

CLIMATE CHANGE RESEARCH AT THE EXASCALE

Giovanni Aloisio^{*,§}, Italo Epicoco^{*,§}, Silvia Mocavero[§] and Mark Taylor[^]

(*) University of Salento, Lecce, Italy

(§) Euro-Mediterranean Centre for Climate Change – CMCC, Italy

(^) Exploratory Simulation Technologies Dept, Sandia National Labs, Albuquerque -USA

Abstract

Climate change needs both high-resolution simulations for short-term prediction and the integration/coupling of several models for long-term climate projection. Exascale can be considered as a deep revolution on climate change applications, allowing highest resolutions of climate models that will surpass the resolution of today's operational weather forecast models. This allows for the creation of a unified next-generation atmospheric simulation system, while sharing the expertise and know-how of both climate and weather communities. Climate and weather scientists can then collaborate on unified models, that strongly reduce uncertainty in climate prediction. However, legacy climate applications, in order to exploit exascale, require models re-engineering, improving computational kernels, new parallel model design and algorithms. Moreover, new models, dynamic grids and solution methods that have to be conceived on exascale computers, in charge of carrying out efficient operations.

Climate change research

Climate research community established that the configuration of a climate model consists in the definition of a trade-off among (i) complexity, as the range of model components involved (i.e., physical, chemical, and biological) and their interactions, (ii) prediction term/ensemble size and (iii) resolution (Figure 1) [1].

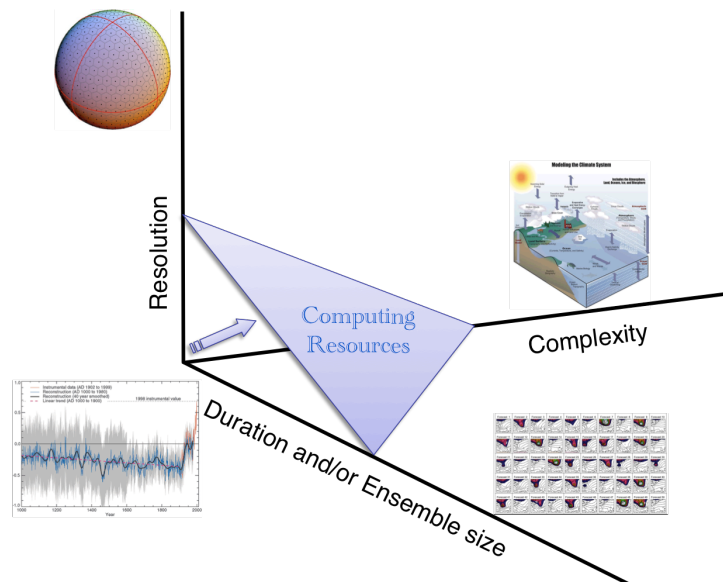


Figure 1 - Resources trade-off. Image of Jim Kinter (COLA)

The IPCC (Intergovernmental Panel on Climate Change) Working Group I (WG-I) focused on the assessment of the physical scientific aspects of the climate system and climate change, has established

that current computational resources and advances in climate model allow for the selection of two different classes of next experiments to be collected in fifth Assessment Report (AR5) [2], to be delivered by the end of 2013:

- short-term climate predictions (to 2030), for which single scenario and high-resolution coupled (atmospheric/ocean) GCMs (General Circulation Models) are used. These predictions are designed for impacts study and mainly directed to policy and decision makers. The multi-model approach applied for the AR4 will have to be further developed. The availability of science and methods from related fields, such as weather prediction, leads to various novel ways in which more information can be drawn from the rapidly growing database of climate model simulations
- long-term climate projections extending beyond 2100, for which single/multi-scenario and medium resolutions models are used. These projections include GCMs of different complexities, taking into account physical and biogeochemical feedbacks from climate change impacts on land use and economy.

The Working Group on Coupled Modelling (WGCM) provided a new set of experiments, representing the fifth phase of the Coupled Model Intercomparison Project (CMIP5). CMIP5 is intended to cover the next five years, including simulations for assessment in both the AR5 and beyond it. Taylor et al. [3] have defined the protocol for the CMIP5 that will promote a standard set of model simulations in order to:

- evaluate how realistic the models are by simulating the recent past
- provide projections of future climate change on two time scales, near term (out to about 2035) and long term (out to 2100 and beyond)
- understand some of the factors responsible for the differences in model projections, including quantifying some key feedbacks such as those involving clouds and the carbon cycle
- define a standard for data and metadata description.

The CMIP5 experiment design has been finalized with the following suites of experiments:

- I. "near-term" (decadal) prediction simulations, initialized with observed ocean state and sea-ice
- II. "long-term" (century) simulations, initialized with the output of GCM past period simulations
- III. "atmosphere-only" high-resolution stand-alone simulations performed on a short period, generally the time-slices 1979-1988 (AMIP period) or 2026-2035, due to the limit of available computational resources.

