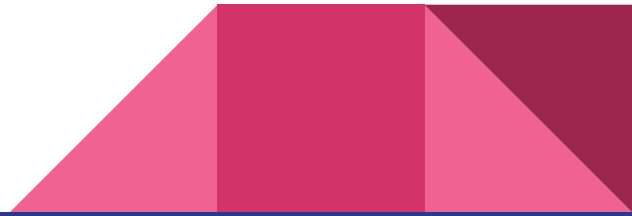
SGeMS HPC

Péricles Lopes Machado
PhD candidate - LPM/UFRGS
Advisor: Prof. PhD João Felipe C. L. Costa

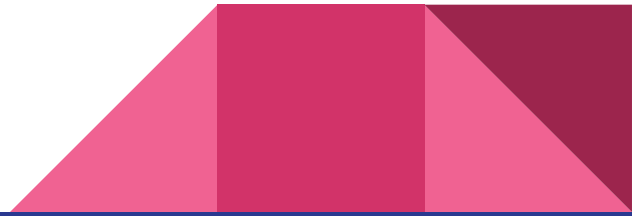
Why High Performance Computing (HPC)?

- Time = $$$.
- Scalability is important to improve the quality of results.
- High resolution models becomes less time expensive.
- High precision algorithms demand many computational resources.
- It’s cheap to implement a computational cluster in an industrial environment or in a laboratory.
- New uncertainty methodologies or stochastics decision algorithms demand a big number of simulations or estimations.



What has been implemented?

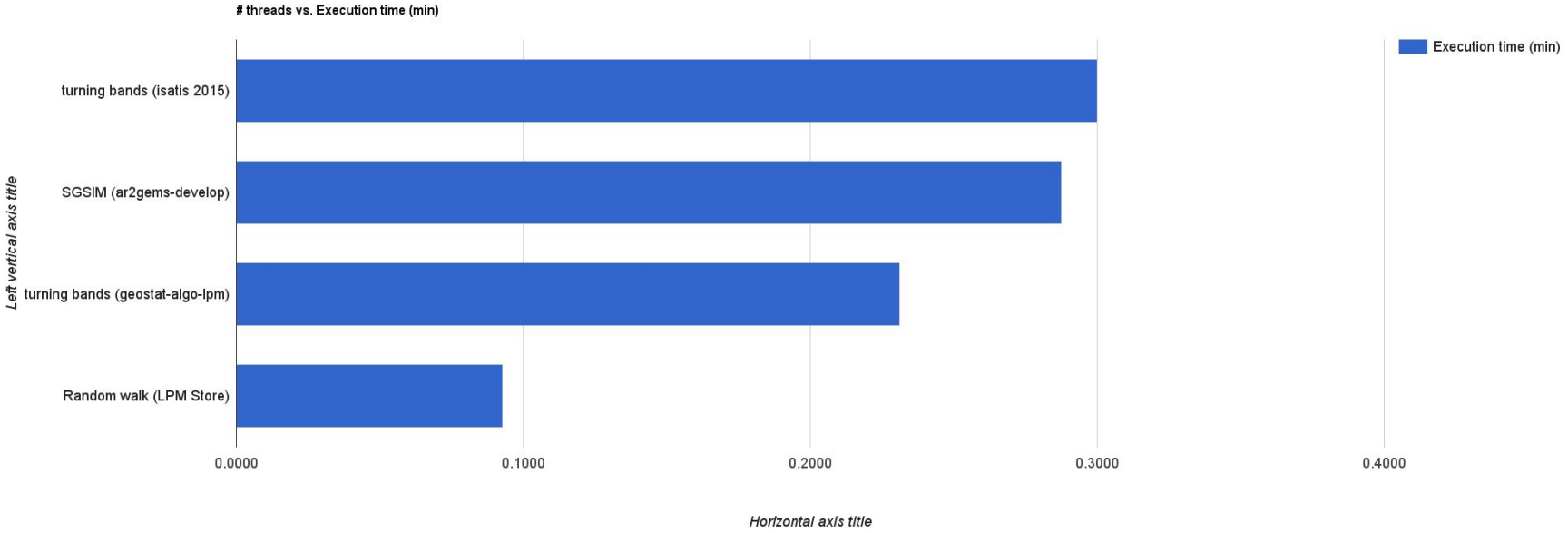
- New simulation algorithms developed to use efficiently the power of HPC: Random Walk and PGP-SS (Parallel Grid Partition to Sequential Simulations)
- A new version of SGeMS. Now, SGeMS has a task manager to handle many operations running simultaneously.
- Support to python plugins in SGeMS. Now, it’s easier (and faster) to develop new algorithms using the SGeMS back-end.
- A computational grid using Hydra (our main server) and the desktop available in our laboratory.



