Bulletin of Geophysics and Oceanography Vol. 63, n. 2, pp. 189-214; June 2022
DOI 10.4430/bgo00382

A Convolutional Neural Network-Monte Carlo approach for petrophysical seismic inversion

M. Aleardi and C. De Biasi

Earth Sciences Department, University of Pisa, Italy

(Received: 23 September 2021; accepted: 3 December 2021; published online: 22 March 2022)

ABSTRACT We implement a machine-learning inversion approach to infer petrophysical rock properties from pre-stack data that combines a convolutional neural network and a discrete cosine transform of data and model spaces. This transformation is used for model and data compression. The network learns the inverse mapping between the compressed seismic and the compressed petrophysical domain. A theoretical rock-physics model relates elastic and petrophysical properties, while the exact Zoeppritz equations map the elastic properties onto the seismic domain. Training and validation examples are generated under the assumption of a Gaussian variogram model and a non-parametric prior. A Monte Carlo simulation strategy is employed for uncertainty assessment. We present synthetic inversions on a realistic subsurface model and the outcomes of the proposed approach are compared with those achieved by a standard linearised inversion. The network predictions are assessed in case of errors in the calibrated rock-physic model, in the estimated source wavelet, and in the assumed noise statistics. We also demonstrate that transfer learning avoids retraining the network from scratch when the target and training properties differ. Our experiments confirm that the implemented inversion successfully solves the petrophysical seismic inversion, opening the possibility to get instantaneous predictions of reservoir properties and related uncertainties.

Key words: seismic inversion, reservoir characterisation, neural network.

1. Introduction

Pre-stack seismic inversion is extensively used to infer petrophysical reservoir properties (e.g. clay content, porosity, water saturation), and to obtain a geometric description of the investigated subsurface area (Doyen, 2007; Avseth *et al*., 2010). Such petrophysical seismic inversion is often solved through a two-step cascaded workflow: first elastic properties (i.e. P-, S-wave velocities, and density) are usually estimated from pre-stack data through a linearised inversion in which a linearisation of the full Zoeppritz equations is the forward operator (Buland and Omre, 2003). Then, a properly calibrated Rock-Physics Model [RPM (Bachrach, 2006; Avseth *et al*., 2010; Rimstad and Omre, 2010; Bosch *et al*., 2015; Aleardi and Ciabarri, 2017; Mavko *et al*., 2020)] estimated from well logs and/or core data is used to convert the estimated elastic characteristics to the sought petrophysical parameters. This two-step approach has several drawbacks, for example, it is not guaranteed that the estimated petrophysical model is fully consistent with the observed seismic data (Doyen, 2007). To overcome this issue a single-step approach has been proposed in which the elastic and petrophysical properties are simultaneously inferred from

pre-stack gathers (Bosch *et al*., 2007; Grana and Della Rossa, 2010; Aleardi and Mazzotti, 2014; Aleardi *et al*., 2018).

Whatever approach is adopted, the petrophysical seismic inversion remains a nonlinear and ill-posed problem in which multiple solutions equally reproduce the observed data. The nonlinearity may arise from the RPM relation that links the petrophysical and the elastic parameters. The ill-conditioning is related to the inherent mathematical properties of the RPM or of the seismic forward operator and is further aggravated by noise contamination in the data, limited data coverage, low data resolution, uncertainties in the estimated RPM and source wavelet. For this reason, geostatistical hard and soft model constraints are usually infused into the inversion framework to reduce the number of possible solutions (Doyen, 1988; Mukerji *et al*., 2001; Dubrule, 2003; Azevedo and Soares, 2017; Azevedo *et al*., 2019), and a probabilistic Bayesian framework (Tarantola, 2005) is usually adopted for accurate uncertainty assessments (Fjeldstad and Grana, 2018; Fjeldstad and Omre, 2019; Aleardi and Salusti, 2020). However, the inclusion of spatial constraints usually increases the computational demand of the inversion procedure, and hence common approaches invert each seismic gather independently thus overlooking the spatial connectivity pattern of the subsurface model. Another difficulty of petrophysical seismic inversion arises from the complex distribution of rock properties in the subsurface. Although a Gaussian prior allows an analytical and computationally fast derivation of predicted model and associated uncertainties (Tarantola, 2005), it is known that such a probabilistic model oversimplifies the actual distribution of reservoir parameters because it neglects the influence exerted by the litho-fluid facies. From a mathematical point of view, this means that a Gaussian model does not capture the multimodality of the actual distribution of petrophysical parameters and hence multimodal priors are often employed [i.e. Gaussian-mixture model, or even non-parametric priors (de Figueiredo *et al*., 2019; Grana, 2020)].

Over the last years, machine learning algorithms have been extensively used to solve geophysical inverse problems. In this context, neural networks are useful when the forward relation is known, but the inverse mapping is either expensive to compute analytically or numerically approximate. Therefore, the network aims to approximate the inverse mapping between the seismic space and the petrophysical domain. Many different types of machine learning algorithms exist such as Artificial Neural Networks (ANNs) and Convolutional Neural Networks (CNNs) (Bishop, 2006; Goodfellow *et al*., 2016; Géron, 2019). ANNs are theoretically capable of learning any nonlinear function linking the output and the input response. Some drawbacks of ANNs are the computational cost, the high number of learnable parameters, the vanishing/exploding gradient problems, and the overfitting issue. Besides, in cases of image applications, ANNs lose the spatial relationships between the pixels of the image and cannot capture sequential information in the data. CNN uses convolutional kernels to extract the relevant features from the input and, hence, they are commonly used in image classification problems. Differently from ANN, CNN captures the spatial variability pattern from an image and exploits parameter sharing to decrease the number of internal parameters to optimise. In brief, compared to conventional fully connected networks, CNNs are regularised networks with two advantages: sparse connectivity and sharing weights among convolutional layers, which reduce the computational cost and improve the generalisation ability (LeCun *et al*., 2015; Schmidhuber, 2015; Krizhevsky *et al*., 2017). Training a CNN usually needs a sufficiently large training set to iteratively refine and update the internal network parameters. After training, the network converts an input into the corresponding output in real-time. In recent years, CNNs have been successfully employed to solve geophysical problems such as seismic full-waveform inversion (Richardson, 2018; Wu and McMechan, 2019; Yang and Ma, 2019), velocity estimations from

seismic data (Araya-Polo *et al*., 2018; Park and Sacchi, 2020), data interpolation and noise attenuation (Wang *et al*., 2019; Sun *et al*., 2020), seismic interpretation (Waldeland *et al*., 2018), impedance inversion (Das *et al*., 2019), elastic inversion (Biswas *et al*., 2019; Aleardi, 2020a; Aleardi and Salusti, 2021), electromagnetic inversion (Puzyrev, 2019; Moghadas, 2020). One well-known drawback of machine-learning methods is that their performances degrade if the statistical properties of the target and training data differ. To overcome this problem transfer learning can be used (Li *et al*., 2020; Park and Sacchi, 2020) in which a small portion of new training data are used to adjust the internal weights of the previously trained network.

Here the petrophysical seismic inversion is solved through a trained CNN, and the porosity, clay content, and water saturation are estimated from pre-stack seismic data. We employ the Discrete Cosine Transform [DCT (Britanak *et al*., 2010; Lochbühler *et al*., 2014] reparameterisation to reduce data and model spaces. This transformation also guarantees the preservation of the assumed spatial and temporal variabilities in the retrieved solution. Indeed, the order of the retained non-zero coefficients determines the wavelength of the recovered model along the spatial and temporal directions. In this context, the network is trained to predict the nonlinear mapping between the compressed seismic data, and the compressed petrophysical model, which constitute the input and output responses of the network, respectively. We combine the implemented CNN inversion with a Monte Carlo (MC) simulation to propagate onto the final estimates both the uncertainties related to noise contamination and the modelling error introduced by the CNN. Indeed, the trained network introduces additional uncertainties related to the learned approximated mapping between the seismic domain and the petrophysical space. For the lack of available field data, we describe synthetic inversions in which the test model simulates a realistic geological setting where a gas-saturated reservoir is hosted in a turbiditic sequence. To assess the applicability and robustness of the presented approach, we perform several inversion tests that realistically simulate errors in the estimated wavelet, in the assumed noise distribution, and in the calibrated RPM. The CNN predictions are also benchmarked against the results of a more standard linearised inversion that uses a petrophysical reformulation of the Zoeppritz equation to directly infer the petrophysical properties from the seismic data (Aleardi *et al*., 2017). To maintain the computational cost affordably, this method inverts each seismic gather separately.

The proposed machine-learning inversion approach includes four steps. 1) Generation phase: define an ensemble of petrophysical models and compute the associated pre-stack responses. The Direct Sequential Co-Simulation method with joint probability distribution (Horta and Soares, 2010) is employed to impose spatial, temporal, and mutual dependencies of the petrophysical properties in the training and validation sets. This geostatistical method considers a non-parametric prior that properly models the facies dependency of the petrophysical parameters. The prior is derived from available well log data and is used to draw 5000 petrophysical models forming the training set. A theoretical RPM based on granular media models (Mavko *et al*., 2020) constitutes the link between the petrophysical and the elastic parameters (Aleardi *et al*., 2018), whereas a convolutional forward operator based on the full Zoeppritz equations provides the nonlinear mapping between the model and the data space. 2) Network design: define the CNN architecture to convert the DCT compressed seismic response into the DCT compressed petrophysical model. 3) Training phase: train the network by minimising the differences between the predicted and desired output. 4) Model prediction and MC error propagation. Fig. 1 schematically represents the proposed CNN inversion framework that infers the petrophysical parameters from the observed data.

We show that the CNN inversion provides more accurate predictions of the petrophysical

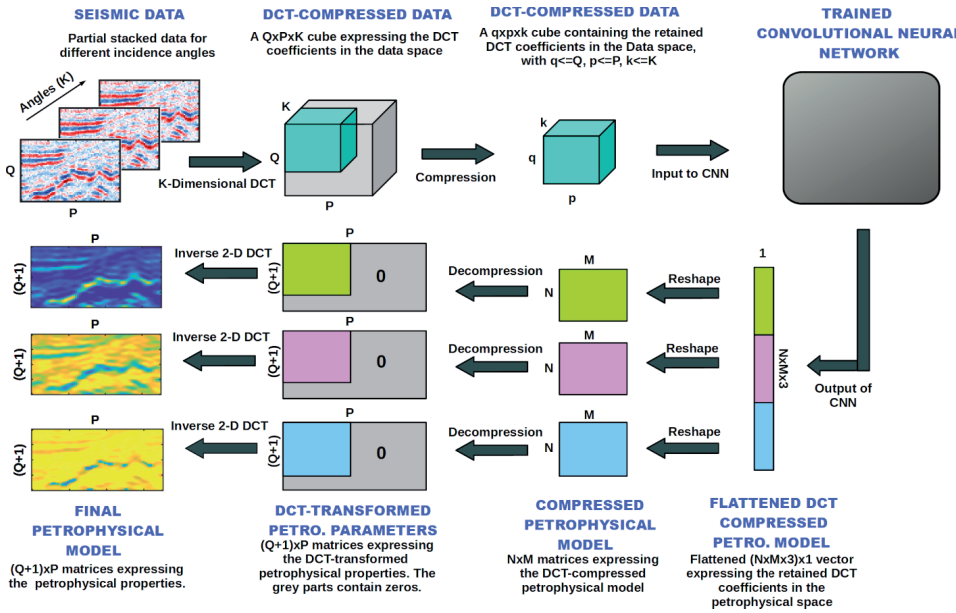

Fig. 1 - Schematic workflow of the implemented CNN inversion.

properties than the standard approach (SA). We also show that the proposed machine-learning approach is quite robust against errors in the estimated source wavelet, in the assumed noise statistics, and in the estimated RPM. We will also illustrate that transfer learning can be employed to adjust the learnable network parameters when the prior petrophysical assumptions or the RPM of the target differ from those assumed during the learning stage. The novelty of our contribution is the combination of DCT, CNN, and a MC simulation to probabilistically solve the petrophysical seismic inversion.

2. Methods

2.1. Discrete cosine transform

In this section, we briefly introduce the DCT compression used to reduce the model and data domains. Additional information about this popular compression strategy and its application to solve geophysical inversions can be found in Britanak *et al*. (2010), Lochbühler *et al*. (2014), and Aleardi (2020b). In particular, we employ this strategy because it exhibits superior compression power over other linear compression methods (Lochbühler *et al*., 2014).

The basis functions employed by the DCT are cosine functions oscillating at different frequencies. This transformation can be applied to mono or multidimensional signals. For example for a 2-D density model $\rho(x, y)$ with $x = [1, \ldots, M_x]$ and $y = [1, \ldots, M_y]$ the transformation can be computed as follows:

$$\mathbf{R} = \mathbf{B}_y \boldsymbol{\rho} \mathbf{B}_x^T, \tag{1}$$

where $\mathbf{B}_x$ and $\mathbf{B}_y$ are the matrices with dimensions M_x x M_x and M_y x M_y, expressing the basis

functions, whereas the M_y x M_x matrix **R** contains the DCT coefficients. The DCT concentrates most of the information of the original signal into the low-order coefficients, and, hence, an approximation of the 2-D density model can be computed as follows:

$$\widetilde{\boldsymbol{\rho}} = \left(\mathbf{B}_y^q\right)^T \mathbf{R}_{qp} \mathbf{B}_x^p, \tag{2}$$

where $\widetilde{\boldsymbol{\rho}}$ is the approximated model, $\mathbf{B}_y^q$ is a [q x M_y] matrix with only the first q rows of $\mathbf{B}_y$; $\mathbf{B}_x^p$ is a [p x M_x] matrix with only the first p rows of $\mathbf{B}_x$; the matrix $\mathbf{R}_{qp}$ represents the first q rows and p columns of **R**. The q and p values are the retained number of basis functions along the y and x directions. Therefore, the DCT transformation reduces the full (M_y x M_x)-D model space to a (q x p)-D DCT-compressed domain (with $p < M_x$ and $q < M_y$). In this context the p x q non-zero numerical coefficients of the $\mathbf{R}_{qp}$ matrix becomes the inverted parameters in the compressed space. Fig. 2 illustrates examples of basis functions of different orders in a 2-D space. Note that the coefficient orders determine the variability along the vertical and horizontal directions.

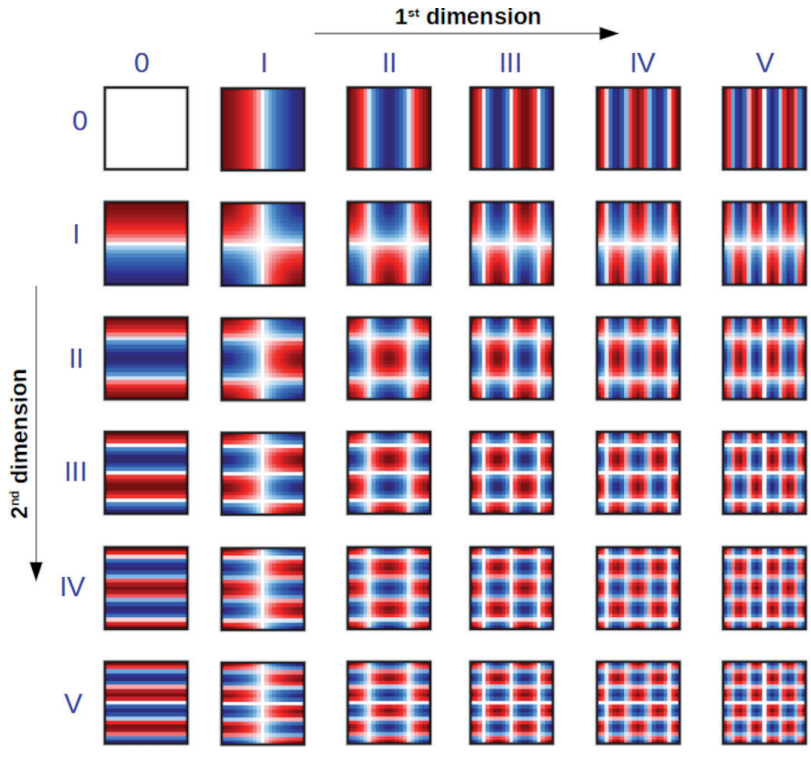

Fig. 2 - 2-D DCT basis functions of different orders. Blue and red codes indicate negative and positive values, respectively.

2.2. Convolutional neural network and MC error propagation

We employ a CNN to learn the mapping between the DCT compressed partially stacked data and the compressed 2-D petrophysical model. The CNN can be expressed as a function F that computes the output **O** from the input **I** through the internal parameters **W**:

$$\mathbf{O} = F(\mathbf{W}, \mathbf{I}). \tag{3}$$

Generally speaking, a CNN uses convolutional filters and fully connected layers to extract features from 1-D, 2-D, or 3-D inputs. The filters are called 'kernels' that slide over the input with a specified stride, whereas the extracted features are called the features maps. This means that each pixel of the input matrix is multiplied by the corresponding value in the kernel [Hadamard product (Laloy *et al*., 2017)]. The number of filters in each convolutional layer, their size, and the stride are user-defined parameters. To include nonlinearity in the mapping process, the ouput of a given layer is also passed through an activaction function before being fed into the next layer. Typical activation functions are the Rectified Linear Unit (ReLU) or the Leaky ReLU (Puzyrev, 2019). Once the features of the input are extracted by convolutional layers, they are flattened to a 1-D vector and fed into the fully connected layer. The learnable CNN parameters are first initialised and successively updated during the learning procedure, in which an error function quantifying the difference between desired and computed outputs is minimised. The updating process is driven by a back-propagation algorithm, whereas the length of the updating step is proportional to the selected learning rate value.

Several hyperparameters define the CNN architecture: number of convolutional layers and filters, activation function, stride and size of the convolutional filters, strategy to initialise the weights, number of epochs, learning rate value, method that minimises the error function. Personal preferences and experiences usually determine the hyperparameter setting. Here we employ a trial-and-error procedure in which different hyperparameters are modified and the final net architecture (presented in the following sections) has been determined according to its accuracy on the validation set.

For what concerns the error propagation, we must project onto the model space both the noise affecting the data and the so-called modelling error introduced by the CNN approximation. To this end we adopt a MC approach. Let **M** denote the models forming the training set, whereas **N** represents the associated CNN predictions. According to Hansen and Cordua (2017), a sample of the modelling error can be computed as **E = M - N**. Under a Gaussian assumption, the modelling error can be written as $\mathcal{N}(0, \mathbf{C}_e)$, where $\mathbf{C}_e$ represents the covariance of **E**, and $\mathcal{N}$ is the Gaussian distribution. We also assume Gaussian noise in the data $\mathcal{N}(0, \mathbf{C}_n)$, and an iterative MC approach is used to propagate both the modelling and the noise errors in the final CNN predictions:

1) use the trained network to compute the predicted model $\mathbf{m}_b$ from the observed data **d**;
2) use the calibrated RPM to compute the elastic properties associated to $\mathbf{m}_b$ and use the Zoeppritz equations to derive the associated seismic response $\mathbf{d}_b$;
3) draw **n** from $\mathcal{N}(0, \mathbf{C}_n)$ and compute $\mathbf{d}_n = \mathbf{d}_b + \mathbf{n}$;
4) use the trained CNN to derive $\mathbf{m}_n$ from $\mathbf{d}_n$;
5) draw **e** from $\mathcal{N}(0, \mathbf{C}_e)$ and compute $\mathbf{m}_e = \mathbf{m}_n + \mathbf{e}$;
6) store $\mathbf{m}_e$ and iterate from 3) to 6) for Q times.

Each vector $\mathbf{m}_e$ is a petrophysical model in agreement with the trained CNN, the assumed noise and modelling errors, and the observed data. An approximated uncertainty quantification can be numerically derived from the Q MC simulations. For simplicity, we assume Gaussian-distributed noise and modelling errors, but the MC approach can handle whatever parametric or non-parametric error distribution. Note that the MC approach is extremely fast because the network instantaneously predicts a model from the input data.

To reduce the computational cost of the MC simulation, we derive the modelling error in the compressed domain and considering all the model parameters simultaneously. This means that for qxp retained coefficients per petrophysical property, we get a $\mathbf{C}_e$ matrix with qxpx3 rows and

columns. This matrix correctly codes the lateral, vertical, and mutual parameter correlations characterising the modelling error. Compared to other uncertainty propagation approaches proposed in machine learning applications [e.g. the use of dropout in the prediction phase (Das and Mukerji, 2020)], the MC approach can easily manage different sources of uncertainties (e.g. errors in the RPM, errors in the estimated source wavelet). For example, errors in the calibrated RPM can be added to the elastic parameters associated with the predicted petrophysical model before computing the $\mathbf{d}_b$ vector, and hence steps 1 and 2 have to be included in the iterative MC procedure. As an alternative, the different sources of errors and uncertainties (noise in the data, imperfection in the forward model) can also be included in the training ensemble and accounted for in the prediction phase (Pradhan and Mukerji, 2020).

3. Results

3.1. Generating and compressing the training and validations sets

For the lack of available field data, we focus on synthetic inversions in which the seismic data set is generated on a 2-D test model, derived by integrating a geologic interpretation with information from five wells. This model represents a gas-saturated clastic reservoir within a turbiditic sequence (Figs. 3a to 3c). We consider the three petrophysical properties of porosity, clay content, and water saturation and three partial angle stacks at incidence angles of 0°, 15°, and 30° (i.e. near, mid, and far stack, respectively). This configuration translates into 50×71×3 = 10650 data points and a full petrophysical space composed of 51×71×3 = 10863 unknown parameters.

Before generating the training and validation examples, the expected statistical properties of the petrophysical parameters and the RPM must be defined. The petrophysical prior assumptions are derived from the available well data, previously used to build the test model. The kernel density estimation algorithm (Parzen, 1962) with a Gaussian kernel has been applied to derive the non-parametric prior (Figs. 3d to 3f). We also assume a stationary 2-D Gaussian variogram model in which the lateral and vertical ranges have been inferred from the lateral variability of the observed seismic data and the vertical variability of the upscaled well log data, respectively (Figs. 3g and 3h). The ranges of the variogram are equal to 160 m and 0.008 s along the spatial and temporal directions, respectively. We consider a stationary variogram and prior model and this assumption simplifies the actual distribution of the petrophysical properties in the quite complex test model. The direct-sequential co-simulation with joint probability distribution has been used to generate 6000 2-D elastic models from the prior. We employ a theoretical RPM based on granular media models (i.e. Hertz-Mindlin contact theory, Hashin-Shtrikman lower bound, and Gassmann model). Figs. 3i to 3m show the rock-physics template associated with the considered RPM: we observe a decrease of P-wave impedance and an increase of *Vp*/*Vs* ratio as the water saturation, and clay content decreases and as the porosity increases. A convolutional forward operator based on the exact Zoeppritz equations provides the observed seismic gathers according to a 30-Hz, zero-phase Ricker wavelet. We contaminate the observed data with uncorrelated Gaussian noise with a standard deviation corresponding to the 25% of the standard deviation of the noise-free seismic gathers. We generate 6000 petrophysical models and associated seismic data: 5000 out of the 6000 models will be used to train the network while the remaining constitute the validation set. Table 1 lists the information defining the training and validation examples. All the considered petrophysical properties are normalised in the range [0, 1], thus ensuring similar variances for all the parameters and better performances during the learning process.

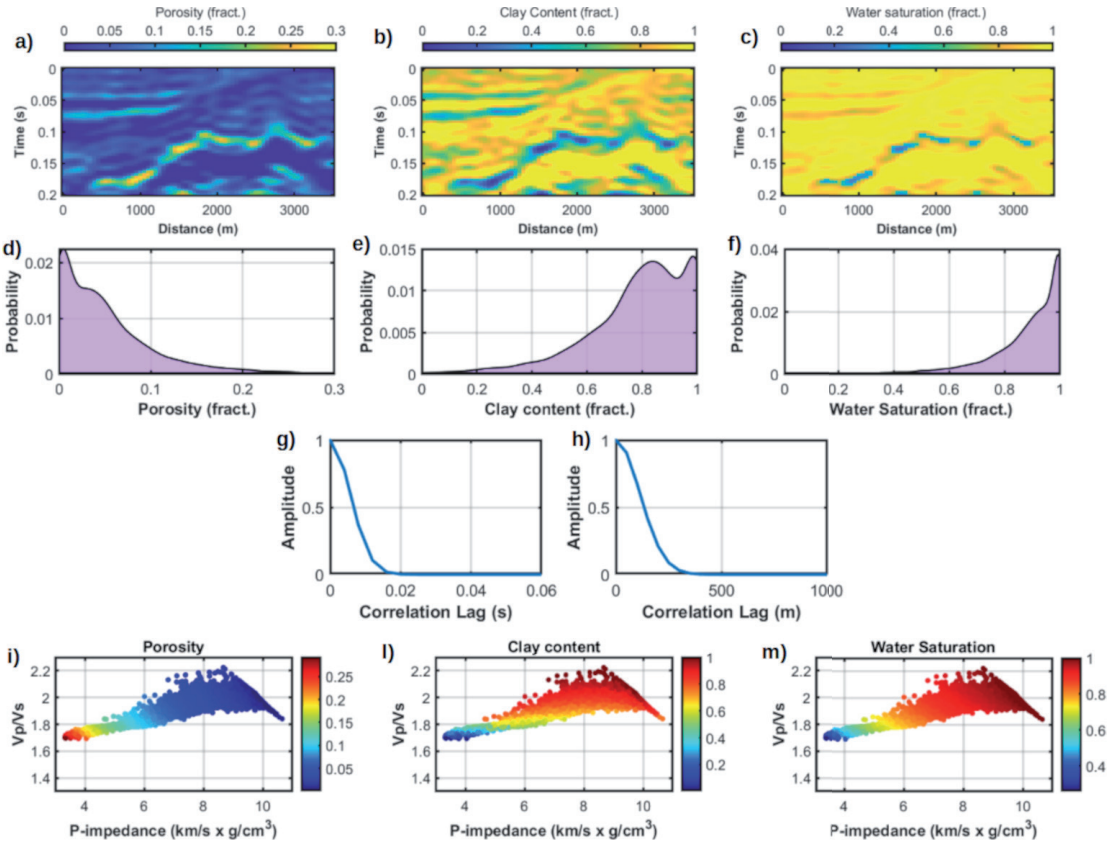

Fig. 3 - Syntetic inversions: a) to c) the petrophysical properties of the test model; d) to f) the marginal prior distribution for the three petrophysical parameters; g) to h) temporal and spatial correlograms associated with the assumed Gaussian variogram model; i) to m) rock-physics template generated from the considered RPM and showing the influence of the three petrophysical parameters on the elastic properties.

Table 1 - Characteristics of the observed data and of the petrophysical prior model assumed in the learning phase.

Seismic data generation		**Petrophysical model generation**	
Standard deviation of uncorrelated noise compared to the standard deviation of the noise-free data	25%	Type of prior distribution	Non-parametric
Correlated noise standard deviation	0	Vertical range of the Gaussian variogram model	0.008 s
Source wavelet phase	0	Lateral range of the Gaussian variogram model	160 m
Source wavelet peak frequency	30		

Before training the network, we must define the number of retained DCT coefficients that approximate the petrophysical model. To this end, we quantify how the explained variability of prior simulations changes as the number of DCT coefficients changes. This selection is a delicate step because the retained coefficients should guarantee data fitting and model resolution comparable to those achieved by an inversion running in the uncompressed domain. A detailed discussion on this topic can be found in Grana *et al*. (2019). Fig. 4 shows the explained variability

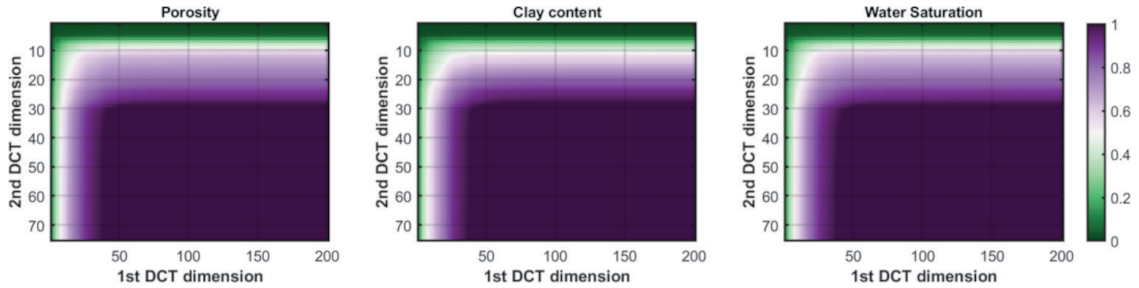

Fig. 4 - Examples of explained model variability for a petrophysical model extracted from the prior and as the number of DCT coefficients along the 1st and 2nd DCT dimension increases. In each plot, the numerical value with coordinate (x, y) indicates the explained variability if the first x, and y DCT coefficients along the 1st and 2nd DCT dimensions, respectively, are used for model compression. It emerges that 30 DCT coefficients along both DCT dimensions explain almost the 100% of the variability of the uncompressed model.

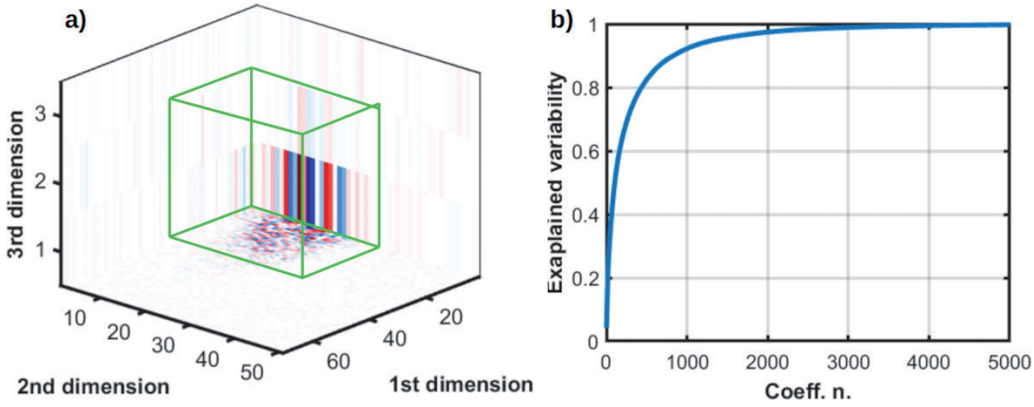

Fig. 5 - Analysis carried out on some seismic gathers generated from prior realisations: a) DCT decomposition of a seismic gather computed on a petrophysical model drawn from the prior. Blue and red colours code low and high values, respectively, while the green rectangles enclose the retained DCT coefficients in the data space; b) explained data variability as the number of considered basis functions increases.

for a petrophysical model drawn from the prior when the number of coefficients changes. It emerges that 30 coefficients per property (q = p = 30) along the two DCT dimensions recover almost the total variability of the original petrophysical model. Therefore, the DCT compression reduces the 10863-D full petrophysical space to 30×30×3 = 2700-D domain. A similar analysis has been carried out on some seismic gathers generated from prior realisations. The green rectangle in Fig. 5a encloses the 35×40×2 = 2800 retained coefficients in the data space that explains almost the total variability of the uncompressed seismic gather (Fig. 5b). Therefore, the full 10650-D data space can be conveniently reduced to a 2800-D domain. The compression of the data and model space significantly reduces the number of unknown parameters and also decreases the number of training examples needed for the learning phase (see the discussion section). Moreover, the use of the DCT significantly reduces the complexity of the CNN architecture (i.e. number of convolutional filters and hidden layers) needed to approximate the inverse operator (Aleardi and Salusti, 2021). This translates into a faster learning phase and an easier hyperparameter setting. Finally, the compression guarantees the preservation of the assumed vertical and lateral variability patterns in the recovered model.

3.2. The selected network architecture

As previously mentioned, the learning phase employs training and validation sets of 5000, and 1000 models, respectively. This means that the input for the training is constituted by a 35×40×2×5000 tensor with 5000 being the number of training examples, and 35×40×2 = 2800 being the number of the retained DCT coefficients in the data domain. The network output includes 30×30×3×5000 coefficients where 30×30×3 = 2700 are the retained basis functions in the petrophysical domain (900 coefficients per petrophysical parameter). The finally selected network (Fig. 6) includes two convolutional blocks and a fully connected layer. The convolutional blocks employ 5 filters of size 3×3, and a stride of 1, the Leaky ReLU with a leakage value of 0.1 is used as the activation function, whereas batch normalisation is employed as a regularisation operator (Santurkar *et al*., 2018). Before the fully connected layer, the dropout regularisation is also used to avoid overfitting. Dropout is a methodology where randomly selected neurons are ignored during the training phase (i.e. in our case the 20%). We employ the He method to initialise the network weights (He *et al*., 2015) and the Adam optimiser (Balles and Hennig, 2018) to minimise the root-mean-square-error (RMSE) between predicted and desired outputs. We employ a batch size of 24, whereas 0.001 is the initial learning rate that is scaled by 0.9 every epoch. Fig. 7 represents the evolution of the RMSE during the learning process. We observe that convergence is achieved in only four epochs, whereas the similar RMSE on the training and validation indicates that overfitting has been prevented. For additional information about the computational cost of the learning stage see the discussion section.

Fig. 8 shows box plots representing the correlation between the true models forming the training and validation sets and the corresponding CNN predictions. The similar correlation coefficients obtained on both validation and training prove the generalisation capability of the trained network and that the overfitting has been avoided. As expected, the quality of the results decreases moving from porosity to clay content, and water saturation in agreement with the different sensitivity of seismic data to petrophysical parameters. Indeed, it is known that the porosity is the parameter that exerts the major influence on determining the elastic properties and, then, the reflected seismic amplitudes, while *Vp*, *Vs,* and density are poorly affected by

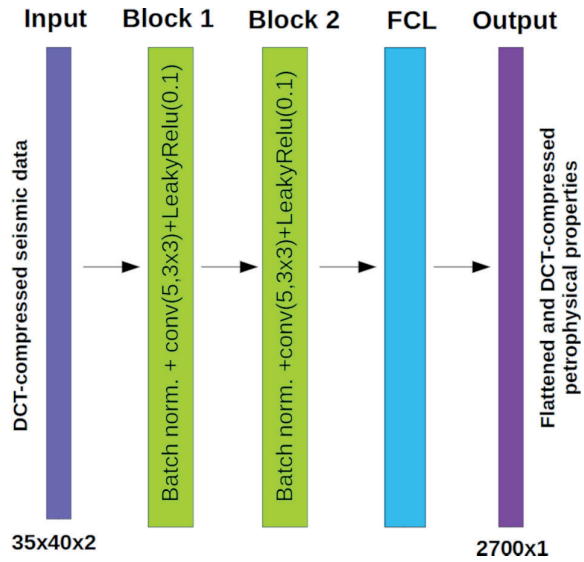

Fig. 6 - The employed network architecture. The image on the far left and far right represent the input and output of the network. The green rectangles represent the convolutional layers with annotated key parameters: the initial value in brackets (e.g. 5) indicates the number of filters, which is followed by the filter size (i.e. 3×3). The cyan rectangle represents the fully connected layer (FCL).

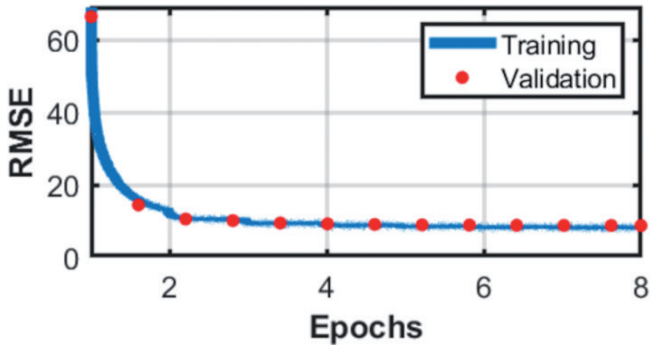

Fig. 7 - Evolution of the RMSE value during training.

water saturation changes. The direct comparison of the network predictions with the associated true models from the validation set (Fig. 9) proves the reliability of the CNN inversion. However, these results refer to petrophysical models that honour both the calibrated RPM and the statistical assumptions made during the learning phase. Therefore, in the following, we discuss the application of the CNN inversion to the more realistic test model shown in Figs. 3a to 3c. The CNN predictions are also benchmarked against those provided by a more standard linearised petrophysical seismic inversion that assumes Gaussian distributed petrophysical parameters and that inverts each seismic gather independently. This method includes the RPM within the Zoeppritz equations to directly infer petrophysical properties from the seismic response (Aleardi *et al*., 2017), while the Gauss-Newton approach minimises an error function defined as a linear combination of L2 norm data misfit and L2 norm difference between predicted and *a-priori* petrophysical model. Hereafter, we refer to this inversion as the SA. The Gaussian prior for the SA inversion has been extracted from the same well log data previously used. For the sake of comparison, all the following tests assume for the SA approach the same noise statistics used in the generation of the training examples. This method also uses the same RPM employed by the CNN inversion and imposes the same vertical variogram to the recovered model. We will also evaluate the stability of the CNN predictions when realistic conditions are simulated: erroneous assumptions about the statistical properties of the noise, errors in the estimated source wavelet, uncertainties in the calibrated RPM. In the final part, an application of transfer learning is discussed.

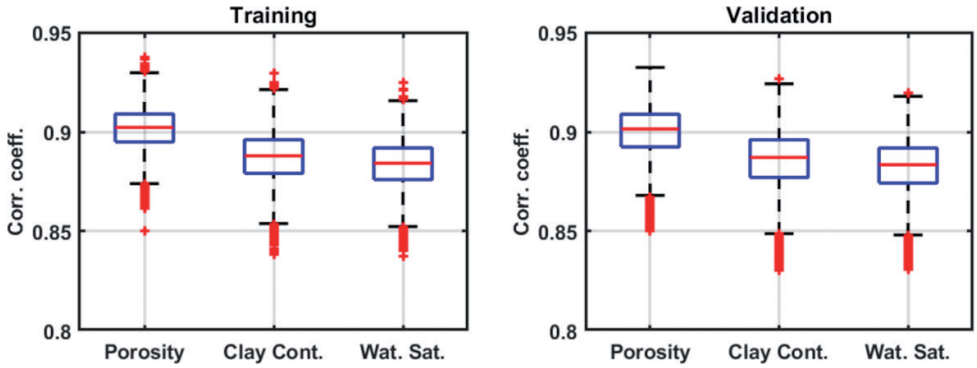

Fig. 8 - Box plot showing the correlation coefficients between actual and predicted models forming the training and validation set (left and right plot, respectively). The red crosses represent outliers.

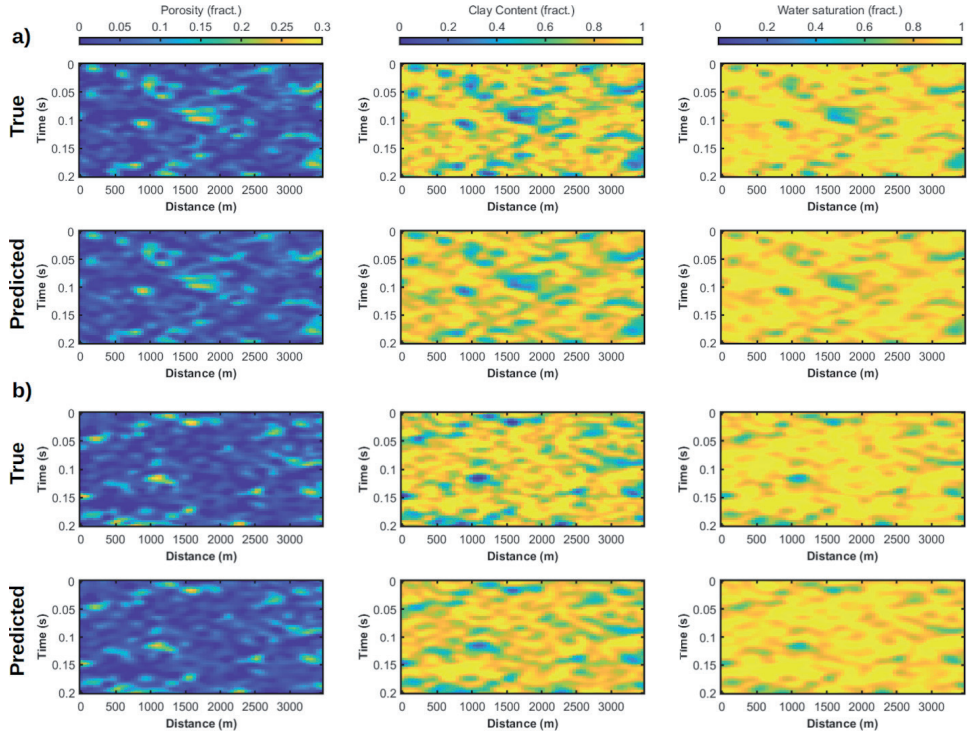

Fig. 9 - Panels a and b compare two petrophysical models extracted from the validation set and the associated CNN predictions. From left to right we represent porosity, clay content, and water saturation.

3.3. Application to the test model

We perform five different inversion experiments in which the CNN and SA predictions are compared. The main characteristics of these tests are listed in Table 2 and discussed hereafter:

Test 1: the source wavelet and the standard deviation of the noise contaminating the observed data (computed on the test model: Figs. 3a to 3c) are the same assumed during the training phase (see Table 1). This test assesses the network performances in the most optimal case of perfectly-known noise properties, RPM, and source wavelet.

Test 2: the standard deviation value of the Gaussian and uncorrelated noise affecting the observed data is twice that assumed for training. This test illustrates the network performances when the actual noise contamination is underestimated.

Test 3: we simulate both uncorrelated and spatially-temporarily correlated Gaussian noise in the observed data. The noise standard deviation values are the same assumed in Test 2. The temporal and spatial variability pattern of the coherent noise is equal to the assumed variogram model. In real data applications, coherent noise can be ascribed to diffractions and/or multiple reflections not successfully attenuated by the processing sequence.

Test 4: similar to Test 3 but we also add errors in the estimated source wavelet. This means that the test data are computed according to a source wavelet different from that assumed in the learning phase. We include errors in the peak amplitude, peak frequency, and phase: the peak frequency is 25 Hz, the phase is 30°, and we simulate a 20% error in the peak amplitude (see Fig. 10).

Test 5: similar to Test 4 but now errors in the estimated RPM are also simulated. This means that the RPM used to project the petrophysical test model onto the elastic space differs from that used

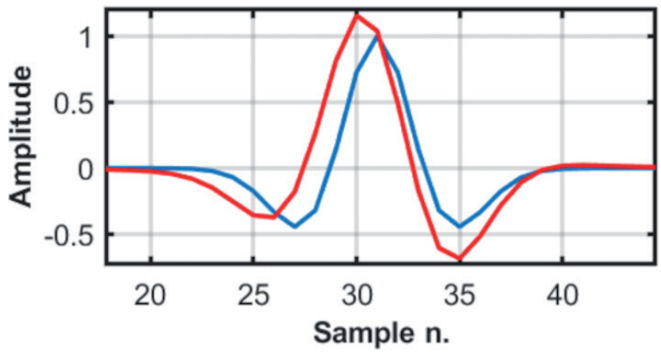

Fig. 10 - Comparison between the wavelet used to compute the observed data in tests 4 and 5 (red curve), and that used to generate the training examples (blue curve).

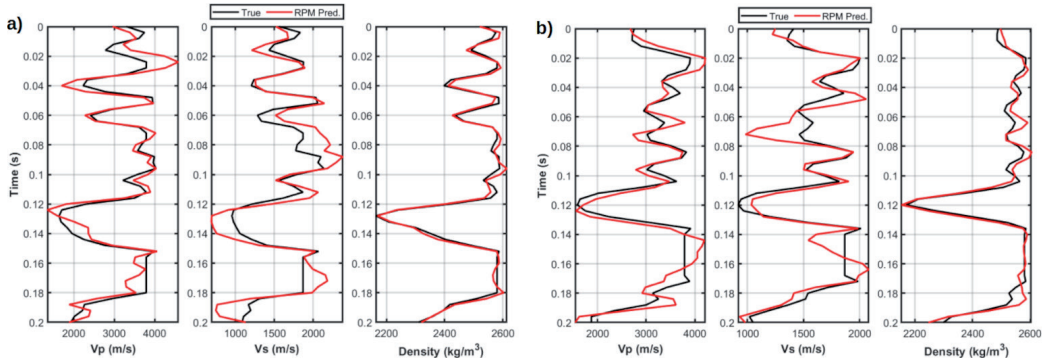

Fig. 11 - Comparison at two different horizontal coordinates along the test model, between the elastic properties computed when the RPM is perfectly known (black curves) and those derived when errors in the RPM calibration are simulated (red curves).

in the learning phase. Realistically, we assume that the RPM errors increase moving from density to P-wave velocity (*Vp*) and S-wave velocity (*Vs*). Indeed, when calibrating an RPM, it is known that the density is the simplest parameter to fit, while it is difficult to achieve good *Vs* reproductions due to the theoretical limitations of granular media models (Sajeva *et al.*, 2019). Fig. 11 shows, for two horizontal coordinates along the test model, a comparison between the elastic properties derived from the petrophysical parameters according to the correct RPM (the one used in the learning phase) and those obtained after introducing errors in the calibrated rock-physics relation.

Table 2 - Characteristics of the five inversion experiments.

Parameters	Training	#1	#2	#3	#4	#5
Uncorrelated noise standard deviation	25%	25%	50%	50%	50%	50%
Correlated noise standard deviation	0%	0%	0%	50%	50%	50%
RPM error	No	No	No	No	No	Yes
Source wavelet phase	0°	0°	0°	0°	30°	30°
Source wavelet peak frequency	30 Hz	30 Hz	30 Hz	30 Hz	25 Hz	25 Hz
Vertical range of correlated noise	/	/	/	8 ms	8 ms	8 ms
Lateral range of correlated noise	/	/	/	160 m	160 m	160 m

In all tests, the correlation coefficients and the Structural Similarity Index (SSI) are used to compare the CNN and SA predictions with the ground truth (Wang *et al*., 2004).

In Test 1 both CNN and SA give similar predictions in good agreement with the test model (Figs. 12a and 13a). However, we observe some lateral scattering in the SA solution that can be ascribed to noise in the data. On the contrary, the DCT parameterisation better preserves the continuity in the CNN estimates. The lateral scattering in the SA predictions severely increases if we simulate an underestimation of the noise (Figs. 12b and 13b). This underprediction is expected to generate overfitting with the observed seismic. Biased petrophysical predictions are obtained when coherent noise is added to the observed data (Figs. 12c and 13c). Errors in the source wavelet increase the bias in the final estimates (Figs. 12d and 13d) that are now affected by either overpredicted or underpredicted petrophysical contrasts. The simultaneous presence of coherent noise in the data and errors in both the source wavelet and in the calibrated RPM significantly decrease the quality of the estimated petrophysical models (Figs. 12e and 13e). However, from just a visual inspection of Figs. 12 and 13, we observe that in all the tests the CNN inversion better recovers the actual variability of the petrophysical properties in the true model, while also preserving the vertical and lateral continuity. Fig. 14 compares the lateral and vertical correlation functions assumed in the learning stage with those computed on the network prediction of Fig. 12a. It should be noted that the DCT compression preserves the assumed continuity patterns in the estimated model.

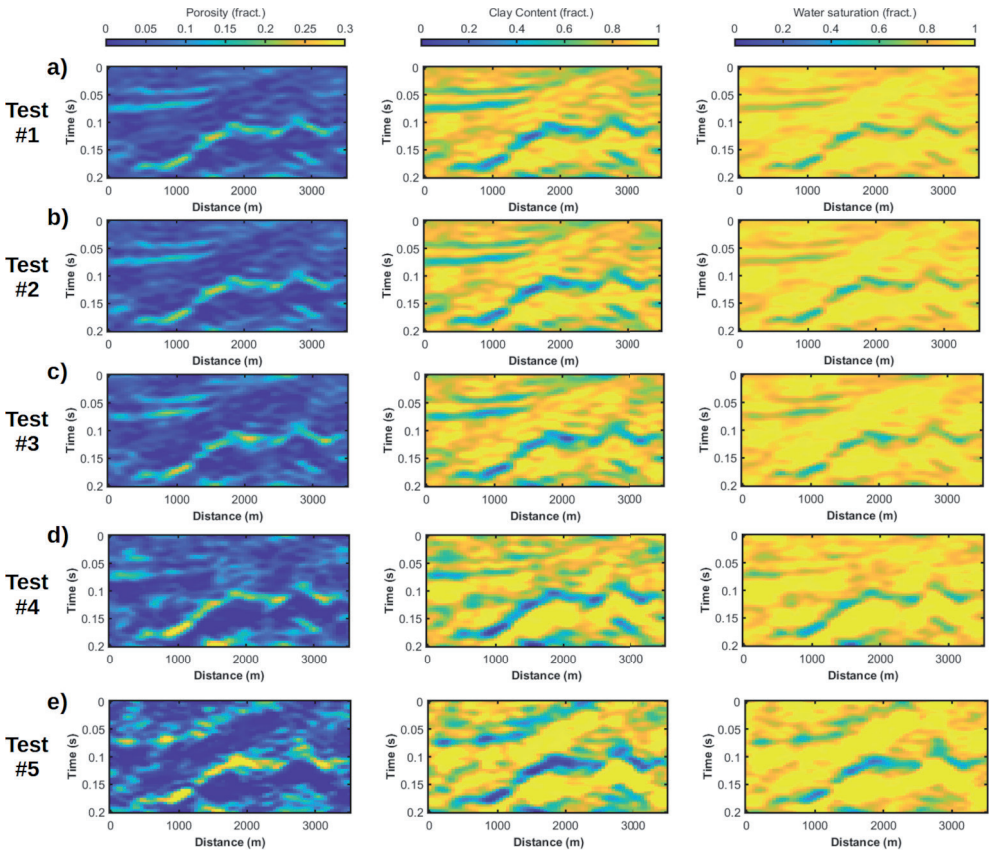

Fig. 12 - CNN predictions for the different tests. Porosity, clay content, and water saturation are represented from left, to right. The true model is shown in Figs. 3a to 3c. Table 2 lists the details of the five inversion tests.

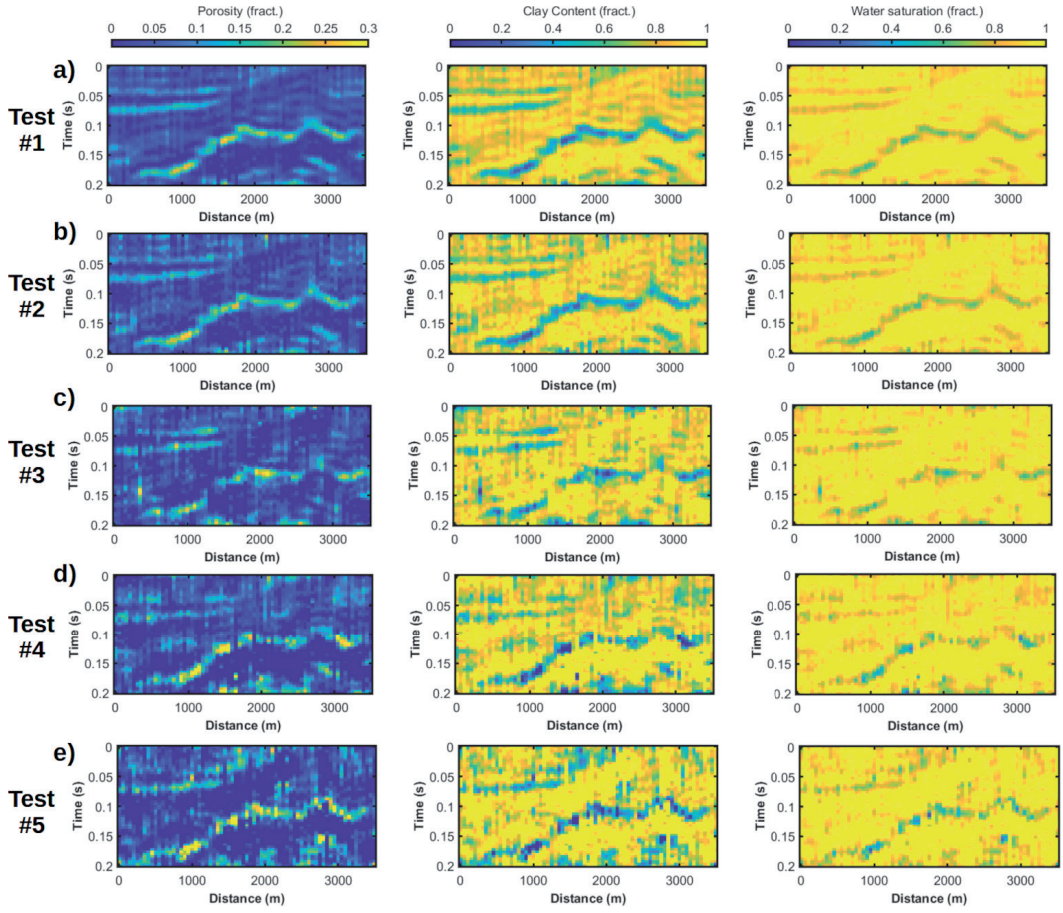

Fig. 13 - SA inversion results for the different tests. Table 2 lists the details of the five inversion tests.

Fig. 15 provides a more quantitative assessment of the CNN and SA predictions and represents the normalised cross-correlation and the SSI values obtained in the five inversion tests and for all the petrophysical properties. The quality of the CNN and SA results is similar for Test 1, while for all the other experiments the CNN outperforms the SA inversion. As expected, the quality of the predictions decreases moving from porosity to clay content and water saturation. This analysis illustrates that the combined use of DCT compression and CNN provides final estimates less affected by errors in the assumed noise model, in the calibrated RPM, and in the estimated source wavelet.

Figs. 16 to 19 compare for tests 3 and 5 the observed seismic gathers and the data computed on the CNN and SA solutions. For Test 3 we get L2 norm data misfit values equal to 6.2 and 5.8 for the CNN and SA inversion, respectively. Therefore, the data difference appears slightly lower for the SA example. However, we point out that the SA prediction corresponds to the minimum of an error function, while the CNN solution is not directly driven by a minimisation process. Differently, for Test 5 the poor prediction provided by the SA approach results in a larger data misfit compared with the CNN inversion (L2 norm data misfit values equal to 19.3 and 21.2 for CNN and SA, respectively).

Figs. 20a to 20c illustrate three MC petrophysical realisations generated for Test 3. As expected, the variability of the realisations increases moving from porosity to clay content

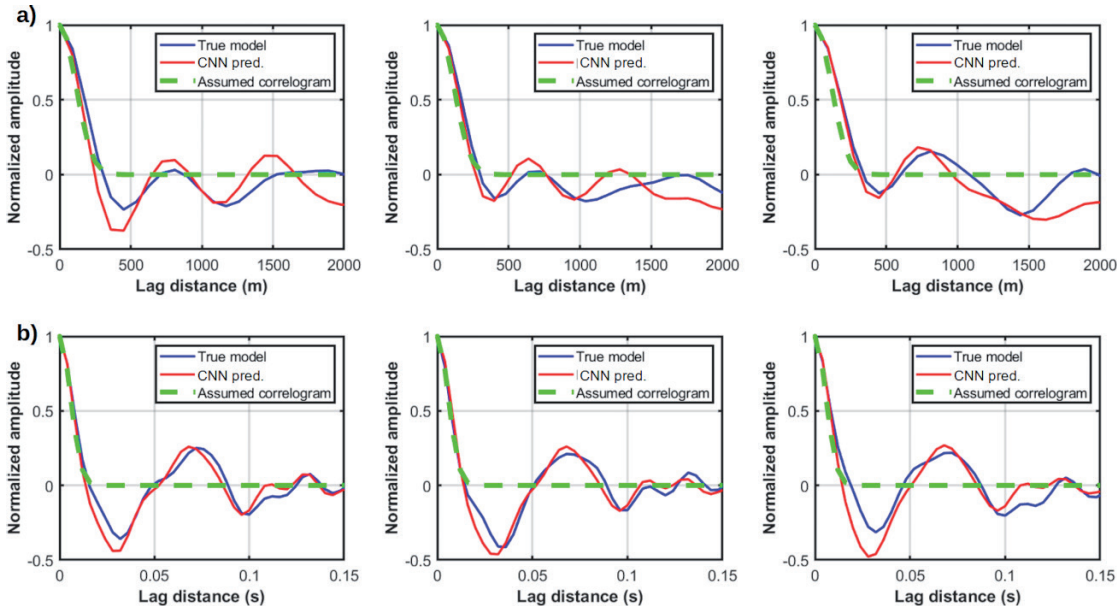

Fig. 14 - Comparison between the assumed lateral (a) and vertical (b) correlograms and the average correlograms derived from the true model (blue lines) and the CNN result (red lines). From left to right we represent porosity, clay content, and water saturation.

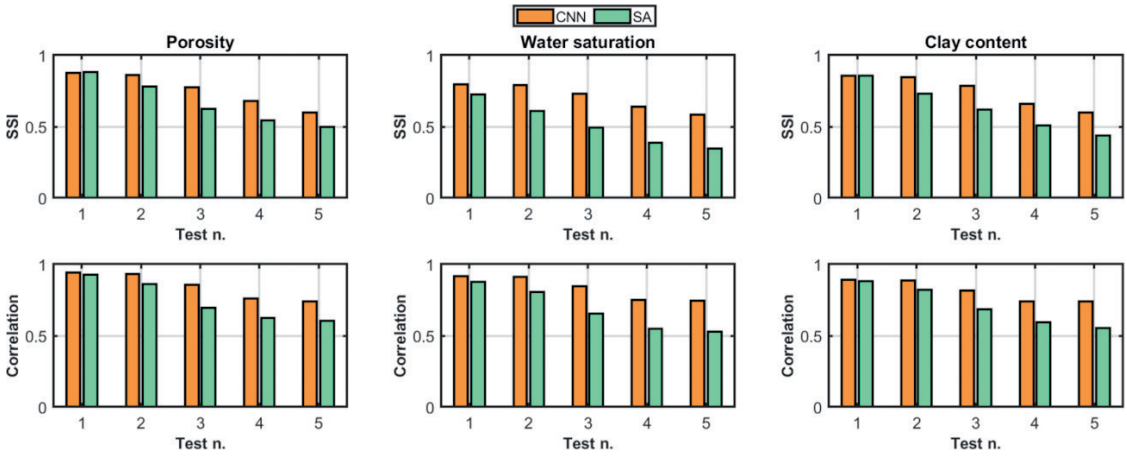

Fig. 15 - SSI and correlation coefficients computed for the different tests and for the CNN and SA results.

and water saturation. Fig. 20d shows, for the same test, the model standard deviation for the three petrophysical properties computed from 1000 MC simulations. The nonlinearity of both the Zoeppritz equations and the RPM produces uncertainties that vary over the model space: usually higher for sand and lower for shale. In a standard inversion approach, this would indicate that the curvature of the error function changes over the model space. We also observe that, as expected, the porosity is the best-resolved parameter, while the water saturation shows higher uncertainties. Finally, Fig. 21 shows for the horizontal coordinate of 1500 m, a direct

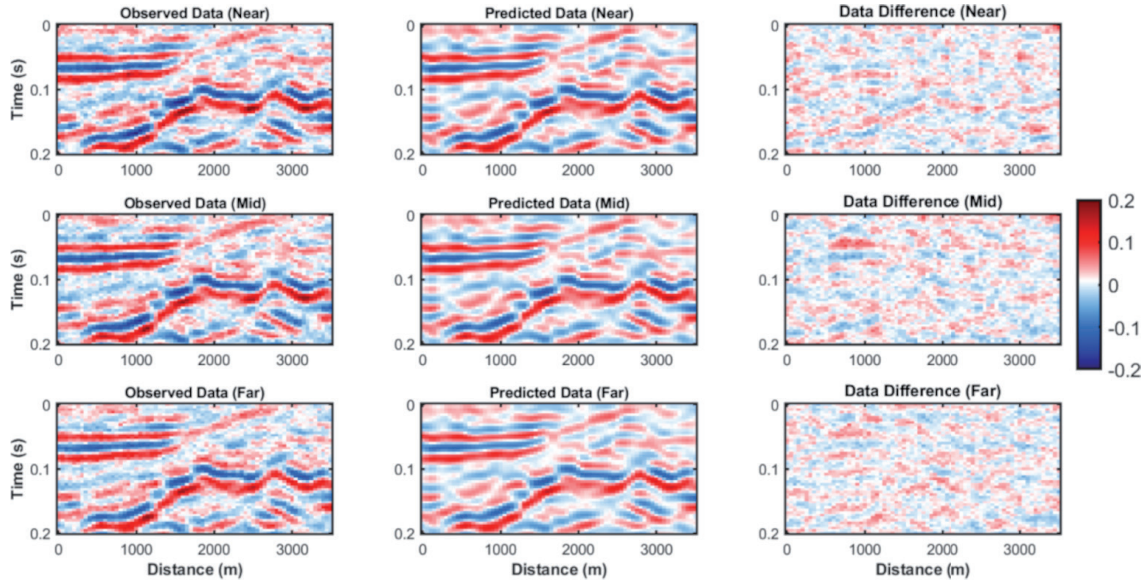

Fig. 16 - Observed data (left column), predicted data (central column), and their difference (right column). This comparison refers to Test 3 and CNN predictions.

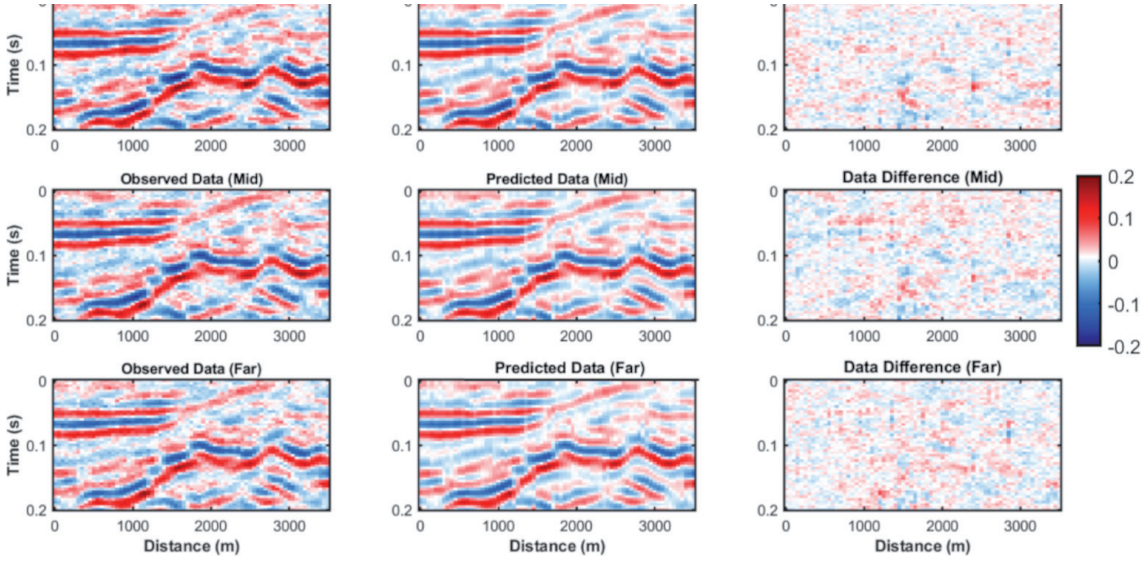

Fig. 17 - As in Fig. 16 but for the SA approach.

comparison between the true petrophysical profiles, 1000 MC simulations, and the associated 95% confidence interval. Even, in this unfavorable scenario with underpredicted and correlated noise in the data, the 95% confidence interval always encloses the true petrophysical values.

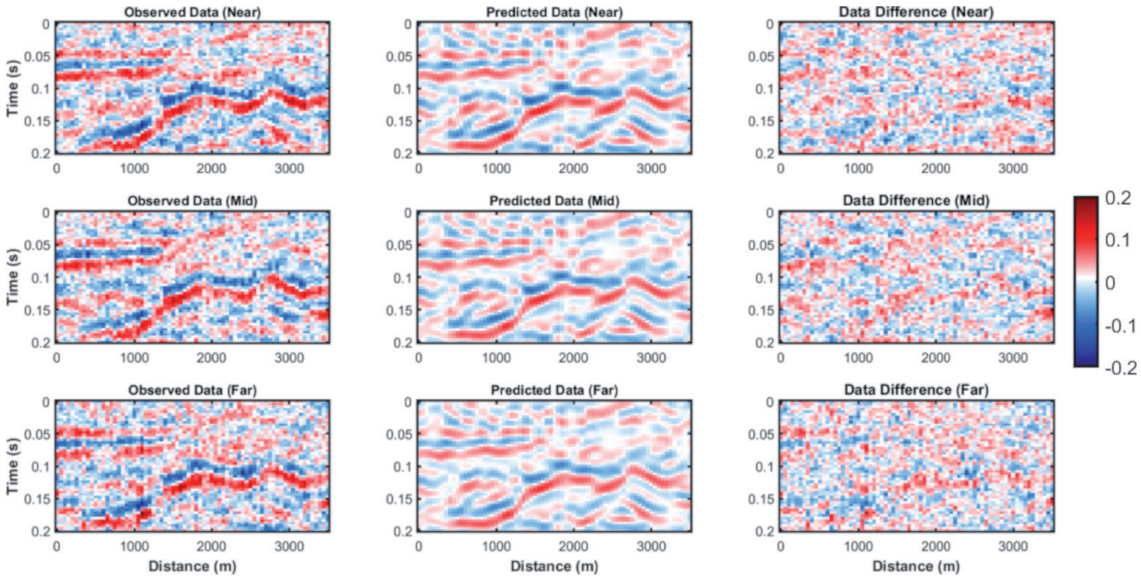

Fig. 18 - Observed data (left column), predicted data (central column), and their difference (right column). This comparison refers to Test 5 and CNN predictions.

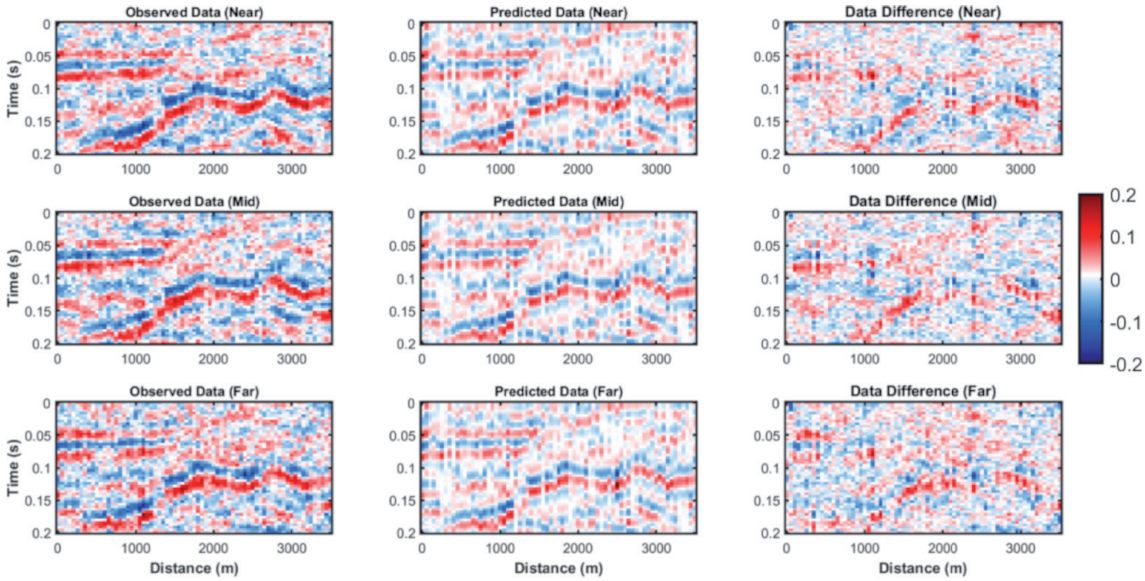

Fig. 19 - As in Fig. 18 but for the SA approach.

3.4. Transfer learning

We now apply the trained network on a modified version of the test model and employing a different RPM. In more detail, the modified test set has been derived by a 50% increasing of the standard deviation of the petrophysical properties of the original model. In addition, we now adopt a theoretical RPM that is better suited for a deeper reservoir with higher P-impedance values and lower *Vp*/*Vs* ratios. The modified prior distribution and the rock-physics template generated by the modified RPM are shown in Fig. 22.

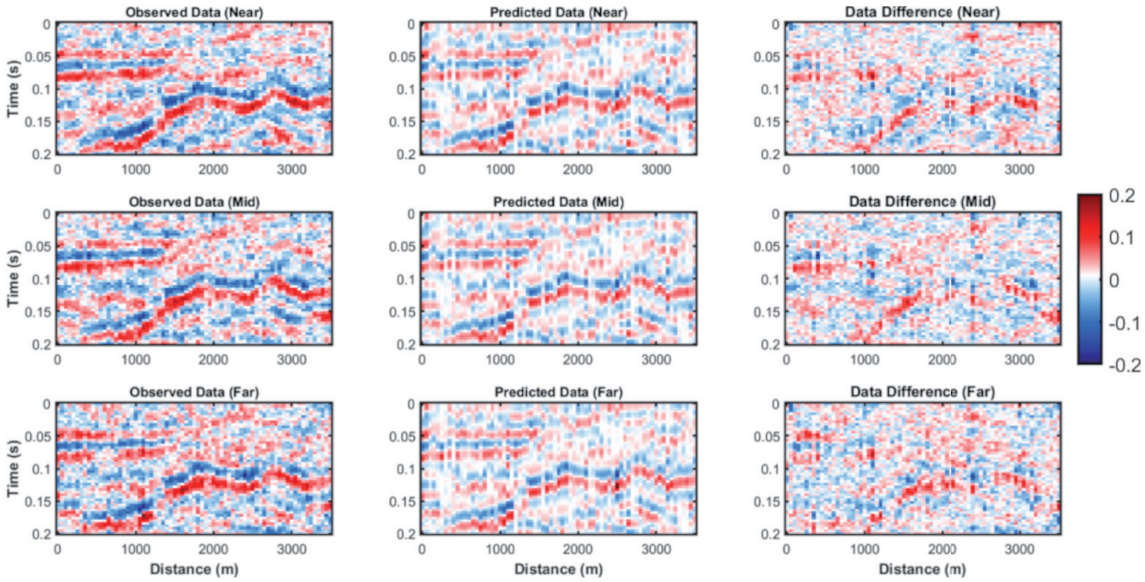

Fig. 20 - Three MC simulations of porosity (a), clay content (b), and water saturation (c) associated to Test 3, respectively. Standard deviation models (d) estimated from 1000 MC simulations for the same test (porosity, clay content, and water saturation are represented from left to right).

Figs. 23a and 23b illustrate the new test model and the CNN results obtained with the previously trained network when the noise model and the source wavelet are perfectly known. We observe that the geological structure of the reservoir is satisfactorily preserved, but the actual petrophysical contrasts at the reflecting interfaces are significantly underpredicted.

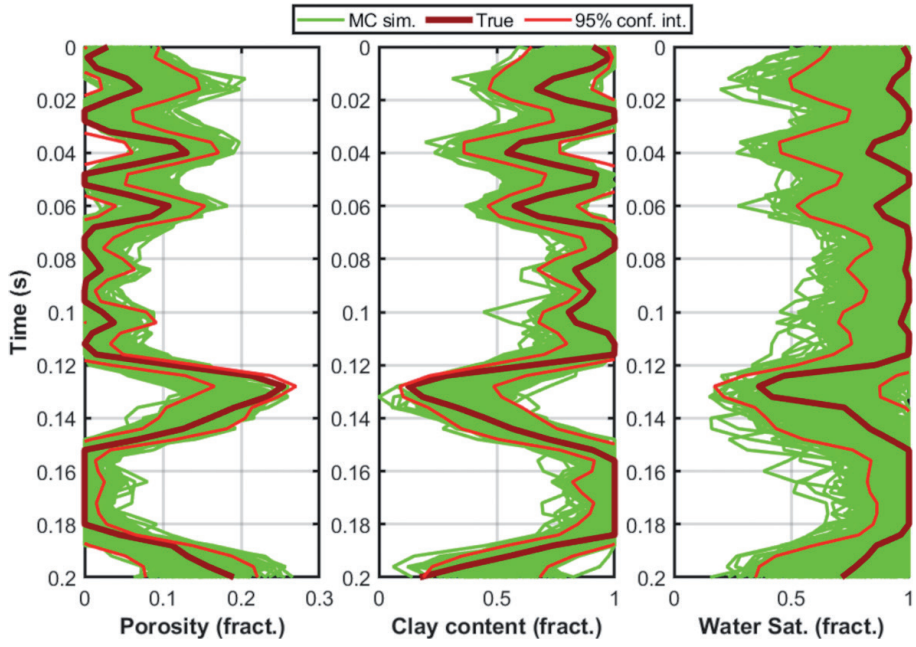

Fig. 21 - Comparison between the true model (brown curves), 1000 MC simulations (green lines), and the associated 95% confidence interval (red lines). This comparison refers to the horizontal coordinate of 1500 m.

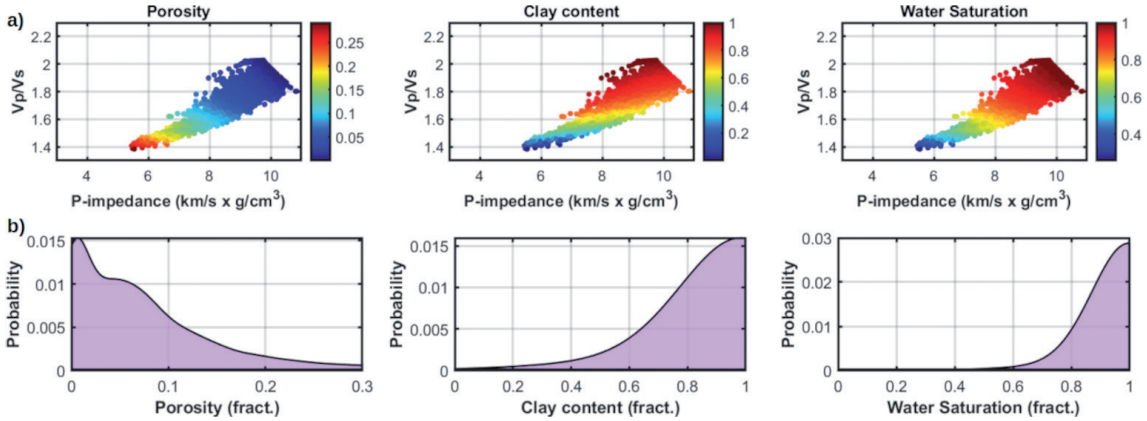

Fig. 22 - Results obtained by the modified RPM: a) the rock-physics template; b) the modified marginal prior for the three petrophysical parameters. Compare with Fig. 3.

Transfer learning avoids retraining the network from scratch when applied to a different data set, associated with different prior assumptions and RPM equation (Park and Sacchi, 2020; Sun *et al*., 2020). This technique uses a new, small training set to adjust the internal weights of a previously trained network. Transfer learning needs a faster training phase than a network with randomly initialised internal parameters. Before applying transfer learning, we must establish which part of the already trained network must be adjusted and this choice is dictated by the difference between the target and the training set; only the internal parameters of the fully connected layers have to be adjusted if this difference is small. Otherwise, the entire pre-trained CNN model must be modified (Huot *et al*., 2018; Park and Sacchi 2020). Therefore, we repeat the previous CNN inversion after the application of transfer learning. We employ the modified prior and RPM to generate new 500 training and 50 validation examples. Some tests (not shown here for brevity) suggested only for an update of the internal parameters of the fully-connected layer. Fig. 23c demonstrates that transfer learning improves the network results especially for the unknowns better informed by the data (porosity and clay content). Fig. 24 displays the evolution of the RMS on the validation and training sets during transfer learning. Note that the final RMSE value is very similar to that previously shown in Fig. 7. This example demonstrates that transfer learning successfully adjusts the network weights when training and target properties differ. However, accurate RPM and prior model estimations are still crucial ingredients to generate new training examples that exhaustively represent the target properties and their relation with the observed seismic response.

4. Discussion

On the one hand, the inclusion of spatial constraints into petrophysical seismic inversion is often discouraged by the increased computational cost of the estimation procedure. On the other hand, it is known that machine-learning approaches usually require a very large training set to appropriately learn the relation linking the input to the output response. Having this in mind, the regression ability of CNN and the compression ability of the DCT have been exploited to implement an alternative petrophysical seismic inversion characterised by a limited computational cost and also capable to preserve the assumed spatial and temporal continuity

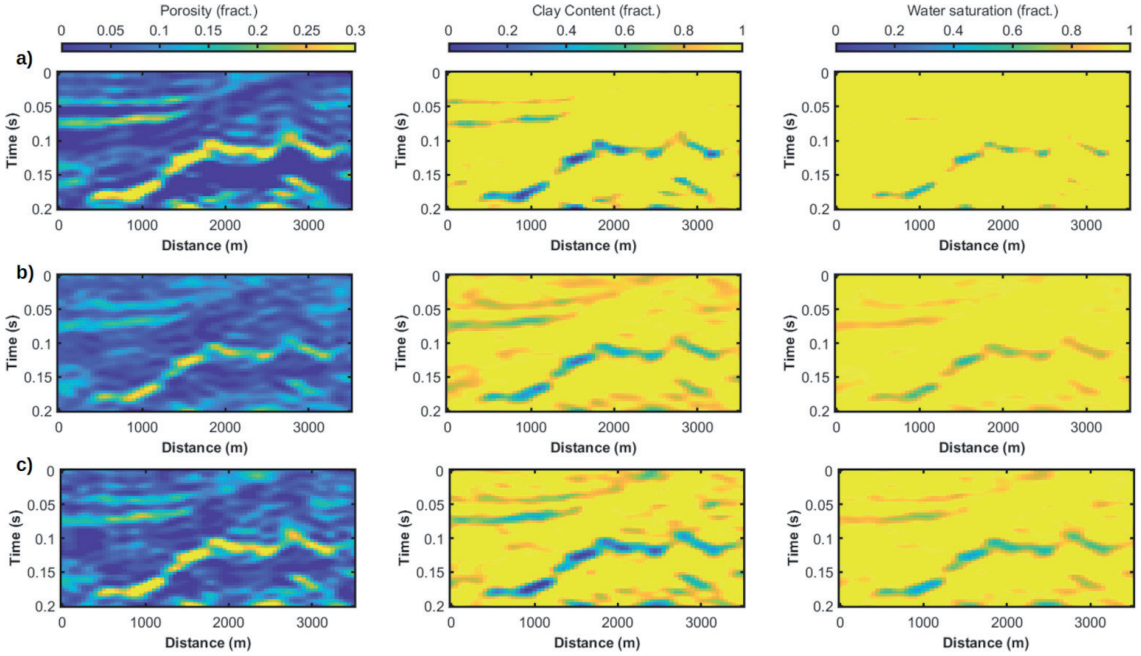

Fig. 23 - Results obtained when the noise model and the source wavelet are perfectly known: a) the modified test model; b) the CNN predictions yielded by the previously trained network (before transfer learning); c) the CNN results after transfer learning.

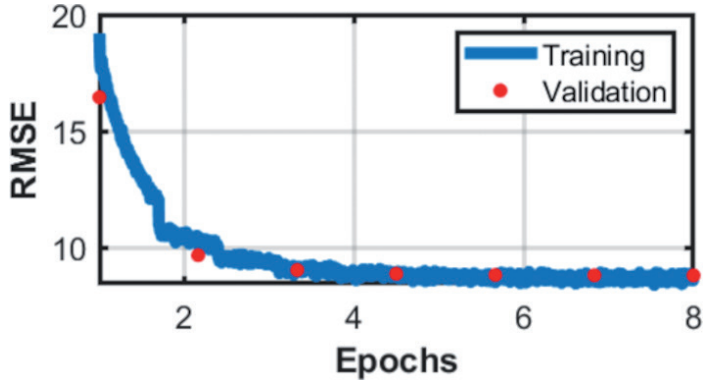

Fig. 24 - RMSE evolution during transfer learning.

in the recovered solution. In other terms, the combination of DCT and CNN provides a fast, albeit approximate, computation of the generalised inverse of a forward operator that includes temporal and spatial constraints in the model space. The DCT guaranteed a significant reduction of the dimensions of the input and output of the network. This reduced both the complexity of the CNN architecture and the computational cost of the generation and training phases. The number of retained DCT coefficients constitutes a compromise between the desired model resolution and the dimensionality reduction of data and model spaces. We demonstrated that DCT projections of prior realisations and associated seismic responses can be used to properly set such a threshold level.

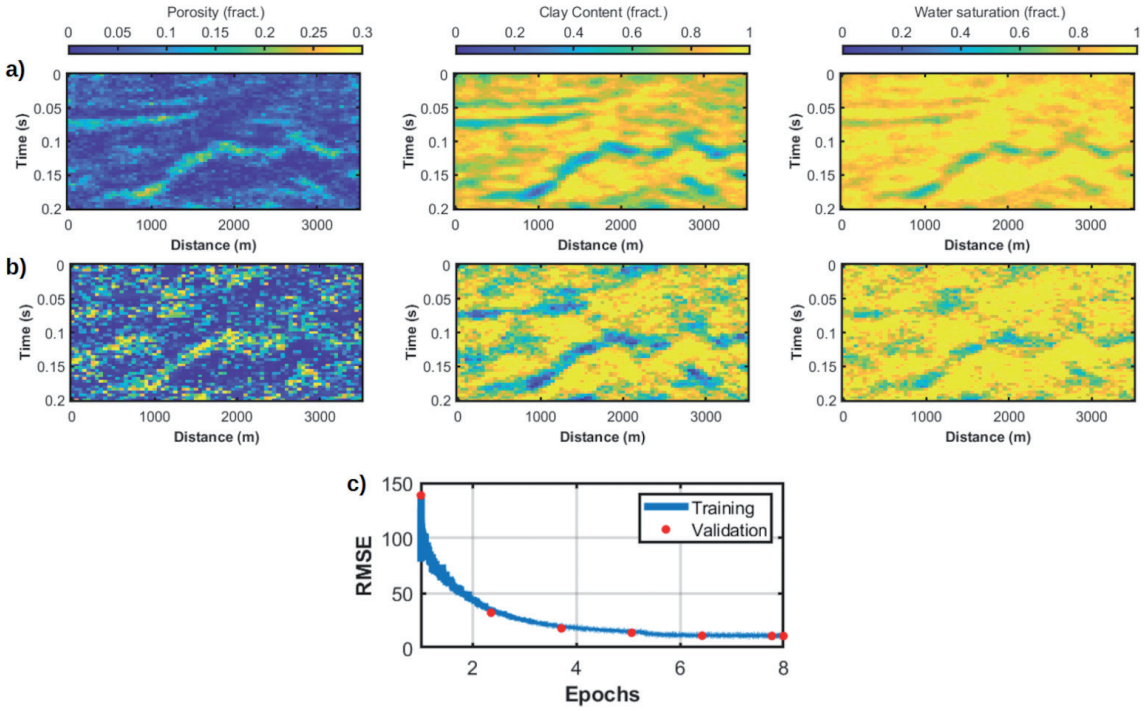

Fig. 25 - CNN predictions without DCT compression: a) results for perfectly known RPM, wavelet, and noise model (see Test 1); b) results for coherent noise in the data, errors in the assumed noise model, in the estimated source wavelet, and in the calibrated RPM (Test 5). Compare with Fig. 12; c) evolution of the RMSE on the validation and training sets for a training without DCT.

Differently from standard inversion approaches, the implemented method does not include any model constraints into the error function, but the network is trained on a data set containing realistic subsurface scenarios and learns how to reproduce a similar model to fit the input seismic data. It is known that an accurate uncertainty appraisal is a key aspect of any inversion approach. For this reason, the implemented CNN inversion framework has been combined with an MC simulation to properly propagate onto the estimated model the uncertainty related to noise contamination and the modelling error introduced by the network approximation.

The main advantage of the implemented inversion is its modest computational demand. We employed Matlab codes running on a notebook equipped with a quad-core intel Core i-7 7700HQ CPU at 2.80 GHz and 16 Gb RAM. Our approach includes four phases: data generation, network design, training, model prediction + MC uncertainty propagation. Different geostatistical simulation tools can be used for the data generation, and here we used the direct sequential co-simulation with joint probability distribution because it handles non-parametric prior models, but the method can be easily extended to whatever prior distribution and correlation pattern. This first stage is very computer intensive but also easily parallelisable. The computing time to draw the 6000 training + validation examples was 3 hours approximately, while only 5 minutes are needed for training. A computer cluster can drastically reduce the computational workload of the generation phase. We used common geostatistical tools to generate the training and validation sets according to the assumed marginal prior and the analytical variogram model. However, more advanced strategies [e.g. generative adversarial neural networks (Laloy *et al.*,

2018)] could be adopted to generate training examples that better reproduce the stratigraphic setting of the investigated area.

The DCT allowed a drastic reduction of the dimension of the input and output response of the network, thereby reducing not only the dimension of the training set but also the human and computational effort needed to select the best network architecture. Indeed, many experiments with different hyperparameter settings take only a short amount of time. Both the CNN inversion and the subsequent MC simulation give predictions in near real-time. For example, the 1000 MC simulations of Fig. 21 can be generated in less than 10 seconds. The many experiments we carried out (not shown here for brevity), demonstrated that without the DCT compression a much deeper network, larger training set, and a significantly longer training time, are needed to achieve similar predictions to those shown here. For the same network described in Fig. 6, the training of the CNN without the DCT takes more than two hours, on the same hardware configuration previously mentioned. As an example, we show here some results obtained without DCT compression on the Tests 1 and 5 previously described (Figs. 25a and 25b): when the RPM, wavelet, and noise model are perfectly known, the quality of the results is acceptable although the predictions are largely affected by lateral and vertical scattering. This scattering can be ascribed to the large dimension of the full petrophysical domain and to the fact that many model parameters are not informed by the data and lie in the null space of solutions. For coherent noise in the data, errors in the estimated wavelet, in the calibrated RPM, and in the assumed noise model (Test 5), the quality of the CNN predictions dramatically worsens with respect to the results obtained with DCT compression (compare Fig. 25 with Fig. 12). Now it would be difficult to map the layer boundaries with reasonable accuracy. This is a further demonstration that DCT compression provides more accurate and stable predictions. Fig. 25c also indicates that the convergence rate during the learning phase is much slower without DCT compression (compare with Fig. 7).

We showed that transfer learning can be used to update the internal network parameters when the target properties differ from those assumed for the training phase. The authors suggest retraining the network whenever the prior assumed for training differ from the prior of the target. In this context, transfer learning is extremely useful because it allows adjusting the network parameters with a very limited computational effort. Less than a minute is needed to re-training the CNN model with transfer learning. In real data applications, blind well log tests and comparisons between predicted and observed seismic gathers can provide important indications on the accuracy of the predictions after transfer learning and on their ability to reproduce the observed data. This comparison is also useful to choose which part of the previously trained CNN model must be adjusted.

Anyway, as for any Bayesian application, high-quality borehole data are needed to derive accurate prior assumptions and a proper RPM calibration. Indeed, the efficacy of the proposed approach critically depends on the validity of the prior assumptions made to generate the training ensemble. This means that the observed data and the seismic data sets used for training can be considered realisations of the same random variables. In other terms, the statistical properties of the observed data must be captured by the data generated by prior realisations. Statistical tools can be used to verify this assumption and to falsify the prior model. We refer the interested reader to Pradhan and Mukerji (2020) for more theoretical insights. However, we deem that the good quality of the results in all the different tests suggests that in our case the prior assumptions hold. The method can be easily extended to 3-D cases by applying a 3-D DCT in the petrophysical space and a 4-D DCT in the data domain. However, extension to 3-D would be prohibitive in our case given the limited computational resources available.

5. Conclusions

We proposed an alternative petrophysical seismic inversion based on CNNs and MC simulations. A trained CNN projects the observed pre-stack seismic gathers onto the petrophysical space. The DCT reparameterisation reduces the computational effort of the generation and training phases and also preserves the assumed spatial variability pattern in the recovered solution. The CNN inversion is combined with a MC simulation that propagates into the petrophysical estimates the modelling error introduced by the network approximation and the noise affecting the data.

