Development Index Prediction Through Big Data Analysis for QX Ultra-Deep Permian Marine Carbonate Gas Reservoir in Sichuan Basin, China

Xiaohua Liu[1(✉)], Xuliang Liu[2], Zhenhua Guo[1], Jichun Zhou[3], and Daolun Li[2]

[1] PetroChina Research Institute of Petroleum Exploration and Development, Xueyuan Road, No. 20, Haidian District, Beijing 100083, China
lxh69@petrochina.com.cn

[2] Hefei University of Technology, Hefei, Anhui, China

[3] Northwestern Sichuan Gas District, PetroChina Southwest Oil & Gas Company, Jiangyou, Sichuan, China

Abstract. Uncertainties in the characterization of new-found, ultra-deep, thin and low porosity Permian gas reservoir reduce feasibility for development index (DI) prediction through reservoir simulation. DI prediction with big data analysis approach are studied. Geology and production data from 30 mature gas fields are reviewed and 13 parameters are selected to represent geological features, deliverability and DI of individual reservoir. Based on the BP neural network algorithm, proxy models are established to correlate DI with geology and deliverability data, and the bagging method is used to effectively improve the experimental accuracy and stability while avoiding over-fitting phenomenon in the case of limited sample data. The coefficient of determination coefficient (R^2) are selected to evaluate the prediction effect of DI. The mean value of the prediction results of the model with higher R^2 value in 2000 numerical experiments was selected as the final prediction result. With the established proxy model, DI for QX reservoir in Permian formation are predicted and the influence of heterogeneity are also evaluated.

Keywords: Development index prediction · BP neural network · big data analysis · ultra-deep gas reservoir · Permian marine carbonate

J. Lin (Ed.): IFEDC 2023, SSGG, pp. 99–112, 2024.
https://doi.org/10.1007/978-981-97-0272-5_8

1 Introduction

The Ultra-deep (>7000 m) Permian marine carbonate formation is a prospective conventional gas exploration and development domain in Sichuan Basin, China, and in recent years, significant breakthrough have been made in some appraisal wells with testing gas rates above 1.0 MMm^3/D from Middle Permian QX gas reservoir in Sichuan Basin, China (hereinafter referred to as QX reservoir). To meet market needs for clean energy, the full development of this reservoir is put on agenda.

The QX reservoir is structurally located at the Longmen Mountain buried thrust front zone, and contains many NE-SW trending faulted anticline and faulted nosing structures, and currently, 6 NE-SW trending tectonic high belts have been defined through seismic and a rough structure map is given in Fig. 1. Due to limited drillings and complex structure, the extension of faults, communication between each belt and gas and water distribution in QX reservoir are still uncertain, which reduce feasibility for a full reservoir modeling and simulation.

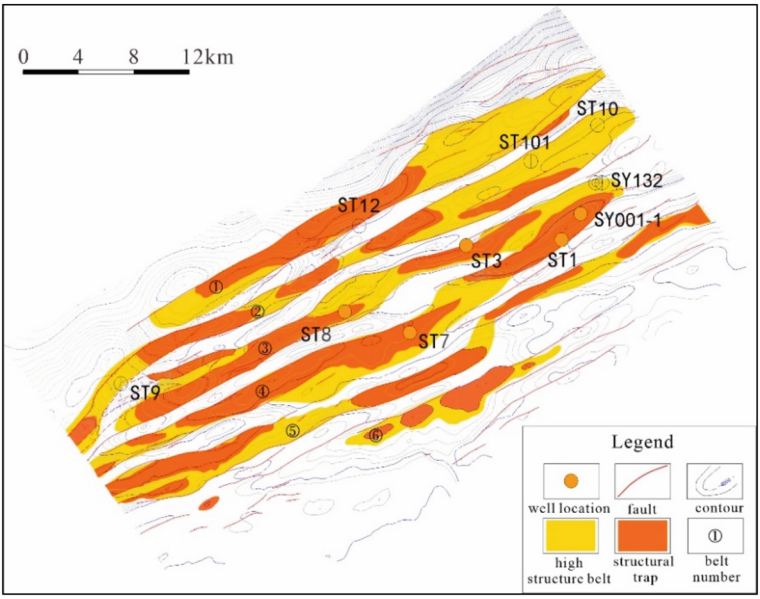


Fig. 1. Top structure map of Permian QX reservoir, SYS Block, Sichuan Basin

Current drilling, geology and geophysics studies show that QX reservoir is featured with ultra-deep buried depth (7200–7800 m), high pressure (>93 MPa), low porosity (3.9% in average) and thin layer (average pay zone thickness 20 m). Due to uneven development of natural fractures and vugs, heterogeneity exists in this reservoir with permeability ranging among 0.01–10 mD. The uncertainty and heterogeneity in this ultra-deep, thin layered and low porosity reservoir pose risks in cost-effective development, to lower risks in initiating exploitation activities, proper Development Index (DI) for

guiding the commercial and steady development of the new findings are the key concerns of management.

