Energy Proceedings
Vol 42, 2024
ISSN 2004-2965

Machine Learning Evaluation Method of Inter-well Sand Body Connectivity by CatBoost Model: A Field Example from Tight Oil Reservoir, Ordos Basin, China

Baobiao Pu [1], Renyi Cao [1*], Gaofei Yan [1], Zhihao Jia[1], Linsong Cheng[1], Yongchao Xue[1]

1 China University of Petroleum (Beijing), Beijing, 102249, PR China

(*Corresponding Auther: caorenyi@126.com)

ABSTRACT

Tight oil reservoirs are mainly developed by water injection. The conventional method could not effectively characterize the sand body connectivity between oil and water wells since it only uses static geological parameters or dynamic production data to identify the connectivity between wells. In this paper, a novel machine learning evaluation method for inter-well sand body connectivity using both static geological and dynamic production data is constructed. This method is based on the CatBoost algorithm. Three types of sand body connectivity are classified by analysis of four geological factors including porosity, permeability, shale volume and net-to-gross ratio, and three dynamic production factors including oil production total (OPT), liquid production total (LPT) and water injection total (WIT). Finally, the novel method proposed in this paper is applied to predict the sand body connectivity between oil and water wells using the data from a tight oil reservoir located in the Ordos Basin, China. The results show that the proposed method can improve the forecast accuracy of inter-well sand body connectivity from 50% to 85%.

Keywords: Tight oil reservoir; Sand body connectivity; CatBoost model; Machine Learning Evaluation Method; Parameters optimization

NONMENCLATURE

Symbols	
a	Prior term weight
i	Number of trees
j	Return to leaf node region
k	Number of leaf nodes
n	Parameter number
P	Prior value
x	Normalized data
$x*$	Data before normalization
x^{*}_{min}	The minimum value of the data before normalization
x^{*}_{max}	The maximum value of the data before normalization
$\overline{x}_i$	The average value of the *I*-th feature over the entire dataset
$\overline{x}_i^{(j)}$	The average value of the *I*-th feature on a class *j* dataset
$\tilde{x}_{k,i}^{(j)}$	The eigenvalue of the *I*-th feature on the *k*-th sample point of class *j*

1. INTRODUCTION

Tight oil reservoirs have been the key of crude oil production in the future because of the exhaustion of traditional petroleum resources[1]. Tight oil reservoirs are mainly developed by complement producing energy using water injection. However, it is challenging to accurately divide sub-layers and create effective displacement due to reservoir heterogeneity and unknown sand-body connectivity between oil and water wells.

The reasonable fine division of sub-layers, which is based on the accurate characterization of sand body connectivity, is very important for the development of tight oil reservoirs. The empirical technique of the division of sub-layers, analysis method of oilfield static geology parameter, the analysis method of dynamic production data, the stratal slice method, the machine learning method, etc. are the primary prediction methods for sand body connectivity. The machine learning method can be used to predict samples with unknown sand body connectivity by training and learning from samples with known sand body connectivity. Particularly, the Catboost algorithm, one of the machine

