Performance study of OpenIFS: towards a more efficiently scalable model

Mario C. Acosta[1], Xavier Yepes-Arbós[1,2], 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

Weather and climate models are **improving the accuracy of their simulations** with some factors such as the reduction of the dependence on parametrizations due to the increase of grid resolution. However, this accuracy improvement will need **more computational resources** through a **new generation of supercomputers**. To take advantage of this new generation, **performance analysis** could help to know in detail the **computational behavior** of the models and the information obtained could be used to introduce optimizations. The **optimizations** will improve the **energy efficiency** of the models when millions of CPU hours are used for their parallel execution.

OpenIFS is a model used by a large community and new optimizations will be needed for its massive parallel execution in the future. Similarly to previous experiences using profiling tools for **EC-Earth, NEMO and IFS** [1][2][3], in this study we present our **methodology** to know more about the computational behavior of OpenIFS. Additionally, we present how we are planning to improve the **functionality and performance of the Input/Output (I/O)** of OpenIFS.

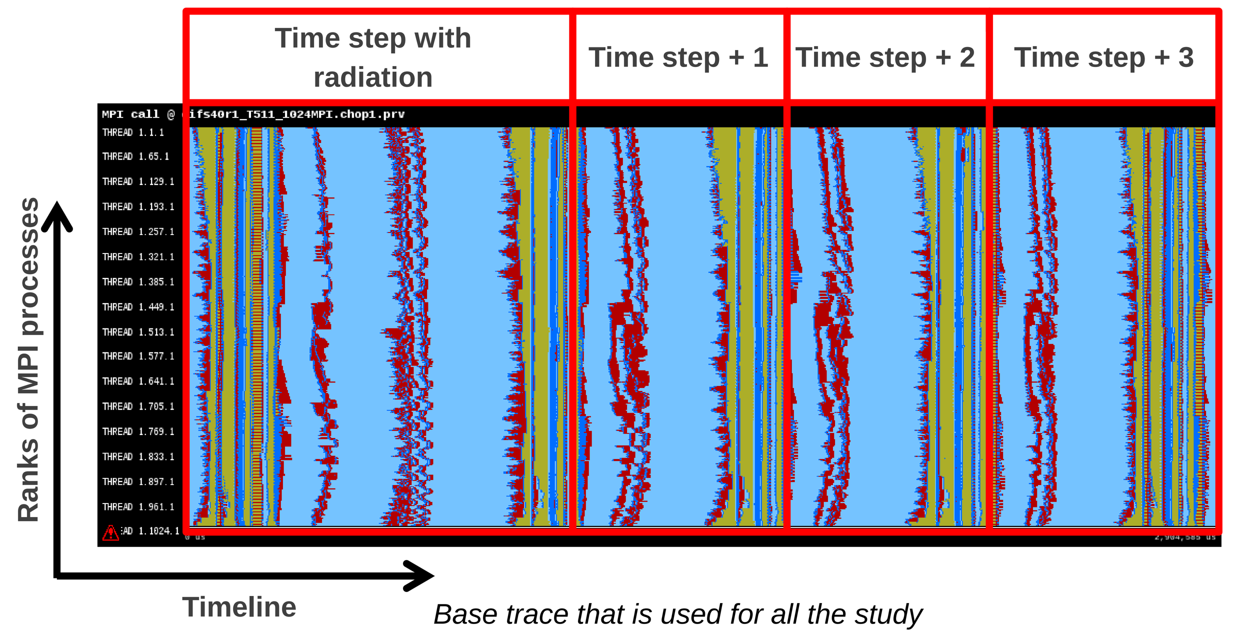
2. BSC tools

Performance tools [4] are essential to study the behavior of OpenIFS:

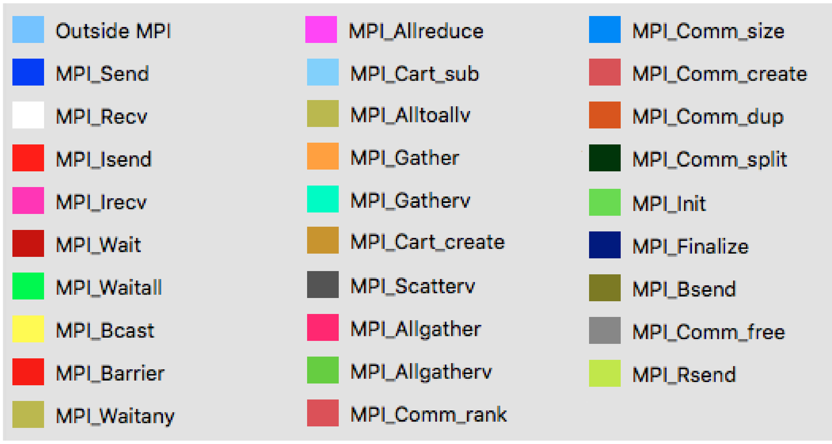
- **Extrae**: is a package used to instrument the code. It generates trace-files with hardware counters, MPI messages and other information.
- **Paraver**: is a browser used to analyze both visually and analytically trace-files.
- **Dimemas**: is a simulator based on traces to predict the behavior of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

3. What can you see in a trace?

The view of a trace consists of **threads or processes** on the Y axis and the **timeline** on the X axis. Particularly, in this poster some views are of **MPI functions**, where each type of call is identified by a color. Furthermore, all views are obtained from an execution that uses 1024 processes. The trace on the right has 1 time step with radiation plus 3 regular ones.



Base trace that is used for all the study



Legend with the colors representation of the MPI functions

4. Our methodology

Complex models require complex studies. Several approaches are needed to optimize an Earth System Model (ESM):

Mathematical study

- Discretization used (explicit, implicit, semi-implicit…)
- Parallel adaptation (solvers, preconditioners…)

Computational study

- Achieve load balance among components
- Reduce overhead introduced by parallel applications
- Assure that the computational algorithm takes advantage of the architecture

Profiling Study

- General profiling
- Profiling applied to ESMs

Introducing optimizations

- Improvement of the algorithm
- Change the method using a new approach

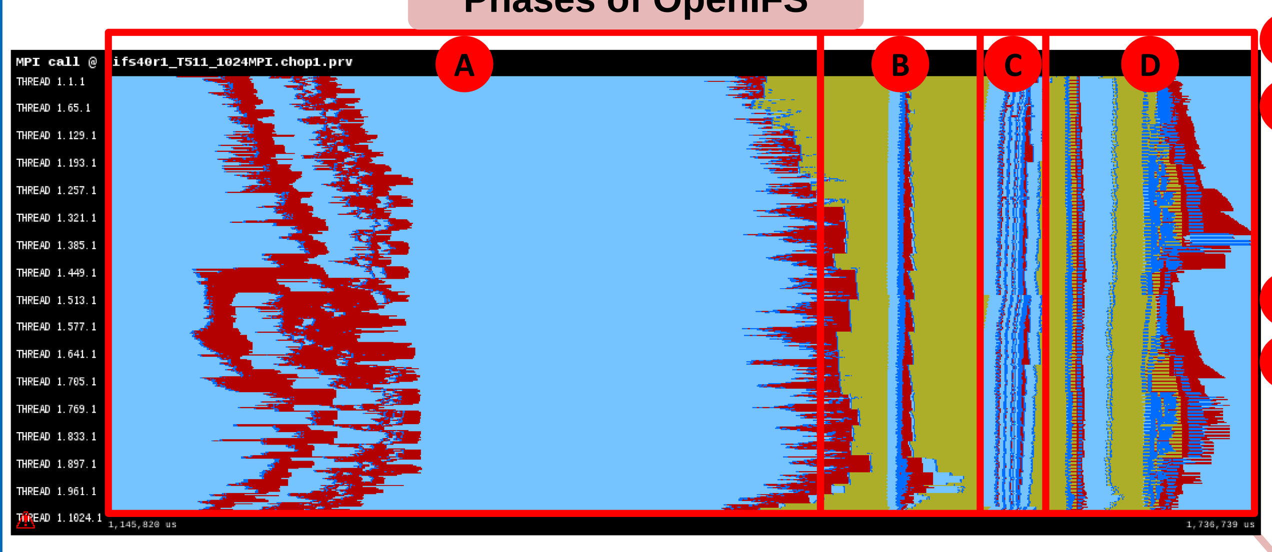
Reproducibility study

- Take into account the chaotic nature of climate models, evaluate if the accuracy and reproducibility of the model is similar using or not the optimizations proposed

