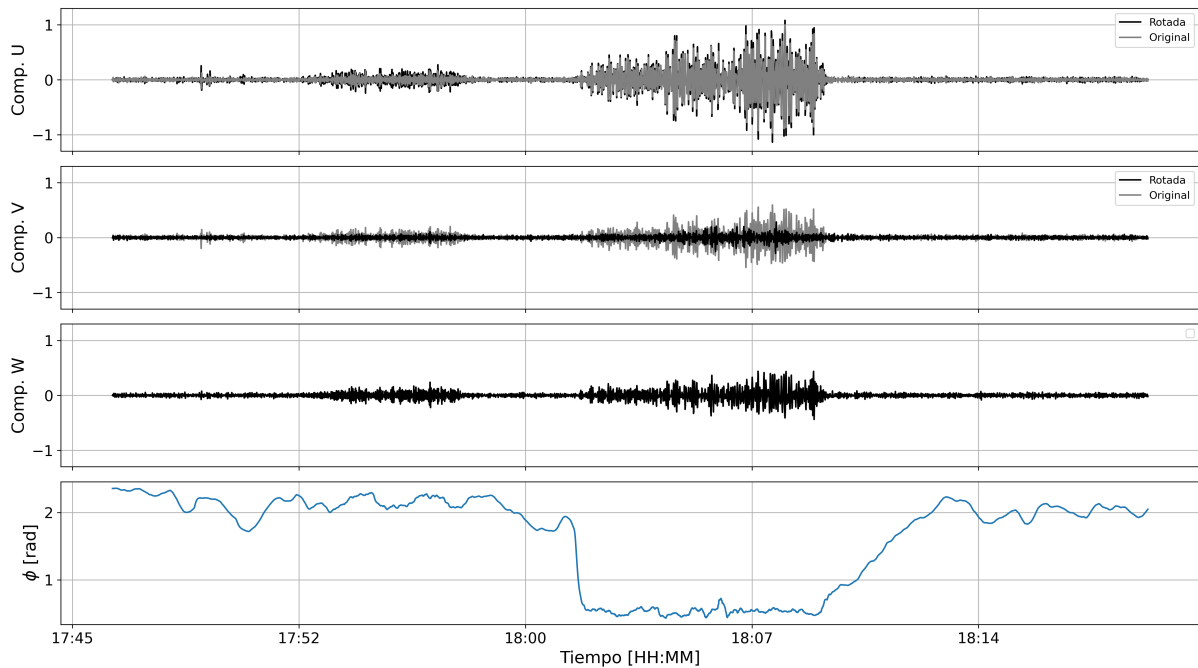


Figura 3.7: Gráficos 1 a 3: componentes U , V y W antes (gris) y después (negro) de la primera rotación. La componente W no sufre cambios. Abajo: ángulo óptimo obtenido muestra a muestra.

La Figura 3.7 muestra que, efectivamente, la transferencia de la energía de la señal de la componente V a la componente U fue en su mayor parte exitosa, aunque aún se observa señal en la componente V . En el último tren de ondas en particular, la energía del evento parece no haber podido rotarse en su totalidad. Para verificar esto se re-calcula la relación señal/ruido componente a componente. Luego de la primera rotación, toma los siguientes valores:

$$SNR_U^{(1)} \approx 179, \quad (3.16)$$

$$SNR_V^{(1)} \approx 8, \quad (3.17)$$

$$SNR_W^{(1)} \approx 26. \quad (3.18)$$

De estos valores se pueden hacer las siguientes observaciones: en primer lugar, la relación señal/ruido de la componente U no aumentó sustancialmente con la rotación. Esto puede deberse al hecho de que en términos relativos, ya había gran parte de la energía en esa componente, o también a que parte del ruido presente en la ventana 3 también se trasladó a la componente U . En segundo lugar, la relación señal/ruido de la componente V disminuyó sustancialmente (un orden de magnitud), pero sigue siendo bastante alta, lo que indica que, efectivamente, hubo señal que no pudo ser rotada hacia el plano (U, W) .

3.3.2. Segunda rotación

La segunda rotación es completamente análoga a la primera en cuanto a su implementación, con la salvedad de que ahora se minimiza la energía en la componente U , rotando respecto de la componente V , cuya energía ya fue minimizada. La Figura 3.8 muestra el comportamiento del ángulo θ junto con los resultados de aplicar la rotación a los sismogramas. Se utilizaron los mismos parámetros que en la primera rotación (ver Tabla 3.2).

Luego de la rotación, la relación señal/ruido componente a componente toma los siguientes valores:

$$SNR_U^{(2)} \approx 6, \quad (3.19)$$

$$SNR_V^{(2)} \approx 8, \quad (3.20)$$

$$SNR_W^{(2)} \approx 226. \quad (3.21)$$

Es importante notar que $SNR_U^{(2)}$ no es cercana a la unidad, como tampoco lo es $SNR_V^{(2)}$. Esto es así porque, como ocurrió también en la primera rotación, no toda la energía de la señal fue transferida a la componente vertical. Sin embargo, hubo un aumento muy grande de la relación señal ruido en la componente W , al punto tal de que ésta supera la relación señal/ruido observada en la componente U antes de las rotaciones. La Figura 3.9 resume el comportamiento de la masa del sismómetro luego de cada rotación. Se ilustra claramente que la reducción a la componente vertical fue exitosa.

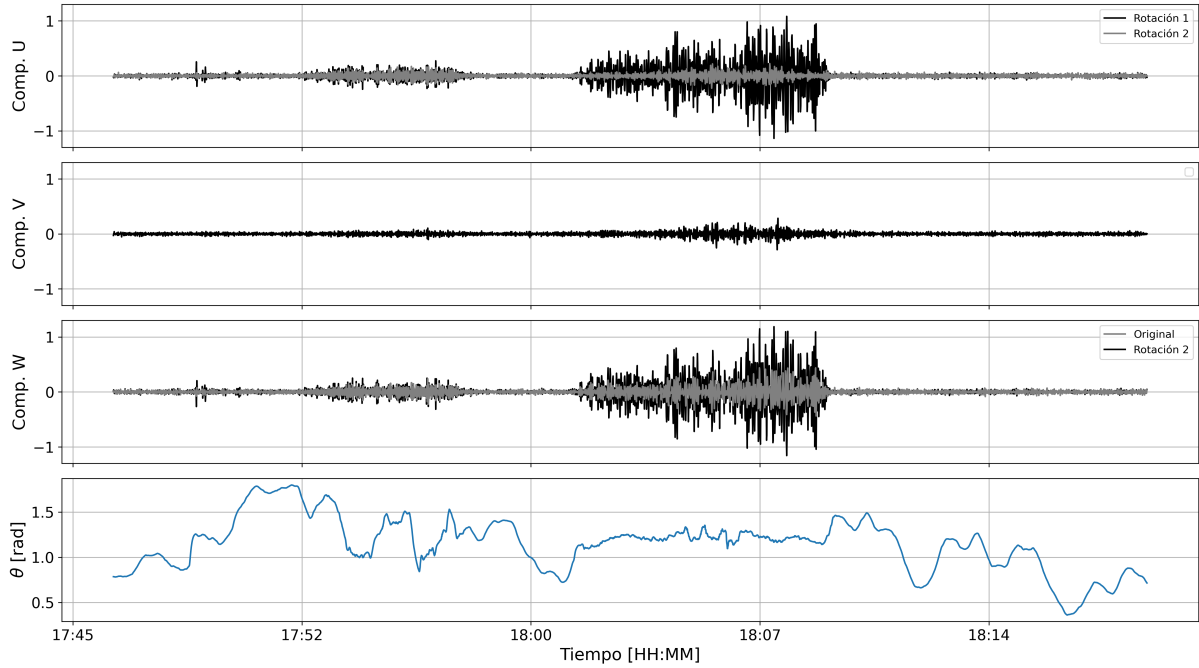


Figura 3.8: Gráficos 1 a 3: Componentes U , V y W antes (negro) y después (gris) de la segunda rotación. La componente V no se rota en esta etapa. Abajo: ángulo óptimo obtenido muestra a muestra.

3.3.3. Observaciones de los resultados

Gráficamente, pueden hacerse varias observaciones acerca del comportamiento de los eventos marcados en la Figura 3.6 luego de la aplicación del algoritmo. El comportamiento es distinto para cada uno de los tres eventos seleccionados, por lo que es conveniente separar estas observaciones por ventana.

■ Ventana 1:

Una estimación de la relación señal/ruido para este evento en la componente W previo a la rotación y luego de la rotación arroja los siguientes valores:

$$SNR_W^{(0)} \approx 2 \quad (3.22)$$

$$SNR_W^{(2)} \approx 4. \quad (3.23)$$

Esto implica que este arribo, el cual previo a la rotación estaba prácticamente enmascarado por ruido incoherente en la componente vertical, ahora es visible y detectable en dicha componente.

