Feasibility of Moment Tensor Inversion From a Single Borehole Data Using Artificial Neural Networks*

Oleg Ovcharenko[1], Jubran Akram[1], and Daniel Peter[1]

Search and Discovery Article #42212 (2018)**
Posted May 21, 2018

*Adapted from extended abstract based on oral presentation given at GEO 2018 13th Middle East Geosciences Conference and Exhibition March 5-8, 2018, Manama, Bahrain

[1]King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia (oleg.ovcharenko@kaust.edu.sa)

Abstract

We discuss the feasibility of using artificial neural networks for moment tensor inversion of three-component microseismic data from a single vertical well. For this purpose, we solve a nonlinear regression problem using a feed-forward artificial neural network (ANN). We use a synthetic dataset from a homogeneous velocity model to train the network. The trained network is applied on another set of synthetic data for performance evaluation. We also investigate the effect of number of hidden layers and choice of optimization algorithm on the accuracy of inverted results. Our results indicate that for this simple case, neural network is able to invert reasonably for the full moment tensor without explicitly specifying microseismic event locations and velocity model.

Seismic source mechanisms given as moment tensors provide useful information to understand the geometry and nature of buried faults. These earthquake mechanisms can also be used to reconstruct the stress state in a fracture zone (Rebetsky et al., 2012). Although moment tensor inversion is done on a routine basis in global seismology (Tsuruoka et al., 2009), this is not as common in most microseismic applications, e.g., hydraulic fracturing. Also, this is more challenging because microseismic data often have low S/N waveforms and are acquired using limited aperture geometries (Eaton and Forouhideh, 2011). According to Shearer (2009), a seismic source inversion for all six moment tensors is a well-posed problem for high-quality surface acquisitions. However, in borehole acquisitions the inversion becomes ill-posed due to narrow-angle and irregular directional apertures (Grechka et al., 2016). This leads to large condition numbers of the inversion matrices and high sensitivity to noise in the data. Extraction of microseismic source parameters normally requires multi-well measurements leading to good source coverage but use of such configuration increases the cost and depends upon the availability of nearby wells. Even though such multi-well monitoring has been done in the past (Kolinsky et al., 2009), single-well microseismic acquisitions remain the most common approach to monitor hydraulic fracturing. Vavryčuk (2007) indicates the number of independent moment tensor parameters that can be directly inverted depends on acquisition design. For single-well measurements, only up to 5 of all 6 moment tensor components could be resolved as shown in Table 1.

In this article, we investigate the feasibility of single-well moment tensor inversion using artificial neural networks (ANNs), which can provide effective models from data-driven learning for solving complex and ill-posed problems. In the subsequent sections, we present the theory for ANN and illustrate selection of optimal ANN architecture, in particular the number of hidden layers and choice of optimization method for error minimization. Finally, we show the inversion results from a synthetic dataset that is generated using a homogeneous velocity model.

Artificial Neural Networks in Practice

Artificial neural networks recently have become tremendously popular, although they were investigated several decades ago in different fields (Röth and Tarantola, 1994; Van der Baan and Jutten, 2000). The increased interest has mainly been sparked by industrial applications, which benefit from significant improvements in several areas: (1) hardware advances that allow to perform large-scale training stages much faster; (2) software improvements with algorithms that foster new application areas for neural networks; and (3) society that has become amenable towards the use of artificial neural networks in daily life.

In practice, we implement a Multilayer Perceptron (Minsky and Papert, 1969) architecture of the ANN which could be considered as a series of successive matrix-vector multiplications, where each vector a_n contains output values of neurons from the previous layer and matrix W_n that contains weights of links between previous and next layers. In matrix form, values of elements in a following layer could be expressed depending on weights and values from their previous layer

$$a_{n+1} = W_n\, a_n$$

where multiplication represents signals going from neurons in layer n to the ones in layer $n + 1$. In a real brain, these electric signals go through brain synapses which have each a characteristic own bandwidth value. In artificial networks such values are stored as the elements of weight matrices W_n.

A neural network has to be trained to recognize the underlying relationship in data. The most common way to do this is by using a back-propagation algorithm (LeCun et al., 1998). Learning starts with random values of the weights to break the symmetry of back-propagation. The goal of training is to minimize the error between true output and predicted output by the network. Given multiple samples, the ANN adjusts weights between layers in successive iterations. At the end, the trained network is just a set of matrices whose elements were tuned to be able to solve a specific problem. Prediction is performed by subsequent multiplication of the input data with these matrices, followed by substituting of the product values into neuron's activation functions. This described routine is being repeated till reaching the output layer.

Creation of a Synthetic Microseismic Source Database

First of all, in order to use supervised learning techniques, one has to create a fairly representative base of training models - in our case this is a set of peak amplitudes and polarization of P and S wave arrivals with corresponding six moment tensor components. Figure 2 shows one

training dataset sample consisting of peaks and corresponding polarities of P and S arrivals, recorded at each receiver for one specific event. Thus, every new training sample has another location, source mechanism and magnitude.

