



HAL
open science

Near-Surface Seismic Arrival Time Picking with Transfer and Semi-Supervised Learning

Ngo Nghi Truyen Huynh, Roland Martin, Thomas Oberlin, Bastien Plazolles

► **To cite this version:**

Ngo Nghi Truyen Huynh, Roland Martin, Thomas Oberlin, Bastien Plazolles. Near-Surface Seismic Arrival Time Picking with Transfer and Semi-Supervised Learning. *Surveys in Geophysics*, 2023, 10.1007/s10712-023-09783-y . hal-03778597v2

HAL Id: hal-03778597

<https://hal.science/hal-03778597v2>

Submitted on 28 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Near-surface seismic arrival time picking with transfer and semi-supervised learning

Ngo Nghi Truyen Huynh · Roland Martin · Thomas Oberlin · Bastien Plazolles

Abstract The understanding of subsurface information on the Earth is crucial in numerous fields such as economics of oil and gas, geophysical exploration, archaeology and hydro-geophysics, particularly in a context of climate change. The methodology consists in reconstructing the seismic velocity model of the near surface, that contains information about the basement structure, by solving the inverse problem and resolving the related complex nonlinear systems with the data collected from seismic experiments and measurements. In the last few years, many deep neural networks have been proposed to simplify the seismic inversion problem based, for instance, on automatic differentiation of the adjoint operator, or on automatic arrival time picking. However, such approaches require a large amount of labeled training data, which are hardly available in real applications. We present here a deep learning approach for arrival time picking, aimed to deal with unlabeled data. The main building blocks are transfer learning, as well as a semi-supervised learning strategy where the pseudo-labels are greedily computed with robust regression, and classification algorithms. The hybrid method showcases very high scores when evaluating on synthetic data, and its application to a real dataset containing a limited amount of labeled data shows the computational efficiency and very accurate results.

Keywords Seismic modeling, Near-surface seismic imaging, Arrival time picking, Deep convolutional neural networks, Transfer learning, Semi-supervised learning

Highlights

- Automatic arrival time picking in the context of near-surface seismic data
- Transfer and semi-supervised learning for both synthetic and real data
- Hybrid post-processing method based on support vector machine for enhancing pseudo-labels

1 Introduction

Seismic exploration is the use of seismic imaging to investigate the subsurface structures of the Earth (Talwani and Kessinger, 2003). It plays a crucial role in the delineation of near surface geology for economic deposits of oil, gas, or minerals, but also for engineering purposes, archaeological, geophysical and geotechnical scientific studies. Among the parameters used for seismic exploration purpose, seismic velocities are some of the most important ones. They can be defined as the speed with which an elastic wave propagates through a medium,

Ngo Nghi Truyen Huynh

INSA Toulouse, Département Génie Mathématique et Modélisation, 31400 Toulouse, France

Present address: INRAE, Aix Marseille Université, RECOVER, 13100 Aix-en-Provence, France

Roland Martin

Laboratoire GET, Université Toulouse 3 Paul Sabatier, IRD, CNRS UMR 5563, Observatoire Midi-Pyrénées, 31400 Toulouse, France

E-mail: roland.martin@get.omp.eu

Thomas Oberlin

ISAE-SUPAERO, Université de Toulouse, 31400 Toulouse, France

Bastien Plazolles

Laboratoire GET, Université Toulouse 3 Paul Sabatier, IRD, CNRS UMR 5563, Observatoire Midi-Pyrénées, 31400 Toulouse, France

and thus considered to be seismic properties. Seismic inversion consists in reconstructing the subsurface velocity model by processing the data collected from seismic experiments (Fig. 1).

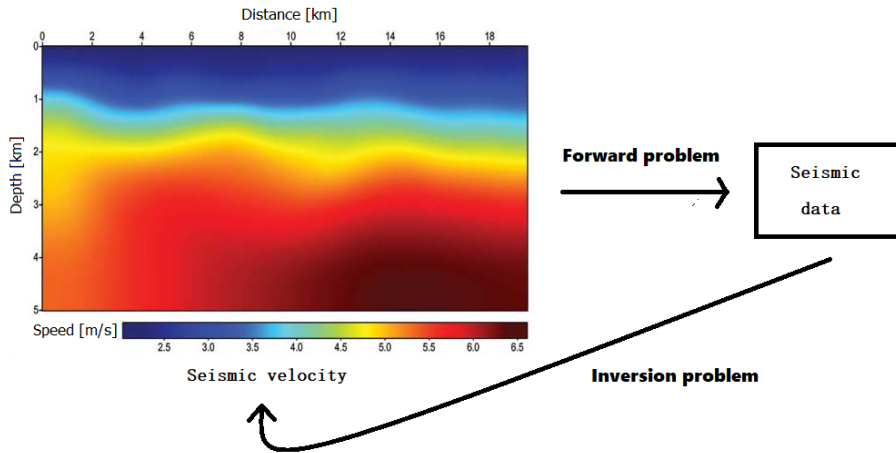


Fig. 1. Illustration of the seismic inversion problem.

Basically, such velocity information can be derived by travel-time tomography (Hole, 1992; Hobro et al., 2003; Bording et al., 1987; Zelt and Barton, 1998; Rawlinson et al., 2003) or full-waveform inversion (FWI) (Virieux et al., 2017; Xu et al., 2012; Virieux and Operto, 2009; Tarantola, 1987, 1988, 1984; Tarantola and Valette, 1982; Baeten et al., 2013; Plessix, 2006; Zhu et al., 2009; Tromp et al., 2005), that have proven to be efficient for acoustic model building in many geological scenarios. However, these techniques are sophisticated to implement and still have limitations affected by human interventions (Kosloff et al., 1996; Jones, 2010). Currently, most seismic inversion problems are addressed by:

- physics-driven seismic inversion based on adjoint theory (commonly used in the geophysical community). This method attempts to minimize iteratively a cost function defined by the differences between the observed and calculated data (e.g., l^2 -norm). The inversion process is non-linear and the model corrections are thus generally estimated iteratively by computing the gradient of the cost function at each step of the inversion algorithms. This gradient is based on some model parameters such as the density ρ and the speed of seismic phases V_P , V_S , and calculated by the adjoint state method, permits the computation of the first derivative of a physical observable or an associated objective function with respect to model parameters (Fichtner et al., 2006). These gradients are also called sensitivity kernels (Liu and Tromp, 2006; Peter et al., 2011; Tromp et al., 2008, 2005; Zhu et al., 2009; Plessix, 2006).
- data-driven seismic inversion/tomography based on deep learning techniques, that uses the supervised learning method to reconstruct the velocity model directly from recorded seismic data via deep convolutional neural networks (CNN) (Adler et al., 2021; Yang and Ma, 2019; Zheng et al., 2019; Li et al., 2020; Araya-Polo et al., 2018; Bianco et al., 2019; Earp et al., 2020; Kong et al., 2019; Yu and Ma, 2021). Apart from those end-to-end deep inversion models, deep learning can also be used to help or improve a physical inversion, we will term those approaches *hybrid*. Several works for instance considered the automatic differentiation power of deep learning to help computing the adjoint operator (Cao and Liao, 2015; Richardson, 2018; Zhu et al., 2021; Baydin et al., 2018). Other approaches used deep neural networks for extracting dispersion curves in the frequency domain (Dai et al., 2021) to get information on the different seismic modes (surface wave modes, guided modes, etc.), or directly in the time domain for arrival time picking (Wang et al., 2019; Zhu and Beroza, 2018).

Although appealing, the end-to-end deep learning techniques fully resolving the inverse problem require a huge amount of labeled data, which either costs a lot of time and money, or massively relies on simulations. On the contrary, the hybrid approaches require less complex models and smaller datasets, while keeping relying on well grounded physical models. Among them, automatic arrival time picking is one of the solutions that require the fewest data labels, as shown in Table 1. It consists in detecting the first arrival time of body or surface waves (Rayleigh waves) and thus helps reconstruct the P and S seismic velocity model thanks to the inversion of the P and S wave phases arrival times picked in the time series (seismograms). The P velocity phases are first arrival times that are picked and the Rayleigh waves are slower than the P-waves but are

more energetic with high amplitudes and can also be detected to give information on the S velocity model close to the (free-) surface. Namely, these arrival times allow us to access to the time windows and thus compute the gradient of the cost function by using adjoint-based inversion techniques for FWI or travel-time inversion methods (tomography) based on ray tracing method approaches (Lecomte et al., 2015; Podvin and Lecomte, 1991; Qian et al., 2007; Fomel et al., 2009; Huang et al., 2019; Billette and Lambare, 1998).

However, all the components (both vertical and horizontal) of the recorded signals are generally available in teleseismic and geophysical exploration setup configuration contexts and allow to detect first P and S travel times as well as different converted waves and even Rayleigh waves. But, as it is often the case in many near-surface geophysics surveys (like in our present seismic experiment for instance), only the vertical component is available and the P-wave first arrivals can be essentially and accurately picked while the first S-wave arrival times and Rayleigh train-waves can not be accurately picked and can rather only be partially and qualitatively detected. Generally, in the near-surface community, only the P-waves are picked manually or automatically, while the S and Rayleigh (dispersive)-train waves are not picked in the time domain due to too many converted waves generated by the heterogenous medium under study and they are thus rather studied in the frequency domain through dispersion analysis. However, these qualitative arrival times can sometimes provide an estimate of the time windows giving information at far offset on the dispersive train-waves corresponding to surface (Rayleigh) waves that could be inverted in the frequency domain and provide estimates of Rayleigh and S-wave speeds at depth. Here in this study we will thus show how the P-waves arrival times can be very well detected for only one vertical component using deep learning techniques when only one component is available while dispersive wave-trains can only be detected qualitatively and for receivers far away enough from the source. We also show here, as a matter of preliminary information, what our deep learning approach can provide (even partially or inaccurately) in terms of surface (dispersive) wave detection, even if deeper investigations should be done in the future to process them better (in the time or frequency domain).

Method	Unit data sample
Arrival time picking	1 trace
Extraction of dispersion curves	r traces
End-to-end inversion	$s \times r$ traces

Table 1. Illustration of the unit data sample for different deep learning approaches. A seismic experiment with s sources and r receivers makes a single data sample for an end-to-end inversion, but $r \times s$ samples for time picking.

Note that arrival time picking is not a novel technique in seismic inversion. Manual time picking has been used a lot in the past decades, as well as automatic time picking methods such as AR-picker (Akazawa, 2004). But none of them were fully satisfactory, manual techniques being too much expensive in terms of time spent in human supervision, while automatic techniques suffered from a limited accuracy. Data-driven techniques have successfully solved those drawbacks in some specific seismic configurations, especially with the use of deep neural networks that led to a significant improvement in terms of accuracy. State-of-the-art techniques are PickNet (Wang et al., 2019) and PhaseNet (Zhu and Beroza, 2018). PickNet is constructed from VGG-16 net (Simonyan and Zisserman, 2014), a very deep CNN for large-scale image recognition. PhaseNet on the other hand is based on U-Net (Ronneberger et al., 2015), that was introduced for biomedical image segmentation and particularly, has shown its efficiencies for segmentation of neuronal structures in electron microscopic stacks. The numerical experiments performed in Mousavi et al. (2020) have shown that PhaseNet has a better score and standard deviation of error than PickNet. PhaseNet has been designed for three component (one vertical and two horizontal components) teleseismic data and we want to use and improve it for our near surface application (active seismic) even when only the vertical component is available. In this paper, we thus focus on an automatic arrival time picking method based on PhaseNet and we improve this method as well as discuss different strategies in transfer learning complemented by support-vector-machine-based semi-supervised techniques for pseudo-labeling.

