Author Response RC2

Curlew 1.0 - Spatio-temporal geological modelling with neural fields in python

Akshay V. Kamath*[1], Samuel T. Thiele[1], Marie Moulard[2], Lachlan Grose[3], Raimon Tolosana-Delgado[1], Michael J. Hillier[4], Florian Wellmann[5,6], and Richard Gloaguen[1]

[1]Helmholtz-Zentrum Dresden-Rossendorf, Helmholtz Institute Freiberg, Chemnitzer Str. 40, 09599 Freiberg, Germany.

[2]École Nationale Supérieure de Géologie (ENSG), Université de Lorraine, 54000 Nancy, France.

[3]School of Earth, Atmosphere and Environment, Monash University, Clayton, VIC Australia.

[4]Geological Survey of Canada, Natural Resources Canada, Ottawa, ON K1A 0E8, Canada.

[5]Institute of Computational Geoscience, Geothermics and Reservoir Geophysics (CG3), RWTH Aachen University, 52074 Aachen, Germany.

[6]Fraunhofer Research Institution for Energy Infrastructures and Geothermal Systems (IEG), 44801 Bochum, Germany.

Correspondence: Akshay Kamath (a.kamath@hzdr.de)

Dear Reviewer,

We thank you for your time and effort reviewing the submitted manuscript, and are pleased that you appreciated our results. We have incorporated your suggestions into the revised manuscript, as detailed in the following pages. Please note that to facilitate the evaluation of our revision, the line numbers of the reviewers' comments refer to the originally submitted manuscript while line numbers of our responses refer to our revised manuscript. The green text refers to our reply whereas the blue text refers to the updates in the manuscript.

Kindest regards,
Akshay Kamath (on behalf of the authors)

Kamath et al. present *curlew*, a new open-source Python package designed for structural geological reconstructions using measured geological constraints. They use a neural-fields approach which allows for the incorporation of different local and global loss functions. This is a flexible framework for imposing different kinds of geological data constraints. The manuscript also demonstrates how Random Fourier Features (RFFs) can be used both to evaluate model uncertainty and to improve convergence. The codes related to this work are publicly available in a Zenodo repository. All of the files that were used to produce the diagrams of the manuscript are present in this repository. I tried reproducing the figures by running them in a Jupyter notebook. The Jupyter notebooks included pre-computed outputs, showing that the code had already been run and reproduced the manuscript's figures. However, rerunning the Jupyter cells yielded results different compared to the original outputs (see attached file). On several occasions the results could alter the interpretation (for example Figure 5d). Therefore, it is unclear whether the random seed was applied correctly. Could this issue be related to the fact that the *NumPy* and *PyTorch* operations have not been controlled by the random seed? Additionally, it would be important to assess the extent to which the algorithm's inherent randomness affects the final geological reconstruction.

We thank the reviewer for running our codes in order to reproduce our results, and are pleased that everything ran without issue. The random seed issue is likely due to differences in the *numpy* and *pytorch* versions used (as well as underlying software, e.g., CUDA), as these influence all subsequent random processes. We have updated the figures using the newer versions of the underlying packages (with the version numbers for the environment used to create the updated figures is now present in the requirements.txt file within the updated curlew_examples repository). The safest way to ensure exact replication of results is saving the *state_dict* of the trained model. This way, once trained, the model can be re-run anywhere to get the exact same results. We have also provided trained models (as joblib .pkl files) for our examples. As for the effect of the inherent randomness on the result, close to the data, the results are quite consistent (as seen in the uncertainty plot), and gets less consistent as we move away from the data. This is expected and is one of the important strengths of curlew: to be able to explore the solution space in a much more complete manner with different projections of the RFF mapping.

A question that arose while reading the manuscript concerns the comparison between *curlew* and existing structural modeling packages such as Aspen-SKUA, 3D-GeoModeller, Leapfrog and GemPy. The computational benefits of *curlew* are clear since it is a differentiable and adaptable code that can run in parallel on multi-CPU and GPU systems. However, it would be very insightful if the manuscript elaborated on whether *curlew* also provides qualitative improvements in geological detail or interpretability due to the flexible incorporation of multiple loss functions. For example, would a model produced with *curlew* give comparable results to the previously mentioned algorithms? What differences could be expected between the available codes for geological reconstructions?

This is an excellent and very interesting question, although we consider such a comparison / benchmarking exercise to be complicated enough that it warrants a publication of its own. The central theme of this current contribution is 1) To showcase the power of RFF encodings to bypass spectral bias, 2) To generate complicated models with multiple fields (something which has not been explored in detail in the aforementioned algorithms), 3) To utilise the gradient of the scalar fields to model kinematic structures such as faults and dykes, and 4) The possibility of adding additional neural networks to predict ancillary datasets, which can update geometry: a semi-supervised geological modelling approach. Comparison with existing methods is thus considered to be out of scope, and will be explored in a possible future study.

Another point that could benefit from further discussion is the tuning of the loss-function weighting factors. The authors briefly mention the use of SoftAdapt (Heydari et al., 2019) but also point out that the results are mixed. Because the weights (or hyperparameters) can significantly influence the final reconstruction, additional insight into this challenge would be welcome. For instance, what alternative methods (e.g., Bayesian optimization, gradient-based methods) could be suitable in future improvements of the algorithm and the tuning of the weighting factors?

We agree that the balancing of loss-function weighting factors is a significant challenge in multi-objective optimization. While we explored dynamic weighting strategies like **SoftAdapt**, which attempt to solve a **Pareto optimization** problem by updating hyperparameters each epoch, our preliminary results showed a tendency toward non-convergence. We suspect this is due to the highly non-convex loss surface and the distinct geometries of the different loss terms—for example, a monotonicity loss (L_g) targeting the second derivative may conflict with a local value loss (L_l) in regions of high data density.

In our tests, we have observed that certain loss terms often either "overpower" others or are ignored entirely. Regarding future improvements, we are planning on exploring several avenues:

1) **Hard-coding Geological Rules:** Instead of relying solely on loss terms, we are investigating alterations to the **Random Fourier Feature (RFF)** mapping and network architecture to inherently satisfy certain geological constraints, such as periodicity for folding.
2) **Gradient-based Surgery:** Methods like **Gradient Surgery** (e.g., PCGrad) could be suitable to prevent conflicting gradients from interfering with the optimization of separate geological events. However, these methods need to be explored/altered before applying to the complex problem of interpolating geological models.

While we consider the implementation of these methods to be future work, we now mention their possible application on L511.

Finally, I have a few minor remarks regarding the text. I would recommend simplifying some of the sentence structures. In several places the use of long sentences or multiple commas makes the text a bit difficult to follow. Shorter sentences would largely improve the readability. In addition, I noticed a few typographical errors throughout the manuscript. Careful proofreading would be helpful. For example:

Equation (1) should have v_s instead of v

Rectified.

Line 48, “do not” instead of “don’t”

Rectified.

Line 272, there is a noun missing after “from”

Rectified.

We have proofread the text and updated certain sentences to make the manuscript more readable (please see the track changes document).