



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

DIVISIÓN DE INGENIERÍA EN CIENCIAS DE LA TIERRA

FACULTAD DE INGENIERÍA

REPRODUCIBILIDAD EN REPRESENTACIÓN SÍSMICA

T E S I S

QUE PARA OBTENER EL GRADO DE

INGENIERO GEOFÍSICO

P R E S E N T A

OZIEL GONZÁLEZ VERDEJO

DIRECTOR DE TESIS

SERGIO CHÁVEZ PÉREZ



MÉXICO DF, FEBRERO DE 2010

JURADO ASIGNADO

Presidente	Dr. Martín Cárdenas Soto
Secretario	M.C. David Escobedo Zenil
Vocal	Dr. Sergio Chávez Pérez
1 ^{er.} Suplente	Dr. Francisco José Sánchez Sesma
2 ^{do.} Suplente	Dr. Cinna Lomnitz Aronsfrau

RESUMEN

En el campo de la sismología de exploración hay, a nivel nacional, un creciente interés en incorporar la reproducibilidad computacional en las actividades prácticas, de investigación y docencia afines al procesamiento y representación de datos sísmicos (modelado, migración, tomografía e inversión). Esta reproducibilidad computacional implica la descripción y distribución de los elementos involucrados en un experimento numérico o una secuencia de procesamiento, y así, de manera independiente, un profesional de la práctica, investigador, instructor o estudiante, pueda repetirlo y comprobarlo. De ese modo, el usuario de una plataforma de cómputo reproducible puede realizar tres actividades importantes: estudiar y verificar los programas, reproducir los resultados, y modificar los programas y/o parámetros originales. La reproducibilidad representa beneficios académicos, técnicos y económicos porque i) disminuye la dependencia de licencias de programas comerciales para llevar a cabo investigación, ii) posibilita la adaptación, el desarrollo y transferencia de tecnología, y iii) auxilia la administración y transmisión eficiente del conocimiento. Las herramientas de cómputo reproducible, de código abierto, son un gran apoyo para investigar, educar y capacitar a estudiantes y profesionales, e incluso para la práctica profesional en las geociencias.

En este trabajo documento, adapto y divulgo la reproducibilidad en procesamiento y representación de datos sísmicos con el fin de contribuir a generar conciencia en los diferentes niveles académicos y profesionales acerca de sus beneficios, así como favorecer la asimilación de los programas de código abierto en esta área. Presento ejemplos de representación sísmica, de interés en la academia e industria, usando datos sísmicos mexicanos. Como resultado de esta investigación compruebo que es posible asimilar, adaptar, desarrollar y transferir tecnología de interés en la industria petrolera a bajo costo, empleando programas de código abierto y siguiendo un esquema de cómputo reproducible.

ABSTRACT

Within the field of exploration seismology, at national level, there is interest in integrating reproducibility in practical, educational and research activities related to processing and seismic imaging. This reproducibility implies description and distribution of elements involved in numerical experiments. Thus, a researcher, teacher or student can repeat it and prove it independently. Users of a reproducible research platform can accomplish three main activities: study and verify the programs; reproduce results independently, and modify the original programs and/or parameters. Reproducibility has important academic, technical and economical benefits because i) it diminishes dependence on commercial licenses, ii) it enables the development and transfer of technology, and iii) it is useful for the administration and efficient transmission of knowledge. In this work I document, adapt and divulge reproducibility in seismic processing and imaging in order to spread this concept and its benefits within our academic and professional environment, and to encourage the use of open source software in this area within our academic and professional environment. I present seismic imaging examples, of interest in both academic and professional environment, using Mexican seismic data. As a result of this research, I proved that it is possible to assimilate, adapt, develop and transfer technology at low cost, using open source software and following the reproducible research scheme.

CONTENIDO

RESUMEN	v
ABSTRACT	vi
CONTENIDO	vii
INTRODUCCIÓN	9
Capítulo 1. Reproducibilidad Computacional	12
1.1 Panorama de la Investigación Reproducible	12
1.2 Ventajas y desventajas de la Reproducibilidad	15
1.3 El Código Abierto, importancia y presencia en Geociencias	17
Capítulo 2. Reproducibilidad en representación sísmica	19
2.1 Programas de Código Abierto para Sismología de Exploración	19
2.2 Características principales de algunos Programas para Sismología de Exploración	21
2.2.1 <i>SEPLib</i>	21
2.2.2 <i>Seismic Unix</i>	23
2.2.3 <i>FreeUSP</i>	23
2.2.4 <i>DDS</i>	24
2.2.5 <i>Madagascar</i>	25
2.3 Documentos reproducibles en representación sísmica	26
Capítulo 3. Estudios de Caso	27
3.1 Acondicionamiento, manipulación y graficación de datos sísmicos	27
3.1.1 Lectura y acondicionamiento de datos	27
3.1.2 Manipulación y graficación de datos sísmicos	28
3.2 Postprocesamiento de datos sísmicos	30
3.2.1 Obtención de atributos sísmicos instantáneos	31
3.2.2 Enfatización de la imagen sísmica e incremento de resolución vertical	34

3.2.2.1	<i>Negativo de la segunda derivada</i>	34
3.2.2.2	<i>Multiplicador de fase instantánea</i>	38
3.2.3	Conversión tiempo a profundidad y viceversa	41
Capítulo 4. Discusión		47
CONCLUSIONES		49
REFERENCIAS		51
APÉNDICES		53
1.	Programas de código abierto útiles en geociencias	54
2.	Programas de código abierto que satisfacen las necesidades de investigación en el procesamiento de datos sísmicos y representación sísmica	55
3.	Herramientas complementarias de código abierto para sismología de exploración	56
4.	Programas de código abierto para sismología de exploración con restricciones para uso comercial	57
5.	Bibliografía sobre sismología de exploración reproducible	58
6.	Documentos reproducibles y datos disponibles en línea	59
7.	Secuencia de instrucciones <i>SConstruct</i> para lectura, manipulación, graficación de un volumen sísmico, cálculo de atributos sísmicos e incremento de resolución vertical	60
8.	Secuencia de instrucciones <i>SConstruct</i> para conversión tiempo a profundidad de datos sísmicos y viceversa	68
AGRADECIMIENTOS		71

INTRODUCCIÓN

La reproducibilidad es un principio del proceso de investigación científica. Ese principio implica la descripción detallada de los experimentos para que un investigador independiente pueda repetirlos y comprobarlos. Por su parte, la reproducibilidad computacional se logra al describir y proporcionar los elementos de un experimento numérico, es decir: el artículo de investigación, los datos, los códigos desarrollados, la plataforma computacional, la secuencia de instrucciones (*script* en inglés), la lista de parámetros y materiales auxiliares. De ese modo, un profesional de la práctica, investigador, instructor o estudiante usuario de una plataforma de cómputo reproducible puede realizar tres actividades importantes: estudiar y verificar los programas, reproducir los resultados, y modificar los programas y/o parámetros originales.

La reproducibilidad computacional representa beneficios académicos, técnicos y económicos porque i) disminuye la dependencia de licencias de programas comerciales para llevar a cabo investigación, ii) posibilita la adaptación, el desarrollo y transferencia de tecnología, y iii) auxilia la administración y transmisión eficiente del conocimiento.

Los recientes avances en las tecnologías de información y comunicación, como Internet, han dado lugar al surgimiento de políticas de acceso abierto a la información encaminadas a la difusión adecuada y actualizada de publicaciones científicas. Al tener acceso abierto a las publicaciones, el lector puede usarlas, distribuir las y mostrarlas de manera libre, bajo la condición de citarlas adecuadamente. El acceso abierto a la información es de particular importancia para los países en vías de desarrollo, ya que los problemas de distribución y difusión de sus revistas científicas impiden el acceso y uso adecuado del conocimiento que se genera localmente. Algunos proyectos que atienden la necesidad de difusión de la información son: *Public knowledge project* (disponible en: <http://pkp.sfu.ca/>), y a nivel Iberoamérica *Latindex* (disponible en: <http://www.latindex.unam.mx>) y *Scientific electronic library online* (disponible en: <http://www.scielo.org>).

En el contexto del acceso abierto al conocimiento, la reproducibilidad computacional tiene un papel importante. Su práctica comienza a volverse común, al grado de convertirse en un estándar para las publicaciones en universidades en el extranjero como la Universidad de Stanford (California, EUA), la Escuela de Minas de Colorado (Colorado, EUA) o la Universidad de Texas (Austin, Texas, EUA), por mencionar algunas, e incluso en publicaciones científicas como la revista *Geophysics*. Sin embargo, en nuestro país, a nivel académico e industrial, no ha sido suficientemente explorada.

En el campo de la sismología de exploración hay, a nivel nacional, un creciente interés en incorporar la reproducibilidad computacional en las actividades prácticas, de investigación y docencia afines al procesamiento y representación de datos sísmicos (modelado, migración, tomografía e inversión). En el Instituto Mexicano del Petróleo (IMP) y en PEMEX Exploración y Producción (PEP) tal interés ha quedado de manifiesto con el apoyo otorgado para la realización de investigaciones específicas para este caso, así como en diversas reuniones de trabajo. El presente trabajo representa parte de ese esfuerzo para impulsar dicho proyecto.

En este trabajo planteo como objetivos generales documentar, adaptar y divulgar la reproducibilidad en representación sísmica, con el fin de contribuir a generar conciencia en los diferentes niveles académicos y profesionales acerca de sus beneficios, tales como disminuir la dependencia de licencias comerciales, adaptar tecnología y auxiliar la transmisión y administración del conocimiento.

Para lograr lo anterior, fijé los siguientes objetivos particulares: a) estudiar las características de algunos programas de representación sísmica, aptos para la reproducibilidad, cuyo potencial en las áreas académicas y de investigación es destacable; b) analizar la importancia de los documentos reproducibles en representación sísmica; c) realizar tareas de representación sísmica con datos sísmicos mexicanos, similares a aquellas realizadas en la industria de exploración sísmica, para mostrar los beneficios académicos, técnicos y económicos que implica la reproducibilidad.

En el Capítulo 1, *Reproducibilidad computacional*, presento el panorama actual de la investigación computacional reproducible, expongo sus ventajas y desventajas, y destaco la importancia de los programas de código abierto en geociencias. En el Capítulo 2, *Reproducibilidad en representación sísmica*, describo algunos programas útiles para llevar a cabo la reproducibilidad en representación sísmica y enfatizo la importancia de los documentos reproducibles, por ser auxiliares valiosos en esa materia. En el Capítulo 3, *Estudios de caso*, presento resultados reproducibles de actividades de representación sísmica, de interés en la academia e industria petrolera, utilizando datos sísmicos mexicanos. Mi equipo de trabajo fue una computadora personal con sistema operativo *Linux* y el programa de cómputo reproducible, de código abierto, *Madagascar*. En el Capítulo 4, *Discusión* presento la discusión de los resultados obtenidos. Enseguida presento las conclusiones. Al final incluyo apéndices sustanciales: 1) las tablas resumen de las plataformas de cómputo reproducible, 2) las tablas resumen de documentos reproducibles, y 3) los códigos desarrollados para reproducir los resultados del presente trabajo en *Madagascar*.

CAPÍTULO 1

REPRODUCIBILIDAD COMPUTACIONAL

1.1 PANORAMA DE LA INVESTIGACIÓN REPRODUCIBLE

La reproducibilidad es uno de los principios del trabajo de investigación científica, e implica la descripción detallada de los experimentos para que un investigador independiente pueda repetirlos. Así, éste contará con los elementos necesarios para comprobar los resultados expuestos y formular sus propias interpretaciones; como consecuencia, podrá juzgar imparcialmente las implicaciones de dichos resultados por un lado y estará en condiciones de ampliarlas o refutarlas.

En aras de la reproducibilidad, la publicación de la investigación computacional, acompañada de versiones impresas de los códigos empleados, ha sido una práctica relativamente común desde finales de la década de los setenta. No obstante, en los años recientes los experimentos numéricos han alcanzado tal grado de complejidad que las descripciones escritas no bastan para que sean plenamente reproducibles; por ello, autores como Fomel y Claerbout (2009), Donoho et al. (2009), LeVeque (2009) y Stodden (2008) consideran que para garantizar la reproducibilidad se ha vuelto necesario que el investigador proporcione, de manera adecuada (actualmente a través de medios electrónicos como Internet) y debidamente documentada, los siguientes elementos derivados de la investigación computacional: 1) el artículo de investigación; 2) los datos empleados, incluyendo la descripción de los mismos y del preprocesamiento; 3) el experimento, incluyendo códigos documentados, la secuencia de instrucciones para generar los resultados, la lista de parámetros requeridos y una descripción de la metodología; 4) los resultados del experimento, como figuras y datos; 5) materiales auxiliares, como la plataforma de desarrollo o el código para su presentación en Internet. En conjunto, dichos elementos se denominan *compendio*. Stodden (2008) detalla la documentación necesaria para los elementos del *compendio* y Vandewalle (2009) propone un esquema para su publicación en línea del mismo.

La investigación reproducible tuvo como principales pioneros, hace poco más de dos décadas, a Jon Claerbout, profesor de geofísica en la Universidad de Stanford, y su equipo de investigadores del Proyecto de Exploración de Stanford (SEP, por sus siglas en inglés), un consorcio académico patrocinado por la industria petrolera. Gracias a la firme convicción de Claerbout de que la transferencia efectiva del conocimiento se lograba al hacer disponible el ambiente de desarrollo en conjunto con las instrucciones necesarias para obtener los resultados, surgió *SEPlib*, una plataforma de procesamiento de datos sísmicos, en la cual se preparaban documentos electrónicos con la característica particular de tener hipervínculos a los códigos y datos necesarios para obtener cada uno de los resultados presentados. De ese modo, el usuario tenía la posibilidad de realizar tres actividades importantes: 1) estudiar y verificar los programas; 2) reproducir de manera automática los resultados; y 3) variar los parámetros originales para analizar diferentes escenarios.

Actualmente, la reproducibilidad resulta más accesible para el usuario que al momento de su aparición gracias al uso de ambientes de cálculo numérico como *MATLAB* (Donoho et al., 2009) o herramientas de compilación de programas como *SCons* (Fomel y Hennenfent, 2007) y lenguajes de programación como *Python* (LeVeque, 2009). Incluso se aplica en otras áreas del conocimiento como: el análisis armónico (Donoho et al., 2009), la solución de problemas hiperbólicos (LeVeque, 2009) y el procesamiento y visualización de imágenes médicas (*The Insight Journal*, disponible en: <http://www.insight-journal.org/>).

Existen cuatro condicionamientos recientes en las publicaciones científicas que deben tomarse en cuenta:

1) La revista *Geophysics* incluyó, a partir de 2004, la sección “*Geophysical software and algorithms*” (en español, Programas y algoritmos geofísicos; disponible en: <http://software.seg.org/>), en la cual se publican artículos que describen un problema a resolver y el algoritmo para resolverlo. Para su reproducción, el código fuente se publica en dicho sitio de Internet. La guía para la publicación en esta sección contiene recomendaciones para elaborar un código bien documentado (disponible en: <http://software.seg.org/GEOindex.html>).

2) Revistas especializadas como *Nature* o *The Insight Journal*, incluso comienzan a solicitar el código empleado o los datos como requisito para la publicación (Stodden, 2009).

3) Algunas instituciones académicas, como la Universidad de Harvard (Stodden, 2009) y entidades gubernamentales, como el Departamento de Energía en Estados Unidos (LeVeque, 2009), solicitan que los trabajos que financian sean hechos públicos de manera abierta con el propósito de promover el aprovechamiento de los resultados y ahorrar recursos económicos. Lo anterior implica, en el caso de trabajos computacionales, que el código completo se ponga a disposición de la comunidad para su uso y modificación, bajo los términos del código abierto.

