Petroleum Science 18 (2021) 1662–1674

Contents lists available at ScienceDirect

Petroleum Science

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

Original Paper

# Machine learning seismic reservoir prediction method based on virtual sample generation

Kai-Heng Sang, Xing-Yao Yin, Fan-Chang Zhang*

*School of Geosciences, China University of Petroleum (East China), Qingdao, Shandong 266580, China*

ARTICLE INFO

*Article history:*
Received 27 May 2020
Accepted 5 March 2021
Available online 21 September 2021

Edited by Jie Hao and Teng Zhu

*Keywords:*
Virtual sample
Machine learning
Reservoir prediction
Hypersphere characteristic equation

ABSTRACT

Seismic reservoir prediction plays an important role in oil exploration and development. With the progress of artificial intelligence, many achievements have been made in machine learning seismic reservoir prediction. However, due to the factors such as economic cost, exploration maturity, and technical limitations, it is often difficult to obtain a large number of training samples for machine learning. In this case, the prediction accuracy cannot meet the requirements. To overcome this shortcoming, we develop a new machine learning reservoir prediction method based on virtual sample generation. In this method, the virtual samples, which are generated in a high-dimensional hypersphere space, are more consistent with the original data characteristics. Furthermore, at the stage of model building after virtual sample generation, virtual samples screening and model iterative optimization are used to eliminate noise samples and ensure the rationality of virtual samples. The proposed method has been applied to standard function data and real seismic data. The results show that this method can improve the prediction accuracy of machine learning significantly.


## 1. Introduction

The seismic reservoir prediction method developed in 1970s and popularized after 1990s. It has been successfully applied in various oil and gas fields, and has played an important role in improving exploration efficiency and economic benefit (Ahmed et al., 2010; Liu et al., 2014; Gan et al., 2018). With the development of high precision oil and gas exploration, traditional seismic reservoir prediction methods based on linear hypothesis have been unable to meet the requirements of fine reservoir description (Yin et al., 2012). Machine learning, as a research hotspot in the big data mining field, is an effective tool to building prediction models from data (Peng et al., 2009; Gao et al., 2015; Zeng et al., 2018). Along with the rapid development of machine learning in other fields, many achievements have been made in machine learning seismic reservoir prediction.

Liu et al. (2015) used the neural network and multi-scale seismic to achieve the goal of improving inversion resolution. Zhang et al. (2014) applied the convolution neural network to seismic reservoir prediction and got good results. Liu et al. (2020) proposed a lithofacies classification method based on local deep multi-kernel learning support vector machine. In addition, many other machine learning algorithms are also applied in geophysics successfully, such as long short-term memory networks (Cai et al., 2019), generative adversarial networks (Kaur et al., 2020), and multilayer long short-term memory (Chen et al., 2020). To sum up, most of the machine learning reservoir prediction methods need to establish mapping relationship between seismic data and borehole data. However, in many cases, the training samples do not meet the requirements of sample number and data completeness (Lin et al., 2018), which is also called small sample problem in the field of data science. Direct use of these small sample data for modeling will lead to poor generalization ability and therefore will affect the prediction accuracy.

The small sample problem has always been a hot topic in academia and industrial communities. And now it is still a big challenge (Lu et al., 2018; Jayadeva et al., 2018). In order to overcome this problem, researchers have developed various methods to improve the prediction accuracy. These methods can be divided into two categories. One is based on grey system theory (Wu et al., 2012; Chang, 2015). The other method is virtual sample generation (VSG), which fills the information gap among the original samples

* Corresponding author.
*E-mail address:* zhangfch@upc.edu.cn (F.-C. Zhang).

https://doi.org/10.1016/j.petsci.2021.09.034

by adding virtual samples (Zheng et al., 2017; Tang et al., 2018).

The concept of virtual sample was proposed initially by Niyogi et al. (1998), who used the prior information from small sample sets to generate virtual samples. And they also proved mathematically that the process of VSG is equivalent to incorporating prior knowledge into model. Li et al. (2007) used adaptive network-based fuzzy inference system (ANFIS) and mega-trend-diffusion (MTD) to obtain the scheduling knowledge of the management system. To avoid the disadvantage of single distribution hypothesis, Zhu et al. (2016) proposed multi-distribution MTD method to generate virtual samples. The results showed that the prediction accuracy can be improved significantly after adding virtual samples.

In this paper, we study the machine learning seismic reservoir prediction method based on virtual samples. To solve the small sample problem, a supervised virtual sample generation method is proposed. First, we use the deep extreme learning machine, data trend estimation and hypersphere characteristic equation to generate the initial virtual samples. Then, the mechanism of virtual sample screening and model iterative optimization is used to eliminate noise samples and ensure the rationality of virtual samples. This method is performed on standard function data and real seismic data to demonstrate its feasibility. The results show that the proposed method is effective on the numerical model and real seismic data.

## 2. Theory

### 2.1. Deep extreme learning machine

#### 2.1.1. Extreme learning machine

Extreme learning machine (ELM), proposed by Huang et al. (2006a), is a kind of single hidden layer feedforward neural network for classification and regression. Different from usual forward feed neural network based on gradient descent, ELM adopts a fast and new learning mechanism. It randomly assigns input weights and thresholds, and uses the least square method to calculate the output weights. The structure of ELM is shown in Fig. 1.

The output of the ELM model can be represented as follows:

$$Y = H\beta = \begin{bmatrix} f(W_1X_1 + b_1) & \cdots & f(W_{Nh}X_1 + b_{Nh}) \\ \vdots & f(W_jX_i + b_j) & \vdots \\ f(W_1X_N + b_1) & \cdots & f(W_{Nh}X_N + b_{Nh}) \end{bmatrix} \times \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{Nh} \end{bmatrix} \tag{1}$$

where $Y = \{Y_1, Y_2, \cdots, Y_N\}$ represents the output matrix, $H$ is an output matrix of the hidden layer, $N$ is the number of samples, $\beta = \{\beta_1, \beta_2, \cdots, \beta_{Nh}\}^T$ is the weight matrix between the hidden neurons and the output neurons, $f(\cdot)$ is the activation function, $W_j = \{w_{j1}, w_{j2}, \cdots, w_{jp}\}$ denotes weight vector between the hidden neuron and the input neurons, $X_i = \{x_i^1, x_i^2, \cdots, x_i^p\}^T$ respectively represent the input vector, $Y_i = \{y_i^1, y_i^2, \cdots, y_i^q\}^T$ represent the output vector, and $b_j$ is the threshold of hidden neuron.

According to Huang et al. (2006b), the single hidden layer feedforward neural network can assign the input weights randomly, and the hidden layer output matrix $H$ can remain unchanged in the learning process. Thus, the optimization of the single hidden layer neural network will transform to the least square solution of the linear system $H\beta = Y$:

$$\left\|(W_1, \cdots, W_{Nh}, b_1, \cdots, b_{Nh})\widehat{\beta} - Y\right\|^2 = \min_{\beta}\|(W_1, \cdots, W_{Nh}, b_1, \cdots, b_{Nh})\beta - Y\|^2 \tag{2}$$

where $\widehat{\beta}$ is the calculated weight matrix between the hidden neurons and the output neurons.

#### 2.1.2. Deep extreme learning machine

The classical extreme learning machine has only one hidden layer, so it is difficult to extract the deep characteristic information from sample set, even if there are a large of neurons. Therefore, in

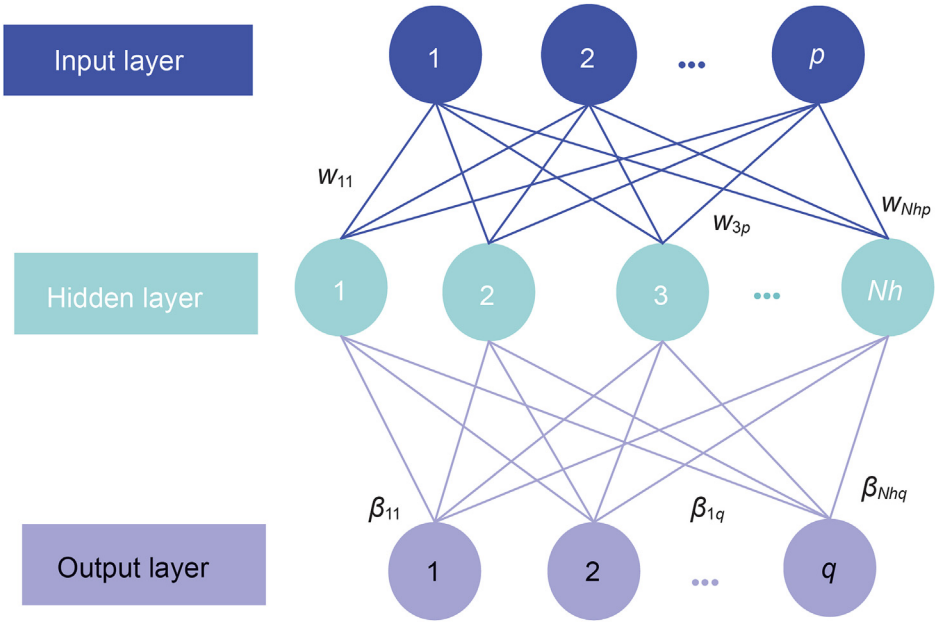


**Fig. 1.** The structure of ELM. $w_{Nhp}$ represents the weight between the hidden neuron $Nh$ and the input neuron $p$. $\beta_{Nhq}$ represents the weight between the hidden neuron $Nh$ and the output neuron $q$. $p$,$Nh$,$q$ respectively represent the number of neurons in the input layer, hidden layer, output layer.

this paper, we combine ELM with a deep neural network to establish the deep extreme learning machine (DELM). The essence of DELM is still a multi-layer feedforward neural network, and the structure of DELM is shown in Fig. 2. The part in the red box is the main structure of DELM. It still consists of an input layer, multiple fully connected hidden layers, and an output layer. Those ELM models in black boxes are used for DELM training.