The geometry for the experiment was chosen to be a 300 m thick slice of a homogeneous, elastic medium at 1200 m depth, spanning rectangle of 1 km^2. Compressional- and shear-wave velocities were Vp = 5.0 km/s and V_s = 2.941 km/s correspondingly, density was ρ = 2200 kg/m^3. A set of 50 receivers was evenly distributed inside the observation well placed in the middle of the model (Figure 1). Training events were randomly generated inside the volume, both in terms of location, mechanism and magnitude. We created a database of 2, 000 events, each of which corresponded to a specific moment tensor. The dataset was split in proportion 75 - 15 - 15% to build training, validation and testing data sets correspondingly.

To get amplitudes and polarities of first arrivals, one has to pick them. However, arrival picking in microseismic acquisition is a difficult task and lies beyond the scope of this work. Assuming the availability of a robust picking approach (Akram et al., 2017), we use arrival times of P and S waves obtained directly from simulation for picking. An example of a seismic trace together with indicators of the maximum amplitudes and polarities of P and S arrivals are shown in Figure 2. Finally, the generated 1,000 data samples have the dimensions given in Table 2.

Ann Architecture

The neural network architecture plays a crucial role in the reconstruction of desired output parameters. The number of hidden layers, amount of neurons in each of them, weight initialization and type of activation function all have an important influence on the performance of a neural network. An optimal layout of hidden layers is usually found by trial-and-error techniques (Baum, 1989). So far, there is no unique recipe of how to design them. The problem of ambiguity in the net- work configuration has already been discussed in impressive amount of papers and books, as an example one could refer to the book by Demuth et al. (2014). In general, one has to solve an optimization problem in a multidimensional space to find the optimal configuration of the network that minimizes a certain loss function. It has been proven in Ito (1992) that a non-linear perceptron with one hidden layer is able to approximate any continuous function with a given precision if the number of hidden neurons tends to infinity. In the scope of this work, a one-layer configuration did not perform sufficiently well so that we had to increase the number of hidden layers. This might indicate a violation of the smoothness assumption as well as some data redundancy, which needed to be accommodated by the network.

Normalization of input and output data dramatically affects quality of reconstruction and duration of training. There is a variety of data normalization approaches, in this work we use normalization which maps input and output data onto the range [−1,1]. Due to this normalization, we have chosen the hyperbolic tangent function as activation function in neurons of all hidden layers,

$$Tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

where the output range is the same as the given input range. Mathematically, the neural network then maps a set of input data onto a set of corresponding outputs:

$$L\left(\begin{bmatrix} P_x^{11} & P_y^{11} & P_z^{11} & \dots & P_z^{1N} \\ \vdots & & & & \\ P_x^{K1} & P_y^{K1} & P_z^{K1} & \dots & P_z^{KN} \end{bmatrix}\right) = \begin{bmatrix} M_1^1 & M_2^1 & \dots & M_6^1 \\ \vdots & & & \\ M_1^K & M_2^K & \dots & M_6^K \end{bmatrix}$$

where P_i^{KN} refers to a peak value of i^{th} waveform component recorded at N^{th} receiver during generation of K^{th} training sample, and M_i^K is the i^{th} independent moment tensor component of source mechanism used to generate K^{th} dataset.

Throughout this work, all the ANNs used are feed-forward, fully-connected networks with hyperbolic tangent activation functions in neurons of hidden layers and linear functions in the output ones. An example of a trial ANN configuration is shown in Figure 3.

Number of Hidden Layers

In this section, we show how the number of hidden layers affects the accuracy of moment tensor inversions. The accuracy here is defined as the fraction of correctly predicted moment tensors among all used testing examples. A correct prediction is achieved if all six components do not differ by more than 5% of the maximum value from the reference set.

As mentioned above, the use of networks with more than 2 or 3 hidden layers is uncommon. Normally, networks of that size are able to approximate any complex function. In our study the networks with more than three-layers were able to achieve the wanted target of at least 80% correct predictions on synthetic data.

The best in terms of accuracy/costs ration multi-level architecture in our tests included 3 hidden layers, decreasing gradually in terms of size. This configuration was found by a mix of manual and automated approaches and appeared to be [158,79,52]. We compute initial estimate of number of neurons in the first hidden layer following (Huang, 2003) despite the authors investigate a two hidden-layer network architecture. Figure 4 provides an idea of how the prediction accuracy changes depending on the number of used hidden layers.

Visual comparison of resulting source mechanisms corresponding to a batch of outputs for a 3 hidden layers network is given in Figure 5. Although the corresponding “beach balls” in Figure 5 look very similar between true and predicted results, we notice that for fewer layers involved, the less accurate the predictions become in total according to Figure 4. That is also due to the fact that these beach ball visualizations do not express absolute values of moment tensor components, but are diagrams of the characteristic focal mechanisms. Absolute values could differ significantly, but would not become visible. Here we are interested in correct absolute values, so these visual representations are only complementary indicators rather that principal ones.

