Earth Science Activity

David Weissenbach (IPSL)
weissenb@ccr.jussieu.fr

EGEE is funded by the European Union under contract IST-2003-508833

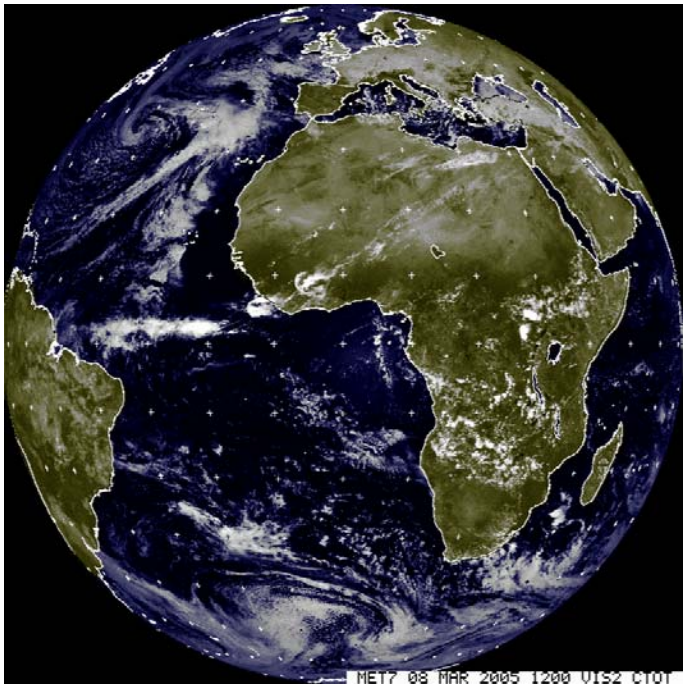
EARTH SCIENCE

The Earth is a complex system, centre of a large variety of phenomena.
Its description is separated in various independent domains with interfaces between them

Solid Earth Ocean Atmosphere

The approach could be based on physics, chemistry and/or biology

➔ The community is constituted by many small groups, that aggregate for projects (and separate afterwards).



Earth Science Applications

- Earth Observations by Satellite

Ozone: ESA(IT), KNMI(NL), IPSL(FR), UTV(IT)

SAR: IGUASSU sotfware system(CZ), ESA(IT)

- Seismology:

Earth Quakes: IPGP(FR)

SPECFEM3D (MPI): IPGP(FR), MIGP(FR)

- Hydrology

Swimed: CRS4(IT), UNINE(CH), INAT(TU), IMFT(FR)

Flood: IISAS(SK)

- Climate: DKRZ(DE)
- Geosciences: CGG(FR)
- Pollution: BAS(BG)
- Meteorology, Space Weather: GCAS(RU)
- Mars Atmosphere: IPSL/CETP(FR)
- Database Collection: NTUA(GR)

EARTH SCIENCE ORGANISATION

- Coordination Earth Science: Monique Petitdidier (IPSL)
  - Earth Science Research :

    M. Petitdidier and W. Som de Cerff (KNMI)
  - Expanding Geosciences on Demand:

    D. Thomas, G. Youinou (CGG)
- 12 Laboratories and research Institutes ; 1 organisation (ESRIN) ; three private companies (CGG, DutchSpace and Iguassu Software System)
- In Bulgaria, France, Germany, Greece, Italy, Netherlands, Russia, Switzerland, Slovaquia, Czech Republic

VIRTUAL ORGANISATION (VO)

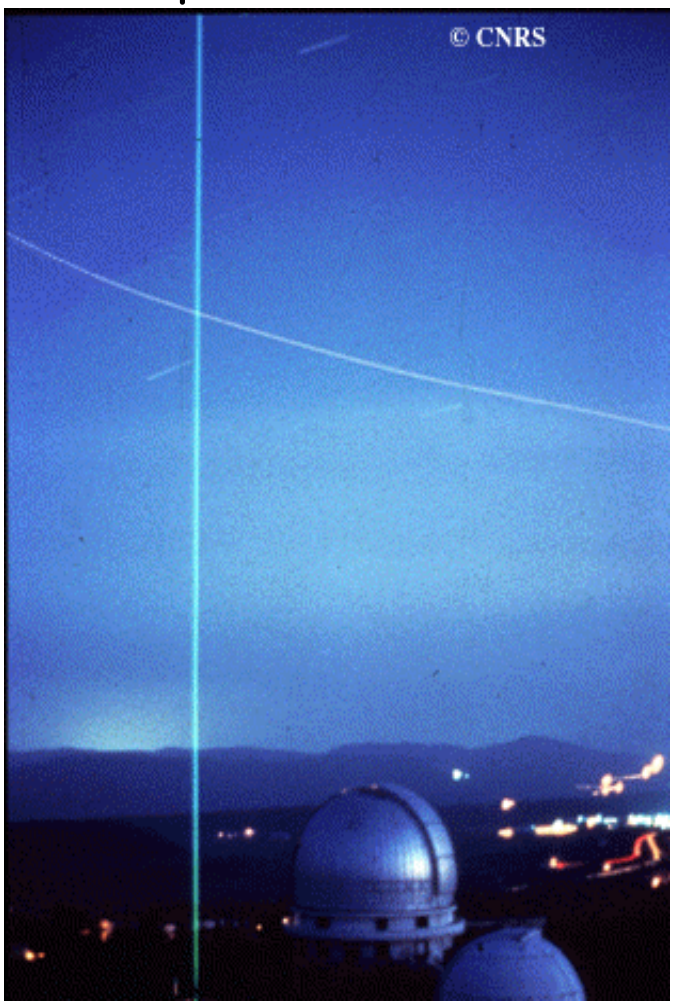
- 2 Virtual Organisations : Due to different data policy for Academic and private research
- ESR (Earth Science Research) All ES research activities – 35 persons (1BG, 1CH, 1CZ, 2DE, 10FR, 1GR, 6IT, 3NL, 4Ru, 6 SK)
- EGEODE (Expanding Geosciences On Demand): centered around Geocluster (software from Compagnie Générale de Géophysique) around 5-10 persons

- ESR has a Wiki web site http://datagrid.nadc.nl/twiki/bin/view/ESR/WebHome
- EGEODE: web site http://www.egeode.org

EARTH OBSERVATION BY SATELLITE [1]

Typical satellite case: large number of files

7 years of Ozone from GOME/ERS2 experiment



CHALLENGE

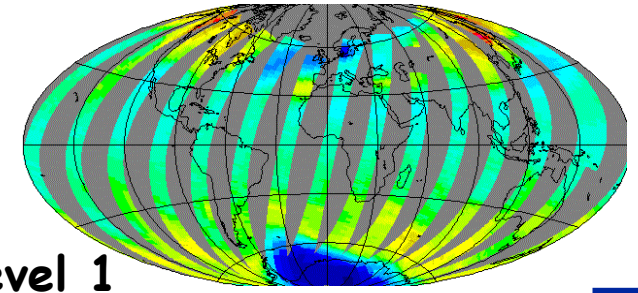
PRODUCTION: Complex Algorithms

(neural algorithm and inversion algorithm)

VALIDATION: For a given day find 10 profiles among 26000.

