EUROPEAN COMMISSION

Community research

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

IS-ENES2 DELIVERABLE 9.1
HR ESM Initial performance analysis

File name: {IS-ENES2_D9_*1*.pdf}

Reporting period: *01/04/2013 – 30/09/2014*

Author(s):	Eric Maisonnave Uwe Fladrich

Reviewer(s):	Irina Fast Graham Riley

Release date for review: 2014/09/16

Final date of issue: 2014/09/29

Revision table			
Version	**Date**	**Name**	**Comments**
0.1	2014-09-16	Uwe Fladrich	Original version for review
0.2	2014-09-18	Uwe Fladrich	Including review comments from Irina
0.3	2014-09-18	Uwe Fladrich	Including review comments from Graham
1.0	2014-09-29	Uwe Fladrich	Final version

Abstract

The list of software components participating in the the multi-model multi-member, high-resolution coupled climate simulation experiments is established. A new set of metrics for the computational performance of Earth System Models is developed and used for an initial performance analysis of the models contributing to JRA1.

Project co-funded by the European Commission's Seventh Framework Programme (FP7; 2007-2013) under the grant agreement n°312979		
Dissemination Level		
PU	Public	
PP	Restricted to other programme participants including the Commission Services	
RE	Restricted to a group specified by the partners of the **IS-ENES2** project	
CO	Confidential, only for partners of the **IS-ENES2** project	

Table of contents

Executive Summary

1. Objectives

The main objective of JRA1 is to define, set up, and run a multi-model multi-member high-resolution (M4HR) Earth System Model (ESM) ensemble experiment. This deliverable establishes the prerequisites for the planned M4HR experiments, which are threefold: (i) a list of software components and participating ESMs contributing to JRA1; (ii) a new set of metrics for the computational performance of climate models; (iii) an initial performance analysis of the participating ESMs using the aforementioned set of metrics.

2. Results

The list of software components and participating ESMs has been established, which comprises five European ESMs in high-resolution configuration as well as seven associated components, which are partly developed and used collaboratively among the contributing institutions. Reflecting the shortcomings of conventional metrics for computational performance, a new set of metrics was developed based on the ideas of V. Balaji. This new development provides for a much more specific analysis of climate models. Using this new set of performance metrics, the participating ESMs are subjected to an initial analysis, which shows that the ESMs under consideration form, indeed, a distinct class of computational models.

3. Perspectives

The list of software components and ESMs will be at the basis of further work in JRA1, particularly when defining the specific M4HR experiments. The newly developed set of performance metrics will be suggested for wider use in the climate community and particularly in other IS-ENES2 work packages. The results of the initial performance analysis indicate need for optimisation and will be used to make efficient use of the HPC resources needed for the M4HR experiments. Moreover, the analysis results can be used in association with (or be compared to) the activities in JRA2.

Objectives

The main objective of JRA1 is to define, set up, and run a multi-model multi-member high-resolution (M4HR) earth system model (ESM) ensemble experiment. As a first step, this deliverable aims at the definition and set up phases. In order to come up with a working multi-model configuration a list of participating ESMs along with required software components needs to be established. Furthermore, documentation of the participating ESMs regarding their capabilities in the context of a multi-model experiment is needed. The research institutions are required to secure computer resources for their respective ESM and go through the porting and installation process for their ESM on the respective computing platform.

One prerequisite for the M4HR experiments is an initial and individual assessment of the computational performance of the participating ESMs. However, it turns out that the computational performance of ESMs is not easily determined, mainly because traditional metrics for computational performance – as well as related performance measurement tools – are often failing to reflect the particular behaviour of ESMs and the needs of the scientific community running this type of software. One reason for this is the inherent multi-physics and multi-scale nature of the earth system, which is often dealt with by loosely coupling domain-specific codes through a coupling software and by running very long simulations. Another reason is the complexity and long heritage of ESM software. Another key issue is that traditional performance analysis often focuses on individual computational kernels, whereas ESM experiments often suffer from bottlenecks that arise when the complete workflow is assembled.

Consequently, a new set of metrics for the computational performance of ESMs needs to be developed in order to complement the traditional metrics. The new metrics should be based on the experience with previous ESM experiments (such as CMIP5) and should answer practical questions that commonly emerge in larger ESM experiments. The initial performance analysis of the participating ESMs is then based on the newly developed metrics.

Results

The results reported in this deliverable comprise three parts: The lists of participating ESMs and further software components, a new set of computational performance metrics for ESMs, and the results of an initial analysis of the participating ESMs using these metrics.

