

Ocean Sci. Discuss., referee comment RC2
<https://doi.org/10.5194/os-2022-11-RC2>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Too much focus on the diagnostics, too little on the perturbation method.

Anonymous Referee #2

Referee comment on "Ensemble quantification of short-term predictability of the ocean dynamics at a kilometric-scale resolution: a Western Mediterranean test case" by Stephanie Leroux et al., Ocean Sci. Discuss., <https://doi.org/10.5194/os-2022-11-RC2>, 2022

Review of Leroux et al. "Ensemble quantification of short-term predictability of the ocean dynamics at kilometric-scale resolution: A Western Mediterranean test-case."

The manuscript presents an analysis of an ensemble of ocean model simulations at very high resolution using a novel idea for intrinsic model errors based on concepts of location errors. The article uses very solid and interesting concepts and methodology and makes both a refreshing and useful contribution to the operational ocean forecasting community (where I belong).

The exploitation part of the research is very well developed and thoroughly explained, which will certainly help popularise probabilistic diagnostics into the oceanographic community, but is so extensive as to almost entirely eclipse the core stochastic model developments, which constitute the novel aspect of the paper. It is indeed seldom that one sees a theoretical advance (that from the papers from Mémin and Chapron) brought into a realistic ocean model, so it is of general interest to see for the first time the effects of the stochastic perturbations on the model solution. However there is no discussion of the numerical effects of these perturbations and no visual from the perturbed model (illustrations are disappointingly always extracted from the CI control simulation without stochastic noise), leaving an uncomfortable impression that something is hidden from the readers. Another aspect that is not discussed is the somewhat binary response of the model to the amplitude of the stochastic noise. The 1% case corresponds to 15m/d displacements (according to my own back-of-the-envelope calculation) and is most often indistinguishable from the CI (0%) case. On the contrary, the 5% perturbations corresponds to 75 m/d, which also seems tiny, turns out completely different from the other two cases and generates kilometres of feature location uncertainties within one single day. What happens between 1 and 5% that causes such a binary response? I believe that tidal amplification is the culprit and suggest an additional experiment in the detailed comments below, where the stochastic noise is turned off in the model nesting zone.

The doubts on the stochastic perturbation method do not impair the main findings of the paper, because the latter probably stand with the CI control ensemble alone, and the diagnostic methods can be applied to any stochastic model, but there is a risk that the manuscript is used to advocate for a stochastic model perturbation method that it does

not truly validate.

Another general remark about the use of the probabilistic diagnostics is that some of them can be generalised to deterministic forecasts under ergodicity assumption: spatially averaged statistics (CPRS, PSD) can be interpreted as expectations and could be applied to forecast systems that have invested in high model resolution rather than in ensembles.

Overall the paper is very good and makes a very enjoyable read. I am impressed by the enormous amount of thoughts and work that went into it. The structure, the style and the illustrations are all excellent, and will certainly make a splash in the operational community. So I recommend its publication after revisions that I would call "major" because of a possible problem in the implementation of the stochastic method. The paper is maybe a little on the long side but I will suggest some reduction of the illustrations and point out a few repetitions in the text. Ideally the manuscript should be split into two separate papers, one demonstrating a new stochastic perturbation method and the other on the ensemble forecast diagnostics, but I will not insist on this if the authors can shed more lights on the stochastic perturbation method without adding pages of text.

Detailed comments:

Title, abstract and introduction: no remark. All are representing well the actual contents of the paper.

Section 2

- Figure 1: Why do you need to define as many as 3 subregions?
- Line 90: I understand that the eNAT60 configuration is not only a boundary condition but a baseline to which the different experiments should revert if there were no stochastic perturbations at all. Please make it explicit and come back to it whenever the different experiments are compared to eNAT60.
- indicate which method is used to impose lateral boundary conditions (the Flather conditions?).
- Line 95-98: a) and c) are not strictly a "difference" and b) should not lead to any difference as long as the model is numerically stable. Please rephrase.
- Line 114-119: This argument is contorted. Any intrinsic or extrinsic errors (in the vertical mixing or winds for example) may as well affect the smallest scales of the ocean, if they are set up to do so. It would clarify the argument if you state upfront that you consider location errors exclusively and that other types of errors can be added at will.
- Line 134: Indicate the physical scales of 1% and 5% with respect to the temporal autocorrelation: displacements of 15 m/d and 75 m/d respectively.
- Line 139: "quite consistent" does not sound too good. Can you recall which conclusion of Mémin (2014) is comforted by the present study?