Usually, in Field Development Plan (FDP), full reservoir modeling and simulation will be performed to predict DI which consists of a series of parameters including Field Annual Production Rate (FAPR), Field Plateau Period (FPP) at certain FAPR, Well Spacing Density (WSD), Well Average Daily Production (WADP) during FPP, and field Ultimate Recovery Factor (URF). And these key index will direct operators to make drilling plan and development policies. But this common approach is less reliable in QX reservoir due to uncertainties in reservoir characterization. Recently, big data analysis technique (Safavian et al., 1991; Quinlan, 1986; Rao et al., 2019; Franco-Lopez et al., 2001; Gou et al., 2019; Thierry et al., 2019; Burges et al., 1998; Chapelle et al., 1999; Janik et al., 2006; Torkaman et al., 2015) provides novel, efficient and economical tools for reservoir engineering and has been proved to be a powerful tool in production forecast. Some researchers use big data technology to build proxy model by correlating the complex, non-linear relationship among parameters to forecast flow rates and hydrocarbon recoveries (Panja et al., 2017; Zhong et al., 2020; Ng et al., 2021; Li et al., 2021; Shen et al., 2022; Zha et al., 2021; Zhou et al., 2014), and in some literatures, big data technology have been utilized to facilitate reservoir simulation in saving run time and cost, or improving accuracy in history matching (Ke et al., 2017; Cheng et al., 2019; Luciana et al., 2020; Feng et al., 2019), and some researchers use big data technology to guide stimulation design by correlating the fracturing parameters into post stimulation oil production prediction model (Zhu et al., 2015).

But less literature is presented to forecast overall DI for a raw gas reservoir with big data analysis technique. The purpose of this paper lies in the point that how we utilized the geology and production history data in developed reservoirs to facilitate the exploitation of new findings. Geology, dynamic and DI data from 30 mature gas fields are collected and processed, and 13 parameters are selected to represent geological features, deliverability and DI of individual reservoir. Through BP Neural Network, proxy models are established to correlate DI with geology and deliverability data. Moreover, the stability of the predicted results are also considered, to avoid randomness in a single experiment, Bagging method (Eugene et al., 2022) is used to make the results more stable for cases with limited samples. With the established models, overall DI for QX reservoir are then given based on current drilling and testing information, and risks caused by heterogeneity are also discussed. The results can serve as a criteria for directing the successful development of this ultra-deep marginal pools.

2 Data Acquisition and Processing

Geology, deliverability and DI data of 30 major mature gas reservoirs from different gas-bearing basins in China are reviewed. With per capita porosity among 3.4%–28.6% and per capita dynamic permeability ($K_{dynamic}$, permeability from well test interpretation) ranging from 0.1–38.5 mD, these reservoirs contain sandstones and carbonate rocks with or without natural fractures. Based on post FDP implementation evaluation of DI, these reservoirs, with 15–691 10^9 m^3 in OGIP and 0.3–10.7 10^9 m^3/a in actual plateau gas production, are all believed to be successfully developed reservoirs. In data

preparation, logging and dynamic data of 1500^{+} wells in these reservoirs are reviewed to better understand the productivity and its dominating factors of individual reservoir. And 13 parameters are selected and listed in Table 1 to represent geological features, deliverability and DI of individual reservoir. To make sure that these parameters fully represent reservoir characteristics, a lot of reservoir engineering study are conducted, especially in the selection of productivity related parameters, such as permeability and well AOF. We have two sets of reservoir permeability, which are matrix permeability (K_{matrix}) obtained in well logging interpretations or core testing and dynamic permeability ($K_{dynamic}$) calculated in well test interpretation. The correlations of permeability (both K_{matrix} and $K_{dynamic}$) *vs* porosity, and $K_{dynamic}$ *vs* K_{matrix} shown in Fig. 2 indicate that the porosity for most reservoirs are quite low (<10%), but the permeability varies considerably, and inconsistency exists between $K_{dynamic}$ and K_{matrix} due to the development of natural fractures. Figure 3 indicates that one of the DI parameters—FPR is more dependent on $K_{dynamic}$ than K_{matrix}, and it can be seen in Fig. 4 that $K_{dynamic}$ also dominate well deliverability (AOFP).

