Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 261-271
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

Prediction and Analysis of Shale Gas Well Pressure based on Multimodal Elman

Ju Li
Chongqing Open University & Chongqing Technology and Business Institute, Chongqing, 400052 China

Abstract:

In the actual shale gas exploitation process, there are strong nonlinear, multi-characteristic, time series problems among the production data, which make the hidden information among variables not fully expressed, resulting in low pressure prediction accuracy. Therefore, a prediction method based on multimodal Elman is proposed to analyze the pressure of shale gas wells. The main contributions are as follows: (1) constructing the maximum covariance feature set as the direction axis for projection, so as to maximize the degree of heterogeneity between the original data. (2) Based on this, a multimodal Elman model is constructed for prediction and analysis. Compared with Elman algorithm, it can obtain more potential variables, thus increasing the prediction accuracy of the model. The experimental results show that the application in shale gas production proves the effectiveness and feasibility of the proposed method.

***Keywords**: shale gas; multimodal Elman; time series; covariance.*

I. INTRODUCTION

In the major deployment process of promoting the national energy transformation and fully implementing the national goal of "emission peak and carbon neutrality" [1], the next few years will still be the golden period of clean energy. The low-carbon fossil energy with the most room for development, natural gas, will be closely linked with renewable energy for a long time, and shale gas will be the main force in the future main battlefield. China's shale gas storage has exceeded one trillion cubic meters, and China has become the leading country in the world in this regard. However, under the pressure of development, use and environmental protection, how to use the current effective data link to ensure the service life of shale gas wells is indeed a current difficulty and focus.

In the process of shale gas extraction, the data has the characteristics of time series, nonlinearity and dynamics, so it has become the focus to effectively extract the inevitable value in the data. Because of the dynamic, seasonal and periodic characteristics, it is difficult for traditional methods to build a

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 261-271
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

suitable mathematical model to describe shale gas time series data. Therefore, modeling and prediction of time series data has gradually become a hot topic in data science and engineering [2,3]. At present, the forecasting methods for time series can be divided into two categories: statistics and learning methods. Statistics-based methods mainly establish statistical analysis mechanism forecasting models, such as moving average model [4], multivariate statistical method [5], artificial neural network [6-8], support vector machine [9-10], etc. However, these models have some limitations, such as over-reliance on data distribution and stability assumptions, and lack of generalization ability when applied to industrial process production. With the rapid growth of data volume and variability in production process, learning-based modeling methods such as fuzzy similarity-based method [11] and deep neural network [12-14] are showing their promising capabilities including strong data fitting ability and parallelism. However, in terms of shale gas wells, the complex seepage characteristics with multi-scale and multi-flow spaces are different from those of conventional gas reservoirs. The established mechanism model can not fully consider the unique seepage, adsorption and diffusion phenomena of shale gas, and it is difficult to accurately describe the characteristics of shale reservoir. On the one hand, with the continuous development of shale gas, the gas well pressure shows a downward trend. On the other hand, in actual production, the output of gas wells needs to be adjusted frequently with the change of natural gas demand and production level, which is called shale gas adjustable well. Limited by the existing sensing technology, only several typical daily data such as air content, water content and casing pressure can be obtained. Generally, casing pressure is the main factor to guide production scheduling, and the increase of water content in casing often leads to severe fluctuation of casing pressure [15].

To solve the nonlinear and time series problems of data in casing production process, this paper first proposed a new multimodal Elman algorithm for data prediction for shale gas pressure prediction system. Firstly, the data are projected by covariance projection method to obtain a visual distribution map, and whether there are many overlapping variables among the data and whether the projected data can well reflect the hidden dynamic characteristics among the variables are observed. Secondly, based on the multiple correlation characteristics of actual shale gas casing pressure data, a projection matrix under various modes is constructed to project the data in all directions, so that it can extract the potential feature set to the maximum extent. Finally, in the shale gas pressure system, there is a more significant correlation between multi-parameter variables, and it is used for prediction and analysis.

