IS-ENES – WP 4 – D4.7

Original title in the DoW: **General coupled model assembling guide based on practical experience gained when providing dedicated user support**

Revised title: **OASIS Dedicated User Support**

Abstract:

Within IS-ENES, work package 4 "Strengthening the European Network on Earth System Modelling" proposed to provide technical help to implement new coupled models or improve existing configurations based on the OASIS coupler. Calls for applicants have been opened at the beginning of each year of the project (2009-2012). A total of 3 person-months were available each year. 31 applications, from 8 different countries, have been assessed by an IS-ENES Selection Committee.

Technical reports of the 12 OASIS Dedicated User Support (ODUS) are provided in the Appendix. Relying on their results, we described here our vision of the “multiple executable” coupling strategy adopted by the European climate community and how our project contributed to enhance it.

Grant Agreement Number:	**228203**	**Proposal Number:**	**FP7-INFRA-2008-1.1.2.21**
Project Acronym:	**IS-ENES**		
Project Co-ordinator:	**Dr Sylvie JOUSSAUME**		

Document Title:	OASIS Dedicated User Support			**Deliverable:**	D 4.7
Document Id N°:		**Version:**	1.0	**Date:**	March 13th 2013
Status:	Final				
Filename:	ISENES_Deliverables_4.7.docx				
Project Classification:	Public				

Approval Status		
Document Manager	**Verification Authority**	**Project Approval**

REVISION TABLE

Version	Date	Modified Pages	Modified Sections	Comments
0.1	2013/01/10			First version by E. Maisonnave
0.2	2013/03/08			Revision by S. Valcke
0.3	2013/03/11			Revision by M.-A. Foujols

Executive Summary

Within IS-ENES, work package 4 "Strengthening the European Network on Earth System Modelling" proposed to provide technical help to implement new coupled models or improve existing configurations based on the OASIS coupler. Calls for applicants have been opened at the beginning of each year of the project (2009-2012). A total of 3 person-months were available each year. 31 applications, from 8 different countries, have been assessed by an IS-ENES Selection Committee.

This one person.year work has produced the implementation of HPC compliant interfaces (ECHAM for NEMO and FEOM), interfaces for regional modelling (COSMO-CLM-Parflow), an interface with already integrated Earth-system (COSMO-CESM), an interface for 3D regional/global two way nesting (COSMO-ECHAM) and an interface for global model with zoom (NEMO/ERNA-ARPEGE).

Fact reports were produced on OASIS4 and OASIS3-MCT behaviour and performances on HPC configurations. It helps to optimise climate model performances on supercomputers (Ec-Earth, COSMO-CLM), developing a specific tool for performance measurement (“lucia”).

As expected, the ODUS program served IS-ENES partners, other community laboratories and CERFACS, via bug reports and coupler enhancement suggestions based on practical experiments.

Relying on these results, we describe here our vision of the “multiple executable” coupling strategy adopted by the European climate community and how our project contributed to enhance it.

1. Coupling strategies

If we want to understand how important is “multiple executable” coupling for climate modelling in Europe, let’s characterise how a coupler (OASIS) for MPMD (Multiple Program Multiple Data) is used and by which kind of laboratory.

The IS-ENES ODUS program has been organised in 7 different laboratories. If we look at the different granted institutions (see Table 1), we can notice that only 2 are national climate modelling centres (Rossby and Hadley centres, associated to SMHI and Met Office Meteorological agencies). Others laboratory activities are focused on ocean (AWI, LOCEAN, UBO), land surfaces (ETHZ and University of Bonn, also supported but outside the IS-ENES program) or atmosphere (BTU).

Laboratory	Models
SMHI (Sweden)	IFS (atmosphere) – NEMO (ocean), *Ec-Earth* RCA (atmosphere) – RCO (ocean) RCA (atmosphere) – NEMO (ocean)
LOCEAN-IPSL (France)	ECHAM (atmosphere) – NEMO (ocean) WRF (atmosphere) – NEMO (ocean)
AWI (Germany)	ECHAM (atmosphere) – FEOM (ocean)
ETHZ (Switzerland)	COSMO (atmosphere) – CLM/CESM (soil) – *Parflow (hydrology)*[1]
BTU (Germany)	COSMO (atmosphere) – ECHAM (atmosphere) – MPI-OM (ocean)
The Met Office (UK)	UnifiedModel (atmosphere) – NEMO (ocean) – WaveWatch (waves)
UBO (France)	ARPEGE (atmosphere) – NEMO (ocean)

Table 1: Supported coupled models in IS-ENES granted laboratories

OASIS is used in climate laboratories to assemble the various climate model components developed inside or outside the institution. But in most of the specialized laboratories, it can be seen as a tool that allows to replace, in the laboratory main model, forcing boundary quantities by an exogenous model that produces those quantities.

This practice, facilitated by the community free movement of models, is widely spread in a European context, where national and continental scientific programs[2] encourage, organise and fund such activity. Furthermore, laboratories can share model development and support[3], which help creating communities: within those groups, additional components to the main model can also be exchanged.

But it is obvious that this or those additional models cannot be supported as the main model is. To support and develop a model is such an expensive task (which requires advanced competences) that it cannot be done with several components by all research groups, but only in bigger climate modelling laboratories.

To make possible the use of an extra model, without developing it, a modular interface is required. Most of the ODUS activity focused on writing, extend or improve performances of those interfaces

1 For COSMO community, in collaboration with Bonn University
2 One can cite, non exhaustively, Pulsation (France), TR32 (Germany), EMBRACE, SPECS and other IS-ENES WPs (FP7)
3 COSMO (atmosphere), Ec-Earth (climate), NEMO (ocean) ...

in the targeted models. The reader can find in table 1 the various coupled systems implemented or enhanced during those 4 years. Initially implemented for two components, an interface can be extended for other exchanges[4]. When the coupling interface is modified by a model community member to allow the coupling with a new module, this module is theoretically made available for all the model community users: this highly increases the diffusion of a module through the community.

Let's immediately notice that to use a module in a coupled system does not necessarily imply to fully understand its how it works (even less its co-functioning with the main model).

2. How is OASIS used ?

One of the ODUS benefit is the transfer of knowledge about how a coupling is set and how a coupling interface is implemented.

Coupling is an exercise that must be considered complex. Its principle is simple. Its implementation is not. To split it into two parts, one theoretical (implementation of mathematical formula, mainly devoted to computing scientists) and one empirical (validation with observations comparison, mainly devoted to geophysical scientists) is an obstacle to its understanding.

To simplify, we can report from ODUS two kinds of coupling management. When it is an integrative activity of the laboratory (Climate centres) or of an (even single) scientist long term research program. And when it is an exploratory activity (like trying, for the first time, to switch from fixed boundary conditions to coupled quantities).

In the first case, most of the time, the OASIS functioning is well known. ODUS has served to extend its use or performances of new configurations on new supercomputers. The knowledge of external modules depends on the research program maturity.

In the other case, a time limited project is often the occasion of the interface implementation. The laboratory takes benefit of ODUS to set up a single configuration of the model, that will serve to provide scientific results in a restricted study. Even though few things are known of the exogenous module, non intrusiveness and simplicity of OASIS coupling theory leads to fast results (it is the goal) but also to weak implementations[5].

3. How to enhance coupled model implementation?

There is no clear solution that can be recommended to any kind of project. Nevertheless, any partner in a coupling implementation, from coupling implementers and laboratory institutions to OASIS developers, should avoid some basic traps, that we will mention. In conclusion, we will suggest how an extended ODUS could contribute to this.

It is generally admitted that the theoretical (implementation) part of a coupling is completed when a

[4] For example, OASIS gives the possibility to exchange coupling fields through the single master processor. If user does so, it won't be necessary for him understand models parallelism strategy and but this would lead to dramatically slow down performances when his model resolution will increase

[5] For example, the possibility to exchange coupling fields through the single master processor avoids to understand models parallelism strategy but leads to dramatically low performances when resolution increases

given number of coupling time steps[6] has been achieved successfully (without obvious drifts). Coupling implementers should take care not to entrust anyone with the task of starting coupling validation[7] at this point but to carry out by themselves.

All details of the initial implementation are important and some of them will necessarily be modified during validation (or when the model will be later used in another configuration). They can not be easily changed by a third party because (i) they are spread on different parts of, at least, two different (and, most of the time, legacy) programs, (ii) their parametrization is difficult[8] and sometimes depends on different sets of parameters (one per coupled model) and (iii) the relative simplicity of use and adaptability of OASIS could lead to subtle but esoteric implementations that get documentation work more expensive.

Unsurprisingly, the author strongly suggests that both computing and geophysical validation skills were associated to coupling implementation work. He still does not lose hope that an even better solution, the double skills for the same person, would be, at last, efficiently encouraged at any level of our institutions.

More practical aspects can also be enhanced when a model coupling implementation is considered. First, it is important to consider it as a complex operation, that must be organised and which requires adequate tools.

When an exogenous model is added, it implies that the existing workflow is modified. This technical aspect of a coupling operation is neglected, although it can be highly time consuming. It is usually a strong reason of frustration for coupling implementers. We suggest to consider workflow adaptation as a work *per se*. Enabling an OASIS coupling on models is complex enough for being done separately from the workflow modification.

In a first step, it is crucial to work in a simplified environment. To start with a workflow already set for production is definitely not a good idea. Once completed a first coupling validation run, the merging of modules workflow could be done accordingly to this initial simplified environment.

Performance tuning can also be a crucial part of the work and can even be the condition for production[9]. Sometimes, model results can be changed by this tuning. For example, when the sequence of coupling exchanges is modified. The impact of this change must then be evaluated[10].

Advices for interface implementation are expressed in the exhaustive OASIS documentation. Let's emphasize again on the necessity to call OASIS exchange routines (prism_put and prism_get) at each time step. Interface implementers are reluctant to let OASIS calculate averages or accumulations. We remind that those operations are performed locally (no need of MPI exchanges, then no extra cost). The advantages strongly appear when more than two models are coupled with different coupling time steps. For the same reason, when necessary, it is recommended to let OASIS save coupled fields on separated restart files.

6 For an ocean-atmosphere coupling, a one month long run meets the needs of a simple validation

7 For example with observation comparison

8 OASIS namcouple gathers some (but not all) of them. For example, how to simply parametrize the length of the boundary zone between parent and child grid coupling fields, in an coupling model including a zoom (see ODUS #12)

9 To be convinced, the author will have a look to the different performances of the same ECHAM-COSMO model as observed during ODUS #10

10 See ODUS ## 7 and 9

Other benefits of this program arose through collaborations with computing centres (organized, on the last 2 years, in the PRACE IP programs). It gave us the possibility to emphasize the importance of different MPI characteristics for climate modelling in Europe such as:

- MPMD mode (several executables launched on the same MPI execution environment)
- multi-threading (more than one MPI process can be run on a single resource) that can be interesting when models are running sequentially
- mapping (to explicitly choose the position of a given process on a particular resource)
- mixed OpenMP-MPI mode

and any combination of those 4 features, that most of the time require a particular and subtle MPI parameterization from supercomputer administrator.

4. How to enhance OASIS ?

When one start testing his recently implemented interface, the first question is always: which module of my coupled system has failed ? The difficulty is that we cumulate at least three problems: the lack of model error handling (sometimes), the lack of model error handling on coupling interface (it depends on the interface implementer himself) and the lack of coupler error handling.

During the recent writing of OASIS3-MCT, a more precise error handling has been included. The ODUS program gave the opportunity to intensively test the coupler on a lot of different configurations and a lot of easy-to-do mistakes has been reported. However, to imagine new mistakes is one of our civilization favourite activities and preventing them all is such an important work that its implementation in a FORTRAN software like OASIS could lead to multiply line code number by a factor 2 or 3 and definitely darken its algorithm.

We noticed that OASIS3-MCT includes a much better error handling than on past OASIS versions. Nevertheless, an extra effort would be appreciated to avoid that people wrongly attribute bugs to coupler and not to their own coupling implementation[11].

To conclude this chapter, one can observe during the past 4 years the increasing importance of regional modelling. Naturally, OASIS has been selected to set up numerous European regional configurations.

As we proved it, OASIS can be used for regional coupled modelling. Nevertheless, some missing important features would be very useful to simplify the definition of coupled region boundaries[12].

5. How ODUS could contribute to OASIS better use ?

The purpose here is not to propose enhancements to the ODUS (mostly because this program stopped) but to make an assessment, made by its main contributor.

This one person.year work has produced the implementation of HPC compliant interfaces (ECHAM for NEMO and FEOM), interfaces for regional modelling (COSMO-CLM-Parflow), an interface with

[11] That could lead to the complete implementation withdrawal

[12] Most of the time, there is a geographic mismatch between regional grid extensions of the two models. This could at least, slow down coupling performances, but, sometimes, even forbid interpolation weight computations

already integrated Earth-system (COSMO-CESM), an interface for 3D regional/global two way nesting (COSMO-ECHAM) and an interface for global model with zoom (NEMO/ERNA-ARPEGE).

Fact reports were produced on OASIS4 and OASIS3-MCT behaviour and performances on HPC configurations. It helps to optimise climate model performances on supercomputers (Ec-Earth, COSMO-CLM), developing a specific tool for performance measurement ("lucia").

As expected, the ODUS program served IS-ENES partners, other community laboratories and CERFACS, via bug reports and coupler enhancement suggestions based on practical experiments.

Within the limits of the previously described community means and common practices, it helped to facilitate the implementation and the use of coupled models in Europe, identifying or resolving practical issues during coding, or simply by discussion[13]. Focusing and isolating the work of both model and coupler specialists on a given time period, these one month long format of each ODUS gave enough time to precisely identify issues, sometimes with more than one modelling group. One month to complete a coupled model implementation from A to Z is clearly not enough[14]. But ODUS target was clearly to bring an help and not to deliver a ready-to-use coupled system.

Sometimes, it has been possible to collaborate with an host laboratory to develop tools that should benefit to the whole community. This is what happened with the load balancing tool "lucia", even though its lack of robustness and the work overload of the OASIS development team still defers its official release jointly with OASIS.

ODUS program gave us a clearer idea about present and future model community requirements. One major result is the survey of the different OASIS version limits toward model parallelism increase. It contributes to drive OASIS supporting thousands of cores configurations, which actually satisfy the needs of the community in Europe.

Mission	Date	OASIS version	Distribution (cores)	Supercomputer
LOCEAN-IPSL	10/2009	4	500	IBM P6
AWI	11/2009	4	n/a	IBM P6
SMHI	02/2010	3	50	HP Nehalem
SMHI	10/2010	3	1000	Cluster Opteron
ETHZ	11/2010	4	150	CRAY XT5
LOCEAN-IPSL	03/2011	4	150	IBM P6
ETHZ	07/2011	3	150	CRAY XT5
Bonn University	11/2011	3	n/a	n/a
SMHI	02/2012	3	1200	Cluster Opteron
BTU	08/2012	3-MCT	60	IBM P6
The Met Office	11/2012	3-MCT	2000	IBM P7
UBO	12/2012	3	140	IBM P6

Table 2: Observed coupled model infrastructure during ODUS program

[13] Only a few hours have been necessary to help Bonn University people setting up their own coupling interfaces.

[14] Its documentation has been problematic: more time was needed to report the different level of information to the laboratory user (implementation and user guide) and to IS-ENES community (coupling specificity, elements to share)

Made on real models, coupler tests reveal present needs. Identification of more model characteristics impacting their coupling[15] helped to set-up more realistic toy models, used for coupler validation.

ODUS program has increasingly contributed to disseminate OASIS best practice through laboratories. We hope that it will contribute too, within associated EU projects such as Embrace or IS-ENES2, to favour interactions between not only laboratory managers but also coupling implementers.

Laboratory	Year	Title
AWI	2009	OASIS3-OASIS4 coupling for FEOM
ETH/Meteo Swiss	2011	COSMO-CLM with OASIS
Bonn University	2011	COSMO-CLM with OASIS
DWD	2012	IS-ENES OASIS Dedicated Support
BTU	2012	COSMO-ECHAM with OASIS3-MCT
NCAR	2013	OASIS Dedicated Support

Table 3: Seminar list

It has been easily carried on until its end, despite its length and its cost[16]. An extension of this activity, in a different framework, has been submitted to CERFACS direction.

Conclusion

ODUS successes are the results of a conjunction of efforts:

- from host laboratories, with personal involvement of laboratory applicants, sharing computing and other infrastructures means, with sometimes effective participation to lodging and food (ETHZ, AWI) and, always, the warm and friendly atmosphere that my hosts knew how to create.
- from OASIS developers, that bring an additional and real-time support to this program. It is important to emphasize that no OASIS enhancement could be possible without their goodwill.
- from CERFACS authorities, in addition to the standard OASIS development they offer to the community
- from IS-ENES selection committee who, again voluntarily, tried to estimate the scientific potentiality of each application. We notice that the difficulties associated to this aspect were also probably underestimated. It lead for example to grant important laboratories, that certainly offered good conditions for an efficient ODUS collaboration, to the detriment to the smallest ones.

Despite of this goodwill or, rather, because of it, it seems obvious that this activity cannot continue by its own. Our results are weak: a coupler (OASIS4) development has been forsaken, some coupled models still need to be set up (ETH, BTU, UBO) by their users and a new support could be necessary for that.

Even though it is clear that a lot of contemporaneous European collaborations are including such long dedicated missions, it seemed not possible to rely on larger existing European

[15] Representative decompositions, use of restart, high resolution ...
[16] What it involved for the main contributor: journey length, activity interruptions at CERFACS ...

infrastructures[17], for things as different (but essential) as lodging or expertise networking[18].

Generally speaking, an activity such as ODUS is efficient when it supplements and enriches existing activities and not when it substitutes for a local missing manpower. However, we hope that we contributed to identify and strengthen the existing network of OASIS implementers.

Thanks to Kerstin Fieg (AWI), Marco Giorgetta, Monica Esch (MPG-M), Uwe Fladrich, Klaus Wyser, Colin Jones, Martin Evaldsson, Wang Shiyu, Ralf Döscher, Robinson Hordoir (SMHI), Chandan Basu, Torgny Faxén (NSC), Laurent Brodeau (Stockholm University), Edouard Davin, Sonia Seneviradne, Anne Roches (ETHZ), Oliver Fuhrer (MeteoSwiss), Andy Döbler (Francfort University), Matthieu Masbou, Prabakhar Shresta, Mauro Sulis (Bonn University), Andreas Will, Stefan Weiher, Eberhard Schaller (BTU), Markus Thuerkow, Ingo Kirchner (FUB), Jennifer Brausch (DWD), Jean-Guillaume Piccinalli (CSCS), Richard Hill, Omar Jamil, François-Xavier Bocquet, Mick Carter, Catherine Guiavarch, Chris Harris, Adrian Hines, Mike Hobson, Matthew Mizielinski, Steve Mullerworth, David Pearson, Jean-Christophe Rioual (the Met Office), Anne Marie Tréguier, Claude Talandier (UBO), Julie Deshayes, Eric Machu (IRD), Clément de Boyer Montaigut (Ifremer), Stéphane Sénési, Silvana Buarque (Météo-France), Marie-Alice Foujols (IPSL), Olivier Marti, Arnaud Caubel, Yann Meurdesoif (LSCE), Sébastien Masson, Guillaume Samson, Claire Lévy and Rachid Benshila (LOCEAN) for their support and their interest to our work. Thanks to the IS-ENES WP4 Selection Committee, among whom Reinhard Budich (MPG-M), Wilco Hazeleger (KNMI), Sylvie Joussaume (LSCE) and Enrico Scoccimarro (CMCC). Thanks to the OASIS development team, Sophie Valcke, Laure Coquart (CERFACS), Moritz Hanke (DKRZ), Rene Redler (MPI-M) and Anthony Craig.

Estimated carbon emissions for 12 continental journeys by terrestrial and collective means of transport: 1110 KgEqCO2

[17] Except the EU terrestrial and maritime transport infrastructure, but mostly composed of independent national networks. Allowing to prepare and document the different missions, the convoying between laboratories (with adequate connection times) has been always punctual and nice. It demonstrated (if needed) the continental aerial network pointlessness

[18] or even simple wireless connections in laboratories like Eduroam

Fig 1: Cloud cover record, November 2012

Appendix

The 12 OASIS User Support Reports

Mission #1
Sept 28- Oct 23 2009

Host: Sébastien Masson
Laboratory: IPSL-LOCEAN, Paris (France)

Main goal: set up the high resolution model on scalar machine

Main results

- Set up of a new high resolution model on an SMP machine, reaching OASIS3 memory limits (this OASIS3 limitation was lifted afterwards)
- Implementation of a new OASIS4 interface in NEMO

Main task 1: NEMO-ECHAM OASIS3 at high resolution on IBM Power6 supercomputer

After checking previous attempts to update ECHAM interface for OASIS4 (Luis Kornblueh, Rene Redler, Stephanie Legutke), we decided to start from official release of ECHAM version 5.4 (thanks to Monika Esch, MPI).

We integrated ECHAM on the IPSL LMDZ-NEMO compiling and running environment on F-IDRIS IBM Power6 (thanks to Marie-Alice Foujols, IPSL), exchanging LMDZ by ECHAM within the IPSL environment.

We began ECHAM interface modifications, adapting coupling fields to NEMO needs (update with 5.3 modifications already implemented in Japan on Earth Simulator). On the other hand, with the previously developed standard NEMO interface, there was nothing to do on oceanic side for ECHAM compliance.

To set up the first OASIS3 pseudo-parallel runs, we implemented a low resolution configuration (NEMO ORCA2 - ECHAM T106) in complement of the targeted resolution (NEMO ORCA1/4 - ECHAM T319).

To be able to create, from a mono-processor namcouple, several namcouples for OASIS pseudo parallel use, the Earth Simulator existing tool has been improved for maximum parallel coupling (not necessary any more to declare at least one coupled field from each source model).

We provided a modified version of OASIS3, including a clock count: with the corresponding shell script analysis tool, it is then possible to precisely measure load balancing between each component of the coupled configuration (ocean and atmosphere).

Even with maximum OASIS3 parallelization (one OASIS per coupling field), EXTRAP analysis memory requirements (for high resolution ORCA ¼ grid) **oversize F-IDRIS IBM Power6 limits** (memory limit: 3.2Gb per processor). However, thanks to further modification in OASIS3 options (allowing the use of the nearest non masked nearest neighbour for the target points having all original source neighbours masked), the EXTRAP functionality is not mandatory anymore.

Concerning performances, the fastest configuration took **2 hours 30 minutes** to complete a one month long run of ORCA1/4-T319 coupled model, with a 2 hours coupling step and using 512 processors (13 days to complete 10 years, using 160.000 CPU hours)

Warning/ Issue: an atmosphere model **extra cost (+25%)** is observed at each time step (not only at coupling time step) on coupled mode, compared to stand alone mode (same problem with ARPEGE on NEC SX9 and SGI Altix)

Main task 2: ECHAM-parallel OASIS3 interface implementation

To prepare ECHAM for a fully parallel OASIS4 interface, we changed from Box to Orange the domain decomposition as seen by OASIS3. The difficulty lays in ECHAM's special partitioning: there are two box domains per partition.

We also activated prism_put and prism_get routine calls at each time step, letting OASIS accumulate the coupling fields at chosen frequency.

Even though an extra cost was observed at low resolution (+10 %), this improvement has no effect at high resolution (- 1%).

Main task 3: NEMO OASIS4 interface

To test the chosen interpolations on ECHAM and NEMO grids, two OASIS4 toy models have been set up on the F-IDRIS IBM Power6. This facilitated the SCC and the SMIOC XML configuration files definition and tests of our running environment on the machine.

Once NEMO – OASIS4 interface updated, following the NEMO3 new coupling “per field” interface style, NEMO-ORCA2 tests have been processed coupled with a ECHAM-T106-like toy component.

But **2 weeks** with two OASIS3-4 “experts” (thanks to Laure Coquart) and a NEMO developer **were not enough to switch from OASIS3 to OASIS4**, even on NEMO low resolution configuration

Mission #2
Oct 26- Nov 20 2009

Host: Kerstin Fieg
Laboratory: Alfred Wegener Institute, Bremerhaven (Germany)

Main goal: ECHAM-FEOM coupling with OASIS4

Main results

- Enhancement of OASIS4 functionalities to allow ECHAM-NEMO coupled configuration setup at low resolution
- Similar performances of ECHAM-NEMO with OASIS3 and OASIS4 at low resolution

Task 1: Set up of the OASIS4 interface within ECHAM

OASIS4 interface has been defined and implemented within ECHAM, taking benefit of new OASIS3 interface implemented previously at IPSL (every process is now involved in coupling send/receive).

These developments have been tested with ECHAM-NEMO configuration. Grid declaration done through the OASIS4 communication library API was validated using a NEMO-like toy coupled to ECHAM (developed by Laure Coquart, CERFACS).

These tests revealed 2 issues within OASIS4:

- OASIS4 IO library is not supporting a grid sub-partition (necessary to process the particular ECHAM grid partitioning – 2 non contiguous domains on 1 process). This problem, which concerns restart read/write and interpolation global search, needs to be addressed by OASIS team.

- Using the interpolation weights, a bug has been identified (and recently fixed) linked to the NEMO multi-grid (mid-point discretization) and parallel decomposition characteristics.

A performance measurement was done with this low resolution (t106-orca2) configuration (including test of OASIS4 parallelization): we concluded that performances reached the OASIS3 performances level but only with a high number of processors for the OASIS4 transformer (10).