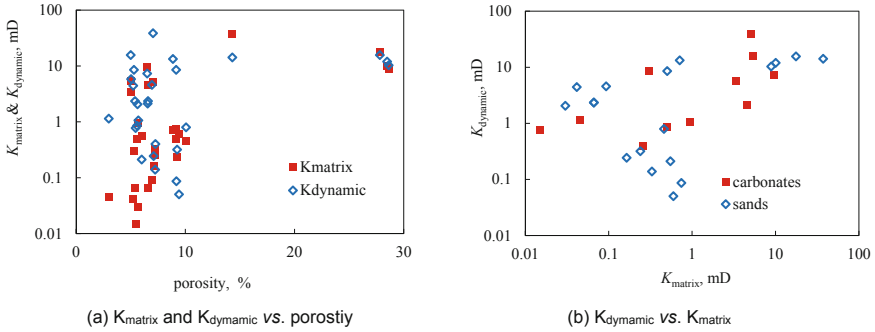
Table 1. Geology, deliverability and DI parameters for individual reservoir

Parameter type	No. of parameters	parameters	scope of values in 30 reservoir samples	Values in QX reservoir
Geology	6	reservoir depth, m	910–6800	7500
		reservoir pressure, MPa	9.8–115.5	96
		pressure coefficient, MPa/100 m	0.85–2.12	1.28
		reserves abundance, $10^9 m^3/km^2$	0.1–5.9	0.32
		average porosity, %	3.4–28.6	3.7
		average K_{matrix}, mD	0.01–37.3	0.51
Deliverability	2	average $K_{dynamic}$, mD	0.1–38.5	2.0
		well average AOFP, $10^3 m^3/d$	68–9695	1420

(*continued*)

Table 1. (*continued*)

Parameter type	No. of parameters	parameters	scope of values in 30 reservoir samples	Values in QX reservoir
DI	5	Field Annual Production Rate (FAPR), %	0.18–4.11	2.5
		Field Plateau Period (FPP), a	5–20	9–11
		Well Spacing Density (WSD), km^2/well	0.4–10.5	5–6
		Ultimate Recovery Factor (URF), %	37.4–75.0	62
		Well Average Daily Production, $10^3m^3/d$	4–1907	280–300



(a) K_{matrix} and $K_{dymamic}$ *vs.* porostiy (b) $K_{dymamic}$ *vs.* K_{matrix}

Fig. 2. K_{matrix}, $K_{dymamic}$ and porostiy in 30 sample fields

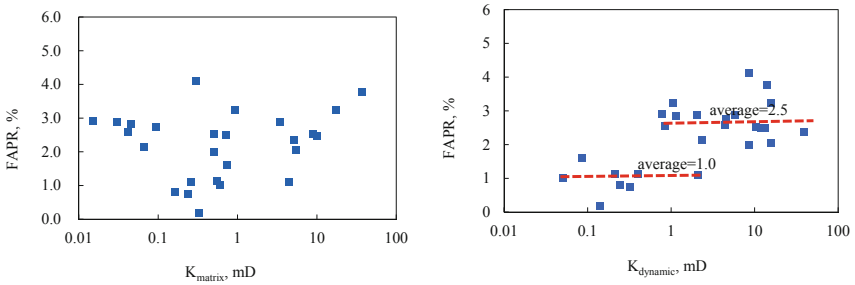


Fig. 3. FAPR *vs.* K_{matrix} and $K_{dynamic}$ in 30 sample fields

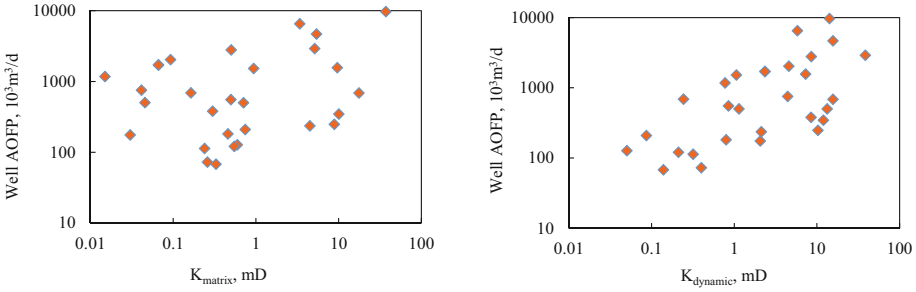


Fig. 4. Well AOFP *vs.* K_{matrix} and $K_{dynamic}$ in 30 sample fields

3 BP Neural Network Algorithms

The BP (Back Propagation) neural network is a data mining technique in developing correlation models between input variables and output variables in big data analysis. In the following we would briefly describe the main algorithms.

