Extrapolating low-frequency prestack land data with deep learning

Oleg Ovcharenko*, Vladimir Kazei*, Pavel Plotnitskiy*, Daniel Peter*, Ilya Silvestrov[†], Andrey Bakulin[†], Tariq Alkhalifah*

* King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

[†] EXPEC Advanced Research Center, Saudi Aramco, ebDhahran, Saudi Arabia

SUMMARY

Missing low-frequency content in seismic data is a common challenge for seismic inversion. Long wavelengths are necessary to reveal large structures in the subsurface and to build an acceptable starting point for later iterations of full-waveform inversion (FWI). High-frequency land seismic data are particularly challenging due to the elastic nature of the Earth contrasting with acoustic air at the typically rugged free surface, which makes the use of low frequencies even more vital to the inversion. We propose a supervised deep learning framework for bandwidth extrapolation of prestack elastic data in the time domain. We utilize a Convolutional Neural Network (CNN) with a UNet-inspired architecture to convert portions of band-limited shot gathers from 5-15 Hz to 0-5 Hz band. In the synthetic experiment, we train the network on 192x192 patches of wavefields simulated for different cross-sections of the elastic SEAM Arid model with free-surface. Then, we test the network on unseen shot gathers from the same model to demonstrate the viability of the approach. The results show promise for future field data applications.

INTRODUCTION

The primary goal of seismic exploration is to locate hidden targets in the subsurface. FWI seeks to build a velocity model by iteratively adjusting the model parameters so that the respective synthetic wavefields match the recorded ones (e.g., Tarantola, 2005). As a nonlinear optimization problem solver, FWI depends on a starting model, a stepping strategy for iterations and the seismic data. Conventional FWI works well, but only if provided with either a good starting model or low frequencies in the data. Applications of FWI with extensive customization (Kazei et al., 2016; Ovcharenko et al., 2018a) and parameter tuning (Kalita et al., 2019) extend its domain of applicability, yet expert knowledge and parameter tuning are still typically required.

Data-driven approaches seek to limit the need for expert knowledge by implicitly deriving the concept of wave phenomena from a corpus of relevant seismic data. Therefore, they can potentially deliver good starting models for FWI by replacing tomography (Araya-Polo et al., 2018; Kazei et al., 2019b, 2020; Zwartjes, 2020) or even gradually reproducing the FWI resolution for small models (Lin and Zhang, 2019; Wu and Lin, 2019; Yuan et al., 2020; Araya-Polo et al., 2020; Siahkoobi et al., 2020). Supervised learning approaches for the task of initial velocity model estimation require large datasets of realistic subsurface models. Early attempts of building such deep learning (DL) solutions (Mosser et al., 2018; Kazei et al., 2019a) had limited success. Therefore, the construction of

suitable starting models for high-frequency FWI remains a significant challenge.

If the recorded waveforms had sufficient low-frequency content, there would be less need for good starting models (e.g., Baeten et al., 2013). The acquisition of such data is feasible and beneficial, but remains expensive and requires a specialized low-frequency source (Ten Kroode et al., 2013; Brenders et al., 2018). Li and Demanet (2016) synthesized realistic low-frequency waveforms by phase and amplitude extrapolation of individual body waves in seismograms. Ovcharenko et al. (2017) proposed and developed (Ovcharenko et al., 2018b, 2019a) a DL approach operating on data in the frequency-domain to derive low-frequency representations of full shot gathers. Alternative trace-by-trace approaches for bandwidth extrapolation in the time domain was pursued by Sun and Demanet (2018, 2019). Hu et al. (2019a) proposed and developed (Hu et al., 2019b) the concept of progressive learning to extend the seismic data bandwidth and showed initial field data results of such a technique. While allowing to adjust a training set for particular data, the progressive learning approach yet requires retraining the network and multiple FWI runs for each new application. There are two main advancements, that we present here.

First, existing approaches for low-frequency extrapolation operate on acoustic data, which is suitable for marine acquisition setups. Land data are more challenging for FWI due to the strong noise contamination and the presence of surface waves caused by the free surface (Bakulin et al., 2018). In this study we explore the feasibility of bandwidth extrapolation for elastic land data.

