Prediction of Comprehensive Water Cut in Periodic Waterflood Reservoir Based on Variable Weight Combination Model

Quyang LIU, Xiaodong LIU, Dong Hao, Qiang XUE *
No.7 oil production plant of PetroChina Changqing Oilfield Company, Xi’an, 710018, China
*Corresponding author e-mail: liuhuif_cq@petrochina.com.cn

Abstract: The comprehensive water cut of periodic waterflood reservoirs is random and volatile, and it is a non-stationary time series. In order to solve the problems of poor prediction results of conventional methods or heavy workload and long time-consuming in reservoir numerical simulation, an EMD- LSTM neural network model and a smoothing spline regression (Smoothing Spline) variable weight combined prediction model are proposed. This method introduces empirical mode decomposition (EMD) to process non-stationary and non-linear data, and combines the advantages of machine learning and curve regression to improve the prediction accuracy of the model. The variable weight combination model is used to predict the water content of the periodic water injection reservoir. The results show that the prediction accuracy of the variable weight combination model is significantly higher than that of the single model. The variable weight combination model can reduce the prediction error and can effectively and accurately predict the periodic water injection oil. The comprehensive water content of the reservoir can guide the adjustment of water injection parameters for periodic water injection reservoirs.

1. Introduction
Reservoir numerical simulation is the most common method for comprehensive prediction of oil fields, but it relies on accurate geological modeling. It takes a long time to fit and has a large workload. In order to overcome the shortcomings of numerical simulation, scholars have realized the dynamic prediction of oil well production by means of machine learning methods such as BP neural network [1-2], support vector machine [3] and long-term memory neural network (LSTM) [4-5]. The integrated water cut of the reservoir is a non-stationary data, so the Empirical Mode Decomposition (EMD) method is introduced to process non-stationary and non-linear data to further improve the prediction accuracy of the LSTM neural network model. Smoothing spline regression is a local modeling method, which is a piecewise polynomial connected according to a certain degree of smoothness. It considers two aspects of the goodness of fit and the smoothness of the regression curve, and has certain advantages. Although the single model has its own advantages in forecasting, it is also susceptible to the influence of various factors in the forecasting process, which makes the forecast results at some moments unsatisfactory. Therefore, this paper selects two types on the basis of previous studies. A single prediction model, a variable weight combination model based on the principle of the minimum absolute value of the error is established. It is verified that the prediction accuracy of the combined model is significantly higher than that of the single model, and the combined model can effectively
and accurately predict the comprehensive water cut of the periodic water injection reservoir, and can guide the adjustment of the water injection parameters of the periodic water injection reservoir.

2. Principle and method of prediction model

2.1 Basic theory of EMD decomposition

EMD method \[6\] was proposed by n.e.huang. It can decompose complex nonlinear signal into several IMF components and trend terms with different scales (frequencies), and the decomposed IMF components contain the eigenvalues of signals with different fluctuation frequencies in the original signal.

The decomposition steps of EMD method \[6\] are as follows:

1. Determine all maxima and minima of time series \(x(t)\);
2. The upper envelope maximum \(x_{max}(t)\) and lower envelope minimum \(x_{min}(t)\) of the original data were obtained through third-order spline function interpolation. Calculate its mean value: \(D(t) = \frac{x_{max}(t) + x_{min}(t)}{2}\);
3. \(E^1(t)\) is a new signal, that is, \(E^1(t) = x(t) - D(t)\). Determine whether \(E^1(t)\) meets the condition of the eigenmode function (IMF). If so, it is defined as \(E^1(t) = c^1(t) = IMF1(t)\); If not, \(E^1(t)\) will be used as a new signal and repeated 1-2 steps, iterating \(k\) times until \(E^k(t)\) meets the definition of eigenmode function. Then the first IMF component with the highest frequency will be extracted, namely \(E^k(t) = c^1(t) = IMF1(t)\).
4. Deduct \(IMF1(t)\) from the original signal \(x(t)\) to get a new signal \(R(t)\), that is, \(R(t) = x(t) - IMF1(t)\). Repeat the above steps 1-3 to get \(IMF2(t)\) with the next highest frequency.
5. Repeat the above steps until the residual value meets the requirement that \(\delta\) is less than 0.2, and the empirical value of \(\delta\) is generally 0.2-0.3.

\[
\sum \left| E^k - E^{k-1} \right|^2 \leq \delta
\]  

(1)

In this case, \(x(t)\) can be expressed as:

\[x(t) = \sum_{i=1}^{n} IMF_i(t) + R_n(t)\]  

(2)

\(R_n(t)\) is the trend term, and \(x(t)\) represents the original signal. After EMD decomposition, N IMF components with high to low frequency and one trend item were obtained.