5. Computational phases of OpenIFS

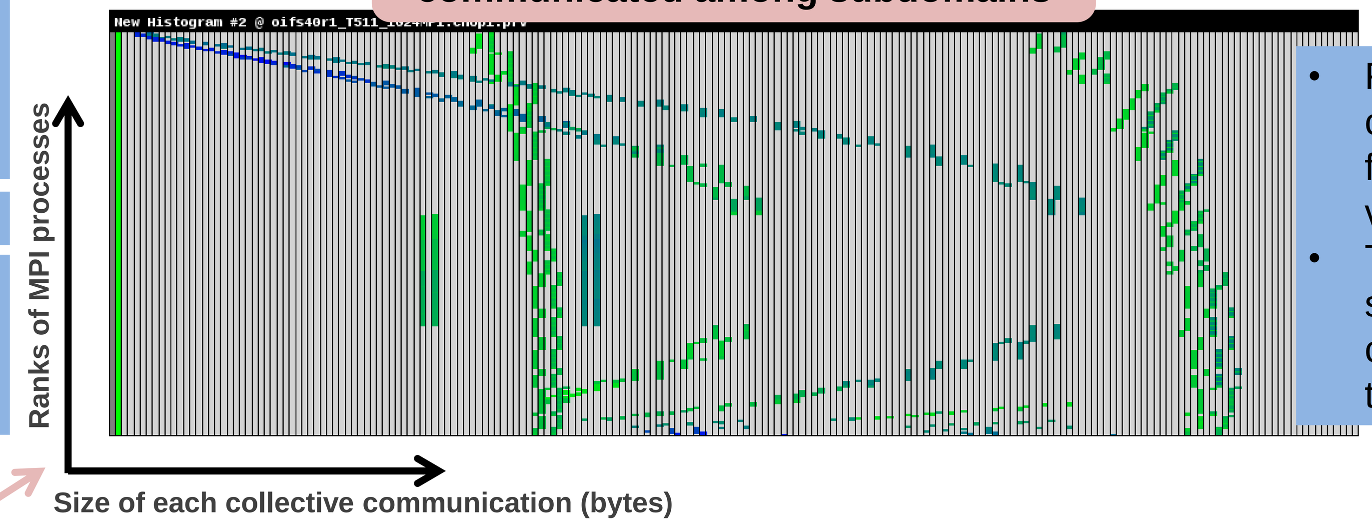
The computational algorithm of OpenIFS can be divided into different phases. Profiling tools can be used to highlight the main bottlenecks in the parallel execution of each phase.

Phases of OpenIFS



- A: Grid point calculations
- B: Direct transformations and transpositions (Fourier + Legendre)
- C: Spectral calculations
- D: Inverse transformations and transpositions (Fourier + Legendre)

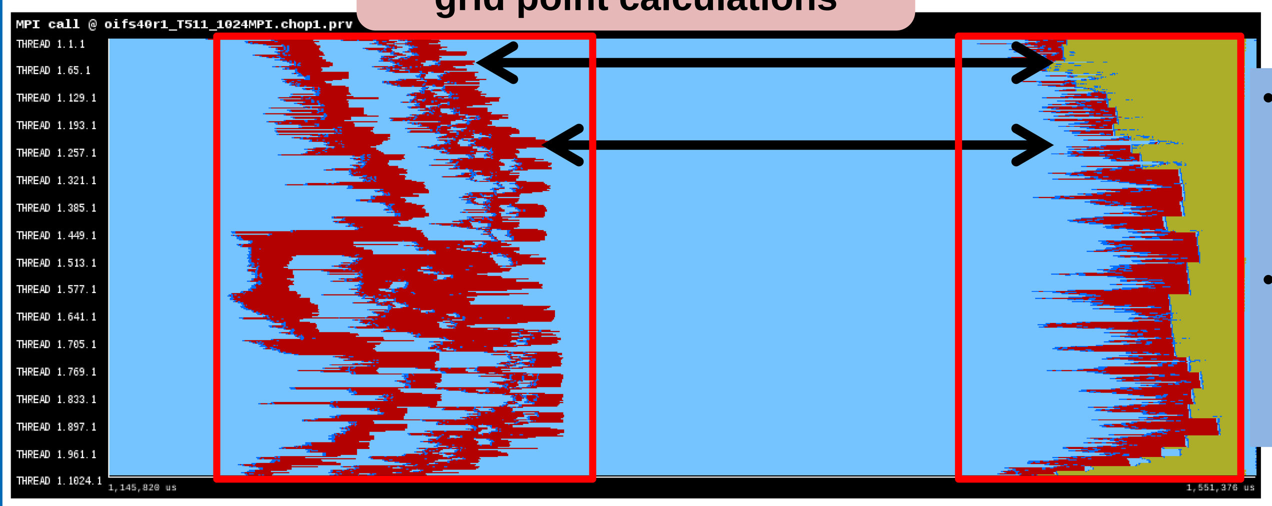
Imbalance in the number of bytes communicated among subdomains



