A Climate Application using Grid Environment Solutions

Roberto P. Souto, Eduardo R. Rodrigues,
Philippe O. A. Navaux, Nicolas Maillard
Universidade Federal do Rio Grande do Sul
Instituto de Informática
{rpsouto,errodrigues,navaux,nicolas}@inf.ufrgs.br

Abstract

Simulations of the atmosphere, usually demand high processing power and large storage resources. In this context, we present two purposes of grid environment in order to perform climate modeling for a long-term period: the GBRAMS and RECLIRS projects. In the former, a grid infrastructure was built to manage a 10-year period split into 1-year integration jobs, performed in three different regions of Brazil, and also for three set o boundary conditions (three initial conditions). Therefore, there is up to nine independent jobs, submitted to three clusters spread over Brazil. Moreover, three distinct middlewares, Globus Toolkit, OurGrid and OAR/CIGRI, were compared in their ability to manage these jobs, and results on the usage of each node of the grid are provided. The central contribution are how to use grid computing to speed-up climatology generation and the middleware impact on this enterprise. In its turn, in RECLIRS project, instead of integrating one job for each simulated year, one integrates 12 three-month periods for each year, which is usually employed in numerical prediction. A three-month job starts every month of the year, and such executions are repeated for every year. In spite of demanding more computational power, the jobs in this framework are totally independent, allowing for better load balancing in the grid.

1. Introduction

Better weather and climate forecast are currently one of the principal objectives of governments around the world. Simulations of atmosphere, however, usually demand high processing power and large storage resources. In this context, this work presents two projects that apply grid computing to speed up the generation of a long term monthly average of atmospheric forecasts produced by a specific model, *i.e.*, the model climatology

The first one is the GBRAMS project [7, 2, 9], an effort of research groups of four academic and research intitutes of Brazil: CPTEC/INPE, LAC/INPE, II/UFRGS, IAG/USP, and also HP/Brasil and Somar Meteorologia, two partners from industry. This was a project supported by FINEP.

During the project, a 10-year climatology was obtained for different regions in Brazil, using the mesoscale - or regional - model BRAMS (Brazilian Regional Atmospheric Modeling System) [8, 10, 1]. RAMS, developed by the Atmospheric Science Department at the Colorado State University, is a multipurpose numerical prediction model designed to simulate atmospheric circulations spanning from hemispheric scale down to large eddy simulations of the planetary boundary layer. In its turn, BRAMS differs from RAMS due to the development of computational modules more suitable to tropical atmospheres.

As any mesoscale forecast model, BRAMS needs boundary conditions during the integration period, which are previously generated by the CPTEC's global model, with lower spatial resolution. The number of independent jobs which can be assigned to any grid site, is as many as the number of initial conditions provided by the global model, multiplied by the number of regions. This is the only source of job independence, since each year depends on the last one to start its job.

RECLIRS is a new project, similar to GBRAMS in many aspects, but with peculiar features. This is an initiative of the following institutions: CRSPE/INPE, II/UFRGS, IPH/UFRG, IF/UFSM, Inf/UFSM, MET/UFPEL and FEPAGRO.

As opposed to GBRAMS, due to the nature of the adopted method of climatology, jobs in the RECLIRS project are fully independent. Instead of integrating one job for each simulated year, one integrates 12 three-month periods for each year, which is usually employed in numerical prediction. A three-month job starts every month of the year, and such executions are repeated for every year.

In this work, details about similarities and differences between these two projects are presented in Section 2. An presention concerning the grid architecture and a sample of products generated by the grid infrastructure, is in Section 3. It is focused in Section 4, analysis previously showed in [9], regarding to Grid middleware comparisions, and the climatology reached in GBRAMS project.

2. GBRAMS and RECLIRS: a brief comparison

Both projects have the same objective, that is, to obtain a 10-year climatology. However, the main difference regards to method chosen to reach it. In GBRAMS was used the methodology in production at International Research Institute for Climate and Society (IRI), where the whole long-term period is performed continuously.

