# Data assimilation needs and options in earthquake modeling

Jean-Bernard H. Minster(1) and Michael M. Watkins(2)

(1) Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, Universioty of California Sand Diego, La Jolla California, 92093-0225 (e-mail jbminster@uscd.edu phone +1 619 534-5650; fax +1 619 534-2902)

(2) Jet Propulsion Laboratory, Satellite Geodesy and Geodynamics Group NASA/Caltech Jet Propulsion Laboratory, 4800 Oak Grove Drive, Mail Stop 238-600 Pasadena, CA 91109-8099 (Email: mmw@cobra.jpl.nasa.gov; phone +1 818 354-7514, fax +1 818 393-4965).

## Abstract

**The rate of seismic, geodetic, and remote sensing data accumulation has increased enormously in recent years. At the same time, our modeling capability has grown as fast as Moore's law of growth in computing capacity. A complete description of the earthquake phenomenon requires us to deal with many orders of magnitude in spatial and temporal scales, and with departures from linear behavior of the system. Under these conditions, the well-established methods for the inversion of geophysical data may no longer be practical or satisfactory. We propose to animate a discussion of alternate approaches, ranging from global optimization methods such as *evolutionary strategies* to *data assimilation* methods that make use of advanced compiler techniques.**

## Objective

Our objective is to launch and stimulate discussion of how best to accommodate very large data sets in order to improve parametric models of the various stages of the earthquake cycle so as to satisfy these observations. This discussion should examine a range of possibilities drawn from other fields, and assess their potential for improving the match between numerical simulations, field observations and laboratory data.

## Methods

The inversion of geophysical data is a problem that enjoys the fruit of several decades of theoretical research. Inverse methods of various are now well described in several texts. The inverse method has been a spectacular success which has spilt over in a number of other disciplines.

Nevertheless, the enormous increase in the rate of data acquisition over the past couple of decades, and the simultaneous increase in complexity of the models has made data inversion an increasingly challenging computational problem. Thus, adapting a model to incoming data in such a way that the model "keeps up" with the new data can be a daunting challenge, and will certainly be as massive volumes of remotely sensed observations become available. In addition, for complex nonlinear models, the ubiquitous issue of local minima of the cost function to be minimized becomes ever more difficult to manage. Many well established methods will probably not be up to the challenge, especially in cases where data do not depend on model parameters in a differentiable fashion.

On the other hand, with the advent of powerful computational tools, a number of new techniques have become practical. They include global optimization techniques which range from Genetic Algorithms to a rather rich variety of Evolutionary Programming approaches, often coupled to simulated annealing schedules to avoid the trap of local minima. These methods, known as *Evolutionary Strategies* draw their heritage from Monte Carlo techniques, and depend critically on the ability to compute forward problems quickly and repeatedly. Luckily, they are "embarrassingly" parallel, and extremely well suited to any number of computing environments involving a large number of processors.

At the same time, climate modelers and oceanographers have developed a suite of techniques to accomplish *Data Assimilation* whereby dynamic models are continuously adjusted to accommodate new observations. These techniques are often formulated in terms of an *Adjoint Problem*, which is solved by rather standard techniques. For models which involve a very large phase space, however, and for complicated cost functions (which may involve hundreds of terms), the effort of deriving and coding the adjoint problem could be an insurmountable practical obstacle. New automatic methods, based on advanced compiler techniques, can produce the adjoint code from the forward modeling code with quite attractive efficiency. They offer a solution which has gained considerable credibility in recent years, and is now being tested and verified in several applications.

We suggest that the ACES workshop should include an informal discussion of the various options which might be explored to improve earthquake models in response to the accumulation of new data. The discussions should attempt to identify which methods appear to be applicable to different problems, and formulate pilot projects to validate the selected approaches.