CERFACS

EUROPEAN CENTRE FOR RESEARCH AND ADVANCED TRAINING IN SCIENTIFIC COMPUTING

Oasis

is-enes

INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

esiwace

CENTRE OF EXCELLENCE IN SIMULATION OF WEATHER AND CLIMATE IN EUROPE

Latest developments of the OASIS3-MCT coupler for improved performance

S. Valcke, L. Coquart, A. Craig, G. Jonville, E. Maisonnave, A. Piacentini

Outline

Historical overview

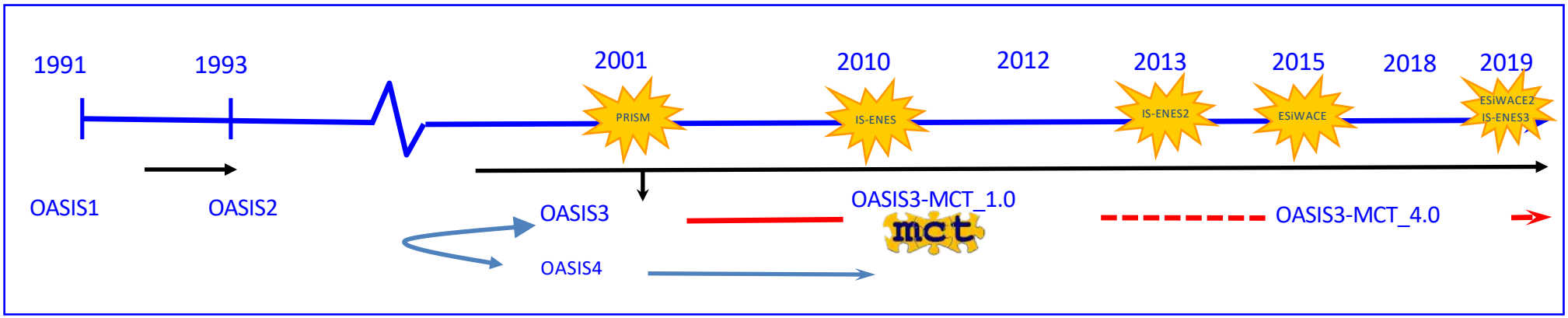
User community

Latest developments

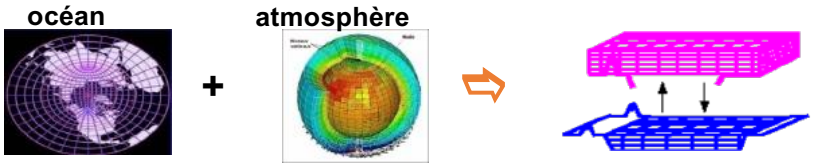
On-going developments

Conclusion

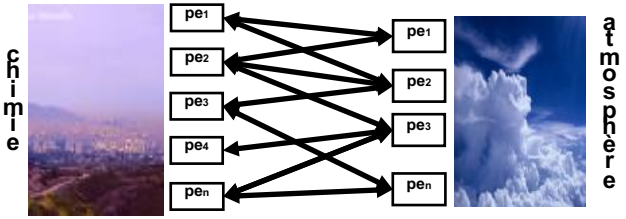
Historical overview



- OASIS1 -> OASIS2 -> OASIS3:
  - 2D ocean-atmosphere coupling
  - low frequency, low resolution :
  - → Flexibility, 2D interpolations



- OASIS4 / OASIS3-MCT:
  - 2D/3D coupling of high-resolution parallel components
  - →Parallelism, performance



- F90 & C, LGPL licence, public domain libraries (MPI, NetCDF, SCRIP, MCT)

User community

2019 survey

67 climate modelling groups around the world use OASIS3-MCT …

…. to assemble more than 80 coupled applications !!

