# Deep-learning-based local wavefront attributes and their application to 3D prestack data enhancement

Kirill Gadylshin[1], Ilya Silvestrov[2], and Andrey Bakulin[2]

## ABSTRACT

The work presents a novel workflow to accelerate the estimation of local wavefront attributes (LWAs) from massive 3D prestack seismic data using deep learning (DL) focusing on data enhancement. A standard estimation method based on a semblance-based brute-force optimization provides good results but is time consuming. A modification of the U-net convolutional neural network, commonly used in image processing applications, is proposed to link the seismic data with the wavefront attributes. Color pixel image input for the neural network is generated through a straightforward seismic data regularization based on supergrouping followed by red, green, and blue encoding. The proposed workflow can be adapted to any 3D prestack seismic volume. Conventional semblance-based attributes estimation is required for the training step but only for approximately 1% of the total data. The prediction step is very efficient and reduces the overall run time significantly. The verification of the proposed approach is performed on challenging real land and marine data sets. As a result, DL-based estimation of LWAs accelerates computation up to 200 times compared to the standard method. The attributes from the proposed DL-based approach indicate an acceptable match compared with the brute-force semblance-based optimization results. Conventional and proposed estimation methods result in comparable prestack data enhancement results for more reliable seismic processing in challenging areas.

## INTRODUCTION

Extracting kinematic traveltime information from prestack seismic wavefields can enhance seismic data processing and imaging. Routinely, a conventional velocity analysis is used for this purpose by fitting the reflection traveltimes with a global hyperbolic curve. An estimated normal-moveout velocity represents the reflection wavefield's most commonly used kinematic wavefront attribute. Such a global attribute is defined at zero offset. It approximates the reflection moveout for all offsets within a small-spread approximation and all azimuths at an arbitrary common-midpoint position. In contrast, local kinematic wavefront attributes are estimated for each particular offset, azimuth, or any other point in a 5D prestack data volume. The most known local kinematic attributes are dips of local events (first-order traveltime derivatives) estimated from prestack data in common-shot or common-receiver gathers, which definition can be traced back to the method of controlled directional reception (Rieber, 1936; Riabinkin, 1957). Wavefront curvatures defined by the second-order traveltime derivatives represent more advanced attributes characterizing the kinematics of the wavefield.

Kinematic wavefront attributes are used by a variety of seismic data processing methods. The first-order traveltime derivatives are used in slope- or stereo-tomography methods (Billette and Lambaré, 1998; Lambaré, 2008; Lambaré et al., 2014; Bakulin et al., 2021a) to estimate depth velocity models. Fomel (2007) shows how local event slopes can be used to accomplish different time-domain imaging tasks, from normal-moveout correction and Dix inversion to prestack time migration. Curvatures of normal and normal-incidence-point wavefronts at zero offset are used in the common-reflection-surface (CRS) method to produce stacked seismic sections with an improved signal-to-noise ratio (S/N) (Mann et al., 1999). An alternative approach for such enhanced stacking is the multifocusing (MF) method (Berkovitch et al., 2008), formulated in terms of common-reflection-element and common-reflection-point wavefronts' curvatures. In addition to stacking, the zero-offset kinematic parameters from these two methods can be used for reflection tomography (Duveneck, 2004), diffraction imaging (Berkovitch et al., 2009; Rad et al., 2018), data interpolation (Xie and Gajewski, 2017), and prestack data enhancement (Baykulov and Gajewski, 2009). Common-offset generalizations of the CRS and MF methods use a larger number of local

[1]Institute of Petroleum Geology and Geophysics, Novosibirsk, Russia. E-mail: gadylshin@gmail.com (corresponding author).
[2]EXPEC Advanced Research Center, Saudi Aramco, Dhahran, Saudi Arabia. E-mail: ilya.silvestrov@aramco.com; andrey.bakulin@aramco.com.

attributes to describe the kinematics of the wavefront at an arbitrary offset. Thus, they do not rely on global hyperbolic assumptions of the zero-offset approximations and provide better stacking and data enhancement capabilities (Zhang et al., 2001; Müller and Spinner, 2010; Berkovitch et al., 2011). A more general representation of the local stacking operators as the second-order traveltime approximations is used by Hoecht et al. (2009) to interpolate seismic data and by Buzlukov and Landa (2013) for prestack signal enhancement in the common-offset domain. Bakulin et al. (2018b, 2020) extend this approach to arbitrary seismic gathers and develop nonlinear beamforming (NLBF) technique to efficiently enhance massive 3D prestack land seismic data for different applications, such as imaging, first-break picking, and full-waveform inversion (FWI).

Estimating the kinematic attributes is usually the computationally demanding part of data processing algorithms discussed previously. For modern high-density 3D seismic surveys, the size of the prestack data can reach thousands of terabytes. Consequently, computing resources and runtime become a bottleneck when the tedious estimation of the kinematic attributes must be executed on such massive data volumes. This often requires millions of core hours and becomes an insurmountable limitation for real-data applications. In addition, their multiple recomputations might be necessary at different spatial scales for adaptive multiscale processing (Bakulin et al., 2019). Therefore, different approaches are proposed to optimize the estimation stage. For example, the earliest applications of the CRS method rely on a search in a dimension-reduced data set, in which different kinematic attributes are estimated by optimizing a semblance-based coherency function in sequential order (Mann et al., 1999). To facilitate the coherency optimization, more advanced approaches use global search methods, such as the evolutionary-based Nelder-Mead algorithm (Xie and Gajewski, 2016), very fast simulated annealing (Garabito, 2018), or an efficiency-improved genetic algorithm (Sun et al., 2022a). In addition, different strategies can be applied for estimating the kinematic parameters based on quality versus performance trade-offs (Bakulin et al., 2021b). Furthermore, modern hardware architectures also can improve performance (Sun et al., 2022b). Other schemes are proposed to estimate the kinematic parameters as an alternative to the coherency-based optimization approaches described previously (using semblance or other coherency measures as a cost function). They include a plane-wave destruction method (Fomel, 2002; Santos et al., 2011), a multiple linear regression representation (Hellman, 2014), and a structure-tensor-based approach (Waldeland et al., 2018). However, when one faces the low quality of seismic data, which often is the case in land seismic applications, the standard semblance optimization approaches generally demonstrate the most robust results.

Recently, a strong interest in machine-learning-based approaches has emerged in seismic data processing and interpretation (Yu and Ma, 2021). Different techniques are proposed to attenuate the strong noise in the data. Zhu et al. (2019) use DeepDenoiser neural network to separate noise and signal by learning a nonlinear regression. Yu et al. (2019) apply an approach based on a deep convolutional neural network (CNN) initially proposed for image denoising by Zhang et al. (2017) to remove three kinds of seismic noise: random noise, linear noise, and multiples. Kaur et al. (2020) propose a CycleGAN algorithm to suppress ground roll. A physics-constrained solution based on a combination of unsupervised and supervised deep-learning (DL) approaches for ground-roll attenuation is proposed by Pham and Li (2022). In seismic interpretation, the locali-

zation of faults, dips, and layers is similar to object detection problems in computer vision. Deep neural networks (DNNs) for image classification are used in seismic attribute analysis (Das et al., 2019; Feng et al., 2020; You et al., 2020). Zu et al. (2021), Huang et al. (2021), and Gadylshin et al. (2021) present approaches for local slope estimation by CNNs directly from the prestack data. A method for interpolation of wavefront attributes using inpainting with a partially CNN is presented by Gadylshin et al. (2020).

This paper explores a DL-based approach to significantly accelerate the estimation of local wavefront attributes (LWAs) with a focus on prestack data enhancement using the NLBF method. The outline of the paper is the following. First, the NLBF method and the LWAs are briefly discussed. Then, we describe the neural network architecture used to derive the attributes from the prestack data and explain its training and application workflow. Finally, the obtained results for challenging marine and land data are presented and discussed, focusing on prestack data enhancement.

## METHOD

### Local wavefront attributes

First, let us briefly define the local kinematic wavefront attributes. Because the main focus of this work is the enhancement of prestack seismic data using NLBF (Bakulin et al., 2020), the definition is tied to this approach; however, the extension of the results to other approaches and applications is straightforward. According to the following formula, enhancement with NLBF constitutes a local weighted summation of neighboring traces using local time-shift corrections:

$$u(x_0, y_0, t_0) = \sum_{x \in B_0} w(x, y) u(x, y, t_0 + \Delta t(x, y; x_0, y_0)), \quad (1)$$

where $u(x, y; t)$ is a trace with spatial coordinates $x$ and $y$ and time $t$ defined at each point of a 3D $X$-$Y$-$T$ prestack data subvolume. The two spatial coordinates are arbitrary and depend on the type of input seismic gather. It could be either receiver $X$ and $Y$ coordinates if we consider common-shot data, shot $X$ and $Y$ coordinates in the case of common-receiver data, or shot $X$ and receiver $Y$ coordinates if we deal with the cross-spread gathers. The enhanced trace's coordinates after beamforming are given by $x_0$ and $y_0$. The summation is accomplished within a local rectangular region $B_0$ around the position of the enhanced trace along a traveltime surface with a moveout $\Delta t(x, y; x_0, y_0)$. In NLBF, it is assumed that a second-order surface can locally approximate this moveout as follows:

$$\Delta t = t(x, y) - t_0(x_0, y_0)$$
$$= A\Delta x + B\Delta y + C\Delta x\Delta y + D\Delta x^2 + E\Delta y^2, \quad (2)$$

where $A$, $B$, $C$, $D$, and $E$ are the unknown wavefront attributes and $\Delta x = x - x_0$ and $\Delta y = y - y_0$ represent the spatial shifts of the summed trace with respect to the output trace. The wavefront attributes are defined independently for each 3D input seismic gather on a regular estimation grid in the $X$-$Y$-$T$ volume. They are functions of the prestack time and the spatial point coordinate and are local in this regard. The estimation grid does not necessarily coincide with the data grid. Usually, it is chosen to be sparser to achieve better performance. More technical details about the grids and the usage of the wavefront attributes in the NLBF method can be found in Bakulin et al. (2020). In the standard estimation approach, the unknown coefficients $A$ and $B$

(wavefront first spatial derivatives or dips) and $C$, $D$, and $E$ (wavefront second derivatives or curvatures) are estimated by brute-force searching over different beamforming surfaces and picking the one with the best coherency defined by a maximum value of a semblance function $S$ (Taner and Koehler, 1969). The conventional NLBF method uses a so-called "2 + 2 + 1" estimation strategy (Buzlukov and Landa, 2013; Bakulin et al., 2021b). In the beginning, a search is performed for parameters $A$ and $D$ in the $X$-$T$ plane and independently for parameters $B$ and $E$ in the $Y$-$T$ plane using 2D cross sections of the seismic data extracted from the 3D subvolume. In the final stage, the last parameter $C$ is estimated in the 3D $X$-$Y$-$T$ cube after fixing the previous four parameters. Likewise, this study focuses on the similar plane-by-plane estimation of parameters $A$, $B$, $D$, and $E$, but using machine-learning approach. The remaining parameter $C$ is not estimated (this is outside the scope of the paper) and is set to zero in all the examples.

