Grid Activity in Earth Science

M. Petitdidier (IPSL),

Monique.petitdidier@cetp.ipsl.fr

H. Schwichtenberg (FhG/SCAI)

Horst.schwichtenberg@fhg.scai.de

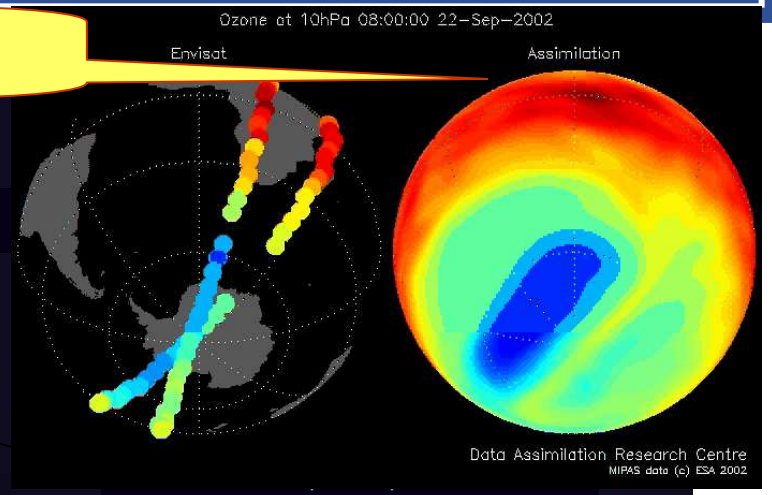
In collaboration with EGEE and DEGREE EU-project partners

www.eu-egee.org

Information Society and Media

Planet Earth : a complex system

GOME total ozone assimilation



Stratospheric Ozone

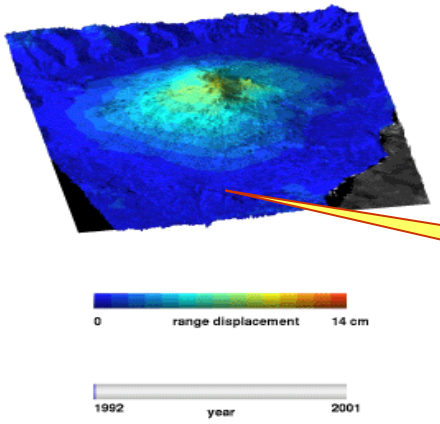
Industrial Emissions

Topography & Motion

Land cover & vegetation

Marine SST, SSH& colour

Currents, bathymty & ice



10 y displacement of Etna 1992-01

Why GRID for Earth Science?

- **GRID infrastructure**
  - an “open platform” for handling computing resources, data, tools…
    - § Partner can use a lot more resources than the ones he (she) brings in
  - Impressive number of shared resources
    - § EGEEII around 60,000 CPUS distributed in 250 sites
    - § 20 PB storage
  - A collaborative possible platform among teams and/or countries
    - § interactive collaboration to avoid effort duplication
  - Secure and restricted access to resources, data, tools…
    - § Same data and software policy as outside Grid
  - Grid will open new fields of investigation
  - on Earth Science

Why Grid for Earth Science?

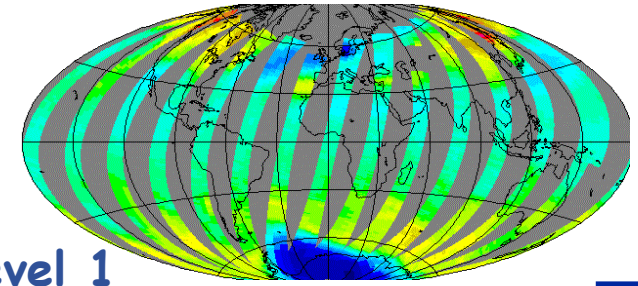
- **Intensive computing**
  - Massive parallel jobs like Climate models, simulation of an earthquake in Los Angeles….=> HPC
  - Large number of CPUs mostly independent at least MPI jobs with few CPUs like data set exploitation, Statistical approach, Monte Carlo, parametric studies, job on alert…
- **Data and/or algorithm sharing**
  - Same date set used by several teams
  - Algorithm or software deployed on Grid and used for different purposes
- **Trusted Resources**
  - Security
  - Confidentiality