2.2 Long short term memory neural network (LSTM)

The traditional neural network can not save and use the information of the previous time, and can not predict the time series data. Therefore, based on the traditional recurrent neural network, LSTM neural network is proposed \[7\]. LSTM neural network introduces the mechanism of memory unit and hidden layer state to control the information transmission between hidden layers. It improves the memory module of traditional recurrent neural network. There are three gates in the memory unit of LSTM neural network, which are input gate, forgetting gate and output gate. The input gate controls the state input of hidden layer; Forgetting gate can keep or choose information; The output gate determines the information that can be output. LSTM model can effectively solve the problem of long-time dependence and gradient explosion caused by long sequence. LSTM neural network can mine the change rule of data \[8-11\], and realize the accurate prediction of comprehensive water cut of cyclic water injection reservoir.

2.3 LSTM neural network based on EMD decomposition

Reservoir comprehensive water cut is a non-stationary series, which changes with time and various influencing factors. Therefore, when LSTM neural network is used to predict the data, the prediction result is not very ideal. In order to improve the prediction accuracy, EMD decomposition is introduced, which can reduce the influence of its non-stationarity on the prediction accuracy. Each component decomposed by the EMD method was taken as the input variable, and the comprehensive water cut value of the reservoir was taken as the only output variable. The predicted value of each IMF
component and trend term was reconstructed to get the final predicted value. The optimal prediction model of LSTM neural network based on EMD can be obtained by repeated simulation to obtain the parameters of LSTM neural network with the minimum relative error.

2.4 Smoothing Spline Regression
Smooth spline regression \cite{12} is actually a local modeling method, a piecewise polynomial connected according to a certain smoothness. The commonly used spline function is the cubic spline function, that is:

Suppose there are real numbers \( t_1, t_2, \ldots, t_n \) on the interval \([a, b]\) and satisfy \( a < t_1 < t_2 \cdots t_n < b \), \( f(x) \) is a function in the interval, and the following conditions are satisfied:
1) In the interval \( f(x) \) is a cubic polynomial
2) \( f(x) \) and its second derivative are continuous at \( t_i (i = 1, 2 \cdots n) \).

The expression of cubic smooth spline can be expressed as follows:

\[
 f(x) = d_i (x - t_i)^3 + c_i (x - t_i)^2 + d_i (x - t_i) + a_i \tag{3}
\]

Smooth spline regression fitting is to find a smooth function to minimize the residual sum of squares. Introducing a penalty function to minimize the sum of squares of penalty can make the error fitting curve smooth. The expression is:

\[
 \min \sum_{i=1}^{n} (y_i - f(x_i))^2 + \alpha \int_{a}^{b} (f''(x))^2 \, dx \tag{4}
\]

\( \sum_{i=1}^{n} (y_i - f(x_i))^2 \) is the loss function, which is used to fit the data; \( \alpha \int_{a}^{b} (f''(x))^2 \, dx \) is the penalty for the fluctuation of the function, and the second derivative represents the roughness of the function; \( \alpha \) is a smooth parameter, \( \alpha \) The larger the value, the smoother the function.

3. Establishment of variable weight combination model
In order to make full use of the advantages of emd-lstm model and smooth spline regression, a variable weight combination prediction model based on emd-lstm model and smooth spline regression is proposed for comprehensive water cut prediction of cyclic water injection reservoir. Let the prediction value of emd-lstm model at time \( t \) be \( E(t) \), and the prediction value of smooth spline regression be \( S(t) \), \( t = 1, 2 \cdots n \), and give the two models variable weight value respectively \( w_1 \) and \( w_2 \). The final variable weight combination forecasting model is as follows

\[
 H(t) = w_1 E(t) + w_2 S(t), \quad w_1 + w_2 = 1, \quad t = 1, 2 \cdots n \tag{5}
\]

4. Case Analysis
The comprehensive water cut data of a reservoir from January 2012 to December 2020 were selected for EMD decomposition to reduce the influence of its non-stationary on the prediction accuracy. Then, 90% of the data were used as the training data for the fitting model and 10% as the validation data. Finally, the trained model is used to predict the comprehensive water cut and guide the adjustment of reservoir water injection.

