Author's Response for comments of review 1

Dear Anonymous Referee:

We very much appreciate the overall positive attitude of the referee to our manuscript and thank you for your time and very useful comments! We give below responses to some of these. And we also give a marked-up version and a revised version.

1. The sample selection for learning is not clearly described in the paper. Please add the sample selection instructions, and specify the source of the samples and the selection principles.

Reply: In the training experiment, we collected 2400 periods of transient electromagnetic method secondary field signals from the same collection location, and selected 434 data points in per period. Meanwhile, 100 periods of signals are randomly acquired as a test and validation set for the improving the robustness of the model.

2.Please describe the parameters of the experimental platform, the hyper-parameter indexes in the model, and the code of the main model module.

Reply: Thank the reviewer for this comment about hyper-parameters, so we accept your suggestion to add more experiment details about learning rate, batch-size and so on. More details can be find in marked-up manuscript. Meanwhile, we upload the code of the main model module on github: https://github.com/tonyckc/SFSDSA.

3.Please add more description of the specific parameters and experimental details of kalman filter and wavelet transform.

Reply: This question was described in detail in the fifth part of the original manuscript.

4. The related work section over introduces the encoder related literature. Please elaborate the literature of transient electromagnetic signal noise reduction and signal filtering.

Reply: Yes, we introduce fewer related works about transient electromagnetic method signal noise reduction and signal filtering. Therefore, we describe more related works about TEM denoising using other Kalman filter and Wavelet transform method, and then adding more references. More details can be find in marked-up manuscript.

5. The loss in figure 1 is calculated as two regularized losses. What does that mean? It should be explain detail.

Reply : We are very sorry that make you think that we calculate two regularized losses because of our error figure. In fact, in Page 4 Line 14 we described that the theoretical signal is used to get the model loss with the output using the activation function. Meanwhile, we

have replaced Figure 1 to a clear version. And the new figure is shown as below.



6.Please explain the relationship between figure 5 and the corresponding explanatory text. And whether the coordinates in Figure 5 is correct ?

Reply : In the part 4, we narrated about the figure 5 detailedly. Small-scale deep learning model, and the training times can be less. As is shown in Figure 5, we set each batch to 8 and every 10 training steps output MAE value as a iteration sampling point. We found that the MAE values near the 150th sampling point which will start oscillate.

Both coordinates of the figure is logarithmic, the two coordinates have been modified. That can be seen in modified manuscript and is shown as below.



Author's Response for comments of review 2

Dear Anonymous Referee:

We very much appreciate the overall positive attitude of the referee to our manuscript and thank you for your time and very useful comments! We give below a first response to some of these. Meanwhile, according to your comments, we revised this manuscript. All of the changes were made in a marked-up manuscript version and a clear revised version.

1. Comment from Reviewer: How exactly you are planning to train the network on realistic geophysical problems?

Reply:

We agree with your comment. We explained how to train the network on realistic geophysical problems in the section 4 (Experiment and Analysis). However, we found that the explanation of this process was not clear after we carefully read the fourth part of the manuscript again.

We briefly described the process: As described in the third part (Mathematical Derivation of SFSDSA), we can obtain an actual detection signal sample and a theoretical signal sample, and then we build a model for training. Meanwhile, Figure 1 shows the network structure and training process in a more vivid way. For realistic geophysical problems such as transient electromagnetic method secondary field signals, we carried out experiments in the fourth part according to the process proposed in the third part .We collect the actual detection signal of the secondary field, the dimensions are 1 * 434 (this dimension can reflect the attenuation process of transient electromagnetic method secondary field secondary field signal), the inversion theory signal and the actual detection signal of the secondary field has the same dimensions. Finally, we used two samples for training according to training process of the Figure 1(we added more details, such as training platform, hyper-parameters etc.). Meanwhile, inversion theory signals play a semi-supervisory role in the model. In the end, SFSDSA can map the signal points of the noise interference to the high probability points with clean signal as reference according to the deep characteristics of the signal, so as to realize the signal noise and reduce noise interference.

2. Comment from Reviewer: Is this method can be generalized in the sense that the training on one data can be used on different datasets? Reply:

Yes, this method has a good generalization in a certain sense. Our method has good generalization for different collection points of the same geological feature area. As shown in Figure 9, we use the same model for 7 collection points. However, if the acquisition areas of the two data have large differences in geological features, this will inevitably lead to different deep features of the forward and inversion signals that cause the secondary field. The model will perform noise reduction based on the geological features represented by the previous training dataset. Therefore, it is necessary to analyze the known geological features more carefully and apply the model according to the actual geological conditions before using our method. At the same time, this view is consistent with machine learning theory(**Neyshabur et**

al,2017). If the model will be well generalized, it must be built to varying degrees of similarity problems. If we do not analyze the principle of the problem and ignore the huge differences in features, it is unrealistic to try to achieve a high degree of generalization. According to this comment, we added this view to the part of conclusion the marked-up manuscript.

3. Comment from Reviewer: Any comments on using supervised learning since that seems work better than the unsupervised learning ? Reply:

Recently, we have noticed that supervised learning performs well in classification problems such as image recognition and semantic understanding (He et al., 2016, Long et al.,2014) . At the same time, unsupervised learning also has a good performance in clustering and association problems (Klampanos et al., 2018.), and the goal of unsupervised learning is usually to extract the distribution characteristics of the data in order to understand the deep features of the data (Becker et al., 1996, Liu et al., 2015). Both supervised learning and unsupervised learning have their own well-behaved areas, so we need to choose different learning styles and models for different problems. For the noise suppression problem of the secondary field signal in transient electromagnetic method, our goal is to extract the deep features of the secondary field signal, and map the data points affected by noise to the estimated high probability points according to their own signal features. We also found that the purpose of extracting the distribution characteristics of the secondary field signal data is similar to that of unsupervised learning. Meanwhile, unsupervised learning models are widely used in different signal noise reduction problems, some of which perform well such as gravitational waves, power transmission equipment status signals, etc. According to this comment, we added this view to the part of relate work in marked-up manuscript.

4. Comment from Reviewer: If noise is not random as shown in the examples, will this method still work?

Reply: Yes, Our model can extract features of the secondary field signal, so as the signal points of noise interference are mapped to the estimated high probability points according to their own signal characteristics. From a very natural point of view, noise can be seen as an interference whether it is random or not. At the same time, the deep learning model has a good generalization feature to support our point of view (**Neyshabur et al., 2017**), we also added measures to improve generalization in SFSDSA such as regularization (**Nowlan and Hinton., 1992**), so our method has a better performance in actual tests such as the results of Figure 8 and Figure 10. Therefore, this method is still work in a certain sense if the noise is not random.

5. Comment from Reviewer: What the runtime cost of the proposed method compared to other denoising methods?

Reply: Our runtime cost are less at the end of training compared to other denoising methods such as wavelet transform. We can use the data with noise to achieve end-to-end denoising (as described in the process of Figure 1) using the trained model, without having to spend a lot of time to adjust the wavelet threshold and wavelet base like wavelet transform.

For small sample data sets, the time consumption difference between SFSDSA and other

denoising methods is small, but when the number of data samples reach a certain quantity, the model has a higher advantage in time consumption after training.

We appreciate all the comments, which we will use to improve the manuscript.

References

He K , Zhang X , Ren S , et al. Deep Residual Learning for Image Recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2016, DOI:10.1109/CVPR.2016.90

Long J., Shelhamer E., Darrell T., Fully Convolutional Networks for Semantic Segmentation. IEEE Transactions on Pattern Analysis & Machine Intelligence, 39(4):640-651, 2014.

Klampanos I A , Davvetas A , Andronopoulos S , et al.: Autoencoder-driven weather clustering for source estimation during nuclear events, Environmental Modelling & Software, 102:84-93, 2018,

Becker S , Plumbley M.: Unsupervised neural network learning procedures for feature extraction and classification. Applied Intelligence: The International Journal of Artificial, Intelligence, Neural Networks, and Complex Problem-Solving Technologies,6(3):185-203. 1996.

Liu J H , Zheng W Q , Zou Y X . A Robust Acoustic Feature Extraction Approach Based on Stacked Denoising Autoencoder IEEE International Conference on Multimedia Big Data. IEEE Computer Society, 2015.

Neyshabur B , Bhojanapalli S , Mcallester D , et al. Exploring Generalization in Deep Learning. 2017

S. J. Nowlan and G. E. Hinton. Simplifying neural networks by soft weight-sharing. Neural Computation, 4(4), 1992.

Author's Response for comments of review 3

Dear Anonymous Referee:

Thank for you positive comments to our manuscript!

We give below responses to some of these. Meanwhile, according to your comments, we revised this manuscript overhaul. All of the changes were made in the supplement files, which are a marked-up version and a revised version.

1. Comment from Reviewer: "The paper aims to denoise a signal with autoencoders (unsupervised manner). However, the authors did this by putting a theoretical signal as output. This is not an unsupervised manner to proceed. Why did the authors put an output? Is it a traditional way to proceed in geophysics?"

Reply: We are very sorry that make you think that we put a theoretical signal as output because of our unclear expression in Figure 1(in original manuscript). In fact, in **Page 4 Line 14**, we described that the theoretical signal is used to get the model loss with the output using the loss function, to realize back propagation. Meanwhile, we have replaced Figure 1 with a clear version. For another question, naturally, it is not a traditional way to proceed in geophysics. We have modify manuscript according to this comment.

2. Comment from Reviewer: "Page 2 line 3: can you please explain little bit why PCA is cumbersome and what could be the effect on the signal used as case study?

Reply: According to the references (Wu et al., 2014), the process of PCA can be divided into 5 steps.

- (1) Normalize the obtained data
- (2) Calculate the covariance matrix for obtaining multidimensional data
- (3) Decompose the covariance matrix to obtain the eigenvalue matrix and eigenvector
- (4) Obtain the corresponding main components after dimensionality reduction according to the PCA calculation method

(5) Selecting the representative principal components by the trend comparison method and the L-curve method, and performing reconstruction to obtain the denoised secondary field signal waveform.

By using the PCA method, we do the experiment to verify the effect of noise reduction. But the process of programming is more complicated using mathematical derivation, so we use scikit-learn library to realize noise reduction. However, the underlying structure is not easy to modify resulting in scikit-learn library is unable to adjust parameters adaptively based on signal characteristics. Meanwhile, we found that the filtering effect is not ideal. More details can be find in revised manuscript.

3. Comment from Reviewer: "Page 3, in which the SELU activation function and the ADAM optimization algorithm are introduced, a justification of choice is needed."

Reply: The problem of too many nodes dying is a general disadvantage for RELU activation function and improved RELU activation functions like leaky RELU all consistently outperform

the RELU in some tasks (Xu et al.2015). Therefore, it is necessary to apply the improved RELU function to reduce the impact of the shortcomings of the RELU function. We choose the SELU that have the preponderances of overcoming vanishing and exploding gradient problems in a sense and the best preforming in full connection networks (Klambauer et al., 2014).

Adam algorithm have the advantages of calculating different adaptive learning rates for different parameters and requiring little memory(Kingma et sl.,2014). Through **Table 1**(in original manuscript), we find that the combination of models using SELU is better than the combination of models using RELU in the MAE indicator. Similarly, we find that the combination of models using ADAM optimization algorithm outperform compared with not using ADAM in the MAE indicator. More details can be find in revised manuscript.

4. Comment from Reviewer: "Page 3 line 24: "SELU activation function is utilized to prevent too many of depth": please put a reference for that? Same page line 12, authors said: SELU and ADAM optimization algorithm are used to solve the problem of over-fitting. How? Need references for this point or a good justification."

Reply: we are very sorry that the sentence of "SELU activation function is utilized to prevent too many of depth" has a spelling mistake (the word 'depth' should be replaced to 'death') to lead to an unclear and incorrect description. In fact, this sentence wants to express that SELU is utilized to reduce the impact of too many dying nodes problem(Xu et al.2015, Klambauer et al., 2014).

For the second question in page 3 line 12, our description of function of SELU and Adam is unclear because of the poor grammar. In fact, we chose Adam algorithm, which have the advantages of calculating different adaptive learning rates for different parameters and requiring little memory(Kingma et sl.,2014). And SELU have the preponderances of overcoming vanishing and exploding gradient problems in a sense and the best preforming in full connection networks (Klambauer et al., 2014). We changed the description of the part to a correct expression. More details can be find in revised manuscript.

5. Comment from Reviewer:"Please add other criteria in addition to the MAE"

Reply: In fact, we analyzed and compared the selection of the two loss functions of MAE and MSE in the previous experiments as shown in figure 1. Meanwhile, according to the previous work and the secondary field signal denoising task of transient electromagnetic method, we think that MAE is a better choice.

First, our task is to map the outliers affected by noise to the vicinity of the theoretical signal point, in other words, model should ignore the outliers affected by noise to make it more consistent with the distribution of the overall signal. We know that MAE is quite resistant to outliers(Rishabh, 2015), so we choose it.

Second, the squared-error is going to be huge for outliers, which tries to adjust the model according to these outliers on the expense of other good-points(Rishabh, 2015). For signal that are subject to noise interference in the secondary field of transient electromagnetic method, we don't want to over-fitting outliers that are disturbed by noise, but we want to treat them as noise interfered data.

Finally, observing the secondary field signal of transient electromagnetic method, we

found that the amplitude of the early track data points is very large, but the amplitude of the late track data is small, and the squared-error will inevitably give the early points of the abnormal points more weight to result in Ignoring the difference in late-channel data, this is very unfair. This question may lead to inaccurate model and late-channel signals will be ignored.



We have modify manuscript according to this comment.



6. Comment from Reviewer: "The data splitting need more explanations. The experimental case study needs also some explanation with some exploratory analysis"

Reply: In the previous experiment, we randomly collected 2400 periods of transient electromagnetic method secondary field signals from the same collection location and we collected 434 signal points per period. Meanwhile, 100 periods of signals are randomly acquired as a test and validation set. In the meantime, we accept the second suggestion to do some explanation with some exploratory analysis in reply 7 and we update the manuscript for adding more details. We have modify manuscript according to this comment.

7. Comment from Reviewer: "For choosing only 2 hidden layers, did you take into account the other hyper parameters. I suggest a grid search, which is possible to do using TensorFlow library or Keras in Python".

Reply: Thank the reviewer for this precious and professional comment about hyper parameters, and we're so sorry that this paper doesn't list some important hyper parameters such as learning rate, regular parameter and so on completely. We have added those key hyper-parameters in marked-up manuscript.

In the previous experiments, we set hyper-parameters (batch-size=8, learning-rate=0.1, regularization-rate=0, epochs=20) based on experience but we initially take the measure of a small number of epochs (epochs=2) according to experiment. We added the experiment as shown in Figure 2 to support our standpoint. The model oscillates quickly and converges. Training with fewer epochs can avoid useless training and over-fitting, maintaining the distribution characteristics of the signal itself. As shown in Figure 5(in original manuscript), the reconstruction error oscillates and converges as the training progresses. This phenomenon is similar to the tail of the actual signal. We try stopping training when the convergence occurs, the idea similar to early-stopping makes the model more robust(Caruana ,2000).





At the same time, we got the result of stacking two AEs with good effect as shown in Figure 4(in original manuscript). We guess that the size of the AE hidden layer is too small after multiple stacks (for instance, the 4th AE only has 27 nodes because the size of latter AE is half of the previous AE in order to extract the better feature), and the representation of signal characteristics are not complete resulting in large reconstruction costs. If we want to get a better result, more iterations may be used but this tends to cause overfitting. Meanwhile, we found that the reconstruction loss of the second AE is already very small shown in Figure 2. And it is not necessary to stack more AEs.

We accept the reviewer's suggestion to do a grid search, and we get the good parameter combination of learning rate and regularization rate (learning rate=0.001 and regularization rate=0.15).

Table 1 The training cost of combination of learning rate and regularization rate. The value represents the MAE of the first fifty data points. According to the experience, about the first fifty data points have better effect for extracting time-domain order waveforms.

regularization rate learning rate	0.05	0.1	0.15	0.2
0.1	61515.3	12670.3	14448.9	11112.1
0.01	1735.2	1918.1	2126.6	1825.7
0.001	1526.6	1669.5	1377.3	1780.6
0.0001	1493.2	1678.1	1392.3	1955.5

8. Comment from Reviewer:" For the comparison with traditional methods, please add PCA."

Reply: We have already added in the manuscript about the comparison of PCA algorithm in transient electromagnetic signal denoising. After the filtering test, and then the MAE corresponding to the calculation of the theoretical data, it can be seen that the effect of pca filtering is lower than SFSDSA. Please see the fifth part of the article for details.

9. Explain how the traditional methods were applied (mother wavelet …). Reply :

A denoising algorithm utilizing wavelet threshold method and exponential adaptive window width-fitting(Ji et al.,2016). An exponential fitting algorithm was used to achieve the attenuation curve for each window, and the data contaminated with non-fixed electromagnetic noise was replaced by their results.

Another algorithm utilizes multi-resolution analysis via a stationary wavelet transform of the data(Li et al.,2017). The measured data are decomposed into detailed coefficients and approximated coefficients. Then, the logarithmic slope of measured data and a threshold are calculated to identify the noise in the detailed coefficients; the corresponding detailed coefficients are processed to reduce the noise. Finally, the undisturbed data are reconstructed using inverse stationary wavelet transform.

The third method presents an exponential fitting-adaptive Kalman filter to remove mixed electromagnetic noises(Ji et al.,2017), while preserving the signal characteristics. It consists of an exponential fitting procedure and an adaptive scalar Kalman filter. The adaptive scalar Kalman uses the exponential fitting results in the weighting coefficients calculation.

Another wavelet-based baseline drift correction method for grounded electrical source airborne transient electromagnetic signals(Wang et al.,2013), through simulations, this method can improve the signal-to-noise ratio. Simulation results show that the wavelet-based method outperforms the interpolation method. All above were added in manuscript at the part of related work.

Response for some remarks:

1. Put more explanation on the caption of figure 1 if possible.

Reply: We accept this suggestion to put more explanation on the caption of figure 1. More details can be find in revised manuscript.

2. Equation 9: put bracket. In addition, explain it little bit (m, X, h …) if possible.

Reply: We are so sorry that we miss bracket on the right of 'x', and the input value of MAE should revised to 'x' and 'y'. x denotes the noise interference data, m denotes the number of sampling points, h denotes the model and y denotes theoretical data. The revised formula can be find in marked-up manuscript.

$$MAE(\mathbf{x}, y) = \frac{1}{m} \sum_{i=1}^{m} |h(x)^{(i)} - y^{(i)}|$$

3. Page 6 line 10: the authors used Tensorflow, please put a figure of the architecture of the used model.

Reply: The figure is exporting from TensorBoard GRAPHS to show the architecture of used model. https://github.com/tonyckc/SFSDSA/blob/master/The%20model%20structure%20.png

4. Since the journal is open source, think to put your code on an open source platform (e.g. GitHub ...)

Reply: Code can be find: https://github.com/tonyckc/SFSDSA.

We appreciate all the comments, which we will use to improve the manuscript.

References:

Caruana R, Lawrence S, Giles L.,: Overfitting in neural nets: backpropagation, conjugate grad ient, and early stopping, International Conference on Neural Information Processing System s. MIT Press, 2000.

Ji, Y., Li, D., Yu, M.,Wang,Y.,Wu,Q.,Lin,J. :A de-noising algorithm based on wavelet thresholdexponential adaptive window width-fitting for ground electrical source airborne transient electromagnetic signal, Journal of Applied Geophysics., 128,1-7,DOI: 10.1016/j.jappgeo.2016.03.001,2016.

Ji, Y., Wu, Q., Wang, Y.,Lin,J.,Li,D.,Du,S.,Yu,S.,Guan,S.: Noise reduction of grounded electrical source airborne transient electromagnetic data using an exponential fitting-adaptive Kalman filter, Exploration Geophysics., DOI: 10.1071/EG16046 ,2017.

Klambauer, G., Unterthiner, T., Mayr, A., Hochreiter, S.: Self-Normalizing Neural Networks, arXiv preprint, arXiv:1706.02515:1706.0251, 2014.

Li, D., Wang, Y., Lin, J.,Yu,S.,Ji,Y.: Electromagnetic noise reduction in grounded electricalsource airborne transient electromagnetic signal using a stationary wavelet- based denoising algorithm, Near Surface Geophysics.,1(-9),DOI: 10.3997/1873-0604.2017003 2017.

Rishabh.: L1 vs. L2 Loss function, http://rishy.github.io/ml/2015/07/28/l1-vs-l2-loss, 2015

Xu B, Wang N, Chen T, et al. Empirical evaluation of rectified activations in convolutional network[J]. arXiv preprint arXiv:1505.00853, 2015

Wang, Y., Ji, Y.J, Li,S.Y., Lin, J., Zhou, F.D., Yang, G.H.,: A wavelet-based baseline drift correction method for grounded electrical source airborne transient electromagnetic signals. Exploration Geophysics,44,229–237,2013.

Wu, Y., Lu, C. D., Du, X. Z., Yu, X.D.: A denoising method based on principal component analysis for airborne transient electromagnetic data, Computing Techniques for Geophysical and Geochemical Exploration., 36(2), 170-176, DOI: 10.3969/j.issn.1001-1749.2014.02.08, 2014.

List of relevant changes made in manuscript

List of comments modified according to the reviewer 1 in manuscript :

1. The sample selection for learning is not clearly described in the paper. Please add the sample selection instructions, and specify the source of the samples and the selection principles.

Reply: This change can be seen in page 7 line 5.

2. Please describe the parameters of the experimental platform, the hyperparameter indexes in the model, and the code of the main model module.

Reply: This change can be seen in page 12 line 12. Meanwhile, we upload the code of the main model module on github: https://github.com/tonyckc/SFSDSA.

3. Please add more description of the specific parameters and experimental details of kalman filter and wavelet transform.

Reply: This change can be seen in page 12 line 5.

4. The related work section over introduces the encoder related literature. Please elaborate the literature of transient electromagnetic signal noise reduction and signal filtering.

Reply: This change can be seen in page 2 line 19.

5. The loss in figure 1 is calculated as two regularized losses. What does that mean? It should be explain detail.

Reply : This change can be seen in page 6.

6. Please explain the relationship between figure 5 and the corresponding explanatory text. And whether the coordinates in Figure 5 is correct ?Reply : This change can be seen in page 11 Figure 7.

List of comments modified according to the reviewer 2 in manuscript :

1. Comment from Reviewer: How exactly you are planning to train the network on realistic geophysical problems?

Reply: This change can be seen in page 7 line 5 and page 8 line 9.

2. Comment from Reviewer: Is this method can be generalized in the sense that the training on one data can be used on different datasets?

Reply: This change can be seen in page 15 line 4.

3. Comment from Reviewer: Any comments on using supervised learning since that seems work better than the unsupervised learning ?

Reply: This change can be seen in page 3 line 14.

4. Comment from Reviewer: If noise is not random as shown in the examples, will this method still work?

Reply: This question is explained in response 2 for the review 2.

5. Comment from Reviewer: What the runtime cost of the proposed method compared to other denoising methods?

Reply: This question is explained in response 2 for the review 2.

List of comments modified according to the reviewer 3 in manuscript :

1. Comment from Reviewer: "The paper aims to denoise a signal with autoencoders (unsupervised manner). However, the authors did this by putting a theoretical signal as output. This is not an unsupervised manner to proceed. Why did the authors put an output? Is it a traditional way to proceed in geophysics?"

Reply: This change can be seen in page 6 (Figure 1 and it's caption).

2. Comment from Reviewer: "Page 2 line 3: can you please explain little bit why PCA is cumbersome and what could be the effect on the signal used as case study?
Reply: This change can be seen in page 2 line 5, page 12 line 10 and more detail in response 3 for review 3.

3. Comment from Reviewer: "Page 3, in which the SELU activation function and the ADAM optimization algorithm are introduced, a justification of choice is needed."Reply: This change can be seen in page 4 line 1-4.

4. Comment from Reviewer: "Page 3 line 24: "SELU activation function is utilized to prevent too many of depth": please put a reference for that? Same page line 12, authors said: SELU and ADAM optimization algorithm are used to solve the problem of over-fitting. How? Need references for this point or a good justification."

Reply: This change can be seen in page 4 line 5-9 and page 4 line 21.

5. Comment from Reviewer: "Please add other criteria in addition to the MAE" Reply: This change can be seen in page 7 line 11 and page 8 line 2.

6. Comment from Reviewer: "The data splitting need more explanations. The experimental case study needs also some explanation with some exploratory analysis" **Reply:** This change can be seen in page 7 line 5.

7. Comment from Reviewer: "For choosing only 2 hidden layers, did you take into account the other hyper parameters. I suggest a grid search, which is possible to do using TensorFlow library or Keras in Python".

Reply: This change can be seen in page 8 line 9, page 7 line 8 and Figure 6 and page 11 Table 1.

8. Comment from Reviewer:" For the comparison with traditional methods, please add PCA."

Reply: This change can be seen in page 12 line 10, Figure 10(c) and page 2 (Related work)

9. Explain how the traditional methods were applied (mother wavelet ...).

Reply: This change can be seen in page 2(Related work) and added some references.

Response for some remarks(Review 3):

1. Put more explanation on the caption of figure 1 if possible.

Reply: We accept this suggestion to put more explanation on the caption of figure 1. More details can be find in revised manuscript.

2. Equation 9: put bracket. In addition, explain it little bit (m, X, h …) if possible.

Reply: This change can be seen in page 8, formula 9($MAE(\mathbf{x}, y) = \frac{1}{m} \sum_{i=1}^{m} |h(\mathbf{x})^{(i)} - y^{(i)}|$)

3. Page 6 line 10: the authors used Tensorflow, please put a figure of the architecture of the used model.

Reply: The figure is exporting from Tensorboard GRAPHS to show the architecture of used model. https://github.com/tonyckc/SFSDSA/blob/master/The%20model%20structure%20.png

4. Since the journal is open source, think to put your code on an open source platform (e.g. GitHub ...)

Reply: The code can be find: https://github.com/tonyckc/SFSDSA.

A denoising stacked autoencoders for transient electromagnetic signal denoising

Fanqiang Lin^{1,4}, Kecheng Chen^{2,1}, Xuben Wang^{3,4}, Hui Cao^{3,4}, Danlei Chen¹, and Fanzeng Chen¹ ¹School of Information Science and Technology, Chengdu University of Technology, Chengdu,610059,China ²School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu,611731,China ³College of Geophysics, Chengdu University of Technology, Chengdu 610059,China ⁴Key Lab of Geo-Detection and Information Techniques of Ministry of Education, Chengdu 610059,China *Correspondence to:* Kecheng Chen(15775118240@163.com)

Abstract.

Transient electromagnetic method (TEM) is extremely important in geophysics. However, the secondary field signal(SFS) in TEM received by coil is easily disturbed by random noise, sensor noise and man-made noise, which results in the difficulty in detecting deep geological information. To reduce the noise interference and detect deep geological information, we apply

- 5 autoencoders, an unsupervised learning model in deep learning, on the basis of analyzing the characteristic of SFS, to denoise SFS. We introduce SFSDSA, a Secondary Field Signal Denoising Stacked Autoencoders, based on deep neural networks of feature extraction and denoising.SFSDSA maps the signal points of the noise interference to the high probability points with clean signal as reference according to the deep characteristics of the signal, so as to realize the signal denoising and reduce noise interference. The method is validated by the measured data comparison, and the comparison results show that the noise
- 10 reduction method can effectively reduce the noise of SFS, in contrast with the Kalman and wavelet transform methods, and strongly support the speculation of deeper underground features.

1 Introduction

20

Through the analysis of SFS in TEM, the information of underground geological composition can be obtained and has been widely used in mineral exploration, oil and gas exploration and other fields (Danielsen et al., 2003, Haroon et al., 2015) Due

15 to the small amplitude of the late field data in the secondary field, it may be disturbed by random noise, sensor noise, human noise and other interference(Rasmussen et al., 2017) which leads to data singularities or interference points, and thus the deep geological information can not be reflected well. Therefore, it is necessary to make full use of the characteristics of the secondary field data to reduce the noise in the data and increase the effective range of the data.

Many methods have been developed for noise reduction of transient electromagnetic method. These methods can be broadly categorised into three groups:

- (1)Kalman filter algorithm(Ji et al.,2017)
- (2)Wavelet transform algorithm(Ji et al., 2016, Li et al., 2017)

(3)Principal component analysis(PCA) (Wu et al., 2014)

Kalman filtering is an effective method in linear systems, but it has little effect in nonlinear fields such as transient electromagnetic signals. The acquisition of wavelet threshold is cumbersome, and wavelet base selection is very difficult. In order to achieve the desired separation effect, to design an adaptive wavelet base is necessary. Likewise, the PCA algorithm is cumber-

5 some too. According to the references(Wu et al., 2014), the process of PCA is composed of 5 steps.

However, deep learning has been used to reduce noise from images, speech, and even gravitational waves(Jifara et al.,2017, Grais et al.,2017, Shen et al.,2017). Meanwhile, the representative model of deep learning Autoencoder(AE)(Bengio et al.,2006) has been successfully applied in many fields (Hwang et al.,2016). AE with noise reduction capability(Denoising Autoencoders,DAE)(Vincent et al.,2008) has been widely used in image denoising (Zhao et al.,2014), audio noise reduction (Dai et al.,2015), the recon-

10 struction of holographic image denoising(Shimobaba et al.,2017) and other fields.

Nevertheless, in the field of geophysics, the application of deep learning model is limited (Chen et al., 2014) The use of deep learning model to reduce the noise of geophysical signals has not been applied.

Therefore, in this paper, the Secondary Field Signal Denoising Stacked Autoencoders (SFSDSA) is proposed to reduce noise, based on a deep neural network with SFS feature extraction.

15 SFSDSA will be affected by noise interference signal point according to the deep characteristics of the signal mapping to the high probability of points by reference to SFS geophysical inversion signal, so as to realize the signal denoise and reduce noise interference.

2 Related Work

A denoising algorithm utilizing wavelet threshold method and exponential adaptive window width-fitting(Ji et al., 2016). An

20 exponential fitting algorithm was used to achieve the attenuation curve for each window, and the data contaminated with non-fixed electromagnetic noise was replaced by their results.

Another algorithm utilises multi-resolution analysis via a stationary wavelet transform of the data(Li et al.,2017). The measured data are decomposed into detailed coefficients and approximated coefficients. Then, the logarithmic slope of measured data and a threshold are calculated to identify the noise in the detailed coefficients; the corresponding detailed coefficients are

25 processed to reduce the noise. Finally, the undisturbed data are reconstructed using inverse stationary wavelet transform. The third method presents an exponential fitting-adaptive Kalman filter to remove mixed electromagnetic noises(Ji et al., 2017), while preserving the signal characteristics. It consists of an exponential fitting procedure and an adaptive scalar Kalman filter. The adaptive scalar Kalman uses the exponential fitting results in the weighting coefficients calculation.

Another wavelet-based baseline drift correction method for grounded electrical source airborne transient electromagnetic
 signals(Wang et al., 2013), through simulations, this method can improve the signal-to-noise ratio. Simulation results show that the wavelet-based method outperforms the interpolation method.

The aforementioned kalman filter and wavelet transform are universal traditional filtering methods, and have their own defects.

However, the SFS itself has distribution characteristics, and the distortion of the waveform generated by the noise causes deviation from the signal point of the distribution.

The theoretical research indicates that (Bengio et al.,2006), the incomplete representation of autoencoders will be forced to capture the most prominent features of the training data, and the high order feature of data is extracted, so autoencoders can be applied to the feature extraction and abstract representation of SFS.

5

- Theoretical research also shows that (Vincent et al.,2008)), Denoising Autoencoders (DAE) can map the damaged data points to the estimated high probability points according to the data characteristics, and achieve the target of repairing the damaged data. Therefore, DAE can be appplied to map the SFS data points that will be disturbed by noise to the estimated high probability points, to achieve the purpose of SFS noise reduction.
- 10 Studies have found (Vincent et al.,2010)the stacked DAEs (SDAE) have a strong feature extraction capability, and can improve the feature extraction and enhance the ability of calibrating the deviation points disturbed by noise. SDAE is also commonly used in the compression encoding of the pre-processing height of complex images (Ali et al.,2017).

We also noticed that supervised learning performs well in classification problems such as image recognition and semantic understanding(He et al., 2016, Long et al., 2014). At the same time, unsupervised learning also has a good performance in

- 15 clustering and association problems (Klampanos et al., 2018.), and the goal of unsupervised learning is usually to extract the distribution characteristics of the data in order to understand the deep features of the data (Becker et al., 1996, Liu et al., 2015). Both supervised learning and unsupervised learning have their own well-behaved areas, so we need to choose different learning styles and models for different problems. For the noise suppression problem of the SFS in TEM, our goal is to extract the deep features, and map the data points affected by noise to the estimated high probability points according to their own
- 20 signal features. We also found that the purpose of extracting the distribution characteristics of the SFS data is similar to that of unsupervised learning. Meanwhile, unsupervised learning models are widely used in different signal noise reduction problems.

Therefore, based on the study of the distribution characteristics of the Secondary Field data and autoencoder denoising method, we propose SFSDSA, a Secondary Field Signal Denoising Stacked Autoencoders, which is a deep learning model of transient electromagnetic signal denoising

25 (1)SFSDSA will be stacked by multiple AEs to form a deep neural network of multilayer owe complete coding, and multiple AEs are used as a higher-order feature extraction part, which can utilize its deep structure to maximize the characteristics of secondary field data.

(2)Based on the principle of DAE, SFSDSA will set the secondary field measured data (received data)as the input data, and geophysical inversion method is used to process the measured data of the secondary field to obtain the inversion signal as the

30 clean signal data. SFSDSA maps the signal points of the noise interference to the high probability points with clean signal as reference according to the deep characteristics of the signal. Because maintaining the original data dimension is especially important for the undistorted and post-processing of the signal, it is necessary to set the original dimension after the last coding as the output layer dimension. Although the output method may produce the decoding loss, it can have high abstract retention of the secondary field data characteristics, and map the affected signal points to the high probability position points. (3)Adam algorithm(Kingma et sl., 2014), introducing regularized loss and scaled exponential linear units(SELU) (The problem of too many nodes dying is a general disadvantage for RELU activation function and improved RELU activation functions like Leaky RELU all consistently outperform the RELU in some tasks(Xu et al., 2015). Therefore, it is necessary to apply the improved RELU function to reduce the impact of the shortcomings of the RELU function. We choose the SELU

- 5 that have the preponderances of overcoming vanishing and exploding gradient problems in a sense and the best preforming in full connection networks(Klambauer et al., 2017)as activation function, are applied to SFSDSA model,2014). We chose Adam algorithm, which have the advantages of calculating different adaptive learning rates for different parameters and requiring little memory(Kingma et sl.,2014). Meanwhile, introducing regularized loss to solve the problem problems of over-fitting due to increased depth and the problem of too many dying nodes causing the signal to be unusable, and the problem of SFSDSA
- 10 only learning an identity function.

3 Mathematical Derivation of SFSDSA

Firstly, the secondary field data(actual detection signal) are treated as a noisy input. Since the secondary field data are mainly a time-amplitude value, we can sample the signal as a point-amplitude value, in the form of matrix A, the dimensions are $1 \times N$:

15
$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n-1} & a_{1n} \end{bmatrix}$$
 (1)

Secondly, the geophysical inversion method is used to obtain the theoretical signal, which can be used as a clean signal, then the theoretical signal is sampled as point-amplitude value, in the form of matrix \tilde{A} , the dimensions are $1 \times N$:

$$\tilde{A} = \begin{bmatrix} \tilde{a}_{11} & \tilde{a}_{12} & \cdots & \tilde{a}_{1n-1} & \tilde{a}_{1n} \end{bmatrix}$$
(2)

Thirdly, SFSDSA training model can be built, and Adam, which is a stochastic gradient descent (SGD) method, is applied to
prevent gradient disappearance, and regularization loss is used to prevent over-fitting and SELU activation function is utilized to prevent too many points of depthdeath.

$$g_{\theta}(a_{1n}) = f_{SELU}(Wa_{ln} + b) \tag{3}$$

$$g_{\theta}(a_{1n}) = \lambda \begin{cases} Wa_{1n} + b & a_{1n} > 0\\ \alpha e^{Wa_{1n} + b} - \alpha & a_{ln} <= 0 \end{cases}$$

$$\tag{4}$$

Where $\theta = (w, b)$, w denotes the $N \times N'$ parameter matrix (N' < N), b denotes the offset of the N' dimensions. After the first compression coding layer, the signal is extracted features to $1 \times N'$. In order to extract high-level features while removing as much noise as possible and other factors, we can compress again.

$$g_{\theta'}(a_{1N'}^{'}) = \lambda \begin{cases} W^{'}a_{1N'}^{'} + b^{'} & a_{1N'}^{'} > 0\\ \alpha e^{W^{'}a_{1N'}^{'} + b^{'}} - \alpha & a_{1N'}^{'} <= 0 \end{cases}$$
(5)

w denotes the $N' \times N''$ parameter matrix (N'' < N'), and b denotes the offset of the N'' dimensions, and features of actual detection signal is extracted again, after more feature extraction layers can be stacked. For the secondary field signal, it is necessary to maintain the same input and output dimensions to ensure that the signal is not distorted and later processed. When feature extraction reaches to a certain extent, it is necessary to reconstruct back to input dimensions.

5

Reconstruction can be regarded as the process that the noisy signal points map back to the original dimensions after features being highly extracted .At the same time, reconstruction is the process of signal characteristic amplification. Finally output matrix \overline{A} with the same dimensions as the inputs can be got :

$$\bar{A} = \begin{bmatrix} \bar{a}_{11} & \bar{a}_{12} & \cdots & \bar{a}_{1n-1} & \bar{a}_{1n} \end{bmatrix}$$

$$\tag{6}$$

The output A we obtained can be used to get the loss from the clean signal A using the loss function. The general loss function
has square loss, which is mostly used in the linear regression problem. However, the secondary field data are mostly non-linear, and absolute loss is used in this paper:

$$L(\bar{A},\tilde{A}) = |\bar{A} - \tilde{A}| \tag{7}$$

In the meantime, regularization loss optimization is used in this paper, in order to avoid over-fitting model, then:

$$loss = \theta^*, \theta^{'*} = \arg_{\theta, \theta^{'}} \min \frac{1}{n} \sum_{i=1}^n L(x^i, g_{\theta^{'}}(f_{\theta}(x^i))) + \lambda R(w)$$

$$\tag{8}$$

15 After the loss is calculated, Adam algorithm is used to reverse optimization of parameters.

Figure 1 is the algorithm structure diagram of SFSDSA. With reference to the theory of DAE, SFSDSA maps the signal points of the noise interference to the high probability points with clean signal as reference according to the deep characteristics of the signal, so as to realize the signal noise and reduce noise interference. This high probability position is determined by the theoretical clean signal and the multi-layer model and the feature extraction ability. The multi-layer feature extraction makes

20 the deep feature of secondary field data be preserved, and the effect of noise is reduced.

4 Experiment and Analysis

In this paper, the secondary field signal of a certain place is used as the experimental analysis signal. Usually, the secondary field signals can be obtained continuously for a period of time, so a large number of signals can be extracted conveniently as the training samples.

The secondary field actual signals are extracted as 1×434 as input signals of noise pollution, as is shown in the Figure 2(a). At the same time, based on the secondary field actual signals, the geophysical inversion method is used to obtain the theoretical detection signal as clean signal uncontaminated by noise, as is shown in the Figure 2(b).

In order to be able to highlight the differences between the data, data are expressed in a double logarithmic form(loglog), as is shown in the Figure 3(a) and Figure 3(b).





The deep features of original data are abstracted by features extraction layers(compression coding layers). As the number of layers increases, SFSDSA can be a more complex abstract model with limited neural units, (to get higher-order features for this small-scale input in this paper), and the more features extraction layers will inevitably lead to over-fitting. Moreover, the reconstruction effect can be affected by the number of features extraction layer nodes. If SFSDSA model has too few nodes, the characteristics of the data can not be learned well. However, if the number of features extraction layer nodes are too large,

5

the designed lossy compression noise reduction can not be achieved well and the learning burden is increased.

Therefore, based on the aforementioned questions, we design the SFSDSA model(Figure 1), and the number of nodes in the latter features extraction layer is half the number of nodes in the previous features extraction layer, until finally reconstructed back to the original dimension. SFSDSA model is a layer-by-layer features extraction, which can be regarded as a stacked AE process. Low dimensions are represented by the high-dimensional data features, which can learn the input features.

10

At the same time, since the reconstruction loss is the loss of the output related with the clean signal, it can also be said that the input signal can be regarded as a clean signal based on the noise, the training measure of DAE model increases the



Figure 2. (a)Actual signal curves. (b) Inversion of theoretical signal curves.

robustness of the model and reconstructs the lossy signal, and mapping the signal point to its high probability location can be viewed as a noise reduction process.

In the training experiment, we collected 2400 periods of transient electromagnetic method secondary field signals from the same collection location, and selected 434 data points in per period. Meanwhile, 100 periods of signals are randomly

5 acquired as a test and validation set for the improving the robustness of the model. We use Google's deep learning framework – Tensorflow. The parameter settings for the model are as follows: batch-size = 8, epochs = 2. We do a grid search and get the good parameter combination of learning rate and regularization rate as shown in table 1(learning rate=0.001 and regularization rate=0.15).

We analyzed and compared the selection of the two loss functions of MAE and MSE in experiments as shown in figure
4. Meanwhile, according to the previous work and the SFS denoising task of transient electromagnetic method, we think that MAE is a better choice. On the one hand, our task is to map the outliers affected by noise to the vicinity of the theoretical signal point, in other words, model should ignore the outliers affected by noise to make it more consistent with the distribution of the overall signal. We know that MAE is quite resistant to outliers(Rishabh, 2015). On the other hand, the squared-error is going to



Figure 3. (a)Actual detection signal. (b) Inversion of theoretical detection signal.

be huge for outliers, which tries to adjust the model according to these outliers on the expense of other good-points(Rishabh, 2015). For signal that are subject to noise interference in the secondary field of transient electromagnetic method, we don't want to over-fitting outliers that are disturbed by noise, but we want to treat them as noise interfered data.

The evaluation index is the mean absolute error(MAE) of output reconstruction data and clean input data. The smaller the MAE, the closer the output reconstruction data is to the theoretical data. The model also performs better in noise reduction.

$$MAE(X,h) = \frac{1}{m} \sum_{i=1}^{m} |h(x^{(i)} - y^{(i)})|$$
(9)

In this paper, we use Google's deep learning framework, Tensorflow.

5

In the previous experiments, we set hyper-parameters (batch-size=8, learning-rate=0.1, regularization-rate=0, epochs=20) based on experience but we initially take the measure of a small number of epochs (epochs=2) according to experiment. We

10 added the experiment as shown in Figure 2 to support our standpoint. The model oscillates quickly and converges. Training with fewer epochs can avoid useless training and over-fitting, maintaining the distribution characteristics of the signal itself.



Figure 4. the training cost comparison of MAE and MSE, the left is the training cost of fine-tuning, and the right is the training cost of reconstructing.

As shown in Figure 5(in original manuscript), the reconstruction error oscillates and converges as the training progresses. This phenomenon is similar to the tail of the actual signal. We try stoping training when the convergence occurs, the idea similar to early-stopping makes the model more robust(Caruana ,2000).

By analyzing Figure 45, the relationship between MAE and the number of hidden layers, we found that MAE was the
smallest when the hidden layer was 2, the result of stacking two AEs have good effect. We guess that the size of the AE hidden layer is too small after multiple stacking(for instance, the 4th AE only has 27 nodes because the size of latter AE is half of the previous AE in order to extract the better feature), and the representation of signal characteristics are not complete resulting in large reconstruction costs. If we want to get a better result, more iterations may be used but this tends to cause over-fitting. Meanwhile, we found that the reconstruction loss of the second AE is already very small shown in Figure 6. And it is not necessary to stack more AEs.

Small-scale deep learning model, and the training times can be less. By analyzing Figure 2(a), we found that because the amplitude of the tail of the actual signal is small, and the influence of the noise is significant, so the tail of the signal oscillates violently. As is shown in Figure 5, we set each batch to 10 and every 10 training steps output MAE value as a sampling point. We found that the MAE values near the 150th sampling point start to oscillate. Based on these two oscillation relationships,

15 we can make some conclusions that training and reverse update parameters, and reconstruction loss is oscillating, so learning will fall into the bottleneck.

Meanwhile, after the feature extraction and noise reduction to a certain extent, the noise interference can not be completely removed, and the reconstruction can not completely present the clean signal, and it is only possible to map the signal points as high probability points as possible to reduce reconstruction loss.



Figure 5. SFSDSA hidden layer number and MAE values.





4.1 Training results

After several experiments, the MAE of actual signals fell from 534.5 to about 215.Compared with the secondary field actual signals and signals denoised by SFSDSA model, the noise reduction effect of SFSDSA is obvious in Figure 6.

The 35th to 55th points are selected for specific analysis in Figure 7. Through noise reduction of training good SFSDSA model, the singular points(large amplitude deviation from theoretical signal) affected by the noise map to the high probability



Figure 7. Error-Iteration sampling pointpoints.



Figure 8. Actual secondary field data after SFSDSA model noise reduction

Table 1. The training cost of combination of learning rate and regularization rate. The value represents the MAE of the first fifty data points.

 According to the experience, about the first fifty data points have better effect for extracting time-domain order waveforms

regularization rate	0.05	0.1	0.15	0.2
<u>0.1</u>	61515.3	12670.3	14448.9	11112.1
0.01	1735.2	<u>1918.1</u>	2126.6	1825.7
0.001	1526.6	1669.5	1377.3	1780.6
0.0001	1493.2	1678.1	1392.3	1955.5



Figure 9. Specific points analysis.

positions(e.g., no. 38, no. 51 points). This process is the process of damage reconstruction that the DAE model has verified. At the same time, our stacked AEs model also keeps on extracting the features, and the singular points are restored to the corresponding points according to the characteristics of the data. The whole process realizes the noise reduction of the secondary field actual signal based on the secondary field theoretical signal, and the model maps the singular points to locations where there is a high probability of occurrence, which is also similar to the most estimative method based on observations and model

5 there is a high probability of occurrence, which is also similar to the most estimative method based on observations and mode predictions by Kalman filtering.

5 Comparison with traditional noise reduction methods

We also conducted wavelet transform and Kalman filter experiments, in which the number of layers of the wavelet transform is three, DWT () and construction function IDWT () is called in Matlab.

10 Kalman filtering is implemented in Python, where the system noise Q is set to 1e-4 and the measurement noise R is set to 1e-3. The Figure 8 shows the absolute error distribution for that method. We can find from the figure model of noise reduction based on SFSDSA of secondary field data, SFSDSA is better than kalman filter and wavelet transform. At the same time as the kalman filter is a linear filter, its noise reduction effect is so poor in this paper.

By using the PCA method, we do the experiment to verify the effect of noise reduction. But the process of programming 15 is more complicated using mathematical derivation, so we use scikit-learn library to realize noise reduction. However, the underlying structure is not easy to modify resulting in scikit-learn library is unable to adjust parameters adaptively based on signal characteristics. After the filtering test, and then the MAE corresponding to the calculation of the theoretical data, it can be seen that the effect of pca filtering is lower than SFSDSA.



Figure 10. (a)Kalman filter. (b)Wavelet transform filter. (c)PCA filter. (d)SFSDSA denoising.

Model Name	Parameter Setting	MAE
SFSDSA+SELU+REGULARAZTION+ADAM	Learning_rate_base:0.00103	150.36
SFSDSA+RELU+REGULARAZTION+ADAM	Learning_rate_base:0.00103	1500.20
SFSDSA+SELU+ADAM	Learning_rate_base:0.00103	164.30
SFSDSA+SELU+REGULARAZTION	Learning_rate_base:0.00103	5112.30
Wave Transform	Three layers of wavelet transform	451.20
Kalman Filter	Q=e-4 R=e-3	503.20

At the same time, we compared the optimization results of various models using the traditional method with those of the SFSDSA model in table $\frac{1.2}{2}$.



Figure 11. The geographic distribution of the collection points(1th to 7th).

Figure 9-11 is the diagram of the mine where the exploration experiment was conducted. The red thick curve is the actual mine vein curve. A data collection survey line, which is the southwest-northeast pink curve shown in the figure, is designed with seven points marked as number 1 to 7 along it, and the distance between each point is 50 meters.

In the data analysis, we analyzed the first 50 points in the second field which collected in actual mine. The early signal of the secondary field is stronger than later, and it is not easy be disturbed by the noises. So in the Figure 10, we take out the later 21 points in each collection point, which is used for further analysis. Figure 10 (a) is extracted time-domain order waveforms formed by the actual data acquired at the seven collection points at the same time. Figure 10 (b) extracted time-domain order

10 waveforms formed by the data denoised by SFSDSA model. By comparing the two images in Figure 10, it can be clearly seen that the curves in Figure 10(a) have obvious intersections, and the intersections in Figure 10(b) can't be seen almost.In



Figure 12. (a)The Original 30th to 50th points from seven actual detecting locations. (b)The denoising 30th to 50th points from seven actual detecting locations.

transient electromagnetic method, the intersected curve can't indicates the information of the deeper underground geological information. It can be explained that the curve after the de-noising model can reflect the deep geological information.

6 Conclusions

Based on the transient electromagnetic method, deep-level surveys are conducted. The deep-seated information is reflected in the late-stage data of the second field. The signals are very weak and easily contaminated by noise. The measured data are modeled to obtain the theoretical model. The model will perform noise reduction based on the geological features represented by the previous training data-set. Therefore, it is necessary to analyze the known geological features more carefully and apply the model according to the actual geological conditions before using our method. At the same time, this view is consistent with machine learning theory(Neyshabur et al., 2017). This method has good generalization for different collection points of the

10 <u>same geological feature area</u>. By introducing the deep learning algorithm integrated with the characteristics of the secondary field data, it can map the contaminated data in late track data to a high probability position. By comparing several filtering algorithms, in which the same sample data are used, the stack noise reduction from the encoder method can reduce the MAE,

thereby reducing the noise, and it is conducive to the subsequent pumping processing to further improve the effective detection depth.

Code availability. The code of the SFSDSA are available upon request(15775118240@163.com or linfq@cdut.edu.cn)

Data availability. The data used in this paper are available upon request(15775118240@163.com or linfq@cdut.edu.cn)

5 Appendix A

A1

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This paper is supported by National Key R&D Program of China (NO.2018YFC0603300).

References

Ali, A., Fan, Y., Shu, L.:Automatic modulation classification of digital modulation signals with stacked autoencoders, Digital Signal Processing., 71,108-116,DOI: 10.1016/j.dsp.2017.09.005,2017.

Becker S, Plumbley M.: Unsupervised neural network learning procedures for feature extraction and classification[J]. Applied Intelligence:

- 5 The International Journal of Artificial, Intelligence, Neural Networks, and Complex Problem-Solving Technologies, 6(3):185-203,1996.
 - Bengio, Y., Lamblin, P., Dan, P., Larochelle, H.: Greedy layer-wise training of deep networks, International Conference on Neural Information Processing Systems., MIT Press, 153-160, 2006.

Caruana R, Lawrence S, Giles L.,: Overfitting in neural nets: backpropagation, conjugate gradient, and early stopping, International Conference on Neural Information Processing Systems. MIT Press, 2000.

- 10 Chen, B., Lu, C.D., Liu, G.D.: A denoising method based on kernel principal component analysis for airborne time domain electro-magnetic data. Chinese Journal of Geophysics., 57(1), 295–302, DOI: 10.1002/cjg2.20087, 2014.
 - Danielsen J E, Auken E, Jørgensen F, et al. The application of the transient electromagnetic method in hydrogeophysical surveys, Journal of Applied Geophysics., 53(4),181-198,DOI: 10.1016/j.jappgeo.2003.08.004,2003.

Dai, W., Brisimi, T.S., Adams, W.G., Mela, T., Saligrama, V., Paschalidis, I.C.: Prediction of hospitalization due to heart diseases by supervised

15 learning methods, International Journal of Medical Informatics...84,189-197,2015. Grais, E. M., Plumbley, M. D: Single Channel Audio Source Separation using Convolutional Denoising Autoencoders,Proceedings of the IEEE GlobalSIP Symposium on Sparse Signal Processing and Deep Learning, 5th IEEE Global Conference on Signal and Information Processing,14-16,DOI:10.1109/GlobalSIP.2017.8309164 ,2017.

Haroon, A., Adrian, J., Bergers, R., Gurk, M., Tezkan, B., Mammadov, A.L., Novruzov A. L.: Joint inversion of long-offset and central-loop tran-

20 sient electromagnetic data: Application to a mud volcano exploration in Perekishkul, Azerbaijan, Geophysical Prospecting.,63.2,478-494, DOI: 10.1111/1365-2478.12157,2014. Rasmussen, S., Nyboe, N. S., Larsen J. J. :Extraction and Use of Noise Models from Production-Mode Transient Electromagnetic Data, Geophysics., 83(1), 1-40,DOI:10.1190

He K , Zhang X , Ren S , et al. Deep Residual Learning for Image Recognition. IEEE Conference on Computer Vision and Pattern Recognition, CVPR., 2016, DOI:10.1109/geo2017-0299.1, 2017. Ji, CVPR.2016.90

- 25 Hwang, Y., Wu, Q., Wang, Y.,Lin,J.,Li,D.,Du,S.,Yu,S.,Guan,S.:Noise reduction of grounded electrical source airborne transient electromagnetic data using an exponential fitting-adaptive Kalman filter,Exploration Geophysics.,DOI: 10.1071/EG16046 ,2017. Tong, A., Choi, J.:Automatic construction of nonparametric relational regression models for multiple time series,International Conference on Machine Learning, 3030-3039,2016.
 - Ji, Y., Li, D., Yu, M., Wang, Y., Wu, Q., Lin, J. : A de-noising algorithm based on wavelet threshold-exponential adaptive window
- width-fitting for ground electrical source airborne transient electromagnetic signal, Journal of Applied Geophysics., 128,1-7,DOI: 10.1016/j.jappgeo.2016.03.001,2016. Li, D., Wang,
 - Ji, Y., Wu, Q., Wang, Y., Lin, J., Yu, S., Ji, Y.: Electromagnetic noise reduction in grounded electrical-source airborne transient electromagnetic signal using a stationarywavelet- based denoising algorithm, Near Surface Geophysics., 1(-9), DOI: 10.3997/1873-0604.2017003 2017. Wu, Y., Lu, C. Li, D., Du, X.Z.S., Yu, X.D.: A denoising method based on principal component analysis for S. Guan, S.; Noise
- 35 reduction of grounded electrical source airborne transient electromagnetic data , Computing Techniques for Geophysical and Geochemical Exploration .,36(2),170-176,DOI: 10.3969using an exponential fitting-adaptive Kalman filter, Exploration Geophysics.,DOI: 10.1071/j.issn.1001-1749.2014.02.08,2014. EG16046 ,2017.

- Jifara, W., Jiang, F., Rho, S.,Liu,S.: Medical image denoising using convolutional neural network: a residual learning approach, Journal of Supercomputing, The Journal of Supercomputing, 6,1-15,DOI:10.1007/s11227-017-2080-0,2017. Grais, E. M., Plumbley, M.D: Single Channel Audio Source Separation using Convolutional Denoising Autoencoders, Proceedings of the IEEE GlobalSIP Symposium on Sparse Signal Processing and Deep Learning, 5th IEEE Global Conference on Signal and Information
- 5 Processing,14-16,DOI:10.1109/GlobalSIP.2017.8309164 ,2017. Shen, H., George, D., Huerta, E.
 Kingma, D. P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint,arXiv:1412.6980,2014.
 Klambauer, G., Unterthiner, T., Mayr, A.,Zhao, Z.: Denoising Gravitational Waves using Deep Learning with Recurrent Denoising Autoencoders, arXiv Hochreiter, S.: Self-Normalizing Neural Networks, arXiv preprint.,arXiv:1706.02515:1711.09919v11706.0251,2014.
- 10 Klampanos I A, 2017. Hwang, Y., Tong, A., Choi, J. : Automatic construction of nonparametric relational regression models for multiple time series, International Conference on International Conference on Machine Learning. 3030-3039,2016. Vincent, P., Larochelle, H., Bengio, Davvetas A, Andronopoulos S, et al. Autoencoder-driven weather clustering for source estimation during nuclear events. Environmental Modelling & Software, 102:84-93,2018.

Li, D., Wang, Y., Manzagol, P.A.: Extracting and composing robust features with denoising autoencoders, International Conference

- 15 on Machine Learning., ACM, 1096-1103, DOI: 10.1145/1390156.1390294, 2008. Zhao, Mingbo., Chow, T.W.S., Zhang, Z., Li, B.Lin, J., Yu, S., Ji, Y.: Electromagnetic noise reduction in grounded electrical-source airborne transient electromagnetic signal using a stationary wavelet-based denoising algorithm, Near Surface Geophysics., 1(-9), DOI: Automatic image annotation via compact graph based semi-supervised learning, Knowledge-Based Systems., 76,148-165, DOI: 10.101610.3997/j.knosys.2014.12.014,1873-0604.2017003 2017.
- 20 Liu J H , Zheng W Q , Zou Y X:A Robust Acoustic Feature Extraction Approach Based on Stacked Denoising Autoencoder, IEEE International Conference on Multimedia Big Data. IEEE Computer Society, 2015. Dai, W.,Brisimi, T
 - Long J., Shelhamer E., Darrell T.: Fully Convolutional Networks for Semantic Segmentation, IEEE Transactions on Pattern Analysis & Machine Intelligence., 39(4):640-651,2014.

Neyshabur B, Bhojanapalli S, Mcallester D, et al.: Exploring Generalization in Deep Learning, NIPS. 2017

35

- 25 Rasmussen, S., Nyboe, N. S., Adams, W. G. Larsen J. J.: Extraction and Use of Noise Models from Production-Mode Transient Electromagnetic Data, Geophysics., 83(1), 1-40, DOI: 10.1190/geo2017-0299.1, Mela, T., Saligrama, V., Paschalidis, I. C.: Prediction of hospitalization due to heart diseases by supervised learning methods, International Journal of Medical Informatics. 2017.
 - Shen, H., George, D., Huerta, E. A., 84, 189-197, 2015. Zhao, Z.: Denoising Gravitational Waves using Deep Learning with Recurrent Denoising Autoencoders, arXiv:1711.09919v1, 2017.
- 30 Shimobaba T, Endo Y, Hirayama R, et al. Autoencoder-based holographic image restoration, Applied Optics.,56(13),F27,DOI: 10.1364/AO.56.000F27,2017. Chen, B., Lu, C.D., Liu, G.D.:A denoising method based on kernel principal component analysis for airborne time domain electro-magnetic data, Chinese Journal of Geophysics., 57(1), 295–302,DOI: 10.1002/ejg2.20087

Vincent, P., Larochelle, H., 2014. Bengio, Y., Lamblin Manzagol, P., Dan, P., Larochelle, H.: Greedy layer-wise training of deep networks A.: Extracting and composing robust features with denoising autoencoders, International Conference on Neural Information Processing

Vincent, P., Larochelle, H., Lajoie, I., Manzagol, P.A.: Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion, Journal of Machine Learning Research., 11(12), 3371-3408, 2010. Ali, A., Fan

Systems., MIT Press, 153-160,2006. Machine Learning., ACM, 1096-1103,DOI: 10.1145/1390156.1390294,2008.

- Wang, Y., Ji, Y.J. Li, S.Y., Lin, J., Zhou, F.D., Yang, G.H.,: A wavelet-based baseline drift correction method for grounded electrical source airborne transient electromagnetic signals. Exploration Geophysics,44,229–237, Shu,L.:Automatic modulation elassification of digital modulation signals with stacked autoencoders, Digital Signal Processing., 71, 108-116, DOI: 10.10162013.
- Wu, Y., Lu, C. D., Du, X. Z., Yu, X.D.: A denoising method based on principal component analysis for airborne transient electromagnetic data. Computing Techniques for Geophysical and Geochemical Exploration..36(2),170-176,DOI: 10.3969/i. dsp.2017.09.005,2017. Kingma,D.
- P. j. issn. 1001-1749.2014.02.08, Ba, J.: Adam: A method for stochastic optimization.arXiv preprint, arXiv: 1412.6980, 2014. Klambauer, G.

5

Zhao, M.B., Chow, Unterthiner, T.W.S., Zhang, Z., Li, B.: Automatic image annotation via compact graph based semi-supervised learning, Knowledge-Based Systems., 76, Mayr, A., Hochreiter, S.: Self-Normalizing Neural Networks, arXiv preprint, arXiv: 1706.02515:1706.0251, 2014. 148-165, DOI: 10.1016/j.knosys.2014.12.014.2015.