D2.3 Fully optimised and scientifically validated version of CNRM-CM-HR model (T359, ~ 1/4°) implemented using a subset of the platform (XIOS, OASIS3-MCT, ...)

D2.5 CNRM-CM version at high resolution (T359 ~ 1/4th°) using (at least large part) of the platform

Sophie Valcke, Marie-Pierre Moine
November 2018

Cerfacs Technical Report #TR-CMGC-18-210

Table of content

1. Introduction

This document reports the work done at CNRM and at Cerfacs to develop the high-resolution version of CNRM-CM6 climate model, CNRM-CM6-1-HR, using different tools from the CONVERGENCE common modelling platform. The high-resolution (HR) version strongly follows the low-resolution (LR) version with of course few modifications, which are specifically described hereafter.

There were two CONVERGENCE deliverables concerning CNRM-CM6-1-HR, i.e. “D2.3 Fully optimised and scientifically validated version of CNRM-CM-HR model (T359, ~ 1/4°) implemented using a subset of the platform (XIOS, OASIS3-MCT, ...)”, originally due at month 40 then delayed month 46, and “D2.5 CNRM-CM version at high resolution (T359 ~ 1/4th°) using (at least large part) of the platform”, originally due month 44 then delayed month 60. As the optimisation and the scientific validation comes necessarily after the technical set-up of a model, the two deliverables originally described in CONVERGENCE were gathered into one.

The current version of CNRM-CM6-1-HR, used to run HighResMIP (High-Resolution Model Intercomparison Project) simulations in the framework of the 6th phase of the Coupled Model Intercomparison Project (CMIP6) and EU project PRIMAVERA, is available since May 2018 (i.e. month 53 of the project) and is described hereafter. Like the other French climate model, IPSL-CM6, CNRM-CM6-1-HR uses the main bricks of CONVERGENCE modelling platform, i.e. the OASIS3-MCT coupler, the XIOS IO server and the dr2xml python environment for configuring XIOS.

In this document, we present the composition of CNRM-CM6-1-HR, describing first the 3 components in terms of binary, namelist and input files (section 2.1), then the coupling exchanges between these components managed by OASIS3-MCT (section 2.2), and then our specific I/O workflow managed by XIOS (section 2.3). In section 3, we present some

performance analysis and optimisations that helped increasing CNRM-CM6-1-HR throughput and we finish in section 4 with a first scientific validation of CNRM-CM6-1-HR.

2. CNRM-CM6-1-HR composition

Most components of CNRM-CM6-1-HR are available in CNRM official stable directory structure for CMIP6, i.e. /scratch/CMIP6/V2/ on Météo-France bullx computing platform Beaufix. This directory will be referred to as $CMIP6 hereafter. Other directories mentioned hereafter are all on Beaufix.
The low resolution (LR) version of CNRM climate model, CNRM-CM6-1, will be described in detail in a coming paper Voldoire et al 2019. However, as this paper is not published yet, we copied in Appendix A, the current status of its section "2 Model description". Hereafter, we give a global description of CNRM-CM6-1-HR components, insisting on the differences with the LR version.

2.1. Components (binary, namelist, input files)

We provide here some details on the binary used in CNRM-CM6-1-HR coupled model. We also provide some details on the namelist and input files but their precise name and location for each component can be found in Appendix B, which provides the ECLIS configuration file for CNRM-CM6-1-HR historical, param_CPLHR_HIST1950_P1_v1.

Atmospheric and surface component ARPEGE-SURFEX

The atmospheric component of CNRM-CM6-1-HR is ARPEGE-climat v6.3.2, i.e. the same than in the LR version. It contains the SURFEX module that calculates surface fluxes over all surface types. The corresponding binary is available in $CMIP6/bin/atm . The high-resolution version uses a linear triangular truncation T359 with a corresponding reduced Gaussian grid at ~50 km resolution; the number of vertical levels in the atmosphere is 91 as in the LR version.
As configuration file for ARPEGE, we use the standard T359 namelist from $CMIP6/namelist/atm, except that the cloud simulator COSPSIM is no activated. For SURFEX, the namelist used is very close to the low resolution one except for some parameters linked to the resolution and to the sea/coast treatment (see section 2.2.3 below).

The different input files used are the standard ones for the surface fixed conditions such as orography and for the land cover and surface parameters (ecoclimap*.bin)

Hydrology TRIP

The hydrology model TRIP, routing the water to the ocean, forms a separate executable. The standard binary used is available in $CMIP6/bin/riv. The HR and LR coupled models use the same TRIP version with 720x360 grid points.

The grid files for TRIP are the standard ones from ${CMIP6}/data/riv

Ocean and sea-ice NEMO-Gelato

The ocean component of CNRM-CM6-1-HR is based, as the LR one, on the version 3.6 of NEMO (SVN revision 9455 tagged 3_6_STABLE) with few evolutions registered under git (revision f845920e7f2169d) in CNRM Open Source Site (https://opensource.umr-cnrm.fr). It uses the extended eORCA025 tripolar grid with 1442x1050 grid points horizontally, resulting in a resolution of ¼ degree at the equator (27 km), and the same vertical resolution than in the

LR coupled model, i.e. 75 levels with a depth of 1m near the surface to 200 m at a depth of 6000 m.

The sea ice component, Gelato 6, is also available through git (revision 8250e198106a168) in CNRM Open Source site. Gelato 6 is a parallel code completely embedded in NEMO from which it inherits the horizontal grid and global domain decomposition.

The ocean namelist is quite similar to the LR but for some specifications linked to the grid resolution and grid folding at the North (not detailed here) and for the following parameters:

- rn_dt : the time step for the dynamics: 900 seconds (instead of 1800 seconds);
- nn_fsbc: the sea-ice and surface boundary computation is called every 4 time steps (instead of 2);
- sn_chl: the file used to prescribe the chlorophyll is merged_ESACCI_BIOMER4V1R1_CHL_REG05.nc and no light penetration is activated;
- rn_dep_max: river runoffs are spread vertically depending on the flow intensity to a maximum depth of 150m (instead of 10m);
- rn_aht_0 and rn_aeiv_0: the horizontal eddy diffusivity for tracers and the eddy induced velocity coefficient are 300 m^2/s (instead of 1000 m2/s);
- namdyn_ldf sub namelist: a horizontal laplacian eddy viscosity of 20000 m^2/s is used instead of a bilaplacian operator

Gelato namelist is the same than the LR version except that only 5 vertical layers are activated (instead of 9).

The input files for NEMO come from /home/gmgec/mrgi/chevalli/DATA/NO_SAVE/NEMO3.6/eORCA025L75/data_eORCA025L75_cmip6.3, except for the restart temperature and salinity fields, which come from EN4 (Good et al 2013) to follow the PRIMAVERA protocol. The original EN4 data had to be processed to produce conservative temperature and absolute salinity, as we use the TEOS-10 thermodynamic approach (http://www.teos-10.org) to determine the properties of sea water, and to detect and suppress density inversions. The resulting files are conservative_temperature_clim_EN4_eORCA025L75_CERFACS_corrected_extrap.nc and absolute_salinity_clim_EN4_eORCA025L75_CERFACS_corrected_extrap.nc in the $DATA_OCE directory (see Appendix B).

2.2. Coupling ARPEGE-SURFEX, TRIP and NEMO-GELATO through OASIS3-MCT

2.2.1 Coupling exchanges

The coupling fields and remappings managed by OASIS3-MCT_3.0 in CNRM-CM6-1-HR between the 3 executables ARPEGE-SURFEX, TRIP and NEMO-GELATO are practically the same than in CNRM-CM6-1. The only difference is the option “opt” activated for the “CONSERV” operation, which carries out the global conservation with an optimal algorithm using less memory and a faster algorithm[1]. The OASIS3-MCT_3.0 version used corresponds to revision 1983 of the SVN branch OASIS3-MCT_3.0_branch.

NEMO sends the ocean and sea-ice surface properties, i.e. sea surface temperature over water (O_SSTSST) and ice (O_TepIce), ocean currents (O_OCurx1 & O_OCury1), sea-ice fraction

[1] The drawback of this algorithm is that it does not ensure bit-for-bit reproducibility when the grid decomposition or number of processes of the component are changed.

(OIceFrc) and sea-ice albedo (O_AlbIce) to ARPEGE-SURFEX. These are used in SURFEX to calculate the surface turbulent momentum fluxes (COZOTAUX & COMETAUY) and the non-solar (latent, sensible, longwave) total flux (CONSFTOT) and over ice flux (CONSFICE), which are returned back to NEMO. The water fluxes (precipitation COTOLIPR, sublimation COSUBLIM , snow COTOSOPR, evaporation COTHSHSU) as well as the total (COSHFTOT) and ice (COSHFICE) solar fluxes are also calculated in ARPEGE-SURFEX and transferred to NEMO. The heat fluxes are redistributed over ice and water in NEMO-Gelato as a function of the current ice cover in each cell so to ensure global conservation.

In addition, the excess runoff, deep soil drainage, and calving (SXRUNOFF, SXDRAIN:=, SXCALV) are sent from ARPEGE-SURFEX to TRIP with also the freshwater flux to the atmosphere over floodplain open water (SXSRCFLD). TRIP send to ARPEGE-SURFEX the water table depth of the groundwater (TRWTD), the grid-cell groundwater fraction (TRFWTD), floodplain grid-cell fraction (TRFFLD), and floodplain potential infiltration (TRPIFLD). TRIP routes continental water to the ocean and sends the resulting runoff (TRRIVDIS) to NEMO, where it is spread vertically depending on the flow intensity to a maximum depth of 150m (instead of 10m in the LR version). TRIP also sends Greenland and Antarctic calvings (TRCALVGR and TRCALVAN) to NEMO; the water corresponding to calving is incorporated at the ocean surface over the globe for the Greenland ice sheet and only south of 60S for the Antarctic ice sheet. Finally, ARPEGE-SURFEX calculates net water budget over lakes (LKWATBUD) and sends it to NEMO where it is distributed over the globe to close the water budget.

2.2.2 Cell corner definition for the NEMO grid

Some of the coupling exchanges described above require conservative remapping. A definition of the grid cell corners is therefore needed by OASIS3-MCT. The definition of these corners was not straightforward for the NEMO grid.

