



Documentation of the METEO-FRANCE seasonal forecasting system 7

C3S_330_Météo-France - Operational Seasonal Forecasts

Issued by: Météo-France / Lauriane Batté

Date: 30/06/2019

Ref: C3S_D330_3.1.1_201906_System 7 documentation_v1

Official reference number service contract: 2018/C3S_330_Météo-France/SC1

This document has been produced in the context of the Copernicus Climate Change Service (C3S). The activities leading to these results have been contracted by the European Centre for Medium-Range Weather Forecasts, operator of C3S on behalf of the European Union (Delegation Agreement signed on 11/11/2014). All information in this document is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability. For the avoidance of all doubts, the European Commission and the European Centre for Medium-Range Weather Forecasts has no liability in respect of this document, which is merely representing the authors view.



Contributors

METEO-FRANCE

Lauriane Batté

Laurent Dorel

Constantin Ardilouze

Jean-François Guérémy



Table of Contents

1. Basic Facts	5
2. Modeling components of the forecast system	8
2.1 Introduction	8
2.2 The atmospheric model, ARPEGE V6.4	8
2.3 The ocean model NEMO V3.6 and sea-ice model GELATO V6	10
2.4 The Land Surface model Surfex V8 and river routing model CTRIP	10
2.5 The coupler OASIS3-MCT	11
2.6 ECLIS	12
2.7 XIOS	12
3. Generation of stochastic perturbations	13
3.1 Introduction	13
3.2 The nudged pre-hindcasts: generating the perturbation fields	13
3.3 Random calendar and definition of ensembles	15
3.4 Model implementation	16
4. Generation of initial conditions	18
4.1 Fetching the initial conditions (hindcast)	19
4.2 Fetching the initial conditions (forecast)	20
5. Environment for CLimate Simulation (ECLIS)	22
5.1 ECLIS organization	22
5.2 Use of ECLIS	22
5.3 Design a param_ file	23
5.4 Standard plugins	23
6. The basic set of scripts	24
7. Post processing	25
8. References	27



Introduction

Météo-France has been involved in real-time seasonal forecasting activities since the late 1990s. As part of the Eurosip consortium, the seasonal forecast system at Météo-France has evolved through the development of several consecutive versions, including the latest developments of the components of the coupled GCM as well as new ensemble initialization and generation techniques.

The current operational system is Météo-France System 6, developed in the framework of the previous Copernicus Climate Change Service C3S-433 contract for the pre-operational phase of the provision of seasonal forecasts.

The aim of this document is to present the forthcoming Météo-France System 7, which is planned to be operational from autumn 2019 onwards. We will first present a brief overview of this new system and describe the main features of each individual component. Then, similarly to the documentation for System 6, we will detail the procedures for the ensemble generation as well as the initial states for the atmosphere, ocean, sea ice and land surface components. We will then describe the environment software ECLIS and basic setup scripts used to launch the forecasts and re-forecasts, as well as the forecast post-processing now mostly managed by XIOS.

This document does not address the atmosphere and ocean data assimilation techniques which are carried out respectively by ECMWF and MERCATOR-OCEAN.



1. Basic Facts

The table below gives the main features of Météo-France System 7.

1. Table 1 : basic facts

Météo-France System 7	
Ensemble version	
Ensemble version identifier code	CNRM-CM 6
Short description	Global ensemble system using a lagged-average and a stochastic scheme to take into account initial state and model uncertainties. Based on 51 members, run once a month up to 7 months.
Status	Operational
Data time of the first forecast run	01/10/2019
Configuration of the EPS	
Is the model coupled to an ocean model?	Yes from day 0
Short description of the ocean model	NEMO v3.6 , ORCA 0,25° grid, 75 model levels
Is the model is coupled to a sea ice model?	Yes
Short description of the sea ice model	GELATO v6 (Salas y Melia, 2002), embedded in the ocean model. It is initialized from unperturbed 0.25 degree resolution MERCATOR-OCEAN Ocean and Sea-Ice Analysis
Is the model coupled to a wave model?	No
Horizontal resolution of the atmospheric model	TL359
Number of atmospheric model levels	91
Top of the model	0.01 hPa
Type of model levels	Hybrid sigma-pressure
Forecast length	7 months
Run frequency	Once a month
Is there an unperturbed control forecast included	No
Number of perturbed ensemble members	51



Integration time step	Atmosphere/surface: 10' Ocean/sea-ice: 15' Coupling frequency: 1 hour
Initial conditions and perturbations	
Data assimilation method for analysis	4D Var
Resolution of the model used to generate Control Analysis	O1280L137 (IFS operational analysis)
Ensemble initial perturbation strategy	Lagged-average and in-run perturbations
Model uncertainties perturbations	
Is model physics perturbed?	No
Do all ensemble members use exactly the same model version?	Yes
Is model dynamics perturbed?	Yes (Batté and Déqué 2016)
Are the above perturbations applied to all forecast members?	Yes
Surface boundary perturbations	
Perturbation to sea surface temperature?	No
Perturbation to soil moisture?	No
Perturbation to surface stress or roughness?	No
Any other surface perturbation?	No
Other details of the models	
Description of the model grid	Reduced Gaussian Grid
List of model levels	From top to bottom (Pa) 1, 3, 6, 10, 17, 27, 42, 63, 92, 129, 177, 237, 311, 401, 508, 633, 779, 946, 1136, 1349, 1587, 1850, 2139, 2454, 2796, 3165, 3560, 3982, 4430, 4904, 5403, 5927, 6475, 7046, 7638, 8252, 8890, 9555, 10251, 10982, 11752, 12566, 13431, 14350, 15326, 16362, 17461, 18628, 19865, 21176, 22565, 24037, 25595, 27245, 28990, 30836, 32788, 34851, 37031, 39333, 41763, 44324, 47007, 49787, 52637, 55531, 58441, 61342, 64222, 67067, 69864, 72601, 75262, 77836, 80312, 82678, 84922, 87035, 89011, 90842, 92520, 94042, 95408, 96616, 97666, 98562, 99329, 99983, 100512, 100914, 101204
What kind of large scale dynamics is used ?	Spectral semi-lagrangian



