

Reproducibilidad en representación sísmica

Reproducibility in Seismic Imaging

González-Verdejo O.

*Dirección de Exploración y Producción
Instituto Mexicano del Petróleo
Correo: oziel.gonzalezv@gmail.com*

Chávez-Pérez S.

*Dirección de Exploración y Producción
Instituto Mexicano del Petróleo
Correo: sergio.chavezp@gmail.com*

Información del artículo: recibido: marzo de 2010; aceptado: mayo de 2011

Resumen

En el campo de la sismología de exploración, actividad fundamental en la industria petrolera, existe interés en el ámbito nacional por incorporar la reproducibilidad computacional en la aplicación, investigación y docencia del procesamiento y representación de datos sísmicos (modelado, migración, tomografía e inversión). Esta reproducibilidad computacional implica la estructuración y descripción de los elementos de un experimento numérico o secuencia de procesamiento. Así, de manera independiente, un profesional de la práctica, investigador, instructor o estudiante puede estudiarlo, verificarlo, reproducirlo y modificarlo. En este trabajo documentamos y adaptamos la reproducibilidad en representación sísmica para contribuir a generar conciencia académica y profesional acerca de sus beneficios, así como para favorecer la asimilación de los programas de código abierto en esta área. Presentamos ejemplos de enfatización de imágenes sísmicas, de interés académico e industrial usando datos sísmicos mexicanos. Comprobamos que es posible asimilar, adaptar 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, there is interest at national level of integrating reproducibility in applied, educational and research activities related to seismic processing and imaging. This reproducibility implies the description and organization of the elements involved in numerical experiments. Thus, a researcher, teacher or student can study, verify, repeat, and modify them independently. In this work, we document and adapt reproducibility in seismic processing and imaging to spread this concept and its benefits, and to encourage the use of open source

Descriptores

- reproducibilidad
- sismología de exploración
- representación sísmica
- representación enfatizada
- código abierto

Keywords

- reproducibility
- exploration seismology
- seismic imaging
- enhanced imaging
- open source

software in this area within our academic and professional environment. We present an enhanced seismic imaging example, of interest in both academic and professional environments, using Mexican seismic data. As a result of this research, we prove that it is possible to assimilate, adapt and transfer technology at low cost, using open source software and following a reproducible research scheme.

Introducción

La reproducibilidad es un principio del proceso de investigación científica e implica la descripción detallada de los experimentos para que, de manera independiente, un investigador, profesional, instructor o estudiante pueda repetirlos y comprobarlos.

En 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, 2009) coinciden en que la transferencia efectiva del conocimiento en las áreas que involucran cómputo se logra al hacer disponible el programa de cómputo o paquetería en conjunto con las instrucciones y datos necesarias para obtener los resultados. Por lo tanto, para garantizar la reproducibilidad computacional es recomendable proporcionar y documentar los siguientes elementos:

- a) El artículo o informe técnico,
- b) El programa de cómputo o paquetería de trabajo,
- c) El experimento numérico o flujo de trabajo, incluyendo los códigos y la secuencia de instrucciones para generar los resultados,
- d) Los datos empleados y
- e) Los resultados del experimento, como figuras y datos.

De ese modo, el usuario de una paquetería 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:

1. Disminuye la dependencia de licencias de programas comerciales para llevar a cabo investigación.
2. Posibilita la adaptación, el desarrollo y transferencia de tecnología y
3. Auxilia la administración y transmisión eficiente del conocimiento.

Acceso abierto a la información

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, distribuirlas y mostrarlas de manera libre, bajo la condición de citarlas adecuadamente. Algunos proyectos que atienden la necesidad de difusión de la información son: *Public knowledge project* (disponible en: <http://pkp.sfu.ca/>), en 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 a la información, la reproducibilidad 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 en Austin (Texas, EUA), por mencionar algunas, e incluso en publicaciones científicas como la revista *Geophysics*, una de las publicaciones más importantes en geofísica aplicada. Sin embargo, en nuestro país la reproducibilidad en sismología de exploración, a nivel académico e industrial, no ha sido suficientemente explorada.

