An encoder-decoder deep surrogate for reverse time migration in seismic imaging under uncertainty

Rodolfo S. M. Freitas · Carlos H. S. Barbosa · Gabriel M. Guerra · Alvaro L. G. A. Coutinho · Fernando A. Rochinha

Received: 19 June 2020 / Accepted: 18 March 2021 / Published online: 6 April 2021
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2021

Abstract
Seismic imaging faces challenges due to the presence of several uncertainty sources. Uncertainties exist in data measurements, source positioning, and subsurface geophysical properties. Reverse time migration (RTM) is a high-resolution depth migration approach useful for extracting information such as reservoir localization and boundaries. RTM, however, is time-consuming and data-intensive as it requires computing twice the wave equation to generate and store an imaging condition. RTM, when embedded in an uncertainty quantification algorithm (like the Monte Carlo method), shows a many-fold increase in its computational complexity due to the high input-output dimensionality. In this work, we propose an encoder-decoder deep learning surrogate model for RTM under uncertainty. Inputs are an ensemble of velocity fields, expressing the uncertainty, and outputs the seismic images. We show by numerical experimentation that the surrogate model can reproduce the seismic images accurately, and, more importantly, the uncertainty propagation from the input velocity fields to the image ensemble.

Keywords Reverse time migration · Deep learning · Surrogate modeling · Uncertainty quantification

1 Introduction

Seismic imaging is employed to delineate the salient geological features of the Earth subsurface. Imaging methods are popular in the Oil & Gas industry as they are designed to be focused on the more essential characteristics: the horizons bounding the regions of interest. They can also be used in conjunction with inverse methods such as Full Waveform Inversion [16]. Imaging methods are designed and built departing from the integration of specialized optical (illuminating) principles and physics-based models describing the wave propagation through heterogeneous media. A critical aspect arising from such arrangement is the potential computational cost required, as a large domain is to be illuminated, which implies solving partial differential equations (PDEs) associated with the wave models in that area. The situation tends to be more complicated as the excitation signals include high-frequency content, which demands very fine grids in space and time. Such time-consuming tasks often hamper the use of high-fidelity codes constructed upon physics-based models. That becomes a more critical issue whenever one faces many-query applications like sensitivity analysis, design, optimization, or uncertainty quantification.

In this work, we develop a machine-learning model to alleviate computational costs and provide seismic images with quantified uncertainty [4]. In this context, we propose a Monte Carlo method (MC) to sweep a large ensemble of plausible velocity fields obtained by approximate methods, and, therefore, prone to errors. We compute an ensemble of images aiming at characterizing the propagated uncertainties along with the seismic image processing. Moreover, we embed the MC sampling as an outer loop of a larger computational workflow proposed in [4] and detailed in Algorithm 1. This algorithm is structured in three sequential stages, enabling a probabilistic framework
for seismic imaging. At this point, it is worth emphasizing that we consider here the velocity field estimation as the only uncertainty source. Other important aspects like noise in the measured signals or model discrepancies like the linearization underlying the RTM are not considered in the present phase of the development of our work.

The first stage aims at generating plausible subsurface velocity fields honoring seismic data. Probabilistic inversion, such as Bayesian tomography [4, 5, 10, 11], and stochastic FWI [9, 20, 32, 58, 60] can provide a velocity field ensemble used as input to the second stage. Hence, in Stage 2, an imaging technique migrates the seismogram information using each velocity field sample. This strategy wraps a seismic migration tool into an MC algorithm aiming to build a set of migrated seismic images. We have chosen the Reverse Time Migration (RTM) as the seismic migration technique to localize the seismic reflectors in the correct depth location in the subsurface [59]. RTM is a depth migration approach based on the two-way wave equation, frequently used in industry, that provides reliable subsurface high-resolution seismic images useful for seismic interpretation and reservoir characterization [59].

The last stage of the workflow post-processes the ensemble of RTM seismic images, calculating uncertainty maps and extracting features, such as horizons and faults, that characterize uncertainty in the resulting images. Our approach is conceived to maintain RTM structure represented by the three stages alluded above of typical workflows used by the industry. In that sense, it differs from [19, 41, 47] where seismic imaging is cast as a Bayesian inference problem, and the final images are samples from a probability density function. In both situations, uncertainties are considered, either in a forward propagation perspective or, as usual of Bayesian approaches, as an inverse problem.

Due to its flexibility by design, it is possible to generate different workflow versions, by, for instance, replacing components within the stages (e.g., different strategies for the velocity fields estimation in Stage 1) targeting to accommodate different demands or efficiency requirements. Nonetheless, we would still be facing a time-consuming computational task in the many-query Uncertainty Quantification (UQ) analysis of Stage 2. This motivates us to follow a consolidated trend, replacing the original physics-based model by a cheap-to-evaluate surrogate. Recently, Machine Learning techniques, like Gaussian Processes [7, 8, 37–39, 42] and Deep Neural Networks (DNNs), [26, 34, 35, 49, 50, 55, 61, 62] have deployed efficient surrogates for UQ analysis. Gaussian Processes have achieved considerable success with computer models using inputs and outputs of moderate dimensionality, including the ability to blend data of different sources, leading to multi-fidelity approaches [37, 39].

On the other hand, DNNs have gained popularity due to their unique profile of being flexible and scalable nonlinear function approximators. Another aspect worth highlighting is the substantial amount of computer libraries and tools available to enable their use.

Here, we apply the deep learning surrogate architecture proposed in [61] for systems governed by PDEs cast as an image-to-image problem. The performance of such architecture was tested in uncertainty quantification of flows in heterogeneous media [35], extended to semi-supervised learning in [62], and inverse problems in [34], with excellent results. This architecture comprises convolutional layers and dense blocks, following an encoder-decoder neural network arrangement to handle the potentially high-dimensional inputs and outputs. More specifically, we employ the deep learning architecture for constructing efficient proxies for RTM imaging by avoiding the high costs of solving a wave propagation equation twice in a heterogeneous medium. Such surrogates are nonlinear mappings linking the uncertain velocity field to the seismic images. It is worth highlighting that in contrast to typical surrogates, we do not replace only a forward solver associated to a PDE, but the entire expensive imaging process. The surrogate can handle the high-dimensional inputs (velocity fields) and outputs (seismic images), leading to cost savings in processing and memory storage. We demonstrate through three examples that such an approach enables producing seismic images with quantified uncertainty. Indeed, it can accurately reproduce the ensemble of images resulting from the uncertainty propagation via an MC sampling, and derived quantities as the structure tensor used to characterize seismic horizons.

---

**Algorithm 1** Workflow for seismic imaging with quantified uncertainty.

**Input**: source signals, seismograms, and spatial domain (raw data).

**Output**: ensemble of seismic images.

- **Stage 1**: Generate an ensemble of velocity fields:
  - Bayesian inversion with simplified physics-based models
- **Stage 2**: Propagate uncertainties – migrate the seismograms for the velocity field ensemble using RTM, producing a corresponding ensemble of seismic images
  - Monte Carlo loop over samples produced in Stage 1
- **Stage 3**: Post-process the RTM seismic images
  - Uncertainty maps;
  - Automatic features (horizons) detection;
  - Probabilistic characterization of such features;
reflecting boundary condition \( \partial \Omega_{D} \) applied on \( \partial \Omega_{D} \) the free-surface. The operator \( \partial \Omega_{D} \) represents the non-overlapping boundary partitions. The number of spatial dimensions, \( n_{sd} = 2, 3 \) the number of spatial dimensions, \( \partial \Omega_{D} = \partial \Omega_{D} \cup \partial \Omega_{inf} \subset \mathbb{R}^{n_{sd} - 1} \) is the domain boundary, and \( \partial \Omega_{D} \) and \( \partial \Omega_{inf} \) are non-overlapping boundary partitions. \( \partial \Omega_{D} \) is the portion of the boundary where Dirichlet boundary conditions are applied, representing, for instance, the free-surface. The operator \( \partial \Omega_{D} \) represents the non-reflecting boundary condition \( [12] \) applied on \( \partial \Omega_{inf} \). The pressure (the forward-propagated source wavefield) \( p(r, t) \) is defined at the position \( r = \{ r_i \} \in \Omega \) where \( i = 1, \ldots, n_{sd} \), and time \( t \in [0, T] \). Moreover, \( v(r) \) is the background compressional wave velocity spatial field, and \( f(r, t) \) is the seismic source. The vector \( r \) represents the seismic source position. The backward-propagated wavefield is calculated solving,