In this paper, we choose the ELM auto-encoder (ELM-AE) learning algorithm (Cheng et al., 2017; Uzair et al., 2018; Yang et al., 2018) to obtain the weights $\{V_1, V_2, \cdots, V_{h+1}\}$, where $h$ is the number of hidden layers. As shown in Fig. 2, use the ELM-1 to calculate the output weight matrix $\widehat{\beta}^1$ by Eq. (2). In ELM-1, the output vector $\widehat{X}_i$ is equal to the DELM input vector $X_i$. Take the transpose of the output weight matrix $\widehat{\beta}^1$ as the DELM weight matrix $V_1$, and calculate the first hidden layer output vector $\gamma^1$. Then use $\gamma^1$ as the input vector and output vector of ELM-2 to get $\widehat{\beta}^2$. In the same way, take the transpose of $\widehat{\beta}^2$ as the weight matrix $V_2$ of DELM. And so on, until the last hidden layer, calculate $\widehat{\beta}^{h+1}$ with $Y_i$ as the objective output by $\widehat{\beta}^{h+1} = (\gamma^h)^{-1} Y_i$, where $(\gamma^h)^{-1}$ is the generalized inverse matrix of $\gamma^h$ and $Y_i$ is the objective output of DELM.

DELM can extract the feature information hidden in the sample data through supervised learning. Compared with the conventional back propagation neural network, it does not need the iterative optimization process, which can improve the learning efficiency greatly (Tang et al., 2016). At the same time, it solves the problem that the ELM with single hidden layer structure cannot get deep information, and it has a strong non-linear mapping ability.

### 2.2. Data trend estimation

To fill the information gap among small samples, it is important to have a correct estimation to the sample distribution (Li et al., 2006). In this paper, we use the data trend estimation method to calculate the extension domain of each attribute.

For sample $X_i = \{x_i^1, x_i^2, \cdots, x_i^p\}^T$, $x_i^p$ is one data attribute of this sample. So, $\{x_1^p, x_2^p, \cdots, x_N^p\}$ can be one data attribute for this sample

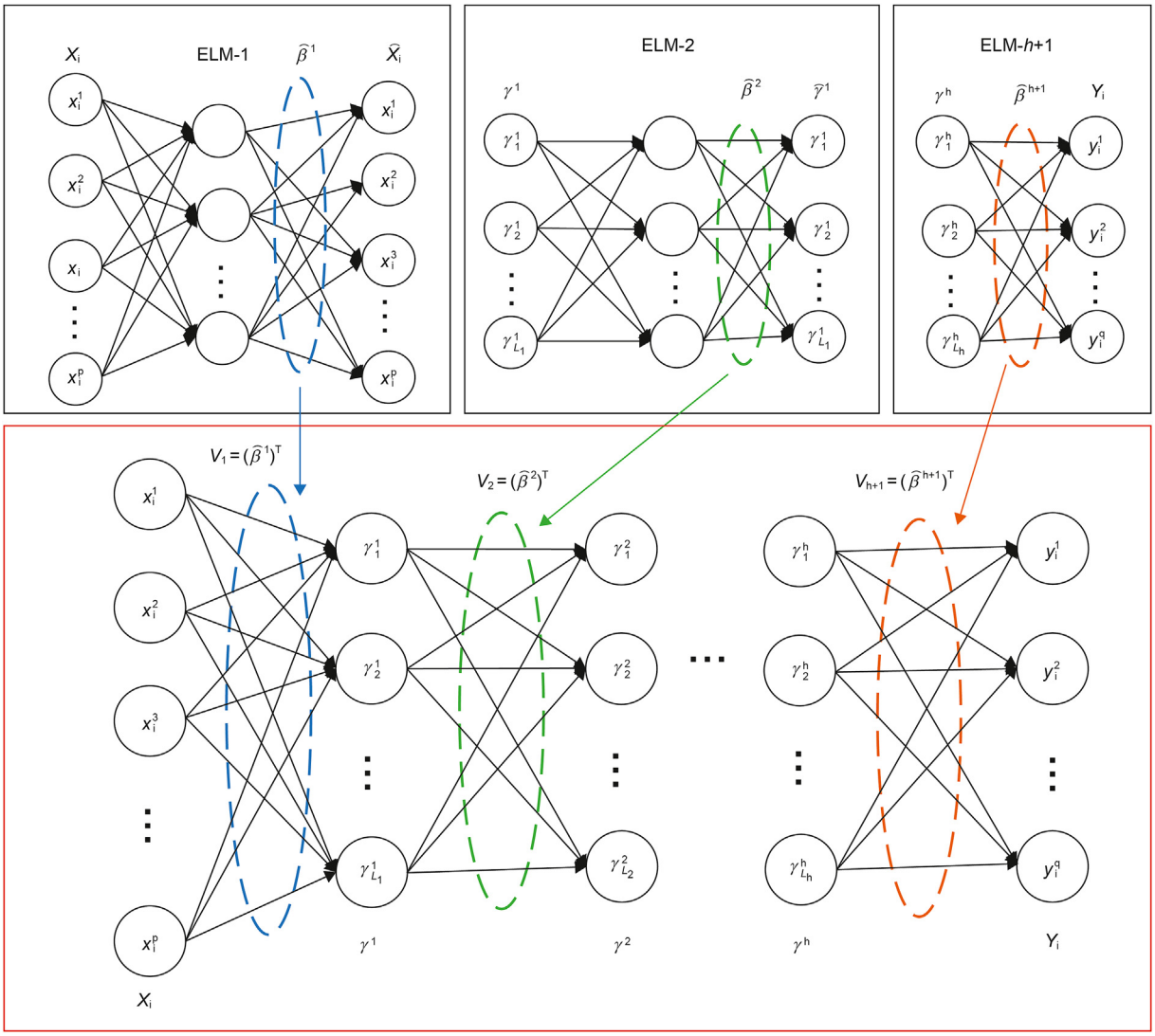


**Fig. 2.** The structure and training mechanism of DELM. The part in the red box is the structure of DELM. The rest are ELM models for DELM training. We choose the ELM auto-encoder learning algorithm to obtain the weights $\{V_1, V_2, \cdots, V_{h+1}\}$. The dotted lines and arrow marks indicate the relationship of weight matrixes between DELM and ELM.

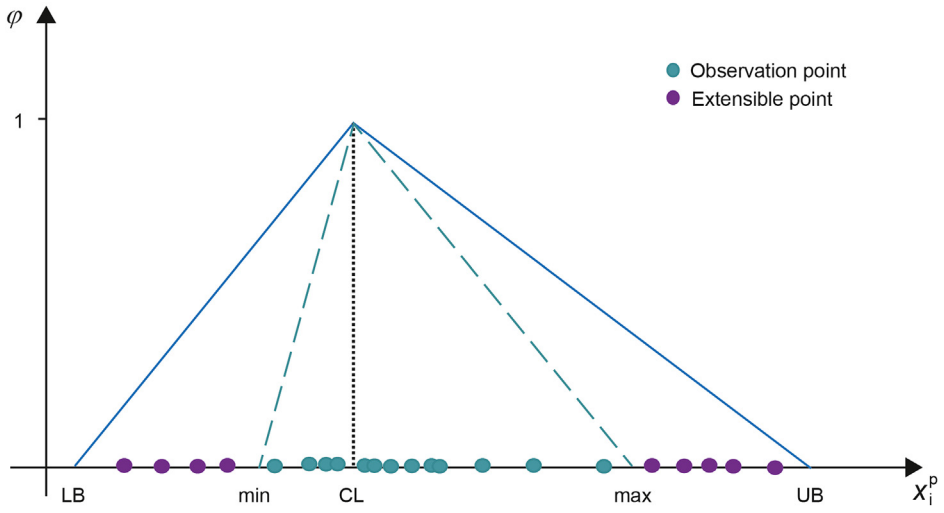


**Fig. 3.** Determine acceptable extended domain of an attribute. *CL* is the data attribute center point. *LB* is acceptable lower bound and *UB* is acceptable upper bound of the attribute. min and max are the minimum and maximum values for current attribute value. $\phi$ is membership function.

set. The center point of this data attribute can be calculated via:

$$CL = \frac{1}{N}\sum_{i=1}^{N} x_i^p \tag{3}$$

We use triangle membership function to calculate the occurrence possibility. The mathematical expressions of attribute left and right diffusion skewness are:

$$Sk_L = \frac{N_L}{N_L + N_U + \sigma} \tag{4}$$

and

$$Sk_U = \frac{N_U}{N_L + N_U + \sigma} \tag{5}$$

where $N_L$ and $N_U$ respectively represent the samples number those are less or greater than the data attribute center point *CL*, $\sigma$ is shape correction factor, taken as 1 in this paper.