A continuación, veremos que está a nuestro alcance gracias a la disponibilidad de programas de código abierto, especialmente diseñados para la investigación reproducible.

Importancia del código abierto en representación sísmica

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, permitiéndole al usuario trabajar con los programas y modificarlos a su gusto, siempre y cuando se apegue a condiciones tales como: no revenderlos directamente o intentar impedir el acceso a los mismos.

La existencia de programas de código abierto representa un gran apoyo en investigación y docencia. Al investigador le permiten llevar a cabo su trabajo sin la necesidad de comprar licencias de paquetes comerciales de altos precios. 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.

En sismología de exploración, el procesamiento de datos experimenta innovaciones constantes a nivel académico e industrial e implican un alto grado de complejidad computacional. A escala industrial, dichas innovaciones suelen desarrollarse con miras a crear programas comerciales o mejorar los sistemas internos de las mismas compañías petroleras y de servicios; por razones de confidencialidad 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 aprovechar el potencial de las herramientas de código abierto y de tipo académico.

El panorama de los programas de código abierto en representación sísmica 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 paqueterías de cómputo reproducible, disponibles de manera gratuita para la comunidad. Lo anterior ha dado lugar a la realización continua de foros para la difusión y discusión del tema, 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>) o la publicación de artículos de difusión sobre la importancia del código abierto en Ciencias de la Tierra (Hall, 2010) y tesis académicas (e.g., González, 2010).

Un número importante de dichas paqueterías se distribuyen con licencias que permiten su uso gratuito para fines académicos y comerciales, y se pueden ejecutar en una computadora personal convencional con el sistema operativo adecuado (generalmente *Unix* o *Linux*). En el apéndice 1 se presenta en una tabla resumen, algunas paqueterías que cumplen con dicha característica y que satisfacen la mayoría de las necesidades de investigación en el procesamiento y representación de datos sísmicos. Estos programas de código abierto, a diferencia de sus equivalentes comerciales, aún tienen limitaciones, por ejemplo, el manejo espacial de los datos sísmicos (coordenadas geográficas de los datos) o la falta de interfaces

gráficas que faciliten su uso. No obstante, las comunidades que los desarrollan trabajan actualmente en esas tareas y tienen como meta alcanzar la funcionalidad de los programas comerciales.

La gratuidad de las paqueterías del apéndice 1 resulta muy importante, en términos económicos para las instituciones nacionales de educación superior e investigación, como el Instituto Mexicano del Petróleo (IMP), puesto que las licencias de los programas de procesamiento y representación sísmica están cotizadas hasta en decenas de miles de dólares y suelen estar limitadas a un número reducido de equipos de cómputo.

Documentos reproducibles en representación sísmica

Paqueterías como *SEPLib* y *Madagascar* (descrito más adelante) están encaminadas a la publicación de documentos reproducibles. Este tipo de documentos, disponibles en su mayoría en formato electrónico a través de Internet, tienen la característica de integrar los resultados vinculados a las secuencias de instrucciones (*script*) con las que se generaron. Además, las secuencias de instrucciones cuentan con vínculos a los códigos fuente de los programas empleados. Por lo tanto, el uso de este tipo de documentos permite una mejor comprensión y asimilación del cómputo requerido para generar los resultados.

En el apéndice 2 presentamos bibliografía fundamental de procesamiento y representación de datos sísmicos con la particularidad de ser reproducible. Esta bibliografía es de gran utilidad para quien desea incursionar de manera teórica y práctica en esta materia y los estudiantes de geociencias lo encontrarán particularmente valioso.

En el apéndice 3 presentamos una selección de documentos reproducibles como informes de consorcios académicos, tesis, artículos de investigación y datos públicos. Estos materiales incluyen algunos de los avances más relevantes en materia de procesamiento y representación de datos sísmicos y su disponibilidad permite acortar la brecha entre la teoría y la práctica. Para familiarizarse con el concepto de reproducibilidad le sugerimos al lector consultar la sección en línea de documentos reproducibles de la paquetería *Madagascar* [http://www.reproducibility.org/wiki/Reproducible_Documents].

