OASIS Developer Meeting
October 12th -14th , 2009, DKRZ, Hamburg
S.Valcke
CERFACS Working Notes WN-CMGC-09-154

Participants:

DKRZ: Joachim Biercamp, Moritz Hanke, Kerstin Ronneberger
MPI-M: René Redler, Mathis Rosenhauer
AWI: Kerstin Fieg
CERFACS: Sophie Valcke

OASIS4 collaborative development procedures

- Implication of the different partners

- CERFACS:
  - Two persons working full time on OASIS at CERFACS, Sophie and Laure (CNRS engineer). An important amount of work still goes on OASIS3, especially as all OASIS users in CMIP5 will use OASIS3 and not OASIS4.
  - Jean Latour will be hired for 2 years and another engineer for 18 months during the IS-ENES project.
  - Jean-Marie Epitalon will also be hired as a consultant to finalize the GUI and other XML related tasks.

- DKRZ:
  - New partner in the collaboration on OASIS development, currently in the framework of IS-ENES with 36 pm (Moritz Hanke).
  - One of DKRZ task is to provide software support and development for German climate modellers groups (MPI-M and others).
  - Implication on the long term in the OASIS development is not clear; it will depend on the needs of the climate modeller groups (MPI-M, AWI and others).

-MPI-M:
  - René's tasks will include transfer of knowledge on OASIS4 to Moritz at DKRZ and some OASIS4 development depending on MPI-M needs.
  - At MPI-M: current coupled models are:
    - ECHAM T63 –MPI-OM 1 deg (OASIS3)
    - ECHAM T159 – MPI-OM 0.4 deg, on O(100) pes (switch to OASIS3 para by René); will be used for CMIP5; then test with OASIS4 by René
  - At MPI-M, it is planned to develop a high-resolution coupled model for the STORM project (based on the CMIP5 models): ECHAM ~T255 or ~T311 – MPI-OM with OASIS3, then OASIS3 para, then OASIS4 (O(1000) pes)). Planned for 2010.
  - On the longer term (5 year and +), the HICOM coupled model with ocean and atmosphere using icosahedral grids will be developed; the grids will maybe have a different resolution, and so interpolation will be needed; it is however not clear if this coupled model will use OASIS4 or not; an integrated system might be a preferred option.

-AWI:
  - Some groups in AWI use some COSMOS configuration based on OASIS3
  - AWI is also involved in the STORM project.

- Besides this, 3 projects will include use of OASIS4: ScalES (coupling of ECHAM5 to finite element –unstructured- FEOM ocean model), the integration of FEOM back in COSMOS (with OASIS4 supporting finite element grids in 2D – with only one vertical level), and THAURUS (decadal prediction project with ScalES coupled model, currently submitted to BMBF for funding).

- BoM from Melbourne:
  - the Bureau of Meteorology from Melbourne seems very interested in interacting on OASIS4 development
  - We welcome this interaction and will make sure to integrate back their developments and bug fixes, if any.

- ICE-INFRA FP7 proposal
  - Sophie has been contacted by CSC from Finland to possibly contribute to a proposal that will be submitted to the EU INFRA-2010-1.2.2 call on the "Integration of Ice-sheet Models into Climate and Earth System Models". For this project, coupling of dynamic unstructured grids seems to be required. This comes somewhat too soon as the main objective with IS-ENES will be to stabilize the current developments. Of course, CSC and other project partners will be welcome to use and implement new functionality in OASIS4.
- Organisation of an OASIS4 training course:
  - The idea to organize a one-week training course on OASIS4 sources and structure for a reduced number of people (4 or 5) is discussed. René and Moritz will discuss this possibility in more detail. Hubert could also be involved. It would take place in the first half of 2010. This could be an important step in a transfer of expertise on OASIS4 to DKRZ and CERFACS.
- Use of e-mail exchanges, SVN, TRAC tickets
  - For e-mail exchanges on OASIS4 development, it is agreed to always state clearly in the subject or in first line who is directly concerned with the mail and who should take related actions.
  - For e-mail exchanges, it is agreed to treat only one issue per e-mail/ticket and to use a significant subject for each email; if any, the TRAC ticket number should be mentioned in the e-mail subject (and corresponding SVN check in).
  - The interaction on OASIS4 development will be done by mail for daily developments; the mail should be sent to whoever is interested in the development. René, Moritz, Kerstin, Sophie are interested to be in copy of all mails if the first rule above is respected (Sophie to ask Hubert, Jean, Laure, Justin if they are interested too).
  - If an issue lasts for more than few e-mails, a TRAC ticket should be opened and a summary of the discussion should be regularly posted on the ticket.
  - Important developments should be done on separate branches (userdef by Jean, unstructured grids by Kerstin, etc.) and merged regularly to avoid diverging developments.
  - Sophie to give writing access to SVN + TRAC to Kerstin and Moritz and to ensure that oasis4_developers list receives TRAC log messages.
- Other general notes:
  - GATEWAYS project has been funded as a Marie Currie project. This project will develop a coupled ocean-atmosphere model including NEMO with AGRIF nested part over the Agulhas region. It is not clear if the coupler will be OASIS3 or OASIS4.

