nature portfolio

Peer Review File

ONLINE CALIBRATION OF DEEP LEARNING SUB-MODELS FOR HYBRID NUMERICAL MODELING SYSTEMS

Corresponding Author: Dr Said Ouala

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
This work addresses the problem of creating data-driven subgrid models for numerical simulations. This is a highly active area of research, with broad applications across science and engineering. A novel approach to construct online trained models, but circumventing the requirement of differentiable numerical schemes is introduced. The method revolves around the idea of approximating the term in the online loss function which requires differentiable numerics, allowing for online training without a differentiable simulator. The method is demonstrated on two toy models of chaotic dynamical systems, Lorenz 63 and quasi-gesotrophic fluid flow, and shown to improve performance with respect to an offline-trained model. Additionally, the authors demonstrate fine-tuning an offline trained model using the new approximate online method improves performance.

This is a highly relevant and topical work, and introduces an innovative approach to tackling the problem of online learning. The results are comprehensive and compelling, with sensible experiments, baseline model comparisons, and metrics chosen. I recommend to the editor publication of the work, after minor revisions:

1. I was going to suggest some relevant references: Nonnemacher & Greenebrg 2021 (https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2021MS002554), Frezat et al. 2023 (https://arxiv.org/pdf/2310.19385) and Pedersen et al. 2023 (https://arxiv.org/abs/2307.13144) as relevant works that explore replacing the differentiable solver with an approximate model - I see the first 2 in the bibliography, but cannot see where they are referenced in the main text? Some short discussion of the proposed work in comparison to these published works should be in the main text.

2. It seems the second part of Proposition 3.1, in equation 18, assumes that previous timesteps have been integrated using the full numerical solver, not the explicit Euler. Training over multiple timesteps, we need to compose multiple Euler timesteps together, which would lead to some accumulation or error that doesn't seem to be accounted for here.

3. Appendix f1 title typo

4. Some additional description of the computational cost of the various methods would also be interesting. For example - training an online model with RK4 integration requires long gradient graphs due to recursive calls to the neural network at each numerical timestep. This problem is circumvented in the Euler gradient approximation. However, the Jacobian approximation may lead to additional non-trivial computation. This could lead to some trade-offs in terms of compute & model performance that are important considerations.

Reviewer #3

(Remarks to the Author)
The major claims of the paper introduce an innovative method known as Euler Gradient Approximation (EGA) for enhancing online calibration in hybrid modeling systems with non-differentiable components. These claims are novel and potentially influential for the community, particularly in the numerical modeling and machine learning integration fields. However, the novelty and broader impact would be more compelling if the authors could compare their method against more established or recent approaches with appropriate citations.

While the methodology is promising, the paper could be strengthened with a more rigorous statistical analysis and additional

empirical validations across diverse systems to demonstrate robustness and reproducibility. Details on the experimental setup and parameters are somewhat lacking, which might hinder reproducibility.

The paper has the potential to influence thinking in fields that involve dynamic modeling and real-time data assimilation but would benefit greatly from a deeper theoretical exploration and clarity in model assumptions. Further evidence in the form of expanded case studies or real-world applications would enhance the work's credibility and applicability.

In terms of statistical analysis, more detailed information on the methods used for data handling and analysis should be included to assess the appropriateness and validity of the statistical conclusions drawn.

In summary, with additional details, rigorous testing, and clearer exposition, the paper could significantly impact the field. I have chosen to remain anonymous for this review process.

Version 1:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
Thanks to the authors for their comprehensive response. My comments are thoroughly addressed in the new submission, and I recommend publication of this revised version to the editor.

Online Calibration of Deep Learning Sub-Models for Hybrid Numerical Modeling Systems

Communications Physics

September 10, 2024

Said Ouala[1], Bertrand Chapron [2],
Fabrice Collard [3], Lucile Gaultier [3], Ronan Fablet [1],
Contact: said.ouala@imt-atlantique.fr

[1]Lab-STICC, IMT Atlantique, University of Bretagne Loire, France.
[2]Ifremer, LOPS, Brest, France.
[3]ODL, Brest, France.