A high resolution (t319-orca¼) model has been set up to try to measure performances, because it is only at such level of parallelism that OASIS4 is supposed to be fully efficient.

This test revealed another issue: at such resolution, and using Netcdf-3 within the OASIS4 IO library, a memory limit per processor has been reached on IBM Power6. To address this issue, Netcdf-4 option, implemented in OASIS4 but not fully tested, has to be used. At the end of this user support period, another issue forbid us to do planned measurements. OASIS4 performances, within a high resolution coupled configuration (including models), have still to be measured.

An oral report of the work on ECHAM interface took place at Max Planck Institute with ECHAM

(Marco Giorgetta & Monika Esch), OASIS (Rene Redler & Moritz Hanke) and AWI developers (Kerstin Fieg).

A seminar was given at AWI (OASIS3-OASIS4 coupling for FEOM) with Kerstin Fieg.

Task 2: ECHAM-FEOM coupling

We transformed FEOM launching script to allow ECHAM coupling and we defined OASIS4 smioc/scc xml configuration files (former OASIS3 namcouple).

Within ECHAM model, we identified needed quantities for FEOM coupling and implemented a new ECHAM coupling interface for FEOM coupling fields exchange (cpl_feom CPP key).

A preliminary debugging simulation of FEOM-ECHAM coupled model has been launched on D-DKRZ IBM Power6.

Mission #3
Feb 1- Feb 19 2010

Host: Uwe Fladrich
Laboratory: SMHI, Nörrköping (Sweden)

Main goal: Increase EC-Earth performances using OASIS3

Main results
• Improvements on IFS-NEMO coupling exchanges strategy • Gain of 15% in EC-Earth model performances re-ordering field exchange and using OASIS3 pseudo-parallel version

Task 1: Coupling strategy

In a first step, the “model by model consumption measurement” functionality was integrated in the local OASIS version (now available with the last OASIS3 svn trunk version) for performance measurement and the associated shell-script tool for analysis was made available on the “gimle” NSC supercomputer (HP Proliant Intel Nehalem cluster). This tool helped us to determine respective speeds of the two components constituting the EC-Earth2 SMHI configuration (IFS and NEMO)

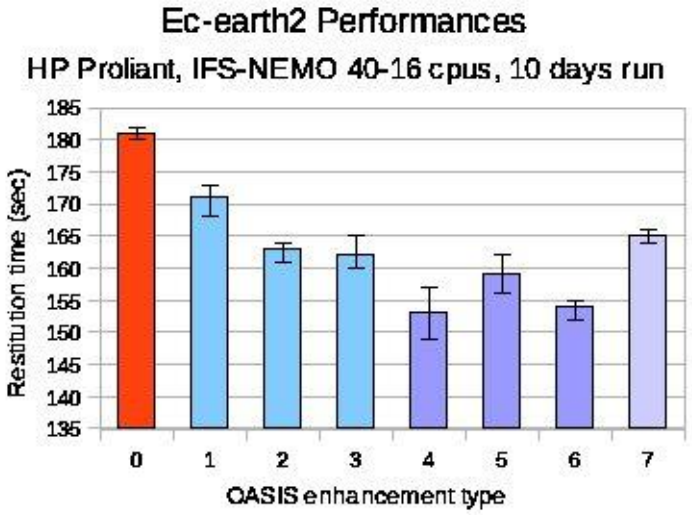
In a second step, we considered a simple sequencing of the coupling fields within the namcouple: firstly, ocean-to-atmosphere fields have to be sent by the (fast) ocean, interpolated by OASIS and then made available for the atmosphere, so that as soon as it is ready, the atmosphere receives these fields, sends its atmosphere-to-ocean fields, and goes on; OASIS then interpolates these atmosphere-to-ocean fields (while the atmosphere is running) and sends them to the ocean. This forced sequentiality allows OASIS communication and interpolation time to be spend in parallel to atmosphere computations.

A precise examination of prism_put/prism_get call strategy within both atmospheric (IFS) and oceanic (NEMO) EC-Earth components revealed that a deadlock appeared in case of such coupling field exchange imposed sequence. An inversion of prism_put/prism_get calls within NEMO model was coded to address this deadlock problem (prism_snd routine called first at the beginning of taumod routine instead of at the end of flxmod).

Further code analysis allowed us to suggest additional improvements in EC-Earth coupling sequence: prism_get/prism_put called at each time step within IFS and average operation done by OASIS, possible use of OASIS restart files …

Task 2: Performance improvements

The total elapsed time of a 10 days run was then measured for a series of different coupled configurations. For configurations 4 to 7 below, an upgrade of OASIS used in the coupled system to the new OASIS3 pseudo-parallel version (on a field-per-field basis) was first realized. Results are shown on the figure below.



0: initial state, 1: inverting get/put sequence within nemo, 2: same than 1 but seq=1 or 2 only within namcouple, 3: same than 2 but without NOBSEND, 4: same than 3 but using 8 oasis3 // and coupled fields splitted in the same order than on initial namcouple, 4 by 4 blocks, 5: same than 4 but coupled fields splitted mixing o2a and a2o, 6: same than 4 but with seq=1 within namcouple, 7: same than 4, 5 or 6 but using only 1 oasis3 //

0. Initial configuration
1. Coupling field order inversion as described in §1
2. Sequentiality (namcouple SEQ option) is reduced to 2, instead of 39 (number of coupling fields). For all ocean-to-atmosphere coupling fields (those which are able to slow down the slowest model, and the whole coupled configuration), the exchange between NEMO and OASIS are first performed , then all interpolations are done. OASIS is now able to communicate the coupling fields to IFS. When the atmosphere is ready to get these coupling fields, OASIS performs MPI sending. In the previous configuration (1), sequentiality was defined to 39. For ocean to atmosphere coupling fields, the exchange of the first field was performed, it was then interpolated and sent to IFS. When IFS was ready to receive all the ocean-to-atmosphere fields, OASIS was therefore only able to communicate the first coupling fields to IFS. So OASIS had then to receive the second field, interpolate it and send it to IFS, and so on for all the fields : that's the reason of (1) extra cost regarding to (2).
3. Without namcouple NOBSEND option (Buffered send used), but with a sequentiality of 2, no significant improvement measured.
4. With OASIS3 pseudo parallel option and 8 OASIS instances (performances are the same with 16 instances). Each OASIS is exchanging 4 or 5 coupled fields instead of 39. Each OASIS is exchanging ocean to atmosphere fields only or atmosphere to ocean fields only. Sequentiality is set to 4 or 5. 8 coupling fields are immediately available to IFS model instead of 1 in (2) configuration, which explains better performances in this configuration.
5. Slowest performances if namcouples mix ocean to atmosphere and atmosphere to ocean coupling fields (to be explained !)
6. No significant improvement measured if sequentiality reduced to 1 within (4) configuration.
7. Verification: OASIS3 version for pseudo parallel mode differs from OASIS3 used in mono processor mode. But OASIS3 pseudo parallel version used in mono processor mode (1

OASIS) exhibits quite same performances than (1) with the same experimental setup.

Other various actions:

- Discussions with OASIS users at Rossby Center about MPI buffered send and coherence with MPI model parallelism, and about future of OASIS4 for RCM coupling. Discussion with NEMO users and NSC people about Baltic sea configuration and possible improvements for load balancing. Report of EC-Earth performances to Stockholm University OASIS users.
- Tests of last OASIS3 release
- Adaptation of Rene Redler's script for namcouple splitting (OASIS3 // configuration)

Mission #4
Oct 18- Nov 12 2010

Host: Uwe Fladrich
Laboratory: SMHI, Norrköping (Sweden)

Main goal: Test performances of the OASIS3 based Ec-earth high resolution model

Main conclusion
The OASIS3 mono-process cost (~6 s per coupling step) is not negligible compared to Ec-earth3 T799-ORCA025 coupling step duration (e.g. 33.2 sec in the 800-256-1 configuration) leading to a coupling overhead of 11% on an Opteron-Infiniband cluster. But an optimal use of parallel version of OASIS3 on 10 processes could reduce this overhead down to 1.3%. Limits of OASIS3 are not yet hit but could be reached on this machine, if IFS model continues to scale for more than 3000 cores (as described by ECMWF), or on MPP architectures (CRAY XT, IBM BG) with equivalent number of resources, or on any other machine if the component models can be distributed so to reach a quasi perfect load-balancing.

Model / machine description

SMHI's coupled model (high resolution version) deals with:

- IFS, cycle 36: T799, 843.490 grid points, ˜25Km, 62 vertical levels, time step: 720s
- NEMO, v3.2: ORCA025, 1.472.282 grid points, ˜40Km, 45 vertical levels, time step: 1200s
- OASIS v3 (pseudo parallel)

20 coupling fields are exchanged between the two components at a coupling frequency of 3 hours. The model is available on Ekman supercomputer, 1.268 compute nodes of 2 quadripro AMD Opteron (# 10.144), Infiniband interconnection, located at Royal Institute of Technology (KTH), Stockholm, centre for parallel computers (PDC).

Evaluation of Oasis additional cost

At such high resolution, on a scalar machine, the limited parallelism of OASIS could become a bottleneck for the coupled simulation. We propose to determine the OASIS cost (communications and interpolation calculations) and its impact on the global performances of the model.

The best load balancing between ocean and atmosphere models has to be reached first. At this point, the coupling overhead can be measured as the difference between the elapse time of the slowest stand alone model and the elapse time of the whole coupled system[19].

[19] In order to measure those quantities, CERFACS' sh_balance tool (see OASIS Dedicated User Support 2009, Annual report) is launched on working directory (using *.prt files and cplout_0). On Ekman, it is installed on /afs/pdc.kth.se/home/e/emaison/Public/Projects/ecearth3/util/balance_oasis/sh_balance and can be launched on any working directory. This script computes clock time written by OASIS and model PSMILE libraries at each prism_get or prism_put call. This script is exploitable only for OASIS mono-processor mode, or on parallel mode with at least 1 coupling field on any coupling direction (ocean to

The resource ratio found for NEMO/IFS load balancing is about ¼ if IFS runs on 512 cores (i.e. NEMO should then run on 128 cores). Increasing the number of cores, NEMO seems less scalable than IFS, and ideal ratio reaches 1/3, i.e. NEMO should run on 256 cores when IFS runs on 800 cores.

Note that a a limit of 1546 cores (1280-256-10) was reached (results are not shown). This limitation is probably due to machine implementation of scaliMPI (SMHI and NSC are currently working on an OpenMPI version).

On the following table, all figures represent the elapse time for one coupling time step, mean of the 15 first coupling time steps (45 hours), excluding initialization (particularly MPI starting, which cost increases with partitioning) and finalization phases. To fix the issue of machine load dependent results, several realizations of the same experiment are processed.

IFS-NEMO-OASIS nb of cores	512-128-1	512-128-10	800-256-1	800-256-10
1-IFS standalone	41.	41.	29.9	29.9
2-EC-Earth3	45.7	42.3	33.2	30.3
2.1-IFS component	41.8	n/a	32.7	n/a
2.2-NEMO component	38.5	n/a	24,6	n/a
2.3-OASIS	5.5	n/a	6	n/a
Coupling overhead (2-1)	4.7(13.4%)	1.3 (3%)	3.3 (11%)	0.4 (1.3%)

Table 1: 2 hour long simulation response time (in seconds) for the different components and for EC-Earth3 coupled model. The coupling overhead is calculated as the difference between EC-Earth and IFS standalone elapse time.

In this configuration, IFS and NEMO run in parallel and not sequentially. We can observe here that OASIS elapse time is non negligible when it runs in mono-processor mode (respectively 5.5 seconds and 6 seconds for the 512-128-1 and the 800-256-1 configurations). In this case, the coupling induces significant overhead in elapse time with respect to the IFS standalone run (respectively 13.4% and 11%); this is true even if OASIS3 interpolates the fields when the fastest component waits for the slowest as OASIS3 cost itself is larger than the component imbalance.

But when the parallelism of OASIS3 increases (going from 1 to 10 processes, i.e. with 2 coupling fields per OASIS3 process), OASIS3 elapse time decreases and its cost can almost be “hidden” in the component imbalance. Even if we do not have direct measures of OASIS elapse time in these cases, this can deduced by EC-Earth3 elapse time which decreases from 45.7 to 42.3 seconds (512-128-1 -> 512-128-10 configurations) and from 33.2 to 30.3 seconds (800-256-1 -> 800-256-10 configurations). Therefore, it can be concluded that OASIS3 pseudo-parallelisation can be an efficient way to reduce the coupling overhead (which goes from 13.4% to 3% in the 512-128 configuration and from 11% to 1.3% in the 800-256 configuration).

Of course, this way of “hiding” the cost of OASIS3 works only if there is some imbalance of the components elapse time which allows OASIS3 to interpolate the fields when the fastest component waits for the slowest. If the components were perfectly load balanced, then OASIS3 cost, even if

atmosphere and atmosphere to ocean). NOBSEND option has to be disabled (buffered send needed).

lower when OASIS3 is used in the pseudo-parallel mode, would be directly added in the coupled model elapse time.

Surprisingly, a slow down was observed at each time step of IFS model (even when coupling is not performed) when 20 OASIS are used instead of 10, and performances dramatically decreased (results not shown). No explanation was found to the problem : the degradation is not linked to the mapping of coupler processes on the machine, neither to the size of attached MPI buffers (and same behaviour is observed with or without NOBSEND option).

In conclusion, the OASIS3 mono-process cost (~6 sec per coupling step) is not negligible compared to Ec-earth3 T799-ORCA025 coupling step duration (e.g. 33.2 sec in the 800-256-1 configuration), leading to a coupling overhead of 11%. An optimal use of parallel version of OASIS3 on 10 processes could_reduce this overhead down to 1.3%. In this case, it is still possible to perform coupling operations when the fastest component model waits for the slowest component model. But this strategy could become inapplicable :

- on the same architecture, if IFS model continues to scale for more than 3000 cores
- on thin node or MPP architectures (CRAY XT, IBM BG) with equivalent number of resources
- on any platforms, if the component models can be distributed so to reach a quasi perfect load-balancing.

NEMO outputs

A new IO library is available within NEMO (IOM). This library could be used by NEMO identically to what was previously done by IOIPSL. Or be activated within a separate executable (ioserver): NEMO communicates through MPI with this module, which operates the output asynchronously.

Launching a separate ioserver could enhanced performances on machines where massively parallel concurrent writing on disk is an issue. Moreover, output fields are already gathered on the global grid (if only one occurrence of ioserver is used).

Both solutions use new XML and namelist parameters files[20]. A new compilation has to be done, this time adding key_iomput to the NEMO CPP list. New libraries are created, linked to NEMO and an executable is available at the same place than the NEMO one[21].

NEMO with IOM

NEMO embedding IOM (each NEMO processes are involved in output, no ioserver) exhibits significative slow down compared to the initial NEMO model using IOIPSL.

[20] Files iodef.xml and xmlio_server.def. Iodef.xml allows a more flexible definition of output fields (possibly at different frequencies). xmlio_server.def parametrizes MPI buffer size and indicates if server has to be used with or without OASIS.

[21] To use IOM without a separate ioserver, simply set using_server namelist parameter to false, and launch your coupled system as usual. To be able to run a separate ioserver, set using_server to true, modify OASIS namcouple to declare the ioserver as an uncoupled part of the system and launch the new executable with usual MPMD command, considering it as a normal component of the coupled system.

	IOIPSL output	IOM output
NEMO coupled	38.5	44.0
Ec-earth3	47.5	48.9

Table 2: NEMO output library version effect on Ec-earth3 performances

Number of cores: NEMO #128, IFS#512, OASIS#1. The amount of daily data produced is slightly lower with IOM option, but it is compensated by additional monthly diagnostics.

To gather local fields split on several files into one single file global field, the “rebuild” tool has been installed on the supercomputer. Its cost is nearly the same than the total time needed to process the climate simulation, which could also become problematic on more parallel architectures.

NEMO and ioserver

It appears that the parallel version of OASIS3 was not fully compatible with the ioserver. Bug has been reported, fixed by Arnaud Caubel (IPSL) and added by Sophie Valcke (CERFACS) to the next OASIS3 official release.

Optimal ioserver MPI buffer size has to be found, to be able to perform a run of this configuration. Anyway, considering the additional slowing down observed, it finally seems not possible to use the external ioserver for Ec-earth on this machine. Those performance issues are well identified at IPSL, and mainly due to unwanted calls to former IOIPSL library. They recommend to test the next c++ version of the IOM library as soon as it will be available.

Toward an OASIS4 interface in IFS for Ec-earth3

An OASIS4 interface has been coded and used in IFS some years ago (GEMS project) by Johannes Flemming, with Kristian Mogensen (ECMWF). Some interesting features such as the use as coupling fields of documented and easily identifiable IFS arrays, a namelist-based switch to selected subsets of coupling fields (similar to NEMO interface), and an unique routine for both prism_get and prism_put calls, were tested on the current Ec-earth configuration.

As a first step, each OASIS4 call is replaced by an OASIS3 one. The current Ec-earth OASIS3 interface is switched off. Furthermore, the model driven accumulation is also switched off, both prism_put and prism_get routines are called at each time step but the prism_get routine is followed by the filling of the corresponding IFS array only at coupling time step. Modification list for each routine is available on annex 2.

A first run validated the possibility to use this new interface for NEMO coupling within Ec-earth system. In particular,

- the partitioning and the coupling field declarations
- the exchanges synchronization (prism_put/get at the appropriate time)
- the validity of input fields and of some output fields

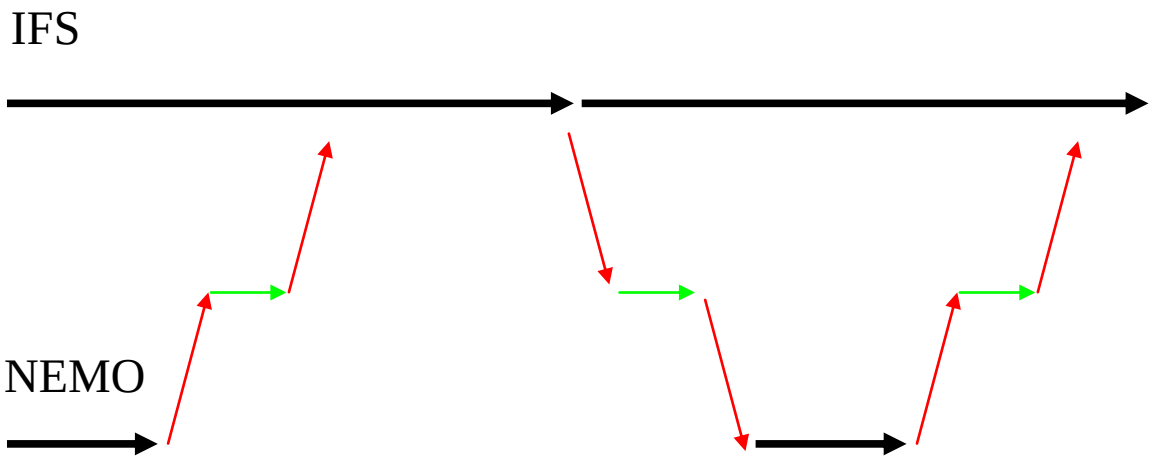
This interface is ready now to be tested with OASIS4 calls (OASIS4 is also available on NEMO) but some questions remain :

- How to find the information within IFS to fill some output coupled fields needed by OASIS ?
- Which IFS arrays can be filled with which input coupling fields ?
- Are the untested OASIS4 initialization routines able to fully describe the Ec-earth grid ?

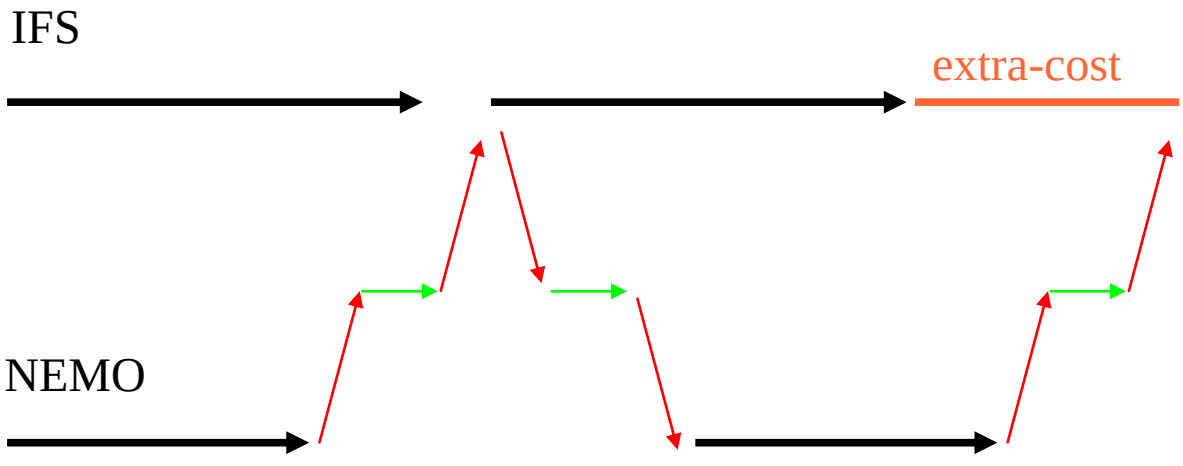
All those questions have to be addressed first to be able to know if OASIS4 is susceptible of driving efficiently an Ec-earth high resolution / highly parallel configuration on MPP supercomputers.

Annex 1

Coupling fields exchange synchronization in different cases, with OASIS monoprocessor configuration, sequential mode correctly defined (SEQ=1 for NEMO to IFS fields, SEQ 2 for IFS to NEMO fields)



Case 1: OASIS cost (red+green) smaller than IFS-NEMO running time difference



Case 2: OASIS cost (red+green) greater than IFS-NEMO running time difference

Annex 2

Modified routines

master.F90	Call to couplo4_inimpi
sumpini.F90	LCOUPLO4 set to=.true.
couplo4_mix.F90	Change IFS OASIS identifier
couplo4_inimpi.F90 couplo4_endmpi.F90	OASIS3 instead of OASIS4 calls
couplo4_definitions.F90	-Same than previous -Declaration of a new set of variables (Ec-earth namcouple compliant)
couplo4_exchange.F90	-Same than previous -prism_get/put at any time step no 3D coupling allowed

Mission #5
Nov 16- Dec 9 2010

Host: Edouard Davin
Laboratory: ETH, Zürich (Switzerland)

Main goal: Implement and validate an OASIS4 interface for a regional atmosphere-land model

Main conclusion
A regional atmosphere (COSMO-CLM, DWD) and a land scheme (CLM, NCAR) model have been coupled with OASIS4, at low resolution on a MPP scalar machine (on 100 cores), in order to simplify version updates, allow the use of different time stepping and different grids at different resolutions, and prepare other model components plugging. The first OASIS4 limitation observed is that OASIS4 does not currently support the CLM original partition; an exchange of the coupling fields through the master process only had to be implemented and will probably be a major bottleneck at higher resolution. Note that the mpp_io library, used for reading and writing restarts in OASIS4, does not support this type of partitioning either. Also, a strong slowing down is observed in OASIS4 exchanges when the number of cores used for the component models increases, even if these additional cores are not involved in the coupling exchanges. This problem, occurring on a particular configuration of grids and partitioning, has to be further investigated.

Model / machine description

COSMO-CLM (here called COSMO)

This regional atmosphere model (v4.8, and its climate version, v11) is used by a large community in several central Europe countries (in which ETHZ). DWD and several other meteorological agencies host the operational version of the model. Grid size: 109x121x32, 0.44 degrees. Parallelism reaches 100 MPI tasks on the targeted supercomputer.

CLM

This land scheme is developed at NCAR (v3.5). It is used within the integrated CCLM climate model. Initially, CLM is launched on the same grid that COSMO.

The model is available on CRAY XT5 supercomputer, with 22,128 compute cores (2 six-core AMD Opteron 2.4 GHz Istanbul processors per node), CRAY SeaStar 2.2 interconnect. Peak performance of 212 Teraflop/s. The machine is located at CSCS, Manno, Ticino, Switzerland.

Rationale

This user support task proposed to upgrade an existing coupled system with the latest version of the OASIS coupler.

The previous coupling (called integrated coupling) gathered two models on one executable. In this integrated version, CLM is called as a subroutine of COSMO at each time step (sequential coupling). CLM reads input file describing its grid. This input file is written once by COSMO and contains COSMO grid specifications: CLM grid points are COSMO land points.

CLM and COSMO run on the same processes (used sequentially by one model and the other) which means that CLM and COSMO partitioning differs (CLM processes only land points, COSMO both land and sea points).

Exchanges between models do not need interpolation (CLM and COSMO land points are located at the same place), just communications between processors (a grid point could be located on different processes during CLM or COSMO computation).

During the dedicated user support task, OASIS has been evaluated for its capacity to:

1. non intrusively be implemented on COSMO and CLM codes
2. let user choose the best time step for each model
3. launch models on a different number of processors (best number according to models distinct scalability)
4. investigate possibility of non sequential coupling

Implementation on models

To let user decide which coupling method he wants to use, we kept the possibility to choose at compilation stage (by CPP key) between stand alone, existing integrated coupling method (COSMO calling CLM as a subroutine) or OASIS4 coupling.

As previously implemented in several models (see NEMO interface, mission #1 and #6), a distinct OASIS4 interface has been written.