After obtaining the right and left diffusion skewness, we calculate the attribute acceptable extension domain by

$$LB = CL - Sk_L \times \sqrt{-2 \times \hat{s}_x^2 / N_L \times \ln(\phi(LB))} \tag{6}$$

and

$$UB = CL + Sk_U \times \sqrt{-2 \times \hat{s}_x^2 / N_U \times \ln(\phi(UB))} \tag{7}$$

where *LB* is acceptable lower bound and *UB* is acceptable upper bound of the attribute. As shown in Fig. 3, $\phi(LB)$ and $\phi(UB)$ are the membership values of *LB* and *UB*. $\hat{S}_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i^p - \bar{x}^p)^2$ is the attribute variance. $\bar{x}^p$ is the attribute mean.

As shown in Fig. 3, the attribute acceptable extended domain can be calculated by data trend estimation method, $x_i^p \in [LB \ \min] \cup [\max \ UB]$. It can be used as constraint to virtual samples, which makes the virtual samples more accurate.

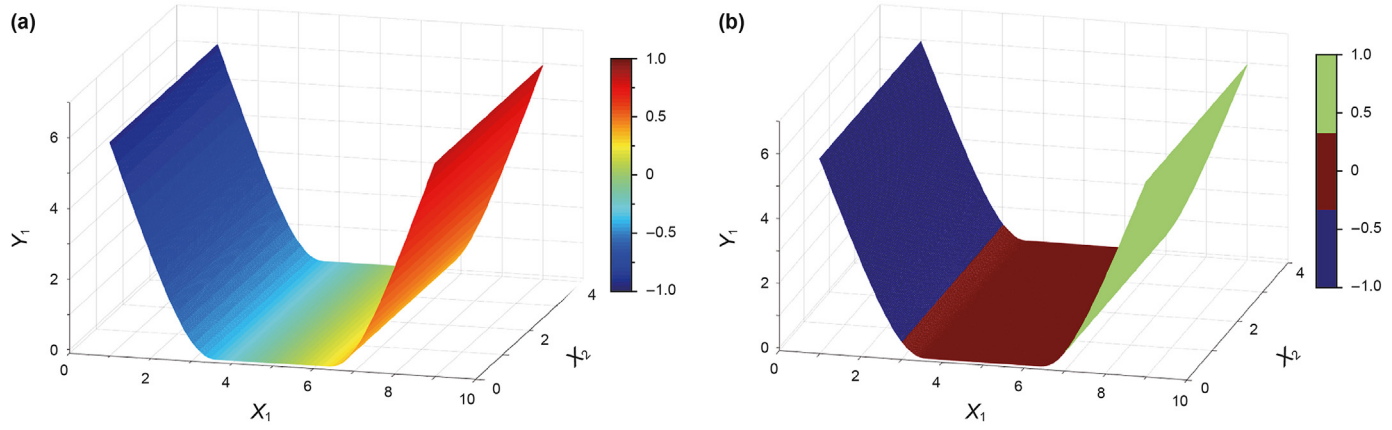


**Fig. 4.** Nonlinear manifold subspace partition representation. (a) a nonlinear surface; (b) three locally linear surfaces divided by using the nonlinear surface in (a). In the locally linear space, problem can be analyzed and solved using the linear methods.

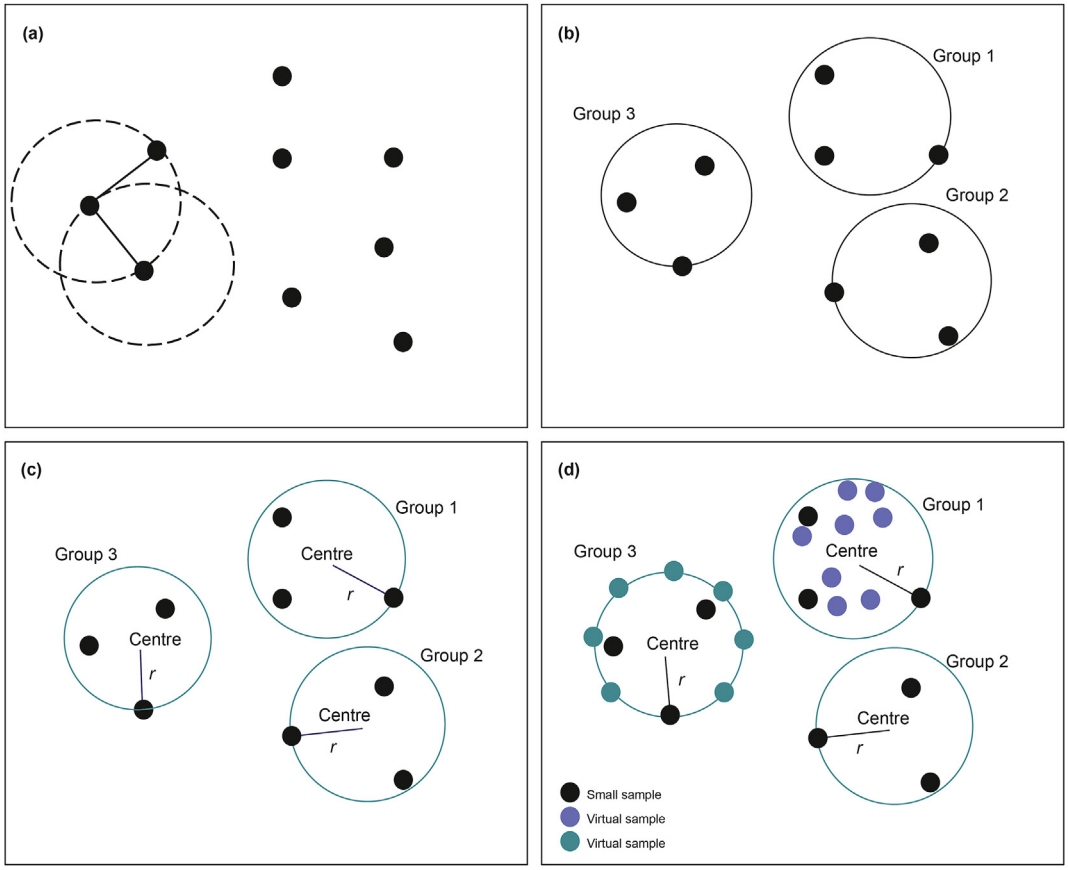


**Fig. 5.** The process of virtual sample generation. (a) Small sample grouping process; (b) Small sample grouping result; (c) Building hypersphere; (d) Generating virtual samples.

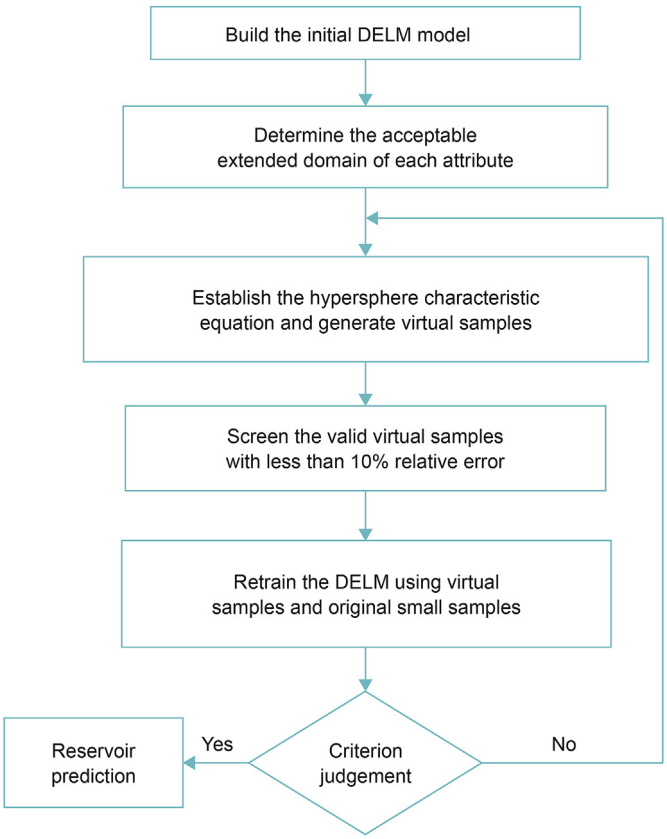


**Fig. 6.** Flowchart of small sample modeling.

### 2.3. Hypersphere characteristic equation

#### 2.3.1. Manifold subspace division theory

In computing science, the function space, which varies with time or other factors, is defined as nonlinear manifold (Lin and Zha, 2008; Jamshidi et al., 2011). So, manifold is a space with Euclidean properties locally (Tenenbaum et al., 2000). As shown in Fig. 4, a nonlinear manifold can be divided into several linear spaces, in which we can analyze and solve problems using linear methods. Generally, the more manifold subspaces are, the higher linearity of subspaces have (Wang et al., 2008). So, in this paper, we divide the nonlinear manifold that contains the small sample set into several local linear subspaces.

#### 2.3.2. Establishment of hypersphere characteristic equation

According to function theory, any sample is contained in one base of topological space. And a finite number of samples are included in a linear superposition space, which is composed by multiple bases (Suhubi, 2003). That means, in small sample problem, the small sample space is a subspace of the real model space. And it can be considered as a linear combination of spheres in the entire space. These spheres are hyperspheres that can describe the distribution of samples. The hypersphere has only two descriptive parameters: hypersphere center and hypersphere radius, which uses the following mathematical expression:

$$S^p = \{X_i \in \mathbb{R}^p : \ ||X_i - X_0|| \leq r\} \tag{8}$$

where $X_0$ and $r$ represent hypersphere center and hypersphere radius respectively, $X_i$ is the sample point in hypersphere.