The BP neural network, a concept introduced by scientists in 1986, is a multilayer feed-forward neural network trained according to an error back propagation algorithm and is one of the most widely used neural network models (Burks et al., 2000, Meinel et al., 2010). Currently, the vast majority of neural network models used in the practical application of artificial neural networks are in the form of BP networks and variations of it.

The BP algorithm consists of two processes: the forward propagation of the signal and the backward propagation of the error (Hecht-Nielsen, 1989). In forward propagation, the input samples are passed in from the input layer, processed in turn by the hidden layer and then passed to the output layer. If the actual output of the output layer does not match the desired output, it moves to the back propagation of error stage. The BP network consists of an input layer, an output layer and a hidden layer, and the structure of the BP neural network is as follows (Fig. 5).

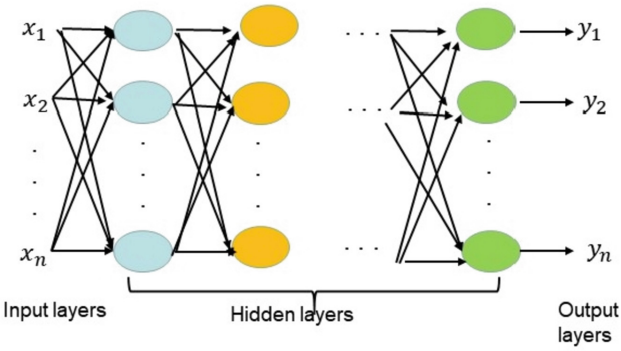


Fig. 5. Structure of BP neural network

The specific steps are: let the input vector is $X = (x_1, x_2, \ldots, x_n)$, the input vector of the hidden layer is $hi = (hi_1, hi_2, \ldots, hi_p)$, the output vector of the hidden layer is $ho = (ho_1, ho_2, \ldots, ho_p)$, the input vector of the output layer is $yi = (yi_1, yi_2, \ldots, yi_q)$, the output vector of the output layer is $yo = (yo_1, yo_2, \ldots, yo_q)$, the desired output vector is $d_o = (d_1, d_2, \ldots, d_q)$.

The input and output of each neuron in the hidden layer are calculated by randomly selecting the kth input sample.

$$hi_h(k) = \sum_{i=1}^{n} w_{ih} x_i(k) - b_h \quad h = 1, 2, \ldots, p \tag{1}$$

$$ho_h(k) = f(hi_h(k)) \quad h = 1, 2, \ldots, p \tag{2}$$

$$yi_o(k) = \sum_{h=1}^{p} w_{ho} ho_h(k) - b_o \quad o = 1, 2, \ldots, q \tag{3}$$

$$yo_o(k) = f(yi_o(k)) \quad o = 1, 2, \ldots, q \tag{4}$$

where w_{ih} is the connection weight of the input layer to the middle layer, w_{ho} is the connection weight of the hidden layer to the output layer, b_h is the threshold of each neuron in the hidden layer, and b_o is the threshold of each neuron in the output layer, $f()$ is the activation function.

Initialize the error function with a random number within $(-1, 1)$ and set the precision ε. With a maximum number of iterations M, the error function is

$$e = \frac{1}{2} \sum_{o=1}^{q} (d_o(k) - yo_o(k))^2 \tag{5}$$

Calculate the partial derivatives of the error function with respect to each neuron in the output layer and calculate the parameters of each layer with following equation:

$$\frac{\partial e}{\partial w_{ho}} = \frac{\partial e}{\partial yi_o} \frac{\partial yi_o}{\partial w_{ho}} = \delta_o(k) ho_h(k) \tag{6}$$

$$\frac{\partial e}{\partial yi_o} = \frac{\partial\left(\frac{1}{2}\sum_{o=1}^{q}\left(d_o(k) - yo_o(k)\right)\right)^2}{\partial i_o}$$
$$= -(d_a(k) - yo_o(k))yo'_o(k) \cdot \overline{n}e(d_o(k) - yo_o(k))f(yi_o(k)) = \delta_o(k) \quad (7)$$

$$\frac{\partial yi_o(k)}{\partial w_{ho}} = \frac{\partial\left(\sum_h^p w_h ho_h(k) - b_o\right)}{\partial w_{ho}} = ho_h(k) \quad (8)$$

