



Documentation of the METEO-FRANCE Pre-Operational seasonal forecasting system

Issued by:

Date: 12/01/2017

Ref: SERVICE CONTRACT N° 2015/C3S_433_Lot1-

METEO-FRANCE (Deliverable D3.1)



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.



Documentation of the METEO-FRANCE Pre-Operational seasonal forecasting system

METEO-FRANCE

Laurent Dorel
Constantin Ardilouze
Michel Déqué
Lauriane Batté
Jean-François Guérémy

Date: 12/01/2017

REF.: SERVICE CONTRACT N° 2015/C3S_433_Lot1-
METEO-FRANCE (Deliverable D3.1)

Contents

1 Basic facts	3
2. Modeling components of the forecast system.....	7
2.1 Introduction	7
2.2 The atmospheric model, Arpege V6.2.5.....	8
2.3 The ocean model Nemo V3.6 and sea-ice model Gelato V6	9
2.4 The Land Surface model Surfex V8.1 and river routing model TRIP	10
2.5 The coupler Oasis MCT v3.0	11
2.6 Eclis.....	12
3. Generation of stochastic perturbations.....	14
3.1 Introduction	14
3.2 The nudged pre-hindcasts: generating the perturbation fields.....	14
3.3 Random calendar	17
3.4 Definition of the ensembles.....	19
3.5 Model implementation.....	19
4. Generation of initial conditions.....	22
4.1 Fetching the initial conditions (hindcast).....	23
4.2 Fetching the initial conditions (forecast)	25
5. Environment for CLImate Simulation (Eclis).....	26
5.1 Eclis organization.....	26
5.2 Use of Eclis	26
5.3 Design a 'param_' file	27
5.4 Standard plugins.....	27
6 The basic set of scripts	28
7. Post processing	29



Introduction

Météo-France built its own model for numerical weather prediction as soon as the early 1970's, and for climate simulations in the early 1980's. However, seasonal forecasting became a fully operational activity in the late 1990's. The seasonal forecast system at Météo-France, imbedded in the Eurosis consortium, has evolved by successive versions, named “systems” as they include modifications of several models as well as initial conditions. A new system is issued every 3-4 years. The presently operational system is Eurosis system 5. This 2-year old forecast system is used in the proof of concept phase of the C3S-433 contract. During the first year of the contract, we have assembled different components to build up a new forecast system which will be used in the pre-operational phase.

After a brief overview of the new system, the present document describes the features of its the individual components. In a second part, more technical, the document details the first stage of a forecast, consisting in building initial states for the atmosphere, ocean, sea-ice and land surface. It also features the preliminary operations which generate the stochastic perturbations and the hindcast ensembles. The procedure managing the forecasts (and the hindcasts) is then detailed. Finally, the post-processing of raw output files from the model integration into standardized time series and monthly-mean files is shortly described.

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 the pre operational forecast system.

Table 1 : basic facts

	PO forecast system
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	Pre-Operational
Data time of the first forecast run	01/03/2017
Configuration of the EPS	
Is the model is coupled to an ocean model ?	Yes from day 0
Short description of the ocean model	Nemo v3.6 , ORCA 1° grid, 75 model levels
Is the model is coupled to a sea ice model ?	Yes
Short description of the sea ice model	Sea-ice model is Gelato v6 (Salas y Melia, 2002), embedded in the ocean model. It is initialized from unperturbed 1 degree resolution MERCATOR-OCEAN Ocean and Sea-Ice Analysis
Is the model is coupled to a wave model ?	No
Horizontal resolution of the atmospheric model	TL359
Number of 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: 7'30" Ocean/Sea-Ice: 30' Coupling frequency: 3 hours
Initial conditions and perturbations	
Data assimilation method for control 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 the control forecast ?	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	23 years (1993-2015)
Produced on the fly or fix re-forecasts ?	fix re-forecasts



Frequency	monthly
Ensemble size	25 members
Initial conditions	ERA Interim (T255L60) for atmospheric and land surface MERCATOR-OCEAN reanalyses for ocean
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 in 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.