## LWA deep neural network

The wavefront attributes in the NLBF method are estimated at each point of a 3D regular spatial and temporal grid covering each input seismic gather, so the semblance optimization problem must be solved many times, resulting in computationally costly operation. This study proposes using a DNN to accelerate the expensive standard semblance-based optimization workflow. CNNs, which are usually applied to the analysis of visual imagery, rely on local coherency (LeCun et al., 2015) and, hence, have the potential to extract coherency information from seismic data as well. A particular case of CNN, initially proposed for biomedical image segmentation, is a U-Net (Ronneberger et al., 2015), an encoder-decoder CNN with skip connections. Our LWA DNN is a modification of the original U-net with additional layers. LWA DNN directly links prestack seismic data with wavefront attributes. The actual network architecture (Figure 1) is based on the one used by Gadylshin et al. (2020) for attributes' inpainting, with the main difference in exploiting a conventional convolutional layer instead of partial convolutions. The LWA DNN encoder part contains eight convolutional layers followed by batch normalization with rectified linear unit (ReLU) activation function (Fukushima, 1969). The decoder part has eight upsampling layers followed by a convolution operation with LeakyReLU activation (Glorot et al., 2011). The LWA DNN input and output are $512 \times 512$ red-green-blue (RGB) images, similar to conventional image-oriented applications. By design, every output image point contains all information from the input image. The TensorFlow (Abadi et al., 2016) library with graphical processor unit (GPU) support is used for the numerical implementation.

## LWA DNN training

The training step is used to estimate the neural network weights according to Figure 2a. We use approximately 1% of the total number of input gathers uniformly distributed over the survey area to create the training data set. First, the LWAs are estimated for this data set using a conventional semblance-based optimization approach on a regular estimation grid, discussed previously. Although
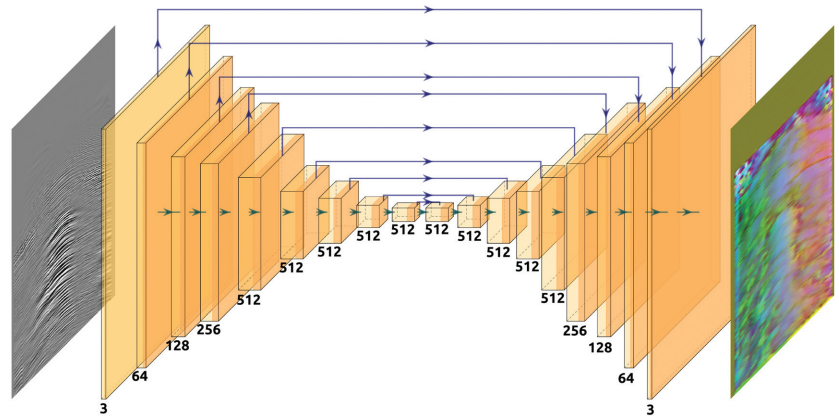


Figure 1. The schematical drawing of the LWAs DNN architecture for direct kinematic attributes estimation. Regularized prestack seismic gather (on the left) flows through the encoder-decoder CNN with skip connections (shown by arrows) and is converted to RGB image (on the right) encoding two output LWAs and semblance. The number of channels in the corresponding layers is shown at the bottom.
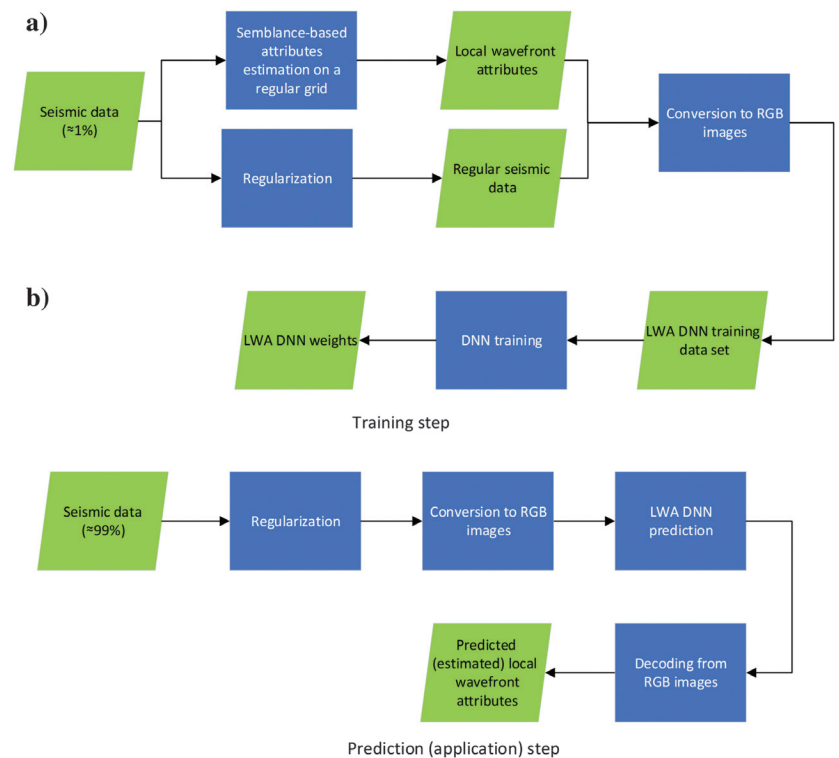


Figure 2. A flowchart of the proposed DNN-based workflow including (a) the training step to calculate the weights of the DNN and (b) the prediction step at which the network is applied to estimate the wavefront attributes.

the attributes are defined on the regular grid, the irregularity in the acquisition usually occurs in practice. Because the neural network's input and output are images with constant pixel size, data regularization is a required step in the proposed workflow. In the current work, we use a so-called supergrouping — a simplified 3D local stacking (after introducing the moveout corrections using legacy velocities) of nearby traces in a sliding window with a small summation aperture (Bakulin et al., 2018a). An example of this step is shown in Figure 3. It shows that supergrouping acts as an efficient regularization that collects signals onto a rectangular grid. Although we use supergrouping for simplicity here, other more advanced regularization techniques also can be applied similarly. The regular seismic data are naturally converted to the RGB image space using a grayscale colormap. This transformation is achieved by rescaling the data onto interval [0,1] via linear mapping. Then, the input image for the training step is created by generating a pixel triplet as $\langle u(x,t), u(x,t), u(x,t) \rangle$, with each colored component (red, green, and blue) having the same value after rescaling. In the case of the wavefront attributes, for each time-space point of the regular estimation grid, a scaled triplet of dip and curvature in one spatial direction and semblance ($\langle A, D, S \rangle$ or $\langle B, D, S \rangle$) is transformed into

a <Red, Green, Blue> color image pixel (Figure 4). Each pair of color images — input seismic gather and corresponding color-coded attributes — form a single training sample, as shown in Figure 1.

To calculate the size of the training data set, let us consider a single 3D seismic gather. The NLBF estimation grid has the size of $N_t^{est} \times N_x^{est} \times N_y^{est}$ grid points, where $N_t^{est}$ is the number of time samples and $N_x^{est}$ and $N_y^{est}$ are the number of estimation grid points along the x- and y-directions, respectively. Performing the NLBF estimation step for the dip and curvature pairs in the X-T plane (A and D attributes) and Y-T plane (B and E attributes) results in $N_x^{est} + N_y^{est}$ training samples. If we assume the fixed size of the seismic gathers and hence the constant size of the estimation grid and $N_G$ is the number of seismic gathers used for the creation of the LWA DNN training data set, then the total number of training samples in the data set equals to $N_G \cdot (N_x^{est} + N_y^{est})$.

The DNN training is performed on a multi-GPU node in parallel mode. For optimization, we use the Adam algorithm with an initial learning rate of $2 \times 10^{-4}$. The loss is mean-square error function. To avoid overfitting, we use the early stopping regularization technique.

## LWA DNN application

After completing the training step and obtaining the neural network weights, we are ready to apply the trained LWA DNN to predict the LWAs for the remaining 99% of the data (Figure 2b). Similar to the training step, the data are first regularized via supergrouping and converted to RGB images using a grayscale colormap. Then, the data go through the trained LWA DNN, which outputs the predicted wavefront attributes encoded in the RGB images. Decoding these images provides the LWAs on the regular estimation grid, similar to the standard semblance-based estimation approach. Finally, these attributes are used for prestack data enhancement following the typical beamforming used by NLBF.