\[
\nabla^2 \bar{p}(r, \tau) - \frac{1}{v^2(r)} \frac{\partial^2 \bar{p}(r, \tau)}{\partial \tau^2} = s(r, \tau),
\]

\[
\bar{p}(r, \tau) = 0 \quad \text{on} \ \partial \Omega_{D} \quad \text{and} \quad \partial \bar{p}(r, \tau) = 0 \quad \text{on} \ \partial \Omega_{inf},
\]

\[
p(r, 0) = 0 \quad \text{and} \quad \frac{\partial p(r, 0)}{\partial \tau} = 0, \quad r \in \Omega, \quad (1)
\]

which is an equation similar to Eq. 1, but with a different source \( s(r, \tau) \), that is, the recorded signals at the receivers positioned in \( r \). Besides, the evolution in Eq. 2 is over the reverse time \( \tau = T - t \). Thus, the backward-propagated wavefield \( \bar{p}(r, \tau) \) is defined in \( \Omega \) and \( \tau \in [0, T] \).

The IC dictates the quality and fidelity of the final RTM image. There are several possibilities, for instance, excitation ICs \([13, 14, 36] \), extend ICs \([44, 45, 51] \), wavefield decomposition ICs \([31] \), and the zero-lag cross-correlation ICs \([15, 59] \). We have chosen, without loss of generality, the simple zero-lag cross-correlation between the forward and backward propagated waves at each point in \( \Omega \),

\[
I(r) = \int_{0}^{T} p(r, \tau) \bar{p}(r, \tau) \, dt. \quad (3)
\]

The IC amplitudes in Eq. 3 do not provide an explicit physical relationship with the reflection coefficients. In \([15] \), we find a detailed explanation of the relation between the imaging condition and the reflection coefficient. Nevertheless, the resulting image provides the correct amplitude contrast locations of the geological interfaces of rocks with different physical properties \([59] \). The amplitude contrast patterns are the main feature of the migrated seismic images explored in the present work.

We apply an explicit 2\textsuperscript{nd}-order in time and 4\textsuperscript{th}-order in space finite difference numerical scheme \([48] \) to approximate (1) and (2), leading to the vector \( \mathbf{v} \), the discrete version of the velocity field, and similarly the vectors \( \mathbf{p}, \hat{\mathbf{p}}, \mathbf{s}, \mathbf{f} \) at each time step. Here, we implemented the numerical scheme in MATLAB \([33] \). Note that the vectors \( \mathbf{p}, \hat{\mathbf{p}}, \mathbf{s}, \mathbf{f} \) have the same dimension, that is \( N = N_x \times N_y \) (or \( N = N_x \times N_y \times N_z \) in 3D), where \( N_x, N_y \) (and \( N_z \)) are the number of grid points in each Cartesian direction. Each discrete seismogram is a vector of size \( N_{rec} \times (N_\tau + 1) \), where \( N_{rec} \) is the number of receivers, and \( N_\tau = T/\Delta t \), with \( \Delta t \) is the time step. The seismic source \( \mathbf{f} \), here a Ricker-type wavelet \([43] \), has dimension \( N_t \). RTM is not only computationally intensive but it is also data-intensive due to the high dimensional inputs, the amount of data to manage, for instance, storing and retrieving \( \mathbf{p}, \hat{\mathbf{p}} \) and the computational
costs associated with imposing the stability and dispersion conditions in the discrete two-way wave equation [59]. The dispersion relation takes into account the number of grid points per wavelength and the medium properties, which in our isotropic acoustic case is the P-wave velocity. Thus, complexity in estimating the migrated image increases with high heterogeneous media and seismic source cutoff frequency. Least-squares RTM (LSRTM) extends its classical approach to an iterative method by linearizing the wave equation with respect to the background velocity field [46, 57]. Besides, the LSRTM can be made either in the data domain or model domain [59]. In the present work, we restrict ourselves to the standard RTM technique that corresponds to the LSRTM’s first iteration in the data domain [46, 59] as described in Eqs. 1, 2 and 3.

As we wrap RTM into a sampling method in Algorithm 1, for taking into consideration the input uncertainties, the overall computational cost of Stage 2 rises proportionally to the number of samples for the MC method to achieve the statistical convergence, \( N_{MC} \), the cardinality of the ensemble of possible velocity fields. Typically, seismic raw data recording sets are split into multiple steps (\( N_{shots} \)) to cope with the potentially high spatial dimensions to be covered and to enhance the signal to noise ratio (SNR) in processing stages. Each step covers fully or partially the imaging domain corresponding to different arrangement of sources and receivers. It is essential to mention that RTM sweeps over the number of shots producing partial migrated images per shot, computed by Eq. 3. When this loop ends, a process called stacking sums the partially migrated seismic images into a single stacked seismic image [27, 56]. Algorithm 2 details the generation of the ensemble of RTM images, where a set of seismograms, \( \{ s_1, \cdots, s_{N_{shots}} \} \), a set of velocity fields, \( \{ v_1, \cdots, v_{N_{MC}} \} \), and a seismic source (\( f \)) are given as inputs. The indexes \( N_{shots} \) and \( N_{MC} \) represent the number of shots and the number of samples for the MC method. For each MC iteration, we solve the wave equation twice, firstly for the seismic source and secondly for the seismograms associated with it. The computation of the imaging condition uses both solutions (source wavefield, and receiver wavefield), retrieving from persistent storage the source wavefield to build the migrated seismic section and stacking the partial results over time (\( I_{\sum I} \)), and over the number of seismograms (\( I_{\sum shotjd} \)). At the end of Algorithm 2, we obtain the discrete seismic image set \( \{ I_1, I_2, \cdots, I_{N_{MC}} \} \) where each \( I_s \) is a vector in \( \mathbb{R}^N \). Often, each image is filtered to sharpen the resulting features. Nevertheless, we do not apply any filter to the ensemble of seismic images. This enables us to verify if the surrogate model can reproduce the raw seismic image, which may contain noise related to the high frequency in the source term. Summarizing, migrations of seismograms for the set of velocity fields produce the corresponding set of migrated seismic images, where each has a direct relation with one velocity sample.

**Algorithm 2** Reverse time migration under uncertainty.

**Require:** \( \{ v_1, \cdots, v_{N_{MC}} \}, \{ s_1, \cdots, s_{N_{shots}} \} \), and \( f \)

1: function RTM\_UQ( vectors \( \{ v_1, \cdots, v_{N_{MC}} \} \), vectors \( \{ s_1, \cdots, s_{N_{shots}} \} \), vector \( f \) )

2: for \( s = 1 \) to \( N_{MC} \) do

3: \hspace{1em} read \( v_s \), and \( f \)

4: \hspace{1em} initialize image condition \( I_{\sum shotjd} = 0 \)

5: \hspace{1em} for \( shotjd = 1 \) to \( N_{shots} \) do

6: \hspace{2em} initialize \( n_t = 0 \)

7: \hspace{2em} apply initial conditions for \( i_t = 0 \)

8: \hspace{2em} for \( i_t = 1 \) to \( N_t \) do

9: \hspace{3em} \( n_t = n_t + i_t \cdot \Delta t \)

10: \hspace{3em} solve Eq. 1 \( \triangleright \) source wavefield

11: \hspace{3em} store \( p_n \)

12: \hspace{2em} end for

13: \hspace{1em} initialize \( n_t = 0 \), and \( I_{\sum n_t} = 0 \)

14: \hspace{1em} apply initial conditions for \( i_t = 0 \)