OASIS3-MCT is used in 5 of the 7 European ESMs participating to CMIP6

Latest developments

OASIS3-MCT_4 .0 (June 2018 release)

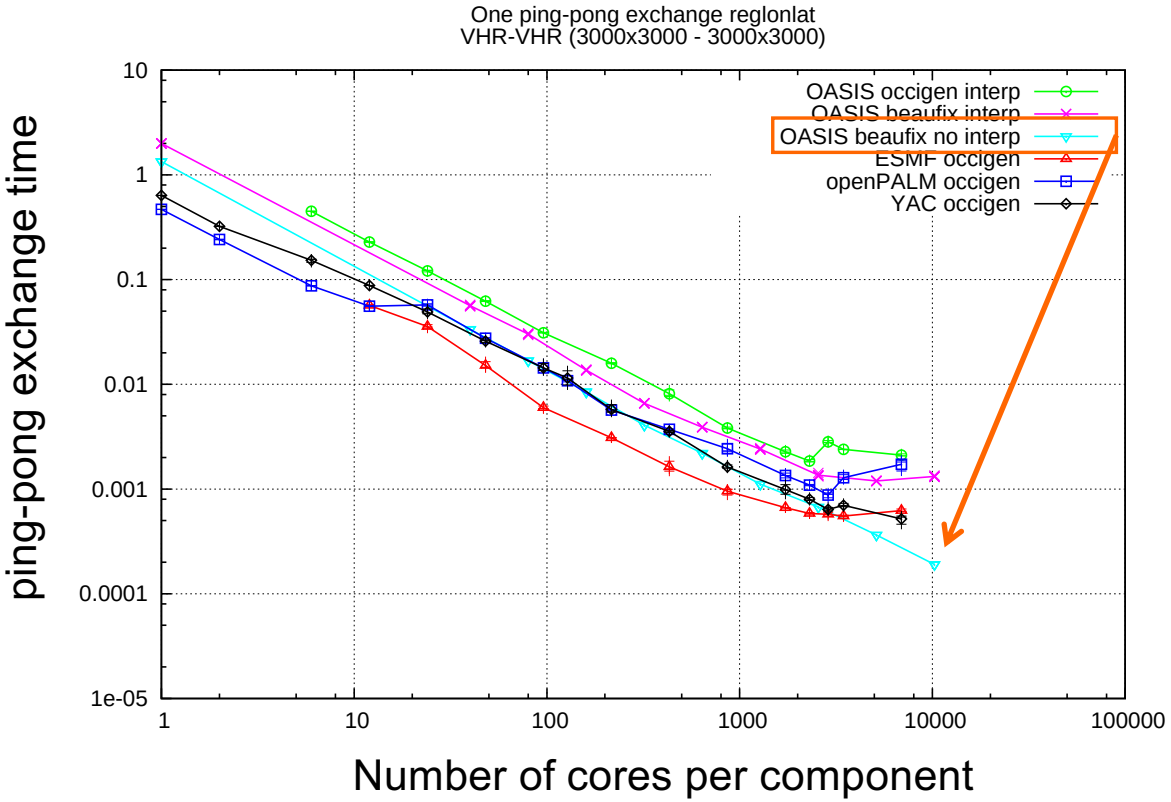
- Bundle fields
- Activation of **« nointerp » for identical grids – impact on IS-ENES2 benchmarks**
- New **more performant algorithms for the global CONSERV operation**
- Upgrade to 2.10beta1 MCT library : reduces by O(10) – O(100) MCT router initialization cost
- **Debugging of the coupling initialisation**
- **Optimisation of the communication using the mapping weights**
- **Hybrid MPI+OpenMP parallelisation of the SCRIP library**

Latest developments

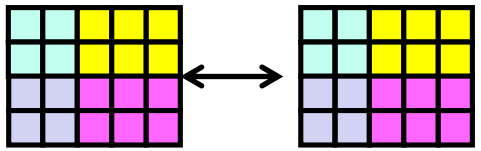
Since June 2018(trunk version):

- Fractional masks for the global conservation operation CONSERV
- New options in global CONSERV to conserve fields with average value close to zero
- « True » area normalisation in conservative remapping
- Possibility to deactivate the “additional nearest-neighbour” option (BILINEARNF, BICUBICNF, DISTWGTNF, and GAUSWGTNF)
- Bugfix for local distance calculation in GAUSWGT interpolation
- More systematic test of NetCDF returned error code

"nointerp" : bypass matrix-vector multiplication for identical grids



IS-ENES2 benchmark VHR:
ping-pong exchange between
3000x3000 regular lat-lon grids
same decomposition