The DataGrid Ozone processing and validation test-case [2]

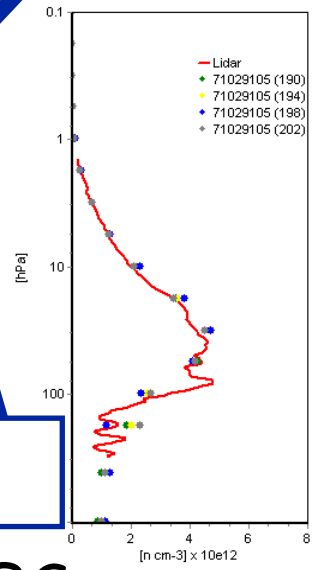
Raw satellite data
from the GOME instrument
(~75 GB - ~5000 orbits/y)



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

EGEE environment

Visualization & Analyze

Additional GRID demonstration: GOMOS, OMI, CEOS, …

OBSERVATION DE LA TERRE PAR SATELLITE [3]

Production of Ozone profiles from satellite data GOME/ESR by UTV (Italie), ESRIN (Italie), KNMI(Hollande)

Validation of the profiles with lidar data by IPSL, UTV

NNO: Algorithm « Neural network »

2 versions (F. del Frate, M. Iapaolo, S. Casadio)

8 years (1995-2003) on EGEE:

38489 files/algorithm with ~ 1800 ozone profiles/file

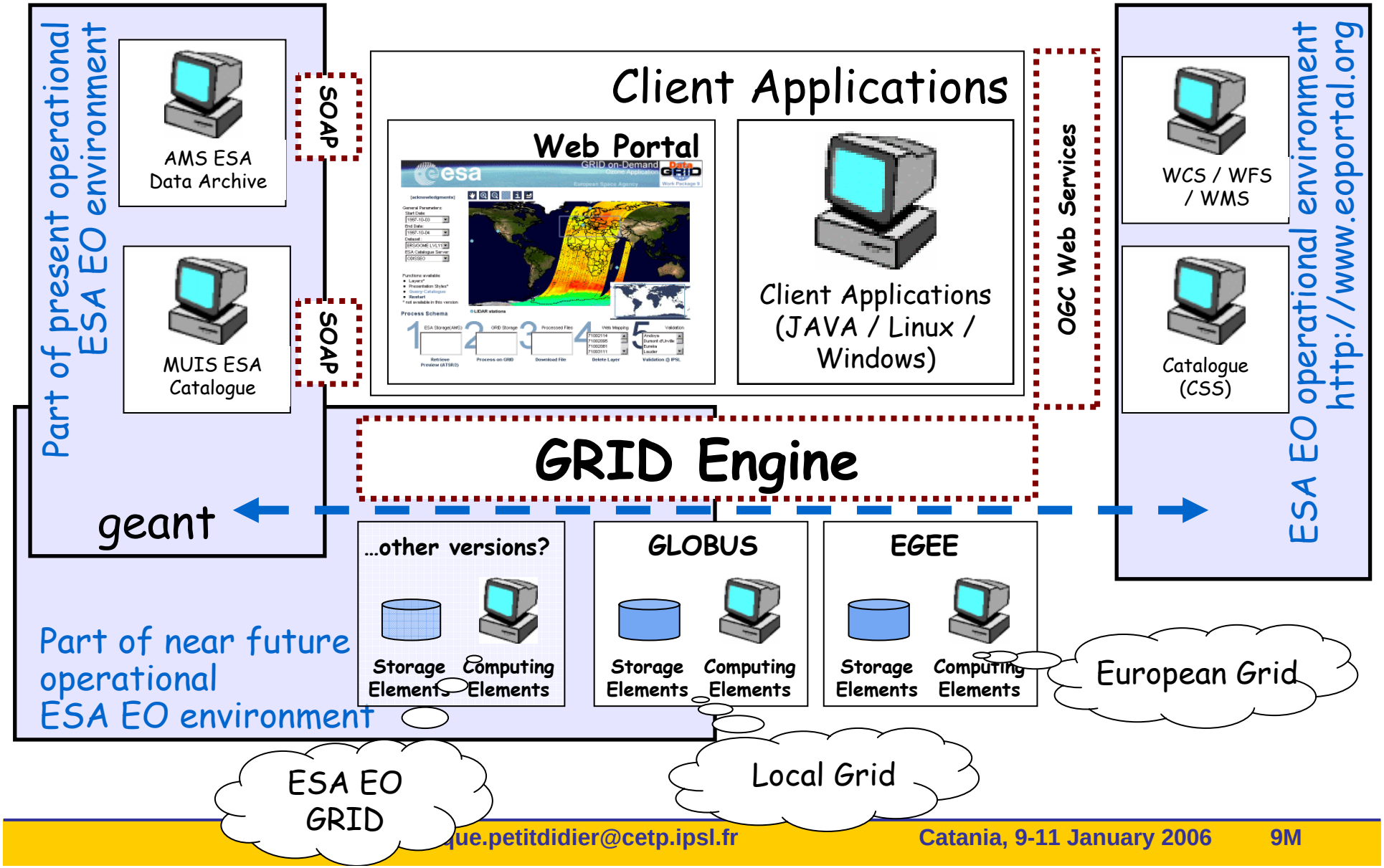
OPERA: Algorithm – Inversion (R. van derA et al., 2002)

Several months onEGEE: 76000 Files, 1 profile/file (W. Som de Cerff)

Database: metadata taking into account the orbits footprint

GRID integration with EO Web Services

luigi.fusco@esa.int

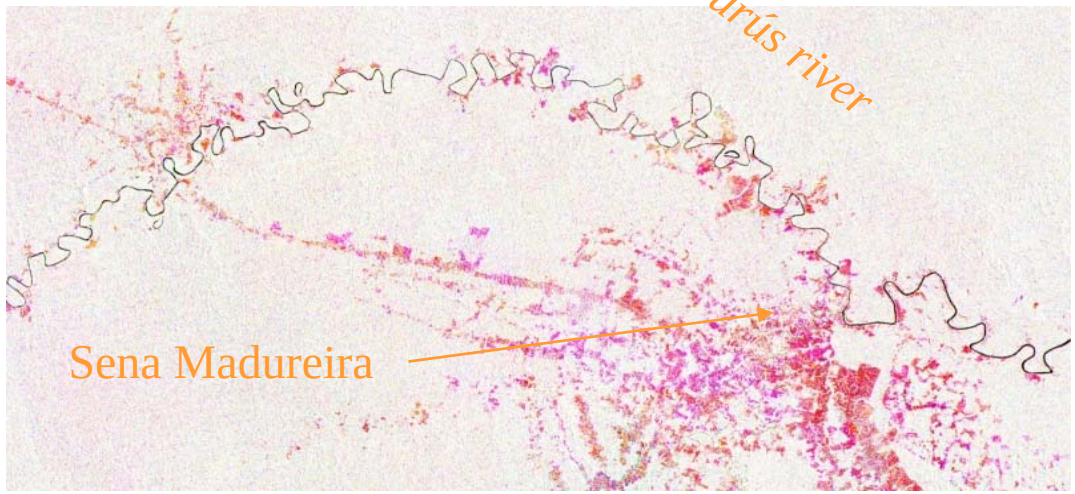


ESA PECS-GRID project [1]

martin.paces@iguassu.cz

Earth Survey through Radar Imaging

- Assessment and validation of the porting of SAR data processing SW to EGEE
- Example: Multitemporal Analysis **– automated backscatter image generation, calibration, coregistration, geocorrection and creation of RGB composites from ENVISAT ASAR products using BEST (Basic ENVISAT SAR Toolbox)**
- Early stage of the project **– only preliminary prototypes are available.**