Calculate the partial derivatives of the error function for each neuron in the hidden layer, the connection weights that follow, and the input values for that layer,

$$\frac{\partial e}{\partial hi_h(k)} = -\left(\sum_{h=0}^{q}\delta_0(k)w_{ho}\right)f'(hi_h(k)) = \delta_h(k) \quad (9)$$

$$\frac{\partial hu_h(k)}{\partial w_{ih}} = x_i(k) \quad (10)$$

$$\frac{\partial e}{\partial w_{ih}} = \delta_h(k)x_i(k) \quad (11)$$

Use (6) (7) (8) to correct the output layer connection weights,

$$\Delta w_{ho}(k) = -\mu\frac{\partial e}{\partial w_{ho}} = \mu\delta_o(k)ho_h(k) \quad (12)$$

$$w_{ho}^{N+1} = w_{ho}^{N} + \eta\delta_o(k)ho_h(k) \quad (13)$$

Use (9) (10) (11) to correct the hidden layer connection weights,

$$\Delta w_{ih}(k) = -\mu\frac{\partial e}{\partial w_{ih}} = -\mu\frac{\partial e}{\partial hi_h(k)}\frac{\partial hi_h(k)}{\partial w_{ih}} = \delta_h(k)x_i(k) \quad (14)$$

$$w_{ih}^{N+1} = w_{ih}^{N} + \eta\delta_h(k)x_i(k) \quad (15)$$

Finally, calculate the global error,

$$E = \frac{1}{2}\sum_{k=1}^{m}\sum_{o=1}^{q}\left(d_o(k) - y_o(k)\right)^2 \quad (16)$$

4 Bagging

Bagging is a parallel method of ensemble learning, where data is sampled and the results are voted on. For a given data set containing multiple samples, we randomly select one sample into the sampling set and put that sample back into the initial data set, making it still possible for it to be selected for the next sampling. Combining Bagging with BP neural network. The model is trained several times to get the average of the predicted values. It can improve the accuracy and stability of prediction while avoiding the over-fitting phenomenon.

5 DI Prediction Proxy Model Development Through Big Data Analysis

5.1 Correlation Coefficient Calculation

As we want to build DI prediction model through big data analysis, geology and deliverability parameters listed in Table 1 are categorized as characteristic data and DI parameters in the table are defined as target output variables. In proxy model development, initially, the coefficient of correlation r (Bookbinder et al., 1987) between characteristic data and target output data are calculated and those characteristic data with high absolute r values are selected as input parameters. Table 2 presents the calculated r values between characteristics data and DI, and those characteristic data with underlined values are selected as inputs.

Table 2. Calculated r values between Characteristic Data and DI

DI parameters	Depth	Pressure	Pressure Coefficient	Reserves Abundance	K_{matrix}	Porosity	$K_{dynamic}$	AOFP
URF	**0.358**	**0.333**	0.204	0.298	0.273	−0.176	**0.401**	**0.522**
FPP	−0.006	**0.146**	**0.247**	0.065	**0.168**	0.139	−0.127	0.136
FAPR	−0.023	0.12F9	0.334	**0.385**	0.190	0.204	**0.392**	**0.416**
WSD	0.223	0.225	0.101	**−0.318**	-0.091	**−0.450**	0.123	0.167
WADP	0.326	0.494	0.624	**0.735**	**0.745**	−0.007	0.388	**0.948**

Note:those characteristic data with underlined values are selected as inputs

5.2 Proxy Model Development

In proxy model development, BP Neural Network is used to establish the relationship between input variables and output variables. We design different combinations of correlated variables as input models. For example, we design four input models for predicting UFR and three input models for predicting FPR, as shown in Tables 3 and 4 respectively.

