imath.cesga.es, the VO for the European mathematicians

M.T. Sánchez[1], R. Cao[2], J. Fernández[4], M. García-Magariños[3], F. García-Torre[4], J.M. Gesto[5], A. Gómez[1], W. González-Manteiga[3], J.M. Gutiérrez[4]

[1]CESGA, Fundación Centro Tecnológico de Supercomputación de Galicia. Santiago de Compostela, Spain
[2]Departamento de Matemáticas. Universidad de A Coruña. A Coruña, Spain.
[3]Departamento de Estadística e Investigación Operativa. Universidad de Santiago de Compostela. Spain.
[4]Departamento de Matemática Aplicada y Ciencias de la Computación. Universidad de Cantabria. Spain.
[5]Departament de Matemàtica Aplicada III. ETSECCPB, UPC, Barcelona, Spain.

http://mathematica.nodo.cesga.es

tsanchez@cesga.es

Outline

- i-MATH Project
  - Main objectives
  - Computing Platform
- GRID-MATHEMATICA Activity
  - imath.cesga.es VO
  - Infrastructure
- Three study cases
- Conclusions

- Ingenio MATHEMATICA (i-MATH) is a CONSOLIDER singular research project.

  http://www.i-math.org

  - To increase mathematical presence on the Spanish and worldwide scene.
    - From basic research to applications
    - How to understand the physical world
    - The essential computational support
    - Direct applications to society

  - COMPUTING Platform
    - To promote the use of computational methods.
    - To increase the involvement of Spanish researchers in strategic and hot subjects in scientific computation

GRID-MATHEMATICA

- Objectives:
  - To design a new Virtual Organization VO to use in i-MATH.
  - To create a first validation prototype by means of executing in a new GRID infrastructure three cases of mathematical and computational interest.
  - To expand this infrastructure to all the members of i-MATH.

Virtual Organization imath.cesga.es

- Creation of VO imath.cesga.es in EGEE infrastructure:
  - First VO for Mathematics Research.
  - 1,600 people
  - 300 research groups

http://imath.cesga.es

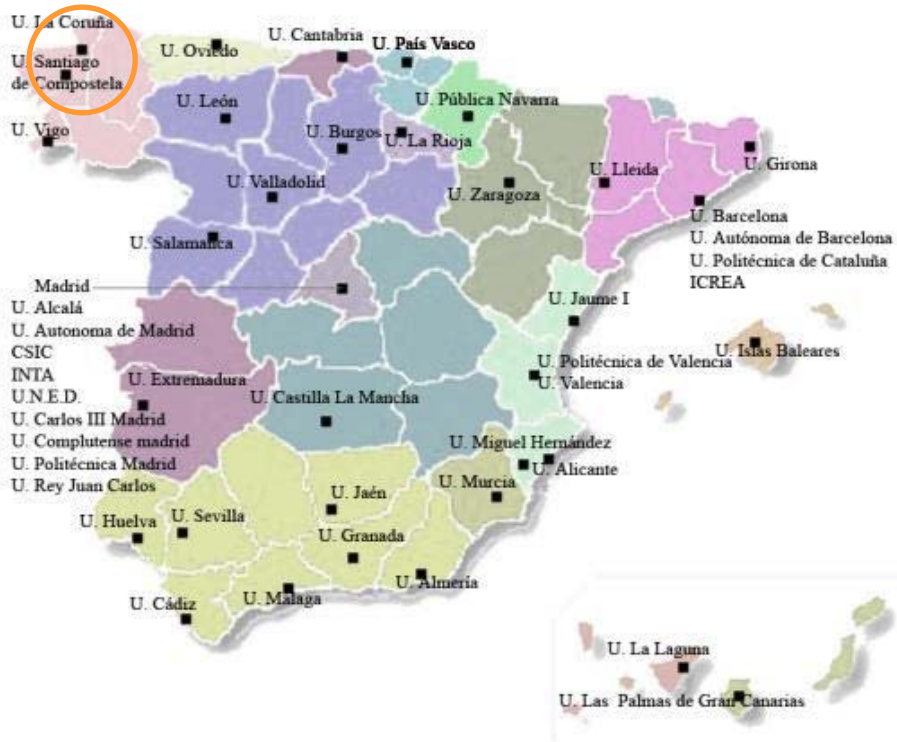
Enabling Grids for E-sciencE

CESGA

Infrastructure:

Three sites:

CESGA

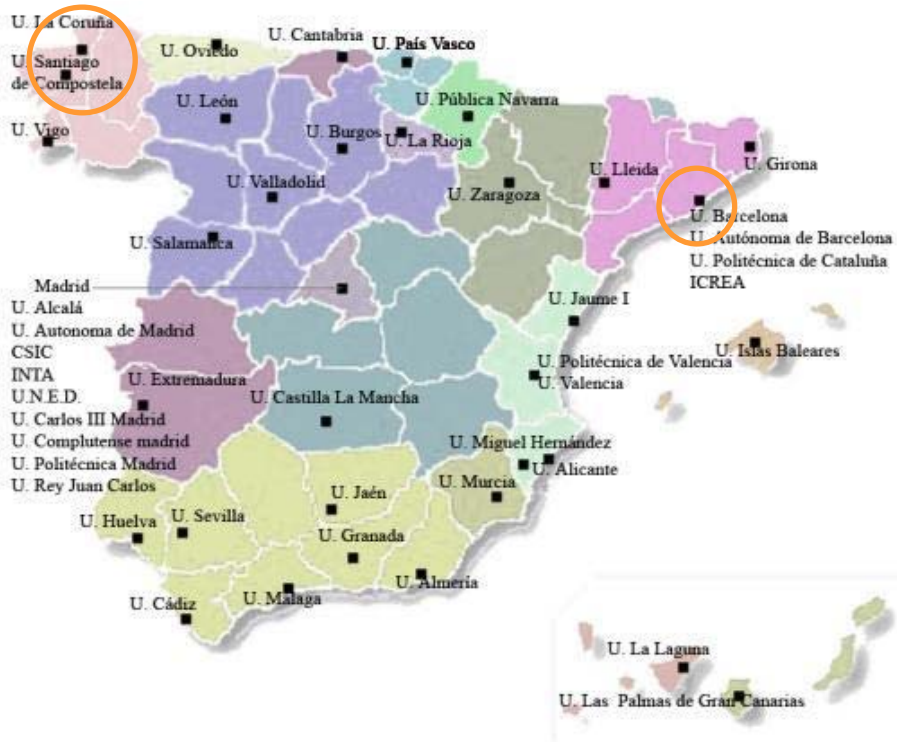


Infrastructure:

Three sites:

CESGA

IMUB



egee
Enabling Grids for E-sciencE

CESGA

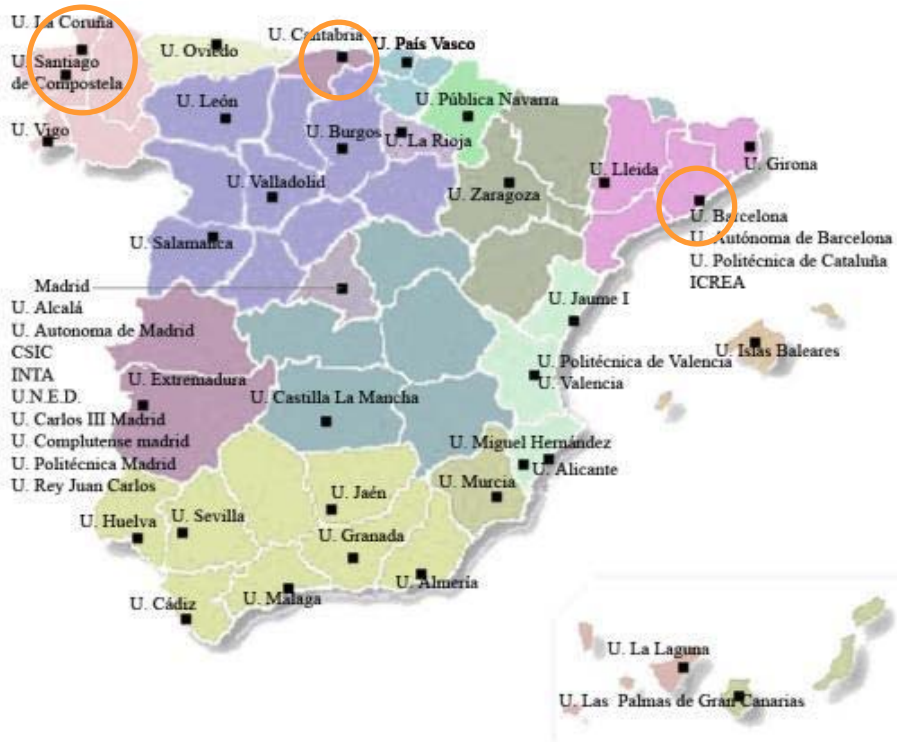
Infrastructure:

Three sites:

CESGA

IMUB

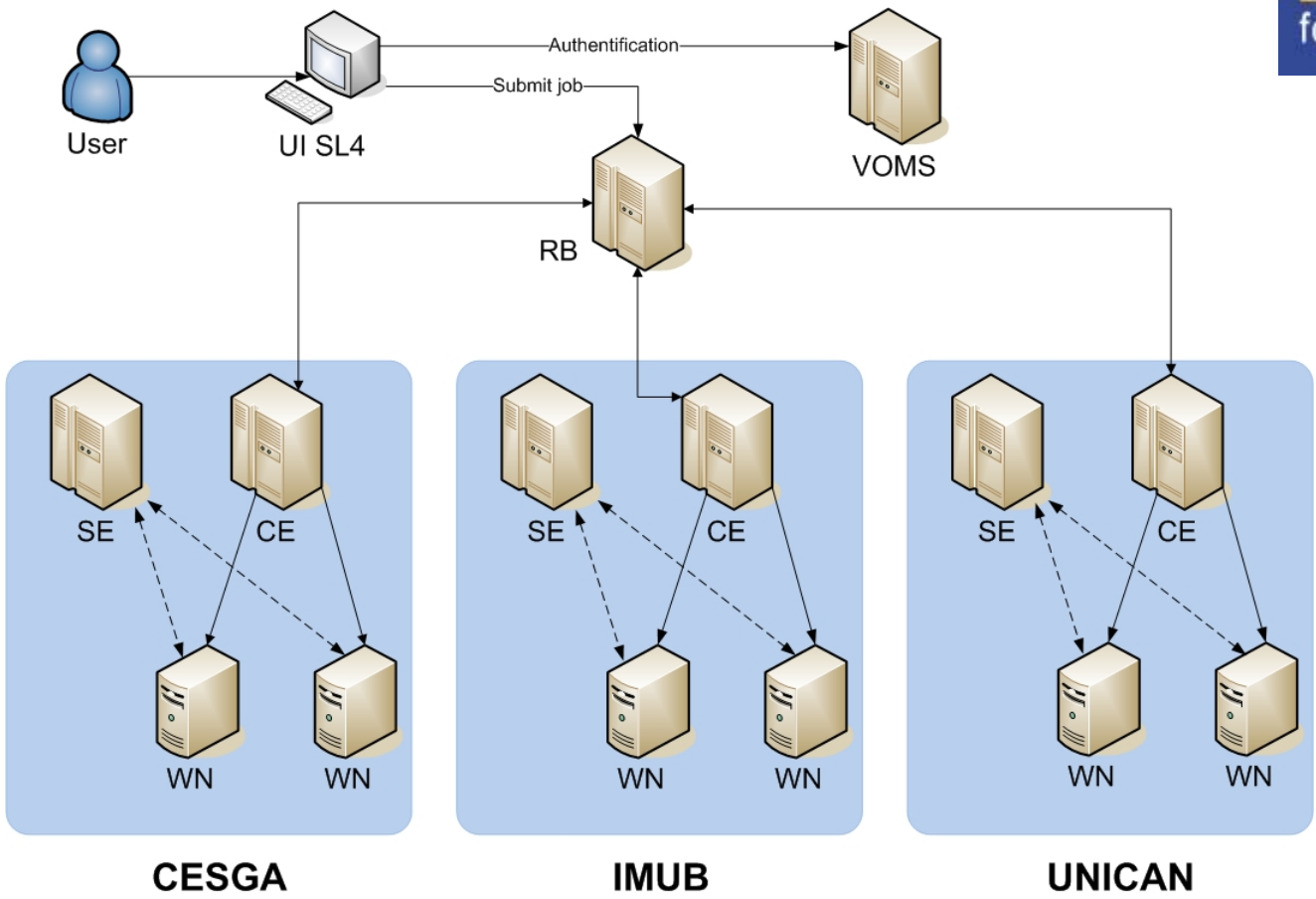
UNICAN



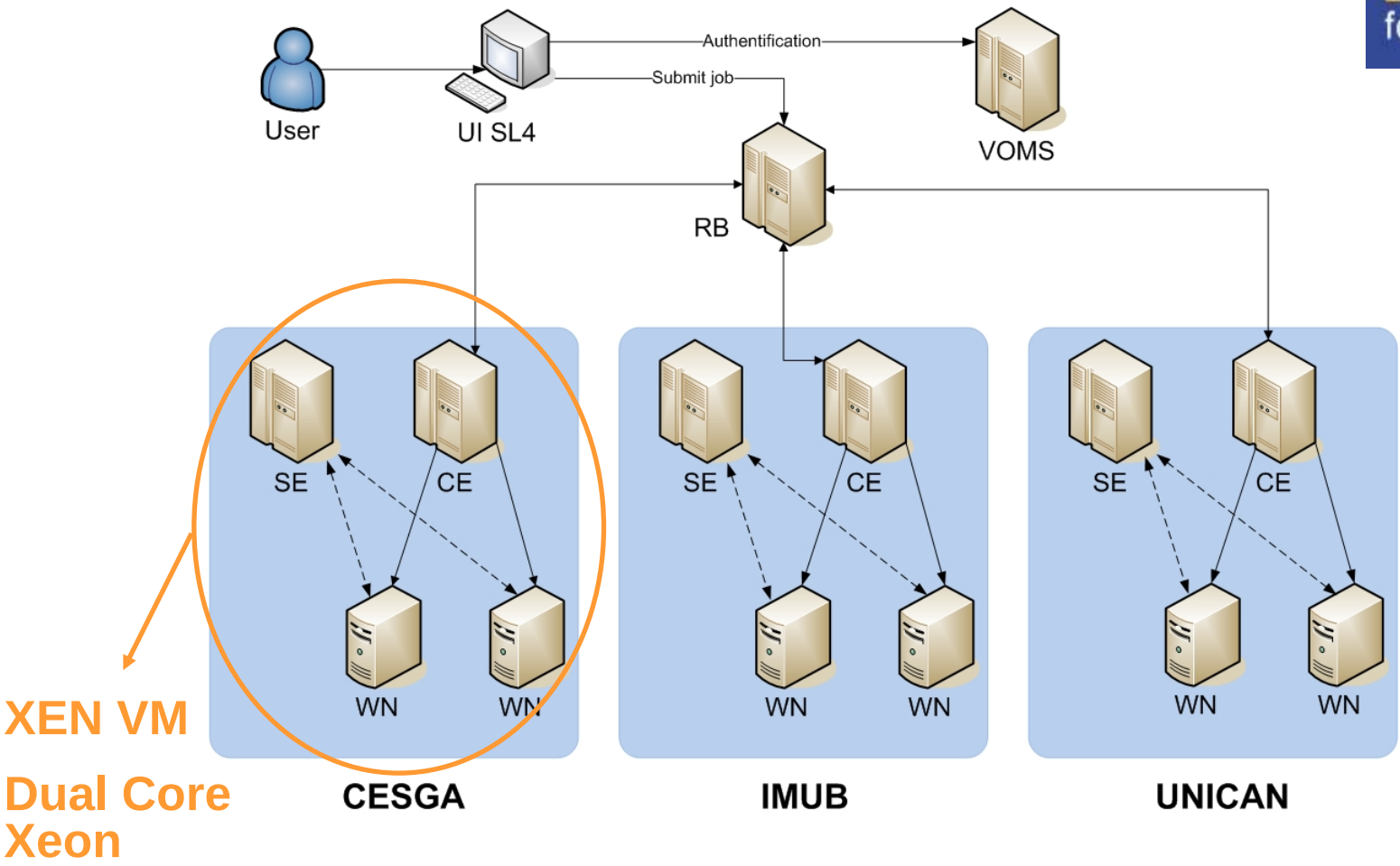
Enabling Grids for E-sciencE

CESGA

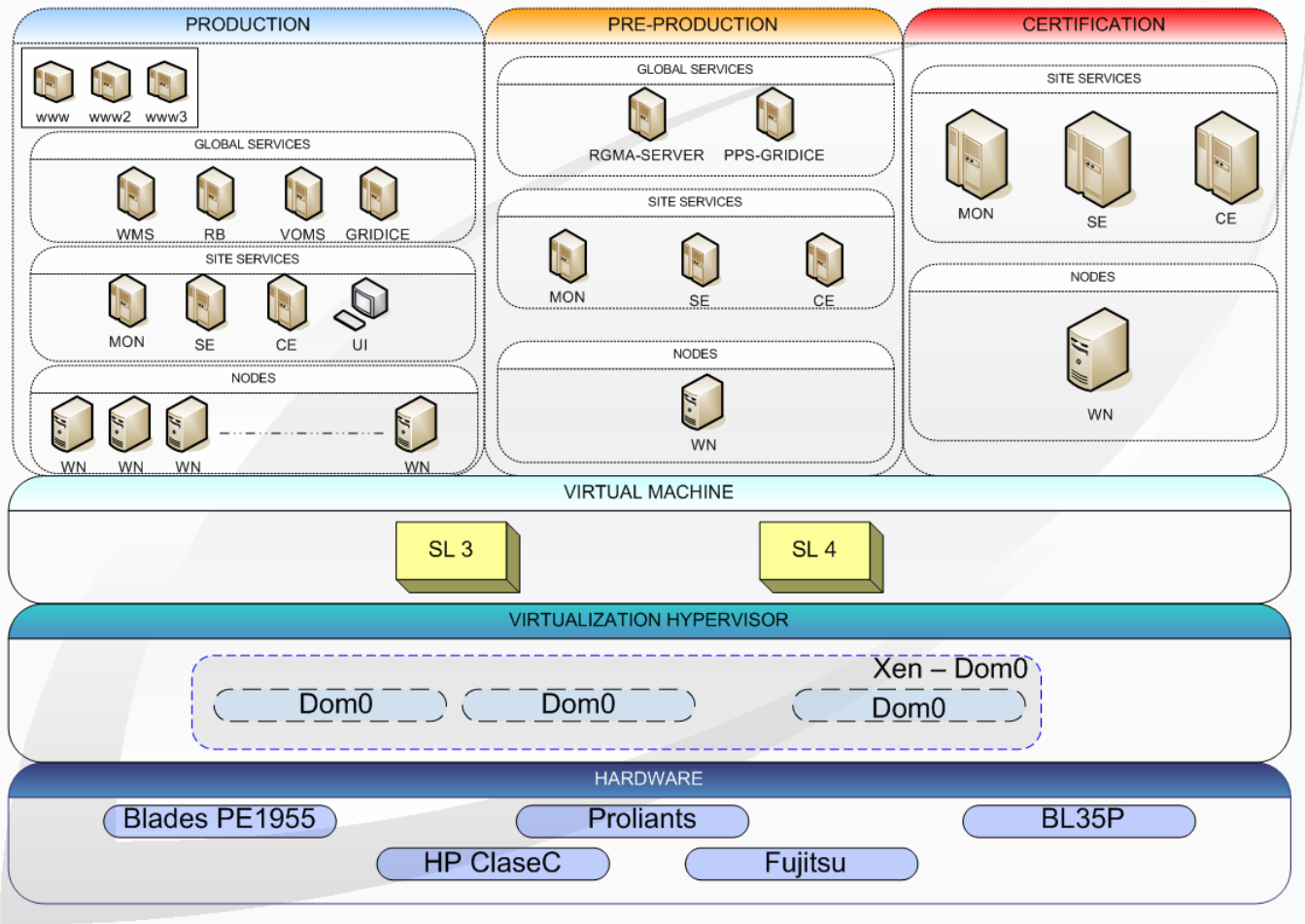
Infrastructure:



Infrastructure:



Infrastructure: CESGA-EGEE



356 cpus

1TB storage

VM nodes

Study cases

- To evaluate the applicability in Mathematics research of GRID infrastructure
  - Case 1: Support Vector Machines (SVMs) adaptation to case-control studies with SNP data
  - Case 2: Monte Carlo simulation of climate sensitivity
  - Case 3: Estimation of Fekete points

Study cases

- To evaluate the applicability in Mathematics research of GRID infrastructure
- Make easy to use GRID infrastructure
  - Testing different solutions (Nimrod/G, GridWay, …)
  - P-Grade Grid Portal
    - simplicity
    - powerful workflow

http://uisl4.imath.cesga.es:8080/gridsphere/gridsphere

Study cases

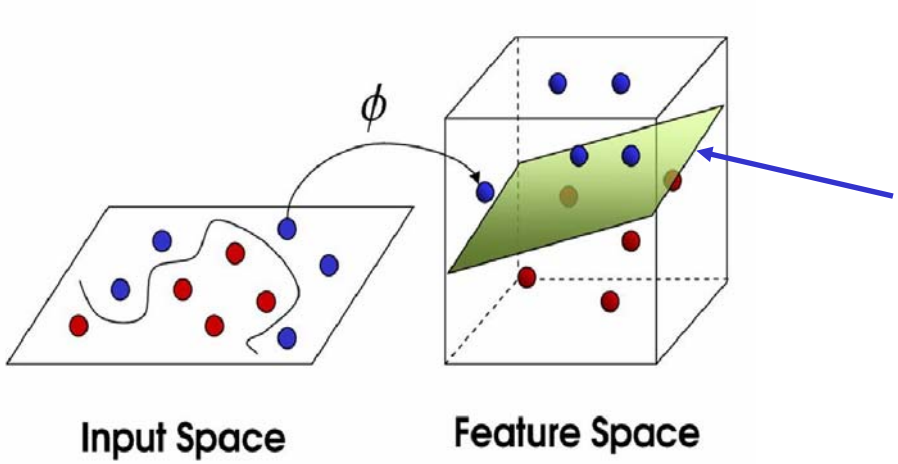
Case 1: Support Vector Machines (SVMs) adaptation to case-control studies with SNP data.

- Genetic association studies are thought as the biggest promise facing the discovery of the genetic basis of common diseases (diabetes, cancer, psychiatrics,…).
- Single Nucleotide Polymorphisms (SNPs) are responsible for around 90% of genetic variability among humans.
- Need to develop new statistical methodologies qualified to dissect (since the genetic point of view) these complex diseases.

Study cases

Case 1: SVMs, how do they work?

- Construction of a linear classifier in a high-dimensional feature space.
- This is made by means of kernels, similarity measures between individuals obtained from their SNP values.
- Compulsory adaptation to work with SNPs (categorical data).



Study cases

Case 1: SVMs, need of GRID techniques

- SVMs are itself computationally demanding.
- Our adaptation, carried out to work with SNPs, transform each individual genotype in a number of configurations that is two up to the power of a great number, increasing dramatically the computational efforts required.
- Large amounts of genetic data to be studied

Study cases

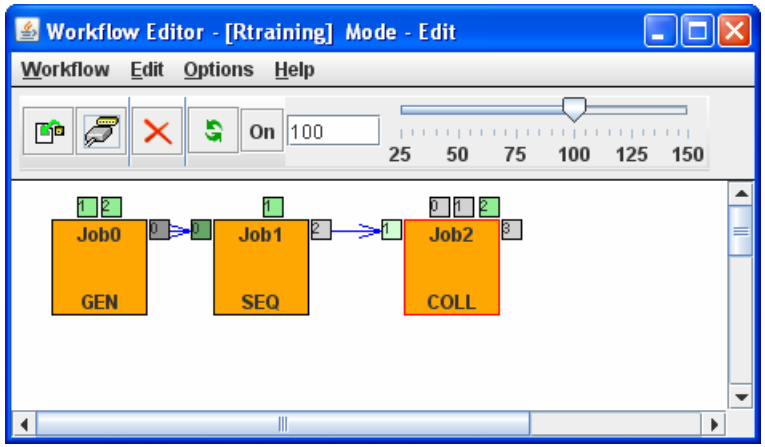
Case 1: SVMs, need of GRID techniques

- SVMs are itself computationally demanding.
- Our adaptation, carried out to work with SNPs, transform each individual genotype in a number of configurations that is two up to the power of a great number, increasing dramatically the computational efforts required.
- Large amounts of genetic data to be studied

- Needed software: Statistical Free Software R.
- Completely installed in CESGA-EGEE

