Bi-Directional LSTM-based non-causal deconvolution

G. Roncoroni[1], I. Deiana[2], E. Forte[1], M. Pipan[1]*
[1] University of Trieste, MIGE, Trieste, Italy
[2] Stanford University, Stanford, USA

Summary
The paper introduces a novel methodology utilizing a bidirectional Long Short-Term Memory (LSTM) neural network for seismic data deconvolution. By training the network on synthetic seismic traces, it achieves consistent and accurate deconvolution results without requiring model retraining. The approach is compared to Spiking and Ricker-compliant deconvolution results. The methodology is adaptable to different signal-to-noise ratio scenarios and can be applied to various seismic datasets without the need for retraining, making it robust and versatile.

Introduction
Deconvolution aims at improving the resolution of seismic images by eliminating unwanted effects such as reverberations, multiples, and source wavelets (see e.g. Dondurur, 2010) and is a critical step in seismic data processing. Its ultimate goal is to recover the reflectivity sequence from the recorded seismic data by eliminating the source wavelet (Dondurur, 2010; Santamaria et al., 1999). Deconvolution techniques have been used extensively in seismic exploration to enhance the interpretation of the geological structures (Dasgupta & Nowack, 2008). This method is rooted in the principles of spike deconvolution, which seeks to enhance the resolution of seismic data by transforming seismic wavelets into shorter transients (in theory approaching to spikes) while suppressing undesired events (Menanno & Mazzotti, 2011).
Additionally, the integration of neural networks in deconvolution processes has attracted more and more attention due to their ability to discern even intricate patterns and relationships in data, leading to enhanced deconvolution outcomes (Iqbal et al., 2019).
We developed a bidirectional Long Short-Term Memory (LSTM) neural network (NN) methodology to deconvolve seismic data. Our approach produces consistent and accurate deconvolution results without the need for model retraining. Its implementation is user-friendly and does not require specific parameter adjustments.
Notably, the approach surpasses Spiking deconvolution results and, in some cases, matches and even outperforms refined algorithms like Ricker-compliant deconvolution - RDC (Clearbout and Guitton, 2014) being fast and computationally efficient. These characteristics qualify our method as an asset for seismic data processing.

Method
In this work, we propose a methodology to map the non-linear transformation between band-limited seismic data and its deconvolved counterpart.
We achieve this outcome by training a bidirectional Long Short-Term Memory (LSTM) neural network (NN). To generate a large training dataset, we create synthetic seismic traces and their corresponding deconvolved versions. The deconvolved wavelet is computed by numerically summing Ricker wavelets within a specific frequency range based on the input trace's central frequency. In detail, the range is defined considering the central frequency (*f*) of the input trace (with f varying randomly between *f/2* and *2f*), retaining frequencies between *f/2* and $f*2^4$. This varying range of frequency in the input dataset gives us the ability to deal with different frequencies.

Our training approach utilizes a convolutional model. The model takes the convolved seismic traces as input and aims to predict the deconvolved time series derived from the summed Ricker wavelets. This approach enabled the generation of 100,000 random traces for the NN training within a reasonable timeframe (approximately 670 seconds). To generate the training dataset, we used a modified convolutional model defined as:

$$\mathrm{trace_i}(t) = w_i(t) * (r(t) + n_1(t)) + n_2(t)$$
$$\mathrm{trace_o}(t) = w_o(t) * (r(t) + n_1(t)) + n_2(t)$$

where $\mathrm{trace_i}$ and $\mathrm{trace_o}$ represent the input and the reference output traces, respectively; r(t) is a randomly distributed reflection coefficient series and $n_1(t)$ represents a random noise component added before convolution on these coefficients. $n_1(t)$ is obviously the term that affects more the SNR of the dataset, while $n_2(t)$ represents a random noise component added after the convolution process. Since the only difference between the two-training datasets is the wavelet, the model aim is to learn a nonlinear transformation between $\mathrm{trace_i}$ and $\mathrm{trace_o}$, i.e. learn the deconvolution operator.
The random trace generation process allows us to explore the impact of noise characteristics (i.e., noise distribution in the coefficient series), as well as the number of reflections on the deconvolution process. This flexibility enables us to adapt the methodology to different signal-to-noise ratio (SNR) scenarios, if needed. However, for this study, we focus on a unique bi-LSTM neural network model, trained on data with a SNR ranging from 0.01 to 0.2, with respect to the maximum amplitude of each trace. Importantly, due to the learned mapping in the randomly generated synthetic training dataset, retraining for different real-world dataset