RED October 2006

GREEN September 2006

BLUE August 2006

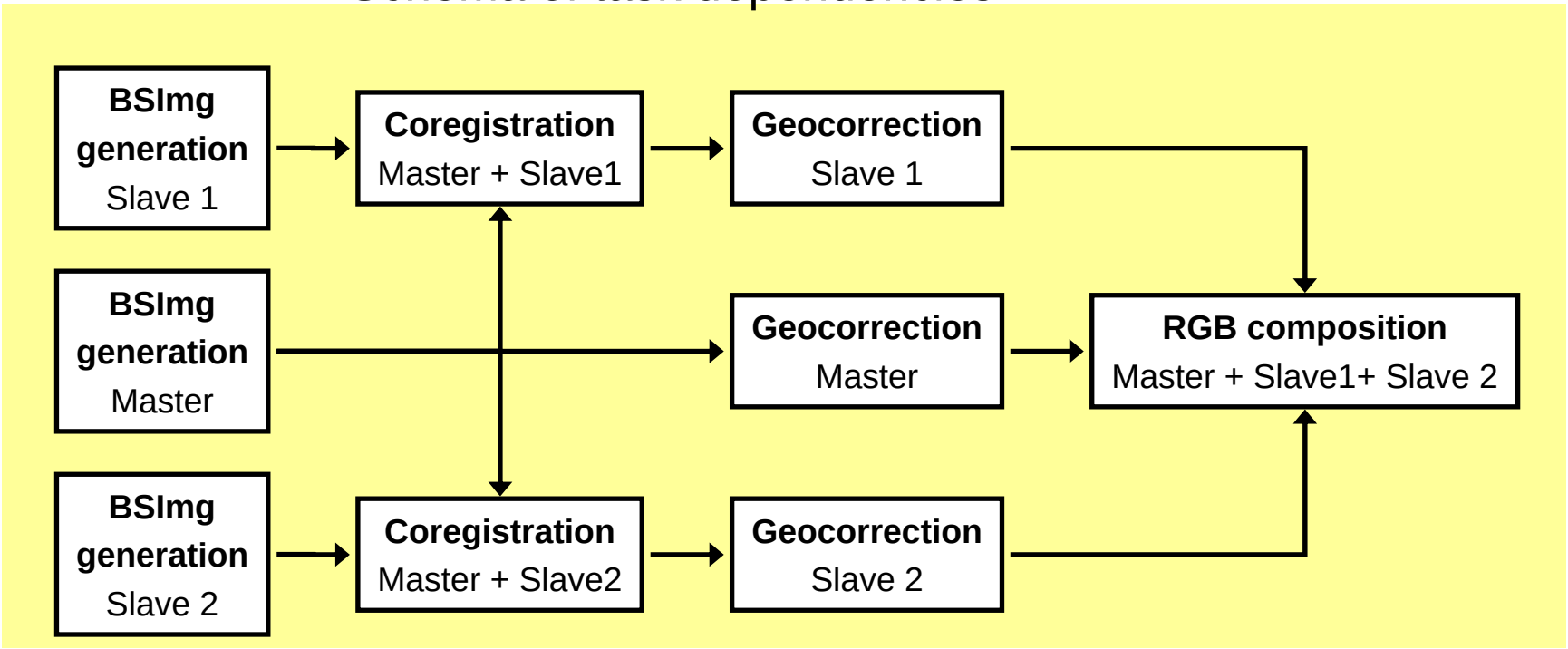
RGB composite generated from ENVISAT ASAR WSM products.

Sena Madureira and Purús river, western Brazil

ESA PECS-GRID project [2]

- Difficulty encountered **– synchronisation of concurrent tasks, notifications to UI about the state of running tasks.**
- LCG2 **– limited support for workflow management**
- **Can be solved by repeated queries from UI to logging and bookkeeping server**.

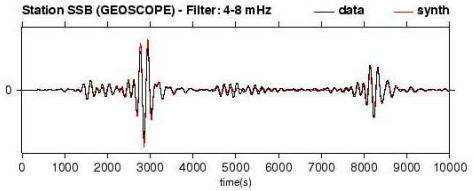
Schema of task dependencies



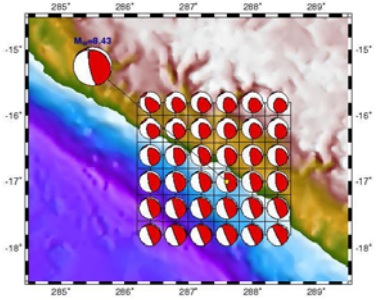
Determination of Earthquake mechanism [1]

clevede@ipgp.jussieu.fr

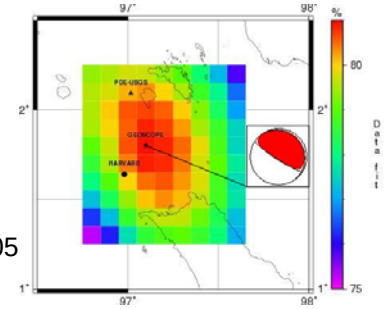
- **Rapid determination of the centroid moment tensor of large earthquakes (location and mechanism in the point source approximation)**
- **The application uses long-period seismic data (T<100s). The centroid (lat.,long.,depth and duration) is determined by a parameter space exploration, while the possible mechanisms are obtained by linear inversion for each point of the centroid parameter space. The final solution is selected through data fit criteria.**
- **To fulfill the '*fast*' aspect of the determination (i.e. within hours after earthquake occurence), requires availability of about a hundred WN for about an hour of computation.**
- **Successfully ported on EGEE by IPGP (Fr)**



Peru, June 23, 2001
Mw=8.4



Sumatra, March 28, 2005
Mw=8.5



Determination of Earthquake mechanism [2]

- **The difficulties encountered are linked with the need to handle a lot of jobs: job failure due to matching CEs that are actually not available, long time response from some CEs for status update.**
- **Shell scripts are used to create and manage the jobs from the UI: (re-)submission, output retrieval,...**
- **gLite features expected: parametric jobs (this application fits exactly in this category) and DAG (suitable for further developments scheduled for the application).**

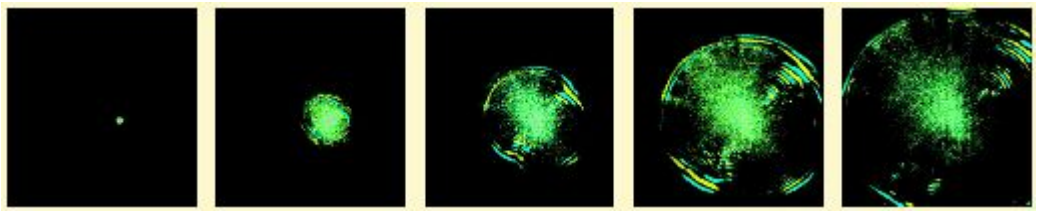
SpecFEM3D [1] (moguilny@ipgp.jussieu.fr)

- Resolution of regional scale seismic wave propagation problems to model wave propagation at high frequencies and for complex geological structures, with use of the spectral-element method.