Optimization Techniques

Optimization is the process of finding the set of parameters that minimize a certain loss function. The choice of optimization technique used in a neural network training process may lead to results that differ significantly. In application to the neural network designed in this work, we use Stochastic Gradient Descent (SGD) and provide performance comparisons for several, most popular gradient descent optimization techniques. The following algorithms are compared: Adaptive Moment Estimation – *Adam* (Kingma and Ba, 2014), Nesterov accelerated gradient (Nesterov, 1983), Momentum (Qian, 1999), RMSProp (Hinton et al., 2012) and Adagrad (Duchi et al., 2011).

Figure 6 shows a comparison of the listed techniques in terms of prediction accuracy. RMSProp, that is an adaptive learning rate method, showed up to 90% correct predictions on a synthetic dataset, even with a 2% inaccuracy assumption. This result is not unusual for such a simple model geometry. All other optimization techniques perform relatively well, but they get stuck earlier on in local minima during the iterative optimization process. Note that RMSProp performs well with our specific dataset, however each of the methods listed above has its advantages when being applied to other problems, where the size of training dataset, size of input and output data may differ.

Results and Discussion

We tested various network configurations with one and many hidden layers. Our results suggest that it is difficult to find an optimal neural network architecture that would always provide the best performance. In our testing framework, the ANN configuration with three hidden layers and [158, 79, 52] neurons in each of them showed the best accuracy, reaching about 90% correct results with a 2% error assumption. A random batch of focal mechanism solutions from the testing set is shown in Figure 7.

The focal mechanism that represents isotropic expansion (full blue beach ball in Figure 7) seems to be poorly reconstructed, but this is due to the visualization issue. These source types hold ones on moment tensor's diagonals, but as soon as even very small values appear at off-diagonal spots the beach ball ceases to be "pure". Direct value by value comparisons of the moment tensor components show that in most cases the value alternation are within the margin of error. Robust method for determining an optimal feed-forward ANN architecture is still missing, so our search approach is mostly based on a trial-and-error technique as suggested by Baum (1989).

As optimization technique we used RMSProp with learning rate, batch size and $L2$ weight regularization parameter of $1E-3$, 100 and $\alpha = 5E-4$ correspondingly. Due to relatively small size of input data training the network took just a few minutes on a laptop with 4-core CPU. Practical implementation was using the Tensorflow library.

Conclusions

Machine learning is getting increasingly popular in geophysics due to the ability to solve complex nonlinear problems without specifying physical laws explicitly. In this work, we explored the applicability of artificial neural networks to be used for full moment tensor inversion in case of single-well synthetic three-component wavefield recordings. Input for the feed-forward ANN used in this work consisted of amplitudes and polarities of P and S waves arrivals, whereas output was given by six moment tensor components. We showed for simple case that the

neural network is able to satisfactory invert for the full moment tensor without explicitly specifying microseismic event locations and velocity model. Overall maximum prediction accuracy reached with a three-layer ANN was about 90%, assuming moment tensor value alterations of 2%. Although, the prediction capability of such a neural network is high, the network architecture itself is mainly found by trial-and-error. Future research is needed in guiding an optimal network search and to explore the applicability of this approach to more complex cases.

Acknowledgements

The research reported in this publication was supported by funding from King Abdullah University of Science and Technology (KAUST). For computer time, this research used the resources of the Information Technology Division and Extreme Computing Research Center (ECRC) at KAUST.

References Cited

Akram, J., O. Ovcharenko, and D. Peter, 2017, A robust neural network-based approach for microseismic event detection: SEG Technical Program Expanded Abstracts 2017, p. 2929-2933.

Baum, E.B., 1989, A proposal for more powerful learning algorithms: Neural Computation, v. 1, p. 201-207.

Demuth, H.B., M.H. Beale, O. De Jess, and M.T. Hagan, 2014, Neural network design: Martin Hagan.

Duchi, J., E. Hazan, and Y. Singer, 2011, Adaptive subgradient methods for online learning and stochastic optimization: Journal of Machine Learning Research, v. 12, p. 2121-2159.

Eaton, D.W., and F. Forouhideh, 2011, Solid angles and the impact of receiver-array geometry on microseismic moment-tensor inversion: Geophysics, v. 76, WC77-WC85.

Grechka, V., Z. Li, B. Howell, and V. Vavrycuk, 2016, Single-well moment tensor inversion of tensile microseismic events: Geophysics, v. 81, KS219-KS229.

Hinton, G., N. Srivastava, and K. Swersky, 2012, Lecture 6a overview of mini-batch gradient descent.

Huang, G.-B., 2003, Learning capability and storage capacity of two-hidden-layer feedforward networks: IEEE Transactions on Neural Networks, v. 14, p. 274-281.

Ito, Y., 1992, Approximation of continuous functions on r d by linear combinations of shifted rotations of a sigmoid function with and without scaling: Neural Networks, v. 5, p. 105-115.

Kingma, D., and J. Ba, 2014, Adam: A method for stochastic optimization: arXiv preprint, 1412.6980.

Kolinsky, P., L. Eisner, V. Grechka, D. Jurick, and P. Duncan, 2009, Observation of shear-wave splitting from micro-seismicity induced by hydraulic fracturing - a non-vti story: 71st EAGE Conference and Exhibition incorporating SPE EUROPEC 2009.