- For a balanced communication, the figure should show vertical lines
- This means that some processes could spend more time communicating

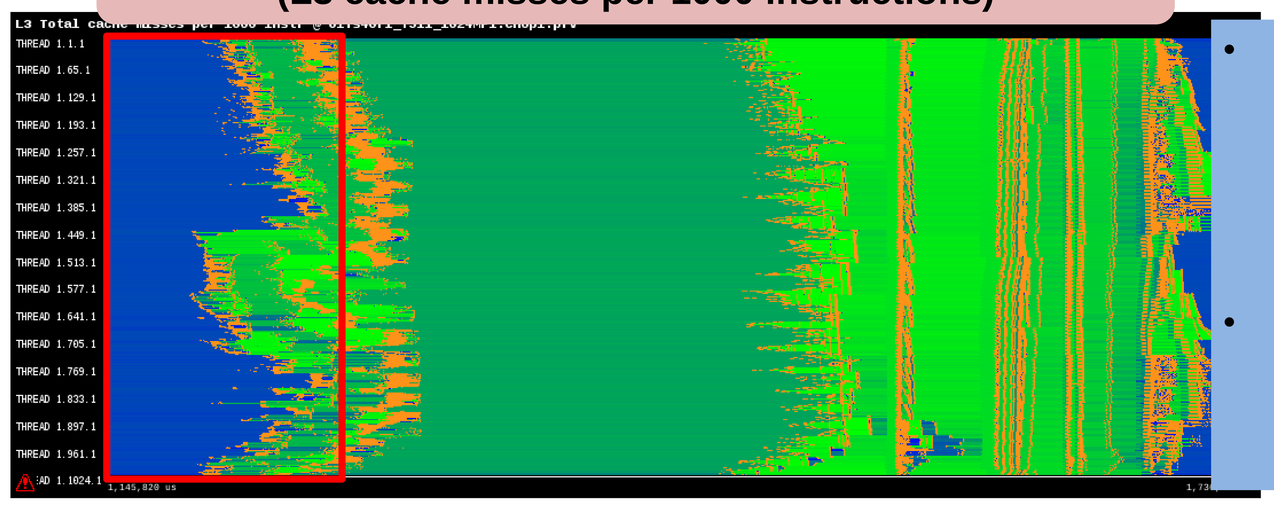
Some drawbacks can be highlighted from the study

Load imbalance due to the grid point calculations



- Black arrows of different lengths represent the load imbalance
- This means that some processes are working more than others

Memory access slowness in the grid point calculations (L3 cache misses per 1000 instructions)



- Blue color means that the access to memory is slower than other calculations (green)
- This means that calculations for this region are slower than others