New algorithms for the global CONSERV operation

Different algorithms with different trade-offs between performance and reproducibillty

- *gather* : field gathered and summed on master process: costly but bit-for-bit reproducibility
- *lsum8*: local sum by each process, sent to all other processes, then global sum by all: fastest but no bit-for-bit reproducibility
- *lsum16 :* as *lsum8* with quadruple precision : 2 x slower than *lsum8* but higher chance of reproducibility
- *ddpdd:* parallel double–double algorithm with single scalar reduction (He &Ding, 2001)
- *reprosum*: fixed point method based on ordered double integer sums (Mirin &Worley, 2012) : bit-for-bit results except in extremely rare cases

cores, mapping	CONSERV unset	CONSERV *lsum8*	CONSERV *lsum16*	CONSERV *ddpdd*	CONSERV *reprosum*	CONSERV *gather*
48, *src*	4.00	8.27	16.78	10.65	17.34	117.72
48, *dst*	4.39	8.02	16.59	10.42	16.98	142.12
180, *src*	1.25	2.21	4.59	2.87	4.85	126.91
180, *dst*	1.56	2.26	4.62	2.92	4.90	130.01

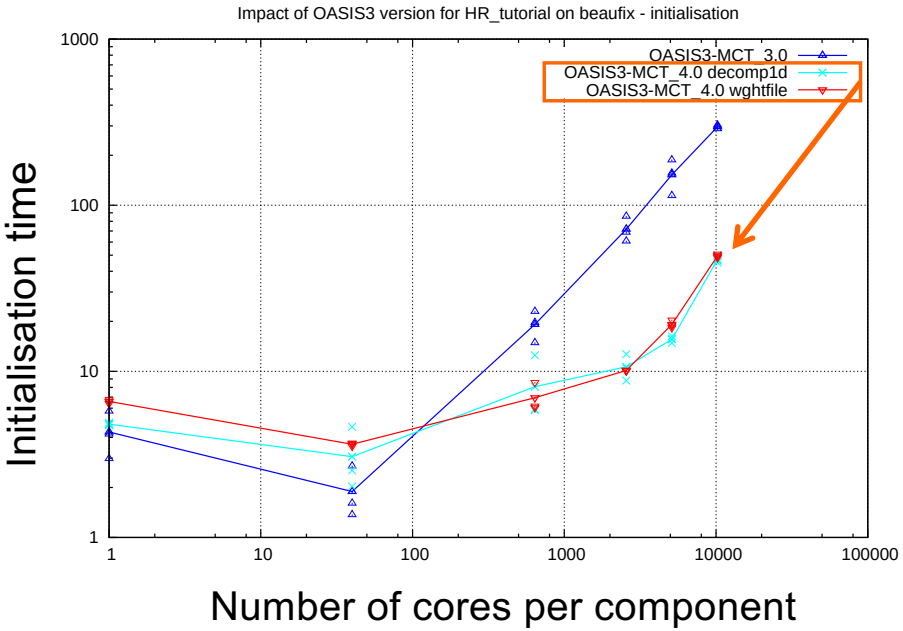
ORCA025 - T799
Cerfacs Lenovo

➢ *reprosum* is the new default: bit-for-bit reproducibility (except in extremely rare cases) and 7/26 times faster than *gather* on 48/180 cores

Debugging of the coupling initialisation

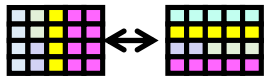
Bugfix: removal of concurrent writing into the OASIS3-MCT debug files at initialization