■ Ventana 2

Esta ventana es la que menos cambio en su amplitud sufre luego de la aplicación del algoritmo. En este caso, la relación señal/ruido en la componente W previo y luego de la rotación muestran una variación menor:

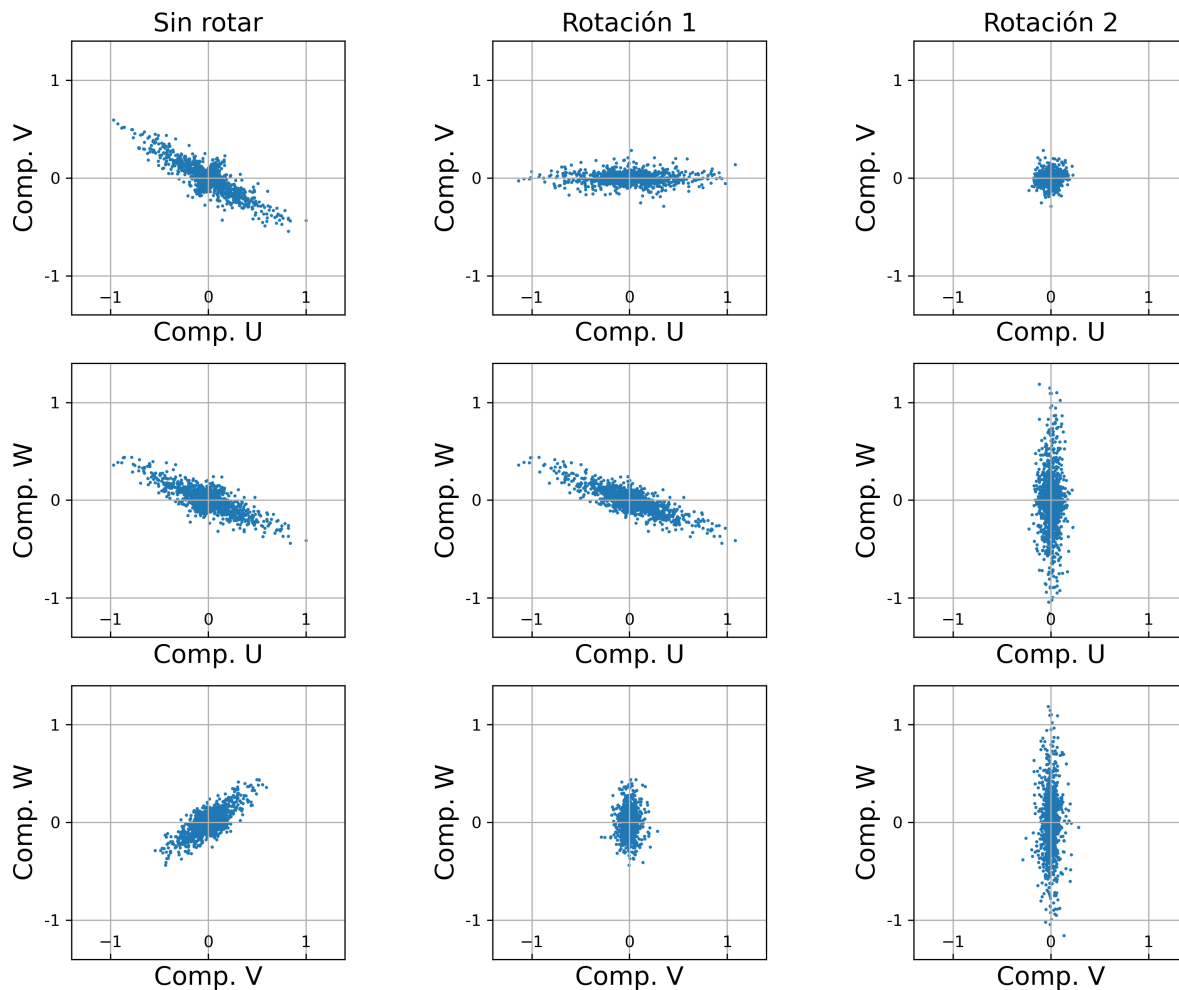


Figura 3.9: Trayectoria de la masa del sensor durante la duración de la ventana seleccionada. Columna 1: previo a la rotación. Columna 2: luego de la primera rotación. Columna 3: luego de la segunda rotación.

$$SNR_W^{(0)} \approx 7 \quad (3.24)$$

$$SNR_W^{(2)} \approx 12. \quad (3.25)$$

Esto es consistente con el comportamiento de los ángulos en este evento. Ambos ángulos (especialmente θ) presentan variaciones importantes, con lo que los ángulos de rotación no son constantes durante el evento. Sin embargo, la primera de las rotaciones fue exitosa, puesto que trasladó gran parte de la energía de la componente V a la componente U . Es decir, las ondas correspondientes a la ventana 2 pueden concentrarse en el plano (U, W) , pero no en una única dirección. Esto puede deberse a que el evento en cuestión se compone principalmente de ondas elípticamente polarizadas superpuestas, con lo cual la polarización no es lineal y el algoritmo es incapaz de reducir el evento a una única com-

ponente. Si este es el caso, un aspecto rescatable de este algoritmo es que, cuando trabaja con ondas que no están linealmente polarizadas, la rotación **no afecta negativamente la relación señal/ruido de las ondas**. Esto es importante porque no se corre el riesgo de que el algoritmo actúe de una manera tal que se enmascare algún evento visible *previo* a la rotación.

■ Ventana 3

Este evento es el que más variación presenta en su relación señal/ruido luego de la reducción a la componente vertical. Esto es predecible si se observa el comportamiento de los ángulos, mostrado en las Figuras 3.7 y 3.8. En este caso, los ángulos óptimos son casi constantes, presentando muy poca variación cuando se los compara con el resto de la serie de tiempo, y toman valores estimados de

$$\phi \approx 28^\circ, \quad (3.26)$$

$$\theta \approx 71^\circ. \quad (3.27)$$

3.3.4. Detección de eventos

Esta sección comprende la etapa de detección de los eventos antes y después de la rotación. Para la comparación, y para verificar la eficacia de los métodos en el ejemplo en cuestión, se aplican los algoritmos antes y después sobre la componente vertical. En ambos casos, se aplicaron los mismos parámetros, cuyos valores están resumidos en la Tabla 3.3. Los valores asignados están basados en el trabajo de Sabbione *et al.* (2011), con modificaciones adecuadas que se ajustan mejor a los datos sintéticos marcianos. Para ver la definición de cada parámetro, referirse la Sección 1.2.

Detección en sismograma sin rotar

La aplicación del método con los parámetros seleccionados a la componente vertical sin rotar arroja resultados variables de acuerdo al método. En la Figura 3.10 se muestra el comportamiento de los cocientes STA/LTA, junto con los tiempos de los eventos detectados para cada uno de los tres métodos analizados.

- **Método de Allen (MAM):** Este método no detectó eventos. Esto puede deberse a la baja relación señal/ruido que hace que, con los parámetros seleccionados, el método no pueda resolver los arribos. Gráficamente, se observa que el STA/LTA tiene un comportamiento diferente cuando se tiene un primer arribo, pero el umbral no es suficiente para que éstos sean detectados. Si bien es claro que si el umbral hubiese sido menor la detección de algunos eventos habría sido exitosa, la decisión de tomar $THR = 3$ resultó de la necesidad de evitar falsas detecciones en otros sismogramas del mismo set de datos.

Método	Parámetro	Valor
MAM	C_3	0.05
	C_4	0.01
	T_{taper}	120 s
	THR	3
ESM	T_{STA}	5 s
	T_{LTA}	50 s
	T_{Hann}	10 s
	T_{taper}	120 s
	THR	3
MAM	T_{STA}	5 s
	T_{LTA}	50 s
	T_{Hann}	10 s
	T_{taper}	250 s
	THR	8

Tabla 3.3: Valores asociados a los utilizados para los tres métodos de detección automática.

- **Método de Earle-Shearer (ESM):** Detectó un único evento, cercano al primer arribo del sismo 3. No detecta eventos en la ventana 1 ni en la ventana 2.
- **Método de Allen Modificado (MAM):** Detectó un único evento, cercano al primer arribo del sismo 3 y muy cercano a la detección hecha por el método ESM. La diferencia en tiempo entre el evento obtenido por ESM y el obtenido por MAM es del orden de 10^{-6} s. No detecta eventos en la ventana 1 ni en la ventana 2.

Detección en sismograma rotado

Las detecciones automáticas post-rotación arrojan resultados que varían significativamente con respecto a las aplicadas sobre la componente vertical sin rotar. La Figura 3.11 muestra el comportamiento de los atributos y las detecciones en cada uno de los métodos. En este caso, los tres métodos arrojaron resultados similares, detectando tres eventos, uno en la ventana 1 (que no había sido detectado en el sismograma sin rotar) y dos cercanos al primer arribo en la ventana 3. Ninguno de los tres métodos detecta eventos en la ventana 2.