4.1 EMD-LSTM model prediction

4.1.1 EMD decomposition
The boundary continuation method is used to deal with the decomposition boundary problem, and the comprehensive water-bearing data is decomposed by EMD, and the \( \delta \) value is 0.2. The decomposition obtains 3 IMF components and 1 trend item `Res`. It can be seen from Figure 1 that IMF1 has the largest amplitude and the highest frequency, and the amplitude and frequency of the remaining components decrease or decrease sequentially. It shows that the decomposed time series reduces the randomness and volatility of the original data, and provides stable time series data for LSTM neural network prediction.
4.1.2 LSTM neural network prediction
Matlab2019 was applied to build the LSTM neural network model. The number of hidden layers of the neural network was randomly set as 1, the time series step size was 12 months, the number of neurons was 100, the number of training cycles was 100, the batch size was 4, and the initial learning rate was 0.005. The gradient threshold is 1; The solver is' ADAM '. ADAM is the most commonly used algorithm at present, which has fast convergence speed and good learning effect.

Through times of simulation training to get the optimal parameter combination: hidden layers to 2, the Numbers of hidden layer neurons respectively, 30, 50 and time step for 6 months, in which information to predict the future in the past six months 1 months of production, batch size is 3, which is updated once every three sample network parameters, training cycle 300 times.

The IMF component and trend term obtained by EMD decomposition were taken as the input variables, and the comprehensive water cut value of the reservoir was taken as the output variables, which were brought into the LSTM neural network prediction model that had been trained. Finally, the predicted results of each IMF component and trend item were reconstructed to form the final comprehensive predicted value of water cut, which is shown in Table (2).

4.2 Smooth spline regression
The smooth spline regression program was written by Matlab2019. After several parameter adjustment fitting (see Table 1), the value of smooth factor $\alpha$ in Equation (4) was determined, that is, when $\alpha=0.9$, the fitting effect was the best. The predicted value is shown in Table (2).

| Indicators      | Smooth factor |
|-----------------|---------------|
|                | 0.2 | 0.4 | 0.6 | 0.8 | 0.9 |
| SSE (variance)  | 23.29| 17.80| 14.07| 10.08| 6.93 |
| R (Fitting coefficient) | 0.992 | 0.994 | 0.995 | 0.997 | 0.998 |
| RMSE (Root mean square) | 0.579 | 0.531 | 0.499 | 0.465 | 0.432 |

![Fig.1 IMF components and trends](image_url)
From the prediction results of single model, single model shows different advantages and disadvantages at different times, and the results have certain errors. By giving each single model a variable weight, we can get a variable weight combination model with higher prediction accuracy than single model.

### Table 2 Comparison of predicted value and actual value of each model (partial value)

| Time    | Actual value (%) | Smooth spline regression Predicted value (%) | Relative error (%) | EMD-LSTM Predicted value (%) | Relative error (%) | Polynomial regression prediction value (%) | Relative error (%) |
|---------|-----------------|---------------------------------------------|--------------------|-------------------------------|--------------------|--------------------------------------------|--------------------|
| 201201  | 37.60           | 37.99                                       | 1.03               | 37.93                         | 0.87               | 38.62                                      | 2.72               |
| 201202  | 38.44           | 38.47                                       | 0.06               | 38.33                         | 0.30               | 38.29                                      | 0.41               |
| 201203  | 39.13           | 38.84                                       | 0.72               | 38.20                         | 2.37               | 38.21                                      | 2.33               |
| 201204  | 39.68           | 39.04                                       | 1.61               | 38.92                         | 1.90               | 38.34                                      | 3.38               |
| 201205  | 39.25           | 39.02                                       | 0.58               | 39.50                         | 0.63               | 38.60                                      | 1.67               |
| 201206  | 38.68           | 38.91                                       | 0.60               | 40.00                         | 3.41               | 38.95                                      | 0.69               |
| 201207  | 38.28           | 38.88                                       | 1.57               | 38.71                         | 1.12               | 39.36                                      | 2.82               |
| 201208  | 38.93           | 39.03                                       | 0.24               | 39.39                         | 1.18               | 39.79                                      | 2.19               |
| 201209  | 39.56           | 39.35                                       | 0.53               | 40.25                         | 1.75               | 40.22                                      | 1.67               |
| 201210  | 39.73           | 39.83                                       | 0.24               | 39.17                         | 1.41               | 40.64                                      | 2.29               |
| 201211  | 40.26           | 40.44                                       | 0.45               | 39.70                         | 1.38               | 41.03                                      | 1.92               |
| 201212  | 40.95           | 41.17                                       | 0.52               | 41.62                         | 1.62               | 41.39                                      | 1.07               |

