WEP Warwick Evans Publishing

International Journal of Natural Resources and Environmental Studies

ISSN: 3006-2012 (Print), ISSN: 3006-0834 (Online) | Volume 3, Number 1, Year 2024
DOI: https://doi.org/10.62051/ijnres.v3n1.09
Journal homepage: https://wepub.org/index.php/IJNRES/index

Predicted Open Flow Rate Based on Capacity Test Well Data

Yu Fu, Qinghui He

Petroleum Engineering School, Southwest Petroleum University, Chengdu, China

ABSTRACT

Open flow measurement in carbonate reservoirs directly affects the effectiveness of gas well production evaluation, assessment of reservoir potential, and guidance of safe production. In order to overcome the limitations of the traditional non-resistance flow calculation method, which relies on human adjustment and cannot meet the field demand in real time. In this study, based on the historical data and geological data of the blocks of the target gas reservoir, the production capacity test data of 8 wells and 13 wells were obtained. By using the K nearest neighbor method for data processing, a carbonate rock open flow prediction model based on the data from the capacity well test data was established, with a correlation coefficient of more than 0.92. The results of this study show that the predicted open flow rate is able to simplify the traditional complex mathematical calculation methods. This study provides support for capacity evaluation, assesses the potential of the gas reservoir, and helps the gas reservoir to develop appropriate safety production measures.

KEYWORDS

Carbonate Reservoirs, Open Flow Prediction, Data Processing, Production Measures.

1. INTRODUCTION

With the continuous growth of energy demand and the increasing depletion of traditional energy resources, the development and utilization of unconventional energy resources have attracted more and more attention. Among them, carbonate gas reservoirs, as an important unconventional natural gas resource with huge reserves and wide distribution, have received extensive attention and research from the global energy industry. However, the development and evaluation of carbonate gas reservoirs face a series of challenges due to their complex reservoir characteristics and difficult development.[1]

In the development process of carbonate gas reservoirs, as an important evaluation technology, open flow measurement plays a crucial role in the capacity evaluation of gas wells, the assessment of gas reservoir potential and the guidance of safe production. Currently, the traditional non-resistance flow calculation methods include back-pressure well test method and isochronous well test method[2], etc. Although they are widely used in the field of reservoir capacity evaluation, their limitations of relying on human adjustment and not being able to meet the field demands in real time make them somewhat deficient in terms of the accuracy and practicability of the non-resistance flow rate. Therefore, to address the limitations of the traditional non-resistive flow calculation methods, this study focuses on the carbonate rocks in Block X of the Sichuan Basin, and combines the capacity test well data, non-resistive flow historical data, and geologic data from multiple gas wells, as well as the K nearest-neighbor method of data processing, to establish a predictive model for the non-resistive flow of carbonate reservoirs based on the data from the capacity test well data. The establishment of this prediction model provides a reliable support for the evaluation of gas well production capacity, effectively evaluates the potential of the gas reservoir, and provides a scientific basis for the formulation of corresponding safety production measures in the gas reservoir. The results of this study will be able to more accurately carry out the open flow measurement of gas reservoirs, provide more reliable technical support for the development and management of carbonate gas reservoirs, and promote the scientific, efficient and sustainable development of gas reservoir development.[3]

2. DATA PROCESSING OF PRODUCTION WELL TEST DATA

The production capacity well test data, unhindered flow history data and geological data used in this study were taken from the carbonate rocks of the Lamp IV section in the Gaoshitian~Muoxi area of the Sichuan Basin, with stratigraphic lithology dominated by dolomite, in consolidated contact with the Lamp III section below it, and in unconformable contact with the overlying mudstone-dominated Maidiping and Uichengzhusi formations.[4] West of Gao Shiti~Muoxi area is the Deyang-Anyue rift trough, and the Lamp IV stratum is rapidly extinguished. The Lamp IV section has experienced uplift by the Tongwan movement, and the residual thickness of the strata after undergoing denudation is generally distributed in the range of 280-380 m. In order to ensure the accuracy and reasonableness of the data, we preprocessed the data using the K-nearest-neighbor machine-learning algorithm.

