

Reproducibility as a Community Effort

Lessons from the Madagascar Project



Sergey Fomel

Jackson School of Geosciences

The University of Texas at Austin

12/13/2012

ICERM Reproducibility in Computational and Experimental Mathematics

What is Science?



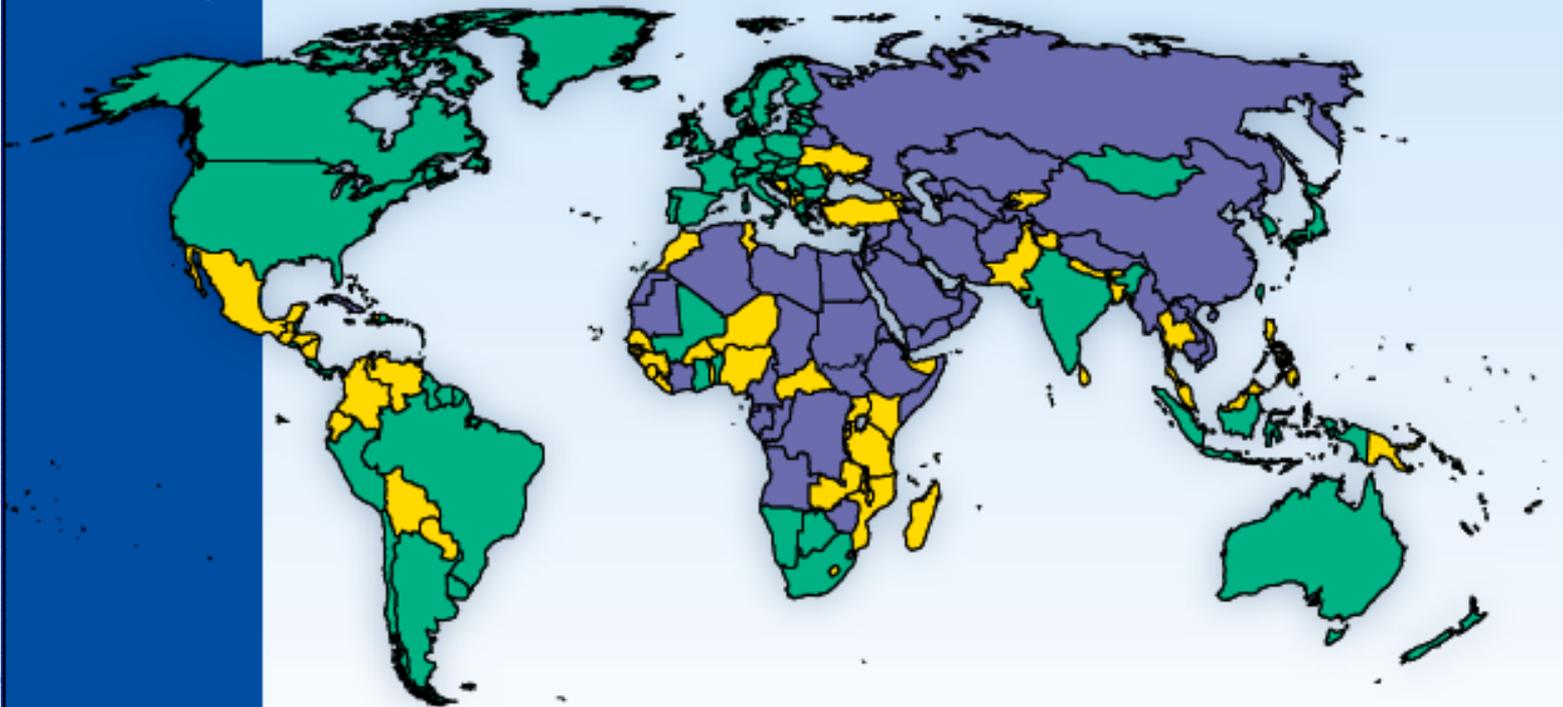
What is Science?