Desarrollo

En la práctica petrolera mexicana resulta de gran importancia el post-procesamiento de datos sísmicos, par-

ticularmente la enfatización de imágenes sísmicas para mejorar su interpretación. Los datos sísmicos convencionales tienen resolución vertical limitada a decenas de metros (el criterio convencional de límite de resolución vertical es un cuarto de la longitud de onda sísmica) por lo que la identificación de rasgos geológicos de espesores menores al límite de resolución vertical es difícil.

Actualmente existe interés en incrementar la resolución sísmica vertical para mejorar la definición de las características geométricas de los yacimientos de hidrocarburos. La resolución sísmica depende del ancho de banda de la señal sísmica, por lo que un incremento de frecuencia, consistente con la geología, puede resultar en un incremento de resolución útil. Si el incremento de frecuencia no es consistente con la geología es posible que aparezcan artefactos numéricos.

Mostraremos el uso de la paquetería de código abierto *Madagascar* para probar dos operadores matemáticos de incremento de frecuencia de la señal sísmica y analizar los resultados.

El primer operador es el negativo de la segunda derivada de los datos; el segundo es un multiplicador de fase instantánea.

Nuestro equipo de trabajo fue una computadora personal, convencional, con sistema operativo *Linux*. El volumen de datos sísmicos que empleamos en nuestros ejemplos fueron proporcionados por PEMEX (Agua Fría-Coapechaca-Tajín, en Chicontepec) y tiene formato SEG-Y, el estándar de la industria sísmica.

Madagascar

Madagascar es una paquetería que integra elementos de sus precursores *SEPlib* y *Seismic Unix* (apéndice 1) empleando esquemas de cómputo moderno. La paquetería consta de tres niveles:

- 1) *Programas de procesamiento*. Más de 750 programas desarrollados principalmente en lenguaje C (también en C++, Fortran, Python, etc.) que actúan como filtros sobre los datos y pueden encadenarse mediante “*pipes*” de Unix. La lista de programas según el tipo de tareas de procesamiento y representación sísmica (modelado, migración, tomografía e inversión) se encuentra en: [http://www.reproducibility.org/wiki/Seismic_task-centric_program_list]. La clasificación de los programas de acuerdo al tipo de tareas está en: [http://www.reproducibility.org/wiki/Task-centric_program_list]. La documentación y los códigos fuente están en: [<http://www.reproducibility.org/RSF/>]

- 2) *Flujos de procesamiento SConstruct* o *scripts*. Escritos en lenguaje de programación *Python* y ejecutados mediante el programa *SCons*. Pueden ser verificados, intercambiados y modificados por los usuarios. Los códigos *SConstruct* están compuestos con las instrucciones

Flow- Flujo de procesamiento. Estructura: *Flow* (‘*Archivo de salida*’, ‘*Archivo de entrada*’, ‘*Programas y parámetros de procesamiento*’).

Plot- Gráficas intermedias. Estructura: *Plot* (‘*Gráfica de salida*’, ‘*Archivo de entrada*’, ‘*Programas y parámetros de graficación*’).

Result- Gráficas finales, compuestas de gráficas intermedias. Estructura: *Result* (‘*Grafica final*’, ‘*Gráficas de entrada*’, ‘*Parámetros de graficación*’).

- 3) *Documentos reproducibles*. Conjunto de libros y artículos técnicos. La lista está en: [http://www.reproducibility.org/wiki/Reproducible_Documents].

El formato interno de datos sísmicos de *Madagascar* se denomina RSF (regularmente muestreado, por sus siglas en inglés) y es compatible con el formato SEG-Y. La lectura del volumen de datos sísmicos y cambio de formato SEG-Y a formato RSF se hace con la instrucción *Flow* de la línea 12 y 13 del flujo *SConstruct* que presentamos en el apéndice 4. La manipulación del volumen de los datos, así como la selección de una sección vertical del volumen está indicada en las líneas 15 a 19 del flujo *SConstruct*.