15: \hspace{1em} for \( i_t = 1 \) to \( N_t \) do

16: \hspace{2em} \( n_t = N_t - (n_t + i_t \cdot \Delta t) \triangleright \) reverse time

17: \hspace{2em} read \( p_n \), and \( s_{shotjd} \)

18: \hspace{2em} solve Eq. 2 \( \triangleright \) receiver wavefield

19: \hspace{2em} calculate \( I_{\sum n_t} = I_{\sum n_t} + p_n \hat{p}_{n_t} \triangleright \) imaging condition

20: \hspace{2em} end for

21: \hspace{1em} stack \( I_{\sum shotjd} = I_{\sum shotjd} + I_{\sum n_t} \triangleright \) stacking

22: end for

23: \( I_s \leftarrow I_{\sum shotjd} \)

24: \hspace{1em} store \( I_s \forall s \in [1 \leq s \leq N_{MC}] \)

25: end for

26: end function

Different strategies can be pursued in order to make Algorithm 1 practical, by reducing the inherent computational costs of processing and storage. They could rely, for instance, on data compression [2, 28, 30, 53] or more effective stochastic sampling [3], but here, as mentioned before, our choice is to use deep learning surrogates for the RTM imaging, described in the following section.

### 3 Deep learning surrogate

The main goal of surrogate models is to replicate the multivariate input-output mapping provided by physical models governed by PDEs to save computational costs. Performing uncertainty quantification in such conditions is often hampered whenever one faces high-dimensionality, the so-called curse of dimensionality. As pointed out
in the literature, DNNs have proved successful in such situations by exploiting low-dimensional latent spaces and sophisticated training methods [22]. We aim to construct and evaluate the performance of DNNs acting as a surrogate model for the RTM imaging under uncertainty, using as baseline the encoder-decoder architecture proposed by [61] and designed for problems cast as image-to-image regressions. We briefly review the building blocks of the network in this section.

We present in Fig. 1 a schematic view of the network architecture, that follows the basic structure proposed in [61]. It provides the big picture of our end-to-end solution, depicting the main components of the encoder-decoder network. In our particular application, the input for the deep learning surrogate is the ensemble of heterogeneous velocity fields, and the outputs are the corresponding seismic images, calculated by Algorithm 1. Inputs and outputs are high-dimensional spatial fields, and the surrogate training is cast as a field-to-field regression. This approach is image-inspired, it relies on connecting each pixel of the input field to an output pixel, where pixels correspond to grid points in the input computational mesh and output fields. The trained network maps the velocity $v \in \mathbb{R}^N$ into the seismic image amplitudes $I \in \mathbb{R}^N$.

The encoder-decoder architecture displayed in Fig. 1 consists of two sequential main phases. The first is the encoder, where dimensionality reduction occurs, followed by the decoder that reconstructs the network output with its original dimension. Alternating dense blocks and transition layers constitute both phases. Indeed, this architecture merges key characteristics of fully connected networks with convolutional networks. On one side, convolutional networks are quite effective in dimensionality reduction [22] and are capable of capturing spatial correlations present in the input velocity fields. Fully connected layers enhance the information transmitted across the network, improving the overall efficiency reflected in reasonable sizes of the needed training data set [23].

Dense-blocks connect all layers directly to each other, helping the training process with the improvement of information flow and gradients across the network [23]. Inputs of the $l$-th layer are the concatenated outputs from the previous layers, that is, $z^l = H^l([z^0, z^1, \ldots, z^{l-1}])$, with $z^l$ the output of $l$-th layer, and $[z^0, z^1, \ldots, z^{l-1}]$ refers to their concatenation, and $[0, \ldots, l-1]$. $H^l$ is a non-linear transformation. Here, $H^l$ results from applying three consecutive operations, batch normalization [25] followed by a ReLU [21] and, convolution. The dense-block has two design parameters, the number of layers $L$, and the growth rate, $K$, the number of output features of every single layer. The transition layers here, in the encoder (decoder), are convolutional (deconvolutional) and, therefore, handle the dimension inputs or outputs of dense blocks. As shown in Fig. 1, during the encoder phase, a convolutional layer is used to extract feature maps from the high dimensional velocity fields. The extracted feature maps are processed by an alternating series of dense blocks and encoding layers. The last layer of the encoder phase produces low-dimensional feature maps that characterize the high dimensional field, as shown in the purple maps in Fig. 1. Such maps are passed to the decoder phase, which is composed of an alternating series of dense-blocks and decoding layers, returning the seismic images to its (original) dimension.

The surrogate $g$ is expressed formally in a compact notation as,

$$\hat{I}_i = g(v_i; w),$$

where $\hat{I}_i$ is the surrogate output (seismic image, $I_i$) for an input $v_i$, and $w$ contains the parameters of the neural network. Training the neural network means learning parameters $w$ using data from the set $D = \{(v_i, I_i)\}$, $i = 1 \cdots N_{\text{train}}$ obtained from simulations with the RTM algorithm, where $N_{\text{train}}$ is the number of samples in the training set. The stochastic gradient descent algorithm estimates the parameters $w$ of the network for a given loss function [17]. We consider the following $L_2$ regularized mean squared error (MSE) function,

$$L_{MSE} = \frac{1}{N_{\text{train}}} \sum_{i=1}^{N_{\text{train}}} \|\hat{I}_i - I_i\|_2^2 + \alpha \Omega(w).$$  \hspace{1cm} (5)

Here the penalty function is given by $\Omega(w) = \frac{1}{2} w^T w$ for the $L_2$ regularization. Moreover, the root mean squared error (RMSE) is used for monitoring the convergence of the training error. The RMSE is given by,

$$RMSE = \sqrt{\frac{1}{N_{\text{train}}} \sum_{i=1}^{N_{\text{train}}} \|I_i - \hat{I}_i\|_2^2}.$$  \hspace{1cm} (6)
4 Numerical experiments

Here, we present three examples to demonstrate the ability of the encoder-decoder surrogate to replace the original two-way wave equation RTM algorithm efficiently. To facilitate assessing the surrogate response we replace the Bayesian inversion in the first stage of Algorithm 1 by generating the velocity ensemble with an imposed uncertainty level and form. Hence, we assign to the different geological layers random velocity magnitudes in order to produce synthetic data to train the neural network and perform the uncertainty analysis. The first example deals with a medium containing two flat geological layers of constant velocity. We increase the difficulty for the surrogate in the second example, by employing a more complex medium, in which the five geological layers are no longer flat, implying horizontal heterogeneity. The final example has six geological layers with two tilted interfaces and three-flat interfaces. Two geological layers are wedge-shaped and used to test the resolution limit and the surrogate model’s capacity to replicate this feature.

The encoder-decoder networks are constructed using the open platform Tensorflow [1]. The Adam optimizer algorithm [18] is employed for parameter learning, considering a weight decay of $1 \times 10^{-5}$, and an initial learning rate of $1 \times 10^{-3}$, with a learning rate scheduler, that is used to drop two times on plateau of the rooted mean squared error. We compute a total of 1300 samples by considering the velocity magnitude constant in the interior of each geological layer.

We assign for all examples the following form of the stochastic velocity field

$$v = \sum_{l=1}^{n_l} \bar{v}_l (1 + \sigma_l \xi_l) \mathbf{P}_l \tag{7}$$

where $n_l$ is the number of geological layers, $\bar{v}_l$ is the mean velocity within each geological layer, $\sigma_l$ is the standard deviation, here assumed as 5%, $\xi_l \sim U[-1, 1]$ is a random variable following a uniform distribution, and $\mathbf{P}_l$ is an $N$-dimensional vector containing 1 in the components corresponding to the $l$-th geological layer grid points and 0 otherwise. After that, we apply a moving harmonic average to $v$ with a window length of 20 grid points to mimic the background velocity fields computed by parameter model building techniques like tomography or full-waveform inversion [46]. To analyze the surrogate training convergence, out of 1300 samples computed with the original RTM model, we create four training datasets with 200, 400, 600, and 800 samples each. Such an approach allows us to evaluate the convergence of the surrogate training. We used the remaining $N_{test} = 500$ samples to test the trained network.