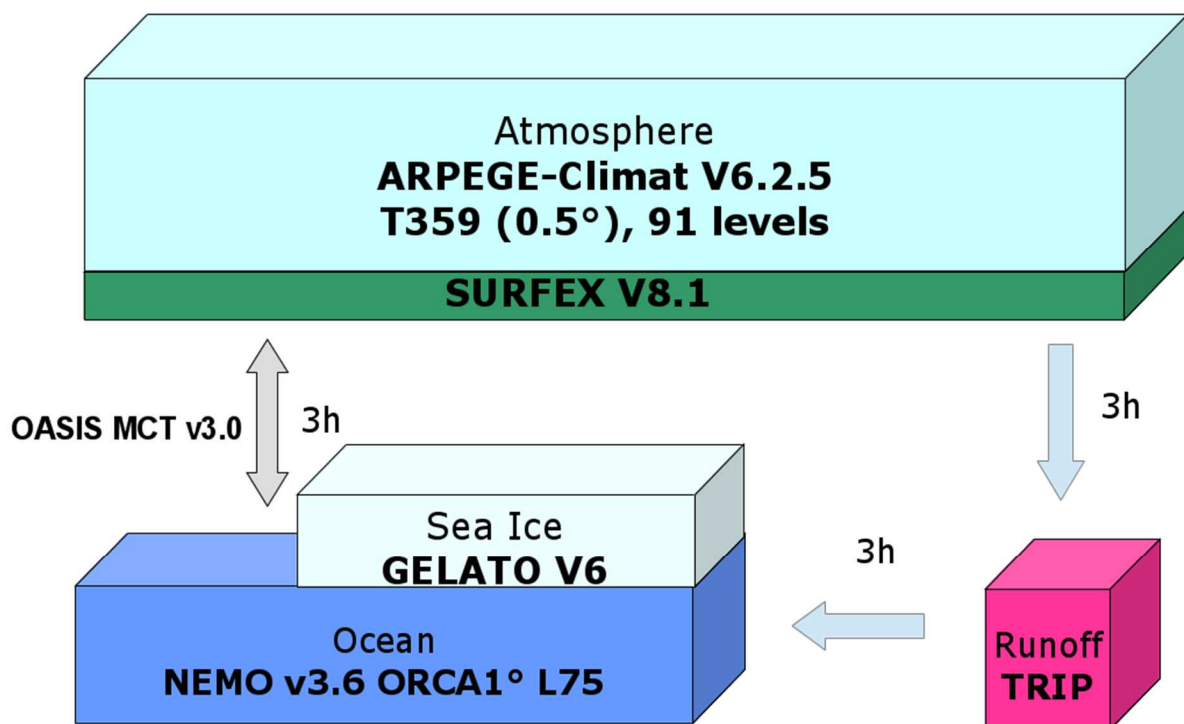


Figure 1 - CNRM-CM model components of the seasonal forecast system



2.2 The atmospheric model, Arpege V6.2.5

Arpege-IFS forecast model has been jointly developed by ECMWF and Météo-France for numerical weather prediction in the late 1980s (Courtier and Geleyn, 1988). A climate version of this model, 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. As far as the parameterization packages are concerned, ECMWF and Météo-France use different schemes, except for radiation. As far as the climate version is concerned, the major change with respect to the version used in system 5 (used in CMIP5 as well) is the adoption of a fully new set of physical parameterizations prepared for the next CMIP6 experiment.

The physics used in system 5 and the previous systems back to the Demeter project is named “diagnostic physics”. This set of physical parametrizations was based on the laws of thermodynamics and of statistical mechanics. The main assumption was that the diabatic processes reached a statistical equilibrium for each grid point at each time step. This assumption is reasonable when the time step is 30 min or longer or when the grid size is 100 km or larger. However, the limitations of the “diagnostic physics” appeared when used for short-range weather prediction or in a 1-dimensional model driven by observations: the time step by time step evolution is chaotic. This is because in this approach each time step ignores the diabatic calculations of the previous one. One of the most noticeable consequence was the fact that the daily cycle of convective precipitation reached its maximum at noon. When considering monthly averages, the model chaotic behaviour is not that detrimental. This explains why Météo-France used the “diagnostic physics” in CMIP5 scenarios as well as is Eurosis system 5, while developing a new physics.