Negativo de la segunda derivada

Calcular el negativo de la segunda derivada de los datos sísmicos incrementa la frecuencia y en ciertos casos mejora la continuidad y definición de los reflectores 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 la 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 corrimiento del espectro 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 obtener un mayor corrimiento hacia altas frecuencias. En ambos casos hay un incremento en el ruido, sin embargo, el resultado final es interpretable.

El operador del negativo de la segunda derivada está incluido en paquetes comerciales de procesamiento de datos sísmicos. Implementamos el flujo de procesamiento en *Madagascar* para obtener resultados equivalentes. La figura 1 presenta el resultado de aplicar los operadores de negativo de la segunda y cuarta derivadas, seguidos por un control automático de ganancia para una línea longitudinal del volumen de datos sísmicos.

El resultado de la segunda derivada es satisfactorio si se compara la continuidad y definición de los reflectores sísmicos con los datos originales, como se aprecia en los reflectores de la zona de 1.5 s; el resultado de la cuarta derivada también presenta mejor continuidad y defini-

ción, pero presenta un incremento de ruido más notorio. El código para obtener los resultados lo presentamos en el apéndice 4, en las líneas 21 a 33.

La figura 2 muestra 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 un corrimiento del espectro hacia las altas frecuencias. Este corrimiento es un producto matemático y aunque los resultados pueden ser interpretables, la consistencia con la geología resulta cuestionable. Presentamos el código para la obtención de los espectros de amplitud en el apéndice 4, en las líneas 35 a 43.

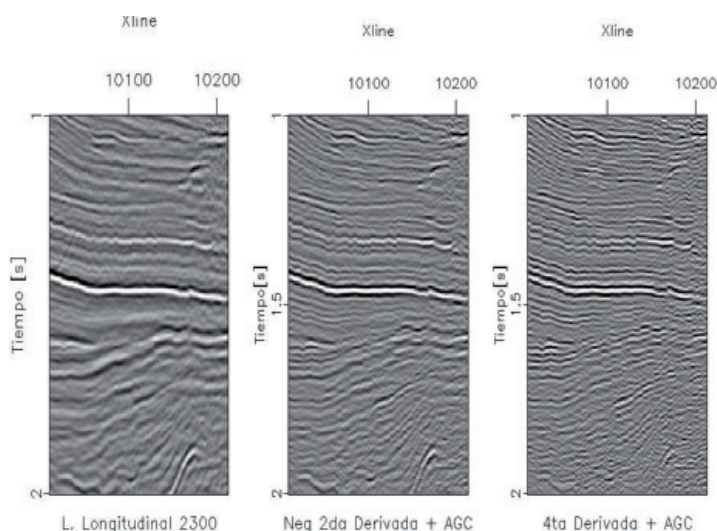


Figura 1. Izquierda: sección original. Centro: negativo de la segunda derivada; es notable la mayor definición y continuidad de los eventos. Derecha: cuarta derivada; también hay incremento en la resolución, pero el ruido aumenta. La distancia entre trazas sísmicas es de 25 m

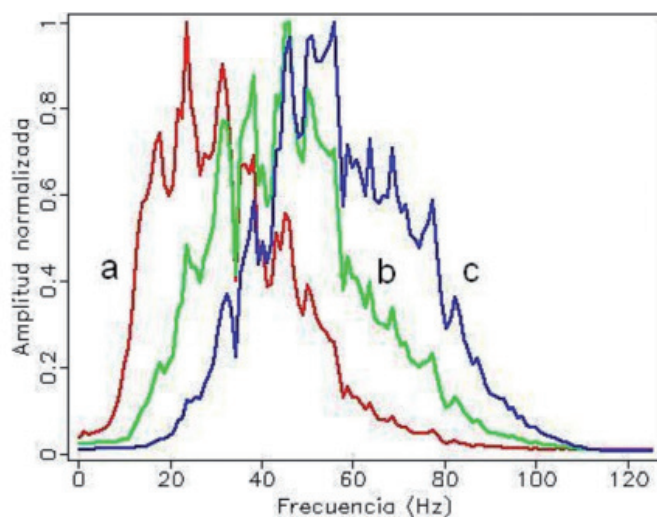


Figura 2. Espectros de amplitud, promediados y normalizados, antes (a) y después de obtener el negativo de la segunda derivada (b) y cuarta derivada (c). Nótese el corrimiento de las frecuencias dominantes hacia las altas frecuencias después de aplicar los operadores