## MARINE DATA EXAMPLE

As a first example, let us consider the modern marine multiazimuth ocean-bottom node (OBN) data set. It is designed for FWI and possesses
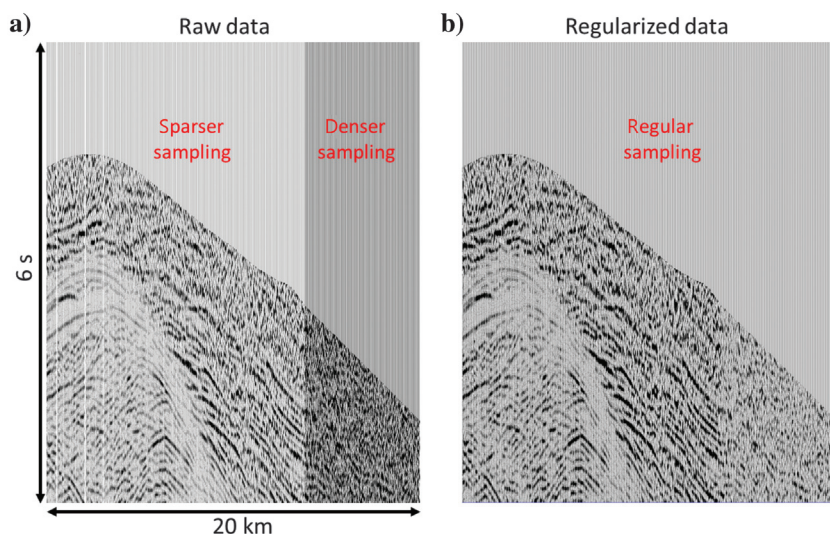


Figure 3. An example of a common-receiver gather from a long-offset marine OBN data set (a) before and (b) after regularization via supergrouping.
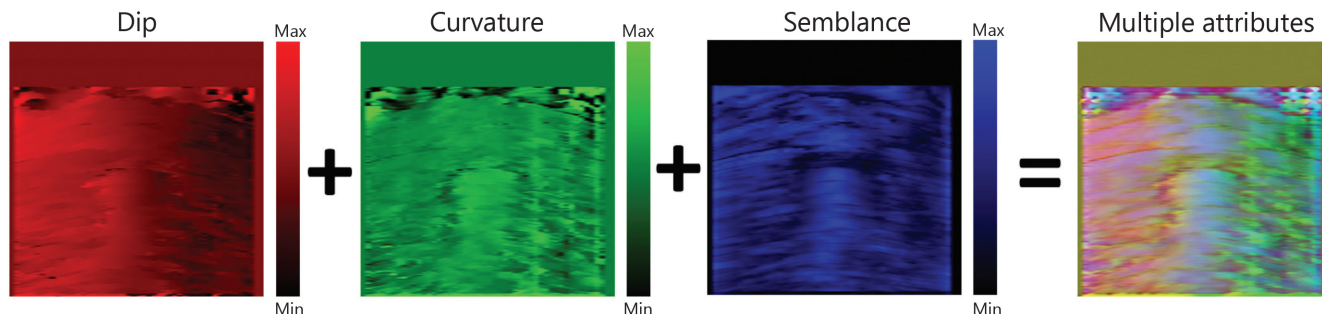


Figure 4. Example of coloring (encoding stage) for two kinematic wavefront attributes and semblance required for NLBF. The multiple attributes are encoded into a single RGB image comprised of three channels: red — dips, green — curvatures, and blue — semblance. The schematically shown min and max values in the dip and curvature colorbars specify the estimation intervals used for search in the dimension-reduced data set. The semblance colorbar is limited by the estimated minimum and maximum semblance values.

ultralong offsets up to 20 km. Complex bathymetry and near-surface velocity variation lead to significant distortions of the recorded wavefield, as shown in Figure 3a. As a result, even early arrivals often are obscured and incoherent. Enhancement with NLBF is required to precondition these data for first-break picking and FWI, as discussed by Kim et al. (2019, 2020).

The data are acquired with a 50 m × 50 m shot spacing and 100 m × 300 m receiver (OBNs) spacing in inline and crossline directions. The survey area is 20 km × 31 km, containing 95 receiver lines or 11,027 common-receiver gathers (Figure 5). The resulting full prestack volume size is approximately 15 TB. The LWA DNN method is applied to this data set to estimate local dips and curvatures for NLBF enhancement with a focus on early arrival events. First, a conventional estimation of local wavefronts attributes is performed based on brute-force semblance optimization as a performance reference test. The estimation and summation apertures (defined as half-offsets from the estimation grid point) equal 500 m and 250 m, respectively. The Shaheen-II HPC cluster from KAUST (King Abdullah University of Science and Technology, 2023a) is used with a single CPU node having a dual-socket 16 core Intel Haswell processor. The running time of the conventional algorithm is 2500 s per one seismic gather on a single CPU node. Hence, the attributes estimation on the entire data set requires at least 25 h using 300 Shaheen-II CPU nodes, making it challenging for a regular production processing flow.

For DL-based workflow, we also start with the standard attributes estimation based on the brute-force semblance-based optimization technique but only using 1% of all common-receiver gathers (Figure 5). Then, estimated dips, curvatures, and semblance are converted to colored RGB images. Trace density in this data set often changes within one gather, creating additional challenges for implementing LWA DNN. By applying the supergrouping-based regularization, the data are obtained on a regular grid with slightly enhanced S/N (Figure 3b). Next, the LWA DNN training and verification are performed. For this OBN data set, approximately 50,000 training samples (pairs of the seismic data and the corresponding wavefront attributes on a 512 × 512 grid) are obtained. As a rule of thumb, the data set is randomly split in a proportion of 80/20. Here, 80% of the data are used directly to update the LWA DNN weights during the training. In contrast, the remaining 20% are used to control the generalization error (validation loss). As the loss function on the validation set starts to grow, we halt the training and work with the DNN weights obtained after the 20th epoch (Figure 6). The LWA DNN training is performed on the Ibex cluster also from KAUST (King Abdullah University of Science and Technology, 2023b). The learning process on a single GPU node with four Nvidia Tesla P100 takes approximately 1900 s.

An example of the predicted-encoded attributes is shown in Figure 7. A visual comparison of the DNN prediction and the semblance-based result, considered as ground truth, shows a reasonable visual match between the two in terms of the RGB images. Then, attributes are decoded into physical values allowing us to make a more quantitative comparison (Figure 8). Although the results of the two methods are not identical, there is a good correspondence between them, especially in terms of dips and semblance values.
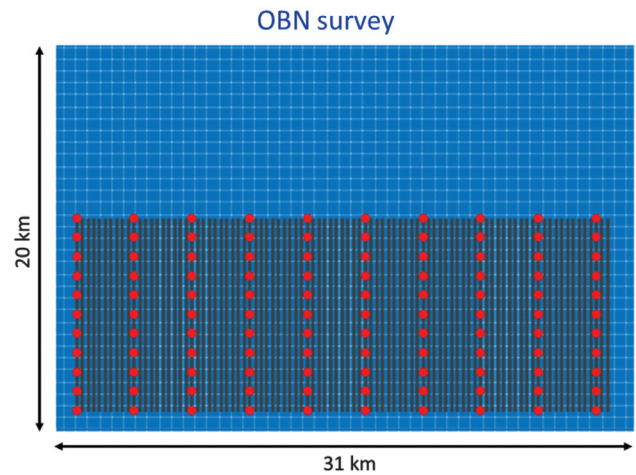


Figure 5. The marine 3D OBN data set acquisition geometry. Positions of receivers are shown in dark gray, whereas shots are in blue. The training data set is generated using common-receiver gathers with the receivers' positions indicated by the thick red points.
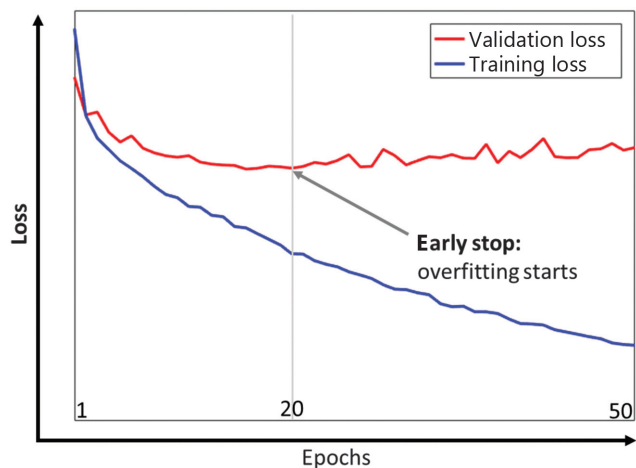


Figure 6. The training loss (red) and the validation loss (blue) for the 3D marine OBN example are plotted versus epoch number. The arrow indicates the epoch when the early stop regularization is applied (generalization error starts to grow).
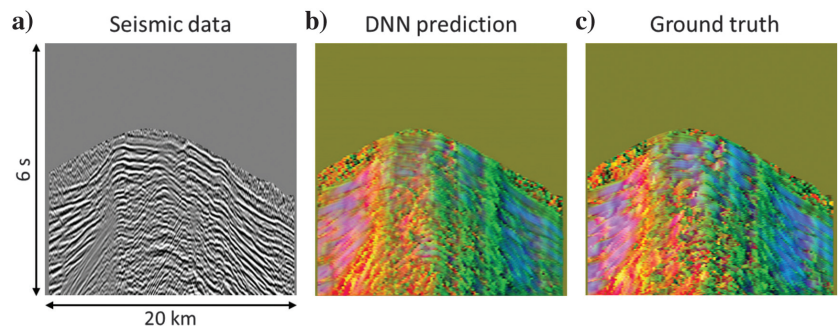


Figure 7. The LWA DNN prediction results for the 3D marine OBN data set: (a) the seismic data, (b) the predicted-encoded LWAs, and (c) the ground truth computed by conventional semblance-based optimization method.

The curvatures are the least reliable estimation parameters, which are typical for a standard semblance-based optimization. Despite some differences, we conclude that the DNN prediction results reasonably match the standard semblance-based method.

Estimating LWAs is an intermediate step in the signal enhancement procedure. To evaluate the influence of the predicted attributes on the beamforming result itself, the original raw data (Figure 9a)

are compared with the data after enhancement using conventional (Figure 9b) and DNN-based NLBF (Figure 9c). Because, in this case, data enhancement is focused on the early arrivals for FWI application, the ranges (minimum and maximum) of possible dips and curvatures are chosen correspondingly. Consequently, some dipping events visible at later arrivals are partially suppressed. Although the internal regularization step is used in the DNN-based estimation workflow, the raw input and the enhanced output data have the same irregular geometry. This is confirmed by the geometry gap visible as a vertical stripe in Figure 9a–9c. The difference between the two versions of the enhanced data obtained by a simple subtraction is reasonably small for practical purposes (Figure 9d). To quantitatively assess the enhancement results, we calculate the normalized root mean square (NRMS) (see Appendix A) between two sets of data obtained with the standard and DL-based NLBF implementations (see Figure 9e and 9f). The NRMS is a sensitive repeatability metric of phase and amplitude similarity between two traces extensively used in 4D seismic (Kragh and Christie, 2002). The NRMS plot is calculated in a trace-by-trace manner using a sliding window of 200 ms. In the areas where the coherent signal dominates (i.e., high S/N), almost perfect repeatability (low NRMS) is achieved. In contrast, in the areas where the noise prevails (low S/N), lower repeatability (high NRMS) is observed. The NRMS distribution histogram for the entire gather (Figure 9f) shows the mean NRMS value of approximately 31%, which often is acceptable even for different vintages of 4D data.
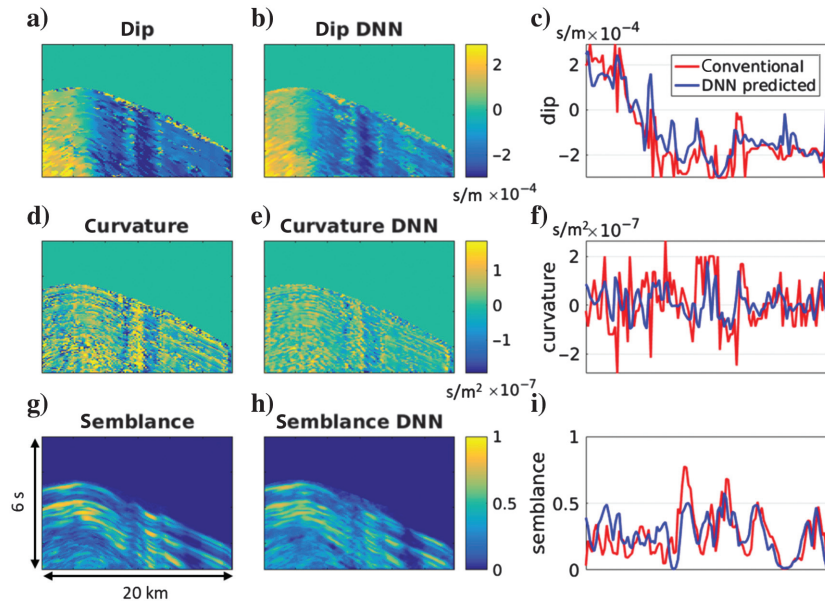