4) Stodden (2008) propuso la *Licencia de Investigación Abierta* (ORL, por sus siglas en inglés) y el *Estándar de Investigación Reproducible* [RRS, por sus siglas en inglés, (Stodden, 2009)]. Ambos describen marcos legales para cada uno de los elementos del compendio que aseguran las atribuciones y promueven investigación reproducible.

1.2 VENTAJAS Y DESVENTAJAS DE LA REPRODUCIBILIDAD

El esquema de la investigación computacional reproducible implica ventajas y desventajas para el investigador. Aquí retomo los aspectos más relevantes con base en las experiencias de los siguientes autores: Claerbout y Karrenbach (1992), Stockwell (1997), Donoho et al. (2009), LeVeque (2009) y Stodden (2009).

En primera instancia, es necesario tomar en cuenta que llevar a cabo investigación reproducible implica tiempo y esfuerzo. Esto se debe a que la producción y documentación de los elementos del compendio debe realizarse de manera cuidadosa. Además, hay que considerar que los programas deben estar en constante mantenimiento para evitar su obsolescencia. De manera adicional a la publicación en revistas técnicas o científicas, se sugiere la creación de un sitio en Internet para hacer disponibles los elementos necesarios para la difusión completa de la investigación.

A pesar del tiempo y esfuerzo requerido, los autores antes mencionados coinciden en que la reproducibilidad permite una administración eficiente del conocimiento. En primer término, el autor obtiene los mayores beneficios, puesto que en una etapa posterior a la publicación le resulta más sencillo revisar o retomar su investigación. Adicionalmente, esta práctica es de suma utilidad durante la interacción que el autor mantiene con:

- 1) coautores, quienes deben mantenerse informados acerca del progreso del proyecto;
- 2) actuales y futuros investigadores y estudiantes en el equipo de trabajo, quienes podrán asimilar el conocimiento eficientemente; además, sus contribuciones reproducibles, las cuales pueden revisarse con mayor facilidad, crearán un capital valioso incluso si se presenta la movilidad de los integrantes;

3) lectores, quienes no tendrán que implementar por su cuenta y desde el inicio el trabajo del autor antes de verificarlo y utilizarlo;

4) comités evaluadores de artículos.

El aspecto de la publicación del código computacional puede presentar objeciones por parte del investigador, por considerarlo una pieza valiosa de capital intelectual cuyo uso ajeno no se puede controlar una vez publicado (Donoho et al., 2009). No obstante, en la mayoría de los casos esto representa tres ventajas, tanto para el desarrollo del conocimiento como para el mismo investigador:

1) permite que un investigador independiente detecte errores en el código; además, se pueden disminuir los errores pequeños provocados por las diferencias entre el experimento actual y el de referencia.

2) promueve el desarrollo del conocimiento. Por ejemplo, desde la perspectiva de LeVeque (2009), hacer público el código y/o plataforma de trabajo significa una invitación a hacer uso de herramientas cuidadosamente diseñadas sin costo alguno. Al respecto comenta: “En mi experiencia cada experimento computacional da lugar a más preguntas que respuestas [...] Probablemente cualquier código valioso puede dar lugar a más publicaciones que las que el propio autor podría hacer. Si otro investigador aprovecha el código y lo aplica en una dirección distinta, debería verse como un desarrollo positivo, tanto para la ciencia como para el autor original, puesto que recibe crédito en el proceso”.

3) En efecto, hay evidencias que indican que los documentos reproducibles reciben más citas que aquellos que no lo son (Donoho et al., 2009).

Dadas las ventajas de la publicación con formato reproducible, considero que la investigación computacional en geociencias, en el ámbito nacional, debe atender las tendencias internacionales para hacer eficiente la administración y divulgación del conocimiento.

1.3 EL CÓDIGO ABIERTO, IMPORTANCIA Y PRESENCIA EN GEOCIENCIAS

Un esquema propicio para la reproducibilidad es el código abierto (*open source*, en inglés). Este término se aplica a aquellos programas a cuyo código fuente se tiene acceso, gracias a su distribución de carácter libre (los términos pueden consultarse en: <http://www.opensource.org/docs/definition.php/>), permitiéndole al usuario trabajar libremente con los programas, siempre y cuando se sujete a condiciones tales como: no revenderlos directamente o intentar impedir el acceso a los mismos. Más aún, la disponibilidad del código fuente otorga la posibilidad de modificarlo y extender sus capacidades, así como usar partes del mismo código para el desarrollo de aplicaciones propias, incluso de tipo comercial. El código abierto representa un movimiento cuya filosofía tiene por principio la mejora continua de los programas.¹

La existencia de programas de código abierto representa un gran apoyo en investigación y docencia. Al investigador le permite llevar a cabo su trabajo libremente sin compra de licencias de paquetes comerciales de altos precios, que en el caso de representación sísmica pueden ascender a miles de dólares. Asimismo, el beneficio se extiende a profesores y estudiantes, quienes así pueden contar con la posibilidad de trabajar con herramientas a las que no podrían tener acceso de otra manera. Stockwell (1997), por ejemplo, hace referencia al código abierto como una forma de conocimiento público, al igual que los libros de texto o revistas técnicas.

En el ámbito de las geociencias el desarrollo de este tipo de herramientas ha sido exitoso. Basta mencionar que tan sólo en el sitio de desarrollo de programas de código abierto *Sourceforge.net* (disponible en: <http://www.sourceforge.net>) existen poco más de cuatrocientos proyectos disponibles en este rubro (la sección de geociencias se encuentra en: http://sourceforge.net/softwaremap/trove_list.php?form_cat=567). En el Apéndice 1, en una tabla resumen, incluyo una breve selección de programas que son útiles en geociencias;

¹ Aquellas personas interesadas en la historia del código abierto encontrarán útil el *Capítulo I: The origins of open source software* en (Rosenberg, 2000).

asimismo, en el Capítulo 2 abordaré las herramientas útiles para el procesamiento de datos sísmicos y representación sísmica.

También debe destacarse que el código abierto ha tenido consecuencias positivas en el desarrollo del cómputo de alto rendimiento, al crear alternativas al alcance de universidades y centros de investigación. En la industria petrolera esta actividad implica equipos sofisticados y presupuestos elevados. Gracias al trabajo conjunto de la comunidad del código abierto se han cubierto satisfactoriamente necesidades primordiales en este ámbito, como: 1) el alto desempeño, logrado mediante comunicación eficiente entre grupos de computadoras (*cluster* en inglés), incluso de escritorio, que actúan como supercomputadora; 2) la escalabilidad, es decir, la capacidad de continuar añadiendo poder de cómputo y ganar el correspondiente incremento en productividad; y 3) la disponibilidad, o sea, mantener el sistema trabajando durante periodos prolongados sin problemas, incluso ante riesgos de posibles fallas en los componentes.² Todo ello a costos muy reducidos.

² Los detalles se pueden encontrar en el *Capítulo 8: The drive toward the mainframe* en (Rosenberg, 2000).

CAPÍTULO 2

PANORAMA DE LA INVESTIGACIÓN REPRODUCIBLE

2.1 PROGRAMAS DE CÓDIGO ABIERTO PARA SISMOLOGÍA DE EXPLORACIÓN

En la sismología de exploración, actividad fundamental en la industria petrolera, el procesamiento de datos experimenta innovaciones constantes que implican un alto nivel de complejidad computacional. A escala industrial, dichas innovaciones suelen desarrollarse con miras a crear programas comerciales o mejorar los sistemas privados de las mismas compañías petroleras; en ambos casos, no se tiene acceso a sus códigos. Por lo tanto, una estrategia económica para evitar el rezago tecnológico y la dependencia de programas comerciales en materia de educación, investigación, transferencia tecnológica e innovación es el uso de herramientas de código abierto y de tipo académico.

El panorama de los programas de código abierto en el procesamiento de datos sísmicos resulta prometedor en cuanto a su uso y con ello también el de la reproducibilidad. El esfuerzo realizado por parte de investigadores en universidades extranjeras y empresas se manifiesta hoy en día como un conjunto de paquetes disponibles para la comunidad. Lo anterior ha dado lugar a la realización continua de foros para la difusión y discusión del tema, como por ejemplo: el taller “*Open Source Exploration and Production Software*”, en el marco de la 68ª Reunión de la Sociedad Europea de Geocientíficos e Ingenieros, realizada en 2006 (disponible en: <http://www.eage.org/files/viennaworkshop2.pdf>)

Un número importante de dichos sistemas se distribuyen con licencias que permiten su uso de manera gratuita para fines académicos y comerciales. En el Apéndice 2 presento, en una tabla resumen, los programas de código abierto que cumplen con dicha característica y que satisfacen la mayoría de las necesidades de investigación en el procesamiento de datos sísmicos y representación sísmica. Dicha gratuidad resulta

particularmente importante, por una parte, para las instituciones de educación superior nacionales y por otra, para organizaciones como el Instituto Mexicano del Petróleo (IMP).

Sin embargo, en el caso del IMP, el tipo de servicios que presta a Petróleos Mexicanos (PEMEX) se considera comercial, por lo que es importante tomar en cuenta el tipo de licencia de los programas y las restricciones para su uso comercial. Dada esta particularidad, entonces, resulta conveniente, retomar, modificar y adaptar el contenido de la clasificación que se presenta en la página *Other open-source geophysical packages* (disponible en: http://www.reproducibility.org/wiki/Other_open-source_geophysical_packages).

En el Apéndice 3 presento algunas herramientas complementarias para sismología de exploración, con licencias de código abierto o equivalentes.

En el Apéndice 4 presento algunos programas de código abierto para sismología de exploración, con restricciones para su uso comercial. El lector que desee profundizar en el tema puede acudir a la página electrónica citada previamente, o consultar las memorias del taller “*Open Source Exploration and Production Software*” antes mencionado; o consultar la sección de geociencias de *Sourceforge.net* (ver inciso sección 1.3).

2.2 CARACTERÍSTICAS PRINCIPALES DE ALGUNOS PROGRAMAS PARA SISMOLOGÍA DE EXPLORACIÓN

Como se mencionó en el inciso anterior, existen programas de código abierto que satisfacen la mayoría de las necesidades de investigación reproducible en el procesamiento de datos sísmicos y representación sísmica (ver Apéndice 2). Describo las características de dichos programas, con particular énfasis en los sistemas *SEPlib* y *Madagascar* por representar la vanguardia en investigación reproducible debido a su enfoque hacia la elaboración de documentos electrónicos.

2.2.1 *SEPlib*

SEPlib fue la primera plataforma computacional desarrollada para la investigación reproducible y ha sido la base para la investigación en el Programa de Exploración de Stanford (SEP). Sus orígenes se remontan a principios de los ochentas, cuando Einar Kjartansson y Shuki Ronen, dos estudiantes en el SEP, programaron SY, un conjunto de bibliotecas en lenguaje C para procesamiento de datos sísmicos. SY tenía la característica especial de permitir el acceso a los datos y al encabezado de un archivo con formato SEG-Y, un estándar en la exploración sísmica, motivo por el cual tuvo éxito dentro del SEP. A raíz de lo anterior, comenzaron a surgir importantes contribuciones por parte de los demás colaboradores del SEP. Al poco tiempo, Jon Claerbout, fundador y encargado del SEP en aquel entonces, decidió reunir las herramientas desarrolladas y de ese modo surgió la primera versión de *SEPlib*.

Los programas que incluye *SEPlib* están escritos en lenguaje C y *Fortran*. Los parámetros necesarios para dichos programas se especifican mediante líneas de comando. Los programas actúan como filtros sobre los datos de entrada y se pueden encadenar mediante *pipes* de *Unix*.

Una de las ideas revolucionarias que engloba este sistema fue la vinculación de los documentos de texto con la secuencia de instrucciones necesarias para crear los resultados presentados en dicho texto (ver Claerbout y Karrenbach, 1992). En este sistema, dicha vinculación se logra utilizando archivos *GNUmake* para ejecutar las secuencias de procesamiento (la publicación de Schwab et al. (2000) describe las instrucciones que deben incluir los archivos *GNUmake* y las mismas se encuentran en: <http://sepwww.stanford.edu/doku.php?id=sep:research:reproducible>). De esta manera un artículo, reporte o curso en formato electrónico, escrito en el procesador de palabras *Tex*, puede incluir imágenes que están ligadas al archivo *GNUmake* correspondiente, mediante una interfaz gráfica, permitiéndole al usuario ejecutar alguna de las siguientes tareas: eliminar, generar o ver la figura. De manera adicional, el usuario puede eliminar del directorio los archivos intermedios o revisar de manera individual los archivos asociados al documento. En conjunto, las actividades anteriores le permiten al usuario comprender mejor la relación existente entre los programas que fueron utilizados y los resultados.

Otra idea trascendente en torno al desarrollo de este proyecto fue el interés por hacer disponible el ambiente computacional local del SEP a quienes contaran con una computadora con sistema operativo tipo Unix. Una lista de las herramientas, distribuidas de manera libre, necesarias para trabajar con dicho ambiente se encuentra disponible en: <http://sepwww.stanford.edu/doku.php?id=sep:software:septour>.

Actualmente, *SEPlib* es un sistema completo que permite realizar una gran variedad de tareas de procesamiento de datos sísmicos y representación sísmica. Permite manipular datos bidimensionales y tridimensionales, con geometrías regulares e irregulares. Es importante mencionar que tiene soporte para cómputo paralelo mediante el protocolo MPI (*Message passing interface*) y OpenMP (*Open Multi-Processing*).

2.2.2 Seismic Unix

Es un sistema para plataformas tipo *Unix* desarrollado por el Centro para Fenómenos de Onda (CWP, por sus siglas en inglés) de la Escuela de Minas de Colorado (CSM, por sus siglas en inglés), Golden, Colorado, EUA. Comenzó cuando Shuki Ronen, estudiante del SEP, llevó SY a la CSM durante una estancia en 1983. Junto a Jack K. Cohen transformaron el programa, renombrándolo *SU*, para hacerlo soportable, extensible y exportable. En un principio se concibió como un ambiente sencillo de procesamiento y desarrollo para el geofísico con habilidades en la programación, sin embargo, no se trataba de una versión lo suficientemente fácil de manejar como para hacer una distribución pública.

Gracias a la adopción del ANSI C (Instituto Americano de Estándares Nacionales) como el lenguaje estándar, y el trabajo cuidadoso y bien documentado de Dave Hale, *SU* se convirtió en un ejemplo a seguir de estilo de programación y administración en el cómputo científico.

A partir de 1992, *SU* comenzó a distribuirse de manera pública, dejando de ser desarrollado exclusivamente por el CWP. En la actualidad cuenta con más de 300 archivos ejecutables y se usa con mucho éxito por distintos grupos en el mundo de la academia e industria.

2.2.3 FreeUSP

FreeUSP es la versión libre de *USP* (*Unix Seismic Processing*), desarrollado por el Centro de Investigación de Tulsa, de Amoco (TRC, por sus siglas en inglés), Oklahoma, EUA. El *USP* se comenzó a desarrollar en 1985, como alternativa al Sistema de Información Sísmica del TRC, por un equipo integrado originalmente por Don Wagner, Paul Gutowski y Martin Smith, del TRC, y el profesor Robert Herrman, de la Universidad de Saint Louis, Missouri, EUA.

El *USP* ganó simpatía rápidamente por ser un sistema eficaz de reproducibilidad en el ambiente productivo debido a sus capacidades e integración eficiente de nuevas herramientas. En el transcurso del primer año ya contaba con 50 programas y era capaz de procesar los datos sísmicos bidimensionales. En 1993 Amoco tomó la decisión de usar de manera exclusiva el *USP* como sistema interno, demostrando tanto su efectividad técnica como su impacto económico positivo. A mediados de los noventas el *USP* contenía herramientas de procesamiento de datos tridimensionales, datos multicomponente y para el cálculo de atributos sísmicos.