Study cases

Case 1: Support Vector Machines, training-test

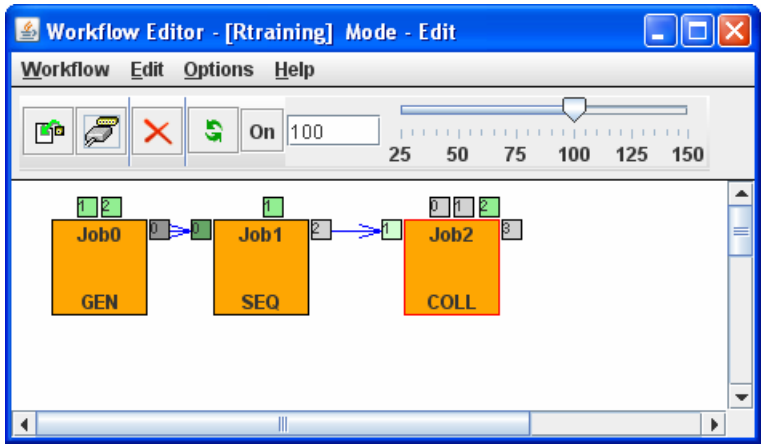


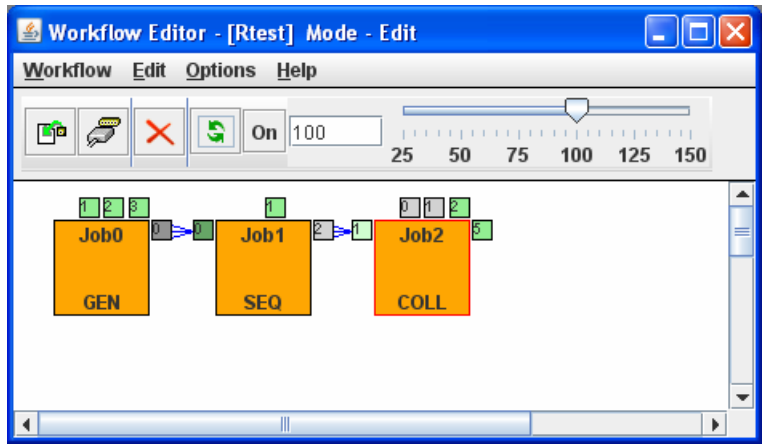
Training phase

Prediction case-control

Study cases

Case 1: Support Vector Machines, training-test





Training phase

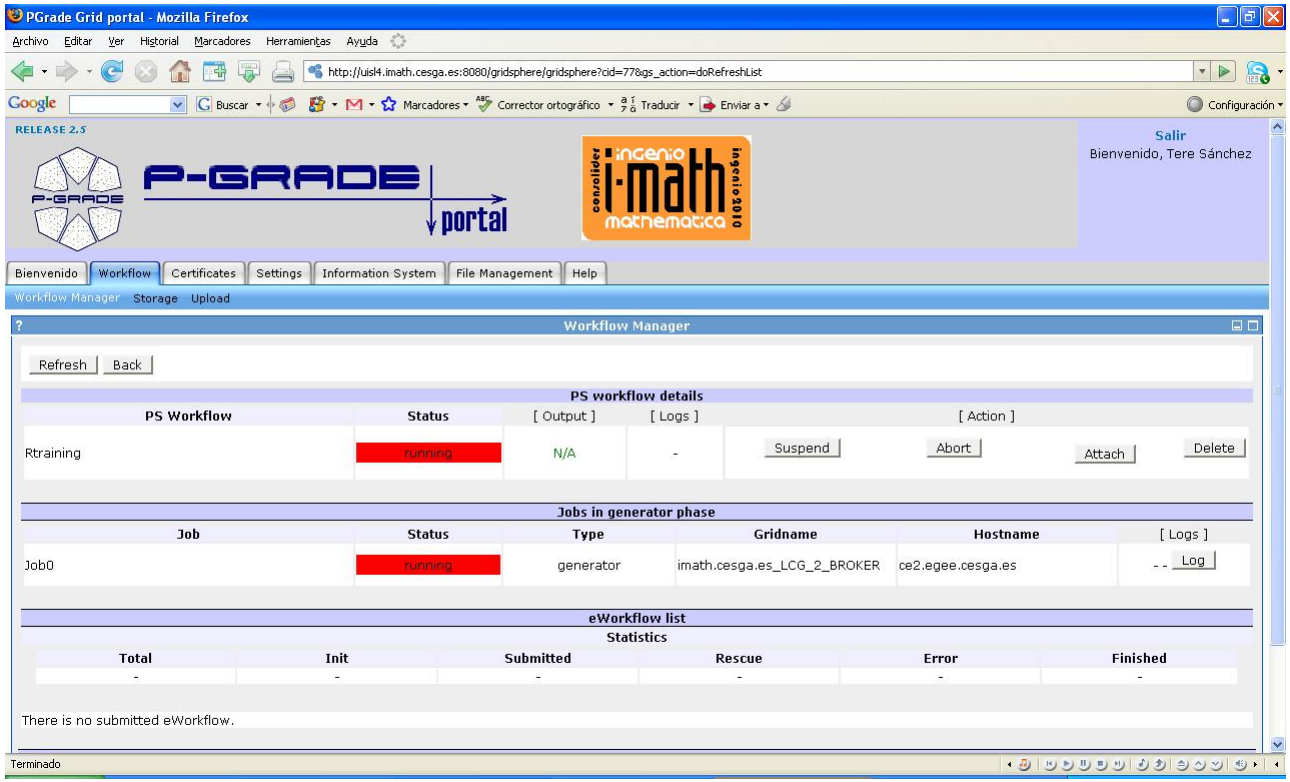
Prediction case-control

Test phase

Method check with data

Study cases

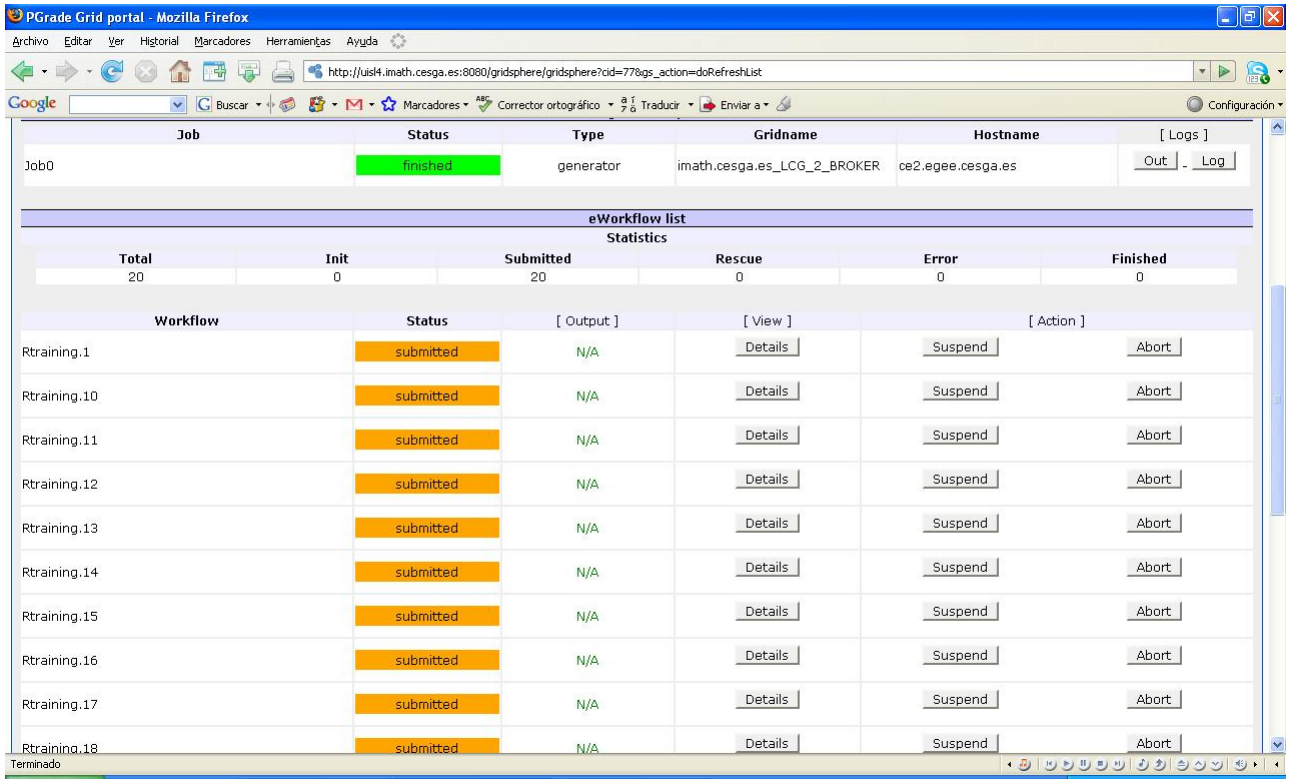
Case 1: Support Vector Machines, training



cpus: 20 Time: 140 min

Study cases

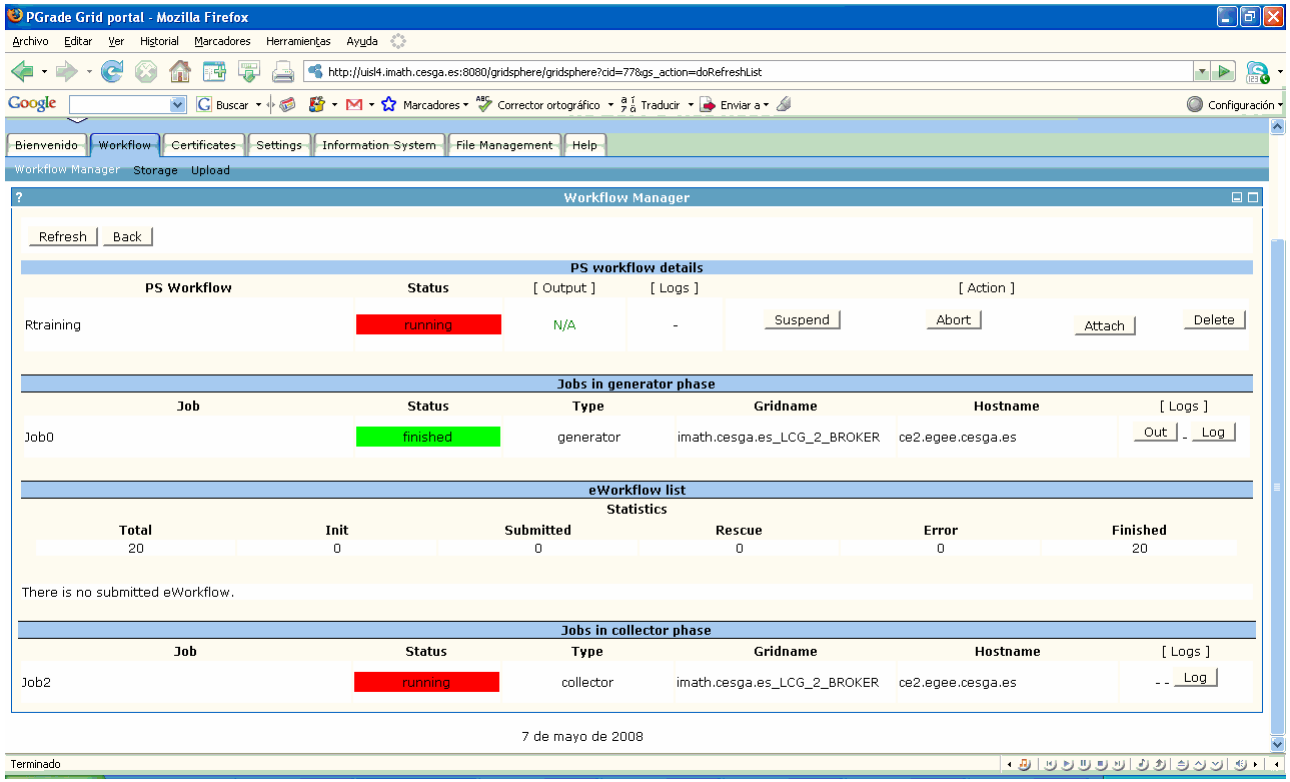
Case 1: Support Vector Machines, training



cpus: 20 Time: 140 min

Study cases

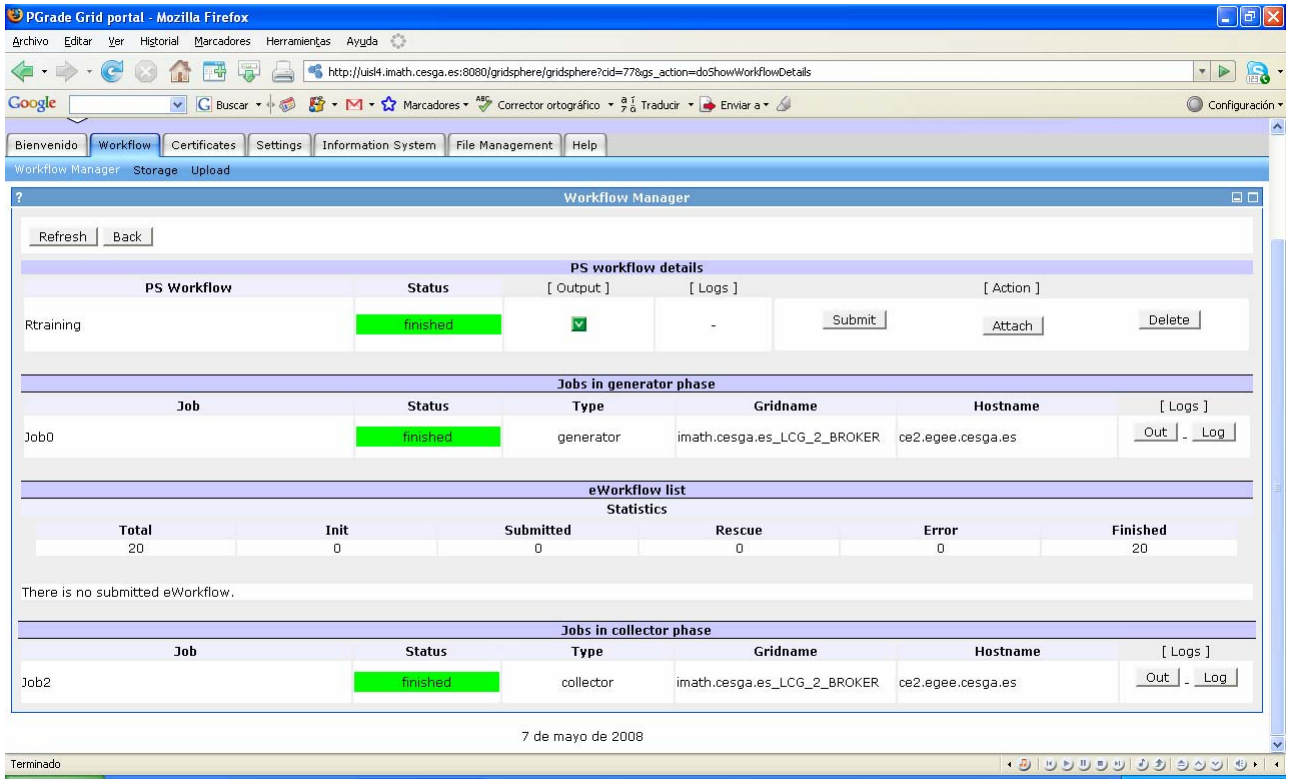
Case 1: Support Vector Machines, training