# This is a paper for 15th International Conference on Applied Energy (ICAE2023), Dec. 3-7, 2023, Doha, Qatar.

learning methods, has the advantages of low sample demand, high training accuracy, quick calculation speed, and high training result accuracy. Allen[2],[3] originally proposed the idea and theory of the crucial net-to-gross ratio value and established the probability model of sand body connection, emphasizing that the connectivity of sand bodies is impacted by the geometry of sand bodies. In the experiment, Hovadik[4] came to the conclusion that connectivity increased significantly with increasing net-to-gross ratio within a specific data range, and the curve of connectivity vs. net-to-gross ratio was shape of "S". Based on the porous flow theory, King P.R.[5] has focused on the spatial connectivity of complex sand bodies. Both present meander river sediments and prehistoric fluvial reservoir outcrops were described hierarchically by Andrew D. Miall[6]. A three-dimensional spatial simulation model was established by Paola C. and Mohrig D.[7] taking into account the thickness of various reservoir sand bodies. Thin and narrow channel outcrops on the surface would nonetheless offer geological data for the description of sand bodies in subsurface rivers, according to Eschard R., Lemouzy P., Bacchiana C.[8] et al. To describe the sand bodies, Webb E. K. and Davis J.M.[9] proposed a series of simulation methods for research on the paleogeomorphic environment and sedimentary channel of the basin. The net-to-gross ratio model and the numerical simulation model were combined into a new model, which David K.L. and Hovadik J.[10] then used to carry on statistical calculations on the values of various net-to-gross ratios. By examining the sedimentary characteristics and evolution rules of multi-stage single sand bodies, Minh N.H.[11] determined the distribution characteristics and scale range of sand bodies. When the net-to-gross ratio is less than 20%, the connectivity is weak, according to Pranter M.J. and Sommer N.K.[12]. The connectivity increases quickly when the net-to-gross ratio is more than 30%. In order to characterize the inter-well sand body connectivity of actual oil fields, Ford G.L. and Pyles D.R.[13] developed a fine sand body model for complex block reservoirs of fluvial facies. A modeling method with a compression algorithm and multi-point statistics (MPS) was proposed by Walsh D. A. and Manzocchi T.[14] to create a system model with a high net-to-gross ratio and low connectivity. In order to construct a forward formation model that takes into account outcrop and subterranean river systems, Colombera L. and Mountney N. P.[15] incorporated matrix and fracture model data from numerous river systems, quantified fracture aperture, and highlighted the coupling relationship between river channels and fractures.

The objective of this paper is to establish a evaluation method of the sand-body connectivity of tight oil reservoirs based on the CatBoost algorithm, which could improve the prediction accuracy and efficiency.

2. METHODOLOGY

2.1 CatBoost model

The CatBoost algorithm, which could handle categorical variables, has low dependence on hyperparameters and good accuracy, is typically used to address the issue of efficiently processing categorical information, which is shown in Fig. 1. Through preceding terms, CatBoost lessens the negative effects of noise and low-frequency data, and its basic formula is as follows:

$$\hat{x}_k^i = \frac{\sum_{j=1}^{p-1}\left[x_{\sigma j,k} = x_{\sigma p,k}\right]\cdot Y_{\sigma j} + a\cdot P}{\sum_{j=1}^{p-1}\left[x_{j,k} = x_{\sigma p,k}\right]\cdot Y_{\sigma j} + a} \tag{1}$$

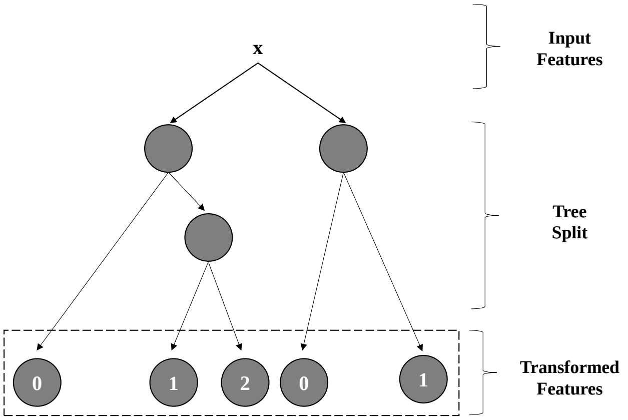
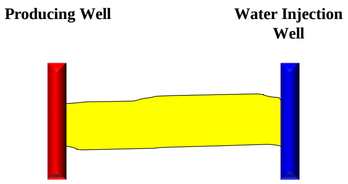


Fig. 1. Schematic diagram of CatBoost algorithm

2.2 Optimization method of parameters

In this paper, the continuous quantitative characterization problem was changed into a qualitative analysis model and classification, and it defines sand body connection in terms of sand body connectivity judgment. The connectivity of the sand body is classified into three levels: level I, which indicates strong connectivity; level II, which indicates moderate connectivity; and level III, which indicates weak connectivity. The schematic diagram of different connectivity types are shown in Figs. 2 (a)–(c).