- **Application first written by D. Komatitsch (Université de Pau), used in more than 75 laboratories in the world,especially for the study of earthquakes.**

- **Very scalable application using** F90 + MPI **needing NFS mounted homes and requiring the successive launch of two mpirun on the SAME nodes, allocated in the SAME order.**

- **Ported to EGEE by IPGP (CNRS)**



SpecFEM3D [2]

- **Difficulties encountered:**
  - LRMS's do not differentiate Nodes and CPUs.
  - MPI enabled queues contains CPU with varying power
    ➔ most powerful wait for the less powerful**.**

- **Some successful tests done with gLite 1.1.**

- **Perspectives:**
  - CPUs reservation.
  - Inter-sites MPI (MPICH-G2, MPICH-V).
  - High performance networks (Myrinet, Infiniband...).
  - MPI on RCs that have hundreds of nodes.

GRADIOMETRY [1] (pajot@ipgp.jussieu.fr)

- GOCE gravity gradiometry data inversion (PhD work).
- Infer the subsurface density distribution in the Earth from spatial derivatives of the gravity field measured by the satellite GOCE (ESA, 2006), or onboard a ship (Bell Geospace, 2005).
- Method: pseudo Monte Carlo inversion – neighbourhood algorithm.
- Needs to generate a big amount of models (density distribution – C language) almost simultaneously, to be compared with the data (FORTRAN 77).
- Porting on EGEE: starting. Forward modelling OK.

Gradiometry [2]

- Needs that the models have been generated to go on with next step. Each (group of)model(s) generation is one job.

  Would be useful if it could be checked automatically, that all those jobs are done and then the following is launched.

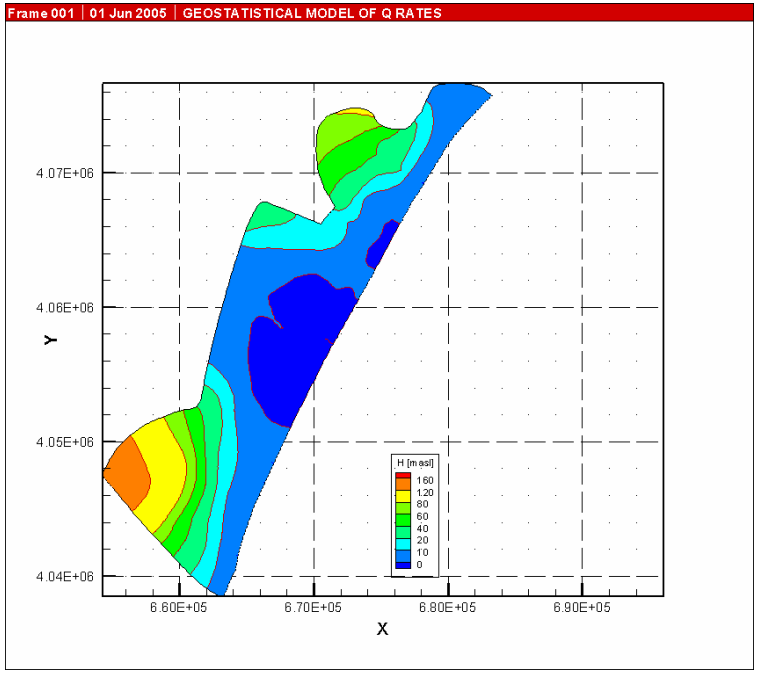
- Possible need to encrypt the data and their access for privacy reasons (if VOMS is not enough)

SWIMED [1] (giuditta@crs4.it; philippe.renard@unine.ch)

- **Goal of the application :** Modeling seawater intrusion in coastal aquifer – assessing water sustainability under uncertainty.
- **Description of the application :** 3D Finite element code solving the groundwater flow and transport partial differential equations including variable density fluids and coupled with a Monte Carlo analysis to quantify uncertainty.

Cap Bon Peninsula

Korba aquifer – eastern coast



Calculated groundwater levels for one simulation

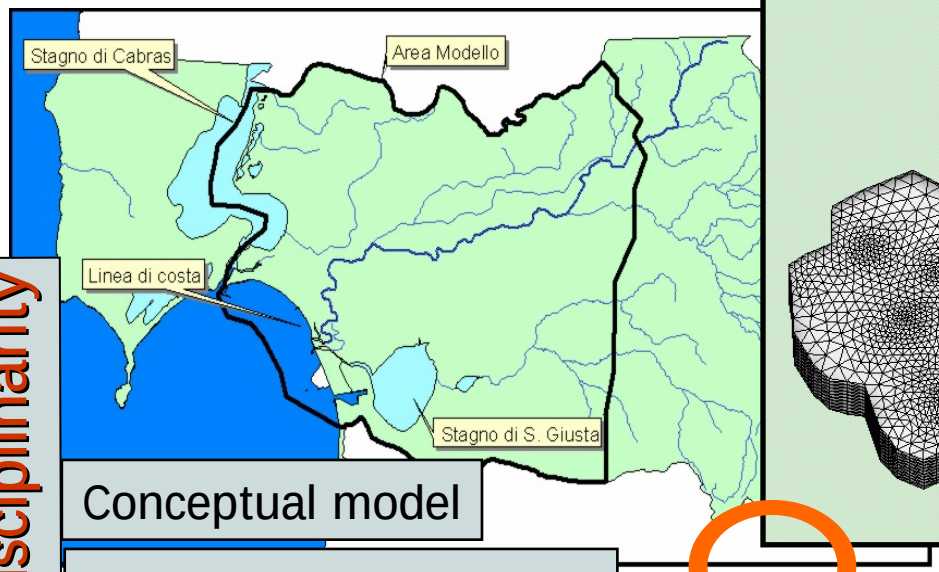
SWIMED [2]

- Specific requirements : none for CODESA3D
- Porting on EGEE: this code is running without problem
- Difficulty encountered: manpower was still an issue until november 2005 but funds have been raised by CHYN and CRS4 that will allow to increase significantly our productivity in 2006
- Open problems : licensing issues for commercial codes (FEFLOW), interactive Xwindow runs

