UNIVERSITÀ
DEGLI STUDI
DI TRIESTE

UNIVERSITÀ DEGLI STUDI DI TRIESTE

XXXVI CICLO DEL DOTTORATO DI RICERCA IN

**Earth Science, Fluid-Dynamics, and Mathematics.
Interactions and Methods**

Long-Short-Therm Memory in Active Wavefield Geophysical Methods

Settore scientifico-disciplinare: **GEO11**

DOTTORANDO / A
Giacomo Roncoroni

COORDINATORE
Stefano Maset

SUPERVISORE DI TESI
Prof. Michele Pipan

CO-SUPERVISORE DI TESI
Prof. Emanuele Forte

ANNO ACCADEMICO 2022/2023

Long-Short-Therm Memory in Active Wavefield Geophysical Methods

Giacomo Roncoroni

XXXVI cycle of PhD program in
Earth Science, Fluid Dynamics, Interaction and Methodology
Department of Mathematics and Geosciences
University of Trieste
Trieste, Italy

January 2024

Supervisor Prof. Michele Pipan
Co-Supervisor Prof. Emanuele Forte

Contents

Abstract

The PhD works discuss the application of Long Short-Term Memory (LSTM) networks in active wavefield geophysical methods. In this work we emphasize the advantages of Deep Learning (DL) techniques in geophysics, such as improved accuracy, handling complex datasets, and reducing subjectivity. The work explores the suitability of LSTM networks compared to Convolutional Neural Networks (CNNs) in some geophysical applications. LSTM networks have the ability to capture temporal dependencies and are well-suited for analyzing geophysical data with non-stationary behavior. They can process both time and frequency domain information, making them valuable for analyzing both Seismic and Ground Penetrating Radar (GPR) data. The PhD work is described in five main chapters covering the methodological development, regression, classification, data fusion, and frequency domain signal processing. In this thesis we introduce the concept of "Deep Attributes", which are derived from the analysis of LSTM's internal Hidden Layers (HLs) demonstrating that they can be used to improve the interpretation of Seismic or GPR data. The stability and robustness of these attributes are evaluated on both synthetic and real data, including time-lapse applications. We discussed the use of LSTM in simulating forward modeling for geophysical surveys, which is crucial for inversion methodologies like Full Waveform Inversion (FWI). While LSTM reduces computational costs, we found severe limitations due to the 1-D layered assumptions necessary for the velocity field. We found the LSTM's effectiveness in classification tasks for reflection events, both binary and multi-class, and highlighted the significance of the training dataset and statistical robustness in these tasks. Applications on real data to objectively extract lunar subsurface structures from GPR Chang'E4 Chinese mission are also presented. Furthermore, the research explores data fusion using LSTM, with a focus on GPR data from the NASA Mars2020 mission, addressing the need for merging different radar datasets to improve interpretation accuracy and depth assessment. Furthermore, we found LSTM's strengths in regression tasks, frequency manipulation, and its potential impact on various geophysical applications, including seismic deconvolution, Normal Move Out (NMO) corrections, and Amplitude Versus Offset (AVO) analysis. In summary, this PhD thesis highlights the power of LSTM networks in active wavefield geophysics, emphasizing their ability to handle complex, time-dependent data and to extract valuable information for interpretation, modeling and inversion.

Extended Abstract

With the surge in big data and computational power in recent years, the application of Machine Learning (ML) to geophysical data has gained popularity. In fact, the data-intensive nature of geophysics makes it a prime candidate for the application of Deep Learning (DL) techniques. The advantages of DL methods in geophysics are multiple: improved accuracy (Kim et al., 2020), capability to handle vast and complex datasets (Caique et al., 2022), reduced subjectiveness in many traditionally labor-intensive tasks (Feng et al., 2022), among the others. However, as with all DL applications, care must be taken in model selection, training, and validation to ensure that the results are both accurate and physically meaningful. Since geophysics is inherently a study of complex, time- and space-dependent processes, a main question arises: Which neural architecture is the most suited to the task?. While Convolutive Neural Networks (CNNs) have seen significant success in image and spatial data processing (e.g. Alzubaidi et al., 2021), Long Short-Term Memory (LSTM) networks offer distinct advantages for certain specific datasets (e.g. Hochreiter and Schmidhuber, 1997). Although CNNs are a common choice over the geophysical community (Yu and Ma, 2021), some recurrent neurons could better represent the physics behind several geophysical applications. In order to provide a complete answer to this question, the current PhD work aims at studying in depth the strength, potentiality and limitations of the use of Recurrent neurons in active wavefield geophysics, widening one of the less studied branches of DL application to the topic.

LSTM networks, i.e. a type of recurrent neural network (RNN), have shown great potential for specific applications: their use in geophysics is starting slowly, due to some severe computational limitations, e.g. Graphical Processing Units (GPUs) are crucial for their training (Zheng et al., 2021), and gradient issues, e.g. gradient explosion and vanishing (Staudemeyer and Morris, 2019). Although this, LSTM can get important advances since it has got the ability to remember patterns over long sequences and to capture temporal dependencies, which are critical when predicting,

for instance, the next seismic event based on the previous wavefield. CNNs, being primarily spatial processors, may miss these intricate temporal patterns especially when introducing a batch division in the dataset. Additionally, geophysical data often exhibits non-stationary behavior (Kjartansson, 1979), where statistical properties change over time: LSTM recurrent nature allows to adapt their neurons to such non-stationarities, making it particularly well-suited for different issues of active wavefield geophysics. Furthermore, geophysical datasets, such as active seismic or electromagnetic Ground Penetrating Radar (GPR) data, have information in both the time and frequency domains: while CNNs can handle spectral data using spectral bands as channels (akin to RGB channels in images), LSTM can process the time domain and then feed into a frequency domain analysis, making the process more cohesive.

Current PhD thesis work is focused on the creation, tests and critical discussions of a wide set of methodologies covering some of the main applications of DL in active wavefield geophysics. The work is divided into 5 main chapters, i.e. 1. Methodological Development, with focus on explainable DL, 2. Regression, 3. Classification, 4. Data Fusion, and 5. Frequency Domain Signal Processing. The main aim of current work is to test LSTM and prove the importance of a conscious choice of the neurons, activation functions and NN geometry. Especially when using a RNN, with the severe constraints of the causality processes, a careful evaluation on the use of Bi-Directional wrapper (Graves and Schmidhuber, 2005) is crucial to gain good (i.e. realistic, stable and robust) results. Moreover, understanding how this neuron works is pivotal to the explainable AI, opening the "black box" and giving statistical explanations to the results. To do this, an extensive analysis of the internal Hidden Layer (HL) of a trained Neural Network based on LSTM has been carried on and presented in Roncoroni et al., submitted A. To better generalize the problem and to focus on the neuron itself, we exploited a Encoder-Decoder (ED) geometry, which aims at reducing the complexity of the transmitted information, i.e. the encoder restore it in the decoder. Since the input and the output are fixed, i.e. are the same traces, the analysis of the HLs of a trained ED-LSTM points out some interesting features of the trace, e.g. frequency changes, noisy areas, phase continuities, and reflection strengths anomalies, among the others. We therefore propose the use of such information as helpful seismic (or GPR) signal attributes (defined as "Deep Attributes") for interpretation purposes (Roncoroni et al., submitted A). The current approach has been evaluated both on synthetic data and on real data. One of the main concerns on this approach was the stability of the derived attributes: in order to prove such stability, we trained the ED-LSTM on a 4-D dataset, the Sleipner seismic dataset (Equinor, 2020) using, as training data, a single 3-D post-stack seismic profile acquired in 2008 and inferring the Deep Attributes, i.e.

the HL features, on the trained ED-LSTM applied to a 3-D post-stack survey acquired later in time in the same area. The robustness of the results has shown promising stability even for time lapse (i.e. 4-D) applications.

To probe the capability of LSTM in the recurrent tasks, we propose a methodology based on a Bi-Directional LSTM architecture to simulate the forward modeling of an active wavefield geophysical acquisition. Since this task is often crucial for inversion methodologies, e.g. Full Waveform Inversion (FWI) (Virieux et Operto, 2009), the reduction of the computing time of the forward modelling and of the memory requirement are the main criticisms of current state of the art techniques. In Roncoroni et al., 2021 we introduced this approach to reduce the computational costs and time of a 2-D forward modelling in a simple 1-D layered media. Although the computational costs are heavily reduced and the methodology can be easily scaled to different acquisition geometry and waveform, a severe limitation lays in the 1-D layered assumption, which can be considered true only in few specific geological settings and do not generalize well the actual complexity of the problem in most real scenarios.

Another common task of the application of DL in geophysics is the classification topic, which has been widely explored especially in the interpretation branch of seismic. We investigated the potential of LSTM both as binary and as multi-class classifiers, focusing on reflection events. Roncoroni et al., 2022A proposed a binary classification of the reflection events on an unbalanced dataset, i.e. 0 for no reflection events and 1 for the presence of a reflection. In this paper we also focused on filling the gap between field data and synthetic ones by proposing a modified version of the convolutive approach of the seismic traces introducing a pre-convolutional noise component that mimics the anisotropy of the geologic materials. The methodology has been applied successfully on both GPR and seismic real data, proving the performances of the approach and its robustness. Starting from the proposed approach, in Roncoroni et al., 2022B we introduced also a multi-class classification, using a simple LSTM-NN to infer the reflection polarity. Since the task is more complex, we implemented a combined use of the real part of the seismic trace (i.e. the recorded data) and of some complex attributes, like the cosine of the instantaneous phase (Chopra and Marfurt, 2005).

Both the approaches had the key strength of the training dataset, generated as a simple convolution of random reflection coefficients series, and in the statistical robustness assured by the 1-D analysis. This is a crucial point of the whole implementation: since

no information on lateral continuity is provided to the NN, the reflection continuity is a direct and not pre-constrained measure of the quality of the results. The methodology developed in this works was successfully tested in Roncoroni et al., submitted B with an application on a complex GPR dataset acquired by the Chinese Chang'E4 mission on the far side of the moon. Thanks to the integration between GPR attributes and the DL implemented Horizon Extraction we were able to provide a new and objective interpretation of several lunar subsurface complex structures.

Another important class of algorithms are the data fusion methods: we tested LSTM on this complex task using GPR data from the NASA Mars2020 mission, Roncoroni et al., 2023a. The need of a fusion methodology directly arises from the peculiar GPR mounted onboard the mission, RIMFAX, which is a gated frequency-modulated continuous-wave instrument. Due to this choice, the mission produced 3 separated radar datasets with different information and time windows driven by the central frequency of the data. Merging this information is crucial to manage the data and for an accurate interpretation both in depth (exploiting lower frequencies), and in the shallower part (exploiting higher frequencies). A further generalization was made by extending the methodology from RIMFAX to classical multi-frequency GPR acquisitions performed using different antennas, Roncoroni et al., 2023b. This latter paper shows results on 3-D synthetic data (Koyan and Tronicke, 2020) and on 2-D field data acquired in Antarctica and extends the methodology to further tests like the fusion of different data holding different information.

Through all these tests, we found out that LSTM has a great power into regressions tasks and that a simple convolutive approach can generalize good both the field seismic and GPR data, without remarkable differences or constraints. Furthermore, the recurrent nature on the neuron is perfect to extrapolate information about the frequency content. Roncoroni et al., Submitted C exploited this ability to extend the frequency spectrum toward low frequencies, resulting into a sort of denoise on low frequency deep seismic data (Audhkhasi and Singh, 2022). Finally, the use of this potential to fill the frequency null space has been proven into the challenging tasks of low frequency extraction from recorded seismic time series (i.e. traces). The methodology proved to gain great results in reconstructing the low frequency counterpart of the data not only in comparison with complex synthetic scenarios (Versteeg, 1994), but also to enlarge the spectrum of recorded field data. The applications of this technique open up new interesting possibilities not only into frequency FWI, as discussed into this thesis work, but also on algorithms like seismic deconvolution using DL, i.e. filling the null space

towards higher frequencies, as well as in Normal Move Out (NMO) corrections for Common Shot Gather (CSG) in order to reduce the stretching and gain information even on large offsets at shallow depth. These applications could be crucial for future advances not only in wavefield geophysical data interpretation, but also in Amplitude Versus Offset (AVO) applications on shallow targets.

In summary, the use of recurrent neurons has shown promising results in different active wavefield applications and for various issues. Although there are some limitations, e.g. the 1-D assumption for the synthetic seismic and the need of GPUs for training, some promising results were achieved especially in terms of reduction of time-consuming tasks like seismic forward modeling. Moreover, the study of HL statistics turned out to be very important not only to understand the behavior of this neuron with wave-based signals, but also for the extraction of several additional and somehow embedded information in a statistical way from a geophysical dataset. The application of LSTM in the classification tasks led to the introduction of a powerful tool that aims at constraining the seismic and GPR interpretation to a more data-driven solution. Furthermore, the fusion of different types of data has unlocked the potential of frequency merging and some other applications are now under tests, like e.g. merging of seismic 3-C recordings and Distributed Acoustic Sensing (DAS) data. The main power of LSTM over classical CNN-based algorithms is the embedded frequency information and this could potentially open a completely new branch in addition to the classical processing techniques, allowing a frequency manipulation that is impossible to achieve using conventional methods.

Acknowledgements

My deepest gratitude to *Michele Pipan*, Supervisor, for not only discussing ideas and providing opportunities for travel and discussions but also for entrusting me with the leadership of a fieldwork in Iraqi Kurdistan. Your unwavering support and belief in my capabilities have been pivotal in my personal and academic growth.

Emanuele Forte, Co-Supervisor, deserves special recognition for engaging in scientific discussions, closely following the progress of my work, and providing meticulous reviews of the thesis. Your expertise and dedication have significantly enhanced the quality of this research.

Nicola Bienati from ENI, your insightful discussions and contributions have been a crucial aspect of this research. Your expertise and willingness to engage in discussions greatly enriched my understanding of the subject matter.

Satish C. Singh from IPGP, Paris, deserves a special mention for hosting me and imparting a wealth of knowledge regarding the scientific approach. Your hospitality and mentorship were pivotal in broadening my scientific horizons.

I am also deeply grateful to *Tor Arne Johansen* at University of Bergen for granting me the remarkable opportunity to join the Svalbard fieldwork, a transformative experience that greatly enriched my research. Additionally, your kind invitation to the Lofoten seminar was an honor, and I am thankful for the insights gained from it.

This thesis could not have been completed without the unwavering support and encouragement of these individuals. I am also grateful to my friends and family for their constant belief in my abilities. Their emotional support sustained me throughout this journey.

Finally, I extend my sincere appreciation to the academic community and institutions that made this research possible. This thesis represents the collective effort of many, and I am deeply thankful for the contributions and collaboration that have made it a reality.

List of Figures

List of Tables

List of Papers

G. Roncoroni, C. Fortini, L. Bortolussi, N. Bienati, M. Pipan, *Synthetic seismic data generation with deep learning*, Journal of Applied Geophysics, Vol. 190, 2021, 104347, doi.org/10.1016/j.jappgeo.2021.104347.

G. Roncoroni, E. Forte, L. Bortolussi, M. Pipan, *Efficient extraction of seismic reflection with Deep Learning*, Computers & Geosciences, Volume 166, 2022, 105190, ISSN 0098-3004, doi.org/10.1016/j.cageo.2022.105190.

G. Roncoroni, E. Forte, L. Bortolussi, L. Gasperini, M. Pipan, *Polarity assessment of reflection seismic data: a Deep Learning approach*, BGO, Vol. 63, 2022, n. 4, 693-700, doi.org/10.4430/bgo00409.

G. Roncoroni, E. Forte, M. Pipan, *Merging gated frequency-modulated continuous-wave Mars2020 RIMFAX ground-penetrating radar data*, GEOPHYSICS, 2023, 88: A7-A12, doi.org/10.1190/geo2022-0466.1.

G. Roncoroni, E. Forte, I. Santin, M. Pipan, *Deep Learning based multi-frequency GPR data merging*, Geophysics, Just-Accepted Articles, doi.org/10.1190/geo2023-0215.1.

G. Roncoroni, E. Forte, M. Pipan, *Deep Attributes: innovative LSTM-based seismic attributes*, submitted to GJI, Under review.

G. Roncoroni, E. Forte, I. Santin, A. Černok, A. Rajšić, A. Frigeri, W. Zhao, M. Pipan, *Deep Learning driven interpretation of Chang'E4 Lunar Penetrating Radar*, submitted to Journal of Geophysical Research: Planets, Under review.

G. Roncoroni, E. Forte, I. Santin, A. Černok, A. Rajšić, A. Frigeri, M. Pipan, *High frequency Lunar Penetrating Radar quality control, editing and processing of Chang'E-4 lunar mission*, submitted to Scientific Data, Under Review.

I. Santin, **G. Roncoroni**, E. Forte, M. Pipan, *GPR modelling and inversion to quantify the debris content within ice.* Near Surface Geophysics, 2023, 1–15. doi.org/10.1002/nsg.12274.

M. Venier, L. Ziberna, L. Mancini, A. Kao, F. Bernardini, **G. Roncoroni**, S. Milani, M. Youbi, M. Yondon, A. De Min, D. Lenaz, *Multi-scale and multi-modal imaging study of mantle xenoliths and petrological implications*, 2023, doi.org/10.2138/am-2022-8866.

List of Extended Abstract

E. Forte, I. Santin, R. R. Colucci, M. Dossi, M. Guglielmin, M. Pipan, **G. Roncoroni**, M. Žebre, *GPR data analysis for cold and warm ice detection and characterization in polythermal glaciers*, SEG Global Meeting Abstracts, 2020: 69-72, doi.org/10.1190/gpr2020-019.1.

G. Roncoroni, M. Dossi, E. Forte, M. Pipan, L. Bortolussi, *Velocity analysis on common offset GPR data: A deep learning approach*, SEG Global Meeting Abstracts, 2020, 388-391, doi.org/10.1190/gpr2020-101.1.

N. Fuji, N. Dhabaria, **G. Roncoroni**, R. Myhill, S. Durand, A. Borgeaud, P. Tackley, T. Nakagawa, F. Deschamps, *Towards waveform seismic filtering of mantle convection models*, EGU22-3554, doi.org/10.5194/egusphere-egu22-3554.

I. Santin, **G. Roncoroni**, E. Forte, M. Pipan, *GPR inversion and modelling for glacial internal debris estimation and characterization, SEG Global Meeting Abstracts*, 2022: 39-42, doi.org/10.1190/gpr2022-064.1.

G. Roncoroni, E. Forte, M. Pipan, *Multi frequency data merging with bi-directional LSTM*, SEG Global Meeting Abstracts, 2022 : 151-154, doi.org/10.1190/gpr2022-069.1.

Open Data

As one of the results of this PhD work we published an open dataset from the Chinese mission on the moon (Chang'E4 Lunar Penetrating Radar (LPR) data, Channel 2). The dataset is published, cleaned from positioning errors and repeated traces, both in raw and processed version and is described in Roncoroni et al., Submitted C and is at the base of Roncoroni et. al., Submitted B.

Data are available in **SEG-Y format**, the standard format for the geophysical community at figshare.com/s/9ce7f1cb8ff0fb8d90c9.

Data are available in **PDS-4 format**, the standard format for the Planetary Science community at figshare.com/s/41cee6210832d1b7a713.

Chapter 1

Overview

Current PhD work aims to study and improve the use of Recurrent Neural Network (RNN), with specific focus on Long Short-Term Memory (LSTM) for active wavefield Geophysics.

1.1 Motivations

With the surge in big data and computational power in recent years, the application of Machine Learning (ML) to geophysical data has gained popularity. The data-intensive nature of geophysics makes it a prime candidate for the application of Deep Learning (DL) techniques.

The advantages of DL methods in geophysics are multiple: improved accuracy Kim et al. (2020) the capability to handle vast and complex datasets Marques et al. (2022), reduced need for manual intervention in many traditionally labor-intensive tasks Feng et al. (2022), among the others. However, as with all applications of DL, care must be taken in model selection, training, and validation to ensure that the results are both accurate and physically meaningful.

Although the clear advantages that applying Machine Learning (ML) methods in active geophysics lead, there are some severe limitations, e.g. the lack of large amount of high-quality labeled data, which can be challenging.
Geophysics is inherently a study of complex, time-dependent processes. As machine learning techniques gain traction in geophysical applications, a question arises: which neural architecture is most suited to the task? While Convolutive Neural Networks

(CNNs) have seen significant success in image and spatial data processing (Alzubaidi et al., 2021), LSTM networks, a type of recurrent neural network (Hochreiter and Schmidhuber, 1997), offer distinct advantages for certain specific datasets.
LSTM networks have shown immense potential for certain applications but their use in geophysics is starting slowly, as shown in Figure 1.1.1; due to some severe computational limitations, i.e. Graphical Processing Units (GPUs) are crucial to make them computationally efficient, but can get important advances.

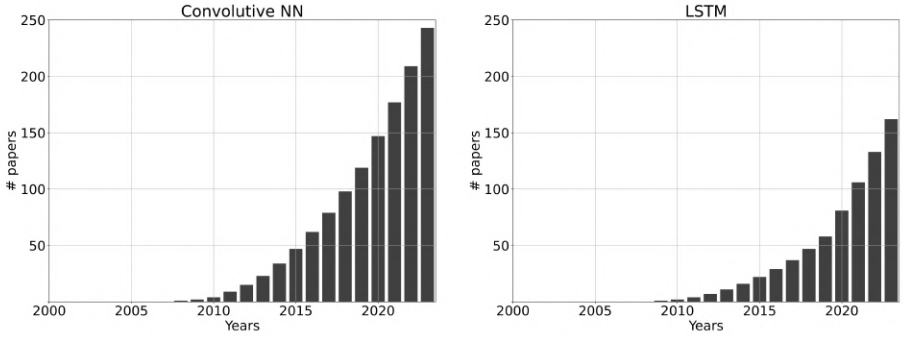

Figure 1.1.1: Comparison between the use of CNN (left side) and LSTM (right side) in active wavefield geophysics according to Journal Citation Reports (JCR)

One of the fundamental differences between LSTM and CNN lies in their inherent design and applications: while CNNs excel at spatial data processing, making them a perfect tool for image recognition tasks, LSTMs are designed to handle sequences. The time-dependent nature of many geophysical datasets - such as for wavefield methods - could make LSTM a more fitting choice.
LSTMs have the ability to remember patterns over long sequences and can capture temporal dependencies, which are critical when predicting, for instance, the next seismic event based on the previous waves behaviour. While CNNs, being primarily spatial processors, e.g. Alzubaidi et al. (2021), may miss these intricate temporal patterns, LSTM's design inherently allows for the backpropagation of errors not just across layers, but also back through time, thanks to its gating mechanisms.
Geophysical data often exhibits non-stationary behavior (Kjartansson, 1979), with statistical properties changing over time. LSTM's recurrent nature allows it to adapt to such non-stationarities, making this tool particularly well-suited for tasks like earthquake characterization Berhich et al. (2023), where patterns might evolve due to changes in tectonic behavior.
Many geophysical datasets, such as seismic or electromagnetic data, can be analyzed in both the time and frequency domains: while CNNs can handle spectral data using

spectral bands as channels (akin to RGB channels in images), LSTM can process the time domain and then feed into a frequency domain analysis, making the process more cohesive.
Furthermore, in geophysics, not all data sequences are of equal length. Some events, like certain seismic recording, may span from one second to one minute, while others, like GPR signals, lasts just some nanoseconds. LSTMs inherently deal with sequences of variable lengths, making them robust choices for the diverse dataset lengths encountered in geophysics.
To conclude, geophysical data is often replete with noise and outliers, often due to environmental factors or instrumental errors. LSTM's sequential nature and its memory capabilities enables them to recognize and even filter out noises, making possible more accurate and meaningful interpretations of geophysical phenomena.

1.2 Chapter List

A comprehensive description of the chapters of this thesis is provided hereafter.

Chapter 2
In this chapter we theoretically introduce the main concepts of DL then focusing on the specific neuron under study: LSTM. After defining the complexities and the problems, we furthermore provide a complete state of the art for the use of LSTM networks in wavefield methods.

Chapter 3
In this chapter, starting from a very simple example, i.e. to implement a LSTM-based NN to perform a convolution, we point out strengths and limitations of using this kind of architectures. We take into consideration not only the causal nature of a Neuron like this type and its implications in predicting maximum- and minimum- phase signals, but also frequencies. Furthermore, at the end of the chapter we present a work based on the understanding the Hidden Layer of a LSTM network and its possible use as seismic attributes, as presented in Roncoroni et al. (tteda): **G. Roncoroni**, E. Forte, M. Pipan, *Deep Attributes: innovative LSTM-based seismic attributes*, submitted to GJI, Under review.

Chapter 4
We here focus on a specific recurrent task, i.e. the generation of synthetic seismic data

with LSTM. Results of this approach are presented in Roncoroni et al. (2021): **G. Roncoroni**, C. Fortini, L. Bortolussi, N. Bienati, M. Pipan, *Synthetic seismic data generation with deep learning*, Journal of Applied Geophysics, Vol. 190, 2021, 104347, doi.org/10.1016/j.jappgeo.2021.104347 .

Chapter 5

The classification tasks are one of the most popular applications of LSTM. We performed a binary classification of seismic and GPR reflections in Roncoroni et al. (2022a): **G. Roncoroni**, E. Forte, L. Bortolussi, M. Pipan, *Efficient extraction of seismic reflection with Deep Learning*, Computers & Geosciences, Volume 166, 2022, 105190, ISSN 0098-3004, doi.org/10.1016/j.cageo.2022.105190. .

On the other hand Roncoroni et al. (2022b) presents a multi class classification to define the polarity of the reflections: **G. Roncoroni**, E. Forte, L. Bortolussi, L. Gasperini, M. Pipan, *Polarity assessment of reflection seismic data: a Deep Learning approach*, BGO, Vol. 63, 2022, n. 4, 693-700, doi.org/10.4430/bgo00409 .

Starting from the methodology presented before, we applied the algorithm on some GPR data from Chinese mission Chang'E4, Roncoroni et al. (ttedb), increasing the objectiveness of the interpretation (Roncoroni et al., ttedc): **G. Roncoroni**, E. Forte, I. Santin, A. Černok, A. Rajšić, A. Frigeri, W. Zhao, M. Pipan, *Deep Learning driven interpretation of Chang'E4 Lunar Penetrating Radar*, submitted to Journal of Geophysical Research: Planets, Under review.

Chapter 6

Another complex task that could fit well the strengths of LSTM is data fusion: this class of algorithms gains importance with the increasing amount of available data which becomes crucial to exploit properly the different resolution dataset.

Roncoroni et al. (2023a) propose this approach to NASA mission Mars2020 in **G. Roncoroni**, E. Forte, M. Pipan, *Merging gated frequency-modulated continuous-wave Mars2020 RIMFAX ground-penetrating radar data*, GEOPHYSICS, 2023, 88: A7-A12, doi.org/10.1190/geo2022-0466.1 .

Due to the wide applicability of this approach, Roncoroni et al. (2023b) extended the methodology with further tests to classical GPR instruments digging deeper into the theory in **G. Roncoroni**, E. Forte, I. Santin, M. Pipan, *Deep Learning based multi-frequency GPR data merging*, submitted to Geophysics, accepted.

Chapter 7

One of the main advantages of the use of LSTM is the capability to deal with

frequencies, as we did in Chapter 6 for data fusion. We here exploited this potential for imaging purposes: we present the use of this methodology for real data imaging. Furthermore, an overview on the current work is introduced in this chapter with focus on two articles in preparation: one to enhancement low frequency in a processed post-stack data for deep seismic imaging and the other one is a three-step approach to fill the low frequency null space gap on raw CSG.

Chapter 8

An overview of the work done during this work and a way forward for LSTM neurons is presented in this chapter, with a critical analysis of its future perspectives.

Chapter 9

Although the main focus of the project was to understand the use of LSTM in geophysical WaveField methods, some other works were jointly performed within the PhD program. Here we point out some of them.

Bibliography

Laith Alzubaidi, Jinglan Zhang, Amjad J. Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, J. Santamaría, Mohammed A. Fadhel, Muthana Al-Amidie, and Laith Farhan. 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data* 8, 1 (March 2021), 53. https://doi.org/10.1186/s40537-021-00444-8

Asmae Berhich, Fatima-Zahra Belouadha, and Mohammed Issam Kabbaj. 2023. An attention-based LSTM network for large earthquake prediction. *Soil Dynamics and Earthquake Engineering* 165 (2023), 107663. https://doi.org/10.1016/j.soildyn.2022.107663

Tian Feng, Saeed Mohanna, and Lingsen Meng. 2022. EdgePhase: A Deep Learning Model for Multi-Station Seismic Phase Picking. *Geochemistry, Geophysics, Geosystems* 23, 11 (2022), e2022GC010453. https://doi.org/10.1029/2022GC010453 _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2022GC010453.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Computation* 9, 8 (Nov. 1997), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735 _eprint: https://direct.mit.edu/neco/article-pdf/9/8/1735/813796/neco.1997.9.8.1735.pdf.

Taeyong Kim, Junho Song, and Oh-Sung Kwon. 2020. Probabilistic evaluation of seismic responses using deep learning method. *Structural Safety* 84 (May 2020), 101913. https://doi.org/10.1016/j.strusafe.2019.101913

Einar Kjartansson. 1979. Constant Q-wave propagation and attenuation. *Journal of Geophysical Research: Solid Earth* 84, B9 (1979), 4737–4748. https://doi.org/10.1029/JB084iB09p04737 _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/JB084iB09p04737.

Caique Rodrigues Marques, Vinicius Guedes Dos Santos, Rafael Lunelli, Mauro Roisenberg, and

Bruno Barbosa Rodrigues. 2022. Analysis of Deep Learning Neural Networks for Seismic Impedance Inversion: A Benchmark Study. *Energies* 15, 20 (Oct. 2022), 7452. https://doi.org/10.3390/en15207452

Giacomo Roncoroni, Emanuele Forte, Luca Bortolussi, Luca Gasperini, and Michele Pipan. 2022b. Polarity assessment of reflection seismic data: a Deep Learning approach. *BGO* 63 (Nov. 2022), 693–700. https://doi.org/10.4430/bgo00409

Giacomo Roncoroni, Emanuele Forte, Luca Bortolussi, and Michele Pipan. 2022a. Efficient extraction of seismic reflection with Deep Learning. *Computers & Geosciences* 166 (Sept. 2022), 105190. https://doi.org/10.1016/j.cageo.2022.105190

Giacomo Roncoroni, Emanuele Forte, and Michele Pipan. 2023a. Merging gated frequency-modulated continuous-wave Mars2020 RIMFAX ground-penetrating radar data. *GEOPHYSICS* 88, 2 (2023), A7–A12. https://doi.org/10.1190/geo2022-0466.1 arXiv:https://doi.org/10.1190/geo2022-0466.1

Giacomo Roncoroni, Emanuele Forte, and Michele Pipan. Under reviewa. Deep Attributes: innovative LSTM-based seismic attributes. (Under review). Submitted to GJI.

Giacomo Roncoroni, Emanuele Forte, Ilaria Santin, Ana Cernok, Andrea Rajsic, Alessandro Frigeri, and Michele Pipan. Under reviewb. High frequency Lunar Penetrating Radar quality control, editing and processing of Chang'E-4 lunar mission. (Under review). Submitted to Scientific Data.

Giacomo Roncoroni, Emanuele Forte, Ilaria Santin, Ana Cernok, Andrea Rajsic, Alessandro Frigeri, Wenke Zhao, and Michele Pipan. Under reviewc. Deep Learning driven interpretation of Chang'E4 Lunar Penetrating Radar. (Under review). Submitted to Journal of Geophysical Research: Planets.

Giacomo Roncoroni, Emanuele Forte, Ilaria Santin, and Michele Pipan. 2023b. Deep Learning based multi-frequency GPR data merging;. *GEOPHYSICS* 0, ja (2023), 1–35. https://doi.org/10.1190/geo2023-0215.1 arXiv:https://doi.org/10.1190/geo2023-0215.1

Giacomo Roncoroni, Carlo Fortini, Luca Bortolussi, Nicola Bienati, and Michele Pipan. 2021. Synthetic seismic data generation with deep learning. *Journal of Applied Geophysics* 190 (July 2021), 104347. https://doi.org/10.1016/j.jappgeo.2021.104347

Chapter 2

Introduction

2.1 Neural Network

2.1.1 What is a Neural Network

Neural networks (NN) Nielsen (2018) form the foundation of many modern Machine Learning and Artificial Intelligence systems. They are a broad class of algorithms that can capture intricate statistical patterns in input data.
The fundamental unit of a neural network is the neuron: it receives inputs, processes them using a weighted sum followed by a non-linear function (typically an activation function), and produces an output.
Functions like the sigmoid, hyperbolic tangent (tanh), and rectified linear unit (ReLU) introduce non-linearity to the network, enabling it to capture complex relationship.
Neurons are organized into layers: input layer, hidden layers, and output layer. The input layer receives data, hidden layers process it, and the output layer provides the final output.
Each connection between neurons has an associated weight. During training, these weights are adjusted to minimize the difference between the predicted and actual outputs. Biases are additional parameters that allow neurons to shift their output.
In order to minimize the error between the actual and predicted outputs, in a supervised learning algorithm, we use backpropagation Rumelhart et al. (1986). It calculates the gradient of the error function with respect to the weights of the network for a given input-output pair and adjusts the weights accordingly.
The aim is to minimize a function, i.e. the Loss Function, which measures the discrepancy between the predicted and actual outputs.

In the years, different types of Neural Networks have developed, main type of neurons have been divided in:

- **Feed-forward Neural Networks** (FNN) Bebis and Georgiopoulos (1994): The simplest type where information flows in one direction, from input to output, for each hidden layer every neurons of the previous layer is connected to every neuron on the second: often referred to as fully connected NN.
- **Convolutional Neural Networks** (CNN) Lecun and Bengio (1995): Primarily used for image processing, they have convolutional layers that can capture spatial hierarchies in data.
- **Recurrent Neural Networks** (RNN) Sherstinsky (2018a): Used for sequence data like time series or natural language, they can capture temporal dependencies.

When neural networks have many hidden layers, they are defined as Deep Neural Networks (DNN) Nielsen (2018). DNN can capture extremely intricate patterns in large datasets but require more data and computational power.

Specifically, in geophysical applications, some of the state of the art applications of DL and Geophysics are:

- Seismic Interpretation with Deep learning algorithms. (CNNs) have been typically applied to improve the accuracy and speed of seismic interpretation, identifying subsurface features such as faults Wu et al. (2019), horizons Roncoroni et al. (2022a), and salt bodies Muller et al. (2023), while reducing the overall subjectivity of the process.

- Prediction of reservoir properties such as porosity, permeability, and fluids saturation Zhang et al. (2022a).

- ML methods to assist in the automatic picking of seismic arrival times, crucial for earthquake location and tomographic imaging Feng et al. (2022a).
- Reduce computational requirements in both Electromagnetic (EM) Dai et al. (2022) and seismic Finite Difference simulations. Moseley et al. (2019), Roncoroni et al. (2021)

2.1.2 Neural Network tasks

Commonly, NN applications can be summarized in regression and classification tasks (Nielsen, 2018; Perotin et al., 2019), an overview on the two differences is presented in Figure 2.1.1.

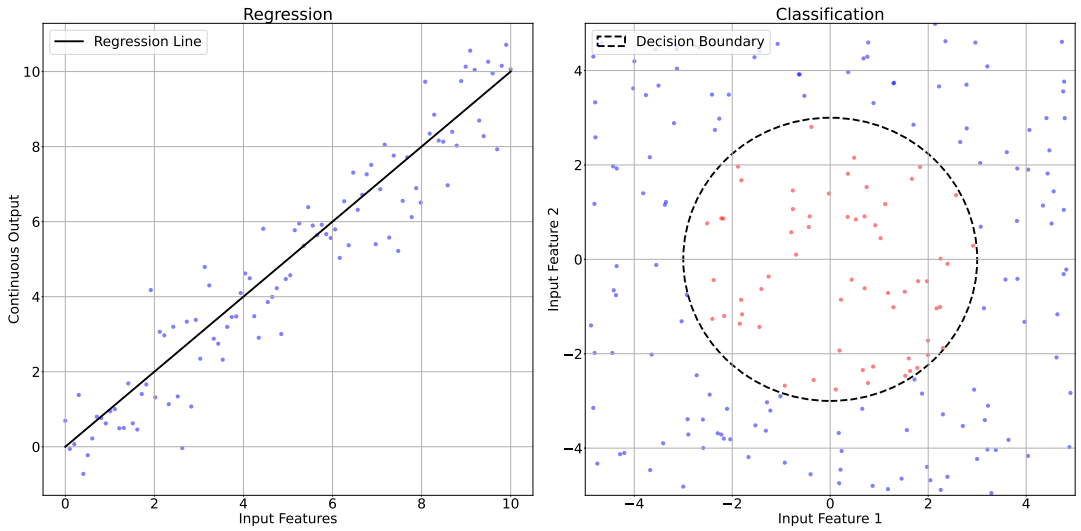

Figure 2.1.1: Description of the two main classes of NN algorithms: regression (left) and classification (right).

First class of algorithm is regression (Nielsen, 2018), which aims to predict a continuous value based on input features, e.g. predicting house prices based on features like square footage, number of bedrooms, location, etc.
The other main task is classification (Zhang, 2000), on the other hand, categorize data into n different classes, e.g. email spam detection (two classes: spam or not-spam).

2.2 Neural Network ingredients

Training a neural network (NN) is a multi-step process (Nielsen, 2018) that involves preparing data, designing the network architecture, specifying training parameters, and then using data to iteratively adjust the network's weights.

We can summarize the whole process in some pivotal steps:

- **Data Preparation**:
  - Data Collection: Gathering data relevant to the problem
  - Data Pre-Processing: Normalization/Standardization, i.e. Scale features to a small range; data Augmentation, i.e. artificially increase dataset's size by creating modified versions of the existing data, used especially in tasks like image recognition.
  - Dataset Split, i.e. Divide the dataset into: Training dataset, Validation dataset, Test dataset.

- **Model Design**:
  - Architecture Selection: Decide on the number of layers and neurons in each layer based on the complexity of the problem.
  - Activation Functions: Choose activation functions for each layer, e.g., ReLU (Rectified Linear Unit) for hidden layers and Softmax for multi-class classification output layers.
  - Weight Initialization: Initialize the weights, often using methods like Xavier or He initialization to ensure that activations don't become too small or too large in the initial stages of training.

- **Training Configuration**:
  - Loss Function: Select a loss function that quantifies the difference between the predicted and actual values.
  - Optimizer: Choose an algorithm to adjust the weights of the network based on the loss gradient.
  - Learning Rate: Set a step size for weight updates. This parameter may require tuning, and sometimes learning rate schedules or adaptive learning rates are used.
  - Regularization: Implement techniques to prevent overfitting, such as dropout, L1/L2 regularization, or early stopping.

  - Batch Size: Decide on the number of samples that will be used to compute the gradient in each iteration.

- **Training Process**:
  - Forward Propagation: For each input in the training set, compute the predicted output by passing it through the network layers.
  - Loss Computation: Compute the loss by comparing the predicted output to the actual output.
  - Backward Propagation: Use the chain rule to calculate the gradient of the loss concerning each weight in the network.
  - Weight Update: Adjust the weights in the direction that minimizes the loss, using the selected optimization method.

- **Evaluation and Testing**:
  - Validation Performance: Monitor performance on the validation set during training to tune hyperparameters and prevent overfitting.
  - Testing: After training is complete, evaluate the model's performance on the test set to get an estimate of its performance on unseen data.

2.2.1 Neurons

At the heart of every neural network there are *Neurons* (Minsky and Papert, 1969; Nielsen, 2018). Each neuron receives one or more inputs, processes them, and produces an output. A basic Neuron performs a weighted sum of its inputs, adds a bias term, and then applies an activation function.
Given inputs $[x_1,\ x_2,\ \ldots,\ x_n]$, weights $[w_1,\ w_2,\ \ldots,\ w_n]$, and a bias b: The linear transformation is defined as:

$$z = w_1 \cdot x_1 \ + \ w_2 \cdot x_2 \ + \ \ldots \ + \ w_n \cdot x_n \ + \ b \tag{2.1}$$

and the activation function f is applied to output z:

$$y = f(z) \tag{2.2}$$

This can be represented in matrix form as:

$$y = f(w^T \cdot x + b) \tag{2.3}$$

where w is the weight vector and x is the input vector. The activation function, f, defines the output of a neuron given its input. It introduces non-linearity into the network, which allow to deal with non-linear relationship.

2.2.2 Optimizers

Optimizers (Abdulkadirov et al., 2023) are algorithms used to update and adjust model parameters (usually the weights and biases of neurons) in response to the output of the loss function during neural network training. The goal is to effectively minimize the loss function, improving the model's accuracy.
The most fundamental optimization algorithm is the Gradient Descent. It updates parameters in the negative direction of the gradient according to:

$$\mathbf{w} = \mathbf{w} - \eta \nabla J(\mathbf{w}) \tag{2.4}$$

where w represents the parameters, η is the learning rate and $\nabla J(w)$ is the gradient of the loss function J with respect to the parameters w.

Instead of computing the true gradient using the entire dataset (as in vanilla GD), Stochastic Gradient Descent (SGD), Ruder (2016), replaces the actual gradient calculated from the entire data set, by a one calculated over a randomly selected subset of the data. This method introduces a momentum term to dampen oscillations and speed up convergence. The update rule becomes:

$$\begin{aligned} v &= \beta v - \eta \nabla J(\mathbf{w}) \\ \mathbf{w} &= \mathbf{w} + v \end{aligned} \tag{2.5}$$

where v is the velocity term and β is the momentum coefficient.

Further development was introduced with Adagrad (Duchi et al., 2011), which adapts the learning rate for each parameter based on the historical gradient information:

$$\mathbf{w} = \mathbf{w} - \frac{\eta}{\sqrt{G + \epsilon}} \odot \nabla J(\mathbf{W}) \tag{2.6}$$

Where G is a diagonal matrix where each diagonal entry i, i is the sum of the squares

of the gradients with respect to w_i up to the current iteration, and ϵ is a small smoothing term to prevent division by zero and $\odot$ is the element-wise multiplication.

The main optimizer used in this work was Adam (Adaptive Moment Estimation), a combination of the previous presented optimizer (Kingma and Ba, 2015):

$$\begin{aligned} m =& \beta_1 m + (1-\beta_1)\nabla J(\mathbf{w}) \\ v =& \beta_2 v + (1-\beta_2)(\nabla J(\mathbf{w}))^2 \\ \mathbf{w} =& \mathbf{w} - \frac{\eta}{\sqrt{v+\epsilon}} m \end{aligned} \tag{2.7}$$

where m is the estimate of the first moment (mean) and v is the estimate of the second moment (uncentered variance).

2.2.3 Loss

The Loss Function is a crucial component, for DL methodologies. It quantifies how well the predictions of a model fits the actual data. During the training phase, the goal is to minimize the value of the loss function.

Mean Squared Error (MSE) is a commonly used loss for regression problems. It measures the average squared difference between the estimated and the actual values. It is defined as:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{2.8}$$

where: y_i is the actual value, $\hat{y}_i$ is the predicted value and N is the number of samples.

Cross-Entropy Loss (Murphy, 2022) is used for binary and multi-class classification problems. It quantifies the difference between two probability distributions, for binary classification is called binary cross entropy, while for multi-class classification, with C classes, it is called Categorical Cross Entropy.
Categorical Cross-Entropy, used in multi-class classification problems is defined as:

$$Categorical\ Cross\ Entropy = -\sum_{c=1}^{C} y_c \log(\hat{y}_c) \tag{2.9}$$

where: C is the number of classes, y_c is the ground truth probability for class c (usually 1 for the correct class and 0 for others) and $\hat{y}_c$ is the predicted probability for class c.

2.3 Recurrent Neural Networks

Recurrent Neural Networks (RNNs) and their more advanced counterparts, like Long Short-Term Memory networks (LSTMs) and Gated Recurrent Units (GRUs), are particularly well-suited for sequence-based data. While the primary domains of application for RNNs have been time series forecasting, natural language processing, and speech recognition. Their potential applications in geophysics, especially in time-dependent or sequential data scenarios, are being explored.

2.3.1 General description

A Recurrent Neural Network (RNN) (Sherstinsky, 2018b; Staudemeyer and Morris, 2019) is a type of artificial neural network designed for processing sequences of data. Unlike traditional feedforward neural networks that process inputs in a single pass, RNNs have a feedback loop that allows information to be passed from one step in the sequence to the next, as shown in Figure 2.3.1. This loop enables RNNs to maintain a form of memory or context, making them particularly useful for tasks involving sequential data like time series, natural language, speech.

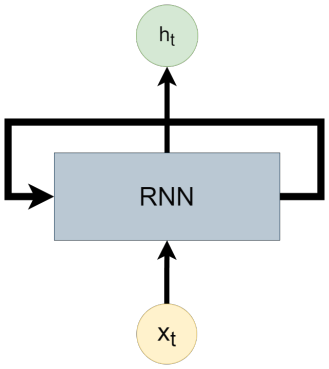

Figure 2.3.1: Schematic sketch of a single Recurrent Neuron that loops over time.

The key feature of an RNN is its ability to take the output from a previous step and use it as an input for the current step, effectively allowing the network to incorporate information from previous steps into its current decision-making process. This process of incorporating previous information makes RNNs suitable for tasks where the order and context of the data matter.
However, traditional RNNs have some limitations, especially when they come to capturing long-range dependencies in sequences. In fact, they suffer from the "vanishing gradient" problem Staudemeyer and Morris (2019), which makes it difficult for them

to retain and propagate information over long distances in the sequence.
To address this limitation, various advanced RNN architectures have been developed. One of them is the Long Short-Term Memory (LSTM) network, which includes specialized mechanisms to control the flow of information and mitigate the vanishing gradient problem, Figure 2.3.2. Another architecture is the Gated Recurrent Unit (GRU), which is a simplified version of LSTM that also addresses the gradient vanishing problem while requiring fewer parameters.

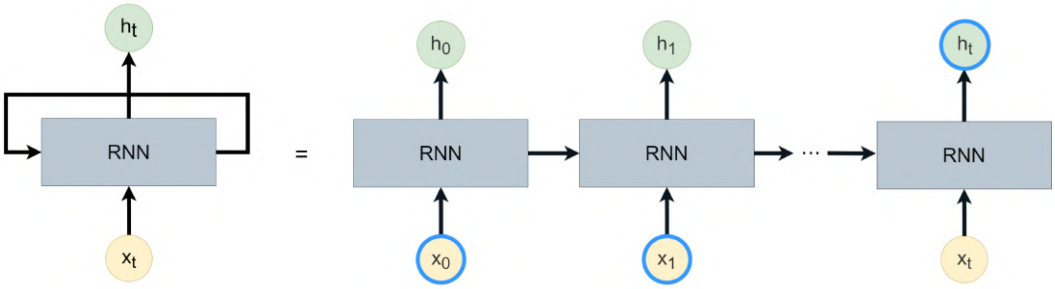

Figure 2.3.2: Unrolling the time loop we can see the memory concept: output h_t gets information from all the previous inputs, depicted from x_0 and x_1.

At its core, a RNN processes sequences of data by introducing a hidden state that is updated at each time step. The hidden state at time step t is denoted as h_t, and it is calculated using the current input x_t and the previous hidden state h_{t1} along with some learnable parameters W and biases b:

$$h_t = f(W \cdot [h_{t1}, x_t] + b) \tag{2.10}$$

Here, f is the activation function like the hyperbolic tangent (tanh) or Rectified Linear Unit (ReLu), w represents the weight matrix, and $[h_{t1},\ x_t]$ is the concatenation of the previous hidden state and the current input.

2.3.2 LSTM neuron

In LSTM (Staudemeyer and Morris, 2019), the hidden state is split into two parts: the cell state c_t and the hidden state h_t. LSTM introduces three gating mechanisms that control the flow of information: the forget gate, input gate, and output gate. A schematic representation of the neuron is depicted in Figure 2.3.3

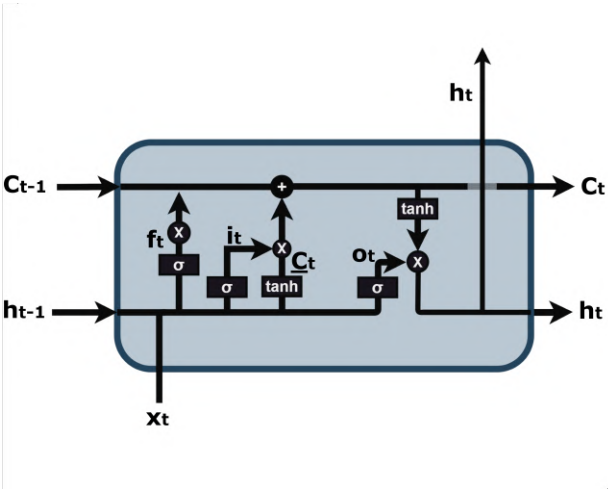

Figure 2.3.3: General representation of a single LSTM neuron at a single time step t

.

The key point is the cell state c_t, depicted in Figure 2.3.4 , which gives to LSTM the capability to selectively add or remove information from its cell state, a process meticulously managed by specialized structures known as gates.

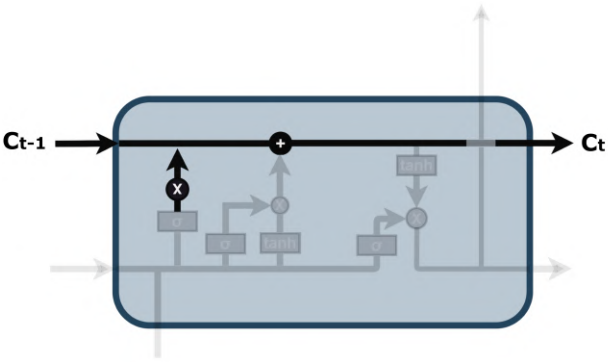

Figure 2.3.4: Schematic representation of the cell state line.

Gates serve as a mechanism to choose whether or not to permit information to flow through. They are constructed using a sigmoid neural network layer in combination with a pointwise multiplication operation.
The sigmoid layer generates values within the range of zero to one, indicating the extent to which each component should be allowed to pass through. A value of zero means "block everything," whereas a value of one means "allow everything". Within a

LSTM, there are three of these gates, which serve to safeguard and regulate the cell state and are: the forget gate, the input gate, and the output gate.
The initial step within LSTM involves determining which information must be discarded from the cell state; this decision-making process is carried out by a sigmoid layer referred to as the "forget gate layer", Figure 2.3.5 . This layer examines both h_{t-1} and x_t and produces a numerical output ranging from 0 to 1 for each element within the cell state C_{t-1}. We can formalize the forget gate process as:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{2.11}$$

where f_t is the part of information to be a forgotten, W_f is the actual weight of the NN, h_{t-1} is the previous timestep hidden state while x_t is the actual input.

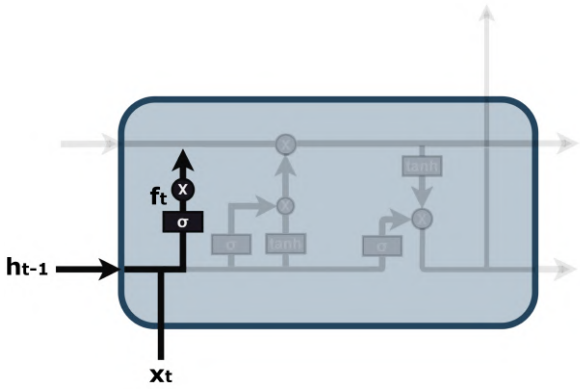

Figure 2.3.5: Schematic representation of the forget gate.

The input gate decides what new information to store in the cell state c_t, depicted in Figure 2.3.6. Similar to the forget gate, it takes h_{t-1} and x_t as inputs and produces a value between 0 and 1 for each element in the cell state. It also produces a candidate cell state (c_t) that will be added to the cell state.

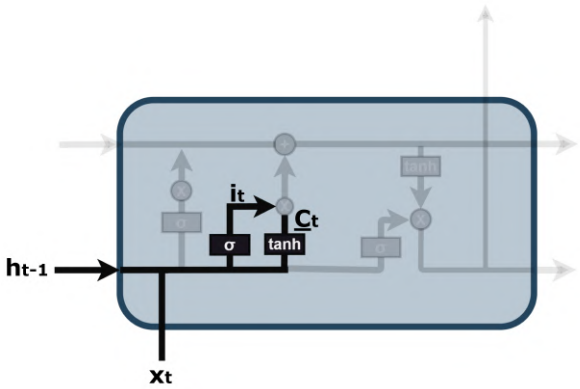

Figure 2.3.6: Schematic representation of the input gate.

The main target of the input gate is to determine which new information to incorporate into the cell state. This process comprises two components:

$$i_t = \sigma(W_t \cdot [h_{t-1}, x_t] + b_i) \tag{2.12}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{2.13}$$

where W_t and W_C are the weights, h_{t-1} is the previous timestep hidden state while x_t is the actual input.
Initially, a sigmoid layer referred to as the "input gate layer", denoted as i_t determines which values have to be modified or updated. Following this, a hyperbolic tangent (tanh) layer generates a vector of potential new values denoted as $\tilde{C}_t$, which could be appended to the current state.

Once we have computed forget and input gates, the cell state C_t can updated by combining the information selected by the forget gate and the information from the input gate's candidate cell state, Figure 2.3.7.

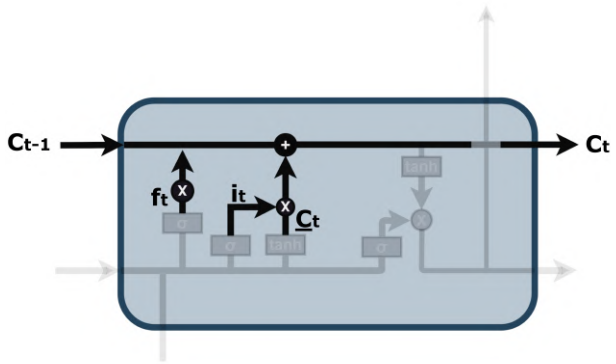

Figure 2.3.7: Schematic representation of the update of the cell state.

We apply this transformation to the former cell state, C_{t-1}, into the updated cell state, C_t.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{2.14}$$

where the first term removes the elements we previously elected to forget from the previous cell state $C_{t-1}$and the product of i_t and $\tilde{C}_t$ represents the new candidate values scaled by the extent to which each element of the state has to be modified.

Lastly, LSTM determine the information we intend to output. This output will be derived from our cell state, but will represent a refined version of it, Figure 2.3.8.

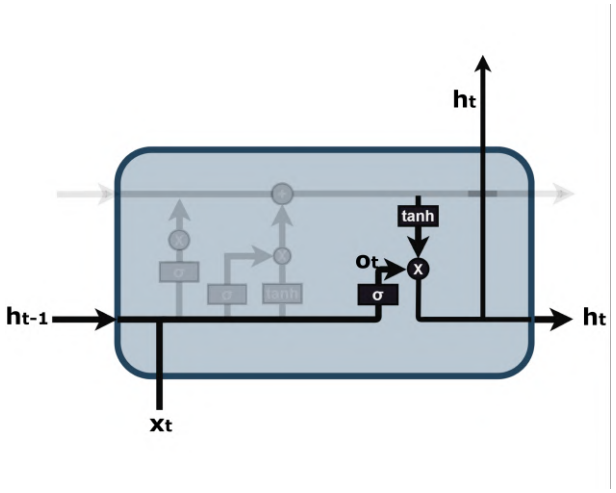

Figure 2.3.8: Schematic representation of the output gate.

The process involves two key steps: first, we apply a sigmoid layer, which makes decisions about which components of the hidden state cell state have to be included in the output and we pass the cell state through a hyperbolic tangent (tanh) layer. We then multiply the output of the sigmoid gate by the result of the tanh layer. This multiplication ensures that we only output the portions of the cell state that we have chosen to retain.

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{2.15}$$

$$h_t = o_t * \tanh(C_t) \tag{2.16}$$

2.3.3 LSTM modifications

What has been elucidated thus far pertains to a conventional Long Short-Term Memory (LSTM) network. However, it is essential to recognize that the LSTM landscape is characterized by a rich tapestry of various modifications.

One salient variant in the realm of LSTMs, as first proposed by Gers and Schmidhuber (2000), incorporates "peephole connections." This innovation entails permitting gate layers to incorporate information from the cell state, thus enabling a more intricate information flow. Although our earlier diagram depicted the inclusion of peepholes across all gates, many scholarly works selectively introduce peepholes for specific gates, rendering the model more adaptable to specific contexts.

Another departure from the conventional LSTM architecture involves the adoption of coupled forget and input gates. Instead of independently deciding what information to retain and what to introduce, these decisions are intertwined. The concept revolves around the concept that forgetting takes place only when fresh information is introduced, and new values are integrated into the state only when antiquated information is erased. This approach aligns the memory management process more closely with the information introduction, thus facilitating enhanced information flow control.

A more substantial departure from the LSTM framework is embedded in the Gated Recurrent Unit (GRU), an innovation introduced by Chung et al. (2014). The GRU consolidates the forget and input gates into a singular "update gate," thereby simplifying the model. This design modification also combines the cell state and the hidden state, resulting in an architectural refinement that has witnessed an upsurge in popularity over time.

The catalog of LSTM variants is far more extensive than the aforementioned examples. It includes models like Depth Gated Recurrent Neural Networks (RNNs) by Yao et al. (2015) and divergent strategies to address long-term dependencies, such as Clockwork RNNs proposed by Koutník et al. (2014).

The critical inquiry pertains to the optimality of these variant models and the significance of their disparities. In this regard, Greff et al. (2015) undertook a comprehensive comparative analysis of popular LSTM variants, ultimately discerning their relative performance parity. Jozefowicz et al. (2015), in a deep analysis, assessed over ten thousand RNN architectures and uncovered certain models that exhibited superior performance compared to LSTMs in specific task domains.

2.3.4 LSTM problems and solutions

Standard Recurrent Neural Networks (RNNs) face a limitation in handling sequences longer than approximately 5 to 10 time steps, as described in Gers et al. (1999). This limitation arises from the tendency of back-propagated error signals to either grow exponentially or diminish significantly with each time step. Over an extended sequence, this behavior causes the error to either blow up or vanish, making learning challenging or even impossible (Hochreiter and Schmidhuber, 1996; Bengio et al., 1994).
To understand this problem more thoroughly, let's examine the back-propagation algorithm and the vanishing error analysis. When updating the weights of an RNN after it has processed data from time t0 to time t, the weight update formula is given by:

$$\triangle W_{[u,v]} = -\eta \frac{\partial E_{total}(t',t)}{\partial W_{[u,v]}} \tag{2.17}$$

where, $\frac{\partial E_{total}(t',t)}{\partial W_{[u,v]}}$ represents the gradient, and η is the learning rate. The back-propagated error signal at time t (with $t_0 < t$) of the unit u is:

$$\vartheta_u(\tau) = f'_u(z_u(\tau)) \left(\sum_{v \in U} W_{vu} \vartheta_v(\tau + 1) \right) \tag{2.18}$$

where τ is the time unit, $W_{v,u}$ is the weight that connects the unit v to the unit u and z_u is the weighted input of the unit u.
Therefore, in the context of a fully recurrent neural network featuring a collection of non-input units U, the error signal occurring at a specific output-layer neuron $o \in O$ (being O the set of output units at time-step t) is retroactively propagated through time for a duration equal to $t - t_0$ time-steps (where $t_0 ¡ t$), ultimately influencing an arbitrary neuron v. This process results in the error being modified by the subsequent scaling factor:

$$\frac{\partial \vartheta_v(t')}{\partial \vartheta_o(t)} = \begin{cases} f'_v(z_v(t')) W_{[o,v]} & \text{if } t - t' = 1, \\ f'_v(z_v(t')) \left(\sum_{u \in U} \frac{\partial \vartheta_u(t'+1)}{\partial \vartheta_o(t)} W_{[u,v]} \right) & \text{if } t - t' > 1 \end{cases} \tag{2.19}$$

To solve this equation, we need to unroll it over time. For $t' \leq \tau \leq t$, consider u_τ as a non-input-layer neuron in one of the replicated networks at time τ. By setting $u_t = v$ and $u_{t'} = o$, Staudemeyer and Morris (2019) derives the equation:

$$\frac{\partial \vartheta_v(t')}{\partial \vartheta_o(t)} = \sum_{u_{t'} \in U} \cdots \sum_{u_{t-1} \in U} \left(\prod_{\tau=t'+1}^{t} f'_{u_\tau}(z_{u_\tau}(t - \tau + t')) W_{u_\tau, u_{\tau-1}} \right) \tag{2.20}$$

From this equation, if the magnitude of the term $f'_{u_\tau}(z_{u_\tau}(t - \tau + t'))W_{u_\tau,u_{\tau-1}}$ is consistently greater than 1 for all τ, the product grows exponentially, leading to errors blow-up. Conversely, if this term is consistently less than 1 for all τ, the product decreases exponentially, causing the error to vanish, hindering effective learning within a reasonable time frame.

A more comprehensive theoretical analysis of the challenges with long-term dependencies in RNNs is discussed in Kremer and Kolen (2001), which also briefly introduces potential solutions to address this problem.

According to Gers et al. (2003), Long Short-Term Memory (LSTM) networks are particularly effective in tasks where a small amount of information needs to be retained over an extended period. This remarkable ability is attributed to the use of memory blocks within LSTMs. Memory blocks are unique structures that offer control over the flow of information through input and output gates, preventing irrelevant data from entering or leaving the memory block. Additionally, memory blocks feature forget gates that selectively weight the information stored within the cell. When previous information becomes obsolete, these forget gates can reset the state of individual cells inside the block. The presence of forget gates also enables the continuous prediction of data, as they can help cells entirely forget their previous states, preventing any biases in future predictions.

However, like other neural network algorithms, LSTMs require the network's architecture to be predefined, with the number of memory blocks remaining constant throughout the network's operation. Consequently, the memory capacity of the network is ultimately limited. Reference Gers et al. (2003) also points out that this limitation is unlikely to be resolved by simply increasing the network's size uniformly and suggests that modularization can facilitate more effective learning.

LSTM offers a broad range of variants and topologies to suit different applications. In recent years, the foundational LSTM method has often been referred to as "vanilla" LSTM and is commonly used with various extensions and modifications.

Conventional Recurrent Neural Networks (RNNs) typically analyze data in a unidirectional manner, considering only the past information during processing. However, the work presented in reference Graves and Schmidhuber (2005) explores the concept

of bidirectional analysis in the context of LSTM networks. In fact, in a bidirectional LSTM, data is processed in two directions: both forward and backward through two separate LSTM networks. These two networks share the same output layer, and this bidirectional approach is especially useful for tasks like phoneme classification, as noted in Graves and Schmidhuber (2005).
One significant advantage of bidirectional LSTM is that it eliminates the one-step truncation limitation that was present in traditional LSTM networks. Instead, it implements a complete error gradient calculation, making it more capable of handling sequences effectively. This approach simplifies the implementation of bidirectional LSTM and allows it to be trained using the standard back-propagation Through Time (BPTT) algorithm.

2.4 The 1-D approach

Most of the approaches based on CNN exploit the 2-D or even 3-D nature of the data; on the contrary, LSTM is a purely 1-D approach, except for some modifications, e.g. grid-LSTM Greff et al. (2015).
This thesis work aims not to overtake this limitation, but to exploit it. Since the classical DL approach is not based on the physics of the phenomena but on statistical relationship inside the data, we want to test if further statistics could be exploited by considering a pure 1-D approximation.
If we take into consideration the waves propagation inside a medium, a 2-D assumption should be done to take into consideration the physics of the wave propagation. In Roncoroni et al. (2021) we had to simplify the actual relationship to obtain a problem. We therefore decided to work under the assumption of a pure 1-D velocity model: this is a clear limitation introduced by the nature of the neuron.
On the other hand, Roncoroni et al. (2022a) and Roncoroni et al. (2022b) focus on the specific task of extracting events from a seismic recording and the 1-D approach exploited in the two articles gives more robustness to the method without simplifying the problem. If we would lay under a pure 2-D assumption, we would probably increase the extracted horizon continuity but we should be very careful about the obtained results, possibly missing reflections in very low Signal-to-Noise ratio data or introducing some untrue and spurious correlations. Since we forget about lateral continuity and we rather focus on the information held by a single trace, we can make further evaluations on the performances achieved by the method by looking for the lateral continuity of the reflectors which is totally independent by the extraction procedure since, as pointed out, it is completely 1-D. Furthermore, in Roncoroni et al. (tted), the geological setting itself was not fitting a pure a 2-D approach, since the

reflection events in the data were not due to laterally-continuous changes, but in most of the cases by local scattering and diffraction points. In that setting, the look for lateral continuity would possibly have laid to incorrect results simulating lateral continuities which no longer exist.
A final remark on the strengths of the 1-D approach should be done on the topic of signal processing with DL: in the data fusion tasks (Roncoroni et al., 2023a,b) as well as in the frequency manipulations, the main strength specifically lays on the 1-D approach. In fact, in both cases we can skip any lateral continuity, just focusing on the frequency-amplitude content behaviour.

2.5 LSTM and Geophysics: State of the art

Current state of the art proposes several LSTM application topics, covering different steps of velocity analysis, inversion, data interpretation, data processing and other analysis related to the exploration geophysics an more.
Velocity analysis and inversion are getting more and more important in actual geophysics: due to this, the creation of DL-based methodology has gained popularity. Pereg et al. (2020) proposes a new method for pre-stack time migration velocity analysis (MVA) of seismic data using RNNs. The approach leverages RNNs' ability to learn structural information in time and space to map local data sub-volumes to velocity points, generating the root-mean-square (RMS) velocity field. The approach is highly efficient computationally and can reduce the workload for analyzing large volumes of 3D seismic data, although this methodology could have problems in low Signal to Noise ratios (S/N) scenarios.
Gao et al. (2022) introduce a new model incorporating a deep-learning-based data correction term to account for the portions of the data that cannot be adequately represented by the convolutional model alone. The data correction term is implemented using a LSTM-based DL architecture, with the parameters learned from a dataset comprising several well logs. The results demonstrate that the new model effectively characterizes complex seismic data and the proposed SD method exhibits significant advantages over traditional methods when it comes to building high-fidelity new reflectivity model in complex scenarios.
Cova and Liu (2023) works with S-wave velocity in reservoir characterization and rock physics modeling. To address the lack of this information, a supervised data-driven approach is proposed using a graph convolutional network with a bidirectional gated recurrent unit (GCN-BiGRU).
The method leverages the total information coefficient to capture non-linear dependencies among well-log data and utilizes graph embedding to extract spatial relationship.

The bidirectional gated mechanism is employed to incorporate depth relationship in both upward and backward directions. Additionally, the method incorporates an unsupervised graph neural network to handle outliers and generates additional features using the complete ensemble empirical mode decomposition with the additive noise method.

Comparing the GCN-BiGRU network with other machine learning methods, (such as Castagna's empirical velocity formula, support vector regression, long-short-term memory (LSTM), GRU, and BiGRU) using a North Sea reflection seismic open dataset, the results demonstrate that the proposed method outperforms the other approaches in predicting S-wave velocity.

Zhang et al. (2022b) uses A Spatio-Temporal Neural Network (STNN) for post-stack impedance inversion in seismic data. The STNN consists of a convolutional neural network (CNN) block, and a recurrent neural network (RNN) block, enabling the network to capture the dynamics and correlations of seismic series in both space and time. The STNN outperforms traditional CNN-based methods in terms of inversion accuracy and geologic reliability.

Wang et al. (2022) proposes a novel method for robust and convenient S-wave velocity inversion and prediction. The S-wave velocity is an important parameter in reservoir characterization, seismic inversion, and geo-mechanical analysis. Obtaining accurate S-wave velocity measurements from field data is challenging, and traditional petrophysical modeling can be influenced by subjective factors. The proposed method combines a convolutional neural network (CNN) with a long short-term memory (LSTM) network to establish a deep hybrid neural network. The CNN captures the nonlinear mapping relationship between conventional logging data (acoustic and density) and S-wave velocity, while the LSTM network incorporates the vertical variation trend of the stratum. The proposed method offers a novel and practical approach for S-wave velocity prediction with significant implications in academia and industry.

Due to the great power of the RNNs in dealing with causal signals, an extensive use has been proposed for signal processing purposes.

He et al. (2022) start from wavelet transform, widely used in seismic waveform analysis, and address the computation issue on its application on big datasets. To address this problem, a novel wavelet selection method is proposed, considering the big dataset for intelligent processing of seismic signals. The method calculates the relevance (r) based on the correlation coefficient and variance contribution rate of seismic waveforms. By calculating r values for all seismic signals in the dataset, a set is formed. The decomposition stability (w) is defined as the mean value (m) divided by the variance value (s^2) of the set. The wavelet that maximizes w for the dataset is considered

the optimal wavelet. The proposed method is applied to automatic mining-induced seismic signal classification and seismic P arrival picking. The results demonstrate that the proposed method effectively enhances the intelligent processing of seismic signals, improving accuracy in classification, and reducing errors in picking seismic P arrivals.