II. COVARIANCE PROJECTION

2.1 t- Visualization of Embedded Data in Distributed Domain

In the original data set of actual shale gas production, there are multiple correlations among the data. Without knowing any characteristics among the data, the t-SNE method is used to process the data by

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 261-271
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

means of dimension reduction and project it into three-dimensional space. The mainstream methods include Principal Components Analysis (PCA), Multi-Dimensional Scaling, MDS), etc. These are the linear dimensionality reduction methods that keep the original distance information as much as possible in conventional manifold learning, and cannot explain the complex polynomial relationship between features. T-SNE method is based on the probability distribution of random walk on neighborhood graph, and its structural relationship can be found in data. Due to the existence of some outlier information, the relative distance between outlier information and other general information remains in the data after dimension reduction, so the similar distance may be too concentrated and the dissimilar distance too scattered after dimension reduction. To ensure that the distance after dimension reduction does not lose the original information at the same time, the t-SNE method based on probability distribution is used for dimension reduction and visualization. The process production data of shale gas well 4-3HF is taken as an example:

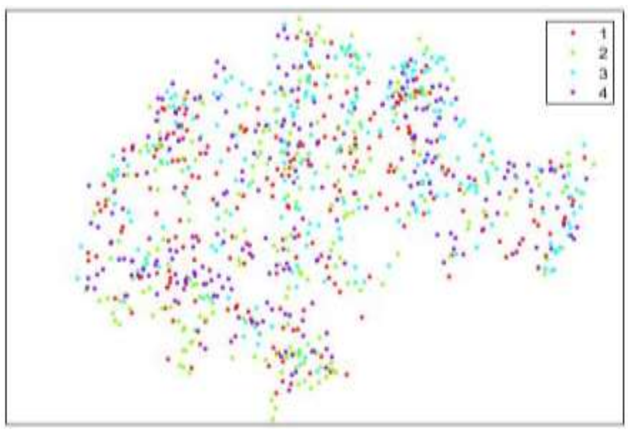


Fig 1: Visualization of production pressure data in shale gas 4-3HF process

It can be seen from Figure 1 that, after the original data related to the well pressure is visually displayed by using t-SNE method, the data of each dimension has a clearer distribution in the t-SNE feature space, and it is found that there is a large amount of data overlap between data sets. To get more potential variables for correlation analysis and effectively predict shale gas production, this paper proposes a new projection method to get more potential variables and useful data sets.

2.2 Covariance Projection

In this section, the covariance projection method is proposed. To effectively extract information from the original data, it is usually necessary to standardize the data. Standardization can greatly highlight the correlation between process variables, eliminate the influence of different measurement dimensions on

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 261-271
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

the model and simplify the structure of data model. However, considering that it is more difficult to extract the hidden information between variables, it is necessary to increase the maximum heterogeneity of data structure while eliminating dimensions, so the covariance projection method is proposed. By obtaining the covariance matrix of the original matrix and taking this matrix as the projection axis, the original data can be projected, so as to increase the maximum variation information between the data and provide important support for subsequent correlation analysis and prediction.

The basic idea of covariance projection analysis is to consider the influence of covariates when analyzing the observed variables that may affect the dependent variables, and then analyze the influence of control variables after eliminating the influence of covariates. The covariance matrix of X is:

$$c = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n-1} \tag{1}$$

Each dimension data in covariance is projected as the corresponding axis in the original data, so as to obtain a new data matrix: Γ :

$$\Gamma = Xc \tag{2}$$

The following figure 2 shows that under the maximum feature covariance projection, the structural separation of each feature variable in the original data is maximized, so that the potential features can be extracted to the maximum extent.

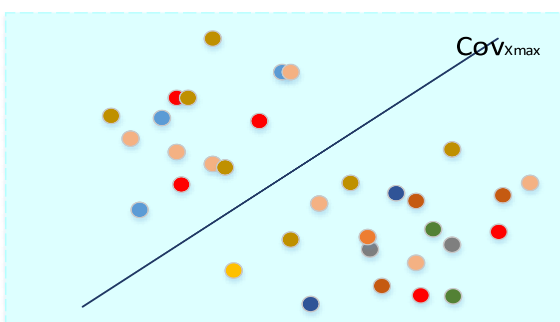


Fig 2: Covariance projection of pressure data set

III. MULTIMODAL NEURAL NETWORK

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 261-271
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