For the RECLIRS project, it was decided for the European Center Medium-Range Weather Forecasts (ECMWF) climatology scheme. In this case, it is obtained an average for twelve 3-month model integration period, one for each month of the year. The advantage of this strategy, is the fully independence of all jobs, where each job means a 3-month period performed.

Figure 1 shows examples for both methods, where the Feb-Mar-Apr period climatology is performed by the ECMWF scheme. It is important to note that the first month (January in ECMWF, and December 1991 in IRI) is used only as a *spin-up* period, concerning to necessary numerical model adjustments in the beginning of integration.

In the GBRAMS project, the 10-year period was split into ten 1-year consecutive integrations. This scheme was possible since BRAMS provides a checkpoint/restart mechanism that allows one to interrupt and resume the integration at every simulated year. This is done for 3 different spatial sub-domains of Brazil, showed in Figure 2(a).

Since a job is created for each region and initial condition, each year of integrated time demands 3 simulations, one per initial condition, for each of the 3 sub-domains, thus providing 9 independent 1-year BRAMS parallel jobs.

In RECLIRS project, only in the shoutern region of Brazil, showed in Figure 2(b), will be performed the model climatology. However, as it use ECMWF scheme, all jobs are fully independent each other.

3. Grid Architecture

A grid infrastructure 3 containing three clusters geographically distributed in Brazil. A web portal was designed to submit a n-year climatology simulation to be executed. This submission requires that the user fills the simulation parameters in the corresponding RAMSIN file (RAMS input file), such as the time extension of the simulation and

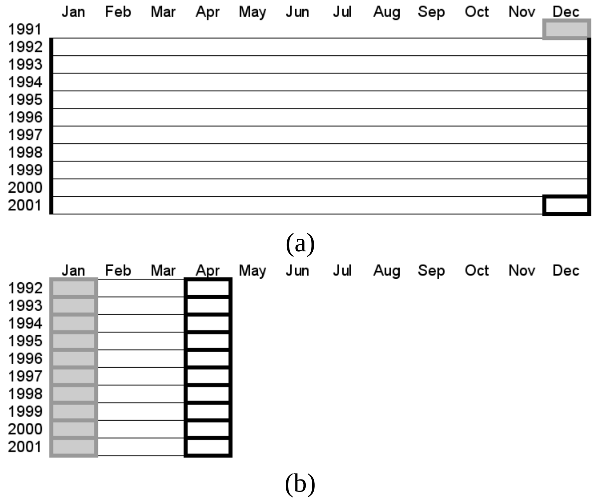


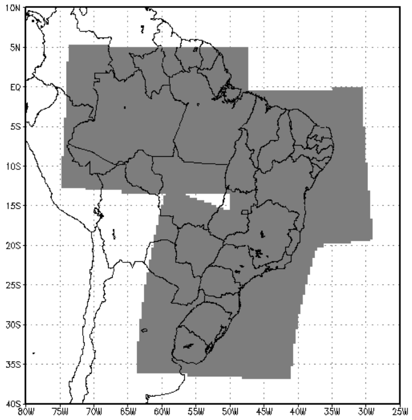
Figure 1. Examples of jobs in both methods employd in GBRAMS and RECLIRS project, for the 10-year period using :
(a) IRI climatology, where the whole period is split into 1-year jobs, and
(b) ECMWF climatology, where the Feb-Mar-Apr climatology period is performed.

the chosen domain. Each new job is inserted into a database, from where it is taken later for execution. A scheduler is in charge of taking the ready jobs from the database and executing them in a grid computer. After the execution, the results of post processed analysis are available for viewing in the web portal, as such the monthly average temperature and accumulated precipitation (Figure 4).

