Petroleum Science 21 (2024) 1611–1631

Contents lists available at ScienceDirect

Petroleum Science

journal homepage: www.keaipublishing.com/en/journals/petroleum-science

Original Paper

Probabilistic seismic inversion based on physics-guided deep mixture density network

Qian-Hao Sun [a, b], Zhao-Yun Zong [a, b, *], Xin Li [c]

[a] National Key Laboratory of Deep Oil and Gas, China University of Petroleum (East China), Qingdao, 266580, Shandong, China
[b] Laoshan Laboratory, Qingdao, 266580, Shandong, China
[c] CNOOC Research Institute Ltd., Beijing, 100028, China

ARTICLE INFO

Article history:
Received 21 February 2023
Received in revised form
11 October 2023
Accepted 20 December 2023
Available online 28 December 2023

Edited by Jie Hao and Meng-Jiao Zhou

Keywords:
Deep learning
Probabilistic inversion
Physics-guided
Deep mixture density network

ABSTRACT

Deterministic inversion based on deep learning has been widely utilized in model parameters estimation. Constrained by logging data, seismic data, wavelet and modeling operator, deterministic inversion based on deep learning can establish nonlinear relationships between seismic data and model parameters. However, seismic data lacks low-frequency and contains noise, which increases the non-uniqueness of the solutions. The conventional inversion method based on deep learning can only establish the deterministic relationship between seismic data and parameters, and cannot quantify the uncertainty of inversion. In order to quickly quantify the uncertainty, a physics-guided deep mixture density network (PG-DMDN) is established by combining the mixture density network (MDN) with the deep neural network (DNN). Compared with Bayesian neural network (BNN) and network dropout, PG-DMDN has lower computing cost and shorter training time. A low-frequency model is introduced in the training process of the network to help the network learn the nonlinear relationship between narrowband seismic data and low-frequency impedance. In addition, the block constraints are added to the PG-DMDN framework to improve the horizontal continuity of the inversion results. To illustrate the benefits of proposed method, the PG-DMDN is compared with existing semi-supervised inversion method. Four synthetic data examples of Marmousi II model are utilized to quantify the influence of forward modeling part, low-frequency model, noise and the pseudo-wells number on inversion results, and prove the feasibility and stability of the proposed method. In addition, the robustness and generality of the proposed method are verified by the field seismic data.

1. Introduction

Seismic inversion is an excellent tool for converting seismic data into subsurface parameters, such as impedance, formation velocity, density, and even petrophysical parameters. The accurate estimation of model parameters is helpful to understand the underground geological structure and properties, which can be used to guide reservoir characterization and fluid discrimination. At present, probabilistic inversion based on Bayesian theory is the mainstream method to estimate model parameters. As one of the popular inversion methods, Bayesian deterministic inversion has been widely utilized in exploration geophysics due to its high computational efficiency and ease of application (Buland and More, 2003; Downton, 2005; Li et al., 2017; Zong et al., 2012). In deterministic inversion, it is usually assumed that the model parameters and noise obey Gaussian prior distribution, and then MAP (maximum a posterior) solution is obtained by gradient-based method. However, seismic data usually lacks low-frequency and contains noise (Sun and Demanet, 2020; Zong et al., 2018). Utilizing narrowband seismic data to predict broadband impedance is an ill-posed and nonlinear problem (Gholami, 2016). To address this issue, the regularization constraint is usually added to the objective function to reduce the solution space (Gholami, 2015; Hamid and Pidlisecky, 2015; Sun and Zong, 2019). Compared with deterministic inversion, geostatistical inversion usually assumes that the prior distribution is Gaussian mixture distribution. Under the Bayesian framework, a posterior probability distribution is randomly sampled to estimate the optimal interval of parameters

* Corresponding author.
E-mail address: zongzhaoyun@upc.edu.cn (Z.-Y. Zong).

https://doi.org/10.1016/j.petsci.2023.12.015

and the optimal geological facies (Grana, 2016; Grana and Rossa, 2010; Li et al., 2020; Spikes et al., 2007). Although both methods are mature enough to solve many practical problems, simplified physical models will lead to undesirable results. Deep learning (DL) methods can learn the complex physical system directly from training data. DL can overcome the problem of insufficient accuracy of physical models in traditional seismic inversion and geostatistical inversion. Therefore, it is a very suitable solution to those problems without transparent knowledge background or accurate physical model.

With the emergence and rapid development of deep neural network (DNN), the DL algorithm has been successfully used to solve various problems in exploration geophysics. The DL method can more accurately learn the nonlinear forward and inverse mapping between seismic data and model parameters, thus reducing the solution space. According to different training data, seismic inversion methods based on DL can be divided into three categories: unsupervised learning, supervised learning, and semi-supervised learning. Unsupervised learning seismic inversion requires only seismic data. For instance, Biswas et al. (2019) developed a physics-constrained DL inversion method. However, the prediction accuracy of this method largely depends on the quality of seismic data, and the low-frequency models are not involved in the training process. As a special type of unsupervised learning, self-supervised learning is also a label-free learning method, which usually includes pretext tasks and downstream tasks (Hu et al., 2020). In general, in self-supervised learning, the pretext task is designed to mine the characteristics of seismic data itself as supervisory information to improve the feature extraction capability of the network. Then, the pre-trained network is used to solve the concerned downstream tasks. Based on the above process, self-supervised learning is usually used to predict low-frequency information of seismic data (Wang et al., 2020). Supervised learning usually requires logging curves and corresponding seismic traces to establish inversion network, such as velocity estimation based on full connected neural networks (FCNNs) (Röth and Tarantola, 1994), seismic impedance inversion based on convolutional neural networks (CNNs) (Das et al., 2019), petrophysical parameters seismic inversion based on CNNs (Das and Mukerji, 2020), and pre-stack inversion based on U-net (Cao et al., 2022). In their research, stack or pre-stack seismic data are inputs to the network, and model parameter curves obtained from logging is output labels. The accuracy of supervised learning model largely depends on the richness of training data set. Compared with the above machine learning methods, semi-supervised learning not only learns the nonlinear relationship between logging curves and corresponding seismic data, but also learn the physical laws of forward modeling. For example, Alfarraj and AlRegib (2019) proposed a physics-constrained semi-supervised seismic inversion, which can fully mine seismic data and alleviate the effect of fewer labels. Because of the advantages of semi-supervised learning, it has been widely utilized to predict reservoir parameters (Chen and Saygin, 2021; Di and Abubakar, 2021; Sun et al., 2021; Wu et al., 2021). To improve the performance of networks, low-frequency model is introduced in training process to better compensate for the missing low-frequency information (Song et al., 2022).

Although the conventional seismic inversion method based on DL has good performance, it cannot directly evaluate the inversion uncertainty of model parameters. In seismic inversion method based on DL, different statistical methods are generally utilized to analyze the uncertainty of predicted parameters, such as approximate Bayesian computation (Das et al., 2019). As a powerful tool to quantify the uncertainty of results, Bayesian neural network (BNN) has been applied to signal discrimination (Maiti and Tiwari, 2010), seismic inversion (Ge et al., 2021), and well-log prediction (Feng et al., 2021). However, the parameters of BNN are twice as large as those of traditional networks, which will lead to larger computational cost and longer training time (Feng et al., 2021). Network dropout is also a method to quantify the output uncertainty (Junhwan et al., 2022), but it requires multiple realizations of the network. In order to directly and quickly quantify the uncertainty of the target variable conditioned on observed data. Bishop (1994) proposed mixture density network (MDN), which can learn the nonlinear mapping between the input data and the posterior probability distribution of the target variable. Based on this advantage, MDN is utilized instead of Monte Carlo method to resolve nonlinear problems in petrophysics (Shahraeeni and Curtis, 2011; Shahraeeni et al., 2012; Wang et al., 2021). MDN is also utilized to quantify the uncertainty of S-wave velocity from the Rayleigh wave dispersion curve (Earp et al., 2020). In addition, MDN is also utilized to evaluate the uncertainty of travel-time tomography (Earp and Curtis, 2020).

Inspired by MDN and semi-supervised learning seismic inversion, we propose a novel probabilistic seismic inversion method based on physics-guided deep mixture density network. The construction of PG-DMDN (physics-guided deep mixture density network) is based on the sequence modeling neural network (SM-net) proposed by Alfarraj and AlRegib (2019). In this paper, impedance inversion is taken as an example to introduce proposed method. The proposed method can be easily extended to the case of prestack data. Marmousi II model test verifies the feasibility of the PG-DMDN in the case of small samples. First, we compare PG-DMDN with SM-net to illustrate the advantages of the proposed method. At the same time, we quantify the influence of forward model part, noise, low-frequency model, and the number of pseudo-wells on inversion results. To illustrate the reliability of the prediction uncertainty, we provide a comparison to other uncertainty estimation methods, such as Stein variational gradient descent method (SVGD) (Zhang and Curtis, 2020, 2021; Zhao et al., 2022). Finally, the potential of PG-DMDN to solve the small-sample problem is illustrated by field data, and the uncertainty of seismic inversion is also evaluated.

Compared with SM-net, the proposed PG-DMDN has four differences: (1) PG-DMDN can predict the posterior probability distribution of model parameters rather than the deterministic value. (2) In the training process of PG-DMDN, low-frequency model is introduced to help the network learn the nonlinear relationship between narrowband seismic data and low-frequency impedance. (3) Since the output of the PG-DMDN is a posteriori probability distribution, we change the corresponding loss function. The loss function of forward modeling part is reconstructed based on Bayesian theory. (4) To improve the horizontal continuity, we added a block constraint to the framework and modify the corresponding network structure. Similar training strategies can also be found in the literature (Gao et al., 2020). Compared with BNN, the training parameters of the PG-DMDN are half of those of BNN, so its training cost is lower. The proposed method can not only obtain more accurate inversion results, but also predict the probability distribution of model parameters.

2. Theory

2.1. Physics-guided deep mixture density networks

For the conventional inversion problem based on DL, the parametric DNN is utilized to learn the nonlinear mapping between seismic data $\mathbf{x}$ and model parameter $\mathbf{R}$, which is generally expressed as

$$F_{\Theta}^{\dagger}(\mathbf{x}) \approx \mathbf{R}. \tag{1}$$

$F_{\Theta}^{\dagger}: \mathbf{x} \rightarrow \mathbf{R}$ represents the DNN inverse operator parameterized by Θ.

To evaluate the uncertainty of seismic inversion, the target variable in Eq. (1) is replaced by the probability distribution of model parameter conditioned on seismic data. Then, Eq. (1) can be rewritten as

$$F_{\Theta}^{\dagger}(\mathbf{x}) \approx p(\mathbf{R} \mid \mathbf{x}). \tag{2}$$

The MDN can be used to emulate the approximation of $p(\mathbf{R} \mid \mathbf{x})$.The conditional probability of $\mathbf{R}$ with respect to $\mathbf{x}$ can be represented by the known probability (Bishop, 1994):

$$p(\mathbf{R}|\mathbf{x}) = \sum_{i=1}^{k} \alpha_i(\mathbf{x}) \kappa_i(\mathbf{R}|\mathbf{x}), \tag{3}$$

where $\kappa_i(\mathbf{R}|\mathbf{x})$ is a known probability density function (PDF), called the kernel; k represents the number of kernels; and $\alpha_i(\mathrm{x})$ represents the weighting coefficient of each kernel.

It is well known that given a sufficient number of Gaussian kernels, the Gaussian mixture model can approximate any PDF with the desired accuracy (Shahraeeni and Curtis, 2011). Therefore, we assume that the kernel is Gaussian and $\kappa_i(\mathbf{R}|\mathbf{x})$ can be written as (Bishop, 1994)

$$\kappa_i(\mathbf{R}|\mathbf{x}) = \frac{1}{\prod_{j=1}^{h}\left(\sqrt{2\pi}\,\sigma_{ij}(\mathbf{x})\right)} \exp\left\{ -\frac{1}{2} \sum_{j=1}^{h} \frac{\left(R_j - \mu_{ij}(\mathbf{x})\right)^2}{\sigma_{ij}^2(\mathbf{x})} \right\}, \tag{4}$$

where h represents the number of target variables $\mathbf{R} = (R_1, \cdots, R_h)$, μ_{ij} represents the jth value in the mean vector of the ith kernel, and σ_{ij}^2 denotes the jth diagonal value in the covariance matric of the ith kernel. It is obvious from Eqs. (3) and (4) that the output of MDN is $\mathbf{M} = \{\alpha_i, \mu_i, \sigma_i (i = 1, 2, \cdots, k)$. We take $\mathbf{M}$ as a function of the input seismic data $\mathbf{x}$, denoted as $F_{\Theta}^{\dagger}(\mathbf{x}) \approx p(\mathbf{R} \mid \mathbf{x}) = \mathbf{M}$. Because of the complex nonlinear relationship between $\mathbf{M}$ and $\mathbf{x}$, DNN is used to establish mapping $F_{\Theta}^{\dagger}: \mathbf{x} \rightarrow \mathbf{M}$. The DNN and MDN constitute DMDN.

The coefficients α_i must satisfy $\sum_{i=1}^{k} \alpha_i(\mathrm{x}) = 1$, and the standard deviations σ_i must be positive. The outputs of DNN are transformed as follows (Bishop, 1994):

$$\alpha_i = \frac{\exp\left(\mathrm{y}_i^{\alpha}\right)}{\sum_{i=1}^{k} \exp\left(\mathrm{y}_i^{\alpha}\right)}, \; i = 1, \cdots, k, \tag{5}$$

$$\sigma_{ij} = \exp\left(\mathrm{y}_{ij}^{\sigma}\right), \; i = 1, \cdots, k, \text{ and } j = 1, \cdots, h, \tag{6}$$

$$\mu_{ij} = \mathrm{y}_{ij}^{\mu}, \; i = 1, \cdots, k, \text{ and } j = 1, \cdots, h, \tag{7}$$

where y_i^{α}, y_{ij}^{σ} and y_{ij}^{μ} denote the outputs of DNN. For each seismic trace $\mathbf{x}$, DMDN can be utilized to estimate the corresponding PDF of impedance $\mathbf{R}$.

Next, the structure of DNN in DMDN is introduced and PG-DMDN is constructed. As shown in Fig. 1, PG-DMDN includes forward modeling part and inversion part. The forward modeling part consists of convolution model and likelihood function. The inversion part consists of feature fusion layer, upscaling and regression layer. The feature fusion layer consists of multiple feature extraction layers. The multiple feature extraction layers are used to replace a single layer to simultaneously extract the spatial information from $i - n$ trace to $i + n$ trace. This layer will improve the horizontal continuity of the results. The single feature extraction layer consists of low-frequency information extraction layer, sequence model layer and local feature extraction layer. Hyperbolic tangent function (Tanh) is an activation function used to increase network nonlinearity. PG-DMDN proposed in this work is an improvement of SM-net (Alfarraj and AlRegib, 2019), which has four differences:

(1) The first difference is that the output of PG-DMDN is probability distribution not deterministic value. In the regression layer, MDN is used to replace the full connection layer to output the posterior probability distribution of model parameters.
(2) The second difference is that in feature extraction, a low-frequency information extraction layer is added to the feature extraction layer. This layer can enhance the network's ability to capture low-frequency information of parameters.
(3) The third difference is that In the forward modeling part, the Gaussian likelihood function is used to replace the mean squared error (MSE), because MSE cannot be used to train the probabilistic networks.
(4) The fourth difference is that in the feature fusion layer, multiple feature extraction layers are used to replace a single layer to simultaneously extract the spatial information from i-n trace to $i + n$ trace. The features of seismic trace i and its adjacent $2n$ seismic traces are extracted by $2n + 1$ feature extraction layers, and they are adaptively fused to obtain the spatial information of seismic data and low-frequency models. This part will improve the horizontal continuity of the inversion results.

The inputs of PG-DMDN are block seismic data $\mathbf{x}_{i-n\sim i+n} (i = 1, 2, \cdots, m)$ and block low-frequency model $\mathbf{R}_{i-n\sim i+n}^{\mathrm{low}} (i = 1, 2, \cdots, m)$, and the output is PDF $p(\mathbf{R}_i | \mathbf{x}_i)$ $(i = 1, 2, \cdots, m)$, where m represents the trace number. The mean value of model parameters can be estimated with $R_i = \sum_{j=1}^{k} \alpha_j \mu_j$. Finally, the mean value is fed into the forward modeling part to evaluate the difference between the synthetic data and the real data. The architecture of sequence modeling layer and local feature extraction layer will be introduced in the next section.