cpus: 20 Time: 140 min

Study cases

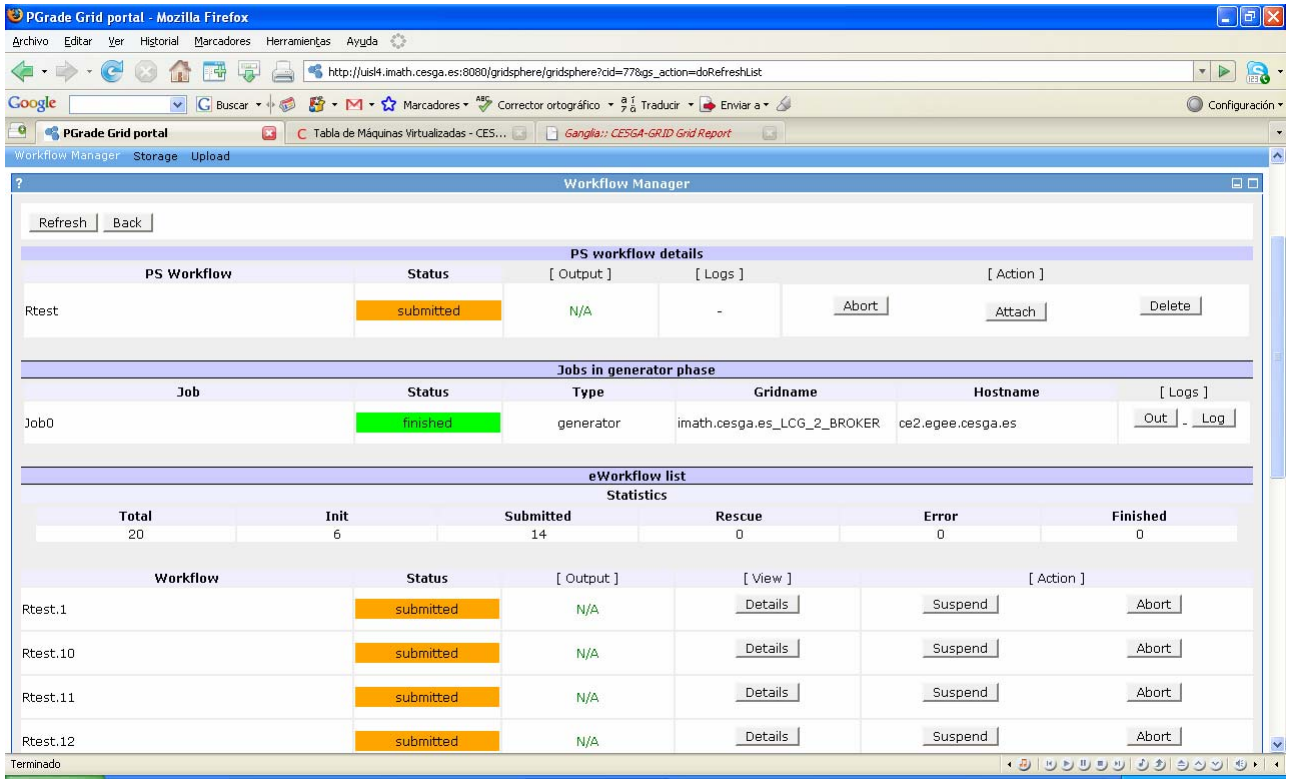
Case 1: Support Vector Machines, training



