Middle-East Journal of Scientific Research 19 (6): 769-774, 2014
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2014.19.6.12533

A Test Case on the Usage of Grid Infrastructure in Regional Climate Simulation

[1]Mohamed ElKharrim , [2]Lahcen Bahi, [1]Abdellah Mokssit,
[3]Abderrahman ElKharrim and [4]Redouane Merrouch

[1]Direction de la Météorologie Nationale, Casablanca, Morocco
[2]Mohammed V University, Ecole Mohammadia d'Ingénieurs, Rabat, Morocco
[3]Mohammed Premier University, Oujda, Morocco
[4]Centre National pour la Recherche Scientifique et Technique, Rabat, Morocco

Abstract: Computing Grids are emerging as a new paradigm for sharing and aggregation of geographically distributed resources for solving large-scale compute and data intensive problems in science. Climate simulation is very time consuming and thus it's one of the important fields that can be addressed by Grids. The aim of this paper is to describe the implementation of the Regional Climate Model (RegCM4.3) in the Moroccan Computing Grid Infrastructure MaGrid and the tests performed using MPI parallel jobs. Also the computational efficiency of RegCM4.3 simulations on the Grid is assessed using different domain resolutions and CPUs number.

Key words: Grid computing · Regional climate model · RegCM4.3 · MPI parallel jobs

INTRODUCTION

The effects of climate change will mostly be felt on local to regional scales. To develop better forecast skill in regional climate change, an integrated multi-scale modeling capability (i.e., a pair of global and regional climate models) becomes crucially important in understanding and preparing for the impacts of climate change on the temporal and spatial scales that are critical to all countries and nation's future environmental quality and economical prosperity. Thus the interest in regional climate modeling has been steadily increasing in the last two decades and a number of regional climate models (RCMs) has been developed with a wide user base. Among them there is the RegCM (Regional Climate Model). The RegCM system is a community model and in particular it is designed for use by a varied community composed by scientists in industrialized countries as well as developing nations [1]. Regional Climate Simulation model needs a High Performance Computing (HPC) platform to reduce running time. Grid becomes an ideal solution when no HPC resources available especially for testing and offers a great amount of cheap and available resources. In this work, the process of porting the RegCM Model to the Grid is described and results of a test case simulation using advanced MPI jobs are presented. The model release used in this work is the RegCM4.3.4 and initial runs were completed using a 30km and 60km grid spacing forced by ERA-Interim Reanalysis.

Regional Climate Model (REGCM4.3): The RegCM is the first limited area model developed for long term regional climate simulation originally developed in the late 1980s, with a wide base of model users. It has participated to numerous regional model inter-comparison projects and it has been applied by a large community for a wide range of regional climate studies, from process studies to paleo-climate and future climate projections [2], [3].The RegCM is supported through the Regional Climate research NETwork (RegCNET) a widespread network of scientists coordinated by the Earth System Physics section of the Abdus Salam International Centre for Theoretical Physics (ICTP). The mission of the RegCNET is to promote climate research and ameliorate problems of scientific isolation in economically developing nations by fostering research partnerships, developing collaborative projects and providing an internationally recognized scientific and educational research forum.

Since the release of RegCM3, the model has undergone a substantial evolution both in terms of software code and physics representations and this has led to the development of a fourth version of the model

Corresponding Author: Abderrahman ElKharrim, Mohammed Premier University, Oujda, Morocco.

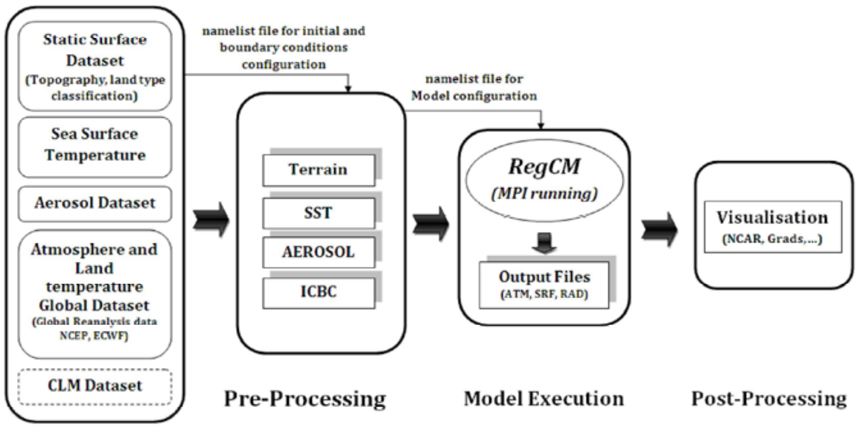

Fig. 1: RegCM4.3 workflow

RegCM4 that shows maturity by incorporating increasingly comprehensive physics packages and interactively coupled components of the climate system (chemistry/aerosol, ocean, lake, biosphere) and an improved performance compared to previous versions, although further testing by the user community is needed to fully explore its sensitivities and range of applications.