The K-Nearest Neighbors (KNN) algorithm is an instance-based learning method, and its core idea is to classify or regress unknown samples by the information of their nearest neighbors in the feature space. In the KNN algorithm, the class or value of an unknown sample is determined by the class or value of its nearest neighbor. The basic idea of the KNN algorithm can be summarized as follows：

Distance metric: first, the distance between the unknown sample and each sample in the sample set is calculated by the selected distance metric (e.g., Euclidean distance, Manhattan distance, etc.).

Determine the value of K: Determine the number of neighbors K, i.e., select the K samples in the feature space that are closest to the unknown sample as the neighbors of the prediction target.

Voting: for classification problems, count the frequency of occurrence of each category among the K neighbors, and select the category with the highest frequency of occurrence as the predicted category of the unknown sample. For regression problems, the average or weighted average of the K neighbors is usually used as the prediction result.

The formula works as follows

$$\hat{y} = \frac{1}{K} \sum_{x_i \in N_K(x)} y_i \tag{1}$$

where $\hat{y}$ denotes the predicted value of the unknown sample x and y_i denotes the true value of the sample x_i.

Based on the KNN algorithm we can see the difference between the data before and after processing, where the missing values have been automatically filled in.

Table 1 Open flow data processing results

Well Number	Open flow before treatment (104m³)	Open flow after treatment (104m³)
1	88.51	88.51
2	89.41	89.41
3	87.61	87.61
4	0	85.42
5	0	84.52
	…	
15	82.12	82.12
16	83.56	83.56

3. OPEN FLOW PREDICTION IN CARBONATE GAS RESERVOIRS

There are a lot of researches on the prediction of non-resistance flow, which are mostly based on traditional mathematical and statistical algorithms, simulation calculation methods, etc. However, the above methods have strict requirements on data and the proportion of human factors in the technology is large, and their calculation steps are cumbersome, practicality is not strong and they can't specifically analyze the effects of various stratigraphic parameters on the non-resistance flow. In this paper, two algorithms, gradient boosting regression and random forest regression, are selected to construct a machine learning model to predict the open flow rate of target wells.[5]

The root mean square error MSE, correlation coefficient R2 and mean absolute error MAE in the evaluation index are used as the evaluation indexes of the accuracy of the output results of the gradient boosting regression model, in which R2 is suitable for comparing the fitting degree between the simulation results and the real results, and the higher number represents that the results are more close to the reality, and the MSE and MAE represent the model fitting error, and the lower the number indicates that the model fitting effect is more good. The expressions are respectively:

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

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y - \hat{y}_i)^2 \tag{3}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}| y - \hat{y}_i | \tag{4}$$

Gradient Boosting is a powerful integrated learning method for regression and classification tasks. It gradually builds a powerful predictive model by progressively reducing the prediction error of the previous model. Specifically, Gradient Boosting starts with a simple initial model, and then in each round of iterations, fits the residuals (errors) of the current model and adds new weak models to the current model with certain weights. This method can effectively deal with nonlinear relationships and complex datasets and improve the overall model prediction performance. The fitting formula is as follows:

$$F_M(x) = F_0(x) + \sum_{m=1}^{M}\gamma_m h_m(x) \tag{4}$$

Establish a gradient lifting prediction model based on the XGBOOST algorithm, select the relevant parameters in the production capacity test well data as the characteristic variables, and the open flow rate as the target variable for prediction, resulting in a model R2 (correlation coefficient) of 0.92, which has a high accuracy rate, and can be used as a guide for the production time. (See Figure 1)

