Summary of 2019 OASIS dedicated supports (IS-ENES3)

E. Maisonnave, CERFACS

Abstract

Following the original IS-ENES-1 program, a dedicated support is provided again to the climate modelling community in Europe, in order to set up, upgrade or enhance coupled systems based on OASIS. After a selection process, ETH Zürich, MetOffice and GEOMAR Kiel laboratories were granted with a total of 3 person-months. During this support, we could upgrade OASIS3-MCT to the current version 4 and make available new coupler functionalities in the coupled systems, such as the parallel computing of interpolation weights. Interfaces are modified to allow single precision computations (ETHZ), concurrent coupling of ocean and ice (MetOffice) or full ocean zoom coupling (GEOMAR). A significant performance improvement is always obtained. Set up configurations are already used for studies (GEOMAR, ETHZ) and our modifications saved in community repositories (FOCI, NEMO). This should facilitate the diffusion of our work to a larger number of laboratories (e.g. AWI, and COSMO & NEMO users) and contribute to the spreading of coupled modelling in our community. In addition, one can find in Appendix the detailed Carbon footprint of this work, with the hope that it could contribute to understand how to make the best benefit of our infrastructure in a sustainable way.

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

The Horizon 2020 European infrastructure project IS-ENES3 (2019-2022) is organising the extension of the existing OASIS support (hotline, training) to a dedicated support, at user site. The 6th work package, named "Services on European ESMs and Software Tool", proposes to provide a technical help to design, upgrade or enhance the implementation of OASIS3-MCT interfaces in models and/or set up a tailored and computationally efficient coupled system. In 2019, a total of 3 person-months of Dedicated User Support was offered to 3 different groups. This service excludes the scientific development tuning, analysis and evaluation of the coupled model components themselves and of the coupled model as a whole, i.e. any comprehensive geophysical study in link with the implemented coupling.

Applicants had to briefly describe their project and their needs filling a questionnaire, available on-line. The collection and a preliminary technical analysis were done at CERFACS and transmitted to a panel of the OASIS Advisory Board members, that had to make the selection, taking into account:

- The originality of the problem: e.g. new physics (ice sheets, hydrology, atmosphere/ocean boundary layer, regional modelling, ...), increased task parallelism (extraction and concurrent running of sub-components e.g. sea-ice), etc.
- The quality of the methodology proposed
- The expected scientific impact of the target coupled system and its long-term support by the applicant group
- The opportunity of development of cooperation with communities outside ENES
- Potential training aspects for new or young users
- The synergy with the OASIS3-MCT development plan

The panel selected the 3 proposals appearing to bring clear benefits to a wide community, and for OASIS testing and demonstration: ETH Zürich, Met Office and GEOMAR Kiel. The panel also suggested that the support should work mainly on maintainable aspects useful for the community. Discussion followed on possible clarification to bring to the call, to help the panel making a more informed choice in the future. Improvements in the call could include:

- more details about the task complexity, to decide whether help is needed or not
- what could be beneficial for the coupler itself in the requested support
- more information about the community that is supposed to use the newly built coupled system: what goes beyond a single application for a given institution
- what is support actually helping

Conclusions of the panel meeting were transmitted to IS-ENES head, which validated the selection.

In the present summary, we give a brief description of the coupled models we set up and we mention the most important result. In a larger report[1], a detailed description of the three technical collaborations is provided, together with their main results. In appendix, we also tried to provide a summary of the support costs, and particularly its carbon footprint.

---

[1]Maisonnave, E., 2019: OASIS Dedicated Support, 4th annual summary , Technical Report, **TR/CMGC/19/149**, CECI, UMR CERFACS/CNRS No5318, France

ETH Zürich (Switzerland), Land-Climate Dynamics group

June 24 – July 19, 2019
Main Goal: Upgrading and performance enhancement of atmosphere-land coupled model on GPU based supercomputer

Summary

To take a maximum benefit of the GPU compliant COSMO model, the last version of OASIS and ClandM were included in the coupled system, the existing OASIS interface in COSMO was modified to allow single precision compiling, and a coupling between heterogeneously (PGI/Intel) compiled components was set up successfully. Altogether, this upgrade leads practically to the multiplication by a factor 1.5 of the COSMO-CLandM coupled system speed.

