The GM(1,1) grey model for the Total Suspended Solids Emission in Waste Water caused by shale gas Production

Xiaofeng Li [a], Zhibing Liu [b]

School of Science, Southwest Petroleum University, Sichuan 610500, China

[a]yearn1993@163.com, [b]50825372@qq.com

Abstract

Grey theory is one of the most common methods for solving uncertain problems using limited data, due to its high performance in time series prediction. However, the inappropriate background value and initial value are the main factor affecting prediction accuracy of the Grey Model GM(1,1).

Keywords

Grey theory; time series prediction; Total Suspended Solids Emission.

1. Introduction

The grey theory was put forward by professor Deng Julong of huazhong university of science and technology. Grey forecasting method can contain both known information, and predict the unknown or uncertain information system, is to change within a certain amount of azimuth and time dependent in the process of grey prediction. Due to the existence of noise, the historical data sequence tends to be disordered, and the gray theory is to generate an orderly regeneration sequence through the accumulation of this irregular original data sequence. Gray modeling is to use this regeneration sequence to establish the differential equation model.

2. The GM (1,1) prediction model

The $GM(1,1)$ model is the simplest and most commonly used prediction model in the grey prediction model, and it is the basis of other grey prediction models.

Definition of gray accumulation.: Set the original sequence $X^{(0)} = X^{(0)}(1), X^{(0)}(2), \cdots, X^{(0)}(r)$, After a cumulative sequence is $X^{(1)} = X^{(1)}(1), X^{(1)}(2), \cdots, X^{(1)}(r)$,

$$X^{(1)}(k) = \sum_{j=1}^{k} X^{(0)}(j), k = 1, 2, \cdots, r \tag{1}$$

In the above sequence, $X^{(1)}$ is called the first order accumulation sequence of $X^{(0)}$, called $1 - AGO$.

Definition of gray accumulation: Set the original sequence. $1 - AGO$ is $X^{(1)} = X^{(1)}(1), X^{(1)}(2), \cdots, X^{(1)}(r)$, then

$$X^{(0)}(k) = X^{(1)}(k) - X^{(1)}(k-1) \tag{2}$$

is called the first order of sequence $X^{(1)}$.

Consider the original sequence $X^{(0)} = X^{(0)}(1), X^{(0)}(2), \cdots, X^{(0)}(r)$, The $1 - AGO$ $X^{(1)} = X^{(1)}(1), X^{(1)}(2), \cdots, X^{(1)}(r)$ Satisfy the definition of cumulative generation, then $X^{(0)}(k)$ The ordinary differential equation of albinism is.

$$\frac{dX^{(1)}(k)}{dt} + aX^{(1)}(k) = b \tag{3}$$

Discrete the type

$$\Delta^{(1)}(X^{(1)}(k)) + aZ^{(1)}(X^{(1)}(k)) = b \tag{4}$$

$\Delta^{(1)}(X^{(1)}(k))$ is the cumulative sequence of $X^{(1)}(k)$ at time k.

$Z^{(1)}(X^{(1)}(k))$ is $\frac{dX^{(1)}(k)}{dt}$ the background value at time k.

As

$$\Delta^{(1)}(X^{(1)}(k)) = X^{(1)}(k) - X^{(1)}(k-1) = X^{(0)}(k) \tag{5}$$

$$Z^{(1)}(X^{(1)}(k)) = \frac{1}{2}(X^{(1)}(k) + X^{(1)}(k-1)) \tag{6}$$

Substituting (5)(6) into (4)

$$X^{(0)}(k) = b - \frac{1}{2}[X^{(1)}(k) + X^{(1)}(k-1)] = a \cdot [-\frac{1}{2}(X^{(1)}(k) + X^{(1)}(k-1))] + b \tag{7}$$

Unfold the above equation is

$$\begin{bmatrix} X^{(0)}(2) \\ X^{(0)}(3) \\ \vdots \\ X^{(0)}(n) \end{bmatrix} = \begin{bmatrix} -\frac{1}{2}(X^{(1)}(1) + X^{(1)}(2)) & 1 \\ -\frac{1}{2}(X^{(1)}(2) + X^{(1)}(3)) & 1 \\ \vdots & \vdots \\ -\frac{1}{2}(X^{(1)}(k) + X^{(1)}(k-1)) & 1 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} \tag{8}$$