3.4. Conclusiones

Se seleccionaron tres eventos correspondientes al set de datos sintéticos marcianos. Estos eventos tienen la particularidad de comportarse de manera diferente uno respecto a otro, lo que permite estudiar la eficacia de la reducción a una componente en tres casos diversos. La eficacia de la reducción a la componente vertical se analizó para cada

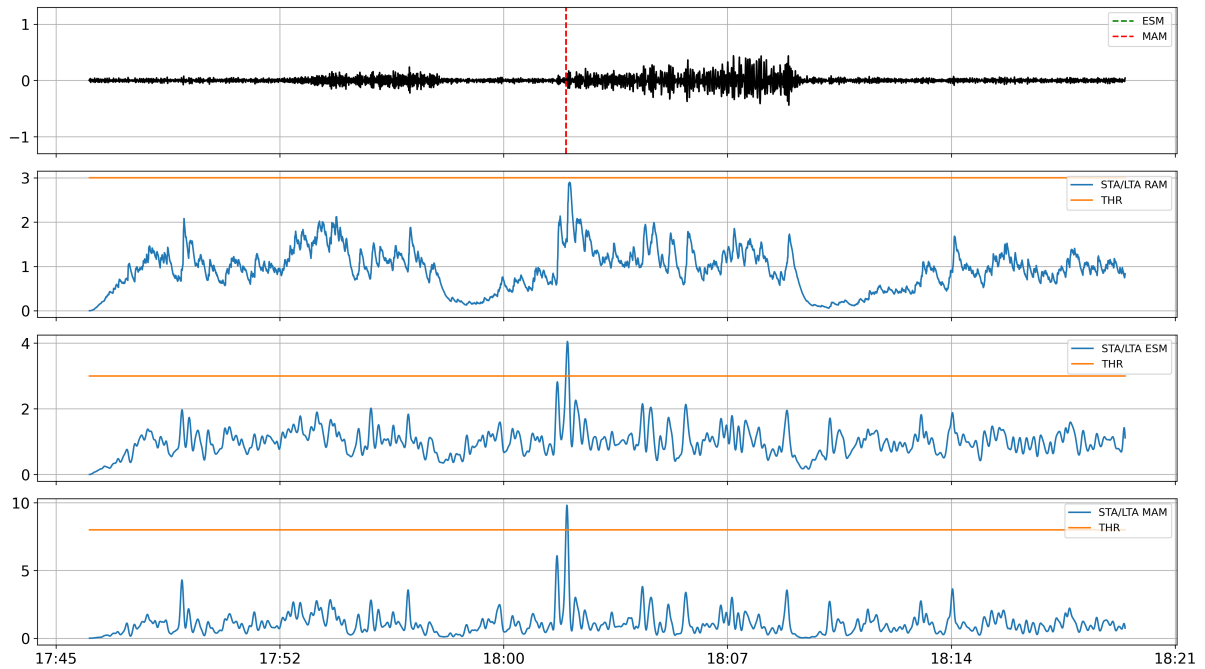


Figura 3.10: Gráfico 1: sismograma con los eventos detectados por cada método de detección aplicado a la componente vertical sin rotar. El método RAM no detectó eventos; los métodos ESM y MAM detectaron el mismo evento con una diferencia de 10^{-6} s. Gráfico 2 a 5: Atributos de RAM, ESM y MAM, respectivamente.

ventana por separado calculando la relación señal/ruido antes y después de la aplicación de cada rotación por separado. Para verificar si la reducción significaba una mejora en los resultados de los métodos de detección, se aplicaron estos métodos al sismograma de ejemplo.

La aplicación del algoritmo de reducción a una componente resultó muy favorable en dos de los tres casos de estudio. Esto se refleja en los valores calculados para las relaciones señal/ruido calculadas en cada componente. En el caso particular del tercer evento, la relación señal/ruido en la componente W luego de las dos rotaciones es significativamente mayor que las relaciones señal/ruido de cada componente por separado.

La energía del primer evento también fue transferida exitosamente a la componente vertical. Tal es así, que todos los métodos de detección pudieron detectarlo. Si bien el aumento en la relación señal/ruido previo y luego de la rotación no es tan marcado como en el tercer caso, esto puede deberse a que la primera ventana podría ser una fase de onda interna, que en eventos tectónicos suelen tener menos energía.

Por su parte, el segundo evento no mostró aumentos significativos en su relación señal/ruido luego de aplicada la reducción. No obstante, esto no significa que no se puedan extraer conclusiones de este caso. Lo que se observa es que, en la primera rotación, la transferencia de energía de la componente V a la componente U es muy buena, pero no así el caso de la componente U a la W en la segunda rotación. Esto implica que la amplitud

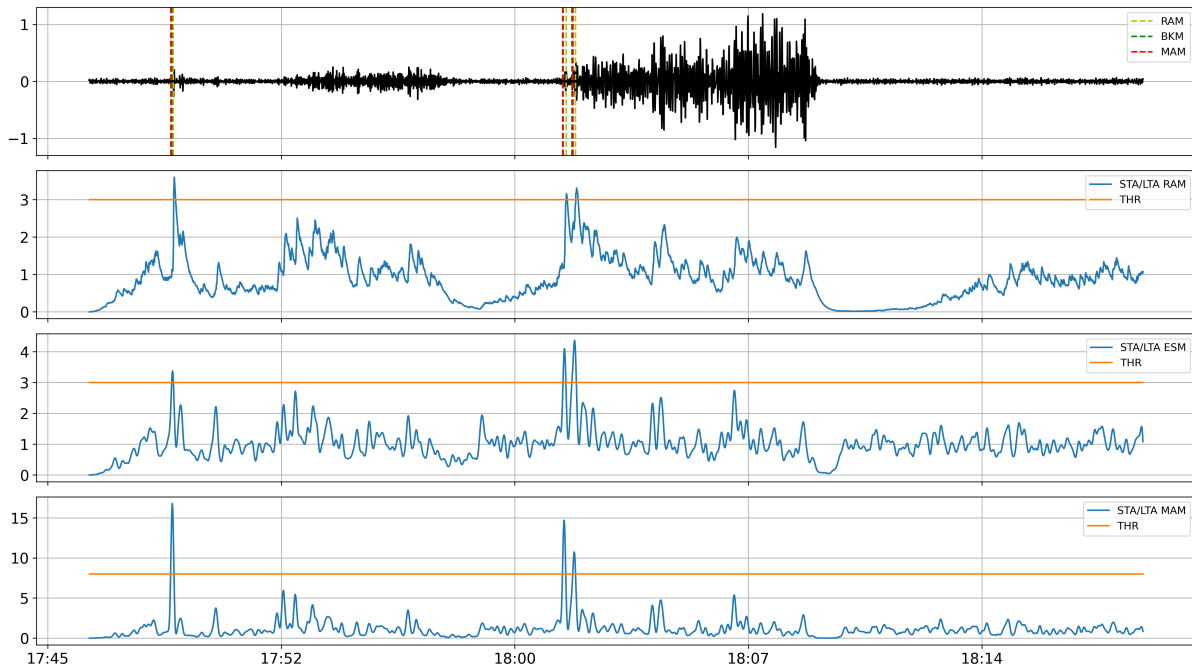


Figura 3.11: Gráfico 1: sismograma con los eventos detectados por cada método aplicado a la componente vertical rotada. Los tres métodos detectaron tres eventos cada uno. Gráficos 2 a 5: Atributos de RAM, ESM y MAM, respectivamente.

de la señal fue proyectada a un plano, pero no a una única dirección. Esto es consistente con el comportamiento de ondas superficiales, que no presentan una polarización lineal y por ende no es posible reducirlas a una única componente. Incluso en este caso desfavorable para el método, donde la hipótesis principal de polarización lineal no se cumple, la reducción pudo aumentar en un cierto grado la relación señal/ruido, mostrando que no hubo efectos adversos que afecten negativamente la energía o la forma de las ondas superficiales.

Por lo tanto, puede concluirse que el algoritmo tiene un funcionamiento acorde a lo deseado, reduciendo la señal distribuida en las tres componentes del sismómetro a una única componente cuya relación señal/ruido es mayor si la hipótesis de polarización lineal se cumple. Si esta hipótesis no es válida, el método no muestra mejoras significativas, pero no disminuye la relación señal/ruido de la señal.

En relación a las detecciones, y dado que la relación señal/ruido es mayor luego de la reducción a una componente, es de esperar que la capacidad de detección de los algoritmos aumente (provisto que los valores de sus parámetros sean sensatos) cuando se aplican sobre el sismograma rotado.

Previo a la rotación, los métodos muestran un desempeño y un comportamiento variado según la ventana. Los métodos ESM y MAM son los únicos que detectan el inicio del evento en la ventana 3. RAM, por su parte, no detectó ningún evento. Ninguno de los tres métodos detecta ningún evento en la ventana 2.

En este punto es importante reconocer que la detección de ciertos eventos pudo haber ocurrido si los parámetros hubiesen sido diferentes. Por ejemplo, si el THR de RAM hubiese sido menor, la detección del evento 3 habría ocurrido. Sin embargo, los parámetros elegidos son el resultado de pruebas en otros eventos del mismo set de datos sintéticos, que fueron hechas con el objetivo de llegar a una relación adecuada entre la cantidad de eventos detectados correctamente y la cantidad de falsas detecciones. No obstante, es importante realizar un análisis más fino de la selección de los parámetros de detección, a fin de reconocer si existe un conjunto de parámetros tales que las detecciones sean mejores a las obtenidas.