Each research group on climate change has established its own roadmap for AR5 simulations. Starting from the experiments suggested by the CMIP5, a plan has been defined by trading-off the number of years to be simulated and the resolution for short and long-term simulations; the aim is to deliver the simulations outcomes by the end of 2010 (models with a low-resolution are used as pilot for high-resolution simulations). Considering the current computational architectures, almost all of the main climate centers have planned to contribute to the AR5 by using models at 50Km - 10Km horizontal resolution for short-term predictions. As for today, available resources allow to scale the models execution to O(10K) cores. At higher resolutions, there are several bottlenecks to scalability:

- each routine must be improved, even those which were formerly considered as computationally insignificant
- models performance degrade due to I/O operations: on modern high performance architectures, both efficient parallel I/O libraries and a re-design of output strategy (all variables at all points at

- regular intervals may not scale) are necessary
- pre and post processing packages for models are not scalable
- the current models write the diagnostics information and the post-processing is carried out at a different job-step. It is necessary to explore how the post-processing can be carried out as the model runs
- the numerics based dynamical core of an atmospheric component in a climate model, doesn't scale if a 1D domain decomposition is used: it limits the scalability to O(1K) cores. The use of advanced grids (i.e. cubed-sphere, geodesic, yin-yang) allows for the change of domain decomposition by increasing scalability to O(10K) cores.

Computational challenges at the exascale

In the future, exascale will be able to provide the computational resources needed to increase resolution, complexity and ensemble size. Simulation duration can benefit from exascale whereas models are able to scale. There are several challenges for the current climate models to face when scaling at an order of 10^{18} . At higher resolution, new physical aspects can be taken into account and integrated into the climate models; it is necessary to design scalable computational kernels and algorithms, as well as considering new approaches and paradigms in the parallel programming in order to follow the features of the exaflops architectures. Some of these aspects are presented as follows.

Model complexity: in many climate models, details in the representation of clouds can substantially affect the model estimates of cloud feedback and climate sensitivity [4]. The cloud feedbacks remain the largest source of uncertainty in climate sensitivity estimates. Yesterday, the GCMs were the most comprehensive models available, including a higher number of components of the climate system, designed to provide the best representation of the system and its dynamics. Today, the Cloud Resolving Model (CRM) replaces the conventional convective and stratiform cloud parameterizations, as well as allowing for explicit computation of the global cloud fraction distribution for radiation computations. They are integrated within the Global Cloud Resolving Model (GCRM)[5], that represents a global atmospheric circulation model with a grid-cell spacing of approximately 3 km, capable of simulating the circulations associated with large convective clouds. The major limitation is its high computational cost. In general, there are three issues that have to be addressed: which equations should be used; what are the right methods to solve these equations how get a computer capable to compute the solution fast enough to be useful [5]. Exascale architectures could provide a solution to the last issue. High-resolutions needed for cloud solving are limited by the convergence of the mesh at the poles, using the classical Cartesian grid. There are many successful techniques to handle this pole problem; however, most of them substantially degrade parallel scalability by requiring too much inter-processor communication. The use of advanced grid, such as the spherical geodesic or the cubed-sphere grid, permits very high-resolutions. The spherical geodesic grid-based atmospheric model developed with SciDAC [5] support at Colorado State University is only one of several advanced grid technologies that would permit resolutions of order 1 km.

Data Assimilation: one of the main aspects that will be influenced by the availability of high-resolution models (O(1km)) models is the management of the huge amount of observed data needed to evaluate the predictions reliability. Most of these data come from the Earth Observation System (EOS) communities, especially those coming from remote sensing applications that should be efficiently exploited for data assimilation into the Earth system models. The high spatial resolution and worldwide extent of the remote sensing data are essential qualities can be efficiently used to describe the interactions among the local, regional and global scales of meteorological and climate aspects. The volume of the data sets, which currently reaches the order of magnitude of 1 terabyte per day of

multispectral imagery, poses a significant challenge to its application namely for assimilation of land-surface data into Earth system models at high resolutions. Then, a full data assimilation is not feasible with the available computational capabilities.

