Journal of Computational Science 4 (2013) 423–437

Contents lists available at ScienceDirect

Journal of Computational Science

journal homepage: www.elsevier.com/locate/jocs

Marine ecosystem model calibration with real data using enhanced surrogate-based optimization

M. Prieß [c], S. Koziel [b], T. Slawig [a,*]

[a] Department of Computer Science, Christian-Albrechts Universität zu Kiel, 24098 Kiel, Germany
[b] Engineering Optimization & Modeling Center, School of Science and Engineering, Reykjavik University, Menntavegur 1, 101 Reykjavik, Iceland
[c] GEOMAR – Helmholtz Centre for Ocean Research Kiel, Düsternbrooker Weg 20, 24105 Kiel, Germany

ARTICLE INFO

Article history:
Received 23 July 2012
Received in revised form 2 April 2013
Accepted 8 April 2013
Available online 18 April 2013

Keywords:
Marine ecosystem model
Surrogate-based optimization
Model calibration
Parameter identification
Response correction
Low-fidelity model

ABSTRACT

We have already shown in a previous methodological work that the surrogate-based optimization (SBO) approach can be successful and computationally very efficient when reconstructing parameters in a typical nonlinear, time-dependent marine ecosystem model, where a one-dimensional application has been considered to test the method's functionality in a first step. The application on real (measurement) data is covered in this paper. Essential here are a special model data treatment and further methodological enhancements which allow us to improve the robustness of the algorithm and the accuracy of the solution. By numerical experiments, we demonstrate that SBO is able to yield a solution close to the original model's optimum while time savings are again up to 85% when compared to a conventional direct optimization of the original model.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Numerical simulations play a key role to simulate and predict processes in the earth's climate system, ranging from fluid dynamics (in atmosphere and oceans), thermodynamics, radiative transfer to bio- and biogeochemical interactions, e.g., in marine or other type of ecosystems. The underlying models are typically formulated as time-dependent partial differential equations (PDEs) [1–3].

Since many important processes are non-linear, the numerical effort to simulate the whole or parts of the climate system with a satisfying accuracy and resolution is quite high. This motivates the development and use of reduced order models by e.g. coarser discretizations (in time and/or space) or by parametrizations to reduce the system size and thus the computational effort [3]. Through those parametrizations, several additional parameters enter the system. Many of them are not known beforehand and not directly measurable.

Growth and mortality rates in marine ecosystem models [4,5], one of which is taken as a test case for the proposed optimization methodology, are examples for such unknown parameters. Marine ecosystem models describe the transport, interactions and biogeochemistry among ocean biota. The modeled processes comprise the marine biogeochemical cycles among carbon and the major nutrients (see, e.g., [4–6]). Marine ecosystem models are of great importance for understanding the oceanic uptake of carbon dioxide and for projections of the marine ecosystem's responses to climate change.

Generally, before a transient simulation is possible, a marine ecosystem model has to be calibrated, i.e., the relevant parameters have to be identified such that the simulated tracer concentrations ideally resembles the actual physical and biogeochemical processes. Moreover, the ability to forecast future dynamics within the marine ecosystem crucially depends upon parameterizations of the desired biogeochemical processes. Thus, since there is no general consensus on what is *the* "correct" ecosystem model or model structure to represent the observed quantities under consideration, an assessment of the different models/parametrizations highly depends on their validation against the given observed quantities. Mathematically, this parameter identification can be classified as a least-squares type optimization or inverse problem (see, e.g., [7]). This optimization or calibration process requires a substantial number of typically expensive function and optionally sensitivity or even Hessian matrix evaluations.

Straightforward attempts by employing the *high-fidelity* or *fine* model under consideration directly in an optimization loop using conventional optimization techniques are therefore tedious or even beyond the capability of modern computer power, especially when

* Corresponding author. Tel.: +49 431 880 4490; fax: +49 431 880 7617.
E-mail addresses: mpriess@geomar.de (M. Prieß), koziel@ru.is (S. Koziel), ts@informatik.uni-kiel.de (T. Slawig).

1877-7503/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jocs.2013.04.001

using traditional, gradient-based techniques. The need for an accelerated optimization process, which especially becomes important while handling complex three-dimensional models, becomes critical.

Surrogate-based optimization (SBO) addresses this issue by shifting the computational burden from the accurate and expensive high-fidelity model to its fast but yet reasonably accurate surrogate. More specifically, the idea of SBO is to replace the fine model in the optimization process in the sense of providing predictions of the model optimum. The surrogate can be created by approximating sampled fine model data (so-called *function-approximation surrogates*, see [8–10]) or by employing a so-called physics-based *low-fidelity* or *coarse* model, a computationally cheap representation of the fine one. The latter approach is used in this paper. Since the accuracy of the coarse model is usually not sufficient to directly use the latter in lieu of the fine model in an optimization loop, it is often necessary to use suitable alignment/correction techniques to reduce the misalignment between the coarse and fine model responses. Popular correction/alignment techniques include response correction [11] and space mapping [12]. Surrogate-based optimization is widely and very successfully used in engineering sciences (see, e.g., [12–14,8]).