En 1999 Amoco y Bristish Petroleum (BP) se fusionaron, y el proyecto USP continuó desarrollándose. En 2000 se propuso hacer una versión pública del sistema con el objetivo de proporcionar a los estudiantes y científicos un sistema funcional apto para el desarrollo y procesamiento; y dentro del esquema del código abierto, agregar las contribuciones de externos para mejorar el sistema.

2.2.4 DDS

El *DDS (Data Dictionary System)* es un paquete que ha probado su utilidad en la investigación en el área de representación sísmica en Amoco y BP por más de una década. Fue diseñado y puesto en marcha por Randall Selzer, del TRC, en 1995, cuando fue indispensable conseguir un óptimo desempeño en el cómputo paralelo. Su modelo a seguir fue *SEPlib*, debido a que el manejo de la información se hace a través de volúmenes grandes de datos a diferencia de los que trabajan con las trazas de manera individual, como *SU* y *USP*.

El *DDS* incluye un subconjunto de programas de representación sísmica, desarrolladas por el Grupo de Exploración y Tecnología de BP en Houston, Texas, EUA. El sistema es portable y permite el acceso eficiente a los datos entre las plataformas de trabajo. Una de sus ventajas clave es la capacidad de leer y escribir en varios formatos: *SEPlib*, *USP*, *SU*, SEG-Y, y aquellos definidos por el usuario. Dicha capacidad

resulta muy útil puesto que permite hacer operaciones intermedias con otros sistemas de procesamiento, sirviendo como un puente entre distintos ambientes. El *DDS* también permite la documentación de los datos para que puedan ser leídos y entendidos. Se distribuye como código abierto desde 2003.

2.2.5 Madagascar

Madagascar es un sistema completo que comenzó a desarrollarse en 2003 por un grupo de geofísicos, ingenieros petroleros y computólogos. Integra los elementos de *SEPlib* y *Seismic Unix* bajo un esquema de desarrollo de cómputo moderno.

El sistema consta de dos niveles:

1) El primero está compuesto por poco más de 300 programas de bajo nivel, desarrollados en el lenguaje *C*, que actúan como filtros sobre los datos y que pueden encadenarse mediante *pipes* de *Unix*. La lista de programas con la documentación y los códigos correspondientes están disponibles en: <http://www.reproducibility.org/RSF/>. La clasificación de los programas de acuerdo al tipo de tareas está disponible en: http://www.reproducibility.org/wiki/Task-centric_program_list. La lista de los programas de acuerdo al tipo de tareas de procesamiento de datos sísmicos de reflexión y representación sísmica se encuentra en: http://www.reproducibility.org/wiki/Seismic_task-centric_program_list.

2) El segundo está compuesto por flujos o historias de procesamiento de alto nivel, descritos con el lenguaje de programación *Python* y ejecutados mediante *SCons* (de manera alternativa a los archivos *GNUmake* utilizados en *SEPlib*). Dichas historias de procesamiento pueden ser verificadas, intercambiadas y modificadas por los usuarios (Fomel y Hennenfet, 2008).

Madagascar usa un formato simple, flexible y universal de datos que puede operar con conjuntos de datos muy grandes. El formato RSF (regularmente muestreado, por sus siglas en inglés) se tomó de *SEPlib*.

La universalidad del formato permite compartir herramientas de procesamiento de propósito general con científicos de otras disciplinas, como ingenieros petroleros que trabajan en simulaciones de yacimientos de gran escala.

2.3 DOCUMENTOS REPRODUCIBLES EN REPRESENTACIÓN SÍSMICA

Si bien la investigación se puede hacer reproducible al hacer público el *compendio* derivado de la investigación (Capítulo 1.1), autores como Claerbout y Karrenbach (1992), Schwabb et al. (2000) y Fomel y Hennenfet (2008) promueven una idea que lleva al límite esta práctica: los documentos reproducibles. Plataformas como *SEPlib* y *Madagascar* están encaminadas a la publicación con dicho formato. Este tipo de documentos tiene la característica de tener los resultados vinculados a las secuencias de instrucciones (*script*), que a su vez tienen vínculos a los códigos de los programas empleados. El aprendizaje apoyado en este tipo de documentos permite una mejor comprensión del cómputo implicado.

En el Apéndice 5 presento las referencias de algunos libros de sismología de exploración que tienen la particularidad de ser reproducibles.

En el Apéndice 6 presento algunos documentos reproducibles que incluyen ejemplos de los reportes mediante los cuales los consorcios académicos difunden su investigación con los patrocinadores, así como tesis académicas y artículos de investigación.

CAPÍTULO 3

ESTUDIO DE CASOS

En el ambiente de exploración petrolera, nacional e internacional, hay gran interés en temas de acondicionamiento, manipulación, graficación y postprocesamiento de datos sísmicos. En este Capítulo presento ejemplos que ilustran la reproducibilidad en esas actividades, con datos sísmicos de la región de Chicontepec y del Golfo de México, con el propósito de destacar los beneficios del uso de programas de código abierto y documentos reproducibles.

3.1 ACONDICIONAMIENTO, MANIPULACIÓN Y GRAFICACIÓN DE DATOS SÍSMICOS

El acondicionamiento, manipulación y graficación de datos sísmicos son actividades básicas en la sismología de exploración. Utilicé *Madagascar* y el volumen de datos sísmicos terrestres Agua Fría – Coapechaca - Tajín, en Chicontepec, proporcionado por PEMEX Exploración y Producción (PEP), para efectuar las siguientes tareas: 1) lectura y acondicionamiento de datos; y 2) manipulación y graficación del subvolumen, cortes verticales y en tiempo.

3.1.1 Lectura y acondicionamiento de datos

Los archivos de datos sísmicos suelen tener formato SEG-Y, un estándar en la industria de la exploración sísmica. Los programas de código abierto como *Madagascar*, *SEPlib* o *Seismic Unix* utilizan formatos internos para el procesamiento y visualización de los datos, por lo que es necesario hacer el cambio de formato correspondiente.

Utilicé *Madagascar* para la lectura del volumen sísmico Agua Fría-Tajín-Coapechaca, con formato SEG-Y y cambio al formato interno RSF (formato regularmente muestreado, por sus siglas en inglés). El formato RSF almacena los valores numéricos de los datos en formato binario, y utiliza encabezados con formato de texto para almacenar las propiedades de los datos (el lector interesado puede consultar la Guía del formato RSF, disponible en: http://www.reproducibility.org/wiki/Guide_to_RSf_file_format). En el Apéndice 7 presento el código correspondiente (líneas 11 a 13); el programa que utilicé para la conversión de formato es *sfsegypread*.

3.1.2 Manipulación y graficación de datos sísmicos

Una vez hecha la conversión de formato, modifiqué el encabezado del archivo RSF, con el programa *sfput*, para especificar las dimensiones del volumen. El archivo final se puede procesar y visualizar con los programas de *Madagascar*. Para ejemplificar el uso de programas, obtuve subvolúmenes, cortes horizontales y verticales con el programa *sfwindow* y apliqué control automático de ganancia (AGC, por sus siglas en inglés) al subvolumen original con el programa *sfagc*. Para generar gráficas de densidad variable usé el programa *sfgrey*. Los resultados se grafican en pantalla con el programa *sfxtpen*.

En la Figura 1 presento las gráficas de un subvolumen de Agua Fría-Tajín-Coapechaca, antes y después de aplicar el control automático de ganancia. En la Figura 2 presento la gráfica correspondiente a la línea longitudinal 2300 y un corte horizontal para el tiempo 1.448 s. El código para generar los resultados lo presento en el Apéndice 7, y corresponde las líneas 15 a 46.

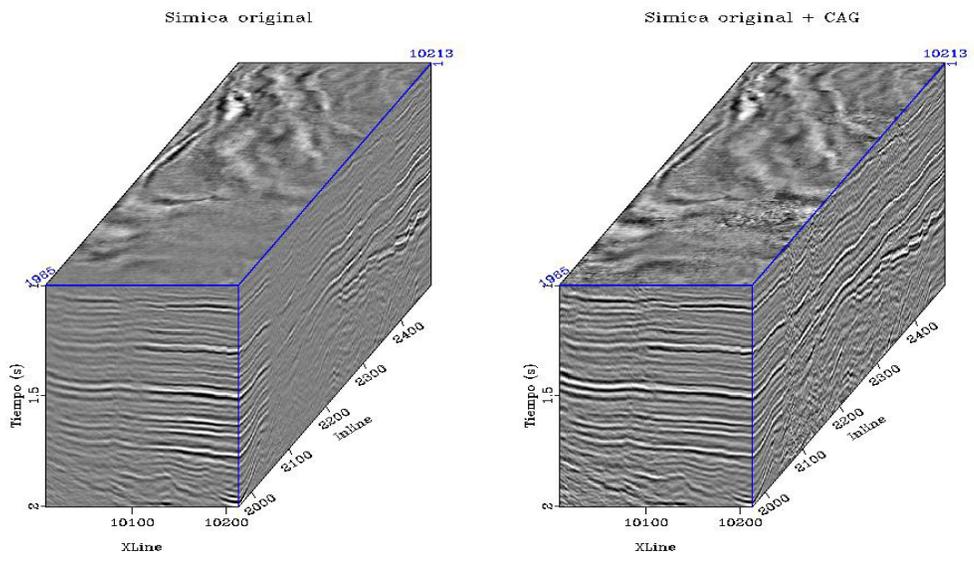


Figura. 1 Subvolumen Agua Fría-Coapechaca-Tajín, en Chicontepec, antes y después de aplicar control automático de ganancia. La distancia entre trazas es de 25 m.

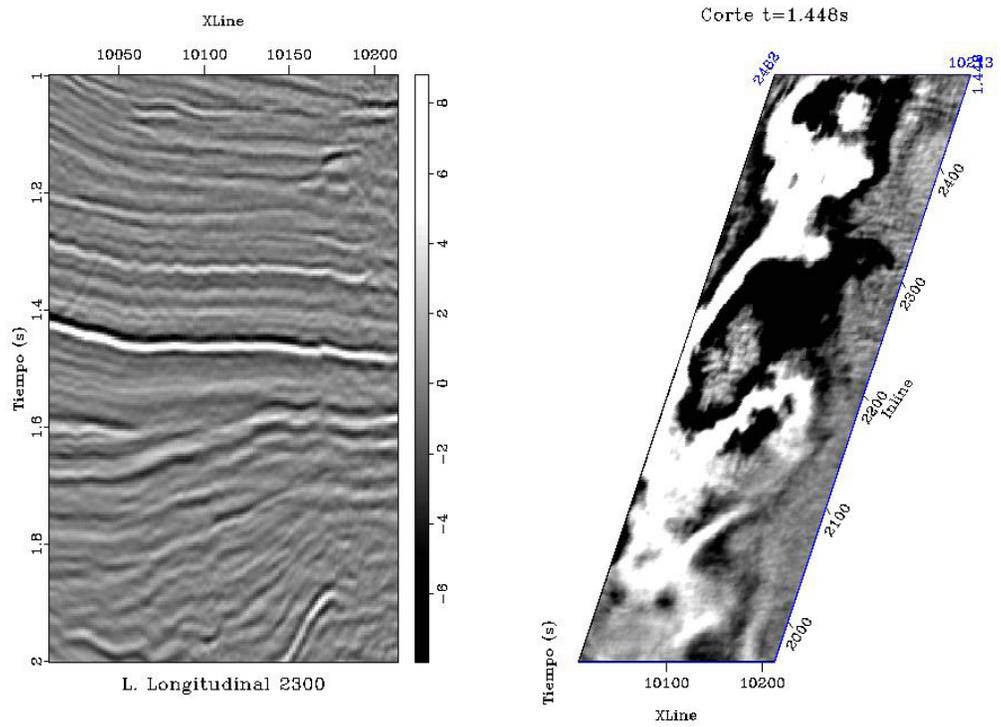


Figura. 2 Línea longitudinal 2300 del subvolumen y corte en tiempo 1.448 s. La distancia entre trazas es de 25 m.

Una manera de presentar los datos que permite estudiar patrones estructurales es la animación o graficación consecutiva de cortes del volumen. En la Figura 3 presento cuatro tomas instantáneas de una animación de cortes en tiempo. El código lo presento en el Apéndice 7 (líneas 47 a 67).

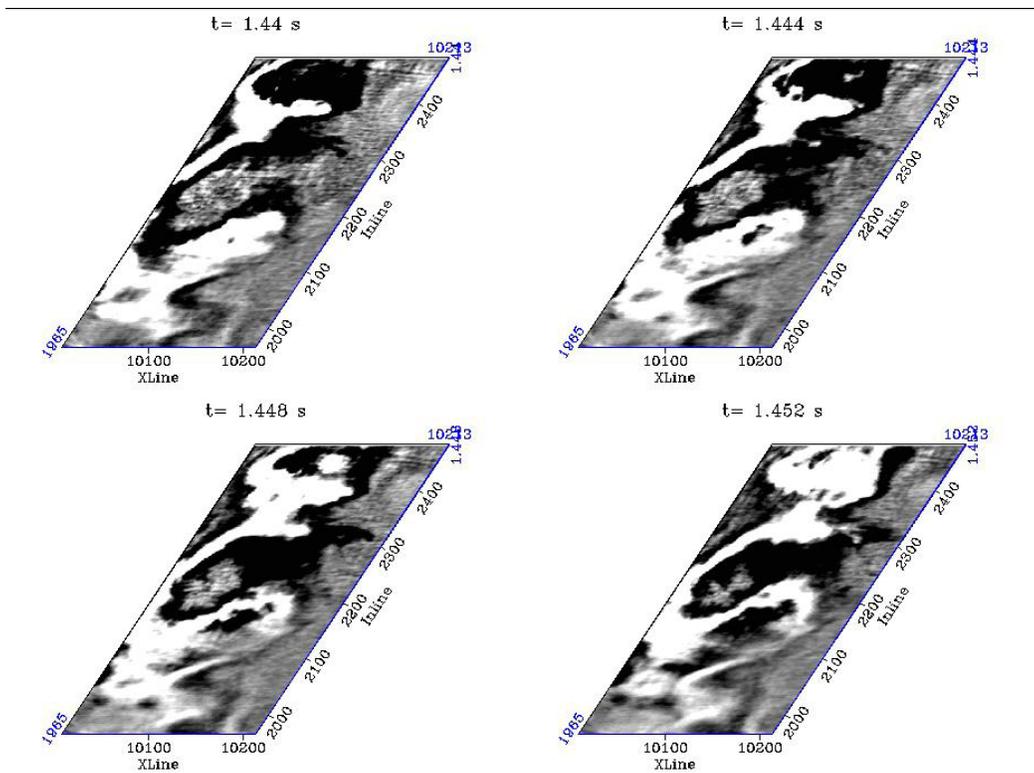


Figura 3. Imágenes instantáneas de una animación de cortes horizontales del volumen Agua Fría-Coapechaca-Tajín, en Chicontepec.

3.2 Postprocesamiento de datos sísmicos

El tema de postprocesamiento de datos sísmicos es de interés en el IMP y PEP. En este apartado presento ejemplos que ilustran la reproducibilidad en los siguientes temas:

- 1) obtención de atributos sísmicos instantáneos;
- 2) enfatización de la imagen sísmica e incremento de resolución vertical;
- 3) conversión tiempo a profundidad y viceversa.

3.2.1 Obtención de atributos sísmicos instantáneos

Los atributos sísmicos instantáneos se utilizan para reforzar visualmente o aislar rasgos de interés durante las etapas de interpretación estructural y estratigráfica de la información sísmica. Estos atributos se derivan del análisis complejo de la traza sísmica, que denoto como $\mathbf{x}(t)$, que consiste en representarla como la parte real de una traza compleja, $\mathbf{X}(t)$. La parte imaginaria de la traza compleja, $\mathbf{x}^*(t)$, se conoce como cuadratura y se obtiene al aplicar un cambio constante en la fase de la traza sísmica original. En la ec. (1) presento la relación entre la traza compleja, la traza sísmica y la cuadratura:

$$\mathbf{X}(t) = \mathbf{x}(t) + i \mathbf{x}^*(t) \quad (1)$$

Con excepción del cambio de fase en la cuadratura, un geofísico observaría las mismas características, es decir la misma coherencia y la misma relación señal-ruido. Un método para calcular la cuadratura es la transformada de Hilbert (artículos como Taner et al. (1979) y Barnes(2007) contienen información detallada al respecto).