Multiplicador de fase instantánea N

Recientemente, Stark (2009) propuso un algoritmo para incrementar el número de cimas y valles en una traza sísmica. Los resultados son interpretables y pueden enfatizar información geológica, 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. 1), y se expresa como en la ec. (2):

$$x(t) = A(t) \cos(\Theta(t)), \quad (1)$$

$$x_N(t) = A(t) \cos(N \Theta(t)), \quad (2)$$

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.

La figura 3 presenta los resultados para $N=2$ y $N=3$ en la sección vertical del volumen sísmico que utilizamos anteriormente. Los resultados para $N=2$ son aparentemente satisfactorios en términos de definición y de continuidad de eventos. No obstante, hay aparición de reflectores sísmicos espurios que no corresponden necesariamente a la geología del subsuelo. Estructuras

como pliegues y discontinuidades pueden hacerse más notorios con este operador. El caso $N=3$ introduce más reflectores espurios y degrada la imagen. El código para obtener los resultados lo presentamos en el apéndice 4, en las líneas 45 a 76.

La figura 4 muestra espectros de amplitudes, normalizados, correspondientes a las secciones de la figura 3 después de aplicar los operadores. Hay un corrimiento del espectro hacia las altas frecuencias, pero dicho incremento no se apegaba al sentido físico de los datos sísmicos. Presentamos el código correspondiente en el apéndice 4, en las líneas 79 a 84.

Discusión

Los resultados presentados ejemplifican la posibilidad de asimilar y adaptar tecnología de interés en la industria petrolera, en un contexto académico o industrial, a bajo costo, empleando paqueterías de cómputo reproducible instaladas en una computadora personal convencional con un sistema operativo tipo *Linux*.

Adaptamos una propuesta para enfatizar la imagen sísmica basada en el incremento de frecuencia de la señal sísmica, obteniendo el negativo de la segunda derivada de los datos sísmicos. Los resultados son satisfactorios en términos de definición y continuidad de los reflectores sísmicos. Debido a la naturaleza matemática del operador los resultados son limitados y no son necesariamente consistentes con la geología.

Adaptamos una propuesta reciente para el incremento de resolución vertical basado en un multiplicador N de fase instantánea. Los resultados para incrementos únicos (*e.g.*, $N=2$) mejoran la continuidad de los eventos y destacan rasgos geológicos como fallas y plie-

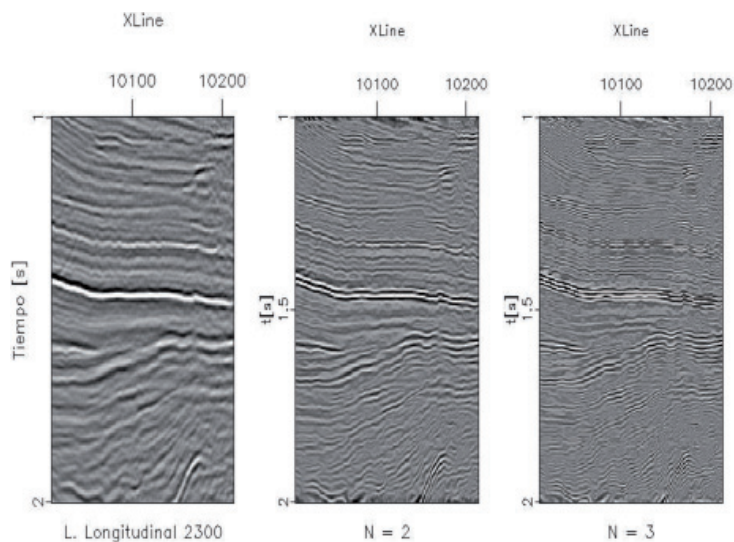


Figura 3. Derecha: sección original. Centro: incremento de resolución vertical con $N=2$; se enfatizan pliegues y discontinuidades. Derecha: incremento de resolución vertical con $N=3$; este operador introduce artefactos numéricos. La distancia entre trazas es de 25 m

