Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Preparing NEMO and EC-Earth models for very high-resolution production experiments

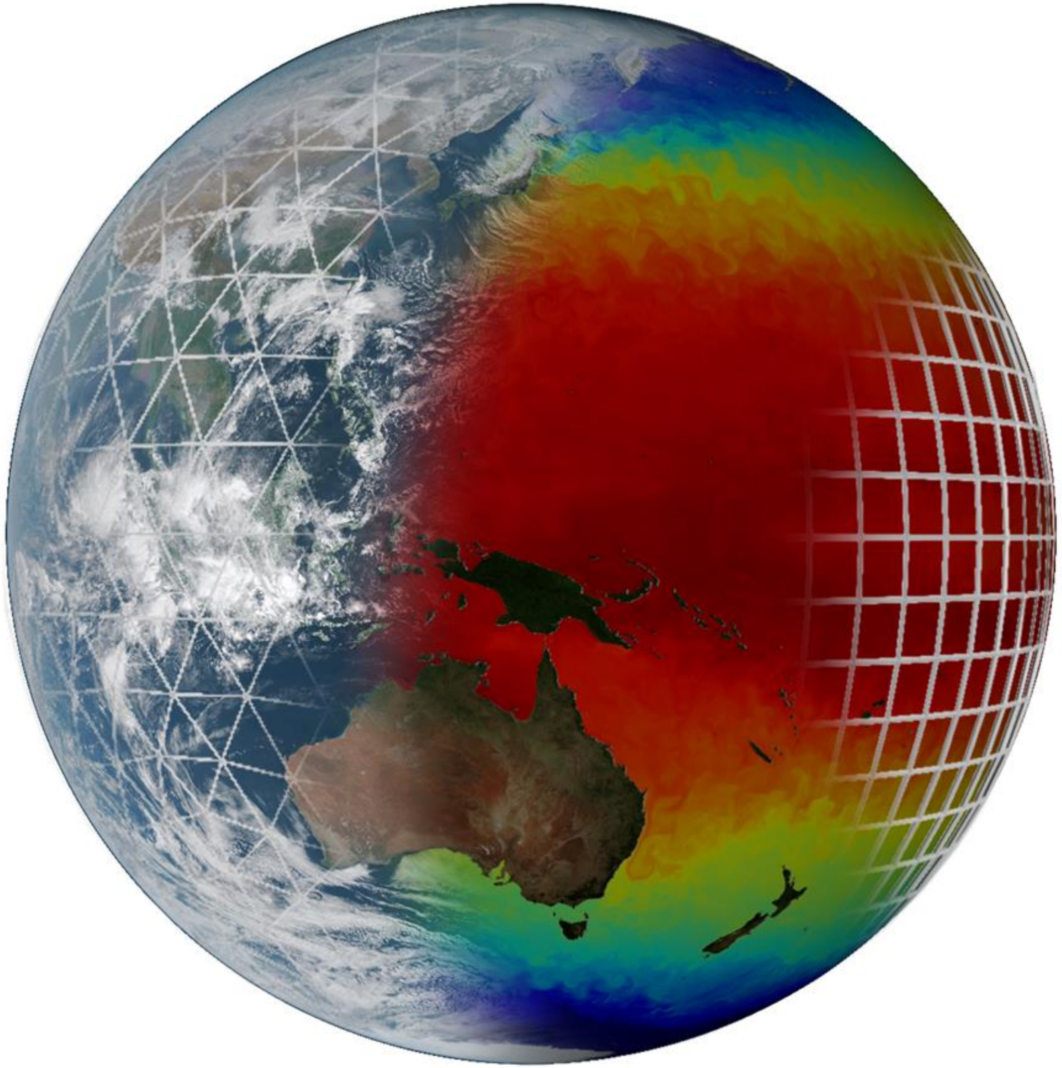
Miguel Castrillo (BSC), Dorotea Iovino (CMCC), Clement Bricaud (Mercator Ocean)

27/05/2021

Summer School on Effective HPC for Climate and Weather

The EC-Earth model

EC EARTH



Atmosphere:

IFS

ECMWF

Ocean - ICE:

NEMO - LIM

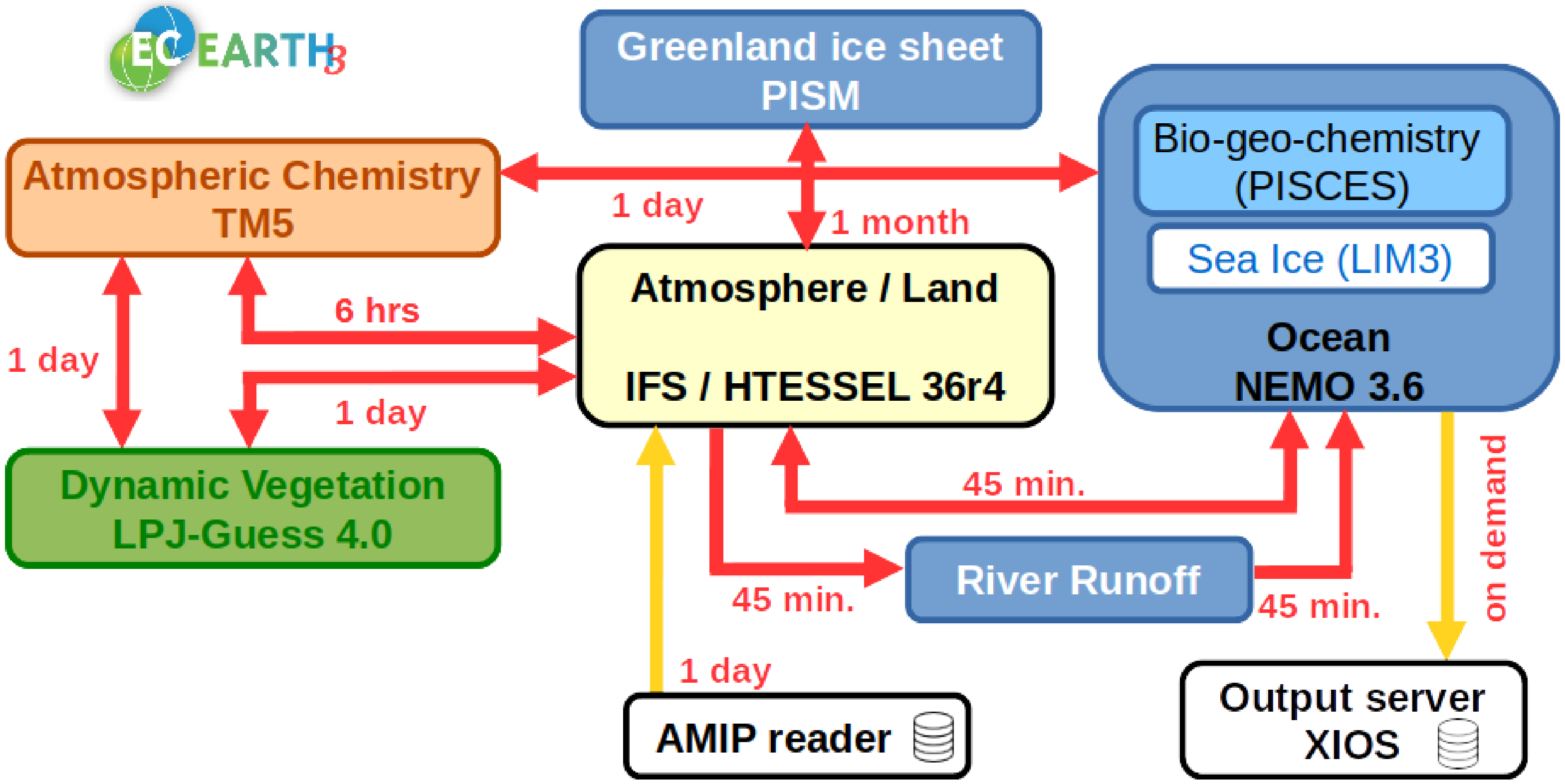
NEMO

Coupler:

Oasis

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EC-Earth structure



~30 partner institutes

8 core partners

KNMI, AEMET, DMI, Met Éireann, FMI, IPMA, CNR-DTA, SMHI

Workgroups

Technical
Tuning and CMIP
Atmospheric Composition
Climate Prediction
Land Vegetation
Ocean
Paleoclimate
EC-Earth 4

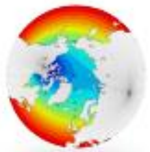
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Efficiency in Earth Science models

- Especially **critical** in Earth science models
- Simulations use a huge amount of computational **resources**
  - **Data** handling becoming an increasing bottleneck.
- Future simulations will need much more resources

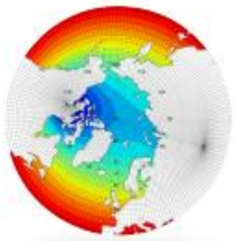
ORCA 2
550 MB of memory
8 CPU hours
10 Gigabytes of output (daily)

0.84M points



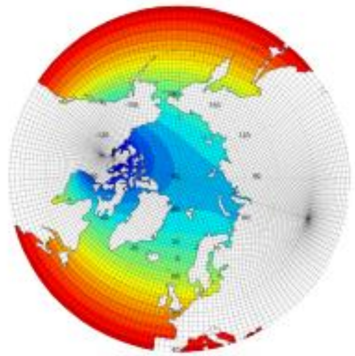
ORCA 1/4
47 Gigabytes of memory
3500 CPU hours
120 Gigabytes of output (daily)

67.72M points



ORCA 1/12
414 Gigabytes of memory
90 000 CPU hours
1 Terabyte of output (daily)

991M points



ORCA 1/36
> 1 Terabytes of memory
~4 000 000 CPU hours
> 5 Terabytes of output (daily)

8,917M points

EC-Earth 3 coupled ~10 km

ESiWACE: EC-Earth ~10km coupled demonstrator

- **IFS** cycle 36r4 for atmosphere
    - T1279L91: ~16 km grid point distance, **2.1 M** grid points
- **NEMO-LIM3** v3.6 for ocean & sea-ice
    - ORCA12L75: ~9 km grid point distance, **13.2 M** grid points*
- Total 3D space points: **1,181kM vertices**

* Including land points (up to 1/3 of the total)

