Towards better biogeochemical ocean models

Assessment of global biogeochemical ocean models via multi-objective optimisation

V. Sauerland, I. Kriest, C. Leonhard, A. Srivastav, *Kiel University, GEOMAR Helmholtz Centre for Ocean Research Kiel*

In Short

- Calibration of biogeochemical ocean models facilitates model intercomparison, but which parameters are optimal depends on tracers of interest
- It is difficult to pose proper bounds on the model-data misfit of some tracer while minimising the misfit of another tracer or to properly join different tracer misfits into one objective function
- Multi-objective optimisation eases the task to find good compromise solutions, which allows for better model assessment
- We integrate and apply multi-objective optimisation into a framework using an efficient tool for global simulation of oceanic biogeochemical processes

Biogeochemical models mirror important processes of the global marine ecosystem.They have a similar influence on model predictions [1] as ocean circulation [2] but depend on roughly constrained parameters. There is a plethora of biogeochemical models of different complexity, each of which appears to be particularly skilled to meet certain (types of) observations. However, due to the parametric uncertainties, there seems to be no superior model for all research questions.

Therefore, the parameters of any biogeochemical model must be adjusted until model predictions meet corresponding observations. Optimisation upto nearly optimal model-data misfit values facilitates model assessment [3] but is computational expensive in the face of non-convex processes and their expedient coupling to global ocean circulation.

Investigating the effect of data distribution and availability on model assessment with single misfit functions requires decisions about weights applied to different data sets, or a particular form of misfit function, which may be very influential for the optimal parameter choice [4]. But optimisation with respect to a certain target can affect further important properties either positively or negatively.

Examining trade-offs between such opposing effects will improve our understanding about the appropriateness of diverse biogeochemical models for particular research questions.

Optimisation Framework Parameter optimisation facilitates model intercomparison. However, good parameters with respect to one type of observations can be bad for another type of observations due to the parametric uncertainties. An assessment of these undesirable effects would be supported by a couple of compromise parametrisations that offer well distributed trade-offs between different targets. Single objective optimisation can consider such matters in terms of constraints on certain model-data misfits while minimising other misfits. But it is difficult to pose proper bounds on the model-data misfit of some tracer while minimising the misfit of another tracer or to properly join different misfits into one objective function.

We address more than one objective function, say $f_1, \dots, f_k$, instead of combining k objectives in a single function. Two solutions $x \neq y$ are said to be incomparable if $f_i(x) > f_i(y)$ but $f_j(x) < f_j(y)$ for some $i \neq j$. Multi-objective optimisation algorithms aim to find good incomparable solutions such that the user (a) obtains (nearly) optimal solutions for each objective, (b) can choose a best compromise solution in his/her opinion, (c) can draw more accurate conclusions about the interplay of parameters with regard to their influence on model skills.

The field of multi-objective (MO) optimisation is treated with EAs [5] and EDAs [6], including CMA-ES. As we made good experience with our parallelisation of a single-objective CMA-ES [7], we parallelised a MO CMA-ES [8] for firstly two objectives. Parallelisation and handling of the framework are similar to the single-objective case. It is, thus, generic and universally applicable with a high level of portability among different architectures.

Concerning model simulation, we keep using the “Transport Matrix Method” (TMM) for easy and generic coupling between different biogeochemical models and circulation [9] available via Github [10]. It allows easy switching between transport matrices.

Achievements and further goals In the first year of the project we dealt with two objectives, addressing an important research question which is centered around how well models represent oxygen minimum zones (OMZs) [11,12]. A posteriori analysis of OMZs simulated by individuals of the optimisation trajectory (obtained in HLRN project shk00025) showed that this metric can be inversely related to the root-mean-square-error (RMSE) applied so far. Facing this situation we compared single-objective model calibration with bi-objective model calibration, the