Second, previous approaches deal with significantly subsampled nonoverlapping sets of individual shot gathers (either individual traces or sparse frequency subsets). Here, we design a preprocessing and training pipeline for a deep neural network to convert highly overlapping patches of band-limited prestack shot gathers into their low-frequency pairs, which should improve the robustness of the procedure.

EXPERIMENTAL SETUP

The 3D SEAM Arid model (Oristaglio, 2015) offers a suitable framework for experiments on land data. It exhibits complex near-surface geology with relatively simple targets in the lower part of the model. We extract 80 parallel vertical sections and use them to generate a 2D seismic dataset. We numerically simulate synthetic wavefields in each of these elastic models with free surface and then use these data for training of the DL model.

We consider a 4 km deep and 10 km long cross-sections sampled on $[1.2 \times 3.2] \times 10^3$ grid points with step of 3.125 m (Figure 1). Sources and receivers are evenly distributed over the surface with 100 m and 12.5 m spacing, respectively. Sparse distribution of sources helps to diversify the dataset of recorded waveforms.

Figure 1: A cross-section from SEAM Arid elastic model. Red and green strokes indicate locations of sources and receivers, respectively.

To generate the synthetic data we use the finite-difference modeling in time domain (Köhn et al., 2012) with peak source frequency of 15 Hz. The recording time is 5 sec with time sampling of 0.25 ms. We then low-pass the data using a 15 Hz filter and downsample traces to 8 ms.

DEEP LEARNING FRAMEWORK

We approach the problem of bandwidth extrapolation as image to image mapping. In particular, we train a deep learning model to operate on portions of time-domain seismic data. Training on portions of images allows inference for arbitrary size of inputs. Therefore, the inference might be performed in a sliding window manner as long as the time-step, frequency content and spacing between receivers are the same.

Dataset We apply high- and low-pass filters to each of the generated shot gathers to create high- and low-frequency versions of the data gathers, respectively. Input data gathers have frequency content from 5 to 15 Hz, whereas respective labels contain low-frequency data in the 0-5 Hz range. A single data pair is then created by randomly cropping patches of 192x192 from the filtered data (Figure 2). The entire dataset is split into partitions of 80/10/10 for training, validation and testing, respectively.

Figure 2: Sample patches, 192 x 192 pixels, of input and target data from training dataset. The network converts monochromatic patches of band-limited data (5-15 Hz, left) into respective low-frequency targets (filtered by a low-pass filter with 5 Hz corner frequency, right).

Pre-processing Manipulations on input and target data prior to training define the success of a DL application. The goal of pre-processing is to equalize the impact of features in the data on training. Amplitudes of seismic data naturally decrease at longer offsets and times due to attenuation and geometrical spreading. To compensate for these phenomena Kazei et al. (2019b) used standard scaling of individual gathers of the data. We use automatic gain control (AGC) from the Madagascar open-source package (Fomel et al., 2013) to compensate for the natural amplitude decay. AGC evaluates root-mean-squared amplitudes in sliding windows of 25×25 equal to the same constant value. This value is irrelevant in our workflow so we leave the default settings of the Madagascar implementation. At the pre-processing stage, the real amplitude of the waveform after the AGC correction joins the constant scaling coefficient for each particular patch and thus can be selected arbitrarily. After that we extract patches from random locations and scale them to fall into a range of $[-1, 1]$ by dividing every patch by its maximum absolute value. The saved coefficient is applied then to the patch of target data. All pre-processing operations are reversible given the respective scaling coefficients.

Architecture UNet (Ronneberger et al., 2015) is an architecture of common choice for image-to-image mapping, it consists of encoder and decoder parts, interconnected by skip-connections. Skip-connections copy outputs from encoder layers and concatenate them to respective inputs of decoder layers. This internal data transfer leads to enhanced resolutions of the resulting output data and regularizes the training process (Ronneberger et al., 2015). We set the numbers of convolutional 3x3 kernels to 32, 64, 128, 256 and 512 in the encoder layers of the network. The decoder shares the same set of layers as the encoder, but in reversed order.

Metric We use the L1 norm as pixel-wise loss for training the image-to-image model. In the example, we quantify the mismatch between true and predicted low-frequency data as a normalized root-mean-squared score (NRMS, Figure 3) often used in 4D seismic (Kragh and Christie, 2002). The perfect match then corresponds to zero value. NRMS is a rather strict metric sensitive to small time shifts and amplitude variations. Values of 10-30% are often acceptable for 4D seismic allowing subtraction and correlation of the data.

