Inversion for shear-tensile focal mechanisms using an unsupervised physics-guided neural network

Hongliang Zhang, Kristopher A. Innanen, and David W. Eaton

ABSTRACT

We present a novel physics-guided neural network to estimate shear-tensile focal mechanisms for microearthquakes using displacement amplitudes of direct P-wave. Compared with conventional data-driven fully connected (FC) neural networks, our physics-guided neural network is implemented in an unsupervised fashion and avoids the use of training data, which may be incomplete or unavailable. We incorporate three FC layers and a scaling & shifting layer to estimate shear-tensile focal mechanisms for multiple events. Then, a forward modeling layer is added, which generates synthetic amplitude data based on the source mechanisms emerging from the previous layer. The neural network weights are iteratively updated to minimize the mean squared error between observed and modeled normalized P-wave amplitudes. We apply this machine-learning approach to a set of 530 induced events recorded during hydraulic fracture simulation of Duvernay Shale west of Fox Creek, Alberta, yielding results that are consistent with previously reported source mechanisms for the same dataset. A distinct cluster characterized by more complex mechanisms exhibits relatively large Kagan angles ($10°$ – $25°$) compared with the previously reported best double-couple solutions, mainly due to model simplification of the shear-tensile focal mechanism. Uncertainty tests demonstrate the robustness of the inversion results and high tolerance of our neural network to errors in event locations, velocity model and P-wave amplitudes. Compared with a single-event grid-search algorithm to estimate shear-tensile focal mechanisms, the proposed neural network approach exhibits significantly higher computational efficiency.

INTRODUCTION

Moment tensors are widely used to describe earthquake source mechanisms. They not only characterize the shear slip of the fault through the double-couple (DC) component, but also can reveal more complex faulting process (e.g., shear slip on the non-planar fault, tensile faulting, co-slipping of multiple shear events) via their non-DC components (Julian et al., 1998). The range of inversion approaches for moment tensors can be classified into three major categories: amplitude-based approaches (Ebel and Bonjer, 1990; Rögnvaldsson and Slunga, 1993; Šílený et al., 2009; Pesicek et al., 2016), amplitude ratio-based approaches (Kisslinger, 1980; Hardebeck and Shearer, 2003; Jechumtálová and Šílený, 2005) and waveform-based methods (Spikin, 1986; Zhu and Helmberger, 1996; Dreger et al., 2000; Zhang et al., 2016). Waveform-based algorithms are often used for the source-mechanism inversion of earthquakes that are recorded by regional and global-scale networks. Low-magnitude earthquakes or microearthquakes (e.g., associated with microseismic events or induced seismicity), recorded by local arrays, are sometimes challenging to accurately invert for moment-tensor parameters through waveform fitting, due to typically low signal-to-noise ratios (Eaton and Forouhideh, 2011; Li et al., 2011). Instead, amplitude- or amplitude-ratio based inversion approaches have typically been applied for these events (e.g., Staněk et al., 2014; Shang and Tkalčić, 2020).

In addition to the theory-based inversion approaches mentioned above, recent studies have also documented applications of machine-learning algorithms to source-mechanism inversion. For example, Käufl et al. (2015) developed a pattern-recognition-based probabilistic method to estimate earthquake source parameters using near-field and regional seismic waveforms, in which a neural network was calibrated using a synthetic dataset. In addition, Ovcharenko et al. (2018) successfully applied a FC neural network with five hidden layers to moment-tensor inversion of microseismic events recorded by a single well. Like most supervised learning techniques, these machine-learning approaches are data-driven and require labeled data to train the neural network weights, after which the trained networks are then applied to the validation/test dataset. In cases of limited or unavailable training data, these machine-learning algorithms may fail to yield reliable source-mechanism solutions. To overcome this issue, a number of physics-guided neural networks, in which theoretical, or model-based, constraints on the relationship between geophysical data and subsurface unknowns are incorporated into neural networks, have been formulated (e.g., Biswas et al., 2019; Downton et al., 2019; Ren et al., 2020; Sun et al., 2020; Zhang et al., 2020. To date, theory-guided neural network design has not been applied to the problem of source-mechanism inversion.

Instead of implementing the inversion for full moment tensors that describe seismic sources, here we use a physically based shear-tensile model (Vavryčuk, 2001) to directly characterize the slip along the fault with an off-plane component. Compared with the moment tensor, the use of shear-tensile model can yield more robust result for low-magnitude events due to less model parameters and can ease the requirement for azimuthal coverage of stations (Šílenỳ et al., 2014; Šílenỳ, 2018; Petružálek et al., 2018). The shear-tensile mechanism has four independent parameters, strike, dip, rake and slope. The slope represents the angle between the slip direction and the fault plane and has a value ranging between -90° and 90°. A value of 0° for slope denotes the pure shear source, and values of -90° and 90° represent tensile closure and tensile opening, respectively (Vavryčuk, 2001). We choose to use the shear-tensile model for two main reasons: (1) notable and spurious non-double-couple (non-DC) components may be caused within the moment tensors of microseismic events or low-magnitude earthquakes due to the low quality of input data (Ou, 2008; Godano et al., 2011; Zhang et al., 2016), and constraining the source mechanism for these events to a simpler shear-tensile model has been proven convenient and effective (e.g., Vavryčuk, 2011; Stierle et al., 2014; (2) compared with the six independent elements of the full moment tensor, all the four parameters of the shear-tensile model have definite and relatively small value ranges, which is thus straightforward for the neural network to scale values into these ranges.

In this study, we present and validate a novel physics-guided neural network for simultaneous estimation of shear-tensile focal mechanisms for a set of microearthquakes recorded on a dense network of stations. Our proposed neural network design combines FC layers and seismic source theory to simultaneously estimate shear-tensile focal mechanisms for multiple microearthquakes. In contrast to previous data-driven machine learning techniques for source-mechanism determination, here training data is no longer required in our neural network since a forward modeling layer is incorporated into the FC neural network to guide the learning process. In this paper, we first introduce the architecture and implementation details of the proposed physics-guided neural network. Then, we apply it to the

source-mechanism inversion of 530 events induced during a four-well hydraulic-fracturing completion program in Duvernay Shale. To test the robustness of the source-mechanism inversion results, we investigate the sensitivity of retrieved focal mechanisms to the velocity model perturbation, event location errors and different levels of noise. The method proposed here is tailored for dense surface/shallow-borehole monitoring arrays (e.g., Eaton et al., 2018), but is easily generalizable to other types of sensor networks. Due to the inclusion of physics-based forward modeling layer to guide the learning, labeled data is not required and the training process also inverts for multiple source mechanisms.

