SnT 2019
CTBT: SCIENCE AND TECHNOLOGY CONFERENCE

# Study of variants for seismic data pre-processing which are not leading to significant losses of information that may be needed

*Kislov K.V.*[1] (kvkislov@yandex.ru),

*Gravirov V.V.*[1, 2] (gravirov@rambler.ru)

1. Institute of Earthquake Prediction Theory and Mathematical Geophysics Russian Academy of Science (IEPT RAS), http://www.mitp.ru

2. The Schmidt Institute of Physics of the Earth of the Russian Academy of Sciences (IPE RAS), http://www.ifz.ru

T3.5-P68

## INTRODUCTION

Brain waves are out there carrying thoughts about a unified configuration tool for analysis of seismic data. We mean a multi-task Deep Neural Network (DNN) [*Kislov, Gravirov,* 2018; *Ross et al.,* 2018; *Yuan et al.,* 2018; *Meyer et al.,* 2019]. Due to the specifics of the seismic data and the range of tasks which cannot be identified in advance a very complex network will be required.

To simplify the network architecture and facilitate its training, it is necessary to carry out the data pre-processing. The STA/LTA method is widely used for it. Other technologies are also known [*Gentili, Michelini*, 2006; *Kislov*, *Gravirov*, 2010; *Kislov*, *Gravirov*, 2011]. The second stage depends on the objectives of the analysis. Each time, a new method of data preprocessing is developed. This creates an obstacle to the using of DNN's multi-task.

Thirty numerical parameters are enough to characterize a ten-second seismic recording. If we could measure them correctly, we would the mysteries of the Earth, but not the secrets of signals. If the pre-processing marks out features of a signal and allows it to be viewed in a smaller space, this will simplify the post-processing algorithms and their learning.

It is necessary to create the pre-processing algorithm, which will not lose the informative features of the signal. It can be used in any analysis, but the multi-task DNN is primarily meant.

## DATA PRE-PROCESSING METHODS

The data pre-processing includes data formatting, data deletion and/or restoration, extraction of informative features and scaling of the feature vector [*Kong et al.,* 2018; *Kotsiantis et al.,* 2006].

Most of all, we are interested in extracting informative features. For this, signal transformation can be used without dimension reduction [*Gravirov et al.,* 2013]. The way of extraction features can have physical interpretation. This increases the visibility of the method, provides opportunity to control the processing of data. On the other hand, features that do not have a clear physical meaning can provide better recognition capability. The method of data pre-processing should provide:

- noise reduction,
- data compression,
- signal reconstruction without noise,
- the ability to work in a real time.

## THE EXTRACTING OF INFORMATIVE FEATURES OF DATA

Let a seismic signal $w_i \in W^{*N}$ has a length of $N$ samples. As a result of the transformation $v_i = f(w_i)$, we get another representation having $M$ counts long, which shows the features of the seismic signal. All set of possible signals $W^{*N}$ is mapped onto a set of representations $V^M = F(W^{*N})$. There is an inverse transform $\acute{w}_i = g(v_i)$; $W^{*N} = G(V^M)$.

Errors of transformations cause that $\acute{w}_i$ is slightly different from $w_i$. This difference should not exceed some value $E$. First and foremost, the reconstructed signal should show the same seismic event, with the parameters in the corresponding tolerance. The formal procedure for determining the error E, for example, the ordinary least squares technique, is ill-suited here.

Since the original signal contains noise, the real set W exceeds the ideal W*. In addition we agree with some loss of information at direct transformation in a pursuit of the compact representation of $v_i$, i.e., close input signals will have the same representation. Direct and inverse transformations have to be stable with rather small disturbances at the inputs. Then the analysis which is carried out in a time domain can be equally possible in the transformation domain.

## COMPRESSIVE SAMPLING

Consider a method called Compressed Sensing, Sparse Sampling, or Compressive Sampling (CS). In geophysics, this method has been used for about 40 years[*Taylor et al.*, 1979; *Oldenburg et al.*, 1983; *Gan et al.*, 2016].

