This presentation participates in OSPP
Outstanding Student Poster & PICO Contest

Simulation-based performance analysis of EC-Earth 3.2.0 using Dimemas

Xavier Yepes-Arbós[1,2], Mario C. Acosta[1], Kim Serradell[1], Alicia Sanchez Lorente[1], Francisco J. Doblas-Reyes[1,3]

[1]Earth Sciences Department, Barcelona Supercomputing Center (BSC), Spain, [2]Universitat Politècnica de Catalunya (UPC), Spain, [3]Institució Catalana de Recerca i Estudis Avançats (ICREA), Spain

1. Introduction

EC-Earth [1] is a global coupled climate model, which integrates a number of component models in order to simulate the earth system. The two main components are **IFS** 36r4 as atmospheric model and **NEMO** 3.6 as ocean model, both coupled using **OASIS3-MCT**. There are other small components like LIM3 as sea-ice model and runoff-mapper to distribute runoff from land to the ocean through rivers.

Coupling consists in connecting and synchronizing different components to exchange data. Since each component has a **different scalability**, to find a good balance between them to reduce inefficiencies is a challenge. This happens in EC-Earth, where to achieve **a good efficiency is a complex issue** [2][3]. For example, the scalability of this model using the T255L91 grid with 512 MPI processes for IFS and the ORCA1L75 grid with 128 MPI processes for NEMO achieves 40.3 of speedup, or 15.7 simulated years per day (SYPD).

Therefore, a **performance analysis** is required to find the **bottlenecks** of the model to then apply the correct optimization techniques. There are previous works using profiling and tracing [2][3], but in this study we present a different approach based on **simulation**. Using traces of EC-Earth, Dimemas can simulate its message-passing behavior to predict the impact of hardware changes.

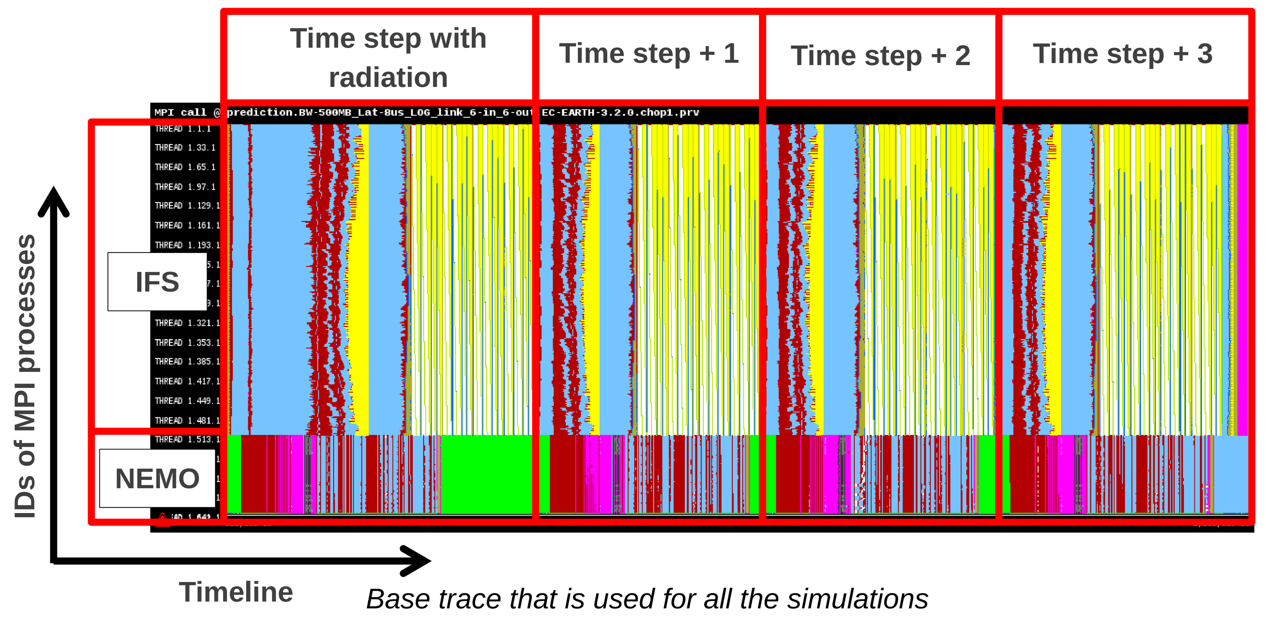
2. BSC tools

Performance tools [4] are essential to study the behavior of EC-Earth:

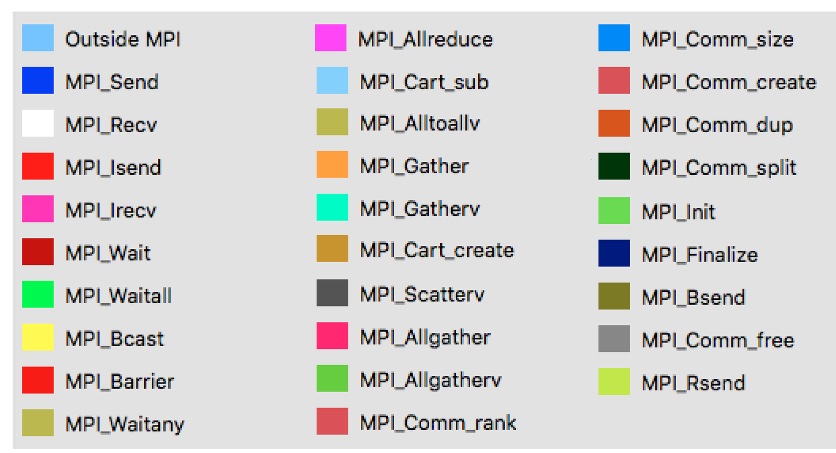
- **Extrae**: is a package used to instrument the code. It generates trace-files with hardware counters, MPI messages and other information.
- **Paraver**: is a browser used to analyze both visually and analytically trace-files.
- **Dimemas**: is a simulator based on traces to predict the behavior of message-passing programs on configurable parallel machines.

3. What can you see in a trace?

The view of a trace consists of **MPI processes** on the Y axis and the **timeline** on the X axis. Particularly, in this poster all views are colored by **MPI functions**, where each type of call is identified by a color. Furthermore, all views contain 4 time steps, using 512 processes for IFS and 128 for NEMO. There are 3 regular time steps and 1 IFS time step with radiation. All the scenarios are compared with regard to a **simulated base trace** that is adjusted to the actual trace. This base trace uses an interconnection network with 8 µs of **latency** (delay of a message between the sender and the receiver) and 500 MB/s of **bandwidth** (the maximum rate that data can be transferred).



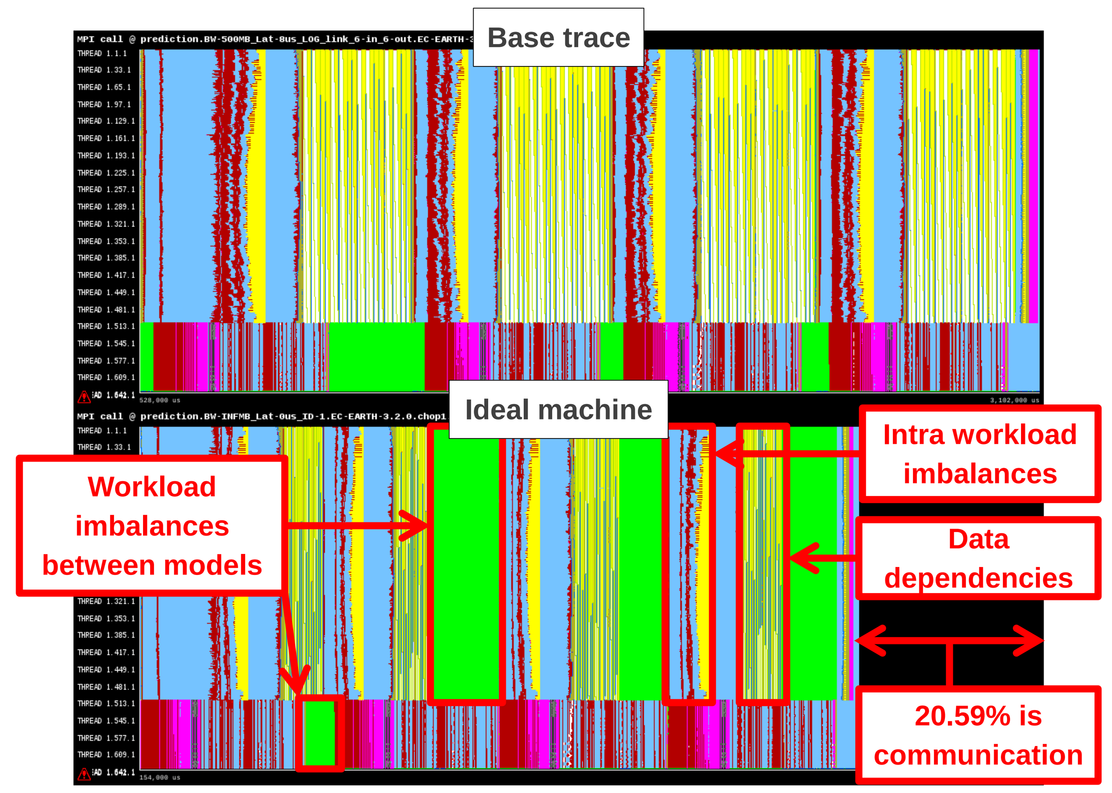
Base trace that is used for all the simulations