Pan et al. (2020) faces conventional sparse spike deconvolution algorithms based on the iterative shrinkage-thresholding algorithm (ISTA) and its huge limitation, i.e. when the accuracy of the wavelets is compromised. To address this, a novel approach is proposed, combining recurrent neural networks (RNNs) and ISTA to develop an RNN-like ISTA as an alternative sparse spike deconvolution algorithm. The algorithm is tested using synthetic and real seismic data. The proposed algorithm improves the accuracy of deconvolution results by reducing the impact of incorrect seismic wavelets provided by conventional ISTA. The study presents the mechanism and derivation of the algorithm and verifies its effectiveness through examples both on synthetic and real data.

Biswas et al. (2019) successfully applies RNNs to seismic processing tasks, including Normal MoveOut (NMO) correction. Traditional methods involve human intervention and extensive computation time. In this study, a novel approach using recurrent neural networks (RNNs) is developed for NMO velocity estimation directly from seismic data. The RNN is trained using supervised learning, where input gathers (before NMO correction) and corresponding pre-calculated NMO velocities are used as labeled data. The Adam optimization algorithm is employed to train the RNN, adjusting the weights of the neurons to minimize the mean squared error between the estimated and correct NMO velocities. Once trained, the RNN can predict NMO velocities for the rest of the seismic gathers.

The method is validated on a noisy dataset from Poland, where only 10% of the CMPs are used for training. The trained network is then utilized to predict NMO velocities for the remaining CMP locations. The stack section obtained using the RNN-generated NMO velocities closely resembles the stack obtained using the conventional semblance method.

Liu and Ma (2022) address the challenge of predicting sharp and accurate images in spatio-temporal sequences. They introduce the concepts of Short time Interactive Memory (SIM) and Long erm Interactive Memory (LIM) to enhance the long term interaction between states in the network. The SIM block updates the current input and hidden states, while the LIM block updates memory cells and hidden states. By stacking the I-LSTM units, the IM-LSTM model is constructed. Experimental results demonstrate the effectiveness and flexibility of the proposed method, outperforming other advanced methods in terms of predicting future frames and other representations,

as evaluated by the Learned Perceptual Image Patch Similarity (LPIPS) index.
The high complexity of exploration projects, e.g. 4-D seismic datasets, requires DL for dealing with such a complexity, especially in the exploration geophysics.
Feng et al. (2022b) focuses on 4-D seismic monitoring, widely used in CO2 sequestration projects, to monitor the fluid flow in the subsurface through time. To address the limited data availability and high surveys costs, resulting in sparse seismic data, spatio-temporal neural network-based models have been developed to effectively and efficiently interpolate and extrapolate high-fidelity images . The models are based on an autoencoder architecture and incorporate the LSTM structure with a new loss function regularized by optical flow. Real 4-D post-stack seismic imaging data from the Sleipner CO2 sequestration field is used to validate the models. Both the numerical and expert evaluations demonstrate that the models can generate high-quality 2-D/3-D seismic imaging data at a reasonable computational cost. This offers the potential for real-time monitoring and near-future forecasting of CO2 storage reservoirs. The developed models provide valuable insights for understanding reservoir behavior and assessing potential risks in CO2 sequestration projects.
Zhang et al. (2022a) study reservoir parameters prediction, which is crucial for oil and gas exploration. To address the lack of complete reservoir description using elastic parameters obtained from seismic inversion, a deep learning method based on a gated recurrent neural network (GRNN) is developed for simultaneous prediction of porosity, saturation, and shale content. The GRNN, based on Gated Recurrent Unit (GRU), updates and resets the hidden state automatically. The input model parameters are: compressive wave velocity, shear wave velocity, and density. The model is trained to capture nonlinear relationship between these input parameters and multiple physical parameters. The input data is pre-processed using z-score whitening, and the model hyperparameters are optimized based on the mean absolute error (MAE) of the validation set. Although the prediction is weaker than methods using logging data due to differences in input parameters, the proposed GRNN method holds higher practical value for exploration work.
Sun et al. (2020) introduces a theory-designed recurrent neural network (RNN) for single- and multi-dimensional scalar acoustic seismic forward modeling, allowing the forward propagation to be set up. Training this network using measured seismic data effectively solves the seismic inverse problem and is equivalent to gradient-based seismic full-waveform inversion (FWI).
The RNNs are refined by comparing different optimization methods and learning rates, including standard deep-learning optimization, nonlinear conjugate gradient, and limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithms. Numerical analysis indicates that adaptive moment (Adam) optimization with a learning rate

matching standard FWI updates produces stable and reliable waveform inversion results. This finding is confirmed by a multidimensional 2D Marmousi experiment. Main limitation is linked to the limited exploitation of this methods on real data.
In conclusion, RNN were used to perform reflection seismic data interpretation and classification.
Júnior et al. (2023) proposes a methodology to aid geoscientists in analyzing and interpreting seismic data with RNN tools focusing on direct hydrocarbon indicators. The methodology, for gas accumulation detection, uses a combination of the convolutional LSTM model and particle swarm optimization scheme. The proposed method achieved impressive performance metrics, including an F1-score of 84.22% and an accuracy of 99.42% in the best scenario.
Nie et al. (2023) faces the problem of automated interpretation of seismic data. However, traditional deep learning methods solely rely on labeled data for model training, which limits their utilization of a vast amount of unlabeled data. To address this severe issue, self-supervised learning, commonly used in Natural Language Processing (NLP) and Computer Vision, is proposed as an effective approach to learn from unlabeled data.
In this study, a pretext task inspired by Masked Autoencoders (MAE) is designed for self-supervised pre-training using unlabeled seismic data. The pre-training phase aims to extract valuable information from the unlabeled data. Subsequently, the pre-trained model is fine-tuned for downstream tasks. Experimental results demonstrate that the model effectively captures information from unlabeled data through the pretext task, leading to improved performance in downstream tasks. One of the main limitation of this approach is the degradation of performances when the Signal-to-Noise ratio decreases.

2.6 Conclusion

In this introductory chapter, we have pointed out the fundamental concepts of LSTM networks and their various modifications for different geophysical applications exploited only in very recent years. We have discussed how LSTMs are particularly effective in tasks where retaining information over an extended period is essential. The unique structure of LSTMs, with memory blocks and forget gates, allows for precise control over the flow of information and enhances their suitability for several tasks in exploration geophysics.

We've also delved into the significance of LSTM algorithms modifications, such as peephole connections, coupled forget and input gates, and the Gated Recurrent Unit

(GRU). These innovations offer increased adaptability and improved information flow control, addressing some of the limitations faced by standard Recurrent Neural Networks (RNNs) when dealing with long sequences.

Furthermore, we've highlighted the importance of LSTM models in solving specific geophysical problems, such as seismic data interpretation, velocity analysis, and reservoir parameters prediction. The application of LSTM networks in addressing the complexities of geophysical exploration projects, including 4-D seismic imaging, has demonstrated their potential in real-time monitoring and forecasting.

Additionally, we've discussed the application of RNNs in signal processing, specifically dealing with seismic waveform analysis, where they have shown promise in improving accuracy and reducing computation times. Self-supervised learning using LSTM-based approaches has enabled the utilization of unlabeled data in automating seismic data interpretation.

In conclusion, this introductory chapter has laid the foundation for understanding the role of LSTM and RNN-based models in geophysics, showcasing their potential to revolutionize data analysis, interpretation, and prediction in the field. As we progress through this work, we will delve deeper into the applications, methodologies, and limitations of these neural network architectures in addressing complex geophysical challenges.

Bibliography

Ruslan Abdulkadirov, Pavel Lyakhov, and Nikolay Nagornov. 2023. Survey of Optimization Algorithms in Modern Neural Networks. *Mathematics* 11, 11 (2023), 2466. https://doi.org/10.3390/math11112466

George Bebis and Michael Georgiopoulos. 1994. Feed-forward neural networks. *IEEE Potentials* 13, 4 (1994), 27–31. https://doi.org/10.1109/45.329294

Yoshua Bengio, Patrice Simard, and Paolo Frasconi. 1994. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks* 5, 2 (1994), 157–166. https://doi.org/10.1109/72.279181

Reetam Biswas, Anthony Vassiliou, Rodney Stromberg, and Mrinal K. Sen. 2019. Estimating normal moveout velocity using the recurrent neural network. *Interpretation* 7, 4 (2019), T819–T827. https://doi.org/10.1190/INT-2018-0243.1 arXiv:https://doi.org/10.1190/INT-2018-0243.1

Junyoung Chung, Çaglar Gülçehre, KyungHyun Cho, and Yoshua Bengio. 2014. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling, In NIPS 2014 Workshop on Deep Learning, December 2014. *CoRR* abs/1412.3555, 1412.3555. arXiv:1412.3555 http://arxiv.org/abs/1412.3555

David Cova and Yang Liu. 2023. Shear wave velocity prediction using bidirectional recurrent gated

graph convolutional network with total information embeddings. *Frontiers in Earth Science* 11 (Feb. 2023), 1101601. https://doi.org/10.3389/feart.2023.1101601

Qiqi Dai, Yee Hui Lee, Hai-Han Sun, Jiwei Qian, Genevieve Ow, Mohamed Lokman Mohd Yusof, and Abdulkadir C. Yucel. 2022. A Deep Learning-Based GPR Forward Solver for Predicting B-Scans of Subsurface Objects. *IEEE Geosci. Remote Sensing Lett.* 19 (2022), 1–5. https://doi.org/10.1109/LGRS.2022.3192003

John Duchi, Elad Hazan, and Yoram Singer. 2011. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research* 12, Jul (2011), 2121–2159.

Shihang Feng, Xitong Zhang, Brendt Wohlberg, Neill P. Symons, and Youzuo Lin. 2022b. Connect the Dots: *In Situ* 4-D Seismic Monitoring of CO $_2$ Storage With Spatio-Temporal CNNs. *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), 1–16. https://doi.org/10.1109/TGRS.2021.3116618

Tian Feng, Saeed Mohanna, and Lingsen Meng. 2022a. EdgePhase: A Deep Learning Model for Multi-Station Seismic Phase Picking. *Geochemistry, Geophysics, Geosystems* 23, 11 (2022), e2022GC010453. https://doi.org/10.1029/2022GC010453 _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2022GC010453.

Zhaoqi Gao, Sichao Hu, Chuang Li, Hongling Chen, Xiudi Jiang, Zhibin Pan, Jinghuai Gao, and Zongben Xu. 2022. A Deep-Learning-Based Generalized Convolutional Model For Seismic Data and Its Application in Seismic Deconvolution. *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), 1–17. https://doi.org/10.1109/TGRS.2021.3076991

Felix A. Gers and Jürgen Schmidhuber. 2000. Recurrent nets that time and count. In *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, Vol. 3. IEEE, 189–194 vol.3. https://doi.org/10.1109/IJCNN.2000.861302

Felix A. Gers, Jürgen Schmidhuber, and Fred Cummins. 1999. Learning to forget: continual prediction with LSTM. In *1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470)*, Vol. 2. 850–855 vol.2. https://doi.org/10.1049/cp:19991218

Felix A. Gers, Nicol N. Schraudolph, and Jürgen Schmidhuber. 2003. Learning Precise Timing with Lstm Recurrent Networks. *J. Mach. Learn. Res.* 3, null (mar 2003), 115–143. https://doi.org/10.1162/153244303768966139

Alex Graves and Jürgen Schmidhuber. 2005. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks* 18, 5 (2005), 602–610. https://doi.org/10.1016/j.neunet.2005.06.042 IJCNN 2005.

Klaus Greff, Rupesh Kumar Srivastava, Jan Koutník, Bas R. Steunebrink, and Jürgen Schmidhuber. 2015. LSTM: A Search Space Odyssey. *CoRR* abs/1503.04069 (2015). arXiv:1503.04069 http://arxiv.org/abs/1503.04069

Zhengxiang He, Shaowei Ma, Liguan Wang, and Pingan Peng. 2022. A Novel Wavelet Selection Method for Seismic Signal Intelligent Processing. *Applied Sciences* 12, 13 (June 2022), 6470. https://doi.org/10.3390/app12136470

Sepp Hochreiter and Jürgen Schmidhuber. 1996. LSTM Can Solve Hard Long Time Lag Problems. In

Proceedings of the 9th International Conference on Neural Information Processing Systems (Denver, Colorado) *(NIPS'96)*. MIT Press, Cambridge, MA, USA, 473–479.

Rafal Jozefowicz, Wojciech Zaremba, and Ilya Sutskever. 2015. An Empirical Exploration of Recurrent Network Architectures. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37* (Lille, France) *(ICML'15)*. JMLR.org, 2342–2350.

Domingos Alves Dias Júnior, Luana Batista Da Cruz, João Otávio Bandeira Diniz, Aristófanes Corrêa Silva, Anselmo Cardoso De Paiva, Marcelo Gattass, Carlos Rodriguez, Roberto Quispe, Roberto Ribeiro, and Vinicius Riguete. 2023. Detection of potential gas accumulations in 2D seismic images using spatio-temporal, PSO, and convolutional LSTM approaches. *Expert Systems with Applications* 215 (April 2023), 119337. https://doi.org/10.1016/j.eswa.2022.119337

Diederik Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *International Conference on Learning Representations (ICLR)*. San Diega, CA, USA.

Jan Koutník, Klaus Greff, Faustino J. Gomez, and Jürgen Schmidhuber. 2014. A Clockwork RNN. *CoRR* abs/1402.3511 (2014). arXiv:1402.3511 http://arxiv.org/abs/1402.3511

Stefan C. Kremer and John F. Kolen. 2001. *Field Guide to Dynamical Recurrent Networks* (1st ed.). Wiley-IEEE Press.

Yann Lecun and Yoshua Bengio. 1995. *Convolutional networks for images, speech, and time-series*. MIT Press.

Guixin Liu and Zhonghua Ma. 2022. Prediction of spatiotemporal sequence based on IM-LSTM. In *2022 2nd International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI)*. IEEE, Nanjing, China, 247–250. https://doi.org/10.1109/CEI57409.2022.9950135

Marvin Minsky and Seymour Papert. 1969. *Perceptrons*. MIT Press, Cambridge, MA.

Benjamin Moseley, Andrew Markham, and Tarje Nissen-Meyer. 2019. Fast approximate simulation of seismic waves with deep learning.

A.P.O. Muller, J.C. Costa, B. Fraga, and C.R. Bom. 2023. Deep-salt: 3D salt segmentation from inaccurate migrated subsurface offset gathers using convolutional LSTM layers. 2023, 1 (2023), 1–5. https://doi.org/10.3997/2214-4609.202310124

Kevin P. Murphy. 2022. *Probabilistic Machine Learning: An introduction*. MIT Press. probml.ai

Tongyu Nie, Sirui Wang, Yuqi Wang, Xunqian Tong, and Feng Sun. 2023. An effective recognition of moving target seismic anomaly for security region based on deep bidirectional LSTM combined CNN. *Multimedia Tools and Applications* (Jan. 2023). https://doi.org/10.1007/s11042-023-14382-5

Michael A. Nielsen. 2018. Neural Networks and Deep Learning. http://neuralnetworksanddeeplearning.com/

Shulin Pan, Ke Yan, Haiqiang Lan, José Badal, and Ziyu Qin. 2020. A Sparse Spike Deconvolution Algorithm Based on a Recurrent Neural Network and the Iterative Shrinkage-Thresholding Algorithm. *Energies* 13, 12 (June 2020), 3074. https://doi.org/10.3390/en13123074

Deborah Pereg, Israel Cohen, Anthony A. Vassiliou, and Rod Stromberg. 2020. Prestack time migration velocity analysis using recurrent neural networks. *Journal of Applied Geophysics* 181 (Oct. 2020), 104160. https://doi.org/10.1016/j.jappgeo.2020.104160

Lauréline Perotin, Alexandre Défossez, Emmanuel Vincent, Romain Serizel, and Alexandre Guérin. 2019. Regression Versus Classification for Neural Network Based Audio Source Localization. In *2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. 343–347. https://doi.org/10.1109/WASPAA.2019.8937277

Giacomo Roncoroni, Emanuele Forte, Luca Bortolussi, Luca Gasperini, and Michele Pipan. 2022b. Polarity assessment of reflection seismic data: a Deep Learning approach. *BGO* 63 (Nov. 2022), 693–700. https://doi.org/10.4430/bgo00409

Giacomo Roncoroni, Emanuele Forte, Luca Bortolussi, and Michele Pipan. 2022a. Efficient extraction of seismic reflection with Deep Learning. *Computers & Geosciences* 166 (Sept. 2022), 105190. https://doi.org/10.1016/j.cageo.2022.105190

Giacomo Roncoroni, Emanuele Forte, and Michele Pipan. 2023a. Merging gated frequency-modulated continuous-wave Mars2020 RIMFAX ground-penetrating radar data. *GEOPHYSICS* 88, 2 (2023), A7–A12. https://doi.org/10.1190/geo2022-0466.1 arXiv:https://doi.org/10.1190/geo2022-0466.1

Giacomo Roncoroni, Emanuele Forte, Ilaria Santin, Ana Cernok, Andrea Rajsic, Alessandro Frigeri, Wenke Zhao, and Michele Pipan. Under review. Deep Learning driven interpretation of Chang'E4 Lunar Penetrating Radar. (Under review). Submitted to Journal of Geophysical Research: Planets.

Giacomo Roncoroni, Emanuele Forte, Ilaria Santin, and Michele Pipan. 2023b. Deep Learning based multi-frequency GPR data merging;. *GEOPHYSICS* 0, ja (2023), 1–35. https://doi.org/10.1190/geo2023-0215.1 arXiv:https://doi.org/10.1190/geo2023-0215.1

Giacomo Roncoroni, Carlo Fortini, Luca Bortolussi, Nicola Bienati, and Michele Pipan. 2021. Synthetic seismic data generation with deep learning. *Journal of Applied Geophysics* 190 (July 2021), 104347. https://doi.org/10.1016/j.jappgeo.2021.104347

Sebastian Ruder. 2016. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747* (2016).

David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. 1986. Learning representations by back-propagating errors. *nature* 323, 6088 (1986), 533–536.

Alex Sherstinsky. 2018a. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network. *ArXiv* abs/1808.03314 (2018).

Alex Sherstinsky. 2018b. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network. http://arxiv.org/abs/1808.03314 cite arxiv:1808.03314Comment: 39 pages, 10 figures, 66 references.

Ralf C. Staudemeyer and Eric Rothstein Morris. 2019. Understanding LSTM - a tutorial into Long Short-Term Memory Recurrent Neural Networks. *CoRR* abs/1909.09586 (2019). arXiv:1909.09586 http://arxiv.org/abs/1909.09586

Jian Sun, Zhan Niu, Kristopher A. Innanen, Junxiao Li, and Daniel O. Trad. 2020. A theory-guided

deep-learning formulation and optimization of seismic waveform inversion. *GEOPHYSICS* 85, 2 (March 2020), R87–R99. https://doi.org/10.1190/geo2019-0138.1

Jun Wang, Junxing Cao, Shuang Zhao, and Qiaomu Qi. 2022. S-wave velocity inversion and prediction using a deep hybrid neural network. *Science China Earth Sciences* 65, 4 (April 2022), 724–741. https://doi.org/10.1007/s11430-021-9870-8

Xinming Wu, Luming Liang, Yunzhi Shi, and Sergey Fomel. 2019. FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation. *Geophysics* 84, 3 (04 2019), IM35–IM45. https://doi.org/10.1190/geo2018-0646.1 arXiv:https://pubs.geoscienceworld.org/geophysics/article-pdf/84/3/IM35/4682641/geo-2018-0646.1.pdf

Kaisheng Yao, Trevor Cohn, Katerina Vylomova, Kevin Duh, and Chris Dyer. 2015. Depth-Gated LSTM. *CoRR* abs/1508.03790 (2015). arXiv:1508.03790 http://arxiv.org/abs/1508.03790

Guoqiang Peter Zhang. 2000. Neural networks for classification: a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 30, 4 (2000), 451–462.

Jiajia Zhang, Zhuofan Liu, Guangzhi Zhang, Bin Yan, Xuebin Ni, and Tian Xie. 2022a. Simultaneous prediction of multiple physical parameters using gated recurrent neural network: Porosity, water saturation, shale content. *Frontiers in Earth Science* 10 (Aug. 2022), 984589. https://doi.org/10.3389/feart.2022.984589

Jian Zhang, Hui Sun, Wentao Yuan, Chen Yang, and Yiran Xue. 2022b. Post-Stack Impedance Inversion Based on Spatio-Temporal Neural Network. *IEEE Geoscience and Remote Sensing Letters* 19 (2022), 1–5. https://doi.org/10.1109/LGRS.2022.3227071

Chapter 3

Explainable AI

Explainable AI (XAI) refers to the concept of designing and developing artificial intelligence systems and machine learning models in a way that makes their decision-making processes understandable to the user. XAI aims to bridge the gap between the black-box nature of many advanced AI models and the need for humans to trust, check, interpret, validate, and control these systems. XAI is particularly important in scenarios where AI is used for critical applications and decisions.

Some of the key principles and techniques of Explainable AI are interpretability, Feature Importance, Model Transparency and both local and global explanations.
Making AI models more interpretable by providing insights into how they arrive at their predictions or decisions making the methodology stronger and more reliable. This can involve using simpler models, like decision trees or linear models, that are easier to understand. Although this, for some complicated tasks, these simple models do not grant the required complexity to provide affordable results.
Identifying the most influential features or factors that the AI model uses to make predictions can help users to understand which variables contribute the most to a decision and evaluate the physical reasons that were leading the model to give certain output.
Furthermore, providing both local explanations for individual predictions and global explanations for the overall behavior of the model is crucial to all the AI based techniques. Local explanations help users to understand why a specific prediction was made, while global explanations offer insights into the model's general behavior.
Explainable AI is essential not only for increasing trust in AI systems but also for debugging and improving models, uncovering biases, and identifying potential concerns. XAI allows the user to have more control over AI systems and fosters

responsible AI development and deployment.

In this chapter, starting from the basic knowledge provided into the Chapter 2, we focus on some XAI-related topics, starting from the results obtained throughout this whole research work. Actually, understanding the results of given models is not only linked to its complexity, but also to the training data and on the internal feature of the neuron itself.

3.1 Training dataset

One of the fundamental challenges in real-word DL application is the gap between training and testing data. This section explores the importance of closing this gap, discusses the challenges involved, and outlines potential strategies and approaches to address this issue.

One of the central challenges in active wavefield geophysics, as well as in many other machine learning applications, is the gap between training and testing data. In the context of seismic data analysis, this gap arises from differences between the data distributions encountered during training and those encountered in the real-world testing dataset. Furthermore, for some specific tasks, it is difficult, when possible, to acquire field data to train the model. This gap can lead to model inaccuracies, and in turn, limit the practical utility of machine learning algorithms.

Understating which is the statistical distribution that a model need to be trained to be applicable on field data is crucial and is heavily task-dependent. The best approach would be the use of real data with precise labels: even though this would be the best choice, most of the times it is not applicable since the data does not generalize properly the geological complexity, are biased by subjective intervention (e.g. in seismic interpretation), or are very expensive to acquire.

Due to this limitation we were testing the best way to generate synthetic data that would lead to accurate applications on real-case-scenarios.

The best approach is to perform a Finite Difference modelling of the wavefield: this is crucial for some tasks, likesynthetic data generation Moseley et al. (2020), Roncoroni et al. (2021), but can be too expensive and time demanding for some other applications like reflections picking (Roncoroni et al., 2022), due to the huge amount of different scenarios that are mandatory to generalize well the real world complexity.

One of the key modifications introduced to the classical seismic convolutional model Russell (1988) related to the noise: while from literature Yilmaz (2001) the seismic trace was introduced as:

$$S(t) = r(t) * w(t) + n(t) \tag{3.1}$$

where $S(t)$ is seismic trace, $r(t)$ is the reflection coefficients series over time, $w(t)$ is the wavelet and $n(t)$ is the noise.
In Roncoroni et al. (2022) we found out that, for NN training purposes and for specific tasks, a better generalization for both real seismic, or even GPR data is represented by:

$$S(t) = (n_1(t) + r(t)) * w(t) + n_2(t) \tag{3.2}$$

where $S(t)$ is seismic trace, $r(t)$ is the reflection coefficients series over time, $w(t)$ is the wavelet and $n_2(t)$ is the noise added post convolution, while $n_1(t)$ is the noise added directly to the reflection coefficients, both having a random distribution.

In order to show the data preparation, we provide the code for this section in https://github.com/Giacomo-Roncoroni/LSTM-wavefield/tree/main/00_data_generation.
In Figure 3.1.1 we show a starting velocity model, Figure 3.1.1A, the reflection coefficient computed as in Yilmaz (2001) 3.1.1B and the final convoluted data 3.1.1C. Velocity model is created according to the one introduced by Wu et al. (2019).

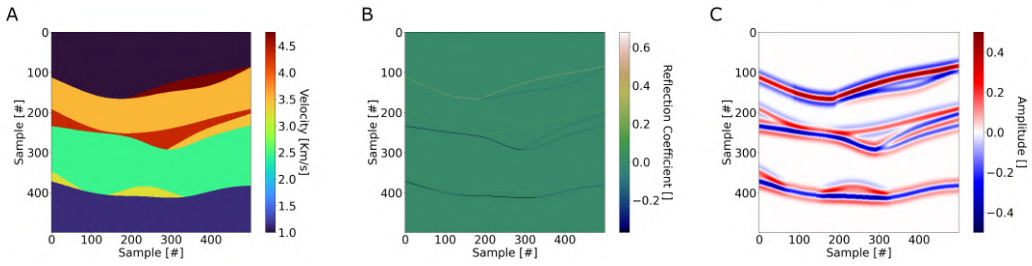

Figure 3.1.1: 2-D Data preparation process: random velocity model (A), reflection coefficient series (B) and final convoluted trace without any noise (C).

By adding random noise to the reflection coefficients series, as described in Eq. 3.2, we get results provided in Figure 3.1.2

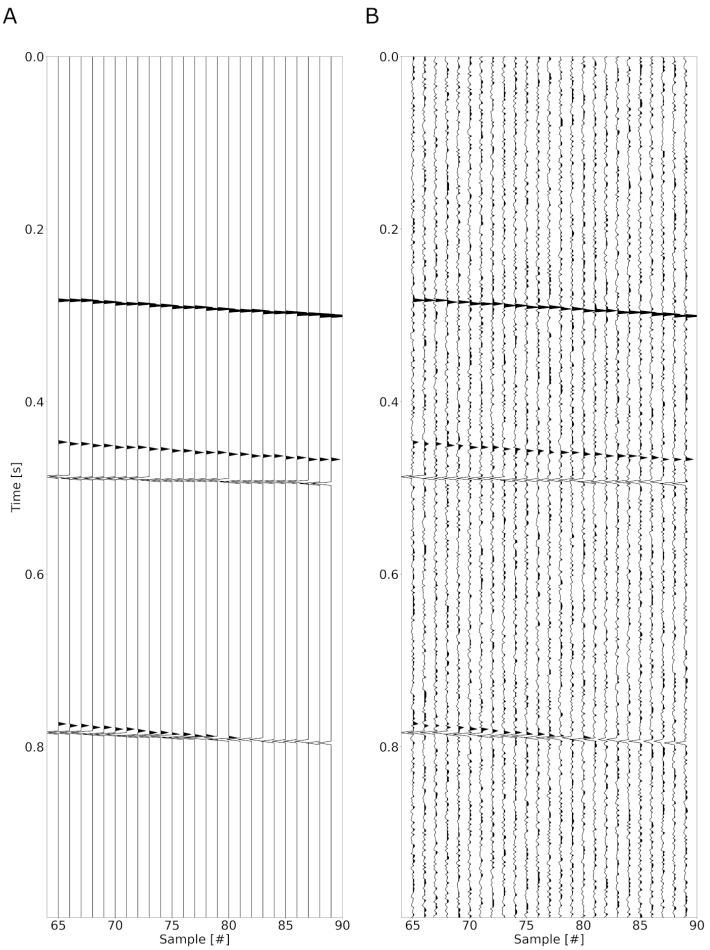

Figure 3.1.2: Noise free reflectivity series (A) and the same series with random noise added (B).

If we now convolve a wavelet (a Ricker wavelet in used in all these examples) we get to the final 2-D seismic data simulation. In Figure 3.1.3 we can compare the results: as we can see, the noise introduced by the convolution is better representative of a real seismic dataset.

The noise shown in Figure 3.1.3B is still strong, but it does not affect the reflections. Training performed on a dataset like the one in Figure 3.1.3B would led to worst performances in applications on real data.

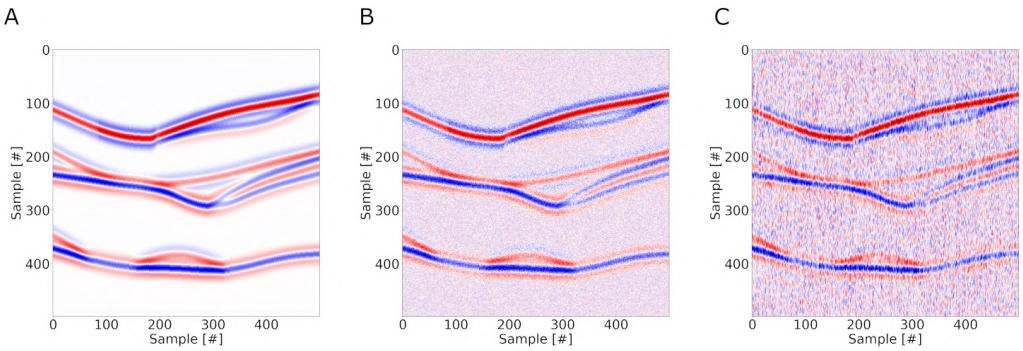

Figure 3.1.3: Convoluted data without noise (A), as described in Eq. 3.1 (B) and as proposed in Eq. 3.2 (C).

Looking more carefully at the waveform in Figure 3.1.4, we can clearly see that the pre-convolution noise acts on the same frequency area as the actual signal, introducing random noise that can also be, by chance, coherent between some adjacent traces.
A data like in Figure 3.1.4C can generalize better the noise component that we can find on real seismic data.

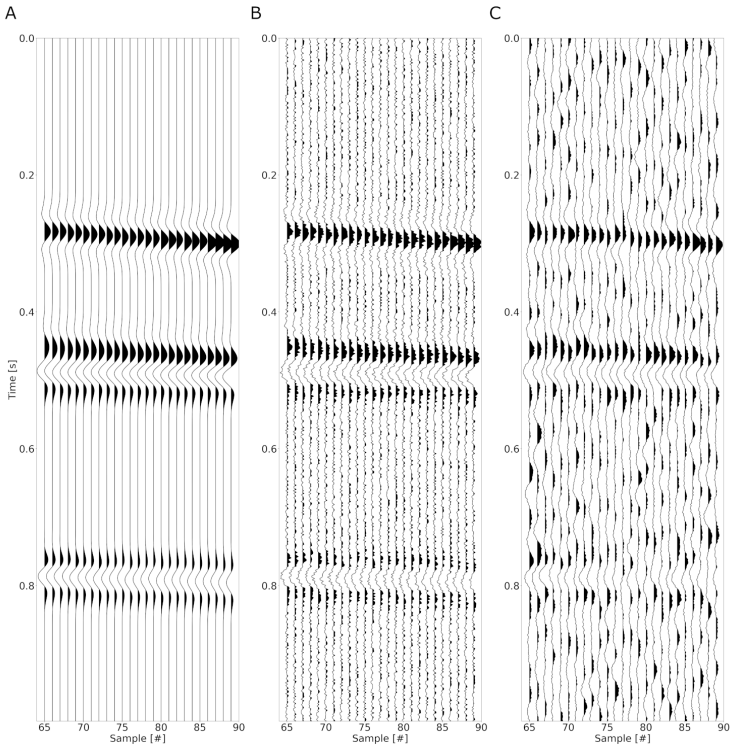

Figure 3.1.4: Wiggle plot of the convoluted data without noise (A), as described in Eq. 3.1 (B) and as proposed in Eq. 3.2 (C).

3.2 Convolution simple examples

Since one of the key concepts of this work is to establish when LSTM could be good to develop specific methods in active-wavefield geophysics, we provide here a simple example on the use of LSTM and on the eventual use of the Bi-Directional wrapper. All codes related to this section are available in https://github.com/Giacomo-Roncoroni/LSTM-wavefield/tree/main/01_LSTM-1D-Conv.

If we assume that we want to train a LSTM-based NN to perform the convolution of a fixed wavelet (a Ricker wavelet with a fixed frequency f_0 in the present case), to a randomly distributed reflectivity coefficients series, we need to define the training dataset.
We can easily create a dataset by using a convolution operator on randomly distributed reflection coefficients. Before we train the NN let's consider the two cases depicted in Figure 3.2.1 namely: a maximum phase convolution, i.e. 3.2.1B, and a zero phase convolution i.e. 3.2.1C

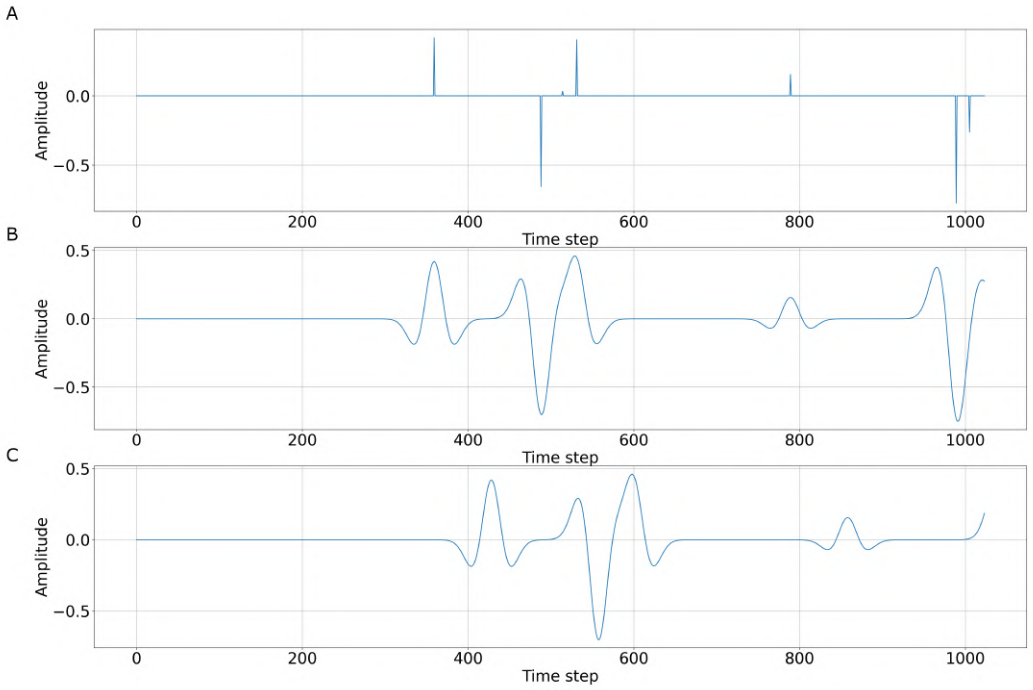

Figure 3.2.1: Task introduction: random reflectivity coefficients (A), maximum phase convolution with a Ricker wavelet (B) and zero phase convolution (C)

We define two different NNs. A plain LSTM is a NN composed by four layers with 8, 4, 2 and 1 neurons, respectively, defined *mono-Directional* in the next example.

A second NN is composed by four layers: Bi-directional wrappers with 4, 2 and 1 neurons, and one with a single LSTM neuron at the end, we will refer to this NN as *Bi-Directional* in this section.
As in the previous chapter, LSTM is a causal layer, so if we perform a maximum phase convolution we have to take into consideration that for the *mono-Directional* it will be impossible to generate the information of the half Ricker wavelet coming before the reflection coefficient. The *Bi-directional* will perform better in this case, since half of the neurons will see the reverse-in-time version of the input, being able to reconstruct the whole waveform.
On the other hand, if we face a zero phase convolution, while the *mono-Directional* will have the perfect geometry to achieve the task, half of the information in the *Bi-Directional* will not be important for the final result and will have to be discarded. This will make the training process more complicated and even achieving worst results.
As we can see in Figure 3.2.2, the expectations are confirmed by the training curves. While in the maximum phase training, Figure 3.2.2A, the *Bi-Directional* manages to get quite low values of loss, the *mono-Directional* get stuck to a local minima after around 50 Epochs.
On the other hand, in the zero phase convolution, Figure 3.2.2B, *Bi-Directional* shows three different local minima with a loss reduction at around 60 and 130 Epochs. The *mono-Directional* performs better in this case with a smooth loss, still decreasing at epochs 300.

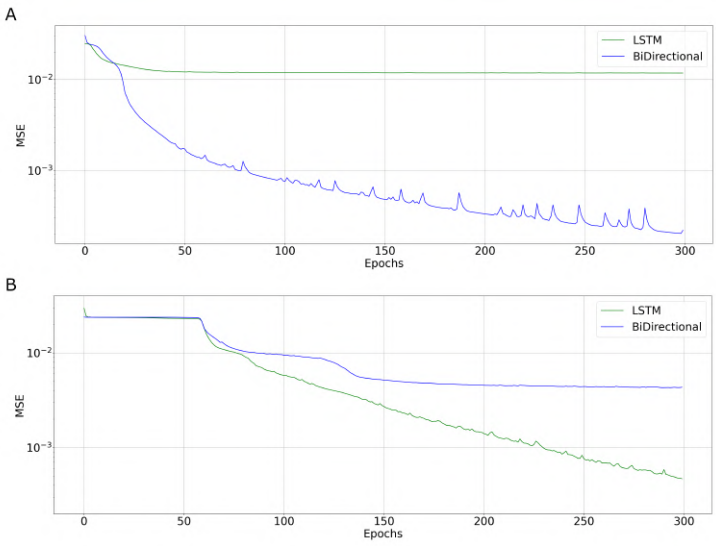

Figure 3.2.2: Training results of the *mono-Directional*, in green, and the *Bi-Directional*, in blue. Loss vs Epochs is shown for the maximum phase convolution (A) and for the zero phase convolution (B).

An exemplary prediction on the maximum phase training is shown in Figure 3.2.3. As we can see, the *Bi-Directional* performs quite good, while the *mono-directional* is not able to predict half of the wavelet.

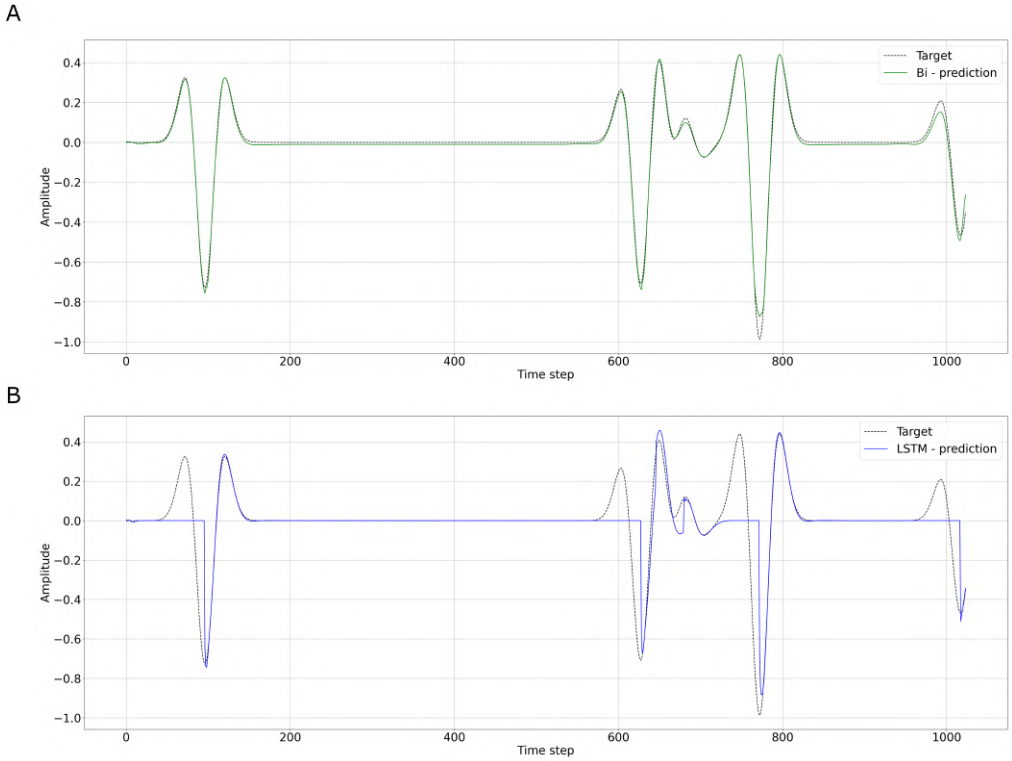

Figure 3.2.3: Maximum phase prediction: target curve is depicted in black. *Bi-Directional* prediction is shown in A, while *mono-directional* is presented in B.

A sample prediction on the zero phase training is shown in Figure 3.2.4. While in this case the *mono-directional* performs good, the *bi-directional* shows, as expected, worst performance.

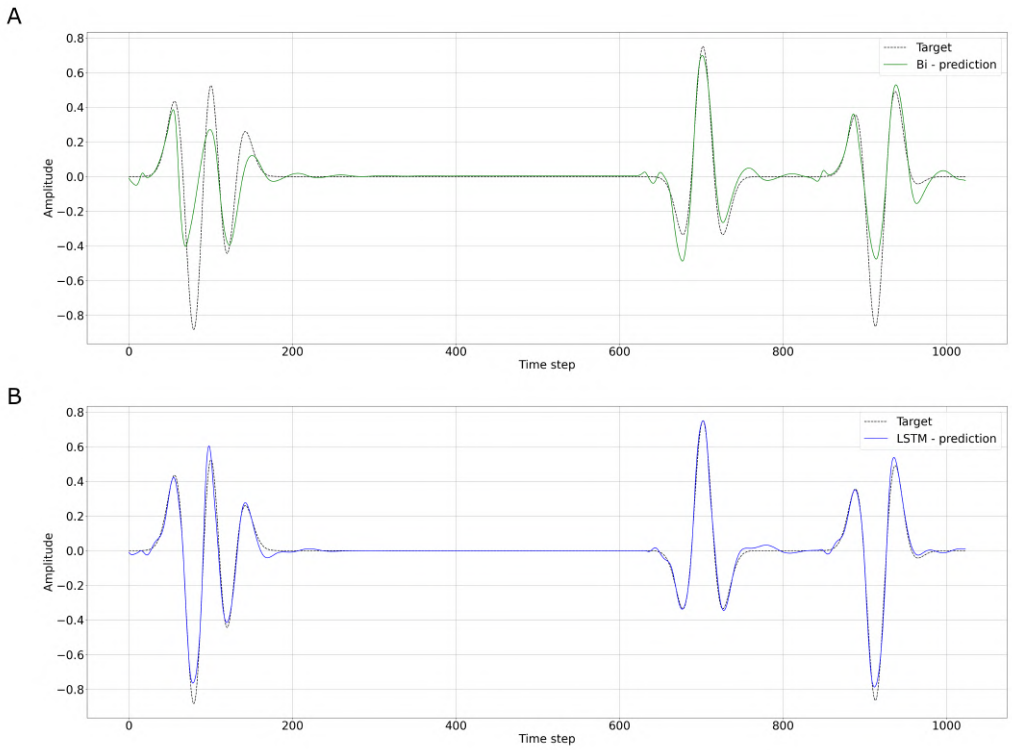

Figure 3.2.4: Zero phase prediction: target curve is depicted in black. *Bi-Directional* prediction is shown in A, while *mono-directional* is presented in B.

3.3 Deep Attributes: innovative LSTM-based seismic attributes

In this article we propose a new approach to help reflection seismic interpretation using deep learning algorithms, focusing on the potential of exploiting the hidden layers of neural network predictions as new reflection seismic signal attributes (defined in the article as "Deep Attributes"). We tested this procedure on synthetic and field datasets evaluating its time-monitoring performance on a 4-D seismic dataset.

The article specifically focuses on the use of LSTM based neural networks in seismic interpretation. The choice of LSTM is motivated by the causal nature of the RNN, which can provide a more reliable representation of the different signal components embedded in the seismic data. The article provides a detailed explanation of the structure of an LSTM-based neural network and how it can be used to extract hidden features from seismic data.

We applied the analysis on both synthetic and field 2-D and 3-D reflection seismic datasets, demonstrating the affordability and versatility of the proposed procedure. We also evaluated the performance of DA on a 4-D seismic dataset to assess the applicability for time-monitoring purposes. The results showed that the proposed approach is effective in reducing the number of attributes panels while keeping both high and low frequency features. We conclude that the proposed approach has the potential to improve seismic interpretation accuracy and efficiency.

Deep Attributes: innovative LSTM-based seismic attributes

G. Roncoroni, E. Forte, M. Pipan

Submitted to: Geophysical Journal International, *Under review.*

Key Words

Deep Learning; *Deep Attributes*; *Seismic Attributes*; *Sleipner Dataset*

Highlights

- We define Deep Attributes as the internal prediction of the Neural Network.
- The study of the Hidden Layer of an Encoder-Decoder LSTM gives us information about the input dataset.
- We applied the study on Marmousi model with comparison with the velocity model.
- We applied the methodology on a seismic 2-D section and on the Sleipner 4-D dataset.

Introduction

Seismic attributes have been developed since the 1970s to help reflection seismic interpretation exploiting additional quantities to classical reflection amplitude and multi-trace horizons correlation. In fact, the basic principle of seismic data interpretation is the detection of reflections which are related to subsurface impedance contrasts (e.g. Anstey (2013)). Therefore, seismic horizons are linked not only to any stratigraphic or structural contacts, but also to porosity or fluid content variations within the same geological units.

An incredibly large number of seismic attributes have been developed since the first applications. While most of the attributes are well-defined from the mathematical point of view their correlation with specific physical parameters or geological elements is still not apparent (Li and Zhao, 2014). Since the 1990s, the further development of new seismic attributes or the more sophisticated calculation of already available ones benefited from 3-D seismic surveys, characterized by a spatial data coverage which was previously inconceivable (Chopra and Marfurt, 2005). In addition, 3-D based seismic attributes made possible remarkable advantages in terms of (semi) automated data

interpretation and 3-D (auto) picking and surface/volume extraction, as well as more accurate estimates of reservoir parameters and their spatial variations.

Following the definition of Chopra and Marfurt (2007), seismic attributes are *any measure of seismic data that helps us visually enhance or quantify features of interpretation interest*; we here propose a new approach to automatically extract seismic attributes from any reflection seismic dataset based on Neural Network (NN) calculations.

In recent years, NN, and in a broader sense, Deep Learning (DL) techniques have been extensively applied in different branches of geophysics (a comprehensive review is provided by Yu and Ma, 2021), and particularly in different steps of reflection seismic surveys with several different objectives:

- Optimize data acquisition maximizing the information while avoiding redundant data (e.g. Lu et al. (2019); Jadhav et al. (2016a))
- Perform data simulations creating realistic synthetic data sets (e.g. Moseley et al. (2020), Roncoroni et al. (2021))
- Enhance data quality by means of interpolation, processing and imaging algorithms and techniques (e.g. Jia and Ma (2017); Wang et al. (2019b); Hou and Hoeber (2020); Klochikhina et al. (2020))
- Improve interpretation, data classification, information extraction and seismic horizon/surface picking (Sain and Kumar, 2022)
- Perform data inversion and parameters extraction (e.g. Wang et al. (2019a); Wang et al. (2020); Ruiz et al. (2021)).

Alongside all these applications, we propose a totally new way of using Hidden Layer (HL) predictions, which are usually the 'transparent' steps of any NN lying in between data input and the expected output. The idea is based on the inference process of a LSTM-based NN. In its HL a NN extracts information related to some features (i.e. attributes) embedded in the input data and uses them to get the required inference (i.e. the output). DL applications typically ignore the information from the intermediate steps because the main interest lies in the final output which is considered the manageable result of the entire process. In other words, we analyse the possibility to exploit the intermediate prediction steps to highlight features and emphasize characteristics embedded in the data but neither recognizable by traditional interpretation, nor made apparent by classical attributes or multi-attribute approaches.

Since the obtained features are produced by a DL procedure and they are full-fledged attributes we decided to refer to them as *Deep Attributes* (DA). In addition, such a

definition follows similar ones already used, even for slightly different issues, in other fields of research and technology like face recognition (Jadhav et al., 2016b) and, more in general, to image analysis and pattern recognition (Kim et al., 2022).

To avoid any confusion, we remark that our procedure does not combine, mix, classify, fuse previous attributes. Therefore, it is different from the ones recently proposed by Qian et al. (2018)and further modified by Li et al. (2022) in which DL is used to *extract deep features from seismic waveforms and combine multiple attributes for seismic attribute fusion at the same time* so typically exploiting the NN output and not considering its HL.

We applied our analysis on both synthetic and field 2-D and 3-D reflection seismic thus demonstrating the affordability and versatility of the whole procedure. Furthermore, we evaluated the performance of DA on a 4-D seismic dataset to assess the applicability for time-monitoring purposes.

Methods

In this paper we focus on LSTM-based NN, however, the proposed approach could be applied, in principle, to any NN geometry. The choice of LSTM is motivated by the causal nature of the Recurrent Neural Network (RNN) which can provide a more reliable representation of the different signal components embedded in the seismic data. The simplest RNN is made up of a single neuron that receives an input, produces an output, and sends that output to itself and to the output vector. At each time step t, the recurrent neuron receives the inputs x(t), as well as its own output from the previous time step y(t-1). Special neurons have been introduced to handle long-term dependencies: one of the most used is the LSTM, at first introduced by Hochreiter and Schmidhuber (1997)). LSTMs are in fact explicitly designed to avoid the long-term dependency problems: they have a chain-like structure, as for the RNN, but the repeating module has a different architecture (see Fig. 3.3.1), with the peculiar ability to keep on information even from x(t)<x(t-1). The capability to spot long time dependencies within a series (in our case a time series) is the reason for the choice of this type of layers for the proposed extraction method.

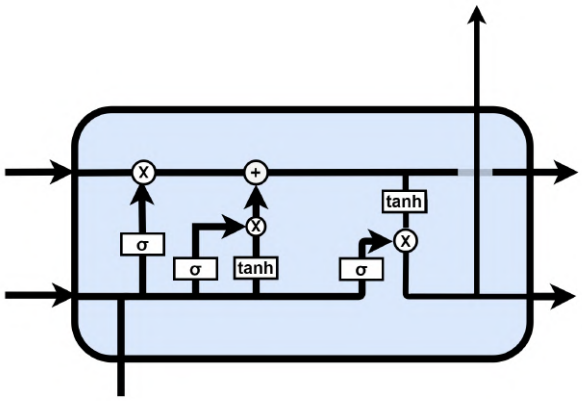

Figure 3.3.1: Conceptual scheme of a neuron in a LSTM which contains four interacting layers.

Another crucial point for the methodology is the NN geometry: the use of an encoder-decoder structure is directly linked to the typical Encoder-Decoder Convolutive NN (ED-CNN). A classical geometry of ED-CNN is made of a chain of couples of CNN layer, linked with pooling layers in the encoder and with up-sampling layer in the decoder. A pooling layer takes values in an interval (defined as kernel) and gives out a single value, i.e. the maximum values in this case; doing this we can perform a reduction of the length of the trace equal to the kernel size. An up-sampling layer makes exactly the opposite: it takes a single value at and replicates it kernel-size times.

The mix between this kind of neurons and the encoder-decoder geometry grants us that each values of the hidden layer can be directly linked to the correct position in time or, in some way, to what happened at previous times. This assure the causality typical of seismic waves propagation.

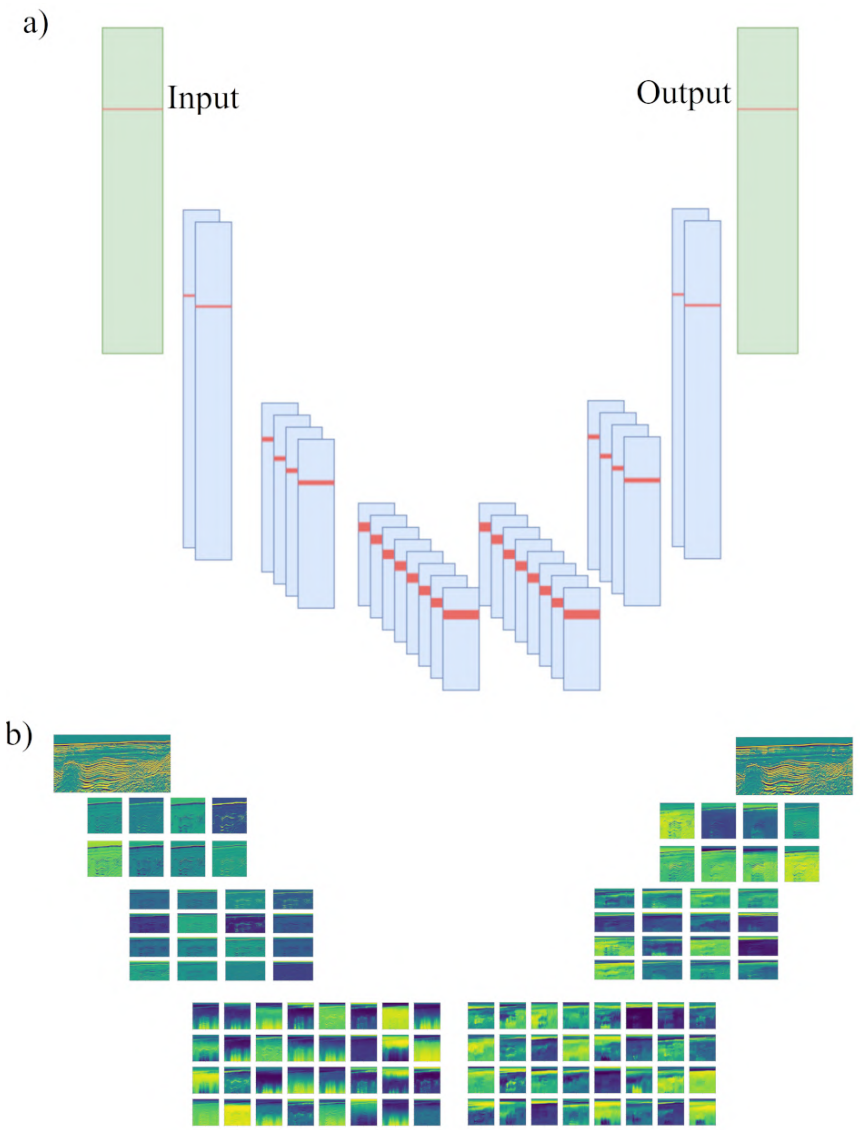

Figure 3.3.2: a) Base structure of the proposed method. The geometry is based on an encoder-decoder LSTM NN with input equal to output. Recurrent layers are depicted in blue, while a sample at time t is depicted in red to show the loss of resolution of DA. b) Example of applications of the LSTM ED to the real data shown in Figure 6. We can see the loss of resolution and the implementation of DA (lower part). We can also evaluate the good reconstruction of the input, comparing the two top right pictures (for a better visual comparison see also Figure 3.3.7).

In Figure 3.3.2 we explain the main NN geometry used. If we try to focus on the change of a single time step, assuming kernel-size = 2 both for the pooling and for the up-sampling layers, we have before the first pooling layer point D1, which refers to time t. After the first pooling layer we get point D2, referring to time [t, t+1] and at the third

layer we have point D3 referring to time [t, t+3]. In the encoded version of the signal we have point E1 and E2 referring both to time [t, t+5] and the information for their creation comes just from time<t+5, since the recurrent neuron doesn't go backwards. In the Decoder part we have exactly the inverse of what described above.

Since the base of the methodology is the down sampling of the seismic wavelet, we have to take care of the input sampling rate. In order to estimate the optimal sampling rate for each signal (beside the Nyquist-Shannon theorems and related issues, see e.g. Dossi et al. (2018)) we compute the amplitude spectrum in the Fourier domain with a unit sampling interval. We then resample it in order to get a desired frequency with this unitary value. This approach allows to work without changes on every input frequency without considering the sample rate. This allows to generalize the values we will state later in this paper to any source frequency.

The best value was estimated taking into account the minimum depth of the Neural Network needed to lose the wave nature during down sampling: if we oversample a trace, e.g. a wavelet is discretized in 64 points, after three pooling layers we will still have 8 sample per wavelet and in turn the task of reconstructing the signal will be easy and no DA will be needed. On the other hand, if we use a too low sampled trace, we will have to create a very deep NN and we will lose a large amount of information.

By taking into account the maximum discretization and the loss value, since we still have to get a good reconstruction of our input data, we found as an optimal value 16 samples per wavelet. We tested this value on different data with various complexities and we got good results.

As we can see in Figure 3.3.3, the loss function has a plateau between 4 and 8 samples, while it starts decreasing with higher values. From 32 samples the number of pooling layers is not enough to make the NN able to lose the waveform during the training: this means that we get very good results in terms of prediction (which is not the main objective of DA extraction), but poor results in term of feature extraction.

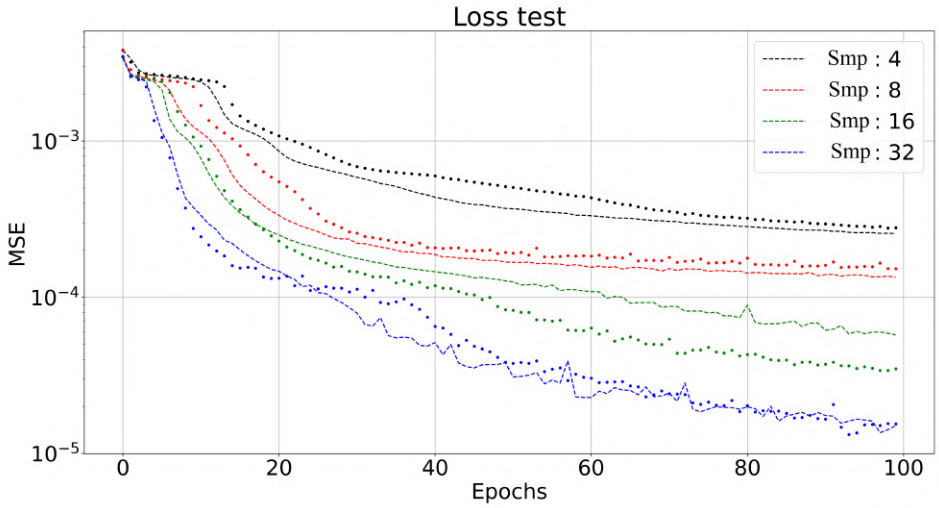

Figure 3.3.3: Loss function values with varying discretization steps of the input: with 4 and 8 samples we got to a plateau, while increasing the sampling rate we get better results but to avoid overfitting we chose 16 (green) as a good value since it can get acceptable.

Once we have defined the procedure, the main objective of the NN is to obtain an exact copy of the input, but by processing it into the Encoder part so that it can retrieve the original trace in the Decoder part.

The proposed workflow is:

- Create a LSTM-based NN that fits a specific problem.
- Train the NN on the data.
- Use the HL predictions as set of additional information (i.e. DA) for improved seismic analysis and interpretation purposes.

Once we have a trained NN we can start working on the prediction of each single neuron from each HL. In Figure 3.3.4 we show all the DA produced by the trained NN during the horizons extraction step.

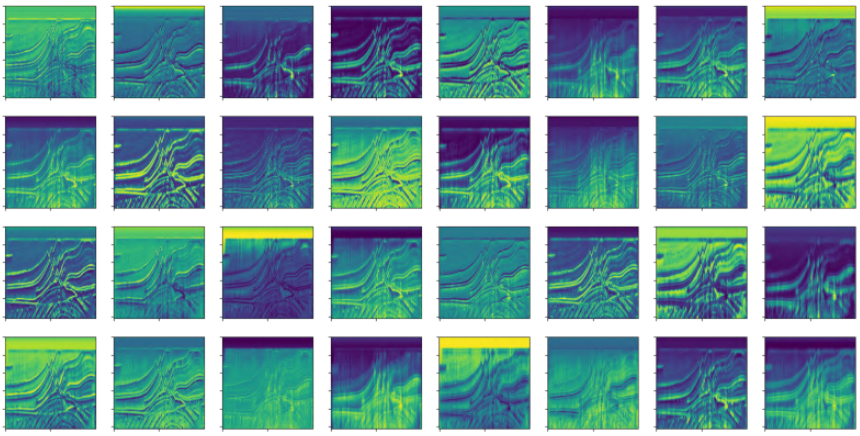

Figure 3.3.4: 32 DA of hidden layer 3 of the Encoder phase applied to a seismic profile of the Marmousi data set.

The proposed methodology can be applied to several different datasets and with slightly different purposes. In this paper we propose two different approaches to evaluate its potential.

At first, we used the method to infer complex geometries in 2-D seismic profiles. We tested the effectiveness of the proposed DA approach on synthetic and field data considering the obtained DA then computing a Principal Component Analysis (PCA) in order to reduce the number of attributes, while summarizing the whole information content.

We further exploit a single DA applying the same NN to a 3-D seismic survey repeated in different years (i.e. a 4-D data set) to monitor a CO_2 storage field. This allows to evaluate the stability and the repetitiveness of the methodology and its use for monitoring purposes.

Results and Discussion

Application on 2-D data: Synthetic example

In order to test the methodology we performed at first an analysis on the elastic Marmousi model, (Martin et al., 2002) which is a classical data set used for evaluation purposes. An example HL of the Encoder part of the NN applied to a profile of this dataset is shown in Figure 3.3.4.

From Figure 3.3.4 it is apparent that some of the DA provide information about the single reflectors geometry, while other are more related to the different seismic signatures of the various zones. Moreover, it is interesting to notice that the resolution level of the DA panels is variable and spans over a wide range. Since it is difficult to get and summarize the whole information contained in many DA, we decided to apply a PCA to each HL to get less attributes that could condense all the extracted features. In order to better understand how much information we are losing with the PCA analysis we compute the *explained variance ratio*. As we can see in Figure 3.3.5, with just 9 components we can explain 0.91 (i.e. 91%) of the total variance of the first HL of the encoder.

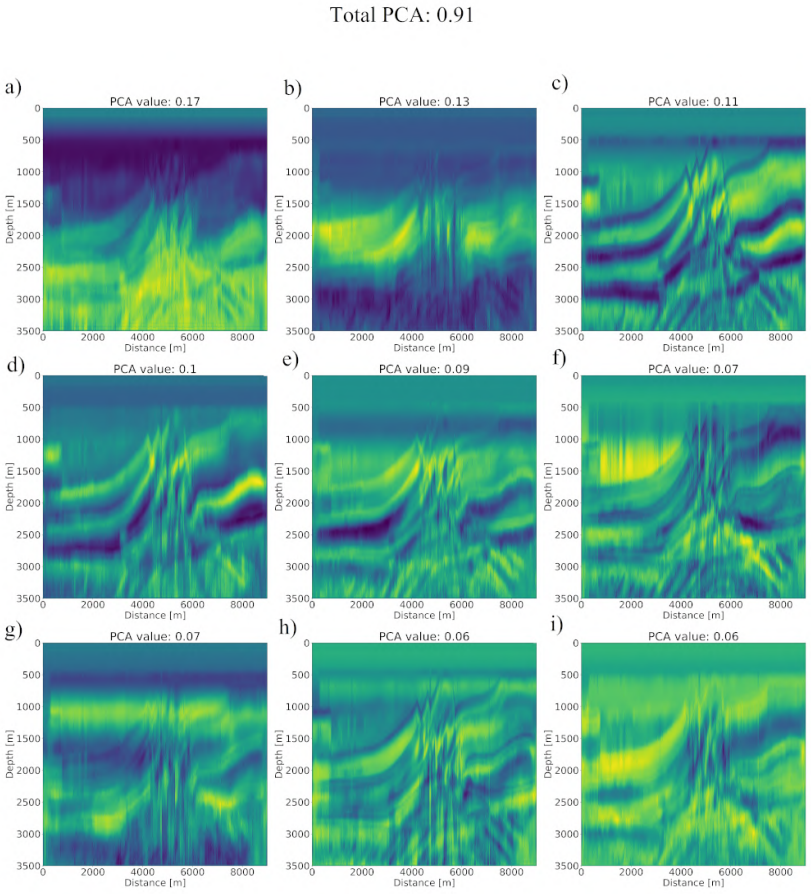

Figure 3.3.5: Test of the application of PCA on the HL. The total explained variance ratio is plotted on the top of each picture. As we can see, 9 PCA results can explain 0.91 of the total variance.

From the PCA panels it is apparent that they encompass not only the information

related to the geometry and location of the reflections, but even more important, there are additional details related to materials in-between reflectors (i.e. transmission attributes) or to changes in the frequency content or main geologic domains. In order to better understand the prediction performances of the proposed methodology we directly compare some of the velocity models used to generate the Marmousi dataset with some of the obtained DA (Figure 3.3.6).

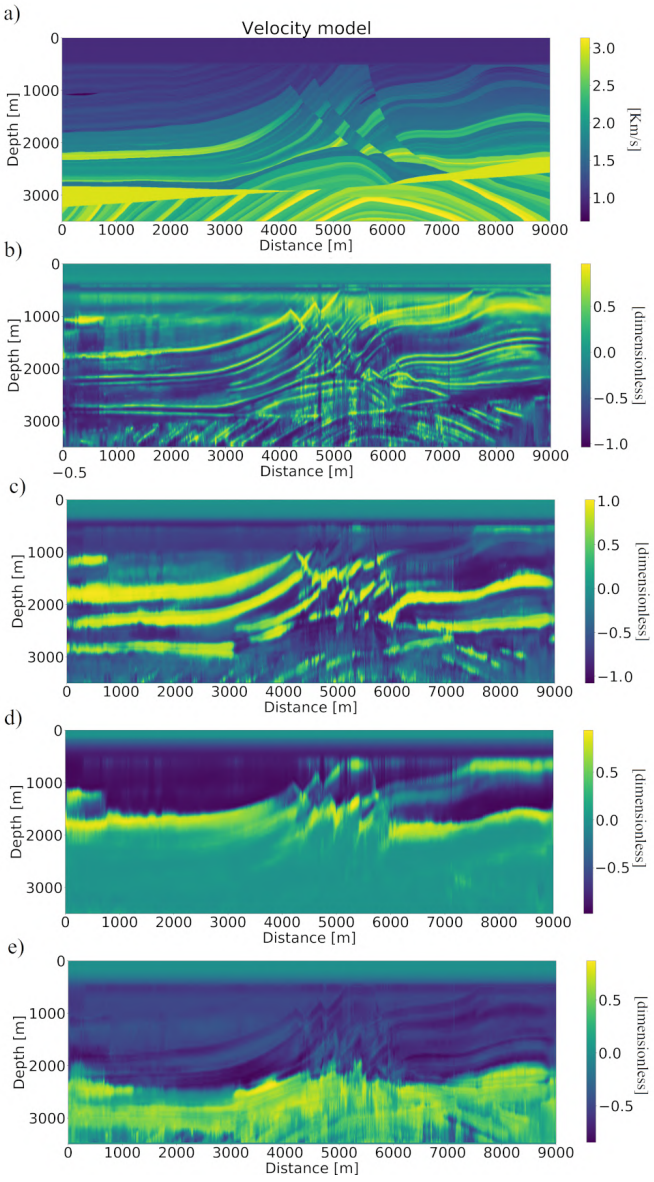

Figure 3.3.6: Velocity model a), and DA b) to e) for an example of seismic profile from the Marmousi model.

It is interesting to notice that DA can highlight elements at different resolution and

with different geological meaning. For instance, while in Figure 3.3.6-b several horizons related to the main velocity variations are apparent, in Figure 3.3.6-c a general overview of the stratigraphy is provided. Moreover, Figure 3.3.6-d and Figure 3.3.6-e provide a clear distinction between the shallow lower velocity zone and the high velocity one in depth, with some slightly different details.

Application on 2-D data: Field Data example

We test the proposed methodology on field data of the 2-D marine seismic profile of the WS10 exploration project, obtained in autumn 2010 in the west Mediterranean Sea by the Istituto Nazionale di Oceanografia e Geofisica Sperimentale (OGS), which also performed the data processing (Geletti et al., 2014). The selected portion of the seismic profile images a rifted margin of the eastern Sardo-Provençal Basin characterized by a faulted salt dome and by a portion of an almost undisturbed sedimentary sequence (Figure 3.3.7).