1.1 List of Participating Earth System Models

Five European ESMs have been identified for participation in JRA1's M4HR experiments. However, this does not imply that all of them contribute to the same experiments since there will be different degrees of coordination and integration between individual ESMs and their respective institutions. The following table lists all five ESMs along with their representing institution in JRA1 and their major software components:

ESM name	Institution	Components
ARPEGE-NEMO	Météo-France / CERFACS	ARPEGE, NEMO, OASIS3-MCT
EC-EARTH3	SMHI	IFS, NEMO, OASIS3
HadGEM3	UK Metoffice	GA6.0, GL6.0, NEMO, CICE, OASIS3
CESM-NEMO	CMCC	CAM, NEMO, CLM, CICE, RTM, CPL7
NorESM	MET.no	CAM-Oslo, NCC-MICOM, CLM, CICE, CPL7

There are, as the table shows, some common software components, for example the OASIS coupler and the NEMO ocean model. Another common component to some of the ESMs (although not listed in the table) is the I/O subsystem. Both the OASIS coupler and the I/O subsystem are subject to further investigation in other tasks of JRA1. Moreover, infrastructure components, such as job control and data analysis tools, will be different across models and institutions and will be looked at in separate tasks of JRA1.

1.2 List of Further Software Components

The software infrastructure for M4HR experiments does not only include the actual ESMs but also further components needed to complete the workflow. The list of software components that receive special attention in JRA1 reads:

Component name	Institution	Component type
OASIS	CERFACS	Coupler
XIOS	CNRS-IPSL	I/O subsystem
CDI-PIO	DKRZ	I/O subsystem
CDO	MPG	Postprocessing/Data analysis tool
Autosubmit	IC3	Job control tool
Cylc	UREAD-NCAS	Job control tool
Rose	MetOffice	Job control tool

Even though the coupler and the I/O subsystem can be seen as integral part of actual ESM, they are listed separately here because they are subject to investigation in tasks 2 and 3, respectively.

1.3 Metrics for performance evaluation of ESMs

The computational performance of a single member ESM experiment can be assessed by traditional performance metrics as usually applied by computing centres. These metrics typically include single-core measures, such as the rate of floating point operations (FLOP), and parallelism measures, such as speed-up and parallel efficiency. Although these are important metrics for performance analysis and optimisation, they mostly reflect the computing centre's view and provide the climate scientist with a limited picture of how the ESM's behaves on the computing platform. Typical questions that ESM users have when they plan or run an experiment include:

- How long will the experiment take (including data transfer and post-processing)?
- How many nodes (cores+memory) can be efficiently used in different phases of the experiment?
- Are there bottlenecks in the experiment workflow?
- How much short-term/medium-term/long-term disk space is needed?
- Can/should the experiment be split up in parallel chunks (e.g. How many ensemble members should be run in parallel?)?

Although these questions are clearly related to the computational performance of ESMs, they are not answered by examination of FLOP rates or speed-up curves. In order to provide better

answers to the above questions, a new set of performance metrics has been developed, which is specific to the needs of ESMs and climate model experiments. The performance metrics are based on the ideas of V. Balaji, as presented at the IS-ENES Workshop "Exascale Technologies and Innovation in HPC for Climate Models" [1] and were adjusted and extended to fit the needs of JRA1.

The following list of performance metrics was developed and used for the initial analysis of the participating ESMs in JRA1:

Resolution: Spatial resolution of the computational grid for each of the physical domains (typically atmosphere and ocean) complemented by the total number of grid points. The resolution can be specified in a domain-specific way, for example, the average horizontal spacing and the number of vertical levels for atmospheric grids.

Complexity: Different measures were discussed before compromising on the number and dimension of variables in the ESM's restart files. The assumption is that the restart files represent the internal state of the ESM, thus allowing an estimate of the complexity to be deduced from the size dimension of the internal state space. The main advantage of this measure is that it is easy to obtain from any ESM.

Simulated years per day (SYPD): The number of years that can be simulated by the ESM in a given configuration on a given computing platform during a 24-hour period, assuming dedicated computing resources. Practically, this number is often deduced from shorter (than one day) test runs.

Actual simulated years per day (ASYPD): The number of years that can be simulated by the ESM in a given configuration on a given platform in a multi-user environment (i.e. *not* assuming dedicated resources). This metric is usually measured using a long simulation with restarts, thus including queueing time between chunks, and Workflow cost (see this term below).

Core hours per simulated year (CHPSY): This metric measures the actual computational cost of the ESM simulation. It is usually determined by the product of the model run time and the number of cores used.

