Data-driven methods for the initialization of full-waveform inversion

Dissertation by

Oleg Ovcharenko

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

King Abdullah University of Science and Technology

Thuwal, Kingdom of Saudi Arabia

November, 2021

EXAMINATION COMMITTEE PAGE

The dissertation of Oleg Ovcharenko is approved by the examination committee

Committee Chairperson: Prof. Martin Mai

Committee Members: Prof. Daniel Peter, Prof. Tariq Alkhalifah, Prof. Xiangliang Zhang, Prof. Sergey Fomel

ABSTRACT

Data-driven methods for the initialization of full-waveform inversion

Oleg Ovcharenko

This thesis is dedicated to the study of methods for the initialization of seismic full-waveform inversion. Full-waveform inversion (FWI) is a non-linear optimization technique for high-resolution imaging of the subsurface. While being widely accepted in the industry it still struggles when applied to band-limited seismic data in the absence of a realistic starting model for inversion.

In this thesis, I propose three methods to improve the initialization of FWI which focus on improvements in model, data, and joint model and data domains. The first method aims to improve velocity model building in the presence of salt bodies by measuring variance between mono-frequency inversion results. High variance then indicates problematic areas where we introduce corrections and facilitate the convergence of non-linear optimization. The second method approaches the same problem of initialization of FWI, but in the data domain rather than the model domain. I extrapolate low-frequency data from its respective higher-frequency components of seismic wavefield by using deep learning. The method operates in the frequency domain and aims at mono-frequency extrapolation by a dedicated neural network. We observe that lower frequencies are easier to extrapolate due to smooth variations in the long-wavelength signal. The third method aims to jointly recover both low-wavenumber initial model and low-frequency data to enable successful elastic FWI in marine streamer data setups. This way, the reconstructed tomographic model of the subsurface compensates for the missing ultra-low frequencies in reconstructed low-frequency data. Altogether, this leads to successful elastic FWI in synthetic and field data surveys. I conclude the thesis by discussing the benefits and drawbacks of the proposed methods as well as give an outlook on future research.

ACKNOWLEDGEMENTS

What a journey! When starting my Ph.D. I would not have imagined that this endless research road would show me the world, both literally and metaphorically, as well as put me together with a number of brilliant people.

First of all, I would like to express my gratitude to those who shaped me into the researcher I have become. My Ph.D. supervisor, Prof. Daniel Peter who has always been beside me while letting me follow my interests. You gave me the vision, freedom of research and travel, so I spent these years doing what I love and never got burnt out. Prof. Tariq Alkhalifah, who backed me up from the side of exploration geophysics and who introduced me to weekly discussions in his SWAG group that together with SMI group meetings greatly contributed to the development of ideas proposed in this thesis. Dr. Vladimir Kazei, without your advice and collaboration, I cannot even imagine this adventure. And it would be boring as well.

I would like to thank my committee members for their feedback and efforts to improve this manuscript. Prof. Xiangliang Zhang, who asked important questions early in my studies, giving me enough time to incorporate answers into this thesis. Prof. Martin Mai and Prof. Sergey Fomel who were kind to dedicate their time and efforts into evaluating my work. The layout and typesetting of this manuscript are inspired by Dr. Siarhei Khirevich and Dr. Claire Birnie and their dissertations.

I owe the industrial internship experience to my supervisors. Dr. Song Hou from CGG who taught me a lot about seismic data processing. Dr. Anatoly Baumstein from ExxonMobil who showed me a research team-play in a big company and who put enormous efforts into bringing me to Texas, USA, during the pandemic.

Finally, I would like to express my gratitude to my family who has always been somewhere nearby unconditionally supporting me. My wonderful wife, Yana Ovcharenko, owes this thesis as much as I do. Without you behind me this just wouldn’t be possible.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Introduction

Hydrocarbons and ores have accompanied humankind throughout its history. From at least the Iron Age until the nineteenth century, the use of oil was limited and oil was obtained almost exclusively from surface seeps and mine workings. With the introduction of new technologies in both production and refining in the 1860s, the oil era has begun [1]. Nowadays, public attention is often focused on fossil fuels used as a source of energy. However, this is only the tip of the iceberg as products made using petroleum derivatives have become indivisible parts of our daily life. We unconsciously boost oil production when we brush our teeth with toothpaste, carry purchases from the store in plastic bags (hopefully for not much longer anymore), or go out for a run in new sneakers. This is why *the demand for oil and gas naturally grows* with the population increase on our planet.

Searching for new oil reservoirs is a costly multi-step procedure where one of the final stages is seismic imaging and inversion. The imaging aims to reveal the geological structure of the subsurface by looking primarily at acoustic waveforms recorded on the surface. Successful imaging requires numerous hours of human expertise and computational resources. Altogether, these efforts contribute to the raw cost of oil. In particular, slow turnover and inaccurate imaging of subsurface targets might infer significant financial losses due to using excessive human and computational resources, drilling in the wrong locations, or overestimating potential resources. Thus, increasing accuracy, delivering faster turnover, and reducing expert intervention into seismic imaging and inversion would increase the efficiency of oil exploration.

1.1 Background

From the variety of methods for recovering the model of the subsurface from recorded acoustic waveforms, seismic full-waveform inversion (FWI) is the modern industry standard for high-resolution imaging. This thesis covers three different methods for the data-driven initialization of FWI. Two of the proposed methods are tailored for frequency domain data manipulations while the other one works in the time domain. One method pivots on a conventional FWI engine while the other two have deep learning as their backbone. Such a diversity of definitions makes it hard to comprehend everything in a shared introduction. For this reason, below I provide a brief overview of FWI and relevant concepts from deep learning. Please refer to detailed introductions in each chapter for a deeper dive into the relevant domain for each application.

Below I explain the concept of FWI as well as cover relevant topics from deep learning. Then, I explain the evolution of ideas in the manuscript and provision the content of its chapters.

1.1.1 Full-waveform inversion

Exploration for oil starts with a seismic survey. Acoustic waves propagating in the Earth encode information about rock properties along their propagation paths and deliver these data to the surface. Surface oscillations are then recorded by receivers in the form of seismograms which are later used to reconstruct the original distribution of rock properties. The ultimate goal of seismic exploration is to build an accurate model of the subsurface which then might be used as a road map for the development of hydrocarbon reservoirs or mining fields. Methods for imaging the subsurface by seismic data evolved over time from zero-offset time migration to Full-Waveform Inversion (FWI), which is the modern industrial standard [2, 3, 4]. Despite being sharpened by decades of successful applications, FWI still requires an expert intervention which includes preparation of seismic data, building a realistic starting velocity model, and modifying the inversion objective which drives optimization. Improvements in each of these stages would facilitate the convergence of the method to a realistic model of the Earth's interior.

As the name suggests, FWI attempts to match both phase and amplitude information in recorded waveforms by exploring the space of possible subsurface realizations. The core idea of FWI is to iteratively build a model of the subsurface which would produce seismic recordings in a synthetic experiment identical to the field observations. For this purpose, synthetic and observed seismograms are matched to compute values of a misfit, or objective, functional. As a *non-linear optimization problem*, FWI might converge to a minimum of the objective functional which does not guarantee the fidelity of the resulting distribution of the model parameters. The success of FWI, as for any problem of nonlinear optimization, mainly depends on the formulation of the objective function, the initial point for the optimization, and the optimization strategy. The latter is usually implemented with gradient-based optimization schemes [5], whereas the first two parameters are subject to extensive research. When the geological target is simple and the setup is synthetic, the objective function tends to be smooth and convex, which ultimately results in quick convergence to a true parameter distribution in the model. However, this is never true in realistic scenarios where the objective function suffers from multiple local minima due to sinusoidal nature of wavefield and the extensive scattering. Meanwhile, natural noise and imperfection of seismic acquisition affect the ill-posedness of the inverse problem and limit the reachable resolution of subsurface imaging. Derivatives of the misfit functional with respect to model parameters provide a model update for each iteration. The misfit functional tends to zero as the synthetic model approaches the true distribution of geological parameters.

Methods for misfit and gradient modifications

Modifications introduced into the misfit function of FWI generally assume two groups of methods. The first group is focused on the re-formulation of the metric to calculate the difference between synthetic and observed data. The second group aims to regularize the inversion by reducing the number of local minima on the objective function. Smoother objective functions e.g., envelope or cross-correlation-based functions lead to fewer local minima and therefore FWI is more likely to converge to the global minimum of the misfit functional [6, 7, 8, 9, 10]. The headwind is that these methods are typically associated with increased computational costs. Furthermore, adding regularizations [11] and/or constraints [12] to the inversion also can lead to better convergence toward the global minimum. Examples of constrained inversions include works by [13, 14] and [15], who used a total variation norm to invert a salt-affected subsurface using FWI. Regularized inversion methods could be computationally expensive when multiple runs on a fine mesh are needed and these methods often require prior assumptions about the subsurface structure [16].

An alternative for modifications with misfit functional is to manipulate image or gradient with single iteration updates of FWI. Gradient conditioning can lead to shorter paths toward the global minimum without being trapped in the local minima [17, 18]. Image-guided inversion [19], gradient optimization, [20] or gradient conditioning through scattering angle-based filters [21, 22] can serve the same purpose.

Methods for initial model building

While research on the formulation of the objective functional and related regularizations is mostly analytical, the problem of building a realistic initial model for optimization remains an empirical task. For a geologically simple model, the starting model for FWI might be built using travel time tomography which fails to deliver a sufficiently accurate model for FWI when highly reflective salt bodies are present in the subsurface. This is due to strong reflections and multi-scattering which occur when seismic waves interact with the surface of a buried salt body. Starting inversion from a poor initial point often leads to the convergence of optimization to a local minimum that projects as cycle-skipping artifacts in the model domain.

Cycle-skipping occurs when the initial model of the subsurface model significantly deviates from the true geological distribution. Then, the algorithm mismatches wiggles of synthetic and field data which ultimately leads to localized clusters of unrealistically high or low values of target parameters. While given an accurate starting point, optimization quickly converges to a realistic distribution of geological parameters. The conventional approach for building an accurate subsurface model for FWI for a salt-induced medium is manual salt-edge picking on migrated images [23]. Numerous automated approaches for constrained and regularized inversion have been proposed since the discovery of promising oil reservoirs beneath salt bodies in the Gulf of Mexico [24, 14, 11]. Whereas the inversion could naturally converge to a true subsurface model even from a poor starting point if broadband seismic data with very low-frequency content are available. Otherwise, when there are not contrastive objects in the medium, traveltime tomography allows building an accurate kinematic model based on picked first arrivals [25].

When low-frequency data are available in the dataset the multiscale inversion strategy [26] might compensate for the poor initial model. This multiscale inversion strategy assumes a successive series of inversions of band-limited subsets of seismic data. In the frequency domain, this is equivalent to running a number of mono-frequency inversions to gradually produce the subsurface model with fine details recovered at higher frequencies. In the time domain, entire bands rather than individual frequencies are inverted at each stage of the multiscale strategy. Thus, the lowest available frequency in the dataset defines the scale of the largest structures recovered by inversion at early iterations of the multiscale workflow.

The key takeaway from this subsection is that available high-fidelity initial model or low-frequency content present in the data are major contributors to the success of FWI application. Meanwhile, the initial model and low-frequency data are interchangeable when considered in the framework of FWI. When a robust initial model is available, inversion might be started from a higher frequency. Alternatively, when low-frequency data are accessible, the inversion might be initiated from a poor initial guess.

1.1.2 Deep learning

The transition to the information age in the 21st century was also marked by the exponential growth of amounts of data generated by humanity. The data-rich environment, in turn, inevitably leads to the development of methods which would "learn" dependencies in the data to address a given task.

Deep learning (DL) is a subset within the science of Artificial Intelligence primarily attributed to the usage of deep neural networks [27]. The core idea of DL is built around algorithms that modify themselves to produce the desired output while given a set of structured or unstructured data. An artificial neuron is considered a principal building block for deep learning models. Each neuron performs a non-linear operation on the input data and the produced output then becomes an input to the next neuron or to many neurons in a chain-like architecture. There are more sophisticated elements within the DL toolbox but the intention here is to give a bird-view on structures relevant for this manuscript. Typically, such artificial neurons are arranged into layers which then shape multi-layer architectures of artificial neural networks (NN). The standard NN consists of several layers, each of which creates a new representation of the source data. This way, the source data are inherently decomposed into low- and high-level representations which are then assembled together in a non-linear way to produce the target output. The mismatch between produced and target outputs is then used to adjust the parameters of the model. Over numerous iterations, the NN is able to learn any non-linear function which makes it a universal approximator [28].

Unsupervised learning is the paradigm of training neural networks on unlabeled data. Meaning that the algorithm attempts to identify clusters and dependencies in the input data where the target is either unknown or ambiguous. Unsupervised methods enable the truly data-driven way for retrieval of the desired properties from available seismic data. While not used in this manuscript I strongly believe that unsupervised solutions for the problems of low-frequency extrapolation and velocity model building are somewhere nearby and we will find them soon. However, in this work, I focus on supervised methods to address the same problems. Meanwhile, existing unsupervised solutions showed promising results in horizon tracking and segmentation [29, 30], facies classification [31, 32], first-arrival picking and event detection, [33, 34, 35] and others. Notice, that these applications assume that sufficient information is already available in the input data. So the problem converges to extracting this hidden information from the inputs. On the contrary, there are such problems where the sought-for quantities are not present in the input data due to the inability to accurately collect those target properties in the field. For example, low-frequencies < 3 Hz are challenging to acquire in the field as well as directly "measuring" the geological structure of deep subsurface is impossible. While possible to be addressed in an unsupervised fashion (e.g. within the generative adversarial framework [36, 37]), these problems with limited access to real-world targets still rely on supervised methods and synthetic datasets.

Supervised learning assumes the availability of a target pair for each given input data sample. The NN then learns patterns by being fed with multiple pairs of available inputs and desirable outputs. This is the to-go paradigm when a relevant dataset is available. Such supervised methods are also actively developed for facies classification

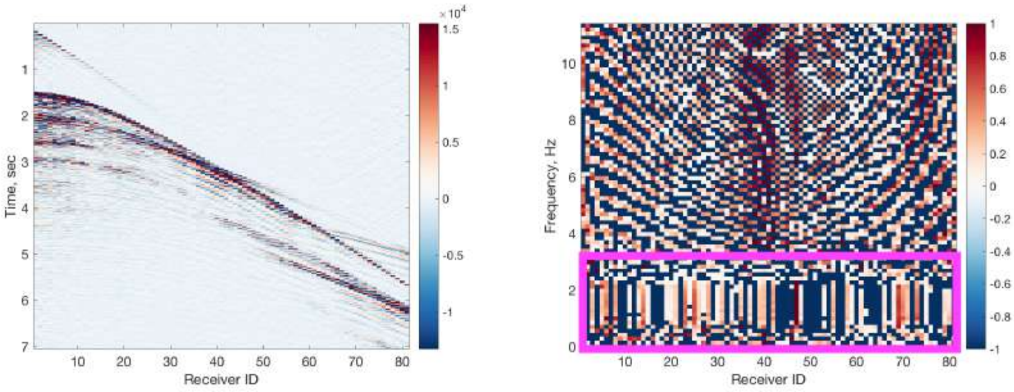
[38, 39, 40], fault mapping [41, 42], first-arrival picking [43, 44], source mechanism inversions [45], and various kinds of seismic inversions [46, 47, 48, 39, 49, 50, 51]. However, real-world applications of deep learning for exploration geophysics struggle from limited available datasets for supervised training. The reason is that a typical problem of interest requires solving an inverse problem that is a costly and labor-intensive task. For example, such regression tasks as velocity model building directly from field data or low-frequency extrapolation from high-frequency seismic data require running inversion for the subsurface and costly low-frequency acquisition to collect target data for supervised training. The affordable workaround, however, is to generate synthetic data which would be sufficient to enable the knowledge transfer between the synthetic and real-world domains. In this thesis, all deep learning applications are tailored within the supervised learning paradigm where training datasets are generated in synthetic setups.

Methods for reconstruction of low-frequency seismic data

Low frequencies present in seismic data are a valuable asset in multiple applications [52]. The long-wavelength data scatter less and penetrate deeper into the subsurface. It also features fewer minima during non-linear optimization and increases the resolution of wavelet-involving operations. Also, the wider bandwidth of seismic data improves the decoupling of parameters in multiparameter FWI [53, 54].

The desired range of low-frequencies is limited by traveltime tomography from one side and the machinery of seismic acquisition from another. In particular, the background velocity model derived from tomography compensates for frequencies below 1 Hz [55]. While the frequency content recorded in a generic marine airgun survey drops below noise level at about 4 Hz [56]. The frequencies falling in-between these estimates correspond to the gap in the model wavenumber spectrum [57]. Reconstruction of these low frequencies is the general objective of the proposed low-frequency extrapolation methods.

Acquiring low-frequency data is a costly venture because it requires low-frequency sources [58] and sensitive receivers [59]. Nevertheless, the state-of-the-art seismic surveys allow recording such data in the field environment. The ultra-long offset seabed acquisitions employing an array of ocean-bottom nodes are capable of registering frequencies as low as 1.5 Hz [60, 61]. A similar range of ultra-low frequencies might be detected in the marine buoy survey setup [62] where hydrophones are floating under the water surface attached to a buoy. The advanced marine streamer survey [63] might record the robust signal with frequency content down to 2.5 Hz. When low frequencies, e.g. below 1 Hz, are present in the data, the FWI is likely to converge even from a homogeneous prior. Unfortunately, these data are rarely available in production datasets due to hardware limitations of seismic sources and receivers or due to the high costs of an ultra-broadband acquisition (Figure 1.1). The data sample shown in Figure 1.1 is acquired during marine streamer survey by CGG [63].



(a) (b)

Figure 1.1: Example of a common-shot gather in time (a) and frequency (b) domains. Low-frequency part of data is missing (magenta box).

Deep learning is a young topic in geophysics compared to physics-based methods that sometimes date back centuries ago. For this reason, there is a wide scope of new research directions. Over the past decade, the research community proposed a variety of methods aiming at data-driven bandwidth extrapolation of seismic data. [64] introduced a beat-tone inversion to extract low frequencies from the interference of waveforms at neighboring frequencies. [65] considered a seismogram as a modulated signal and used a de-modulation operator to extract the low-frequency envelope. [66] analytically extrapolated the low frequencies by decomposing selected seismic records into elementary events and taking into account inter-trace relations. [67] addressed frequency extrapolation as a convex optimization problem with a total-variation regularization that accounts for the spatial correlation between traces. Discussions on bandwidth extrapolation in data-space and image-space approaches are provided in [68], who extrapolated frequency bandwidth of the data by wavenumber extrapolation of the extended images with subsequent extended Born modeling.

The feasibility of frequency bandwidth extrapolation was discussed from the wavenumber illumination point of view [69] and by the sparse nature of seismic signal [70]. The feasibility of trace-to-trace extrapolation in time domain data was explained as a by-product of super resolution [71, 72]. A frequency-domain approach for data-driven bandwidth extrapolation was proposed by [73]. This work considers multi-offset seismic data (shot gathers) in the frequency domain where each frequency might be represented as a complex-valued vector. For such a data view, authors designed a non-linear approximator (deep neural network) to extrapolate single low-frequency data from observed multiple high-frequency data. In general, the authors showed that smooth mono-frequency data at multiple offsets is a suitable target for shot-to-shot low-frequency extrapolation with the subsequent use at acoustic FWI [74, 69]. Meanwhile, [75, 76] built a CNN for trace-by-trace extrapolation and applied it to time-domain data. Another approach has been proposed by [77], who demonstrated the extrapolation by jointly processing beat-tone and raw waveform data using a deep CNN based on the Inception neural network [78]. An example of low-frequency extrapolation approached by the knowledge transfer from training on real-world OBN data to towed-streamer data was shown by [79]. I provide more detailed breakdown of deep learning methods in relevant chapters of this manuscript.

Methods for reconstruction of the low-wavenumber velocity model

A robust initial velocity model is crucial for FWI convergence when low-frequency data is not available. Deep learning methods typically focus on end-to-end pipelines for velocity model building from seismic data since no other framework yet offers similar capabilities. Specifically, raw or processed seismic data goes as input while the target is the distribution of model parameters. An ideal pipeline then would eliminate the need for running costly FWI. Despite promising concepts, these methods are not yet capable of replacing the traditional inversion pipeline.

Tomography typically assumes a lower resolution of the constructed model of the subsurface than FWI. Reaching the accuracy of FWI, by a learning-based method, in recovering the geological structure is an ambitious goal while not yet been achieved. Deep learning tomography was proposed [46] and further developed in [80, 81]. Authors trained deep learning models in synthetic environments to infer subsurface structure from sparse survey data. Other approaches operating on datasets sampled from and representing the entire survey were explored in [82, 83, 84, 85, 50]. [86] developed a hybrid approach where two networks are trained simultaneously to ensure minimization of the model related functional while meeting wave-field physics constraint. Alternatively, gradients derived from early iterations of poorly initialized FWI might also be used for initial model building [87]. The headwind of this approach is primarily due to the high computational costs of building the training dataset. Generation of each training sample is such a setup requires running a few iterations of FWI.

The major challenge for these methods is their dependency on the geometry of specific seismic acquisition. Meaning that the relative position of shot and receiver locations constraints the generalization capability of these methods on other surveys. An alternative to this approach would be using a limited-aperture design of input data that would lift the exact survey geometry requirement but still demand identical spacing and source-receiver configuration within the aperture. Assuming that the data from the limited aperture is sufficient to illuminate the local subsurface, [88] proposed to map a set of neighboring common-midpoint-gathers into the central vertical elastic profile. This approach manipulates shot-gather data rather than full-survey data that improve its applicability to a broader range of domains.

To sum up, the geophysical industry accumulated tremendous amounts of seismic data during decades of oil exploration. The data-driven framework, where the new knowledge is extracted either directly from data or its derivatives, offers promising opportunities for faster processing of seismic data with fewer assumptions made on the physics of wave phenomena. In the next section, I explain how ideas evolved from my first encounter with seismic inversion toward the ideas proposed in this thesis.

1.2 Evolution of ideas

My first encounter with FWI was for educational purposes and in application to a geological domain with a buried salt body. In particular, I had to run inversion in the frequency domain and experience how fragile and difficult is the process when started from a naive assumption about the underlying subsurface. It is known that severe cycle-skipping artifacts are commonly observed in the regions with massive salt bodies present in the subsurface due to large mismatches between high-velocity salt bodies and smooth background velocity model. This was the starting point where I observed that a frequency-domain FWI initialized from a linear velocity model persistently fails with a typical pattern of periodic low-velocity spots in the model domain. These spots might be associated with a local minimum of the objective function for that particular frequency. The FWI kept failing at each of sufficiently-high mono-frequencies, however, the pattern of cycle-skipped velocity anomalies was migrating for each of them. Alternatively, when started from a realistic initial model, the inversion converged well for all frequencies. This observation led to an assumption that cycle-skipping artifacts shifted against their counterparts from other frequencies might indicate unresolved regions of the subsurface. By iteratively identifying and correcting areas with high variance between mono-frequency updates, the inversion converged to a satisfactory result.

Another observation from the frequency domain was related to the data, whilst the previous case focused on the model. This observation was driven by the shape of mono-frequency common-shot-gather in this domain, which is a complex-valued vector where each value describes an amplitude at a receiver. In particular, the data at lower frequencies features long-wavelength oscillations rather than those at higher frequencies. This leads the lower frequencies to be generally more simple and suitable for extrapolation. At the same time, I also got intrigued by deep learning and decided to implement it as a mapping between a set of available high frequencies and a single unknown low frequency. This would assume independence between individual low-frequencies and a rather simple smooth target. In terms of deep learning, I wanted to translate a 2-channel image (real and imaginary parts for multiple frequencies) into a 2-channel vector. This approach showed promise in synthetic setup, however, I decided to switch to time-domain since this is a natural domain for seismic data and frequency-domain methods are widely used in real-world applications.

Aiming for a field data application, I spent some time setting up the time-domain FWI for streamer survey and put my hands on real marine data. At the same time, I switched to low-frequency extrapolation in the time domain where the entire shot gathers from available high frequencies was mapped in its low-frequency counterpart. The major challenge in this domain is the amplitude variability within the target low-frequency range. In particular, amplitudes of data filtered below 3 Hz are almost negligible compared to those from below 5 Hz. Accounting for these weak amplitudes ended up being the key task in this problem formulation. In terms of deep learning, the application was equivalent to image-to-image translation. Exploring the importance of long-wavelength signals, I found that inaccurately extrapolated data at such tomographic frequencies($<$ 2 Hz) might make FWI fail even worse, compared to a simple linear initial model. However, reaching superior accuracy in bandwidth extrapolation alone would require impractical tuning of synthetic dataset generation. I had to modify the initial idea and explored other ideas such as using various combinations of frequency bands as inputs and targets to the network, using a generative adversarial framework, and others. Since marine surveys feature regular geometry for each shot gather, I decided to set the velocity model underlying each shot location as an auxiliary target. The motivation was to constraint the solution space of the network by multi-task formulation of the objective function for training. Another contribution I had to make was the generation of synthetic training dataset which would be suitable for inference on field data. For this purpose, I made a wrapper over the existing numerical wave propagation engine and used it to simulate waveforms in a set of random subsurface initializations using the source wavelet and survey design extracted from field data. While training data was sufficient, the joint model and data reconstruction were still unstable. The solution I found was to introduce another term into the objective function of neural network training that was measuring the trace-wise correlation between prediction and target. Effectively, this compensated for geometrical spreading and other amplitude-sensitive differences by comparing normalized traces. Altogether, the multi-task objective aiming to recover both smooth initial model for inversion as well as low-frequency data delivered generalizability sufficient to transfer knowledge learned from synthetic data to field data application.

1.3 Thesis layout

There are three principal parts of the thesis each explaining a method to improve the initialization of full-waveform inversion. The first part explains a method that acts in the frequency domain and focuses exclusively on recovering an accurate initial model for inversion. The second part is dedicated to a method designed for the reconstruction of missing low-frequency data by a set of weak deep learning models. The last part combines the objectives of the first two methods and describes a method that aims to jointly reconstruct time-domain low-frequency seismic data and a smooth background model sufficient to enable elastic FWI on marine field data.

The detailed breakdown of chapters is as follows:

Chapter 2 introduces a variance-based method for building an initial model for FWI for scenarios when the subsurface contains a salt body. The proposed iterative approach manipulates mono-frequency updates from frequency-domain FWI and identifies areas prone to cycle-skipping. Then it corrects these areas by averaging from the robustly inverted surroundings. The procedure of detection and correction of the corrupted parts of the subsurface continues until mono-frequency inversions act coherently well.

Chapter 3 explains a deep learning method for the extrapolation of missing low-frequency content in seismic data. Specifically, the proposed approach operates mono-frequency representations of common-shot-gathers aiming to extrapolate one frequency at a time domain. In practice, it is equivalent to mapping a complex-valued matrix of available high-frequency data into a vector for single target mono-frequency data at each receiver position. Moreover, we explain the feasibility of low-frequency extrapolation from a wavenumber illumination perspective.

Chapter 4 describes a multi-task learning approach for joint recovery of the low-wavenumber model of the subsurface and low-frequency seismic data. The key idea is to enable time-domain FWI when neither the initial model for optimization nor low-frequency data are available. We show that the recovered background velocity model effectively compensates for inaccuracies in extrapolated ultra-low-frequency data that ultimately result in successful inversion in both synthetic and field data setups.

Chapter 5 summarizes contributions reported in the thesis. I also discuss the benefits and challenges of time and frequency domain formulations when projected onto the task of low-frequency extrapolation. In the end, I overview potential directions of future research development.

Chapter 2

Variance-based model interpolation for improved full-waveform inversion in the presence of salt bodies

When present in the subsurface salt bodies impact the complexity of wave-equation-based seismic imaging techniques. Typically, the Born approximation used in every iteration of least-squares-based inversions is incapable of handling the sharp, high-contrast boundaries of salt bodies. This chapter introduces a method for building initial model for FWI when applied in an environment with the presence of salt bodies. We introduce the idea of measuring variance between different mono-frequency FWI updates to identify the problematic regions in the initial subsurface model. The areas with high variance between several mono-frequency FWI updates highlight areas that are likely to be corrupted by cycle-skipping artifacts. We then correct those areas to improve the initial point for inversion. The procedure repeats interactively until the coherency between mono-frequency FWI updates improves.

Results presented in this chapter are adaptation of the paper published in Geophysics Journal.

Ovcharenko, Oleg, Vladimir Kazei, Daniel Peter, and Tariq Alkhalifah. "Variance-based model interpolation for improved full-waveform inversion in the presence of salt bodies." *Geophysics* 83, no. 5 (2018): R541-R551.

2.1 Introduction

Challenges in seismic imaging created by the presence of salt are primarily from the high acoustic impedance contrast and the sharp and steep boundaries between a salt body and its sedimentary environment. Salt is a mechanically weak and light rock, it thus “flows” along the weakest conduit when stress is applied [24]. Usually the direction of this flow is toward the surface where the salt later forms large salt canopies surrounded by relatively young sediments. Complex salt structures with sharp and steep flanks with high-velocity areas overriding the slower ones cause illumination issues when imaging salt-affected targets. Most of the propagating energy is reflected back from the surface of the salt, which results in insufficient illumination of the areas beneath the salt body. Successful salt and sub-salt imaging usually requires the integration of geologic knowledge with geophysical data [23] as well as broadband, long-offset and multiazimuth acquisition [14].

The objective of FWI is to minimize both amplitude and phase differences between observed and modeled seismic data. For successful inversion the Born approximation requires the initial velocity model to deliver mismatches in travel times less than half the period [89]. Larger errors cause cycle-skipping artifacts on the wavepaths specific to each source-receiver pair. These artifacts in the image domain can be identified as repeated contrast velocity anomalies that, in turn, lead the inversion to convergence toward a local minimum on an objective function [90]. Inversion of mono-frequency data allows features to be resolved at a specific scale. The inversion results from different mono-frequencies therefore do not match exactly.

When mono-frequency data are modeled, the phase of a harmonic plane wave at a given point on a wavepath depends on the frequency and travel time. Hence, at different frequencies, the phases of the waves propagating through the same wavepaths differ. While minimizing the misfit between observed and modeled data, FWI amends the velocity model to match the phase shift with its nearest zero crossing at a multiple of 2π. This phase shift varies at each frequency, causing the data to be fitted differently. These shifts cause cycle-skipping artifacts to vary from frequency to frequency for a particular wave.

Here, we extend the group of gradient-based methods to improve FWI in the presence of salt bodies. In particular, we propose manipulating updates acquired in parallel from several FWI iterations at different mono-frequencies rather than just at gradients. We identify problematic regions based on the variance distribution between these updates. Then, following [91], we iteratively introduce corrections into these problematic regions in the model, leading FWI to better convergence. As a result, the proposed technique provides a more robust initial velocity model for conventional FWI.

In the following, we first discuss the general features of cost functions. We then present a method for variance-based salt body reconstruction using FWI. Finally we demonstrate the applicability of our method using crops from the left and central parts of the BP 2004 velocity benchmark model [92] and using a cross section of the SEAM Phase I model [93].

2.2 Theory

Full-waveform inversion is a technique that performs high-resolution subsurface imaging by minimizing the difference between modeled and observed seismic data. Theoretically, FWI performs a search in the model space to find a subsurface velocity model that provides a minimum value of the so-called misfit or objective function, J [3, 90]. The most familiar and simple cost function is defined by the squared L_2 norm of the differences between observed, $\mathbf{d}_0$, and generated, $\mathbf{d}$, seismic data for model $\mathbf{m}$:

$$J\left(\mathbf{m}\right) = ||\mathbf{d}(\mathbf{m}) - \mathbf{d}_0||_2^2. \tag{2.1}$$

The function, J, is generally nonlinear [94] and non-convex with multiple local minima [95]. Gradient-based optimizations therefore pose a major challenge due to the fact that a search could prematurely stop in one of minima (Figure 2.1) and cause selection of an improper velocity model.

A popular technique to improve data fit delivered by FWI is the multiscale approach proposed by [26]. The multiscale approach implies successive inversions of low-pass-

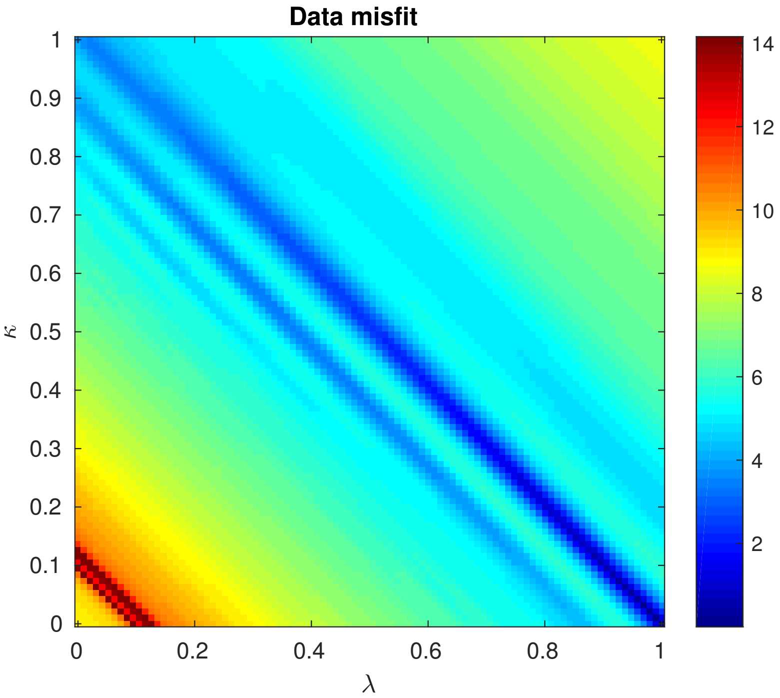


Figure 2.1: Data misfit of the linear velocity model approaching the true central part of the BP 2004 model as $\kappa v_0 + \lambda v$. Dark zones are the minima where inversion could stall prior to reaching the proper result.

filtered data, $\mathbf{d}_{0,\omega}$, from lower to higher frequencies, ω:

$$J\left(\mathbf{m}\right) \approx \sum_{\omega} \tilde{J}\left(\mathbf{m},\omega\right) = \sum_{\omega} ||\mathbf{d}(\mathbf{m},\omega) - \mathbf{d}_0(\omega)||_2^2 \tag{2.2}$$

Cost functions, $\tilde{J}$, in inversions of low-frequency data tend to have local minima that are farther apart than cost functions at higher frequencies, as schematically illustrated in Figure 2.2. This allows an optimization technique to converge to a lower minimum of J and, thus, to recover an improved low-wavenumber model. Later, this low-wavenumber subsurface image is used as a starting model for FWI at higher frequencies to recover high-resolution features of the model. However, in the absence of low frequencies this approach is less reliable for complex velocity models. Even in the presence of low frequencies, the complexity of the model may require an impractical number of FWI iterations.

Within the multiscale approach, it is common practice to simultaneously invert for several neighboring frequencies instead of a single one [96, 97]. According to this practice, the whole range of frequencies is split into groups (bands) and then gradients from different frequencies in the same band are summed at each iteration. This provides

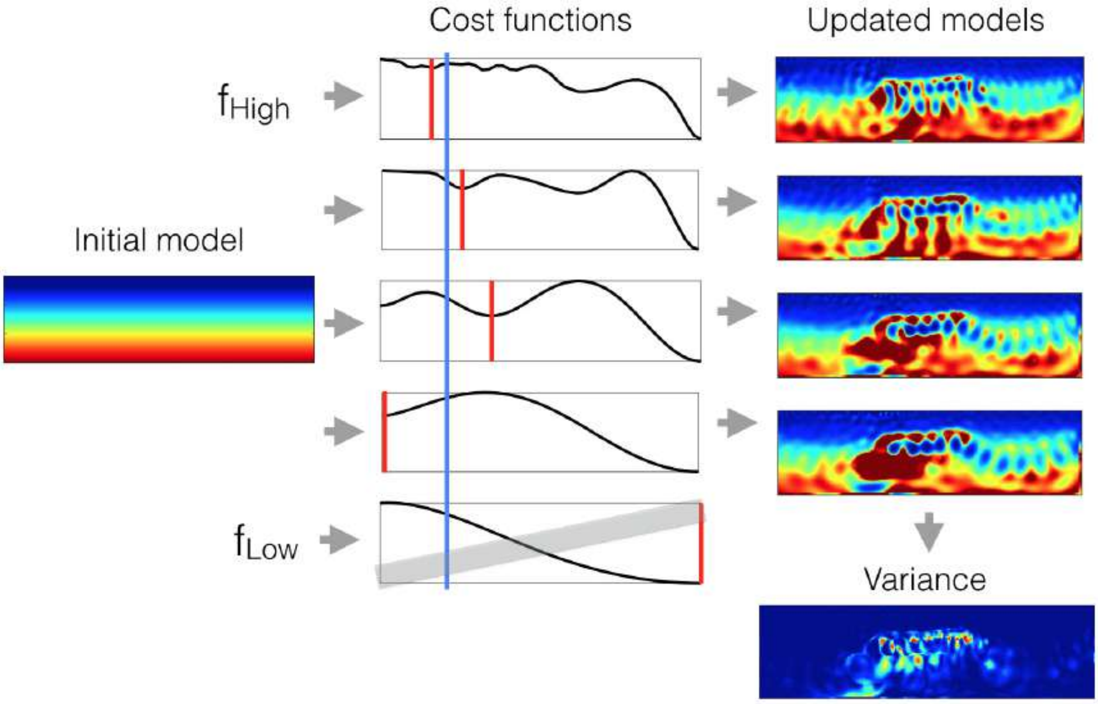


Figure 2.2: Schematic projection of local minima of the objective function onto the model domain: Given an initial model (left), objective functions, $\tilde{J}$, at different frequencies (central column) will guide inversions from the initial model (indicated by the blue line) to the closest local minimum (red line), where the inversion stops. Corresponding final velocity models and variance distribution between them are plotted in the images in the right column.

an enhanced signal-to-noise ratio in the gradient, thus improving the convergence when dealing with models that generate complex wave phenomena.

