



**Barcelona
Supercomputing
Center**
Centro Nacional de Supercomputación



eScience center

(Open)IFS-XIOS integration for the future EC-Earth 4 version: the benefits of outputting CMORized netCDF files

Xavier Yepes-Arbós, Mario C. Acosta, Gijs van den Oord and Glenn Carver



The research leading to these results has received funding from the EU H2020 Framework Programme under grant agreement no. 641727

This material reflects only the author's view and the Commission is not responsible for any use that may be made of the information it contains

23/10/2018

Index

1. Introduction
2. How will EC-Earth benefit from XIOS?
3. IFS-XIOS integration and optimization overview
4. Performance evaluation
5. Conclusions

1. Introduction

eScience center



**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación

Introduction

- Earth system models have benefited of the exponential growth of supercomputing power
- This allows to use more complex computational models to find more accurate solutions
- As a consequence, the generated amount of data has grown considerably
- However, since the I/O was not significant enough in the past, not much attention was paid to improve it

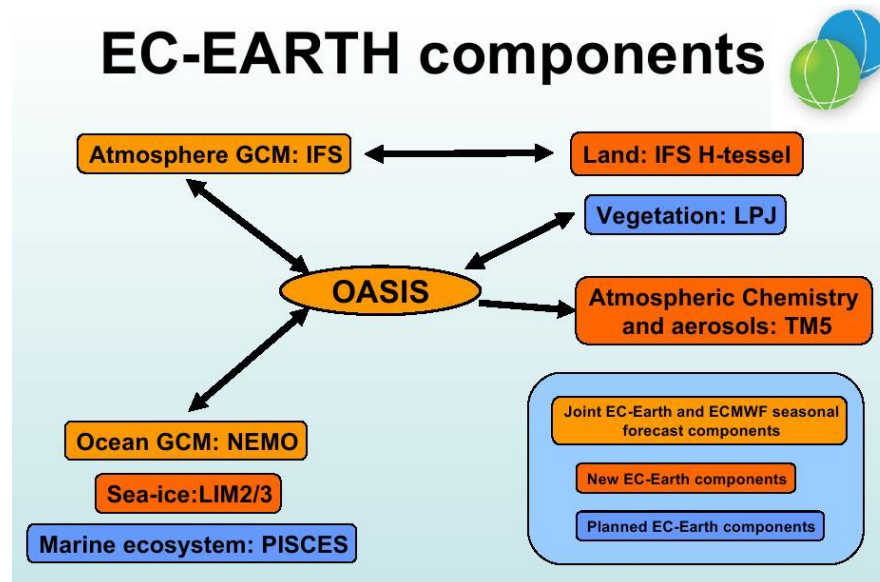
Introduction

- Earth system models have benefited of the exponential growth of supercomputing power
- This allows to use more complex computational models to find more accurate solutions
- As a consequence, the generated amount of data has grown considerably
- However, since the I/O was not significant enough in the past, not much attention was paid to improve it

 is no exception!

EC-Earth overview

- EC-Earth 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 as the atmospheric model and NEMO as the ocean model



The I/O problem in EC-Earth

- In particular, the IFS version of EC-Earth is experiencing an I/O bottleneck
- EC-Earth has been recently used to run experiments using the T511L91-ORCA025L75 configuration under the H2020 PRIMAVERA project
- Experiments require to output a lot of fields, causing a considerable slowdown in the EC-Earth execution time
- I/O in IFS represents about 30% of the total execution time

IFS overview

- The Integrated Forecast System (IFS) is a global data assimilation and forecasting system developed by the European Centre for Medium-Range Weather Forecasts (ECMWF)
- It has two different output schemes:
 - The Météo-France (MF) I/O server which is fast and efficient from a computational point of view. It is only used at ECMWF, such its operational forecasts
 - A sequential I/O scheme which is slow and inefficient from a computational point of view. It is used by non-ECMWF users, this is, in OpenIFS and in the IFS version of EC-Earth

IFS overview

- The inefficient sequential I/O scheme of IFS requires a serial process:
 - Gather all data in the master process of the model
 - Then, the master process sequentially writes all data
- This is not scalable for higher grid resolutions, and even less, for future exascale machines

