
Response to reviewers on manuscript gmd-2016-174 : "Development of a probabilistic ocean modelling system based on NEMO 3.5: application at eddy resolution" submitted to Geoscientific Model Development.

We wish to thank the reviewers for their interest in our paper, for their constructive comments and suggestions that lead to useful improvements in the manuscript.

N.B: In the following, the reviewers' questions and comments are shown in '***bold-italic***' type, our answers appear in 'standard type' and the modified text of the manuscript is given in '*italic type*'.

REVIEWER 1:

1) *No line for "median" in the legend for Fig 6.*

Fig. 6: Thank you for noting this. In fact the ensemble median was plotted as a thin white line, which is why it is not seen in the caption. Most of the time, the median is anyway hidden below the line of the ensemble mean (thick yellow) in the figure, since the ensemble PDF is quasi-gaussian. So to avoid confusion we have now removed 'median' from the figure and caption.

REVIEWER 2:

1) *The online diagnostics is one of the most useful developments proposed in the paper but authors don't provide any information on the computational cost of these online diagnostics. As these diagnostics need several global mpi communications, the cost should be important. Could you provide this cost at least for an example of this statistic?*

We have computed online the ensemble mean and variance of four 2D fields (i.e mixed layer depth, sea-surface temperature, sea-surface zonal and meridional velocities) at both daily and monthly frequencies during the whole OCCIPUT global run; the same diagnostics were also computed at hourly frequency for 6 specific months. Our tests with and without this small set of online ensemble diagnostics showed that they did not induce any noticeable increase of the cost, which remained undistinguishable from the slight, random run-to-run variations of CPU costs.

More generally, the cost of these diagnostics depends at first order on their call frequency and on the dimension of the treated arrays: their cost may indeed become noticeable or even substantial if online ensemble diagnostics were applied on all 3D fields at all time steps. We did not perform such an extreme test and thus we cannot evaluate its computational cost. We have added the following paragraph in section 4.3 to address and summarize these points:

"The cost of online ensemble diagnostics depends on the call frequency, number and size of the concerned fields, on the architecture of the machine and the performance of communications. Our online ensemble diagnostics concerned a few two-dimensional fields at hourly to monthly frequencies, and had a negligible cost. "

2) *Authors suggest that the ensemble online method could be useful for relaxation of the ensemble mean toward a climatology for example. Could you explain more precisely the way this could be done, is there already work and references about such method? It is not obvious that it will work properly. Is there a way to keep a good spread of the ensemble ?*

The comment you refer to in section 3.2 is an example that illustrates how the online ensemble diagnostics may be used. By “relaxation of the ensemble mean toward a climatology” we mean computing a correction term based on the simulated ensemble mean, and then applying this term identically to all members. By construction, each member would be corrected by the same amount: this would not directly affect the ensemble spread but only “translate” the entire ensemble distribution toward the climatological value. We are not aware of any reference about such an approach, but implementing it would be straightforward. Note that we discuss a variant of this method in our response to your question 5) about the surface fluxes. We have tried to clarify this example in section 3.2.

Therefore, the following paragraph:

“This may be useful for certain applications: a simple example would be for example the relaxation (nudging) of the model simulation towards some climatological data. In this case, indeed it could be much better to relax the ensemble mean than the individual ensemble members, to avoid damping the intrinsic variability of the system by the relaxation.”

has been replaced by:

“For instance, it may be interesting to relax the modeled forced variability towards reference (e.g. reanalyzed or climatological) fields, with no explicit damping of the intrinsic variability: the nudging term would involve the current ensemble mean and be applied identically to all members at the next time step, resulting in a simple “translation” of the entire ensemble distribution toward the reference field.”

3) *Could you explain why do you use the NATL experiment for the gulf stream study and the ORCA one for the MOC?*

The objective of the paper is mainly to present new, generic model developments implemented in NEMO. The NATL and ORCA experiments are only given here as examples, to validate the ensemble modeling system in both its regional and global configurations, and to illustrate different types of results it can provide on various ocean quantities and scales (monthly temperatures, annual MOC, etc). Given the high computational cost of such ensemble experiments, the choice of using regional instead of global configurations may be judicious in some cases, depending on the scientific questions that are addressed. Some studies require the global configuration anyway; for example, the global ensemble simulation shows that part of the interannual intrinsic variability (ensemble spread) of the AMOC is generated in the South Atlantic, and is thus missing in the regional ensemble experiment. We are currently working on a publication dedicated to this subject (Leroux et al, 2017, in prep for J. of Climate) as also mentioned in the last paragraph of section 5.

4) *Could you provide more information of the restoring which is done in the simulation? Is there a sea surface salinity restoring, a sea surface temperature restoring?*

As in most oceanic hindcasts, there is a SSS restoring within each member that corresponds to a 300 days timescale over 50m (166.67 mm/day piston velocity; see Griffies et al, Ocean Modelling 2009).

This information is now given in Table 1. There is no explicit SST relaxation; please see our answer to the next question regarding the implicit SST relaxation due to the use of bulk formulae.

5) As you use bulk formulae to compute your atmospheric fluxes and to constrain your model, it is not true that you have strictly the same atmospheric forcing in all the members. Could you provide quantified informations of the variance of the atmospheric fluxes in the experiment? It will be useful to know if this variability is negligible or not.

Indeed, the atmospheric fluxes computed through bulk formulae somewhat differ among the ensemble because SSTs do so. However all members “see” the exact same atmospheric evolution, as we wrote in section 4.2 (“forced by the exact same atmospheric conditions”). We have clarified this point in the last paragraph of section 5.3, which now reads:

“This is expected from the design of these ensemble simulations: each ensemble member is driven through bulk formulae by the same atmospheric forcing function, but turbulent air-sea heat fluxes somewhat differ among the ensemble because SSTs do so. This approach induces an implicit relaxation of SST toward the same equivalent air temperature (Barnier et al, 1995) within each member, hence an artificial damping of the SST spread. These experiments thus only provide a conservative estimate of the SST intrinsic variability. Note that alternative forcing strategies may alleviate or remove this damping effect: ensemble mean air-sea fluxes may be computed online at each time step and applied identically to all members (see section 3.2). This alternative approach is the subject of ongoing work and will be presented in a dedicated publication.”

6) There is no discussion about impact of the number of members in the study, as you have 50 members in your global simulation it will be interesting to know how each member gives information and if the ensemble spread converges? This point could at least discussed in the perspectives.

Our paper mostly aims to present the probabilistic version of NEMO; choosing an appropriate number of ensemble members is an important concern for users, depending on their specific applications. This question is now shortly discussed in the conclusion, as follows;

The size N of the ensemble simulation depends on the objectives of the study, the desired accuracy of ensemble statistics, and the available computing resources. Our choice $N=50$ allows a good accuracy ($1/\sqrt{50} \approx \pm 14\%$) for estimating the ensemble means and standard deviations. Moreover, this choice allows the estimation of ensemble deciles (with 5 members per bin) for the detection of possibly bimodal or other non-gaussian features of ensemble PDFs; such behaviors were indeed detected in simplified ensemble experiments (e.g. Pierini, 2014) and may appear in ours. Given our preliminary tests with E-NATL025, $N=50$ appeared as a satisfactory compromise between our need for a long global $1/4^\circ$ simultaneous integration, our scientific objectives, PRACE rules (expected allocation and elapse time, jobs' duration, etc), and CURIE's technical features (processor performances, memory, communication cost). Our tests also indicate that the convergence of ensemble statistics with N depends on the variables, metrics and regions under consideration. For all these reasons, N must be chosen adequately for each study.

Fig. 3: Keeping the same color or symbol code between fig 3a and 3b could be more clear for reader

As suggested, the same color code has been kept between fig 3a and 3b.

Fig. 6 : There is no legend line for the Median

See our answer to reviewer 1 question 1.

REVIEWER 3:

We thank the reviewer for his/her interesting comments about the mathematical formulation in our paper. They helped us improving the generality and accuracy of our statements. We did our best to take them into account.

1) Please refer to interactive discussion for referee#3 comment#1

We agree that the mathematical background given in section 2 does not provide the most general mathematical framework, and does not encompass all possible ways of simulating model uncertainties. However, on the one hand, it is general enough to include everything that it is possible to do with our specific implementation. Our implementation indeed only introduces autoregressive processes (time-correlated and/or space-correlated), possibly followed by a nonlinear transformation to make them non-Gaussian, and all this is included in the given mathematical framework (by changing the definition of x and M as granted by the reviewer). And, on the other hand, using a more general mathematical framework would make this section more difficult to follow for non-mathematicians.

To answer the reviewer comment, we have thus decided to keep a simple mathematical framework, but we have added a word of caution explaining that this is only one possible approach, and an explanation that the ensemble method could also be used if another type of stochastic parameterization had been implemented.

(i) Before Eq. 3, the following sentence has been added:

"One possibility is for instance to modify Eq. 1 as follows:"

(ii) After Eq. 3, the following sentence has been added:

"Eq. (3) does not include all possible ways of introducing a stochastic parameterization in a dynamical model, but it is sufficient to include the implementation that is described in this paper (in particular, to include the use of space-correlated or time-correlated autoregressive processes by expanding the definition of $\{x\}$ in Eq. (3))."

(iii) The last sentence of the 4th paragraph of section 2 (before the final summary paragraph) has been rewritten:

"This Monte Carlo approach is very general and be also applied to any kind of stochastic parameterization (not only the particular case described by Eq. 3). It was first applied (...)."

2) Please refer to interactive discussion for referee#3 comment#2

Yes, we agree that the ensemble usually provides only a rough approximation of the probability distribution, and that the ensemble can give estimates of quantities (like time correlations) that are not contained in the marginal pdfs (for each time t) provided by the Fokker-Planck equation.

To clarify this, the following changes have been made in the text of the paper:

(i) the last sentence of section 2 has been replaced by:

"in this paper, this problem is solved using an ensemble simulation, which provides identically-distributed realizations from the probability distribution, and thus a way to compute any statistic of interest."

(ii) in the 3rd paragraph, *"a solution to Eq. (2) or (4)"* has been replaced by *"an approximate description of the probability distribution"*.

(iii) in section 3.3, the first sentence has been replaced by:

"Ensemble simulation (...) parameterization (as introduced in Eq. 3)."

3) Please refer to interactive discussion for referee#3 comment#3

Yes, we agree with the reviewer, the term *"equiprobable"* was not used correctly. It has been replaced by *"independent and identically distributed"* as suggested by the reviewer, or just suppressed where it was not useful.

Development of a probabilistic ocean modelling system based on NEMO 3.5: application at eddying resolution

Laurent Bessi res¹, St phanie Leroux^{2,3}, Jean-Michel Brankart^{2,3}, Jean-Marc Molines^{2,3}, Marie-Pierre Moine¹, Pierre-Antoine Bouttier^{2,3}, Thierry Penduff^{2,3}, Laurent Terray¹, Bernard Barnier^{2,3}, and Guillaume S razin^{1,2,3}

¹CNRS/CERFACS, CECI UMR 5318, Toulouse, France.

²CNRS, LGGE, F-38041 Grenoble, France.

³Univ. Grenoble Alpes, LGGE, F-38041 Grenoble, France.