COSMO-Climate Limited-area Model v5 of the regional atmosphere model, https://www.clm-community.eu From 6 to 2Km, 383x328 (LR) to 801x801 (HR), 60 vertical levels	Community Land Model (CLM, here CLandM) v5, as part of the CESM v2.0.0 coupled framework, http://www.cesm.ucar.edu/ From 25 to 2Km, 288x128 (LR) to 1332x811 (HR)

Supercomputer:

"piz daint", CSCS, Manno, Switzerland

5704 nodes of 12 cores - Xeon E5-2690v3 12C 2.6GHz, and 1 NVIDIA Tesla P100 – Memory per node: 64 Gb

https://www.top500.org/system/177824

Model description

The reference coupled configuration is derived from the first implementation of an OASIS based coupled system developed at ETH Zürich. The atmosphere is represented by the climate compliant version of the DWD-MeteoSwiss COSMO limited area operational model. This model was recently fully re-written to be efficiently handled on GPGPU architectures. A Domain-Specific Language (STELLA) was used to hide the complexity of C++/CUDA coding to the end-users and efficiently perform the dynamical core computations on GPU. In the 5th version of COSMO we used, only I/O and OASIS coupling are still performed on one single core of the CPU host.

The stand-alone version of the model leaves unused 11 or the 12 CPU of the node. The OASIS coupled configuration takes benefit of these idled resources to perform the calculations of an alternative land surface model (CLandM), before disabling the existing TERRA land surface subroutines of the COSMO model. CLandM is the land component of the NCAR Earth System (usually named CLM). A previous modification of the CESM structure ensured the two-way communication of surface fields between the two models across the internal CESM coupler and atmosphere forcing modules. CLandM MPI processes are mapped to the 11 available CPU cores of the node and coupled via OASIS to the COSMO MPI process also located on one CPU core.

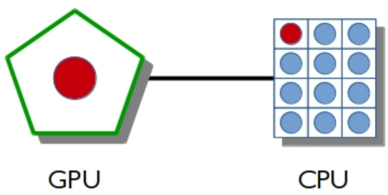

Figure 1: COSMO (red) and CLandM (blue) task binding on GPU (CUDA threads) and CPU (MPI processes)

To take the maximum of performance of GPU resources, the PGI compiler v18.10.0 was used, along with the MPI cray-mpich/v7.7.2 library.

The replacement of TERRA original soil model by CLandM proved its relevance and this additional value is obtained with the same amount of energy, considering that CPUs are still consuming energy while standing. However, the sequentiality of COSMO/CLandM calculations necessarily increases the restitution time compared to COSMO stand-alone (even including TERRA subroutine cost).

Main result

The model speedup, resulting from the 4 improvements we proposed during the support period, is measured during a 2 days long simulation, using the LR version of the model on 9 nodes, in a close to the optimum parallel decomposition. Initialisation and termination phases of the execution are excluded. Variability of restitution time, not fully estimated, is below 5%.

The 2D CLandM computations, also performed at lowest resolution than COSMO (25Km vs 6Km), run much quicker (1 order of magnitude). The consequence is that the improvements made on the CLandM side have a minor impact to the overall coupled system speed. The COSMO calculation precision effect is as expected close to a 40% increase of the speed. The same effect, but smaller, is obtained with the HR configuration (20%).