### 4.3 variable weight combination model prediction

The key of combination forecasting model is to determine the optimal weight value, which includes the weight within the sample and the weight outside the sample. Predecessors have summarized a variety of methods to determine the weight value in the sample, such as the objective function of minimizing the sum of square error, relative error and absolute error to determine the best weight coefficient; Such as fuzzy variable weight combination forecasting method. The methods to determine the weight value outside the sample include average value method, error reciprocal method, etc.

1) In-sample weights

Based on the principle of minimizing the absolute value of the combination prediction error at the sample points, the weight value in the sample was calculated through Equation (6) on the basis of satisfying the weight coefficient itself. The variable weight value of each model is shown in Table 3.

\[
\min \sum_{t=1}^{n} |f(t) - \sum_{i=1}^{m} w_i(t)f_i(t)|
\]

\[
s.t. \sum_{i=1}^{m} w_i(t) = 1, \text{ and } w_i(t) \geq 0
\]

2) Out-of-sample weights

According to the above method, the weight in the sample can be obtained, namely the weight value in the past. The purpose of the variable weight combination prediction model is to predict the future, and the weight should be extrapolated. At present, the common average method is to average all the weight values in the sample to determine the weight values out of the sample. According to the weight data in Table 3, the weight of smooth spline regression with out-of-sample weight value is 0.68, and the weight of EMD-LSTM is 0.32, as calculated by Equation (7).

\[
w_i^{n+1} = \frac{w_1^n w_2^n + \cdots + w_m^n}{n-1}
\]
Table 3 Weight values of each model in the combined model (partial values)

| Time   | Smooth spline weight | EMD-LSTM weight |
|--------|----------------------|-----------------|
| 201201 | 0.00                 | 1.00            |
| 201202 | 0.79                 | 0.21            |
| 201203 | 1.00                 | 0.00            |
| 201204 | 1.00                 | 0.00            |
| 201205 | 0.52                 | 0.48            |
| 201206 | 1.00                 | 0.00            |
| 201207 | 0.00                 | 1.00            |
| 201208 | 1.00                 | 0.00            |
| 201209 | 0.77                 | 0.23            |
| 201210 | 0.85                 | 0.15            |
| 201211 | 0.76                 | 0.24            |

The predicted values of each model and the above determined weight values in and out of samples are put into Equation (5) for calculation, and the predicted values of the variable weight combination model can be obtained.

The variable weight combination model is used to predict the comprehensive water cut in the next 8 months. The predicted results are shown in Figure 2. It can be seen that the actual comprehensive water cut will show a downward trend in January 2021, and the predicted value will show the same downward trend as the actual value.

The model is used to predict the comprehensive water cut of each production well, and advance prediction and advance parameter adjustment are adopted, such as increasing the water injection amount during the injection period or extending the water injection time to improve the effect of periodic water injection and improve the oil recovery rate of corresponding production Wells.

![Figure 2 Comparison of actual and predicted comprehensive water cut](image)

4.4 Evaluation of indicators

Root mean square error (RMSE) and mean absolute percentage error (MAPE) were used to evaluate the prediction accuracy of each model. The smaller the prediction error is, the higher the prediction accuracy is. It can be seen from Table 4 that the prediction accuracy of the variable weight combination model is higher than that of the other two single models. The prediction accuracy of smooth spline regression is close to that of combination model, while the prediction accuracy of polynomial regression is much lower than that of other methods.
Table 4 Evaluation indexes of each model

| Prediction method                     | RMSE  | MAPE  |
|--------------------------------------|-------|-------|
| Smooth spline regression             | 1.72  | 0.06  |
| The EMD - LSTM prediction            | 2.06  | 0.09  |
| Variable weight combination model    | 1.16  | 0.04  |
| Polynomial regression (8 times)      | 14.34 | 0.39  |

5. Conclusion and recognition

1. The LSTM neural network prediction model based on EMD overcomes the randomness of the comprehensive water cut of the reservoir, and can decompose the comprehensive water cut into a number of IMF components and trend items for the stability treatment. According to the IMF components and trend items, the LSTM neural network prediction model can be made to improve the prediction accuracy of the LSTM neural network prediction model.