Order $\Phi = \begin{bmatrix} a \\ b \end{bmatrix}, Y = \begin{bmatrix} X^{(0)}(2) \\ X^{(0)}(3) \\ \vdots \\ X^{(0)}(n) \end{bmatrix}, B = \begin{bmatrix} -\frac{1}{2}(X^{(1)}(1) + X^{(1)}(2)) & 1 \\ -\frac{1}{2}(X^{(1)}(2) + X^{(1)}(3)) & 1 \\ \vdots & \vdots \\ -\frac{1}{2}(X^{(1)}(k) + X^{(1)}(k-1)) & 1 \end{bmatrix}$

So the (4) equation is simplified to.

$$Y = B\Phi \tag{9}$$

The parameter vector Φ can be solved by least square estimation. The equation is rewritten as

$$Y = B\hat{\Phi} + E \tag{10}$$

To make the

$$\min\left\|Y - B\Phi\right\| = \min(Y - B\hat{\Phi})(Y - B\Phi) \tag{11}$$

Use the matrix derivative formula $\hat{\Phi} = (B^T B)^{-1} \cdot B^T Y = \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix}$ (12)

substitute the parameters for the first order linear differential equation. $\frac{dX^{(1)}(k)}{dt} + \hat{a}X^{(1)}(k) = \hat{b}$, and solve it。It can be obtained by using constant variation method.

$$\hat{X}^{(1)}(k) = ce^{-a(k-1)} + \frac{\hat{b}}{\hat{a}} \tag{13}$$

Order $k = 1$， $\hat{X}^{(1)}(1) = c + \frac{\hat{b}}{\hat{a}}$， solved $c = \hat{X}^{(1)}(1) - \frac{\hat{b}}{\hat{a}}$， tnen

$$\hat{X}^{(1)}(k)=(X^{(0)}(1)-\frac{\hat{b}}{\hat{a}})\cdot e^{-\hat{a}(k-1)}+\frac{\hat{b}}{\hat{a}} \quad (14)$$

Restore the original data

$$X^{(0)}(k)=X^{(1)}(k)-X^{(1)}(k-1)=(1-e^{\hat{a}})\left[X^{(1)}(1)-\frac{\hat{b}}{\hat{a}}\right]\cdot e^{-\hat{a}(k-1)}. \quad (15)$$

3. Accuracy of model

$e(i)$ is the residual of the original sequence $x^0(i)$ and the predicted data column $\hat{x}^0(i)$, namely $e(i)=x^0(i)-\hat{x}^0(i)$. The overall accuracy of the model adopts the three most commonly used methods. The mean absolute error, mean square error and mean absolute percentage error are evaluated.

The mean absolute error (MAE) is the mean of absolute error; MAE can better reflect the actual situation of the predicted error.

$$MAE=\frac{\sum_{i=1}^{n}\left|e_i\right|}{n}=\frac{\sum_{i=1}^{n}\left|y_i-\hat{y}_i\right|}{n} \quad (16)$$

The mean square error(MSE) is the expected value of the difference between the parameter estimation and the actual value of the parameter. MSE can evaluate the degree of change of data, and the smaller the value of MSE, the more accurate it is to describe the experimental data.

$$MSE=\frac{\sum_{i=1}^{n}(e_i)^2}{n}=\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n} \quad (17)$$

Mean absolute percentage error (MAPE) is mainly a review of different models to the same set of data, which can be used to measure a model to predict the results of good or bad, look at the size of the MAPE is meaningless, because the MAPE is a relative, not absolute value.

$$MAPE=\frac{1}{n}\sum_{i=1}^{n}\frac{\left|y_i-\hat{y}_i\right|}{y_i} \quad (18)$$

4. The simulation

This paper, taking a southern domestic shale gas field exploitation of shale gas in 10 consecutive months back drainage contains the total number of total suspended solids (TSS) based on the data of original data GM(1, 1) prediction model for fitting, and then according to the result of fitting compared with existing data, and demonstrate the short-term forecasting precision of the model. At the same time, it will provide valuable decision-making information for environmental protection, and help decision-makers to make rational development of shale gas plan.

Table1 TSS content of shale gas in a southern shale gas field for 10 consecutive months.

	1	2	3	4	5	6	7	8	9	10
TSS（mg/L）	618	664	776	889	968	1163	1324	1466	1635	1698

The GM(1,1) prediction model USES the traditional least-squares principle to estimate the model parameters. The first five sets of data were modeled and the remaining five sets of data were predicted and validated. The prediction model of TSS is,

TSS： $\hat{X}^{(1)}(k+1)=5811.1e^{0.1k}-5193.1$.