LeCun, Y., L. Bottou, G.B. Orr, and K.-R. Muller, 1998, Efficient backprop: Neural networks: Tricks of the trade, Springer, p. 9-50.

Minsky, M., and S. Papert, 1969, Perceptrons.

Nesterov, Y., 1983, A method of solving a convex programming problem with convergence rate o (1/k2): Soviet Mathematics Doklady, p. 372-376.

Qian, N., 1999, On the momentum term in gradient descent learning algorithms: Neural networks, v. 12, p.v145-151.

Rebetsky, Y.L., O. Kuchai, N. Sycheva, and R. Tatevossian, 2012, Development of inversion methods on fault slip data: Stress state in orogenes of the central asia: Tectonophysics, v. 581, p. 114-131.

Röth, G., and A. Tarantola, 1994, Neural networks and inversion of seismic data: Journal of Geophysical Research: Solid Earth, v. 99, p. 6753-6768.

Shearer, P. M., 2009, Introduction to seismology: Cambridge University Press.

Tsuruoka, H., H. Kawakatsu, and T. Urabe, 2009, Grid mt (grid-based real-time determination of moment tensors) monitoring the long-period seismic wavefield: Physics of the Earth and Planetary Interiors, v. 175, p. 8-16.

Van der Baan, M., and C. Jutten, 2000, Neural networks in geophysical applications: Geophysics, v. 65, p. 1032-1047.

Vavryčuk, V., 2007, On the retrieval of moment tensors from borehole data: Geophysical Prospecting, v. 55, p. 381-391.

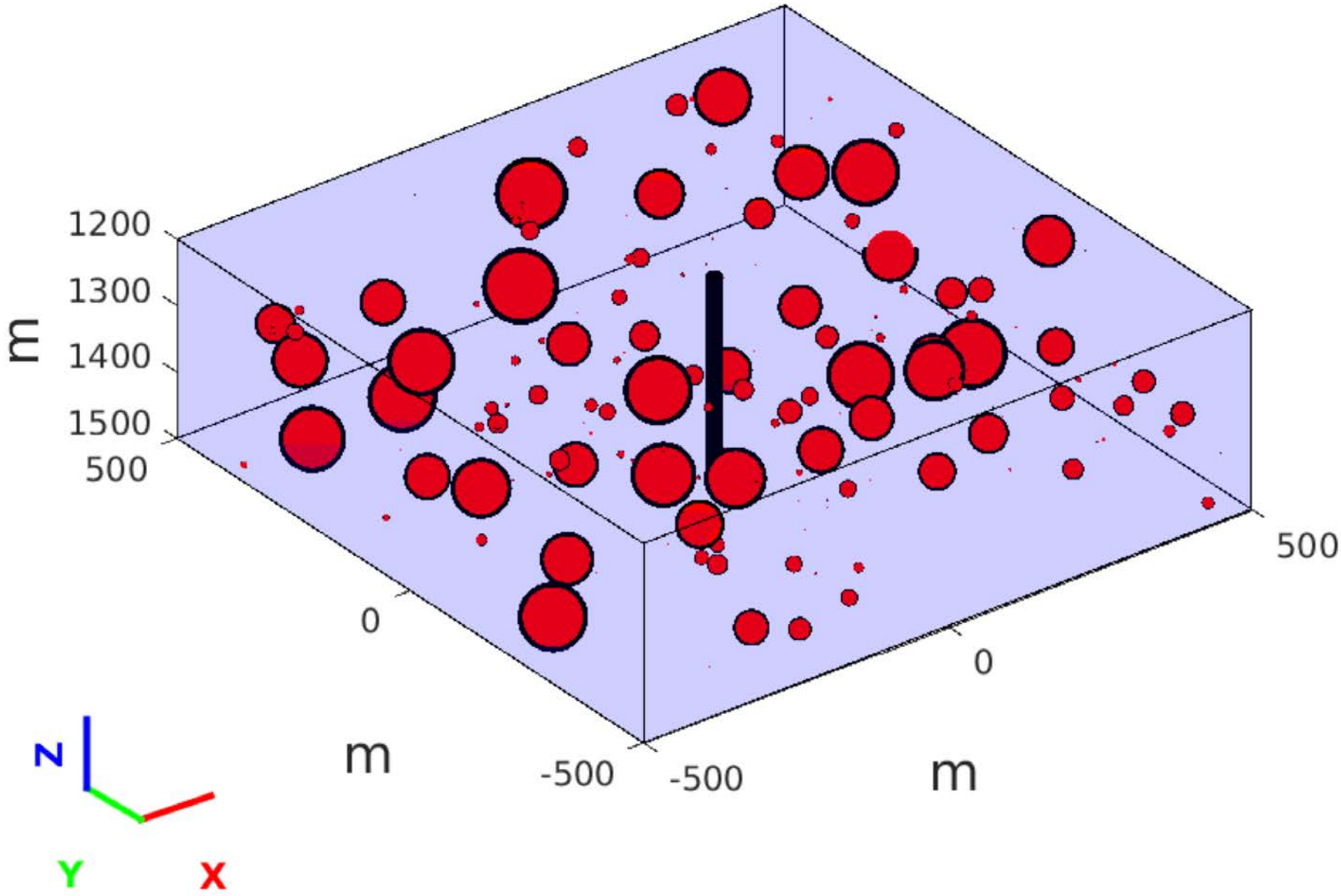

Figure 1. Homogeneous isotropic velocity model for data generation. Black line in the middle indicates acquisition well with 50 evenly distributed receivers. Red dots indicate location of microseismic events and their size is relative to magnitude.

