

“Use of Genetic Algorithms for Ocean Model Parameter Optimisation”

By Marcus Falls, Raffaele Bernardello, Miguel Castrillo, Mario Acosta, Joan Lloret, Martí Galí

In this paper the authors propose a Biased Random Key Genetic Algorithm (BRKGA) for the estimation of parameters of the Earth system models that ensure the optimal model performance. The method is tested using the one dimensional configuration of PISCES-v2, the biogeochemical component of NEMO, a global ocean model. In particular, a test case of particulate organic carbon is examined. First, the optimisation is done against the output of the default model. Second, the subset of selected parameters is optimised against the POC estimates from BGC-Argo floats. They find that the algorithm is faster and more computationally efficient than a brute-force or random approach to tune model parameters.

The work is very important and the approach is novel. The paper is well structured and well written. I still have some questions, answering which needs minor revision of the manuscript, I guess.

Major comments

1. In lines 322-323 you say “The results of experiment D9, plus additional analyses that we report in the Supplemental Information (SI), provided the criteria to select the 5 parameters that were used in subsequent PO experiments.” I did not find Supplemental Information (SI). Therefore, I could not understand why *wsbio2* or *wsbio2max* are not selected. *wchldm* in experiment a27e converges to a much higher value than in the other experiments, thus exhibiting large spread. In the biogeochemical point of view, I can understand selection of *wchldm*. A parameter that controls sinking of bPOC is relevant and should be selected, also. In figure 14, it can be seen that bPOC has positive default and optimised bias, which could show that bPOC is not removed from the system rapidly enough. Parameter(s) that control sinking of bPOC are not optimised and the other processes do not compensate this. Please, see also comment on Figure 3 below. Therefore, I would like to see more clear justification why *wsbio2* and/or *wsbio2max* are not selected for PO.

2. My major concern is difference of optimised parameter's values in D5 and O5_LAB1 experiments. In case of O5_LAB1 experiment *wsbio* was 2-times smaller, *xremip* 3-times larger and *grazflux* almost two orders of magnitude smaller than corresponding parameters in case of D5. *Wchld* and *wchldm* had comparable values. Both of the experiments are compared at the location LAB1 (Table 3). D5 is compared to model simulation with default parameter set and O5_LAB1 with observations. Authors say that model simulation with default parameter set compared well with observations. In my understanding D5 and O5_LAB1 point to the dominant role of different processes in the same POC system, although D5 and O5_LAB1 both provide relatively good results. In D5, sPOC sinking is relatively fast and the sinks of sPOC consist of the GOC fragmentation upon mesozooplankton flux feeding (controlled by *grazflux*), mesozooplankton flux feeding on sinking POC (controlled by *grazflux*) and POC degradation (controlled by *xremip*). In O5_LAB1, sPOC sinking is two times slower and POC degradation (controlled by *xremip*) is much higher. The role of GOC fragmentation upon mesozooplankton flux feeding (controlled by *grazflux*) and mesozooplankton flux feeding on sinking POC (controlled by *grazflux*) is negligible. Concerning bPOC, the same tendency is true, except that sinking speed parameter of bPOC had default value in both cases. For instance, for me, it is difficult to decide either to use parameters' values from D5 or O5_LAB1 in the model simulation.

This rises for me a question how robust and consistent the BRKGA approach is? Is it justified to use a subset of parameters for optimisation or how the subset should be selected? Some more discussion about these issues should be provided. I acknowledge the authors discussion about the relevance of different biogeochemical processes in case of the O5 experiments.

Also, I would like the authors to mention the differences in the values of optimised parameter sets in case of O5_LAB1 in comparison to D5 in section 3.2.1.

3. L485-497 According to my opinion the inclusion of the Tasman Sea case is not natural part of this paper. Therefore, I suggest to remove this part and the figures S6-S8. If you include this part, then readers might like to have more detailed description of the experiment, etc.

Minor comments

L88: I would suggest to use term “a set of parameters” instead of “an ideal set of parameters”.

L93-94: The authors claim “Finally, we discuss how our approach can become the first step towards assimilating new kinds of observations into existing Earth system models.” I did not find such discussion. I suggest to remove this sentence. There is enough material in the paper even without discussion on data assimilation.

L144-145: “...here we focus on 9 parameters expected to strongly influence mesopelagic POC dynamics (table 2).” Expected by whom or why? Could you provide reference to the choice of 9 parameters or formulate it better? Is it how POC dynamics is formulated in PISCES-v2? In the next sentence you list the processes that these parameters control, and I fully agree with your choice. For readers who are not familiar with PISCES-v2, it is rather time consuming to go through mathematical formulation of the processes by Aumont et al (2015). Maybe reference to Aumont et al (2015) is sufficient.

L170: Instead “three-fold” should be “two-fold”, parameter *wchldm*.

L230: I guess S_3 should be instead of S_T .

L 265: There is no reference to figure 2 in the text.

L318: Should be *wsbio2max*

Figure 3: In the experiment a27m, *wsbio2* is larger than *wsbio2max*. I seem like these two parameters “have changed their values”. Also, *wchldm* in experiment a27e converges to a much higher value than in the other experiments. Could the authors have comments or discuss these cases in Discussion part? What do these cases tell about BRKGA? *caco3r* behaves differently. It shows large spread, but the end values are distributed more evenly between min and max.

L322: I did not find Supplemental Information (SI). The figures that are referred as S1, S2, etc. can be found in Appendixes, but no text on additional analyses.

L329: Should be figures A1 and A2; and B1-B3.

L341: Please specify which 3 parameters.

L357-368: Is this analysis necessary? “Both the D5 and D5_rand experiments reached the breakpoints with a similar speed, in 5–15 generations (average of 8).” “This analysis illustrates the greater efficiency of the BRKGA compared to the RS.” I guess that authors mean computational efficiency. So, I would not say that the BRKGA is more efficient than the RS. No doubt that the RS outperforms the BRKGA in terms of the optimised parameter set. “Breakpoint detection can be used to stop GA

experiments in an adaptive manner, thus saving computation.” Could the authors be more specific? Stopping the computation while breakpoint is reached (or after n-number of iterations) does not seem to be good idea. This paragraph, i.e. calculation of the breakpoints, does not provide additional information. Convergence of the cost function and calculated statistics can be seen in Figures and each reader could decide how many iterations are feasible and sufficient.

L371-389: I would suggest to remove the analysis related to the breakpoints. I agree that parameter values at the breakpoints are very close to their final values. But this is not the case in D5 experiments. If I would use the BRKGA, I would not stop calculations at the breakpoint or even not close to it.

L383-384: Could you provide the values of mean absolute biases (default and optimised; sPOC and bPOC) additionally to the plots in figures 13 and 14?

L412-414: See my comment L357-368. I do not see that GA produces parameter sets quicker than the RS, i.e. breakpoints were reached at the same number of iterations.

Table 9: Caption is wrong

L421: Word “figure” is missing

L451: Should reference be “table 11” instead of “figure 11”?

L467: Should be “figure”

L482-483: “...an increase in *grazflux* could also improve model skill, as shown in fig. 10 and 11.” How can I see that in figures 10 and 11?