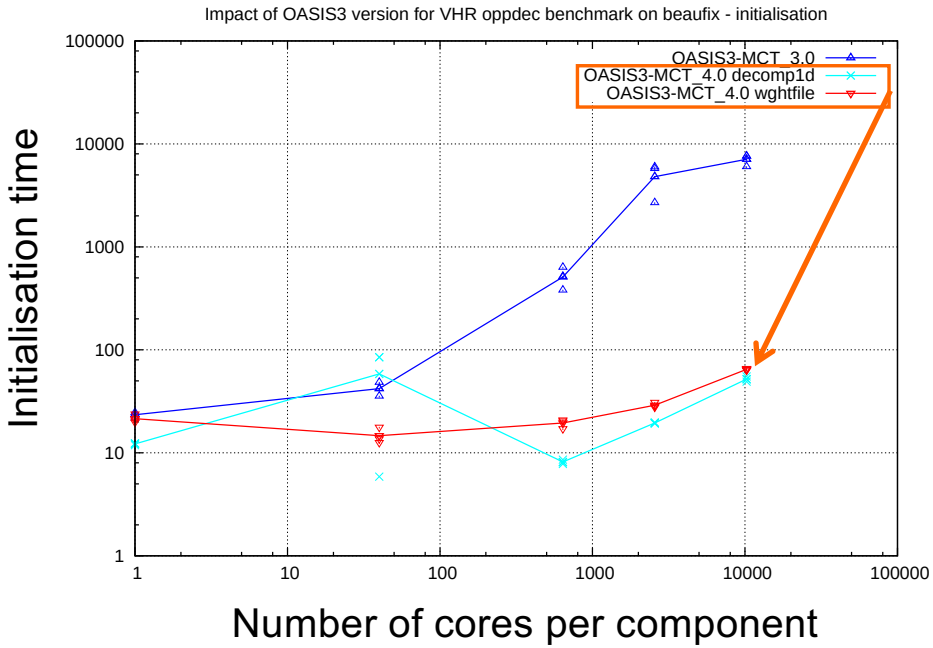
NEMO ORCA025 grid (1021x1442) –
Gaussian Reduced T799 grid (843 000)



⇒ 82% reduction in init time at 10240 cores

IS-ENES benchmark VHR: 3000x3000
reg lat-lon grids, opposite decompositions





⇒ 99% reduction in init time at 10240 cores

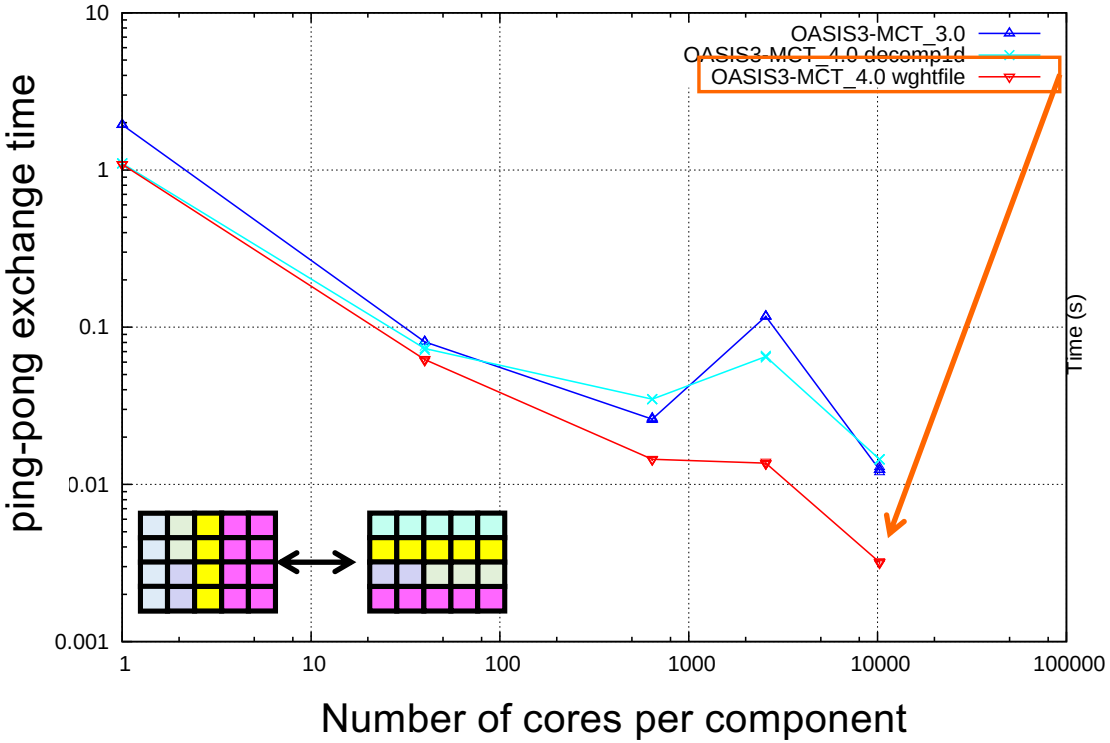
Optimisation of the communication using the mapping weights