Computational kernels: the largest scalability bottleneck in today's climate system models is within the dynamical core (dycore) of the atmosphere model component, due to the numerical methods used to solve the pole problem mentioned above. Many modeling centers are thus developing more scalable dynamical cores based on unstructured or block structured quasi-uniform grids for the sphere which avoid clustering points at the poles. One example, using the cubed-sphere grid and spectral finite elements, is being pursued by scientists from Sandia National Laboratories, the National Center for Atmospheric Research (NCAR), and the Oak Ridge National Laboratory. This dycore is part of NCAR's High-Order Method Modeling Environment (HOMME) and has been integrated in the CAM atmospheric component of the Community Climate System Model (CCSM), improving its scalability and performance. As a result, a version of the CCSM with unprecedented scalability has been run efficiently on 86,000 processors with a horizontal average grid spacing of 25 km [6]. However, obtaining high quality solutions on these quasi-uniform grids with the desired accuracy, conservation, advection and dissipation properties, is as important as performance and scalability and remains an active area of research. HOMME is addressing some of these issues through the use of high-order numerics, a *compatible* discretization (making it the first dycore in the CCSM to locally conserve mass, energy and 2D potential vorticity) [7] and conservative non-oscillatory tracer advection scheme [8]. All these improvements allowed for the development of a dynamical core in the CCSM with unsurpassed scalability and very competitive numerics.

Load balancing: climate models on scalable parallel computer systems can suffer from evident load unbalancing among its components, for example: in the sea-ice models the only computational load is related to the pole grid points; whereas in the ocean models the computational load is negligible on land grid points; finally in the atmosphere models the computational load related to some physical parameterization (solar radiation, convective adjustment, etc.) is related only on specific grid points. The load balancing plays a key role in the performance of a parallel algorithm and it is one of the limitation factors to the scalability of the modern global circulation models. Several efforts have been conducted to improve the load balancing algorithms: space-filling curves partitioning is one of the promising efficient approaches, as it allows for the elimination of load imbalance in the computational grid due to land points. Improved load balance combined with code modifications within the conjugate gradient solver significantly increases the simulation rate of the Parallel Ocean Program at high resolution [9].

Parallel programming: the upcoming computer architectures for petaflops are characterized by the enhancement of on-chip level of parallelism, an increasing number of cores inside a single CPU or node as well as an even more sophisticated hierarchical structure of memory. The race to exaflops will bring us to even more improved high-end architectures that will have an increasing reliance on software-managed on-chip parallelism. These architectural trends bring into question the message-passing programming model that has dominated high-end programming for the past decade. As we move towards massive on-chip parallelism, the consideration of models in addition to message passing turns out necessary. Several different approaches can be considered, among which the Partitioned Global Address Space (PGAS) languages are an alternative to both message passing models like MPI and shared memory models like OpenMP. PGAS languages offer the possibility of a programming model that will work well across a wide range of shared memory, distributed memory and hybrid

platforms. Some of these languages, including UPC, CAF and Titanium, are based on a static model of parallelism, which gives programmers direct control over the underlying processor resources [10]. The parallel models adopted within the current scenario of climate community exploit the message-passing paradigm. Several efforts are currently being carried out to migrate to a hybrid approach with OpenMP and MPI, in order to better exploit the shared memory among an ever increasing number of cores per node.

Reference

- [1] “Scientific grand challenges: challenges in climate change science and the role of computing at the extreme scale”, Report from the Workshop Held November 6-7, 2008, Washington D.C. (<http://www.sc.doe.gov/ober/ClimateReport.pdf>)
- [2] “Scoping Meeting for the IPCC Fifth Assessment Report (AR5)”, Venice, Italy, 13-17 July 2009. (http://www.ipcc.ch/scoping_meeting_ar5/documents/doc02.pdf)
- [3] K. E. Taylor, R. J. Stouffer, and G. A. Meehl, “A Summary of the CMIP5 Experiment Design”, 31 December 2008. (http://www.clivar.org/organization/wgcm/wgcm-12/reports/Taylor_CMIP5_expts7.pdf)
- [4] “Climate Change 2007: The Physical Science Basis” IPCC Fourth Assessment Report (AR4), August 2007. (http://www.ipcc.ch/publications_and_data/publications_ipcc_fourth_assessment_report_wg1_report_the_physical_science_basis.htm)
- [5] D. A. Randall, “Design and Testing of a Global Cloud-Resolving Model”, Colorado State University. (http://kiwi.atmos.colostate.edu/gcrm/Midterm_Review.pdf)
- [6] Taylor, Edwards, St.Cyr, “Petascale Atmospheric Models for the Community Climate System Model: New Developments and Evaluation of Scalable Dynamical Cores”, J. Phys. Conf. Ser., 125, 2008.
- [7] Taylor, Edwards, Thomas, Nair, “A mass and energy conserving spectral element atmospheric dynamical core on the cubed-sphere grid”, J. Phys. Conf. Ser., 78, 2007.
- [8] M. A. Taylor, A. St.Cyr and A. Fournier, “A non-oscillatory advection operator for the compatible spectral element method”, Springer Lecture Notes in Computer Science 5545, 2009.
- [9] J. M. Dennis, “Inverse Space-Filling Curve Partitioning of a Global Ocean Model”, IPDPS, 2007
- [10] K. A. Yelick, “Programming models for petascale to exascale”, IEEE IPDPS, 2008