The NEMO grid is of Arakama-C type; this means for example that a T point (onto which temperature is defined) at (i,j) is surrounded by F points (onto which vorticity is defined) at (i,j), (i-1,j), (i-1,j-1), (i,j-1). In the NEMO code, routine cpl_oasis3.F90 can be used to define the “corners” of the T grid, in the OASIS sense, exploiting this structure. This has to be done in a run not excluding land processes (see section 3.1) in order to avoid holes in the grid definition. However, this routine was initially producing wrong corners for the before-last j line (j=1049), which resulted in holes in the grid as shown on figure 1.

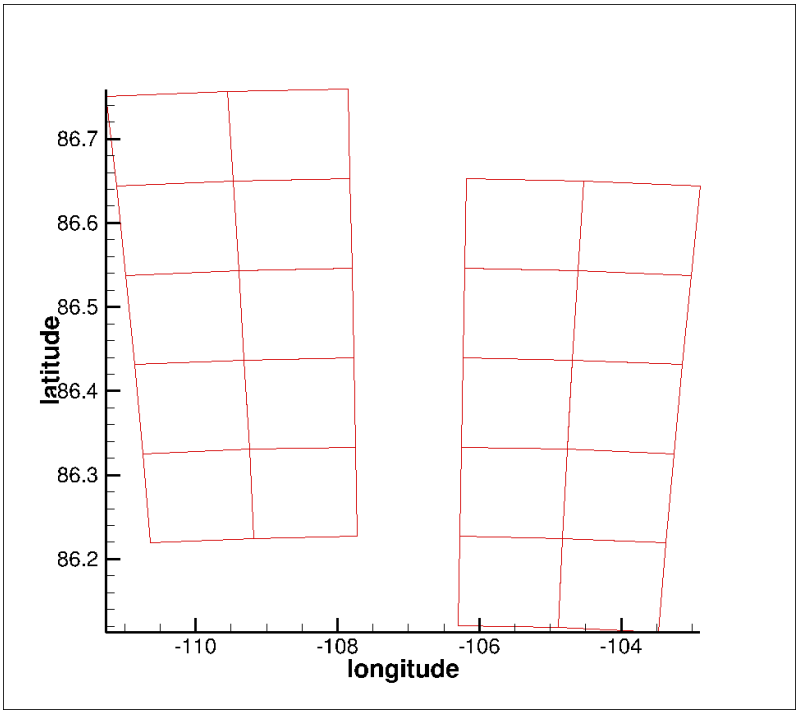

Figure 1: grid cells as originally defined by routine cpl_oasis3.F90 for the eORCA025 grid for 4 top rows (j=1047 to 1050) and for i=413 to 417.

After different tests and code examination, the problem was fixed by modifying the arguments of the call to lbc_lnk routine for the corners longitude and latitude as follows:

CALL lbc_lnk(zclon(:,:,jc), 'T', 1.,'mpp')

whereas it was “CALL lbc_lnk(zclon(:,:,jc), 'F', 1.)” originally. With these arguments, routine lbc_lnk completes the halos but does not reorder the array, as could be needed with the special North Fold of the NEMO grid.

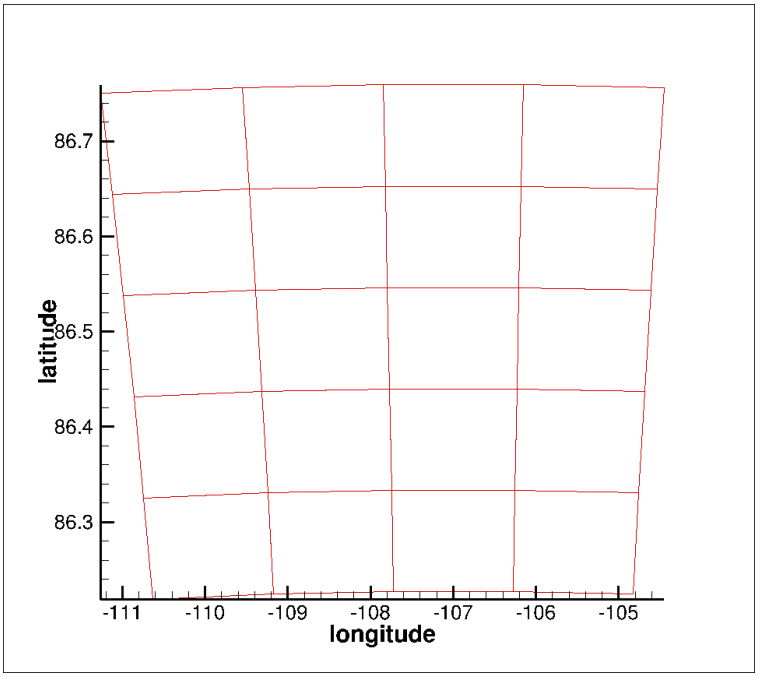

Figure 2: grid cells as defined after the bugfix in routine cpl_oasis3.F90 for the eORCA025 grid for 4 top rows (j=1047 to 1050) and for i=413 to 417.

With this modification, the grid cells are now all properly defined and there are no holes in the grid anymore as shown on figure 2.

Once the corners had been properly defined by the routine, another difficulty was encountered when we tried to use OASIS3-MCT routines to write the points, corners, masks and areas to the NetCDF files grids.nc, masks.nc and areas.nc expected by the coupler. In parallel, the code would abort not directly when calling those routines but in another NEMO routine performing specific parallel operations. The reasons for this bug were never found and we finally decided to bypass the problem by using directly NetCDF calls so to make each process write its part of the grid. We then had to use the NEMO rebuild script rebuild_nemo_outputs.sh from the ECLIS environment to recreate one file gathering the partial files.

2.2.3 Bugfix for the bicubic remapping

When creating the grid, area, and mask files for the coupler, an issue regarding the bicubic remapping was encountered. Initially, the land-sea coupling masks of ARPEGE-SURFEX and NEMO were defined independently. This implied that for some remappings, some target grid points near the coast could have only masked points among their original neighbours on the source grid. In that case, the “additional non-masked nearest neighbour” option was activated which means that an extra search was performed so to assign the nearest non-masked neighbour to these target points.

It was observed that for the bicubic remapping from ARPEGE-SURFEX to NEMO, this option did not properly work for a series of non-masked target grid points located at the very South of the domain, which did not get any value. An inspection of the code revealed that the “additional non-masked nearest neighbour” was effectively activated but was considering only source grid points from a restricted portion of the source grid (that restriction is used to

accelerate the original search) which, in that specific case, contained only masked points. That bug was corrected in OASIS3-MCT, extending the search to the whole grid for these cases. This problem did not show up for the low-resolution grids because the masks were coinciding better.

2.2.4 Land-sea coastline and mask issues

Even with the bug fix described above, the remapping error was still observed to be quite high in some regions, e.g. 11.5% for the conservative remapping for some ARPEGE-SURFEX non-masked grid points on the Asian continent corresponding to masked regions on the NEMO grid, or of about 1.8% form ARPEGE-SURFEX to NEMO at the very South of the NEMO domain corresponding to a region entirely masked on the ARPEGE-SURFEX grid.

It was then decided to modify the proportion of land and sea in ARPEGE-SURFEX cells (in input files ecoclimap*.bin) so to fit to the land-sea coastline defined by the NEMO mask. The NEMO mask was remapped to the ARPEGE-SURFEX grid with a conservative algorithm and the proportion of land and sea in each atmospheric grid cell was adapted so to fit the remapped NEMO mask. For the ARPEGE-SURFEX coupling mask, a cell is declared unmasked as soon as it gets some proportion of sea. This means that the coastlines in ARPEGE-SURFEX and in NEMO are now coherent and the coupling problem is not ill-posed anymore. This implies that the "additional non-masked nearest neighbour" option is not required anymore and that we should not have to apply any global conservation operation anymore.

Another issue concerned the overlapping grid cells. For the conservative remapping in OASIS3-MCT it is essential, when two cells are overlapping, to mask one so that their contribution would not be counted twice. However, we have to be cautious to mask the cell that NEMO expects to be masked as it is NEMO that will assign proper values to masked cells. In the eORCA025 grid, the columns of the before-last and last i (i.e. i=1441, 1442) overlap the first and second columns (i=1, 2). Also the last j (j=1050) overlaps the before-before j (j=1048) in opposite direction, and the before last j (j=1049) overlaps to itself. NEMO expects to receive a coupling field into which i=1 and i=1442 and the last j=1050 are masked. A bit surprisingly, the line with j=1049 does not need to be half-masked. We were of course careful to follow these constraints.

2.2.5 Detailed analysis of the remapping quality

Once the coupling grids, masks and areas were defined, solving all the issues mentioned above, a detailed quality of the different remapping involved in CNRM-CM6-HR was realised. These tests were done using the test_interpolation environment, available with the OASIS3-MCT release, that couples two component models, model1 and model2, and evaluates the quality of the remapping between the source grid of model1 and the target grid of model2. The field sent by model1 is defined by an analytical function on model1 grid. The remapping error is defined as the difference between the interpolated values of the field received and the values of the analytical function on the target grid points, divided by the interpolated field (and multiplied by 100 to have it in %).

The grids acronyms used in the OASIS3-MCT configuration file namcouple for CNRM-CM6-HR and a description of the corresponding grids are given in Table 1.

Grid	Description
ssea	ARPEGE Gaussian Reduced T359 over sea
torc	NEMO eORCA025 T grid over sea
slan	ARPEGE Gaussian Reduced T359 over land
tlan	TRIP over land
tfld	TRIP for flooding plains
tsea	TRIP over runoff discharge points
nort	NEMO ORCA025 T grid over a narrow region near the coast for runoff discharge
tgw	TRIP for groundwater
tant	TRIP over Antarctica
noat	NEMO ORCA025 T grid over sea south of 60S for the Antarctica water budget discharge
trge	TRIP over Greenland
slak	ARPEGE Gaussian Reduced T359 over lakes

Table 1: Acronym and description of the grids used in CNRM-CM6-1-HR

Table 2 provides the results for the remappings involved in CNRM-CM6-1-HR. In all cases, we checked that all unmasked target point receive a value.

