WP10/JRA2: Performance benchmarks for coupled climate models

Work Package Number	WP10/JRA2		**Start Date or Starting Event** **End Date**						**Month 1** **Month 48**	
Work Package Title	**Performance benchmarks for coupled climate models**									
Activity Type	RTD									
Participant Number	2	1	3	4	12	10	6	15	7	**TOTAL**
Participant Short Name	**DKRZ**	**CNRS-IPSL**	**CREFACS**	**CMCC**	**UNIMAN**	**MPG**	**MetO**	**LiU**	**STFC**	
Person months per participant	**18**	**11**	**9**	**12**	**7**	**3**	**11**	**6**	**5**	**82**

Objectives:

Climate models are highly complex systems that combine a variety of algorithms within one application. It is thus not possible to assess the performance characteristics of climate models just by using idealized "mini-applications" (often referred to as kernels) or performance modelling approaches for benchmarking, as is popular in other communities. Therefore, benchmarking computer systems with respect to climate applications is cumbersome and involves high costs for application providers, computer vendors and computer centres.

We will provide a framework and a suite of real application benchmarks freely available (upon registration) via the ENES Portal, e.g. to computer vendors and high performance computing centers. This will

- Facilitate and improve benchmarking for procurements
- Provide vendors, manufacturers of CPUs and compiler builders a better way to assess performance characteristics of climate applications, thus fostering co-design and innovation
- Provide the potential to better compare the performance of different climate models on different computer systems, which might eventually lead to a « climTOP10 » for computers

Many climate models today are not single executable programs but coupled systems of often very different individual models. However coupled systems are usually not part of benchmarks because of their complexity and high adaptation costs on new systems. To address this issue we will include coupled models in our benchmark suite and we will compare and evaluate different coupling strategies. This will improve the ability to assess the expected performance of specific coupled models on existing or planned computers, e.g PRACE machines.

<u>The work package is organised into three tasks:</u>

In Task 1 we will set up the organizational and technical framework to distribute and evaluate performance benchmarks for climate modelling.

In Task 2, a suite of models of varying complexity ranging from simple kernels to full coupled models will be put together. Indeed, different Earth System Model components show diverse demands regarding the floating point, memory bandwidth, and I/O capabilities of a HPC system. Since no one model will accurately describe the workload on a production system, a suite of models has to be prepared for benchmarking.

In Task 3 different coupling strategies and coupling tools will be evaluated by defining a suite of coupled benchmarks based on simplified model components which capture the essence of the coupling without the complexities of the science.

Links with other work packages

This work has a strong cross cutting dimension within ISENES2. It will be beneficial to "innovation on HPC"

(NA2 and NA5), since it will help to get improved insight into the performance characteristic of ESM codes. It will substantially count on the results of NA3 "ESMs and their environment" and will also interact with JRA1 "high resolution ensembles" with respect to performance assessment and scalability.

Description of work

Task 1: Framework and benchmarking guide
DKRZ (lead) (6 pm), CNRS-IPSL (co-lead) (5 pm), MetO (2 pm), UNIMAN (3pm)

In this task, we will create the technical and organizational framework to set up, run, distribute and evaluate performance benchmarks. This framework will include a library for fine granular timing of model runs and for
assistance in standardized check for correctness of results. Relevant performance metrics for the type of models provided through task 2 will be identified and a matching performance analysis workflow will be defined and used to actually evaluate the performance of those models. We will leverage the experience gained within the RAPS initiative with respect to licensing, best practices in setting up benchmarks and contact to interested vendors (RAPS: "Real Applications on Parallel Systems", see also raps.enes.org; RAPS is an unfunded effort, only including a limited set of stand-alone models).

Task 2: Suite of base benchmarks
DKRZ (9 pm), CMCC (12 pm), MPG (3 pm), CNRS-IPSL (6 pm), LiU(6 pm)

Using the framework set up in task1, we will provide a suite of typical climate modelling applications. The suite will be comprised of kernels representing compute intensive algorithms that are frequently used in and are characteristic for climate models; full models of individual components of the climate system (including versions with very high resolution to test scalability) and full coupled models that can be run on existing architectures and have been used within the CMIP5 experiments.