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 expose their ideas and results to **independent testing and replication** by other scientists. This requires the **complete and open exchange of data, procedures and materials.**





Freedom in the World 2012



Second Paper Published in *Geophysics*

ON SEISMIC COMPUTATIONS, WITH APPLICATIONS, I¹

M. M. SLOTNICK²

INTRODUCTION

The problem of depth determinations from data obtained by the method commonly called "reflection shooting" is an interesting one and one which, very often, assumes an importance that justifies a good deal of study. It is the purpose of this paper to indicate a method

¹ Published by permission of the Board of Directors, Humble Oil and Refining Company.

² Mathematician, Geophysics Department, Humble Oil and Refining Company.

First Paper Published in *Geophysics*

VOLUME I

JANUARY, 1936

NUMBER 1

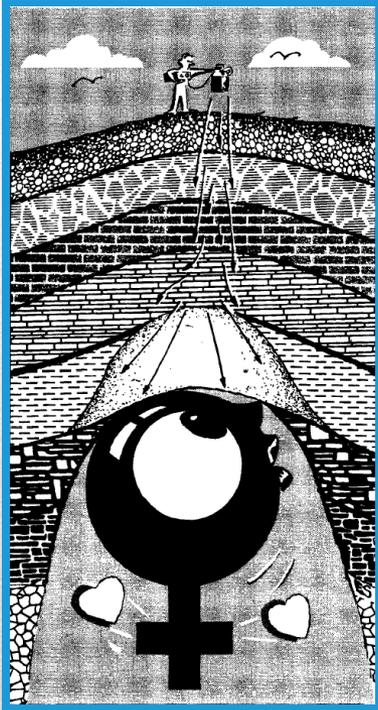
GEOFYSICS

A Journal of General and Applied Geophysics

BLACK MAGIC IN GEOPHYSICAL PROSPECTING¹

L. W. BLAU²

EDITOR'S NOTE: *The term "doodle-bug" is coming more and more to mean proposed methods of geophysical prospecting that are neither based upon scientific fact nor upon known or proven properties of oil, minerals and geologic formations. The geophysicist is often consulted concerning the reliability of such a proposed method, and his task then is to explain scientifically just why the proposed method fails and is unsuitable for the intended purpose.*