	COSMO timing (s)	CLandM timing (s)	Coupled model speedup (%)
reference	530	32.	
gpts bounding	530	8.2	4 %
single precision	305	8.2	41 %
Intel compiling	305	3.9	1 %
Overall improvement (sequential) :			45 %
concurrent cpl	305	0.1	46 %

Table 1: elapsed time of 2 simulated day long run of COSMO-CLandM-LR coupled system, after modification of the reference configuration by (i) bounding the CLandM active grid points to COSMO domain limits, (ii) changing for single precision real variable in COSMO, (iii) compiling CLandM with a different compiler (Intel) and (iv) performing both COSMO and ClandM calculations concurrently

However, a load balanced configuration will better benefit from the CLandM related improvements. The Intel compiling can potentially divide by two the CLandM restitution time. The removal of off limit grid points on reasonably well fitted latitude-longitude domains can also significantly speed up the land model. Finally, the concurrent coupling mode, which physical results have to be preliminarily validated, gives the possibility to use the idled CPU of a COSMO-GPU model without any additional cost, nor from respect to energy or restitution time. In its HR configuration, the coupled system speed exhibits a 10% increase.

~

MetOffice Exeter (UK), Climate Science IT Applications

August 26 – September 20, 2019
Main Goal: Setup an OASIS coupling between NEMO ocean and SI3 sea ice components

Summary

Coming with the new NEMO 4.0 version, the recent upgrade of the sea-ice component from LIM to SI3 makes necessary a check-up of the ocean/surface_module coupled interface. Few code modifications, included in a development branch for later trunk update, were necessary to perform test simulations at ORCA1 and ORCA12 resolution and roughly check its validity. Improvement of NEMO speed and cost is real but limited to 10 to 20% and observed with sufficiently high decomposition only. At its best, our coupled configuration is faster (x2) and cheaper (-25%), but, since it is spread on a larger number of resources, it could reduce the actual speed (simulation + scheduling time) of production runs.

NEMO, ocean (OPA)	NEMO, sea ice (SI3)
v4 of the global ocean model, https://www.nemo-ocean.eu	Included in the surface module (SAS) together with flux computations and icebergs
From ORCA1 to ORCA12, 31 to 75 vertical levels	Same resolutions, same grid, but possibly different decompositions

Supercomputer:

CRAY XC40, “xce”, MetOffice/Science Park, Exeter, UK

2496(+6636) nodes of 36 cores - Xeon E5-2695V4, 18C, 2.1GHz

https://www.top500.org/system/178925

Model description

The NEMO ocean model is currently included in several configurations of coupled models, for climate modelling but also operational purposes. A large range of resolutions from ORCA1 to ORCA12, and coupled components (from atmosphere-ocean only to ESM components) makes necessary the modularity of the OASIS coupler and the choice of interpolation offered by the SCRIP and ESMF libraries.

In various contexts (CMIP6 exercise, operational forecasting …), speed and cost are more than crucial quantities that this support proposes to deals with. Met Office IT group identified the ocean/sea-ice interface as a potential source of computing performance enhancement. As demonstrated previously, the separation of the ocean surface module (SAS), including sea ice model, from the ocean related routines of NEMO (OPA) in two executables, and their coupling via OASIS, allows the concurrent performing of the two set of calculations, increasing the model speed and, in some appropriate conditions of load balancing, reducing its cost.

The recent upgrade of the sea-ice component from LIM to SI3 makes necessary the check-up of the OPA/SAS coupled interface. Component scalability changes resulting in particular from recent Polar folding communication improvements may have jeopardized the expected advantages of the OASIS coupling. This is what this study proposes to verify.

Main result

A set of performance simulations, led during 100-time steps, is produced with a high-resolution configuration (ORCA12, 1/12 degree horizontal resolution). A version incompatibility between OASIS and XIOS libraries prevent to take a comprehensive measure of the full IO configuration at high scalability. Consequently, the results given below do not include the 12 nodes usually allocated to XIOS at ORCA12 resolution.

The coupled setup reduces simulation speed and increase its cost if a relatively few numbers of resources are allocated. The reason why a good load balancing cannot be achieved at lower decomposition is the total memory requirement of the model (the SAS module, which includes most of the memory bound 3D NEMO variables, was impossible to launch on less than 8 nodes).

Using a larger number of resources, the load imbalance can be reduced and the coupled mode slowly become faster and cheaper than the reference configuration. However, this superiority occurs for such a high number of resources that they are usually difficult to obtain from the batch scheduler. At its best, our coupled configuration is faster (x2) and cheaper (-25%), but its extra resource requirement could reduce the actual speed (simulation + scheduling time) of production runs.