Figure 8. Example of attributes prediction by LWA DNN for the 3D marine OBN data set. (a) The conventionally estimated dips, (d) curvatures, and (g) semblance and (b) the DNN-predicted dips, (e) curvatures, and (h) semblance. (c, f, and i) The horizontal profiles (at $t = 2.5$ s) comparing conventionally estimated (red) and DNN-predicted (blue) attributes: (c) dips, (f) curvatures, and (i) semblance. The conventionally estimated and predicted attributes are plotted using the identical colorbar.

The DNN-based prediction time for one common-receiver gather takes approximately 3 s in contrast to the 2500 s required for the standard semblance-based estimation approach. If training and prediction time are counted together, the proposed method speeds up the conventional wavefront attributes estimation scheme by two orders of magnitude (Table 1). A regular NLBF estimation time (second column in Table 1) and LWA DNN training data set generation time (third column) are evaluated for a single Shaheen-II CPU node. For example, suppose one uses 300 CPU nodes and spends approximately 25.5 h to perform a regular NLBF estimation. In this case, to normalize the output per unit of computing resources, we report elapsed time of 7650 h (25.5 h multiplied by 300 nodes) to quote the time required for a single node to complete the workflow. In contrast, the LWA DNN training time (fourth column) and prediction time (fifth column) are calculated using a single Ibex GPU node, so no normalization is required. The DL-based NLBF estimation time (sixth column in Table 1) is a sum of LWA DNN training data set generation time, training time, and prediction time. The last column in Table 1 (speed-up factor) is a ratio
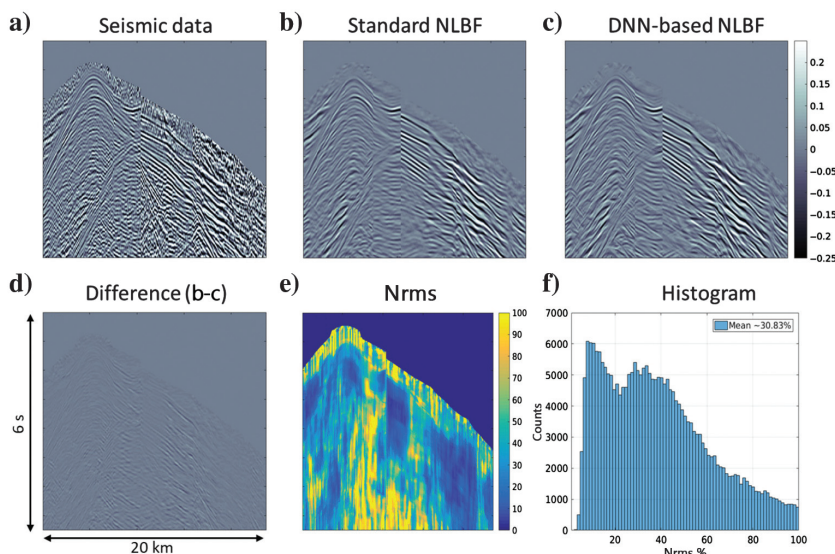


Figure 9. Comparisons of the marine OBN data enhanced with standard and DNN-based NLBF: (a) input data, (b) standard NLBF, (c) DNN-based NLBF, (d) the difference obtained by a simple subtraction between two versions of NLBF data shown with the same scale, (e) the corresponding NRMS, and (f) its histogram.

between a regular NLBF estimation time and DL-based NLBF estimation time. We emphasize that 99% of the data were processed using only a single GPU node. At the same time, to process 15 TB of the 3D prestack data in a reasonable time, one has to use hundreds of CPU nodes. These results demonstrate the ability of LWA DNN to predict good-quality wavefront attributes, bypassing the expensive standard semblance-based estimation scheme.

# LAND DATA EXAMPLES

## Synthetic land data from SEAM Arid data set

Before moving to a more challenging real land data example, the proposed DNN-based workflow is applied to realistically complex synthetic elastic data from the SEAM Arid model data set (Oristaglio, 2012). The model represents complex near-surface conditions typical to a desert environment that significantly deteriorate seismic wavefield and may lead to extreme challenges during data processing (Regone et al., 2017; Bakulin and Silvestrov, 2021c). The near surface consists of alternating high- and low-velocity layers creating complex wavetrains of near-surface arrivals overlaying the weak reflections. In addition, the model contains shallow and deep karst fields creating volumetric scattering of all arrivals. The model is 10 km × 10 km in horizontal extent and 3.75 km in depth (Figure 10). A typical field 3D orthogonal acquisition geometry is extracted from the original dense SEAM Arid data set with inline and crossline spacing for receivers of 25 m and 150 m, respectively. Likewise, source inline and crossline spacing intervals are 50 m and 100 m. Nine-geophone arrays are further stimulated to mimic typical onshore acquisition in a desert environment. The data have been passed through a conventional data processing flow, including linear and random noise attenuation. This experiment aims to assess the feasibility of LWA DNN to predict the wavefront attributes without consequent data enhancement. As in the previous example, we use the estimation aperture of 500 m and the summation aperture of 250 m.

Estimation of the attributes is performed in the cross-spread domain benefiting from the densest sampling available in the acquisition geometry at hand. Centers of cross-spread gathers (intersec-

tions of receivers lines and orthogonal to them sources lines) are depicted as blue points in Figure 10. As in the previous example, 1% of the data are used for LWA DNN training (red points in Figure 10). The selected cross-spread centers are extracted from an evenly spaced regular subgrid to achieve the diversity and representativeness of the training data set. Similar to the previous example, we randomly split the training data set and use 80% of the images for training and 20% to manage the generalization error.

After the training stage, the LWA DNN is applied to a testing data set not previously seen by the neural network. Figure 11a shows an example of input gather from this testing set and the LWA DNN output colored image with encoded attributes. For comparison, a standard semblance-based optimization approach is applied to obtain ground truth results and encode these attributes using the same coloring scheme. Figure 11b shows a good correspondence between RGB images from the LWA DNN output and the standard approach. Then, colored images are decoded back to compare the individual attributes (Figure 12). We can see a reasonable match between the dip and curvature attributes obtained with the traditional and the proposed LWA DNN approaches. For example, Figure 12i compares values extracted for all offsets at approximately 1.2 s with good agreement, especially in the middle section. Although in the conventional algorithm, the semblance is computed as an objective function (using 3D ensembles and not 2D slices; Bakulin et al., 2020), in the proposed method, the semblance values are predicted
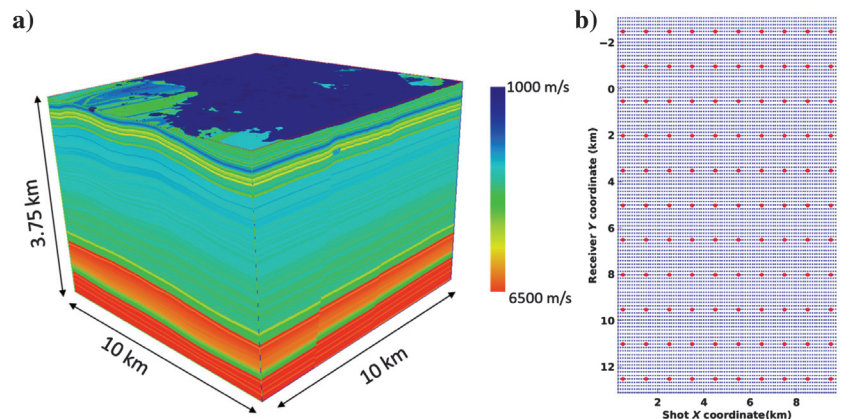


Figure 10. (a) The SEAM Arid velocity model and (b) the corresponding acquisition's cross-spread centers are plotted as blue points in the shot X/receiver Y domain. Red points indicate the positions of the selected cross spreads used for LWA DNN training.

Table 1. Compute time for different DL-based NLBF steps compared with the regular NLBF estimation time derived from various numerical experiments.

| Numerical experiment | Regular NLBF estimation time (h) | LWA DNN training data set generation time (h) | LWA DNN training time (h) | LWA DNN prediction time (h) | DL-based NLBF estimation time (h) | DL-based NLBF speed-up factor |
|---|---|---|---|---|---|---|
| Marine OBN data set | 7650 | 77 | 0.52 | 9.2 | 86.72 | 88 |
| Synthetic land SEAM data set | 1323 | 14 | 0.24 | 8.7 | 22.94 | 57 |
| Real land single-sensor experiment 1 | 76 | 0 | 0 | 0.37 | 0.37 | 205 |
| Real land single-sensor experiment 2 | 76 | 0.78 | 0.11 | 0.37 | 1.26 | 60 |
| Real land single-sensor experiment 3 | 76 | 0.78 | 0.3 | 0.37 | 1.45 | 52 |

To normalize the comparison, standard NLBF estimation time and training data set generation time were estimated as if a single CPU node on the Shaheen-II HPC cluster is used. Likewise, the LWA DNN training and prediction time was estimated on a single GPU node of the Ibex cluster.

by the LWA DNN along the 2D slices without actual computation from the data. To more fairly compare the quality of the estimated attributes, we directly recalculate semblance values using the estimated $A$ and $D$ parameters by these methods using the same actual 2D slice of the data extracted from the $X$-$T$ plane (Figure 12j and 12k). This more direct like-for-like comparison reveals that semblance values from the standard method are higher and more

spatially consistent. In other words, the conventional method finds slightly more coherent events than the LWA DNN. Nevertheless, the event kinematics are still similar, and the differences are unimportant for the data enhancement application.