METHOD

Figure 1 shows the architecture of the proposed physics-guided neural network, in which the network uses three FC layers and a Scaling & Shifting layer to map the normalized P-wave displacement amplitudes (i.e., the observations) into the four parameters for the shear-tensile model. Unlike the conventional data-driven neural networks that require input-label pairs, we implement the neural network in an unsupervised way by incorporating a forwarding modeling layer, i.e., the Radiation Pattern & Normalization layer, to calculate the theoretical normalized P-wave amplitudes for any given focal mechanism. Then, a loss function is used to evaluate the data residual between observed and modeled amplitudes, and a proper optimization algorithm is used to iteratively update network weights. The followings show the implementation details of the proposed neural network.
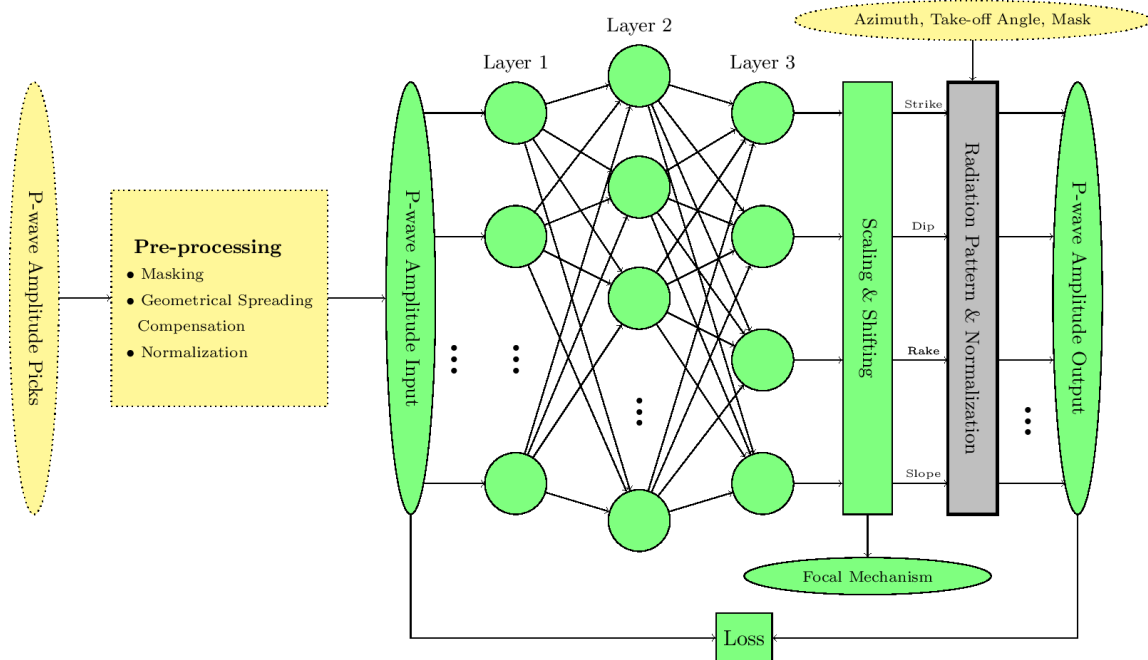

FIG. 1. Architecture of the physics-guided neural network. Inputs and outputs of the network are represented by elliptic blocks. Data preparation works prior to network training are indicated by dotted-line blocks, and the physics-based forward modeling is denoted by the bold-line block.

Data pre-processing

Prior to the network training, some pre-processing steps (the square block in Figure 1) should be applied to the picked P-wave amplitudes, including data masking, geometrical-

spreading compensation and normalization. The purpose of applying a mask to the amplitude data is to avoid effects caused by individual stations with unavailable or unreliable picks, and to ensure that all the events considered in the inversion have the same input size. The masking operation can be represented as

$$\hat{\mathbf{A}} = \mathbf{A} \odot \mathbf{M}, \tag{1}$$

where $\mathbf{A}$ and $\hat{\mathbf{A}}$ denote P-wave amplitude matrices before and after masking, respectively, and $\mathbf{M}$ is the mask. Matrices $\mathbf{A}$, $\hat{\mathbf{A}}$ and $\mathbf{M}$ have same the dimension $N_1 \times N_2$, in which N_1 and N_2 represent total numbers of events and stations, respectively. Each element in $\mathbf{M}$ has a value of either 0 or 1, representing whether the corresponding station has available (the value 1) or unavailable (the value 0) amplitude pick. The symbol represents the Hadamard product operator. The pre-processing also compensates for amplitude attenuation due to geometrical spreading in the layered isotropic medium (Ursin and Hokstad, 2003). Next, for each event, P-wave amplitudes are normalized by dividing amplitude values by the maximum amplitude (in the absolute sense) for that event. Then the pre-processed amplitudes are used as inputs for the subsequent FC layer.

Fully connected layers and Scaling & Shifting layer

There are three FC layers in our model. The numbers of neurons for first and third layers are equal to total numbers of stations (N_2) and shear-tensile-model parameters (four), respectively. For the neuron number in the second layer, a series of representative values (32, 48, 64, 80, 96, 112 and 128) within a rule-of-thumb range, $< 2N_2$ (according to Heaton, 2008), are examined. Base on the test result, we choose to use 128 neurons which yields the lowest loss value. In order to introduce nonlinearity into the network, activation functions are applied to the three FC layers. The mathematical operation at the FC layer can be represented as

$$\mathbf{y}^{(k)} = f(\mathbf{w}^{(k)}\mathbf{x}^{(k)} + \mathbf{b}^{(k)}), (k = 1, 2, 3), \tag{2}$$

where $\mathbf{x}^{(k)}$, $\mathbf{y}^{(k)}$ and $\mathbf{b}^{(k)}$ are the input, output and bias vectors, respectively, at FC layer k; $\mathbf{w}^{(k)}$ is the weight matrix, and f represents the activation function for that layer. In this study, the rectified linear unit (ReLU) is used as the activation function at both FC layers 1 and 2, which is linear for positive values and zero otherwise (i.e., $\mathbf{y}^{(k)} = \max(0, \mathbf{w}^{(k)}\mathbf{x}^{(k)} + \mathbf{b}^{(k)})$, $k = 1, 2$). To output values within the range of 0–1, a sigmoid function is applied at the third FC layer, expressed as

$$\mathbf{y}^{(3)} = \frac{1}{1 - e^{-(\mathbf{w}^{(3)}\mathbf{x}^{(3)} + \mathbf{b}^{(3)})}}. \tag{3}$$

To generate values within ranges of $0° - 360°$, $0°–90°$, $-180° - 180°$ and $-90° - 90°$ for strike, dip, rake and slope, a Scaling & Shifting layer is connected to the third FC layer, represented as

$$\mathbf{y} = \mathbf{s}_0 + \mathbf{s}_1 \odot \mathbf{x}. \tag{4}$$

In this layer, $\mathbf{s}_0 = [0°, 0°, \text{-}180°, \text{-}90°]^T$ and $\mathbf{s}_1 = [360°, 90°, 360°, 180°]^T$ are the shifting and scaling vectors, which correspond to the lower limits and distances between upper and lower limits of the four parameters of the shear-tensile model, respectively.

Forward modeling layer

In contrast to conventional data-driven machine learning approaches, we incorporate a forward modeling layer, called Radiation Pattern & Normalization, into the proposed network to generate theoretical normalized P-wave amplitudes for any given shear-tensile focal mechanism. Then, the network uses the physics implied by the radiation patterns to guide the training for model parameters, by minimizing the residual between observed and modeled data.

