Sequential Gaussian Simulation
Application Porting

GISELA/EPIKH School for
Application Porting

Álvaro Parra
alvaro.parra@alges.cl
ALGES

- ALGES aims at developing tools and models for the characterization of geological phenomena
  - Environmental applications
  - Geological modeling
  - Mining / petroleum

- Goals
  - Prediction (estimation) of categorical and continuous variables
  - Uncertainty quantification for decision making
  - Multivariate modeling (estimation and simulation)
  - Software development and mathematical modeling:
    - Image analysis of geological, mining, and metallurgical data
    - Supercomputing for geostatistical modeling
    - Optimization and decision making

- AMTC – Advanced Mining Technology Center
- www.alges.cl
Mining Industry
Infrastructure

• Cluster (hosted at CEC datacenter)
  - IBM BladeCenter 8852 with 4 servers
  - Each server: 2 Intel Xeon Quad Core processors
  - Total of 32 cores
  - Fiber Optics communications
  - 2 TB Hard Drive for Storage

• Office space for 14 full time researchers

• Software licences:
  - Matlab
  - Isatis
  - Gemcom
  - Visual studio
  - Productivity software
Estimation vs. Simulation

Plans based on average assumptions are wrong on average

The drunken state of their average position is ALIVE

But the drunken average state is DEAD
Algorithm

• Geoestatistical Simulation
  • Variability estimation
  • Best reproduction of geological phenomena
• Sequential Simulation
• Implementations
  • Most used method in simulation of mineral resources in mining industry
  • Broadly known algorithm and present in many mining softwares
• Challenges
  • Enormous amount of data
  • Multiple realizations
Estimation and Variability Measure
SISIM Example (1)
SISIM Example (2)
Sequential Simulation

- Each simulated data becomes conditioning data
  - Sequential algorithm
  - No direct parallelization strategies
- Simulation → Multiple realizations (hundreds)
  - Each realization can be obtained Independently
Approach

- **Parametric Job**
  - Simulations are performed by Parametric Jobs
- **OpenMP**
  - Each Working Node performs multiple simulations using OpenMP
Pre-processing

Upload Samples
(samples.csv.tar.gz)

Upload APP
(gs.tar.gz)

Configuration

Algorithm parameters
(xml file)

Execution

Submit experiments
(specify a experiment name)

Retrieve realizations
(for a specified experiment)
Approach

• Each experiment is identified by a name
  • Identify the job id
  • Output data of each simulation is stored into a directory with the experiment name
  • Allow run multiple experiments, each one for multiple realizations

• A bash script (gs-grid) to perform all operations
  • Submit
  • Cancel
  • Status
  • Upload input data
  • Download input data
  • Output
Scritp: gs-grid

• Usage
  • gs-grid COMMAND [ARGS]

• List of Commands:
  • **submit** Submit a job
  • **status** Print the job status
  • **output** Retrieve the generated files of a job. Should receive the -N parameter with the number of simulated realizations. If --all argument is passed is retrieved the standard output and error for each realization.
  • **cancel** Cancel a job
  • **clean-se** Delete the output data on the storage
  • **upload-app** Upload the app (gs.tar.gz) to the storage element
  • **upload-input** Upload samples.csv.tar.gz to the storage element
  • **download-input** Download to the UI the files gs.tar.gz and samples.csv.tar.gz from the storage
Scritp: gs-grid

• **ARGS**
  
  • **--config** File with the simulation parameters. If not passed the default is simulation.xml
  
  • **--vo** Set the virtual organization. If not passed the default is prod.vo.eu-eela.eu
  
  • **--ce** Set the computing element. If not passed the default is prod-se-02.ct.infn.it
  
  • **--se** Set the storage element. If not passed the default is prod-se-02.ct.infn.it
  
  • **--name** The name of the experiment
  
  • **-N** Number of realizations. The default value is 1
  
  • **--all** Specify if retrieve the standard output and error
<simulation type='1' option='0' realizations='2' seed='123456'>
  <input filename='muestras.csv' nscores='1'>
    <columnPositions x='0' y='1' z='2' var='3' wt='1' />
    <trimmingLimits start='0' end='1e20' />
  </input>