A lack of low-frequency data or of a decent initial model might result in synthetic data that are greater than a half-cycle away from observed data [18]. This leads inversions to converge to a local minimum, which in the model domain could be associated with the presence of cycle-skipping artifacts [90]. Long-offset data that illuminates deep parts of the model are generally the most vulnerable to cycle-skipping issues [95]. This happens due to larger phase shifts accumulated when seismic energy follows longer wavepaths. Deep targets are therefore more likely to be inverted erroneously.

In salt-body models, a significant part of the wave energy is reflected back from the top of the salt without propagating in the salt body. This reflection contributes to the finding that the top of the salt is well resolved in most FWI cases whereas the sub-salt areas become corrupted. Another scenario occurs when the margins of the salt body are properly identified from reflections but the inner content of the salt cannot be uniquely resolved. Corresponding objective functions of inversions of salt-body models at different single frequencies differ significantly in the same model. The inversion could prematurely stall in a local minimum of the objective function rather than searching for the best model. An important observation is that, in the synthetic case, the global minimum remains the same for misfit functions at all frequencies, whereas the local minima are different. Whenever the data are corrupted by noise the global minimum can shift when the frequency is changed, but variations in the local minima will, most likely, be larger. When the L_2 norm of the noise is higher than half of the difference between the misfit functional values at global and local minima, several global minima can potentially exist for a single misfit functional. Stacking several neighboring frequencies helps in such a case [96].

In our approach, we exploit the invariance of the global minimum for a selection of carefully determined frequencies. Consider a set of successive mono-frequency data, starting from 3 Hz which is feasible to be collected in the field with good signal-to-noise ratio [59]. Without very low-frequency data, such as below 2 Hz, independent inversions at each monofrequency will stop at their own local minimum in the vicinity of a point on the objective function corresponding to the initial guess, as is schematically shown in Figure 2.2. These minima are different due to the lack of global convexity of standard FWI functionals. The mismatch in the stopping points in the objective function of different frequencies indicates the convergence of FWI to different subsurface models. The key concept of our proposed approach is the measurement of mismatches between corresponding model updates. To this end, we use the variance between velocities at every point in the models.

We assume that the presence of cycle-skipping artifacts in the image domain indicates a local minimum of the objective function. We compare model updates computed in parallel for different single frequencies, using a variance criterion to identify regions in the image where the optimization has stalled with different model results. For salt-body reconstructions, we correct selected variable areas using the proposed “flooding" technique in the next section and repeat the process to obtain an improved initial model for FWI.

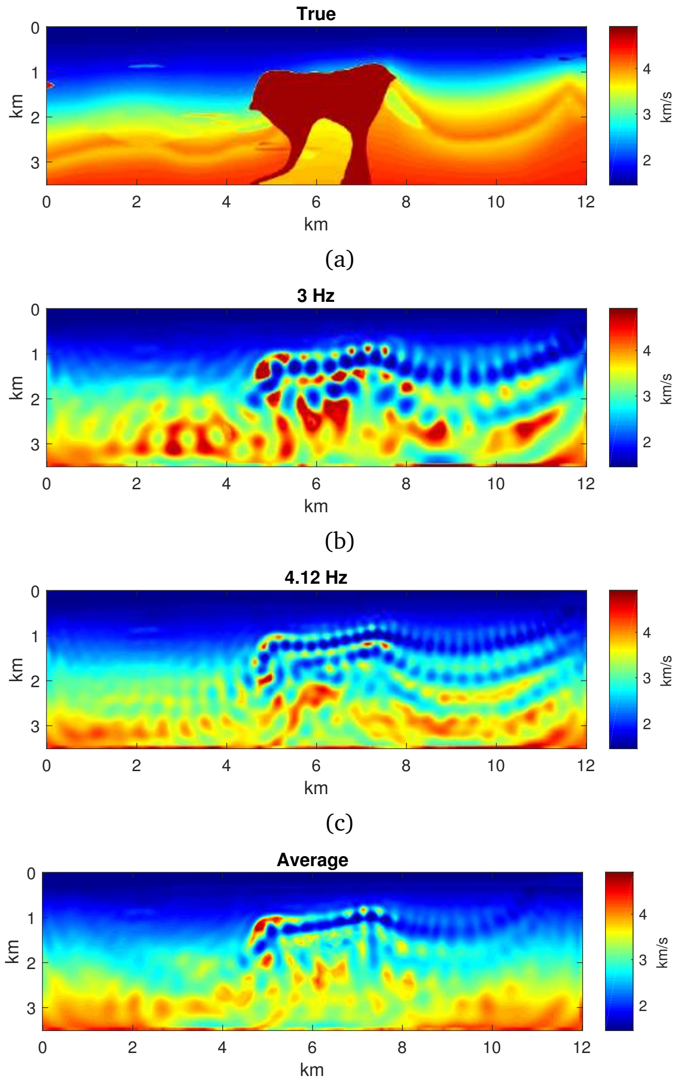
2.3 Method

In our method we use the variance between the resulting models at different frequencies to guide the flooding process. This includes identifying a sufficiently accurate starting model.

2.3.1 Variance-based salt flooding

Model updates from different frequencies deviate from each other. By comparing these updates, we can localize regions associated with global and local minima in the model domain. For this purpose, we compute in parallel updated subsurface models, $\mathbf{m}_k$, for a number, K, of distinct frequencies using FWI. Figure 2.3b-c shows such models, $\mathbf{m}_k$, for one of the chosen test cases. The number of distinct frequencies is arbitrary, but our tests revealed that four distinct frequencies are usually sufficient.

The selection of frequencies is motivated by the need to have variation in the cycle-skipping artifacts between the updates, while the updates for the easily resolvable areas should be similar. The frequencies therefore should not be too close to each other to detect the variation and not too far apart to have similar wavenumber coverage. Because a constant ratio between subsequent frequencies is natural for frequency-domain inversions [94, 95], we scaled the frequencies with a small coefficient inside our bands.



(a)

(b)

(c)

(d)

Figure 2.3: Central part of the BP 2004 benchmark velocity model: (a) the true velocity model; updated velocity models from FWI at minimum 3 Hz (b) and maximum 4.12 Hz frequencies (c) used for flooding; (d) weighted average between model updates from different single frequencies. Repeated contrast velocity intrusions are cycle-skipping artifacts caused by a significant mismatch between the initial and true velocity models.

We constrained the scaling ratio such that the maximum frequency used for flooding, f_{max}, would not exceed double the minimum frequency f_{min}. This allows some similarity of updates from multiples of f_{min} to be mitigated, which we observed in our experiments. This similarity can be roughly explained in the following way. When the frequency is doubled the phase shifts between the observed and modeled data are doubled. As a consequence, if the phase shift was very close to 2π for frequency f, it would be close to 4π for the frequency $2f$ and the updates would be very similar in approximately half of the cases. This suggests that the updates from these frequencies are very likely to be similar even for cycle-skipped areas.

This requirement suggests that the frequencies, f_k, required for flooding should be distributed in the range $f_k \in [f_{min} \ldots 2f_{min})$. The coefficient that allows $K = 4$ frequencies, f_k, in this band is then defined

$$\frac{f_{max}}{f_{min}} < 2, \quad f_{k+1}/f_k < \sqrt[3]{2} \approx 1.259. \tag{2.3}$$

To satisfy equation 2.3, we take

$$f_{k+1}/f_k = 0.9^{-1} \ldots. \tag{2.4}$$

An example of this approach is visualized in Figure 2.4.

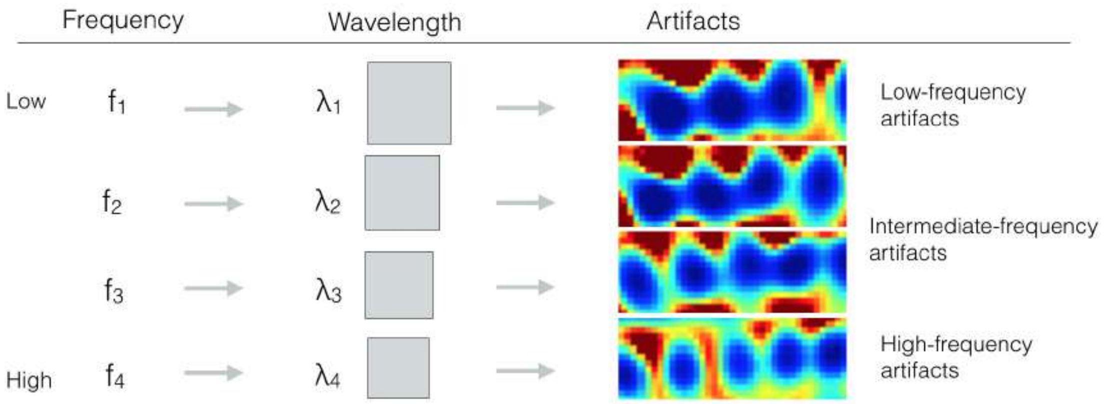


Figure 2.4: The dependence of cycle-skipping artifacts on local wavelengths. Frequencies (left column) increase from top to bottom with corresponding wavelengths (central column) and associated cycle-skipping artifacts (right column) in velocity models.

Updated models, $\mathbf{m}_k$, are together passed through the following three steps of the flooding procedure:

1. **Weighted model average.** Due to the local nature of cycle-skipping artifacts, they should vary in updates from different frequencies. To enhance regions that match among most models and to diminish those that are different, we sum up all models and find the weighted average of the FWI updates. As a result, we create a background model, $\mathbf{m}_b$, in which cycle-skipping artifacts are reduced and well-defined features are enhanced:

$$\mathbf{m}_b = \frac{\sum\limits_{k=1}^{K} \mathbf{m}_k w_k}{\sum\limits_{i=1}^{K} w_i}, \qquad w_k = \frac{1}{f_k}, \tag{2.5}$$

where $\mathbf{m}_k$ is the model update from FWI at frequency f_k. The weighting term, w_k, obtains its largest value at the lowest frequency. Such averaging permits data from the lower frequencies to contribute more to the averaging because low-frequency content is less corrupted by numerical artifacts and provides, in general, smoother models. Figure 2.5 presents a weighted average model for the BP test case.

2. **Weighted model variance.** Because variance is a measure that indicates how much a variable alternates from its weighted average value, high values indicate large differences in model updates. High-variance areas mostly appear near or within salt bodies where cycle-skipping is the most prominent (Figure 2.6(a)). These uncertain areas are prone to be problematic and thus require additional processing to weaken the artifacts and to help to avoid local minima. We find the variance distribution, $\mathbf{V}$, point-wise among the model updates:

$$\mathbf{V} = \frac{\sum\limits_{k=1}^{K} w_k \left(\mathbf{m}_k - \mathbf{m}_b\right)^2}{\sum\limits_{i=1}^{K} w_i}, \tag{2.6}$$

where w_k has the same weighting term that was used to find the background model, $\mathbf{m}_b$, and $\mathbf{m}_k$ are the model updates for frequencies f_k.

Variable variance threshold. Let $\epsilon \in [0, 1]$ be a coefficient that separates a normalized variance distribution into normal and anomalous parts (Figure 2.6(b)). Anomalous parts are those where the variance among the updated models is relatively high, implying that these areas could be associated with local minima. Anomalous regions shape a "variance mask", $\mathbf{V_m}$, which overlaps "suspicious" regions:

$$\mathbf{V}_m = \begin{cases} 1, & \text{if } \frac{\mathbf{V}}{V^{max}} > \epsilon \\ 0, & \text{otherwise} \end{cases} \tag{2.7}$$

We change the threshold, ϵ, on the fly depending on the ratio between the mean V^{avg} and the maximum V^{max} values of the original variance distribution, $\mathbf{V}$. When the optimization technique finds the global minimum, the minimum is often stationary for all frequencies, f_k. This implies that the difference between updates is small and therefore the average and maximum values of the variance distribution are close to each other. For all flooding iterations, subscripted as i, we store the mean variance distribution, V_i^{avg}, normalized by its maximum, V_i^{max}. Then, we modify the threshold, ϵ_n, depending on this ratio and its history. We define anomalous areas in the model as those that meet the relation given by equation 2.7, where:

$$\epsilon_n = \epsilon_0 \frac{\max\limits_{i=1,n} \widehat{V}_i^{avg}}{\widehat{V}_n^{avg}}, \qquad \widehat{V}^{avg} = \frac{V^{avg}}{V^{max}} \tag{2.8}$$

Threshold, ϵ_n, is computed at each flooding iteration, n, and increases when the mean and maximum values of the variance distribution are close. Generally, ϵ_n increases with iterations as updates at given different frequencies become increasingly similar as model converges and $\widehat{V}_n^{avg}$ decreases. The initial threshold, ϵ_0, is set to the empirical value 0.2 found by the examination of the initial variance distribution, $\mathbf{V}$.

3. **Flooding of high-variance areas.** All regions covered by the variance mask, $\mathbf{V_m}$, are point-wise filled with the power mean [98] of the values found within a circle with a radius of half a local wavelength. We assume that such a radius corresponds to the average size of a local cycle-skipping artifact. The power of the mean, p, defines how biased the result will be toward higher values. The empirical rule is to set p equal to the signal-to-noise ratio (SNR) in the observed data. SNR is defined as the ratio between the energies of the signal and the noise. The maximum local wavelength at each point

of the model is determined by

$$\mathbf{\Lambda}_{max} = \frac{\mathbf{m}_b}{f_{min}}, \tag{2.9}$$

where $\mathbf{m}_b$ is the weighted average background velocity model and f_{min} is the minimal frequency at which inversions are performed. The point-wise expression for flooding with a power mean is

$$\begin{gathered} \underset{i,j \in \mathbf{V_m}}{m_b^{(i,j)}} = \sqrt[p]{\frac{\sum_k \sum_l \left(m_b^{(k,l)}\right)^p}{\sum_k \sum_l 1}}, \; k,l \in S \\ S: \; \sqrt{(i-k)^2 + (j-l)^2} \leq \frac{\lambda_{ij}^{max}}{2}. \end{gathered} \tag{2.10}$$

Depending on SNR, there are three natural outcomes from the equation 2.10. When the data are noise free, the power of the mean $p \to \infty$ delivers a maximum value from circle S to model update $m_b^{(i,j)}$. Whereas when the signal and noise are indistinguishable, $p = 1$, the output is an arithmetic average within the circle. The higher the SNR, the closer the value of the update to the maximum within the circle. The high-variance areas in Figure 2.6(a) are processed according equation 2.10. This allows the algorithm to average the cycle-skipping artifacts with relevant content from inside the circle region. A sample of a resulting flooded velocity update after the first flooding iteration is shown in Figure 2.7.

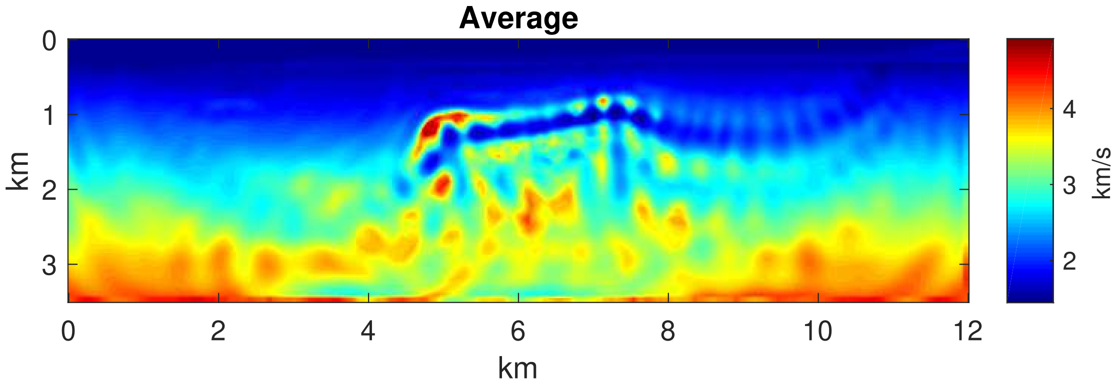
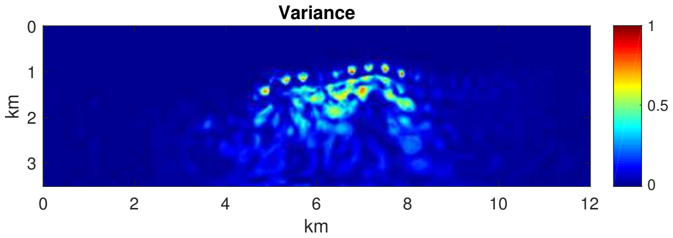
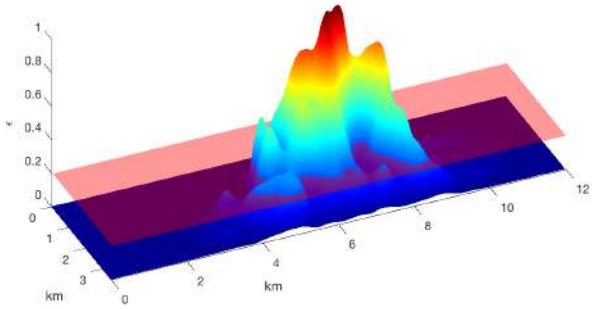


Figure 2.5: Model average $\mathbf{m}_b$ built as weighted average between model updates $\mathbf{m}_k$ from different single frequencies f_k.

We applied this three-step flooding algorithm to results from FWIs computed in



(a)



(b)

Figure 2.6: Model variance distribution: (a) the variance distribution among selected FWI updates at 3.0, 3.33, 3.7 and 4.12 Hz after the first 15 L-BFGS iterations; (b) the variance threshold, $\epsilon = 0.2$, where all areas with values above the threshold ϵ (red plane) are considered as anomalous.

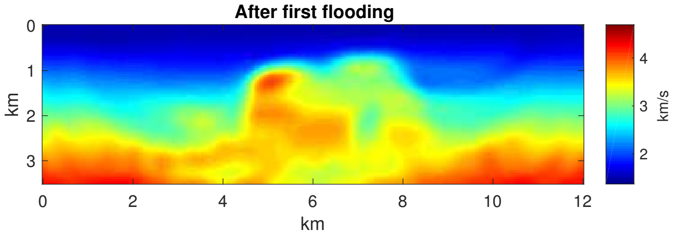


Figure 2.7: Salt flooding: the velocity model after the first run of salt flooding procedure on the central part of the BP 2004 velocity model.

parallel at single frequencies. We iteratively repeated the three steps until the introduced V^{avg}/V^{max} ratio stagnated.

2.3.2 Selection of the starting model

The initial model for FWI should provide a fair approximation of the general background of the area under study. Here, we used a one-dimensional velocity model with a water layer on top as the initial model. The suitable slope, β, of the linear model was found by brute-force search. This minimizes the misfit between observed and modeled data. An analytical function describing the set of initial models can be written as

$$v(z) = v_w + \beta \, max \left(0, z - w\right), \tag{2.11}$$

where v_w is the velocity in the water layer, w is the known water depth, and β is the variable slope of the one-dimensional velocity model. The gradient could start either from a flat sea floor, or from real bathymetry or be a combination of two linear models.

2.4 Results

We demonstrate the applicability of our variance-based salt flooding technique in a two-dimensional, isotropic acoustic medium mimicking geological structures of the deep-water area of the Gulf of Mexico. Our first and second numerical examples demonstrate implementation of the method with synthetic noise-free and noisy data (SNR=10) generated from sections from the central and left parts of the BP 2004 model [92], respectively. In the last example, we operate with noise-contaminated data (SNR=5) generated for the east-west cross section of the SEAM Phase I velocity model [93]. Moreover, we assume the water depth and the velocity to be known.

2.4.1 Central part of BP 2004

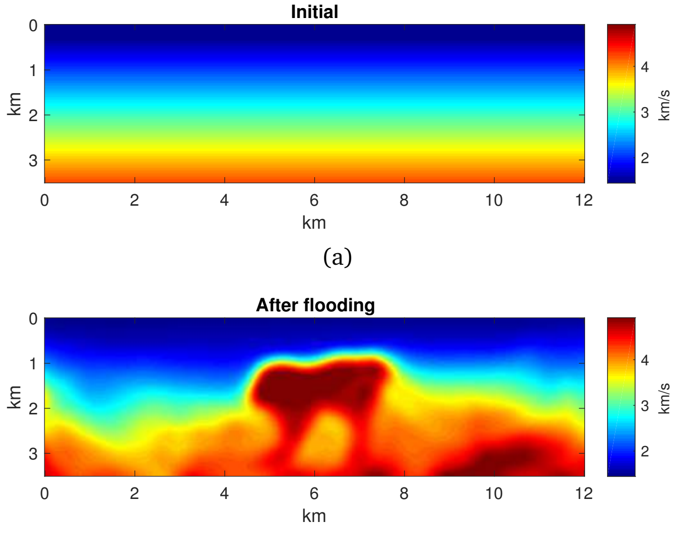
The central part of the BP 2004 model is challenging for FWI due to a massive salt body with steep flanks and a hidden low-velocity intrusion (Figure 2.3(a)). This intrusion is a potential trap for hydrocarbons. The primary goal is therefore to gain a satisfactory image of the area below the salt body. The acquisition involved 241 receivers and 121 sources uniformly distributed on the surface with respective spacing of 40 and 80 meters. The size of the submodel was 3.5 km $\times$ 12 km; the model was discretized by a Cartesian uniform grid 175×600 with 20 m spacing in both directions. The frequency range for multiscale FWI was from 3 to 7.46 Hz with a multiplicative increment of 1.2.

The initial model for mono-frequency multiscale FWI was built using the proposed salt flooding technique. As input, the algorithm was given velocity updates from four single frequencies that were expected to provide optimal variance content. These frequencies were sampled according to equation 2.4 from the minimum available frequency of 3 Hz with the step size equivalent to the change in local wavelengths of 10%. The resulting set of frequencies was 3.0, 3.33, 3.7 and 4.12 Hz. The initial variance threshold ϵ_0 was 0.2. Any variation between the updates above this threshold was considered to be anomalous. This initial value was roughly selected by examination of the variance distribution (Figure 2.6(b)) such that it would separate the dominant peak.

Several input models to the flooding routine are shown in Figure 2.3b-c. These are the updates from FWI at the lowest and highest single frequencies after 15 iterations using a limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm [99]. Cycle-skipping artifacts (repeated velocity anomalies) are clearly visible on each updated model.

At the lowest of available frequencies used for the flooding, we performed forward modeling on a coarser mesh to ensure that there were at least ten grid points per shortest wavelength in the model. This is equivalent to 88×300 grid points with 40 m spacing. This problem size reduction leads to faster computation for FWI only within the three-step flooding procedure. We start our flooding approach considering the background model shown in Figure 2.8(a).

The variance distribution between updates is the key ingredient of the proposed approach. This distribution pinpoints regions in the model that were reconstructed in the most ambiguous way. Thus, we compute the variance between the updates from a set of single frequencies according to equation 2.6. Following the iterative flooding workflow described in the previous section, we reconstruct the shape of the salt body with its principal features, such as its internal cavity. The resulting starting model for multiscale FWI is shown on Figure 2.8(b). It took 20 runs of the three-step flooding procedure to build this model. The proposed technique requires more iterations to converge when applied to noise-free data. With such data, a variance mask likely covers accurately a high-variance region of the velocity model where most of the cycle-skipping artifacts occurred. The flooding procedure performs weighted averaging inside a circle area with a radius of half a local wavelength. Thus, the more homogeneous area is covered by the circle, the less changes are introduced by weighted averaging. As a result, it either takes longer to populate the anomalous areas according to equation 2.10 or the process stagnates when the artifacts from distinct frequencies align completely.

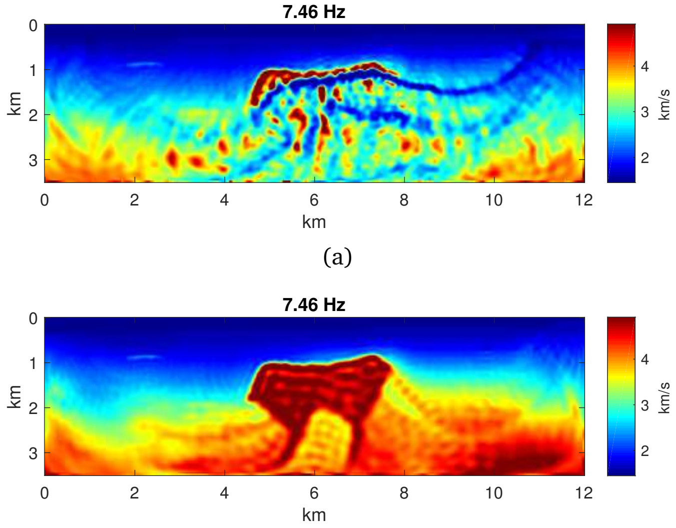


(a)

(b)

Figure 2.8: Starting models for multiscale FWI for the central part of the BP 2004 model: (a) the initial, one-dimensional velocity model with a constant gradient; (b) the velocity model obtained after 20 flooding iterations using our variance-based salt flooding technique. Numerous flooding runs are needed for noise-free data.

Non-regularized conventional multiscale FWI including a series of successive mono-frequency inversions fails when started from a one-dimensional initial model (Figure 2.8(a)). As expected, the top of the salt is correctly constructed whereas other parts are corrupted by cycle-skipping artifacts (Figure 2.9(a)). In contrast, our proposed iterative flooding approach allows for the preparation of an adequate starting model (Figure 2.8(b)) that leads FWI to a better final result (Figure 2.9(b)). As shown in Figure 2.9(b), the low-velocity anomaly beneath the salt body surrounded by thin salt “legs” is well resolved. Usually, such thin salt intrusions beneath a massive salt body can hardly be identified by conventional FWI. The mismatch between the inverted model and the true model in the right bottom corner is caused by poor illumination in that area.



(a)

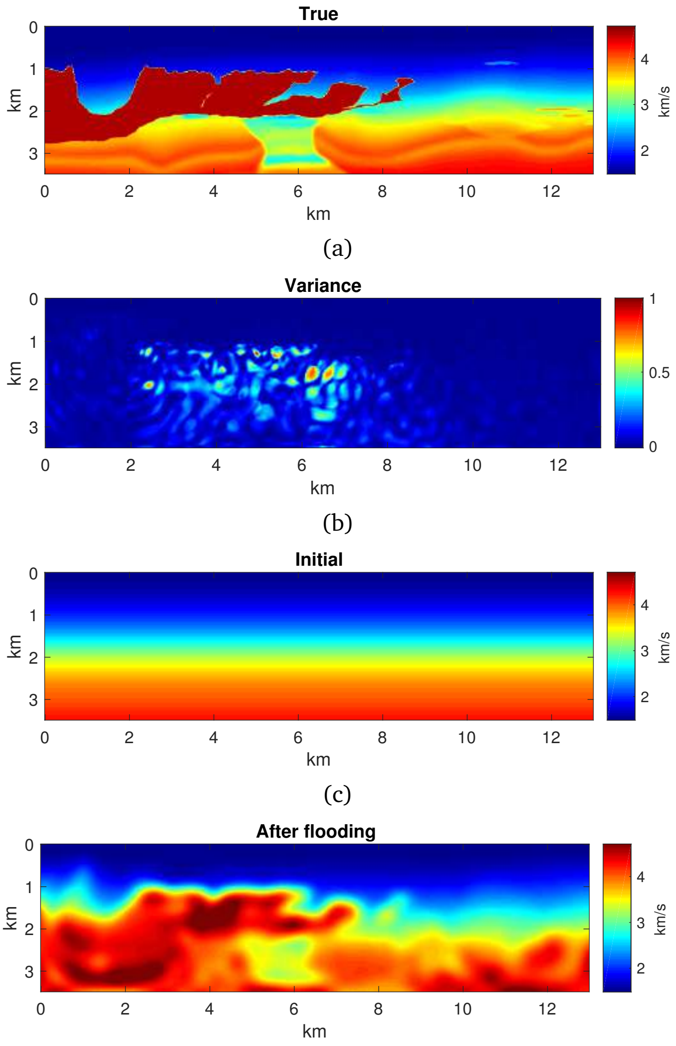
(b)

Figure 2.9: The resulting velocity model for the central part of the BP 2004 model after 30 iterations of conventional FWI at a range of frequencies from 3 to 7.46 Hz, starting from the (a) linear and (b) flooded initial models.

2.4.2 Left part of BP 2004

Potential hydrocarbon reservoirs in the left part of the BP 2004 velocity model (Figure 2.10(a)) can be found immediately beneath an elongated salt body [92]. The impedance contrast at these areas is higher than elsewhere in the model, which suggests multiple difficulties when imaging these targets. The crop is 2.5 km × 13 km in size and it is also discretized by a uniform Cartesian grid with 20 m spacing (175×650 grid points). The acquisition involved 314 receivers and 158 sources uniformly distributed along the surface with respective spacing of 40 and 80 m. SNR was set to 10, which is relatively high for surface seismic.

We start our flooding approach based on the background model shown in Figure 2.10(c). The highest variance between the updates occurred in the areas below the salt (see Figure 2.10(b)). This variance indicates that there is uncertainty in the inversion results in these areas. After 12 runs of the flooding procedure, the mean and maximum values of the variance distribution grew close and the stopping criterion for the flooding loop was reached. The flooded velocity model is shown in Figure 2.10(d). In the noise-free case, the number of FWI iterations at each flooding step varies because the exit criterion is the stagnation of the functional decrease. In the presence of noise (SNR=10), this approach does not work well. We therefore set a constant number of FWI iterations preceding each flooding iteration. In this example, each flooding run required 10 L-BFGS iterations as



(a)

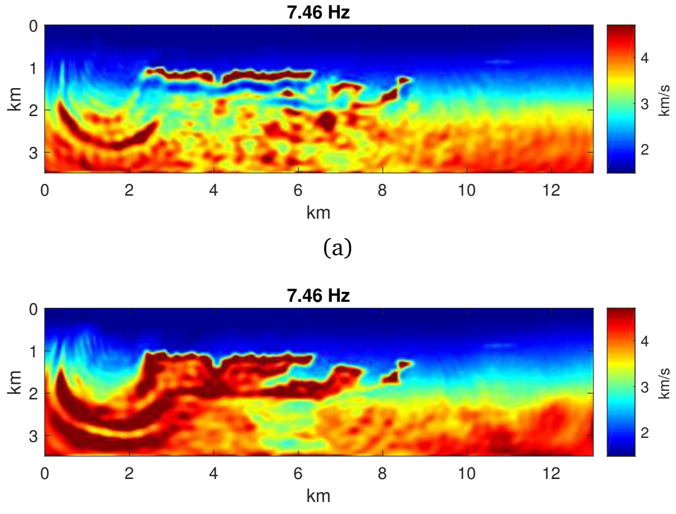
(b)

(c)

(d)

Figure 2.10: Left part of the BP 2004 benchmark velocity model: (a) the true velocity model; (b) the variance distribution among updates from 3.0 to 4.12 Hz used for salt flooding; (c) the initial, one-dimensional velocity model with a constant gradient; (d) the velocity model obtained after 12 flooding iterations using our variance-based salt flooding technique.

this many iterations delivered stagnation of the gradient. The flooding was performed on a sparse, resampled computational grid with 88×325 grid points due to the involvement of only long-wavelength data.



(a)

(b)

Figure 2.11: The resulting velocity model for the left part of the BP 2004 model after 30 iterations of conventional FWI at 3–7.46 Hz, starting from the (a) linear and (b) flooded initial models.

Conventional multiscale FWI with total variation (TV) regularization [100] starting from 1D velocity model (Figure 2.10(c)) succeeded in imaging the top of the salt, but failed to reveal the sub-salt structures (Figure 2.11(a)). Cycle-skipping artifacts were easy to discern within the salt body where the initial velocity model differs significantly from the true one. Given that we used the flooded velocity model (Figure 2.10(d)) as the initial model for multiscale FWI, we could recover both the salt body and the area beneath the salt. Despite the relatively low frequencies used for FWI, most short-wavelength features could be distinguished in the final inversion result (Figure 2.11(b)). One of the targeted low-velocity areas below the salt, likely a reservoir, is inverted in its correct location and can clearly be identified. The small reservoir to the left in the area below the salt could not be recovered probably due to the presence of noise in the data.

2.4.3 Cross-section of SEAM Phase I

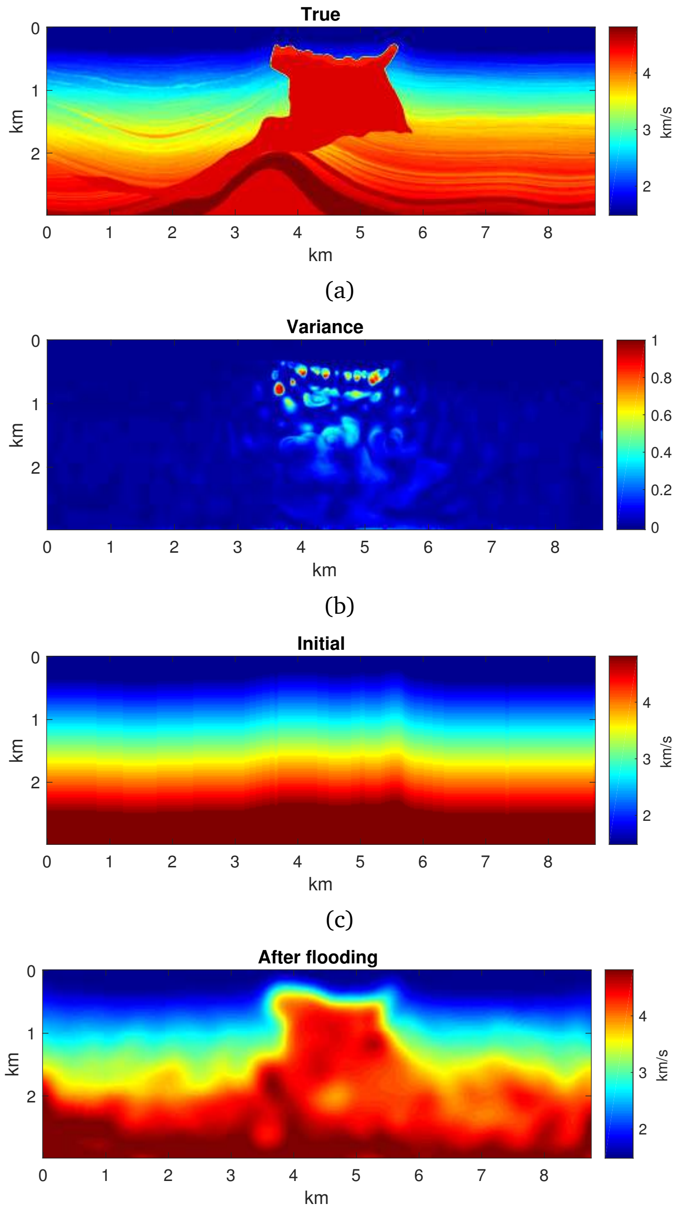
The subsurface model built on the SEAM Phase I dataset mimics a realistic geology of a salt-containing region in the Gulf of Mexico [93]. There is a massive salt body with steep flanks embedded into a layered sediment environment (Figure 2.12(a)). We implement the variance-based technique on a resampled 2D subset of the model taken in the east-west direction from North 23.9 km with dimensions of 3 km × 8 km. We run FWI from 3 to 10 Hz. The target area was the sedimentary basin hidden beneath a hanging salt flank.

The model was discretized with 150 × 438 grid points with 20 m spacing. The acquisition involved 106 receivers and 53 sources uniformly distributed along the surface with respective spacing of 40 and 80 m. The presence of noise is ubiquitous in real data applications, we therefore added random Gaussian noise with SNR of 5 to the data.