Correspondence to: Laurent Bessi res (bessieres@cerfacs.fr)

NOTE TO REVIEWERS:

* Additional text has been highlighted in Chartreuse Green.

* Removed text has been striped.

Abstract. This paper presents the technical implementation of a new, probabilistic version of the NEMO ocean/sea-ice modelling system. Ensemble simulations with N members running simultaneously within a single executable, and interacting mutually if needed, are made possible through an enhanced MPI strategy including a double parallelization in the spatial and ensemble dimensions. An example application is then given to illustrate the implementation, performances and potential use of this novel probabilistic modelling tool. A large ensemble of 50 global ocean/sea-ice hindcasts has been performed over the period 1960-2015 at eddy-permitting resolution ($1/4^\circ$) for the OCCIPUT project. This application is aimed to simultaneously simulate the intrinsic/chaotic and the atmospherically-forced contributions to the ocean variability, from meso-scale turbulence to interannual-to-multidecadal time scales. Such an ensemble indeed provides a unique way to disentangle and study both contributions, as the forced variability may be estimated through the ensemble mean, and the intrinsic chaotic variability may be estimated through the ensemble spread.

1 Introduction

Probabilistic approaches, based on large ensemble simulations, have been helpful in many branches of Earth-system modelling sciences to tackle the difficulties inherent to the complex and chaotic nature of the dynamical systems at play. In oceanography, ensemble simulations have first been introduced for data assimilation purposes, in order to explicitly simulate and, given observational data, reduce the uncertainties associated to e.g. model dynamics, numerical formulation, initial states, atmospheric forcing (e.g. Evensen, 1994; Lermusiaux, 2006). This type of probabilistic approach is also used to accurately assess ocean model simulations against observations (e.g. Candille and Talagrand, 2005), or to anticipate on the design of satellite observational missions (e.g. Ubelmann et al., 2009).

21 Performing ensemble simulations can be seen as a natural way to take into account the internal variability inherent to any
 1 chaotic and turbulent system, by sampling a range of possible **and equiprobable** trajectories of this system
 2 **(independent and identically distributed).**

3 For example, long-term climate projections, or short-term weather forecasts, rely on large ensembles of atmosphere-ice-
 4 ocean coupled model simulations to simulate the probabilistic response of the climate system to various external forcing
 5 scenarii, or to perturbed initial conditions, respectively (e.g. *Palmer, 2006; Kay et al., 2015; Deser et al., 2016*).

6 The ocean is, like the atmosphere or the full climate system, a chaotic system governed by non-linear equations which couple
 7 various spatio-temporal scales. A consequence is that, in the turbulent regime (i.e. for $1/4^\circ$ or finer resolution), ocean models
 8 spontaneously generate a chaotic intrinsic variability under purely climatological atmospheric forcing, i.e. same repeated an-
 9 nual cycle from year to year. This purely intrinsic variability has a significant imprint on many ocean variables, especially in
 10 eddy-active regions, and develops on spatio-temporal scales ranging from mesoscale eddies up to the size of entire basins, and
 11 from weeks to multiple decades (*Penduff et al., 2011; Grégorio et al., 2015; Sérazin et al., 2015*). The evolution of this chaotic
 12 ocean variability under repeated climatological atmospheric forcing is sensitive to initial states. This suggests that turbulent
 13 oceanic hindcasts driven by the full range of atmospheric scales (e.g. atmospheric reanalyses) are likely to be sensitive to initial
 14 states as well, and their simulated variability should be interpreted as a combination of the atmospherically-forced and the
 15 intrinsic/chaotic variability.

16 On the other hand, NEMO climatological simulations at $\sim 2^\circ$ resolution (in the laminar non-eddy regime) driven by a
 17 repeated climatological atmospheric forcing are almost devoid of intrinsic variability (*Penduff et al., 2011; Grégorio et al.,*
 18 *2015*). Because $\sim 1/4^\circ$ -resolution OGCMs are now progressively replacing their laminar counterparts at ~ 1 - 2° resolution used
 19 in previous CMIP-type long-term climate projections (e.g. HighResMIP, *Haarsma et al., 2016*), it becomes crucial to better
 20 understand and characterize the respective features of the intrinsic and atmospherically-driven parts of the ocean variability,
 21 and their potential impact on climate-relevant indices.

22 Simulating, separating and comparing these two components of the oceanic variability requires an ensemble of turbulent
 23 ocean hindcasts, driven by the same atmospheric forcing, and started from perturbed initial conditions. The high computational
 24 cost of performing such ensembles at global or basin scale explains why only few studies have carried out this type of approach
 25 until now, and with small ensemble sizes (e.g. *Combes and Lorenzo, 2007; Hirschi et al., 2013*).

26 Building on the results obtained from climatological simulations, the ongoing OCCIPUT project (*Penduff et al., 2014*) aims
 27 to better characterize the chaotic low-frequency intrinsic variability (LFIV) of the ocean under a fully-varying atmospheric
 28 forcing, from a large (50-member) ensemble of global ocean/sea-ice hindcasts at $1/4^\circ$ resolution over the last 56 years (1960-
 29 2015). The intrinsic and the atmospherically-forced parts of the ocean variability are thus simulated simultaneously under
 30 fully-varying realistic atmosphere, and may be estimated from the ensemble spread and the ensemble mean, respectively. This
 31 strategy also allows to investigate the extent to which the full atmospheric variability may excite, modulate, damp, or pace
 32 intrinsic modes of oceanic variability that were identified from climatological simulations. OCCIPUT mainly focuses on the
 33 interannual-to-decadal variability of ocean quantities having a potential impact on the climate system, such as Sea Surface
 34 Temperature (SST), Meridional Overturning Circulation (MOC), and upper Ocean Heat Content (OHC).

35 This paper presents the technical implementation of the new, fully probabilistic version of the NEMO modelling system
 1 required for this project. It stands at the interface between scientific purposes and new technical developments implemented
 2 in the model. The OCCIPUT project is presented here as an application, to illustrate the system requirements and numerical
 3 performances. The mathematical background supporting our probabilistic approach is detailed in section 2. Section 3 describes
 4 the new technical developments introduced in NEMO to simultaneously run multiple members from a single executable (allow-
 5 ing the online computation of ensemble statistics), with a flexible input/output strategy. Section 4 presents the implementation
 6 of this probabilistic model to perform regional and global $1/4^\circ$ ensembles, both performed in the context of OCCIPUT. The
 7 strategy chosen to trigger the growth of the ensemble spread, and the numerical performances of both implementations are
 8 also discussed. Section 5 finally presents some preliminary results from OCCIPUT to further illustrate potential scientific
 9 applications of this probabilistic approach. A summary and some concluding remarks are given in section 6.

10 2 From deterministic to probabilistic ocean modelling: mathematical background

11 The classical, deterministic ocean model formulation can be written as

$$12 \quad d\mathbf{x} = \mathcal{M}(\mathbf{x}, t)dt \quad (1)$$

13 where $\mathbf{x} = (x_1, x_2, \dots, x_N)$ is the model state vector; t is time; and \mathcal{M} is the model operator, containing the expression of the
 14 tendency for every model state variable. An explicit time-dependence is included in the model operator since the tendencies
 15 depend on the time-varying atmospheric forcing.

16 Computing a solution to Eq. (1) requires the specification of the initial condition at $t = 0$, from which the future evolution
 17 of the system is fully determined. OCCIPUT investigates how perturbations in initial conditions evolve and finally affect the
 18 statistics of climate-relevant quantities. This problem may be addressed probabilistically by solving the Liouville equation:

$$19 \quad \frac{\partial p(\mathbf{x}, t)}{\partial t} = - \sum_{k=1}^N \frac{\partial}{\partial x_k} [\mathcal{M}(\mathbf{x}, t) p(\mathbf{x}, t)] \quad (2)$$

20 where $p(\mathbf{x}, t)$ is the probability distribution of the system state at time t . Eq. (2) shows that this distribution is simply advected in
 21 the phase space by local model tendencies. In chaotic systems, small uncertainties in the initial condition ($p(\mathbf{x}, 0)$ concentrated
 22 in a very small region of the phase space) yield diverging trajectories; such systems are poorly predictable on the long range.
 23 In the turbulent ocean, the mesoscale turbulence and the low-frequency intrinsic variability (LFIV) have a chaotic behaviour.
 24 They both belong to what we will call more generally intrinsic variability in sections 4 and 5, in the sense that they do not
 25 result from the forcing variability but spontaneously emerge from the ocean even with constant or seasonal forcing. Because
 26 this intrinsic variability is chaotic, it can only be described in a statistical sense and the probabilistic approach of Eq. (2) is
 27 required.

28 In addition to uncertainties in the initial condition, it is sometimes useful to assume that the model dynamics itself is un-
 29 certain. This leads to a non-deterministic ocean model formulation, in which model uncertainties are described by stochastic

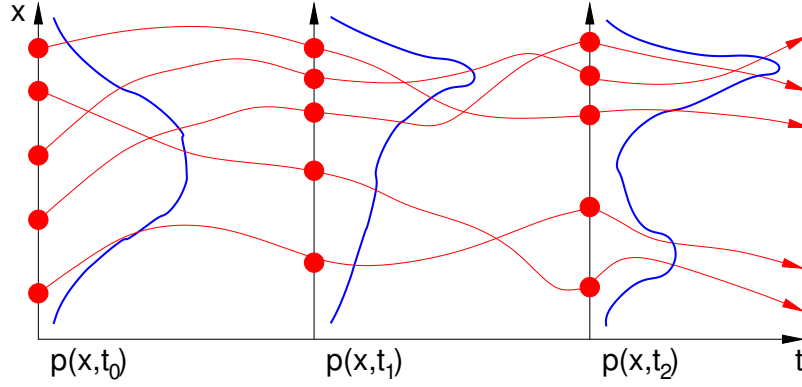


Figure 1. Schematic of an ensemble simulation (red trajectories), as an approximation to the simulation of an evolving probability distribution (in blue).

processes. One possibility is for instance to modify Eq. (1) as follows:

$$d\mathbf{x} = \mathcal{M}(\mathbf{x}, t)dt + \Sigma(\mathbf{x}, t)d\mathbf{W}_t \quad (3)$$

In this equation **very-general**, \mathbf{W}_t is an M -dimensional standard Wiener process, and $\Sigma(\mathbf{x}, t)$ is an $N \times M$ matrix, describing the influence of these processes on the model tendencies.

Eq. (3) does not include all possible ways of introducing a stochastic parameterization in a dynamical model, but it is sufficient to include the implementation that is described in this paper (in particular, to include the use of space-correlated or time-correlated autoregressive processes by expanding the definition of \mathbf{x} in Eq. (3)).