Si $\mathbf{x}(t)$ y $\mathbf{x}^*(t)$ son conocidas, la traza compleja se puede representar como en el término de la derecha de la siguiente expresión:

$$\mathbf{X}(t) = \mathbf{x}(t) + i \mathbf{x}^*(t) = A(t) e^{i\Theta(t)} \quad (2)$$

El término $A(t)$ se conoce como envolvente o amplitud instantánea, y es el módulo de la traza sísmica:

$$A(t) = (\mathbf{x}^2(t) + \mathbf{x}^{*2}(t))^{1/2} = |\mathbf{X}(t)| \quad (3)$$

La fase instantánea, $\Theta(t)$, se define como:

$$\Theta(t) = \tan^{-1}\left(\frac{x^*(t)}{x(t)}\right) \quad (4)$$

A partir de la ecuación 2 se puede obtener una relación entre la traza sísmica y los atributos sísmicos instantáneos:

$$x(t) = A(t) \cos(\Theta(t)) \quad (5)$$

La amplitud instantánea $A(t)$ contiene la información asociada a los contrastes litológicos y la fase $\Theta(t)$ se asocia a la continuidad de los eventos.

Barnes (2007) resume el concepto de los atributos sísmicos instantáneos de la siguiente manera: 1) la amplitud instantánea se puede entender como el valor máximo de amplitud que puede obtenerse, para un tiempo dado, como consecuencia del cambio constante de fase de la traza sísmica; 2) la fase instantánea representa el cambio de fase de la traza sísmica necesario para que alcance un valor máximo en un instante de tiempo dado; es independiente de la amplitud. Una manera práctica de representarla es graficando su coseno, lo cual funciona como control automático de ganancia.

A partir de una sección vertical del subvolumen del volumen Agua Fría-Coapechaca-Tajín (ver Figura 2), obtuve las secciones de coseno de la fase instantánea y envolvente (Figura 4). La envolvente la generé con base en la partir de la ec. (3), para ello, calculé la cuadratura de la traza sísmica con el programa *sfenvelope* de *Madagascar*, con la opción de Transformada de Hilbert. El coseno de la fase instantánea lo obtuve a partir de la relación expresada en la ec. (5). La fase instantánea la calculé a partir del coseno de la fase instantánea, con la función coseno inverso. Presento el código en el Apéndice 7, en las líneas 68 a 90.

La figura 5 es una comparación entre una traza sísmica original, la cuadratura de la traza sísmica y la envolvente, con polaridad positiva y negativa, para el intervalo de tiempo 1.0 a 1.2 s. Presento el código en el Apéndice 7, en las líneas 91 a 102.

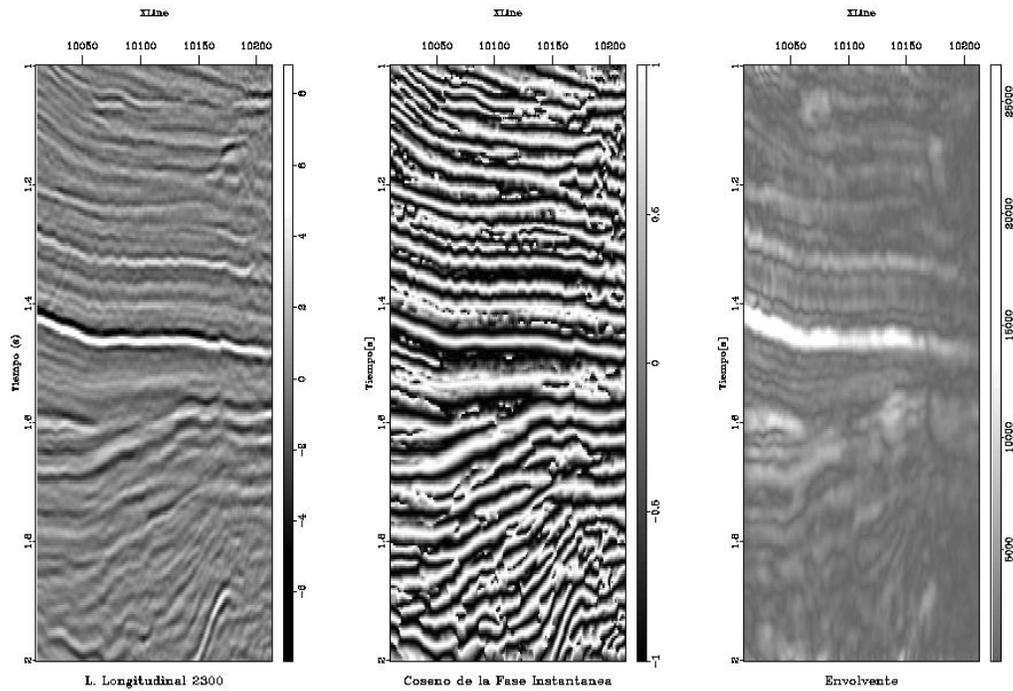


Figura 4. De izquierda a derecha: línea longitudinal 2300 del subvolumen Agua Fría-Coapechaca-Tajín, coseno de la fase instantánea y envolvente. La distancia entre trazas es de 25 m.

Traza Original; Cuadratura; Envolverte [Pos. y Neg.]

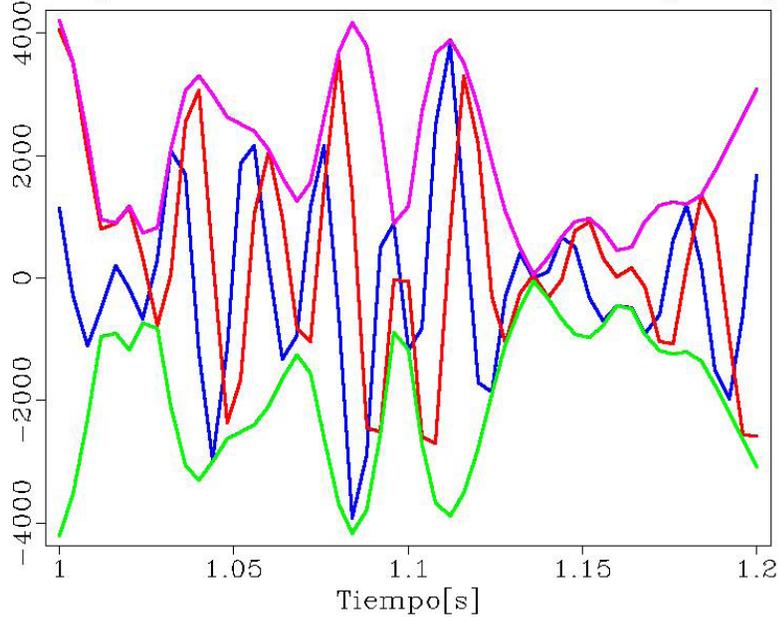


Figura 5. Detalle de traza original (azul), cuadratura (rojo) y envolverte (positiva en rosa y negativa en verde). El eje vertical es la amplitud de la señal.

3.2.2 Enfatización de la imagen sísmica e incremento de resolución vertical

Una de las actividades en las que hay gran interés es la enfatización de imágenes sísmicas con la finalidad de mejorar la interpretación posterior. Presento ejemplos de dos métodos que se emplean para esa tarea: el primero, la obtención del negativo de la segunda derivada de los datos; y el segundo la incorporación de un multiplicador de fase instantánea N , que tiene carácter heurístico.

3.2.2.1 *Negativo de la segunda derivada*

Calcular el negativo de la segunda derivada de los datos sísmicos permite mejorar la continuidad y definición de los eventos en los datos sísmicos. El fundamento es que el operador de derivada en el dominio de la frecuencia implica la multiplicación de la Transformada de Fourier por el producto del número imaginario i

y frecuencia ($i\omega$, por lo tanto, la segunda derivada es igual a multiplicar por $(i\omega)^2 = -\omega^2$; como la polaridad se invierte, se debe multiplicar por -1 para obtener la polaridad original). El resultado de esta operación en datos sísmicos, en el dominio de las frecuencias, es un cambio en la amplitud de las frecuencias, proporcional a la amplitud original y a la frecuencia, que produce un efecto de corrimiento de la frecuencia dominante hacia las altas frecuencias. De modo similar, el operador de cuarta derivada (que implica la multiplicación de la Transformada de Fourier por $(i\omega)^4 = \omega^4$ permite correr más la frecuencia dominante hacia altas frecuencias. En ambos casos hay un incremento en el ruido, sin embargo, el resultado final es razonable.

El operador del negativo de la segunda derivada está incluido en paquetes comerciales. Implementé el flujo de procesamiento en *Madagascar* para obtener los mismos resultados, utilizando datos proporcionados por PEMEX (Agua Fría-Coapechaca-Tajín, en Chicontepec).

En la Figura 6 presento el resultado del negativo de la segunda derivada y de la cuarta derivada, con control automático de ganancia, para la línea longitudinal 2300 del subvolumen que seleccioné previamente. El resultado de la segunda derivada es satisfactorio en términos de continuidad y definición de eventos; el de la cuarta derivada también lo es, pero presenta un incremento de ruido más notorio. El código lo presento en el Apéndice 7, en las líneas 104 a 115.

En la Figura 7 presento la comparación de los espectros de amplitudes, normalizados, de las secciones antes y después de obtener el negativo de la segunda derivada y la cuarta derivada. Como resultado de la aplicación de los operadores hay corrimiento de la frecuencia dominante hacia las altas frecuencias. Presento el código en el Apéndice 7, en las líneas 117 a 125.

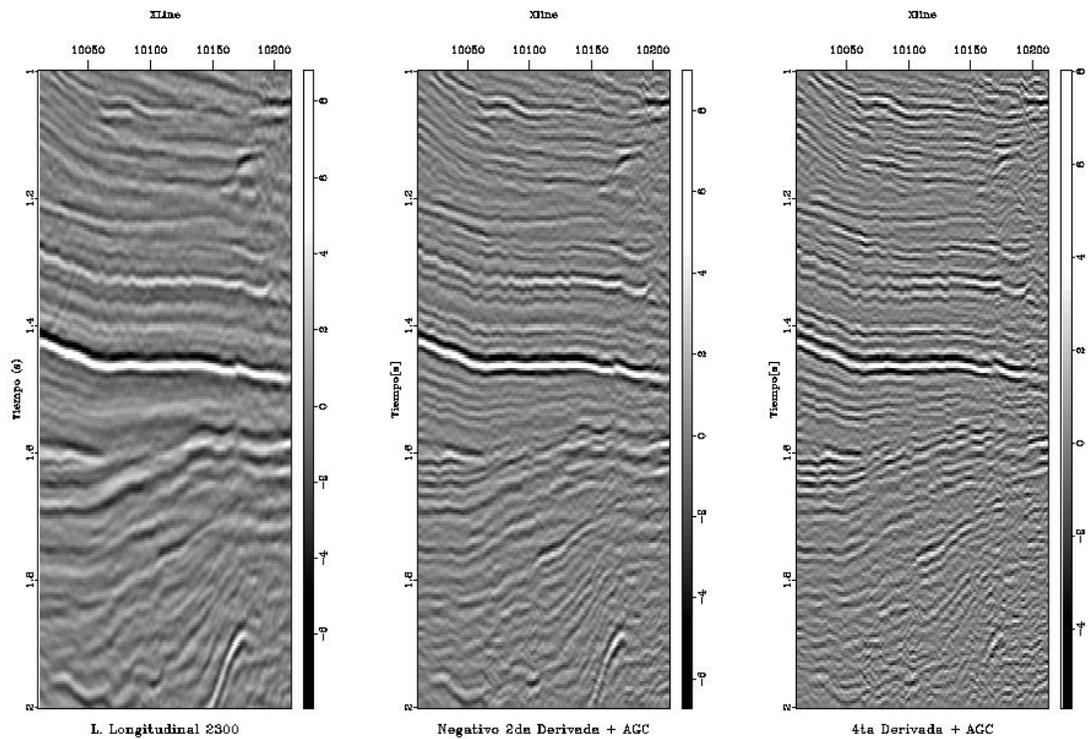


Figura. 6. Izquierda: sección original. Centro: negativo de la segunda derivada; es notable la mayor definición y continuidad de los eventos. Derecha: cuarta derivada; también hay incremento en la resolución, pero el ruido aumenta. La distancia entre trazas es de 25 m.

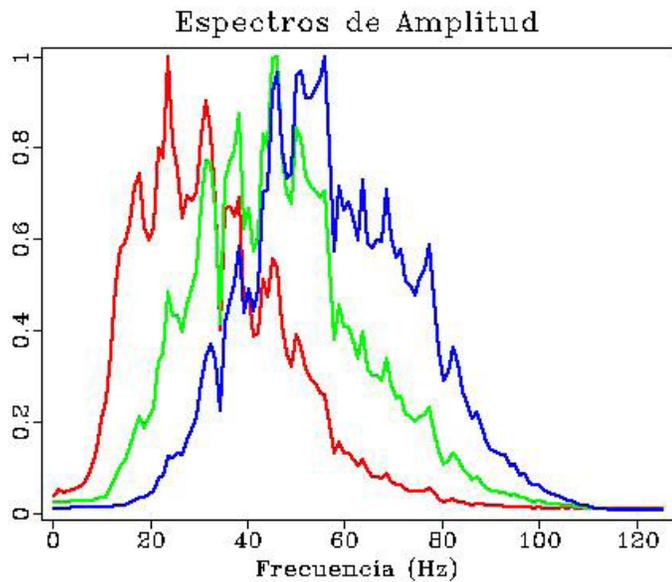


Figura. 7. Espectros de amplitud, promediados y normalizados, antes (rojo) y después de obtener el negativo de la segunda derivada (verde) y cuarta derivada (azul). Nótese el corrimiento de las frecuencias dominantes hacia altas frecuencias después de aplicar los operadores. El eje vertical es la amplitud normalizada.

Consideraré utilizar la media aritmética de los datos originales y los derivados para incrementar el ancho de banda de los datos sísmicos. Sin embargo, dicha operación no genera una mayor enfatización de los eventos con respecto al negativo de la segunda derivada.

En la Figura 8 presento la comparación de la sección original, el negativo de la segunda derivada y la media aritmética ($\frac{1}{2}$ original + $\frac{1}{2}$ negativo de la segunda derivada). El código lo presento en el Apéndice 7, en las líneas 127 a 135. En la Figura 9 presento la comparación de los espectros de amplitudes, normalizados, de las secciones antes y después de obtener el negativo de la segunda derivada y la media aritmética. Presento el código en el Apéndice 7, en las líneas 137 a 144.