Memory bloat: This metric indicates the ratio of actual to ideal memory consumption of the ESM. The ideal consumption memory is deduced from Complexity, as being the total memory needed to fit restart file variables. The actual memory is the only figure that requires a generic measurement tool, usually provided with the scheduler.

Coupler cost: Ratio of time spent in the coupler *doing calculations* to the overall run time of the model. This needs either a thorough performance analysis (tracing/profiling) or support in the coupler software. For OASIS, support for the coupler cost metric has been developed [2].

Load imbalance: Ratio of the time spent *waiting* in the coupler for one of the components to the overall run time. Again, this can be obtained by carefully examining messages sent within the coupled model using a tracing tool. Alternatively, the coupler software can directly collect and present this metric, as OASIS does.

Data output cost: Extra time that an ESM needs to write the model output to the file system. This is measured as the ratio of the run time for a standard run (including standard model output) to the run time for a run with model output switched off.

Data intensity: Amount of data that is read or written by an ESM in a given time during a typical run. For global climate models, it is mostly the written data that contributes to the data transfer, which is why the I/O speed metric may be limited to the output data.

Workflow cost: Additional time is often needed to process and/or transfer model data into the form and place such that what is considered to be the result of the ESM run is achieved.

The workflow cost is the ratio of this additional time to the run time of the ESM. This metrics needs a certain formalisation of the overall workflow to be able to separate the ESM run from the rest of the workflow steps. It is worth noticing that part of the workflow tasks could be done in parallel (concurrently) with the ESM run. This metric is (only) concerned with the extra (consecutive) part of the time needed for workflow tasks.

Parallelisation: The number of computational units (cores or nodes as applicable) that is used for a certain ESM run. This number can be specified separately for the components of a coupled model and complemented by information about the parallelisation paradigm.

Of the above metrics, resolution and complexity are static measures (meaning that they can be obtained by a static analysis of the model), whereas the other metrics are dynamic. Dynamic metrics can only be determined during actual runs of the model and differ usually between any two runs. Due to the computational costs of testing the ESMs, no averaging over several runs is required at this stage, thus accepting a certain level of inaccuracy. All metrics (except Memory bloat) can be simply obtained by scientists without the need for any generic tool provided by computing centres.

Contrary to traditional metrics for computational performance, some of the above measures (namely SYPD and ASYPD) do not only depend on the ESM implementation and computational platform as such, but also on the configuration of the model and, notably, on the usage pattern of the computing platform. This latter fact has been the subject of lively discussion, both within JRA1 and with representatives of computing centres (which oppose the use of these two metrics). Nevertheless, it has been agreed that these metrics provide important information needed by scientists in order to plan and monitor large ESM experiments. In fact, the metrics include valuable information about the capability of a certain computing platform to perform a given experiment in a certain time frame, given its integration in a computing centre.

1.4 Initial results of performance evaluation

The following table lists, as the first part of the initial performance analysis, the static performance metrics for the participating ESMs. The numbers have been reported by the respective modelling groups and include different configurations for some of the ESMs. For comparison, a high-end configuration of the GFDL model is included.

ESM	Component	Resolution		Complexity	Grid points[1]
GFDL- CM2.6S	atmosphere	50 km	0.5L32	18[2]	331,372,800
	ocean	10 km	0.1L50		
ARPEGE-NEMO	atmosphere	50 km	T359L31	53	118,456,350
	ocean	25 km	ORCA025L75	38	
	sea-ice			47	
	coupler			20	
EC-Earth	atmosphere	40 km	T511L91		158,131,358
	ocean	25 km	ORCA025L75	31	

1) Sum over component grids.
2) The complexity for GFDL-CM2.6S is given by the number of prognostic variables and not by the definition used in this document (number of variables in restart files)

ESM	Component	Resolution		Complexity	Grid points[1]
NorESM (A)	sea-ice			78	~5,000,000
	coupler			22	
	atmosphere	100 km	0.9×1.25°	72	
	ocean	100 km	1°	38	
	sea-ice			25	
	land			59	
NorESM (B)	atmosphere	50 km		72	
	land			59	
HadGEM3-GC2 (A)	atmosphere	60 km	0.83×0.55°L85	133	~122,000,000
	ocean	25 km	ORCA025L75	34	
	sea-ice			25	
	land			~36	
	coupler			16	
HadGEM3-GC2 (B)	atmosphere	25 km	0.35×0.23°L85		~178,000,000
	ocean	25 km	ORCA025L75		
	sea-ice			25	
	land			~36	
	coupler			16	
CESM-NEMO	atmosphere	28 km	0.25L30	138	~115,086,100
	ocean	25 km	0.25L50	38	
	sea-ice			46	
	land			121	
CESM1.2	atmosphere		0.25°		~27,000,000
	ocean		1°		

It has to be noted that some of the models or components fall short of the definition for high-resolution (HR) as given by the term “in the range of 20-50 km (0.25-0.5°)” in the text of JRA1 objectives. Nevertheless, it was chosen to relax this requirement slightly in order to allow for a larger group of participating ESMs and to give a measure of the gap between standard and HR configurations.

