The Establishment and Analysis of Gas Yield Prediction Model

Xianyu Meng*a, Yanping Cuia, Xiaoyan Cal and Jie Gaoa

a Ordnance Non-commissioned Officer Academy of PLA, Wuhan, China
* Corresponding author: Xianyu Meng, master,849934279@qq.com

Abstract. Accurate prediction of production in gas production area is the basis of scientific management and science distribution in gas fields. In the process of gas field development, accurate yield prediction plays an important role in developing adjustment and deployment, and improving the efficiency of operation, which determines the scale of production investment and decision direction of gas field. Based on the analysis of the production law of gas field and the investigation of related data, this paper utilizes the principle of linearity distribution, and the prediction model of gas yield fitting is established based on the density function.

1. Introduction
In the long run, gas production is a process from increment to decreasing. In the early stage of mining, the main factors affecting production are mining capacity, and the greater the mining capacity, the higher the yield. However, with the continuous development of gas fields, the restriction effect of reserves will become more and more obvious; when it becomes the leading role, gas production will be reduced by year to zero. The annual gas production data of a gas field in 1957~1976 years are as follows. Based on this data, the yield prediction model of gas field will be established in this paper.

Table 1. Output of a gas field.

| Year | Production ($m^3$) |
|------|-------------------|
|      | 1957   | 1958  | 1979  | 1960  | 1961  | 1962  | 1963  |
|------|------|------|------|------|------|------|------|
| 1957 | 19    | 43    | 59    | 82    | 92    | 113   | 138   |
| 1964 | 148   | 151   | 157   | 158   | 155   | 137   | 109   |
| 1971 | 89    | 79    | 70    | 60    | 53    | 45    |

2. Description of Symbols and Assumptions of the Model
φ(x) B Distribution density function
ν Modification coefficient of model
Q(t) Gas field production of year t
N_p Accumulated production of gas fields
N_g Total recoverable reserves in gas fields


$C, \alpha, \beta$ Undetermined parameters of the model

Suppose that the 1957 year is the first year of mining and assume that the gas field production process is safe and accident free.

3. Establishment of the Model

3.1. Data Preprocessing

Assume that the gas field 1957~1976 annual gas production data (Table 1) are divided into two groups, one for the known value, for the estimation of model parameters, and the other as a reference for the predictive effect of the model. We consider that the early and medium-term gas production in the gas field is better, and the late gas production decreases year by year, because the fixed cost is constant, so the prediction of gas yield in the later period is more important; this paper takes the data of the first 16 years as the known value, and constructs the mathematical model to predict the gas yield in the latter four years. The results of the specific data processing are shown in Table 2 and Table 3.

| Year | 1957 | 1958 | 1959 | 1960 |
|------|------|------|------|------|
| Production | 19 | 43 | 59 | 82 |

| Year | 1961 | 1962 | 1963 | 1964 |
|------|------|------|------|------|
| Production | 92 | 113 | 138 | 148 |

| Year | 1965 | 1966 | 1967 | 1968 |
|------|------|------|------|------|
| Production | 151 | 157 | 158 | 155 |

| Year | 1969 | 1970 | 1971 | 1972 |
|------|------|------|------|------|
| Production | 137 | 109 | 89 | 79 |

Table 2. Known datasheets

Table 3. Data sheets to be tested

| Year | 1973 | 1974 | 1975 | 1976 |
|------|------|------|------|------|
| Production | 70 | 60 | 53 | 45 |

3.2. Model Establishment

The variation law of oil and gas production is gradually decreasing after the rapid growth reaches the peak, and the decreasing rate decreases. The distribution of this change law is similar to the distribution of $\Gamma$. This article presents a similar one which is a new distribution called $\mathcal{B}$ Distribution$^{[1]}$, and its density function is:

$$\varphi(x) = \frac{1}{B(\alpha, \beta)} \frac{x^{\alpha-1}}{(1+x)^{\alpha+\beta}}$$ (1)

