Reproducible Geophysics
Archiving Experiments in the MADAGASCAR Project

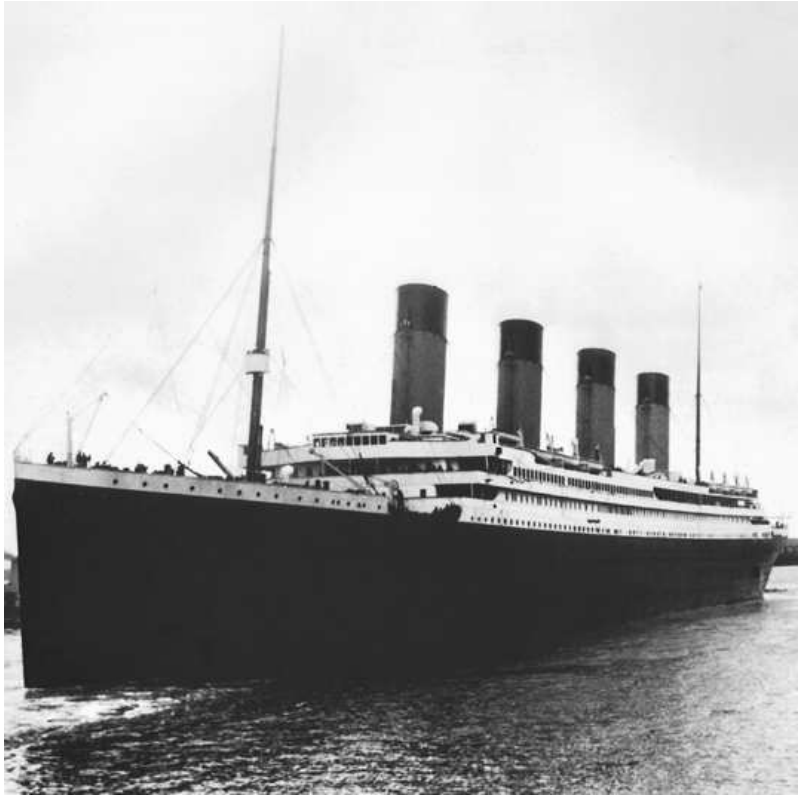
Sergey Fomel

Jackson School of Geosciences
The University of Texas at Austin

May 25, 2010

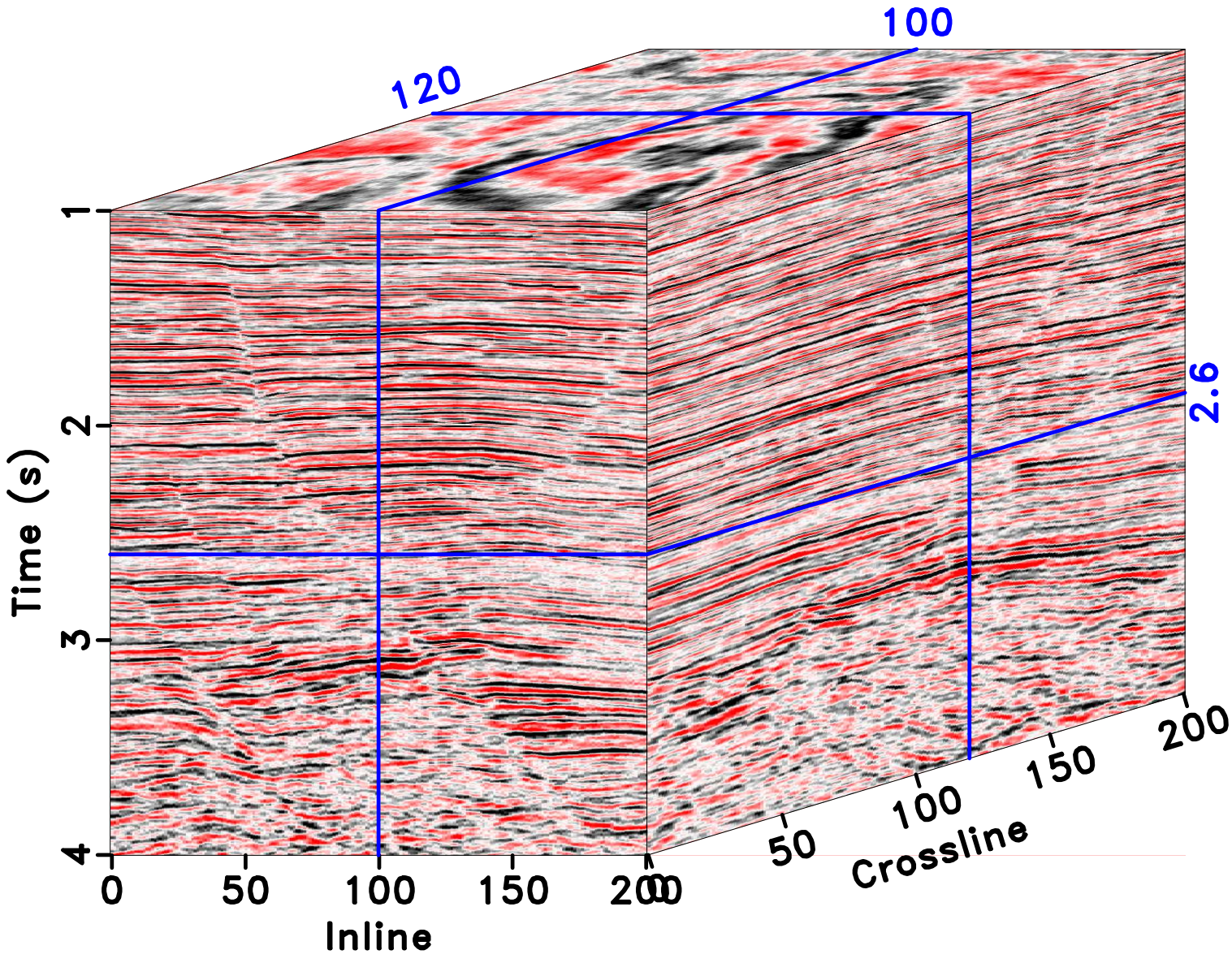
Outline

Largest Moving Object on Earth



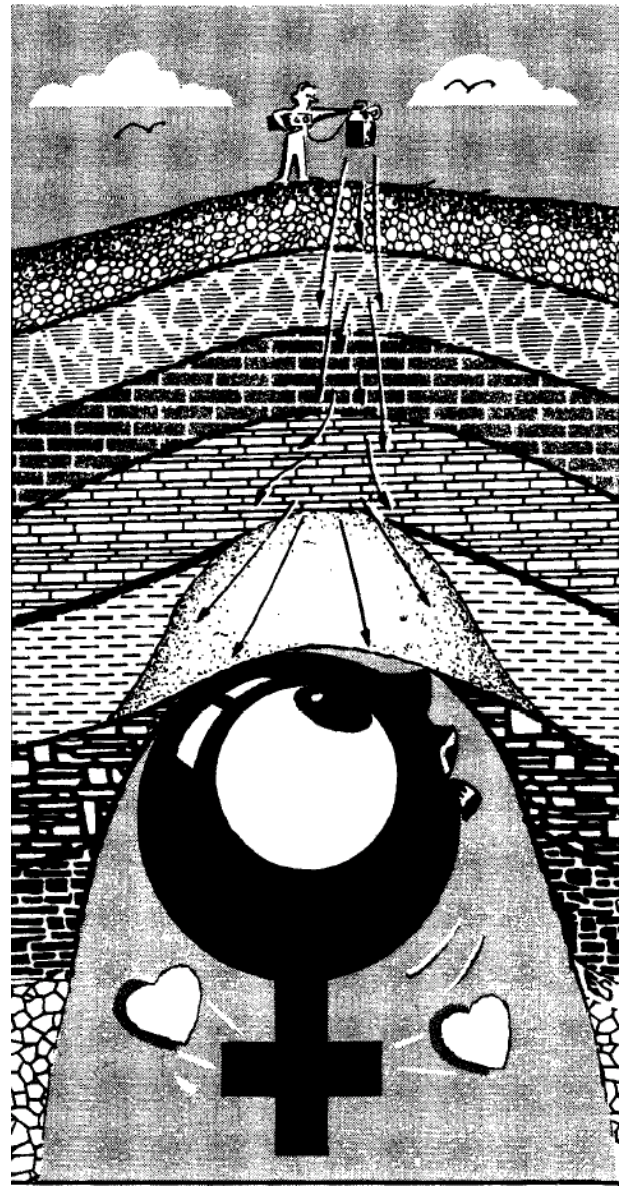
1912