A) Source grid	B) Target grid	C) Remapping	D) Max error (%)	E) Grid point of max error	F) Mean error (%)
nogt	ssea	FRACNNEI	0.32	(10,1)	0.044
nogt	ssea	BILINEAR	0.31	(140137,1)	0.0032
ssea	nogt	BICUBIC	0.15	(595,992)	0.000021
ssea	nogt	FRACNNEI	0.23	(989,17)	0.0048
slan	tlan	FRACNNEI	0.41	(187,276)	0.021
slan	tfld	FRACNNEI	0.21	(203,185)	0.014
tlan	slan	FRACNNEI	11.9	(35035,1)	0.041
tsea	nort	DISTWGT	12.6	(1021,699)	0.43
tgw	slan	FRACNNEI	40.6	(177168,1)	2.22

Table 2 – Maximum error (D), grid point of maximum error (E), and mean error (F) for the different remappings (C) involved in CNRM-CM6-1-HR for the different source (A) and target (B) grids (see Table 1 for the acronym signification).

We have inspected plots of the 2D error fields in detail and we conclude that all remappings are valid, even if the error for the last 3 remappings, which involve specific grids, seems to be high. For the FRACNNEI remapping from *tlan* to *slan*, the maximum errors are located on few South Pacific points for which an island exist in the target *slan* grid but not in the source *tlan* grid; the non-masked nearest neighbour option therefore uses the value from the nearest non-masked island. The DISTWGT is implemented to match TRIP discharge grid points (on land) to the nearest non-masked NEMO point over a narrow oceanic region near the coast; it is therefore normal that the remapping error appears quite high in this matching of land to ocean points. For the FRACNNEI remapping from *tgw* to *slan*, the average error is quite high because of high error over Antarctica and Greenland which are masked on the source grid and not on the target grid; this should have no impact on the remapping of TRIP groundwater.

Finally, we note that there is in fact no interpolation from TRIP to NEMO for the Greenland and Antarctica calvings (TRCALVGR on *trge* -> OCalvigr on *nogt* ; TRCALVAN on *tant* -> OCalvian on *noat* in the namcouple), and from ARPEGE-SURFEX to NEMO for the water budget over lakes (LKWATBUD on *slak* -> OLakeWat on *nogt* in the namcouple). For these fields, it is the only global conservation operation that results in a uniform repartition of the source field over the non-masked target grid.

2.3. Data processing and I/O via XIOS

When the development of CNRM-CM6-1 and CNRM-CM6-1-HR started, the objective was to use these models in many different simulations, in particular in the framework of the on-going 6th phase of the international Coupled Model Intercomparison Project (CMIP6) and in other European projects such as PRIMAVERA, APPLICATE or COPERNICUS. These projects have extremely high requirements in terms of outputs. For CMIP6, these output requirements form the CMIP6 Data Request (DR); for the other projects, we created what we call a “home data request” so to use the same workflow.

One important aspect of the restructuring between the CMIP5 and the CMIP6 data management workflow was to avoid the use of CMOR library to produce CMIP6-compliant data, since it requires heavy data handling and is very time consuming. To reach this objective, XIOS, one of the main brick of the CONVERGENCE platform, was implemented by CNRM in each component of CNRM-CM6-1 (ARPEGE, SURFEX, TRIP - NEMO being since a long time XIOS-enabled). XIOS, developed by IPSL, is a parallel Input/Output server that allows a very flexible and sophisticated configuration of the output files to produce; it also offers online field operations, thereby reducing, or even completely removing, the need of post-processing. XIOS operations include time sampling and averaging, spatial remapping and reduction, vertical interpolation and simple arithmetic. These operations are configured at run time using an XML syntax. To exploit XIOS functionalities in the CMIP6 framework, an additional tool, called dr2xml (https://github.com/senesis/dr2pub), translates the CMIP6 and home DR for each year of each experiment in a set of XIOS conformant XML definitions. These definitions then activate the outputs in the coupled model components thanks to an alias table describing the correspondence between model variable names and CMIP6 DR variable names, called the “ping” files. The operations on the data are then performed online by the client part of XIOS on the model MPI tasks and sent to XIOS additional MPI tasks, called servers. Two levels of servers are used to aggregate and redistribute the parallel output fields so that any given field is gathered into one file written to disk by one single MPI task (thus avoiding parallel writing). Special care was taken in XIOS to allow for an overlap of

model computation and I/O operations using communication buffers so to allow for scaling at high number of cores.

Thanks to these tools and environment, CNRM-CM6-1 generates NetCDF output files, which are directly conformant with CMIP6 and/or PRIMAVERA requirements and ready to publish on the Earth System Federated Grid (ESGF). This approach is applicable to any other model and is currently also used in the Institut Pierre Simon Laplace (IPSL) climate model.

Thanks to this environment, our historical and control runs using CNRM-CM6-1-HR produce for example, without any post-processing, about 180 GO of output data for each simulated year involving 436 geophysical variables which can be invoked in 42 so-called tables MIP tables (which combine a specific physical component, e.g. the ocean, with a specific frequency, e.g. monthly) thus leading to 590 'CMOR variables' covering a variety of spatial shapes and frequencies.

3. Performances and optimisation

3.1. Elimination of processes over land for NEMO

The first optimisation was to activate the option in NEMO to remove the processes that would cover only land points and do not have any real calculation to perform.

To do this, the total number of local domains (*jpnij* in the namelist and *NPROC_OCE* in the ECLIS configuration file) has to be set to the effective number of domains (i.e. excluding ones that would cover only land) and this number has to be different from and smaller than the product *jpni* x *jpinj* where *jpni* and *jpinj* are namelist parameters giving the number of domains for the decomposition in the i and j directions respectively. The parameter *ln_fillcoordinates* also has to be set to *true* in the namelist.

A tool OptimizeProcs.sh delivered with NEMO sources can be used to define the possible numbers. In our case, we did performance tests for two different numbers of processes:

- jpnij = NPROC_OCE = 1094 while *jpni* x *jpinj* = 43x34=1462
- jpnij = NPROC_OCE = 530 while *jpni* x *jpinj* = 25x27=675

3.2. Load balancing for different I/O volumes

While activating the elimination of land processes, different trial-and-error tests were performed to determine the optimal number of processes for each component so to have a load balanced system. To that end, the *lucia* tool, provided with the OASIS3-MCT coupler, was activated to measure the waiting and calculation time in each component and different XIOS measures were also examined. Some of these numbers are reported in Table 1 for some of the tests performed.

Practically without I/O (A in Table 3 with 0.0005 GO/year), the tests lead to the conclusion that a configuration over 48 nodes with 768, 1094, 1 and 48 cores for ARPEGE, NEMO, TRIP and XIOS respectively is relatively well load-balanced. In that case, the coupled model takes 2374 sec to run one month, which leads to 3.0 SYPD (Simulated Years Per Day). ARPEGE-SURFEX waits about 4.8 % of the time (113.04/2375) and NEMO slightly more i.e. 15% of the time. We have also checked (thanks to other measures available in XIOS log files, not shown here) that the number of XIOS servers (75% of 48 i.e. 36) does not cause any bottleneck as the client part is almost never waiting for server buffers to get free

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Expe	NPROC_ARP	NPROC_OCE	NPROC_IOS	ratio_server2	Volume/an (GO)	Elapsed per month (sec)	Arpege: total time spent for XIOS	Nemo: total time spent for XIOS	Trip: total time spent for XIOS	lucia surfex calculations	lucia nemo calculations	lucia trip calculations	lucia surfex waiting	lucia nemo waiting	lucia trip waiting
A	768	1094	48	75	0.0005	2375.0	43.3	29.9	8.7	1553.8	1309.8	613.79	113.04	357.02	1093.24
B	768	1094	48	75	996.0	7328.0	470.8	129.0	31.3	5106.3	2755.8	648.5	1191.23	3541.77	5712.28
C	768	1094	48	75	360.0	5548.0	312.4	117.3	90.3	3778.6	1775.8	770.42	472.49	2475.31	3684.71
D	768	530	48	75	300.0	4759.0	289.0	1289.0	163.2	3067.7	3230.8	698.49	382.55	219.39	3533.89

Table 3 – Time spent in different parts of the components running on different number of cores (columns 2, 3 & 4 for ARPEGE, NEMO and XIOS respectively – TRIP is always run on 1 core) for specific I/O loads (0.0005, 996, 360 and 300 GO/year on lines A, B, C & D respectively). Column 7 gives the elapsed total time for one month of simulation. Columns 8, 9 & 10 give the time spent on XIOS client side for calculations in ARPEGE, NEMO & TRIP; columns 11, 12 &13 the time spent in each component for its proper calculations, while columns 14, 15 & 16 provide the time spent waiting in the initialisation and during the coupling exchanges.

But the performance and the load balance of the coupled model drastically drops when the I/O volume is increased. With an extremely heavy load of 996 GO per year (corresponding in fact to the requirements of both CMIP6 HighResMIP and the PRIMAVERA EU project), i.e. line B in Table 3, the throughput of the coupled model drastically falls down to ~1 SYPD (7328 sec/month). We verified again that the slow down comes from the XIOS client part and not from the XIOS server part. The increased calculation load in XIOS client part (columns 8, 9 & 10 in Table 3) for ARPEGE-SURFEX or NEMO-Gelato is not the only reason for the increased elapse time. We see that the waiting time both in ARPEGE-SURFEX (column 11, 1191.23 sec) or NEMO-Gelato (column 12, 3541.77 sec) increases drastically. It was observed separately that this increase in the waiting time is in part due to the initial generation

by XIOS of a file containing NetCDF metadata "distribute_file_surfex_server.dat" [2]. Improvement of this aspect in XIOS is currently going on.

In an effort to diminish the elapse time, we started to reduce the volume of I/O eliminating all variables that we did not consider essential for CMIP6 or PRIMAVERA and were able to get to 360 GO per year (line C) or even 300 GO per year (line D). But in the C case, we see that the components are still very much unbalanced with NEMO waiting 45% of the time (2475.31/5548, i.e. columns 15/7) and ARPEGE-SURFEX waiting 8.5% of the time (472.49/5548, i.e. columns 14/7).

Based on the C case, we decided to reduce the number of cores for NEMO going down to 530 (still eliminating land processes as *jpni* x *jpinj* = 25x27=675) and got the numbers of the D line in Table 1. With ARPEGE-SURFEX waiting 8% and NEMO waiting 4.6% of the time, we consider that the load balance of this configuration is acceptable, as we know that some of the waiting time linked to the initialisation phase is incompressible with the current XIOS version. In that case, one month of simulation now requires 4759 sec, corresponding to ~1.5 SYPD. We run now on 34 nodes (with 40 cores/node) leading to 21760 cores.hrs/SY. We kept this configuration for the final volume of data produced (~180 GO/year).

