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

Final General Assembly

16-18 January 2023 Toulouse

CPMIP: Synthesis and main lessons for CMIP7

Mario Acosta (BSC)
Jean-Claude André (CERFACS)

IS-ENES3 has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

Final overview

- **IS-ENES3 provided a unique opportunity to exploit CPMIP metrics**, performing for the first time a complete computational/energy analysis and the creation of a novel data-base based on CMIP6 experiments, using the different models and platforms available all across Europe.

- The outcome of this work was published in the D4.3 deliverable (1) including:
  - The possibilities for collaboration with other groups (ES-DOC, HPC-TF and Carbon Footprint G.)
  - The analysis illustrating some practical examples, and proving the usefulness of the metrics to the community.
  - Main difficulties encountered in the coordination of the collection, including general recommendations on how to solve these problems for future collections and analyzes.

(1) Mario Acosta et al. 2021, ISENES3 D4.3: CPMIP performance metrics evaluation for CMIP6 and community advice. https://doi.org/10.5281/zenodo.6394049

The IS-ENES3 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

CPMIP: Community List

Model / Institution	People Involved
CNRM-CM6	Sophie Valcke, Marie Pierre Moine
IPSL-CM	Arnaud Caubel
EC-Earth	Mario Acosta, Uwe Fladrich, Philippe Le Sager
MetO	Harry Shepherd, JC Rioual
CMCC	Italo Epicoco, Silvia Mocavero
MPI-M-DKRZ	Maria Moreno, Reinhard Budich, Joachim B.
U. Read	Grenville Lister, Bryan Lawrence
Nor-ESM	Alok Kumar Gupta
TOPAZ/MOM5	Paulo Nobre
GFDL	Niki Zadeh

Include 11 models with 32 CMIP6 configurations (AMIP, OCE, Coupled, different resolutions…)

The IS-ENES3 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

Final overview

- The outcome of this work was published in the D4.3 deliverable including:
  - **The possibilities for collaboration with other groups (ES-DOC, HPC-TF and Carbon Footprint G.)**
  - The analysis illustrating some practical examples, and proving the usefulness of the metrics to the community.
  - Main difficulties encountered in the coordination of the collection, including general recommendations on how to solve these problems for future collections and analyzes.

The IS-ENES3 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

Final overview

- The outcome of this work was published in the D4.3 deliverable including:
  - The possibilities for collaboration with other groups (ES-DOC, HPC-TF and Carbon Footprint G.)
  - **The analysis illustrating some practical examples, and proving the usefulness of the metrics to the community.**
  - Main difficulties encountered in the coordination of the collection, including general recommendations on how to solve these problems for future collections and analyzes.

The IS-ENES3 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

Analysis and results summary

- The analysis illustrated some practical examples, and proving the usefulness of the metrics to the community.
  - Resolution impact
  - Complexity impact
  - Data output impact
  - ASYPD: Queue time and interruptions impact
  - Coupling impact

- In a previous work, we also studied a specific model (EC-Earth) to evaluate the computational efficiency on different machines or configurations
  - Complexity Impact: Identify which component is the bottleneck of the coupled version
  - ASYPD: Queue time could differ between machines, due to the different set-up of the queue systems
  - Comparison through machines: Detect bottlenecks according to the limitations of each hardware

The IS-ENES3 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

Final overview

- The outcome of this work was published in the D4.3 deliverable including:
  - The possibilities for collaboration with other groups (ES-DOC, HPC-TF and Carbon Footprint G.)
  - The analysis illustrating some practical examples, and proving the usefulness of the metrics to the community.
  - **Main difficulties encountered in the coordination of the collection, including general recommendations on how to solve these problems for future collections and analyzes.**

The IS-ENES3 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

CMIP7 path

- **Main difficulties encountered in the coordination of the collection, including general recommendations on how to solve these problems for future collections and analyzes.**
  - **Performance metrics collection is secondary. Facilitate the collection should help to the institutions.**
  - **A coordinated collection has been proved useful to ensure to get the metrics and solve possible inconsistencies or gaps during the process.**
  - **Create a finer granularity for some of the metrics could be important to improve the analysis in the future.**

The IS-ENES3 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

CMIP7 path

- Performance metrics collection is secondary. Facilitate the collection should help to the institutions.
  - Integrate the collection through automatic methods and workflow managers
  - Perform this work for real models (IFS, NEMO, FESOM, ICON) and real workflow managers Autosubmit, ECflow for projects as ESiWACE3, EERIE and other contracts.
  - Facilitate the integration for other institutions as a service through ENES-RI, covering models and platforms as much as possible.

