Geoscientific Model Development Discussions

Geosci. Model Dev. Discuss., referee comment RC1
https://doi.org/10.5194/gmd-2021-175-RC1, 2021

Comment on gmd-2021-175

Anonymous Referee #1

Referee comment on "A derivative-free optimisation method for global ocean biogeochemical models" by Sophy Oliver et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2021-175-RC1, 2021

General Comments

The article examines the ability and performance of the mathematical optimisation software DFO-LS (Cartis et al., 2019) to calibrate global biogeochemical ocean models --- an important task in the field of earth system modelling.
The study is exemplified with the BGC model MOPS (Kriest et al, 2015), coupled to a 2.8° configuration of the MIT general circulation model (Marshall et al., 1997), using the transport matrix method (TMM, Khatiwala, 2007) as a "physical linearisation" in order to enable more efficient model evaluations.
The objective function to be optimised is a weighted sum of squared differences between observed tracer values and their modeled counterparts (a measure frequently applied to global ocean models). Since DFO-LS is a derivative-free method for least squares problems it appears to be a promising candidate for the model calibration task. Consequently, the paper is tightly focused on the calibration performance of DFO-LS and its ability to deal with noise and sparse observational data. It is compared to a "general purpose" derivative-free stochastic search algorithm, namely an implementation of a popular variant (Hansen, 2016) of the covariance matrix adaption evolution strategy (CMA-ES) by N.~Hansen. CMA-ES has been applied to the same TMM coupled BGC model by Kriest et al., 2017, as one of the first calibration attempts w.r.t. globally coupled BGC models.
The article is well laid out. It gives a quick overview about both optimisation tools and presents a tight experimental design with TWIN experiments that yield informative results and discussions.
The performance results of DFO-LS are quite promising. At least for the given model setup with TWIN experiments, there is a clear decrease of the number of function evaluations, (i.e., of expensive model simulations) required to recover the selected set of BGC parameters (and the model-data misfit value) within sufficient accuracy. Even the number of DFO-LS iterations is less than with the applied CMA-ES version in the setting by Kriest et al.\ (there, a CMA-ES iteration means 10 model simulations in parallel while a DFO-LS iteration evaluates the model for a single parameter vector, which is derived by constructing a quadratic "surrogate model" and minimising it within a trust region). Further, a good convergence property of DFO-LS with respect to both noise and sparse

data is empirically confirmed.
It remains to be investigated if the observed good performance is sustained with real-world observations (which I would expect) and for other globally coupled BGC models (as designated by the authors).

Specific Comment

I am curious about the impact of the cost function definition, which is explained in quite some detail in section 2.5. Did you initially work without a partition into 27 biomes and 3 tracers? Does the required number of function evaluations (e.g. to reach "baseline optimality") significantly increase if the objective function is provided as a single sum of squared differences, only?

Technical Corrections

- The CMA-ES subsubsection and the DFO-LS subsection have different depth levels (2.3.1 and 2.4). Using 2.3.1 and 2.3.2 (or 2.4 and 2.5) would be more consistent.
- The sentence beginning in line 157 seems incomplete. Is there a word missing?
- I would remove the first three words "Results table of" from the captions of Tables 3 and 4.
- In Figure 4 the parameter boundary lines are not "red dotted" as stated in the caption but black thin lines.
- Line 320: "verses" => "versus"