Abstract

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions. In this document, we address the issues raised as best as possible. We have made some modifications to the structure of the main document so that it broadly complies with the style of CommsPhys. As a result, the numbering of the sections, theorems and other items might have being changed from the initial submission. In order to have concise answers to the reviewers comments, we use in this rebuttal the numbering of the tracked change version of the manuscript in blue next to the initial numbering of the first manuscript.

1 Reviewer 2

Reviewer Comment 1

This is a highly relevant and topical work, and introduces an innovative approach to tackling the problem of online learning. The results are comprehensive and compelling, with sensible experiments, baseline model comparisons, and metrics chosen. I recommend to the editor publication of the work, after minor revisions:

Response We appreciate the feedback on our work and the manuscript. Every comment is addressed carefully below, and the modifications can be found in blue in the tracked-changes version of the manuscript.

Reviewer Comment 2

I was going to suggest some relevant references: Nonnemacher and Greenebrg 2021, Frezat et al. 2023 and Pedersen et al. 2023 as relevant works that explore replacing the differentiable solver with an approximate model - I see the first 2 in the bibliography, but cannot see where they are referenced in the main text? Some short discussion of the proposed work in comparison to these published works should be in the main text.

Response

We thank the reviewer for rising this point and for the suggested references. We have a section in the Supporting Information Appendix B that discusses alternative solutions of the online learning problem. We discuss in this section the possibility of using emulators and we cited the first two references provided by the reviewer. We also discussed the possibility of designing models for the computation of the gradient of the online learning problem using optimal control formulations.

Following the reviewer's comment, we moved appendix B to the main paper. This new subsection is entitled **2.5 Related works on online learning for hybrid systems** and discusses recent, state-of-the-art, alternative solutions to the online learning problem that use i) emulators (where we also included the third reference provided by the reviewer) ii) Optimal control methods and iii) Gradient free (ensemble based) methods for which recent works also emerged. This new subsection can be found in blue in the tracked-changes version of the manuscript.

Reviewer Comment 3

It seems the second part of Proposition 3.1, in equation 18, assumes that previous timesteps have been integrated using the full numerical solver, not the explicit Euler. Training over multiple timesteps, we need to compose multiple Euler timesteps together, which would lead to some accumulation or error that doesn't seem to be accounted for here.

Response

As observed by the reviewer, the second part of the proposition 3.1 (2.1) assumes that the previous time-steps are integrated with the full numerical solver. This is particularly important since we don't want to modify the forward solver of the numerical model. Training over multiple time-steps does lead to error accumulation and these errors are being accounted for in our framework.

Specifically, when we derive the formula for the Euler Gradient Approximation (theorem 3.1 (2.1)), we show that there is an error in the gradient approximation that is equal to (please refer to appendix C.2 (B.2), equation (39) (44)):

$$\sum_{j=1}^{j=n-1} \Big(\prod_{i=1}^{i=n-j} \frac{\partial \Psi(\Psi^{n-i}(\mathbf{u}_t))}{\partial \Psi^{n-i}(\mathbf{u}_t)} \Big) O(h^2) + O(h^2)$$

Please note the sum over n time-steps which represents the error accumulation of the gradient estimation.

Due to this error accumulation, we have two versions of the EGA method, the one given in theorem 3.1 (2.1) (equation (14) (17)), which assumes that the number of training steps n is fixed. In this context, the error above is bounded by $O(h^2)$. The second version is given by corollary 3.1.1 (2.1.1) (equation (15) (18)). In corollary 3.1.1 (2.1.1), the number of training steps n changes to correspond to some final time t_f and in this context, the error accumulation above is only bounded by $O(h)$.

We have added the following text after corollary 3.1.1 (2.1.1) to explain the difference between the two versions of the EGA approximation and to link this with the accumulation of errors in the computation of the gradient.

The difference between the EGA formulation of Theorem 2.1 and the one of Corollary 2.1.1 lies in the convergence rate of the accumulated error in the gradient approximation. Specifically, we show (see equation (44) in the Supporting Information Appendix B.2) that this error accumulation term is bounded by $nO(h^2)$. In Theorem 2.1, the number of time steps n is assumed to be fixed, resulting in the error accumulation term being bounded by $O(h^2)$. In Corollary 2.1.1, n increases as h decreases, i.e., $n = (t_f - t_0)/h$, so the error accumulation is only bounded by $O(h)$.