(a)

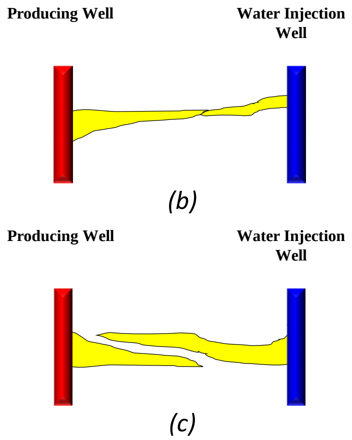


Fig. 2. Sand body connectivity evaluation sample diagram. (a) Sand body connectivity is strong; (b) Sand body connectivity is moderate; (c) Sand body connectivity is weak.

According to geological factors including reservoir physical characteristics, sand body distribution and heterogeneity, which are reflected in OPT, LPT, water cut increasing rate (WCIR), and WIT, the sand body connectivities are totally different. Therefore, it can be forecasted using static parameters like porosity, permeability, width-to-thickness ratio, shale volume, net-to-gross ratio, and permeability variation coefficient, as well as dynamic parameters like OPT, LPT, WCIR, and WIT. More input features may result in more thorough information, but in fact, too many features might slow machine learning, decrease prediction accuracy, and degrade model performance. Therefore, it is essential to optimize the static and dynamic parameters that impact the connectivity of the sand body while retaining the model's accuracy.

The static and dynamic parameters associated with sand body connectivity were optimized using the enhanced F-score method. Give an explanation of the training sample set $X_k \in \mathrm{R}_m$ (k = 1, 2,..., n); l ($l \geqslant 2$) is the number of sample categories, and n_j is the number of samples of class j, where j = 1, 2,..., l. The I-th feature's F-score in the training sample is then defined as follows:

$$F\left(i\right)=\frac{\sum_{i=1}^{l}\left(\bar{x}_i^{(j)}-\bar{x}_i\right)^2}{\sum_{j=1}^{l}\frac{1}{n_j-1}\sum_{k=1}^{n_j}\left(x_{k,j}^{(j)}-\bar{x}_i^{(j)}\right)^2} \tag{2}$$

The higher the F-score number, the bigger the difference between categories and the smaller the difference within categories, meaning the feature could better differentiate itself according to the standard of classification.

2.3 Data set construction

Each set of sand bodies corresponding to oil wells and water wells can be regarded as an independent dataset. The input variables of samples are made up of both static and dynamic parameters. The results of logging interpretation make up the static parameters. The production parameters obtained by dividing the liquid production profile and the water injection profile make up the dynamic parameters. Finally, 101 samples in total were collected. Fig. 3 shows the flow chart of sand body connectivity evaluation using the dynamic and static combined based on the CatBoost model.

It is necessary to normalize the sample data prior to training and then use the CatBoost model for training in order to increase the convergence speed during training, increase the accuracy of the evaluation, and avoid operation difficulties and even errors caused by unbalanced data distribution, which will affect the training results and lead to errors in the subsequent results. The data of the sample is limited to 0~1 by the normalization formula, and data normalization processing can reduce the training computational complexity and improve the accuracy of the training results. The normalization formula can be expressed by:

$$x=\frac{x^*-x_{\min}^*}{x_{\max}^*-x_{\min}^*} \tag{3}$$

There were 38 Level III samples with poor sand body connectivity, 34 Level II samples with medium sand body connectivity, and 29 Level I samples with good sand body connectivity among them. A training sample set was randomly chosen from the 101 samples, consisting of 23 Level I samples, 28 Level II samples, 29 Level III samples, and the remaining 21 samples as test samples.