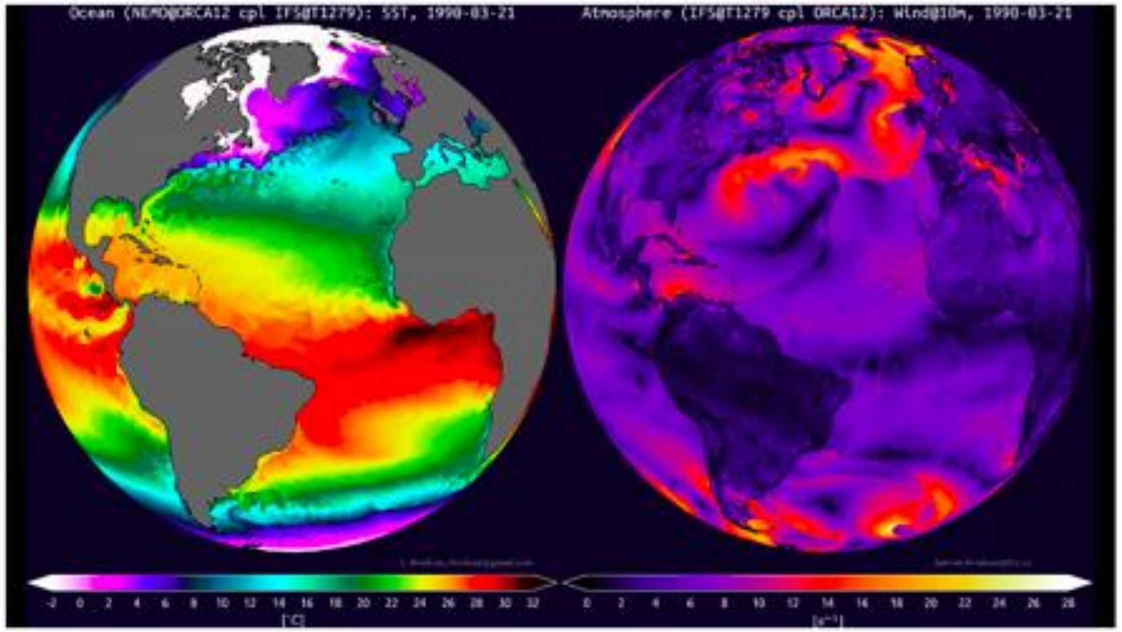
EC-Earth 3 - T1279-ORCA12 in MareNostrum 3

First global, coupled ~10km simulations

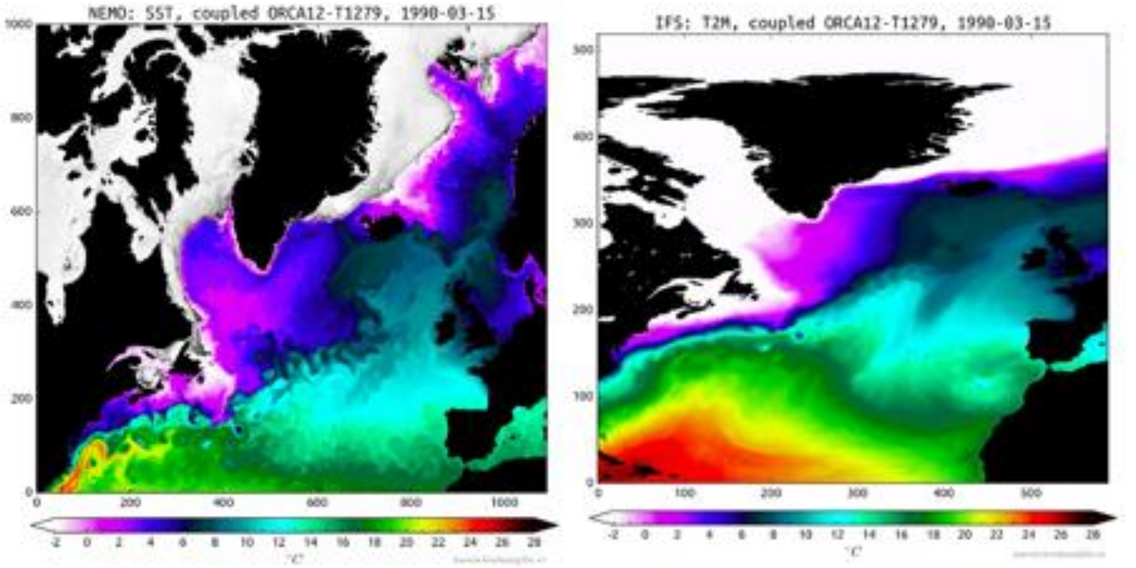
- **EC-Earth 3.2** (IFS36r4 + NEMO 3.6 + OASIS3-MCT)
- 2,035 MPI tasks - 60 SDPD
    - 1,170 NEMO
    - 848 IFS
    - 16 XIOS (I/O server)
    - 1 runoff mapper
- **MareNostrum3** @ BSC

First EC-Earth T1279-ORCA12 results

First global, coupled ~10km simulations



Left, Global Sea Surface Temperature of the ocean component NEMO. Right, Global Speed Wind at 10m of atmosphere component IFS.



Left, regional crop Sea Surface Temperature of the ocean component NEMO. Right, regional crop Temperature at 2m of the atmosphere component IFS.

MareNostrum evolution

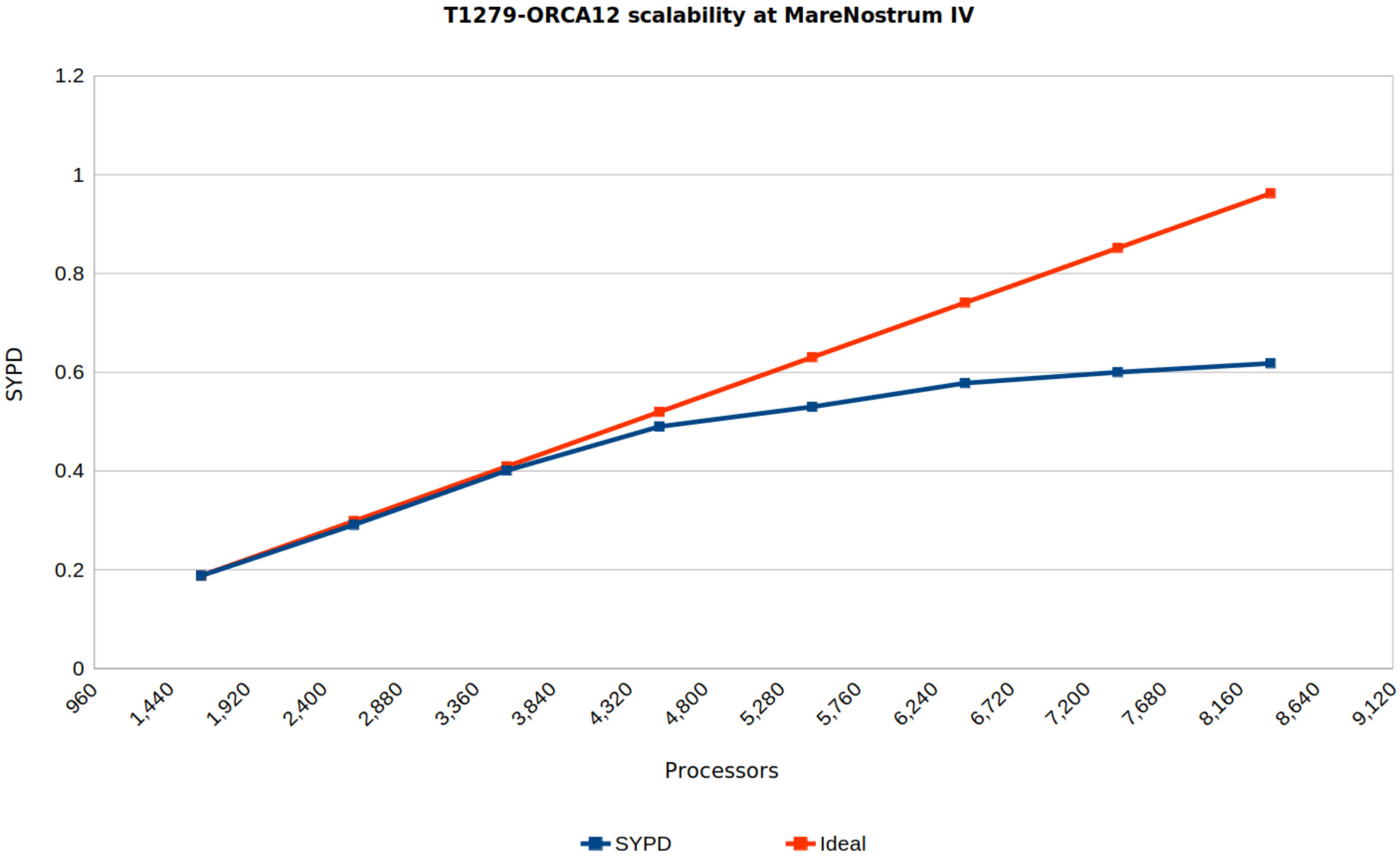
	MareNostrum III (2012)	MareNostrum IV (2017)
Processor	Intel Xeon E5-2670 2.6 GHz	Intel Xeon Platinum 8160 2.1 GHz
#Cores per socket / #Sockets	8 / 2	**24** / 2
Memory	32Gb DDR3-1600	96Gb DDR4-2667
Interconnection	Infiniband FDR10 10Gb	Intel Omni-Path **100Gb**

MareNostrum III – 1.1 petaFLOPS

MareNostrum IV – 11.15 petaFLOPS

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EC-Earth 3 - T1279-ORCA12 in MareNostrum 4



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EC-Earth 3 - T1279-ORCA12 in MareNostrum4

Operational global, coupled ~10 km simulations:

- **EC-Earth 3.2** (IFS36r4 + NEMO 3.6 + OASIS3-MCT)
- **5,040 MPI tasks** - 0.44 SYPD, 160 SDPD
    - 3,209 NEMO
    - 1,584 IFS
    - 69 XIOS (I/O)
    - 1 runoff mapper

100 year exp
~27M computing hours!!!