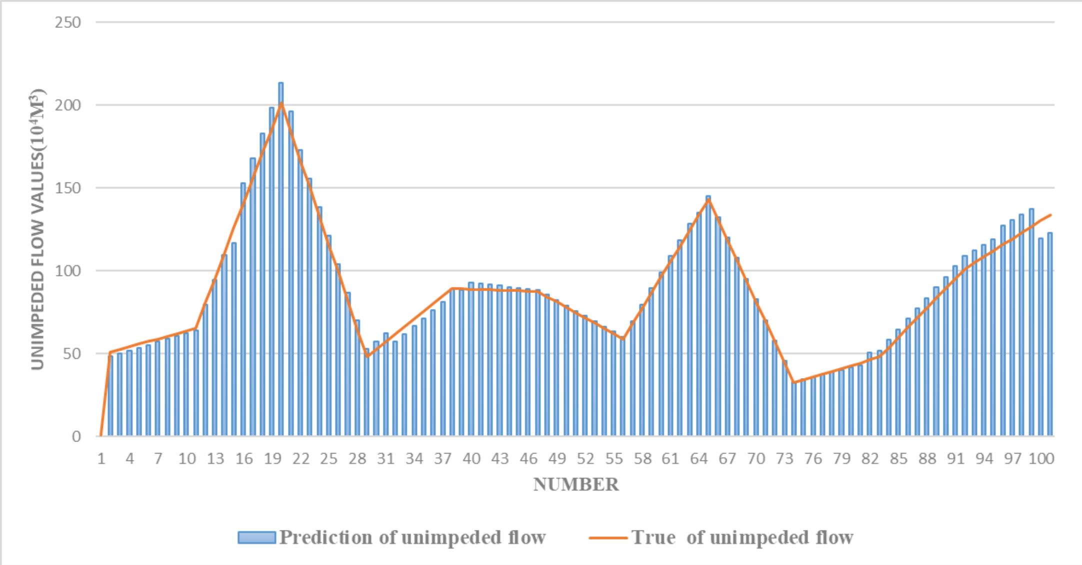

Figure 1 Gradient boosting model predicts open flow results

Random Forest (Random Forest) is a powerful integrated learning method for classification and regression tasks. It builds a powerful predictive model by constructing multiple decision trees and combining their predictions. Specifically, Random Forest generates multiple different decision trees by introducing two kinds of randomness: randomly sampled data and randomly selected features.

First, multiple subsample sets are randomly sampled from the original training dataset with putback, and each subsample set is the same size as the original dataset. Then, during the construction of each tree, for each node, a subset of features is randomly selected for splitting from the full set of features, instead of using the full set of features. These two types of randomness increase the variability between trees and enhance the generalization ability of the model. When predicting, the final prediction of the random forest is a combination of the predictions of all the trees. For the categorization task, each tree votes and selects the category with the most votes as the final prediction. For the regression task, the prediction results of all trees are averaged as the final prediction. This method can effectively deal with high-dimensional data and complex nonlinear relationships, and improve the prediction performance of the overall model.

Establish a gradient lifting prediction model based on the Random Forest algorithm, select the relevant parameters in the production capacity test well data as the characteristic variables, and the open flow rate as the target variable for prediction, and obtain the model R (correlation coefficient) of 0.86, with a high accuracy rate. (See Figure 2)