In terms of performance, the LWA DNN training takes approximately 850 s. For this numerical experiment, combining the training data set generation time with the LWA DNN training and prediction time, we arrive at 23 h to perform DL-based NLBF estimation. In comparison, the regular NBLF estimation time is approximately 1323 h (see Table 1). The obtained speed-up factor of approximately 57 is significant for practical applications.

### Field single-sensor data from the desert environment

In the final example, the proposed LWA DNN approach is applied to broadband single-sensor single-source land 3D data from the desert environment (Pecholcs et al., 2012; Cordery, 2020). Figure 13 shows an example of a common-receiver gather after passing through a comprehensive data processing flow. Despite the significant effort in processing, reflected waves remain largely invisible. Furthermore, such single-sensor data from the desert environment are known to have the lowest S/N and are very challenging to process (Bakulin et al., 2020; Cordery, 2020). Therefore, signal enhancement algorithms, such as NLBF, are essential for such data sets.

In acquisition, orthogonal source and receiver lines 75 m and 150 m apart are used, each with 12.5 m inline station spacing. As a result, cross-spread gathers with a high density 12.5 m × 12.5 m grid require approximately 1 h per gather for the standard attributes estimation based on a brute-force semblance optimization. Similar to the previous examples, we use the estimation aperture of 500 m and the summation aperture of 250 m. Therefore, significant computational resources are required to process the entire data set with many gathers. As in the previous examples, the DNN-based workflow provides computational savings for such high-density 3D data. For this example of the LWA DNN workflow, a single-source line consisting of 460 cross-spread gathers is used.

Similar to the previous examples, we create a sparse training data set (approximately 1% of the total data) to calculate the weights of the DNN. However, in this case, three different experiments are considered to develop the optimal DL-based workflow. The first two are applications of the transfer learning approach with and without additional limited training using the initial LWA DNN weights acquired in the synthetic SEAM Arid model experiment. The third experiment is LWA DNN training from scratch using only field data.
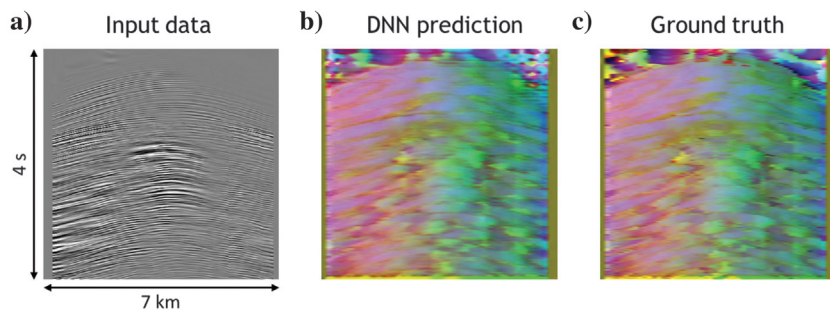


Figure 11. The LWA DNN prediction results on SEAM Arid data set: (a) the seismic data, (b) the predicted-encoded attributes, and (c) the ground truth computed by conventional semblance-based optimization method.
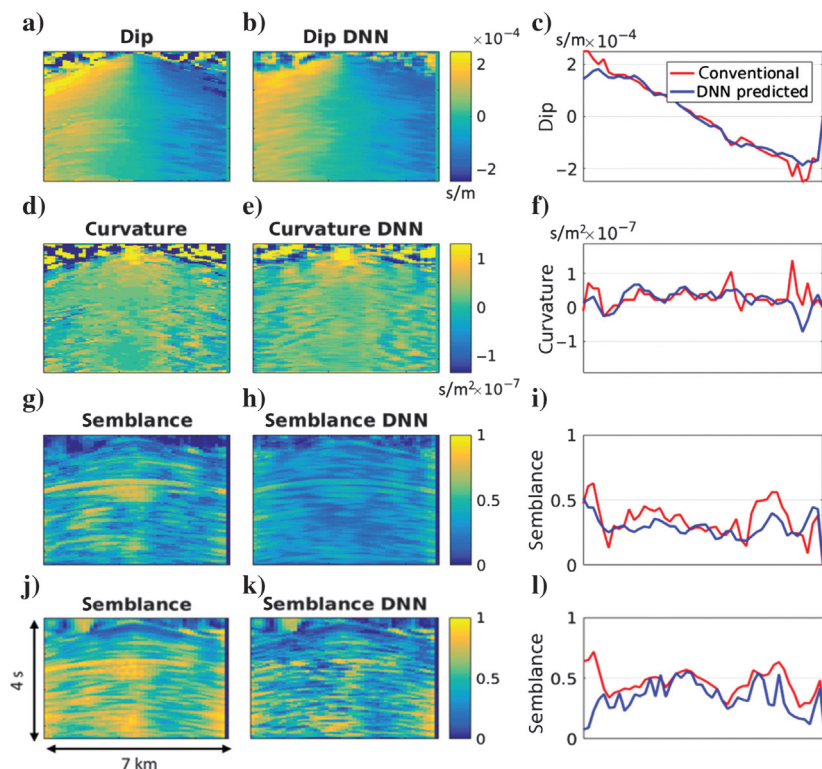


Figure 12. Example of attributes prediction by LWA DNN for the 3D marine OBN data set. (a) The conventionally estimated dips, (d) curvatures, and (g) semblance and (b) the DNN-predicted dips, (e) curvatures, and (h) semblance. (c, f, i, and l) The horizontal profile of conventional and DNN-predicted values at $t = 1.2$ s. (c) Dips, (f) curvatures, and (i) semblance with DNN-predicted attributes in blue and conventionally estimated in red. Please note that (g) semblance is computed using an entire set of optimal parameters (dips and curvatures) and using 3D ensembles of the data. In contrast, panel (h) is predicted by DNN using a 2D slice of the data (i.e., 2D ensembles). For a more direct comparison, (j–l) show semblance panels recalculated using the exact same $X$-$T$ slice of 2D data and substituting actual $A$ and $D$ parameters estimated by each of the methods. The result for the conventional method is shown in (j) and for the DNN-based method in (k).

*Transfer learning with no additional training (experiment 1)*

First, we only use LWA DNN weights obtained after training on the synthetic SEAM Arid data. It allows determining to what degree the DNN trained on synthetic data can handle challenging field data that was not seen before. Such a question is reasonably justified for this case because the synthetic model and data replicate, to a large extent, data features typical for an arid environment in which the real data are acquired. In addition, these data sets have similar bandwidths and these are sorted into the cross-spread domain. Next, we apply the pretrained LWA DNN to predict wavefront attributes for the field single-sensor data with these similarities in mind. The predicted dips and curvatures are shown in Figure 14a and 14e. Comparing them with the conventional semblance-based optimization solution (Figure 14d and 14h), one can see that the LWA DNN resolves some general features of the attributes, such as correctly predicting positive dips on the left and negative dips on the right. However, the absolute values and more detailed features remain unresolved. This can be seen from Figure 15a and 15d, in which the dips and curvatures attributes are extracted along the horizontal profile at 1 s traveltime. The magnitudes of the DNN-predicted dips and curvatures are lower than the reference solution, and overall, their behavior is smoother.

*Transfer learning with additional limited training (experiment 2)*

In the next experiment, the same pretrained neural network from the synthetic data is used, but additional limited training is applied using the field single-sensor data to further adapt the DNN weights for this specific case. Only four training epochs are run, taking approximately 400 s. The prediction results obtained with this new neural network are shown in Figures 14b, 14f, 15b, and 15e. We clearly observe much better correspondence to the conventional reference solution. In case, additional training allows DNN LWA to resolve attributes' magnitudes and outline further details not revealed in the previous experiment.

*Training LWA DNN from scratch using only field data (experiment 3)*

In the final experiment, the LWA DNN is trained from scratch using only a small subset (1%) from the field data and randomly initialized DNN weights before the training. The training process takes approximately 1100 s, more than two times the training time spent in the previous experiment with the pretrained network. The corresponding predicted attributes from the testing data set are shown in Figures 14c, 14g, 15c, and 15f. Overall, the results are very close to the ones obtained in experiment 2.

To judge the quality of estimated beamforming trajectories, we recalculate the semblance values using the estimated wavefront attributes from all three experiments. The semblance obtained after the standard NLBF estimation subroutine is used as a reference. The corresponding average semblance coherency measures per training seismic gather are provided in Table 2 for different experiments. As expected, the best average semblance is

for the reference test (0.0150). Out of three DNN versions, the second experiment gives the highest value of 0.0124, slightly lower than the reference. Converting these semblance values to S/N values (see Bakulin et al., 2022), we obtain a very low S/N value of approximately −20 dB, which illustrates the challenges faced during the processing of such data. Finally, the mean absolute percentage error (MAPE) value is calculated between the parameters from the reference test and three LWA DNN training experiments (see Table 3). In this case, the best MAPE value also is obtained in experiment 2
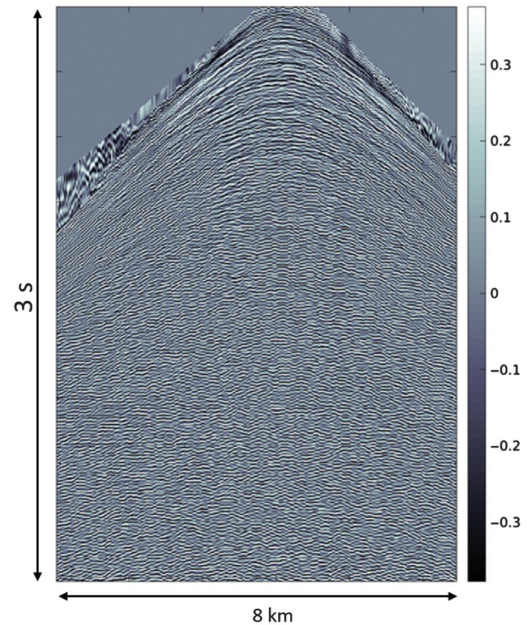