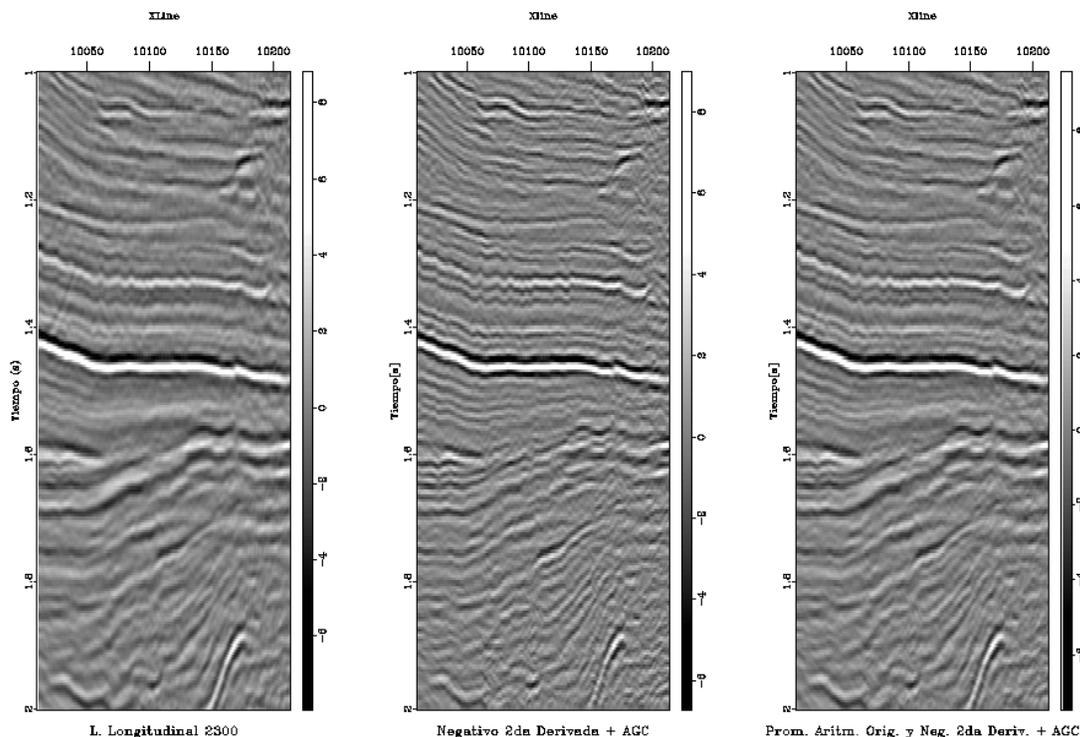


Figura. 8. Izquierda: sección original. Centro: negativo de la segunda derivada. Derecha: media aritmética de la sección original y el negativo de la segunda derivada; algunos detalles se enfatizan con respecto a la sección original. La distancia entre trazas es de 25 m.

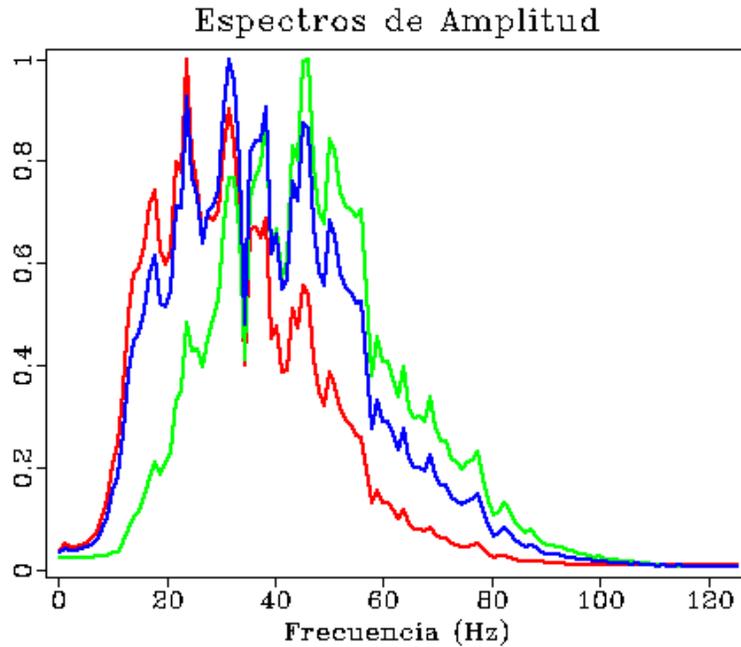


Figura 9. Espectros de amplitud antes (rojo) y después de obtener el negativo de la segunda derivada (verde) y la media aritmética (azul). Después de obtener la media hay incremento en el ancho de banda, pero no implica mayor enfatización con respecto a la segunda derivada. El eje vertical es la amplitud normalizada.

3.2.2.2 Multiplicador de fase instantánea

Recientemente, Stark (2009) propuso un algoritmo que permite incrementar el número de cimas y valles en una traza sísmica. Los resultados son de mayor resolución y enfatizan estructuras geológicas, como fallas y pliegues. El algoritmo consiste en incorporar un multiplicador N , o acelerador de la fase instantánea en la ecuación que convierte la amplitud y fase instantánea en una traza sísmica (ec. 5), y se expresa como en la ec. (6):

$$x_N(t) = A(t) \cos(N\Theta(t)) \quad (6)$$

donde $x_N(t)$ es la traza sísmica modificada, $A(t)$ es la amplitud instantánea, $\theta(t)$ es la fase instantánea y N es el multiplicador o factor de aceleración del giro de la fase instantánea. Si $N=1$, entonces la traza

generada corresponde a la traza original, i.e. $x(t) = x_N(t)$. Si $N = 2$, la frecuencia se duplica y el número de crestas y valles se incrementa el doble; si $N=3$ se incrementa en tres, y así sucesivamente.

Adapté la idea de Stark (2009) con formato reproducible. A manera de introducción, en la Figura 10 presento un caso sintético, partiendo de una función coseno con frecuencia de 10 Hz, y el resultado para $N = 2$. En este caso, la señal modificada tiene una frecuencia del doble de la original, es decir, 20 Hz.

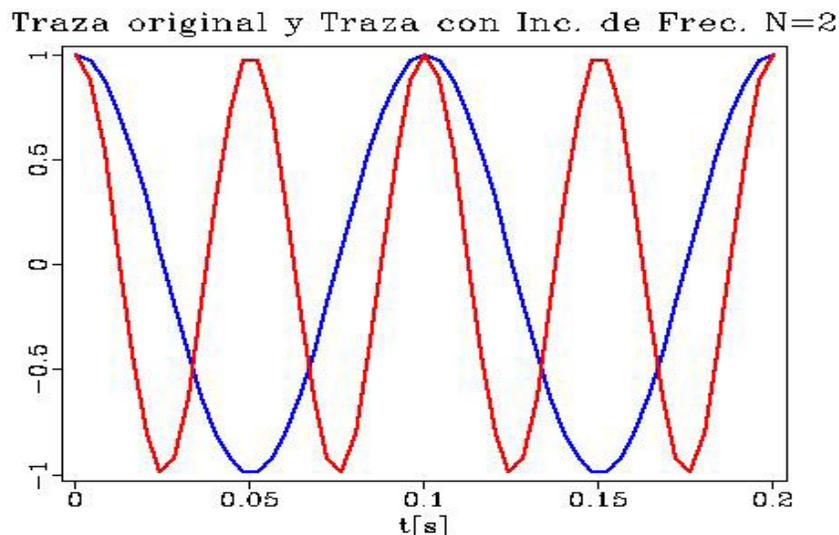


Figura. 10 Coseno antes (azul) y después del incremento de frecuencia con $N = 2$ (rojo). El eje vertical es la amplitud.

En la Figura 11 presento los resultados para $N = 2$ y $N = 3$ en la sección del subvolumen sísmico Agua Fría-Coapechaca-Tajín, que he utilizado anteriormente. Los resultados para $N = 2$ muestran mayor resolución y mejoran la continuidad de eventos. No obstante, debe tomarse en cuenta la aparición de artefactos e incremento de ruido que no corresponden necesariamente a la geología del subsuelo. Estructuras como pliegues y discontinuidades pueden hacerse más notorios con este operador. El caso $N = 3$ no resultó muy convincente. En la Figura 12 muestro espectros de amplitudes, normalizados, correspondientes a las secciones de la Figura 11, después de aplicar los operadores. Hay un incremento en las amplitudes de las altas frecuencias, pero dicho incremento no se apega al sentido físico de los datos sísmicos. Presento el código en el Apéndice 7, en las líneas 147 a 195.

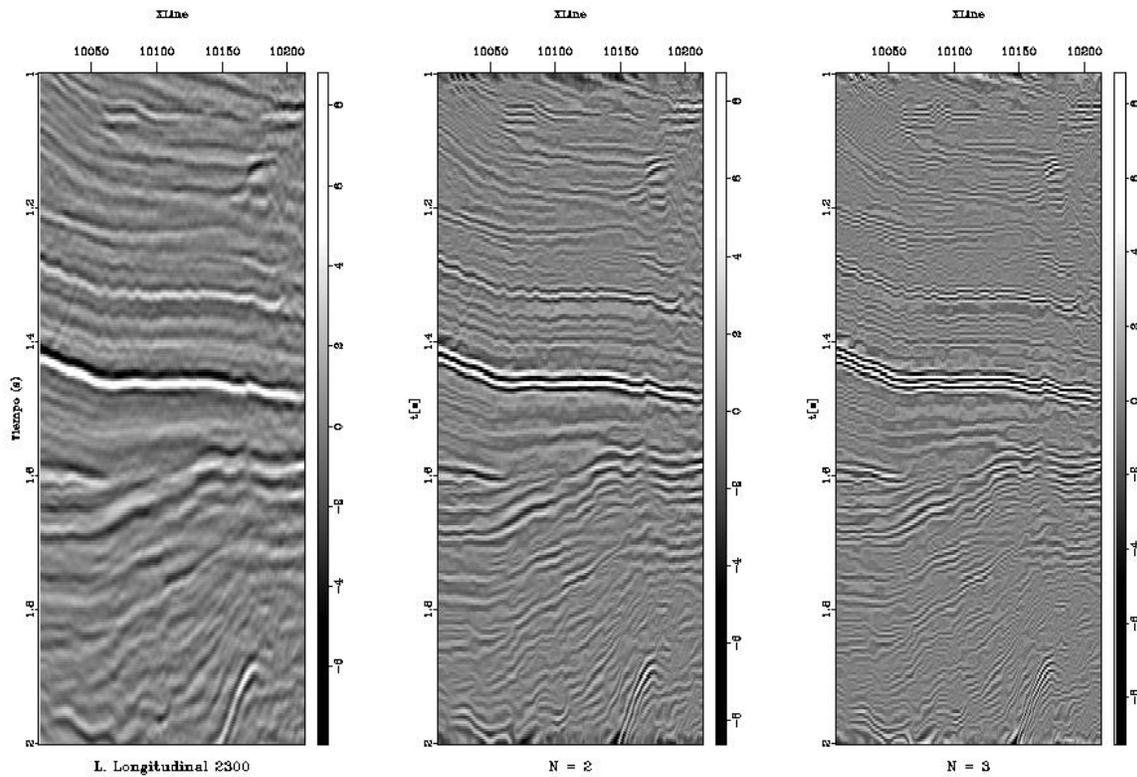


Figura. 11 Derecha: Línea longitudinal 2300 del subvolumen Agua Fría-Coapechaca-Tajín; centro: incremento de resolución vertical con $N = 2$; derecha: incremento de resolución vertical con $N = 3$. El incremento en el número de crestas y valles es un efecto numérico y no corresponde a la geología, sin embargo puede enfatizar rasgos como pliegues y fallas. La distancia entre trazas es de 25 m.

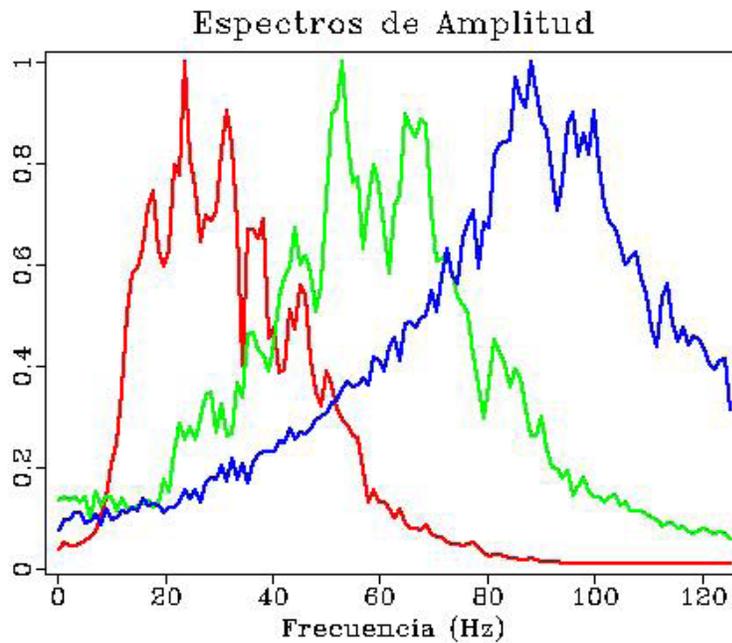


Figura. 12. Espectros de amplitud antes (rojo) y después de aplicar el operador de incremento con $N=2$ (verde) y $N=3$ (azul). Los corrimientos de las frecuencias dominantes hacia las altas frecuencias son artefactos numéricos y no corresponden necesariamente a la geología. El eje vertical es la amplitud normalizada.

3.2.3 Conversión tiempo a profundidad y viceversa

La conversión tiempo a profundidad y viceversa de secciones sísmicas es de interés en la interpretación sísmica. El vínculo entre el tiempo y la profundidad es un modelo de velocidad; estimar un modelo que represente de manera adecuada la relación entre tiempo y profundidad ha sido una actividad central en representación sísmica.

Dependiendo de los datos con los que se cuente se pueden estimar distintos modelos de velocidad. Generalmente, a partir de la sísmica se obtiene un modelo inicial de velocidades que debe ser ajustado con datos de pozo; de este modo se alcanza el detalle suficiente para obtener la conversión con buena precisión.

A partir de información sísmica de reflexión convencional se estiman velocidades de apilamiento y de migración, que son tipo *rms* (raíz de los cuadrados promediados). Las velocidades rms no representan la velocidad verdadera del subsuelo, por lo que se deben procesar para obtener modelos adecuados.

A diferencia de las velocidades rms, la velocidad promedio se obtiene de manera directa con información de pozo, como tiros de prueba y perfiles sísmicos verticales. Bartel, D. et al. (2006) indican que el modelo de velocidad promedio permite la conversión directa de tiempo a profundidad.

En ausencia de datos de velocidades de pozo, una primera aproximación para la conversión tiempo a profundidad es el uso del modelo de velocidad de intervalo (Amery, G. B., 1993). La velocidad de intervalo es una propiedad intrínseca de las rocas y se calcula a partir de las velocidades rms con la ecuación de Dix:

$$V_n = \frac{V^{2rms_n} T_n - V^{2rms_{n-1}} T_{n-1}}{T_n - T_{n-1}} \quad (7)$$

Donde V_n es la velocidad del intervalo, cuya cima y base corresponden a los tiempos T_{n-1} y T_n respectivamente, y V_{rms} es la velocidad rms asociada a los reflectores que delimitan dicho intervalo. Las velocidades rms se definen como:

$$V_{rms} = \frac{\sum V_i^2 t_i}{\sum t_i} \quad (8)$$

En la Figura 13 se muestra como se construye una gráfica de profundidad en función de la velocidad de intervalo: la profundidad entre reflectores se calcula multiplicando la velocidad de intervalo por el intervalo de tiempo correspondiente, y la suma de los intervalos de profundidad da la profundidad (Z) a cada reflector.

Para hacer un modelo apropiado para la conversión tiempo a profundidad se deben conocer las velocidades de intervalo desde primera capa del modelo hasta la profundidad a la que se localiza el objetivo. Si estos modelos se utilizan para migrar o convertir de tiempo a profundidad, se suavizan o promedian.