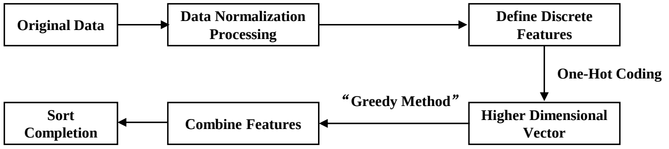
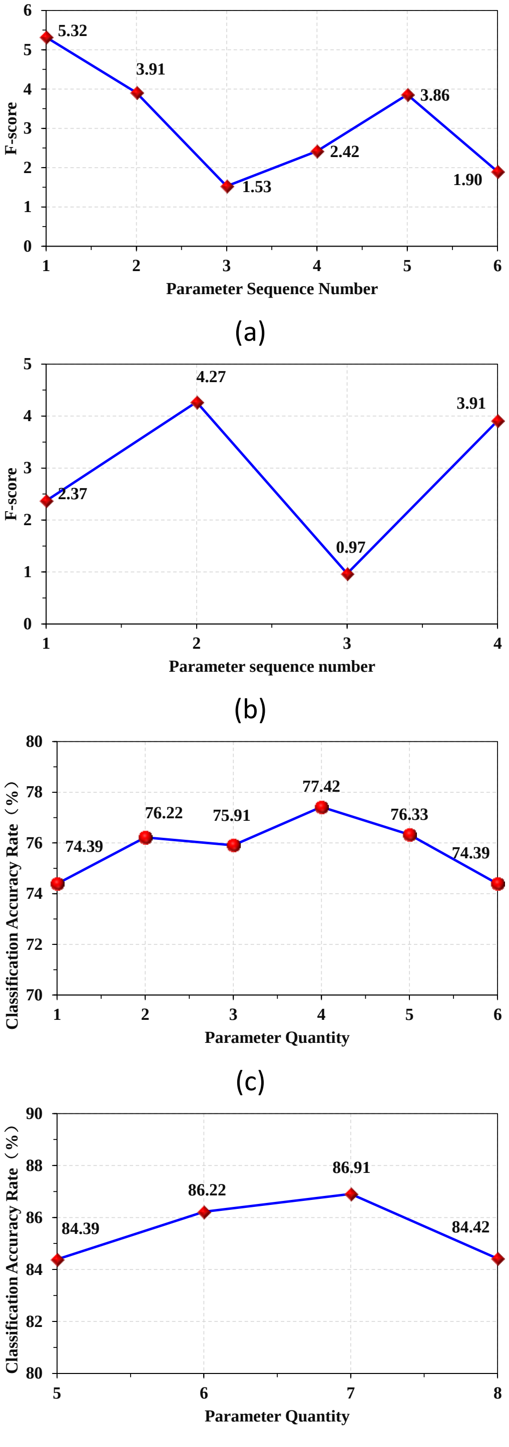


Fig. 3. Flow chart of sand body connectivity evaluation

3. *RESULTS AND DISCUSSION*

3.1 Optimal selection of dynamic and static parameters

Figs. 4 (a)–(d) show the F-score values of dynamic and static dynamic parameters as well as the correlation between the number of parameters and classification accuracy rate using the enhanced F-score method.



(a)

(b)

(c)

(d)

Fig. 4. Optimization results of static and dynamic parameters. (a) and (b) are the F-score values of static and dynamic parameters, respectively; (c) and (d) are the classification accuracy of static and dynamic parameters, respectively.

The results show that X = {porosity, permeability, shale volume, net-to-gross ratio, OPT, LPT and WIT}. When the dynamic and static parameters are both used to evaluate sand body connectivity, the classification accuracy of sand body connectivity is greatly increased compared to only using static parameters. Therefore, the evaluation method of sand body connectivity, both using dynamic and static parameters, is more complete and precise.

3.2 *Prediction results of sand body connectivity*

The prediction results of sand body connectivity show that 80 sample sets are trained by the model, with a training accuracy rate of 87.5%. The model is then tested using the remaining 21 samples, and the predicted results are contrasted with the actual sand body connectivity, as shown in Fig. 5. Fig. 6 shows the confusion matrix heat map of the anticipated results based on SPSSPRO. It can be observed from the examination of Figs. 5 and 6 show that 18 of the 21 test samples agreed with the actual results, and 3 of them had incorrect judgments. The accuracy of the prediction results of the model proposed in this paper is significantly better than the traditional method only using static data or experience to predict sand body connectivity. In addition, the model proposed in this paper has an accuracy of about 85%.