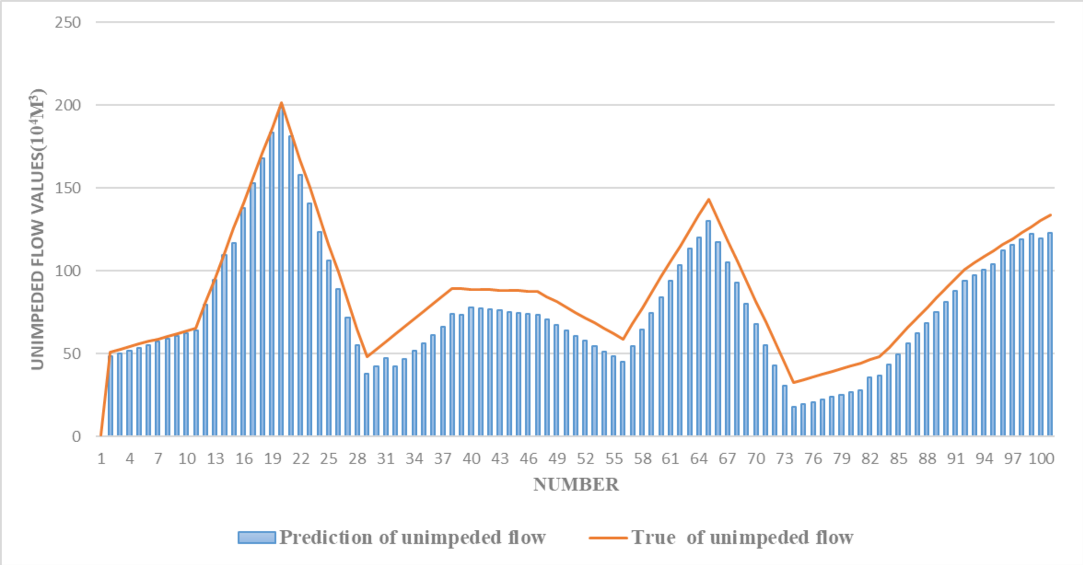

Figure 2 Random Forest Model Predicts Open Flow Results

By the model automatically selected 30% of data for testing, the experimental results from the 2 regression models and the comparison between the test samples and the actual values are shown below:

Table 2 Results of two regression prediction models

regression prediction model	Evaluation indicators		
	R2	MSE	MAE
XG-boost	0.92229	0.0453	0.3398
random forest	0.86178	0.0514	0.4436

In a comprehensive assessment of the provided predictive models, we first note that the Random Forest regression model performs reasonably well on the R^2 metric (0.86) but slightly less well on the MSE and MAE metrics (0.05 and 0.44) . Despite its better generalization ability and less susceptibility to overfitting, the Random Forest model is still limited in its predictive performance. The random forest model is better for large-scale datasets, but in some cases, the prediction performance may be poor due to insufficient number of trees or other improperly set hyperparameters.

The gradient boosted regression model (XG-Boost), on the other hand, overcomes the shortcomings of the above models and thus achieves better prediction performance. This is because the gradient boosting model forms a strong classifier by integrating multiple weak classifiers, which is able to better capture the complex relationships between the data and improve the prediction accuracy Therefore, based on its excellent performance and powerful features, the gradient boosting regression model performs the most prominently on this dataset, and is the best choice and the XG-boost regression model has a similar R^2 value of the mean square error and mean absolute error are small, and the fitting results are closer to the real value, which indicates that the prediction accuracy of the open flow can reach 92%, and also proves that the XG-boost regression model has a good prediction ability.

4. CONCLUSIONS AND OUTLOOK

In this study, based on the historical and geological data of the unhindered flow in each block of the target gas reservoir, the carbonate unhindered flow prediction model based on the data from the capacity test wells of 8 wells and 13 wells obtained was successfully established by using the K-nearest neighbor method for data processing. The gradient boosting and random forest algorithm were utilized to predict the unhindered flow rate, and the correlation coefficient exceeded 0.92, showing the high accuracy and reliability of the prediction model.

The results show that the predicted unhindered flow using gradient boosting and random forest algorithm can significantly simplify the traditional complex mathematical calculation methods. Through this method, not only can the potential of gas reservoirs be quickly and accurately assessed, but also provide strong support for the capacity evaluation of gas reservoirs. This research result provides a scientific basis for the formulation of corresponding safety production measures and helps to improve the efficiency and safety of gas reservoir management and development.

Future research will be devoted to expanding and diversifying the dataset, optimizing and integrating more advanced algorithms, conducting in-depth feature engineering and data processing, developing real-time prediction and automation systems, and promoting their application to more gas reservoirs and oil fields. Through these measures, we aim to further improve the accuracy and application value of the open flow prediction model and provide stronger technical support for the efficient and safe development of gas reservoirs.

REFERENCES

[1] Ellmerer M, Schaupp L, Trajanoski Z. Continuous measurement of subcutaneous lactate concentration during exercise by combining open-flow microperfusion and thin-film lactate sensors[J], 1998.

[2] Hu F, Hao Q, Bao K. A Survey on Software-Defined Network and OpenFlow: From Concept to Implementation[J], 2014.

[3] Stoessell RK, Ward WC, Ford BH. Water chemistry and CaCO3 dissolution in the saline part of an open-flow mixing zone, coastal Yucatan Peninsula, Mexico[J], 1989.

[4] Xie X, Pasquali M. A new, convenient way of imposing open-flow boundary conditions in two- and three-dimensional viscoelastic flows[J], 2004.

[5] G, P, Rigby. Stabilized needle electrode system for in vivo glucose monitoring based on open flow microperfusion.[J], 1996.