3.1 Elman Neural Network

Elman neural network can not only reflect the dynamic characteristics of the system but also store information. There is information delay in the network, and there is feedback of delay information. The characteristics of recursive network storage information come from the feedback of network signals. The recursion of signals makes the entry and exit state of the network at a certain moment k not only related to the input state at the moment k, but also related to the signals before the moment k, which fully shows the dynamic performance of the network system and is perfectly suitable for modeling and forecasting shale gas production data with time series. The structure diagram is as follows figure 3:

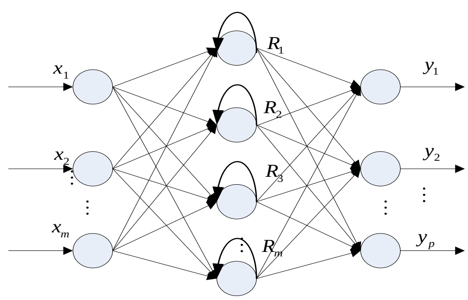


Fig: 3 Schematic diagram of Elman neural network

The relational expression of each layer is:

Input layer:

$$x_i^0(k) = x_i(k) \qquad (i = 1, 2, \cdots, n_0) \tag{3}$$

Hidden layer:

$$\begin{cases} s_i^1(k) = \sum_{j=1}^{n_0} w_{ij}^0 x_j^0(k) + \sum_{j=1}^{n_1} w_{ij} c_i^2(k) \\ x_i^1(k) = f1(s_i^1(k)) \end{cases} \qquad (i = 1, 2, \cdots, n_1) \tag{4}$$

Association layer:

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 261-271
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

$$\begin{cases} s_i^2(k) = x_i^1(k-1) \\ c_i(k) = s_i^2(k) \end{cases} \quad (i = 1, 2, \cdots, n_1) \tag{5}$$

Output layer:

$$\begin{cases} s_i^3(k) = \sum_{j=1}^{n_1} w_{ij}^1 x_j^1(k) \\ y_i(k) = f2\left(s_i^3(k)\right) \end{cases} \tag{6}$$

Wherein, $x_i^0(k)$ is the input of the first node of Elman network; $s_i^1(k)$ and $x_i^1(k)$ respectively represent the input and output of the first node of the hidden layer; $s_i^2(k)$ and $c_i(k)$ respectively represent the input and output of the first node of the correlation layer; $s_i^3(k)$ and $y_i(k)$ respectively represent the input and output of the first node of the output layer; $f1(\cdot)$ and $f2(\cdot)$ respectively represent the activation functions of the hidden layer and the output layer; n_0 is the number of input layer nodes; n_1 is the number of hidden layer nodes; $y(k)$ represents the output of the output layer node; $f(\cdot)$ represents the activation function of the hidden layer; w_{ij}^0, w_{ij}^0 and w_{ij}^1 represent hidden layer, correlation layer, output layer and join weights, respectively.

$x_i^0(k)$ is the input of the i-th node; $s_i^1(k)$ and $x_i^1(k)$ respectively represent the input and output of the i-th node of the hidden layer; $s_i^2(k)$ and $c_i(k)$ respectively represent the input and output of the i-th node of the correlation layer; $s_i^3(k)$ and $y_i(k)$ respectively representing the input and output of the i-th node of the output layer; $f1(\cdot)$ and $f2(\cdot)$ respectively represent the activation functions of the hidden layer and the output layer; n_0 is the number of input layer nodes; n_1 is the number of hidden layer nodes; $y(k)$ represents the output of output layer nodes; $f(\cdot)$ represents the activation function of the hidden layer; w_{ij}^0, w_{ij}^2 and w_{ij}^1 represent hidden layer, correlation layer, output layer and join weights, respectively.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 261-271
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

IV. EXPERIMENTAL RESEARCH

In this section, the actual production pressure data of shale gas wells are used to verify the effectiveness of the proposed algorithm multimodal Elman. In actual industrial production, shale gas pressure is mainly determined by variables such as water, oil pressure, casing pressure and daily output, and cannot be directly modeled as a separate data source. We extracted 468 sets of data from December 15th, 2019 to July 15th, 2020. The training data set consists of 463 sets of data and the test data set consists of 5 sets of data. In the experiment, the production data of shale gas process is modeled by multimodal Elman algorithm, and the accuracy and reliability of the algorithm is analyzed.