The RegCM Modeling System Has Four Components: Terrain, ICBC, RegCM and Postprocessor. Terrain and ICBC are the two components of RegCM preprocessor. Terrestrial variables (including elevation, landuse and sea surface temperature) and three-dimensional isobaric meteorological data are horizontally interpolated from a latitude longitude mesh to a high-resolution domain on a Rotated (and Normal) Mercator, Lambert Conformal, or Polar Stereographic projection. Vertical interpolation from pressure levels to the ó coordinate system of RegCM is also performed. ó surfaces near the ground closely follow the terrain and the higher-level ó surfaces tend to approximate isobaric surfaces. Since the vertical and horizontal resolution and domain size can vary, the modeling package programs employ parameterized dimensions requiring a variable amount of core memory and the requisite hard-disk storage amount is varied accordingly [4].

As regards physical parameterization, RegCM4 uses the radiation scheme of the NCAR CCM3, which is described in Kiehl *et al*. (1996) and use by default the BATS (Biosphere-Atmosphere Transfer Scheme) as land surface package designed to describe the role of vegetation and interactive soil moisture in modifying the surface-atmosphere exchanges of momentum, energy and water vapor (Dickinson *et al*. 1993). RegCM4 uses also (optional) the Community Land Model (CLM; Oleson et al. 2008) which is the land surface model developed by the National Center of Atmospheric Research (NCAR) as part of the Community Climate System Model (CCSM), (Collins et al. 2006). CLM version 3.5 was coupled to RegCM for a more detailed land surface description option. Complete reference to the model and downloads are available here [4].

Grid Technology: The term Grid is referring to a distributed computing infrastructure, able to provide resources based on the needs of each client [5]. Grid computing is the collection of computer resources from multiple locations to reach a common goal. The grid can be thought of as a distributed system with non-interactive workloads that involve a large number of files. What distinguishes grid computing from conventional high performance computing systems such as cluster computing is that grids tend to be more loosely coupled, heterogeneous and geographically dispersed. The requirements of grid computing technology are: security, fault tolerance, global name spaces, scalability, adapting heterogeneity, persistence and non persistence, extensibility, complexity management and autonomy.

Grid technology offers several advantages for the users:

- Grid computing saves financial resources both in capital and operating costs. This positive aspect is achieved by using all the computing resources of all components of the grid.
- Grid provides access to an enormous amount of storage space and computing resources difficult to reach by a single institution.

- Users take advantage of resources not fully used. In general, clusters are used just a few hours per day or during some months in the year.
- Grid technology provides security mechanisms that manage the access to shared resources. System administrators find Grid technology helpful because they can rely on its security mechanisms to grant access to users. [6]
- Users can access to geographically distributed resources, this allow working with data or computing resources of other institutions.

In order to provide users access to these distributed resources, Grid technology uses a service called *middleware* which is a system software between applications and operating system that provide services to application (Discovery, storage, execution…) and a standardized interfaces to services and hide heterogeneous of the Grid environment.

Grid Requirements: The requirements of grid computing technology are: security, fault tolerance, global name spaces, scalability, adapting heterogeneity, persistence and non persistence, extensibility, complexity management and autonomy. The most important requirements for a successful climate Grid application are [6]:

Failure Awareness: The application has to foresee all the possible sources of failure being able to face them or at least detect them and act in consequence.

Checkpointing for Restart: In case of failure, due to the computational cost of climate applications, one would want to restart the simulation in a different working site from the point it was interrupted.

Monitoring: Since climate simulations last for a long time, the user requires to know the current status of the experiment and their associated simulations: which percentage of the experiment is complete, whether there are simulations running, which time step is being calculated by a simulation, which data sets have been produced and in which storage elements are they, which is the last checkpointing/restarting point.

Data and Metadata Storage: The goal of the climate model experiments is the generation of (large amounts of) simulated climatic information. This information needs to be post-processed and analyzed by the different tools used by the climate researcher. Therefore, the data has to be easily accessed by users. A data and metadata management system has to be developed to handle all the information generated.