In homogenous media, the far-field P-wave displacement caused by a shear-tensile fault can be expressed as (Ou, 2008)

$$\mathbf{u}^P = \frac{\mu A \Delta \dot{u}(t - r/\alpha)}{4\pi \rho \alpha^3} \frac{\hat{\mathbf{r}}^T \mathbf{S} \hat{\mathbf{r}}}{r} \hat{\mathbf{r}}, \tag{5}$$

where μ, A, ρ and α represent shear modulus, rupture area, density and P-wave velocity, respectively; r is the distance between source and receiver; $\hat{\mathbf{r}}$ is a unit vector pointing from source to receiver; $\mathbf{S}$ is the source dislocation tensor; and $\Delta \dot{u}(t)$ is the time derivative of the magnitude of dislocation. The six independent elements of $\mathbf{S}$ are

$$
\begin{aligned}
S_{11} &= [2\sigma/(1 - 2\sigma) + 2\sin^2 \delta \sin^2 \phi] \sin \gamma - (\sin \delta \cos \lambda \sin 2\phi + \sin 2\delta \sin \lambda \sin^2 \phi) \cos \gamma \\
S_{22} &= [2\sigma/(1 - 2\sigma) + 2\sin^2 \delta \cos^2 \phi] \sin \gamma + (\sin \delta \cos \lambda \sin 2\phi - \sin 2\delta \sin \lambda \cos^2 \phi) \cos \gamma \\
S_{33} &= [2\sigma/(1 - 2\sigma) + 2\cos^2 \delta] \sin \gamma + \sin 2\delta \sin \lambda \cos \gamma \\
S_{12} &= -\sin^2 \delta \sin 2\phi \sin \gamma + (\sin \delta \cos \lambda \cos 2\phi + \sin 2\delta \sin \lambda \sin 2\phi/2) \cos \gamma \\
S_{13} &= \sin 2\delta \sin \phi \sin \gamma - (\cos \delta \cos \lambda \cos \phi + \cos 2\delta \sin \lambda \sin \phi) \cos \gamma \\
S_{23} &= -\sin 2\delta \cos \phi \sin \gamma - (\cos \delta \cos \lambda \sin \phi - \cos 2\delta \sin \lambda \cos \phi) \cos \gamma
\end{aligned}
\tag{6}
$$

where ϕ, δ, λ and γ are the strike, dip, rake and slope, respectively, and σ is the Poisson's ratio, which is fixed and assumed to be 0.25 in the inversion. As elaborated below, the take-off angles used in the inversion are based on ray tracing.

As mentioned previously, the observed P-wave amplitudes are scaled over the maximum amplitude. In terms of the modeling data, when the same normalization is applied, the first term in equation (5) for different stations are approximately the same, and thus can be canceled out. Considering that geometrical spreading correction for the layered medium is already applied to the observed data, the geometrical spreading term, $1/r$, in equation (5) for isotropic media can be neglected, and the normalization of the modeled data only needs to be applied to the radiation pattern, $\hat{\mathbf{r}}^T \mathbf{S} \hat{\mathbf{r}}$.

In this study, an isotropic, 1-D layered velocity model is used for ray tracing in order to estimate the take-off angles. The vector $\hat{\mathbf{r}}$ is calculated based on the azimuth measured from source to receiver and the take-off angle at the source obtained from ray tracing. To be consistent with the observed data, the same mask is also applied to the theoretical normalized P-wave amplitudes. As shown in the elliptical block above the forward modeling layer in Figure 1, these data can be prepared independently outside of the designed neural network.

Loss function and optimization algorithm

In our neural network, the mean squared error (MSE) between observed and theoretical P-wave amplitudes is adopted as the loss function and represented as

$$\text{Loss} = \frac{1}{N_2} \sum_{k=1}^{N_2} \|\mathbf{a}_k^{obs} - \mathbf{a}_k^{mod}\|_2^2, \tag{7}$$

where $\mathbf{a}_k^{obs}$ and $\mathbf{a}_k^{mod}$ represent the observed and theoretical normalized P-wave-amplitude vectors for event k, N_2 is the total number of events, and the operator $\|\|_2^2$ denotes squared L_2-norm. In the training phase, the physics-guided network utilizes the adaptive momentum (Adam) algorithm (Kingma and Ba, 2014) to update network weights and biases, and to output optimal shear-tensile focal mechanisms for a set of events. In the Adam algorithm, the learning rates is updated adaptively for each individual model parameter from estimates of first and second moments of the gradients.

APPLICATION TO A TEST DATASET

Dataset

In this section, we present an application of our proposed neural network to passive seismic data from the Tony Creek Dual Microseismic Experiment (ToC2ME) (Eaton et al., 2018; Rodríguez-Pradilla and Eaton, 2019), in which a diverse suite of sensors were utilized to monitor a four-well hydraulic-fracturing program within the Duvernay Shale formation west of Fox Creek, Alberta. The acquisition system consists of one strong-motion accelerometer, six broadband seismometers and a 68-station shallow-borehole local network. As presented by Eaton et al. (2018), over 25,000 putative events were detected with a maximum magnitude of M_W 3.2, of which 4,083 events with relatively high signal-to-noise ratio were located.

Using a least-squares inversion based on 3C amplitudes from the dense array observations of the direct P-wave, Zhang et al. (2019) obtained moment-tensor solutions for 530 high-quality events with moment magnitudes greater than 0.2. According to Zhang et al. (2019), four groups of mechanisms are evident. Events within groups one and two exhibit predominantly strike-slip mechanisms along subvertical N-S/E-W or NE-SW/NW-SE trending nodal planes; events within group three exhibit either oblique dip-slip on subvertical NW-SE trending nodal plane or shallow-dipping slip on NE-SW trending nodal plane. Group four contains events with diverse set of source mechanisms that do not fit into any

of the previous three major groups. Figure 2 shows the distributions of 530 events, best DC mechanisms for the three major groups and outliers in group four.
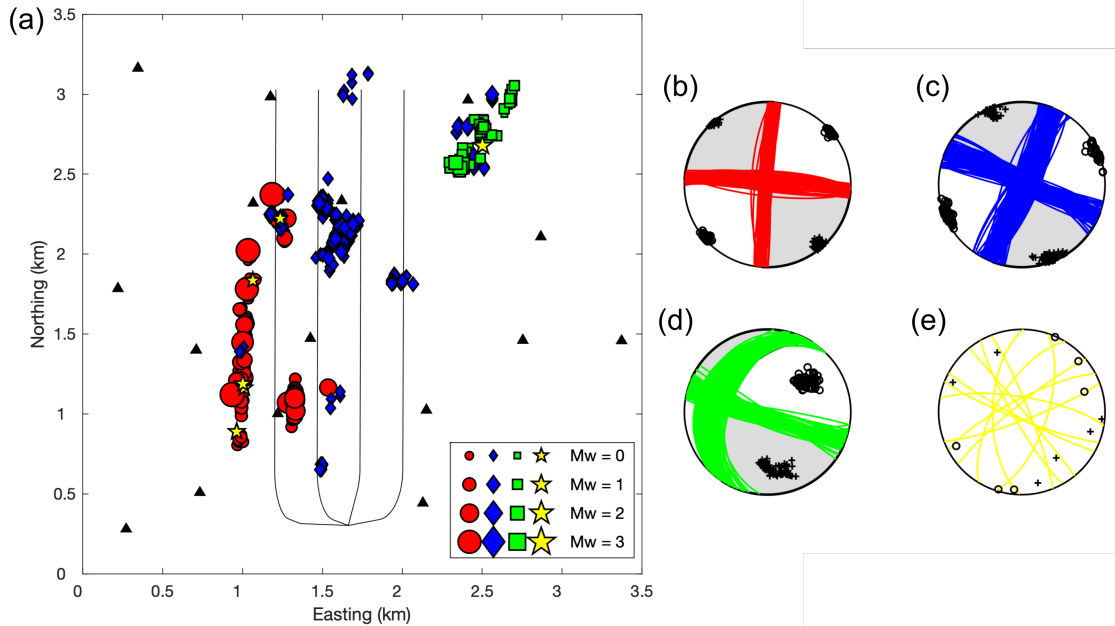

FIG. 2. (a) Spatial distributions for the four groups of events used in this study, and best double-couple components of the moment tensors obtained by Zhang et al. (2019) for (a) group one, (b) group two, (c) group three and (d) group four. In figure (a), the circle, diamond, square and pentagram represent events within group one, group two, group three and group four, respectively, and these events symbols are scaled based on moment magnitude. In the beachball diagrams, black circles and crosses denote P and T axes, respectively (modified from Zhang et al., 2019).