The paper is organized as follows. In Section 2, we first describe PhaseNet, including its architecture and its use for passive seismic data. Following this, in Sections 3 and 4, we describe our dataset and the transfer and semi-supervised methodology we use to adapt PhaseNet to this specific context. Experimental results and discussion are then provided in Section 5, while Section 6 concludes this paper.

2 PhaseNet

PhaseNet (Zhu and Beroza, 2018) is built on a U-Net, a deep CNN strategy that was previously developed for biomedical image segmentation in 2015 (Ronneberger et al., 2015). It consists of two connected symmetrical branches containing downsampling and upsampling layers (Fig. 2). Each layer is made of a convolution followed by ReLU activation function. The downsampling branch is similar to a classification CNN. Its role is to extract interesting features, it reduces the spatial size while expands the number of channels or features. Conversely, the upsampling branch uses transpose convolution to reduce the number of features and expands the spatial dimension, in order to obtain a segmentation map with same size as the input image. To ease this upsampling process, U-Nets contain skip connections that link each corresponding layers of the two branches. Skip connections are generally introduced to avoid the overfitting and they improve the training stability by keeping information of the down-sampling CNN layers. Those connections are concatenative: some channels of the downsampling layer are directly passed to the corresponding upsampling layer. They can also be seen as a way to keep high-frequency coefficients, which help reconstructing the segmentation map with the same resolution as the input image. The inputs are three-component seismograms while the outputs are the predicted probability distributions of observing a P-wave, S-wave and noise/background.

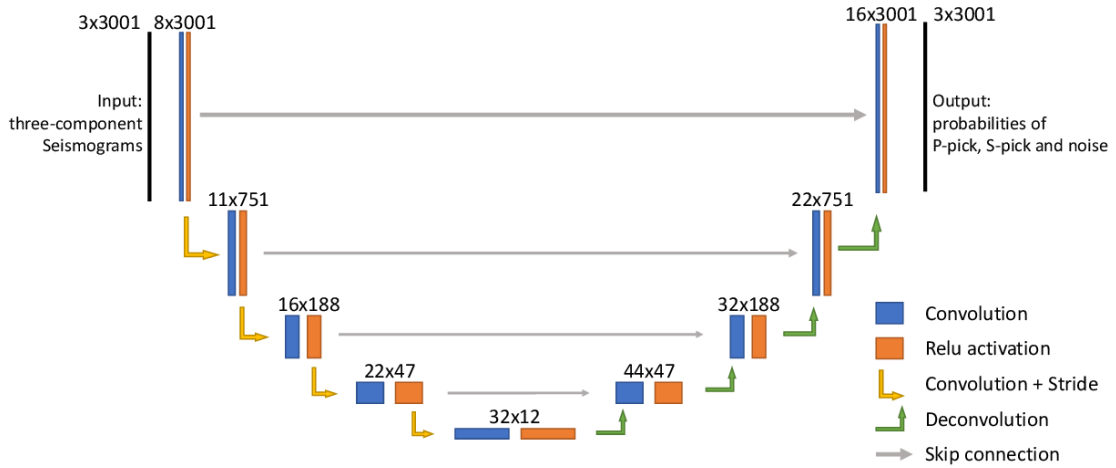


Fig. 2. PhaseNet architecture, taken from Zhu and Beroza (2018).

Note here that the problem has been changed from a detection task to a segmentation task. This means that instead of directly detecting the arrival times of the P and S waves, PhaseNet aims to predict probabilities of such events to occur at each time. A pre-processing is thus needed to create those labels in the training data from the time stamps, which is done simply by generating Gaussians centered at the arrival time and with a fixed standard deviation, chosen to be 10% of the total length of the signal. Conversely, the network output needs to be post-processed, in order to detect arrival times from the segmentation maps (Fig. 3). This is achieved with a simple peak detection algorithm, where several parameters can be used to impose a minimum peak height or a minimum delay between two successive peaks (Duarte and Watanabe, 2021).

3 Data

Before describing the seismic data acquisition setup, we summarize the kind of waves we aim at detecting for the recorded dataset. Indeed, seismic sources like earthquakes generate primary P-waves that are travelling faster than other waves through the earth and are the first waves to arrive and be recorded by the seismic stations located at the surface, closely followed by their reflection from the surface and S-waves arrive and are recorded next. The surface waves like the Rayleigh waves are travelling along the free surface with an elliptic polarization pattern of the wave field. They are slightly slower than the S-waves propagating close to the surface, with a seismic velocity lower than the S-wave velocity by a factor of around 0.9. These Rayleigh waves have a lot of energy compared to the P-waves and can penetrate enough until some depth into the

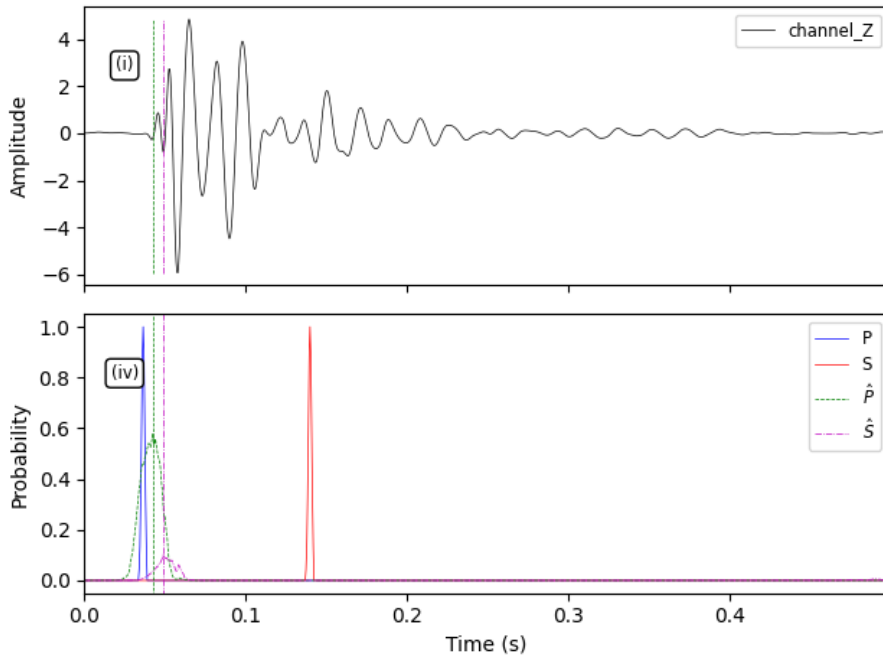


Fig. 3. Top: a seismogram from our real data. Bottom: input segmentation map or probability distribution of P-wave and S-wave provided by analysts (continuous lines) and output segmentation map predicted by pre-trained model in PhaseNet. The vertical lines are the detected peaks from the segmentation maps.

shallow structures to give some information on their S-wave velocities. For more detailed explanations about the physics of Rayleigh waves and without being exhaustive, the reader can be referred to the classical textbook of [Aki and Richards \(1980\)](#) among many other authors.

Now, we describe here a typical near-surface acquisition performed by the real experimental setting done by members of the METIS laboratory at Paris-Sorbonne (Paris, France) to record and detect such seismic wave phases for the characterization of near-surface layers (first 40 meters). In this experiment seismic data were collected in a dry land area located in the Vaucluse region (south of France), and have been originally processed with classical picking techniques ([Dangeard, 2019](#)) inspired by the methodology of [Bauer et al. \(2003\)](#); [Baumann-Wilke et al. \(2012\)](#); [Bauer et al. \(2010\)](#). This is a test case before applying the data processing techniques to more complex media and monitoring water-saturated media, etc. as in [Dangeard et al. \(2018, 2021\)](#). For this purpose, a seismic survey has been deployed along a studied profile of around 120 m length ([Carriere et al., 2016](#)), on which the sources and geophones (receivers) are lined up as described in [Fig. 4](#). A 96-channel seismic recorder with 14 Hz vertical component geophones was used with a 1 m spacing. The source is generated by a 1.25 kg hammer hitting vertically a metal plate. Sources are located between two geophones, and two sources are located before and after the geophone line (the first source is located at 10 m before the first geophone and the last source at 10 m from the last geophone). At each receiver position, seismograms are repeatedly recorded for a source triggered several times at the same location and are stacked in the time-domain to increase signal-to-noise ratio. The sampling rate is 0.5 ms for a 2 s total recording time and a pre-triggering delay of -0.02 s (i.e. the time shift between the recording starting time and the actual beginning of the seismic signal) to include the full surface-wave trains and background noise. Once the source takes place, each receiver records the wave signals into a seismogram for the vertical component which is very sensitive to the P-waves and also to the Rayleigh-waves. Besides, only the time windows involving the dispersive Rayleigh train waves (which are much slower than the first P-wave arrival times) can be recovered at receivers located far from the sources (around more than 40 receivers away from the sources in this experiment). These time pickings can only give a qualitative estimate of the time windows of those dispersive waves, help extracting them from some seismic traces and possibly invert their related dispersion curves (essentially the fundamental mode as commonly done in the community) in the frequency domain. However, here, it is very interesting to see how powerful our automatic picking method can be to retrieve the signals even with only one vertical component, and more specifically the first P-wave travel times. In this experiment, we are considering 96 sources and 96 receivers, so the data consist of 9,216 ($= 96 \times 96$) seismograms recorded by a single (vertical) component.

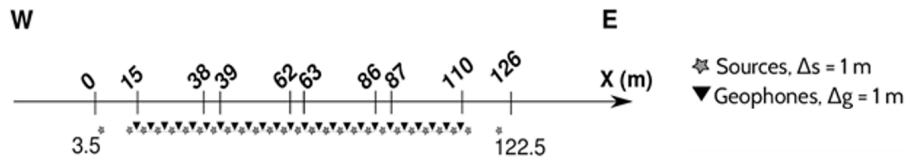


Fig. 4. Measuring experiment method.

4 Methodology

We present here how we adapted PhaseNet to process our data presented above. This data-set of a near-surface active seismic experiment is indeed very different from the data used to train PhaseNet in [Zhu and Beroza \(2018\)](#). The characteristics of both datasets are shown in [Table 2](#), showing in particular that our samples are shorter and contain only the z component.

	NCEDC	Our real data
Type	Passive imaging	Active imaging
Number of samples	700,000	9,216
Labels	Yes	No
Input channels	3 (x, y, z)	1 (z)
Input length	9,000 time samples	4,000 time samples
Sampling rate	100 Hz	2000 Hz

Table 2. Differences between the dataset in PhaseNet (NCEDC) and our real dataset

This section explains how we managed to adapt the PhaseNet model to our real dataset, where various transfer learning settings are tested on different simulated data ([4.1-4.3](#)) and then the real application for near surface data is detailed in [4.4](#).