- **MareNostrum4** @ BSC

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EC-Earth 3 - T1279-ORCA12: production runs

- **PRIMAVERA** is a **Horizon 2020** project which aims to develop a **new generation of advanced and well-evaluated high-resolution global climate models**, capable of simulating and predicting regional climate with **unprecedented fidelity**, for the **benefit** of governments, business and society in general.

- The **High Resolution Model Intercomparison Project (HighResMIP)** is a **CMIP6** endorsed MIP that applies, for the **first time**, a **multi-model approach** to the systematic investigation of the **impact of horizontal resolution**.

H2020: European HPC & science integration case

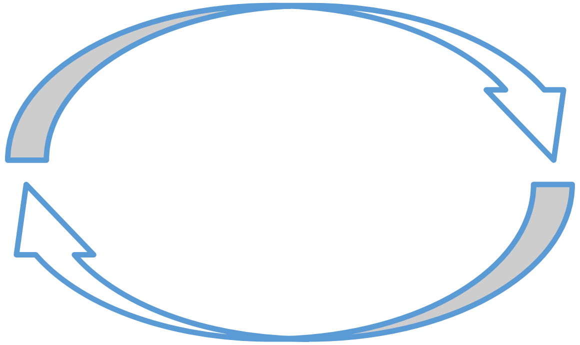
PRACE

Research infrastructure

esiwace

CENTRE OF EXCELLENCE IN SIMULATION OF WEATHER AND CLIMATE IN EUROPE

HPC applications (CoEs)



PRIMAVERA

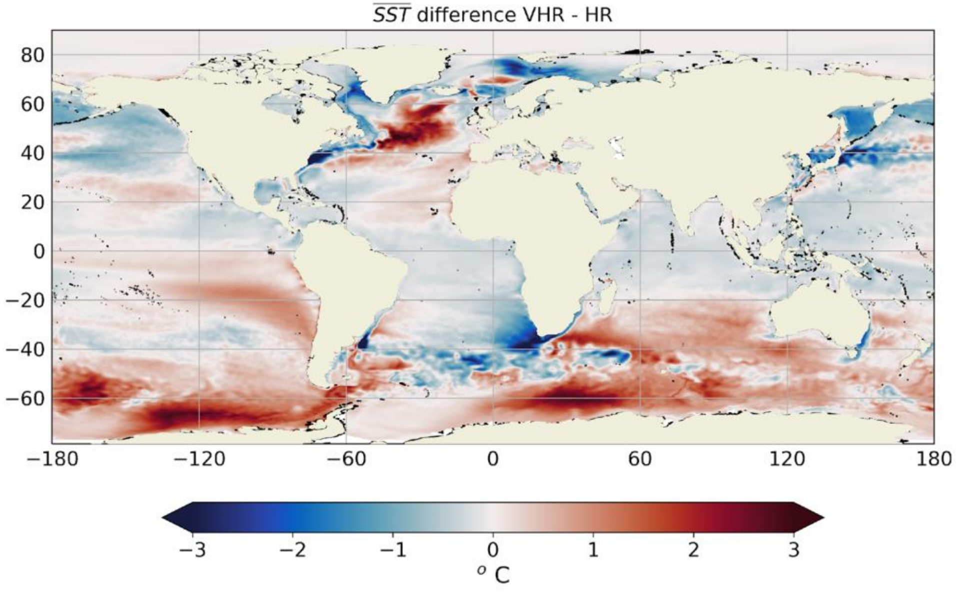
Climate science and HPC

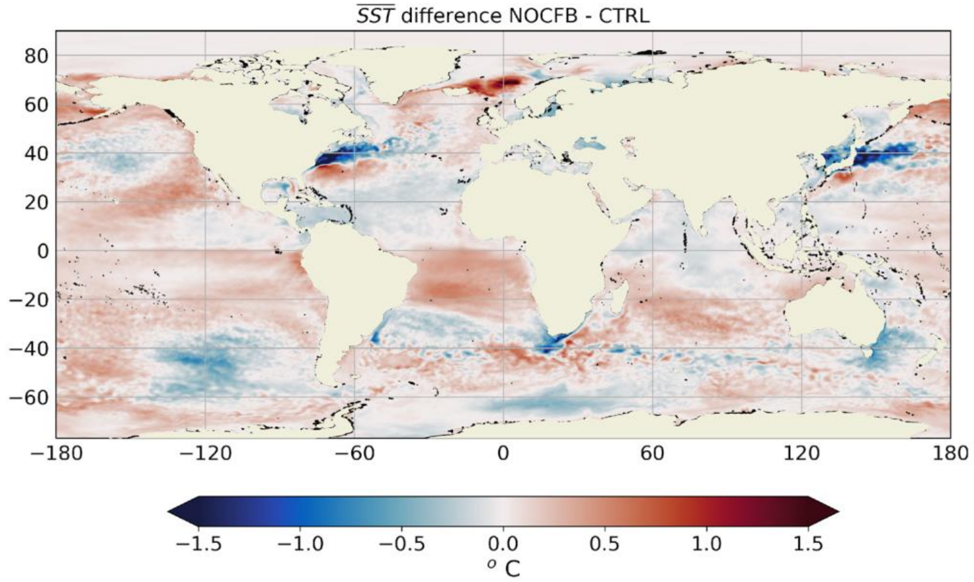
Community

BSC
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Scientific objectives

- Develop and prepare a **new generation** of **global high-resolution** climate models
- Evaluating global high-resolution climate models at a **process level**
- Focus on **air-sea interactions** at oceanic mesoscale:
  - Thermal feedback
  - Evaluate the role of the mechanical interactions between oceanic surface currents and atmospheric winds ("**current-feedback**")