With this additional stochastic term, Liouville equation transforms to Fokker-Planck equation:

$$\begin{aligned} \frac{\partial p(\mathbf{x}, t)}{\partial t} = & - \sum_{k=1}^N \frac{\partial}{\partial x_k} [\mathcal{M}(\mathbf{x}, t)p(\mathbf{x}, t)] \\ & + \frac{1}{2} \sum_{k=1}^N \sum_{l=1}^N \frac{\partial^2}{\partial x_k \partial x_l} [D_{kl}(\mathbf{x}, t)p(\mathbf{x}, t)] \end{aligned} \quad (4)$$

where $D_{kl}(\mathbf{x}, t) = \sum_{p=1}^M \Sigma_{jp}(\mathbf{x}, t)\Sigma_{lp}(\mathbf{x}, t)$. The probability distribution $p(\mathbf{x}, t)$ is thus affected by the stochastic diffusion tensor $\mathbf{D}(\mathbf{x}, t)$ during its advection in the phase space by local model tendencies.

However, since Eqs. (2) and (4) are partial differential equations in an N -dimensional space, they generally cannot be solved explicitly for large size systems. Only an approximate description of $p(\mathbf{x}, t)$ can be obtained in most practical situations. A common solution is to reduce the description of $p(\mathbf{x}, t)$ to a moderate size sample, which can be viewed as a Monte Carlo approximation to Eqs. (2) and (4). This approach is illustrated in Figure 1. The computation is initialized by a sample of the initial probability distribution $p(\mathbf{x}, t_0)$ (on the left in the figure), and each member of the sample is used as a different initial condition to Eqs. (1) and (3). The classical model operator can then be used to produce an ensemble of model simulations (red trajectories in the figures), which provide a sample of the probability distribution at any future time, e.g. $p(\mathbf{x}, t_1)$, or $p(\mathbf{x}, t_2)$.

1 This Monte Carlo approach was first applied to ocean models in the framework of the ensemble Kalman filter (*Evensen*, 1994)
2 to solve ocean data assimilation problems.

3 This Monte Carlo approach is very general and can be also applied to any kind of stochastic parameterization (not only the
4 particular case described by Eq. (3)). It was first applied to ocean models in the framework of the ensemble Kalman filter
5 (*Evensen*, 1994) to solve ocean data assimilation problems.

6 In summary, Eq. (1) describes the problem that is classically solved by the NEMO model; Eq. (3) is a modification of this
7 problem with stochastic perturbations of the model equations that explicitly simulate model uncertainties;

8 Eqs. (2) and (4) represent the kind of problems that we propose to solve in this paper using ensemble NEMO simulations.
9 in this paper, this problem is solved using an ensemble simulation, which provides identically-distributed realizations from the
10 probability distribution, and thus a way to compute any statistic of interest.

11 3 Performing ensemble simulations with NEMO

12 The NEMO model (Nucleus for a European Model of the Ocean), described in *Madec* (2012), is used for oceanographic
13 research, operational oceanography, seasonal forecasts and climate studies. This system embeds various model components
14 (see <http://www.nemo-ocean.eu/>), including a circulation model (OPA, Océan PARallélisé), a sea-ice model (LIM, Louvain-la-
15 Neuve Ice model), and ecosystem models with various levels of complexity. Every NEMO component solves partial differential
16 equations discretized on a three-dimensional grid using finite-difference approximations. The purpose of this section is to
17 present the technical developments introduced in our probabilistic NEMO version, and to make the connection between these
18 new developments and existing NEMO features.

19 3.1 Ensemble NEMO parallelization

20 The standard NEMO code is parallelized with MPI (Message Passing Interface) using a domain decomposition method. The
21 model grid is divided in rectangular subdomains ($i = 1, \dots, n$), so that the computations associated to each subdomain can be
22 performed by a different processor of the computer. Spatial finite-difference operators require knowledge of the neighbouring
23 grid points, so that the subdomains must overlap to allow the application of these operators on the discretized model field.
24 Whenever needed, the overlapping regions of each subdomain must be updated using the computations made for the neigh-
25 bouring subdomains. The NEMO code provides standard routines to perform this update. These routines use MPI to get the
26 missing information from the other processors of the computer. This communication between processors makes the connection
27 between subdomains in the model grid.

28 In practice, upon initialization one MPI communicator is defined with as many processors as subdomains, each processor is
29 associated with a subdomain and knows which are its neighbours.

30 Ensemble simulations may be performed with NEMO by a direct generalization of the standard parallelization procedure
31 described above. In other words, our ensemble simulations are performed from one single call to the NEMO executable, simply
32 using more processors to run all members in parallel. This technical option is both natural and unnatural.

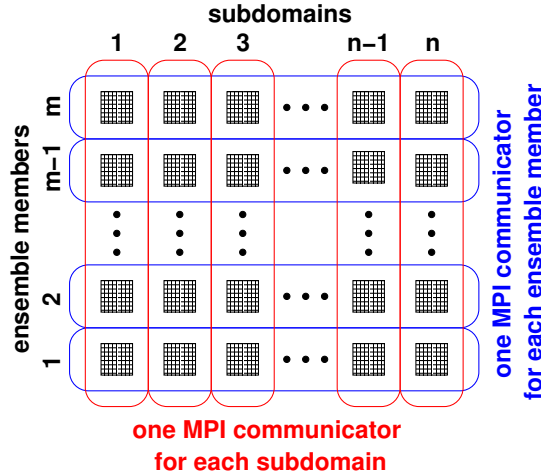


Figure 2. Schematic of the double parallelization introduced in NEMO: each processor (black squares) is dedicated to the computations associated to one model subdomain and one ensemble member. There is one MPI communicator within each ensemble member (in blue) to allow communications between neighbouring subdomains as in the standard NEMO parallelization; and there is one MPI communicator within each model subdomain (in red) to allow communication between ensemble members (e.g. to compute ensemble statistics online if needed). The total number of processors is thus equal to the product of the ensemble size by the number of subdomains ($m \times n$).

30 It is natural since an ensemble simulation provides a solution to Eq. (2) or (4); it is thus conceptually appealing to advance
 31 all members together in time.

It is natural since an ensemble simulation provides an approximate description of the probability distribution; it is thus conceptually appealing to advance all members together in time.

2 It is unnatural since independent ensemble members may be run separately (in parallel, or successively) using independent
 3 calls to NEMO. However, the solution we propose is so straightforward that there is virtually no implementation cost, and
 4 is more flexible since the ensemble members may be run independently, by groups of any size, or all together. Furthermore,
 5 running all ensemble members together provides a new interesting capability: the characteristics of the probability distribu-
 6 tion $p(\mathbf{x}, t)$ in Eq. (2) or (4) may be computed online, virtually at every time step of the ensemble simulation. This has been
 7 done using MPI to gather the required information from every member of the ensemble. These MPI communications make a
 8 natural connection between ensemble members, as a sample of the probability distribution $p(\mathbf{x}, t)$.

9 In practice, this implementation option only requires that at the beginning of the NEMO simulation, one MPI communicator
 10 is defined for each ensemble member, each one with as many processors as subdomains, so that each processor knows to
 11 which member it belongs, on which subdomain it is going to compute and what are its neighbours. Inside each of these
 12 communicators, each ensemble member may be run independently from the other members, without changing anything else
 13 in the NEMO code. However, all members are obviously not supposed to behave exactly the same: the index of the ensemble
 14 member must have some influence on the simulation. This influence may be in the name of the files defining the initial
 15 condition, parameters or forcing, or in the seeding of the random number generator (if a random forcing is applied, as in Eq. 3).

16 The index of the ensemble member must also be used to modify the name of all output files, so that the output of different
 17 members is saved in different files. As it appears, this implementation of ensemble simulations does not require much coding
 18 effort (a few tens of lines in NEMO, partly because most of the basic material was already available in the original code). More
 1 technical details about this can be found in appendix 7.

2 In summary, the NEMO ensemble system relies on a double parallelization, over model subdomains and over ensemble
 3 members, as illustrated in Figure 2. In this algorithm, ensemble simulations are thus intricately linked to MPI parallelization.
 4 There is no explicit loop over the ensemble members; this loop is done implicitly through MPI; running more ensemble
 5 members means either using more processors or using less processors for each member.

6 3.2 Online ensemble diagnostics

7 As mentioned above, one important novelty offered by the ensemble NEMO parallelization is the ability to compute online any
 8 feature of the probability distribution $p(\mathbf{x}, t)$. This can be done within additional MPI communicators connecting all ensemble
 9 members for each model subdomain (in red in Fig. 2). MPI sums in these communicators are for instance immediately sufficient
 10 to estimate:

11 – the mean of the distribution

$$12 \quad \mu_k = \int x_k p(\mathbf{x}, t) d\mathbf{x} \quad \sim \quad \tilde{\mu}_k = \frac{1}{m} \sum_{j=1}^m x_k^{(j)} \quad (5)$$

13 where x_k is one of the model state variable; $x_k^{(j)}$ is this variable simulated in member j ; μ_k the mean of the distributon
 14 for this variable; and $\tilde{\mu}_k$, the estimate of the mean obtained from the ensemble. It is interesting to note that the sum over
 15 ensemble members in Eq. (5) is not explicitly coded in NEMO, it is performed instead by a single call to MPI, which
 16 computes the sums over all processors of the ensemble communicators (in red in Fig. 2). The same remark also applies
 17 to the sums in the following equations.

18 – the variance of the distribution

$$19 \quad \sigma_k^2 = \int (x_k - \mu_k)^2 p(\mathbf{x}, t) d\mathbf{x}$$

$$20 \quad \sim \quad \tilde{\sigma}_k^2 = \frac{1}{m-1} \sum_{j=1}^m \left(x_k^{(j)} - \tilde{\mu}_k \right)^2$$

21 where σ_k^2 is the variance of the distributon for variable x_k ; and $\tilde{\sigma}_k^2$, the estimate obtained from the ensemble. The
 22 ensemble standard deviation is simply the square root of $\tilde{\sigma}_k^2$.

23 – Ensemble covariance between 2 variables at the same model grid point:

$$24 \quad \gamma_{kl} = \int (x_k - \mu_k) (x_l - \mu_l) p(\mathbf{x}, t) d\mathbf{x}$$

$$25 \quad \sim \quad \tilde{\gamma}_{kl} = \frac{1}{m-1} \sum_{j=1}^m \left(x_k^{(j)} - \tilde{\mu}_k \right) \left(x_l^{(j)} - \tilde{\mu}_l \right)$$

where γ_{kl} is the covariance between variables x_k and x_l , and $\tilde{\gamma}_{kl}$, the estimate obtained from the ensemble.

This is directly generalizable to the computation of higher order moments (skewness, kurtosis), which is then reduced to MPI sums in the ensemble communicators. Moreover, simple MPI algorithms can also be designed to compute online many other probabilistic diagnostics, as for instance the rank of each member in the ensemble, and from there, estimates of quantiles of the probability distribution. Specific applications of this feature are discussed in section 5.

This online estimation of the probability distribution, via the computation of ensemble statistics, opens another interesting new capability: the solution of the model equations may now depend on ensemble statistics, available at each time step if needed.

This may be useful for certain applications: a simple example would be for example the relaxation (nudging) of the model simulation towards some climatological data. In this case, indeed it could be much better to relax the ensemble mean than the individual ensemble members, to avoid damping the intrinsic variability of the system by the relaxation.

For instance, it may be interesting to relax the modeled forced variability towards reference (e.g. reanalyzed or climatological) fields, with no explicit damping of the intrinsic variability: the nudging term would involve the current ensemble mean and be applied identically to all members at the next time step, resulting in a simple “translation” of the entire ensemble distribution toward the reference field.