Legend with the colors representation of the MPI functions

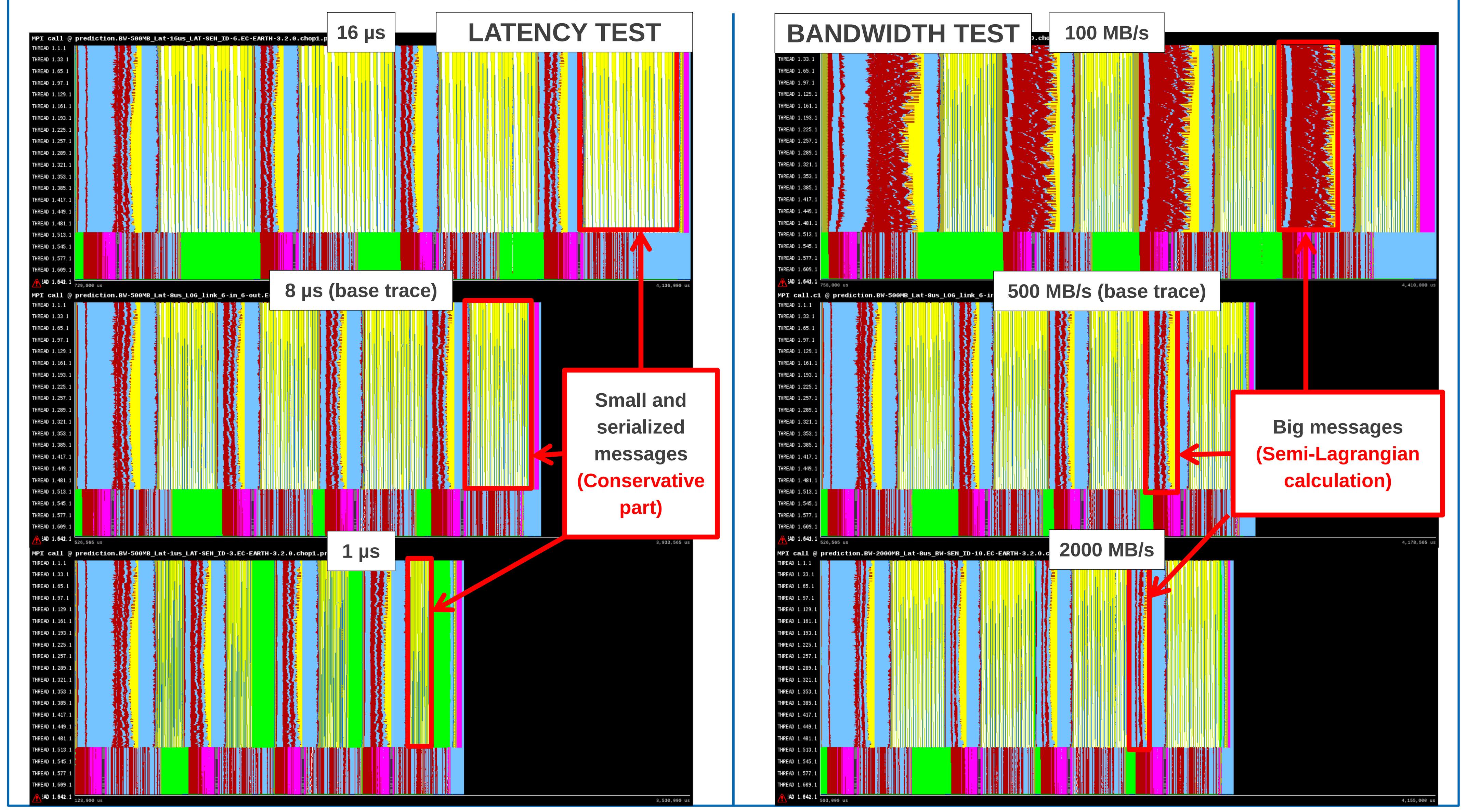
4. Ideal machine

The first test consists in simulating the **ideal machine**, which has an interconnection network with **infinite bandwidth** and **no latency**. Thus, it is possible to study the communication's overhead, the workload imbalance and potential data dependencies. The simulation shows that the 20.59% of execution time is communication, there are workload imbalances and data dependencies.