The next set of performance metrics is concerned with the computational speed of the ESMs in a given configuration, on a given computational platform. The participating groups where asked to run their respective ESM and report the *simulated years per day* (SYPD), *actual simulated years per day* (ASYPD), and *core hours per simulated year* (CHPSY) metrics. It is important to note that no further requirements, other than being practically relevant, were placed on the configuration of the test experiment at this stage. This included the parallelisation of the model, the amount of output or the lengths of the experiment. Most groups chose to report two sets of results: one for a *capability*-type experiment and one for a *capacity* run. In a capability-type experiment, it is tried to minimise the time-to-result, using whatever computational resources needed. A capacity experiment takes the efficient use of the

computational platform into account, which usually requires fewer computational resources. In theory, a capacity run can be defined by setting a threshold value for the parallel efficiency, however, this is not easily done for an ESM as the scalability analysis is often prohibitively expensive. Instead, the definition of capacity and capability configuration is left to the judgement of the modelling groups. The following table lists the performance metrics for computational speed:

ESM	Configuration	SYPD	ASYPD	CHPSY
GFDL-CM2.6S	capability	2.2	1.6	212,465
ARPEGE-NEMO	capacity	5	1.5	5,190
	capability	5.2		
EC-Earth	capacity	2.6		10,353
NorESM (A)	capacity	15.4		
	capability	17.2		1,369
NorESM (B)	capacity	6		
	capability	8		4,129
HadGEM3-GC2 (A)	capacity	1.8	1.8	
	capability			14,745
HadGEM3-GC2 (B)	capacity	0.6	0.6	
	capability			64,056
CESM-NEMO	capacity	0.31	0.063	
	capability			163,111
CESM1.2		0.21		119,422

The remaining performance metrics could be used as indications for potential computational bottlenecks. However, not all of the ESMs provide all metrics, although it is planned to complete the following table over the course of JRA1:

ESM	Memory bloat	Coupler cost	Load Imbalance	Data output cost	Workflow cost
GFDL-CM2.6S	12%	5.7%	20%		
ARPEGE-NEMO	1.2%	4%	13%	<1%	30%
EC-Earth	~5%	3.5%	24.7%		
NorESM (A)		12%			
NorESM (B)		1%			
HadGEM3-GC2 (A)				~10%	
HadGEM3-GC2 (B)				~10%	
CESM-NEMO	1%	3.3%	2.5%	25%	
CESM1.2		~8%			

The Data intensity, although part of the performance metrics set, is not listed because it was provided for only one of the participating ESMs.

The previous table concludes the initial performance analysis of the participating ESMs in JRA1. The numbers given above show a first quantitative assessment of the computational performance of the ESMs in their respective configurations and in their given computational environment. No effort is made to document the characteristics of the computational platforms. On the one hand is this left to conventional performance analysis (such as in JRA2), and on the other hand the assumption is that the chosen metrics are self-consistent even without taking hardware details into account. No ranking is implied and it is even difficult to draw conclusions about computational bottlenecks at this stage. Nevertheless, the results seem to be coherent across ESMs, which confirms the self-consistency of the specific selection of the metrics.