Other applications, such as ensemble data assimilation, may also require an online control of the ensemble spread, which is hereby made possible within NEMO.

3.3 Connection with NEMO stochastic parameterizations

Ensemble simulations are directly connected to stochastic parameterizations: they provide a solution to Eq. (4) rather than Eq. (2) when a stochastic forcing is applied to the model (as in Eq. 3).

Ensemble simulations are directly connected to stochastic parameterizations (as introduced in Eq. 3).

In NEMO, stochastic parameterizations have recently been implemented to explicitly simulate the effect of uncertainties in the model (*Brankart et al.*, 2015). In practice, this is done by generating maps of autoregressive processes, which can be used to introduce perturbations in any component of the model. In *Brankart et al.* (2015), examples are provided to illustrate the effect of these perturbations in the circulation model, in the ecosystem model and in the sea ice model. For instance, a stochastic parameterization was introduced in the circulation model to simulate the effect of unresolved scales in the computation of the large scale density gradient, as a result of the nonlinearity of the sea water equation of state (*Brankart*, 2013). This particular stochastic parameterization is switched on during one year in order to initiate the dispersion of the OCCIPUT ensemble simulations started from a single initial condition (see section 4).

3.4 Connection with NEMO data assimilation systems

Ensemble model simulations are also key in ensemble data assimilation systems: they propagate in time uncertainties in the model initial condition, and provide a description of model uncertainties in the assimilation system (e.g. using stochastic perturbations). Data assimilation can then be carried out by conditioning this probability distribution to the observations whenever

1 they are available. The ensemble data assimilation method that is currently most commonly used in ocean applications is the
2 Ensemble Kalman filter (*Evensen, 1994*), which performs the observational update of the model probability distributions with
3 the assumption that they are Gaussian. However, it has been recently suggested that the Gaussian assumption is often insuf-
4 ficient to correctly describe ocean probability distributions, and that more general methods using for instance anamorphosis
5 transformations (*Bertino et al., 2003; Brankart et al., 2012*) or a particle filtering approach (e.g. *Van Leeuwen, 2009*) may be
6 needed. One of the purpose of the SANGOMA European project is precisely to develop such more general methods for ocean
7 applications, and to implement them within NEMO-based ocean data assimilation systems (e.g. *Candille et al., 2015*). In these
8 methods, the role of ensemble NEMO simulations is even more important since they require a more detailed description of the
9 probability distributions (as compared to the Gaussian assumption, which only requires the mean and the covariance). The im-
10 portance of ensemble simulations in data assimilation certainly explains why the ensemble NEMO parallelisation (introduced
11 above) has been first applied within SANGOMA, to assimilate altimetric observations in an eddying NEMO configuration of
12 the North Atlantic (*Candille et al., 2015*).

13 **3.5 Connection with the NEMO observation operator and model assessment metrics**

14 Another important benefit of the probabilistic approach is to consolidate and objectivate statistical comparisons between actual
15 observations and model-derived ensemble synthetic observations. Probabilistic assessment metrics are commonly used in the
16 atmospheric community (e.g. *Toth et al., 2003*) but are quite new in oceanography. Briefly speaking, these methods generally
17 quantify two attributes of an ensemble simulation: the *reliability* and the *resolution*. An ensemble is reliable if the simulated
18 probabilities are statistically consistent with the observed frequencies. The ensemble resolution is related to the system ability to
19 discriminate between distinct observed situations. If the ensemble is reliable, the resolution is directly related to the information
20 content (or the spread) of the probability distribution. A popular measure of these two attributes is for instance provided by
21 the Continuous Rank Probability Score (CRPS), which is based on the square difference between a cumulative distribution
22 function (cdf) as provided by the ensemble simulation and the corresponding cdf of the observations (*Candille and Talagrand,*
23 *2005*).

24 In OCCIPUT, such probabilistic scores will be computed from real observations and from the ensemble synthetic obser-
25 vations (along-track Jason-2 altimeter data and ENACT-ENSEMBLE temperature and salinity profile data) generated online
26 using the existing NEMO observation operator (NEMO-OBS module). NEMO-OBS is used exactly as in standard NEMO
27 within each member of the ensemble, hence providing an ensemble of model equivalents for each observation rather than a sin-
28 gle value. Probabilistic metrics (i.e. CRPS score) will then be computed to assess the reliability and resolution of the OCCIPUT
29 simulations.

30 **3.6 Connection with NEMO I/O strategy**

31 Our implementation of ensemble NEMO using enhanced parallelization is technically not independent from the NEMO I/O
32 strategy. In NEMO indeed, the input and output of data is managed by an external server (XIOS, for XML IO Server), which
33 is run on a set of additional processors (not used by NEMO). The behavior of this server is controlled by an XML file, which

1 governs the interaction between XIOS and NEMO, and which defines the characteristics of input and output data: model fields,
2 domains, grid, I/O frequencies, time averaging for outputs, . . . To exchange data with disk files, every NEMO processor makes
3 a request to the XIOS servers, consistently with the definitions included in the XML file. In this operation, the XIOS servers
4 buffer data in memory, with the decisive advantage of not interrupting NEMO computations with the reading or writing in
5 disk files. One peculiarity of this buffering is that each XIOS server reads and writes one stripe of the global model domain
6 (along the second model dimension), and thus exchanges data with processors corresponding to several model subdomains. To
7 optimize the system, it is obviously important that the number of XIOS servers (and thus the size of these stripes) be correctly
8 dimensioned according to the amount of I/O data, which may heavily depend on the model configuration and on the definition
9 of the model outputs.

10 To use XIOS with our implementation of ensemble NEMO for OCCIPUT, we thus had to take care of the two following
11 issues. First, different ensemble members must write different files. This problem could be solved because XIOS was already
12 designed to work with a coupled model, and can thus deal with multiple contexts, one for each of the coupled model compo-
13 nents. It was thus directly possible to define one context for each ensemble member, just as if they were different components
14 of a coupled model. Second, in ensemble simulations, the amount of output data is proportional to the ensemble size, so that
15 the number of XIOS servers must be increased accordingly, with some care however, because the size of the data stripe that is
16 processed by each server should not be reduced too much.

17 **4 Example of application: the OCCIPUT project**

18 The implementation of this ensemble configuration of NEMO was motivated to a large extent by the scientific objectives of
19 the OCCIPUT project, described in the introduction. In this section, we present two ensemble simulations, E-NATL025 and
20 E-ORCA025, performed in the context of this project. We focus on the model set-up, the integration strategy, the numerical
21 performances of the system, followed by a few illustrative preliminary results in section 5.

22 **4.1 Regional and global configurations**

23 E-ORCA025 is the main ensemble simulation aimed by OCCIPUT. It is a 50-member ensemble of global ocean/sea-ice hind-
24 casts at $1/4^\circ$ horizontal resolution, run for 56 years. Before performing this large ensemble, a smaller (20-year \times 10-member)
25 regional ensemble simulation, E-NATL025, was performed on the North Atlantic domain in order to test the new system
26 implementation and to validate the stochastic perturbation strategy for triggering the growth of the ensemble dispersion. The
27 global and the regional ensemble configurations are both based on version 3.5 of NEMO, and use a $1/4^\circ$ eddy-permitting quasi-
28 isotropic horizontal grid (~ 27 km at the equator), the grid size decreasing poleward. Table 1 summarizes the characteristics
29 of ensembles. The model parameters are very close to those used in the DRAKKAR-ORCA025¹ one-member setups (*Barnier*
30 *et al.*, 2006), the present setup using a greater number (75) of vertical levels (see table 1). They are also close to those used for

¹DRAKKAR-ORCA025 website: <http://www.drakkar-ocean.eu/global-models/orca025>

	REGIONAL ENSEMBLE	GLOBAL ENSEMBLE
ENSEMBLE NAME:	E-NATL025	E-ORCA025
SPATIAL DOMAIN:	North Atlantic (21°S-81°N)	Global
HORIZONTAL RESOLUTION:	1/4° (486×530 grid-points)	1/4° (1441×1021 grid-points)
VERTICAL RESOLUTION:	46 levels	75 levels
ENSEMBLE SIZE:	10 members	50 members
TIME PERIODE COVERED BY ENSEMBLE:	1993-2012	1960-2015
STOCHASTIC PERTURBATION PHASE:	1 year (1993)	1 year (1960)
SURFACE BOUNDARY CONDITIONS:	DFS5.2 (<i>Dussin et al.</i> , 2016) Turbulent air-sea fluxes : bulk formula. (with absolute wind) SSS relaxation in each member: 50m/(300 days) piston velocity.	

Table 1. Main characteristics of the NEMO 3.5 set-up used for the regional and global OCCIPUT ensembles.

1 the 327-year ORCA025-MJM01 one-member climatological simulation used in *Penduff et al.* (2011), *Grégorio et al.* (2015)
2 and *Sérazin et al.* (2015) to study various imprints of the LFIV under seasonal atmospheric forcing.

3 **4.2 Integration and stochastic perturbation strategies**

4 A one-member spin-up simulation is first performed for each ensemble. For the regional ensemble (E-NATL025), it is per-
5 formed from 1973 (cold start) to 1992, forced with DFS.5.2 atmospheric conditions (*Dussin et al.*, 2016). For the global
6 ensemble (E-ORCA025), the spin-up strategy has to be adapted to match the OCCIPUT objective to perform the ensemble
7 hindcast over the longest period available in the atmospheric forcing DFS5.2 (i.e. 1960-2015). The one-member spin-up sim-
8 ulation is thus performed as follows: (1) it is first forced by the standard DFS5.2 atmospheric forcing from January 1st, 1958
9 (cold start) to December 31st, 1976; (2) this simulation is continued over January 1977 with a modified forcing function that
10 linearly interpolates between the 1st of January 1977 to the 31st of January 1958; (3) the standard DFS5.2 forcing is applied
11 again normally from February 1st, 1958 to the end of 1959. This 21-year spin-up (1958 to 1977, then 1958 to 1959) thus
12 includes a smooth artificial transition from January 1977 back to January 1958. This choice was made as a compromise to
13 maximize the duration of the single-member spin-up simulation and of the subsequent ensemble hindcast, while minimizing
14 the perturbation in the forcing during the transition, since 1977 was found to be a reasonable analogue of 1958 in terms of key
15 climate indices (El Niño Southern Oscillation, North Atlantic Oscillation, and Southern Annular Mode).

16 The N members of both ensemble simulations (i.e. $N=10$ for E-NATL025 and $N=50$ for E-ORCA025) are started at the end
1 of the single-member spin-up; a weak stochastic perturbation in the density equation, as described by equation 3 and in section
2 3.3 (see also *Brankart et al.*, 2015) is then activated within each member. This stochastic perturbation is only applied for one
3 year to seed the ensemble dispersion (during 1993 for E-NATL025, during 1960 years for E-ORCA025). It is then switched off
4 throughout the rest of the ensemble simulations. Once the stochastic perturbation is stopped, the N members are thus integrated
5 from slightly perturbed initial conditions (i.e. 19 more years for E-NATL025 and 55 more years for E-ORCA025), but forced by
6 the exact same atmospheric conditions (DFS5.2, *Dussin et al.*, 2016). The code is parallelized with the double-parallelization
7 technique described in 3.1 so that the N members are integrated simultaneously through one single executable.