Current implementation requires to treat mapping and cross-component communication separately; therefore, a “mapping decomposition” of the target grid on the source tasks is created:

- *decomp_1d :* each target grid point is assigned to a source task in a trivial 1-D way
- *decomp_wghtfile:* a target grid point is associated with the source task that holds the source grid points needed for calculation of its interpolated value:
  - mapping communication reduced (number and size of messages)
  - same cost for sparse matrix multiply
  - same cost communication between source and target

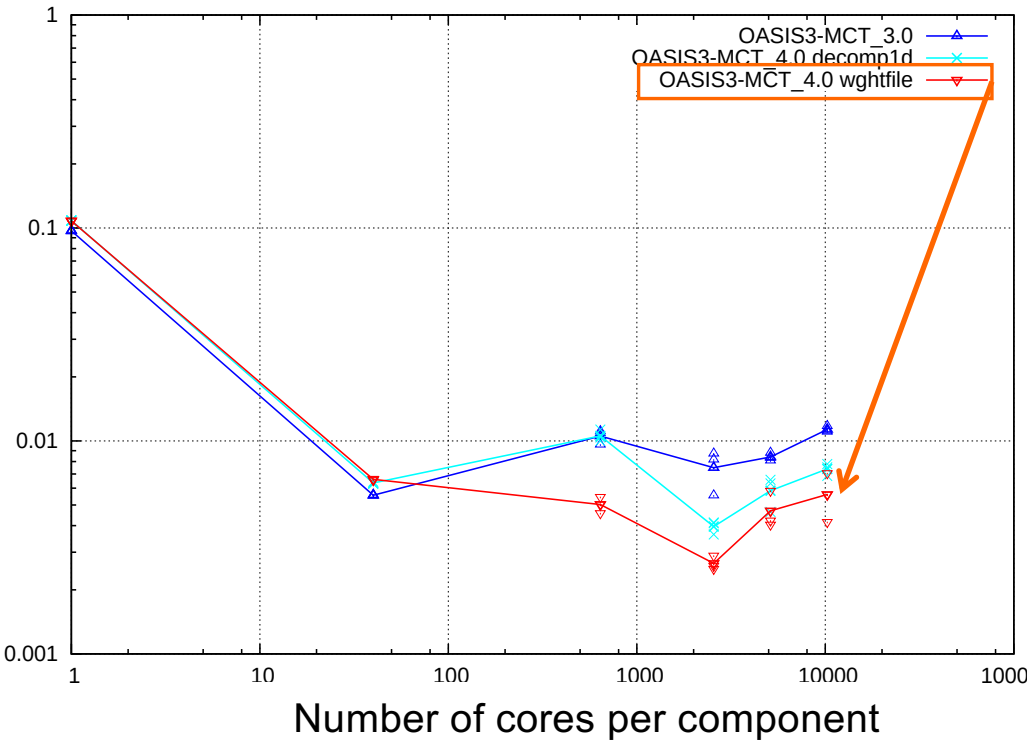
Optimisation of the communication using the mapping weights

