Biogeosciences Discuss., 9, C4676–C4685, 2012
www.biogeosciences-discuss.net/9/C4676/2012/

Biogeosciences Discussions

Interactive comment on "Reducing the model-data misfit in a marine ecosystem model using periodic parameters and Linear Quadratic Optimal Control" *by* M. El Jarbi et al.

E. J. Mustapha

mej@informatik.uni-kiel.de

Received and published: 13 October 2012

[11pt,a4paper]article amsfonts amsmath graphicx wrapfig graphicx amsmath graphicx epstopdf pdfpages tabularx

algorithmic longtable lscape amsthm caption epstopdf multirow multicol url tabularx font=small,labelfont=bf,singlelinecheck=false float subfig color ruled plain

equationsection

Comments to referee 2: We thank you for your comments, and are sorry for the inconsistencies. We give answers here, referring to your numbering.

Answers to general comments:

(1) To clarify the sections 2 and 3, we have made the following changes in the mathematical notation:

- The model variable/ state variable/ tracer vector – as defined in eq (1) – is denoted by $\mathbf{y} := (y^l)_{l=N,P,Z,D}$ instead of $\mathbf{x} := (x^l)_{l=1,2,3,4}$.
- The observational data is denoted by $\mathbf{y}^{obs} := (y^{l,obs})_{l=N,P,Z,D,PP}$ instead of $\mathbf{y}^{obs} := (y^{obs}_m)_{m=1,...,5}$.
- The aggregated values of the model output (that is used in the original cost function F) is denoted by $\mathbf{y}^{mod} := (y^{l,mod})_{l=N,P,Z,D,PP}$ instead of $\mathbf{y}^{mod} := (y^{mod}_m)_{m=1,...,5}$.
- The reference trajectory used in the linearization is denoted by $\mathbf{y}^{ref} := (y^{l,ref})_{l=N,P,Z,D}$ instead of $\mathbf{x}^{ref} := (x^{l,ref})_{l=1,2,3,4}$.
- The biogeochemical coupling terms are denoted by $\mathbf{q} = (q^l)_{l=N,P,Z,D}$ instead of $\mathbf{q} = (q^l)_{l=1,2,3,4}$.
- The linearized state in section 3 is denoted by $\mathbf{x}_k$ instead of $\mathbf{z}_k$.

The notation used for the linearization in section 3 has been corrected and made clearer. See the detailed answers below.

(2) The model fit is much better at the surface than in deeper layers. We believe that the reason for this is the following.

- The non-linear model must be linearized using a reference trajectory. The reference trajectory must have values in every spatial layer. Thus we have to interpolate the spatially sparse observational data.

- Since there are less data available in deeper layers, the interpolation error there is bigger than in upper layers where the database is denser. This will affect the quality of the linearized model and consequently also the quality of the optimal parameters.

These points will be mentioned in the final version. We will also add a figure of the model output compared to the data for one or more deeper layers in the final version. Due to the lack of data in deeper layers, a comparison of the cost function per layer is hard to compute.

(3) This is a good point, we will include such kind of example in the final version.

(4) We will discuss this point and give some possible reasons in the final version. Possible reasons for the good quality of the fit using this method here are

- the use of observational data for the reference trajectory in the linearization
- the application of the method separately on each time interval between two succeeding observation points (in time)
- the usage of a very short temporal discretization on each of these intervals
- the usage of exact derivative information for the system matrices in linearization (using Automatic Differentiation, a point we did not mention in the first version).

Still the result in this application is impressive, and might not be obtained in other examples.