The hypersphere characteristic equation can be expressed as

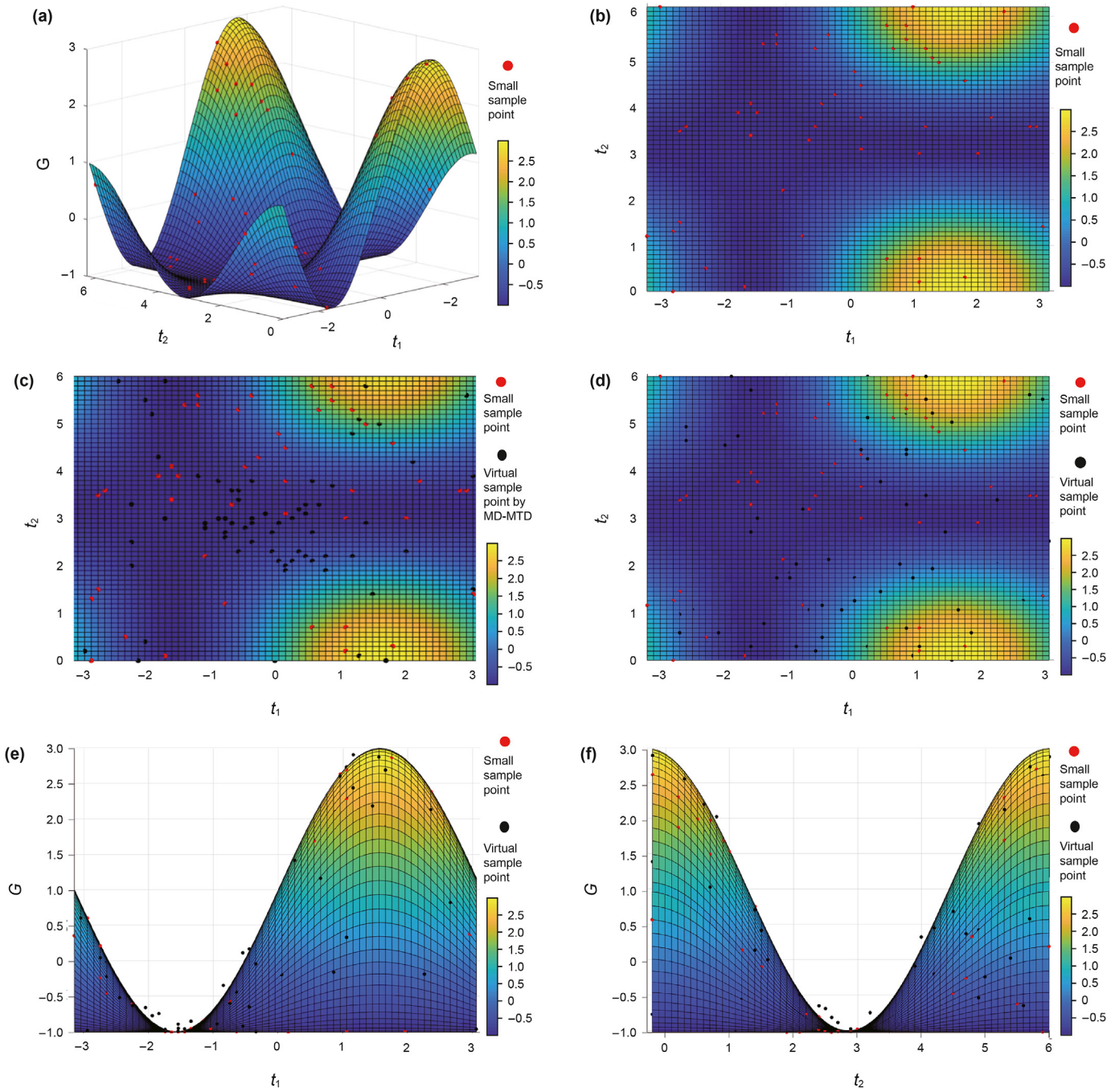


**Fig. 7.** Standard function surface (a) and sample projection information on the plane of $t_1 - t_2$ (b, c, d), $G - t_1$(e), $G - t_2$(f).

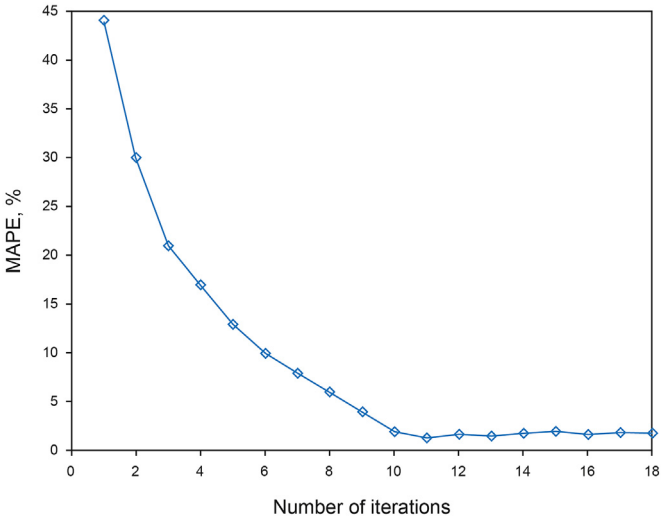


**Fig. 8.** The prediction error of standard function example.

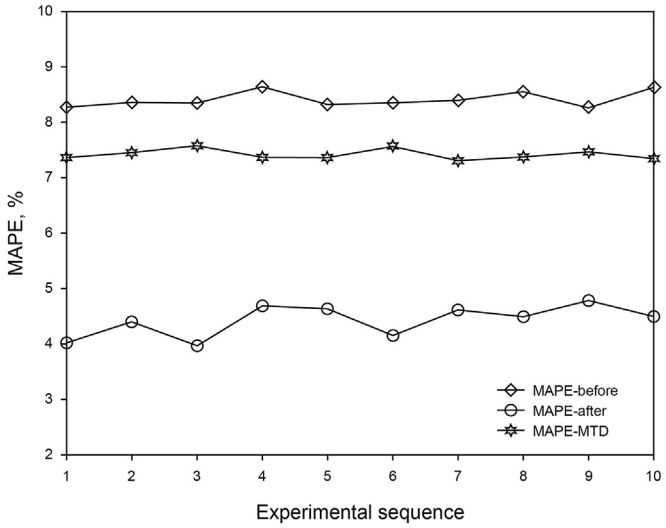


**Fig. 9.** The comparison of prediction error. MAPE-before is the mean absolute percent error using small samples. MAPE-after is the mean absolute percent error using small samples and virtual samples. MAPE-MTD is the mean absolute percent error using the MD-MTD method according Zhu et al. (2016).

**Table 1**
The comparison of prediction effect.

| Experimental sequence | $MAPE_{before}$, % | $MAPE_{after}$, % | $MAPE_{MTD}$, % | *EIR*, % |
|---|---|---|---|---|
| 1 | 8.2659 | 4.0192 | 7.3608 | 51.38 |
| 2 | 8.3505 | 4.3978 | 7.4441 | 47.34 |
| 3 | 8.3445 | 3.9664 | 7.5706 | 52.47 |
| 4 | 8.6348 | 4.6892 | 7.3645 | 45.69 |
| 5 | 8.3110 | 4.6361 | 7.3592 | 44.22 |
| 6 | 8.3429 | 4.1506 | 7.5625 | 50.25 |
| 7 | 8.3920 | 4.6160 | 7.3048 | 45.00 |
| 8 | 8.5415 | 4.4933 | 7.3740 | 47.39 |
| 9 | 8.2593 | 4.7867 | 7.4617 | 42.04 |
| 10 | 8.6232 | 4.4946 | 7.3452 | 47.88 |

$$\begin{cases} \tilde{x}_i^1 = r\sin\theta_1\sin\theta_2\cdots\sin\theta_{p-1} \\ \tilde{x}_i^2 = r\sin\theta_1\sin\theta_2\cdots\cos\theta_{p-1} \\ \tilde{x}_i^3 = r\sin\theta_1\sin\theta_2\cdots\cos\theta_{p-2} \\ \tilde{x}_i^4 = r\sin\theta_1\sin\theta_2\cdots\cos\theta_{p-3} \\ \vdots \\ \tilde{x}_i^{p-1} = r\sin\theta_1\cos\theta_2 \\ \tilde{x}_i^p = r\cos\theta_1 \end{cases} \tag{9}$$

where $\theta_1, \theta_2, \theta_3, \cdots, \theta_{p-1} \in [0, 2\pi]$ are hypersphere angles, and $\tilde{X}_i = \{\tilde{x}_i^1, \tilde{x}_i^2, \tilde{x}_i^3, \cdots, \tilde{x}_i^p\}$ represents one virtual sample, while $\tilde{x}_i^1, \tilde{x}_i^2, \tilde{x}_i^3, \cdots, \tilde{x}_i^p$ are the values of virtual sample $\tilde{X}_i$ in different dimensions.