The IS-ENES3 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

Our workflow

EC EARTH

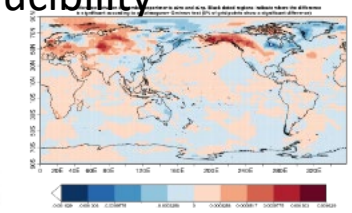
GitLab

Develop

Checkout

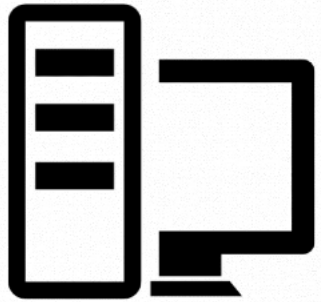
Continuous Delivery Testing – Reproducibility

Jenkins



production

trunk

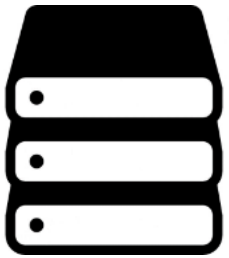


Run experiment

AUTOSUBMIT

Performance metrics

ESGF Earth System Grid Federation



Archive & publish

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

CMIP7 path

- A coordinated collection has been proved useful to ensure to get the metrics and solve possible inconsistencies or gaps during the process.
  - Provide a new coordinated collection for CMIP7 through ENES-RI
  - Use the new data-base and work done as starting point
  - Learn from the experience and main gaps during the process to improve the collection for the next iteration
  - Ensure the collection through different frameworks to facilitate and improve the work (ESiWACE3, EERIE and other contracts).
    - Create a common framework using real models to evaluate different platforms
      - HPCW: High Performance Climate and Weather benchmark suite (ESCAPE2)
        - ESiWACE3 will make possible the integration of the CPMIP for a common benchmark suite

The IS-ENES3 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

CMIP7 path

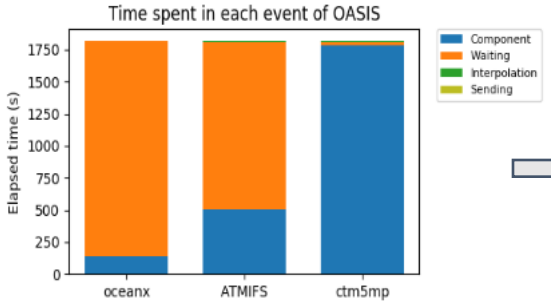
- Create a finer granularity for some of the metrics could be important to improve the analysis in the future.
  - Save information about the different components individually (for a coupled model) could be important for the analysis and improvement.
  - Store data of sub-metrics using ES-DOC to be part of the new data-base.
  - Create new metrics, methods or tools to provide this finer granularity in the most simple and automatic approach for the users, being compatible with the models and couplers used by the community.