Using the solutions of the above models, the results of 1-ago sequence can be calculated, and the simulation and predictive value of the original sequence can be obtained by using the reduction and reduction.

Table2 The comparison between the predicted value of TSS and the actual value.

季度	实际浓度	GM(1,1)模型	
		预测值	相对误差
1	618	618	0
2	664	710.8	0.070
3	776	797.8	0.028
4	889	895.4	0.007
5	968	1004.9	0.038
6	1163	1127.8	-0.030
7	1324	1265.8	-0.044
8	1466	1420.6	-0.031
9	1635	1594.4	-0.025
10	1698	1789.4	0.036
MAE			0.056
MSE			0.035
MAPE			0.023

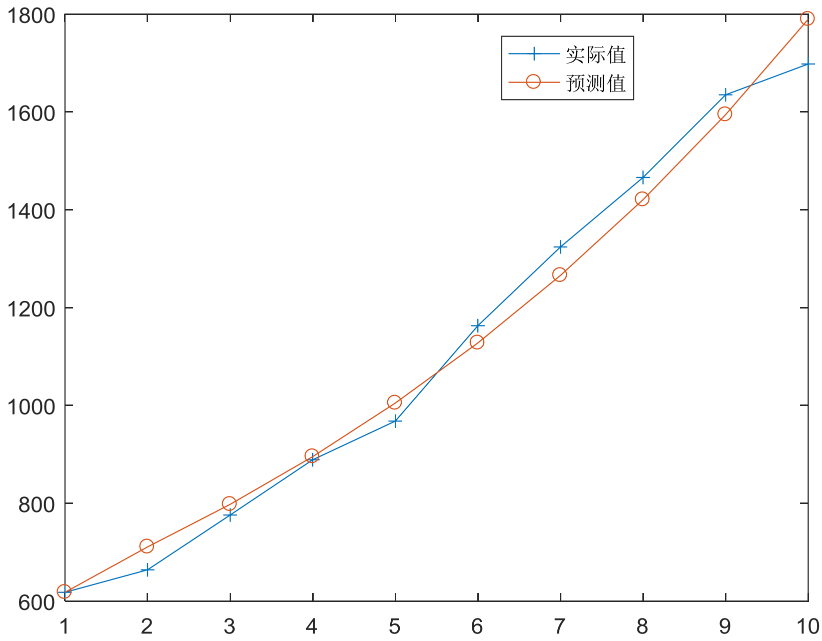

Figure 1 The comparison between the predicted value of TSS and the actual value

5. Conclusion

In this paper, the GM(1,1) model for the TSS in industrial waste water was constructed. In addition, MAE, MSE and MAPE were as evaluation indexs of prediction error to compare the model prediction effect. The GM(1,1) model could meet the requirement of high precision prediction.

References

[1] Jyh-Horng W, Huang Y F, Chen Z L, et al. Comments on the minimum number of data required

for GM(1,1) modeling[J]. 1999.
[2] Deng J L. Control problems of grey systems[J]. Systems & Control Letters, 1982, 1(5):288-294.
[3] Wang Q, Fang Z, Guo Y, et al. Grey Systems: Theory and Application[J]. Grey Systems Theory & Application, 2004, 4883(4):44-45.
[4] Olmstead S M, Muehlenbachs L A, Shih J S, et al. Shale gas development impacts on surface water quality in Pennsylvania[J]. Proceedings of the National Academy of Sciences of the United States of America, 2013, 110(13):4962.
[5] Wang Z X. Nonlinear Grey Prediction Model with Convolution Integral NGMC (1, n) and Its Application to the Forecasting of China's Industrial SO2 Emissions[J]. Journal of Applied Mathematics, 2014, 2014:174-178.
[6] Xie N M, Liu S F. Discrete grey forecasting model and its optimization[J]. Applied Mathematical Modelling, 2009, 33(2):1173-1186.
[7] Kumar U, Jain V K. Time series models (Grey-Markov, Grey Model with rolling mechanism and singular spectrum analysis) to forecast energy consumption in India[J]. Energy, 2010, 35(4):1709-1716.
[8] Pao H T, Fu H C, Tseng C L. Forecasting of CO 2, emissions, energy consumption and economic growth in China using an improved grey model[J]. Energy, 2012, 40(1):400-409.