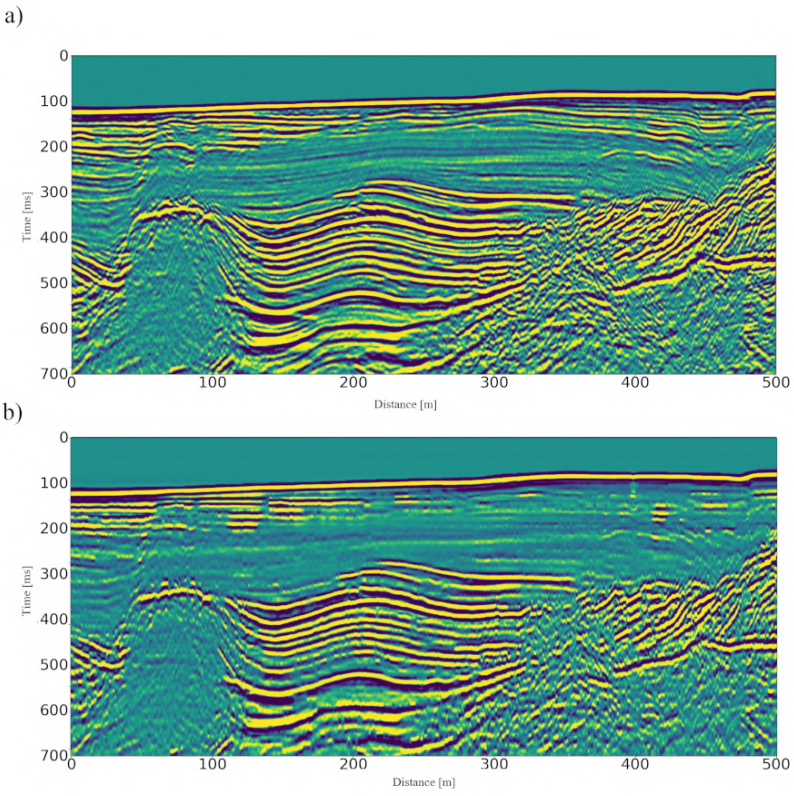

Figure 3.3.7: Example of a seismic profile of the WS10 exploration project: a) field processed data; b) NN reconstruction (output).

The PCA analysis performed allow to condense the main part of DA information (i.e. 92%) in just 4 panels (Figure 3.3.8). Beside the effectiveness in reducing the number of DA panels PCA is quite efficient in keeping both high and low frequency features. In fact, while panels a) and b) provide information about the general geological (and velocity) trends, panels c) and d) give evidence of local structures like the salt dome as well as of the main horizons.

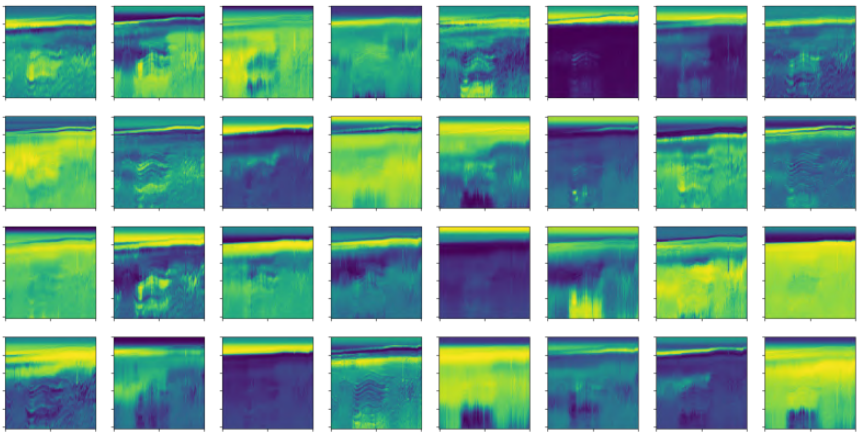

Figure 3.3.8: Results of the PCA analysis on the last layer before the Encoding part of the geometry applied on the same dataset as in Figure 3.3.7a.

Application on a 4-D dataset

The Sleipner 4-D Seismic dataset is a reference dataset from the Sleipner CO_2 storage site. Several seismic surveys have been performed at the storage site since 1994, when the first survey was completed before the commencement of fluid injection. This dataset contains seismic cubes from 1994 to 2010. In 2007 PGS performed a full reprocessing of 1994, 2001, 2004 and 2006 data sets collected with the same geometry, acquisition parameters,with an identical processing flow (EQUINOR, 1999). An example of crossline for the four data sets is presented in Figure 3.3.9.

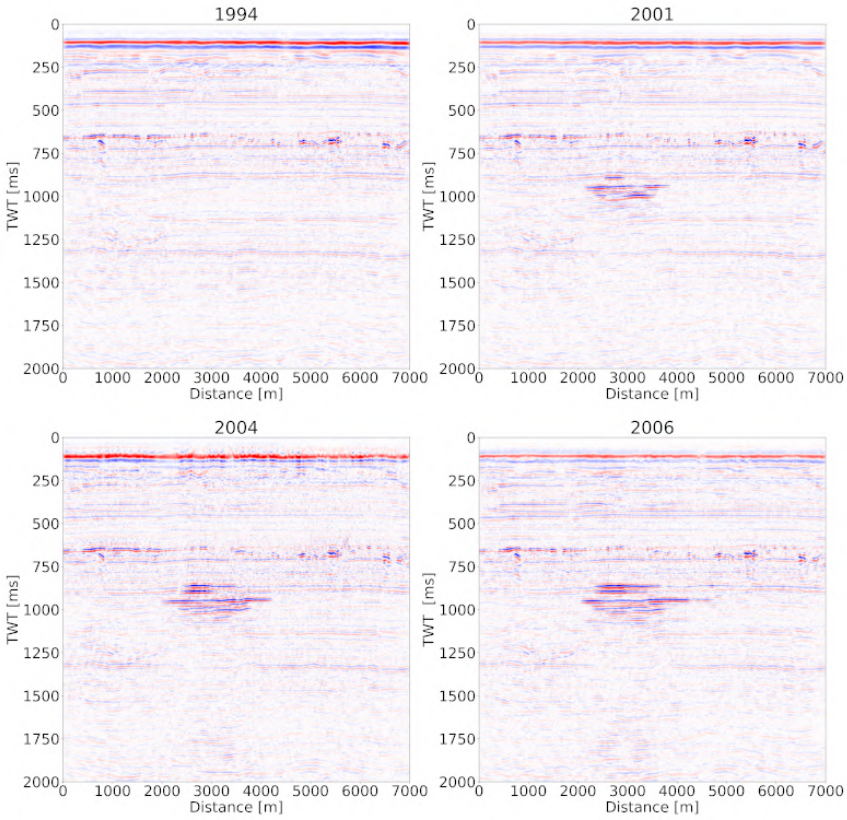

Figure 3.3.9: Example of crossline of 1994, 2001, 2004, 2006 data set reprocessed by PGS in 2007 (EQUINOR, 1999)

We applied the technique to the processed data cube from 2001 acquisition and we selected just a DA that enhances the strong reflection of the injection area. We applied the same trained NN on the data set of all the years from 1994 to 2006 to test the stability of the approach. A time slice showing some of the results is provided in Figure 3.3.10.

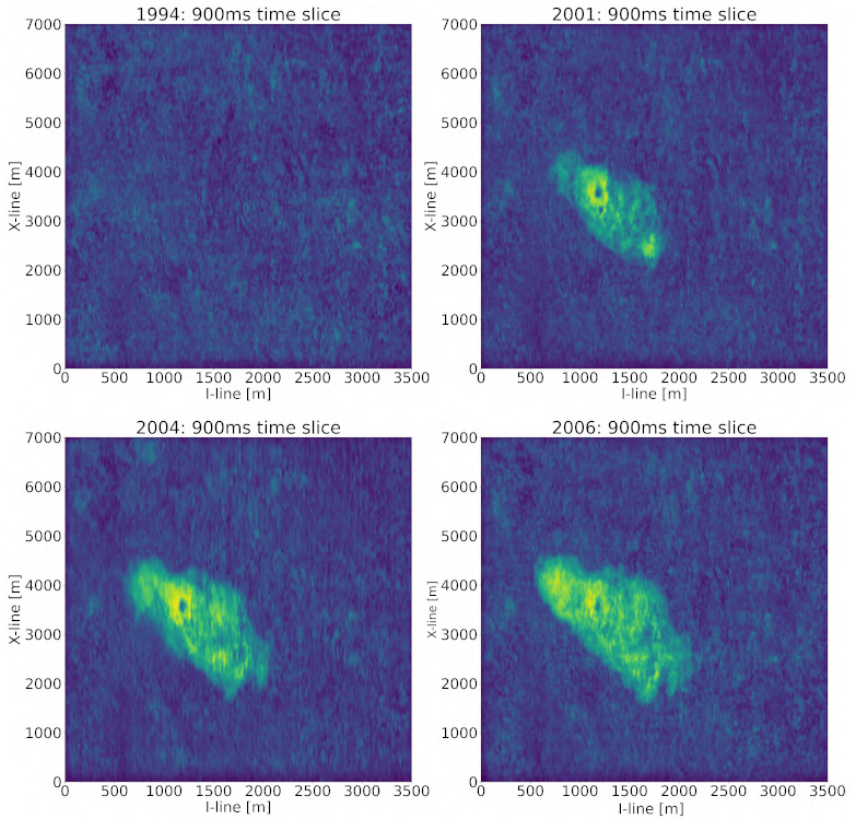

Figure 3.3.10: 900 ms time slice extracted from the 1994, 2001, 2004 and 2006 3-D datasets representing the same single DA. The NN training was performed on 2001 data and then applied to all the other data as a stability test of the whole DA calculation procedure.

As we can see, the methodology is stable over different and independent data sets collected in different years and the single DA plotted is able to highlight the CO_2 saturated area and its limits. Values plotted in Figure 3.3.10 are scaled to the maximum values of more recent 2006 time slice: results evidence that also the magnitude of the prediction is independent from the training data. This is a crucial and not obvious result that points out the high stability and affordability of the whole approach, which is helpful not only for qualitative analyses but also for detailed monitoring purposes.

Moreover, a simple threshold (values > 0) on the DA on the 4-D cube is effective to enhance the evolution of the injected area through time (Figure 3.3.11). While plots in Figure 3.3.10 are very helpful in estimating the real boundaries of the CO_2 saturated

volume, as already discussed, the purpose of the DA threshold is rather aimed at testing the stability of the approach.

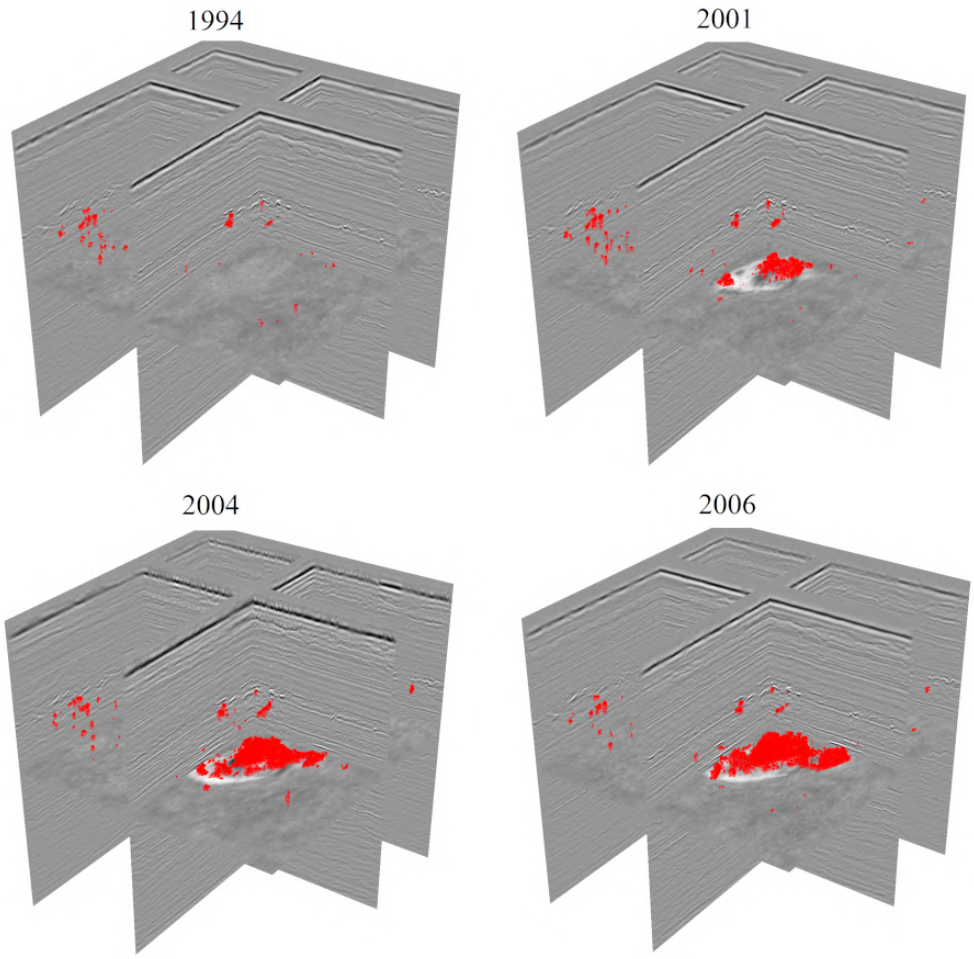

Figure 3.3.11: In red values of the DA plotted in Figure 3.3.10 with values greater than 0 used as a stability test. See text for details.

We remark that the proposed approach is quite different from all the existing seismic attribute analyses, including the ones exploiting different strategies to combine and condense attributes (see e.g. Sain and Kumar (2022)). In fact, PCA, clustering, cross plot, neural network methods are effective in decreasing the total number of attributes but all have to calculate the attributes beforehand and are somehow subjective in terms of the selection of attributes to be combined (Meldahl et al., 2001). The proposed technique, on the contrary, exploits LSTM networks using their intermediate prediction steps, which are usually disregarded, as a way to recover additional information through quantities here referred as DA. Meta-attributes, which are the aggregation of seismic attributes combined with the interpreter's insight through *artificial intelligence*

techniques (Sain and Kumar, 2022) are indeed something different, but future research could be directed on theone hand to combine DA with the experience of the interpreter and on the other to infer a one–to-one correlation between DA and specific physical parameters and/or geological features.

From the computational point of view the methodology is profitable on GPU machines, since LSTM gets a great improvement from cuda implementation. For a single 3-D Sleipner data cube, training time on a laptop with a RTX-1080 with 8Gb of memory takes only few hours for a 116,532 traces training. Prediction time on the same number of traces is just of 2.4 seconds for the whole dataset.

Conclusions

We exploit the Hidden Layers (HL) of LSTM networks as Deep Attributes (DA) of reflection seismic data sets. The proposed methodology is definitely different fromthe ones in which NN are applied to combine/condense previously calculated attributes.

The results obtained from synthetic and field data show that the new method is able to manage even complex geometries highlighting not only single seismic reflectors (i.e. features with high spatial frequency) but, even more important, the main geological and geophysical features related for instance to the low spatial frequency seismic velocity trend. PCA can be successfully applied on DA in order to reduce the total number of attributes and is effective in retaining the information content.

The stability tests performed demonstrate the high affordability of the procedure that can be trained on a single data set (or just on a portion) and then applied to larger seismic volumes. The application of DA to a 4-D data set collected to monitor a controlled CO_2 injection in a underground gas storage further demonstrate the stability and the repetitiveness of the methodology and in turn its full applicability for monitoring purposes.

Future research will be directed to infer specific correlations between DA and single or integrated physical parameters which would make the DA a new possible strategy for the quantitative subsurface petrophysical characterization at different scale and resolution levels.

Computer code availability

The implementation of the Deep Attributes extraction can be found at `https://github.com/Giacomo-Roncoroni/DeepAttributes`.

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 furthermore thanks the PNRA projects IPECA (PNRA18_00186) and CRIOVEG (PNRA18_00288)

Bibliography

Nigel A. Anstey. 2013. *Seismic Interpretation: The Physical Aspects*. Springer Netherlands. https://books.google.it/books?id=beTsCAAAQBAJ

Satinder Chopra and Kurt J. Marfurt. 2005. Seismic attributes — A historical perspective. *GEOPHYSICS* 70, 5 (Sept. 2005), 3SO–28SO. https://doi.org/10.1190/1.2098670

Satinder Chopra and Kurt J. Marfurt. 2007. *Seismic Attributes for Prospect Identification and Reservoir Characterization*. Society of Exploration Geophysicists. https://books.google.it/books?id=dP2iACuzq34C

Matteo Dossi, Emanuele Forte, and Michele Pipan. 2018. Quantitative Analysis of GPR Signals: Transmitted Wavelet, Amplitude Decay, and Sampling-Related Amplitude Distortions. *Pure and Applied Geophysics* 175, 3 (March 2018), 1103–1122. https://doi.org/10.1007/s00024-017-1752-2

EQUINOR. 1999. *Sleipner dataset*. Technical Report. DOI : 10.11582/2020.0000..

Riccardo Geletti, Fabrizio Zgur, Anna Ben, F. Buriola, Silvana Fais, Maurizio Fedi, Emanuele Forte, Arianna Mocnik, Valeria Paoletti, Michele Pipan, Riccardo Ramella, Roberto Romeo, and Alessandro Romi. 2014. The Messinian Salinity Crisis: New seismic evidence in the West-Sardinian Margin and Eastern Sardo-Provençal basin (West Mediterranean Sea). *Marine Geology* 351 (2014), 76–90. https://doi.org/10.1016/j.margeo.2014.03.019

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Computation* 9, 8 (Nov. 1997), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735 _eprint: https://direct.mit.edu/neco/article-pdf/9/8/1735/813796/neco.1997.9.8.1735.pdf.

Sang Hou and H. Hoeber. 2020. Seismic Processing with Deep Convolutional Neural Networks: Opportunities and Challenges. 2020, 1 (2020), 1–5. https://doi.org/10.3997/2214-4609.202010647

Aishwarya Jadhav, Vinay Namboodiri, and K. Venkatesh. 2016a. *Deep Attributes for One-Shot Face Recognition*. 516–523. https://doi.org/10.1007/978-3-319-49409-8_44

Aishwarya Jadhav, Vinay P. Namboodiri, and K. S. Venkatesh. 2016b. Deep Attributes for One-Shot Face Recognition. In *ECCV Workshops*. https://api.semanticscholar.org/CorpusID:35906367

Yongna Jia and Jianwei Ma. 2017. What can machine learning do for seismic data processing? An interpolation application. *GEOPHYSICS* 82, 3 (May 2017), V163–V177. https://doi.org/10.1190/geo2016-0300.1

Elena Klochikhina, Sean Crawley, Sergey Frolov, Nizar Chemingui, and Tony Martin. 2020. Leveraging deep learning for seismic image denoising. *First Break* 38, 7 (2020), 41–48. https://doi.org/10.3997/1365-2397.fb2020048

Kunhong Li, Jingjing Zong, Yifeng Fei, Jiandong Liang, and Guangmin Hu. 2022. Simultaneous Seismic Deep Attribute Extraction and Attribute Fusion. *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), 1–10. https://doi.org/10.1109/TGRS.2021.3113075

Ming Li and Yimin Zhao. 2014. Chapter 5 - Seismic Attribute Analysis. In *Geophysical Exploration Technology*, Ming Li and Yimin Zhao (Eds.). Elsevier, Oxford, 103–131. https://doi.org/10.1016/B978-0-12-410436-5.00005-8

Ping Lu, Yuan Xiao, Yanyan Zhang, and Nikolaos Mitsakos. 2019. Deep learning for 3D seismic compressive-sensing technique: A novel approach. *The Leading Edge* 38, 9 (Sept. 2019), 698–705. https://doi.org/10.1190/tle38090698.1

Gary S. Martin, Kurt J. Marfurt, and Shawn Larsen. 2002. Marmousi-2: An updated model for the investigation of AVO in structurally complex areas. In *SEG Technical Program Expanded Abstracts 2002*. Society of Exploration Geophysicists, 1979–1982. https://doi.org/10.1190/1.1817083

Ben Moseley, Tarje Nissen-Meyer, and Andrew Markham. 2020. Deep learning for fast simulation of seismic waves in complex media. *Solid Earth* 11, 4 (Aug. 2020), 1527–1549. https://doi.org/10.5194/se-11-1527-2020

Feng Qian, Miao Yin, Xiao-Yang Liu, Yao-Jun Wang, Cai Lu, and Guang-Min Hu. 2018. Unsupervised seismic facies analysis via deep convolutional autoencoders. *GEOPHYSICS* 83, 3 (May 2018), A39–A43. https://doi.org/10.1190/geo2017-0524.1

Giacomo Roncoroni, Emanuele Forte, Luca Bortolussi, and Michele Pipan. 2022. Efficient extraction of seismic reflection with Deep Learning. *Computers & Geosciences* 166 (Sept. 2022), 105190. https://doi.org/10.1016/j.cageo.2022.105190

Giacomo Roncoroni, Carlo Fortini, Luca Bortolussi, Nicola Bienati, and Michele Pipan. 2021. Synthetic seismic data generation with deep learning. *Journal of Applied Geophysics* 190 (July 2021), 104347. https://doi.org/10.1016/j.jappgeo.2021.104347

Roberto Ruiz, Anna Roubickova, Cyrille Reiser, and Neelofer Banglawala. 2021. Data mining and machine learning for porosity, saturation, and shear velocity prediction: recent experience and results. *First Break* 39, 7 (2021), 71–76. https://doi.org/10.3997/1365-2397.fb2021055

Brian H. Russell. 1988. The Convolutional Model. In *Introduction to Seismic Inversion Methods*. Society of Exploration Geophysicists. https://doi.org/10.1190/1.9781560802303.ch2

Kalachand Sain and Priyadarshi C. Kumar. 2022. *Meta-attributes and Artificial Networking: A New Tool for Seismic Interpretation*. Wiley. https://books.google.it/books?id=IWJ6EAAAQBAJ

Benfeng Wang, Ning Zhang, Wenkai Lu, and Jialin Wang. 2019b. Deep-learning-based seismic data interpolation: A preliminary result. *GEOPHYSICS* 84, 1 (Jan. 2019), V11–V20. https://doi.org/10.1190/geo2017-0495.1

Wenlong Wang, George A. McMechan, and Jianwei Ma. 2020. Elastic full-waveform inversion with recurrent neural networks. In *SEG Technical Program Expanded Abstracts 2020*. Society of Exploration Geophysicists, Virtual, 860–864. https://doi.org/10.1190/segam2020-3425921.1

Yuqing Wang, Qiang Ge, Wenkai Lu, and Xinfei Yan. 2019a. Seismic impedance inversion based on cycle-consistent generative adversarial network. In *SEG Technical Program Expanded Abstracts 2019*.

Society of Exploration Geophysicists, San Antonio, Texas, 2498–2502. https://doi.org/10.1190/segam2019-3203757.1

Xinming Wu, Luming Liang, Yunzhi Shi, and Sergey Fomel. 2019. FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation. *Geophysics* 84, 3 (04 2019), IM35–IM45. https://doi.org/10.1190/geo2018-0646.1 arXiv:https://pubs.geoscienceworld.org/geophysics/article-pdf/84/3/IM35/4682641/geo-2018-0646.1.pdf

Oz Yilmaz. 2001. *Seismic Data Analysis* (2 ed ed.). Investigations in Geophysics, Vol. 10. Society Of Exploration Geophysicists, Tulsa, USA. https://doi.org/10.1190/1.9781560801580

Chapter 4

Recurrent Task

4.1 Synthetic seismic data generation with deep learning

In this article, we describe a new methodology for generating synthetic seismograms using deep learning methods. This approach is based on training a LSTM-NN on a large dataset of synthetic seismic data, generated using Finite Differences, to generate synthetic seismograms from 1D models of the subsurface without solving the wave equation.

The methodology can also be made more flexible by allowing changes in wavelet and acquisition geometry. This enables us to generate synthetic seismograms that are tailored to specific subsurface models and to different acquisition scenarios.

In this article, we compare the results of this methodology with some published ones to demonstrate its effectiveness. We also discuss the potential applications of our methodology, including its use in solving the inverse problem of seismic wave propagation.

Synthetic seismic data generation with deep learning

G. Roncoroni, C. Fortini, L. Bortolussi, N. Bienati, M. Pipan

Journal of Applied Geophysics, Volume 190, July 2021, 104347
doi: doi.org/10.1016/j.jappgeo.2021.104347

Key Words

Reflection Seismic; *Seismic Modelling*; *Deep Learning*; *Machine Learning*

Highlights

- DL to generate 1-D acoustic synthetic seismograms without solving wave equation
- Solution to the 1-D problem through custom Recurrent Neural Network
- Retraining strategy to improve flexibility and applicability
- Computational complexity analysis

Introduction

The main objective of this work is the implementation of Deep Learning (DL) solutions to generate synthetic seismograms from 1D acoustic models without solving the wave equation. This is done by training a NN model which after training is able to predict common shot gathers from 1-D velocity models.

The wave equation, is non linear with respect to velocity. Numerical solutions schemes, such as finite difference or pseudo-spectral schemes, are computationally demanding and the results may be affected by dispersion error and boundary reflections.

The NN approach is computational demanding only during NN training and is affected by the problems in the training dataset. The proposed methodology can solve the forward problem of seismic wave propagation faster than classical methods, especially when high-frequency source wavelets are considered.

Although direct application and computational advantages, being able to simulate data with a simple 2D geometry but taking into account the multiples reflection, leads the path to possible future works for a better and faster solution of the inverse problem.

The generation from text data of an audio track that simulates the human voice, is a similar problem that has been already tackled by DL approaches (see e.g. Oord et al. (2016)). The application of DL to sciences, and in particular to geophysics, is

a new branch that has been rapidly developing in the last years benefiting to the big improvements in the cost-to-performance ratio of computational resources. The power of Deep Learning is due to the capacity of finding recurrent patterns or causality without the need of user judgment or explicit coding by capturing statistical relations to provide generative or discriminative models.

Øye and Dahl (2019) proposed a method based on Convolutive Neural Network (CNN) to estimate velocity models from Raw Shot Gathers, with possible applications to Full Waveform Inversion. This is a new approach that can overtake the computational limit of the old methods (such as genetic algorithms) used to estimate a velocity model. Also Yang and Ma (2019) developed a technique to build seismic velocity models using U-net. Guo et al. (2019) used Bi-directional Long Short-Term Memory (LSTM) for seismic impedance inversion, an ill-posed and non-linear problem. The Bi-directional LSTM recurrent neural network applied to the inverse problem of P-impedance estimation can be attractive because it does not need a prior estimate of the wavelet, one of the main problems in this type of studies.

Moseley et al. (2018) proposed a wavenet-like net Oord et al. (2016), a stack of causal CNN layers, for a fast approximate simulation of acoustic waves.

The methodology proposed in this work considers the seismogram as a time-series (as in speech generation) rather than a matrix of pixels (as in image processing). We have used a custom made Recurrent Neural Network Sherstinsky (2018) model to implement such strategy.

The result is a model able to learn how to solve the problem in 1-D-layered media and capable of an approach to increase flexibility and to reduce the computational cost in the training process.

The final goal is to speed up forward modelling for the wave equation with respect to the classical methodology and to provide a starting point for future applications to more complicated scenarios, e.g. elastic approximation, or to a new approach to the inverse problem.

Methods

Datasets

We used two different datasets to train NN and to evaluate the importance of the dataset on different training and prediction features: the first taken by Moseley et al. (2018), hereinafter referred to as MOS, which was used to tackle the same task, the other generated by numerical simulation [with the codes provided in the Devito Project by Louboutin et al. (2019)] , hereinafter referred to as DEV.

Both datasets consists of two tensors (training/validation): two for the input and two for the output data. The input tensor is made of a number of velocity functions

(10,000) and the output tensor of Common Shot Gather panels (CSG; 10,000 panels of 11 seismic traces for MOS data and 257 seismic traces for DEV data).
We split both test datasets, made of 10,000 input velocity functions and 10,000 output CSG, in training and validation datasets (80% the former and 20% the latter).
We will hereafter use the notation tensor for multi-dimensional arrays: in particular, the output of the NN is a multi-dimensional array with size [10000, 257, 1000].
MOS is available on GitHub Moseley et al. (2018). It is made of a tensor with velocity functions and a tensor with the associated CSG panels. The 50% reduction in size of the dataset, compared to the original 20,000 sample Moseley's dataset, is due to the fact that our model is based on a reduced number of weights , compared to Moseley's one, and does not require large datasets. We have applied a random data selection and splitting procedure.
The CSG panel was computed by solving the wave equation by using finite differences over a 3Km x 3Km grid in a horizontally layered model. There are 11 evenly spaced receivers, with horizontal spacing of 200m, and the source is at the 6th receiver position.
MOS and DEV mainly differ in number of receivers and source-receiver offset, i.e, DEV is the result of numerical simulation with the parameters shown in Table 4.1.
We randomly generated different subsurface models by using the velocity and depth constraints shown in Table 4.1.
In order to generate the input velocity function for FD (in depth) and for NN (in time) we used the following workflow:

for each velocity function:
n = generate a casual number of horizons
generate n random depth values
sort depth values and append 0 at the beginning
generate n + 1 velocity values
build velocity function in depth
convert depth to time
build velocity function in time
Normalize data between 0 (1000 m/s) and 1 (6000 m/s) "

Data in DEV dataset has 257 evenly spaced receivers, with horizontal spacing of 5m, and the source at the 129^{th} receiver position.

Parameter	Set values
Sampling rate [s]	0.001 s
Listening time [s]	1 s
Grid point [x, y]	[600, 600]
Grid spacing [min, max]	[5, 5] m
Ricker peak f [Hz]	20 Hz
Velocity [min, max]	[1500 m/s, 5000 m/s]
Depth [min, max]	[100m , 3000 m]
Number of layers	[3-8]

Table 4.1: Parameters used to produce the second dataset

NN architecture

We tested several architectures, varying the number of Hidden Layers (HL), neurons and with different number of filters and different kernel sizes in the convolutional layer. After an extensive tests phase, we were able to define the best performing architecture, shown in Fig. 4.1.1. This NN consists of two LSTM layers Hochreiter and Schmidhuber (1997), a convolutional layer Simard et al. (2003), and a final LSTM layer. We choose LSTM because they are recurrent NN, which means they can deal with causal tasks, and they have a long short-term memory, so they should better represent the physics behind the task.

All of the LSTM layers have a number of neurons equal to the number of sensors in the recording array to be simulated. The LSTM used in this model are implemented by the Keras layer CuDNNLSTM, a Fast LSTM implementation with NVIDIA CUDA Deep Neural Network library (cuDNN), a GPU-accelerated library of primitives for Deep Neural Networks (DNN) Chetlur et al. (2014). In the convolutional layer the kernel size is [64, 4] and it has 4 filters with a ReLu activation function Glorot et al. (2011). This gives the NN the flexibility to get values in the window [t - 4 : t] from the 64 closer neurons, whit t the time steps, and to store it in 4 different filters.

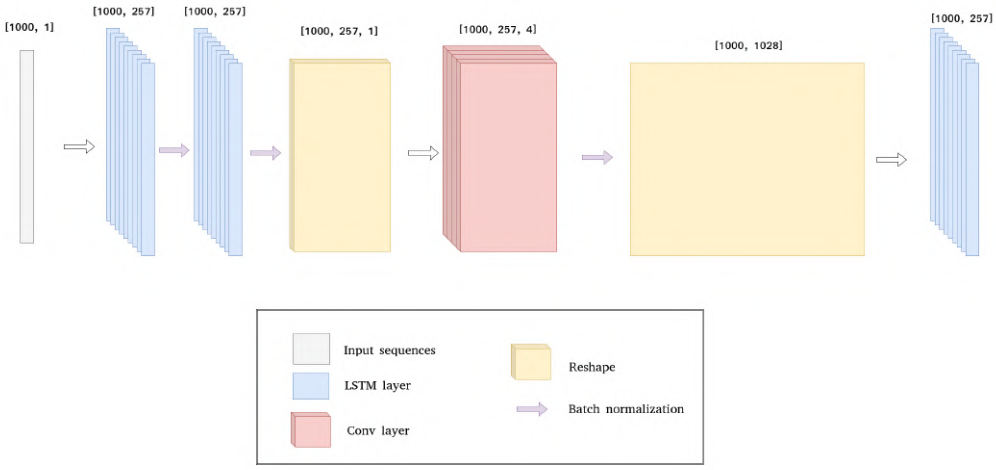

Figure 4.1.1: The NN is made of two Long Short Term Memory (LSTM) layers, a Convolutive layer, and a final LSTM layer.

The proposed DL architecture (Fig. 4.1.1) provides an output h at time t from three different input, namely: Xt, h(t-1) and the cell state line. The latter is from the method used to obtain the long term memory. The output is calculated from a series of weights, activation functions and simple mathematical operations. In case of overlapping signals, such as e.g. multiple/primary reflections, such architecture is unable to perform correct predictions at times larger than the overlap point. This is the reason for the introduction of the CNN layer: it increases the number of input to the same neuron thus allowing a correct prediction even in the post-overlap part of the record. Fig. 4.1.2 A,B,C illustrate the performance of such solution during the training: in Fig. 4.1.2 A,B (without Convolutive Layer) the loss function is stable from epoch 320 and the problem is not solved. Fig. 4.1.2 C (with Convolutive Layer) shows a successful prediction from epoch 180.

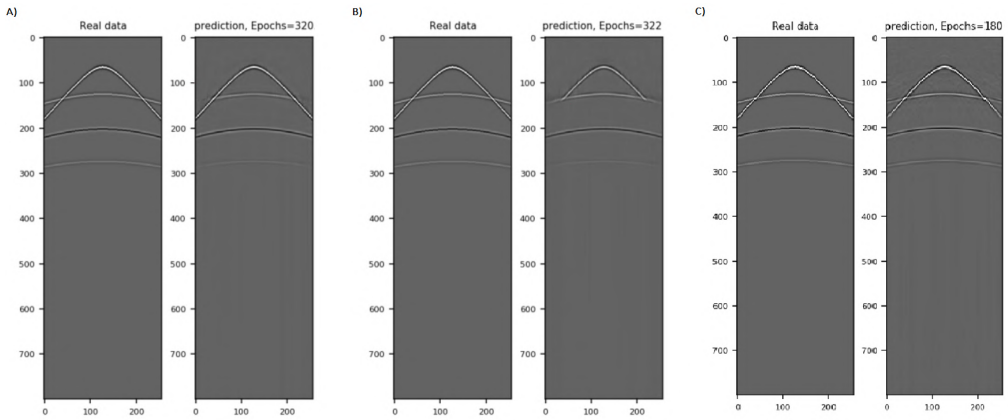

Figure 4.1.2: Performance on overlapping signals during the training. Prediction without Convolutive Layer (A, B), at epochs 320 and 322 respectively, and with Convolutive Layer (C), at epochs 180.

In order to train the NN, we need to minimize the norm of the difference between the prediction and the reference output, this is called loss function: the most used loss functions in regression tasks are the mean square error and the mean absolute error. However, in our scenario, geometrical spreading and transmission losses cause an amplitude decay with time, such that the loss function is sensitive to the direct wave only due to its dominant amplitude and NN outputs zeros. Accordingly, in the first implementation, the NN was not able to predict reflections because they did not affect the loss function enough.
In order to sort this problem out we redefined the loss function as

$$loss \ = || \ y_{true} - y_{pred} \ ||_1 \cdot G \tag{4.1}$$

with G defined as a quadratic function $G \ = \ \dfrac{x^2}{\tilde{x}}$ where x is the two-way traveltime $\tilde{x}$ the median on the whole domain of the function $y = x^2$.
After computing the loss function, the error is backpropagated and the weights on each neuron are updated by using the backpropagation algorithm. For this task we tested several algorithms, namely Adam Kingma and Ba (2014), Adamax Kingma and Ba (2014), AdaGrad Kingma and Ba (2014) and SGD Nemirovski et al. (2009) and we eventually chose a special case of the Adam algorithm, AdaMax Kingma and Ba (2014), which is the most used for regression tasks and exhibits better performance for this problem. We tested also different batch sizes and we got best stability in loss function with batch size 512.

NN training

We implemented the NN by leveraging on the Keras library Chollet et al. (2015). We

trained the model on two NVIDIA Tesla P100 GPU accelerators.
All the images shown in this paper are predicted from blind data, which means that the net was not trained on them.
Hidden layer geometry is the same, but we used different numbers of neurons in the output layer to reproduce the number of geophones.
Fig. 4.1.3 shows the structure of the training process with the NN seen as a black box. We split the generated dataset, made of 10,000 samples, in training and validation datasets (80% the former and 20% the latter). With samples we are referring to a the vector of velocity values and its associated CSG panel. The proposed model is based on few weights respect to Mosley's model, so we decided to use half dataset. Data selection and splitting is done randomly and no specific samples were selected.

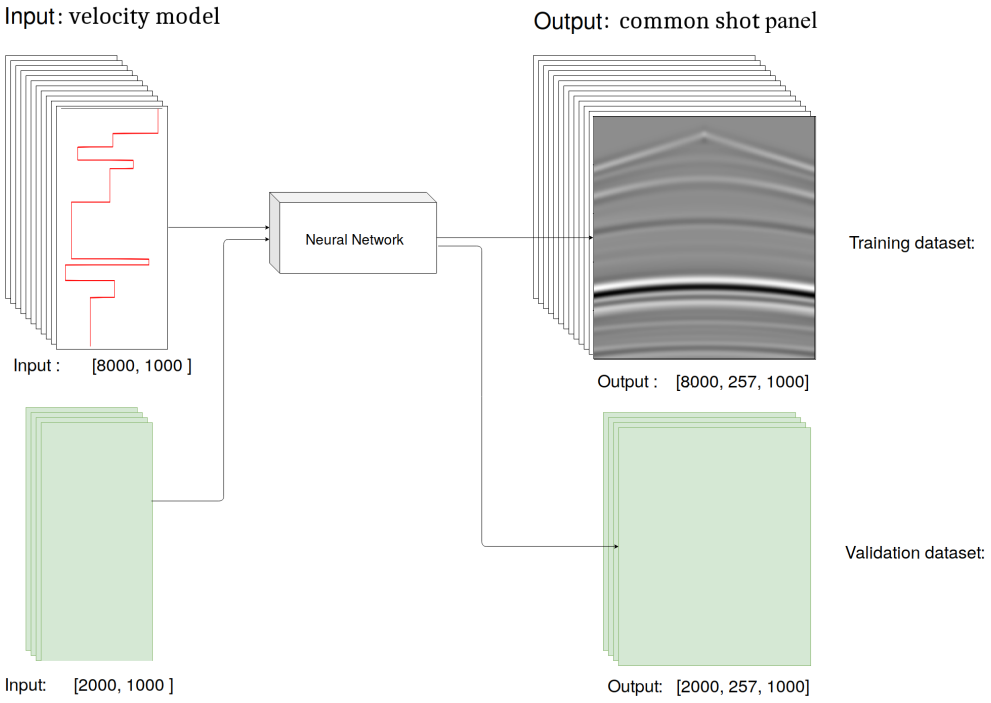

Figure 4.1.3: The dataset is split in four tensors: two for the input and two for the output data (training and validation dataset, respectively). Input tensors are made of a number of velocity functions and output tensors of Common Shot Gather panels (CSG).

A major drawback of such approach is the limited flexibility, i.e. the trained NN is able to predict a number of different situations but source-receiver offset spacing and input wavelet are limited to the ones it was trained with. This can overly extend the training time and reduce the profitability of the method.

Therefore, we developed an alternative strategy by retraining the model, first trained with one of the two test datasets, with smaller datasets that included new features to avoid the need of a complete training for each modified feature.
Due to the fact that the NN has to internally encode information like input wavelet and offset spacing into its weights, a change of simulation parameters should not vary deeply the main distribution of weights.
By doing so the function to be minimized by the NN should be more linear Kavzoglu (2009) and it should need less epochs to converge to a good result.
The training process on MOS dataset is shown in Fig. 4.1.4, in which the loss is plotted against epochs of training, we applied Early stop criteria as Keras Callback.

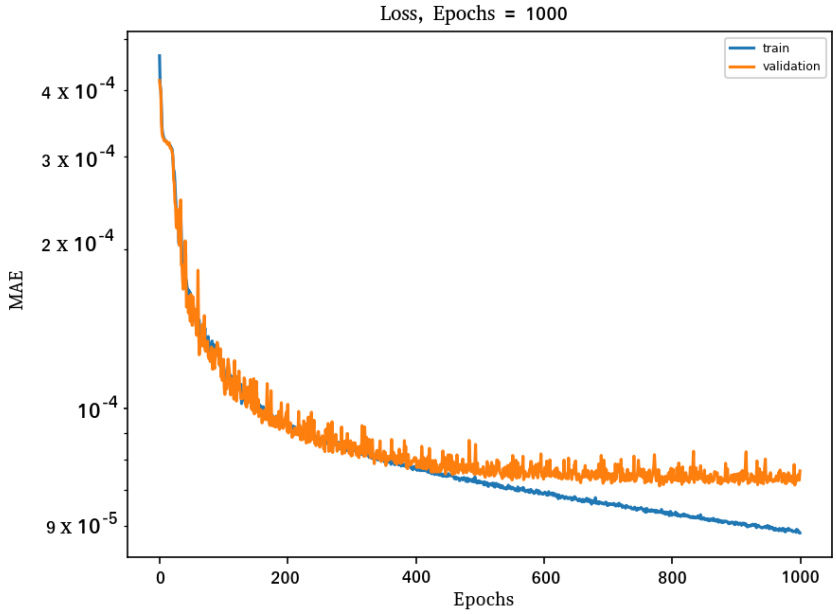

Figure 4.1.4: Training of the proposed net on the MOS dataset: Loss vs Epochs plot. Orange line represents the trend of validation dataset prediction while blue line shows the training one.

The training process on DEV dataset is shown in Fig. 4.1.5: validation and training loss do not diverge too much and they are close enough to each other, which means prediction is good and it is not overfitting.

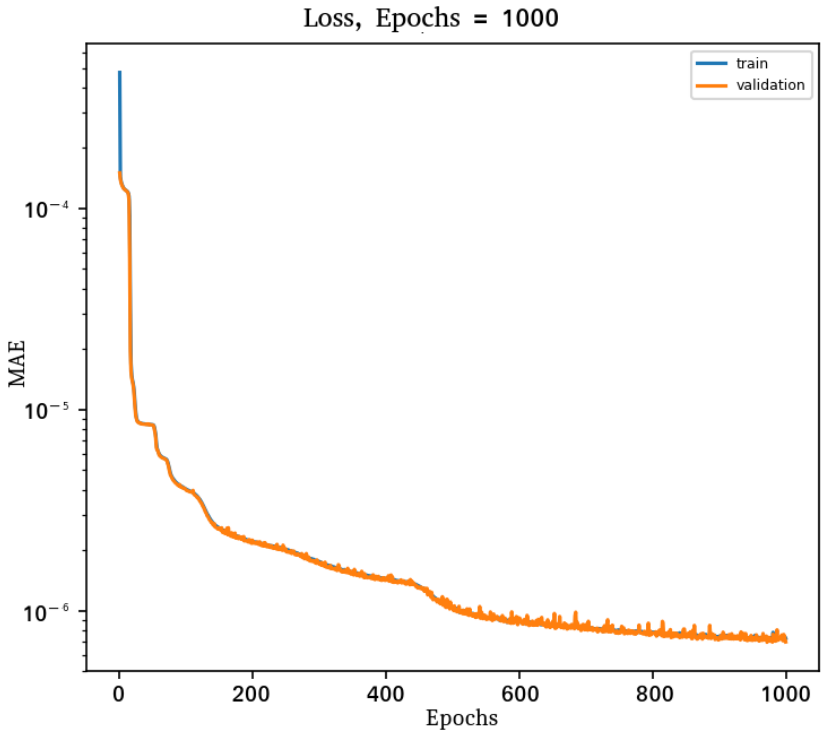

Figure 4.1.5: Loss vs Epochs of the training on DEV dataset. Orange line represents the trend of validation dataset prediction while blue line shows the training one.

Results

MOS Dataset

After the training phase, we tested the net on the dataset generated by Moseley et al. (2018) to evaluate the capability of the net to generalize the problem, i.e. to make correct predictions from datasets different from the training ones.

In this case the training lasted 40 hours on two NVIDIA Tesla P100 GPU accelerators and the loss decreased quite smoothly.

After the training, we tested the NN on a case not utilized in the training phase: an example of this prediction and of the corresponding residuals is shown in Fig. 4.1.6. The direct wave is well predicted; there are still small errors in the magnitude of the gained reflections, but the peak positions are correctly predicted.

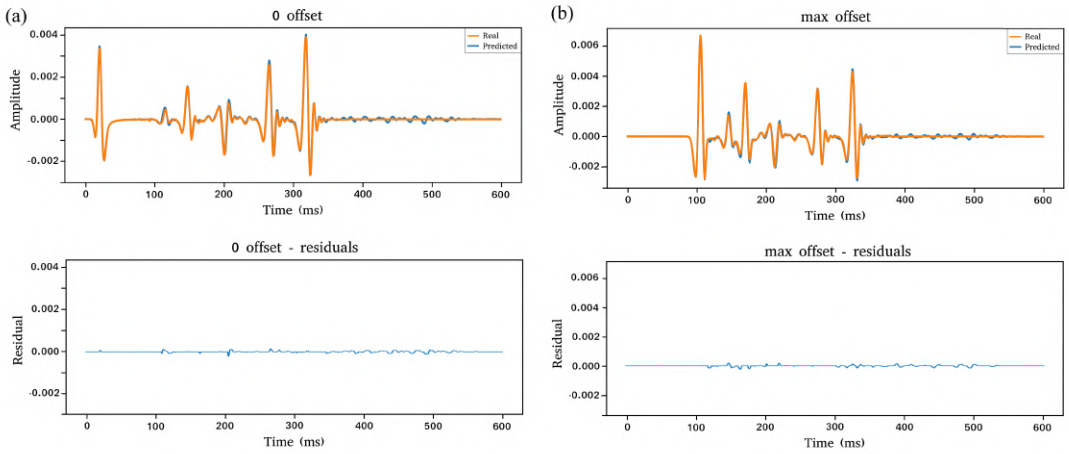

Figure 4.1.6: Prediction and residuals of a predicted CSG panel. 0-offset prediction and residuals (a), max-offset prediction and residual (b). In red we highlighted the phase shift in the residuals.

In Fig. 4.1.6 we can split the analysis of the results in two parts, focusing the part from 0 *ms* to 350 *ms* and from 350 *ms* to the end.

The first part of the data contains primary reflections and we find some small phase shifts, in red boxes in Fig. 4.1.6, but the prediction is good. It is able also to accurately reproduce interference between signals.

The second part is characterized by the presence of multiple reflections only and noise and no primary reflections: in this case the NN does not reproduce anything. This problem can be found in all the dataset and it is present also in the Moseley et al. model Moseley et al. (2018) . This lack of prediction is due to an inadequate statistics on such event in the training dataset.

In order to evaluate prediction performances of NN, we predicted 1000 panels from the dataset and we analyzed the error by plotting the residuals of both nets on a histogram with a logarithmic vertical axis (Fig. 4.1.7) and the horizontal axis divided in 50 bins linearly spaced from 0 to 0.001: we can see that using LSTM we got a lower error.

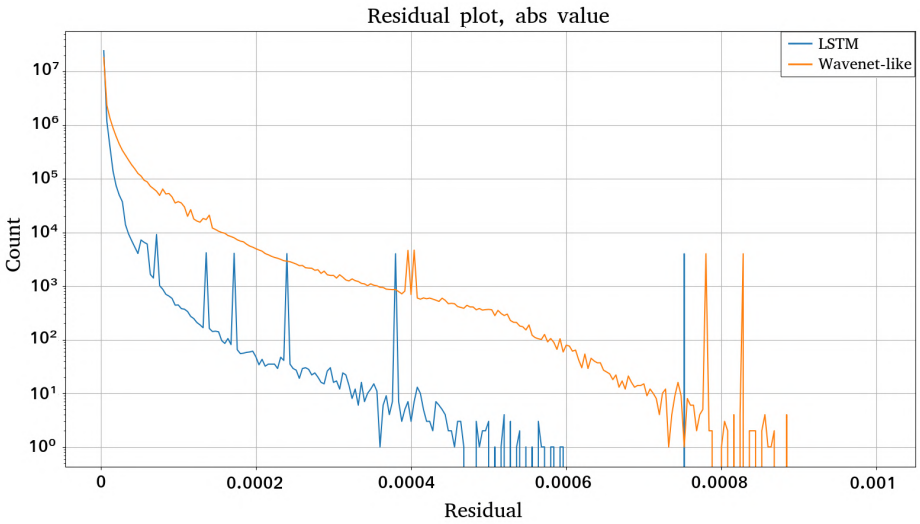

Figure 4.1.7: Residual plot of both models: LSTM model in blue, WaveNet-like Oord et al. (2016) model in orange. Vertical axis is logarithmic. Horizontal axis, called residual, is made of 50 linearly divided bins in a range of 0 to 0.001 (MAE).

We further compared the reference data and the results of prediction of our net in the frequency domain. In Fig. 4.1.8 we can see the amplitude spectrum of reference data (in blue) and predicted data (in orange) and residual of 1,000 CSG panels. Results are satisfactory as the mean absolute error is 0.89%.

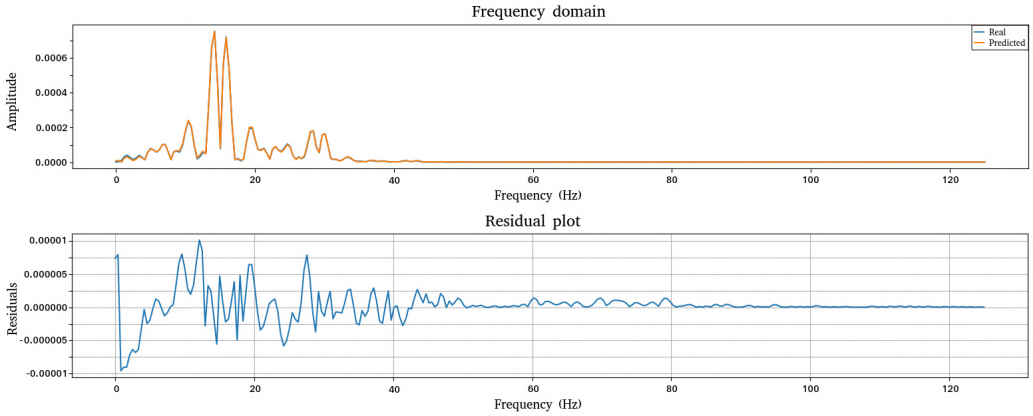

Figure 4.1.8: Reference dataset (in blue), predicted data (in orange) and residual of 1,000 CSG panels in frequency domain

DEV Dataset

We trained the net on another dataset obtained from numerical simulation to have better control of the training process. That allow us to study in depth the importance

of the dataset and test different critical situations. In order to do this, we kept the offset spacing and wavelet fixed, but we varied the number of layers in the 1-D model. In Fig. 4.1.9 are presented two predictions with 256 offsets on two independent velocity profiles: we show the input profile on the left, the desired output in the middle, and the predicted output on the right.

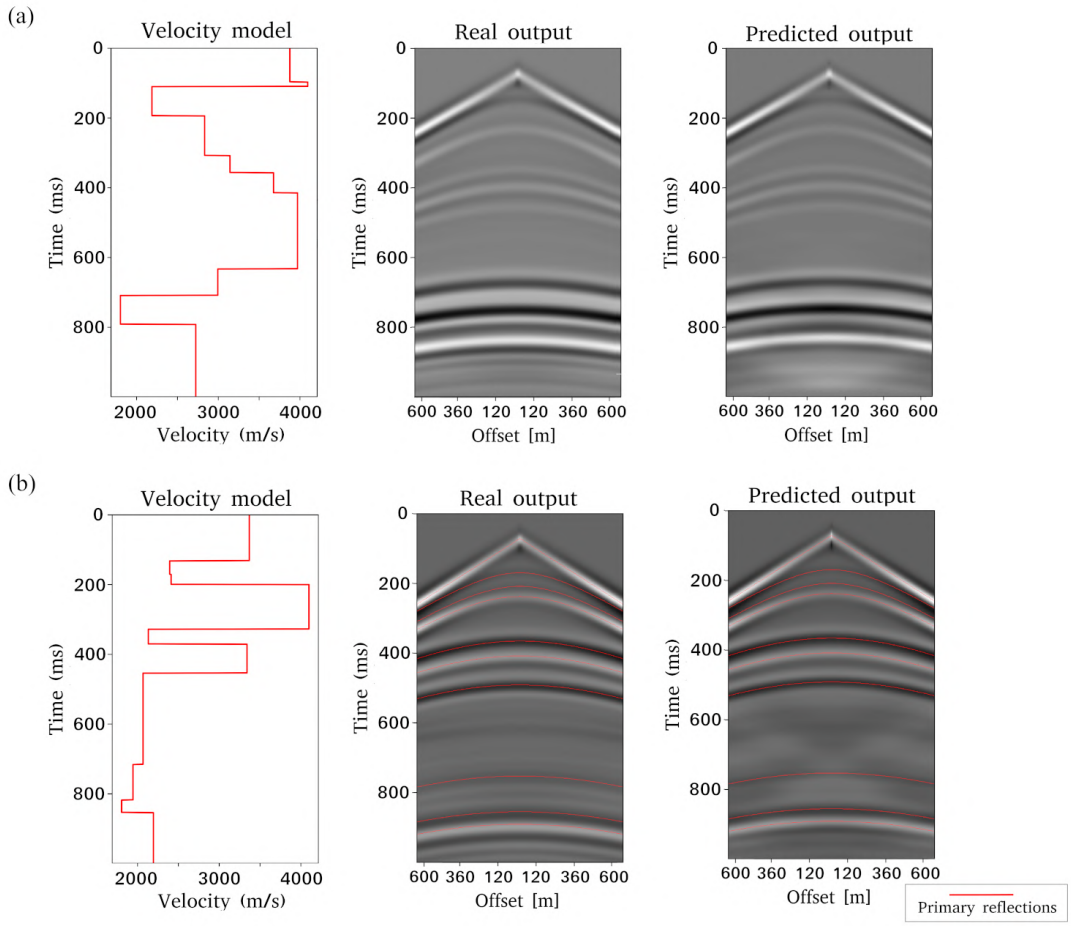

Figure 4.1.9: Two predictions with all 257 offset on two independent profiles. Both are with the input profile, on the left, the desired output, in the middle, and the predicted output, on the right.

In Fig. 4.1.9-a, prediction is very close to the target, except for the central area at 950 *ms*, in which we find a small positive amplitude error.
In Fig. 4.1.9-b, we have a CSG with multiple reflections: we have plotted the 2-way traveltime of the primary reflections over the predicted data, so that we can easily identify non-primary reflections. We can see a multiple reflection predicted from the model at 600 *ms*.

In the panel we can also see a signal, similar to the one generated by reflection from edges, reproduced by the net. Prediction time for 257 offset CSG (DEV dataset) is 0.064s.

The role of the training dataset

Because the NN does not model data from equations, it has to learn from the output of a forward model how an input of the forward model is related to its response. This leads to a crucial role of the dataset used: a bad dataset will lead to a bad model as the net will learn all the features that are in training data.

In Fig. 4.1.10, we have a critical level of numerical dispersion, marked with A and B in Fig. 4.1.10. Despite this, the net is able to accurately reproduce such phenomena.

It is interesting that such predictions come from the same trained net that made the prediction shown in Fig. 4.1.9: this means that the net is also able to discern when and how dispersion has to be predicted. The NN can discern different scenarios and predicts numerical dispersion quite accurately.

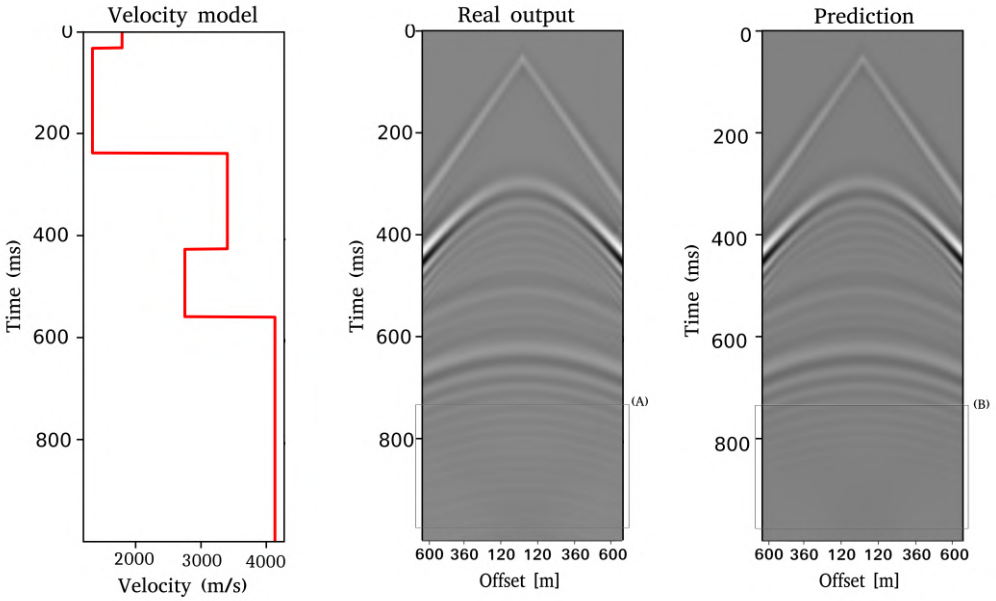

Figure 4.1.10: Reference output and prediction of a CSG panel with numerical dispersion problems due to Finite Differences grid dispersion (A and B).

The model had problems in predicting multiple reflections, as presented in the previous section, and we found a problem due to inadequate statistics on such events in the training dataset.

At the beginning the model was trained on a subsurface model with 7 layers: this lead

to a very low probability in finding multiple reflections and the net seemed not to learn what these signals were due to. In order to make an event important in a dataset we had to find a lot of examples of this, and it has to be of significant amplitude compared to others: in this dataset multiple reflections were too sparse and of low amplitude to affect it significantly.
To try to solve this problem, we retrained the already trained model on a dataset generated on purpose to contain several multiple reflections. We imposed a 3-layers subsurface model with the first two layers closer to each other than the third one. This led to a dataset with lots of statistics on multiple reflections.
In Fig. 4.1.11 we can see: the input, the reference output, the prediction of the retrained model and the prediction on the original model.
As we can see in the retrained model prediction, a multiple reflection at 500 ms is close to the reference signal, while the old net did not predict this signal.

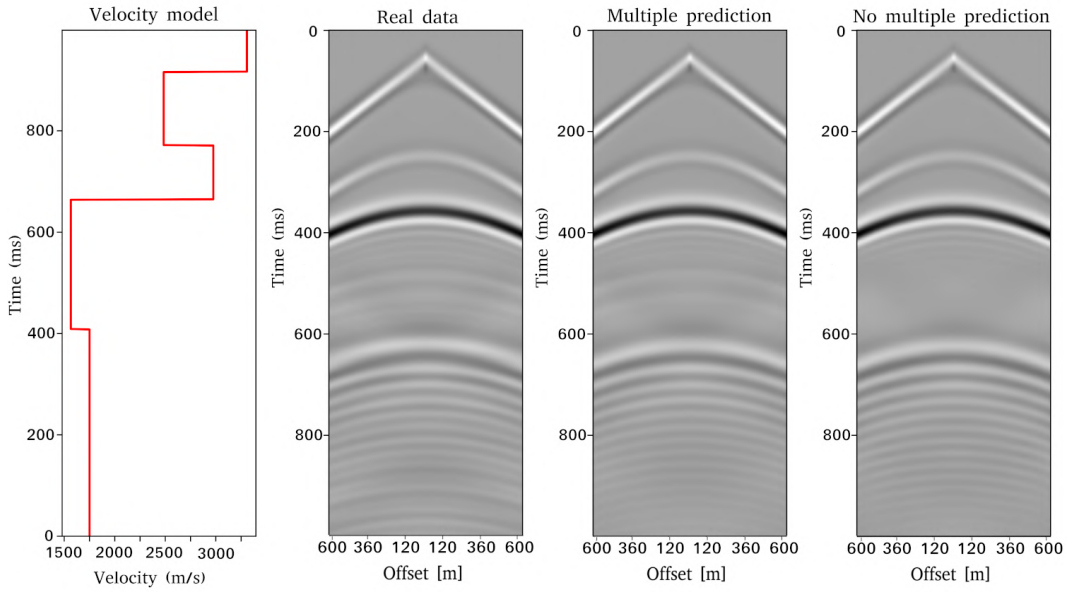

Figure 4.1.11: Input, reference output, prediction of the retrained model and prediction of the original model

Retraining

In order to improve the applicability of the proposed NN, we had to make it more flexible by allowing rapid inclusion of information like waveform and offset spacing.
A short scheme of this process follows:

1. Generate a new small dataset $\sim 10\%$ of the dimensions of the first dataset with the new features.

2. Train the previously trained model on the new dataset.

As we can see from loss vs epochs plot, in Fig. 4.1.12, the new training converges quickly to a minimum and the space in which the gradient is computed seems to be smooth.
It is apparently not necessary to create new set of weights, because the old prediction is already in the range of the new right solution: a limited amount of additional work is enough to reach the best solution as it reaches a minimum in less than an hour.

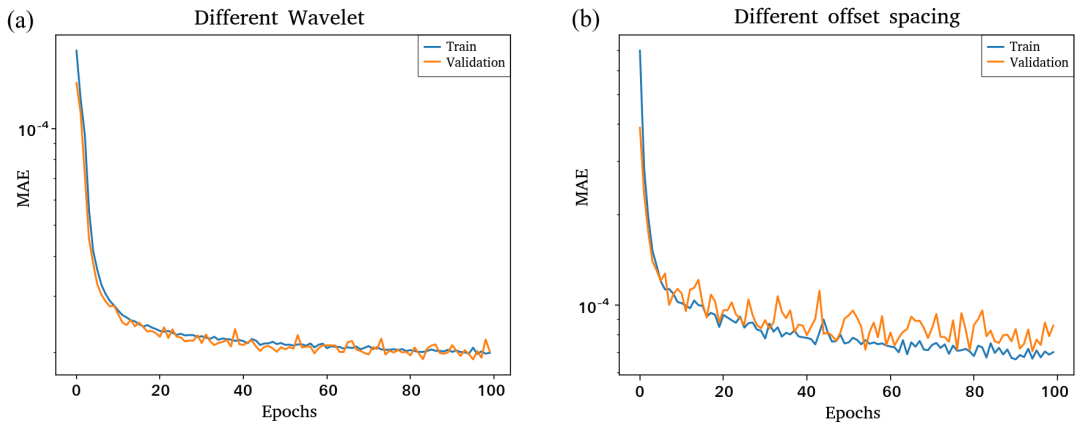

Figure 4.1.12: Loss vs epochs for the retraining of a model with a different waveform (a) and different offset spacing (b).

Fig. 4.1.13-a compares the old prediction (on the right) and the new prediction (in the middle), after the retraining, with a different waveform. While in the old model the source wavelet was a Ricker with peak frequency 10 Hz, in the retrained model we used a higher 20 Hz frequency.
As we have done for the waveform, we have trained a model also to predict a different offset spacing: the results are shown in Fig. 4.1.13-b.

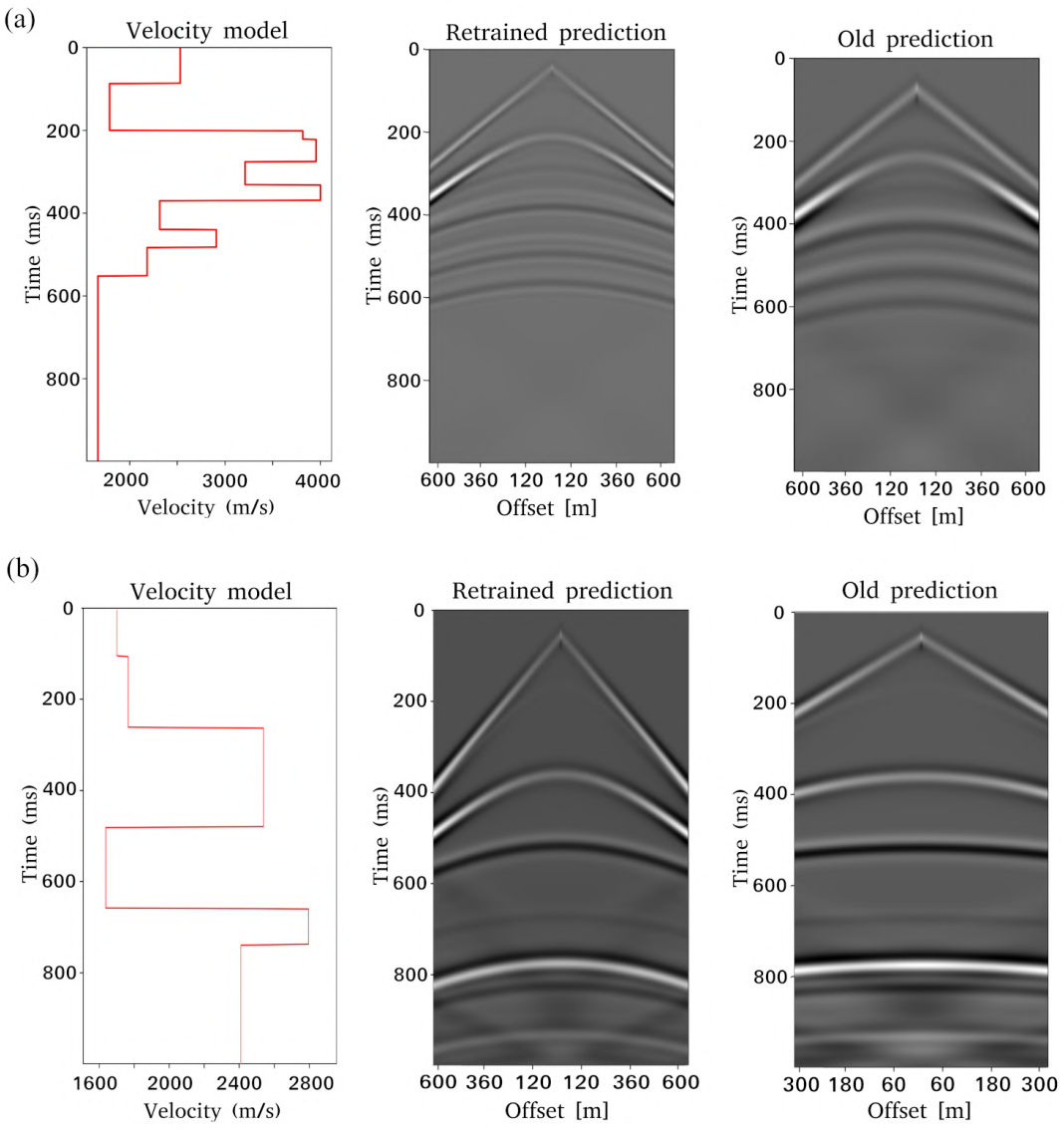

Figure 4.1.13: Prediction of an independent velocity function of the retraining of a model with a different waveform (a) and different offset spacing (b). Old prediction (on the right) and the new prediction (in the middle).

Computational Complexity

We measured the time required to train 500 epochs as a function of the dimension of the desired output. To do this we trained the same net increasing the number of columns (offsets) of the output matrix as $2^n + 1$.

In order to evaluate the net performance, we checked also the trends of the loss as a function of time: this is plotted in Fig. 4.1.14.