Transform the signal $w^N$ into a space of low dimensionality $y_i = \Phi w_i$. Matrix of linear map $\Phi^{M \times N}$ acts as an encoder. Let's represent the signal via some basis $w = \Psi x$, where the columns of the matrix $\Psi^{N \times N}$ are the coordinates of the basis $\Psi$, and column vector $x^N$ contains the coefficients that represent the signal $w^N$ in the basis $\Psi$. The representation $x^N$ will be called s-sparse if $x$ is a linear combination of $s$ basis vectors only, i.e. with the condition $\|x\|_{\ell 0} \le s$. The vector $x$ is called compressible if the $s$ components are large enough, and the others are small. The CS uses the rule that all natural signals are rare when they are expressed properly [*Candès, Wakin*, 2008]. When M <N, the equation $y - \Phi z = 0$ has an infinite number of solutions $z$. To restore the original signal, CS uses the information that x is compressible.

The signal coding is a simple linear procedure $y = Ax$, where the measurement matrix $A = \Phi\Psi$ must satisfy the Restricted Isometry Property (RIP). The fulfillment of this condition with a high probability can be achieved by choosing a matrix A with random elements distributed, for example, according to the normal law or Bernoulli's law (randomization of the process). Random matrices are largely incoherent with any fixed basis $\Psi$ [*Bourgain et al.*, 2011]. The less coherence, the less $M$ can be. But random matrices are not optimal or stable and other methods are often used [*Gurbuz et al.*, 2009; *Robucci et al.*, 2010].

Compression reduces the noise (small level noise), which in most cases has no sparse representation [*Ben-Haim et al.*, 2010].

Whether the signal can be correctly restored from the compressed vector $y$ depends on the measurement matrix, the compressed vector $y$ itself and the reconstruction algorithm. Usually, $\ell$1-optimization is used as a measure of the correspondence of the recovered signal to the original one or the size of the error. The reconstructed compressible signal $\hat{x} = \arg\min \|x'\|_{\ell 1}$ will be correspond to the initial one with a high probability when $M \ge C_1 \cdot s \cdot \log(N/s)$, where $C_1$ is a constant depending on the method of creating the measurement matrix. This is a convex optimization problem, which can be considered as well as the linear programming problem, known as basis pursuit [*Foucart, Rauhut*, 2013]. To take into account data noise and transformations noise, use $\ell$1-optimization with relaxed constraints: $\min \|\hat{x}\|_{\ell 1} : \|A\hat{x} - y\|_{\ell 2} \le \varepsilon$, where $\varepsilon$ describes noise. Other algorithms are used for recovery too [*Blumensath, Davies*, 2009].

The recovery error consists of two components: the first is proportional to the data noise, the second is the approximation error. The quality of reconstruction considerably decreases with increasing the noise. It is possible to estimate the error on a wide range of parameters only by comparing the original and restored seismograms (Fig. 1).

Many types of analysis are possible without signal reconstruction [*Davenport et al.*, 2010]. This removes the recovery error. For example, the CS algorithm can easily be embedded into the neural network as an extra layer [*Li et al.*, 2018].

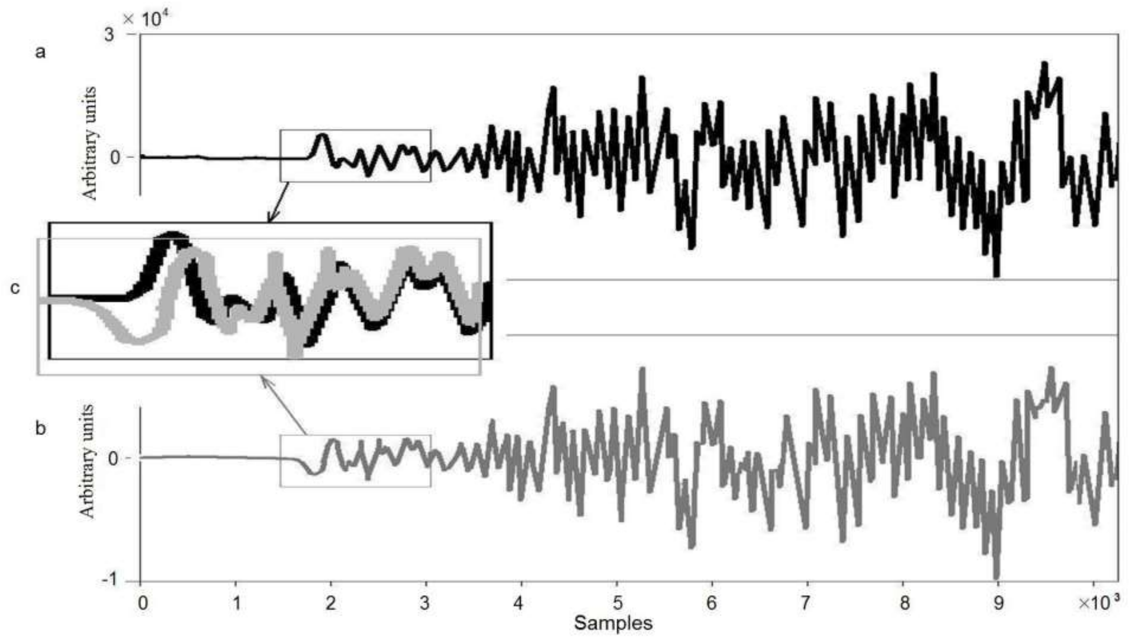


Fig. 1 a - the first arrival of the earthquake; b - an example of signal recovery; N = 1024, M = 60; c - the comparison.

CS is increasingly used in seismic exploration works and data processing of seismic networks [*Yao et al.*, 2011; *Herrmann*, 2019].

CS- method is a very effective. But its main advantages will be affected when seismic sensors will take random observations. The obstacles to using CS as the seismic data pre-processing are the requirement of low noise and the need for high sparsity [*Davenport et al.*, 2012]. In addition CS algorithms have large computational costs and their application is difficult in real-time systems.

## COMPRESSION CODING WITH TRANSFORMATION

For audio and video compression, coding with transformation is used, which also uses *s*-sparse representations. A discrete wavelet transform (DWT), discrete Fourier transform (DFT) or discrete cosine transform (DCT) are commonly used for coding with transformation.

Even at small *s*, the use of these algorithms requires a large computational cost. All *N* transform coefficients must be calculated, but most of them (excluding *s* each) are useless. The values of *s* and the numbers of their positions in the vector *w* must be encoded [*Granichin, Pavlenko*, 2010]. Compression coding with transformation clearly loses over CS.

## AUTOENCODER

In recent years, interest in autoencoders (AE) has increased again [*Creswell, Bharath*, 2019; *Shi et al.*, 2019]. In [*Valentine, Trampert*, 2012], AE was used to compress seismic signals of a length of 512 counts into a 32-dimensional vector.

The simplest AE is a neural network with one hidden layer, trained by the back-propagation method to reconstruct its own inputs ($y = x$). The input $x$ is mapped to the latent variable layer $z = h_1(Wx + \beta_0)$. Then $z$ is mapped to the output layer $y = h_2(W'z + \beta_1)$. Here h is the element-by-element activation function (for example, a sigmoidal function, a ReLU, or another); adjustable parameters: β is bias weights, W is weight matrix. The AE is trained to minimize recovery errors (for example, root-mean-square error) $E(x,y) = \|x-y\|^2 = \|x - h_2(W'(h_1(Wx+\beta_0)) + \beta_1)\|$ and is the input data mining tool. [*Kislov, Gravirov*, 2017]. Like principal component analysis (PCA), the AE translates data to a low dimensionality, but the transformation is non-linear. The AE is more promising than PCA or other known methods.

Deep AE can compress the data better. Let's define an exit of layer $i$ as $z_i$, $i \in \{1,\ldots,q+1\}$. At that, $z_0 \equiv x$; $y \equiv z_{q+1}$, where $q$ is the number of hidden layers. Then $z_i = h_i(z_{i-1}W_i + \beta_i)$. Usually deep AE contains 5 - 7 layers. In our case, layer sizes are 1000, 500, 250, 60, 250, 500, 1000 neurons. The middle layer, the 60-dimensional vector, is the compressed representation of the signal.

AE's education needs to adjust the matrix W and the biases $\beta$ to minimize some loss function $L(x, y)$. The process is simplified if the symmetric autoencoder is used, i.e. the encoding and decoding weight matrices are transposed for each other $W_i = W^T_{q-i+2}$. On the other hand, if you abandon this principle, you can reduce the error of signal reconstruction [*Majumdar*, *Tripathi*, 2017].

We are interested in Stacked Denoising Autoencoder (SDAE) which has the possibility of data compression with noise reduction. When training, an artificially noisy signal is inputted. Noise is made out as adding Gaussian noise or a random process of zeroing some of the inputs. This forces AE to reveal the strongest interrelations, at the same time not building identical maps.

Training of deep AE can be solved using a greedy layer-by-layer algorithm and unlabeled data, i.e., unsupervised learning. For long, it was assumed that these are frills, and that the random initialization of the weights, the use of the ReLU activation function, and some tricks give the chance to train network by fine-tuning only. Only that's not enough in order to get the maximum effect from AE [*Erhan et al.*, 2010].

A well-trained AE with an enhanced denoiseing version can solve many important seismological tasks. However, it does not provide effective suppression of various anthropogenic noises. The trained AE works for orders quicker than CS.

## WAVELET TRANSFORM

The wavelet transform is a convolution of the analyzed signal $x(t)$, $t \in (-\infty, \infty)$ with wavelet functions $\psi\left(\frac{t-b}{a}\right)$, $a, b \in (-\infty, \infty)$, $a \neq 0$ [*Saad et al.*, 2019]. Seismic data $x_k = x(t_k)$ is a time series of consecutive values with a constant step $\Delta t$ : $t_k = \Delta t \cdot k$, $k = 0,1,\ldots N-1$. Their discrete wavelet transformation is defined by the following expression:

$$W(a,b) = \frac{1}{n(a,b)} \sum_{k=0}^{N-1} x_k \psi\left(\frac{t_k - b}{a}\right), \quad n(a,b) = \sum_{k=0}^{N-1} e^{-\frac{1}{b}\left(\frac{t_k - b}{a}\right)^2}$$

, where $a, b \in R$; $a \neq 0$; $x(t)$ is the signal; $\psi(t)$ is the wavelet. The parameter $a$ is scale. It determines the size of the wavelet function $\psi\left(\frac{t-b}{a}\right)$.

Its analogue in Fourier analysis is the period of harmonic oscillation. The value b is shift along the time axis.

It is convenient to use elements of the local energy spectrum or scalogram as the input to the DNN . At the bottom of the scalogram (Fig. 2) there are coefficients with small numbers, giving a detailed picture of the signal, and from above – with large numbers, giving a coarse picture of the signal.

In Fig. 2a you can clearly see added complexity of the spectrum at the edges. We have developed an algorithm that implements an adaptive filter bandwidth selection and allows processing and filtering of input digital seismic data [Gravirov et al., 2014]. You can isolate the interfering low-frequency components and the splashes associated with technogenic noise and remove them from the original process (see Fig. 2b). We have stopped at fourth-order Daubechies wavelets "db4". If using them, the form of the reconstructed signal is the least different from the original.

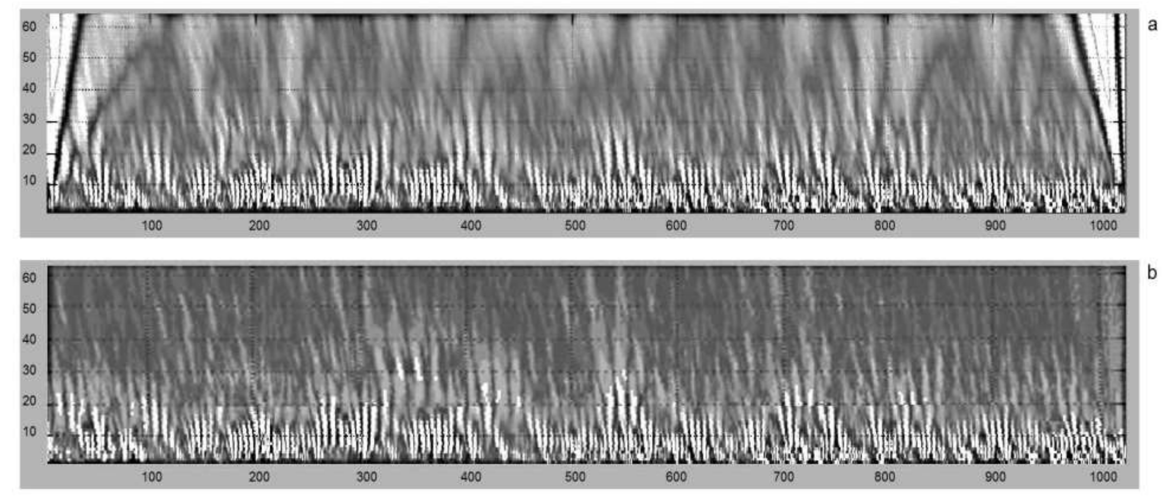


Fig. 2. Representation of a test seismic signal with a depth of decomposition of 64: a - the scalogram of the seismic window with a length of 1024 samples; b - processed scalogram of the same section of the recording.

Usually for further processing choose 5-8 the most characteristic levels of decomposition. The spectral bands can be selected based on the power distribution [*Altunkaynak*, 2014]. Decomposition on each band is a kind of signal feature map. Reconstruction of a signal according to feature maps gives the signal cleaned from noise. Reconstruction by one of the feature maps makes it possible to consider the features highlighted by this card (Fig. 3).

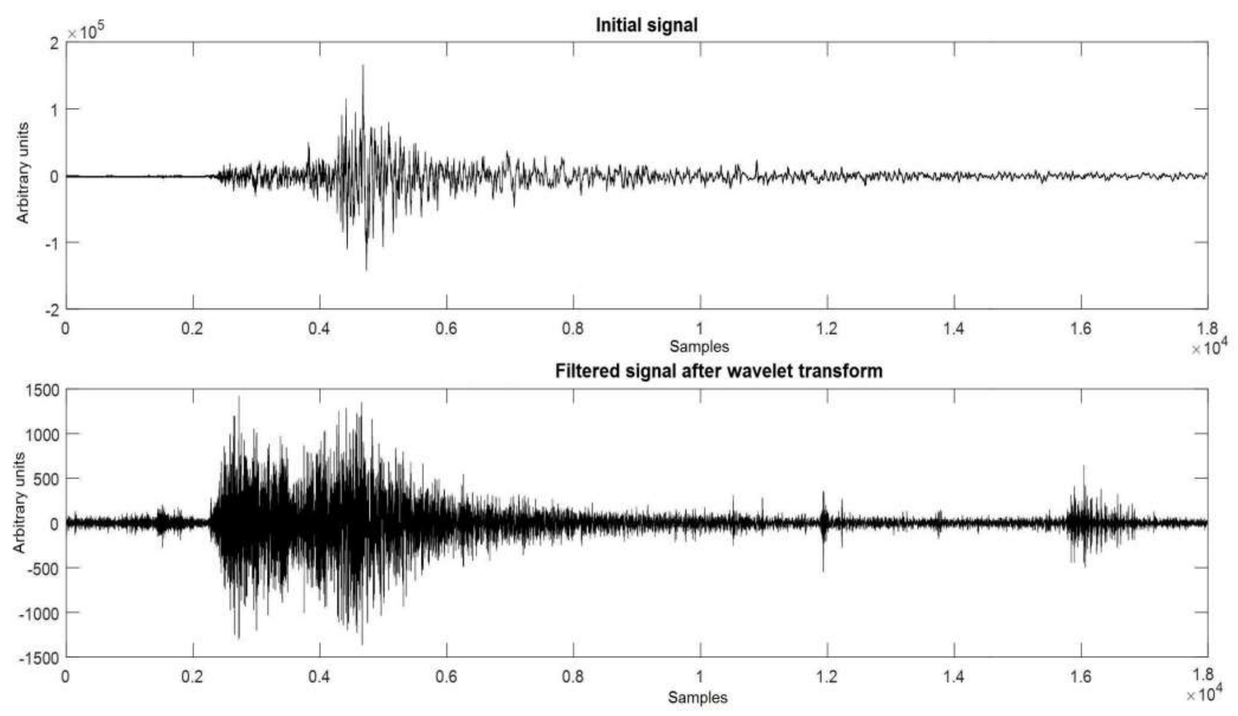


Fig. 3. a - Earthquake recording; b - The same signal, restored after the wavelet transform by feature map number 64.

The approach when the wavelet transformation applies to feature extraction is quite useful if you use the convolutional layer as the input of the DNN.

## CONCLUSION

We have explored the most promising methods, although there may be other approaches. It is not possible to make certain conclusions on the effectiveness or ineffectiveness of this or that method due to the diversity of seismological tasks, especially since new tasks may appear. The coding part of AE compresses data well, but loses in noise reduction. The wavelet as data pre-processing increases the input data for DNN, but well marks out features of a signal. CS eliminates the "symmetrical" noise and compresses the data well, but requires additional control of saving information.

At reconstruction of a signal, the error of recovery is added, and the computational cost increases. It is better to use reconstruction only for visualization and control, and to work only with the transformed data. Only the noise components of the recordings are subject to removal: instrumental and technogenic noise.

The seismic data pre-processing is necessary not only for the analysis by the DNN, but also for other methods of analysis.

It seems to us that with the present level of development of the DNN, the preferred method of seismic recordings pre-processing is the wavelet transformation. In addition to the excellent ability of this method to filter noises of various nature, it allows to clearly localize the signal changes over time, and it very visual method.

The purpose of this work was comparison of various methods for seismic data pre-processing and providing reference materials for future research.

## REFERENCES

1. *Altunkaynak A.* Predicting Water Level Fluctuations in Lake Michigan-Huron Using Wavelet-Expert System Methods // Water Resources Management. 2014. 28(8). P. 2293–2314. https://doi.org/10.1007/s11269-014-0616-0
2. *Ben-Haim Z., Eldar Y.C., Elad M.* Coherence-based performance guarantees for estimating a sparse vector under random noise // IEEE Trans. Signal Process. 2010. 58(10). P. 5030–5043. DOI: 10.1109/TSP.2010.2052460
3. *Blumensath T., Davies M.* Iterative hard thresholding for compressive sensing // Appl. Comput. Harmon. Anal. 2009. 27(3). P. 265–274. https://arxiv.org/pdf/0805.0510.pdf
4. *Bourgain J., Dilworth S., Ford K., Konyagin S., Kutzarova D.* Explicit constructions of rip matrices and related problems // Duke Math. Journal. 2011. 159. P. 145-185. arXiv:1008.4535v3
5. *Candès E.J., Wakin M.B.* An Introduction to Compressive Sampling [A sensing/sampling paradigm that goes against the common knowledge in data acquisition] // IEEE Signal Processing Magazine. 2008. 25(2). P. 21 – 30. DOI: 10.1109/MSP.2007.914731
6. *Creswell A., Bharath A.A.* Denoising Adversarial Autoencoders // IEEE Trans. Neural Netw. Learn. Syst. 2019. 30(4). P. 968 – 984. DOI: 10.1109/TNNLS.2018.2852738
7. *Davenport M.A., Boufounos P.T., Wakin M.B., Baraniuk R.G.* Signal Processing With Compressive Measurements // IEEE Journal of Selected Topics in Signal Processing. 2010. 4(2). P. 445 – 460. DOI: 10.1109/JSTSP.2009.2039178
8. *Davenport M.A., Laska J.N., Treichler J.R., Baraniuk R.G.* The pros and cons of compressive sensing for wideband signal acquisition: noise folding versus dynamic range // IEEE Trans. Signal Process. 2012. 60(9). P. 4628–4642. https://doi.org/10.1109/TSP.2012.2201149
9. *Erhan D., Bengio Y., Courville A., Manzagol P.-A., Vincent P., Bengio S.* Why Does Unsupervised Pre-training Help Deep Learning? // Journal of Machine Learning Research. 2010. 11. P. 625-660.
10. *Foucart S., Rauhut H.* A mathematical introduction to compressive sensing / New York. Springer. 2013. xviii+625 p. ISBN 978-0-8176-4948-7
11. *Gan S., Wang S., Chen Y., Chen X., Huang W., Chen H.* Compressive sensing for seismic data reconstruction via fast projection onto convex sets based on seislet transform // Journal of Applied Geophysics. 2016. 130. P. 194-208. https://doi.org/10.1016/j.jappgeo.2016.03.033
12. *Gentili S., Michelini A.* Automatic picking of P and S phases using a neural tree // J Seismol. 2006. 10(1). P.39-63. https://doi.org/10.1007/s10950-006-2296-6
13. *Granichin O.N., Pavlenko D.V.* Randomization of data acquisition and ℓ1ℓ1-optimization (recognition with compression) // Autom. Remote Control. 2010. 71(11) . P. 2259–2282. https://doi.org/10.1134/S0005117910110019
14. *Gravirov V.V., Kislov K.V., Gravirova L., Vinberg F.* The Use of Wavelet Transformation Techniques in Structure of an Artificial Neural Network for Recognition of Early Arrival of Earthquakes on Strongly Noisy Seismic Records // Book of Abstracts CTBT: Science and Technology 2013, Vienna, Austria, T3-P132, 2013. P.139. http://www.ctbto.org/fileadmin/user_upload/SnT2013/bookofabstracts.pdf
15. *Gravirov, V.V., Kislov, K.V., Vinberg, F.E.* Wavelet transform as a tool for processing and analysis of seismograms / Proceedings of the 10th International Conference Problems of Geocosmos - 2014, St. Petersburg State University, St. Petersburg, Petrodvorets, 2014, P. 168-172. http://geo.phys.spbu.ru/materials_of_a_conference_2014/S2014/30_Gravirov<br><br>.pdf
16. *Gurbuz A.C., McClellan J.H., Scott W.R.* A compressive sensing data acquisition and imaging method for stepped frequency GPRs // IEEE Trans. Signal Process. 2009. 57(7). P. 2640-2650. DOI: 10.1109/TSP.2009.2016270
17. *Herrmann F.J.* Compressive Sensing / Seismic Laboratory for Imaging and Modeling. https://www.slim.eos.ubc.ca/research/compressive-sensing. Accessed 2019/04/04.
18. *Kislov K.V., Gravirov V.V.* Deep Artificial Neural Networks as a Tool for the Analysis of Seismic Data // Seismic Instruments. 2018. 54(1), P. 8–16. DOI: 10.3103/S0747923918010073
19. *Kislov K.V., Gravirov V.V.* Earthquake arrival identification in a record with technogenic noise // Seismic Instruments. 2011. 47(1). — P. 66–79. https://doi.org/10.3103/S0747923911010129
20. *Kislov K.V., Gravirov V.V.* The use of artificial neural networks for classification of noisy seismic signal // Seismic Instruments. 2017. 53(1). — P. 87–101. DOI: 10.3103/S0747923917010054
21. *Kislov K.V., Gravirov V.V.* Neural network techniques for earthquake detection in high noise // Исследовано в России. 2010. № 070е. С. 837–848.
22. *Kong Q., Trugman D.T., Ross Z.E., Bianco M.J., Meade B.J., Gerstoft P.* Machine Learning in Seismology: Turning Data into Insights // Seismological Research Letters. 2018. 90(1). P. 3-14. https://doi.org/10.1785/0220180259
23. *Kotsiantis S.B., Kanellopoulos D. Pintelas P.E.* Data Preprocessing for Supervised Learning // International Journal of Computer Science. 2006. 1(1) ISSN 1306-4428. P. 111-117. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.104.8413&rep=rep1&type=pdf
24. *Li Y., Song B., Kang X., Du X., Guizani M.* Vehicle-Type Detection Based on Compressed Sensing and Deep Learning in Vehicular Networks // Sensors. 2018. 18(12). pii: E4500. DOI: 10.3390/s18124500
25. *Majumdar A., Tripathi A.* Asymmetric stacked autoencoder // 2017 International Joint Conference on Neural Networks (IJCNN). 2017. DOI: 10.1109/IJCNN.2017.7965949
26. *Meyer M., Weber S., Beutel J., Thiele L.* Systematic identification of external influences in multi-year microseismic recordings using convolutional neural networks // Earth Surf. Dynam. 2019. 7. P. 171–190. https://doi.org/10.5194/esurf-7-171-2019
27. *Oldenburg D.W., Scheuer T., Levy S.* Recovery of the acoustic impedance from reflection seismograms // Geophysics. 1983. 48(10). P.1318–1337. http://dx.doi.org/10.1190/1.1441413
28. *Robucci R., Gray J.D., Chiu L.K., Romberg J., Hasler P.* Compressive sensing on a CMOS separable-transform image sensor // Proc. IEEE. 2010. 98(6). P. 1089-1101. DOI: 10.1109/JPROC.2010.2041422
29. *Ross Z. E., Meier M.-A., Hauksson E.* P Wave Arrival Picking and First-Motion Polarity Determination With Deep Learning // J. Geophys. Res.-Sol. Ea. 2018. 123. P. 5120–5129. https://doi.org/10.1029/2017JB015251
30. *Saad O.M., Shalaby A., Sayed M.S.* Automatic discrimination of earthquakes and quarry blasts using wavelet filter bank and support vector machine // J Seismol. 2019. 23(2). P. 357-371. https://doi.org/10.1007/s10950-018-9810-5
31. *Shi Y., Lei M., Ma R., Niu L.* Learning Robust Auto-Encoders With Regularizer for Linearity and Sparsity // IEEE Access. 2019. 7. P. 17195 – 17206. DOI: 10.1109/ACCESS.2019.2895884
32. *Taylor H.L., Banks S.C., McCoy J.F.* Deconvolution with the l1 norm // Geophysics. 1979. 44 (1). P. 39–52. https://doi.org/10.1190/1.1440921
33. *Valentine A.P., Trampert J.* Data space reduction, quality assessment and searching of seismograms: autoencoder networks for waveform data // Geophys. J. Int. 2012. 189(2). P. 1183–1202. https://doi.org/10.1111/j.1365-246X.2012.05429.x
34. *Yao H., Gerstoft P., Shearer P.M., Mecklenbräuker C.* Compressive sensing of the Tohoku-Oki Mw 9.0 earthquake: Frequency- dependent rupture modes // Geophys. Res. Lett. 2011. 38, L20310. DOI:10.1029/2011GL049223
35. *Yuan S, Liu J, Wang S, Wang T., Shi P.* Seismic waveform classification and first-break picking using convolution neural networks // IEEE Geosci. Remote Sens. Let. 2018. 15(2). P. 272-276. ISSN 1545-598X. DOI: 10.1109/LGRS.2017.2785834