We initiated the flooding procedure on a sparse grid (75 × 298 grid points) starting from a linear velocity model with known sea floor bathymetry (Figure 2.12(c)). During seven flooding runs we performed 15 iterations of mono-frequency non-regularized FWI in parallel at 3.0, 3.33, 3.7 and 4.12 Hz. The algorithm computed variance between updated models from listed frequencies and applied corrections according to equation 2.10, which ultimately resulted in the low-wavenumber initial model shown in Figure 2.12(d). The variance distribution before the first iteration of the flooding procedure was dominated by inconsistencies that were located primarily in the top part of the salt body (Figure 2.12(b)). This was because the most significant velocity contrasts occurred there. The variance threshold ϵ_0 was 0.2. The number of FWI iterations was selected after examination of the data misfit functions and remained constant for all inversions at all frequencies (Figure 2.13).

To provide full-scale FWI, we ran successive inversions of mono-frequency data at 3.00, 3.6, 4.32, 5.18, 6.22, 7.46 and 10.75 Hz in the framework of conventional multiscale FWI [94]. Conventional FWI regularized with TV failed due to the lack of low frequencies (Figure 2.14(a)). However, it could capture top of the salt although deeper parts, such as the complex basement and the salt body itself were not captured.

When started from the initial model produced by the flooding procedure, the final result from the TV-regularized inversion was improved (Figure 2.14(b)). While the resulting model is still mildly blurry, all boundaries of the main salt body, and the high-velocity arch in the basement are resolved in the final model.



(a)

(b)

(c)

(d)

Figure 2.12: Cross section of the SEAM Phase I velocity model: (a) the true velocity model; (b) the variance distribution among updates from 3.0 to 4.12 Hz used for salt flooding; (c) the initial, one-dimensional velocity model with a constant gradient; (d) the velocity model obtained after 7 flooding iterations using our variance-based salt flooding technique.

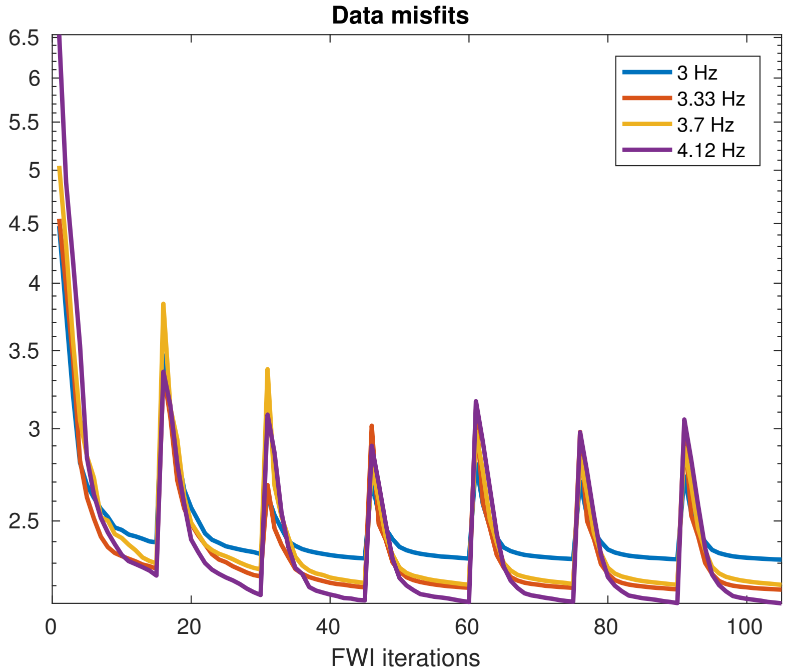
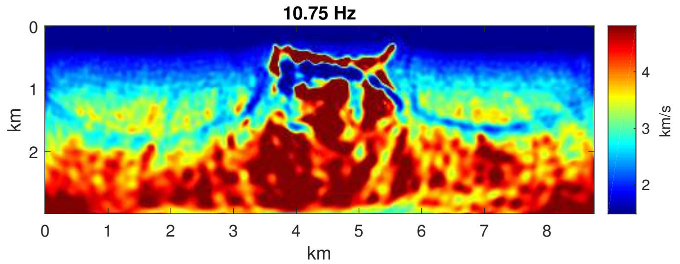
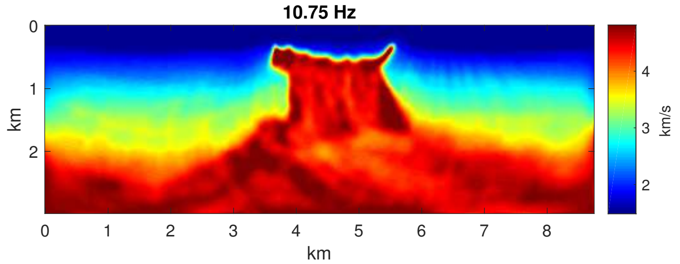


Figure 2.13: Misfit functions decay during flooding iterations. Each spike except the first one reflects a flooding run with the preceding 15 iterations of mono-frequency FWI on a coarse grid at 3.0, 3.33, 3.7 and 4.12 Hz for a cross section of the SEAM Phase I model.



(a)



(b)

Figure 2.14: The resulting velocity model for the cross section of the SEAM Phase I model after a few iterations of FWI with total-variation regularization at each listed frequency ranging from 3 to 7.46 Hz and SNR of 5, starting from the (a) linear and (b) flooded initial models.

2.5 Discussion

The proposed variance-based reconstruction technique extends the applicability of conventional FWI to complex salt bodies. As input, it uses model updates at certain discrete frequencies. As output, the technique provides an alternative model update that can be used as an initial model for the next FWI iteration. The proposed procedure can be easily embedded into any conventional FWI routine, regardless of the solver or computational domain because only model updates for corresponding single frequencies are required as input.

When forward modeling is performed in the time domain, on-the-fly discrete Fourier transform [101] can be used to extract specific frequency content from the data. Apart from a set of standard parameters involved in frequency-domain FWI, we used the following additional tuning parameters in the proposed procedure:

Number of FWI iterations before flooding. We used L-BFGS with a stopping criterion given by the relative decrease in the functional that was less than 10^{-4}. This empirically resulted in 10 to 20 L-BFGS iterations per flooding cycle. This parameter is dependent on the FWI realization and problem. It should therefore be determined directly from examining a specific misfit function. The number of iterations depends on the choice of optimization algorithm. In a reasonably complex model, quasi-Newton methods are typically the best choice for FWI implementation. A truncated Newton method in some cases can be a better option to handle multiscattering [102]. Potentially, a single but computationally more expensive iteration of the truncated Newton could replace several L-BFGS iterations currently used for variance map construction. Another option would be to include multi-scattering in the inversion as suggested by [103].

Step size between single frequencies. The step size has to be sufficient to cause a visible shift of the cycle-skipping artifacts. We scaled the frequencies such that the wavelength at the same location for the next frequency would be 0.9 of the previous one, which provides a bound on the frequency scaling factor that depends on the number of frequencies.

Initial variance threshold. The parameter $\epsilon_0 \in [0, 1]$ separates normal and anomalous values in the variance distribution. An arbitrary case can be found by algorithmic or visual examination of the variance distribution such that the ϵ plane cuts off the dominating high-amplitude values. From our tests, an empirical value of 0.2 is recommended. This threshold should be amended if the flooding process does not converge.

Power of the mean. According to equation 2.10, parameter p indicates how biased the flooding will be toward higher values. We set p equal to SNR in the observed data as noise adds ambiguity to the inversion, which then suggests that the averaging in unstable areas be milder to avoid unrealistic high values in the model.

Number of flooding iterations. The process should stop automatically when the high-variance regions are gone. Equation 2.8 introduces a variance-dependent function that stagnates with iterations when the maximum and mean values of the variance distribution converge. The relative decrease in the result of equation 2.8 serves as a stopping criterion for flooding iterations.

In total, there is only one purely user-defined parameter – the initial variance threshold, $\epsilon_0 \in [0, 1]$, whose recommended default value is 0.2. The number of FWI iterations before flooding depends on the optimization technique used. The step between frequencies is selected from the empirical wavelength ratio (equation 2.4). The iterative flooding procedure terminates automatically when the variance distribution in the model appears to be close to homogeneous.

The computational cost of our proposed variance-based salt-flooding procedure is determined by the lowest frequency available in the data. The flooding routine itself only computes the variance between the updated models and then performs correction on the most unstable areas. The execution time of three-step flooding iteration in most cases is therefore negligible. Even for large models this process takes only a few seconds. The major computational cost is introduced by the preceding iterations of the optimization technique that searches for minima of the objective functions in the implemented FWI routine. The numerous FWI iterations required for the salt flooding procedure should be performed on a sparse computational grid. This is because only data in the vicinity of the lowest available frequency are required to prepare the starting velocity model for full-scale FWI. The proposed approach is only expected to produce an adequate initial model for full-scale FWI, therefore only lowest available frequencies need to be modeled on a coarse mesh. Model coarsening dramatically reduces the computational costs of FWI iterations preceding the flooding procedure, making the proposed workflow feasible.

All results presented in this work were computed on a laptop with a quad-core CPU. The complete iterative salt flooding procedure on a sparse grid for each of the examples took between five and ten minutes whereas the subsequent multiscale FWI on a dense grid took about 40 minutes.

When there is noise in the observed data, the performance of the variance-based approach is improved. This is because more flooding iterations are likely necessary with the noise-free data for the algorithm to interpolate from the robustly defined regions into the ones covered by the variance mask.

To verify if the model created by the flooding procedure is ready for FWI, some quality control (QC) procedures might be necessary. These may include examination of extended images or angle gathers [104] to avoid misinterpretation of salt bodies in practice.

As a final remark, the starting one-dimensional velocity model for FWI has to be reasonable (as is the case for any other model used for FWI). A completely inconsistent initial guess of the velocity would inhibit FWI from recovering any part of the salt body that should be partly resolved in realistic cases, such as the top of the salt. In a worst case scenario, the variance-based approach will not work properly when an adequate variance distribution among the updates cannot be retrieved, which may happen when several iterations of FWI are not able to provide any information about the salt body. Nevertheless, the use of robust data at the lowest available frequency together with the salt flooding procedure on a corresponding coarse mesh allows us to obtain an adequate initial model for FWI in a cost-effective way.

2.6 Conclusions

Conventional FWI fails in the presence of the salt bodies without the aid of low-frequency data. Here, we proposed a new approach to frequency-domain FWI that allows us to perform inversions without low frequencies, without significant modifications to the optimization workflow and with only a moderate increase in computational complexity. The novelty of the approach lies in looking at nonlinear updates of FWI from several single frequencies rather than just at gradients or migration images. These updates show significant similarity if the model is mature enough to be handled by Born approximation, unlike reverse-time migration images that are different for different frequencies in a complicated medium. The variance distribution between these updates is used as an indicator to reveal areas where FWI fails. We update these areas through a simple, automated flooding process with the power mean of velocities from nearby locations.

The only inputs for a single iteration of the proposed procedure are several model updates from conventional FWI. Incorporating our flooding procedure into any existing inversion routine is therefore simple. Designed to work with the lowest frequencies available, the procedure allows the use of sparser grids for modeling and inversion for the preparation of an improved initial model to be used at later FWI stages. Moreover, these additional computational costs are negligible within a complete multi-scale FWI scenario. The proposed technique could be easily extended to 3D and more realistic physics as it operates exclusively in the model space.

Chapter 3

Deep learning for low-frequency extrapolation from multi-offset seismic data

Extrapolation of low-frequency data is an alternative to initial model building discussed in the previous chapter. During inversion, the low-frequency seismic data might compensate for the missing accurate initial model for FWI by recovering the low-wavenumber features of the subsurface. However, it is challenging to acquire field data with an appropriate signal-to-noise ratio in the low-frequency part of the spectrum. This chapter explains a method for low-frequency extrapolation in frequency domain seismic data using deep learning. We address the same problem of FWI initialization in the presence of salt bodies as in previous chapter but approaching it from data rather than model prospective. In wavenumber illumination framework we explain the connection between high- and low-frequency data at different offsets. We also show that lower frequencies are easier to extrapolate since these tend to be smooth long-wavelength functions. Motivated by this observation we train an ensemble of weak learners each extrapolating data for single frequency.

Results presented in this chapter are adaptation of the paper published in Geophysics Journal.

Ovcharenko, Oleg, Vladimir Kazei, Mahesh Kalita, Daniel Peter, and Tariq Alkhalifah. "Deep learning for low-frequency extrapolation from multioffset seismic data." *Geophysics* 84, no. 6 (2019): R989-R1001.

3.1 Introduction

Delineation of subsurface structures from a complicated seismic dataset, in particular of salt and subsalt regions, is a very challenging task [105, 106], with challenges at every stage, starting from the velocity model building process to migration, followed by its interpretation [107, 108]. To retrieve subsurface parameters from seismic data, full-waveform inversion (FWI) [2, 3, 4] proves to be a powerful tomography technique to invert for high-resolution images. It updates an initial model, that often lacks salt-body information, using the gradient of a data misfit functional, through an iterative optimization strategy. These nonlinear gradient-based optimization schemes allow to substantially deviate from the initial model assumptions and introduce strong subsalt features. Nevertheless, due to cycle-skipping problems, they often fail to retrieve reliable models as they converge to local minima rather than finding the global one, specifically when the initial model is too far from the true subsurface structure. Among the many challenging factors such as strong multiples, incomplete acquisition geometries and poor illumination due to complex overburden velocities, the absence of low frequencies in recorded seismic data contributes the most to the failure of FWI [26]. In this study, we aim to assist FWI by reconstructing the missing low frequencies, using an artificial neural network (ANN), which is a promising data-driven approach within the machine learning framework.

Seismic acquisition procedures and logistics have been significantly improved over the last two decades. However, they often fail to record the temporal low-frequency seismic data with good signal-to-noise ratio (SNR) [59]. Seismic vibrators and their required mechanical and hydraulic systems are still limited in their ability to transmit sufficient seismic low frequency energy into the subsurface. In addition to ambient noise, the recording unit has its own system noise (thermal and quantization) that aggravates SNR of low-frequency components in data. This absence of usable low frequencies in seismic datasets leads FWI to reconstruct inaccurate long-scale features, eventually yielding to a local, rather than global, minimum model. To avoid local minima solutions, conventional FWI algorithms aim either to change the misfit function [7, 10], modify the gradient [21, 22, 109], incorporate model domain regularization [110, 11], or extrapolate the missing low-frequency part of the spectrum.

With the recent advent of deep learning and artificial intelligence applications, machine learning approaches have emerged as powerful tools in data analytics. Our proposed method aims at extrapolating missing low frequencies by employing convolutional neural networks (CNN), using a deep learning approach.

Here, we build on our previous work [73] by extending the neural network approach to deep CNNs and by further investigating its theoretical limits. Since it is well established in the FWI community, that long offsets can, to some extent, replace low frequencies [111, 112], we discuss the advantages of multi-offset (shot gather) data extrapolation from a theoretical point of view, and evaluate a 2D CNN algorithm for this purpose.

By training a deep CNN on general subsalt models, we highlight the subtleties in such deep learning techniques, using on physics-based data inferences and the sampling limitations we may face in representing the associated model space. We then extrapolate missing low frequencies of seismic shot gathers in the frequency domain for different acquisitions. With those artificially enhanced datasets with low frequencies, we investigate the potential within a numerical benchmark model to invert long-scalelength features of a specific subsalt model. We start with the theory section discussing the relation between model-wavenumber and data-frequency spectra, followed by a detailed description of the deep learning framework in our study. To show the versatility of our proposed method, we consider datasets of BP-2004 and SEAM models in which the minimum available frequency is 2 Hz, and where the data inference is conducted with the same deep CNN to extrapolate lower frequency sampling points.

3.2 Theory

For low-frequency extrapolation, any data inference technique is not only limited by acquisition geometry and instrumentation, but also by the physics of seismic wave propagation. Thus, let us first derive a simple theoretical model for frequency extrapolation, based on the wavenumber illumination theory. To simplify the theoretical framework, we only consider acoustic waves hereafter.

In the frequency domain, the pressure wavefield $p(\mathbf{x}, \omega)$ of a point source located at

$\mathbf{x}_s$ satisfies the Helmholtz equation

$$(\Delta_{\mathbf{x}} + \frac{\omega^2}{v^2(\mathbf{x})})p(\mathbf{x}, \omega) = \delta(\mathbf{x} - \mathbf{x}_s)s(\omega), \tag{3.1}$$

with $\Delta_{\mathbf{x}} \equiv \sum_i \frac{\partial^2}{\partial x_i}$, and where ω denotes the angular frequency. Given a source signature $s(\omega)$ and a velocity model $v(\mathbf{x})$, equation 3.1 determines the solution of p. Source signatures can, in principle, be estimated from data [113]. Therefore, we assume $s(\omega)$ to be known and normalized at frequencies used for the extrapolation. Without further restriction, we set $s(\omega) = 1$ for each frequency ω. In general, $s(\omega)$ can be set to different values at lower frequencies, e.g., to mimic a Ricker wavelet. Exact locations of source and receivers are also known in most exploration setups. Therefore, if the true velocity model $v(\mathbf{x})$ and exact physics of wave propagation in the associated rock materials are known, low frequencies can simply be modeled by numerical methods.

Analysis of fully-nonlinear seismic inverse problem is a major challenge. In the following, we consider the Born approximation [114] for perturbed wavefields in a homogeneous model where data and model spectra can be directly related [115, 57, 111, 116, 117, 18, 118, 54, 119]. We find the perturbed trace at receiver $\mathbf{g}$ by

$$\delta p(\mathbf{s}, \mathbf{g}, \omega) \propto \delta v(\mathbf{K}) \tag{3.2}$$

where $\mathbf{s}$ and $\mathbf{g}$ are unit vectors pointing towards the source and the receiver, respectively. In seismic exploration, most subsurface models are dominated by horizontal structures. Thus, vertical wavenumbers in the model spectrum define the variations in these horizontal structures along the vertical direction. Focusing on vertical wavenumbers, equation 3.2 leads to

$$\delta p(\theta, \omega) \propto \delta v(K_z), \tag{3.3}$$

$$K_z(\theta, \omega) = 2\frac{\omega}{v}\cos\frac{\theta}{2} \tag{3.4}$$

where K_z is the vertical wavenumber and θ is the opening angle between $\mathbf{s}$ and $\mathbf{g}$, which encodes the offset. Equation 3.3 effectively connects data collected at different frequencies with the velocity model. It is evident, from equation 3.3, that if the angle θ between the source and the receiver is fixed, every wavenumber in the model is then illuminated by a unique frequency, which makes a trace-by-trace extrapolation impossible. Furthermore, if the wavefield perturbation $\delta p(\theta_1, \omega_1)$ is a known entity, we can construct the model at the wavenumber $K_z(\theta_1, \omega_1)$ and subsequently model any wavefield $\delta p(\theta_2, \omega_2)$, so that $K_z(\theta_2, \omega_2) = K_z(\theta_1, \omega_1)$, according to equation 3.3. The extrapolation process can thus be summarized as

$$\delta p(\theta_1, \omega_1) \rightarrow \delta v(K_z) \rightarrow \delta p(\theta_2, \omega_2). \tag{3.5}$$

For direct extrapolation according to equation 3.5, a perturbed model δv would need to fall within the limits of the validity regime of the Born approximation [114]. Otherwise, the extrapolation would lead to an inaccurate wavefield at the target frequency. Accuracy of the reconstructed perturbation δv depends on coverage of model wavenumber spectrum and thus can be improved by including data for more frequencies ω and offsets (implicitly introduced by scattering angle θ) to the left part of equation.

The lowest frequency that can be extrapolated without any assumptions about the model spectrum can be derived from equation 3.4. This frequency is proportional to the lowest wavenumber K_{low} illuminated by the available seismic data

$$\omega_{low} = \min_{\theta} \frac{K_{low}}{2 \cos \frac{\theta}{2} v}, \tag{3.6}$$

where K_{low} is defined by the minimum frequency ω_{min} available in the dataset as well as the opening angle θ_{max} for the longest offset

$$K_{low} = 2 \frac{\omega_{min}}{v} \cos(\theta_{max}/2). \tag{3.7}$$

.

For zero-offset data, $\cos \frac{\theta}{2} = 1$, the estimate 3.6 reaches its mimimum and leads to the limit of the lowest frequency that can be extrapolated

$$\omega_{low} = \omega_{min} \cos(\theta_{max}/2). \tag{3.8}$$

The equation 3.8 defines how far the frequency can be extrapolated without any additional assumptions about the model. This limiting frequency is defined by the minimum frequency ω_{min} available in the dataset as well as the opening angle θ_{max} for the longest offset. For example, we can estimate the lowest frequency for extrapolation at different locations in the subsurface model, based on equation 3.8, for a seismic acquisition limited by an available minimum frequency of 2 Hz (Figure 3.1).

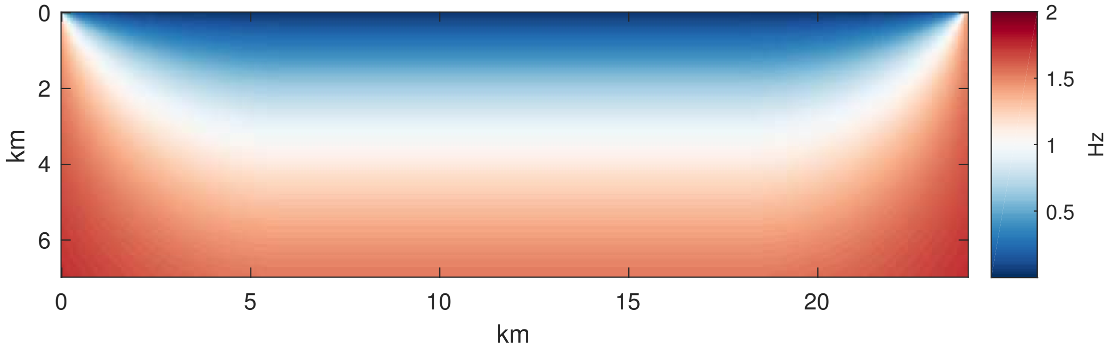


Figure 3.1: Frequency estimation ω_{lowest} for extrapolation based on wavenumber illumination theory, for a given minimum signal of 2 Hz available in the acquisition dataset and a maximum source-receiver offset of 12 km. This model size setup will be used in the subsequent benchmark study.

Typically subsurface velocities increase with depth, which provides larger illumination angles at depth from refractions and diving waves, and makes it possible to estimate lower frequencies. Additionally, we assume that the geological model has layered, block-like structures, and we build this a priori geological knowledge into the training set of subsurface models. This allows to push further down the limits of the lowest frequency that can be extrapolated.

Finally, limited applicability of the Born approximation leads us towards data-driven approach with non-linear estimators such as deep neural networks. Artificial neural network does not require explicit manipulations with the model perturbations and allow to infer a non-linear relation from numerous pairs of high-frequency data and low-frequency data. Unlike Born scattering, this relation is non-linear and does not rely on particular background velocity model. Therefore, the domain of applicability is defined only by the training data set and architecture of the network.

3.3 Deep learning framework

Neural network architectures date back as early as [120]. Since then, major advances, supported by powerful computational hardware progresses, have been made, including the recent progress in architectural representations, training capacities and data inferences in various scientific fields and engineering applications [121]. Artificial neural networks represent a mathematically simplified model of neurons connected in a biological brain. The simple mathematical unit of a neuron is called a perceptron. It mimics a neuron that accumulates charge and passes it to a non-linear thresholding output function. Multiple perceptrons assembled into a layered structure shape a simple, fully-connected artificial neural network where all neurons within a layer are connected to each other. Each incoming connection to a neuron has a weight that defines the contribution of this connection to the total output. Neural networks have become one of the most powerful and appealing data analysis tools for non-linear adaptive data regression, using multiple input and multiple output data [122].

In supervised learning, the neural network is given pairs of input and target data. The training stage consists of a two-step optimization problem of tuning the variable weights of all neurons. In a first step, input data are passed through all layers of the multi-layer network before a misfit is computed between the results from its final output layer and the desired target data. The formulation of misfit is optional, although in generic cases, it is set to a L_n - norm. In a second step, the misfit is back-propagated through each connection in the network, changing each weight according to the respective gradient (for a more complete overview we refer to [123].

3.3.1 Convolutional Neural Network (CNN)

Based on our previous experience with feed-forward neural networks [73, 74], we see that the major challenge for a fully-connected neural network architecture is the steep growth of the number of trainable parameters, when increasing the size of input and target data. This impedes scalability and leads to an impractical amount of data samples needed for the training stage. A convolutional neural network efficiently addresses this challenge by implementing a concept of local spatial connectivity, meaning that only data within a receptive field are fully-connected. The size of a receptive field corresponds to the spatial extent of a kernel, which always shares the same depth with the input volume, but is usually smaller along width and height axes, as shown schematically in Figure 3.2.

The convolution of the input data with a kernel results in a feature map that maximizes locations where a pattern in the data matches the kernel.

Feature maps from convolutions with different kernels are stacked together to build

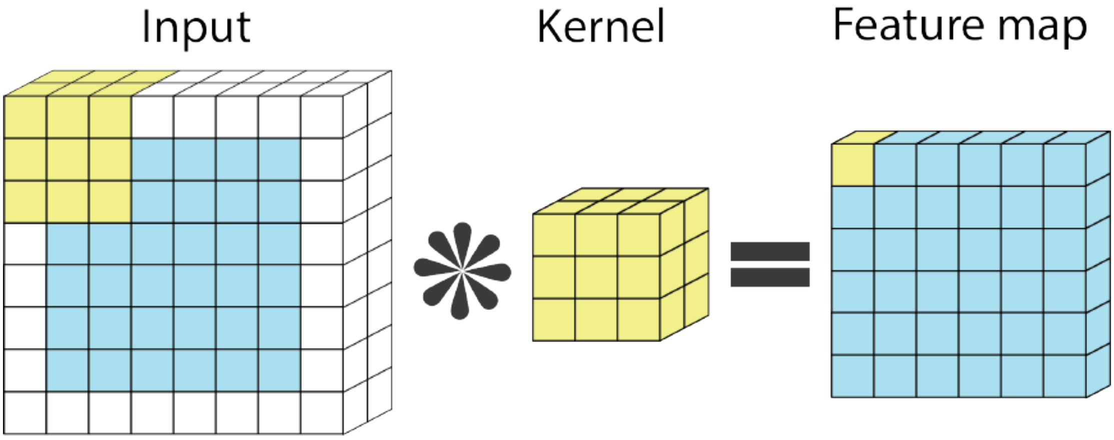


Figure 3.2: Convolution of input volume with a kernel resulting in a feature map. Depth of the kernel is always equal to the depth of the input volume, whereas its spatial extent may differ.

an input volume for the next layer. This implies that the number of kernels used in each convolutional layer is equal to the depth of the output volume from this layer. Kernels are usually initiated randomly and are trained by back-propagation of misfits between expected and predicted outputs from the network. After training, each kernel matches a certain feature in the data, with more complex features learned at deeper layers of the network [124]. For the purpose of low-frequency extrapolation, we discretize each shot gather in the frequency domain, and treat it as a digital image for feature detection. Adding several convolutional layers together will then allow to generalize these features across all shot gathers and ultimately increase the robustness of the low-frequency extrapolation with CNNs.

3.3.2 Input and target data

Marine seismic acquisition delivers a suitable framework for supervised ML applications. Towed streamers carrying a fixed number of hydrophones naturally matches the limitation imposed on the dataset shape for a neural network. The restriction is that all pairs of training and testing data should share the same dimensions throughout the dataset. Input and target data for the network designed in this study are high- and low-frequency parts of the spectrum of a shot gather, respectively. For each shot, we extract the observed high-frequency part from the frequency spectrum of seismic data recorded by a linear array of hydrophones and treat it as the input, whereas a single low-frequency representation of the shot gather is the target.

In general, neural networks in supervised learning are trained on pairs of input and target data. So far, most neural network setups are determined empirically, meaning that a trial-and-error approach is then necessary to find an optimal network architecture that will lead to the best inference result. In that sense, both training and validation data are of primary importance, whereas a method to treat these data is optional. This means that the same dataset might be processed by a number of different methods to find the best option. Thus, it is crucial to have a representative initial dataset that will constitute a solid basis to build a ML framework around it. Let us therefore explain in more details how we generate synthetic data, based on physical approximations, to determine a neural network able to generalize across different seismic acquisitions.

To investigate our deep learning approach to extrapolate low-frequencies, we generate a synthetic dataset that mimics a marine seismic acquisition. We solve the Helmholtz wave equation to obtain a pressure wavefield for a single frequency. For each frequency and source-receiver pair, we obtain a complex value describing the pressure field excited by the source and recorded at the receiver location. Data from a set of receivers sharing the same source are shaping a single-frequency shot-gather, which is a complex-valued vector with as many elements as there are receivers (Figure 3.3).

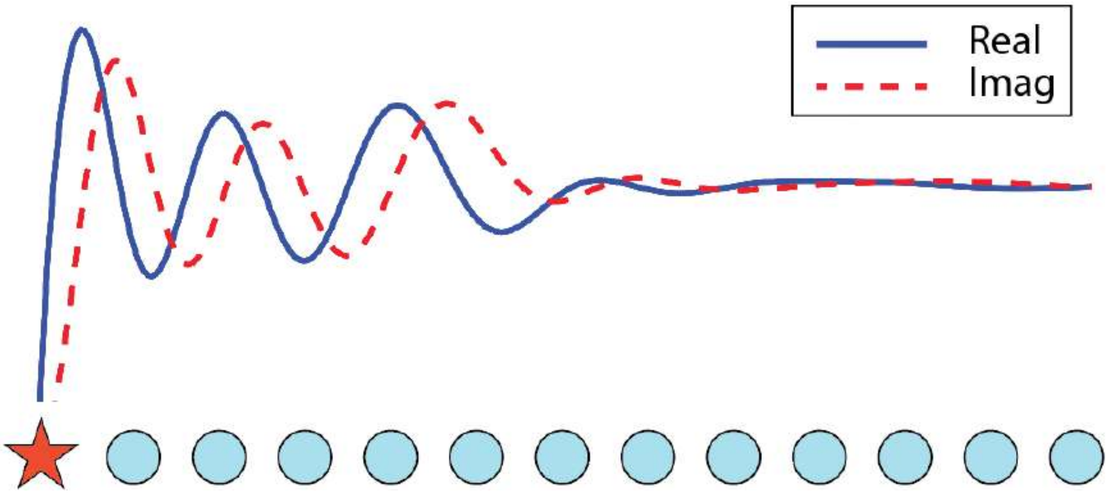
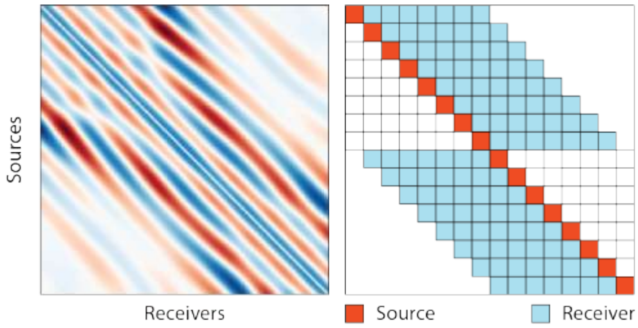


Figure 3.3: Example of mono-frequency representation of a shot gather. This is a complex-valued vector with as many elements as there are receivers in the acquisition. The red star indicates the source position and blue dots represent receivers.

The entire set of observed mono-frequency data $p(x_s, x_r)$ shapes a data matrix (Figure 3.4(a)). We assume an evenly-spaced acquisition, so the matrix is symmetric and squared. When the seismic acquisition covers the entire target area, the data matrix is densely populated. On the other hand, in marine seismic acquisition, the maximum offset is limited by the length of the streamer, which results in missing off-diagonal elements in the data matrix (Figure 3.4(b)).



(a) (b)

Figure 3.4: Real part of a mono-frequency data matrix (a) and its mapping onto fixed-offset streamer acquisition layout (b).

A stack of mono-frequency data matrices for a set of consecutive single frequencies shapes a data cube (Figure 3.5). Top and bottom surfaces of the cube are the data matrices for the highest and the lowest frequency, respectively. A slice of the cube, normal to the source axis, represents a shot gather with its spectrum content along the vertical. To preserve continuity in the wavenumber domain, we sample the frequency spectrum as a geometrical set, meaning that every following frequency is obtained from the previous one as the product with a constant coefficient. A similar selection of frequencies is used in the multi-scale approach [26], which is a common practice in frequency-domain FWI.

Exploring the data cube, one could notice continuous stripes extending within a single shot gather, as shown in Figure 3.6. These features correspond to transmission and reflection modes of the wavefield. Event tracking [66] essentially stretches data at a short offset and a high frequency to the data at a long offset and a low frequency, along the depicted lines. For a simple homogeneous velocity model, their shape could be derived analytically from a wavenumber analysis, whereas, when the velocity model is inhomogeneous, these features become distorted in a non-linear way. This makes their extrapolation challenging.

3.4 Method

The standard workflow for a machine learning algorithm consists of only three stages:

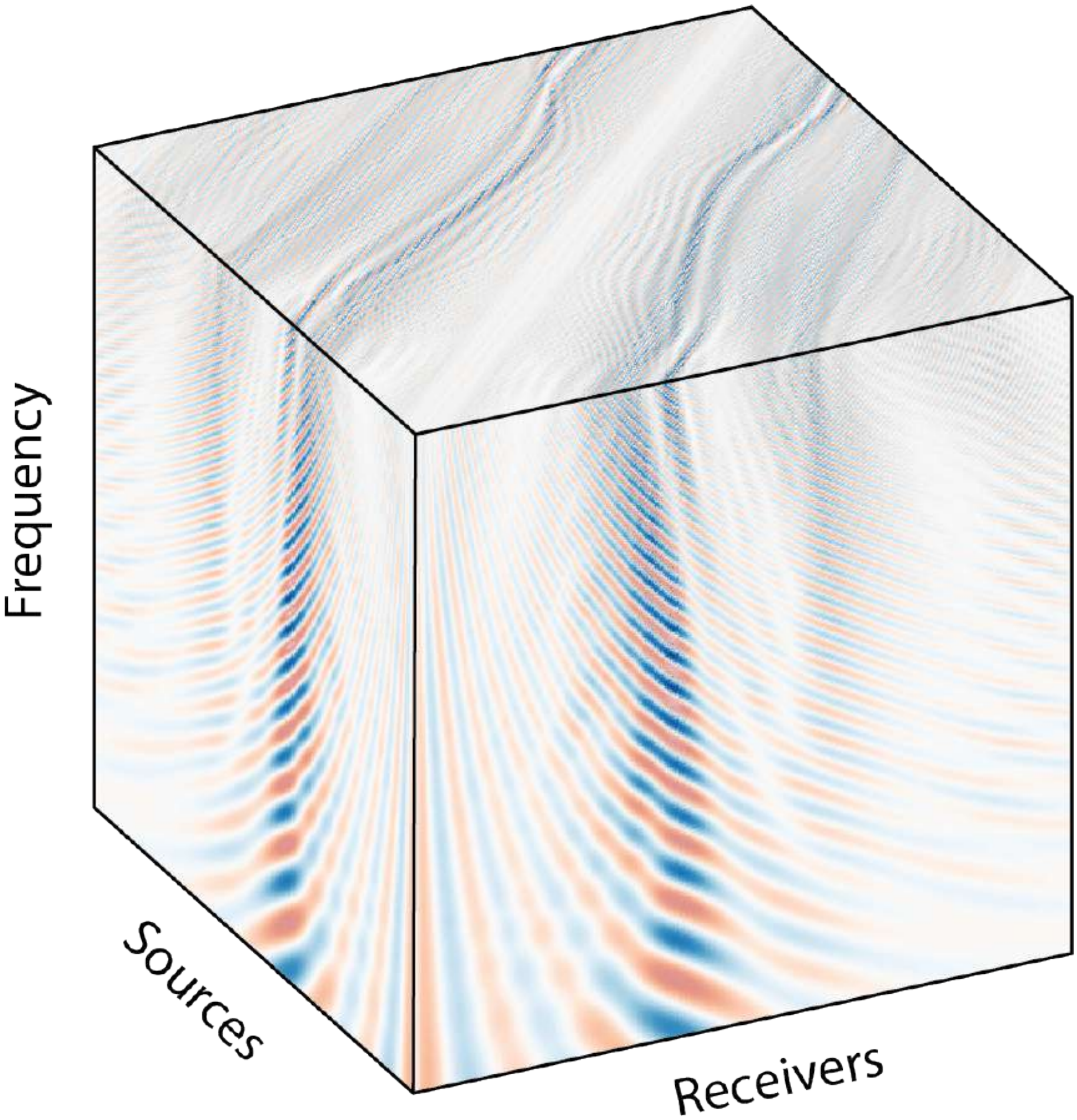


Figure 3.5: A data cube created from a stack of mono-frequency data matrices. Each side slice of the cube is a shot gather, decomposed into a number of frequencies.