Accuracy is measured using the distance between predictions with the surrogate model and those computed with the RTM original model. To evaluate the surrogate model quality, we consider the coefficient of determination ($R^2$-score) metric [52]. We define the coefficient of determination as,

$$R^2 = 1 - \frac{\sum_{i=1}^{N_{test}} \| \mathbf{I}_i - \hat{\mathbf{I}}_i \|^2}{\sum_{i=1}^{N_{test}} \| \mathbf{I}_i - \bar{\mathbf{I}} \|^2} \tag{8}$$

where $\mathbf{I} = \frac{1}{N_{test}} \sum_{i=1}^{N_{test}} \mathbf{I}_i$ is the mean of test samples. The $R^2$-score metric represents the normalized error, allowing the comparison between surrogate models trained by different datasets, with values close to 1 corresponding to the surrogate models best accuracy. Here, we consider 0.95 a good target. Also, we intend that the surrogate model returns not only good predictions of seismic images but also accurate estimations of quantities that characterize the uncertainties in such images. To measure the degree of uncertainty in the seismic images, we follow the approach proposed in [29]. In their approach, the degree of uncertainty is expressed by a confidence index that represents the pointwise normalized standard deviation, where low values represent high variabilities and high values the opposite. The confidence index is,

$$c(r) = \frac{\sigma_{max} - \sigma(r)}{\sigma_{max} - \sigma_{min}}, \tag{9}$$

where $c(r)$ is the confidence index at position $r$, $\sigma(r)$ is the standard deviation at position $r$, and $\sigma_{min}$ and $\sigma_{max}$ are the minimum and maximum field standard deviations, respectively. Another form of measuring the degree of uncertainty is the coefficient of variation, defined as the pointwise ratio between the standard deviation and the mean,

$$c_v(r) = \frac{\sigma(r)}{\mu(r)} \tag{10}$$

![Fig. 2 Simple geologic setup: two horizontal geological layers](image-url)
Table 1 RTM numerical parameters

| Parameter | Value | Description (Unit) |
|-----------|-------|--------------------|
| $h$       | 20    | Spatial discretization (m) |
| $\Delta t$ | $2.22 \times 10^{-3}$ | Temporal discretization (s) |
| $t_a$     | 0.5   | Total acquisition time (s) |
| $N_x \times N_y$ | 50 × 50 | Number of grid points |
| $i_{srcx}, i_{srcy}$ | (5:5:45), 5 | Source positions |
| $i_{isx}, i_{isy}$ | ([1:1:50], 5) | Receiver positions |

where $\mu(r)$ is the expected value at position $r$.

4.1 A simple geologic scenario: Efficiency and convergence analysis

In this first example, designed to evaluate the efficacy and efficiency of the proposed surrogate, we assume a simple geologic scenario in which two horizontal homogeneous geological layers separated by a flat interface parallel to the surface composes the subsurface, as shown in Fig. 2. This domain has 1000 m of depth and 1000 m of lateral extension, where the velocity in the first geological layer is 3000 m/s, and the velocity in the deeper geological layer is 4500 m/s.

We produce synthetic data running the RTM solver using the reference velocity field of Fig. 2, and a seismic source with a cutoff frequency of 30 Hz. In such modeling, we simulate a fixed split-spread acquisition [27] comprising nine shots, where receivers are positioned near the surface for each shot and equally spaced at 20 meters. The seismic source is also placed near the surface and moved 100 meters for each shot, covering the entire domain with nine shots. Table 1 shows the parameters used in the numerical modeling of the acoustic wave equations and the positioning of the seismic source and receivers given by the index ranges $[i_{srcx}, i_{srcy}]$ for the sources, and $[i_{isx}, i_{isy}]$ for the receivers. The grid size and time step respect the numerical dispersion, and stability criteria [48].

Table 2 details the neural network architecture. The neural network is constructed by combining trial-and-error and Hyperopt algorithm [6] to search for the hyperparameters that give the best $R^2$-score. The first layer is convolutional, with kernel size equal to 4 and stride 2. This first layer captures spatial relations from the velocity input. The number of features maps after the first convolutional layer is 48. The neural network has 3 dense-blocks with $L = 4$ and $K = 16$. Dense-blocks have a kernel size equals 3, and a stride of 1. Encoder-decoder layers have a kernel size of 3 and a stride of 2. In the decoding layer, we introduce a transposed convolution, allowing the output to be reconstructed to its original dimensionality, equal to the computational grid. Also, data scaling is a highly recommended pre-processing step to improve deep learning models’ stability and performance [22]. To enhance the network training, we normalize the data such that the image condition at each grid point only assumes values between 0 and 1.

Table 2 Neural network architecture

| Layers      | Output | Dimension |
|-------------|--------|-----------|
| Input       | 1      | $50 \times 50$ |
| Convolution | 48     | $24 \times 24$ |
| Dense-block 1 | 112   | $24 \times 24$ |
| Encoding    | 56     | $12 \times 12$ |
| Dense-block 2 | 120   | $12 \times 12$ |
| Decoding 1  | 60     | $24 \times 24$ |
| Dense-block 3 | 124   | $24 \times 24$ |
| Decoding 2  | 1      | $50 \times 50$ |
| ReLU        | 1      | $50 \times 50$ |

“Outputs” represents the number of features maps and “Dimension” is the dimension of the features maps.

Fig. 3 RMSE decay with number of epochs

Fig. 4 Test $R^2$-score and efficiency in function of the number of training samples
and 1. Consistently with that choice, a final ReLU layer [21] imposes that the outputs are positive. The resulting neural network has 218,425 parameters to be estimated along the training. That number was furnished by a built in function of Tensorflow.

Figure 3 shows the decay of the RMSE as a function of the number of epochs during the training process, for training data sets ranging from 200 to 800 samples. Note that the RMSE stabilizes after 150 epochs for all cases and that for smaller data sets, we see higher error values. The key characteristics that one is seeking when replacing the original expensive computational model by a surrogate are efficiency and accuracy. The encoder-decoder surrogate is conceived for alleviating computational costs of a Monte Carlo (MC) analysis involving the ensemble of plausible velocity fields. To evaluate the surrogate performance in such a context, we introduce the index in Eq. 11. The MC’s overall cost relies on the number of runs of the wave propagation solver, two (forward and adjoint) for each sample. On the other hand, as a run of the surrogate is inexpensive when compared to the one of the two-way solver, the main component of the cost of the encoder-decoder lies on the training, which is split into two stages: generating the training and solving the optimization problem to obtain the weights of the neural network. Our experience demonstrates that the second is not relevant. Therefore, $N_S$ and $N_{MC}$ are good proxies for each option’s computational costs, and their ratio makes the core of our efficiency-index.

$$\text{Efficiency} = \left(1 - \eta \frac{N_S}{N_{MC}}\right) \times 100 \quad (11)$$

The index in Eq. 11 estimates the percentage of the saved computational costs, and $\eta$ is an adjustment factor accounting for the time spent in the construction, training, and making predictions with the surrogate model. Without loss of generality, we assume for now $\eta = 1.0$. For less optimistic conditions, we recognize that $\eta > 1.0$. We calculate the coefficient of determination $R^2$ to assess the accuracy of the neural network with the remaining 500 samples. We observe that the surrogate model accuracy is good even in small training data scenarios, reaching $R^2 \geq 0.95$, as shown in Fig. 4. Furthermore, Fig. 4 also depicts the surrogate efficiency. Here it is worth highlighting that 600 RTM runs are necessary to compute the variance with a relative error of $1 \times 10^{-3}$. Thus, we see that the efficiency is higher than 90%, even for the larger data set with 800 samples.