EGEEII – ES VO

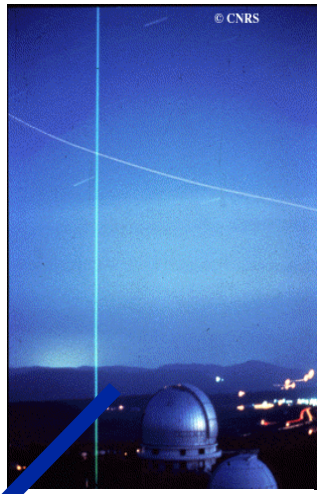
- **2 Virtual Organisations :**
  - Due to different data policy for Academic and private research
- **VO ESR (Earth Science research)**
  - As an average around 50 – not always the same persons
  - Mainly scientists, few tests carried out by Company
  - Bulgaria, Finland, France, Germany, Greece, Italy, Russia, Slovakia,Spain, Switzerland, The Netherlands
  - Around 2000CPUs/day for ESR
- **VO EGEODE (**Expanding GEOsciences on DEmand**)**
  - centered around Geocluster (software from Compagnie Générale de Géophysique) around 30 persons
  - Recently French Academic institutes already using Geocluster join the VO

Sharing data: GOME

Raw satellite data
from the GOME instrument
(~75 GB - ~5000 orbits/y)

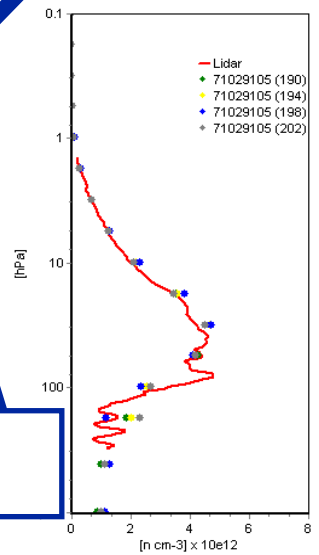


Level 1
(example of 1 day total O_3)



ESA(IT) – KNMI(NL)
Processing of raw GOME
data to ozone profiles.
2 alternative algorithms
~28000 profiles/day

IPSL(FR)
Validate some of the
GOME ozone profiles (~10^6/y)
Coincident in space and time
with
Ground-Based measurements



Level 2

Visualization
& Analyze

Meta Database
server

PosgreSQL
– geospatial
search

EGEE environment

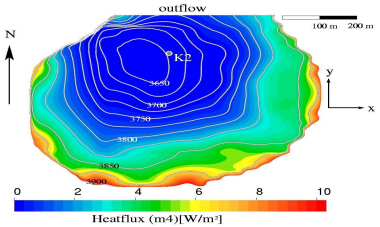
GOME

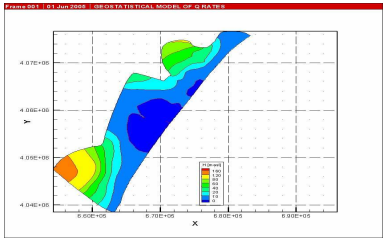
- **7 years of data, 14,5 orbits/day**
- **Algorithms:**
  - Neural network, NNO, (ESA, UTV) using IDL - 2 versions
  - Inversion Algorithm (KNMI) –data, O3 climatology, ECMWF..
- **Lidar data (NSDC)**
  - 7 stations maximum (IPSL)
- **Number of filed: 70000 for both 2 versions of NNO**
- **Common development**
  - Metadata base on a server with security and restricted access
  - Query by Geolocalisation in time and space of orbits passing over a lidar site by using PosgreSQL)
- **Results**
  - Unique case of validation of a whole satellite data set with all the data taken by other instruments
  - Once the application was ported the validation of another algorithm or version is very fast

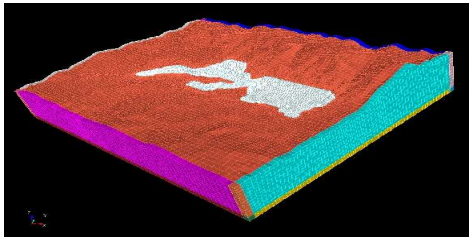
ES applications

Sharing Algorithm

GEOCLUSTER	ELMER	CODESA-3D	3DSEM_UNSTRUCT
CGG-Veritas	CSC - Finland	CRS4 -Italy	IPGP- France