Objective

- Taking advantage that NEMO is already outputting data through XIOS, we chose to integrate XIOS into IFS as well
- The XML Input/Output Server (XIOS) is an asynchronous MPI parallel I/O server developed by the Institute Pierre Simon Laplace (IPSL)
- The use of XIOS has the objective of improving the computational performance and efficiency of IFS (by extension EC-Earth), and thus, reduce the execution time
- Moreover, it has a series of additional benefits (explained in next section)

European collaboration

- Netherlands eScience Center (NLeSC)/Koninklijk Nederlands Meteorologisch Instituut (KNMI)
- ECMWF

2. How will EC-Earth benefit from XIOS?

eScience center



**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación

Current EC-Earth workflow

- EC-Earth runs experiments that have different tasks in their workflows
- Post-processing task is sequentially executed after the simulation task



Critical path = Pre-processing + Simulation + Post-processing

Post-processing in EC-Earth

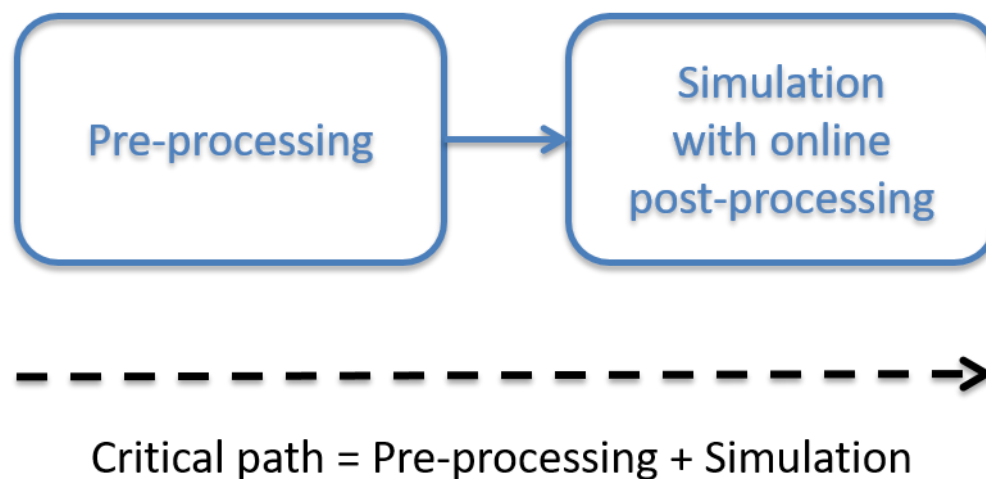
- IFS is originally developed for weather forecasting, so it writes using the GRIB format to meet critical time-to-solution requirements
- When IFS is used in EC-Earth for climate modeling, post-processing is needed to:
 - Convert GRIB files to netCDF files
 - Transform data to be CMIP-compliant (CMORization)
 - Compute diagnostics
- Post-processing turns into an expensive process

XIOS' key role

- The current operations performed in the EC-Earth post-processing task are avoidable, since XIOS has these features:
 - Output files are in netCDF format
 - Written data is CMIP-compliant (CMORized)
 - It is able to post-process data online to generate diagnostics
- Thus, the use of XIOS in EC-Earth has a twofold effect:
 - Improve the computational performance and efficiency of the model, and thus, reduce the execution time (previously mentioned)
 - Reduce the critical path of its workflow by avoiding the post-processing task

Future EC-Earth workflow

- Critical path will be shortened by concurrently running post-processing with the simulation
- Simple workflows are less prone to have configuration errors



Other benefits

- The configuration of XIOS is more intuitive than the current approach. It is done through an XML file
- Data compression
- Save storage space because it will be only stored processed data ready to be used

