Research on Time Series of Gas Trend Based on R Language

Peng Wang[1], Xuewen Li[2,*] and Xintan Chang[1]
[1]Xi'an University of Science and Technology, Shanxi, China
[2]Xi'an ChangKe Electronic Science and Technology Co., Ltd, China
*Corresponding author

Abstract—**Use the history monitoring data of the gas of Shaanxi Huangling No.2 coal mine, the time series of monitoring data are decomposed and smoothed by using R language. The abnormal and missing values in the raw data are analyzed and three methods of time series smoothing (Simple exponential smoothing method, Holt exponential smoothing method and Holt-Winters exponential smoothing method) are used to predict the variation law of gas concentration. The actual value and the predicted value are compared to verify the effectiveness of the forecasting method, and the conclusion can make practical significance for the safe production of the mine.**

Keywords-R language; time series; gas trend warning; simple exponential smoothing method; holt exponential smoothing method; holt-winters exponential smoothing method

I. INTRODUCTION

Although the state of coal mine safety production work increasingly attention, intelligent unmanned face security technology is still partial to the conventional, mainly in: (1)The detection data of the coal mining face is difficult to enter the security information system into the real-time analysis and information sharing; the safety inspector of the coal mining face, such as the inspection staff, lacks the access channel for the monitoring data and the analysis conclusion of the mine safety monitoring system in the whole mine, the face area and even the working area. ②Ventilation software features is single, inadequate application support, and impact on the whole mine and face ventilation optimization and control. (3) Lack of basic tracking analysis of the coal mining face, it is difficult to objectively evaluate the actual effect of various safety measures and ensure the gradual optimization of safety measures, which is not conducive to the gradual improvement of the level of work safety.

Data prediction is based on the analysis of historical data, forecasting the characteristics of new data or data trends in the future and so on. The methods used for prediction are based on classification prediction of neighborhood, classification regression tree prediction, neural network prediction, support vector machine prediction, time series prediction and so on.

Time series prediction refers to the analysis of time series variables, based on the time series presented the development process, dynamic direction and trends, the use of mathematical methods to establish a mathematical model, its analogy or extension, to predict the next Time or the next period of time may reach the level.

In the study of gas concentration prediction, some experts and scholars have studied the gas concentration time series prediction method, and achieved good application effect.

These methods include autoregressive moving average model[1][2], chaos and fractal[3][4][5], neural network[6][7][8], support vector machine[9][10], data mining and so on.

The R language is a locale designed for statistical calculations and graphical displays. It is an implementation of the S language developed by Bell Labs, providing a series of statistical and graphical display tools [11]. The R language is also an open, shared software platform for statistical analysis and data mining. On this platform, software researchers can continually expand their capabilities to develop software for specific applications.

Therefore, the R language is applied to the coal mine intelligent comprehensive mining comprehensive analysis of the trend of gas, the effective use of safety monitoring to monitor real-time collection of large gas data, the collected gas data for in-depth analysis and regular extraction, to achieve large data environment The intelligent and systematic management of mining surface safety is the inevitable trend of modern mine development.

II. DECOMPOSITION OF TIME SERIES

The decomposition of time series is mainly to decompose the regularity of the sequence into different components, including the long-term trend part, the seasonal variation part and the irregular fluctuation part.

Among them, the long-term trend is the time series in a long period of time shown by the general situation; seasonal changes that affect the seasonal changes in the natural and seasonal changes occur; irregular fluctuations due to accidental, unpredictable sequence changes.

The model decomposition has a basic addition model and a multiplication model. The basic form of the additive model is: $x_t=m_t+s_t+y_t$. The basic form of the multiplication model is: $x_t=m_t*s_t*y_t$. The time series values of the various components can be obtained by decomposition.

In the R language, the decompose function and the STL function can be used for the decomposition of time series. Taking the upper corner of 205 coal mining face of Huangling No.2 coal mine as an example:

```
> w<-read.table("yuanshi.prn",header=T)
```

> which(is.na(w))

[1] 180

> w[180,]=0.1645

> w1=w[1:288,]

> q<-ts(w1,start=1,frequency=12)
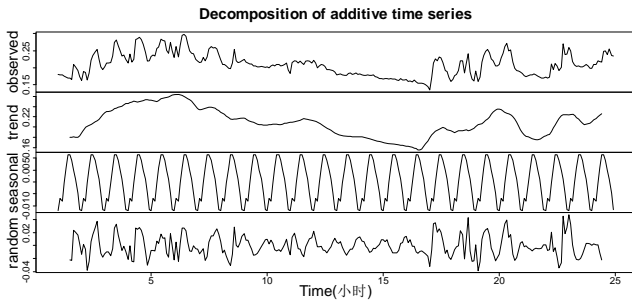
> plot(decompose(q),xlab="Time(hour)")