## 3. Methodology

### 3.1. Virtual sample generation method

The detailed steps of the virtual sample generation method based on hypersphere characteristic equation are as follows:

(1) Small sample data grouping: as shown in Fig. 5a and Fig. 5b, group small sample data set according to manifold subspace partition method. Detailed steps were performed as previously described (Jia, 2009).
(2) Establish the hypersphere characteristic equations using grouped data: firstly, calculate the centres of these groups using the following equation:

$$g_j^k = \frac{1}{n}\sum_{i=1}^{n} x_{ij}^k \quad j = 1, 2, \cdots, m \ ; \ k = 1, 2, \cdots, p \tag{10}$$

where $n$ and $m$ are respectively dimension of the number of samples in the group and the number of sample groups, $g_j^k$ is centre value of $k-$th dimension in $j-$th hypersphere, $x_{ij}^k$ is $k-$th dimension value of $i-$th sample in $j-$th hypersphere.

As shown in Fig. 5c, $r_j = \{g_j^1, g_j^2, \cdots, g_j^p\}$ is the centre of the $j-$th hypersphere, and set the maximum distance between centre and all samples in one group as radius. Then use Equation (9) to establish the hypersphere characteristic equations.

(3) Generate virtual samples: as shown in Fig. 5d, based on hypersphere parameter equation, there are two methods for generating virtual samples. Method 1 generates virtual samples randomly inside the hyperspheres; Method 2 generates virtual samples uniformly on the edges of hyperspheres.

### 3.2. Training and prediction method based on virtual data and machine learning

The Fig. 6 displays the workflow of machine learning seismic reservoir prediction based on VSG. The detail steps are as follows:

(1) Building the initial DELM model: set up DELM model, and train DELM model using small sample data.
(2) Virtual sample generation: calculate the acceptable extended domain of each attribute by DTE method, and use it as

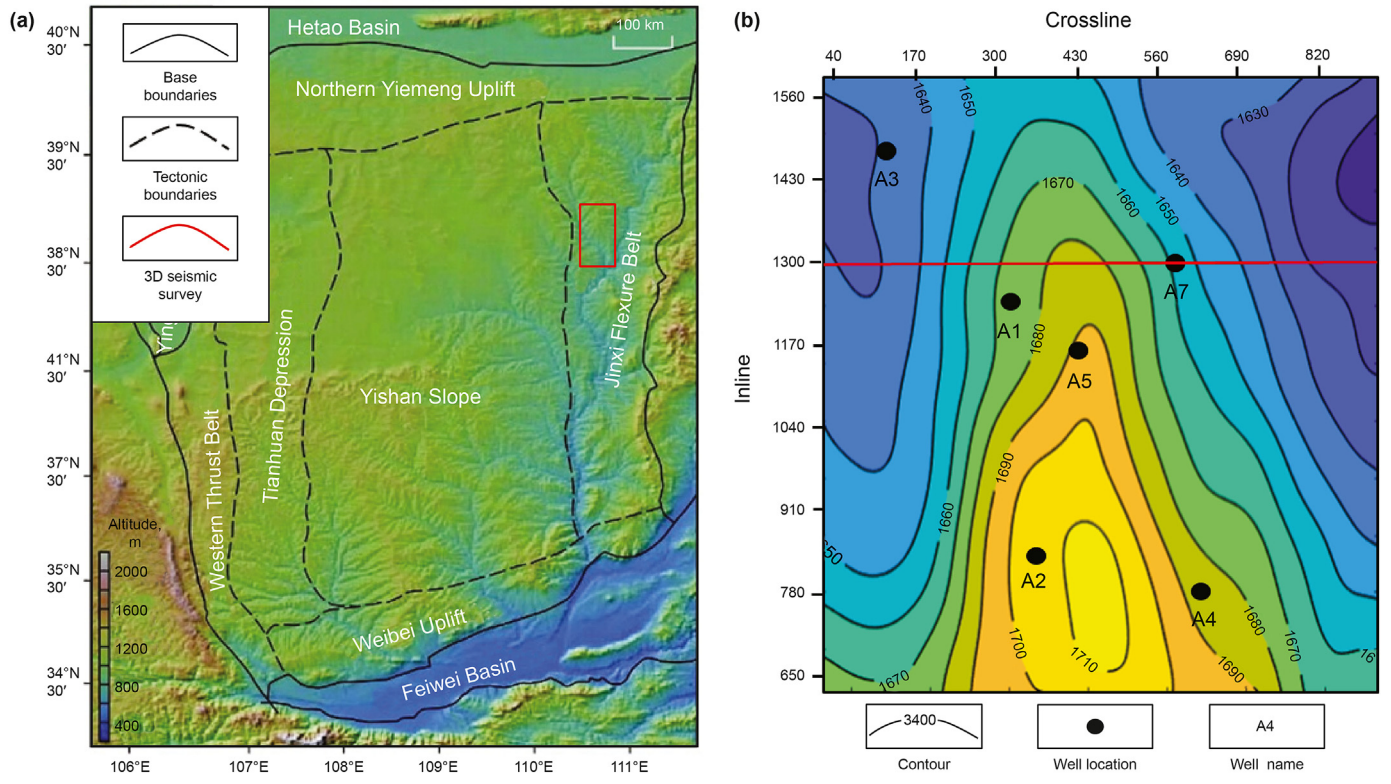


**Fig. 10.** (a) Regional map of the Ordos Basin (modified from Wu et al., 2019); (b) The structure map of the target layer.

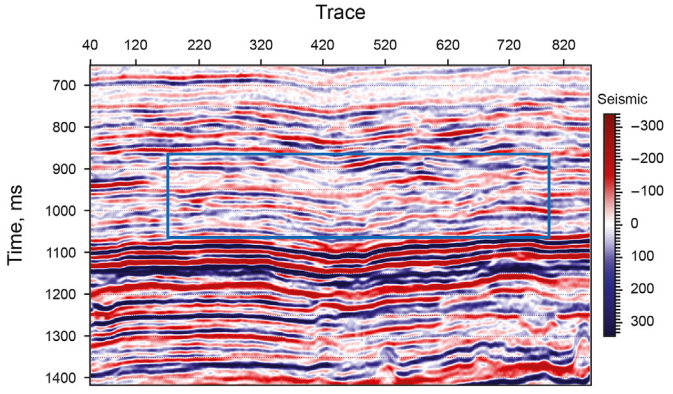


**Fig. 11.** Seismic data section. Location of the seismic data is shown in Fig. 10b (red line). The blue box indicates the range of Fig. 13a.

constraint to establish hypersphere characteristic equation. Then generate virtual samples.

(3) Virtual sample screening: according to Li and Wen (2014), Li and Lin (2006), when relative error is less than 10%, the virtual samples are acceptable. Therefore, we determine 10% as the virtual samples screening standard. Virtual samples, whose relative error are less than 10%, are considered as valid virtual samples, and the relative error equation is as follows:

$$\frac{\|Y_i - Y_i'\|}{\|Y_i'\|} < 10\% \tag{11}$$

where $Y_i$ represents output variable. $Y_i'$ represents objective output variable. But in application, we cannot get the real $Y_i'$. So, we use the virtual input data and the initial DELM to get an approximation objective output data. $\|\cdot\|$ represents the norm of vector.

(4) Update the DELM: retrain the DELM using virtual samples and original small samples.

(5) Prediction error judgment using test set: calculate prediction error using test samples. If prediction error meets the termination criterion, the training process completes; otherwise, return to step 2) to continue generating virtual samples.

(6) Reservoir prediction: use the DELM and post-stack seismic data for reservoir prediction.

## 4. Applications

### 4.1. Standard function example

In this section, we use the standard function to verify the validity of the proposed method. As shown in Eq. (12), the function has two inputs and one output. In this paper, set the input value: $t_1 \in [-\pi : 0.1 : \pi]$, $t_2 \in [0 : 0.1 : 2*\pi]$. The function is shown in Fig. 7a, indicating that the function is highly nonlinear.

$$G(t_1, t_2) = \sin(t_1) + \cos(t_2) + \sin(t_1) \times \cos(t_2) \tag{12}$$

In this section, 50 samples are chosen as small sample set, and we also use 20 samples as test set. Fig. 7a and b shows the distribution of small samples in space and plane. From the distribution on the plane, it is not uniform, and there are no effective samples in many areas.

