Summary of Session 4.1:
Inversion and assimilation of geodetic data

B. Minster(1) and T. Sagiya(2)

(1) Institute of Geophysics and Planetary Physics (IGPP), Scripps Institute of Oceanography, University of California at San Diego, La Jolla, CA 92093-0225, USA (e-mail: jbminster@ucsd.edu; phone: 619-534-5650; fax: 619-534-2902).
(2) Crustal Dynamics Laboratory, Geographical Survey Institute of Japan, 1 Kitasato, Tsukuba, Ibaraki 305-0811, Japan (e-mail: sagiya@gsi-mc.go.jp; phone: +81-298-64-6939; fax: +81-298-64-2655).

At Noosa Heads (Queensland, Australia) inaugural meeting of the ACES international cooperation on earthquake simulations, Working Group 4.1 was charged with discussion of the issue of data inversion and data assimilation. Very quickly, the discussions reduced their focus on two main topics.

1. The essential need for efficient, large scale database management tools
2. The need for a gradual evolution from the well developed concept of geophysical data inversion to the emerging one off data assimilation, for which no generally accepted model exists in the domain of earthquake modeling...

Database needs and requirements

The requirements for the implementation of advanced databases in Earthquake simulations

The first general issue discussed by working Group 4.1 pertained to the need to organize observations and model results in such a way that they can be managed effectively. Of special importance was the recognition that large-scale numerical models used for earthquake simulations will require computer-compatible databases both for inputs and outputs. Thus for any simulation there will be a need to document and manage:

1. *Input databases.*
   Examples include:

   - Earth structure (especially geometry),
   - Geology (rock types, faults),
   - Earth properties used in the simulation (rheology, friction)

2. *Output databases:*

   Upon completion of large-scale simulations, results of interest will be available in the form of very large output files, which may be used in toto or in part by the end-users. Browsing these output files will be an important function, until such time as the simulation in question is superseded by a better or more relevant one.

   For each simulation, the input database should be unambiguously identifiable and fully documented, as well as permanently attached to the simulation. This is in order that one can replicate the simulation at a future time, as new technologies or new numerical techniques become implemented. It was duly noted that both input and output databases should not be considered as "static" information. They will evolve with time, will be augmented and improved, and the mechanisms for managing and maintaining them should be able to accommodate such evolution. There remains therefore an unresolved issue of management and control. This is essential if the use of earthquake simulations is to achieve a transition from a "data explanation" role to "system evolution forecasting" role. Such transition was recognized unanimously by the group as a worthy long-term goal.

Major database questions to be addressed

In this context, Working Group 4.1 identified the following questions for which answers (or at least better answers!) should be sought:

Q1: What information should be incorporated in the databases?

- Geometry, rheology, fault properties..
- Observations (geodetic, seismic, other...)
- Interpretations (intermediate models?)

Q2: Who should be placed in charge of maintenance and control of the databases?

Q3: Is the database design capable of lasting 10-20 years in the face of rapid evolution of both hardware and software environment?

Q4: Are existing database management tools offered commercially or through government agencies adequate for the task?

Recommendation

In addition, the Working Group suggested that the following recommendation be considered by the ACES workshop participants:

Recommendation: In the same way that the global seismology research community is using a Preliminary Reference Earth Model, PREM, for any region in which we plan to use earthquake simulations to explain observations and ultimately to forecast future behavior, researchers should agree on a REGIONAL REFERENCE EARTH MODEL to serve as a basis for comparison of different simulations.

Simulation, data inversion and data assimilation

The second general topic of discussion pertained to the incorporation of fundamental geophysical observations in the overall process of developing earthquake models. The discussion included two major facets:

1. The inversion of geophysical data (in the "classical" sense) aimed at deriving or improving models after a data set has been collected,
2. The so-called "assimilation" of incoming data within models that are by nature time-dependent, such as models which forecast the future evolution of an earth system.

Combination of geodetic data

The Working Group discussed at some length the issue of inverting geodetic data. Such data are acquired continuously, but at a rate such that repeated inversions at selected epoch times is a very attractive approach. In particular, the Working Group noted that temporally-continuous geodetic data (such as those collected by a GPS network) are highly complementary to spatially-continuous but temporally scarce data (such as those collected by Synthetic Aperture Radar satellites, especially in interferometric -InSAR- mode). Already, in recent years, the inversion of dense, continuous, three-component geodetic observations has led to the identification of post-seismic deformation compatible with cumulative aftershock deformation, but up to 10 times larger in amplitude, and compatible with afterslip (Analyses of the Landers, Northridge, and Sanriku earthquakes). This appears to indicate that both effects (aftershock sequences and afterslip) respond to the same stress field.

Data assimilation

On the other hand, the Working Group concurred that in the long-term, data assimilation -within a model that evolves with time- is as important to earthquake simulations as it is nowadays to weather forecasting...but there was very considerable discussion about the following question: how do we do data assimilation in the case of earthquakes? A major presentation of "Bayesian" inversion of leveling data (using the Akaike method) in the Nankai Trough-Shikoku area showed how one can achieve refinement of an inversion result by taking account of the rheology distribution. The outcome of such inversions is quite appropriate for insertion in simulations. Nevertheless, ultimately, our goal should be to achieve better forecasting through a physical model, improved over time by data assimilation, instead of relying on mere phenomenology, which is a common characteristic of many earthquake forecasting schemas at the present time.

The natural comparison with the weather forecasting problem was explored at considerable length by the Working Group. In the case of earthquake observations, the relevant data acquisition rate is much slower, which gives researchers the luxury of using a deliberate approach, and to conduct multiple simulations, instead of having to use computationally efficient, but physically and numerically questionable approximations in the simulations. The obvious issue in such a comparison is that earthquake physicists do not enjoy a commonly accepted model for the evolution equations that govern the system (e.g. the Navier-Stokes field equations for fluid systems used in climate and weather research). The Working Group concluded that such a gap should be an insurmountable deterrent to the concept of data assimilation, and that initial steps should be taken now, instead of later. Nevertheless, a recurring question was left without answer: How do we assimilate seismicity data (in the form of an earthquake catalog)?

Global optimization techniques

Perhaps as a side issue-but one that is enabled by the recent increase in massive computation power- the Working Group discussed the fact that unlike most problems studied to sufficient maturity in geophysical inversion to date, many problems in earthquake modeling are not easily amenable to "classical" geophysical inversion. In particular, not all geophysical observables in the earthquake modeling problem can be handled using the usual techniques, which often require continuity of data functionals, and even differentiability to sufficient order. Many observables (e.g. earthquake catalogs) are not continuous in the state variables, not differentiable with respect to these variables, and subject to complex constraints. In at least some of these cases, we may want to resort to global optimization techniques. Examples of such techniques include Genetic Algorithms (GA) and Evolutionary Strategies (ES). Such techniques have been applied successfully to the solution of geophysical problems-particularly in seismology. These methods exhibit the characteristic that they become very computationally intensive for large phase space dimension. However, it is clear that they are "embarrassingly" parallel. Besides, as we agreed, earthquake scientists are not operating under the same time pressure as weather scientists...yet!