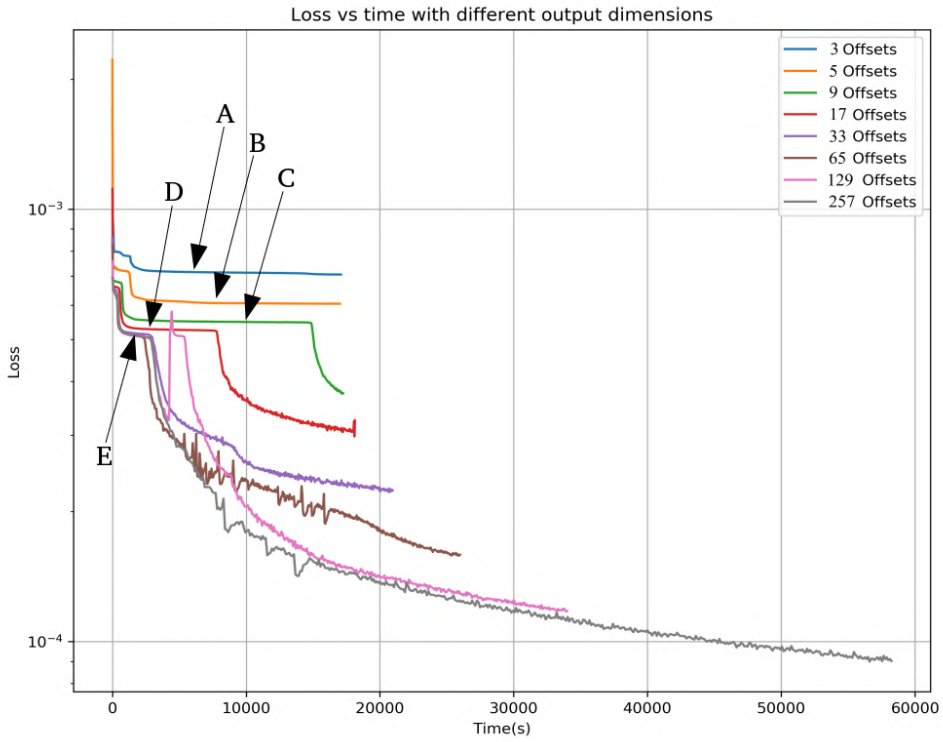

Figure 4.1.14: Loss vs time for different offsets. We can notice a plateau at $(5-8)\times10^{-4}$. With letter A, B, C and D we referred to the plateau of, respectively, 3,5, 9, 17 offsets. With letter E we show the plateau of the remainder data.

As we can see we have different trends but all of them show a plateau. If we analyze the behavior of the function in the surroundings of such plateau we find out that it is the point at which the net really learns to solve the problem, as shown in Fig. 4.1.15. Up to a 9-offset threshold the net is not able to solve the problem in 500 epochs; then it seems to learn faster when the number of offsets increases.
Total training time and time required to solve the problem are reported in Table 4.2.

Offsets	Total time (h)	Solving time (h)	Prediction time (s)
3	4.80h	not solved	0.017s
5	4.74h	not solved	0.015s
9	4.80h	4.00h	0.015s
17	5.05h	2.16h	0.019s
33	5.83h	0.80h	0.021s
65	7.24h	0.60h	0.024s
129	9.46h	0.80h	0.033s
257	16.21h	0.80h	0.064s

Table 4.2: Total training time, the moment where the problem starts to be solved and prediction time for each number of offsets considered.

In Fig. 4.1.15 we can see the exact timing when the problem was understood with 5, 129 and 257 offsets. We got this timing by looking at the predictions made during the training of the NN at the time the loss exited the plateau. We can also see that 500 epochs are not enough to solve the problem if the number of offset is 5.

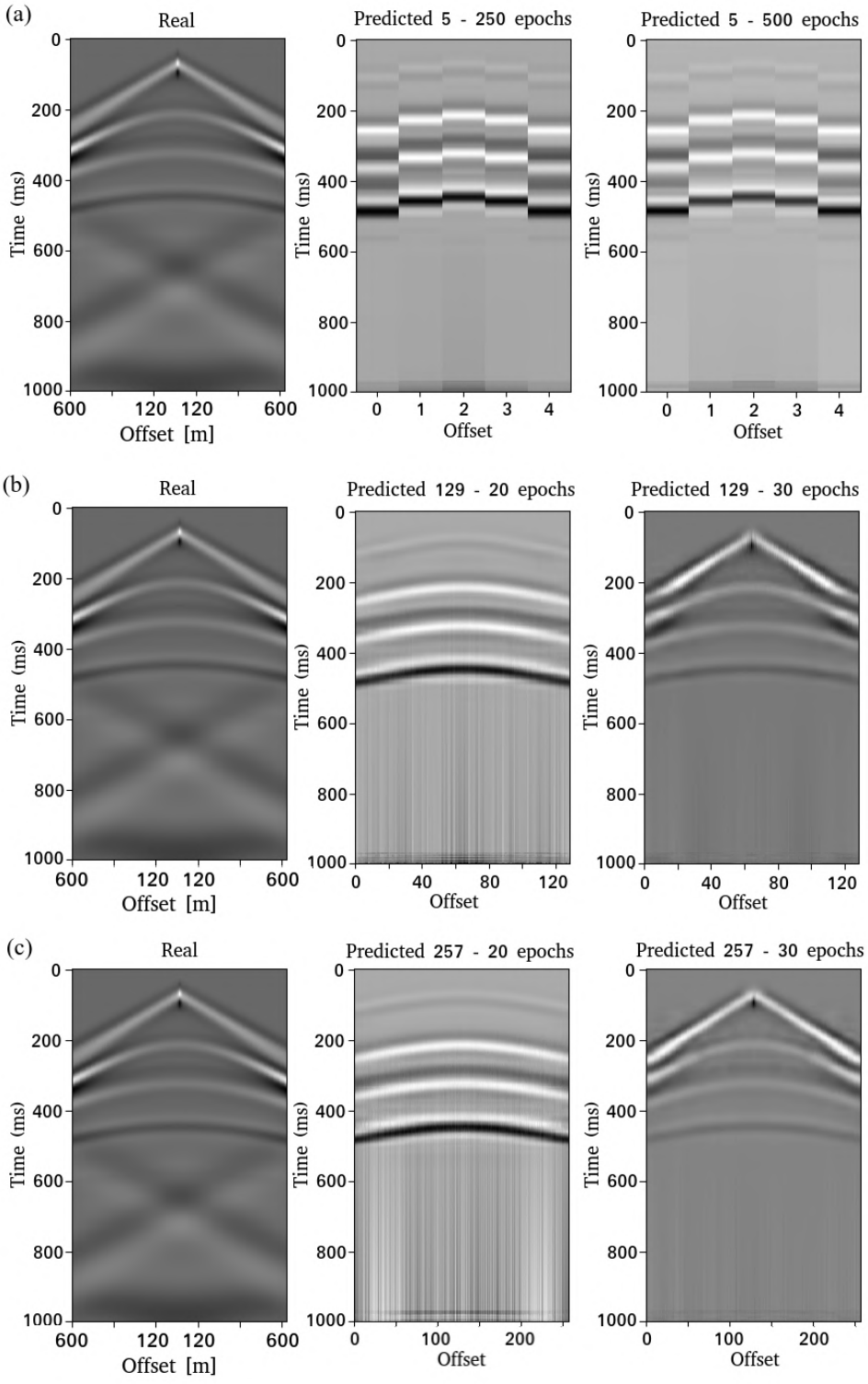

Figure 4.1.15: Comparison between synthetic (FD simulation, column left) and predicted data for 5 (a), 129 (b) and 257 (c) offsets

In order to evaluate prediction time we used each trained model to predict 100 panels and we took the average time for each model. Results are shown in Fig. 4.1.16.

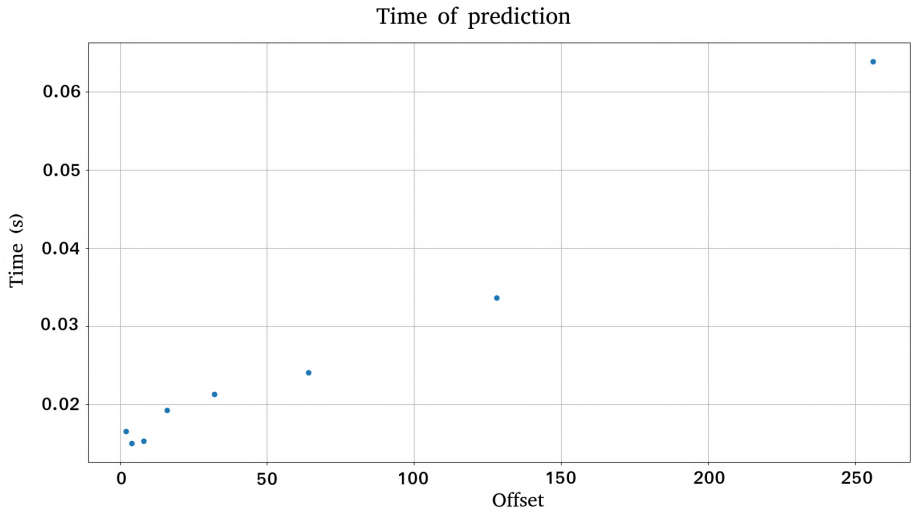

Figure 4.1.16: Prediction time vs offsets. Data are reported in Table 4.2.

Prediction time for a 257 offset CSG panel, under constraints shown in Table 4.1 and with the same hardware, is of 0.08s compared to 2.17s of the FD algorithms used in this paper. According to this benchmark, this would lead to a method that is 27 times faster than the classical one.

Even though prediction time is smaller compared to that obtained by classical methods, it is affected by different factors.

Prediction time for this net depends only on the number of offsets used, on temporal discretization and on record length in time.

The FD generation time, instead, does not depend on number of offsets but on model extension and discretization.

Conclusions

The potentialities of Neural Network (NN) in the solution of non linear problems are well known and the successful application to the present case, i.e. synthetic seismic data generation in the 1-D acoustic case, demonstrates such potential.

The proposed methodology is robust and can solve the forward problem, i.e. numerical modeling of seismic wave propagation starting from the velocity model faster than classical methods, especially when high-frequency source wavelets are considered.

For the 1-D acoustic case NN are a good alternative to the classical Finite Difference (FD) method. They are able to rapidly and accurately solve the problem.

The waveform is well reproduced and the net manages to predict multiple reflections.

While prediction time is much lower than classical methods (e.g. FD), the training

process needs large times and datasets. Such large training effort may imply that the applicability of the method is limited.
In order to tackle such issue and increase flexibility and applicability we propose a strategy based on retraining.
The proposed method gives good results and is able to make adjustments of the NN parameters in a short time (less than one hour on two NVIDIA Tesla P100 GPU accelerators) and with a small dataset.
In this way, even though we are still linked to FD data generation, we are able to modify parameters in the NN by using just a small amount of data generated by FD.
Our results show that the proposed method allows fast and easy parameter change.
The specific architecture and the use of LSTM allowed to obtain seismograms with lots of offsets without substantially increasing the training time.
One possible future work is the application of this methodology for the inverse problem solution and may offer great benefits in the solution of challenging cases, such as e.g. velocity function inversion in presence of multiple reflections.
This technique may open new perspectives of development, in particular the extension to the 2-D and 3-D elastic case. Due to the performance demonstrated in the present work, such extension may be feasible and they can lead to efficient alternative to classical methods.
Nonetheless, the complexity of such extension will require a complete revision of the design of the net.

Acknowledgments

The authors would like to thank Eni, for the opportunity to pursue this work, also through the use of the computational resources made available at ENI's HPC4 [1], and for the approval to bring it to publication.

Bibliography

Sharan Chetlur, Cliff Woolley, Philippe Vandermersch, Jonathan Cohen, John Tran, Bryan Catanzaro, and Evan Shelhamer. 2014. cuDNN: Efficient Primitives for Deep Learning. *CoRR* abs/1410.0759 (2014). arXiv:1410.0759 http://arxiv.org/abs/1410.0759

Francois Chollet et al. 2015. Keras. https://keras.io.

Xavier Glorot, Antoine Bordes, and Yoshua Bengio. 2011. Deep Sparse Rectifier Neural Networks. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (Proceedings of Machine Learning Research, Vol. 15)*, Geoffrey Gordon, David Dunson, and Miroslav Dudík (Eds.). PMLR, 315–323.

R. Guo, J.J. Zhang, D. Liu, Y.B. Zhang, and D.W. Zhang. 2019. Application of Bi-directional Long

[1] HPC4 is a supercomputer owned by ENI spa that has a peak performance of 18.6 Petaflops

Short-Term Memory Recurrent Neural Network for Seismic Impedance Inversion. 2019, 1 (2019), 1–5. https://doi.org/10.3997/2214-4609.201901386

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Comput.* 9, 8 (11 1997), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Taskin Kavzoglu. 2009. Increasing the accuracy of neural network classification using refined training data. *Environmental Modelling & Software* 24 (07 2009), 850–858.

Diederik Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations* (12 2014).

Mathias Louboutin, Michael Lange, Fabio Luporini, Navjot Kukreja, Philipp Witte, Felix Herrmann, Paulius Velesko, and G. Gorman. 2019. Devito (v3.1.0): an embedded domain-specific language for finite differences and geophysical exploration. *Geoscientific Model Development* 12, 3 (2019), 1165–1187. https://doi.org/10.5194/gmd-12-1165-2019

Benjamin Moseley, Andrew Markham, and Tarje Nissen-Meyer. 2018. Fast approximate simulation of seismic waves with deep learning. arXiv:1807.06873 [physics.geo-ph]

Arkadi Nemirovski, Anatoli Juditsky, Guanghui Lan, and And Shapiro. 2009. Robust Stochastic Approximation Approach to Stochastic Programming. *Society for Industrial and Applied Mathematics* 19 (01 2009), 1574–1609.

Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. 2016. WaveNet: A Generative Model for Raw Audio. (2016). http://arxiv.org/abs/1609.03499 cite arxiv:1609.03499.

Alex Sherstinsky. 2018. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network. *ArXiv* abs/1808.03314 (2018).

Patrice Simard, David Steinkraus, and John Platt. 2003. Best practices for convolutional neural networks applied to visual document analysis. In *Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings.* 958–963. https://doi.org/10.1109/ICDAR.2003.1227801

Fangshu Yang and Jianwei Ma. 2019. Deep-learning inversion: A next-generation seismic velocity model building method. *GEOPHYSICS* 84, 4 (2019), R583–R599. https://doi.org/10.1190/geo2018-0249.1 arXiv:https://doi.org/10.1190/geo2018-0249.1

Ola Øye and E.K. Dahl. 2019. Velocity model building from raw shot gathers using machine learning. https://doi.org/10.3997/2214-4609.201900039

Chapter 5

Classification Task

Wavefield geophysics has seen significant advancements with the integration of DL for classification tasks. These cutting-edge methods have revolutionized the way geophysicists analyze seismic and radar data, enabling more accurate, efficient, and automated interpretations of subsurface structures and properties. In this context, we explored the application of LSTM methods for classification tasks focusing on binary classification (Roncoroni et al., 2022a) with an algorithm designed to classify reflections and multi-class classification (Roncoroni et al., 2022b) focusing on the polarity assessment of a reflection data.

Furthermore, we applied a modification of Roncoroni et al. (2022a) to a state of the art GPR data acquired on the Moon by the Chinese mission Chang'E4, providing a completely new and more objective interpretation of the whole dataset (Roncoroni et al., ttedb).

5.1 Efficient extraction of seismic reflection with Deep Learning

This research paper presents a novel approach for interpreting horizons in seismic reflection data using a Deep Learning-based methodology. The proposed flow is fast, accurate, and reduces subjectivity when compared to traditional manual or control-points based methods. The LSTM architecture is used to handle causality and can be applied to input data of different dimensions. The preliminary splitting of the input into patches of predefined dimension, commonly required by Convolutive Neural Networks, is not necessary in the case of LSTM. The results obtained for both synthetic and field datasets show that the proposed approach can be an effective tool for applications that require accurate automated identification/extraction of reflection signals from large datasets, such as for velocity analysis, tomographic, and interpretation purposes. The proposed methodology offers soft constraints for velocity model geometry in all the inversion methodologies that require an accurate input model, like for instance Full Waveform Inversion techniques. This paper contributes to the growing body of research on the application of Deep Learning in geophysical methods and highlights the great potential of LSTM in active wavefield geophysical methods.

Efficient extraction of seismic reflection with Deep Learning

G. Roncoroni, E. Forte, L. Bortolussi, M. Pipan

Computers & Geosciences, Volume 166, September 2022, 105190.
doi: doi.org/10.1016/j.cageo.2022.105190

Key Words

Horizon Extraction; *Deep Learning*; *Neural Network*; *Reflection Seismic*; *GPR*

Highlights

- NN trained on synthetic data -convolutive approach- that works on field data.
- 1-D approach is applicable on both 2D and 3D datasets
- Fully automatic, no need of parameters setting.
- Fast methodology that can deal both with seismic and GPR data.

Introduction

With the increase of 3-D seismic surveys, the quantity of geophysical data used in the interpretation has grown at a phenomenal rate Dorn (1998). Therefore, new picking strategies and algorithms have been implemented in order to make data interpretation faster, less subjective and more accurate at the same time.

There is a plethora of different methods used to interpret seismic horizons and their classification is almost impossible since they are intrinsically different in terms of adopted strategy, type of data application (e.g. 2-D or 3-D), expected results (line drawing, horizon extraction, simple geometrical definition of the structures). Typical approaches include manual picking, interpolation, auto-picking, voxel tracking, or surface slicing Dorn (1998), often starting on the manual picking of a few control points (a.k.a. seeds) then connected by means of different interpolation algorithms across selected portions of the analyzed data set.

Such an approach is not so trivial especially for 3-D seismic data due to several issues and critical steps including, among the others: erroneous and not coherent control points selection; autopicking of data having low reflection coherence and/or limited signal-to-noise ratio (S/N); intrinsic directivity of data and/or of the selected algorithm; complex geology preventing effective autotracking Herron (2014). Multiples,

lateral amplitude variations, lateral phase discontinuities related or not to actual geological features, and overall low S/N ratios are further challenges Hoyes and Cheret (2011). Most of the existing picking strategies exploit reflection amplitudes and are as a consequence very sensitive to S/N but also to lateral amplitude variations, not always related to acoustic impedance changes. The applied processing flow represents another potential issue because it can introduce both distortions in the horizon's continuity and not real coherence, which lead to erroneous results.

Recently, different strategies have been suggested in order to extract the horizons directly from the seismic dataset without manual picking Cubizolle et al. (2015). These new algorithms track surfaces throughout the entire data volume (or within selected portions of it) and are global in nature because they do not need any previous definition of seeds (Stark, 2004; Lomask and Guitton, 2007). In fact, they are able to simultaneously define multiple horizons exploiting the full dimensionality of the data thus offering, at least in favorable conditions faster and less subjective results Hoyes and Cheret (2011). Different approaches have been tested: dip-driven algorithms exploit continuity/variations of dips, global optimization methods focus on the entire available dataset, while horizon patching solutions divide the problem into two different steps i.e. the picking and the horizon grouping or definition. They are based on different assumptions and strategies but are generally time consuming due to their computational cost. A comprehensive review with their pros and cons is provided, for instance, in Hoyes and Cheret (2011).

We use a Neural Network (NN) approach to interpret seismic horizons and to reduce the subjectivity of the control points-based procedures, while assuring accurate and robust horizon extraction. This can be done by a preliminary training on a large synthetic dataset (100,000 traces with 256 timesteps, i.e. samples each). The Long Short Term Memory (LSTM) architecture fits well the problem because it can handle causality and can be applied to input data of different dimensions. The preliminary subdivision of the input into patches of predefined dimension, commonly required by Convolutive Neural Network, is not necessary in the case of LSTM.

Training the NN with synthetic seismograms offers additional benefits, such as the ability to introduce arbitrarily predefined characteristics that are difficult to find in a single seismic dataset. They can also reproduce different levels of resolution and can be generated at a low computational cost. Last but not least, they overcome the implicit limit of an unknown underground model, as in the case of field datasets, thus providing complete control over the position of the reflectors. We present and critically discuss all the phases of the new proposed flow, namely: the data generation, the training, and the application to both synthetic and field data.

Methods

Generation of the data for the NN training We train the NN on synthetic data to avoid any link to the field dataset to be interpreted and to have a complete control over the NN performance through the knowledge of the subsurface model that generated the training data. We start from the classical convolutional model of the seismic traceYilmaz (2001a) to generate the data:

$$f(t) \;=\; w(t) * r(t) \;+\; n(t) \tag{5.1}$$

where *f(t)* is the seismic trace, *w(t)* is the wavelet produced by the seismic source, *r(t)* is the reflectivity function (i.e. the series of reflection coefficients) and *n(t)* is the noise. We modified the terms in equation (5.1) to generate a variety of synthetic data that may effectively represent that a wide range of conditions commonly encountered in field datasets, in terms of signal and signal-to-noise characteristics.

Number of samples and sampling interval are 256 and 0.002s respectively: we have chosen 256 as the dimension of the data vector representing each seismic trace to reduce the training time and the computational load, without limiting the effectiveness of the training phase.

Reflectivity (*r(t)*) takes as random values in the range $\pm[0.04, 1]$ to avoid reflection coefficient labeled with unrealistic signal-to-noise ratio. Each trace can have a maximum of 7 (seven) reflection coefficients in different and random positions in time ranging from sample 10 to 246.

w(t) is a Ricker wavelet (Wang, 2015), defined in the time domain as:

$$r(\tau) = \left(1 - \frac{1}{2}\omega_p^2\tau^2\right) \exp\left(-\frac{1}{4}\omega_p^2\tau^2\right) \tag{5.2}$$

where τ is time (in seconds) and ω_p is the most energetic (or dominant) frequency in radiant per second.

The dominant frequency is chosen between $[30Hz$ and $70Hz]$ with a random uniform distribution, in order to cover the common seismic source frequencies.

To better approximate a real seismic acquisition we need to add noise *n(t)*, as seen in Equation (5.1).

We added the noise component n(t) in (1) to better reproduce realistic field datasets and introduced different types of noise with different random magnitudes, as shown Fig. 5.1.1. Testing of the NN on field data was performed after introduction of the following types of noise:

- Noise 1: defined as a pure random noise that is added to the convolved trace. It

is *n(t)* defined in Equation (5.3)

$$f(t) \;=\; w(t) * r(t) \;+\; n_1(t) \tag{5.3}$$

- Noise 2: it represents an "anisotropic noise" and it is a noise added before the convolution with the reflection coefficient series.

$$f(t) \;=\; w(t) \;*\; (\; n_2(t) \;+\; r(t)\;) \tag{5.4}$$

- Noise 3: use of both Noise 1 and Noise 2 in order to simultaneously reproduce the effects that could affect a seismic trace, namely random noise and anisotropic noise.

$$f(t) \;=\; w(t) \;*\; (\; n_2(t) \;+\; r(t)\;) \;+\; n_1(t) \tag{5.5}$$

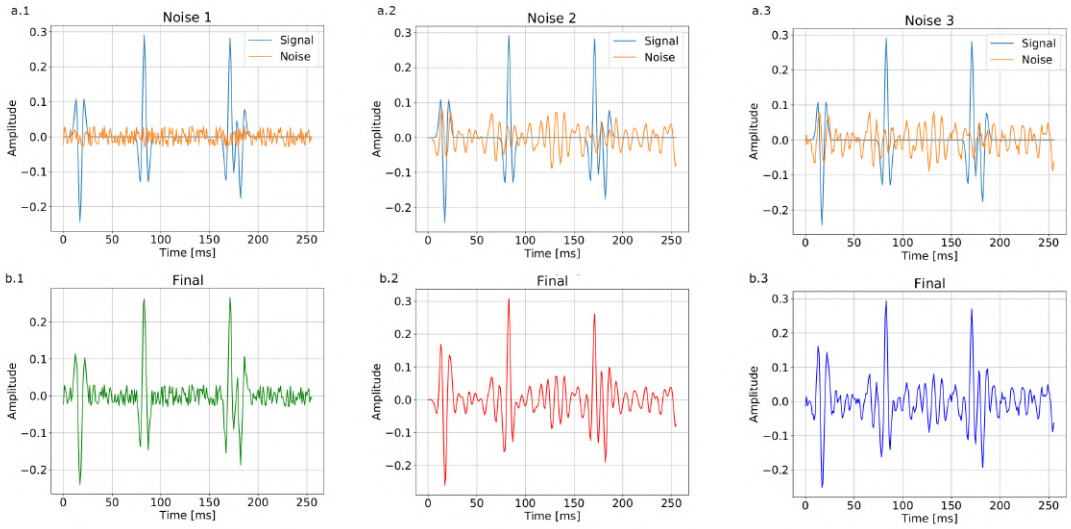

Figure 5.1.1: Three different types of noise (orange) superimposed on a synthetic seismic trace with four reflections (light blue) [a.1, a.2, a.3]; seismic traces resulting from the sum of synthetic trace and noise [b.1, b.2, b.3] (see text for details).

The reference output is defined by a binary indicator (1,0) used to label each sample as reflection / no reflection respectively.

We used the following algorithm to generate the input seismic trace and the reference output for the NN training:

for *i in range(number of desired data:)*

n = generate a casual number of horizons

generate n ***reflectivity values***

generate a random ***wavelet***

generate n random ***depth values***

sort depth values

build the ***reflection coefficient series*** *with depth and reflectivity values*

build ***seismic trace*** *by convolving the reflection coefficient series and the wavelet*

Build the ***categorical reference output*** *with the depth values*

This algorithm is very fast and can be iterated to generate large numbers of seismic traces and reference outputs.

Neural Network Geometry

We chose the network geometry in Fig. 5.1.2 after a grid search to estimate NN hyper-parameters, e.g. number of neurons and depth of the NN.

LSTM maintains the causality and the Long-Term memory better fits the physics behind the wave propagation Hughes et al. (2019).

The Bi-Directional LSTM is able to improve the accuracy of NN classification Guo et al. (2019): in the present study, it can in particular reduce false positives in the final layer by improving the identification of the correct shape of the wavelet.

The output is driven by a Dense layer with a Softmax activation function that outputs a probability value equal to 1 on the phase with the maximum amplitude of a reflection.

We used the CuDNNLSTM Hochreiter and Schmidhuber (1997a) implemented in Keras Chollet et al. (2015), a fast approximation of LSTM on Nvidia CUDA Chetlur et al. (2014): this means that both the training and the model must be used on a CUDA-compatible GPU.

As optimizer we used AdaMax Kingma and Ba (2014), a modified version of Adam with infinity norm, because it performs better on this problem compared to Adam and SGD Kingma and Ba (2014).

As loss function we used Categorical Crossentropy Mannor et al. (2005a), defined as:

$$loss = -\sum_{i=0}^{C} t_i log(s_i) \tag{5.6}$$

Where t_i and s_i are the expected classes and the NN score for each class i in C, respectively summed on the time steps: in this work $C = 2$.

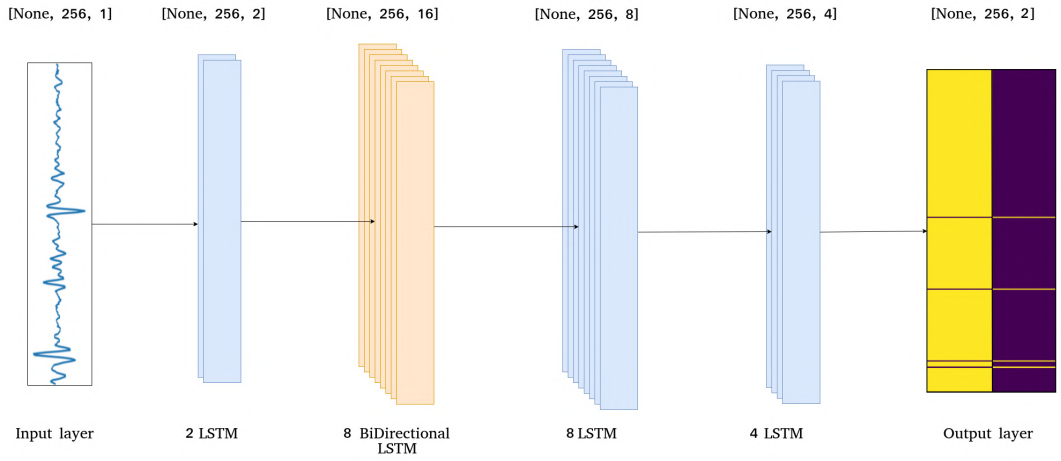

Figure 5.1.2: NN geometry: we set a layer with 2 LSTM (blue, second block from left), 8 BiDirectional LSTM (orange, third block), 2 layers with 8 and 4 neurons each (blue, fourth and fifth blocks) and a final Dense layer with 2 neurons. On the left the input seismic trace and on the right the corresponding reference output.

Training

The training phase was split in two steps Kavzoglu (2009): an initial training on a noiseless dataset and a subsequent training on a noisy dataset. This choice was due to the unbalanced output solutions: a direct training on the noisy trace would led to a huge local minimum where the NN outputs only 0s. To avoid this, we initially train the NN on a noiseless dataset. We got good results from such training, as shown in the additional materials: the training phase lasts 2 hours and it reaches an accuracy of 0.9995 in test and 0.980 in validation.

We trained the NN with Adamax optimizer with a learning rate of 0.01, categorical crossentropy as loss function and batch size of 512.

After the training on the noiseless dataset, we retrained this model 3 times on 3 different datasets with the noise levels and the noise types described in the previous section.

The main objective of this training phase is to evaluate the performances of the trained NN on synthetic and field datasets to infer the optimum solution to generate synthetic data for the application of NN on field seismic data.

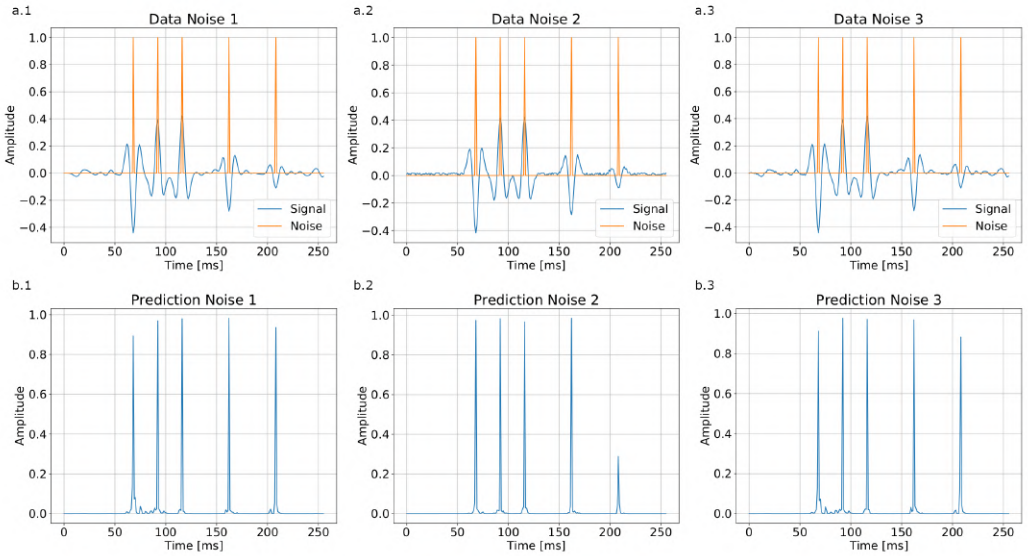

Figure 5.1.3: Noisy traces (blue; see Fig. 5.1.1 for the different types of noise) and expected output (orange); (b.1-2-3): results of the NN prediction.

Prediction

A 1-D approach can be adversely affected by a high level of random noise, although the prediction method is quite robust.

We use the ensemble learning technique to reduce the prediction uncertainty: this technique uses multiple learning algorithms to obtain better predictive inferences (Mendes-Moreira et al., 2009b).

In detail, we tested two solutions, namely prediction with different NNs trained on a dataset with the same characteristics and prediction with the same NN on a single trace and on its inverted version in time.

The two approaches produced similar results and we thus decided to use a single NN to reduce the required training effort.

This methodology generates 2 different predictions, and we combined them with the geometric mean, as it gives better results than the arithmetic mean.

This can be explained by the nature of the prediction: we predict values of probability [0-1], so if NN1 predicts 0.5 (a mid value due to a noisy area) and NN2 0.001 (a very low value), the arithmetic mean would produce a value of $\simeq 0.25$, while the geometric approach would penalize more this prediction ($\simeq 0.02$). An example is shown in Fig. 5.1.4. We can see how two wrong peaks at 25ms and at 780ms, in the first prediction, are muted by the geometric mean.

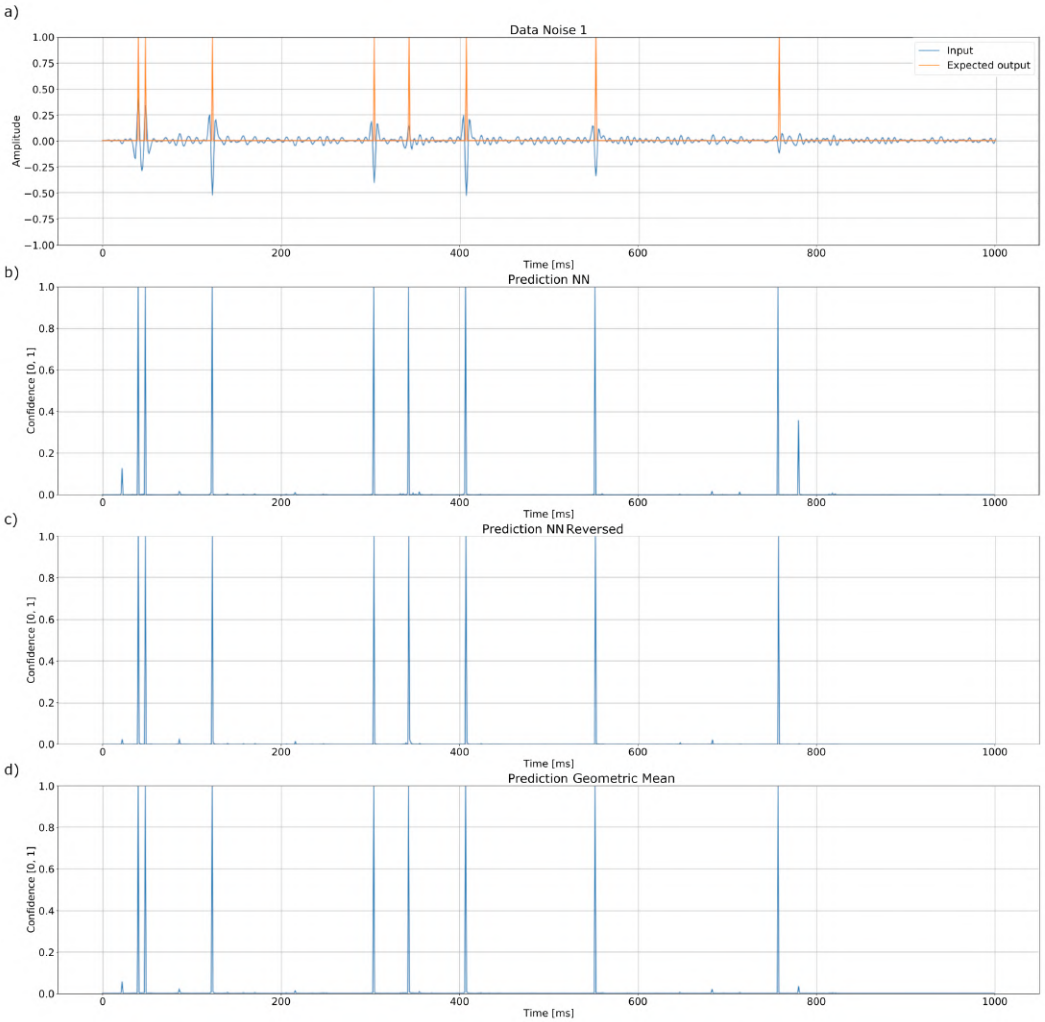

Figure 5.1.4: (a): Input data and expected output, (b): results of the NN prediction, (c): results of the prediction performed on the input trace reversed in time, (d): geometric mean of the two predictions.

We further normalized the maximum value of the first reflection to one to avoid exceedingly large values that can affect the NN performance.

Results

Test on 2-D synthetic data

The first test of the NN performances on 2-D datasets generated with the same workflow as the training dataset produced almost perfect results, as expected. So we moved onto a more complicated example using the Marmousi model Martin (2004).

Fig. 5.1.5 shows part of the data and the prediction, which is given as a probability set that associates a probability value to each point in the section: the value indicates the probability of the point to be a reflector, i.e. to belong to a reflecting surface. Therefore a threshold above which the point is labelled as a reflector has to be set.

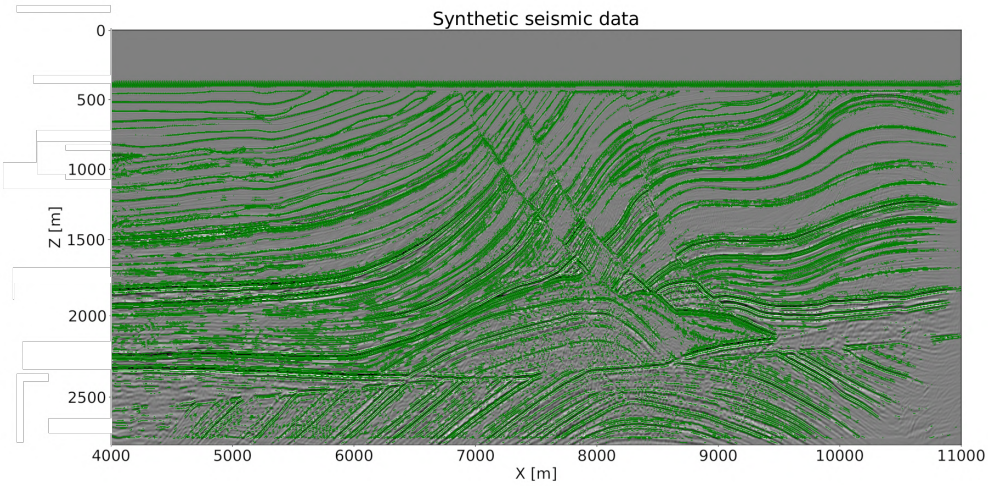

Figure 5.1.5: Results of the horizon extraction (green) performed by the NN on the Marmousi Dataset, (green), superimposed on the input data.

We directly estimate the optimum threshold by evaluating the number of points classified as reflectors vs the threshold (Fig. 5.1.6). We perform this task by using the algorithm described in Satopaa et al. (2011a). The threshold is set at the sharp inflection point clearly visible in the resulting curve, thus limiting the subjectivity of the choice (Fig. 5.1.6).

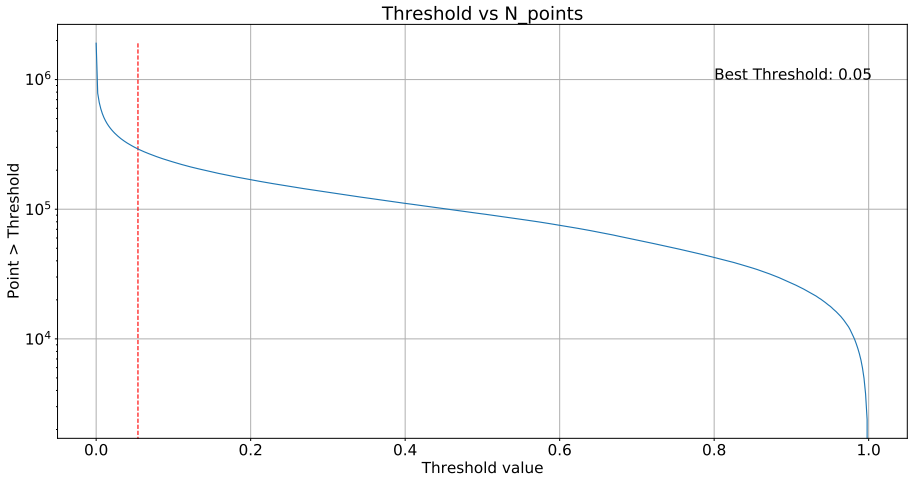

Figure 5.1.6: Number of points classified as reflectors (vertical axis) vs. Threshold value: optimal threshold indicated in red at the knee point.

Test on 2-D field data

In order to test the proposed methodology we use at first a 2-D marine seismic profile of the WS10 exploration project, obtained in autumn 2010 in the west Mediterranean Sea by the Istituto Nazionale di Oceanografia e Geofisica Sperimentale (OGS), which also performed the data processing (Geletti et al., 2014). The selected portion of the seismic profile images a rifted margin of the eastern Sardo-Provençal Basin characterized by a faulted salt dome and by a portion of an almost undisturbed sedimentary sequence (Fig. 5.1.7). For such reason, the analyzed data represent an interesting and complex test for the proposed procedure. We focused on this portion also because it is exactly the same used in Forte et al. (2016) (figure 5a therein) to test an automated picking and phase assessment approach based on phase seismic attributes. The NN is able to properly extract all the main horizons, both where they are sub-horizontal (i.e. in the shallow part) and where they exhibit a significant dip (i.e. along the flanks of the salt dome). As desired, horizons interrupt at the fault location (f labels in Fig. 5.1.8),while correctly no horizons are detected within the salt dome (sd labels in Fig. 5.1.7). The latter result is definitely not easily achievable with traditional picking methods based on 1-D, 2-D or 3-D approaches, because some lateral coherent events can be detected even if they are not actually related to real reflectors (see results in Fig. 5a in Forte et al. (2016)).

NN is also able to extract the high amplitude reflector below the salt dome, while no significant features are detected in the deepest part of the section below the salt dome where reflectors are not continuous also because of some coherent noise due to

over-migration effects.

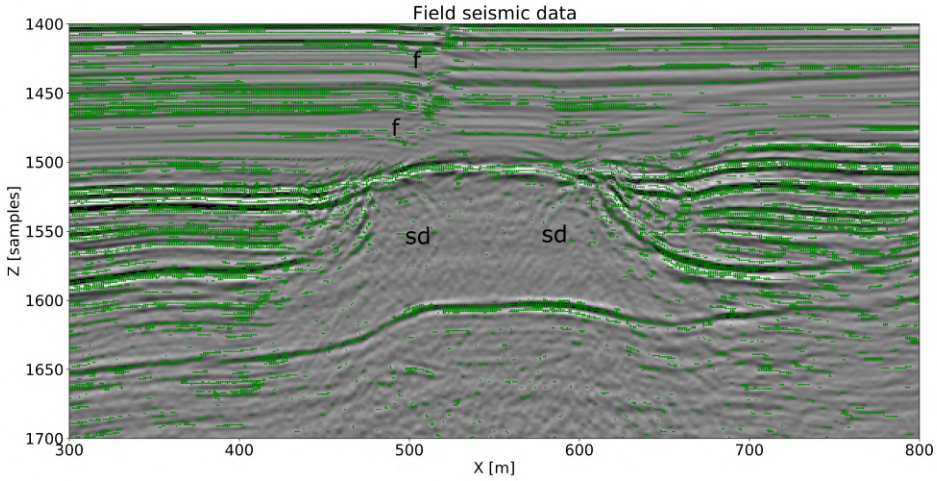

Figure 5.1.7: Example of application of the NN to field seismic data: the data are part of a seismic line from the western Sardinian margin of the WS10 exploration project [see text and Geletti et al. (2014) for details]. The prediction results are shown in green: "sd", salt dome; "f", main fault.

We point out that the above described prediction procedure can be applied to automatically extract the reflectors from any kind of dataset after the training phase is concluded on a set of randomly generated data containing various types and levels of noise.

In order to verify this capability and evaluate the performances, we tested the NN on totally different data, i.e. a Ground Penetrating Radar (GPR) dataset collected in a Glacier in the Eastern Alps (for further details about this data please refer to Colucci et al. (2014)). Although GPR and reflection seismic techniques differ radically concerning sources characteristics and physical parameters, they are both based on the same physical wave theory Ursin (1983). Therefore, processing and analysis techniques used for seismic data may be adapted, at least from the theoretical point of view, also to GPR datasets.

The test data have 802 samples per trace with a 0.454 ns sampling interval, while the spatial sampling interval (0.15 m) is not relevant in a 1-D procedure.

In the example provided in Fig.5.1.8 the NN is able to extract all the relevant reflectors, which are related, from top to bottom to completely different glaciological and geomorphological units, namely: snow and firn (sf), firn (f), debris (d) ice (i) and bedrock (b). The procedure performs quite well on continuous reflectors (like for instance sf-f

contact) but also on highly scattered levels (like d). The extraction of the bedrock is much more difficult mainly because it is not continuous, it is characterized by diffuse scattering and it has a lower signal-to-noise ratio than the shallower units.

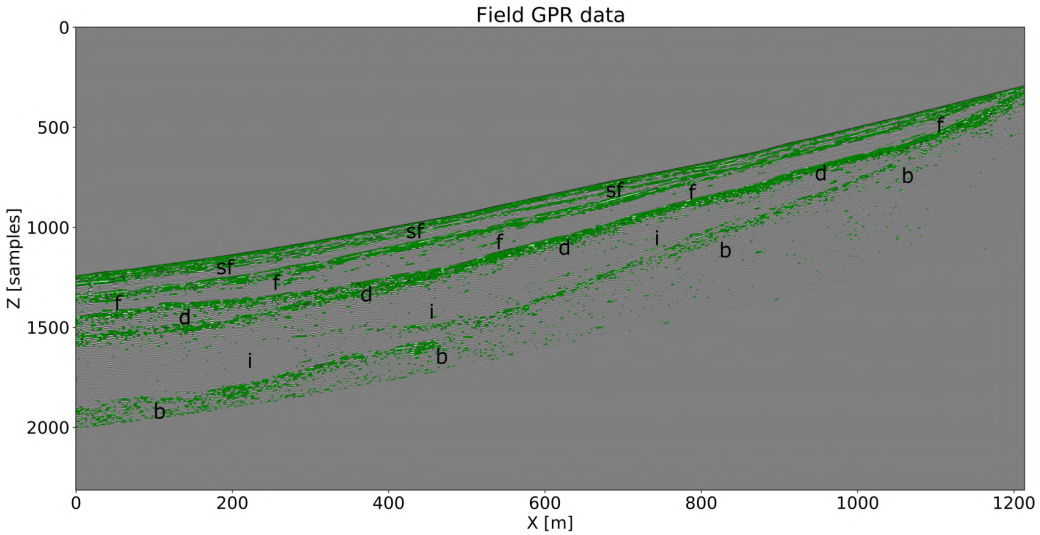

Figure 5.1.8: Example of application of the NN to field GPR data: the GPR profile crosses a glacier in the Eastern Alps [see text and Colucci et al. (2014) for details]. The results of the prediction are shown in green: "sf", snow and firn; "f", firn; "d", debris; "i", ice; "b", bedrock.

Test on 3-D synthetic data

We also tested the approach on 3-D synthetic datasets to deepen the analysis of the proposed method.

The datasets used are taken from Wu et al. (2019b, 2020) and were originally generated to train a Convolutive Neural Network for fault recognition (Wu et al., 2019b), and for paleokarst detection (Wu et al., 2020). We chose these datasets to test our methodology because they are a good approximation to field data.

In order to test the flexibility of our methodology we directly applied our method to the datasets without preliminary processing or additional training to check its flexibility and robustness. The dataset taken from Wu et al. (2019b) has been generated with a convolutive approach to simulate a faulted environment Wu et al. (2019b).

We tested the NN on a 128x128x128 samples seismic volume obtained by random selection from this dataset, The results are quite good, despite the huge number of close reflections; as shown in Figure 5.1.9 we can see the continuity of the horizons in the 3 dimensions and also easily identify the main faults.

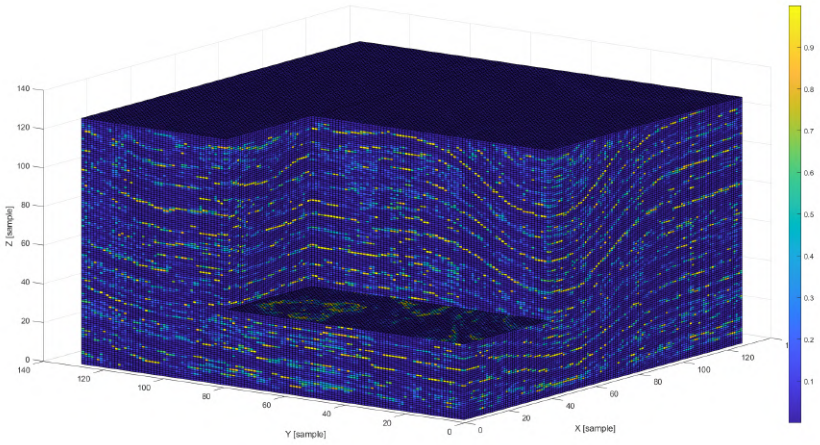

Figure 5.1.9: Example of results of NN prediction on a 3-D synthetic dataset generated from a subsurface model corresponding to a faulted environment (Wu et al., 2019b): the color scale is associated to the probability level (1 = reflection, see text for details).

The second dataset was also generated with a convolutive approach to simulate a paleokarst environment (Wu et al., 2020). From this dataset, we extracted a 256 x 256 x 256 seismic volume. The prediction is fast and quite accurate (see Figure 5.1.10) and we can identify the collapsing structures as high noise area, the 3-D continuity of the horizons and their interruptions in the karstified areas.

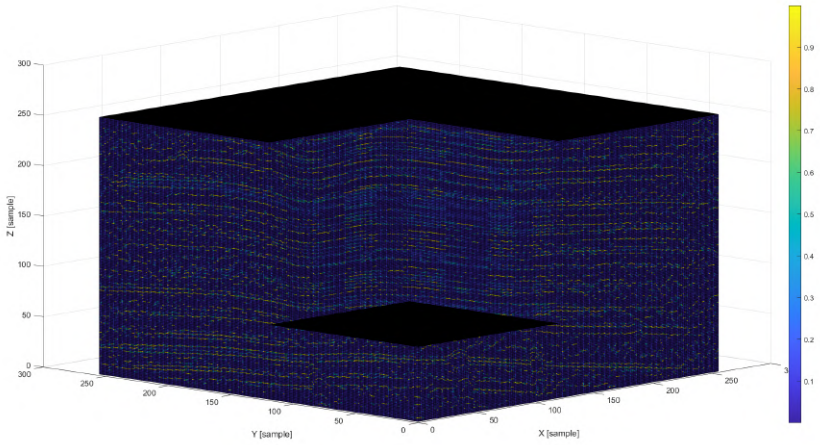

Figure 5.1.10: Example of results of NN prediction on a 3-D synthetic dataset generated from a subsurface model corresponding to a paleokarst environment (Wu et al., 2020): the color scale is associated to the probability level (1 = reflection, see text for details).

Discussion

The proposed methodology substantially reduces the computational costs of the generation process compared to other approaches. The use of Finite Differences methods for the forward modeling typically requires huge amounts of memory and computational time for the generation of datasets like the ones here considered. We therefore choose the convolutive approach, which has the added advantage of better handling different types of superimposed noise (see subsection 2.1).

Another key factor is using fully synthetic datasets for the training phase: avoiding field data at this stage prevents possible bias due to manual selection of horizons and implicit limitations in data characteristics due to the fact that field data can only include a limited range of possible geological situations. Furthermore, the more classic approach, in which a portion of the available data is used to train the NN and subsequently applied to the entire data set, has the obvious inherent constraint that the previous data is excluded from the extraction procedure. An example of such an approach can be found in Tschannen et al. (2020), where the authors trained a Convolutive Neural Network on a small portion of the dataset after a manual interpretation, then applying the trained model on the remaining data.

For validation purposes, we tested the proposed procedure on the synthetic Marmousi dataset from Martin (2004). For this dataset all the model details are known, including reflector's geometry, velocity model and seismic wavelet. We can evaluate our results considering both the seismic section (Fig. 5.1.11a) and the real velocity model (Fig. 5.1.11b). The shallowest part of the profile, from 0 to about 1500 meters, is well predicted and matches very well both the seismic section and the velocity model; also the faults are correctly identified.

The analysis of the results of prediction in the deep part of the section gives useful indications about the performance of the network as a function also of the noise level. The complexity of Marmousi's velocity model gives rise to several non-primary events, such as multiple reflections, which appear more and more at increasing depths. This is particularly evident within the wedge located between 2250-2500 m and 4000-7000 m in vertical and horizontal coordinates, respectively.

The comparison between the seismic velocity model and the seismic section reveals that several coherent events (similar to reflectors) occur in the section although the wedge is actually homogeneous and characterized by a single velocity value. Despite the increasing number of primary and non-primary interfering events, the NN makes accurate predictions and correctly retrieves the main elements of the structural model.

A remarkable result of the proposed approach is that the NN is able to deal also with the wavelet stretching due to the well known low-pass filtering effect of the geological materials (Yilmaz, 2001a).

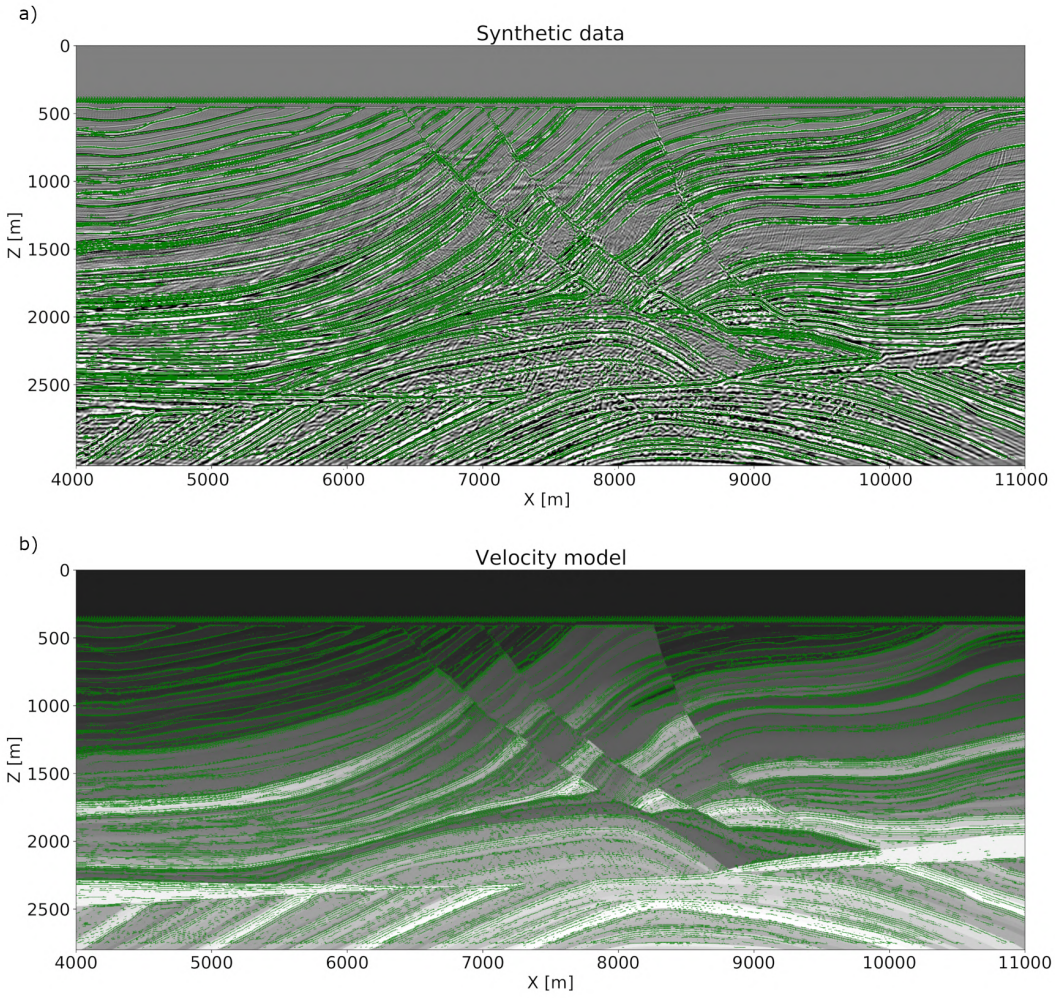

Figure 5.1.11: Marmousi dataset: example of prediction a) on the seismic profile and b) on the true velocity model.

The data about the computational load and performance of the NN refers to a machine with a 2 core Intel(R) Xeon(R) CPU, 2.20GHz, 12Gb Ram and a Nvidia Tesla T4 GPU. We performed 100 tests to test the generation time of the algorithm: each test produced a dataset of 10,000 traces 256 samples long. The total time for each test was $2.80s \pm 0.13s$.

Training time is not an issue, as the NN does not need to be retrained for each specific task, as above highlighted, but can be applied directly to most field datasets. The first training, without noise, lasts 100 epochs on a training dataset of 100,000 x 256 samples and this takes less than 1 hour. The training with noise is longer, due to the increasing complexity of the task: it lasts 4000 epochs for a total training time of about 26 hours.

A great strength of this method is the significantly shorter prediction time than traditional approaches. We tested it on 1000 datasets with variable time dimensions from 4 to 4096 samples. LSTM is able to predict any temporal length and we analyzed also the change in prediction time depending on the record length, resulting, as expected, in a linear relationship between temporal length and time of prediction.

Conclusions

We propose a Deep Learning-based methodology that is fast and accurate in extracting different types of reflectors from reflection seismic dataset. The results obtained from its application to synthetic and field datasets show that it can be an effective tool for applications that require accurate automated recovery of reflection signals from large datasets, such as e.g. velocity analysis, tomography, interpretation. One of the main advantages introduced is the prediction with a correlated probability: this is crucial in the recovery of horizons (i.e. reflecting surfaces) from both 2-D and 3-D data because this parameter can weigh the estimate of the optimum surface, decreasing the level of subjectivity of the whole procedure. Other possible applications of this methodology could be the phase assessment of reflections, which in turn represents a very important task for velocity estimation. In particular, in the field of velocity modeling it may offer soft constrains for velocity model geometry in all the inversion methodologies that require an accurate input model, like for instance Full Waveform inversion. The proposed algorithm is fast, versatile, and does not require any input from the user, as the threshold is estimated automatically using a Knee function (see section 3.1 for details) while the training phase is done on synthetic data thus avoiding possible bias due to manual horizons picking and intrinsic limitations and peculiarities of a specific field dataset. Further researches will focus on horizons patching and automated phase recognition.

Computer Code Availability

Codes used in this article are available on github: https://github.com/Giacomo-Roncoroni/Efficient_horizons_extraction

Acknowledgements

This research was supported by PNRA projects IPECA (PNRA18_00186) and CRIOVEG (PNRA18_00288) and by 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.

5.2 Polarity assessment of reflection seismic data: a Deep Learning approach

This paper presents a novel methodology for the polarity assessment of reflection seismic data using a fully 1-D approach based on a LSTM architecture. The proposed methodology is able to accurately extract seismic horizons and track polarity information (i.e. signal phase behaviour), even in complex conditions, without requiring any input besides the seismic data. The algorithm performances are quite good since it is able to spot reflection events with the right polarity and it just misses a few points on low-amplitude horizons, in which the spatial continuity is not complete. The effectiveness of the proposed procedure is demonstrated through synthetic and field datasets testing, including a 2-D marine seismic profile of the WS10 exploration project and marine seismic data acquired on the Ionian continental platform. The results show that the Bi-directional LSTM architecture is an effective solution for automatic polarity assessment and can be applied to any type of seismic dataset, including large 3-D volumes thanks to its moderate computational load. This approach has a great potential to significantly improve the accuracy and efficiency of polarity assessment in reflection seismic data.

Polarity assessment of reflection seismic data: a Deep Learning approach

G. Roncoroni, E. Forte, L. Bortolussi, L. Gasperini, M. Pipan

BGO, Volume 63, November 2022, 0098-3004.
doi: doi.org/110.4430/bgo00409

Key Words
polarity assessment; *seismic phase*; *Deep Learning*

Introduction

The time delay of a reflected event between adjacent traces is a crucial parameter in the interpretation of seismic reflection data. Several seismic attributes help to analyze the lateral continuity of the reflected events, such as e.g. coherency and phase-based attributes. Polarity is a further crucial element in such analysis and attributes such as apparent polarity can help tracking reflections and variations related to fluid substitutions. Polarity is related to the reflection coefficient values which are in turn related to the contrasts of acoustic impedance and so to the petrophysical parameters of the subsurface materials. Therefore, polarity changes with depth and polarity reversals along a reflector are essential indicators in the quantitative analysis of seismic data.

Besides the well-known problems related to phase distortions due to both seismic data acquisition, analysis and processing (Brown, 2009) the polarity assessment (a.k.a. phase detection) is often a not trivial step during reflection seismic data interpretation. In fact, phase assessment is a common problem affecting autopicking methods (Forte et al., 2016). In the ideal case, the final interpretation objective is to pick (and extract) the first phase of each reflection and the related peak amplitude. In this way, the subsurface geometries and the seismic impedance contrasts can be extracted and inversion procedures can be properly performed.
Picking of reflected events can be somehow related to techniques applied in seismology, in particular first-break picking (Sabbione and Velis, 2010). The phase assessment is a generalization of such an issue, with the main difference related to the extremely higher number of reflections and records (i.e. traces) typical of reflection seismic data.

While in recent years several phase picking techniques have been developed and

implemented exploiting machine learning techniques for both seismological (e.g. Cano et al. (2021)) and reflection seismic (e.g. Tschannen et al. (2020)) applications, less effort has been specifically addressed to the polarity assessment.

In this paper, we describe and test a procedure for the polarity assessment of reflection seismic data based on a fully 1-D Neural Network (NN) approach, requiring in input only the seismic data, thus minimizing the subjectivity level and the intervention of the interpreter. The procedure is at first tested on 1-D and 2-D synthetic data, with various noise levels. We successively apply it on real cases previously used for testing other published methods, in order to allow an objective assessment of the effectiveness of the proposed procedure. In particular, we thoroughly analyze the benefits of using a 1-D strategy which can be applied to any type of seismic dataset including large 3-D volumes thanks to its moderate computational load.

Methods

The algorithm is based on a Long Short-Term Memory (LSTM) architecture because we want to keep the causality of the data and the Long-Term memory better fits the physics behind the wave propagation (Hughes et al., 2019). In fact, the Bi-Directional LSTM is a strategy able to improve the accuracy of NN classification (Guo et al., 2019) and in the present case it can help the NN to find the correct shape of the wavelet by working on both sides of it (Figure 5.2.1).

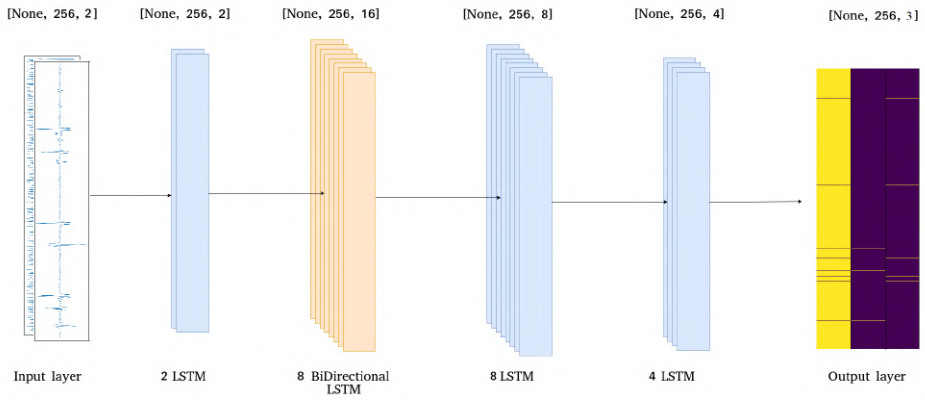

Figure 5.2.1: Example of Neural Network architecture of the proposed algorithm. From left to right: input, hidden layers and output with three possible categories: no reflection, positive reflection, negative reflection.

The output is driven by a dense layer with 3 neurons and a SoftMax activation function that outputs a probability value equal to 1 on the maximum phase of a reflection in

time: the three classes represent negative polarity, no reflection and positive polarity. We use the CuDNNLSTM, a fast approximation of LSTM (Hochreiter and Schmidhuber, 1997a) that works on Nvidia CUDA (Chetlur et al., 2014) and AdaMax as optimiser (Kingma and Ba, 2014), a modified version of Adam with infinity norm and categorical cross-entropy as loss function (Mannor et al., 2005a). We adopt a double input, namely synthetic data and its cosine of the instantaneous phase, that should give more information to the algorithm and exploits the possibility given by the NNs to deal with a layered information.

The training is fully performed on synthetic data obtained from a convolutional model-based scheme, while the subsequent horizon extraction step can be applied to any type of field seismic dataset.

The training dataset is often a crucial issue for the performances of the algorithm on field datasets: we train the NN on synthetic data to avoid any link to a specific field dataset and to have a complete control over the NN performance through the knowledge of the subsurface model that generates the training data. After tests, we found out that the best way to simulate field data for NN is to add pure random noise to the convolved trace and also to the reflection coefficient series before the convolution:

$$f(t) = (w(t) * (n_1(t) + r(t))) + n_2(t) \tag{5.7}$$

where $r(t)$ is a randomly uniform distributed coefficient series with random values, while $n_1(t)$ and $n_2(t)$ are noise series randomly generated for each data. $n_1(t)$ has been added to simulate realistic cases. While $n_2(t)$ could represent an instrumental or environmental noise, $n_1(t)$ is a noise linked to the nature of the material and can be seen in the seismic trace as random noise convoluted to the seismic wavelet. The other term is the wavelet: we use different frequency ranges to generate the wavelet in order to simulate the stretching, spectral variations, and variability that occur under natural conditions. We do not need to train on a previously defined wavelet since by feeding a NN with a recorded signal, the link between its temporal discretisation and the actual recording time is unknown for the NN, unless specified, as the NN takes as input only amplitude information. This allows us to define the temporal discretisation of the desired wavelet and just resample our data to the desired discretisation value, equal to 36 in this training. In order to reduce the prediction uncertainty, we use the ensemble learning technique that exploits multiple learning algorithms to obtain better predictive inferences. In particular, we tested two solutions, namely: prediction with different NNs trained on a data set with the same characteristics, and prediction with the same NN on a single trace and on its time-inverted version. The two approaches produced similar results; we thus decided to use a single NN to reduce the required training effort.

With the application of ensemble learning, we obtain two different predictions: one on the single trace and the other on its inverted version in time. We successively combine them with the geometric mean, as it gives better results than, for instance, the arithmetic mean.
The prediction is given as a probability set that associates a probability value to each point: the value indicates the probability of the point to be a reflector, i.e. to belong to a reflecting surface. A minimum threshold above which a point is labelled as a reflector can be set. The optimum threshold is estimated by evaluating the number of points classified as reflectors vs. the threshold. We perform this task by using the algorithm described in Satopaa et al. (2011b). The threshold is set at the sharp inflection point clearly visible in the resulting curve, thus limiting the subjectivity of the choice. We apply this methodology both for positive and negative values (Figure 5.2.2).

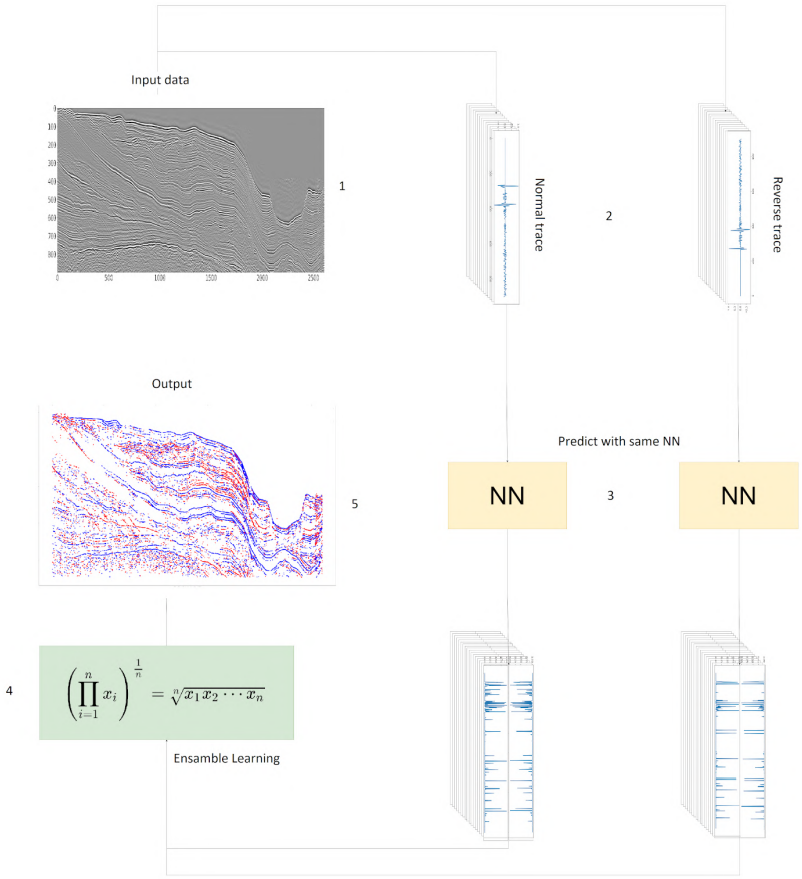

Figure 5.2.2: Example of the proposed workflow: 1) input data; 2) normal and reversed-in-time version of the input, reshaped as trace-by-trace matrix; 3) NN prediction; 4) ensemble learning, i.e. a geometrical mean of the input; 5) final results.

Results and Discussion

In this section we present the application of the method to both synthetic and real data.