2009


100
120
2.6
Time (s)
1
2
3
4
0
50
100
150
200
Inline
50
100
150
200
Crossline

Black Magic in Geophysical Prospecting

L. W. Blau, 1936

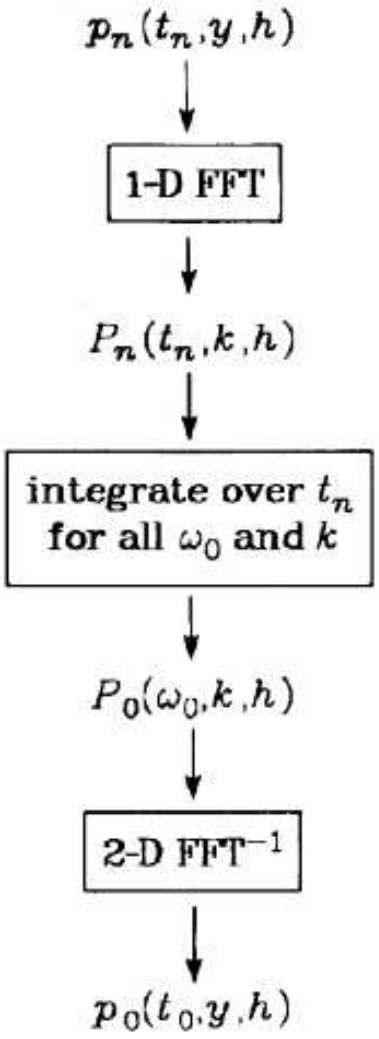


Black Magic in Computational Science

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)



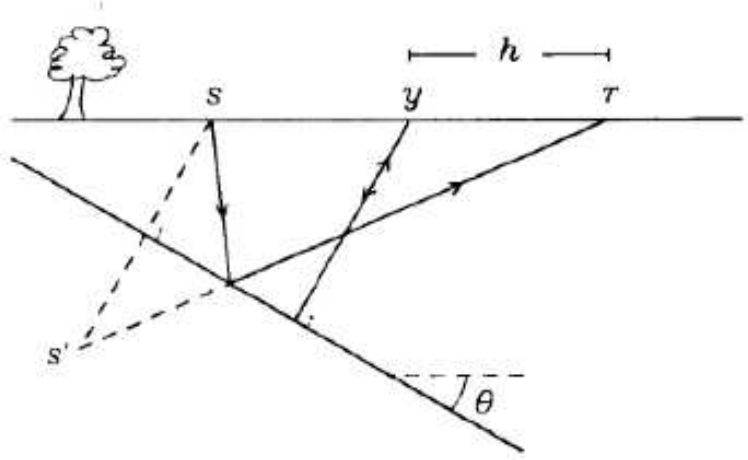


FIG. 1. The seismic experiment, conducted over a simplified subsurface with one dipping reflector. Applying the law of cosines to triangle $s'sr$, one may express the travel time t from source s to receiver r in terms of zero-offset time t_0, half-offset h, velocity v, and dip θ. The result is equation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