Synthetic inversion experiments on a realistic subsurface model and simulating realistic scenarios (very low signal-to-noise ratio, errors in the estimated source wavelet, errors in the calibrated RPM) showed promising results. The CNN efficiently and effectively approximated the inverse of a nonlinear operator, which is very expensive to compute analytically. In the optimal scenario (i.e. for RPM, noise statistics, and source wavelet perfectly known), the trained network provided predictions comparable to those achieved by a more standard linearised inversion. Differently, the CNN outperformed the SA in all the other more realistic tests. This indicates that the combined use of CNN and DCT guarantees final predictions less affected by errors in the estimated source wavelet, in the calibrated RPM, and in the assumed noise model. The implemented approach needs a relatively small training set for the learning process and once trained, the CNN and the MC simulation provide model predictions and associated uncertainty estimations instantaneously. Transfer learning can be employed to rapidly update the network weights when the target and training exhibit different statistical characteristics.

Data availability. Codes and synthetic data are available on request from the authors.

REFERENCES

Aleardi M.; 2020a: *Combining discrete cosine transform and convolutional neural networks to speed up the Hamiltonian Monte Carlo inversion of pre-stack seismic data*. Geophys. Prospect., 68, 2738-2761.

Aleardi M.; 2020b: *Discrete cosine transform for parameter space reduction in linear and non-linear AVA inversions.* J. Appl. Geophys., 179, 104106.

Aleardi M. and Mazzotti A.; 2014: *A feasibility study on the expected seismic AVA signatures of deep fractured geothermal reservoirs in an intrusive basement.* J. Geophys. Eng., 11, 065008, doi: 10.1088/1742-2132/11/6/065008.

Aleardi M. and Ciabarri F.; 2017: *Assessment of different approaches to rock-physics modeling: a case study from offshore Nile Delta*. Geophys., 82, MR15-MR25.

Aleardi M. and Salusti A.; 2020: *Markov chain Monte Carlo algorithms for target-oriented and interval-oriented amplitude versus angle inversions with non-parametric priors and non-linear forward modellings.* Geophys. Prospect., 68, 735-760, doi: 10.1111/1365-2478.12876.

Aleardi M. and Salusti A.; 2021: *Elastic prestack seismic inversion through discrete cosine transform reparameterization and convolutional neural networks*. Geophys., 86, R129-R146, doi: 10.1190/geo2020-0313.1.

Aleardi M., Ciabarri F. and Mazzotti A.; 2017: *Probabilistic estimation of reservoir properties by means of wide-angle AVA inversion and a petrophysical reformulation of the Zoeppritz equations.* J. Appl. Geophys., 147, 28-41.

Aleardi M., Ciabarri F. and Gukov T.; 2018: *A two-step inversion approach for seismic-reservoir characterization and a comparison with a single-loop Markov-chain Monte Carlo algorithm.* Geophys., 83, R227-R244.

Araya-Polo M., Jennings J., Adler A. and Dahlke T.; 2018: *Deep-learning tomography*. The Leading Edge, 37, 58-66.

Avseth P., Mukerji T. and Mavko G.; 2010: *Quantitative seismic interpretation: applying rock physics tools to reduce interpretation risk*. Cambridge University Press, Cambridge, UK, 375 pp.

Azevedo L. and Soares A.; 2017: *Geostatistical methods for reservoir geophysics*. Springer, Berlin, Germany, 174 pp.

Azevedo L., Grana D. and Amaro C.; 2019: *Geostatistical rock physics AVA inversion*. Geophys. J. Int., 216, 1728-1739.

Bachrach R.; 2006: *Joint estimation of porosity and saturation using stochastic rock-physics modelling*. Geophys., 71, O53-O63.

Balles L. and Hennig P.; 2018: *Dissecting adam: the sign, magnitude and variance of stochastic gradients.* In: Proc. 35th International Conference on Machine Learning, Stockholm, Sweden, Vol. 80, pp. 404-413.

Bishop C.M.; 2006: *Pattern recognition and machine learning*. Springer, Berlin, Germany, 758 pp.

Biswas R., Sen M.K., Das V. and Mukerji T.; 2019: *Prestack and poststack inversion using a physics-guided convolutional neural network*. Interpretation, 7, SE161-SE174, doi: 10.1190/INT-2018-0236.1.

Bosch M., Cara L., Rodrigues J., Navarro A. and Díaz M.; 2007: *A Monte Carlo approach to the joint estimation of reservoir and elastic parameters from seismic amplitudes*. Geophys., 72, O29-O39.

Bosch M., Bertorelli G., Álvarez G., Moreno A. and Colmenares R.; 2015: *Reservoir uncertainty description via petrophysical inversion of seismic data*. The Leading Edge, 34, 1018-1026.

Britanak V., Yip P.C. and Rao K.R.; 2010: *Discrete cosine and sine transforms: general properties, fast algorithms and integer approximations*. Academic Press Inc., Elsevier Science, Amsterdam, The Netherlands, 715 pp.

Buland A. and Omre H.; 2003: *Bayesian linearized AVO inversion*. Geophys., 68, 185-198.

Das V. and Mukerji T.; 2020: *Petrophysical properties prediction from prestack seismic data using convolutional neural networks*. Geophysics, 85(5), N41-N55.

Das V., Pollack A., Wollner U. and Mukerji T.; 2019: *Convolutional neural network for seismic impedance inversion*. Geophys., 84, R869-R880.

de Figueiredo L.P., Grana D., Roisenberg M. and Rodrigues B.B.; 2019: *Multimodal Markov chain Monte Carlo method for nonlinear petrophysical seismic inversion*. Geophys., 84, M1-M13.

Doyen P.M.; 1988: *Porosity from seismic data: a geostatistical approach*. Geophys., 53, 1263-1275.

Doyen P.; 2007: *Seismic reservoir characterization: an earth modelling perspective*. EAGE Education Tour publications, Utrecht, the Netherlands, 255 pp., doi: 10.3997/9789073781771.

Dubrule O.; 2003: *Geostatistics for seismic data integration in earth models*. Society of Exploration Geophysicists, Houston, TX, USA, 274 pp., doi: 10.1190/1.9781560801962.

Fjeldstad T. and Grana D.; 2018: *Joint probabilistic petrophysics-seismic inversion based on Gaussian mixture and Markov chain prior models*. Geophys., 83, R31-R42.

Fjeldstad T. and Omre H.; 2019: *Bayesian inversion of convolved hidden Markov models with applications in reservoir prediction*. IEEE Trans. Geosci. Remote Sens., 58, 1957-1968.

Géron A.; 2019: *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems*. O'Reilly Media Inc., Sebastopol, MA, U.S.A., 819 pp.

Goodfellow I., Bengio Y. and Courville A.: 2016; *Deep learning*. The MIT Press, Cambridge, MA, U.S.A., 800 pp.

Grana D.; 2020: *Bayesian petroelastic inversion with multiple prior models*. Geophys., 85, M57-M71.

Grana D. and Della Rossa E.; 2010: *Probabilistic petrophysical-properties estimation integrating statistical rock physics with seismic inversion*. Geophys., 75, O21-O37.

Grana D., de Figueiredo L.P. and Azevedo L.; 2019: *Uncertainty quantification in Bayesian inverse problems with model and data dimension reduction*. Geophys., 84, M15-M24.

Hansen T.M. and Cordua K.S.; 2017: *Efficient Monte Carlo sampling of inverse problems using a neural network-based forward-applied to GPR crosshole traveltime inversion*. Geophys. J. Int., 211, 1524-1533.

He K., Zhang X., Ren S. and Sun J.; 2015: *Delving deep into rectifiers: surpassing human-level performance on imagenet classification*. In: Proc. IEEE International Conference on Computer Vision, Santiago, Chile, pp. 1026-1034.

Horta A. and Soares A.; 2010: *Direct sequential co-simulation with joint probability distributions*. Math. Geosci., 42, 269-292.

Krizhevsky A., Sutskever I. and Hinton G.E.; 2017: *Imagenet classification with deep convolutional neural networks*. Commun. ACM, 60, 84-90.

Laloy E., Herault R., Lee J., Jacques D. and Linde N.; 2017: *Inversion using a new low-dimensional representation of complex binary geological media based on a deep neural network*. Adv. Water Resour., 110, 387-405.

Laloy E., Hérault R., Jacques D. and Linde N.; 2018: *Training-image based geostatistical inversion using a spatial generative adversarial neural network*. Water Resour. Res., 54, 381-406.

LeCun Y., Bengio Y. and Hinton G.; 2015: *Deep learning*. Nature, 521, 436-444.

Li S., Liu B., Ren Y., Chen Y., Yang S., Wang Y. and Jiang P.; 2020: *Deep-learning inversion of seismic data*. IEEE Trans. Geosci. Remote Sens., 58, 2135-2149.

Lochbühler T., Breen S.J., Detwiler R.L., Vrugt J.A. and Linde N.; 2014: *Probabilistic electrical resistivity tomography of a CO2 sequestration analog*. J. Appl. Geophys., 107, 80-92.

Mavko G., Mukerji T. and Dvorkin J.; 2020: *The rock physics handbook: tools for seismic analysis of porous media, 2nd ed.* Cambridge University Press, Cambridge, UK, 523 pp., doi: 10.1017/CBO9780511626753.

Moghadas D.; 2020: *One-dimensional deep learning inversion of electromagnetic induction data using convolutional neural network*. Geophys. J. Int., 222, 247-259.

Mukerji T., Avseth P., Mavko G., Takahashi I. and González E.F.; 2001: *Statistical rock physics: combining rock physics, information theory, and geostatistics to reduce uncertainty in seismic reservoir characterization*. The Leading Edge, 20, 313-319.

Park M.J. and Sacchi M.D.; 2020: *Automatic velocity analysis using convolutional neural network and transfer learning*. Geophys., 85, V33-V43.

Parzen E.; 1962: *On estimation of a probability density function and mode.* The Ann. Math. Stat., 33, 1065-1076.

Pradhan A. and Mukerji T.; 2020: *Seismic Bayesian evidential learning: estimation and uncertainty quantification of sub-resolution reservoir properties*. Comput. Geosci., 24, 1121-1140, doi: 10.1007/s10596-019-09929-1.

Puzyrev V.; 2019: *Deep learning electromagnetic inversion with convolutional neural networks*. Geophys. J. Int., 218, 817-832.

Richardson A.; 2018: *Seismic full-waveform inversion using deep learning tools and techniques*. Cornell University, Ithaca, NY, U.S.A. arXiv:1801.07232 (physics.geo-ph), 18 pp.

Rimstad K. and Omre H.; 2010: *Impact of rock-physics depth trends and Markov random fields on hierarchical Bayesian lithology/fluid prediction*. Geophys., 75, R93-R108.

Sajeva A., Scarpellini D., Tempone P. and Capaccioli S.; 2019: *A semi-empirical approach to model pressure dependence of elastic moduli in granular media accounting for variations of coordination number and Poisson ratio*. Geophys. Prospect., 67, 872-887.

Santurkar S., Tsipras D., Ilyas A. and Madry A.; 2018: *How does batch normalization help optimization?* In: Proc. 32nd Conference on Neural Information Processing Systems, Montréal, QC, Canada, pp. 2483-2493.

Schmidhuber J.; 2015: *Deep learning in neural networks: an overview*. Neural Networks, 61, 85-117.

Sun J., Slang S., Elboth T., Larsen Greiner T., McDonald S. and Gelius L.J.; 2020: *A convolutional neural network approach to deblending seismic data*. Geophys., 85, WA13-WA26.

Tarantola A.; 2005: *Inverse problem theory and methods for model parameter estimation*. SIAM, Philadelphia, PA, USA, 348 pp., doi: 10.1137/1.9780898717921.

Waldeland A.U., Jensen A.C., Gelius L.J. and Solberg A.H.S.; 2018: *Convolutional neural networks for automated seismic interpretation*. The Leading Edge, 37, 529-537.

Wang B., Zhang N., Lu W. and Wang J.; 2019: *Deep-learning-based seismic data interpolation: a preliminary result*. Geophys., 84, V11-V20.

Wang Z., Bovik A.C., Sheikh H.R. and Simoncelli E.P.; 2004: *Image quality assessment: from error visibility to structural similarity*. IEEE Trans. Image Process., 13, 600-612.

Wu Y. and McMechan G.A.; 2019: *Parametric convolutional neural network-domain full-waveform inversion*. Geophys., 84, R881-R896, doi: 10.1190/geo2018-0224.1.

Yang F. and Ma J.; 2019: *Deep-learning inversion: a next-generation seismic velocity model building method*. Geophys., 84, R583-R599.

Corresponding author: Mattia Aleardi,
Earth Sciences Department, University of Pisa
Via Santa Maria 53, 56126 Pisa, Italy
Phone: +39 050 2215722; e-mail: mattia.aleardi@unipi.it