IS-ENES benchmark VHR: 3000x3000
reg lat-lon grids, opposite decompositions



⇒ 75% reduction in exchange time at 10240 cores

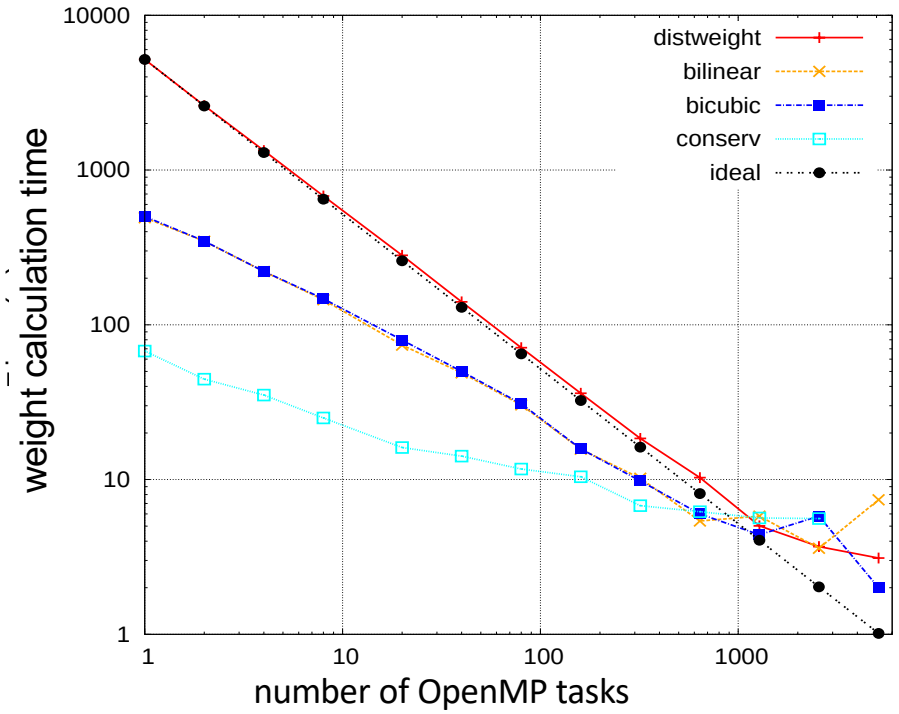
NEMO ORCA025 grid (1021x1442) –
Gaussian Reduced T799 grid (843 000)



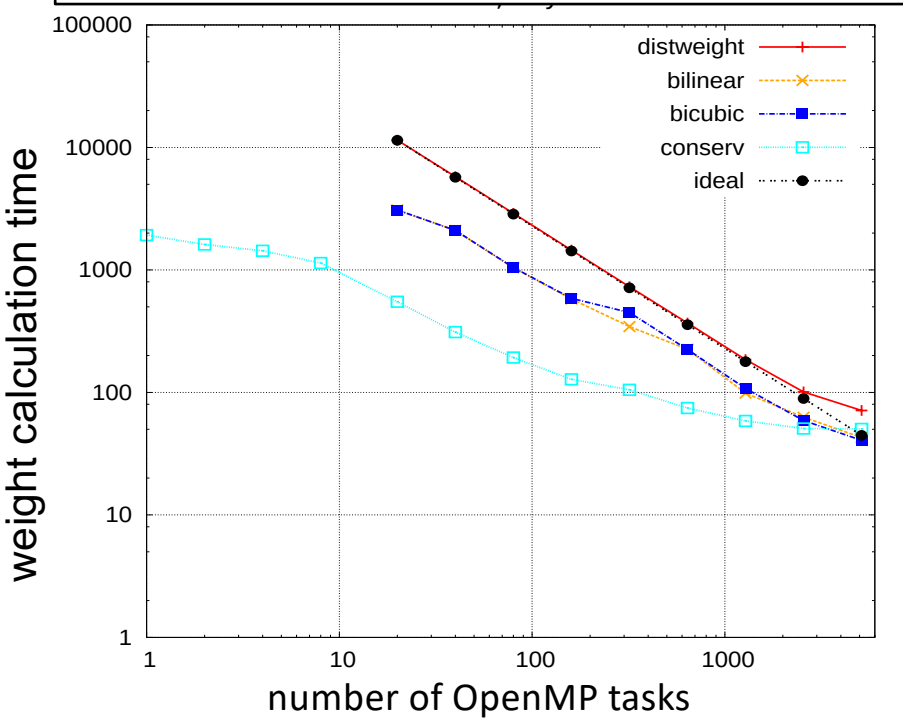
⇒ 24% reduction at 10240 cores

Hybrid MPI+OpenMP parallelisation of the SCRIP library

ORCA025 (1442x1050) – T359 (181724)



ORCA12 grid (4322x3147) – T799 grid (843490)