3. IFS-XIOS integration and optimization overview

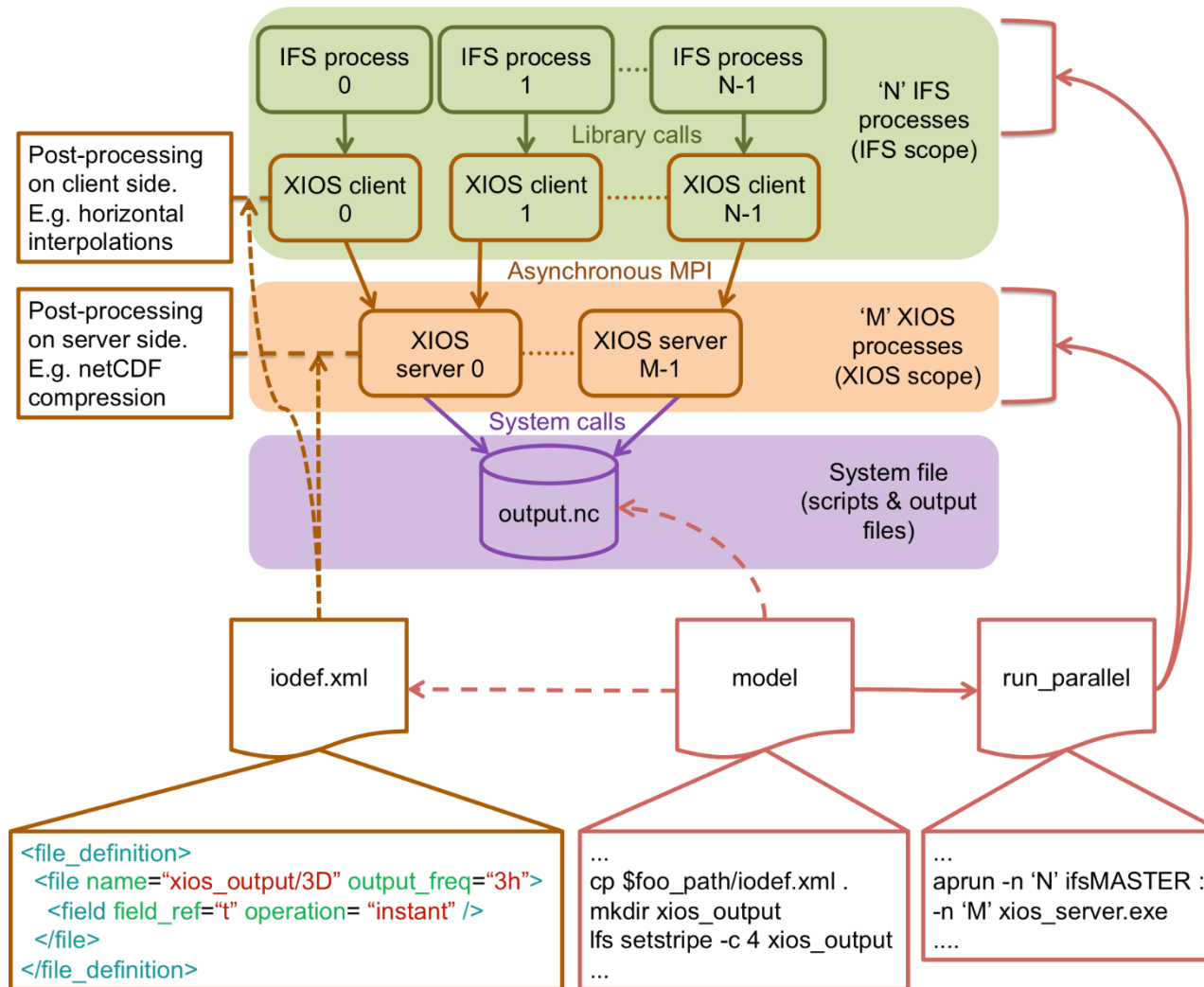
eScience center



**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación

Scheme of the IFS-XIOS integration



Development steps

- XIOS setup
 - Initialization
 - Finalization
 - Context: calendar and geometry (axis, domain and grid)
 - *lodef.xml* file
- Grid-point fields transfer
 - NPROMA blocks gather
 - Send fields
- Environment setup
 - XIOS compilation
 - Include and link XIOS, netCDF and HDF5
 - Model script
 - Supporting MPMD mode
- FullPos integration to support vertical post-processing: grid-point fields only