FIGURE I. TIME SERIES DECOMPOSITION OF GAS CONCENTRATION IN SHANGYUJIAO

From the overall trend of the time series, the sequence has no obvious upward trend, so the sequence belongs to the additive model. Where trend represents the trend, seasonal represents the season, and random represents irregular fluctuations. From Figure 1, the overall trend is from rise to decline, and then to rise, the fluctuation range is between 0.15 to 0.25; Compared with the trend fluctuation and random fluctuation, the seasonal factors have little influence and can be neglected because the fluctuation range is between -0.01 and 0.005, although the seasonal variation trend is seasonal. In the random trend, the fluctuation is between -0.04 ~ 0.04, the first 10 hours and 16 to 24 hours of these two fluctuations are more intense, the middle of the period of time fluctuations are more regular. The decompose (q) command allows you to see the specific size of each section.

III. TIME SERIES SMOOTHING METHOD

Time series prediction of the two basic methods, namely, smoothing and fitting method, there are moving average method, exponential smoothing method, auto-regressive moving average method. Moving average method takes n as a simple average, for example, it is assumed that the influence of the sequence observations on the t period is the same regardless of the distance of time; The exponential smoothing rule is to consider the distance of time, and the weight from far to near observations decreases exponentially with the increase of time interval. The auto-regressive moving average rule is to analyze the sequence structure by considering the sequence dependency of the observed values.

A. *Simple Exponential Smoothing Method*

This method assumes that the time series data does not contain seasonal variation and there is no trend of the system, and the basic formula is the sequence smoothing value of the period t (the long term trend estimation period) is equal to the sequence observed value of period t and the sequence smoothing value of t-1 period. That is:

$$\tilde{x}_t = a x_t + (1-a)\tilde{x}_{t-1}, 0 < a < 1$$

Where a is the smoothing coefficient, $0 < a < 1$, after repeated iterations,

$$\begin{aligned}\tilde{x}_t &= a x_t + (1-a)\tilde{x}_{t-1}\\ &= a x_t + (1-a)(a x_{t-1} + (1-a)\tilde{x}_{t-2})\\ &\cdots\\ &= a x_t + a(1-a)x_{t-1} + \cdots a(1-a)^{t-1}x_1 + a(1-a)^t \tilde{x}_0\end{aligned}$$

The weight a (1-a) [i] decreases exponentially with i.

In the R language, the smoothing coefficient a is estimated according to HoltWinters, taking the time series w1 as an example:

> cf.hw1=HoltWinters(w1,beta=F,gamma=F)

> cf.hw1

Smoothing parameters:

alpha: 0.9999572

Coefficients:

[,1]

a 0.234

The results show that the slope coefficient is estimated to be 0. 9999572, and the smoothing value of the upper corner at the end of August 3 is 0.234.

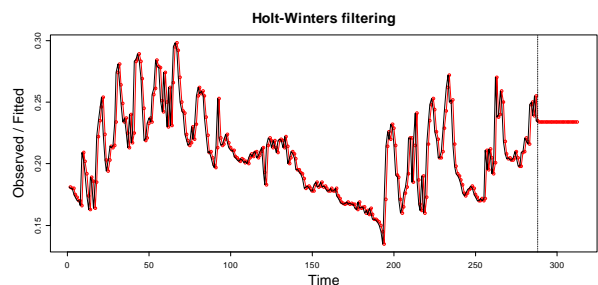For 24 time points in the first two hours of August 4 to predict, and draw the forecast curve Figure II:

FIGURE II. SIMPLE EXPONENTIAL SMOOTHING AND PREDICTION

Calculate the mean absolute error (mae), root mean square error (rmse), and mean absolute percentage error (mape) of the predicted results:

>w2=w[289:312,]# actual value

> mae=mean(abs(yuce-w2))# average absolute error

> mae

[1] 0.03504169

> rmse=mean((yuce-w2)^2)# root mean square error

> rmse

[1] 0.001482126

> mape=mean(abs(yuce-w2)/w2)# average absolute percentage error

> mape

B. Holt Exponential Smoothing Method

The Holt exponential smoothing method considers the change of the system trend on the basis of simple exponential smoothing, and assumes that there is a relatively fixed linear trend in the time series. Assuming that a_t is the horizontal value of x_t (ie, the long-term trend estimate), b_t is the slope change value of t to t + 1 period, then the sequence prediction value of t + 1 period is:

$$\hat{x}_{t-1} = a_t + b_t$$

Among them,

$$a_t = \alpha x_t + (1-\alpha)(a_{t-1} + b_{t-1})$$
$$b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1}$$

α, β are greater than 0, less than 1.