4.1 Generating simulated datasets

SPECFEM2D ([Komatitsch et al., 2012](#)) is a well-known and robust computational software for 2D and 2.5D (i.e. axisymmetric) simulations of acoustic, elastic, viscoelastic, and poroelastic seismic wave propagation in heterogeneous media as well as for data gradient computations used in FWI or adjoint tomography applications ([Tromp et al., 2008](#)). SPECFEM2D is based on the spectral-element method (SEM) ([Komatitsch, 1997](#); [Komatitsch and Vilotte, 1998](#); [Komatitsch and Tromp, 1999](#); [Peter et al., 2011](#)). It performs a parallel programming based upon the Message Passing Interface and also includes support for GPU graphics card acceleration, that permits to remarkably speed up the simulation process.

We aim to simulate the seismograms with 2 channels (2 components in x /horizontal and z /vertical) that are similar enough in terms of complexity to our real data, and we can thus create a training dataset for our training model. Some specific model parameters can be adjusted in order to achieve a diverse dataset for our database. For instance, we can simulate the waveforms and seismograms with a range of natural states of the material depending on their densities, their absorbing boundary elements, the number of layers and the thickness between them, the seismic velocities V_P and V_S associated to each layer as well as the nature of each layer (acoustic or elastic), etc.. Generally a V_P/V_S ratio between $\sqrt{3}$ and 2.5 is taken depending of the level of partial fluid saturation or unconsolidated granular compaction of the solid medium. We perform two series of synthetic simulations to build a learning database: one with a flat topography and the other with a realistic topography. We thus first realize simple simulations by introducing flat interfaces and a flat free surface that defines the topography, typically as in [Fig. 5](#).

The input parameters must respect a dispersion relation. This relation describes the effect of dispersion on the properties of waves in a medium. To avoid too much dispersion in the simulated signals, a minimum number of grid points per minimum wavelength must be defined: here around 5 discretization mesh points per minimum wavelength are needed in SPECFEM2D to describe and sample accurately the wavelengths

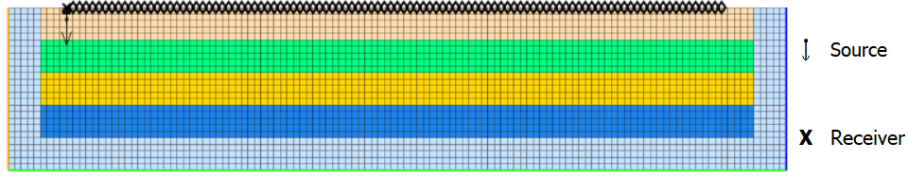


Fig. 5. Simple interface model consisting of 5 layers and 6 flat interfaces with a length of 120 m and a depth of 25 m. We have 101 receivers in total and 1 source (located at the free surface). Every two consecutive receivers are separated by 1m. The dominant source frequency is 80 Hz and the source type is a Gaussian wavelet in time. The total number of time steps is 24,000 and the time step of recording is 0.000025 s. The element size is around 1 m. The P and S seismic velocities vary from 900 up to 4,000 m/s and from 500 up to 2,300 m/s respectively.

of the seismic signal. The parameters must also respect the Courant–Friedrichs–Lewy numerical stability condition (in SPECFEM2D, usually we take $C_{CFL} \leq 0.68$) while solving the elastodynamic partial differential equations numerically. Absorbing PML boundary conditions optimized at grazing incidence (Martin et al., 2008) are introduced to attenuate spurious reflected waves that could come back into the inner computational domain from the outer boundaries. This allows to mimic a semi-infinite medium.

Now by adding a topography to the simulation, we can increase the complexity of the basement model. Based on the arrival times obtained by hand-picking methods, we will thus be able in a near future to solve the seismic inversion problem that involves the forward and adjoint problems solved for instance through ray tracing methods (Lecomte et al., 2015; Podvin and Lecomte, 1991; Qian et al., 2007; Fomel et al., 2009; Huang et al., 2019; Billette and Lambare, 1998). In order to build the synthetic database we build a complex velocity model that mimics the typical kinds of models encountered in the real field with seismic velocity variations increasing with depth from a few hundred meters per second close to the topography up to 3,900 m/s at 30 m depth. The variations in depth are taken into account by introducing 28 layers and 29 interfaces which are enough in our 30 m deep case. In this case, Fig. 6 shows the model with the real topography of the field experiment with the 1 m spacing between geophones of Fig. 4.

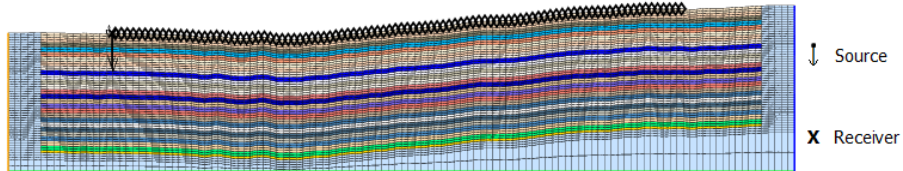


Fig. 6. Interface model consisting of 28 layers and 29 interfaces with a length of 126 m and a depth of 23.5 to 31 m. We have 96 receivers in total and 1 source located at the free surface. Every two consecutive receivers are separated by 1m. The dominant source frequency is 80 Hz and the source type is Gaussian. The total number of time steps is 60,000 and the time step of recording is 0.0000125 s. The element size is around 1 m. The P and S seismic velocities are from 640 up to 3,900 m/s and from 370 up to 2,200 m/s respectively.

4.2 Data augmentation

To increase the size of our dataset, we apply three different transforms to our simulated database: time shifting, waveform sign reversal and noise. Firstly, for each trace (a sample of the training set), we create for instance some additional time-shifted versions with random translation parameters, so that a created trace consists of the arrival times of both P and S-wave with a probability of 0.95. A version excluding these arrival times can also be created with a probability of 0.05, that permits the network learn to detect noise from the original trace. We then duplicate the samples by reversing the waveform sign through a multiplication of the signals by -1. Ultimately, we add white Gaussian noise (AWGN) to all the samples with signal-to-noise ratio (SNR) randomly chosen between 15 and 30 dB. Recall that, if we denote by S and N the signal and the

noise, respectively, the SNR is defined as:

$$\text{SNR}(S, N) = 10 \log \left(\frac{\|S\|_2^2}{\|N\|_2^2} \right)$$

An illustration given in Fig. 7 shows a simulated trace and its augmented versions.

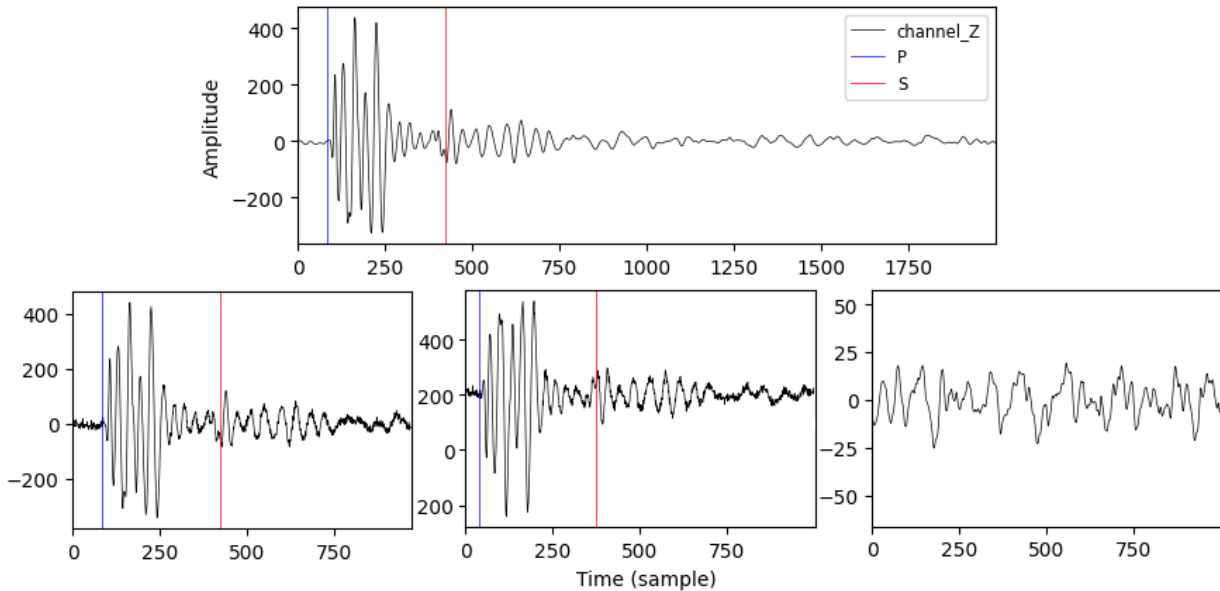


Fig. 7. Top: a simulated trace. Bottom, from left to right: noisy sample, noisy and reversed and shifted sample, pure noise (not containing P-wave nor S-wave).

4.3 Transfer Learning

To maximize the performance of the network, we aim to use the features already learnt by PhaseNet when trained on the huge *Northern California Earthquake Data Center Catalog* (NCEDC 2014) (Zhu and Beroza, 2018). Yet, this dataset is quite different from ours, in terms of scale, frequencies, experimental conditions, sensors, etc.. We thus propose to use part of the original PhaseNet features while retraining part of the network on our simulated data, a process referred to as *transfer learning*. Such a methodology has already been used in Chai et al. (2020), where the pre-trained PhaseNet is successfully fine-tuned with only 3,500 seismograms, which represent 0.45% of the data used to train the original PhaseNet model.

To this end, we started by rewriting the code of PhaseNet from Tensorflow to Keras. Tensorflow is a Python-based open-source library that particularly showcases its efficiency when working with large datasets requiring an excellent functionality and a high performance. Keras on the other hand is a high-level neural network library that runs on top of Tensorflow. Keras is more user-friendly and more convenient to implement transfer learning strategies, since the user can easily initialize and fine-tune each layer separately.

Let us remark that contrary to classification networks, whose final layers can be easily fine-tuned, transfer learning with U-Nets is more challenging. The main reason is that the corresponding layers from the down-sampling and up-sampling stage need to be consistent, thus it is not possible to only fine-tune the last layers. We investigated different transfer learning strategies that could work well given this setting, by considering different settings for both the initialization and the fine-tuning. For each layer, we thus decide i. whether its weights should be initialized randomly or with the pre-trained weights from PhaseNet, and ii. if we allow those weights to change during the training or if we *freeze* them. The model trained from scratch (scenario 1 in Fig. 8) will be compared to different transfer learning settings. For example, in scenario 2, we load all weights from the pre-trained model into our new model and fine-tune all layers. In scenario 3, we do the same thing but the weights of the two deepest layers are frozen. Namely, the weights in frozen layers are not updated during the training. In the next setting, we load weights only from downsampling and upsampling

layers (i.e. the weights of input and output layers are randomly initialized) and freeze weights of the two deepest layers. These three strategies were those that provided the best results. Other strategies were also performed as those depicted in cases 5 and 6.

	Input layer	Down conv layer 1	Down conv layer 2	Down conv layer 3	Down conv layer 4	Up conv layer 4	Up conv layer 3	Up conv layer 2	Up conv layer 1	Output layer
Case 1	Random									
Case 2	Pre-trained									
	Fine-tune									
Case 3	Pre-trained									
	Fine-tune				Freeze		Fine-tune			
Case 4	Random	Pre-trained						Random		
		Fine-tune		Freeze		Fine-tune				
Case 5	Pre-trained					Random				
	Freeze									
Case 6	Pre-trained									
	Fine-tune	Freeze			Fine-tune	Freeze			Fine-tune	

Fig. 8. Different transfer learning settings. Two criteria are defined for each case: i. weight initialization, that is either random (yellow) or pre-trained (pink), ii. if the weights are initialized from a pre-trained model, then they are either fine-tuned (green) or frozen (blue).