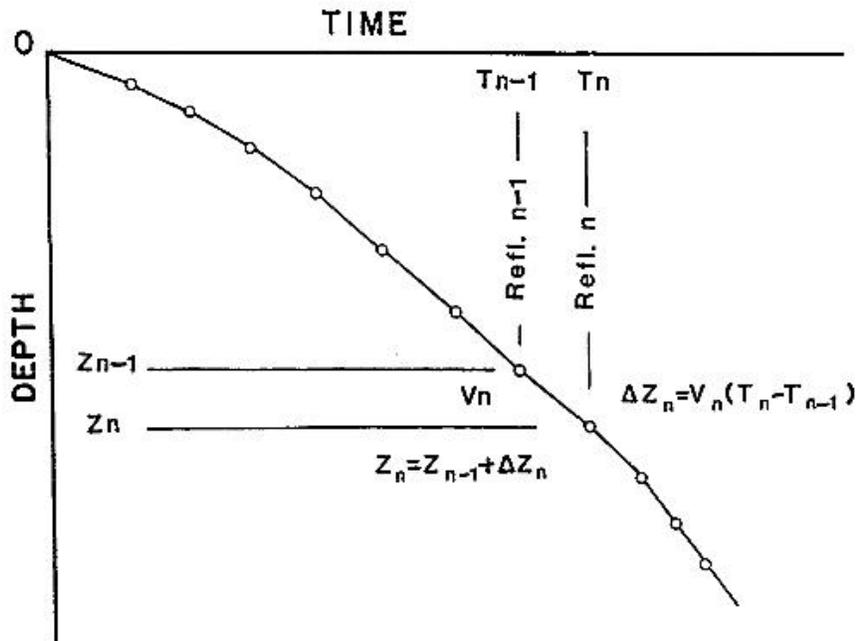


Figura. 13. Gráfica de profundidad-tiempo generada a partir de la velocidad de intervalo (tomada de Amery,1993).

Por su parte, la conversión profundidad a tiempo es importante puesto que diversos procedimientos en procesamiento e interpretación de datos sísmicos se llevan a cabo en el dominio del tiempo. Es común que dichos procedimientos se lleven a cabo después del proceso de representación sísmica en profundidad. Jones (2009) sugiere que el campo de velocidades empleado para esta conversión sea una versión suavizada del campo de velocidades de migración en profundidad.

Madagascar cuenta con programas para la conversión de modelos de velocidad tipo rms a velocidad de intervalo; asimismo, existen programas para la conversión tiempo-profundidad y viceversa.

A partir de una sección vertical del volumen de datos sísmicos marinos Nautla, localizado en Lankahuasa, y un modelo de velocidad de intervalo (Figura 14), realicé la conversión profundidad a tiempo (Figura 15). Para fines de ilustración, se hizo la conversión tiempo a profundidad nuevamente (Figura 16).

El código para generar los resultados que presento está en el Apéndice 8. La lectura, cambio de formato y extracción de sección vertical del volumen sísmico corresponde a las líneas 4 a 19. La construcción del modelo de velocidad, y el ajuste dimensional corresponde a las líneas 21 a 37. Aunque no es indispensable para la conversión profundidad a tiempo, a manera de ilustración, presento el código para la conversión profundidad a tiempo del modelo de velocidad en las líneas 41 a 46; el programa que utilicé fue *sfdepth2time*. La conversión profundidad a tiempo de la sección sísmica corresponde a las líneas 49 a 57; el programa que utilicé fue *sfdepth2time*. Por último, la conversión tiempo a profundidad de la sección sísmica corresponde a las líneas 60 a 68; el programa que utilicé fue *sftime2depth*.

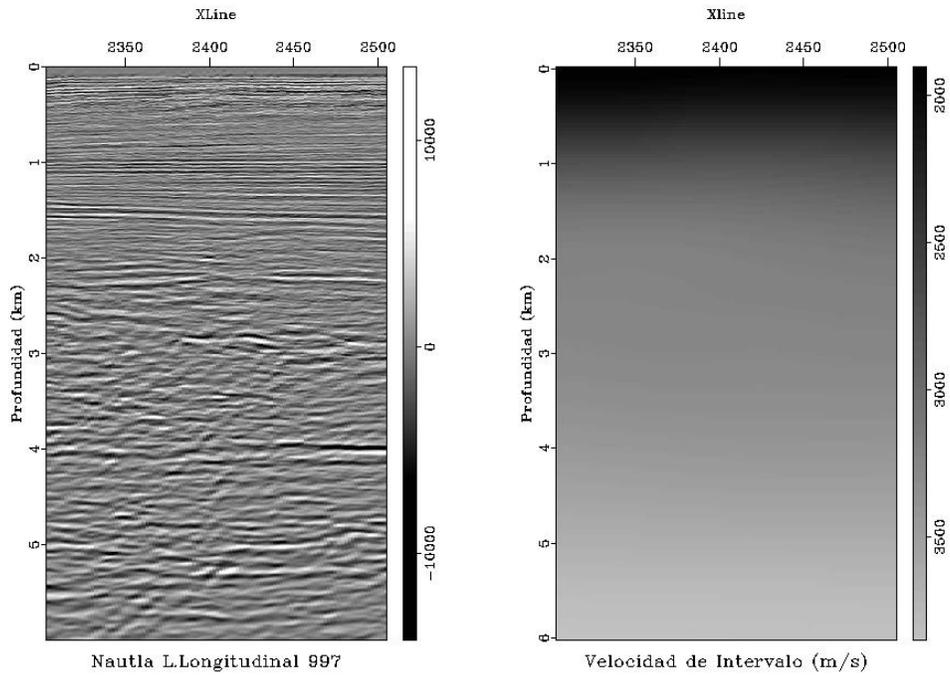


Figura. 14. Sección original, en profundidad, y modelo de velocidad de intervalo correspondiente. La distancia entre trazas es 12.5 m.

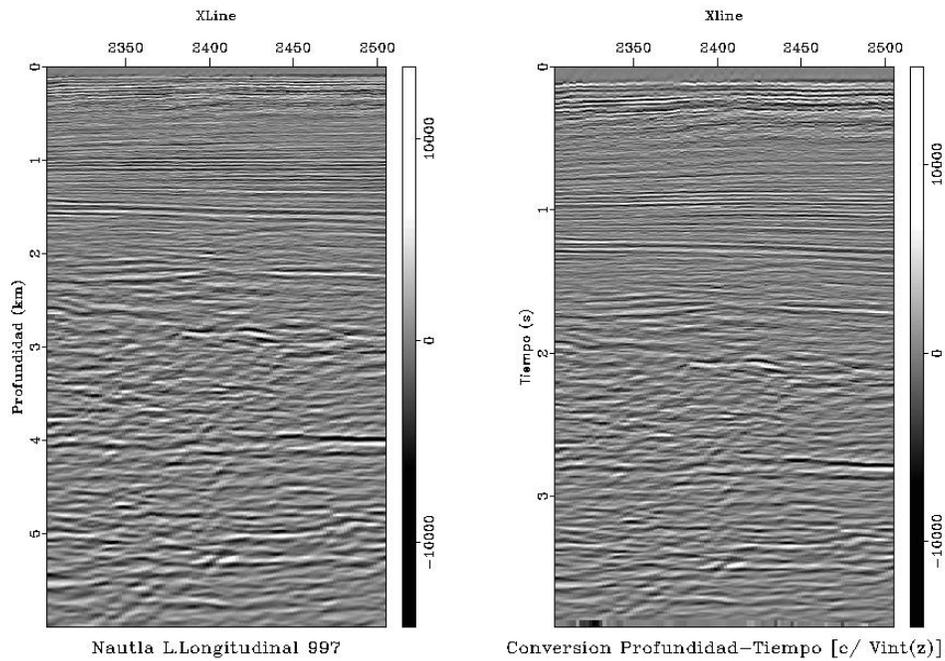


Figura. 15. Derecha: sección sísmica original, en profundidad. Izquierda: conversión profundidad a tiempo. La distancia entre trazas es 12.5m.

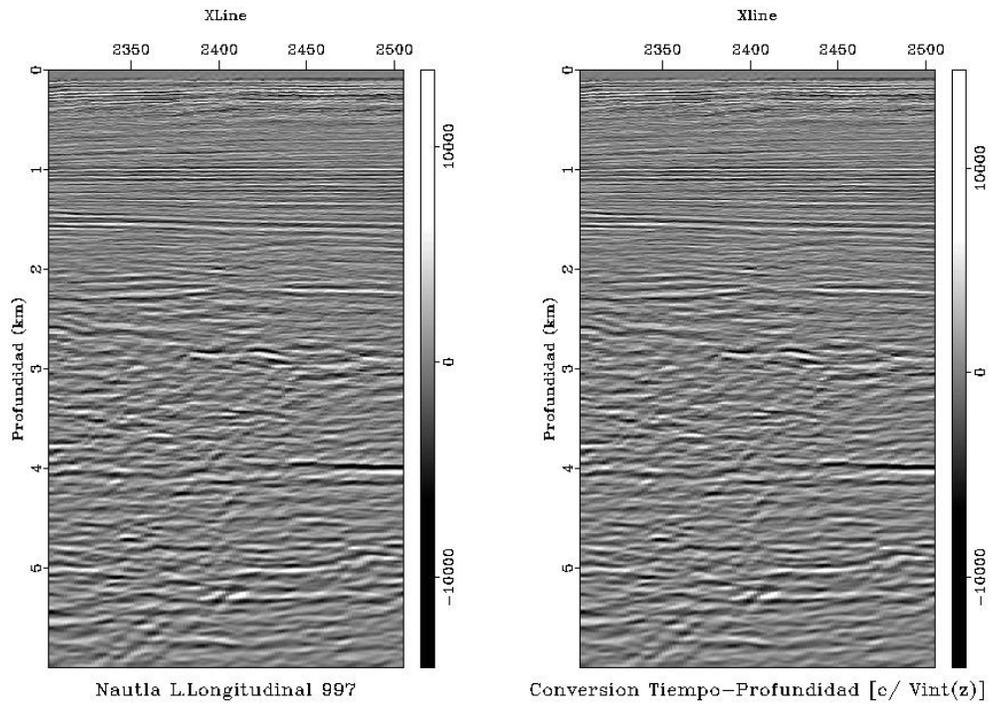


Figura. 16 Derecha: Línea sísmica original. Izquierda: conversión tiempo a profundidad de la sección en tiempo de la Figura 15. La distancia entre trazas es 12.5 m.

CAPÍTULO 4

DISCUSIÓN

Los resultados presentados en el Capítulo 3 ejemplifican la posibilidad de asimilar y adaptar tecnología de interés en la industria petrolera, en un contexto académico, a bajo costo. En primera instancia, es relevante que en la academia se pueda recurrir a plataformas de cómputo reproducible y prescindir el uso de herramientas comerciales para llevar a cabo tareas de representación sísmica similares a las que se realizan en la industria. El uso de documentos reproducibles adaptar las herramientas de cómputo que empleé fue de gran utilidad y permitió desarrollar el trabajo de manera eficiente. Además, es destacable que dichas tareas puedan realizarse incluso en computadoras personales que cuenten con un sistema operativo tipo *Unix* (p. ej. *Linux*, que cuenta con distribuciones públicas y gratuitas). La consecuencia de los puntos anteriores es que un estudiante, profesor, investigador o profesional involucrado en representación sísmica, que sea usuario de una plataforma de cómputo reproducible puede estar en condiciones de aprovechar las ventajas de la reproducibilidad (ver sección 1.2).

Desarrollé flujos de trabajo para realizar las tareas de acondicionamiento, manipulación y graficación de datos sísmicos en formato SEG-Y, y para obtener atributos sísmicos instantáneos. Las anteriores son tareas esenciales en la sismología de exploración.

Adapté una propuesta para la enfatización de la imagen sísmica basada en la obtención del negativo de la segunda derivada de los datos sísmicos (Figura 8). Los resultados muestran mejor definición y continuidad de los eventos. Esta es una manera sencilla, con resultados limitados, de resolver el problema de incremento del contenido de altas frecuencias.

Adapté una propuesta reciente para el incremento de resolución vertical basado en un multiplicador N de fase instantánea. Los resultados para incrementos únicos (p. ej. $N = 2$, como se muestra en la Figura 11) mejoran la continuidad de los eventos y destacan rasgos geológicos como fallas y pliegues; debe tomarse en cuenta este método genera artefactos numéricos. Los resultados para $N = 3$ (Figura 11) no fueron convincentes. Una explicación es que el incremento en la frecuencia de provoca el submuestreo de las señales con frecuencia mayor a la frecuencia de Nyquist dividida entre el valor N . Con este ejemplo se demuestra que aunque es sencillo obtener un incremento de frecuencia, se debe tener cautela con los operadores utilizados para ello.

Desarrollé un flujo de trabajo para realizar la conversión tiempo a profundidad y viceversa. Este tema es de interés puesto que lo datos sísmicos migrados tiempo se convierten a profundidad, con fines interpretativos, y a su vez, los datos migrados en profundidad se convierten al dominio del tiempo para efectuar cierto tipo de tareas, como el filtrado en tiempo.

Los flujos de trabajo o secuencias de instrucciones con formato reproducible (p. ej. Apéndices 7 y 8) pueden adaptarse, con relativa sencillez, a datos nuevos. Dichos flujos facilitan la detección de errores durante las implementaciones con nuevos datos. Además, debido a la estructura que se requiere para su reproducción, facilitan la comprensión de la relación entre los programas a ejecutar y los archivos de entrada y salida.

La reproducibilidad en representación sísmica es útil en las actividades de docencia, investigación y la práctica profesional. Sus beneficios académicos, técnicos y económicos son importantes, por ello debemos difundir el uso y generación de herramientas reproducibles. Queda mucho trabajo por delante, ya que existe una gran diversidad de programas de código abierto para representación sísmica que pueden ayudar a resolver problemas de interés actual.

CONCLUSIONES

1) La reproducibilidad es de gran utilidad en la docencia, investigación, capacitación de profesionales, e incluso para la práctica profesional. Representa beneficios académicos, técnicos y económicos como:

- i) disminución en la dependencia de licencias de programas comerciales para llevar a cabo investigación,
- ii) posibilita la adaptación, el desarrollo y transferencia de tecnología, y
- iii) auxilia la administración y transmisión eficiente del conocimiento.

2) Hay cuatro recientes condicionamientos en las publicaciones científicas que denotan el interés por promover la reproducibilidad:

i) La revista *Geophysics* incluyó la sección “*Geophysical software and algorithms*” (en español, Programas y algoritmos geofísicos; disponible en Internet en el sitio: <http://software.seg.org/>), en la cual se publican artículos cuyos algoritmos y códigos fuente se ponen a disposición del público a través del Internet.

ii) Revistas especializadas, como *Nature* o *The Insight Journal*, incluso comienzan a solicitar el código empleado o los datos como prerequisite para la publicación.

ii) Algunas instituciones académicas y entidades gubernamentales en Estados Unidos solicitan que los trabajos que financian sean hechos públicos bajo el esquema del código abierto.

iv) La *Licencia de Investigación Abierta* (ORL, por sus siglas en inglés) y el *Estándar de Investigación Reproducible* describen marcos legales para cada uno de los elementos del compendio que aseguran las atribuciones y promueven la investigación reproducible.

3) Del estudio realizado puede advertirse que la publicación con formato reproducible cobrará mayor importancia en el futuro; por ese motivo, considero que la investigación en representación sísmica, en el ámbito nacional, debe atender dicha tendencia para hacer eficiente la administración y divulgación del conocimiento.

4) La existencia de programas de código abierto representa un gran apoyo tanto en la investigación como en la docencia. Al investigador le permiten llevar a cabo su trabajo sin la necesidad comprar licencias de paquetes comerciales. A los profesores y estudiantes, les permite trabajar con herramientas a las que no podrían tener acceso de otra manera. En el ámbito de las geociencias el desarrollo de este tipo herramientas ha sido exitoso y un número importante de dichos sistemas ya se distribuyen con licencias que permiten su uso de manera gratuita tanto para fines académicos, como comerciales. Dada la diversidad de programas de código abierto para representación sísmica, considero que existen herramientas que permiten ayudar a resolver algunos de los problemas de representación sísmica, de interés actual.

5) Como prueba de lo anterior, demostré lo siguiente: i) actividades realizadas en la industria de la exploración sísmica las efectué en una computadora personal convencional a la que le instalé una plataforma de cómputo reproducible, ii) asimismo, con dicho equipo convencional desarrollé flujos de trabajo para enfatizar la imagen sísmica, iii) incrementé la resolución vertical de datos sísmicos y iv) llevé a cabo conversión tiempo a profundidad de datos sísmicos, y viceversa.