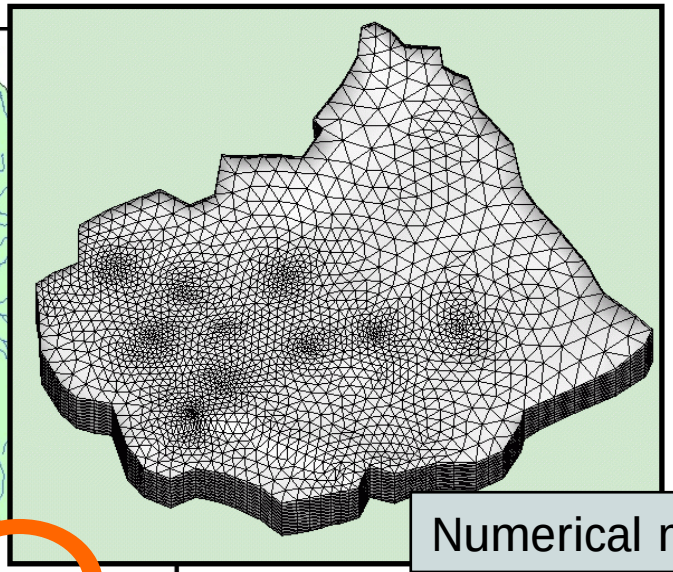
Enabling Grids for E-science in Europe

Saltwater Intrusion in Coastal Aquifers

Interdisciplinarity

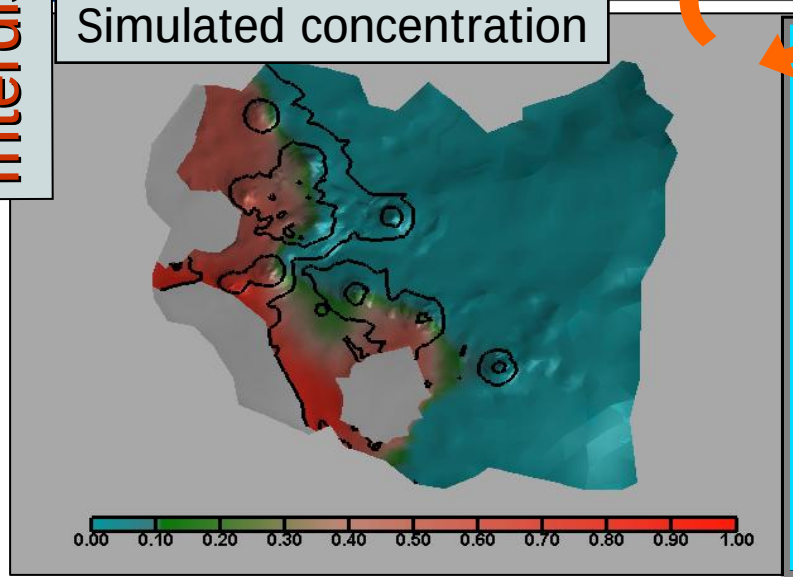


Conceptual model

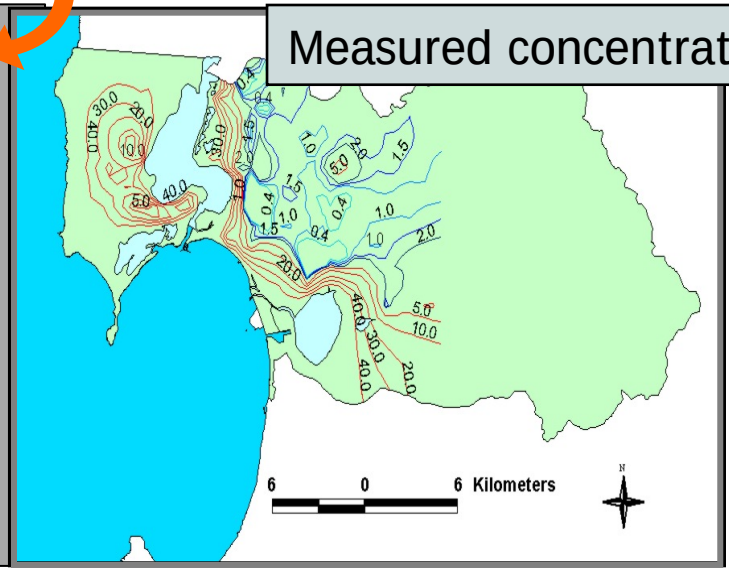


Numerical model

Simulated concentration



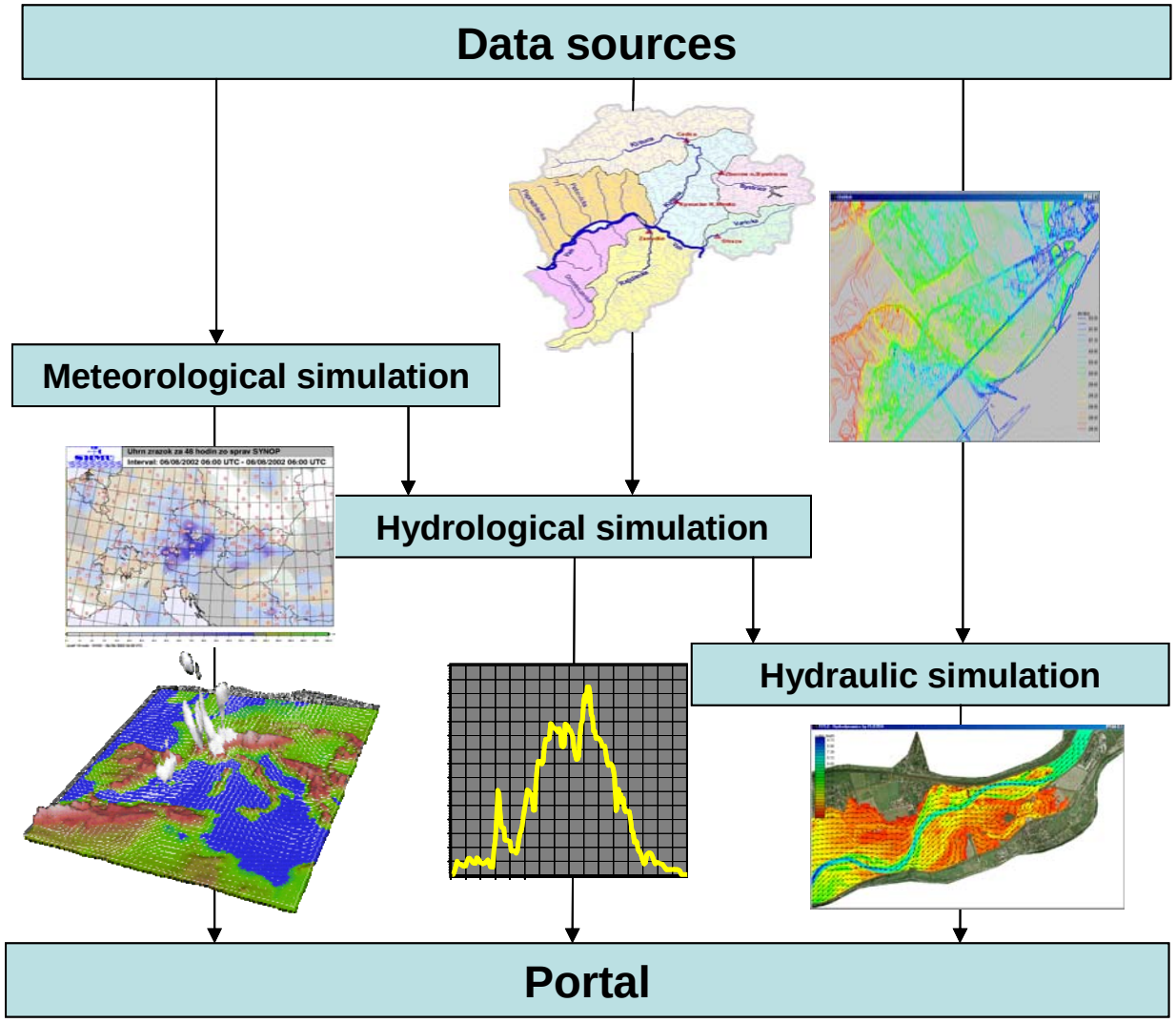
Measured concentration



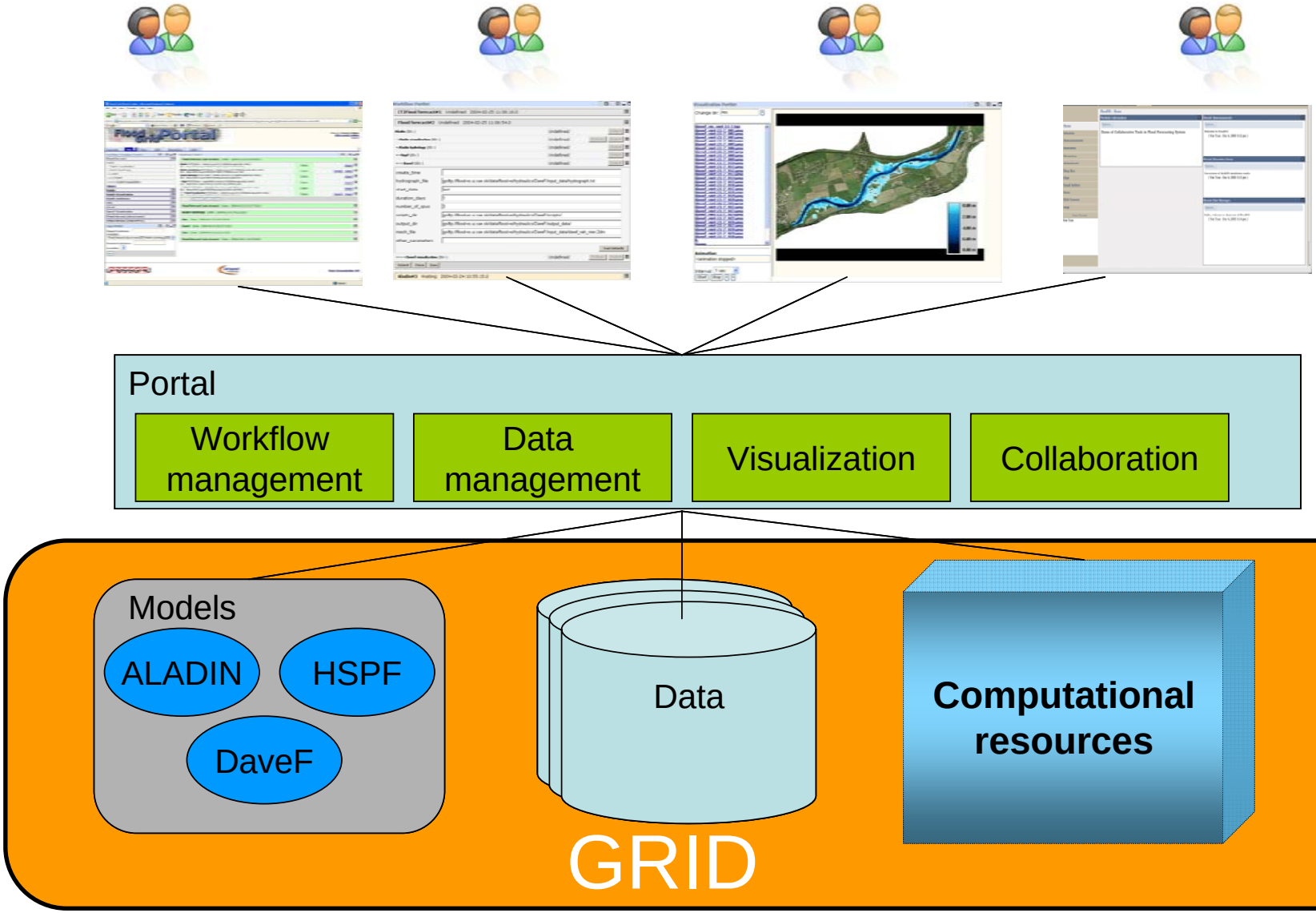
Flood application[1] (T. Viet viet.ui@savba.sk)

- The aim of Flood application is to forecast flood using advanced technologies in computer science (numerical modeling, Grid computing, visualization, ...)
