Interactive comment on "Calibrating a global three-dimensional biogeochemical ocean model (MOPS-1.0)" *by* Iris Kriest et al.

Anonymous Referee #1

Received and published: 4 October 2016

These authors demonstrate the use of an evolutionary algorithm to optimize parameters for a marine ecosystem model to observed distributions of nitrate, phosphate and oxygen using offline advection from stored circulation fields. This is a worthwhile experiment, and I congratulate the authors on a (mostly) well written paper.

My most important objection to the present text is that the description of the methodology is somewhat difficult to follow. The authors have tried to present some very mathematical material in accessible terms, and I applaud them for making the effort. But I am not convinced that they have succeeded. This is something that has plagued this field from the beginning. I reviewed some of the very first ocean biogeochemical model parameter-estimation papers, at the very beginning of my reviewing career. I distinctly remember advising authors to try a bit less hard to use 'lay' language, because they weren't really succeeding at making the material accessible or understandable.

Try this: write out the method as if you were writing it for an audience that understands all of the mathematics, using both words and equations and fully integrating the mathematics into the prose, but avoiding jargon as much as possible. Then go through it and look for points where the language could be made more accessible, without leaving the intended meaning vague and without changing the overall structure (much).

I think this exercise will produce a more understandable description of the method, at least understandable by what I would consider the target audience for this paper, i.e., oceanographers like myself who have a basic understanding of, but not a deep familiarity with, the mathematics employed here. To put it bluntly: biological and chemical oceanographers whose comfort level with mathematics does not extend much beyond $Y=a+bX$ are not going to read this paper, or at least they aren't going to read the Methods in any depth. The important thing is to document exactly what was done so that others can reproduce it.

The weighting scheme expressed by w_i could be better explained. From Table 1, Section 2.2.4 and the algorithm schematic (p. 9) we can conclude the following: (a) The w's do not change over the course of the optimization; (b) there is some sort of a priori ranking that allows these weights to be defined us a function of the index i; and (c) the basis for the latter is not explained. Section 2.2.4 states that samples should get more reliable for larger lambda, via a regression-to-the-mean argument. But this logic does not really tell the reader why $w_1>w_2>w_3$, when in fact everything presented here is, or could be, for a single value of lambda. As the text is currently written, the rolling a dice (sic) analogy is fatuous. Obviously the sample mean will on average be closer to the population mean for larger samples. But the present text seems to imply that a sample of e.g., n=5 will be more reliable than a sample of the same size if you draw more of them. I don't doubt that the methodology is valid, but the present description is confusing and incomplete.

The captions to Figures 1 and 2 are not very informative. Figure 1 caption does not explain the meanings of the symbols that appear within the shapes or of the shapes

themselves (circles vs rectangles). Figure 2 does not have axis labels. It is difficult to guess what is meant by "fitness values are shown as dots" when these dots fall directly on the function they are sampled from. Is fitness the x or the y axis? I have a hard time envisioning it as either.

Minor points:

(1) The Conclusion is unfocused, meanders among topics, and repeats points already stated in the Discussion. I think it could be cut to about half its current length, if it were clearly focused on what are the key take-home messages of this work.

(2) The sinking model could be better explained, given its significance to the main points of this work. I understand that it is fully explicated in the previous publication cited, but one or two sentences spelling out exactly what are the assumptions and functional relationships used will make it easier on the reader. I would also advise to state what the primary currency of the model is (in section 2.1.2), i.e., are the biological compartments denominated in N or P units (see Figure 9).

(3) The experiment codes in Section 2.6 should be explained, and not just in the sense that the abbreviations are defined. This paragraph should be expanded to include an explanation of what the purposes of the different experiments are, in a conceptual sense. For example, WIDE appears to indicate broad limits on what values the parameters can take in the optimization (vs narrow a priori imposed limits). The reader will eventually figure this out, but it is good practice to clearly state it up front.

(4) Section 3.1 emphasizes the reasons for slow convergence of K_PO4, but glosses over the fact that convergence of the zooplankton growth and death rates is not much faster (Figure 3). It is clear that all 3 of these parameters are quite strongly correlated with each other (Figure 5), so the slow convergence is not very surprising, as the misfit function surface will be more or less flat over a large area of the parameter space. This correlation is also apparent in the subsequent sections (e.g., 13/11-12, 15/3, Table 3).

(5) Figure 6 is not adequately explained. The caption simply states that what is shown here is "a region ±2% around the average parameter value of the last generation" while the text states it shows a region of parameter space "close to the optimum". The most plausible interpretation I can think of here is that "the last generation" represents the one prior to convergence having been declared and the optimization terminated. But this could be spelled out more clearly in the caption.

(6) Figure 16 could use some summary statistics. In some cases it looks as if the optimized parameters are worse than the reference case, but there is a lot of regional variation. It would be good if the global integrals of the misfit function were stated in each case. It would also be a very good idea to include some statement of what defines the 'best' individual.

(7) I don't think Figure 17 is necessary. If this material is really necessary I think it would be better to format it as text, similar to the algorithm schematic on p. 9.

Details:

1/14 change "model's" to "model"

2/6-8 Move Orr ref to the end in same parenthesis as Najjar. Current wording confuses OCMIP1 and OCMIP2, i.e., refers to protocols for OCMIP1 and then cites results from OCMIP2.

2/9 delete "global"

2/19 delete "rather sluggish"

2/26 change "insufficient" to "inappropriate"

2/26-28 "The establishment of an automatic optimization of global biogeochemical ocean models is aimed for in this current study that should enable ..." The development of automatic optimization of global ocean biogeochemical models that is the goal of this study should enable...