- Almost perfect scalability for nearest-neighbour and bilinear (-> 1280 tasks for HR;-> 2560 tasks for VHR)
- Good scalability for bicubic remapping
- Less scalability for conservative remapping, due to better sequential performance (bin restriction)
- **Reduction in the weight calculation time of O(10)-O(100) at high number of cores**

On-going developments

- Beta version of API for python codes (STFC)
- Additional and improved diagnostics
- Additional pre- and post-processing transformation
- Redesign of the LUCIA load-balancing tool (with BSC)
- Analysis of ESMF, XIOS, YAC, ATLAS, MOAB-TempestRemap interpolation library
- Replacement of SCRIP with one of the above
  - ➢ OASIS3-MCT_5.0, 12/2021
- Support of grids with mask evolving with time (?)

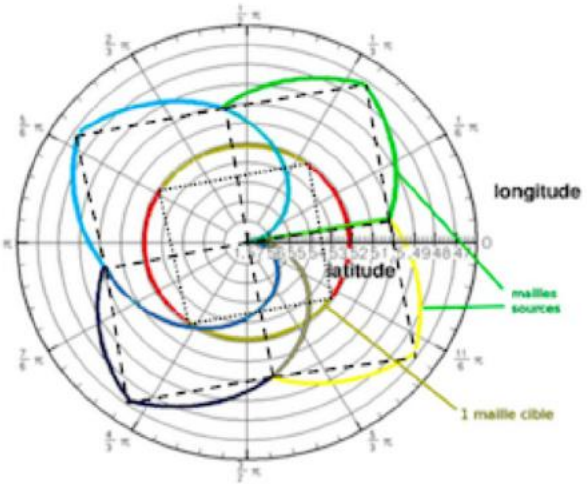
Analysis of SCRIP library

Detailled analysis of the quality of the SCRIP library
(CERFACS tech reports: Jonville & Valcke 2019, Valcke & Piacentini 2019)

- 4 grid types : lon-lat, logically-rectangular, icosahedral, Gaussian-reduced
- Two normalisation options : FRACAREA (intersected area) and DESTAREA (full area)
- Impact of Lambert equivalent azimuthal projection

For lon-lat, logically-rectangular, icosahedral grids:

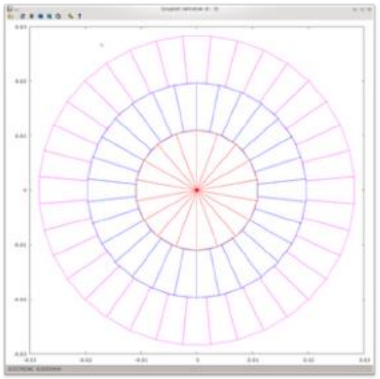
- FRACAREA OK for all grids with and without Lambert projection
- DESTAREA OK for all grids but
  - log.rect <-> lon-lat: only if Lambert projection is activated
  - icos -> log.rect: only if Lambert projection is not activated
  - log.rect -> icos: Lambert projection does not change the results



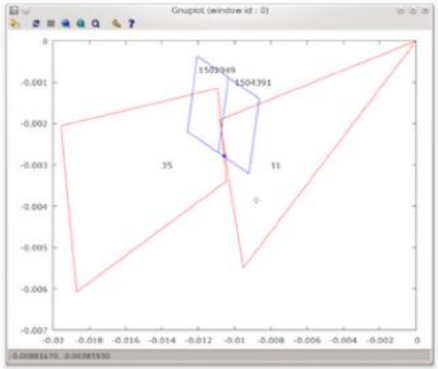
Cell distortion near the pole
without Lambert projection

Analysis of SCRIP interpolation library

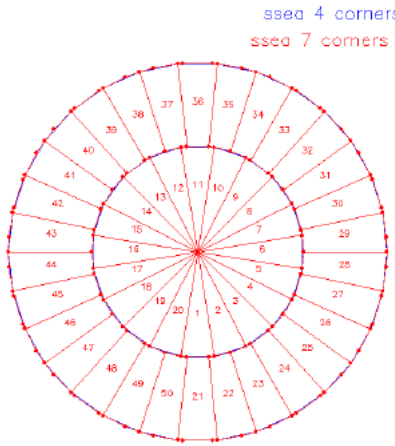
Detailled analysis of the SCRIP remapping quality for the Gaussian-reduced grid :