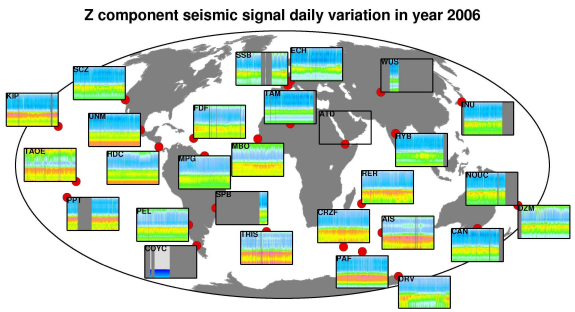
Partners:	Partners:	Partners:	Partners:
VO - EGEODE	VO- ESR	EUMEDGrid	EELA

Exploring Large Set of Data

Geoscope: (http://geoscope.ipgp.jussieu.fr) IPGP-France

- 28 seismological stations and data center
- 25 years of data available on EGEE
- Processing of the whole data set on EGEE

Impact on other seismological data Center design



Processing and/or production of large data sets

- **Risk evaluation of contamination of water resources by pesticides at different time and spatial scales.**
  - Weather scenarios (data base by Meteo France), soil scenarios and 100 pesticides

⇒ **12 millions de run 1-2h each**

⇒ **10 Toctets**

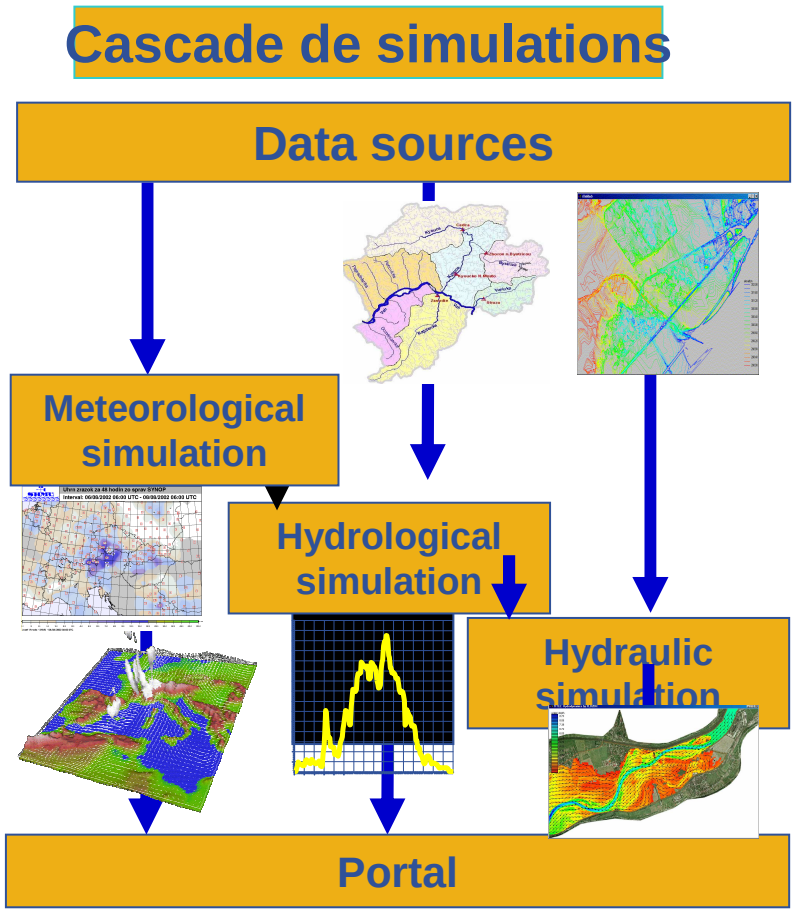
⇒ **Interest to use EGEE:**

- 24h/24h and 7 days/7
- Possibility to run simultaneously hundreds of jobs

Complex workflow: Forecasting of flood

L. Hluchy, Viet Tran, M. Ciglan …. (II-SAS, Bratislava Slovaquie)

- Danube river
- Data :meteorology, river network rivières, landscape
- Meteorology model ALADIN (MPI-parallel), MM5 (MPI-parallel)
- Hydrology HSPF (sequential-parametric), NLC (sequential-parametric),
- Hydraulic. DaveF (MPI-parallel), FESWMS (MPI-parallel)
- output: weather, precipitations, hydrography, water level and flood speed forecast
- Cascade of jobs managed by dynamical workflow