In Figure 5.2.3 we can see a synthetic 1-D test. It shows the performances of the NN on a dataset which is similar to training data. As we can see the NN performs very good and the output expressed as a probability in range [0, 1] for classes positive and negative shows very high values in almost all the cases. At time step 590, the approach is able to predict a very close overlapping between positive and negative reflections and it correctly predicts a very low probability value for a wrong negative polarity signal at 596 time step.

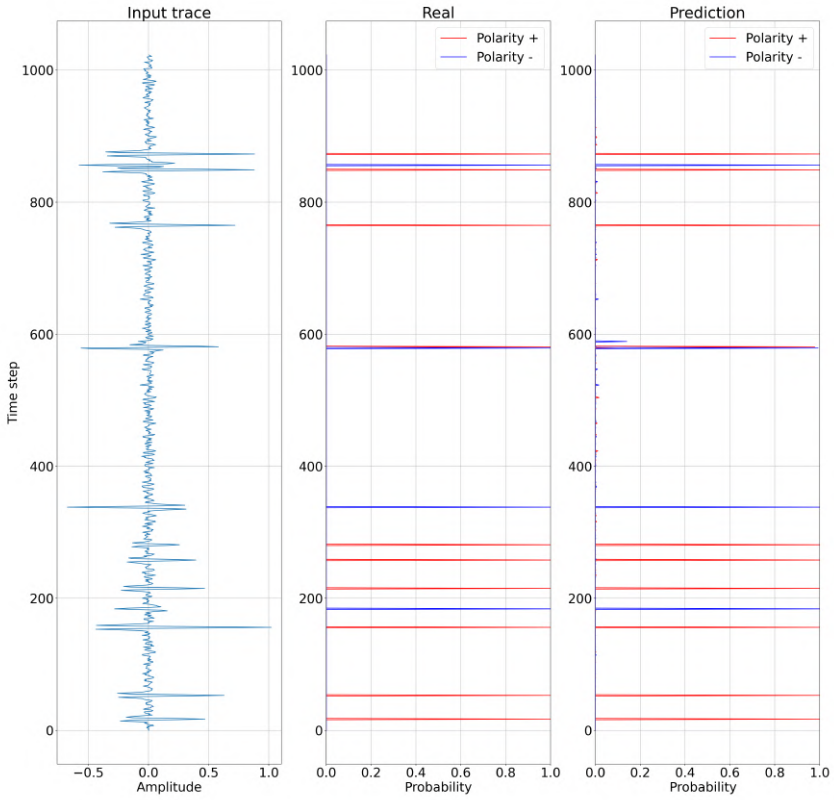

Figure 5.2.3: Synthetic example of prediction: from left to right we have the input trace, the real output and the NN prediction.

In order to test the 2-D potentiality of our trace by trace approach we generated some

random 2-D models. In Figure 5.2.4 we can see the target data (left) and the NN prediction (right). As we can see, here we applied a threshold to data to decide if one point should be considered a reflection or not: in this case the minimum threshold was set to 0.9 probability.
The algorithm performances are quite good since it is able to spot reflection events with the right polarity and it just misses some points on low amplitude horizons, in which continuity is not complete, e.g. along the event marked with A in Figure 5.2.4. In the prediction we can spot also some wrong randomly distributed predictions, which could easily be deleted in a post processing step, since they do not show any lateral coherency.

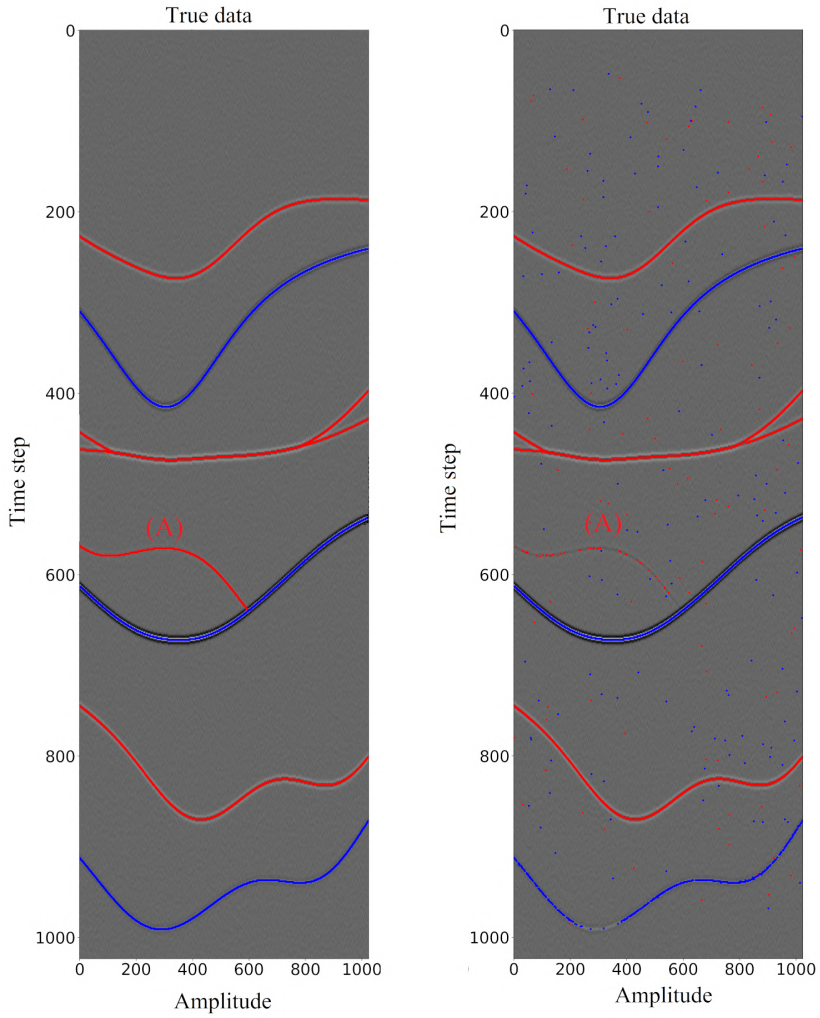

Figure 5.2.4: Prediction on a 2-D synthetic seismic profile generated with a convolutive approach. Real polarity, on the left and related prediction, on the right. With A we marked horizon with a low amplitude that was not continuously reconstructed by the methodology.

In order to test the proposed methodology we use a 2-D marine seismic profile of the WS10 exploration project, obtained in autumn 2010 in the west Mediterranean Sea by the Istituto Nazionale di Oceanografia e Geofisica Sperimentale (OGS), which also performed the data processing (Geletti et al., 2014). The selected portion of the seismic profile crosses a rifted margin of the eastern Sardo-Provençal Basin characterized by a faulted salt dome and by a portion of an almost undisturbed sedimentary sequence (Figure 5.2.5). For such reason, the analyzed data represent an interesting and complex test for the proposed procedure. We focused on this portion also because it is exactly the same used in Forte et al. (2016) (Figure 5.2.5a therein) to test an automated picking and phase assessment approach based on phase seismic attributes. The NN is able to properly extract all the main horizons, both where they are sub-horizontal (i.e. in the shallow part) and where they exhibit a significant dip (i.e. along the flanks of the salt dome), and is able to keep track of the inversion from positive to negative polarity under the salt dome (marked with P and L in Figure 5.2.5, respectively).

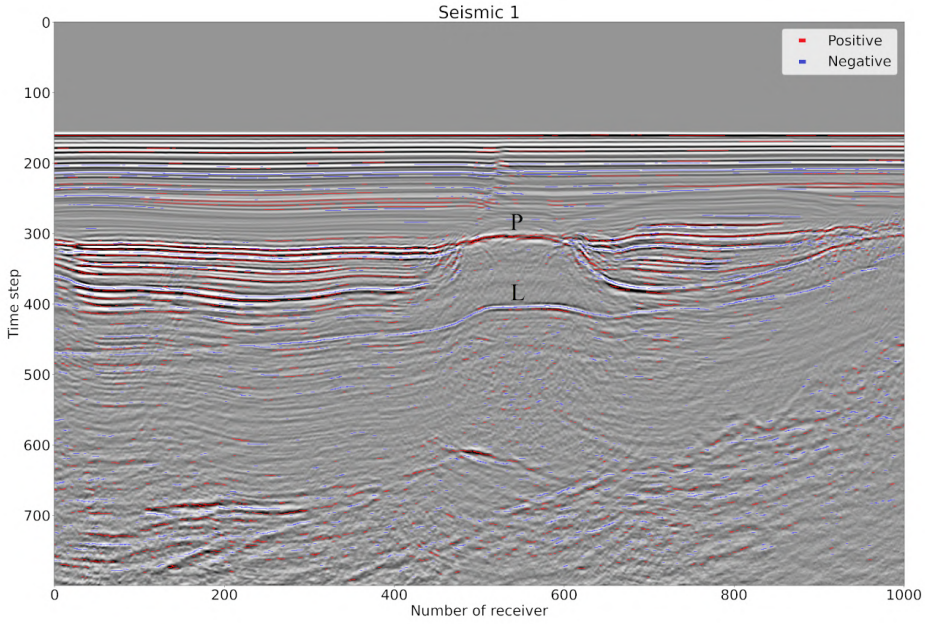

Figure 5.2.5: Prediction on a 2-D marine seismic profile of the WS10 exploration project in the west Mediterranean Sea acquired by the Istituto Nazionale di Oceanografia e Geofisica Sperimentale (OGS), a salt dome with top at time step 300 is located between positions 450 and 650.P and L mark the positive and negative polarity, respectively at the top and at the bottom of the salt body.

The second test considers marine seismic data acquired on the Ionian continental platform (seismic source boomer).
Over a total length of 25 km, the profile samples an extremely irregular sea bottom topography along with complex sub-bottom structures that include steep and conflicting dips, faulted horizons, thin layers and sedimentary wedges. The algorithm correctly follows horizons and polarity reversals even in the steeply sloping parts (see A and B in Figure 5.2.6).

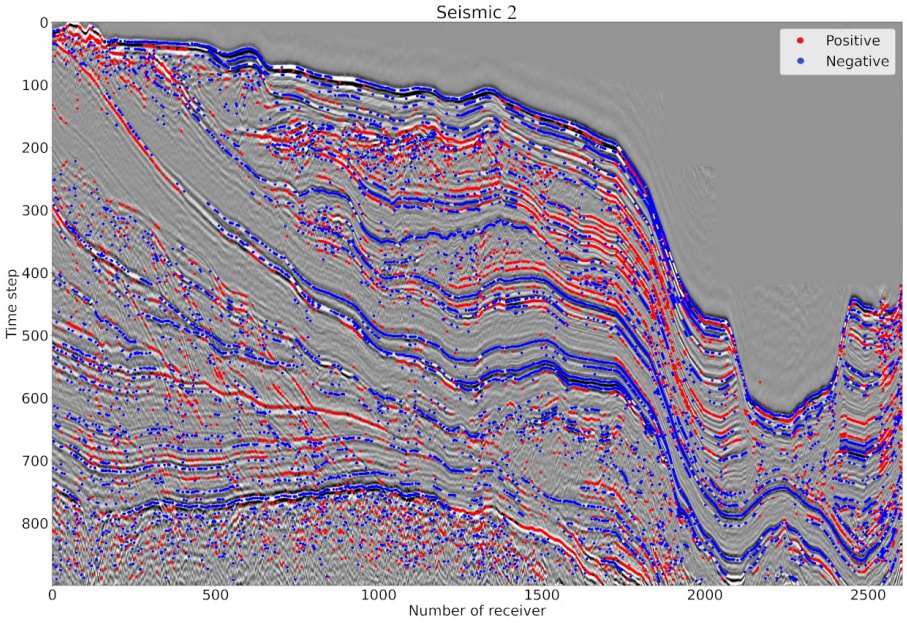

Figure 5.2.6: Example of prediction performed by the proposed NN on marine seismic data from the Ionian continental platform.

Conclusions

Bi-directional LSTM architecture exhibits good prediction performance and is an effective solution for polarity assessment as demonstrated by synthetic and field dataset testing. The double input strategy (amplitude data and cosine of instantaneous phase) apparently provides a further improvement in performance probably related to the NN characteristic of exploiting the layering of information and benefiting from the relative insensitivity to variations in the amplitude of the instantaneous attribute. Training on

synthetic data allows great flexibility as well as to prepare NNs for real data applications of any complexity. The training phase for the NN presented in this work required a total time of 6 hours on a machine with GPUs[1]. which can be considered an acceptable computational cost when compared with the subsequent performance of the NN in the analysis of the field data. The training effort is further reduced by the solution of a single NN working on direct and reverse time data. The threshold for labeling the reflectors is set through a user-independent procedure which greatly reduces the subjectivity of the whole process. The application of the algorithm to the seismic data of the eastern margin of the Sardinian-Provençal Basin and of the Ionian continental platform shows that the performance of the NN is not influenced by the complexity of the structural conditions and by the topography of the seabed.
Further analysis is needed to test the proposed strategy on different types of data (land seismic, GPR) but the tests performed to date show that the 1-D approach and the complete adaptability to different wavelets and time scales make the algorithm robust and able to cope with a virtually unlimited range of applications.

Acknowledgements
This research was partially supported by PNRA projects IPECA (PNRA18_00186) and CRIOVEG (PNRA18_00288) and by 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 the anonymous reviewers for their fruitful comments and suggestions.

[1]On Cineca Marconi 100: 2x16 cores IBM POWER9 AC922 at 3.1 GHz, 4 x NVIDIA Volta V100 GPUs, 256 GB RAM

5.3 Deep Learning driven interpretation of Chang'E4 Lunar Penetrating Radar

This paper presents a groundbreaking achievement in lunar exploration through the use of Deep Learning driven interpretation of Chang'E4 Lunar Penetrating Radar. The team reprocessed and interpreted data collected by the high frequency channel of the lunar radar, using a new algorithm that allows to automatically obtain objective and detailed results. The algorithm was based on Long-Short-Term Memory (LSTM), which is particularly well-suited for processing sequential data. The team was able to image dipping layers and shallow buried crateriform structures within the lunar regolith, as well as recognize deeper structures similar to craters. Several previously not recognized structures were imaged and characterized. This result demonstrates the potential of LSTM-based applications in active wavefield geophysical methods, and opens up new roads for exploration and discovery in the field.

Supporting information for this paper can be found in https://figshare.com/s/8830a91ac0c51a4c9bf7.

Deep Learning driven interpretation of Chang'E4 Lunar Penetrating Radar

G. Roncoroni, E. Forte, I. Santin, A. Černok, A. Rajšić, A. Frigeri, W. Zhao, M. Pipan

Submitted to Journal of Geophysical Research: Planets.

Under Review

Preprint in ESS Open Archive.

doi: doi.org/10.22541/essoar.169504584.44883108/v1

Key Words

Deep Learning; *Lunar Penetrating Radar*; *Data integration*

Highlights

- Processing and analysis of a more than 1400 m long Chang'E4 Lunar Penetrating Radar profile collected on the farside of the Moon.
- For the first time a Deep Learning based algorithm is exploited on Lunar radar data to extract the subsurface horizon probability.
- Improved subsurface geometry was obtained and new elements were detected, including craterform structures and related deposits.

Introduction

The aim of the Chinese lunar landing mission Chang'E-4 (CE-4) is to unravel the causes of irregular volcanic products and regolith between the near and far side of the Moon. The Yutu-2 rover landed on $3^{rd} January 2019$, on the lunar far side, in the ancient Van Kármán crater (diameter D = 185 km; 177.5991degE, 45.4446degS), located within the South Pole-Aitken Basin (SPA), the largest and likely the oldest impact structure on the Moon (Byrne, 2008), Fig. Sup. S1, S2. The two major scientific targets of the Yutu-2 rover are: to study the mineralogy of the SPA by collecting in situ reflectance spectra; to image the subsurface shallow geology using a subsurface penetrating radar system. For these reasons, in addition to reflectance spectra and several other sensors, the Yutu-2 rover is equipped with a dual frequency Lunar Penetrating Radar (LPR) with central frequencies centered at 60 and 500 MHz (respectively CH-1 and CH-2). LPR is the first radar moving directly on the surface

of the Moon's farside (Dong et al., 2021). As in the Chang'E-3 (CE-3) mission, the fundamental goal of the LPR surveys in CE-4 was the exploration of the lunar subsurface structures along the rover's path down to several tens, or even hundreds of meters (Fang et al., 2014; Jia et al., 2018; Wu et al., 2019a). This is the first instrument traveling on the Moon's surface capable of radar sounding at such depths, with horizontal and vertical spatial resolutions up to about 0.1 meters.
Since it landed, the rover has been moving along an irregular path (Figure 5.3.1, Fig. Sup. S4), broken into sectors separated by many stops and turnarounds points. The initial studies focused on the first hundreds of meters of the path by applying further analysis, processing, and inversion algorithms (Giannakis et al., 2021; Wang et al., 2021; Zhou et al., 2021) before data interpretation (Dong et al., 2021, 2020; Lai et al., 2020; Li et al., 2020). These early studies revealed a horizontally layered subsurface with an almost constant regolith thickness of $\approx 10 - 12$ meters and several ejecta layers just below it, as well as deeper basalt layers (Lai et al., 2021; Li et al., 2020).
As the mission progressed, new LPR datasets were released, and new evidences for buried structures emerged from the data, such as a paleocrater from a meteorite impact (Zhang et al., 2021), dipping features (Feng et al., 2022)), a "sandwich structure" within a paleocrater (Zhou et al., 2022), and some faults (Chen et al., 2022).

Up to now, most of the studies used visual interpretation to detect horizons by only considering the reflection amplitude, while just in two cases, single and straightforward signal attributes like the instantaneous amplitude (Zhou et al., 2022) and the signal central frequency (Giannakis et al., 2024) were exploited. In this way, an unavoidable subjectivity is introduced into the interpretation process, and other analyses are needed to support the interpretation, in particular numerical simulation and velocity analysis (Giannakis et al., 2021; Wang et al., 2021; Zhou et al., 2022; Chen et al., 2022). Diffraction hyperbolas analysis can be effective in estimating the EM velocity field, from which properties such as dielectric permittivity and mean density can be derived. However, there are intrinsic problems in addition to the limited number of diffractions (most of them concentrated in the shallower part of the profile, Figure 5.3.2), their interference and their often-irregular shape. A noteworthy issue is related to the not-rectilinear travel path of the rover on the Moon's surface which features abrupt changes of direction along a highly irregular route, as well as varying speeds (Figs. Sup. S8, S10-S13.). Other studies estimated the dielectric constant from the reflection amplitude (e.g. Dong et al. (2020); Giannakis et al. (2024)). This inversion approach is undoubtedly effective in some cases indeed there are both theoretical and practical limitations (Forte et al., 2014).

The low-frequency data of the LPR system (i.e. CH-1) are affected by interference phenomena first described for the CE-3 mission (Li et al., 2018) and then reported also for the CE-4 mission (Pettinelli et al., 2021). The debate is still open (Zhang et al., 2021) and some recent studies continued to exploit the low frequency dataset (Cao et al., 2023). Our work focuses on the high-frequency LPR dataset (i.e. CH-2) due to its high quality and potential information content already emerged from the preliminary analysis.

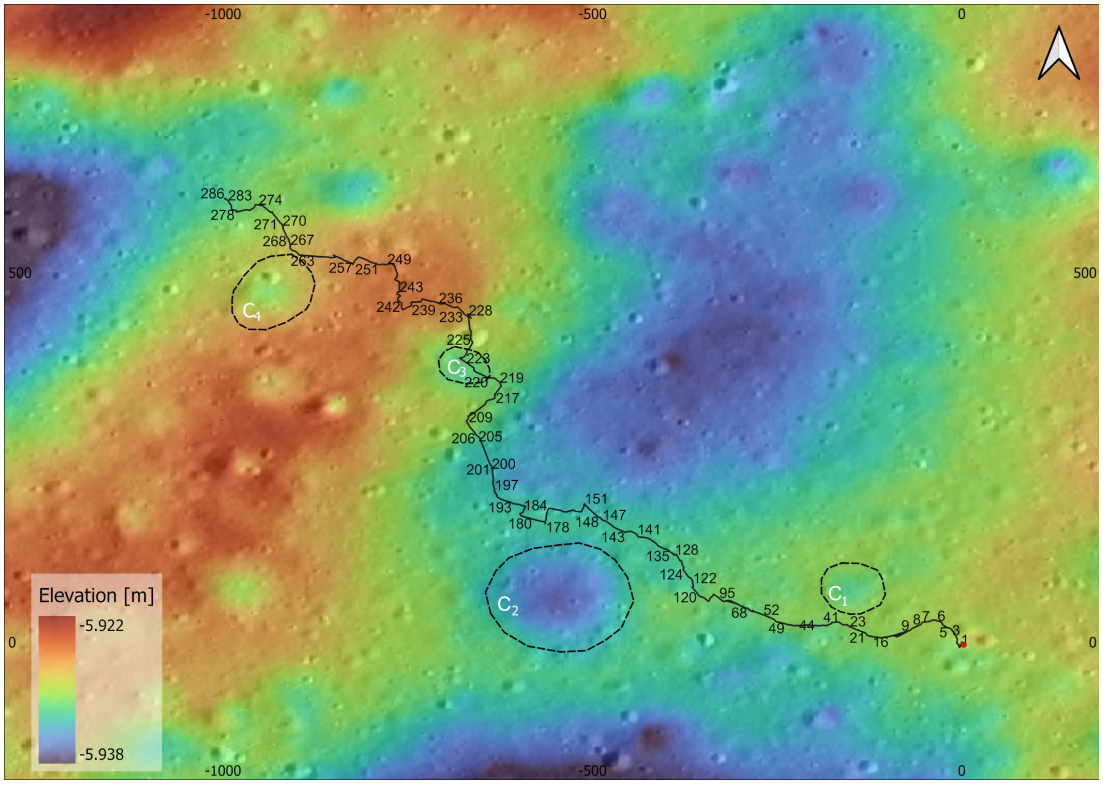

Figure 5.3.1: Rover path with waypoint numbers superimposed on a fusion of topography and satellite image. C_n refer to craters and related structures described in the text. The red dot marks the landing point. Coordinates are in meters relative to the landing point (0,0). Data from: https://quickmap.lroc.asu.edu/.

We first address the problem of reflectors extraction by applying a new automated method based on Deep Learning techniques, which provides objective and reliable results and has proven its effectiveness in different environments, datasets, and signal-to-noise ratios (Roncoroni et al., 2022a) (Figure 5.3.2 bottom). We then applied a combination of signal attributes that have already been successfully exploited on GPR datasets, e.g. Sénéchal et al. (2000); Forte et al. (2012); Zhao et al. (2018), to further constrain and improve data interpretation.

For the first time, we here show the high-frequency LPR data (CH-2) recorded until 27^{th} March 2023, representing the longest dataset ($\approx 1,440m$) openly available at the time of writing, adding more than 700 m to the longest high frequency profile published so far (Chen et al., 2022). We show new structures previously non considered or not imaged, summarizing, and partially re-interpreting the ones already described.

Overview of the landing site and geological context The entire CE-4 landing region exhibits a superposition of complex impact morphologies spanning from the pre-Nectarian to Copernichan epochs. The oldest structure is the SPA Basin, interpreted as one of the oldest, if not the oldest recognizable lunar basin ($\approx 4.3Ga$, Fernandes et al. (2013); White et al. (2020)). The Von Kármán crater was predominantly dated as pre-Nectarian, i.e. $\approx 4.2Ga$ (Lu et al., 2021) though other references describe it as Nectarian ($\approx 4Ga$) (Feng et al., 2022). The neighboring impacts, notably Finsen, Alder, Leibniz, Maksutov, and Von Kármán L and L' (Fig. Sup. S1), produced ejecta materials that filled in and affected the bottom of the Von Kármán crater (Lu et al., 2021; Huang et al., 2018; Chang et al., 2021). The northern and eastern parts of Von Kármán are covered by ejecta from Lebnitz and Finsen craters, respectively, while the western part is flooded by mare basalts (Lu et al., 2021; Huang et al., 2018), (Fig. Sup. S1, S2). The timing and relative sequence of these ejecta depositions and basalt flows are relevant for interpreting the local stratigraphy at the CE-4 landing site (Lai et al., 2020). However, the studies up to date have shown a persistent inconsistency in the interpretation of the local stratigraphy (Chang et al., 2021).
The Yutu-2 rover LPR profiles have been interpreted to show that the post-mare deposits at the CE-4 landing site are up to $\approx 45m$ thick, while a recent article (Giannakis et al., 2024) suggests a shallow basaltic lava layers starting at $\approx 10m$ depth. the Finsen crater has been unequivocally described as the dominant source of ejecta that covers the landing site (Lu et al., 2021; Huang et al., 2018; Xiao et al., 2021; Xu et al., 2021). Less agreement has been reached on the exact age of the Finsen crater, as it was reported to be either Eratoshenian (Fortezzo et al., 2020) ($\approx 3.0 - 3.1Ga$ (Lu et al., 2021; Chang et al., 2021)) or Late Imbrian ($\approx 3.5Ga$, Guo et al. (2021) or $\approx 3.6Ga$, Ivanov (2018)). Further inconsistencies include the significance of the Alder crater Imbrian (Lu et al., 2021) or Nectarian (Chang et al., 2021) ejecta in the topmost ($> 45m$) layer. While the interpretation of the early (and therefore shorter) Yutu-2 data considered it a prominent component (Lai et al., 2020), subsequent studies found Alder crater ejecta to be negligible in the topmost layer, in agreement with remote sensing interpretations (Huang et al., 2018), and to possibly

only occur beneath the youngest mare basalts at greater depths (more than 50m, Lu et al. (2021); Chang et al. (2021); Xu et al. (2021)). Most recently, four craters have been identified as principal sources of primary ejecta at the CE-4 landing site, and their most likely emplacement sequence from older to younger (Xu et al., 2021) is: Maksutov, Von Kármán L', Von Kármán L (all late-Imbrian), and then Finsen. The ejecta delivered by larger and older impacts like Leibnitz and Schrödinger, or as distant as Imbrium or Orientale, are expected at greater depth, not accessible by CH-2, beneath mare basalts (Xiao et al., 2021). The mare basalts flooded the floor of the Von Kármán in several episodes, namely between $\approx$ 3.15 and 3.75 Ga (Ling et al., 2019). Those deeper structures and stratigraphy, for instance, the oldest basalt flows, that occur at depths greater than $\approx 50m$ have been assessed using the lower-frequency CH-1 (Lai et al., 2020). However, their reliability is still debated (Cao et al., 2023).

Methods

Radar data pre-processing is a crucial step before data analysis and interpretation of subsurface structures. In addition to the normal processing flow, that is performed also on the earth GPR data, we observed problems related to duplicated traces and data file stitching (Lai et al., 2020). Redundant data removal is a critical step due to the acquisition system, since the rover Yutu-2 stops to acquire other measurements like panoramic cam or visible near infrared spectroscopy without interrupting the acquisition of LPR data. This process generates raw data with local redundancies that need to be removed. We have designed an algorithm capable of performing this removal automatically and minimizing the subjectivity of the procedure, saving time, and avoiding residual duplications (Fig. Sup. S13); see Roncoroni et al. (tteda) and `https://github.com/Giacomo-Roncoroni/LPR_CE4/` for further detail about the algorithm and its application on data.

Moreover, data acquired in different days are stored separately in different files (SOL) and need to be merged to get a manageable full dataset. 634,419 A-scans (i.e. traces) for a total length of the path equal to $\approx 1440m$ within the SOL range between 01 (4^{th} January 2019) and 286 (27^{th} March 2023) have been released at the moment of writing (August 2023). The complete original dataset after duplicated trace removal encompasses 40022 traces and is made available in different formats and versions in Roncoroni et al., submitted where all the data parameters, processing details, and data at various steps are also provided.

Beside standard processing steps, one of the commonly applied GPR processing

algorithms is migration: its purpose is to correct for the distortions that can occur in the recorded signals due to both subsurface dipping reflectors, and diffraction of the electromagnetic waves (scattering).
The migration changes the reflector dip, location and length only if they are not horizontal, while in the latter case they are not modified anymore (Yilmaz, 2001b). Since for migration EM velocity model is the most crucial parameter, we chose not to apply it since the rover's non-linear path (Fig. Sup. S8, S9) and the out-of-plane hyperbolas (Jiao et al., 2000) did not allow us to retrieve a trustful velocity model. On the other hand, migration can surely focus diffraction hyperbolas, but such a peculiar shape is very helpful in localizing scatterers.
Therefore, the migration procedure would not allow us to retrieve better resolution on the horizon and would potentially heavily degrade the imaging of deeper horizons, for which there are no reliable constraints on the velocity model and signal degradation is expected due to border effects (Yilmaz, 2001b).

For similar reasons, time-to-depth conversion was done using a constant EM velocity equal to 0.16 m/ns (Roncoroni et al., tteda). It is certainly true that a more detailed velocity field could be reconstructed exploiting diffraction hyperbolas, but as previously pointed out, there are relevant and insuperable issues and limitations.

LPR Horizon extraction

For the automatic horizon extraction, Fig. 2b and b', we modified the workflow proposed in Roncoroni et al. (2022a) and Roncoroni et al. (2022b), for GPR measurements, implementing and exploiting a Neural Network (NN) that takes both the data amplitude and the cosine of the instantaneous phase, as input. The entire train model and codes can be found in `https://github.com/Giacomo-Roncoroni/CE4-HrEx`.
The algorithm utilizes a Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber, 1997b) architecture to maintain the causality of the data and take advantage of its ability to better fit the physics behind wave propagation. The use of Bi-Directional LSTM is also employed to improve the accuracy of NN classification. The output of the NN is driven by a dense layer with two neurons and a SoftMax activation function (Mannor et al., 2005b) that outputs a probability value indicating the presence of reflections as a function of time.
We trained the NN using a synthetic dataset to eliminate potential biases arising from the field dataset and to exert full control over the NN performance through the known subsurface model that generated the training data. Therefore, the first prediction output is given as a probability set, where each point is associated with a probability

value indicating its likelihood of belonging to a reflecting surface. Then, to obtain a binary indicator we set a threshold. The reference output in this way represented by a binary indicator $[0-1]$ labeling each sample as either reflection or no reflection, respectively. The optimum threshold is estimated by evaluating the number of points classified as reflectors at various threshold values, and selecting the sharp inflection point visible in the resulting curve. This method minimizes the subjectivity of the choice and is applied as a constant on all data.
Since we are working with a 1-D methodology, to reduce the noise effect we trained the NN to predict the whole wave package and not only its maximum phase, as performed in Roncoroni et al. (2022a). To mitigate the uncertainty in predictions, an ensemble learning strategy was employed. This strategy leverages multiple learning algorithms to achieve improved predictive outcomes (Mendes-Moreira et al., 2009a). This methodology resulted in two separate predictions, which were then combined using their geometric mean, as it provided better results compared to the arithmetic mean. This can be attributed to the nature of the prediction, where probabilities in the range $[0-1]$ are being predicted.

LPR attributes analysis

Attribute analysis is a technique used to extract features and information from GPR data to support interpretation and data analysis and at first exploited for reflection seismic data (Chopra and Marfurt, 2007). In this paper we used several attributes to get a more detailed and constrained LPR interpretation and to verify and validate the results obtained with the automated horizon extraction. in particular we calculated:

- Cosine of the instantaneous phase (Chopra and Marfurt, 2007) (a.k.a. cosine of phase): it is a complex and amplitude independent attribute that clearly displays bedding (Fig. E1).
- Instantaneous Frequency (Chopra and Marfurt, 2007) it is a complex attribute, commonly used for highlighting specific events, such as abnormal attenuation and thin bed tuning (Fig. E4).
- Sweetness (Oliveros and Radovich, 1997): It is an attribute computed by dividing the trace envelope by the square root of the instantaneous frequency (Fig. E3, E8, E9). It is able to characterize and emphasize differences between various facies.

Results

Topography and DTM analysis of the landing site

The topography of the terrain where the rover landed is dominated by sub-parallel ejecta rays interpreted to be originated mainly from the Finsen crater (Fig. Sup. S3). This is reflected in the distribution of alternating $\approx 400 - 500m$ wide topographic low and high zones, occasionally connected by lower-lying bridging material (Figure 5.3.1, Fig. Sup. S5). The rover has landed on a relatively high zone and the first 400 m of the path covered these ejecta-rich strata. The path then continues across a lower zone about 500 m wide that is followed by another high zone from about 1,000 to 1,400 m along the path (Figure 5.3.1). The final $\approx 100m$ of the path, disclosed until now, are towards another low elevation zone. This topography plays an essential role in ejecta distribution, but earlier studies have not fully considered it.

LPR profile interpretation and stratigraphy

We describe different stratigraphic units interpreted by exploiting LPR data of the $\approx 1,440m$ long rover path, focusing on the first $\approx 50m$ depth. The rover path is presented in Figure 5.3.1 (the landing site is marked by the red dot), and the radar profile in Figure 5.3.2 (the landing site is on the left side). In discussing the results, we outline the general stratigraphic units as obtained by automated NN extraction integrated by radar attributes evaluation (see Methods).

The new Deep Learning horizon extraction (Figure 5.3.2b) interpreted some reflectors which are almost continuous along the entire profile, as well as other horizons present only in specific locations. By integrating the reflector probability (Figure 5.3.2b) with the reflection amplitude (Figure 5.3.2a) and integrated attribute analyses (Figure 5.3.3; Fig. Sup. S15-S20) we can interpret single horizons and both their spatial correlation and facies: similar colours represent the same stratigraphic level along the radar profile (Figure 5.3.2b, c; Figure 5.3.3).

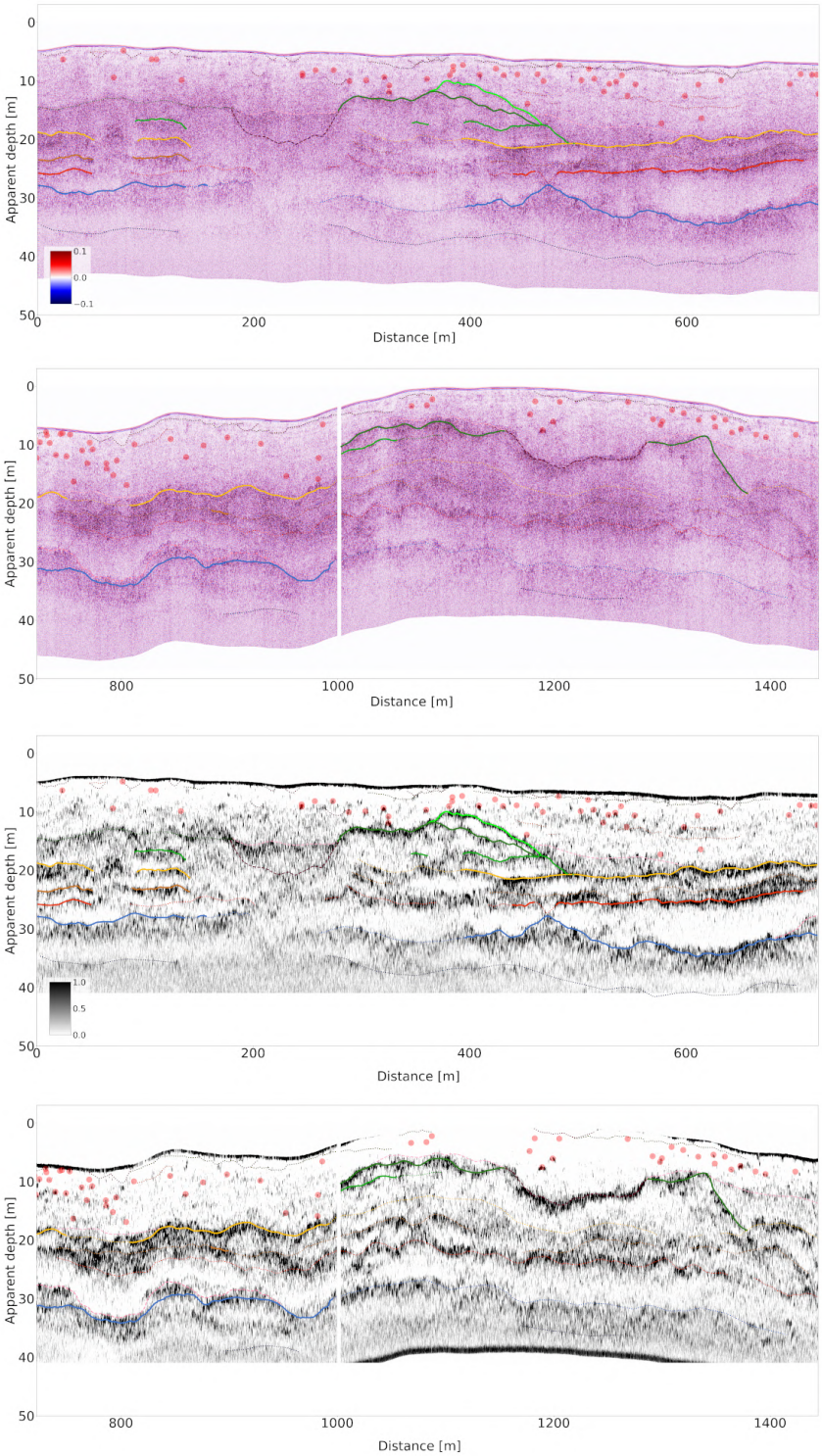

Figure 5.3.2: Total interpreted LPR dataset in amplitude (a, a') and automated Deep Learning horizons extraction (b, b'). Light red dots represent localized scatterers, while continuous, dashed and dotted lines follow the main recognized reflectors.

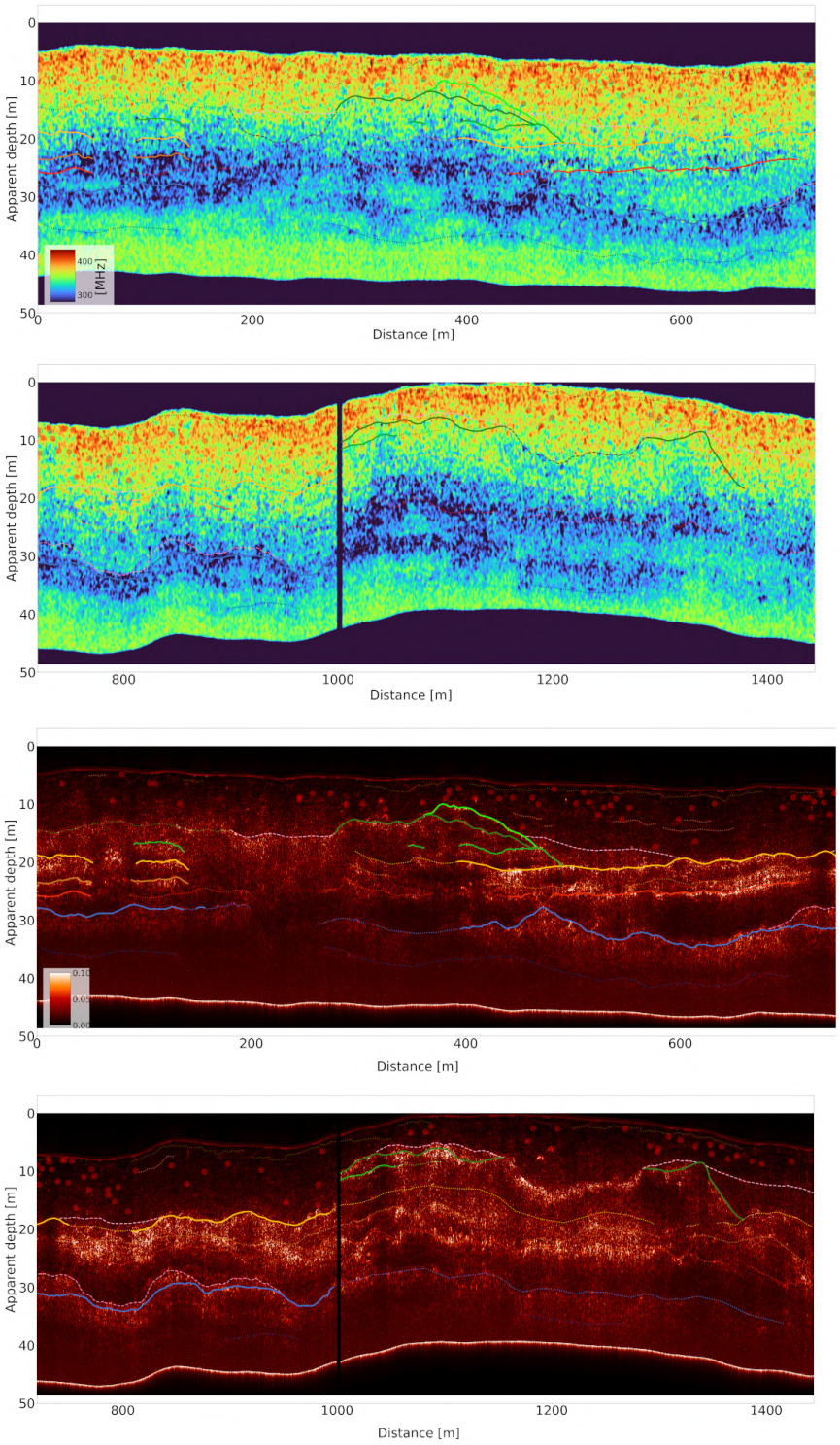

Figure 5.3.3: LPR total interpreted dataset: smoothed dominant frequency (a, a'); sweetness (b, b').

Moreover, we interpreted beneath the entire rover path different electromagnetic (EM) units (U1-U4) on the basis of their EM signature, geometry and relative location, as detailed below (Figure 5.3.4).

U1 has a low overall reflectivity without clear and coherent high amplitude reflections. This unit can be observed from the topographic surface down to an almost constant depth of $\approx 12m$, seen from the landing site position to a distance of 275 m, decreasing down to as low as 6 m at a distance of 380 m and then approaching 15 m between 500 and 950 m (Figure 5.3.4). Its thickness is again lower (5-10m) until 1,350 m then increases again at the end of the profile. Within U1 we imaged two new types of structures, namely: several (at least 20) with a concave shape, while several others are local sub-horizontal reflectors. None of these features have been previously reported, probably because they have an overall low signal amplitude, but appear very clear when phase or other composite attributes like sweetness (see Methods) are considered (Figure 5.3.3, S17). The concave structures appear close to the surface and have a mean width equal to 16.4 m (maximum 23.2 m; minimum 8.3 m) and some of them are partially overlapping. We interpret all those structures as filled craters produced by either small meteorites or, most likely, as secondary craters that are very frequent in this area. Some of these craters appear concealed at the top with quite discontinuous but still recognizable sub-horizontal reflector. In addition to these reflectors, some other deeper and significantly longer ones have been imaged within U1 and specifically between 450 and 780 m. They show a maximum (apparent) equal to 6^{deg} between 450 and 500 m where they lie over dipping layers of U2. The maximum lateral extension of a single reflector reaches 70 m, demonstrating that the regolith is not entirely chaotic but, at least locally, layered and showing stratification that follows the former (i.e. deeper) morphology.

Stronger reflectors are beneath $\approx 12m$ depth from the beginning of the profile, including discrete horizontal and slightly dipping reflectors down to a depth of $\approx 30m$. These layered zones can be classified into two separate stratigraphic units based on their amplitude, signature and lateral continuity (Fig. Sup. S16-S18): the top one (U2) that is $\approx 8 - 10m$ thick and entails two roughly equally thick layers (green in Figure 5.3.2 and 5.3.4), and the slightly thicker beneath ($\approx 12m$, U3) that contains up to three layers (orange to red in Figure 5.3.2, light blue in Figure 5.3.4). U2 is present in the first 480 m of the profile and from $\approx$ 950 to 1350 m, while U3 can be observed throughout the profile. At a depth of $\approx 30m$, a strong reflector appears (with local lower reflectivity) with significantly different characteristics from those of the facies above (see e.g. Fig. Sup. S18), and persists laterally throughout the observed profile (U4 in Figure 5.3.4). Actual stratigraphic structure is also determined by the excavated local materials mixed with that ejecta and reworked by multiple impacts.

The final stratigraphic layers rather reflect the mixture of the primary ejecta and the excavated local materials (i.e., ejecta deposits) (Xu et al., 2021) and the thicknesses of the U1-U4 layers described here are broadly in agreement with previously interpreted thicknesses. In this regard, the U1 has been interpreted as fine-grained regolith (e.g. Lai et al. (2020); Chen et al. (2022); Zhang et al. (2021)) dominated by Finsen ejecta which was then reworked, mixed and overturned by numerous impacts but compositionally it is very similar to the ejecta itself (Dong et al., 2020; Ling et al., 2019; Guo et al., 2021). In addition to several low amplitude interfering events made clear by phase analysis (see e.g. Figs.S16), there are some localized scatterers having different amplitudes, alternatively interpreted as decimeter-sized boulders ejected during the formation of Finsen crater, including an unknown fraction of local rocks (Chen et al., 2022), or as broken pieces of glass-bearing breccia projectiles excavated from pre-existing small craters on the lunar far side Ling et al. (2019). Some authors further divide U1 into two sub-units: the topmost is more homogeneous with weaker amplitude because the surface materials have undergone a longer weathering period, while the lower portion has a high overall reflectivity interpreted as a less weathered material (Zhang et al., 2021). This division is not apparent in our analysis, even if the shallower part of U1 seems to have higher numbers of scatterers than the deeper one. Scatterers produce diffractions hyperbolas on radar sections which have been exploited to estimate the EM velocity (and from it the dielectric permittivity) of this shallow zone (e.g. refs. (Dong et al., 2020, 2021; Lai et al., 2020; Chen et al., 2022) even if the rover path is not straight and its speed is not constant thus resulting in distorted hyperbolic patterns, as previously pointed out.

Discrete layers within U2 and U3 ($\approx 12 - 30m$ deep) correspond to what has been previously described as different coarse ejecta deposits (i.e., the mixture zone of Finsen's primary ejecta, pre-Finsen primary ejecta (Maksutov, Von Kármán L', and Von Kármán L), and local basalt materials) (Xu et al., 2021)).

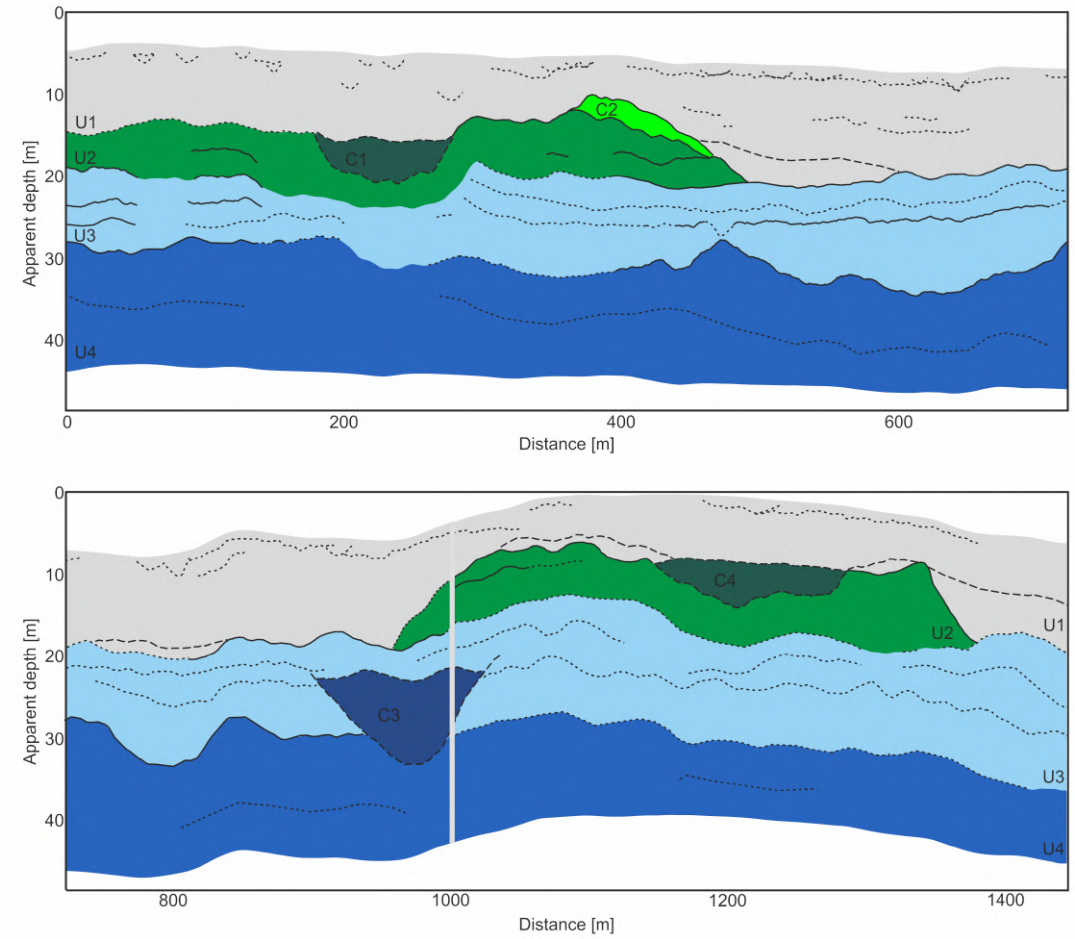

Figure 5.3.4: Subsurface units assessment from LPR data interpretation (Figures 5.3.2, 5.3.3). U1 (grey), U2 (in green), U3 (in light blue), U4 (in blue) represent different macro units. C1, C2', C3 and C4 are interpreted as paleo-crater filling and related materials, while dotted and dashed lines mark layering and peculiar structures within the main units (see text for description and details).

The lowest U4 most probably represents a mare basalt layer ¿ 30 m deep, as already pointed out by several studies (e.g. Zhang et al. (2021); Guo et al. (2021)), with some discontinuous highly attenuated internal layers also due to low overall signal-to-noise ratio. This seems to be confirmed by the signal frequency behaviour (Figure 5.3.3); however, a conclusive interpretation of such a unit is not possible just from the analysis of LPR data. In general, the observed stratigraphy is quite in agreement with strata described earlier (Xu et al., 2021; Zhang et al., 2021; Lai et al., 2021), while notable discrepancies and new imaged structures will be discussed especially regarding U1, U2 and paleo craters and related structures C1 to C4.

Results

The landing site of Chang'E-4 shows a morphology with alternating topographic lows and highs reflecting Finsen ejecta rays, which are transected by the Yutu-2 rover (Figure 5.3.1, 5.3.5 + supplementary). In particular, based on a fine-scale DTM map we detected four crater shapes (C1 to C4 in Figure 5.3.1, 5.3.5) crossed or very close to the rover path. Crater C1 is buried just below the regolith and developed within U2. It was first interpreted by Zhou et al. (2021) and then confirmed by several other authors. We estimated its maximal excavation depth (d) as a function of the crater diameter (D) using the relation $d = 0.084D$ (Melosh, 1989; Warner et al., 2017) (Figure 5.3.4). The obtained result of $8.8m$, being $D = 105m$, is in very good agreement with the interpreted paleo-crater which extends to a maximum depth based on the LPR data of $8.0m$. If we consider the entire zone in which the layering is absent as the crater diameter, having an extension at its top of about $125m$, an excavation depth of 10.5 m is obtained: it matches the vertical extension in which the layers are absent i.e., from the white dashed line (top of the crater filling materials) down to the red dashed line (Figure 5.3.2).

The same analysis was performed on C3 and C4. C3 has a diameter equal to about $142m$ and an estimated excavation depth from LPR data of $12.4m$ while C4 has a similar width and a depth of just $6m$. By applying the previously reported relation, we obtain a d value equal to $11.9m$ which is in good agreement with C3 while it is not with C4 possible because this latter crater was modified after the main impact, as suggested also by its very irregular shape.

A similar analysis performed on the 20 shallow craters (some of them being coalescent), (Figure 5.3.2) gives a mean diameter of $16.4m$ and a consequent excavation depth of $1.4m$; also, in this case, it is quite similar to the one imaged by the LPR data, which values range from 0.9 to $2.4m$. The similar size and their close and regular spatial distribution suggest they were created as secondary craters.

A peculiar structure is labelled as C2' in Figure 5.3.4. It lies on the top of U2 and is apparent between $360 - 460m$ on the profile. In this portion LPR path crosses the rim of crater C2 whose center is to the south of the profile. C2 is an elliptical crater evident of the surface; it is $\approx 145m$ (North-South) by $\approx 195m$ (East-West) wide.

This smaller (max wideness in NW-SE direction equal to about 115 m) and younger crater (Figure 5.3.6) superimposed on main crater C2 rim (estimated to be younger than $100Ma$, REF) appears quite fresh. In particular, it is apparent when considering the surface azimuth (Figure 5.3.6c), even if the surface slope is smooth (Figure 5.3.6d). Indeed, from the satellite imagery (Figure 5.3.6a) this structure is not recognizable.

Very high reflectivity and reflection continuity of the LPR horizon (light green in Figure 5.3.2) imply that it entails, at least partially, possible impact melt that was created on

the rim of C2.

Deep Learning driven interpretation of the LPR data, linked with integrated attribute analysis and satellite imagery, can extract the stratigraphic horizons, correlate them spatially and group their main units, instead of obtaining by a more subjective manual line drawing. While the overall structure revealed by this method agrees reasonably well with previous observations (Feng et al., 2022; Chen et al., 2022; Zhang et al., 2021; Lai et al., 2021), among the others, it allows for the first time the recognition of unexpected and less evident sub-surface structures. Importantly, using this method, we were able to distinguish stratigraphic units with different electromagnetic characteristics, as well as recognize their correlation with the present-day topography. For instance, U2 is not present in the central part of the LPR profile ($480 - 950m$). Notably, based on topography and distribution of lows and highs, it can be observed that this segment of missing U2 unit corresponds to the low terrain where Finsen ejecta were originally less deposited (Figure 5.3.5). Based on this observation, we suggest that the U2 unit, whose top is the first strong layered reflector beneath the regolith, corresponding to the top of the Finsen ejecta. At the end of the released data, U2 is no longer present since the path is approaching another low topographic area in which Finsen ejecta is not expected (Figures 5.3.1, 5.3.5). Therefore, U2 layer can be identified as directly deposited by Finsen event.

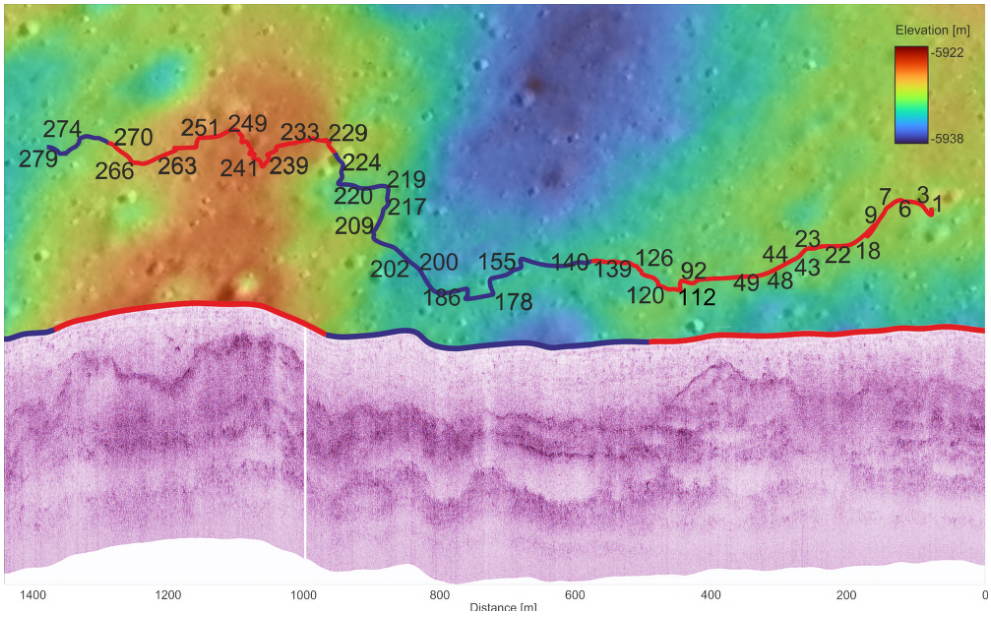

Figure 5.3.5: Correlation of surface and sub-surface structures. Plain view of the Rover path and correlation with LPR processed profile (in amplitude). Red and blue segments highlight high and low topography zones, respectively. The red dot marks the landing point. Data from: https://quickmap.lroc.asu.edu/.

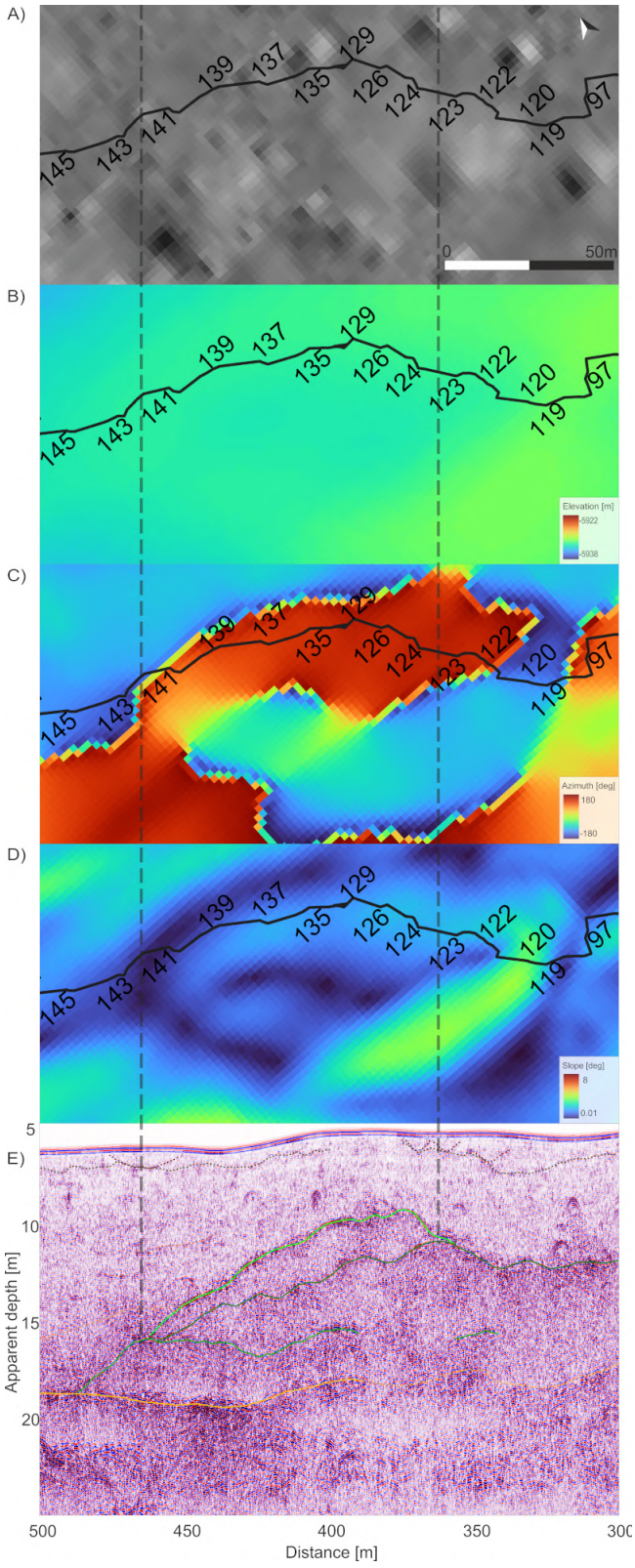

Figure 5.3.6: Analysis of the LPR portion between SOL 95 and 145 and comparison with surface morphology. a) satellite photograph; b) surface elevation; c) aspect; d) slope; e) LPR interpreted data. See discussion for details.

All available studies consider a homogeneous regolith without clear internal reflectors, only with local high amplitude scatterers with the exception of Feng et al. (2023) which did not directly recognize layering within the regolith but evidenced lateral and vertical macro electrical permittivity changes within it, on the base of an algorithm for permittivity estimation exploiting a new approach for diffraction hyperbolas fitting, limited to the shallower $150ns$ (i.e. $12m$ considering a constant EM velocity equal to $0.16m/ns$). In any case, their proposed model is 1-D and so, by definition, the obtained layers are perfectly parallel and horizontal. We here show not only that the layering is visible and clearly imaged on the radar profile but that it is slightly dipping and follows the deeper paleo-topography. This observation is an independent support toward the fact that the Finsen ejecta were deposited in alternating lows and highs which formed a paleo-topography of the terrain as early as $\approx 3Ga$ ago, or even earlier. The regolith layers were then deposited following this paleo-relief and were not entirely annihilated by the subsequent impacts. As layers do not show complete homogeneous mixing, it is expected that the Finsen ejecta would not completely mix with ejecta from previous craters and the layers dominated by previous ejecta may still exist; however the topmost layer is a mixture of dominantly Finsen ejecta with pre-materials, in agreement with geological mapping suggesting dominance of the Finsen ejecta in this entire portion of the VK crater, crossed by SW-NE ejecta rays from Finsen and dominated by characteristic orthopyroxene (LCP) (Huang et al., 2018), Fig. Sup. S3. In addition to regolith internal layering, for the first time at least 20 shallow buried craters have been detected directly on the LPR dataset. They are not apparent on the base of reflection amplitude (Figure 5.3.7) suggesting, as expected, that the filling material is very similar to the surrounding one. However, using signal attributes and in particular phase attributes (Figure 5.3.7) the lateral limit of such a crater is evident and can be quite easily laterally recognized even when the Deep Learning horizon extraction does not clearly recognize apparent structures.

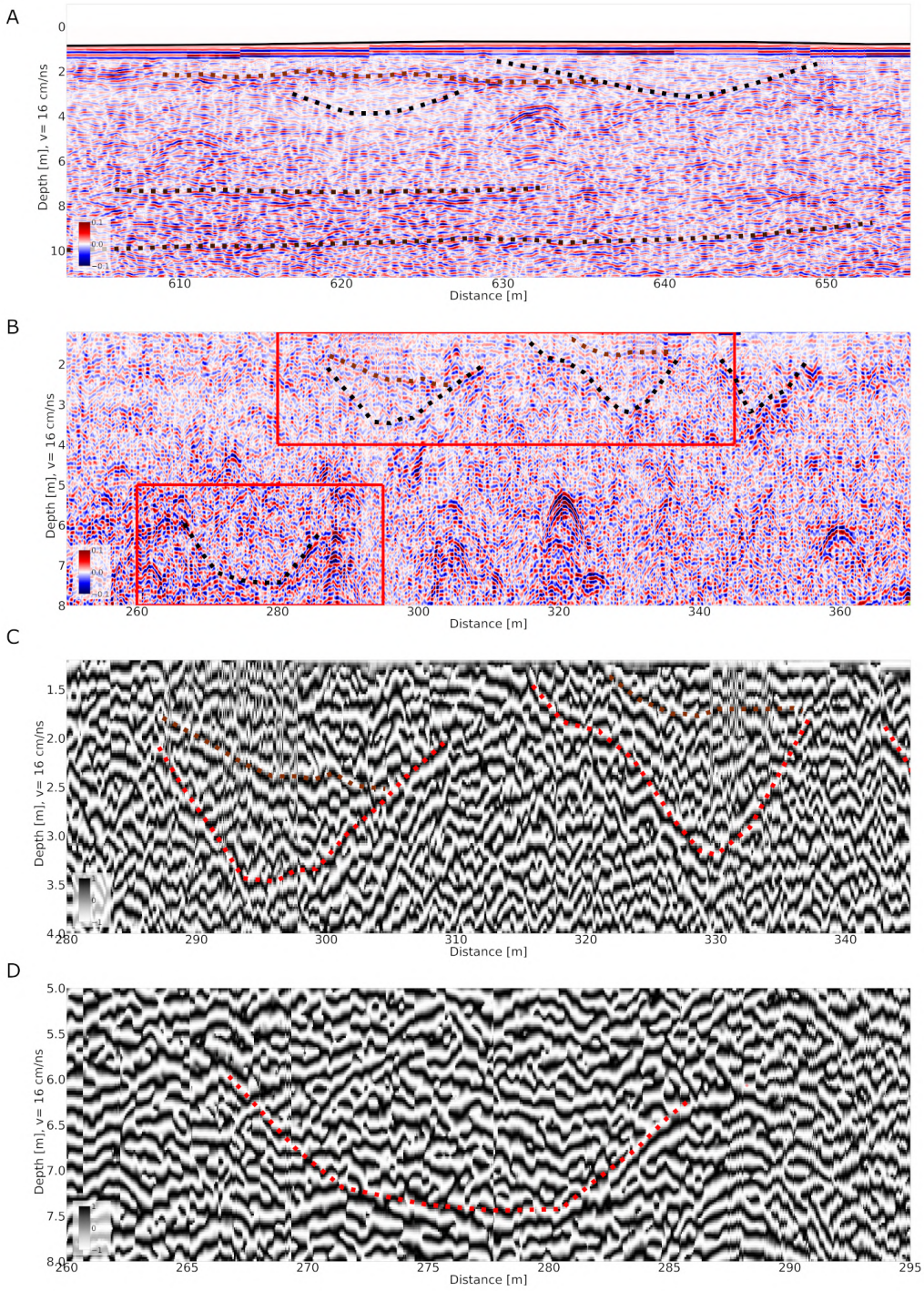

Figure 5.3.7: Details of layering and shallow crateriform structures within the regolith (U1), a) and b). c) and d) show the cosine of instantaneous phase within red boxes in b).

Conclusions

New LPR data and Deep Learning-based interpretation allowed to identify new and somewhat unexpected subsurface structures on the far side of the Moon. In particular, while available studies consider a homogeneous regolith with only local scatterers and no apparent reflectors, we not only show that several layers are visible on the LPR data, but also that they are not always horizontal but rather follow the deeper paleo-topography. The DL automated horizon probability procedure integrated with the analysis of combined signal attributes allowed the discovery of fine-scale features in regolith and ejecta layers which were not previously imaged, probably due to their low overall amplitude and elusive nature. In particular, we recognized at least 20 shallow buried crater-like structures within the regolith and further four developed deeper within different stratigraphic units. We made a relevant step forward in correlating layers and defining their different geological meaning. The LPR dataset was not interpreted as a stand-alone information, but it was fully integrated with satellite-derived information, and specifically to surface photographs and detailed elevation models, finding that subsurface units are well correlated with the present-day topography. This observation is an independent support toward the fact that the Finsen ejecta was deposited in alternating lows and highs which formed a paleo-topography of the terrain. Thanks to the exploited integrated data approach, we assigned specific geological and geomorphological meaning to the identified subsurface reflectors, defining four different units along the considered rover path, and describing their relationship.

The obtained results proved the importance of integrated analysis of Lunar data for subsurface assessment and structure identification, which are in turn crucial for possible resources evaluation. Further research will be addressed to the calculation of DL-based attributes

Acknowledgments

A. Č. acknowledges Rita Levi Montalcini fellowship by the Italian Ministry of University and Research (MUR). We thank the Chang'E-4 payload team for mission operations and China National Space Administration for providing the Chang'E-4 data that made this study possible. This work was supported by the National Natural Science Foundation of China (11773023, 11941001, U1631124) and the Civil Aerospace Pre-research Project (D020302). The Chang'E-4 data used in this work is processed and produced by "Ground Research and Application System" (GRAS) of China's Lunar and Planetary Exploration Program, it can be downloaded at http://moon.bao.ac.cn/ce5web/searchOrder_dataSearchData.search.

Open Research

Data can be found on 2019 DOI: https://dx.doi.org/10.12350/CLPDS.GRAS.CE4.LPR-2B-2019.vA 2020 DOI: https://dx.doi.org/10.12350/CLPDS.GRAS.CE4.LPR-2B-2020.vA. Ground Research and Application System of China's Lunar and Planetary Exploration Program. Chang'E-4 Lunar Penetrating Radar Level 2B scientific Dataset. China National Space Administration, 2020.

The codes related to the horizon extraction algorithm can be found at https://github.com/Giacomo-Roncoroni/CE4-HrEx, while the codes for data pre-processing at https://figshare.com/s/36c46ad26ab1aadcfcd7. Processed data are available at https://figshare.com/s/9ce7f1cb8ff0fb8d90c9.

Bibliography

Alistair R. Brown. 2009. Phase and Polarity Issues in Modern Seismic Interpretation. *AAPG Search and Discovery* 40397 (2009), V67–V76.