- Flood application consisted from several components:
  - Numerical models (meteorology, hydrology, hydraulics)
  - Workflow management and Data management modules
  - Visualization and Collaboration modules
  - Portal
- Requirements
  - Job execution with priority
  - Data access control
- Porting on EGEE
  - Job execution has been ported to gLite
  - Data management in progress

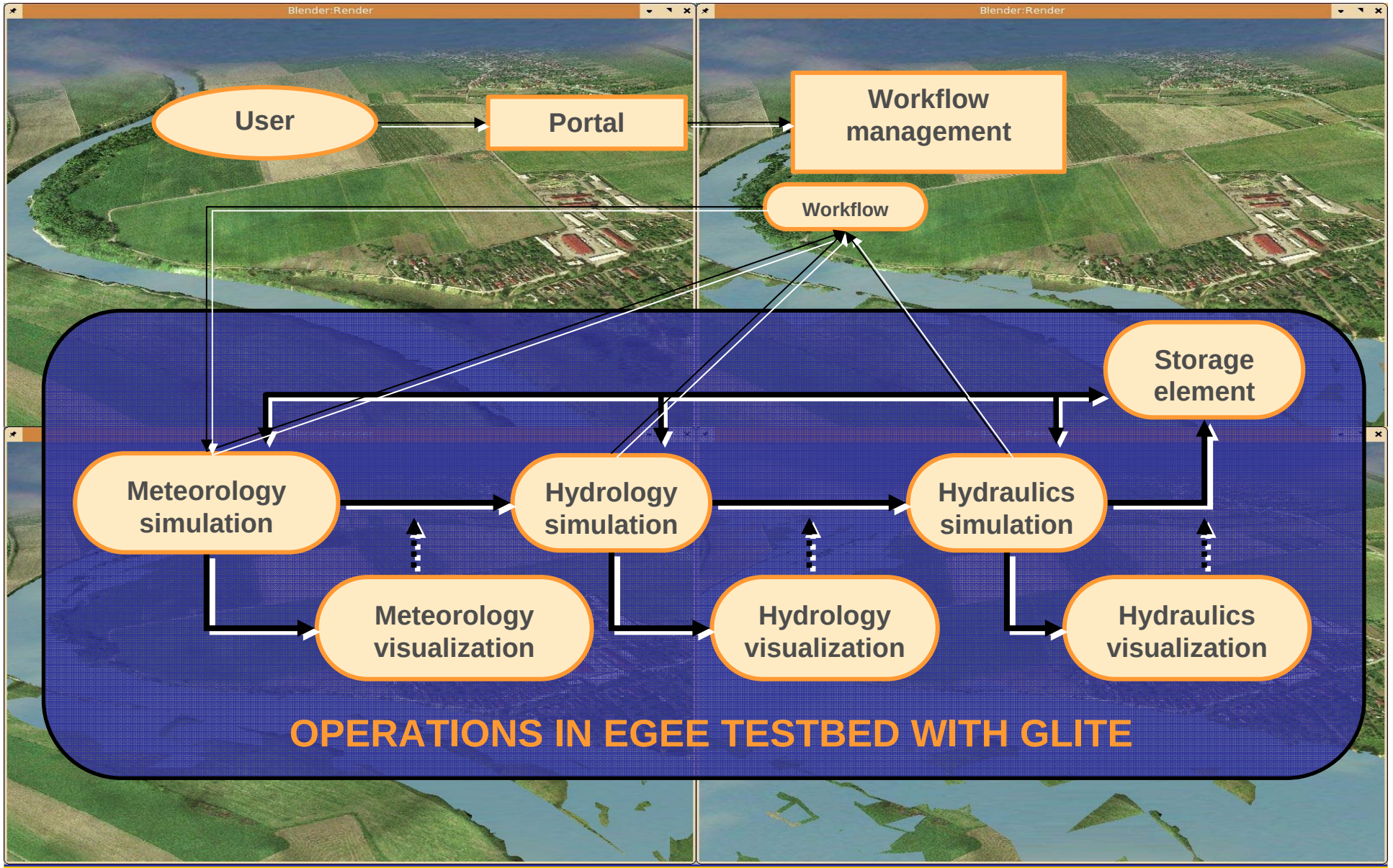
Flood forecasting problem [2]



FloodGrid components [3]



Simplified scenario [4]



Flood application[5]