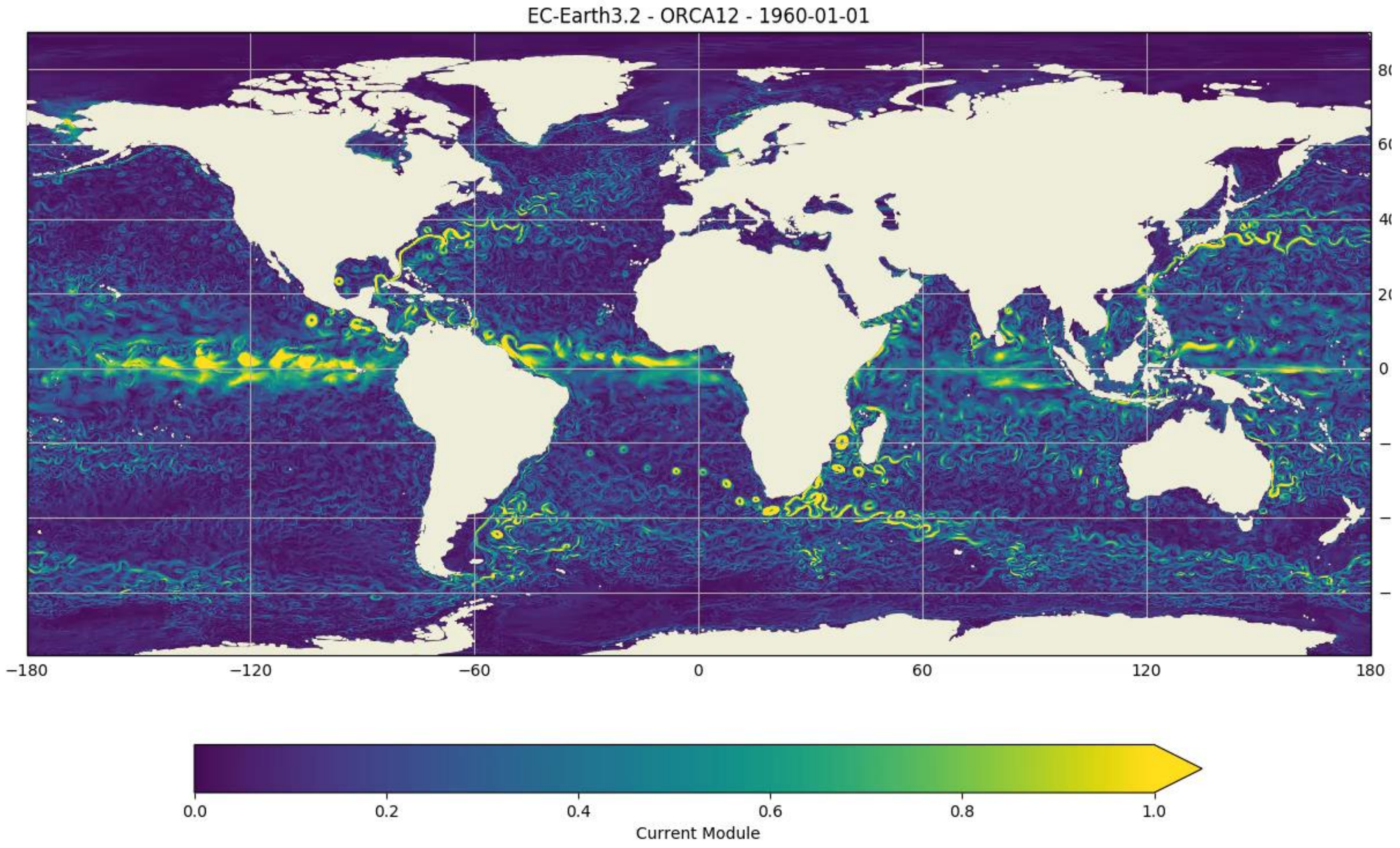




BSC Supercomputing Center
Centro Nacional de Supercomputación

Courtesy of: Thomas Arsouze

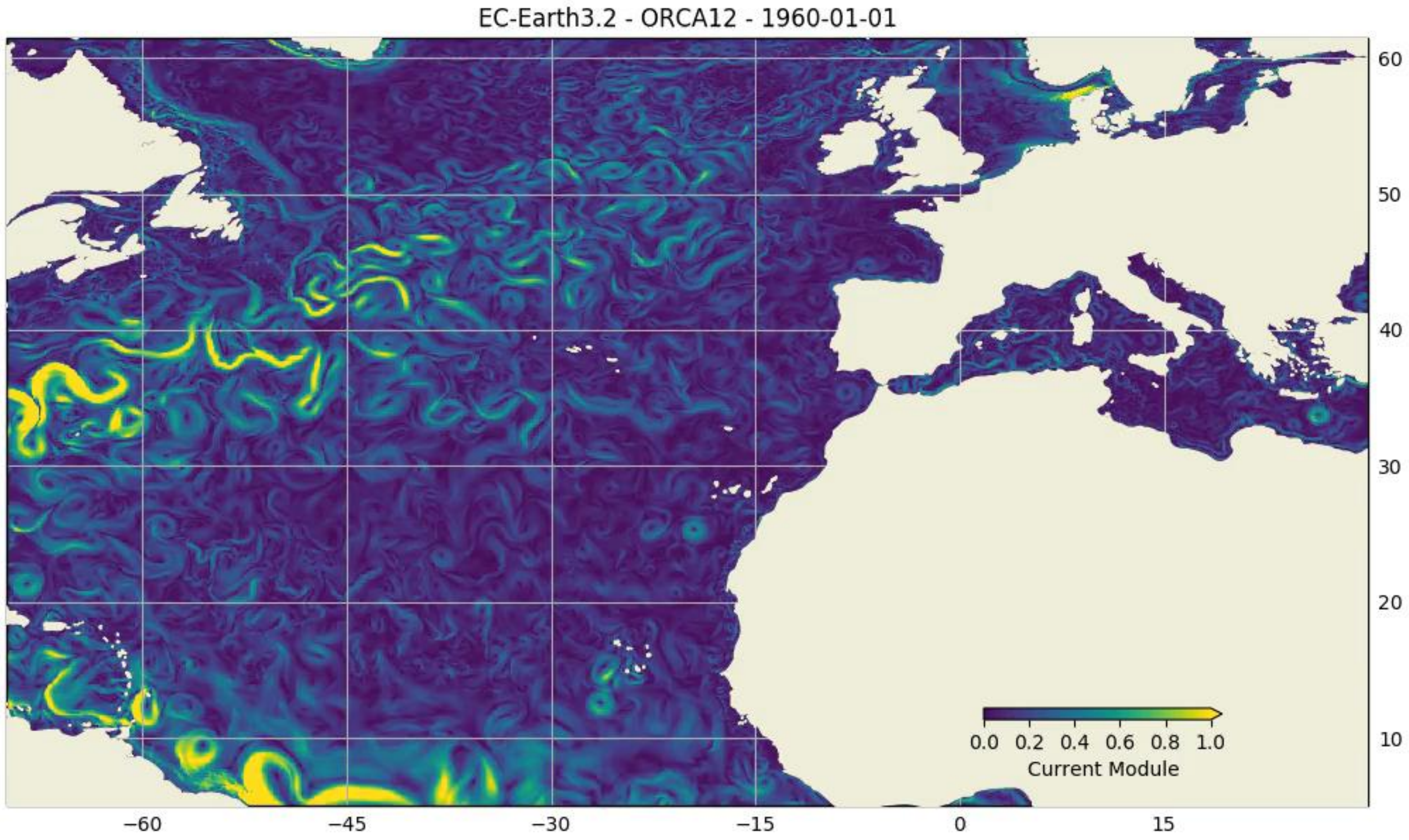
EC-Earth 3 - T1279-ORCA12: production runs



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Courtesy of: Thomas Arsouze

EC-Earth 3 - T1279-ORCA12: production runs



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

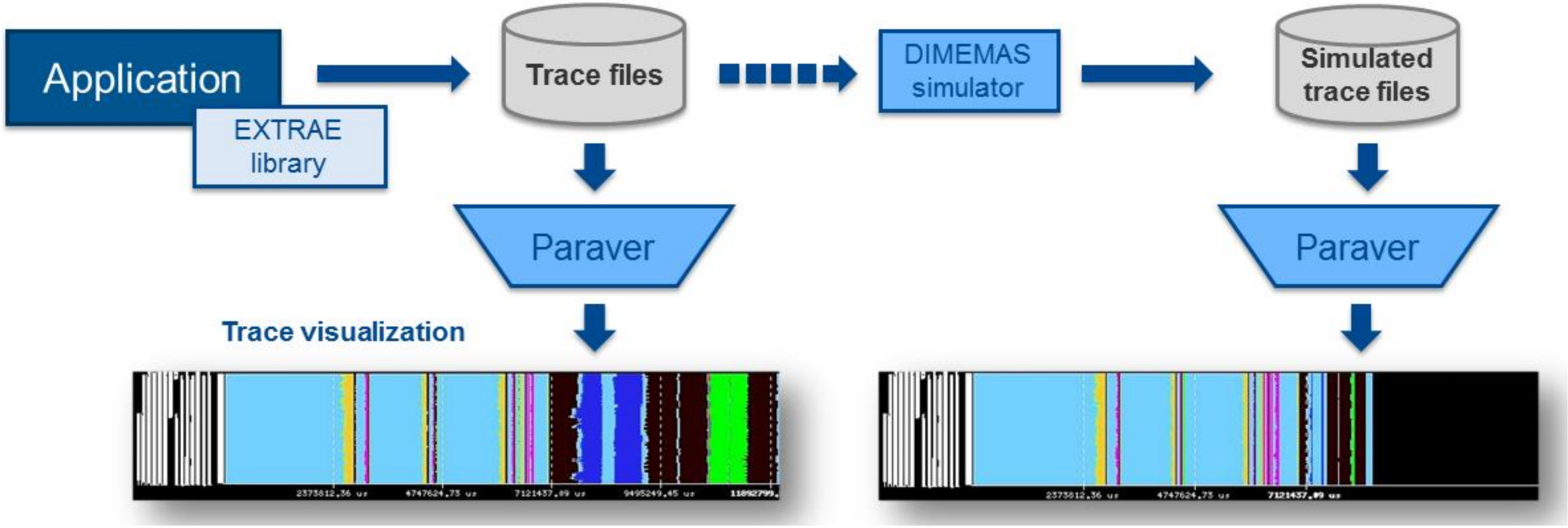
Courtesy of: Thomas Arsouze

EC-Earth 3 - T1279-ORCA12: Main bottlenecks

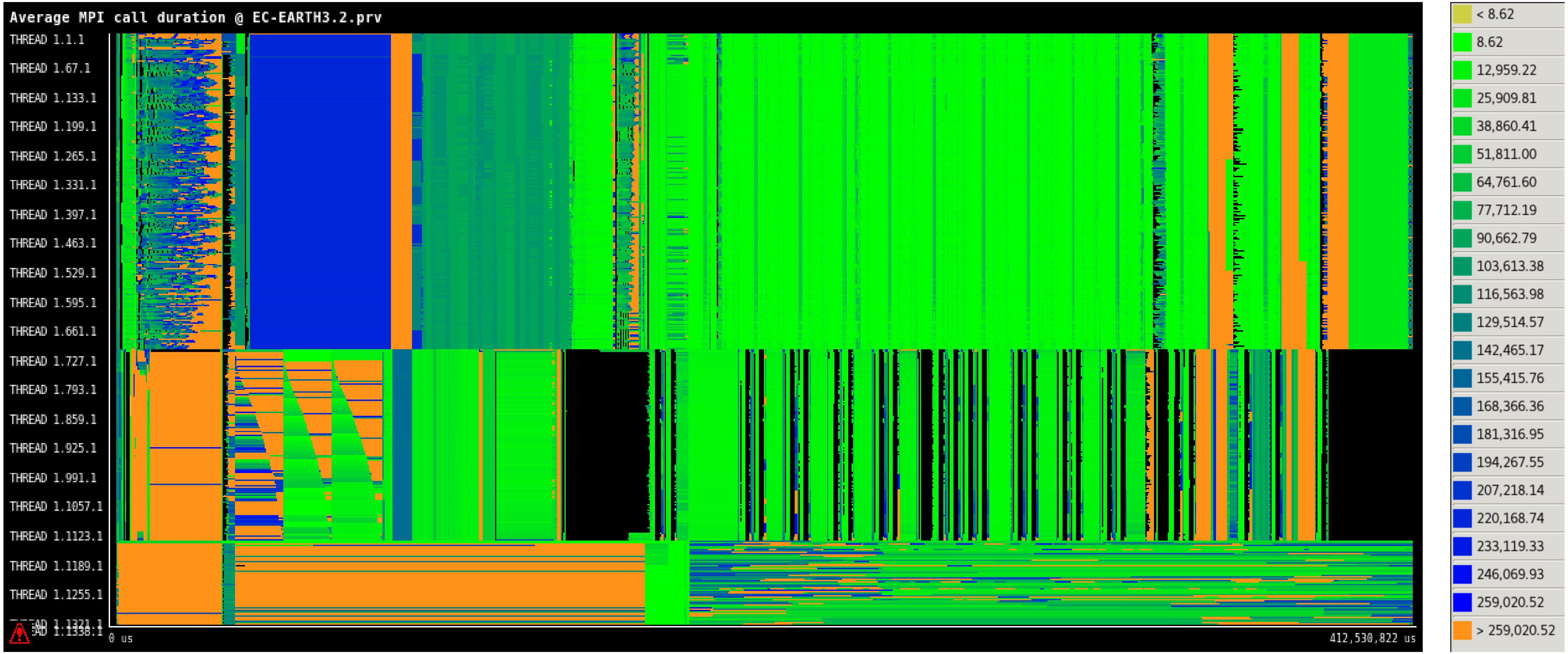
- **I/O overhead** → Interface IFS with **XIOS**
- **Sea-ice scalability** → Reduce **global** communications, couple through **OASIS**
- **Legacy atmospheric model (2010) →** Update **IFS** to newest cycle, using octahedral grid

Performance analysis

- From 1991
- Based on **traces**
- Open Source: https://tools.bsc.es
- **Extrae**: Package that generates Paraver trace-files for a post-mortem analysis
- **Paraver**: Trace visualization and analysis browser
- **Dimemas**: Message passing simulator