- Sophie: Add a note on the documentation on a problem with pgcc 8.0.5, 9.0.4 when compiling C routine for XML reading (include of libxml parser.h)
- Partial MPI2 implementations on 64-bit machines:

  René has provided a workaround for partial MPI2 implementations on 64 bits machines. To use this workaround implemented in revision 2101, one has to compile with the use_MPI2 and DONT_HAVE_STDMPI2 CPP keys but with "not_spawn" in the XML SCC file. The workaround:

  - provides the dummy MPI_Comm_Spawn routine
  - ensures that MPI_Comm_spawn is not used
  - ensures that MPI_Finalized, MPI_Allreduce with MPI_IN_PLACE as first argument, and MPI_Waitall with MPI_STATUSES_IGNORE as 3rd argument will not be used either

  All other MPI2 routines are effectively called (e.g. other MPI2 routines are in fact provided on the IBM power6).

  Note: Using MPI1 32 bits addressing on 64-bit machines will not necessarily work even if the application is less than 2GB because addresses do not necessarily start at 0.

  Note: Kerstin tells us that IBM announced that a full MPI2 release to be available in September.

- Interaction with DEISA:
  - DEISA platforms include IBM power6 in Garching, Blue Gene in Garching, CRAY XT5 in Edinburgh; Mare Nostrum is also in DEISA.
  - Sophie to ask Laure and Eric about their plans to test coupled applications on DEISA machines.
  - It is agreed that it would be useful to invite IS-ENES participants to dedicated workshop during DEISA-PRACE symposium in Barcelona in May to discuss the test of coupled applications on DEISA platforms and also the identification of a coupled benchmark in PRACE.
- Sophie to ask Thierry Morel or Michel Valin about the bypass of loadlever regarding mixed openMP-MPI parallelisation for Mathis.

On-going developments and development planned during IS-ENES

Different OASIS4 developments were discussed. A priority (Px) was identified for each task and the tasks with P1 were assigned to one or more developers were. The different priorities mean the following:

- P1 the task will be addressed during IS-ENES
- P2 the task will be addressed during IS-ENES if time permits
- P3 the task relates to an interesting issue but will not be addressed during IS-ENES
- P4 the task is dropped

Sophie will review the wiki pages and TRAC tickets based on these discussions.

On a general basis, it was also discussed if we should offer more options (in particular to address ill-based problems) to the users (with the risk that they do not really realize what they are doing when they use them) or less options forcing them to explicitly address their specific problems. As a first step, Sophie will ask OASIS3 users about their namcouple to see what are the OASIS3 specific functionalities are really used and to help give a priority to the OASIS4 developments. In particular, Sophie will investigate which coupled models have non matching sea-land masks and need forced global conservation.

- Validation, optimisation and tests
  - (P1) Full validation of 2D global parallel search (for point-based and cell-based algorithms, for reglonlatvert, irrlonlat_regvrt, gaussreduced_regvrt grids):
    - (P1, Laure, René, Hubert) CICLE remaining problems
    - (P1, Moritz, René) Additional off-line tests for the ECHAM-MPI-OM coupled set-up with realistic subblock partitioning
  - (P1) Analysis of OASIS4 PSMILe and T performance and scalability for a high number of processes (centralisation of results on a wiki page)
    - (P1, Kerstin et al) performance analysis of ECHAM-FEOM within SCALes
    - (P1, René et al) performance analysis of ECHAM-MPIOM within STORM project (based on the CMIP5 models): ECHAM ~T255 or ~T311 – MPI-OM 0.4 deg with OASIS3, then OASIS3 para, then OASIS4 (O(1000) pes)).
    - (P1, Eric & Laure) ARPEGE T359 – ORCA ¼deg, currently with OASIS3 para (adaptation to OASIS4 planned in 2010)
  - (P1, Laure & Sophie) Comparison of OASIS3 and OASIS4 interpolation results and analysis of reproducibility.
  - (P1, Laure & Mathis) : Development of a suite of automatic tests