Reviewer Comment 4

Appendix f1 title typo

Response

We thank the reviewer for noticing this typo. It was corrected.

Reviewer Comment 5

Some additional description of the computational cost of the various methods would also be interesting. For example - training an online model with RK4 integration requires long gradient graphs due to recursive calls to the neural network at each numerical timestep. This problem is circumvented in the Euler gradient approximation. However, the Jacobian approximation may lead to additional non-trivial computation. This could lead to some trade-offs in terms of compute and model performance that are important considerations.

Response

We thank the reviewer for rising this very relevant point. Using the Euler Gradient Approximation instead of the gradient of a numerical scheme that requires N functions evaluations ($N = 4$ when using RK4) leads to a simpler computational graph. Regarding the Jacobian approximation, the EGA can use a first order Jacobian approximation with allows to keep the computation of the backward pass trivial (it will only require in practice using e.g. a torch.nograd on the forward operations that are not required by the Euler Gradient Approximation).

We have added a subsection 2.9 that discusses these aspects:

2.9 EGA for an improved computational cost in training long roll-outs

1.1 EGA for an improved computational cost in training long roll-outs

Increasing the number of training steps n strongly influences the inference accuracy of both surrogate physical simulators [1, 2] and hybrid models [3, 4, 5]. When large number of time steps are required, The EGA can be seen as an algorithm that can be used to reduce the computational complexity of the computational graph in the training phase. For instance, if we assume that the numerical solver requires N Number of Function Evaluations (NFE) per time step, the EGA can be used with a first order Jacobian approximation and would require only backpropagation through a single function evaluation. This allows to have larger roll-outs at a smaller computational cost. We evaluate the ability of using the EGA on fully differentiable hybrid models to reduce the time and memory complexity of the backward pass, as well as the corresponding gradient errors in section 2.11.2 and SI appendix E.

We also added an experimental section 2.11.2 for the Lorenz 63 system compares the memory and computation performance of the backward pass computed using the computational graph of the Runge-Kutta 4 solver (and DOPRI8 in the Supporting Information Apendix E) with respect to the Euler Gradient Approximation. Overall, the EGA leads to a speedup and to a reduced memory cost of a factor of 4 with respect to standard backprop through RK4 while keeping the error levels smaller than 33% for all the tested simulation time steps n.

2 Reviewer 3

Reviewer Comment 1

The major claims of the paper introduce an innovative method known as Euler Gradient Approximation (EGA) for enhancing online calibration in hybrid modeling systems with non-differentiable components. These claims are novel and potentially influential for the community, particularly in the numerical modeling and machine learning integration fields. However, the novelty and broader impact would be more compelling if the authors could compare their method against more established or recent approaches with appropriate citations.

Response We would like to thank the anonymous reviewer for this positive feedback on our work. We acknowledge the reviewer's concern regarding the comparison of our method with state-of-the-art approaches and appreciate the opportunity to clarify this point.

We have compared our approach with both established and more recent methods. Specifically, in the QG experiment, we demonstrate that online learning, (including online learning with the EGA) overcomes the known limitations of widely used eddy-viscosity-based subgrid-scale models such as the Dynamic Smagorinsky model. Furthermore, we have compared our online learning approach with EGA to offline learning, which is the most commonly employed training method for hybrid models. Our results show that EGA avoids typical issues associated with offline learning, such as unphysical behaviors and instabilities in the simulations. Additionally, our method performs comparably to online learning with a differentiable numerical model.

In the Supporting Information Appendix G.1 (D), we extend our analysis to an other system, the slow/fast Lorenz 96 system, where we observe similar results to those obtained in the Lorenz 63 and QG experiments. In the Lorenz 96 experiment, and in addition to the fully differentiable online learning and offline learning baselines, we also compare our method to the use of a differentiable emulator (which is very recent state-of-the-art methodology for online learning). Our results show that EGA achieves similar outcomes while bypassing the need for emulator calibration.

