

Interactive comment on “A comparison between gradient descent and stochastic approaches for parameter optimization of a coupled ocean–sea ice model” by H. Sumata et al.

Anonymous Referee #2

Received and published: 6 February 2013

General comments:

The manuscript "A comparison between gradient descent and stochastic approaches for parameter optimization of a coupled ocean–sea ice model" compares two existing methods for parameter estimation and compares them on a non-linear system of interest, the Arctic ice-ocean NAOSIM model. The issues raised are applicable to other ice-ocean models independently of the physics they actually contain. I even expect the issues of irregularity of the cost function will become more and more relevant as coupled ice-ocean improve their physics and horizontal resolution because their increasing nonlinearity. The approach is conceptually appealing to a sea ice modeler in need of

C1583

a rigorous approach to tune the model parameters, because this is a recurrent activity needed every time the model resolution is increased or a new set of forcing fields is made available. The article is very well structured, makes a good overview of the literature in the field but the criteria used to compare the two methods do not make a fair trial in my view: The gradient descent stopping criterion is based on the classical norm of the gradient but the micro-Genetic Algorithm (mGA) has a fixed number of iterations based on offline idealized tests, and generally much larger than the FD method (100 to 400), which makes its total cost much larger. For that cost, the users could instead apply an advanced data assimilation method (type 4DVAR or EnKF with 100 members running in parallel) and carry out state estimation over several years. The mGA method is not strongly parallel either, since only 5 or 8 individuals run in parallel, but the 400 iterations have to be run sequentially. In contrast, the FD method converges in 44 iterations and thus seems much more acceptable in terms of both wallclock time and total cost, while still delivering reasonably good parameter values. So I am not convinced that the proposed winner will be appealing to the community at large.

The article also lacks a few precisions that would make the mGA method more intuitive. I understand that the method has been published elsewhere but the three pages devoted to the presentation of the method are often vague (Section 2.5, see specific comments below). So the conclusions seem to be heavily biased towards the mGA method but I may have misunderstood some explanations in the paper. In the latter case, I believe the paper should benefit from a few clarifications.

Specific comments:

1) If I understand correctly, a user of the mGA method would need to run 10 experiments \times 5 individuals \times 400 iterations = 20.000 model years in order to fine tune model parameters. This seems like an overkill, since state estimation can be performed with 100 times the model cost or less (see Sakov et al. 2012, Mathiot et al. 2012, Fenty 2010). I can put the question in different terms: given a limited computer time allocation, which method will provide the best parameters?

C1584

2) Description of the mGA method. Is the natural selection process keeping only one - fittest - individual or several of them?

3) The recombination of genes (exchanging bits of the seven 8-bits real parameters) does not seem to have a notion of "distance" between the parameter values, so that parameters are always explored across the full range of admitted values. The fittest individual does make it to the next generation, but I do not understand how the other individuals can bear any resemblance to it although this seems to be needed in the "assessment of convergence" in Fig. 2.

4) The information about the total cost of each method is misleading because it does not make the distinction between the wallclock time (duration of one model run times number of sequential iterations) and the total computational cost (wallclock time times number of parallel runs). Both are critically important numbers for potential users and should appear as extra lines in table 3.

5) The choice of initial parameters for the gradient descent method (Table 4) could be better explained. They appear as an arbitrary -but not random - sets of numbers.

Typos and minor remarks:

- p.5 l. 29: mulTi-processor.

- P.13, Eq. 3: Is the exponent really 20, or is it a square with a typo?

- P.14. l3: ontrol parameter.

- P.17 l.3 parameterspace

- P.18 l.16: What is a "pseudo function"?

- Figure 3 does not make the mGA much clearer than the text does. It would be more illustrative to show examples of the recombination of genes and renovation of generation.

C1585

- Fig. 13 is hard to read. Line plots showing converging parameter estimates could be clearer.

Suggested reading:

Fenty, I., PhD Thesis. MIT. 2010.

Mathiot, P., König Beatty, C., Fichet, T., Goosse, H., Massonnet, F., and Vancoppenolle, M.: Better constraints on the sea-ice state using global sea-ice data assimilation, *Geosci. Model Dev.*, 5, 1501-1515, 2012

Sakov, P., F. Counillon, L. Bertino, K. A. Lisæter, P. R. Oke, and A. Korabely, TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic, *Ocean Sci.*, 8, 633-656, 2012.

Interactive comment on *Ocean Sci. Discuss.*, 9, 3593, 2012.

C1586