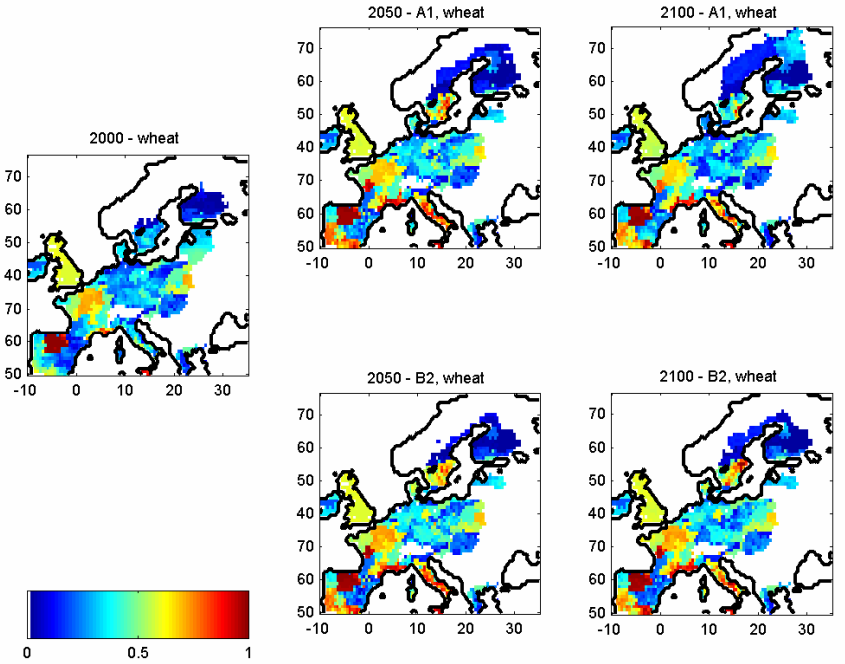
- Tools developed
  - Flood application was developed in CrossGrid project and many components were created during the project
- Difficulty during porting to EGEE
  - Bug in WMS Java API caused error when submitting job directly from workflow manager
    - Have to use command line interface to workaround
    - Bug is recently fixed in CVS
- Useful features of gLite
  - Access control (ACL) for data
  - VOMS support

KLUM@LPJ-C [1]

climate impacts on agricultural land-use changes

K.Ronneberger (ronneberger@dkrz.de)

- The dynamic global vegetation&crop model LPJ-C is extended by a global agricultural land-use model KLUM to simulate future crop patterns under climate change for Europe
- Spatial units are simulated independently from each other, allowing an easy division of the model into several parallel jobs
- The model is currently running on EGEE for different climate and socio-economic scenarios
- By using EGEE the time to solution can be significantly reduced.



Example result: Future land shares of wheat for the two different IPCC climate scenarios A1 and B2

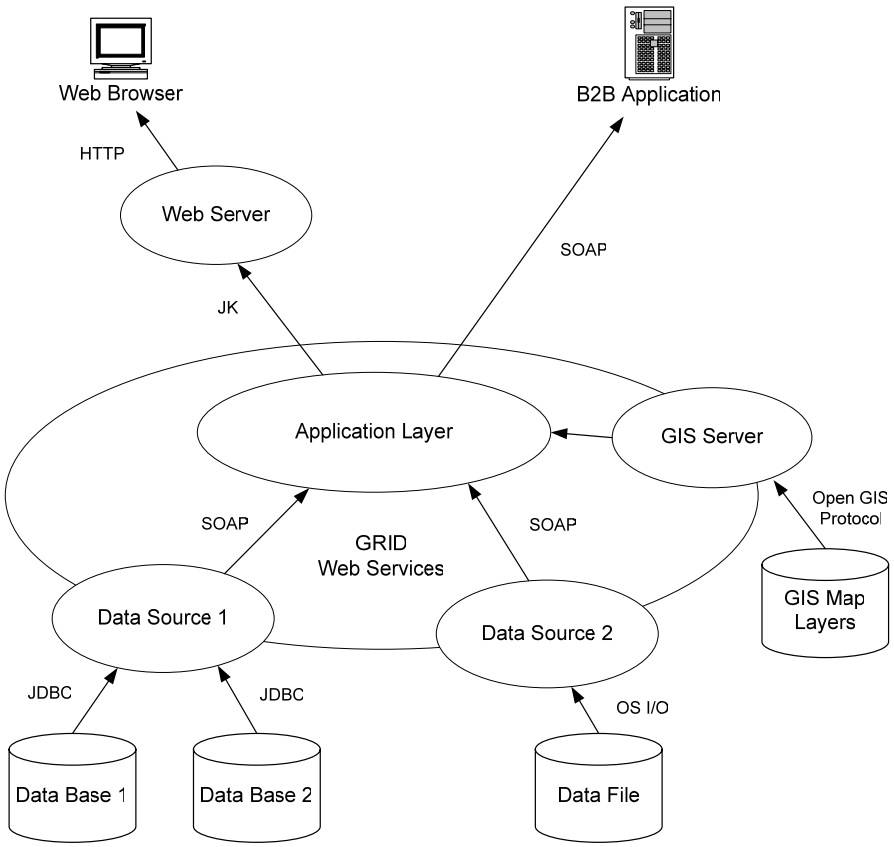
KLUM@LPJ-C [2]

- Tools used:
  - LFC-catalogue
  - Myproxy
  - LCG-tools for data management
- Difficulties encountered:
  - Debugging: tedious and slow due to limited monitoring tools
  - No ‘run-time guaranty’: some jobs took twice the expected time
  - Data management: In one case the upload of the results to the grid failed, so the complete output was lost and the job had to be redone
- Comments:
  - This is only a test application. For the larger models/applications direct and secure access to traditional data storage (e.g. centralized databases) is essential
  - To reach the broader climate-modelling community the processes and tools need to become more transparent and user friendly
  - Models with a broad user community are needed to justify the time, invested to adopt the model to scientific Linux, prepare and upload the data

Meteorology and Space Weather [1]

ZHIZHIN Mikhail jjn@wdcb.ru

- Two environmental data processing projects, SPIDR and IDEAS
  - Already working on GTK-4 (uses WSRF)
  - OGSA-DAI used for interaction with data sources
  - Distributed data mining using Grid infrastructure (with support from the Microsoft Research Europe): project Environmental Scenario Search Engine. The aim is to develop tools to do fuzzy logic queries on terabyte datasets.