Charles J. Byrne. 2008. *The South Pole-Aitken Basin and the South Polar Region.* Springer New York, New York, NY, 60–93. https://doi.org/10.1007/978-0-387-73206-0_8

Eduardo Valero Cano, Jubran Akram, and Daniel B. Peter. 2021. Automatic seismic phase picking based on unsupervised machine-learning classification and content information analysis. *GEOPHYSICS* 86, 4 (2021), V299–V315. https://doi.org/10.1190/geo2020-0308.1 arXiv:https://doi.org/10.1190/geo2020-0308.1

Huaqing Cao, Yi Xu, Luyuan Xu, Ling Zhang, Roberto Bugiolacchi, and Feng Zhang. 2023. From Schrödinger to Von Kármán: An Intriguing New Geological Structure Revealed by the Chang'e-4 Lunar Penetrating Radar. *Geophysical Research Letters* 50, 2 (Jan. 2023), e2022GL101413. https://doi.org/10.1029/2022GL101413

YiRen Chang, ZhiYong Xiao, YiChen Wang, ChunYu Ding, Jun Cui, and YuZhen Cai. 2021. An updated constraint on the local stratigraphy at the Chang'E-4 landing site. *Earth and Planetary Physics* 5, 1 (2021), eepp2021007. https://doi.org/10.26464/epp2021007 arXiv:https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.26464/epp2021007 eepp2021007 epp2021007.

Ruonan Chen, Yi Xu, Minggang Xie, Ling Zhang, Shengli Niu, and Roberto Bugiolacchi. 2022. Subsurface stratification and dielectric permittivity distribution at the Chang'E-4 landing site revealed by the lunar penetrating radar. *Astronomy & Astrophysics* 664 (Aug. 2022), A35. https://doi.org/10.1051/0004-6361/202142677

Sharan Chetlur, Cliff Woolley, Philippe Vandermersch, Jonathan Cohen, John Tran, Bryan Catanzaro, and Evan Shelhamer. 2014. cuDNN: Efficient Primitives for Deep Learning. *CoRR* abs/1410.0759 (2014). arXiv:1410.0759 http://arxiv.org/abs/1410.0759

Francois Chollet et al. 2015. Keras. https://keras.io.

S. Chopra and K.J. Marfurt. 2007. *Seismic Attributes for Prospect Identification and Reservoir Characterization.* Society of Exploration Geophysicists. https://books.google.it/books?id=dP2iACuzq34C

Renato Colucci, Emanuele Forte, Chiara Boccali, M Dossi, L Lanza, Michele Pipan, and Mauro Guglielmin. 2014. Evaluation of Internal Structure, Volume and Mass of Glacial Bodies by Integrated LiDAR and Ground Penetrating Radar Surveys: The Case Study of Canin Eastern Glacieret (Julian Alps, Italy). *Surveys in Geophysics* 36, 2 (2014), 231–252. https://doi.org/10.1007/s10712-014-9311-1

Fabien Cubizolle, Thomas Valding, Sebastien Lacaze, and Fabien Pauget. 2015. *Global method for seismic-well tie based on real time synthetic model.* 1776–1781. https://doi.org/10.1190/segam2015-5862834.1

Zehua Dong, Guangyou Fang, Di Zhao, Bin Zhou, Yunze Gao, and Yicai Ji. 2020. Dielectric Properties of Lunar Subsurface Materials. *Geophysical Research Letters* 47, 22 (Nov. 2020), e2020GL089264. https://doi.org/10.1029/2020GL089264

Zehua Dong, Guangyou Fang, Bin Zhou, Di Zhao, Yunze Gao, and Yicai Ji. 2021. Properties of Lunar Regolith on the Moon's Farside Unveiled by Chang'E-4 Lunar Penetrating Radar. *Journal of Geophysical Research: Planets* 126, 6 (2021), e2020JE006564. https://doi.org/10.1029/2020JE006564 arXiv:https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020JE006564 e2020JE006564 2020JE006564.

Geoffrey A. Dorn. 1998. Modern 3-D seismic interpretation. *The Leading Edge* 17, 9 (1998), 1262–1262. https://doi.org/10.1190/1.1438121

Guang-You Fang, Bin Zhou, Yi-Cai Ji, Qun-Ying Zhang, Shao-Xiang Shen, Yu-Xi Li, Hong-Fei Guan, Chuan-Jun Tang, Yun-Ze Gao, Wei Lu, Sheng-Bo Ye, Hai-Dong Han, Jin Zheng, and Shu-Zhi Wang. 2014. Lunar Penetrating Radar onboard the Chang'e-3 mission. *Research in Astronomy and Astrophysics* 14, 12 (Dec. 2014), 1607–1622. https://doi.org/10.1088/1674-4527/14/12/009

Jianqing Feng, Matthew. A. Siegler, Yan Su, Chunyu Ding, and Iraklis Giannakis. 2023. Layered Structures in the Upper Several Hundred Meters of the Moon Along the Chang'E-4 Rover's First 1,000-m Traverse. *Journal of Geophysical Research: Planets* 128, 8 (2023), e2022JE007714. https://doi.org/10.1029/2022JE007714 arXiv:https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2022JE007714 e2022JE007714 2022JE007714.

Jianqing Feng, Matthew A. Siegler, and Mackenzie N. White. 2022. Dielectric properties and stratigraphy of regolith in the lunar South Pole-Aitken basin: Observations from the Lunar Penetrating Radar. *Astronomy & Astrophysics* 661 (May 2022), A47. https://doi.org/10.1051/0004-6361/202143015 arXiv:2203.02840 [astro-ph, physics:physics].

Vera Fernandes, Jörg Fritz, Ben Weiss, Ian Garrick-Bethell, and David Shuster. 2013. The bombardment history of the Moon as recorded by 40Ar-39Ar chronology. *Meteoritics & Planetary Science* 48, 2 (2013), 241–269. https://doi.org/10.1111/maps.12054 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/maps.12054

Emanuele Forte, M. Dossi, Michele Pipan, and Renato Colucci. 2014. Velocity analysis from common offset GPR data inversion: theory and application to synthetic and real data. *Geophysical Journal International* 197, 3 (June 2014), 1471–1483. https://doi.org/10.1093/gji/ggu103

Emanuele Forte, Matteo Dossi, Michele Pipan, and Anna Del Ben. 2016. Automated phase attribute-based picking applied to reflection seismics. *Geophysics* 81, 2 (2016), V141–V150. https://doi.org/10.1190/GEO2015-0333.1.

Emanuele Forte, Michele Pipan, D. Casabianca, Raffaele Di Cuia, and Alberto Riva. 2012. Imaging and characterization of a carbonate hydrocarbon reservoir analogue using GPR attributes. *Journal of Applied Geophysics* 81 (June 2012), 76–87. https://doi.org/10.1016/j.jappgeo.2011.09.009

C M Fortezzo, P D Spudis, and S L Harrel. 2020. RELEASE OF THE DIGITAL UNIFIED GLOBAL GEOLOGIC MAP OF THE MOON AT 1:5,000,000-. (2020).

Riccardo Geletti, Fabrizio Zgur, Anna Ben, F. Buriola, Silvana Fais, Maurizio Fedi, Emanuele Forte, Arianna Mocnik, Valeria Paoletti, Michele Pipan, Riccardo Ramella, Roberto Romeo, and Alessandro Romi. 2014. The Messinian Salinity Crisis: New seismic evidence in the West-Sardinian Margin and Eastern Sardo-Provençal basin (West Mediterranean Sea). *Marine Geology* 351 (2014), 76–90. https://doi.org/10.1016/j.margeo.2014.03.019

Iraklis Giannakis, Javier Martin-Torres, Yan Su, Jianqing Feng, Feng Zhou, Maria-Paz Zorzano, Craig Warren, and Antonios Giannopoulos. 2024. Evidence of shallow basaltic lava layers in Von Kármán crater from Yutu-2 Lunar Penetrating Radar. *Icarus* 408 (Jan. 2024), 115837. https://doi.org/10.1016/j.icarus.2023.115837

Iraklis Giannakis, Feng Zhou, Craig Warren, and Antonios Giannopoulos. 2021. Inferring the Shallow Layered Structure at the Chang'E-4 Landing Site: A Novel Interpretation Approach Using Lunar Penetrating Radar. *Geophysical Research Letters* 48, 16 (2021), e2021GL092866. https://doi.org/10.1029/2021GL092866 arXiv:https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2021GL092866 e2021GL092866 2021GL092866.

Dijun Guo, Wenzhe Fa, Xiaojia Zeng, Jun Du, and Jianzhong Liu. 2021. Geochemistry of the Von Kármán crater floor and thickness of the non-mare ejecta over the Chang'e-4 landing area. *Icarus* 359 (May 2021), 114327. https://doi.org/10.1016/j.icarus.2021.114327

R. Guo, J.J. Zhang, D. Liu, Y.B. Zhang, and D.W. Zhang. 2019. Application of Bi-directional Long Short-Term Memory Recurrent Neural Network for Seismic Impedance Inversion. 2019, 1 (2019), 1–5. https://doi.org/10.3997/2214-4609.201901386

Donald Herron. 2014. Pitfalls in horizon autopicking. *Interpretation* 3 (08 2014), SB1–SB4. https://doi.org/10.1190/INT-2014-0062.1

Sepp Hochreiter and Jürgen Schmidhuber. 1997a. Long Short-Term Memory. *Neural Comput.* 9, 8 (11 1997), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Sepp Hochreiter and Jürgen Schmidhuber. 1997b. Long Short-Term Memory. *Neural Comput.* 9, 8 (11 1997), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Jack Hoyes and Thibaut Cheret. 2011. A review of "global" interpretation methods for automated 3D horizon picking. *The Leading Edge* 30, 1 (2011), 38–47. https://doi.org/10.1190/1.3535431

Jun Huang, Zhiyong Xiao, Jessica Flahaut, Mélissa Martinot, James Head, Xiao Xiao, Minggang Xie, and Long Xiao. 2018. Geological Characteristics of Von Kármán Crater, Northwestern South Pole-Aitken Basin: Chang'E-4 Landing Site Region. *Journal of Geophysical Research: Planets* 123, 7 (2018), 1684–1700. https://doi.org/10.1029/2018JE005577 arXiv:https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2018JE005577

Tyler W. Hughes, Ian A. D. Williamson, Momchil Minkov, and Shanhui Fan. 2019. Wave physics as an analog recurrent neural network. *Science Advances* 5, 12 (12 2019). https://doi.org/10.1126/sciadv.aay6946

Boris A. Ivanov. 2018. Size-Frequency Distribution of Small Lunar Craters: Widening with Degradation and Crater Lifetime. *Solar System Research* 52, 1 (Jan. 2018), 1–25. https://doi.org/10.1134/S0038094618010021

Yingzhuo Jia, Yongliao Zou, Jinsong Ping, Changbin Xue, Jun Yan, and Yuanming Ning. 2018. The scientific objectives and payloads of Chang'E-4 mission. *Planetary and Space Science* 162 (Nov. 2018), 207–215. https://doi.org/10.1016/j.pss.2018.02.011

Yuren Jiao, George A McMechan, and Elena Pettinelli. 2000. In situ 2-D and 3-D measurements of

radiation patterns of half-wave dipole GPR antennas. *Journal of Applied Geophysics* 43, 1 (Jan. 2000), 69–89. https://doi.org/10.1016/S0926-9851(99)00048-8

Taskin Kavzoglu. 2009. Increasing the accuracy of neural network classification using refined training data. *Environmental Modelling & Software* 24 (07 2009), 850–858.

Diederik Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations* (12 2014).

Jialong Lai, Yi Xu, Roberto Bugiolacchi, Xu Meng, Long Xiao, Minggang Xie, Bin Liu, Kaichang Di, Xiaoping Zhang, Bin Zhou, Shaoxiang Shen, and Luyuan Xu. 2020. First look by the Yutu-2 rover at the deep subsurface structure at the lunar farside. *Nature Communications* 11, 1 (July 2020), 3426. https://doi.org/10.1038/s41467-020-17262-w

Jialong Lai, Yi Xu, Roberto Bugiolacchi, Hon Kuan Wong, Luyuan Xu, Xunyu Zhang, Ling Zhang, Xiaoping Zhang, Long Xiao, Bin Liu, and Kaichang Di. 2021. A Complex Paleo-Surface Revealed by the Yutu-2 Rover at the Lunar Farside. *Geophysical Research Letters* 48, 20 (2021), e2021GL095133. https://doi.org/10.1029/2021GL095133 arXiv:https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2021GL095133 e2021GL095133 2021GL095133.

Chunlai Li, Yan Su, Elena Pettinelli, Xing Shuguo, Chunyu Ding, Jianjun Liu, X. Ren, Sebastian Lauro, F. Soldovieri, Xingguo Zeng, Xingye Gao, Wangli Chen, Shun Dai, Dawei Liu, Guangliang Zhang, Zuo Wei, Weibin Wen, Zhoubin Zhang, Xiaoxia Zhang, and Hongbo Zhang. 2020. The Moon's farside shallow subsurface structure unveiled by Chang'E-4 Lunar Penetrating Radar. *Science Advances* 6, 9 (2020), eaay6898. https://doi.org/10.1126/sciadv.aay6898 arXiv:https://www.science.org/doi/pdf/10.1126/sciadv.aay6898

Chunlai Li, Shuguo Xing, Sebastian E. Lauro, Yan Su, Shun Dai, Jianqing Feng, Barbara Cosciotti, Federico Di Paolo, Elisabetta Mattei, Yuan Xiao, Chunyu Ding, and Elena Pettinelli. 2018. Pitfalls in GPR Data Interpretation: False Reflectors Detected in Lunar Radar Cross Sections by Chang'e-3. *IEEE Transactions on Geoscience and Remote Sensing* 56, 3 (March 2018), 1325–1335. https://doi.org/10.1109/TGRS.2017.2761881

Zongcheng Ling, Le Qiao, Changqing Liu, Haijun Cao, Xiangyu Bi, Xuejin Lu, Jiang Zhang, Xiaohui Fu, Bo Li, and Jianzhong Liu. 2019. Composition, mineralogy and chronology of mare basalts and non-mare materials in Von Kármán crater: Landing site of the Chang'E-4 mission. *Planetary and Space Science* 179 (Dec. 2019), 104741. https://doi.org/10.1016/j.pss.2019.104741

Jesse Lomask and Antoine Guitton. 2007. Volumetric flattening: an interpretation tool. *The Leading Edge* 26, 7 (2007), 888–897. https://doi.org/10.1190/1.2756869

Yu Lu, Yunzhao Wu, Gregory G. Michael, Jinsong Ma, Wei Cai, and Nannan Qin. 2021. Chronological sequence of Chang'E-4 landing zone within Von Kármán crater. *Icarus* 354 (Jan. 2021), 114086. https://doi.org/10.1016/j.icarus.2020.114086

Shie Mannor, Dori Peleg, and Reuven Rubinstein. 2005a. The Cross Entropy Method for Classification. In *Proceedings of the 22nd International Conference on Machine Learning (ICML '05)*. 561–568. https://doi.org/10.1145/1102351.1102422

Shie Mannor, Dori Peleg, and Reuven Rubinstein. 2005b. The Cross Entropy Method for Classification

(ICML '05). Association for Computing Machinery, New York, NY, USA, 561–568. https://doi.org/10.1145/1102351.1102422

G. S. Martin. 2004. The Marmousi2 model, elastic synthetic data, and an analysis of imaging and AVO in a structurally complex environment. (2004). Master's thesis. University of Houston.

H.J. Melosh. 1989. *Impact Cratering: A Geologic Process*. Oxford University Press. https://books.google.it/books?id=nZwRAQAAIAAJ

João Mendes-Moreira, Alipio Mario Jorge, Carlos Soares, and Jorge Freire de Sousa. 2009a. Ensemble Learning: A Study on Different Variants of the Dynamic Selection Approach. In *Machine Learning and Data Mining in Pattern Recognition*, Petra Perner (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 191–205.

João Mendes-Moreira, Alipio Mario Jorge, Carlos Soares, and Jorge Freire de Sousa. 2009b. nsemble Learning: A Study on Different Variants of the Dynamic Selection Approach. In *Machine Learning and Data Mining in Pattern Recognition*. Springer Berlin Heidelberg, 191–205. https://doi.org/10.1007/978-3-642-03070-3_15

R. Burnet Oliveros and Barbara J. Radovich. 1997. Image-processing display techniques applied to seismic instantaneous attributes over the Gorgon gas field, North West Shelf, Australia. In *SEG Technical Program Expanded Abstracts 1997*. Society of Exploration Geophysicists, 2064–2067. https://doi.org/10.1190/1.1885862

Elena Pettinelli, Sebastian E. Lauro, Elisabetta Mattei, Barbara Cosciotti, and Francesco Soldovieri. 2021. Stratigraphy versus artefacts in the Chang'e-4 low-frequency radar. *Nature Astronomy* 5, 9 (Sept. 2021), 890–893. https://doi.org/10.1038/s41550-021-01432-x

Giacomo Roncoroni, Emanuele Forte, Luca Bortolussi, Luca Gasperini, and Michele Pipan. 2022b. Polarity assessment of reflection seismic data: a Deep Learning approach. *BGO* 63 (Nov. 2022), 693–700. https://doi.org/10.4430/bgo00409

Giacomo Roncoroni, Emanuele Forte, Luca Bortolussi, and Michele Pipan. 2022a. Efficient extraction of seismic reflection with Deep Learning. *Computers & Geosciences* 166 (Sept. 2022), 105190. https://doi.org/10.1016/j.cageo.2022.105190

Giacomo Roncoroni, Emanuele Forte, Ilaria Santin, Ana Cernok, Andrea Rajsic, Alessandro Frigeri, and Michele Pipan. Under reviewa. High frequency Lunar Penetrating Radar quality control, editing and processing of Chang'E-4 lunar mission. (Under review). Submitted to Scientific Data.

Giacomo Roncoroni, Emanuele Forte, Ilaria Santin, Ana Cernok, Andrea Rajsic, Alessandro Frigeri, Wenke Zhao, and Michele Pipan. Under reviewb. Deep Learning driven interpretation of Chang'E4 Lunar Penetrating Radar. (Under review). Submitted to Journal of Geophysical Research: Planets.

Juan I. Sabbione and Danilo Velis. 2010. Automatic first-breaks picking: New strategies and algorithms. *GEOPHYSICS* 75, 4 (2010), V67–V76. https://doi.org/10.1190/1.3463703 arXiv:https://doi.org/10.1190/1.3463703

Ville Satopaa, Jeannie Albrecht, David Irwin, and Barath Raghavan. 2011a. Finding a Kneedle in a Haystack: Detecting Knee Points in System Behavior. 166 – 171. https://doi.org/10.1109/ICDCSW.2011.20

Ville Satopaa, Jeannie Albrecht, David Irwin, and Barath Raghavan. 2011b. Finding a "Kneedle" in a Haystack: Detecting Knee Points in System Behavior. In *2011 31st International Conference on Distributed Computing Systems Workshops*. 166–171. https://doi.org/10.1109/ICDCSW.2011.20

Tracy Stark. 2004. Relative geologic time (age) volumes—Relating every seismic sample to a geologically reasonable horizon. *Geophysics* 23 (09 2004). https://doi.org/10.1190/1.1803505

Pascale Sénéchal, Hervé Perroud, and Guy Sénéchal. 2000. Interpretation of reflection attributes in a 3-D GPR survey at Vallée d'Ossau, western Pyrenees, France. *GEOPHYSICS* 65, 5 (Sept. 2000), 1435–1445. https://doi.org/10.1190/1.1444832

Valentin Tschannen, Matthias Delescluse, Norman Ettrich, and Janis Keuper. 2020. Extracting horizon surfaces from 3D Seismic Data using Deep Learning. *Geophysics* (03 2020), 1–43. https://doi.org/10.1190/geo2019-0569.1

Bjørn Ursin. 1983. Review of elastic and electromagnetic wave propagation in horizontally layered media. *GEOPHYSICS* 48, 8 (1983), 1063–1081. https://doi.org/10.1190/1.1441529

Ruigang Wang, Yan Su, Chunyu Ding, Shun Dai, Chendi Liu, Zongyu Zhang, Tiansheng Hong, Qing Zhang, and Chunlai Li. 2021. A Novel Approach for Permittivity Estimation of Lunar Regolith Using the Lunar Penetrating Radar Onboard Chang'E-4 Rover. *Remote Sensing* 13, 18 (Sept. 2021), 3679. https://doi.org/10.3390/rs13183679

Yanghua Wang. 2015. Frequencies of the Ricker wavelet. *Geophysics* 80 (02 2015), A31–A37.

N. Warner, M. Golombek, J. Sweeney, R. Fergason, Rhona Kirk, and Claire Schwartz. 2017. Near Surface Stratigraphy and Regolith Production in Southwestern Elysium Planitia, Mars: Implications for Hesperian-Amazonian Terrains and the InSight Lander Mission. *Space Science Reviews* 211, 1-4 (Oct. 2017), 147–190. https://doi.org/10.1007/s11214-017-0352-x

Lee White, Ana Cernok, James Darling, M. Whitehouse, K. Joy, Cyril Cayron, Joseph Dunlop, Kimberly Tait, and Mahesh Anand. 2020. Evidence of extensive lunar crust formation in impact melt sheets 4,330 Myr ago. *Nature Astronomy* 4, 10 (May 2020), 974–978. https://doi.org/10.1038/s41550-020-1092-5

Weiren Wu, Chunlai Li, Wei Zuo, Hongbo Zhang, Jianjun Liu, Weibin Wen, Yan Su, Xin Ren, Jun Yan, Dengyun Yu, Guangliang Dong, Chi Wang, Zezhou Sun, Enhai Liu, Jianfeng Yang, and Ziyuan Ouyang. 2019a. Lunar farside to be explored by Chang'e-4. *Nature Geoscience* 12, 4 (April 2019), 222–223. https://doi.org/10.1038/s41561-019-0341-7

Xinming Wu, Luming Liang, Yunzhi Shi, and Sergey Fomel. 2019b. FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation. *Geophysics* 84 (02 2019), IM35–IM45. https://doi.org/10.1190/geo2018-0646.1

Xinming Wu, Shangsheng Yan, Jie Qi, and Hongliu Zeng. 2020. Deep Learning for Characterizing Paleokarst Collapse Features in 3-D Seismic Images. *Journal of Geophysical Research: Solid Earth* 125 (09 2020). https://doi.org/10.1029/2020JB019685

Zhiyong Xiao, Chunyu Ding, Minggang Xie, Yuzhen Cai, Jun Cui, Ke Zhang, and Juntao Wang. 2021. Ejecta From the Orientale Basin at the Chang'E-4 Landing Site. *Geophysical Research Letters* 48, 3 (2021), e2020GL090935. https://doi.org/10.1029/2020GL090935

arXiv:https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020GL090935 e2020GL090935 2020GL090935.

Luyuan Xu, Xunyu Zhang, Le Qiao, and Jialong Lai. 2021. Evaluating the Thickness and Stratigraphy of Ejecta Materials at the Chang'e-4 Landing Site. *The Astronomical Journal* 162, 1 (July 2021), 29. https://doi.org/10.3847/1538-3881/abf8b0

Oz Yilmaz. 2001a. *Seismic Data Analysis* (2 ed ed.). Investigations in Geophysics, Vol. 10. Society Of Exploration Geophysicists. https://doi.org/10.1190/1.9781560801580

Oz. Yilmaz. 2001b. *Seismic Data Analysis* (2 ed ed.). Investigations in Geophysics, Vol. 10. Society Of Exploration Geophysicists. https://doi.org/10.1190/1.9781560801580

Ling Zhang, Yi Xu, Roberto Bugiolacchi, Bin Hu, Cai Liu, Jialong Lai, Zhaofa Zeng, and Zhijun Huo. 2021. Rock abundance and evolution of the shallow stratum on Chang'e-4 landing site unveiled by lunar penetrating radar data. *Earth and Planetary Science Letters* 564 (June 2021), 116912. https://doi.org/10.1016/j.epsl.2021.116912

Wenke Zhao, Emanuele Forte, Federica Fontana, Michele Pipan, and Gang Tian. 2018. GPR imaging and characterization of ancient Roman ruins in the Aquileia Archaeological Park, NE Italy. *Measurement* 113 (Jan. 2018), 161–171. https://doi.org/10.1016/j.measurement.2017.09.004

Haoqiu Zhou, Xuan Feng, Chunyu Ding, Zejun Dong, Cai Liu, and Wenjing Liang. 2022. Heterogeneous Weathering Process of Lunar Regolith Revealed by Polarimetric Attributes Analysis of Chang'E-4 Lunar Penetrating Radar Data Acquired During the Yutu-2 Turnings. *Geophysical Research Letters* 49, 13 (July 2022), e2022GL099207. https://doi.org/10.1029/2022GL099207

Haoqiu Zhou, Xuan Feng, Zejun Dong, Cai Liu, and Wenjing Liang. 2021. Application of Denoising CNN for Noise Suppression and Weak Signal Extraction of Lunar Penetrating Radar Data. *Remote Sensing* 13, 4 (Feb. 2021), 779. https://doi.org/10.3390/rs13040779

Chapter 6

Data Fusion

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

This paper presents a novel approach to merging gated frequency-modulated continuous-wave Mars2020 RIMFAX ground-penetrating radar data using a semi-unsupervised DL algorithm based on Bi-didirectional LSTM. The proposed methodology allows the automatic merging of a varying numbers of dataset at different frequencies, providing more comprehensive subsurface analysis.
The algorithm is trained 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 introduction of a user-defined merging area provides additional control over the final merged profile, but the algorithm can also handle different merging areas. This approach has the potential to be applied to other types of ground-penetrating radar data and can contribute to the development of more efficient and accurate active wavefield geophysical methods.

The supplementary material related to this article is available at: https://library.seg.org/doi/suppl/10.1190/geo2022-0466.1 and at https://figshare.com/s/c6a22136b8005003bced.

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

G. Roncoroni, E. Forte, M. Pipan

Geophysics, Volume 88, March 2023, 1942-2156
doi: doi.org/10.1190/GEO2022-0466.1

.

Key Words

Ground-penetrating radar; *Artificial Intelligence*; *Frequency Domain*

Highlights

- Merging Mars2020 GPR data
- Deep Learning-based merging of Frequency-Modulated Continuous-Wave GPR

Introduction

Based on the spatial decay rate of electromagnetic waves propagating through lossy dielectrics, higher frequency ground-penetrating radar (GPR) 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 Imager for Mars' Subsurface Exploration (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.

The state-of-the-art GPR data fusion approaches integrate data at different frequencies performing semiautomatic 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 2D 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., 2021a). De Coster and Lambot (2018) propose a method and apply it after removing the specific effects of antennas from GPR data, whereas Soldovieri and Orlando (2009) suggest a strategy based on tomographic inversions and then combining partially overlapping frequency bands. Interestingly, data fusion approaches

also have been proposed and tested to combine different data from nondestructive 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 nonphysical amplitude equalization and amplitude balance (De Coster and Lambot, 2018) as well as 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 also can 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), Figure 18), having three different bandwidths (equal in free space to 1050, 750, and 450 MHz, respectively) (Hamran et al. (2020), Table 5).
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 data set.

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 data set. This approach can be classified as a semi-unsupervised DL procedure that exploits a fully 1D approach, i.e., is made trace by trace. To merge data sets characterized by different frequency bands and by different time windows locations 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 three layers of a bidirectional LSTM (Schuster and Paliwal, 1997), with four, two, and one bidirectional layer 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 data set. A graphical representation of this training scheme is provided in Figure 6.1.1. In addition, the user can set an optional input parameter, hereafter referred to as "merging interval". This parameter, $val^k_{[0,1]}$ 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 in which the signal-to-noise ratio is very low.

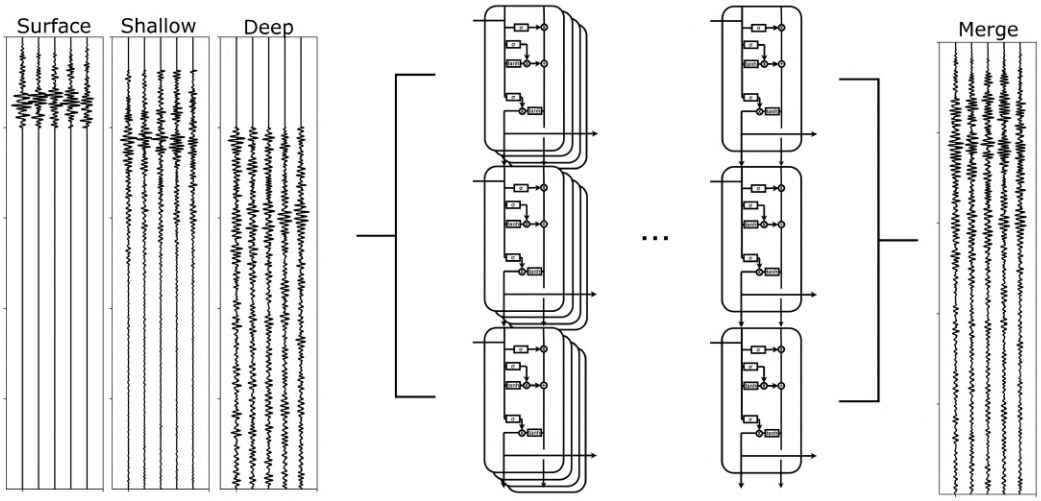

Figure 6.1.1: Graphical representation of the training scheme: on the left side, i.e., input, we have the three input data namely surface, shallow, and deep, respectively, with the merging interval marked with the solid lines. The green lines represent val^k_0 for each input, whereas the red lines represent val1k for each input (Equation 6.3). NN represents the LSTM neurons into a bidirectional wrapper, and in output, we find the prediction i.e., the merged data.

The loss function is expressed by the equation:

$$loss = \sum_{k=1}^{n} \lambda_k loos_k \tag{6.1}$$

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

$$loss_k = \|target_k \; - \; prediction\|_2^2 \tag{6.2}$$

where the merged prediction is compared with the $k^{th}_{[1,2]}$, whereas $\lambda_k$is defined as a

piece-wise function:

$$\lambda_k(t) = \begin{cases} 0 & \text{if } t < val_0^k, \\ 1 & \text{if } val_0^k < t < val_1^k, \\ 0 & \text{if } t > val_1^k \end{cases} \tag{6.3}$$

where val_0^n and val_1^n are the lower and the upper boundary of the k^{th} data, respectively, as shown in Figure 6.1.1.

First, we tested such a method on synthetic data simulating the expected subsurface of Mars (Figure 6.1.2).

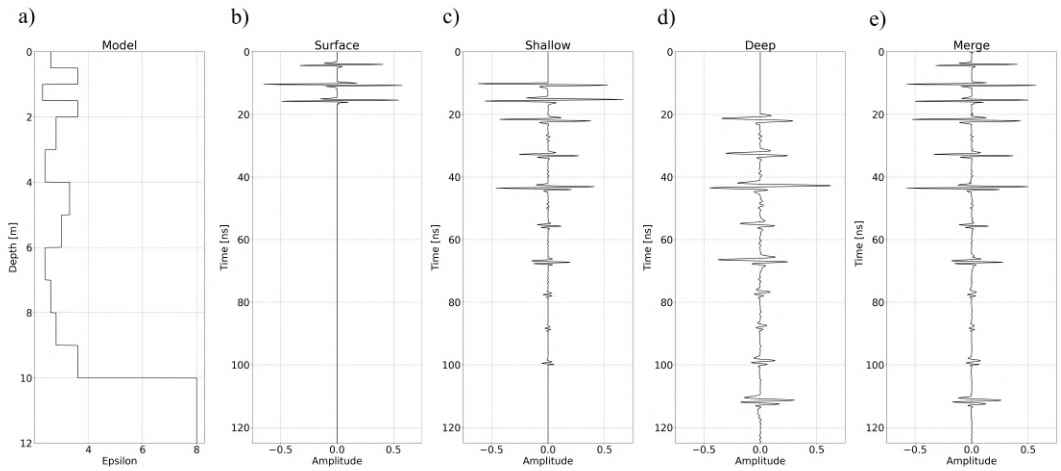

Figure 6.1.2: Synthetic simulation and merging obtained on a random model. (a) The model and the (b–d) surface, shallow, and deep simulation modes, respectively. We applied the same time windows imposed by the RIMFAX Radar system, and we merged the data with the proposed methodology obtaining the results in (e). See text for further details.

We defined a random 1-D model, as shown in Figure 6.1.2a, defining two main layers, i.e., 10 m of sedimentary cover with $\epsilon = 2.8$ and a bedrock with $\epsilon = 7.8$, as proposed by Hamran et al. (2020). 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 four layers in the first $2m$ 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 a source, we used a Ricker wavelet with the central frequencies reported in Hamran et al. (2020) for surface, shallow, and deep acquisition modes. To simulate our target acquisition, we deleted data outside of the receiver window for each specific frequency, as summarized by Hamran et al. (2020).

The merged results (Figure 6.1.2e) are very accurate in time as compared with the surface, shallow, and deep traces (Figure 6.1.2b - d, respectively) and well balanced in amplitude. For instance, the reflection at approximately $45ns$ 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 Discussion

After selecting data from the entire RIMFAX data set (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. We chose to perform the training on only the 10% of the data set only for computational efficiency: in theory, it would be possible to use the whole data set, as we do not overfit the problem with this approach.

Because the entire data set is quite long and contains a total of 23,199 traces (after removing repeated ones as previously described; supplemental Figures S1, S2, and S3), we analyze two specific portions characterized by slightly different features and structures, namely: sub-horizontal layers with lenses (Figure 6.1.3) and monocline dipping reflectors (Figure 6.1.4). Merged results are in both cases quite good (for the amplitude spectra, see supplemental Figures S4, S5, S6, S7, and S8). In particular, no artifacts are introduced, and the continuity of reflectors is preserved. Moreover, merged data in Figure 6.1.3 clearly show the vertical sequence of layers, which is difficult to understand by separately analyzing the three RIMFAX acquisition modes (for the interpretation, also see supplemental Figures S11 and S12). In fact, from approximately 20 and 55 ns, two separated sub-parallel reflectors are imaged, as well as two superimposed lenses. In Figure 6.1.4, the dipping reflectors are perfectly reconstructed with no vertical gaps from only a few nanoseconds down to 250 ns.

Abrupt lateral amplitude variations in the merged profiles are not introduced by the application of the fusion algorithm because they also are present (and sometimes even more apparent) in 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)).

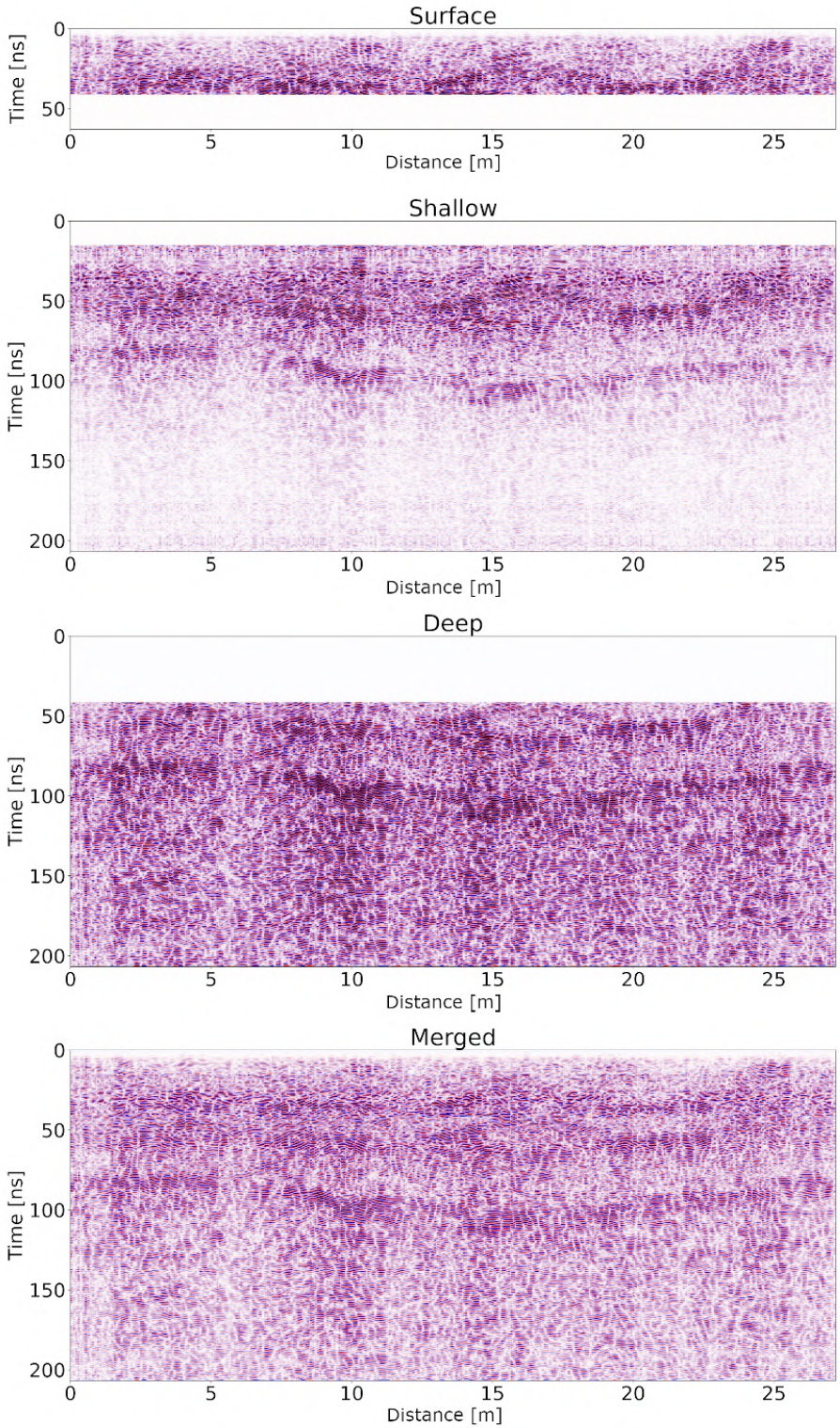

Figure 6.1.3: *SOL 113–116* RIMFAX data. (a–d) Surface, shallow, deep, and merged data, respectively.

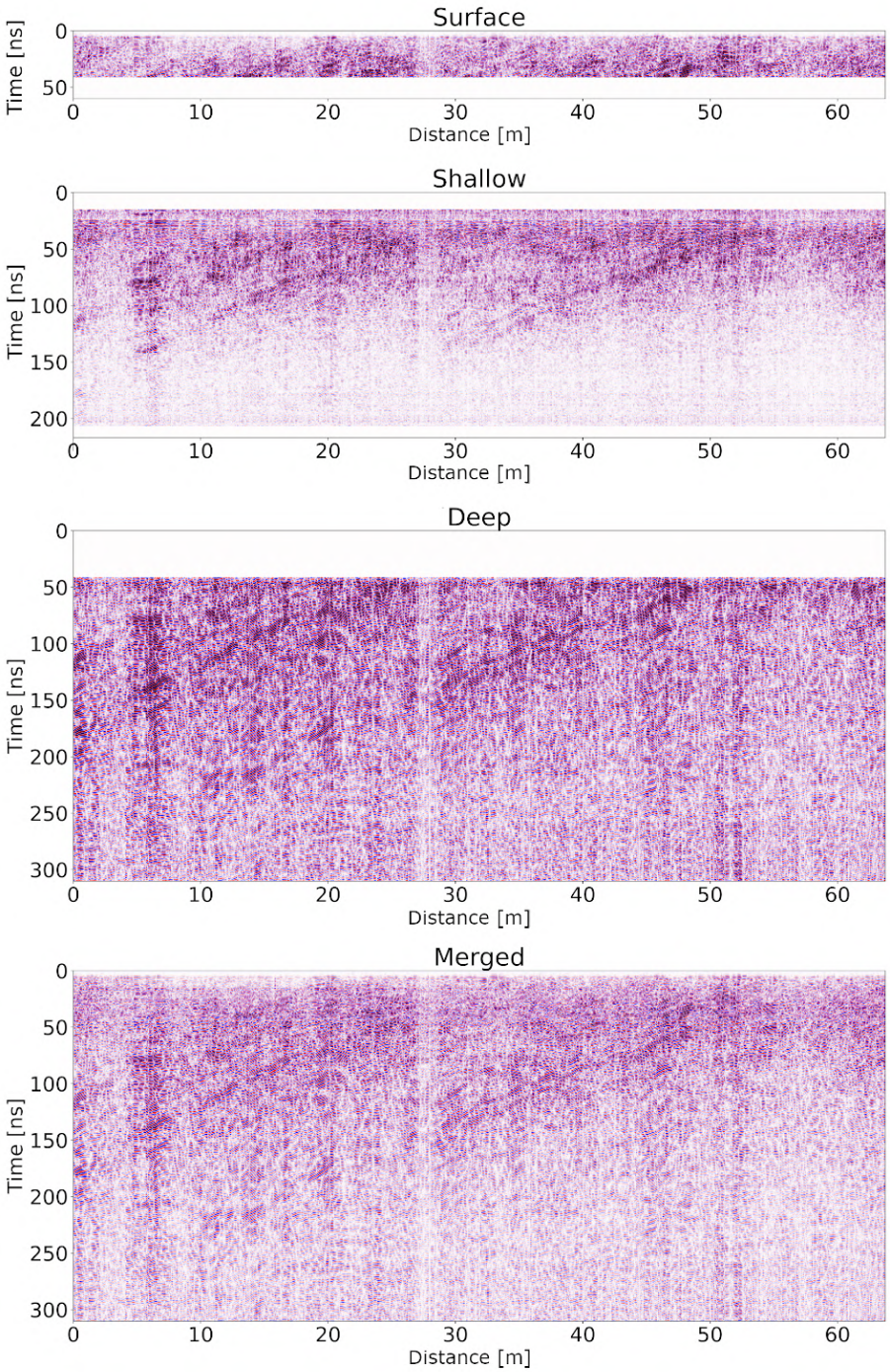

Figure 6.1.4: *SOL 200–203* RIMFAX data. (a–d) Surface, shallow, deep, and merged data, respectively.

However, lateral phase continuity is high and no trace gaps such as the previously described cut-and-paste effects are introduced by the merging algorithm (supplemental

Figures S9 and S10); 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 6.1.5, 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 of the traces (Figure 6.1.5c1 and 6.1.5d1) in which an abrupt frequency variation is present. This does not occur with the proposed procedure (Figure 6.1.5c2 and 6.1.5d2), obtaining a smoothed variation of the spectral content. Moreover, some horizons (h labels) are clearer and show higher lateral continuity.

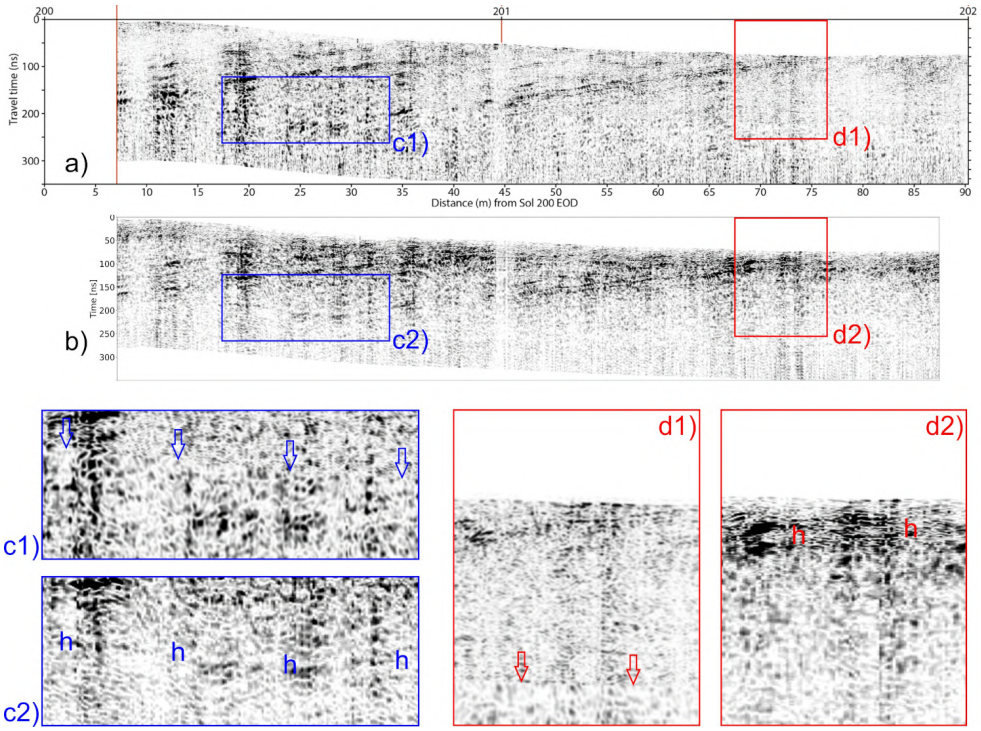

Figure 6.1.5: Comparison between data published in Hamran et al. (2022), (b) Figure 4b in Hamran et al. (2022) and the results of the proposed algorithm. (c1 and d1) Two close up of Hamran et al. (2022) figure, whereas (c2 and d2) the same portion of the data obtained with the proposed merging strategy. Vertical arrows highlight the cut-and-paste effect apparent in Hamran et al. (2022), whereas "h" labels mark horizons made cleared on data merged with the new proposed procedure.

Regarding the computation times over the entire data set, a matrix with 23,000 traces, 5000 time samples for the three modes, and training time with 10% of the data, as described previously, takes only $970s$ on a laptop with Intel(R) Core (TM) $i7-10875H$, $32GBRAM$ and an Nvidia GeForce RTX 2070 Super with $8GB$. On the same laptop, the prediction time for the entire data set takes $24s$

Conclusion

We succeeded in creating a DL-based methodology that can merge RIMFAX multifrequency GPR data that is robust, fast, and does not require particular data preprocessing and conditioning.
The introduction of specific user-defined merging windows allows for making 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 because they do not suffer from the possible introduction of amplitude/phase gaps or local/coherent artifacts (Figures S13, S14, and S15). 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 similar to in the case of the RIMFAX data set but can be applied to merge any type of data set with multiple spectral components and recording windows.

Acknowledgement

This research was partially supported by 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 Halliburton Landmark through their academic grants. We further thank three anonymous reviewers for their fruitful comments and suggestions.

Data and Materials Availability

Codes related to this article can be found at: https://github.com/Giacomo-Roncoroni/merging_RIMFAX.

6.2 Deep Learning based multi-frequency GPR data merging

This paper presents an approach to merge GPR data using artificial intelligence. We provide a generalization on what presented in Roncoroni et al. (2023), with a efficient and effective way to analyze subsurface structures of data acquired using different antennas.

With this method, we can obtain more accurate and comprehensive information about subsurface structures. The proposed approach is based on a new training procedure based on unsupervised learning, using a Bi-Directional LSTM network. The results of this study demonstrate the potential of LSTM networks in data fusion, providing a starting point for future work in seismic, i.e. merge of DAS and active seismic data.

Deep Learning based multi-frequency GPR data merging

G. Roncoroni, E. Forte, I. Santin, M. Pipan

Geophysics, Just-Accepted Articles, 0016-8033
doi: doi.org/10.1190/geo2023-0215.1

Key Words
Ground-penetrating radar; *Artificial Intelligence*

Highlights

- Semi-supervised Deep Learning signal merging
- New multi frequency GPR datasets merging procedure
- Wide applicability without restrictions on number and range of frequencies demonstrated after extensive synthetic and field data tests
- Wide applicability and robustness to combine A-, B-, C-scans datasets is shown

Introduction

Ground penetrating radar (GPR) exploits electromagnetic (EM) waves to investigate the shallow subsurface. It is a versatile technique that can be adapted to different applications spanning from geological, engineering, environmental, archaeological, glaciological, and forensic investigations, among others (Lombardi et al., 2022). The achievable resolution is a function of the physical characteristics of subsurface materials, being inversely proportional to the EM velocity and unavoidably decreases for increasing depths, as for all the geophysical techniques. Resolution can be improved by increasing the source frequency, i.e. the frequency content of the EM wavelet used for sounding. However, this leads to a reduced depth of investigation due to the stronger attenuation of high frequency signals. A trade-off between the desired resolution and depth of investigation is therefore mandatory in all the GPR systems based on a single central frequency. Moreover, by using a single central frequency signal, as is usually done, imaging performance is limited, in particular when the subsurface geometry is complex and the targets have different dimensions (Lu et al., 2020). Modern multi-channel (or multi-array) systems can provide multi-frequency datasets

but pose severe logistical constraints, especially when they operate on steep or rough terrains. In addition, issues arise in interpretation due to the large amount of data and their intrinsically different nature when collected with different central-frequencies antennas.

In order to combine together at least two signals having different central frequencies, several algorithms and procedures have been proposed. Dougherty et al. (1994) were the first who suggested to exploit such a possibility. Since that time, several other methods have been implemented and tested. The final objective is referred as multi-frequency compositing (e.g. Endres et al. (2004)), multi-frequency fusion (Xu et al., 2019; Zhao et al., 2021b), data fusion (e.g. Lu et al. (2020)) or merging (e.g. Roncoroni et al. (2023)). A comprehensive review of the state-of-the-art of the different proposed approaches and their peculiarities can be found in Alemdağ et al. (2022).

A detailed classification of the proposed methods to combine together data with different frequency components is out of the scope of this study, but, in general, there are two different strategies either based on 1-D fusion (i.e. considering each A-scan - or portions of it - as separated entities), or based on 2-D fusion (i.e. combining B-scans or portions of them as "images", see e.g. Lu et al. (2020)). The latter are often also referred to as image fusion (Zhao et al., 2021a).Here, for the first time, we propose a multi-frequency merging algorithm based on semi-supervised Bi-Directional Long-Short Term Memory to automatically merge varying numbers of data sets with different central frequencies. Training and inference of the Neural Network (NN) are made on the same dataset to be merged.

In Roncoroni et al. (2023) merging of separated windows (i.e. partially overlapping portions of B-scans) of the peculiar RIMFAX Mars radar dataset is presented; as an improvement of the previous approach, we here propose a general procedure that can be applied to merge any multi-frequency data without constraints in terms of time window length and overlap, number and range of frequencies to be combined, sampling interval and number of samples.The performance of the proposed algorithm is here discussed considering both synthetic and field A-, B-, C-scans (i.e. 1-D, 2-D and 3-D data, respectively). We demonstrate that the merged results are robust and do not require any amplitude balance or other specific processing steps. The bandwidth is expanded and the resolution is increased, while the phases are not distorted and no noise or spikes are introduced. Such a strategy may prevent possible misinterpretations,

making easier data interpretation from large to small scales, in turn providing a more comprehensive understanding of the shallow subsurface. Merged results can be further processed, used for automated advanced interpretation techniques or attribute-based analyses, as well as exploited for data inversion.

Method

The proposed approach is based on a new training procedure based on unsupervised learning: this allows us to use the same dataset for both training and inference: the kernel of the methodology is a Bi-Directional Long Short-Term Memory (LSTM, see e.g. Schuster and Paliwal (1997)). The choice of this type of neuron is crucial for our goals. Bi-Directional LSTM, represented in a simplified form in Figure 6.2.1, is a recurrent neuron with long memory that takes, as input, both the time-dependent signal and its reversal in time versions. This neuron is pivotal for the results and performances of the methodology. Let's consider a simple signal, i.e. a Ricker wavelet, at $t = 10ns$ discretized with sampling interval dt

$$W(t) \;=\; [1 - \frac{1}{2}(\omega(t - dt))^2]e^{-\frac{1}{4}(\omega(t-dt))^2} \tag{6.4}$$

With a normal LSTM neuron, we can get information from everything that happened before (t), but we cannot get any information on what is happening at $t > t + \Delta t$. Therefore, we need to consider the signal from both sides with respect to each sample to allow the Neural Network (NN) to correctly understand the waveform, its changes, and how it moves over time.

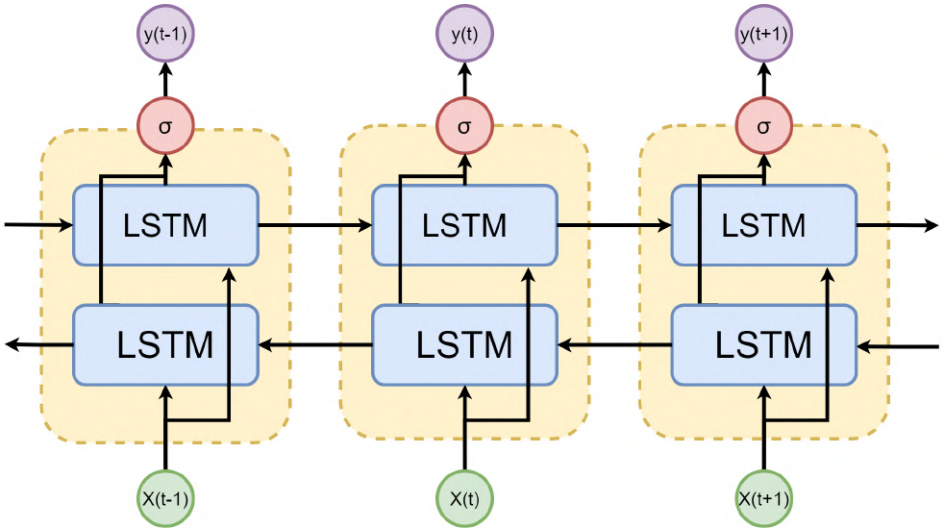

Figure 6.2.1: Unwrapped neuron of a Bi-directional LSTM. We can see both the recurrent nature of the neurons and the union between normal and reversed in time version with a sigmoid function.

After the definition of the neuron characteristics, we have to define a target for the training. Since we do not have a priori information on what we want to get and a synthetic simulation of the merged target could be tricky and not easy to define, we exploit a different strategy that mixes a supervised and an unsupervised methodology, we can define this approach as hybrid learning.

Before merging, we need to apply some essential pre-processing: one important step is the data amplitude recovery, usually exploiting exponential or more complex functions. This phase is crucial to correctly weight events when computing Mean Squared Errors (MSE). Another crucial step is to apply a normalization of the data, thus obtaining comparable amplitudes. We tried different approaches and it turned out that the optimum solution was to scale the integral of the amplitude spectrum over the frequencies of interest to the same value, e.g. equal to 1. Another possibility could be just to normalize the whole dataset to 1, but this would put excessive weight on the low frequency components.

The proposed methodology is based on the training and inference of a NN, trained on the same dataset to be merged. Specific limits and threshold can be inserted by the operator, e.g. the definition of weighting window for each frequency - as discussed in Roncoroni et al. (2023)-, but the procedure can also be applied to merge entire traces (i.e. A-scans) collected with different central frequency antennas.

In order to merge data sets at different frequency bands, we used a procedure implemented for NN training, namely the Gradient Descent optimizations algorithms (Kingma and Ba (2014)), and the power given by a custom transformation based on only a few neuron weights.The base of this approach is that merging should be performed by a few layers of a Bi-Directional LSTM (Schuster and Paliwal (1997)), 3 in this examples, with 4, 2 and 1 Bi-Directional layers and a single LSTM neuron for the output, aiming at minimizing the custom loss function with a single prediction from all input data. Exploiting only a few parameters for all data – as a result of using a very compact NN - we implemented a methodology that is robust to outliers and noise. A representation of this training scheme is depicted in Figure 6.2.2. Furthermore, the user can provide an input parameter, which we will refer to as the merging interval. This parameter (val_1 and val_2 in the following equations), is introduced into the algorithm directly in the loss function and should allow the NN to disregard areas where no useful input information is provided. If we want to keep information inside all input signals, we can just set the full trace as the merging interval, obtaining in output a broader

frequency spectrum signal:

$$loss \;=\; \lambda_1 loss_1 + \lambda_2 loss_2 \tag{6.5}$$

Where $loss_n$ is defined as:

$$loss_n \;=\; \|input_{frequency_n} \;-\; prediction\|_2^2 \tag{6.6}$$

while $\lambda_n$is defined as a piece-wise function:

$$\lambda_1(t) \;=\; \begin{cases} 1 & \text{if } t_1 < val_1, \\ 0 & \text{if } t_0 \geq val_1 \end{cases} \tag{6.7}$$

$$\lambda_2(t) \;=\; \begin{cases} 1 & \text{if } t_1 \geq val_2, \\ 0 & \text{if } t_0 < val_2 \end{cases}$$

With λ_1 and λ_2 referring to high and low frequency data, respectively. The introduction of val_1 and val_2, if $val_1 > val_2$, leads to the creation of 3 different zones, referring to Figure 6.2.2, within the green box, yellow box, and overlapping part; in detail:

- High frequency zone (HFZ), where $t < val_1$, green box.
- Merging area (MA), where $val_2 < t < val_1$, overlapping area.
- Low frequency zone (LFZ), where $t > val_2$, yellow box.

In the HFZ the loss function is evaluated only on the highest frequency data. In MA both data are weighted with the same value, while in LFZ only low frequency data are taken into account. After defining the two parameters val_1, val_2, we train the NN with the same data we want to merge. The loss values as a function of training epochs is the measure of the convergence of the algorithm. So we have to consider these values to decide whether we have reached a stable model. After few iterations (see discussion paragraph), a minimum in the loss function is reached and the same trained NN can be applied to infer the merged data.

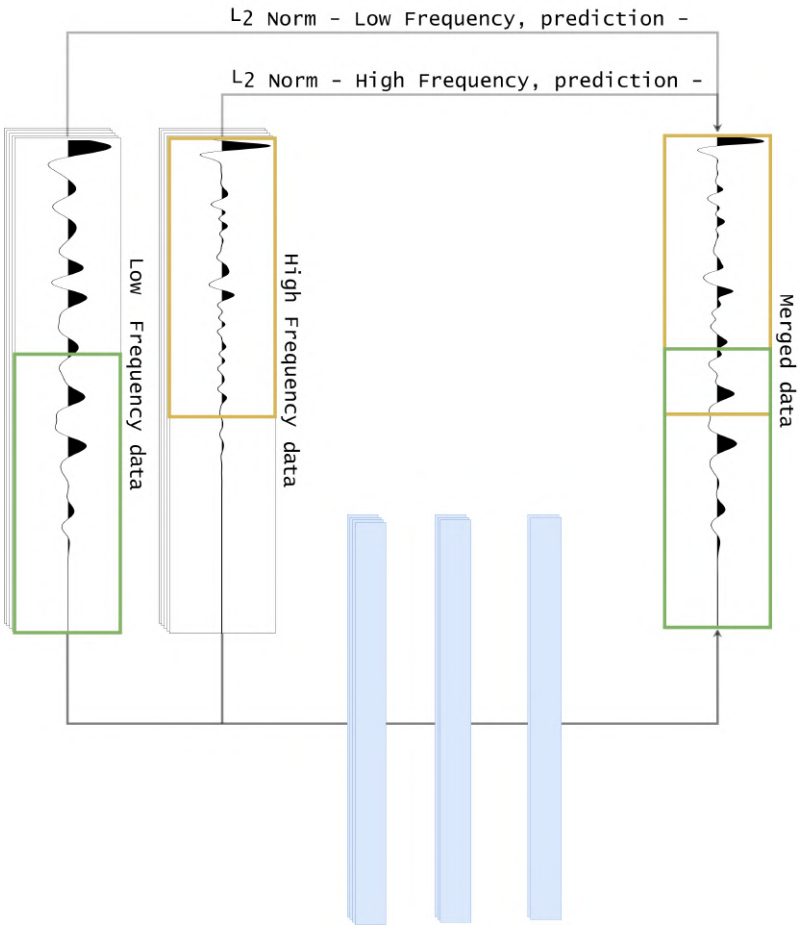

Figure 6.2.2: Neural Network training workflow. Starting from n inputs (just two are here depicted) we retrieve a single mixed output. The network is trained to minimize the L2 norm between each input and the output. A further mixing area constrain can be given to enhance higher frequencies in the shallower data.

Results

Test on synthetic A-scan As each data set is different, depending on the specific subsurface conditions being investigated, we may want different features to be enhanced and we let this specific option to the interpreter. A user-defined area is the kernel of the adopted strategy. Let us consider a double central frequency GPR signal, i.e. acquired with two different antennas, e.g. $400MHz$ and $200MHz$. We may want to retrieve shallow information from the entire bandwidth and we may want to keep only the lower frequency components in the deepest section of the record; we therefore can define this as an information merging from now on.

If we are in a low loss medium, e.g. ice or dry sand, a higher central frequency can carry the most information, but adding lower frequency components can extend the bandwidth and improve resolution and information content. In Figure 6.2.3 we pointed

out a typical simplified situation where we can see both HFZ (grey, left column), LFZ (blue, right column), and the merging area (grey-blue in the central column). The synthetic dataset was generated by convolving a maximum phase Ricker wavelet with just 3 reflection coefficients and we set the training so that in the HFZ we just want to get the $400MHz$ component (Figure 6.2.3 , column A), in the LFZ just the $200MHz$ one (Figure 6.2.3 , column C), and in the merging area both of them (Figure 6.2.3 , column B). As we can see, results are very good in terms of bandwidth widening of the merged dataset, and the residuals are very low in the HFZ and LFZ for higher and lower frequency components, respectively. As we can see, the merging interval can retrieve a broader frequency version of the input wavelet, as apparent in the amplitude spectra (Figure 6.2.3b).

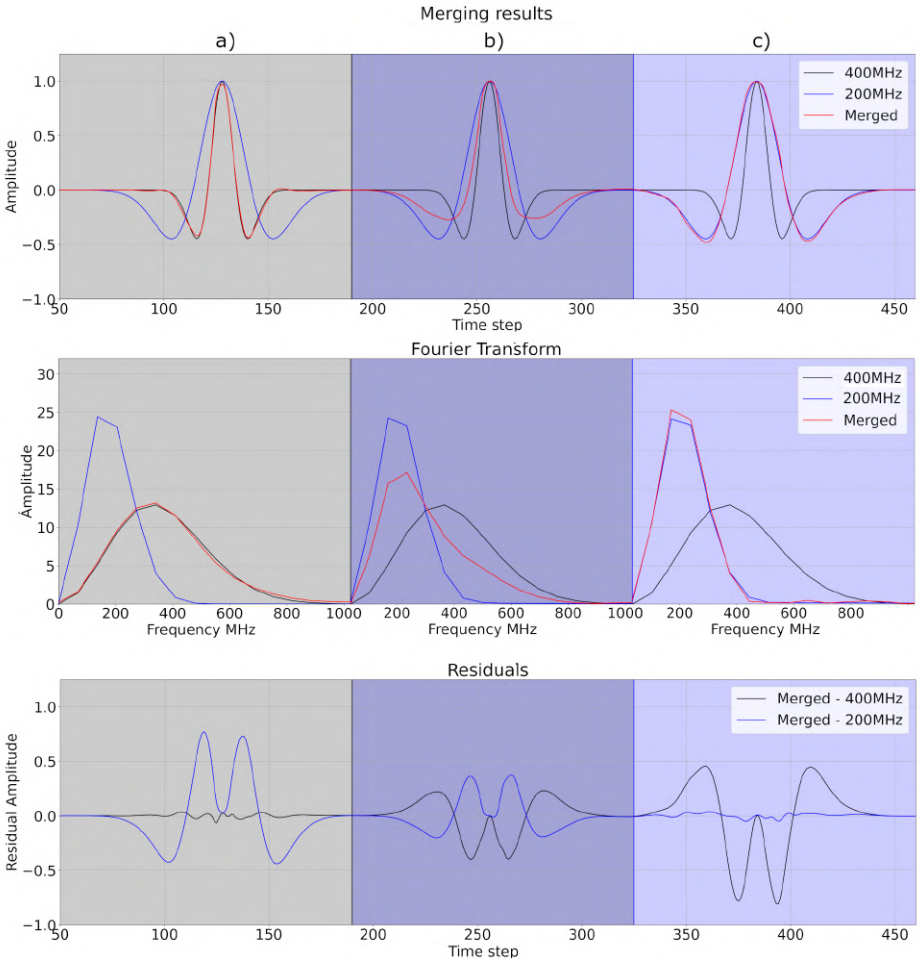

Figure 6.2.3: Frequency merging on a simple 3-reflectors synthetic model. The figure is divided into three columns: HFZ (column a), depicted in grey, LFZ (column c) in blue, and the merging area (column b) is in grey-blue. In the first raw we can see input signals and the merged result, in the second one the amplitude-frequency spectrum of each area and in the last raw the residuals.

Indeed, the residuals of the merged frequencies (Figure 6.2.3c, central column) are almost perfectly symmetric and are null at the point at the peak of the wavelets, thus demonstrating the effective combination of the two frequency components, without any phase shift.
For the methodology the phase consistency between input and output data is a crucial step. As we can see from the residuals, the phase is perfectly preserved, since residuals have zero errors in both the wavelets at the maximum peak.

Test on Synthetic C-scan

As a more complex and realistic synthetic application, we tested the methodology on the $3-D$ synthetic *Herten dataset* (Koyan and Tronicke, 2020a), based on a publicly available hydrofacies data set, which has been derived from sedimentological outcrop mapping, direct sampling, and geostatistical simulations of an aquifer-analogue at the Herten gravel pit (SW-Germany). The model is very realistic as it includes several different sedimentary features, including accretionary structures, cut-and-fill sequences, well-sorted sand-gravel bodies, and matrix-supported gravels. The dataset originates from six parallel digitized outcrop images, which are interpreted in terms of lithological facies resulting in six $16m$ long rasterized $2-D$ facies sections.
These sections cover a depth range of $7m$ and with constant cell dimension equal to $0.05m$. Laboratory measurements of facies-specific hydrogeological properties including porosity and hydraulic conductivity lead to a subdivision of the mapped lithological facies into 10 different units, termed hydrofacies. The authors use these $2-D$ hydrofacies sections to perform geostatistical modeling resulting in a $3-D$ subsurface model discretization of $0.05m$ in all spatial dimensions, from which the electrical parameters are derived. Such a model was used as input to model $3-D$ GPR data using gprMax (Warren et al., 2016, 2019). The *Herten* subsurface model was used to simulate a GPR Common Offset (CO) dataset with a $100MHz$ central frequency source (Koyan and Tronicke, 2020b) then integrated with two other datasets obtained on the same subsurface model, but with central frequencies of 50 and $200MHz$ (Koyan and Tronicke, 2020a). The three synthetic GPR datasets are perfect for testing the proposed merging procedure because, from one side, the data image different geometries and include diffractions and interference phenomena, from the other the model parameter and geometries are perfectly known. All the details about model discretization and parameters used to derive the 50, 100 and $200MHz$ datasets are reported in Koyan and Tronicke (2020a,b).

In Figure 6.2.4, wiggle-variable area plots of just a few input traces ($A-scans$) and

merged data are reported. We highlight not only the excellent phase preservation, but also the introduction of low frequency components in the shallower part of the merged plot. For instance, if we compare the 3 wavelets of the event between 25 and 50ns we can clearly see the low- high- frequency mixing in the merged data (Figure 6.2.4c).

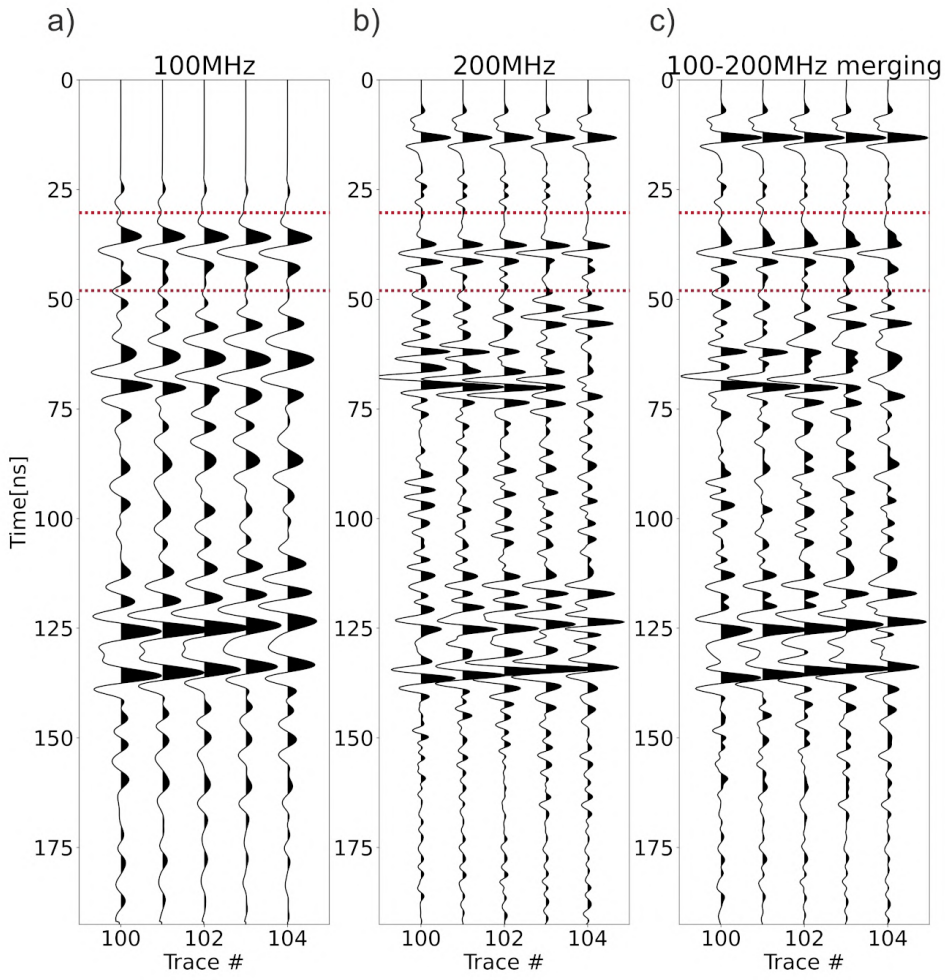

Figure 6.2.4: Merging test on 5 exemplary traces. Wiggle plots of the 100 MHz (a) and 200 MHz (b) inputs, and the obtained merging (c). Red lines mark an area where the low- high- frequency mixing is apparent.

