Nonlin. Processes Geophys. Discuss., https://doi.org/10.5194/npg-2018-11
Manuscript under review for journal Nonlin. Processes Geophys.
Discussion started: 9 March 2018

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

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

Xianqiong Cheng[1],Qihe Liu[2], Pingping Li[1], Yuan Liu[1]

[1] College of Geophysics, Chengdu University of Technology, Chengdu , P.R. China
[2] School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu , P.R. China

Correspondence to: Xian-Qiong Cheng(chxq@cdut.edu.cn)

Abstract. Crustal thickness is an important factor affecting lithosphere structure and deep geodynamics. In this paper, we propose to apply deep learning neural networks called stacked sparse auto-encoder to obtain crustal thickness for eastern Tibet and western Yangtze craton. Firstly taking phase and group velocities of Rayleigh surface wave simultaneously as input and theoretical crustal thickness as output, we construct twelve deep neural networks trained by 70,000 and tested by 30,000 theoretical models. We then invert observed phase and group velocities by these twelve neural networks. Based on test errors and misfits with other crustal thickness models, we select the optimized one as crustal thickness for study areas. Compared with other ways detected crustal thickness such as seismic wave reflection and receiver function, we adopt a new way for inversion of earth model parameters, and realize that deep learning neural network based on data driven with the highly nonlinear mapping ability can be widely used by geophysical inversion method, and our result has good agreement with high-resolution crustal thickness models. We conclude that deep learning neural network is a promising, efficient and believable tool for geophysical inversion.

Keywords: Crustal thickness; Phase and group velocities; Surface wave; Stacked sparse auto-encoder;Deep learning ; Neural network

1 Introduction

Discontinuity between crust and mantle called moho discontinuity varying greatly over small length scales is an important factor for geodynamics including crustal evolution, tectonic activities, in addition to the correcting gravity for the crustal effects, seismic tomography and geothermal modeling. Many researches focus on obtaining depth of moho discontinuity called crustal thickness by various data and different methods.

Often crustal thickness can be inverted from many types of data, such as inverting deep seismic sounding profile for Chinese continent to get crustal thickness(Zeng et al.,1995), inverting satellite gravity data to get whole global crust and lithospheric thickness(Fang et al.,1999), inverting Bouguer gravity and topography data to get crustal thickness for China and its surrounding areas(Huang et al.,2008;Guo et al., 2012),inverting receiver function to get crustal thickness and Possion's ratio for Chinese continent (Chen et al.,2010;Zhu et al.,2012;Xu et al.,2007;). Especially, a newest crust model called crust1.0 at $1^{o}\times1^{o}$(Laske et al.,2013;Stolket al., 2013) are based on refraction and reflection seismology as well as receiver function studies. Besides these data related to crustal thickness mentioned above, crust thickness has significant effects on fundamental mode surface waves (Meier et al.,2007,Grad et al.,2009). Dispersion characteristic of surface wave provides a powerful tool to research structure of crust and upper mantle (Legendre, C. P. et al.,2015). So far phase and group velocity measurements of fundamental mode surface waves are most commonly used to constrain shear-velocity structure in the crust and upper mantle on a global scale (Zhou et al. 2006;Shapiro &Ritzwoller ,2002) or on regional scale (Zhu et al.,2002;Zhang et al.,2011;Yi et al.,2008), also the newly developed ambient noise surface wave tomography has been used to constrain shear-velocity structure(Sun et al.,2010;Yaoet al.,2006;Zheng et al.,2008;Zhou et al.,2012),while a few works to invert fundamental mode surface wave data for global or regional crustal thickness and to present a global or regional crustal thickness model(Devile et al.,1999;Meier et al.,2007; Das &Nolet 2001; Lebedev et al.,2013). As periods and method measured differently between group velocity and phase velocity, which the probing depths are different and measure error are largely independent, the

Nonlin. Processes Geophys. Discuss., https://doi.org/10.5194/npg-2018-11
Manuscript under review for journal Nonlin. Processes Geophys.
Discussion started: 9 March 2018

simultaneous inversion of group velocity and phase velocity is substantially better than the use of either alone (Shapiro & Ritzwoller,2002).

There are several inverse methods to get crustal thickness, and these methods can be broadly classified into two classes: (1) model-driven methods and (2) data-driven methods. For model-driven methods, researchers mainly consider physical relation between earth parameters space and data space to calculate inverse function. Most methods based on model-driven treat crustal thickness inversion as a linear problem, and most importantly, their results are heavily depended on initial earth model. In contrast to model-driven methods, another kind of fully non-linear data-driven method called neural network to put forward to get crustal thickness (Devile et al.,1999; Meier et al.,2007). Neural network with the highly nonlinear mapping ability is widely used by geophysical inverse method based on data-driven, which apply the actual seismic, logging data and its attribute to predict earth parameters. Compared with model-driven inversion, data-driven inversion maps and predicts an arbitrary nonlinear relationship fast and accurately without considering about physical relations between earth model parameters and data space. As such, neural networks can be very useful in situations where the forward relation is known, but the inverse mapping is unknown or difficult to establish by more conventional analytical or numerical methods(de Wit et al.,2013). So the target of neural network inversion is to find the mapping from a set of training data. Neural networks have been widely used in different geophysical applications well summarized by van der Baan &Jutten (2000) such as in electrical impedance tomography(Lampinen&Vehtari ,2001), in seismic processing including trace editing, travel time picking, horizon tracking, and velocity analysis. Devilee et al.(1999) were the first to use a neural network to invert surface wave velocities for Eurasian crustal thickness in a fully non-linear and probabilistic manner. Meier et al.(2007) further develop the methods of Devilee et al. (1999), then invert surface wave data for global crustal thickness on a $2^\circ \times 2^\circ$ grid globally using a neural network.

As seismology points out that there are many factors affect phase and group velocity, inverting phase and group velocity for discontinuities within the earth forms a non-linear inverse problem (Meier et al.,2007). Because of strong non-linear relations between crust thickness and surface wave dispersion, we cannot treat it with a linear inverse problem as Montagner&Jobert (1988) stated. Although shallow neural network with less number of hidden layers, can present nonlinear inverse function, it maybe cannot learn or approximate the true inverse function well when the true inverse function is too complicated. In contrast, deep learning neural network can overcome this defect since it has powerful representation abilities and can discover intricate structures in large data sets, because it take use of the back-propagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer (LeCunet.al.,2015).

In this paper, considering the advantages and characteristics of deep learning neural network, a new fast inverse method based on data-driven, called deep stacked Sparse Auto-encoders (sSAE) neural network is introduced to solve the nonlinear geophysical inverse problems. We focus on deep learning neural networks to solve the non-linear inverse problem, and then apply them to retrieve the crustal thickness for eastern Tibet and western Yangtze craton from newest and high-resolution phase and group velocity maps. Based on normal mode theory we compute phase and group velocities for the sampled radially symmetric earth models to generate 100,000 theoretical models. Firstly taking theoretical phase and group velocities of Rayleigh surface wave simultaneously as inputs and corresponding theoretical crustal thickness as outputs, we construct twelve deep neural networks trained by 70,000 and tested by 30,000 theoretical models. We then invert observed phase and group velocities by these twelve neural networks. Based on test errors and misfits with other crustal thickness models, we select the optimized one as crustal thickness for study areas.

To the best of our knowledge, we are the first to introduce deep learning neural networks to learn and invert crustal thickness, and our result reveals that crustal thickness is strong nonlinear with respect to phase and group velocity. The merits of our methods include: Firstly, since deep learning neural networks can represent complex functions, it is possible to learn the crustal thickness inverse function precisely. Secondly, inverse mapping based on neural network is of high efficiency because new observations can be inverted instantaneously once well-trained deep learning neural networks with multiple hidden layers are constructed. Moreover, our deep learning neural networks are trained on vast synthetic models. 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, which indicates deep learning neural networks are robust to neural network structures with suitable layers. In what follows, we first give a short introduction to deep learning neural networks.

2 Deep Learning Neural Networks

Nonlin. Processes Geophys. Discuss., https://doi.org/10.5194/npg-2018-11
Manuscript under review for journal Nonlin. Processes Geophys.
Discussion started: 9 March 2018

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

In geophysics the true inverse function is usually a very complicated one between data space and model space. Traditional linear inverse methods treating the true inverse function as linear one can resolve linear relation problems. However, they depend on physical relation between two parameter spaces and initial earth models. Neural network has its origins in attempts to find mathematical representations of information processing in biological systems (Bishop ,1995). The more deep strength of Artificial Neural Networks (ANNs) is, the more capabilities learn to infer complex, non-linear, underlying relationships without any a priori knowledge of the model(Bengio,2009). Shallow neural network has gained in popularity in geophysics last decade and has been applied successfully to a variety of problems such as well-log, interpretation of seismic data, geophysical inversion, etc. Although shallow neural network can present nonlinear inverse function, it can only learn the relatively simple inverse function. In contrast, Many research results indicate that deep learning neural network has powerful representation ability and can apply a big geophysical observable data to learn and approximate the complicated inverse function well[Lecun et al.,2015 Bengio et al.,2006; Liu et al.,2015].

Based on the analysis above, we design deep learning neural network to obtain crustal thickness for eastern Tibet and western Yangtze craton. Compared with shallow neural networks, deep learning neural network allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction and can learn complex functions. The essence of deep learning is building an artificial neural network with deep structures to simulate the analysis and interpretation process of human brain for data such as image, speech, text, etc. However, many research results suggest that gradient-based training of a deep neural network gets stuck in apparent local minima, which leads to poor results in practice (Bengio, 2009). Fortunately, the greedy layer-wise training algorithm proposed by Hinton et.al 2006 overcomes the optimization difficulty of deep networks effectively. The training processing of deep neural networks is divided into two steps. Firstly, unsupervised learning methods are employed to pre-train each layer parameters with the output of the previous layer as input, giving rise to initialize parameter values. After that, the gradient-based method is used to finely tune the whole neural network parameter values with respect to a supervised learning criterion as usual. The advantage of the unsupervised pre-training method at each layer can help guide the parameters of that layer towards better regions in parameter space(Bengio,2009).There are multiple types of deep learning neural network, such as convolutional neural networks, deep belief net and stacked Sparse Auto-encoders(sSAE). sSAE works very well in learning useful high-level feature for better representation of input raw data. Since sSAE learning algorithm can automatically learn even better feature representations than the hand-engineered ones, sSAE is used widely in many domains such as computer vision, audio processing, and natural language processing[Hinton,2006; Deng,J et al.,2013]. Similar to these problems, we need extract earth feature representation from dispersion of surface wave. Here we introduce Sparse Auto-encoder briefly, and detailed description of the network training method is given by Liu et al.(2015).

The structure of sSAE is stacked by sparse auto-encoders to extract abstract features. A typical Sparse Auto-Encoder (SAE) can be seen as a neural network with three layers, as shown in Figure 1, including one input layer, one hidden layer, and one output layer. The input vector and the output vector are denoted by v and $\hat{v}$, respectively. The matrix W is associated with the connection between the input layer and the hidden layer. Similarly, the matrix $\widehat{W}$ connects the hidden layer to the output layer. The vector b and $\hat{b}$ are the bias vectors associated with the units in the hidden layer and the output layer, respectively. The SAE is trained to encode the input vector v into some representation so that the input can be reconstructed from that representation. Let $f(x)$ denote the activation function, and the activation vector of the hidden layer then is calculated (with an encoder) as:

$$z=f(Wv+b), \qquad (1)$$

where z is the encoding result and some representation for the input v. The representation z, or code is then mapped back (with a decoder) into a construction $\hat{v}$ of the same shape as v. The mapping happens through a similar transformation, e.g.:

$$\hat{v} = f(\widehat{W}z + \hat{b}) \qquad (2)$$

Nonlin. Processes Geophys. Discuss., https://doi.org/10.5194/npg-2018-11
Manuscript under review for journal Nonlin. Processes Geophys.
Discussion started: 9 March 2018

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

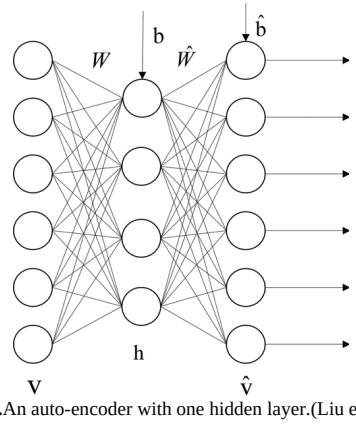

Figure 1. An auto-encoder with one hidden layer.(Liu etal.,2015)

SAE is an unsupervised learning algorithm which sets the target values to be equal to the inputs and constrain output of hidden layer which are near to zero and most hidden layer are inactive, the cost function is expressed as:

$$J_{sparse}(W,b) = J(W,b) + \beta \sum_{j=1}^{S_2} \rho \log \frac{\rho}{\hat{\rho}_j} + (1-\rho)\log \frac{1-\rho}{1-\hat{\rho}_j} \quad (3)$$

Here $J(W,b)$ is cost function without sparsity constrain, β controls the weight of the sparsity penalty term,S_2 is the number of neurons in the hidden layer, and the index j is summing over the hidden units in our network.$\hat{\rho}_j$ is the average activation of hidden unit j, ρ is a sparsity parameter, typically a small value close to zero.

Further, a stacked Sparse Auto-Encoder (sSAE) is a neural network consisting of multiple layers of SAE in which SAE are stacked to form a deep neural network by feeding the representation of the SAE found on the layer below as input to the current layer. Using unsupervised pre-training methods, each layer is trained as sSAE by minimizing the error in reconstructing its input which is the output code of the previous layer. After all layers are pre-trained, we add a logistic regression layer on top of the network, and then train the entire network by minimizing prediction error as we would train a traditional neural network. For example, a sSAE with two hidden layers is shown in Figure 2. This sSAE is composed of two SAEs. The first SAE consists of the input layer and the first hidden layer, and the representation or code of the input v is $h_1 = f(W_1 v + b_1)$. The second SAE comprises of two hidden layers, and the code of h_1 is $h_2 = f(W_2 h_1 + b_2)$. Each SAE is added to a decoder layer as shown in Figure 1, and we can then employ unsupervised pre-training methods to train each SAE by expression (1). Finally, the matrixW_1,W_2,bias vector b_1 and $b_1$are initialized. We then apply supervised fine-tuning methods to train entire network. Since our aim is calculating crustal thickness and this is a regression problem, we attach a layer connected fully with last layer of the encoder part (the matrix W_s). After that, we train this network as done in a traditional neural network.

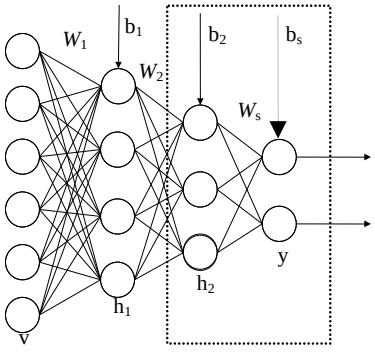

Figure 2. Stacked Sparse Auto-Encoder with two hidden layers.

Nonlin. Processes Geophys. Discuss., https://doi.org/10.5194/npg-2018-11
Manuscript under review for journal Nonlin. Processes Geophys.
Discussion started: 9 March 2018

3 Inverting surface wave data for crustal thickness

As Meier et al. (2007) demonstrated that the neural network approach for solving inverse problems is best summarized by three major steps as shown in Figure 3: (1) forward problem. In this stage we proceed by randomly sampling the model space and solve the forward problem for all visited models based on seismic wave normal mode theory. (2) designing a neural network structure. In this stage taking phase and group velocities as inputs and theoretical crustal thickness as outputs we train the deep learning neural networks and get an optimized one. (3) inverse problem. Base on trained networks we invert crustal thickness from observed phase and group velocities.

In what follows we show how to train a sSAE deep learning neural networks to model surface wave dispersion based on synthetic seismogram, then invert dispersion curves based on the trained networks. Finally we compare our crustal model with other crustal thickness models, and discuss the geodynamic consequences implied by our model.

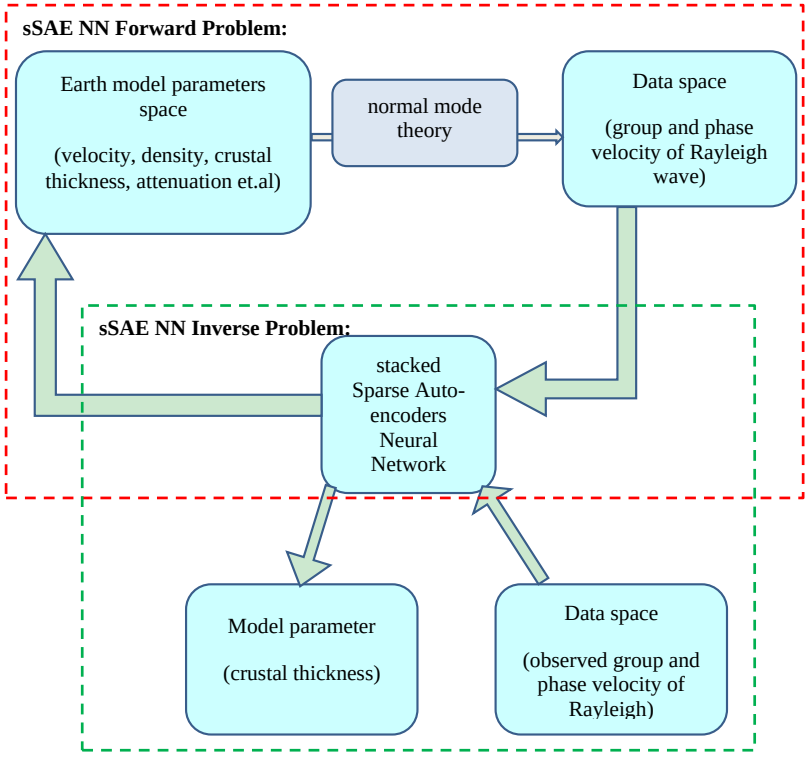

Figure 3. Crustal thickness inversion based on sSAE neural network composed of two parts: Forward Problem and Inverse Problem.

3.1 data preparation

We closely follow the model parametrization methodology outlined in de Wit et al. (2014), which is based on the Preliminary Reference Earth Model(PREM, Dziewonski and Anderson,1981)and is parameterized on a discrete set of 185 grid points used by Mineos package(Masters et al., 2014). In addition, these models we have got show no correlations between physical parameters such as velocity, density, η and attenuation profiles. As the model parametrization method mentioned above, we generate 100,000 synthetic models based on the 1-D reference models PREM, which are randomly drawn from the prior model distribution, also prior ranges for the various parameters in our model are given in tables A.2–A.4. of de Wit et al.(2014).We use the Mineos package to compute phase and group velocity for fundamental mode Rayleigh waves for all 100,000 synthetic 1-D earth models.

As for observation data used in stage of inversion below, group and phase velocities carry the

Nonlin. Processes Geophys. Discuss., https://doi.org/10.5194/npg-2018-11
Manuscript under review for journal Nonlin. Processes Geophys.
Discussion started: 9 March 2018

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

same information, although group velocities are more sensitive to the shallow structure. 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(Devilee et al.,1999). The two are related by

$$U(T) = \frac{c(T)}{1+\frac{T}{c(T)}\frac{dc(T)}{dT}} \qquad (4)$$

Where U denotes group velocity, C denotes phase velocity and T is period. Based on Rayleigh wave phase velocity from ambient noise(Xie et.al,2013) shown in Figure 4 averaged from 10 to 35mHz, we compute corresponding group velocity according (4) shown in Figure 5 averaged from 10 to 30mHz.

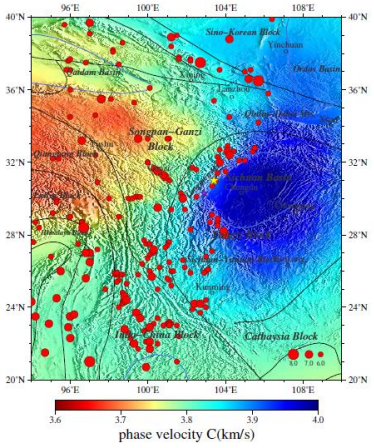

Figure 4. Averaged phase velocity of western Yangtze craton(Xie et al.,2013) from 10 to 35mHz.The black lines in the figure show structure lines. The blue lines show boundaries of sedimentary basins. The red dots show seismic events in this region from 1975 to 2015, and size of dot demonstrates size of magnitude from Ms 6.0 to Ms 8.0. The yellow and purple stars demonstrate Wenchuan and Lushan earthquakes respectively. These are same to Figure 4, Figure 6 and Figure 7.

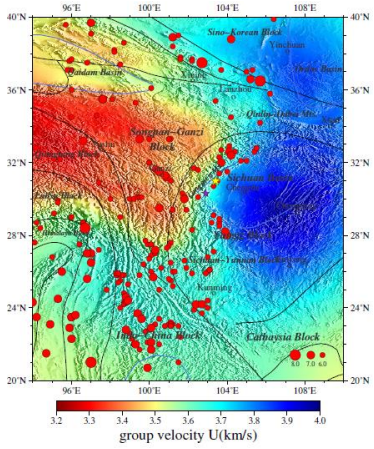

Figure 5. Averaged group velocity of western Yangtze craton according to Xie et al., 2013 from 10 to 30mHz.

3.2 training sSAE deep learning neural network

As we all know, using a set of examples of corresponding input–output pairs, artificial neural networks can approximate an arbitrary non-linear function to solve the non-linear inverse problem. These examples are presented to a network in a so-called training process, during which the free parameters of a network are modified to approximate the function of interests (de Wit et al. 2014). Here adopting sSAE deep learning neural network, detailed methods presented in 2 above, we pre-train the neural network taking theoretical group and phase velocity of Rayleigh wave as inputs and outputs and attain the initial weights and bias for neural network. And then we take theoretical group and phase velocity of Rayleigh wave as input, and crustal thickness as output to fine-tune neural network as done in a traditional neural network.

Neural network training is sensitive to the random initialization of the network parameters. Therefore, it is common practice to train several neural networks with different initializations, and subsequently choose the network which performs best on a given synthetic test data set, and the network which performed best on the test set is used to draw inferences from the observed data. After trying many times, we find the proportion of training data set to test one is 3:1 is reasonable (Figure 6).

Nonlin. Processes Geophys. Discuss., https://doi.org/10.5194/npg-2018-11
Manuscript under review for journal Nonlin. Processes Geophys.
Discussion started: 9 March 2018

We have got final test errors which may be produced not only by different neural network structure decided by the number of inputting neuron, hidden layers and neuron in middle layer, also optional parameters such as number of train epochs and size of batch. What's more, type of activation function, value of learning rate, zero masked fraction, and value of non-sparsity penalty can affect final test errors. We give twelve cases and their corresponding test errors in table 1.

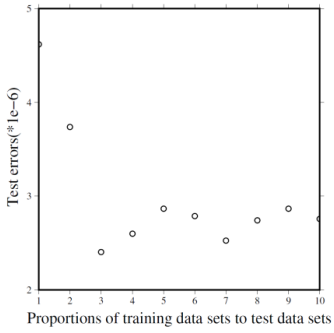

Figure 6. The relationship between proportions of training data sets to test data sets and test errors.

3.3 inverting crust thickness

Based on our all twelve neural networks, we invert Rayleigh phase velocities (10~35.0mHz) and group velocities (10~30.0mHz) to attain twelve crustal thickness models for eastern Tibet and western Yangtze craton. Considering not only the test errors of sSAE networks, also misfits and correlation coefficients of our twelve models with crustal thickness models from other researches, we select network structure given in table 1 shown in ※. We find the best fit crustal thickness model from sSAE (Figure 7).We compare our model with crustal thickness model from receiver function(Zhu et al.,2012),and the other two global crustal thickness models, CRUST2.0 from Bassin et al. (2000) and the CUB2 model from Shapiro&Ritzwoller (2002)(Figure 8) in the same region. The correlation coefficients and scatter plots of our model versus ZJS, our model versus CRUST2.0 and our model versus CUB2 (Figure 9) indicate that overall agreement between the three models. However, the agreements of our model with CUB2 and CRUST2.0 are better than with ZJS, since model ZJS attained from Zhu et.,al(2012) has relatively sparse stations with poor data coverage and lower resolution.

Table1 deep learning neural network structures taking in this article

sSAE Structure	parameters				Error	CUB2		CRUST2.0		ZJS	
	Layers	D	E	F	$\times 10^{-6}$	G	H	G	H	G	H
[21 50 10 1]	Layer 1	0.3	10	1e4	170.4	7.32	0.78	7.60	0.79	8.66	0.72
[21 50 10 1]	Layer 1	0.3	10	1e3	48.36	6.66	0.76	7.29	0.77	6.62	0.73
[21 50 10 1]	Layer 1	0.3	10	1e2	20.09	7.00	0.75	7.18	0.76	6.02	0.68
[21 50 10 1]	Layer 1	0.3	100	1e3	73.19	6.58	0.77	7.88	0.79	7.65	0.73
[21 50 10 1]※	Layer 1	0.3	100	1e2	8.40	6.62	0.78	6.70	0.80	6.63	0.69
[21 50 10 1]	Layer 1	0.01	100	1e2	6.64	6.42	0.77	6.97	0.81	6.86	0.68
[21 10 2 1]	Layer 1	0.01	100	1e2	7.47	7.20	0.78	7.43	0.78	8.15	0.72
[21 100 50 20 1]	Layer 1	0.5	100	1e2	4.77	8.07	0.74	9.87	0.79	9.63	0.63
[21 200 50 20 10 1]	Layer 1	0.5	100	1e2	2.73	13.1	0.71	14.8	0.78	16.0	0.63
[21 200 100 50 20 10 5 1]	Layer 1	0.5	100	1e2	3.33	8.93	0.77	10.5	0.83	11.6	0.66
[21 200 100 50 20 10 5 1]	Layer 1	0.5	100	50	2.53	12.6	0.79	13.7	0.85	16.5	0.67
[21 50 40 30 20 10 5 1]	Layer 1	0.5	100	50	2.54	12.3	0.77	14.3	0.80	15.2	0.73

In this article, we fixed the following four parameters in every situation: A-type of activation function(sigma); B-learning rate(1); C-zero masked fraction(0.5).

Nonlin. Processes Geophys. Discuss., https://doi.org/10.5194/npg-2018-11
Manuscript under review for journal Nonlin. Processes Geophys.
Discussion started: 9 March 2018

Various parameters: D-non-sparsity penalty, which is zero except for layer 1 in every sASE structure; E-number of epochs; F-batchsize.
G-RMS misfit of our result with other model; H-correlation coefficient of our result with other model.
※- selected sSAE neural network structure

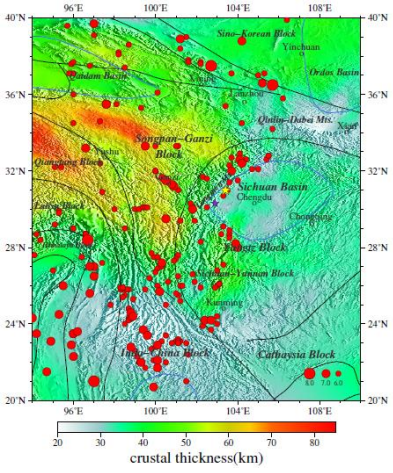

Figure 7. Crustal thickness of western Yangtze craton from this paper.

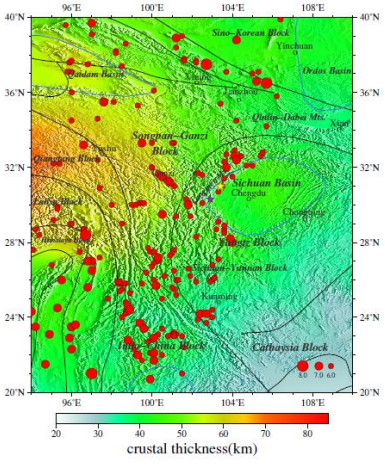

Figure 8. Crustal thickness of model CUB2 from Shapiro&Ritzwoller (2002)

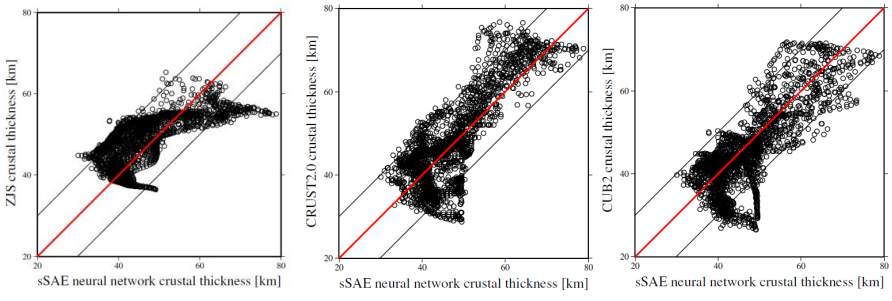

Figure 9.(From left to right) scatter plots of our model versus ZJS, our model versus CRUST2.0 and our model versus CUB2

4 Discussion

On the one hand, we can attain the crustal thickness and resultant geodynamic consequences in research region from our result. We find the relatively good agreement of our result (Fig.7) with CUB2(Fig.8),CRUST2.0 (Fig.9). All these three models indicate that crustal thickness is deeper in the west of Longmen mountain than in the east of Longmen mountain. Moreover, our result reveals more details: the eastern Tibetan Plateau crustal thickness is complex and changes largely. The average crust thickness is about above 60km, especially about 70-75km at Qiangtang block, under which there is a northward-dipping moho gradient zone. There is relatively shallow crust at Songpa-Ganzi block and is characteristic of decreasing in northwest-southeast orientation. Model CUB2.0 tells us the crustal thickness of Sichuan basin is about 40km and is relatively smooth, however our model reveals there are some changes about crustal thickness in this region, that is crustal thickness is thin around Chengdu especially northeastward to Chengdu, in addition there is about 50km thick crust under Qinlin-Dabei fold belt,also we can get that crustal thickness of northeast to Sichuan basin is about 45~48km.What's more, crustal thickness around Xi'an and Ordos basin is shallow about 35km. Conversely, change of

crustal thickness in Sichuan-Yunnan block is sharp, where crustal thickness is 60km in northwest and 35km in southeast. All detailed information is consistence with Wang et.al(2010) who attained the crustal thickness estimated by the H-k stacking method based on the broad band teleseismic data recorded at 132 seismic stations in Longmen mountains and adjacent regions(26°~35°N,98°~109°E). In addition, after analysing the distribution of the epicenters during 1970-2015, we realize that great earthquakes in Sichuan and Yunnan have occurred in brittle upper crust in Longmen mountain fault zone, where crustal thickness changes sharply as to about 10km, and great Ms 8.0 Wenchuan earthquake in 2008 and Ms 7.0 Lushan earthquake in 2013 occurred. The reason may be that main fault cut moho discontinuity where materials exchange between crust and mantle and accumulating press induce a series of earthquakes frequently.

On the other hand, our results show deep learning neural networks can invert crustal thickness effectively due to their owning capability to represent complex inverse functions:

Test errors of deep learning neural network may be influenced by the number of layer in networks which shows more layers induce smaller test errors, which we can attain from Table 1 when the number of layer in networks adds from three to six, test error decreases from 1.7e-4 to 2.5e-6. In addition, training parameters as batchsize decrease from 1e4 to 1e3 and test error decreases from 1.7e-4 to 2.5e-5. Also when epochs increase from 10 to 100, corresponding test error decreases from 2.0e-5 to 8.4e-6.

The robustness of deep learning neural networks is strong. When the number of layers in network achieves six, changes of the number of neurons in each layer have little influence on test errors which is about 2.5e-6.

The neural network structure shown in ※ from table 1 reveals misfits of our model with model CUB2, CRUST2.0 and ZJS are relatively low with 6.62,6.70 and 6.63, and corresponding correlation coefficients are relatively high with 0.78, 0.80 and 0.69 respectively, however, test errors is 8.4e-6 and is not minimum. This tells us test error may be not the only criterion determining which neural network is best because small test error may be induced by overfit.

5 Conclusion and remarks

Taking use of sSAE deep learning network, we present crustal thickness map of eastern Tibet and western Yangtze craton(Fig.7). The data sets consist of phase velocities of Rayleigh waves from Xie(2013) at discrete frequency of 10.0, 12.5, 15.0, 17.5, 20.0, 22.5, 25.0, 27.5, 30.0, 32.5,35.0mHz and derived group velocities of Rayleigh waves at discrete frequency of 10.0, 12.5, 15.0, 17.5, 20.0, 22.5, 25.0, 27.5, 30.0mHz.We conclude that:

(1) For all our simulations we use sSAE with different neural network structures which are decided by many factors such as the number of layers and neurons in neural networks, optional parameters such as the number of epoch and batchsize, type of activation function, values of learning rate and non-sparsity penalty and so on. We find that the number of hidden units is not a crucial parameter and networks with different number of hidden units give similar results, however batchsize is an important factor for results.

(2) After inverting these twelve networks, different networks produced different results. When test errors achieve some value, misfits are high and correlation coefficients are low, which we think it is maybe caused by overfit. This means networks fit well training data set, but generalization ability does not increase. In our future work, we'll focus on how to resolve this problem in using sSAE.

(3) We present a crustal thickness model for eastern Tibet and western Yangtze craton. Compared our model with current knowledge about crustal structure as represented by ZJS,CRUST2.0, CUB2. The overall agreement with these three models is very good, and agreement is generally better with CUB2 and CRUST2.0 attained from relatively dense stations with rich data coverage and higher resolution.

(4) The results are obtained using a neural network approach called sSAE which is widely and successfully used in pattern recognition. As we all know, geophysical inversion is so complex that we should analysis and enhance neural network to apply to these complicated problems.

Acknowledgements. Our work was supported by the National Natural Science Foundation of China (Grant No. 41774095) and Active Fault Study Team of Chengdu University of Technology(Grant No. 10912-KYTD201505).The authors are grateful to Xie for providing the phase velocity maps, and Zhu for making the model available.

References

Bassin, C., Laske, G. & Masters, G.: The current limits of resolution for surface wave tomography in north America, EOS, Trans. A. geophys.Un., F897,2000.

BengioY.: Learning deep architectures for AI, Foundations and trends in Machine Learning, 2(1): 1-127,2009.

Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H.: Greedy layer-wise training of deep networks. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, pp.153–160,2006.

Bishop, C.M.: Neural Networks for Pattern Recognition, Oxford University Press, Oxford, UK,1995.

Burchfiel B. C., Chen Z., Liu Y. and Royden L. H.: Tectonics of the Longmen Shan and adjacent regions, Central China. International Geology Reviews 37: 661–735,1995.

ChenY.,F.NiuR.Liu.: Crustal structure beneath China from receiver function analysis, J. geophys. Res.,115(B3):B03307,2010.

Clark M. K., Schoenbohm L. M., Royden L.H.,et.al.: Surface uplift, tectonics, and erosion of eastern Tibet from large-scale drainage patterns. Tectonics 23:tc1006, doi:10.1029/2002TC001402,2004.

Cornford, D., Nabney, I.T. & Bishop, C.M.: Neural network based wind vector retrieval from satellite scatter meter data, Neural Computing and Applications, 8, 206–217,1999.

Das, T.&Nolet, G.: Crustal thickness estimation using high frequency Rayleigh waves, Geophys. Res. Lett, 123, 169–184, 2001.

de Wit, R.W.L., et al.: Bayesian inversion of free oscillations for Earth's radial (an)elastic structure. Phys. Earth Planet.In. http://dx.doi.org/10.1016/j.pepi.2014.09.004,2014.

de Wit, R.W., Valentine, A.P. & Trampert, J.: Bayesian inference of Earth's radial seismic structure from body-wave travel times using neural networks, Geophys. J. Int., 195(1), 408–422,2013.

Deng, J., Zhang, Z., Marchi, E., Schuller, B: Sparse Autoencoder-Based Feature Transfer Learning for Speech Emotion Recognition. In Proceedings of the 5th Biannu Humaine Association Conference on Affective Computing and Intelligent Interaction, Geneva, Switzerland, pp. 511–516,2013.

Devilee, R., Curtis, A. & Roy-Chowdhury, K.:An efficient, probabilistic neural network approach to solving inverse problems: inverting surface wave velocities for Eurasian crustal thickness, J. geophys. Res.,104(B12), 28841–28857,1999.

Dziewonski, A.M., Anderson, D.L.: Preliminary reference Earth model. Physics of the Earth and Planetary Interiors. 25, 297-356,1981.

Fang J.: Global crustal and lithosphere thickness inverted by using satellite gravity data, Crustal deformation and Earthquake, **19**,1,26–31 (in Chinese) ,1999.

Gao X., Wang W.M, Yao, Z. X.: Crustal structure of China mainland and its adjacent regions. Chinese J .Geophys, (in Chinese) , 48,3,591-601,2005.

Grad M.,Tiira T.: The moho depth map of the European Plate, Geophys. J. Int., 176(1), 279–292,2009.

Guo, L.H., et al.: Preferential filtering method and its application to Bouguer gravity anomaly of Chinese continent. Chinese J .Geophys, (in Chinese). 55 (12), 4078–4088,2012.

Hinton, G. E. and Salakhutdinov, R. R.: Reducing the dimensionality of data with neural networks, Science, 313,5786, 504 - 507,2006.

Huang J. P., Fu R. S. Xu P., et al.: Inversion of gravity and topography data for the crust thickness of China and its adjacency, Acta Seismologica Sinica(in Chinese),28,3,250-258,2006.

Lampinen, J. &Vehtari, A.: Bayesian approach for neural networks -review and case studies, Neural Networks, 14(3), 257–274,2001.

Laske, G., Masters., G., Ma, Z. and Pasyanos, M.: Update on CRUST1.0 - A 1-degree Global Model of Earth's Crust, Geophys. Res. Abstracts, 15, Abstract EGU2013-2658,2013.

Lebedev, S., Adam, J. M. C., & Meier, T.: Mapping the moho with seismic surface waves: a review, resolution analysis, and recommended inversion strategies. Tectonophysics, 609, 377–394,2013

LeCun, Y., Bengio, Y. and Hinton, G. E.: Deep Learning. Nature, 521, 436-444,2015.

Legendre, C. P., Deschamps, F., Zhao, L., & Chen, Q. F.: Rayleigh-wave dispersion reveals crust-mantle decoupling beneath eastern Tibet. Scientific reports,2015.

Masters G., Barmine M.P., Kientz S. Pasadena.: Calif. Inst. of Technol., Mineos User's Manual. Computational Infrastructure for Geodynamics,2014.

Masters, G., Barmine, M., Kientz, S.: Mineos. User Manual Version 1.0.2. Calif.Inst. of Tech., Pasadena, CA,2011.

Montagner, J.-P.&Jobert, N.:Vectorial tomography—ii. Application to the indian ocean, Geophys. J. Int., 94, 309–344,1988.

Meier U., Curtis A., Trampert J.: Global crustal thickness from neural network inversion of surface wave data, Geophys. J. Int., 169,706-722,2007.

Liu, Q., Hu, X., Ye, M., Cheng, X. and Li, F.: Gas Recognition under Sensor Drift by Using Deep Learning. Int. J. Intell. Syst., 30: 907–922. doi: 10.1002/int.21731,2015.

Shapiro, N. M. and Ritzwoller,M. H.: Monte-Carlo inversion for a global shear-velocity model of the crust and upper mantle，Geophys. J. Int. ,151, 88–105,2002.

Stolk, et al.:High resolution regional crustal models from irregularly distributed data: application to Asia and adjacent areas. Tectonophysics, 602, 55–68,2013.

Sun, X., X. Song, S. Zheng, Y. Yang, and M. Ritzwoller: Three dimensional shear wave velocity structure of the

crust and upper mantle beneath China from ambient noise surface wave tomography, Earthquake Sci., 23, 449–463,2010.

Teng, J. W., Wang, Q. S., Wang, G. J., et al.: Specific gravity field and deep crustal structure of the ‘Himalayas east structural knot’. Chinese J .Geophys, (in Chinese), 49,4,1045-1052,2006.

van der Baan, M. &Jutten, C.: Neural networks in geophysical applications, Geophysics, 65, 1032–1047,2000

Wang C Y,LouH,Yao Z X,etal.: Crustal thicknesses and poisson’s ratios in Longmen shan mountains and adjacent regions. Quaternary Sciences(in Chinese),30(4):652-661,2010.

Xie, J., M. H. Ritzwoller, Shen W., Yang Y., Zheng Y., and Zhou L.: Crustal radial anisotropy acrossEastern Tibet and the Western Yangtze Craton, J. geophys. Res.118(8), 4226-4252, doi:10.1002/jgrb.50296,2013.

Xu, L., Rondenay, S. & Van Der Hilst, R. D.: Structure of the Crust beneath the southeastern Tibetan Plateau from teleseismic Receiver Functions. Physics of the Earth and Planetary Interiors. 165, 176–193,2007.

Yi G X,Yao HJ,Zhu J S,et a1.:Rayleigh-wave phase velocity distribution in China continent and its adjacent regions.Chinese J.Geophys.(in Chinese).51(2):402-411,2008.

Yao, H., R.D. Van Der Hilst, and M. V. De Hoop.: Surface-wave array tomography in SE Tibet from ambient seismic noise and two-station analysis–I. Phase-velocity maps, Geophys. J. Int., 166(2), 732–744,2006

Zeng R. S., Sun W. G. et.al.: The depth of Moho in the mainland of China. Acta Seismologica Sinica(in Chinese), 17,3,322-327,1995.

Zhang P.,Zhu L. B.,Chen H. P.,Wang Q. D.,Yang Y.: Crustal structure in China from teleseismic receiver function, Acta Seismologica Sinica,36,5,850-861.doi:10.3969/j.issn.0253-3782.2014.05.009,2014.

Zhang, Q., E. Sandvol, J. Ni, Y. Yang, and Y. J. Chen.: Rayleigh wave tomography of the northeastern margin of the Tibetan Plateau, Earth Planet. Sci. Lett.,304(1-2), 103–112,2011.

Zheng, S., X. Sun, X. Song, Y. Yang, and M. H. Ritzwoller.: Surface wave tomography of China from ambient seismic noise correlation, Geochem. Geophys.Geosyst., 9, Q05020, doi:10.1029/2008GC001981,2008.

Zhou, L., J. Xie, W. Shen, Y. Zheng, Y. Yang, H. Shi, and M. H. Ritzwoller.: The structure of the crust and uppermost mantle beneath South China from ambient noise and earthquake tomography, Geophys. J. Int., 189(3), 1565–1583,2012.

Zhou,Y., Nolet, G., Dahlen, F.&Laske, G.: Global upper-mantle structure from finite-frequency surface-wave tomography, J. geophys. Res.,111, B04304,2006.

Zhu J. S., Zhao J. M., Jiang X. T. et.al.: Crustal flow beneath the eastern margin of the Tibetan plateau, Earthquake Science,25,5-6, 469-483,2012.

Zhu J.S., Cao J. M., Cai X. L. ,et.al.: High resolution surface wave tomography in east asia and west pacific marginal seas,Chinese J . Geophys, (in Chinese).45,5,646-664,2002.