Control of a lot of jobs: Earthquake Characteristics

Fast Determination of mechanisms of important earthquakes
(IPGP: E. Clévédé, G. Patau; IPSL: D. Weissenbach)

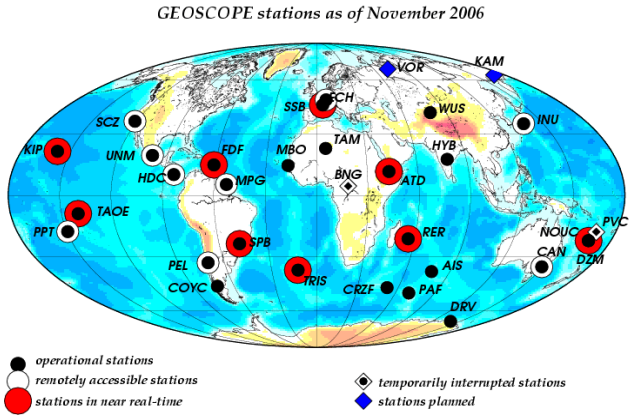
Application to run on alert

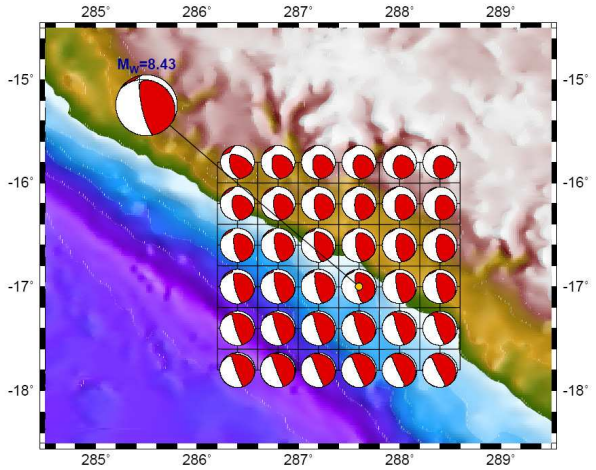
- ✓Collect data of 30 seismic stations from GEOSCOPE worldwide network
- ✓Select stations and data
- ✓Define a spatial 3D grid +time based on the assumed earthquake location
- ✓Run for each grid point or group of grid points a job => ~ 50-100jobs

Results obtained ~6hr after the earthquake

Important for emergency action and other related researches

All major earthquakes so treated: 21/24 in 2006 => catalogue





Parametric job: Geomorphology

Objective : understanding landscapes formation and evolution

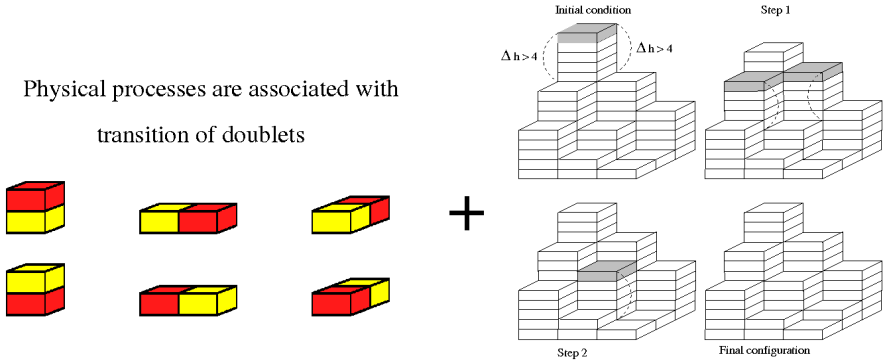
IPGP: C. Narteaux and O. Rozier

Examples : erosion of the mountains, dynamic of dunes ...

Algorithm : 3D Cellular automaton for geomorphological research very simple transition rules between cells of different states (transport, deposition ...).