The new “prognostic physics” relies upon the same 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 along 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.

The time step has been reduced to improve the short time evolution of the fluxes: 1h for radiation and 7.5 min for the rest of the physics, instead of 3h and 15 min in the former version.

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 V6.2.5.



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).

2.4 The Land Surface model Surfex V8.1 and river routing model TRIP

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/>

The reference paper for the last version of Surfex is in preparation while writing this document. However, most novel features have been published in Le Moigne et al. (2016), Decharme et al. (2015), Vergne et al. (2014), Decharme et al. (2013) and Decharme et al. (2012).



2.5 The coupler Oasis MCT v3.0

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 platforms¹. Oasis sustained development is ensured by a collaboration between CERFACS and the Centre National de la Recherche Scientifique (CNRS).

The current OASIS3-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).

OASIS3-MCT is a portable set of Fortran 77, Fortran 90 and C routines. Low-intrusiveness, portability and flexibility are OASIS3-MCT key design concepts. After compilation OASIS3-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. OASIS3-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). OASIS3-MCT also supports file I/O using netcdf.

The new version, OASIS3-MCT_3.0 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 OASIS3 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) OASIS3-MCT coupling library. The namcouple configuration file is also largely unchanged relative to OASIS3, 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.

In this system, the ensemble members are generated by small differences in their initial conditions but also during the integration, in order to take into account uncertainties in the initial state as well as model errors.



3. Generation of stochastic perturbations

3.1 Introduction

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 previous 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:

$$dX/dt = \text{true_equations}(X, t) + \text{model_errors}(X, t)$$



where X is the vector describing the prognostic variables (temperature, moisture, wind and surface pressure) and t is time. We have coded $\text{model_equations}(X,t)$ and 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 ERA-interim reanalysis (ERA-Interim) (ERA-Interim in the following). There is a big dilemma:

if the relaxation is strong, the model will follow ERA-Interim 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. Because weather regimes have a characteristic time of circa 5 days, a natural value for the length between two drawings is 5 days. We tested 6 hours, 5 days and 1 month with a previous low resolution version of the model, and did not observe a systematic impact on scores. The choice of the relaxation time is less obvious. In previous studies, we have tried values between 1 and 100 days. Values below 6 h are not realistic because ERA-Interim is not available at a higher frequency. 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. The results show that a value of 10 days is a reasonable compromise for the scores of the different variables in the different regions. See Batté and Déqué (2016) for more details on this technique.

The nudged experiment is based on 12*23 7-month simulations. There are indeed 12 calendar months and 23 available hindcast years (1993-2015). Note that a start date on 01 December 2015 implies that the model is nudged toward ERA-Interim till June 2016.



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 ERAI interim prognostic variables
3. the postprocessing is highly simplified
4. each day, the differences ERAI minus model variables, times the relaxation factor, are saved in a file.

The script:

```
~deque/mitraillette.sh
```

launches 23 simulations for a given starting month. In fact this script simply updates another script

```
~deque/traite_nudging_359191.c3s
```

which prepares the environment for the model runs. The model script is

```
~deque/arpege/mccV6.2.5.t1359191r_nud_c3s
```

The daily correction files are FA files (special Arpege format based on grib) named CR\$YYYY\$MM\$DD and saved on

```
/scratch/work/deque/corrections/corr_N$S
```

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

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 ERAI, the relaxation is progressively dialled down to zero in the lowest 5 vertical levels (the lowest level 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

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 for all and are valid whatever a hindcast or a forecast, whatever the month or the year. They will be changed only when the hindcast period is modified (not just extended by a couple of years), or when the method of taking a new random sequence every 5th day is modified.