3/1 change "environments" to "resolution"

3/16-18 "This efficient "offline" method for ocean passive tracer transport represents the advective and diffusive components of an ocean circulation model in form of transport matrices, that have been extracted prior to the biogeochemical simulations performed here from a physical global circulation model." This efficient "offline" method for ocean passive tracer transport represents advection and mixing in the form of transport matrices that have been calculated from an ocean circulation model simulation prior to the biogeochemical simulations performed here.

3/20-23 I don't think the "see also" or the multiple references to the same paper within the same sentence are necessary.

3/26 MOPS should be defined at first use

4/2-4 "Both aerobic and anaerobic remineralization are parameterized as a saturation curve, using half-saturation constants to regulate the affinity of these processes to either oxidant, as well as the inhibition of denitrification through oxygen." Aerobic and anaerobic remineralization are parameterized as saturation (Monod-type) curves that regulate the rates of these processes using either oxidant, as well as the inhibition of denitrification by oxygen.

4/4 delete "accomplished" or change it to "actual"

4/12 via a parameterization of river runoff? I doubt that this model has explicit river inputs.

4/17 and elsewhere CMAES is sometimes hyphenated, sometimes not

4/29 change "opposite" to "contrast"

4/33 "searchspace" should be "search space" ("eigenvalue", "eigenvector", and "univariate", by contrast, are actual words (see 6/4-9))

5/2-3 "QiEA versions for continuous problems have also been investigated in the liter-

ature." Could use a literature reference

5/13 "therefor" (this misspelling appears repeatedly throughout the text)

5/12 "pseudo code" I assume this refers to the algorithm outline, which is useful, but I don't think this term is appropriate here.

5/17-18 "Gaussian bell" I don't think this term is useful or necessary. A Gaussian distribution is sometimes colloquially referred to as "bell curve", but the term is not normally used in the scientific literature. You have defined the distribution as Gaussian, so most subsequent references to "the bell" could just refer to the "the distribution". You might have to finesse the wording in a few places, but I would prefer if this term were not used. "the mean of the bell is attracted towards the good samples" is a good example of the kind of writing I critiqued in my general comments: it tries too hard to be accessible and ends up just being vague.

6/22 I think I understand what sort of vector multiplication is implied here but I'm not sure the terminology is correct (see http://mathworld.wolfram.com/VectorDirectProduct.html). If you multiply x*y' in Matlab for example, it represents a scalar product, which is clearly not what is meant here (see also algorithm outline on p. 9, 3rd to last line of while loop).

6/23 change descend to descent (this misspelling appears several times, in the text and Figure 2 caption)

7/28-29 "the minimum of the penalized fitness function lies within the feasible box" Shouldn't this penalty function be 0 for points inside the boundaries?

11/12 "different random selection of the parameters from the distribution" A different random selection of parameter values; the parameters sampled are fixed.

9/6 Why not state what the "termination criterion" is? (see also "stopping criterion" in algorithm outline above, 11/17, 13/4)

11/31 delete "and large ocean volumes"

12/6 "do not decrease monotonously" monotonically

12/9 delete "obviously"

13/26 Change "a phenomenon that does not occur in the real ocean" to something like "a statistically optimal but physically meaningless solution"?

14/6 "a closer fit to biogeochemical fluxes" based on what? There are no fluxes in the misfit function. Perhaps Table 4 provides support for this assertion but it is not cited.

14/15 "organic tracer concentrations" I think this refers to biological tracers like zoo-plankton, as opposed to "inorganic" tracers like nitrate (15/12-14). These are some-times referred to as "abiotic" tracers (15/1). I would suggest just referring to "tracers" generically and "biological" tracers where appropriate, perhaps with "(e.g., phytoplank-ton)" at first occurrence for clarification. Choose your wording but I strongly recommend that "abiotic" not be used.

14/27 "for some parameters it is quite insensitive to changes" changes in what?

15/15 "not improved on cost of any other tracer" not improved at the cost of any other tracer (see also 16/21)

16/30 change "resembles" to "represents"

17/11 "Another possibility to avoid undesired effects like nearly extinct zooplankton is to bring in further objectives which consider that issues." Another possible way to avoid undesired effects like nearly extinct zooplankton is to introduce further criteria that take account of this issue.

17/12 "the cost function" This term appears out of the blue and is not defined until much later. I don't care if you say cost function or misfit function but be consistent.

17/19 "The topic of multi-objective optimization is intensively regarded" I can't tell what

this means.

17/25 "It remains to be investigated, whether this is related to the lack of temporal solution, or to phosphate not being too tightly related to dissolved or particular organic matter." It remains to be investigated whether this is related to the lack of seasonal data, or to phosphate concentration being weakly dependent on dissolved or particular organic matter concentration.

18/5 change "cure for" to "solution to"

20/12-14 "However, it is also related to the biogeochemical model structure itself, as the mapping of simulated to observed tracers and diagnostics can depend strongly on the biogeochemical model structure." If one is looking for opportunities to shorten the text this would seem to be a good place to start.

20/20 add "in" before "the appendix"

20/21 "refer the reader to that website" Doesn't this refer to a published paper? If it doesn't then we need a lot more detail, because the reader is referred to KO15 for all of the details of the biogeochemical model.

20/28 delete "vectors of"

23/17 Something is wrong here. Why is "reprint of" necessary?

24/12 why is a Discussion paper from 2014 cited? Was the final paper not accepted? (see also Seferian et al)

26/20 Srokosz misspelled

In Table 2 the term used to define the upper boundary differs between the caption (and the footnote) and the table headers.

In Table 3 caption change "brackets" to "parentheses" and delete first comma.

In Table 4 the depth for export is given as 120 m in the caption and 130 in the column

header.

In Figure 2 caption change "then" to "than" and "standard derivation" to "standard deviation"

In Figure 4 caption specify log10 or ln