To further illustrate how the surrogate model approximates the predictions of the original model with good accuracy, Fig. 5 shows comparisons for two realizations chosen randomly from the test set. We observe that the surrogate presents good results, returning good predictions of the seismic image amplitudes. We also depict a comparison between the standard deviation, $\sigma(r)$, the confidence index,
Fig. 6 UQ indexes - standard deviation, \( \sigma(\mathbf{r}) \), confidence index, \( c(\mathbf{r}) \), and coefficient of variation, \( c_v(\mathbf{r}) \) - predicted by the RTM (left) and surrogate models (right). The relative errors to the RTM model are lower than 2%.

\[ c(\mathbf{r}) \text{, and the coefficient of variation, } c_v(\mathbf{r}) \text{, predicted by the original and surrogate models with } N_{\text{testing}} = 500 \text{ testing samples, see Fig. 6. Besides, we introduce a discrete version of a } L_2 \text{ relative error between two spatial fields } g, \text{ one computed with the RTM and the other by the surrogate as,} \]

\[
\epsilon_g^2 = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{g_i^\text{RTM} - g_i^s}{g_i^\text{RTM}} \right)^2
\]

where the subscripts refer to how we compute the field \( g \). This index is an average of the pointwise relative error for all \( N \) grid points. The visual resemblance of the images in Fig. 5 is quantified using Eq. 12, leading to relative errors that stay below 2%. For the indexes in Fig. 6, the relative errors computed with Eq. 12 are less than 2%.

We also investigate the impact of using a limited number of data to construct the surrogate. It is well known that this entails a new source of epistemic uncertainty. For carrying out such an analysis, we follow the idea introduced in [40]

| Samples | \( \rho_{N_s} \) |
|---------|---------------|
| 200     | 0.4156        |
| 400     | 0.1400        |
| 600     | 0.0668        |
| 800     | 0.0651        |

\( N_s \) represents the number of samples in the training process.
of using Spearman’s correlation. This coefficient expresses the relation of two random variables. Here, the first is $\epsilon_i$, the difference between IC values at each grid point computed by the surrogate and the original model, for each velocity field sample. The second random variable is representative of the resulting uncertainty. We take $\gamma_i = (\hat{I}_i - \mu_I)^2$, the contribution of each sample for the variance of the ensemble of images at the $j$-th pixel. Therefore, the Spearman’s coefficient reads as

$$
\rho_{N_S}^j = \frac{\sum_{i=1}^{N_S} (\epsilon_i - \mu_{\epsilon})(\gamma_i - \mu_{\gamma})}{\sqrt{\sum_{i=1}^{N_S} (\epsilon_i - \mu_{\epsilon})^2 \sum_{i=1}^{N_S} (\gamma_i - \mu_{\gamma})^2}}
$$

(13)

where $\mu_{\epsilon}$ and $\mu_{\gamma}$ stand, respectively, for the expected values for both random variables. The superscript $N_S$ is included to refer to the number of samples used for the particular surrogate. The rational of the analysis lies on observing how Spearman’s coefficient evolves with the increasing of the number of samples employed for obtaining the surrogate. As we observe, in Table 3 where we present the averaged value over all grid points, the correlation decreases, and, consequently, the epistemic uncertainty also diminishes.

### Table 4 RTM numerical parameters

| Parameter | Description (Unit) |
|-----------|--------------------|
| $f_{cutoff}$ | Spatial discretization (m) |
| $f_{cutoff}$ | Temporal discretization (s) |
| $t_a$ | Total acquisition time (s) |
| $N_x \times N_y$ | Number of grid points |
| $i_{srcx}, i_{srcy}$ | Source positions |
| $i_{rx}, i_{ry}$ | Receiver positions |

4.2 A non-flat medium with five geological layers

To challenge the encoder-decoder surrogate, we use a synthetic geologic model with five homogeneous layers similar to the one proposed in [24]. The 2D velocity model consists of a water layer with velocity 1500 m/s, and four mini sedimentary basins with velocities of 2000 m/s, 2500 m/s, 3000 m/s, and 4000 m/s, respectively. Figure 7 display a schematic view of the velocity field with 1000 m of depth and 1000 m of lateral extension.

Two synthetic seismogram sets are generated for the velocity fields shown in Fig. 7 considering now the seismic sources with cutoff frequencies of 30 and 45Hz. Table 4 shows the RTM parameters and the positioning of sources and receivers. For the cutoff frequency of 45Hz, due to the imposition of the stability and dispersion conditions in the discrete two-way wave equation, there is a significant

### Table 5 Neural network architecture

| Layers | Output | Dimension |
|--------|--------|-----------|
| Input  | 1      | 100 × 100 |
| Convolution | 48 | 48 × 48 |
| Dense-block 1 | 112 | 48 × 48 |
| Encoding | 56 | 24 × 24 |
| Dense-block 2 | 120 | 24 × 24 |
| Encoding | 60 | 12 × 12 |
| Dense-block 3 | 124 | 12 × 12 |
| Decoding | 62 | 24 × 24 |
| Dense-block 4 | 126 | 24 × 24 |
| Decoding | 63 | 48 × 48 |
| Dense-block 5 | 127 | 48 × 48 |
| Decoding | 1 | 100 × 100 |
| ReLU | 1 | 100 × 100 |

“Outputs” represents the number of feature maps and “Dimension” is the spatial dimension of the features maps.
Fig. 9 Test $R^2$-score and efficiency in function of the number of MC samples. (a) $R^2$-score for the trained networks. (b) Efficiency

![Graphs showing test $R^2$-score and efficiency vs. number of MC samples.](image)

Fig. 10 Randomly selected images from the test data set computed by the RTM model (a) and the relative error between the seismic images computed by the RTM and surrogate model (b). The relative errors in the seismic image amplitudes, $e_I$, are lower than 6%
Fig. 11 UQ indexes - standard deviation, $\sigma(r)$, confidence index, $c(r)$, and coefficient of variation, $cv(r)$ - predicted by the surrogate model (left) and the relative errors between the surrogate predictions to the RTM model (right). The relative errors between the surrogate predictions to the RTM model for the UQ indexes are less than 6%.
Table 6 Spearman’s correlation coefficient between the absolute error and variance

| Samples | ρ    |
|---------|------|
| 400     | 0.1098 |
| 600     | 0.0999 |
| 800     | 0.0529 |

“Samples” represents the number of samples in the training process.

increase in the input dimensionality, that bears the potential to hamper the neural network training. The next subsections present results for the scenarios involving the two frequencies of excitation.

4.2.1 Cutoff frequency - 30Hz

We detail the architecture of the neural network for this scenario in Table 5. It contains five dense blocks, leading to 412,210 parameters to be trained. We can see in Fig. 8 the RMSE decay as the number of epochs increase for all training sets. We verify the accuracy of the surrogate by computing the $R^2$ score for the 500 testing samples. We find that, as expected, for networks trained with larger data sets, $R^2$ values are closer to 1.0, as shown in Fig. 9a. Due to limitations imposed by the high cost of generating samples using the full RTM model for this example, we develop a different efficiency analysis extrapolating from the basic conditions used for the network training. We assume conservatively that $N_{MC}$ is of the same order of the case in Section 4.1. Thus, we start from a scenario where only 5,000 samples are needed to characterize uncertainties in the seismic images and extrapolate to more expensive potential scenarios requiring hypothetically up to 50,000 samples. Here, $N_S$ is equal to 1100 samples, where 600 samples to train the neural network with an accuracy of $R^2 \geq 0.95$, and 500 to test the surrogate model. Figure 9b depicts the efficiency analysis in function of $N_{MC}$. We note, for the worst scenario, an efficiency of around 78%, and for the scenarios where $N_{MC}$ is higher than 10,000 samples, the efficiencies reach values greater than 90%. For the most expensive scenario, we see an efficiency close to 98%.

