Author response to referee comments

Deep learning for fast simulation of seismic waves in complex media

Introduction:
This document contains our responses and proposed manuscript changes resulting from the two reviews of our "Deep learning for fast simulation of seismic waves in complex media" discussion paper.

We would like to thank the reviewers (Andrew Curtis and Andrew Valentine) for their in-depth and valuable comments on the paper and hope our responses below address their comments.

Where we have deemed appropriate, we have grouped similar comments from both reviewers together and provided a single response.

Ben Moseley, Tarje Nissen-Meyer & Andrew Markham

Reviewer comments:

Author response:

We can see how our message comes across as though this is a toy problem that cannot be extended to elastic media and that the paper does not consider enough how it advances Solid Earth science.

The message we want to convey is that we are taking the first steps in understanding whether deep learning can aid seismic simulation tasks relevant to the Solid Earth sciences, and in that sense proposing a new research direction.

Whilst we only present methods for "toy" 2D acoustic simulation, and whilst we believe there are challenges to extending our methods to 3D, more heterogeneous media and elastic simulations tasks relevant to the Solid Earth, based on our first steps we believe further research is needed and the avenue could eventually yield practical and useful tools for Solid Earth science. We believe this ought to be done before any decision on whether this approach has the potential to scale up at production level for realistic Earth science applications.

We currently believe that extending our method to 3D is mostly a computational challenge, rather than requiring an entire conceptual recast to our network design. Similarly, our

network structure is readily extendable to elastic simulation; the dimensionality of the network's inputs and outputs just needs to be increased. Whilst further research is needed to understand if these extensions are possible, we believe that our main contribution is to show that deep neural networks can simulate more complex 2D models for which no solutions other than numerically discretized ones exist and not just simple 1D models, which is a positive step towards using deep neural networks in real applications.

Proposed manuscript changes:

We propose to "reframe" the manuscript such that it includes a sort of "manifesto" or roadmap proposing and investigating deep learning for aiding Solid Earth simulation tasks:
1. Add more emphasis in the introduction on practical Solid Earth seismic simulation tasks, their challenges and how deep learning and our method fits into them
2. Add more detail in the discussion section on future research directions which could extend our methods so they can eventually be used in real Solid Earth simulation tasks. We would include new subsections discussing what we see are the main challenges towards this: 1) extending to more heterogeneous models, 2) extension to elastic simulation and 3) extension to 3D simulation.

Reviewer comments:

[Reviewer 1] • What do the authors foresee as the primary application(s) of their approach? The discussion seems to mainly envisage inversion-related use cases. Some comments on this:
1. *The 'fast seismic inversion' approach discussed in Section 3 is essentially a variant of the 'prior sampling' approach discussed in detail by Käufl et al (2016). The key strength and weakness of this as an inversion strategy is that all samples (i.e. the training set) are generated without reference to any observed data. This enables very rapid inference once data becomes avail- able, but it means that most training samples lie far from the observed data and are largely wasted from the perspective of any one inference. The end result is that inferences are considerably less well-constrained than would be possible with posterior sampling (see Fig. 9 of Käufl et al 2016). The bottom line is that prior sampling only seems a worthwhile strategy for (a) problems where time is of the essence, e.g. earthquake early warning, or (b) problems where the 'same' inference task needs to be solved many thousands of times with different data vectors.*
2. *Using the learned model in Monte Carlo simulations seems superficially attractive, but comes with significant caveats. Fundamentally the inference remains entirely based on the information contained within the training dataset, and so all the limitations of prior sampling remain. The random walk would need to be constrained to only generate models compatible with the training data, if results are to be meaningful. Perhaps it would be possible to progressively retrain the learned simulation as the McMC proceeds, to ensure accuracy in relevant parts of the model space: this starts to move towards the Bayesian optimisation approaches discussed in (e.g.) Wang et al (2013). To play devil's advocate: if a problem is too complex to tackle using an McMC approach using physical simulations, can we really be confident that a learned model is sufficiently accurate to yield meaningful results? How big a training set is required to capture the full complexity of the physical problem?*

Some discussion and commentary on these issues, and other potential applications, would be appreciated.

Author response:

We agree that discussing in more detail the potential applications of our method would be valuable, as described in our previous response above. We also think that the specific issues highlighted on inversion in this comment are important to consider and we propose to expand our discussion on the pros and cons of our inversion technique to include these.

Proposed manuscript changes:

1. Add more detail in the discussion section on potential applications (as described in our response above)
2. Discuss in more detail the issues of prior sampling and limitations of the training dataset when discussing the pros and cons of our inversion technique.