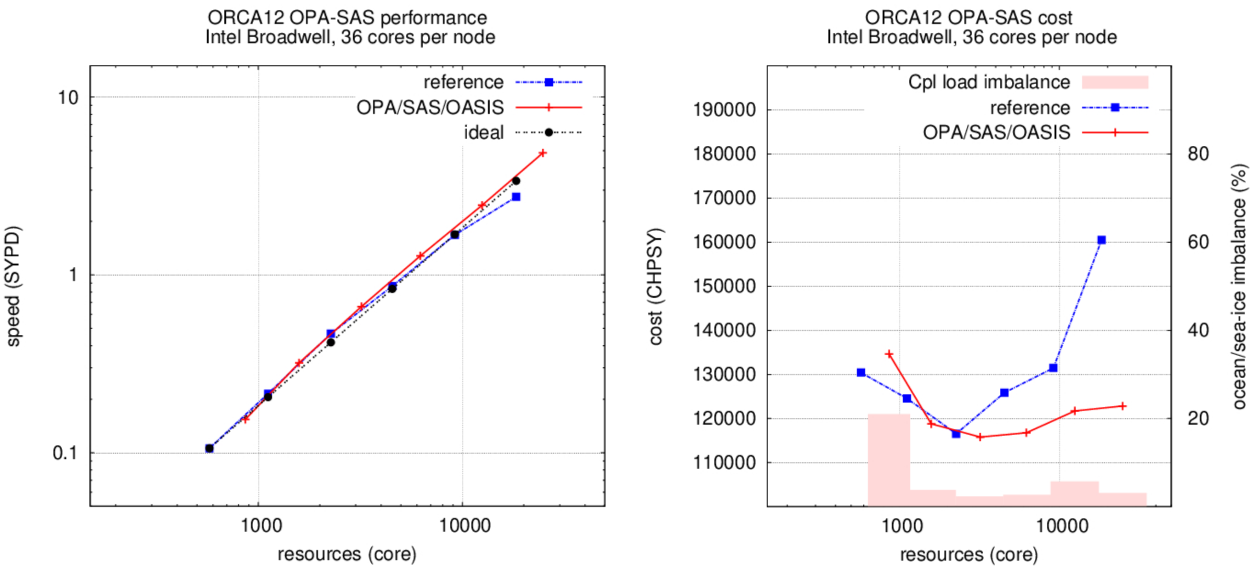

Figure 2: Speed and cost of the NEMO 4.0 model, ORCA12 grid, no land-only subdomain removal, with ocean and surface module (including ice) components coupled via OASIS (red) and in regular uncoupled mode (reference, blue). On right plot, load imbalance of the coupled experiments, i.e. ratio between fastest component waiting time in coupling and total elapsed time (in time loop). No XIOS output. In regular mode, 16-31-63-126-255-510 nodes are allocated to NEMO. In coupled mode, the same node number is allocated to the OPA and 8-12-24-44-92-180 additional nodes are allocated to SAS

GEOMAR Kiel (Germany), Marine Meteorology team

September 30 – October 19, 2019

Main Goal: Extending OASIS coupling between OpenIFS and NEMO ocean to an AGRIF zoom

Summary

The necessary removal of the on-disk coupling procedure of the ocean zoom surface fields in the GEOMAR OpenIFS-NEMO-AGRIF coupling (FOCI) required the upgrade of both NEMO and OpenIFS interfaces. The unexpensive OASIS coupling that has been set up allowed to increase the OpenIFS horizontal resolution to 25Km. The CPU cost of the coupled system, that includes the North Atlantic zoom AGRIF, is estimated to approximately 30 time less than the CPU cost of the corresponding global ORCA12 based configuration.