4.4 Semi-supervised learning for real near-surface seismograms

In near-surface active imaging, the seismograms are measured by an array of sensors, so that each experiment generates thousands of seismograms. Manually picking the entire dataset would require a huge effort. Instead, we propose here a semi-supervised strategy, where the CNN time picking previously described can be applied with only a few known labels. The main idea is to exploit the strong correlation between the seismograms from close receivers, through robust regression and outlier detection.

The first step is to label for a small dataset, that can be manually done by hand picking, combining with the predictions from the model trained on the simulated dataset with topography (especially for S-wave picking, that could be hardly achieved only with hand picking method). For instance, this labeling process can be performed with 2 leftmost sources and 2 rightmost sources, that are enough to scan all the lateral heterogeneities inside the whole physical domain between the center of the domain and its endings. Subsequently, we train a model on this dataset using transfer learning with the pre-trained model in PhaseNet. In the next step, we perform a pseudo-labeling process, that aims at building up iteratively pseudo-labels in order to retrain the model on a dataset combining the real- and pseudo-labels. In each iteration, we use the model trained on the previous labeled dataset to generate pseudo-labels for an unlabeled dataset of equal size. Then we make an effort to improve the quality of pseudo-labels using robust regression methods. Finally, the size of labeled dataset is doubled until we obtain all labels for the entire real dataset (Algorithm 1). Table 3 illustrates the computational efficiency of this algorithm, applied to our real context. The whole process takes only 2 hours and 32 minutes, handled by the GPU, which is a remarkable improvement compared to hand-picking methods.

In our semi-supervised procedure, enhancing the quality of pseudo-labels is crucial to achieve a model with accurate predictions. For this, we successively use parametric and non-parametric robust regression techniques, to first get rid of the main outliers, and correct them with the least possible assumptions thanks to the non-parametric formulation. Physically, the larger the distance between the source and the receiver, the larger the arrival times, which also helps to individualize the wave arrivals (P waves and more particularly

Algorithm 1 Proposed semi-supervised algorithm

Let $(S_i)_{i=1..N}$ be N unlabeled sets of seismograms.

For $j = 1..[\frac{N}{2}]$, denote: $D_j = \left(\bigcup_{l=1..j} S_l\right) \cup \left(\bigcup_{r=N-j+1..N} S_r\right)$ be the set of j -leftmost and j -rightmost sources. We assume here that $D_j = \bigcup_{i=1..N} S_i$ if $j > [\frac{N}{2}]$.

Denote M_0 be the pre-trained model with the NCEDC data.

Step 1. Given a number $n \ll N$, label for small dataset D_n .

Step 2. Train model, build up pseudo-labels and retrain model:

$k \leftarrow 1$

while $k < \log_2([\frac{N}{2}])$ **do**

Train model M_k on D_{2^k} using data augmentation and transfer learning with pre-trained model M_{k-1}

Generate pseudo-labels for $D_{2^{k+1}} \setminus D_n$ using model M_k

Correct these pseudo-labels using robust regression methods

$k \leftarrow k + 1$

end while

Step 3. Train a final model on $D_{[\frac{N}{2}]}$ and generate labels for $D_{[\frac{N}{2}]} \setminus D_n$ using this model.

		Dataset	Traces	Epochs	Execution time (s)
Iteration 1	P1	D_2	$96 \times 4 \times 3$	30	69
	P2	$D_4 \setminus D_2$	96×4		2.61
	P3				0.07
Iteration 2	P1	D_4	$96 \times 8 \times 3$	40	173
	P2	$D_8 \setminus D_2$	96×12		3.07
	P3				0.19
Iteration 3	P1	D_8	$96 \times 16 \times 3$	50	416
	P2	$D_{16} \setminus D_2$	96×28		3.39
	P3				0.49
Iteration 4	P1	D_{16}	$96 \times 32 \times 3$	60	980
	P2	$D_{32} \setminus D_2$	96×60		4.36
	P3				1.15
Iteration 5	P1	D_{32}	$96 \times 64 \times 3$	80	2,549
	P2	$D_{48} \setminus D_2$	96×92		5.35
	P3				1.78
Final training	P1	D_{48}	$96 \times 96 \times 3$	100	4,939
	P2	$D_{48} \setminus D_2$	96×92		5.35
Total computation time: 2 hrs 32 min 33 sec					

Table 3. Computation time when performing Algorithm 1 applied to our real data, where $N = 96$ and $n = 2$, processed by the NVIDIA GeForce RTX 3080 Graphics card. Each iteration is conducted in three phases: training (P1), predicting or generating pseudo-labels (P2) and correcting pseudo-labels (P3). Note that the number of traces are multiplied by 3 in P1 due to the data augmentation process.

the S and Rayleigh waves). For each source, let us consider a set of seismograms recorded by all the receivers. As we can see in Fig. 9, some points (represented by red dots) are not coherent with the main trend of all these arrivals and we aim to detect them and correct them.

Robust regression methods (e.g. Huber regression (Huber, 1964), Theil-Sen regression (Theil, 1950; Sen, 1968), RANSAC regression (Yoo et al., 2019)) can be chosen to make such classification. They provide an alternative to least squares regression by requiring less restrictive assumptions. These methods attempt to dampen outlier influences in the model to provide a better fit to the majority of the data. For example, Huber regression (Huber, 1964) uses a particular loss function that penalizes the points having a large residual. To detect the outliers of pseudo-labels, we employ logarithmic regression model for the first P-wave arrival time and linear regression model for the S-wave (the regression models are selected depending on the different trends of the arrival times of the two waves related to the receivers).

Although we observed that the data almost take the shape of a line (or a curve), we have no reason to assume that the linear regression model (or logarithmic regression model) is the best fit to the data. But at least, robust regression methods are helping us detect the anomalies that are unusually far from other

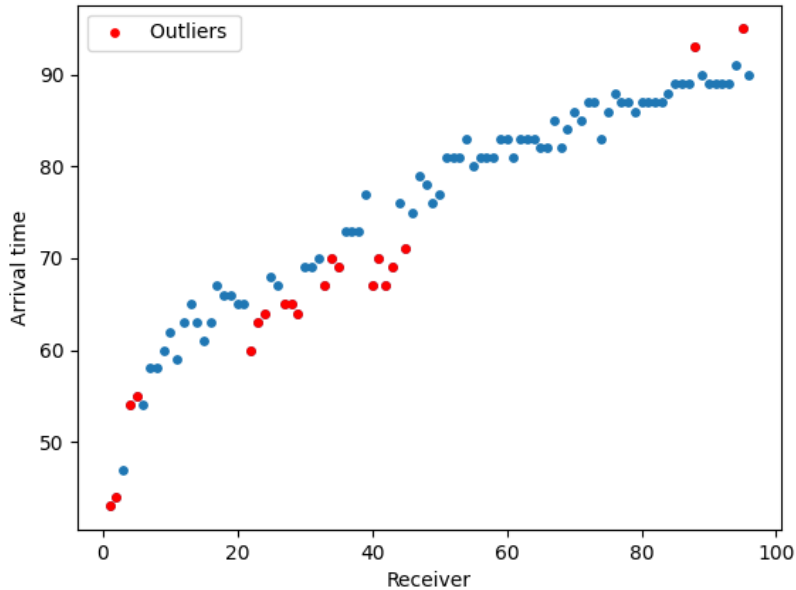


Fig. 9. Outlier detection of P-wave pseudo-labels (time arrivals normalized by the 0.5 ms sampling time step) for the leftmost located source (i.e. source 1) using Huber logarithmic regression, where $\epsilon = 1.55$ is a threshold controlling the number of samples that should be classified as outliers.

observations. We thus take here the Huber regression method which gives better results than classical linear regression methods. Next, we apply SVR to the data with standard observations. The SVR is an application of support vector machine (SVM) in regression and was first developed in [Drucker et al. \(1997\)](#). While many regression algorithms attempt to minimize the l^2 -norm errors, that aims at reducing the number of features used in the final model, SVR tries to minimize the l^2 -norm of the coefficient vector (not the squared error) within a constraint on the error term. This constraint is handled with a boundary line based on a maximum error and support vectors to find an appropriate line or hyperplane (in higher dimensions) to fit the data.

5 Results and Discussion

In this section, we will discuss the performance of the network (with and without using transfer learning) on two datasets. The first data set is the one simulated without adding realistic topography, that is divided into a training-validation set and a test set with 3,890 (88%) and 480 (12%) samples respectively (after implementing data augmentation). The second one is the dataset simulated by adding realistic topography that is divided into a training-validation set and test set with 2,880 (90%) and 288 (10%) samples respectively (also after implementing data augmentation). Then we will apply our semi-supervised framework to the real dataset.

Let us consider the model trained on the synthetic data without realistic topography. [Fig. 10](#) displays the training loss for the 6 different transfer learning scenarios described in [Fig. 8](#). The results indicate that all transfer learning strategies (corresponding to cases 2 to 6) converge faster than the training from scratch (case 1). The best curve is the one of case 3 while case 2 seems to converge to a bad local minimum. [Fig. 11](#) showcases the training and validation loss in these different cases. We don't nearly have the problem of over-fitting or under-fitting in cases 1, 2 and 3 while the three remaining cases may run into the problem of over-fitting in which the model memorizes the training samples and does not generalize well to samples belonging to the test set. To avoid this phenomena, we can, for instance, employ some regularization parameters in order to simplify the complexity of the model. [Table 4](#) shows the scoring metrics in different transfer learning strategies. We can interpret that both settings 1 and 3 give the best scores among transfer learning strategies (best F1-score for P-phase detection in case 1 and best F1-score for S-phase detection in case 3). We thus recommend the third configuration as the optimal setting with adequate scores (particularly for precision scores), fastest convergence rate and smooth training/validation loss curves. In this case, the two deepest layers of the neural network have demonstrated the ability to learn crucial hidden features from the original

data, which are relevant for the new data. By freezing these two layers, the network only needs to be trained on the features of the surface layers for optimal adaptation to the new data. However, the results for P-phase are not as favorable as for S-phase, likely due to the weaker amplitudes of P-waves in our simulations, making it harder for the network to detect the arrival times of P-waves.

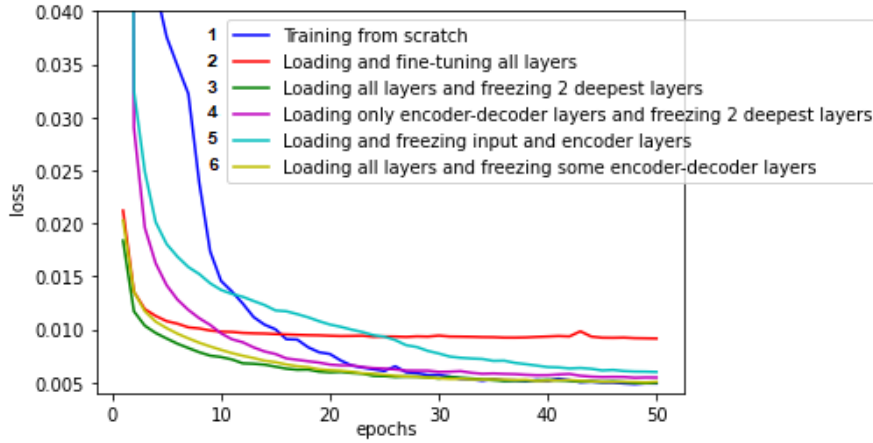


Fig. 10. Comparison of training loss for different transfer learning strategies without adding realistic topography to the simulated dataset.