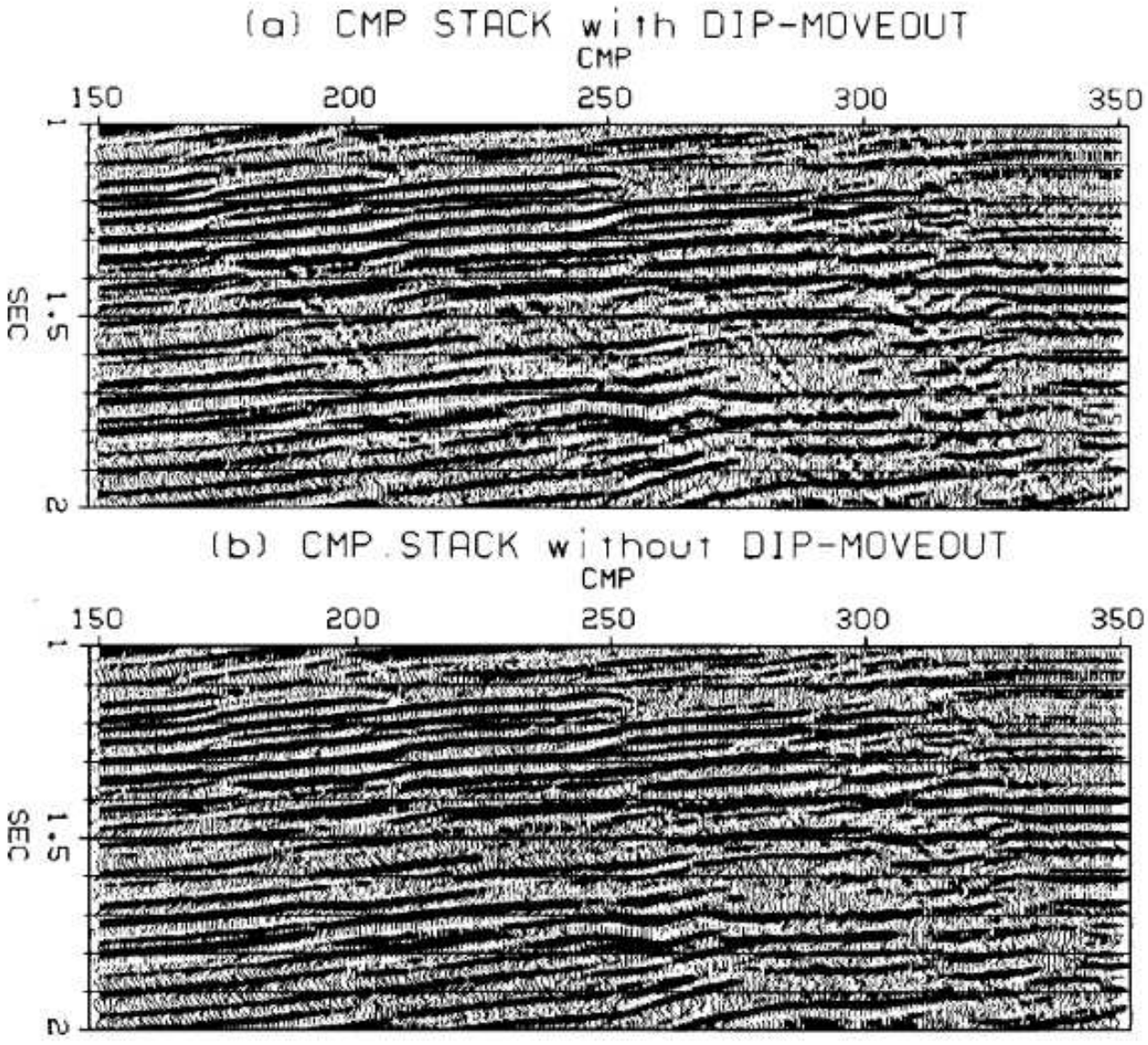
Defining

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

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

$$P_0(\omega_0, k, h) = \int dt_n \, A^{-1} \, e^{i\omega_0 t_n A} \int dy \, e^{-iky} p_n(t_n, y, h). \qquad (12a)$$

(Hale, 1984)


(a) CMP STACK with DIP-MOVEOUT
CMP
150
200
250
300
350
1
1.5
2
SEC
(b) CMP STACK without DIP-MOVEOUT
CMP
150
200
250
300
350
1
1.5
2
SEC

Outline

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?

What is Reproducible Research?

- Attaching software code and data to publications

  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*

Reproducible Research Discussions

- http://www.reproducibleresearch.net

ICASSP 2007

Berlin-6 2008

CiSE 2009

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

IEEE Signal Processing Magazine 2009

- Vandewalle et al.

Yale Roundtable 2009

Personal Experience

1991–2001 Jon F. Claerbout

- Stanford Exploration Project
- Generations of Ph.D. students
- The principal beneficiary is the author

2003–Present MADAGASCAR package

- Software code requires continuous maintenance
- Maintenance requires an open community
- http://www.reproducibility.org/

Outline

Facts

http://www.ahay.org/
http://www.reproducibility.org/

- Publicly released in 2006 (GPL)
- Approaching 1.0 release in 2010
- 25+ developers
- 250,000+ lines of code
- 10,000 downloads from SourceForge
- 80 reproducible papers; 3,000 reproducible results
  - http://www.ahay.org/wiki/Reproducible_Documents

MADAGASCAR architecture

Recipes

- Main programs operating on data files
- C, C++, Fortran, Java, Python, Matlab
- Regularly Sampled Format
- Unix pipes
- SCONS data processing flows

Inputs

- Data repository

Results

- Figures included in publications
- Archived to serve as regression tests
- Links from programs to source code and recipes
- Links from recipes to programs and data
- Hierarchy `book/chapter/project`
- LaTeX2HTML

Thanks

- Vladimir Bashkardin, Jules Browaeys, William Burnett, Cody Brown, Maria Cameron, Lorenzo Casasanta, Joseph Dellinger, Jeff Godwin, Gilles Hennenfent, Trevor Irons, Jim Jennings, Long Jin, Roman Kazinnik, Siwei Li, Guochang Liu, Yang Liu, Doug McCowan, Henryk Modzelewski, Colin Russell, Paul Sava, Jeffrey Shragge, Xiaolei Song, Eduardo Filpo Silva, Ioan Vlad, Jia Yan

Conclusions

- Reproducible research
  - Attaching software and data to publications
  - Computational experiments
  - Reproducibility is not the goal!
- Lessons from experience in computational geophysics
  - The principal beneficiary is the author
  - Software code requires continuous maintenance
  - Maintenance requires an open community!