Tien Shan (Chine)

Modèle



avalanche mechanism

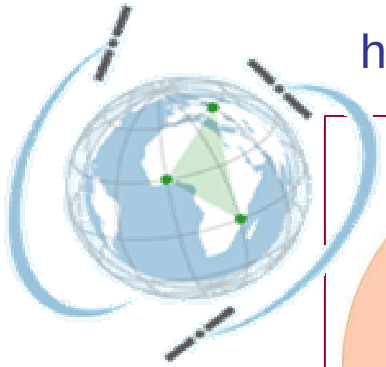
Final remarks

- **In EGEEIII improvement and/or implementation of functionalities more adapted to Earth Science**
  - License server
  - Data management
  - Workflow
  - Portal
- **New tools needed to use the whole Grid potential**
  - Due to Change in scale of computing power
  - Need of Exploration of huge data sets
  - Creation of Platform integrating web services, computing power, information systems….
- **New conceptual approach of Earth Science**
  - Role of Scientist
    - Interactive collaboration -> less duplication of development and/or adaptation
    - More time for new ideas, new research
    - Confidentiality of the research
  - Application development (access to several large data sets, more CPUs…)

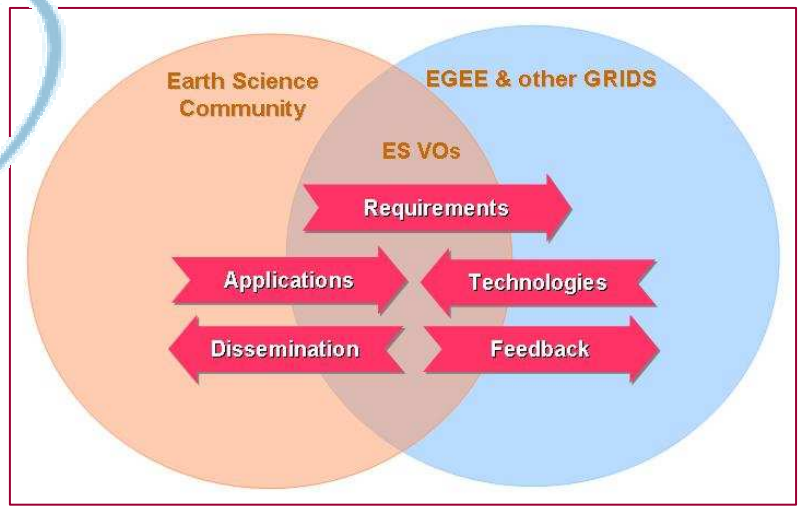
- Strategic objectives
  - **Bridge the ES and GRID communities throughout Europe**
  - **Ensure that ES requirements are satisfied in next Grid generation**
  - **Ensure the integration of emerging technologies for managing ES knowledge**

The DEGREE team:

IISAS, Slovakia (Coordinator)
CNRS, France
KNMI, The Netherlands
UNINE, Switzerland
CRS4, Italy
SCAI, Germany
GCRAS, Russia
ESA-ESRIN, Italy
CGG, France
Dutch Space, The Netherlands



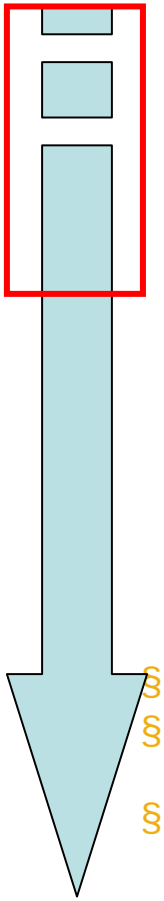
http://www.eu-degree.eu



Project Vision

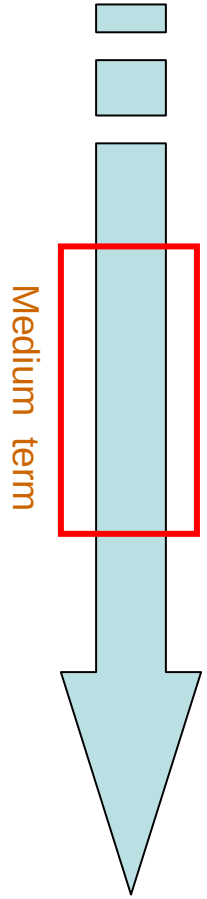
Build a bridge linking the ES and Grid communities