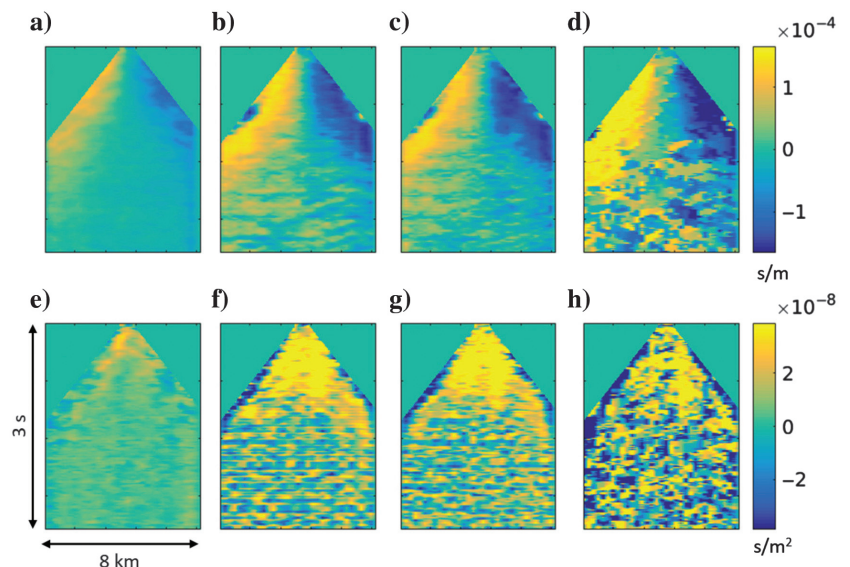Figure 13. The example of cross-spread gather from single-sensor land data (AGC applied).



Figure 14. (a–d) Dips and (e–h) curvatures for single-sensor land data predicted by the LWA DNN using different scenarios: (a and e) transfer learning experiment 1, (b and f) transfer learning experiment 2, (c and g) training from the scratch experiment 3, and (d and h) the reference from conventional estimation used as ground truth.

(31%). The third experiment comes closely behind (35%), whereas the worst MAPE of 46% characterizes the first experiment.

Whereas comparing attributes and semblance values provides insights into the performance of the DL-based estimation approaches, the ultimate goal is to obtain enhanced prestack data. Therefore, let us compare the enhanced data generated by these three experiments (transfer learning, transfer learning with additional short training, and LWA DNN training from scratch) and contrast them to the conventional NLBF (Figure 16). Despite the predicted attributes' differences, the final enhanced prestack seismic data appear very similar, even for transfer learning experiment 1. Finally, the NRMS metric is calculated over the area shown by a red rectangle in Figure 16e. NRMS metric quantifies the sample-by-sample similarity between DL-based and conventionally enhanced data. For example, the average NRMS for experiment 1 is approximately 59% (Figure 17a). Likewise, it is approximately 32% for experiment 2 (Figure 17b) and approximately 37% for experiment 3 (Figure 17c). Although 59% may be somewhat high for 4D processing, such results could still be acceptable for conventional 3D processing. However, it is clear that additional training on the field data gives a significant edge, with the reduction in NRMS to 32% representing a considerable improvement in similarity. NRMS metric suggests that the optimal DL-based scenario is experiment 2, leading to the best semblance value and prestack data closest to what is achieved with conventional NLBF.

The overall DL-based NLBF performance is summarized together with the regular NLBF estimation in Table 1. Although the regular NLBF is performed on hundreds of CPU nodes of the Shaheen-II HPC cluster, in the summary table, the time for a single CPU node is presented to compare it with LWA DNN computation on a single GPU node of the Ibex cluster. The total estimation time within the LWA DNN workflow comprises training data set generation, training, and prediction time. For marine and land data testing data sets, the DNN-based workflow provides a significant speed-up factor compared to a regular NLBF estimation procedure varying from 52 to 205, depending on the estimation scenario.

## DISCUSSION

We have shown that direct estimation of wavefront attributes from 3D prestack data using LWA DNN can significantly speed up data enhancement without unduly sacrificing the quality of the results. Although the DNN-predicted attributes are acceptable for enhancement, their accuracy tends to be lower than those obtained by the standard semblance-optimization approach. Other geophysical applications of the wavefront attributes, such as tomography or diffraction separation, can be more sensitive to the attributes' quality and may require more detailed analysis. Another potential application of the DL-based approach is to reduce the search intervals for the standard semblance-based methods by using the predicted attributes as a preliminary estimated value.

Currently, we use colored RGB images representing 2D slices of the encoded dips, curvatures, and corresponding semblance values as output. Although the semblance value can be directly recalculated once the dips and curvatures are estimated, predicting it with DNN gives us additional confidence in the results and allows us to quality control them. An alternative implementation may exclude semblance from the training/prediction and focus only on dips together with curvatures or dips only. Moreover, it also is possible to avoid converting the attributes to RGB images. In this study, we primarily do it to achieve easy visual correspondence with the previous image-based DNN applications and to simplify the visual quality control during the training. In fact, any properly normalized tensor with an arbitrary number of channels (wavefront attributes, their combinations, or some functions of wavefront attributes) could be used. The same reasoning also could be applied to the input data, meaning that the regularization and the gray-scale-based color encoding steps might be avoided or implemented differently than in the current workflow using another network architecture.

The optimization of LWA DNN architecture that could lead to more accurate attributes could
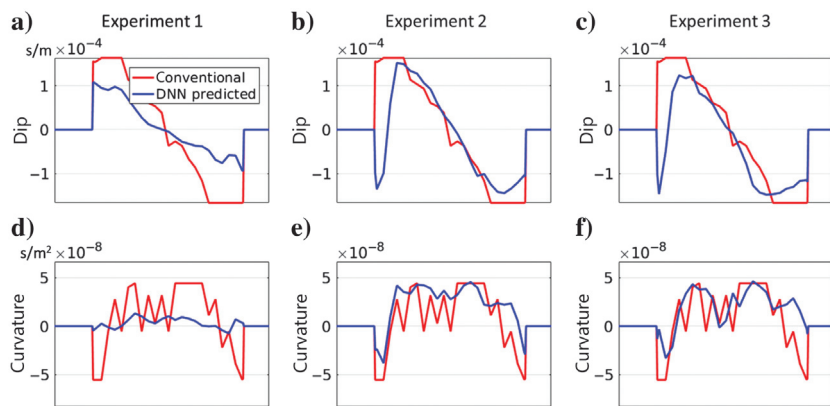


Figure 15. The horizontal profiles of conventional and DNN-predicted values at $t = 1$ s for single-sensor land data. The comparison of the (a–c) predicted dips and (d–f) curvatures for three numerical experiments: transfer learning (experiment 1 — a and d), transfer learning with additional short training (experiment 2 — b and e), and LWA DNN training from scratch (experiment 3 — c and f). Conventional estimation results are shown in red, whereas DNN predictions are in blue.

**Table 2. Average semblance per 3D *X-Y-T* estimation grid over the testing data set for three LWA DNN training scenarios and the reference standard NLBF test.**

| Experiment 1 | Experiment 2 | Experiment 3 | Reference value (standard NLBF) |
|---|---|---|---|
| 0.0090 | 0.0124 | 0.0116 | 0.0150 |

**Table 3. The MAPE value computed for each pair of semblance cubes obtained from the reference case and each of the three training scenarios.**

| Experiment 1 | Experiment 2 | Experiment 3 |
|---|---|---|
| 46% | 31% | 35% |

The reference cube is obtained using a conventional estimation procedure.

be a topic of further research. There are different possible options, including the usage of residual blocks instead of the conventional convolutional blocks, which in many cases, significantly improve the quality of the DNN predictions. Other architectures, different from U-net and ResUnet, also could be evaluated. Particular attention should be paid to creating an optimal training data set. Whereas currently, we use approximately 1% of the total seismic data uniformly distributed over a survey area to create representative samples, more advanced options based on gathers' similarity functions may be used to provide more diversity in training.

Overall, whereas the results presented in this study show the great potential of DNN-based workflows to improve the wavefronts' attribute calculations significantly, many potential ways exist for further improvements. We hope that this work will be a good starting point for other research studies in this direction.
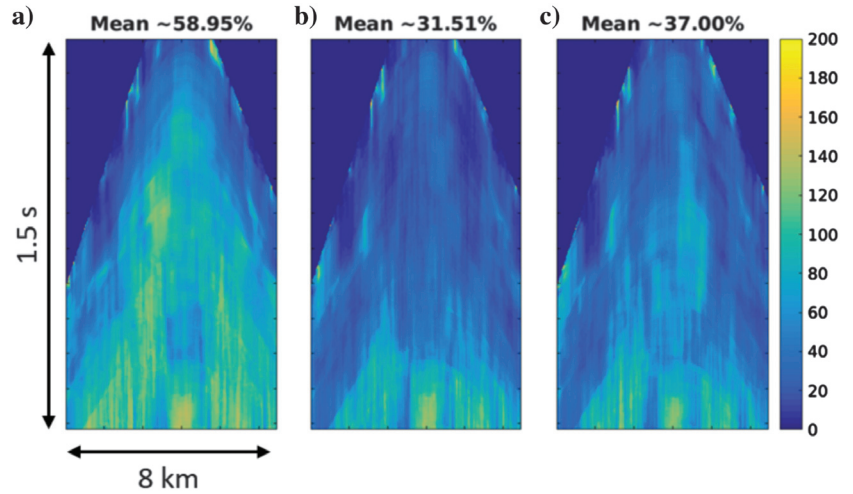