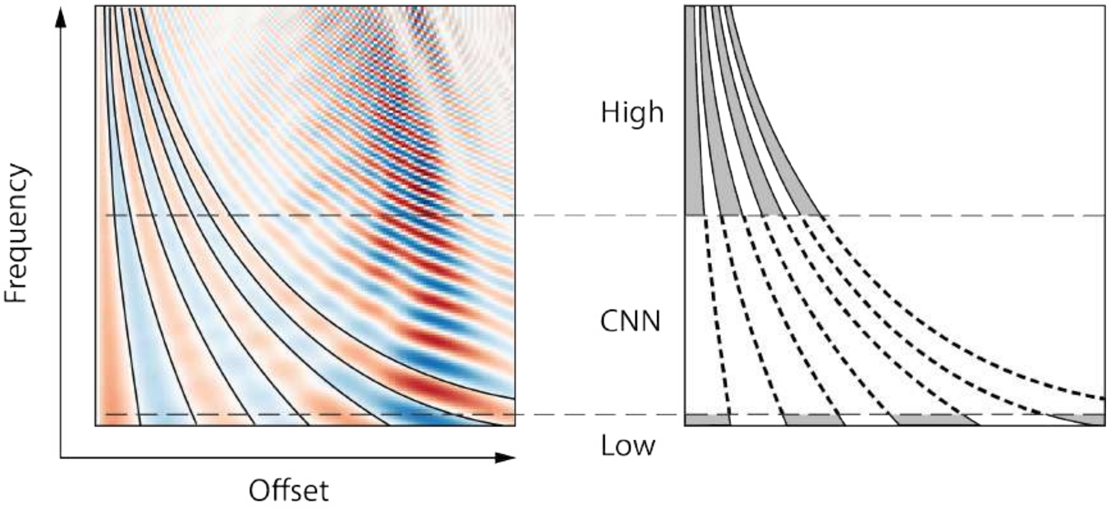


Figure 3.6: Single shot gather in the frequency domain. Lines indicate the stationary phase for transmitted waves. For reflections, the stationary phases are in the opposite direction. The data regime of the convolutional neural network (CNN) connects high to low frequencies.

- The data. The whole dataset is separated in two parts, one to serve for training of the algorithm, and the other part to be used for validation and performance evaluation. Data preprocessing is also an important step within the preparation stage, as it helps to equalize the contribution of features in the training.
- Selection and training of the algorithm. The dataset is a constant which can be manipulated in a variety of ways, so our goal is to identify an approach that delivers the best performance result on the validation data.
- Inference. The algorithm infers about the unseen target data based on its training experience.

For a streamer acquisition with R hydrophones, the input data has the shape of $F \times R \times 2$. The last factor 2 is due to the real and imaginary parts of the wavefield in the frequency domain, and F indicates the number of frequencies in the high-frequency part in which the shot gather spectrum has been split. Similarly, the shape of the mono-frequency target data is $1 \times R \times 2$. We discretize the known frequency range into $F = R$ frequencies to get a squared volume with two channels. Equal spatial dimensions in the image assure equal contributions from both vertical and horizontal features within the training stage of the CNN. Figure 3.7 shows a schematic representation of our input data matrix and output data vector. Note that the desired target output is specific to a single low frequency predicted at R receivers, i.e., each single low-frequency extrapolation is associated with an individual neural network.

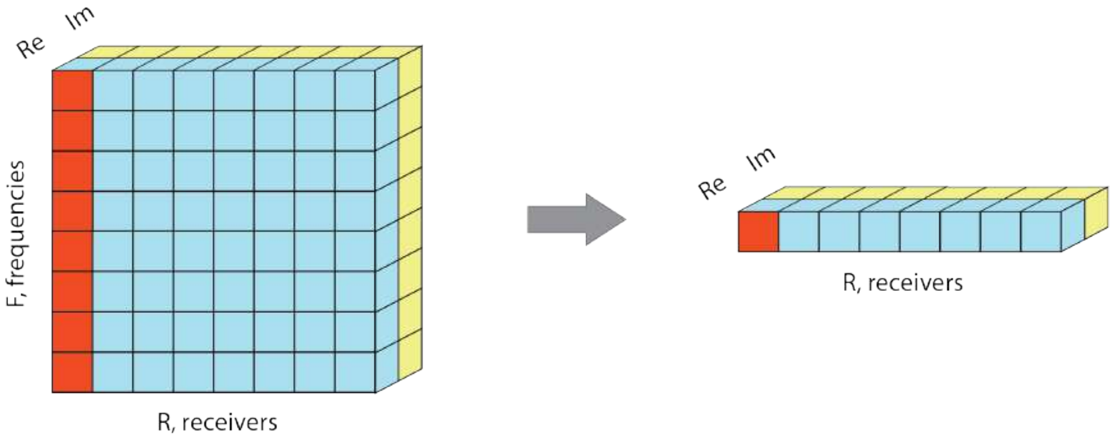


Figure 3.7: Input data is a sampled high-frequency part of a shot gather spectrum, whereas target data is a single low-frequency shot gather representation.

3.4.1 Random model generation

For all artificial neural network applications, the training stage is the most crucial step. In supervised learning, multiple pairs of data are needed to train the network, where massive datasets are often required for more complex data analysis applications. For our purposes, many real-world seismic data acquisitions would be required to train a CNN capable of generalizing subsurface features, which would become unfeasible. Instead, we generate all synthetic data for training based on approximate physics, all synthetic data for training and then use the trained CNN for inference on samples of true data generated for benchmark models. All synthetic seismic datasets are modeled for marine seismic acquisitions in randomly generated velocity models, and are used to train and validate the neural network.

The generation of realistic velocity models itself is a challenging task and deserves separate consideration [125, 126, 127]. Here, we only list empirical practices that outperformed others in our particular framework. We find that approaches for random model generation such as random Gaussian fields, layered models, and linear gradient models with salt-body intrusions, do not succeed in building a representative set of velocity models that would lead to successful training. A principal component analysis of those random velocity models shows that, often, only a few principal components dominate all others, which may be the result of improper parameter tuning or an implementation bias. However, when building random models from interpolation in-between random 1D velocity profiles (Figure 3.8), the total model variance is spread among a larger amount of principal components, and thus results in a better network inference on validation data.

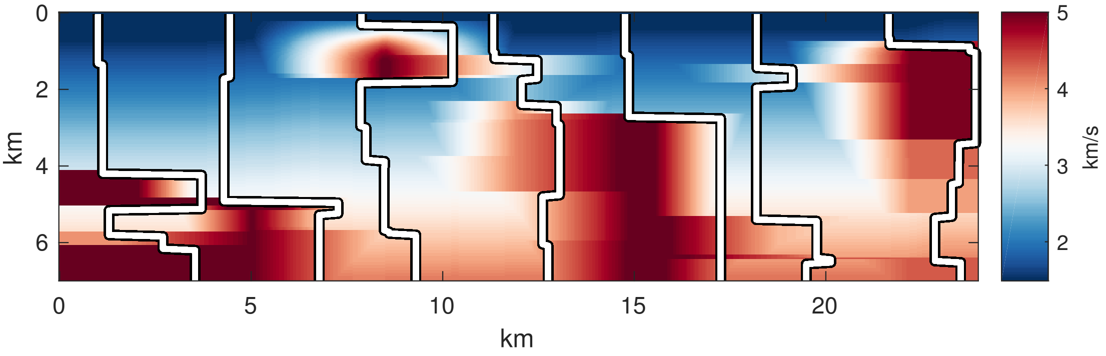


Figure 3.8: Random velocity model created by interpolation in-between random 1D velocity profiles.

The number of points to sample a random 1D profile, as well as the number of profiles needed to produce a velocity model with dominant lateral structures, is defined by the minimum and maximum frequencies to be extrapolated. We compute the range of expected wavelengths within the physical dimension of the velocity model, and then use these wavelengths to limit the thicknesses of layers within a profile. Random profiles are then stacked side-by-side and reshaped to match the size of the original velocity model where the acquisition takes place.

We also introduced a linear background trend into all generated random subsurface models such that it serves as a baseline to limit the lowest velocities in the generated models. To that end, we conducted a rough grid search for trends that minimize the misfit between observed and modeled data. Each newly generated random model is then compared, in terms of Euclidean distance, with the set of all previously built models to ensure a more even sampling of the model space.

Normalization of features within the dataset is crucial for successful training of any deep learning model. Proper scaling of the data accelerates training and equalizes contributions from each feature in the training stage. We split complex valued wavefield into its real and imaginary parts and scale each of them individually to fit the range $[-1, 1]$. Scaling coefficients are saved and then applied when needed to restore the data values to the original scale that is suitable for FWI applications.

3.4.2 CNN architecture

We design and train each CNN to serve as a non-linear extrapolator from high- to low-frequency representation of a shot gather. Figure 3.9 shows our CNN architecture, which consists mainly of four convolutional blocks with two fully-connected layers at the end. Input and target data to the network are represented by a complex-valued matrix and a vector, respectively (as shown in Figure 3.7). For the CNN, each convolutional block includes two consecutive convolutional layers sharing the same set of hyperparameters. Each block is followed by a batch-normalization and a max-pooling layer aimed at reducing the dimensionality and equalizing contributions from each block to the learning process. A stack of convolutional layers with small-sized kernels is more efficient in capturing larger features in the dataset rather than an equivalent single layer with a larger kernel size [128]. This means that a sequence of two consecutive convolutional layers with a stride of 1 and receptive field of 3×3 is equivalent to a single layer with a receptive field of 5×5, whereas fewer trainable parameters which are beneficial for preventing overfitting. A set of convolutional blocks compresses input data into a latent space which is then fully-connected by two dense layers to shape the target output.

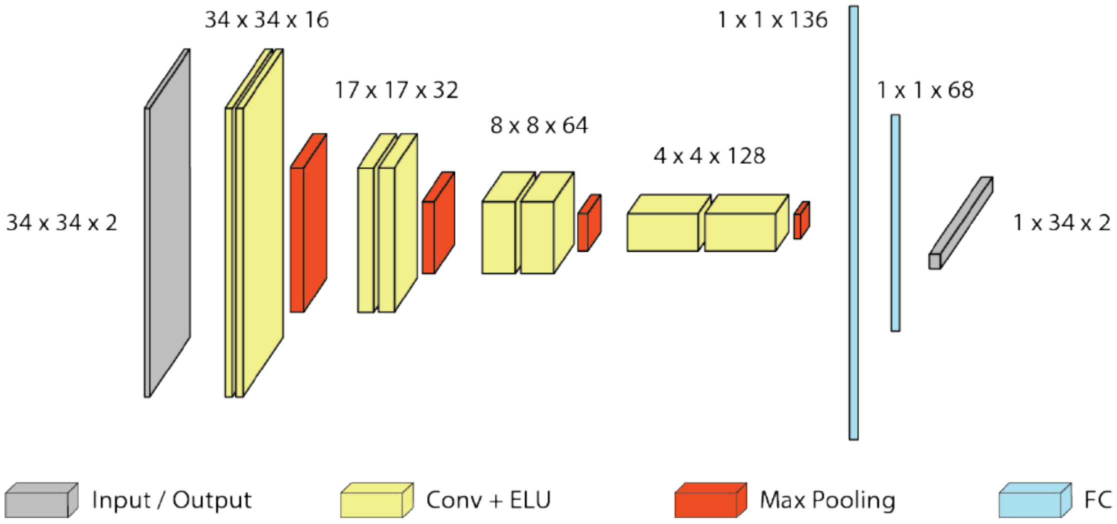


Figure 3.9: Architecture of a convolutional neural network designed for low-frequency extrapolation. There are four convolutional blocks followed by two fully-connected layers.

In more details, there are 16, 32, 64 and 128 kernels used in each of convolutional blocks. We pad each convolution with zeros and use exponential linear units [129] as an activation function for all convolutional layers. The output from the last convolutional layer is flattened into a vector of 512 units, which is then passed to the fully-connected part of the network composed of layers with 136 and 68 neurons, respectively. At the latest stage, we reshaped the output from the last layer of the network to be 34×2, which stands for one complex value for each of the 34 receivers in the streamer line. To meet the range of normalized training data ($[-1, 1]$), we equip the first of two dense layers with a hyperbolic tangent activation function that spans the same interval. The last dense layer, which has a linear activation, serves as a summator for its inputs. Such architecture leads to a total of 373,212 trainable parameters.

Finally, we train the network using Adam's optimizer [130], with a batch size of 32 and a learning rate dropping down to 10^{-5}. We initialize random weights following [131], and use early stopping to prevent overfitting when training does not advance for more than six iterations. For the numerical implementation of the CNN, we used Keras [132] with a TensorFlow [133] backend.

3.5 Examples

In this section, we illustrate our low-frequency extrapolation approach for the central and left parts of the BP 2004 [92] and SEAM Phase I [93] benchmark models. To demonstrate its validity, we conduct a multi-scale acoustic FWI with the extrapolated low frequencies and show results for the central part of BP 2004 velocity model.

The central segment of the BP 2004 model is 24 km long and 7 km deep. Surface acquisition involves 68 collocated sources and receivers evenly placed with 320 m spacing. Input and target data for the network are assembled from samples of equal size extracted from the acquisition data matrix, according to the scheme shown in Figure 3.4(b) where we limit the maximum offset in the data is limited to half of the model width (which is 12 km). This mimics a standard streamer acquisition and allows us to process all the shots in an efficient manner, using reciprocity to process all data that are not inferred directly. All benchmark sections share the same acquisition geometry and model dimensions, the only difference being in the distribution of acoustic parameters.

3.5.1 Dataset

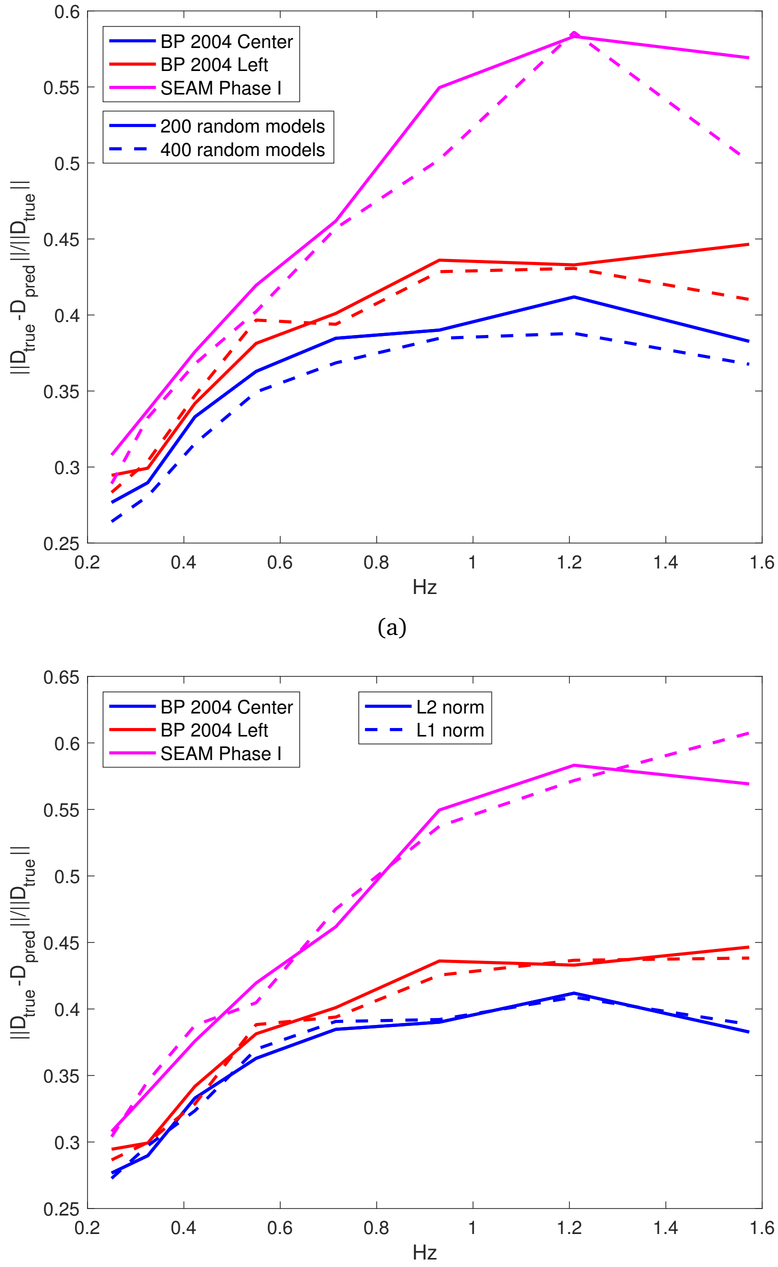
The synthetic dataset includes shot gathers from marine surveys generated for 400 random subsurface velocity models. With 68 unique receiver locations, whereas the network extrapolates mono-frequency data for only 34 of them at a time. Thus, to shape a squared input volume for the network, we need to model 34 mono-frequencies for each shot gather, within the known range from 2 Hz to 4.5 Hz and a single target low-frequency for output.

Figures 3.10(a) and 3.10(b) compare the extrapolation accuracies of the network for multiple models, different training set sizes and different misfit definitions, respectively. From a total of 27,200 data pairs, 20% (5,440 samples) are used as a validation dataset, whereas the rest of 21,760 samples serve as a training dataset. The extrapolation accuracy is then measured as the total sum of all misfits between the target and inferred outputs for the validation set. When compared with a training dataset of twice this size, we find that the extrapolation accuracy of the network improves, although only slightly. This may be due to a relatively homogeneous dataset among the generated random velocity models, where more models only lead to little extra information.

Furthermore, we do not observe significant differences when solving the network optimization for an L_1 or L_2 misfit function. In general, we find that very low frequencies are better extrapolated. This is somewhat expected, as for longer wavelength signals, the effects of complex subsurface features become less pronounced, thus easier to detect for a general regression system.

3.5.2 Low-frequency extrapolation results

The ability to generalize data inference is one of the most beneficial features of artificial neural networks [134]. Complex non-linear relations learned by the neural network enable it to produce also reasonable results when applied to data it has never seen. We check this generalization ability of our networks, trained each for a single low-frequency extrapolation, by feeding them high-frequency data generated for several benchmark velocity models. All benchmark models are mimicking a realistic geology of salt-induced media, thus the same neural network should be able to infer the low-frequency signal for each of them.



(a)

(b)

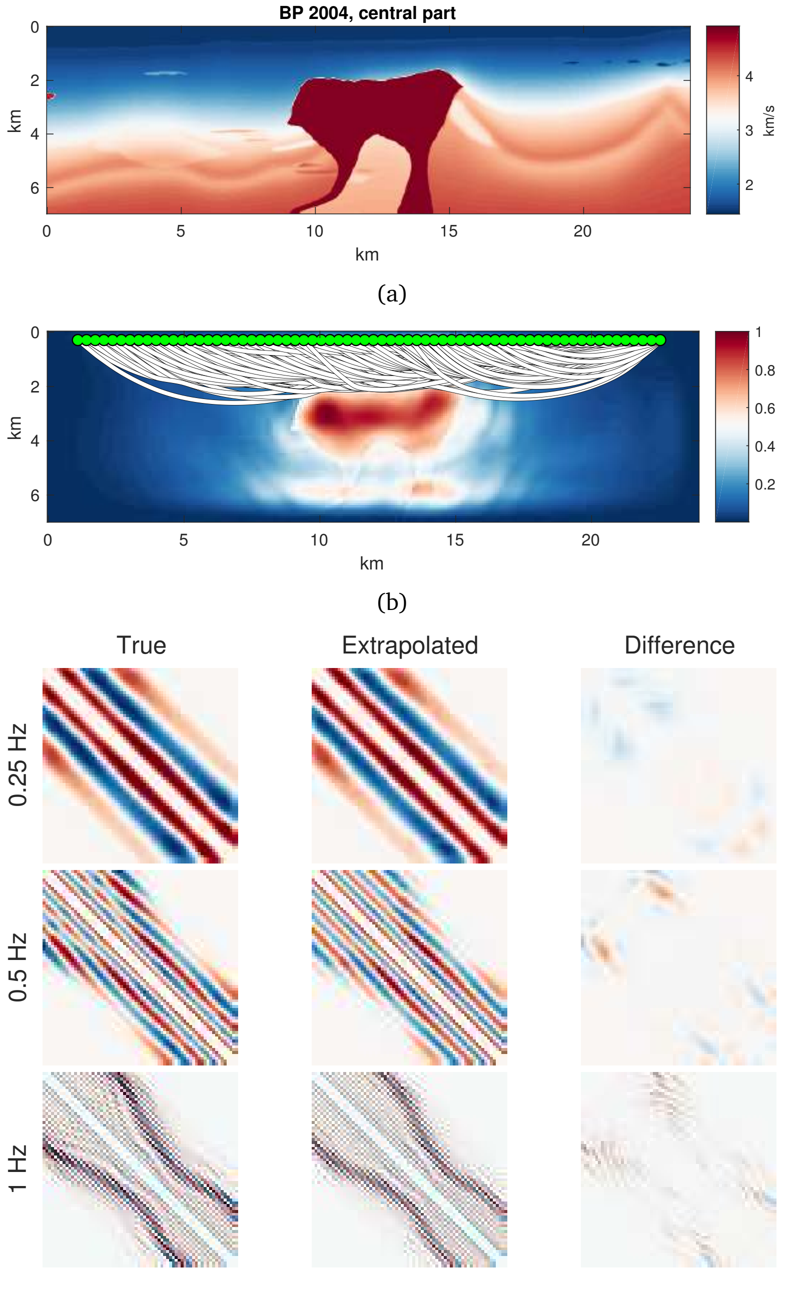
Figure 3.10: Comparison of misfits between target and inferred data for the networks trained on datasets built from 200 and 400 random velocity models (a), and optimization for L_1 and L_2 loss functions (b).

In the following, we compare real parts of true and extrapolated data matrices and their differences for 0.25, 0.5 and 1 Hz. We also compare data for central and left parts of the BP 2004 benchmark model, as well as for a section from the SEAM Phase I model. Extrapolated data for each shot gather in the acquisition, at different frequencies, are generated by individual CNNs. As previously shown in Figure 3.7, amplitudes of the extrapolated low-frequency wavefield for a shot gather are produced from high-frequency data within the known frequency range. The known frequency range is sampled by as many individual frequencies as there are receivers.

BP 2004 central. The massive salt body in the middle of the BP 2004 benchmark model is one of its distinctive features (Figure 3.11(a)). We extract this salt-containing section to show the benefits from using the extrapolated low-frequency data in such a complex geology. The main complexity comes from the presence of both flat and oblique flanks of the salt body, which cause strong reflections and multi-scattering. In particular, the strong reflections from the salt-sediment surface lower the signal energy propagating through the salt body, which then causes tomographic imaging techniques to fail. Additionally, the steep flanks cause multi-scattering, which becomes a challenge for methods relying on the Born approximation. Finally, there is a low-velocity anomaly below the salt, which may be a hydrocarbon collector. Targeting the inversion to this collector, we first extrapolate low-frequency data and will show its use in a multi-scale FWI.

As noted above, the maximum offset in the data is limited to half of the model width (12 km). Due to this limited acquisition offset, the illumination in deep parts of the model becomes poorer (Figure 3.11(b)). We can therefore expect that the neural network training will mostly corroborate connections in the data, due to variations in the upper part of the velocity model, where data coverage is the highest.

Extrapolated low-frequency data (Figure 3.11(c)) follows the major trends in the true data. The mismatch between data matrices increases at higher frequencies, possibly caused by more complex contributions of subsurface features into the total misfit. During the training stage of the network, we search for the minimum of the L_2 loss function that is mostly sensitive to the largest deviations in the data. However, we find that



(a)

(b)

(c)

Figure 3.11: Central section from the BP 2004 benchmark velocity model (a). Normalized illumination in the model accounting for truncated half-offset acquisition. White lines indicate ray paths of diving wave, green points indicate locations of collocated sources and receivers (b). Real part of data matrices for extrapolated low-frequencies at 0.25, 0.5 and 1 Hz (c).

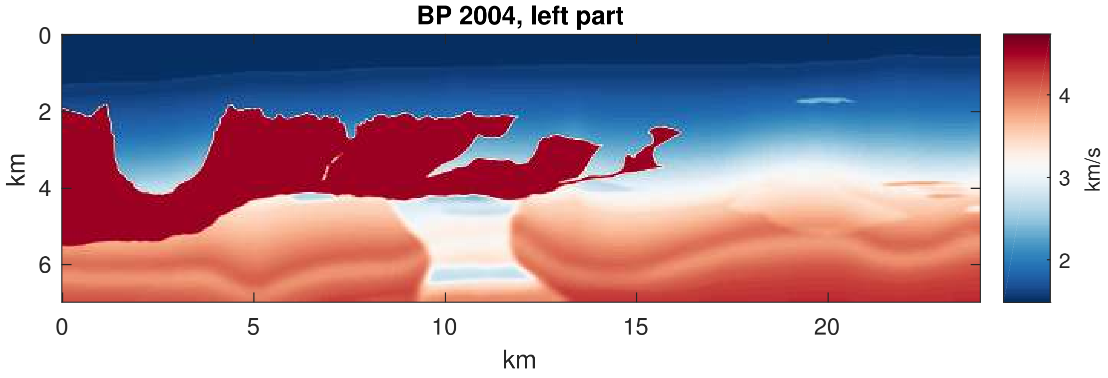
optimizing for a more sensitive L_1 misfit does not improve the overall extrapolation results (Figure 3.10(b)). From other side, parametrization of the input data defines features of the wavefield that contribute to the norm. In this study, we treat the complex wavefield by explicitly feeding its real and imaginary parts to the network. A different parametrization of the wavefield by its amplitude and phase didn't lead to significantly different extrapolation results, however there is still a room further exploration. For this reason, in future studies, we plan to investigate both misfit functions that are more sensitive to phase variations and other parametrizations of the data.

BP 2004 left. In the section from the left-side of the BP 2004 model, there is an elongated salt body covering almost half of the section's bottom part (Figure 3.12(a)). Without low-frequency data, conventional FWI stagnates, retrieving only the top of the salt and corrupting the rest of the image, due to a lack of illumination. This is another case where low-frequencies would highly contribute to the success of full-waveform inversions.

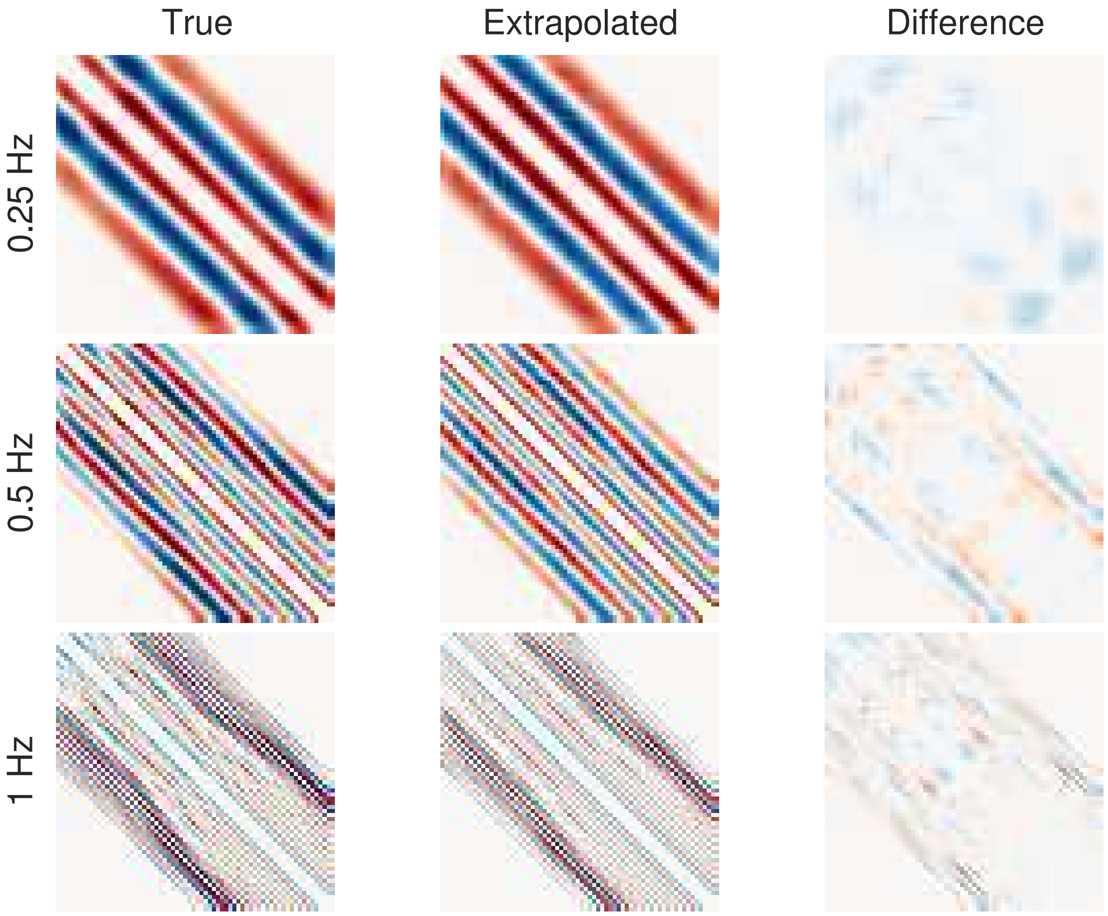
Distinctive features in the data matrix in Figure 3.12(b) are parallel stripes, shaped by off-diagonal elements. These are built by strong reflections coming from the top of the salt in which its surface is lying below water and a thin layer of sediments. Extrapolation results show a mismatch in fine details at higher frequencies, whereas smooth data is well-extrapolated at lower frequencies.

SEAM Phase I. Despite some similarities in the placement of the salt body, at the center of the velocity model, data matrices from the SEAM Phase I (Figure 3.13(a)) and the central section from the BP 2004 model (Figure 3.11(a)) look very different. The shape of the salt surface, as well as the water depth, strongly affect the structure of data matrices, due to reflections of different amplitudes arriving at receivers with different phase shifts.

Extrapolated data for the section from the SEAM Phase I model (Figure 3.13(b)) show a fit at low frequencies that is worse than the fit for other test examples listed above. The largest errors occur in parts of the data matrices where seismic sources are placed right above the salt surface (centers of top and bottom edges of the data matrix). A reason for this is the resizing (for testing purposes) of the original model, in order

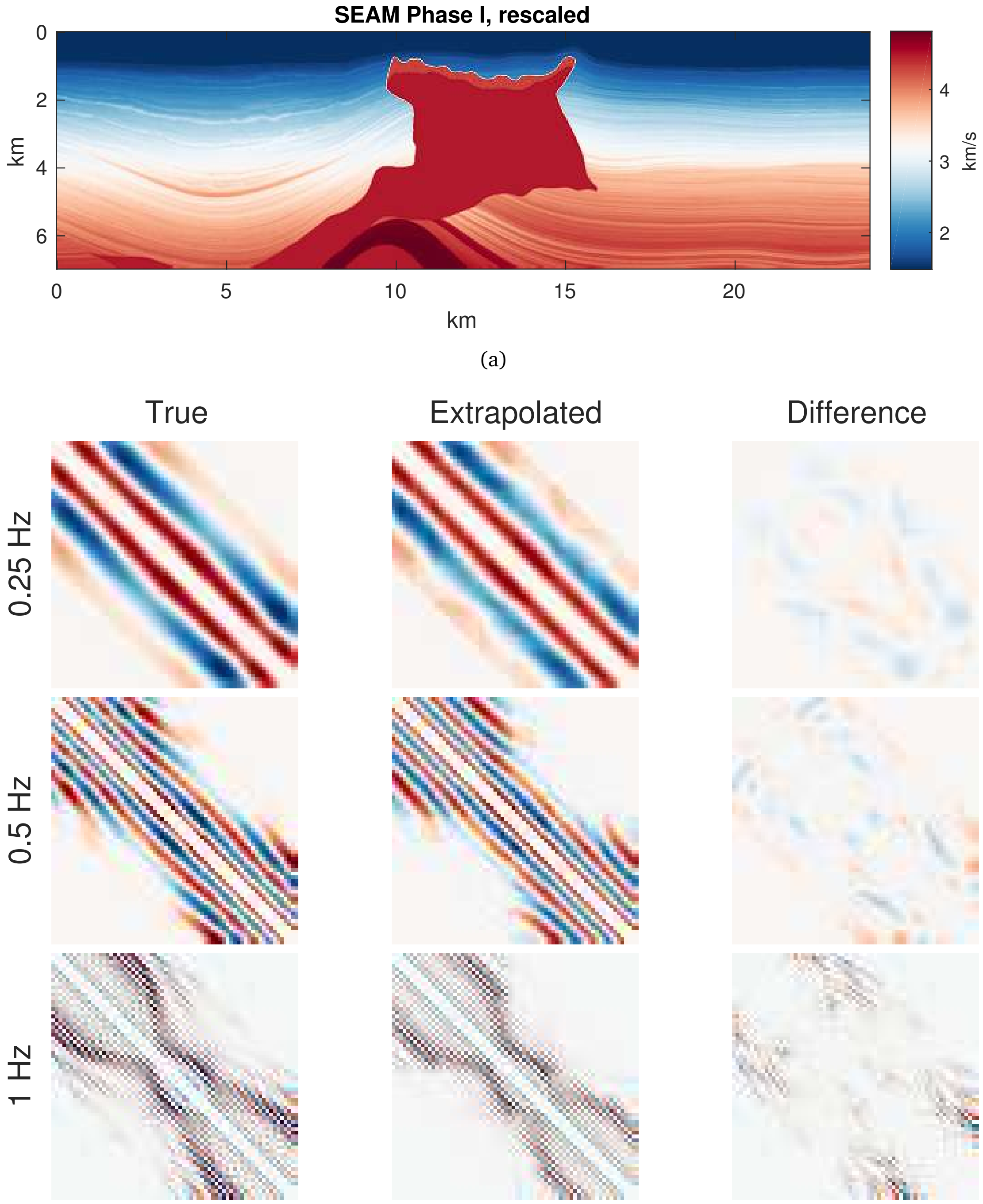


(a)



(b)

Figure 3.12: Left section from BP 2004 benchmark velocity model (a). Real part of data matrices for extrapolated low-frequencies at 0.25, 0.5 and 1 Hz (b).



(a)

(b)

Figure 3.13: A section from SEAM Phase I benchmark velocity model (a). Real part of data matrices for extrapolated low-frequencies at 0.25, 0.5 and 1 Hz (c).

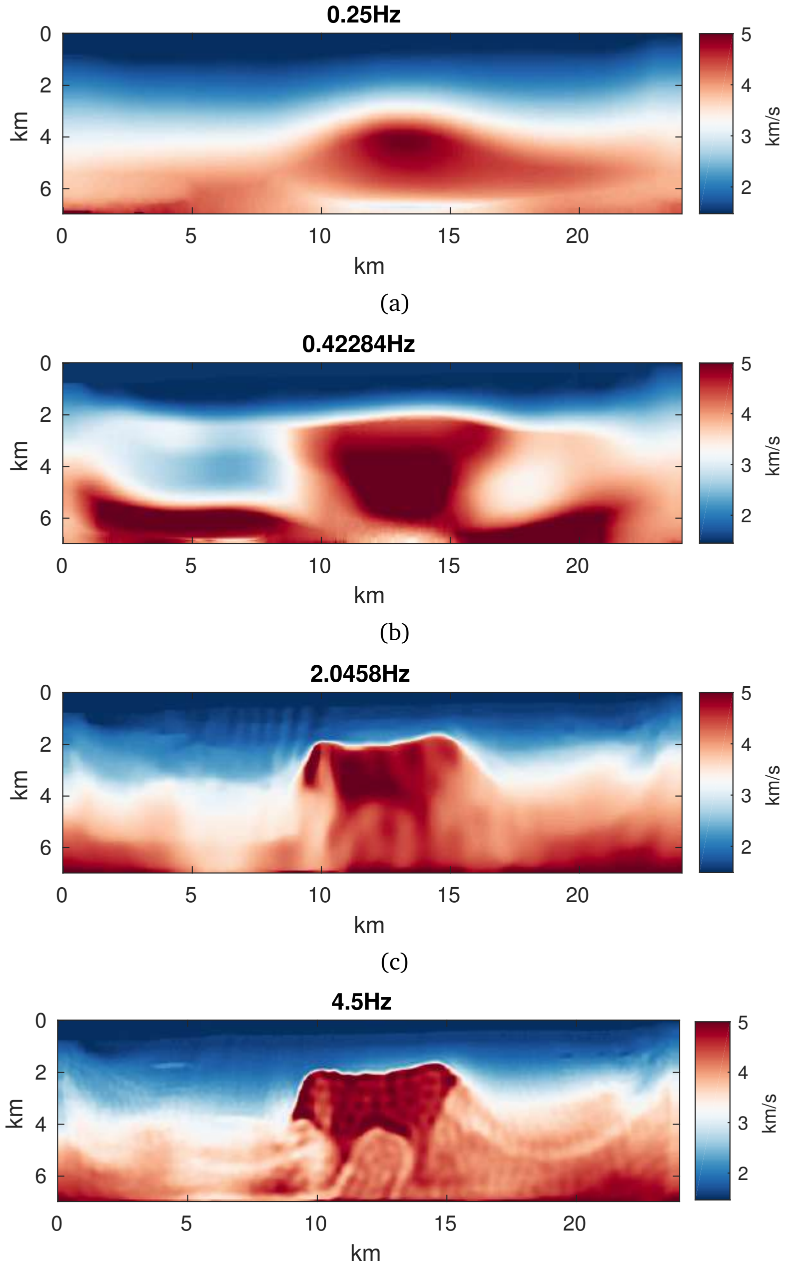
to fit the same dimensions as those for the randomly generated data, which results in a very shallow water layer above the salt. For low frequencies, such a small gap is less than a quarter wavelength in the water, and thus becomes negligible, leading both the modeling of the synthetic training data, as well as the extrapolation by the network, to poor reconstruction results for those particular areas.