What kind of boundary layer parametrization is used ?	Cuxart, Bougeault and Redelsperger (2000)
What kind of convective parametrization is used ?	Gueremy (2011), Piriou et al. (2007)
What kind of large scale precipitation scheme is used ?	Lopez (2002)
What cloud scheme is used ?	Sommeria and Deardorff (1977)
What kind of land-surface scheme is used ?	Explicit multilayer snow and soil scheme as described in Masson et al. (2013)
How is radiation parametrized ?	Long wave radiation : Mlawer et al. (1997) Short wave radiation : Morcrette (1990)
Re-forecast configuration	
Number of years covered	24 years (1993-2016)
Produced on the fly or fixed re-forecasts ?	fixed re-forecasts
Frequency	monthly
Ensemble size	25 members
Initial conditions	ERA5 (TI639L137) for atmosphere and land surface MERCATOR-OCEAN reanalyses for ocean (NEMO-GELATO ORCA0.25 forced run nudged toward NEMO 1/12° re-analysed (T,S))
Is the model physics and resolution the same as for the real-time forecasts?	Yes
Is the ensemble generation the same as for the real-time forecasts?	Yes



2. Modeling components of the forecast system

2.1 Introduction

As any long-range forecast system, Météo-France system consists of an ensemble forecast operational production together with an ensemble re-forecast dataset, also called hindcast.

Both ensembles come from integrations of the global coupled atmosphere/ocean/sea ice/land surface model CNRM-CM illustrated in Figure 1.

CNRM-CM runs at Météo-France on the beaufix and prolix supercomputers.

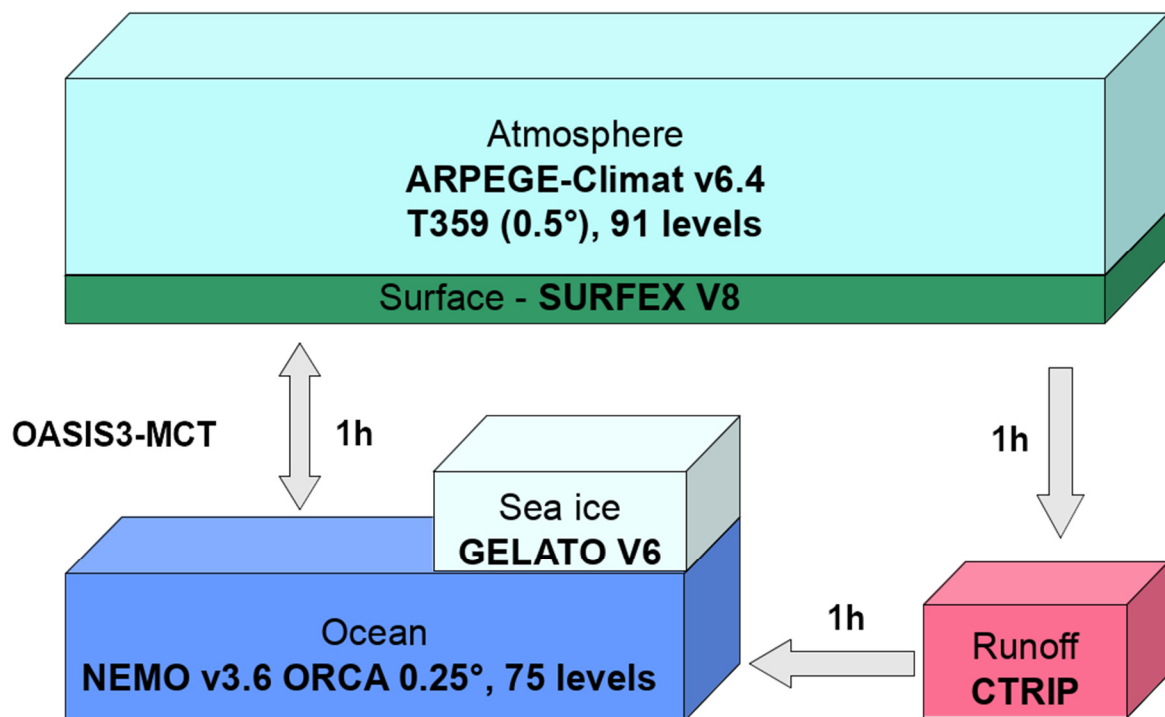


Figure 1 - CNRM-CM model components of the seasonal forecast system

2.2 The atmospheric model, ARPEGE V6.4

ARPEGE-IFS forecast model has been jointly developed by ECMWF and Météo-France for numerical weather prediction since the late 1980s (Courtier and Geleyn, 1988). A climate version of this model



was especially designed for climate change and seasonal predictability experiments by Déqué et al. (1994). The weather and climate versions have evolved in parallel since this date, including major improvements like semi-lagrangian advection and high performance parallelization. Except for radiation, ECMWF and Météo-France use different schemes for the physical parameterizations of the model. As for System 6, System 7 uses the physical parameterization schemes developed for the CMIP6 version of the CNRM-CM model (Voldoire et al. 2019).

This new set of parameterizations, called “prognostic physics”, rely upon the assumption that the diabatic terms are calculated for all atmospheric columns independently. But it includes a memory and a horizontal exchange of the diabatic processes through new prognostic variables: turbulent kinetic energy, convective vertical velocity, solid and liquid water in different states (cloud, large-scale precipitation and convection). These variables are advected and memorized during the whole simulation. Of course the parametrizations of convection (Guérémy, 2011), of turbulence (Cuxart et al., 2000) and of cloud and large-scale precipitation (Lopez, 2002) have been re-written to take into account the time evolution of these variables.