    This suite would automatically be run on a regular basis on different platforms and their results would automatically be analysed (e.g. to check if a modification has no side effect); candidate platforms are blizzard IBM power6, tornado opteron PC cluster, yuki SX9, AWI SX8R). See with open source build BOT automated testing system (server and clients) based on python that will be installed at MPI-M and used for COSMOS (tornado and blizzard); this could be used for OASIS itself too; contact at MPI-M on this issue is Monika Esch.
  - (P2) 3D interpolations:
    - (P2, Sophie & Laure) Validation of current 3D and 2D1D interpolations (see also TRAC ticket #13)
    - (P2, Sophie & Laure) Comparison of nneighbour2d and nneighbour3d for 3D degenerated grids with one vertical level only (TRAC ticket 15)

- Bug fixing
  - (P1, Sophie) PSMILe-Transformer synchronisation (TRAC ticket #9)
  - (P1, René and Hubert) Treatment of periodicity in i and j (TRAC ticket 23)
    - Currently in the code, “cyclicity” means periodicity is the j dimension and “periodicity” is intended in the i dimension.
    - Periodicity in j should not be supported as this is not applicable on the sphere. However, current code treating periodicity in j should not be removed but simply hidden under a special CPP key (for possible later use in Cartesian domains). Following a problem reported by ECMWF, a fix was implemented by René; this fix needs clean up.
    - Periodicity of the global grid in i is supported; whether the grid is periodic in i or not must be defined by the user in the SMIOC XML file.
  - (P1, Laure) Support of changing model timestep
    - Check changing model timestep is effectively supported. There should not be any particular problem with besides for accumulation and average
  - (P1, Jean) Support of sub blocks for gridless grid
    - This was observed by Jean not to work

        - Jean will produce a test case reproducing the problem
    - (P2) Support of run not covering a finite number of coupling timestep
        - Currently, the time axis of the exchanges is constructed supposing that each run covers a finite number of coupling timesteps. If this is not the case, some information about the end time of a run would need to be stored at the end of the run (together with e.g. some accumulation or averaged data) and read in at the beginning of the subsequent run.
        - Implication for coupling restart (with constant coupling frequency) (TRAC ticket #20);
            - Currently, the coupling restarts are supported only if the run covers a finite number of coupling periods.
            - Restart should written only if the put is the last active of the run and if the time of the put + lag > end of the run (for now do not take into account lag > 1 coupling period where restart would have two temporal instance – write would be OK, read maybe not)
        - Implication for accumulation and averaging operation: currently not supported, needs to be implemented (information will then have to be transferred from one run to the other)
    - (P2) Bug in running more than two components in parallel into one application (TRAC ticket #49). The problem seems to be related to the set up of an internal communicator in mpp_io (note: pe_list is the list of processes included in the communicator)
    - (P2) PSMILe I/O bugs (see TRAC tickets #18, #24, #42)

- Improvements of current functionality
    - (P1, Moritz & René) nneighbour option for 2D conservative remapping (TRAC ticket #32)
        - For a target cell not intersecting any source cell, finding the closest source cell or the cell of the source point closest to the target point in the target cell is not easy because the information about the point is currently not available in the cell based search
        - Extrapolation on the target side would imply additional calculation; extrapolation on the source side would maybe be better because parallel calculations already done on the source side
    - (P1, Jean-Marie) Analysis of the efficiency of XML ingestion in the initialisation phase
    - (P1, Sophie & Jean) Review coding of Transformer interpolation routines
    - (P1, Sophie) Change masks in Transformer from integers to logical (TRAC ticket #20)
    - (P1, Sophie) For I/O, check time operation below the prism_get and related returned info code.
    - (P1, Sophie) Improve documentation on interpolation schemes (1)
    - (P1, Sophie) Provide database of users, FAQs, CPU statistics on the web/wiki, forum, on-line tutorial
    - (P3) Implement more SCC and SMIOC validity checks in Driver (TRAC ticket 20)
    - (P4) Support of more than one temporal instance of a coupling field in a restart.
    - (P4) Timestamp on coupling fields within OASIS4 PSMILe and T

        - We will not implement this as we currently impose and check that the date_bounds of the prism_put/prism_get calls cover exactly the whole run duration without any gap and any overlap.