cpus: 20 Time: 140 min

Study cases

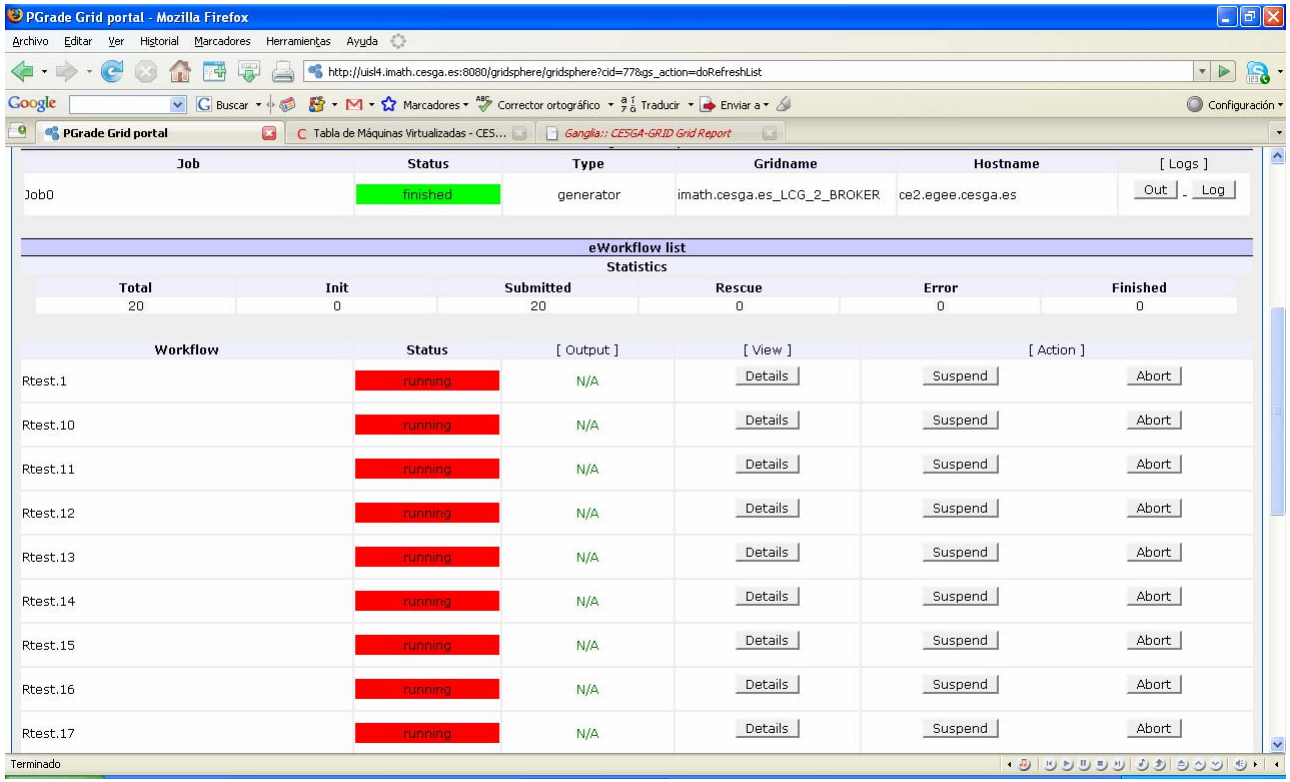
Case 1: Support Vector Machines, test



cpus: 20 Time: 60 min

Study cases

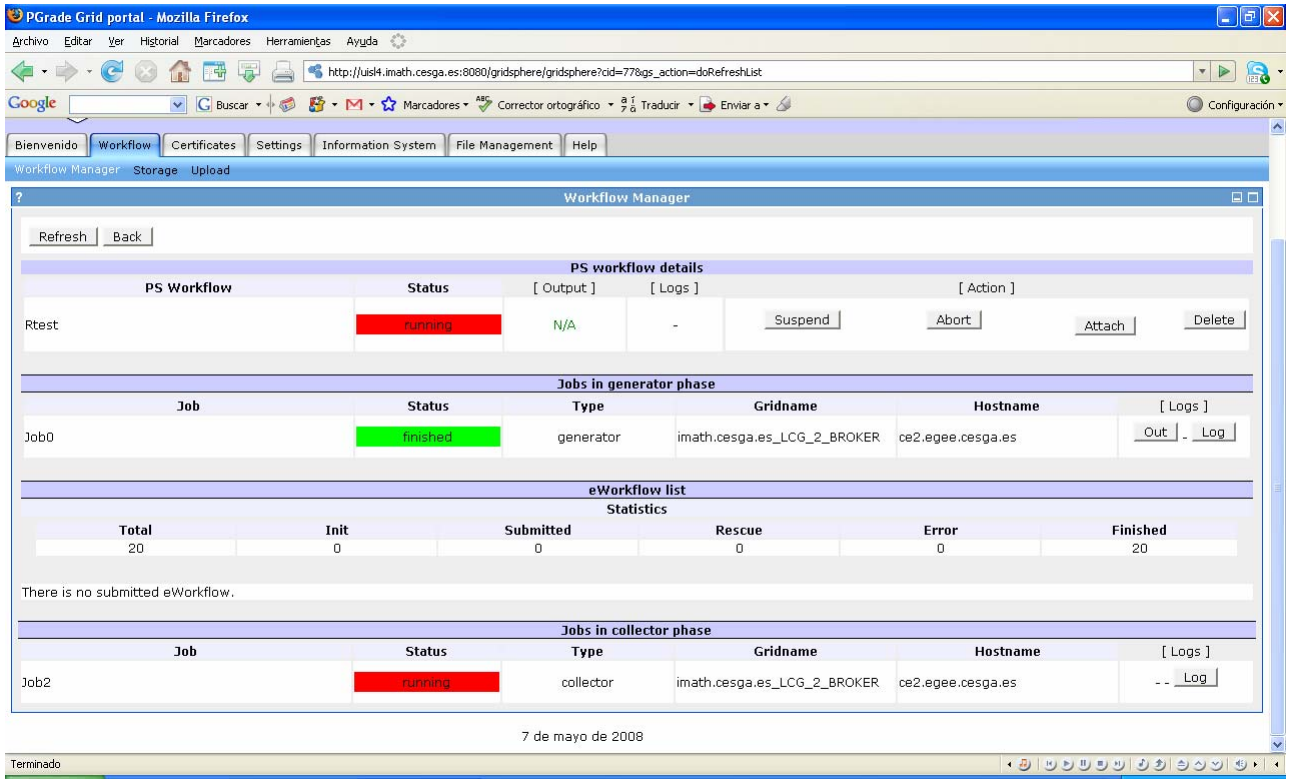
Case 1: Support Vector Machines, test



cpus: 20 Time: 60 min

Study cases

Case 1: Support Vector Machines, test

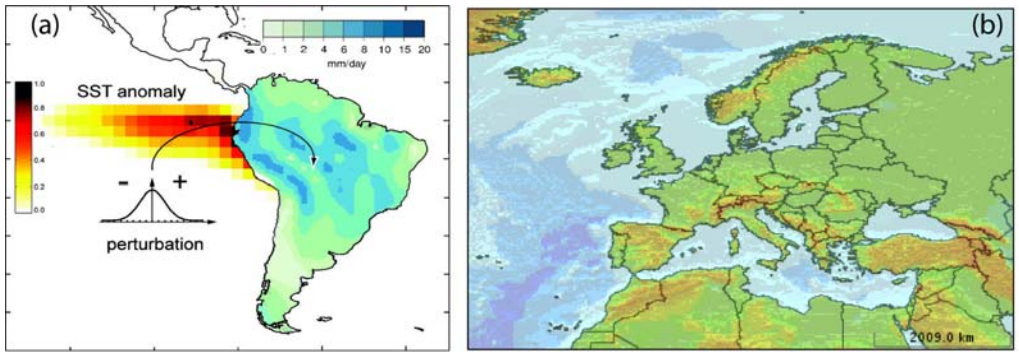


cpus: 20 Time: 60 min

Study cases

Case 2: Monte Carlo simulation of climate sensitivity

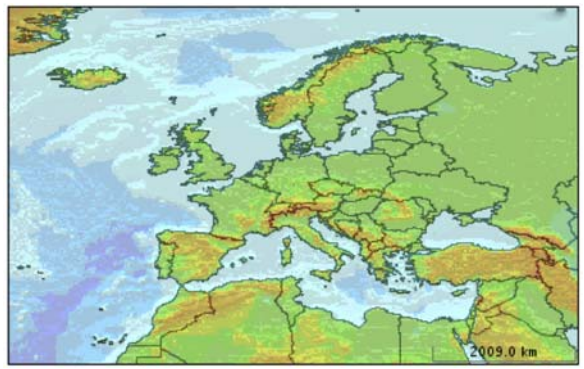
- EELA Project: Develop a simulation and analysis tool to predict local impacts of “El Niño” in Latin America
  - Sensitivity study SST-Precipitation
- Goal: To analyze the problem of soil moisture content over Europe



Study cases

Case 2: Monte Carlo simulation of climate sensitivity

- CAM: Global Atmospheric Model.
  - Produce information such as Precipitation or Temperature in a grid of points over the globe.
- Different realizations of the perturbation in a Monte Carlo approach.

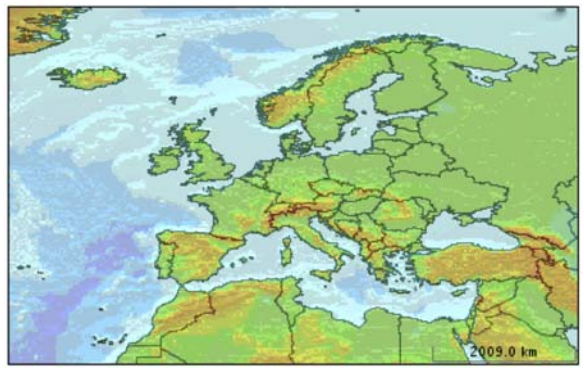


Study cases

Case 2: Climate sensitivity, requeriments

- Runs from days to weeks.
- AMGA metadata catalogs
- 500 GB storage

- Work in progress



Study cases

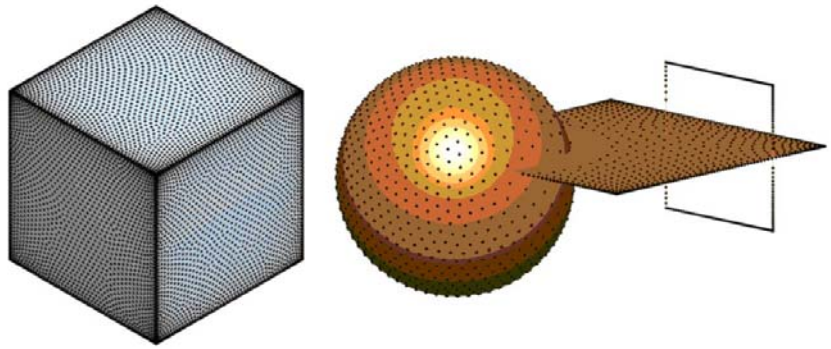
UNIVERSITAT POLITÈCNICA DE CATALUNYA

Case 3: Estimation of Fekete points

N^{th} order Fekete points are the N-uples $\omega_N = \{x_1, \ldots, x_N\},\ x_i \in \mathbb{R}^d$ which minimize on a compact $S \subset \mathbb{R}^d$ different potential energy functionals $\mathcal{I}_N$ depending on the relative distances among N points.

$$\mathcal{I}_N(x) = \sum_{1 \le i < j \le N} \log \frac{1}{|x_i - x_j|}$$

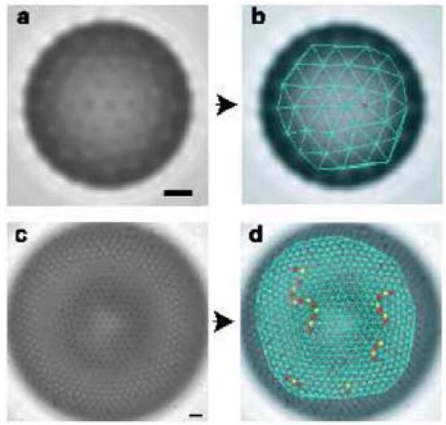
$$\mathcal{I}_N(x) = \sum_{1 \le i < j \le N} \frac{1}{|x_i - x_j|^s}$$



CESGA

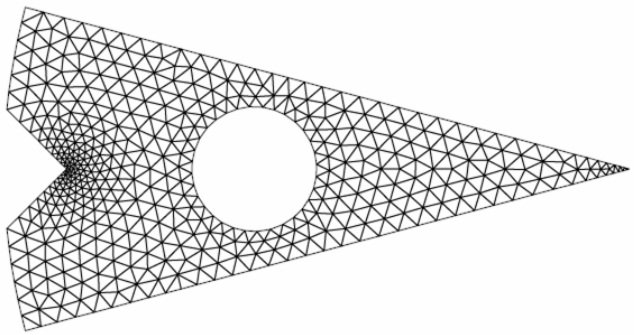
Study cases

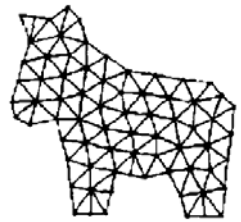
Case 3: Estimation of Fekete points



Applications in:

- Physics
- Chemistry
- Numerical Methods
- Complexity Theory

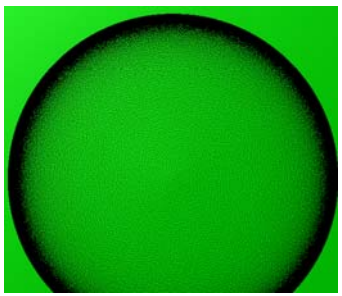




Study cases

Case 3: Estimation of Fekete points

- An algorithm developed for the numerical estimation of local minima.
  - Versatile, robust and efficient
  - Sphere case profusely analyzed
  - Considerable amount of experimental data needed for different geometries and energies.



Bendito, Carmona, Encinas, Gesto, Estimation of Fekete points. *Journal of Computational Physics* 225 (2007) 2354-2376.

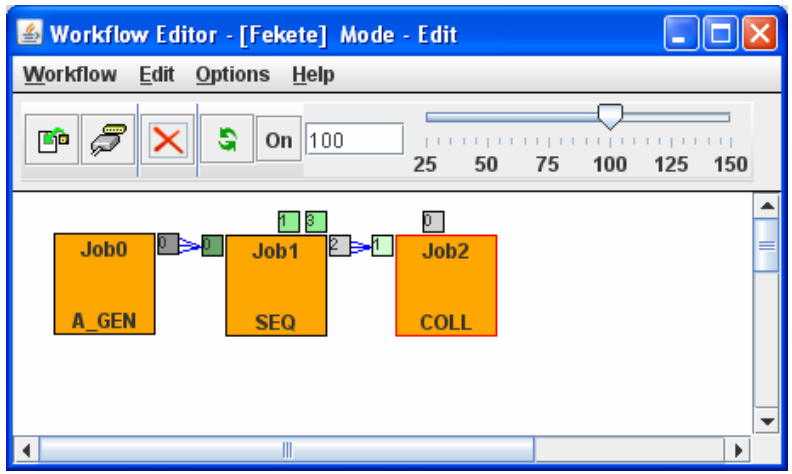
Study cases

Case 3: Estimation of Fekete points, requirements.