In this study, we only focus on the data from the dense shallow-borehole array and estimate shear-tensile focal mechanisms using the proposed neural network for the 530 high-quality events shown in Figure 2. The adopted 1-D layered velocity model and work-flow for P-wave-amplitude picking were described by Zhang et al. (2019). It is worth noting that the proposed neural network can be easily extended to include S-waves in the inversion. In the case of the ToC2ME dataset, however, P-wave coda signals (mainly re-verberations within the sedimentary layers) interfere with precise picking of direct S-wave, especially for low-magnitude events. This interference leads to unstable amplitude picks for the S-wave. In addition, a test implemented by Zhang et al. (2019) showed that, com-pared with the source-mechanism solutions obtained with both P- and S-waves, the use of only P-wave amplitudes yielded more consistent nodal-plane solutions with the linear alignments of microseismic events and lower misfits. Thus, S-wave amplitudes are not included in the source-mechanism inversion in this study.

Results

Since a total of 530 events recorded by 68 shallow-borehole stations are used to train the physics-guided neural network, the input dimension for the first FC layer is 530×68 (i.e., $N_1 \times N_2$). The network is trained with Adam algorithm and a fixed learning rate of 0.0001. With the batch size of 530 events, the training process takes approximately 220 seconds on Intel Core i7-8700 CPU (3.20 GHz) for 10,000 epochs. Then, the source mechanisms

generated at the 10,000th iteration are the optimized shear-tensile focal mechanisms for the 530 events. Figure 3 (a) shows the loss curve during training, in which at the first 1,000 epochs the loss decreases rapidly and significantly with a reduction of $\sim 97\%$ of the initial loss value. Then the loss decreases slowly, and the training converges to an optimal solution after 6,000 epochs. To evaluate the retrieved focal mechanisms, the percentage misfit between the modeled and observed displacement amplitudes is calculated for each individual event, which is defined as

$$\Delta a_k = \frac{(\mathbf{a}_k^{obs} - \mathbf{a}_k^{mod})^T (\mathbf{a}_k^{obs} - \mathbf{a}_k^{mod})}{(\mathbf{a}_k^{obs})^T \mathbf{a}_k^{obs}}, (k = 1, 2, 3, , 530). \tag{8}$$
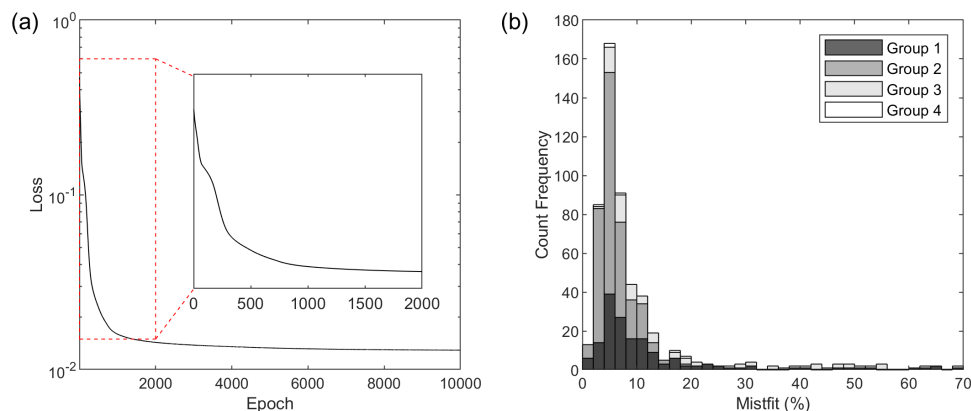
FIG. 3. (a) Loss curve using the physics-guided neural network during training and (b) histogram of misfits between the observed and modeled normalized P-wave displacement amplitudes.

As shown in Figure 3 (b), the misfits for most events within the four groups are relatively small, and 494 out of 530 events have misfits less than 30%. The overall misfit yielded here is comparable to that by Zhang et al. (2019) for the same dataset. This is significant, since the previous study used a full moment-tensor model which provided additional degrees of freedom. Figures 4 (a)-(d) show the DC focal mechanisms for the four groups of events, which exhibit similar features to the those in Figures 2 (b)-(e). Namely, events within both groups one and two have strike-slip dominated mechanisms along subvertical nodal planes but with different strike directions, and that events within group three exhibit either oblique dip-slip on subvertical plane or shallow-dipping slip. To evaluate the difference between the two sets of results in Figures 2 (b)-(e) and Figures 4 (a)-(d), the Kagan angle (Kagan, 1991) is adopted here, which represents the minimum 3D rotation angle between two DC focal mechanisms. The value of Kagan angle ranges between $0°$ and $120°$, and a value of $120°$ indicates significant difference between two sets of focal mechanisms. Figure 4 (e) shows the histogram of Kagan angles between results in Figures 2 (b)-(e) and Figures 4 (a)-(d). It can be observed that most events within both group one and group two have Kagan angles less than $10°$, and most events within group three have relatively large Kagan angles between $10°$ and $25°$. According to previous studies (e.g., Lee et al., 2014; Nakamura et al., 2016), a Kagan angle of less than $30°$ indicates that two focal mechanisms are similar. Based on our calculation, all the retrieved Kagan angles are less than $30°$, indicating a significant similarity between the two sets of results and confirming the effectiveness of our proposed physics-guided neural network.
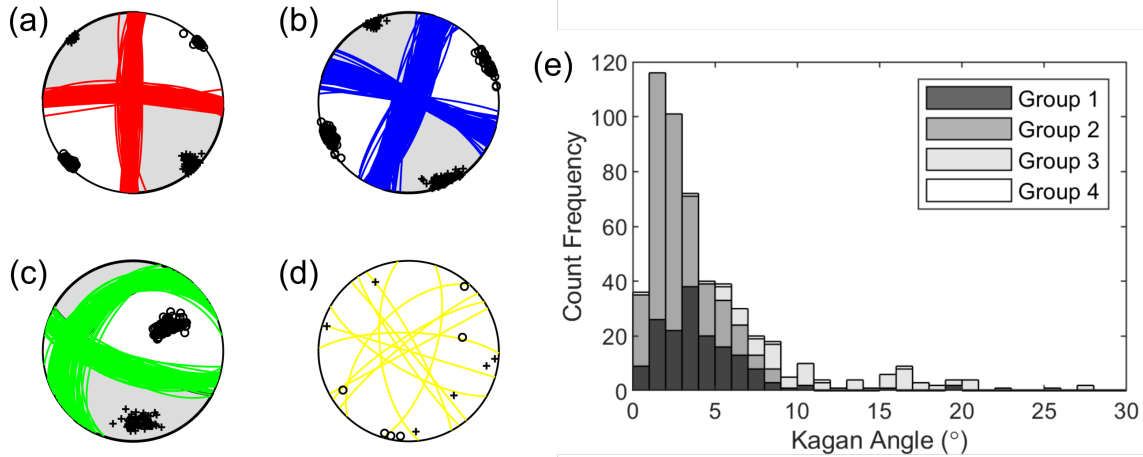
FIG. 4. DC components of the shear-tensile focal mechanisms for events within (a) group one, (b) group two, (c) group three, (d) group four and (e) histogram of the Kagan angles between the DC components estimated using physics-guided neural network and those obtained in Figures 2 (b)-(e) by Zhang et al. (2019). Black circles and crosses in the beachball diagrams denote P and T axes, respectively.