In conclusion, we can state that even some non-negligible work was done regarding the load balancing of CNRM-CM6-1-HR, additional gain can certainly still be achieved, especially in XIOS initialisation phase. Another important improvement would be to be able to run CNRM-CM6-1-HR with chunks of 1 year instead of one month (obligatory today): the whole initialisation phase would be done 12 times less often and this would necessarily significantly reduce the overall simulation time. Work is currently going on in those two directions at IPSL and CNRM.

[2] Indeed, in its current version, lucia does not make the difference between waiting time in the initialisation phase and waiting time during the coupling exchanges (i.e. load imbalance).

4. Scientific validation

CNRM-CM6-1_HR is used by Cerfacs for the HighResMIP simulations of CMIP6 and PRIMAVERA. In this framework, we ran one 1950 spinup of 30 years (starting from rest and EN4, see section 2.1), one 1950 control of 200 years, and three 1950-2014 historical simulations . In addition, we also ran ten AMIP (i.e. ARPEGE-SURFEX-TRIP only) 1950-2014 historical simulations. And we are now starting three 2014-2050 fully coupled runs and AMIP scenario runs following the Shared Socioeconomic Pathways SSP5-85.

We used CliMAF, the CONVERGENCE framework for climate models evaluation and analysis, to monitor the simulations while they were running, in order to make sure that the results were right at the first order. But while being trained to master CliMAF for deeper analysis, a first validation of CNRM-CM6-1-HR scientific results of these runs was achieved using the Climate Variability Diagnostics Package (CVDP) developed at NCAR.

The results of the CVDP for CNRM-CM6-1-HR control and historical simulations are available on Cerfacs intranet and on demand for external people. The analysis of CVDP results lead to conclude positively on the technical and scientific validity of CNRM-CM6-1-HR set up. Without presenting here an extensive analysis of all aspects of the model, which per se could form the content of few scientific papers, we present here a very reduced subset of these analyses for CNRM-CM6-1-HR historical (1950-2014) simulation so to give first elements of such a technical and scientific validation.

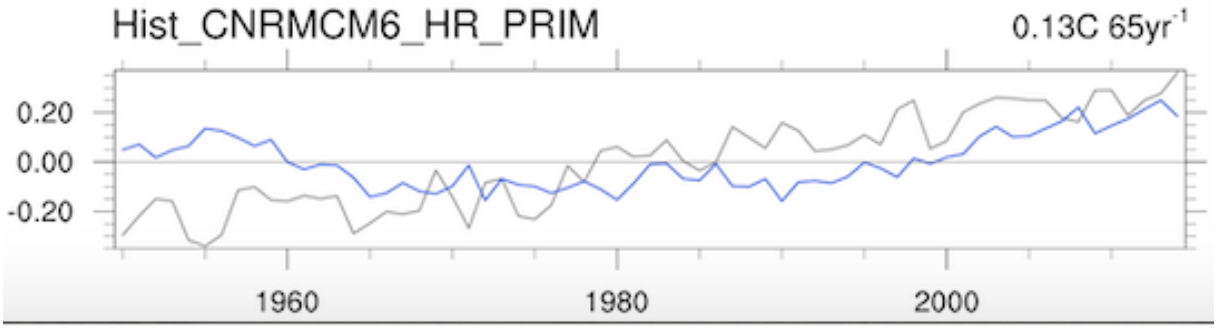

Figure 3 - Sea Surface Temperature (SST) annual global mean for CNRM-CM6-1-HR historical (1950-2014) simulation in blue, and for the Extended Reconstructed Sea Surface Temperature (ERSST) v5 data set, in grey.

Figure 3 presents a timeseries of Sea Surface Temperature (SST) annual mean for the 1950-2014 period (in blue) that can be compared to the Extended Reconstructed Sea Surface Temperature (ERSST) v5 dataset (in grey). CNRM-CM6-1-HR reproduces a global warning especially since 1970; its value of 0.13C/65yr, which is less than for ERSST (0.59C/65yr), seems to be due to a warm initial condition (i.e. our 30-year spinup ended in a warm phase).

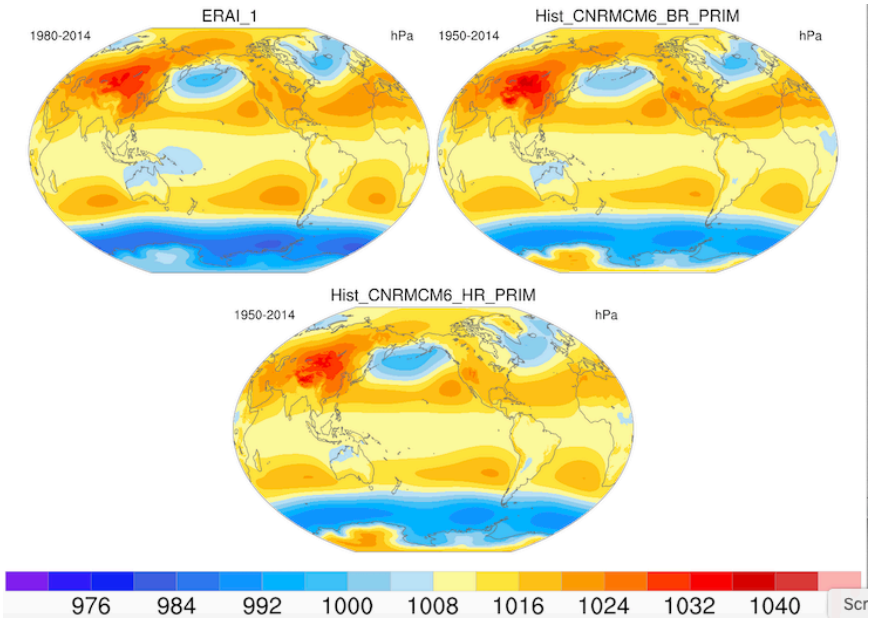

Figure 4 Mean DJF sea-level pressure over the 1950-2014 period for CNRM-CM6-1-HR (bottom), for the LR version CNRM-CM6-1 (top right) and for ERAI (top left).

Figure 4 represents the mean December-January-February (DJF) sea-level pressure over the 1950-2014 period for CNRM-CM6-1-HR, for the LR version of the model and for ERA-Interim reanalysis. We see that the global patterns are in general well represented for both the low and the high-resolution model and that CNRM-CM6-1-HR seems to be less zonal in the Atlantic, which is in better agreement with the reanalysis.

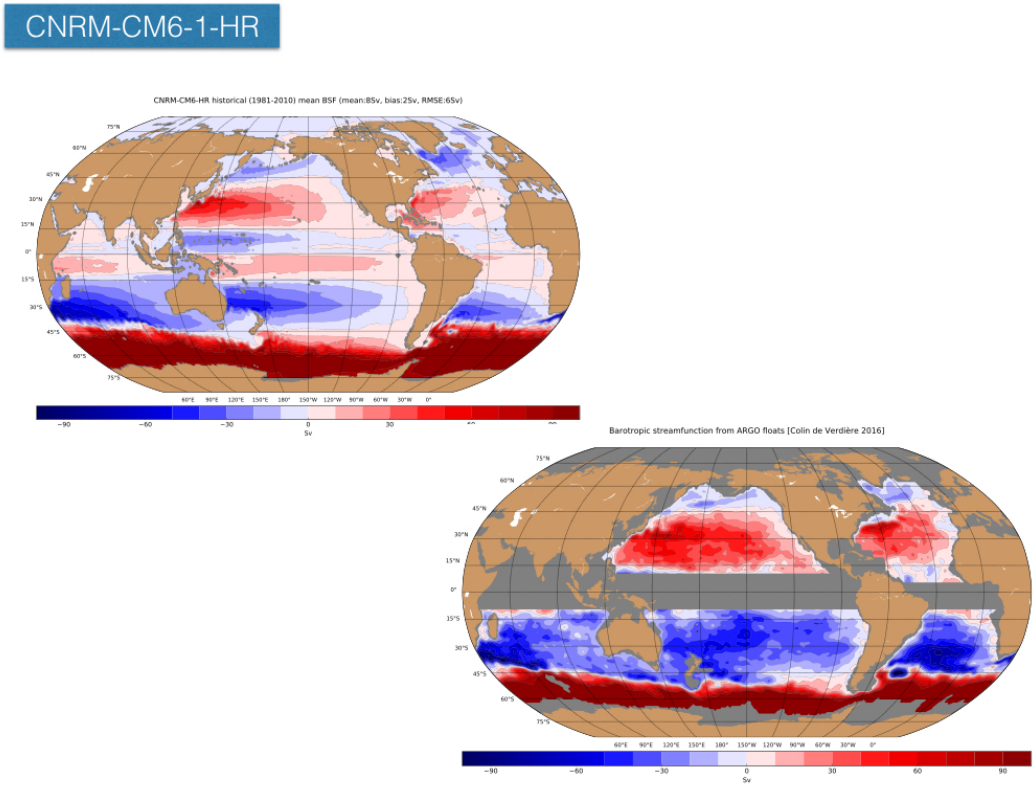

Figure 5 – Mean barotropic streamfunction over the 1981-2010 period for CNRM-CM6-1-HR (top) and from Argo floats (Colin de Verdière, 2016)

The mean barotropic circulation represented on Figure 5 shows that overall the gyres are well reproduced even if they are shifted slightly southward compared to observations from Argo floats (Colin de Verdière, 2016). The flow is also too zonal but there is an improvement in the western boundary regions with respect to the low-resolution results (not shown).