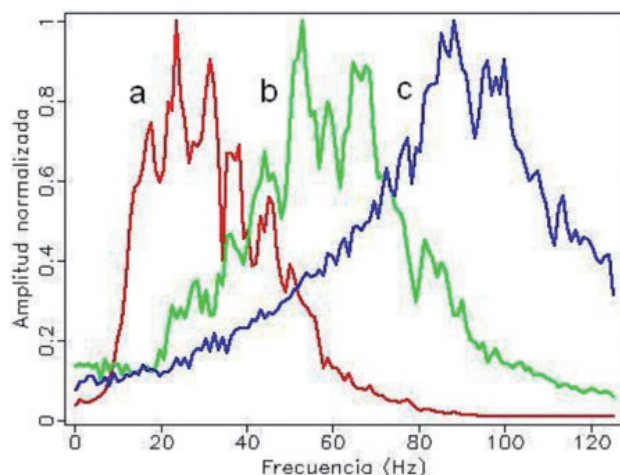


Figura 4. Espectros de amplitud antes (a) y después de aplicar el operador de incremento con $N=2$ (b) y $N=3$ (c). Nótese el corrimiento de las frecuencias dominantes hacia las altas frecuencias después de aplicar el operador

gues; debe tomarse en cuenta que este método genera artefactos numéricos. Los resultados para $N=3$ introducen muchos reflectores sísmicos espurios. Una explicación es que el incremento en la frecuencia provoca submuestreo de las señales con frecuencia mayor a la frecuencia de Nyquist dividida entre el valor N . Con este ejemplo mostramos que aunque es sencillo obtener un incremento de frecuencia, se debe tener cautela con los operadores utilizados para ello, pues no se apegan necesariamente al sentido físico de los datos.

Los flujos de trabajo o secuencias de instrucciones con formato reproducible (*e.g.*, el apéndice 4) pueden adaptarse, con relativa sencillez, a datos nuevos.

Dichos flujos facilitan la detección de errores durante las implementaciones con nuevos datos. Además, por su estructura facilitan la comprensión de la relación entre los programas a ejecutar y los archivos de entrada y salida.

Conclusiones

1. La reproducibilidad es de gran utilidad en la docencia, investigación, capacitación de profesionales e incluso en la práctica profesional. Representa beneficios académicos, técnicos y económicos como:
 - a) disminuir la dependencia de licencias de programas comerciales para llevar a cabo investigación,
 - b) posibilitar la adaptación, el desarrollo y transferencia de tecnología y

c) auxiliar la administración y transmisión eficiente del conocimiento.

2. Puede advertirse que la publicación con formato reproducible cobrará mayor importancia en el futuro. Por ese motivo, consideramos que la investigación en representación sísmica, en el ámbito nacional, debe atender dicha tendencia para hacer eficiente la administración y divulgación del conocimiento.
3. 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 de 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. Dada la diversidad de programas de código abierto, consideramos que existen herramientas que ayudan a resolver algunos de los problemas básicos de representación sísmica y de interés actual.
4. Como prueba de lo anterior, efectuamos actividades realizadas en la industria de la exploración sísmica en una computadora personal convencional a la que le instalamos una paquetería de código abierto de procesamiento de datos sísmicos y desarrollamos flujos de trabajo para enfatizar imágenes sísmicas.
5. Los documentos reproducibles fueron un apoyo fundamental para la aplicación de las herramientas que empleamos y nos permitieron desarrollar las actividades que presentamos.
6. Comprobamos que es posible asimilar, adaptar, desarrollar y transferir tecnología de interés en la industria petrolera a bajo costo, empleando paqueterías de código abierto y siguiendo un esquema de cómputo reproducible.

Agradecimientos

A Marco Vázquez García, Gerente de Estudios Geofísicos de PEMEX Exploración y Producción, por proporcionar financiamiento, datos sísmicos y permiso para publicarlos. A Liliana Vargas Meleza, del Instituto Mexicano del Petróleo, cuyas observaciones y revisión fueron un gran apoyo y a Cinna Lomnitz, del Instituto de Geofísica de la UNAM, por sus comentarios en torno al acceso abierto a la información.

Apéndice 1. Programas de código abierto que satisfacen las necesidades de investigación básicas en el procesamiento de datos sísmicos y representación sísmica

Nombre (Año de aparición)	Descripción y características especiales	Escrito/ mantenido por	Licencia ¹	SO ²	Sitio en Internet
Madagascar (2006)	Paquetería 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, SEPlib y SU.	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, SEPlib y SU.	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
Seismic Unix (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
SEPlib (1991)	Primera paquetería para elaboración y uso de documentos reproducibles. Permite leer documentos reproducibles del SEP (1989 - 2009; documentos 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)	"SEPLIB se puede usar o modificar para uso propio; no se puede distribuir, 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 2. Bibliografía reproducible sobre sismología de exploración

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

Apéndice 3. Documentos reproducibles y datos disponibles en línea

Tipo	Institución / autor	Descripción	Paquetería	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 velocidad	Madagascar	http://www.reproducibility.org/wiki/Reproducible_Documents
Datos	Varios	Registros de campo 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 4. Flujo de procesamiento *SConstruct*

```

1  from rsfproj import *
2
3  #####CAMBIO DE FORMATO SEG Y 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 SEG Y a RSF
12 Flow('sismica_original','sergio_sub.segy', '''segyread tfile=tfile1.rs
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 ***ENFATIZACION DE IMAGEN CON NEGATIVO DE SEGUNDA DERIVADA*
22
23 #Negativo de segunda derivada
24 Flow('neg2da','insub','deriv | deriv | scale dscale=-1')
25 Plot('neg2da','neg2da','''agc | grey title='Negativo 2da Derivada + AGC'
26                                     label1='Tiempo[s]' label2='Xline' scalebar=y''')
27
28 #Cuarta Derivada
29 Flow('4da','insub','deriv | deriv | deriv | deriv')
30 Plot('4da','4da','''agc | grey title='4ta Derivada + AGC'
31                                     label1='Tiempo[s]' label2='Xline' scalebar=y''')
32
33 Result('F1','insub neg2da 4da','SideBySideAniso')
34
35 #Espectros de Amplitud
36 Flow('espinsub','insub','spectra all=y | scale axis=1')
37 Flow('espneg2da','neg2da','spectra all=y | scale axis=1')
38 Flow('esp4da','4da','spectra all=y | scale axis=1')
39
40 Flow('espcompderiv','espinsub','cat espneg2da.rs
41                                     f esp4da.rs
42                                     f axis=2 ')
43
44 Result('F2','espcompderiv','''graph label1='Frecuencia (Hz)'
45                                     title='Espectros de Amplitud' plotfat=7 plotcol=5,3,6''')
46
47 #####INCREMENTO DE FRECUENCIA CON MULTIPLICADOR DE FASE ***
48
49 #####OBTENCION DE ATRIBUTOS SISMICOS INSTANTANEOS#####
50
51 #Cuadratura
52 Flow('hilb','insub','envelope hilb=y order=75')