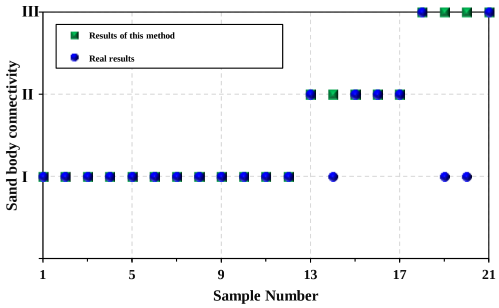


Fig. 5. Comparison of the method proposed in this paper and the traditional method

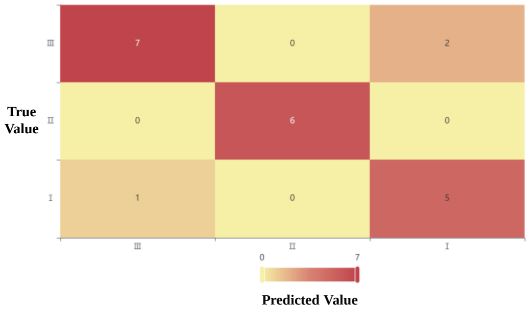


Fig. 6. Confusion matrix heat map of CatBoost classification test results

3.3 *Method validation*

Oil and water wells are chosen based on the target area inter-well sand body connectivity, and the prediction results of inter-well sand body connectivity are subsequently confirmed based on the dynamics between oil and water wells.

It is predicted by the CatBoost algorithm that the sand body between W1 and W2 has Level I connectivity. The curves of W1 water injection rate and W2 dynamic production performance, which are respectively shown in Figs. 7 and 8, are analyzed. The results demonstrate that the sand body connectivity between W1 and W2 Wells is strong, validating the efficacy of the method proposed in this paper.

Moreover, the sand body connectivity between W1 and W2 using the traditional method (only static parameters) and the method proposed in this paper (both static and dynamic parameters) is evaluated, respectively. The results show that the sand body connectivity between W1 and W2 is judged to be moderate connectivity using the traditional method, while it is judged to be strong connectivity using the method in this paper. The accuracy of the method proposed in this paper is further demonstrated.

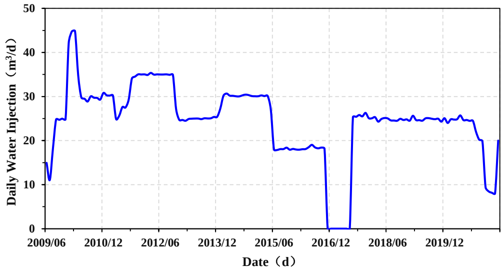


Fig. 7. Curve of W1 water injection rate

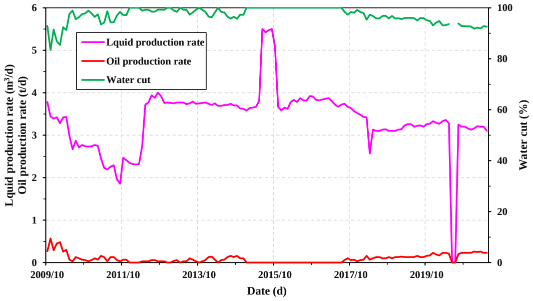
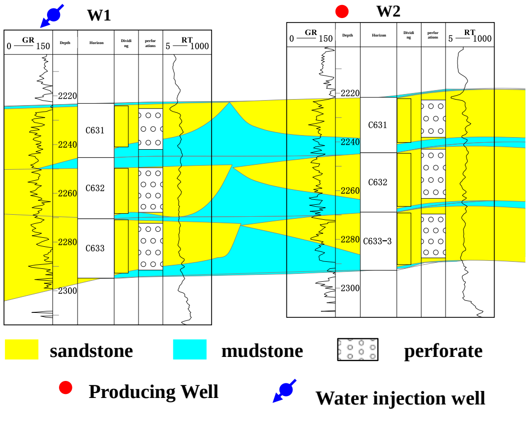
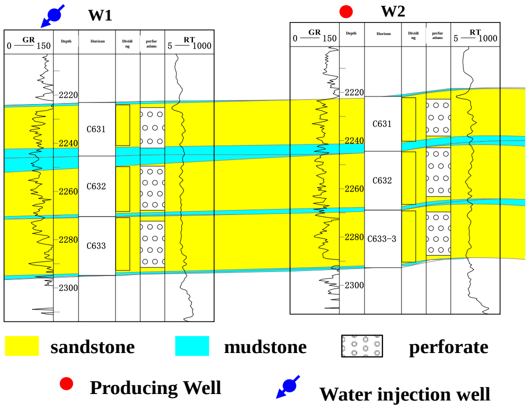