2.2. Sequence modeling layer based on bidirectional-gated recurrent units

The traditional feed-forward neural network has been widely used in intelligent inversion of model parameters. However, the feed-forward neural network only captures the deterministic relationships and ignores the temporal correlation of seismic data and model parameters (Alfarraj and AlRegib, 2019). Recurrent neural networks (RNNs) can effectively capture the temporal correlation of time series, but fails to learn the long-term correlation. RNNs have the vanishing gradient problem when learning the time dependence of long sequence data (Werbos, 1990). As a more complex form of RNNs, long short-term memory (LSTMs) and gated recurrent units (GRUs) can be utilized to avoid gradient disappearance (Cho et al., 2014; Hochreiter and Schmidhuber, 1997). For small sample problems, such as seismic inversion, there is almost no difference in performance between LSTMs and GRUs. Compared with LSTMs, GRUs has fewer parameters, faster convergence speed

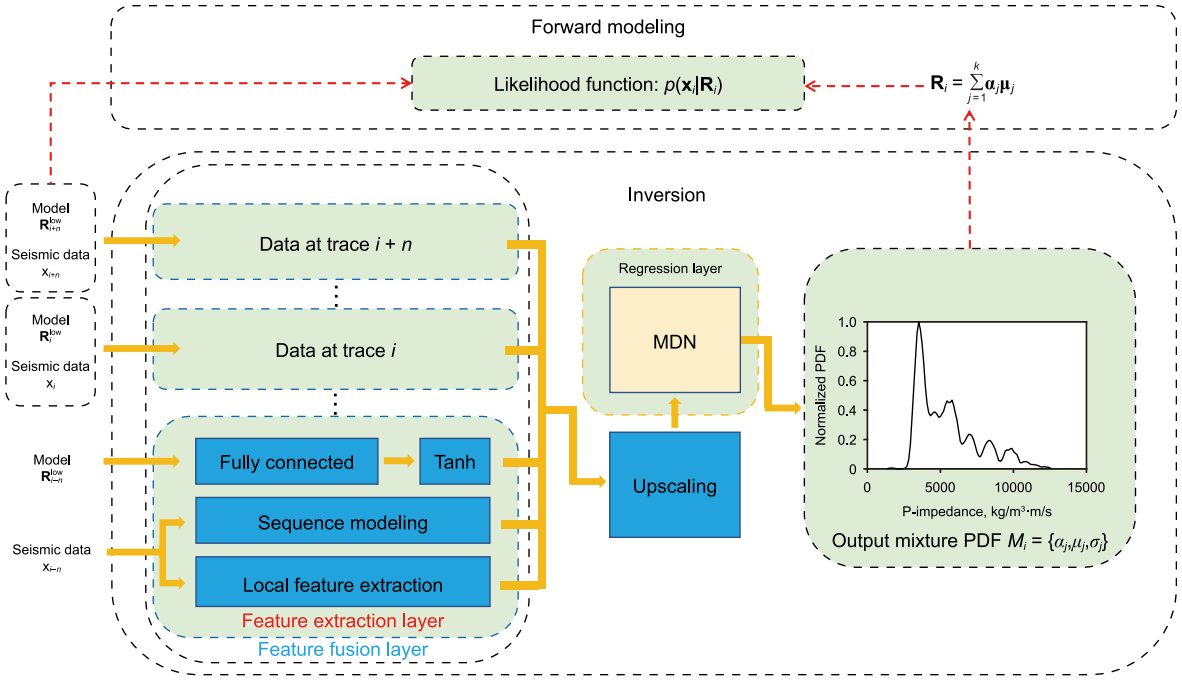

Fig. 1. The network structure of PG-DMDN.

and lower training cost. Therefore, in this study, bidirectional GRUs (B-GRUs) is selected as the sequence modeling layer to establish PG-DMDN. As shown in Fig. 2, a single-layer bidirectional GRUs (B-GRUs) consists of two GRUs. The forward process through a GRUs is as follows:

$$\begin{cases} \mathbf{u}_i^{(t)} = \text{sigmoid}\left(\mathbf{W}_{uh}\mathbf{h}_i^{(t-1)} + \mathbf{W}_{ux}\mathbf{x}_i^{(t)} + \mathbf{b}_u\right) \\ \mathbf{r}_i^{(t)} = \text{sigmoid}\left(\mathbf{W}_{rh}\mathbf{h}_i^{(t-1)} + \mathbf{W}_{rx}\mathbf{x}_i^{(t)} + \mathbf{b}_r\right) \\ \widehat{\mathbf{h}}_i^{(t)} = \tanh\left(\mathbf{W}_{hh}\left[\mathbf{r}_i^{(t)} * \mathbf{h}_i^{(t-1)}\right] + \mathbf{W}_{hx}\mathbf{x}_i^{(t)} + \mathbf{b}_h\right) \\ \mathbf{h}_i^{(t)} = \left(1 - \mathbf{u}_i^{(t)}\right) * \mathbf{h}_i^{(t-1)} + \mathbf{u}_i^{(t)} * \widehat{\mathbf{h}}_i^{(t)} \end{cases}, \tag{8}$$

where $\mathbf{u}_i^{(t)}$, $\mathbf{r}_i^{(t)}$, $\mathbf{x}_i^{(t)}$, $\mathbf{h}_i^{(t)}$, $\widehat{\mathbf{h}}_i^{(t)}$ and $\mathbf{y}_i^{(t)}$ $(i = 1, 2, \cdots, m;\ t = 1, 2, \cdots, T)$ represent the update-gate vectors, reset-gate vectors, input seismic data, hidden state variable, candidate hidden state variable, and final hidden state variable of the ith sample, respectively. m represents the batch size and T represents the number of sampling points. The operator * denotes element multiplication. The term $\mathbf{W}$ and $\mathbf{b}$ are learning parameters of the network, which will be updated during training. The subscript represents the direction of connection between different variables. For example, $\mathbf{W}_{uh}$ represents the weight used to connect update vectors and hidden state variable. Reset and update gate control the amount of information transferred from the previous hidden state to the current hidden state. B-GRUs not only captures the previously stored information (e.g., $\mathbf{h}_i^{(t-1)}$), but also extracts the later stored information (e.g., $\mathbf{h}_i^{(t+1)}$).

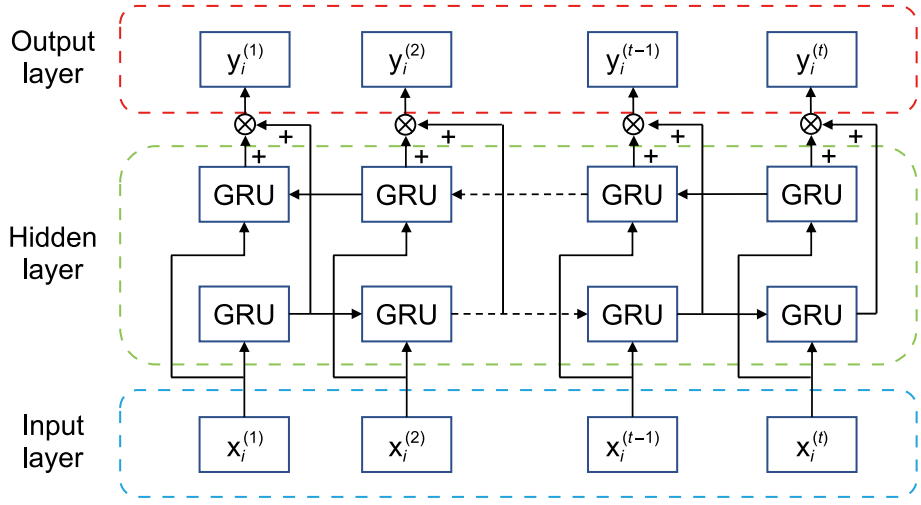

Fig. 2. The architecture of a single-layer B-GRUs.

In this paper, we use three layers of B-GRUs to construct the sequence modeling layer, which is the same as the SM-net (Alfarraj and AlRegib, 2019), to extract the long-term background trend from the seismic data. As shown in Fig. 3, the input data of the sequence modeling layer is a batch of seismic data slices with the size of $b \times T \times 1$. When seismic data is fed into the sequence modeling layer, the output size becomes $b \times T \times 2c$, where c is the dimension of the final hidden state in the B-GRUs. It is worth noting that the number of features in the output layer is $2c$, which is determined by the bidirectional propagation property of B-GRUs.

2.3. Local feature extraction layer based on CNN

Convolutional block (ConvBlock) generally includes convolution layer, batch normalization layer and nonlinear activation function. The core of CNN is the convolution layer, in which the feature map is convolved with the convolution kernel. The sliding convolution kernel can capture the local features of seismic data in different time windows. The forward process is the following form (Lecun et al., 1998; Gao et al., 2016):

$$\mathbf{y}_j^p = \sigma\left(\sum_{i=1}^{I} \mathbf{x}_i^{p-1} * \mathbf{W}_j + \mathbf{b}_j\right), \tag{9}$$

where $\mathbf{y}_j^p$ represents the jth feature map in the pth layer, $\sigma()$ is the nonlinear activation function, I represents the number of feature maps in the $(p-1)$th layer, $\mathbf{x}_i^{p-1}$ is ith feature map in the $(p-1)$th layer, $\mathbf{W}_j$ denotes the jth convolution kernel of the pth layer, and $\mathbf{b}_j$ is the bias matrix of the pth layer. The CNN learns the features contained in seismic data by updating these two parameters. Taking Eq. (9) as an example, the convolution layer is described in detail. Assume that the number of output maps in the $(p-1)$th is I, expected output in the pth is J, and the size of each convolution kernel is $m \times 1$, the size of the final output is $(m \times 1 \times I + 1) \times J$. Note that different convolution kernels capture different features from the input seismic data. The convolution kernels with different dilation factors can be regarded as different convolution kernels. The dilation factor controls the distance between points of the convolution kernel (Yu and Koltun, 2016). The multiple dilation factors enable convolution kernel to extract information not only near the reference sample, but also far away from it. For one-dimensional convolution kernel with size 3 and dilation = 1, 3, and 5, the dilated convolution is displayed as Fig. 4. The batch normalization layer normalizes the output of the convolution layer. See Appendix A for details of batch normalization (BN).

The nonlinear activation function is an important part of the neural network, which enables the neural network to approximate the nonlinear function. Tanh is utilized as the activation function. The tanh is as follow:

$$\sigma(\mathbf{x}) = \tanh(\mathbf{x}) = \frac{e^{2\mathbf{x}} - 1}{e^{2\mathbf{x}} + 1}, \tag{10}$$

where $\mathbf{x}$ is the input feature maps.

In this paper, three parallel one-dimensional convolution blocks with different dilation factors are utilized to capture the multi-scale features of seismic data. Then, multi-scale features are fused by using the full connected layer and convolution layer. The architecture of the local feature extraction layer is similar to that in the SM-net (Alfarraj and AlRegib, 2019), as shown in Fig. 5. The input data is a batch of seismic data slices of size $b \times T \times c_1$.When the seismic data is fed into the local feature extraction layer, the output size becomes $b \times T \times c_2$, where $c_2 = 2c$ mentioned in B-GRUs is the dimension of the final feature map. The network parameters c_1 is the number of input feature map and c_2 is the number of output feature map. Normally, $c_2 = 2c_1$. m and T are batch size and time sampling points.

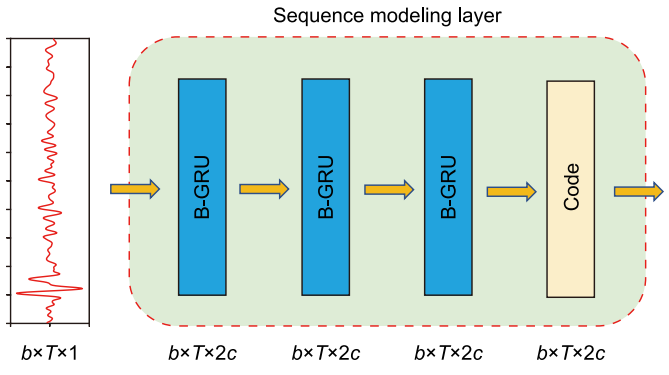

Fig. 3. The architecture of sequence modeling layer.

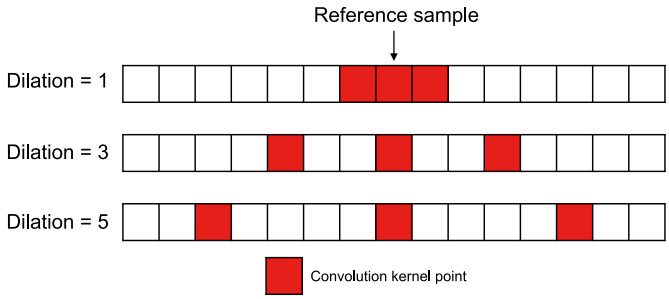

Fig. 4. Dilated convolution. (kernel size = 3, dilation factors dilation = 1, 3, 5).

2.4. Semi-supervised probabilistic seismic inversion method

The conventional physics-guided DL inversion includes two parts: forward modeling and inversion. The forward part is a physical model or a trainable neural network (Yuan et al., 2021), which can be written as

$$\mathbf{x} = F(\mathbf{R}) + \mathbf{n},$$

or

$$\mathbf{x} \approx F_\Theta(\mathbf{R}), \tag{11}$$

where $F : \mathbf{R} \rightarrow \mathbf{x}$ represents physical modeling operator and $F_\Theta : \mathbf{R} \rightarrow \mathbf{x}$ represents DNN parameterized by Θ. For the inverse problem, the parameterized DNNs is utilized to learn the nonlinear mapping between seismic data and model parameter, which is generally expressed as Eq. (1).

Compared with conventional physics-guided DL seismic inversion, the proposed probabilistic seismic inversion method has two differences. On the one hand, for the inverse operator, the target model parameter $\mathbf{R}$ is replaced by $\mathbf{M}$ in Eq. (1). On the other hand, since the output of inverse operator is the probability distribution of model parameter, the relationship between the probability distribution of model parameter and seismic data should be considered in the forward modeling. Therefore, we introduce Gaussian likelihood function into forward part to constrain the inverse operator:

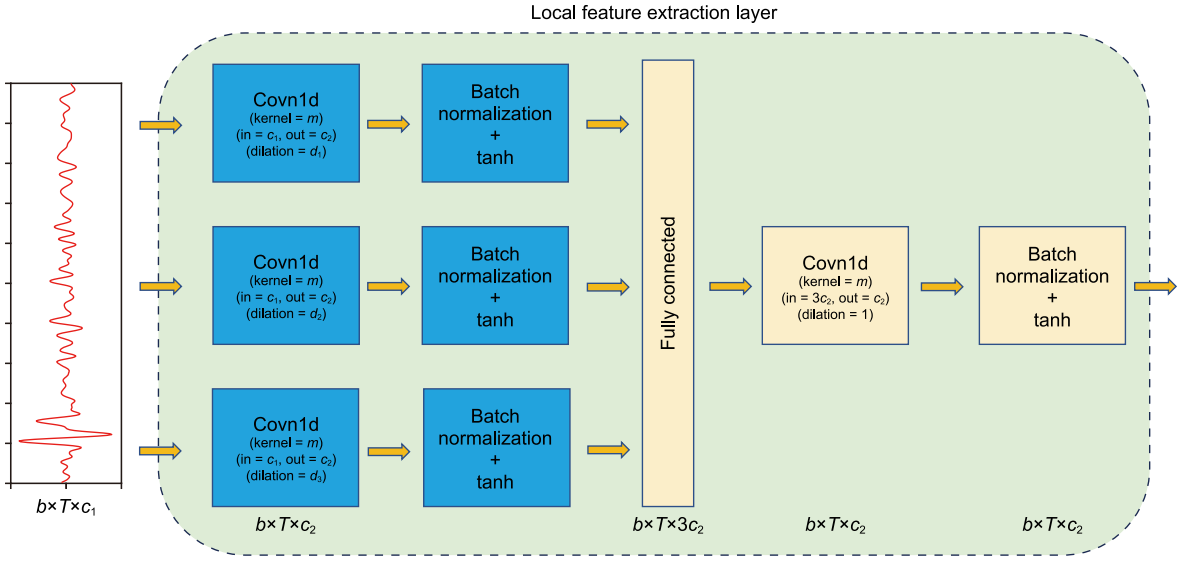

Fig. 5. The architecture of local feature extraction layer. (Covn1d denotes one-dimensional convolutional block).

$$p(\mathbf{x}\mid\mathbf{R})=\frac{1}{\sqrt{2\pi}\sigma_n}\exp\left[-\frac{(\mathbf{x}-F(\mathbf{R}))^{\mathrm{T}}(\mathbf{x}-F(\mathbf{R}))}{2\sigma_n^2}\right], \tag{12}$$

where $p(\mathbf{x}\mid\mathbf{R})$ is likelihood function, which describes the probability distribution of the difference between the real data and synthetic data, and σ_n^2 is the noise variance.

In this work, the modeling operator $F(\mathbf{R})$ in Eq. (12) is a convolution model, which can be written as

$$F(\mathbf{R})=\mathbf{Gm}, \tag{13}$$

where $\mathbf{m}=0.5\mathbf{D}\log(\mathbf{R})$ is the reflection coefficient of the model parameter. Then, Eq. (13) can be rewritten as

$$F(\mathbf{R})=\frac{1}{2}\mathbf{GD}\log(\mathbf{R}), \tag{14}$$

where $\mathbf{G}$ is the wavelet matrix, and

$$\mathbf{D}=\begin{bmatrix}-1 & 1 & 0 & \cdots\\ 0 & -1 & 1 & \ddots\\ 0 & 0 & -1 & \ddots\\ \vdots & \ddots & \ddots & \ddots\end{bmatrix}$$

is a differential matrix, which converts model parameter to reflectivity.