CLM interface

Due to a lack of appropriate option in present OASIS4 partitioning (see OASIS development paragraph), coupling fields have to be gathered on the master processor, which then communicate the coupling fields to COSMO. Consequences on coupling performances has to be evaluated but is obviously an issue for further massively parallel configuration setting.

The possibility to define distinct grids for regional land and atmosphere models implies necessarily a geographic mismatch between global domains: some grid points of the larger grid cannot receive information from the narrowest grid.

A strategy to mask the points of the largest sub-domain (CLM) falling outside the COSMO domain had to be designed. In the present User Support solution, on a first step, CLM global domain has to be larger than COSMO's one (CLM latitude and longitude limits have to include COSMO limits). The first received field has to be saved and the simulation stopped. A new mask is deduced from this interpolated coupling field and CLM is restarted with this new mask.

Interface routines, driving exchanges with OASIS, are really non intrusive. As usual in such kind of implementation, the off-line mode routines which read forced fields in external files are

switched off and replaced by our coupling fields receiving routines. Coupling fields sending is called as soon as coupling fields are available.

COSMO interface

The main originality of OASIS interface implementation lies in the possibility to involve a subset of model processes in the coupling, some of them providing no information to land model (all grid points are masked ocean grid points).

At definition stage, prior to any prism_def operation, the number of not masked points is calculated and OASIS initialization routines are called only if this number is non zero.

The subroutine call of CLM in COSMO was **easily changed for the OASIS interface**. Prism_put and prism_get routines (in this order) are called one immediately after the other. Gather/scatter operations needed in the previous integrated coupling (interpolation on the whole domain) are now disabled, and communication time is saved at this stage.

Results & performances

Due to the inability of the mpp_io embedded output library to support component processes not involved in the coupling, and thanks to OASIS4 developer Moritz Hanke (DKRZ), an alternate Netcdf based parallel output algorithm has been implemented in both interfaces. Received coupling fields are written (and overwritten) at each coupling time step.

Those fields could be:

- used to re-built the CLM adjusted mask (see above)
- used to check interpolations validity at implementation stage
- compared to arrays exchanged in the previous coupling at validation step.

Two examples of CLM and COSMO received coupled fields, produced after 17 days of simulation are shown on figure 1.

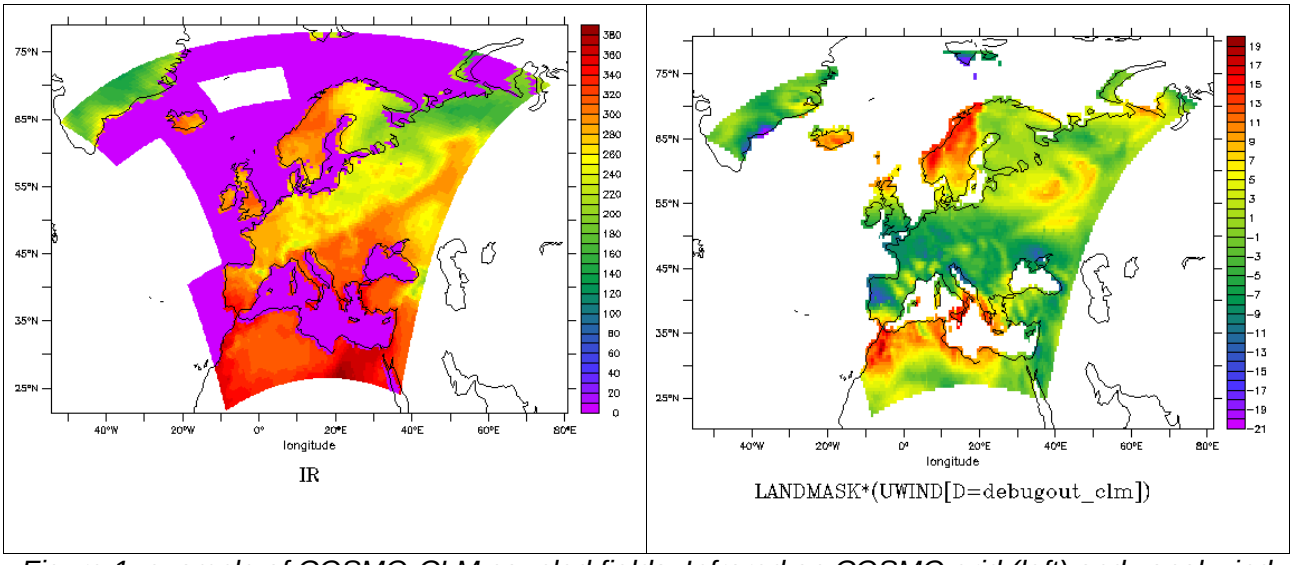


Figure 1: example of COSMO-CLM coupled fields: Infrared on COSMO grid (left) and zonal wind

on CLM grid (right) after 17 days

Geographic discretization of CLM and COSMO can be now totally independent. OASIS performs an interpolation (bicubic) between those two grids, giving the possibility for the COSMO-CLM user to define, if necessary, a finer resolution on one or the other model.

Also interesting for performance improvement: **each model parallelism can be set at its own optimal level**. Saved resources should compensate the fact that both COSMO and CLM models need their own processors [22].

Performances also take benefit of the possibility to set different time step for each model (and coupling time step different from model time steps).

The newly OASIS4 interface allows now, with some additional development, **plugging of other models** used by COSMO and COSMO-Climate communities. NEMO or ECHAM are possible candidates to complement the regional system model.

At the end of the user support period, Andy Dobler (Francfort University) starts with us an adaptation (following NEMO example) of the COSMO-OASIS4 interface for OASIS3, to couple COSMO to a Mediterranean NEMO configuration.

Concerning performances, figure 2 shows scalability of previous integrated coupled model (red curb) and newly implemented OASIS coupled configuration,

(a) using the same time step (240s) on both models and the same coupling frequency (cyan curb)

(b) decreasing down to 1 hour the land model timestep and the coupling frequencies (blue curb).

On this graphic, for OASIS based configurations, the resources number is the total number of cores used for both models and coupler. 12 cores (1 node) are devoted to OASIS, 12 cores to CLM (land model calculations are much less expensive than the atmosphere ones) and the number of cores for COSMO varies.

Compared to the previous integrated configuration, a strong slow down is observed due to an increase of the time spent in the coupling communications (about the same than the time needed for one time step calculations). Surprisingly, this slow down increases when CLM is spread on more processes, even if these processes are not involved in the coupling, see CLM interface section above).

Reducing the coupling time step to 1 hour, OASIS slow down is less visible and curb fits COSMO scalability. For a total number of processes greater than 75, response time becomes even better than with the previous integrated coupling approach. But it is important to realize here that this reduction in the response time is partially due to the fact that CLM is called less often, and changes on model results have to be evaluated to conclude if this configuration is or not equivalent to the existing one.

[22] On CSCS machine, or on every machine where number of processes on one core is limited to one, it is impossible to launch processes of the two executables on the same resources. If models are called sequentially, some resources are wasted while the processes of one executable waits while the other model performs its calculations.

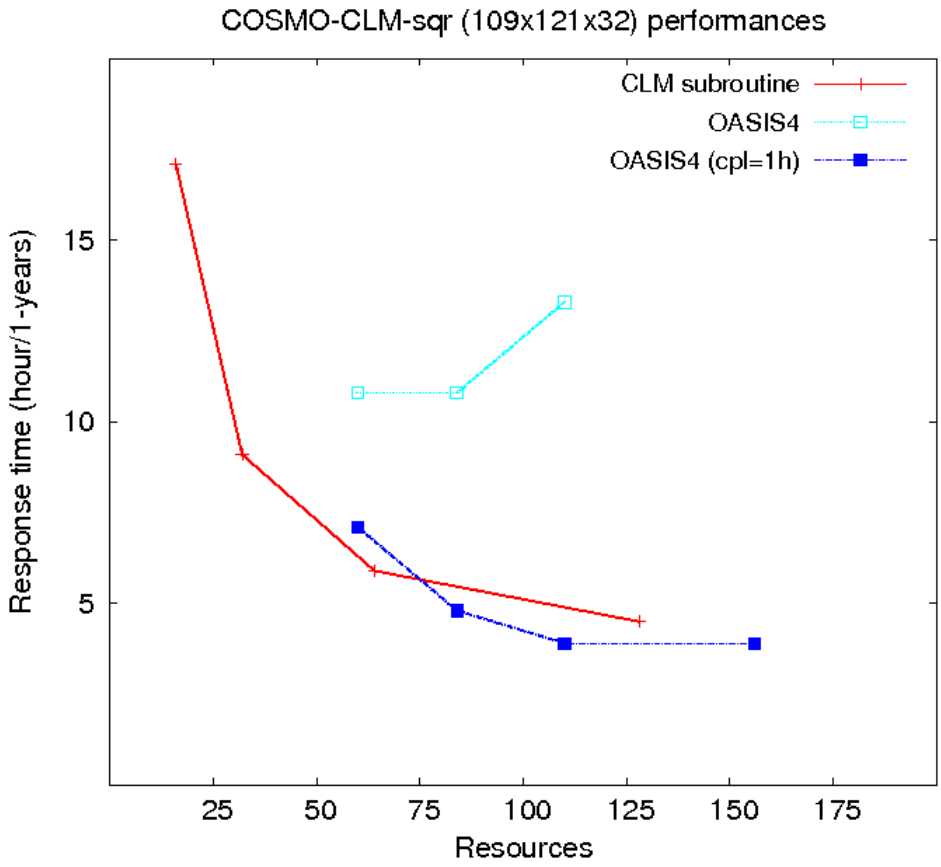


Figure 2: COSMO-CLM coupled model performances on CRAY XT5 for different coupling interfaces

OASIS4 developments

Exploring OASIS4 coupler capacity to fully satisfy the community needs is one of the important OASIS Dedicated User Support goal.

A strong day by day interaction with OASIS4 developers allowed us to identify, address or bypass related issues.

1D partitioning

A first analysis of CLM partitioning let us identify a restriction on OASIS use for some rectangular grids[23]. CLM calculations only occurs on land points and are independent: those points can be considered as a 1D vector, split onto processes without any geographic consideration (same kind of partitioning could be observed on models like LIM sea ice or SURFEX land scheme).

The temporary solution implemented (exchange of gathered field on master processor) can

23 For the moment, OASIS assumes that (i plus 1) and (i minus 1) points are geographically neighbours, which is not the case on our partitioning. A non trivial development is needed to address this issue.

be considered as a major bottleneck for further high resolution (and massively parallel) CLM configuration.

Masked partitions

In our particular partitioning, some sub-domains are not involved in the coupling because they don't intersect any unmasked points (COSMO) or also because they are not the master processor (CLM).

During coupling definition phase (halo detection), we identified some OASIS incapacity to consider separately MPI communications of process involved in the coupling and MPI waiting state in which non involved process lies. This problem has been addressed on-the-fly and solved by OASIS developers.

Mpp_io library

To validate coupling fields, OASIS debug outputs have been turn on but without any success, because the mpp_io library seemed to be unable to support a configuration into which some processes are not involved in the coupling and therefore do not write any part of the coupling fields in the output file.

This restriction also concerns coupling restart read/write operations and therefore prevented us to further test a parallel execution of the land and atmosphere models (instead of their sequential execution as reported here).

To visualize the coupling fields, a **workaround solution has been** suggested by OASIS team and **implemented**, each process calling in turn Netcdf within model interfaces (see above in Results & performances).

Scalability issues

As mentioned above, the most important problem identified during this support task concerns **abnormal coupling slow down** and its **increase with the number of model processes**, even when those processes are not involved in the coupling (see table 3).

CRAY XT5 (12 cores/node)				SGI Altix (8 cores/node)			
OASIS	CLM	COSMO	Comm(s)	OASIS	CLM	COSMO	Comm(s)
12	12	36	0.20	8	8	32	0.01
12	12	60	0.30	8	8	64	0.02
12	12	96	0.37	8	8	128	0.05
12	12	132	0.43				
12	12	60	0.30	8	8	64	0.02
12	36	60	0.33	8	64	64	0.53
12	60	60	0.56	8	128	64	1.80

Table 3: mean duration of a total OASIS coupling sequence

Further investigations are needed to identify which service can be at the origin of the slowing down. A toy model is implemented to try to reproduce the problem, without success. This could suggest that an interaction between model calculations and MPI communication could occur in this configuration (memory ? MPI buffers ?)

The coupled model has been ported on an SGI Altix platform and the same performance tests processed (4 right columns of the table 3). Even if calculations are achieved during the same time with about the same number of processors, the time spent within OASIS calculations and communications is slower.

But again, if we increase the number of CLM processors (not involved in the coupling), this duration amazingly increases. This identical behaviour on both machines suggests that the particular CSCS MPI installation on the CRAY machine (or some default MPI characteristics) is not at the origin of the noticed slowing down[24].

Needed improvements

On top of those OASIS improvements, some additional tasks have to be tackled to be able to guarantee an functional and efficient coupling between COSMO and CLM models.

Needed on every coupled system exchanging fluxes, conservative interpolation has to be tested.

Then, coupled fields have to be compared with those of integrated coupling. Characteristics of those fields have to be checked when coupling field frequency decreases. A special care should be taken on the behaviour of exchange coefficient, recomputed in the atmosphere when fluxes are changed by land model. Models characteristics such as diurnal cycle or long term means should be compared.

Finally, we hope that scalability tests with higher definition (operational) models could begin, to possibly investigate limitations of OASIS4 parallelism with such configuration.

24 Various code instrumentation (TAU & Scalasca) have been tested to try to identify the bottleneck. TAU profiling only reveals MPI_Wait or MPI_Barrier excessive durations. Scalasca full tracing was not possible to produce on our MPI MPMD configuration (the software new version could address the problem, contacts are ongoing with Jean Guillaume Piccinalli from CSCS).

Mission #6
Feb 28- Mar 25 2011

Hosts: Guillaume Samson & Sébastien Masson
Laboratory: LOCEAN, Paris (France)

Main goal: Implement and validate an OASIS4 interface for a regional atmosphere-ocean model

Main conclusion
Even though model interfaces have been successfully adapted from OASIS3 to OASIS4 in the WRF and NEMO component models, too many issues remains to attribute a scientific validity to the coupling: impossibility to use OASIS reading and writing mechanism for coupling restart files, issue with masked partitions, non completion of the run with different interpolations above a certain number of cores. In fact, even if OASIS4 computational cost appears significantly low (less than 1% of the total duration), failures on parallel interpolation weight calculation forbid to fully validate coupler scalability at serious level of parallelism for the different interpolations (e.g. more than 128 resources for the nearest neighbour interpolation).

Model / machine description

WRF
The NCAR regional atmosphere model (v3.2.1) is used by a large community in several countries. This easy to use model becomes more and more popular on several European laboratories. Grid size: 469x256x28. Parallelism reaches 128 MPI tasks on the targeted supercomputer.

NEMO
The well known European ocean model is embedded on most of the continental CMIP5 coupled systems. We used a regional configuration (Indian ocean) developed at LOCEAN with the 3.3 version. Grid size: 463x273x46. No parallelism needed up to 24 MPI tasks.

The model is available on IBM Power6 supercomputer, with 3,584 compute cores (16 dual-core IBM P6 4.7 GHz processors per node), Infiniband x4 DDR interconnect. Peak performance of 67.3 Teraflop/s. The machine belongs to IDRIS CNRS supercomputing centre, Orsay, France.

Rationale

This user support task proposes to adapt interfaces of an existing coupled system (based on OASIS3) to be able to use the new version of the OASIS coupler.

This OASIS4 coupling will be evaluated for its capacity to:
1. be operational without any important interface modification
2. provide a very simple interpolation (WRF and NEMO spatial discretizations are very close so a nearest neighbour interpolation is satisfactory)

3. address coupler scalability issues with highly parallel configuration (high number of process-to-process communications during coupling field message passing)

OASIS interfaces

NEMO interface:

This work starts from developments produced during #1 OASIS User support mission completed last year on NEMO model, based on ORCA global configuration. It includes corrections added during high resolution ARPEGE-NEMO coupling tests (IS-ENES WP8-JRA2).

Slight improvements have been made on existing interface:

1. possibility to switch off a process when all its grid points are masked (“masked partition”). 2 options:
   - the model is launched only with sub-domains containing at least some non masked points (optional on NEMO only)
   - the model is launched on all sub-domains but coupling is not effective on masked partitions (only prism_init, prism_init_comp, prism_enddef and prism_terminate OASIS primitives are called by the corresponding processes).

2. pseudo-parallel coupling field writing (thanks to Moritz Hanke, DKRZ). To compensate for OASIS IO deficiency, coupling field is written immediately after receiving. Writing is pseudo parallel, which means that every process writes its sub-domain in turn (and not simultaneously).

Same features has been implemented on WRF interface.

WRF interface:

Following NEMO implementation, a module_cpl_oasis4.F has been written, to be able to call the same wrapping procedures with OASIS3 (module_cpl_oasis3.F):

cpl_prism_init: initialization of coupled mode communication
cpl_prism_define: definition of grid and fields
cpl_prism_snd: snd out fields in coupled mode
cpl_prism_rcv: receive fields in coupled mode
cpl_prism_finalize: finalize the coupled mode communication
cpl_prism_update_time: update date sent to OASIS

This last routine only has to be called when OASIS4 coupling is active. The others are called in both configuration (OASIS3 or OASIS4 coupling).

On both WRF and NEMO interfaces, basic timing measure (using MPI_Wtime) has been implemented:

1. after last coupling field receive and before first coupling field receive.
2. before last coupling field send on (slowest model) WRF interface and after last receive on (fastest model) NEMO interface.

Two shell scripts collect information on standard output files and provide:

1. total time spent by model for calculation. This information is needed to be able to balance processor allocation.
2. time spent for WRF to NEMO coupling fields communications and interpolation (6 coupling fields).

OASIS improvements

Restart

OASIS coupling restart read/write is based on mpp_io library (same library used for coupling restart read/write in OASIS3).

Issues occurs frequently on mpp_io with any kind of non regular grid or partitioning (see report on mission #2). This time, a deadlock appears on simple NF_GET_VARA_DOUBLE function. A switch from standard netcdf calls to p-netcdf (parallel netcdf) was not successful (error on reading arrays shape declaration).

Moreover, we found two error sources that could easily mislead inattentive programmers and cause large waste of time:

1. A bad declaration of time bounds (not verified by OASIS) could leads to a mismatch between restart file netcdf read (that should occur at time step 0) and effective information exchange (that is delayed at coupling time step number 2)
2. LAG declaration must not be done in second (as OASIS3 required it) or, in that case, a deadlock (for a quite difficult reason to identify) will occur at 2nd coupling time step. Here, a specific error would be appreciated.

Simplification of restart read/write strategy is urgently required. Mpp_io library substitution should be a 2011 priority for OASIS developers.

Masked partitions

On a broad range of model types (surfaces, ocean, etc.), calculations take place only on a subset of the grid points. When parallelism increases, some partition could be made up of masked points only.

For a sub-domain composed entirely with masked points, it was observed that calling the ordinary API routines (prism_def_grid, prism_set_corners, prism_set_mask, prism_def_partition, prism_set_points, prism_def_var) lead to a deadlock.

We therefore decided to switch off coupling on these masked partitions (see NEMO interface paragraph).

For example, on figure 3 (left), NEMO parallelism along X (longitude) is 6, parallelism along Y (latitude) is 4, but total number of allocated processors is 23 (no calculation over Australia region).

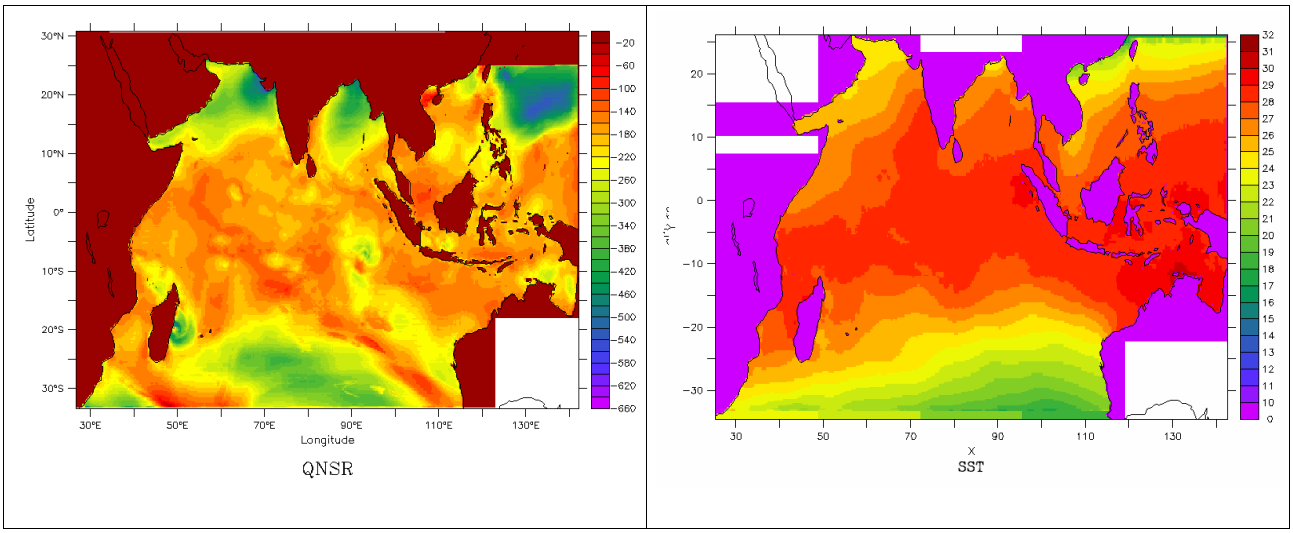


Figure 3: example of NEMO-WRF-OASIS4 coupled fields:
non solar on NEMO grid (left), 1 uncoupled partitions (white box)
and SST on WRF grid (right), 11 uncoupled partitions

Implementation of our coupling interfaces must take into account those particularities (see “OASIS interfaces” paragraph). One of the difficulties is the necessity to synchronize involved and masked processors.

Unfortunately, possibility to switch off partition have side effect on parallel interpolation functionality : target/source point matching research could not be done on empty partitions.

Consequence could be the non definition of some values (particularly near coastline and domain boundaries). It was observed with the “nearest neighbour” (NN) interpolation. Use of bilinear or bicubic interpolation leads to deadlock, when parallelism increases (up to #128). NN is preferred to those two other interpolations.

Moreover, WRF and NEMO discretization are the same, except on northern hemisphere were NEMO latitude circles begin to fold due to pole duplication: in this case, NN is the most appropriate interpolation.

To bypass problem of non definition near coastline and domain boundary, we decided to choose first strategy of masked partition switching off: model is launched on all sub-domains but coupling is not effective if all grid point of the domain are masked (see NEMO interface paragraph).

Nearest neighbour interpolation

The interpolation we prefer to use at this stage suffers from the impossibility to find neighbours on source grid if target grid point centre position does not belong to a non masked source grid point area. This occurs near masks, masked partitions or whole domain boundaries.

To compensate for this default, a simplified and domain restricted nearest neighbour algorithm has been implemented on model interfaces.

Results/performances

Restitution time comparison between coupled model and slowest model in forced mode (WRF) do

not show any extra cost.

Resources #	NEMO-WRF-OASIS4	WRF stand alone
18-6-1	1220	1220
121-6-1	330	330

Table 4: restitution time (for 1 simulated day) on coupled and forced mode (s). On forced mode, WRF has the same resources # than in coupled mode.

Evaluated OASIS cost remains significantly slower (less than 1%) than computational time. This cost increases with model parallelism but could be reduced with coupler parallelism.

WRF-NEMO resources #	1 OASIS4	6 OASIS4
18-6	0.025 / 50.9 (0.05%)	
52-6	0.107 / 25.0 (0.42%)	0.027 / 25.0 (0.10%)
110-12	0.119 / 15.4 (0.77%)	0.065 / 15.4 (0.42%)

Table 5: estimated OASIS time / model restitution time (ratio %). In seconds. OASIS time represents communication and interpolation time needed for 6 coupling field exchanges.

Unfortunately, unidentified issues occurring during neighbour identification phase (prism_enddef routine, parallel interpolation weight calculation) at higher level than #128 prohibit more serious test on OASIS4 scalability.

Mission #7
Jun 20- Jul 21 2011

Host: Edouard Davin
Laboratory: ETH, Zürich (Switzerland)

Main goal: Optimize OASIS interfaces on regional atmosphere and land models

Main conclusion
To overcome a performance default of our previous OASIS4 interfaces, the COSMO model has been coupled again with CLM (CCSM), but using OASIS3. Several optimizations made it as fast as the initial integrated COSMO-TERRA model. In complement, taking advantage of both OASIS and CCSM (CESM) modularity, our coupled model easily integrated a version upgrade of CLM (v3.5 to v4), preparing the way for other possible CESM/OASIS couplings.

Model / machine description

COSMO-CLM (here called COSMO)
This regional atmosphere model (COSMO v4.8, and its climate version, COSMO-CLM v11) is used by a large community in several central Europe countries (from which ETHZ). DWD, MeteoSwiss and several other meteorological agencies host the operational version of the model. Grid size: 109x121x32, 0.44 degrees. Parallelism reaches 100 MPI tasks on the targeted supercomputer.

CLM
This land model is developed at NCAR (v4). It is used within the integrated CESM climate model. Initially, CLM3.5 was coupled as a stand alone model through OASIS4 with COSMO (see Dedicated User Support #5).