- Lots of relatively small jobs
  - From minutes to some hours
- 2,000 computational hours for each geometry and energy
- 5 GB storage

Study cases

Case 3: Estimation of Fekete points



N=482

3000 runs

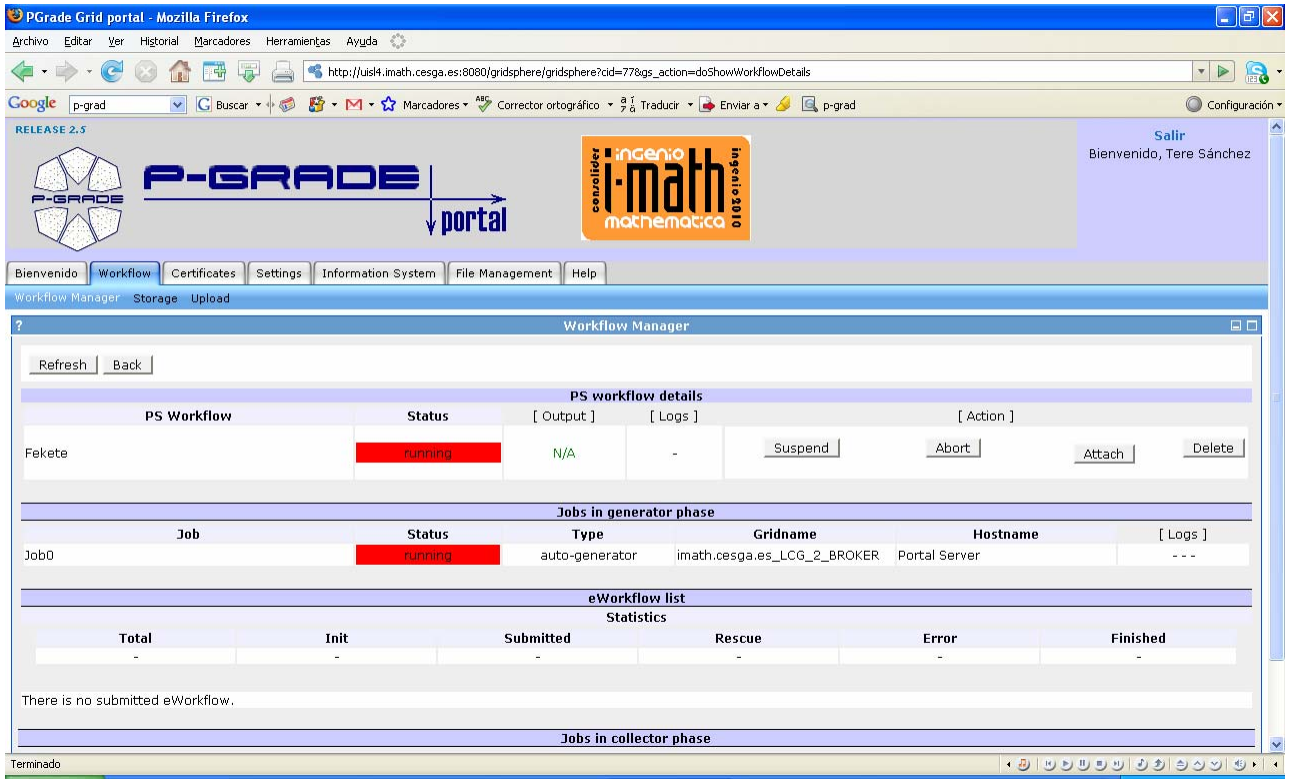
30 cpus

100 runs each cpu

Parametric studies

Study cases

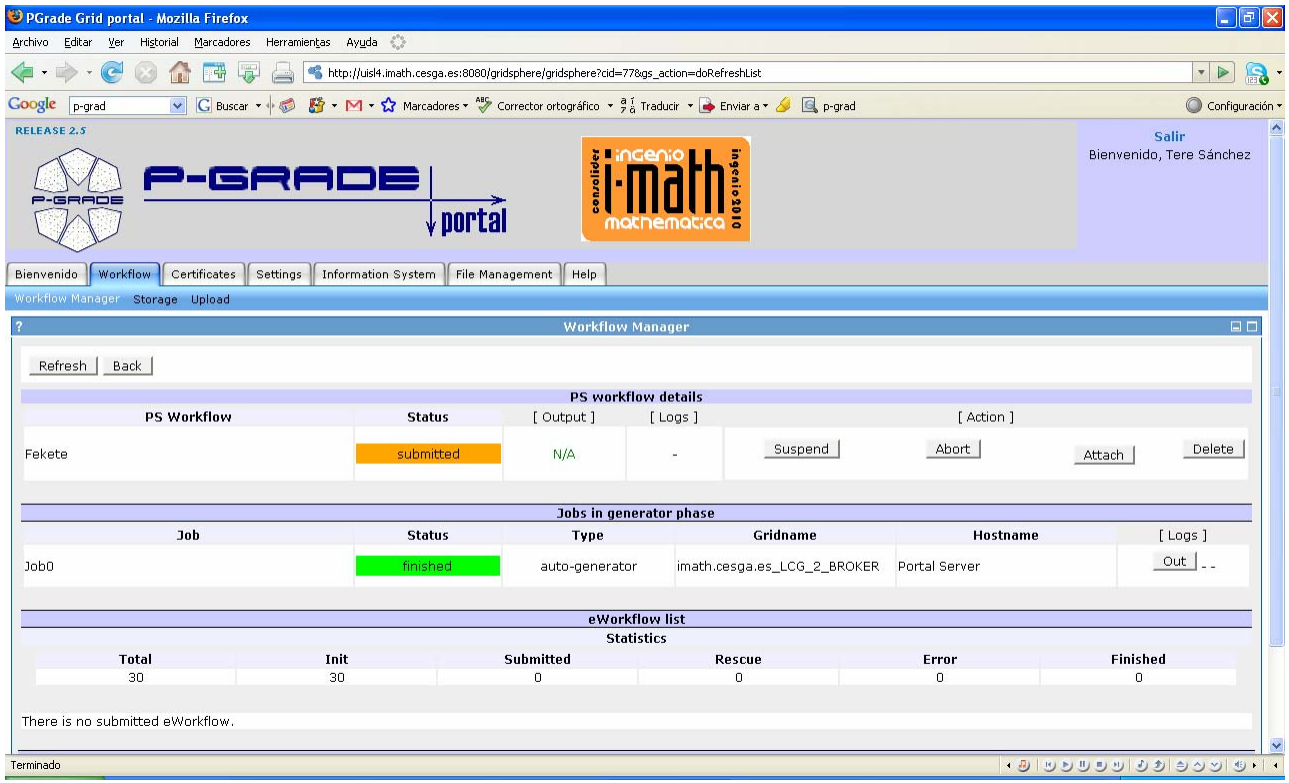
Case 3: Estimation of Fekete points



cpus: 30 Time: 210 min (in only a computer 4500 min)

Study cases

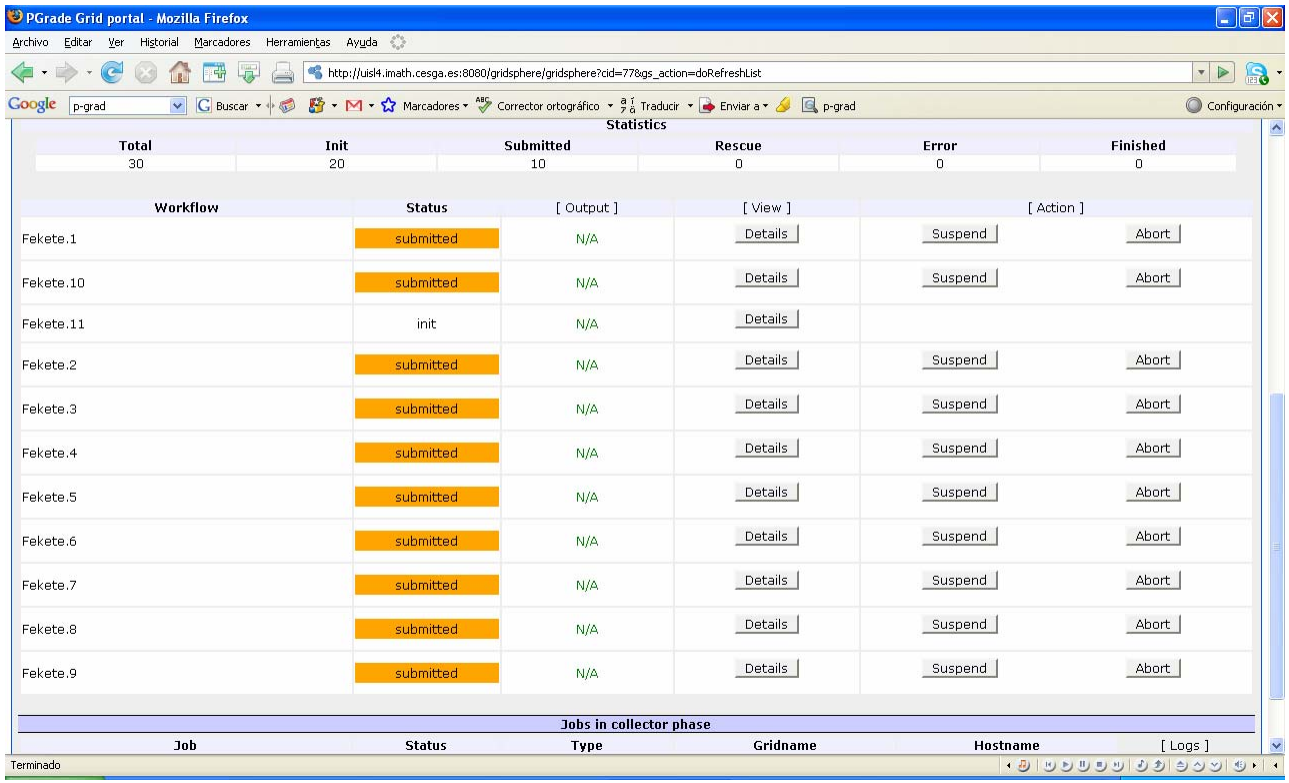
Case 3: Estimation of Fekete points



cpus: 30 Time: 210 min (in only a computer 4500 min)