(both land or sea) is not necessary, making the methodology robust, versatile and almost site independent.
To perform the training, we weighted the loss function of the trace to obtain a map that is compliant not only with the time-amplitude information, but also with the frequency counterpart. Since the network initially has no physical information about how the amplitude, phase and frequency contents are related, unfolding the loss term can enhance the training and gradient backpropagation. Our proposed loss function can be summarized as:

$$loss(X,\ Y) = \left\| X\ -\ Y \right\|_2 +\ \lambda \left\| \mathbb{R}(\hat{X})\ -\mathbb{R}(\hat{Y}) \right\|_2$$

where X is the reference output, Y is the NN prediction and $\mathbb{R}(\hat{X})$ and $\mathbb{R}(\hat{Y})$ are the real parts of their Fourier transforms, respectively; λ is a regularize parameter, set to 0. During training. The first term is the classical loss function, while the second term is its frequency domain counterpart. The addition of a frequency related term to the classical loss function allows to give a higher contribution to the frequency dependence of the traces, which would have less impact in early stages of the gradient descent-based optimization of a classical mean square error.

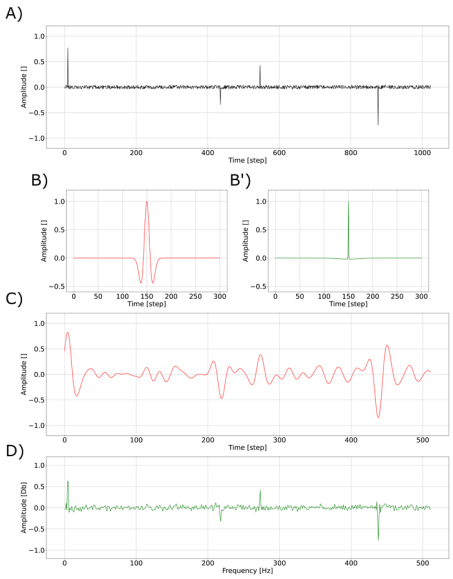

Figure 1: Training dataset generation for the proposed methodology: A) is the reflection coefficient series with noise. B) and B') are the input and the reference wavelets (w_i and w_o in Eq. 1) respectively. C) is the input data generated by convolution of reflection coefficient series A) with wavelet B). D) is the reference training data, generated by convolution of reflection coefficient series A) with wavelet B').

Our methodology addresses the non-linear mapping between seismic signals within a limited frequency band (arbitrarily choosing 0.1-0.4 Hz) with a unitary sampling interval. This choice simplifies computations but does not restrict the applicability of the method to a specific frequency range. In fact, to overcome this limitation, we introduce a resampling step that effectively scales the input data to the input frequency range, established during the NN training. This is possible because the underlying mapping process remains agnostic to the original sampling interval of the input data.
This resampling step aligns the input data with the network's frequency input training range. We achieve this by computing a resampling factor that minimizes the Pearson correlation coefficient (PCC) (Schober et al., 2018) between the target spectrum (the average spectrum of the input training data) and spectra representing resampled versions of the input data. PCC measures the linear correlation between variables, in this context highlighting the spectral similarity. By iteratively evaluating the PCC values for various resampling factors, we identify the factor corresponding to the minimum value (i.e. closest to zero), signifying minimal spectral difference. This minimum point therefore represents the optimal resampling factor for the specific input data.

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

where: X_i and Y_i are the individual observations of variables X and Y, while $\bar{X}$ and $\bar{Y}$ are the means of variables X and Y, respectively.
We developed an algorithm that automates this process to avoid subjectivity and for user convenience. It explores candidate resampling factors, resamples the data accordingly, calculates the spectral similarity using PCC, and refines the search based on the initial best value. Ultimately, the function returns the resampling factor that leads to a resampled version with a frequency spectrum closely matching the target spectrum. This ensures optimal data preparation for the subsequent actual deconvolution process. By leveraging the learned mapping in the random domain and performing an optimal data resampling step, we achieve a methodology that is generalizable to a wide range of seismic deconvolution problems, encompassing both land and sea datasets obtained using different impulsive seismic sources and even Ground Penetrating Radar (GPR) data. Furthermore, by modifying the resample factor it is possible to fine tuning the results of the deconvolution.
The entire methodology can be summarized in a simple workflow, as follow:

- Amplitude balance: apply gain and/or other amplitude normalization algorithms
- Compute Pearson resample factor and resample the data
- Apply frequency filters (Optional)
- Deconvolve using the NN operator
- Apply frequency filters (Optional)

Results

To evaluate the prediction quality, we performed a comparison of our approach results -referred as LSTM deconvolution from now on- with the RCD (Claerbout and Guitton, 2014) process on the Debubbled version of the Gulf of Mexico Dataset used in the cited paper to evaluate the RCD process (Figure 2). The LSTM approach performs a deconvolution comparable to the RCD case, obtaining spikes at the center lobe of the Ricker wavelet in the reflections.