Con respecto a la eficacia de los métodos cuando se realiza previamente la reducción de las componentes, ésta aumenta en todos los métodos. RAM, ESM y MAM detectan los arribos correspondientes a las ventanas 1 y 3, lo cual representa una mejora puesto que en la componente sin rotar no eran capaces de hacerlo. Estos resultados son consistentes con los observados en el Capítulo 4, donde tanto el algoritmo de reducción como los métodos de detección son aplicados a todo el set de datos.

De estas observaciones es posible concluir que el algoritmo de reducción a una componente resulta, en los tres métodos aplicados en este sismograma de ejemplo, en un aumento en la cantidad de detecciones correctas.

Capítulo 4

Resultados

En este capítulo se presentan los resultados de las detecciones obtenidas mediante los métodos descritos en los capítulos anteriores, aplicados a los datos sintéticos marcianos correspondientes al proyecto *Preparing for InSight*. Más específicamente, este capítulo compara la cantidad de detecciones hechas por los métodos con el sismograma sin rotar sobre la componente W , con las detecciones obtenidas utilizando el sismograma construido a partir del algoritmo de reducción a una componente. Los datos fueron separados en 365 sismogramas de 24 horas de longitud cada uno. Se aplicó un filtro pasa-altas con frecuencia de corte en 0.2 Hz, como se describe en la Sección 2.3. A cada día se aplicaron los métodos de detección antes de la rotación, y luego de la rotación, a la componente W .

Para evaluar los resultados en todos los casos se utilizaron los parámetros ya mostrados en la Tabla 3.3, pero con $T_{taper} = 240$ s, a fin de reducir los efectos de borde.

Como referencia, se utilizaron las detecciones hechas por el equipo de Géoazur (Rolly *et al.*, 2019). El equipo de Géoazur utilizó dos métodos en paralelo para una detección de primer orden de candidatos a eventos:

- Método 1: los sismogramas se filtraron con un filtro pasa-altas de frecuencia de corte de 0.2 Hz. Los eventos fueron detectados con un método del tipo STA/LTA tradicional, con una longitud de ventana STA de 5 s, y una longitud de ventana LTA de 50 s.
- Método 2: se generó una señal auxiliar a partir de valores espectrales sumados entre las frecuencias 0.2 Hz y la frecuencia de Nyquist (1 Hz). El método utilizado fue del tipo STA/LTA con una longitud de ventana STA de 1 s y una longitud de ventana LTA de 200 s.

Estos métodos arrojaron 778 candidatos, que luego fueron picados y analizados manualmente hasta reducirlos a 214 eventos enviados al equipo del proyecto *Preparing for*

InSight (Rolly *et al.*, 2019). Las 214 fases picadas por Géoazur fueron comparadas con las detecciones resultantes de los métodos aplicados en este trabajo, previo y luego de la reducción a una componente. Cabe destacar que las detecciones hechas por Géoazur no son perfectas, ya que algunas pueden ser falsas, mientras que otras fases no fueron detectadas (van Driel *et al.*, 2019). Sin embargo, dado que no se contó con los picados correctos, para el análisis se consideran las detecciones de Géoazur como referencia para estudiar la eficiencia de los métodos aplicados en este trabajo.

Para determinar picados detectados tanto por Géoazur como por la detección propia, se aplicaron diversas tolerancias. De esta manera, se define un evento *correctamente detectado* a todo evento detectado por los métodos cuya diferencia en valor absoluto respecto a un arribo picado por Géoazur es menor a la tolerancia especificada. Las tolerancias utilizadas para la comparación fueron de 12 s, 18 s y 24 s. Cabe señalar que el objetivo fue realizar detecciones, y no determinaciones precisas de los tiempos de arribo de las fases obtenidas por los métodos automáticos.

4.1. Detección sin reducción a una componente

La detección sin reducción a una componente involucra aplicar los métodos de picado automático a cada una de las componentes por separado.

Componente U

Los eventos detectados al aplicar los tres métodos de detección a la componente horizontal U sin rotar se resumen en la Tabla 4.1. Con respecto a la cantidad de detecciones coincidentes con Géoazur, se observa que el método RAM tiene el peor desempeño de los tres métodos. Por su parte, los métodos ESM y MAM tienen desempeños mucho mejores, que son consistentes a lo largo de las tres componentes, pero siguen siendo menores a los resultados provistos por la componente W , como se verá en secciones posteriores.

Método	Detecciones	12 s	18 s	24 s
RAM	190	4	8	11
ESM	180	6	9	20
MAM	356	13	16	19

Tabla 4.1: Número de detecciones y porcentaje de coincidencias con los resultados de Géoazur para diferentes tolerancias, para el caso de la componente U del sismograma sin rotar.

Los resultados se analizan estadísticamente mediante la estimación de la precisión y la exactitud de cada método según la componente elegida. La exactitud de un método puede cuantificarse mediante el valor medio de la diferencia entre las detecciones propias y las

detecciones de Géoazur (dentro de las tolerancias estipuladas), mientras que la precisión se estima mediante la desviación estándar de dicha diferencia.

Para el caso de la componente U , los resultados obtenidos se muestran en la Tabla 4.2. Se observa que todos los métodos presentan sesgos considerables, reflejado por sus respectivos valores medios. Por otro lado, las dispersiones son sumamente altas.

Método	12 s	18 s	24 s
RAM	-2.9 ± 4.7	-1.8 ± 9.8	-1.4 ± 14.3
ESM	-2.7 ± 6.4	-2.4 ± 10.4	-11.0 ± 14.1
MAM	-4.3 ± 6.2	-4.4 ± 8.8	-3.7 ± 11.5

Tabla 4.2: Valor medio y desviación estándar de los métodos utilizados a partir de las detecciones realizadas sobre la componente U sin rotar, para tolerancias de 12 s, 18 s y 24 s.

La Figura 4.1 muestra el comportamiento de las detecciones, graficando para cada método y para cada tolerancia la distribución de las diferencias. Lo que se observa es que, en general, los métodos no presentan una distribución normal, esperable para los errores observacionales. En particular, el método RAM muestra una distribución mucho más uniforme que una distribución normal. El método ESM muestra un sesgo importante hacia los valores negativos, lo cual se observa también en el valor medio y en la dispersión calculadas. El método MAM también muestra un sesgo hacia los valores negativos. Esto se aprecia en los valores medios, que indican que los métodos tienen una tendencia a adelantarse a los tiempos determinados por Géoazur. Por otro lado, en general las distribuciones tienen una dispersión importante, lo cual es consistente con los valores altos de la desviación estandar calculada en cada tolerancia.

Componente V

La Tabla 4.3 resume el resultado de aplicar los métodos de detección a la componente horizontal V sin rotar. Los resultados son similares a los obtenidos en la componente U . Esto es esperable, puesto que ambas componentes son horizontales y tienen características similares.

Método	Detecciones	12 s	18 s	24 s
RAM	156	5	8	11
ESM	172	7	9	21
MAM	304	11	17	19

Tabla 4.3: Número de detecciones y porcentaje de coincidencias con los resultados de Géoazur para diferentes tolerancias, para el caso de la componente V sin rotar.

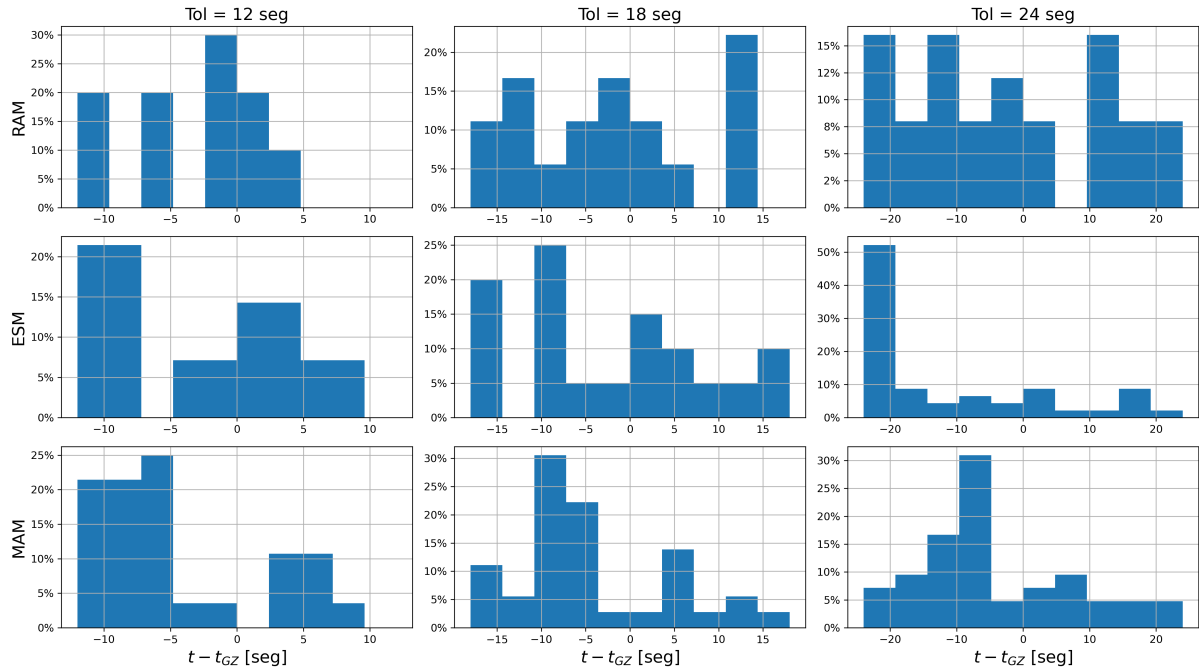


Figura 4.1: Distribución de las diferencias entre los tiempos picados por los métodos en la componente U sin rotar y los tiempos de Géozur, para las distintas tolerancias consideradas.