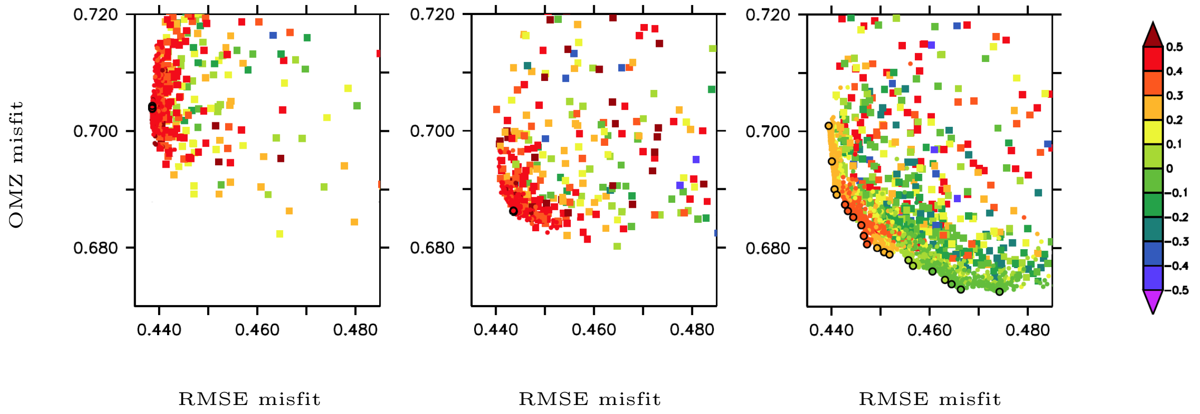

Figure 1: *Projection of parameter values (normalised, colour scale) onto model misfits, in the vicinity of the optimal solution of optimisations. Parameter values have been scaled by* $(\theta - \theta_c)/(\theta_u - \theta_l)$*, where* θ *is the parameter value,* θ_u *and* θ_l *are upper and lower boundary constraints, and* $\theta_c = (\theta_u + \theta_l)/2$ *is the center of the allowed interval. Lined circles denote the parameters of the last generation. Left panel: minimising the RMSE between simulations and observations of oxygen and nutrient distributions. Mid panel: minimising of the sum of the RMSE misfit and the mismatch between simulated and observed OMZs. Right panel: bi-objective (simultaneous) optimisation of RMSE misfit and OMZ mismatch.*

latter providing a clearly better set of compromise solutions. Fig. 1 compares results of two single-objective model-data misfit optimisations regarding both, OMZ only and the sum of OMZ and RMSE, with a bi-objective optimisation regarding OMZ and RMSE, simultaneously. The bi-objective optimisation converges to a collection of solutions, representing well distributed and improved compromises between the respective single-objective optima. The obtained compromise solutions somewhat follow the so called Pareto-principle meaning that we only need to sacrifice 20% of the best value with respect to one objective in order to reach 80% of the other objectives' optimum.

While bi-objective model calibration experiments go on in a broader context within HLRN project shk00033, we aim to expand our optimisation framework to more than two objectives. Compromise model calibrations with respect to three or four types of target observations should further help to reveal reasons behind parametric uncertainties.

Past model calibration studies observed diverging results with respect to different ocean sites, too (see, e.g., section 7.2 in [13]). Revisiting this issue with multi-objective optimisation is an another application we aim to address in the second year of this project.

WWW

http://www.uni-kiel.de

More Information

[1] I. Kriest, S. Khatiwala, A. Oschlies, *Prog. Oceanogr.* **86**, 337–360 (2010). doi: 10.1016/j.pocean.2010.05.002

[2] R.G. Najjar et al., *Glob. Biogeochem. Cycl.* **21**, GB3007 (2007). doi: 10.1029/2006GB002857

[3] I. Kriest, *Biogeosci. Discuss.* (2017). doi: 10.5194/bg-2017-71

[4] G.T. Evans *J. Mar. Sys.* **40–41**, 49–54 (2003). doi:10.1016/S0924-7963(03)00012-5

[5] K. Deb, *Multi-objective optimization using evolutionary algorithms* John Wiley & Sons (2009)

[6] M. Hauschild, M. Pelikan, http://medal-lab.org/files/2011004_rev1.pdf (2011)

[7] I. Kriest, V. Sauerland, S. Khatiwala, A. Srivastav, A. Oschlies, *Geosci. Model Dev.* **10**(1), 127–154 (2017). doi:10.5194/gmd-10-127-2017

[8] C. Igel, N. Hansen, S. Roth, *Evolutionary Computation*, **15**(1), 1–28 (2007). doi: 10.1162/evco.2007.15.1.1

[9] S. Khatiwala, *Glob. Biogeochem. Cycl.* **21**, GB3001 (2007). doi:10.1029/2007GB002923

[10] https://github.com/samarkhatiwala/tmm

[11] A. Cabre, I. Marinov, R. Bernadello, D. Bianchi, *Biogeosciences* **15**, 5429–5454 (2015). doi:10.5194/bg-12-5429-2015

[12] V. Cocco, et al. *Biogeosciences* **10**, 1849–1868 (2013). doi:10.5194/bg-10-1849-2013

[13] M. Schartau, et al. *Biogeosciences* **14**, 1647–1701 (2017). doi:10.5194/bg-14-1647-2017