OpenIFS, atmosphere	NEMO, ocean
cy40 of the global model, https://confluence.ecmwf.int/display/OIFS	v3.6 of the global ocean model, https://www.nemo-ocean.eu
From T159 (85Km) to T799 (25km), 91 vertical levels	ORCA05, 46 vertical levels. Includes AGRIF zooms (e. g. North Atlantic)

Supercomputer:

"mistral", DKRZ, Hamburg, Germany

1368 nodes of 24 cores - Xeon E5-2680v3 12C 2.5GHz – Memory per node: 64 Gb

https://www.top500.org/system/178567

Model description

The ocean-atmosphere coupled model currently used at GEOMAR Kiel substantially differs from the standard NEMO based coupled models of the climate community. It is derived from the former Kiel Climate Model System, including ECHAM atmosphere. The NEMO global ocean horizontal resolution is 0.5 degrees (ORCA05) but it can be increased in selected areas (Southern Ocean, North Atlantic …) taking benefit of the AGRIF functionality. In the new FOCI (Flexible Ocean and Climate Infrastructure) configuration, ECHAM is replaced by the licensed software of ECMWF IFS, OpenIFS. Our study will focus on two horizontal resolutions: a coarse T159 (~125km) and an accurate T799 (~25km). The global ocean NEMO v3.6 in ORCA05 grid includes a zoom over North Atlantic Ocean (VIKING) at 1/10 degree resolution. A runoff remapping tool complements this climate model. NEMO outputs are speed up by the XIOS I/O server (detached mode) but OpenIFS still ensures its own output (serial mode, GRIB format).

This model can be handled, in various configurations (different resolution, different zoom location) thanks to the community working environment ESM-Tools developed at AWI, Bremerhaven by Dirk Barbi. In this environment, several other OASIS based coupled models are available, from which the former KCM model and the FESOM based system AWI-CM.

The existing FOCI configuration relies on the version 2 of the OASIS3-MCT coupler. Only coupled fields discretised on the NEMO global grid are exchanged through the coupling library. The existing OpenIFS and NEMO coupling interfaces are not prepared to allow the exchange of AGRIF child grid fields. Consequently, AGRIF needs to read its forcing flux condition in a file (NEMO forced mode). This file is updated online by fluxes provided by OASIS, thanks to the EXPOUT option. The main drawback of this solution is the prohibitive cost of disk writing at high resolution. The purpose of this dedicated support is to substitute to this on-disk coupling solution a full OASIS coupling, through standard MPI communications.

Main result

The necessary removal of on-disk coupling procedure required the upgrade of both NEMO and OpenIFS interface, and the modification of OASIS parameters. As shown in Fig 3, the coupling exchanges must take place not only in NEMO parent grid related subroutines, but also on the AGRIF part of the code.

Figure 3: Sequence of coupling exchanges between OpenIFS and NEMO, including an AGRIF zoom

A three-month long simulation gives a sufficient garantee of model stability and validates the technical implementation of the full OASIS coupled FOCI-AGRIF model. After a computing resource tuning of the 2 models (load balancing), a speed of 1 SYPD is measured, for a cost of 17,000 CHPSY. The AGRIF zoom multiplies by a factor 10 the cost of the global ORCA05 model but this cost has to be compared with the cost of the ORCA10 (global 1/10 degree) that would be necessary to set AGRIF spatial resolution of the North Atlantic Ocean globally. This cost could be roughly estimated to 300 times the ORCA05 one. In conclusion, the set-up of this configuration will help to lead selected studies relying on high resolution modelling that costs approximately 30 times less than the same studies relying on the standard ORCA12 configuration.

Conclusion

During the selection procedure, the panel emphasised the importance of not restraining the support to a one to one collaboration but to rather prefer actions that could have a broader impact on communities. We tried to quantify this community impact, in a table that summarises (i) the oral communications organised and the origin of the participant/audience, (ii) code updates in official centralised repositories, from which OASIS gitlab and (iii) written communications (emails) to laboratories making part of the hosting laboratory working network. This counting necessarily neglects any action in link with our work, organised by the hosting laboratory, that could take place after the dedicated support period.