The IS-ENES3 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

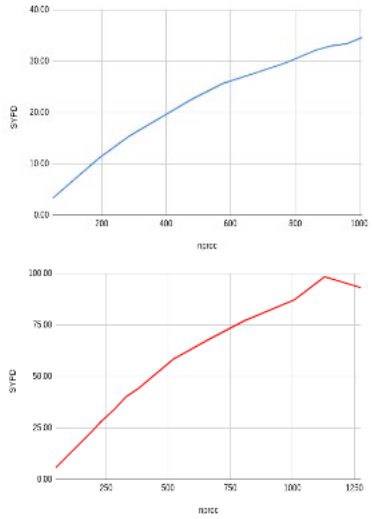
CMIP7 path

load-balancing method

An experiment with a poor load-balance



Get components' scalability properties



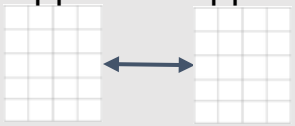
Automatic iterative process

1. Run the ESM on the HPC machine

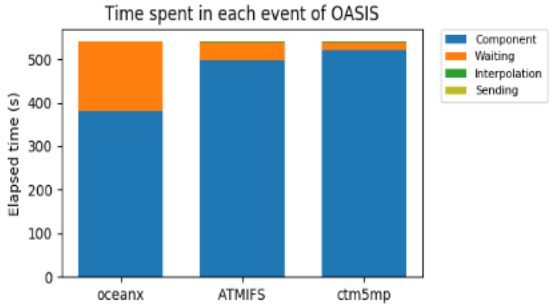
2. Get performance metrics

3. Modify the processors allocation

App. 1 App. 2



A balanced experiment



The IS-ENES3 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

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

Final General Assembly

16-18 January 2023 Toulouse

A synthetic view of compute resources used for CMIP6

Jean-Claude ANDRE (ENES HPC Task Force)

IS-ENES3 has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

The synthetic view

CMIP6 experiments Institutions/Models	SY sent to ESGF	Total SY including additional runs	SY (%) (ESGF/total)	Data sent to ESGF (PB)	Total data produced (PB) including additional data	Data (%) (ESGF/total)	CH (Mh) for SY sent to ESGF	Total CH (Mh) including additional runs	CH (%) (ESGF/total)
EC-Earth	28105	38854	72.3	0.8	1.405	56.9	31.3	46.5	67.3
CNRM-CERFACS	47000	110000	42.7	0.8	2.48	32.2	160	365	43.8
IPSL	75000	165000	45.5	1.8	7.6	23.7	150	320	46.9
CMCC	965	1926	50.1	0.27	1.46	18.5	1.99	4.34	45.8
UKMO	59000	117764	50.1	1.2	13.96	8.6	683	1491	45.8
NERC	640	1277	50.1	0.46	2.49	18.5	55.50	121.2	45.8
NCC-NORESM2	34443	68749	50.1	0.32	1.1	29.1	27.23	80	34.0
MPI	24175	35000	69.1	1.92	10.38	18.5	18.5	35.61	45.8
DKRZ	1276	1321	96.6	0.29	1.57	18.5	5.52	5.90	93.6
	270604	539891	50.1 %	7.86	42.445	18.5 %	1131	2470	45.8 %
CMIP5	93000		?	1.05	6.9+	15.2	46.8		?

Additional runs/data: development, tuning, extras, not sent to ESGF
In red: reconstructed missing number (using ratios from other groups)

Caviat: other groups contributed to the effort and sent data to ESGF (SY: +20%), but this could not be accounted for here. The global picture is hopefully still valid

CMIP6, between CMIP5 and CMIP7: some remarks

CMIP6 - When compared to output finally sent to ESGF
- it takes twice as much resources for additional work (CH, SY)
- less than ¼ of the data are finally sent to ESGF

CMIP6 - Energy aspects (MWh, rough order-of-magnitude estimates !)
- Runs production: 10^4; Data transfer: 10^3; Local data storage: 10^3

CMIP5 to CMIP6
- number of SY multiplied by ~3 (larger ensembles ?)
- number of CH multiplied by 20 to 50, *i.e.* a mean SY requires 10-to-20 more computing resources (increased resolution, more complex models)
- amount of data transfered multiplied by ~8 , *i.e.* a mean SY produces 3 times as much data (increased resolution)

CMIP7 ?
- up to 1-to-2 order of magnitude increase of CH (exascale for climate models ?)
- relatively stable number of SY, but likely much larger increase of data amount

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

THE CONSORTIUM

Coordinated by **CNRS-IPSL**, the IS-ENES3 project gathers **22 partners** in **11 countries**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°824084

Our website
https://is.enes.org/

Follow us on Twitter !
@ISENES_RI

Contact us at
is-enes@ipsl.fr

YouTube
Follow our channel
IS-ENES3 H2020