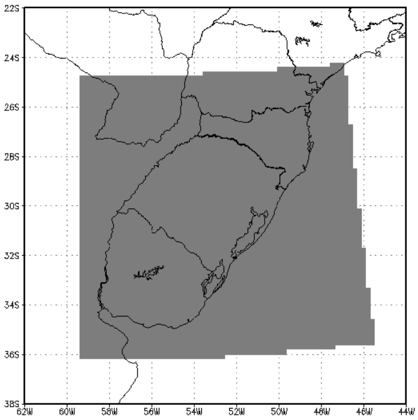
4. Analysis of Grid Middleware and GBRAMS Climatology Results

Three grid computing platforms were used in the GBRAMS project: the Globus Toolkit [5, 6], OAR/CIGRI [3] and OurGrid [4]. These grid middleware have been tested on their capacity to solve the meteorological application detailed before, being evaluated in aspects such as the management of large amounts of data and job scheduling.

One can expect that most performance variations in the computation of the climatology should be explained by differences of implementation in the three employed middleware. In a recent work [9] it was made an analysis, in order to test this assumption for the application. It was also discussed the importance of the job distribution observed for the grid computing performance reached in the GBRAMS project.



(a)



(b)

Figure 2.
(a) Three regions of Brazil where was obtained model climatology in GBRAMS project, and
(b) the southern region of Brazil where will be performed model climatology in RECLIRS project.

An evaluation of the climatology obtained using the grid, consists in verifying whether the bias due to the BRAMS model has been correctly removed from the climate forecast.

4.1. Grid Middleware comparison

For each initial condition, a 10-year simulation was integrated for the years from 1989 up to 1998. This whole period is then split into shorter 1-year jobs, and executions are distributed among the three different middleware. Additionally, a period of 2 months (November and December) immediately preceding the middleware set of jobs must be integrated. This is because numerical models require a *spin up* period of simulation. These additional results were not used in the climatology calculation. Thus, for a complete 10-year BRAMS simulation in 3 sub-domains in Brazil, and with 3 ensemble members, we have integrated 1134 months.

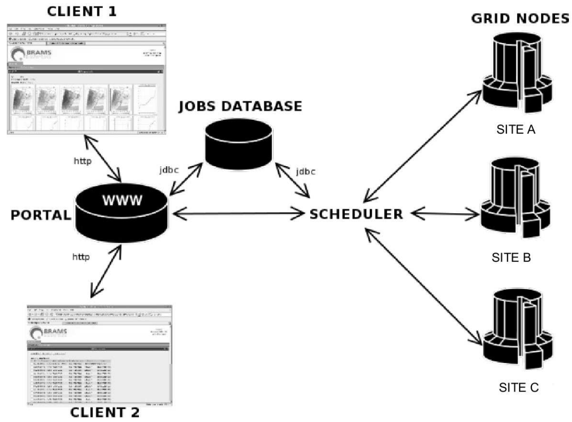


Figure 3. Grid architecture implemented in GBRAMS and RECLIRS.

	T_p – cluster time (days)	Integrated months (m)	T_p/m (hh:mm)
Globus	43.7	414	02:32
OurGrid	33.7	342	02:30
CIGRI	39.6	378	02:30
	117.0	1134	

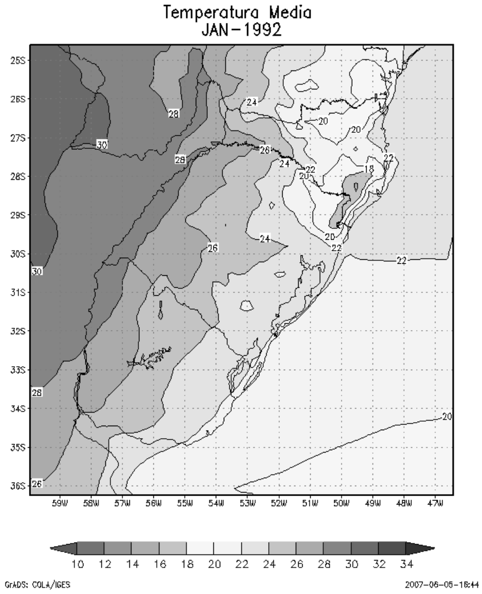
Table 1. Accumulated cluster time (T_p)

It is defined here as accumulated middleware cluster time (T_p), the overall runtime of the jobs submitted by the middleware to grid nodes. Since computations are eventually overlapped in time (up to the number of available grid nodes), the time of grid usage demanded to perform the whole set of jobs is defined here as grid elapsed time (T_g), wich is naturally shorter than T_p. As shown in Table 1, 117 days of accumulated cluster time were spent during jobs execution submitted by all middlewares. Regarding to the grid elapsed time, it was almost half of the overall accumulated cluster time, with a total of 61.5 days of grid usage.

