Maximilian Gelbrecht
Earth System Modelling Group
TUM School of Engineering and Design
Technical University Munich
Lise-Meitner-Straße 9, 85521 Ottobrunn, Germany
email: maximilian.gelbrecht@tum.de

Munich, April 12, 2023

Dear Dr Ham,

Thank you very much for your in-depth look at our paper. We agree that we were brushing over some potential concerns and questions that geoscientific model developers might have a bit too quickly in our previous revision. Therefore we have revised the manuscript further.

In our revised article we are much clearer on past advances that have been made with tools such as TAF and Tapenade, as well as the FEM models that can use dolphin-adjoint for deriving differentiable models. Certainly these approaches can contribute massively to the potential benefits of differentiable ESMs. We also agree with the editor that for a comprehensive ESM to be differentiable it might not be feasible for AD to tape each elementary operation. This is why approaches like the one taken by dolphin-adjoint are extremely valuable. Alternatives to that depend on the AD system in question.

With JAX, we see a strong vectorization of the code together with checkpointing schemes as an alternative to that. For their differentiable FVM PDE solver Kochkov et al remark that every time step is checkpointed. As the editor remarked, this is also the approach taken by most ML libraries. Adapting an existing model e.g. to the JAX framework wouldn't even be possible without vectorization. Häffner et al deliver a detailed account of that process in their GMD paper on the JAX-based Veros ocean model (albeit their full model is not differentiable yet). Enzyme works on the level of the LLVM IR which enables highly optimized gradient code. It will attempt to recompute most values in the reverse pass per default and cache (tape) only what is necessary (Moses, 2020). Zygote uses static single assignments form that is more memory efficient than "normal" tapes / Wengert lists and also recomputes (checkpoints) values in the reverse mode (Innes, 2019).

In the revised version we also reference more directly the recent success of differentiable CFD solvers of Kochkov et al that employ a FVM and pseudo-spectral approach in their JAX-CFD library. Their study using this framework clearly showcases the enormous potential differentiable models can have as they show an enormous speed-up in their simulations. All of this comes with the additional co-benefit of GPU acceleration as well. Ongoing projects such as. the DJ4Earth and CliMA project also work on a differentiable version of the FVM ocean model Oceananigans. We also remark that the current generation of AD tools needs much less user intervention as the editor is also absolutely correct in remarking the massive manual intervention needed for adjoint models like MITgcm.

However, it is also important to stress that not all ESM components are FEM models and not all ESM components are AO-GCMs. Let's consider vegetation models for example. To

our knowledge currently there is no differentiable vegetation model. Therefore, other tools are needed for those components and these can be vectorized models using frameworks such as JAX or Julia's SciML ecosystem.This is something that we are indeed also working on, as well as differentiable pseudo-spectral atmosphere models.

Indeed we were also a bit too quick to reject the challenges imposed by flow control. AD can without major problems differentiate many functions that are continuous but not differentiable, as e.g. the ReLU function. Differentiating through discontinuous functions does indeed involve the problems the Editor is mentioning. We are fully aware that achieving differentiable comprehensive ESMs is a very challenging task that also involves further research and a lot of tedious model translations. This is a perspective article, its purpose is also to offer a perspective, a personal viewpoint and look forward. We were motivated by the recent success of projects such as those of Kochkov et al Google Research group and Um, Holl and Thurey et al. at with their differentiable PDE solvers based on frameworks such as JAX, the ongoing work on the fully Julia-based CliMA model and our own ongoing work on a differentiable pseudo-spectral atmosphere model to propose this perspective. We believe that future model development should take differentiable programming into account to get the tremendous benefits that we outlined in this article.

Section 3 that mostly explains basic principles of AD has already been revised in response to Reviewer 2s suggestions, we added to Section 4 and 6 to explain in more detail what practitioners have to do and how this compares to past efforts with tools such as TAF and Tapenade.

We also checked our manuscript again carefully for the issues raised by Reviewer 2 and have addressed them point-by-point as follows:

- We added, compared to the original manuscript, many references, especially those referring to FEM such as dolphin-adjoint for the FEniCs and Firedrake frameworks, and past efforts for differential ESMs with tools such as TAF and Tapenade. We also extended the references to differentiable PDE solvers for CFD such as JAX-CFD of Kochkov et al and PhiFlow of Holl et, and Um et al. We hope that this reflects past and ongoing work on differentiable models more broadly

- We refer to slope limiters in the Challenges for differentiable ESMs section

- Additional references to differentable PDE solvers and studies that showcase the enormous potential of ANNs to accelerate and complement them have been added. In Section 5.3 we explain several studies that show the enormous potential for hybrid approaches that combine physics-based models with ANNs, e.g. an ANN subgrid scale parameterization. We believe that these approach would not be possible with more "old-school statistical methods",

- We revised Section 3 already in our last revision to make it much clearer what the "how" and the "what" is from the perspective of an Earth system modeler. In this revision we extended upon that in Section 4 and Section 6, to make it more clear what practical steps would have to be undertaken for a differentiable ESM

- We agree with the reviewer that computing second derivatives can have considerable benefits. While there are also computationally very intensive to compute, we already added more references to this possibility in the last revision

- We added a reference Noemi Petra's work to our manuscript as the Reviewer recommended

On behalf of the authors,

sincerely,

Maximilian Gelbrecht