Those models are available on CRAY XT5 supercomputer, with 22,128 compute cores (2 six-core AMD Opteron 2.4 GHz Istanbul processors per node), CRAY SeaStar 2.2 interconnect. Peak performance of 212 Teraflop/s. The machine is located at CSCS, Manno, Ticino, Switzerland.

OASIS3 interface for CLM3.5

Initial issue

The previously developed OASIS4 interfaces on CLM3.5 (part of CCSM climate model) land model reveal a lack of scalability at relative low level of parallelism. As shown on figure 1, the CLM model (as part of the coupled system) response time increases dramatically when parallelism reaches 60 PE (light blue curb). A code tracing revealed that time was mainly spent on the OASIS4 receiving routine (prism_get). Unfortunately, this default could not be reproduced with toys.

Dedicated User Support duration is limited to a few weeks: to switch from OASIS4 to OASIS3 is the quicker solution we found to overcome this scalability issue. It took a few days to adapt

interfaces and bypass the issue (orange curb). Scalability of our interfaces is now limited by COSMO-OASIS3 communications cost only (red curb).

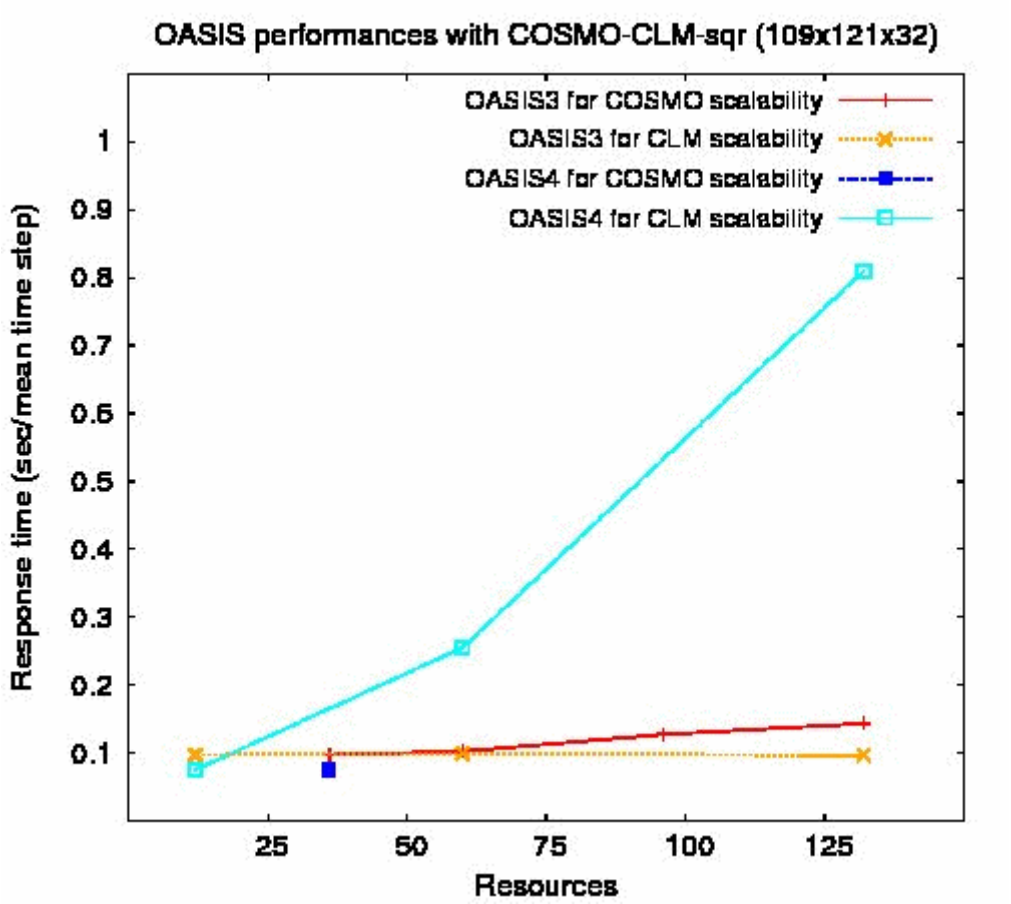


Figure 1: OASIS3 and OASIS4 interfaces performance

Implementation

Consequently, our new OASIS3 interfaces on both CLM and COSMO models are derived from the previous implementation designed for OASIS4 (see Dedicated User Support report #5). One of the PRISM/OASIS4 specifications was the compatibility of PSMILE interface library (routines called by models) with the OASIS3 one: it explains why it was quite easy to adapt the previously implemented interface to OASIS3 specificity.

In addition, we took benefit of an existing OASIS3 interface on COSMO, developed by Andy Dobler (Francfort University) to couple this model with NEMO.

Our final implementation on COSMO differs from U. Francfort's one:

- coupling fields are different
- we add the possibility to produce auxiliary files (masks, grids, areas) during the definition phase (this characteristic is inherited from our previous OASIS4 interface)

Nevertheless, the similarity of both implementations should encourage the COSMO community to merge them into an unified interface, extending the coupling system to a land/ocean-atmosphere configuration. In this way, a clean package including our interface and the associated input files has been communicated to COSMO-CLM administrator. It should contribute to help COSMO community to build a modular coupled system based on OASIS standard (possibly with IS-ENES help during a new Dedicated User Support).

As the great majority of present supercomputers, CSCS CRAY XT5 requires that a minimum number of PE (12, one node) was allocated for each executable of our coupled system. Consequently, to use OASIS3 on its “pseudo”-parallel mode was mandatory, but has some side effect on the interface implementation[25].

Optimization

Measurement tool

To be able to measure the impact of the following optimizations, the previously developed OASIS option (CPP key “balance”), using MPI_Wtime routine, has been activated (see OASIS Dedicated User Support #4). It delivers informations such as relative duration of each module of the system and OASIS communications + calculations time. But CSCS machine characteristics forbid a simple use of this measurement tool:

- each node has different clock times
- measures writing (Fortran WRITE on standard output) at each time step significantly slows down the simulation execution

The first problem has been addressed measuring the clock differences at the beginning of the run but, again, calling an MPI_Barrier on both OASIS and model routines.

The second one makes necessary a complete re-rewriting of our measurement tool: the different informations measured must be synthesized and written at the end of the run (the mean/min/max values), and not at each time step. Due to a lack of time, this development was postponed: the French ANR project “PULSATION”[26] is supposed to address this issue (2012). A first implementation has been designed and is described on the last Dedicated User Support mission report of this document. For the moment, ratios between the measured quantities are supposed to be the same with or without measurement tool enabling. Absolute values are deduced from one of those quantities (from the total run duration, for example).

Coupling fields number reduction

Five coupling fields (see annex 2) are exchanged from CLM to COSMO, which is less than the number of available OASIS driver PEs. It means that all those coupling fields are processed at the same time by one OASIS instance: the parallelism is almost ideal.

On the way back (COSMO to CLM), the initial coupling fields number was 13. It means that 1 OASIS instance has 2 fields to process, which slowed down the whole coupling sequence. A brief analysis of how coupling fields were used by CLM showed that convective rain and snow, as well as grid scale rain, snow and mist, could be merged into two coupling fields only (total convective and total grid scale precipitations). Gain on performances (on OASIS total time) is about 20%.

---

25 The auxiliary file writing routines, launched during the definition phase of the interface, was not compatible with the pseudo-parallel mode (neither with OASIS performances measurement pre-compiling option). Some code modifications within OASIS were necessary to bypass the issue. On the code interface, MPI_Barrier (on the MPI_COMM_WORLD communicator, which manages the coupled exchanges) has been called. This implementation is temporary, and has to be redefined for an official release.

26 http://www.locean-ipsl.upmc.fr/~pulsation

Raw performances

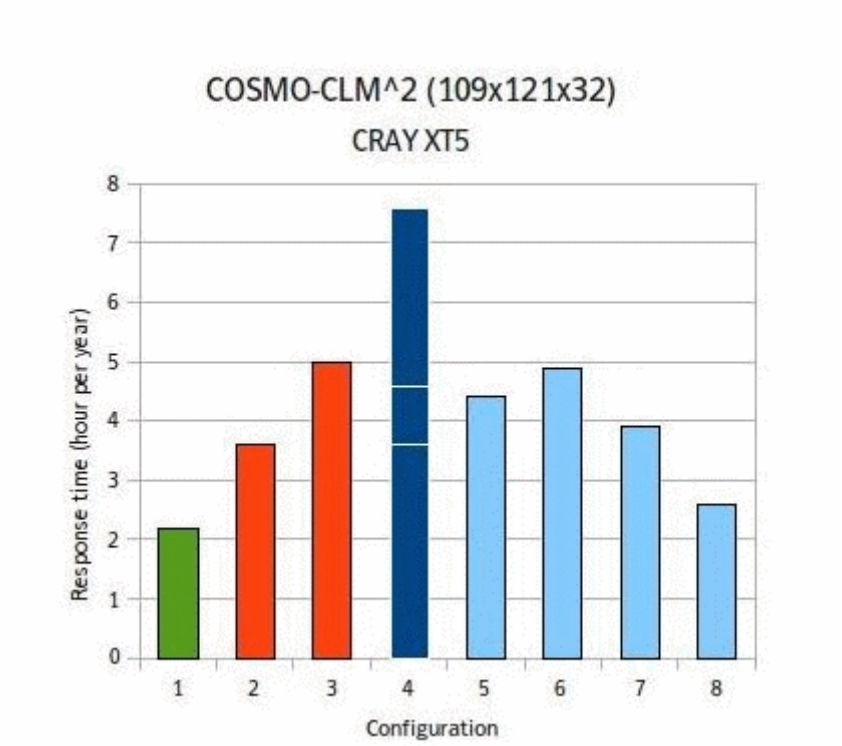


Figure 2: Performances of various COSMO-CLM couplings (OASIS and non OASIS)

This figure shows the compared performances (elapsed time) between:

- the very first COSMO configuration, without CLM, using the native land model TERRA, called as a subroutine on the same grid than COSMO (green box, n° 1 on 132 cores).
- the COSMO/CLM existing implementation (developed by ETHZ), where the transformed CLM model is called as a subroutine by COSMO (orange boxes, n°2 on 132 PE and n°3 on 60 cores)
- a serie of COSMO/CLM OASIS coupled configurations, tested during the present Dedicated User Support period. The dark blue box n°4 shows the performances of the very first configuration, using 12 PE for OASIS, 60 for COSMO and 60 for CLM, for a total of 132 cores. White lines represent, from bottom to top, respective COSMO, CLM and OASIS contributions to the total time.

Configurations 1, 2 and 4 are using 132 cores but, on the OASIS-based configuration 4, COSMO and CLM calculations are done on 60 cores only. The extra cost due to OASIS coupling is then deduced from the comparison with configuration n°3, where COSMO and CLM models also compute on 60 cores: this extra cost reaches 50%.

However, on a CPU consumption point of view, it is more suitable to compare the performances of the OASIS-based / OASIS-less configurations on the same number of total resources (configuration 1 and 2): then, the OASIS configuration is more than 2 times slower than the previous COSMO/CLM coupling (configuration 2) and more than 3 times slower than the initial COSMO/TERRA run (configuration 1).

Optimizations are definitely necessary to reduce this important extra cost.

Coupling frequency

For practical reasons, the CLM time step model was initially the same than the atmosphere time step (240s). This similarity could have a scientific justification but an increase of the land model time step could also be tested: CLM time step (and COSMO-CLM coupling) were set to 1h[27].

Designing our interface, we chose to call, at each time step, coupling fields sending and receiving routines. This permits to change easily[28] the value of the coupling frequency.

Blue box n°5 of figure 2 shows a significant improvement: most of the time is now spent on COSMO, essentially because CLM and OASIS are called 15 times less often.

Obviously, this modification must have an impact on model results. Those consequences would have to be checked by ETHZ users, if this configuration is chosen.

Coupling sequence

Another heritage of the previous COSMO/CLM “by subroutine” coupled configuration is the sequentiality of calls (model calculations are done one after the other).

Again, it is particularly simple to change the OASIS coupled sequence and do both model calculations in parallel[29]. COSMO and CLM could process calculations of a given time step at the same time.

The corresponding performances are shown on blue box n°6 of figure 2. It represents the total duration of the slower model, increased by a fraction of the time necessary for coupling.

As for coupling frequency, model behaviour modifications, induced by the new coupling strategy, has to be further investigated.

The last two optimizations can be jointly set and performances enhanced again (blue box n°7 on figure 2).

Compared to the previous “by subroutine” coupling, the OASIS multi executable approach let us choose the best parallelism for each model (according to their own scalability). Launching COSMO on 132 cores, CLM on 60 (and OASIS still on 12), we reach the most efficient configuration at this resolution (blue box n°8 on figure 2). The total duration is now comparable to the initial COSMO stand alone configuration.

---

[27] When a model is called as a subroutine of the other (coupling previously developed on configuration 2), both models should have the same time step (or buffers have to be implemented to accumulate coupling fields). With OASIS, model time steps are independent and coupling frequency can be changed with a simple directive on parameter file

[28] Just modifying “namcouple” OASIS parameter file (second section, coupling period per field and lag index). See OASIS3 user guide. If OASIS “prism_put” sending routine is called at each time step and LOCTRANS-AVERAGE option is activated, OASIS ensures the necessary accumulations of coupled quantities

[29] Both models are now using coupling fields calculated by the other model at the previous coupling time step. At the first time step, CLM has now to read the initial coupling fields on a file. This file could be created by COSMO on an previous independent run, activating an optimization option on the OASIS interface (oas_cos_vardef.F90 file). This operation has to be done once: at the end of each run, OASIS creates a restart file with coupling fields of the last coupling time step. This is this file that has to be used at the beginning of the next run.

Current limitations

Limits of our Dedicated User Support exercise forbid the tuning of all the possible parameters of our implementation.

1. An explicit process mapping is possible on CSCS Cray XT5 machine[30]. Given that a sensible spread has been observed in our performance measures, it is possible that a mapping which would take into account communication density between PEs and their position on the machine would change those performances.

2. Theoretically, coupling frequency could be different for each coupling field but some light modification will be necessary on the code to ensure it.

3. OASIS proposes a large variety of interpolations. The conservative one has to be chosen for some quantities (fluxes). Others could be tested to enhance performances.

But the main limitation affects perspective on resolution increase, particularly for meteorological applications (MeteoSwiss is one possible user of the COSMO-CLM OASIS configuration), given that OASIS3 already exhibits lack of performances on some previously developed configurations (see for example Dedicated User Support report #4 on EC-EARTH high resolution CGCM).

OASIS3 interface for CLM4

Rationale

Coupling modularity is one the most appreciable feature of OASIS. Once an interface is written on a model, due to the implementation non intrusiveness, it is relatively easy to maintain it on the successive versions of the model. In addition, if one model has to be upgraded, nothing has to be done on the other side to keep using the coupled system.

Version 4 of CLM is available through CESM integrated system. The land model stand alone configuration is no longer available, and the whole system (land model + coupler + driver + atmospheric variable forcing module) has now to be coupled with OASIS.

Strategy

Popularity of the OASIS framework mostly relies on its capacity to make the use of an external model as simple as the reading of a forcing dataset.

That is exactly the philosophy of this new CLM-COSMO coupling.

Build from a CLM stand alone configuration case (I_TEST_2003), our CESM coupled model mainly consists on the driver, the prognostic land model and a “data models” (DATM for atmosphere). The main function of data models is to read forcing files. Modules are linked to the driver using the

[30] Each process of the coupled configuration could be assigned to one particular core, among nodes reserved through SLURM batch scheduler. Notice that, if the machine has been initially configured for such purpose (on SGI Altix “jade” CINES machine, for example), a multi-threading (using more than one process on one core) could significantly reduce the amount of necessary resources without changing performances: actually, two sequentially coupled models can share the same resources, because calculations are processed one after the other.

CPL7 internal coupler, which ensures remapping or interpolations, if necessary.

The only CESM code modifications necessary for an OASIS coupling consists in:

- defining DATM module grid on the original CLM grid, through a forcing file which holds variables not provided by COSMO atmosphere (aerosols)
- organizing OASIS exchanges through this DATM module

The CESM code, coupled with OASIS, still consists on its original components. Code modifications (communications with OASIS) mostly take place on DATM module. As shown on figure 3, the red arrows, which represent the OASIS connections, only connect the atmosphere data model (DATM) rectangle.

Implementation

An exhaustive description of our Fortran interface implementation and input files modification/addition is given in annex 1. This paragraph only summarizes the principle of the CESM modifications needed to build the OASIS interface.

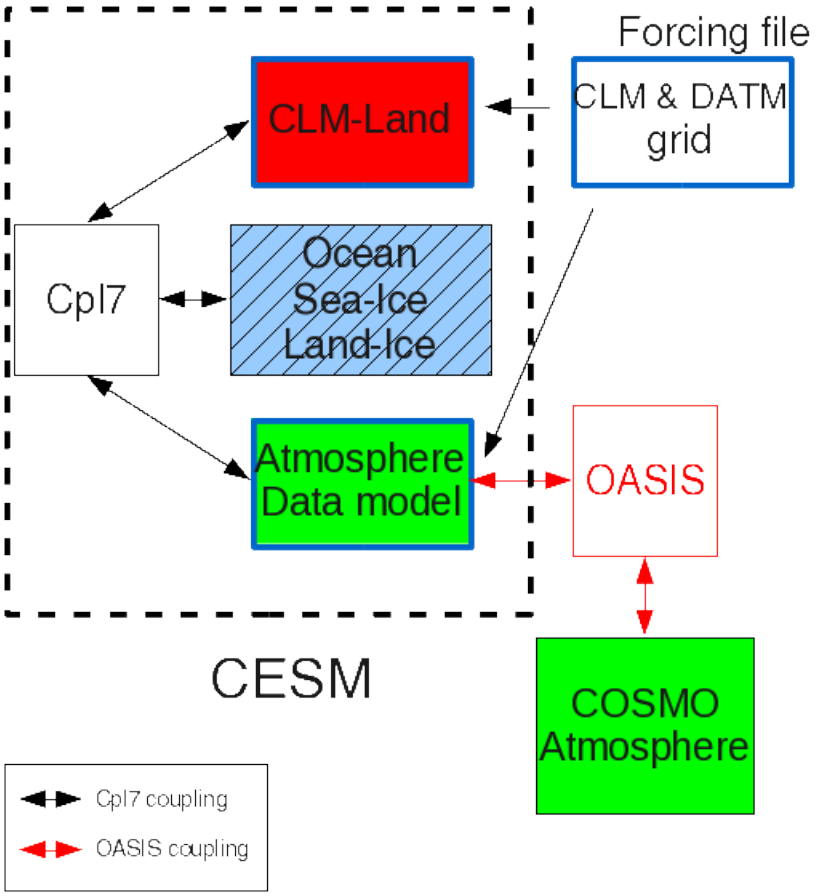


Figure 3: Description of CLM (CESM) / COSMO coupling using OASIS

We choose to start from a CLM stand alone CESM configuration (I_TEST_2003).As a first step of the OASIS interface implementation, we modify the file which contains the atmospheric forcing fields. We interpolate the input file variables, describing them on the CLM original grid. Aerosols (not given by COSMO) are the only variables actually used by the model: the other variables will be overwritten by the OASIS coupling fields.

In this way, DATM module has now the same spatial discretization than CLM land model (it defines its own grid according to the dimensions read on the forcing files). Then, the CPL7 functions will be limited to remapping (no interpolation between land and data atmosphere models) and only if decompositions of both components (CLM and DATM) are different.

OASIS interpolations (with atmospheric grid) are defined for the CLM discretization:

- On an initialization phase, grid mask and coordinates are communicated to OASIS, at the same time than names of exchanged coupling fields.
- On main temporal loop, and at each time step, OASIS “send” and “receive” primitives are called through the DATM module. To convey land model variables there, a driver modification is necessary: those variables have to be (potentially) remapped as if the prognostic atmosphere model was active.

Notice that the standard MPI management has to be slightly changed at initialization phase: CESM is not supposed to use the MPI_COMM_WORLD communicator, and its driver is forced to work with a local communicator (provided by OASIS). Consequently, a predefined OASIS routine is called by the CESM driver to let it switch off MPI.

Advantages

1. Simplicity: As previously said, rapidity and non-intrusiveness of implementation are a strength of OASIS. To call a set of initialization, declaration, sending, catching and ending OASIS interface routines adapted to CESM, we only had to modify 2 driver subroutines (ccsm_driver.F90 and ccsm_comp_mod.F90) and 1 DATM file (datm_comp_mod.F90).
2. Modularity: No other modification is required on COSMO and OASIS code or on their input files (in use on the previously set up CLM3.5 / COSMO / OASIS coupled model).
3. Scalability: taking advantage of the internal DATM parallelism, which could be adjusted independently of the CLM one, just changing a namelist parameter, OASIS exchanges could be made on a variable number of PEs. Figure 4 shows that the OASIS exchanges cost remains constant with parallelism (but expected to grow significantly at higher resolution with decomposition of more than 100 sub domains). On the contrary, it appears much more efficient to parallelize the DATM module (less than 0.01s on 122 PE but 0.3s when DATM runs on only 1 PE). Slowing down (reducing parallelism) occurs on remapping between CLM and DATM through the driver (driver_l2c, driver_a2c, driver_c2l, driver_c2a) but, above all, during DATM reading (“strdata_advance”) and scattering (“datm_scatter”).

4. Extensibility: on figure 4, a blue box represents different CESM modules, disabled in the present configuration. Theoretically, the same OASIS coupling interface (on DATM module) should allow us to exchange, with COSMO, information coming from (and given to) ocean, sea-ice or land-ice modules. To go further, the same interface may be implemented on other data modules (like DOCN) to ensure an OASIS coupling of the only CESM module that could not be plugged in the present configuration: the CAM atmosphere model.

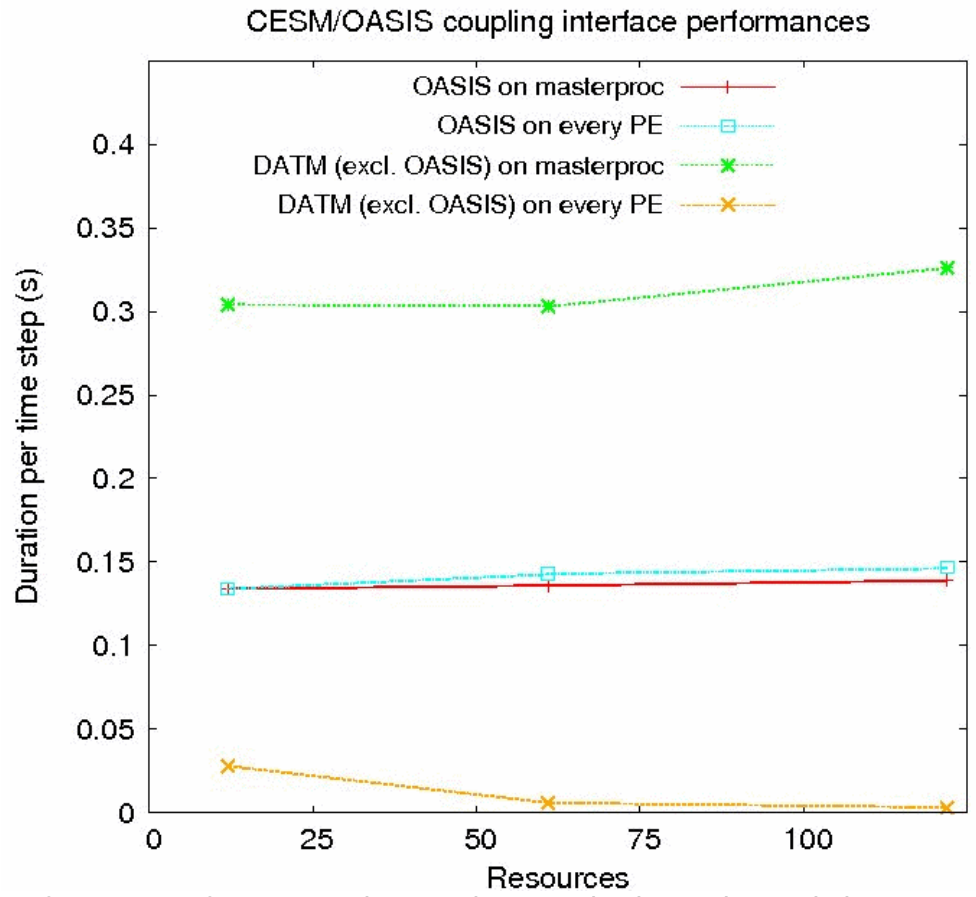


Figure 4: Performances of OASIS exchanges (communications + interpolations, two ways) and of DATM routines (excluding OASIS send/receive calls)

Current limitations

The present implementation only addresses problems of version update, allowing ETHZ to keep using their CLM model in an OASIS coupled system, though the new CLM version cannot be used easily without the whole CESM framework.

Considering low size of the targeted configuration, we prefer to focus on implementation facility rather than on performances, in order to facilitate management, by user, of next version updates.

Consequently, the system general design (presented in figure 3) strongly suggests that an further increase of parallelism (with higher resolution model) would lead to a lack of performances.