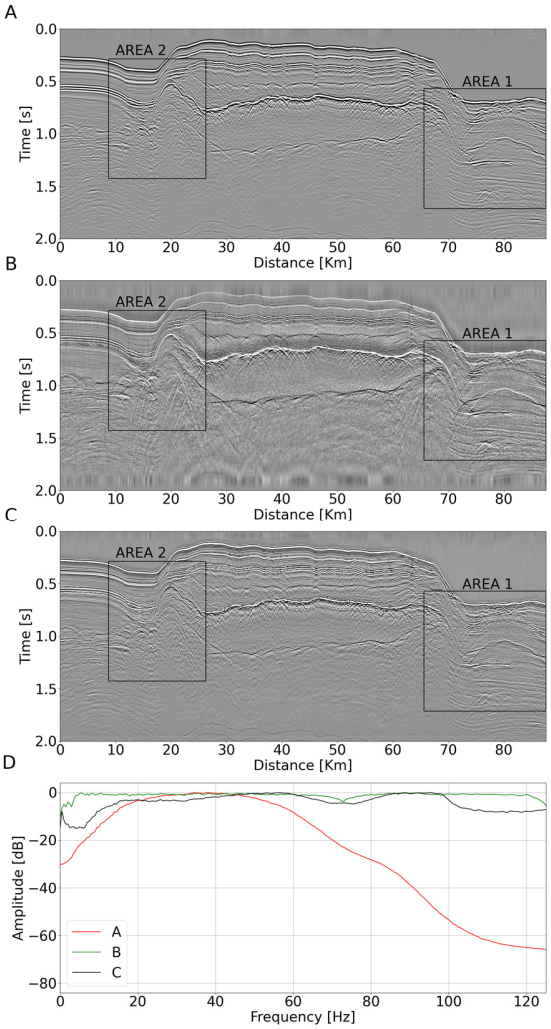

Figure 2: Comparison between raw data (A), RCD (B) and LSTM deconvolution (C). In D we can see the power spectrum of input (in red), of LSM deconvolution (in black) and of RCD (in green).

Both algorithms preserve the reflection polarities, even in complex zones like subsalt reflections and the multiple reflections at around 2s. The power spectrum of the LSTM output is not as white as in the RDC, mainly due to a more limited amount of low and high frequencies. This behavior is compliant to the training wavelet and it could be further improved just slightly modifying the target wavelet. The artifacts in the lower part of the RCD are neglected as they are an operator artefact that can be solved by extending the model borders.

Figure 3 shows a zoom on the right portion of the Gulf of Mexico section. The increase in resolution with respect to the raw section on both algorithms is very clear. In this scenario, the top right part of the LSTM outputs more potential reflections than the RCD approach and the hyperbolas are also preserved better in the LSTM deconvolution.

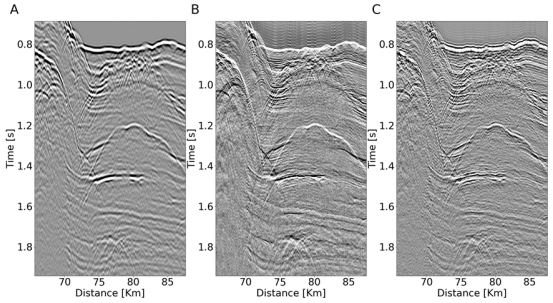

Figure 3: Zoom of Area 1 depicted in Figure 2. A) Raw data, B) RCD and C) LSTM Deconvolution. See text for further details.

In Figure 4 we can see a detailed zoom on the left flank of the salt dome. Here the LSTM results shows again a high resolution section comparable to the results of the RCD in this complex geological setting. The LSTM does not introduce artifacts or high frequency noise, and again well preserves the shape and definition of hyperbolic events.

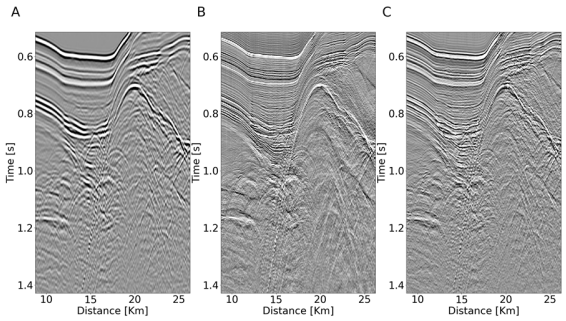

Figure 4: Zoom on Area 2 as depicted in Figure 2. A) Raw, RCD B) and LSTM Deconvolution C). See text for further discussion.

LSTM non-causal deconvolution

To further test the methodology, we apply the same NN, i.e. without any retraining, on a vintage land data from USGS, (NPRA, 2001).

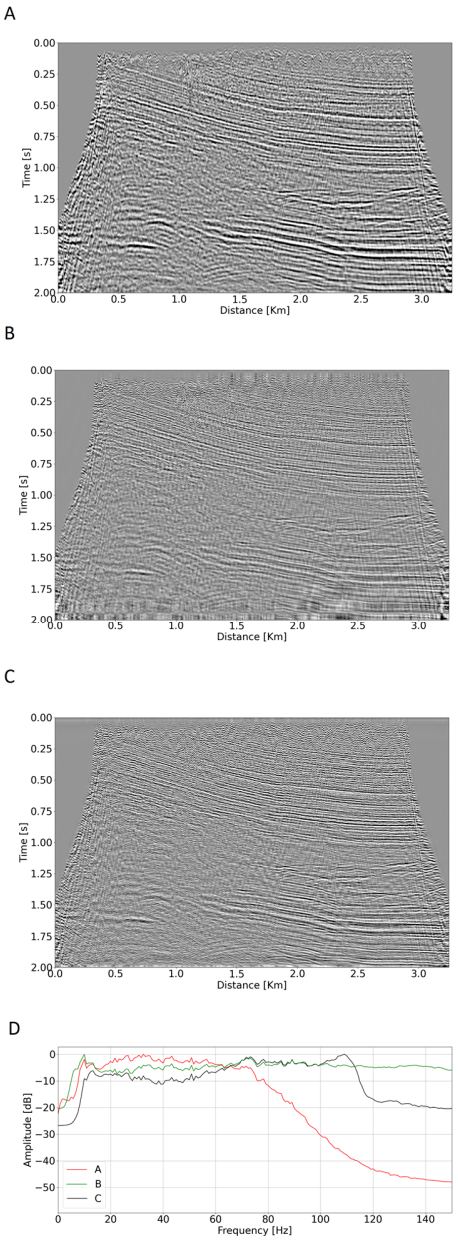

Figure 5: Comparison between raw data (A), Spiking Deconvolution (B) and LSTM deconvolution (C). In D we can see the power spectrum of input (in red), of Spiking Deconvolution (in green) and of LSM deconvolution (in black).

In this test we performed a comparison with a classical commercial Spiking deconvolution to compare our methodology to a standard processing procedure, in an acquisition setting where Spiking deconvolution can deliver good results.

In Figure 5 we can see the raw data, A, the LSTM deconvolution, B, the Spiking deconvoluted data, C, and in D the power spectrum of the 3 data mentioned above.

If we compare the results obtained by the Spiking deconvolution with the LSTM deconvolution, we can appreciate a similar performance in the lower part of the section, where both methodologies manage to correctly deal with the strong reflections. In the first part, where the SNR is lower, we can see that horizontal reflection are better enhanced in the LSTM deconvolution and the SNR is increased by the proposed methodology.

Conclusions

We introduce a novel seismic deconvolution approach leveraging a bidirectional LSTM neural network trained on synthetic seismic traces. Compared to traditional methods, our methodology offers several advantages. It delivers consistent and accurate results without the need for model retraining, promoting efficiency, generalization, and user-friendliness. Additionally, it can surpass established techniques like Spiking deconvolution and, in specific scenarios, even rivals RCD in terms of resolution, being fast to implement, requiring little data analysis and barely eliminating user parameters tuning. Furthermore, the method boasts computational efficiency, yielding rapid results, making it a valuable asset for seismic data processing workflows.

Crucially, our methodology adapts to various signal-to-noise ratios and frequency bands, offering flexibility for real-world applications. Its inherent robustness and versatility enable seamless application across diverse seismic datasets without retraining, streamlining the processing pipeline. Beyond accurate deconvolution, the approach has the potential to enhance seismic data resolution and interpretation, aiding in improved subsurface structure delineation and geological feature identification.

Future work will focus on further optimization of the method and exploration of its applicability to a broader range of geological settings, source signatures, and data types. We believe this approach holds significant promise for enhancing seismic imaging and interpretation across various applications, ultimately contributing to a deeper understanding of geological structures

Acknowledgments

The authors would like to acknowledge Antoine Guitton for his help in performing the comparison with the RCD method and Nicola Bienati for his insightful discussions and advices on the proposed methodology.