Answers to specific comments:

(We will repeat the comments here, followed by our answers thus marked by *Answer:*)

Abstract: Include "one-dimensional" in the model description. Mention upfront that time-varying parameters are introduced and also mention how many of the parameters are optimized.
Answer: Will be included.

p10208, l 14: Please divide this large paragraph up, it is hard to read.
Answer: We will divide the introduction to the following paragraphs:

- Description of the marine ecosystem models and motivation for parameter estimation/ optimization.
- Different methods of parameter optimization used so far in the literature, and brief discussions of results.
- Goals of this paper.
- Structure of this paper.

p10209, l 5: Again, add "one-dimensional", e.g. "use one-dimensional models which simulate a single water column only".
Answer: Will be added.

p10209, l 11: Define "NPZD" and other abbreviations right away.
Answer: Will be added.

p10209, l 17: "This reflects the aim to obtain a model that is applicable for arbitrary time intervals." There are other practical reasons why parameters are assumed to be constant, it is easier to obtain them and the risk of over fitting the data is much lower. These points could be mentioned here.
Answer: A remark on this will be included.

p10210, l 11: This sounds a bit misleading, the goal of "improved model parametrizations that can model these variations even with constant model parameters" is not achieved in this study, please rephrase.
Answer: We will omit this sentence, since – as you noted correctly – this was not achieved here.

p10210, l 13: What is "optimal control", the term control is only defined later.
Answer: We will give a short definition of the term ”control” and a reference, but will later on avoid it.

p 10210, l 15: "This avoids the process of parametrization in the sense that we do not have to know or assume how the above mentioned periodic functions look like." Sentence is hard to understand.
Answer: Will be changed to:

> As a consequence, we do not introduce a prescribed periodic parametrization of the unknown quantities as growth rates etc.

p 10210, l 17: Point out the two aims more explicitly in the text above.
Answer: We will include this in the third paragraph of the introduction.

p 10211, l 5: Break up this long sentence.
Answer: Will be changed.

p 10211, l 25: Rephrase, $\mathbf{u}$ is a vector containing the biological model parameters.
Answer: Will be changed.

p 10212, l 2: What is q_n?
Answer: Has to be $\mathbf{q}$ which denotes the vector of the biogeochemical coupling terms, will be changed.

p 10212, eq (2): So $x_1 = P$. It would be more consistent to stick to one notation, e.g., x_P. The x_l notation could still be used by writing $l = N, P, Z, D$. It would also avoid confusion with z (depth) and$Z$ (zooplankton).
Answer: Will be changed, we will replace N, P, Z, D in the model equations (2) by y^P etc. The sinking term for detritus will be omitted. It is already contained in equation (1).

p 10212, eq (2): The sinking term for detritus is included here again.
Answer: Will be changed, see also last comment.

10213, l 9: Now ZOO is introduced in addition to Z and x_3, please stick to one notation.
Answer: Will be changed accordingly.

p 10213, l 11: Make it more obvious that $x_3 (= x_Z)$ is dependent on z(depth).
Answer: This will be done. We will rewrite the description of the aggregated data to fit to the new notation (see general comment (1) above).

p 10213, l 17: Why is J dependent on μ and u which were not defined previously?
Answer: Will be made clear. Additional symbols μ, u will be suppressed in the notation, all dependencies will be mentioned in the text when omitted in notation.

p 10214, l 4: "Note that these numbers may be quite different for the different years". Why are the F_{mj} in eq 4 then normalized by N_{mj}? This approach appears to emphasize observations in those years with little data, is there a reason for it?
Answer: Actually F_{mj} is the averaged annual misfit per model output/tracer, weighted using the vector σ of measurement uncertainties, and by the number $N_{l,j}$ of observational data per tracer and year. We will modify the Section 2.3 accordingly to emphasize this point.

p 10214, l 9: What motivates this choice measurement uncertainties?
Answer: We will add a reference for the actual values used here.

p 10215: The use of "controls" and "control variables" is confusing. Maybe mention once that the control variables are the parameters, then just stick to the term parameters.
Answer: Will be changed to ”parameters” throughout the paper.

p 10216, l 9: Is it a "time-step of length τ"?
Answer: τ meant time-step here, will be changed to Δt.

p 10216, eq 6: Is it an l or an I and what does it mean? Do the n signify time coordinates? Please introduce each new symbol. Without this information, the rest of the description remains guesswork.
Answer: Here I is the identity matrix, will be noted right away. Time index will be k now, according also to the linearization process later on.

p 10216, l 23: Mention the "4x4" and other properties right away in line 11.
Answer: Will be done.

p 10217, l 15: Now commas are used to separate the indices in the subscript, please use this throughout the paper.
Answer: Will be changed.

p 10217, l 16: The first 3 bullet points are not consistent: if there are N_j observations, the steps cannot go from 0 to $N_j + 1$.