Reviewer comments:

[Reviewer 1] • The main 'selling point' of the author's approach is that it enables seismograms to be generated significantly faster than would be possible using 'traditional' forward models. However, this comes with a number of caveats that I think need to be discussed more carefully.
1. *(As above) the author's approach is (I suspect) only effective for models that are sufficiently similar to those in the training set. The numerical forward code does not suffer from this restriction, and can handle complexities that aren't present in the authors' setup (e.g. anisotropy, variable density). How much of a speedup could be achieved by using a numerical code that had been designed with prior knowledge of the characteristics of the authors' training set? Put another way: the speedup could be made to seem even more impressive by using a code designed for a vastly more complex setting (e.g. SPECFEM3D) to build the training set. How fair is the comparison that is being presented?*
2. *The headline speed comparisons ignore the costs of building a training set and then training the neural networks, which are significant. How many simulations does a user need to envisage performing before the author's approach becomes cost-effective overall? I think this is going to be a rather large number. Again, some discussion of the pros and cons of the author's approach would seem desirable.*

[Reviewer 1] • In general, we can make numerical simulations faster by introducing physical approximations. In such cases we typically have some intuition for how those approximations will impact upon results. Learned models offer a speedup without explicit physical approximations, but come with uncertainties that are difficult to quantify rigorously, and which may vary considerably depending on the particular set of inputs chosen. Would the authors like to comment on the pros and cons of the two different strategies for reducing computational costs?

[Reviewer 2] I also agree with most (but not all) of the comments from Reviewer 1 (Andrew Valentine), in particular his comment that a fair comparison for the layered-medium example would be between neural networks and other modelling methods that intrinsically assume a 1D Earth structure rather than full finite difference methods.

<u>Author response:</u>

For the first case where we consider simple layered 1D Earth models we agree that a fairer speed comparison would be against existing numerical methods which intrinsically assume 1D layers. For the 2D faulted Earth models we consider, to the best of our knowledge FD methods are the most efficient tools for these types of models, and therefore we believe this is a fair comparison.

We agree that there is more nuance to the comparison than just speed; for example, a discussion on where our approach break downs and the cost of training dataset generation and training would be useful.

<u>Proposed manuscript changes:</u>

1. For the case of 1D layered Earth models compare our approach (our WaveNet model) against the speed of an existing numerical method which intrinsically assumes 1D layers.
2. Include a discussion of where our approach breaks down and the relative cost of generating the training data and training our models.

<u>Reviewer comments:</u>

[Reviewer 2] 4. The paper appears to have committed the equivalent of an 'inverse crime'. If I understand correctly, the authors have trained networks in the forward and inverse directions using models with a certain parametrisation, and have tested the networks on models of exactly the same parametrization (Reviewer 1 touched on this too). While this would be reasonable if the models were themselves reasonably realistic, in a practical field like Geophysics I think it is necessary to test the networks on examples that lie outside of the range of the training set – not only using different models from those in the training set, but models that are not within the span of the algorithm used to generate the training set (as in the real Earth).

For example, (a) Earp and Curtis (2019, arXiv – https://arxiv.org/abs/1907.00541) and (b) Earp et al. (2019, arXiv – https://arxiv.org/abs/1908.09588) perform (a) probabilistic travel time tomographic imaging, and (b) probabilistic surface wave inversion for averaged shear wave structures with depth, using deep neural networks. The test examples using in both cases are created using a finer parametrization than was used in the network training set – thus the actually structure of the (synthetically) 'true' Earth is not attainable by the networks; nevertheless, they can be used as a useful check of whether in such cases (as in the real Earth) the networks behave sensibly – giving results that are spatial averages in some sense of the 'true' structure. To be clear I do not think that the above references are perfect in this regard and could certainly be improved (e.g., could use even more complex models for tests); nevertheless, the authors could usefully think about such tests for their work as it would strengthen the conclusions.

[Reviewer 1] • In any ML-based approach, the training data is central to the applicability of the method. The author's trained network appears effective for simulating waveforms in models that are generated using the same criteria as were used to make the training set. However, I suspect performance will be significantly worse

for models that have significantly different character. This is something that deserves more discussion than it receives, perhaps with some examples. A particular concern in practical settings may be how an end-user can assess whether their input model is 'sufficiently close' to the training set.

Author response:

Whilst we were careful not to test the performance of the network using the same examples which were used to train the network, we only used examples drawn from the same distribution as the training distribution to test our networks. Because the training distribution only contains simple models we agree with the reviewers that this does not inform us on the performance of the network for models outside of the training set, and furthermore our current networks are likely to perform worse for such models.

