## Author Reply

Authors Name: Xian-QiongCheng , Qi-He Liu, Ping-Ping Li
Paper Name: Inverting Rayleigh surface wave velocities for eastern Tibet and western Yangtze craton crustal thickness based on deep learning neural networks

Revision Date: 11 / 15/ 2016

Summary of Responses:

We thank Ceri Nunnrefereefor her working for this paper, who has given many good suggestions, which we are incorporated in this revised work.

Below are the responses of work we have done.

| **Comments and Suggestions** | **Response** | **Page Reference ( Origin)** | **Page Referred ( New)** |
|---|---|---|---|
| **Major points** | | | |
| (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. | According to prior ranges for the various parameters in our model given intables A.2–A.4.of de Wit et al.(2014), we determine earth model parameters. For example , since the range of moho depth is between 20km and 70km, we can set the moho depth between this range randomly, based on the formula h=h1+(h2-h1)×rand, in which h1=20km,h2=70km and rand is a random data between 0 and 1. Which is same to other parameters. So we can attain 100,000 or even more synthetic models. | | |

| | | | |
|---|---|---|---|
| (2). p4, 27-29 I wondered whether the choice of PREM as a starting point for training themodels was a good choice. PREM is very different from much of the Tibetan plateau and surrounding areas. On the other hand, 100,000 synthetic modelssounds 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 probabilisticmodel). Then the authors should either justify their use of PREM as a good choice – orinclude some other models which take into account larger crustal thickness. | As stated in the paper(Meier U,2007), the neural network approach for solving inverse problems is best summarized by three major steps: (1) proceed by randomly sampling the model space and solve the forward problem for all visited models (i.e. compute phase and group velocities for the sampled radially symmetric earth models using normal mode theory). This results in a collection of earth models and corresponding phase and group velocities (called the training data set). (2) design a neural network structure that can accept phase and group velocities as input and compute the earth model as output, then use the training data to train the network (i.e. change the parameters of the network such that the network output represents the desired output, the earth model).<br>(3) Once the network is trained it represents the nonlinear inverse mapping from phase and group velocities to earth structure. For any observed dispersion curve the trained network will give an output that is close to the “real earth”.<br><br>In the first stage, when we trainedthe neural network, we focused on the relationship between earth models and corresponding phase and group velocities. furthermore, we show this relationship by neural network structure(that are type of activation function, learning rate, zero masked fraction, non-sparsity penalty, number of epochs; batchsize ).So we require the inputting synthetic earth model should be coincide with the real earth roughly. Eventually, we can select PREM as a starting point for training neural network. | | |
| (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. | Crustal thickness in this paper meant depth below sea level | | |
| (4). Topographic effects. I think the authors should mention whether they are think there are errors associated with the surface | In this paper we did not think about topographic effects. | | |

| | | | |
|---|---|---|---|
| topography - and if anything can be done aboutthese 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 | We rename our title as referee suggested | | |
| 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 assumethat the other models are any better than yours! So Iwasn't sure why you would want to choose yourfavourite model based on a comparison (This comment doesn't apply to the later sectionwhere you compare your models with other models - I thought that was interesting). | Other models such as CRUST2.0, CUB2 ,ZJS are of high resolution with good data coverage based on other methods such as receiver function and so on. Especially these models are consistency with many other researches. | | |
| 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 | We change Moho discontinuity to crustal thickness | | |

| | | | |
|---|---|---|---|
| correspondence with it). | | | |
| p1, 33 'has significant effects on fundamental model surface wave' - you could reference otherpapers as well here | We added the corresponding reference | Page 1, line 33 | Page 1, line 31 |
| p1, 36 adjants =>surrounding areas | We corrected the corresponding description as referee suggested | Page 1, line 36 | Page 1, line 34 |
| p1, 42 defaults => defects | We corrected the corresponding description as referee suggested | Page 1, line 42 | Page 1, line 40 |
| p1, 48-50 - be careful with the wording. For example Shapiro and Ritzwoller 2002 is shear-wave velocity model - not directly about crustal thickness. | We corrected the corresponding reference | Page 1, lines 48-50 | Page 1, lines 46 |
| p2, 18, traditional shallow neural networks => please explain what these are first | We explained traditional shallow neural networks have has less number of hidden layers. | Page2, line 18 | Page 2, line 20 |
| p2, 29 deep learning neural networks. Moreover, our deep learning neural networks train on vast synthetic models. => please explain | Deep learning neural networks have multiple hidden layers.<br><br>Moreover, we use vast synthetic models as training data set to train our deep learning neural networks. | Page2, line 29 | Page2, line 31 |
| 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 | "Layers" in neural network mean some neurons(computational units) in the same position. For example, in figure 1 the leftmost neurons consist of input layer, and rightmost neurons consist of output layer. | | |
| p2, 57 'shallow neural networks' => this still hasn't been explained | Shallow neural networks have been explained on P2,20 | | |
| p3, 12, Add references | We added the corresponding references | Page 3, line 12 | Page3, line 14 |

| | | | |
|---|---|---|---|
| p4, 28, PREM - please explain what PREM and Mineos are, and reference them | We explained PREM and Mineos and reference them | Page 4, line 28 | Page4, line 30 |
| p4, 39 'Since a larger part of the signal is affected by the crustal structure, combination twotypes of data will constrain crustal thickness better in the presence of noise.' Reference this statement if you can | We added a reference to this statement | Page 4, line 39 | Page 5, line 2 |
| p5, 3 Based on Rayleigh wave phase velocity from ambient noise(Xie et.al,2013), we compute corresponding group velocity<br>This sentence didn't make sense the first time I read it - I didn't understand that you were using data from Xie 2013) | This meant Xie et.al(2013) gave Rayleigh wave phase velocity, and we compute corresponding group velocity according to formula (4) in this paper | | |
| p5, 18 After trying many times, we find the proportion of training data set to test one is 3:1 is reasonable.<br>A figure would be helpful here. | We gave a figure (Figure5)in the revised paper to demonstrate the proportion of training data set to test one is 3:1 is reasonable | | Page 6 , figure 5 |
| p5, Please explain this table further. Which is a good result and why, and give a bit more explanation about the table headings | The parameters for neural network shown in ※in table 1 are good results, since test error is relatively low and correlation coefficient of our this model with crustal thickness models from other research is relatively high. | | |
| p5, 28 as shown in table 1 shown in ※. What is this symbol?? It is referred to several times. | this symbol※ meant various parameters for deep learning neural network (called neural network structure)we taken in this paper | | |
| p5, 25-35 This table mixes method, results and discussion. It | We separate figure6,figure7 and figure8 | Page 5, lines 25-35 | Page 6, lines 12-15 |

| | | | |
|---|---|---|---|
| would be better to separate them. | | | |
| Fig 3 - Indicate that the left side is your work | We indicate figure 6 is our work | Figure 3 | Figure 6 |
| p5 - Add a new figure showing some velocity profiles across the region - as explained above Icannot see how you have arrived at Fig 3. | We added the phase and group velocities profiles shown in figure 3 and figure 4 | | Page 5, Figure 3-4 |
| p7, 49-53 - Interesting point, but explain a bit more | Overfit in neural network mean our hypothesis may fit the training set very well, but fail to generalize to new example, which means test errors may be low, but inversion errors may be high. | Page 7, line 49-53 | Page 8, line 39-42 |
| | | | |
| | | | |
| | | | |