1. OASIS3 restricted parallelism (one OASIS process per coupling field) is not sufficient when problem size increases
2. CLM/DATM internal coupling, though efficient, increases the total time needed to exchange information between CLM and COSMO
3. As on any other OASIS coupling, MPI process (from the different executables) mapping on the reserved resources could significantly affect performances, which makes mandatory a fine and, possibly, difficult tuning
4. In addition, OpenMP could not be used, at least without code and/or MPI launcher settings modification

It is obvious that extra developments are necessary to be able to increase resolution and parallelism of the system. Will they be sufficient ? This question is the concern of the larger debate

of compared advantages of integrated/composite coupling.

Anyway, the OASIS capacity to make the *use* of an external model as simple as the reading of a forcing dataset should not hide that, once a technically validated configuration is available, a substantial work, including modifications of models parametrization, is then necessary to take into account the newly created coupled phenomena.

Annex 1: OASIS3 interface implementation on CESM

Added routines

oas_clm_vardef.F90	CLM/OASIS interface global variable definition
oas_clm_init.F90	Let OASIS organize MPI initilization
oas_clm_define.F90	Communicate model information to OASIS at initial step
oas_clm_finalize.F90	Let OASIS organize MPI ending
send_fld_2cos.F90	Fill arrays with coupling fields and call oas_clm_snd for each coupling field sending
oas_clm_snd.F90	Send one coupling field to OASIS
receive_fld_2cos.F90	Call oas_clm_rcv for each coupling field catching and fill model arrays
oas_clm_rcv.F90	Receive one coupling field from OASIS

Modified routines

To find these modifications on the code, see CPP key “COUP_OAS ”

ccsm_driver.F90 - Let OASIS close MPI communications , calling oas_clm_finalize
ccsm_comp_mod.F90 - Let OASIS define local MPI communicator (instead of MPI_COMM_WORLD) , calling oas_clm_init - Launch internal coupling routines to bring information to DATM module from other modules (force the prognostic atmosphere case) and particularly from land model to be able to use this information, sending it to OASIS
datm_comp_mod.F90 - Communicate model characteristics (grid lat/lon and mask, sub-domain distribution per process) to OASIS, calling oas_clm_define - Prepare coupling fields and send coupling field to OASIS (through send_fld_2cos routine) - Receive coupling fields from OASIS (through receive_fld_2cos routine) after reading complementary forcing fields (aerosols) and overwrite appropriate arrays with corresponding information

Compiling on CSCS system

Change CSCS batch_script (/project/s193/emaison/cesm1_0_3/scripts/batch_cscs.sh)

1. start from "I_TEST_2003" CLM configuration
2. indicate 2 new include directory and library (OASIS) to the CESM compile script:

```
export USER_FFLAGS="-DCOUP_OAS -I/users/emaison/oasis3/CRAYXT/build/lib/psmile.MPI1"
export USER_LDFLAGS="/users/emaison/oasis3/CRAYXT/lib/libpsmile.MPI1.a
                     /users/emaison/oasis3/CRAYXT/lib/libmpp_io.a"
```

3. copy OASIS interface fortran files (and CESM modified routines) into scratch compiling directory from /project/s193/emaison/CLM4/src/ directory to /scratch/rosa/emaison/testclm4/SourceMods/

Running on CSCS system

Launching directory: /users/emaison/COSMO4.8-CLM11-CLM3.5/run/clm4_EXP

Prepare new input files:

1. No need to change any Oasis and Cosmo parameter and input files (could be the same than CLM3.5 coupling)

2. Change some values within input_clm/lnd_in parameter file:

finidat = ' ' -> no restart
fatmgrid = 'data/surfdata_0122x0276.nc' -> CLM uses Europe grid, modified to match COSMO mask
fatmlndfrc = 'data/surfdata_0122x0276.nc'
fsurdat = 'data/surfdata_0122x0276.nc'

3. Change values of oatm_input/datm_atm_in parameter file:

dataMode = 'CLMNCEP' -> same option than initial CESM config
domainFile = 'data/surf_datm.nc' -> read DATM data (= OASIS coupling fields) on the same Europe grid than CLM, modified to match COSMO
streams = 'OASIS.stream.txt 1 1 1 ' -> take same forcing information at any time step from parameter file OASIS.stream.txt
vectors = 'null'
mapmask = 'nomask'
tintalgo = 'linear' -> those last 3 info for CLM/DATM interpolations (should not be used).

4. Build the "oatm_input/OASIS.stream.txt" fake parameter file:
This file allows to:
- read aerosols forcing file
- read other forcing variables. Those forcing values will be replaced by the coupling fields: they could be a simple copy of aerosols (or zero).
- define DATM model grid reading this file
The aerosol file defines the DATM grid. OASIS cpl fields are exchanged following this grid: That means that aerosol file defines the CLM grid (as seen by OASIS).

5. Build Netcdf input files:

oatm_input/surf_datm.nc: this file holds lat/lon information for DATM grid. It could be built copying CLM variables from file: /project/s193/emaison/preproc_CLM/surfdata_0122x0276.nc

Original variables	LONGXY	LATIXY	LANDMASK	AREA	LANDFRAC
Copy names	XC	YC	MASK	AREA (converted to radian squared, x2.464E-08)	FRAC

oatm_input/aero_dummy.nc: with correct aerosols data on CLM grid. WARNING: for the moment, aerosols values are not correct. Build them with NCAR tools.

6. Change some values on original drv_in parameter file:
   - the start date start_ymd (WARNING: COSMO/CLM calendars could be inconsistent)
   - the total duration (in time step and not in days)
   - the total task for both CLM and DATM modules. DATM task number could be equal to CLM total tasks (every PE are involved in the OASIS coupling), or equal to 1 (only master PE exchanges information through OASIS). WARNING: the number of PE involved in the coupling must be changed consistently on namcouple parameter files.

7. To build OASIS auxiliary files, DATM task number must be set to 1. Once the files are created, they can be saved and copied on the working directory before launching the next simulation. Then, the OASIS auxiliary files procedure won't be activated no more. This second production phase appears more efficient if DATM task number is then set to CLM total tasks.

Annex 2: CESM/COSMO coupling fields

Coupling field	OASIS naming rule (CESM interface)	Sent by
surface temperature	CLMTEMPE	COSMO
surface winds	CLMUWIND, CLMVWIND	COSMO
specific water vapour content	CLMSPWAT	COSMO
thickness of lowest level	CLMTHICK	COSMO
surface pressure	CLMPRESS	COSMO
direct shortwave downward radiation	CLMDIRSW	COSMO
diffuse shortwave downward radiation	CLMDIFSW	COSMO
longwave downward radiation	CLMLONGW	COSMO
total convective precipitations	CLMCVPRE	COSMO
total gridscale precipitations	CLMGSPRE	COSMO
wind stresses	CLM_TAUX, CLM_TAUY	CESM
total latent heat flux	CLMLATEN	CESM
total sensible heat flux	CLMSENSI	CESM
emitted infrared (longwave) radiation	CLMINFRA	CESM
albedo	CLMALBED	CESM

Bonus mission
Nov 3 2011

Host: Matthieu Masbou
Laboratory: Bonn University (Germany)

Main goal: Provide support on the previously designed COSMO-CLM OASIS coupling and tutorial on general OASIS use

Main conclusion
Bonn University users of COSMO-CLM model ended installing their configuration and start coupling Parflow hydrography model to the OASIS based system.

Model / machine description

COSMO-CLM (here called COSMO)
This regional atmosphere model (COSMO v4.8, and its climate version, COSMO-CLM v11) is used by a large community in several central Europe countries (from which Bonn University). DWD, MeteoSwiss and several other meteorological agencies host the operational version of the model. Grid: centered on West Germany settlements. High resolution (10km) is targeted.

CLM
This land model is developed at NCAR (v3.5)

ParFlow
Hydrological model developed at Bonn University. Finer resolution are targeted (100m)

The described configuration is developed for the German project TR32 (joining Aachen, Bonn, Braunsweig, Köln and Juelich Universities). TR32 is focused on soil/atmosphere interactions at spatial scale from Km to cm square. Possible extensions could lead to include WRF and ICON to the initial coupled configuration.

Model is available on the Bonn University local cluster.

OASIS3 interfaces for COSMO-CLM and CLM-ParFlow

OASIS3 interfaces on both CLM and COSMO models have been derived by Prabakhar Shresta (Bonn University) from the previous implementation described on Dedicated User Support reports #5 and #6)

He implemented a new functionality on OASIS (COOKING stage) to ensure efficient downscaling between coupling field exchanged between highly different spatial discretization scales (Schomburg at al. 2010). This development has been proposed to the OASIS development team.

He is currently modifying COSMO spatial discretization to match perfectly CLM grid requirements

(due to no possibility of grid stretching on CLM model).

For this coupling , IS-ENES support only consists in useful bypasses or advices and corresponding report to OASIS users such as:

- mpp_io / OpenMPI 1.2 mismatch on previously designed cluster (and Intel compiler). Solution consists in disabling mpp_io features and providing Moritz Hanke's bypass (see Dedicated User Support reports #5 and #6)
- NOBSEND option disabling for large buffer exchanges
- incompatibility of OASIS3 pseudo parallel mode and OASIS grid writing functionalities (solved by OASIS3 ETHZ modified version providing)

Concerning CLM-Parflow coupling, a first implementation is currently developed by Mauro Sulis and a quick overview of the implementation state has been done. OASIS3 version use allows parallel coupling on CLM3.5 (instead of master-processor-only coupling, implemented at ETHZ). This option should allows TR32 users to enhance coupling performances when a fully parallel version of OASIS3 will be practically available.

Parflow C-language written code benefits from a C encapsulated version of the PSMILE routines (also developed at Bonn University).

Difficulties have been expressed by developers on topics such as:

- prism_put/get positioning on the newly coupled Parflow model and on CLM (for Parflow exchanged coupling fields)
- OASIS restart functionality
- prism_def_var_proto argument characteristics

The OASIS support gives us opportunities to:

- better explain characteristics of the previously developed OASIS interface and ensure diffusion of IS-ENES realization
- identify usual difficulties consecutive to OASIS interface implementation and parametrization
- report unknown malfunctions
- evaluate Bonn University OASIS related work and possible contributions to OASIS further enhancements

Mission #9
Feb 6- Mar 2 2012

Host: Uwe Fladrich
Laboratory: SMHI, Norrköping (Sweden)

Main goal: Measure and enhance performances of the OASIS3 based Ec-Earth model

Main conclusion
A portable and OASIS3 pseudo-parallel mode compliant version of our OASIS performance measurement tool has been developed and tested on several SMHI models. Thank to it, it could be soon possible to measure and compare the coupling extra cost of the standard OASIS3 based version and the presently implemented OASIS3-MCT based version of the Ec-Earth high resolution model. At the same time, different interactions contributed to set up two new OASIS3 coupling with regional models (RCA-NEMO and RCA-RCO).

Model / machine description

SMHI's coupled model (high resolution version) originally deals with:

- IFS, cycle 36: T799, 843.490 grid points, ˜25Km, 62 vertical levels, time step: 720s
- NEMO, v3.3: ORCA025, 1.472.282 grid points, ˜40Km, 45 vertical levels, time step: 1200s
- OASIS v3 (pseudo parallel)

20 coupling fields are exchanged between the two components at a coupling frequency of 3 hours. The model is available on Ekman supercomputer, 1.268 compute nodes of 2 quadripro AMD Opteron (# 10.144), Infiniband interconnection, located at Royal Institute of Technology (KTH), Stockholm, centre for parallel computers (PDC).

OASIS3-MCT upgrade

Set-up during #4 Dedicated User Support[31], the Ec-Earth OASIS3 based configuration was still slowed down by coupler, and its performances supposed to be strongly reduced on machine allowing massive parallelism.

For several reasons, the replacement of OASIS3 by OASIS3-MCT has been preferred to the firstly envisaged OASIS4 upgrade.

Started on the ekman machine, the replacement process consisted in a very few operations:

- into code interfaces (a single mod_prism module has to be called instead of a suite of specialized module)
- on namcouple (simplified due to the fact that OASIS3-MCT is currently not able to calculate interpolation weight, but only to read it on a file, which name must now be specified on

[31] Maisonnave, E. and Valcke, S.: OASIS Dedicated User Support 2010, Annual Report ,Technical Report, TR/CMGC/11/28, SUC au CERFACS, URA CERFACS/CNRS No1875, France (2011)

namcouple)

Several FORTRAN philosophy related inaccuracies (argument array dimensions) have been corrected on IFS and NEMO coupling interface to be able to reproduce, with OASIS3-MCT, the identical coupling process, including coupling field restart read/write.

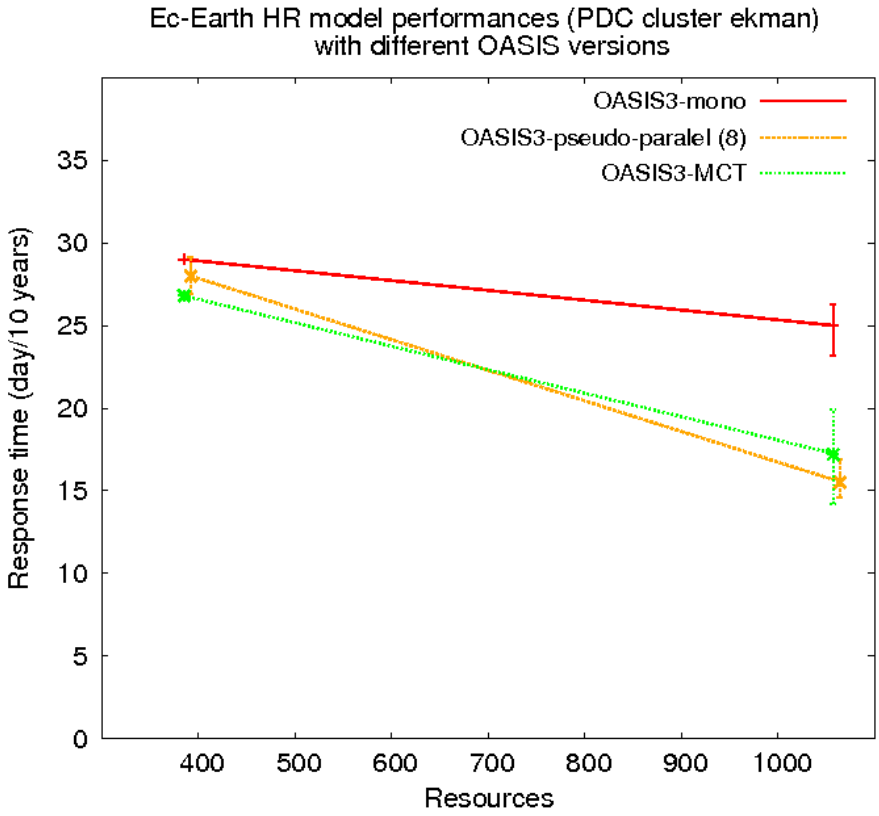


Illustration 1: Ec-Earth performances with coupler upgrade

On Figure 5, measurements of coupled model performances (total simulation time) are shown for both OASIS3 and OASIS3-MCT based configurations. As a reminder, performances taken with an OASIS3 mono-process coupling (all coupling fields are exchanged through a single OASIS3 process) are also displayed.

Due to the Ec-Earth coupling sequence (IFS and NEMO run in parallel), a large amount of the time needed for coupling (MPI message exchanges + interpolations) is hidden by the speed difference between the two components: most of the coupling operations are performed when the fastest model has ended its computations and before the slowest model has ended theirs.

Combined to the relatively low parallelism level of the configuration (reducing the amount of exchanged messages during coupling and consequently the time needed to perform them), this particularity forbids to get a clear idea of how profitable our coupling technique enhancement is. Nevertheless, a small fastening is observed comparing our o(1000) parallel tests.

To better estimate the OASIS3-MCT benefit, several new experiments have been performed. The first idea was to increase parallelism (to increase coupling exchanges and test both coupler ability to manage them). To do so, a larger machine was required: PRACE tier-0 machine “curie” has been targeted and the model ported on it by John Donners (SARA) through the IS-ENES/PRACE

IP1 joint project. Unfortunately, the results of this work has been delay by several “curie” operating defaults (machine upgrade) and can not be presented here.

The second idea was to change the coupling strategy and run atmosphere and ocean sequentially: this technique has the advantage to clearly make appearing all the time needed to perform coupling, as each model is waiting calculation results of the other one and OASIS operations needed to bring the information to its receiving interface. But this configuration could be difficult to set up and has no particular scientific interest for Ec-Earth teams. Consequently, test has been done (on “curie”) using the widely used CERFACS coupled model ARPEGE-NEMIX. Results clearly show the interest of an OASIS3-MCT upgrade at such level of parallelism[32].

Even though such result could be easily extended to the ocean-like and atmosphere-related model Ec-Earth, we preferentially would like to show the reduction of coupling time induced by the OASIS upgrade on the Ec-Earth model itself.

OASIS3 performance measurement toolkit

To reach this goal, a precise evaluation of the coupling communication and interpolation cost is required. Such quantities could be evaluated using the OASIS dedicated support development presently available on current coupler release. Activating the “balance” CPP key during OASIS3 compiling, MPI_Wtime clock time measures are printed on “prt” OASIS output files.

During the simulation, each time than a model sends or receives a coupling field, a clock time is output before and after the corresponding PSMILE library call. Symmetrically, the same measures are printed from coupler side.

On a post processing phase, a shell script (sh_balance) is used to convert the different measures into synthetic informations. Despite several advantages, proved by its capacity to provide all the previous performance related informations published in the previous OASIS Dedicated User Reports, the increasing level of models parallelism, but also coupler parallelism (OASIS3 pseudo parallelism) reveals the limit of a shell based development.

For those reasons, we decided to entirely re-write our tool in FORTRAN-90, ensuring its portability at the same time than its capacity to process results produced on massively parallel systems. Nevertheless, given that each model process, at each coupling time step and for each coupling field, writes a certain amount of ASCII format information on files, performances could be affected by a relatively large amount of disk access.

OASIS instrumentation

To partly avoid such drawback, an simple enhancement on OASIS implementation consists in suppressing FORTRAN “flush” routine call after each write file access call. But possibility must be given to the user to keep this functionality when an on-line analysis is required.

A second enhancement on OASIS implementation is necessary to measure the possible time shift between the different clocks of nodes allocated to our coupled model. This issue is particularly difficult to address, and an exact synchronization impossible to achieve. We assume that a simple

[32] See PRACE IP1/IS-ENES WP8 joint project web site: https://redmine.dkrz.de/collaboration/projects/prace/wiki/_OASIS4_upgrade_

measure after the coupling initialization phase MPI_Wait call (common to all process) will fit our precision requirements.

Those two enhancements will be soon available on OASIS3 official distribution.

Post-processing tool

The FORTRAN executable is called through a simple shell script, which ensures portable compiling (-c option) at the same time than execution. Completing the analysis, the graphical tool "gnuplot" is used (if available) to produce a simple EPS format visualisation of the main results.

As previously described, each time than a coupling field is exchanged by any process involved in the coupling, two clock measures are produced on the corresponding "prt" file[33]: one before calling the PRISM sending or receiving routine, and one after.

Those two standard measures are read twice by our FORTRAN program.

On a first step, our program identifies which coupling field is exchanged by which model and counts how many time it is. This first reading allows us to determine the arrays dimensions which will contain the information to process:

- the number of exchanged fields[34]
- how many times are they exchanged

The field exchange sequence (as seen by coupler) is deduced and displayed on standard output[35]. This information will help the user to check whether this sequence matches the sequence defined on the models. If not, it means that buffered MPI communications are probably activated.

Program also checks that each field is received (by coupler or by a model) as often as it is send (by a model or by coupler). If not, a message is displayed to inform the user that simulation did not end correctly. Consequently, further analyses will be systematically done excluding the last two coupling step. Symmetrically, the first coupling step is also excluded to not take into account restart operations duration that could slow down the simulation beginning.

For those reasons, the total simulation elapsed time (as observed with a simple UNIX "time" command) is greater than the figures given by our program.

On a second step, information available on "prt" files is read again. Now, the purpose is to fill the two different arrays allocated with the previously defined dimensions.

One dimension of those two arrays is 8. This figure matches the measure number necessary to trace every operation done to carry a given coupling field from one model to the other, i.e. before (a) and after (b) source model send, before (c) and after (d) coupler receive, before (e) and after (f) coupler send and before (g) and after (h) target model receive.

[33] There is one "prt" file per coupled model process (model or coupler)
[34] Equal to half the number of fields exchanged on all coupler executable (they could be several if OASIS pseudo-parallel mode is enabled)
[35] On OASIS pseudo parallel mode, the first fields are those described on namcouple_1 file

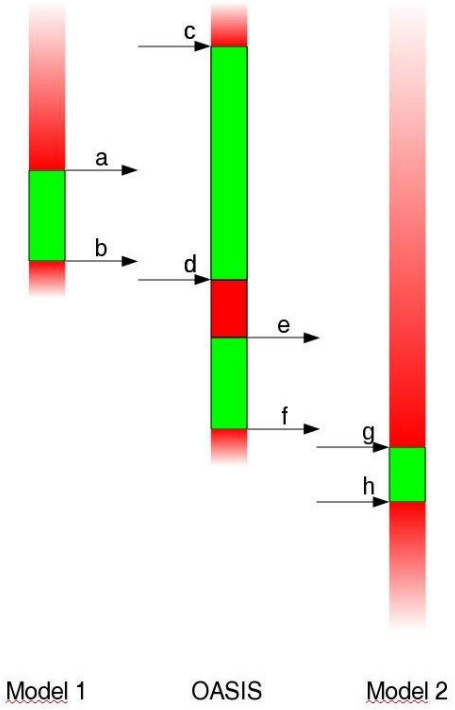


Illustration 2: Naming convention for coupling field exchanges

Load balancing

For each process involved in coupling and for each coupling field, our first data array is filled with the cumulated value, along all the valid coupling steps, of 3 quantities:

- the time needed to send the coupling field: timing (b) – timing (a)
- the time needed to receive the coupling field: timing (h) – timing (g)
- the time needed to perform all the other operations:

timing (a) – timing_from_previous_cpl_time_step (b) (for source model)
timing (e) – timing (d) + timing (c) – timing_from_previous_cpl_time_step (f) (for coupler)
timing (g) – timing_from_previous_cpl_time_step (h) (for target model)

The first two operations gather the time needed to write or read coupling fields on MPI buffers but also the time spend to wait the moment when those exchanges are allowed. After adding quantities for all coupling fields, this time could be seen as the time that models need to exchange the coupling fields. Consequently, the third operation could be seen as a time when coupling independent operations are performed[36]. For more convenience, we will call it “computation” time.

As different model (and coupler) MPI process could start or end the different operations at different moments, we choose to selected the maximum duration from every process.

On coupler side, the important figure lies on this computation time: it measures interpolations and other operations speed. It cumulates the time spent by each coupling field: on OASIS pseudo-parallel mode, it does not take into account the fact that several coupling field are processed at the same time. On any mode, if SEQ namcouple option is the same for each coupling fields, our

[36] It is not exactly equal to the total time needed by the model to perform a forced simulation, on a stand alone mode because, in this case, coupling field reading is replaced by forcing field reading

calculated quantity adds several times the duration when OASIS is performing all the interpolation of fields with same SEQ parameter. For all those reasons, this OASIS calculation time could generally not be considered as the total elapsed time needed to perform the various calculations, but better as a measure of how fast a subset of coupling fields is computed. On pseudo parallel mode, this quantity must be compare with itself for several machines, or for several model parallelism. On OASIS mono-processor mode *only*, it could be seen as the total time needed to perform interpolations, and only when SEQ parameter differs from one coupling field to the other.

Even though adaptations are needed to take into account special (and quite non standard) cases[37], we could test these analysis on two different SMHI OASIS3 based coupled models: Ec-earth and RCA-NEMO. For the first example, simulation has been arbitrarily stopped before the end defined in the different namelists and namcouple. Our tool don't need a complete simulation to give its results and can be launched on working directory even during the simulation.

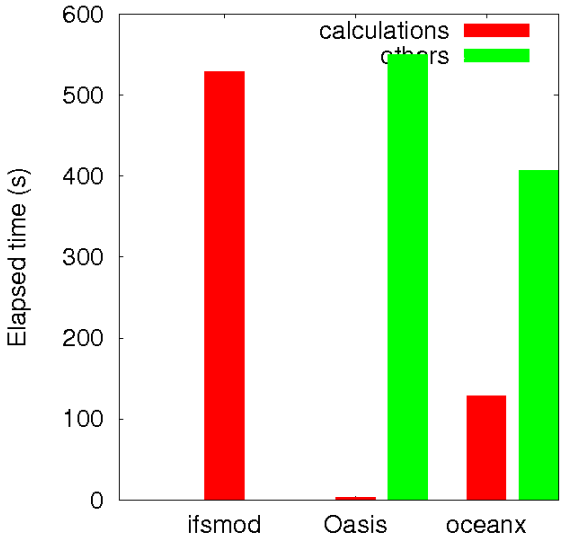


Illustration 3: Balance analysis for IFS/NEMO coupled model

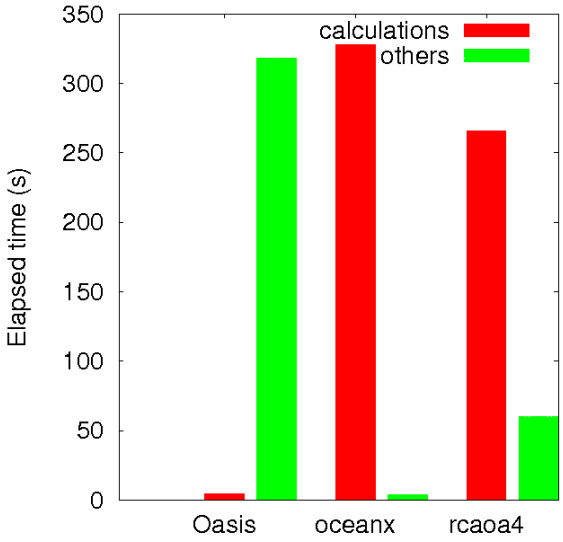


Illustration 4: Balance analysis for RCA/NEMO coupled model

On graphics proposed above[38], red boxes represent “total” calculation time, and green boxes “total” time needed to send, receive or wait the coupling fields.