What has been implemented?

- Improved the Turning Bands performance in SGeMS.
- Developed a set of algorithms and visualization tools (LPM Store)

Some results...

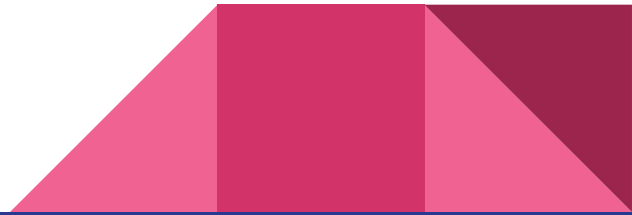


Some results...

Algorithm	Execution time (min)	Speed-up
turning bands (isatis 2015)	0.3000	1
SGSIM (ar2gems-develop)	0.2876	1.04
turning bands (geostat-algo-lpm)	0.1971	1.52
Random walk (LPM Store)	0.0928	3.23

Work in progress

- SGeMS HPC: a REST API to provide easy access to HPC resources, geostatistics primitives (kriging, simulation, estimation, variograms, etc...). With this API, it will be possible to create interfaces to other softwares like Data Mining, Isatis, GSLib or other environments (mobile, web, clusters). The main objective is to create a standard way to access SGeMS resources to develop distributed algorithms (algorithm running in many separated machines).

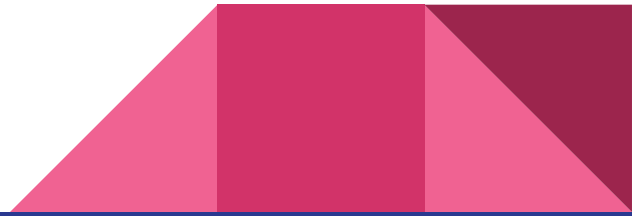


Work in progress

- SGeMS API: a subset of SGeMS HPC. This API is being developed to provide a set of geostatistics primitives and data handling methods to stand-alone python scripts and C++ applications.
- SGeMS Data Sources: A set of classes to provide access to different production Database Managers or Key-Values Servers like MySQL, SQLite, Oracle, REDIS, PostGreSQL, etc. This API is important to implement data parallelism. We can use this API to simulate/estimate huge dataset containing billions of nodes (very high resolution models). Data parallelism is as important as process parallelism.

Work in progress

- LPM Store HPC: implement a set of state-of-art High Performance estimation and simulation algorithms using SGeMS HPC, SGeMS API and SGeMS Data Sources. This is the final product of my work.



Publications

- Rasera, L. G; Machado, P. L.; Costa, J.F.C.L. *A conflict-free, path-level parallelization approach for sequential simulation algorithms*. Computers & Geosciences, vol. 80, July 2015, pg. 49-61
- Caixeta, R.M.; Costa, J.F.C.L.; Machado, P. L.; Ribeiro, D. T.; Dias, D. H.; Pires, R.O. *How precise and accurate is the space of uncertainty derived from simulations?* In: APCOM 2015, Fairbanks AK. Proceedings of the 37th APCOM, 2015.