Gaussian-Reduced cells defined with 4 vertices : corners of a cell do not necessarily match the corners of a neighbour cell



Gaussian-Reduced cells defined with 4 vertices (in red) in Lambert space : the grid cells do not completely cover the globe



Gaussian-Reduced cells defined with 7 vertices: corners of a cell match corners of a neighbour cell

For Gaussian-reduced grids:

- FRACAREA OK without Lambert projection (4-corner and 7-corner grid)
- DESTAREA not OK: significant error with & without Lambert projection (4-corner and 7-corner grid)

Analysis of other interpolation libraries: on-going …

ESMF:

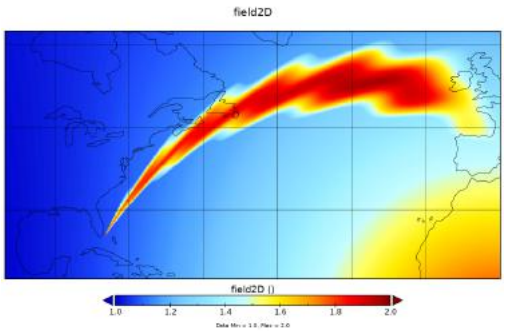
- Very good candidate: complete, efficient, very good user support
- Still some issues with ORCA grid on the North-fold

ATLAS (ECMWF) (CERFACS Tech Rep, Piacentini 2020) :

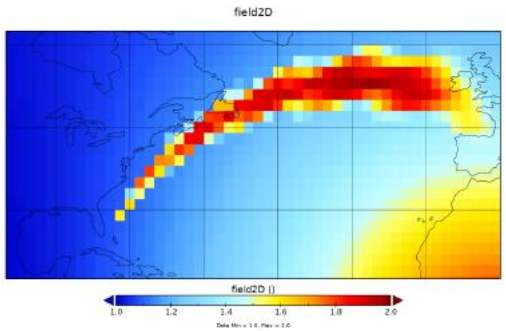
- Appealing, provides a useful portable toolkit for the best usage of new heterogeneous architectures
- But cannot be used in OASIS3-MCT **on the short term** (no handling of masked values, no conservative regrid, low support for geometries and representations from existing models, …)

XIOS (IPSL):

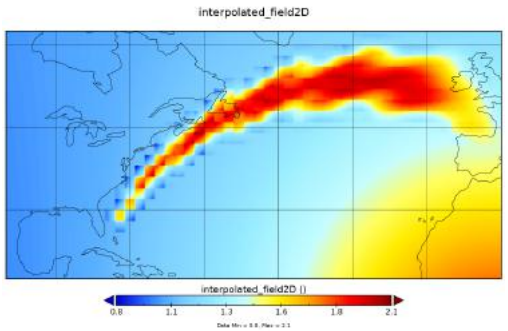
- Very interesting 2nd O conservative regrid (no need to provide field gradient, Kritsikis 2015)



Analytical field on ORCA025



Analytical field on 2^0 reg lat-lon



Interpolated field from 2^0 reg lat-lon to ORCA025

Conclusions

- OASIS is still there and lively
- Large and always growing community
- Active developments and user support (IS-ENES3 and ESiWACE2 funding)
- Needs some reshape of its interpolation library (OASIS3-MCT_5.0 December 2021)
- Next challenges:
  - use of MCT : good enough for next big coupled systems?
  - support of dynamic grids, recalculation of regridding weights during the simulation

Conclusions

- OASIS is still there and lively
- Large and always growing community
- Active developments and user support (IS-ENES3 and ESiWACE2 funding)
- Needs some reshape of its interpolation library (OASIS3-MCT_5.0 December 2021)
- Next challenges:
  - use of MCT : good enough for next big coupled systems?
  - support of dynamic grids, recalculation of regridding weights during the simulation

Acknowledgements

The work presented here has been funded by the IS-ENES3 and the ESiWACE2 projects in the framework of the European Union's Horizon 2020 research and innovation programme.

ESiWACE2 has received funding under the grant agreement N° 823988.

IS-ENES3 has receive funding under the grant agreement N°824084.

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

Join the community on ZENODO !