Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2016-174-AC2, 2016

Geoscientific Model Development Discussions

Interactive comment on "Development of a probabilistic ocean modelling system based on NEMO 3.5: application at eddying resolution" *by* Laurent Bessières et al.

Laurent Bessières et al.

bessieres@cerfacs.fr

Received and published: 9 December 2016

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.

Response to 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 negligeable 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 replace 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 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 ($1/\sqrt{50} = \pm 14\%$) 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° 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.

Other comments on figures

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.

NB: 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.

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-174, 2016.