In addition to the DC components, our neural network also estimates the slope for each individual event. Figure 5 displays the histogram of slopes for the 530 events. In general, most events exhibit positive slope values, indicating a preference for tensile opening rather than closing. In addition, most events within groups one and two have small slope values, ranging between -2° and 10°, indicative of almost pure double couple. In comparison, relatively large slopes are evident for events within group three; nevertheless, the slope range between 10° and 20° for this group still reveals DC-dominated source mechanisms.
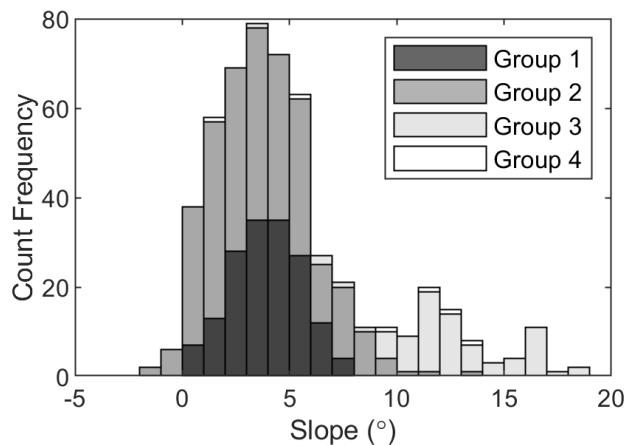
FIG. 5. Histogram of slopes for the four groups of events.

To illustrate the training process for individual events, we select one representative event from each of the three major groups. Figure 6 shows misfit versus epoch plots for these three events; the results for only the first 2,000 epochs are shown for the purpose of demonstration. Starting from the 100th epoch, the beachball diagrams for the three events are shown after every 200 epochs, and Kagan angles between the DC components at the selected epoch and final epoch (i.e., epoch 10,000) are also presented below beachball

diagrams. For all three events, the misfit decreases rapidly initially, followed by a relatively slow decline in the misfit curve. Although the three events share similar features in the misfit curves, different convergence speed can be observed for the three events in Figure 6. For the two events within group one and group two, no noticeable changes in the beachball diagrams can be observed after ~500 epochs, and Kagan angles between the DC source mechanisms at 1,900th epoch and 10,000th epoch are 1.9° and 0.7°, respectively. Whereas, the beachball diagram for the event in group three still keeps updated after 500 epochs, and the difference between focal mechanisms at 1,900th epoch and 10,000 epoch is relatively large (with a Kagan angle of 4.1°), which indicates relatively slow convergence speed than the other two events, and requires more epochs of training to get the optimized solution. Figure 7 shows beachball diagrams of the retrieved shear-tensile focal mechanisms for the three events in Figure 6, in which observed normalized P-wave amplitudes are also projected and represented by the circles with different size. It is evident that the focal-mechanism solutions are well constrained by the P-wave-amplitude data due to the good azimuthal coverage of stations and generally high signal-to-noise ratio for these events.
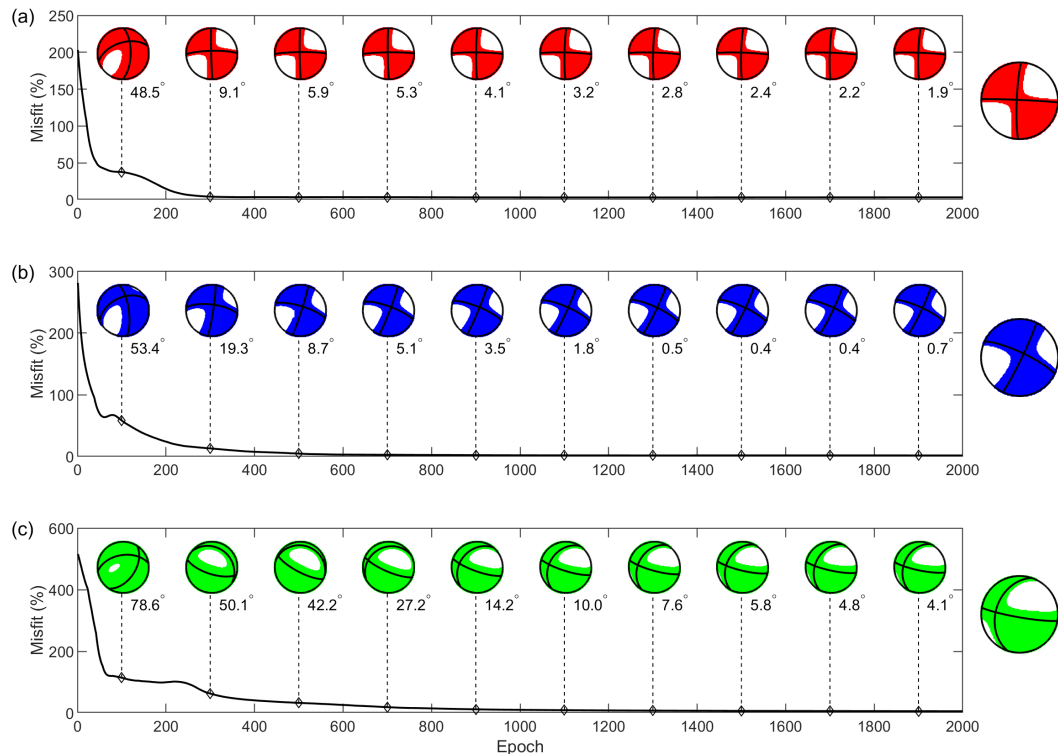
FIG. 6. Misfit curves for the first 2,000 epochs for three representative events from (a) group one, (b) group two and (c) group three. Beachball diagrams and Kagan angles between focal mechanisms at selected epochs and the final epoch are also indicated. Beachball diagrams at the last epoch are shown on the right side of the three subplots.