In this test we did not use the lowest frequency data (i.e. 50MHz) since their information is very limited compared to the higher frequency ones (i.e. 100 and 200MHz), with a significantly lower resolution and without any addition of information in the deeper part of the record.

The merged profile (i.e. B-scan), (Figure 6.2.5c) shows not only enhanced overall resolution but also continuity of reflectors (see e.g. yellow arrows in the figure). The proposed method does not introduce neither coherent nor random noise and "cut and paste" effects (Roncoroni et al., 2023). Furthermore, we have a general broadening of the spectrum that can help inversion processes.

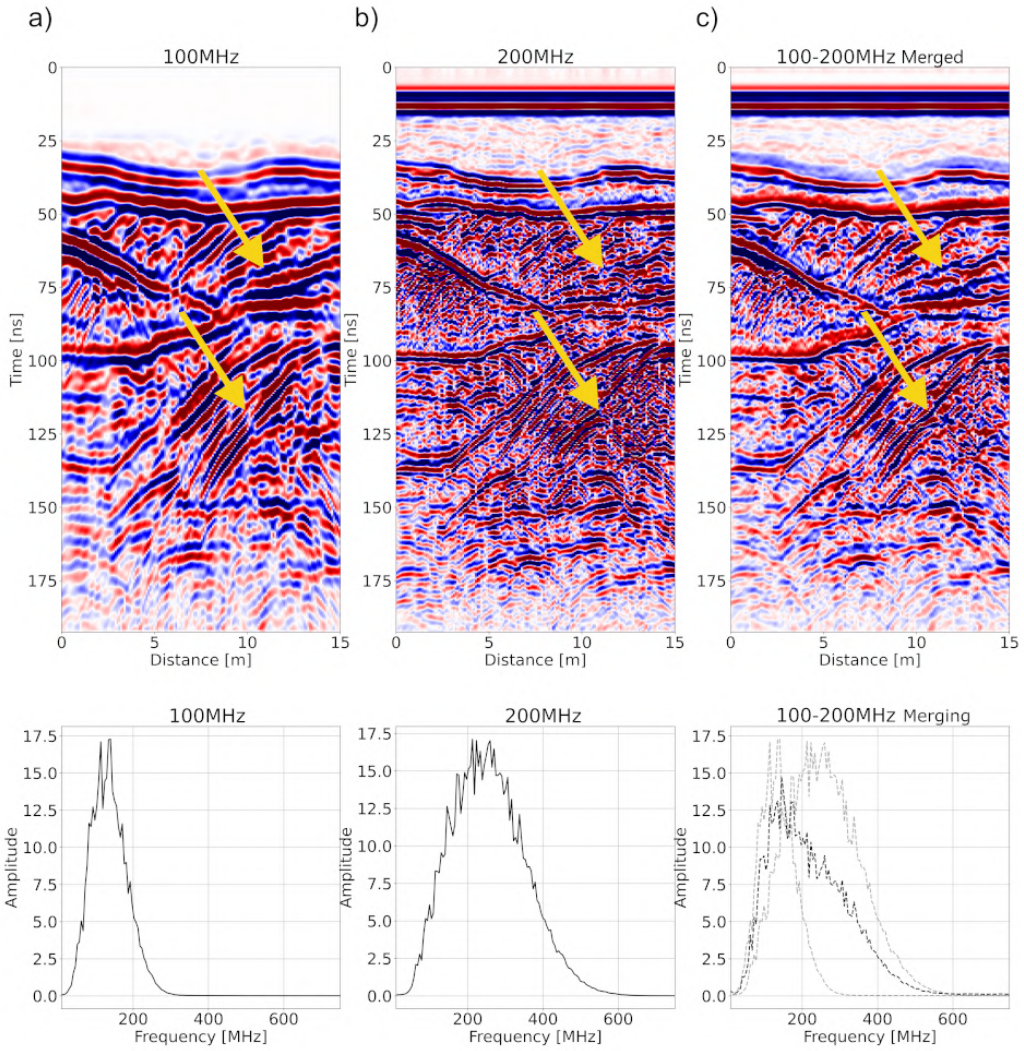

Figure 6.2.5: Merging test on a double-central frequency synthetic profile extracted from the 3-D Herten dataset. 100 MHz profile (a); 200 MHz profile (b); merged profile (c), and their respective amplitude spectra. In the merged spectral panel, the 100 and 200 MHz spectra are reported in light grey for a better visual comparison with the merged one.

We applied the workflow to the whole *3-D Herten dataset* (Figure 6.2.6) to further test the effectiveness and stability of the procedure.

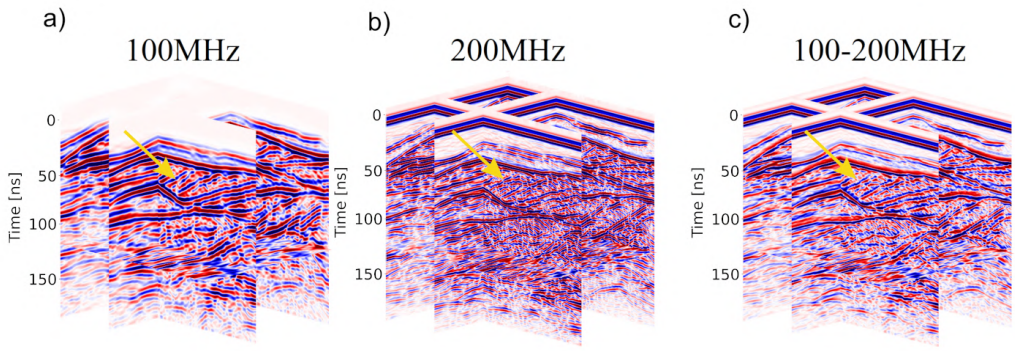

Figure 6.2.6: Merging of the whole 3-D 100 and 200 MHz Herten datasets. 100 MHz (a); 200 MHz (b); a100-200 MHz merged data (c).

The test proves the stability and robustness of the method and its effectiveness in preserving not only the signal amplitude contrasts, but also its phases. As we can see in the 3-D plot (Figure 6.2.6c) we obtained a good lateral continuity and reflectors matching at all the crossing points. Also, the imaging of relatively high dipping reflectors in the merged dataset is overall better than the one on the single-central frequency volumes (e.g. yellow arrow in Figure 6.2.6).

Test on Field B-scan

The proposed merging method was finally applied to field GPR data. During the *XXXVIII Italian Expedition to Antarctica* (austral summer 2022-2023), a multi-frequency GPR dataset was acquired using a ProEx GPR instrument (Malå Geoscience) first connected with 250MHz, then 500MHz and finally 800MHz shielded antennas. The dataset is therefore composed of three GPR profiles collected on Boulder Clay glacier (Victoria Land - East Antarctica) along the same path with the three different antenna pairs. Details about the acquisition zone are reported in Azzaro et al. (2022). The profiles are collected along a straight line with an approximate E-W direction, perpendicular to the glacier flow and with a 0.10 m trace interval. This dataset was chosen to test the merging procedure because the different frequencies have a completely different penetration depth and highlight different glacial features including diffractions and reflections with a large dip range.

The typical characteristics of GPR datasets, in which high central frequency antennas ensure high resolution but have a limited penetration depth while the opposite is true

for low frequency ones, is well highlighted by Figure 6.2.7, where the 800MHz GPR profile (Figure 6.2.7a) provides clear identification of reflectors in the first $60-70ns$. In particular, it images a strong continuous horizon very close to the surface, which can be associated to the transition between firn and ice. In the 500MHz data (Figure 6.2.7b) some internal glacial structures can be recognized, as well as the ice-bedrock high amplitude reflection, imaged with higher continuity and toward its maximum depth (i.e. about $400ns$) in 250MHz data (Figure 6.2.7c).

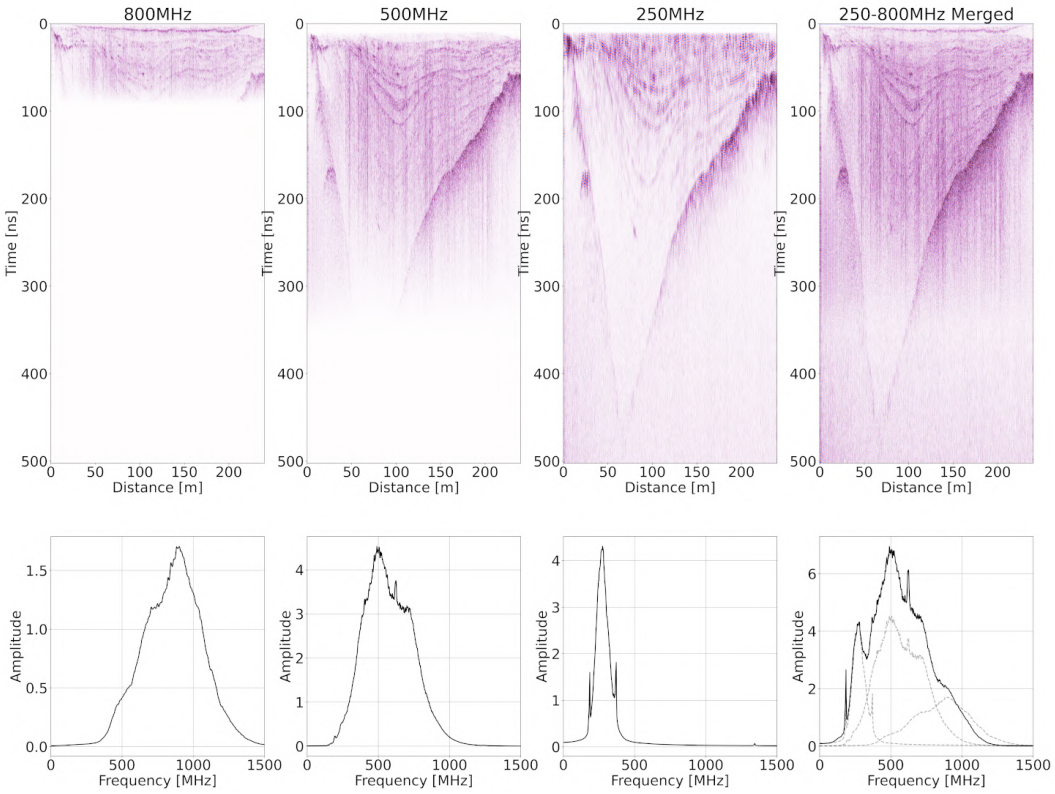

Figure 6.2.7: Merging test on a field dataset: (a) 800 MHz profile; (b) 500 MHz profile; (c) 250 MHz profile; (d) merged profile, and their respective amplitude spectra. In the merged spectral panel, the 250, 500 and 800 MHz spectra are reported in light grey for a better visual comparison with the merged one.

The corresponding amplitude spectra (bottom row of Figure 6.2.7) show the widened frequency band of the merged data. The comparison between single-central frequency profiles (Figure 6.2.7a,b,c) and the merged data (Figure 6.2.7d) highlights the overall improvement of resolution while ensuring a signal penetration depth which is comparable with the 250MHz one. In fact, all features characterizing each GPR profile can be recognized in the merged data, from the firm-ice transition to all internal high dipping layers and the ice-bedrock interface.

The 250MHz antenna allows for optimal imaging of the bedrock, but the imaging offered by the merged data extends simultaneously from surface to deep targets ensuring for all of them optimal detectability and a very good resolution. In addition, low frequencies favour the identification of several dipping layers especially in the first 100m of the profile, which were not clearly resolved from the 800 and 500MHz antennas.

Discussion

Most GPR data merging studies focus on synthetic datasets based on models that are very simple in terms of geometry and physical properties, i.e. planar reflectors with constant dip and constant electromagnetic properties of each layer (Xiao and Liu, 2016; Bi et al., 2020; Alemdağ et al., 2022). In some cases, random noise or spikes are added after forward modeling (e.g. Alemdağ et al. (2022)). On the contrary, we tested our procedure on one of the most realistic synthetic 3-D GPR dataset at present available, and derived from a real outcrop through mapping, direct sampling and geostatistical simulations (Koyan and Tronicke, 2020a).As far as real data is concerned, a typical problem is not only that low central-frequency GPR data yields lower resolution, but the shallower portion is masked by the air and ground waves arrivals (Jol, 2009)
Figure 6.2.8 shows the same merged data as in Figure 6.2.7: the shallow unconformity (black arrows) cutting the dipping layers (red arrows) which onlap on it, can be easily identified especially in the first 100m of the profile.

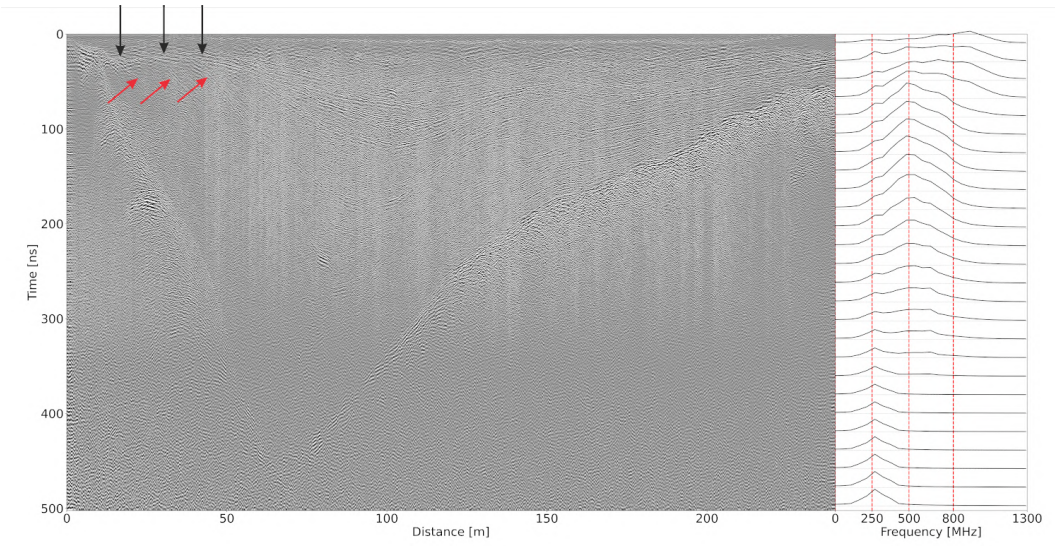

Figure 6.2.8: Merged data profile and amplitude spectra of the data in Figure 6.2.7d calculated with 20ns long sliding time windows. The profile has a gradual frequency decay with depth down to about 300ns. At greater depths the frequency content is more stable showing a dominant frequency of about 250MHz. Red and black arrows mark specific reflectors described in the text.

At least 10 distinct internal layers, including continuous and discontinuous horizons, can also be displayed and recognized, whereas in a single central-frequency acquisition could not be possible to image such internal features. The amplitude spectra on the right of Figure 6.2.8 demonstrate the spectral shift with depth due to the quick decay of high frequencies components, but also the effectiveness of merging which increased the resolution especially in the shallow portion of the profile, where the bandwidth is remarkably wide and the peak amplitude is greater than 800MHz. In the bottom part the amplitude spectra are more constant and are characterized by a dominant frequency of about 250MHz. The stability of the method was tested by calculating some signal attributes and by migrating the merged data. The cosine of the instantaneous phase attribute (Figure 6.2.9) allows us to better evaluate the continuity of reflectors and it is therefore optimal to assess the effectiveness and robustness of the merging method. In Figure 6.2.9, the ice-bedrock reflection can be traced clearly and continuously even where it has greater inclination (cyan arrows in Figure 6.2.9) as well as all the internal layers of the glacier. This example highlights the performance of the proposed method in preserving the continuity of the events without creating distortions or jumps related to the juxtaposition of different frequency bands.

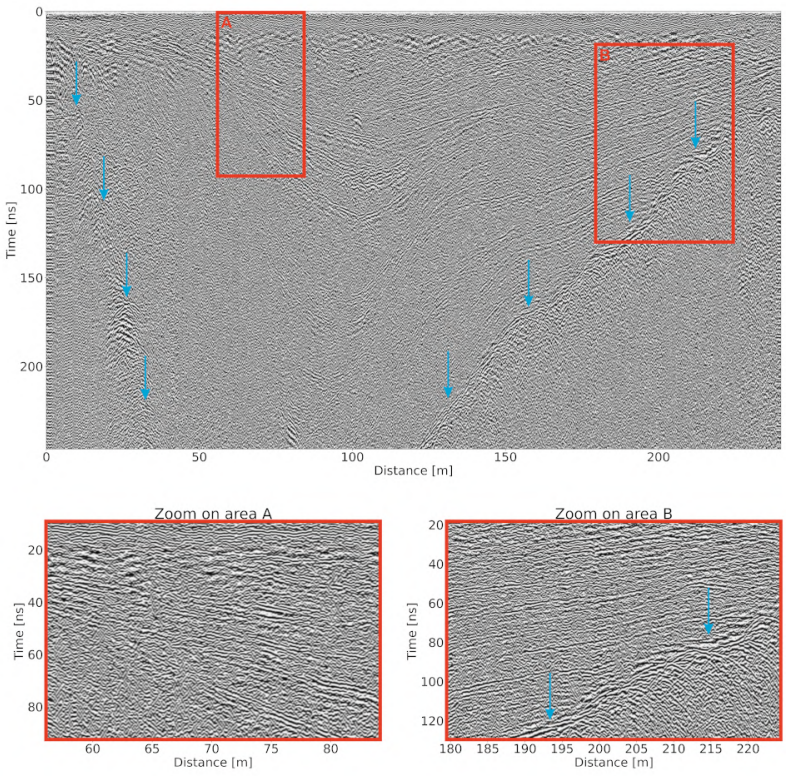

Figure 6.2.9: Cosine of the instantaneous phase calculated on the merged data of Figure 6.2.7d and 6.2.8. Time is limited to 250ns with two close-ups on areas A and B, respectively.

Another test of the validity and efficiency of the merging methodology is the application of a migration algorithm to the merged data (Figure 6.2.10). Migration focuses the diffractions' energy to the apex and places the reflectors at their correct position. All migration algorithms are indeed very sensitive to noise and abrupt spectral (both in frequency and phase) variations of the signal (see e.g. Lehmann and Green (2000)). After applying the topographic correction with a constant velocity equal to $0.17m\ ns^{-1}$, we applied a *Stolt F-K migration* algorithm with a constant velocity equal to $0.17m\ ns^{-1}$ which corresponds to a 3.2 relative electrical permittivity, typical of pure ice. This is surely an oversimplified velocity field but it does not represent a crucial parameter for our test. The merged depth-migrated section, (Figure 6.2.10), provides the most realistic imaging of the internal structure of the glacier, allowing us to better appreciate the correlation between the morphology of the bedrock and the shape of the internal layers, onlapping on the almost horizontal unconformity that cuts the deeper glacial strata. No apparent over- or under-migration effects are present, while the lateral continuity of the reflectors is always preserved thus demonstrating once again the stability, robustness and validity of the merging procedure. Merged data can be further processed and analyzed, considering that no artifacts, outliers or coherent noise components are added by the data merging process.

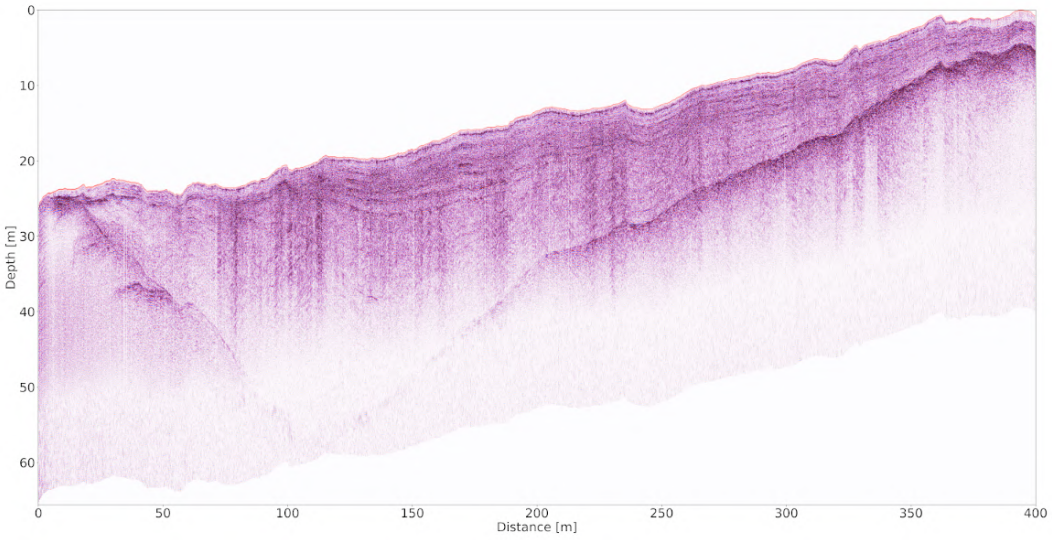

Figure 6.2.10: Depth-migrated merged field data as in Figure 6.2.7D and 6.2.7.

Many of the existing merging algorithms need to be applied to datasets after intensive pre-processing. In fact, in several procedures, amplitude recovery (i.e. gain) and spectral balance are mandatory (Xiao and Liu, 2015; De Coster and Lambot, 2018; Alemdağ et al., 2022). This introduces some subjectivity in the merged output and the sections

obtained by combining different frequency profiles present abrupt transitions between areas dominated by different central frequencies. In addition, data fusion procedures often require non-physical amplitude balance and pre-equalization, user-defined time window selections and post-processing procedures to minimize unwanted cut and paste effects (Lu et al., 2020; Roncoroni et al., 2023). On the contrary, our approach does not require any specific pre/post-processing and can be applied even on raw data when the signal-to-noise ratio is high enough. In our procedure, the integral of the amplitude spectra is scaled to the same normalization value before data merging, thus obtaining a spectral balance and normalization at the same time. In addition, the proposed procedure does not have specific requirements on the frequency range to be merged, nor to the number and sampling characteristics of the GPR profiles. The tests on real data (Figure 6.2.9, 6.2.10) show that merged data can be further processed without any remarkable issue, both in time and frequency domains. The test of the proposed merging procedure on the field dataset (i.e. the Antarctic glacier) took the following computation times on a laptop with Intel(R) Core (TM) $i7-10875H$, $32GbRAM$ and a Nvidia GeForce RTX 2070 Super with $8Gb$ of memory: training over 50% of each dataset (i.e. every two traces) corresponding to 1200 by 4000 matrix took $1635s$, while merging (i.e. prediction) time less than $4s$. The trained NN, once calculated, can be applied to any dataset with similar acquisition parameters. The computational costs are small enough to be managed by a laptop/pc.

Conclusion

Merging GPR data with different central frequencies can optimize the unavoidable trade-off between resolution and penetration depth when using the standard approach exploiting just a single couple of antennas. We proposed a new merging algorithm that can effectively manage the complementarity of information at different central frequencies, properly exploiting data redundancy. We tested the method on different synthetic and real cases in totally different situations and with different acquisition parameters and frequency ranges. Stability tests demonstrated the robustness of the algorithm and the improved quality of the results which can be further processed and analyzed. The proposed procedure is helpful not only to improve the response of reflectors but also of diffractors and in complex subsurface conditions, such as e.g. heterogeneous materials and conflicting dips. It can successfully manage even interfering events without introducing noise or artifacts. Another strength of the methodology is that it is totally data driven and can deal with noisy data. The training model requires a three order of magnitude longer computational time than the prediction. However, the training can be done just once and can then be used to merge any dataset with similar spectral, geological and geometrical characteristics. The proposed procedure is not limited to specific frequency components or geological

setting and can be potentially exploited to merge any type of dataset having different spectral components. Further research will be focused on the application of this methodology to merging data coming from 3D GPR datasets. We would also focus on possible applications on Common Shot Gather (CSG) stacking procedures, exploiting a modified version of this methodology.

Acknowledgments

The acquisition of Antarctic data is supported by a grant from the National Antarctic Research Program (PNRA), Research Project PNRA 2018_00186 – E. Authors would like to thank Dr. Philipp Koyan (Univ. of Potsdam) for kindly giving us the 3-D Herten synthetic GPR dataset. Cineca consortium is further acknowledged for making available its high-performance computing facilities.

Data and materials availability

B-Scan data are available at https://doi.org/10.6084/m9.figshare.23993688.v1.
C-Scan data can be found at http://dx.doi.org/10.17632/by3yh79hx4.1
Codes related to the methodology are presented in https://github.com/Giacomo-Roncoroni/merging_GPR

Bibliography

Hilal Alemdağ, Aysel Şeren, and Hakan Karslı. 2022. Combining ground-penetrating radar sections with different antenna frequencies including time-frequency domain noise suppression filters. *GEOPHYSICS* 87, 4 (2022), WB41–WB54. https://doi.org/10.1190/geo2021-0710.1 arXiv:https://doi.org/10.1190/geo2021-0710.1

Maurizio Azzaro, Maria Papale, Carmen Rizzo, Emanuele Forte, Davide Lenaz, Mauro Guglielmin, and Angelina Lo Giudice. 2022. Antarctic Salt-Cones: An Oasis of Microbial Life? The Example of Boulder Clay Glacier (Northern Victoria Land). *Microorganisms* 10, 9 (Aug. 2022), 1753. https://doi.org/10.3390/microorganisms10091753

Wenda Bi, Yonghui Zhao, Ruiqing Shen, Bo Li, Shufan Hu, and Shuangcheng Ge. 2020. Multi-frequency GPR data fusion and its application in NDT. *NDT & E International* 115 (Oct. 2020), 102289. https://doi.org/10.1016/j.ndteint.2020.102289

Adam D. Booth, Anthony L. Endres, and Tavi Murray. 2009. Spectral bandwidth enhancement of GPR profiling data using multiple-frequency compositing. *Journal of Applied Geophysics* 67, 1 (Jan. 2009), 88–97. https://doi.org/10.1016/j.jappgeo.2008.09.015

Alberic De Coster and Sebastien Lambot. 2018. Fusion of Multifrequency GPR Data Freed From Antenna Effects. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11, 2 (Feb. 2018), 664–674. https://doi.org/10.1109/JSTARS.2018.2790419

M. E. Dougherty, P. Michaels, J. R. Pelton, and L. M. Liberty. 1994. Enhancement of Ground Penetrating Radar Data through Signal Processing. In *Symposium on the Application of Geophysics to Engineering and Environmental Problems 1994*. Environment and Engineering Geophysical Society, 1021–1028. https://doi.org/10.4133/1.2922053

Anton Efremov, Oleksii Karpenko, and Lalita Udpa. 2022. Generalized multifrequency fusion algorithm for defect detection in eddy current inspection data. *NDT & E International* 129 (July 2022), 102654. https://doi.org/10.1016/j.ndteint.2022.102654

A L Endres, A Booth, and T Murray. 2004. Multiple frequency compositing of spatially coincident GPR data sets. (2004).

Svein-Erik Hamran, David A. Paige, Abigail Allwood, Hans E. F. Amundsen, Tor Berger, Sverre Brovoll, Lynn Carter, Titus M. Casademont, Leif Damsgård, Henning Dypvik, Sigurd Eide, Alberto G. Fairén, Rebecca Ghent, Jack Kohler, Michael T. Mellon, Daniel C. Nunes, Dirk Plettemeier, Patrick Russell, Matt Siegler, and Mats Jørgen Øyan. 2022. Ground penetrating radar observations of subsurface structures in the floor of Jezero crater, Mars. *Science Advances* 8, 34 (Aug. 2022), eabp8564. https://doi.org/10.1126/sciadv.abp8564

Svein-Erik Hamran, David A. Paige, Hans E. F. Amundsen, Tor Berger, Sverre Brovoll, Lynn Carter, Leif Damsgård, Henning Dypvik, Jo Eide, Sigurd Eide, Rebecca Ghent, Øystein Helleren, Jack Kohler, Mike Mellon, Daniel C. Nunes, Dirk Plettemeier, Kathryn Rowe, Patrick Russell, and Mats Jørgen Øyan. 2020. Radar Imager for Mars' Subsurface Experiment—RIMFAX. *Space Science Reviews* 216, 8 (Dec. 2020), 128. https://doi.org/10.1007/s11214-020-00740-4

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Comput.* 9, 8 (11 1997), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Harry M. Jol. 2009. . Elsevier, Amsterdam. iv pages. https://doi.org/10.1016/B978-0-444-53348-7.00016-8

Diederik Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations* (12 2014).

C. Kohl, M. Krause, Christiane Maierhofer, and J. Wöstmann. 2005. 2D- and 3D-visualisation of NDT-data using data fusion technique. *Materials and Structures* 38, 283 (April 2005), 817–826. https://doi.org/10.1617/14293

Philipp Koyan and Jens Tronicke. 2020a. 3D modeling of ground-penetrating radar data across a realistic sedimentary model. *Computers & Geosciences* 137 (April 2020), 104422. https://doi.org/10.1016/j.cageo.2020.104422

Philipp Koyan and Jens Tronicke. 2020b. *Analyzing 3D multi-frequency ground-penetrating radar (GPR) data simulated across a realistic sedimentary model.* 275–278. https://doi.org/10.1190/gpr2020-073.1 arXiv:https://library.seg.org/doi/pdf/10.1190/gpr2020-073.1

Frank Lehmann and Alan G. Green. 2000. Topographic migration of georadar data: Implications for acquisition and processing. *GEOPHYSICS* 65, 3 (May 2000), 836–848. https://doi.org/10.1190/1.1444781

Chunlai Li, Wei Zuo, Weibin Wen, Xingguo Zeng, Xingye Gao, Yuxuan Liu, Qiang Fu, Zhoubin Zhang, Yan Su, Xin Ren, Fang Wang, Jianjun Liu, Wei Yan, Xu Tan, Dawei Liu, Bin Liu, Hongbo Zhang, and Ziyuan Ouyang. 2021. Overview of the Chang'e-4 Mission: Opening the Frontier of Scientific Exploration of the Lunar Far Side. *Space Science Reviews* 217, 2 (March 2021), 35. https://doi.org/10.1007/s11214-021-00793-z

Federico Lombardi, Frank Podd, and Mercedes Solla. 2022. From Its Core to the Niche: Insights from GPR Applications. *Remote Sensing* 14, 13 (June 2022), 3033. https://doi.org/10.3390/rs14133033

Guoze Lu, Wenke Zhao, Emanuele Forte, Gang Tian, Yong Li, and Michele Pipan. 2020. Multi-frequency and multi-attribute GPR data fusion based on 2-D wavelet transform. *Measurement* 166 (Dec. 2020), 108243. https://doi.org/10.1016/j.measurement.2020.108243

Mostafa Naghizadeh and Mauricio Sacchi. 2009. Making FX Interpolation More Robust by Spectrum-guided Reconstruction. (2009).

Giacomo Roncoroni, Emanuele Forte, and Michele Pipan. 2023. Merging gated frequency-modulated continuous-wave Mars2020 RIMFAX ground-penetrating radar data. *GEOPHYSICS* 88, 2 (March 2023), A7–A12. https://doi.org/10.1190/geo2022-0466.1

David Rumelhart, Geoffrey Hinton, and Ronald Williams. 1986. Learning representations by back-propagating errors. *Nature* 323 (1986), 533–536. https://api.semanticscholar.org/CorpusID:205001834

M. Schuster and K.K. Paliwal. 1997. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing* 45, 11 (Nov. 1997), 2673–2681. https://doi.org/10.1109/78.650093

W.R. Scott, Kangwook Kim, G.D. Larson, Ali Gurbuz, and J.H. McClellan. 2004. Combined seismic,

radar, and induction sensor for landmine detection, Vol. 3. 1613 – 1616 vol.3. https://doi.org/10.1109/IGARSS.2004.1370637

Francesco Soldovieri and Luciana Orlando. 2009. Novel tomographic based approach and processing strategies for GPR measurements using multifrequency antennas. *Journal of Cultural Heritage* 10 (Dec. 2009), e83–e92. https://doi.org/10.1016/j.culher.2009.09.001

Craig Warren, Antonios Giannopoulos, and Iraklis Giannakis. 2016. gprMax: Open source software to simulate electromagnetic wave propagation for Ground Penetrating Radar. *Computer Physics Communications* 209 (Dec. 2016), 163–170. https://doi.org/10.1016/j.cpc.2016.08.020

Craig Warren, Antonios Giannopoulos, Alan Gray, Iraklis Giannakis, Alan Patterson, Laura Wetter, and Andre Hamrah. 2019. A CUDA-based GPU engine for gprMax: Open source FDTD electromagnetic simulation software. *Computer Physics Communications* 237 (April 2019), 208–218. https://doi.org/10.1016/j.cpc.2018.11.007

Jianping Xiao and Lanbo Liu. 2015. Multi-frequency GPR signal fusion using forward and inverse S-transform for detecting railway subgrade defects. In *2015 8th International Workshop on Advanced Ground Penetrating Radar (IWAGPR)*. IEEE, Florence, Italy, 1–4. https://doi.org/10.1109/IWAGPR.2015.7292680

Jianping Xiao and Lanbo Liu. 2016. Permafrost Subgrade Condition Assessment Using Extrapolation by Deterministic Deconvolution on Multifrequency GPR Data Acquired Along the Qinghai-Tibet Railway. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9, 1 (Jan. 2016), 83–90. https://doi.org/10.1109/JSTARS.2015.2487970

Xianlei Xu, Junpeng Li, Xu Qiao, and Gui Fang. 2019. Fusion of multiple time-domain GPR datasets of different center frequencies. *Near Surface Geophysics* 17, 2 (2019), 141–150. https://doi.org/10.1002/nsg.12033 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/nsg.12033

Wenke Zhao, Lin Yuan, Emanuele Forte, Guoze Lu, Gang Tian, and Michele Pipan. 2021a. Multi-Frequency GPR Data Fusion with Genetic Algorithms for Archaeological Prospection. *Remote Sensing* 13, 14 (July 2021), 2804. https://doi.org/10.3390/rs13142804

Wenke Zhao, Lin Yuan, Emanuele Forte, Guoze Lu, Gang Tian, and Michele Pipan. 2021b. Multi-Frequency GPR Data Fusion with Genetic Algorithms for Archaeological Prospection. *Remote Sensing* 13, 14 (Jul 2021), 2804. https://doi.org/10.3390/rs13142804

Chapter 7

Frequency Inference

The research for low-frequency seismic data, crucial for deep subsurface understanding, remains a fundamental challenge. These low-frequency components carry vital information about geological features, but noise and natural Earth filters hinder their acquisition. The Earth acts as a low-pass filter for seismic waves, attenuating high-frequency components more than low-frequency ones. These low-frequency signals are invaluable for deep seismic studies, imaging beneath high-velocity layers, and exploring complex geological areas. However, capturing reliable low-frequency data below $10Hz$ is complex due to increasing noise and marine ghost reflections.

Efforts to address this issue include developing low-frequency energy sources and innovative streamer configurations. Data processing techniques aim to extend seismic bandwidth and enhance signal-to-noise ratios, yet challenges persist, especially in the presence of ghost reflections. Deep learning, using neural networks, has emerged as a promising solution for enhancing low-frequency seismic data.

Our proposed methodologies employ Bi-Directional LSTM neural networks to enhance low-frequency seismic data. In this chapter we present two approaches, where the loss function considers both time and frequency domains: one to enhancement low frequency in a processed post-stack data for deep seismic imaging and the other one is a three-step approach to fill the low frequency null space gap on CSG.
By applying this methodologies, we aim to bridge the gap in low-frequency seismic data, advancing our understanding of Earth's geological features.

7.1 Low frequency enhancement for deep seismic imaging

The Earth acts as a low-pass filter for seismic waves, where high frequency components of energy are attenuated more than low frequency components which travel much further into the earth. Therefore, the lower end of the seismic spectrum is important for deep seismic studies, seismic imaging beneath high-velocity layers such as salt bodies (Etgen et al., 2016) or basalts (Ziolkowski and Kipple, 2003), as well as in complex geology areas such as thrust-fold belt (Brenders, 2011). Furthermore, low frequency components are crucial for quantitative imaging using seismic full waveform (Li and Demanet, 2016; Ovcharenko et al., 2019). However, extending the useful bandwidth of seismic data much below $10Hz$ is not trivial Ten Kroode et al. (2013). in fact, the level of noise tends to increase exponentially as frequency decreases. Furthermore, in marine environments, source- and receiver-side reflections from the sea surface (ghosts) produce periodic frequency notches (the first one at $0Hz$) and cause a low-cut filtering effect making low frequencies propagation and imaging challenging (Singh et al., 1996).

Efforts have recently been made to develop low frequency energy sources (Reust et al., 2015; Abma and Ross, 2013; Ronen and Chelminski, 2017; Morozov, 2021), but these sources are complicated to use and thus have had limited success up to now. For example, to increase the low frequency content down to $1.5 - 4Hz$, Ronen and Chelminski (2017) proposed sources with enormous volume (28,000 cubic inches) and low firing pressure, which pose significant operational challenges and are difficult to effectively implement in scientific or industrial surveys.

To address the problem associated with streamer ghost, different solutions have been proposed. Instead of deploying the streamer at a constant depth, slant (Ruan et al., 2014) or curved (Amundsen and Reitan, 2014) streamer configurations were proposed. As receivers are at different depths along the streamer, the notches would arrive at different frequencies and the resulting spectrum would be broadened (Soubaras and Whiting, 2011). Another strategy was to deploy two streamers at two different depths, in under/over configurations, allowing to simulate a third streamer: the sum of all the streamers would broaden the spectrum as well (Singh et al., 1996). These ideas have had some success but extending the lower frequencies below 2-3 Hz still remains a challenge. More recently, multi-sensor technology has been developed, where both pressure and vertical components data is recorded at the same depth (Farouki et al., 2010). As the notches on the pressure sensor coincide with the peaks on the vertical component sensor, adding these two data sets would cancel the notch, broaden

the spectrum, and enhance frequencies towards zero. However, noise on geophone components is very high at frequencies below 10 Hz, hence only pressure data are used below 10 Hz (Olofsson, 2010).
In addition, there have been efforts to enhance low frequency components during processing, which can be broadly divided into two categories: extension of seismic bandwidth towards the low frequency side (Sun et al., 2021) and improvement in signal-to-noise ratio in low frequency regime (Chen et al., 2019). For example, the application of zero-phase source deconvolution (Woodburn et al., 2011) and spectral whitening (Masoomzadeh et al., 2005) can boost the low frequencies but the signal to noise ratio tends to be poor. Another approach, called wavelet transform method (Smith et al., 2008), is based on the convolution of a time-series (seismic trace) with a scaled and translated wavelet (*Morlet wavelet*), resulting in the summation of several sub-harmonic frequencies to the original time-series, and hence boosting the low-frequency part of spectrum. However, frequencies available for harmonic prediction and notch due to ghost reflections during seismic acquisition pose a barrier in realizing this technique to its full potential. The application of a demodulation operator to extract the envelope of a seismic trace can further provide low-frequency information (Wu et al., 2014). Li and Demanet (2016) analytically extrapolating the low frequencies by decomposing selected seismic records into elementary events and considering inter-trace relations whereas Wang and Herrmann (2016) addressed frequency extrapolation as a convex optimization problem with a total-variation regularization that accounts for spatial correlation between traces. While these methodologies have been successfully applied to synthetic data, they have failed to yield useful results in their application to real data because these algorithms are very sensitive to signal-to-noise ratio in the recorded data as they often track strong events and treat weak events as noise.

Recently, Deep Learning based on neural network (NN) has been used to enhance the low frequencies components of seismic records.
Aharchaou and Baumstein (2020) proposed to measure ultra-low frequency $(1 - 4Hz)$ on the seabed, then using such a data to train a NN to determine a bandwidth extension function to enrich streamer data with low frequencies. However, as mentioned above, acquiring low-frequency data can be very challenging and costly. Ovcharenko et al. (2019) used high frequency data to train the NN to predict the low frequencies. The main issue with this methodology is that NN can have generalization problems since the training is linked to data acquired in a specific geological setting. With these premises, we propose a new method based on Bi-Directional

Long Short-Term Memory (LSTM) to enhance the lo frequency components of reflection seismic data and we show its efficiency on synthetic and real deep seismic datasets.

We chose Bi-directional LSTM (Hochreiter and Schmidhuber, 1997) neurons because of their capacity to deal with temporal dependencies and to predict any temporal length due to the recurrent nature of the layer. There are three different conceptual steps in our approach, namely: synthetic data computation, choice of the NN, and computation of error functions. The first step is dedicated to generate synthetic data. To allow the methods to deal with different amplitudes, frequency contents and noise levels, the dataset ($T(t)$) is randomly generated using a convolutional model of the form (Yilmaz, 2001),

$$T(t) \;=\; w(t) * (R(t) + n_1(t)) + n_2(t) \tag{7.1}$$

Where R(t) is a uniformly distributed reflection coefficient time series with random values, $n_1(t)$ and $n_2(t)$ are random noise with uniform distribution, $w(t)$ is the wavelet. $n_1(t)$ has been added to simulate real data noise due to wave propagation or scattering and $n_2(t)$ represent instrumental or environmental noise. In our case, both $n_1(t)$ and $n_2(t)$ randomly vary between 0 and 5% of the maximum amplitude of the signal.

We used two types of wavelets to generate two different traces. For the first trace we used a Ricker wavelet with a dominant frequency of $4Hz$, whereas for the second trace we used a broadband Ricker wavelet with frequency from 0.5 to $8Hz$. These two wavelets are used as input and output for the NN training, respectively. We choose a broader frequency spectrum for the NN target to train and predict a very broad set of frequencies.

The training dataset is intended to be a maximum phase: this is a crucial choice since we need to avoid phase shifts in the training dataset to properly position the reflections in the prediction. For example, a minimum phase wavelet will introduce a phase shift that would be difficult to take into account in the field data prediction due to the introduction of wide side lobes and, as a result, shifting the maximum phase of the wavelet. Since LSTM works on the data only in one direction, we would not be able to correctly reconstruct events after their maximum phase: to solve this problem, we applied a Bi-Directional wrapper that allows the neuron to work both on the main trace and on its time-reversed version.

After defining the training dataset and the types of neurons, we choose a NN geometry

based on three hidden layers with 4, 2 and 1 Bi-Directional neurons (Figure 7.1.1). The geometry of the NN was chosen after several tests by varying the complexity of the NN. We find that very complex NN geometries required a significant increase in training and prediction times with negligible reduction in errors, and therefore we chose the above describe simple geometry. The last important step of the methodology is the computation of the loss function. The loss function is defined by the sum of the Mean Squared Error (MSE) between the real and predicted traces. Normally, the loss function is computed in the time domain, but here we also include the frequency domain. This double loss function can be written as:

$$loss(X,\ Y) = \left\| \frac{X}{\max(|X|)} - \frac{Y}{\max(|X|)} \right\|_2 + \left\| \frac{\mathbb{R}(\hat{X})}{\max(\mathbb{R}(\hat{X}))} - \frac{\mathbb{R}(\hat{Y})}{\max(\mathbb{R}(\hat{X}))} \right\|_2 \tag{7.2}$$

where X is the reference output, Y is the NN prediction and $\mathbb{R}(\hat{X})$ and $\mathbb{R}(\hat{Y})$ are the real part of their Fourier transforms, respectively. The first term is the classical loss function, normalized according to the reference output, while the second term is its frequency domain counterpart. We added a normalization according to the reference output to avoid the use of a scaling coefficient, that would need further evaluation on the best value and could introduce some subjectivity into the choice. Furthermore, this kind of normalization allows the methodology to give the same weight to traces with very high and very low reflectivity values, since they are both scaled to one in the loss function. The addition of a frequency part to the classical loss function allows to give an higher contribute to the lower frequencies, which is essential for further calculations. Another advantage of considering both the time and frequency domain amplitudes is that the low frequency content affects less the time domain part of the loss function than the frequency domain part. For instance, if we compute the Mean Square Error (MSE) between Figure 7.1.1 a-c and Figure 7.1.1 a'-c' we get values of $1.43e^{-05}$ for the time-amplitude MSE, and $4.31e^{-04}$ for the amplitude-frequency counterpart. To allow the NN to predict signals independent of amplitudes, the training dataset is scaled by a random factor in the half interval of [0, 1]. This scale factor is balanced in the loss function by normalizing the true data to 7.1.1 and then scaling the prediction according to the true data re-normalization factor. In the second term of Equation 7.2, both $R(\hat{X})$ and $R(\hat{Y})$ are normalized according to $\max(R(\hat{X}))$.
This normalization approach has two advantages:

- It gives the same weight to signals with different amplitudes, which is crucial when working with real data;

- The two loss functions can be summed without requiring any further regularization coefficient, since they have the same order of magnitude.

We trained the NN using 100,000 randomly traces generated obtained using an optimizer Adamax (Kingma and Ba, 2014).

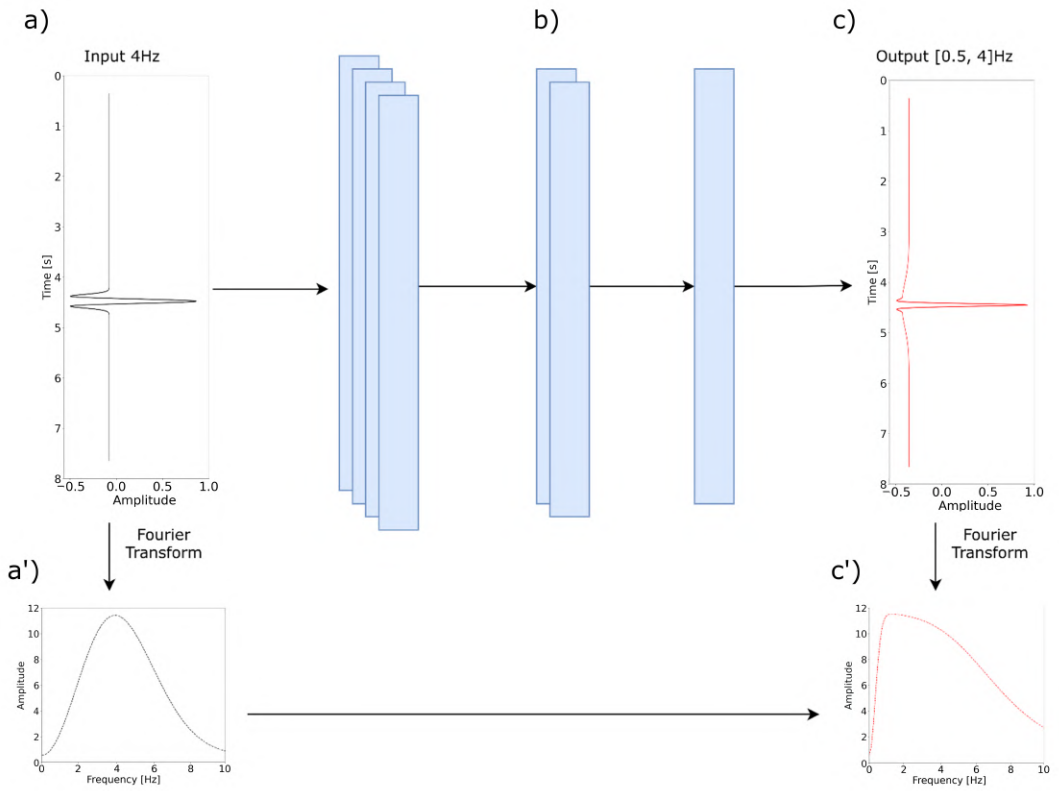

Figure 7.1.1: Schematic diagram showing the setup of the methodology: (a) Input trace with a dominant frequency of 4Hz. The LSTM neurons and the bi-directional wrapper are depicted (b). In (c) the broadband frequency output. Over the normal loss, computed through input (a) and output (c), i.e. the first term in Equation 2, we also have the amplitude-frequency counterpart (a' and c'), i.e. the second term in Equation 7.2.

The NN was first tested on a synthetic data set. The synthetic trace contains only three events: one at $1.5s$ and the other two at $3.5s$ and $4.5s$, interfering with each other (Figure 7.1.2a). The time sampling is constant and equal to 2ms. The seismic source is a Ricker wavelet with a dominant frequency of $3Hz$ (Figure 7.1.2b). We added 5% of random noise before and 10% after convolution (see Equation 7.1). The first event in Figure 2a has the maximum amplitude of -0.3 and the second event

has two interfering reflections with maximum values of 0.87 and 0.92. We generated $100,000$ data with randomly chosen number of events at random times and with random amplitudes according to in Equation 7.1. The synthetic test on the blind dataset of the trained NN shows that the NN has allowed to extend the frequency bandwidth from 1 to 7 Hz. The side lobes present in the original signal have been reduced and the events after the application of the NN prediction are sharper. From this result we can also see that the phases are correctly retained by the method and the amplitudes are preserved. Moreover, the two events at $3.5s$ and $4.5s$ are distinct and clear. To quantitatively evaluate the obtained result we estimated the MSE both in the time and in the frequency domains: The time-amplitude MSE was $6.7x10^{-4}$, while the frequency-amplitude MSE was $1.2x10^{-3}$, about twice that of time-amplitude MSE, indicating that the frequency amplitude has larger impact on the prediction. As described, this is the key point that led us to include frequency-amplitude MSE in the loss function.

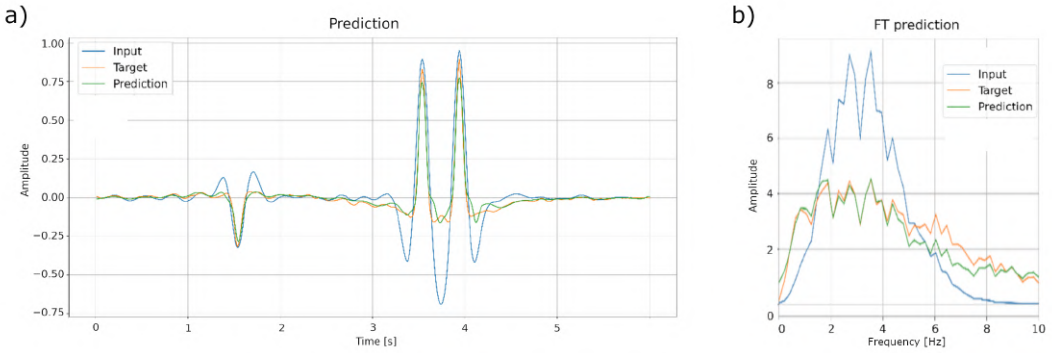

Figure 7.1.2: Blind test of the NN for a synthetic data set. (a) Input at $4Hz$ (blue), target data (orange) and prediction (green) and (b) the corresponding amplitude-frequency spectra.

To test this methodology on field data we used processed deep seismic reflection data from the Atlantic Ocean (Audhkhasi and Singh, 2022). The original data were acquired using a $12km$ long multi-sensor streamer and a 10170 cubic inch source. The shot interval was $75m$, and the record length was $31s$. The data were processed to enhance the low frequency energy reflected from the lithosphere-asthenosphere boundary and the Gutenberg discontinuity. We have applied the trained NN in a time window from 20 to $31s$ along a $220km$ long profile portion covering the oceanic lithosphere from $27Ma$ to $47Ma$ (Audhkhasi and Singh, 2022). Further details about data processing and interpretation can be found in Audhkhasi and Singh (2022). Figure 7.1.3a shows a comparison between the full-processed seismic profile obtained

by Audhkhasi and Singh (2022) and (Figure 7.1.3b) the results of the NN prediction. To make possible a comparison of the signal amplitudes with synthetic data, we applied a global normalization, setting to one the maximum amplitude of the whole dataset . We then applied the NN trained as previously described on the field data. The NN was applied trace by trace, i.e. strictly in 1-D thus inherently considering each trace completely independent by all the others. The higher continuity of the two target reflectors related to the lithosphere-asthenosphere boundary (lying between about $22s$ and $26s$) and to the Gutenberg discontinuity (at about $29s$) on the enhanced frequency profile is apparent. The resulting image is overall sharper and the signal to noise has also increased. We can also observe small scale structural variations in the two reflection events. Since our approach is locally 1-D, the lateral coherency is real, and cannot be relater to artefacts introduced by the applied methodology. The frequency spectrum shows the enhancement of low frequencies down to at least $1Hz$ and the bandwidth was clearly enlarged.

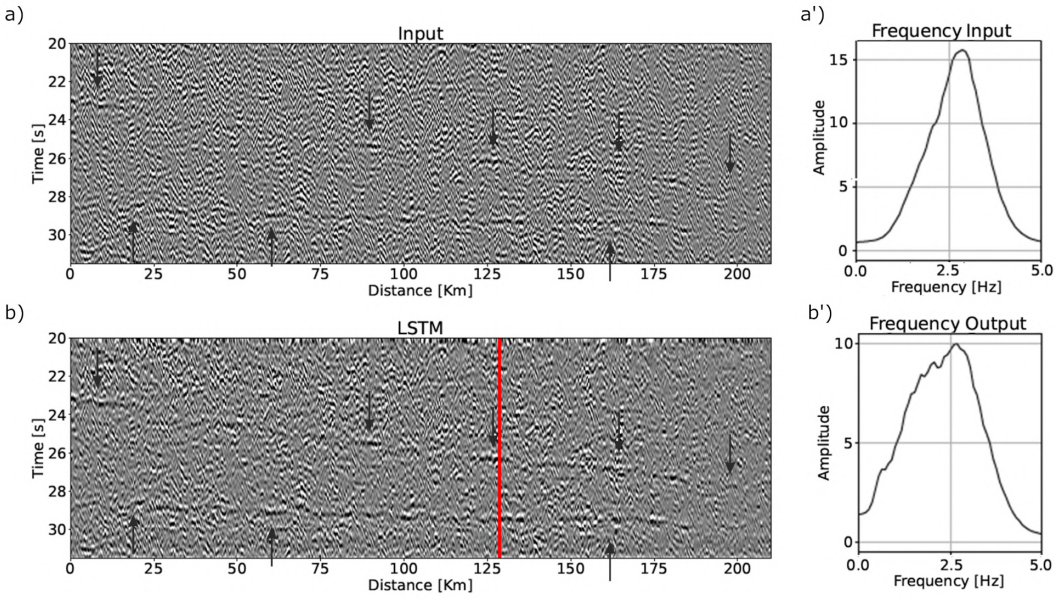

Figure 7.1.3: Processed data (a) and enhanced frequency data (b) and the corresponding amplitude-frequency spectra (a' and b', respectively). The two horizons described in Audhkhasi and Singh (2022) are pointed out by black arrows. Solid red lines mark the area analyzed in Figure 7.1.4.

To better highlight the low frequency enhancement, we vertically stacked 80 traces to reduce random noise and to improve the interpretability of the reflection events in a portion of the profile where the two reflection events are nearly flat and clearly visible on the original processed section, i.e. at $120km$ along the profile. Figure 7.1.4 shows

both the original and NN results. We can clearly see that the amplitudes of lobes have been reduced and the signal-to-noise ratio has been improved. Our method is able to preserve amplitudes, as we apply the inverse of the normalisation factor, and it does not introduce phase shifts.

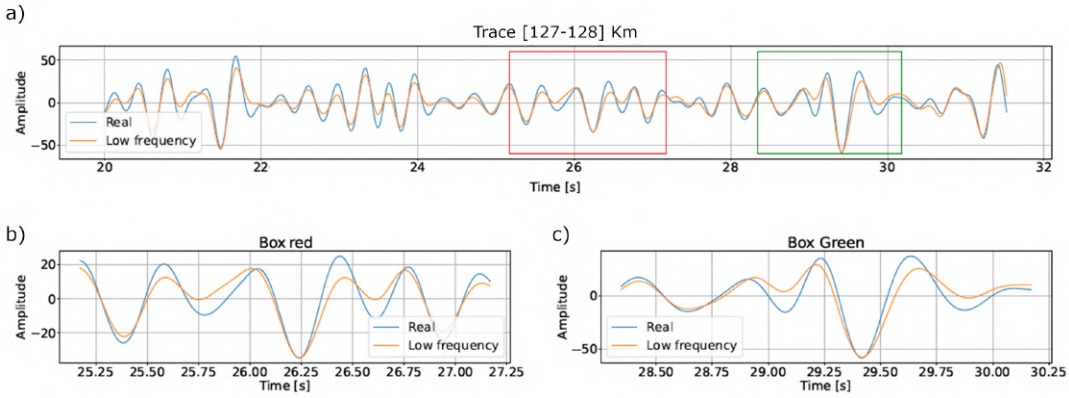

Figure 7.1.4: Plot of the waves from the processed (light blue) and the enhanced low frequency data (orange). 12 seconds of the recordings stacked in one Km (a) and two close up on the deep reflections analyzed in Audhkhasi and Singh (2022) ((b), the shallower and (c) the deeper).

The procedure took 11 hours to train the network on the Cineca cluster Marconi100, with the operation parallelized on a single node with 2x16 cores IBM POWER9 AC922 at 3.1, 4 NVIDIA Volta V100 GPUs and 256 GB RAM. After the training, the NN prediction was performed on a laptop with 32Gb Ram and a Nvidia RTX 1080 with 8GB of GPU memory and took only 12 s for the whole dataset, i.e., 71,000 traces, each with 3,500 samples.

We propose a methodology based on a LSTM NN for low frequency enhancement of seismic data. The training of the NN is based on a synthetic dataset, similar to real data, allowing to control different features of the algorithm and check its affordability. We have used an innovative loss function based both on the minimization of MSE between trained and predicted data (amplitude), and their Fourier transforms (amplitude spectra). This allows not only to have a faster and more precise convergence during the training, but also to obtain a more accurate and reliable output. The methodology is purely 1-D, i.e. independently applied trace by trace, thus allowing a successful application even on very noisy datasets, while avoiding artifacts that 2-D algorithms may introduce. The application of the method to real deep seismic data shows that it can be successfully applied and provides sharper and more interpretable

deep images. The methodology is very fast after the initial training and can be applied to a large variety of seismic data in terms of scale, resolution, source type, dimensions.

Acknowledgments

This research was supported by IPGP (Paris, France). Our method exploited CINECA HPC, within the agreement CINECA-University of Trieste.

Open Research

All data needed to evaluate the conclusions in the paper are present in the paper. The seismic data used in this study can be found at http://get.iedadata.org/doi/329809.

7.2 Low-frequency inference: filling null space gap with LSTM.

The absence of low-frequency signals in seismic data can lead to the emergence of local minima, making the extraction of reliable information a challenge. To tackle this issue, DL techniques have shown promise in improving the accuracy and effectiveness of reconstructing missing low-frequency signals (Ovcharenko et al., 2019).
As seismic waves propagate through the Earth, they undergo a low-pass filtering effect, where high-frequency energy components are rapidly attenuated, scattered, and dispersed, while low-frequency signals travel more extensively through the subsurface (Etgen et al., 2016). This low-frequency spectrum is of particular importance for deep seismic studies in various geological contexts, including areas with high-velocity layers such as salt bodies (Etgen et al., 2016), basaltic structures in sedimentary environments (Ziolkowski and Kipple, 2003), and thrust-fold belt regions (Brenders et al., 2010). Low-frequency energies also play a crucial role in quantitative imaging using seismic full waveform inversion (FWI), enabling the characterization of rock and fluid properties and reservoirs (Li and Demanet, 2016). However, extending the useful bandwidth of seismic data well below 10 Hz is a challenging endeavor due to the exponential increase in noise levels at lower frequencies (Ten Kroode et al., 2013).

In marine environments, the presence of source and receiver-side reflections at the water surface (known as "ghosts") introduces periodic frequency notches and imposes a low-cut filtering effect, further complicating low-frequency imaging (Singh et al., 1996). Recent efforts have been made by various companies to develop low-frequency energy sources to address this challenge (Aznar et al., 2022; Chelminski et al., 2019). Solutions proposed to mitigate the steamer ghost problem include deploying slant or curved streamers to introduce frequency diversity (Provenzano et al., 2020) and using multi-sensor technology to cancel out notches (Mellier and Tellier, 2018).
Both the scientific and industrial communities have invested substantial efforts in expanding the usable bandwidth of seismic data and improving low-frequency data quality. These efforts involve modifications in both source and receiver configurations (Chelminski et al., 2019) . On the source side, increasing the source volume and adjusting the towing depth have been explored, albeit with certain limitations (Ronen and Chelminski, 2017; Chelminski et al., 2019). Hyperclusters of air guns have been tested to decrease the characteristic frequency, although they come with amplitude trade-offs (F. Hopperstad et al., 2012) Asynchronous firing of air guns for constructive interference has been proposed, with the "Popcorn" method showing promising results (Abma, 2018).

On the receiver side, various broadband streamer solutions have been proposed, including deep-towed solid streamers, slanted streamers, over/under streamers, and multi-component streamers (Singh et al., 1996; Soubaras and Whiting, 2011; Mellier and Tellier, 2018).
Multi-sensor streamer technology, incorporating hydrophones and accelerometers, has emerged as an advanced solution for improving seismic data quality Carlson et al. (2007); Firth et al. (2018). It offers improved wavefield separation and deghosting capabilities, enhancing the quality of seismic data (Firth et al., 2018). This technology is expected to drive future seismic imaging and reservoir characterization efforts.

In seismic data processing, efforts to address low-frequency challenges are categorized into two main areas: extending seismic bandwidth toward lower frequencies and improving signal-to-noise ratio at low frequencies (Masoomzadeh et al., 2005; Woodburn et al., 2011; Li and Demanet, 2016). Techniques such as zero-phase source deconvolution and spectral whitening have been applied to boost low frequencies, but they often yield poor signal-to-noise ratios (Woodburn et al., 2011; Masoomzadeh et al., 2005). Continuous Wavelet Transform has been used to add sub-harmonic frequencies to the original time-series, enhancing low-frequency content (Smith et al., 2008). Techniques for extending frequency bandwidth include demodulation operators, analytical extrapolation of low frequencies, and convex optimization approaches (Wu et al., 2014; Li and Demanet, 2016; Wang and Herrmann, 2016).

Machine learning approaches, such as artificial neural networks, have also been developed to extrapolate low frequencies from limited bandwidth seismic data Ovcharenko et al. (2019). These methods offer potential solutions to enhance the quality of low-frequency seismic data (Fang et al., 2020).

One of the ongoing challenges in low-frequency seismic data is the low signal-to-noise ratio caused by surface noise contamination. Dense receiver spacing and high-fold coverage have been suggested to mitigate this issue (Farouki et al., 2010). Additional techniques, such as Empirical Mode Decomposition and Variational Mode Decomposition, aim to separate signal and noise components for better data quality (Liu et al., 2023). Mathematical morphological filtering has also demonstrated improved results with reduced computational demands (Liu et al., 2019).

In conclusion, the quest to enhance low-frequency seismic data quality involves innovations in acquisition technologies, data processing techniques, and the application of advanced machine learning methods. While challenges persist, ongoing research and development efforts continue to push the boundaries of what is achievable in the field of seismic exploration.We tried to address the well-known and crucial problem of the null space in the frequency domain with a methodology based on LSTM and a training performed only on synthetic data.

The chosen neurons are Bi-directional LSTM due to their capacity to deal with temporal dependencies and to predict any temporal length due to the intrinsic recurrent nature of the layer.
If we feed a NN with a time series, the link between its temporal discretization and the actual recording time is an unknown for the NN, because it takes as input only amplitude information disregarding the time occurrence.
Starting from the theoretical description of LSTM, provided in Chapter 2, it is clear that we provide, as input, the amplitude value at time t. This feature is crucial for the approach and allows a novel formulation when dealing with signal discretization.
Since the NN has no information on the signal time-discretization, we can think at the sampling frequency (i.e. the inverse of time-discretization) information as strictly dependent to the Δt chosen. This is quite typical in signal processing and we can both assume a conventional and constant $\Delta t = 1$ (i.e. applying an arbitrary time sampling normalization, in turn producing a sampling frequency normalization) or as an alternative approach, we can consider the signal nominal frequency as the length in samples of the source wavelet, called *sample duration* from now on.

$$\text{sample duration} = \frac{T}{\Delta t} \tag{7.3}$$

where T is the duration of a wavelet and Δt is the sampling interval.
What we do is training a NN to take wavelets with variable *sample duration*, e.g. from 100 samples, and return it to wavelets at maximum phase all with 200 samples.
Since the *sample duration* is often difficult to retrieve on real data, we can set the this therm as related to the second zero of the auto-correlation of the trace (Yilmaz, 2001; Treitel, 1969), as often done in common seismic processing algorithms (Robinson and Treitel, 2000). By doing this we are able to generalize the whole process and we can train on a arbitrary frequency signal (e.g. equal to 32Hz with $\Delta t = 0.0008s$) and generalize this to a $Lag = 22$ as obtained from the second zero of the auto-correlation this signal.
If we take such a signal and we compare it to a 8Hz signal sampled at $\Delta t = 0.0008s$,

as shown in Figure 7.2.1, we can compute the second zeros of the auto-correlation, resulting in a $Lag = 91$ for $8Hz$.

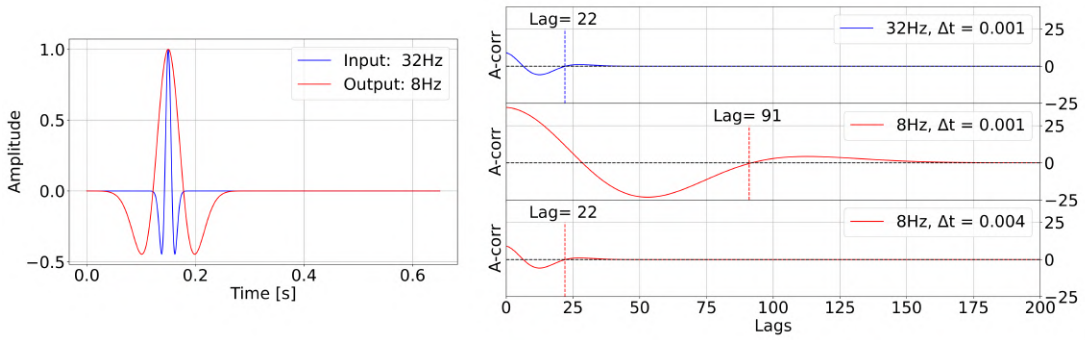

Figure 7.2.1: General description of the auto-correlation process. On the left we have 2 Ricker wavelets, $32Hz$, $\Delta t = 0.001s$ (Blue) and $8Hz$, $\Delta t = 0.001s$ (red). On the left we can see the second zero of the auto-correlation for this two signals (top and middle) and for the $8Hz$ resampled to $\Delta t = 0.004s$

If we now resample the 8Hz signal with a $\Delta t = 0.0032s$ (i.e. one sample every four), the resampled signal has still 8Hz frequency, but a $Lag = 22$. In this way, the NN will treat as the same sampling frequency both the $32Hz$ the resample 8Hz signals. This can be extended to any input signal and any resampled version of it.

With this generalization, we can train the NN with the same strategy previously introduced in Section 3.2, but with the possibility to scale and adapt the methodology to, virtually, any wavelet by simply resampling it to the Lag used in the training, i.e. $Lag = 22$ in the previously described example.
In Figure 7.2.2 we can see an exemplary trace from the training: Figure 7.2.2-A shows the noisy reflection coefficient, as described in Equation 3.2, Figure 7.2.2-B shows the computed trace at 32Hz with a sampling rate of 0.0008s, i.e. the input of the training, and Figure 7.2.2-C shows the reference output, i.e. a trace at 8Hz with the same sampling rate as before.
According to the Nyquist–Shannon sampling theorem (Shannon, 1949), we can compute the Nyquist frequency, which is equal to 625 Hz: this very large value is crucial for avoiding the introduction of possible aliasing and to make possible a signal decimation (to obtain the desired Lag) without substantial loss in amplitude (Dossi et al., 2018).
The use of the simple pure convolutive approach allows us to produce in 24s, 100,000 randomly simulated traces which can be use for any training purpose.