2. When single model EMD-LSTM model and smooth spline regression were used for prediction, all models had certain errors, while the variable weight combination model could realize the complementary advantages among all models. In this paper, a variable weight combination forecasting method based on the principle of minimum absolute error at sample points is proposed, which makes the predicted value of the variable weight combination model more consistent with the actual value.

3. The prediction accuracy of the variable weight combination model is significantly higher than that of the single model, and it can effectively and accurately predict the comprehensive water cut of the periodic water injection reservoir. Although the model lacks physical significance, it enriches the prediction method of comprehensive water cut, and can provide reliable basis for the adjustment of water injection parameters.

Symbolic Notes:
- \( x(t) \) —— time series of comprehensive water cut;
- \( x_1(t), x_2(t) \) —— maximum and minimum values of the envelope;
- \( M(t) \) —— mean value of envelope;
- \( E^1(t), E^k(t) \) —— new signal, \( k \) —— number of iterations;
- \( n \) —— the number of frequencies;
- \( R(t) \) —— high frequency component;
- \( \delta \) —— Critical value;
- \( R_n(t) \) —— trend term;
- \( f(t) \) —— the actual value;
- \( f_p(t) \) —— the predicted value;
- \( w_j(t) \) —— weight value;
- \( m \) —— number of single models;
- \( t \) —— time series;
- \( n \) —— sample number;
- \( w_i^{t+1} \) —— the weight out of sample.

Reference

[1] ZHOU Cailan, LIU Min. Application research on oil production forecasting based on BP neural network [J]. Journal of Wuhan University of Technology, 2009, 31 (3): 125-129.
[2] FAN Ling, ZHAO Mengmeng, YIN Cuan, et al. Analysis method of oilfield production performance based on BP neural network[J]. Fault-Block Oil & Gas Field, 2013, 20 (2): 204-206.
[3] SEEMANNND, WILLIAMSONM, HASANS. Improving reservoir management through big data technologies[R].SPE167482, 2013.
[4] LIU W, LIU W D, GU J. Predictive model for water absorption in sublayers using a machine learning method[J]. Journal of Petroleum Science and Engineering, 2019, 182: 106367.
[5] GERSFA, SCHMIDHUBERJ, CUMMINSF. Learning to forget: Continual prediction with LSTM [J]. Neural Computation, 1999, 12 (10): 2451-2471.
[6] Huang N E, Shen Z. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society London, 1998, A, 454: 903-995.
[7] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9 (8): 1 735-1 780.
[8] PANJA P, VELASCO R, PATHAK M, et al. Application of artificial intelligence to forecast
hydrocarbon production from shales [J]. Petroleum, 2018, 4 (1):75-89.

[9] WANG S, CHEN Z, CHEN S. Applicability of deep neural networks on production forecasting in Bakken shale reservoirs [J]. Journal of Petroleum Science and Engineering, 2019 (179): 112-125.

[10] HOU Chunhua. Prediction method of oil production in new wells based on long and short term memory neural network[J]. Oil and Gas Geology and Recovery, 2019, 26(3): 105-110

[11] GU Jianwei, ZHOU Mei, LI Zhitao, et al. Prediction method of oil well production based on data mining based on long-term and short-term memory network model[J]. Special Oil and Gas Reservoirs, 2019, 26(2): 77-81.

[12] Wang Fengxue. The Study on Smooth Spline Regressionand Its Application[J]. Statistics and Applications, 2019, 8(4):604-612.

[13] ZHANG Xiaolin, CHEN Jiajia, ZHANG Zhenhua. The Weight Determination in Variable Weight Combination Forecast Based on Compositional Data[J]. Journal of Shanxi University (Nat.Sci.ed), 2014, 37(2): 188-194. 94.