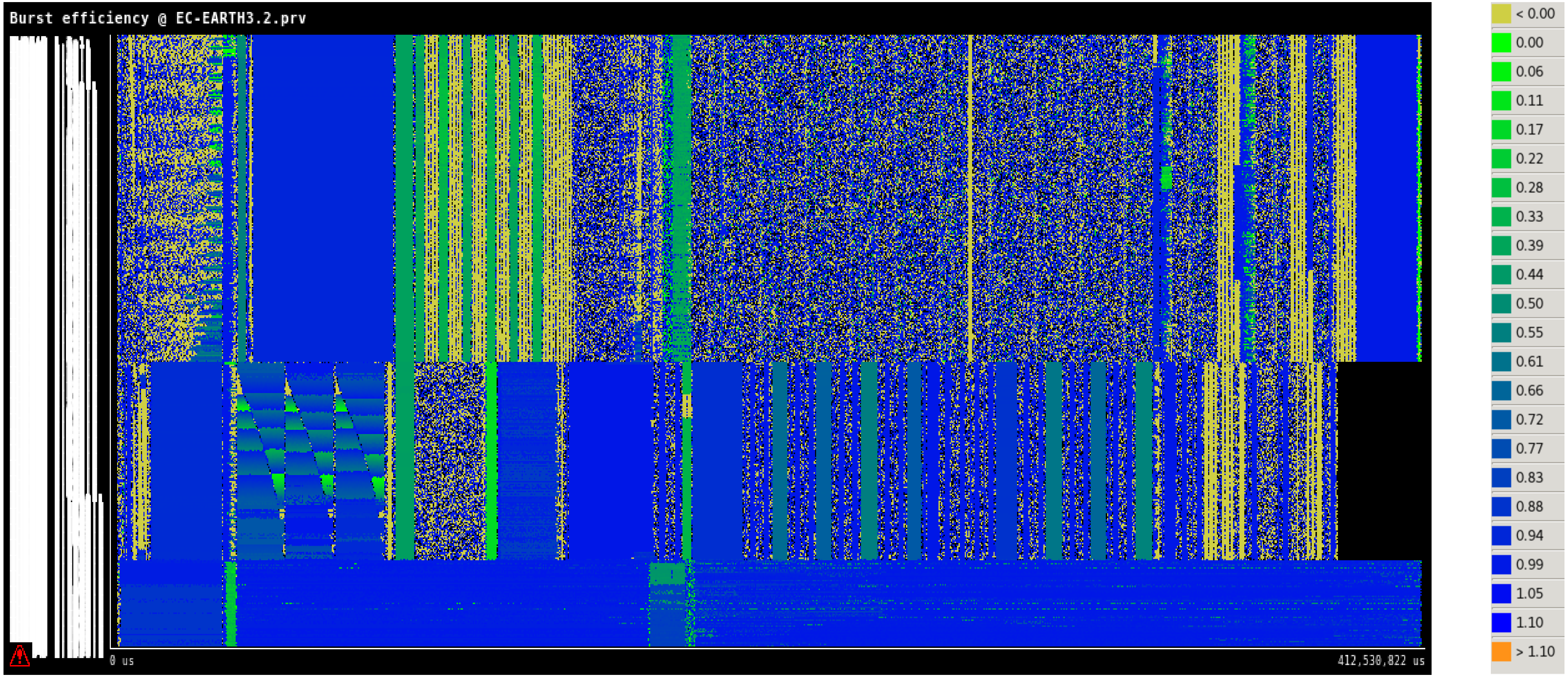


EC-Earth - T1279-ORCA12: Performance analysis



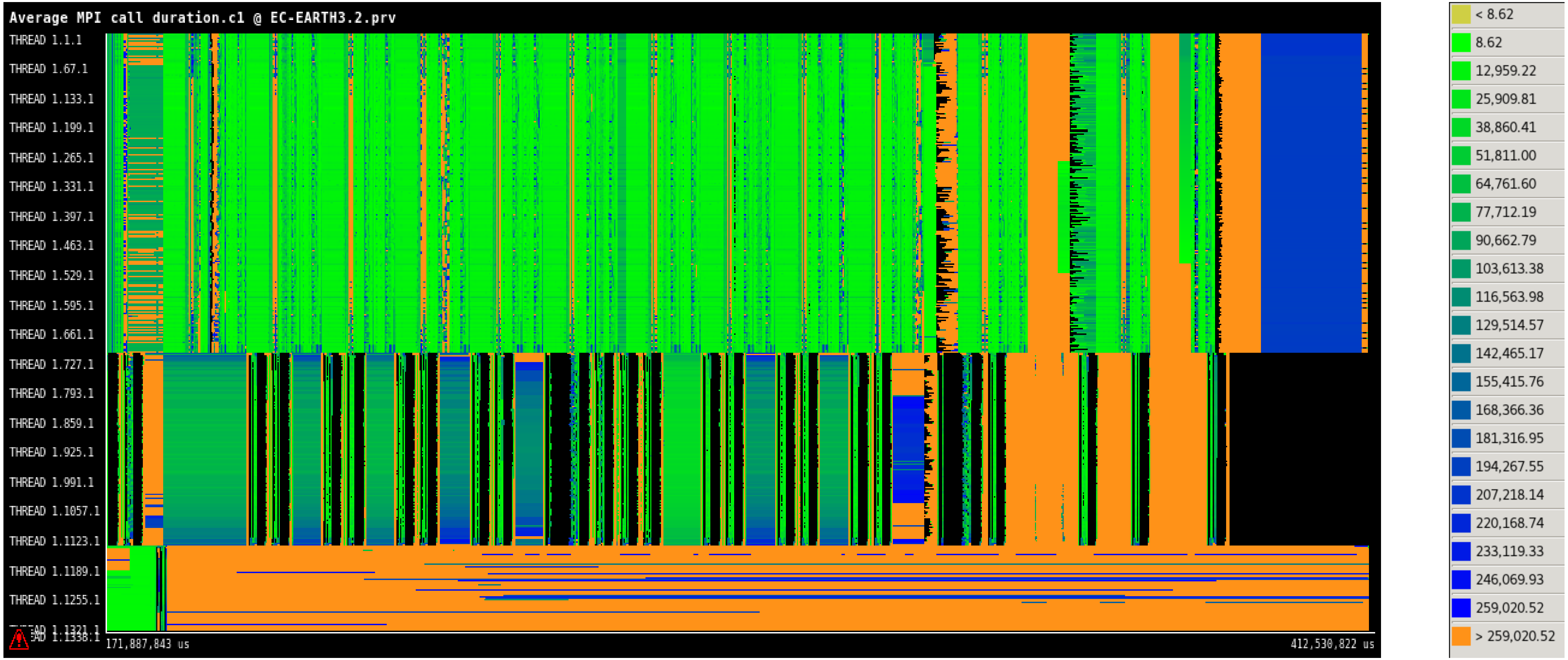
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EC-Earth - T1279-ORCA12: Performance analysis



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EC-Earth - T1279-ORCA12: Performance analysis



Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación

ESiWACE 2: EC-Earth coupled ~10 km production-mode

- Develop **infrastructure** for production-mode configurations
  - **Coupling** infrastructure **(OASIS)**
  - Improvement of **I/O (XIOS)**
  - **NEMO** for high-resolution
  - Infrastructure for **high-resolution data**
- Develop production-mode **configurations**
- Port models to **pre-exascale** EuroHPC **systems**

ESiWACE 2: EC-Earth coupled ~10 km production-mode

- Develop **infrastructure** for production-mode configurations
- Develop production-mode **configurations**
- Port models to **pre-exascale** EuroHPC **systems**

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

esiwace
CENTRE OF EXCELLENCE IN SIMULATION OF WEATHER AND CLIMATE IN EUROPE

EC-Earth coupled ~10 km production-mode

ESiWACE2: EC-Earth 4 VHR coupled demonstrator

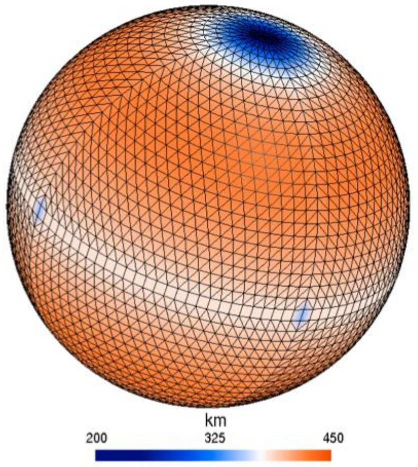
- **OpenIFS** cycle 43r3 for atmosphere
  - Tco639L91: ~16 km grid point distance, **1.66 M** grid points
- **NEMO-SI3** v4 for ocean & sea-ice
  - ORCA12L75: ~9 km grid point distance, **13.2 M** grid points*
- Total 3D space points: **1,141kM vertices**

* Including land points (up to 1/3 of the total)

EC-Earth coupled ~10 km production-mode

Main assets

- Brand **new ESM: EC-Earth 4**
  - **OpenIFS cycle 43r3** (2020)
  - **NEMO v4.0.2** (2019) (incl. **SI3** sea-ice model)
  - **OASIS3-MCT 4**
- **Common** asynchronous **I/O server** (XIOS v2.5)
- **Octahedral reduced** gaussian grid for OpenIFS
- Possibility to **switch** numeric **precision**