  <backTransformation tailExtrapolationMinimum='0.1'
    tailExtrapolationMaximum='7' lowerTailOption='1' lowerTailParameter='0'
    upperTailOption='1' upperTailParameter='0' />

  <output filename='output/realization.csv'>
    <grid>
      <nodes x='40' y='60' z='13' />
      <start x='5' y='5' z='5' />
      <size x='10' y='10' z='10' />
    </grid>
  </output>

  <neighborSearchParameter minData="1" maxData="24" maxOctant="0" visitedNodes="24">
    <radii hmax="100" hmin="100" vert="100" />
    <angles azimuth="0" dip="0" plunge="0" />
  </neighborSearchParameter>

  <variographicModel nuggetEffect='0.05'>
    <structure type='2' sll='0.95' parameter=''>
      <ranges hmax="100" hmin="100" vert="100" />
    </structure>
  </variographicModel>
</simulation>
Upload binaries and Input data

- App (gs.tar.gz) including:
  - `usr/include/`, `usr/lib/`, `bin/`, `log4cplus.properties`
- Input data (tar.gz):
  - `samples.csv.tar.gz`
Submit an experiment

```
[valparaiso06@cluster30 ~]$ gs-grid.sh submit --name exp2 -N 100

Connecting to the service https://wms-eela.ceta-ciemat.es:7443/glite_wms_wmproxy_server

Warning - Unable to delegate the credential to the endpoint: https://wms-eela.ceta-ciemat.es:7443/glite_wms_wmproxy_server
Stack dump

Switching to next WMProxy Server...

Connecting to the service https://wms.eela.ufrj.br:7443/glite_wms_wmproxy_server

===================== glite-wms-job-submit Success ======================

The job has been successfully submitted to the WMProxy
Your job identifier is:
https://lb2.eela.ufrj.br:9000/f4mo_BRfHtn7Eixm7YmftA

The job identifier has been saved in the following file:
/home/tutor/valparaiso06/exp2.id

====================================================================
Status

```
[valparaiso06@cluster30 ~]$ gs-grid.sh status --name exp2

Status info for the Job : https://lb2.eela.ufrj.br:9000/qXOi2LG0BFVljdRuW4qkyQ
Current Status: Running
Status Reason: Job successfully submitted to Globus
Destination: grid012.ct.infn.it:2119/jobmanager-lcglslf-prod
Submitted: Wed Dec 8 12:57:32 2010 CLST

Status info for the Job : https://lb2.eela.ufrj.br:9000/rux6PtFESTD1KUGmsRxxYw
Current Status: Scheduled
Status Reason: Job successfully submitted to Globus
Destination: grid012.ct.infn.it:2119/jobmanager-lcglslf-prod
Submitted: Wed Dec 8 12:57:32 2010 CLST
```
[valparaiso06@cluster30 ~]$ gs-grid.sh output --name exp2 -N 100

Connecting to the service https://wms.eela.ufrj.br:7443/glite_wms_wmproxy_server

=============================================

JOB GET OUTPUT OUTCOME

Output sandbox files for the DAG/Collection:
https://lb2.eela.ufrj.br:9000/f4mo_BRfHtn7Eixm7YmftA
have been successfully retrieved and stored in the directory:
/home/tutor-valparaiso06/valparaiso06_f4mo_BRfHtn7Eixm7YmftA

=============================================

Retrieving gs_output_data0.tar.gz to download/
Retrieving gs_output_data1.tar.gz to download/
Retrieving gs_output_data2.tar.gz to download/
Retrieving gs_output_data3.tar.gz to download/
Realization Example
Conclusions

- The application was migrated successfully
- The developed script permit run simulations easily
- Complicated compilation → Develop a better makefile
- The School open new possibilities for ALGES
- Desired Improvements
  - Have development environment → Fast develop
  - Improve the wms response
  - Access to more resources
Future Work

- Benchmark testing
- Perform simulations from an hybrid or web software
- Migrate other more resource consuming simulation algorithms
- ALGES Grid Integration
Sequential Gaussian Simulation
Application Porting

GEISELA/EPIKH School for
Application Porting

Álvaro Parra
alvaro.parra@alges.cl
Www.alges.cl
Using one node of the ALGES's Cluster