Tal como se hizo en la componente U , se calcula el valor medio y la desviación estándar de las diferencias entre las detecciones resultantes de los métodos automáticos y las detecciones de Géozur. Los resultados se muestran en la Tabla 4.4. Tanto el valor medio como la dispersión muestran valores similares a los obtenidos con la componente U . Como ya se observó en la componente U , los métodos tienen una tendencia a adelantarse a los tiempos picados por Géozur. Ésto queda ilustrado por los valores negativos de las medias de las diferencias. La dispersión es también muy alta, lo que muestra que las diferencias toman un gran rango de valores.

Método	12 s	18 s	24 s
RAM	-1.6 ± 6.6	-2.0 ± 10.2	-2.5 ± 14.1
ESM	-1.1 ± 7.5	-1.9 ± 10.3	-9.8 ± 15.0
MAM	-1.6 ± 7.2	-3.8 ± 10.3	-3.9 ± 12.0

Tabla 4.4: Valor medio y desviación estándar de los métodos utilizados a partir de las detecciones realizadas sobre la componente V sin rotar, para tolerancias de 12 s, 18 s y 24 s.

Finalmente, la Figura 4.2 muestra la distribución de las diferencias en función del método y de la tolerancia. Al igual que con la componente U , las diferencias no presentan una distribución simétrica centrada en el origen, sino que se muestran sesgadas hacia los tiempos negativos y con una dispersión importante.

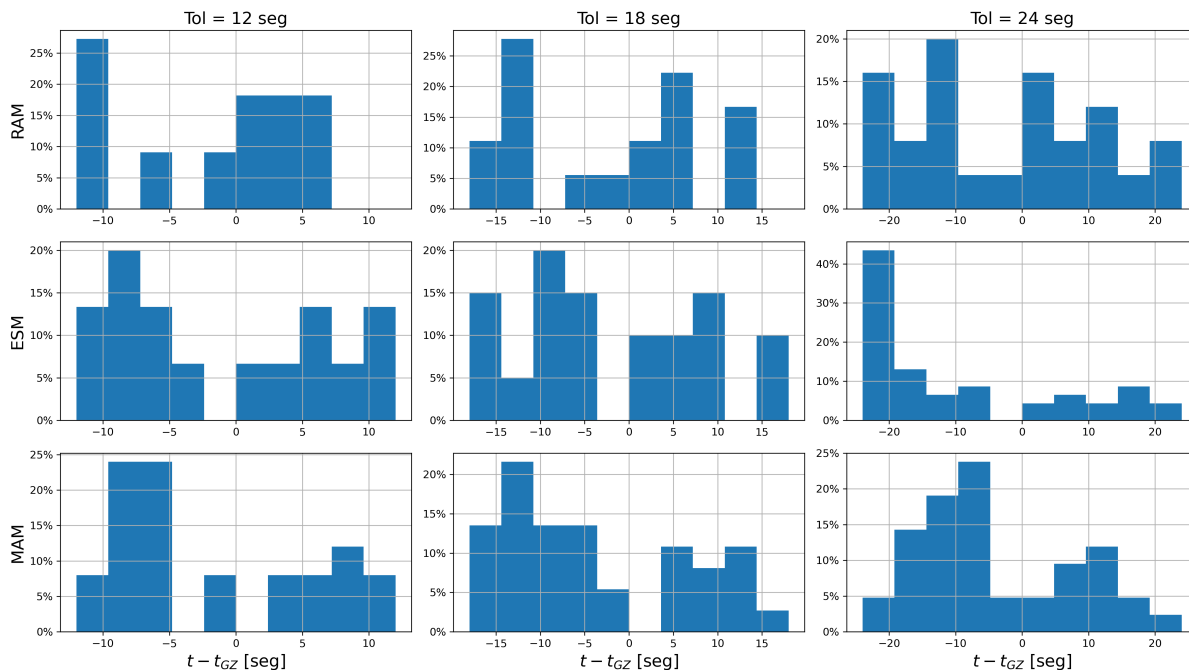


Figura 4.2: Distribución de las diferencias entre los tiempos picados por los métodos en la componente V y los tiempos de Géozur, para las distintas tolerancias consideradas.

Componente W

La Tabla 4.5 resume los resultados obtenidos de aplicar los algoritmos de detección a la componente W sin rotar. Es de esperar que ésta sea la componente con mayor relación señal/ruido, sobre todo para los arribos de ondas internas. En este caso, todos los métodos aumentan la cantidad de eventos correctamente detectados conforme aumenta la tolerancia máxima. El método MAM es el que más detecciones correctas obtuvo para bajas tolerancias, excepto para la tolerancia de 24 s, donde el número de detecciones coincide con la de ESM. Sin embargo, también tiene una gran cantidad de detecciones en comparación con RAM y con ESM, lo que indica que contiene muchos picados falsos. Este resultado es consistente en todas las componentes.

Método	Detecciones	12 s	18 s	24 s
RAM	193	5	10	13
ESM	213	8	11	21
MAM	430	14	19	21

Tabla 4.5: Número de detecciones y porcentaje de coincidencias con los resultados de Géozur para diferentes tolerancias, para el caso de la componente W sin rotar.

Los valores medios y las desviaciones estándar se resumen numéricamente en la Tabla 4.6. Como es de esperar, las desviaciones estándar aumentan conforme la tolerancia aumenta. Los valores medios del método RAM son los más cercanos a cero para todas

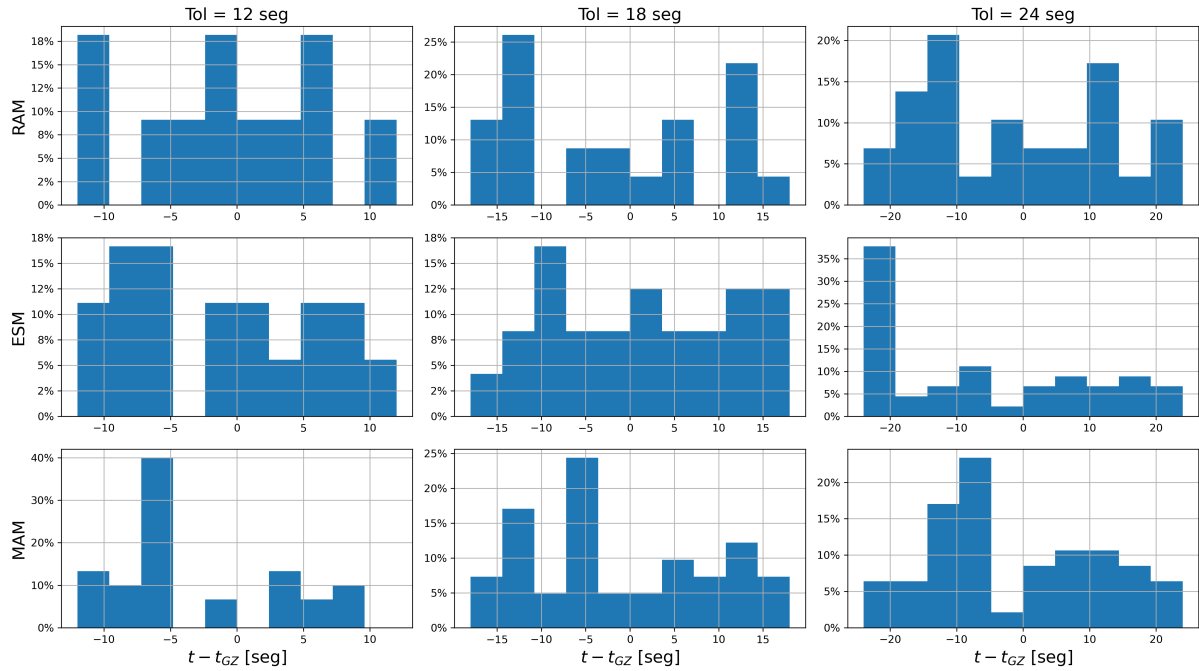


Figura 4.3: Distribución de las diferencias entre los tiempos picados por los métodos en la componente W sin rotar y los tiempos de Géozur, para las distintas tolerancias consideradas.

las tolerancias, y el método MAM es el que tiene el mayor valor medio. El método ESM supera al resto de los métodos tanto en valor medio como en dispersión para tolerancias altas.