- Major improvements and new functionality
    - (P1, Kerstin & Sophie) Connectivity implementation and support of finite-element grids
        - Connectivity of the grid points and cells is needed for proper interpolation near the pole (when the grid extends to the pole), for optimisation of the global search, and for interpolation from finite-element grids
        - Currently, the prism_def_partition gives some connectivity information but only for reglonlatvrt grids for the interior of the global grid domain; for irrlonlat_regvrt, this is not even always the case because some connected cells can be arbitrarily stretched and the connectivity should tell that such a neighbour cell is in fact invalid.
        - In general, connectivity at the edge of the global domain is required e.g. for proper interpolation near the pole. For gaussred and reglonlat grids, this information can automatically be deduced; this was implemented done for gaussred grids for bicubic and bilinear interpolation for the upper and lower edges of the grid domain, therefore for these interpolations near the N and S pole (see TRAC ticket #XXXX), although it has not been fully validated for the bilinear interpolation. For reglonlat and irrlonlat grids, let's wait for a proper declaration of the connectivity before improving the bilinear and bicubic interpolation near the poles.
        - For the global search, the connectivity between the partitions is needed to search for missing source neighbours for the target points falling near the border of a partition domain. Currently, when the source grid is gaussred, the list of missing points is exchanged between all processes. For other grids, the geographical information about the envelope of the partition exchanged at the beginning is used to pre identify the processes that are likely to contain the missing points.
        - The support of finite (triangular) element grids is planned within 2 years in SCAles. For such finite elements, the fields (e.g. temperature) are defined on the node and there is a grid function for the value of the field everywhere on the triangle. An index is associated to each triangle.
        - For the point based search on such finite-element grids, the connectivity is needed for the mapping of the finite-element to the auxiliary regular grid. Then the multigrid algorithm could be used on the auxiliary grid. Once the cell of the auxiliary grid into which the target point falls is identified, the triangles intersecting this auxiliary cell will investigated. When the triangle containing the target point is finally identified, the "edge grid functions" could then be used to calculate the exact value of the field at the target point location. Contact at AWI regarding edge grid functions: Sergej Danilov, Dimitri Sidorenko, Sven Harig. For cell based search, the multigrid algorithm could be used to identify in which triangle the corners of the cell fall, then Sergej, Dimitri and Sven should be consulted on how to calculate the resulting value for the cell.
        - For end of January within ScalES, Kerstin to provide a first proposition for a prism_set_connectivity API for iteration with other people. Support of unstructured grids in the sense of finite-elements grids (not clouds of points) is planned.

