# Reviewer Response

JGR-MLC: 2024JH000264-T

## Manuscript Information

***Title***: "RocMLMs: Predicting Rock Properties through Machine Learning Models"
***Corresponding Author***: Buchanan Kerswell
***Corresponding Author Affiliation***: Geosciences Montpellier, University of Montpellier, CNRS, University of Antilles, Place Eugène Bataillon, 34095 Montpellier, France
***Corresponding Author Email***: buchanan.kerswell@umontpellier.fr
***Journal Information***:

- Name: Journal of Geophysical Research: Machine Learning and Computation
- ISSN: 2993-5210
- Editor-in-Chief: Dr. Enrico Camporeale and Dr. Raffaele Marino
- Associate Editor Assigned to Review: Dr. Yangkang Chen
- Review Received Date: July 2, 2023
- Review Response Date: July 19, 2023
- Manuscript ID: 2024JH000264-T

## Reviewer #1

### General Comments

*"In this paper Kerswell et al. used the GFEM program Perple_X to predict thermodynamically self-consistent rock properties at arbitrary PTX conditions between 1-28 GPa, 773-2273 K, and mantle compositions ranging from fertile (lherzolitic). I think the machine learning models (RocMLMs) is good, the effectiveness has been well tested.*

*The authors demonstrated that RocMLM predictions against reference mantle models based on observations of seismic waves were found good agreement. RocMLMs are therefore capable of fast and highly-accurate predictions of changes in rock properties and can be implemented in high-resolution computer simulations of mantle convection. As shown in Figures 3-5 and Figure 7, the predicted profiles of density, Vp and Vs are not always consistent with the PREM or STW105. The authors should explain the reasons for these differences. On the other hand, the authors did not discuss the effect of volatile on density, Vp and Vs, I suggest to clarify this point. In a word, I recommend to publish this paper after a minor revision."*

---

Thank you for the feedback, we agree that the difference between predicted elastic properties and the geophysical reference models PREM and STW105 (i.e., RocMLM accuracy) is a major point of interest and discussion in this work. We have therefore evaluated and described the major effects on RocMLM accuracy (at various depths) in Sections 3.1 and 4.2, including the sensitivities to: 1) regression algorithms (Section 3.1, lines 312–335), 2) mantle geotherms (Section 3.1, lines 336–349 and 354–363), 3) the presence or absence of particular phase transitions (Section 3.1, lines 349–355), and 4) bulk mantle compositions (Section 4.2). We have revised Section 3.1 and 4.2 for clarification, but otherwise believe that these sections adequately describe the most important sensitivities of RocMLM accuracy with references to Figures 3–5 and Figure 7. Therefore we have not added anything substantial to the revised manuscript in this regard.

We also agree that the sensitivity of predicted elastic properties to volatile content is an important point to consider. However, a discussion on this point is outside the scope of the present study because we only test and evaluate our RocMLMs on dry mantle rocks. As mentioned at the end of Section 5 (lines 525–529), we intend to expand the RocMLM approach to include hydrated mantle rocks (a more complex system to model) in future work. A larger discussion of the effects of volatiles will be more appropriate in later work when we can demonstrate volatile effects quantitatively. For now, we have added a few words to clarify our rationale for modeling dry mantle rocks in Section 2.1.3:

*"Note that the procedure above assumes a dry and volatile-free mantle, which precludes the presence of hydrous phases and free-fluid that could significantly impact density and seismic wave velocities predicted for our training dataset. Since this work was intended to demonstrate a proof-of-concept, we adopted a dry volatile-free mantle assumption with the following rationale: dry mantle rocks are a simpler system to model thermodynamically with well-constrained thermodynamic data and equations of state [@stixrude2022], yet provide an adequate framework for comparing RocMLM, Lookup Table, and GFEM approaches. Future works will expand the RocMLM approach to hydrous systems with a wider variety of potential mantle end-member compositions."*

## Specific Comments

None

# Reviewer #2

## General Comments

*"The manuscript describes the machine learning method of replacing the look-up table method in connecting chemical properties (P-T-X) with physical properties (density-Vp-Vs) in the mantle. The*