3.5.3 FWI application

To test the usefulness of the extrapolated low-frequency data, we follow a multi-scale approach running an acoustic FWI [26] for the central part of the BP 2004 benchmark velocity model (Figure 3.11(a)). We initiate the inversion at the lowest extrapolated frequency, and use the inversion result at that frequency as an initial model for FWI at the next higher frequency. The set of frequencies used in this strategy is that of a geometric sequence. Assuming that observed data are available within the range from 2 Hz to 4.5 Hz, we extrapolate the missing data needed for this multi-scale approach, starting from 0.25 Hz with a geometric multiplier of 1.3. The corresponding total sequence consists of twelve frequencies, with eight extrapolated ones {0.25, 0.33, 0.42, 0.55, 0.72, 0.93, 1.21, 1.57} Hz, and four within the known range {2.05, 2.66, 3.46, 4.50} Hz.

The CNN learns to map a single low frequency from a set of high frequencies. Therefore, to extrapolate eight target low frequencies used for the multi-scale approach, we train eight neural networks sharing the same architecture and input training data set. The only part that differs is the target data, different for each target low frequency. The inversion is then sensitive both to phase and amplitude of the extrapolated data. However, since these values may not be entirely accurate, we compensate for inaccuracies of poorly extrapolated data by adding an additional regularization term to the misfit function. As shown in the Figure 3.10(b), the extrapolation errors increase for higher frequencies, meaning that using poorly predicted data at later inversion stages will hamper inverted velocity models. Thus, to successfully run the inversion we use the first five extrapolated low frequencies only, and fill the remaining gap by adding regularization terms to the misfit function [135].

The largest velocity anomalies in the resulting subsurface model are placed in their correct locations when inverting the lowest extrapolated frequency data of 0.25 Hz. According to Figure 3.10(b), the best match of true and extrapolated data is reached



(a)

(b)

(c)

(d)

Figure 3.14: Multiscale full-waveform inversion of low-frequency data extrapolated by CNN (a-b) and data from the known interval (c-d).

for the lowest frequencies, which is confirmed in the image domain, when running an inversion of the extrapolated data. The resulting image from FWI, at the highest frequency of 4.5 Hz, recovers most features in the central part of the model. We also see the target low-velocity arch, below the massive salt body. Examples of constrained FWI applied to similar benchmark models are shown in [14, 110, 11].

To summarize, FWI applied to extrapolated low-frequency data converges to a reasonable initial model, which then leads the inversion at higher frequencies to a more robust final velocity model than that obtained without extrapolated frequencies. As a result, we see the current neural network approach to low-frequency extrapolation as being potentially interesting for FWI applications in complex subsalt models.

3.6 Application to noisy data

Seismic data are always corrupted by noise in real-world scenarios, thus we want to examine the capability of the network to treat noisy data.

We train the network on synthetic data for two scenarios. First, we add an artificial Gaussian noise, SNR=14, to the input part of the training dataset and subsequently test the network on a dataset with SNR=7, for the central part of BP 2004 benchmark model (Figure 3.11(a)). A reduced level of added noise, during training acts as a regularization [136], which helps to prevent overfitting and to handle noise at the inference stage. Second scenario demonstrates the case when the network uses a noise-free dataset for training, whereas noisy data is given as input for inference.

Low-frequency data reconstructed from noisy data exhibit features similar to those in true data for both scenarios (Figure 3.15). Noise in the high-frequency data affects mostly the amplitude of the extrapolated low-frequency data, whereas the phase remains close to the inference results from noise-free data. However, the network trained on clean data did not succeed in extrapolating for the higher frequencies from noisy input data.

3.7 Discussion

Deep neural networks are powerful data-driven mathematical models able to derive relations directly from data. This allows accounting for realistic physics such as visco-

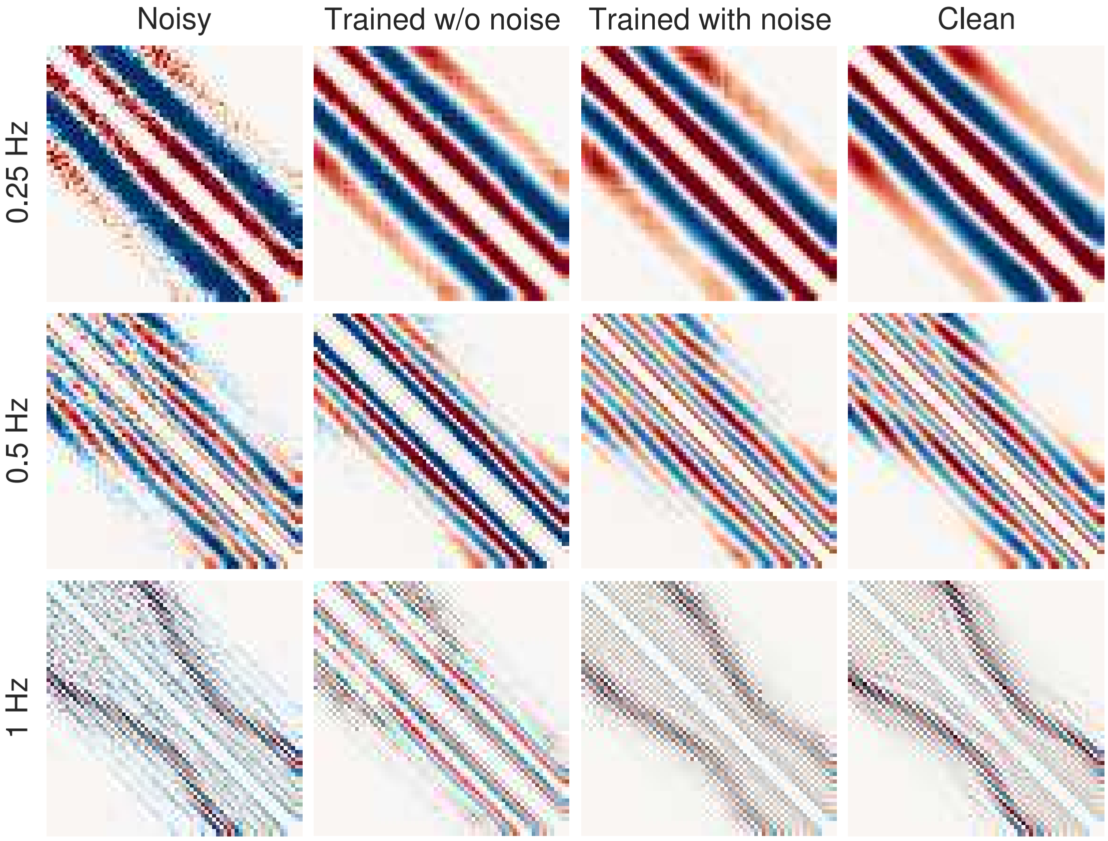


Figure 3.15: Extrapolation of low-frequency data for 0.25, 0.5 and 1 Hz from noisy high-frequency data within 2-4.5 Hz range, and for a SNR value of 7. The network was trained for two scenarios, being given input training data without noise and with noise. The later application delivers better inference result.

elastic anisotropic effects, as no explicit modeling is involved at the inference stage. We use an acoustic wave equation for modeling, which is most likely not realistic. A common drawback is that there is no unambiguous approach to determine the architecture and set of hyperparameters of the neural network. The architecture plays an important role, but there are only empirical practices on how to design it. The structure of the CNNs we implemented in this study was motivated by the need to compress input data into a latent space, and then map it onto a space with the dimensions of the target data. It is therefore an exploratory setup, where the current experience can only guide to future improvements in such a machine learning framework. Transfer learning paradigm also offers promising options to facilitate the search for a deep learning setup for frequency extrapolation [137]. Under this approach, a deep learning model designed and trained for one task might be applied for another conceptually similar task, however at lower re-training cost.

Generation of synthetic training data can be prohibitively expensive when attempting to train a network to extrapolate a complete low-frequency dataset from entire survey data recorded for numerous seismic sources. In such scenario of survey-to-survey mapping, one random velocity initialization only delivers one training sample. However, targeting to interpolate for a shot gather in a single pass, rather than for entire survey, drastically reduces the time needed to generate a training dataset, while preserving a promising extrapolation capability. For a single random subsurface model, we generate as many training data pairs as there are sources. We benefit from this approach even more when we run the wave propagation modeling in the frequency domain where computational costs do not scale as fast as that in the time domain, when modeling for multiple sources. Still, it remains unclear which particular features from the synthetic dataset generated with approximate physics – purely acoustic in this study – can help in training neural networks for them to be able to generalize data inference across different subsalt acquisitions.

Time-domain data is natural for real-world seismic surveys. However, diversity in duration and sampling rates of recorded traces lead to variable input dimensions of data and make it challenging to directly use such time-domain data in ML applications, unless these are designed to process time series data. Proper preprocessing and compression of time-domain data seems to be the key point to address prior to developing an deep learning model. Otherwise, conversion of time-domain data to frequency domain also leads us to the sufficient framework for frequency bandwidth extrapolation.

Although low-frequency extrapolation results are not entirely accurate, we were still able to demonstrate that, if used within a multi-scale inversion strategy combined with additional regularizations, they can help the FWI to correct large scale-length velocity features in the initial model. The neural network approach demonstrated here has the advantage of being a purely data-driven regression tool, able to generalize across different seismic acquisitions. It therefore minimizes assumptions on data connectivity or physics-related extrapolations. However, implicitly, we still include *a priori* geological information in form of the dataset generated for the training stage. Future research needs to evaluate how this sampling of the model space affects, and may improve the network inference accuracy. That said, the current neural network extrapolation results help to further enhance the robustness of FWI in complex subsurface regions.

With recent advances in the domain of multiparameter scattering in the wavenumber domain the theory can easily be extended to elastic and anisotropic cases. Wave propagation on HPC architectures, in frequency and time domains, paves the way for the extension of the numerical part of the studies. In future studies, we also plan to apply the method to real data.

3.8 Conclusions

We explain the feasibility limits of the frequency bandwidth extrapolation from a wavenumber illumination perspective. Through wavenumber analysis, we show that high-frequency data recorded at long offsets are linked, through the subsurface velocity distribution, to low-frequency data at short offsets. Therefore, bandwidth extensions of individual traces are not viable without strong constraints on the model. However, extensions of full data sets are possible, assuming that the inverse problem is, in principle, resolvable with the available data.

We propose the framework for multi-offset low-frequency data extrapolation. Treating the entire acquisition as a collection of independent shot gathers is beneficial for several reasons. First, shot-wise low-frequency data extrapolation makes the proposed neural network-based technique applicable without retraining to a range of exploration setups with fixed-offset geometry such as in marine streamer acquisition. Second, the generation of the synthetic training dataset in the frequency domain drastically reduces computational costs as we are able, from a single velocity model, to produce as many training data pairs as there are sources.

We designed a deep convolutional neural network to generate a low-frequency representation of a shot gather, given its high-frequency part, and trained it on data generated from random velocity models. Inputs and outputs from the network are multiple high-frequency and single low-frequency representations of a shot gather, respectively. We subsequently apply the trained network to extrapolate the spectrum, for several benchmark data sets. Extrapolation excels at capturing general trends in the data, which is enough for extrapolating very low frequencies. On the other hand, fine features at higher frequencies are missing. Finally, we run an acoustic FWI for a benchmark model, using the extrapolated low-frequency data. The synthetic inversion tests show that the artificial data of low frequencies are accurate enough to correct most of the large scale-length error in the initial model, and help FWI to converge.

Chapter 4

Multi-task learning for low-frequency extrapolation and elastic model building from seismic data

Time domain is natural for seismic data. However, the wide range of amplitude variations within the desired low-frequency range in time domain causes challenges for data-driven methods. In particular, amplitudes of long-wavelength signals in the lower part of a selected frequency range might be negligible compared to those from the later part of the same range. Thus, the key problem for low-frequency extrapolation in time domain is to focus the training of a deep learning model on weak long-wavelength signals while keeping higher frequencies also accurately reconstructed. Rather than aiming to reach superior accuracy in bandwidth extrapolation, this chapter proposes to jointly reconstruct low-frequency data and a smooth background subsurface model within a multi-task deep learning framework. We automatically balance data, model and trace-wise correlation loss terms in the objective functional and show that this approach improves the extrapolation capability of the network. We also design a pipeline for generating synthetic data suitable for field data applications. Finally, we apply the same trained network to synthetic and real marine streamer datasets and run an elastic full-waveform inversion from the extrapolated dataset, highlighting the potential of this novel approach.

Results presented in this chapter are adaptation of the paper submitted to IEEE Transactions on Geoscience and Remote Sensing.

Ovcharenko, Oleg, Vladimir Kazei, Tariq Alkhalifah and Daniel Peter. "Multi-task learning for low-frequency extrapolation and elastic model building from seismic data" *Undergoing review in IEEE Transactions on Geoscience and Remote Sensing*, 2021.

4.1 Introduction

Seismic waveforms recorded at the surface are the primary source of information about the Earth's interior [138]. If the subsurface elastic properties are known, then seismic waveforms can be simulated and compared to the ones recorded in a field experiment. Full-waveform inversion (FWI) is a technique that optimizes hypothetical subsurface properties such that every wiggle in the recorded seismic data matches the data simulated using these subsurface properties [90]. Due to its versatility and ability to handle any realistic data, FWI is dominating seismic imaging in the past decade in both seismological [139] and exploration communities [140, e.g.]. FWI optimizes the model of the Earth so that it explains the data using local search for optimal parameters. For such a non-linear optimization problem the two key requirements for success are the right data to optimize towards and the right starting point or initial model [95]. When none of these is available the inversion suffers from cycle-skipping phenomena when wiggles of synthetic data are matched with their wrong counterpart from observed data [141]. This is the case in reality when all that is typically available is band-limited seismic data contaminated by noise. Therefore there are two tasks that need to be completed to set up FWI:

1. Estimation of a realistic initial model and source wavelet
2. Conditioning the data or re-defining features that should fit between simulated and recorded data

While these tasks are separate, they are tangled with each other. Namely, the solution for the first task leads to requirements for the second one. For example, if the initial model for inversion is very close to the actual subsurface, then conventional FWI would work without low frequencies. On the other hand, if the initial model has a major seismic shadow in a wrong place (for example if a salt body is mispositioned), then there is almost no way that available band-limited data can correct the non-illuminated areas of the model as there is no sensitivity from the synthetic model.

The vast majority of FWI formulations assume the availability of a decent approximation of velocity trend (e.g. extracted from regional logs) and focus on FWI and/or data improvements. The data misfit for FWI might be constructed to compensate for the missing low-wavenumber information [8, 142, 143, 144, 145]. Alternatively, constraints applied alongside the main objective might serve the same goal [12, 146, 147, 148, 11]. Conditioning and smoothing of gradients for model updates also help inversion to tolerate inaccurate initial starting model for inversion [149, 110, 18, 148].

Classic approaches brought the dawn of FWI in seismic exploration, however, they either require a different inversion engine to be used (e.g. advanced traveltime tomography) or require more iterations of high-frequency seismic simulations. With multi-task learning, we aim at correcting the data and estimating the initial model so that elastic FWI works.

4.1.1 Low-frequency seismic data

In a perfect world, where a broadband signal illuminates the target domain from wide angles, FWI would converge to a high-fidelity reconstruction of the medium. In reality, only offset- and band-limited data recorded at the Earth's surface is available for inversion. Altogether, this makes FWI a highly nonlinear optimization procedure suffering from the non-uniqueness of the solution. Numerically, this corresponds to the presence of multiple local minima of the objective function [150] where the optimization algorithm might get stuck on its way down to the optimal solution resting in the global minimum.

Recall, that the background velocity model derived from tomography compensates for frequencies below 1 Hz [55]. While the frequency content recorded in a generic marine airgun survey drops below noise level at about 4 Hz [56]. The frequencies falling in-between these estimates correspond to the gap in the model wavenumber spectrum [57]. Reconstruction of these low frequencies is the objective of this study.

4.1.2 Reconstruction of missing low-frequency data

Data-driven reconstruction of missing low-frequency energy might be approached in time and frequency domains. A time-domain seismic signal is a composition of multiple independent mono-frequencies. Thus, bandwidth extrapolation in the time domain aims to simultaneously recover a range of such frequencies. The headwind of working in the time domain is that amplitudes of the signal at the lower end of the target frequency range are significantly weaker than those at the higher end. While amplitude balancing techniques such as automatic gain control (AGC) or spectral whitening might help in addressing the issue, these methods are sensitive to noise yet their applicability is still to be explored. Unlike the time domain, frequency domain representation of seismic data offers a discretized framework where every frequency component might be considered in its own context. Dimension of each mono-frequency data is less by one, compared to multi-frequency time domain data, thus allowing to train a weak dedicated network for each individual frequency [151, 69]. However, the application of such methods to time-domain data might be impractical since bandwidth extrapolation of each frequency would require training of its own small network.

Time-domain representation of seismic data is natural both in acquisition and processing. In particular, the majority of deep learning methods for bandwidth extension are focusing on the time/offset format of the data. [76, 152] proposed and then developed a trace-by-trace approach for frequency extrapolation in the time domain. The method operates on full-duration time series and is powered by WaveNet architecture. The approach is suitable for elastic waveform inversion in marine survey layout. [153] and [154] extrapolated low frequencies by training convolutional networks on AGC-balanced patches of time-domain seismic data in marine and land setups, respectively. [79] avoided using synthetic data and trained the UNet to translate knowledge of low-frequency data from OBN survey to band-limited shallow streamer data. [155] proposed to iteratively halve the central frequency of a gather by a recursive convolutional network. Since synthetic data does not accurately represent the field data, [156] developed the self-supervised learning pipeline where predicted low-frequency data is iteratively injected into FWI engine. The approach was further developed in [157]. [158] also trained a network in a self-supervised fashion to retrieve similarity in transition between high and low-frequency bands.

4.1.3 Reconstruction of the low-wavenumber initial model

A low-wavenumber model available for the target subsurface might compensate for missing low-frequency content in seismic data. A general requirement is that such a model should be sufficient to avoid cycle-skipping phenomena at the lowest frequency present in the data. Deep learning offers opportunities to estimate realistic initial models for FWI directly from the data. Most commonly, the objective for training is formulated as finding a non-linear mapping between a complete set of seismic data from a synthetic experiment and respective subsurface model. Meaning that the whole low-wavenumber model of the subsurface is predicted from seismic data representing the entire seismic survey [80, 81, 83, 86, 84, 85, 50]. In a synthetic setup, a subsurface model predicted in such a way might have a resolution similar to FWI since the training and application domain are close to each other. These methods are often tailored for specific survey geometry and should be re-trained when experimental setup changes. Assuming that the data from the limited aperture is sufficient to illuminate the local subsurface, [88] proposed to map a set of neighboring common-midpoint-gathers into the central vertical elastic profile. This approach manipulates shot-gather data rather than full-survey data that improves its applicability to a broader range of domains. This way, authors apply the same trained network to synthetic and field data with the requirement of having an identical configuration of input seismic data. Motivated by the broad application domain of limited-aperture problem formulation, we select a common-shot-gather as the minimal source of input data for a deep learning model and explore limitations of such formulation.

We attempt to jointly address tasks of time-domain low-frequency data extrapolation and building a tomographic model of the subsurface. Specifically, we design a deep neural network architecture and formulate a multi-task learning objective to simultaneously deliver two outputs.

The key contributions of this work include:

1. Method for generation of a synthetic dataset based on real-world marine data.
2. An architecture of a neural network for joint prediction of low-frequency data for a common-shot gather and smooth initial velocity model under it.
3. Multi-task training objective for joint data and model prediction.
4. Case study for the application of elastic FWI with ML inputs

The paper is organized as follows. First, we introduce the multi-task objective for joint low-frequency extrapolation and building a smooth background model. Then we explain the generation of synthetic datasets tailored for specific real-world marine streamer data. Finally, we showcase the application of the trained network on the synthetic Marmousi II model and real-world marine streamer data and run elastic FWI in these models.

4.2 Multi-task learning framework

Multi-task learning (MTL) is inspired by the human ability to indirectly deduct knowledge from related tasks [159]. For example, extracting the summary of a book can help identify the genre. From the deep learning point of view, MTL is equivalent to training a network to simultaneously perform several tasks while partially or completely sharing trainable weights in the neural network branches leading to each of these tasks. With hard parameter sharing, the MTL formulation forces the optimization to accommodate the common knowledge for the main and auxiliary tasks in the shared weights of the network. This potentially improves the generalization capability of the network on the main task by constraining the domain of suitable solutions [5]. Thus, MTL generally solves a multi-objective optimization problem, which requires the definition of a combined objective functional that is typically a weighted sum of functionals related to each task.

4.2.1 MTL loss design

Our objective is to jointly recover the low-frequency information for the entire common-shot gather (CSG) and to reconstruct the low-fidelity subsurface model covered by the streamer at the moment of shot excitation. We further denote it as a local subsurface model for each shot. Two widespread choices for the FWI misfit are the point-wise accumulated norms, and trace-wise correlation coefficients (a.k.a. L2 comparison of the normalized traces). Therefore, we measure the low-frequency seismic data fit by the sum of the point-wise L1 norm of the data difference (L_d) and correlation coefficient of the predicted and labelled traces (L_c). The predicted low-wavenumber velocity model serves as a starting FWI model. The quality of the prediction is constrained by the model loss measuring the closeness of the reconstructed subsurface to the ground truth synthetic model (L_m). Successful completion of this task can compensate for the incomplete reconstruction of ultra-low frequencies. Furthermore, the completion of these two tasks delivers both a starting point and extrapolated data for FWI.

The complete objective function for our MTL implementation is as following:

$$L = \mathbb{W}(L_d, \sigma_d) + \mathbb{W}(L_c, \sigma_c) + \mathbb{W}(L_m, \sigma_m) + L_{reg} \tag{4.1}$$

where L_d is the difference-based data loss, which treats each pixel in the data independently. The second term, L_c, also operates in the data domain but it promotes scale-independent trace-wise correlation. The last term, L_m, is responsible for fitting the local subsurface model. $\mathbb{W}$ is the weighting operator with trainable parameter σ_i that dynamically balances training in between tasks, described below. Moreover, we add a total-variation regularization L_{reg} into training to promote both sparsity in the predicted model, as typically done in FWI [14, 11, e.g.], and ultra-low frequency components in seismic data, as in compressive sensing [160, e.g.]. Let us explain each component of the multi-task loss in more detail.

Weighting the tasks

Following [161], we define the weighing operator for a regression loss L_i as

$$\mathbb{W}(L_i, \sigma_i) = \frac{1}{2\sigma_i} L_i + \log \sigma_i, \tag{4.2}$$

with a task-uncertainty related parameter σ_i, i.e., for the data, σ_d, model, σ_m, and trace-wise correlation, σ_c. The value of σ_i quantifies the error/uncertainty associated with the prediction for the task i. When the uncertainty σ_i increases, the weight for the respective loss term L_i decreases. This effectively reduces the contribution of the gradient with respect to L_i into the minimization of the multi-objective functional. By this, contributions of multiple loss terms can be adjusted on-the-fly to enable an uncertainty-driven automatic loss balancing. In practice, each σ_i is a scalar that is trained alongside the network weights.

4.2.2 Difference-based data loss

The first objective of the training is to predict time-offset low-frequency data. In particular, we aim to reconstruct the wavefield recorded for frequencies below 5 Hz given input seismic data filtered above 4 Hz (Figure 4.2). The intentional overlap of corner frequencies of band-pass filters eliminates the gap between known and unknown bands caused by filter design.

We use a mean-absolute-error (MAE) loss for element-wise comparison of predicted and target volumes of seismic data. Hereafter, we refer to a general pair of target and predicted data as x and y, respectively

$$L_d = |d - \hat{d}|_1, \quad MAE(x, y) = \frac{\sum_{i=1}^{N} |x_i - y_i|}{N}, \tag{4.3}$$

where d is the true data and $\hat{d}$ is predicted data by the deep neural network. The network is trained to reconstruct the CSG data and deliver a high correlation of individual traces in the CSG. Unfortunately, far-offset traces typically have lower amplitudes and thus smaller contributions, which challenges their predictions. Still, far-offset traces carry diving waves with important information for low-frequency FWI [116] and thus need to be reconstructed more carefully. To address this issue, we add in the following a *correlation* loss term.

4.2.3 Correlation loss

The MAE loss treats seismic data as a collection of independent data points, linearly attributing more weight to larger amplitude mismatch. To reduce the amplitude dominance of short offsets over far offsets where the signal is generally weaker, we add an auxiliary loss term that measures the trace-wise correlation of the signal along the time axis.

Pearson coefficients quantify the linear relationship between two variables, ignoring bias and scale. For seismic data with a zero mean, the Pearson coefficient is equivalent to a cosine similarity measure. Cosine similarity is a commonly used scale-independent metric, popular in the computer vision community. Effectively, it normalizes a pair of traces by their norms and finds angles as if traces were vectors in a multi-dimensional space. Since the zero-mean assumption might not be met when multiple arrivals are recorded by the same receiver, we use the Pearson coefficient, as a more general formulation of cosine similarity, defined by

$$\rho = \frac{cov(x,y)}{s_x s_y}, \quad cov(x,y) = \frac{\sum_{t=0}^{N}(x_t - \mu_x)(y_t - \mu_y)}{N}, \tag{4.4}$$

where $cov(x,y)$ stands for the covariance between two traces, with μ_α and s_α denoting the mean and standard deviation for a given trace, respectively,

$$s_\alpha = \sqrt{\frac{\sum_{t=0}^{N}(\alpha_t - \mu_i)^2}{N}}, \quad \mu_\alpha = \frac{\sum_{t=0}^{N}\alpha_i}{N}. \tag{4.5}$$

The summation over t means counting over N temporal samples in a trace, α stands for either x or y.

When applied for each trace in a CSG, ρ becomes a vector of coefficients ranging from -1 to 1, between a perfect phase mismatch and match, respectively. The correlation loss between ground truth, d, and predicted seismic data, $\hat{d}$, might be formulated in a straightforward way

$$L_c = 1 - \frac{\sum_{k=0}^{K}\rho(d_k, \hat{d}_k)}{K}, \tag{4.6}$$

where the subscript k denotes an individual trace from K traces in a CSG. The range of L_c is from 0 to 2, for perfect match and mismatch, respectively.

4.2.4 Model loss

Without additional amplitude balancing the quality of extrapolated ultra-low-frequency seismic data is insufficient for FWI to converge from a 1-D initial model. In particular, reflections at frequencies < 2 Hz are poorly resolved. The lack of such low frequencies might be generally compensated by the availability of low-wavenumbers in the initial model, reducing the effect of the unresolved low-frequency content.

A single CSG contains sufficient data to recover a layered structure of the subsurface in the vicinity of the shot location. Based on that, we formulate the subsurface fitting term similar to the data matching term by

$$L_m = |m - \hat{m}|_1. \tag{4.7}$$

It seeks to optimize the MAE loss by recovering the sparse smooth background model of the underlying subsurface, which intends to replace the presence of tomographic frequencies. Since a subsurface model for an individual shot gather can not be accurately recovered, we linearly average overlapping areas from all shots to build an initial model for FWI. The resulting smooth background model then has a resolution approximately equivalent to traveltime tomography, compensating for the missing ultra-low frequencies in predicted data.

Finally, we add as regularization a total-variation loss term that promotes smoothness in the predicted subsurface model as well as the ultra-low frequency component of seismic data,

$$L_{reg} = L_{TV} = |\hat{m}|_{TV} + |\hat{d}_u|_{TV}. \tag{4.8}$$

4.2.5 Architecture

We design a fully-convolutional architecture for joint prediction of the local subsurface model together with two cascaded bands of seismic data (Figure 4.1). The network consists of an encoder as well as data and model decoders. As an encoder, E, we use a modified multi-column structure by [162]. In particular, we keep the three branches that accommodate dilated convolutional layers with kernel sizes of 3, 5, and 7 and create a bottleneck by dropping all upsampling layers. The outputs from each branch are concatenated along the channel dimension and passed through another convolutional layer to shape the final encoded representation of the input data. We find that such a multi-scale decomposition of the input data is crucial to capture the weak-amplitude and long-wavelength trends in the input volume. The benefit is primarily due to the large perceptive field of dilated convolutional kernels.

There are two decoders, a data decoder and a model decoder. The data decoder is a stack of transposed convolutional layers spatially upscaling the encoder bottleneck into the dimensions of seismic data. The model decoder has a purely convolutional structure that preserves spatial dimensions of the encoder output and maps it into the single-channel model of the subsurface. Such a design decision of the model decoder promotes sparse subsurface velocities and reduces the number of trainable parameters.

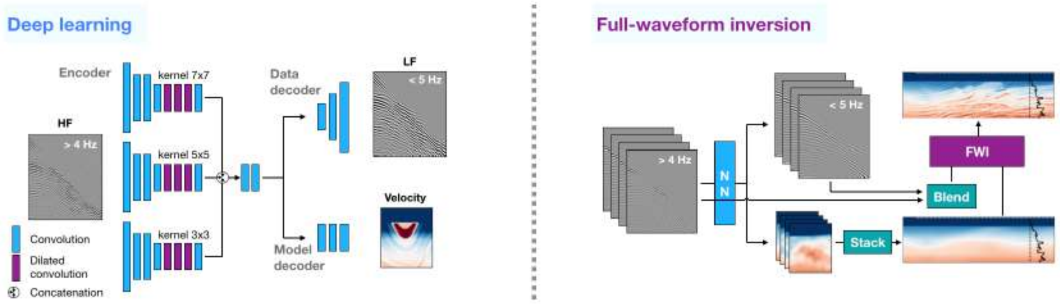


Figure 4.1: Multi-task network architecture (left) and inference workflow for extrapolated FWI (right). The input high-frequency data, HF, maps into low-frequency data, LF, as well as a sparse local subsurface velocity model. The blending block then substitutes extrapolated frequencies below 4 Hz into available seismic data. Local subsurface models are stacked into the initial subsurface model for inversion.

4.2.6 Implementation details

Similar to full-waveform inversion, the training step of a deep neural network is a non-linear optimization problem that is sensitive to the initial weights in layers of the network as well as to the set of hyper-parameters selected for training. Here we list the practical aspects that we found significant to this deep learning application.

Ensemble learning

Artificial neural networks are high-variance approximators prone to overfitting data [163]. Also, the nature of non-linear stochastic optimization makes training sensitive to the initial random state of network weights. As a result, the mapping between inputs and outputs learned by a deep neural network depends on weight initialization. Averaging predictions produced by having the same architecture, but differently initialized models, helps to reduce this bias from the initial set of weights. Consistent features are shared by all ensemble members, while initial weight-related errors in predictions from different models cancel out [164]. We average predictions from 10 identical networks initialized by different random seeds and notice that the cumulative prediction consistently outperforms individual predictions of the ensemble members.

Batch size

Larger batch sizes typically lead to a better load of computational units and reduce the generalization power of the network [123]. [165] analytically showed that large-batch implementations are prone to converging to so-called sharp local minima of the objective function, while small-batch implementations converge to smooth local minima. The sharp minimum implies the impossibility of an optimization method to escape the attraction basin. We performed tuning of the batch size and empirically found that a batch size of 4 delivers the lowest MTL loss on the testing dataset, and hence it is the most suitable for the proposed architecture and dataset size.

Weight initialization

Before training a convolutional neural network, its weights are typically set to small random values. However, the initialization method depends on the type of activation function used in the layer. Since we use the Leaky ReLU activation in all convolutional layers of our network, we follow the initialization method proposed by [166], also known as the Kaiming initialization. A proper range of initial weights prevents gradient vanishing problems as well as the problem of exploding gradients.

Learning rate

A properly selected learning rate policy improves the convergence rate and leads the iterative non-linear optimization to a deeper minimum of the objective function. The concept of super-convergence introduced by [167] suggests using the one-cycle strategy, which changes the learning rate for every batch, gradually increasing it from the initial warm-up pace to a large maximum rate, and then decreases it back to an ever lower value than the initial rate. The authors show that the large learning rate serves as an auxiliary regularizer when approached following the one-cycle strategy. We too observe in our experiments that the training reaches a deeper minimum of the objective function when using the one-cycle policy, rather than the reduce-on-the-plateau strategy that quickly overfits our data.

4.3 Seismic data

In a supervised learning framework, it is critical to minimize the domain gap between training and application datasets. For this reason, we copy acquisition design, noise imprint, and the source signature from a particular seismic dataset when generating our synthetic training data.

4.3.1 Marine streamer data

The target 2D marine streamer dataset was acquired in the North-Western part of the Australian continental shelf. Figure 4.2 shows a common-shot gather from the dataset and its average power spectrum. There are 1,824 CSGs in the survey, excited successively along a line with approximately 18.75 m spacing. The waveforms were then recorded by 648 hydrophones placed along the towed streamer every 12.5 m. We use the signal recorded for 6.2 seconds with 2 ms temporal sampling. The survey included the Broadseis acquisition system with a variable depth streamer [63], capturing a robust seismic signal above 2.5 Hz. Assuming that data below 4 Hz are unavailable, we use the frequency band [2.5, 4] Hz as a real-world reference to evaluate the bandwidth extrapolation results.

The source wavelet for each source location is estimated in the frequency domain following [168]. For the sake of simplicity, we assume the average wavelet signature to be shared among all sources. The general experimental layout for the marine streamer data experiment is following the description in [169].

4.3.2 Generation of synthetic data for a specific marine dataset

The workflow for synthetic data generation utilizes the survey geometry, the seafloor bathymetry, and the average source wavelet derived from the target field data. The general knowledge about this deep-water area of seismic exploration defines dimensions and elastic parameter distributions for a set of random subsurface velocity models. Meanwhile, the source signature and source-receiver configuration determine the layout for elastic wave propagation in each random subsurface model.

