Faster earthquake source mechanism determination on egee

E. Clévédé[1] D. Weissenbach[2]

[1]Institut de Physique du Globe de Paris, CNRS, France

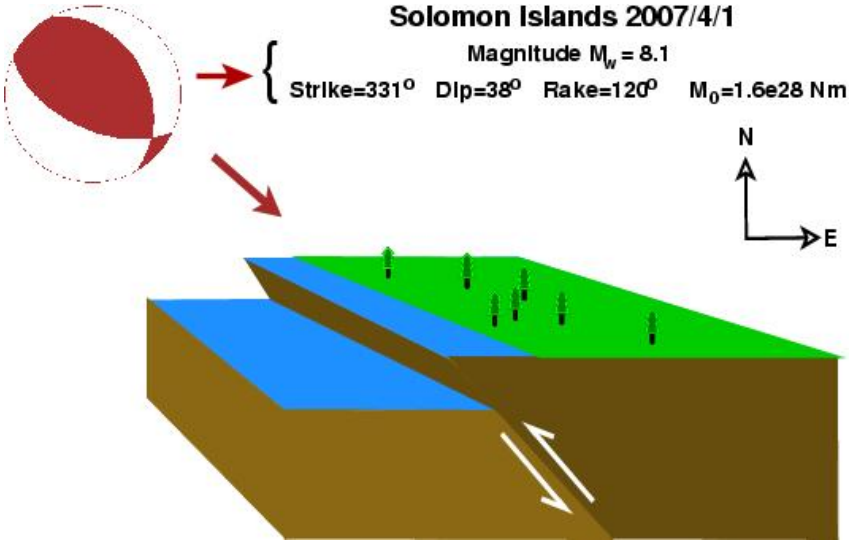
[2]Institut Pierre-Simon Laplace, CNRS, France

Sept. 2007 — Annecy

Earthquake source mechanism software

A good description of an earthquake is given by the Seismic Moment Tensor at the centroid.

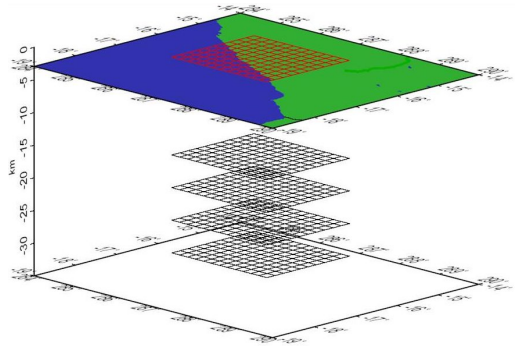
Simulations placing the centroid at several locations and calculating the resulting seismic moment tensor are compared to the real seismic data generated by the earthquake.



The best solution for the location is determined according to the data fit.

Specifications:

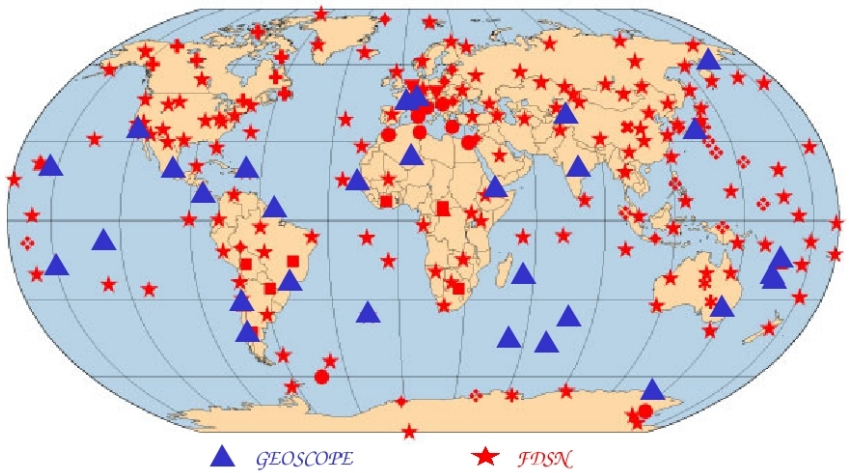
- Non linear inversion on a 3D. spatial grid, each **location** being **independent** from the others. The same set of scripts and binary executables (compiled from Fortran) is used at each position.
- Static/dynamic data and soft weight 16MB/1.5MB. Output is only tens of kB.
- Overall computation time ranges from **300 to 600 hours** and may need preliminary calibration run(s) on a 5% sample.