Fig. 7d shows the projection of virtual samples on plane $t_1 - t_2$. From the distribution, the virtual samples fill among small samples, so that makes the samples distribution more uniform on the plane, which will improve the generalization ability of prediction model. It also proves that the hyperspheres are good enough to capture the

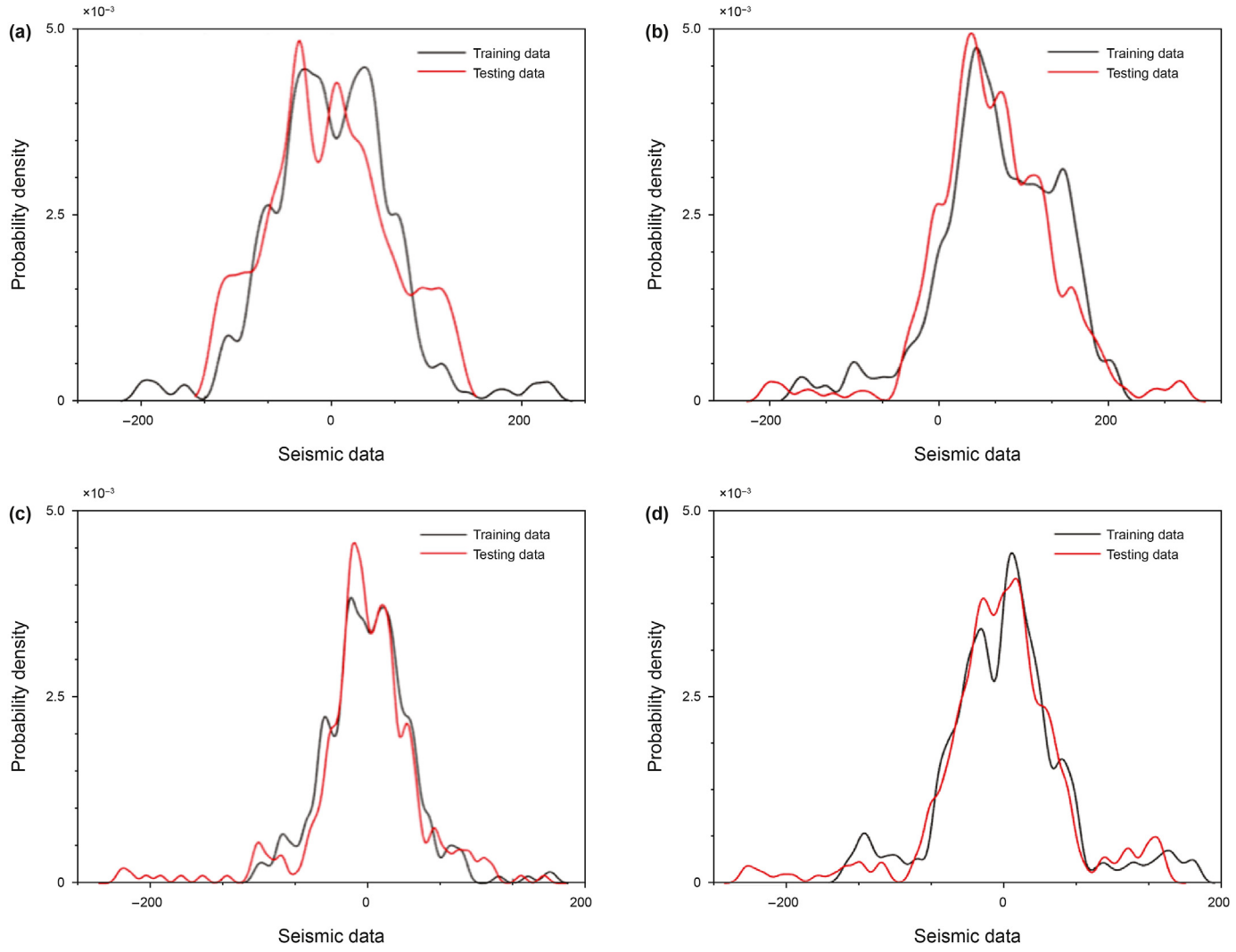


**Fig. 12.** The probability density of training and testing data. From the 30 elements of the input variable, we select the 5th (a), 15th (b), 20th (c), 25th (d) elements and calculate the probability density of the input variable. The red line is the probability density distribution of the test data, and the black line is the probability density distribution of the training data.

multi-modal distributions of the input data. And Fig. 7e and f displays the projections on the plane $t_1 - G$ and $t_2 - G$. The virtual samples can basically fall on the surface or fit well with the surface, which indicates the correctness of the virtual samples. Fig. 8 shows the prediction error curve along iterations. After a certain number of iterations, the error curve can converge quickly.

In order to further verify the improvement of virtual samples on prediction model, we do 10 independent experiments and compare the prediction error. We employee two metrics, i.e. mean absolute percent error (MAPE) and error improvement rate (EIR),

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left\|\frac{Y_i' - Y'}{Y'}\right\| \times 100\%$$
$$EIR = \frac{MAPE_{\text{before}} - MAPE_{\text{after}}}{MAPE_{\text{before}}} \times 100\% \quad (13)$$

where $n$ denotes the number of test samples, $MAPE_{\text{before}}$ and $MAPE_{\text{after}}$ are the mean absolute percent error before and after the virtual samples generated.

Table 1 and Fig. 9 respectively show the prediction results using different methods and different samples. They show that the prediction error decreases significantly after adding virtual samples. The prediction error is about 8.5% before adding virtual samples. After adding the virtual samples, the error reduces to 4% approximately, and the error correction rate is about 50%.

We also compare the method proposed in this paper with MD-MTD method (Zhu et al., 2016). Fig. 7c shows the projection of virtual samples generated by MD-MTD. The virtual samples are gathered in center area of the plane. For many areas, the MD-MTD method does not generate virtual samples. As shown in Fig. 9, the MAPE of MD-MTD method increases but the effect is limited.

### 4.2. Real data example

In order to study the feasibility of the proposed method in real seismic reservoir prediction, we use a practical work area for test.

(1) Geologic background: The Ordos Basin is a large, intracratonic sedimentary basin in central China, which comprises a total area of $23 \times 10^4$ km$^2$ (Fig. 10a, Wu et al., 2019). The study area is located on northeastern of Ordos Basin, in which the red box is the 3D seismic survey area. There is a

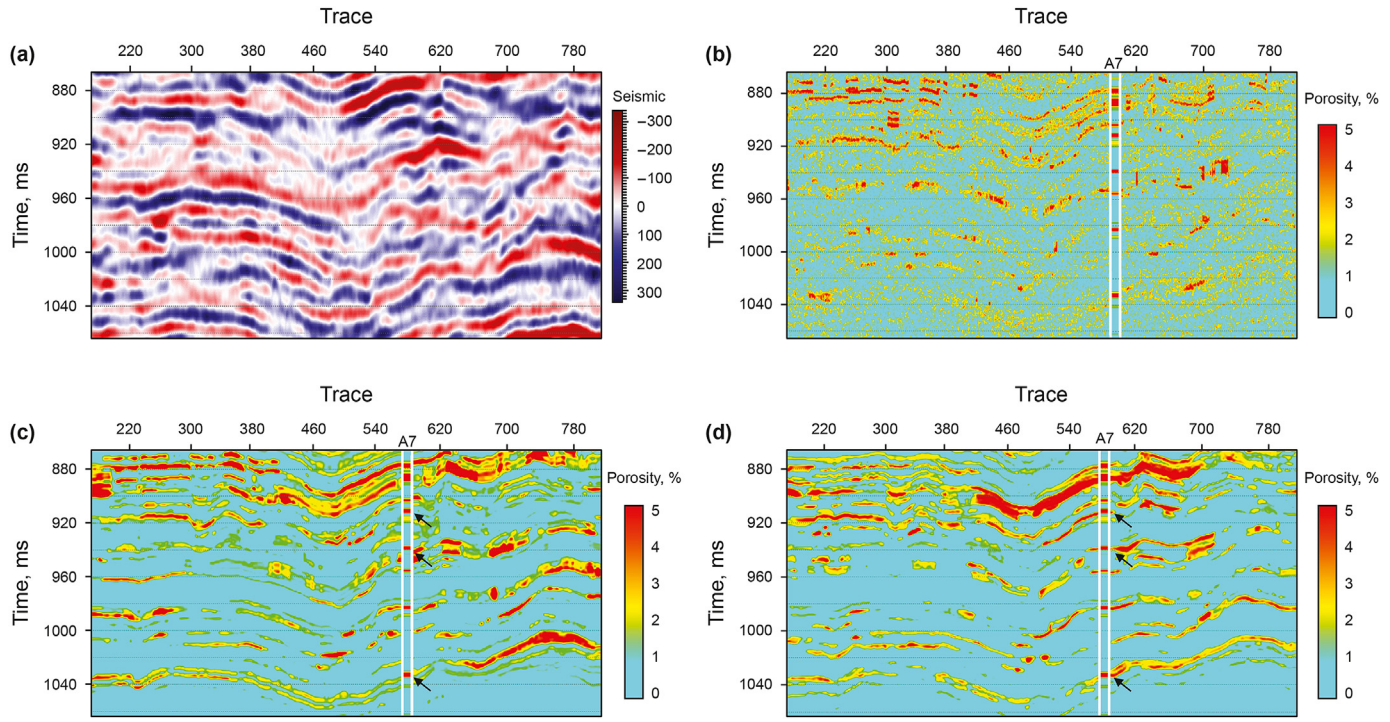


**Fig. 13.** The effect comparison of real data application. (a) Seismic section; (b) Predicted porosity section by small samples; (c) Predicted porosity section by small samples and unscreened virtual samples; (d) Predicted porosity section by small samples and screened virtual samples.

large tight sandstone gas reservoir in the Upper Paleozoic succession, which is characterized by strong horizontal heterogeneity, thin interbeds in vertical, low porosity, low permeability, low saturation and generally less than 10 m in thickness. The gas reservoir is not controlled by geological structure and is dominated by large stratigraphic traps. The seismic profile of target layer is shown in Fig. 11. Due to these special geological conditions, it is difficult to identify the horizontal and vertical distribution of the reservoir using pre-stack seismic inversion and other conventional post-stack seismic inversion. And because this area is in the initial development stage, the well data is limited. Accurate reservoir prediction in the case of few wells is of great significance for further exploration and development. The well distribution and geological structure of the Upper Paleozoic are shown in Fig. 10b.