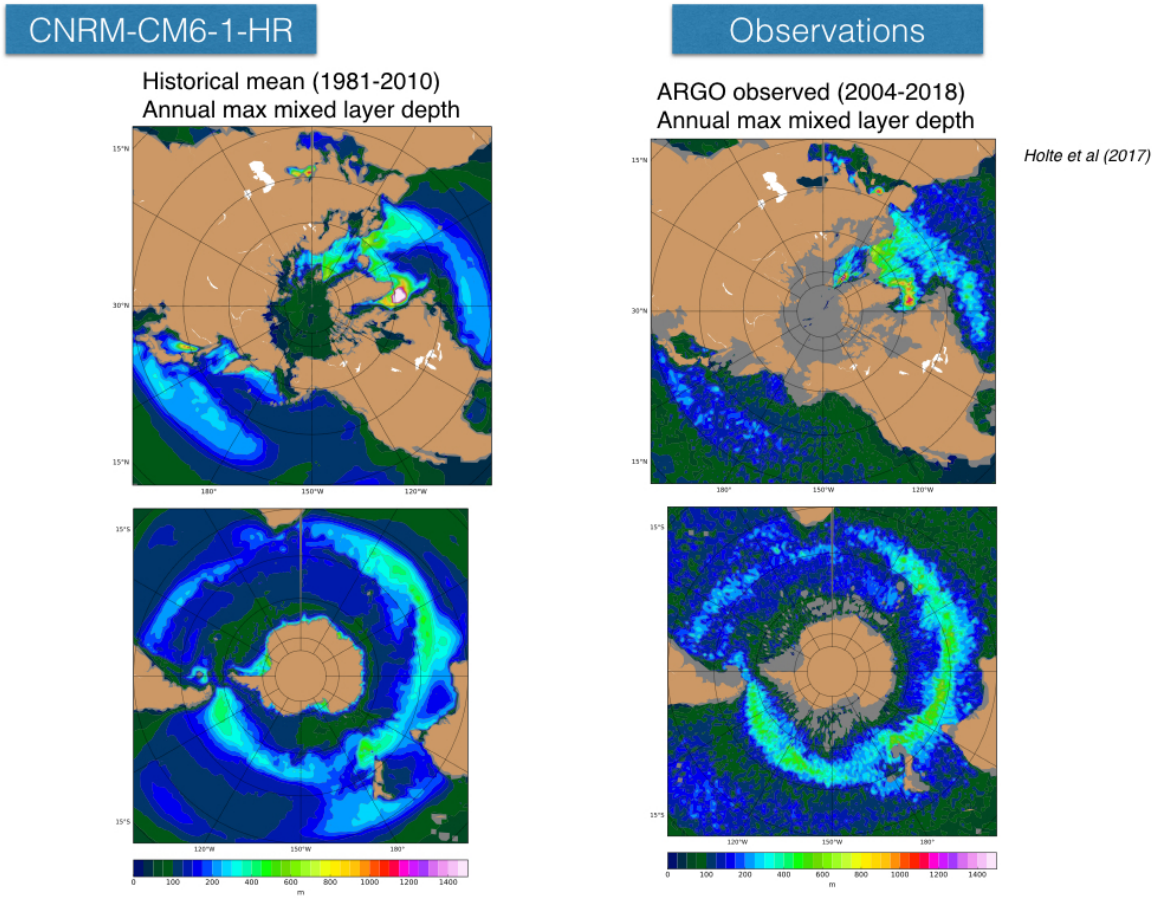

Figure 6 – Mean annual mixed layer depth over 1981-2010 for CNRM-CM6-1-HR (left) and over 2004-2018 from ARGO float observations, for the Arctic (top) and for the Antarctic (bottom)

Figure 6 shows that the deep convection sites in the Northern North Atlantic are realistic in CNRM-CM6-1-HR with deep water formation occurring in the Labrador and GIN Seas consistent with observations. The mixed layer depth in the Labrador Sea is deeper than in the low-resolution model, which represents an improvement. There is no deep water formation in the Southern Hemisphere in HR, which is also the case in observations during the Argo period. We note that we had an unrealistic polynia in the Southern Ocean the low-resolution model that has disappeared in CNRM-CM6-1-HR.

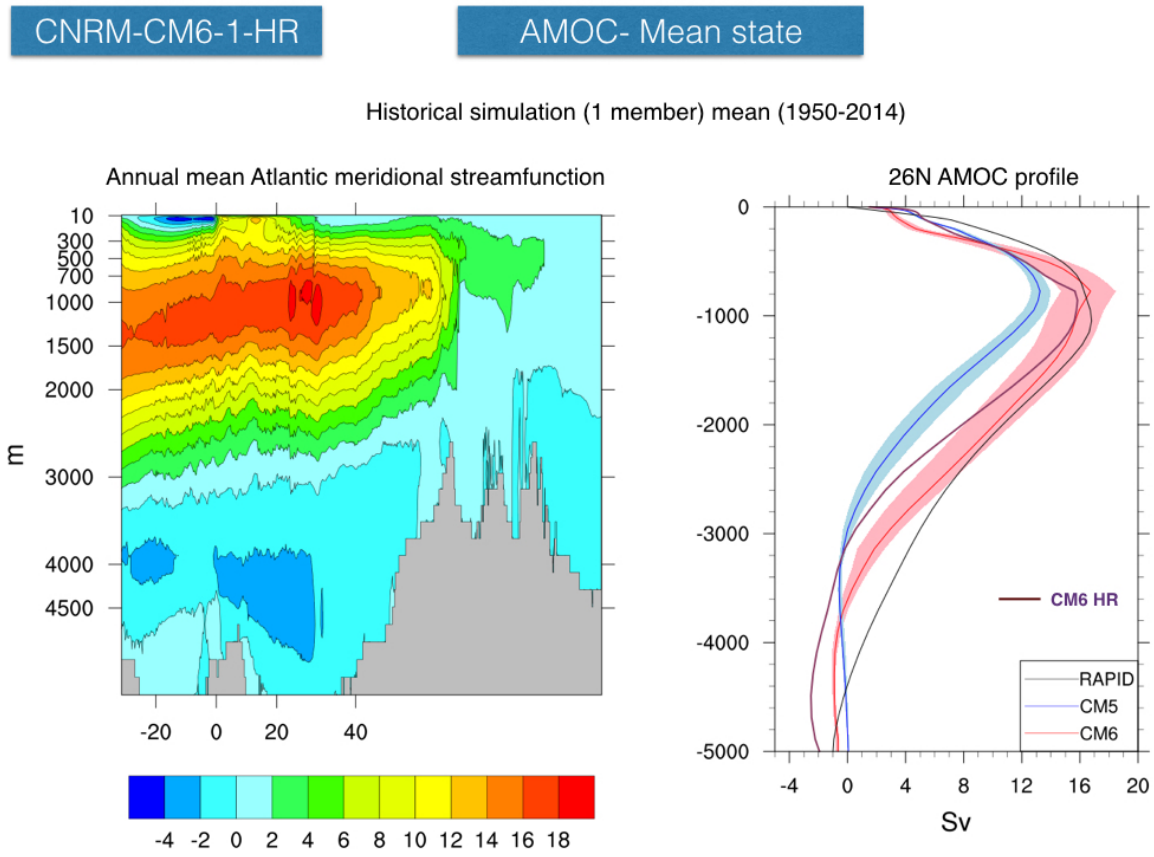

Figure 7 – AMOC stream function annual mean (left) and profile at 26° N (right) for CNRM-CM6-1-HR (purple line). The envelop and the mean profile for 10 historical members of the low-resolution model (in pink and red) and for the previous version of the model CNRM-CM5 (in blue and dark blue) are also shown on the left

The Atlantic Meridional Overturning Circulation (AMOC) shown on figure 7 indicates that the maximum is located around 25-30N and about 1000m depth. The strength of the AMOC at 26N is close to the observed value (16Sv compared to 18Sv). Plotting the 10 historical members of the LR model (pink envelop) and comparing to CNRM-CM6-1-HR indicates that in the upper 2000m the results are close and comparable to observations but below 2000m the AMOC in CNRM-CM6-1-HR is too weak. This might be due to the lack of parameterization of mesoscale eddies in CNRM-CM6-1-HR.

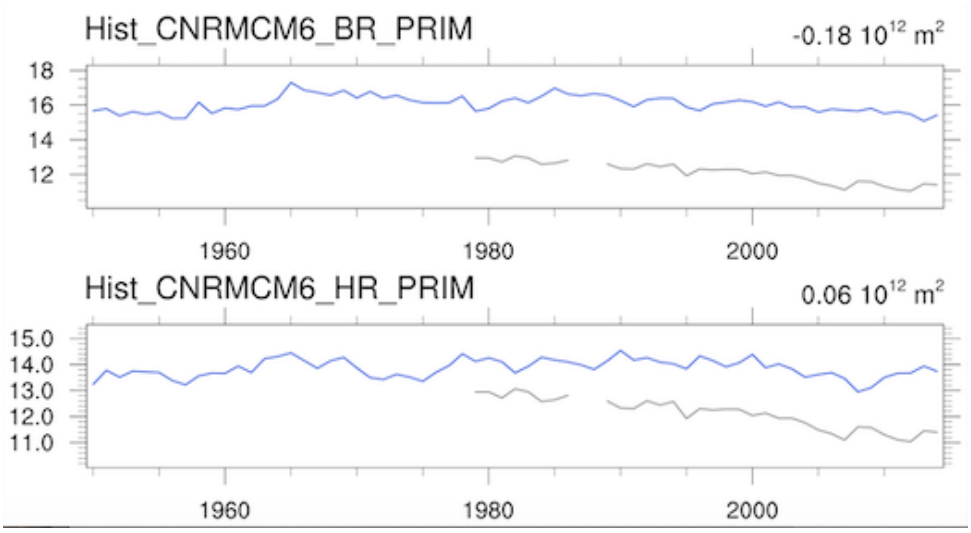

Figure 8 - Timeseries of Sea Ice Extent (SIC) annual mean for the LR model (top) and for CNRM-CM6-1-HR. NASA bootstrap observations are also shown in grey.

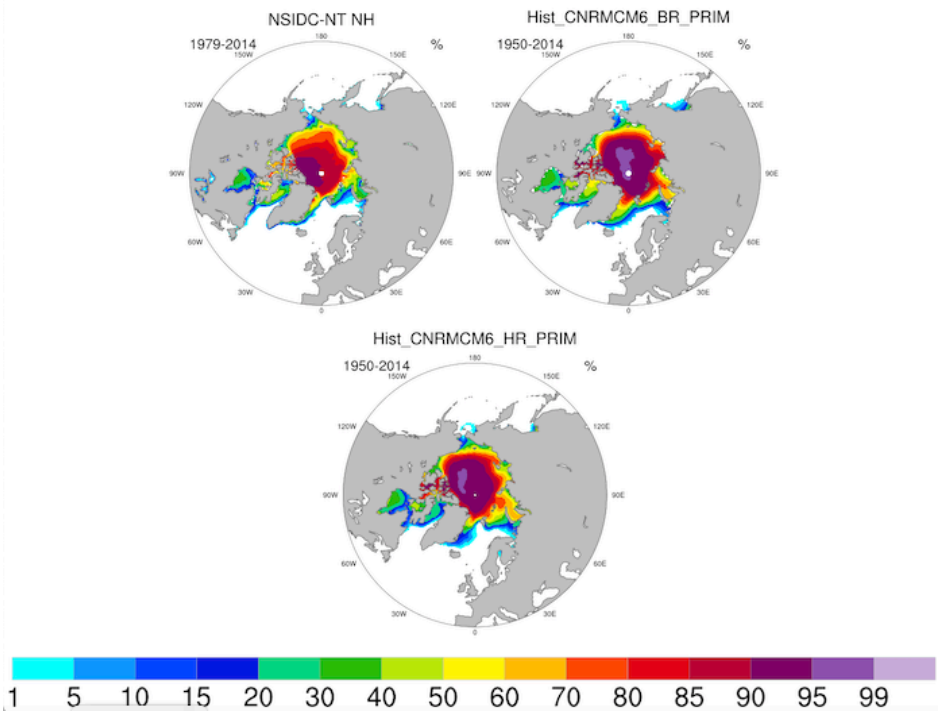

Figure 9 – Sea Ice Extent over 1950-2014 for CNRM-CM6-1-HR (bottom), for the low-resolution version (top right), and over 1979-2014 from NSIDC observation (top left).

