Nonlin. Processes Geophys. Discuss.,
https://doi.org/10.5194/npg-2018-11-RC1, 2018

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

Interactive comment on "Inverting Rayleigh surface wave velocities for crustal thickness in eastern Tibet and the western Yangtze craton based on deep learning neural networks" *by* Xianqiong Cheng et al.

H. Matchette-Downes (Referee)

hrmd@mit.edu

Received and published: 11 May 2018

1 General comments

To the best of my knowledge, the authors are the first to attempt to use deep neural networks (as opposed to shallow neural networks) to invert observations of surface wave dispersion for crust thickness (or any similar problem in seismology). However, major revisions are necessary to demonstrate that the method is working as intended,

Printer-friendly version

Discussion paper

and to show that it is an improvement over shallow neural networks (e.g. Meier et al., 2007), which are simpler. The manuscript also has some misleading statements and omissions. I suggest revisions below.

2 Specific comments

2.1 Inclusion of noise in training data

The authors do not mention whether or not the synthetic training data contain noise. A neural network trained on noise-free synthetic data will perform very poorly on real data containing noise (e.g. Meier et al., 2007, figure 8b). If noise was included in the training data, the authors should describe this. Otherwise, it should be included.

2.2 Conversion from group velocity to phase velocity

The authors calculate group velocity from a published phase velocity map using the standard formula (their equation 4). However, including both phase and group velocity will only add new information if the phase and group velocity are measured independently (as is commonly the case). Therefore it is misleading to include the calculated group velocity in this paper. The group velocity data should be removed from the study or replaced by group velocity data measured independently. (Generally phase velocity is more sensitive to deeper structure so it is easier to infer deep structure from phase velocity measurements.)

2.3 Benefit of deep neural network versus shallow neural network

A deep neural network is one with more than one hidden layer, whereas a shallow neural network has just one hidden layer. The additional complication of using a deep neural network is justified if the mapping has a hierarchical structure. For example, in image processing, it is common to move from the more elementary aspects of the input data (e.g. the values of the individual pixels) to intermediate parts (such as the distribution of edges) and finally to the most abstract aspects (such as the subject of the image). While it is undoubtedly true that the Earth has a hierarchical structure, ranging from individual grains to entire continents, the authors do not demonstrate that the dispersion data contain sufficiently complicated information to justify a deep neural network. The paper does not currently demonstrate that a deep neural network offers any improvement over a shallow neural network, such as that used by Meier et al. (2007). A comparison should be given.

2.4 Non-unique solutions

The authors focus on the non-linearity of the inverse problem, but they do not mention that it is also non-unique. Conventional optimisation of a neural network can lead to meaningless outputs for a non-unique mapping, as shown in figure 3b of Meier et al. (2007). Ideally, the method should be changed to solve for a probability distribution, for example using histogram or median networks (Devilee et al., 1999) or a mixture density network (Meier et al., 2007). Otherwise, the authors should attempt to quantify the range of non-uniqueness, or at least mention it in their discussion.

2.5 Unattributed quotations

Some explanatory sections are taken verbatim from other work, for example the paragraph beginning at 3.2:19 is identical to the second paragraph of section 3 of de Wit et al. (2014). These sections should be attributed, and either paraphrased or written in quotation marks.

2.6 Meaning of ‘data-driven’

It is misleading to say that the method ‘data-driven’ (e.g. lines 1:9–11). The inversion is model-driven; it is trained using a large number of synthetic data which are generated using a known forward mapping (in this case, the calculation of dispersion by normal mode summation). The role of the neural network is to approximate the inverse relation apparent in the synthetic dataset. The description ‘data-driven’ is appropriate when the forward mapping is not known (or not used). An example would be speech processing, where the meaning of a word cannot be calculated from its audio waveform.

2.7 Lateral resolution of crust thickness

Figures 7 and 8 show a comparison of the crust thickness model in this study with the crust thickness model in Xie et al. (2013). Although the two models are based on the same data, the result in this study appears to resolve much finer features. The authors should explain how this higher resolution is achieved and whether it is justified.

3 Corrections to the writing

There are some errors in the writing, but I have not listed them in detail, in the expectation that the body of the text will change.

4 References

Devilee, Curtis & Roy-Chowdhury (1999), https://doi.org/10.1029/1999JB900273
Meier, Curtis & Trampert (2007), https://doi.org/10.1111/j.1365-246X.2007.03373.x
de Wit, Valentine & Trampert (2014), https://doi.org/10.1016/j.pepi.2014.09.004
Xie et al. (2014), https://doi.org/10.1002/jgrb.50296

Printer-friendly version

Discussion paper