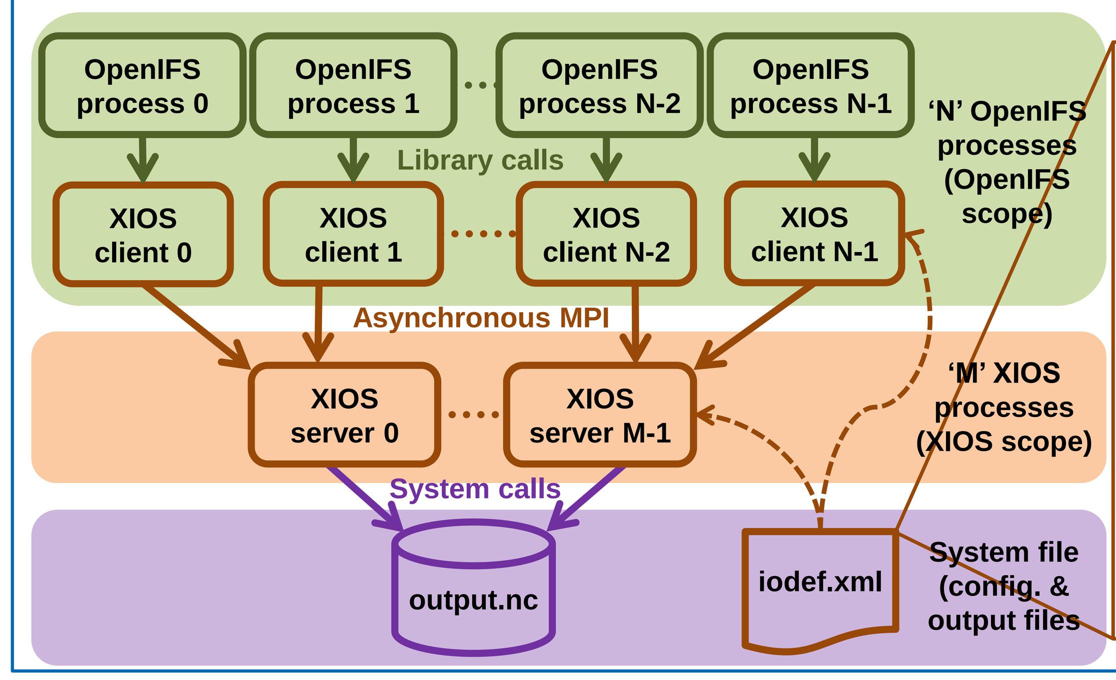
6. Improvements in the I/O

Although this last task is not strictly related to the performance study presented on the others points, they share the end-goal of improving OpenIFS. Particularly, it focuses on the I/O, and not only from a performance point of view, but also from a more user-friendly setup. Therefore, this ongoing task pursues to **improve the I/O** of OpenIFS with a twofold goal:

- Increase the **functionality and usability of EC-Earth**, since in a future it will use OpenIFS as the atmospheric component.
- **Speedup the throughput** (more performance).

The approach to achieve it is using **XIOS** (XML Input/Output Server). It is able to decouple the direct writing of a model using several servers communicated through **asynchronous MPI** communications. Thus, the **model does not have to wait** for the I/O to be done and can continue with its simulation. Depending on the amount of data to be written, the use of XIOS can lead to an increase of performance.

On the other hand, the XIOS **output configuration** is done through an **XML file**, offering simplicity to users: a friendly **post-process** definition and online diagnostics generation. In addition, unlike the current solution (**FullPos**), XIOS writes in **NetCDF format**, which is a **requirement of CMIP6**.




```
<grid id="grid_3d">
  <domain id="domain_2d" type="rectilinear" ni_glo=… />
  <axis id="axis_1d" n_glo=… />
</grid>

<field_definition>
  <field id="temp" grid_ref="grid_3d" />
  <field id="pressure" grid_ref="grid_3d" />
</field_definition>

<file_definition>
  <file name="output" freq_output="1mo">
    <field field_ref="temp" />
    <field field_ref="pressure" />
  </file>
</file_definition>
```

References

[1] 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 (BSC)
[2] Tintó Prims, O. et al., 2015: "Optimization of an ocean model using performance tools". BSC
[3] Acosta, M.C. et al., 2016: "Performance analysis of EC-Earth 3.2: Coupling". BSC
[4] Barcelona Supercomputing Center. BSC performance tools, 2017: https://tools.bsc.es/

The research leading to these results has received funding from the EU H2020 Framework Programme under grant agreements no 675191 (ESiWACE) and no 641727 (PRIMAVERA).
This poster can be downloaded at **https://earth.bsc.es/wiki/doku.php?id=library:external:posters**
Corresponding authors: **mario.acosta@bsc.es** and **xavier.yepes@bsc.es**

Barcelona Supercomputing Center – Centro Nacional de Supercomputación
EXCELENCIA SEVERO OCHOA
esiwace – CENTRE OF EXCELLENCE IN SIMULATION OF WEATHER AND CLIMATE IN EUROPE
PRIMAVERA