Table 3. Combinations of correlated variables as inputs for predicting URF

Input models	Depth	Pressure	$K_{dynamic}$	AOFP
Model 1	1	1	1	1
Model 2	1	0	1	1
Model 3	0	1	1	1
Model 4	0	0	1	1

Note:1 means the variable is used, 0 means it is not used

Table 4. Combinations of correlated variables as inputs for predicting FAPR from gas fields

Input models	Reserves Abundance	$K_{dynamic}$	AOFP
Model 1	1	1	1
Model 2	0	1	1
Model 3	1	0	1

Note:1 means the variable is used, 0 means it is not used

As limited sample data may introduce randomness and occasionality in model development, thus weaken model credibility, to avoid these disadvantages, samples data are disordered in each training with 80% and 20% being selected randomly for model training and model validating respectively. For fixed inputs and outputs, the risks of occasionality caused by limited sample data also exist if only single numerical test is conducted, to tackle this problem, 2000 numerical tests are performed and those models with high coefficient of determination (R^2) of test set are selected as best fit models. All best fit models are used to predict the DI value, and then the average is calculated to obtain the final prediction result. R^2 is generally used in regression models to evaluate the degree of conformity between predicted values and actual values, and R^2 is defined as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

where: $\bar{y}$ denotes the average of the true target values. The higher the score of the R^2, the closer the predicted value of the sample is to the true value.

In summary, we calculate the correlation coefficients between the predictor variables and other variables, find the most relevant variables to the predictor variables. Then use the bagging-based BP Neural Network to establish the relationship between the predictor variables and the relevant variables. Finally, the training effect is evaluated by R^2. And the final prediction value is obtained by averaging from the better prediction results. Whole process of the algorithm is described in the Appendix, and Table 5 shows the network parameter settings for WADP prediction experiments.

Table 5. Optimal values for each parameter in the predicted WADP

Parameters	Setting
Units	10, 64, 128, 256 …, 64, 1
Epochs	Period: 500
activation	Relu
optimizer	Adam

Figure 6 shows, as an example, the prediction results of the two experiments with higher R^2 in 2000 prediction experiments of WADP. As depicted in Fig. 6, sound fitting

can be observed between prediction and actual values. The predicted value in Fig. 6(a) is basically consistent with the measured value, while the predicted value in Fig. 6(b) is slightly deviated from the measured value, but the error is still small in the case of a small amount of data. Experimental results show that the bagging-based BP neural network has high precision in DI prediction. In addition, this analysis method is easily scalable with the addition of the latest machine learning methods.

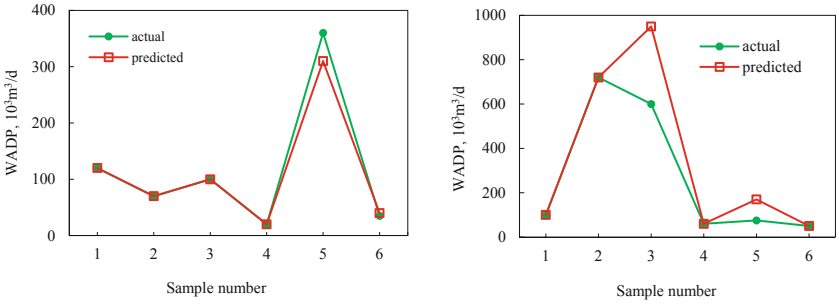


Fig. 6. WADP prediction models validation. (a) and (b) represent different models with different numerical testing samples

5.3 DI Prediction for QX Reservoir

Current drilling, geology and well testing data in QX reservoir are reviewed, and characteristic parameters including geology and deliverability data are evaluated based on our understanding of the reservoir. The quantifying of these parameters will be discuss below and their values are presented in Table 1. Reservoir mid-depth based on drilling wells is 7500m with initial reservoir pressure of 96 MPa and pressure gradient 1.28 MPa/100 m; reservoir porosity from both core analysis and logging interpretations are among 2.0%–6.0%, with 3.7% in average; K_{matrix} from core analysis range from 0.01 mD to 53 mD, and 0.51 mD in mean; K_{dynamic} obtained through 9 wells' test interpretations are ranging in the scope of 0.1–10 mD, with 2.0 mD in average, reflecting the improvement of mobility with the development of natural fractures; AOFP from both horizontal wells and vertical wells are among 0.15–3.85 10^6 m^3/d, with average 1.42 10^6 m^3/d; based on logging and pressure data, average reserve abundance is evaluated as 0.32 10^9 m^3/km^2.