As a case study, in order to investigate the applicability of an SBO methodology to the optimization of marine ecosystem models, we consider a representative of the class of one-dimensional models. Clearly, the computational effort in a one-dimensional simulation is significantly smaller than in the three-dimensional case. However, the complexity of the response of this specific model is comparably high. Thus, although one-dimensional, this model serves as a suitable and computationally affordable test example to initially check the feasibility of the proposed optimization methodology. These initial experiments are indispensible to gain essential information on the principal functionality of the SBO algorithm. Clearly, to further demonstrate the efficiency of the SBO approach, future work will have to investigate applicability of this method to other, also three-dimensional biogeochemical models as well as distinct measurement data.

Exhaustive optimization runs by using both local, gradient-based and global, genetic algorithms have been previously performed for this specific model (see, e.g., [15–17]). However, it is not the focus of this paper to further assess the quality of the optimal solution obtained there. For the purpose of this paper we tentatively accept the previously found minima. Also, we do not seek a quantitative interpretation of the solutions obtained by SBO in the biogeochemical context. Our aim clearly is to demonstrate the applicability of the proposed methodology to the parameter optimization of the considered model. More specifically, the focus is to demonstrate that, by exemplary optimization runs, SBO is able to yield a solution close to the one obtained by a direct fine model optimization at low optimization costs.

One straightforward way to introduce a physics-based coarse model is to reduce the spatial and/or temporal resolution, whereas the latter is used for the selected model in this paper. Moreover, we use a *multiplicative* response correction technique for the alignment of the coarse and fine model response.

In our previous work [18], a basic formulation of this surrogate was sufficient to create a reliable approximation, yielding a remarkably accurate solution at low computational costs. This was verified by *model generated, attainable data*.

In this paper, the application on *real* data is covered. Utilizing additionally fine and coarse model sensitivity information ensures the zero- and first-order consistency conditions between the fine model and the surrogate, i.e., agreement in function values and first-order derivatives. Embedding the algorithm in a trust-region framework [19,20] allows us to further improve the robustness of the SBO and accuracy of its solution. The trade-offs between the solution accuracy and the extra computational overhead related to sensitivity calculation will be addressed. We show the results of an exemplary SBO run and compare the solution to those obtained by a *direct* fine and coarse model optimization. We demonstrate that, on the one hand, a direct optimization of the fine model requires a substantial number of comparably expensive fine model evaluations. On the other hand, a direct coarse model optimization is computationally cheap but yields a rather inaccurate solution only. We subsequently show that the multiplicative response correction technique substantially helps to improve the coarse model's accuracy while the optimization costs are still comparably low. Using this approach within SBO, a solution close to the one obtained by a direct fine model optimization while reducing the optimization costs down to 15%.

The structure of the paper is as follows: In Section 2.1, we briefly highlight the special properties of marine ecosystem models, our model example, and the corresponding optimization problem. We introduce the basic idea of surrogate-based optimization in Section 3. The coarse model that we use as a basis to create a surrogate, is recalled in Section 4. The response correction approach used to obtain the surrogate is motivated and described in Section 5. The optimization setup, numerical results and discussion of exemplary optimization runs are provided in Sections 6 and 7. Section 8 concludes the paper with a summary and an outlook.

2. Marine ecosystem models

Marine ecosystem models mainly consist of two parts, namely the ocean circulation and the biogeochemical model (see, e.g., [4,5,21]). The coupling between ocean circulation and the biogeochemical interactions such as photosynthesis is mostly regarded as a one-way coupling. This means that the influence of the biota on the circulation (including temperature and maybe salinity distribution) is assumed to be negligible and thus is often omitted. Velocity and temperature fields are computed beforehand by an ocean circulation model and only used as *forcing data* (so-called *off-line* mode) for the biogeochemical simulations which significantly reduces the computational effort [22]. See for example [23] where such an off-line computation has been thoroughly described and investigated for an atmospheric model. Our example model (cf. Section 2.1) is simulated in such an off-line mode.

The model equations consist of a system of coupled advection-diffusion-reaction equations, where the reaction terms (also called *source minus sink*, or *sms* terms) are given by the biogeochemical interactions between the biogeochemical tracers. A system of these *transport equations* for n_t tracers then generally reads

$$\frac{\partial y_i}{\partial t} = \operatorname{div}(\kappa \nabla y_i) - \operatorname{div}(v y_i) + q_i(y, \mathbf{u}), \quad i = 1, \ldots, n_t \tag{1}$$

where $y_i(t, x) : I \times \Omega \to \mathbb{R}$, $I = [0, T]$, denotes the concentration of tracer i at time t and the spatial location x. If no interactions with the atmosphere is taken into account, homogeneous Neumann conditions on the boundary $\Gamma = \partial\Omega$ for all concentrations are employed, i.e.,

$$\frac{\partial y_i}{\partial n} = n \cdot \nabla y = 0 \quad \text{on } I \times \Gamma, \quad i = 1, \ldots, n_t, \tag{2}$$

where n denotes the normal vector. The time dependent turbulent mixing/diffusion coefficient $\kappa(t, x) : I \times \Omega \to \mathbb{R}$ as well as the velocity vector field $v(t, x) : I \times \Omega \to \mathbb{R}^3$ with $v = (v_i)_{i=1,2,3}$, both satisfy the Navier–Stokes equations. Since, here, the parameters $\mathbf{u} \in \mathbb{R}^{n_p}$, which are subject to the parameter optimization, are scalar coefficients in the nonlinear biogeochemical coupling terms q_i, we use a boldfaced notation.

Download English Version:

https://daneshyari.com/en/article/10332849

Download Persian Version:

https://daneshyari.com/article/10332849

Daneshyari.com