> cf.hw2=HoltWinters(w1,gamma=F)

Smoothing parameters:

  alpha: 1

  beta : 0.001458547

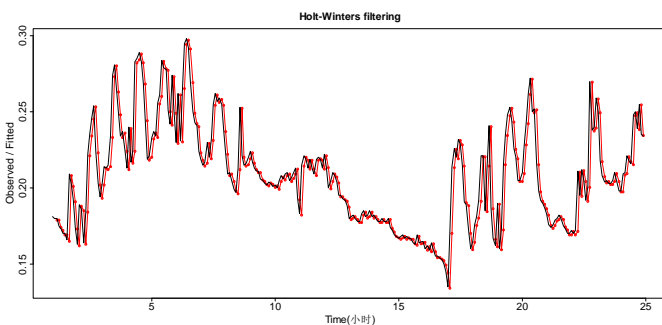 gamma: FALSE

> plot(cf.hw2,type="o")

FIGURE III. HOLT EXPONENTIAL SMOOTHING FITTING

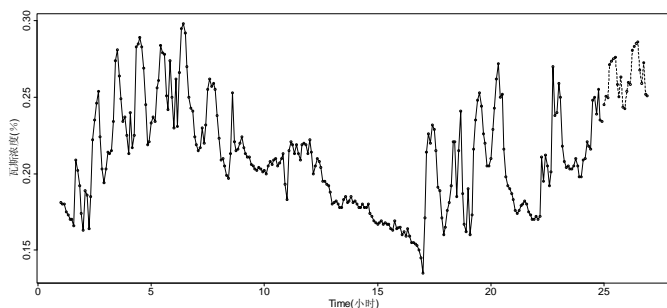>ts.plot(w1,yuce2,type="o",lty=1:2,xlab="Time(hour)",ylab="Gas concentration(％)")

FIGURE IV. HOLT EXPONENTIAL SMOOTHING PREDICTION

The predicted values for the 24 time points are as follows: 0.2450825, 0.2506909, 0.2495842, 0.2711427, 0.2734068, 0.2753326, 0.2762909, 0.2584911, 0.2502290, 0.2633309, 0.2435037, 0.2424353, 0.2538908, 0.2596739, 0.2585009, 0.2808008, 0.2831168, 0.2850821, 0.2860455, 0.2675906, 0.2590119, 0.2725466, 0.2520008, 0.2508705.

Calculate the prediction error, the mean absolute error is 0.05173379, the mean square error is 0.003211549, and the average absolute percentage error is 0.2549301.

C. Holt-Winters Exponential Smoothing Method

Holt-Winters exponential smoothing method is based on Holt exponential smoothing method to consider the impact of seasonal changes can be used to add the model or multiplication model.

Among them, the additive model is as follows:

$$\begin{cases} a_t = \alpha(x_t - s_{t-p}) + (1-\alpha)(a_{t-1} + b_{t-1}) \\ b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1} \\ s_t = \gamma(x_t - a_t) + (1-\gamma)s_{t-p} \end{cases}$$

Where, a_t, b_t, s_t represent the sequence level value, slope value and seasonal factors value of the t period, respectively; α, β, γ are the smoothing coefficients, the values are between 0 and 1, and p is the periodicity of the seasonal variation.

> w1<-ts(w1,frequency=12)

> cf.hw3=HoltWinters(w1,seasonal="additive")

> cf.hw3

Smoothing parameters:

  alpha: 0.7870092

  beta : 0.007560822

  gamma: 1

> win.graph(width=4.875,height=2.5,pointsize=8)

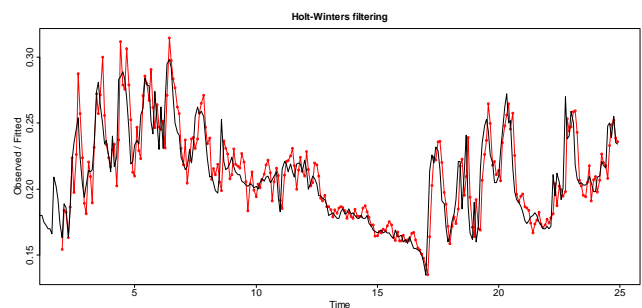> plot(cf.hw3,type="o")