Coupler calculation time is negligible compared to model computation time. Matrix multiplication (interpolation) of global grid can be quickly processed. On RCA-NEMO configuration, OASIS is used on mono-processor mode and each coupling field has different SEQ parameter: the calculated quantity adds calculation time of each coupling field. On IFS-NEMO, the OASIS pseudo parallel mode is enabled and all coupling field of one coupler instance have the same SEQ parameter. This could explain that IFS-NEMO value of OASIS computation time remains lower than RCA-NEMO one, despite its finer resolution.

Comparing computation time for model of a same configuration allows us to conclude that RCA and NEMO response times are much more balanced than IFS and NEMO ones. To avoid that a model spends too much time waiting to the other, it is recommended to allocate more resources to

[37] For example, when coupling time step differs from one model to the other

[38] Those pictures are automatically produced by our tool (EPS format), if gnuplot is available on the processing machine

the slowest model. To make his (her) choice, the user has also to take into account the relative scalability of the two (or more) coupled system components.

One interesting aspect of our tool is that, comparing the different computation performances measured with several resource numbers, one could establish, simultaneously, scalabilities of the different components. This method is much more rapid and accurate than measuring those results with the stand-alone models, mainly because forced and coupled configuration scalabilities could differ.

Coupling efficiency

A second array is produce during the second part of our program.

For each coupling time step, and for each coupling field, we select, for each model or for coupler, the latest timing produced by any of its MPI process for the following quantities:

- the time for the source model needed to write the coupling field to MPI buffer + the times needed for the coupler to read it and write the interpolated field + the time for the target model to read it.
- The time spent by any model (or coupler) to wait coupling field, needed to keep calculating.

The first quantity is equal to:

(timing (d) – timing (a)) + (timing (h) – timing (e))

But timing (c) – timing (b) must be subtracted from this quantity each time than timing (b) is prior to timing (c), i.e. each time that OASIS is waiting the model to start interpolations. Identically, timing (g) – timing (f) must be subtracted too from the same quantity each time that timing (f) is prior than timing (g), i.e. each time than source model is waiting the coupler to start its new calculations.

Consequently, in the previously described case, we can increase the total time when target model (and/or coupler) is waiting an information with the quantity:

timing (g) – timing (f) and/or timing (c) - timing (b)

One should notice that, when OASIS is waiting coupling fields from source model, the target model can wait the coupler at the same time. It means that the addition of the two quantities is not a measure of how much the fastest model is slew down by the coupling (and the slowest model)[39] but how much all the coupling operations are slew down by the chosen coupling sequence.

Extension to OASIS3-MCT

Due to time limitation reasons, it has been impossible to design an identical tool for the new OASIS3-MCT version. A preliminary step will be necessary: a coupler instrumentation allowing to print timing on output files.

Consequently, at the end of the Dedicated Support period, it was still not possible to compare OASIS3 and OASIS3-MCT on EC-Earth high resolution model. Nevertheless, CERFACS plans to end this work next time than an efficient OASIS3 / OASIS3-MCT comparison will be necessary (ANR PULSATION project, for example).

[39] i.e. timing (g) – timing (f)

Other debugging activities

In addition to EC-Earth, two other OASIS based couplings are currently implemented at SMHI: the regional RCA-NEMO model and the previously OASIS4 based regional RCA-RCO model.

At this occasion, we could expand our data base with OASIS possible misleading features, from which the possibility given to the user to:

- read OASIS restart file with wrong dimension (unclear failure)
- send and receive halos in addition to working arrays (high possibility of shifts in the coupling field communication)
- open existing interpolation weigh file in UNIX “write” mode (could be protected)
- spend a lot of time searching the FORTRAN north_thresold non parametrizable variable and its right value (1.6) to avoid weigh calculation anomaly at north pole (SCRIP conservative interpolation)
- hang his/her simulation only when using NOBSEND exchanges at some high parallelism level
- choose between misleading ways to globally conserve (CONSERV operation) spatially weighted and cumulated values
- use mask values (“masks” file) different from those previously used to calculate weights
- choose SCRIP related “bins” number without a clear diagnostic on how seriously an interpolation could be affected near sub-domain boundary if this number is insufficiently high (and, at the opposite, how too much time consuming is an insufficiently low number)

Those different questions, asked by five different persons (mainly with permanent positions) all along the Dedicated User Support period, give a good information on general users strategy chosen to set up an OASIS coupling.

Mission #10
Aug 14- Sep 7 2012

Host: Andreas Will
Laboratory: BTU, Cottbus (Germany)

Main goal: Set-up OASIS3-MCT interfaces in global and regional atmosphere models to enable a 2-way nesting

Main conclusion
Based on OASIS3 previously developed interfaces (for MPI-OM, CLM and NEMO coupling), a 3 dimensional OASIS3-MCT coupling between COSMO and ECHAM has been set up. Observed performances (without advanced optimisation, the coupling overhead on a mono-node IBM Power6 simulation is about 6% of the total elapse time) reveal that an OASIS3-MCT tight 3 dimensional coupling (the coupling frequency is equal to the ECHAM time step) is inexpensive at low resolution, and therefore most probably acceptable at high resolution.

Model / machine description:

COSMO-CLM (here called COSMO)
This regional atmosphere model (COSMO v4.8. In its Climate Mode: COSMO-CLM v11) is used by a large community in several central Europe countries, from which Eidgenössische Technische Hochschule Zürich (ETHZ). Deutscher Wetterdienst (DWD), MeteoSwiss and several other meteorological agencies host the operational version of the model. The grid size is 221x111x47, i.e. ~2 degrees. A decomposition on 32 MPI tasks was used for tests on the target supercomputer.

ECHAM/MPI-OM
The Max Planck Institut für Meteorologie (MPI-M) global atmospheric model is here used in its 6th version. The starting configuration already includes an ocean model (MPI-OM), coupled at Deutsche Klimarechenzentrum (DKRZ) through OASIS3-MCT. The grid size is 192x96x47 (T63) for ECHAM and 254x220 for MPI-OM. A decomposition on 32 MPI tasks was used for tests on the targeted supercomputer.

OASIS:
Both OASIS3 (version 3.3) and OASIS3-MCT (version 1.0) has been used during this work.

Those models are available on IBM Power6 supercomputer, which has 8,448 compute cores (16 dual-core processors per node) and an Infiniband interconnect (peak performance of 158 Teraflop/s). The machine is located at DKRZ, Hamburg, Germany.

Rationale

An increasing number of studies presently require models with more accurate horizontal and vertical resolution. But global high resolution CGCM are, in many cases, not affordable in terms of required human and computing resources. The solution consisting on a local increase of resolution on regions of interest can satisfy most of the present scientific project requirements.
Such zoom can be defined on a sub-domain of the model grid, where calculations are refined. WRF atmosphere or NEMO ocean models, for example, both proposed integrated solutions to

allow one way (the inner model is forced by boundary conditions provided by the largest model) or two way nesting (in addition, the largest model considers updated information produced by the inner model).
Researchers of the Cottbus Brandenburgische Technische Universität (BTU), in collaboration with the Berlin Frei Universität (FUB), both belonging to the COSMO community, proposed to use two different models for global (ECHAM) and regional (COSMO) modelling and to take benefit of OASIS to exchange the information necessary for a 2 ways nesting between these models.
Independently of various scientific issues (buffering, filtering ...) not addressed in the present support, a clear challenge of such coupling is the huge amount of information (3D fields) exchanged at a very high frequency (largest model time step). This requirement (exchange of large volume of data) is the major reason why integrated coupling, i.e. the two models are merged into one executable and the data is exchanged through the memory, is presently preferred instead of an external OASIS coupling, in different laboratories such as Météo-France (coupling between ARPEGE atmosphere and SURFEX land surface) or LOCEAN (for the coupling between ocean and LIM sea ice, or PISCES Biogeochemistry).

OASIS interfaces

Both COSMO and ECHAM models already included OASIS3 interfaces.
Previous OASIS Dedicated User Supports (see missions #5 and #7) led to the implementation of OASIS3 interfaces in the Community Land Model (CLM) and in NEMO ocean model (see mission #9).
In ECHAM, a recent DKRZ upgrade to OASIS3-MCT of the atmosphere-ocean coupling with MPI-OM gave a base for an extension to the presently described ECHAM-COSMO coupling.

First implementation
A preliminary study was hosted by the Centre Suisse de Calcul Scientifique (CSCS), gathering three COSMO community members, Edouard Davin (ETHZ), Jennifer Brausch (DWD) and Andreas Will (BTU) as COSMO coordinator.

In order to ensure COSMO capability (I) to exchange simultaneously coupling information with three different models (NEMO or MPI-OM ocean, CLM land surface and ECHAM AGCM) and (ii) to possibly extend coupling to other models, it was decided to implement a single and modular OASIS3 interface for all coupling.

Following NEMO example, the two coupling operations (collecting information to send out and filling model variables with received information) are achieved by common routines for all the coupling fields. The different operations are launched following the type of coupling chosen by namelist (ocean coupling, land surface coupling and/or atmospheric 2 ways nesting). Unfortunately, despite numerous but probably unclear advices of the author, developers did not follow NEMO example, and coupling choices are hard coded with respect to a limited set of models. Consequently, it is not possible to select by namelist different coupling fields for a given type of coupling. For example, the planned coupling between COSMO and MPI-OM ocean model will not be possible through interface implemented for the coupling to NEMO ocean model. A duplication of the code implementing the interface will then be unavoidable.

The simultaneous use of several models in our coupled system imposes the choice of the same version of the coupler for all components (given that OASIS3 and OASIS3-MCT cannot be used together in one coupled configuration). For performance reasons, it seemed more promising to choose OASIS3-MCT as a common version of coupler.

Nevertheless, the existing OASIS3 interfacing was kept as an option in COSMO, through a CPP

key. This ensures, for the moment, the possibility of coupling COSMO and version 4 of the Community Land Model.[40] The high similarity between OASIS3 and OASIS3-MCT Application Programming Interface (API) makes the amount of additional code necessary to propose this modularity very limited.

After one week of joint developments with A. Will, S. Weiher (BTU), M. Thürkow (FUB), J. Brausch (DWD), E. Davin (ETHZ), J.-G. Piccinali (CSCS) for model performance measurement and the author for OASIS support, a unified interface is now available in COSMO for coupling to ECHAM, NEMO and CLM.

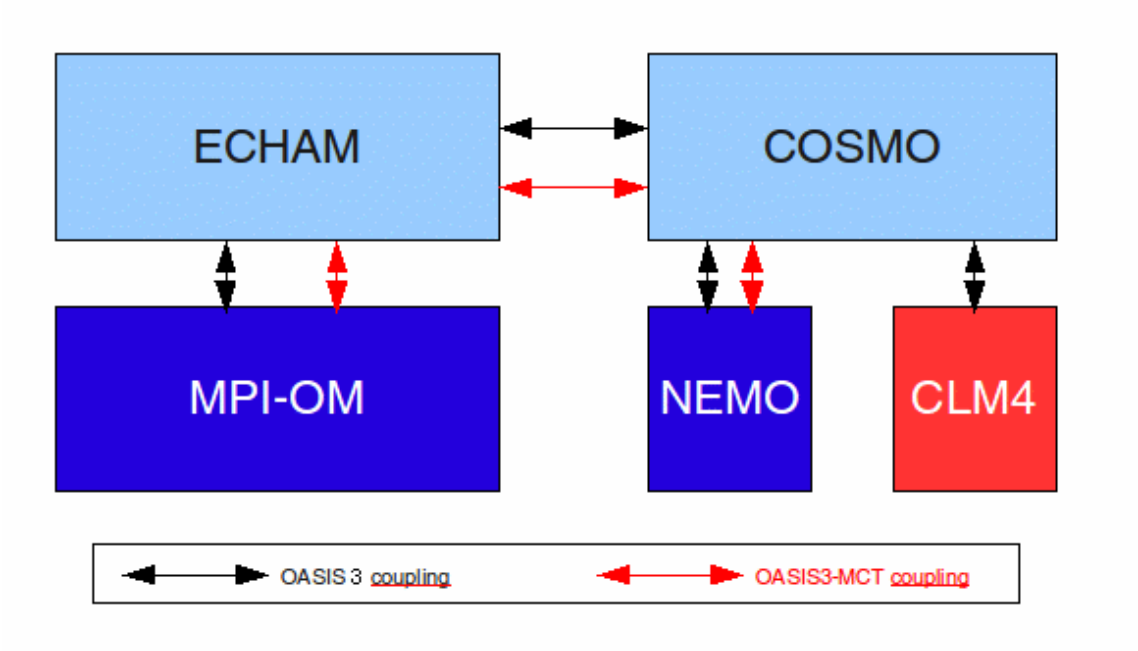


Fig 1: COSMO-CLM coupled constellation currently implemented

Two ways nesting

The ECHAM interface

The targeted global/regional coupled model also includes coupling with the MPI-OM ocean, thanks to the pre-existing ECHAM-MPI-OM coupling based on OASIS3-MCT. In ECHAM, the existing OASIS3-MCT interface was extended to exchange necessary information with COSMO.

Duplication of the existing interface was chosen by ECHAM developers. In addition to the existing MPI-OM related mo_couple.f90 routine, a similar routine was created for COSMO coupling (mo_couple_two_ways_nesting.f90)

As previously reported (see mission #1), the existing interface does not take advantage of OASIS per-coupling-field accumulation option and accumulates independently the coupling field values according to a single pre-defined coupling frequency.

A mixed system of coupling field declaration (oasis_vardef) depending on namcouple (MPI-OM part) and model namelist (COSMO part) allows the simultaneous use of the two interfaces. Nevertheless, a high level of code duplication and the impossibility to use this code in the previous

[40] CLM4 is a module of the CESM coupled model, which also uses MCT for internal parallelism. For the moment, no clear solution was found to allow simultaneous use of MCT library in CESM and OASIS.

configuration (ECHAM/MPI-OM coupled model without COSMO) strongly foresees the future (and advisable) rewriting of a common interface, following NEMO or, now, COSMO examples.

Launching script

As usual in such coupling exercise, when models are coming from different communities, one of the two compiling and running environments must be chosen and adapted to the other model needs.

Concerning compiling, due to time limitations, both existing environments were kept, even though it appeared clearly that ECHAM's “Integrating Model and Data Infrastructure” environment (IMDI) is much more adapted to a common production workflow than to such development/debugging exercise. MPI-M team must be contacted to discuss a solution.

Thanks to Ingo Kirchner (FUB) developments, a COSMO style launching script (template) was delivered during the OASIS support period and a COSMO/ECHAM/MPI-OM coupled simulation could be quickly launched: after a first COSMO stand alone run, ECHAM and MPI-OM input files and adapted namelists are created in the COSMO working directory, and the coupled simulation is launched with the COSMO additional component.

A modification was necessary in the coupled template script to take into account the additional two-way nesting coupling fields in the initial ECHAM/MPI-OM namcouple. Optionally, our script allows to exchange the 3D fields as single 2D fields at each level or, together, via a single communication.

Coupling strategies

In our configuration, regional and global models perform their calculations at the same time: the two models are running sequentially (see fig 2.) It means that, unlike most of the presently implemented OASIS coupling in our climate community, any slow down during the coupling operation has an equivalent impact on the total coupled model restitution time. It is then particularly important to try to minimize any coupling overcost.

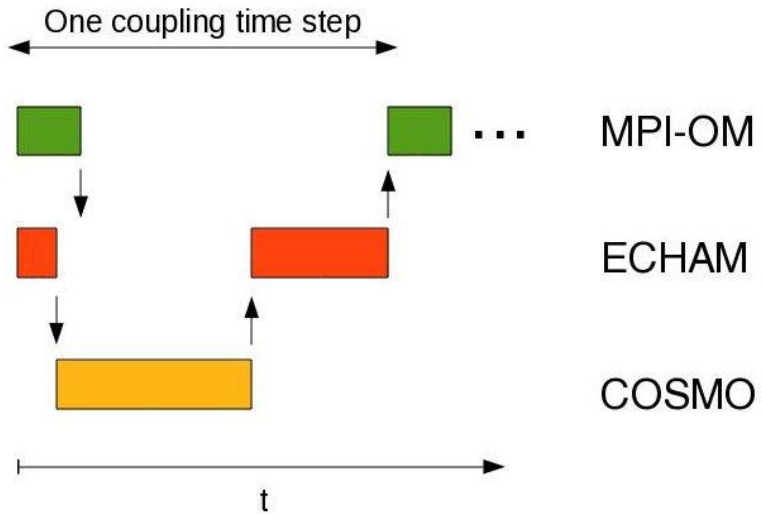


Fig 2: COSMO/ECHAM/MPI-OM coupling sequence

Pseudo 3D coupling

Present COSMO/ECHAM main coupling characteristic is the 3 dimensional size of most of its exchanged fields (see Annex 1). A clear concern of such configuration is the time needed to interpolate and exchange the whole information at each time step of the global model. Compared to a standard coupling, several improvements were necessary to reach an acceptable level of performances.

One usual way of improving the coupling performance is to reduce the amount of exchanged information. Considering that models have a different number of vertical levels, one improvement could be to perform the vertical interpolation by the source processes before the exchange if the target model has less vertical levels, or by the target processes after the exchange if the source model has less vertical levels in order to minimise the amount of information transferred. For the moment, the described coupling configuration does not include this vertical interpolation but only *n* horizontal interpolations, with *n* equal to the number of ECHAM vertical levels (47).

On a preliminary intent of coupling optimisation, an OASIS3-MCT namcouple parameter was adapted to our particular case. Its functionality consists in coupling multiple fields via a single communication (see §B.4 of OASIS3-MCT User guide). Inside OASIS3-MCT, fields are stored together and a single mapping and a single send or receive instruction are executed for the whole group of fields. We will call this technique a “pseudo 3D coupling”.

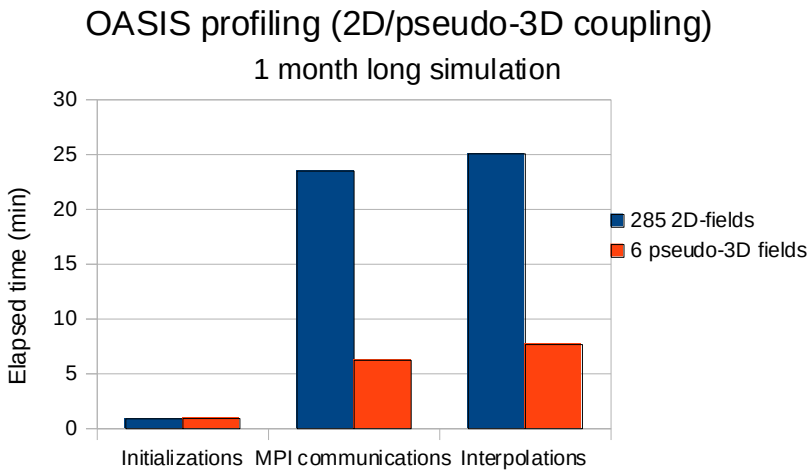


Fig 3: OASIS profiling

The benefit of such improvement is clearly visible in the given profiling (fig. 3) made with the embedded OASIS3-MCT measurement tool (routine set from mod_oas_timers file). Measurements can be turned on by changing a code variable (TIMER_debug) before compiling. With TIMER_debug=1 or 2 timing files are produced at the end of the simulation (one for the root process of each model, one for the other processes of each model). In each file, one paragraph summarizes the total elapsed time spent by several group of OASIS3-MCT routines, that can be gathered into 4 categories:

- initialization (map_definition, cpl_setup, cpl_smatrd and advance_init groups, called only once at the beginning)
- communications (psnd_* and prcv_*, one per coupling field)
- interpolations and other operations (pmap_*, pcpy_* and pavg_*)
- coupling restart writing (wrst_*) if any (it is no more the case in our coupled configuration)

These diagnostics are made on each process. Minimum, maximum and mean values, considering all process values, are displayed in the root process result file. We choose to only consider the maximum value, even though this strategy could lead to over-estimate the OASIS3-MCT impact on

performances[41].
Initialization is negligible in our case. Interpolations and communications during coupling exchanges share the total time spend on OASIS (see figure 4).
We compared such quantities for 2 one-month long simulations that only differs by the way the coupling fields are exchanged: 285 2D-fields or 6 pseudo 3D-fields, each pseudo 3D field being composed of 47 2D fields (+ 3 2D-fields). Considering figure 3, it clearly appears that the huge number of MPI messages required in the first case strongly slows down the simulation: communications then takes 24 min. The solution that consists in grouping the small messages into a bigger one must be preferred; the communication then takes only 6 min. Similarly, an interpolation (matrix multiplication) performed with bigger arrays seems faster (7 min) than several small sequential operations (25 min).

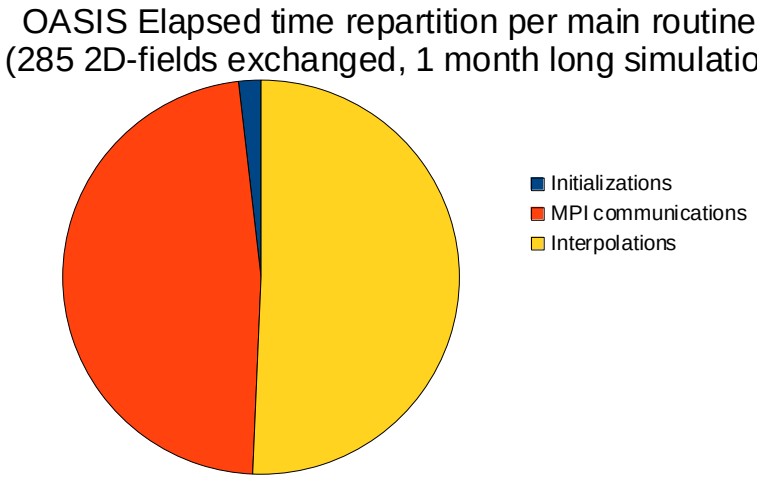


Fig 4: Profiling repartition between OASIS routine categories

The total extra-cost of OASIS3-MCT coupling could be deduced doing a difference between:

- a coupled run where exchanges are limited to only 1 2D-coupling field (where OASIS coupling cost is reduced to less than one second)
- and our full 285 2D-fields or 6 pseudo 3D-fields configurations.

From figure 5, we can evaluated the OASIS3-MCT overhead to 50% in the first case (285 2D-fields) and 6% in the second (6 pseudo 3D-fields). This last figure tends to prove that OASIS3-MCT is quite suitable for such 3D coupling necessary at each time step of our 2 way nesting. This result must be extended to any OASIS3-MCT coupling composed of many similarly interpolated fields (and not necessarily only different vertical levels of the same variable).

[41] Profiling of “receive' configuration is ambiguous for some coupling fields: it can both include communication time and waiting time (the target model is waiting for the source model to finish its calculations and provide the coupling fields)

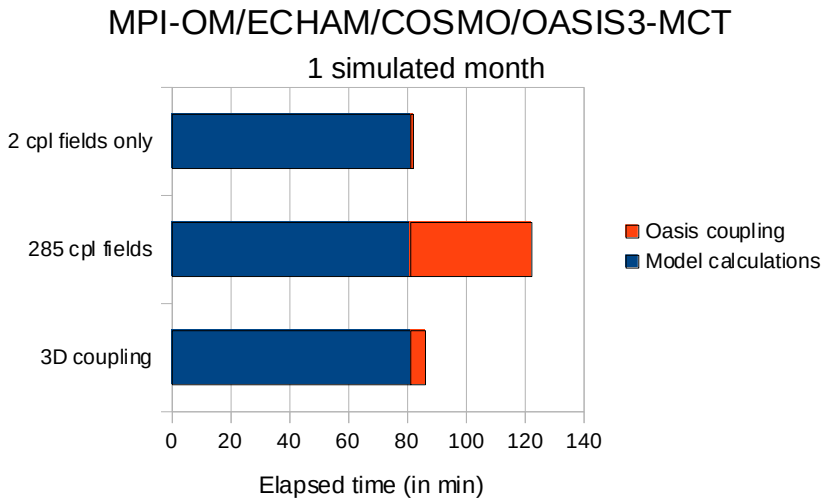


Figure 5: OASIS extra cost

An additional experiment made on more than a single node (64 processes spread on two different 32-cores nodes instead of 64 processed launched on the same 32-cores node) does not exhibit notable increase in OASIS communication (nor interpolation) time. It suggests that, even though we would increase parallelism for higher resolution configurations, using greater number of nodes, our experimental OASIS setup should still keep its good performances[42].

Hyper-threading

Another information can be deduced from this complementary experiment: an hyper-threading of our model processes gives comparable performances than allocating a core to each of them. This could be explained by the sequencing of our coupled system (see fig. 2).