N24 octahedral Gaussian grid

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

esiwace
CENTRE OF EXCELLENCE IN SIMULATION OF WEATHER AND CLIMATE IN EUROPE

EC-Earth Tco639-ORCA12 production-mode

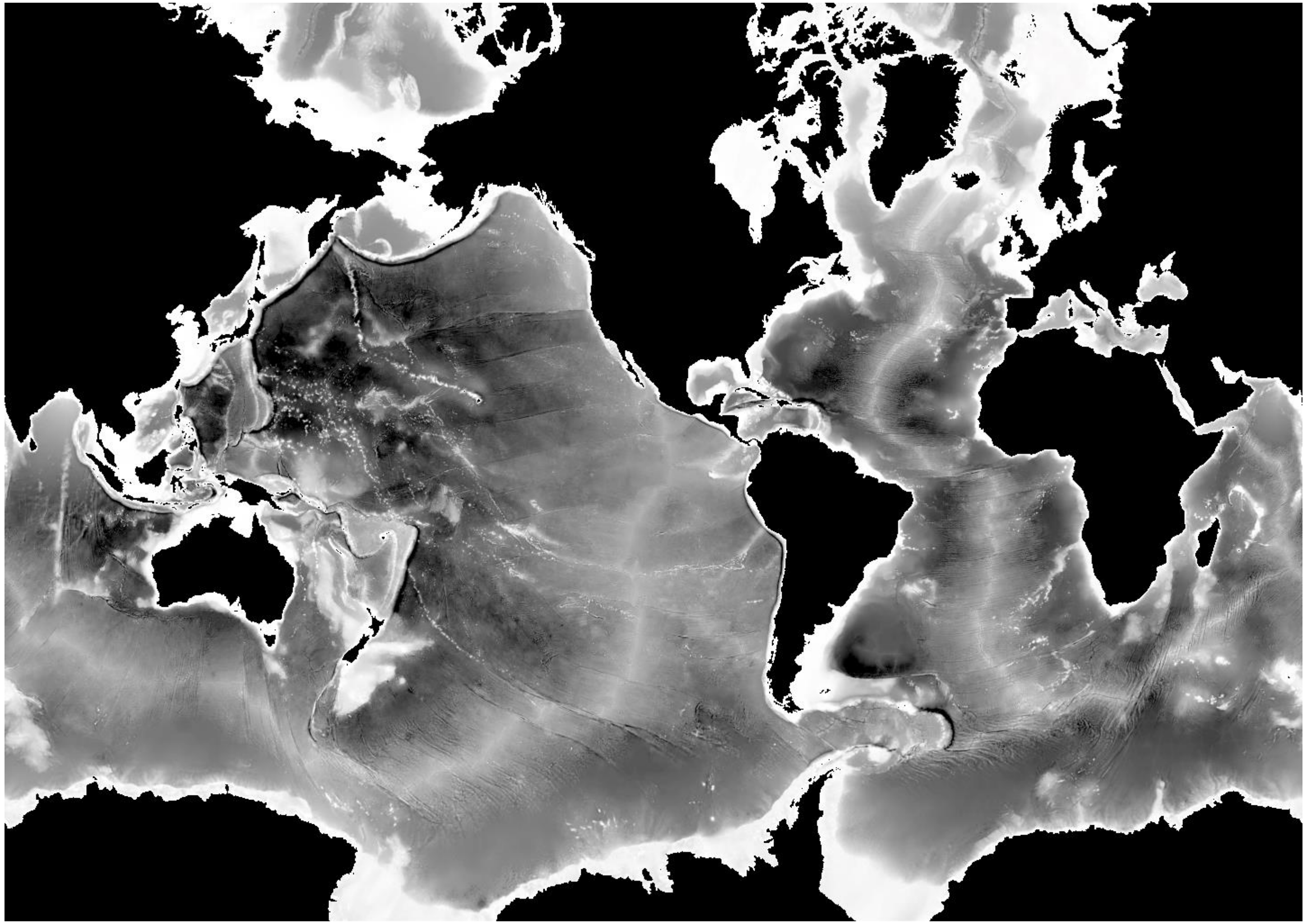
New VHR configuration development

- **Ocean**
    - **ORCA12 adapted** from T1279-ORCA12 configuration in EC-Earth 3.
- **Atmosphere**
    - **Initial conditions** for the **Tco639 grid.**
- **Coupler**
    - **OASIS** coupler grids, masks and areas information using the **OCP**[1] tool.
    - **OASIS remapping weights** generated in parallel (OMP)[2].

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

[1] https://github.com/JanStreffing/ocp-tool
[2] OASIS3-MCT4 new feature

EC-Earth Tco639-ORCA12 production-mode

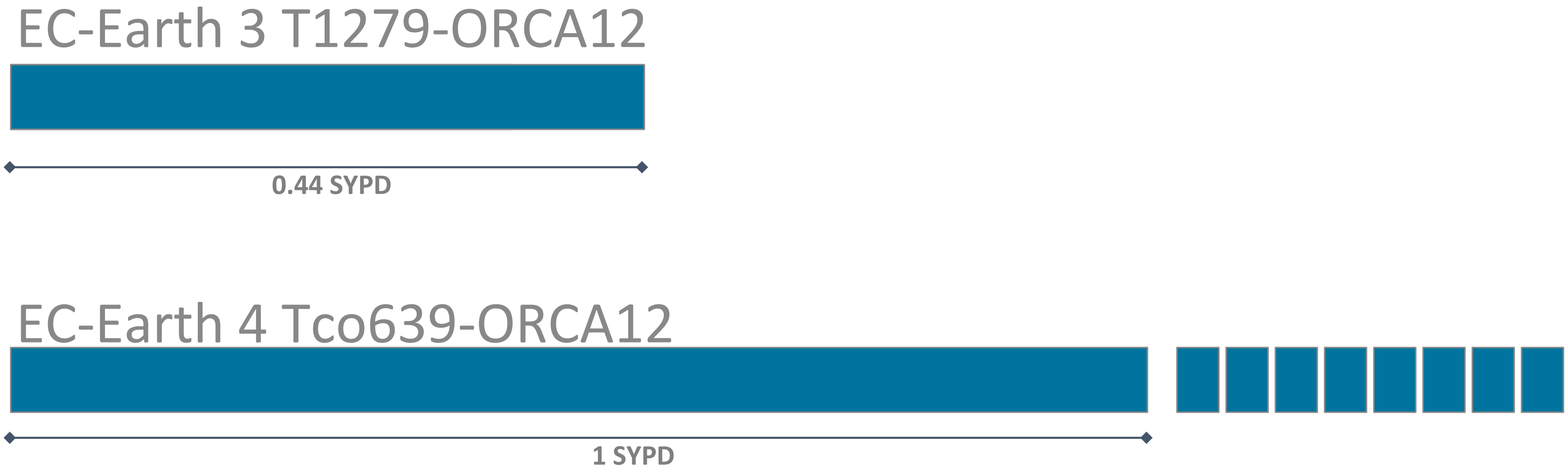


ORCA12 bathymetry

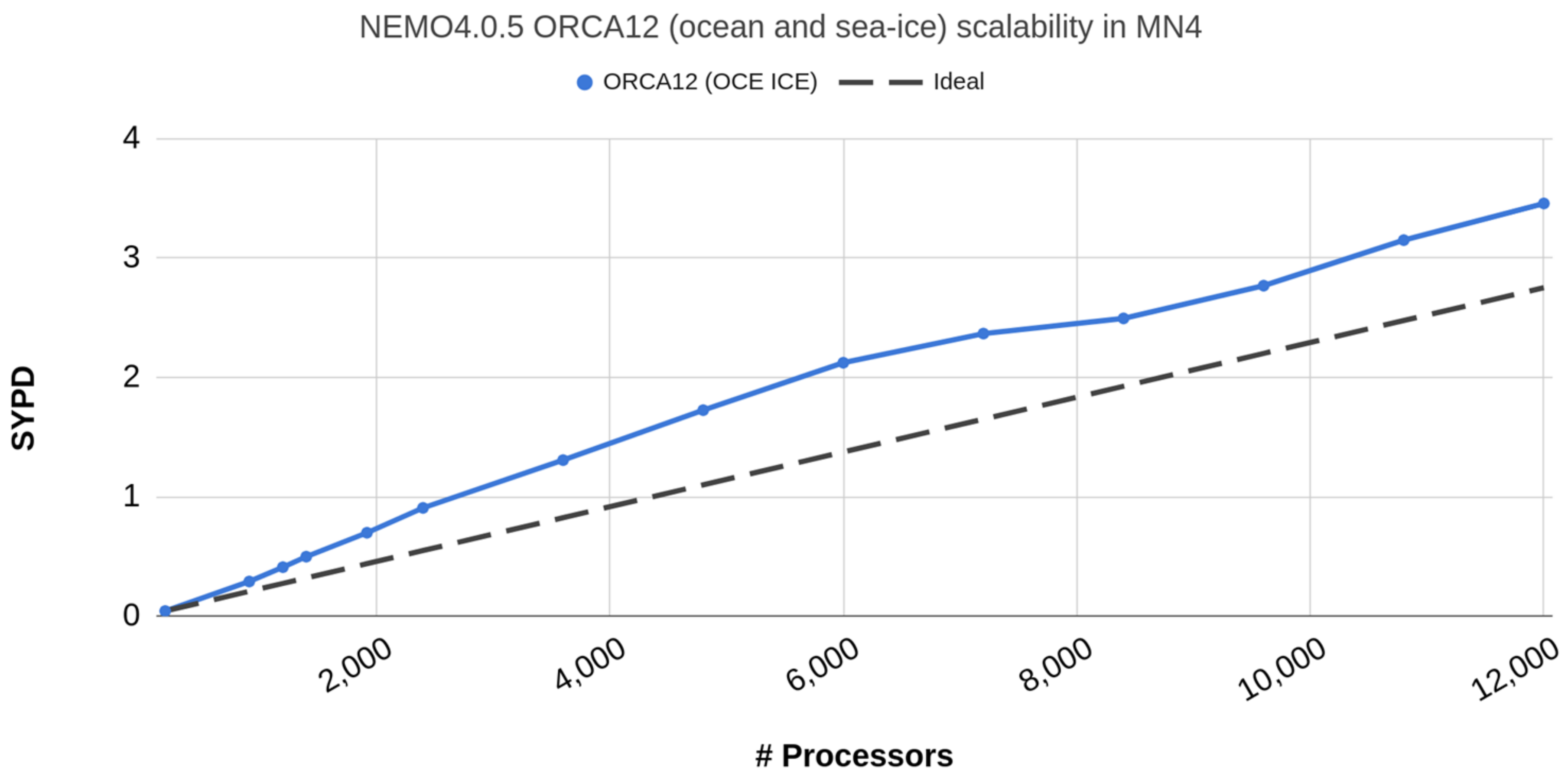
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EC-Earth Tco639-ORCA12 production-mode