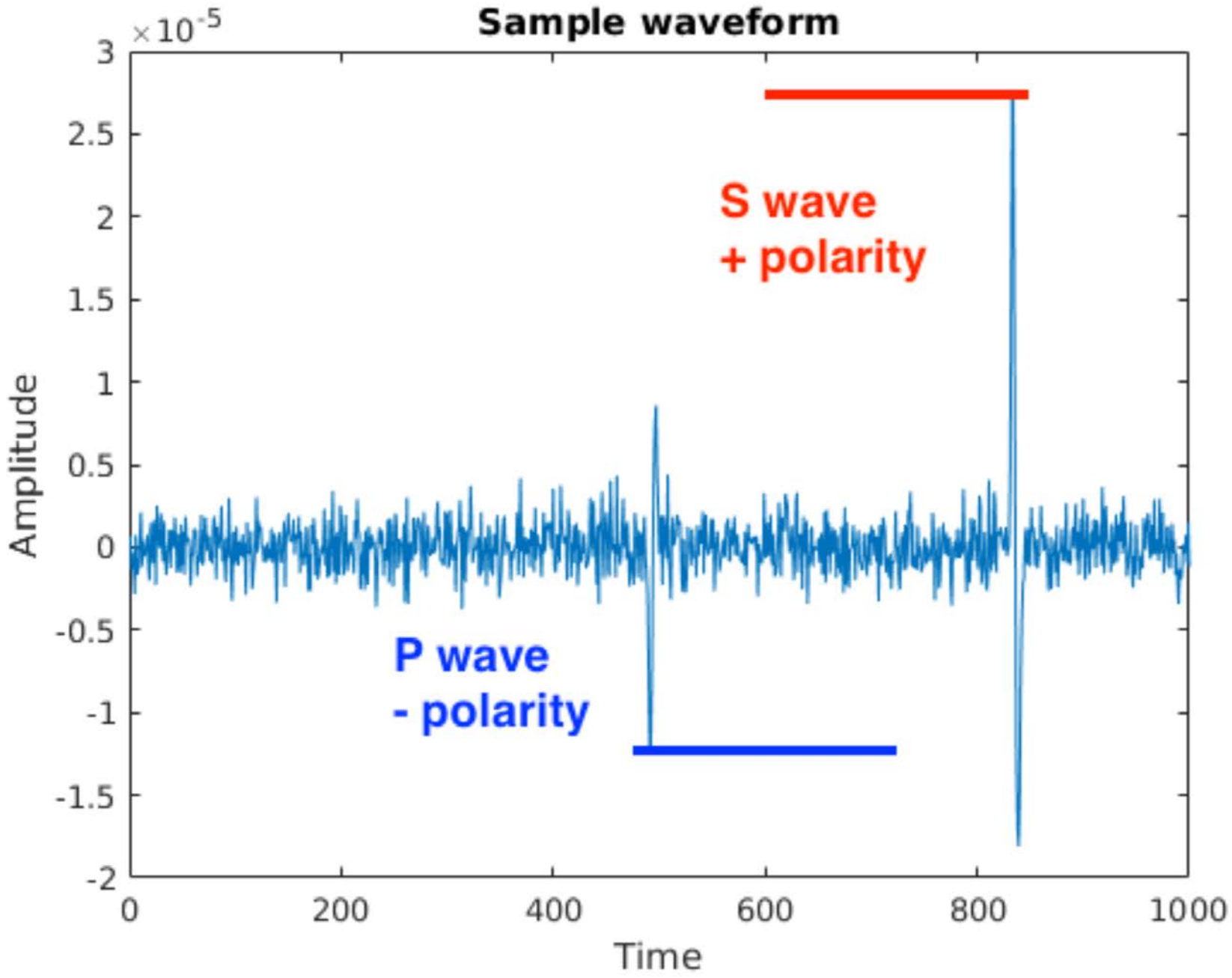

Figure 2. Sample waveform recorded at an in-well receiver.