Random subsurface models

The generation of realistic subsurface models remains a non-trivial task. In detail, the properties of random models in Earth sciences were explored in [125]. [126] generated realistic seismic models using shuffling of coefficients in a wavelet-packet domain and

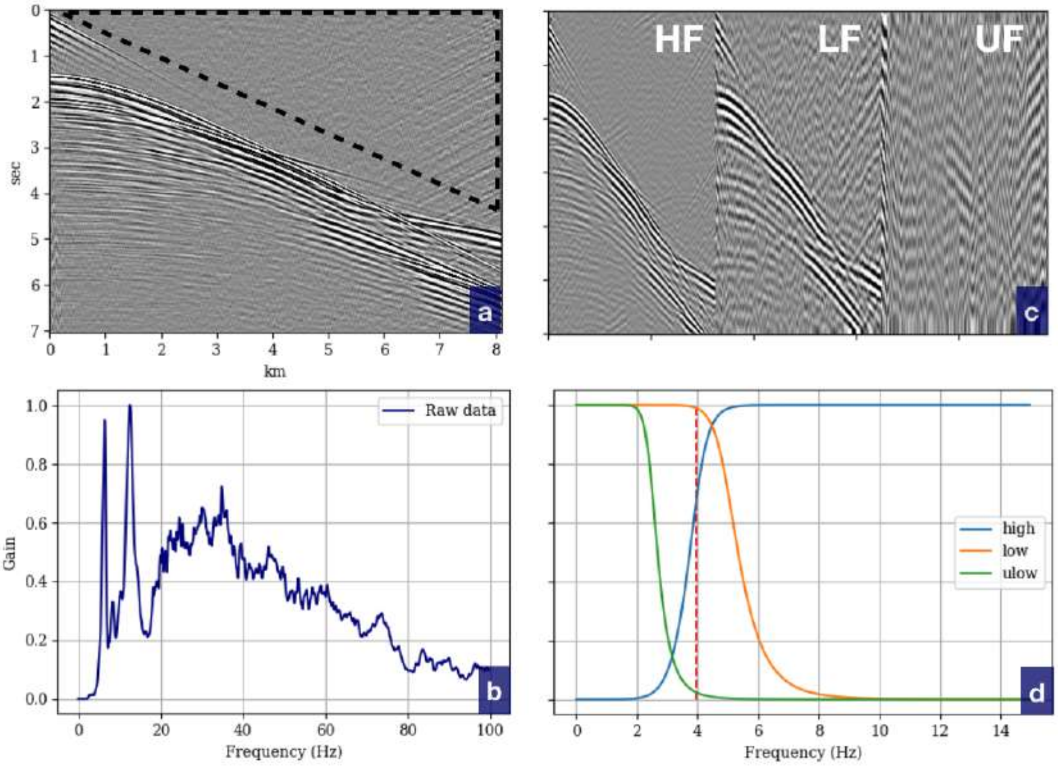


Figure 4.2: Marine streamer data (a) with its power spectrum (b). The dashed box outlines the pre-arrival area that serves as a donor of noise. Butterworth bandpass filters (d) are used to split seismic data into high-frequency (HF) and low-frequency (LF) partitions (c). The ultra-low frequency range (UF) where field data was not captured. Zero amplitudes are set in the input data for frequencies below the dashed line at 4 Hz.

separate trend randomization. [127] showed that a style-transfer approach is capable of transferring the layered structure from a geological reference to a smooth velocity background, but at high computational costs. [170] trained a generative adversarial network to accelerate style transfer between real-world images and geological models. [171] formulated a pipeline for building 3D synthetic models with salt intrusions. Alternatively, [88] demonstrated that an elastic transformation applied to a simple layered model might be sufficient for generating a diverse dataset of seismic waveforms. We follow and modify this approach by creating random models that approximately follow a user-defined background trend.

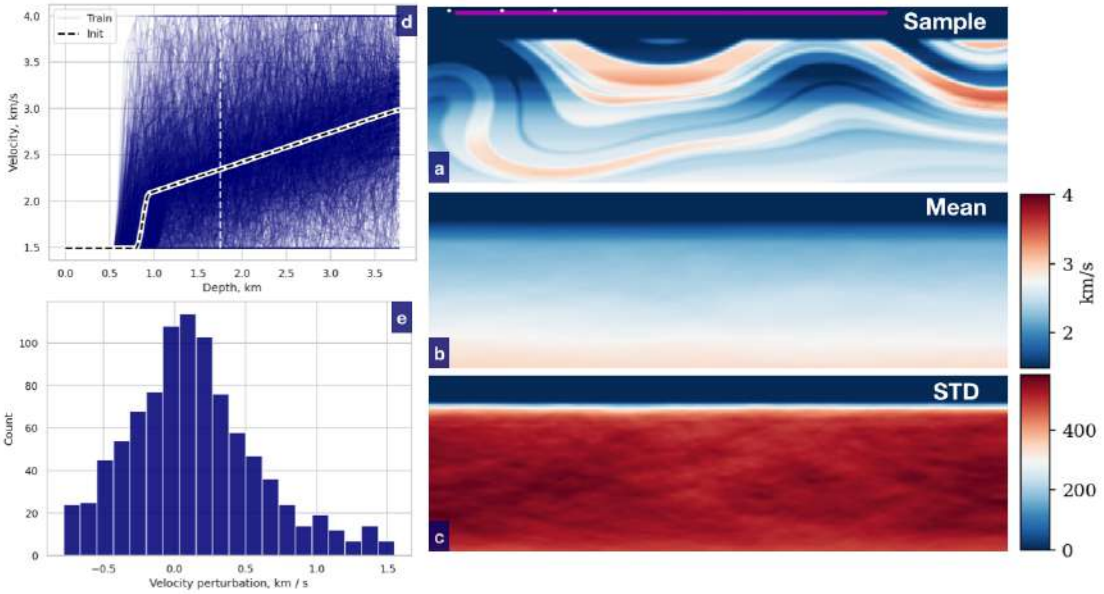


Figure 4.3: A sample random velocity model (a), the mean (b) and standard deviation (c) for a dataset of 1,024 synthetic subsurface model realizations. The stack of central well-logs (d) for the dataset of random subsurface models and the histogram of velocity perturbations (e) at the depth marked by the white dashed line.

The workflow for the generation of realistic subsurface models includes 4 steps: (i) we generate a sparse sequence of random values ranging from -1 to 1 to emulate the distribution of impedance in depth; (ii) we build a dimensionless velocity profile $v(z)$ by integrating and fitting the result into the range from 0 to 1; (iii) we replicate the $v(z)$ profile to make a laterally-homogeneous layered model $v(x, z)$; and (iv) we finally apply the elastic transform to distort the layered model and randomly re-scale the model velocities within 1.5 to 4 km/s.

Quantitatively, the amount of distortion is controlled by the mean and variance of 2D random Gaussian fields. Alternating those, one might produce models ranging from slightly variable layered models to salt-containing initializations. To ensure that the produced model is following a selected background trend, $v_0(x, z)$, we first remove the original trend from the generated model. Effectively, this creates a perturbation distribution $\delta v(x, z)$, centered around zero. Then, we add the randomized trend into the perturbation map, $v_0(x, z) + \delta v(x, z)$ (Figure 4.3). The background trend models also feature the fluctuating depth of the sea bottom. This makes the deep learning model trained on such a dataset applicable to marine datasets with variable water depth. The set of generated models then follows the specified trend while evenly spanning the domain in terms of feature variability.

This procedure generates random samples of compressional wave velocities, V_p . We then scale shear-wave velocity V_s and density ρ by empirical relations derived in [172],

$$\rho = 310 * V_p^{0.25}, \qquad V_s = V_p/\sqrt{3}. \tag{4.9}$$

Forward modeling in elastic media

The choice between an acoustic and elastic formulation of wave propagation defines the fidelity of the phenomena as well as the computational costs involved in the generation of the training dataset. [173] showed that for a large-offset marine data there is significant energy attributed to wave mode conversion, regardless of the water-bottom type. The discrepancies between acoustic and elastic modeling increase with offset and get more prominent in media exhibiting severe scattering [174, 175]. We thus employ the elastic formulation of wave propagation in both training dataset generation and FWI of marine data.

To generate synthetic seismograms for our set of random subsurface models, we use the elastic finite-difference implementation in time domain by [176] and [177]. The source wavelet extracted from the observed seismic data is the key component to enable the generation of training data and subsequent FWI. Unlike land data, the marine streamer data features sufficient wave propagation in the water layer, which makes it possible to estimate the source signature reliably [113]. In particular, we use the average source signature from all shots in the survey for forward modeling in random models.

Realistic seismic noise from field data

The lack of realistic noise in the training dataset is another contributor to the domain gap between synthetic and field data. The amplitude spectrum of seismic noise varies at different frequencies whereas common random Gaussian noise is prone to affect only high-frequencies when applied directly in the data domain [178]. To mimic realistic noise, we extract the noise imprint directly from the field data, which turned out to be a simple and efficient way to focus training on removing a particular noise pattern from predicted data. Instead of training the network to cope with diverse random synthetic noise, we train the network to remove the real-world noise specific to the target dataset.

An example of such a donor-area of representative noise is shown in Figure 4.2. We collect such triangular noise patches from all field shot gathers intended for use in FWI. Then, we tile and replicate these triangles into rectangles large enough to cover twice the entire target shot gather in the training dataset. The double coverage of data shape by noise allows augmenting the dataset on-the-fly during training by shifting the noise pattern and reversing its polarity. The limited amount of field noise samples extracted from target data makes the network accurately remove the data-specific noise at the inference stage.

Pre-processing

After adding realistic noise, we split the generated synthetic data into inputs and targets for training. We apply a set of 8th-order Butterworth band-pass filters, as shown in Figure 4.2. Before that, we normalize raw seismic data by dividing it by the maximum of its absolute value.

The input high-frequency data (HF) is made from full-band synthetic data by high-pass filtering with a corner frequency of 4 Hz. Moreover, we explicitly set zeros in the frequency domain for all data below 4 Hz to avoid signal leakage into the target. The low-frequency (LF) target data are built by applying a low-pass filter with a corner frequency of 5 Hz. We evaluate the accuracy of extrapolated data on the ultra-low frequency sub-band (UF) of target data, constructed by low-pass filtering below 3 Hz.

The overlap between corner frequencies of the target LF and the input HF range roughly accounts for the shape of band-pass filters and ensures the lossless pass of the target data below 4 Hz. Finally, we map LF into the [-1, 1] range, making it suitable for a deep learning application. This way the amplitude information about targets is lost and we use the amplitude of data in the available range from 4 to 5 Hz to reconstruct the scale of predicted data prior to the FWI application. The local velocity model underlying the shot location is another target for training. These model targets are scaled into the range [-1, 1] using velocity box conditions and the linear transform mentioned earlier.

4.4 Numerical experiments

To reduce the gap between applications to synthetic and real data, we conduct both a synthetic and field experiment, using the same survey design and source signature extracted from the real-world marine streamer data. First, we detail the dimensions of input and target data as well as specific hyper-parameters of the training runtime. Then, we show extrapolation results and apply FWI to synthetic and field data.

4.4.1 Network training

The density of the marine survey described earlier offers redundant information for the extrapolation of data below 4 Hz. We reduce the dimensionality of the data by sampling receivers in the streamer with 25 m offset as well as increasing the time sampling rate of seismograms from 2 ms to 16 ms. This way, each high-frequency CSG (input data) measures 324×376 data points along offset and time dimensions, respectively. The target low-frequency data shares the same dimensions while the target subsurface model is 4 times more sparse and measures 81×94 model points with 100 m spacing along the offset dimension (Figure 4.4). For the training dataset, we generated 3,072 synthetic shot gathers by modeling 3 shots in each of the 1,024 initializations of random velocity models. The final dataset was further split into training, validation, and testing partitions of 2,765, 154, and 153 samples, respectively.

To compensate for the ambiguity caused by random weight initialization [164], we spawn an ensemble of 10 identical networks initialized from different random seeds and average their predictions. The training strategy for each network includes 81 epochs with batch size 4, guided by an Adam optimizer [130] with variable learning rates. In particular, we utilize the one-cycle learning rate schedule by [167] with learning rate

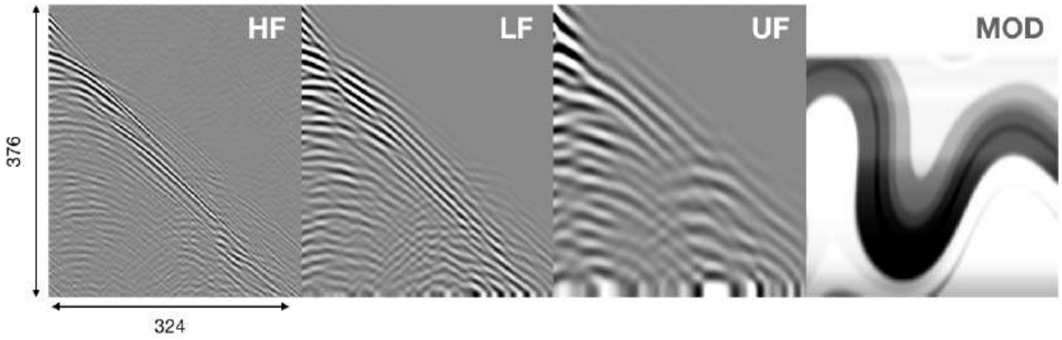


Figure 4.4: Sample input and target data used for training. High-frequency input data (HF), low-frequency target for training (LF) and local subsurface model (MOD, upscaled 4 times for visualization) for a synthetic shot gather. The ultra-low frequency data (UF) was used for the evaluation of bandwidth extrapolation.

bounds of 10^{-5} and 10^{-3} (Figure 4.5).

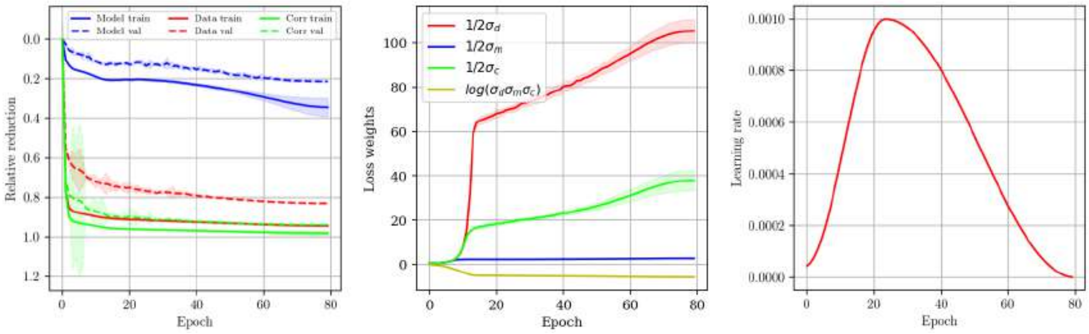


Figure 4.5: Training and validation loss curves (left), weights of respective losses (center) and learning rate schedule (right). A larger weight for a certain loss term indicates higher confidence associated with the task (lower uncertainty).

4.4.2 Marmousi II benchmark model

The Marmousi II model [179] is a standard benchmark for inversion and imaging algorithms. We modify the model to match the field survey setup by cropping its dimensions and rescaling its range (Figure 4.11). In particular, we reduce the maximum velocity in the model to 4 km/s in order to meet the model bounds introduced earlier for the generation of the synthetic dataset. The computational domain is discretized onto a regular mesh with 25 m spacing, measuring 152×600 model points. This is equivalent to the velocity model of 3.8×15 km along with depth and offset axes, respectively. We place 128 sources every 200 m along the offset dimension, from 1.125 km to 13.825 km. The marine streamer is about 8 km long and the first offset used in the inversion is 175 m. In this setup, we record the data using 324 hydrophones evenly spaced at 25 m along the streamer. For simplification purposes we ignore the variable depth of the Broadseis survey system and approximate it by placing the streamer line at a depth of 50 m below the free surface.

Contributions of loss terms

In MTL, we seek to optimize the distribution of weights in the layers of the network which minimize several loss terms at the same time. To understand the contribution of each loss term, we gradually add loss terms one after the other to training and plot the predicted data Figure 4.6. There, we compare the synthetic data reference against predicted LF data (top row) and its subset matching the UF range (bottom row) inferred by several network configurations.

Experimental setups described here pivot on two network architectures as well as the increasing complexity of the loss formulation. The baseline architecture is the UNet [180] that is a common choice in numerous applications dedicated to image translation and segmentation. To construct an inference reference, we set UNet to directly predict LF data from input data HF (UNet L). All other configurations are based on a multi-column network layout, which shares the prefix "MTL". The first setup, given by a multi-column encoder, aims to directly reconstruct one target that is low-frequency data (L). The dynamic loss weighting is implemented in the remaining configurations where the objective function for training combines more than one term. The first objective is to fit the data loss together with the correlation loss term (LC). The last experimental setup involves adding the subsurface model loss term (LCM) thus using all terms from the equation 4.1.

Analyzing the inference results for a single CSG (Figure 4.6), we observe that the generic UNet manages to recover strong events in the data while it fails at weak events, such as reflections. The multi-column architecture with a single data target (L) shows promise in recovering weak events. Adding the trace-wise correlation term (LC) boosts the amplitudes of predicted data at later times. An intuition behind using the Pearson coefficient as a loss term is similar to the one for the cosine similarity, where the target for optimization is the angle between two vectors rather than their amplitude matching. Lastly, by adding the local subsurface as a target we guide the training toward a unified solution that would accommodate a weak connection between waveforms and

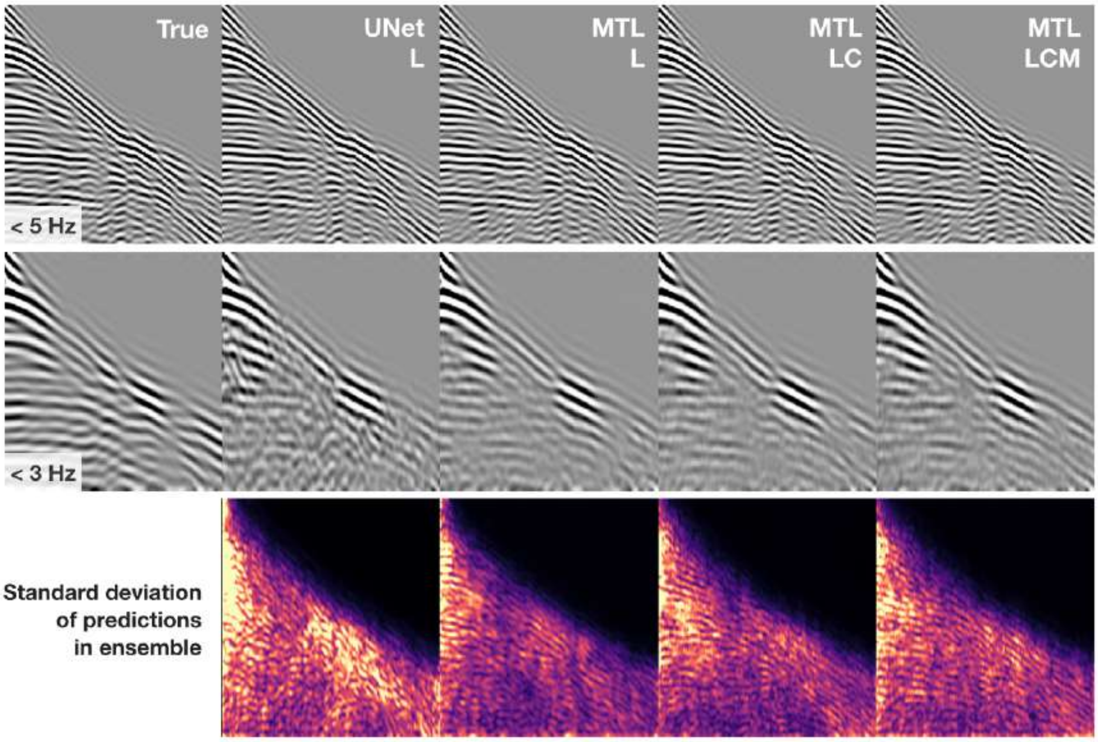


Figure 4.6: Low-frequency data (< 5 Hz, first row) predicted for a single synthetic shot gather by UNet and multiscale network configurations. Same data after low-pass filtering < 3 Hz (second row). Subscripts indicate the objectives of training: low-frequency data (L); previous with correlation loss term (LC); previous with the local subsurface model (LCM). The standard deviation of predictions by an ensemble of 10 network initializations (third row)

Obj	Arch	R2	SSIM	Pearson
True	-	1	1	1
L	UNet	0.38 ± 0.25	0.79 ± 0.02	0.63 ± 0.25
L	MTL	0.44 ± 0.24	0.81 ± 0.02	0.69 ± 0.23
LC	MTL	0.44 ± 0.26	0.81 ± 0.02	0.68 ± 0.25
LM	MTL	$-2.21 \pm \ldots$	$0.0 \pm \ldots$	$0.0 \pm \ldots$
LCM	MTL	0.44 ± 0.25	0.81 ± 0.02	0.69 ± 0.23

Table 4.1: Evaluation metrics for selected network configurations. Mean R2 score, Structural Similarity Index (SSIM) and Pearson coefficient measured for 80 extrapolated shots (< 3 Hz) used for FWI in Marmousi II model. Abbreviations for experiments are explained in Figure 4.6.

the subsurface (LCM). Fitting the multi-objective loss is a more challenging task with more variables involved. However, the LF data predicted in this way appears to be more accurate than the one predicted without model and correlation loss terms (Table 4.1).

For each configuration we also plot the trace-wise Pearson correlation coefficient (Figure 4.7). The intention is to understand how the proposed architecture and loss terms affect the linear correlation between predicted and true data when compared to the baseline configuration. Specifically, we compute the mean and standard deviation from the low-pass inference results, filtered below 3 Hz, on 80 shot gathers used later for the inversion in the Marmousi II model, and compare these quantities with the mean trace-wise Pearson coefficient computed for predictions by UNet (dashed line). There is a minimum value in the near-offset range shared among all experiments which increases again when moving away from the source location. The proposed approach (LCM) shows a higher linear correlation with the target data at all offsets as compared to UNet.

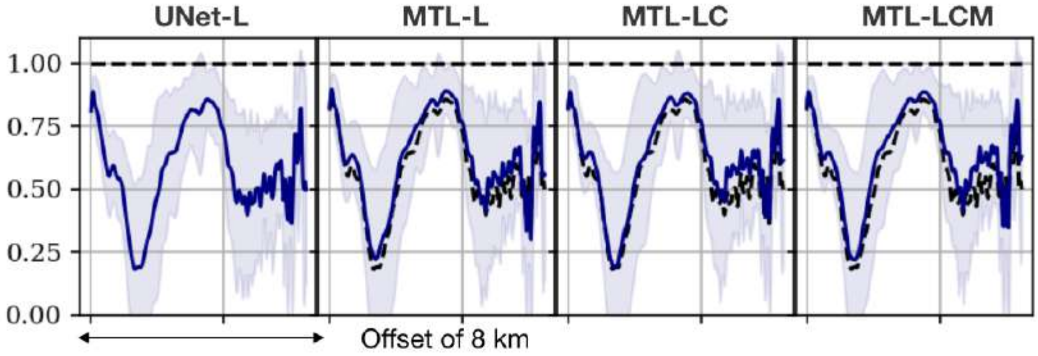


Figure 4.7: Pearson correlation coefficient for the set of 80 shots used for FWI on synthetic data. A horizontal dashed line indicates the perfect correlation score. The dashed line in correlation plots indicates the performance of UNet. Abbreviations for experiments are explained in Figure 4.6.

Table 4.1 shows the mean metric scores of inference on the same shots of synthetic data mentioned earlier. Due to the mild overlap between target and input data, the data < 5 Hz predicted by all network setups shows a nearly perfect correlation with the true data, not shown in the table. Lower frequencies, in turn, have a higher value for early iterations of FWI so we focus attention on the subset below 3 Hz. Aside from the Pearson coefficient, we use a set of common metrics to compare the performance of the algorithms. The R2 metric [181] measures how much more variance the model describes compared to the mean of the dataset. The structural similarity index measure (SSIM) quantifies the perceptual similarity between two images [182]. Unlike other metrics reported here, it depends on the window size where evaluation happens. We set it to be 0.1 from the minimum dimension of the data or 35 pixels along each side.

All experiments with multi-column architecture show improved fit in the ultra-low frequency range compared to the UNet application. Otherwise, reported metrics indicate similar performance among all formulations of multi-task objectives. A direct recovery of the data and model (LM) was not successful without using the correlation loss term. In such cases, the training fell into the local minimum equivalent to recovering of the subsurface model only, giving up the data fitting turn. Despite marginal differences in performance of data only (L) and data with trace-wise correlation (LC) objectives, the correlation term appears to be crucial for simultaneously fitting data and model objectives (LCM). Together with the recovered subsurface model, the predicted low-frequency data seems sufficient to guide FWI to a better minimum.

Inference on synthetic data

The trained MTL neural network accepts a band-limited shot gather, HF, $>$4 Hz, as input and produces two outputs - low-frequency data, LFp, < 5 Hz and the sparse representation of the local subsurface model Mp. Numerically, this is equivalent to translating the input volume dimension of $[1, 324, 376]$ into $[1, 324, 376]$ and $[1, 81, 94]$.

The first output of the model is low-frequency data. Weak amplitudes of the predicted signals at later times make it difficult to evaluate the presence of signals there. We apply an automatic gain control (AGC) to several low-pass subsets of predicted synthetic data to visualize how the complexity of extrapolation increases at lower frequencies (Figure 4.8). As can be seen, the reconstruction is more accurate at higher frequencies since we intentionally introduce the overlap window between 4 and 5 Hz to recover the amplitudes of the predicted data. The predicted data below 3 Hz matches the target in shallow parts while the signal gets scattered at later times.

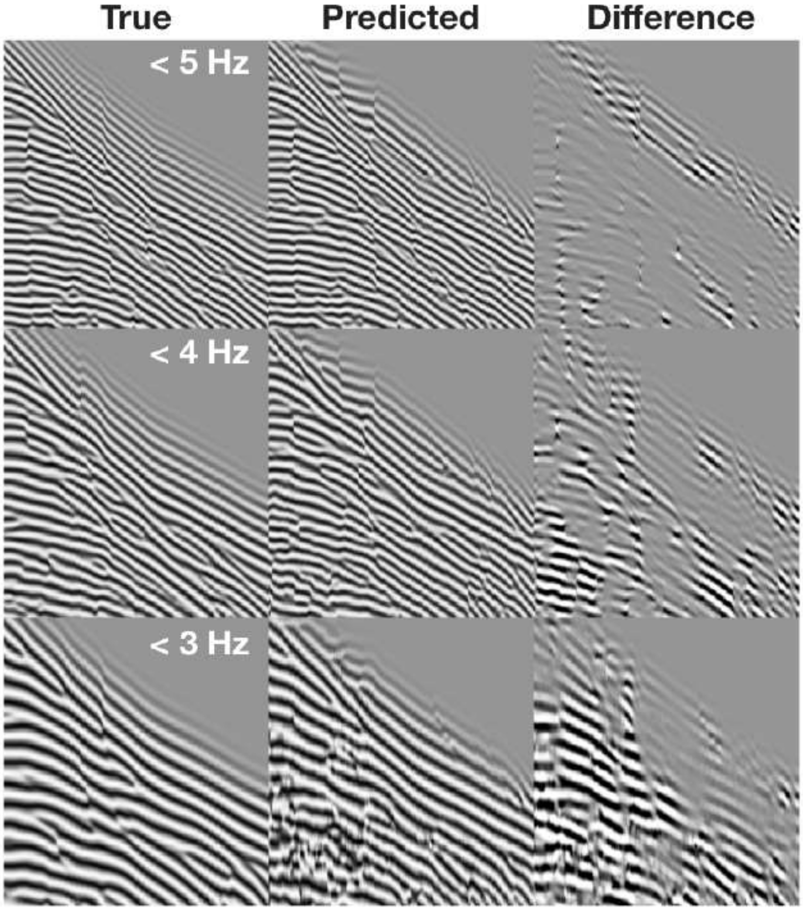


Figure 4.8: Example interval comparison of true and predicted low-frequency data after low-pass filtering and automatic gain control.

The second output is the local subsurface model. The network learns a direct mapping between input data and the geological structure underlying the shot location. To understand the data-to-model translation, we upscale the predicted velocity model by a factor of 4 (the ratio is built into the network design) and overlap it with the input data Figure 4.9. Given that the synthetic dataset used for training was generated assuming a flat water bottom at variable depth, this approximation of variable-depth marine subsurface geometry appears to be sufficient for the network to match strong water bottom reflections in the data with the depth of the underlying seafloor. Also, predicted velocity anomalies seem to be spatially limited by the offset reached by the propagating signals. To improve the spatial coverage in the predicted model we accommodate this observation by introducing offset-flip data augmentation at the training stage. This effectively doubles the size of training datasets and motivates the weight optimization to search for flip-invariant encoder embeddings of the data.

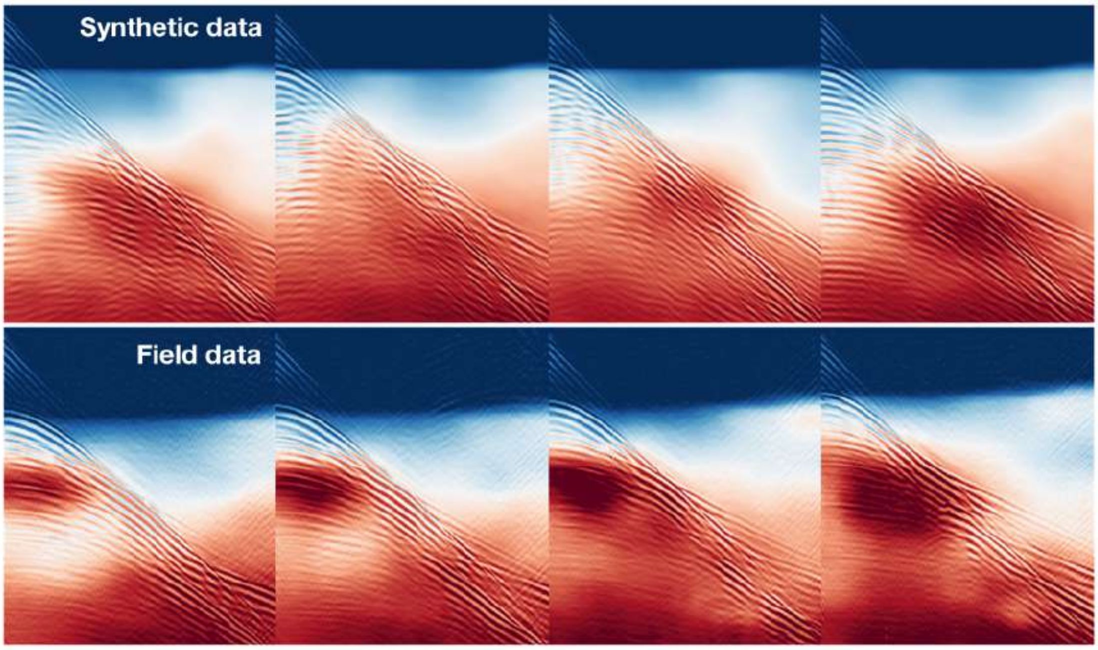


Figure 4.9: Local subsurface velocity models predicted by the same trained network and overlapped with respective input synthetic (top) and field data (bottom). The network pivots on water bottom reflections to recover the depth of the seafloor as well as translates deeper reflections into velocity anomalies.

A single CSG is not sufficient to recover the complete 2D subsurface model. However, it contains enough information about the background velocity trend, water depth, and near-offset geological structures.

Full-waveform inversion

The complete workflow for elastic FWI initiated from predicted data is shown in Figure 4.1. First, we build an initial velocity model for inversion by taking the weighted average of predicted local velocities for each shot. The part of the initial model not covered by the survey (before the first source location) is a mirrored version of the reconstructed model. Before running FWI, we populate the missing frequency content below 4 Hz in the field data with the one of the predicted data. True amplitudes of low-frequencies were lost at the pre-processing stage and we approximately reconstruct those from the overlapping range from 4 to 5 Hz between predicted and available high-frequency data. Specifically, we apply a band-pass filter with the before-mentioned corner frequencies to both predicted and observed data and find the maximum amplitude in the filtered data. Then, we normalize the predicted data to fit the range from -1 to 1 and multiply it by the amplitude extracted from field data. The remaining step is to merge the predicted data with the field data, done in the frequency domain.

The inversion strategy uses an L_2 norm minimization of the difference between observed and simulated data. We successively invert the low-pass filtered data with corner frequencies from 3 to 7 Hz, making 1 Hz steps. We further regularize the inversion of the extrapolated 3 and 4 Hz data by applying a 2D spatial variable Gaussian filter to the gradients [183]. This causes smooth velocity updates for the most uncertain data. We disable this regularization for frequencies starting from 5 Hz, which inserts high-resolution details into the inverted models.

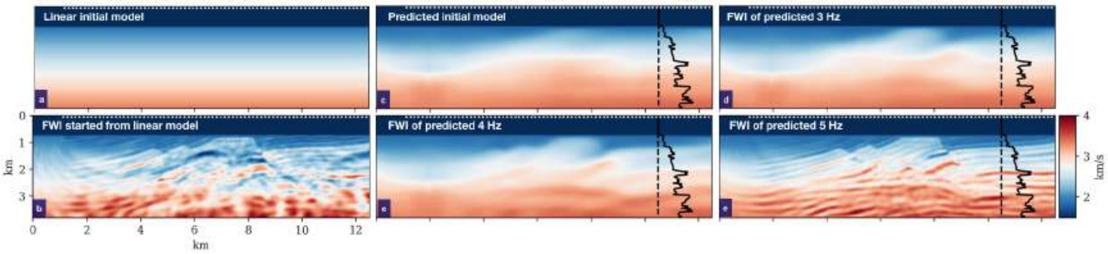


Figure 4.10: Full-waveform inversion of band-limited data 4-7 Hz (b) started from linear initial model (a). Extrapolated FWI initiated from predicted initial model and low-frequency data below 3, 4, and 5 Hz (c-e).

As seen in Figure 4.10, conventional FWI initiated from a linear initial model fails when data below 4 Hz is unavailable even when velocities in shallow sediments are assumed to be known from well-logs. The resulting subsurface model is corrupted by severe cycle-skipping artifacts indicating that the mismatch between target and initial subsurface models can not be inverted by FWI. In the following, we initiate the inversion from a predicted smooth initial velocity model and use predicted data as the target.

The predicted initial model appears to be locally linear (Figure 4.10) as expected due to the limited amount of information encoded in a single CSG. In other words, the predicted model explains the water bottom reflection and background trend at the reference log location, rather than reflects the detailed structure of the subsurface. When followed by the inversion of predicted data below 3 Hz, the shallow part of the subsurface becomes more pronounced. The next iteration of inverting for data filtered below 4 Hz details the complex fault structures in the central part as well as corrects the high-velocity layers around the reference well-log. The inversion of subsequent frequency bands above 5 Hz is dominated by high-frequency field data and thus inserts fine details into the inverted subsurface. Turning off Gaussian smoothing of gradients above 5 Hz is another reason for the higher resolution at later stages.

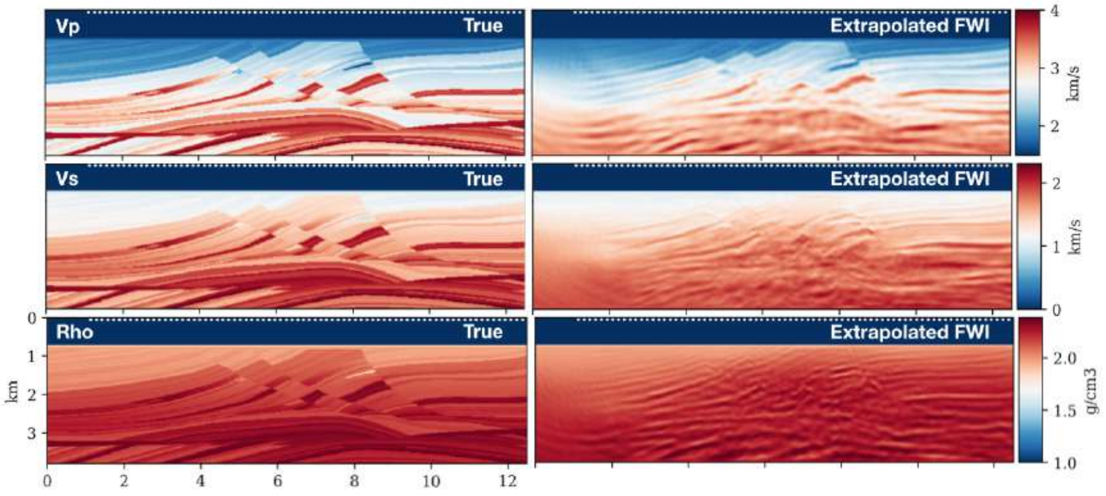


Figure 4.11: Elastic full-waveform inversion of Marmousi II benchmark model. The true distribution of elastic parameters V_p , V_s and ρ are compared with their counterparts inverted by extrapolated FWI.

The isotropic elastic Marmousi II model is parameterized by velocities of compressional waves, V_p , shear waves, V_s , and density, ρ (Figure 4.11). We construct V_s and ρ using the same empirical relations as were used for the generation of random subsurface models for training. The reason is to deliver a homogeneous framework between synthetic and field experiments where these elastic parameters are unknown.

The V_p component reconstructed by the extrapolated FWI closely follows the well-log down to a depth of 3 km and then undershoots the deep high-velocity structure. The potential reservoir hidden in the folds of sedimentary layers in the shallow part is clearly resolved. Reconstructed shear-wave velocity V_s , and density ρ , also exhibit common features with the target model but show worse results. The reason for that is in the nature of the scattering phenomena. In an ideal illumination scenario (infinite offsets and unlimited frequency content) all three isotropic elastic parameters could be resolved from the recorded P-waves [54, 119, 118]. P-waves recorded by hydrophones are most sensitive to V_p so this parameter is resolved most accurately. The range of illumination angles decreases with depth, as well as high frequencies decay faster with distance. Decoupling of perturbations from different parameters becomes more challenging for deeper targets. For these reasons, V_s and ρ are mostly resolved in the shallow parts of the model.