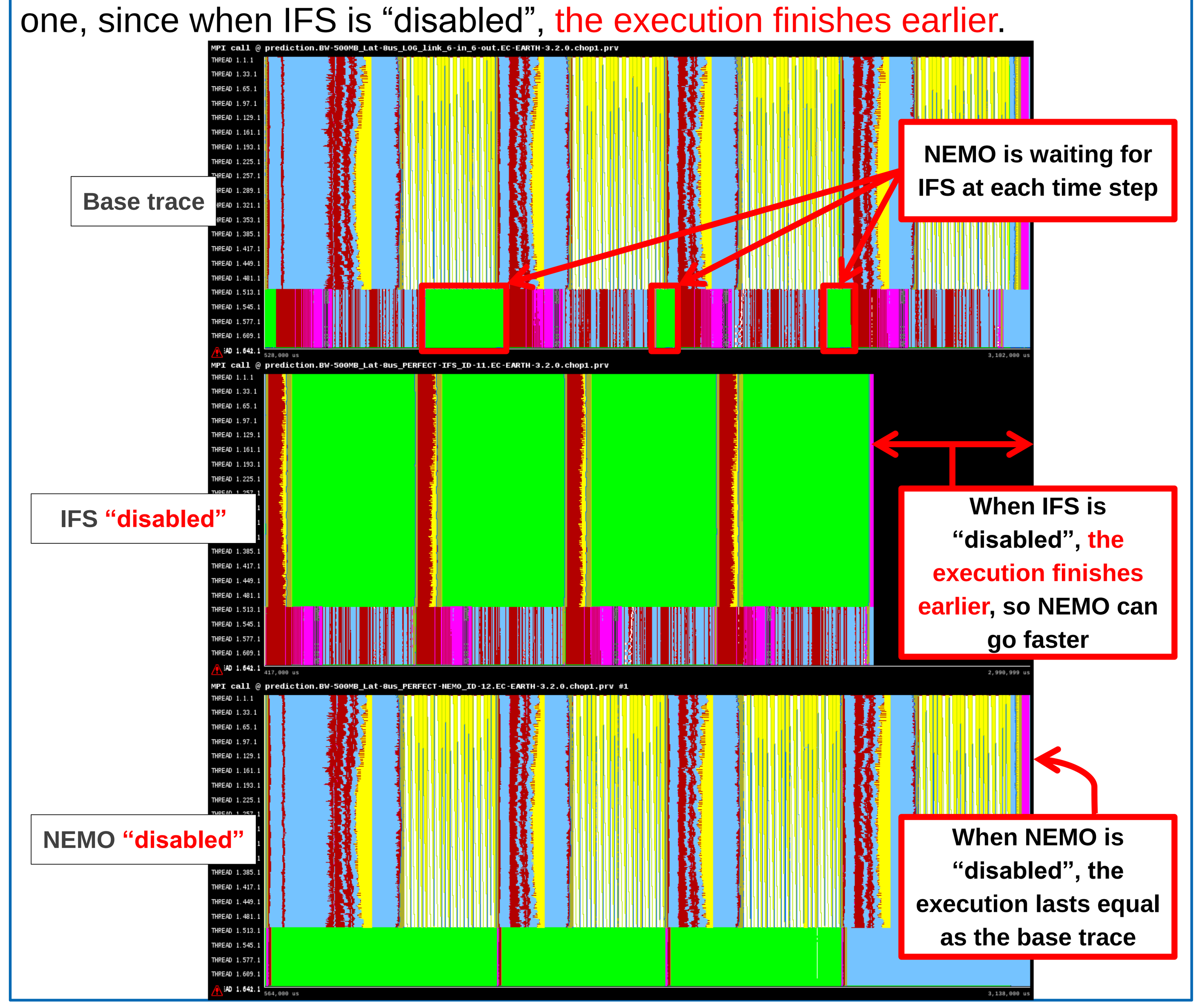
5. Model sensitivity

The second test consists in studying the **communication sensitivity** by changing **latency** and **bandwidth**. This is useful to determine the **efficiency** of communications. It is preferable to have few **big** messages rather than many **small** messages in order to exploit the bandwidth and reduce the latency. The figures suggest that the conservative part in the coupling from IFS to NEMO has small and serialized messages (sensitive to latency) and the Semi-Lagrangian calculation in IFS has big messages (sensitive to bandwidth).



6. Disabling one model

The last test is useful to see the **impact of the coupling** process between IFS and NEMO. The idea is to "disable" one of the models to see how coupling affects the other one. By "disabling" is meant to make one of the models much **faster** than usual and not changing any parameter of the other one. The traces show that the IFS' time step is **slower** than NEMO's one, since when IFS is "disabled", the execution finishes earlier.



7. Conclusions

Each model reacts different to hardware changes.

- The **ideal machine** scenario reveals that there are several sources of inefficiencies: **20.59%** of the execution time is **communication** which means that at least a 20.59% of EC-Earth execution is inefficient; there are not only **workload imbalances** between IFS and NEMO at each time step, but also within each model as shown by MPI waiting functions; and there are **data dependencies** as suggested by chains of messages.
- From a **model sensitivity** point of view the **latency affects** the **conservative part in the coupling** from IFS to NEMO (dependent small messages), whereas the **bandwidth affects** the **Semi-Lagrangian calculation** (big messages). In NEMO, the simulated latencies and bandwidths in this study only affect slightly its execution time, in spite of its well-known computational inefficiency.