```

```

51
52 #Envolvente de amplitud
53 Flow('amp',None,'''math x=insub.rsf y=hilb.rsf
54         output='sqrt(x^2 +y^2)' ''')
55
56 #Coseno de la fase instantanea
57 Flow('cos',None,'''math x=insub.rsf a=amp.rsf output="x/a" | put o2=10010 '
58
59 #Fase instantanea
60 Flow('fase',None,'''math cos=cos.rsf output="acos(cos)" | put o2=10010 '
61
62 #CASO N=2
63 Flow('insubN2',None,'''math amp=amp.rsf phi=fase.rsf
64         output='amp*cos(2*phi)' | put o2=10010''')
65 Plot('insubN2','insubN2','''agc | grey
66         title='N = 2'
67         label2="XLine" label1='t[s]' scalebar=y ''')
68
69 #CASO N=3
70 Flow('insubN3',None,'''math amp=amp.rsf phi=fase.rsf
71         output='amp*cos(3*phi)' | put o2=10010''')
72 Plot('insubN3','insubN3','''agc | grey
73         title='N = 3'
74         label2="XLine" label1='t[s]' scalebar=y ''')
75
76 Result('F3','insub insubN2 insubN3','SideBySideAniso')
77
78
79 #Espectros de Amplitud N=2 y N=3
80 Flow('espinsubN2','insubN2','spectra all=y | scale axis=1')
81 Flow('espinsubN3','insubN3','spectra all=y | scale axis=1')
82 Result('F4','espinsub','''cat espinsubN2.rsf espinsubN3.rsf axis=2 |
83         graph label1="Frecuencia (Hz)"
84         title='Espectros de Amplitud' plotfat=7 plotcol=5,3,6''')
85 End()

```

Nota: Los signos de número (#) son comentarios. La numeración se incluyó sólo para fines de referencia.

Referencias

- Donoho D.L., Maleki A., Morteza S., Ur-Rahman I., Stodden V. Reproducible Research in Computational Harmonic Analysis. *Computing in Science & Engineering*, 11(1):8-18, enero-febrero 2009.
- Fomel S. y Claerbout J. Reproducible Research. *Computing in Science & Engineering*, 11(1):5-7, enero-febrero 2009.
- González-Verdejo O. *Reproducibilidad en representación sísmica*. Tesis (licenciatura en ingeniería geofísica), México DF, Universidad Nacional Autónoma de México, Facultad de Ingeniería, Febrero de 2010, 71p.
- Hall M., The Future is Open for Business: Open Source Tools for the Geoscientist. *First Break*, 28(6):119-123, junio de 2010.
- LeVeque R.J. Python Tools for Reproducible Research on Hyperbolic Problems. *Computing in Science and Engineering*, 11(1):19-27, enero-febrero 2009.
- Stark T.J. Frequency Enhancement Via an Integer Multiplier or Just Another GeoWizardry Attribute?, en: Reunión anual de la Sociedad de Geofísicos de Exploración, (79^a, 2009, Houston, EEUU). Resumen disponible a través de la SEG en: <http://www.seg.org>
- Stodden V. The Legal Framework for Reproducible Scientific Research. *Computing in Science & Engineering*, 11(1):35-40, enero-febrero 2009.
- Stodden V. Enabling Reproducible Research: Open Licensing for Scientific Innovation [en línea]. Universidad de Stanford, California, EUA, 2008. [fecha de consulta marzo de 2010]. Disponible en: <http://www.stanford.edu/~vcs/papers/Licensing08292008.pdf>

Semblanza de los autores

Oziel González-Verdejo. Obtuvo el título de ingeniero geofísico en la Facultad de Ingeniería de la UNAM en 2010. Actualmente trabaja en la empresa de servicio EMGS. Su área de interés profesional es la sismología de exploración.

Sergio Chávez-Pérez. Obtuvo los títulos de ingeniero geofísico en 1984 por la Facultad de Ingeniería de la UNAM, el de maestro en ciencias (geofísica) en 1987 por la Universidad de Carolina del Sur, EUA y el de doctor en filosofía (geofísica) en 1997 por la Universidad de Nevada, Reno, EUA. Actualmente es investigador en el Instituto Mexicano del Petróleo y profesor de asignatura de la Facultad de Ingeniería de la UNAM. Sus áreas de interés profesional son: representación sísmica (migración, modelado, tomografía e inversión), procesamiento de datos sísmicos y transformación de campos de ondas, interpretación de datos sísmicos y propagación de ondas sísmicas.