Each of the benchmarks (with possible exception of the very high resolution versions) will be tested on at least two different computer architectures in at least two different computing centres. All of these benchmarks will share a number of properties which are not commonly found in the production versions of the models. The most important of these additional properties are detailed timing and automated tests for correctness of the results. Each model has to be modified to provide these properties individually. As far as possible this will be done by using the general methods provided by task 1. The following list shows the envisaged individual benchmark programs along with the responsible partners.
a) Computation intensive kernels extracted from NEMO and from ICON (CMCC, MPG, DKRZ)
b) ICON (atmosphere) full model at two resolutions: state of the art and very high (DKRZ, MPG)
c) NEMO ocean and CAM atmosphere coupled within CESM framework (CMCC)
d) Full coupled model (based on COSMOS-CMIP5 model) (DKRZ, MPG)
e) Full coupled model (based on IPSL-CM5 model) (CNRS-IPSL)
f) Full coupled version of EC-Earth (LIU)

Task 3: Evaluation of coupling strategies
CERFACS (9 pm), STFC (5 pm), UNIMAN (4 pm), MetO (9 pm), DKRZ (3 pm)

The intention is to characterize the essential coupling challenges in European climate model components by defining a suite of coupled benchmarks based on simplified model components which capture the essence of

the coupling without the complexities of the science.
a) This will involve capturing the set of functional and performance characteristics that provide the key constraints on any coupling system.
b) A set of simplified components will be coded with the aim of reproducing the priority characteristics identified in a)
c) The coupling between these simplified components will be implemented with a subset of existing widely used coupling technologies, the initial focus being on OASIS and ESMF. Custom built coupled solutions will also be implemented (i.e. hand-crafted, for example, directly using MPI). These will act as a base case for comparison purposes. This benchmark suite will be released to the community, including model developers, developers of coupling technologies and developers of computer systems.
d) Experiments based on the benchmark suite will be done on specific platforms varying key parameters such
as data size, number of fields and domain decompositions. Results will be analysed and presented to the community, for example, at a future IS-ENES coupling workshop (NA2), at EGU and/or through publication in academic publications.
e) In addition a performance modelling exercise will be undertaken to support the analysis of the benchmark results.

Deliverables

D10.1 (Task 1, mo 18)	Documentation and User Guide for benchmarking framework (DKRZ)
D10.2 (Task 1, mo 33)	Interim report on status of benchmark suite (DKRZ)
D10.3 (Task 3, mo 42)	Report on benchmark suite for evaluation of coupling strategies (CERFACS)
D10.4 (Task 2, mo 46)	Report on the suite of base benchmark and comparison of performance on available platforms (DKRZ)

Description of deliverables

D10.1) Documentation and User Guide for benchmarking framework: The deliverable will be a report describing best practices for evaluating the performance of HPC systems with respect to coupled models. [task 1] [month18]

D10.2) Interim report on status of benchmark suite: The deliverable will be a report describing status of the different application benchmarks provided through this work package within the first two years of the project.
The document will include instructions how to obtain and use those benchmarks and how to analyze the results. Reference performance measurements using these benchmarks will be provided. [task 1] [month 33]

D10.3) Report on benchmark suite for evaluation of coupling strategies: The deliverable will be a report describing the benchmark suite provided for the evaluation of coupling strategies. The different coupling strategies will be compared and judged based on the benchmark suite. The document will also include detailed
instructions on how to obtain and use the benchmark suite. [task 3] [month 42]

D10.4) Report on the suite of base benchmark and comparison of performance on available platforms: The deliverable will be a report describing the final versions of the different application benchmarks provided through this work package. The document will include instructions on how to obtain and use the benchmarks and on how to analyze the results. Reference performance measurements using these benchmarks will be provided. [task 2] [month 46]

Milestones	***Comments***
M10.1 (mo 12) Definition of the benchmark suite for evaluation of coupling strategies (CERFACS)	*Task 3 – links with D10.3*
M10.2 (mo 24) Benchmark version of the four coupled models (DKRZ)	*Task 2 – links with D10.2 and D10.4. Benchmarks downloadable including instructions and performance reference*
M10.3 (mo 30) Kernels extracted from ICON and NEMO (CMCC)	*Task 2 – links with D10.2 and D10.4 Kernels downloadable including description instructions and performance reference*
M10.4 (mo 36) Implementation of the benchmark suite for evaluation of coupling strategies (CERFACS)	*Task 3 – links with D10.3*
M10.5 (mo 36) Benchmark versions of ICON (DKRZ)	*Task 2 – links with D10.2 and D10.4 Benchmarks downloadable including instructions and performance reference*