Development steps

- XIOS setup
 - Initialization
 - Finalization
 - Context: calendar and geometry (axis, domain and grid)
 - *lodef.xml* file
- Grid-point fields transfer
 - NPROMA blocks gather
 - Send fields
- Environment setup
 - XIOS compilation
 - Include and link XIOS, netCDF and HDF5
 - Model script
 - Supporting MPMD mode
- • **FullPos integration** to support vertical post-processing: grid-point fields only

FullPos integration

- FullPos is a post-processing package currently used by IFS
- The use of FullPos is necessary to perform vertical interpolations not supported by XIOS
- It is called as usual, and afterwards, data is sent to XIOS
- For now, it is only possible to output grid-point fields

Optimization techniques used

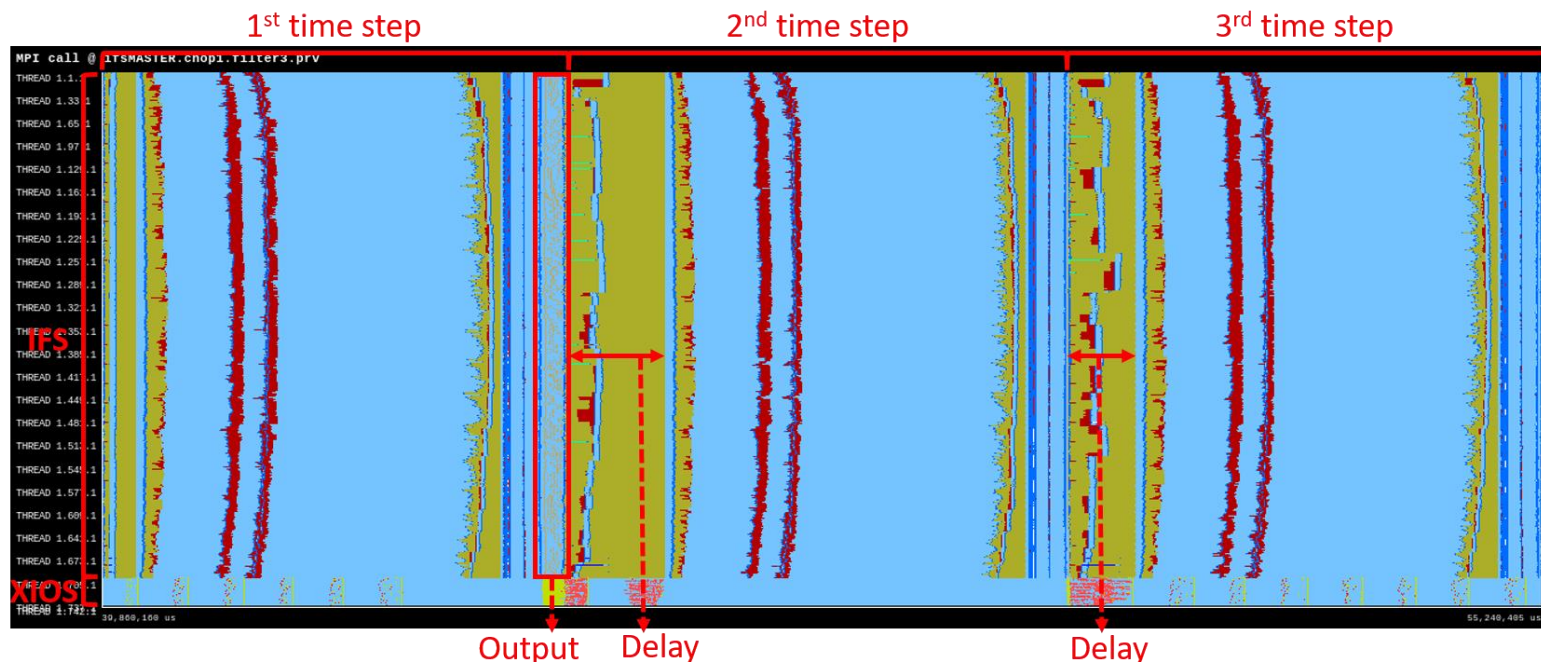
- Parallelization using OpenMP threads
- Optimized compilation of XIOS with `-O3`
- Computation and communication overlap

Optimization techniques used

- Parallelization using OpenMP threads
- Optimized compilation of XIOS with `-O3`
- • Computation and communication overlap

