Author Reply

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

Revision Date: 12 /9/ 2018

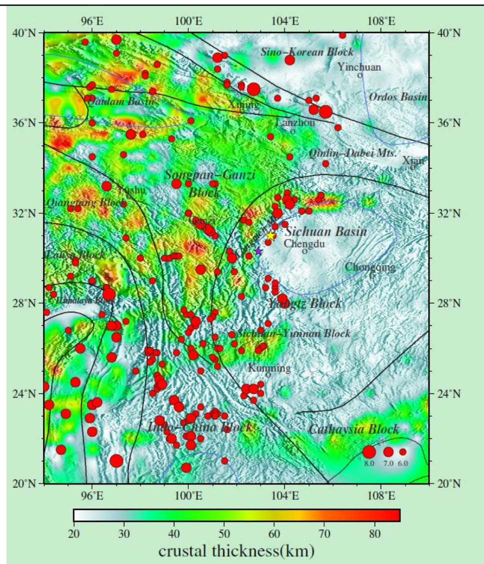
Summary of Responses:

We thank the referee for his 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.

For refree1:

Comments and Suggestions	Response
1) **Grammar mistakes**	Thanks to the reviewer for the suggestions. We check the English sentence by sentence and upload revised manuscript.
2) Is the new method indeed better?	In order to demonstrate if these anomalies are persistent, are mere accidents, or are artifacts of the inversion, we refer to the result of research in the same region from Wang(2010), who attained the crustal thickness estimated by the H-k stacking method based on the broad band tele-seismic data recorded at 132 seismic stations in Longmen mountains and adjacent regions(26°~35°N,98°~109°E)(Figure 9 in the article). Our result reveals similar details with Wang(2010) and indicates these anomalies are persistent.
3) Is the authors’ model indeed better than Shapiro & Ritzwoller (2002)?	Taking the Monte Carlo method (Hansen, 2013) and using four processors for only 1000 iterations, it takes three weeks to invert the Xie (2013) data set to the crust thickness of the same region , and the result shown below indicate that overall agreement between our and this result. Although this result shows singular values in some places , maybe the result is high resolution after many more iterations using Monte Carlo method. However, in our approach, our training process took less than 6 hours and the inversion process took a few minutes. Compared to our method, Monte Carlo method is computational expense.



For refree2:

Comments and Suggestions	Response
A significant result of this manuscript would be if the sSAE neural network achieved better results than a simple 'shallow' neural network (e.g. Meier et al., 2007). However, achieving lower test errors is not sufficient to show that the complicated neural network is better than the simple one. More work is needed to demonstrate that the sSAE is not over-fitting the training data. For example, it is necessary to state the total number of parameters in the neural network compared to the number of training data points. If the total number of parameters is large compared to the number of training data, then further work is necessary to check for over-fitting, such as regularisation.	Thanks to the reviewer for the suggestions. We try our best to avoid overfitting based on neural network development. Our model uses 380,000 data as the training set and 120,000 data as the test set. The two data are separated and the iteration is stopped when the error of the test set is not falling. Therefore, the early stop mechanism we used in this article to avoid over-fitting problems. In addition, in the fine tuning of the model, a second-order norm regularization method is also used to avoid overfitting. On the other hand, the number of our model weight parameter is 30455, the number of bias parameters is 386, a total of 30841 parameters, and the number of parameters is much smaller than the number of training data.