IS-ENES WP4

OASIS Dedicated User Support 2009
Annual report

E. Maisonnave, S. Valcke

TR/CMGC/10/26

Within the IS-ENES Work-package 4, CERFACS proposed to provide personal help to implement a new coupled model or improve an existing configuration based on OASIS.

This document summarizes the results of the 3 one month long missions completed during the 2009-2010 winter in the French, German and Swedish laboratories of Institut Pierre-Simon Laplace (IPSL), Alfred Wegener Institute für Polar und Meeresforschung (AWI) and Sveriges Meteorologiska och Hydrologiska Institut (SMHI).

The last mission was devoted to improve an existing OASIS3 coupling (EC-Earth), and the first two clearly focussed on implementing a new kind of coupling using OASIS4 on a frontier configuration (Sintex ECHAM-NEMO high resolution) and with an innovative mesh (ECHAM-FEOM).

We detail in the chapters below the technical developments achieved for the 3 targeted configurations. If the OASIS3 related works ended in a satisfying way, we obviously were short of time to complete OASIS4 configurations and, in particular, to process reference simulations in order to validate our developments.

In particular, if reference version of NEMO.v3.3 or EC-Earth.v3 seems to integrate our OASIS3 improvements, it will be very difficult to integrate our OASIS4 modifications in NEMO.v3.3 or ECHAM.v6. Concerning FEOM, our very preliminary work should be maintained in the future versions of the model after being tested.

Even if our practical work doesn't need to be fully integrated to reference versions to be useful to granted laboratories, durability of the work will not be ensured otherwise.

It is clear that the next missions duration will have to be extended to take into account the difficulties encountered implementing the new OASIS4 functions in climate models.

The exchanges with our guests were also very fruitful from the OASIS developers point of view.

The OASIS Users Support Team developed his comprehension of the new requirements of version 4 and improved his capacity to replace the version 3 primitive calls by version 4 ones in the structure of different models (ocean and atmosphere).

The OASIS developers received different requests from the support

team and attempted to process real-time modifications on the coupler code. Several OASIS4 problems have been identified (for example on restart read/write) thanks to its use on demanding configurations.

During those 3 months mainly spent with the OASIS users granted by the IS-ENES program, other scientists have been informed of our code coupling activity (a seminar on this topic have been organized by Kerstin Fieg at AWI), even from other laboratories: LOCEAN NEMO system team, MPG ECHAM code developers, Stockholm University researchers, NSC engineers and support technicians from various supercomputing centers.

To conclude, the OASIS Users Support Team wants to thank warmly their guests. Their presence was one of considerable poetry. From an accounting point of view, we have to emphasize that they addressed all the local administrative issues (office, lodging ...) and greatly facilitated the organization of our activities.

Estimated carbon emission diagnostic for 3 travels by traditional means of transport: 400Kg

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 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

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

Mission #2
Oct 26- Nov 20 2009

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

Main goal: ECHAM-FEOM coupling with OASIS4

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 commuication 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.

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

Mission #3
Feb 1- Feb 19 2010

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

Main goal: Increase EC-Earth performances using OASIS3

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)

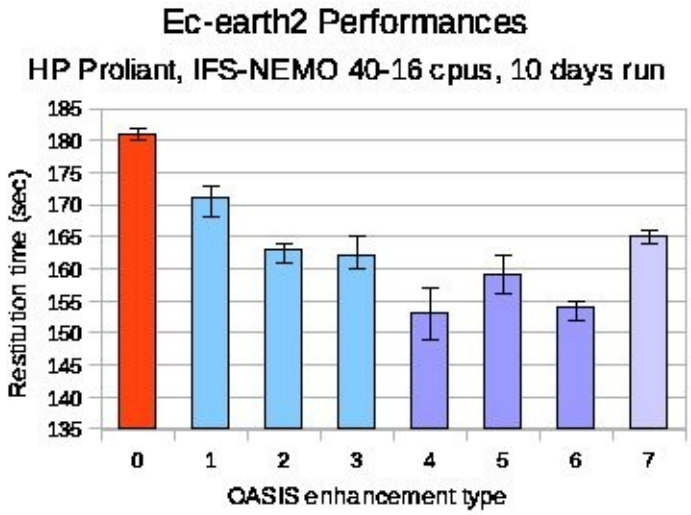
In a second step, we considered a simple sequencing of the coupling fields within the namcouple: firstly, ocean-to-atmosphere fields have to be sent by the (fast) ocean, interpolated by OASIS and then made available for the atmosphere, so that as soon as it is ready, the atmosphere receives these fields, 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-toatmosphere 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 (Bufferized send used), but with a sequentiality of 2, no significative 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 bufferized 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)

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-parallelversion