*machine learning method is 10-1000 times more efficient in computing efficiency, which is crucial to implementing high-resolution geodynamic modeling.*

*I think the work is well-suited for publication in JGR Machine Learning and Computation. The work will attract the attention of audiences in petrology, rock physics, geodynamics, and seismology. The study is sound in terms of method and well-written. I recommend publication after minor revision.*

*There are two primary suggestions from my side:*

*1) the verification dataset is highly related to the training method, which limits the generalizability of the machine learning method. I don't think this harms this study because the machine learning method aims to replace the look-up table method with the same results and higher computation efficiency. Therefore, I suggest the manuscript clarify the goal of generalizability.*

*2) In Figure 6, the current method is not a fair comparison between neural networks and decision trees because the former method sets the epoch constant at 100. In contrast, the latter method increases trees with increased dataset size. That's my interpretation of the manuscript. Please clarify the comparison between the different methods.”*

---

Thank you for the feedback, to address the first comment of Reviewer #2, we have decided to use a more conventional kfold cross-validation approach to validate the accuracy of RocMLMs to “unseen” PTX conditions instead of a separate shifted validation dataset as we presented in the original submission. As the reviewer points out, the exact validation approach in this case does not significantly impact the results of the study because we did not intend to validate the RocMLMs on PTX conditions outside of the training dataset, but we acknowledge that our original approach was unconventional. A more conventional kfold cross-validation approach should be more familiar to JGR:MLC readers, including how to interpret the results. We used k=5 so that each RocMLM was evaluated 5 times on a 80/20 train/test split (the normal train/test split convention). We have revised Section 2.4 accordingly. Figures 3–5 and 6, and Table S1 in the Supplementary Information have also been updated with the revised results for RocMLM accuracy.

For the second comment, we believe that the reviewer is conflating RocMLM training speed vs. RocMLM inference speed. As we mention in Section 3.2 (lines 423–426), training times are irrelevant for this work because we are only interested in the prediction speeds of RocMLMs after training. We believe that some of the text may have caused confusion about this point (including the terms “execution speed” and “elapsed time” rather than “prediction speed” and “inference time”). We have changed this language throughout the manuscript and revised Figure 6 to make this point more clear. We have also added a few words to Section 3.2 to acknowledge that direct comparisons among RocMLM training times is unfair so that the reviewer’s point is also clarified:

*"We note that although directly comparing training times among RocMLMs is unfair in part due to differences in their regression algorithms and initial hyperparameters, training times for NN algorithms are many orders of magnitude larger than DT and KN algorithms (Supplementary Information). However, training times are neither limiting nor critical for geodynamic applications as training is independent from, and precedes numerical simulations. Therefore this study focuses only on RocMLM performance after training (i.e., prediction speed) , which is the most important metric when considering the practicality of coupling RocMLMs to high-resolution numerical geodynamic models."*

## Specific Comments

*"Line 293. Table 3. The Max epoch value is low. For deep learning, normally, a considerable value is given for the epoch, followed by the early stopping method. Once this study defines epoch value at 100, the neural network's computing time is almost fixed, regardless of the dataset size. In comparison, the decision tree method applied here adds more trees with the increased dataset size. In my interpretation, that's why the decision tree method takes more computing time with the increase of dataset, as shown in Figure 6, while neural network computing time is constant with dataset size."*

Thank you for the feedback. Please see the note above on training vs. inference time and the revision made to Section 3.2 to clarify this point. Figure 6 shows RocMLM inference time (prediction speed) rather than training time. For practical purposes, the training time is irrelevant in our intended use case (coupling RocMLMs with high-resolution numerical geodynamic models). The only relevant performance metric is inference time (prediction speed), and to a lesser extent, model size (in Mb) as shown in Figure 6. We believe that the reviewer's point in this case is confusing NN epochs as a hyperparameter that affects inference time, which it does not. For NN, inference speed is primarily dependent on the NN architecture, number of neurons, types of layers, and computer hardware—not on the number of epochs.