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 (1993-2015); 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.

the days are consecutive during a 5-day sequence

A random calendar is a text file of 36 lines in the form:

```
ln -sf $CORR/CR1994${MT}08 RX$YYYY${MM}0100
```

...

```
ln -sf $CORR/CR1999${MT}08 RX$YYYY${MM}3500
```

```
ln -sf $CORR/CR1999${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 35. 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 35 days whatever the month is to facilitate the generation of pentads. Of course, only the actual days are used by the model (e.g. 31 in January, 28 ou 29 in February ...). The 5-day (pentad) sequences are defined by the first day of



each sequence. This number is selected at random between the 01 and 24 of the month. This allows to have a universal random calendar whatever the month. As a side effect the perturbations of day 29, 30 and 31 are never used by the model. Because 23 years are available, there are $(24*23)/7=78$ possible calendars having no date in common: one has to choose 7 pentad start dates (because $7*5=35$) amongst $24*23$ candidate days. This ensures that two ensemble members can never be identical.

The 78 calendars are generated by the script:

```
~deque/fic_cal_rand_c3s_1.sh
```

and gathered in three files:

```
~deque/rand_calend28_n78_$NUM.tar
```

where NUM is a, b or c. Indeed, one can generate much more than 78 calendars, but they might associate 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 20 or 25 of the month, the perturbation of the month after. However the amplitude of the perturbations (which are model errors) significantly increases with the range at the beginning of an initialized run. So, during the first phase of the forecast, the start of the pentad is limited to day 08 of the month. As a result, there are only $(8*23)/7=26$ possible non-overlapping calendars. This is not a limitation, as there are 12 hindcasts or 25 forecasts starting at the same date. The 26 calendars are in:

```
~deque/rand_calend12_n26_$NUM.tar
```



which has been generated (once for all) by the same script as the standard random calendars.

3.4 Definition of the ensembles

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. One could have made a further simplification by using the same association between a member and its random calendar whatever the month, but there is a drawback. In a random calendar, only 7 (and possibly less) years are involved. If during each month a given member uses these same years, then the perturbations do not fully reflect the statistical distribution of model errors. For this reason, a given member will use each month a different random calendar. There are 8 lists: one for month 0 (the end of the month in lagged members 002, 003 ...), month 1, ... month 7. They are gathered in file

```
~deque/rand_liste.tar
```

which has been created by

```
~deque/fic_cal_rand_c3s_2.sh
```

Each list contains 51 numbers (25 numbers for month 0) which associate an ensemble member to the number of a random calendar (1 to 26 for month 0, 1 to 78 for month 1 to 7). In the case of a hindcast, only the first 25 (or the first 12 at month 0) numbers are used.

3.5 Model implementation

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

The file `$HOME/choix_$EXPID` gives the letter a, b or c of the series of random calendars (NUMR). 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” by the script:

```
~deque/repere4.sh
```



automatically submitted when the model crashes. Note that in case of timeout when the model is frozen by the system, this procedure is inefficient (and useless). The security is ensured by the creation-destruction by the model script of a file:

```
$HOME/encours_$(EXPID)
```

which is checked by a half-hourly cron. If the above file exists and the EXPID job is not in the queue, EXPID is resubmitted. This type of failure occurs much more often than the numerical explosions (at Météo-France as well as at ECMWF). But the security mentioned here has nothing to do with the stochastic perturbations.

The variable CORR which appeared in section 3 above indicates where to find the perturbation files. This is a huge directory (about 1 TB for one calendar month). It may be saved on hendrix, but must be loaded on prolix before a forecast or a hindcast starts.

Then the model script needs to decide which type of random calendar to use. If the start day (variable JJ) is not 01, one must use rand_calend12_n26_\$(NUMR).tar (shorter list), otherwise one uses rand_calend28_n78_\$(NUMR).tar, which is the standard case. One needs to know if the number of the month (STEP=0 for lagging, STEP=1 for first month, ... STEP=7 for last month) is IPASS (for member 001) or IPASS-1 (for other members). IPASS is the counter of the automatic launching procedure. Variables MM, MN and MT are current month, next month and true month respectively. They are used in for UNIX variable substitutions in the random calendar file. In general MT=MM, but for STEP=0, MT is the month after. For instance, if the model starts on 25 November, then MM=11 and MT=12 because December perturbations are used. Month MN is associated to year YYYYN because of the possible change of year.

Thanks to the list of calendars associated with the right step, one assigns to member \$MEMBER 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. Note that links such as RX1993022900 or RX1994013500 will never be used.



It is of highest importance that the namelist switch LDYSTO is set to .TRUE. . In the Arpege namelist. With the default value (.FALSE.) the perturbations have the wrong sign. In the nudging runs this switch must be at its default value. If not, the model crashes rapidly, which is less detrimental than consuming hours of computer for a degraded result.



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) are launched. The batches differ by the initial dates which are 20th and 25th 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 23 years from 1993 until 2015. 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 20th of the previous month, 12 more members start from the 25th of 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.

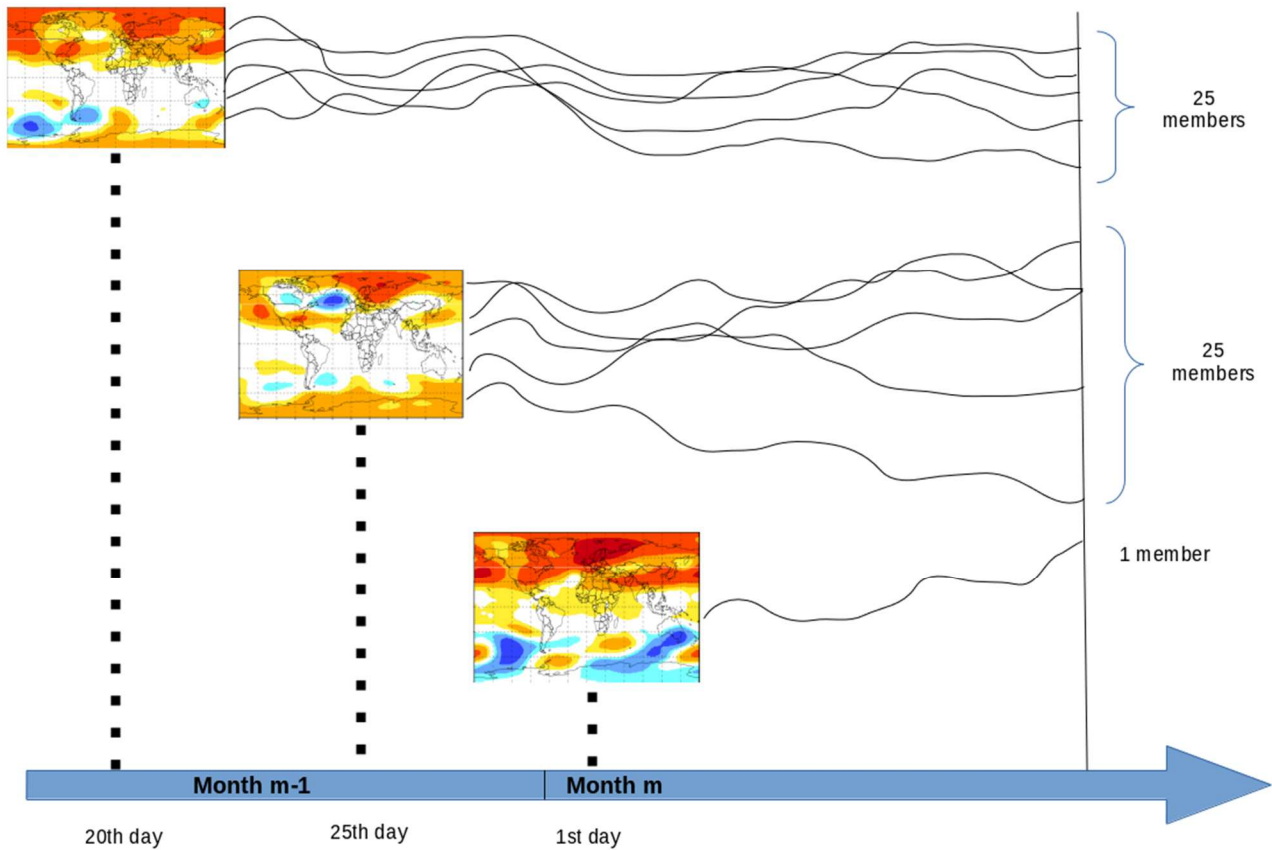


Figure 2 : Generation of the 51-members-ensemble of the seasonal forecast

4.1 Fetching the initial conditions (hindcast)

The initial conditions for Arpege come from ERA-interim reanalyses, extracted at 6-hourly frequency from 1 January 1979 to the latest available date. The files are extracted from MARS archived and converted into Arpege format by

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

The files, on the original grid tl255l60R are tarred by 10-day periods and sent to Météo-France by the ectrans utility