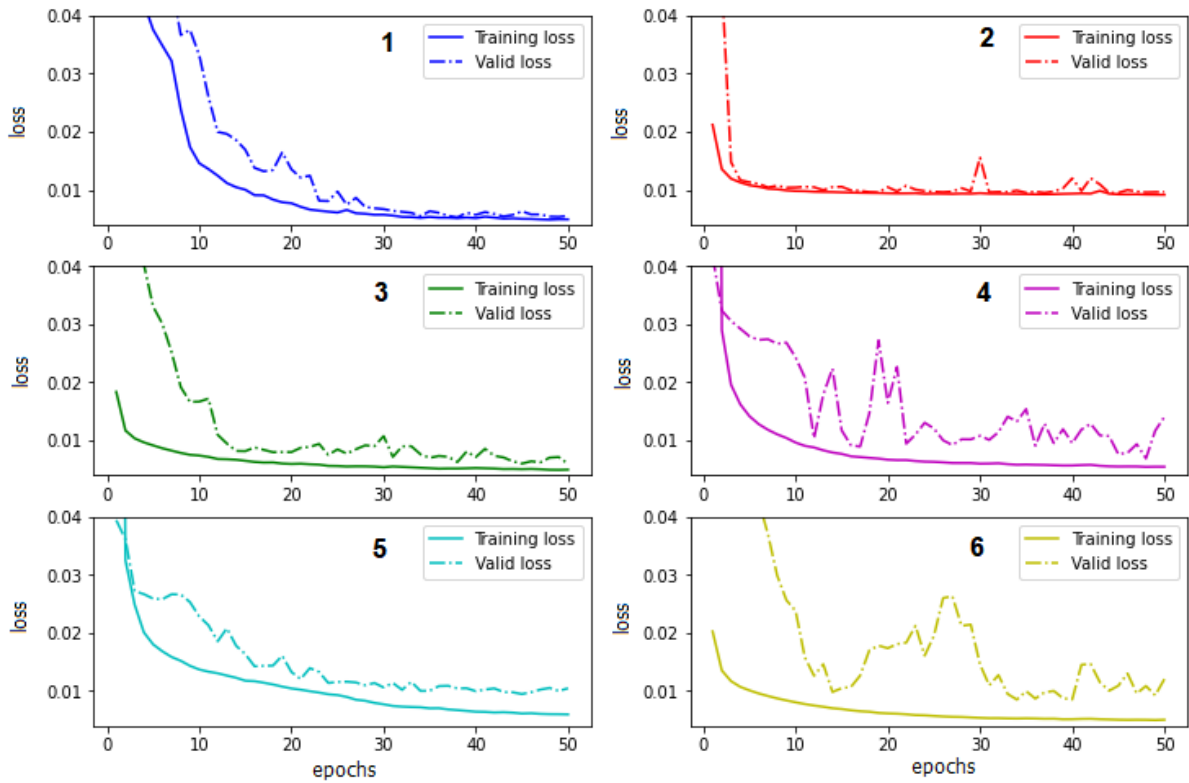


Fig. 11. Training and validation loss for different transfer learning strategies without adding realistic topography to the simulated dataset.

	Phase	Precision	Recall	F1-score
Case 1	P	0.694	0.612	0.650
	S	0.965	0.98	0.973
Case 2	P	0.595	0.535	0.564
	S	0.983	0.987	0.985
Case 3	P	0.636	0.548	0.589
	S	0.983	0.999	0.991
Case 4	P	0.562	0.518	0.539
	S	0.967	0.963	0.965
Case 5	P	0.540	0.414	0.469
	S	0.980	0.985	0.982
Case 6	P	0.630	0.547	0.586
	S	0.998	0.980	0.989

Table 4. Scoring metrics for different transfer learning strategies without adding realistic topography to the simulated dataset. The bold values indicate the highest F1-scores for the detection of both the first P- and S-waves time arrivals.

Now, by training and testing on the synthetic data with realistic topography, we can see the impressive scores (greater than 0.954) for both picking P and S phases in Table 5.

	Phase	Precision	Recall	F1-score
Case 3	P	0.972	0.968	0.970
	S	0.954	0.954	0.954

Table 5. Scoring metrics by adding realistic topography to the simulated dataset.

Next, by using the best model trained on simulated data and the best configuration of transfer learning in our case, we attempt to label for our real data with the methodology mentioned in 4.4. Fig. 12 shows an improvement of pseudo-labels while employing our semi-supervised algorithm (Algorithm 1) at iteration 5 where we observe "smooth" predictions (less outliers) comparing to the predictions at iteration 1 with many outliers. Once all real data are labeled, Fig. 13 shows the seismograms and their arrival times labeled for sources 1, 48 and 96. In general, P-wave detection is much easier than S-wave detection because the first signal of P-wave is more energetic and clearly distinguished from the noise. Furthermore, picking the S-wave is harder when the source and the receiver are too close to each other (this corresponds to near wave field context). The S-wave is thus difficult to detect because it emerges from the noise and is too close to the P-wave. However, the S-wave detection is easier when the source is sufficiently far away from the receiver (far wave field context) but this still remains qualitative due to the different wave conversions and the dispersive nature of the stratified and heterogeneous medium. In this case, the signal of the S-wave (and the highly energetic Rayleigh/surface wave stuck just behind it) can be distinguished from the noise and the first P-wave signal. High frequency surface waves (i.e. Rayleigh waves) are also picked (see Fig. 13) but only at medium or large offset (beyond more than 48 receivers from the sources) because in the near-field they can not be really well individualized. The first arrivals related to the P-waves give more accurate information on medium or larger depths but not on very shallow depths due to the near-field effect close to the sources. Rayleigh-wave labels can be detected only qualitatively (not very accurately) at medium/large offset (clearly visible for sources 1 and 96 in Fig. 13). Those labels have larger values than the ones of P-wave and correspond to smaller velocities and weathered shallow layers. However, all these S and surface (Rayleigh) wave labels can only be estimated qualitatively. Therefore, in this study, we will thus only be able to consider the P-wave first arrivals and estimate them quantitatively.

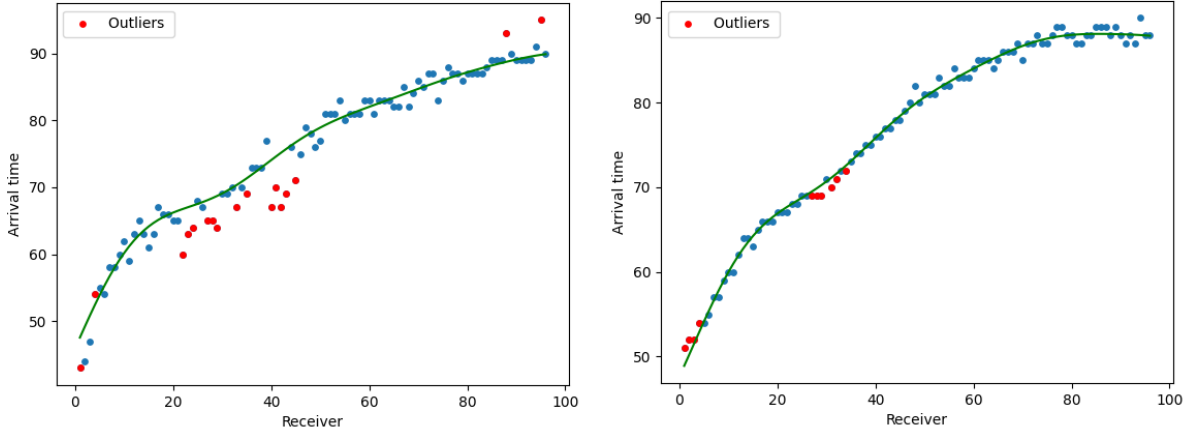


Fig. 12. Correcting P-wave pseudo-labels (time arrivals normalized by the 0.5 ms sampling time step) at iteration 1 (left) and at iteration 5 (right) using SVR, where the maximum error $\epsilon = 0.1$ and the regularization parameter $C = 200$.

In Fig. 14, we represent all the 9126 P-wave picked labels (i.e first time arrivals) in a shot-offset (i.e. source versus receiver) diagram, those picked labels being computed for (i, j) and (j, i) symmetric source-receiver pairs. This travel-time labels distribution for each source-receiver position pair allows to check lateral coherency of the data. The same objects should be seen by the seismic rays coming from opposite ways (from source i to receiver j or source j to receiver i). In the diagonal of the diagram the labels correspond to the zero-offset first arrival times: we have $i = j$ there and both the source and the receiver are located at the same place. Below the diagonal, the labels are mainly influenced by the upstream geological structures and those above the diagonal are influenced by the downstream structures. With such representation we can have preliminary and qualitative estimates of subsurface apparent P-velocities and of course the lateral variations along the seismic line. The apparent P velocities increase with depth. When the source-receiver distance increases, the detected labels are more and more sensitive to the P-velocities of the heterogeneities at depth. More generally, the longer the offset the greater is the depth of probing.

We can also see in Fig. 15 that at first sight the first arrival times of the direct (i -th source/ j -th receiver pair) and reverse shots (j -th source/ i -th receiver pair) can be compared and give some cross-validation information about the influence of significant lateral heterogeneities, time arrivals picking consistency and also another metric of label error estimates. Despite a significant noise level at far offsets, P-waves can be easily identified.

However, the picking uncertainties are estimated in a very constraining way because of the strong variations in data quality observed along the seismic line. As we can see in the source-receiver diagram, some values are not really consistent physically and have poor quality. This can be due to the fact that the data have not been recorded properly due to different instrument responses, poor coupling between sources and geophones, tilt angles of geophones and sources, near-field offset physical effects, changes in waveforms generated by lateral heterogeneities or attenuation of the signal or sensitivity to ambient noise. The data have possibly not been reproduced exactly because the source shape is not the same from a location to another. Even with an automatic source generation, the reproducibility can not be fully guaranteed (in terms of orientation, frequency content, amplitude, etc.), and thus the first P-arrival waveforms neither. These label picking drawbacks can also be due to similar problems for both the sources and geophones in terms of their orientations and acquisition specifications. All these issues are not really discussed in the literature. However, as in other experiments like those of Pasquet et al. (2015a,b); Dangeard et al. (2018); Dangeard (2019), we have very similar configurations of measurement, for a same survey equipment and a same analyst operator on a same site.