In the formula, $\alpha > 0, \beta > 0$ are pending constants and $B(\alpha, \beta)$ is determined by the following formula:

$$B(\alpha, \beta) = \int_0^{+\infty} \frac{t^{\alpha-1}}{(1+t)^{\alpha+\beta}} \, dt$$ (2)

We introduce a model correction coefficient $\nu$, and we apply Formula(1) for the yield prediction of gas fields. Thus get:

$$Q(t) = \frac{\nu}{B(\alpha, \beta)} \frac{t^{\alpha-1}}{(1+t)^{\alpha+\beta}}$$ (3)
$Q(t)$ is the gas field production of year $t$ ($t$ start from 1). So the accumulated production of gas fields $N_p$ can be expressed as:

$$N_p(t) = \int_0^t Q(\tau)d\tau$$

(4)

We put Formula(3) into Formula(4), and we get:

$$N_p(t) = \nu \int_0^t \frac{1}{B(\alpha, \beta)} \frac{\tau^{\alpha-1}}{(1+\tau)^{\alpha+\beta}}d\tau$$

(5)

When $t \to +\infty$, $N_p$ will tend to be total recoverable reserves in gas fields $N_R$:

$$N_R = \nu \int_0^{+\infty} \frac{1}{B(\alpha, \beta)} \frac{\tau^{\alpha-1}}{(1+\tau)^{\alpha+\beta}}d\tau$$

(6)

Based on Formula(1), (2), we can get:

$$\int_0^{+\infty} \frac{1}{B(\alpha, \beta)} \frac{\tau^{\alpha-1}}{(1+\tau)^{\alpha+\beta}}d\tau = 1$$

So,

$$\nu = N_R$$

Based on Formula(3), we can get:

$$Q(t) = \frac{N_R}{B(\alpha, \beta)} \frac{t^{\alpha-1}}{(1+t)^{\alpha+\beta}}$$

(7)

We set

$$C = \frac{N_R}{B(\alpha, \beta)}$$

Based on Formula(7), we can get:

$$Q(t) = C \frac{t^{\alpha-1}}{(1+t)^{\alpha+\beta}}$$

(8)

Formula(8) is the yield prediction model of gas field, $C, \alpha, \beta$ are the pending parameters.

4. Model Solution
Based on the relevant data, we exploit the Formula (8) to carry out proper deformation treatment and turn it into linear relation, and then we use linear regression to solve the undetermined coefficient $C, \alpha, \beta$. 

Based on Formula (8), we can get:

$$Q(t+1) = C \cdot \frac{(t+1)^{a-1}}{(t+2)^{a+\beta}}$$  \hspace{1cm} (9)

We use Formula (9) divided by Formula (8):

$$\frac{Q(t+1)}{Q(t)} = \left(\frac{t+1}{t}\right)^{a-1} \cdot \frac{t+1}{t+2}$$  \hspace{1cm} (10)

We take logarithms on both sides:

$$\ln\left(\frac{Q(t+1)}{Q(t)}\right) = (\alpha - 1) \ln\left(\frac{t+1}{t}\right) + (\alpha + \beta) \ln\left(\frac{t+1}{t+2}\right)$$  \hspace{1cm} (11)

We set

$$Y = \ln\left(\frac{Q(t+1)}{Q(t)}\right)$$

$$X_1 = \ln\left(\frac{t+1}{t}\right), \quad X_2 = \ln\left(\frac{t+1}{t+2}\right)$$

$$A = \alpha - 1, \quad B = \alpha + \beta$$

Based on Formula (11), we can get:

$$Y = AX_1 + BX_2$$  \hspace{1cm} (12)

$A$ and $B$ can be obtained by binary linear regression, so as $\alpha, \beta$.