RegCM4 Grid Implementation: MaGrid (http://magrid.ma/) is the Moroccan Computing Grid Initiative that was established and hosted by the end of the year 2006 at the National Center for Scientific and Technical Research (CNRST). It uses the Moroccan Academic Wide Area Network (MARWAN) (http://marwan.ma/) which is the National Research and Educative Network (NREN) that connects it to GÉANT, the pan-European data network dedicated to the research and education community. The main goal behind building MaGrid was to provide the Moroccan research communities with an operational infrastructure offering computing power and storage capacity to host users' jobs and data and also to encourage collaborations and openings on the global nowadays worldwide research areas.

MaGrid (cf. figure 2) is currently a set of 4 working sites, as seen on Gstat, offering a total of 162 Intel® Xeon® CPU Cores and about 21 TB of storage. All components are running Scientific Linux 5.5 OS and gLite 3.2 Grid Middleware. Several applications were adapted and ported on MaGrid and are now accessible through the MaGrid VO (http://magrid.ma/), a national Virtual Organization created to embrace the growing Moroccan scientific communities.

In order to port RegCM-4.3 on the distributed grid environment, it was compiled along with its required components within a virtual environment similar to the one on the Worker Nodes. To fulfill the Grid requirements for Climate modeling, several script are written in order to monitor the model running process. The required data are stored in the Grid storage element and handled directly by users. Then after setting up correctly the configuration files and running several preliminary tests (jobs) with small geographical domains with different initial conditions, the package was validated and tagged within the MaGrid VO. RegCM-4.3 is currently operational on MaGrid infrastructure and available for researchers on Meteorology. As it supports the parallel mode run (MPI), the Grid advanced MPI jobs are possible.

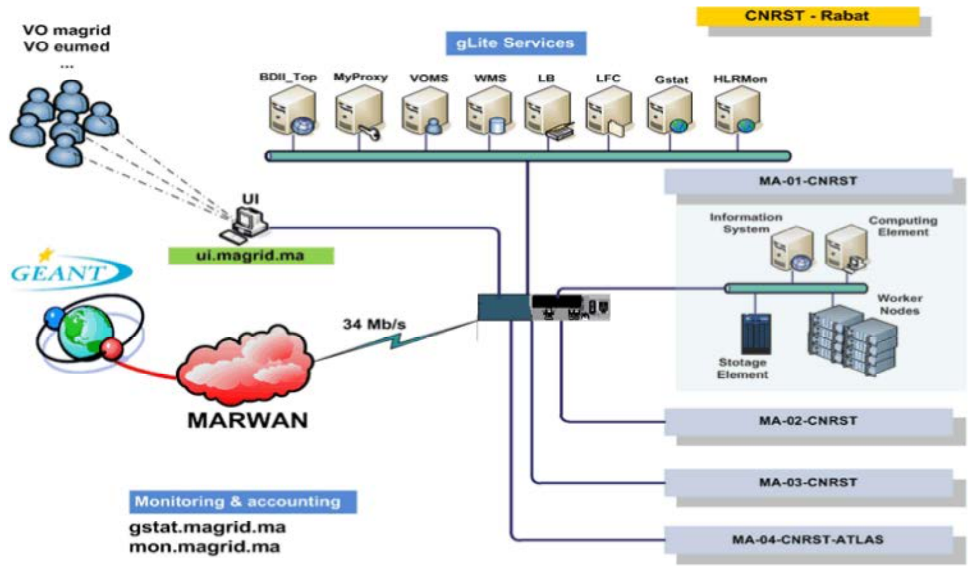

Fig. 2: Schema of Moroccan Grid Infrastructure: MaGrid.

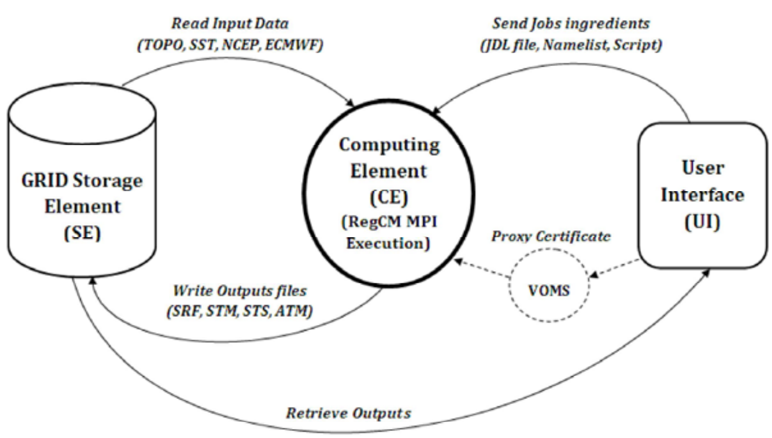

Fig. 3: Schema of interaction of RegCM simulation with data management within MaGrid.