Figure 10 shows comparisons between three images randomly selected from the test set and their relative errors to the RTM model computed by Eq. 12. We observe that the surrogate model returns excellent predictions of the seismic image amplitudes with errors less than 6%. We can also note that the surrogate model predicts the UQ figures - standard deviation, $\sigma(r)$, confidence index, $c(r)$, and coefficient of variation, $c_v(r)$ - with good accuracy, as seen in Fig. 11, where the relative error between the surrogate predictions and the RTM model are lower than 6%. Results in Figs. 10 and 11 show that the encoder-decoder surrogate extrapolates the replication of the IC training targets. Also in this case the Spearman’s correlation coefficient decreases with the increasing number of samples employed for obtaining the surrogate, and, consequently, the epistemic uncertainty also diminishes, as depicted in Table 6.

Next, we deepen our investigation of the surrogate’s ability to reproduce the probabilistic characterization of

Fig. 12 Comparison between PDFs predicted by the RTM model and the surrogate model
Algorithm 1. We provide, using the surrogate, a view of the uncertainties associated with specific seismic targets, the interfaces of geological layers. This view can reveal how the propagated uncertainties can directly impact the images posterior interpretation. Figure 13 provides the seismic image amplitude mean value and associate confidence bands for the four interfaces. In the right part of the figure, we illustrate the uncertainties spatial distribution, having as background a randomly selected image from the ensemble. To promote visual perception, we plotted amplified seismic image amplitude confidence bands associated with each interface. Those bands reflect the seismic image amplitude value dispersion within the images ensemble, and, therefore, might lead to a lack of confidence in the reflector placement.

Table 7  Neural network architecture

| Layers         | Output | Dimension       |
|----------------|--------|-----------------|
| Input          | 1      | 150 x 150       |
| Convolution    | 48     | 72 x 72         |
| Dense-block 1  | 112    | 72 x 72         |
| Encoding       | 56     | 36 x 36         |
| Dense-block 2  | 120    | 36 x 36         |
| Encoding       | 60     | 18 x 18         |
| Dense-block 3  | 124    | 18 x 18         |
| Decoding       | 62     | 36 x 36         |
| Dense-block 4  | 126    | 36 x 36         |
| Decoding       | 63     | 72 x 72         |
| Dense-block 5  | 127    | 72 x 72         |
| Decoding       | 1      | 150 x 150       |
| ReLU           | 1      | 150 x 150       |

“Outputs” represents the number of features maps and “Dimension” is the dimension of the features maps.
ones from the previous example. We do not expect to obtain the surrogate’s optimal performance by employing such a strategy, but it could be useful in practice for saving computational costs. Table 7 shows the neural network architecture for the 45 Hz scenario. The network architecture is the same as in Section 4.2.1 with small changes. The first convolutional layer has a kernel size equal to 7 and a stride of 2. The total number of parameters in the network is 416,390. Figure 14 shows RMSE decay as a function of the number of epochs in the training process.

We can see in Fig. 15a the $R^2$ score for different training sets showing for this more difficult scenario a slight decrease in the neural network quality. Confirming our initial expectations of a non-optimal but acceptable performance, the coefficients of determination $R^2$ for all training datasets are lower than 0.90. Moreover, we estimate the efficiency of the surrogate model in the same manner as in the previous case. However, here we consider values for the adjustment factor $\eta > 1.0$. More precisely, the adjustment factor tries to estimate the time spent in search of the neural network hyperparameters to optimize the surrogate model accuracy and to generate larger training sets. Without loss of generality, we assume that the number of samples $N_S$ to train the neural network is equal to 1100, 600 to train, and 500 to test the surrogate model. Figure 15b shows the efficiency in function of $N_{MC}$, for several adjustment factors. Note that for scenarios with $N_{MC} \leq 10,000$, the efficiency drops significantly for higher adjustment factors. However, for scenarios where $N_{MC} \geq 20,000$ samples the efficiency reaches values close to 80-90%. For scenarios where $N_{MC} \geq 40,000$ we observe an efficiency close to 90% even for the higher adjustment factor.

Despite the lower accuracy presented in this scenario, the surrogate model could reach satisfactory predictions of the seismic image amplitudes, as we can see in Fig. 16. In this Figure, we show three randomly selected images from the test data set computed by the surrogate model and the corresponding errors computed by Eq. 12. Observe that the errors are lower than 10%. Note that the grid is adjusted only to satisfy the stability and dispersion criteria for the 45 Hz cutoff frequency. We do not optimize the domain size for a proper representation of the non-reflecting boundary conditions and source/receiver arrangement. Furthermore,
Fig. 16 Randomly selected images from the test data set computed by the surrogate model (a) and the relative error between the seismic images computed by the RTM and surrogate model (b) trained with 600 samples. The relative errors $e_I$ are lower than 10%.

Fig. 17 shows the standard deviation, $\sigma(r)$, confidence index, $c(r)$, and coefficient of variation, $c_v(r)$, computed by the surrogate model and the respective errors. We observe that the surrogate model predicts the UQ indexes with satisfactory accuracy. The relative errors between the surrogate predictions to the RTM model for the UQ indexes are lower than 8%.

We now investigate the probability density functions (PDFs) of the seismic image amplitudes at the control points in Fig. 7. We use again as reference solution PDFs...
Fig. 17 UQ indexes - standard deviation, $\sigma(r)$, confidence index, $c(r)$, and coefficient of variation, $c_v(r)$ - predicted by the surrogate model (left) and the relative errors between the surrogate predictions to the RTM model (right). The relative errors between the surrogate predictions to the RTM model for the UQ indexes are lower than 8%.
obtained by the RTM model with 500 test samples to verify the accuracy of the surrogate models trained with different datasets to estimate the PDFs at the control points. Figure 18 depicts the seismic image amplitude PDFs at the control points estimated by the surrogate model trained with 200, 400, 600, 800 samples, together with the reference PDFs. We observe that the PDFs obtained with the surrogate model capture well the reference PDFs in all control points, particularly for large training datasets.

4.3 Moving towards a probabilistic characterization of the geological horizons

This final example intends to explore the underlying probabilistic geological horizons’ structure that results from the uncertainty propagation in the migration processes. With that, we demonstrate the end-to-end workflow for images with quantified uncertainty, emphasizing the role played by the proposed surrogate. Producing seismic images is goal-oriented to unveil such geologic structures. In the present context, the spatial curve representing each horizon is considered a realization of an intrinsic spatial random function, more precisely a random process induced by the uncertainties in the velocity field estimation. Such an interpretation follows in line with [47]. Differently from the previous example, instead of considering each point’s

| Samples | ρ   |
|---------|-----|
| 400     | 0.0568 |
| 600     | 0.0139 |
| 800     | 0.0052 |

“Samples” represents the number of samples in the training process.
uncertainties (associated with an image pixel), detecting the horizons involves structure tensors, whose components are the vertical and horizontal entries of the image spatial gradients [54].

We have one migrated image per estimated velocity field containing the target horizon. In a standard RTM workflow, an interpreter would be in charge of picking such a horizon, which is not feasible here as this effort would have to be replicated for the whole ensemble of images. Thus, we apply automatic horizon extractors [54] to the images ensemble, which provides horizon interpretations for a seismic reflector seeded by users in control points. It is worth emphasizing that we expect that capturing the probabilistic structure of the horizon involving spatial correlations requires a large number of samples when compared to the previous examples. Again, the proposed surrogate would work as an enabler of such process.

As a proof-of-concept, we consider the seismic image outcomes for the RTM and encoder-decoder surrogate from on the geologic model with six layers shown in Fig. 19. The geologic model presented in Fig. 19 has one more layer than the example from Section 4.2 and two wedge-shaped geologic structures that impose additional imaging difficulties. Although this example displays significant geological differences compared to the examples presented in Sections 4.1 and 4.2, we employ the same computational configurations. Specifically, the RTM and surrogate model setups are similar to the non-flat medium with five geological layers example, and their numerical parameters and Neural Network architecture can be seen in Tables 4
Comparison among the interpreted horizons based on the seismic images provided by the RTM and surrogate models and 5. However, we did not expose here the convergence analyses as detailed previously for the first two numerical examples. Here, it is worth highlighting that the surrogate model trained with 800 samples has a $R^2$-score equals to 0.9482. Also, Table 8 shows that the Spearman’s correlation coefficient decreases with the increasing number of samples, hence decreasing also the epistemic uncertainty of the surrogate model.