All these drawbacks can also be more clearly seen in Fig. 15 where we show the normalized difference between labels of each symmetric source-receiver pairs (i, j) and (j, i) . As can be seen in Fig. 16, we can see clearly that the cumulated relative differences ("errors") between symmetric (i, j) and (j, i) source-receiver pairs can be considered as small for the picked P-labels. Only 2% of the P-labels have relative errors between 8 and 20%. More important, in general these differences for the P-labels (relative errors lower than 6%

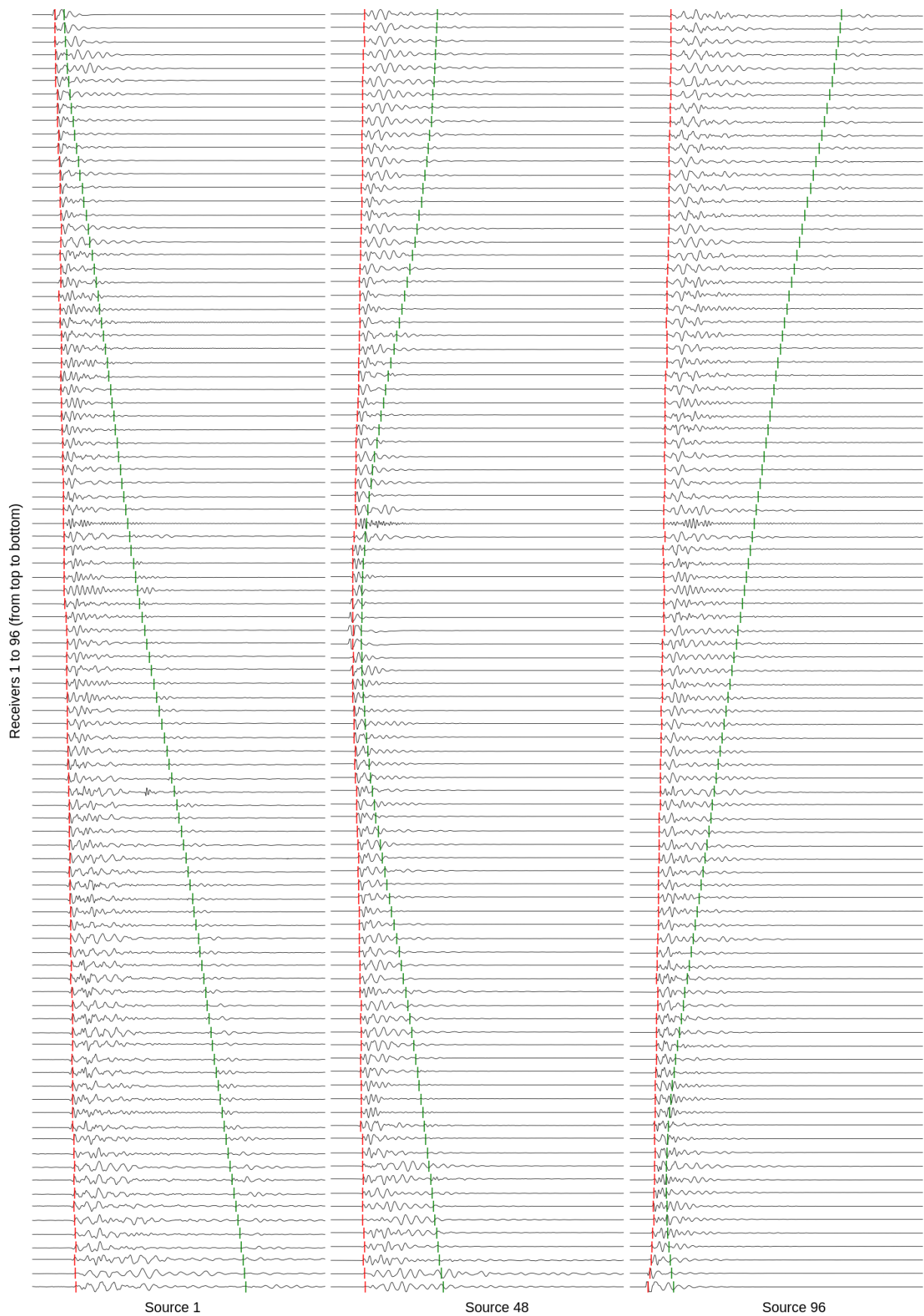


Fig. 13. Seismograms for sources 1 (picked by hand), 48 (labeled using semi-supervised learning) and 96 (picked by hand). The labels of first arrival times are shown (red) and are detected for all seismograms. This is not the case for the surface (Rayleigh) train-waves (green) which are detected very far from the sources because only one component is recorded at all receivers.

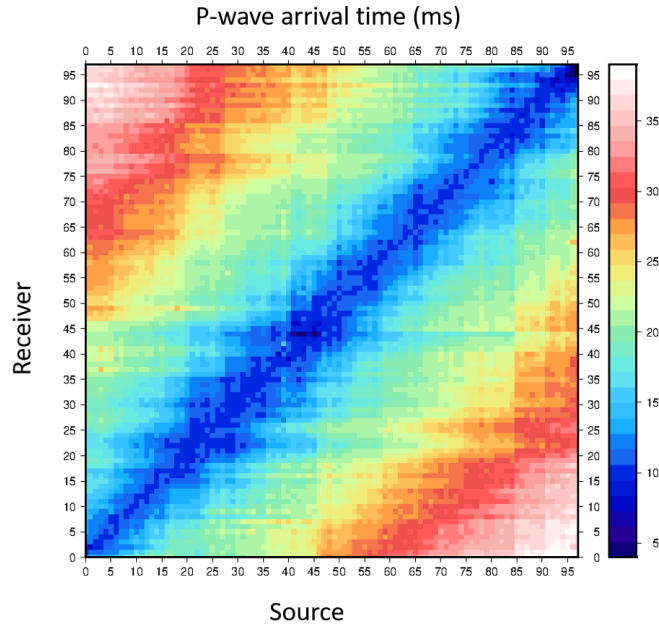


Fig. 14. Source-offset diagram showing the P-wave (first time arrivals) picked labels. For each source-receiver position pair this travel-time labels distribution allows to check lateral coherency of the data. In the diagonal the labels correspond to the zero-offset first arrival times. Below the diagonal, the labels are mainly influenced by the upstream geological structures and those above the diagonal are influenced by the downstream structures. The longer the offset the greater are the depth of probing and the sensitivity to the P-velocities of heterogeneities at depth.

corresponding to maximum variations of 2 to 3 ms time delays) are small for almost 96% of the whole dataset and have almost the same order of time-arrival magnitudes because these pairs are coherently seeing the same medium between them. Those maximum relative errors around 6% are in the same range of relative errors found for other setup configurations (i.e. errors between 5 and 7%), configurations that are generally very similar to those considered here (Pasquet et al., 2015a,b; Bergamo et al., 2016). Similar ranges of relative errors are also retrieved in configurations at laboratory scales for similar near surface studies (Bodet et al., 2010, 2014; Pasquet et al., 2016) (in particular for unconsolidated granular characterization with or without fluidization). Besides, the main differences and strong errors are more important close to the zero-offset source receiver pairs (i, i), at very far offsets and also at some sources (source 44 for instance for the P-wave time arrivals) and also at some receivers (75 to 80, mainly 77 and also 84 for instance). This can be associated to the signal spectra that are not correlated to neighbouring receivers and also to bad contact of sources and stations with the ground. But these inaccurate data still remain very few and can thus be removed from the set of labels before performing any inversion/imaging process without too much lack of accuracy.

Besides, as a matter of information, we can add that the P-labels are generally much better detected with our method with errors generally lower than 6% when compared to the Rayleigh-waves (label errors not shown here). Indeed, this is not really the case for the Rayleigh train-wave labels for which only 60% of the labels show relative errors lower than 8%. These Rayleigh wave time pickings inaccuracy can be due to the fact that the surface waves, which are generally dispersive, can hardly become well enough individualized even at a reasonably far enough distance from the sources (and also maybe because many wave conversions are occurring due to the heterogeneous nature of the medium). To detect them better two or three components should be considered in the whole work flow, but as we already said in a great number of near surface experiments only the vertical component is available. However, it gives us a first qualitative idea of the time window containing the Rayleigh-wave train.

In recent advanced works in near surface geophysics, this time picking process is commonly applied only to the first P-wave time arrivals and is generally done randomly many times (between 15 to 30 times) for each seismogram, a minimum of 15 times being necessary (Dangeard et al., 2018; Dangeard, 2019; Dangeard et al., 2021). A standard deviation (STD) is then estimated to have an idea of the picking errors. Commonly, for seismic setup configurations close to our experiment (in terms of geometry, spatial scale, source frequency content, etc.), first arrival time variations between -0.5 to 0.5 ms or -0.8 ms to 0.8 ms (Dangeard, 2019; Dangeard et al., 2021) are generally considered as insignificant and correspond approximately to P-velocity

error values between 6 and 10 m/s for the kind of surveys configurations we are studying. These values are close to our values ranging around 2 ms mainly, which makes our method competitive in terms of accuracy and speed of the time arrival labels when compared to the more classical methods.

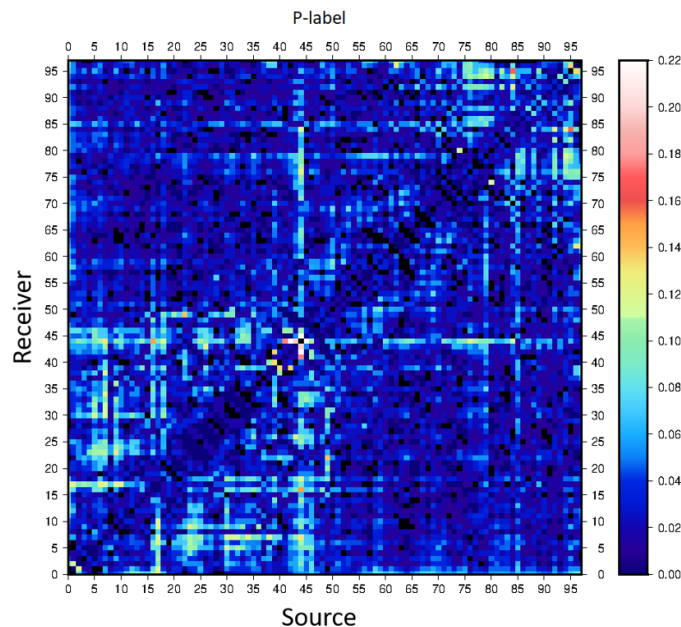


Fig. 15. Source-offset diagrams showing the normalized differences between the P-wave (first time arrivals) picked labels computed for symmetric source-receiver pairs. This metric allows to measure if the symmetric pairs of sources and receivers are able to see the same medium and its lateral variations in a coherent way. In dark blue, the relative errors are very small (generally lower than 0.01). For a few number of sources (like source 44) and receivers (75 to 80, receiver 77 mainly, and also 84) all the corresponding seismic traces are not very well picked due to a bad quality in the data measurements by itself.

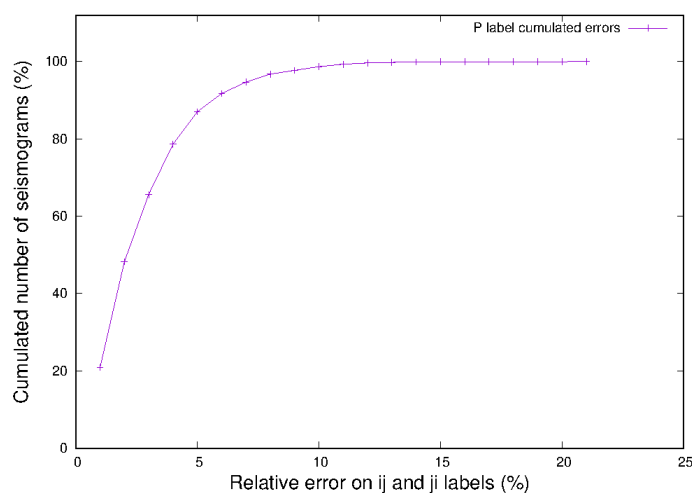


Fig. 16. Cumulated relative errors of the differences of picked labels between (i, j) and (j, i) source-receiver pairs for the P-wave. The relative errors of the picked P-labels are smaller than 6% over almost 96% of the global dataset of seismograms.