With respect to System 6 based on ARPEGE V6.2, only few changes, available in ARPEGE V6.4, have been activated in the physics. First of all, the turbulence has been enhanced in presence of convection through the direct inclusion of convective turbulent kinetic energy (TKE) dynamic and thermodynamic productions in the TKE time evolution equation, instead of the use of an ad-hoc minimum value of a turbulent variable (linked to the turbulence variance, and going into the cloudiness and TKE thermodynamic production computations) in presence of convection only in the upper part of the troposphere. Then, an improved discretisation of the convective vertical velocity equation has been used. Also, a new formulation of the surface momentum flux due to orography has been introduced, based on Beljaars et al. (2004), instead of the previous one taking an orographic roughness length into account. This method includes an analytical dependence to sinusoidal orography, and accounts for the sub-grid scale orography standard deviation.

The time step with respect to System 6 has been slightly lengthened for reasons related to computational cost, from 7.5 minutes to 10 minutes for all the physics (except for radiation which still uses a 1h time step).

A detailed presentation of ARPEGE equations is available here:

<http://www.umr-cnrm.fr/spip.php?article124&lang=en>

<http://www.umr-cnrm.fr/IMG/pdf/arp62ca.pdf>



2.3 The ocean model NEMO V3.6 and sea-ice model GELATO V6

The ocean engine of NEMO (Nucleus for European Modelling of the Ocean) is a primitive equation model adapted to regional and global ocean circulation problems. It is intended to be a flexible tool for studying the ocean and its interactions with the others components of the Earth climate system over a wide range of space and time scales. Prognostic variables are the three-dimensional velocity field, a non-linear sea surface height, the conservative temperature and the absolute salinity. In the horizontal direction, the model uses a curvilinear orthogonal grid and in the vertical direction, a full or partial step z -coordinate, or s -coordinate, or a mixture of the two. The distribution of variables is a three-dimensional Arakawa C-type grid. Various physical choices are available to describe ocean physics, including TKE, and GLS vertical physics. Within NEMO, the ocean is interfaced with a sea-ice model (GELATO V6), and, via the OASIS coupler, with the atmospheric model ARPEGE V6.4.

A more detailed presentation of NEMO is available here:

<http://www.nemo-ocean.eu/About-NEMO/Reference-manuals>

The reference paper for the current version of the model GELATO is Salas y Méliá (2002). Details on new developments in the GELATO model since this reference paper are provided in Voltaire et al. (2019). A different tuning of the turbulent vertical diffusion is used, with increased coefficients for thermodynamic and TKE (through factors included in the namelist) versus the dynamic.

2.4 The Land Surface model Surfex V8 and river routing model CTRIP

Surfex (Surface Externalisée, in French) is a surface modelling platform developed by Météo-France in cooperation with the scientific community.

Surfex is composed of various physical models for natural land surface, urbanized areas, lakes and oceans. It also simulates chemistry and aerosols surface processes and can be used for assimilation of surface and near surface variables.

Surfex has its own initialisation procedures and can be used in stand alone mode and coupled to an atmospheric model.

In Surfex, each model grid box is represented by four surface types : sea or ocean, water bodies (lakes, ...), urban areas and nature (soil and vegetation). Each surface type is modelled with a specific surface model and the total flux of the grid box results from the addition of the individual fluxes weighted by their respective fraction.

A more detailed presentation of Surfex is available here:



<http://www.umr-cnrm.fr/surfex/>

More details on the latest version of Surfex and the ISBA-CTRIP land surface and river routing models can be found in Voldoire et al. (2017) and Decharme et al. (2019).

2.5 The coupler OASIS3-MCT

OASIS is the coupler that manages the exchange of information between the three aforementioned model groups.

In 1991, CERFACS started the development of a software interface to couple existing ocean and atmosphere numerical General Circulation Models. Today, different versions of the OASIS coupler are used by about 45 modelling groups all around the world on different computing platforms. OASIS sustained development is ensured by a collaboration between CERFACS and the Centre National de la Recherche Scientifique (CNRS).

The current OASIS-MCT internally uses MCT, the Model Coupling Toolkit2 (Larson et al. 2005, Jacob et al. 2005), developed by the Argonne National Laboratory in the USA. MCT implements fully parallel regridding, as a parallel matrix vector multiplication, and parallel distributed exchanges of the coupling fields, based on pre-computed regridding weights and addresses. Its design philosophy, based on flexibility and minimal invasiveness, is close to the OASIS approach. MCT has proven parallel performance and is, most notably, the underlying coupling software used in National Center for Atmospheric Research Community Earth System Model (NCAR CESM).

OASIS-MCT is a portable set of Fortran 77, Fortran 90 and C routines. Low-intrusiveness, portability and flexibility are OASIS-MCT key design concepts. After compilation OASIS-MCT is a coupling library to be linked to the component models, and which main function is to interpolate and exchange the coupling fields between them to form a coupled system. OASIS-MCT supports coupling of 2D logically-rectangular fields but 3D fields and 1D fields expressed on unstructured grids are also supported using a one-dimension degeneration of the structures. Thanks to MCT, all transformations, including regridding, are performed in parallel on the set of source or target component processes and all coupling exchanges are now executed in parallel directly between the component processes via Message Passing Interface (MPI). OASIS-MCT also supports file I/O using netcdf.

The new version, OASIS3-MCT supports coupling exchanges between components deployed in much more diverse configurations than before. It is of course possible to implement coupling exchanges between two components corresponding to two different executables running concurrently on separate sets of tasks, as before, but also between two components running concurrently on separate sets of tasks within one same executable, or between different sub-components defined on separate



or overlapping sets of tasks within one executable. It is also now possible to have some or all tasks of a component not participating to the coupling exchanges.

In spite of the significant changes in underlying implementation, usage of OASIS3-MCT in the codes has largely remained unchanged with respect to previous OASIS 3 versions. To communicate with another component, or to perform I/O actions, a component model needs to include few specific calls of the Application Programming Interface (API) OASIS MCT coupling library. The namcouple configuration file is also largely unchanged relative to OASIS 3, although several options are either deprecated, not used or not supported.