Moving towards images uncertainties, Fig. 20 shows samples of horizons, each one extracted from five randomly selected seismic migrated images, either obtained by the RTM technique (Fig. 20a) or by the surrogate model (Fig. 20b). The horizons correspond to the automatic interpretations for the first seismic reflector on different seismic images. Each seismic reflector has a different spatial location that reflects the induced image uncertainty along the seismic migration process. Seismic reflector sample placement is driven by velocity uncertainty, characterized by Eq. (7). To illustrate the dispersion among the horizon samples, we use as a figure of merit the maximum distance in depth between horizons A and E, $\approx 60.0$ meters, as depicted in Fig. 20. Obviously, only five samples are not enough to elucidate the underlying probabilistic structure, but was chosen to achieve visual clearness of the dispersion. Figure 21 compares the seismic results obtained by the surrogate model to the RTM outcomes. Further, it presents the horizon interpretations on the seismic images results provided for both techniques. The surrogate model leads to accurate representations of the RTM seismic images. More importantly, the surrogate model can reproduce fine details of the images like spatial gradients required by automatic horizon extractors.

Seismic images computed with the surrogate allow reliable horizon extractions. In Fig. 22 we compare the horizon interpretations for the RTM and surrogate model outcomes intending to identify potential discrepancies. For improving visual comparison, we show only two pairs of horizons from the randomly selected seismic images. Most differences among the horizons are at the edges of the seismic reflectors, where the seismic signals are misplaced, and interpretations are intrinsically inaccurate. The automatic horizons relative error, evaluated by Eq. 12, is lower than 2%. The relative error considers only the vertical distance between the pair of horizons, and it is computed considering 120 points (see Fig. 22) equally spaced of around 6.7 meters.

5 Conclusions

We propose a deep learning model based on an encoder-decoder architecture to replace the costly RTM technique on producing seismic images. This approach naturally fits the framework of a computational workflow to produce seismic images with quantified uncertainty in [4]. This surrogate model builds a scalable image-to-image mapping, coping with the high dimensionality of both the heterogeneous velocity fields that serve as inputs and images outputs. Such surrogate has revealed to be very efficient in the context of UQ many-query tasks, as demonstrated by our numerical examples. Indeed, that was observed even in cases where we employ a non-optimal neural network architecture.

We place our contribution in the emerging area of physics-informed machine learning, where the final model, in many different ways, blends two main components: often expensive computational models relying on first principles and phenomenological closure equations, and machine learning data-driven tools. Such combination not only suits perfectly to the needs required by the workflow mentioned earlier but also offers a broad spectrum of opportunities to improve performance, like employing more powerful training strategies and automatic hyperparameters optimization.

Acknowledgements This study was financed in part by CAPES, Brasil Finance Code 001. This work is also partially supported by FAPERJ (grant E-26/203.018/2017), CNPq (grant 302489/2016-9), and Petrobras (grant 2017/00148-9).