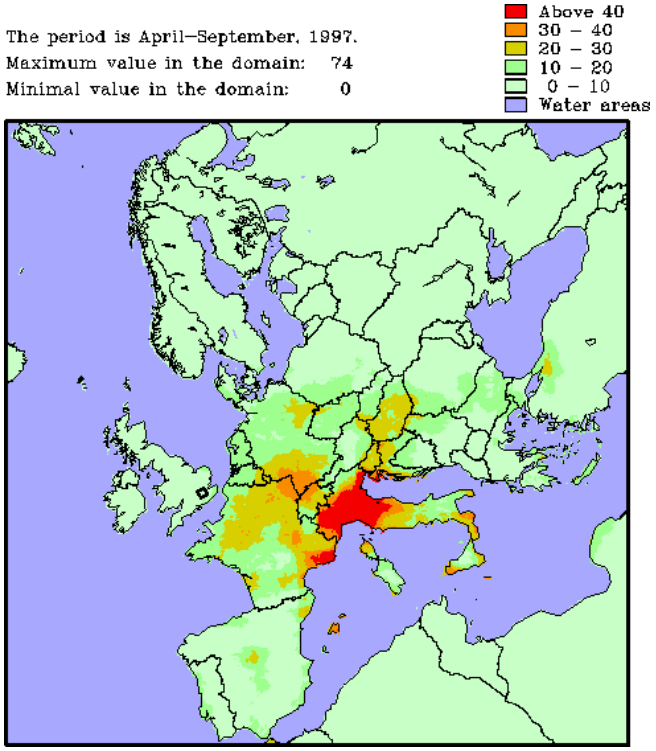


Meteorology and Space Weather [2]

- Difficulties:
  - Slow adoption of the WSRF technology by the GRID community
  - Lack of a standard protocol to exchange scientific data (many different format)
  - Lack of high level Earth Science-oriented metadata scheme

Pollution (T.ostromsky ceco@parallel.bas.bg)

- BAS(Bulgaria): implementation of a large scale air pollution model on EGEE, for evaluation of the concentrations of a large variety of chemical species, responsible for the air pollution
- Model based on a parallel architecture
  - Distribution of computation
  - Development of new advanced splitting scheme (planned)



Difficulties:

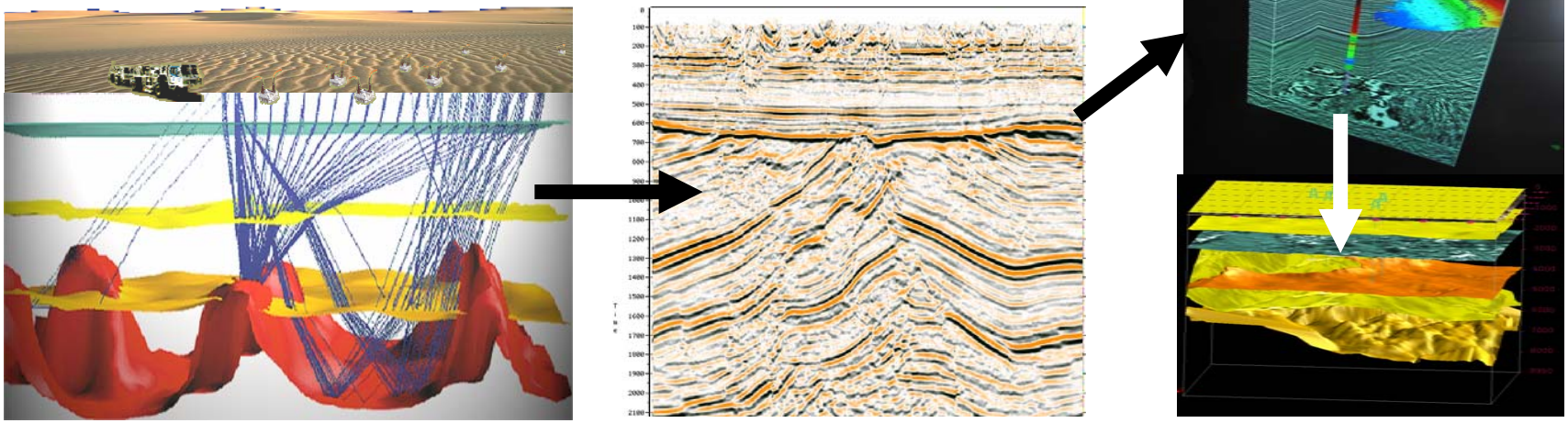
- Large sets of data input and output: need a specific management
- Optimisation to take full advantage of the Grid

Geocluster© [1] (gyouinou@CGG.com)

- Goal of the application
  - **Expanding GEOsciences on DEmand (EGEODE) VO**
    - **Dedicated to research in geosciences**
  - **Seismic processing Generic Platform**
    - **Based on *Geocluster*©, the first industrial application**
    - **Used in production at CGG**
- Description of the application (numerical approaches used, tools…)
  - **Process seismic data and to explore the composition of the Earth's layers**
  - **Include several standard tools for signal processing, simulation and inversion (model optimization)**
- Specific requirements
  - **License management and accounting system**
  - **Compilation server (first industrial version available)**
  - **MPI; Data management linked with the application**
- Porting on EGEE: Status, degree of satisfaction
  - **Batch part of the application fully available on the Grid**
  - **7 sites on production grid + GILDA**

Geocluster© [2]

- The general benefits of grid computing
  - Access to computing resources without investing in large IT infrastructure
  - Lower the total cost of IT by sharing available resources
- And the specific benefits for Research community
  - *Access to a comprehensive, industrial software*
  - Free the researcher from the additional burden of managing IT hardware and software complexity and limitations
  - Have a framework to share data and project resources with other teams across Europe and the rest of the world
  - Share best practices, support and training more easily
  - Enable cross-organizational teamwork and partnership



Geocluster© [3]

- Difficulty encountered: missing functionalities….
  - **Security management of the sites**
  - **No guidelines to accept and welcome new countries**
- Tools developed to fill the gap
  - **Compilation and license server**
  - **Specific tools dedicated to the VO and Geocluster©: eTOOLS**
    - **Follow job's evolution on every stage of the application**
    - **Responsible of the data management (connection with LFC)**
- gLite expectancies:
  - **Successful VOMS test was organised between CC-IN2P3, IPSL-IPGP and CGG**
- Comments
  - **challenges for EGEE-II:**
    - **VO deployment on more EGEE sites**
    - **Attract new users (already started)**
    - **Build an intra-grid from several CGG R&D centres**
    - **Interconnection with our own job manager**

Other APPLICATIONS

(1) Direct Monte Carlo simulation of non thermal sources of hot oxygen in the upper atmosphere of Mars (IPSL): long job (90-120h)

=> Interest for GEANT4 software

- To design and study satellite instruments
- To estimate the collision processes with electron production in the earth upper atmosphere.

(2) Data Collection for Geoscientists (NTUA, Greece)

- Registration of databases available on web
- Need of LFC ….. Pb solved

CONCLUSION[1]

REQUIREMENTS
TO EGEE FROM « Earth Science » COMMUNITY

- **Secure and controlled access to data and metadata (on going)**
- **Prediction and analysis of natural risk need many CPUs (≥100) without reservation.**
- **MPI: many simulations and models need MPI**
- **Interface with Web services specific for application**
- **Interoperability with other grid middleware (SGE) or SW architectures (Web Services)**
- **Filtered access to license server**
- **Accounting**

CONCLUSIONS [2]

- A large variety of applications ported on EGEE ; that incites new users
  - Earth Observation: unique case of validation of a whole dataset (time and space) and several algorithms
  - Hydrology: possibility to fulfill in real or quasi real time the Civilian Society requests.
  - Seismology: example of data interpretation on alert
  - Geocluster: initiative R&D de la Compagnie Générale de Géophysique completely ported
- Success rate (90-100%) on LCG2 – cf Earth Observation
- Interactive Collaboration of the teams around a project
- Difficulties related to funding (non-funded activity)

CONCLUSIONS [3]

- **PERSPECTIVES and ADDED VALUES**

(1) Due to the success rate of LCG2 little by little the ES community is porting application on EGEE and feels confident

(2) Enough good applications to show that EGEE testbed works

(3) gLite will bring some functionalities missing in LCG2 required by ES users

(4) EU Specific support action: DEGREE for Disseminating EGee in Earth sciencE

- Good evaluation, waiting for the amount of funding
- The aim is to contribute to the deployment of Grid technology in the research and industry ES sector