For many seismological applications of forward and inverse modelling, we believe that the Earth models used are typically within a known range of parameters and therefore a training set could be constructed which appropriately spans the expected models, however we believe more research is needed in this area. For instance, dozens of published tomographic models could be used to define a base model range for future tomography modelling where thousands or millions of similar model simulations are needed.

Proposed manuscript changes:

1. We propose to test both our models on more realistic Earth models outside of the span of our training distribution and show the degradation observed. We also propose to suggest ways (such as nearest neighbour analysis) which could help a user determine how close an input model is to the training distribution, and potential future research ideas which could quantify the prediction uncertainty.
2. We will add more discussion on potential applications where we think this limitation is and is not permissible.

Reviewer comments:

[Reviewer 2] 2. The introduction is interesting and reviews some of the appropriate material, but is very sparsely justified, and as also stated by Reviewer 1, it does not include many key references. In my view, every sentence of a scientific work must either be a logical deduction from previous text, must have been deduced/proven in another paper, or may be an argument based on the material in another paper; in that latter two cases that paper needs to be discussed and cited. In this paper, the Introduction cites very few references and therefore contains unjustified (in the sense of, 'not justified') statements. Examples include:

One cannot write a paper on using neural networks to perform full waveform inversion (FWI) without citing Roeth and Tarantola (1994 – J. Geophys. Research). How does the FWI part of this paper improve on their work? That is not at all clear. There are many other papers using neural networks for imaging in Geophysics using waveforms or other types of data; you need to read and cite them, and describe how this work advances the field relative to those works. Currently the latter is not clear. The authors make the case for using neural networks for real-time applications – again first steps in this direction have already been taken (see Cao et al., 2019 – Geophysics, for example) and should be discussed.

[Reviewer 1] • Referencing, especially in the introduction, seems rather haphazard. If citations are to be given for broad, well-established topics such as the utility of seismic simulations in reservoir characterisation, I would expect these to be to major review papers or to 'classics': these are going to be most useful for a reader who is unfamiliar with the field. Without intending any criticism of the cited works, this does not really seem to be the case at present. Moreover, the authors' survey of the history of machine learning in geophysics is very short-sighted, ignoring anything more than a couple of years old. There are neural network papers in the geophysical literature from the late 1970s onwards, and it would be nice to see some acknowledgement of this body of work. Valentine & Trampert (2012) is probably the first instance of 'deep learning' in seismology, though the term had not been invented at that point (and we did not have the benefit of modern computational frameworks).

Author response:

We agree that the referencing in the introduction is sparse and that we only review examples of applications of deep learning to geophysics in the last couple of years, and that the paper would be stronger with more discussion of relevant work.

Proposed manuscript changes:

1. Ensure citations on broad, well-established topics in the introduction are major review papers or "classics".
2. Add and discuss more references where they are sparse, for example Roeth and Tarantola (1994 – J. Geophys. Research) when discussing FWI, and Cao et al., 2019 – Geophysics for real time simulation.
3. Include a fuller review of the applications of deep learning in geophysics, and include earlier examples, such as Valentine & Trampert (2012) and Devilee et al., (1999 – J. Geophys. Res).

Reviewer comments:

[Reviewer 2] 3. The authors promote the fact that they use 'deep learning', and their application certainly fits into that category. However, they must at least discuss why this is a positive feature of the method, and cite previous Geophysical applications of deep learning to support that discussion. Deep learning is usually defined to be the use of 4 or more layers within a neural network. While I agree with Reviewer 1 that his previous work (Valentine and Trampert) was an example of deep learning, the first that I know of in Geophysics was in fact Devilee et al., (1999 – J. Geophys. Res) – which came from the same university.

In my view there is therefore nothing new about the concept of deep learning: we were using it in Geophysics in the '90's. What has changed is the extent to which depth can be used to impose useful structure on networks (as the authors themselves have done in this manuscript – their Figure 9, and also in the paper cited by Reviewer 1); also the number of parameters that can now be used (the width of each layer) has increased hugely. In fact the number of parameters in the authors' application is relatively modest compared to some in machine learning literature, but is certainly comparable to other recent studies in Geophysics; the structure that the authors impose is both sensible and clearly useful in order to help to obtain stable results. These things should be discussed.

Author response:

We agree that deep learning is not a new technique in Geophysics and that two of the enabling recent advancements in this field are the ability to train deeper models with many more parameters. We agree the manuscript does not explicitly make this distinction and we propose to make this clearer.

Proposed manuscript changes:

1. Explicitly acknowledge that deep learning concepts are not new in geophysics and have been used in the past, and explain that, among other factors, such as the availability of more powerful hardware, advancements in training deep models with more parameters have enabled this work.