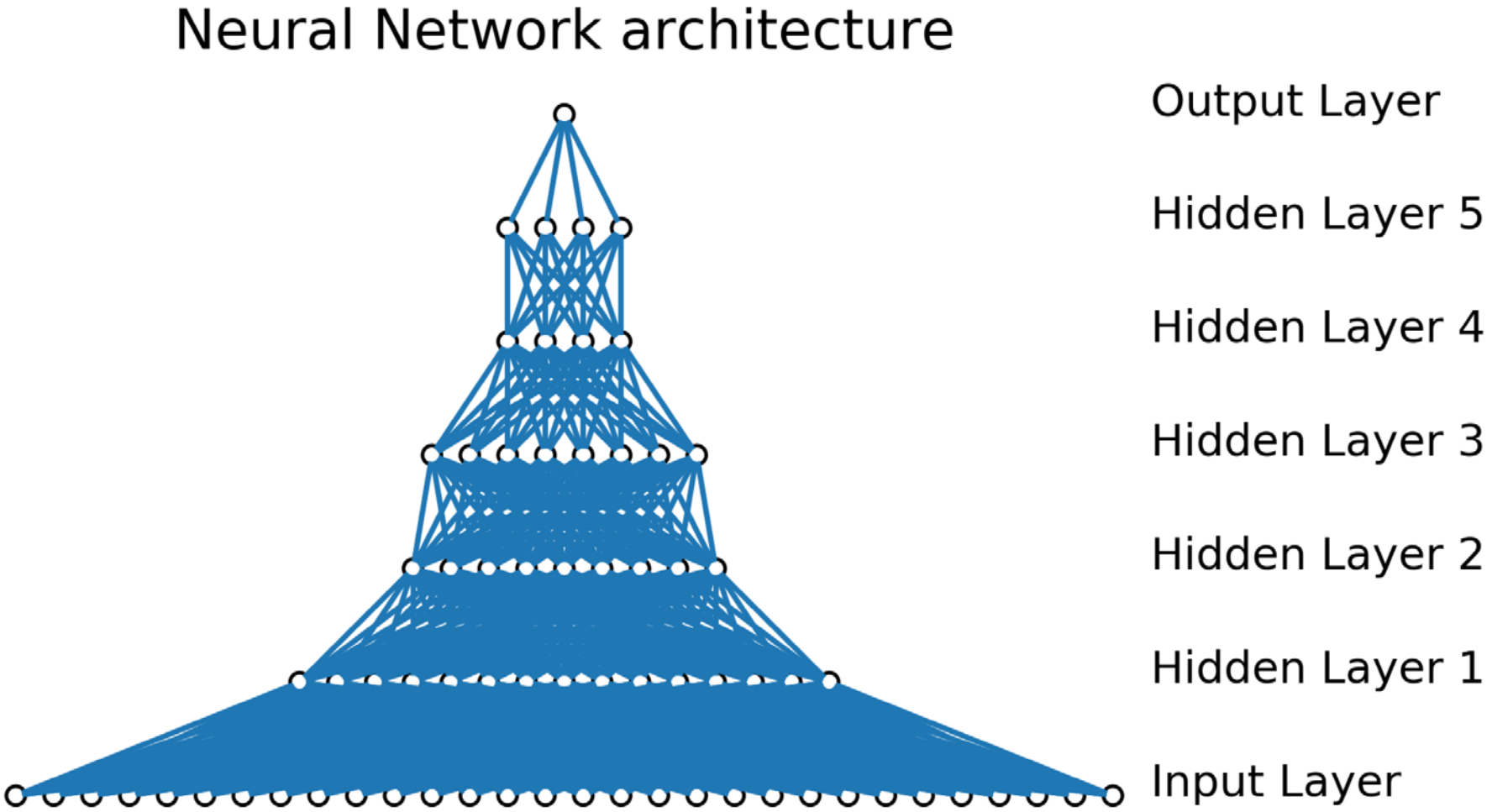

Figure 3. Scaled representation of one of the trial networks tested in this work.