Objective: >1 SYPD



NEMO 4 - ORCA12 in MareNostrum 4



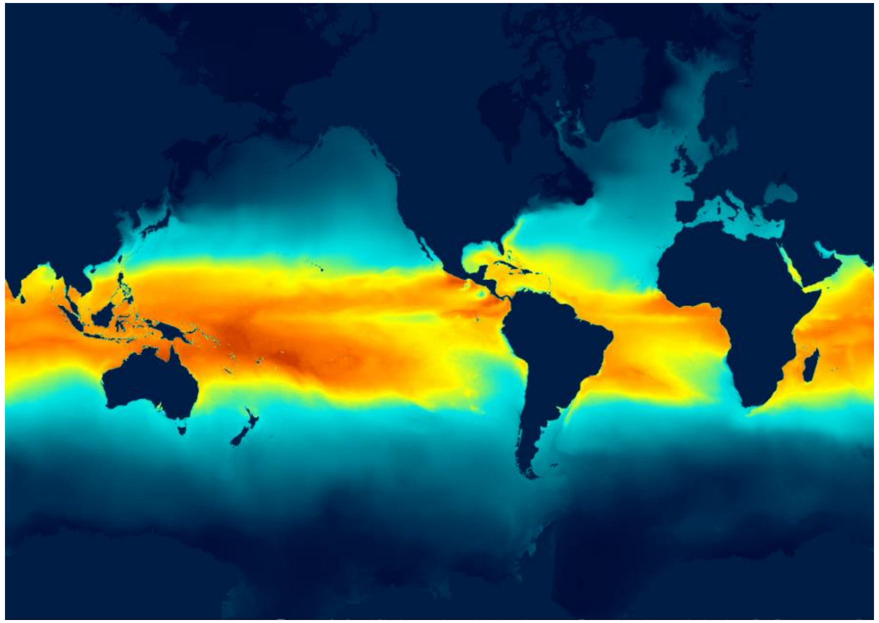
Objective for the coupled version

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EC-Earth coupled ~10 km in production mode

Progress status

- **Deployment** of EC-Earth 4 at MareNostrum 4.
- NEMO (ORCA12) and OpenIFS (Tco639) **configuration data** and initial **parametrizations**.
- Generation and testing of **remapping weights**.
- **Test** runs. Test and tune **output** generation.
- Fine **tuning** of the model’s parameters.
- **Spinup** and generation of initial conditions.
- Load **balance** and **scalability** analysis.
- **Performance** study.
- **Mixed-precision?**



Sea-surface temperature after 1 month

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

An operational EC-Earth workflow in BSC-ES

Model setup

- Configure experiment
- Set up platform-specifics
- Prepare run scripts

Deployment on remote HPC

- Synchronize model files
- Deploy configuration files
- Compile the model

Initial data pre-processing

- Select and deploy initial data
- Pre-process initial conditions

Simulation

- Run simulation
- Assimilation / adaptive ensembles

Output post-processing

- Post-process output data
- Archive restarts / output files
- Transfer results

Verification & diagnostics

- Calculate diagnostics
- Generate visualization plots

Conclusions

- **First coupled ~10km** configuration developed within ESiWACE:
  - Developed and shared among **EC-Earth consortium** partners
  - **Deployed and tested** in the **BSC** HPC systems
  - Used in **production** for **different** projects
  - Used to investigate **very-high resolution scalability** for coupled systems
- **~10 km production-mode** configuration developed within ESiWACE2:
  - Solves the most important **bottlenecks**. Uses **updated** model components
  - Will be deployed and tested in the **pre-Exascale** EuroHPC systems
  - Will allow running **efficient** VHR simulations with a **production throughput**

Preparing NEMO and EC-Earth models for very high-resolution production experiments

Miguel Castrillo (BSC), Dorotea Iovino (CMCC), Clement Bricaud (Mercator Ocean)

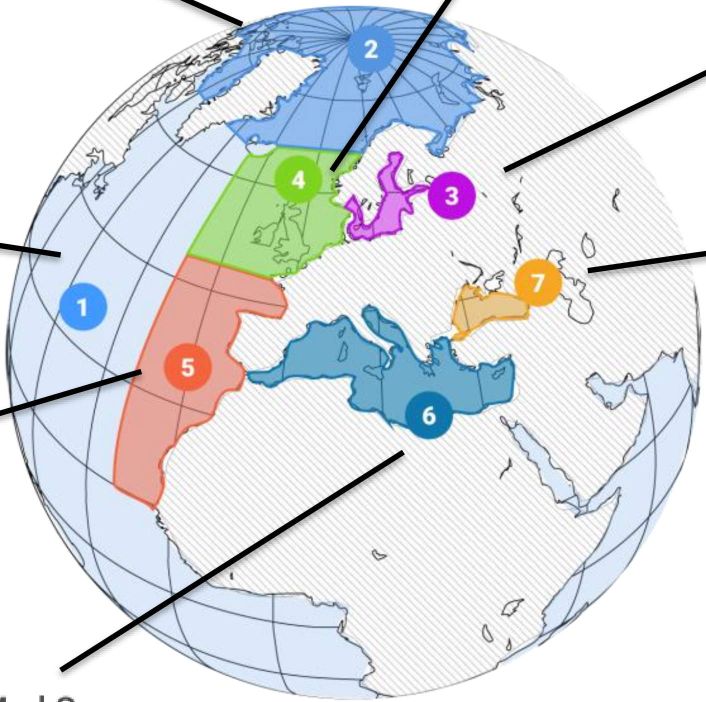
EU Copernicus Marine Service (CMEMS): Improve Ocean forecasting systems

ocean dynamic: HYCOM
sea-ice : EVP model
BGC: ECOSMO

ocean dynamic:
BGC: ERSEM

ocean dynamic: HBM
sea-ice : HBM
BGC: ERGOM

NEMO OGCM for:
ocean dynamic
sea-ice
BGC



ocean dynamic:
BGC: BAMHBI

NEMO OGCM for:
ocean dynamic
BGC

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

ocean dynamic :
BGC: BFM

Copernicus Marine Service

Implemented by MERCATOR OCEAN INTERNATIONAL

Ocean forecasting: improve Ocean numerical representation

- Improve data assimilation systems with new observing platforms

⇒ increase in the **amount** of collected observations

⇒ enhancement of their **accuracy**

⇒ the time and space **resolution** of observations improved

⇒ Description of **finer scales** improved
- SST/microwaves sensors: resolution of about 25 km, whereas
- SST/infrared sensors: resolution down to the kilometric scale.
- SSH/SWOT scales until 15 to 30 km whereas actual altimeters are limited to 150 km.

Ocean forecasting: improve Ocean numerical representation

- Improve numerical model for a better

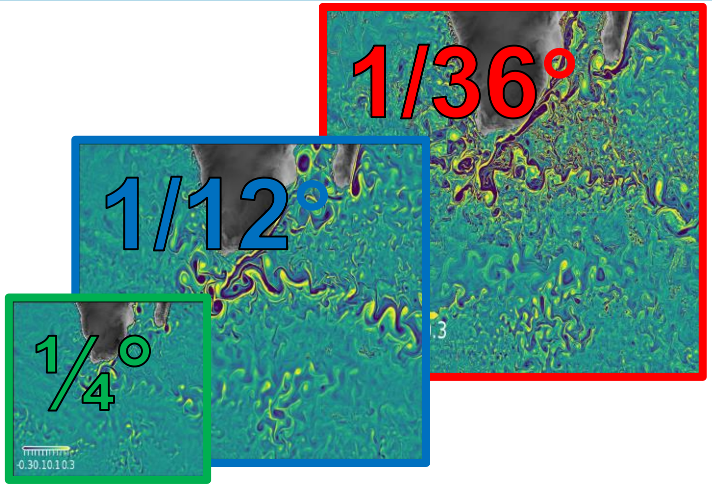
- **representation** of the **circulation** in the open ocean
- **representation** of **energy transfers** between finer and larger scales
- understanding of **scales contributions** (geostrophic flows, tidal motions, waves, inertial currents)

- => **Resolve** scales below 100 kilometers, in particular sub mesoscale processes (1-50 km)

Improve Ocean numerical model

- Improve ocean regime: increase grid **resolution**

regime	Resolution (°)	Resolution (km)	Operated (since)
Eddy-permitting	¼°	20-25 km	2005
Eddy-resolving	1/12°	6-9 km	2009
Submesoscale-permitting	1/36°	2-3 km	2024?



- Improve model physic and **parametrizations**
  - Non-linear free surface (variable volume), multi-category sea ice model,…

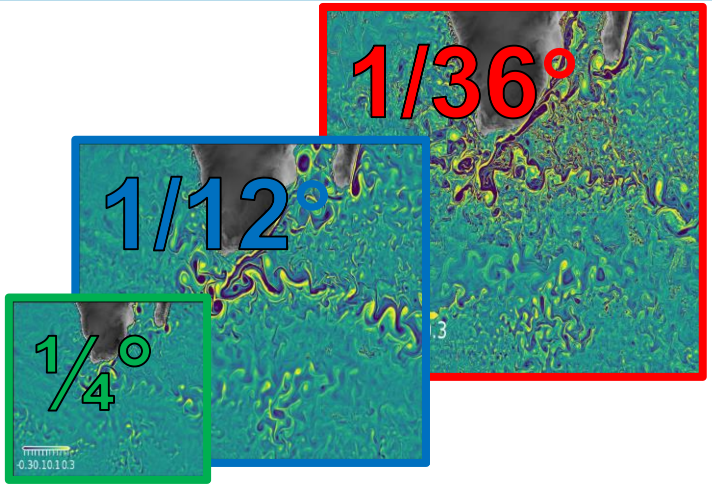
- Improve **numerical representation**
  - higher order numerical schemes for advection
  - Tracers: 4th order FCT scheme Momentum: 3rd order UBS scheme

- Atmospheric **forcing**
  - increase space and time resolution: 16km to 9 km and 3 hours to 1 hour

Improve Ocean numerical model

- Improve ocean regime : increase grid resolution

regime	Resolution (°)	Resolution (km)	Operated (since)
Eddy-permitting	¼°	20-25 km	2005
Eddy-resolving	1/12°	6-9 km	2009
Submesoscale-permitting	1/36°	2-3 km	2024?



- ⇒ More operations
- ⇒ More memory access
- ⇒ More memory capacity
- ⇒ More I/0

- Atmospheric forcing: increase space and time resolution
- 16km to 9 km and 3 hours to 1 hour

Improve numerical models HPC performances

What running a high-resolution model imply ?

⇒ More operations
⇒ More memory access
⇒ More memory needed
⇒ More I/0

What we need to do ?