Figure 17. The NRMS plot computed between reference NLBF data set and various DNN-based enhancement scenarios for single-sensor land data: (a) transfer learning experiment 1, (b) transfer learning experiment 2, and (c) the LWA DNN training from scratch experiment 3. The lowest NRMS is observed in (b), suggesting the best similarity with reference data.
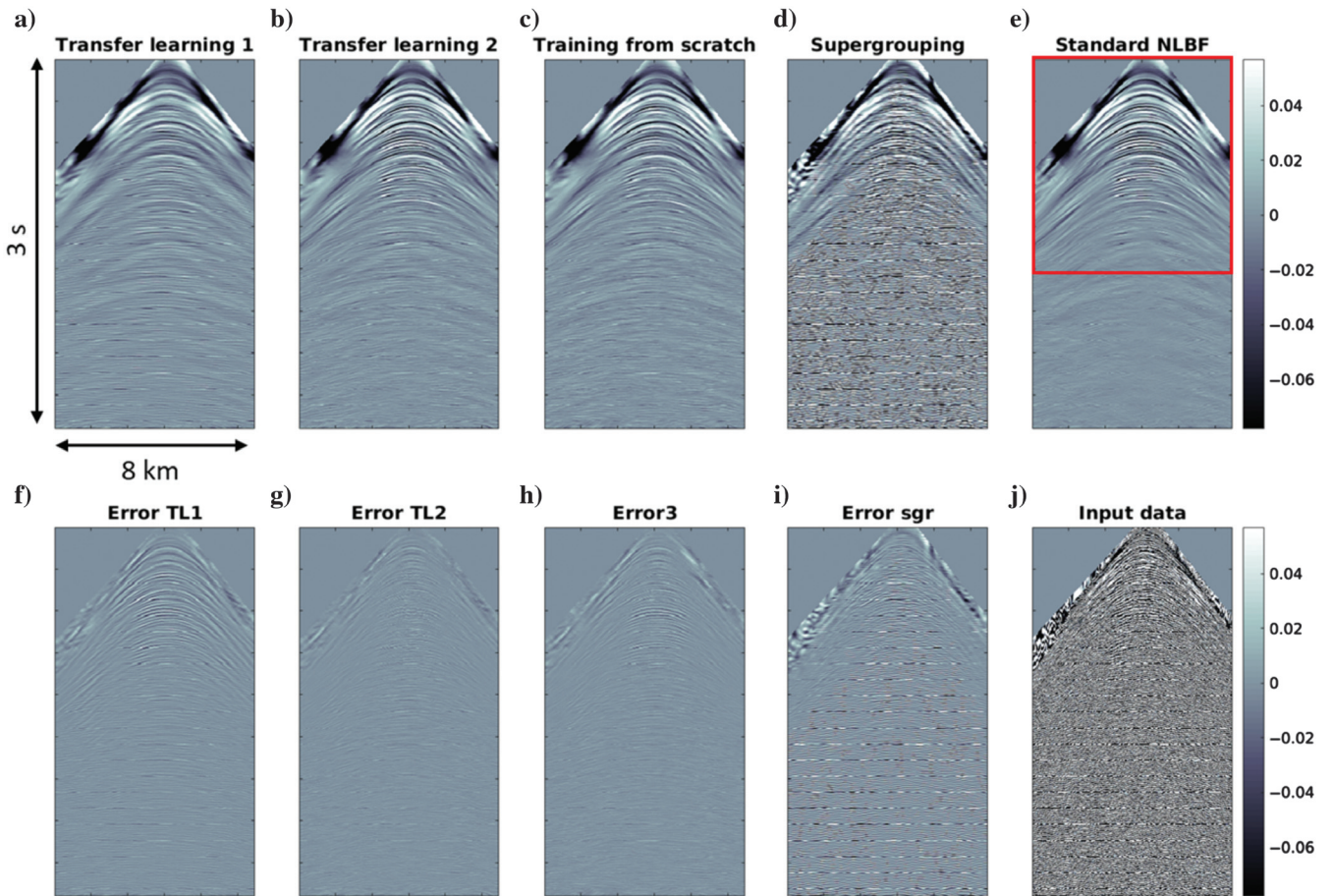


Figure 16. The comparison of the different enhancement scenarios for single-sensor land data: (a) transfer learning experiment 1, (b) transfer learning experiment 2, and (c) LWA DNN training from scratch. For comparison, input data are shown in (j) and supergrouped preconditioned data are shown in (d). The difference between reference test (e) with standard NLBF and various versions of DNN-based enhancement are presented on plots (f–i) for experiment 1, experiment 2, experiment 3, and supergrouping, respectively. All plots are presented using the identical colorbar. The red rectangle indicates the area used for the NRMS comparison (see Figure 17).

## CONCLUSION

A novel DNN-based workflow is presented for the accelerated estimation of LWAs required to enhance massive 3D prestack seismic data sets. It is based on an automatic LWAs estimation using a specially trained convolutional DNN called LWA DNN. During the training, the workflow performs a conventional semblance-based attributes estimation for approximately 1% of the total data. The data regularization step based on supergrouping is incorporated to provide fully populated input images to DNN. The next step is encoding the input data and the output attributes into colored images to derive the inputs analogous to the previous DNN-based applications for visual imagery analysis. The proposed approach is verified on two challenging real land and marine data sets. We demonstrate that significant speed up is achieved using DNN compared to the conventional method — up to the factor of 200 while preserving an acceptable quality of the enhanced data as a final output. This new approach may enable multiscale processing of massive 3D prestack seismic data sets that are currently too computationally demanding.

Although this workflow is adaptive to a particular seismic data set, a set of transfer learning experiments is performed to assess the feasibility of predicting the wavefront attributes using the previously visible input data to the LWA DNN. Numerical tests demonstrate that the best prediction results for challenging 3D land single-sensor data are achieved using the LWA DNN pretrained on a synthetic SEAM Arid data set and supplemented with an additional short training on the field data at hand (i.e., weights adaptation). This observation assumes that the pretrained LWA DNN weights can be used for accelerated local wavefront estimation of analogous data sets, i.e., for data with similar characteristics: sorting, acquisition, S/N, and bandwidth. An initial demonstration of the transfer learning approach is provided using the realistic land synthetic data set as a guide for initial DNN weights with the additional short retraining and subsequent application to the real data. Further investigation of DNN-based workflow is required from the practical point of view to find the best estimation strategy in terms of performance and quality.

## ACKNOWLEDGMENTS

## DATA AND MATERIALS AVAILABILITY

Data associated with this research are confidential and cannot be released.

## APPENDIX A

## DEFINITION OF REPEATABILITY AND FORECASTING ACCURACY METRICS

The NRMS between two traces $a_t$ and $b_t$ at point $t_0$ using a window size $dt$ is the rms of the difference divided by the average rms of the inputs and expressed as a percentage:

$$\text{NRMS}(t_0) = \frac{200 \times \text{rms}(a_t - b_t)}{\text{rms}(a_t) + \text{rms}(b_t)}, \qquad \text{(A-1)}$$

where the rms is defined as

$$\text{rms}(x_t) = \sqrt{\frac{\sum_{t_0 - dt}^{t_0 + dt}(x_t)^2}{N}} \qquad \text{(A-2)}$$

and $N$ is the number of samples in the interval $[t_0 - dt, t_0 + dt]$. The NRMS is usually quoted in percent and ranges from 0% to 200%, with low values representing better similarity or repeatability.

The MAPE is a measure of prediction accuracy of forecasting in statistics, and it is usually expressed as follows:

$$\text{MAPE} = \frac{100\%}{N} \sum_{i=1}^{N} \left| \frac{A_i - F_i}{A_i} \right|, \qquad \text{(A-3)}$$

where $A_i$ is the actual value and $F_i$ is the forecast value.

## REFERENCES

Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, 2016, TensorFlow: Large-scale machine learning on heterogeneous distributed systems: arXiv preprint, doi: 10.48550/arxiv.1603.04467.

Bakulin, A., P. Golikov, M. Dmitriev, D. Neklyudov, P. Leger, and V. Dolgov, 2018a, Application of supergrouping to enhance 3D prestack seismic data from a desert environment: The Leading Edge, **37**, 200–207, doi: 10.1190/tle37030200.1.

Bakulin, A., and I. Silvestrov, 2021c, Understanding acquisition and processing challenges in the desert environment through SEAM Arid and Barrett models: 91st Annual International Meeting, SEG, Expanded Abstracts, 2824–2828, doi: 10.1190/segam2021-3583002.1.

Bakulin, A., I. Silvestrov, and M. Dmitriev, 2019, Adaptive multi-scale processing of challenging 3D seismic data for first-break picking, FWI and imaging: 89th Annual International Meeting, SEG, Expanded Abstracts, 3979–3984, doi: 10.1190/segam2019-3214616.1.

Bakulin, A., I. Silvestrov, M. Dmitriev, D. Neklyudov, M. Protasov, K. Gadylshin, and V. Dolgov, 2020, Nonlinear beamforming for enhancement of 3D prestack land seismic data: Geophysics, **85**, no. 3, V283–V296, doi: 10.1190/geo2019-0341.1.

Bakulin, A., I. Silvestrov, M. Dmitriev, D. Neklyudov, M. Protasov, K. Gadylshin, V. Tcheverda, and V. Dolgov, 2018b, Nonlinear beamforming for enhancing prestack data with challenging near surface or overburden: First Break, **36**, 121–126, doi: 10.3997/1365-2397.n0143.

Bakulin, A., I. Silvestrov, D. Neklyudov, K. Gadylshin, and M. Protasov, 2021a, 3D data-domain reflection tomography for initial velocity model building using challenging 3D seismic data: Journal of Seismic Exploration, **30**, 419–446.

Bakulin, A., I. Silvestrov, and M. Protasov, 2021b, Evaluating strategies for estimation of local kinematic parameters in noisy land data: Quality versus performance trade-offs: Journal of Geophysics and Engineering, **18**, 890–907, doi: 10.1093/jge/gxab060.

Bakulin, A., I. Silvestrov, and M. Protasov, 2022, Research note: Signal-to-noise ratio computation for challenging land single-sensor seismic data: Geophysical Prospecting, **70**, 629–638, doi: 10.1111/1365-2478.13178.

Baykulov, M., and D. Gajewski, 2009, Prestack seismic data enhancement with partial common-reflection-surface (CRS) stack: Geophysics, **74**, no. 3, V49–V58, doi: 10.1190/1.3106182.

Berkovitch, A., I. Belfer, Y. Hassin, and E. Landa, 2009, Diffraction imaging by multifocusing: Geophysics, **74**, no. 6, WCA75–WCA81, doi: 10.1190/1.3198210.

Berkovitch, A., I. Belfer, and E. Landa, 2008, Multifocusing as a method of improving subsurface imaging: The Leading Edge, **27**, 250–256, doi: 10.1190/1.2840374.

Berkovitch, A., K. Deev, and E. Landa, 2011, How non-hyperbolic multi-focusing improves depth imaging: First Break, **27**, 95–103, doi: 10.3997/1365-2397.29.9.53731.

Billette, F., and G. Lambaré, 1998, Velocity macro-model estimation by stereotomography: Geophysical Journal International, **135**, 671–690, doi: 10.1046/j.1365-246X.1998.00632.x.

Buzlukov, V., and E. Landa, 2013, Imaging improvement by prestack signal enhancement: Geophysical Prospecting, **61**, 1150–1158, doi: 10.1111/1365-2478.12047.