$$p_n(t_n, y, h)$$

1-D FFT

$$P_n(t_n, k, h)$$

integrate over t_n
for all ω_0 and k

$$P_0(\omega_0, k, h)$$

2-D FFT⁻¹

$$p_0(t_0, y, h)$$

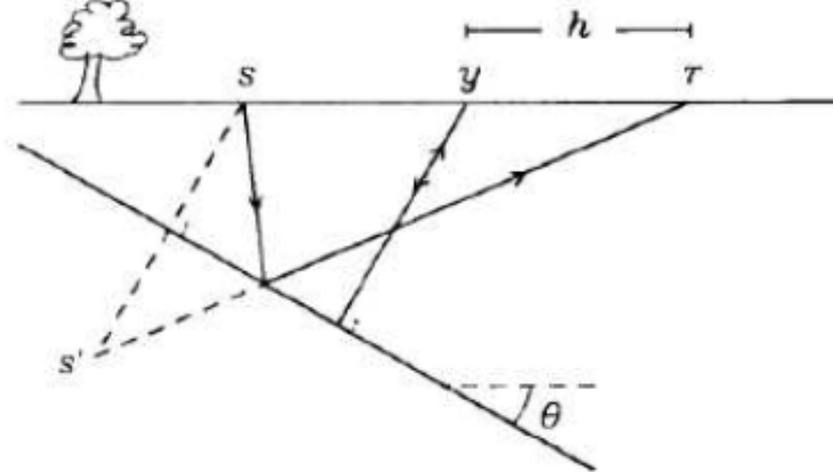


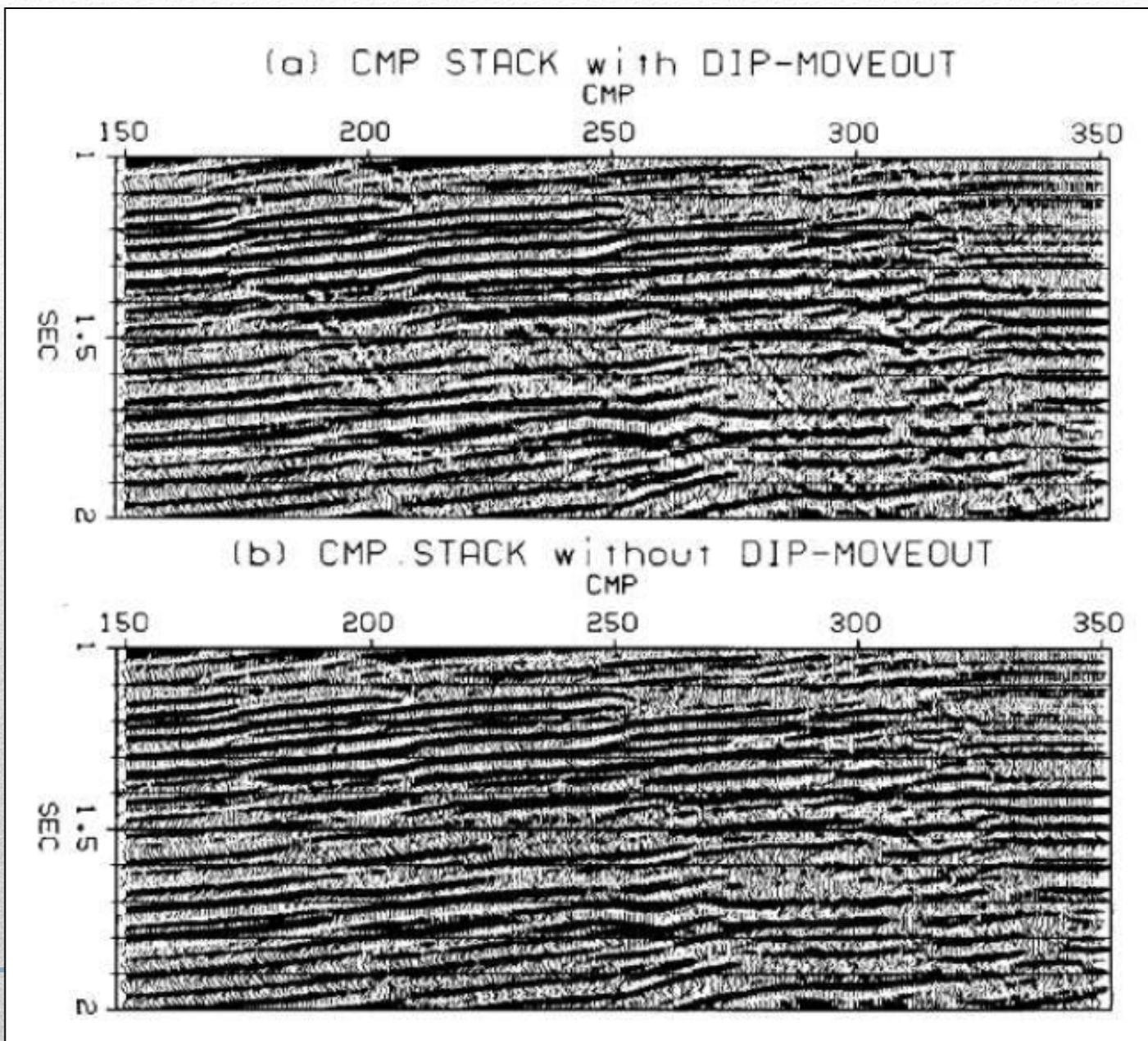
FIG. 1. The seismic experiment, conducted over a simplified subsurface with one dipping reflector. Applying the law of cosines to triangle $s'sr$, one may express the travel time t from source s to receiver r in terms of zero-offset time t_0 , half-offset h , velocity v , and dip θ . The result is equation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