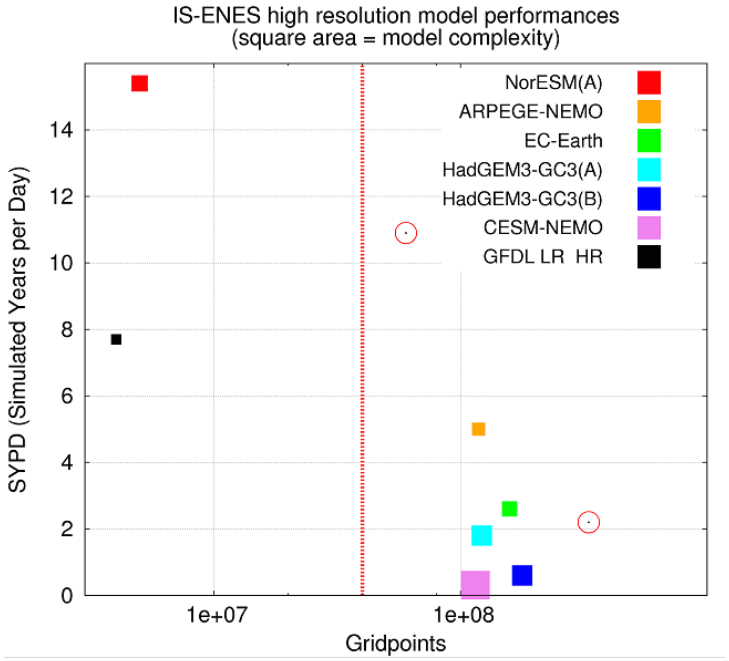

Figure 1: Performance, resolution, and complexity of participating ESMs

Figure 1 illustrates the homogeneity of the chosen complexity and performance metrics across participating ESMs. The models cluster around hundred million grid points (right column) and exhibit a computational performance in the range of 0.3 to 5 simulated years per day. It is also apparent from the figure that the performance decreases with complexity, which seems to validate the complexity criterion.

Perspectives

This deliverable documents the successful start-up phase of JRA1, bringing together the participating institutions with their respective software components. The main results are a new set of metrics for the computational performance of ESMs and its application for an initial analysis of the participating models.

The list of software components – and, more specifically, the particular configurations of the ESMs – will be the basis for the definition of a common set up for the M4HR experiments in JRA1. Moreover, the results of the performance analysis provide the modelling groups with a better insight into the model's behaviour on a given HPC platform as well as early indications of potential bottlenecks. Figure 2 gives an example of how the respective weights of model components in three of the participating ESM can be visualised, thus possibly detecting deficiencies in coupling, load balancing, etc.

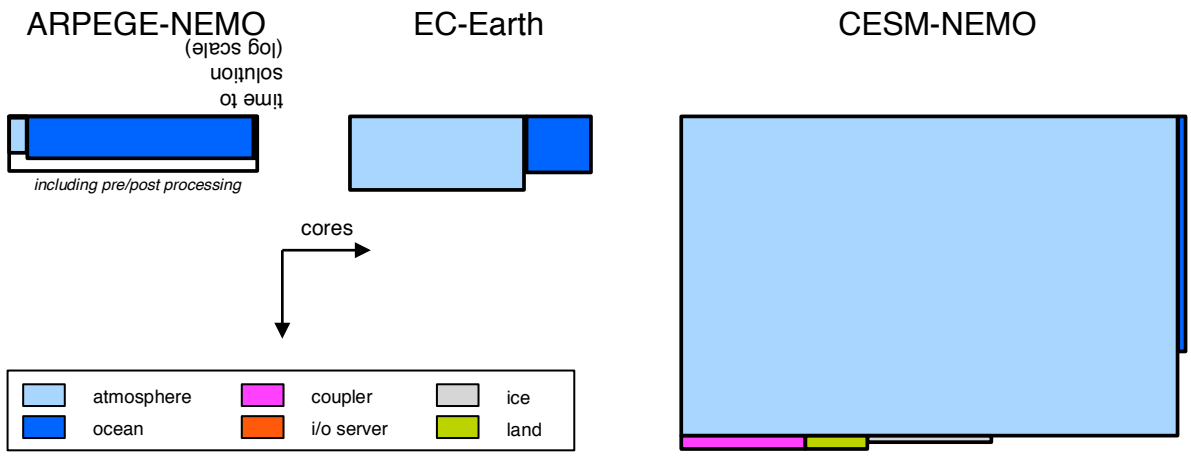

Figure 2: Parallelism and execution time (inverse of SYPD) for three of the participating ESMs

The introduced set of metrics will also help to define particular features of the M4HR experiments. Its extension is necessary to provide information not only about individual member simulations but also ensembles. Particularly, (A)SYPD and data output costs (or I/O speed) seem to be relevant metrics, which should help users to estimate CPU hours, bandwidth and disk storage needed for their future M4HR experiments.

It is expected that the set of performance metrics will be used in a wider context than just within JRA1. Particularly, it is planned to suggest the metrics for use in JRA2 for performance benchmark of coupled climate models.

References

[1] Balaji, V., R. Benson, N. Zadeh, S. Underwood: *Measurements of real model performance.* Exascale Technologies and Innovation in HPC for Climate Models, Hamburg, Germany, 2014

[2] Maisonnave, E., Caubel, A.: *LUCIA, load balancing tool for OASIS coupled systems.* Technical Report, **TR/CMGC/14/63**, SUC au CERFACS, URA CERFACS/CNRS No1875, France, 2014.