MPI-OM and ECHAM run in parallel: the exchanged coupling fields are calculated by the other model at the previous coupling time step. A contrario, ECHAM and COSMO run sequentially: COSMO is called as a subroutine of ECHAM. These characteristics implies that models are idle half of the time. With hyper-threading (32 cores are allocated for 64 processes), 2 processes are sharing the same resource and using it alternatively. Any extra resource (for example when 2 nodes are allocated) is then useless and there is no gain in restitution time.

Current limitation, further developments

Interpolations choice

Default distance weighted nearest-neighbours interpolation (SCRIPR/DISWGT) was chosen in both ways. We approximately calculated that a minimum number of 4 neighbour[43] is needed from ECHAM to COSMO and 10[44] from COSMO to ECHAM. Those figures can be tuned, following the best compromise between quality and performances. On Fig 6 is shown the general aspect of a coupling field (high atmosphere temperature) before and after ECHAM/COSMO interpolation.

[42] Obviously, this also depends on machine interconnect network and MPI implementation.
[43] Minimum number of neighbours for a DISWGT interpolation: less will lead to produce strong gradient on the target grid at source grid point limits
[44] This figure is deduced from the average ratio between source and target grid point areas.

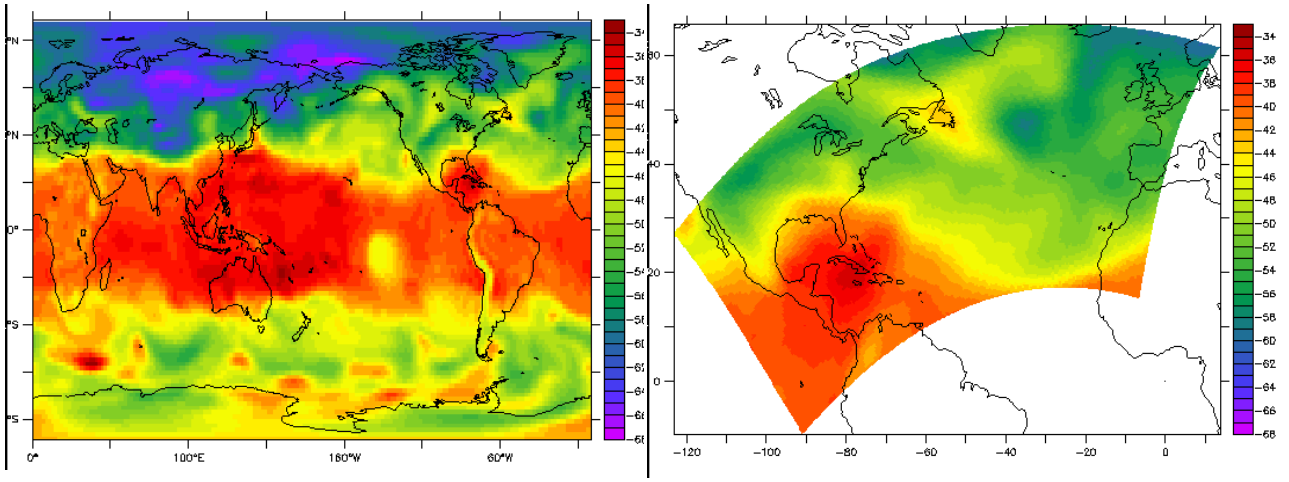


Fig 6: Temperature L28, on ECHAM grid (left) and after 4 neighbours DISTWGT interpolation on COSMO grid (right)

The standard interpolation strategy could be slightly modified to increase even more performances, and, particularly, reduce the amount or the number of MPI communications necessary to exchange the coupling fields.

The easiest improvement consists of changing the location (source or target model) where interpolations are performed, using the MAPPING option, newly offered with OASIS3-MCT. It is expected that performing this operation such that data expressed on the coarser grid (instead of data expressed on the finer grid) is exchanged will reduce the number of exchanges; as observed previously, increasing the size and limiting the number of MPI messages improve the exchange speed.

Another modification is strongly recommended to reduce event further the communication load and the coupling overhead: to pre-select only the ECHAM grid points from the region involved in the coupling, i.e. corresponding to the extension of the COSMO grid. This can be done by building a new coupling mask for ECHAM and adding it on the list of existing variables of masks.nc OASIS auxiliary file. This modification will strongly limit communications and calculations in COSMO-ECHAM exchanges as these will be performed only for ECHAM grid points inside the COSMO region.

Another way to reduce the cost of ECHAM-COSMO exchanges is to declare an ECHAM partition so that only a subset of ECHAM sub-domains, corresponding to the COSMO region, will be involved in the coupling.

Suggestion of OASIS3-MCT improvements

A set of modifications on OASIS3-MCT code were necessary to couple ECHAM and COSMO. Most of them are related to hard coded limits reached in our configuration, due to its unusual high number of coupling fields:

- maximum number of coupling field on mod_oasis_var.F90 and mod_oasis_coupler.F90
- namcouple line length on mod_oasis_namcouple.F90 and mod_oasis_kinds.F90
- number of profiling measures on mod_oasis_timers.F90

Those modifications were reported to the OASIS developers team. A future release should replace hard coded limits by dynamically allocated arrays. This release will then fully support the present COSMO-ECHAM 2 ways nesting coupling.

Some improvements should be investigated by the OASIS developers team to enhance performances of the code. In particular, on the “oasis_coupler_setup” initialization routine, which main task is to found correspondence between model declared and namcouple declared coupling fields. It appears that the routine cost increases with number of coupling fields (up to 1 minute for 285 fields). A better matching algorithm would contribute to speed up our initialization phase and avoid substantial delays at higher parallelism levels.

This support clearly demonstrates that tight 3D coupling is possible with OASIS3-MCT, which should encourage in the future more global to regional couplings. But the current OASIS setup does not really take into account a specificity of such configuration. Given that one domain is larger than the other, interpolation and communication should be limited to the surroundings of the regional area. This cannot be automatically done by OASIS and needs user's intervention.

But this intervention should be, at least, guided. The user should be informed of the non negligible slow down than an automatic creation of interpolation weights (with SCRIPR) can produce. Some advices should be given on how to build an adequate mask on the greater grid and limit exchanges to a subset of global grid sub-domains. Additionally, a tool could be provided to help the user to build this mask. Ideally, an explicit option should be implemented in OASIS so to limit the remapping to the intersection of the two grid domains.

Annex 1: COSMO/ECHAM two ways nesting coupling fields

Coupling field	Sent by
Geopotential height	ECHAM
Surface pressure	ECHAM & COSMO
Temperature (47 levels)	ECHAM & COSMO
Zonal wind (47 levels)	ECHAM & COSMO
Meridional wind (47 levels)	ECHAM & COSMO
Specific humidity (47 level)	ECHAM & COSMO
Specific liquid water content (47 levels)	ECHAM & COSMO
Specific solid water content (47 levels)	ECHAM & COSMO

Mission #11
Oct 29- Nov 23 2012

Host: Richard Hill
Laboratory: Met Office, Exeter (UK)

Main goal: Optimise OASIS-MCT performances in the Met Office high resolution configuration

Main conclusion
OASIS3-MCT implementation in the Met Office high resolution global coupled model has been shown to produce bit-reproducible simulations. Coupling interpolations go about 7 times faster than the previous OASIS3 ones. OASIS3-MCT appears to offer a suitable coupling solution, at least in the medium term, for planned Met Office high resolution models. In addition, a support has been provided to overcome starting issues in adding a wave model on the Met Office climate model.

Model / machine description

Unified Model (UM)
Met Office's global atmosphere model includes JULES soil module. Grid size: 1024x769x85, N512. Parallelism has reached 1500 MPI tasks on targeted supercomputer with more than 50% parallel efficiency.

NEMO
The widely used European ocean model is here associated in a single executable to the Los Alamos National Laboratory CICE (sea-ice) model. They share the same ORCA025 grid (1442x1021) on a 75 vertical levels configuration and are used without NEMO IO server. NEMO parallelism is constrained by load balancing with respect to the UM.

A wave model, called WaveWatch II, is about to be integrated into the OASIS coupled configuration.

Those models are available on IBM supercomputer, which has 33,792 compute cores (4 height-core 3.8 GHz Power 7 processors per node) with Host Fabric Interconnect. 160 extra nodes (Monsoon) are accessible from CERFACS for a total peak performance of 1.194 Petaflop/s. Machines are located at Met Office, Exeter.

Reproducible simulations with OASIS3-MCT

For reasons that cannot be detailed in this report[45], reproducibility is an important requirement for many laboratories, like the Met Office. Any coupled model cannot satisfy this characteristic if any one of its different components, including the coupler, does not.

For this reason, Met Office staff perform extensive regression tests on their successive coupled models to ensure that results are not unexpectedly altered when introducing new versions of the components, including the coupler. Performing the same verification with the new OASIS3-MCT coupler has been one of our first tasks.

To do so, two simple 3-day long runs were performed with two different atmospheric decompositions and their results compared.

A careful check of each component behaviour proved:

- OASIS3-MCT reproducibility, as “bfb” map strategy is set by default for any mapping interpolation used in our namcouple
- UM reproducibility, with any tested parametrization or processor arrangement
- NEMO reproducibility with -O2 optimization option, or -O3 if SOR solver is used instead of PGC. With version 3.4 of our ocean model, a -O3 option combined with the use of PGC solver should give reproducible results if pre-compilation uses the CPP key mpp_rep[46]. This still has to be checked on Met Office supercomputers.

Load balancing

The post-processing tool “lucia”, associated to OASIS for measurement of the coupled model load balancing has been installed on Met Office HPC platform, and saved in svn (FCM at the Met Office) repositories, for both OASIS3[47] and OASIS3-MCT [48].

This tool was initially developed in February[49] 2012 at SMHI (see Oasis Dedicated Users Support # 9) for OASIS3, then adapted to our new OASIS3-MCT version (and called “lucia-mct”).

Both versions require OASIS instrumentation by means of the same CPP key (“balance”) that prints MPI_Wtime based measurements in log files[50], before and after MPI send and receive actions. This key is included in the OASIS3 release, but not yet in the OASIS3-MCT one. Necessary modifications were included in the Met Office dedicated version[51].

---

[45] Even though climate simulation reproducibility is considered crucial by a major part of our community, the OASIS Dedicated Support staff still keeps on thinking that the reason why it is so is not clearly established from a scientific point of view.

From the Met Office point of view reproducibility is a very useful tool which aids development by allowing bugs to be identified more easily and reducing massively the amount of time spent on testing and validation. E.g. if results change then traceability of results in climate science is such an important issue that we have to run extensive scientific comparisons using significant computing (and staff) resources in order to demonstrate that overall scientific evolution has not been affected.

[46] following Rachid Benshila (LOCEAN) instructions.

[47] svn://fcm2/PRISM_svn/PRISM_UKMO/branches/dev/ojamil/r899_load_balance_tool /utilities/oasis3/load_balancer/lucia/lucia

[48] svn://fcm2/PRISM_svn/PRISM_UKMO/branches/dev/ojamil/r899_load_balance_tool /utilities/oasis3-mct/load_balancer/lucia/lucia-mct

[49] a few weeks after the locally popular St Lucia day

[50] “*.prt” files for OASIS3, “debug.*” files for OASIS3-MCT

[51] /data/cr/ocean/emaison/oasis3mct-cottbus on hpc2f supercomputer.

The compiling (-c option) and launching (from the working directory that includes log and namcouple files) of “lucia” and “lucia-mct” are the same but their results differ slightly.

Lucia and lucia-mct graphical output are shown on Fig 1 and 2. Pairs of timings are plotted in red (calculations) and green (waiting or coupling communication time). Where lucia produces three pairs of timing data (one per model and one for OASIS separate executable), lucia-mct only produces two, i.e. one per component which now includes the interpolation time.: what we call calculation time, for the OASIS3 executable, measures the time needed to perform interpolation(fig 1, first box). This time cannot be calculated by lucia-mct, because there is no coupler executable any more: this interpolation time is split into model calculation times. To be able to compare “lucia” and “lucia-mct” measurement, another performance tool measurement must be enabled in OASIS3-MCT (see next §).

On the Met Office IBM Power 7 machine, it has been observed that the production of ASCII output linked to the “balance” option in OASIS log file can significantly change execution time[52]. Timings are supposed to be printed at the very end of the run, which is not the case on this machine (even though no “flush” command is called). This is possibly a consequence of MPI environment variable options, that could be set up later.

For further use of our lucia tools, we emphasise that OASIS3-MCT load balancing cannot be processed with more than 2 models. It is theoretically possible, although untested, with lucia in OASIS3. Use of lucia for OASIS3 in pseudo-parallel mode is not supported either.

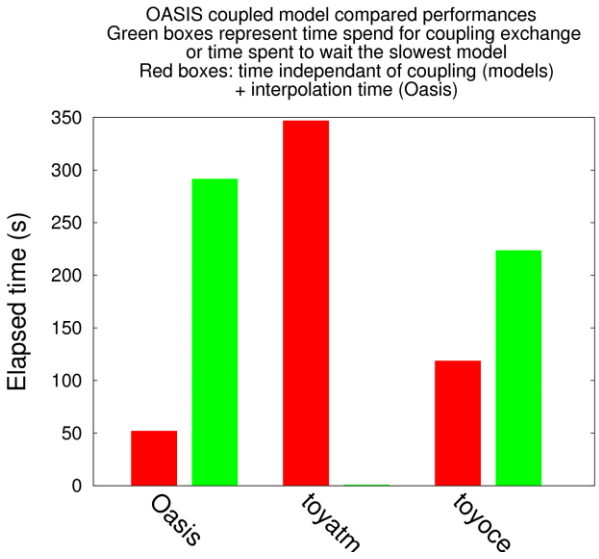


Fig1: OASIS3 coupling

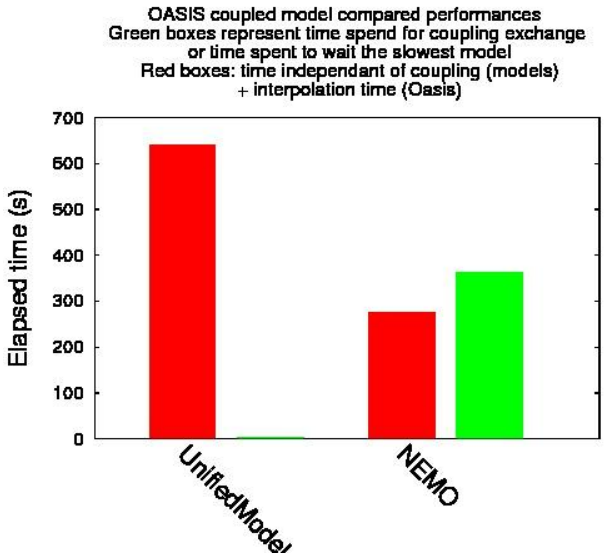


Fig2: OASIS3-MCT coupling

Load balancing of High Resolution UM/NEMO coupled model

Nevertheless, load balancing with the same high resolution UM-NEMO model[53] has been done for two versions of the coupler. In both cases, the UM atmosphere was decomposed on 20x91 partitions and NEMO on 10x12. Fig 1 shows total time for 5 exchanges, fig 2 for 12.

A clear concern is the total time needed by OASIS3 (52 seconds) to perform its interpolations,

[52] A reproducible and significant reduction of elapse time about 10% was measured in high resolution simulations. The fact that enabling timing print makes the simulations faster suggests that synchronization of OASIS communications can probably be enhanced.

[53] Set up and made available by Richard Hill and Omar Jamil

although this tends to be significantly reduced in the pseudo-parallel mode with 8 coupler instances (see next §).

Load balance could not be improved in either cases due to component model limitations in scalability and memory requirements: at this resolution, with the present optimisations, the UM cannot go faster. At the same time, NEMO cannot be decomposed on fewer than 100 sub-domains (which would slow it down) without exceeding the memory limit of a Met Office machine node.

OASIS performance measurement and comparisons

OASIS3-MCT interpolation and communication cost can also be evaluated by in-place measurement in the coupler interface library[54]. It can be enabled in released versions by setting the "TIMER_debug" variable to 1 (to obtain details for the master processor and an overall summary) or 2 (separate measurements for each MPI task) in file mod_oasis_timer.F90.

Time spent on groups of coupling operations, such as interpolations, are displayed in [model_name].timers_0000 files, for the master process. The minimum and maximum duration of all processes is also reported.

The released version of OASIS3-MCT routines at first caused deadlock when these time measurements were activated and gave wrong timing results on Met Office IBM Power7 machine. Bugs were reproduced with a high resolution toy model[55]. This toy was made available to OASIS developers, opening an external access to the "Monsoon" Met Office/NCAS part of the IBM Power 7 system. In parallel, to let us produce our measurements, a workaround was implemented in the Met Office dedicated version of the coupler.

A first analysis of OASIS3-MCT internal timing measurement, done with an N512/ORCA025 high resolution configuration, showed that interpolations (map_smat index) and send/receive operations (psnd_00x/grcv_00x) are significantly bigger than any other action. In particular, initialisation time was found one order of magnitude smaller than the total time needed to exchange coupling field during the simulation. Excluding the receive operation of the first coupling field (which gathers MPI communication and model load unbalance), communication are one order of magnitude smaller than interpolation times.

But overall, in an 8 coupling exchange simulation, which execution time, excluding restart and termination operations, is about 15 minutes, the total cost of the most significant part (interpolations) of OASIS3-MCT, i.e. 0.3 seconds, is clearly negligible. Any further coupling strategy improvement described below has no major impact on simulation execution time.

Interpolation cost was also measured with "lucia" on a similar configuration coupled through OASIS3 (see previous §). This configuration is based on OASIS3 running onto only one process and therefore differs from the most efficient pseudo-parallel case into which OASIS3 runs onto more than one process. An estimation of pseudo-parallel mode performance can be made by dividing the mono-process timing by the ratio of the total coupling field number over the maximum number of coupling fields on a single coupler process, which is about 5 in our case.

Given that the interpolation of 5 coupling exchanges with in the mono-process case takes 52 seconds (see fig 1), we can estimate that interpolation for one coupling exchange in OASIS3

[54] See OASIS3-MCT updated documentation, "Time statistics files" §).
[55] /data/cr/ocean/emaison/oasis3mct-cottbus/examples/toy_eric_pulsation

pseudo-parallel mode would take about 2.3 seconds, i.e. 7 times more than the same operation performed by OASIS3-MCT.

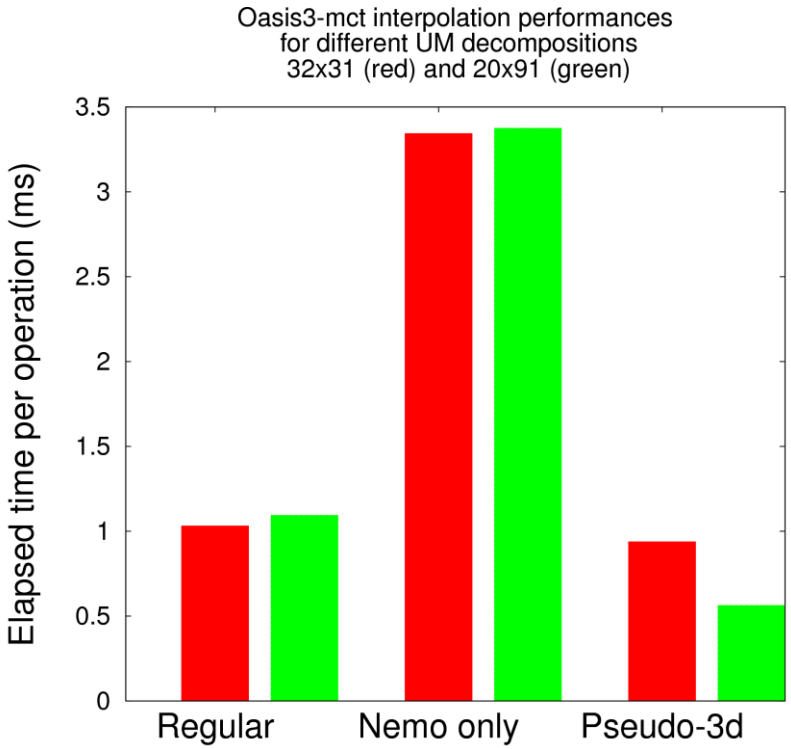


Fig 3: Single interpolation duration for 3 different coupling strategies

In the next step, we tried to further reduce OASIS3-MCT interpolation cost changing two parameters in our namcouple file.

As we know, there is no longer any dedicated coupler executable in an OASIS3-MCT coupling, which means that interpolations, like any coupling operations, are now performed by the models themselves (to be precise, by the linked OASIS3-MCT library routines). These operations can be performed either by the source or destination model processes. This can be defined by the user in the namcouple file (dst or src option of MAPPING operation, src being the default).

Since NEMO always completes its calculations before the UM, our first approach to parameter tuning consisted in performing all interpolations on the NEMO side. Different single interpolation timings (mean of 608) are shown on fig. 3, for two different atmospheric decompositions. Values for default namcouple parameters (the interpolation is done on the source model processes) are indicated by the first two boxes ("Regular"). The following two boxes show duration of interpolations all performed on the NEMO side. Their higher values can be explained by the decomposition difference between the two models: 992 (red box) or 1820 (green) for the atmosphere against 120 for the ocean. When half of the interpolations cannot be done by the most parallel model processes, performance worsens.

A new OASIS3-MCT functionality, so called "pseudo-3D", already tested successfully on the 3D regional/global coupling implemented with COSMO and ECHAM[56], allows several coupling fields sharing the same kind of interpolation to be grouped together in order to perform calculations and field exchanges in bigger arrays[57].

[56] See Oasis Dedicated User Support #10

[57] For details, see OASIS3-MCT User Guide, Appendix B (Changes between previous OASIS3.3 and new

This has been activated for different ice related coupling fields (thickness, cover, temperature …) allowing a reduction in the number of coupling fields from 58 to 22. Results illustrated by the last two boxes of fig. 3 indicate that this tuning slightly speeds up interpolation operations, but only with the higher decomposition. This suggests that it is worth increasing array sizes when UM parallelism gets above a certain level (in this case, gets higher than 992 sub-domains). Better said, we suppose that when parallelism increases, the sub-domain array size gets too small for optimal computing or that the number of MPI messages gets too numerous for the interconnect network capacity. Over some limit, it should be interesting to switch on this OASIS3-MCT pseudo-3D option. But this limit, in our case estimated to be 992 sub-domains or more, needs to be investigated more precisely.

However, the impact of these adjustments was not visible in overall simulation elapsed times. We suppose that the OASIS3-MCT extra cost will be more significant at higher spatial and time resolution, and only if model load balancing can be reached on such future configurations.

Jean-Christophe Rioual started instrumenting OASIS3-MCT with “Dr Hook” measurement routines. This will allow an alternative method of assessing and understanding the different OASIS3-MCT task costs.

Various questions about OASIS

This support also contributed to identifying difficulties and features with OASIS3-MCT and publicising to key Met Office staff some established OASIS3 solutions which may be adapted for use in new model coupling work currently planned or in progress at Met Office.

Global model

OASIS3/OASIS3-MCT upgrade caused an increase in model master processor memory requirement: at high resolution, about 88% of the available memory on the node against 76%. A memory leak has been found and fixed by Richard Hill a few weeks after the presently described OASIS user support.

OASIS3-MCT behaviour could slightly differ from the existing OASIS3. For example, it has been observed that coupling fields output with EXPOUT option are appended to output from previous runs if files pre-exist on the working directory. We suggest to overwrite them instead. The fact that, with OASIS3-MCT, files produced by the EXPOUT option contain data which is immediately viewable even if the model run ultimately crashes is seen as an extremely useful facility, compared with the equivalent situation under OASIS3 whereby such data was only visible if and when the job had completed successfully.

To conclude, even though not actually tested with UM-NEMO, we reported that OASIS3-MCT - Netcdf4 was already used on ARPEGE(or WRF)-NEMO configurations. Armed with this information, the Met Office aims to upgrade all relevant components to use Netcdf4 during 2013.

Regional models

We took the opportunity of a regional model group discussion to outline and clarify a preferred strategy for coupling regional configurations of atmosphere (with non regular grid) and ocean models. We agreed that conservative interpolation must be chosen to take into account the rapidly

varying area of atmospheric grid points. We emphasised that interpolation at boundary can be an issue or, at least, lead to difficulties (see Oasis Dedicated User Support #10). For this OASIS3 based coupling, we informed the regional model group that the “lucia” tool is installed at the Met Office (see first §) for load balancing analysis.

Wave model

After participating to an OASIS training session, François-Xavier Bocquet started an OASIS3 coupling of the WaveWatch II (WW) model with NEMO. As a first step, the ocean model has been replaced by a toy, to check WW-OASIS interface (partition and coupling field declaring, exchange order and frequency, interpolation). After an initial issue due to a non explicit variable naming on both model and OASIS sides, WW interface has been validated.

The next two steps have been defined (NEMO and UM coupling) which clearly pave the way to the first tri-model coupled configuration ever implemented at Met Office.

Mission #12
Nov 26- Dec 21 2012

Host: Anne Marie Tréguier
Laboratory: Université de Bretagne Occidentale, Brest (Brittany)

Main goal: To couple global + regional ocean/sea-ice model to atmosphere

Main conclusion
Model interfaces have been upgraded and special interpolations defined to properly exchange coupling fields between parent/child ocean and atmosphere models. A load balancing analysis of the coupled configuration has been performed during a one month long validation run.