Based on Formula (10), we can get:

$$Q(t) \cdot (1 + t)^{\beta \cdot t} = C \left(\frac{t}{1+t}\right)^{a}$$

We take logarithms on both sides:

$$\ln[Q(t) \cdot (1 + t)^{\beta \cdot t}] = \ln C + \alpha \ln\left(\frac{t}{1+t}\right)$$  \hspace{1cm} (13)

We set

$$Y = \ln[Q(t) \cdot (1 + t)^{\beta \cdot t}]$$

$$A = \ln C$$

$$X = \ln\left(\frac{t}{1+t}\right)$$

We can get:

$$Y = A + \alpha X$$  \hspace{1cm} (14)

$A, \alpha$ can be obtained by linear regression. $\alpha$ can be obtained by linear regression based on Formula (12) or Formula (14). But
Formula (14) is related to $\beta$, $\beta$ is obtained by linear regression, so we should obtain $\alpha$ from Formula (12).

The values of the relevant parameters can be obtained as in Table 4 through program solution with MATLAB (Program code in the Appendix).

**Table 4.** The values of the relevant parameters

| $C$     | $\alpha$ | $\beta$ |
|---------|----------|---------|
| $2.9841 \times 10^{15}$ | 7.8881 | 98.761 |

So the final model of the problem is:

$$Q(t) = \frac{2.9841 \times t^{6.8881}}{(1+t)^{106.6491}}$$

We can predict the gas production when $t = 17, 18, 19, 20$. The result is in Table 5.

**Table 5.** Predictive results

| Year | 1973 | 1974 | 1975 | 1976 | 1980 |
|------|------|------|------|------|------|
| $t$  | 17   | 18   | 19   | 20   | 24   |
| True Value | 70  | 60   | 53   | 45   | ------ |
| Predicted Value | 77.78  | 65.07 | 54.1 | 44.79 | 20.72 |

The model established in this paper is based on the existing data linear simulation, can basically reflect the general law of the oil and gas field output.

**5. Conclusion and Model Optimization**

The model established in this paper is based on the existing data linear simulation. It can basically reflect the general law of the gas field production. The predicted value and the actual value of the comparison diagram is shown below on Figure 1.

![Figure 1. Comparison Plot](image_url)

From the above figure, we can find that the forecasting method based on single model has achieved good fitting effect in the period of obvious decline law. However, there is a large error in predicting the rising period of production, which is due to the complexity of improving natural gas production capacity, and the universality of the single model is not ideal. Therefore, multi-model groups and prediction methods can be used to improve the prediction accuracy. The realization is more complex,
and the specific combination methods are various, limited to the length, this paper only makes a preliminary analysis.

Here we introduce another single model prediction method, Weng's simulation method\cite{2}, for multi-model group and prediction. The model expression is:

$$Q_{o} = at^{b} e^{-ct}$$

The parameters of the improved model system can be obtained by using the method mentioned above. The combination mode adopts piecewise linear superposition and the improved model is:

$$Q(t) = C \cdot \frac{t^{\alpha-1}}{(1+t)^{\alpha+\beta}} u(t-t_0) + at^{b} e^{-ct} u(t_0-t)$$

The function $u(t)$ is a unit step function. Its characteristic is that the function value is 1 when $t > 0$ and the function value is 0 when $t < 0$.

6. Appendix

clear
Q=[19 43 59 82 92 113 138 148 151 158 155 137 109 89 79 70 60 53 45 0 0 0 0 ];
for t=1:19
X1(t)=log((t+1)/t);
X2(t)=log((t+1)/(t+2));
Y(t)=log(Q(t+1)/Q(t));
end
X=[X1',X2'];
Y=Y';
[result,bint,r,rint,s]=regress(Y,X);
a=result(1)+1;
b=result(2)-a;
for t=1:20
R(t)=log(Q(t)*(1+t)^b*t);
T(t)=log(t/(t+1));
end
p=polyfit(T,R,1);
C=exp(p(2))+2984100000000000
b=7.8881
a=12.00+11*b
for t=1:24
product(t)=C*t^(12.00+11*b+b)/(1+t)^(12.00+11*b+b);
end

7. References
[1] Xu Chuansheng. Beta distribution discussion on the properties and application. Journal of Linyi Normal University , 2001 year - volumes 4 period ;
[2] Deng Yong, Du Zhimin, Chen Yanni . Application of neural network optimization combination Prediction model. Journal of Applied Mathematics in colleges and universities . 2008,23(1): 1-6.