(2) Seismic reservoir prediction: there are 6 wells in the study area, and their distribution is shown in Fig. 10b. We select training samples from well A1, A2, A3, A4 and A5. Well A7 is used as test well, without participating in the training process. There are 1000 training samples and 200 test samples. The seismic data beside the wells composes the input variable, while the logging porosity data composes the output variable. The output variable is one-dimensional single element vector, and the input variable is one-dimensional seismic data with 30 elements. In this paper, the sampling interval of seismic data is 1 ms. Setting the sampling position as the midpoint, taking 14 ms up and 15 ms down, a total of 29 ms seismic records (30 samples) are taken as input variable.

To verify that the training and test data follow the same distribution, we calculate the data probability density function. Since the input data are variables with 30 elements, the probability density function is variable with 30 dimensions, which cannot be displayed directly. Therefore, we select the 5th, 15th, 20th, and 25th elements of the input variables to estimate the probability density. And the results are shown in Fig. 12. The red line is the probability density distribution of the test data, and the black line is the probability density distribution of the training data. Based on the probability density results, the training and test data follow the same distribution.

After parameter test, we use the DELM model with structure: 30-20-20-20-20-20-20-1. It means that the DELM has 6 hidden layers, each of which contains 20 neurons. And the input and output layers have 30 and 1 neurons, respectively. After sample set and DELM being established, we predict porosity using the method in Section 3.2.

Fig. 13 shows reservoir prediction results using virtual samples and machine learning. Fig. 13a represents a seismic section, and Fig. 13b, 13c and 13d show the predicted porosity sections. Compared the predicted porosity with seismic, the porosity prediction results have higher vertical resolution and they are consistent well with the seismic morphology in transverse. Fig. 13b displays predicted porosity section by small sample set, and this result has very high randomness. In contrast, as shown in Fig. 13c and d, machine learning with mixed samples can reduce randomness and improve prediction accuracy. Fig. 13c displays predicted porosity result with unscreened virtual samples, while Fig. 13d is predicted porosity result with screened virtual samples. Comparing these two results, there are no difference in resolution, but the latter one is more consistent with logging results. At the positions marked by arrows in Fig. 13c and d, the result using unscreened samples shows a large deviation from logging data, while the result using screened samples is consistent with logging data highly.

In order to make a more intuitive evaluation to the prediction results, Fig. 14 shows the cross plot of predicted porosity with real porosity. Compared with the prediction result by small sample set,

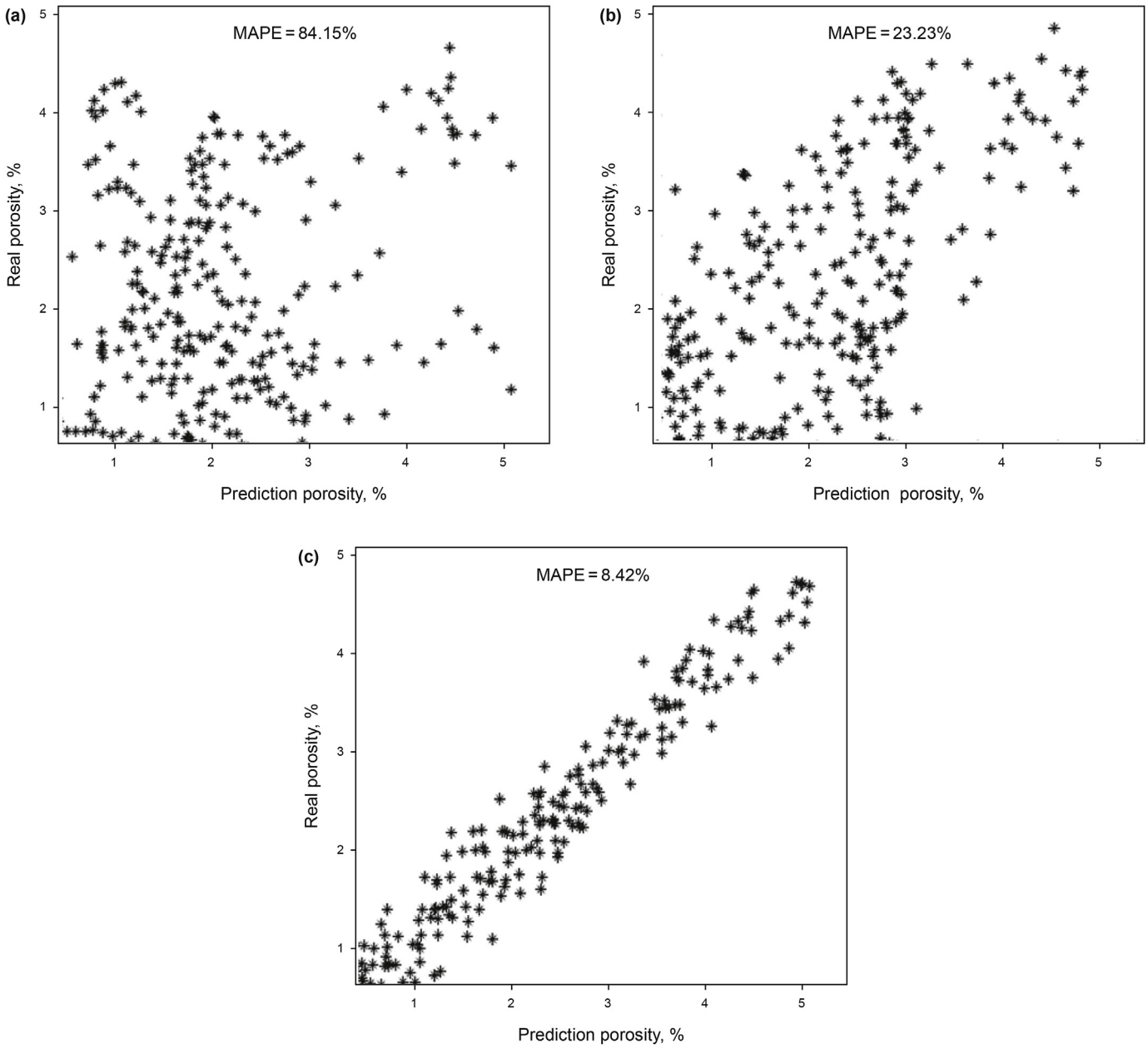


**Fig. 14.** Cross plot of the predicted data with the real data (a) Predicted porosity by small samples; (b) Predicted porosity by small samples and unscreened virtual samples; (c) Predicted porosity by small samples and screened virtual samples.

machine learning with mixed samples can reduce the MAPE to 23.23%. Meanwhile, the method with sample screening can further improve the prediction accuracy to more than 90%.

## 5. Discussion

The machine learning method based on virtual samples can solve the small sample problem to a certain degree. Through the preliminary test, it is found that this method also has certain requirement for the completeness of small samples. Taking reservoir porosity prediction as an example, the proportion of high porosity samples and low porosity samples should be relatively balanced. It means that it shouldn't only have high porosity samples or low porosity samples. When a class of samples is missing, the hypersphere of this class cannot be created, and virtual samples of this class cannot be generated.

About the number of hyperspheres, it is determined by the manifold subspaces. According to the manifold subspace partition theory (Jia, 2009), the number of subspaces is not set in advance. Instead, by setting threshold parameter, the larger the threshold parameter is, the higher nonlinearity of the subspaces have, and the fewer subspaces will be constructed. Conversely, the smaller the threshold parameter is, the more subspaces will be constructed. Based on reference and practical application, the threshold value is related to sample dimension. The higher the sample dimension is, the smaller the threshold value should be to ensure the linearity in subspaces.

This paper is an exploratory work in small sample machine learning problem. We have established the method for small sample reservoir prediction. But like many other machine learning

prediction methods (Lei et al., 2014; Franco, 2006), there is still a problem: how to improve generalization ability. At present, in machine learning reservoir prediction, it can only build sample sets and train model in one geological area. Then predict the reservoir parameter for this area. There is no established method with strong generalization ability, by which reservoir prediction can be made for any geological areas through once learning. This requires further study in future.

## 6. Conclusions

In this paper, machine learning and virtual samples are used to achieve the reservoir prediction in cases of small samples. First, generate the initial virtual samples by using DELM, data trend estimation and hypersphere characteristic equation. Then the mechanism of virtual samples screening and model iterative optimization are used to eliminate noise samples. Finally, the virtual samples are combined with the original small samples to form the training set for machine learning.

Through the standard function example and real data application, the following implications can be recognized:

1) The supervised virtual sample generation method can reduce the influence of noise samples and improve prediction accuracy significantly. Moreover, sample screening is applicable to different virtual sample generation methods. So, we believe that the combination of virtual sample generation and sample screening can achieve the optimal effect in small sample problem.
2) To a certain extent, virtual sample generation method can solve the small sample problem in machine learning. This method can be used to predict the reservoir in less well working area. Minimum requirement for well number can reduce from dozens to several. This can advance the application of machine learning in oil and gas exploration greatly.

## Acknowledgments

This work was supported by National Natural Science Foundation of China under Grants 41874146 and 42030103.

## References

Ahmed, O.A., Abdel-Aal, R.E., AlMustafa, H., 2010. Reservoir property prediction using abductive networks. Geophysics 75 (1), 1–9. https://doi.org/10.1190/1.3298443.

Cai, Y., Shyu, M.L., Tu, Y.X., et al., 2019. Anomaly detection of earthquake precursor data using long short-term memory networks. Appl. Geophys. 16 (2), 257–266. https://doi.org/10.1007/s11770-019-0774-1.