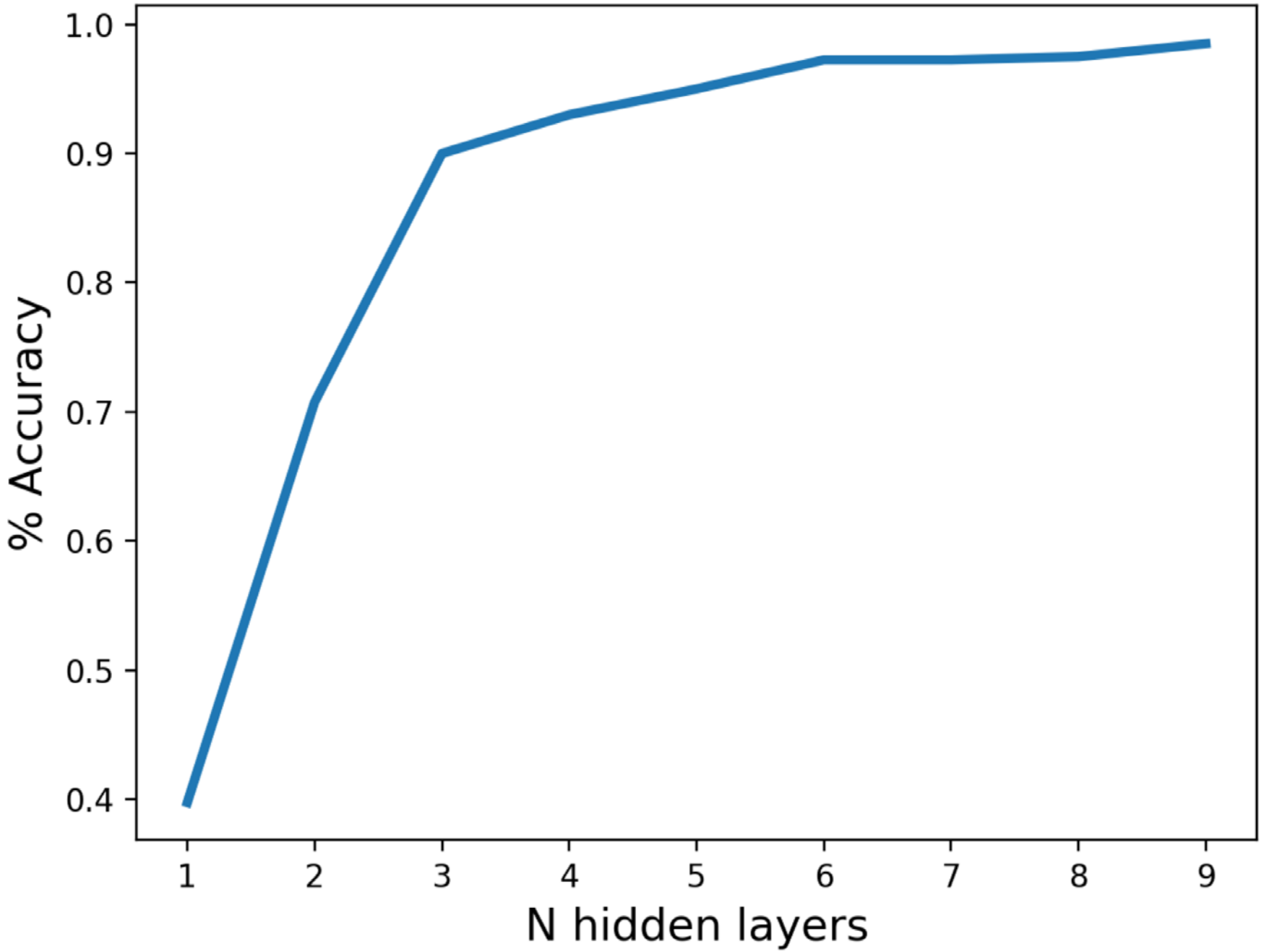

Figure 4. Prediction accuracy of the network depending on the number of cascaded hidden layers, from 1 to 9.

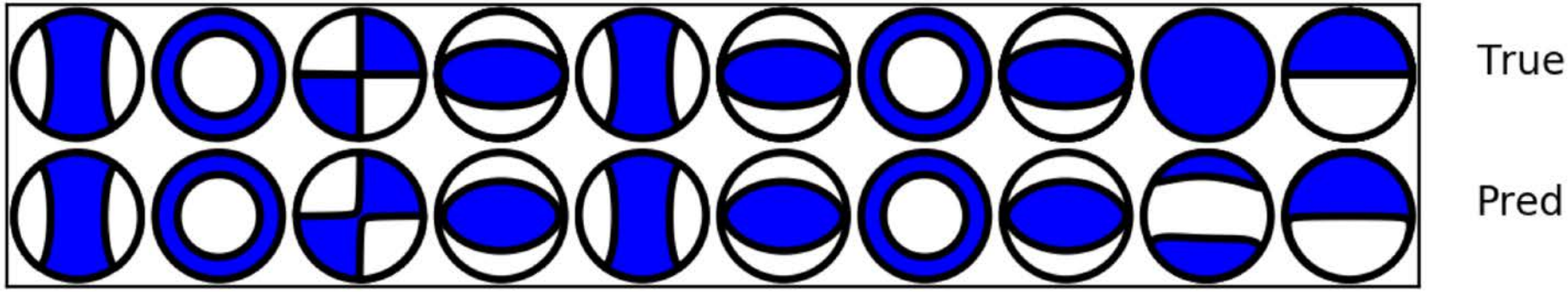

Figure 5. Visual representation of true and predicted focal mechanisms for 3 hidden-layer ANN.