Model / machine description

ARPEGE
Météo-France's AGCM is used in its stretched version (rotated pole centred on Baffin Bay, refined grid around North Atlantic area with stretched factor of 2.5). Grid size: T127 truncation (24,572 grid points), 31 vertical levels, 5.4 version, including SURFEX land model.

NEMO
The popular European ocean model (3.4 version) is here associated to LIM (sea-ice) version 2. This coupled model particularity lies in an so called AGRIF[58] ocean zoom, centred in the North Atlantic region (ERNA[59] configuration). NEMO is defined on 2 different grids: the global (or parent), ORCA05 with 722x511x64 grid points and the regional (or child), 1/8 degree, 724x632x64 grid points. The same computing resources (MPI processes) are used to perform sequentially the computations on the parent and child grids[60]. Boundary conditions of the child grid are provided by the parent grid (through interpolations independent from OASIS). In the other direction, parent grid points belonging to the child area are updated by child calculations at each parent time step.

Before starting the User Support period, those models had been ported on "vargas" IBM supercomputer, 3,584 compute cores (112 thirty-two-cores 4.7 GHz Power 6 processors per node), Infiniband x4DDR Interconnect. Total peak performance of 67.3 Teraflop/s. This machine is located at IDRIS CNRS computing centre, France.

[58] Adaptive Grid Refinement In Fortran, http://www-ljk.imag.fr/MOISE/AGRIF/
[59] Eddy Resolving North Atlantic
[60] In our case, after one time step on parent grid (2160 sec), three time step are performed on the child one (3x720s)

Model interfaces

A preliminary step dealt with an upgrade of the existing model interfaces to allow ocean - atmosphere field exchanges not only with the global ocean but also with its regional zoom.

In NEMO, we mainly relied on an existing implementation provided by LOCEAN experts, developed for a WRF-NEMO high resolution coupling on a tropical belt area (45S-45N)[61]. Our contribution consisted in allowing atmosphere-ocean exchanges from and to the regional zoom in both ways, including for ice related coupling fields.

Implementation principle is relatively simple. As any other NEMO routine in an AGRIF configuration, each OASIS primitive can be called twice, first by the parent and then by the child). Coupling implementation only consisted, as explained in more details below, in choosing to call them twice or once and, in this case, from the parent or from the child.

At initialisation, our sequence of OASIS primitive calls is:

INIT_COMP	GET_LOCALCOMM	DEF_PARTITION	DEF_VAR	def_partition	def_var	enddef

In the time step loop:

PRISM_GET	prism_get	prism_get	prism_get	PRISM_PUT	prism_put	prism_put	prism_put

And at the end:

TERMINATE

Calls from parent grid are shaded in blue and upper-case, from child in red and lower-case. Parent and child share the same MPI process, which means that init_comp, enddef and terminate, that must be called only once per MPI process, must be invoked by one or the other. At the opposite, as parent and child do not have the same grid (even with different dimensions), they both must call def_partition. Consequently, parent and child coupling fields[62] must be also declared separately. As we have chosen to receive and send fields on both grids, prism_get and prism_put primitives are called at each time step of parent and child components[63].

For further extension to configurations including more than one AGRIF zoom, it will be necessary to limit enddef call to the last child only.

Two corrections of the NEMO release version, both related to sea-ice coupling, were necessary (and reported to the NEMO system team): fr1_i0 and fr2_i0 arrays initialisation must be done before (on lim_sbc_init_2) sbc_cpl_init call and dynamical allocation of wind stress on sea-ice must be done in any case, even if this coupling field is not explicitly provided by the atmosphere model.

Both OASIS3 and OASIS3-MCT interfaces have been updated (and validated) at the same time, considering that they only differ on a few CPP key controlled lines (key_oasis3 and key_oasis_mct).

---

[61] PULSATION project (ANR-11-MONU-0010), http://www.agence-nationale-recherche.fr/en/anr-funded-project/?tx_lwmsuivibilan_pi2[CODE]=ANR-11-MONU-0010

[62] In our case, parent or child coupling fields quantities are the same but one could imagine to exchange a different number, or even different coupling fields, from parent and grid child. This is already possible, just changing coupling field configuration on the NEMO parent or child namelist parameter files

[63] According to namcouple defined coupling frequency, MPI coupling exchanges are triggered only on a subset of model time step (see OASIS3-MCT documentation, “Sending a coupling (or I/O) field”). In our case, both parent and child coupling fields are exchanged at the same dates.

Fewer corrections were mandatory on the atmosphere routines, all related to the duplication of outgoing coupling fields, bound to ocean child grid. As soon as an OASIS3-MCT coupling will be possible, according to Météo-France schedule, those modifications will not be necessary any more: our new coupler is supposed to be able to send one coupling field to two different targets, performing different interpolations to different grids, which is exactly what is needed here.

Interpolations

Grids description

OASIS auxiliary files gathering all grid descriptions must be built up first. In NEMO, the information about the parent and child grids was deduced from the “meshmask” output file previously produced by an ERNA stand alone simulation. ARPEGE related data were provided by CNRM from previous coupled configurations.

A first issue appeared when considering the limit between parent and child regions. Child grid consists in two different zones: the main inner 1/8 degree part and a buffer zone, which is usually filled with interpolated information coming from the parent grid.

As suggested by LOCEAN experts, we decided to include the field values coming from this buffer zone into the data sent by the child ocean to the atmosphere: smoothed SST and other ocean/ice coupled quantities coming from this transition zone between parent and child regions are then provided to the atmosphere. We will check that this technique leads to avoid artificial gradient source in this zone[64]. In the other direction, fluxes coming from the atmosphere will also be interpolated on this boundary zone.

Global interpolations

On a second step, interpolations weights for coupling the atmospheric grid with the parent ocean grid must be calculated. Due to the ARPEGE stretched grid characteristics (gaussian grid, with less grid points near poles, combined with a stretched coefficient), meshes have extremely different areas from one pole to the other. SCRIP 'CONSERV' is the only standard OASIS interpolation associating to the target points a different number of source neighbours according to target/source mesh areas ratio and, consequently, to provide relevant source information to any target grid point when this ratio is much bigger than 1.

Unfortunately, SCRIP algorithms are not able to perform 'CONSERV' weights calculations with ARPEGE stretched grid and a new kind of interpolation has to be developed to give a satisfactory solution to this problem. In the meantime, a simple 4 neighbours Gaussian interpolation was preferred, to give more accurate results near the zoomed area than at the antipode. Weights were automatically produced by OASIS at first simulation (taking a few minutes on our supercomputer).

[64] This coupled model configuration is precisely set up to study the impact of physical gradients. That’s why it is particularly crucial to avoid artificial numerically-created gradients at the limit of the zoom.

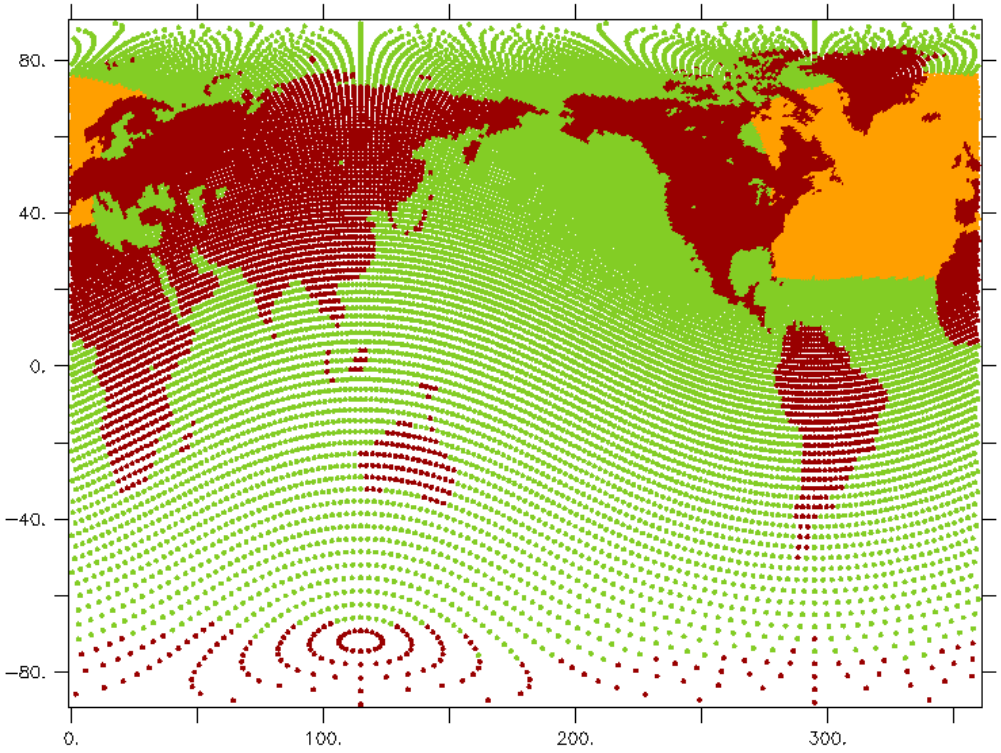


Fig 1: ARPEGE T127-stretched grid, with land mask (brown) and points interpolated from NEMO-AGRIF parent (green) and child grid (orange)

Regional interpolations

The same Gaussian interpolation was chosen to provide information from/to the ocean child grid. As mentioned during previous Dedicated User Supports, OASIS automatic weight calculation was not specially designed for regional coupling. In particular, its algorithm will try to entirely fill a global grid with the information coming from a regional grid. On the other direction (from global to regional), depending on interpolation choice and masking configuration, source grid points located at large distance of regional grid boundaries could be used in calculating coupling field values send to the target grid. It is then strongly recommended to take care on how OASIS calculates its interpolation weight automatically and to modify its behaviour when necessary.

The strategy described below is similar to the one already used with COSMO/CLM and COSMO/ECHAM coupling. OASIS is launched first with initial grid masks. Interpolated fields are then post-processed to better define a mask that reduce exchanges to a smaller region of the global grid.

From the atmosphere (global) to the child ocean (regional), all atmosphere points can be used by the automatic OASIS weight calculation. The closest 4 atmosphere neighbours are associated to each child ocean point. Considering that the target/source grid point area ratio in this region is not too far from 1, far remote atmosphere points are not used at the boundary of the regional grid.

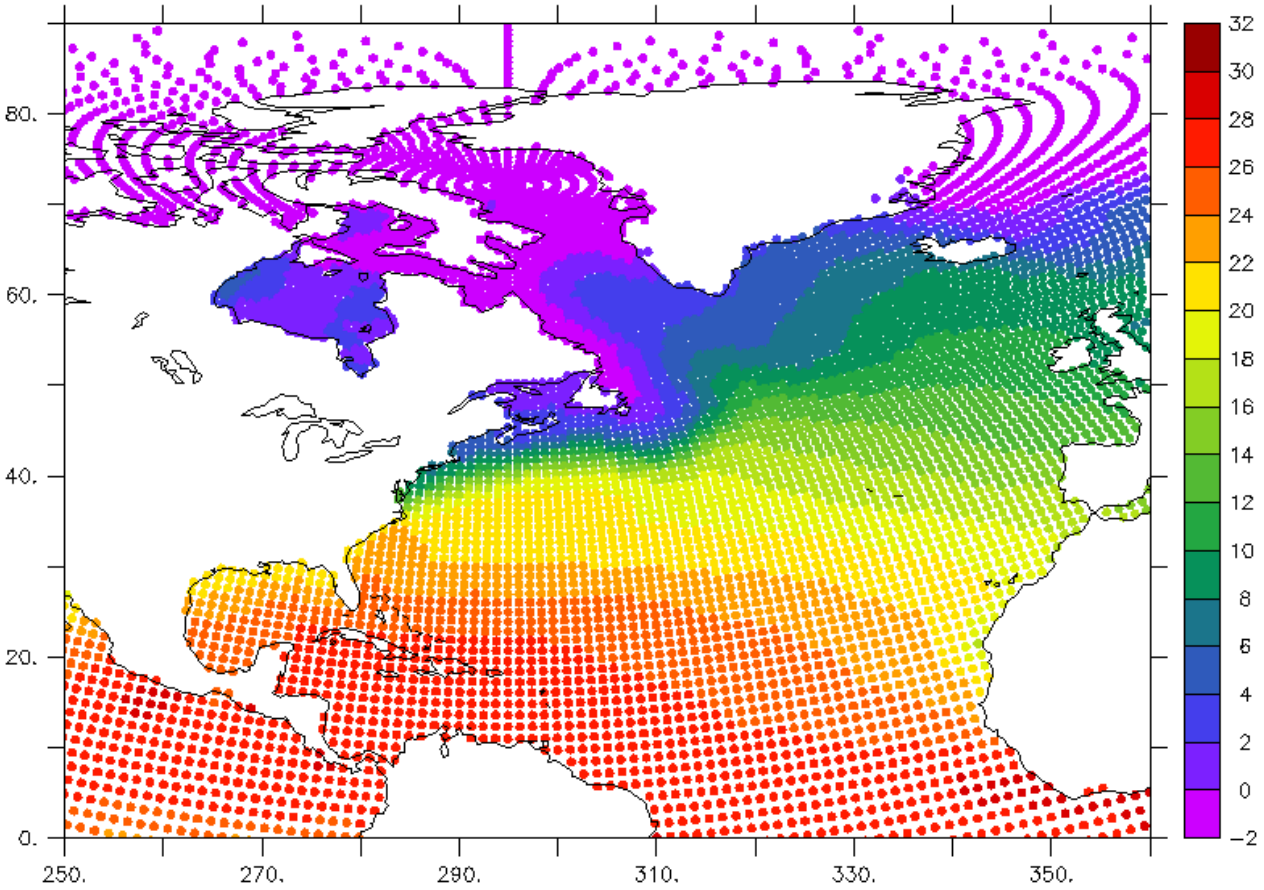


Fig 2: Example of combination on ARPEGE T127-stretched grid of 3h mean SST coupling fields interpolated from NEMO parent and child grids

But in the other direction, OASIS will automatically fill all global atmosphere point with information coming from the ocean zoom region, unless an appropriate mask is defined for the target grid. Different operations must be performed to avoid such problem:

1. A special mask is temporary defined on the child ocean grid: all grid points are unmasked, except those from first and last lines (and columns)
2. A bilinear interpolation is calculated identifying 4 source neighbours (2 in each directions) per target grid point[65]
3. A uniform coupling field is send to OASIS
4. The interpolated field[66] is equal to the source constant value where atmosphere grid points have 4 neighbours on child ocean grid. They are equal to zero on atmosphere masked points or on all grid points slightly outside the limit defined in (1).

We used this information to build two new atmosphere masks as shown on Fig 1. In addition to the original mask (masked points in brown) covering the global domain:

- a mask defining only atmosphere sea points filled by interpolated quantities coming from the child ocean grid (unmasked points in yellow)

[65] OASIS3 bilinear interpolation was slightly modified before as the original algorithm would have used, for the atmosphere points falling outside the regional ocean domain, the non masked points among the 4 nearest neighbours in the regional domain and would have given them a value.

[66] Produced with a restart field with OASIS3 interpolator mode or on an output field with EXPOUT namcouple option

- a mask defining only atmosphere sea points filled by interpolated quantities coming from the parent ocean grid (unmasked points in green)

With a simple OASIS3 BLASNEW operation[67], it is now possible to rebuild a complete global coupling field combining information from the child and parent grids[68] (see example on Fig 2)

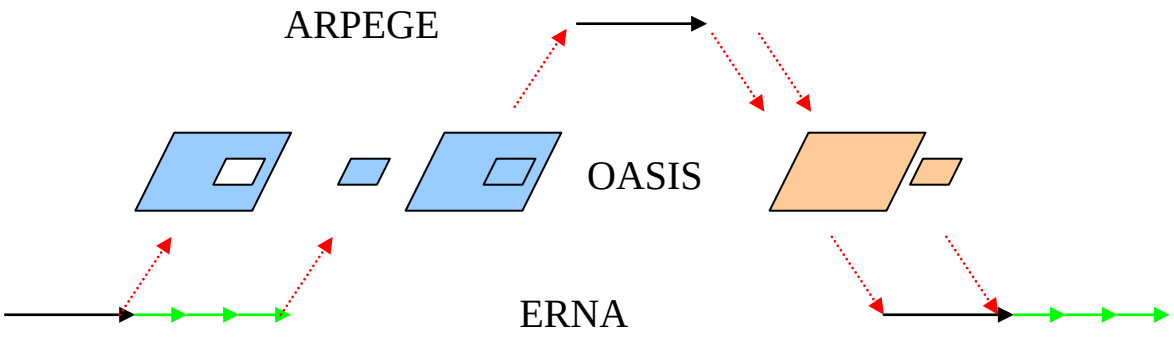


Fig 3: Coupling sequence. Child ocean time steps are represented in green, regular time step in black, OASIS exchanges in red

Namcouple parameter file must be defined accordingly to this new strategy:

- all original coupling fields must now be exchanged twice (in the two directions). Our naming convention follows the NEMO-AGRIF one: the first letter of the global field is replace by the number of the child ocean grid (here 1)
- each child coupling field coming from ocean is added to the corresponding field coming from the parent grid (with BLASNEW option)
- Coupling frequency must be the same on child and parent grid[69].

Parent/child grid re-sticking impact

To evaluate the quality of our interpolated field gathering parent and child grids information, we have compared it, on a 3 hours mean surface temperature (sea/ice mean), with the field coming from the parent grid only (including on the child covered area) and interpolated on the whole atmosphere grid.

Differences are shown on Fig 4. As wished, it clearly appears that no artificial gradient is created at AGRIF zoom domain boundary. One can also noticed that main differences are located on high gradient areas, like Gulf Stream or ice field boundary.

[67] This operation is not yet available on OASIS3-MCT

[68] No modification is then needed in the atmosphere interface compared to when only global field is received

[69] Be careful that LAG value must match model time step of the model from where coupling field is coming (child/parent ocean, or atmosphere, see sequence on Fig 3)

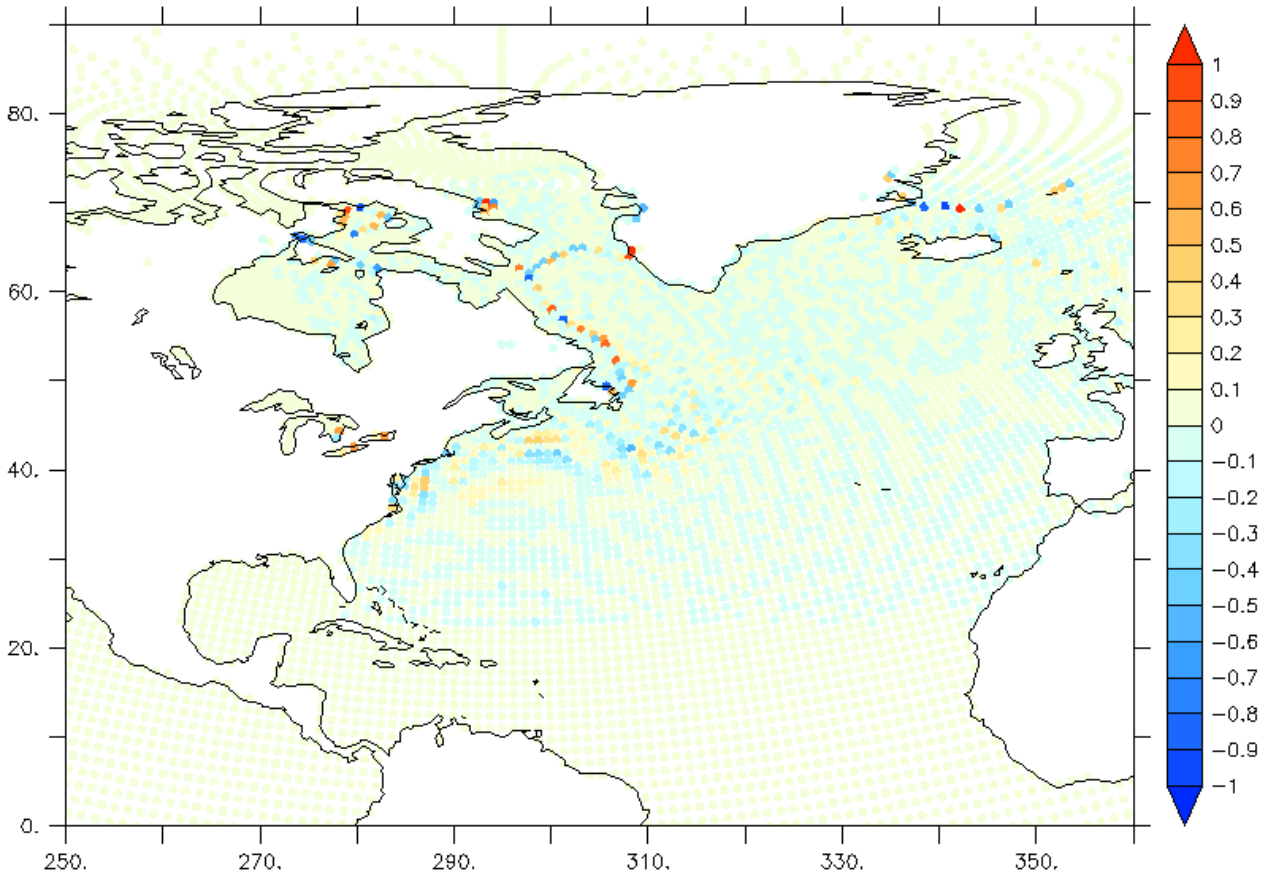


Fig 4: Difference on ARPEGE T127-stretched grid between 3h mean surface temperature field interpolated from parent grid and combination (same as fig 3) of SST fields interpolated from parent and child grid

Validation on model

Due to a lack of time, it was not possible to set up the targeted configuration including stretched gridded atmosphere model. A simple ARPEGE configuration on regular T127 grid (and corresponding OASIS auxiliary and interpolation files) was used to test our coupling.

A one month long run was performed. It still exhibits a pre-existing issue of unrealistic ice temperature biases. Consequently, it was unfortunately not possible to fully certify integrity of our interpolations and interface implementations.

Nevertheless, this regular T127 configuration includes the same grid point number than the stretched one. The load balancing analysis we produced (Fig 5) with the previously developed “lucia” tool[70] will then be equivalent on the targeted configuration[71].

If the main information it shows is robust (speed ratio between ARPEGE on 10 cores and ERNA on 127 is about 3), some information such as OASIS calculation time must be corrected. Our measurement algorithm did not anticipate that during BLASNEW operation (coupling field combination), OASIS has to wait information coming from the child ocean. Consequently, the

[70] For further information on Lucia load balancing analysis, see Dedicated User Support #9
[71] Except if atmosphere time step is different when grid is stretched

OASIS “interpolation time” (Oasis related red box) includes, in our case, the waiting time, at each coupling time step, between parent and child field receiving. Corrected from this bias, the OASIS3 interpolation and exchange times can be considered as negligible compared to the total simulation time.

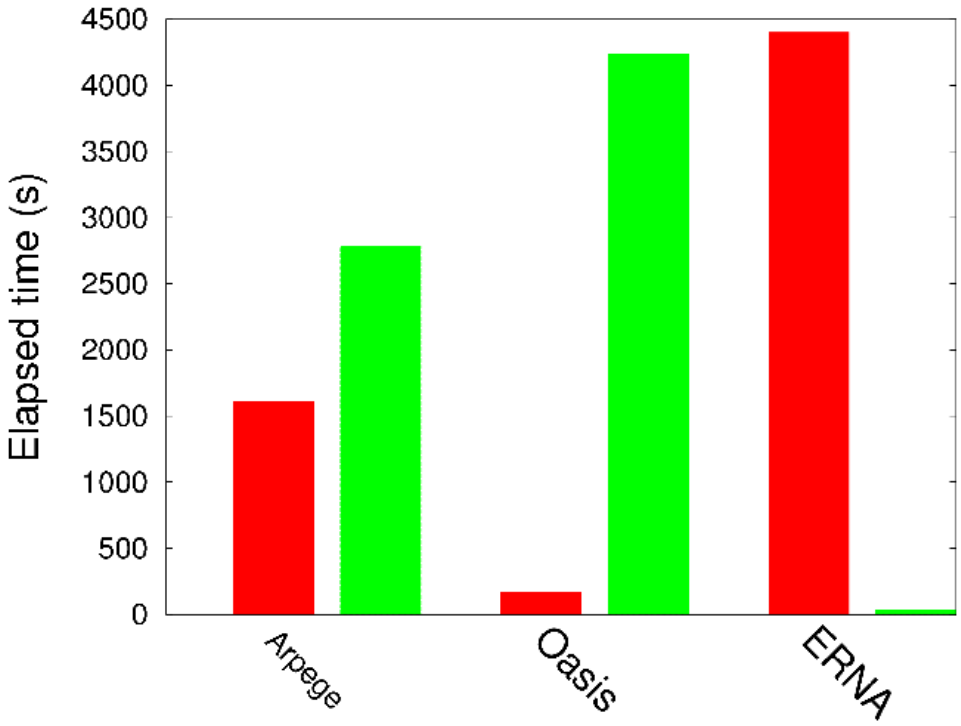


Fig 5: “Lucia” load balancing analysis performed on 26 exchanges (one per day)
ARPEGE T127 on 10 cores, ERNA on 127, OASIS3 on 1