Section 3

- L. 158: what does CI stand for in ENS-CI? Control Integration?
- Figure 2b indicates that even after Laplacian smoothing, the square model grid is distorted and deviates from orthogonality, which may lead to numerical noise and eventually instabilities. The ROMS user community is advised to keep the grid cells orthogonality above 95% in practice, and especially at the lateral boundaries of the model, to avoid errors propagating inside the model grid. My recommendations would therefore be to dampen the model grid perturbations in the nesting zone of the model (in the first 5 or 10 grid cells) to avoid inconsistencies between the outer and inner model solutions, in particular the barotropic mode. I will come back to this at Figure 4.
- Table 2: Define e_1 and e_2 in relation to the appendix.
- Figure 3 shows indistinguishable lines, and no indication of what is good or bad. You could either plot the difference of PSD from the eNATL60 reference or solely indicate the maximum difference in the text and skip the figure altogether. If you keep the figure, I recommend to remove the part for wavelength > 250km because of the small domain.
- Figure 4 exhibits an oscillatory signal in the ensemble spread, whereas intuitively I expect the spread to grow monotonously. The oscillations are most visible in the 5% case but also in the 1% case. I also noted that the oscillations peak at the same time in the 1% and the 5% cases, about 4 times a day. Unless you have used the same random seed in the 1% and the 5% case - which would be odd - the coherent oscillations indicate an amplified resonance of tidal signals, which brings me back to my previous remark about barotropic lateral boundary conditions: the nesting routines (radiation condition or Flather conditions, whichever you use) should allow tidal and other barotropic signals to be evacuated out of the domain, but if the perturbations make this boundary condition imprecise, the tides may be reflected at the lateral model boundary and resonate inside the nested model domain. I have a suspicion that this could be avoided if the perturbations were attenuated near the model boundaries (and maybe in shallow waters as well).
- Line 209: This claim could be confirmed by a look at the accuracy numbers from the MED MFC QuID document on the Copernicus Marine website.
- Figure 5 makes a stunning impression, but is uninformative. I would have preferred to see the 5% case to have a visual impression of the effect of random perturbations (there are otherwise none in the whole paper).

Section 4

- L. 268-280 is a nice introduction of the ensemble diagnostics, but seem like a methodological overkill: the diagnostics are initially intended for location-dependent comparison to observations, but in the absence of observations like in the present study, some more basic diagnostics may be simpler to use than a cross-validation with each ensemble member. This is the case for the CRPS which is aggregated spatially for all members to a single number and does not seem to add more information than a standard deviation. Please replace by the ensemble spread if this is a simpler diagnostics that provides the same insights.
- L298-299 are repeated in the figure caption.
- Figure 8. It would seem fair to mention that beyond 5 days of lead time, the 95% percentile is dependent on the model trajectory and does not make a robust statistic, a larger ensemble or a different perturbation method may improve that.
- The small lines in Figure 10 are not very informative. The three figures could be compressed into one by showing the three 95% quantile only and plotting the differences from the initial CRPS.
- Section 4.2.1: I guess there are technical difficulties with the location score in the presence of islands or complex coastlines. This could be mentioned.
- Figure 11 (top against bottom) is nearly showing the same thing. You could remove the two lowermost panels by adding the 20 isolines in the top panels.

- L. 433: Why choose SSH this time?
- L. 460: scales above 150 km should be removed from the figure.
- L. 461: I would suspect that checkerboarding (numerical noise) would easily cause the correlation of small scales. Numerical noise is ubiquitous in all ocean models although viscosity makes it almost invisible. If the authors use a high-contrast colour scale (like "details" in Ncview), they would probably see some checkerboarding in the model output, which would inevitably appear coherent at the smallest wavelengths of the model output.
- Figure 18: Add the diagonal line for $T=0$.
- L. 485: The authors could indicate which SWOT revisit time would be necessary to maintain the small-scale structures (if the data assimilation were ideally good).

Appendix A1:

- L. 553: "Anamorphic transformation" is a pleonasm.
- L. 582: the link between the theoretical papers from Mémin and Chapron and this one is not obvious. How does the sigma value translate into the stochastic process P?

Appendix A2

- L. 595 to 599: "can be thought", "can be viewed" and "can be argued" make a very embarrassed logical chain to line 600, which I would promote upfront to motivate the approach.
- L. 610: Mention that $a^2 + b^2 = 1$ to maintain the variance constant.
- L. 611: The "assumed independence" of the perturbation is later contradicted by the Laplacian filter in Line 620.
- L. 618: the citation to Garnier et al. (2016) is repeated.
- L. 620: does the Laplacian filter maintain the standard deviation?
- L. 620: is the value of sigma linked to the sigma in Mémin/Chapron?
- L. 629: Transformed to the other grids: do you mean a linear interpolation?
- L. 632: Only here is it possible for the reader to calculate the typical scale of the perturbations (about 15 m/day for 1%). This information is important to realise how much the model amplifies the location noise into location errors (roughly by a factor of 100 to 1000 in a single day, which is mind-boggling) and should be discussed in the main text.

Typos:

-
- l. 133: remove the second "that".
 - L. 239: "characterizing"
 - L. 343 Fussy -> Fuzzy
 - Section 4.3.1: "pf" -> "of"
 - L. 531: Beying -> Beyond