DI for QX reservoir are predicted through our proxy models with input parameters given in Table 1, and the output results shown in Table 1 are as follow: FAPR 2.5%, WSD 5–6 km^2/well, WADP during FPP 280–300 10^3 m^3/d and URF 62%. Heterogeneity caused by lithology change or uneven development of natural fractures can be evidenced from both core samples and deliverability data, as in the low part of structure, well dynamic permeability are in the magnitude of 0.1mD. The influence of heterogeneity on FAPR and URF are also predicted in the proxy models, and results presented in Fig. 7 show that in the "tight" part of the reservoir, the feasible FAPR decreases from 2.5% to

1.5%, and URF declines from 62% to 50%, so economic risk exist in the development of QX reservoir. The effects of horizontal drilling on FAPR and URF are also evaluated and depicted in Fig. 7, and it can be seen that compared with vertical drilling (with average AOFP 1.0MMm^3/d), horizontal drilling (with average AOFP 1.4MMm^3/d) show limited enhancement in both FAPR and URF.

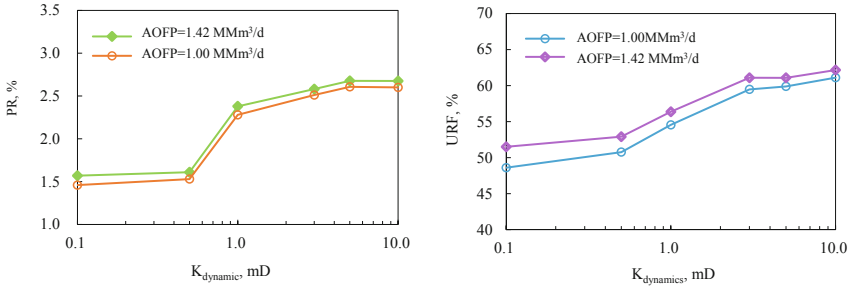


Fig. 7. Predicted *PR* and *URF vs.* K_{dynamic} with different *AOFP*

It should be note the proxy model are based on data from 30 mature, successfully developed major reservoirs, and the DI of these reservoirs contain certain development policies followed currently by the operators. So the predicted DI for QX reservoir can serve as a criteria for directing the successful development of this ultra-deep marginal pools.

6 Conclusions

1. DI prediction models for raw gas reservoirs are established through big data analysis approach. Geology and dynamic data from 30 mature gas reservoirs are reviewed, and 12 parameters are selected to represent geology, deliverability and DI data for individual reservoir, then proxy model are built through bagging-based BP neural network to correlated DI with geology and deliverability data.
2. Experimental results show that the bagging-based BP neural network has high precision in DI prediction in the case of limited sample data.
3. Based on geology and dynamic data, the DI for QX reservoir are predicted in the proxy model with results as following: FAPR 2.5%, FPP 9–11 years, URF 62%, WSD 5–6 km^2/well, WADP during FPP 280–300 10^3 m^3/d. Sensitivity analysis showed that for relative "tight" area, 1.5% of FAPR with UFR of 50% are expected.
4. Through big data analysis, the development polices formed in mature gas fields can provide valuable knowledge in the development of ultra-deep raw gas fields, thus mitigating risks due to uncertainties in reservoir characterization.

Funding. This study was supported by the Scientific and Technology Research Program Funded by CNPC, China (Project No. 2021DJ1505 and 2022KT0905).

Appendix

Algorithm for Prediction Model Devlopment

Input: Select variable combination models as input

Initialize training data and set test number

repeat

$k \leftarrow k+1$

for $j=1$ **to** N **do in parallel**

$\hat{y}_j \leftarrow M_j(X,y)$

Calculate R^2 value c_j, get set $S_j=\{\hat{y}_j, c_j\}$

end for

Select $y_s \in S_j$ where j corresponds to $c_j \geq Average(\sum_{j=1}^{N} c_j)$

until $k=K$

$\hat{y} = Average(\sum y_s)$

return $\hat{y}$

References

Bookbinder, M.J., Panosian, K.J.: Using the coefficient of correlation in method-comparison studies. Clin. Chem. **7**, 1170–1176 (1987)

Burges, C.: A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Disc. **2**(2), 121–167 (1998)

Burks, T.F., Shearer, S.A., Gates, R.S., et al.: Backpropagation neural network design and evaluation for classifying weed species using color image texture. Trans. Asae **43**(4), 1029–1037 (2000)

Chapelle, O., Haffner, P., Vapnik, V.N.: Support vector machines for histogram-based image classification. IEEE Trans. Neural Netw. **10**(5), 1055–1064 (1999)

Cheng, Z., Sankaran, S., Lemoine, V., et al.: Application of machine learning for production forecasting for unconventional resources. Paper URTeC-2019-47 Presented at Unconventional Resources Technology Conference, Colorado (2019)