DEGREE: ES roadmap



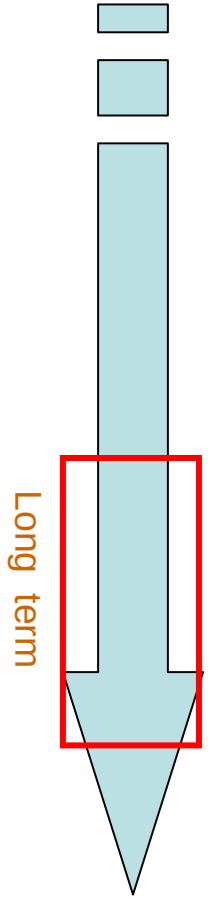
- Community building
  - *Continued support for cross-fertilization and bringing together the ES and Grid developer communities.*
  - *Establishment of a Grid board for ES at large international level*
  - *Identification of ES organizations coordination, working groups, and so on, of key strategic importance, to be kept informed and up to date on the Grid topic,*
  - *Continuation of ES Grid promotion and dissemination activities (i.e. presentations and updates of significant results using Grid).*
  - *Analysis of the ES needs, with reference to large projects (i.e. polar ozone year, International heliophysical Year and thematic projects). So far there is no direct contact with those projects.*
  - *Reference to EGEE where many applications have been ported and it is the base of the ES experience.*
- Support to education, development of Grid training programmes.
- Maximise approach to support large science community (as in G-POD) – single sign-on from all ES data providers.
- Definition and implementation of a standard approach to distributed data and metadata.

DEGREE:ES roadmap



- **Porting to the Grid of new applications**, which may benefit immediately for having access to sensors, data, archives, and so on.
- Integrate in Grid environment a lot more specific **data handling tools**.
- Deploying directly to the **Grid storage** the most used very large environmental databases and file repositories
- Facilitate access to the Grid for **everyday users**
- **Solve** outstanding **blocking issues** for ES applications deployment on Grid
- Increased development of **dedicated ES Grid tools**, based on re-use of newly available generic Grid frameworks, tools and services, based on emerging (WSRF) standards (e.g. OMII).
- Grid should integrate all the features needed for **e-Collaboration** among ES (science and service industry).

DEGREE: ES roadmap



- Enhancement of **ES Grid platform** with additional next-generation dedicated application level tools and services:
  - *ES Grid ontology;*
  - *Grid-based geospatial data handling services;*
  - *Grid metadata management services;*
  - *Grid and WebService programming languages and environments;*
  - *Grid workflow composition and execution using component-based models, search, discovery, access and utilization of SOA-based dedicated ES and non-ES (generic) services;*
  - *Access to large repositories and data holders*
  - *Link to ESFRI Environmental dedicated Research Infrastructures*
  - *Grid services to support geographic data mash-ups and visualization using GoogleMaps/Earth, MS Virtual Earth, and NASA World Wind platforms*

TNA 4.2.3/Goals

Workplan in discussion for:

- **Specific Support**
  - providing and supporting core functionalities (generic tools)
    - § esp. for ES workflow and data models/formats
      - *Workflow tools with automatic annotation have been developed for Flood applications by an ES Cluster partner*
    - § for integration of geographical information systems (GIS)
    - § for access to external web service toolkits like Google maps, MS Virtual Earth
    - § in the context of ES grid-jobs (easy executing, QoS, advanced reservation, steering)

      (different agent based solutions presented on the last EGEE User meeting)
  - direct ES User- and ES Application support (for existing and for new Users)

TNA 4.2.3/Goals

- **Evolution of gLite**
  - Webservices are needed by ES applications

    We will provide Webservice-Standard based interfaces

    esp. for data access and to use gLite grid services
    - § e.c. for OpenGIS as defined by OGC
    - § for OPeNDAP (community standard to access remote data (netcdf, HDF))

    Vision "Service platform for ES applications" (like GEON)

    Example:

    Earth observation Grid Processing on Demand (G-POD) is based on Webservice interfaces
    - § tools from gLite included

    ESA Esrin developed the framework

    The cluster will share technologies with ESA ESRIN and the G-POD and the GENESI-DR project

Earth Science Expectations

- **Pushing frontiers of scientific discovery by exploiting advanced computational methods.**