Método	12 s	18 s	24 s
RAM	-0.5 ± 7.0	-1.5 ± 11.5	-1.1 ± 14.5
ESM	-1.1 ± 7.2	1.2 ± 10.3	-6.1 ± 15.8
MAM	-3.2 ± 6.4	-1.2 ± 10.1	-0.6 ± 12.6

Tabla 4.6: Valor medio y desviación estándar de los métodos utilizados a partir de las detecciones realizadas sobre la componente W del sismograma sin rotar, para tolerancias de 12 s, 18 s y 24 s.

La distribución de las diferencias se ilustran en la Figura 4.3. El comportamiento de las distribuciones es similar a las componentes horizontales. El método ESM muestra el mismo sesgo hacia valores negativos que se observa en las componentes U y V . El método RAM tiene también un comportamiento alejado de una distribución normal, y las diferencias están más uniformemente distribuidas entre los valores positivos y negativos. El método MAM es el que, para una tolerancia de 24 s, muestra una distribución más parecida a una normal, pero sigue habiendo un sesgo hacia las diferencias negativas.

4.2. Detección con reducción a una componente

La reducción a una componente se aplicó sobre la componente W , es decir sobre esta componente se maximizó la energía. Luego, sobre esta misma componente se aplicaron los tres métodos. La Tabla 4.7 resume la cantidad de detecciones por método. Se observa que, en este caso, la cantidad de eventos detectados es significativamente menor que en el caso sin rotación. Por otro lado, se observa que la cantidad de eventos coincidentes con Géozur aumenta considerablemente para todas las tolerancias consideradas.

Método	Detecciones	12 s	18 s	24 s
RAM	172	21	24	26
ESM	217	24	27	30
MAM	348	16	19	24

Tabla 4.7: Número de detecciones y porcentaje de coincidencias con los resultados de Géozur para diferentes tolerancias, para el caso del sismograma rotado.

La Tabla 4.8 resume los valores medios y desviaciones estándar de las diferencias entre los picados obtenidos por los métodos utilizados y los tiempos de Géozur, utilizando la componente W luego de la reducción a una componente.

Método	12 s	18 s	24 s
RAM	-0.7 ± 6.0	-0.0 ± 8.6	1.0 ± 10.7
ESM	1.4 ± 4.9	2.2 ± 7.1	2.7 ± 9.9
MAM	3.2 ± 5.2	2.4 ± 7.7	4.2 ± 11.1

Tabla 4.8: Valor medio y desviación estándar de los métodos utilizados a partir de las detecciones realizadas sobre la componente W rotada, para tolerancias de 12 s, 18 s y 24 s.

Los métodos RAM y ESM muestran mejoras significativas en su valor medio, y las desviaciones estándar disminuyen considerablemente en todas las tolerancias. En particular, para el método RAM los valores medios son más cercanos a cero que en las componentes sin rotar. Por su parte, MAM también presenta dispersiones menores para todas las tolerancias, y los valores medios ahora son positivos.

En la Figura 4.4 se presenta la distribución de las diferencias con el sismograma rotado. Se observan mejoras significativas en el comportamiento de las distribuciones, especialmente para los métodos RAM y ESM. El método RAM presenta una distribución centrada en valores cercanos a cero, con mayor simetría respecto del valor medio. El método ESM abandona el sesgo hacia los valores negativos que está presente en los resultados obtenidos con las componentes sin rotar, lo que robustece la confianza en los picados propios. El método MAM también presenta una distribución más simétrica.

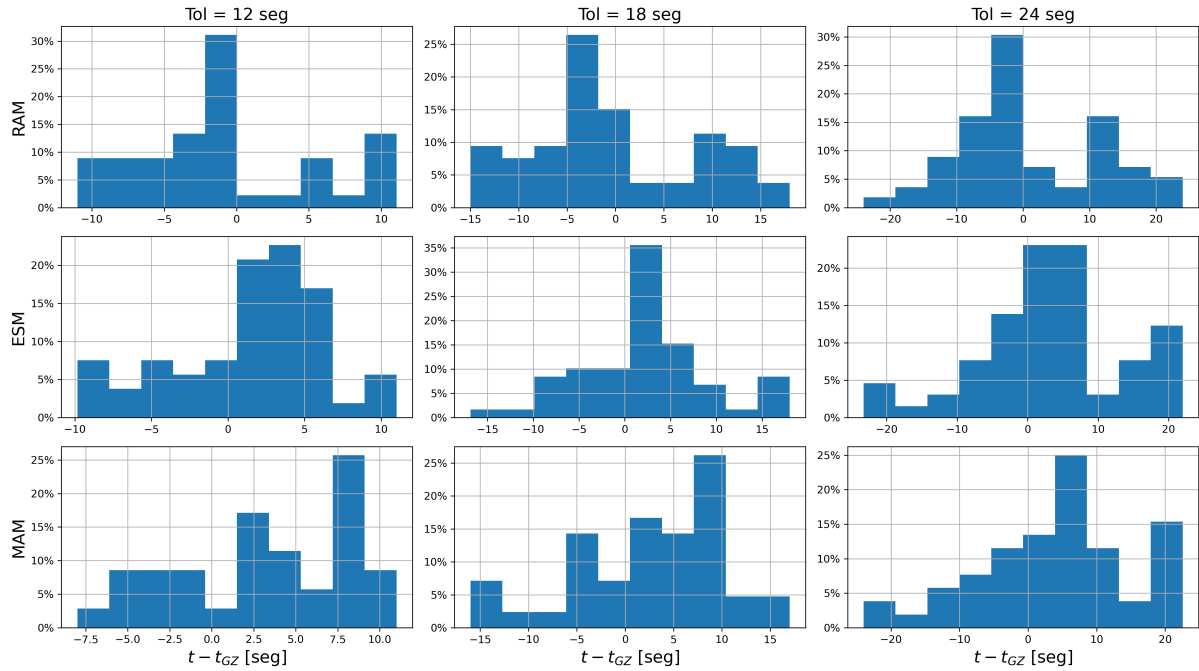


Figura 4.4: Distribución de las diferencias entre los tiempos picados por los métodos automáticos en el sismograma rotado y los tiempos de Géozur, para las distintas tolerancias consideradas.

4.3. Comparación de resultados

La Figura 4.5 muestra el porcentaje de detecciones coincidentes con Géozur para cada componente previo a la rotación (barra blanca), y el porcentaje de detecciones realizadas luego de la rotación (barra azul). Se observa un aumento significativo de las detecciones para todas las tolerancias luego de aplicado el algoritmo de reducción.

El comportamiento estadístico de los resultados se ilustra en la Figura 4.6. Cada método se comporta de manera distinta en cuanto sus valores medios y dispersiones previo y posterior a la aplicación del algoritmo de precondicionamiento. Previo a la reducción a una componente, el método RAM presenta una distribución muy alejada de una distribución normal, lo que provoca valores medios bajos pero dispersiones altas. Se observa una tendencia de los valores medios hacia los valores negativos. Tanto los valores medios como las desviaciones estándar son menores luego de la rotación, lo que es indicativo de una mejora en los resultados del método.

Luego de la rotación, ESM es el método con más cantidad de eventos coincidentes con Géozur, llegando a un 30 % para la tolerancia de 24 s. Antes de la rotación, el método presentaba un sesgo consistente en todas las componentes y en todas las tolerancias; este sesgo se redujo luego de la reducción a una componente. Tanto la media como la dispersión presentan mejoras sustanciales luego de la rotación.

En el caso de MAM, se observan mejoras significativas en cuanto a la dispersión,

el valor medio resultante de este método no disminuye, llegando a ser el método con mayor media para la tolerancia de 24 s. Sin embargo, la rotación provoca un incremento importante en la cantidad de eventos correctamente detectados.

4.4. Conclusiones

Para poder realizar una comparación y estudiar la veracidad de los resultados obtenidos, se utilizaron como referencia los arribos picados por Géozur. Para su obtención, Géozur aplicó dos métodos de detección en paralelo, cuyos resultados fueron catalogados como correctos o falsos mediante una inspección manual. Esto representa una diferencia fundamental con respecto a cómo se realizaron las detecciones en este trabajo, donde solo se realizó la detección automática mediante los algoritmos. En ese sentido, es razonable que el desempeño de los métodos cuando éstos son contrastados con las detecciones de Géozur sea bajo, independientemente del método utilizado.

Sin la rotación de la señal, aplicamos los métodos a las tres componentes y cuantificamos su desempeño para distintas tolerancias. Para las componentes horizontales, la cantidad de detecciones coincidentes con Géozur es similar tanto en U como en V . La componente W muestra mayores coincidencias con Géozur, pero las diferencias con las componentes horizontales son bajas.