Cordery, S., 2020, An effective data processing workflow for broadband single-sensor single-source land seismic data: The Leading Edge, **39**, 401–410, doi: 10.1190/tle39060401.1.

Das, V., A. Pollack, U. Wollner, and T. Mukerji, 2019, Convolutional neural network for seismic impedance inversion: Geophysics, **84**, no. 6, R869–R880, doi: 10.1190/geo2018-0838.1.

Duveneck, E., 2004, Velocity model estimation with data-derived wavefront attributes: Geophysics, **69**, 265–274, doi: 10.1190/1.1649394.

Feng, D. P., K. Fang, and C. P. Shen, 2020, Enhancing streamflow forecast and extracting insights using long-short term memory networks with data integration at continental scales: Water Resources Research, **56**, e2019WR026793, doi: 10.1029/2019WR026793.

Fomel, S., 2002, Applications of plane-wave destruction filters: Geophysics, **67**, 1946–1960, doi: 10.1190/1.1527095.

Fomel, S., 2007, Velocity-independent time-domain seismic imaging using local event slopes: Geophysics, **72**, no. 3, S139–S147, doi: 10.1190/1.2714047.

Fukushima, K., 1969, Visual feature extraction by a multilayered network of analog threshold elements: IEEE Transactions on Systems Science and Cybernetics, **5**, 322–333, doi: 10.1109/TSSC.1969.300225.

Gadylshin, K., I. Silvestrov, and A. Bakulin, 2020, Inpainting of local wavefront attributes using artificial intelligence for enhancement of massive 3-D prestack seismic data: Geophysical Journal International, **223**, 1888–1898, doi: 10.1093/gji/ggaa422.

Gadylshin, K., I. Silvestrov, and A. Bakulin, 2021, Direct estimation of local wavefront attributes using deep learning: 91st Annual International Meeting, SEG, Expanded Abstracts, 1596–1600, doi: 10.1190/segam2021-3583265.1.

Garabito, G., 2018, Global optimization strategies for implementing 3D common-reflection surface stack using the very fast simulated annealing algorithm: Application to real land data: Geophysics, **83**, no. 4, V253–V261, doi: 10.1190/geo2017-0836.1.

Glorot, X., A. Bordes, and Y. Bengio, 2011, Deep sparse rectifier neural networks: Proceedings of the 14th International Conference on Artificial Intelligence and Statistics 15, 315–323.

Hellman, K., 2014, Simultaneous estimation of CRS parameters with multidimensional local slopes: 84th Annual International Meeting, SEG, Expanded Abstracts, 4686–4690, doi: 10.1190/segam2014-0548.1.

Hoecht, G., P. Ricarte, S. Bergler, and E. Landa, 2009, Operator-oriented interpolation: Geophysical Prospecting, **57**, 957–979, doi: 10.1111/j.1365-2478.2009.00789.x.

Huang, W. L., F. Gao, and J. P. Liao, 2021, A deep learning network for estimation of seismic local slopes: Petroleum Science, **18**, 92–105, doi: 10.1007/s12182-020-00530-1.

Kaur, H., S. Fomel, and N. Pham, 2020, Seismic ground-roll noise attenuation using deep learning: Geophysical Prospecting, **68**, 2064–2077, doi: 10.1111/1365-2478.12985.

Kim, Y. S., M. Dmitriev, T. Fei, and Y. Luo, 2020, Velocity model building using FWI and LSRTM: Case study on 3D OBN dataset acquired from offshore Saudi Arabia: 90th Annual International Meeting, SEG, Expanded Abstracts, 815–819, doi: 10.1190/segam2020-3416120.1.

Kim, Y. S., T. Fei, M. Dmitriev, and Y. Luo, 2019, An offshore full waveform inversion with automatic salt flooding: OBN case study: 81st Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: 10.3997/2214-4609.201901232.

King Abdullah University of Science and Technology, 2023a, Shaheen II, accessed 31 March 2023, https://www.hpc.kaust.edu.sa/content/shaheen-ii.

King Abdullah University of Science and Technology, 2023b, Welcome to Ibex, accessed 31 March 2023, https://www.hpc.kaust.edu.sa/ibex.

Kragh, E., and P. Christie, 2002, Seismic repeatability, normalized RMS, and predictability: The Leading Edge, **21**, 640–647, doi: 10.1190/1.1497316.

Lambaré, G., 2008, Stereotomography: Geophysics, **73**, no. 5, VE25–VE34, doi: 10.1190/1.2952039.

Lambaré, G., P. Guillaume, and J. P. Montel, 2014, Recent advances in ray-based tomography: 76th Annual International Conference and Exhibition, EAGE, Extended Abstracts, We G103 01, doi: 10.3997/2214-4609.20141151.

LeCun, Y., Y. Bengio, and G. Hinton, 2015, Deep learning: Nature, **521**, 436–444, doi: 10.1038/nature14539.

Mann, J., R. Jäger, T. Müller, G. Höcht, and P. Hubral, 1999, Common reflection-surface stack: A real data example: Journal of Applied Geophysics, **42**, 301–318, doi: 10.1016/S0926-9851(99)00042-7.

Müller, N. A., and M. Spinner, 2010, Improving prestack migration with CRS techniques — A case study: 72nd Annual International Conference and Exhibition, EAGE, Extended Abstracts, D032, doi: 10.3997/2214-4609.201400734.

Oristaglio, M., 2012, SEAM Phase II — Land seismic challenges: The Leading Edge, **31**, 264–266, doi: 10.1190/1.3694893.

Pecholcs, P., R. Al-Saad, M. Al-Sannaa, J. Quigley, C. Bagaini, A. Zarkhidze, R. May, M. Guellili, S. Sinanaj, and M. Membrouk, 2012, A broadband full azimuth land seismic case study from Saudi Arabia using a 100,000 channel recording system at 6 terabytes per day: Acquisition and processing lessons learned: 82nd Annual International Meeting, SEG, Expanded Abstracts, doi: 10.1190/segam2012-0438.1.

Pham, N., and W. Li, 2022, Physics-constrained deep learning for ground roll attenuation: Geophysics, **87**, no. 1, V15–V27, doi: 10.1190/geo2020-0691.1.

Rad, P. B., B. Schwarz, D. Gajewski, and C. Vanelle, 2018, Common-reflection-surface-based prestack diffraction separation and imaging: Geophysics, **83**, no. 1, S47–S55, doi: 10.1190/geo2016-0445.1.

Regone, C., J. Stefani, P. Wang, C. Gerea, G. Gonzalez, and M. Oristaglio, 2017, Geologic model building in SEAM Phase II — Land seismic challenges: The Leading Edge, **36**, 738–749, doi: 10.1190/tle36090738.1.

Riabinkin, L. A., 1957, Fundamentals of resolving power of controlled directional reception (CDR) of seismic waves (in Russian): Prikladnaya, **16**, 3–36.

Rieber, F., 1936, A new reflection system with controlled directional sensitivity: Geophysics, **1**, 97–106, doi: 10.1190/1.1437082.

Ronneberger, O., P. Fischer, and T. Brox, 2015, U-Net: Convolutional networks for biomedical image segmentation: International Conference on Medical Image Computing and Computer-Assisted Intervention.

Santos, L., J. Schleicher, J. C. Costa, and A. Novais, 2011, Fast estimation of common-reflection-surface parameters using local slopes: Geophysics, **76**, no. 2, U23–U34, doi: 10.1190/1.3553001.

Sun, Y., I. Silvestrov, and A. Bakulin, 2022a, Enhancing 3D land seismic data using nonlinear beamforming based on the efficiency-improved genetic algorithm: IEEE Transactions on Evolutionary Computation, **26**, 1192–1199, doi: 10.1109/TEVC.2022.3149579.

Sun, Y., I. Silvestrov, and A. Bakulin, 2022b, Accelerating the 2+2+1 method for estimating local traveltime operators in nonlinear beamforming using GPU graphics cards: Journal of Geophysics and Engineering, **19**, 389–402, doi: 10.1093/jge/gxac028.

Taner, M. T., and F. Koehler, 1969, Velocity spectra — Digital computer derivation and applications of velocity functions: Geophysics, **34**, 859–881, doi: 10.1190/1.1440058.

Waldeland, A. U., H. Zhao, J. H. Faccipieri, A. H. Schistad Solberg, and L.-J. Gelius, 2018, Fast and robust common-reflection-surface parameter estimation: Geophysics, **83**, no. 1, O1–O13, doi: 10.1190/geo2017-0113.1.

Xie, Y., and D. Gajewski, 2016, Simultaneous estimation of the 3D CRS attributes by an evolutionary-based Nelder Mead algorithm: 86th Annual International Meeting, SEG, Expanded Abstracts, 4326–4330, doi: 10.1190/segam2016-13858160.1.

Xie, Y., and D. Gajewski, 2017, 5-D interpolation with wavefront attributes: Geophysical Journal International, **211**, 897–919, doi: 10.1093/gji/ggx334.

You, N., Y. E. Li, and A. Cheng, 2020, Shale anisotropy model building based on deep neural networks: Journal of Geophysical Research: Solid Earth, **125**, e2019JB019042, doi: 10.1029/2019JB019042.

Yu, S., and J. Ma, 2021, Deep learning for geophysics: Current and future trends: Reviews of Geophysics, **59**, e2021RG000742, doi: 10.1029/2021RG000742.

Yu, S., J. V. Ma, and W. Wang, 2019, Deep learning for denoising: Geophysics, **84**, no. 6, V333–V350, doi: 10.1190/geo2018-0668.1.

Zhang, K., W. Zuo, Y. Chen, D. Meng, and L. Zhang, 2017, Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising: IEEE Transactions on Image Processing, **26**, 3142–3155, doi: 10.1109/TIP.2017.2662206.

Zhang, Y., S. Bergler, and P. Hubral, 2001, Common-reflection-surface (CRS) stack for common-offset: Geophysical Prospecting, **49**, 709–718, doi: 10.1046/j.1365-2478.2001.00292.x.

Zhu, W., S. M. Mousavi, and G. C. Beroza, 2019, Seismic signal denoising and decomposition using deep neural networks: IEEE Transactions on Geoscience and Remote Sensing, **57**, 9476–9488, doi: 10.1109/TGRS.2019.2926772.

Zu, S., J. Cao, and X. Jiang, 2021, Slope estimation by convolutional neural networks: 83rd Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: 10.3997/2214-4609.202010260.

Biographies and photographs of the authors are not available.