⇒ Adapt model to new architectures
⇒ Improve operations speed
⇒ Improve exploitation of memory hierarchies
⇒ Improve communications
⇒ Improve I/Os

Improve numerical models HPC performances

ESIWACE2 project (EU H2020)

- improve **efficiency** and **productivity** of numerical weather and climate **simulation** and prepare them for future exascale systems
- **prepares** the European weather and climate community to make use of future **exascale systems** in a **co-design** effort involving modelling groups, computer scientists and HPC industry

IMMERSE project (EU H2020)

- *Develop a new, efficient, stable and scalable* ***NEMO reference code*** *with improved performances adapted to* ***exploit future HPC technologies*** *in the context of* ***CMEMS systems***
- *Develop NEMO for the challenges of delivering* ***ocean state estimates*** *and* ***forecasts*** *describing ocean dynamics and biogeochemistry at* ***kilometric scale*** *with improved accuracy*

ORCA36= high resolution configuration used as a bench

For developements « dedicated » to HPC

Improve numerical models HPC performances

Project: SWOP: the Submesoscale-permitting World Ocean Project

- 24M CPU hours obtained with the 22th **PRACE** call on the BSC MareNostrum IV

2 objectives:

- **Test HPC development** on NEMO4 with **ORCA36** configurations
- Perform a **multi year hindcast** forced with the ECMWF high resolution / high frequency IFS dataset

- Compare 2 hindcasts without/with **tidal forcing**
- Transfer and **dissemination** on the EU WEKEO DIAS

The ”ORCA36” 1/36° global configuration

- Based on NEMO 4

- Horizontal grid : tripolar ORCA grid, (2-3km), 12960 * 10850 points
- Vertical grid: 75 Z-levels, 1 meter at surface

- Ocean dynamic and SI3 sea-ice models

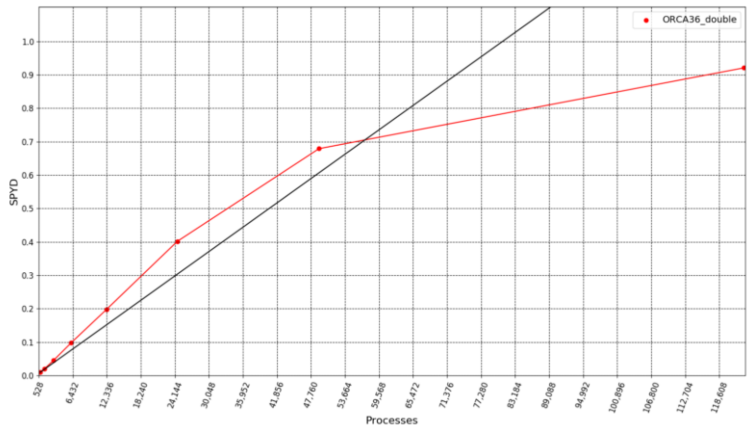
- Bathymetry based on GEBCO 2019

- Forcing datased based on ECMWF IFS (HRES/1 hour) system

ORCA36: first tests of scalability

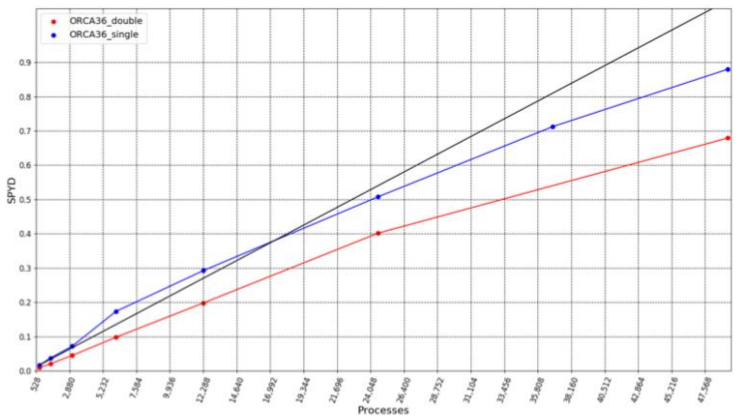
- CMEMS contract with BSC:
- « 87-GLOBAL-CMEMS-NEMO: EVOLUTION AND OPTIMISATION OF THE NEMO CODE USED FOR THE MFC-GLO IN CMEMS » :
- Miguel Castrillo, Mario Acosta, Oriol Tintó Prims, Kim Serradell

ORCA36 scalability
(no forcing, no sea-ice, no outputs)



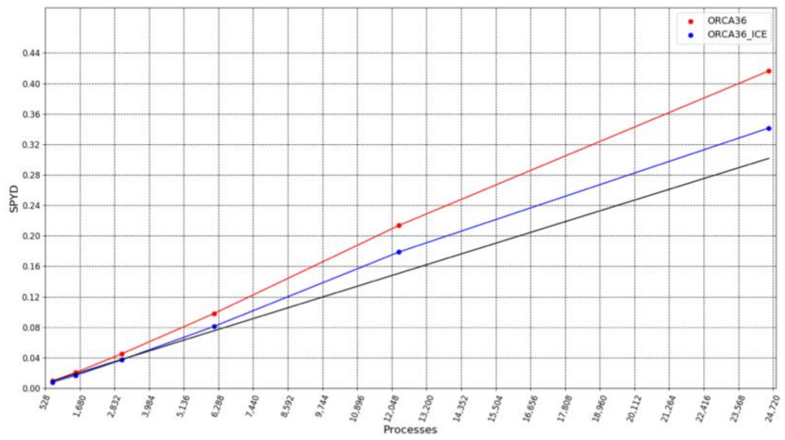
Scalability tested up to 122 000 cores
Good scalability up to **50 000 cores**

ORCA36 scalability
(no forcing, no sea-ice, no outputs)
Simple vs double precision



Single precision improves scalability of the double precision version

ORCA36 scalability
(no outputs)
Sea-ice model impact



ORCA36: first tests of scalability

***performance** metrics*

Number of processes	**768**	**3,072**	**6,144**
Parallel efficiency	93.72	85.74	82.65
Load balance	97.42	92.4	92.74
Communication efficiency	96.2	92.79	89.12
Computation scalability*	100	130.25	110.98
Global efficiency*	**93.72**	**111.67**	**91.72**
IPC scalability*	100	123.47	118.24
Instruction scalability*	100	102.63	94.69
Frequency scalability*	100	102.78	99.12
Speedup*	**1.00**	**4.77**	**2.14**
Average IPC	0.29	0.35	0.42
Average frequency (GHz)	2.09	2.15	2.13

** These values use the column on their left as reference.*

- **parallel efficiency decreases** with the scale
- **gains in computational efficiency** due to the increase in instructions per cycle (IPC).
  better exploitation of the shared resources
  faster memory operations

Proportion of **useful instructions**
(those not involved in MPI communication)

Function	**768**	**3,072**	**6,144**
divhor_m..div_hor_	0.60%	0.61%	0.61%
step_mp_stp_	0.13%	0.16%	0.19%
sbcmod_mp_sbc_	0.05%	0.05%	0.04%
usrdef_s..sbc_oce_	0.49%	0.48%	0.45%
lib_fort.._sum_2d_	0.23%	0.23%	0.23%
eosbn2_mp_rab_3d_	3.52%	3.43%	3.21%
eosbn2_mp_bn2_	0.82%	0.82%	0.84%
zdfphy_m..zdf_phy_	0.20%	0.27%	0.35%
zdfdrg_m..zdf_drg_	0.12%	0.11%	0.10%
zdfsh2_m..zdf_sh2_	0.79%	0.79%	0.78%
zdfgls_m..zdf_gls_	**6.08%**	**6.02%**	**5.86%**
zdfmxl_m..zdf_mxl_	0.27%	0.29%	0.30%
sshwzv_m..ssh_nxt_	0.09%	0.09%	0.11%
domvvl_m.._sf_nxt_	1.86%	1.98%	2.12%
sshwzv_mp_wzv_	0.35%	0.37%	0.45%
eosbn2_m..itu_pot_	1.37%	1.34%	1.26%
zpshde_m..zps_hde_	0.18%	0.17%	0.17%
eosbn2_m..situ_2d_	0.04%	0.04%	0.04%
dynadv_u..adv_ubs_	3.95%	4.07%	4.34%
dynvor_m..vor_een_	1.35%	1.34%	1.42%
dynhpg_m..hpg_sco_	0.75%	0.73%	0.70%
dynspg_t.._spg_ts_	**20.34%**	**21.21%**	**22.63%**
dynzdf_m..dyn_zdf_	2.00%	2.11%	2.24%
trasbc_m..tra_sbc_	0.01%	0.01%	0.01%
traqsr_m..tra_qsr_	**16.59%**	**15.84%**	**14.61%**
traadv_m..tra_adv_	0.26%	0.28%	0.35%
traadv_f..adv_fct_	2.95%	3.05%	3.18%
traadv_f.._nonosc_	4.54%	4.47%	4.40%
traldf_l..ldf_lap_	0.90%	0.97%	1.05%
trazdf_m..tra_zdf_	0.08%	0.09%	0.11%
trazdf_m..zdf_imp_	1.06%	1.14%	1.22%
tranxt_m..tra_nxt_	**18.50%**	**17.71%**	**16.63%**
dynnxt_m..dyn_nxt_	2.66%	2.86%	3.24%
sshwzv_m..ssh_swp_	0.01%	0.01%	0.01%
domvvl_m.._sf_swp_	2.33%	2.42%	2.49%
stpctl_m..stp_ctl_	4.52%	4.41%	4.24%

- most of these instructions are Load and Stores and not floating point operations.
- Impact of code writing or compiler optimization?

ORCA36: first tests of scalability

Time-to-solution for a 5 days run (as for forecasting)

Performed on the new Météo France ATOS/BULL computer

10.000 cores	3H40
20.000 cores	1H55
30.000 cores	1H15
40.000 cores	1H20

Elapsed time for a 5 days run

HPC developments in NEMO 4

- tilling method
- loop fusion

⇒ better use the **memory hierarchy** and improve the **memory access**

- increase halos (=local domain overlapping band)
- Introduction of MPI3 and activation of collective neighbours communications

⇒ **decrease** the cost of **communication** between the different processes

- Mixed precision

⇒ **decrease** the **computational** cost and the **memory** consumption

HPC developments in NEMO 4

- compute diagnostics on GPUs

⇒ improve the **scalability**

- XIOS I/O servers already used with NEMO OGCM only for model outputs
- Activation of model restarts reading and writing with XIOS
- Use 2 levels of servers in XIOS

=> Improve **I/O performances**

Thank you!

The ESiWACE project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 675191

The IMMERSE project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823988

miguel.castrillo@bsc.es dorotea.iovino@cmcc.it cbricaud@mercator-ocean.fr