Description of Gridded RegCM4.3 Test Case: The first step is the selection of model domain and resolution which is mostly determined by the nature of the problem and the availability of computing resources. The domain should be large enough to allow the model to develop its own circulations and to include all relevant forcings and processes and the resolution should be high enough to capture local processes of interest (e.g. due to complex topography or land surface) but avoiding that the boundaries of the domain cross major topographical systems. This is because the mismatch in the resolution of the coarse scale lateral driving fields and the model fields in the presence of steep topography may generate spurious local effects (e.g. localized precipitation areas) which can affect the model behavior, at least in adjacent areas [7].

Required data to run the model (DOMAIN, Topography, Landuse, SST and ICBC files) are stored on the Grid Storage Element and accessed further by the user's script during job runtime. The DOMAIN file localizes the model on a world region and contains the localized topography and landuse databases, as well as projection information and land sea mask. The SST file contains the Sea Surface temperature to be used in generating the Initial and Boundary Conditions for the model for the period specified in the namelist file. Next step is to create the ICBC (Initial Condition, Boundary Conditions) for the model itself. The ICBC files contain the surface pressure, surface temperature, horizontal 3D wind components, 3D temperature and mixing ratio for the RegCM domain for the period and time resolution specified in the input file. The model is forced by

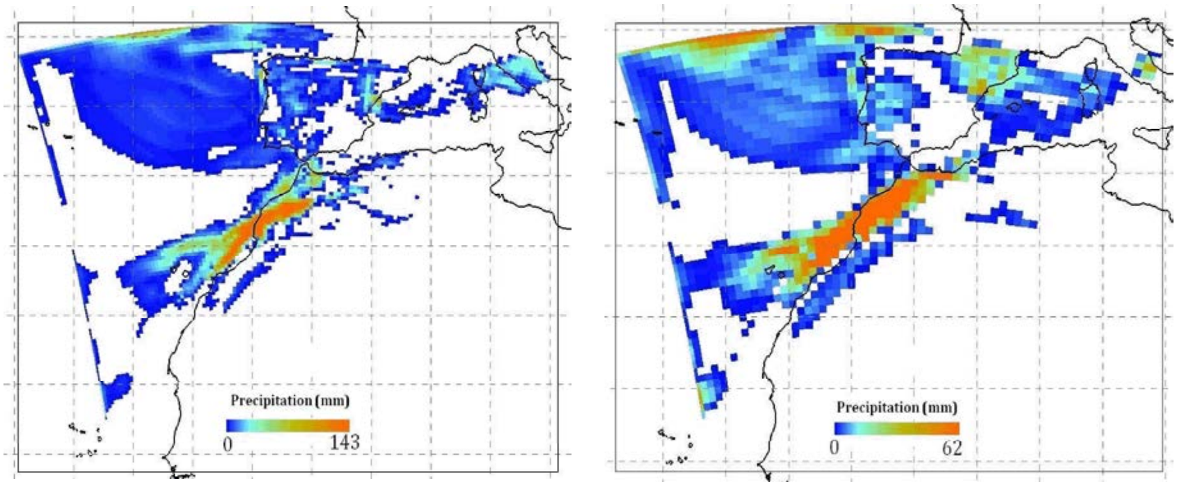

Fig. 4: Total Precipitation simulated by RegCM4.3 for 2010-11-30 with 30km (left) and 60km (right) grid spacing.

ERA-Interim Reanalysis data gridded at 1.5°. we wrote a script which executes all the above steps, beginning with domain creation until running the model. The script need the namelist file edited by the user to configure the model.

Jobs Management: In the RegCM version 4.3 the model parallelizes execution dividing the work between the processors along the jx (longitude) dimension. The minimum work per processor is 3 points along the jx dimension, so the maximum number of processors which can be used in a parallel run depends closely to the horizontal grid dimension [7]. In this case, we have chosen two simulation runs with 120 and 60 as horizontal grid dimension jx to be consistent with the above rule. The script written are executed on the Grid via a JDL job, we have started with 2 CPUs then 4, 6, 8, 16, 20, 40 and 80 CPUs.

