Response to the interactive comment C1583 – C1586 by Anonymous Referee #2 to the manuscript, osd-9-3593-2012, "A comparison between gradient descent and stochastic approaches for parameter optimization of a coupled ocean-sea ice model" by H. Sumata et al.

We would like to thank the anonymous referee #2 for providing a thoughtful assessment of our work, which will certainly help to improve the explanations regarding experimental design and description of the results in the manuscript.

Response to general comments:

The article is dealing with the more general question if a method is applicable to the given problem, i.e. the parameter optimization of a highly non-linear system as a coupled sea ice – ocean model. We admit that the computational cost of the micro Genetic Algorithm (μGA) is high but to our knowledge it is the only algorithm that is able to cope with the very heterogeneous structure of the cost function (see below).

Generally, the FD method is difficult to apply to a minimum search of an ill-shaped function, because it prevents estimating the gradient with sufficient accuracy. As shown in Sec. 3.1, we are dealing with such a function here. In the manuscript we tried to apply the FD method to the problem using a moderate increment size which is not affected by the spiky micro-scale structure but roughly captures the magnitude of the local gradient. This approach can reduce the cost function to some extent (see Sec. 3.2.). However, because of the limitation of the increment size, we cannot estimate the gradient with high accuracy (Sec 2.4: 4th paragraph), and the inaccurate gradient significantly worsens the accuracy of the solution (Sec. 3.2: 2nd and 3rd paragraph). Actually, we cannot increase the accuracy of the solution by increasing the number of iterations due to the inaccurately estimated gradient. After dozens of iterations, the algorithm sometimes provide a cost function slightly larger than the previous estimate, despite it searches a solution in the gradient descent direction. The situation arises due to the nature of the function which is to be minimized, leading to inapplicability of the FD method to find a solution with reasonable accuracy in our case. In addition, because of the very ill-shaped form of the cost function (Sec. 3.1: 2nd paragraph) it is very likely that the FD algorithm has captured around a local minimum. This is tested by applying the FD method with different initial parameters (Sec. 2.6: 2nd paragraph). The posterior parameters vary strongly as is shown in Fig. 9, Fig. 11 and Fig. 15a. At a small number of iterations (less than about 100), the μGA also fails. However, by increasing the number of iterations, the μGA can improve the results, whereas the FD method cannot. This is the reason why we concluded that the FD method fails to estimate parameters while the μGA can do.

If we compare the computational costs between FD and μGA for the same number of iterations the μGA shows better performance than the FD method (see response to the specific comments).

For the adjoint method (we assume that this is meant by 4DVAR – our FD method is also 4DVAR but without using the adjoint), the situation is the same as for the FD method. Both methods utilize the gradient of the cost function with respect to the control vectors to find an optimum. Actually, the motivation for the present work stems from our experience in developing an adjoint model (Kauker et al., 2009). If the cost function is ill-shaped (strongly nonlinear), all the gradient descent methods inevitably encounter the same problem. Note that Fenty (2010) has not used 4DVAR for parameter estimation but for state estimation. In state estimation the control vector is much larger (order of millions) and any gradient descend search algorithm is not that likely to stop at local minima.

EnKF is a powerful method for state estimation (Mathiot et al. (2012) and Sakov et al. (2012) are examples for state estimation). However, it has weaknesses for parameter estimation for strongly nonlinear systems because the error distribution becomes non–Gaussian (Kivman, 2003). Trials of parameter estimation with EnKF are still underway (e.g., Doron et al, 2011; Simon and Bertino, 2012) but we know of no publication about parameter estimation with EnKF for a strong non-linear systems like the sea ice – ocean system. It is not possible to derive the computational costs for parameter estimation with EnKF from the computational costs of state estimation.

Response to specific comments:

1. As Referee #2 pointed out, state estimation by an adjoint method (Fenty, 2010) or by an EnKF (Sakov et al., 2012; Mathiot et al., 2012) can be performed within 100 times the costs of one model run. However, it is not straightforward to compare the computational costs for state estimation with those for parameter estimation.

   In general the μGA performs better than the FD method even in a limited computational time allocation. If we compare the results from the FD method with those from the μGA for the same number of iterations (number of iterations is 32, which is approximately the mean iteration number of the FD method with the standard code, 31.7), the mean estimated cost function values are 15.26 (FD) and 15.14 (μGA – 2), respectively. The total numbers of model runs (10 experiments each) corresponding to these iterations are 2536 (FD) and 1281 (μGA), respectively.