Figure 8 and 9 representing the Sea Ice Extent (SIC) show that the ice is reasonably reproduced and that the excess of ice in the LR model is significantly reduced in CNRM-CM6-1-HR.

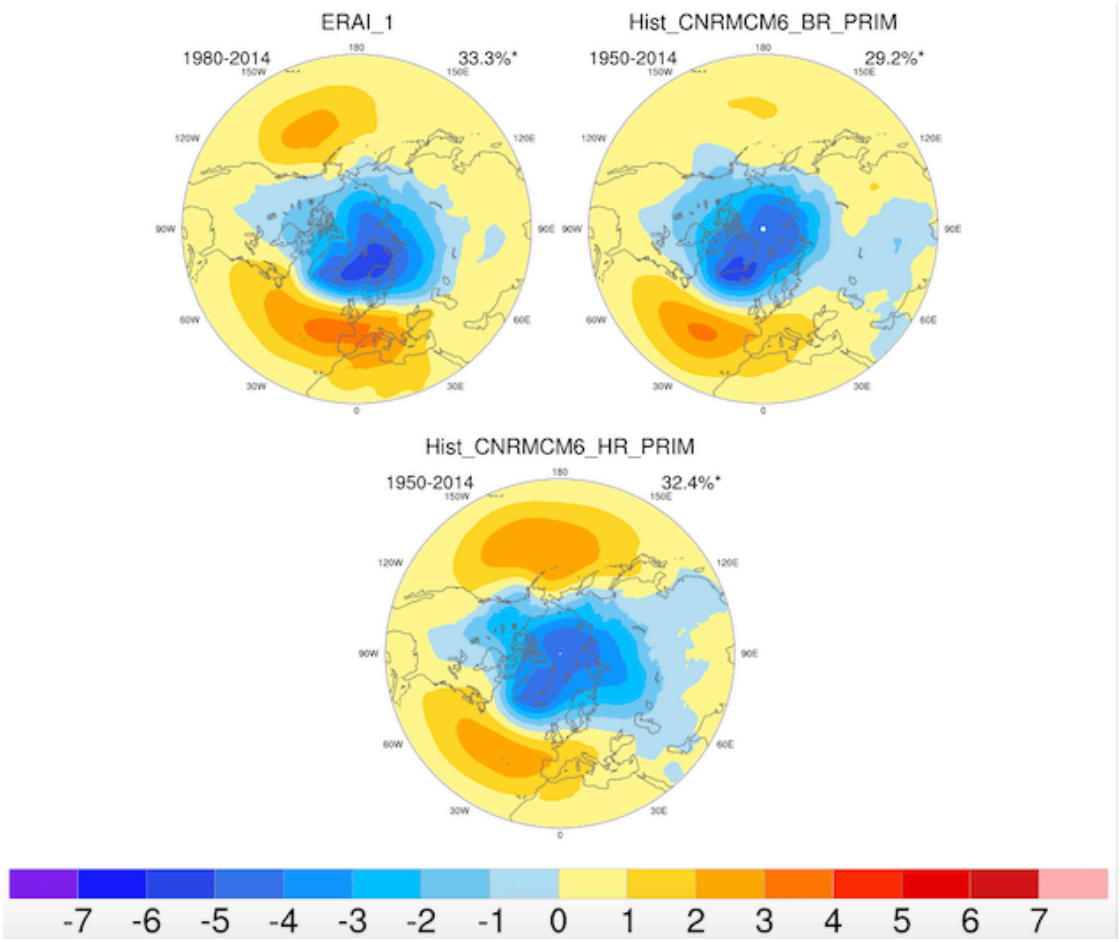

Figure 10 – Northen Annular Mode (NAM) pattern for DJF over the 1950-2014 period for CNRM-CM6-1-HR (bottom), CNRM-CM6-1 (top right) and over 1980-2014 for ERAI (top left)

Figure 10 illustrates the Northern Annular Mode (NAM) (for details of the calculations see http://www.cgd.ucar.edu/cas/asphilli/Results/CVDP-PDO/1920-2100/remove_EM/methodology.html) that characterizes the variability of the model. We see that this mode of variability is present in both versions of the model with a better representation of the Pacific node in CNRM-CM6-1-HR.

5. Appendix A – CNRM-CM6-1 model description

This appendix duplicates the current status of section "2 Model description" of a coming paper Voldoire et al 2019 , describing the low-resolution version of CNRM-CM6-1

2.1 – Atmospheric component ARPEGE-Climat

The atmospheric component of CNRM-CM6-1 is based on the version 6.3 of the global atmospheric model ARPEGE-Climat. Former version 5.2 of ARPEGE-Climat has been described in Voldoire et al. (2013). A summary of the the main characteristics of the new version is presented in the following with the emphasis on major updates. ARPEGE-Climat is a spectral model derived from the ARPEGE/IFS (Integrated Forecast System) numerical weather prediction model developed jointly by Météo-France and European Center for Medium-range Weather Forecast (ECMWF). A linear triangular truncation Tl127 is considered together with a corresponding reduced Gaussian grid (Hortal and Simmons 1991) at a 1.4 degree resolution. The CNRM-CM6-1 atmospheric component is run in a "high-top" configuration with 91 vertical levels (31 for CNRM-CM5.1), following a progressive hybrid σ-pressure discretization; the highest level is set at 0.01 hPa, with 15 levels below 1500 m.

The dynamical core is based on a two time-level semi-Lagrangian numerical integration scheme tagged as cycle 37 of the ARPEGE/IFS system. A 15-minute time step is used except for the radiative transfer module called every 1 h for full computation. In addition to the six prognostic variables already considered in the previous version (temperature, specific humidity, ozone concentration, logarithm of surface pressure, vorticity and divergence), the model includes ten new prognostic variables (cloud and precipitating solid and liquid water contents for both stratiform and convective parts, turbulent kinetic energy (TKE) and convective vertical velocity). As the semi-lagrangian dynamical core is not fully conservative (Lucarini and Ragone 2011), a global mass and water conserving procedure is activated at each time step after application of a local correction to the water contents, following Bermejo and Conde (2002). Note that there is no correction applied on heat.

The same radiation parameterization as the one used in CNRM-CM5.1 is operated, i.e. the Rapid Radiation Transfer Model (RRTM, Mlawer et al. 1997) for the longwave part of the spectrum and a 6-band shortwave scheme originally developed by Fouquart and Bonnel (1980). Some refinements have been included: the major improvement consists of a new aerosol climatology (together with revised optical properties and first indirect effect) based on results coming from an interactive aerosol simulation (paper in preparation). The water content cloud optical properties have also been updated.

The major change versus CNRM-CM5.1 consists of the inclusion of new schemes for the rest of the physics. The convection scheme is based on the work of Piriou et al. (2007) and Guérémy (2011) (new paper in preparation). This scheme provides a continuous and prognostic treatment of convection from dry thermals to deep precipitating events. Entrainment and detrainment depend on the vertical velocity, that is a prognostic variable, and follow a buoyancy sorting mechanism. The closure is based on a dilute CAPE relaxation. The stratiform microphysics scheme was designed

following the work of Lopez (2002). It takes into account autoconversion, sedimendation, ice-melting, precipitation evaporation and collection. Bouteloup et al. (2011) developed a probabilistic approach for the sedimentation, which allows the use of longer time-steps than the original lagrangian one. Turbulence scheme follows the approach of Cuxart et al. (2000), which represents the TKE with a 1.5-order scheme prognostic equation. The non-local mixing length is based on Bougeault and Lacarrere (1989). A specific treatment of the entrainment at the top of the boundary layer is taken into account based on Grenier and Bretherton (2001). This turbulence scheme also diagnoses the grid-scale variance of the local distance to saturation (expressed in terms of total water and liquid-water potential temperature turbulent fluctuations), which in turn allows the computation of both stratiform cloudiness and water content in a consistent way, following the work of Sommeria and Deardorff (1977). This resulting stratiform water content defines the corresponding condensation tendency which enters into the microphysics scheme giving rise to prognostic stratiform water contents. Finally, a non-orographic gravity wave drag parameterization has been introduced in CNRM-CM6-1 (following Lott et al., 2012), in addition to its pre-existing orographic gravity wave drag counterpart.

2.2 – Surface component SURFEX

SURFEX is a numerical platform that deals with surface fluxes over all surface types (Masson et al., 2013). It operates on the same grid and with the same time-step as ARPEGE. Three surface types are considered: land (including urban area treated as rock surface), lakes and ocean.

The land surface is represented using the new ISBA-CTRIP coupled system. ISBA calculates the time evolution of the energy and water budgets at the land surface while CTRIP simulates river discharges up to the ocean from the total runoff computed by ISBA. ISBA explicitly solves the one-dimensional Fourier and Darcy laws throughout the soil using 14-layers down to 12m depth and accounting for the hydraulic and thermal properties of soil organic carbon. The snow is represented by a 12-layers detailed internal-process snow model including a simple ice-sheet runoff to avoid unrealistic snow accumulation over continental glaciers. Plant transpiration is controlled by the stomatal conductance of leaves, which depends on carbon cycling in vegetation. A two-way coupling between ISBA and CTRIP is set up to account for, first, a dynamic river flooding scheme in which floodplains interact with the soil and the atmosphere through free-water evaporation, infiltration and precipitation interception and second, a two-dimensional diffusive groundwater scheme to represent unconfined aquifers and upward capillarity fluxes into the superficial soil. The land surface in one grid-cell is tiled into 12 patches in order to account for the variety of soil and vegetation behavior within a grid point. They aggregate the 500 land cover units at 1-km resolution present in the ECOCLIMAP-II database (Faroux et al., 2013). Mean seasonal cycles of both the snow-free albedo and the leaf area index are prescribed from Moderate Resolution Imaging Spectroradiometer (MODIS) products at 1-km spatial resolution. The soil textural properties (clay, sand, and soil organic carbon content) are given by the Harmonized World Soil Database (HWSD; http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/) at a 1-km resolution. Finally, the mean topography is derived from the 1-km Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010;

https://topotools.cr.usgs.gov/gmted_viewer/). More details can be found in Decharme et al. (2018).