In order to estimate the good ability of prediction by our method, we have trained a model on the whole real dataset (with the labels obtained from semi-supervised learning procedure) and have tested it on other datasets obtained by filtering the real datasets via different strategies: 3 sets of 96 seismograms each are built for sources 24, 48 and 72 through a 10-110 Hz band-pass filter (around 10Hz filtering cutoff frequencies being used in similar survey configurations), and 4 sets of 96 seismograms are built through a 300 Hz low-pass filter for sources 24 and 48, and through 100 Hz and 200 Hz low-pass filters for the source 72. For each set, the labels of the first-arrivals (only for P-wave) have been picked manually and are compared to the labels predicted by the model. As can be seen in Fig. 17, good predictions are obtained with errors between 1% and 10% for both 10-110 Hz band-pass and low-pass frequency filters (subfigures 17-a and 17-b respectively). A maximum average prediction error of around 7% is obtained for the different filters and the three different sources except in the particular case of the 10-100 Hz band-pass filtered data for the source 72. In this particular case, a prediction error of 15% in average is obtained because the first time arrivals were not as easy to pick by hand as for the other sources and filters. But this is just a singular case due to human picking error. More generally, in all cases, the worst prediction errors correspond to bad quality measurements at some stations or at offsets close to the sources. Indeed, very close to the sources, picking errors are almost systematically greater than those at medium offsets because the near-field waveforms are generally not well defined: P, S and Rayleigh waves are all mixed and have not the time to develop and be well individualized. For the far-field (i.e. large enough offset distances), the errors can also contain a significant error because a lot of heterogeneities are located in the ray-path of the seismic wave travelling through all the heterogeneous medium at depth between the source and the receiver. In summary, as we said previously, average prediction errors for those data are lower than 7% and are thus consistent to those presented in Figure 16.

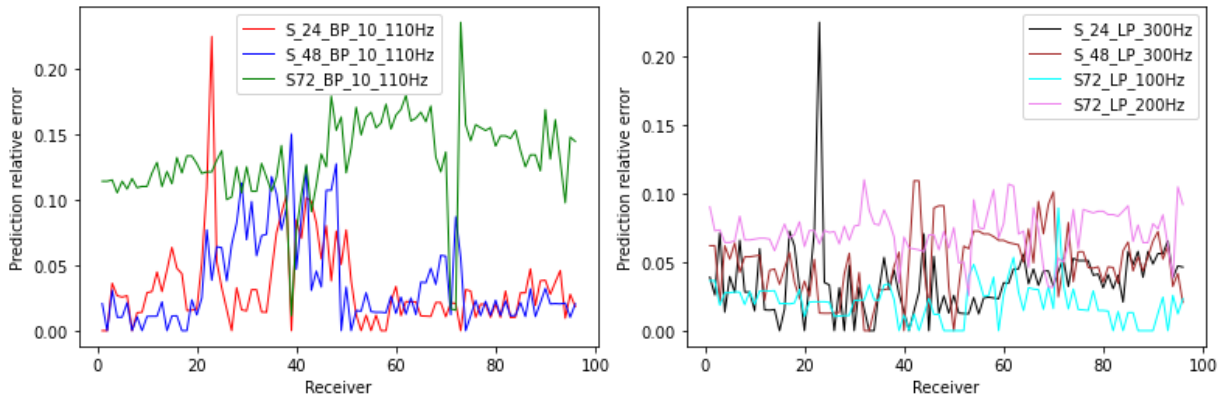


Fig. 17. P-wave label prediction relative errors for different sources at all receivers and for different frequency filters: 10-110 Hz band-pass filter (left), for 100, 200 or 300 Hz low-pass filters (right). The predicted errors are low and acceptable except close to the sources due to near-wave field issues.

6 Conclusion

Deep learning methods for seismic inversion problems are being improved rapidly. An end-to-end deep learning is however unachievable because of the complexity and the lack of labeled datasets. Automatic arrival time picking is one of the less expensive methods in helping resolve the inversion problem. In spite of that, labeling of thousands of seismograms to train a deep neural network is still complicated and labour intensive. We proposed an approach for automatic arrival time picking with unlabeled data based on PhaseNet. By generating a dataset that is close to our real data with SPEC2FEM2D, we tested many transfer learning strategies on the simulated dataset and selected the best one to help the labeling of our real data. The best strategy consisted in loading the pre-existing layers, freezing the 2 deepest layers and fine-tuning all the others. The models trained on the simulated data showcased very high scores for the detection of the first P-wave arrival times and finally helped label for thousands of seismograms of the real data by using semi-supervised learning. Moreover, robust regression methods and SVM are employed in order to detect outliers in the pseudo-labeling and to enhance the quality of pseudo-labels. This can demonstrate the efficiency of

automatic time picking by deep learning in helping to detect time arrivals and to prepare the time arrival datasets to be inverted for tomography applications.

Finally, another advantage of our method is that the semi-supervised algorithm permits to train the model by adding smaller information to the training dataset through relatively small manually picked labels (4 sources and 96 seismograms each) when compared to the more classical detection procedures. Even if recent STD-based detection procedures (Dangeard et al., 2018) can be efficient, they will need random manual picking over 9126 seismograms many times (15 up to 30 times manual pickings for around 6 to 10 sources taken randomly) which is still very time consuming (more than 50 times slower than our method in the best case). Furthermore, all our machine learning database can be enriched little by little through new surveys with new data and by just adding a few more seismograms for a few selected sources. The network will learn more and more and will increase its accuracy in the medium term. We emphasize that our design method could be adapted to new data since it is not time-consuming and does not require a large amount of labeled data. Besides, many acquisition configurations are similar to the one studied here and then similar reasonably low CPU-times will be reached.

As another perspective, future work should still be done to improve for instance time arrival detection related to the different converted waves or surface (Rayleigh) waves detection, which is not trivial to perform due the heterogeneous, stratified and dispersive nature of the media often studied in near-surface geophysics context. Indeed, only qualitative estimates of time windows involving sufficiently high energetic surface waves has been detected here and this can be obtained only far enough from the sources.

Acknowledgements Surface seismic data acquisitions used in this study were supported by the equipex CRITEX equipment granted project N° ANR-11-EQPX-0011 and performed by Marine Dangeard (now at SNCF-Réseau, DGII/DTR/GC/VA/PGRN, La Plaine Saint-Denis, France) and Ludovic Bodet (UMR 7619 METIS, Sorbonne Université, CNRS, EPHE, France) with the help of K. Chalikakis (UMR EMMAH, Avignon Université, France), who we greatly want to acknowledge. This work is part of the MADASSY project funded by the RTRA-STAE via the ENV'IA Network (Toulouse/France).

7 Compliance with Ethical Standards

- Conflict of Interest: All authors declare that they have no conflict of interest.
- Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

References

- Adler A, Araya-Polo M, Poggio T (2021) Deep learning for seismic inverse problems: Toward the acceleration of geophysical analysis workflows. *IEEE Signal Processing Magazine* 38(2):89–119, DOI 10.1109/MSP.2020.3037429
- Akazawa T (2004) A technique for automatic detection of onset time of p-and s-phases in strong motion records. In: Proc. of the 13th World Conf. on Earthquake Engineering, Vancouver, Canada
- Aki K, Richards PG (1980) Quantitative seismology, theory and methods. W. H. Freeman, San Francisco, USA
- Araya-Polo M, Jennings J, Adler A, Dahlke T (2018) Deep-learning tomography. *The Leading Edge* 37(1):58–66
- Baeten G, Maag JAD, Plessix RE, Klaasen R, Qureshi T, Kleemeyer M, ten Kroode APE, Rujie Z (2013) The use of low frequencies in a full-waveform inversion and impedance inversion land seismic case study. *Geophysical Prospecting* 61(4):701–711
- Bauer K, Schulze A, Ryberg T, Sobolev SV, Weber M (2003) Classification of lithology from seismic tomography: A case study from the messum igneous complex, namibia. *Journal of Geophysical Research* 108:2152
- Bauer K, Moeck I, Norden B, Schulze A, Weber M, Wirth H (2010) Tomographic p wave velocity and vertical velocity gradient structure across the geothermal site groß schönebeck (ne german basin): Relationship to lithology, salt tectonics, and thermal regime. *Journal of Geophysical Research: Solid Earth* 115(B8), DOI doi.org/10.1029/2009JB006895
- Baumann-Wilke M, Bauer K, Schovsbo NH, Stiller M (2012) P-wave travelttime tomography for a seismic characterization of black shales at shallow depth on bornholm, denmark. *GEOPHYSICS* 77(5):EN53–EN60, DOI 10.1190/geo2011-0326.1
- Baydin AG, Pearlmutter BA, Radul AA, Siskind JM (2018) Automatic differentiation in machine learning: a survey. *Journal of machine learning research* 18

- Bergamo P, Dashwood B, Uhlemann S, Swift R, Chambers JE, Gunn DA, Donohue S (2016) Time-lapse monitoring of fluid-induced geophysical property variations within an unstable earthwork using p-wave refraction. *GEOPHYSICS* 81(4):EN17–EN27, DOI 10.1190/geo2015-0276.1
- Bianco MJ, Gerstoft P, Olsen KB, Lin FC (2019) High-resolution seismic tomography of long beach, ca using machine learning. *Scientific reports* 9(1):1–11
- Billette F, Lambare G (1998) Velocity macro-model estimation from seismic reflection data by stereotomography. *Geophysical Journal International* 135(2):671–690, DOI 10.1046/j.1365-246X.1998.00632.x
- Bodet L, Jacob X, Tournat V, Mourgues R, Gusev V (2010) Elasticity profile of an unconsolidated granular medium inferred from guided waves: Toward acoustic monitoring of analogue models. *Tectonophysics* 496:99–104
- Bodet L, Dhemaied A, Martin R, Mourgues R, Rejiba F, Tournat V (2014) Small-scale physical modeling of seismic-wave propagation using unconsolidated granular media. *Geophysics* 79(6):T323–T339
- Bording RP, Gersztenkorn A, Lines LR, Scales JA, Treitel S (1987) Applications of seismic travel-time tomography. *Geophysical Journal International* 90(2):285–303, DOI 10.1111/j.1365-246X.1987.tb00728.x
- Cao D, Liao W (2015) A computational method for full waveform inversion of crosswell seismic data using automatic differentiation. *Computer Physics Communications* 188:47–58, DOI doi.org/10.1016/j.cpc.2014.11.002
- Carriere S, Chalikakis K, Danquigny C, Davi H, Mazzilli N, Ollivier C, Emblanch C (2016) The role of porous matrix in water flow regulation within a karst unsaturated zone: an integrated hydrogeophysical approach. *Hydrogeology Journal* 24(7):1905–1918, DOI 10.1007/s10040-016-1425-8
- Chai C, Maceira M, Santos-Villalobos HJ, Venkatakrishnan SV, Schoenball M, Zhu W, Beroza GC, Thurber C, Team EC (2020) Using a deep neural network and transfer learning to bridge scales for seismic phase picking. *Geophysical Research Letters* 47(16):e2020GL088651
- Dai T, Xia J, Ning L, Chaoqiang X, Liu Y, Xing H (2021) Deep learning for extracting dispersion curves. *Surveys in Geophysics* 42:1–27, DOI 10.1007/s10712-020-09615-3
- Dangeard M (2019) Développement d’une approche “ time-lapse ” des méthodes sismiques pour l’hydrogéophysique et la compréhension de la dynamique des hydrosystèmes. Theses, Sorbonne Université
- Dangeard M, Bodet L, Pasquet S, Thiesson J, Guérin R, Jougnot D, Longuevergne L (2018) Estimating picking errors in near-surface seismic data to enable their time-lapse interpretation of hydrosystems. *Near Surface Geophysics* 16(6):613–625, DOI doi.org/10.1002/nsg.12019
- Dangeard M, Rivière A, Bodet L, Schneider S, Guérin R, Jougnot D, Mainault A (2021) River corridor model constrained by time-lapse seismic acquisition. *Water Resources Research* 57(10):e2020WR028911, DOI doi.org/10.1029/2020WR028911
- Drucker H, Burges CJ, Kaufman L, Smola A, Vapnik V, et al. (1997) Support vector regression machines. *Advances in neural information processing systems* 9:155–161
- Duarte M, Watanabe RN (2021) Notes on Scientific Computing for Biomechanics and Motor Control. DOI 10.5281/zenodo.4599319
- Earp S, Curtis A, Zhang X, Hansteen F (2020) Probabilistic neural network tomography across grane field (north sea) from surface wave dispersion data. *Geophysical Journal International* 223(3):1741–1757
- Fichtner A, Bunge Hp, Igel H (2006) The adjoint method in seismology: I. theory. *Physics of the Earth and Planetary Interiors* 157:86–104, DOI 10.1016/j.pepi.2006.03.016
- Fomel S, Luo S, Zhao H (2009) Fast sweeping method for the factored eikonal equation. *Journal of Computational Physics* 228(17):6440–6455, DOI doi.org/10.1016/j.jcp.2009.05.029
- Hobro JWD, Singh SC, Minshull TA (2003) Three-dimensional tomographic inversion of combined reflection and refraction seismic traveltimes data. *Geophysical Journal International* 152(1):79–93, DOI 10.1046/j.1365-246X.2003.01822.x
- Hole JA (1992) Nonlinear high-resolution three-dimensional seismic travel time tomography. *Journal of Geophysical Research: Solid Earth* 97(B5):6553–6562, DOI doi.org/10.1029/92JB00235
- Huang G, Luo S, Ari T, Li H, Nobes DC (2019) First-arrival tomography with fast sweeping method solving the factored eikonal equation. *Exploration Geophysics* 50(2):144–158, DOI 10.1080/08123985.2019.1577110
- Huber PJ (1964) Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics* 35(1):73–101, DOI 10.1214/aoms/1177703732
- Jones IF (2010) Tutorial: Velocity estimation via ray-based tomography. *first break* 28(2)
- Komatitsch D (1997) Méthodes spectrales et éléments spectraux pour l’équation de l’élastodynamique 2D et 3D en milieu hétérogène (Spectral and spectral-element methods for the 2D and 3D elastodynamics equations in heterogeneous media). PhD thesis, Institut de Physique du Globe, Paris, France, 187 pages
- Komatitsch D, Tromp J (1999) Introduction to the spectral-element method for 3-D seismic wave propagation 139(3):806–822, DOI 10.1046/j.1365-246x.1999.00967.x

- Komatitsch D, Vilotte JP (1998) The spectral-element method: an efficient tool to simulate the seismic response of 2D and 3D geological structures 88(2):368–392
- Komatitsch D, Vilotte JP, Cristini P, Labarta J, Le Goff N, Le Loher P, Liu Q, Martin R, Matzen R, Morency C, Peter D, Tape C, Tromp J, Xie Z (2012) Specfem2d v7.0.0 [software]
- Kong Q, Trugman DT, Ross ZE, Bianco MJ, Meade BJ, Gerstoft P (2019) Machine learning in seismology: Turning data into insights. *Seismological Research Letters* 90(1):3–14
- Kosloff D, Sherwood J, Koren Z, Machet E, Falkovitz Y (1996) Velocity and interface depth determination by tomography of depth migrated gathers. *GEOPHYSICS* 61(5):1511–1523, DOI 10.1190/1.1444076
- Lecomte I, Lubrano-Lavadera P, Anell I, Buckley S, Schmid DW, Heeremans M (2015) Ray-based seismic modeling of geologic models: Understanding and analyzing seismic images efficiently. *Interpretation* 3:SAC71–SAC89, DOI 10.1190/INT-2015-0061.1
- Li S, Liu B, Ren Y, Chen Y, Yang S, Wang Y, Jiang P (2020) Deep-learning inversion of seismic data. *IEEE Transactions on Geoscience and Remote Sensing* 58(3):2135–2149, DOI 10.1109/TGRS.2019.2953473
- Liu Q, Tromp J (2006) Finite-frequency kernels based on adjoint methods 96(6):2383–2397, DOI 10.1785/0120060041
- Martin R, Komatitsch D, Gedney SD (2008) A variational formulation of a stabilized unsplit convolutional perfectly matched layer for the isotropic or anisotropic seismic wave equation. *Comput Model Eng Sci* 37(3):274–304
- Mousavi SM, Ellsworth WL, Zhu W, Chuang LY, Beroza GC (2020) Earthquake transformer: an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature communications* 11(1):1–12
- Pasquet S, Bodet L, Dhemaied A, Mouhri A, Vitale Q, Rejiba F, Flipo N, Guérin R (2015a) Detecting different water table levels in a shallow aquifer with combined p-, surface and sh-wave surveys: Insights from vp/vs or poisson’s ratios. *Journal of Applied Geophysics* 113:38–50, DOI doi.org/10.1016/j.jappgeo.2014.12.005
- Pasquet S, Bodet L, Longuevergne L, Dhemaied A, Camerlynck C, Rejiba F, Guérin R (2015b) 2d characterization of near-surface : surface-wave dispersion inversion versus refraction tomography. *Near Surface Geophysics* 13(4):315–332, DOI doi.org/10.3997/1873-0604.2015028
- Pasquet S, Bodet L, Bergamo P, Guérin R, Martin R, Mourgues R, Tournat V (2016) Small-scale seismic monitoring of varying water levels in granular media. *Vadose Zone Journal* 15(7):vzj2015.11.0142, DOI doi.org/10.2136/vzj2015.11.0142
- Peter D, Komatitsch D, Luo Y, Martin R, Le Goff N, Casarotti E, Le Loher P, Magnoni F, Liu Q, Blitz C, Nissen-Meyer T, Basini P, Tromp J (2011) Forward and adjoint simulations of seismic wave propagation on fully unstructured hexahedral meshes 186(2):721–739, DOI 10.1111/j.1365-246X.2011.05044.x
- Plessix RE (2006) A review of the adjoint-state method for computing the gradient of a functional with geophysical applications 167(2):495–503
- Podvin P, Lecomte I (1991) Finite difference computation of traveltimes in very contrasted velocity models: a massively parallel approach and its associated tools. *Geophysical Journal International* 105(1):271–284, DOI 10.1111/j.1365-246X.1991.tb03461.x
- Qian J, Zhang YT, Zhao HK (2007) Fast sweeping methods for eikonal equations on triangular meshes. *SIAM J Numerical Analysis* 45:83–107, DOI 10.1137/050627083
- Rawlinson N, Sambridge M, et al. (2003) Seismic traveltime tomography of the crust and lithosphere. *Advances in geophysics* 46:81–199
- Richardson A (2018) Seismic full-waveform inversion using deep learning tools and techniques. arXiv preprint arXiv:180107232
- Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*, Springer, pp 234–241
- Sen PK (1968) Estimates of the regression coefficient based on kendall’s tau. *Journal of the American statistical association* 63(324):1379–1389
- Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556
- Talwani M, Kessinger W (2003) Exploration geophysics. In: Meyers RA (ed) *Encyclopedia of Physical Science and Technology* (Third Edition), third edition edn, Academic Press, New York, pp 709–726
- Tarantola A (1984) Inversion of seismic reflection data in the acoustic approximation 49:1259–1266
- Tarantola A (1987) *Inverse problem theory: methods for data fitting and model parameter estimation*. Elsevier Science Publishers, Amsterdam, Netherlands
- Tarantola A (1988) Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation 128:365–399
- Tarantola A, Valette B (1982) Generalized nonlinear inverse problems solved using the least squares criterion. *Rev Geophys Space Phys* 20:219–232

- Theil H (1950) A rank-invariant method of linear and polynomial regression analysis. *Indagationes mathematicae* 12(85):173
- Tromp J, Tape C, Liu Q (2005) Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels 160(1):195–216, DOI 10.1111/j.1365-246X.2004.02453.x
- Tromp J, Komatitsch D, Liu Q (2008) Spectral-element and adjoint methods in seismology. *Communications in Computational Physics* 3(1):1–32
- Virieux J, Operto S (2009) An overview of full-waveform inversion in exploration geophysics 74(9):WCC127–WCC152
- Virieux J, Asnaashari A, Brossier R, Métivier L, Ribodetti A, Zhou W (2017) An introduction to full waveform inversion. In: *Encyclopedia of exploration geophysics*, Society of Exploration Geophysicists, pp R1–R40
- Wang J, Xiao Z, Liu C, Zhao D, Yao Z (2019) Deep learning for picking seismic arrival times. *Journal of Geophysical Research: Solid Earth* 124(7):6612–6624, DOI doi.org/10.1029/2019JB017536
- Xu S, Wang D, Chen F, Zhang Y, Lambare G (2012) Full waveform inversion for reflected seismic data. In: 74th EAGE Conference and Exhibition incorporating EUROPEC 2012, European Association of Geoscientists & Engineers, pp cp–293
- Yang F, Ma J (2019) Deep-learning inversion: A next-generation seismic velocity model building method. *GEOPHYSICS* 84(4):R583–R599, DOI 10.1190/geo2018-0249.1
- Yoo J, Borselen R, Mubarak M, Tsingas C (2019) Automated first break picking method using a random sample consensus (ransac). In: 81st EAGE Conference and Exhibition 2019, European Association of Geoscientists & Engineers, vol 2019, pp 1–5
- Yu S, Ma J (2021) Deep learning for geophysics: Current and future trends. *Reviews of Geophysics* 59(3):e2021RG000742
- Zelt CA, Barton PJ (1998) Three-dimensional seismic refraction tomography: A comparison of two methods applied to data from the faeroe basin. *Journal of Geophysical Research: Solid Earth* 103(B4):7187–7210
- Zheng Y, Zhang Q, Yusifov A, Shi Y (2019) Applications of supervised deep learning for seismic interpretation and inversion. *The Leading Edge* 38(7):526–533, DOI 10.1190/tle38070526.1
- Zhu H, Luo Y, Nissen-Meyer T, Morency C, Tromp J (2009) Elastic imaging and time-lapse migration based on adjoint methods 74:WCA167–WCA177
- Zhu W, Beroza GC (2018) PhaseNet: a deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International* 216(1):261–273, DOI 10.1093/gji/ggy423
- Zhu W, Xu K, Darve E, Beroza GC (2021) A general approach to seismic inversion with automatic differentiation. *Computers & Geosciences* 151:104751