    - Note on icosahedral grid: a icosahedron is a regular polyhedron with 20 identical equilateral triangular faces (forming 10 diamonds), 30 edges and 12 vertices.
- (P1, Sophie) Conservative remapping improvement (also in OASIS3)
    - Normalisation by true area (prism_set_areas required with one area value per cell)
    - To improve the calculation near the pole, a rotation could be implemented for cells near the pole (short term solution)
    - On the mid-long term, other algorithms should be considered for precision and efficiency: Monte-Carlo algorithm, IPSL approach, GFDL new scheme, Phil Jones, other.
- (P1, Sophie) Support of vector fields, bundles of vector fields:
    - As it is relatively easy to do in the models the rotation from the local coordinate system to the spherical coordinate system, it is decided to support only zonal and meridional vector components as separate coupling fields. The user will have to indicate in the SMIOC that these fields are vector component and the Transformer will then automatically do the projection in the Cartesian coordinate system based on this SMIOC information (that will have to be transferred to the Transformer)
- (P1, Hanke) Global conservation between the source and the target grid
    - This will require a collective operation (MPI_Allgather) on the source and target sides; the global integral on the source side needs to be transferred to the target side where the difference will be calculated and distributed.
    - This operation should be needed only if there is a mistmatch between land-sea mask; otherwise, the conservative remapping should ensured global conservation.
- (P1, Jean) Possibility to use user-defined weights-and-addresses (J. Latour, CERFACS)
    - Work is completed and tested; needs to be completely validated with different sets of weights-and-addresses
- (P1, Jean-Marie) GUI for SMIOC and PMIOD XML constitution (J.-M. Epitalon, CERFACS)
    - Installation procedure under test by Laure
    - All OASIS4 developers will then be used as beta testers
- (P1, Jean-Marie and Sophie) Simplification of XML file structure (in METAFOR):
    - Descriptive information not needed and it is confusing for the user; remove it.
    - Some information like grid_type is redundant; remove this from the SMIOC
    - Remove pole_covered
    - Still needs to decide what to do with periodicity (René checks its use)
    - Sophie to check if anything else should be simplified.
- (P1, Hubert) Support of sequential components into one application/executable
- (P1, Laure) I/O for non-geographical grids
- (P2, René & Hubert) Support of regional domain
    - The idea is to provide a value for all target points, even the ones falling into source holes (under a specific option activated by the user).
- (P3) Test on hybrid platforms
    - To address on the longer term
    - Note: to use GPUs, part of the code would have to be rewritten with special directives
- (P3) Heterogeneous computing
    - Currently, no strong demand for heterogeneous computing in the community

        - Should be possible with heterogeneous implementation of mpi (e.g. i-mpi)
    - (P3) Storage and reuse of weights-and-addresses by Transformer (TRAC ticket #41)
    - (P3) Transfer of Transformer functionality in the source PSMILe
        - The advantage would be to have one less executables in the coupled system
        - This seems to be desirable for operational centres (both ECMWF and Environment Canada have asked for such functionality)
        - The disadvantages are that the put would become blocking (as the Transformer is acting as a buffer) and that the local memory of the source psmile would be bigger
    - (P3) Support of dynamic grids or partitions (with no changes of total number of processors):
        - This task is evaluated to 12 pm.
        - The mechanism could be as follows: When a source/target model changes its grid, it would send an additional message to the T within the prism_put/prism_get and start the "enddef" steps explicitly before really sending/receiving the field. The corresponding target/source model would receive, when performing its prism_get/prism_put, some indication that the source/target model grid has changed and would automatically start the "enddef" steps below the prism_get/prism_put before really receiving/sending the field. This probably implies that the prism_put would be blocking.
    - (P3) Integration of CISL interpolations from AWI
        - Planned in the 3rd year of ScalES
    - (P3) Simple research algorithm applicable for all grids (TRAC ticket #20)
        - The idea would be to implement a simple and not efficient search algorithm usable only in monoprocessor cases but for all types of grids to make sure that OASIS4 covers at least all OASIS3 functionality
        - This is not considered high priority for now and time should be spent on other OASIS4 specific tasks instead.
    - (P4) vneighbour option for nneighbour2D and nneighbour3D
        - Not needed; close ticket 19

- Other low-priority developments
    - (P3) 2nd order conservative remapping (TRAC ticket #27) (including over partially masked target domain)
    - (P3) More PSMILe function to get SCC and SMIOC info in model code: on demand if any (TRAC ticket #20)
    - (P3) Mixed openMP - MPI parallelisation
    - (P3) Support other exchanges dates than at a regular frequency; not too difficult: description in the SMIOC and some sophistication of the time axis definition
    - (P3) Support dynamically changing frequency:
    - (P3) Combination of more than one source fields for one target field
    - (P3) Support multiple sources for one input with switching between them for the different coupling timesteps

OASIS web site and interaction with vERC

- Kerstin Ronneberger presented a first version of the vERC based on XXXX.
- The OASIS web site could be either developed separately at CERFACS and the vERC would have a link to this web site, or could be fully integrated of the vERC. It should be noted that even in the last case, modification of the content directly by OASIS developers would be possible. This second option seems preferable both for CERFACS, as this would reduce the workload, and for DKRZ, as this would give concrete material for the vERC. This needs to be

confirmed internally at CERFACS. In this case, it should be ensured that CERFACS gets enough visibility on the OASIS page in the vERC. The vERC web site would target the OASIS users and the current OASIS wiki page would be maintained for the developers.

- In both cases, the information to display on a front page needs to be identified.