In response to the reviewer's suggestion, we have included a new section (Section 2.5 in the tracked changes document) that discusses relevant state-of-the-art approaches, along with appropriate citations (e.g., [6]).

Reviewer Comment 2

While the methodology is promising, the paper could be strengthened with a more rigorous statistical analysis and additional empirical validations across diverse systems to demonstrate robustness and reproducibility. Details on the experimental setup and parameters are somewhat lacking, which might hinder reproducibility.

Response

In addition to the Lorenz 63 experiment and to the experiment on the QG model, we also provide in SI document Appendix G.1 (D) an experiment on the Slow-Fast Lorenz 96 model

which is a coupled ODE with slow and fast dynamics. The results are similar to the ones of the Lorenz 63 and to the QG model.

Regarding the experimental setup, we initially presented it in the Supporting Information. However, following the reviewer's suggestion, we have moved the experimental setup to the Methods section of the main paper to enhance readability. We have also included code and data availability statements, which will provide, in the final version of the manuscript, links to the GitHub repository containing the code used to generate the results.

Reviewer Comment 3

The paper has the potential to influence thinking in fields that involve dynamic modeling and real-time data assimilation but would benefit greatly from a deeper theoretical exploration and clarity in model assumptions. Further evidence in the form of expanded case studies or real-world applications would enhance the work's credibility and applicability.

Response We thank the reviewer for this positive feedback on our work. Regarding the theoretical validation of the EGA, our method was derived using error rates from the corresponding numerical schemes, and it provably converges (and experimentally, for instance, in Figure 2) to the true gradients, either quadratically or linearly, depending on how we specify the number of training steps, n. We acknowledge that these error rates do not guarantee the use of the EGA for arbitrary time steps h or an arbitrary number of steps n. However, they do ensure that we can control the error by reducing h and/or n. As for the model assumptions, we clearly state in the manuscript that the EGA is applicable to additive submodels. We also discuss in the Supporting Information Appendix possible extensions to non-additive submodels.

Due to the underlying computational requirements, in terms of compatibility of real-world numerical models and state-of-the-art deep learning frameworks), real-world applications go beyond the scope of this paper, the primary focus of which is the derivation and evaluation of the EGA algorithm. We do agree with the reviewer that real-world applications will support further the credibility and applicability of the EGA approach and our future work will specifically explore applications to the calibration of submodels in the ocean model component of Earth System Models.

Reviewer Comment 4

In terms of statistical analysis, more detailed information on the methods used for data handling and analysis should be included to assess the appropriateness and validity of the statistical conclusions drawn.

Response We understand the reviewer's concern regarding the methods used for data handling. Following the reviewer's comment, we have moved the details of the experimental section from the Supporting Information Appendix to the Methods section in the main paper. This section now includes all the necessary details to ensure the reproducibility of the experiments, along with the evaluation criteria used in the analysis.

References

[1] J. Brandstetter, D. Worrall, and M. Welling, "Message passing neural pde solvers," *arXiv preprint arXiv:2202.03376*, 2022.

[2] P. Lippe, B. Veeling, P. Perdikaris, R. Turner, and J. Brandstetter, "Pde-refiner: Achieving accurate long rollouts with neural pde solvers," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[3] B. List, L.-W. Chen, K. Bali, and N. Thuerey, "How temporal unrolling supports neural physics simulators," *arXiv preprint arXiv:2402.12971*, 2024.

[4] H. Frezat, J. Le Sommer, R. Fablet, G. Balarac, and R. Lguensat, "A posteriori learning for quasi-geostrophic turbulence parametrization," *Journal of Advances in Modeling Earth Systems*, vol. 14, no. 11, p. e2022MS003124, 2022.

[5] D. Kochkov, J. Yuval, I. Langmore, P. Norgaard, J. Smith, G. Mooers, M. Klöwer, J. Lottes, S. Rasp, P. Düben *et al.*, "Neural general circulation models for weather and climate," *Nature*, pp. 1–7, 2024.

[6] C. Pedersen, L. Zanna, J. Bruna, and P. Perezhogin, "Reliable coarse-grained turbulent simulations through combined offline learning and neural emulation," *arXiv preprint arXiv:2307.13144*, 2023.