4.4.3 Marine field data

The same trained neural network as in synthetic experiments is applied for inference on marine field data. We band-pass each of the 80 shots used in FWI between 4 and 15 Hz to produce inputs to the network.

Inference

The network reconstructs low-frequency data below 5 Hz and a set of localized subsurface models (by shot) such as those shown in Figure 4.9. The Broadseis acquisition system used during the survey records frequency components of the wavefield down to 2.5 Hz. Thus, there is a shared band of frequencies below 4 Hz where we can compare predictions and observed waveforms Figure 4.12. We see a match of first arrivals since these are the strongest events. However, since synthetic data does not experience attenuation we observe boosted amplitudes in later arrivals. The network also serves as a denoising operator, and the noise present in the filtered data is due to leakage from higher frequencies. Seismic noise is almost absent in the data at the lowest frequency range for the same reason.

Contributions of the components of the multi-task objective are visualized in Figure 4.13. The correlation term promotes reflections arriving at later times to become more prominent. Adding local subsurface as a target further improves the consistency of predictions within the ensemble of network initializations.

Full-waveform inversion

Full-waveform inversion of the field dataset follows a similar strategy as the inversion of synthetic data. Specifically, we use the L_2 objective function between simulated and observed data and guide the optimization by an L-BFGS algorithm [99]. The target for inversion is a blend of extrapolated low-frequency data and available high-frequency data. We change the corner frequency of low-pass filters to invert for sub-bands of the data in a stage-like fashion [150]. In particular, we partition the full band by filtering

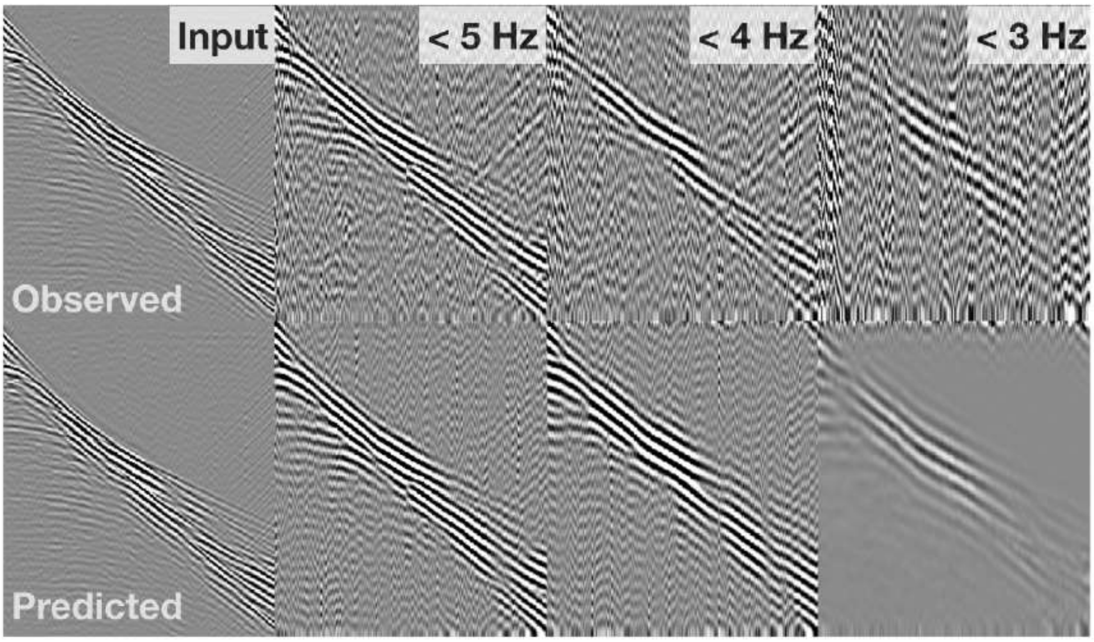


Figure 4.12: Comparison of observed and extrapolated data in frequency ranges of a CSG from marine field data. The synthetic input data for training was set to zero below 4 Hz.

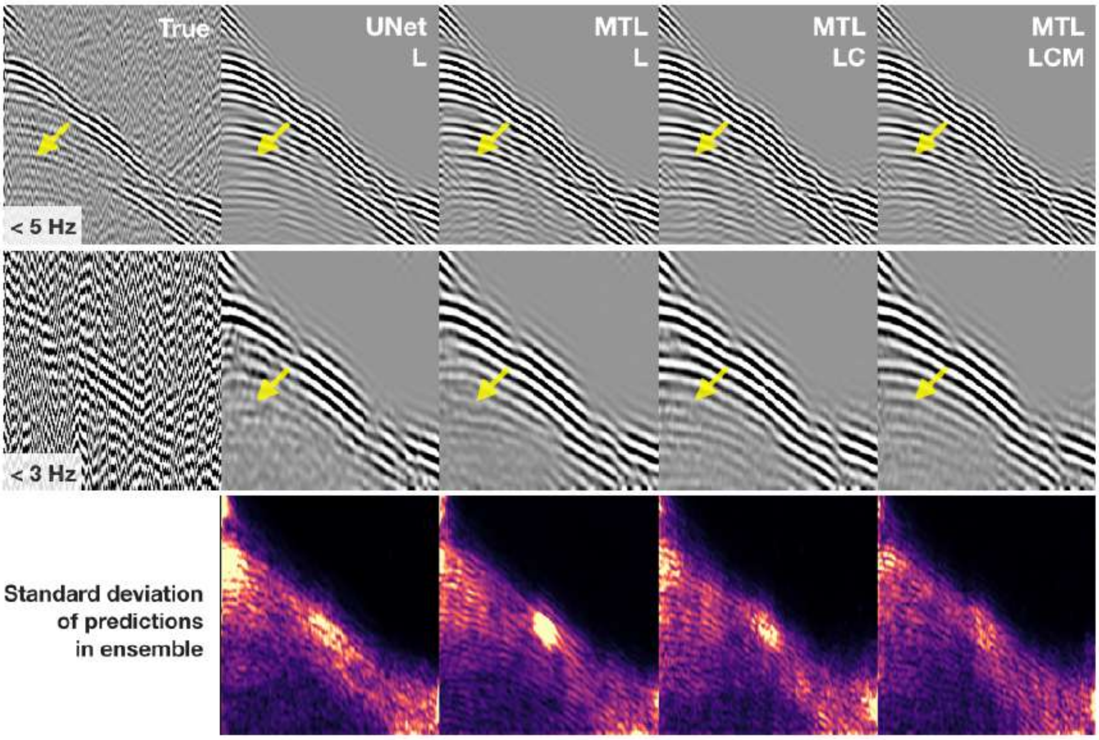


Figure 4.13: Same as Figure 4.6 but for a CSG from marine field data.

it with corner frequencies of 3, 4, 5, 6, and 7 Hz. We account for inaccuracies in the extrapolated data below 4 Hz by Gaussian gradient smoothing at the first two stages. We disable smoothing at later stages. The shape of seafloor found from the earliest reflection arrivals defines a taper mask, disabling model updates in the water column.

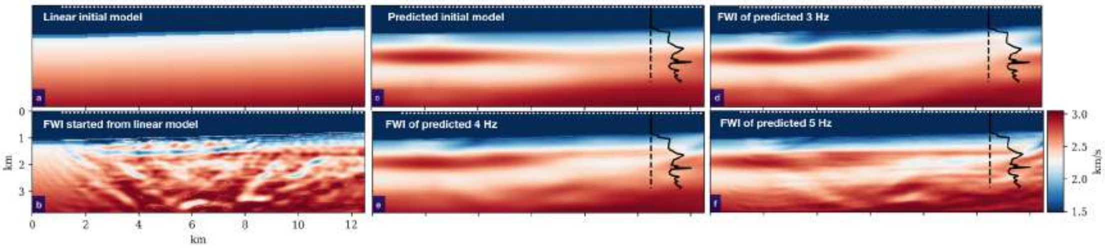


Figure 4.14: Same as Figure 4.10 for real-world marine streamer data.

We examine the capability of FWI without low frequencies by starting it from a 1D velocity model with known water bottom (Figure 4.14, left). Moreover, we assume a known transient layer between water and sediments by using the shallow part of the well-log extracted at an offset of 10.5 km. However, these assumptions appear to be insufficient for the inversion to converge when the low-frequency content in the data is limited by 4 Hz. The resulting model indicates the presence of a high-velocity layer at 2 km depth, unable to build a consistent distribution of elastic parameters.

The predicted background model (Figure 4.14) indicates the presence of a high-velocity anomaly in the left part of the model. Meanwhile, the velocity profile increases gradually with depth in the right part. A notable feature of the predicted model is that the sea floor depth was accurately reconstructed (Figure 4.15). Further FWI iterations of the predicted data at 3 and 4 Hz refine the layered structure of the subsurface. However, the predicted high-velocity layer on the left seems to be overestimated which causes artifacts when inverting higher frequencies.

Figure 4.15 compares an observed well-log recorded at 10.5 km with predicted initial model and final FWI result of 7 Hz data. We find that the peaks of high-velocity layers are shifted in this field data example. A similar shift of the first layer with respect to the well-log was observed in [184]. A reason for that might be the fact that we ignored anisotropy in our simulation. Vertical transverse anisotropy would take into account different velocities of wave propagation along horizontal and vertical directions in such a layered model. Another reason for overestimating velocities deeper than 2 km is that

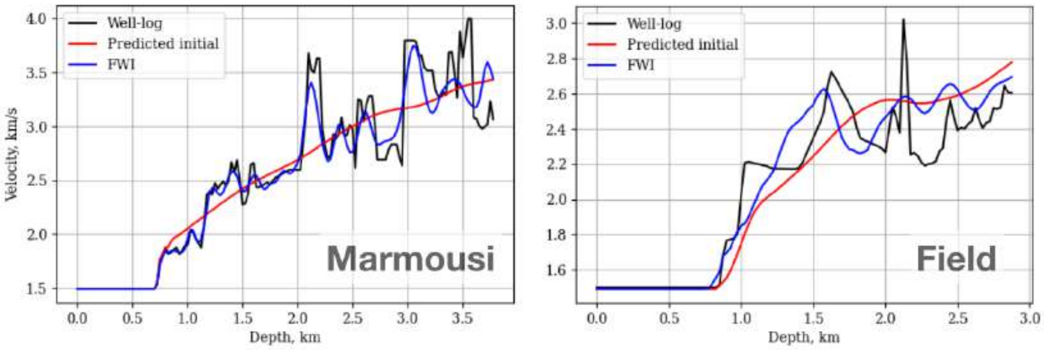


Figure 4.15: Well-logs of V_p velocities compared to predicted initial model and the extrapolated FWI result. Synthetic experiment on Marmousi II model (left) indicated a reasonable fit down to 3 km depth. In the field data experiment (right), the mismatch is partially caused by the isotropic approximation of the medium as well as ignoring the attenuation of the wavefield.

we ignored 3D effects of the field data as well as attenuation in the formulation of the numerical wave propagation. All together, higher amplitudes of the synthetic signal arriving at later times translate into higher velocities of deep reflectors. Still, there is a fair match with the field well-log considering that no well-log information was used when building the initial velocity model and during the inversion itself.

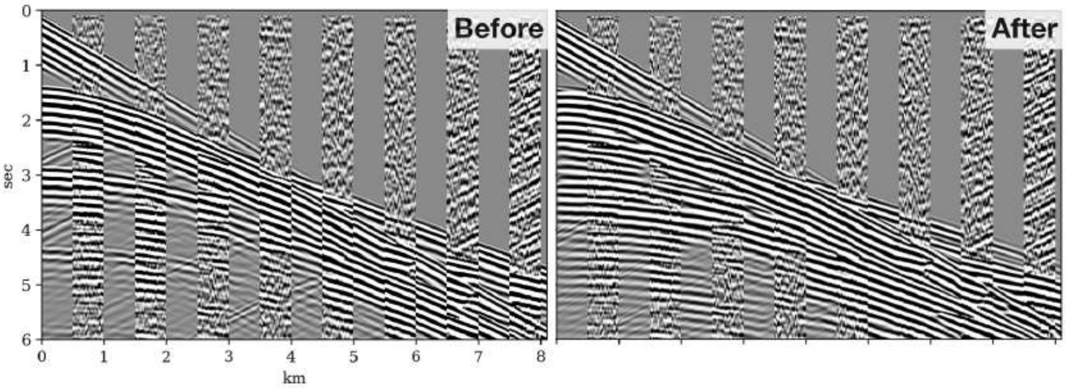


Figure 4.16: Example interval comparison of marine streamer data and synthetic data generated in the final velocity model inverted by FWI, initiated from predicted initial model and low-frequency data.

In our full-waveform inversions, the data match is the only objective. The optimization thus seeks for a distribution of elastic parameters in the subsurface such that the generated synthetic waveforms match every wiggle in the observed seismic data. Figure 4.16 compares the field record from 4 km with its synthetic counterpart from the initial and final model inverted by elastic FWI. We see that phases of the reconstructed events overall improve and are in fair agreement with field recordings, while some amplitudes remain overestimated. Further advances are still possible regarding both modeling and parameterization possibilities as well as FWI strategies, however, this would extend the focus of this study and must be addressed in future research.

4.5 Discussion

In this study, we explore the extension of a data-driven FWI framework to include a multi-task learning approach, combining bandwidth extrapolation and initial subsurface model estimation. Deep learning based on convolutional neural networks is dictated, among others, by the number of trainable parameters and the associated search space. As compared to single-frequency formulations, time-domain signals might be considered as superpositions of multiple monochromatic signals, each featuring its own wavelength and amplitude. A 1D configuration of seismic data thus would be a single seismic trace in the time domain and mono-frequency data in the frequency domain. The dimension of a time-domain trace would be equal to the number of temporal samples while the dimension of a single frequency slice is proportional to the number of receivers. Adding an extra dimension to any of those simplest data types increases the complexity of training since optimization should search for the network weight distribution describing a search space with an extra dimension.

Bandwidth extrapolation becomes more challenging when moving from a trace-to-trace to a shot-to-shot and finally to a survey-to-survey formulation due to the growth of the dimensionality of the outputs. On the other hand, the fidelity of predictable wave phenomena increases thanks to the growth of the input data features. Ultimately, the accuracy of bandwidth extrapolation should improve when adding more data, provided that the training data set is rich and the prediction model has sufficient learning capacity.

Full-waveform inversions within the framework of our study depend on the high-fidelity estimations of either low-frequency data or the initial velocity model. Reaching a required accuracy for each of these estimations independently might be an ambitious task. However, due the inherent trade-off between these tasks, inversions tend to tolerate inaccuracies when fair rather than perfect estimates of both velocity model and low-frequency data are available.

In this work, we extrapolated full shot gathers recorded in a real marine survey. We observe, that none of the predicted low-frequency elastic data are sufficiently accurate to run an elastic FWI starting from a simple, linearly increasing with depth, velocity model. However, when we jointly predict both low-frequency data and the initial model for inversions, the predicted initial model compensates for data inaccuracies, increasing the robustness and fit for a FWI application. Additionally, our test applications further confirm the importance of self-weighing of loss-terms in a multi-task learning formulation. In particular, the auxiliary correlation loss term based on the Pearson coefficient further helps the recovery of reflections in predicted data.

4.6 Conclusions

We developed a data-driven deep learning approach to jointly reconstruct the low-frequency content of an entire common-shot gather together with the respective local subsurface model. The proposed multi-task learning objective with automatic weight balancing aims to simultaneously fit data, initial model estimate, and trace-wise correlation terms. The predicted smooth background model effectively compensates for inaccuracies in recovered ultra-low frequencies by uplifting the lowest frequency required to initiate the inversion. Adding noise from real data to the training dataset and replicating real acquisition parameters (source sweep and source-receiver spacing) allowed us to use the same trained neural network for real and synthetic data, without any transfer learning or retraining. New data acquired by the same streamer and airgun setup in a similar deepwater environment can potentially be processed in seconds now with this trained network.

Chapter 5

Discussion and conclusions

This chapter concludes my dissertation. Here I will summarize the key findings, discuss the limitations of the proposed methods as well as share my outlook on future research.

5.1 Key findings and limitations

Full-waveform inversion is a rather fragile non-linear optimization process that is sensitive to the available bandwidth in the data as well as a starting point. I dedicated these few years to a search for a data-driven way of how to facilitate the convergence of FWI by starting it from a better state. In a nutshell, I showed that model-domain projections of errors that FWI makes can guide to a better model of the subsurface. Lower frequencies are easier to recover when approaching the task in the frequency domain. On the contrary, very low frequencies are more difficult to reconstruct in the time domain since amplitudes there are dominated by higher frequencies within the same target range. Finally, I proposed to loose the requirements for accuracy of extrapolated low-frequency data and smooth background model for time-domain elastic FWI by jointly predicting the missing data and initial model in the multi-task deep learning setup. Below I extend this short summary for each chapter of the work.

5.1.1 Initial model building by the variance-based method

Chapter 2 proposed a way to learn from errors that frequency-domain FWI makes when initiated from a poor initial model. Specifically, I formulate the variance-based method for the initial velocity model building in the presence of salt bodies. High contrasts cause poor illumination of sub-salt and cycle-skipping occurs when velocities inside the salt body are severely underestimated. I showed that these cycle-skipping artifacts in the model domain correspond to the local minima of the objective function for specific mono-frequencies. The model-domain difference between such corrupted inversion results might then be used to locate and correct the unresolved areas of the subsurface. By iteratively interpolating these unresolved areas from the robustly inverted neighborhoods we guide inversion to a better minimum of the objective function.

Discussion. The method is tailored for frequency-domain solvers and I assume that the shape of the artifacts might change depending on the formulation of forward modeling in the frequency domain. The concept should remain valid regardless of the solver but other scenarios should be explored deeper. The beneficial feature of frequency-domain modeling is its computational costs at low frequencies. Long-wavelength data allows sparse sampling of the modeling domain that facilitates running inversions at several mono-frequencies at the same time. Successful termination of the proposed iterative salt-flooding indicates the algorithm constructed a low-wavenumber model of the subsurface suitable for further inversion. Time-domain FWI is the industry standard for 3D problems as solving wave equation in time domain offers scalable and flexible framework. The usage of frequency-domain FWI is typically limited to 2D problems due to high memory requirements associated with solving Helmholtz equation for 3D setups. Thus, for practical applications on field data the proposed variance-based method should be used only for building the initial velocity model followed by time-domain FWI.

5.1.2 Low-frequency extrapolation by deep learning

Chapter 3 explains frequency-domain method for low-frequency data extrapolation by deep learning. I show that the frequency-domain approach is suitable for multi-offset data extrapolation since each mono-frequency data is encoded by a single complex value at each offset, so the entire common-shot-gather in 2D survey setup might be explained by a 1D vector for each frequency. The accuracy of bandwidth extrapolation in such a formulation increases for lower frequencies due to larger wavelengths in respective waveforms. Another benefit of the frequency-domain formulation is that data for each frequency might be scaled independently, mitigating the problem of negligible amplitudes of low frequencies in the time domain. Specifically, the method is based on mapping by a deep neural network from multiple regularly sampled known frequencies to a single target low frequency. This way, a deep learning model becomes specialized in recovering a single frequency by extrapolating patterns in the stack of higher frequencies.

Discussion. The conventional supervised training of the deep learning model is limited by the diversity of the training dataset and the realism of physics of wave phenomena during modeling of seismic wave propagation. Only the acoustic approximation was considered in this work. While sufficient for experiments on synthetic data, it might not be accurate enough to describe complex wave phenomena in field data. Since the approach assumes training a separate convolutional network for each frequency, the practical application of the proposed method is limited by frequency-domain FWI which requires a sparse set of frequencies. The natural representation of a seismic signal is in the time domain so the next method addresses issues outlined in the previous two methods.

5.1.3 Multi-task learning for joint data and model extrapolation

Chapter 3 proposes a deep learning method for joint recovery of low-frequency data and a smooth background model for FWI. I implement the automatic contribution balancing of data, model, and trace-wise correlation loss terms in the objective function and show that training fails when the correlation is not taken into account. Moreover, this study shows that large perceptive fields of convolutional kernels are crucial to capture the long-wavelength imprint of low-frequency data. The key finding of this work is that predicted low-frequency data and the initial background model compensate for inaccuracies of each other by increasing the minimum frequency required for inversion to converge when started from the predicted initial model. This way, the same deep neural network trained on synthetic data produces suitable data/model pairs for elastic FWI applied to synthetic and field data.

Discussion. An interesting observation is that the proposed deep learning application does not break physics. Meaning that the predicted velocity model from a single shot gather, as expected, is missing fine details and is not accurate when considered without the surrounding context of other shot gathers. In particular, near-offset sections of the subsurface are reconstructed in more detail than farther offsets where illumination is far from zero-offset approximation. This supports the assumption that the non-linear function derived in the layers of the network is a reasonable estimator of the interaction of the wave field and the elastic medium. A more physics-backed formulation of the input data for training would be common-midpoint-data arrangement that is natural for velocity analysis. Another note is about the quality of recovered low-frequency data. The FWI is sensitive to the accuracy of extrapolated data at low frequencies. Meaning that mistakes at lower frequencies will strongly affect inversion at higher frequencies. This poses high expectations on extrapolated data at tomographic frequencies. However, joint extrapolation of low-wavenumber partially compensates for inaccuracies in predicted low-frequency data and allows to lift the minimum frequency needed to proceed with elastic FWI.

5.2 Discussion

For deep learning applications of bandwidth extrapolation, there is a trade-off between computational complexity and fidelity of description of wave phenomena. Intuitively, the fidelity of description of wave phenomena also increases when advancing from single trace/frequency formulation toward shot gather and finally reaching the scale of the entire survey where all wave phenomena are realistically captured. However, the complexity of training the deep neural network also increases for the input data of large dimensions. Meaning that single-trace data requires an order of magnitude fewer basis functions to be accurately explained. A shot gather, thus, should be described as a combination of 2D kernels rather than 1D kernels sufficient for trace explanation. The survey-scale setup operating on full-scale spatio-temporal 3D seismic data would then require even more basis functions of higher dimensionality to completely describe common patterns in the data.

The generalization capability of the listed data layouts increases as the fidelity of described physics decreases. Trace-to-trace approaches then might be easily applied to any domain as long as temporal sampling in the data is matched. However, these assume the lowest fidelity limited by information encoded into a single trace. The shot-to-shot formulation requires temporal and spatial sampling to match between shots from different datasets. This formulation already accounts for 2D effects. The generalization of applications operating on entire surveys seems challenging since the knowledge transfer should be possible if an identical acquisition takes place in different geological environments. Such a setup might be viable in marine streamer setups. An alternative solution would be to re-formulate training from physical to reference domain and assume variability in survey parameters (e.g. random shot/receiver locations and delta function as source wavelet).

Both time and frequency domains are each mostly suitable for specific applications of bandwidth extrapolation. Seismic data might be formulated as a 1D vector for single offset (seismogram) in the time domain and as a mono-frequency shot gather in the frequency domain. The single-trace time domain data is equivalent to a composition of multiple independent frequencies each featuring its own amplitude at a particular offset. Generally, lower frequencies have weaker amplitudes and thus the major challenge of time-domain low-frequency extrapolation is balancing the amplitudes within the band of target low-frequency data. On the contrary, the single-frequency data spans over multiple offsets and does not suffer from negligible amplitudes of events at farther offsets since this might be compensated by geometrical spreading correction. The challenge, however, is to ensure consistency between neighboring frequencies. To sum up, the single-trace time domain data is missing spatial connectivity with the data from nearest offsets while single-frequency data is missing temporal connectivity with the data at the next and previous discrete frequencies.

5.3 Future work

Synthetic data is often the only source of data pairs for supervised training of deep learning models when real-world references are not available. However, synthetic data on seismic waveforms is limited by the physics of numerical approximation of wave propagation as well as the geological realism of sample subsurface initializations. The concept of *domain adaptation* [185] offers a feasible way to facilitate the transition between training on synthetic data and application to field data. Our early experiments [186, 187, 188], have already shown a promise in application to near-surface land, marine and microseismic data.

Single common-shot-gather assumes independence of seismic data produced by neighboring shots. This is not true in reality. A more descriptive composition of input data should include a set of shot-gathers within the selected limited aperture. This would cover the pitfalls of time and frequency domains mentioned earlier. Specifically, set *multiple shot-gather data in the input* would ensure connectivity along the offset axis (time-domain data pitfall) while the natural composition of time-domain data as a set of multiple frequencies implicitly makes relations between neighboring frequencies unavoidable (frequency-domain data pitfall). Since deep learning builds the target data as a non-linear combination of basis functions (convolutional kernels) stored in its weights, the connectivity between frequencies (that are independent of each other from the wave-propagation point of view) should be considered as a pattern in the frequency-offset domain.

Deep learning approaches typically treat elements of input data as a collection of independent pixels. This is a valid assumption for processing real-world images while it is not accurate for seismic data that is usually ordered along with offset and time axis. By incorporating knowledge about data structure into the training would constraint optimization on the way to the global minimum of the objective function. The constraints might be directly added into the formulation of misfit functional (e.g. adding trace-wise correlation term) as well as embedded into the design of network architecture (e.g. using dilated convolutions to capture long-wavelength trends). However, an appealing option is to derive a common knowledge about seismic data by using the concept *of self-supervised learning* [189]. This way, the model learns in an unsupervised way to address the so-called pre-text tasks such as classification of rotation applied to the data or inpainting of missing areas. When succeeded in solving these pre-text tasks, the model might be trained to tackle the principal task. Alternatively, physics-informed neural networks [190] directly applied to geophysical problems of being used as a constraint might greatly improve the fidelity of constructed solutions.

5.4 Concluding remarks

Everything is information and diverse topics covered in this work support that. I showed that even mistakes that inversion makes during local optimization can shed the light on how to better initialize it. Another insight is that approximately solving two tasks might deliver better outcomes than trying to reach superior accuracy for only one from these tasks. The entire domain of artificial intelligence evolved around the concept of extracting useful knowledge from vast amounts of data. This is relevant to geophysics in particular since there are large seismic legacy datasets that are still waiting to reveal new insights after being re-processed by modern methods. The future of geophysical research, in my opinion, will focus on unsupervised methods that impose fewer assumptions on the complexity of wave phenomena, unveiling the secrets of the Earth in a truly data-driven way.

Papers published and submitted

Journal articles

1. **Ovcharenko, O.**, V. Kazei, T. Alkhalifah and D. Peter, Multi-task learning for low-frequency extrapolation and elastic model building. *Submitted to IEEE Transactions on Geoscience and Remote Sensing*, 2021.
2. Alkhalifah, T. and **Ovcharenko, O.**, Direct domain adaptation through mutual linear transformations. *Submitted to Journal of Machine Learning Research, 2021*
3. Alkhalifah, T., Wang, H. and **Ovcharenko, O.**, MLReal: Bridging the gap between training on synthetic data and real data applications in machine learning. *Submitted to Geophysical Journal International, 2021*
4. Kazei, V.,**Ovcharenko, O.**, Plotnitskii, P., Peter, D., Zhang, X. and Alkhalifah, T., 2021. Mapping full seismic waveforms to vertical velocity profiles by deep learning. *GEOPHYSICS*, 86(5), pp.1-50.
5. **Ovcharenko, O.**, V. Kazei, M. Kalita, D. Peter, and T.A. Alkhalifah, Deep learning for low-frequency extrapolation from multi-offset seismic data. *GEOPHYSICS*, 2019. 84(6): p. R1001-R1013.
6. **Ovcharenko, O.**, V. Kazei, D. Peter, and T. Alkhalifah, Variance-based model interpolation for improved full-waveform inversion in the presence of salt bodies. *GEOPHYSICS, 2018. 83(5)*: p. R541-R551.

Conference papers

1. **Ovcharenko, O.**, A. Baumstein, and E. Neumann, Surface-related multiple elimination through orthogonal encoding in the latent space of convolutional autoencoder, in *SEG Technical Program Expanded Abstracts* 2021. 2021, Society of Exploration Geophysicists.

2. **Ovcharenko, O.**, V. Kazei, D. Peter, and T. Alkhalifah. Transferring elastic low frequency extrapolation from synthetic to field data. in *83th EAGE Conference and Exhibition 2021*. 2021. European Association of Geoscientists & Engineers.

3. **Ovcharenko, O.**, V. Kazei, D. Peter, I. Silvestrov, A. Bakulin, and T. Alkhalifah, Dual-band generative learning for low-frequency extrapolation of the near-surface land data, in *SEG Technical Program Expanded Abstracts 2021*. 2021, Society of Exploration Geophysicists.

4. Kazei, V., **O. Ovcharenko**, and T. Alkhalifah, Velocity model building by deep learning: From general synthetics to field data application, in *SEG Technical Program Expanded Abstracts 2020*. 2020, Society of Exploration Geophysicists. p. 1561-1565.

5. Kazei, V., **O. Ovcharenko**, P. Plotnitskii, D. Peter, T. Alkhalifah, I. Silvestrov, A. Bakulin, and P. Zwartjes, Elastic near-surface model estimation from full waveforms by deep learning, in *SEG Technical Program Expanded Abstracts 2020*. 2020, Society of Exploration Geophysicists. p. 3872-3876.

6. Kazei, V., **O. Ovcharenko**, P. Plotnitskii, D. Peter, X. Zhang, and T. Alkhalifah. Deep learning tomography by mapping full seismic waveforms to vertical velocity profiles. in *82nd EAGE Annual Conference & Exhibition*. 2020. European Association of Geoscientists & Engineers.

7. **Ovcharenko, O**. and S. Hou. Deep Learning for Seismic Data Reconstruction: Opportunities and Challenges. in *First EAGE Digitalization Conference and Exhibition*. 2020. European Association of Geoscientists & Engineers.

8. **Ovcharenko, O.**, V. Kazei, P. Plotnitskiy, D. Peter, I. Silvestrov, A. Bakulin, and T. Alkhalifah, Extrapolating low-frequency prestack land data with deep learning,

in *SEG Technical Program Expanded Abstracts* 2020. 2020, Society of Exploration Geophysicists. p. 1546-1550.

9. Plotnitskii, P., V. Kazei, **O. Ovcharenko**, D. Peter, and T. Alkhalifah. Extrapolation of Low Wavenumbers in FWI Gradients by a Deep Convolutional Neural Network. in *82nd EAGE Annual Conference & Exhibition*. 2020. European Association of Geoscientists & Engineers.
10. Kazei, V., **O. Ovcharenko**, T. Alkhalifah, and F. Simons. Realistically textured random velocity models for deep learning applications. in *81st EAGE Conference and Exhibition 2019*. 2019. European Association of Geoscientists & Engineers.
11. **Ovcharenko, O.**, V. Kazei, D. Peter, and T. Alkhalifah, Style transfer for generation of realistically textured subsurface models, in *SEG Technical Program Expanded Abstracts* 2019. 2019, Society of Exploration Geophysicists. p. 2393-2397.
12. **Ovcharenko, O.**, V. Kazei, D. Peter, and T. Alkhalifah. Transfer learning for low frequency extrapolation from shot gathers for FWI applications. in *81st EAGE Conference and Exhibition 2019*. 2019. European Association of Geoscientists & Engineers.
13. Peter, D., F. Chen, **O. Ovcharenko**, A.E. Carmona, and Q. Liu. Improving full-waveform inversions using spectral-element seismic wave propagation on emerging HPC architectures. in *Geophysical Research Abstracts*. 2019.
14. Plotnitskii, P., T. Alkhalifah, **O. Ovcharenko**, and V. Kazei, Seismic model low wavenumber extrapolation by a deep convolutional neural network. *ASEG Extended Abstracts,* 2019. 2019(1): p. 1-5.
15. **Ovcharenko, O.**, J. Akram, and D. Peter. Feasibility of moment tensor inversion from a single borehole data using Artificial Neural Networks. in *GEO Bahrain 2018*. 2018. Search and Discovery.
16. **Ovcharenko, O.**, V. Kazei, D. Peter, X. Zhang, and T. Alkhalifah. Low-frequency data extrapolation using a feed-forward ANN. in *80th EAGE Conference and Exhibition 2018*. 2018. European Association of Geoscientists & Engineers.

17. Akram, J., **O. Ovcharenko**, and D. Peter, A robust neural network-based approach for microseismic event detection, in *SEG Technical Program Expanded Abstracts 2017*. 2017, Society of Exploration Geophysicists. p. 2929-2933.

18. **Ovcharenko, O.**, V. Kazei, D. Peter, and T. Alkhalifah. Neural network based low-frequency data extrapolation. in *3rd SEG FWI workshop: What are we getting*. 2017.

19. **Ovcharenko, O.**, V. Kazei, D.B. Peter, and T. Alkhalifah. Super-resolution Time-Lapse Seismic Waveform Inversion. in *AGU Fall Meeting Abstracts*. 2017.

20. **Ovcharenko, O.O.**, V.V. Kazei, D. Peter, and T. Alkhalifah. Variance-based salt body reconstruction. in *79th EAGE Conference and Exhibition 2017*. 2017. European Association of Geoscientists & Engineers.

21. Fuji, N., **O. Ovcharenko**, R. Martin, and C. Cuvilliez. Simple and accurate operators based on taylor expansion for 2D elastic seismogram calculation under geological discontinuities with regular cartesian grids. in *78th EAGE Conference and Exhibition 2016*. 2016. European Association of Geoscientists & Engineers.

REFERENCES

[1] J. Craig, F. Gerali, F. MacAulay, and R. Sorkhabi, “The history of the european oil and gas industry (1600s–2000s),” *Geological Society, London, Special Publications*, vol. 465, no. 1, pp. 1–24, 2018.

[2] P. Lailly, “The seismic inverse problem as a sequence of before stack migrations,” in *Conference on inverse scattering: theory and application*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1983, pp. 206–220.

[3] A. Tarantola, “Inversion of seismic reflection data in the acoustic approximation,” *Geophysics*, vol. 49, no. 8, pp. 1259–1266, 1984.

[4] R. G. Pratt, “Inverse theory applied to multi-source cross-hole tomography.” *Geophysical Prospecting*, vol. 38, no. 3, pp. 311–329, 1990.

[5] S. Ruder, “An overview of multi-task learning in deep neural networks,” *arXiv preprint arXiv:1706.05098*, 2017.

[6] T. Van Leeuwen and W. Mulder, “A correlation-based misfit criterion for wave-equation traveltime tomography,” *Geophysical Journal International*, vol. 182, no. 3, pp. 1383–1394, 2010.

[7] E. Bozdağ, J. Trampert, and J. Tromp, “Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements,” *Geophysical Journal International*, vol. 185, no. 2, pp. 845–870, 2011.

[8] B. Chi, L. Dong, and Y. Liu, “Full waveform inversion method using envelope objective function without low frequency data,” *Journal of Applied Geophysics*, vol. 109, pp. 36–46, 2014.

[9] Y. Choi and T. Alkhalifah, “Unwrapped phase inversion with an exponential damping,” *Geophysics*, vol. 80, pp. R251–R264, 2015.

[10] Y. Yang and B. Engquist, “Analysis of optimal transport and related misfit functions in FWI,” *Geophysics*, vol. 83, no. 1, pp. 1–25, 2017.

[11] M. Kalita, V. Kazei, Y. Choi, and T. Alkhalifah, “Regularized full-waveform inversion with automated salt-flooding,” *Geophysics*, vol. 84, no. 4, pp. 1–74, 2019.

[12] A. Baumstein, “Pocs-based geophysical constraints in multi-parameter full wavefield inversion,” in *75th EAGE Conference & Exhibition incorporating SPE EUROPEC 2013*. European Association of Geoscientists & Engineers, 2013, pp. cp–348.

[13] E. Esser, L. Guasch, T. v. Leeuwen, A. Y. Aravkin, and F. J. Herrmann, “Automatic salt delineation - Wavefield Reconstruction Inversion with convex constraints,” in *SEG Technical Program Expanded Abstracts 2015*. Society of Exploration Geophysicists, 2015, pp. 1337–1343.

[14] E. Esser, L. Guasch, F. J. Herrmann, and M. Warner, “Constrained waveform inversion for automatic salt flooding,” *The Leading Edge*, vol. 35, no. 3, pp. 235–239, 2016.

[15] Z. Guo and M. V. de Hoop, “Shape optimization and level set method in full waveform inversion with 3D body reconstruction,” in *SEG Technical Program Expanded Abstracts 2013*. Society of Exploration Geophysicists, 2013, pp. 1079–1083.

[16] A. Kadu, T. Van Leeuwen, and W. Mulder, “A parametric level-set approach for seismic full-waveform inversion,” in *SEG Technical Program Expanded Abstracts 2016*. Society of Exploration Geophysicists, 2016, pp. 1146–1150.

[17] T. Alkhalifah, “Full model wavenumber inversion,” *ASEG Extended Abstracts*, vol. 2015, no. 1, pp. 1–1, 2015.

[18] ——, “Full-model wavenumber inversion: An emphasis on the appropriate wavenumber continuation,” *Geophysics*, vol. 81, no. 3, pp. R89–R98, 2016.