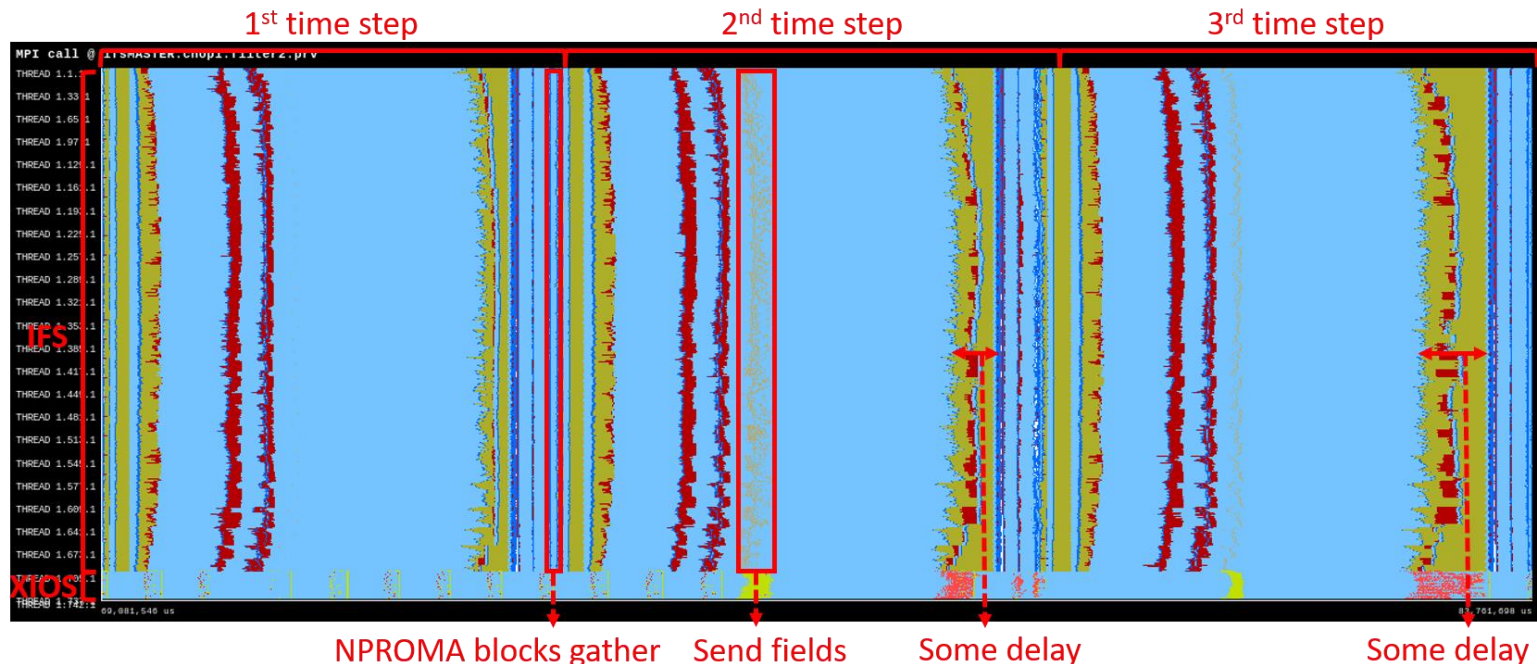
Computation and communication overlap

- The trace shows that after an output time step, there is a delay in the communications of the next two time steps (*MPI_Waitany* and *MPI_Alltoallv*)
- There is a conflict between intra IFS communications and IFS to XIOS communications



Computation and communication overlap

- The sending of data to XIOS is delayed to truly overlap computation and communication
- The trace shows that there is no delay at the beginning of the 2nd and 3rd time steps. However, there is some delay at the end, but it is less significant