The whole workflow of probabilistic inversion method is shown in Fig. 6. The parameters of inverse network are updated by maximizing the following objective function:

$$L(\Theta)=\frac{\lambda_1}{N_\mathrm{p}}\sum_{i=1}^{N_\mathrm{p}}p(\mathbf{R}_i\mid\mathbf{x}_i)+\frac{\lambda_2}{N_\mathrm{s}}\sum_{i=1}^{N_\mathrm{s}}p(\mathbf{x}_i^*\mid\mathbf{R}_i^*), \tag{15}$$

where $\{\mathbf{R}_i,\mathbf{x}_{i-n\sim i+n}\}$ represents the training samples at the well location, $\{\mathbf{R}_i^*,\mathbf{x}_i^*\}$ represents the training samples of all traces. $\mathbf{R}_i^*$ and $\mathbf{x}_i^*$ are the estimated model parameter and corresponding synthetic seismic data, respectively. N_p is the number of available well-log data, N_s is the total number of seismic traces, λ_1 is the weight of inverse model, and λ_2 is the weight of forward model.

In the training phase, the inversion part $F_\Theta^\dagger:\mathbf{x}\to\mathbf{M}$ is trained by using the well-log data, the corresponding seismic data from trace $i-n$ to $i+n$, and low-frequency model from trace $i-n$ to $i+n$. The loss function of this part can be defined as the first term of Eq. (15). The estimated model parameter of the MDN at trace i will be utilized to generate synthetic data. Then, the likelihood function is utilized to estimate the difference between the synthetic and the real data, which can constrain the update of the inversion part. The loss function of this part can be defined as the second term of Eq. (15). To estimate the optimum model parameters, Eq. (15) can be converted to the following form:

$$L(\Theta)=\frac{\lambda_1}{N_\mathrm{p}}\sum_{i=1}^{N_\mathrm{p}}-\ln p(\mathbf{R}_i\mid\mathbf{x}_i)+\frac{\lambda_2}{N_\mathrm{s}}\sum_{i=1}^{N_\mathrm{s}}-\ln p(\mathbf{x}_i^*\mid\mathbf{R}_i^*), \tag{16}$$

By minimizing Eq. (16) with training data sets, the weights and biases of the PG-DMDN can be determined. After several iterations, the hyperparameters of the network model will be optimal. Then, posterior probability distribution of model parameter can be estimated by inputting seismic data into the trained PG-DMDN. Adam algorithm (Kingma and Ba, 2014) is utilized to optimize the training of PG-DMDN.

3. Synthetic data examples

3.1. Synthetic data preparation, network setting, and evaluation of result

Marmousi II model (Martin et al., 2006) is utilized as synthetic data to prove the feasibility and robustness of proposed method. The Marmousi II elastic model has 2720 traces, and each trace has 2070 sampling points. We utilized Marmousi II elastic model to generate absolute P-wave impedance (in time), as shown in Fig. 7(a). A 25Hz zero-phase Ricker wavelet with a sampling interval of 1 ms is convoluted with reflection coefficient obtained from Fig. 7(a) to generate synthetic seismic data. In order to simulate the resolution mismatch, the seismic data is down-sampled by 6 times, as shown in Fig. 7(b). The number of seismic data sampling points is 345. To simulate the low-frequency loss, we filter out the information below 6 Hz in the seismic data. Finally, synthetic seismic data with signal to noise ratio (SNR) of 10, 5, and 2 are used to evaluate the stability of proposed method. The seismic

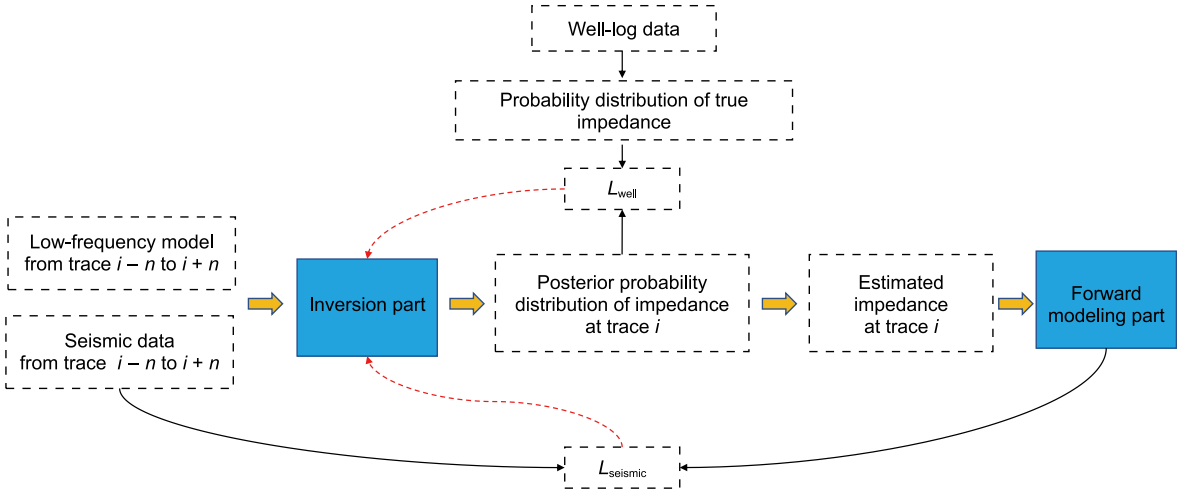

Fig. 6. The workflow of semi-supervised probabilistic seismic inversion method.

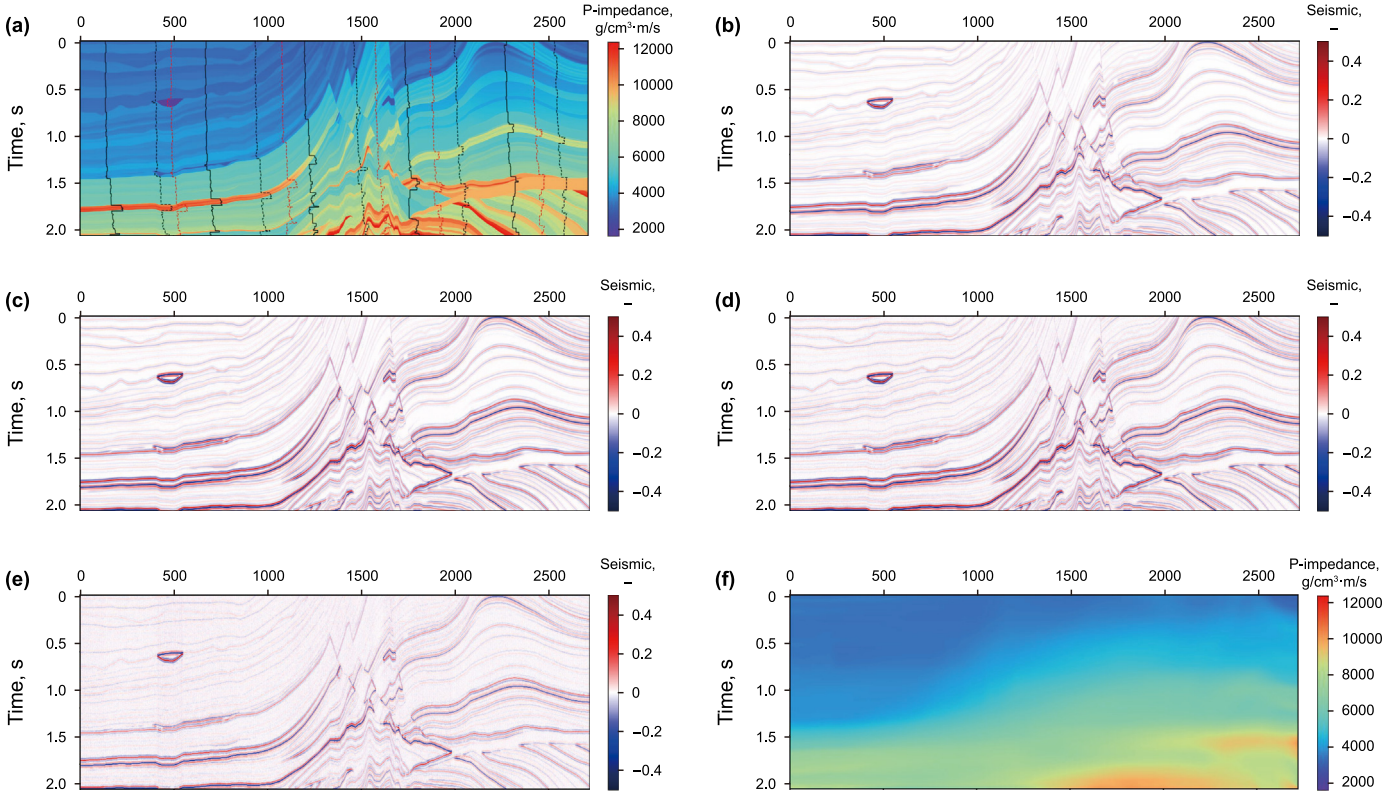

Fig. 7. The P-wave impedance data, seismic data and low-frequency model utilized for the training, validation, or test sets. **(a)** True impedance profile, **(b)** the seismic profile with no noise, **(c)** the seismic data with SNR = 10, **(d)** the seismic data with SNR = 5, **(e)** the seismic data with SNR = 2, and **(f)** the low-frequency model profile.

data with different SNR and low-frequency model are displayed in Fig. 7(c)–(f), respectively. The low-frequency model is obtained by longitudinally and laterally smoothing the true impedance. Five numerical tests are carried out utilizing the seismic data, low-frequency model, and true impedance in Fig. 7. The first test is to illustrate the advantages of the proposed method compared with SM-net inversion method. The second test is to verify the role of the forward modeling part and evaluate the inversion uncertainty. The third test is to demonstrate the robustness of the method. The fourth test attempts to analyze the influence of low-frequency model on inversion results. Finally, we try to test the ability of the proposed inversion method to solve smaller sample problems.

For first three tests, we extract ten impedance curves from the impedance profile (the solid and dashed black lines in Fig. 7(a)) as the training labels. Five pseudo-well curves (the dashed red lines in Fig. 7(a)) at common-depth points (CDPs) 500, 1100, 1600, 1900, 2450 are utilized for validation. For the last test, only the dashed impedance curve is used as the labeled data. Fig. 7(b)–(e) are utilized to test the performance of the PG-DMDN. Before training, the seismic data and impedance curves need to be normalized by the following equation:

$$x' = \frac{x - \mu_x}{\sigma_x}, \tag{17}$$

where x' is the normalized seismic (or impedance) data, x is the original seismic (or impedance) data, μ_x is the mean of x, and σ_x is the variance of x.

After the data sets are prepared, the workflow mentioned above is adopted to train the PG-DMDN. Then, the testing data is fed into PG-DMDN to estimate the impedance result. For the four tests of the synthesis example, the network parameters are set as follows: the number of adjacent traces for training is $n = 2$, the batch size of unsupervised learning part is $b = 10$; the input feature number of B-GRUs is $c = 8$; the input feature number of CNNs is $c_1 = 8$; the output feature number of CNNs is $c_2 = 16$; the Gaussian kernel number is $k = 5$; the kernel size of CNNs is $m = 5$; and the dilation factors are $d_1 = 1$, $d_2 = 3$ and $d_3 = 5$, respectively. Theoretically, the more the number of Gaussian kernels, the higher the accuracy of the inversion results. However, through multiple tests, too many kernels can only bring small improvements, but will increase a large amount of train costs. For four tests of the synthesis example, the Gaussian kernel number is 5, which makes the mixture model consistent with the probability distribution of the true impedance.

Within the testing phase, the inversion results are evaluated by determination coefficient (R^2) and Pearson's correlation coefficient (PCC). σ_{mean} is used to represent the uncertainty of inversion. The PCC is commonly utilized to measure the overall correlation between the original and predicted vectors. PCC is defined as

$$\text{PCC}(x, y) = \frac{1}{N} \frac{\sum_{i=1}^{N} (x(i) - \bar{x})(y(i) - \bar{y})}{\sqrt{\sum_{i=1}^{N} (x(i) - \bar{x})^2} \sqrt{\sum_{i=1}^{N} (y(i) - \bar{y})^2}}, \tag{18}$$

where $\bar{x}$ and $\bar{y}$ are the mean value of x and y, respectively. The R^2 is commonly utilized to evaluate the mean square error between the estimated and target vectors. R^2 is defined as

$$R^2(x, y) = 1 - \frac{\sum_{i=1}^{N} (x(i) - y(i))^2}{\sum_{i=1}^{N} (x(i) - \bar{y})^2}, \tag{19}$$

In addition, we used structural similarity (SSIM) to evaluates the similarity of two images from local statistics (on local windows) using the flowing equation:

$$\text{SSIM}(\mathbf{X}, \widehat{\mathbf{X}}) = [l(\mathbf{x}, \widehat{\mathbf{x}})]^{\alpha} \cdot [c(\mathbf{x}, \widehat{\mathbf{x}})]^{\beta} \cdot [s(\mathbf{x}, \widehat{\mathbf{x}})]^{\gamma}, \tag{20}$$

where $\mathbf{x}$ and $\widehat{\mathbf{x}}$ are patches from estimated and target image, respectively. The terms $l(\mathbf{x}, \widehat{\mathbf{x}})$, $c(\mathbf{x}, \widehat{\mathbf{x}})$, and $s(\mathbf{x}, \widehat{\mathbf{x}})$ are the luminance, contrast, and structure comparison functions, respectively. The terms $\alpha > 0$, $\beta > 0$, and $\gamma > 0$ are constants chosen to adjust the influence of the each of three terms. From SSIM, a single similarity value, denoted as M-SSIM, can be computed by taking the mean of all SSIM scores for all local windows.

3.2. Comparison between PG-DMDN and SM-net

To verify the benefits of the PG-DMDN, it is compared with SM-net. Fig. 8(a) and (c) show the comparison of inversion results based on PG-DMDN and SM-net. Fig. 8(b) and (d) show the absolute difference between the estimated impedance and the true impedance. Because the proposed method uses the block constraints, the lateral continuity of its inversion results is better than that of SM-net. However, SM-net cannot accurately restore the structure in the red rectangular area. In contrast, PG-DMDN-based inversion method has a better performance. The PG-DMDN will first learn the nonlinear relationship between narrowband seismic data and broadband impedance. Then, the modeling operator $F(\mathbf{R})$ is used to generate the narrowband seismic data to reduce the solution space of the inverse part. Besides matching the output of the PG-DMDN with broadband impedance labels, data generated by inputting the PG-DMDN output into the known modeling operator match the narrowband seismic data. The trained PG-DMDN can predict broadband impedance using narrowband seismic data. The low-frequency in the predicted impedance come from logging labels and low-frequency models. If there is no modeling operator to constrain the network, the inaccurate results will be obtained. See the next section for the corresponding results and analysis. To compare the two methods more fairly, the low-frequency model is merge with the results of SM-net, as shown in Fig. 8(e) and (f). The accuracy and lateral continuity of inversion results have been improved, but there are still large errors. A detailed comparison of the inversion results within the red rectangle in Fig. 8(e) and (f) is shown in Fig. 9. It can be seen that the results of SM-net have improved to some extent after merge the initial model. The results prove the superiority of the proposed method. Fig. 10 shows the comparison between the inversion impedance and the validation data. A similar conclusion can be drawn from Fig. 10. The quantitative evaluation of the inversion results is reported in Table 1. PG-DMDN inversion results have higher PCC, R^2, M-SSIM. It can be seen from the above results that PG-DMDN is better than SM-net.

3.3. The role of forward modeling part

To quantify the contribution of forward part in the loss function in Eq. (12), we compare the inversion results for different values of λ_1 and λ_2. Ten black impedance curves (Fig. 7(a)) and seismic data profile (Fig. 7(b)) are utilized as training data. The learning rate is 0.005. We set $\lambda_1 = 1$, $\lambda_2 = 0$ to implement DMDN-based inversion (supervised learning), and $\lambda_1 = 1$, $\lambda_2 = 1$ to implement PG-DMDN inversion (semi-supervised learning). For DMDN-based inversion, only ten impedance curves are used as label data to train the network. The corresponding 10 traces seismic data and low-frequency model are used as input data. The mean impedance models estimated by supervised and semi-supervised learning are show in Fig. 11(a) and (b). Both supervised learning and semi-supervised learning can restore the geological structure, but the inversion results of semi-supervised learning have better horizontal continuity and higher accuracy, especially in red box areas. Fig. 11(c) and (d) show the absolute difference between the original impedance and the estimated impedance. The difference between the estimated impedance of semi-supervised learning and the true impedance is smaller. Fig. 11(e) and (f) show the standard deviation of the estimated impedance. From the figures, the prediction uncertainty indicates where the predictive deviations are relatively large. High uncertainty is observed mainly at impedance boundaries, which is similar to Ge's conclusion (Ge et al., 2021). The standard deviation model obtained by semi-supervised learning (Fig. 11(f)) has a smoother structure than supervised learning (Fig. 11(e)). The smooth structure means that the inversion results are closer to the true model and have less uncertainty (Zhang and Curtis, 2021). To further understand the results, the comparison between the inversion impedance and the validation data is displayed as Fig. 12. The mean impedance and shaded area are in good agreement with the real impedance. The 2σ of shallow simple

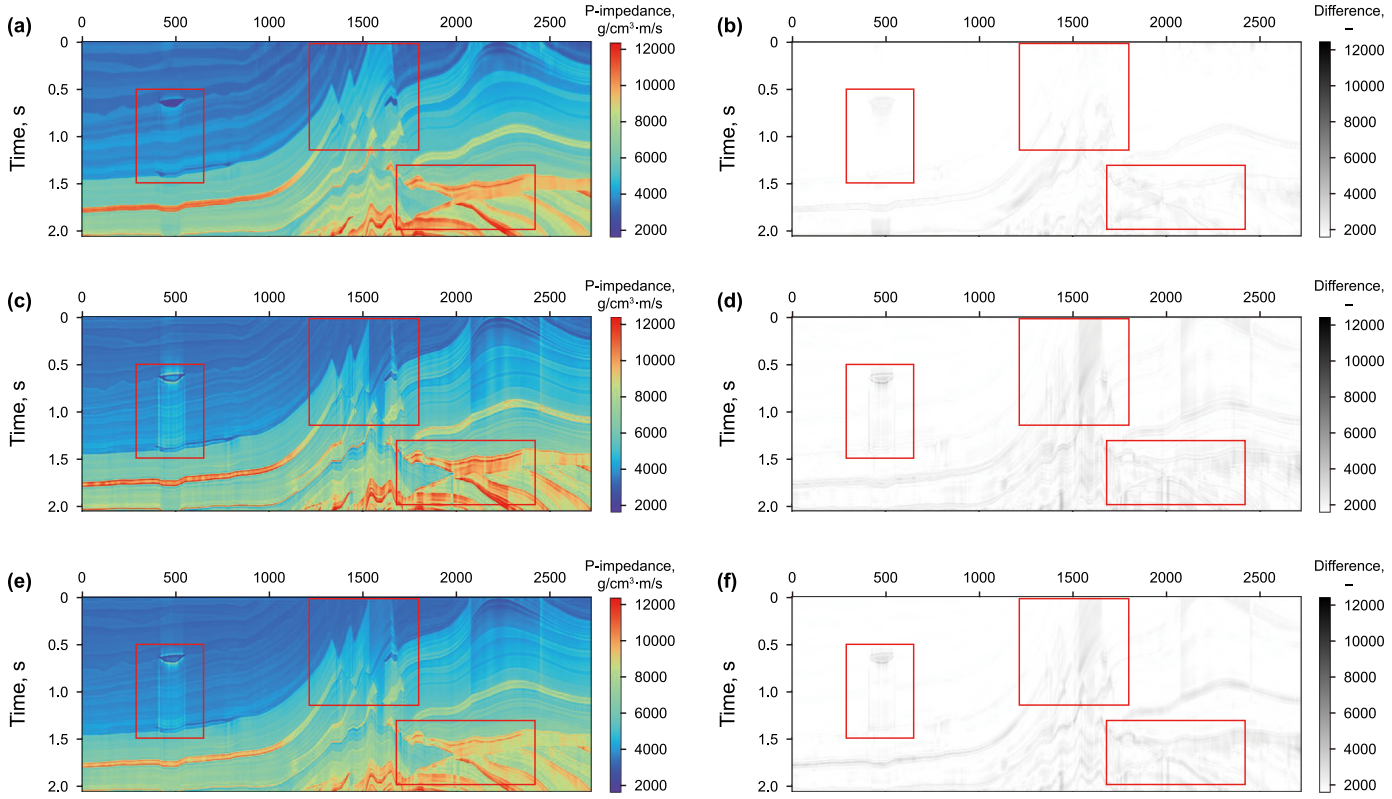

Fig. 8. 2-D synthetic model examples. **(a)** Proposed method. **(b)** Absolute difference between Fig. 8(a) and Fig. 7(a). **(c)** SM-net. **(d)** Absolute difference between Fig. 8(c) and Fig. 7(a). **(e)** SM-net using low-frequency model. **(f)** Absolute difference between Fig. 8(e) and Fig. 7(a).

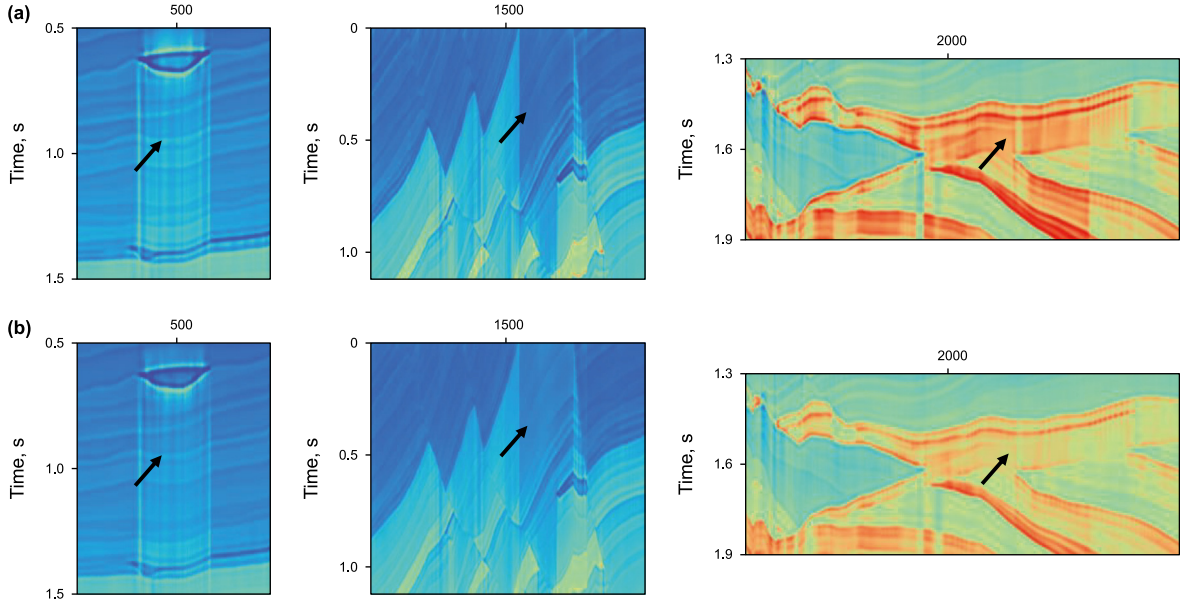

Fig. 9. A detailed comparison of the inversion results within the red rectangle: **(a)** and **(b)** correspond to the results shown in Fig. 8(c) and (e).

structures is small, while that of deep complex structures is large. From Fig. 12, the inversion accuracy of semi-supervised learning is higher than supervised learning, especially at CDP 1600. Because the modeling operator can limit the network to learn the relationship between broadband impedance and narrowband seismic data. In addition, to demonstrate the reliability of the proposed method for uncertainty prediction, we provide comparisons to other uncertainty estimation methods, such as Bayesian variational inference implemented by SVGD (Zhang and Curtis, 2020, 2021; Zhao et al., 2022). For each inversion, we utilize 500 particles that are initially generated from the initial model to implement SVGD (shown in Fig. 7(d)) and updated for 500 iterations. We use a uniform distribution over an interval of 5000 g/cm^3·m/s at each depth as a weaker prior distribution of impedance. Fig. 13 shows the

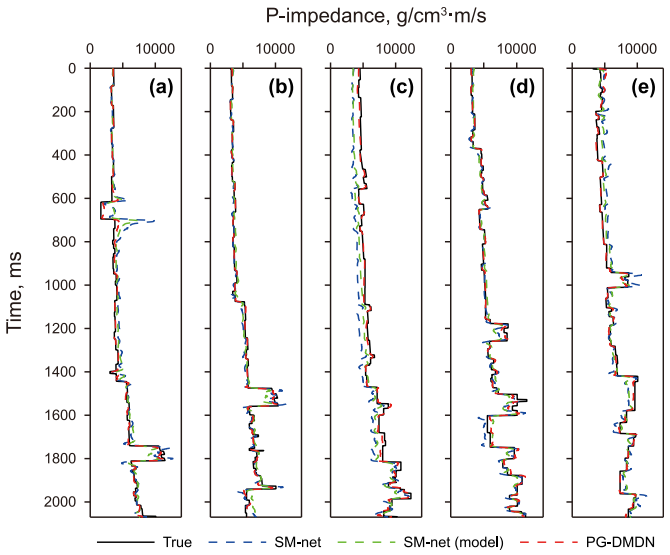

Fig. 10. The comparison between the inversion impedance and the validation data. **(a)** CDP 500. **(b)** CDP 1100. **(c)** CDP 1600. **(d)** CDP 1900. **(e)** CDP 2450.

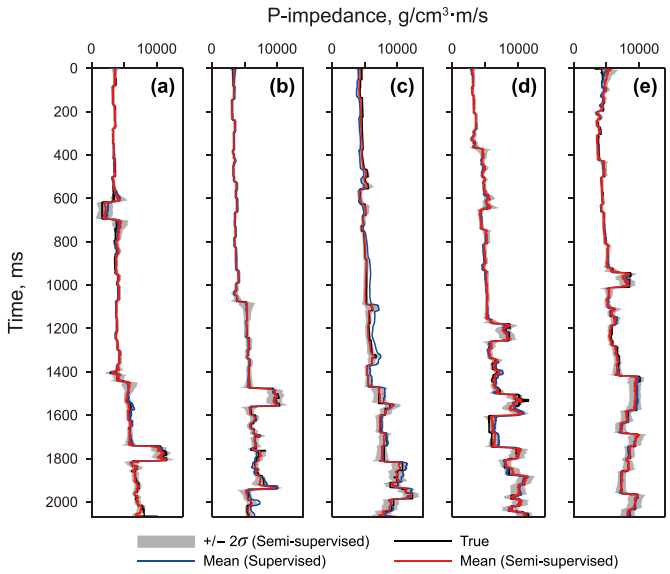

Fig. 12. The comparison between the inverted impedance of supervised and semi-supervised learning scheme and the validation data. **(a)** CDP 500. **(b)** CDP 1100. **(c)** CDP 1600. **(d)** CDP 1900. **(e)** CDP 2450.

Table 1
Quantitative evaluation of inversion results of different methods.

Method	PCC	R^2	M-SSIM
PG-DMDN	0.9853	0.9736	0.9447
SM-net (no model)	0.9746	0.9211	0.8756
SM-net (model)	0.9776	0.9372	0.9033

comparison between the particle density estimated by SVGD and the predicted PDF of the proposed method. In comparison, the posterior PDF predicted by semi-supervised learning is more similar to the particle density. Then, we compare the predicted PDF with the real distribution, as shown in Fig. 14. The posterior PDF

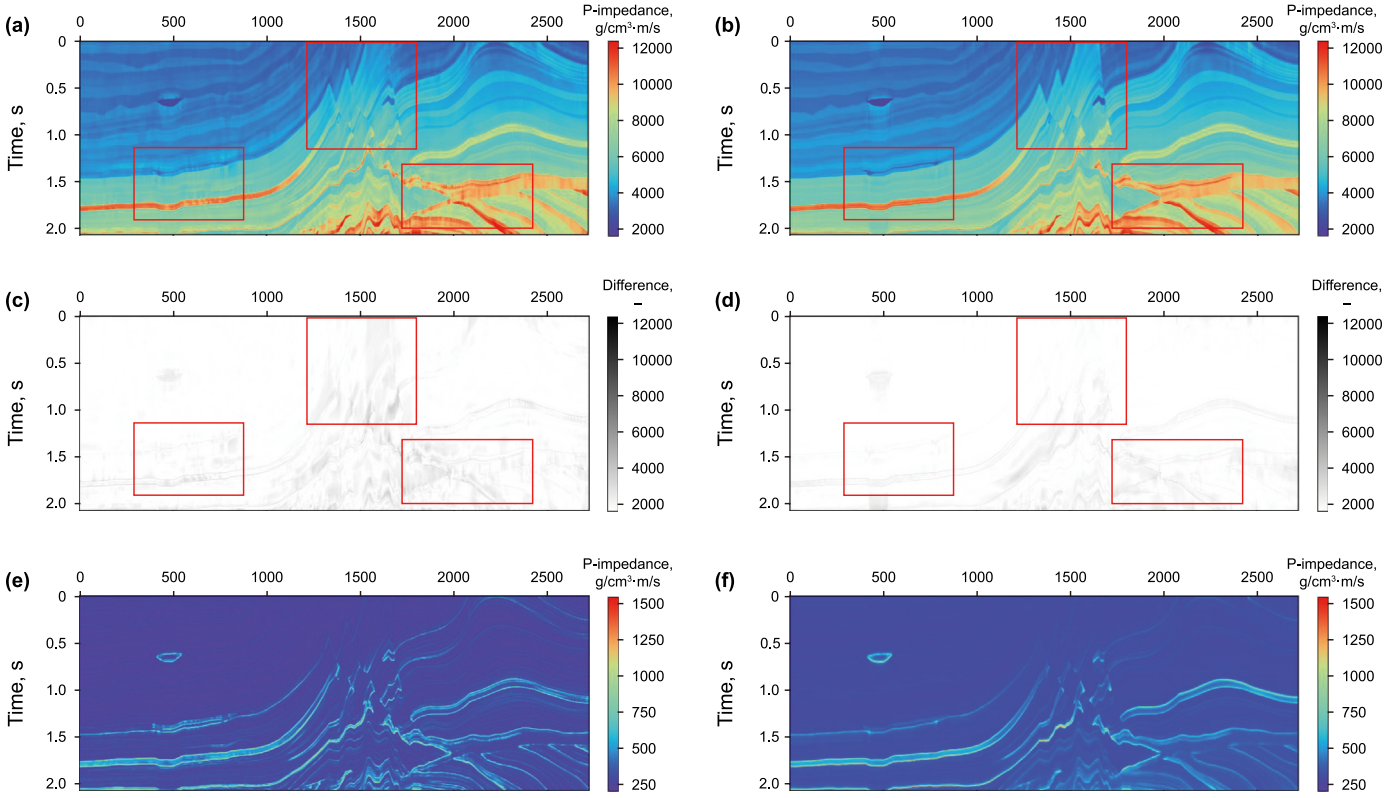

Fig. 11. Comparisons of inversion results between the supervised and semi-supervised learning. **(a)** The estimated impedance of supervised learning. **(b)** The estimated impedance of semi-supervised learning. **(c)** The absolute difference between inverted impedance of supervised learning and true impedance. **(d)** The absolute difference between inverted impedance of semi-supervised learning and true impedance. **(e)** The uncertainty predicted by supervised learning. **(f)** The uncertainty predicted by semi-supervised learning.

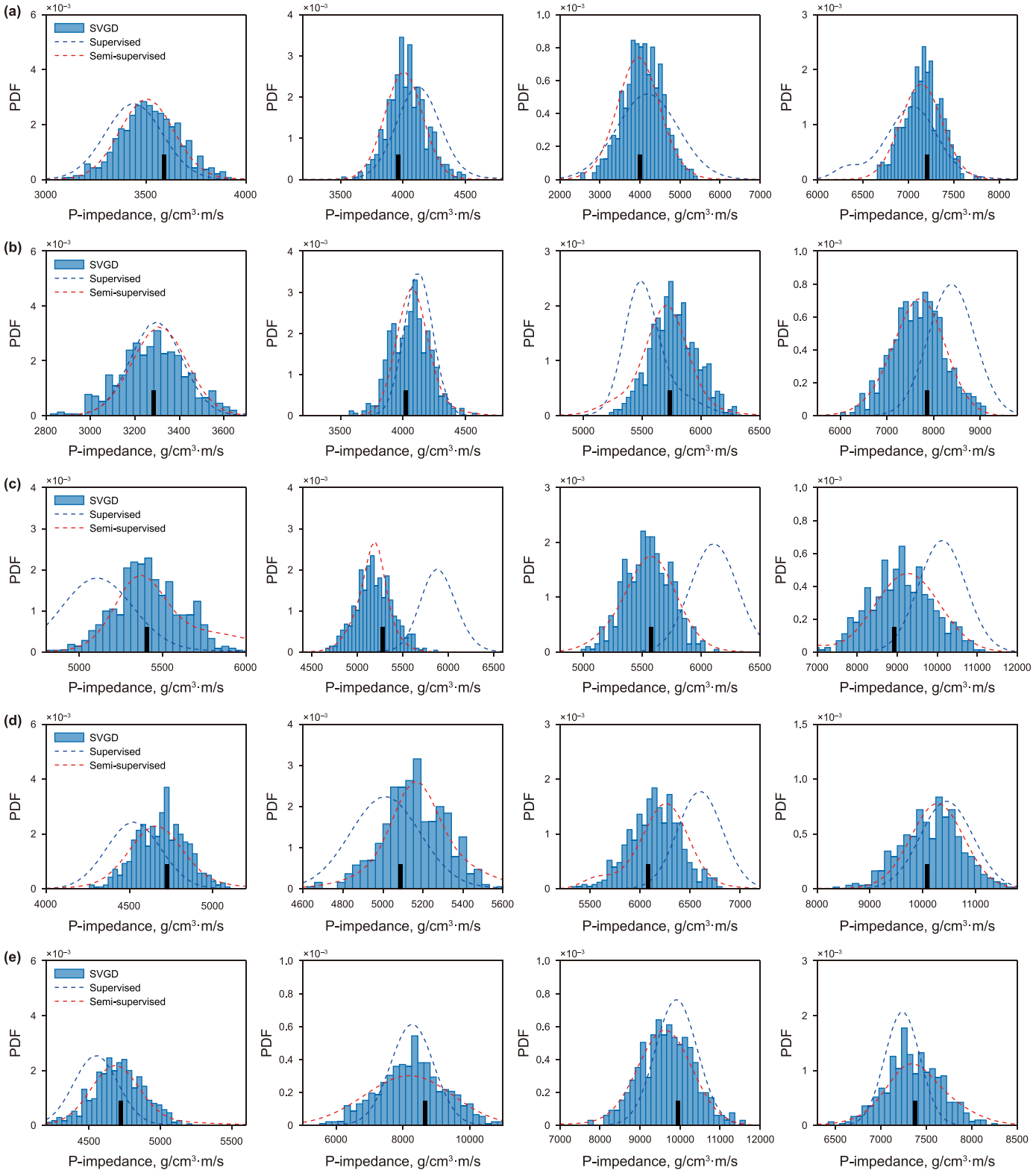

Fig. 13. Four estimated posterior PDFs of impedance at depths of 480, 960, 1440, 1920 ms. **(a)** CDP 500. **(b)** CDP 1100. **(c)** CDP 1600. **(d)** CDP 1900. **(e)** CDP 2450.

predicted by semi-supervised learning are more consistent with the true distribution than supervised learning, especially at CDP 1600. The geological information at CDP 1600 is completely different from other regions. There are few learning samples for supervised learning, and the mapping features of CDP 1600 cannot be learned. In addition, the quantitative evaluation of the role of forward part are reported in Table 2. When $\lambda_1 = 1$ and $\lambda_2 = 0$ (supervised learning), the network learns from only 10 seismic traces and the corresponding impedance traces from well logs. Therefore, compared with the unsupervised scheme, it can produce better

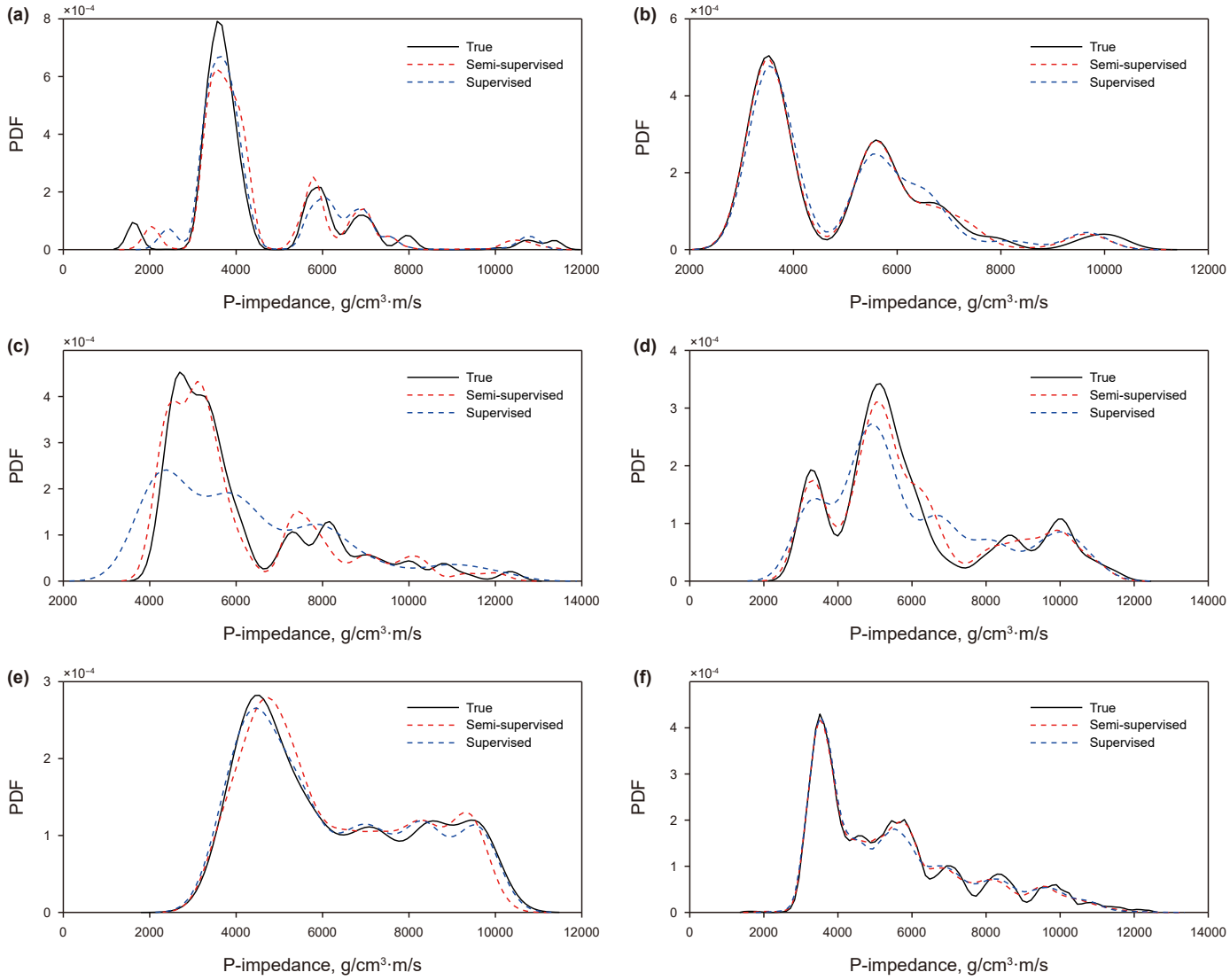

Fig. 14. Comparison between predicted PDF of PG-DMDN and DMDN. **(a)** CDP 500. **(b)** CDP 1100. **(c)** CDP 1600. **(d)** CDP 1900. **(e)** CDP 2450. **(f)** All data.

Table 2
Quantitative evaluation of the role of the forward part.

Training scheme	PCC	R^2	M-SSIM	σ_{mean}
Supervised ($\lambda_1 = 1, \lambda_2 = 0$)	0.9761	0.9586	0.9315	287.1431
Semi-supervised ($\lambda_1 = 1, \lambda_2 = 1$)	0.9853	0.9736	0.9447	274.6873

results. However, training a deep inverse model in a supervised learning scheme requires heavy regularization and careful selection of the training parameters (Alfarraj and AlRegib, 2019). Finally, when $\lambda_1 = 1$ and $\lambda_2 = 1$ (semi-supervised learning), the network is learned from all seismic data and 10 impedance traces from well logs. The average standard deviation σ_{mean} of supervised learning is slightly larger than that of semi-supervised learning. Hence, the semi-supervised scheme improves the performance and regularizes the learning. All the results demonstrate that semi-supervised has better performance than supervised learning in small sample problems.

3.4. The robustness of PG-DMDN inversion

Seismic data with signal to noise ratio (SNR) of 10, 5 and 2 are used to evaluate the robustness of PG-DMDN. Ten black impedance curves (Fig. 7(a)), seismic data profile (Fig. 7(b) and (c)) and low-frequency model (Fig. 7(d)) are utilized as training data. The learning rate is 0.005. Fig. 15 shows the estimated impedance with different noise levels. Fig. 15(a),(c),(e) show the inversion results of the PG-DMDN. Fig. 15(b),(d),(f) show the absolute difference between the estimated impedance and the real impedance. However, the existence of noise reduces the horizontal continuity of inversion results. Fig. 16 shows the estimated impedance curves at CDP 500, 1100, 1600, 1900 and 2450 with different SNR. Fig. 17 shows the predicted posteriori PDF with different noise. The predicted probability distribution is approximately consistent with the true impedance value. The corresponding quantitative evaluation of the inversion results on Marmousi II is reported in Table 3. The existence of noise increases the uncertainty of inversion. From all the results, it can be concluded that PG-DMDN still has good performance even when the SNR is 2.

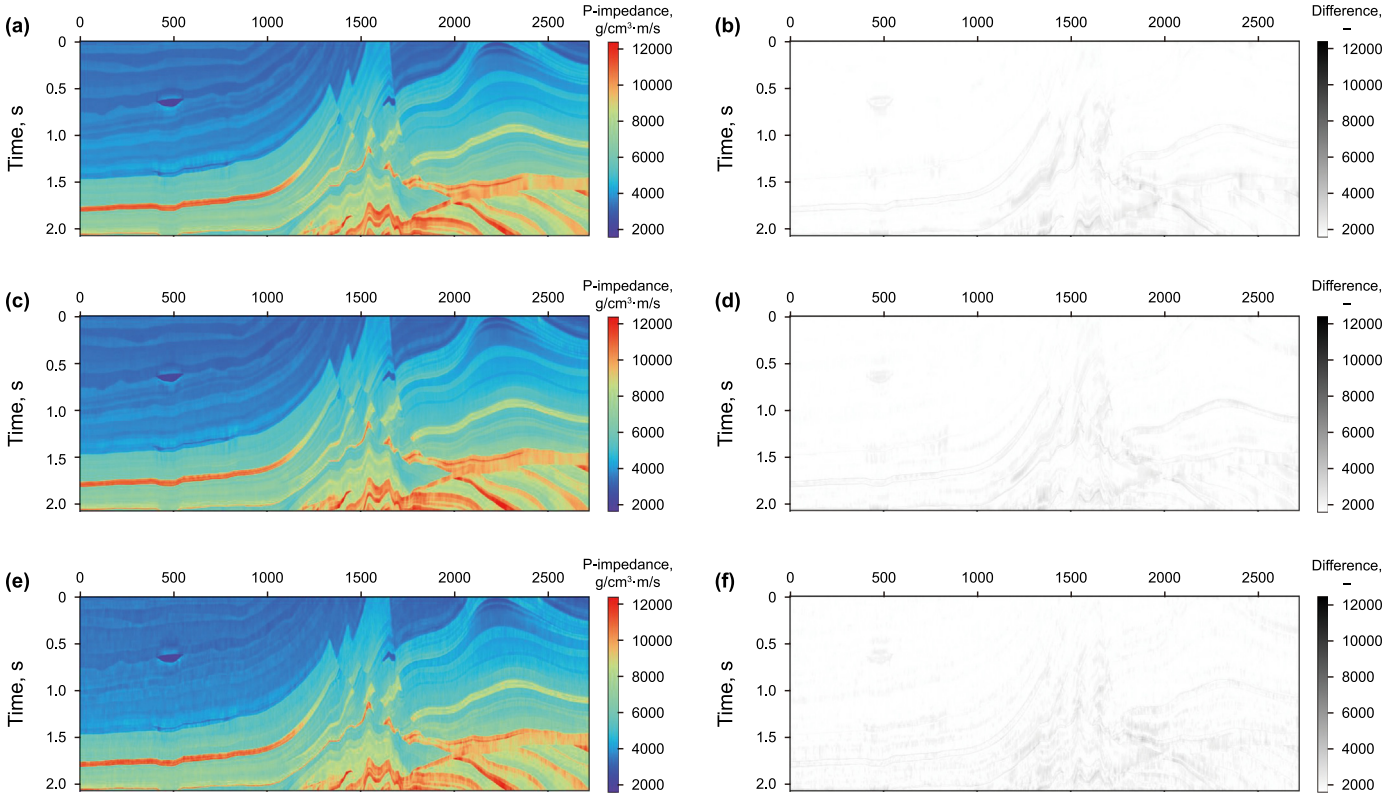

Fig. 15. Inverted impedance of semi-supervised learning inversion with different SNR. **(a)** SNR = 10. **(b)** SNR = 5. **(c)** SNR = 2.

3.5. Influence of low-frequency on PG-DMDN inversion

The influence of low-frequency model on inversion is self-evident. To quantify the effect of low-frequency on PG-DMDN inversion, the results are compared with those without low-frequency model. The setting of network parameters is also consistent with the previous test. Fig. 18(a) shows the estimated impedance profiles utilizing low-frequency models as training data. Fig. 18(c) shows the inversion result without low-frequency model. Fig. 18(b) and (d) show the corresponding absolute difference profiles. From Fig. 18(c) and (d), the inversion results without low-frequency model are basically consistent with the real model, except in the red rectangle. The 10 selected training samples do not contain the relevant low-frequency features from CDP 1520 to 1650. Without the help of low-frequency model, the network does not fully learn the nonlinear relationship between narrowband seismic data and broadband impedance.

However, due to the existence of modeling operator $F(\mathbf{R})$, the network can still learn the some low-frequency features, which is consistent with the conclusion in Section 3. The convolution model is an approximate linear operator, so some features will be lost. The corresponding quantitative evaluation of the inversion results is reported in Table 4. From all the results, it can be seen that the proposed method can still obtain good results without the participation of low-frequency models. Of course, as with the traditional seismic inversion method, the high accuracy low-frequency model is also useful for the DL-based inversion method. How to obtain a reliable low-frequency model in the case of small samples is also a better research direction.

3.6. Smaller samples problem

In seismic inversion based on DL, the small samples problem caused by sparse wells has attracted much attention. In this test, we try to test the ability of the PG-DMDN inversion to solve the smaller sample problem. Low-frequency model (Fig. 7(c)) and only five pseudo-well curves (the dashed red lines in Fig. 7(a)) are used as training data. Fig. 18(a) shows the estimated impedance profiles. The absolute difference between the estimated impedance and the true impedance is shown in Fig. 19(b). The PCC is 0.9805, the R^2 is 0.9656, the SSIM is 0.9404, and the σ_{mean} is 338.9038. The inversion results can still accurately describe the geological structure, and the error is acceptable. Fig. 20 shows the comparison between the estimated parameter and the validation data. The inverted impedance is consistent with the real impedance, but the inversion uncertainty of complex structure area increases. Fig. 21 displays the predicted posteriori PDF of impedance. The predicted posteriori PDF is consistent with the real distribution. Comprehensive analysis demonstrates that the proposed PG-DMDN inversion method still performs well in small sample problems.

4. Field data example

The semi-supervised probabilistic inversion method proposed is validated using the actual seismic data of the South China Sea. The frequency band of seismic data is from approximately 6 to 60 Hz, and the interval sampling is 2 ms. The CDP number is 1899, the sampling point is 500, and sampling interval is 2 ms. As shown in Figs. 22(a) and 6 well-connecting lines are selected as the test profile. The amplitude energy on the right is stronger than that on the left, which is due to the development of limestone. Well A, B, C, D, E and corresponding multi-trace seismic data are used as labeled data, and well F is validation data. The remaining seismic traces are used as unlabeled data. Two types of data are used as training sets for field data example. The sampling interval of well-log data is 0.2 ms, which is one tenth of seismic data. Fig. 22(b) shows the low-

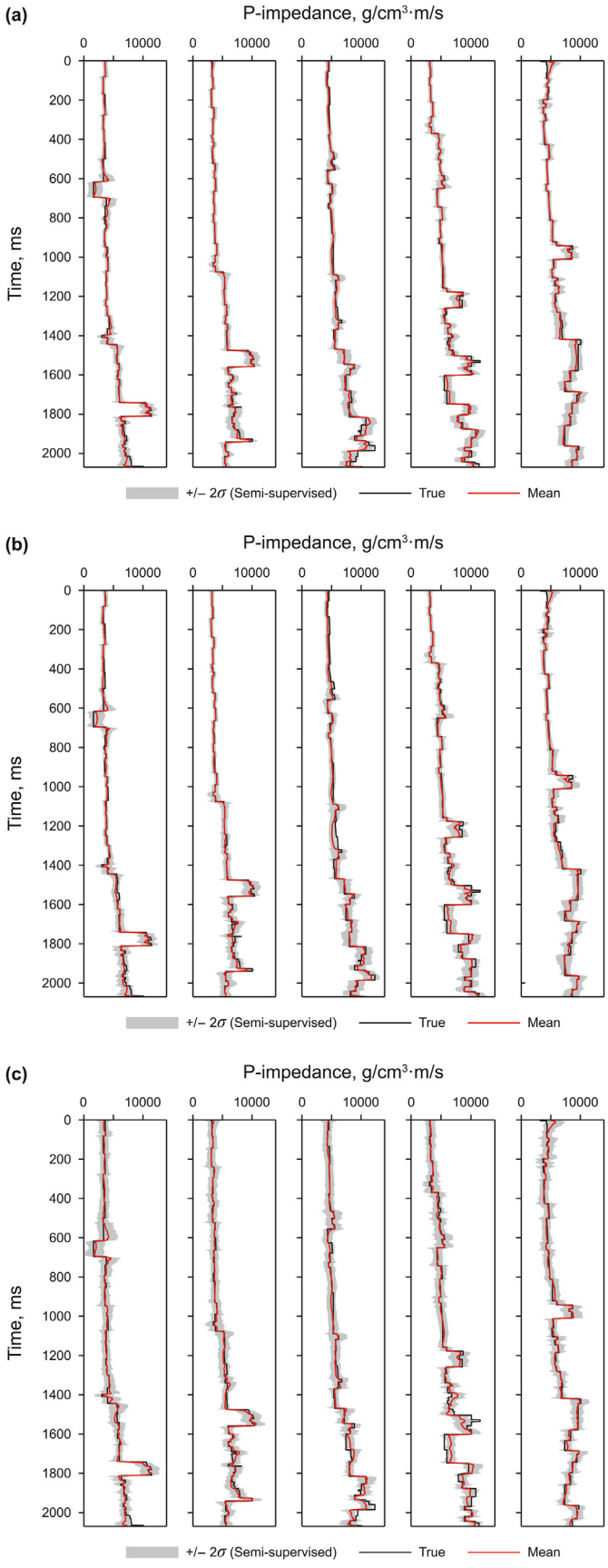

Fig. 16. Comparison between the inverted impedance and validation data with different SNR. **(a)** SNR = 10. **(b)** SNR = 5. **(c)** SNR = 2.

frequency impedance model established by horizontal interpolation of well curves. Well F is not used to establish the low-frequency model. In the training process, the number of iterations is 5000, and the learning rate is 0.003. The rectangular seismic data are used for training network. Therefore, the size of the input data is the same. For missing logging data, we use the inversion results obtain by seismic inversion in the complex domain (Zong et al., 2018) to repair it. The parameters of the network are set as follows: the batch size is $b = 10$; the input feature number of B-GRUs is $c = 8$; the input feature number of CNNs is $c_1 = 8$; the output feature number of CNNs is $c_2 = 16$; the Gaussian kernel number is $k = 3$; the kernel size of CNNs is $m = 3$; and the dilation factors are $d_1 = 1$, $d_2 = 3$ and $d_3 = 5$, respectively. From Eq. (16), it can be seen that parameters λ_1 and λ_2 can be used to control the effect of labeled logging data and unlabeled seismic data on the loss function. Through multiple experiments, the optimal weight coefficient can be obtained to achieve optimal performance of the network. In this test, we set $\lambda_1 = 1$ and $\lambda_2 = 0.8$.

Fig. 23 shows a comparison between PG-DMDN inverted impedance result (Fig. 23(a)) and the SM-net inverted impedance result (Fig. 23(b)). From Fig. 23, the both inverted impedances have good horizontal continuity and high resolution. The inversion results can describe the limestone body well on the right side, that is, the low value part. However, SM-net gets the high value at the red ellipse. The seismic residual between the synthetic and real seismic profile is shown in Fig. 23(c) and (d). The synthetic seismic data is obtained by feeding the PG-DMDN inversion result into the forward modeling part. We can observe that both residuals in Fig. 23(c) and (d) are small in contrast to the seismic data of Fig. 22(a). Thus, both the generated seismic data are close to the real seismic data. The small residual shows that the forward modeling part is effective.

To better comparison, Fig. 24(a) exhibits the comparison between the estimated impedance and the real impedance at well F. The green solid line represents the low-frequency model, the black solid line indicates the true value, the blue solid line denotes the inverted impedance with the SM-net, the red solid line denotes the inverted impedance with the PG-DMDN, and the grey area is predicted probability distribution. The estimated mean impedance of PG-DMDN is well consistent with the real impedance, especially at about 3.85 s, while the estimated impedance of SM-net is inconsistent with the real impedance due to the lack of low-frequency. In addition, the real impedance curve is within the predicted probability distribution range (grey area). Fig. 24(b) shows the comparison between the particle density estimated by SVGD and the predicted PDF of the proposed method. In comparison, the posterior PDF predicted by semi-supervised learning is similar to the particle density. Fig. 24(c) displays the corresponding probability density distribution of impedance. The black solid line denotes the PDF of the real impedance, and the black dashed line represents the PDF of the predicted impedance. Obviously, PG-DMDN can well estimate the posterior probability distribution of impedance. In order to quantify the effect of different combinations of training sets on the inversion results, five additional tests are implemented. Well A, B, C, D and E are used as verification wells, and the remaining wells are used as training labels. At the same time, the verification well does not participate in the construction of low-frequency model to verify the feasibility of the proposed method. Fig. 25 shows the inversion results of impedance of different training sets. It can be seen from Fig. 25 that the inversion results of different training sets are almost the same, which proves the stability of the method. The quantitative evaluation results of different as validation wells are reported in Table 5. It can be seen from Table 5 that the errors of inversion results of different combinations are similar. All the above results verify the effectiveness of the proposed method.

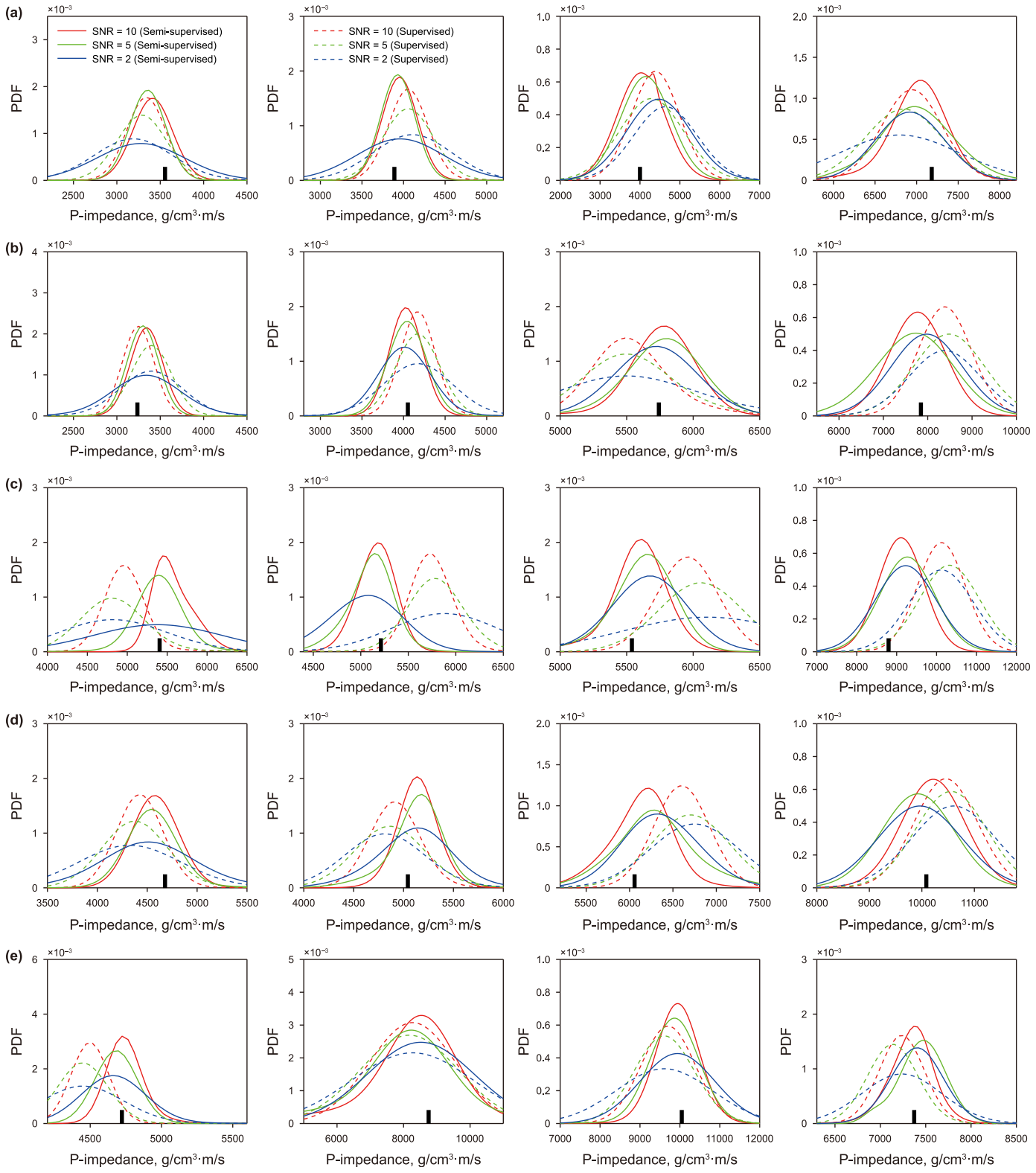

Fig. 17. Four estimated posterior PDFs of impedance at depths of 480, 960, 1440, 1920 ms with different noise. **(a)** CDP 500. **(b)** CDP 1100. **(c)** CDP 1600. **(d)** CDP 1900. **(e)** CDP 2450. Black bars represent the true value.

5. Discussion

Different from some existing semi-supervised deterministic seismic inversion methods, we study a probabilistic DL method combined with convolution model. MDN can simulate the probability distribution of arbitrary data, so we introduce it into semi-supervised deterministic seismic inversion. Correspondingly, the likelihood function is used instead of MSE to describe the difference

Table 3
Quantitative evaluation of the PG-DMDN with different SNR.

Method	SNR	PCC	R^2	M-SSIM	σ_{mean}
PG-DMDN	10	0.9809	0.9710	0.9357	319.9051
	5	0.9806	0.9699	0.9271	347.8565
	2	0.9756	0.9602	0.8797	369.2706

Table 4
Quantitative evaluation of the influence of the low-frequency model.

Model type	PCC	R^2	M-SSIM	σ_{mean}
Low-precision	0.9893	0.9766	0.9447	274.6873
None	0.9805	0.9568	0.9370	346.4596

between synthetic and real seismic data, thus helping to update the network parameters of inversion part. The block low-frequency model is introduced in the training process of the network to learn the nonlinear relationship between narrowband seismic data and broadband impedance. To improve the horizontal continuity, we added a block training strategy to the framework and modify the corresponding network structure. Based on the above four aspects, an inversion method based on PG-DMDN is proposed. Five model tests are utilized to explain the feasibility of the proposed method. The first test verifies the advantages of PG-DMDN. The PG-DMDN can extract spatial information from seismic data and improve the lateral continuity of inversion results. The second test verifies the role of the forward modeling part. This means that PG-DMDN has learned the physical laws of seismic forward modeling. The third test verifies the anti-noise performance of the proposed method. The existence of noise may increase the uncertainty of inversion results. Because there are few well-log data available for training, and other training samples are seismic data. The fourth test demonstrates the importance of the low-frequency information. Low-frequency model is helpful to capture the physical law between seismic data and low-frequency impedance. The fifth tests the ability of the proposed method to solve the problems with only five pseudo-wells. The inversion results show that only using 5 pseudo wells as training data can still obtain good results. Finally, the ability of proposed PG-DMDN inversion method to solve practical problems is verified by a field example. It can be seen from the results that PG-DMDN inversion method can not only accurately estimate the results, but also predict the uncertain information in seismic inversion directly.

Although the proposed method performs well, there are still some limitations. The proposed method still requires an initial model and wavelet. It is difficult to fully learn the nonlinear mapping relationship between narrowband seismic data and broadband impedance only by relying on a small amount of logging data. Therefore, a low-frequency model is needed to assist PG-DMDN in learning this relationship. The accuracy of wavelets also affects the final inversion results. It may be a better method to replace the forward operator with neural operator. In addition, the proposed method can be easily extended to pre-stack inversion to estimate elastic and physical parameters. For three parameter inversion, we only need to set in Eq. (4). The input data becomes pre-stack seismic data, and the label data is a well curve of elastic or physical parameters. The forward operator will become an AVO accurate or approximate reflection coefficient. Based on the above changes, the proposed method can be used for AVO inversion.

6. Conclusion

The conventional deterministic seismic inversion method based on deep learning can only output deterministic value, which cannot reflect the uncertainty of seismic inversion. To address this issue, based on the SM-net, a probabilistic seismic inversion method based on physics-guided deep mixture density network is proposed. The proposed method can capture the uncertainty information of seismic inversion on the premise of ensuring the accuracy. The low-frequency model in the training process can promote the prediction accuracy. The block constraints in the framework can improve the horizontal continuity. The feasibility and stability of proposed method are demonstrated by five synthetic data examples. The first test proves the advantages of the proposed method compared with the SM-net inversion method. The proposed inversion method can obtain more accurate results. The second test proves the validity of the modeling operator. The third test proves the anti-noise performance of the proposed method. The fourth test shows that the proposed method can still obtain a relatively good results without initial model. The fifth test demonstrates that the proposed method still performs well in the case of less label data. The inversion results and the predicted probability distributions are consistent with the validations. In

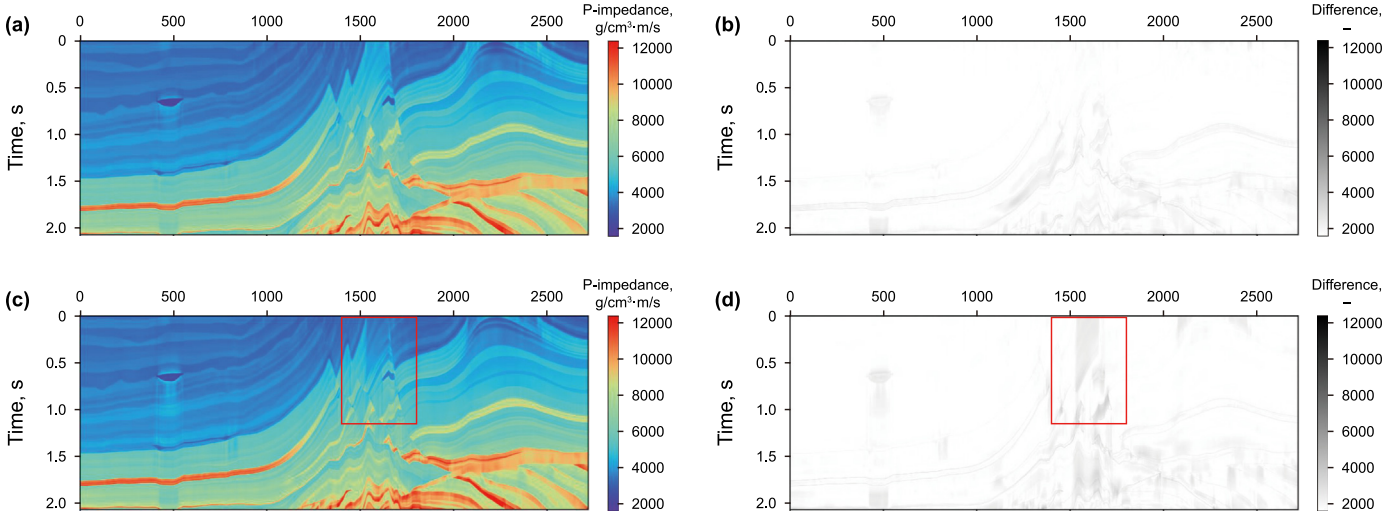

Fig. 18. Inverted impedance profiles. **(a)** Estimated impedance with low-precision model, and **(b)** corresponding absolute difference. **(c)** Estimated impedance without model, and **(d)** corresponding absolute difference.

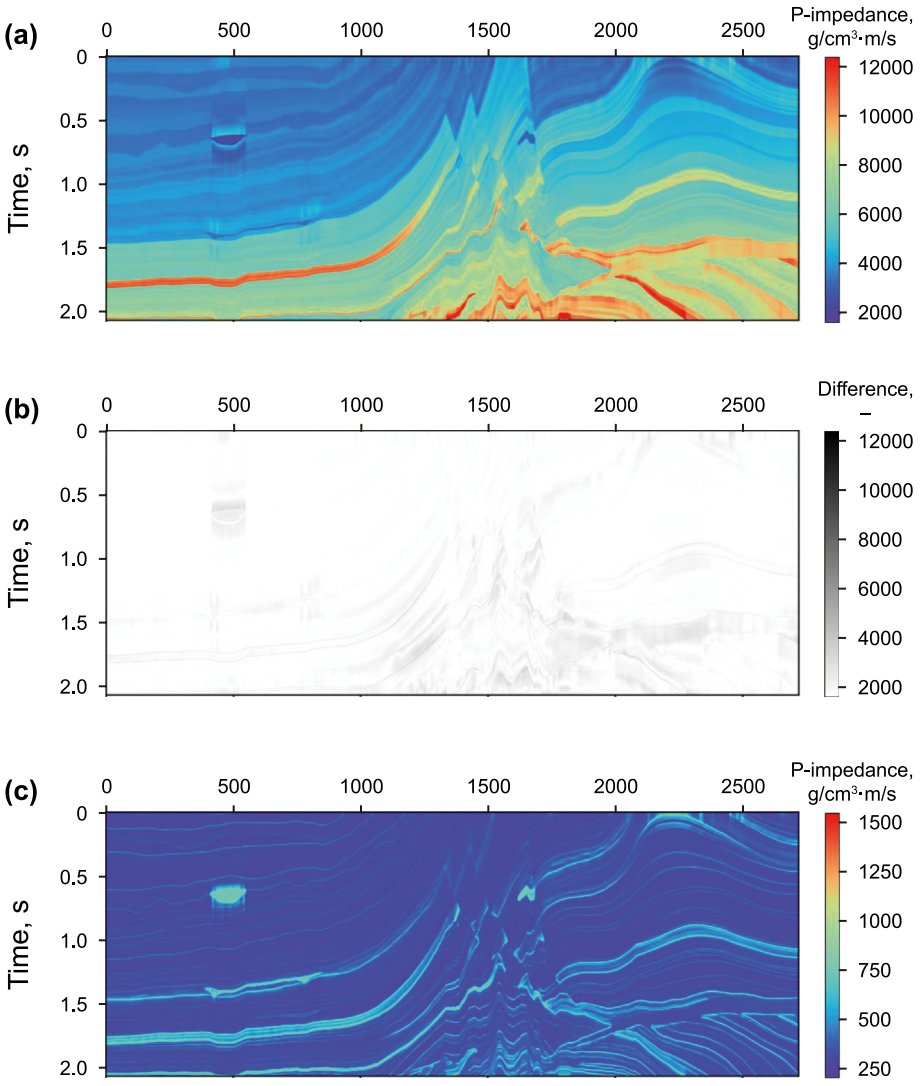

Fig. 19. PG-DMDN-based inversion results. **(a)** Estimated impedance profile using five pseudo-wells curves as training data. **(b)** Absolute difference between Figs. 19(a) and 18(b). **(c)** Predicted uncertainty.

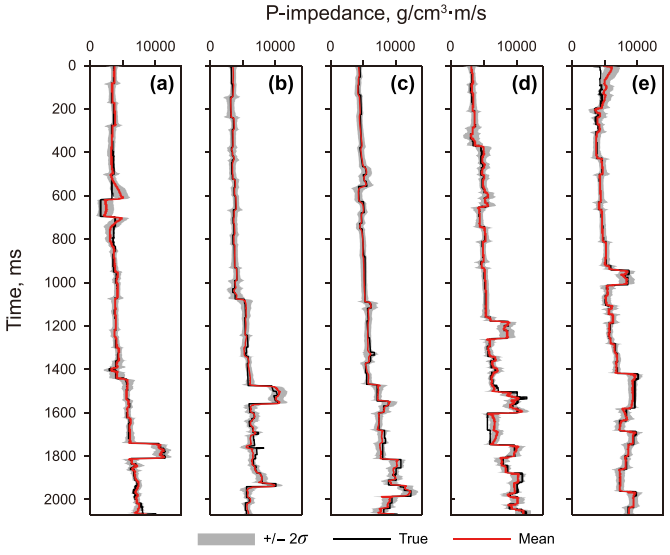

Fig. 20. Comparison between the estimated results and the validation data using five pseudo-wells curves as training data. **(a)** CDP 500. **(b)** CDP 1100. **(c)** CDP 1600. **(d)** CDP 1900. **(e)** CDP 2450.

addition, compared with the results of SVGD method, the reliability of prediction probability is verified. The field data test proves the feasibility of the method in practical application. The results are in good agreement with the real curves, which proves the effectiveness of the proposed method.

CRediT authorship contribution statement

Qian-Hao Sun: Formal analysis, Investigation, Methodology, Validation, Writing – original draft. **Zhao-Yun Zong:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing. **Xin Li:** Provide data, Data analysis, Application, Verification, Writing - editing.

Declaration of competing interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled.

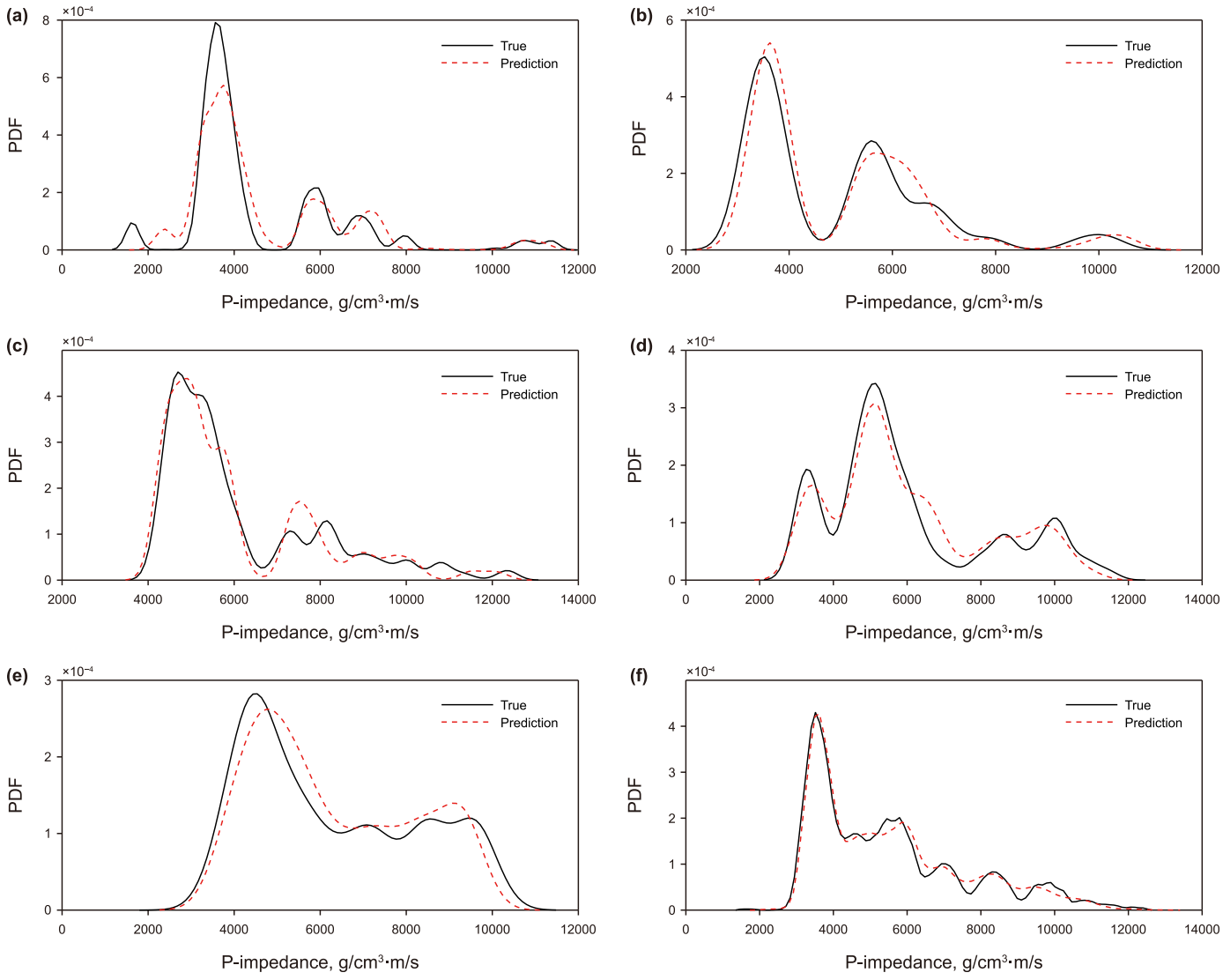

Fig. 21. Comparison between predicted PDF of PG-DMDN and DMDN. **(a)** CDP 500. **(b)** CDP 1100. **(c)** CDP 1600. **(d)** CDP 1900. **(e)** CDP 2450. **(f)** All data. The black line indicates the true probability distribution of impedance, and the red dashed line indicates the predicted probability density distribution of PG-DMDN.

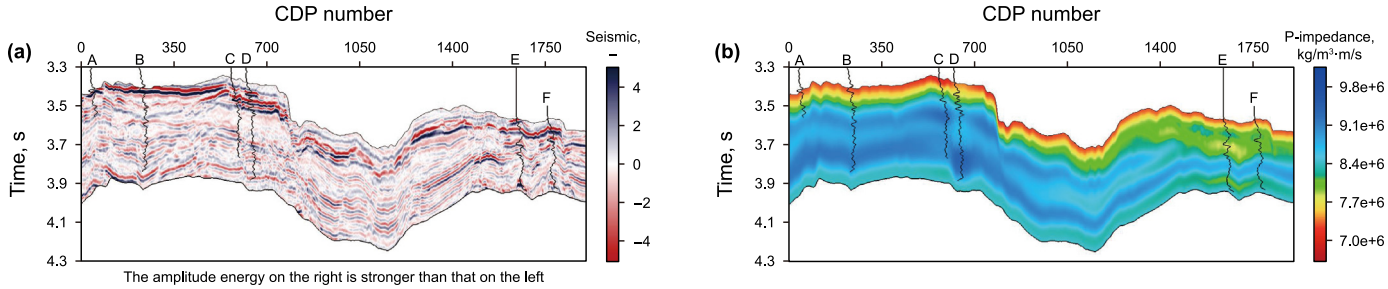

Fig. 22. Training data for PG-DMDN. **(a)** Profile of seismic data. **(b)** Profile of low-frequency impedance model.

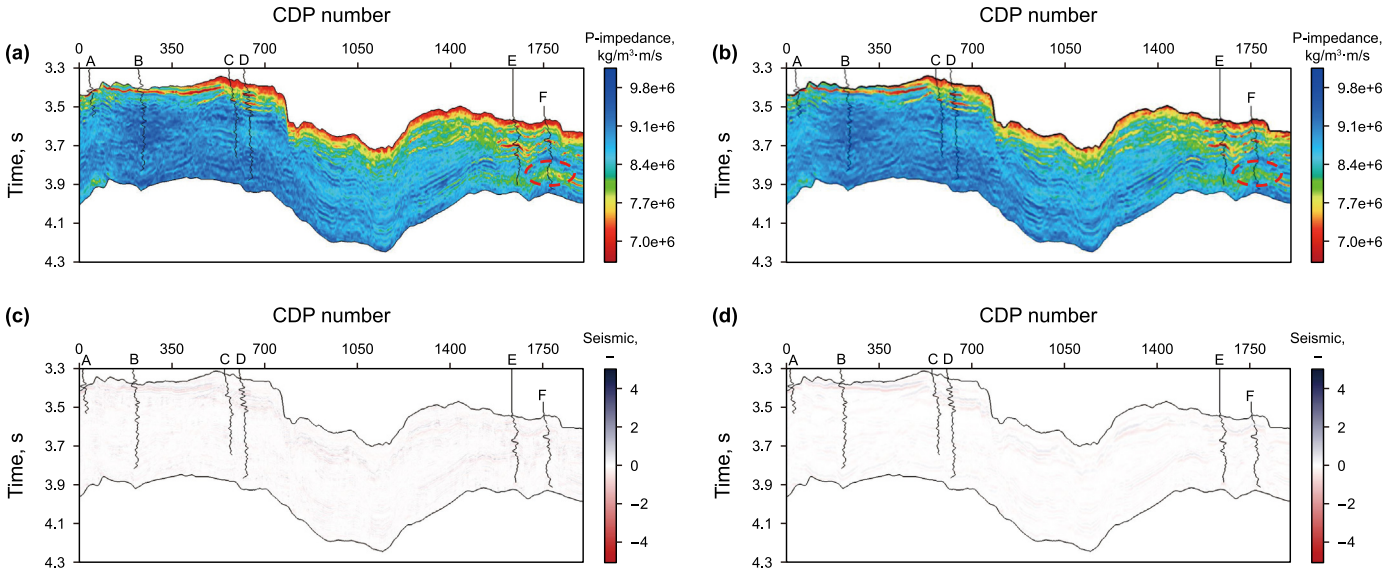

Fig. 23. Comparison between **(a)** the PG-DMDN inversion result and **(b)** the SM-net inversion result. **(c)** The residuals between Fig. 22(a) and the generated seismic data using Fig. 23(a). **(d)** The residuals between Fig. 22(a) and the generated seismic data using Fig. 23(b).

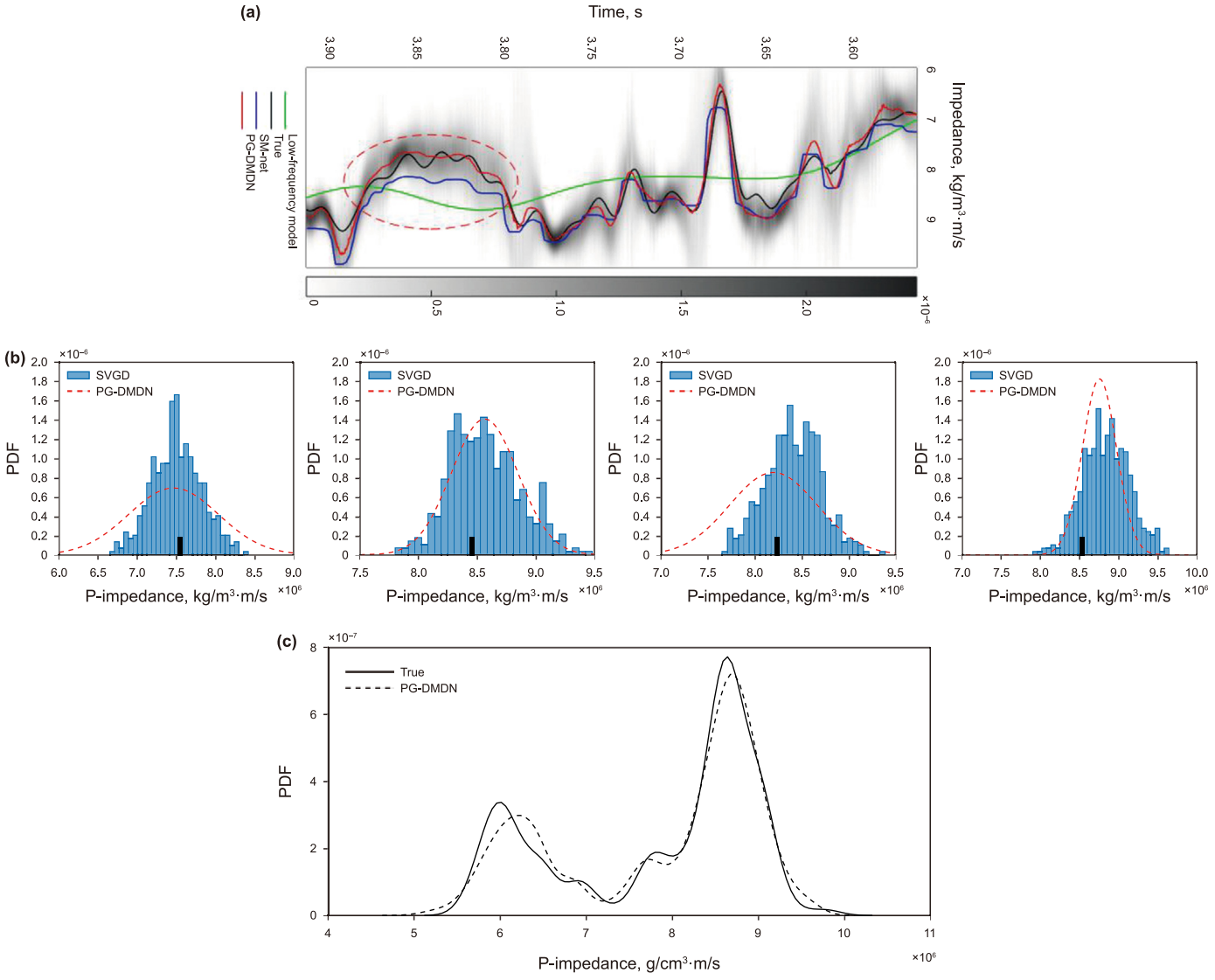

Fig. 24. PG-DMDN inversion results of field data. **(a)** The comparison between the estimated impedance and the real impedance. **(b)** Four estimated posterior PDFs of impedance at depths of 3.6, 3.7, 3.8, 3.9 s. **(c)** The comparison between the predicted probability density distribution and the real probability density distribution.

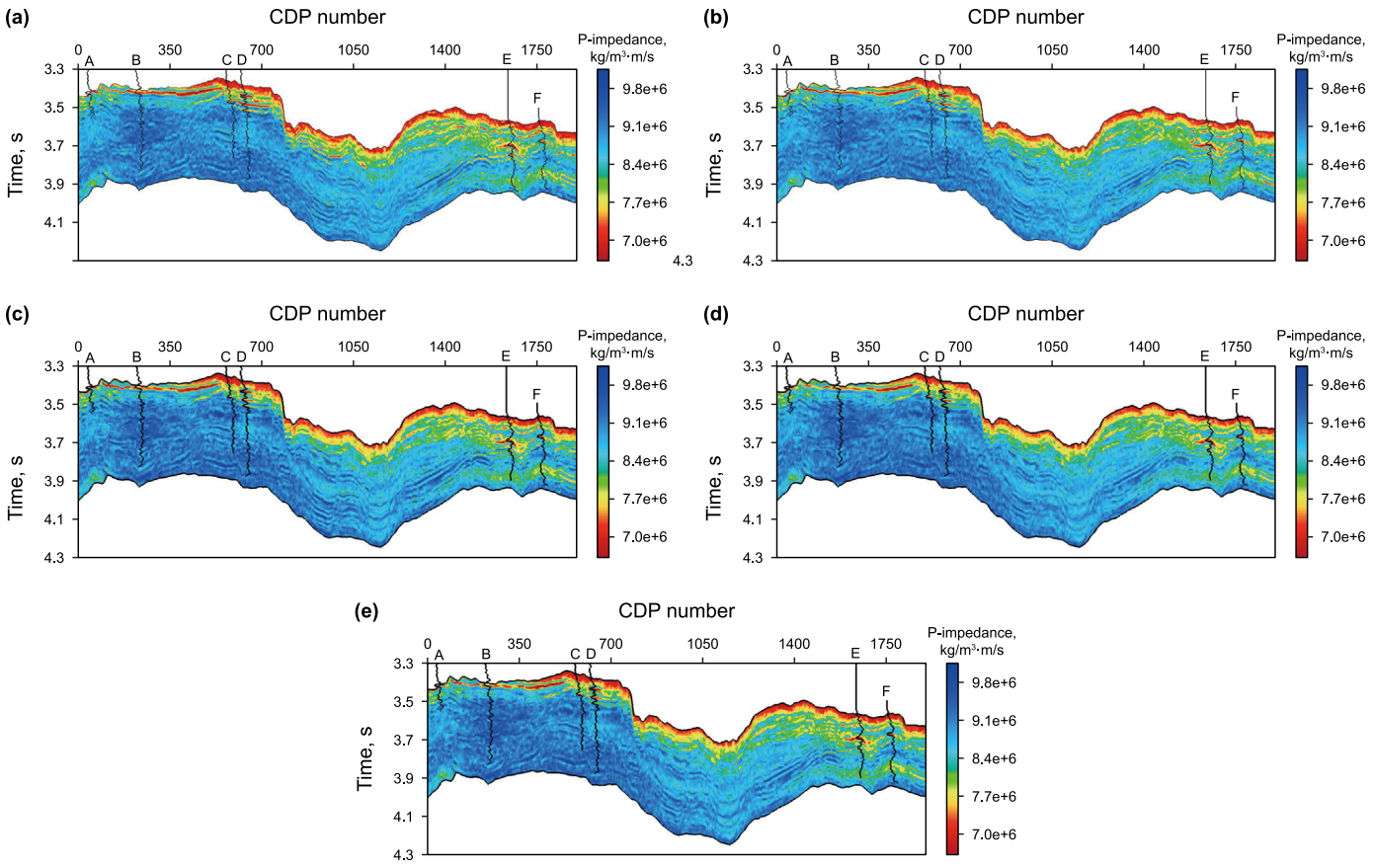

Fig. 25. Impedance profiles inverted from different training sets. **(a)** Well A is a verification well. **(b)** Well B is a verification well. **(c)** Well C is a verification well. **(d)** Well D is a verification well. **(e)** Well E is a verification wells.

Table 5
Quantitative evaluation of the influence of the different combinations of training data.

Verification well	PCC	R^2	MSE
A	0.9656	0.9443	0.0118
B	0.9569	0.9392	0.0168
C	0.9752	0.9505	0.0102
D	0.9587	0.9367	0.01728
E	0.9683	0.9492	0.0112
F	0.9628	0.9494	0.0129

Acknowledgements

We thank the editors and two anonymous reviewers for their comments and constructive suggestions that help to improve our work. We also thank the sponsorship of Shandong Province Foundation for Laoshan National Laboratory of Science and Technology Foundation (LSKJ202203400) and National Natural Science Foundation of China (42174139, 42030103) and Science Foundation from Innovation and Technology Support Program for Young Scientists in Colleges of Shandong Province and Ministry of Science and Technology of China (2019RA2136).

Appendix A

BN makes the distribution of input data in each layer of the network relatively stable and speeds up the learning speed. In addition, BN reduces the sensitivity of the inversion model to network parameters and makes the learning process more stable. Furthermore, BN allows the network to utilize saturation activation functions (e.g. sigmoid, tanh, etc.). BN is divided into the following steps (Ioffe and Szegedy, 2015):

1) Calculate the mean and variance of the batch input $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_b\}$ using the following equation:

$$\mu_{\mathrm{X}} = \frac{1}{b}\sum_{i}^{b}\mathbf{x}_i, \tag{A1}$$

and

$$\sigma_{\mathrm{X}}^2 = \frac{1}{b}\sum_{i=1}^{b}(\mathbf{x}_i - \mu_{\mathrm{X}})^2. \tag{A2}$$

2) Standardize the batch input with Eq. (A3):

$$\widehat{\mathbf{x}}_i = \frac{\mathbf{x}_i - \mu_{\mathrm{X}}}{\sqrt{\sigma_{\mathrm{X}}^2 + \varepsilon}}, \tag{A3}$$

where ε is to prevent invalid calculation when variance is zero.

3) Normalization reduces the representation ability of the network. Therefore, BN introduces two learnable parameters γ and β to solve this problem. A linear transformation is performed on normalized data $\widehat{\mathbf{x}}_i$ (Eq. (A4)).

$$\widehat{\mathbf{y}}_i = \gamma \widehat{\mathbf{x}}_i + \beta. \tag{A4}$$

In particular, when $\gamma^2 = \sigma^2$ and $\beta = \mu$, identity transform can be achieved and the distribution of the original input features can be preserved.

References

Alfarraj, M., AlRegib, G., 2019. Semisupervised sequence modeling for elastic impedance inversion. Interpretation 7 (3), SE237–SE249. https://doi.org/10.1190/Int-2018-0250.1.

Bishop, C.M., 1994. Mixture Density Networks. Neural Computing Research Group Technical Report NCRG/94/004. Aston University.

Biswas, R., Sen, M.K., Das, V., Mukerji, T., 2019. Prestack and poststack inversion using a physics-guided convolutional neural network. Interpretation 7 (3), SE161–SE174. https://doi.org/10.1190/INT-2018-0236.1.

Buland, A., More, H., 2003. Bayesian linearized AVO inversion. Geophysics 68 (1), 185–198. https://doi.org/10.1190/1.1543206.

Cao, D., Su, Y., Cui, R., 2022. Multi-parameter pre-stack seismic inversion based on deep learning with sparse reflection coefficient constraints. J. Petrol. Sci. Eng. 209, 109836. https://doi.org/10.1016/j.petrol.2021.109836.

Chen, Y., Saygin, E., 2021. Seismic inversion by hybrid machine learning. J. Geophys. Res. Solid Earth 126 (9), 126. https://doi.org/10.1029/2020JB021589.

Cho, K., Merrienboer, B.V., Gulcehre, C., Hdanau, D.B., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. Computer Science. https://doi.org/10.3115/v1/D14-1179.

Das, V., Mukerji, T., 2020. Petrophysical properties prediction from prestack seismic data using convolutional neural networks. Geophysics 85 (5), N41–N55. https://doi.org/10.1190/GEO2019-0650.1.

Das, V., Pollack, A., Wollner, U., Mukerji, T., 2019. Convolutional neural network for seismic impedance inversion. Geophysics 84 (6), R869–R880. https://doi.org/10.1190/GEO2018-0838.1.

Di, H., Abubakar, A., 2021. Estimating subsurface properties using a semisupervised neural network approach. Geophysics 87 (1), IM1–IM10. https://doi.org/10.1190/geo2021-0192.1.

Downton, J.E., 2005. Seismic Parameter Estimation from AVO Inversion. University of Calgary.

Earp, S., Curtis, A., 2020. Probabilistic neural network-based 2D travel-time tomography. Neural Comput. Appl. 32 (22), 17077–17095. https://doi.org/10.1007/s00521-020-04921-8.

Earp, S., Curtis, A., Zhang, X., Hansteen, F., 2020. Probabilistic neural network tomography across Grane field (North Sea) from surface wave dispersion data. Geophys. J. Int. 223 (3), 1741–1757. https://doi.org/10.1093/gji/ggaa328.

Feng, R., Grana, D., Balling, N., 2021. Variational inference in Bayesian neural network for well-log prediction. Geophysics 86 (3), M91–M99. https://doi.org/10.1190/geo2020-0609.1.

Gao, L., Chen, P.Y., Yu, S., 2016. Demonstration of convolution kernel operation on resistive cross-point array. IEEE Electron. Device Lett. 37 (7), 870–873. https://doi.org/10.1109/LED.2016.2573140.

Gao, Z., Li, C., Zhang, B., Jiang, X., Xu, Z., 2020. Building large-scale density model via a deep learning based data-driven method. Geophysics 86 (1), M1–M15. https://doi.org/10.1190/geo2019-0332.1.

Ge, M., Wang, W., Zheng, W., 2021. Semi-supervised Impedance Inversion by Bayesian Neural Network Based on 2-d CNN Pre-training. https://doi.org/10.48550/arXiv.2111.10596.

Gholami, A., 2015. Nonlinear multichannel impedance inversion by total-variation regularization. Geophysics 80 (5), R217–R224. https://doi.org/10.1190/geo2015-0004.1.

Gholami, A., 2016. A fast automatic multichannel blind seismic inversion for high-resolution impedance recovery. Geophysics 81 (5), V357–V364. https://doi.org/10.1190/GEO2015-0654.1.

Grana, D., 2016. Bayesian linearized rock-physics inversion. Geophysics 81 (6), D625–D641. https://doi.org/10.1190/geo2016-0161.1.

Grana, D., Rossa, E.D., 2010. Probabilistic petrophysical-properties estimation integrating statistical rock physics with seismic inversion. Geophysics 75 (3), O21–O37. https://doi.org/10.1190/1.3386676.

Hamid, H., Pidlisecky, A., 2015. Multitrace impedance inversion with lateral constraints. Geophysics 80 (6), M101–M111. https://doi.org/10.1190/geo2014-0546.1.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Hu, W., Jin, Y., Wu, X., Chen, J., 2020. Physics-guided self-supervised learning for low frequency data prediction in FWI. SEG Technical Program Expanded Abstracts, pp. 875–879. https://doi.org/10.1190/segam2020-3423396.1.

Ioffe, S., Szegedy, C., 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift. International conference on machine learning, PMLR 448–456. https://doi.org/10.48550/arXiv.1502.03167.

Junhwan, C., Seokmin, O., Joongmoo, B., 2022. Uncertainty estimation in AVO inversion using Bayesian dropout based deep learning. J. Petrol. Sci. Eng. 208, 109288. https://doi.org/10.1016/j.petrol.2021.109288.

Kingma, D., Ba, J., 2014. Adam: a method for stochastic optimization. Computer Science. https://doi.org/10.48550/arXiv.1412.6980.

Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86 (11), 2278–2324. https://doi.org/10.1109/5.726791.

Li, K., Yin, X., Zong, Z., 2017. Pre-stack Bayesian cascade AVA inversion in complex-Laplace domain and its application to the broadband data acquired at East China. J. Petrol. Sci. Eng. 158, 751–765. https://doi.org/10.1016/j.petrol.2017.09.005.

Li, K., Yin, X., Zong, Z., 2020. Facies-constrained prestack seismic probabilistic inversion driven by rock physics. Science China (Earth Sciences) 63 (6), 822–840. https://doi.org/10.1007/s11430-019-9578-1.

Maiti, S., Tiwari, R.K., 2010. Automatic discriminations among geophysical signals via the Bayesian neural networks approach. Geophysics 75 (1), E67–E78. https://doi.org/10.1190/1.3298501.

Martin, G.S., Wiley, R., Marfurt, K.J., 2006. Marmousi2: an elastic upgrade for Marmousi. Lead. Edge 25 (2), 156–166. https://doi.org/10.1190/1.2172306.

Röth, G., Tarantola, A., 1994. Neural networks and inversion of seismic data. J. Geophys. Res. Solid Earth 99 (B4), 6753–6768. https://doi.org/10.1029/93JB01563.

Shahraeeni, M.S., Curtis, A., 2011. Fast probabilistic nonlinear petrophysical inversion. Geophysics 76 (2), E45–E58. https://doi.org/10.1190/1.3540628.

Shahraeeni, M.S., Curtis, A., Chao, G., 2012. Fast probabilistic petrophysical mapping of reservoirs from 3D seismic data. Geophysics 77 (3), O1–O19. https://doi.org/10.1190/GEO2011-0340.1.

Song, L., Yin, X., Zong, Z., Jiang, M., 2022. Semi-supervised learning seismic inversion based on Spatio-temporal sequence residual modeling neural network. J. Petrol. Sci. Eng. 208, 109549. https://doi.org/10.1016/j.petrol.2021.109549.

Spikes, K., Mukerji, T., Dvorkin, J., Mavko, G., 2007. Probabilistic seismic inversion based on rock-physics models. Geophysics 72 (5), R87–R97. https://doi.org/10.1190/1.2760162.

Sun, H., Demanet, L., 2020. Extrapolated full-waveform inversion with deep learning. Geophysics 85 (3), R275–R288. https://doi.org/10.1190/geo2019-0195.1.

Sun, J., Innanen, K.A., Huang, C., 2021. Physics-guided deep learning for seismic inversion with hybrid training and uncertainty analysis. Geophysics 86 (3), R303–R317. https://doi.org/10.1190/geo2020-0312.1.

Sun, Q.H., Zong, Z., 2019. Amplitude variation with incident angle inversion for fluid factor in the depth domain. Ann. Geophys. 62 (5), SE562. https://doi.org/10.4401/ag-7882.

Wang, M., Xu, S., Zhou, H., 2020. Self-supervised learning for low frequency extension of seismic data. Society of Exploration Geophysicists SEG Technical Program Expanded Abstracts, pp. 1501–1505. https://doi.org/10.1190/segam2020-3427086.1.

Wang, Y., Niu, L., Zhao, L., Wang, B., He, Z., Zhang, H., Chen, D., Geng, J., 2021. Gaussian mixture model deep neural network and its application in porosity prediction of deep carbonate reservoir. Geophysics 87 (2), M59–M72. https://doi.org/10.1190/geo2020-0740.1.

Werbos, P.J., 1990. Backpropagation through time: what it does and how to do it. Proc. IEEE 78 (10), 1550–1560. https://doi.org/10.1109/5.58337.

Wu, B., Meng, D., Zhao, H., 2021. Semi-supervised learning for seismic impedance inversion using generative adversarial networks. Rem. Sens. 13 (5), 909. https://doi.org/10.3390/rs13050909.

Yu, F., Koltun, V., 2016. Multi-Scale Context Aggregation by Dilated Convolutions. https://doi.org/10.48550/arXiv.1511.07122.

Yuan, S., Jiao, X., Luo, Y., Sang, W., Wang, S., 2021. Double-scale supervised inversion with a data-driven forward model for low-frequency impedance recovery. Geophysics 87 (2), R165–R181. https://doi.org/10.1190/geo2020-0421.1.

Zhang, X., Curtis, A., 2020. Seismic tomography using variational inference methods. J. Geophys. Res. Solid Earth 125 (4), e2019JB018589. https://doi.org/10.1029/2019JB018589.

Zhang, X., Curtis, A., 2021. Bayesian full-waveform inversion with realistic priors. Geophysics 86 (5), A45–A49. https://doi.org/10.1190/geo2021-0118.1.

Zhao, X., Curtis, A., Zhang, X., 2022. Bayesian seismic tomography using normalizing flows. Geophys. J. Int. 228 (1), 213–239. https://doi.org/10.1093/gji/ggab298.

Zong, Z., Yin, X., Wu, G., 2012. AVO inversion and poroelasticity with P- and S-wave moduli. Geophysics 77 (6), N17–N24. https://doi.org/10.1190/geo2011-0214.1.

Zong, Z., Wang, Y., Li, K., Yin, X., 2018. Broadband seismic inversion for low-frequency component of the model parameter. IEEE Trans. Geosci. Rem. Sens. 56 (9), 5177–5184. https://doi.org/10.1109/TGRS.2018.2810845.