Figure 3: True (left) and predicted low-frequency (middle) data along with their difference (right). NRMS value (title of the right) of 16% suggests excellent sample-by-sample agreement between predicted and true data.

EXAMPLE

We generate training data by simulating 51 sources in 120 2D sections extracted from the SEAM Arid model. This results in 6120 synthetic shot gathers with dimensions of 800×625 (receivers $\times$ time steps) used for training. We run training for 400 epochs with batch size of 16 and learning rate of 10^{-4} using Adam optimization algorithm (Kingma and Ba, 2014). In the example later, we clip the offset for the sample shot to position the source in the middle.

The presence of surface waves is a distinctive feature of land data. These waves dominate the recordings as seen in the Figure 4. The propagation velocity of surface waves is lower than that of the body waves resulting in natural separation between various arrivals at farther offsets. At the near-offset portion of the data, which we consider here (up to 2.5 km), surface waves are visible only at times less than one second. Significant contribution to the error comes from edge effects due to the convolution operation. We account for this source of error by clipping 10 pixels from each side of the image.

The NRMS score describing the cumulative mismatch between normalized predicted data and target data is 21%. The score is dominated by mismatch in surface wave reconstruction. The score decreases when removing the AGC correction (Figure 5). Data after removing AGC suggest that the reconstruction error is largely due to the amplitude mismatch for surface waves reconstruction.

Amplitude spectrum of the target and reconstructed low-frequency data are shown in Figure 6. The corner frequency of 5 Hz used in low-pass filter leaves a room for signal leakage into higher frequencies. This leakage leads to smooth waveforms in the time domain. Otherwise the ripples would occur due to a sharp cut in the frequency spectrum. Amplitude spectrum shows a fair fit at target frequencies of 0-5 Hz while out-of-band higher frequencies are fitted even better. We also compute trace-by-trace cross-correlations to measure time shifts between target and extrapolated data (Figure 7).

Furthermore, we measure the mean relative reconstruction error as a function of recording time. To be representative, it is averaged across all traces in the shot gather. For data after AGC, the largest deviation occurs at later times (Figure 8(a)). Supposedly, this is due to the relative homogeneity of data samples at later recording times. Meaning that the wavefield

Figure 4: Normalized input high-frequency data 5-15 Hz, target 0-5 Hz and predicted low-frequency data with NRMS of respective residual.

Figure 5: Ground truth and extrapolated low-frequency data after de-applying AGC along with their difference (NRMS of 20% dominated by residual from surface waves).

Figure 6: Amplitude spectrum for input high-frequency data (5-15 Hz, downsampled), target and extrapolated low-frequency data after low-pass filtering with corner frequency of 5 Hz.

Figure 7: Trace-by-trace time shifts between target and extrapolated data, measured from their cross-correlation. Extrapolation does not introduce any bias due to near-zero mean (black line) whereas standard deviation (purple) remains relatively low: typical 30 ms standard deviation corresponds to a shift equivalent to 1/7 of the dominant period (200 ms) at the maximum frequency of the 0-5 Hz band.

patterns covered by patches in the bottom of the image appear to be similar to respective patches in other samples. To address this issue, the training dataset might be enriched by transferring elastic parameters from a reference model of subsurface onto a number of synthetic priors (Ovcharenko et al., 2019b). The error distribution migrates toward the early times after removing the AGC (Figure 8(b)). This suggests that waveforms with higher amplitude (surface waves) dominate the error panel.

Figure 8: Mismatch between target and extrapolated low-frequency data computed for every time step by averaging all offsets in a shot gather shown in Figure 4: (a) data extrapolated after applying AGC; (b) extrapolated data but after de-applying AGC. Mean (black line) and confidence range (purple) defined by standard deviation.

The trained UNet network might be used for inference in a sliding-window manner on a seismic data of arbitrary size (Figure 9). In the current example, the training dataset is biased toward image segments dominated by the shorter offsets (sources cover only the central part of the model), which results in higher errors at longer offsets. This could be mitigated by better design of the training set and including sufficient number of representative segments from all ranges of offset-time panels. Also, the reverse AGC operation suffers from introducing the high-amplitude artifacts in areas with zero or weak signal (Figure 9, top). We apply median filter with the window size of $125 \text{ m} \times 80 \text{ ms}$ and mute portions of de-normalized traces prior to the first-break arrival to compensate for these artifacts.