4. Performance evaluation

eScience center



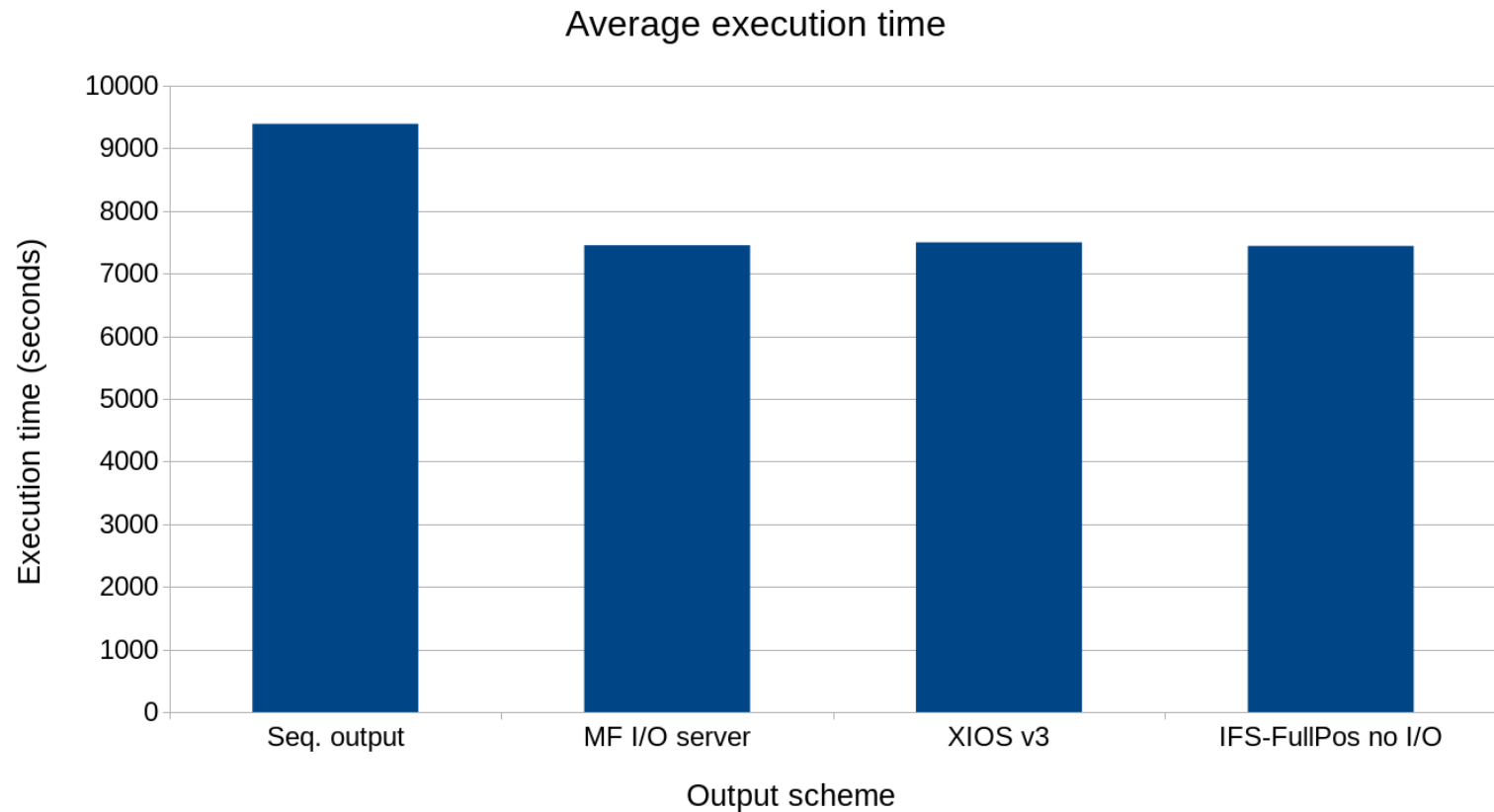
**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación

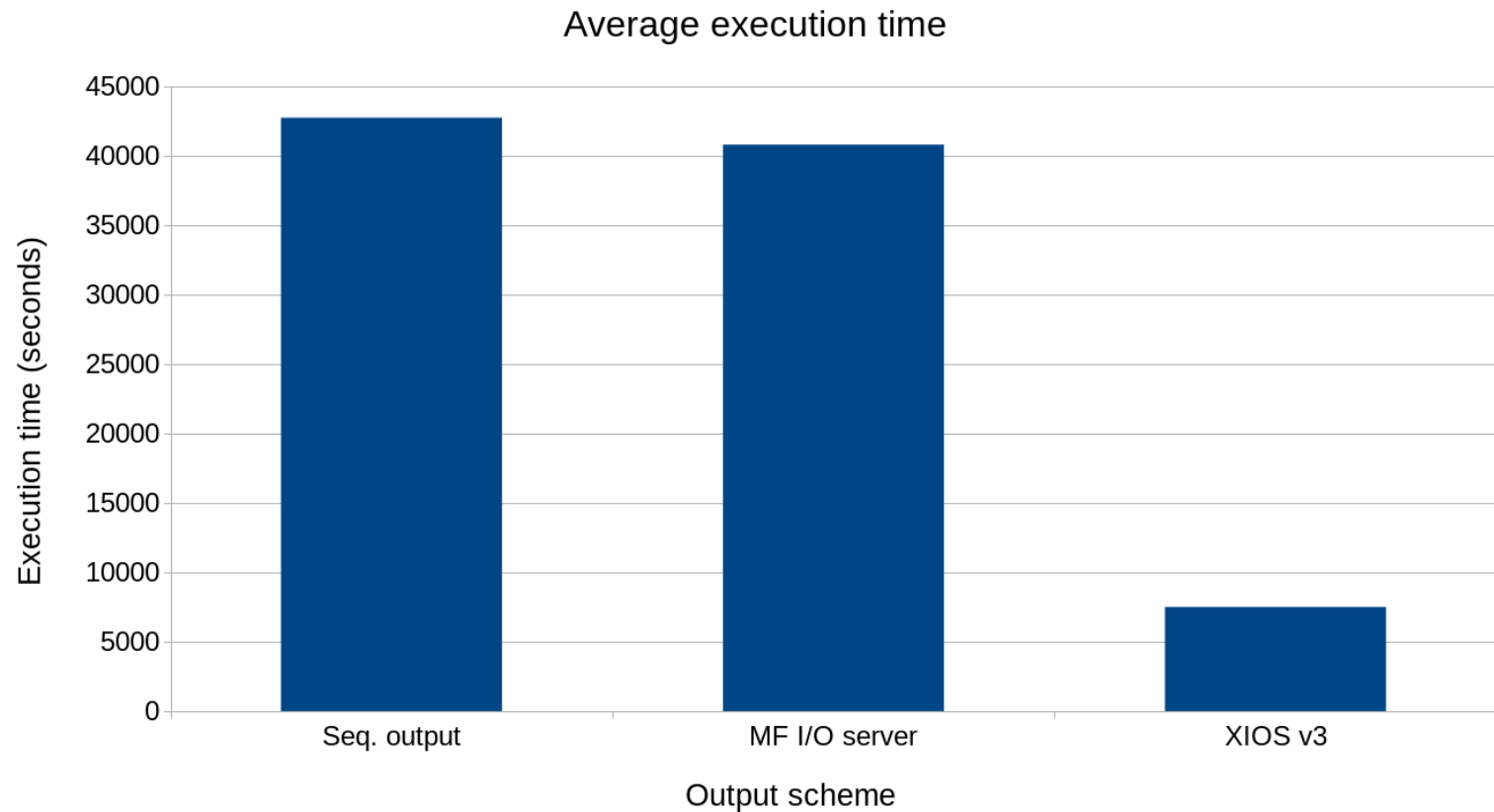
Execution overview

- ECMWF HPC platform (Cray XC40)
- IFS CY43R3
- Octahedral reduced Gaussian grid. Horizontal resolution: T1279 (16 km)
- 702 MPI processes, each with 6 OpenMP threads
- 10 days of forecast with a time step of 600 seconds
- Output frequency: 3 hours
- Only grid-point fields. NetCDF files size: 2.5 TB

Comparison test



Comparison test adding GRIB to netCDF post-processing



5. Conclusions

eScience center



**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación

Conclusions

- The presented development is easy-to-use
- The integration with no optimizations already improved the execution time:
 - Sequential output 9391 seconds (20.7% of overhead) → IFS-XIOS integration 7682 seconds (3.1% of overhead)
- The integration with optimizations is fast and efficient:
 - It only has a 0.7% of overhead
 - Within 56 seconds IFS outputs 2.5 TB of data
- When post-processing to convert GRIB to netCDF files is taken into account:
 - The post-processing takes 9.2 hours (sequentially performed, as in EC-Earth)
 - Thus, the most optimized version is a 5.7x faster than the sequential output and a 5.4x faster than the MF I/O server