8 4.3 Performance of the NEMO ensemble system in OCCIPUT configurations

9 The regional ensemble (E-NATL025) was performed to test the system implementation and to calibrate the global config-
10 uration. The global ensemble simulation E-ORCA025 represents in total 2821 cumulated years of simulation ($56 \text{ yrs} \times 50$
11 $\text{members} + 21 \text{ yrs}$ of one-member spin-up) over 110 million grid points ($\text{Lon} \times \text{Lat} \times \text{Depth} = 1442 \times 1021 \times 75$). As con-
12 firmed thereafter in Figure 3, integrating such a system within one executable with reasonable wall-clock time and managing its
13 outputs lies beyond national or regional European centres computational capabilities (i.e Tier-1 systems) and requires systems
14 that can provide European top capabilities, which are beyond the Petaflops level (i.e Tier-0 systems).

15 All simulations were performed between 2014 and 2016 on the French Tier-0 Curie supercomputer, supported by PRACE
16 (Partnership for Advanced Computing in Europe) and GENCI (Grand Equipement National de Calcul Intensif, French rep-
17 resentative in PRACE) grants (19.10^6 HCPU, see details below). Curie is a Bull system (BullX series designed for extreme
18 computing) based on Intel processors. The architecture used for the simulations is the one of the “Curie thin nodes” con-
19 figuration (Curie-TN), which is mainly targeted at MPI parallel codes and includes more than 80,000 Intel’s Sandy-Bridge
20 computing cores (Peak frequency per core: 2.7 GHz) gathered in 16-cores nodes of 64 GB of memory.

21 Preliminary tests showed that the one-member ORCA025 configuration has a good scalability up to 400 cores on Curie-TN
22 (not shown). In order to test the ensemble global configuration on Curie-TN, short 180-step experiments were run, disregarding
23 the first and last steps (which correspond to reading and writing steps, respectively, that are performed only once during
24 production jobs). The performance of the system was measured in steps per minute by analyzing the 160 steps in between
25 (steps 10 to 170). Figure 3.a shows this measure of the system performance (in step/min) as a function of the number of
26 members, for different domain decompositions (64, 128, 160, 256 and 400 cores/member). It appears that the performance is
27 independent of the ensemble size for domain decomposition up to 160 domains per member. When more than 160 domains
28 per member are used, the performance starts to decrease for increasing ensemble size, from 25 members (resp. 10) for the
29 decomposition with 256 (resp. 400) domains per member. Fluctuations in step per minute may appear (see the performance for
30 the decomposition with 400 domains per member and 25 members on figure 3.a), depending on machine load and files system
31 stability (the performance of this specific point has not been reassessed for CPU cost reasons). The scalability of the global
32 ensemble configuration E-ORCA025 as aimed in OCCIPUT ($N=50$) is shown in Figure 3.b: the efficiency is measured as the

ratio of the observed speedup to the theoretical speedup, relative to the smallest domain decomposition tested, i.e. with 3200 cores (50×64). The efficiency is remarkably good and remains around 90% for 20.000 used cores.

Based on these performance tests, a domain decomposition with relatively few cores was chosen in order to maintain a manageable rate of I/Os. The decomposition with 128 cores per member has been retained (corresponding to the red line on Figure 3) so that $50 \times 128 = 6400$ cores are used for the ensemble-NEMO system.

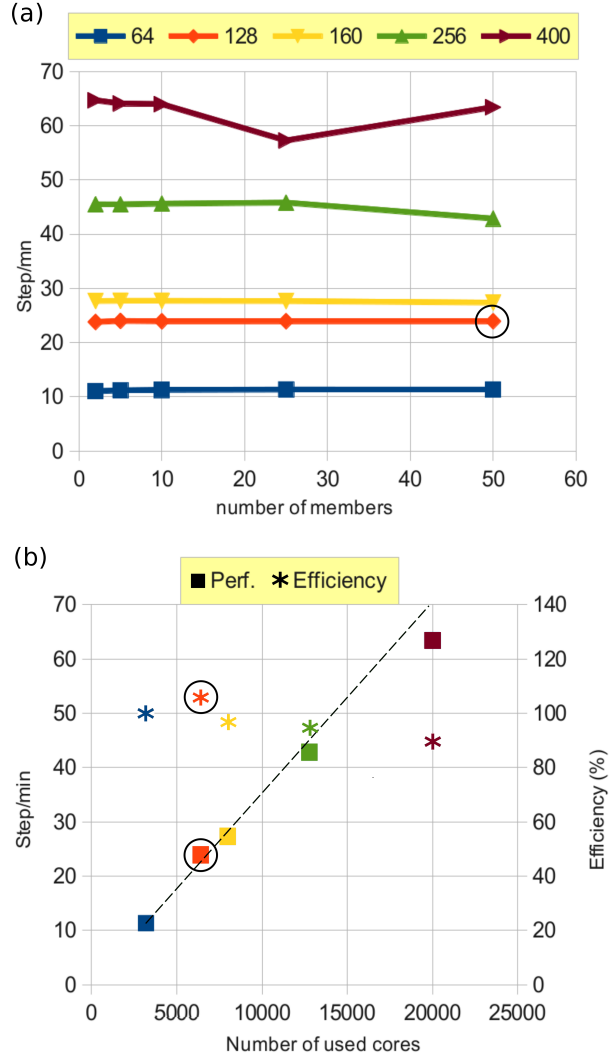


Figure 3. (a) Performance of the global ensemble configuration as a function of ensemble size N , for five domain decompositions: 64, 128, 160, 256 and 400 cores per member (colored lines). (b) Performance in steps per minute and efficiency in % of the global ensemble configuration with 50 members. The dotted line represents the theoretical speedup. The number of cores corresponds here to N times the number of subdomains per member. Our final choice (50 members, 128 cores per member) is indicated with black circles.

As suggested, same color code has been kept between fig 3a and 3b.

4 In order to optimize and to make the I/O data flux management flexible, 40 XIOS servers have been run as independant
5 MPI tasks in detached mode, allowing the overlap of I/O operations with computations. Compared to the 10-member regional
6 case, the 50-member global case required a larger XIOS buffer size. For this reason, each of the 40 XIOS instances was run
7 on a dedicated and exclusive Curie “thin node”, allowing each server to use the entire memory available on each 16-core node
8 (i.e 64 GB); the 40 XIOS servers thus used $16 \times 40 = 640$ cores in total. The integration of the 50-member global E-ORCA025
9 ensemble therefore required the use of $6400 + (40 \times 16) = 7040$ cores.

10 XIOS makes use of parallel file systems capabilities via the Netcdf4-HDF5 format, that allows both online data compression
11 and parallel I/O. Therefore, XIOS is used in “multiple file” mode where each XIOS instance writes a file for one stripe of the
12 global domain, yielding 40 files times 50 members for each variable and each time. At the end of each job, the 40 stripes are
13 recombined on-the-fly into global files.

14 Preliminary tests have shown that the 50-member E-ORCA025 global configuration performs about 20 steps per minute,
15 including actual I/O fluxes and additional operations (e.g. production of ensemble synthetic observations). Since the numer-
16 ical stability of this global setup requires a model time step of 720s, about 2 million time steps, 85 days of elapse time,
17 and about 14.4 million core hours were needed in theory to perform the 56-year OCCIPUT ensemble simulation. The fi-
18 nal CPU cost of the global ensemble experiment was about 19 million CPU hours, due to fluctuations in model efficiency,
19 occasional problems on file systems which required the repetition of certain jobs, the need to decrease the model time step
20 (increased high-frequency variance in the wind forcing data over the last decades) and the online computation of ensem-
21 ble diagnostics (high-frequency ensemble covariances, all terms of the heat content budget ensemble) over the last decade.

The cost of online ensemble diagnostics depends on the call frequency, number and size of the concerned fields, on the
architecture of the machine and the performance of communications. Our online ensemble diagnostics concerned a few two-
dimensional fields at hourly to monthly frequencies, and had a negligible cost.

23 The final E-ORCA025 global database is saved in Netcdf4-HDF5 format (chunked and compressed, compression ratio in
24 *italics* below). The primary dataset produced by the model consists in the following: monthly averages for full-3D fields (56 yr
25 $\times 12 \text{ months} \times 50 \text{ members} \times 2.8 \text{ GB} \times 41.5\% = 39 \text{ TB}$), 5-day averages for sixteen 2D-fields ($56 \text{ yr} \times 50 \text{ members} \times 6.8 \text{ GB}$
26 $\times 30\% = 6 \text{ TB}$), the Jason-2 and ENACT-ENSEMBLES ensemble synthetic observations (5TB), and hourly ensemble statistics
27 for key variables (1 TB). One restart file per member and per year is also archived (about 35 TB after compression). We then
28 computed a secondary dataset, consisting in 50-member yearly/decadal averages of the 3D-fields (2 TB), ensemble deciles of
1 monthly/yearly/decadal 3D-fields (6 TB), and data associated with on-line monitoring (1 TB). The total output amounts to less
2 than 100TB and 100.000 inodes on the Curie-TN file system.

3 5 Preliminary results from the OCCIPUT application

4 We now present some preliminary results from the regional and global OCCIPUT ensemble simulations described in section
5 4.1, in order to illustrate the concepts and the technical implementation presented above.

6 5.1 Probabilistic interpretation

7 Figure 4 shows for the 10-member regional ensemble the 1993-2012 timeseries of monthly temperature anomalies at depth 93
8 m at two contrasting grid points: in the Gulf Stream and in the middle of the North Atlantic subtropical gyre. Panels a. and c.
9 represent a sample of N equiprobable evolutions trajectories of the temperature given the identical atmospheric evolution
10 that forces all members.

11 These temperature anomalies were computed by first removing the long-term non-linear trend of the timeseries derived from
12 a local regression model (as in Grégorio *et al.*, 2015). This detrending step acts as a non-linear high-pass temporal filter with
13 negligible end-point effect (LOESS detrending, e.g. Cleveland *et al.*, 1992; Cleveland and Loader, 1996), which successfully
14 removes the unresolved imprints of very low-frequency variabilities (of forced or intrinsic origin), and possible non-linear
15 model drifts. We focus here on the ocean variability that is fully resolved in the 20-yr regional simulation output; we thus
16 choose to remove the total long-term trend of each member individually prior to plotting/analyzing the ensemble statistics
17 presented here. The mean seasonal cycle computed over the ensemble has also been removed from the monthly timeseries.

18 The ensemble-mean timeseries (thereafter E-mean, also noted $\tilde{\mu}_k$ in section 3) was then computed from these detrended
19 timeseries, and illustrates the temperature evolution common to all members, i.e. forced by the atmospheric variability. The
20 temporal standard deviation (thereafter Time-STD) of this ensemble mean thus provides an estimate of the atmospherically-
21 forced variability.