FIGURE V. HOLT-WINTERS EXPONENTIAL SMOOTHING FITTING

> yuce3<-predict(cf.hw3,n.ahead=24)
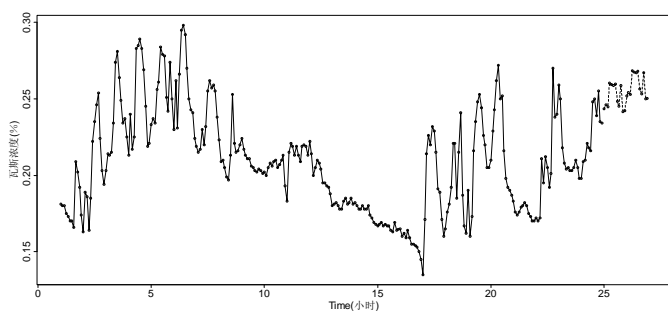
> ts.plot(w1,yuce3,type="o",lty=1:2)

FIGURE VI. HOLT-WINTERS EXPONENTIAL SMOOTHING PREDICTION

The predicted values for the 24 time points are as follows: 0.2436066, 0.2460268, 0.2446355, 0.2602498, 0.2593460, 0.2588089, 0.2597647,0.2483552, 0.2449609, 0.2587713, 0.2416613, 0.2421877, 0.2517943, 0.2542145, 0.2528232, 0.2684375, 0.2675337, 0.2669966, 0.2679524, 0.2565429, 0.2531486, 0.2669590, 0.2498490, 0.2503754

Calculate the prediction error, the mean absolute error is 0.0457605, the mean square error is 0.002451924, and the average absolute percentage error is 0.2248315.

The prediction errors of above three methods are as follows:

TABLE I. THE PREDICTION ERROR OF THREE METHODS

Evaluation index Prediction method	Three Methods		
	mean absolute error	*mean square error*	*average absolute percentage error*
Simple exponential smoothing method	0.0350	0.0015	16.56％
Holt exponential smoothing method	0.0517	0.0032	25.49％
Holt-Winters exponential smoothing method	0.0458	0.0025	22.48％

On the whole, the exponential smoothing does not take into account the seasonal variation and the trend of the system; Quadratic exponential smoothing takes into account the trend of the system; three smooth considerations also take into account the effects of system trends and seasonal changes. The first exponential smoothing prediction error is the smallest, followed by the third exponential smoothing, the second smoothing index prediction error is greatest

IV. CONCLUSION

Based on the prediction results of gas concentration time series forecasting and optimization, it is helpful to improve the accuracy of time series prediction of gas concentration in fully mechanized coal mining face. The gas concentration monitoring and prediction of the gas monitoring data can not only realize the dynamic monitoring of the gas data, but also provide the basis of the decision-making for the managers to obtain the information of the hidden dangers, and make scientific decisions in time to eliminate the hidden dangers of accidents and avoid the personnel casualties.

REFERENCES

[1] Li Suwang. Modeling of Time Series and It's Application in Forecast of Coal Mine Gas[D]. Taiyuan University of Technology,2007

[2] Yang Shengyuan et al. Application of ARMA model in predicting mine gas concentration[J]. ,China Coal, 2008, 34(6): 14~16.

[3] CHENG Jian,BAI Jing-yi,QIAN Jian-sheng,LI Shi-yin. Short-Term Forecasting Method of Coalmine Gas Concentration Based on Chaotic Time Series[J]. Journal of China University of Mining & Technology , 2008, 37(2): 231-234.

[4] LIU Ting-ting,SHI Jiu-gen,HAN Jiang-hong. Prediction of gas concentration chaotic time series based on support vector machines[J].Journal of Hefei University of Technology(Natural Science) , 2009, 32(8): 1150~1153.

[5] DU Xiao-kun. The Application of Improved Chaotic Forecasting Method in Monitoring Gas[J].Microcomputer Information , 2010 , 26(11):190~191.

[6] Zhu Yu. Constructive Neural Network Time Series Prediction in Gas Applications[D].Taiyuan University of Technology,2010

[7] ZHAO Jinxian,YU Guanghua. Model of chaotic sequence and RBF neural network on gas concentration forecast[J]. Journal of Heilongjiang Institute of Science and Technology . 2010,2(2): 131~134.

[8] Zhang Y, Duan X. Based on Cluster Cover Algorithm in Coal Mine Gas Forecast Application[C]// International Conference on Intelligent Computation Technology and Automation. IEEE, 2010:854-857.

[9] Shao Y, Zhang D. Mine Forecast Based on Genetic Annealing Neural Network[C]// International Conference on Information Technology and Computer Science. IEEE Computer Society, 2009:241-244.

[10] ZHAO, Xiao-hu, WANG, et al. On-line least squares support vector machine algorithm in gas prediction[J]. International Journal of Mining Science and Technology, 2009, 19(2):194-198.

[11] Yang Xia. Application of R Language in Large Data Processing[J]. Science & Technology Information 2013, (23): 19-20.