Conclusions

- These numbers denote that the development is scalable and efficient as well as will address the I/O issue
- Summary of benefits for EC-Earth 4:
 - Increase the performance and efficiency of the whole model
 - Online diagnostics computation
 - CMORized netCDF files
 - Data compression
 - Simpler output configuration file using XML syntax
 - Experiments with simpler workflows
 - Save thousands of computing hours and storage space

Conclusions

- These numbers denote that the development is scalable and efficient as well as will address the I/O issue
- Summary of benefits for EC-Earth 4:
 - Increase the performance and efficiency of the whole model
 - Online diagnostics computation
 - CMORized netCDF files
 - Data compression
 - Simpler output configuration file using XML syntax
 - Experiments with simpler workflows
 - Save thousands of computing hours and storage space

Save money!

Ongoing and future work

- Transform fields from spectral space to grid-point space, post-process and send them to XIOS
- The development done for IFS will be ported to OpenIFS (next release)
- Adapt the future EC-Earth 4 version to output fields and compute online diagnostics from OpenIFS and NEMO components through XIOS



**Barcelona
Supercomputing
Center**
Centro Nacional de Supercomputación



eScience center

Thank you!



PRIMAVERA

xavier.yepes@bsc.es

Additional slides

eScience center

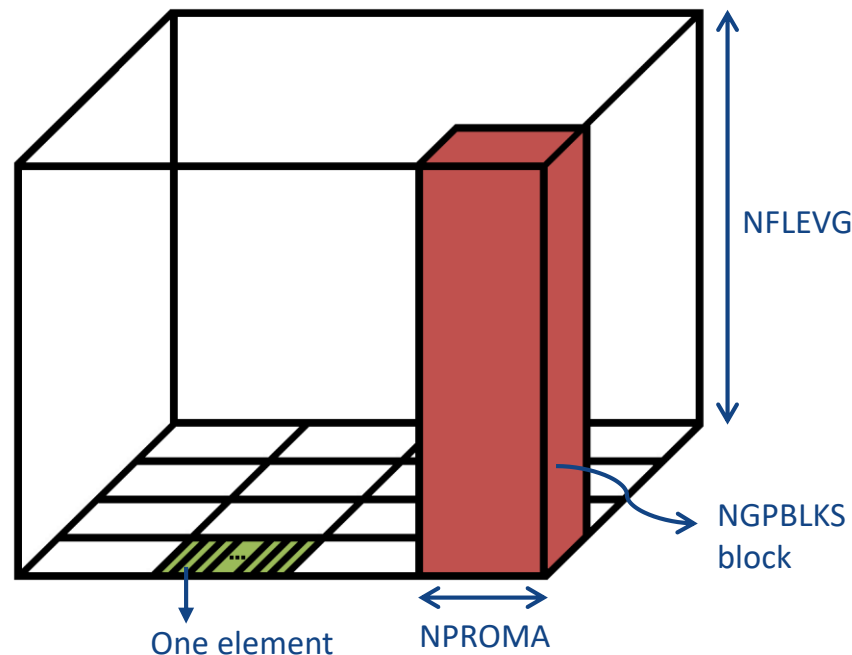


**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación

Subdomain decomposition in IFS

- IFS uses a blocking strategy to efficiently parallelize the manipulation of data arrays using OpenMP
- `IFS_data_array(NPROMA, NFLEVG, NFIELDS, NGPBLKS)`



NPROMA blocks gather

- The IFS data arrays do not match with the XIOS ones:
 - IFS_data_array(NPROMA, NFLEVG, NFIELDS, NGPBLKS)
 - XIOS_data_array(unidimensional 2D domain, NFLEVG)
- It is necessary to re-shuffle fields data before sending them
- According to the blocking strategy used in IFS, an XIOS-style array has to be built by gathering NPROMA blocks

Optimal number of XIOS servers