```
ecgate: /home/ms/fr/tou/transfer_ec_beaufix_multi
```

Then the files are converted to the model geometry tl359l91r and tarred by months by



```
beaufix: /home/gmgec/mrga/deque/aladin/e927/arp2arp.V6.0.ER  
AI_auto
```

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:

```
prolix: /home/gmgec/mrga/deque/prep_previ_359191.c3s
```

Most scripts like this are present both on prolix and beaufix.

The Nemo and Gelato reanalysis has been produced by MERCATOR-OCEAN (Clotile Dubois, 2016, personal communication). They are saved on hendrix on a daily basis.

```
prolix: /home/gmgec/mrga/deque/restartmercator
```

Allows to put the needed restarts at the right place on prolix. Note that the Nemo restarts are necessary to create the Oasis restarts.

The Surfex initial conditions are derived from ERA-Interim surface fields.

The required fields are the land-sea mask, the orography, the temperature and soil water content of the 4 soil layers available in Era-Interim, 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 through the MARS system on the Ecgate computing facility. These grib files are then uploaded on 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 called PREP, and the Ecoclimap database that includes an ecosystem classification and a coherent set of land surface parameters.

More details on this method can be found in Boisserie et al. (2015).



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 aforementioned start dates. 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 a T255 coarser grid.

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/s/spsy/spsy002/PSY2G3R4/db
```

These files describe 14 successive daily ocean states. Two of them are actual ocean analyses resulting from the most recent data assimilation cycle while the other 12 are guesses issued from the ocean model integration.

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 :

- The 20th and the 25th of month M-1
- The 1st of month M

At these start dates, the ECMWF atmosphere analysis, and the most recent 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 CLlimate Simulation (Eclis)

Eclis (Environment for CLlimate 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 'relan' and 'mtool' which are consistent with the need of the main scripts; it 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, 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 `.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 2nd 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 your 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 part 1, the seasonal ensemble forecast is launched 3 times per month. At each start date (the 20th, the 25th and the 1st), a cron job-scheduler executes the script `lance_forsys6.sh`.

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 `param_generique` file which is the main file to launch a forecast with Eclis (see section 4). At the end of `lance_forsys6.sh`, the created 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' CNRM computing facility but launches the script `prolix:/home/gmgec/mrga/ardilouzec/monitrel.sh`

via ssh.