22 The dispersion of individual timeseries about the ensemble mean indicates the amount of intrinsic chaotic variability gener-
23 ated by the model. Its time-varying magnitude may be estimated by the ensemble standard deviation (thereafter E-STD, also
24 noted $\tilde{\sigma}_k$ in section 3). Besides these low-order statistical moments, ensemble simulations actually provide an estimate of the
25 full ensemble probability density function distribution (E-PDF) at any time, with an accuracy that increases with the number
26 of members in the ensemble (see also section 3.2).

27 5.2 Initialization and evolution of the ensemble spread

28 Unlike in short-range ensemble forecast exercises, we do not seek here to maximize the growth rate of the initial dispersion;
29 we let the model feed the spread and control its evolution following its physical laws. Panels b. and d. in Figure 4 confirm
30 that the stochastic perturbation strategy (section 4.2) successfully seeds an initial spread between the ensemble members. The
31 evolution and growth rate of the temperature E-STD depend on the geographical location: it grows faster in turbulent areas
32 such as the Gulf Stream (Fig. 4.b) and slower in less-turbulent areas like the subtropical gyre (Figure 4.d). Note that the
1 spread keeps growing after the stochastic parametrization has been switched off at the end of 1993, and tends to reach some
2 leveled/saturated value after a few years. It is nevertheless still subject to clear modulations of its magnitude on time-scales
3 ranging from monthly to interannual. An additional 8-year experiment (not shown here) has confirmed that when the small
4 stochastic perturbation is applied over the whole simulation instead of one year, the overall evolution, magnitude, and spatial
5 patterns of E-STD, and the ensemble mean solution remain unchanged. In other words, the stochastic parametrization seeds

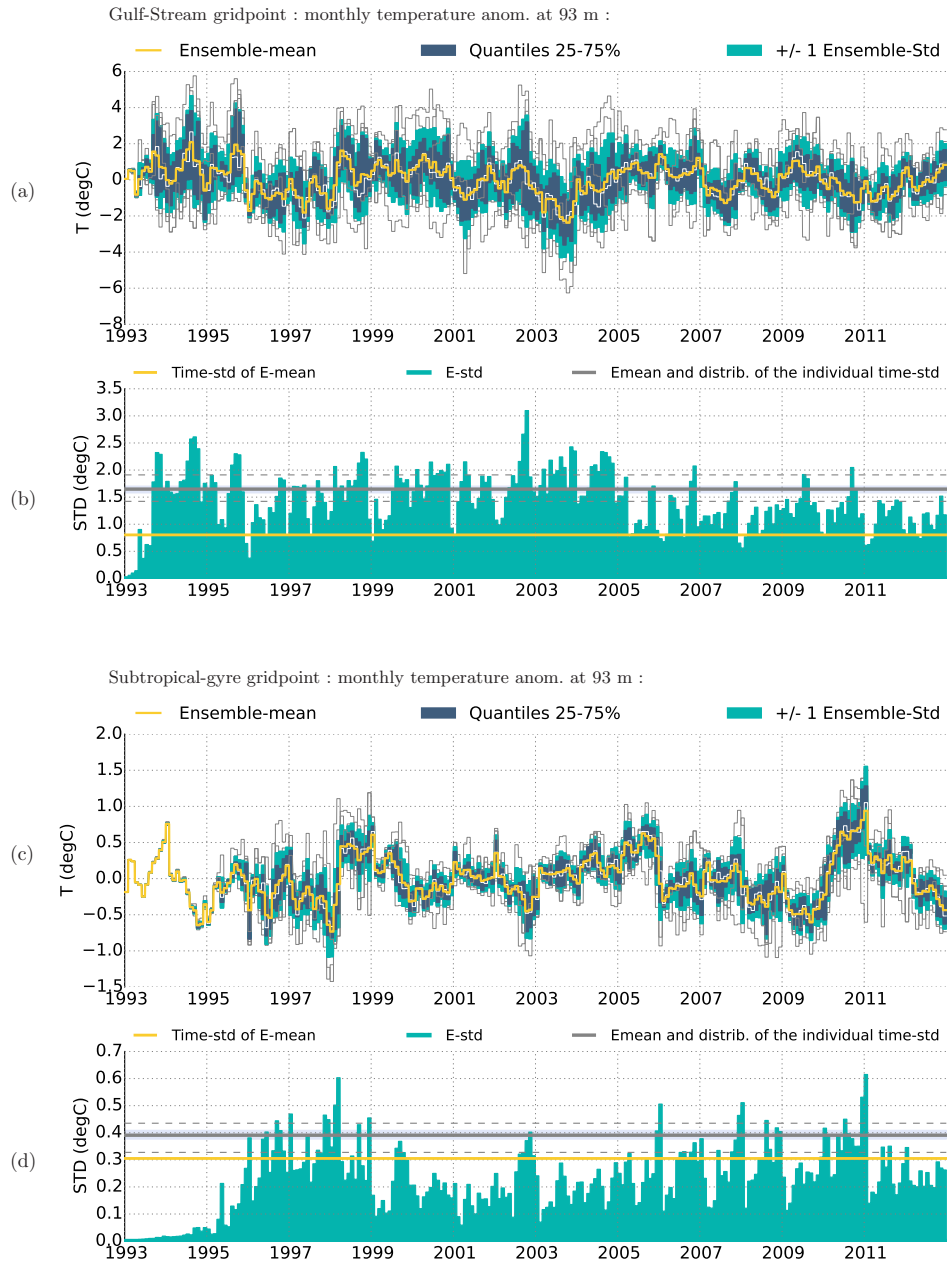


Figure 4. Ensemble statistics of the monthly temperature anomalies from the regional ensemble E-NATL025, at depth 93 m at two grid-points: (a,b) in the Gulf-Stream ($42^{\circ}\text{N};56^{\circ}\text{W}$), and (c,d) in the North-Atlantic subtropical gyre ($22^{\circ}\text{N};42^{\circ}\text{W}$). Anomalies are shown after detrending and seasonal cycle removed (see text for details). (a) The individual trajectories with time of the 10 members appear in thin grey. E-mean is in thick yellow, the interval between quantiles Q1(25%) and Q3 (75%) is filled in dark blue, and the interval E-mean \pm one E-STD is filled in green. (b) E-STD (intrinsic variability, green shading) is compared to the Time-STD of E-mean (forced variability, thick yellow line). Also shown in (b) is the distribution of Time-STD for the 10 members: ensemble mean of the Time-STDs (solid grey), minimum and maximum (dashed grey), and mean \pm one ensemble standard deviation (pale blue shading).

'Median' has been removed from the figure and the caption.

6 the spread during the initialization period, but the subsequent evolution and magnitude of intrinsic variability is subsequently
7 controlled by the model non-linearities regardless of the initial stochastic seeding.

8 5.3 Spatial patterns of the ensemble spread

9 Figure 5 shows maps of E-STD in the regional ensemble E-NATL025, computed from annual-mean anomalies of sea surface
10 height (SSH), sea surface temperature (SST), and temperature at 93 m and 227 m over the last simulated year (i.e. 2012). These
11 maps thus quantify the imprint of interannual intrinsic variability on these variables, and show that after 20 years of simulation,
12 the ensemble spread has cascaded from short (mesoscale-like) periods to long timescales. Annual E-STDs reach their maxima
13 in eddy-active regions like the Gulf Stream (Fig. 5.a) and the North Equatorial Counter Current (Fig. 5.c) where hydrodynamic
14 instabilities are strongest and continuously feed mesoscale activity (i.e. small-scale intrinsic variability), which then cascades
15 to longer time scales. The order of magnitude of this low-frequency intrinsic variability (LFIV) is about 1 °C for SST and 10
16 cm for SSH in the Gulf Stream in 2012. We will compare these amplitudes to those of the atmospherically-forced variability
17 (Time-STD of E-mean) in the next section.

18 Comparing panels b., c. and d. in Figure 5 also illustrates that the ensemble spread of yearly temperature (i.e. its low-
19 frequency intrinsic variability) peaks at subsurface (around the thermocline), and tends to decrease toward the surface in
20 eddy-quiet regions.

21 ~~This is expected from the design of these two simulations where turbulent air-sea fluxes are applied on each member~~
22 ~~independently through bulk formulae, and tend to relax the SST of each member toward the same value set by the air~~
23 ~~temperature. These experiments thus provide a conservative estimate of the upper ocean LFIV and of its imprint on SST.~~
24 ~~This restriction of the SST spread was suppressed in another regional ensemble simulation (not shown); it will be discussed in~~
25 ~~a dedicated publication.~~

26 This is expected from the design of these ensemble simulations: each ensemble member is driven through bulk formulae by the
same atmospheric forcing function, but turbulent air-sea heat fluxes somewhat differ among the ensemble because SSTs do so.
This approach induces an implicit relaxation of SST toward the same equivalent air temperature (*Barnier et al.*, 1995) within
each member, hence an artificial damping of the SST spread. These experiments thus only provide a conservative estimate of
the SST intrinsic variability. Note that alternative forcing strategies may alleviate or remove this damping effect: ensemble
mean air-sea fluxes may be computed online at each time step and applied identically to all members (see section 3.2). This
alternative approach is the subject of ongoing work and will be presented in a dedicated publication.

1 5.4 Magnitudes of forced and intrinsic variability

2 Panels b and d in Figure 4 show how the E-STD evolves at monthly timescale, and how it compares to various Time-STDs
3 (horizontal straight lines). The Time-STD of E-mean (thick solid yellow line) is a proxy for the amount of the forced variability.
4 It turns out to be dominated by the intrinsic variability (E-STD) at the Gulf Stream grid-point. In less turbulent areas like the
5 subtropical gyre, the intrinsic variability is still about 30-50% of the forced part (Fig. 4.d).