A more detailed presentation of OASIS is available here:

https://verc.enes.org/oasis/oasis3-mct_3-0-official-release

2.6 ECLIS

The ECLIS (Environment for CLimate Simulations) environment is used to define and run the seasonal forecast experiments. The different model components (coupler included) each have a namelist, almost all have a restart file, and most have their own output format. Several binaries related to each model group are used. This information is stored in a so-called param file used by ECLIS. The binaries are launched together using a MPI message passing software. A CNRM-CM run is made of an automated sequence of "macro-jobs", where each "macro-job" iterates on a number of months.

2.7 XIOS

A novel aspect of System 7 is the use of the XIOS output manager for online processing of model component outputs (Meurdesoif, 2018). The XIOS software manages the output of each model on each process, and builds a single output file onto the disk. By means of xml files, many post-processing operations can be requested directly to XIOS (e.g. time sampling and/or averaging, regridding, vertical interpolation, ...). More details on the use of XIOS in System 7 are provided in section 7.



3. Generation of stochastic perturbations

3.1 Introduction

Atmospheric stochastic perturbations were introduced in Météo-France System 5. A detailed description of the technique can be found in Batté and Déqué (2016). The role of stochastic perturbations in the forecast system is threefold:

1. to generate equi-probable ensemble members
2. to attempt to partly correct the model imperfections (in particular the systematic error)
3. to address spurious model simulation failures/crashes.

In earlier forecast systems item 1 was managed by introducing a perturbation in an initial atmospheric situation, or by lagging the initial situation by a few days. Item 2 was taken into account by considering forecast anomalies with respect to a model climatology (this is still the case in the new system). Item 3 was managed by slightly modifying the horizontal diffusion coefficient temporarily.

Three ingredients are necessary to activate the stochastic perturbation system in the hindcast as well as in the forecast production:

1. the perturbation fields
2. the random calendars
3. the definition of the ensembles.

These 3 ingredients are then used in the model script.

3.2 The nudged pre-hindcasts: generating the perturbation fields

This is a hard and complex task. Each time the model or the ingredients (e.g. the initial situations) are changed, the perturbation fields need to be re-created. This model dependency is similar to the dependency between forecasts and hindcasts. It is, however, not so crucial: a slight change like a change in computer should not make regeneration necessary.

The fundamental idea is that a model is not perfect, and its equations contain errors:

$$\begin{aligned} dX/dt &= \text{model_equations}(X, t) \\ &= \text{true_equations}(X, t) + \text{model_errors}(X, t) \end{aligned}$$



where X is the vector describing the prognostic variables (temperature, moisture, wind and surface pressure) and t is time. We have coded $\text{model_equations}(X,t)$ as we ignore $\text{true_equations}(X,t)$ or, equivalently, $\text{model_errors}(X,t)$.

If we can estimate $\text{model_errors}(X,t)$ for a set of hindcasts, we cannot fully correct a forecast, but we can re-inject the past errors at random to take into account the uncertainty about our $\text{model_equations}(X,t)$.

A way to estimate this error is to carry out a set of preliminary hindcasts during which the model is linearly relaxed (or nudged) toward the true trajectory, here the ERA5 reanalysis. There is a big dilemma:

- if the relaxation is strong, the model will follow ERA5 closely, but the errors will be errors of a short-range forecast (model starting from an analysis)
- if the relaxation is weak, the model will be in a similar state as in a standard seasonal forecast, but the relaxation terms will correspond to a small fraction of the error.

The method depends on the frequency of the random drawings and the intensity of the perturbation, which is linked to the relaxation time. So as to generate ensemble spread, random drawings are done every day. The choice of the relaxation time is less obvious. In previous studies, we have tried values between 1 and 100 days. Using a 6 h relaxation time for both divergence and vorticity produces dynamical imbalances in the model, because the normal modes of the ARPEGE and ECMWF models are different. Values above 100 days are useless, as we want to account for model error during the first months of the forecasts. While System 6 used sequences of perturbations and random drawings every 5 days, and therefore a 10 day relaxation time, for System 7 a 30 day relaxation time was chosen since random perturbations are changed every day.

The nudged experiment is based on 12*24 7-month simulations. There are indeed 12 calendar months and 24 available hindcast years (1993-2016). Note that a start date on 01 December 2016 implies that the model is nudged toward ERA5 till June 2017. The nudged simulations are as close as possible to members 001 of the hindcasts, with four differences

1. the model does not include stochastic perturbations
2. the model is nudged toward 6-hourly ERA5 prognostic variables
3. the postprocessing is highly simplified
4. each day, the differences ERA5 minus model variables, times the relaxation factor, are saved in a file.



A script (`mitraillette.sh`) launches 24 simulations for a given starting month. In fact this script simply updates another script (`traite_nudging_sys7`) which prepares the environment for the model runs. The model script (`mccV6.3.2.tl359l91r_nud`) produces a one-month run.

Ahead of computing, the daily correction files (special FA ARPEGE format files) are named `CR$YYYY$MM$DD` and placed on

```
beaufix: and prolix:/scratch/work/dorell/corrections/corr_N$S
```

where `$S` is a letter indicating the starting month (A=January, B=February, ... L=December).

Copies of these files are also stored on the hendrix storage system.

There are $4 \times 91 + 1$ records in each file (4 variables in 91 vertical levels + mean sea level pressure). The fields are in spectral coefficients (T359 truncation).

Note that because the surface schemes are based on different empirical approaches in ARPEGE and in ERA5, the relaxation is progressively dialled down to zero in the lowest 19 vertical levels (the lowest 13 levels, from about 1000m, being not relaxed at all). A symmetrical treatment is applied the highest 5 levels (see namelist `NAMNUD` which controls nudging in an ARPEGE run). This reduces the inconsistency at the ocean surface, between a free ocean model and a driven atmosphere.

3.3 Random calendar and definition of ensembles

The random choice of a perturbation is done by selecting an arbitrary day and year (not calendar month). This means that the random choice of the perturbations for a given month during a forecast can be described by a random calendar, because each day we need a new perturbation. Random calendars are created once and for all and are valid whatever a hindcast or a forecast, whatever the initial month or the year.