DISCUSSION

As presented in Figure 4 (e), compared with the other two major groups, the DC components for events in group three have relatively large Kagan angles. This group of events
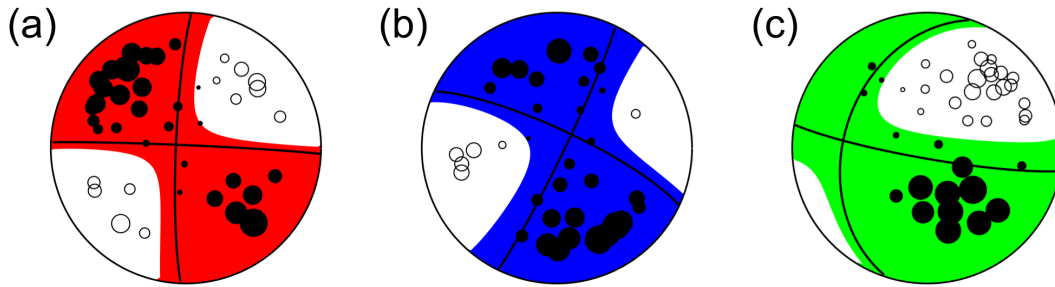
FIG. 7. Beachball diagrams of the retrieved shear-tensile focal mechanisms for three representative events from (a) group one, (b) group two and (c) group three. Solid and open circles represent observed positive and negative P-wave polarities, respectively. The size of the circle is scaled by observed normalized P-wave amplitude in the absolute sense.

is characterized by complex source mechanisms. The NE-SW trending nodal planes for this group of events are consistent with the orientation of linear alignment for this group of events (shown in Figure 2). In addition to the induced-microseismicity revealed NE-SW trending fault, an NNW-SSE trending basement-rooted fault has also been interpreted based on the 3D seismic data (Eaton et al., 2018; Igonin et al., 2018; Weir et al., 2018). The presence of near-synchronous, multi-slip events on two distinct fault planes within the pre-existing fault system was found to be one of the possible causes of the significant non-DC components observed for events within group three (Zhang et al., 2019). Compared with full moment-tensor inversion, a simpler model consisting of four parameters is used to characterize the source mechanism in this study. In the presence of complex source mechanisms (e.g., the co-slipping along pre-existing faults), the results using the proposed neural network may yield relatively large difference from the best DC components of conventional moment-tensor solutions due to its incapability to fit the complex observed amplitude data generated by multiple fault strands.

A number of sources of uncertainty warrant discussion. For example, a simplified 1-D layered isotropic velocity model is used to calculate the take-off angles at sources, and the vertical or lateral deviations from the true velocities may cause errors within the estimated focal-mechanism solutions. In addition, the event-location error and noise within the observed data may also distort the solutions (Vavryčuk and Kühn, 2012; Tan et al., 2018). To examine the robustness of our proposed neural network, similar to the study by Zhang et al. (2019), we conduct a series of uncertainty tests for the retrieved source mechanisms by implementing the inversion using perturbed event locations, velocities and P-wave amplitudes due to noise, respectively. Since the 530 events considered here were located with uncertainties of less than 30 m for epicenters and less than 70 m for focal depths (Eaton et al., 2018), event locations in the uncertainty analysis are perturbed within ranges of $\pm$ 30 m and $\pm$ 70 m for epicenters and focal depths, respectively. For the 1-D layered velocity model, the velocity value of each layer is randomly perturbed with a maximum value of up to 10%. This range of perturbation is reasonable for the well-log-data derived velocities in the microseismic data processing, especially in the case where the velocity model is calibrated with perforation shots (Li et al., 2011; Kuang et al., 2017). In addition, we also examine the effects on the inversion results from noise by adding different levels of noise to the P-wave amplitude picked at each individual station. Six noise levels are in-

vestigated here ranging from 5%, 10% until 30% of the picked P-wave amplitudes (in the absolute sense) at individual stations. For each the uncertainty test presented here, the inversion is performed 50 times, and the results are then compared with those obtained with unperturbed parameters. Table 1 shows the mean Kagan angles and slope deviations between the results with perturbed and unperturbed parameters. It can be observed that the results for slopes are robust for all the perturbation cases, and the mean deviations from slopes obtained with unperturbed parameters are less than 1°. Compared with slopes, the DC mechanisms exhibit relatively large deviations with mean Kagan angles less than 8° for all the cases, nevertheless, this Kagan angle range still indicates similar focal mechanisms according to previous studies (e.g., Lee et al., 2014; Nakamura et al., 2016. In addition, we can also observe from Table 1 that with the increase of noise level, both the mean slope deviation and Kagan angle exhibit only a slight increase, e.g., for noise levels of 5% and 30%, the mean slope deviation and Kagan angle increase from 0.67°, 6.28° to 0.94°, 7.28°, indicating the high tolerance of the neural network to noise. Table 2 shows the DC mechanisms obtained with perturbed parameters for three representative events from three major groups, in which the beachball diagrams for 50 realizations of the source-mechanism inversion are presented. For each of the three events, the beachball diagrams exhibit highly similar features, and only minor variations in the nodal-plane solutions could be observed. In addition, with the increase of noise level, the distribution of nodal planes for the same event becomes relatively sparse, indicating the increased uncertainty.

Table 1. Mean Kagan angles and slope deviations between inversion results using perturbed and unperturbed parameters with standard deviations indicated.

Perturbed Parameters		Slope (°)	Kagan Angle (°)
Event Location		0.84 ± 1.23	6.54 ± 6.72
Velocity		0.83 ± 1.28	7.11 ± 7.04
Noise	5%	0.67 ± 1.19	6.28 ± 6.86
	10%	0.70 ± 1.20	6.42 ± 6.91
	15%	0.74 ± 1.19	6.56 ± 6.87
	20%	0.79 ± 1.20	6.78 ± 6.90
	25%	0.86 ± 1.26	7.01 ± 7.07
	30%	0.94 ± 1.43	7.28 ± 7.02

Our proposed neural network is implemented with PyTorch, and the training for 530 events with 68 shallow-borehole stations takes approximately 220 seconds on an Intel Core i7-8700 CPU (3.20 GHz). To examine the computation time for different input sizes under the same acquisition geometry, we perform source-mechanism inversions separately with different numbers of events ranging from 100, 200 until 1,000 events. As a comparison, we also time the forwarding modeling block for 10,000-time runs with different numbers of events. As shown in Figure 8, the computation time used for the forward modeling block exhibits an approximate linear increase, and it takes ~ 75 seconds for 10,000 times of forward modeling for 1,000 events, which takes up $\sim 20\%$ of the total time used for neural network training. Similarly, the total training time also increases with the increase of input data size, and the training with 1,000 events for 10,000 epochs takes ~ 6 minutes.

Table 2. Beachball diagrams of the DC solutions using perturbed parameters for three representative events.