Study cases

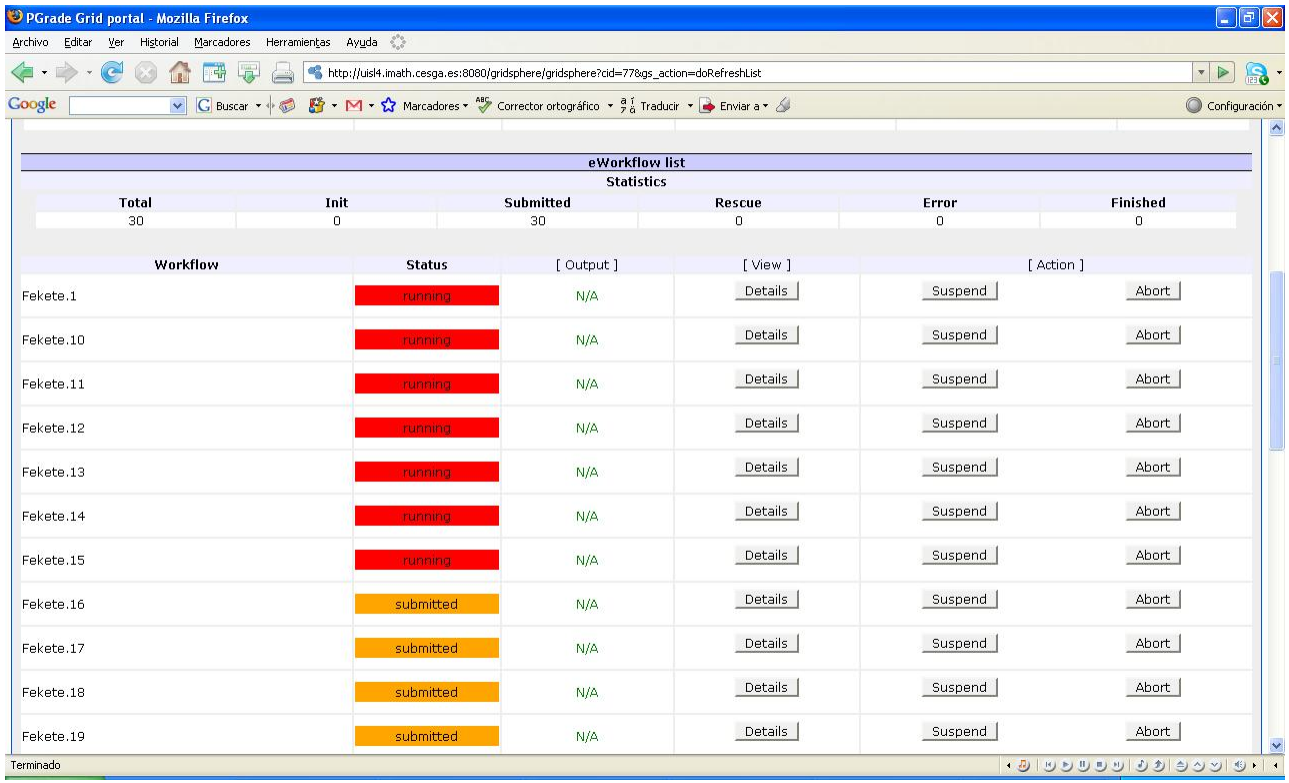
Case 3: Estimation of Fekete points



cpus: 30 Time: 210 min (in only a computer 4500 min)

Study cases

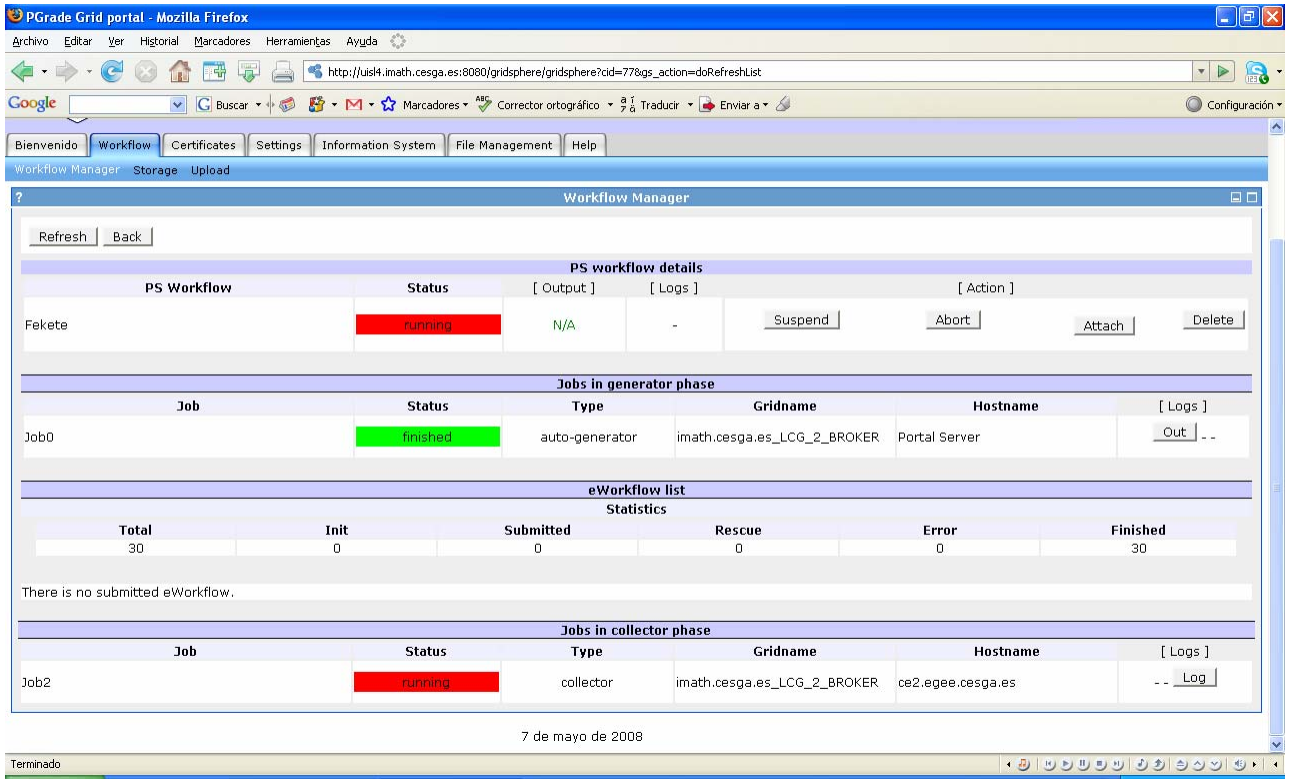
Case 3: Estimation of Fekete points



cpus: 30 Time: 210 min (in only a computer 4500 min)

Study cases

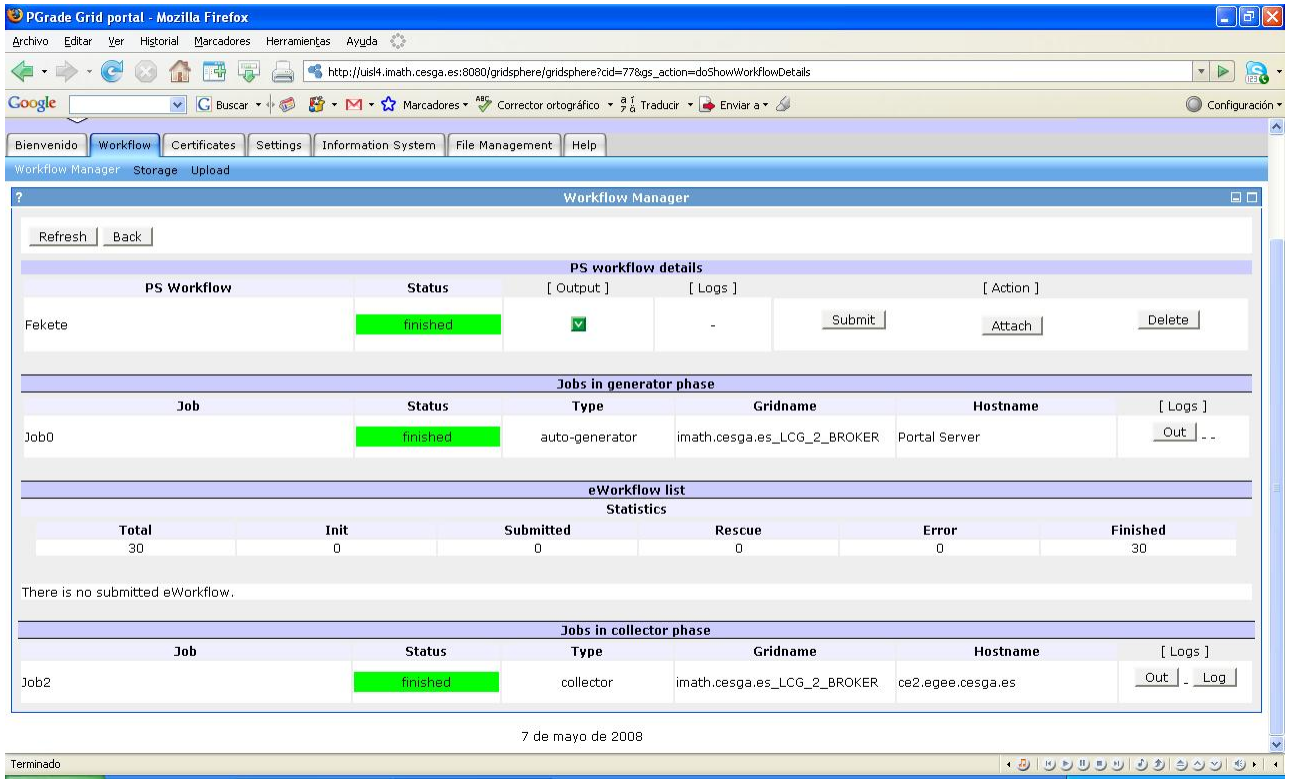
Case 3: Estimation of Fekete points



cpus: 30 Time: 210 min (in only a computer 4500 min)

Study cases

Case 3: Estimation of Fekete points



cpus: 30 Time: 210 min (in only a computer 4500 min)

Conclusions

- imath.cesga.es VO devoted exclusively to Mathematics
- 3 distributed sites, basing front-ends on XEN
- Pilot to prove the applicability of using GRID in mathematical research.
- Three different study cases
  - Statistics, climatology and fundamental maths.
- Make easy the usage of GRID: P-Grade Grid Portal
  - Simplicity, powerful workflow but some limitations
- imath.cesga.es is open to Mathematics research groups in Europe at large!!!

Thanks for your attention!

http://mathematica.nodo.cesga.es

http://imath.cesga.es

tsanchez@cesga.es