A random calendar is a monthly calendar: it allows to associate to each day of the month a random day with the constraint that:

- the same calendar month is used
- the year is taken at random amongst the years of the nudging phase (1993-2016); in the hindcast phase this introduces a small (not significant) positive bias in the scores by the fact that 4% of the perturbations use observed data from the month to be predicted.
- each day a new random perturbation is drawn.

A random calendar is a text file of 32 lines in the form:



```
ln -sf $CORR/CR1994${MT}08 RX$YYYY${MM}0100
...
ln -sf $CORR/CR1999${MT}22 RX$YYYY${MM}3100
ln -sf $CORR/CR2003${MT}05 RX$YYYN${MN}0100
```

The year and the first day in each line come from a random generation process. The second day in each line corresponds to a sequence from 01 to 31. The last line corresponds to the first day of the next month. It is necessary for the linear time interpolation at each time step in the model.

The reason for having 31 days is to cover the entire month whatever the month. Of course, only the actual days are used by the model (e.g. 31 in January, 28 or 29 in February ...).

The 51 calendars (for 51 ensemble members) for each forecast month \$MN are generated by the script `fic_cal_rand_c3s_5.sh` and gathered in three files:

```
beaufix: and prolix:~gueremy/cal/rand_calendm$MN_n51_$NUMR.tar
```

where NUMR is “a”, “b” or “c”. Note that much more than 51 independent calendars can be generated without associating the same random date to a given date. The model uses series “a”. When it crashes because of a numerical (reproducible) instability, it uses series “b” during that month, then goes back to series “a”. Just in case, a third series of calendars is available for a third attempt. This method of “butterfly effect” can fail if the numerical explosion occurs during the first day of the month. In this case, one has to manually go one month backwards and to replace series “a” by series “b”. The series number is automatically reset to “a” at the end of a successful month.

The case of the first partial month (members others than 001) needs a specific care. As there is a single member for the nudged simulations (see section 2 above), one has to use for the runs starting at the end of a given month, the perturbation of the month after.

Once the perturbations and the calendar exist, one has to decide which calendar is used for which ensemble member. To simplify the process the choice will be the same whatever the starting date.

3.4 Model implementation

With the ingredients described above, the implementation in the model script (embedded in ECLIS) used for the hindcasts as well as for the forecasts is straightforward:

At the beginning NUMR=a and NUMR is reset to “a” at the end of a successful run. If the run fails, NUMR is set to “b”, then to “c” using the ECLIS running environment (see section 5) when the model



crashes. Then, the model is automatically re-launched by the script from the beginning of the current month.

In case of a computer system failure (MPI trouble, node failure, preemption ,...), this procedure (NUMR change) is useless, the model being automatically re-launched from the beginning of the current month without any change in the script. This type of failure occurs much more often than the numerical explosions.

The variable \$CORR which appeared in section 3 above indicates the path to the perturbation files. This is a huge directory (about 1 TB for one calendar month). It is saved on a long-term storage machine (hendrix machine at Météo-France, ecfs at ECMWF), but is copied on the high performance computer disks before a forecast or a hindcast starts.

The model script needs to know the current month MM and the true month MT to properly manage the random calendars (from the file rand_calendm\$MM_n51_\$NUMR.tar, see section 3.3). In case of a member different from the first one, there is a lagged start date; as explained in section 3.3, MT is the month after only for the lagging month, and equal to the current month MM thereafter.

Each member \$MEMBER is assigned a calendar number \$NN. This calendar is extracted from the right tar archive and the file lnRANDOM\$NN is executed. This file creates symbolic links between existing perturbations (with random dates) and daily files which will be read each day by the atmospheric model during the run.



4. Generation of initial conditions

A forecast ensemble is based on 51 individual model integrations. These 51 members of the forecast come from a combination of 3 batches of model integrations (aka runs). The batches differ by the initial dates, which are on two Thursdays of the previous month and the 1st of the current month. Within a given batch, the forecast members differ by the application of random perturbations during the integration following the stochastic dynamics technique (see section 3), thus generating the ensemble spread. This is illustrated in Figure 2.

For each calendar month, a series of hindcasts is produced at least two months before the forecast. It accounts for 25 members and spans 24 years from 1993 until 2016. The only differences with the forecast are the size of the ensembles and the preparation of the initial states, which is managed in advance because the reanalyses are already available. In order to maintain the best homogeneity between the forecasts and the hindcasts, 12 members start from the Thursday approximately two weeks before the end of the previous month, 12 more members start from the following Thursday (still in the previous month), and the last member starts from the 1st of the current month. Thus the diagram is very similar to Figure 2 (replacing 25 members with 12 members). As opposed to the 51 forecasts which are launched in three batches (according to the earliest availability of the initial states) the 25 hindcasts are launched in a single batch with 3 initial dates for a given initial month and year.

As an example, the members of the hindcast for the nominal start date of the 1st of January 1993 are launched as follows: 12 members on the 17th of December 1992, 12 on the 24th of December 1992, and one on the 1st of January 1993.

All hindcast and forecast start dates have been predetermined and are stored respectively in:

```
~copernicusps/SAVE/opers7/startdate_199301_201712.txt
```

```
~copernicusps/SAVE/opers7/startdate_201905_202212.txt
```

Note that each batch of runs is shared between the prolix and beaufix machines at Météo-France, with a flexibility to switch more runs onto a given machine depending on the availability of resources.

