Reproducible Computations Using MADAGASCAR Software Package

Sergey Fomel

Jackson School of Geosciences
The University of Texas at Austin

June 12, 2009
### Agenda

<table>
<thead>
<tr>
<th>Friday, June 12</th>
<th>Morning</th>
<th>Afternoon</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Morning</strong></td>
<td>Sergey Fomel</td>
<td>Arnaud &amp; Hélène Huck</td>
</tr>
<tr>
<td><strong>Afternoon</strong></td>
<td></td>
<td>Seismic Interpretation</td>
</tr>
<tr>
<td><strong>Saturday, June 13</strong></td>
<td>Morning</td>
<td>Afternoon</td>
</tr>
<tr>
<td><strong>Morning</strong></td>
<td>Paul Sava</td>
<td>Ivan Vasconcelos</td>
</tr>
<tr>
<td><strong>Afternoon</strong></td>
<td></td>
<td>Seismic Imaging</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Seismic Interferometry</td>
</tr>
</tbody>
</table>

Delft University of Technology

School on Reproducible Computational Geophysics
Outline

History of Reproducible Research
Outline

History of Reproducible Research

History of MADAGASCAR
Outline

History of Reproducible Research

History of MADAGASCAR

MADAGASCAR Components
Outline

History of Reproducible Research

History of MADAGASCAR

MADAGASCAR Components
Black Magic in Geophysical Prospecting

L. W. Blau, 1936
Black Magic in Computational Science

Within the world of science, computation is now rightly seen as a third vertex of a triangle complementing experiment and theory. However, as it is now often practiced, one can make a good case that computing is the last refuge of the scientific scoundrel [...] Where else in science can one get away with publishing observations that are claimed to prove a theory or illustrate the success of a technique without having to give a careful description of the methods used, in sufficient detail that others can attempt to repeat the experiment? *Randall LeVeque, ICM, 2006*
(Hale, 1984)

\[ p_n(t_n, y, h) \]

\[ \text{1-D FFT} \]

\[ P_n(t_n, k, h) \]

\[ \text{integrate over } t_n \text{ for all } \omega_0 \text{ and } k \]

\[ P_0(\omega_0, k, h) \]

\[ \text{2-D FFT}^{-1} \]

\[ p_0(t_0, y, h) \]

**Fig. 1.** The seismic experiment, conducted over a simplified subsurface with one dipping reflector. Applying the law of cosines to triangle \( s't's \), one may express the travel time \( t \) from source \( s \) to receiver \( r \) in terms of zero-offset time \( t_0 \), half-offset \( h \), velocity \( v \), and dip \( \theta \). The result is equation (3) in the text, the

Defining

\[ A = \frac{dt_n}{dt_0} = \frac{t_0}{t_n} = 1 + \left( \frac{\Delta t_0}{\Delta y} \right)^2 \frac{h^2}{t_n^2} \]

and using equation (10) to replace \( p_0(\sqrt{t_n^2 + (\Delta t_0/\Delta y)^2 h^2}, y, h) = p_n(t_n, y, h) \), the Fourier transform becomes

\[ p_0(\omega_0, k, h) = \int dt_n A^{-1} e^{i\omega_0 t_n} \int dy e^{-iky} p_n(t_n, y, h) \]  

(12a)
(Hale, 1984)
What is Science?
What is Science?

Science is the systematic enterprise of gathering knowledge about the universe and organizing and condensing that knowledge into testable laws and theories. The success and credibility of science are anchored in the willingness of scientists to independent testing and replication by other scientists. This requires the complete and open exchange of data, procedures and materials.

American Physical Society, What is science
Abandoning the habit of secrecy in favor of process transparency and peer review was the crucial step by which alchemy became chemistry. In the same way, it is beginning to appear that open-source development may signal the long-awaited maturation of software development as a discipline.

*Eric Raymond, TAUP, 2004*
What is Reproducible Research?

- Attaching code and data to publications
- Code requires continuous maintenance
- Maintenance requires an open community

An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures. Jon Buckheit and David Donoho, WaveLab, 1995
Jon Claerbout’s Story

1987  Sunview experience
  ▶  Interactive programs are slavery

1992  \LaTeX{} + cake
  ▶  Building books by a single command

1990s  Ph.D. students
  ▶  cake to make, CD-Rom to WWW

2001  Reproducible research paper in *CiSE*
  ▶  The principal beneficiary is the author
Moving Forward

ICASSP 2007
Berlin-6 2008
CiSE 2009

- Fomel & Claerbout
- Donoho et al.
- LeVeque
- Ping & Eckel
- Stodden

IEEE Signal Processing Magazine 2009

- Vandewalle et al.

http://www.reproducibleresearch.net
Outline

History of Reproducible Research

History of MADAGASCAR

MADAGASCAR Components
Basic Information

- Started around 2003
- Publicly available since June 12, 2006
- Current version: 0.9.8
- Vladimir Bashkardin, Jules Browaeys, Cody Brown, Maria Cameron, Joseph Dellinger, Sergey Fomel, Gilles Hennenfent, Trevor Irons, Jim Jennings, Long Jin, Guochang Liu, Yang Liu, Doug McCowan, Henryk Modzelewski, Colin Russell, Paul Sava, Jeffrey Shragge, Xiaolei Song, Eduardo Filpo Silva, Ioan Vlad, Jia Yan
- Jon Claerbout, Steve Cole, Dave Hale, Chuck Karish, Stewart Levin, Dave Nichols, Shuki Ronen
- [http://www.ahay.org/](http://www.ahay.org/)
Website Traffic: April 2009/April 2008

Twitter: 1,300%
MADAGASCAR: 970%
Facebook: 220%
Google: 10%
Myspace: -10%

- How Twitter Will Change the Way We Live
- Steven Johnson, *TIME*, June 5, 2009
Access Geography
School and Workshop: Vancouver 2006
School: Austin 2007
Developer Workshop: Golden 2008
Outline

History of Reproducible Research

History of MADAGASCAR

MADAGASCAR Components
One Week Technology Transfer

Monday: Get an idea
Tuesday: Implement it
Wednesday: Test it
Thursday: Communicate it
Friday: Apply it in practice
MADAGASCAR Components

Tuesday: Implement it
- Main programs (C, C++, Fortran, etc)
- 600 modules

Wednesday: Test it
- Data processing flows (Python/SCons)
- 300 scripts $\rightarrow$ 2,400 figures

Thursday: Communicate it
- Books and papers (LaTeX/SCons)
- 100 papers
MADAGASCAR Objectives

- To make computational research efficient
- To make it easy to share computational results
- To maintain an open community
History of Reproducible Research

MADAGASCAR Design Principle

▶ Document computational experiments and use them in the future as regression tests
▶ Reproducible research
▶ Test-driven development
▶ YAGNI (You Ain’t Gonna Need It)

*Always implement things when you actually need them, never when you just foresee that you need them. Ron Jeffries, YAGNI*
Conclusions

- Reproducible computational experiments
- [http://www.ahay.org/](http://www.ahay.org/)
- Help is needed