6) Los documentos reproducibles fueron de gran apoyo para la asimilación de las herramientas que empleé y permitieron desarrollar las actividades que presenté. Con ello compruebo que es posible asimilar, adaptar, desarrollar y transferir tecnología de interés en la industria petrolera a bajo costo, empleando plataformas de código abierto y siguiendo un esquema de cómputo reproducible.

REFERENCIAS

1. Amery, G. B., 1993, Basics of seismic velocities: The Leading Edge, 10, 1087 - 1091.
2. Barnes, A., 2007, A tutorial on complex seismic trace analysis: Geophysics, 72, w33 - w43.
3. Bartel, D., Busby, M., Nealon, J., Zaske, J., 2006, Time to depth conversion and uncertainty assessment using average velocity modeling: 76th Annual International Meeting, SEG, Expanded Abstracts, 2166 - 2170.
4. Claerbout, J. F., 2008, Reproducible computational research: a history of hurdles, mostly overcome, consultado en febrero de 2010; disponible en <http://sepwww.stanford.edu/data/media/public/sep//jon/reproducible.html>.
5. Claerbout, J.F. y Karrenbach, M., 1992, Electronic documents give reproducible research a new meaning: 62th Annual International Meeting., SEG, Expanded Abstracts, 601-604.
6. Donoho, D.L., Maleki, A., Morteza, S., Ur Rahman, I. Stodden, V., 2009, Reproducible research in computational harmonic analysis: Computing in Science & Engineering, 11, 8-18.
7. Fomel, S. y Claerbout, J., 2009, Reproducible research: Computing in Science & Engineering, 11, 5-7.
8. Fomel, S. y Hennenfent, G., 2006, Reproducible computational experiments using Scons, consultado en febrero de 2010; disponible en http://www.reproducibility.org/wiki/Reproducible_computational_experiments_using_SCons.
9. Forel, D., Benz, T. y Pennington, W., 2005, Seismic data processing with Seismic Un*x: SEG.
10. Jones, I. F., 2009, Tutorial: time conversion of depth migrated data: First break, 27, 51 - 55.
11. LeVeque, R.J., 2009, Python tools for reproducible research on hyperbolic problems: Computing in Science & Engineering, 11, 19-27.
12. Rosenberg, D. K., 2000, Open source: The unauthorized white papers: M&T Books.
13. Schwab, M., Karrenbach, N., Claerbout, J., 2000, Making scientific computation reproducible; Computing in Science & Engineering, 11, 61-67.
14. Stark, T. J., 2009, Frequency enhancement via an integer multiplier or just another GeoWizardry attribute?: 79th Annual International. Meeting., SEG, Expanded Abstracts, 1092 - 1096.

15. Stockwell, J., 1997, Free software in education: a case of CWP/SU: Seismic Unix: The leading edge, 16, 1045 -1049.
16. Stodden, V., 2009, The legal framework for reproducible scientific research: Computing in Science & Engineering, 11, 35-40.
17. Stodden, V. 2008, Enabling reproducible research: open licensing for scientific innovation, consultado en febrero de 2010; disponible en <http://www.stanford.edu/~vcs/papers/Licensing08292008.pdf>.
18. Taner, M.T., Koehler, F., y Sherif, R.E., 1979, Complex seismic trace analysis: Geophysics, 44, 1041 - 1063.
19. Vandewalle, P., 2009, How to make a paper reproducible, consultado en febrero de 2010; disponible en http://reproducibleresearch.net/index.php/How_to

APÉNDICES

1. Programas de código abierto útiles en geociencias
2. Programas de código abierto que satisfacen las necesidades de investigación en el procesamiento de datos sísmicos y representación sísmica
3. Herramientas complementarias de código abierto para sismología de exploración
4. Programas de código abierto para sismología de exploración con restricciones para uso comercial
5. Bibliografía sobre sismología de exploración reproducible
6. Documentos reproducibles y datos disponibles en línea
7. Secuencia de instrucciones SConstruct para lectura, manipulación, graficación de un volumen sísmico, cálculo de atributos sísmicos e incremento de resolución vertical
8. Secuencia de instrucciones SConstruct para conversión tiempo a profundidad de datos sísmicos y viceversa

Apéndice 1. Programas de código abierto útiles en geociencias

Nombre (Año de aparición)	Descripción y características especiales	Escrito / mantenido por	Licencia ¹	SO ²	Sitio en Internet
SAGA-GIS (2007)	Sistema de información geográfico (SIG; GIS, por las siglas en inglés). Tiene "Interfaz para programación de aplicaciones" (API, por sus siglas en inglés) para el procesamiento de datos geográficos, lo que facilita la implementación de nuevos algoritmos. Maneja datos en mallas, vectores y tablas.	V. Wichmann, O. Conrad y A. Ringeler. (U. de Goettingen, Alemania)	GPL, LGPL	W / L	http://saga-gis.wiki.sourceforge.net/
SGeMS (2005)	Modelado geoestadístico 3D. Incluye algoritmos clásicos y desarrollados en el Laboratorio del Centro para la Predicción de Yacimientos (SCRF, por sus siglas en inglés) de la U. de Stanford (California, EUA).	N. Remy, A. Boucher, J. Wu, T. Li, J. Rajiv, y M. Sharps (U. de Stanford, California, EUA)	GPL	U / L / M / W	http://sgems.sourceforge.net/
Gplates (2003)	Programa interactivo de visualización y manipulación de reconstrucciones de características geológicas y paleogeográficas de placas tectónicas a través del tiempo geológico.	Proyecto EarthByte (U. de Sidney, Australia), Div. of Geological and Planetary Sciences (Cal Tech, California, EUA) y Centro para Geodinámica (Servicio Geológico Noruego)	GPL	L / M / WXP / V	http://www.gplates.org/
Geoblock (2001)	Modelado 3D, administración de bases de datos y visualización de datos de geociencias. Útil en geología, topografía, modelación minera y cálculo de reservas minerales.	Pavel Vassilliev (U. Belgorod, Rusia)	MPL (Uso gratuito para fines no comerciales)	W	http://geoblock.sourceforge.net/
Gnutek (2001)	Análisis de datos geológicos estructurales. Los datos se pueden presentar en proyección estereográfica (red de Schmidt). El análisis de datos incluye: densidad de polos, isolíneas y estadísticas.	Stefan Krumm (UT, Freiberg, Alemania)	GPL	U / L / M / W	http://sourceforge.net/projects/gnutek/
GMT (1988)	Manipulación de datos geográficos y cartesianos. Producción de imágenes (mapas de contornos, superficies artificialmente iluminadas, perspectivas 3D) con formato	Paul Wessel (U. de Hawaii, Manoa, EUA) y Walter H.F. Smith (Admón. Oceánica y Atmosférica Nacional [NOAA], EUA)	GPL	U / L / M / C	http://gmt.soest.hawaii.edu/

¹ GPL: General Public License; LGPL: Library General Public License ; MPL: Mozilla Public License. Los términos de las licencias pueden consultarse en: <http://www.opensource.org/licenses/alphabetical>

² SO= Sistema operativo; U= UNIX; L= Linux; M= Mac OS-X; C= Cygwin; W= Windows; WXP= Windows XP; V=Windows Vista

Apéndice 2. Programas de código abierto que satisfacen las necesidades de investigación en el procesamiento de datos sísmicos y representación sísmica

Nombre (Año de aparición)	Descripción y características especiales	Escrito/mantenido por	Licencia ³	SO ⁴	Sitio en Internet
<i>Madagascar</i> (2006)	Plataforma para investigación reproducible. Cuenta con una amplia biblioteca de documentos reproducibles actualizados (SEP, CSM, U. de Texas, entre otros) y libros de Jon Claerbout. Incorpora técnicas computacionales modernas. Formato de datos: volúmenes de datos binarios con encabezados ASCII. Compatible con formato SEG-Y, <i>SEPlib</i> y <i>SU</i> .	Sergey Fomel (U. de Texas, EUA) y Paul Sava (Escuela de Minas de Colorado, EUA)	GPL	U / L / M / C	http://www.reproducibility.org
DDS (2003)	Versión limitada del sistema utilizado por BP. Formato de datos: volúmenes de datos binarios con encabezados ASCII. Compatible con formato SEG-Y, <i>SEPlib</i> y <i>SU</i> .	Randall Selzler (BP)	GPL	U / L	http://www.freeusp.org/DDS
CPSeis (2002)	Sistema de procesamiento de ConocoPhillips. Tiene interfaz guiada por el usuario.	ConocoPhillips	GPL	U / L	http://cpseis.org
<i>Seismic Unix</i> (1992)	Utilizado ampliamente en la academia e industria. Formato de datos: trazas con datos binarios y encabezados ASCII. Compatible con formato SEG-Y.	John Stockwell (Escuela de Minas de Colorado [CSM, por sus siglas en inglés], EUA)	BSD	U / L / C	http://timna.mines.edu/cwpcodes
<i>SEPlib</i> (1991)	Primera plataforma para elaboración y uso de documentos reproducibles. Permite leer documentos reproducibles del SEP (1989 - 2009; docs. previos a 2002 son de acceso público) y libros de Jon Claerbout. Formato de datos: volúmenes de datos binarios con encabezados ASCII. Compatible con formato SEG-Y.	SEP (U. de Stanford, CA, EUA) / Robert Clapp (U. de Stanford)	“ <i>SEPLIB</i> se puede usar o modificar para uso propio; no se puede redistribuir, ni sus modificaciones, con fines comerciales, sin previo consentimiento escrito de la Universidad de Stanford”.	U / L / M / C	http://sepwww.stanford.edu/doku.php?id=sep:software:SEPlib

¹ GPL= General Public License; BSD=Berkeley Software Distribution. Los términos de las licencias pueden consultarse en: <http://www.opensource.org/licenses/alphabetical>

² SO= Sistema operativo; U= UNIX; L= Linux; M= Mac OS-X; C= Cygwin; W= Windows.

Apéndice 3. Herramientas complementarias de código abierto para sismología de exploración.

Nombre (año de aparición)	Descripción y características especiales	Escrito/mantenido por	Licencia ⁵	SO ⁶	Sitio en Internet
Delivery (2004)	Plataforma, basada en Java, para inversión sísmica Bayesiana. Útil para caracterización de yacimientos.	CSIRO Petroleum/ BHP Billiton	GPL y BSD para algunos componentes	U / L/ M / C	http://www-old.dpr.csiro.au/StochasticSeismicInversion/top.html
JRG (1999)	Sistema, basado en Java, para procesamiento básico de datos sísmicos de reflexión, capaz de graficar datos 3D. No cuenta con supresión de ruidos, deconvolución o migración.	John Louie (U. de Nevada, Reno, NV, EUA)	“Los programas y métodos son objeto de investigación académica, no comercial. Desearía conocer el uso que se le da y recibir comentarios con respecto a su éxito o fracaso”	U / L/ M/ W	http://www.seismo.unr.edu/jrg/
SeismicLab (ND)	Conjunto de herramientas, para <i>Matlab</i> , para pre-procesamiento, procesamiento y representación sísmica. Es posible trabajar con conjuntos de datos pequeños. Permite leer archivos SEG-Y. Es compatible con <i>SU</i> .	Mauricio Sacchi (U. de Alberta, Canadá)	GPL	U/ L/ W	http://seismic-lab.phys.ualberta.ca/
Kogeo (ND)	Conjunto de herramientas para procesamiento de datos, manejo de bases de datos, interpretación, manipulación de encabezados 3D, y herramientas de visualización 2D y 3D. Compatible con formatos SEG-Y, <i>SU</i> e incluso mapas de bits (imágenes).	Phillipp Konerding (U. de Hamburgo, Alemania)	GPL	W	http://www.kogeo.de
QiWorkbench (ND)	Plataforma extensible, basada en Java, para poner en funcionamiento flujos de trabajo integrados para procesar, analizar y visualizar datos sísmicos	BHP Billiton, G&W Systems, INT, CSM, CSIRO Petroleum (Australia) y U. de California (Sta. Bárbara, CA, EUA)	GPL; BSD para las herramientas comerciales adicionales sin código abierto.	U / L/ M/ W	http://qiworkbench.org

¹ GPL= General Public License; BSD=Berkeley Software Distribution. Los términos de las licencias pueden consultarse en: <http://www.opensource.org/licenses/alphabetical>

² SO= Sistema operativo; U= UNIX; L= Linux; M= Mac OS-X; C= Cygwin; W= Windows.

Apéndice 4. Programas de código abierto para sismología de exploración con restricciones para uso comercial

Nombre (año de aparición)	Descripción y características especiales	Escrito/ mantenido por	Licencia	SO ⁷	Sitio en Internet
Opendtect (2003)	Interpretación de datos sísmicos. Obtención de atributos sísmicos, interpretación auxiliada con redes neuronales. Sistema extensible.	dGB (Holanda)	Gratuito para uso no comercial	U / L/ W	http://ww2.opendtect.org
Jive3D (1997)	Modelado directo e inversión tomográfica. Se puede modelar una gran variedad de tipos de datos sísmicos.	James Hobro (Cambridge, UK)	Gratuito para uso no comercial	U / L	http://bullard.esc.cam.ac.uk/~hobro/Jive3D/
IGeoS-Integrated GeoScience data analysis (1995)	Ejecución de tareas de procesamiento sísmico (reflexión y refracción; simulación de forma de onda; sísmica multicomponente; modelos de velocidad y densidad). Análisis e integración de datos geofísicos y de geociencias.	Igor Morozov (U. de Saskatchewan, Canadá)	Gratuito para uso no comercial.	U / L	http://seisweb.usask.ca/igeos/
CREWES Programas educativos (ND)	Conjunto de herramientas, para <i>Matlab</i> , para procesamiento y representación sísmica. Es un subconjunto del sistema para los patrocinadores.	Consortio CREWES (U. de Calgary, Canadá)	Gratuito para uso no comercial.	U / L/ M/ W	http://www.crewes.org/ResearchLinks/FreeSoftware/

¹ SO= Sistema operativo; U= UNIX; L= Linux; M= Mac OS-X; C= Cygwin; W= Windows.