Lakes, including both the Caspian and the Aral seas, are represented using the bulk FLAKE model which computes the temporal evolution of the vertical lake temperature profile from the surface mixing layer to the bottom, with the lake eventually becoming covered in ice and snow. A skin temperature of a 1-mm thickness has been introduced to simulate a surface temperature representative of the energy budget at the lake surface. The spatial distribution of the lake at the global scale is given by the ECOCLIMAP-II database while the lake depth has been specified from the 1-km Global Lake Depth database (Kourzeneva, 2010). More details can be found in Le Moigne et al. (2016). However, this model does not account for water mass conservation, thus the imbalance between precipitation, runoff and evaporation is artificially spread uniformly over the ocean water in CNRM-CM6-1.

Over the ocean, SURFEX resolves the exchange of energy and water across the air-sea interface. The radiative properties of surface sea water are handled by the ocean surface albedo scheme as proposed by Séférian et al. (2017). The turbulent fluxes of momentum, heat and water are computed using an improved version of the Exchange Coefficients from Unified Multi-campaigns Estimates (ECUME) scheme. It is a bulk iterative parameterization developed at CNRM from in situ measurements. This ECUME formulation is considered to better replicate the variability of the observed turbulent fluxes across a wide range of atmospheric and oceanic conditions.

2.3 – Ocean component NEMO

The ocean component of CNRM-CM6-1 is based on the version 3.6 of NEMO (Nucleus for European Models of the Ocean ; Madec, 2008). It is based on the NOCS-ORCA1 configuration described in details in Danabasoglu et al. (2014). Main differences are highlighted below.

In CNRM-CM6-1, NEMO is run on eORCA1 horizontal grid, which is an extension of the ORCA1° tripolar grid already used in CNRM-CM5.1. The eORCA family differs from the ORCA family by the use of two quasi-isotropic bipolar grids south of 67°S instead of the former Mercator grid, which allows for a more realistic representation of the contours of Antarctic ice shelves (Mathiot et al., 2017). In eORCA1, a nominal resolution of 1° is chosen to which a latitudinal grid refinement of 1/3° is added in the tropics.

CNRM-CM6-1 resolves ocean dynamics on 75 vertical levels using a vertical z-coordinate with partial step bathymetry formulation (Barnier et al., 2006). The level thickness increases from 1m near the surface to 200 m at a depth of 6000 m. The time step is 30 minutes. At the surface, the model uses the split-explicit non-linear free surface formulation proposed by Shchepetkin and McWilliams (2005), with a variable volume. Seawater thermodynamics uses the equation of state defined in the Thermodynamic Equation of State 2010 (TEOS-10, IOC et al., 2010), with conservative temperature and absolute salinity being then the prognostic variables.

Radiative transfer in the water column is resolved using a chlorophyll-dependent three-waveband scheme as described in Lengaigne et al. (2007) and Mignot et al. (2013), using a seasonal climatology of surface chlorophyll concentration derived from a former 60-year long simulation run with NEMO-PISCES (e.g. Lee et al., 2014). A vertical profile of chlorophyll concentration is extrapolated from surface concentrations.

Lateral diffusivity and viscosity are parameterized as in NOCS-ORCA1 (Danabasoglu et al., 2014). Parameterization of vertical mixing is also similar, with two notable exceptions in CNRM-CM6-1: mixing induced by breaking internal waves is parameterized following de Lavergne et al. (2016) and the use of the Fox-Kemper et al. (2011) submesoscale mixed layer restratification scheme.

2.4 – Sea ice component GELATO

Sea ice within CNRM-CM6-1 is represented by Gelato 6. Most upgrades from version 5 of the code (Voldoire et al., 2013, Chevallier, 2012) aimed at improving the overall consistency of the code (in particular, salt, water and energy conservation) and increasing its computational efficiency. In contrast with Gelato 5, Gelato 6 is a fully parallel code. It is completely embedded in NEMO, has the same horizontal grid and inherits the global domain decomposition of the ocean code.

Within the standard configuration of CNRM-CM6-1, Gelato has a time-step of 3600 s. It is used with five thickness categories, based on the World Meteorological Organization classification (less than 0.30 m, 0.3-0.7 m, 0.7-1.2 m, 1.2-2 m and over 2 m thick). The snow and ice parts of every ice category are respectively split vertically into 1 and 9 layers. The sea ice enthalpy formulation of Gelato is based on Notz et al. (2005). As in Hunke et al. (2010), an iterative method is used to solve the vertical heat diffusion equation in sea ice. The solar radiation transmission scheme through the snow pack covering sea ice was upgraded following Grenfell and Maykut (1977). Albedo of dry snow, melting snow and melting ice are model parameters, respectively equals to 0.88, 0.77 and 0.58 in CNRM-CM6-1.

The elastic–viscous–plastic (EVP) sea ice rheology implemented by Bouillon et al. (2009) on an Arakawa C-grid is used, consistent with the formulation of NEMO finite difference scheme, whereas it was solved on an Arakawa B-grid in Gelato 5. The sea ice transport follows an incremental remapping scheme based on an Arakawa-B formulation by Hunke and Dukowicz (1997). It was adapted to an Arakawa C-grid according to previous work from M. Bentsen (Univ. Bergen, Norway, personal communication) to ensure a better consistency between the dynamics and transport, which is crucial to represent sea ice transport through one-grid-cell-wide straits. The transported variables are snow density, volume and enthalpy, and ice surface, volume, enthalpy, salinity and age. Depending on sea ice thickness, rafting and ridging may take place in case of sea ice convergence, as described in Salas y Mélia (2002).

Some aspects of the ice-ocean interface have been revised. In the quadratic bulk formula used to calculate the ice-ocean stress for ice-ocean momentum exchanges, the drag coefficient is set to 1.0x10-3. This value is higher to that used in CNRM-CM5.1 and other models (e.g. Vancoppenolle et al., 2009), in order to take into account that the top ocean level now resides within the relatively thin surface layer (~1-3 m thick), following Roy et al. (2015). The chosen value is a compromise between this rationale and stability constraints. The oceanic heat flux at ice bottom derives from McPhee (1992).

2.5 – Coupling OASIS

The model components are coupled using the OASIS3-MCT software (Craig et al., 2017), which implements exchanges between multiple executables running concurrently using Message Passing Interface (MPI) communication. The OASIS3-MCT approach has the advantage of requiring a minimal amount of modifications in existing

component codes. In CNRM-CM6-1, OASIS3-MCT transfers and interpolates coupling fields between SURFEX, CTRIP and NEMO at a coupling frequency of one hour. In contrast the atmospheric model ARPEGE and SURFEX multi-surface model are coupled inline, i.e. SURFEX is called as a subroutine of ARPEGE at each atmospheric time step, and are thus considered as one executable by OASIS3-MCT; similarly, NEMO and GELATO are also coupled inline at one hour frequency.

NEMO sends the ocean and sea-ice surface properties (sea surface temperature and currents, sea-ice fraction and sea-ice albedo) that are used in SURFEX to calculate surface heat and momentum fluxes which are returned back to NEMO. The same is done for the water cycle with the addition of CTRIP that routes continental water to the ocean. When entering NEMO, river runoffs are spread vertically depending on the flow intensity to a maximum depth of 10m. Additionally, SURFEX provides a net water budget over continental ice sheets and lakes to NEMO to close the water budget. This water is incorporated at the ocean surface over the globe for the Greenland ice sheet and lakes and only south of 60S for the Antarctic ice sheet.

The energy associated to water fluxes is taken into account in NEMO/GELATO. Liquid precipitation and river runoff enthalpy is calculated considering that they are et sea surface temperature. Similarly, ice sheets water flux and snow falls are considered to be solid, with a 0°C temperature; thus the enthalpy flux corresponding to the melting of ice/snow is considered.

As in SURFEX, a tiling approach considering land, ocean and lake surfaces is used to calculate fluxes, the fluxes sent to NEMO are those calculated over the ocean fraction of the grid cell. Due to ocean coastline mismatch between NEMO and SURFEX, a global conservation procedure is applied at each coupling time-step to ensure energy and water conservation in the coupled system.

More details about the OASIS3-MCT implementation in SURFEX can be found in Voldoire et al. (2017).

2.6 – Output server XIOS

CNRM-CM6-1 is used in an increasing number of Model Intercomparison Programs which request numerous diagnostics with strong format requirements. In order to ease their production, the model was interfaced with XIOS [1], an Input/Output parallel server software allowing for a declarative description of output file content and realisation of online field operations, thereby reducing, or even completely removing, the need of post-processing.

Once the distribution of the full grid among processes has been described, the XIOS library has a main interface routine (send_field) which allows each model MPI process to deliver very simply its part of the full grid for any field. Routine send_field can be called from any model routine.

XIOS post-processing operations include time sampling and averaging, spatial remapping and reduction, vertical interpolation and simple arithmetic. They are configured at run time using an XML syntax. The processing is actually shared between the model MPI tasks and a number of XIOS additional MPI tasks, called servers. Two levels of servers are used to aggregate and redistribute the parallel output fields so that any given field is gathered into one file written to disk by one single MPI task (thus avoiding parallel writing). Special care was taken in XIOS to allow for an overlap of model computation and I/O operations using communication buffers so to allow for scaling at high number of cores.