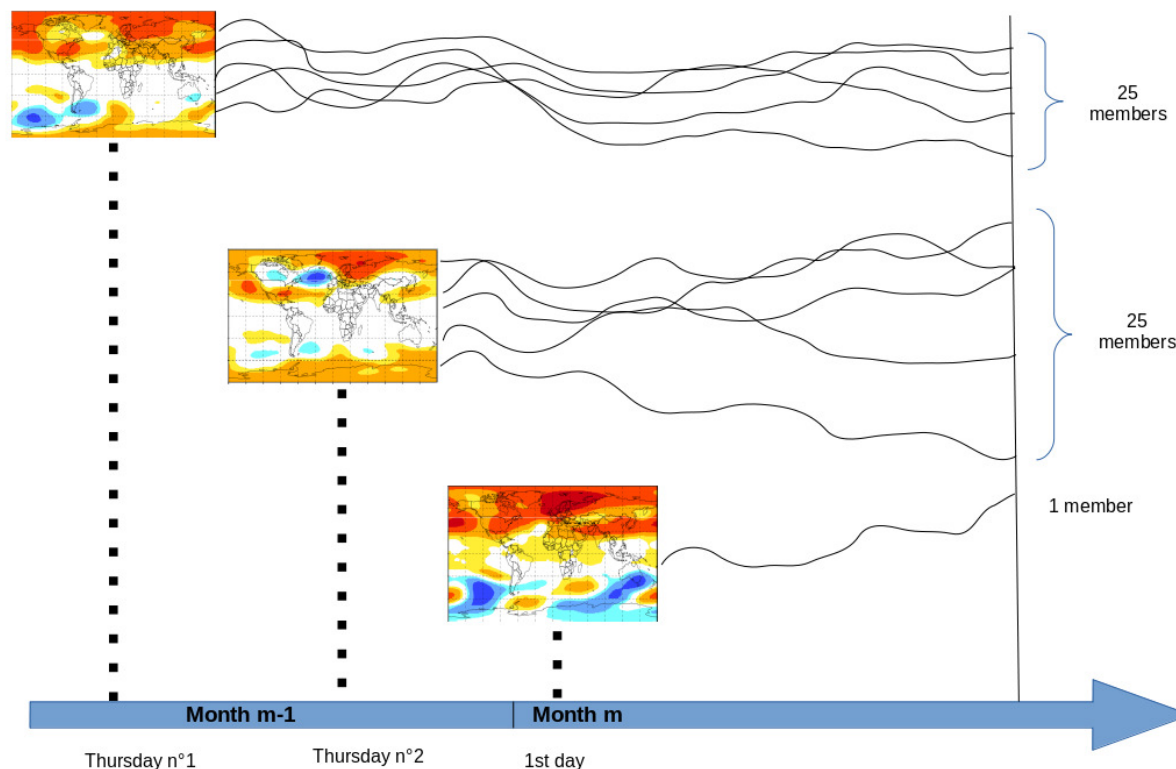


Figure 2 : Generation of the 51-members-ensemble of the seasonal forecast

4.1 Fetching the initial conditions (hindcast)

The initial conditions for ARPEGE come from ERA5 reanalyses, extracted at 6-hourly frequency from 1 January 1979 to the latest available date. The files are extracted from the Climate Data Store and converted into ARPEGE format by

```
cca:/home/ms/fr/top/arpege/grib_to_fa_auto
```

The files, on the grid tl399l137R are tarred by 10-day periods. They are converted to the model geometry tl359l91r and tarred by months by

```
cca:/home/ms/fr/top/arpege/arp2arp_ERA5_step1.cca
```

```
cca:/home/ms/fr/top/arpege/arp2arp_ERA5_step2.cca
```

Then the files are sent to Météo-France by the ectrans utility

```
ecgate:/home/ms/fr/top/SCRIPTS/transfer_restart_GRE_era5
```



```
to beaufix:/scratch/work/dorell/era5/restart/
```

Before launching a hindcast experiment for a given month, it is necessary to extract the three starting situations, and also to prepare the OASIS initial states. This is done by the scripts `prep_rest_sys7` and `prep_rest_sys7_oc`. Note that the NEMO restarts are necessary to create the OASIS restarts.

The NEMO and GELATO reanalysis has been produced by MERCATOR-OCEAN. They are saved on hendrix on a daily basis. ECLIS retrieves these restarts directly on the hendrix storage.

The Surfex initial conditions are derived from ERA5 surface fields. The required fields are the land-sea mask, the orography, the soil type, the temperature and soil water content of the 4 soil layers available in ERA5, the albedo, density and temperature of the snow cover, the low and high vegetation cover as well as the types of high and low vegetation. These fields are retrieved and gathered into grib2 data files on the ecgate computer. These grib files are then uploaded on the Météo-France prolix supercomputer. Finally, these fields are interpolated horizontally and vertically onto the Surfex grid with a dedicated job, requiring a Surfex physiographic data file, an adequate namelist an executable binary file `PREP`, and the Ecoclimap database that includes an ecosystem classification and a coherent set of land surface parameters. More details on soil moisture interpolation can be found in Boisserie et al. (2016).

4.2 Fetching the initial conditions (forecast)

The Surfex initial conditions for the forecast are derived from the 00z IFS operational analysis of the days that match the 3 start dates for a given month. They are processed as soon as the IFS analyses are made available by ECMWF. The processing stage is similar to the one described in the previous section. It should be noted though that due to their very high horizontal resolution (O1280 in Nov. 2016), the IFS operational analyses are retrieved through MARS on the coarser ERA5 resolution.

Every Wednesday our partner MERCATOR-OCEAN uploads a set of 14 files on the Météo-France storage server hendrix in the following directory:

```
hendrix:/home/duboisc/PSY2_G2V3/RUN025/
```

These files describe 14 successive daily ocean states (12h lag). Two of them are actual ocean analyses resulting from the most recent data assimilation cycle while the other 12 are ocean forecasts driven by ECMWF operational forecasts.

Figure 2 shows an example of how the 51 members of the forecast are generated.

The start dates of month M forecasts, as stated above, correspond to:

- Two Thursdays from the second half of month M-1



- The 1st of month M

At these start dates, the ECMWF atmosphere analysis, and the corresponding ocean analysis are fetched and processed. On the first start date, 25 integrations of the model begin, each one with a distinct set of stochastic dynamics perturbations. On the second one, 25 more integrations are initiated similarly. The last integration begins on the 1st day of the month.



5. Environment for CLimate Simulation (ECLIS)

ECLIS (Environment for CLimate Simulation) is the set of scripts and tools necessary to run CNRM-CM. It does not include tools for preparing initial nor boundary conditions. Basically, the script `install` allows to prepare an experiment, and defines another script which will actually run the experiment.