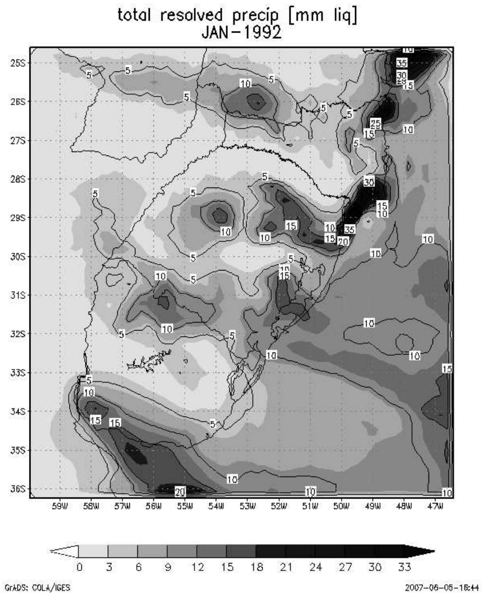
Concerning the monthly integration average times, there are no differences between the systems. This indicates that the execution time of a BRAMS job is middleware-independent. However, the average grid elapsed time per integrated month is not uniform for the

	T_g – grid elapsed time (days)	Integrated months (m)	T_g/m (hh:mm)
Globus	23.9	414	01:23
OurGrid	16.3	342	01:08
CIGRI	21.3	378	01:20
	61.5	1134	

Table 2. Elapsed time (T_g) of all jobs



(a)



(b)

Figure 4. Results of January in 1992: (a) average temperature in Celsius degrees. (b) accumulated precipitation in mm.

three middleware. This value for OurGrid is the smallest one, and this fact suggests a most effective cluster usage. This assumption is confirmed in Figure 5, which shows the percentage of cluster usage during each middleware execution, given by the ratio $T_p^{(s)}/T_g$, where index $^{(s)}$ indicates accumulated time of a single cluster. For OurGrid job executions, clusters always had a better ratio between the accumulated single cluster time ($T_p^{(s)}$) and the grid elapsed time (T_g). As a consequence of this better cluster usage, simultaneous jobs executions were more frequent in OurGrid.

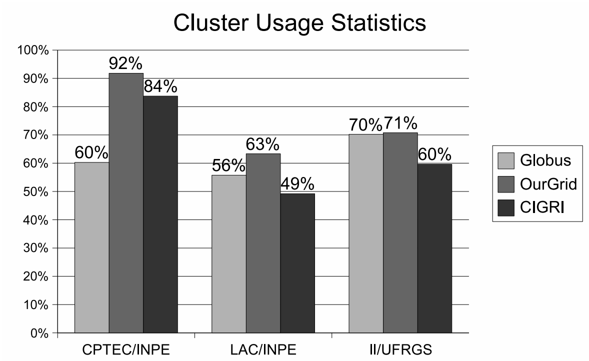


Figure 5. Cluster usage in each middleware. The percentage means the ratio between the accumulated single cluster time ($T_p^{(s)}$) and the grid elapsed time (T_g)

In Table 3, the first three columns inform the portion of accumulated grid elapsed time due to simultaneous job execution in 3 and 2 clusters, and also jobs executed in a single cluster. The average number of clusters used during grid execution is given by the fourth column for each middleware, and these values are compatible with third column of Table 2. Job execution overlapping occurred in three clusters during 44% of elapsed time in OurGrid. In a comparison with Globus and CIGRI this percentage decays to 26% and 31%, respectively.