[19] Y. Ma, D. Hale, B. Gong, and Z. J. Meng, “Image-guided sparse-model full waveform inversion,” *Geophysics*, vol. 77, pp. R189–R198, 2012. [Online]. Available: http://geophysics.geoscienceworld.org/content/77/4/R189

[20] Z. Wu and T. Alkhalifah, “The optimized gradient method for full waveform inversion and its spectral implementation,” *Geophysical Journal International*, vol. 205, no. 3, pp. 1823–1831, 2016.

[21] T. Alkhalifah, “Scattering-angle based filtering of the waveform inversion gradients,” *Geophysical Journal International*, vol. 200, no. 1, pp. 363–373, 2015. [Online]. Available: http://gji.oxfordjournals.org/content/200/1/363.abstract

[22] V. Kazei, E. Tessmer, and T. Alkhalifah, “Scattering angle-based filtering via extension in velocity,” in *SEG Technical Program Expanded Abstracts 2016*. Society of Exploration Geophysicists, 2016, pp. 1157–1162.

[23] I. F. Jones, “Tutorial: Incorporating near-surface velocity anomalies in pre-stack depth migration models,” *First Break*, vol. 30, no. 3, pp. 47–58, 2012.

[24] J. P. Leveille, I. F. Jones, Z.-Z. Zhou, B. Wang, and F. Liu, “Subsalt imaging for exploration, production, and development: A review,” *Geophysics*, vol. 76, pp. WB3–WB20, 2011.

[25] L. R. Lines and R. T. Newrick, *Fundamentals of geophysical interpretation*. Society of Exploration Geophysicists, 2004.

[26] C. Bunks, F. M. Saleck, S. Zaleski, and G. Chavent, “Multiscale seismic waveform inversion,” *Geophysics*, vol. 60, no. 5, pp. 1457–1473, 1995. [Online]. Available: http://geophysics.geoscienceworld.org/content/60/5/1457.abstract

[27] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, “Deep learning,” 2016.

[28] Y. Ito, “Approximation of continuous functions on r d by linear combinations of shifted rotations of a sigmoid function with and without scaling,” *Neural Networks*, vol. 5, no. 1, pp. 105–115, 1992.

[29] Y. Shi, X. Wu, and S. Fomel, "Waveform embedding: Automatic horizon picking with unsupervised deep learning," *Geophysics*, vol. 85, no. 4, pp. WA67–WA76, 2020.

[30] N. Pham, D. Dunlap, and S. Fomel, "Channel facies and faults multisegmentation in seismic volumes," in *First International Meeting for Applied Geoscience & Energy*. Society of Exploration Geophysicists, 2021, pp. 1430–1434.

[31] F. Qian, M. Yin, X.-Y. Liu, Y.-J. Wang, C. Lu, and G.-M. Hu, "Unsupervised seismic facies analysis via deep convolutional autoencoders," *Geophysics*, vol. 83, no. 3, pp. A39–A43, 2018.

[32] V. Puzyrev and C. Elders, "Unsupervised seismic facies classification using deep convolutional autoencoder," *arXiv preprint arXiv:2008.01995*, 2020.

[33] E. Valero Cano, J. Akram, and D. B. Peter, "Automatic seismic phase picking based on unsupervised machine learning classification and content information analysis," *Geophysics*, vol. 86, no. 4, pp. 1–57, 2021.

[34] L. Seydoux, R. Balestriero, P. Poli, M. De Hoop, M. Campillo, and R. Baraniuk, "Clustering earthquake signals and background noises in continuous seismic data with unsupervised deep learning," *Nature communications*, vol. 11, no. 1, pp. 1–12, 2020.

[35] O. M. Saad, G. Huang, Y. Chen, A. Savvaidis, S. Fomel, N. Pham, and Y. Chen, "Scalodeep: A highly generalized deep learning framework for real-time earthquake detection," *Journal of Geophysical Research: Solid Earth*, vol. 126, no. 4, p. e2020JB021473, 2021.

[36] L. Mosser, O. Dubrule, and M. J. Blunt, "Stochastic seismic waveform inversion using generative adversarial networks as a geological prior," *Mathematical Geosciences*, vol. 52, no. 1, pp. 53–79, 2020.

[37] Y. Wang, Q. Wang, W. Lu, Q. Ge, and X. Yan, "Seismic impedance inversion based on cycle-consistent generative adversarial network," *Petroleum Science*, 2021.

[38] A. Vereshagin, T. Wedberg, and A. Stefatos, "Predicting Vertical Resistivity By Machine Learning," in *81st EAGE Conference and Exhibition 2019*, 2019.

[39] Z. Zhang and T. Alkhalifah, "Regularized elastic full waveform inversion using deep learning," *Geophysics, accepted*, 2019.

[40] H. Kaur, N. Pham, S. Fomel, Z. Geng, L. Decker, B. Gremillion, M. Jervis, R. Abma, and S. Gao, "A deep learning framework for seismic facies classification," in *First International Meeting for Applied Geoscience & Energy*. Society of Exploration Geophysicists, 2021, pp. 1420–1424.

[41] X. Wu, L. Liang, Y. Shi, and S. Fomel, "Faultseg3d: Using synthetic data sets to train an end-to-end convolutional neural network for 3d seismic fault segmentation," *Geophysics*, vol. 84, no. 3, pp. IM35–IM45, 2019.

[42] A. Guitton, "3d convolutional neural networks for fault interpretation," in *80th EAGE Conference and Exhibition 2018*, 2018.

[43] J. Akram, O. Ovcharenko, and D. Peter, "A robust neural network-based approach for microseismic event detection," in *SEG Technical Program Expanded Abstracts 2017*. SEG, 2017, pp. 2929–2933.

[44] G. Loginov, D. Anton, D. Litvichenko, and S. Alyamkin, "The First-Break Detection For Real Seismic Data With Use of Convolutional Neural Network," in *81st EAGE Conference and Exhibition 2019*, 2019.

[45] O. Ovcharenko, J. Akram, and D. Peter, "Feasibility of moment tensor inversion from a single borehole data using artificial neural networks," *Search and Discovery*, 2018.

[46] M. Araya-Polo, J. Jennings, A. Adler, and T. Dahlke, "Deep-learning tomography," *The Leading Edge*, vol. 37, no. 1, pp. 58–66, 2018.

[47] L. Mosser, W. Kimman, J. Dramsch, S. Purves, A. De la Fuente Briceño, and G. Ganssle, "Rapid seismic domain transfer: Seismic velocity inversion and modeling using deep generative neural networks," in *80th EAGE Conference and Exhibition 2018*, 2018.

[48] A. Richardson, "Generative adversarial networks for model order reduction in seismic full-waveform inversion," *arXiv preprint arXiv:1806.00828*, 2018.

[49] V. Kazei, O. Ovcharenko, X. Zhang, D. Peter, and T. Alkhalifah, "Mapping seismic data cubes to vertical velocity profiles by deep learning: New full-waveform inversion paradigm?" *Geophysics, submitted*, 2019.

[50] O. Øye and E. Dahl, "Velocity model building from raw shot gathers using machine learning," in *81st EAGE Conference and Exhibition 2019*, vol. 2019, no. 1. European Association of Geoscientists & Engineers, 2019, pp. 1–5.

[51] Z. Geng, Z. Zhao, Y. Shi, X. Wu, S. Fomel, and M. Sen, "Deep learning for velocity model building with common-image gather volumes," *Geophysical Journal International*, 2021.

[52] F. Ten Kroode, S. Bergler, C. Corsten, J. W. de Maag, F. Strijbos, and H. Tijhof, "Broadband seismic data—the importance of low frequencies," *Geophysics*, vol. 78, no. 2, pp. WA3–WA14, 2013.

[53] O. Podgornova, S. Leaney, L. Liang *et al.*, "Analysis of resolution limits of VTI anisotropy with full waveform inversion," in *SEG Technical Program Expanded Abstracts 2015*. Society of Exploration Geophysicists, 2015, pp. 1188–1192.

[54] V. Kazei and T. Alkhalifah, "Waveform inversion for orthorhombic anisotropy with p waves: Feasibility and resolution," *Geophysical Journal International*, vol. 213, no. 2, pp. 963–982, 2018.

[55] G. Baeten, J. W. de Maag, R.-E. Plessix, R. Klaassen, T. Qureshi, M. Kleemeyer, F. t. Kroode, and Z. Rujie, "The use of low frequencies in a full-waveform inversion and impedance inversion land seismic case study," *Geophysical Prospecting*, vol. 61, no. 4, pp. 701–711, 2013.

[56] S. Chelminski, L. M. Watson, and S. Ronen, “Research note: Low-frequency pneumatic seismic sources,” *Geophysical Prospecting*, vol. 67, no. 6-Geophysical Instrumentation and Acquisition, pp. 1547–1556, 2019.

[57] P. Mora, “Inversion = migration + tomography,” *Geophysics*, vol. 54, no. 12, pp. 1575–1586, 1989.

[58] D. Wehner, M. Landrø, and L. Amundsen, “On low frequencies emitted by air guns at very shallow depths—an experimental study,” *Geophysics*, vol. 84, no. 5, pp. P61–P71, 2019.

[59] P. Maxwell and M. Lansley, “What receivers will we use for low frequencies?” in *SEG Technical Program Expanded Abstracts 2011*. Society of Exploration Geophysicists, 2011, pp. 72–76.

[60] S. Sambolian, S. Operto, A. Ribodetti, and L. Combe, “From slope tomography to fwi: Is the conventional workflow viable in complex settings?” in *SEG Technical Program Expanded Abstracts 2020*. Society of Exploration Geophysicists, 2020, pp. 890–894.

[61] A. Brenders, J. Dellinger, C. Kanu, Q. Li, and S. Michell, “The wolfspar® field trial: Results from a low-frequency seismic survey designed for fwi,” in *SEG Technical Program Expanded Abstracts 2018*. Society of Exploration Geophysicists, 2018, pp. 1083–1087.

[62] H. Roende, D. Bate, C. Udengaard, R. Malik, and Y. Huang, “Ultra-long offset sparse node project in deep water GoM for FWI and Imaging,” in *EAGE 2020 Annual Conference & Exhibition Online*, vol. 2020, no. 1. European Association of Geoscientists & Engineers, 2020, pp. 1–5.

[63] R. Soubaras and R. Dowle, “Variable-depth streamer–a broadband marine solution,” *First Break*, vol. 28, no. 12, 2010.

[64] W. Hu, “FWI without low frequency data-beat tone inversion,” in *SEG Technical Program Expanded Abstracts 2014*. Society of Exploration Geophysicists, 2014, pp. 1116–1120.

[65] R.-S. Wu, J. Luo, and B. Wu, “Seismic envelope inversion and modulation signal model,” *Geophysics*, vol. 79, no. 3, pp. WA13–WA24, 2014.

[66] Y. E. Li and L. Demanet, “Full-waveform inversion with extrapolated low-frequency data,” *Geophysics*, vol. 81, no. 6, pp. R339–R348, 2016.

[67] R. Wang and F. Herrmann, “Frequency down extrapolation with tv norm minimization,” in *SEG Technical Program Expanded Abstracts 2016*. Society of Exploration Geophysicists, 2016, pp. 1380–1384.

[68] Y. Li and L. Demanet, “Extrapolated full-waveform inversion: An image-space approach,” in *SEG Technical Program Expanded Abstracts 2017*. Society of Exploration Geophysicists, 2017, pp. 1682–1686.

[69] O. Ovcharenko, V. Kazei, M. Kalita, D. Peter, and T. Alkhalifah, “Deep learning for low-frequency extrapolation from multioffset seismic data,” *GEOPHYSICS*, vol. 84, no. 6, pp. R989–R1001, 2019. [Online]. Available: https://doi.org/10.1190/geo2018-0884.1

[70] W. Hu, Y. Jin, X. Wu, and J. Chen, “Progressive transfer learning for low frequency data prediction in full waveform inversion,” *arXiv preprint arXiv:1912.09944*, 2019.

[71] L. Demanet and N. Nguyen, “The recoverability limit for superresolution via sparsity,” *arXiv preprint arXiv:1502.01385*, 2015.

[72] L. Demanet and A. Townsend, “Stable extrapolation of analytic functions,” *Foundations of Computational Mathematics*, vol. 19, no. 2, pp. 297–331, 2019.

[73] O. Ovcharenko, V. Kazei, D. Peter, and T. Alkhalifah, “Neural network based low-frequency data extrapolation,” in *3rd SEG FWI workshop: What are we getting?* Society of Exploration Geophysicists, 2017.

[74] O. Ovcharenko, V. Kazei, D. Peter, X. Zhang, and T. Alkhalifah, “Low-frequency data extrapolation using a feed-forward ann,” in *80th EAGE Conference and Exhibition 2018*, 2018.

[75] H. Sun and L. Demanet, “Low frequency extrapolation with deep learning,” in *SEG Technical Program Expanded Abstracts 2018*. Society of Exploration Geophysicists, 2018, pp. 2011–2015.

[76] ——, “Extrapolated full waveform inversion with deep learning,” *Geophysics, submitted*, 2019.

[77] Y. Jin, W. Hu, X. Wu, and J. Chen, “Learn low wavenumber information in FWI via deep inception based convolutional networks,” in *SEG Technical Program Expanded Abstracts 2018*. Society of Exploration Geophysicists, 2018, pp. 2091–2095.

[78] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[79] M. Aharchaou and A. Baumstein, “Deep learning-based artificial bandwidth extension: Training on ultrasparse OBN to enhance towed-streamer FWI,” *The Leading Edge*, vol. 39, no. 10, pp. 718–726, 2020.

[80] S. Farris, M. Araya-Polo, J. Jennings, B. Clapp, and B. Biondi, “Tomography: A deep learning vs full-waveform inversion comparison,” in *First EAGE Workshop on High Performance Computing for Upstream in Latin America*, vol. 2018, no. 1. European Association of Geoscientists & Engineers, 2018, pp. 1–5.

[81] M. Araya-Polo, A. Adler, S. Farris, and J. Jennings, “Fast and accurate seismic tomography via deep learning,” in *Deep Learning: Algorithms and Applications*. Springer, 2020, pp. 129–156.

[82] Y. Wu and Y. Lin, “InversionNet: An Efficient and Accurate Data-driven Full Waveform Inversion,” *IEEE Transactions on Computational Imaging*, 2019.

[83] F. Yang and J. Ma, “Deep-learning inversion: A next-generation seismic velocity model building method,” *Geophysics*, vol. 84, no. 4, pp. R583–R599, 2019.

[84] S. Marcus, F. Avelino, d. O. N. Jorcy, D. Antonio, and R. Thomas, “Data-driven full-waveform inversion surrogate using conditional generative adversarial networks,” *arXiv preprint arXiv:2105.00100*, 2021.

[85] Z. Zhang and Y. Lin, “Data-driven seismic waveform inversion: A study on the robustness and generalization,” *IEEE Transactions on Geoscience and Remote sensing*, vol. 58, no. 10, pp. 6900–6913, 2020.

[86] J. Sun, K. A. Innanen, and C. Huang, “Physics-guided deep learning for seismic inversion with hybrid training and uncertainty analysis,” *Geophysics*, vol. 86, no. 3, pp. R303–R317, 2021.

[87] P. Plotnitskii, V. Kazei, O. Ovcharenko, D. Peter, and T. Alkhalifah, “Extrapolation of low wavenumbers in fwi gradients by a deep convolutional neural network,” in *82nd EAGE Annual Conference & Exhibition*, vol. 2020, no. 1. European Association of Geoscientists & Engineers, 2020, pp. 1–5.

[88] V. Kazei, O. Ovcharenko, P. Plotnitskii, D. Peter, X. Zhang, and T. Alkhalifah, “Mapping full seismic waveforms to vertical velocity profiles by deep learning,” *Geophysics*, vol. 86, no. 5, pp. 1–50, 2021.

[89] W. B. Beydoun and A. Tarantola, “First born and rytov approximations: Modeling and inversion conditions in a canonical example,” *The Journal of the Acoustical Society of America*, vol. 83, no. 3, pp. 1045–1055, 1988.

[90] J. Virieux and S. Operto, “An overview of full-waveform inversion in exploration geophysics,” *Geophysics*, vol. 74, no. 6, pp. WCC1–WCC26, 2009.

[91] O. Ovcharenko, V. Kazei, D. Peter, and T. Alkhalifah, “Variance-based salt body reconstruction,” in *79th EAGE Conference and Exhibition 2017*, 2017.

[92] F. Billette and S. Brandsberg-Dahl, “The 2004 BP velocity benchmark,” in *67th EAGE Conference & Exhibition*, 2005.

[93] M. Fehler and P. J. Keliher, *SEAM Phase 1: Challenges of subsalt imaging in tertiary basins, with emphasis on deepwater Gulf of Mexico*. Society of Exploration Geophysicists, 2011.

[94] L. Sirgue and R. G. Pratt, “Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies,” *Geophysics*, vol. 69, no. 1, pp. 231–248, 2004.

[95] W. Mulder and R.-E. Plessix, “Exploring some issues in acoustic full waveform inversion,” *Geophysical Prospecting*, vol. 56, no. 6, pp. 827–841, 2008.

[96] R. Brossier, S. Operto, and J. Virieux, “Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion,” *Geophysics*, vol. 74, pp. WCC105–WCC118, 2009.

[97] F. Sourbier, S. Operto, J. Virieux, P. Amestoy, and J.-Y. L’Excellent, “Fwt2d: A massively parallel program for frequency-domain full-waveform tomography of

wide-aperture seismic data—part 1: Algorithm," *Computers & Geosciences*, vol. 35, no. 3, pp. 487–495, 2009.

[98] P. Bullen, "Handbook of means and their inequalities, vol. 260 Kluwer Academic Publisher," *Dordrecht, The Netherlands*, 2003.

[99] D. C. Liu and J. Nocedal, "On the limited memory bfgs method for large scale optimization," *Mathematical programming*, vol. 45, no. 1, pp. 503–528, 1989.

[100] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[101] G. Goertzel, "An algorithm for the evaluation of finite trigonometric series," *The American Mathematical Monthly*, vol. 65, no. 1, pp. 34–35, 1958.

[102] L. Métivier, R. Brossier, J. Virieux, and S. Operto, "Full waveform inversion and the truncated newton method," *SIAM Journal on Scientific Computing*, vol. 35, no. 2, pp. B401–B437, 2013.

[103] T. Alkhalifah and Z. Wu, "Multiscattering inversion for low-model wavenumbers," *GEOPHYSICS*, vol. 81, pp. R417–R428, 2016. [Online]. Available: https://doi.org/10.1190/geo2015-0650.1

[104] P. C. Sava and S. Fomel, "Angle-domain common-image gathers by wavefield continuation methods," *Geophysics*, vol. 68, pp. 1065–1074, 2003.

[105] J. Etgen, S. H. Gray, and Y. Zhang, "An overview of depth imaging in exploration geophysics," *Geophysics*, vol. 74, no. 6, pp. WCA5–WCA17, 2009.

[106] I. F. Jones and I. Davison, "Seismic imaging in and around salt bodies," *Interpretation*, vol. 2, no. 4, pp. SL1–SL20, 2014.

[107] J. Dellinger, A. Brenders, X. Shen, I. Ahmed, J. Sandschaper, and J. Etgen, "What mistakes are we making while interpreting salt? Could FWI help?" in *79th EAGE Conference and Exhibition 2017-Workshops*, 2017.

[108] J. Dellinger, A. J. Brenders, J. Sandschaper, C. Regone, J. Etgen, I. Ahmed, and K. Lee, "The Garden Banks model experience," *The Leading Edge*, vol. 36, no. 2, pp. 151–158, 2017.

[109] M. Kalita and T. Alkhalifah, "Multiscale full-waveform inversion using flux corrected transport," in *SEG Technical Program Expanded Abstracts 2018*. Society of Exploration Geophysicists, 2018, pp. 1153–1157.

[110] O. Ovcharenko, V. Kazei, D. Peter, and T. Alkhalifah, "Variance-based model interpolation for improved full-waveform inversion in the presence of salt bodies," *Geophysics*, vol. 83, no. 5, pp. 1–60, 2018.

[111] L. Sirgue and R. Pratt, "Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies," *Geophysics*, vol. 69, no. 1, pp. 231–248, 2004. [Online]. Available: http://library.seg.org/doi/abs/10.1190/1.1649391

[112] V. V. Kazei, B. M. Kashtan, V. N. Troyan, and W. A. Mulder, “Spectral sensitivity analysis of FWI in a constant-gradient background velocity model,” in *75th EAGE Conference & Exhibition incorporating SPE EUROPEC 2013*, 2013.

[113] R. G. Pratt, Z.-M. Song, P. R. Williamson, and M. R. Warner, “Two-dimensional velocity models from wide-angle seismic data by waveform inversion,” *Geophysical Journal International*, vol. 124, no. 2, pp. 323–340, 1996.

[114] J. Hudson and J. Heritage, “The use of the Born approximation in seismic scattering problems,” *Geophysical Journal International*, vol. 66, no. 1, pp. 221–240, 1981.

[115] A. J. Devaney, “Geophysical diffraction tomography,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. GE-22, no. 1, pp. 3–13, jan. 1984.

[116] V. Kazei, V. Troyan, B. Kashtan, and W. Mulder, “On the role of reflections, refractions and diving waves in full-waveform inversion,” *Geophysical Prospecting*, vol. 61, no. 6, pp. 1252–1263, 2013.

[117] V. Kazei, B. Kashtan, V. Troyan, and W. Mulder, “FWI spectral sensitivity analysis in the presence of a free surface,” in *SEG Technical Program Expanded Abstracts 2015*. Society of Exploration Geophysicists, 2015, pp. 1415–1419.

[118] O. Podgornova, S. Leaney, and L. Liang, “Resolution of vti anisotropy with elastic full-waveform inversion: Theory and basic numerical examples,” *Geophysical Journal International*, vol. 214, no. 1, pp. 200–218, 2018.

[119] V. Kazei and T. Alkhalifah, “Scattering radiation pattern atlas: What anisotropic elastic properties can body waves resolve?” *Journal of Geophysical Research: Solid Earth*, vol. 124, no. 3, pp. 2781–2811, 2019.

[120] W. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in nervous activity,” *Bulletin of Mathematical Biophysics*, vol. 7, pp. 115–133, 1943.

[121] M. I. Jordan and T. M. Mitchell, “Machine learning: Trends, perspectives, and prospects,” *Science*, vol. 349, no. 6245, pp. 255–260, 2015.

[122] J. Schmidhuber, “Deep learning in neural networks: An overview,” *Neural networks*, vol. 61, pp. 85–117, 2015.

[123] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, “Efficient backprop,” in *Neural networks: Tricks of the trade*. Springer, 2012, pp. 9–48.

[124] R. Eldan and O. Shamir, “The power of depth for feedforward neural networks,” in *Conference on Learning Theory*, 2016, pp. 907–940.

[125] G. Christakos, *Random field models in earth sciences*. Courier Corporation, 2012.

[126] V. Kazei, O. Ovcharenko, T. Alkhalifah, and F. Simons, “Realistically Textured Random Velocity Models for Deep Learning Applications,” in *81st EAGE Conference and Exhibition 2019*, 2019.

[127] O. Ovcharenko, V. Kazei, D. Peter, and T. Alkhalifah, “Style transfer for generation of realistically textured subsurface models,” in *SEG Technical Program Expanded Abstracts 2019*. Society of Exploration Geophysicists, 2019.

[128] A. Karpathy, “CS231n Convolutional Neural Networks for Visual Recognition,” 2018. [Online]. Available: http://cs231n.github.io/convolutional-networks/

[129] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, “Fast and accurate deep network learning by exponential linear units (elus),” *arXiv preprint arXiv:1511.07289*, 2015.

[130] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.

[131] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks,” in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 2010, pp. 249–256.

[132] F. Chollet *et al.*, “Keras,” https://keras.io, 2015.

[133] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, “TensorFlow: Large-scale machine learning on heterogeneous systems,” 2015, software available from tensorflow.org. [Online]. Available: https://www.tensorflow.org/

[134] C. L. Giles and T. Maxwell, “Learning, invariance, and generalization in high-order neural networks,” *Applied optics*, vol. 26, no. 23, pp. 4972–4978, 1987.

[135] V. Kazei, M. Kalita, and T. Alkhalifah, “Salt-body Inversion with Minimum Gradient Support and Sobolev Space Norm Regularizations,” in *79th EAGE Conference and Exhibition 2017*, 2017.

[136] C. M. Bishop, “Training with noise is equivalent to Tikhonov regularization,” *Neural computation*, vol. 7, no. 1, pp. 108–116, 1995.

[137] O. Ovcharenko, V. Kazei, D. Peter, and T. Alkhalifah, “Transfer Learning For Low Frequency Extrapolation From Shot Gathers For FWI Applications,” in *81st EAGE Conference and Exhibition 2019*, 2019.

[138] J. F. Claerbout, *Imaging the earth’s interior*. Blackwell scientific publications Oxford, 1985.

[139] E. Bozdağ, D. Peter, M. Lefebvre, D. Komatitsch, J. Tromp, J. Hill, N. Podhorszki, and D. Pugmire, “Global adjoint tomography: first-generation model,” *Geophysical Journal International*, vol. 207, no. 3, pp. 1739–1766, 2016.

[140] M. Warner, A. Ratcliffe, T. Nangoo, J. Morgan, A. Umpleby, N. Shah, V. Vinje, I. Stekl, L. Guasch, C. Win *et al.*, "Anisotropic 3d full-waveform inversion," *Geophysics*, vol. 78, no. 2, pp. R59–R80, 2013.

[141] W. Hu, J. Chen, J. Liu, and A. Abubakar, "Retrieving low wavenumber information in fwi: An overview of the cycle-skipping phenomenon and solutions," *IEEE Signal Processing Magazine*, vol. 35, no. 2, pp. 132–141, 2018.

[142] M. Warner and L. Guasch, "Adaptive waveform inversion: Theory," *Geophysics*, vol. 81, no. 6, pp. R429–R445, 2016.

[143] F. Chen and D. Peter, "Constructing misfit function for full waveform inversion based on sliced wasserstein distance," in *80th EAGE Conference and Exhibition 2018*, vol. 2018, no. 1. European Association of Geoscientists & Engineers, 2018, pp. 1–5.

[144] Y. Hu, L. Han, R. Wu, and Y. Xu, "Multi-scale time-frequency domain full waveform inversion with a weighted local correlation-phase misfit function," *Journal of Geophysics and Engineering*, vol. 16, no. 6, pp. 1017–1031, 2019.

[145] B. Sun and T. Alkhalifah, "ML-misfit: Learning a robust misfit function for full-waveform inversion using machine learning," in *82nd EAGE Annual Conference & Exhibition*, vol. 2020, no. 1. European Association of Geoscientists & Engineers, 2020, pp. 1–5.

[146] T. Van Leeuwen and F. J. Herrmann, "Mitigating local minima in full-waveform inversion by expanding the search space," *Geophysical Journal International*, vol. 195, no. 1, pp. 661–667, 2013.

[147] Z.-d. Zhang, T. Alkhalifah, E. Z. Naeini, and B. Sun, "Multiparameter elastic full waveform inversion with facies-based constraints," *Geophysical Journal International*, vol. 213, no. 3, pp. 2112–2127, 2018.

[148] V. Kazei, E. Tessmer, and T. Alkhalifah, "Scattering angle-based filtering via extension in velocity," in *SEG Technical Program Expanded Abstracts 2016*. Society of Exploration Geophysicists, 2016, pp. 1157–1162.

[149] Y. Ma, D. Hale, B. Gong, and Z. Meng, "Image-guided sparse-model full waveform inversion," *Geophysics*, vol. 77, no. 4, pp. R189–R198, 2012.

[150] C. Bunks, F. M. Saleck, S. Zaleski, and G. Chavent, "Multiscale seismic waveform inversion," *Geophysics*, vol. 60, no. 5, pp. 1457–1473, 1995.

[151] O. Ovcharenko, V. Kazei, D. Peter, and T. Alkhalifah, "Neural network based low-frequency data extrapolation," in *3rd SEG FWI workshop: What are we getting*, 2017.

[152] H. Sun and L. Demanet, "Deep learning for low frequency extrapolation of multicomponent data in elastic full waveform inversion," *arXiv preprint arXiv:2101.00099*, 2020.

[153] J. Fang, H. Zhou, Y. Elita Li, Q. Zhang, L. Wang, P. Sun, and J. Zhang, "Data-driven low-frequency signal recovery using deep-learning predictions in full-waveform inversion," *Geophysics*, vol. 85, no. 6, pp. A37–A43, 2020.

[154] O. Ovcharenko, V. Kazei, P. Plotnitskiy, D. Peter, I. Silvestrov, A. Bakulin, and T. Alkhalifah, "Extrapolating low-frequency prestack land data with deep learning," in *SEG Technical Program Expanded Abstracts 2020*. Society of Exploration Geophysicists, 2020, pp. 1546–1550.

[155] G. Fabien-Ouellet, "Low-frequency generation and denoising with recursive convolutional neural networks," in *SEG Technical Program Expanded Abstracts 2020*. Society of Exploration Geophysicists, 2020, pp. 870–874.

[156] W. Hu, Y. Jin, X. Wu, and J. Chen, "Physics-guided self-supervised learning for low frequency data prediction in FWI," in *SEG Technical Program Expanded Abstracts 2020*. Society of Exploration Geophysicists, 2020, pp. 875–879.

[157] ——, "Progressive transfer learning for low-frequency data prediction in full waveform inversion," *Geophysics*, vol. 86, no. 4, pp. 1–82, 2021.

[158] M. Wang, S. Xu, and H. Zhou, "Self-supervised learning for low frequency extension of seismic data," in *SEG Technical Program Expanded Abstracts 2020*. Society of Exploration Geophysicists, 2020, pp. 1501–1505.

[159] R. Caruana, "Multitask learning," *Machine learning*, vol. 28, no. 1, pp. 41–75, 1997.

[160] F. J. Herrmann and X. Li, "Efficient least-squares imaging with sparsity promotion and compressive sensing," *Geophysical prospecting*, vol. 60, no. 4-Simultaneous Source Methods for Seismic Data, pp. 696–712, 2012.

[161] A. Kendall, Y. Gal, and R. Cipolla, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7482–7491.

[162] Y. Wang, X. Tao, X. Qi, X. Shen, and J. Jia, "Image inpainting via generative multi-column convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2018, pp. 331–340.

[163] R. Caruana, S. Lawrence, and L. Giles, "Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping," *Advances in neural information processing systems*, pp. 402–408, 2001.

[164] F. Chollet, *Deep learning with Python*. Simon and Schuster, 2017.

[165] N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. P. Tang, "On large-batch training for deep learning: Generalization gap and sharp minima," *arXiv preprint arXiv:1609.04836*, 2016.

[166] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.

[167] L. N. Smith and N. Topin, "Super-convergence: Very fast training of neural networks using large learning rates," in *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*, vol. 11006. International Society for Optics and Photonics, 2019, p. 1100612.

[168] Y. Kim, D.-J. Min, and C. Shin, "Frequency-domain reverse-time migration with source estimation," *Geophysics*, vol. 76, no. 2, pp. S41–S49, 2011.

[169] M. Kalita and T. Alkhalifah, "Efficient full waveform inversion using the excitation representation of the source wavefield," *Geophysical Journal International*, vol. 210, no. 3, pp. 1581–1594, 2017.

[170] S. Feng, Y. Lin, and B. Wohlberg, "Physically realistic training data construction for data-driven full-waveform inversion and traveltime tomography," in *SEG Technical Program Expanded Abstracts 2020*. Society of Exploration Geophysicists, 2020, pp. 3472–3476.

[171] Y. Ren, L. Nie, S. Yang, P. Jiang, and Y. Chen, "Building complex seismic velocity models for deep learning inversion," *IEEE Access*, vol. 9, pp. 63 767–63 778, 2021.

[172] G. Gardner, L. Gardner, and A. Gregory, "Formation velocity and density—the diagnostic basics for stratigraphic traps," *Geophysics*, vol. 39, no. 6, pp. 770–780, 1974.

[173] P. Mora and Z. Wu, "Elastic versus acoustic inversion for marine surveys," *Geophysical Journal International*, vol. 214, no. 1, pp. 596–622, 2018.

[174] Ò. C. Agudo, N. V. da Silva, M. Warner, and J. Morgan, "Acoustic full-waveform inversion in an elastic world," *Geophysics*, vol. 83, no. 3, pp. R257–R271, 2018.

[175] N. Thiel, T. Hertweck, and T. Bohlen, "Comparison of acoustic and elastic full-waveform inversion of 2d towed-streamer data in the presence of salt," *Geophysical prospecting*, vol. 67, no. 2, pp. 349–361, 2019.

[176] D. Köhn, "Time domain 2d elastic full waveform tomography," Ph.D. dissertation, Christian-Albrechts Universität Kiel, 2011.

[177] S. Fomel, P. Sava, I. Vlad, Y. Liu, and V. Bashkardin, "Madagascar: Open-source software project for multidimensional data analysis and reproducible computational experiments," *Journal of Open Research Software*, vol. 1, no. 1, 2013.

[178] C. E. Birnie, "Statistical methods for ambient noise characterisation, modelling and suppression: theory and applications for surface microseismic monitoring." Ph.D. dissertation, University of Leeds, 2018.

[179] G. S. Martin, R. Wiley, and K. J. Marfurt, "Marmousi2: An elastic upgrade for Marmousi," *The Leading Edge*, vol. 25, no. 2, pp. 156–166, 2006.

[180] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *ArXiv*, vol. abs/1505.04597, 2015.

[181] G. Bonaccorso, *Machine Learning Algorithms: Popular algorithms for data science and machine learning*. Packt Publishing Ltd, 2018.

[182] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[183] D. Köhn, D. De Nil, A. Kurzmann, A. Przebindowska, and T. Bohlen, “On the influence of model parametrization in elastic full waveform tomography,” *Geophysical Journal International*, vol. 191, no. 1, pp. 325–345, Oct. 2012.

[184] C. Song and T. A. Alkhalifah, “Efficient wavefield inversion with outer iterations and total variation constraint,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 8, pp. 5836–5846, 2020.

[185] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, “Domain-adversarial training of neural networks,” *The journal of machine learning research*, vol. 17, no. 1, pp. 2096–2030, 2016.

[186] O. Ovcharenko, V. Kazei, D. Peter, I. Silvestrov, B. Andrey, and T. Alkhalifah, “Dual-band generative learning for low-frequency extrapolation of the near-surface land data,” in *2021 SEG Annual Meeting*. Society of Exploration Geophysicists, 2021.

[187] T. Alkhalifah and O. Ovcharenko, “Direct domain adaptation through reciprocal linear transformations,” *arXiv preprint arXiv: 2108.07600*, 2021.

[188] T. Alkhalifah, H. Wang, and O. Ovcharenko, “MLReal: Bridging the gap between training on synthetic data and real data applications in machine learning,” *arXiv preprint arXiv: 2109.05294*, 2021.

[189] A. Bardes, J. Ponce, and Y. LeCun, “Vicreg: Variance-invariance-covariance regularization for self-supervised learning,” *arXiv preprint arXiv:2105.04906*, 2021.

[190] U. b. Waheed, T. Alkhalifah, E. Haghighat, C. Song, and J. Virieux, “Pinntomo: Seismic tomography using physics-informed neural networks,” *arXiv preprint arXiv:2104.01588*, 2021.

APPENDICES

A Journals’ permission and copyright information

This thesis reprints the content of several published research papers authored by me. The foreword for each chapter of this thesis indicates the source and the status of respective publications when necessary. The copyright of these articles belongs to their respective publishers and no copyright infringement is intended.

- *SEG Geophysics Journal* (https://library.seg.org/page/policies/open-access)

  Authors may reuse all or part of their papers published with SEG in a thesis or dissertation that authors write and are required to submit to satisfy criteria of degree-granting institutions.

- *IEEE Transactions on Geoscience and Remote Sensing* (https://journals.ieeeauthorcenter.ieee.org/choose-a-publishing-agreement/avoid-infringement-upon-ieee-copyright/)

  You may reuse your published article in your thesis or dissertation without requesting permission, provided that you fulfill the following requirements depending on which aspects of the article you wish to reuse.

  Text excerpts: Provide the full citation of the original published article followed by the IEEE copyright line: 20XX IEEE. If you are reusing a substantial portion of your article and you are not the senior author, obtain the senior author’s approval before reusing the text. Graphics and tables: The IEEE copyright line (20XX IEEE) should appear with each reprinted graphic and table. Full text article: Include the following copyright notice in the references: “20XX IEEE. Reprinted, with permission, from [full citation of original published article].” When posting your thesis on your university website, include the following message:

"In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [name of university or educational entity]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to

http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink. If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation." Only the accepted version of your article, not the final published version, may be posted online in your thesis.