Con el sismograma rotado, la cantidad de detecciones coincidentes con Géozur aumenta, pasando de una cantidad máxima del 21 % a 30 %. Además la cantidad de detecciones total disminuye, lo que indica una reducción en la cantidad de detecciones falsas, sobre todo para el método MAM. Que el número de detecciones falsas descienda es también una mejora de la eficacia de los métodos de detección, porque le da más confiabilidad a los resultados.

Además de cuantificar la cantidad de eventos, se realizó un análisis estadístico mediante las diferencias entre los tiempos picados por los métodos y los tiempos de Géozur. Idealmente, se espera que las diferencias entre los tiempos provistos por Géozur y los obtenidos mediante los métodos presenten una distribución normal centrada en cero. Sin embargo, las distribuciones presentadas en las Figuras 4.1, 4.2 y 4.3 en general presentan un comportamiento que se aleja del de una distribución normal, y los valores medios asociados no son nulos.

Los resultados de los métodos de detección demuestran que el método de reducción a una componente no solo aumenta la cantidad de eventos detectados coincidentes con Géozur, sino que además conduce a una reducción en los valores medios de las diferencias. Además, la distribución de los picados se vuelve más simétrica, mostrando similitudes con una distribución normal. Los métodos muestran una disminución de su valor medio y dispersión.

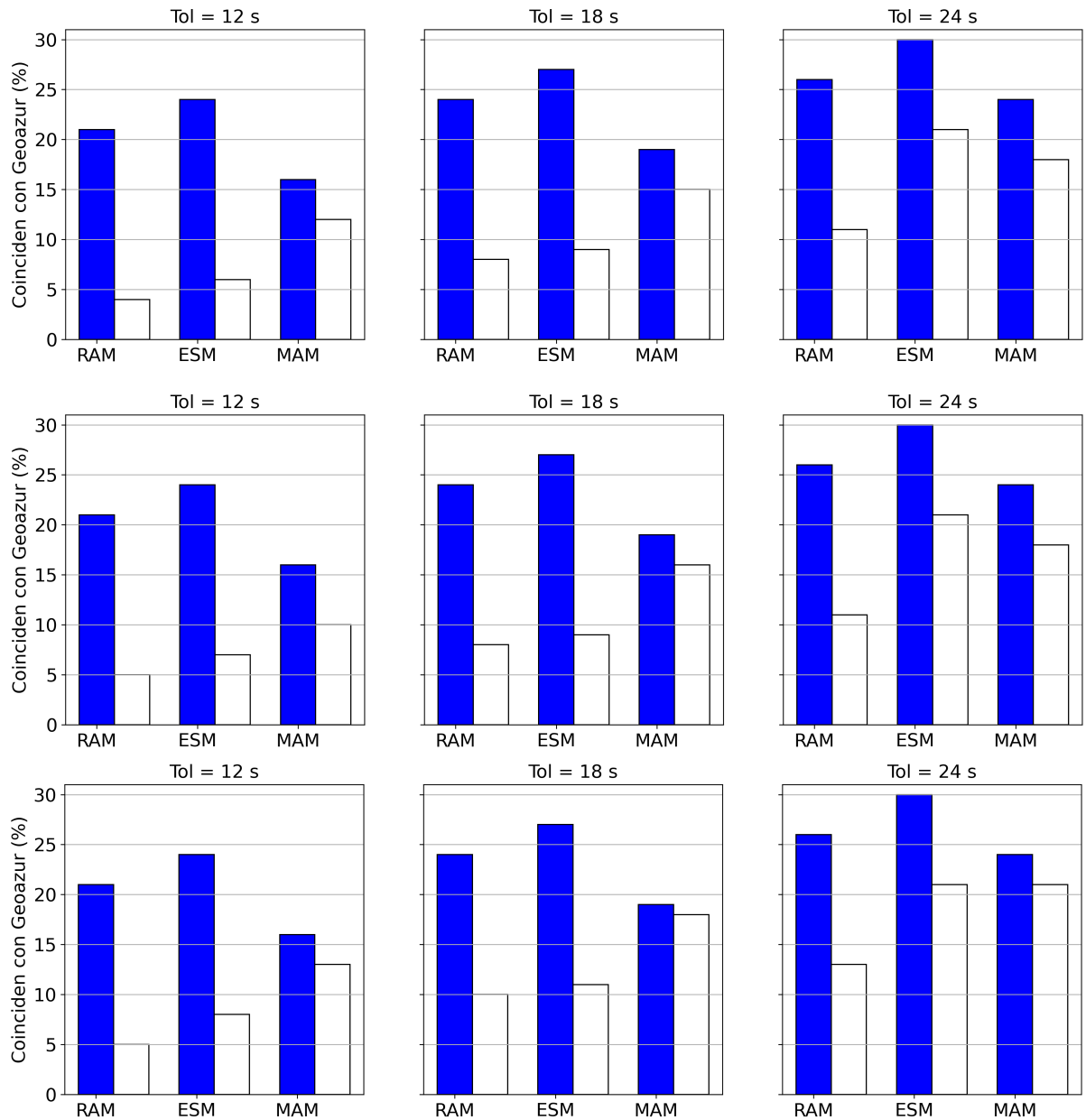


Figura 4.5: Porcentaje de detecciones coincidentes con Géoazur para cada componente sin rotar (blanco) y para la componente W rotada (azul). Fila 1: componente U . Fila 2: componente V . Fila 3: componente W .

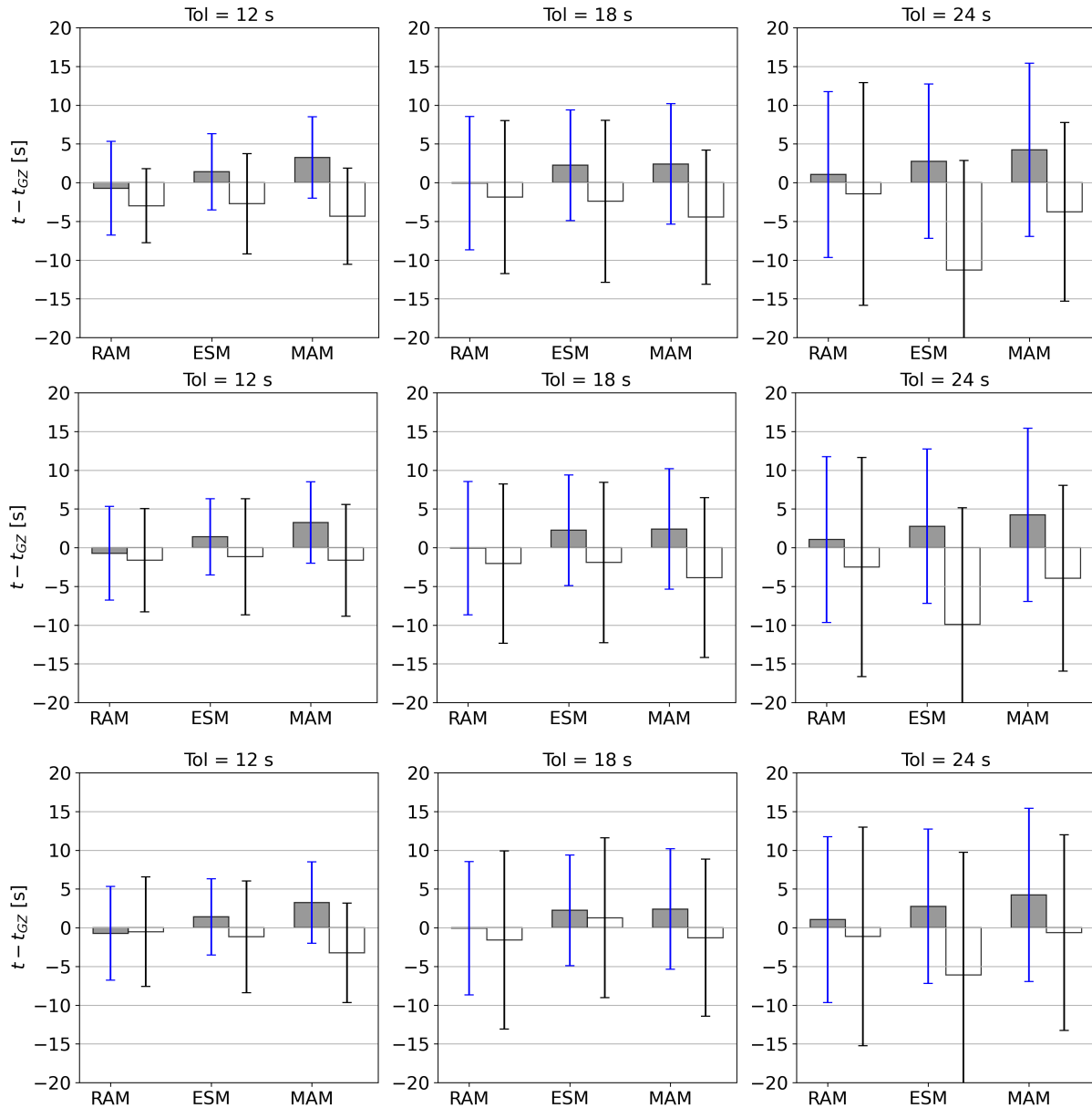


Figura 4.6: Comparación de valor medio y desviación estándar para cada componente sin rotar (blanco) y para la componente W rotada (gris). Fila 1: componente U . Fila 2: componente V . Fila 3: componente W .