Chang, C.J., Li, D.C., Huang, Y.H., et al., 2015. A novel gray forecasting model based on the box plot for small manufacturing data sets. Appl. Math. Comput. 265, 400–408. https://doi.org/10.1016/j.amc.2015.05.006.

Chen, W., Yang, L.Q., Zha, B., et al., 2020. Deep learning reservoir porosity prediction based on multilayer long short-term memory network. Geophysics 85 (4), WA213–W225. https://doi.org/10.1190/geo2019-0261.1.

Cheng, X.Y., Liu, H.P., Xu, X.Y., et al., 2017. Denoising deep extreme learning machine for sparse representation. Memetic Computing 9, 199–212. https://doi.org/10.1007/s12293-016-0185-2.

Franco, L., 2006. Generalization ability of Boolean functions implemented in feedforward neural networks. Neurocomputing 70, 351–361. https://doi.org/10.1016/j.neucom.2006.01.025.

Gan, L.D., Zhang, X., Wang, Y.J., et al., 2018. Current status and development trends of seismic reservoir prediction viewed from the exploration industry. OGP 53 (1), 214–225. https://doi.org/10.13810/j.cnki.issn.1000-7210.2018.01.026 (in Chinese).

Gao, Y., Zhang, B.L., Mao, J.L., et al., 2015. Machine learning-based adaptive very-short-term forecast model for photovoltaic power. Power Syst. Technol. 39 (2), 307–311. https://doi.org/10.13335/j.1000-3673.pst.2015.02.002 (in Chinese).

Huang, G.B., Zhu, Q.Y., Siew, C.K., 2006a. Extreme learning machine: theory and applications. Neurocomputing 70, 489–501. https://doi.org/10.1016/j.neucom.2005.12.126.

Huang, G.B., Chen, L., Siew, C.K., 2006b. Universal approximation using incremental constructive feed forward networks with random hidden nodes. IEEE Trans. Neural Network. 17 (4), 879–892. https://doi.org/10.1109/TNN.2006.875977.

Jamshidi, A.A., Kirby, M.J., Broomhead, D.S., 2011. Geometric manifold learning. IEEE Signal Process. Mag. 28 (2), 69–76. https://doi.org/10.1109/MSP.2010.939550.

Jayadeva, Soman, S., Saxena, S., 2018. EigenSample: a non-iterative technique for adding samples to small datasets. Appl. Soft Comput. 70, 1064–1077. https://doi.org/10.1016/j.asoc.2017.08.017.

Jia, Q.W., 2009. Research on Classification Algorithm with Non-training Pattern Reject Option in High-Dimensional Space. Yanshan University (Qinhuangdao) (in Chinese).

Kaur, H., Pham, N., Fomel, S., et al., 2020. Improving the resolution of migrated images by approximating the inverse Hessian using deep learning. Geophysics 85 (4), 173–183. https://doi.org/10.1190/geo2019-0315.1.

Lei, Y.W., Ding, L.X., Ding, Y.M., 2014. Generalization ability of fractional polynomial models. Neural Network. 49, 59–73. https://doi.org/10.1016/j.neunet.2013.09.009.

Li, D.C., Lin, Y.S., 2006. Using virtual sample generation to build up management knowledge in the early manufacturing stages. Eur. J. Oper. Res. 175, 431–434. https://doi.org/10.1016/j.ejor.2005.05.005.

Li, D.C., Wen, I.H., 2014. A genetic algorithm-based virtual sample generation technique to improve small data set learning. Neurocomputing 143, 222–230. https://doi.org/10.1016/j.neucom.2014.06.004.

Li, D.C., Wu, C.S., Tsai, T.I., et al., 2006. Using mega-fuzzification and data trend estimation in small data set learning for early FMS scheduling knowledge. Comput. Oper. Res. 33 (6), 1857–1869. https://doi.org/10.1016/j.cor.2004.11.022.

Li, D.C., Wu, C.S., Tsai, T.I., et al., 2007. Using mega-trend-diffusion and artificial samples in small data set learning for early flexible manufacturing system scheduling knowledge. Comput. Oper. Res. 34 (4), 966–982. https://doi.org/10.1016/j.cor.2005.05.019.

Lin, N.T., Zhang, D., Zhang, K., et al., 2018. Predicting distribution of hydrocarbon reservoirs with seismic data based on learning of the small-sample convolution neural network. Chin. J. Geophys. 61 (10), 4110–4125. https://doi.org/10.6038/cjg2018J0775 (in Chinese).

Liu, H.J., Lei, X.H., Mao, C.L., et al., 2014. Improving reservoir thickness prediction using seismic attributes and attributes fusion. Acta Geophys. 62 (3), 544–563. https://doi.org/10.2478/s11600-013-0174-5.

Liu, H.Q., Zhang, F.C., Dai, R.H., 2015. Extrapolating of the cross-well seismic data based on the neural network and multi-scale seismic joint inversion. Comput. Tech. Geophys. Geochem. Explor. 37 (3), 348–354. https://doi.org/10.3969/j.issn.1001-1749.2015.03.13 (in Chinese).

Liu, X.Y., Zhou, L., Chen, X.H., et al., 2020. Lithofacies identification using support vector machine based on local deep multi-kernel learning. Petrol. Sci. 17, 954–966. https://doi.org/10.1007/s12182-020-00474-6.

Lin, T., Zha, H.B., 2008. Riemannian manifold learning. IEEE Trans. Pattern Anal. Mach. Intell. 30 (5), 796–809. https://doi.org/10.1109/TPAMI.2007.70735.

Lu, Z.L., Geng, X.H., Chen, G.M., 2018. A Bayesian assumption based forecasting probability distribution model for small samples. Comput. Electr. Eng. 70, 883–894. https://doi.org/10.1016/j.compeleceng.2017.11.025.

Niyogi, P., Girosi, F., Poggio, T., 1998. Incorporating prior information in machine learning by creating virtual examples. Proc. IEEE 86 (11), 2196–2209. https://doi.org/10.1109/5.726787.

Peng, Y., Wang, W.S., Wang, X.R., et al., 2009. Research of machine learning-based risk prediction. Computer science 36 (4), 205–208.

Suhubi, E.S., 2003. Functional analysis. Istanbul, Turkey. https://doi.org/10.1007/978-94-017-0141-9.

Tang, J., Qiao, J.F., Chai, T.Y., et al., 2018. Modeling multiple components mechanical signals by means of virtual sample generation technique. Acta Autom. Sin. 44 (9), 1569–1589 (in Chinese).

Tang, J.X., Deng, C.W., Huang, G.B., 2016. Extreme learning machine for multilayer perceptron. IEEE .Trans. Neural Network.Learn Syst 27 (4), 809–821. https://doi.org/10.1109/TNNLS.2015.2424995.

Tenenbaum, J.B., Silva, V.D., Langford, J.C., 2000. A global geometric framework for nonlinear dimensionality reduction. Science 290 (5500), 2319–2323. https://doi.org/10.1126/science.290.5500.2319.

Uzair, M., Shafait, F., Ghanem, B., et al., 2018. Representation learning with deep extreme learning machines for efficient image set classification. Neural Comput. Appl. 30 (4), 1211–1223. https://doi.org/10.1007/s00521-016-2758-x.

Wang, R.P., Shan, S.G., Chen, X.L., et al., 2008. Manifold-Manifold distance with application to face recognition based on image set. In: IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, USA, pp. 1–8. https://doi.org/10.1109/CVPR.2008.4587719.

Wu, X.C., Pu, R.H., Wang, D.H., et al., 2019. Three-dimensional seismic analysis of late paleozoic coal-bearing series reflections in the hangjinqi, North Ordos Basin, China. Interpretation 7 (1), 67–82. https://doi.org/10.1190/int-2018-0087.1.

Wu, Z.Z., Fan, Y.J., Ma, W.J., 2012. Spot speed prediction model based on Grey neural network. J. Southwest Jiaot. Univ. 47 (2), 285–290. https://doi.org/10.3969/j.issn.0258-2724.2012.02.019 (In Chinese).

Yang, Z.X., Tang, L.L., Zhang, K., et al., 2018. Multi-view CNN feature aggregation with ELM auto-encoder for 3D shape recognition. Cognitive Computation 10,

908–921. https://doi.org/10.1007/s12559-018-9598-1.
Yin, X.Y., Ye, D.N., Zhang, G.Z., 2012. Application of kernel fuzzy C-means method to reservoir prediction. J. China.Univ. Pet 36 (1), 53–59. https://doi.org/10.3969/j.issn.1673-5005.2012.01.009 (In Chinese).
Zeng, Y.Y., Liu, J.Q., Yang, C.Z., et al., 2018. A machine learning based system performance prediction model for small reactors. Nucl. Power Eng. 39 (1), 117–121. https://doi.org/10.13832/j.jnpe.2018.01.0117 (In Chinese).
Zhang, F.C., Liu, H.Q., Niu, X.M., et al., 2014. High resolution seismic inversion by convolutional neural network. OGP 49 (6), 1165–1169. https://doi.org/10.13810/j.cnki.issn.1000-7210.2014.06.019 (In Chinese).
Zheng, R.N., Liu, W.B., 2017. Research on SAR image target recognition based on virtual sample. Mach. Electron. 35 (6), 12–17 (In Chinese).
Zhu, B., Chen, Z.S., Yu, L.A., 2016. A novel mege-trend-diffusion for small sample. CIESC Journal 67 (3), 820–826 (In Chinese).