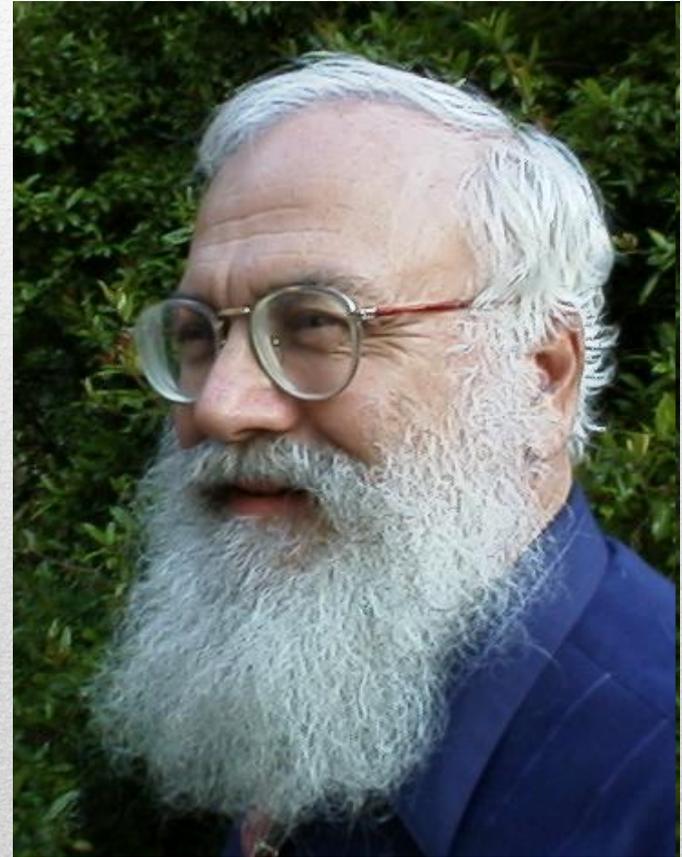
$$P_0(\omega_0, k, h) = \int dt_n A^{-1} e^{i\omega_0 t_n A} \int dy e^{-iky} p_n(t_n, y, h). \quad (12a)$$



Claerbout's principle

“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.”

(Buckheit and Donoho, 1995)



“It is a big chore for one researcher to reproduce the analysis and computational results of another [...] I discovered that this problem has a simple technological solution: illustrations (figures) in a technical document are made by **programs and command scripts that along with required data should be linked to the document itself** [...] This is hardly any extra work for the author, but it makes the document much more valuable to readers who possess the document in electronic form because they are able to track down the computations that lead to the illustrations.” (Claerbout, 1991)



MADAGASCAR
<http://www.ahay.org/>



In a Nutshell, Madagascar...

- ... has had 8,484 commits made by 61 contributors representing 485,143 lines of code
 - ... is mostly written in C with an average number of source code comments
 - ... has a well established, mature codebase maintained by a large development team with increasing year-over-year commits
 - ... took an estimated 129 years of effort (COCOMO) starting with its first commit in May, 2003 ending with its most recent commit 3 days ago
-

Languages



C

62%

TeX/LaTeX

25%

Python

6%

14 Other

7%

Lines of Code

1000k



500k

0k

2004

2005

2006

2007

2008

2009

2010

2011

2012

2013



Code



Comments



Blanks

Contributors

Tariq Alkhalifah, Vladimir Bashkardin,
Jules Browaeys, William Burnett, Cody Brown,
Yihua Cai, Maria Cameron, Lorenzo Casasanta,
Yangkang Chen, Zhonghuan Chen, Jiubing Cheng,
Joseph Dellinger, Esteban Diaz, Sergey Fomel,
Jeff Godwin, Gilles Hennenfent, Jingwei Hu,
Trevor Irons, Jim Jennings, Jun Ji, Long Jin,
Parvaneh Karimi, Roman Kazinnik, Alexander Klokov,
Siwei Li, Guochang Liu, Yang Liu, Xuxin Ma,
Doug McCowan, Henryk Modzelewski, Jack Poulson,
James Rickett, Sean Ross-Ross, Colin Russell,
Christos Saragiotis, Paul Sava, Karl Schleicher,
Jeffrey Shragge, Eduardo Filpo Silva, Xiaolei Song,
Yanadet Sripanich, Junzhe Sun, William Symes,
Ioan Vlad, Robin Weiss, Jia Yan, Lexing Ying



2011年Madagascar计算地球物理暑期学校
2011 Madagascar School on Reproducible Computational Geophysics

Reproducible Research: Lessons from Madagascar

- Reproducibility is not the goal
- The principal beneficiary is the author
- Each computation is a test
- Reproducibility requires maintenance
- Maintenance requires an open community