If the shale gas output of that day is 100,000 hours (4-3HF), the specific number of days predicted is 5, and the historical real data after training are selected as follows (with casing pressure as output):

TABLE I. 4-3 Comparison of HF actual value and predicted value

Press Mpa (Raw data)	14.58	14.46	14.35	14.38	14.25	14.1	14.31	15.32
Forecast result (5 day)								
Press Mpa (Prediction)	14.29	14.75	14.36	13.74	14.54			
Press Mpa (actual value)	14.19	14.79	14.55	14.38	14.23			

It can be seen from Table 1 that after fully mining the hidden variables of shale gas by covariance projection method and carrying out correlation analysis, the predicted values of multi-mode Elman algorithm for different production rates are very close to the actual values, with good prediction results. They can provide reliable basis for the staff in the subsequent production adjustment.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 261-271
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

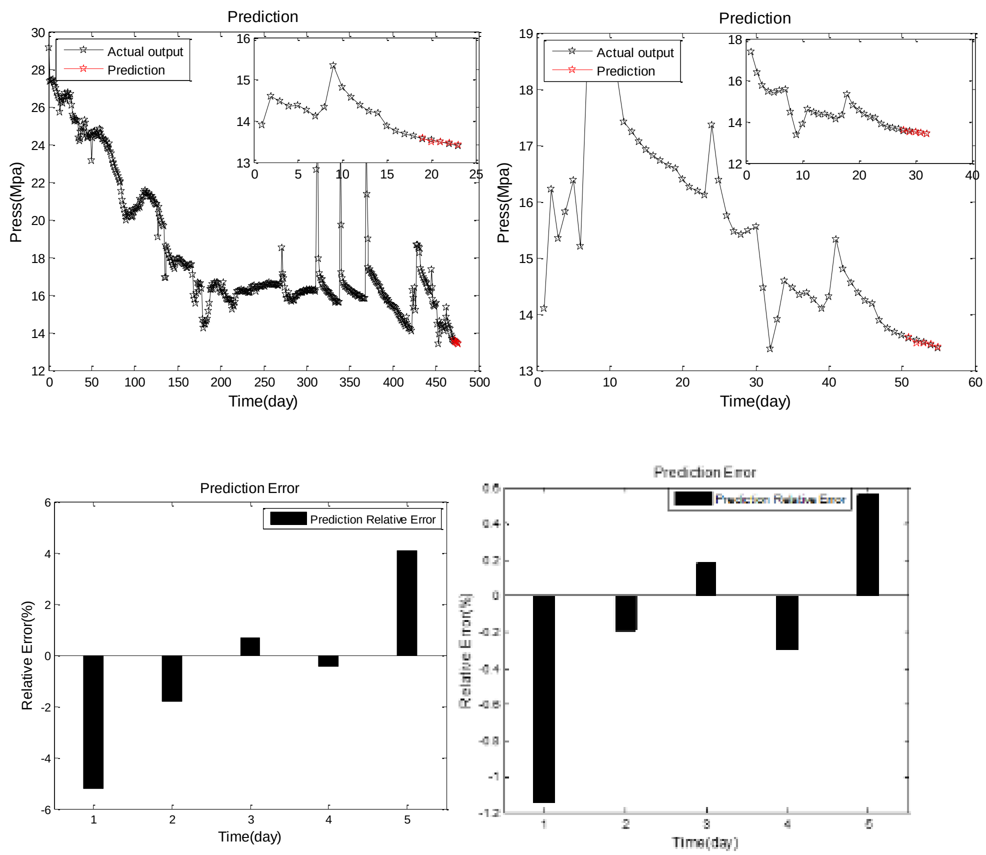


Fig 4: Prediction result and error of multimodal Elman neural network (4-3HF)

Predicted value; actual pressure value; prediction results; pressure; predicted error; time (days) predicted absolute error

As can be seen from Figure 4, when the multi-mode Elman neural network is used to predict the local model of each type of production adjustment well 4-3HF under different types and different production rates, the prediction accuracy can achieve the expected effect by comparing the predicted output value with the actual data of the day.