RESULTS AND DISCUSSION

In order to assess the impact of the model resolution to the running time, several jobs are submitted to the Grid in which the model namelist is configured with different grid spacing (30km and 60km). The results are shown in the figure 4. The CPU time decreases with the increasing number of CPUs used up to 6 CPUs for both grid spacing and start increasing from 6 CPUs for 30km grid spacing and stabilizing from 16 CPUs for 60km grid spacing. This show that the computational efficiency of RegCM4.3 simulation on MaGrid degrades when a number above 6 CPUs are used as a result of increasing overhead on MPI communication. In the case of 30km grid spacing, the computational efficiency decrease rapidly for simulations starting 20 CPUs and above, but remains quasi constant for 60km grid spacing domain simulation.

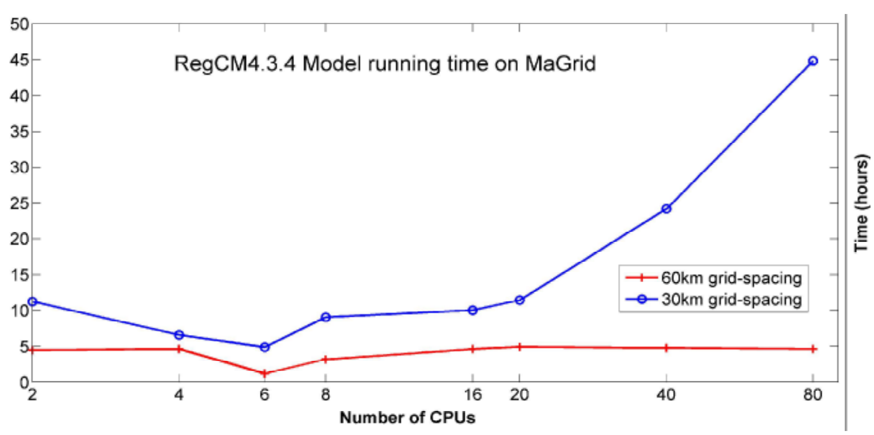

Fig. 5: RegCM4.3.4 Simulations time with CPUs on MaGrid with a grid resolution of 30-km and 60-km.The horizontal grid dimensions are 120 x 120 and 60 x 60 grid points.

CONCLUSION

In this paper, we have described a porting process of a regional climate model on the Grid Infrastructure. The RegCM4.3.4 model has been successfully implemented on Grid infrastructure. The test cases are chosen to finish in reasonable time by selecting limited Domain and short simulation length. The tests indicate that Grid is able to produce reliable solution for Regional Climate simulations. Nonetheless, RegCM's finer resolution simulation which is in particular highly useful for regional climate studies still faces an extra computational barrier; the computational efficiency decrease rapidly with CPUs number for high resolution domain simulations. This limitation makes the Grid unsuitable for massive parallel jobs. Also the data management limits throughout between CPUs and the storage requirement is still an issue especially for climate model which need a large amount of data in order to be run and at the same time, produce a large amount of data. Thus still an effort should be done to allow parallel execution and improve data bandwidth on Grid infrastructure.

REFERENCES

1. Pal, J.S., F. Giorgi, X. Bi, *et al*., 2007. The ICTP RegCM3 and RegCNET: Regional climate modeling for the developing world, Bull. Amer. Meteor. Soc., 88: 1395-1409.
2. Giorgi, F. and L.O. Mearns, 1999. Introduction to special section: Regional climate modeling revisited, J. Geophys. Res., 104: 6335-6352.
3. Giorgi, F., J.S. Pal, X. Bi, L. Sloan, N. Elguindi and F. Solmon, 2006. Introduction to the tac special issue: There gcnet network. Theoretical and Applied Climatology, 86: 1-4.
4. http://gforge.ictp.it/gf/download/docmanfileversion/31/753/ReferenceMan.pdf
5. Georgiana, MARIN, 03/2011. Grid Computing Technology, Database Systems Journal, II: 3.
6. Fernández-Quiruelas, V., *et al*, 2011. Benefits and requirements of grid computing for climate applications. An example with the community atmospheric model / Environmental Modelling & Software, 26: 1057-1069.
7. Filippo Giorgi and all, May 2011. Regional Climatic Model RegCM User's Guide Version, 4-3.
8. Hung-Neng, S. Chin, October 2, 2008. WRF Test on IBM BG/L: Toward High Performance Application to Regional Climate Research, LLNL-TR-407434.
9. Murray, D., J. McWhirter, S. Wier and S. Emmerson, 2003. The Integrated Data Viewer: a Web-enabled application for scientific analysis and visualization. Preprints, 19th Intl Conf. on IIPS for Meteorology, Oceanography and Hydrology.