As an example to clarify the point raised by the reviewer, imagine if the NN epochs were increased by 10x or 100x. In this case, the NN models would take significantly longer to train and may become more accurate, but would not compute a forward pass (inference) any differently after training because the architecture has not changed. On the other hand, the point raised by the reviewer about DT algorithms slowing down with increasing dataset size is correct, which is why we ultimately highlight NN as the most efficient algorithm in the manuscript (Section 4.1, lines 470–475).

We acknowledge that the hyperparameters in Table 3 might not be absolutely optimal. However, we are only comparing RocMLM performance after training, so the hyperparameters need not be optimal as we have demonstrated them to be reasonably accurate (e.g., Figures 3–5). In summary, we believe that the manuscript presents a fair and robust comparison among RocMLM regression algorithms and therefore no major revisions have been made on this point.

---

*"Line 306. Generalizability. This study can clarify that the goal of the machine learning model is not to cover all compositions in all P-T-X fields. The goal of the machine learning model is to agree with the look-up table."*

Thank you for the feedback—this point is also addressed in the general comments above. We have revised Section 2.4 to reflect our updated validation methodology which adopts a conventional kfold cross-validation approach with k=5 so that each validation step is evaluated on a 80/20 train/test split. We believe that this conventional approach will be easier for the readers to interpret. To the reviewers point on the goal of our RocMLM, we revised the text to explicitly state that our intention is to reproduce the Perple_X pseudosections, rather than validate our models on PTX conditions outside of the training dataset:

*"RocMLM accuracy (in terms of RMSE) was evaluated by: 1) the kfold cross-validation approach (k=5) to determine the degree to which RocMLMs can reproduce GFEM predictions (internal accuracy), and 2) comparing RocMLMs predictions with geophysical reference models PREM and STW105 (external accuracy). The kfold cross-validation approach evaluates the capability of RocMLMs to reproduce Perple_X pseudosections by randomly splitting the Perple_X training data into 5 folds, training RocMLMs on 4/5 folds (80% of training data), and evaluating the RMSE on the remaining fold (20% of "unseen" training data). This process is repeated for each of the 5 folds."*.

We have also added these sentences in Section 2.1.1 and Section 2.1.2 to address the intended use and generalizability of our models:

*"The machine learning models developed here are not intended to be used for predictions outside of this PT range and are therefore validated/tested within these PT conditions."*

*"Similarly to the PT conditions above, the machine learning models developed here are not intended to be used for predictions outside of this compositional range and are therefore validated/tested for dry mantle compositions between PSUM and DSUM."*

---

*"Line 316. Are the compositions the same as the training set? Could you mark the range of the P-T verification dataset in Figure 1?"*

Thank you for the questions. Please see the notes above on our new revised strategy for evaluating RocMLM accuracy using a conventional kfold cross-validation approach. We no longer evaluate RocMLM accuracy using a separate shifted validation dataset, which we acknowledge was unconventional. We believe that using a more conventional approach will make the manuscript more clear to the reader in terms of our training procedure and interpretations of RocMLM accuracy. We have revised Sections 2.4, 2.1.1 and 2.1.2 to reflect these changes (see comment above).

*“Line 410. Suggestions for presentation in Figure 6: It is more intuitive for readers to comprehend if (1) the horizontal axis shows the size of the dataset with the unit of MB or GB; (2) the vertical axis in the direction from bottom up shows values from small to large, to be consistent with the horizontal axis.”*

Thank you for the suggestions. The x-axis (model capacity) represents the number of PTX points stored in Lookup Tables or the number of PTX points used to train RocMLMs, so it has no units. However, we have revised Figure 6 to have consistent ordering of the x- and y-axes and have also changed the titles, axis labels, and caption to clarify the Figure as much as possible. Figure 6b now shows model size (in Mb) instead of efficiency (in ms*Mb), which is a more intuitive unit that makes RocMLM performance easier to interpret. We have revised the relevant text associated with Figure 6 to reflect these changes.

# Additional Comments to Editor

Please note that the author affiliations have been updated. The online and print versions of the manuscript should include the author affiliations as they appear in the revised manuscript.

# References

Stixrude, L., & Lithgow-Bertelloni, C. (2022). Thermal expansivity, heat capacity and bulk modulus of the mantle. Geophysical Journal International, 228(2), 1119-1149.