	ETH	MetOffice	GEOMAR
Talks/meetings	* Starting meeting (3 people)	* Starting meeting (8 people, internal, across teams) * Closing presentation (10 people internal, across teams)	* Monday's informal modelling meeting (3 to 4 people, internal, across teams) * Closing presentation (15 people, internal, across teams) * ESM-Tools Workshop meeting (by GEOMAR staff, 10 people, external)
Repository updates	none	NEMO, MetOffice branch OASIS tickets	NEMO, v4 and trunk FOCI git repository (DKRZ) OASIS3-MCT, v4 and master
Networks	MeteoSwiss (1 email)	Exchanges with ESM, HPC and biogeochemistry community of MetOffice (~10 emails)	Exchanges with DKRZ (4 emails)

Table 2: Quantification of community level communications during support

Of course, community impact cannot be fully evaluated on such short period and using such dangerously formal criteria, but it gives an idea on how practically results of the support can be used outside the hosting laboratory.

Appendix

Costs/Sustainability

Budget, energy consumption and carbon footprint are provided in the following table. Computations and train transport are the only two items considered in this summary. Everyday consumption, from which electricity supply for workstation (Intel Atom N270 or Arm Cortex A53) and supercomputer login nodes, is neglected. Energy/CO2 emission conversion (Carbon Intensity) for transport and supercomputing is country and machine dependant.

A comparison of transportation and computation Carbon footprints clearly puts in evidence that a comprehensive effort made to avoid airplane journeys is not sufficient to lower the total amount of greenhouse gaz emissions by more than 50%. The reason is that even short set up tests led with high resolution models, like NEMO-SI3 at ORCA12 resolution, were sufficient to strongly increase the ecological impact of our work.

	Cost	Travel		Computing			Total Carbon footprint (KgCO2e)
	(€)	(Km)	(KgCO2e)[2] [3]	(Core.h)	(kWh)	(KgCO2e)[4]	
ETHZ	3670	2,700	14	1,200	7[5]	0	14
MetOffice	2790	1,200[6]	20	40,000	750[7]	445	465
GEOMAR	1140	1,950[8]	30	8,500	114[9]	86	144
Total	7600	5,850	64	49,700	871	531	623

This quantitative cost analysis makes no sense without the corresponding qualitative analysis of the work produced. In particular, one would pay attention to the capacity of some coupled models, set up during this program, to reduce the computing cost compared to other configurations, while giving the same quality of results, from a geophysical point of view.

[2]SNCF carbon intensity high speed train : 2,4 gCO2equ/Km, intercity : 8.1 gCO2equ/Km, from https://www.oui.sncf/aide/calcul-des-emissions-de-co2-sur-votre-trajet-en-train and https://ressources.data.sncf.com/explore/dataset/emission-co2-tgv/

[3]Paris/London Eurostart journey carbon intensity 4.1 KgCO2equ https://www.eurostar-treadlightly.com/en/environment.php and 41,15 gCO2equ/Km for National Railways https://www.gov.uk/government/publications/greenhouse-gas-reporting-conversion-factors-2019

[4]Carbon intensity of High voltage in Switzerland (~29), UK (593) and Germany (599), according to Moro A., Lonza L., 2018: Electricity carbon intensity in European Member States: Impacts on GHG emissions of electric vehicles, Transportation Research Part D: Transport and Environment, 64 , pp. 5-14.

[5]CSCS CRAY XC50 consumption for 388,000 cores: 2,384kW, PUE < 1.25, see TOP500

[6]Return ticket is not taken into account since it was used for another purpose than this dedicated support (IS-ENES Sea-Ice Workshop)

[7]MetOffice CRAY XC40 consumption for 90,000 cores: 1,348kW, PUE = 1.25, see TOP500

[8]Single ticket is not taken into account since it was used for another purpose than this dedicated support (IS-ENES Sea-Ice Workshop)

[9]DKRZ mistral consumption for 99,000 cores : 1,116kW, PUE = 1.19. See TOP500 and https://www.dkrz.de/communication/news-archive/en-energie-effizienz-des-rechnersystems-mistral