Apéndice 5. Bibliografía sobre sismología de exploración reproducible

Autor (año)	Título	Plataforma	Sitio o editorial
Schuster, G. (2009)	Seismic interferometry	<i>Matlab</i>	Cambridge University Press
Claerbout, J. (2008)	Image estimation by example	<i>SEPLib / Madagascar</i>	http://sepwww.stanford.edu/sep/prof/index.html
Forel, D. Benz, T. y Pennington, W. (2005)	Seismic data processing with Seismic Un*x	<i>Seismic Unix</i>	Soc. Exp. Geophys, Tulsa, OK, EUA
Margrave, G.F. (2003)	Numerical methods of exploration seismology with algorithms in <i>Matlab</i>	<i>Matlab</i>	http://www.crewes.org/ResearchLinks/FreeSoftware/EduSoftware/NumMeth.pdf
Claerbout, J. (2000)	Basic earth imaging	<i>SEPLib / Madagascar</i>	http://sepwww.stanford.edu/sep/prof/index.html
Claerbout, J. (1992)	Earth sounding analysis: processing vs inversion	<i>SEPLib</i>	http://sepwww.stanford.edu/sep/prof/index.html
Claerbout, J. (1985)	Imaging the earth's interior	<i>SEPLib</i>	http://sepwww.stanford.edu/sep/prof/index.html

Apéndice 6. Documentos reproducibles y datos disponibles en línea

Tipo	Institución / autor	Descripción	Plataforma	Sitio
Informes	SEP, U. de Stanford, (CA, EUA), varios autores.	Informes semestrales (1989-2009; sólo son de acceso libre de 1989 a 2002)	SEPlib	http://sepwww.stanford.edu/doku.php?id=sep.research:reports
Tesis	SEP, U. de Stanford, (CA, EUA), varios autores.	Tesis doctorales (1979-2009). De 1992 en adelante incluyen el código de los programas.	SEPlib	http://sepwww.stanford.edu/doku.php?id=sep.research:theses
Artículos	CWP, CSM, (CO, EUA), varios autores.	Artículos recientes	Madagascar	http://www.reproducibility.org/wiki/Reproducible_Documents
Artículos	U. Texas, (TX, EUA), varios autores	Artículos recientes	Madagascar	http://www.reproducibility.org/wiki/Reproducible_Documents
Artículos	Seismic Lab. for Imaging and Modeling (SLIM), (C.B, Canadá), varios autores	Artículos recientes	Madagascar	http://www.reproducibility.org/wiki/Reproducible_Documents
Artículos	SEP, U. de Stanford, (CA, EUA), varios autores.	Artículos recientes	Madagascar	http://www.reproducibility.org/wiki/Reproducible_Documents
Datos	Varios	Modelos de velocidades	Madagascar	http://www.reproducibility.org/wiki/Reproducible_Documents
Datos	Varios	Registros de tiros del libro Seismic Data Analysis de Oz Yilmaz	SU	http://www.cwp.mines.edu/cwpcodes/data/oz.original/
Datos	SEG/ Varios	Datos 2D	Archivos con formato SEG-Y	http://software.seg.org/

Apéndice 7. Secuencia de instrucciones *Sconstruct* para lectura, manipulación, graficación de un volumen sísmico, cálculo de atributos sísmicos e incremento de resolución vertical

```

1  from rsfproj import *
2
3  *****CAMBIO DE FORMTATO SEGY A RSF*****
4
5  #Informacion del volumen sismico original Agua Fria-Coapechaca-Tajin
6  #Rango XLine          Rango InLine          Tiempo
7  #10010 - 10213  1965 - 2482
8  #      nx=204      ni=518      nt=1252
9  #      dx=1      di=1      dt=0.004s
10
11 #CAMBIO DE FORMATO SEGY a RSF
12 Flow('sismica_original','sergio_sub.segy', ''segypread tfile=tfile1.rsf
13          hfile=hfile1 bfile=bfile1 endian=y'')
14
15 *****MANIPULACION DEL VOLUMEN*****
16
17 #SELECCION DE SUBVOLUMEN ORIGINAL
18 Flow('sub','sismica_original',''window min1=1 max1=2 |
19          put n2=204 o2=10010 n3=518 o3=1965'')
20
21 *****GRAFICACION*****
22
23 #Instruccion para graficar volumenes
24 graf=''' byte | grey3 flat=n frame2=10010
25          label1='Tiempo (s)' label2=XLine label3=Inline ''

```

```

26
27 Plot('sub','sub',graf+'title='Simica original'')
28 Plot('subg','sub','agc'|'+graf+'title='Simica original + CAG'')
29 Result('F1','sub subg','SideBySideAniso')
30
31 *****SELECCION DE LINEA LONGITUDINAL*****
32
33 Flow('insub','sub','window min3=2300 n3=1 squeeze=n'')
34 Flow('insubg','insub','agc')
35 Plot('insub','insubg','grey label1='Tiempo (s)' label2=XLine scalebar=y
36         title='L. Longitudinal 2300' '')
37
38 *****SELECCION DE CORTE HORIZONTAL*****
39
40 Flow('csub','sub','window min1=1.448 n1=1 squeeze=n'')
41 Plot('csub','csub',graf+'frame3=2300 title='Corte t=1.448s'
42         point1=0 bias=-104'')
43
44 Result('F2','insub csub','SideBySideAniso')
45
46
47 *****PELICULA DE CORTES TEMPORALES*****
48
49 ni=1          #Inicio
50 nf=6          #Fin
51 ti=1.436      #tiempo inicial
52 dt=0.004      #incremento en tiempo

```

```

53
54
55 cortes = []
56 for inst in range(nf):
57     corte = 'corte%d' % inst
58     cortes.append(corte)
59     tinst = ti+inst*dt
60     cortetiempo='window n1=1 min1=%f squeeze=n | '%tinst
61     Plot(corte,'sub',cortetiempo + "" byte | grey3 flat=n
62             frame2=10010 label2=XLine label3=Inline
63             point1=0 title="t= %g s""%(tinst))
64 Plot('corte',cortes,'Movie',view=1)
65
66 Result('F3', 'corte1 corte2 corte3 corte4', 'TwoRows')
67
68 *****OBTENCION DE ATRIBUTOS SISMICOS INSTANTANEOS*****
69
70 #Cuadratura
71 Flow('hilb', 'insub', 'envelope hilb=y order=75')
72
73 #Envolvente
74 Flow('amp', None, ""math x=insub.rsf y=hilb.rsf
75             output='sqrt(x^2 +y^2)' "")
76
77 Plot('amp', 'amp', ""grey bias=2390.92 title="Envolvente"
78             label1='Tiempo[s]' label2="XLine" scalebar=y "")
79

```

```

80 #Coseno de la fase instantanea
81 Flow('cos',None,'math x=insub.rsf a=amp.rsf output="x/a" | put o2=10010 ')
82
83 Plot('cos','cos','"grey title="Coseno de la Fase Instantanea"
84         label1='Tiempo[s]' label2="XLine" scalebar=y "')
85
86 Result('F4','insub cos amp','SideBySideAniso')
87
88 #Fase instantanea
89 Flow('fase',None,'math cos=cos.rsf output="acos(cos)" | put o2=10010 ')
90
91 #Detalle traza sismica
92 Flow('traza','insub','window n2=1 min1=1 max1=1.2')
93 Flow('traza-amp','amp','window n2=1 min1=1 max1=1.2')
94 Flow('traza-hilb','hilb','window n2=1 min1=1 max1=1.2')
95 Flow('traza-amp-neg','amp','window n2=1 min1=1 max1=1.2 |scale dscale=-1')
96 Flow('traza-fase','fase','window n2=1 min1=1 max1=1.2')
97 Flow('traza-atr',None,'"cat traza.rsf traza-hilb.rsf
98         traza-amp.rsf traza-amp-neg.rsf axis=2"')
99
100 Result('F5','traza-atr','"graph title='Traza Original;
101         Cuadratura; Envolvente [Pos. y Neg.]'
102         plotfat=7 label1='Tiempo[s]' label2=' ' "')
103
104 *****ENFATIZACION DE IMAGEN CON NEGATIVO DE SEGUNDA DERIVADA****
105
106 Flow('neg2da','insub','deriv | deriv | scale dscale=-1')

```

```

107 Plot('neg2da','neg2da','"agc | grey title='Negativo 2da Derivada + AGC'
108             label1='Tiempo[s]' label2="Xline" scalebar=y"')
109
110 #Cuarta Derivada
111 Flow('4da','insub','deriv | deriv | deriv | deriv')
112 Plot('4da','4da','"agc | grey title='4ta Derivada + AGC'
113             label1='Tiempo[s]' label2="Xline" scalebar=y"')
114
115 Result('F6','insub neg2da 4da','SideBySideAniso')
116
117 #Espectros de Amplitud
118 Flow('espinsub','insub','spectra all=y | scale axis=1')
119 Flow('espneg2da','neg2da','spectra all=y | scale axis=1')
120 Flow('esp4da','4da','spectra all=y | scale axis=1')
121
122 Flow('espcompderiv','espinsub','cat espneg2da.rsf esp4da.rsf axis=2 ')
123
124 Result('F7','espcompderiv','"graph label1="Frecuencia (Hz)"
125             title='Espectros de Amplitud' plotfat=7 plotcol=5,3,6"')
126
127 #Promedio Aritmetico (N1+Neg2daDeriv)/2
128
129 Flow('insubprom1neg',None,'"math
130             x=insub.rsf y=neg2da.rsf output='(x+y)/2' "')
131 Plot('insubprom1neg','insubprom1neg','"agc | grey
132             title='Prom. Aritm. Orig. y Neg. 2da Deriv. + AGC'
133             label1='Tiempo[s]' label2="Xline" scalebar=y"')

```

```

134
135 Result('F8', 'insub neg2da insubprom1neg', 'SideBySideAniso')
136
137 #Espectros de Amplitud
138 Flow('espinsubprom1neg', 'insubprom1neg', 'spectra all=y | scale axis=1')
139 Flow('espcomprom1neg', 'espinsub', 'cat espneg2da.rsf
140     espinsubprom1neg.rsf axis=2 ')
141
142 Result('F9', 'espcomprom1neg', 'graph label1="Frecuencia (Hz)"
143     title="Espectros de Amplitud'
144     plotfat=7 plotcol=5,3,6")
145
146
147 *****INCREMENTO DE FRECUENCIA CON MULTIPLICADOR DE FASE N*****
148
149 #CASO SINTETICO 1D A PARTIR DE FUNCION COSENO
150
151 #Senal cosenoidal:  $\cos(2\pi f t)$ 
152 frec=10
153 Flow('cos1D', None, 'math n1=51 d1=0.004
154     output="cos(2*3.1415*d*x1)" "%frec)
155
156 #Atributos senal cosenoidal:
157 #Amplitud instantanea: para el coseno es 1
158 #Fase instantanea: para este caso es el acos
159 Flow('fasec', None, 'math cos=cos1D.rsf output="acos(cos)" ')
160

```

```

161 #Incremento de Frecuencia Caso 1D
162
163 #CASO 1D, N=2
164 Flow('cosN2',None,'" math phi=fasec.rsf output='cos(2*phi)' "')
165 Flow('cosN1-2',None,'cat cos1D.rsf cosN2.rsf axis=2')
166
167 Result('F10','cosN1-2,'" graph
168         title='Traza original y Traza con Inc. de Frec. N=2'
169         label1='t[s]' label2="" plotfat=10"")
170
171 #CASO CON DATOS SISMICOS REALES
172
173 #CASO N=2
174 Flow('insubN2',None,'" math amp=amp.rsf phi=fase.rsf
175         output='amp*cos(2*phi)' | put o2=10010"")
176 Plot('insubN2','insubN2,'"agc | grey
177         title='N = 2'
178         label2=""XLine" label1='t[s]' scalebar=y "')
179
180 #CASO N=3
181 Flow('insubN3',None,'" math amp=amp.rsf phi=fase.rsf
182         output='amp*cos(3*phi)' | put o2=10010"")
183 Plot('insubN3','insubN3,'"agc | grey
184         title='N = 3'
185         label2=""XLine" label1='t[s]' scalebar=y "')
186
187 Result('F11','insub insubN2 insubN3','SideBySideAniso')

```

```
189
190 #Espectros de Amplitud N=2 y N=3
191 Flow('espinsubN2','insubN2','spectra all=y | scale axis=1')
192 Flow('espinsubN3','insubN3','spectra all=y | scale axis=1')
193 Result('F12','espinsub','"cat espinsubN2.rsf espinsubN3.rsf axis=2 |
194         graph label1="Frecuencia (Hz)"
195         title='Espectros de Amplitud' plotfat=7 plotcol=5,3,6"')
196 End()
```

Apéndice 8. Secuencia de instrucciones *Sconstruct* para conversión tiempo a profundidad y viceversa

```
1  from rsfproj import *
2
4  #INFORMACION DE LA SECCION SISMICA
5  #Volumen:293935_NAUTLA_Q_1RA_ETAPA_MIGPRESDMCF CG_WG_997_1688.sgy
6
7  #      INLINE = 997
8  #RANGO  XLINE = 2304-2504
9  #      dz = 0.005
10
12  #CAMBIO DE FORMATO SEG Y a RSF
13
14  Flow('in','Nautla_Inline997_2304-2504.segy','seg yread tfile=tfile.rsf
15          hfile=hfile bfile=bfile endian=y | put
16          o2=2304 d2=2 unit1=km unit2='' '')
17
18  Plot('in','in','grey scalebar=y label1=Profundidad label2=XLine
19          title="Nautla L.Longitudinal 997"')
20
21  #MODELO DE VELOCIDAD V(z)
22  Flow('vi',None,'spike n1=151 n2=26 | put d1=0.040 o1=0 d2=8 o2=2304 |
23          dd form=ascii line=1 out=vint')
24
25  Flow('vint','vi','dd form=native | put d1=0.040 unit1=km unit2='' ')
```

```

27 #Interpolacion: dx=8 -> dx=2
28 Flow('vint-intrp','vint','transp | sinc n1=101 o1=2304 d1=2 | transp')
29
30 #Suavizamiento
31 Flow('vintz','vint-intrp','smooth rect1=20 rect2=5')
32
33 Plot('vintz','vintz','"grey bias=3200 color=j scalebar=y barreverse=y
34     label1=Profundidad label2=Xline
35     title="Velocidad de Intervalo (m/s)"')
36
37 Result('F14','in vintz','SideBySideAniso')
38
39 #Conversion de profundidad a Tiempo
40
41 #Conversion V(z)-> V(t)
42 Flow('vintd2t','vintz','"put d1=40 | depth2time velocity=$SOURCE
43     nt=980 dt=0.004 | put label1=Tiempo unit1=s "')
44
45 Plot('vintd2t','vintd2t','"grey color=j bias=3200 scalebar=y barreverse=y
46     label1=Tiempo label2="F2 Xline" title="Velocidad de Intervalo" "')
47
48
49 #Conversion Profundidad-Tiempo, con V(z)
50 Flow('inm','in','put d1=5 unit1=m')
51 Flow('vintzm','vintz','put d1=40 unit1=m | sinc n1=1200 o1=0 d1=5')
52 Flow('ind2t','inm','"depth2time velocity=vintzm.rsf
53     nt=980 dt=0.004 t0=0 | put unit1=s "')
54 Plot('ind2t','"grey title="Conversion Profundidad-Tiempo [c/ Vint(z)]"')

```

```

55         label1="\F2 Tiempo" scalebar=y label2=Xline")
56
57 Result('F15', 'in ind2t', 'SideBySideAniso')
58
60 #Conversion Tiempo-Profundidad con V(z)
61 Flow('int2d', 'ind2t', "time2depth velocity=vintzm.rsf intime=n
62         nz=1200 dz=5 | put label1="\F2 Profundidad" unit1=km d1=0.005")
63 Plot('int2d', "grey title="Conversion Tiempo-Profundidad [c/ Vint(z)]"
64         label1=Profundidad label2=Xline scalebar=y")
65
66 Result('F16', 'in int2d', 'SideBySideAniso')
67
68 #Result('tprof', 'in vints int2db ind2t', 'TwoColumns')
69 End

```

AGRADECIMIENTOS

A Sergio Chávez Pérez, cuyos comentarios, sugerencias y constante apoyo han sido relevantes en mi formación profesional. Reconozco su interés y compromiso en la formación y capacitación de recursos humanos.

A Liliana Vargas Meleza, del Instituto Mexicano del Petróleo, cuyo trabajo de revisión fue de gran utilidad.

A mis sinodales Cinna Lomnitz Aronsfrau, Francisco J. Sánchez Sesma, David Escobedo Zenil y Martín Cárdenas Soto, cuyas observaciones fueron importantes para complementar el trabajo. Agradezco a Martín por orientarme durante la etapa inicial del proyecto.

A Marco Vázquez García, Gerente de Geofísica de PEMEX Exploración y Producción, por proporcionar datos sísmicos y permiso para publicarlos.

Al Instituto Mexicano del Petróleo por el financiamiento para la realización de este trabajo.

A mis padres, María Eugenia y Carlos, por su valioso apoyo. Le agradezco de manera especial a mi padre por orientarme y motivarme, de principio a fin, durante la elaboración de la tesis, y por auxiliarme con las correcciones y sugerencias de redacción y estilo.

A la Universidad Nacional Autónoma de México, núcleo generador y divulgador de conocimiento, que brinda educación y formación de calidad, pública y gratuita, necesaria para el desarrollo de la sociedad mexicana.