2. This is just an analogue, but we use this as an idealized concept to explain the essence of the selection process. Probably the things occurring in the biosphere are much more complicated than the idealized concept applied here.

3. In a sequence of consecutive search between reinitialization processes, the μGA does not perform a search across the full range of parameter values, but performs a search within a space spanned by the possible combination of binary bits initially contained in the generation. Throughout a sequence of consecutive progress of generations, the order of the binary bits of individuals converge to that of the fittest individual, since the gene of the fittest is always transferred to the next generation without any change, and is repeatedly used for recombination process. Actually, after several generations, the order of the binary bits in the generations become closer to each other, and the diversity of the corresponding parameter values also becomes small (Note that the convergence occurs in the gene's space, not in the parameter space. In the parameter space, only the diversity of parameter values reduces). To demonstrate this, we added a table showing the evolution of the parameter values (see Appendix A).

4. We revised Table 3 according to Referee #2's suggestion.

5. The initial parameter sets are given as follows. The first parameter set is given by the standard parameter setup used in previous studies of NAOSIM, and the second (the third) parameter set is composed of the upper (lower) limits of respective parameter values. The forth parameter set is composed of a combination of parameter values with their upper (h0, cdwat and cdsens) and lower (P*, cdwin, cdlat and albedo) limits. The fifth to the tenth parameter sets are composed of random combinations of selected parameter values, modified in discrete, nearly equidistant steps inside the parameter ranges. We added an appropriate explanation in the revised manuscript (section 2.6, 2nd paragraph).

Response to typos and minor remarks:

-P.5 l.29: The manuscript was revised.

-P.13, Eq. 3: Yes. The exponent is 20. We imposed a large penalty to prevent the search algorithm to leave the prescribed parameter range.

-P.14. L.3: The manuscript was revised.

-P.17. L.3: The manuscript was revised.

-P.18. L.16: The pseudo function is defined as follows;

$$cost = \sum_{i=1}^{M} \left(\frac{m_i}{m_i^{centralvalue}} - 1 \right)^2 \quad ,$$

where M is the number of optimized parameters ($M = 7$), m_i is i-th parameter value and $m_i^{central\ value} = (m_i^{max} + m_i^{min}) / 2$ is the central value of the prescribed range of i-th parameter. The function is smooth and continuous everywhere and has only one minimum, $cost = 0$, at $\boldsymbol{m} = \boldsymbol{m}^{\text{central value}}$. In order to show the definition of the pseudo function and preliminary experiment, we added Appendix B.

-Figure 3: We would like to keep Fig.3 as it is. For readers not familiar with the basis of genetic algorithms it is informative in our view. However, we added Appendix A in which an example of renovation of parameter values by the μGA is provided.

-Figure 13: We would like also to keep Fig. 13 to facilitate the comparison with Fig. 9 and 11. Instead, we provided an additional figure showing the line plots of converging parameter estimates, as Referee #2 suggested (Appendix C).

Reference

Doron, M., Brasseur, P., Brankart, J.- M.: Stochastic estimation of biogeochemical parameters of a 3D ocean coupled physical – biogeochemical model: twin experiments, J. Mar. Sys., 87, 194—207, 2011.

Fenty, I. G.: State estimation of the Labrador Sea with a Coupled Sea Ice – Ocean Adjoint Model, Ph.D. Thesis, MIT, 2010.

Kauker, F., Kaminski, T., Karcher, M., Giering, R., Gerdes, R., Voßbeck, M.: Adjoint analysis of the 2007 all time Arctic sea-ice minimum, Geophys. Res. Lett., 36, L03707, doi:10.1029/2008GL036323, 2009.

Kivman, G. A.: Sequential parameter estimation for stochastic systems, Nonlinear Processes in Geophysics, 10, 253—259, doi:10.1016/j.jmarsys.2011.04.001, 2003.

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

Sakov, P., Counillon, F., Bertino, L., Lisæter, K. A., Oke, P. R., Korablev, A., TOPAZ4: an ocean – sea ice data assimilation system for the North Atlantic and Arctic, Ocean Sci., 8, 633—656, doi:10.5194/os-8-633-2012, 2012.

Simon, E., Bertino, L.: Gaussian anamorphosis extension of the DEnKF for combined state parameter estimation: Application to a 1D ocean ecosystem model, J. Mar. Sys., 89, 1—18, doi:10.1016/j.jmarsys.2011.07.007, 2012.