This approach shows its obvious interest, as compared to the classical offline post-processing approach, in the case of CMIP6 Data Request (DR, https://earthsystemcog.org/projects/wip/CMIP6DataRequest), which addresses 248 experiments; it involves 1274 geophysical variables which can be invoked in 44 so-called tables MIP tables, thus leading to 2027 'CMOR variables' covering a variety of spatial shapes and frequencies. An additional tool, called dr2xml (https://github.com/senesis/dr2pub), translates the CMIP6 DR for each year of each experiment in a set of XIOS conformant XML definitions. These definitions are then used to activate the outputs in CNRM-CM6-1 components, which are all XIOS-enabled thanks to an alias table describing the correspondence between model variable names and CMIP6 DR variable names. At run time, CNRM-CM6-1 then creates NetCDF output files which are directly conformant with CMIP6 requirements and ready to publish on the Earth System Federated Grid (ESGF), except for some climatologies. This approach is applicable to any other model and is currently also used in the Institut Pierre Simon Laplace (IPSL) climate model.

6. Appendix B – ECLIS configuration file for CNRM-CM6-1-HR historical run

This appendix provides ECLIS configuration file for CNRM-CM6-1-HR historical run, param_CPLHR_HIST1950_P1_v1, which can be found in /home/ext/cf/cglo/valckes/SAVE/param/WP6_STREAM1 on beaufix .

==========

```
#!/bin/bash
# ARPEGE V6.3.1, tl359L91
# NEMO V3.6
CMIP6=/scratch/CMIP6/V1
ECLIS=$CMIP6/eclis
CONFIG=AOGCM
[ -z $EXPID ] && EXPID=$(basename $0 | sed -e 's/param_//g')
GROUP=PRIMAVERA
GEOM=tl359l91r
GEOMH=tl359
GEOMO=eORCA025L75
TITRE="$GROUP $CONFIG $GEOM $GEOMO"
RUNMAIL=valcke@cerfacs.fr
INIDATE=19500101 ; ENDDATE=20141231 ; INITIME=0

#               MODEL BINARY USED.
bindir=$CMIP6/bin
ATMEXE=${bindir}/atm/MASTER
UPDCLIARP=${bindir}/atm/UPDCLIARP
UPDCLISFX=${bindir}/atm/UPDCLISFX
RIVEXE=${bindir}/riv/TRIP_MASTER
OCEEXE=/home/ext/cf/cglo/valckes/SAVE/NEMO/bin/nemo_ORCA025_GLT_CPL_CM6
b2_cpl_xios1442-shuffle_f845920e7f2169d_8250e198106a168.exe
IOSEXE=${bindir}/ios/xios_server.exe

#               NAMELISTS USED
ATMNAMREF=~valckes/SAVE/public/arpege/namelist/nam.atm.tl359l91r.AGCM_SV
SFXNAMREF=~voldoire/SAVE/surfex/pgdprep/arp6215_HR/OPTIONS.nam
RIVNAMREF=${CMIP6}/namelist/riv/TRIP_OPTIONS.nam.tl127.${CONFIG}
CPLNAMREF=~valckes/SAVE/oasis/data/namcouple_9_HR.xios.opt.grp
OCENAMREF=~valckes/SAVE/public/nemo/namelist/namoce_eORCA025_primavera_d
iadct.v0_SV_EN4corr_530
ICENAMREF=~chevalli/SAVE/public/NEMO3.6/data_v3_6_STABLE/namelists/gelato/n
amgel_HR_cpl.1

#               XIOS
LIOXOUT=1 ; LRIVIOS=1 ; LSFXIOS=1 ; LOCEIOS=1 ; LICEIOS=1
```

```
XMLS=${CMIP6}/namelist/ios
XMLS_MINE=/home/ext/cf/cglo/valckes/SAVE/public/xios_xmls/f845920e7f2169d43
030aa80477ddfe91b7e33dd
IOSNAMREF="${XMLS}/iodef.xml"
OTHER_FILES="${XMLS}/arpsfx.xml ${XMLS}/atmo_fields.xml ${XMLS}/aero_fields.xml
${XMLS}/chem_fields.xml ${XMLS}/surfex_fields.xml"
OTHER_FILES+=" ${XMLS}/trip.xml ${XMLS}/trip_fields.xml"
OTHER_FILES+=" ${XMLS_MINE}/nemo.xml ${XMLS_MINE}/nemo_fields.xml
${XMLS}/nemo_domains.xml"
IOXSAVEPER=24
#
#                DR
LDR=1
LDR_EXPID=1
DIR_DR2XML=$CMIP6/bin/dr2xml
alt="altdr2xmlpath=/scratch/work/moinemp/CMIP6/V1/bin/dr2xml_V1.8_perso"
DR2XML="$alt $DIR_DR2XML/create_file_defs.sh skip"
#
DIR_SETTINGS=/home/ext/cf/cglo/moinemp/SAVE/public/DR_settings/dr2xml_CMIP
6_V1b_xios1442
DR_EXP_SET=$DIR_SETTINGS/hist-1950_settings_HR.py
DR_LAB_SET=/home/ext/cf/cglo/valckes/SAVE/public/DR_settings/dr2xml_CMIP6_V1
b_xios1442/settings_CNRM-CERFACS_DRTerrayMoine5.py
OTHER_FILES+=" $XMLS/ping_surfex.xml $XMLS/ping_trip.xml"
OTHER_FILES+=" ${XMLS_MINE}/ping_nemo.xml ${XMLS_MINE}/ping_nemo_gelato.xml
${XMLS_MINE}/ping_nemo_ocnBgChem.xml"
#
PFX=${DIR_SETTINGS}/home_data_request
HOMEDR="${PFX}_arpege_GCM.txt ${PFX}_surfex_GCM.txt ${PFX}_trip_GCM.txt
${PFX}_nemo_GCM.txt.MC_22052018 ${PFX}_PRIMAVERA_TerrayMoine.txt"
PATH_EXTRA_TABLES=${DIR_SETTINGS}/Tables_TerrayMoine
$DATA_DRX/areacella_complete_CMIP6_$GEOMH.nc"
$DATA_DRX/xios_interpolation_weights_surfex_FULL_cfsites_domain.nc"
#
#                RESTART FILES
LICEREST=1
LOCEREST=1
EXPREF=CNRM-CM6-1-HR_spinup-1950_r1i1p1f2 ; DATREF=19800101
MACH_RESTART=hendrix
restarts=/home/c/cgie/cgie006/PRIMAVERA/${EXPREF}
ATMREST=${restarts}/atm/restart/${EXPREF}.rst.atm.P${DATREF}
SFXREST=${restarts}/sfx/restart/${EXPREF}.rst.sfx.P${DATREF}.gz
SFXPGD=${restarts}/sfx/restart/PGD.fa.gz
RIVREST=${restarts}/riv/restart/${EXPREF}.rst.trp.P${DATREF}.nc
OCEREST=${restarts}/oce/restart/${EXPREF}.rst.oce.P${DATREF}.nc.tar
ICEREST=${restarts}/ice/restart/${EXPREF}.rst.ice.P${DATREF}.nc.tar
CPLATMRES=${restarts}/cpl/restart/${EXPREF}.cpl.atm.P${DATREF}.nc
CPLOCERES=${restarts}/cpl/restart/${EXPREF}.cpl.oce.P${DATREF}.nc
CPLRIVRES=${restarts}/cpl/restart/${EXPREF}.cpl.trp.P${DATREF}.nc
```

```
#           BC
BCOND=${CMIP6}/data/atm/nclim4/nclim4_${GEOMH}l31r_mMM ; YEAR_BCD=no
DATA_SFX=${CMIP6}/data/sfx/ecoclimap*.bin
DATA_RIV=${CMIP6}/data/riv
DATA_CPL=~voldoire/SAVE/oasis/data/arp6.2.9-sfxmct_nem3.6_HR_4
DATA_OCE=~valckes/SAVE/public/NEMO3.6/data_v3_6_STABLE/data_eORCA025L75_
cmip6.3
LCHECKFOR=0

#           FORCINGS
#
aero=${CMIP6}/data/atm/FORAER/TACTIC2.3/${GEOMH}/AOD550espece_TACTIC2.3_
11avg_YYYYMM.ieee
aero=${CMIP6}/data/atm/FORAER/TACTIC2.3/${GEOMH}/AOD550espece_TACTIC2.3i
nt359_11avg_YYYYMM.ieee
YEAR_SUL=SCEN_1850_2014 ; FORSUL=${aero/espece/S4}
YEAR_BCA=SCEN_1850_2014 ; FORBCA=${aero/espece/BC}
YEAR_ORA=SCEN_1850_2014 ; FORORA=${aero/espece/OM}
YEAR_SDA=SCEN_1850_2014 ; FORSDA=${aero/espece/DD}
YEAR_SSA=SCEN_1850_2014 ; FORSSA=${aero/espece/SS}
YEAR_VOL=SCEN_1850_2014 ;
FORVOL=${CMIP6}/data/atm/FORVOL/${GEOMH}/aod_volcan_strato_v3_YYYYMM_${G
EOMH}r.ieee #MPM
YEAR_OZO=SCEN_1850_2014 ;
FOROZO=${CMIP6}/data/atm/FOROZO/ozone_slarp_${GEOM}/ozone.${GEOM}.7coeffs.
YYYYMM.ieee
YEAR_GHG=SCEN_1850_2014 ; FORGHG=${CMIP6}/data/atm/FORGHG/GHG_HIST.dat

#           JOBS CHARACTERISTICS
NMONTH=12 ; QUEUE=normal64 ;
ELAPSFRONT=08:00:00 ; MEMFRONT=100Mb
ELAPS=18:00:00 ; MEM=60Gb
NPROC_ARP=768
NPROC_RIV=1
NPROC_IOS=48
NPROC_CPL=0
NPROC_OCE=530

SAVE_RESTART_PER=12
SAVE_LISTING="AX" # XXX Remove for optimization?
SAVE_CPL_FILES=0
ACCOUNT=cgie
FTPUTOPT="-u cgie006"
FTGETOPT="-u cgie006"
ARCHIVING=DURING

INSTALLER=${ECLIS?}/scripts/install
[[  "$*"  != *nogo* ]] && . $INSTALLER $*
```

7. Acknowledgements

The work described in this document was partly supported by the CONVERGENCE project funded by the French National Research Agency: ANR-13-MONU-0008

8. References

[Craig et al 2017] A. Craig, S. Valcke, L. Coquart, 2017: Development and performance of a new version of the OASIS coupler, OASIS3-MCT_3.0, Geosci. Model Dev., 10, 3297-3308, https://doi.org/10.5194/gmd-10-3297-2017.

[Good et al 2013] Good, S. A., M. J. Martin and N. A. Rayner, 2013. EN4: quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates, Journal of Geophysical Research: Oceans, 118, 6704-6716, doi:10.1002/2013JC009067