ECLIS manages an environment to run CNRM-CM simulations on supercomputers like `beaufix` and `prolix`. It also manages the storage of the results.

5.1 ECLIS organization

An ECLIS version directory includes a version of tools `reln` and `mtool` which are consistent with the need of the main scripts. The `reln` utility allows to iterate several times a script, whereas the `mtool` utility allows to split a script at time of submission. This directory also contains the following directories:

- **param** which provides useful examples of experiment parameter files; for details on parameters see section 5.4 (Design a 'param_' file)
- **testing** in which sub-directory `prolix` includes parameter files for a very wide variety of experiment configurations; they are provided as a reference of the possible technical settings, but not as scientifically meaningful settings
- **scripts** where there is usually one single, non-numbered, version of the experiment installation script and of each script needed for running a coupled experiment; when a bug is fixed for a script, this appears in the git history of the release directory
- **plugins** which includes the code of the available plugins.

The file `cm_setup` sets up the environment used by ECLIS. It should be sourced from the user's configuration file `.profile`. It is automatically launched by the scripts when needed.

5.2 Use of ECLIS

ECLIS consists in designing an experiment parameter file that ends with `source $INSTALLER $*` and to execute this file with arguments :

- **noask** for avoiding being asked if a previous install should be erased



- **go** for launching the experiment at the end of the (successful) install phase. This is a way to create and launch an experiment at the end of a first one, when there is a logical dependency (e.g. the first experiment creates the restarts that are necessary for installing the second one).

5.3 Design a `param_file`

Defining a CNRM-CM experiment is as simple as designing a so-called `param_file`, which filename must be of the form `param_EXPID`, where `EXPID` is the name given to the experiment. This file, in bash syntax, sets values for ECLIS parameters and then sources the ECLIS install script.

5.4 Standard plugins

Plugins are sets of shell-script commands that ECLIS can take into account during CNRM-CM experiment runs, in order to perform actions atop of the strict running of the coupled or atmosphere-only or ocean-only scripts. Special plugins have been developed for the specificities of seasonal forecast.



6. The basic set of scripts

As stated in section 3, the seasonal ensemble forecast is launched 3 times per month. At each start date (two Thursdays of the previous month and the 1st of the month), a cron job-scheduler executes the scripts `lance_param_F_bfx.sh` and `lance_param_F_plx.sh` on `beaufix` and `prolix` machines.

This first script generates one `param_F????????` file for each member of the ensemble forecast. It contains the initial date and the end date of the forecast. It also sources the generic `param_F` file which is the main file to launch a forecast with ECLIS (see section 5). At the end of `lance_param_F_???.sh`, the current script `param_F????????` is executed. So, it sets values for ECLIS parameters and then sources the ECLIS `install` script.

Another cron scheduled script is executed every 3 hours to monitor the running jobs.

It is installed on the `sxpastel` computing workstation and launches the scripts

```
beaufix: and prolix:/home/gmgec/mrga/copernicusps/opers7/monitrel.sh
```

This monitoring consists in browsing the files

```
beaufix: and prolix:/scratch/work/copernicusps/output/en_cours.txt
```

which contains the IDs of the ongoing jobs. A job ID is automatically written into this file when the job is launched, and deleted from this file as soon as they successfully end. The script `monitrel.sh` checks that every job ID in the file is also present in the supercomputer Slurm scheduling queue. If not, it automatically relaunches the job. It gives up after two trials per job.



7. Post processing

The XIOS output software now manages most of the post-processing of the model files to the requested format for C3S. The inputs to XIOS are sets of xml files for each model component type (Arpege/Surfex, Nemo/Gelato, Trip).

The different fields available, their long names, reference grid and diagnostic operations are listed in separate xml files for each model component (e.g. `surfex_fields.xml` `nemo_fields.xml` and `atmo_fields.xml`). XIOS can manage operations on model outputs such as time averaging, time slicing, or horizontal and vertical interpolation and statistics.

For each requested variable, all global attributes as well as output frequency, vertical levels and horizontal resolution, are defined in a set of files such as `arpsfx_files.xml` in the case of Arpege/Surfex outputs and `nemo_files.xml` for Nemo/Gelato.

All xml files used as input to XIOS for System 7 are stored on the `beaufix` and `prolix` computers in the following path:

```
/home/gmgec/mrga/copernicusps/SAVE/opers7/XIOS
```

XIOS thus generates directly the fields for the atmosphere, surface and ocean on the required $1^\circ \times 1^\circ$ grid with 180 latitudes and 360 longitudes starting at 89.5°S , 0.5°W and extending northwards and eastwards. The data (one file per field and per month) are stored both locally on the HPC machine:

```
beaufix: and
```

```
prolix: /scratch/work/copernicusps/outputs/HINDCAST/$EXPID/
```

and uploaded on the Météo-France storage capacity:

```
hendrix: /home/copernicusps/SYS7/HINDCAST
```

The data format is netcdf (with compression). The files are transferred by the `ectrans` utility to ECMWF on:

```
cca: /sc1/tcwork/zac/lb/copernicus/s7
```

For each starting month and member, a directory is created with a name containing the date and the member number. For example directory `199303_001` contains the files created for member 1 of the March 1993 hindcasts. Note that the start dates are in this case two starts end of February as well as 01 Mar. There is one tar file in each directory containing 51 netcdf files.



Tables 2.1 and 2.2 list the variables. In addition, land-sea mask and surface orography are available.