Capítulo 5

Conclusiones generales y discusión

En este trabajo de Tesis de Grado se propusieron dos objetivos: el primero, investigar la eficacia de tres métodos tradicionales de detección automática de eventos sísmológicos en los datos sintéticos construidos por la misión InSight que modelan los sismogramas que son observados en el planeta Marte. El segundo, utilizar un algoritmo que permita reducir la información a una única componente, a partir de los sismogramas de tres componentes marcianos, cuya relación señal/ruido sea mayor que en el resto de las componentes, y verificar si los métodos de detección muestran una mejoría al ser aplicados sobre el sismograma rotado.

La eficacia del algoritmo de reducción a una componente pudo analizarse mediante el uso de tres eventos sísmicos marcianos concretos con comportamientos distintos. Estos eventos eran observables en las tres componentes, pero la relación señal/ruido era altamente variable dependiendo del evento y de la componente en cuestión. En ese sentido, estos sismogramas eran un ejemplo ideal para verificar si efectivamente el algoritmo propuesto construía una nueva serie de tiempo con una relación señal/ruido mayor que las observadas en cada componente por separado.

Lo observado en ese caso de estudio es que el algoritmo logra rotar efectivamente la señal, concentrando su energía en una única componente y amplificándola por sobre la del ruido, siempre y cuando la señal esté linealmente polarizada. Dos de los eventos considerados mostraban esta polarización, pero uno de ellos no pudo ser efectivamente reducido a una única componente. Sin embargo, este evento en particular sí pudo ser reducido a dos componentes. El hecho de que la energía no haya podido ser reducida a una única dirección pero sí a un único plano, es indicativo de que este evento podría contener principalmente ondas superficiales. No obstante, la relación señal/ruido no empeoró. En ese sentido, se concluye que este método amplifica la relación señal/ruido para fases linealmente polarizadas, y que en caso de que la onda contenga fases superficiales, el algoritmo no disminuye la energía.

Se analizó también la eficacia de los métodos de detección en los datos sintéticos marcianos, previo y luego de la reducción a una componente, comparando los picados resultantes de dichos métodos con los tiempos picados por Géozur. Los tiempos de arribo provistos por Géozur fueron revisados manualmente, además de ser detectados de manera automática. Ésto se contrasta con este trabajo, donde la detección solo fue automática. Por otro lado, estas detecciones contienen eventos omitidos, lo que complejiza la comparación, porque podría haber eventos que los métodos automáticos de este trabajo sí detectaron, pero Géozur no.

Cada método tuvo un comportamiento distinto. Previo a la rotación, los algoritmos tuvieron un desempeño relativamente pobre (asumiendo las detecciones de Géozur como correctas), detectando una baja cantidad de eventos en cada componente. Además, si se consideran solo los eventos correctamente detectados, los errores de detección muestran sesgos y dispersiones considerables, y con distribuciones lejos de comportarse como una distribución normal. Esto es importante, porque da lugar a dos posibilidades que no son mutuamente excluyentes: o bien los parámetros elegidos en este trabajo para los métodos, que fueron elegidos en gran parte a partir de trabajos previos presentes en la bibliografía y en parte a partir de un análisis propio, no son los ideales y existe un conjunto de parámetros tal que la detección es mejor, o bien los métodos utilizados no son los ideales para detectar eventos marcianos en este conjunto de datos. En ese sentido, se requiere más investigación para determinar cuál es el caso.

Finalmente, se comprobó que el desempeño de los tres métodos, tanto en la cantidad de eventos detectados como en el comportamiento de los errores de detección, mejoraba significativamente luego de la aplicación del algoritmo de rotación a una componente.

Los resultados de aplicar los métodos de detección a los sismogramas sin rotar y al sismograma rotado indican que el método de reducción a una componente aumenta la eficacia de los métodos de múltiples maneras: aumentan la cantidad de detecciones coincidentes con Géozur, disminuyen las falsas detecciones y aporta a que la distribución de los errores siga una distribución normal centrada en cero.

Como trabajo a futuro, se propone la aplicación de dicho algoritmo a datos reales marcianos, para verificar su eficacia en datos observados. Si este es el caso, podría generar aportes importantes al estudio del planeta Marte, puesto que la detección de arribos sismológicos es fundamental para la validación de los modelos del interior marciano. También es importante estudiar en mayor detalle los parámetros elegidos para los métodos de detección, puesto que éstos podrían presentar resultados mejores si sus parámetros asociados tienen valores distintos.

Finalmente, y reconociendo que el algoritmo de reducción a una componente muestra resultados prometedores para los métodos utilizados, se propone analizar la eficacia de otros algoritmos de detección más modernos que también son utilizados en el ámbito de la Sismología y de la microsísmica.

Bibliografía

- Aki, Keiiti y Richards, Paul G (2002). *Quantitative Seismology*.
- Allen, Rex V (1978). Automatic earthquake recognition and timing from single traces. *Bulletin of the Seismological Society of America*, 68(5):1521–1532.
- Banerdt, WB, Smrekar, S, Lognonné, P, Spohn, T, Asmar, SW, Banfield, D, Boschi, L, Christensen, U, Dehant, V, y Folkner, W y otros (2013). InSight: a discovery mission to explore the interior of Mars. *Lunar and Planetary Science Conference*, número 1719, p. 1915.
- Cauchy, Augustin (1847). Méthode générale pour la résolution des systemes d'équations simultanées. *Comp. Rend. Sci. Paris*, 25(1847):536–538.
- Clinton, John Francis, Giardini, Domenico, Lognonné, Philippe, Banerdt, B, van Driel, Martin, Drilleau, Mélanie, Murdoch, Naomi, Panning, Mark, Garcia, R, y Mimoun, David y otros (2017). Preparing for InSight: An invitation to participate in a blind test for Martian seismicity. *Seismological Research Letters*, 88(5):1290–1302.
- Earle, Paul S y Shearer, Peter M (1994). Characterization of global seismograms using an automatic-picking algorithm. *Bulletin of the Seismological Society of America*, 84(2):366–376.
- Gómez, Agustín R y Velis, Danilo R (2021). Preacondicionamiento de datos sísmicos 3C para mejorar la relación señal/ruido y favorecer la detección automática de eventos. *Libro de resúmenes de la XXIX reunión científica de la Asociación Científica de Geofísicos y Geodestas (AAGG2021)*, p. 178.
- Kulhánek, Ota (2012). *Anatomy of Seismograms: For the IASPEI/Unesco Working Group on Manual of Seismogram Interpretation*. Elsevier.
- Küperkoch, Ludger, Meier, Thomas, y Diehl, Tobias (2012). Automated event and phase identification. *New Manual of Seismological Observatory Practice 2 (NMSOP-2)*, pp. 1–52. Deutsches GeoForschungsZentrum GFZ.
- Lognonné, Philippe, Banerdt, W Bruce, Giardini, D, Pike, WT, Christensen, U, Laudet, P, De Raucourt, S, Zweifel, P, Calcutt, S, y Bierwirth, Marco y otros (2019). SEIS:

- InSight's seismic experiment for internal structure of Mars. *Space Science Reviews*, 215(1).
- Mimoun, D, Murdoch, N, Lognonné, P, Pike, WT, y Hurst, K y otros (2016). The seismic noise model of the InSight mission to Mars. *Space Sci. Rev.*
- Murdoch, Naomi, Kenda, Balthasar, Kawamura, Taichi, Spiga, Aymeric, Lognonné, Philippe, Mimoun, David, y Banerdt, William B (2017). Estimations of the seismic pressure noise on Mars determined from Large Eddy Simulations and demonstration of pressure decorrelation techniques for the InSight mission. *Space Science Reviews*, 211(1):457–483.
- Nunn, Ceri, Garcia, Raphael F, Nakamura, Yosio, Marusiak, Angela G, Kawamura, Taichi, Sun, Daoyuan, Margerin, Ludovic, Weber, Renee, Drilleau, Mélanie, y Wieczorek, Mark A y otros (2020). Lunar seismology: a data and instrumentation review. *Space Science Reviews*, 216(5):1–39.
- Rolly, Lucie, Ambrois, David, Chèze, Jérôme, Peix, Fabrice, Alemany, Hector, Mercerat, Diego, Balestra, Julien, Deschamp, Anne, y Twardzik, Cédric (2019). Contribution of Géoazur to the InSight blind test.
- Sabbione, Juan I y Velis, Danilo R (2010). Automatic first-breaks picking: New strategies and algorithms. *Geophysics*, 75(4):V67–V76.
- van Driel, Martin, Ceylan, Savas, Clinton, John Francis, Giardini, Domenico, Alemany, Hector, Allam, Amir, Ambrois, David, Balestra, Julien, Banerdt, Bruce, y Becker, Dirk y otros (2019). Preparing for InSight: Evaluation of the blind test for Martian seismicity. *Seismological Research Letters*, 90(4):1518–1534.