Figure 9: Same as Figure 3 but for full-offset acquisition in a synthetic setup. Residual between target and predicted low-frequency data with respective NRMS (right).

CONCLUSIONS

We explore the scenario of frequency bandwidth extrapolation for elastic land seismic data utilizing deep learning. We show that UNet is a suitable architecture for mapping time-domain image segments of prestack elastic seismic data with medium frequency band (5-15 Hz) into the respective portions with low-frequency band (0-5 Hz). We demonstrate reasonable closeness of the reconstructed data to a ground truth using a strict metric of NRMS. Reconstructed low-frequency data do not seem to have any bias in terms of phase as judged from cross-correlation time-shifts with the ground truth target data. Extrapolated data can be useful for enabling full-waveform inversion in cases when data acquisition fails to deliver reliable low frequencies. Alternatively, low-frequency data can be used as a guide to perform signal-to-noise enhancements often required at low frequencies due to insufficient source strength.

ACKNOWLEDGMENTS

We thank Anatoly Baumstein from ExxonMobil for interesting discussion on low-frequency extrapolation at the EAGE Annual Meeting. The research reported in this publication was supported by funding from Saudi Aramco and King Abdullah University of Science and Technology (KAUST).

REFERENCES

- Araya-Polo, M., A. Adler, S. Farris, and J. Jennings, 2020, Fast and accurate seismic tomography via deep learning, *in* *Deep Learning: Algorithms and Applications*: Springer, 129–156.
- Araya-Polo, M., J. Jennings, A. Adler, and T. Dahlke, 2018, Deep-learning tomography: *The Leading Edge*, **37**, 58–66.
- Baeten, G., J. W. de Maag, R.-E. Plessix, R. Klaassen, T. Qureshi, M. Kleemeyer, F. t. Kroode, and Z. Rujie, 2013, The use of low frequencies in a full-waveform inversion and impedance inversion land seismic case study: *Geophysical Prospecting*, **61**, 701–711.
- Bakulin, A., M. Dmitriev, I. Silvestrov, et al., 2018, Enhancement of challenging prestack land data for improved processing and imaging: Presented at the SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition, Society of Petroleum Engineers.
- Brenders, A., J. Dellinger, C. Kanu, Q. Li, and S. Michell, 2018, The Wolfspär® field trial: Results from a low-frequency seismic survey designed for FWI, *in* *SEG Technical Program Expanded Abstracts 2018*: Society of Exploration Geophysicists, 1083–1087.
- Fomel, S., P. Sava, I. Vlad, Y. Liu, and V. Bashkardin, 2013, Madagascar: Open-source software project for multidimensional data analysis and reproducible computational experiments: *Journal of Open Research Software*, **1**.
- Hu, W., Y. Jin, X. Wu, and J. Chen, 2019a, A progressive deep transfer learning approach to cycle-skipping mitigation in fwi, *in* *SEG Technical Program Expanded Abstracts 2019*: Society of Exploration Geophysicists, 2348–2352.
- , 2019b, Progressive transfer learning for low frequency data prediction in full waveform inversion: *arXiv preprint arXiv:1912.09944*.
- Kalita, M., V. Kazei, Y. Choi, and T. Alkhalifah, 2019, Regularized full-waveform inversion with automated salt flooding: *Geophysics*, **84**, R569–R582.
- Kazei, V., O. Ovcharenko, T. Alkhalifah, and F. Simons, 2019a, Realistically textured random velocity models for deep learning applications: Presented at the 81st EAGE Conference and Exhibition 2019.
- Kazei, V., O. Ovcharenko, P. Plotnitskii, D. Peter, X. Zhang, and T. Alkhalifah, 2020, Deep learning tomography by mapping full seismic waveforms to vertical velocity profiles: 82nd EAGE Conference and Exhibition 2020, European Association of Geoscientists & Engineers, 1–5.