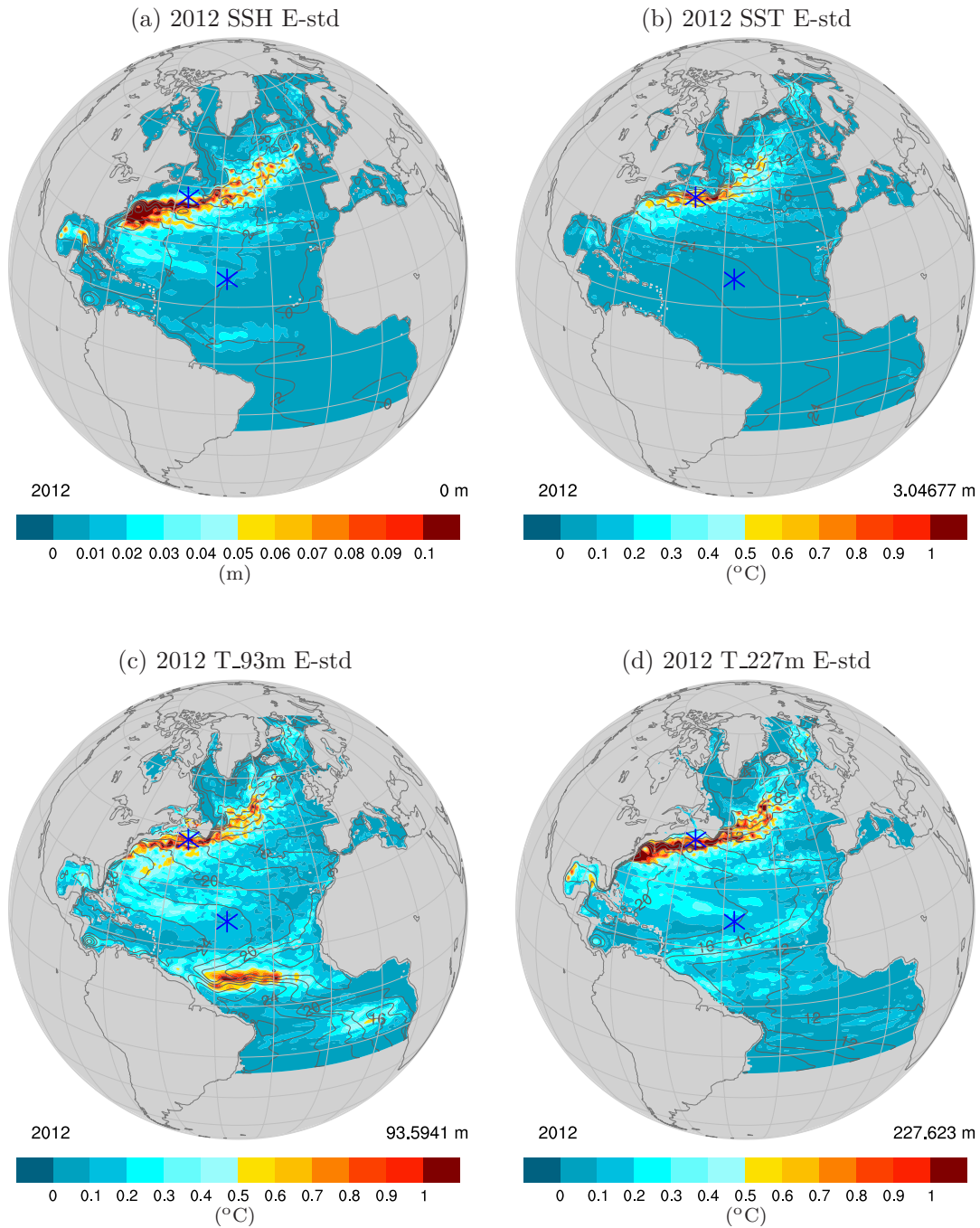


Figure 5. E-STD (shading) for year 2012 of the regional ensemble simulation E-NATL025, computed from annual-means of (a) Sea Surface Height (SSH), (b) Sea Surface Temperature, (c,d) Temperature at depths 93 m and 227 m, respectively. The contours show the corresponding E-mean fields. The blue symbols pinpoint the two grid-points at which timeseries are shown in Figure 4.

6 The E-STD can also be compared to the ensemble distribution of the Time-STDs of the N members (see caption of Fig. 4).
7 By construction, the Time-STD of each member is due to both the forced (shared by all members) and the intrinsic (unique to
8 each member) variability. At the Gulf Stream grid-point (Fig. 4.b), these lines all lie above the Time-STD of E-mean, consistent
9 with a high level of E-STD (i.e. intrinsic variability) contributing significantly to the total variability. At the subtropical gyre
10 grid-point, these lines fall much closer to E-mean since little intrinsic variability contributes to the total variability.

11 5.5 Toward probabilistic climate diagnostics

12 The variability of the Atlantic Meridional Overturning Circulation (AMOC) transport is of major influence on the climate
13 system (e.g. *Buckley and Marshall*, 2016), and is being monitored at 26.5°N since 2004 by the RAPID array (e.g., *Johns et al.*,
14 2008). These observations are shown at monthly and interannual timescales as an orange line in Figure 6, along with their
15 simulated counterpart from E-ORCA025. They were computed in geopotential coordinates as in Zhang (2010) and *Grégorio*
16 *et al.* (2015), and are shown after LOESS detrending and after removing the mean seasonal cycle.

17 The simulated AMOC timeseries are in a good agreement with the observed AMOC variations at both monthly and annual
18 timescales (panels a and c). The total (i.e. combination of forced and intrinsic) AMOC variability is computed as a Time-STD
19 from the observed timeseries and from each ensemble member, and plotted in panels b and d as gray lines. At both time scales,
20 the total AMOC variability simulated by E-ORCA025 lies below the observed variability, consistent with the fact that the
21 model seems to miss a few observed peaks (e.g. 2005, 2009, 2013 on the annual timeseries). Panels b and d also highlight
22 the substantial imprint of chaotic intrinsic variability on this climate-relevant oceanic index at both time scales: at interannual
23 timescale, the AMOC intrinsic variability is weaker than the forced variability, but amounts to about 30% of the latter. A more
24 in-depth investigation of the relative proportion of intrinsic and forced variability in the AMOC and of the variations of the
25 intrinsic contribution with time is currently underway and will be the subject of a dedicated publication.

26 6 Conclusions

27 We have presented in this paper the technical implementation of a new, probabilistic version of the NEMO ocean modelling
28 system. Ensemble simulations with N members running simultaneously in a single NEMO executable are made possible
29 through a double MPI parallelization strategy acting both in the spatial and the ensemble dimensions (Fig. 2), and an optimized
30 dimensioning and implementation of the I/O servers (XIOS) on the computing nodes.

31 The OCCIPUT project was presented here as an example application of these new modelling developments. Its scientific
32 focus is on studying and comparing the intrinsic/chaotic and the atmospherically-forced parts the ocean variability at monthly
1 to multidecadal time scales (e.g. *Penduff et al.*, 2014). For this purpose, we have performed a large ensemble of 50 global
2 ocean/sea-ice hindcasts over the period 1960-2015 at 1/4° resolution, and a reduced-size North Atlantic regional ensemble.
3 These experiments simultaneously simulate the forced and chaotic variabilities, which may then be diagnosed via the ensemble
4 mean and ensemble standard deviation, respectively. The global OCCIPUT ensemble simulation was achieved in a total of 19



Figure 6. Same as Fig.4 but for AMOC anomalies at 26°N in the global ensemble E-ORCA025, from (a,b) monthly- and (c,d) annual-means. In addition, AMOC observational estimates from RAPID at 26°N is shown in orange (see text for details).

5 million CPU hours on the PRACE French Tier-0 Curie supercomputer, supported by a PRACE grant. It produced about 100
6 TB of archived outputs.

7 The members are all driven by the same realistic atmospheric boundary conditions (DFS5.2) through bulk formulae, and
8 represent N equiprobable independent realisations of the same oceanic hindcast. The ensemble experiments performed here
9 have validated our experimental strategy: a stochastic parametrization was activated for one year to trigger the growth of
10 the ensemble spread (see sections 3.3 and 4.2); the subsequent growth and saturation of the spread is then controlled by the
11 model nonlinearities. Our results also confirm that the spread cascades from short and small (mesoscale) scales to large and
12 long scales. The imprint of intrinsic chaotic variability on various indices turns out to be large, including at large spatial
13 and time scales: the AMOC chaotic variability represents about 30% of the atmospherically-forced variability at interannual
14 time scale. These preliminary results illustrate the importance of this low-frequency oceanic chaos, and advocate for the use
15 of such probabilistic modelling approaches for oceanic simulations driven by a realistic time-varying atmospheric forcing.
16 This approach brings in particular new insights on the imprint of this low-frequency chaos on climate-related oceanic indices,
17 and thus helps anticipate the behavior of the next generation of coupled climate models that will incorporate eddying-ocean
18 components. Ongoing investigations focus on these questions and will be the subject of dedicated papers.

19 Our probabilistic NEMO version includes several new features. The generic stochastic parameterization, used here on the
20 equation of state to trigger the growth of the ensemble spread, can be applied to other parameters to simulate model or subgrid-
21 scale uncertainties. The MPI communication between members allows the online computation of ensemble statistics (PDFs,
22 variances, covariances, quantiles, etc) across the ensemble members, which may be saved at any frequency, location and for
23 any variable thanks to the flexible XIOS servers.

24 The size N of the ensemble simulation depends on the objectives of the study, the desired accuracy of ensemble statistics, and
the available computing resources. Our choice $N=50$ allows a good accuracy ($1/\sqrt{50} = \pm 14\%$) for estimating the ensemble
means and standard deviations. Moreover, this choice allows the estimation of ensemble deciles (with 5 members per bin) for
the detection of possibly bimodal or other non-gaussian features of ensemble PDFs; such behaviors were indeed detected in
simplified ensemble experiments (e.g. *Pierini, 2014*) and may appear in ours. Given our preliminary tests with E-NATL025,
 $N=50$ appeared as a satisfactory compromise between our need for a long global $1/4^\circ$ simultaneous integration, our scientific
objectives, PRACE rules (expected allocation and elapse time, jobs' duration, etc), and CURIE's technical features (processor
performances, memory, communication cost). Our tests also indicate that the convergence of ensemble statistics with N de-
pends on the variables, metrics and regions under consideration. For all these reasons, N must be chosen adequately for each
study.

25 More generally, this numerical system computes the temporal evolution of the full PDF of the three-dimensional, multivariate
1 states of the ocean and sea-ice. A very interesting perspective is the online use of the PDF of any state variable or derived
2 quantity (or other statistics such as ensemble means, variances, covariances, skewnesses, etc) for the computation of the next
3 time step during the integration. This would allow for instance distinct treatments of the ensemble mean (forced variability)
4 or the ensemble spread (intrinsic variability) during the integration, e.g. for data assimilation purposes. This NEMO version

5 can therefore solve the oceanic Fokker-Plack equation, which may open new avenues in term of experimental design for
6 operational, climate-related, or process-oriented oceanography

7 **7 Code availability**

8 The ensemble simulations described in this paper have been performed using a probabilistic ocean modelling system based
9 on NEMO 3.5. The model code for NEMO 3.5 is available from the NEMO website (www.nemo-ocean.eu). On registering,
10 individuals can access the code using the open source subversion software (<http://subversion.apache.org>). The revision number
11 of the base NEMO code used for this paper is 4521. The probabilistic ocean modelling system is fully available from the
12 Zenodo website (<https://zenodo.org/record/61611>) with doi:10.5281/zenodo.61611. The authors warn that this provision of
13 sources does not imply warranties and support, they decline any responsibility for problems, errors, or incorrect usage of
14 NEMO. Additional information can be found on NEMO website.

15 The ensemblist features of the model are based on a generic tool implemented in the NEMO parallelization module.

16 The computer code includes one new FORTRAN routine (`mpp_ens_set`, see Algorithm 1) which defines the MPI communi-
17 cators required to perform simultaneous simulations, and to compute online ensemble diagnostics. This routine returns to each
18 NEMO instance: (i) the MPI communicator that it must use to run the model, and (ii) the index of the ensemble member to be
19 run. This index can then be used by NEMO to modify: (i) the input filenames (initial condition, forcing, parameters), (ii) the
20 output filenames (model state, restart file, diagnostics), and (iii) the seed of the random number generator used in the stochastic
21 parameterizations.

22 The online computation of ensemble diagnostics requires additional routines, for instance to compute the ensemble mean or
23 standard deviation of model variables (`mpp_ens_ave_std`, see Algorithm 2). This routine uses the diagnostic communicators
24 defined by `mpp_ens_set` to perform summations over all ensemble members.