Challenge: **Completion a few hours after the occurrence of the earthquake**, including data pre– and post–processing.

Geoscope Stations

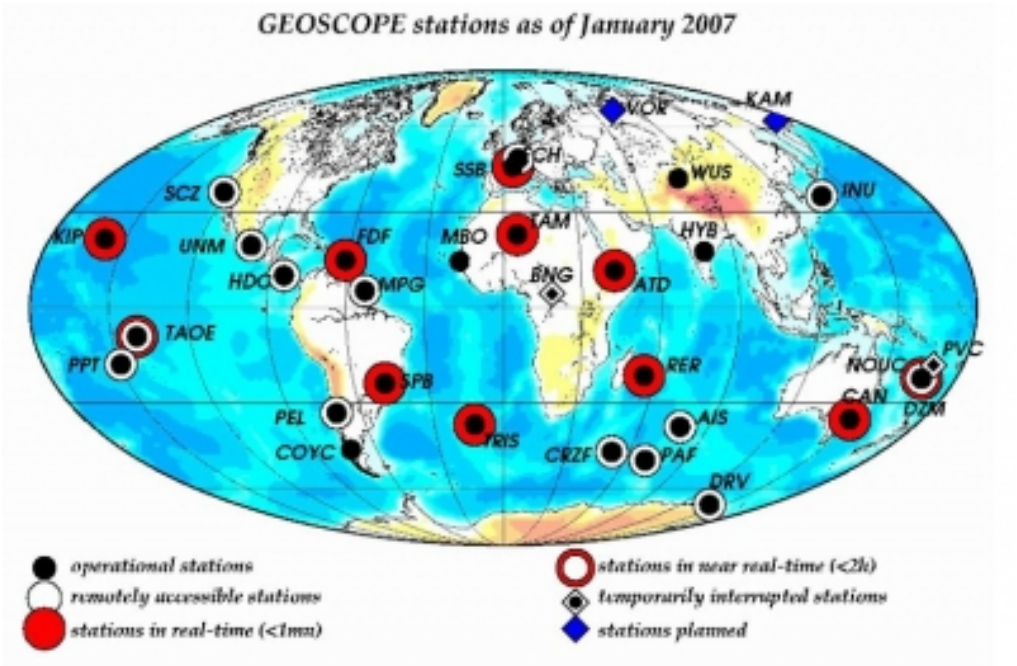
Geoscope is a worldwide network of 28 seismic stations involving several partners: IPGP (CNRS-FR), EOST (CNRS-FR), CEA (FR), CNES (FR), IRD (FR), USGS (US), ...
Member of the Federation of Digital Seismographic Networks
(33 countries, 40 networks) with the lowest noise level.



Collaborations with transnational initiatives: ORFEUS and other Seismic Data Centers, NERIES, CTBTO.

Data access

Some GEOSCOPE data are available in near real-time from the Data Center in Paris.
The whole data set (25 years) can be queried through the NetDC interface: request via `e-mail`, access via `anonymous ftp`.
1 day of SEED data for all stations weights 200 Mb.



Data for interesting seismic events also gathered and made available as a whole 1 to 2 days after the event itself.
Not all available data are used by the software. Selection is handmade.