- Kazei, V., O. Ovcharenko, P. Plotnitskii, D. Peter, X. Zhang, and T. A. Alkhalifah, 2019b, Mapping seismic data cubes to vertical velocity profiles by deep learning: *New full-waveform inversion paradigm?*: *GEOHYPHICS*, submitted.
- Kazei, V., E. Tessmer, and T. Alkhalifah, 2016, Scattering angle-based filtering via extension in velocity: *SEG Technical Program Expanded Abstracts 2016*, Society of Exploration Geophysicists, 1157–1162.
- Kingma, D. P., and J. Ba, 2014, Adam: A method for stochastic optimization: *arXiv preprint arXiv:1412.6980*.
- Köhn, D., D. De Nil, A. Kurzmann, A. Przebindowska, and T. Bohlen, 2012, On the influence of model parametrization in elastic full waveform tomography: *Geophysical Journal International*, **191**, 325–345.
- Kragh, E., and P. Christie, 2002, Seismic repeatability, normalized rms, and predictability: *The Leading Edge*, **21**, 640–647.
- Li, Y. E., and L. Demanet, 2016, Full-waveform inversion with extrapolated low-frequency data: *Geophysics*, **81**, R339–R348.
- Lin, Y., and Z. Zhang, 2019, Learning or memorizing: A study of generalization issues on data-driven seismic full-waveform inversion, *in* *SEG Technical Program Expanded Abstracts 2019*: Society of Exploration Geophysicists, 2644–2648.
- Mosser, L., O. Dubrule, and M. Blunt, 2018, Stochastic seismic waveform inversion using generative adversarial networks as a geological prior: Presented at the First EAGE/PESGB Workshop Machine Learning.
- Oristaglio, M., 2015, Seam update: The arid model seismic exploration in desert terrains: *The Leading Edge*, **34**, 466–468.
- Ovcharenko, O., V. Kazei, M. Kalita, D. Peter, and T. Alkhalifah, 2019a, Deep learning for low-frequency extrapolation from multioffset seismic data: *GEOPHYSICS*, **84**, R989–R1001.
- Ovcharenko, O., V. Kazei, D. Peter, and T. Alkhalifah, 2017, Neural network based low-frequency data extrapolation: Presented at the 3rd SEG FWI workshop: What are we getting.
- , 2018a, Variance-based model interpolation for improved full-waveform inversion in the presence of salt bodies: *Geophysics*, **83**, R541–R551.
- , 2019b, Style transfer for generation of realistically textured subsurface models, *in* *SEG Technical Program Expanded Abstracts 2019*: Society of Exploration Geophysicists, 2393–2397.
- Ovcharenko, O., V. Kazei, D. Peter, X. Zhang, and T. Alkhalifah, 2018b, Low-frequency data extrapolation using a feed-forward ANN: 80th EAGE Conference and Exhibition 2018, European Association of Geoscientists & Engineers, 1–5.
- 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*, Springer, 234–241.
- Siahkoobi, A., G. Rizzuti, and F. J. Herrmann, 2020, A deep-learning based bayesian approach to seismic imaging and uncertainty quantification: *arXiv preprint arXiv:2001.04567*.
- Sun, H., and L. Demanet, 2018, Low frequency extrapolation with deep learning, *in* *SEG Technical Program Expanded Abstracts 2018*: Society of Exploration Geophysicists, 2011–2015.
- , 2019, Extrapolated full waveform inversion with deep learning: *arXiv preprint arXiv:1909.11536*.
- Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation: *siam*, **89**.
- Ten Kroode, F., S. Bergler, C. Corsten, J. W. de Maag, F. Strijbos, and H. Tijhof, 2013, Broadband seismic data The importance of low frequencies: *Geophysics*, **78**, WA3–WA14.

- Wu, Y., and Y. Lin, 2019, Inversionnet: An efficient and accurate data-driven full waveform inversion: *IEEE Transactions on Computational Imaging*.
- Yuan, C., X. Zhang, X. Jia, and J. Zhang, 2020, Time-lapse velocity imaging via deep learning: *Geophysical Journal International*, **220**, 1228–1241.
- Zwartjes, P., 2020, Near surface velocity estimation from phase velocity-frequency panels with deep learning: 82nd EAGE Conference and Exhibition 2020, European Association of Geoscientists & Engineers, 1–5.