Compared with the conventional full moment-tensor inversion, the use of shear-tensile model improves the inversion robustness by decreasing the number of model parameters (i.e., from six to four in this study) and reduction in linearity (Jechumtálová et al., 2014). For the sake of comparison, here we also estimate the shear-tensile focal mechanisms for the 530 events with a two-step grid-search algorithm similar to the one proposed by (Jechumtálová et al., 2014), in which a coarse grid (5°) is first applied across the whole model space, then a fine grid (0.2°) is used in the vicinity of the best solution obtained from the previous step. Figure 9 shows histograms of slope deviations as well as Kagan angles between the two sets of results obtained with the proposed neural network and two-step grid-search method. As shown in Figure 9 (a), the Kagan angles for most events are less than 10°, indicating highly similar double-couple components. In addition, relatively small slope deviations (mainly within the range of -5° – 5°) are also presented in Figure 9 (b). The comparison conducted here further validates the reliability of our solutions. In terms of the two-step grid search, this inversion procedure is repeated for each individual event, which takes ∼ 90 seconds on average for each event. In contrast, our proposed neural network simultaneously estimates source mechanisms for multiple events, and only takes ∼ 220 seconds in total.
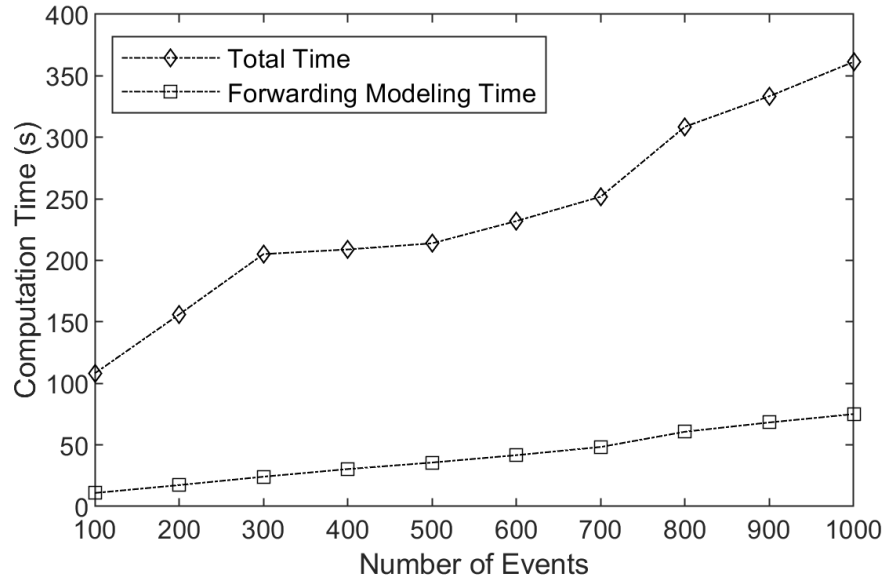
FIG. 8. Computation times for the network training and the forward modeling block with various numbers of input events.
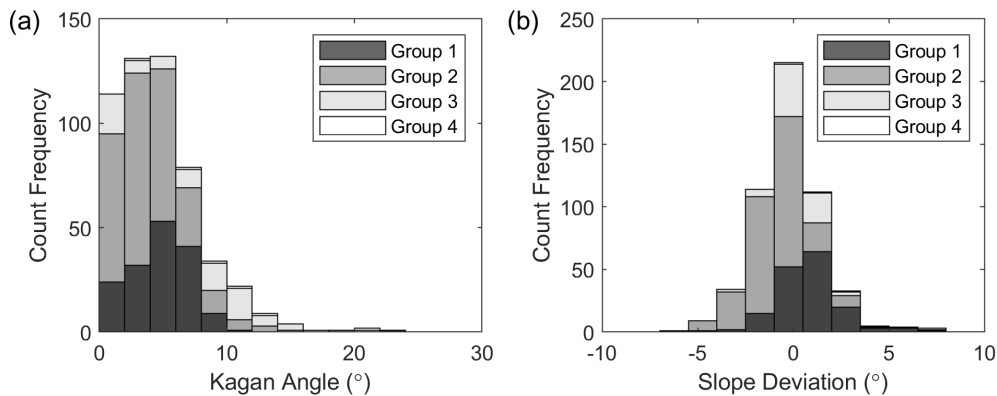

FIG. 9. Histograms of (a) Kagan angles and (b) slope deviations between the two sets of results retrieved with the proposed network and the two-step grid-search algorithm.

CONCLUSIONS

We have developed a physics-guided neural network to simultaneously estimate shear-tensile focal mechanisms for multiple microearthquakes recorded by the dense shallow-borehole array, in which the displacement amplitudes of direct P-wave are used as input. Compared with conventional data-driven neural networks, the proposed neural network doesn't require the training data due to the incorporation of a forward modeling layer (i.e., the Radiation Pattern & Normalization layer). The incorporated physics of the forward problem is used to calculate theoretical normalized P-wave amplitudes for any given source mechanisms from its previous layer, and the mean squared error between the modeled and observed amplitude data is used as the loss function. The network weights and source-mechanism solutions are then optimized by the backpropagation process of the proposed neural network.

We have applied the proposed neural network to 530 induced events recorded by dense shallow-borehole array stations during hydraulic-fracturing operations in the Duvernay Shale. In general, similar source mechanisms to a previous study have been retrieved for these events with relatively low misfits, and three major groups have been identified based on source-mechanism solutions. Compared with the other two major groups, the source mechanisms for group three have exhibited relatively large Kagan angles ($10°$–$25°$) with solutions obtained in previous study in which the source mechanisms for this group were interpreted by the possible co-slipping of pre-existing faults. This relatively large source-mechanism deviation is mainly due to the insufficiency of the incorporated shear-tensile model to characterize complex source mechanisms.

In addition, a series of uncertainty tests presented here indicate that the proposed physics-guided neural network is tolerant to errors in event locations, velocity model and P-wave amplitudes, and only relatively small slope deviations ($< 1°$) and Kagan angles ($< 8°$) have been observed between the results using perturbed and unperturbed data. The comparison with the results using a two-step grid search validates the reliability of our results and demonstrates significant improvement in computational efficiency.

ACKNOWLEDGEMENTS

REFERENCES

Biswas, R., Sen, M. K., Das, V., and Mukerji, T., 2019, Prestack and poststack inversion using a physics-guided convolutional neural network: Interpretation, **7**, No. 3, SE161–SE174.

Downton, J., Collet, O., and Colwell, T., 2019, Rock-physics based augmented machine learning for reservoir characterization, *in* EAGE Subsurface Intelligence Workshop, vol. 2019, European Association of Geoscientists & Engineers, 1–5.

Dreger, D. S., Tkalčić, H., and Johnston, M., 2000, Dilational processes accompanying earthquakes in the long valley caldera: Science, **288**, No. 5463, 122–125.

Eaton, D. W., and Forouhideh, F., 2011, Solid angles and the impact of receiver-array geometry on microseismic moment-tensor inversion: Geophysics, **76**, No. 6, WC77–WC85.

Eaton, D. W., Igonin, N., Poulin, A., Weir, R., Zhang, H., Pellegrino, S., and Rodriguez, G., 2018, Induced seismicity characterization during hydraulic-fracture monitoring with a shallow-wellbore geophone array and broadband sensors: Seismological Research Letters, **89**, No. 5, 1641–1651.

Ebel, J. E., and Bonjer, K. P., 1990, Moment tensor inversion of small earthquakes in southwestern germany for the fault plane solution: Geophysical Journal International, **101**, No. 1, 133–146.

Godano, M., Bardainne, T., Regnier, M., and Deschamps, A., 2011, Moment-tensor determination by nonlinear inversion of amplitudes: Bulletin of the Seismological Society of America, **101**, No. 1, 366–378.

Hardebeck, J. L., and Shearer, P. M., 2003, Using s/p amplitude ratios to constrain the focal mechanisms of small earthquakes: Bulletin of the Seismological Society of America, **93**, No. 6, 2434–2444.

Heaton, J., 2008, Introduction to neural networks with Java: Heaton Research, Inc.

Igonin, N., Zecevic, M., and Eaton, D. W., 2018, Bilinear magnitude-frequency distributions and characteristic earthquakes during hydraulic fracturing: Geophysical Research Letters, **45**, No. 23, 12–866.

Jechumtálová, Z., and Šílený̀, J., 2005, Amplitude ratios for complete moment tensor retrieval: Geophysical Research Letters, **32**, No. 22.

Jechumtálová, Z., Šílený̀, J., and Trifu, C., 2014, Microearthquake mechanism from wave amplitudes recorded by a close-to-surface seismic array at ocnele mari, romania: Geophysical Journal International, **197**, No. 3, 1608–1626.

Julian, B. R., Miller, A. D., and Foulger, G. R., 1998, Non-double-couple earthquakes 1. theory: Reviews of Geophysics, **36**, No. 4, 525–549.

