Review of : Inverting Rayleigh surface wave velocities for eastern Tibet and western Yangtze craton crustal thickness based on deep learning neural networks

Xian-Qiong Cheng[1] , Qi-He Liu[2], Ping-Ping Li[1]

I enjoyed reading the paper, and thought it was a very interesting new application of a technique. Neural networks is not something I know much about, but I thought it was interesting that the authors applied this non-linear technique to the non-linear problem of crustal thickness. In my opinion, it is definitely something that should be published in this journal.

However, there are several points which should be addressed:

Major:

1) I did not understand how the authors built their model of crustal thickness from the data. For example, did they build some layers, and add them up for crustal thickness. What information allowed them to decide they were in the crust or mantle? This was not discussed at all. This discussion needs a figure.

2) p4, 27-29 I wondered whether the choice of PREM as a starting point for training the models was a good choice. PREM is very different from much of the Tibetan plateau and surrounding areas. On the other hand, 100,000 synthetic models sounds pretty impressive! I think this section needs a figure showing the models as a depth profile (either just each model plotted on top of each other - or some kind of probabilistic model). Then the authors should either justify their use of PREM as a good choice - or include some other models which take into account larger crustal thickness.

3) I was not clear during the paper whether crustal thickness really meant thickness, or whether it was depth below sea level. Since the plateau is at 5 km above sea level, this is quite important.

4) Topographic effects. I think the authors should mention whether they are think there are errors associated with the surface topography - and if anything can be done about these errors.

Minor points:

Title: I recommend:
Inverting Rayleigh surface wave velocities for crustal thickness in eastern Tibet and the western

Yangtze craton based on deep learning neural networks

p1, abstract 'Based on test errors 15 and misfits with other crustal thickness models, we select the optimized one as crustal thickness for study areas. ' There is no obvious reason to assume that the other models are any better than yours! So I wasn't sure why you would want to choose your favourite model based on a comparison (This comment doesn't apply to the later section where you compare your models with other models - I thought that was interesting).

p1, 31 The Moho (Mohorovičić discontinuity) is a seismic discontinuity, and may not even be present. It is not the same as crustal thickness (although it has a strong correspondence with it).

p1, 33 'has significant effects on fundamental model surface wave' - you could reference other papers as well here

p1, 36 adjants => surrounding areas

p1, 42 defaults => defects

p1, 48-50 - be careful with the wording. For example Shapiro and Ritzwoller 2002 is shear-wave velocity model - not directly about crustal thickness.

p2, 18, traditional shallow neural networks => please explain what these are first

p2, 29 deep learning neural networks. Moreover, our deep learning neural networks train on vast synthetic models. => please explain

p2, 31, 'Lastly, our results show changes of the number of neurons in each layer have little influence on test errors when the numbers of network layer achieve six and test errors are about 2.5e-6 ' => this didn't really belong here, before you have explained about layers

p2, 57 'shallow neural networks' => this still hasn't been explained

p3, 12, Add references

p4, 28, PREM - please explain what PREM and Mineos are, and reference them.

p4, 39 'Since a larger part of the signal is affected by the crustal structure, combination two types of data will constrain crustal thickness better in the presence of noise.' Reference this statement if you can

p5, 3 Based on Rayleigh wave
phase velocity from ambient noise(Xie et.al,2013), we compute corresponding group velocity

This sentence didn't make sense the first time I read it - I didn't understand that you were using *data* from Xie 2013)

p5, 18 After trying many times, we find the proportion of training data set to test one is 3:1 is reasonable. A figure would be helpful here.

p5, Please explain this table further. Which is a good result and why, and give a bit more explanation about the table headings.

p5, 28 as shown in table 1 shown in ※. What is this symbol?? It is referred to several times.

p5, 25-35 This table mixes method, results and discussion. It would be better to separate them..

Fig 3 - Indicate that the left side is your work

p5 - Add a new figure showing some velocity profiles across the region - as explained above I cannot see how you have arrived at Fig 3.

p7, 49-53 - Interesting point, but explain a bit more

General:

English - There are some problems with the English text which make the paper quite difficult to read. A good revision of the English would be helpful.

I found it difficult to follow your explanation of the development/background of neural networks. More logical explanation - and not introducing terms before they are discussed would be good. (The description of your own method and its application was good and I was able to follow that).

I hope you find these comments useful.

Ceri Nunn

Ludwig-Maximilians-Univesität, Munich

p