25 As can be seen from these routines, this implementation is generic and can be implemented in any kind of model that is
26 already parallelized using a domain decomposition method.

27 *Acknowledgements.* This work is mainly a contribution to the OCCIPUT project, which is supported by the Agence Nationale de la
28 Recherche (ANR) through Contract ANR-13-BS06-0007-01. We acknowledge that the results of this research have been achieved using
29 the PRACE Research Infrastructure resource CURIE based in France at TGCC. The support of the TGCC-CCRT hotline from CEA, France
30 to the technical work is gratefully acknowledged. Some of the computations presented in this study were performed at TGCC under allo-
31 cations granted by GENCI. This work also benefited from many interactions with the DRAKKAR ocean modelling consortium, with the
32 SANGOMA and CHAOCEAN projects. DRAKKAR is the International Coordination Network (GDRI) established between the Centre
1 National de la Recherche Scientifique (CNRS), the National Oceanography Centre in Southampton (NOCS), GEOMAR in Kiel, and IFRE-
2 MER. SANGOMA is funded by the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement 283580.
1 CHAOCEAN is funded by the Centre National d'études Spatiales (CNES) through the Ocean Surface Topography Science Team (OST/ST).
2 The authors are grateful for useful comments from three anonymous reviewers; they also thank the NEMO System Team and Yann Meurdes-

Algorithm 1 mpp_ens_set

Create world MPI group, including all processors allocated to NEMO ensemble simulation (call to MPI_COMM_GROUP)

if (ensemble simulation) **then**

for all (ensemble members $j = 1, \dots, m$) **do**

 Set the list of processors allocated to member j : $r = (j - 1) \times n, \dots, j \times n$

 Create MPI subgroup, including all processors allocated to member j (call to MPI_GROUP_INCL)

 Create MPI communicator, including all processors allocated to member j (call to MPI_COMM_CREATE): $c_{\text{ens}}(j)$

end for

 Get rank of processor in global communicator (call to MPI_COMM_RANK): r

return Index of ensemble member to which it belongs: $j = 1 + r/n$

return MPI communicator to be used for this member: $c_{\text{ens}}(j)$

end if

if (ensemble diagnostic) **then**

for all (subdomains $i = 1, \dots, n$) **do**

 Set the list of processors allocated to subdomain i (across ensemble members): $r = (i - 1) + k \times n, k = 1, \dots, m$

 Create MPI subgroup, including all processors allocated to subdomain i (call to MPI_GROUP_INCL)

 Create MPI communicator, including all processors allocated to subdomain i (call to MPI_COMM_CREATE): $c_{\text{dia}}(i)$

end for

end if

Algorithm 2 mpp_ens_ave_std

Require: Array of model variable: x

Get diagnostic communicator corresponding to this NEMO instance: $c \leftarrow c_{\text{dia}}(i)$

if (ensemble mean) **then**

 Compute sum of x over c (call to MPI_ALLREDUCE, with operation MPI_SUM): s

return Mean: $\mu = s/m$

if (ensemble standard deviation) **then**

 Compute anomaly with respect to the mean: $x' \leftarrow x - \mu$

 Compute squared anomaly: x'^2

 Compute sum of x'^2 over c (call to MPI_ALLREDUCE, with operation MPI_SUM): s

return Standard deviation: $\sigma = \sqrt{\frac{s}{m-1}}$

end if

end if

- 3 oif for interesting discussions about the development of the probabilistic version of NEMO. LB and SL are supported by ANR. JMB, JMM,
- 4 PAB, TP and BB are supported by CNRS. MPM and LT are supported by CERFACS, and GS by CNES and Région Midi-Pyrénées.

5 References

- 6 Barnier, B., L. Siefridt, and P. Marchesiello (1995) Thermal forcing for a global ocean circulation model using a three-year climatology of
ECMWF analyses, *J. Mar. Syst.*, 6, 363–380. doi:10.1016/0924-7963(94)00034-9
- 7 Barnier, B., G. Madec, T. Penduff, J.-M. Molines, A.-M. Treguier, J. Le Sommer, A. Beckmann, A. Biastoch, C. Böning, J. Dengg, C. Derval,
8 E. Durand, S. Gulev, E. Remy, C. Talandier, S. Theetten, M. Maltrud, J. McClean, and B. De Cuevas (2006) Impact of partial steps and
9 momentum advection schemes in a global ocean circulation model at eddy permitting resolution, *Ocean Dynamics*, 56(5), 543–567.
- 10 Bertino, L., G. Evensen, and H. Wackernagel (2003), Sequential data assimilation techniques in oceanography, *International Statistical*
11 *Review*, 71(2), 223–241.
- 12 Brankart, J.-M. (2013), Impact of uncertainties in the horizontal density gradient upon low resolution global ocean modelling, *Ocean mod-*
13 *elling*, 66, 64–76.
- 14 Brankart, J.-M., C.-E. Testut, D. Béal, M. Doron, C. Fontana, M. Meinvielle, P. Brasseur, and J. Verron (2012), Towards an improved
15 description of ocean uncertainties: effect of local anamorphic transformations on spatial correlations, *Ocean Science*, 8(2), 121–142.
- 16 Brankart, J.-M., G. Candille, F. Garnier, C. Calone, A. Melet, P.-A. Bouttier, P. Brasseur, and J. Verron (2015), A generic approach to explicit
17 simulation of uncertainty in the nemo ocean model, *Geoscientific Model Development*, 8(5), 1285–1297.
- 18 Buckley, M. W., and J. Marshall (2016), Observations, inferences, and mechanisms of Atlantic Meridional Overturning Circulation variabil-
19 ity: A review, *Reviews of Geophysics*.
- 20 Candille, G., and O. Talagrand (2005), Evaluation of probabilistic prediction systems for a scalar variable, *Quarterly Journal of the Royal*
21 *Meteorological Society*, 131(609), 2131–2150.
- 22 Candille, G., J.-M. Brankart, and P. Brasseur (2015), Assessment of an ensemble system that assimilates Jason-1/Envisat altimeter data in a
23 probabilistic model of the North Atlantic ocean circulation, *Ocean Science*, 11(3), 425–438.
- 24 Cleveland, W., E. Grosse, and M. Shyu (1992), A package of c and fortran routines for fitting local regression models, *Statistical methods in*
25 *S*.
- 26 Cleveland, W. S., and C. Loader (1996), Smoothing by local regression: Principles and methods, in *Statistical theory and computational*
27 *aspects of smoothing*, pp. 10–49, Springer.
- 28 Combes, V., and E. Lorenzo (2007), Intrinsic and forced interannual variability of the Gulf of Alaska mesoscale circulation, *Progress in*
29 *Oceanography*, 75, 266–286.
- 30 Deser, C., L. Terray, and A. S. Phillips (2016) Forced and internal components of winter air temperature trends over North America during
31 the past 50 years: Mechanisms and implications, *Journal of Climate*, 29, 2237–2258.
- 32 Dussin, R., B. Barnier, L. Brodeau, and J.-M. Molines (2016), The making of Drakkar forcing set DFS5, *DRAKKAR/MyOcean Report*,
33 01-04-16, LGGE, Grenoble, France.
- 34 Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error
35 statistics, *Journal of Geophysical Research: Oceans*, 99(C5), 10,143–10,162.
- 36 Grégorio, S., T. Penduff, G. Sérazin, J. M. Molines, B. Barnier, and J. Hirshi (2015), Intrinsic variability of the Atlantic meridional overturning
37 circulation at interannual-to- multidecadal timescales, *Journal of Physical Oceanography*.
- 1 Haarsma, R., M. Roberts, P. Vidale, C. Senior, A. Bellucci, S. Corti, N. Fučkar, V. Guemas, J. von Hardenberg, W. Hazeleger, et al. (2016),
2 High Resolution Model Intercomparison Project (HighResMIP), *Geoscientific Model Development Discussions*, under review.
- 3 Hirschi, J.-M., A. T. Blaker, B. Sinha, A. Coward, B. De Cuevas, S. Alderson, and G. Madec (2013), Chaotic variability of the meridional
4 overturning circulation on subannual to interannual timescales, *Ocean Science*, 9(5), 805–823.

5 Johns, W., L. Beal, M. Baringer, J. Molina, S. Cunningham, T. Kanzow, and D. Rayner (2008), Variability of shallow and deep western
6 boundary currents off the bahamas during 2004-05: Results from the 26 n rapid-moc array, *Journal of Physical Oceanography*, 38(3),
7 605–623.

8 Kay, J. E., C. Deser, A. Phillips, A. Mai, C. Hannay, G. Strand, J. Arblaster, S. Bates, G. Danabasoglu, J. Edwards, M. Holland, P. Kushner,
9 J. -F. Lamarque, D. Lawrence, K. Lindsay, A. Middleton, E. Munoz, R. Neale, K. Oleson, L. Polvani, and M. Vertenstein (2015), The
10 Community Earth System Model (CESM) Large Ensemble Project: A community resource for studying climate change in the presence of
11 internal climate variability. *Bull. Amer. Met. Soc.*, 96, 1333–1349.

12 Lermusiaux, P. F. (2006), Uncertainty estimation and prediction for interdisciplinary ocean dynamics, *Journal of Computational Physics*,
13 217(1), 176–199.

14 Madec, G. (2012), Nemo ocean engine, *Tech. rep.*, NEMO team.

15 Palmer, T. N. (2006), *Predictability of weather and climate*, Cambridge University Press.

16 Penduff, T., M. Juza, B. Barnier, J. Zika, W. K. Dewar, A.-M. Treguier, J.-M. Molines, and N. Audiffren (2011), Sea level expression of
17 intrinsic and forced ocean variabilities at interannual time scales, *Journal of Climate*, 24(21), 5652–5670.

18 Penduff, T., B. Barnier, L. Terray, L. Bessi eres, G. S erazin, S. Gregorio, J. Brankart, M. Moine, J. Molines, and P. Brasseur (2014), Ensembles
19 of eddy ocean simulations for climate, in *CLIVAR WGOMD Workshop on high*, p. 26, Citeseer.

20 Pierini, S. (2014), Ensemble Simulations and Pullback Attractors of a Periodically Forced Double-Gyre System, in *Journal of Physical
Oceanography*, 44, 3245–3254.

21 S erazin, G., T. Penduff, S. Gr egorio, B. Barnier, J. M. Molines, and L. Terray (2015), Intrinsic variability of sea-level from global 1/12 
22 ocean simulations: spatio-temporal scales, *Journal of Climate*, 28, 4279–4292.

23 Toth, Z., O. Talagrand, G. Candille, and Y. Zhu (2003), *Probability and Ensemble Forecastsg. In Forecast Verification: A Practitioners Guide
557 in Atmospheric Science (pp 137–163)*, ISBN: 0-471-49 759-2, Wiley, UK.

558 Ubelmann, C., J. Verron, J. Brankart, E. Cosme, and P. Brasseur (2009), Impact of data from upcoming altimetric missions on the prediction
559 of the three-dimensional circulation in the tropical atlantic ocean, *Journal of Operational Oceanography*, 2(1), 3–14.

560 Van Leeuwen, P. J. (2009), Particle filtering in geophysical systems, *Monthly Weather Review*, 137(12), 4089–4114.