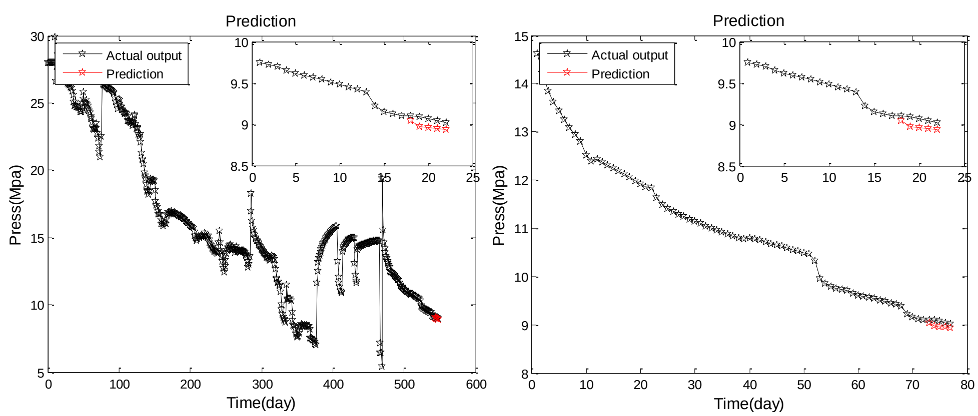
If the shale gas output of that day is 100,000 hours (24-1HF), the specific number of predicted days is 5, and the historical real data after training is selected as follows (with casing pressure as output):

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 261-271
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

TABLE II. 24-1 Comparison of actual value and predicted value of HF

<table>
<tr><td>Casing pressure Mpa (history)</td><td>8.87</td><td>8.85</td><td>8.83</td><td>8.8</td><td>8.8</td><td>8.78</td><td>8.76</td><td>8.73</td></tr>
<tr><td colspan="9">Casing pressure prediction result (5 days)</td></tr>
<tr><td>Casing pressure Mpa (forecast)</td><td colspan="2">8.76</td><td colspan="2">8.80</td><td colspan="2">8.69</td><td>8.74</td><td>8.75</td></tr>
<tr><td>Casing pressure Mpa (true)</td><td colspan="2">8.72</td><td colspan="2">8.71</td><td colspan="2">8.69</td><td>8.67</td><td>8.66</td></tr>
</table>

It can be seen from Table 2 that after fully mining the hidden variables of shale gas by covariance projection method and carrying out correlation analysis, the predicted values of multi-mode Elman algorithm for different production rates are very close to the actual values, with good prediction results.



Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 261-271
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

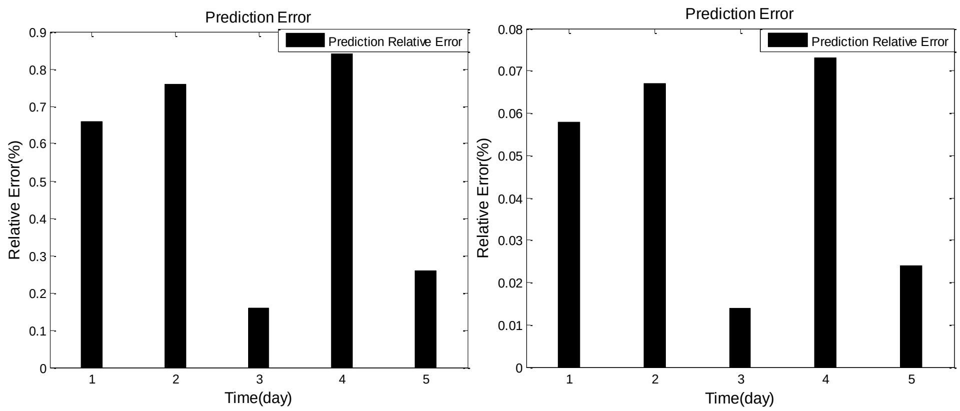


Fig 5: Prediction result and error of multimodal Elman neural network (24-1HF)

As can be seen from Figure 5, when the local model of each type of multimodal Elman neural network is used to predict the production adjustment well 24-1HF under different classification and output, the prediction accuracy is found to achieve the expected effect after comparing the prediction output of neural network with the actual data.

V. CONCLUDING REMARKS

This method is suitable for shale gas production adjustment wells in Jiaoshiba, Fuling. Firstly, a method suitable for production data preprocessing of shale gas production adjustment wells is proposed, and the correlation of different variables is extended by covariance projection algorithm. Secondly, a multimodal Elman neural network model is proposed, which eliminates the influence of abnormal data and local differences on the model, so as to construct a good neural network model for prediction. Compared with the traditional Elman neural network, the prediction accuracy is improved, but there is still room for improvement in the convergence speed of the network.

ACKNOWLEDGEMENTS

This work is funded by Research Foundation of Chongqing Open University (Chongqing Technology and Business Institute), the project No. is NDYB2021-04.

REFERENCES

[1] Editorial Department of Environmental Protection, actively respond to climate change and strive to achieve the new peak goal and carbon-neutral vision, Environmental Protection, no.20, 2020.
[2] Khani H, Farag HEZ, An online-calibrated time series based model for day ahead natural gas demand

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2021 Page No. 261-271
Article History: Received: 22 July 2021 Revised: 16 August 2021 Accepted: 05 September 2021 Publication: 31 October 2021

forecasting, IEEE Trans. Ind. Inf. 15 (4), 2019: 2112–2123.

[3] Zhou Y, Arghandeh R, Zou H, Spanos CJ, Nonparametric event detection in multiple time series for power distribution networks, IEEE Trans. Ind. Electron. 66 (2), 2019: 1619–1628.

[4] Gershenfeld NA, Weigend AS. The Future of Time Series: Learning and Understanding. Addison-Wesley, MA, USA, 1993.

[5] Wang XX, Yao YX, An C, et al. Simultaneous determination of 20 bioactive components in Chuanxiong Rhizoma from different production origins in Sichuan province by ultra-high performance liquid chromatography coupled with triple quadrupole mass spectrometry combined with multivariate statistical analysis. ELECTROPHORESIS, 2020, 6(19): 1-30.

[6] Ren WJ, Tian YC, Zhu YB. Application of CS Neural Network in-Pumping Units' Fault Diagnosis. Journal of Bionic Engineering, 2017, 35(3):324-332.

[7] Zhang L, Tan Z, Liu C, et al. Research on optimal operation method of pumping station based on machine learning. 2017 IEEE Conference on Energy Internet and Energy System Integration, 2017: 1-6.

[8] Sriram LMK, Gilanifar M, Zhou Y, Ozguven EE, Arghandeh R, 2019. Causal markov elman network for load forecasting in multinetwork systems. IEEE Trans. Ind. Electron. 66 (2), 1434–1442.

[9] Li K, Gao XW, Tian Z, et al. Using the curve moment and the PSO-SVM method to diagnose downhole conditions of a sucker rod pumping unit. Petroleum Sci, 2013, 10(1): 73–80.

[10] Yu DL, Zhang YM, Bian HM, et al. A new diagnostic method for identifying working conditions of submersible reciprocating pumping systems. Petroleum Sci, 2013, 10(1): 81–90.

[11] Han M, Zhang S, Xu M, Qiu T, Wang N, Multivariate chaotic time series online prediction based on improved kernel recursive least squares algorithm. IEEE Trans. Cybern. 49 (4), 2019: 1160–1172.

[12] Xing Y, Lv C, 2019. Dynamic state estimation for the advanced brake system of electric vehicles by using deep recurrent neural networks. IEEE Trans. Ind. Electron. 1.

[13] Lei T, Yin X, Zong Z, 2020. Pore pressure prediction in orthotropic medium based
on rock physics modeling of shale gas. J. Natl. Gas Sci. Eng. 74 (103091).

[14] Application of Elman and Cascade neural network (ENN and CNN) in comparison with adaptive neuro fuzzy inference system (ANFIS) to predict key fuel properties of ABE-diesel blends

[15] Zhang R, Zhang L, Tang H, Chen S, Zhao Y, Wu J, Wang K, 2019b. A simulator for production prediction of multistage fractured horizontal well in shale gas reservoir considering complex fracture geometry. J. Natl. Gas Sci. Eng. 67, 14–29.