Answer: The first and the last points were added for technical reasons in the linearization process. They are the beginning and the end of the corresponding year, respectively. There have been some unclear notations in this description of the linearization process. It will be rewritten in the final version.

p 10217, l 21: Why is x suddenly used for observational data, or what is x^{ref}?
Answer: Will be changed, see also answer to general comment (1).

p 10217, eq 9: Is n suddenly defined as a real number, or is $I_{j,i}$ (the interval) really $n_{j,i}, n_{j,i} + 1, ..., n_{j,i+1}$.
Answer: No, n was meant to be integer still. The interval notation is sometimes generalized for a discrete set of integers, but this can definitely be misleading here. It will be changed.

p 10217, eq 9: z_k is a poor choice notation-wise, as $z$was depth previously and Z stands for zooplankton.
Answer: Will be changed to $\mathbf{x}_k$, see above.

p 10217: The notation is confusing: there is n which runs to M but there is also an $n_{j,i}$ while i runs to N_j (previously there was an N_{mj}). I like the fact that j runs to j_{max}, may be this notation can be extended.
Answer: We will follow this idea in the final version where possible.

p 10219, l 4: How does the constraint enforce periodicity, when u_k^{ref} is not periodic (according to eq 11).
Answer: In the first year the choice of the cost function will indeed somehow force ”periodicity” to the constant reference parameters. This effect can be reduced by choosing appropriate small values in the matrices R_k in the first year. In the following years the

difference of the current parameter (vector) u_k to its counterpart from the year before is minimized. This in forces periodicity. The crucial point in adjusting the matrices R_k throughout an optimization run is to both allow

- for sufficiently large deviation from the constant reference parameters in the first year to enable their temporal variation
- and for smaller deviation and thus more or less strict periodicity in the following years.

We will make this clearer in the final version.

p 10219, l 10: The t_n notation does not seem to add anything, please drop it.
Answer: Will be omitted.

p 10219, l 13: Why the new notation? The state was x, the parameters u. What happened to the relationship from p 10218 l 5? Stick to one notation.
Answer: The new quantities arise from the linearization, where now the deviations from the reference model variables ($y_k - y_k^{ref}$) and parameters ($u_k - u_k^{ref}$) are the model variable and parameter. Thus the new notation (in the final version $x_k := y_k - y_k^{ref}$ and $v_k := u_k - u_k^{ref}$) will be used.

p 10220, l 1: Are the parameter values in u or in v now, where does the u come from?
Answer: See above, we will make this point very clear in the final version.

p 10221, l 2: "Bb_n"
Answer: Has to be b_k, will be changed.

p 10222: Are the parameters periodic in Section 4.1 already? Mention this explicitly. It is not quite clear if the results from Fig.1 correspond to the parameters in Fig. 6-8.
Answer: Yes, the parameters are periodic in Section 4.1. We have focused on the fit of model output to observational data in this section. We will emphasize this point in the final version.

Fig. 2,3: For a better comparison it would make sense to join these 2 figures together and have the information from the LQOC and SQP in one plot.
Answer: Will be modified.

Fig. 4: The plots are dominated by a few single pixels with high costs, the rest consists of shades of blue. An adaptation of the colormap would make this figure more informative. Also, if there are pixels without observations, they should be in a color which does not appear in the colormap (e.g., white) to make it distinguishable from a perfect fit.
Answer: Will be changed using a more appropriate colormap.

Fig. 6,7,8: These plots can probably be summarized in one, maybe two plots by showing only 2 years and mentioning the almost perfect periodicity. It would also be beneficial to add the optimized constant (SQP) parameter values and the parameter bounds from Table 1 (as straight lines) into the plot as well for comparison purposes.
Answer: Will be included.

Fig. 9: There appear to be 2 legends for each plot, summarize this information in one.
Answer: Will be changed.