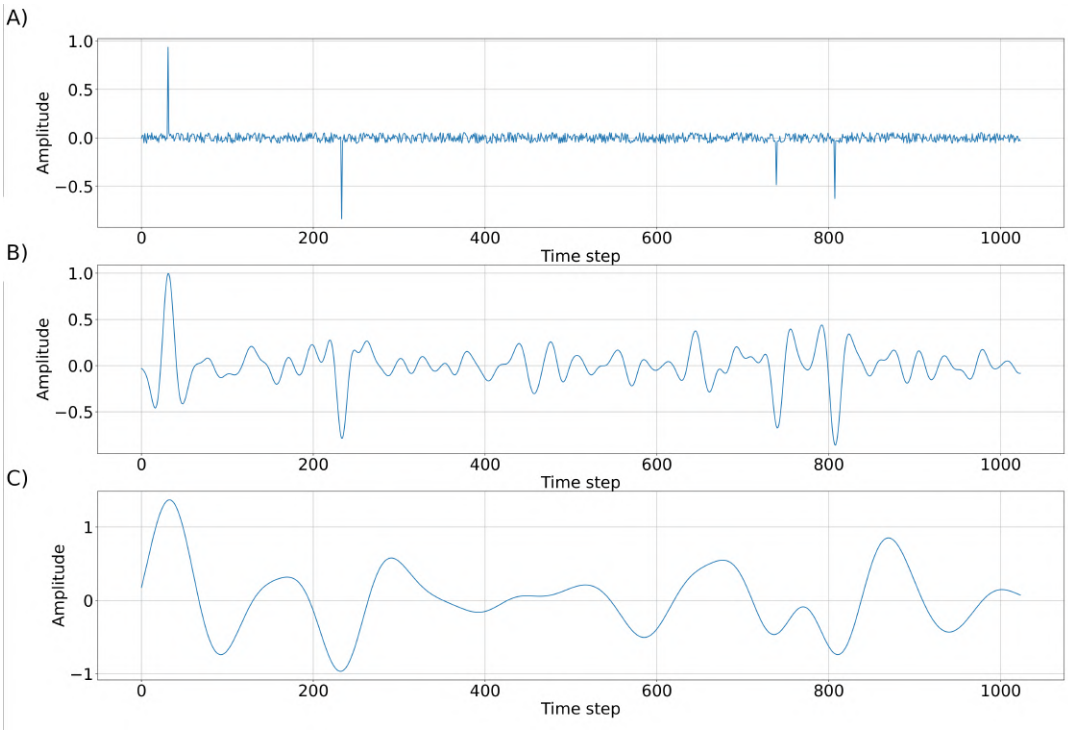

Figure 7.2.2: Example from the 100,000 traces of the training dataset: reflection coefficient (A), high frequency, i.e. 32Hz, input (B) and low frequency, i.e. 8Hz, output (C)

We train the NN with the final aim to predict the $8Hz$ counterpart starting from an input with a frequency of $32Hz$. In order to ensure the NN is able to properly work in the frequency domain, we used as the loss function the same introduced in the previous section, avoiding any amplitude normalization:

$$loss(X,\ Y) = \|X - Y\|_2 + \lambda \left\| \mathbb{R}(\hat{X}) - \mathbb{R}(\hat{Y}) \right\|_2 \tag{7.4}$$

where X is the reference output, Y is the NN prediction, $\mathbb{R}(\hat{X})$ and $\mathbb{R}(\hat{Y})$ are the real and imaginary parts of their Fourier transforms, respectively, while λ is a coefficient to weight the frequency loss.
In the next part of this chapter we will refer to $\|X - Y\|_2$ as the amplitude part and to $\left\| \mathbb{R}(\hat{X}) - \mathbb{R}(\hat{Y}) \right\|_2$ as the frequency part of the loss.
In order to properly understand the stability of the NN, we performed a grid search on the dimensions, training 7 NNs with a different numbers of HL. Assuming the number of HL equal to n, the geometry of the NN is assumed to have a decreasing number of Bi-directional neurons starting from 2^{n-1}, for the first HL, to 2^0, for the last HL. We could label the HL as: $NN(i)\ \ 2^{n-i}\ \ for\ \ i\ \ in\ n$. At the end of each NN we used a single LSTM neuron without the Bi-directional wrapper.
In Figure 7.2.3-A we plot the trends of the loss function, both for training and

validation losses, of the 7 NN.
These were trained on the same dataset of 100,000 traces for 100 epochs, as optimizer we used Adamax with a learning rate of 0.001.

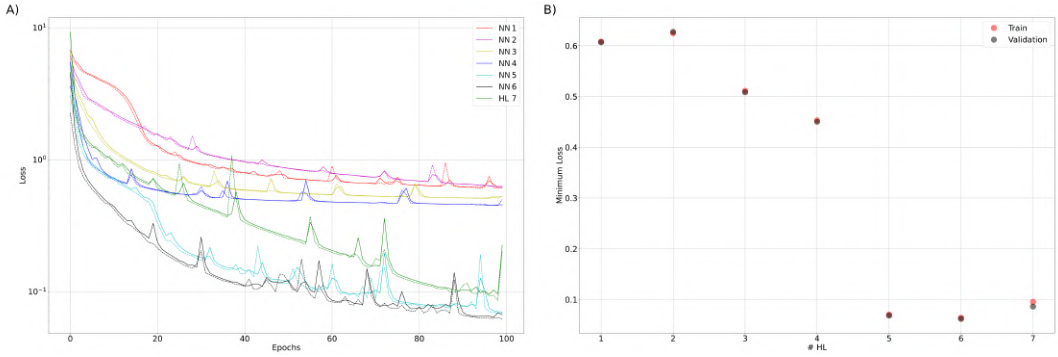

Figure 7.2.3: NN geometry test: in A the training loss (solid line) and validation loss (dashed line) for different number of Hidden Layer, from 1 to 7. In B the minimum value of the loss reached at 100 Epochs.

The results of the training, provided in Figure 7.2.3-A, clearly show that from NN1 to NN4 the training stops in a local minimum, while from NN5 from NN7 we reach lower loss values. If we focus on the trend of NN7, we can also see that the descent direction is more linear, which could be due to the amount of weights that are present in NN7 model, leading to a slower training.
In Figure 7.2.3-B we plot the minimum loss value for each NN. As we can see, we reach a minimum between NN5 and NN6: NN5 is chosen for computational profitability.
If we now look at the prediction of a blind data, Figure 7.2.4, we can see something interesting: NN1 is minimizing the amplitude part of the loss, while NN2 to NN4 minimize just the frequency part, leading to a prediction that has no phase consistency with the reference output. Indeed, from NN5 to NN7 we have good results both in amplitude-phase and in frequency.

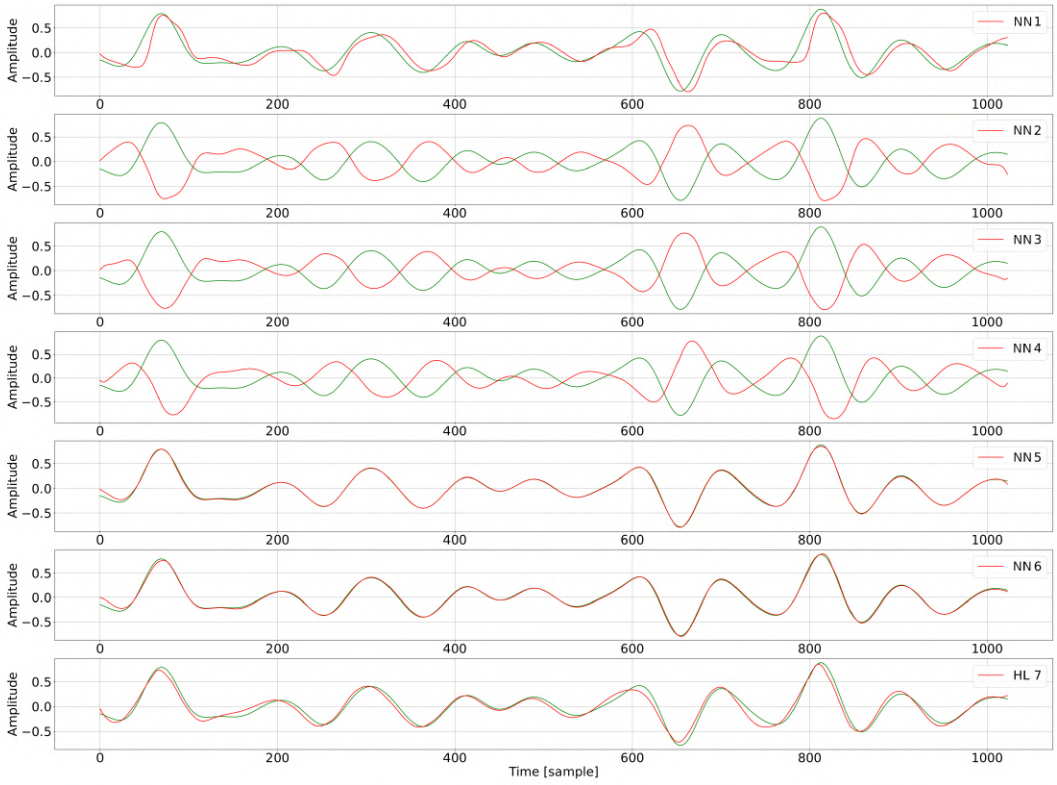

Figure 7.2.4: Prediction on a blind data for the NN trained as in Figure 7.2.3. We can see the amplitude part is minimized in NN1, frequency part is taken into account from NN2 to NN4,m but amplitude is wrong, while from NN5 the NN is able to minimize both counterpart of the loss function, see equation7.4

Figure 7.2.5 shows the current methodology workflow: starting from input A, we compute the second zero of the auto-correlation (a.k.a. data lag) and we compute the scale factor as:

$$scale\ factor\ =\ \frac{data\ lag}{train\ lag} \tag{7.5}$$

where *data lag* is the lag value computed from target data (B), while *train lag* is the value used for NN training (C).

Once we computed the scale factor, we resample the data according to the new value, i.e. $data\ sample \cdot scale\ factor$ and we make the inference with the NN (E) and we resample back to original *data sample* number (F).

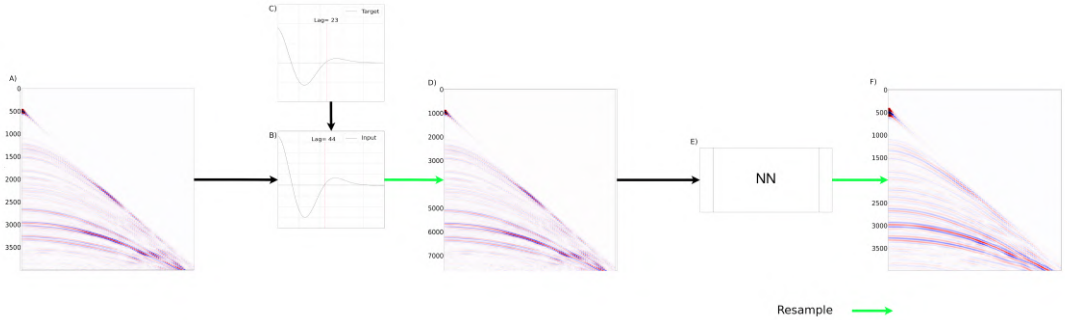

Figure 7.2.5: Representation of the workflow of the presented methodology, in green the resample procedure, first resample according to the factor, i.e. scale factor, computed between data-derived *lag* and the sample *lag* used for training. Second resample is to bring the Δt to the initial one.

Synthetic example: Marmousi model

In order to test the entire above-described methodology, we at first applied it on synthetic data, namely the well-known Marmousi model (Martin, 2004) We performed an acoustic forward modelling on the whole model with, as simulated source, a Ricker with 16Hz, 8Hz and 4Hz central frequencies, respectively (A, B and C in Figure 7.2.6).

As we have described before, the methodology starts from a defined auto-correlation lag, set by the training, and brings the data down to a *target lag*. As set during the training, a higher lag in the data will be always brought back to the *target lag*. This means that, if we apply different resample factors to the input data, the application of the NN would lead to an inference that has always the same *target lag*, but different Δt.

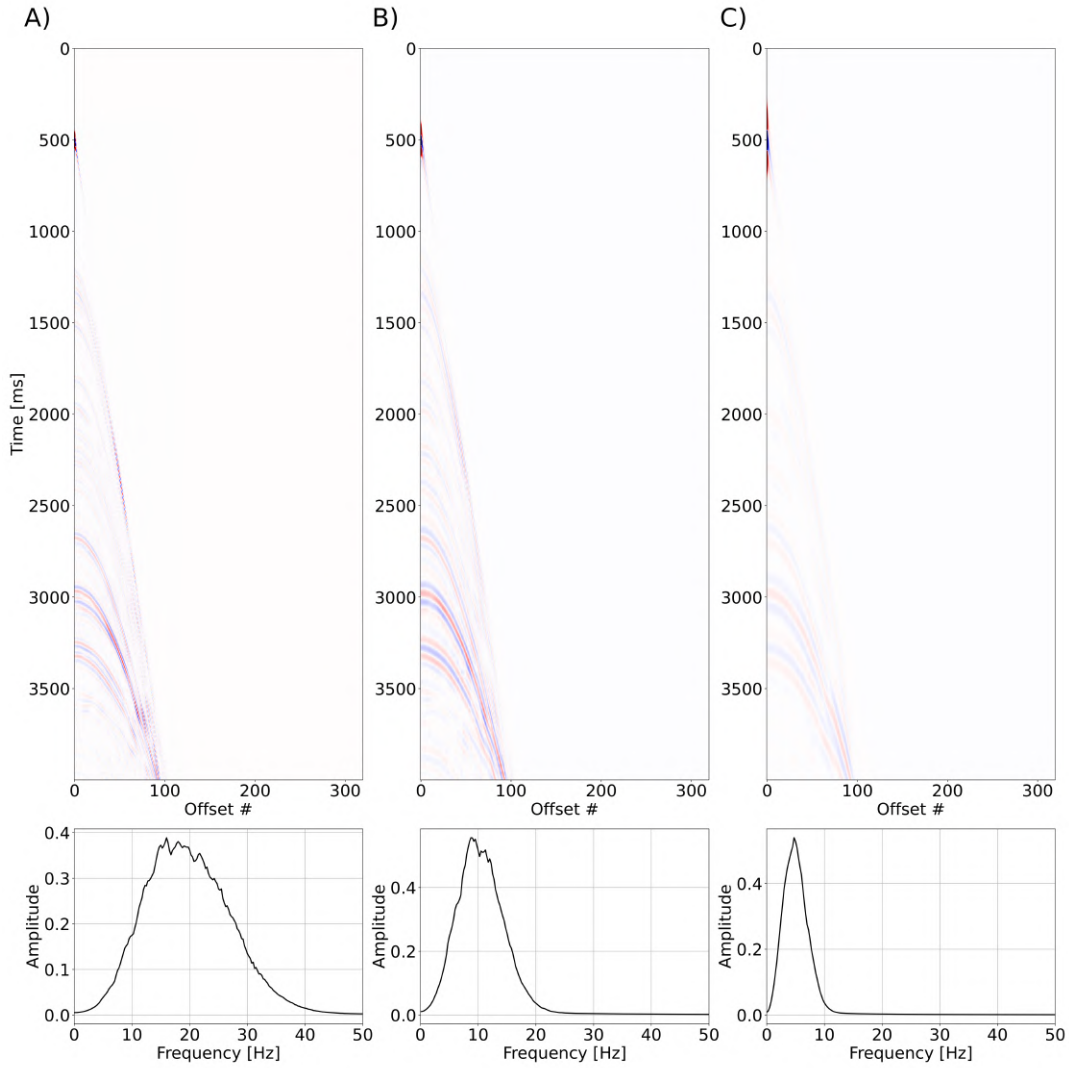

Figure 7.2.6: Synthetic test on Marmousi model (Martin et al., 2002). In A), B), C) a sample CSG at 16Hz, 8Hz and 4Hz, respectively.

Thanks to this, Figure 7.2.7 shows how the prediction, after the last resample depicted in Figure 7.2.5, changes the actual frequency content.
In Figure 7.2.7 we plot the variation of the amplitude-frequency content in the predicted data as a function of the variation of the *scale factor*. As we can see, the frequency shift is consistent with the variation in the *scale factor*.

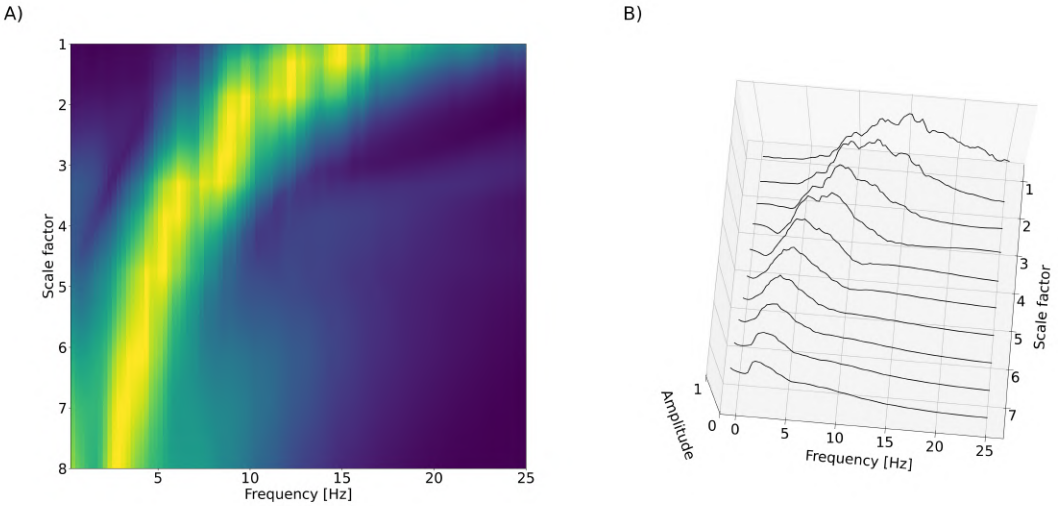

Figure 7.2.7: Resample factor (applied to the input) vs predicted frequency spectrum: all the predicted data are resampled to 0.001s before computing the spectrum. In A a plain view of 200 different resample factor (linearly separated from 1 to 8). In B some amplitude-frequency spectra are depicted.

If we now apply the methodology to the data shown in Figure 7.2.6-A and we predict the data shown in Figure 7.2.6-B, we can see in Figure /7.2.8 that the low frequencies are properly predicted in both domains, i.e. time-amplitude and the amplitude-frequency. If we focus on the thin bed reflections at $\approx 3000ns$ and $\approx 3250ns$ we can appreciate that even their constructive interference occurring at lower frequencies is correctly predicted by the NN.

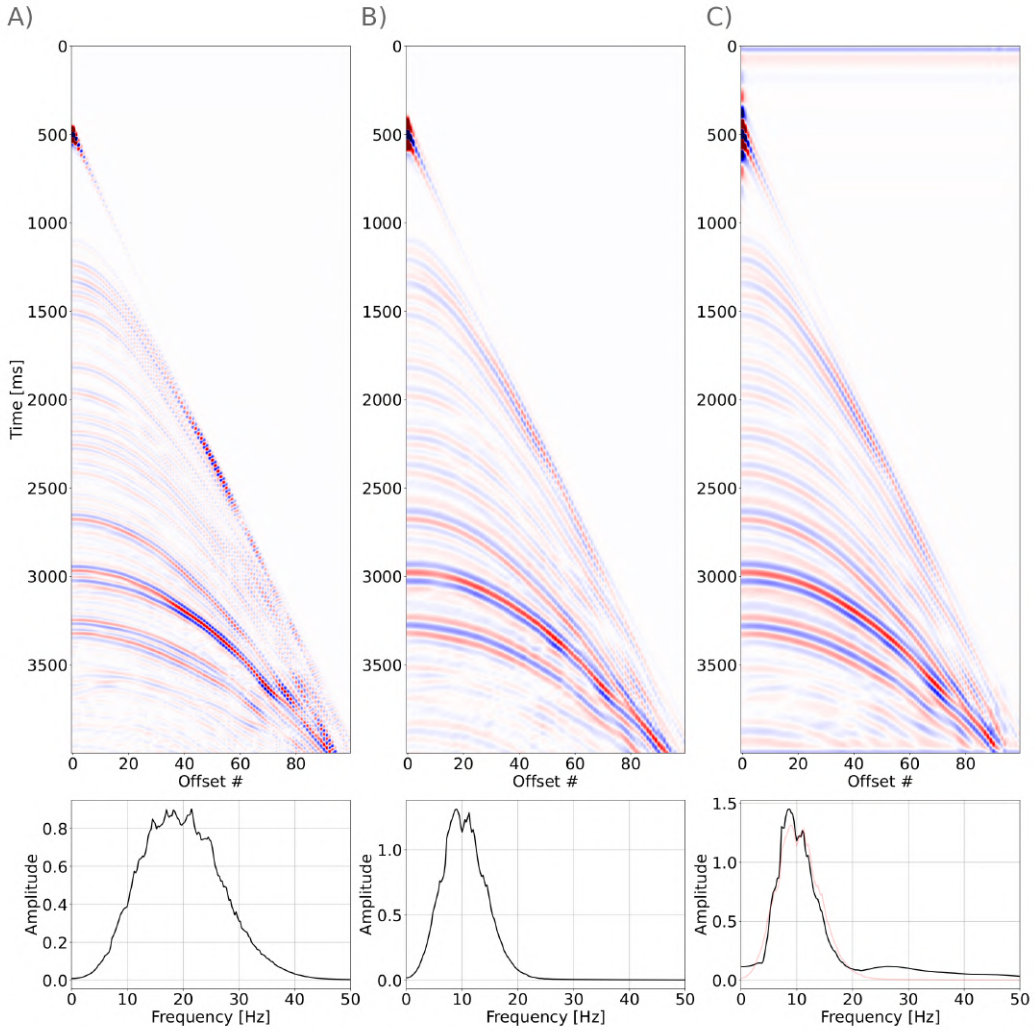

Figure 7.2.8: Input at 16Hz (A), reference output at 8Hz (B) and NN prediction (C). On the bottom the amplitude-frequency plots of each CSG.

If we try to further extent the methodology to lower frequencies, e.g. $4Hz$, we further decimate the $16Hz$ data and we make the prediction on that data. According to Figure 7.2.7 we apply a scale factor of 5 and we make the inference. In Figure 7.2.9 we can see the results: we have a good match in the amplitude-frequency domain but we introduce some errors. In particular, we have a reverberation of the 0-offset direct wave, due to its very high amplitude, and we further introduce some artifacts at the beginning and at the end of the record. However a standard processing can be easily applied to remove this errors, e.g. an horizontal high pass filter.
If we focus again on the thin bed reflections at $\approx 3000ns$ and $\approx 3250ns$, we can see that we are not only reconstructing properly their interference, but we also manage to locate the interference point at $\approx 3500ns$ and offset $\#60 - 70$.

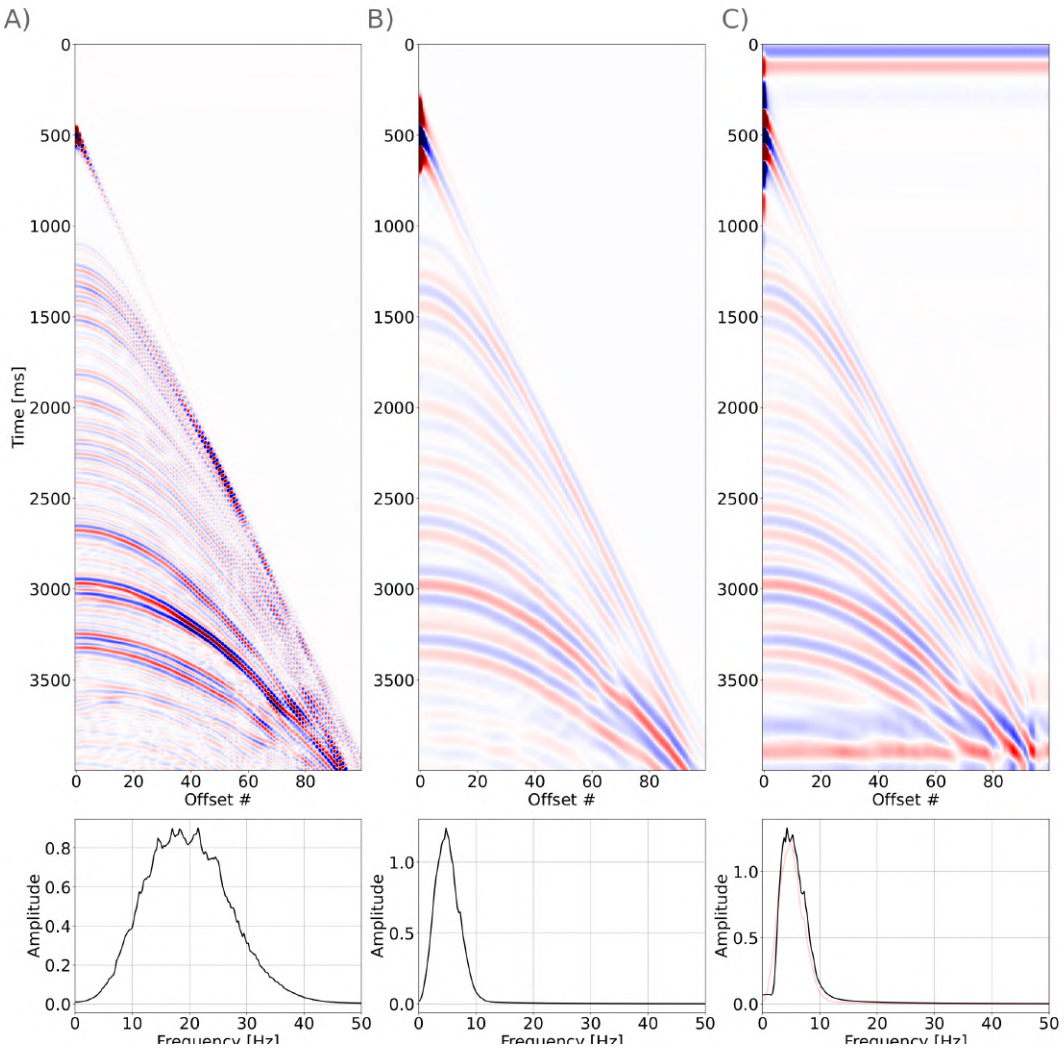

Figure 7.2.9: Input at 16Hz (A), reference output at 4Hz (B) and NN prediction (C). On the bottom the amplitude-frequency plots of each CSG.

If we now focus on a single trace (e.g. offset #10), we can see a comparison between the $16Hz$ input, Figure 7.2.10-A, the computed and the predicted $8Hz$, Figure 7.2.10-B and the computed and the predicted $4Hz$, Figure 7.2.10-C.
Although some amplitude issues, that could be due to errors in the signal decimation (Dossi et al., 2018) or in the interference prediction, we can see that the overall prediction is consistent with the target data and we do no have severe phase-shift introduced by the methodology.
In Figure 7.2.10-C we can also clearly see the artifacts at the beginning of the trace and at the end, that could be due to the absence of signal, for the beginning, since the NN is trained only on noise data and by the abrupt cut of the signals at the end: further investigation will be crucial for clarifying this issues.

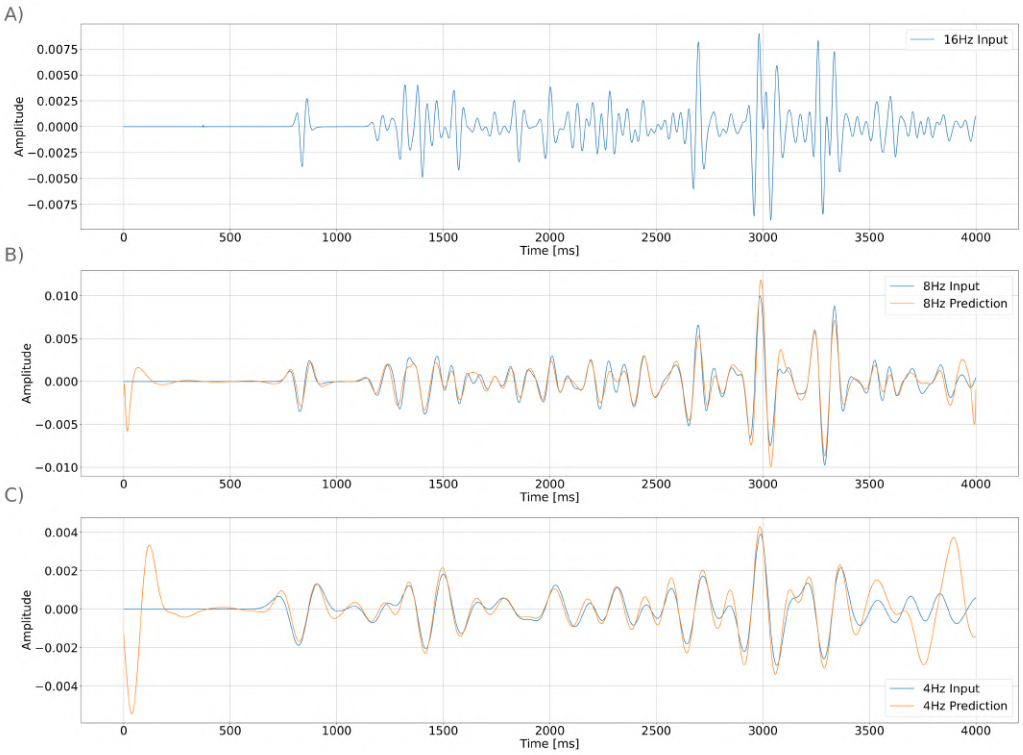

Figure 7.2.10: Single trace plot of the input (A), the 8Hz reference data, blue, and prediction, orange (B) and the 4Hz reference data, blue, and prediction, orange (C)

Field example: Western Sardinia

High-resolution seismic profiles covering a distance of 1430 kilometers were acquired as part of the WS10 project by OGS (Geletti et al., 2014): this data were acquired with a seismic source composed by a pair of GI-guns with a combined volume of 16.3 liters (710 cubic inches). This data were chosen because of the wide frequency spectrum recorded. Specifically, we used one CSG from line 11, applying just an FK filter to remove the swell noise(Claerbout, 1976; Yilmaz, 2001).

After the application of a bandpass Butterworth filter between $8 - 100Hz$, we can see the actually exploited data in Figure 7.2.11.

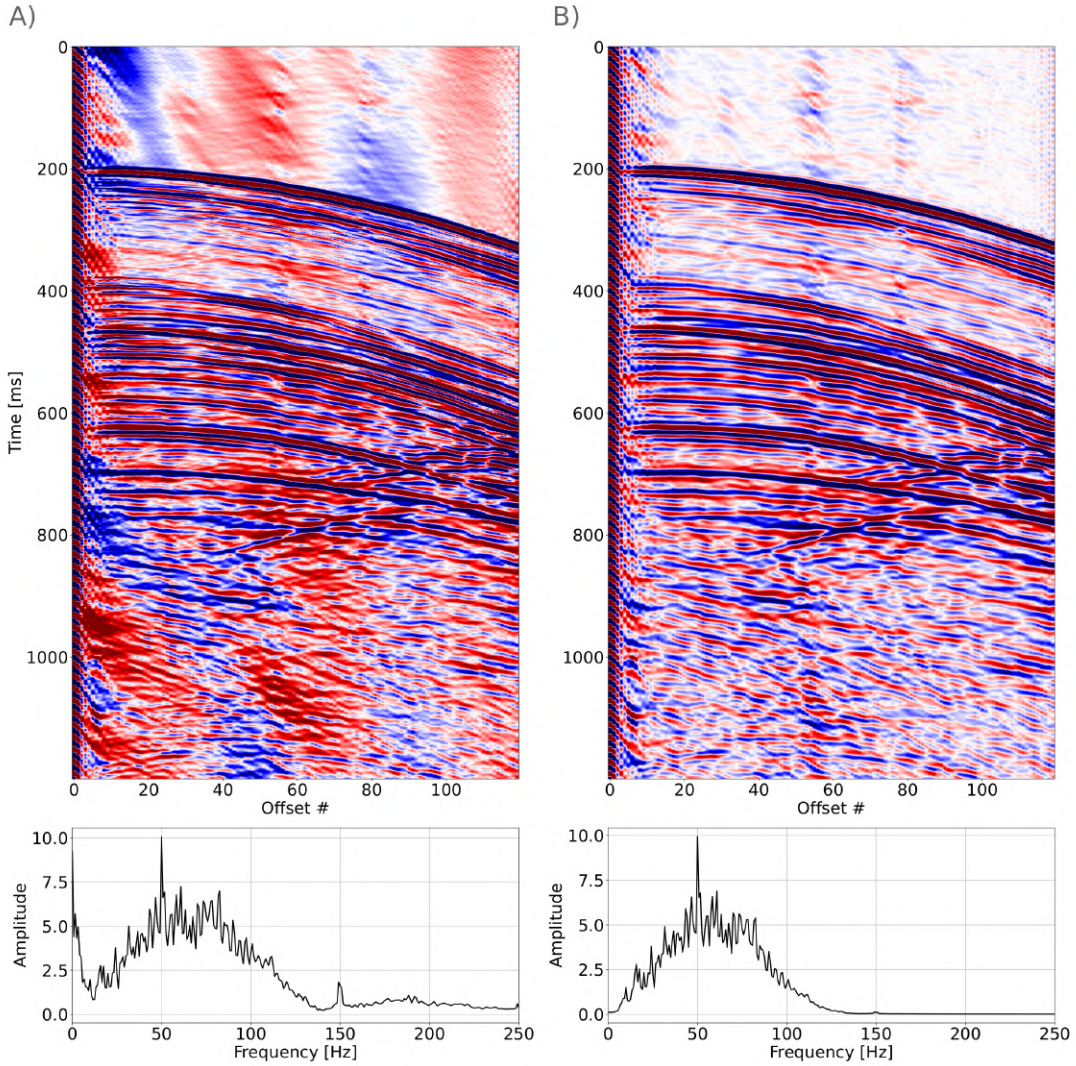

Figure 7.2.11: Sample CSG after F-K filter (A) and after application of a Butterworth filter (B).

Thanks to the wide spectrum of the data we are now able to apply two different Butterworth filters: one to retain the highest part of the spectrum, i.e. $30 - 90Hz$ (hereafter referred to as $60Hz$ data) and a lower one set between $15 - 45Hz$ (hereafter referred to as $30Hz$ data). Results of the filtering procedure are shown in Figure 7.2.12: we will now use $60Hz$ data, Figure 7.2.12-B as input to predict $30Hz$ data, Figure 7.2.12-C.

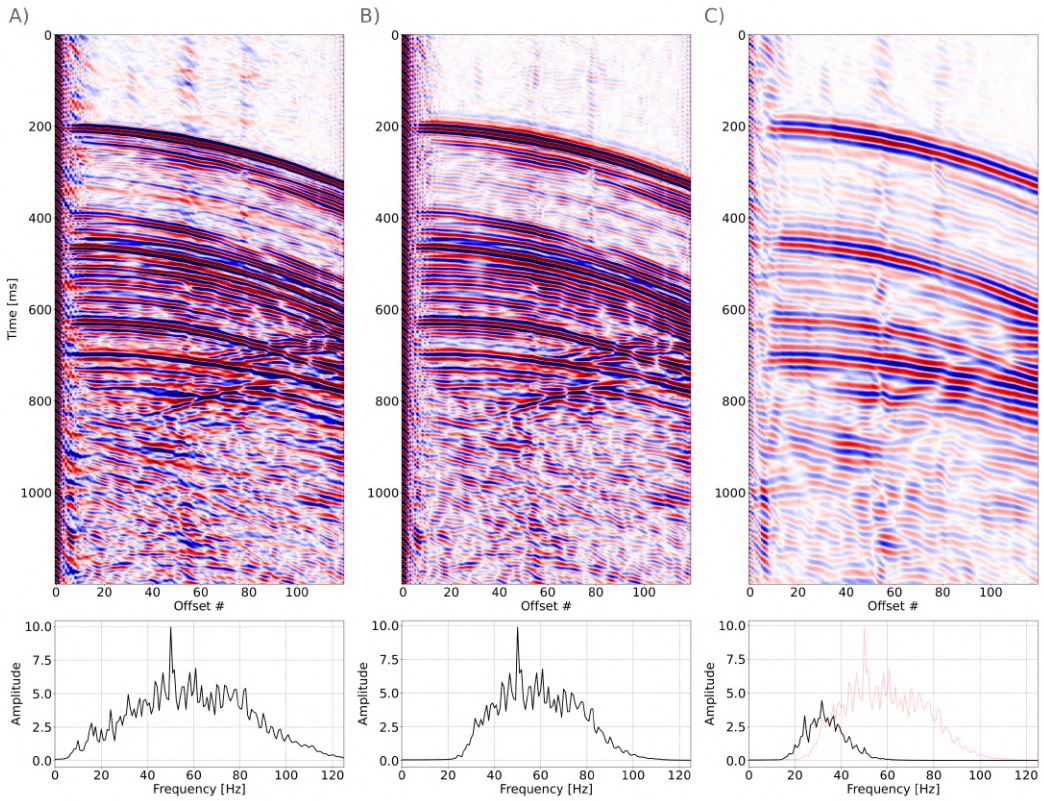

Figure 7.2.12: Filtered data (A) and Butterworth filter applied from 30Hz to 90Hz, lately referred to as 60Hz data (B) and filtered from 15Hz to 45 Hz, a.k.a. 30Hz data.

Figure 7.2.13, shows the results of the prediction: we can see how even some details like the low amplitude part, probably due to destructive interference, between $450 - 600ms$ is properly predicted.
Further attention should be given to the interference: between $600 - 700ms$ at offset $\#75 - 110$ we can see a clearly recognizable destructive interference pattern that was not present in the input, but is perfectly restored in the prediction.

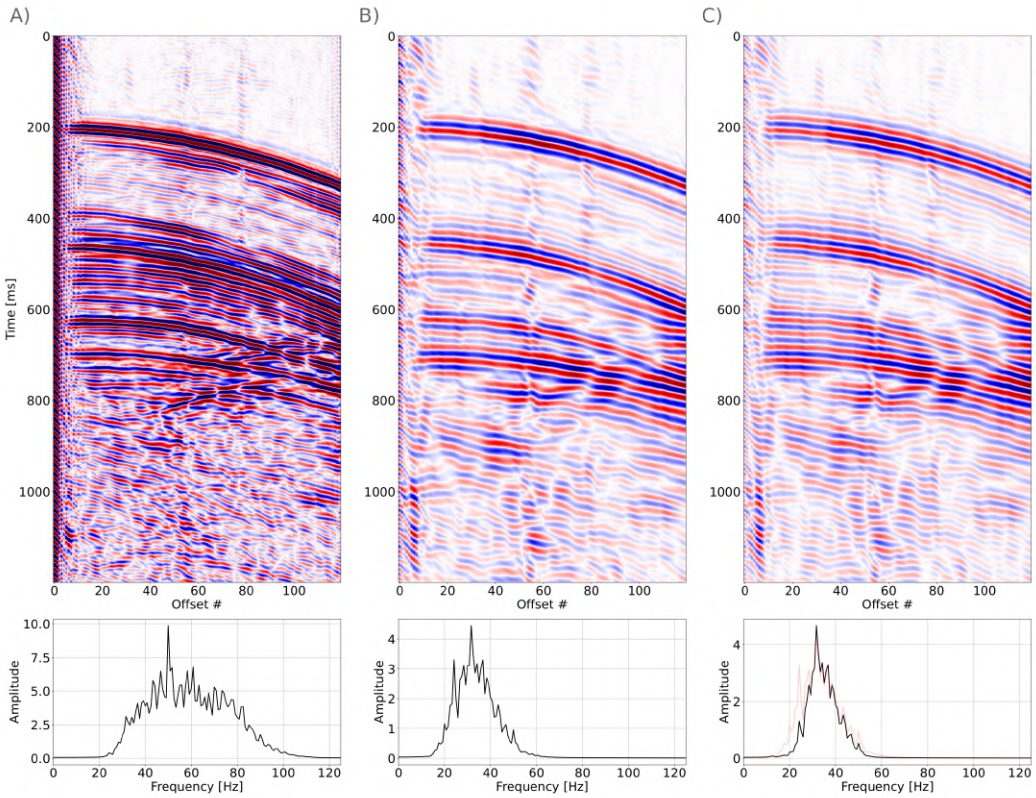

Figure 7.2.13: 60Hz data (A), i.e. the input to the NN, 30Hz data (B) and NN prediction (C).

If we look at the waveform, Figure 7.2.14, we can see that some small amplitude errors persist, but no relevant phase shifts are introduced.

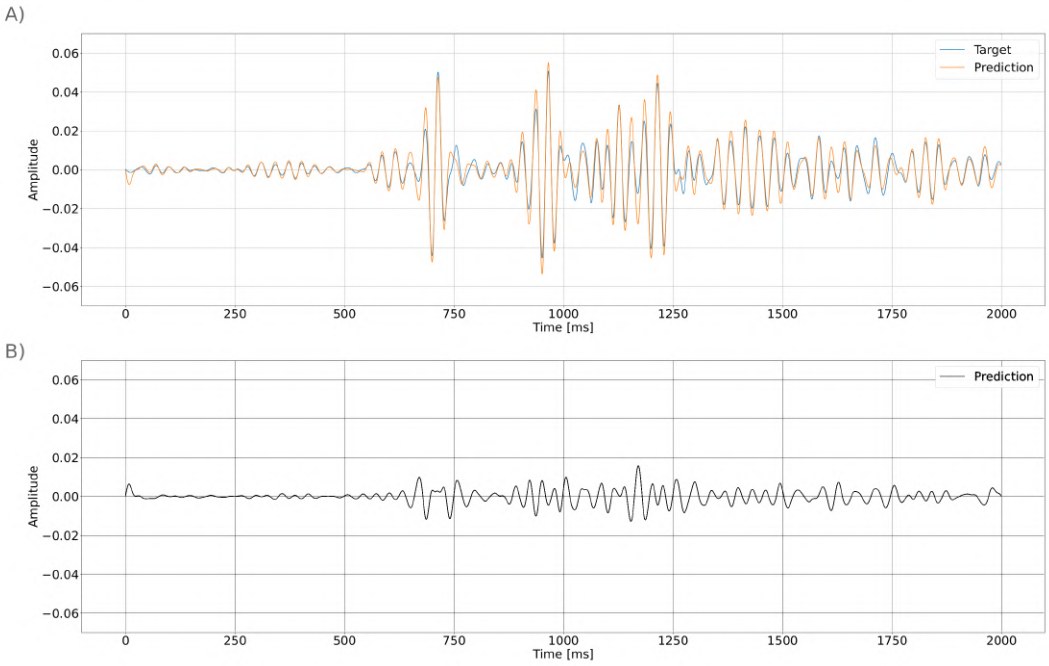

Figure 7.2.14: Single trace plot of the waveform predicted (orange) and filtered from initial data (blue) for the *30Hz data* (A) and residuals (B).

To further provide the applicability of the methodology, we can resample the $30Hz$ prediction and apply again the methodology to infer for instance the $15Hz$ version, Figure 7.2.15.

In this case, without further processing and starting from $60Hz$, the methodology introduces some noises but still predicts coherent events at their correct time position. In particular, the interference between the two reflectors at about 600 and 700ms apparent in both panels A and B is realistically predicted in panel C as a single low frequency event. As far as the spectral content of the prediction, we can see that almost all the components of the predicted spectrum was completely filtered out in the the $60Hz$ data, as previously discussed.

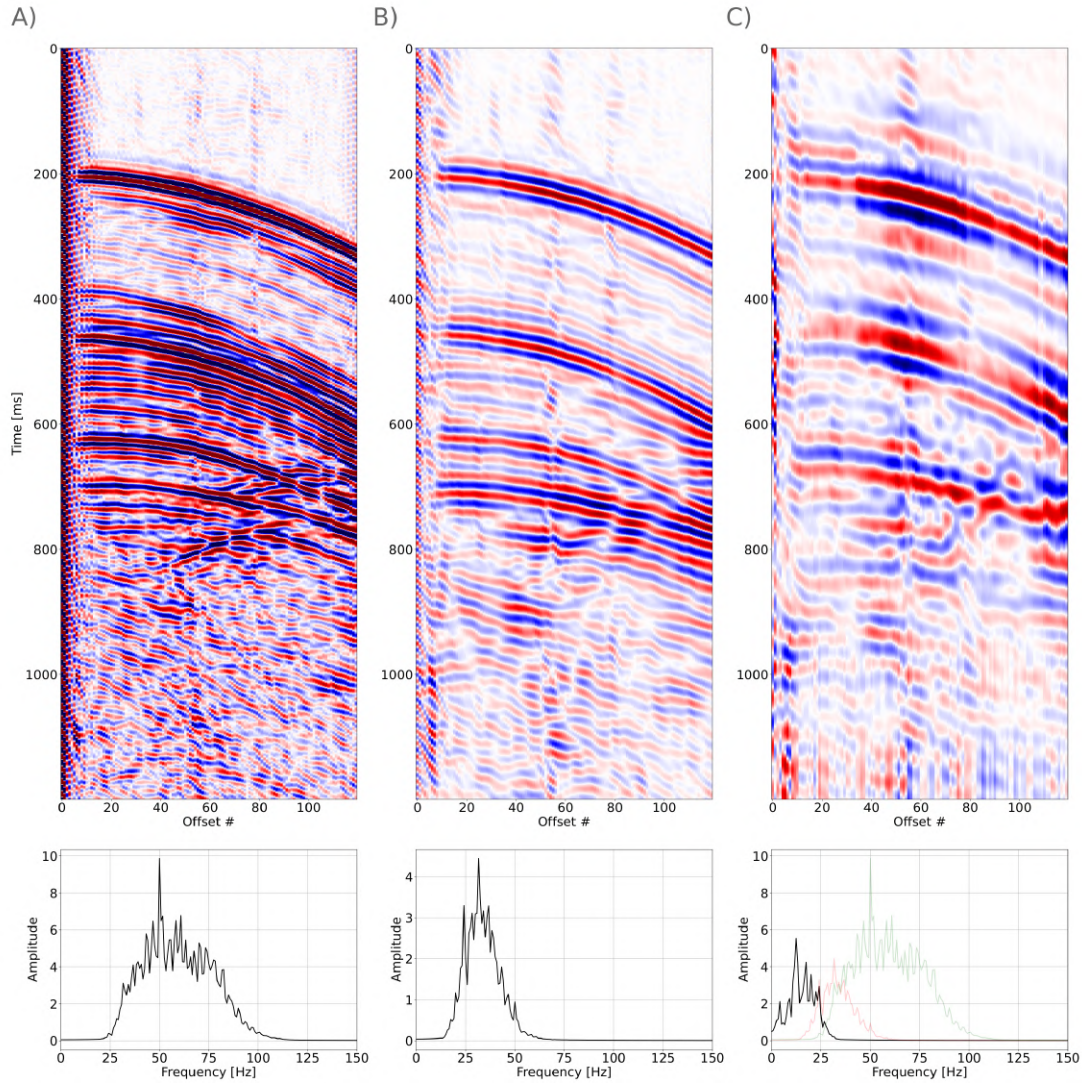

Figure 7.2.15: 60Hz data (A), i.e. the input to the NN, 30Hz data predicted (B) and 15Hz data predicted by the methodology (C).

The methodology is quite profitable, the NN can be trained on a laptop with a GPU: we used a NVIDIA RTX A4500 Laptop GPU with 16Gb of memory, with a total training time of 411 minutes for the training.
Prediction is fast, for predicting 6000 single offsets it takes just 6 seconds (for traces with 8000 time samples, each).

Bibliography

Ray Abma. 2018. Shot coding design for popcorn shooting. In *SEG Technical Program Expanded Abstracts 2018*. Society of Exploration Geophysicists, Anaheim, California, 56–60. https://doi.org/10.1190/segam2018-2995736.1

Ray Abma and Allan Ross. 2013. Popcorn shooting: Sparse inversion and the distribution of airgun array energy over time. In *SEG Technical Program Expanded Abstracts 2013*. Society of Exploration Geophysicists, 31–35. https://doi.org/10.1190/segam2013-0592.1

Mehdi Aharchaou and Anatoly Baumstein. 2020. Deep learning-based artificial bandwidth extension: Training on ultrasparse OBN to enhance towed-streamer FWI. *The Leading Edge* 39, 10 (Oct. 2020), 718–726. https://doi.org/10.1190/tle39100718.1

L. Amundsen and A. Reitan. 2014. Deghosting of Marine Pressure Recordings Along Curved Streamers. Amsterdam, Netherlands. https://doi.org/10.3997/2214-4609.20141588

Pranav Audhkhasi and Satish C. Singh. 2022. Discovery of distinct lithosphere-asthenosphere boundary and the Gutenberg discontinuity in the Atlantic Ocean. *Science Advances* 8, 24 (June 2022), eabn5404. https://doi.org/10.1126/sciadv.abn5404

Jeremy Aznar, Boris Kuvshinov, Guido Baeten, Hamish Macintyre, Julien Large, and Shuki Ronen. 2022. Successful Modelling and Sea Trial of New Low-Frequency Sources Using Standard Onboard Air Supply. *First Break* 40, 11 (Nov. 2022), 77–81. https://doi.org/10.3997/1365-2397.fb2022098

Andrew Brenders, Robert Pratt, and Sylvestre Charles. 2010. Evaluation of preconditioning strategies on waveform tomography of field data: success with long offset seismic data in a thrust-fold belt. 4348–4353. https://doi.org/10.1190/1.3513780

Andrew J Brenders. 2011. Strategies for Waveform Tomography of Long-Offset, 2-D Exploration Seismic Data. (2011).

David Carlson, Walter Söllner, Hocine Tabti, Eli Brox, and Martin Widmaier. 2007. *Increased resolution of seismic data from a dual-sensor streamer cable*. 994–998. https://doi.org/10.1190/1.2792572 arXiv:https://library.seg.org/doi/pdf/10.1190/1.2792572

Steve Chelminski, Leighton M. Watson, and Shuki Ronen. 2019. Research Note: Low-frequency pneumatic seismic sources. *Geophysical Prospecting* 67, 6 (July 2019), 1547–1556. https://doi.org/10.1111/1365-2478.12774

Yangkang Chen, Mi Zhang, Min Bai, and Wei Chen. 2019. Improving the Signal-to-Noise Ratio of Seismological Datasets by Unsupervised Machine Learning. *Seismological Research Letters* (May 2019). https://doi.org/10.1785/0220190028

J F Claerbout. 1976. Fundamentals of geophysical data processing with applications to petroleum prospecting. [Textbook]. (1 1976). https://www.osti.gov/biblio/5417212

Matteo Dossi, Emanuele Forte, and Michele Pipan. 2018. Quantitative Analysis of GPR Signals: Transmitted Wavelet, Amplitude Decay, and Sampling-Related Amplitude Distortions. *Pure and Applied Geophysics* 175, 3 (March 2018), 1103–1122. https://doi.org/10.1007/s00024-017-1752-2

John T. Etgen, Doug J. Foster, and Yu Zhang. 2016. Introduction to the special section: Subsalt imaging. *The Leading Edge* 35, 3 (March 2016), 226–227. https://doi.org/10.1190/tle35030226.1

J. F. Hopperstad, R. Laws, and E. Kragh. 2012. Hypercluster of Airguns – More Low Frequencies for the Same Quantity of Air. Copenhagen, Denmark. https://doi.org/10.3997/2214-4609.20148845

Jinwei Fang, Hui Zhou, Yunyue Li, Qingchen Zhang, Lingqian Wang, Pengyuan Sun, and Jianlei Zhang. 2020. Data-driven low-frequency signal recovery using deep-learning predictions in full-waveform inversion. *GEOPHYSICS* 85, 6 (2020), A37–A43. https://doi.org/10.1190/geo2020-0159.1 arXiv:https://doi.org/10.1190/geo2020-0159.1

Maz Farouki, Andrew Long, and Rune Tenghamn. 2010. Enhanced Resolution, Imaging and Interpretability: Dual-Sensor Towed Streamer Data Examples from Around the World. (2010).

Jo Firth, Gordon Poole, Federico Buriola, Steve McDonald, Paul Fallon, Steve Hollingworth, James Cooper, and Gaeton Mellier. 2018. Benefits of multi-sensor streamers for broadband acquisition. *First Break* 36, 11 (Nov. 2018), 65–69. https://doi.org/10.3997/1365-2397.n0133

Riccardo Geletti, Fabrizio Zgur, Anna Ben, F. Buriola, Silvana Fais, Maurizio Fedi, Emanuele Forte, Arianna Mocnik, Valeria Paoletti, Michele Pipan, Riccardo Ramella, Roberto Romeo, and Alessandro Romi. 2014. The Messinian Salinity Crisis: New seismic evidence in the West-Sardinian Margin and Eastern Sardo-Provençal basin (West Mediterranean Sea). *Marine Geology* 351 (2014), 76–90. https://doi.org/10.1016/j.margeo.2014.03.019

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Comput.* 9, 8 (11 1997), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Diederik Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations* (12 2014).

Yunyue Elita Li and Laurent Demanet. 2016. Full-waveform inversion with extrapolated low-frequency data. *GEOPHYSICS* 81, 6 (Nov. 2016), R339–R348. https://doi.org/10.1190/geo2016-0038.1

Naihao Liu, Shengtao Wei, Shengjun Li, Yang Yang, Yijie Zhang, and Jinghuai Gao. 2023. Sparse unscaled time-frequency transform and its application on seismic attenuation delineation. *GEOPHYSICS* 88, 6 (2023), B355–B368. https://doi.org/10.1190/geo2022-0522.1 arXiv:https://doi.org/10.1190/geo2022-0522.1

Z. Liu, D. Wheaton, and B. Wang. 2019. *Seismic data denoising with mathematical morphological filters.* 4619–4623. https://doi.org/10.1190/segam2019-3216392.1 arXiv:https://library.seg.org/doi/pdf/10.1190/segam2019-3216392.1

G. S. Martin. 2004. The Marmousi2 model, elastic synthetic data, and an analysis of imaging and AVO in a structurally complex environment. (2004). Master's thesis. University of Houston.

Gary S. Martin, Kurt J. Marfurt, and Shawn Larsen. 2002. Marmousi-2: An updated model for the investigation of AVO in structurally complex areas. In *SEG Technical Program Expanded Abstracts 2002*. Society of Exploration Geophysicists, 1979–1982. https://doi.org/10.1190/1.1817083

Hassan Masoomzadeh, Penny J. Barton, and Satish C. Singh. 2005. Advanced processing of long-offset seismic data for sub-basalt imaging in the Faeroe-Shetland Basin. In *SEG Technical Program*

Expanded Abstracts 2005. Society of Exploration Geophysicists, 417–420. https://doi.org/10.1190/1.2142230

G. Mellier and N. Tellier. 2018. Considerations About Multi-Sensor Solid Streamer Design. Oslo, Norway. https://doi.org/10.3997/2214-4609.201802092

Andrey K. Morozov. 2021. Underwater ultra-low frequency seismic source. *The Journal of the Acoustical Society of America* 149, 4 (April 2021), 2163–2172. https://doi.org/10.1121/10.0003953

Björn Olofsson. 2010. Marine ambient seismic noise in the frequency range 1–10 Hz. *The Leading Edge* 29, 4 (April 2010), 418–435. https://doi.org/10.1190/1.3378306

Oleg Ovcharenko, Vladimir Kazei, Mahesh Kalita, Daniel Peter, and Tariq Alkhalifah. 2019. Deep learning for low-frequency extrapolation from multioffset seismic data. *GEOPHYSICS* 84, 6 (Nov. 2019), R989–R1001. https://doi.org/10.1190/geo2018-0884.1

Giuseppe Provenzano, Timothy J. Henstock, Jonathan M. Bull, and Gaye Bayrakci. 2020. Attenuation of receiver ghosts in variable-depth streamer high-resolution seismic reflection data. *Marine Geophysical Research* 41, 2 (June 2020), 11. https://doi.org/10.1007/s11001-020-09407-9

Dennis K Reust, Otis A. Johnston, John A. Giles, and Steve Ballinger. 2015. Very low frequency seismic source. In *SEG Technical Program Expanded Abstracts 2015*. Society of Exploration Geophysicists, New Orleans, Louisiana, 221–225. https://doi.org/10.1190/segam2015-5829955.1

Enders Robinson and Sven Treitel. 2000. *Geophysical Signal Analysis*. Vol. 29. https://doi.org/10.1190/1.9781560802327

Shuki Ronen and Steve Chelminski. 2017. Tuned Pulse Source – a new low frequency seismic source. In *SEG Technical Program Expanded Abstracts 2017*. Society of Exploration Geophysicists, Houston, Texas, 6085–6088. https://doi.org/10.1190/segam2017-w16-04.1

Youyi Ruan, Donald W. Forsyth, and Samuel W. Bell. 2014. Marine sediment shear velocity structure from the ratio of displacement to pressure of Rayleigh waves at seafloor. *Journal of Geophysical Research: Solid Earth* 119, 8 (Aug. 2014), 6357–6371. https://doi.org/10.1002/2014JB011162

C.E. Shannon. 1949. Communication in the Presence of Noise. *Proceedings of the IRE* 37, 1 (jan 1949), 10–21. https://doi.org/10.1109/jrproc.1949.232969

Satish C. Singh, R. W. Hobbs, and D. B. Snyder. 1996. Broadband receiver response from dual-streamer data and applications in deep reflection seismology. *GEOPHYSICS* 61, 1 (1996), 232–243. https://doi.org/10.1190/1.1443944 arXiv:https://doi.org/10.1190/1.1443944

C.B. Smith, S. Agaian, and D. Akopian. 2008. A Wavelet-Denoising Approach Using Polynomial Threshold Operators. *IEEE Signal Processing Letters* 15 (2008), 906–909. https://doi.org/10.1109/LSP.2008.2001815

Robert Soubaras and Peter Whiting. 2011. Variable depth streamer — The new broadband acquisition system. In *SEG Technical Program Expanded Abstracts 2011*. Society of Exploration Geophysicists, 4349–4353. https://doi.org/10.1190/1.3628115

Miaomiao Sun, Zhenchun Li, Yanli Liu, Jiao Wang, and Yufei Su. 2021. Low-Frequency Expansion Approach for Seismic Data Based on Compressed Sensing in Low SNR. *Applied Sciences* 11, 11 (May 2021), 5028. https://doi.org/10.3390/app11115028

Fons Ten Kroode, Steffen Bergler, Cees Corsten, Jan Willem De Maag, Floris Strijbos, and Henk Tijhof. 2013. Broadband seismic data — The importance of low frequencies. *GEOPHYSICS* 78, 2 (March 2013), WA3–WA14. https://doi.org/10.1190/geo2012-0294.1

Sven Treitel. 1969. Predictive deconvolution-theory and practice. *Geophysics* 34, 2 (04 1969), 155–169. https://doi.org/10.1190/1.1440003 arXiv:https://pubs.geoscienceworld.org/geophysics/article-pdf/34/2/155/3154649/155.pdf

Rongrong Wang and Felix Herrmann. 2016. Frequency down extrapolation with TV norm minimization. In *SEG Technical Program Expanded Abstracts 2016*. Society of Exploration Geophysicists, Dallas, Texas, 1380–1384. https://doi.org/10.1190/segam2016-13879674.1

N. Woodburn, A. Hardwick, and T. Travis. 2011. Enhanced low frequency signal processing for sub-basalt imaging. In *SEG Technical Program Expanded Abstracts 2011*. Society of Exploration Geophysicists, 3673–3677. https://doi.org/10.1190/1.3627964

Ru-Shan Wu, Jingrui Luo, and Bangyu Wu. 2014. Seismic envelope inversion and modulation signal model. *GEOPHYSICS* 79, 3 (May 2014), WA13–WA24. https://doi.org/10.1190/geo2013-0294.1

Oz Yilmaz. 2001. *Seismic Data Analysis* (2 ed ed.). Investigations in Geophysics, Vol. 10. Society Of Exploration Geophysicists. https://doi.org/10.1190/1.9781560801580

R.W. Ziolkowski and A.D. Kipple. 2003. Application of double negative materials to increase the power radiated by electrically small antennas. *IEEE Transactions on Antennas and Propagation* 51, 10 (Oct. 2003), 2626–2640. https://doi.org/10.1109/TAP.2003.817561

Chapter 8

Conclusions

Geophysics has witnessed an increased adoption of Machine Learning (ML) techniques, particularly focused on Deep Learning (DL). This trend is driven by the contemporary surge in big data availability and improved computational power, which has enabled more advanced data analysis methods. We outline several advantages of using DL methods in geophysics including improved accuracy, the capability to handle large and complex datasets, the reduced reliance on manual intervention in traditionally labor-intensive tasks. These benefits make DL techniques, such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, highly attractive for geophysical data analysis.
LSTM networks, the focus of this work, can be a suitable DL architecture for geophysics due to their ability to handle time-dependent processes and capture temporal dependencies: these make them a good choice for studying non-stationary behaviors, working on the frequency domain and classifying events, among the others applications.

In this thesis we were focusing on different tasks and applications. In the field of XAI, the use of LSTM could be a fitting choice to establish a methodology for unsupervised information extraction from a wave-based signal. This is due to its intrinsic nature and could lead to the development of Deep Attributes (Roncoroni et al., tteda) for seismic characterization and to a way to understand NN geometry, its limitations due to the HL dimensions, and its overall performance in generalization.

The recurrent task was one of the first applications of LSTM in geophysics we explored (Roncoroni et al., 2021): it got to promising results for the waveform prediction, but it pointed out severe limitations in digesting 2-D information about the

velocity model. Other way should be investigated to fix this issue, e.g. CNN-LSTM or grid-LSTM. Although this, performance was better then with a CNN approach in therms of errors and it turns out to be a faster approximation for 1-D velocity model.

In the classification task results were very good: the approach provides a pure statistical estimation of the reflections, both in the binary task (Roncoroni et al., 2022a) and in the multi-class task (Roncoroni et al., 2022b). The ability to make trace-by-trace predictions can be exploited as a strength when dealing with noisy signals, allowing the final user to have more statistics and a solid base to further apply classical methodologies to extract horizons, or to improve data interpretation (Roncoroni et al., ttedc), as we shown in the data from the Chang'E4 Chinese mission on the moon (Roncoroni et al., ttedb).

Furthermore, this 1-D approach could also be valuable in the field of data fusion. In order to test this method we first started from RIMFAX data, a NASA GPR dataset acquired during the Mars2020 mission, and we succeeded to implement a completely new methodology for the automated and accurate merging the three separated output of the continuous modulated wavelet recorded by the GPR onboard the rover (Roncoroni et al., 2023a). A further extension to "classical" GPR data is provided as a generalization of the previous algorithm in Roncoroni et al. (2023b), with successfully applications on 1-, 2- and 3-D GPR data.
We introduced this methodology for GPR, since it is easier to acquire data with different frequencies: although we are now working to extend this methodology for seismic applications. One of the more intriguing and challenging tasks for this methodology could be its application to data recorded by geophones and by active DAS devices.

In conclusion, the frequency counterpart shows an important branch we think could be crucial in the future of DL-geophysics methodology. After defining the approach and the training procedure, as stated in Chapter 7, we can now apply it to a various number of tasks: Frequency-FWI applications, NMO de-stretching and Seismic or GPR deconvolution, among the others.

Bibliography

Giacomo Roncoroni, Emanuele Forte, Luca Bortolussi, Luca Gasperini, and Michele Pipan. 2022b. Polarity assessment of reflection seismic data: a Deep Learning approach. *BGO* 63 (Nov. 2022), 693–700. https://doi.org/10.4430/bgo00409

Giacomo Roncoroni, Emanuele Forte, Luca Bortolussi, and Michele Pipan. 2022a. Efficient extraction of seismic reflection with Deep Learning. *Computers & Geosciences* 166 (Sept. 2022), 105190. https://doi.org/10.1016/j.cageo.2022.105190

Giacomo Roncoroni, Emanuele Forte, and Michele Pipan. 2023a. Merging gated frequency-modulated continuous-wave Mars2020 RIMFAX ground-penetrating radar data. *GEOPHYSICS* 88, 2 (2023), A7–A12. https://doi.org/10.1190/geo2022-0466.1 arXiv:https://doi.org/10.1190/geo2022-0466.1

Giacomo Roncoroni, Emanuele Forte, and Michele Pipan. Under reviewa. Deep Attributes: innovative LSTM-based seismic attributes. (Under review). Submitted to GJI.

Giacomo Roncoroni, Emanuele Forte, Ilaria Santin, Ana Cernok, Andrea Rajsic, Alessandro Frigeri, and Michele Pipan. Under reviewb. High frequency Lunar Penetrating Radar quality control, editing and processing of Chang'E-4 lunar mission. (Under review). Submitted to Scientific Data.

Giacomo Roncoroni, Emanuele Forte, Ilaria Santin, Ana Cernok, Andrea Rajsic, Alessandro Frigeri, Wenke Zhao, and Michele Pipan. Under reviewc. Deep Learning driven interpretation of Chang'E4 Lunar Penetrating Radar. (Under review). Submitted to Journal of Geophysical Research: Planets.

Giacomo Roncoroni, Emanuele Forte, Ilaria Santin, and Michele Pipan. 2023b. Deep Learning based multi-frequency GPR data merging;. *GEOPHYSICS* 0, ja (2023), 1–35. https://doi.org/10.1190/geo2023-0215.1 arXiv:https://doi.org/10.1190/geo2023-0215.1

Giacomo Roncoroni, Carlo Fortini, Luca Bortolussi, Nicola Bienati, and Michele Pipan. 2021. Synthetic seismic data generation with deep learning. *Journal of Applied Geophysics* 190 (July 2021), 104347. https://doi.org/10.1016/j.jappgeo.2021.104347

Chapter 9

Side Projects

G. Roncoroni, E. Forte, I. Santin, A. Černok, A. Rajšić, A. Frigeri, M. Pipan, *High frequency Lunar Penetrating Radar quality control, editing and processing of Chang'E-4 lunar mission*, submitted to Scientific Data, Under Review.

This paper presents the quality control, editing, and processing of Lunar Penetrating Radar (LPR) data acquired by the Chang'E-4 (CE-4) lunar mission. We describe the pre-processing steps, including the removal of duplicated traces and the correction of timing errors. They also discuss the partial processing, which involves the removal of surface clutter and the application of a band-pass filter to enhance the signal-to-noise ratio. Finally, the full processing is described with an open access python code. The authors provide a comprehensive dataset that can be used for further analysis and integration with other lunar data. The ultimate goal of this work is to improve the interpretation and comprehension of lunar subsurface structures, which in some cases were not previously imaged and properly considered. We demonstrate the effectiveness of the processing techniques by presenting several examples of subsurface structures imaged using the LPR data.

Overall, this paper provides a valuable resource for researchers interested in the study of the Moon's subsurface structure and composition. The authors' processing techniques and dataset can be used to improve the interpretation and comprehension of lunar subsurface structures, and to guide the planning of future lunar missions.

I was working on the conceptualization, the computational part and the codes related to the filtering and the processing of the data.

I. Santin, **G. Roncoroni**, E. Forte, M. Pipan,*GPR modelling and inversion to quantify the debris content within ice.* Near Surface Geophysics, 2023, 1–15. doi.org/10.1002/nsg.12274.

We introduce a new methodology for inversion of GPR data to estimate the debris content within glacier ice. We discuss the methodology, sensitivity tests, and the use of a divide et impera approach for the inversion process. Furthermore, we evaluated the effect of scatterers' dimensions and rock fractions (i.e. volume of scatterers over ice) on the scattering magnitude: sensitivity tests were conducted to analyze the effect of various parameters on the scattering process, narrowing down on scatterers dimensions and rock fractions. The divide et impera approach was used to minimize a defined target function and provide an estimation of the rock fraction responsible for the high-scattering zone (HSZ) in the GPR field data. The methodology provides a completely new data-driven approach with successful application on real data. I was working on the computational part of the approach: implementing both the random input models, the forward model on HPC and the inversion part.

M. Venier, L. Ziberna, L. Mancini, A. Kao, F. Bernardini, **G. Roncoroni**, S. Milani, M. Youbi, M. Yondon, A. De Min, D. Lenaz, *Multi-scale and multi-modal imaging study of mantle xenoliths and petrological implications*, 2023, doi.org/10.2138/am-2022-8866.

This study uses advanced imaging techniques to investigate mantle xenoliths and determine a representative elementary volume (REV) that accounts for their varying properties. The research reveals that a multi-scale approach effectively defines the REV and uncovers the complex and heterogeneous nature of the samples. The authors used micro-computed tomography ($\mu - CT$) to analyze the samples at different scales. Additionally, the study emphasizes the need for high-resolution $3 - D$ reconstructions to obtain accurate modal estimates of mineral phases within the samples, addressing the limitations of $2 - D$ analysis. This work provides insights into the petrological implications and the role of mantle xenoliths in understanding the Earth's mantle and magma formation. I focused on the statistical analysis of the high-resolution $3 - D$ data to provide a modal estimation through $2 - D$ analysis of randomly cut thin sections.