The preceding performance results might indicate that

	3	2	1	avg. use
Globus	0.26	0.34	0.40	1.86
OurGrid	0.44	0.37	0.19	2.25
CIGRI	0.31	0.31	0.38	1.93

Table 3. Cluster usage

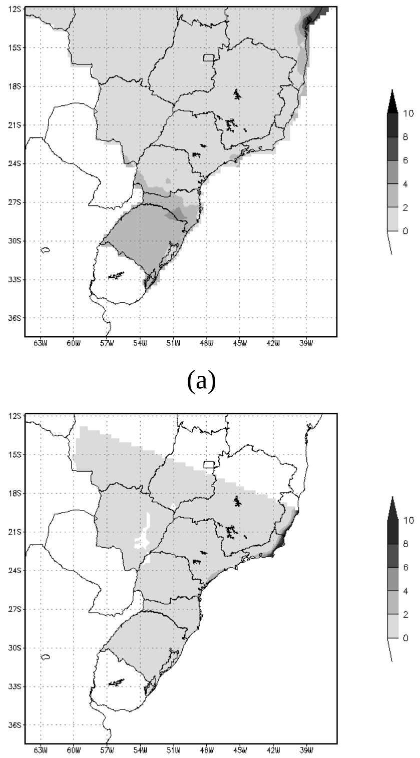
OurGrid is most suitable to schedule jobs in this meteorological application. However, we believe that additional tests should be carried out in order to confirm this apparent tendency. Factors like machine maintenance and the evaluation of job results by a meteorologist have affected the distribution of the jobs. The second factor interrupts some sequences of dependent submitted job executions, since the corresponding previous job results had not been approved in the web portal. We need job executions with more simultaneous use of nodes, enough to stress the usage of grid resources, in order to test whether, in such a scenario, the better grid performance of OurGrid would be reproduced.

4.2. Climatology and Climate Forecast Results

Besides testing the proposal of performing a mesoscale climatology by using grid computing, another key reason to obtain it, is to remove the model bias in the climate forecast done with BRAMS, in order to predict anomalies in meteorological properties in the subsequent season.

Figures 6 and 7 show the effect of removing the model bias in the climate forecast, for the period of June–August 2006, for the south/south-east region of Brazil. The first two graphs show the daily precipitation average (in mm/day), as obtained from observed atmospheric data (Figure 6(a)) and simulated from the BRAMS climate forecast (Figure 6(b)). As can be seen, they do not match each other. Indeed, the clearer gray tones in the southern part of the map indicate that daily precipitation is underestimated in this region. In turn the darker gray levels in the eastern coast (the upper right region of this sub-domain) indicate that the daily precipitation is overestimated. The upper-right corner of Figure 6(b) should not be considered: the straight diagonal line denotes the limit of the integration domain.

However, when the BRAMS model bias is removed from the climate forecast, one obtains a graph of anomalies (Figure 7(b)) that is generally better estimated in most parts of the region, when compared now with the similar observed data graph of anomalies (Figure 7(a)). Some precipitation anomalies still tend to be overestimated in the northeast coast of the map.



(a)

(b)

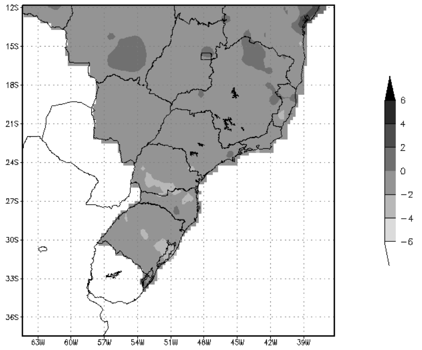
Figure 6. (a) Observed climate; (b) BRAMS climate forecast

4.3. Conclusions

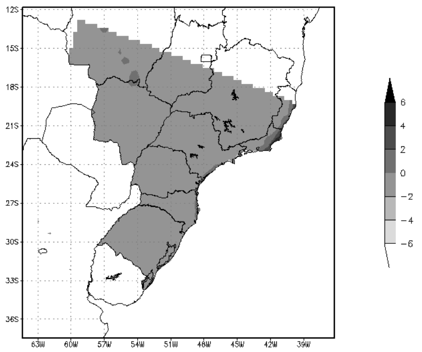
This work was intend to make a review of the grid computing projects, developed by the parallel and distributed computing group at Informatics Institute of UFRGS last two years. Especially talking for the Climate forecast application. The results obtained up to now in GBRAMS project, associated with the increasing interest in the world to better understand the changes in climate, has encouraged us to follow the research. In GBRAMS project was built the basis to be employed in RECLIRS project, that is more suitable to have benefits, like gains of performance and scalability, with the grid computing paradigm.