- The **coupling is a limiting** factor in the model. With the configuration used in this study, the IFS' time step is **slower** than NEMO's one. When IFS is "disabled", the execution finishes earlier, nevertheless, when NEMO is "disabled", the execution time does not change. This means that in coupled models, the whole **system is limited** by the **slowest component**.

References

[1] Hazeleger W. et al., 2011: "EC-Earth V2.2: description and validation of a new seamless earth system prediction model". Clim Dyn, doi:10.1007/s00382-011-1228-5

[2] Yepes-Arbós, X. et al., 2016: "Scalability and performance analysis of EC-Earth 3.2.0 using a new metric approach (Part II)". Barcelona Supercomputing Center

[3] Acosta, M.C. et al., 2016: "Performance analysis of EC-Earth 3.2: Coupling". Barcelona Supercomputing Center

[4] Barcelona Supercomputing Center. BSC performance tools, 2016: https://tools.bsc.es/

Barcelona Supercomputing Center
Centro Nacional de Supercomputación
EXCELENCIA SEVERO OCHOA

The research leading to these results has received funding from the EU Seventh Framework Programme (FP7) under grant agreement no 312979 (IS-ENES2) and from the EU H2020 Framework Programme under grant agreement no 675191 (ESiWACE).
This poster can be downloaded at **https://earth.bsc.es/wiki/doku.php?id=library:external:posters**
The report can be downloaded at **https://earth.bsc.es/wiki/doku.php?id=library:external:technical_memoranda**
Corresponding author: **xavier.yepes@bsc.es**

esiwace
CENTRE OF EXCELLENCE IN SIMULATION OF WEATHER AND CLIMATE IN EUROPE
is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING