GEOPHYSICS®

Merging gated Frequency-Modulated Continuous-Wave Mars2020 RIMFAX GPR data

Journal:	*Geophysics*
Manuscript ID	GEO-2022-0466.R1
Manuscript Type:	Letters
Keywords:	ground-penetrating radar (GPR), artificial intelligence, frequency-domain
Manuscript Focus Area:	Ground-Penetrating Radar

DOI:10.1190/geo2022-0466.1

GEOPHYSICS

erging gated Frequency-Modulated Continuous-Wave Mars2020 RIMFAX GPR data

Right Running Head: Merging Mars2020 RIMFAX data

Roncoroni, Giacomo; Forte, Emanuele; Pipan, Michele

University of Trieste

Merging gated Frequency-Modulated Continuous-Wave Mars2020 RIMFAX GPR data

ABSTRACT

The integration of GPR data at various frequencies, collected with different antennas or with the use of swept-frequency radars opens up interesting perspectives in the study of the subsurface at different resolutions. The proposed methodology is a semi-supervised DL algorithm based on Bi-Directional Long-Short Term Memory to automatically merge varying numbers of data sets at different frequencies. Neural Network training is done directly on the inference data by minimizing a custom loss function based on the L2 norm of all the input data, weighted on the custom merging area and the single output trace. The inference of the trained Neural Network is applied to the same data. The proposed algorithm is tested on synthetic data simulating the Mars conditions and on RIMFAX radar data collected in the Jezero crater during the Mars2020 mission of Perseverance rover, showing successful performances and robustness.

INTRODUCTION

Based on the spatial decay rate of EM waves propagating through lossy dielectrics, higher frequency Ground Penetrating Radar signals allow for higher resolution but lower overall penetration and vice-versa for lower-frequency data. With the increasing availability of instruments offering multi-frequency capabilities, such as the swept-frequency radar RIMFAX (Hamran et al., 2020) adopted during Mars2020 NASA mission or the dual-frequency radar mounted under the Yutu-2 rover in mission Chang E4 (Li et al., 2021) a fast, robust and computationally efficient data fusion methodology is essential to combine and exploit all available information and to overcome the trade-off between penetration and resolution.

At the state of the art, GPR data fusion approaches integrate data at different frequencies performing semi-automatic merging based on statistical methods and probabilistic techniques, in time or frequency (Booth et al., 2009; Bi et al., 2020). Other methods rely on the 2-D wavelet transform to derive a dynamic fusion weighted scheme (Lu et al., 2020) or exploit genetic algorithms adapting the weight of different combined data (Zhao et al., 2020). De Coster and Lambot (2018) proposed a method applied after removing the specific effects of antennas from GPR data, while Soldovieri and Orlando, 2009 suggested a strategy based on tomographic inversions then combined between partially overlapping frequency bands. Interestingly, data fusion approaches have also been proposed and tested to combine different data from non-destructive techniques (e.g. Scott et al., 2004; Kohl et al., 2005) as well as in different other topics such as the detection of cracks based on eddy currents (Efremov et al., 2022).

Although important advances have been achieved in most of the proposed approaches, it is difficult to obtain a smooth transition in the sections obtained by merging different frequency profiles. Indeed, data fusion procedures often require non-physical amplitude equalization and amplitude balance (De Coster and Lambot, 2018) and arbitrary window selections to optimize the output because fusion results typically have unwanted “cut and paste” effects between the different combined windows.

We propose a new approach that uses a Recurrent Neural Network (Rumelhart et al., 1986), in particular a Long-Short Term Memory - LSTM (Hochreiter and Schmidhuber, 1997), to automatically merge variable numbers of data sets at different frequencies, without specific requirements and limitations of input data. The introduction of a user-defined merging area can provide the analyst with additional control over the final merged profile, but the algorithm can also handle different merging areas. RIMFAX data are collected in three different and partially overlapping windows (reported as “Surface”, “Shallow” and “Deep” modes, respectively, Hamran et al., 2020, Fig. 18 therein), having three different bandwidths (equal in free space to 1050, 750 and 450 MHz, respectively, Hamran et al., 2020, Table 5 therein).

The application of the proposed algorithm on RIMFAX radar data is here critically evaluated, analyzed and discussed, demonstrating that the merged results are robust, the bandwidth is expanded and, in turn, the overall resolution is increased allowing a better interpretation of such a unique GPR dataset.

METHODS

The proposed methodology is based on the training and inference of a Neural Network (NN) trained on a small portion (typically 10%) of the analyzed dataset. This approach can be classified as a semi-unsupervised DL procedure and exploits a fully 1-D approach, i.e. is made trace by trace. To merge datasets characterized by different frequency bands and by different time windows location and lengths as for RIMFAX data, we use typical tools implemented for NN training, such as the Gradient Descent optimizations algorithms, and the power given by a custom transformation based only on a few neuron weights.

The basis of this approach is that the fusion should be performed by 3 layers of a Bi-Directional LSTM (Schuster and Paliwal, 1997), with 4, 2 and 1 Bi-Directional layers and a single LSTM neuron for the output, trying to minimize the custom loss function with a single prediction from both input data. Using a few parameters for all data, as a consequence of using a very small NN, offers a robust method for outliers and even for variations of noise and amplitude between adjacent traces, which typically affect the analyzed dataset. A graphical representation of this training scheme is provided in Figure 1. In addition, the user can set an optional input parameter, hereafter referred to as "merging interval". This parameter, $val_{[0,1]}^{k}$ in the following equations, is introduced into the algorithm directly in the loss function and should limit the NN to not consider as input areas lacking useful information, such as typically near the end of some recording windows where the signal-to-noise ratio is very low.

Figure 1

The loss function is expressed by the equation:

$$loss = \sum_{k=1}^{n} \lambda_k loss_k \#(1)$$

where n is equal to 3 for RIMFAX data and $loss_k$ is defined as:

$$loss_k = \|target_k - prediction\|_2^2 \#(2)$$

where the merged prediction is compared to the $k_{[1,2]}^{th}$, while λ_k is defined as a piece-wise function:

$$\lambda_k(t) = \begin{cases} 0 & t < \mathrm{val}_0^k \\ 1 & val_0^k < t < val_1^k \\ 0 & t > \mathrm{val}_1^k \end{cases} \#(3)$$

in which val_0^n and val_1^n are the lower and the upper boundary of the k^{th} data, respectively, as shown in Figure 1.

At first, we tested such a method on synthetic data simulating the expected subsurface of Mars (Figure 2).

Figure 2

We defined a random 1-D model, as shown in Figure 2-a, defining 2 main layers, i.e. 10m of sedimentary cover with $\epsilon = 2.8$ and a bedrock with $\epsilon = 7.8$, as proposed by Hamran et al., 2020. In order to make the merging more meaningful we introduced a vertical variability in the sediments randomizing ϵ within the range 2.8 ± 1 and generating a random model with 4 layers in the first 2 meters and a layer per meter until the bedrock. We set a constant electrical conductivity $\sigma = 0.002Wm$ to see how the merging algorithm would deal with amplitude attenuation.

The modeling is performed with GprMax (Warren et al., 2016) and as source we used a Ricker wavelet with the central frequencies reported in Hamran et al., 2020 for Surface, Shallow and Deep acquisition modes. In order to simulate our target acquisition, we deleted data outside from the receiver window for each specific frequency as summarized in Hamran et al., 2020 – Figure 18 therein.

The merged results (Figure 2-e) are very accurate in time as compared with "Surface", "Shallow" and "Deep" traces (Figure 2-b, c, d, respectively) and well balanced in amplitude. For instance, the reflection at about 45ns appears properly reconstructed with a mix between Shallow and Deep modes. In addition, the procedure does not have any specific problem and does not introduce artifacts when abrupt velocity changes or velocity inversions are present.

RESULTS AND DISCUSSIONS

After selecting data from the entire RIMFAX dataset (from SOL 072 to SOL 204) to avoid redundant data due to continuing data acquisition during rover stops based on the analysis of the amplitude of the raw data, we applied an exponential gain function to compensate for the observed amplitude decay.

We then started the NN training after selecting one trace out of 10 (see (https://github.com/Giacomo-Roncoroni/merging_RIMFAX for the full code). We chose to perform the training on only the 10% of the dataset just for computational efficiency: in theory it would be possible to use the whole dataset, as we do not over fit the problem with this approach.

Since the entire dataset is quite long and contains a total of 23199 traces (after removing repeated ones as above described), (Figure S3 supplementary) we analyze two specific portions characterized by slightly different features and structures, namely: sub-horizontal layers with lenses (Figure 3) and monocline dipping reflectors (Figure 4). Merged results are in both cases quite good (see Figures S4, S7, S8 supplementary for the amplitude spectra). in particular, no artifacts are introduced and the continuity of reflectors is preserved. Moreover, merged data in Figure 3 clearly show the vertical sequence of layers, which is difficult to understand by separately analyzing the three RIMFAX acquisition modes (see also Figures S11 and S12 supplementary for the interpretation). In fact, from about 20 and 55ns two separated sub-parallel reflectors are imaged, as well as two superimposed lenses. In Figure 4 the dipping reflectors are perfectly reconstructed with no vertical gaps from just a few ns down to 250 ns.

Abrupt lateral amplitude variations in the merged profiles are not introduced by the application of the fusion algorithm since they are present (and sometimes even more apparent) also on the three single acquisition modes. In any case, they could be easily reduced by applying lateral trace balance or, for instance, f-x adaptive trace interpolation (e.g. Naghizadeh and Sacchi, 2009). However, lateral phase continuity is high and no trace gaps like the above described "cut and paste" effects are introduced by the merging algorithm (Figures S9 and S10 supplementary); as a consequence, merged data can be further processed and analyzed (or even inverted) without losing information or the risk of introducing outliers, artifacts or coherent noise components. In Figure 5 we provide a comparison between the results obtained with the proposed procedure and the ones recently published by Hamran et al., 2022 where the "cut and paste" effect is apparent on all the traces (Figure 5c1 and 5d1) in which an abrupt frequency variation is present. This does not occur with the proposed procedure (Figure 5c2 and 5d2), obtaining a smoothed variation of the spectral content. Moreover some horizons (h labels) are clearer and show higher lateral continuity.

Figure 3

Figure 4

Figure 5

Regarding the computation times over the entire dataset, a matrix with 23000 traces, 5000 time samples for the 3 modes, training time with 10% of the data, as described above, takes only 970s on a laptop with Intel(R) Core (TM) i7-10875H, 32Gb RAM and a Nvidia GeForce RTX 2070 Super with 8Gb of memory. On the same laptop, prediction time for the entire dataset takes 24s.

CONCLUSION

We succeeded in creating a DL-based methodology able to merge RIMFAX multi frequency GPR data that is robust, fast and does not require particular data pre-processing and conditioning.

The introduction of specific user-defined merging windows, allows to make the methodology more accurate and controlled, according to the selection of those time windows in which the signal-to-noise ratio is higher.

Merged data can be further analyzed, processed and potentially inverted since they do not suffer from possible introduction of amplitude/phase gaps or local/coherent artifacts. Another strength of the methodology is that it is completely data driven and can handle even very noisy data.

The application of the proposed strategy is definitely not limited to three windows and/or frequency components like in the case of RIMFAX dataset, but can be applied to merge any type of dataset with multiple spectral components and recording windows.

ACKNOWLEDGMENTS

This research was partially supported the project “Dipartimento di Eccellenza" of the Department of Mathematics and Geosciences of the University of Trieste. We gratefully acknowledge the support of Shearwater and Hallliburton Landmark through their academic grants. We further thank three anonymous reviewers for their fruitful comments and suggestions.

REFERENCES

Bi, W., Y. Zhao, R. Shen, B. Li, S. Hu, and S. Ge, 2020, Multi-frequency GPR data fusion and its application in NDT: NDT & E International, **115**, 102289, doi: 10.1016/j.ndteint.2020.102289.

Booth, A. D., A. L. Endres, and T. Murray, 2009, Spectral bandwidth enhancement of GPR profiling data using multiple-frequency compositing: Journal of Applied Geophysics, **67**, 1, 88–97, doi: 10.1016/j.jappgeo.2008.09.015.

De Coster, A., and S. Lambot, 2018, Fusion of Multifrequency GPR Data Freed From Antenna Effects:IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing, **11**, 2, 664–674, doi: 10.1109/JSTARS.2018.2790419.

Efremov, A., O. Karpenko, and L. Udpa, 2022, Generalized multifrequency fusion algorithm for defect detection in eddy current inspection data: NDT & E International, **129**, 102654, doi: 10.1016/j.ndteint.2022.102654.

Hamran, S. E., D. A. Paige, H. E. F. Amundsen, et al., 2020, Radar Imager for Mars’ Subsurface Experiment—RIMFAX: Space Sci Rev, **216**, 8, 128, doi: 10.1007/s11214-020-00740-4.

Hamran, S. E., Paige,D. A., A. Allwood, et al., 2022, Ground penetrating radar observations of subsurface structures in the floor of Jezero crater, Mars: J. Science Advances, **8**, 34, doi: 10.1126/sciadv.abp856

Hochreiter, S., and J. Schmidhuber. 1997, Long Short-Term Memory: Neural Computation, **9**, 8, 1735-1780.

Kohl, C., 2005, 2D- and 3D-visualisation of NDT-data using data fusion technique: Mater. Struct., **38**, 283, 817–826, doi: 10.1617/14293.

Li, C., W. Zuo, W. Wen, et al., 2021, Overview of the Chang'e-4 Mission: Opening the Frontier of Scientific Exploration of the Lunar Far Side: Space Sci Rev, **217**, 2, 35, doi: 10.1007/s11214-021-00793-z.

Lu, G., W. Zhao, E. Forte, G. Tian, Y. Li, and M. Pipan, 2020, Multi-frequency and multi-attribute GPR data fusion based on 2-D wavelet transform: Measurement, **166**, 108243, doi: 10.1016/j.measurement.2020.108243.

Naghizadeh M., and M. Sacchi, 2009, Making FX Interpolation More Robust by Spectrum-guided Reconstruction: Frontiers + Innovation, CSPG CSEG, CWLS Convention, Calgary, Alberta, Canada, 376-379. 4,.

Rumelhart, D. E., G.E. Hinton, and R. J. Williams, 1986, Learning representations by back-propagating errors: Nature, **323**, 533-536.

Schuster, M., and K. K. Paliwal, 1997, Bidirectional recurrent neural networks: IEEE Trans. Signal Process., **45**, 11, 2673–2681, doi: 10.1109/78.650093.

Scott, W. R., Kangwook Kim, G. D Larson,.A. C. Gurbuz, and J. H. McClellan, 2004, Combined seismic, radar, and induction sensor for landmine detection: IEEE International Geoscience and Remote Sensing Symposium, IGARSS ’04, Proceedings. 2004, Anchorage, AK, USA, **3**, 1613–1616. doi: 10.1109/IGARSS.2004.1370637.

Soldovieri, F., and L. Orlando, 2009, Novel tomographic based approach and processing strategies for GPR measurements using multifrequency antennas: Journal of Cultural Heritage, **10**, e83–e92, doi: 10.1016/j.culher.2009.09.001.

Warren, C., A. Giannopoulos, and I. Giannakis, 2016, gprMax: Open source software to simulate electromagnetic wave propagation for Ground Penetrating Radar: Computer Physics Communications, doi:10.1016/j.cpc.2016.08.020.

Zhao, W., L. Yuan, E. Forte, G. Lu, G. Tian, and M. Pipan, 2021, Multi-Frequency GPR Data Fusion with Genetic Algorithms for Archaeological Prospection: Remote Sensing, **13**, 14, 2804, doi: 10.3390/rs13142804.

LIST OF FIGURES

Figure 1: Graphical representation of the training scheme: in the left side, i.e. input, we have the 3 input data namely Surface, Shallow and Deep, respectively, with the merging interval marked with solid lines. Green lines represent val_0^k for each input, while the red lines represent val_1^k for each input (Equation-3). NN represents the LSTM neurons into a Bi-Directional wrapper and in Output we find the prediction i.e. the merged data.

Figure 2: Synthetic simulation and merging obtained on a random model. The model a) – and the Surface, Shallow and Deep simulation modes, b), c) and d), respectively are shown. We applied the same time windows imposed by RIMFAX Radar system and we merged the data with the proposed methodology obtaining the results in e). See text for further details.

Figure 3: SOL 113-116 RIMFAX data. From top to bottom: Surface, Shallow, Deep and Merged data, respectively.

Figure 4: SOL 200-203 RIMFAX data. From top to bottom: Surface, Shallow, Deep and Merged data, respectively.

Figure 5: Comparison between data published in Hamran et al. 2022 (a), Figure 4B therein and the results of the proposed algorithm (b). c1, d1 are two close up of Hamran et al., 2022 figure, while c2, d2 shows the same portion of the data obtained with the proposed merging strategy. Vertical arrows highlight the cut and paste effect apparent on the Hamran et al., 2022, while "h" labels mark horizons made cleared on data merged with the new proposed procedure.

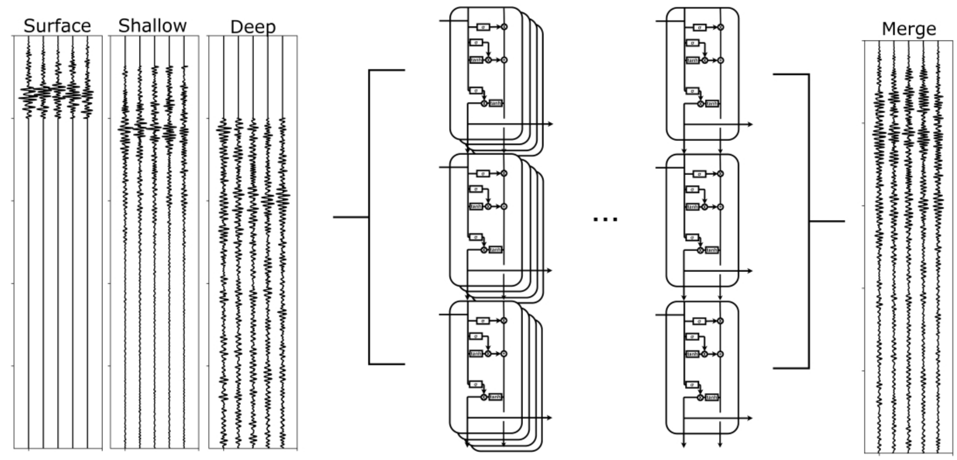

Figure 1: Graphical representation of the training scheme: in the left side, i.e. input, we have the 3 input data namely Surface, Shallow and Deep, respectively, with the merging interval marked with solid lines. Green lines represent 〖val〗_0^k for each input, while the red lines represent 〖val〗_1^k for each input (Equation-3). NN represents the LSTM neurons into a Bi-Directional wrapper and in Output we find the prediction i.e. the merged data.

1206x636mm (118 x 118 DPI)

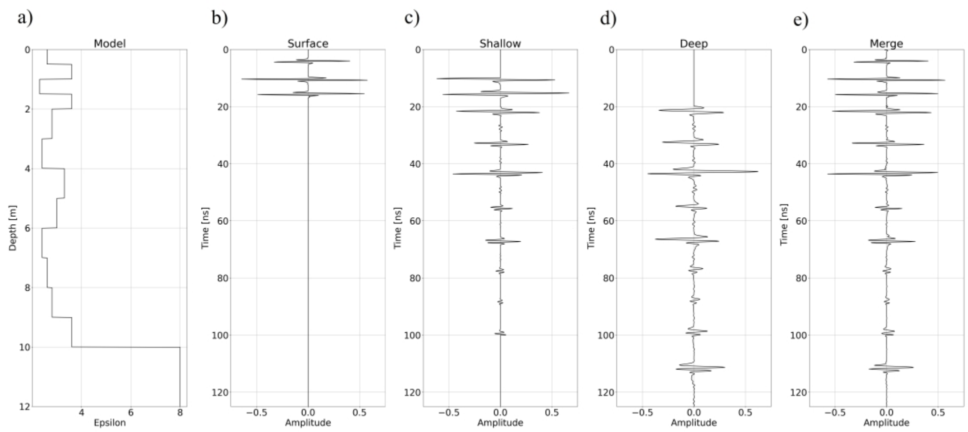

Figure 2: Synthetic simulation and merging obtained on a random model. The model a) – and the Surface, Shallow and Deep simulation modes, b), c) and d), respectively are shown. We applied the same time windows imposed by RIMFAX Radar system and we merged the data with the proposed methodology obtaining the results in e). See text for further details.

4674x2084mm (47 x 47 DPI)

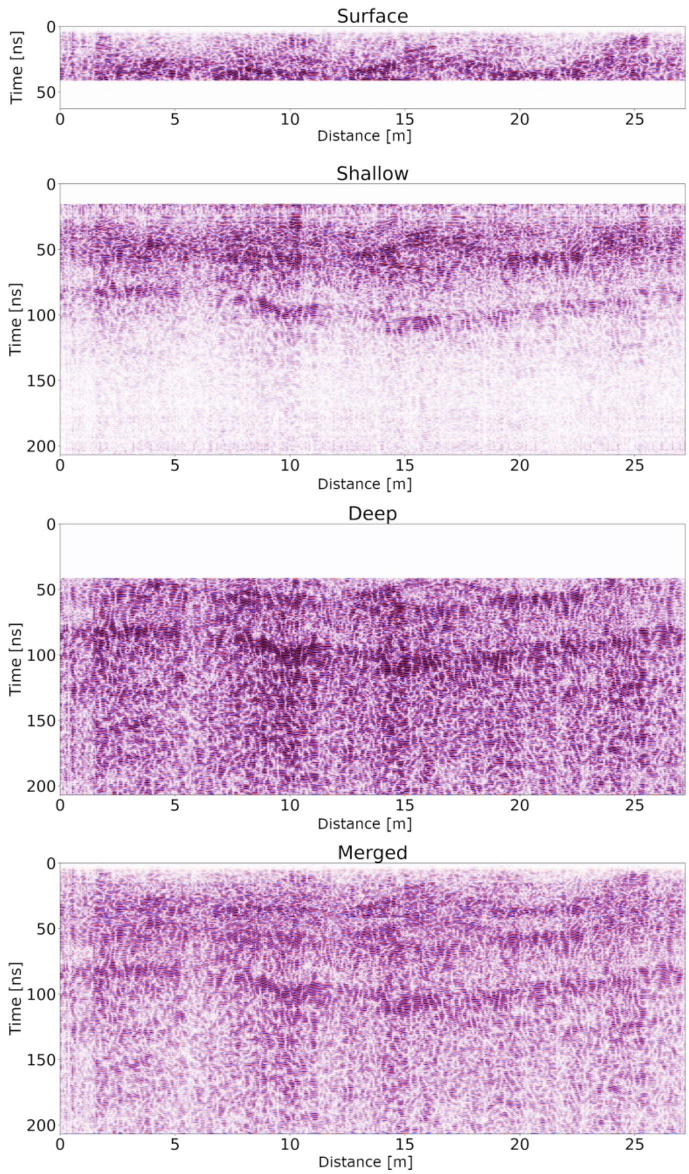

Figure 3: SOL 113-116 RIMFAX data. From top to bottom: Surface, Shallow, Deep and Merged data, respectively.

387x638mm (118 x 118 DPI)

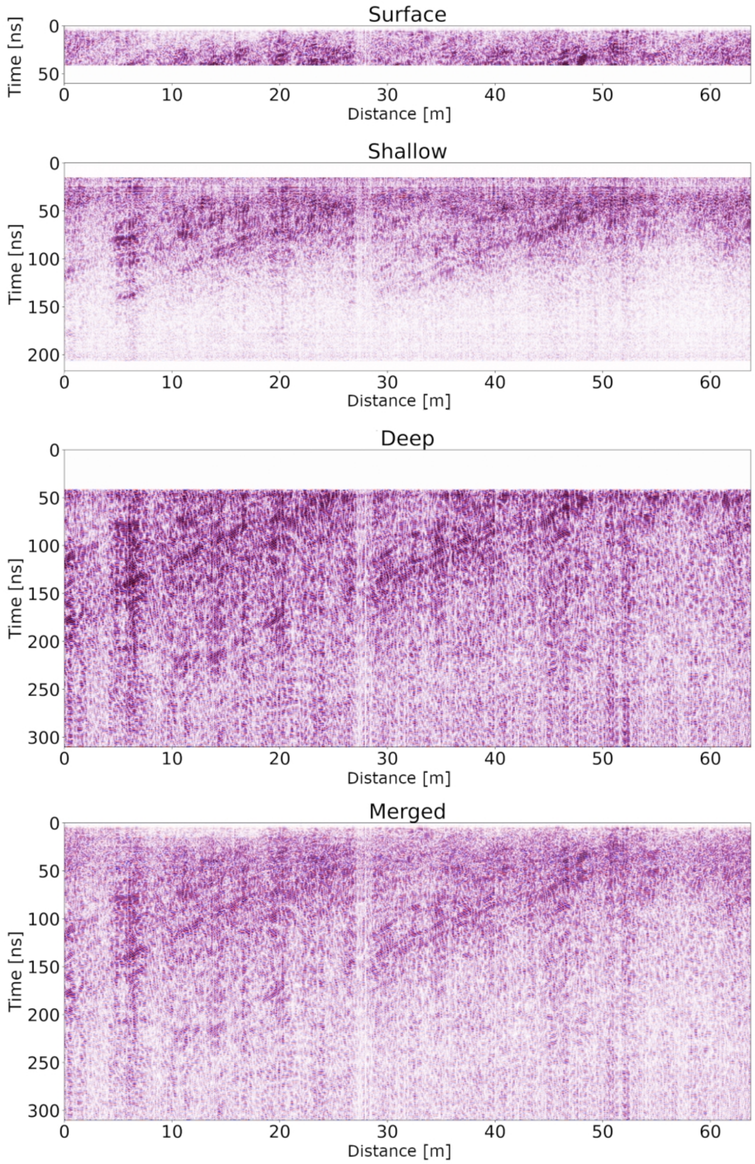

Figure 4: SOL 200-203 RIMFAX data. From top to bottom: Surface, Shallow, Deep and Merged data, respectively.

384x582mm (118 x 118 DPI)

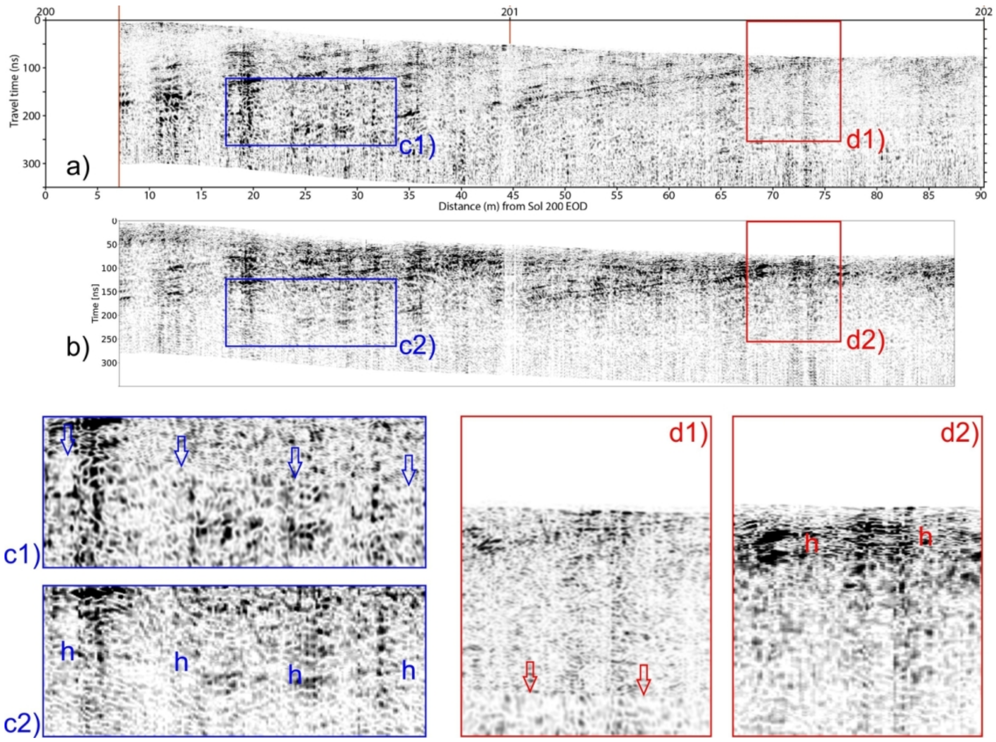

Figure 5: Comparison between data published in Hamran et al. 2022 (a), Figure 4B therein and the results of the proposed algorithm (b). c1, d1 are two close up of Hamran et al., 2022 figure, while c2, d2 shows the same portion of the data obtained with the proposed merging strategy. Vertical arrows highlight the cut and paste effect apparent on the Hamran et al., 2022, while "h" labels mark horizons made cleared on data merged with the new proposed procedure.

131x96mm (300 x 300 DPI)

DATA AND MATERIALS AVAILABILITY

Data associated with this research are available and can be accessed via the following URL:lt;brgt; Note: A digital object identifier (DOI) linking to the data in a general or discipline-specific data repository is strongly preferred.

DOI:10.1190/geo2022-0466.1