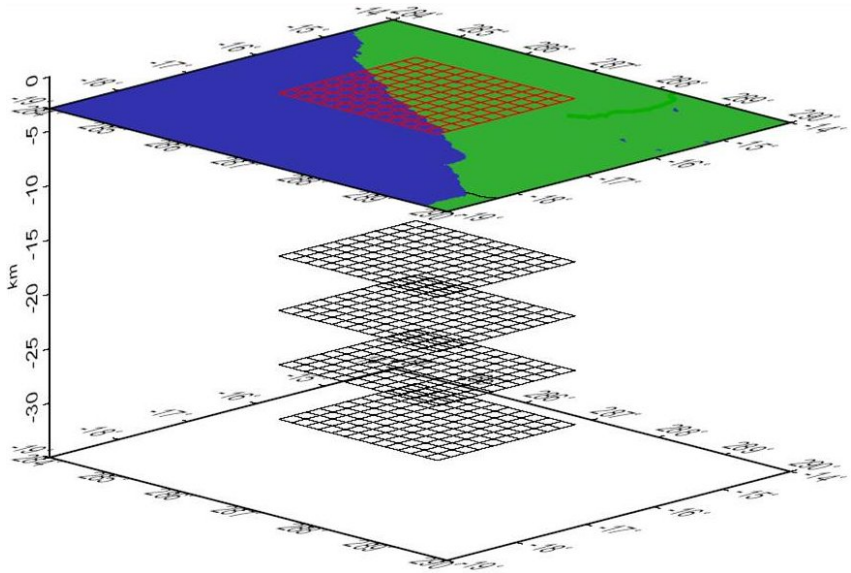
“Naive” egee porting

Canonical parsing:

- Assign [groups of] spatial positions (latitude, longitude, depth) to egee jobs; bound in `jdl`.

```
Executable = "eq.sh";
Arguments = "120.2 35.3 17";
```

- Static data (earth model) are registered on SEs, dynamic part (quake data) comes in the `InputSandbox`.
- RC selection left to the Resource Broker.



Unsatisfying achievements

Limitations:

fallible: inevitable `Aborted` jobs;

slow: jobs eventually queued for hours.

Overcome:
Frequent monitoring of the large number of jobs, "redirecting" blocked and aborted jobs to sites that already executed previous ones.

Results:
Depending on the frequency of human interventions and the size of the earthquake, it used to last 6 to 36 hours untill completion.

Other optimisation issues:
Not only a single job could be submitted maybe 3 or 4 times before it is computed,
but also similar jobs are successively executed on the same job slot.

Needed enhancements

Evolving parameters:

- New middleware: `LFC, gLite ...`
- More RCs and CPUs available to the VO.

General idea:
Defer position assignment: use a centralised **server** to dynamically distribute positions to compute to **running** jobs only.
The core of the job is a **client** requesting a position to compute and wrapping the execution of the software instances.

Requirements:

- Secured connections with fully identified server and client.
- Availability to run several instances at once (e.g. calibration runs).
- Possibility to compute a subset of the spatial positions.

Implementation

Summary of an execution:

- **User** launches a `SSL` **server requesting client certificate** -the proxy authenticating the client job- on a certified host (usually the UI), and submits the **client jobs**.
- **Clients** contact the server when **ready to compute** a [new] position or when **leaving** the pool of running clients.
- **Server** keeps track of the clients, assigns them positions and also makes sure that **all** calculations are **completed**.

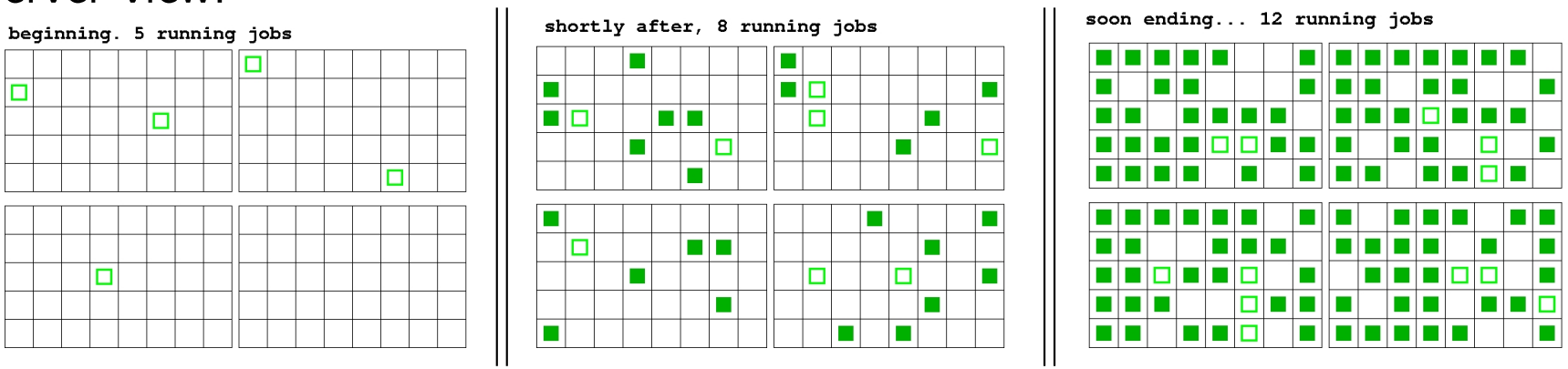
Details:

- Server and client software are written in Python.
- Server config files:

  one describing the quake's geographical location and the positions;
  one describing the CEs to send jobs to and in which amount (the templates need regular updates).
- The server generates the `jdl` files (`parametric jobs` can be generated if a `gLite-WMS` is used). Running it on a UI provides additional facilities.

execution frame

Server view:



Client view:

1. get needed files from SE (data + soft)
2. contact server to get a position to compute
3. computation (former job's duty) + output storage
4. if enough CPU and Wall time remaining (depends on RC/queue) go to step **2.**
5. notify server of job ending

Performances

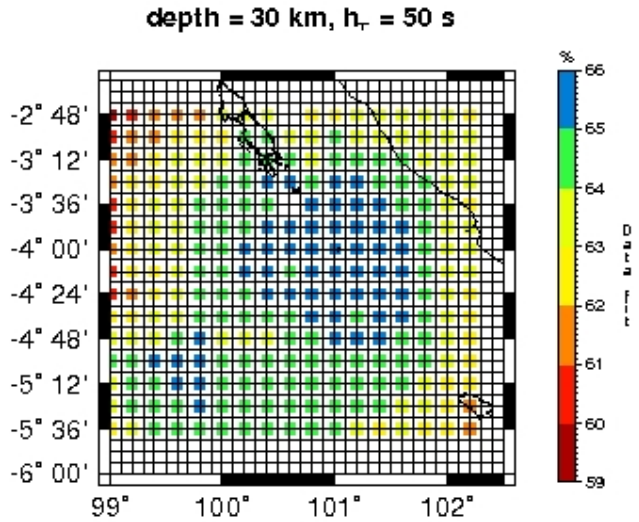
The largest problem (600h, 1386 locations), hardly solved with the first deployment scheme, has been redone in 6 hours by 140 jobs on a total of 230 clients submitted.

Other new features:
The conjuction of more storage availbale to the VO and the `LFC` made it possible to save intermediate results (*synthetic seismograms*) from the first run, now available for other runs concerninq the same earthquake or its replicas, and even other earthquakeas in the same area, sparing up to 3/4 of the computing time.
It is now possible to run very wide simultations (4000 cells) overnight for huge earthquakes, and refine the solution on next day.

Achievements

Improved suitability:

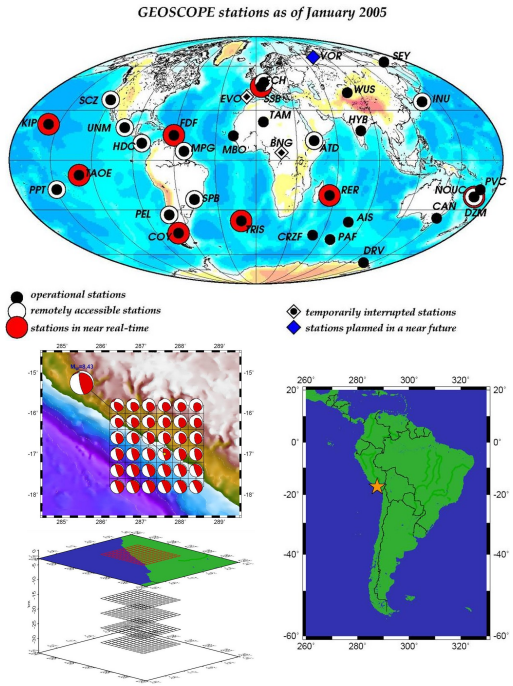
- User side: Faster start, assessable completion time; no further action required but can still submit additional client jobs during the process.
- M/W side: No more monitoring/cancelling/resubmitting operations. Fewer data transfers (same data used with several positions).



- Scientific added value: A software that was only used with a few more carefully studied earthquakes is now intensively running at almost every occurence of a major earthquake.

Credits

- Contacts:
  - GEOSCOPE:
    http://geoscope.ipgp.jussieu.fr
  - Eric Clévédé:
    clevede@ipgp.jussieu.fr
  - David Weissenbach:
    weissenb@ccr.jussieu.fr
- Many Thanks to:
  - Sophie Nicoud — GRID-FR CA
  - Antun Balaz — SEEGRID
  - Geneviève Patau — IPGP



Clévédé *et al.*

Thank you for your attention