Kagan, Y. Y., 1991, 3-d rotation of double-couple earthquake sources: Geophysical Journal International, **106**, No. 3, 709–716.

Käufl, P., Valentine, A., de Wit, R., and Trampert, J., 2015, Robust and fast probabilistic source parameter estimation from near-field displacement waveforms using pattern recognitionrobust and fast probabilistic source parameter estimation using pattern recognition: Bulletin of the Seismological Society of America, **105**, No. 4, 2299–2312.

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

Kisslinger, C., 1980, Evaluation of s to p amplitude rations for determining focal mechanisms from regional network observations: Bulletin of the Seismological Society of America, **70**, No. 4, 999–1014.

Kuang, W., Zoback, M., and Zhang, J., 2017, Estimating geomechanical parameters from microseismic plane focal mechanisms recorded during multistage hydraulic fracturing: Geophysics, **82**, No. 1, KS1–KS11.

Lee, S. J., Liang, W. T., Cheng, H. W., Tu, F. S., Ma, K. F., Tsuruoka, H., Kawakatsu, H., Huang, B. S., and Liu, C. C., 2014, Towards real-time regional earthquake simulation i: Real-time moment tensor monitoring (rmt) for regional events in taiwan: Geophysical Journal International, **196**, No. 1, 432–446.

Li, J., Zhang, H., Sadi Kuleli, H., and Nafi Toksoz, M., 2011, Focal mechanism determination using high-frequency waveform matching and its application to small magnitude induced earthquakes: Geophysical Journal International, **184**, No. 3, 1261–1274.

Nakamura, W., Uchida, N., and Matsuzawa, T., 2016, Spatial distribution of the faulting types of small earthquakes around the 2011 tohoku-oki earthquake: A comprehensive search using template events: Journal of Geophysical Research: Solid Earth, **121**, No. 4, 2591–2607.

Ou, G. B., 2008, Seismological studies for tensile faults: TAO: Terrestrial, Atmospheric and Oceanic Sciences, **19**, No. 5, 4.

Ovcharenko, O., Akram, J., and Peter, D., 2018, Feasibility of moment tensor inversion from a single borehole data using artificial neural networks.

Pesicek, J. D., Cieślik, K., Lambert, M.-A., Carrillo, P., and Birkelo, B., 2016, Dense surface seismic data confirm non-double-couple source mechanisms induced by hydraulic fracturinginduced non-dc source mechanisms: Geophysics, **81**, No. 6, KS207–KS217.

Petružálek, M., Jechumtálová, Z., Kolář, P., Adamová, P., Svitek, T., Šílenỳ, J., and Lokajíček, T., 2018, Acoustic emission in a laboratory: mechanism of microearthquakes using alternative source models: Journal of Geophysical Research: Solid Earth, **123**, No. 6, 4965–4982.

Ren, Y., Xu, X., Yang, S., Nie, L., and Chen, Y., 2020, A physics-based neural-network way to perform seismic full waveform inversion: IEEE Access, **8**, 112,266–112,277.

Rodríguez-Pradilla, G., and Eaton, D. W., 2019, The application of coda and energy methods for magnitude estimation of microseismic events: Seismological Research Letters, **90**, No. 3, 1296–1307.

Rögnvaldsson, S. T., and Slunga, R., 1993, Routine fault plane solutions for local networks: a test with synthetic data: Bulletin of the Seismological Society of America, **83**, No. 4, 1232–1247.

Shang, X., and Tkalčić, H., 2020, Point-source inversion of small and moderate earthquakes from p-wave polarities and p/s amplitude ratios within a hierarchical bayesian framework: Implications for the geysers earthquakes: Journal of Geophysical Research: Solid Earth, **125**, No. 2, e2019JB018,492.

Šílenỳ, J., 2018, Constrained moment tensors: source models and case studies, *in* Moment Tensor Solutions, Springer, 213–231.

Šílenỳ, J., Hill, D. P., Eisner, L., and Cornet, F. H., 2009, Non–double-couple mechanisms of microearthquakes induced by hydraulic fracturing: Journal of Geophysical Research: Solid Earth, **114**, No. B8.

Šílenỳ, J., Jechumtálová, Z., and Dorbath, C., 2014, Small scale earthquake mechanisms induced by fluid injection at the enhanced geothermal system reservoir soultz (alsace) in 2003 using alternative source models: Pure and Applied Geophysics, **171**, No. 10, 2783–2804.

Spikin, S. A., 1986, Estimation of earthquake source parameters by the inversion of waveform data: global seismicity, 1981-1983: Bulletin of the Seismological Society of America, **76**, No. 6, 1515–1541.

Staněk, F., Eisner, L., and Moser, T. J., 2014, Stability of source mechanisms inverted from p-wave amplitude microseismic monitoring data acquired at the surface: Geophysical Prospecting, **62**, No. 3, 475–490.

Stierle, E., Bohnhoff, M., and Vavryčuk, V., 2014, Resolution of non-double-couple components in the seismic moment tensor using regional networks—ii: application to aftershocks of the 1999 m w 7.4 izmit earthquake: Geophysical Journal International, **196**, No. 3, 1878–1888.

Sun, J., Niu, Z., Innanen, K. A., Li, J., and Trad, D. O., 2020, A theory-guided deep-learning formulation and optimization of seismic waveform inversion: Geophysics, **85**, No. 2, R87–R99.

Tan, Y., Zhang, H., Li, J., Yin, C., and Wu, F., 2018, Focal mechanism determination for induced seismicity using the neighbourhood algorithm: Geophysical Journal International, **214**, No. 3, 1715–1731.

Ursin, B., and Hokstad, K., 2003, Geometrical spreading in a layered transversely isotropic medium with vertical symmetry axis: Geophysics, **68**, No. 6, 2082–2091.

Vavryčuk, V., 2001, Inversion for parameters of tensile earthquakes: Journal of Geophysical Research: Solid Earth, **106**, No. B8, 16,339–16,355.

Vavryčuk, V., 2011, Tensile earthquakes: Theory, modeling, and inversion: Journal of Geophysical Research: Solid Earth, **116**, No. B12.

Vavryčuk, V., and Kühn, D., 2012, Moment tensor inversion of waveforms: a two-step time-frequency approach: Geophysical Journal International, **190**, No. 3, 1761–1776.

Weir, R. M., Eaton, D. W., Lines, L. R., Lawton, D. C., and Ekpo, E., 2018, Inversion and interpretation of seismic-derived rock properties in the duvernay play: Interpretation, **6**, No. 2, SE1–SE14.

Zhang, H., Eaton, D. W., Li, G., Liu, Y., and Harrington, R. M., 2016, Discriminating induced seismicity from natural earthquakes using moment tensors and source spectra: Journal of Geophysical Research: Solid Earth, **121**, No. 2, 972–993.

Zhang, H., Eaton, D. W., Rodriguez, G., and Jia, S. Q., 2019, Source-mechanism analysis and stress inversion for hydraulic-fracturing-induced event sequences near fox creek, albertasource-mechanism analysis and stress inversion: Bulletin of the Seismological Society of America, **109**, No. 2, 636–651.

Zhang, R., Liu, Y., and Sun, H., 2020, Physics-guided convolutional neural network (phycnn) for data-driven seismic response modeling: Engineering Structures, **215**, 110,704.

Zhu, L., and Helmberger, D. V., 1996, Advancement in source estimation techniques using broadband regional seismograms: Bulletin of the Seismological Society of America, **86**, No. 5, 1634–1641.