Eugene, L., Chieh-Hsin, L., Hsien-Yuan, L.: A bagging ensemble machine learning framework to predict overall cognitive function of schizophrenia patients with cognitive domains and tests. Asian J. Psychiatr. **69**, 103008 (2022)

Feng, C., Li, J., Feng, Z., et al.: Predict oil production from geological and petrophysical data before hydraulic fracturing using an improved particle swarm optimization based least squares support vector machine. Paper SPE 197250 Presented at Abu Dhabi International Petroleum Exhibition & Conference (2019)

Franco-Lopez, H., Ek, A.R., Bauer, M.E.: Estimation and mapping of forest stand density, volume and cover type using the k-nearest neighbors method. Remote Sens. Environ. **77**, 251–274 (2001)

Gou, J.P., Ma, H.X., Ou, W.H., et al.: A generalized mean distance-based k-nearest neighbor classifier. Expert Syst. Appl. **115**(1), 356–372 (2019)

Hecht-Nielsen, R.: Theory of the backpropagation neural network. Neural Netw. (1989)

Janik, P., Lobos, T.: Automated classification of power-quality disturbances using SVM and RBF networks. IEEE Trans. Power Deliv. **21**(3), 1663–1669 (2006)

Ke, G.L., Meng, Q., Thomas, F., et al.: Light GBM: a highly efficient gradient boosting decision tree. Neural Inf. Process. Syst. (2017)

Li, D.L., Shen, L.H., Zha, W.S., et al.: Physics-constrained deep learning for solving seepage equation. J. Petrol. Sci. Eng. **206**, 1–11 (2021)

Luciana, M.D.S., Guilherme, D.A., Denis, J.S.: Support vector regression for petroleum reservoir production forecast considering geostatistical realizations. SPE Reservoir Eval. Eng. **23**(04), 1343–1357 (2020)

Meinel, L.A., Stolpen, A.H., Berbaum, K.S., et al.: Breast MRI lesion classification: improved performance of human readers with a backpropagation neural network computer-aided diagnosis (CAD) system. J. Magn. Reson. Imaging **25**(1), 89–95 (2010)

Ng, C.S.W., Ghahfarokhi, A.J., Amar, M.N.: Well production forecast in Volve field: application of rigorous machine learning techniques and metaheuristic algorithm. J. Petrol. Sci. Eng. **208**, 109468 (2021)

Panja, P., Velasco, R., Pathak, M., et al.: Application of artificial intelligence to forecast hydrocarbon production from shales. Petroleum 75–89 (2017)

Quinlan, J.R.: Induction of decision trees. Mach. Learn. **1**(1), 81–106 (1986)

Rao, H.D., Shi, X.Z., Rodrigue, A.K., et al.: Feature selection based on artificial bee colony and gradient boosting decision tree. Appl. Soft Comput. **74**, 634–642 (2019)

Safavian, S.R., Landgrebe, D.: A survey of decision tree classifier methodology. IEEE Trans. Syst. Man Cybern. **21**(3), 660–674 (1991)

Shen, L.H., Li, D.L., Zha, W.S., et al.: Surrogate modeling for porous flow using deep neural networks. J. Petrol. Sci. Eng. **213**, 110460 (2022)

Thierry, D., Orakanya, K., Songsak, S.: A new evidential K-nearest neighbor rule based on contextual discounting with partially supervised learning. Int. J. Approximate Reasoning **113** (2019)

Torkaman, M., Safari, et al.: A novel PSO-LSSVM model for predicting liquid rate of two phase flow through wellhead chokes. J. Natural Gas Sci. Eng. **24**, 228–237 (2015)

Zha, W.S., Zhang, W., Li, D.L., et al.: Convolution-based model-solving method for three-dimensional, unsteady, partial differential equations. Neural Comput. 1–23 (2021)

Zhong, Z., Sun, A.Y., Wang, Y., et al.: Predicting field production rates for waterflooding using a machine learning-based proxy model. J. Petrol. Sci. Eng. **194**, 107574 (2020)

Zhou, Q.M., Robert, D., Andrew, K., et al.: Evaluating gas production performances in Marcellus using data mining technologies. J. Nat. Gas Sci. Eng. **20**, 109–120 (2014)

Zhu, L., Li, M.S., Wu, Q.H., et al.: Short-term natural gas demand prediction based on support vector regression with false neighbours filtered. Energy **80**(2), 428–436 (2015)