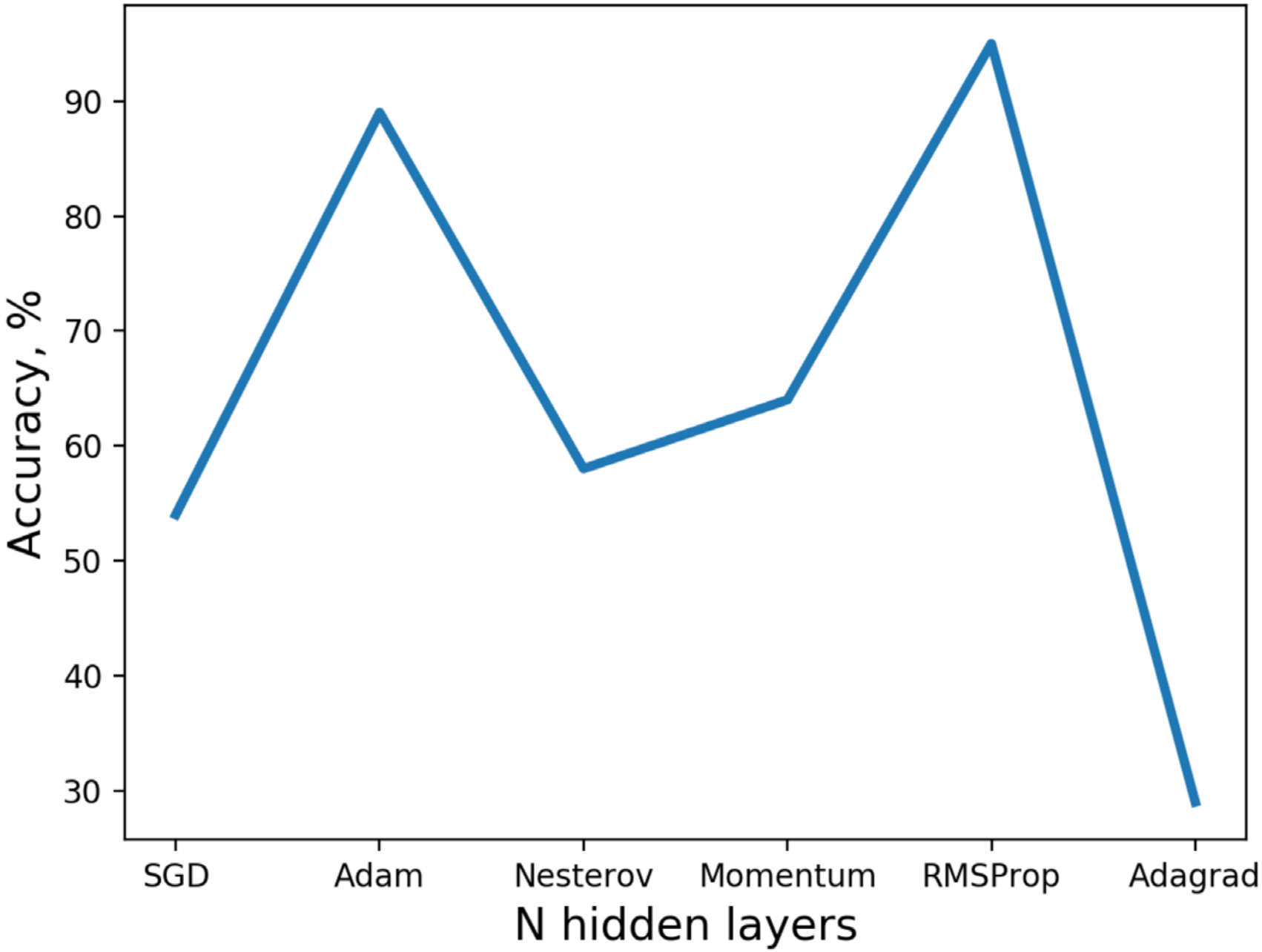

Figure 6. Comparison of prediction accuracy of the network depending on the optimization algorithm used. Assumption is that moment tensor components could differ by 5%.

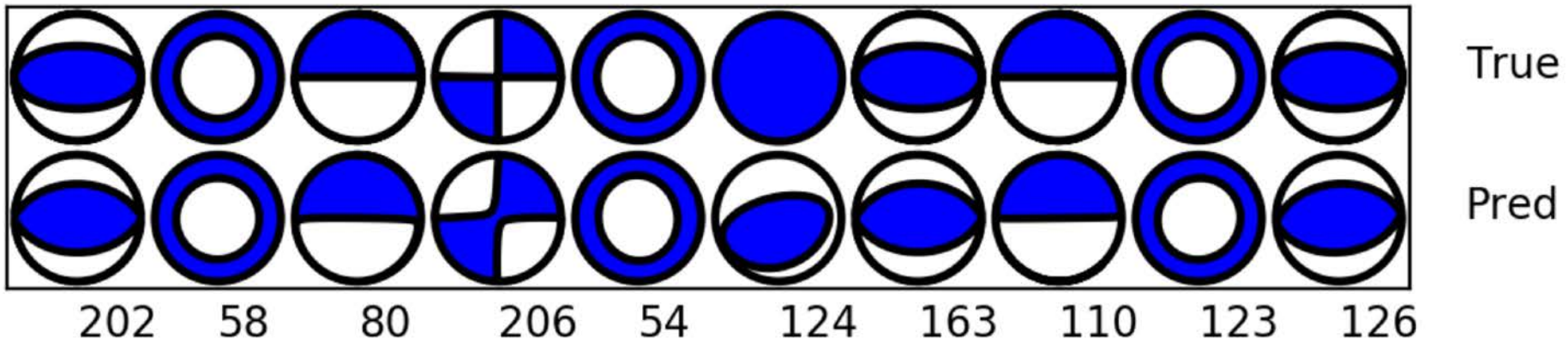

Figure 7. Visual representation of the true focal mechanisms and the ones predicted by the ANN.

	P	S_V	S_H	S_V+S_H	P+S_V+S_H
N	3	3	2	5	5

Table 1. Number of moment tensor components (N) resolved by single-well measurements *(P, S_V , S_H* signals)

	Size	Description
Input:	300	[X, Y, Z] × 50 receivers × 2 wave types
Output:	6	Independent moment tensor components

Table 2. Data used for input and output of the neural network for moment tensor inversions.