Fig. 8. Curve of W2 dynamic production performance



(a)



(b)

Fig. 9. Analysis of sand body connectivity results. (a) is the results of sand body connectivity analyzed by traditional methods; (b) is the results of sand body connectivity analyzed by the method proposed in this paper.

3.4 Field application

In this paper, the data are collected from a tight oil reservoir, Ordos Basin, China. According to statistics, there are 719 sets of Level I (strong connectivity) out of the 2198 sets of single sand bodies in the target area, making up 32.7% of all sand body groups. There are 335 sets of Level II (moderate connectivity), making up 15.2% of all sand bodies. There are 1044 sets of Level III (weak connectivity), making up 47.5% of all sand bodies. The statistical results show that the sand body connectivity of the target reservoir is generally not very good.

Table 1. Statistics on sand body connectivity

Sand Body Connectivity	Quantity	Proportion (%)
Level I (strong connectivity)	719	32.7
Level II (moderate connectivity)	335	15.2
Level III (weak connectivity)	1044	47.5

The distribution of sand bodies parallel to the maximum principal stress direction is examined using the connecting well profile of Wells W125 to W131 as an example. According to Fig. 10, the distribution of sand bodies parallel to the maximum principal stress direction is relatively continuous, while the connectivity of those from top to bottom is getting poor.

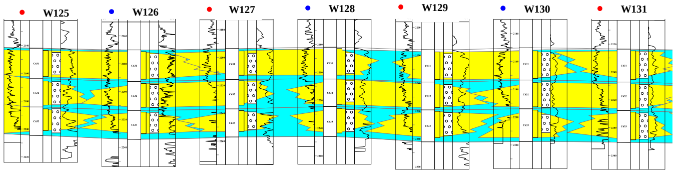


Fig. 10. Connecting well profile of Wells W125 to W131, which is parallel to the maximum principal stress direction

Analysis of the distribution of sand bodies perpendicular to the maximum principal stress direction is done using the connecting well profile of Wells W140 to W146 as an example. According to Fig. 11, the distribution of sand bodies perpendicular to the maximum principal stress direction is more dispersed. In addition, the connectivity of those from top to bottom is getting poor.

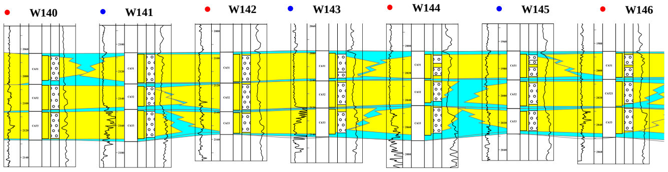


Fig. 11. Connecting well profile of Wells W140 to W146, which is perpendicular to the maximum principal stress direction

4. CONCLUSION

(1) Through the enhanced F-score method, four static parameters are selected, which are porosity, permeability, shale volume and net-to-gross ratio, and three dynamic parameters, which are oil production total, liquid production total and water injection total.

(2) The sand body connectivity prediction method based on the CatBoost model was proposed in this paper, which can be used to predict sand body connectivity, and the prediction accuracy reached 85%, which is better than traditional methods.