2. Table 2: list of atmosphere / land variables archived

Every 6 hours	Every 24 hours	Every 24 hours, accumulated
2 metre temperature	total soil moisture	total precipitation
2 metre dewpoint temperature	snow depth (water equivalent)	convective precipitation
10 metre u wind	snow density	snow fall
10 metre v wind	Tmax at 2 m (over previous 24 h)	surface sensible heat flux
mean sea level pressure	Tmin at 2 m (over previous 24 h)	surface latent heat flux
total cloud cover	wind speed of gust (max over previous 24h)	surface solar radiation downwards
skin temperature		surface thermal radiation downwards
Every 12 hours		surface solar radiation
geopotential		surface thermal radiation downwards
temperature		surface runoff
specific humidity		sub-surface runoff
u/v wind components		top incoming solar radiation
At the following 11 pressure levels:		top net solar radiation
925, 850, 700, 500, 400, 300, 200, 100, 50, 30, 10 hPa		top thermal radiation
		east-west surface stress
		north-south surface stress
		evaporation

3. Table 2.2: list of ocean / sea ice variables archived

Every 6 hours
sea ice temperature
sea surface temperature
Every 24 hours
sea surface height above geoid
sea ice concentration
sea ice thickness
depth of 20 deg isotherm
mixed layer depth
sea surface salinity
heat content over the first 300 m
salinity in top 300m
zonal surface current
meridional surface current



8. References

Batté, L., and Déqué, M. (2016). Randomly correcting model errors in the ARPEGE-Climate v6.1 component of CNRM-CM : applications for seasonal forecasts. *Geosci. Model Dev.* 9 : 2055-2016, doi : 10.5194/gmd-9-2055-2016.

Beljaars, A., Brown, A. and Wood, N, 2004 : A new parametrization of turbulent orographic form drag. *Quart. J. R. Meteorol. Soc.*, 130, 1327-1347.

Boisserie, M., Decharme, B., Descamps, L., & Arbogast, P. (2016). Land surface initialization strategy for a global reforecast dataset. *Quart. J. R. Meteorol. Soc.* 142, 880-888.

Courtier P., J.F. Geleyn (1988). A global numerical weather prediction model with variable resolution: Application to the shallow water equations. *Quart. J. R. Meteorol. Soc.* 114, 1321–1346.

Cuxart, J., Bougeault, Ph. and Redelsperger, J-L., 2000 : A turbulence scheme allowing for mesoscale and large-eddy simulations. *Quart. J. R. Meteorol. Soc.*, 126, 1-30.

Decharme B, Delire C, Minvielle M, et al (2019) Recent Changes in the ISBA- CTRIP Land Surface System for Use in the CNRM- CM6 Climate Model and in Global Off- Line Hydrological Applications. *J Adv Model Earth Syst* 2018MS001545. doi: 10.1029/2018MS001545

Déqué, M., C. Drevet, A. Braun, D. Cariolle (1994). The ARPEGE/IFS atmosphere model: a contribution to the French community climate modelling. *Climate Dynamics* 10, 249-266.

Guérémy, J. F., 2011 : A continuous buoyancy based convection scheme: one and three dimensional validation. *Tellus A*, 63, 687-706.

Jacob, R., J. Larson, and E. Ong, 2005: MxN Communication and Parallel Interpolation in CCSM3 Using the Model Coupling Toolkit. *Int. J. High Perf. Comp. App.*, 19(3), 293-307

Larson, J., R. Jacob, and E. Ong, 2005: The Model Coupling Toolkit: A New Fortran90 Toolkit for Building Multiphysics Parallel Coupled Models. *Int. J. High Perf. Comp. App.*, 19(3), 277-292

Lopez, Ph., 2002 : Implementation and validation of a new prognostic large-scale cloud and precipitation scheme for climate and data-assimilation purposes. *Quart. J. R. Meteorol. Soc.*, 128, 229-257.

Masson, V., Le Moigne, P., Martin, E., Faroux, S., Alias, A., Alkama, R., Belamari, S., Barbu, A., Boone, A., Bouyssel, F., Brousseau, P., Brun, E., Calvet, J.-C., Carrer, D., Decharme, B., Delire, C., Donier, S., Essaouini, K., Gibelin, A.-L., Giordani, H., Habets, F., Jidane, M., Kerdraon, G., Kourzeneva, E., Lafaysse, M., Lafont, S., Lebeaupin Brossier, C., Lemonsu, A., Mahfouf, J.-F., Marguinaud, P., Mokhtari, M., Morin, S., Pigeon, G., Salgado, R., Seity, Y., Taillefer, F., Tanguy, G., Tulet, P., Vincendon, B., Vionnet, V., and Voldoire, A. (2013). The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes, *Geosci. Model Dev.*, 6, 929-960, doi:10.5194/gmd-6-929-2013.

Meurdesoif Y (2018) Xios fortran reference guide. Institut Pierre Simon Laplace (IPSL), http://forge.ipsl.jussieu.fr/ioserver/svn/XIOS/trunk/doc/XIOS_reference_guide.pdf

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research: Atmospheres* (1984–2012), 102(D14), 16663-16682.



Morcrette, J. J. (1990). *Impact of changes to the radiation transfer parameterizations plus cloud optical Properties in the ECMWF model. Monthly Weather Review*, 118(4), 847-873.

Piriou, J. M., Redelsperger, J. L., Geleyn, J. F., Lafore, J. P., and Guichard, F., 2007 : *An approach for convective parameterization with memory : Separating microphysics and transport in grid-scale equations. Journal of the Atmospheric Sciences*, 64, 4127-4139

Salas y Méliá D (2002) *A global coupled sea ice-ocean model. Ocean Model* 4:137–172

Sommeria, G. and Deardorff, J.W., 1977 : *Subgrid-scale condensation in models of nonprecipitating clouds. J. Atmos. Sci.*, 34, 344-355.

Voldoire, A., Decharme, B., Pianezze, J., Lebeaupin Brossier, C., Sevault, F., Seyfried, L., et al. (2017). *SURFEX v8.0 interface with OASIS3-MCT to couple atmosphere with hydrology, ocean, waves and sea-ice models, from coastal to global scales. Geoscientific Model Development*,10(11), 4207–4227.
<https://doi.org/10.5194/gmd-10-4207-2017>

Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., et al (2019). *Evaluation of CMIP6 DECK experiments with CNRM-CM6-1. Journal of Advances in Modeling Earth Systems*, 11.
<https://doi.org/10.1029/2019MS001683>



ECMWF, Shinfield Park, Reading RG2 9AX, UK

Contact: info@copernicus-climate.eu