References

1. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz,
14. Chang, W.F., McMechan, G.A.: Elastic reverse-time migration.
12. Cerjan, C., Kosloff, D., Kosloff, R., Reshef, M.: A nonreflecting
15. Chattopadhyay, S., McMechan, G.A.: Imaging conditions for
11. Brantut, N.: Time-resolved tomography using acoustic emissions
18. Diederik, P.K., Jimmy, B.: Adam: A method for stochastic
5. Belhadj, J., Romary, T., Gesret, A., Noble, M., Figliuzzi, B.:
8. Bilionis, I., Zabaras, N., Konomi, B.A., Lin, G.: Multi-output
6. Bergstra, J., Yamins, D., Cox, D.D.: Making a science of model
7. Bilionis, I., Zabaras, N.: Multi-output local gaussian process
52
Geophysics
12. Cerjan, C., Kosloff, D., Kosloff, R., Reshef, M.: A nonreflecting
5. Belhadj, J., Romary, T., Gesret, A., Noble, M., Figliuzzi, B.:
8. Bilionis, I., Zabaras, N., Konomi, B.A., Lin, G.: Multi-output
5. Belhadj, J., Romary, T., Gesret, A., Noble, M., Figliuzzi, B.:
10. Botero, A., Gesret, A., Romary, T., Noble, M., Maisons, C.:
31. Liu, F., Zhang, G., Morton, S.A., Leveille, J.P.: An effective
24. Huang, T., Zhang, Y, Zhang, H.: The benefit of tti reverse
22. Goodfellow, I., Bengio, Y , Courville, A.: Deep Learning. MIT
25. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep
26. Kahana, A., Turkel, E., Dekel, S., Givoli, D.: Obstacle segment-
27. Kearey, P., Brooks, M., Hill, I.: An Introduction to Geophysical
28. Kukreja, N., Huckleheim, J., Louboutin, M., Hou, K., Luporini,
29. Li, Y , Sun, J.: 3d magnetization inversion using fuzzy c-means
30. Lindstrom, P., Chen, P., Lee, E.J.: Reducing disk storage of full-
31. Liu, F., Zhang, G., Morton, S.A., Leveille, J.P.: An effective
32. Martin, J., Wilcox, L.C., Burstedde, C., Ghattas, O.: A stochastic
33. Liang, Z., Ye, X., Le, A., Mian, Q., Liu, Y.: An adaptive
34. Mo, S., Zhu, Y , Zabaras, N., Shi, X., Wu, J.: Deep convolutional
35. Mo, S., Zabaras, N., Shi, X., Wu, J.: Deep convolutional
36. Nguyen, B.D., McMechan, G.A.: Excitation amplitude imaging
37. Chollet, F.: Deep learning with python manning publications
38. Diederen, P.K., Jimmy, B.: Adam: A method for stochastic
19. Ely, G., Malcolm, A., Poliannikov, O.V.: Assessing uncertainties in
20. Gebraad, L., Boehm, C., Fichtner, A.: Bayesian elastic full-
21. Glorot, X., Bordes, A., Bengio, Y.: Deep sparse rectifier
22. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT
23. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely
24. Huang, T., Zhang, Y., Zhang, H.: The benefit of tti reverse
time migration for subsalt imaging, gulf of mexico. European
25. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep
26. Kahana, A., Turkel, E., Dekel, S., Givoli, D.: Obstacle seg-
27. Kearey, P., Brooks, M., Hill, I.: An Introduction to Geophysical
28. Kukreja, N., Huckleheim, J., Louboutin, M., Hou, K., Luporini,
29. Li, Y , Sun, J.: 3d magnetization inversion using fuzzy c-means
30. Lindstrom, P., Chen, P., Lee, E.J.: Reducing disk storage of full-
31. Liu, F., Zhang, G., Morton, S.A., Leveille, J.P.: An effective
32. Martin, J., Wilcox, L.C., Burstedde, C., Ghattas, O.: A stochastic
33. Liang, Z., Ye, X., Le, A., Mian, Q., Liu, Y.: An adaptive
34. Mo, S., Zhu, Y , Zabaras, N., Shi, X., Wu, J.: Deep convolutional
35. Mo, S., Zabaras, N., Shi, X., Wu, J.: Deep convolutional
36. Nguyen, B.D., McMechan, G.A.: Excitation amplitude imaging
37. Chollet, F.: Deep learning with python manning publications
38. Diederen, P.K., Jimmy, B.: Adam: A method for stochastic
19. Ely, G., Malcolm, A., Poliannikov, O.V.: Assessing uncertainties in
velocity models and images with a fast nonlinear uncertainty
quantification method. Geophysics 83, R63–R75 (2018)
20. Gebraad, L., Boehm, C., Fichtner, A.: Bayesian elastic full-
waveform inversion using hamiltonian monte carlo. EarthArxiv p.
qfitn5 (2019)
21. Glorot, X., Bordes, A., Bengio, Y.: Deep sparse rectifier
neural networks. In: Gordon, G., Dunson, D., Dudik, M. (eds.)
Proceedings of the Fourteenth International Conference on
Artificial Intelligence and Statistics, Proceedings of Machine
Learning Research, vol. 15, pp. 315–323. PMLR, Fort Lauderdale
(2011). http://proceedings.mlr.press/v15/glorot11a.html
22. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT
Press, Cambridge (2016). http://www.deeplearningbook.org
23. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely
connected convolutional networks. In: Proceedings of the IEEE
Conference on Computer Vision and Pattern Recognition (2017)
24. Huang, T., Zhang, Y., Zhang, H.: The benefit of tti reverse
time migration for subsalt imaging, gulf of mexico. European
Assoc Geosci Eng. https://doi.org/10.3997/2214-4609.201400392
(2009)
25. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep
network training by reducing internal covariate shift. In:
Proceedings of the 32nd International Conference on International
Conference on Machine Learning - vol. 37, ICML’15, p. 448–456.
JMLR.org (2015)
26. Kahana, A., Turkel, E., Dekel, S., Givoli, D.: Obstacle seg-
mentation based on the wave equation and deep learning. J.
Comput. Phys. 413, 109458 (2020). https://doi.org/10.1016/j.jcp.
2020.109458. http://www.sciencedirect.com/science/article/pii/
S0021999120302321
27. Kearey, P., Brooks, M., Hill, I.: An Introduction to Geophysical
Exploration, Wiley, Hoboken (2013)
28. Kukreja, N., Huckleheim, J., Louboutin, M., Hou, K., Luporini,
F., Hovland, P., Gorman, G.: Combining checkpointing and data
compression for large scale seismic inversion. arXiv:1810.05268
(2018)
29. Li, Y , Sun, J.: 3d magnetization inversion using fuzzy c-means
clustering with application to geology differentiation. Geophysics
81, J61–J78 (2016)
30. Lindstrom, P., Chen, P., Lee, E.J.: Reducing disk storage of full-
3d seismic waveform tomography (f3dt) through lossy online
compression. Comput Geosci 93, 45–54 (2016)
31. Liu, F., Zhang, G., Morton, S.A., Leveille, J.P.: An effective
imaging condition for reverse-time migration using wavefield
decomposition. Geophysics 76(1), S29–S39 (2011)
32. Martin, J., Wilcox, L.C., Burstedde, C., Ghattas, O.: A stochastic
newton mcmc method for large-scale statistical inverse problems
with application to seismic inversion. SIAM J. Sci. Comput. 34(3),
A1460–A1487 (2012)
33. MATLAB: Version 8.6.0 (R2015b). The MathWorks Inc. Natick,
Massachusetts (2015)
34. Mo, S., Zabaras, N., Shi, X., Wu, J.: Integration of adversarial
autoencoders with residual dense convolutional networks for es-
35. Mo, S., Zhu, Y , Zabaras, N., Shi, X., Wu, J.: Deep convolutional
encoder-decoder networks for uncertainty quantification of
dynamic multiphase flow in heterogeneous media. Water Resour.
Res. 55, 703–728 (2019)
36. Nguyen, B.D., McMechan, G.A.: Excitation amplitude imaging
condition for prestack reverse-time migration. Geophysics 78(1),
S37–S46 (2013)
37. Parussini, L., Venturi, D., Perdikaris, P., Karniadakis, G.: Multi-fidelity gaussian process regression for prediction of random fields. J. Comput. Phys. 337, 36–50 (2017)
38. Pau, G.S.H., Zhang, Y., Finsterle, S.: Reduced order models for many-query subsurface flow applications. Comput. Geosci. 17, 705–721 (2013)
39. Perdikaris, P., Venturi, D., Karniadakis, G.E.: Multi-fidelity information fusion algorithms for high-dimensional systems and massive data sets. SIAM J. Sci. Comput. 38, B521–B538 (2016)
40. Pilikos, G., Faul, A.C.: Bayesian modeling for uncertainty quantification in seismic compressive sensing. Geophysics 84(2), P15–P25 (2019)
41. Poliannikov, O.V., Malcolm, A.E.: The effect of velocity uncertainty on migrated reflectors: Improvements from relative-depth imaging. Geophysics 81, S21–S29 (2016)
42. Rasmussen, C., Williams, C.: Gaussian processes for machine learning (2006)
43. Ricker, N.: The form and laws of propagation of seismic wavelets. Geophysics 18(1), 10–40 (1953). https://doi.org/10.1190/1.1437843
44. Sava, P., Fomel, S.: Least-squares horizons with local slopes and multigrid correlations. Geophysics 83(4), IM29–IM40 (2018)
45. Sava, P., Vasconcelos, I.: Extended imaging conditions for wave-equation migration. Geophys. Prospect. 59(1), 35–55 (2011)
46. Schuster, G.T.: Seismic inversion, 1st edn, Society Exploration Geophysicists (2017)
47. Siahkoohi, A., Rizzuti, G., Herrmann, F.J.: Uncertainty quantification in imaging and automatic horizon tracking: A bayesian deep-prior based approach. arXiv:2004.00227 (2020)
48. Strikwerda, J.C.: Finite difference schemes and partial differential equations. In: SIAM. 2nd edn. (2004). Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9780898717938
49. Sun, J., Niu, Z., Innanen, K.A., Li, J., Trad, D.O.: A theory-guided deep-learning formulation and optimization of seismic waveform inversion. Geophysics 85(2), R87–R99 (2020)
50. Tang, M., Liu, Y.J., durlufsky, L.: A deep-learning-based surrogate model for data assimilation in dynamic subsurface flow problems. J. Comput. Phys. 413, 109456 (2020)
51. Wang, B., Mason, C., Guo, M., Yoon, K., Cai, J., Ji, J., Li, Z.: Subsalt velocity update and composite imaging using reverse-time-migration based delayed-imaging-time scan. Geophysics 74(6), WCA159–WCA166 (2009)
52. Weisberg, S.: Applied Linear Regression. Wiley, New York (2005)
53. Witte, P.A., Louboutin, M., Luporini, F., Gorman, G.J., Herrmann, F.J.: Compressive least-squares migration with on-the-fly fourier transforms. Geophysics 84(5), R655–R672 (2019)
54. Wu, X., Fomel, S.: Least-squares horizons with local slopes and multigrid correlations. Geophysics 83(4), IM29–IM40 (2018)
55. Yang, Y., Perdikaris, P.: Conditional deep surrogate models for stochastic, high-dimensional, and multi-fidelity systems. Comput. Mech. 64, 417–434 (2019)
56. Yilmaz, ¨O.: Seismic data analysis: processing, inversion, and interpretation of seismic data. Society of exploration geophysicists (2001)
57. Zand, T., Siahkoohi, H.R., Malcolm, A., Gholami, A., Richardson, A.: Consensus optimization of total variation–based reverse time migration. Comput. Geosci. https://doi.org/10.1007/s10596-020-09958-1 (2020)
58. Zhao, Z., Sen, M.K.: A gradient based Mcmc method for Fwi and uncertainty analysis. In: SEG Technical Program Expanded Abstracts 2019, pp. 1465–1469 (2019). Society of Exploration Geophysicists
59. Zhou, H.W., Hu, H., Zou, Z., Wo, Y., Youn, O.: Reverse time migration: a prospect of seismic imaging methodology. Earth Sci. Rev. 179, 207–227 (2018)
60. Zhu, H., Li, S., Fomel, S., Studlenger, G., Ghatts, O.: A bayesian approach to estimate uncertainty for full-waveform inversion using a priori information from depth migration. Geophysics 81(5), R307–R323 (2016)
61. Zhu, Y., Zabaras, N.: Bayesian deep convolutional encoder-decoder networks for surrogate modeling and uncertainty quantification. J. Comput. Phys. 366, 415–447 (2018)
62. Zhu, Y., Zabaras, N., Koutsourelakis, P., Perdikaris, P.: Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data. J. Comput. Phys. 394, 56–81 (2019)

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.