(3) The distribution of sand bodies parallel to the maximum principal stress direction is relatively continuous, while the distribution of sand bodies perpendicular to the maximum principal stress direction is more dispersed. In addition, the connectivity of those from top to bottom is getting poor.

ACKNOWLEDGEMENT

The authors acknowledge that this study was partially supported by the National Natural Science Foundation of China (No. 52174038 and No. 52004307), China Petroleum Science and Technology Project-major project-Research on tight oil-shale oil reservoir engineering methods and key technologies in Ordos Basin (ZLZX2020-02-04) and Science Foundation of China University of Petroleum, Beijing (No. 2462018YJRC015).

DECLARATION OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. All authors read and approved the final manuscript.

REFERENCE

[1] Cao R , Jia Z , Cheng L .Using high-intensity water flooding relative permeability curve for edicting mature oilfield performance after long-term water flooding in order to realize sustainable development[J].Journal of Petroleum Science & Engineering, 2022:215PB.

[2] Allen J R L. Studies in fluviatile sedimentation: an exploratory quantitative model for the architecture of avulsion-controlled alluvial suites[J].Sedimentary Geology，1978，21（2）：129-147.

[3] Allen J R L. Studies in fluviatile sedimentation: An elementary geometrical model for the connectedness of avulsion-related channel sand bodies[J]. :Sedimentary Geology，1979，24（3）：253-267.

[4] Hovadik J M, Larue D K. Static characterizations of reservoirs: refining the concepts of connectivity and continuity[J].2007, 13（3）：195-211.

[5] King P R. The Connectivity and Conductivity of Overlapping Sand Bodies[M]. Springer Netherlands, 1990.

[6] Andrew D. Miall. The Geology of Fluvial Depotsits:Sedimentary Facies，Basin Analsis and Petroleum Geology [M]. Berlin，Heidelberg，New York:Spring-Verlag，1996：57-98.

[7] Paola C, Mohrig D. Palaeohydraulics revisited: palaeoslope estimation in coarse-grained braided rivers[J]. Basin Research, 1996, 8(3): 243–254.

[8] Eschard R , Lemouzy P , Bacchiana C , et al. Combining Sequence Stratigraphy, Geostatistical Simulations, and Production Data for Modeling a Fluvial Reservoir in the Chaunoy Field (Triassic, France)[J]. Aapg Bulletin, 1998, 82(4): 545-567.

[9] Webb E K, Davis J M. Simulation of the spatial heterogeneity of geologic properties: an overview Hydrogeologic Models of Sedimentary Aquifers, SEPM (Society for Sedimentary Geology). 1998.

[10] DAVID K L, HOVADIK J.Connectivity of channelzed reservoirs:a modelling approach,Petroleum Geoscience[J].Petroleum Geoscience,2006,12:291-308.

[11] Minh N H. Numerical simulation using disturbed state concept (DSC) model for softening behavior of sand[J] . Geotechnical Engineering，2008，39（1）：25-35.

[12] Pranter M J , Sommer N K . Static connectivity of fluvial sandstones in a lower coastal-plain setting: An example from the Upper Cretaceous lower Williams Fork Formation, Piceance Basin, Colorado[J]. Aapg Bulletin, 2011, 95(6):899-923.

[13] Ford G L , Pyles D R . A hierarchical approach for evaluating fluvial systems: Architectural analysis and sequential evolution of the high net-sand content,

middle Wasatch Formation, Uinta Basin, Utah[J]. Aapg Bulletin, 2014, 98(7):1273-1303.

[14] Walsh D A, Manzocchi T. A method for generating geomodels conditioned to well data with high net:gross ratios but low connectivity[J]. Marine and Petroleum Geology, 2021, 129: 105104.

[15] Colombera L, Mountney N P. Influence of fluvial crevasse-splay deposits on sandbody connectivity: Lessons from geological analogues and stochastic modelling[J]. Marine and Petroleum Geology, 2021, 128.