This monitoring consists in browsing the file

`prolix:/scratch/work/ardilouzec/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 different components of the model (atmosphere, ocean, surface) generate output files to be post-processed. For the atmosphere and surface, the fields are interpolated on a $1^\circ \times 1^\circ$ grid with 181 latitudes and 360 longitudes starting at $90^\circ\text{S}, 0^\circ\text{W}$ and extending northwards and eastwards. For the ocean, the files are provisionally on the Nemo grid. The data (one file per field and per month) are first stored on

```
prolix:/scratch/work/ardilouzec/output/series_nc
```

The format is netcdf (with packing for the surface and atmosphere and with zip compression for the ocean). The files are transferred by the ectrans utility to ECMWF on:

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

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 20 Feb, 25 Feb and 01 Mar. There are one netcdf file per month and per variable in each directory (i.e. 329 files). The file transfer is managed by a script running at ECMWF on computer cca (or ccb):

```
cca:/sc1/home/zac/transfer_prolix_ec3_auto
```

This script is automatically relaunched in a double loop, the innermost running on the ensemble members (forecasts and hindcasts), the outermost running on the years (hindcasts only). The file monitoring this double loop is:

```
cca:/sc1/home/zac/relances/AA3/AA3_his
```

The automatic launching system is the same as used by Eclis (see section 5).



Table 2 lists the daily variables. In addition, land-sea mask and geopotential of surface orography are available.

Table 2: list of daily variables archived

<p>Every 6 hours 2 metre temperature 2 metre dewpoint temperature 10 metre u wind 10 metre v wind mean sea level pressure total cloud cover skin temperature</p> <p>Every 12 hours geopotential temperature specific humidity u/v wind components</p> <p>At the following 11 pressure levels: 1000, 925, 850, 700, 500, 400, 300, 200 hPa</p>	<p>Every 24 hours sea surface temperature total soil moisture surface temperature snow depth (water equivalent) Tmax at 2 m (over previous 24 h) Tmin at 2 m (over previous 24 h)</p>	<p>Every 24 hours, accumulated total precipitation snow fall surface sensible heat flux surface latent heat flux surface solar radiation downwards surface thermal radiation downwards surface solar radiation surface thermal radiation top solar radiation top thermal radiation east-west surface stress north-south surface stress evaporation</p>
---	--	--

<p>For the ocean: Every 24 hours, mean: sea level sea ice concentration sea surface temperature depth of 28 deg isotherm depth of 26 deg isotherm depth of 20 deg isotherm mixed layer depth surface salinity heat content over the first 300 m zonal surface current meridional surface current</p>
--



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.

Boisserie, M., Decharme, B., Descamps, L., & Arbogast, P. (2016). Land surface initialization strategy for a global reforecast dataset. *Quarterly Journal of the Royal Meteorological Society.*

Courtier P., J.F. Geleyn (1988). A global numerical weather prediction model with variable resolution: Application to the shallow water equations. *Quart. J. Roy. Meteor. 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. Met. Soc.*, 126, 1-30.

Decharme, B., Alkama, R., Papa, F., Faroux, S., Douville, H., & Prigent, C. (2012). Global off-line evaluation of the ISBA-TRIP flood model. *Climate dynamics*, 38(7-8), 1389-1412.

Decharme, B., Brun, E., Boone, A., Delire, C., Le Moigne, P., & Morin, S. (2015). Impacts of snow and organic soils parameterization on North-Eurasian soil temperature profiles simulated by the ISBA land surface model. *The Cryosphere Discussions*, 9, 6733-6790.

Decharme, B., Martin, E., & Faroux, S. (2013). Reconciling soil thermal and hydrological lower boundary conditions in land surface models. *Journal of Geophysical Research: Atmospheres*, 118(14), 7819-7834.

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.

Le Moigne, P., Colin, J., & Decharme, B. (2016). Impact of lake surface temperatures simulated by the FLake scheme in the CNRM-CM5 climate model. *Tellus A*, 68.

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. Met. 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.

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.

Vergnes, J. P., Decharme, B., & Habets, F. (2014). Introduction of groundwater capillary rises using subgrid spatial variability of topography into the ISBA land surface model. Journal of Geophysical Research: Atmospheres, 119(19).

Copernicus Climate Change Service

ECMWF, Shinfield Park, Reading RG2 9AX,
UK

Contact: Samuel.almond@ecmwf.int