Acknowledgements

GBRAMS and RECLIRS project has been supported by FINEP.



(a)



(b)

Figure 7. (a) Observed climate anomalies; (b) BRAMS climate anomalies forecast

References

[1] Brazilian Regional Atmospheric Modeling System (BRAMS), 2004. Available in http://www.cptec.inpe.br/brams.

[2] H. CamposVelho, A. Preto, S. Stephany, E. Rodrigues, J. Panetta, E. Almeida, R. Souto, P. Navaux, T. Diverio, N. Maillard, and P. S. Dias. Grid computing for mesoscale climatology: experimental comparison of three platforms. In *Proceedings of the 7th Intl Meeting on HPC for Computational Science (VECPAR'06)*, Rio de Janeiro, Brazil, July 2006. UFRJ.

[3] N. Capit, G. D. Costa, Y. Georgiou, G. Huard, C. Martin, G. M. P. Neyron, and O. Richard. A batch scheduler with high level components. In *Proceedings of the Intl Symposium on Cluster Computing and the Grid (CCGRID 2005)*, pages 776–783, Cardiff, UK, May 2005. ACM/IEEE, IEEE Computer Society.

[4] W. Cirne, F. Brasileiro, D. Paranhos, L. Costa, E. Santos-Neto, and C. Osthoff. Building a user-level grid for bag-of-tasks applications. In L. T. Yang and M. Guo, editors, *HPC: Paradigm and Infrastructure*, chapter 37, page 816pp. John Wiley Sons, nov 2005.

[5] I. Foster. Globus toolkit version 4: Software for service-oriented systems. In H. Jin, D. Reed, and W. Jiang, editors, *Proceedings of the IFIP Intl Conf on Network and Parallel Computing*, volume 3779 of *Lecture Notes in Computer Science*, pages 2–13, Beijing, china, Nov. 2005. Springer.

[6] I. Foster and C. Kesselman. *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann, San Fransisco, 1999.

[7] J. Panetta, E. S. Almeida, E. R. Rodrigues, H. F. de Campos Velho, S. Stephany, A. J. Preto, P. O. A. Navaux, N. Maillard, R. P. Souto, T. A. Diverio, and P. L. S. Dias. Climatology on grid: The g-brams project. In *Proceedings of the 18th Symposium on Computer Architecture and HPC (SBAC-PAD). Workshop ParGov: Applications of Parallel Computing to e-Gov*, Ouro Preto, Brazil, Oct. 2006.

[8] R. Pielke, W. Cotton, R. Walko, C. Tremback, W. Lyons, L. Grasso, M. Nicholls, M. Moran, D. Wesley, T. Lee, and J. Copeland. A comprehensive meteorological modeling system – rams. *Meteorology and Atmospheric Physics*, 49(1-4):69–91, 1992.

[9] R. Souto, R. Avila, P. Navaux, M. Py, N. Maillard, T. Diverio, H. de Campos Velho, S. Stephany, A. Preto, J. Panetta, E. Rodrigues, E. Almeida, P. S. Dias, and A. Gandu. Processing mesoscale climatology in a grid environment. In *Seventh IEEE International Symposium on Cluster Computing and the Grid (CCGrid-2007)*, Rio de Janeiro, Brazil, May 2007. IEEE.

[10] R. Walko, C. Tremback, and R. Hertenstein. *RAMS - The Regional Atmospheric Modeling System - Version 3b - User?s Guide*. ASTER Division, Fort Collins, CO, 1995.