

## Modeling Method for Predicting the Shelf Life of Spare Parts under Automatic Modeling Process

Zhiwei Li\*, Qi Gao, Xinliang Jiang, Shenyang Liu and Tielu Gao

*Mechanical Engineering College, Xinhua District, Shijiazhuang, 050003, China  
arhqs@126.com*

### **Abstract**

*A modeling method which can predict the shelf life of various types of spare parts in a relatively short time is put forward in this article. At present, it is difficult to solve the problem of mass modeling because the shelf life prediction models for different kinds of spare parts are of great diversification. In this paper, the best fitting nonlinear variables are selected by Gram-Schmidt regression method, and the detailed steps of automatic modeling process are given, which have advantages of strong robustness and are easy in programming. Especially, it can eliminate the influence of multicollinearity among alternative models effectively. By using natural rubber heating elongation data, an example is taken to demonstrate the process of automatic modeling. The nonlinear regression models selected by automatic modeling process are consistent in Dakin equation, and the predict values of natural rubber shelf life are included in the storage period given by manufacturing plant.*

**Keywords:** Spare Parts, Shelf Life, Prediction Model, Gram-Schmidt

### **1. Introduction**

The shelf life of spare parts is defined as the duration that the spare parts can meet the qualified requirements in specific storage condition after they leave the factory. Accurately predicting the shelf life of spare parts is of great significance for the scheduling of inventory, making reasonable turnaround strategy and achieving the goal of precise support. Due to the rapid research and development of ordnance equipments, the varieties of spare parts related to these equipments are growing fast. More than tens of thousands of kinds of spare parts are required to complete the repair mission and ensure that the equipment has high availability. Quite a large number of spare parts have no specific shelf life. For the spare parts that have shelf life, on the one hand, the actual storage environment may not meet the requirements specified by manufacturer; on the other hand, the shelf life has different conservative degree. Therefore, we need to revise the shelf life.

The traditional shelf life prediction methods such as time serials analysis, grey prediction method and distribution fitting method can achieve the prediction goal, but when confront with a large amount of spare parts, these methods spend too long time and cannot form the general prediction process. The traditional multiple linear regression model [1] and Fisher discriminate model is the common prediction model in engineering practice, but the precision and effectiveness cannot guarantee when there exists high correlation between variables. Classification and Regression Tree (CART) [2, 3] is an algorithm to select the most influence on outcome variables, however, the algorithm must scan all the original variables for each variable selecting process, and the modeling efficiency is low. With the development of information technology, the field of machine learning becomes a research hotspot. Among these methods, Gram-Schmidt orthogonalization process has obvious advantages in the information decomposition and screening. Chen S [4] has performed data sorting method of linear model based on Gram-

Schmidt process, which reduces the variable selection amount from  $2^p$  to  $p$ . This modeling method has many applications, including RBF (Radial Basis Function) network model, artificial neural network, wavelet network and matching pursuit algorithm [5-8]. Stoppiglia H [9] presents termination criteria of variables screening by defining the probe feature. Other related works [10-19] also proved that Gram-Schmidt process is an effective approach to analyze multi-variables modeling.

In order to establish generous models in a limited time, automatic modeling process is used to solve this problem. The prediction data are gathered in periodic inspection. We present the automatic nonlinear regression modeling process for the shelf life of spare parts, and the final model is validated by experimental data of nature rubber.

## 2. Key Steps of Automatic Modeling Process

Automatic modeling process must address the following questions: one is how to select the key influence factors, the second is how to overcome the defects of multicollinearity in the regression model, the third is how to choose the most suitable model in a number of alternative model.

For the first problem, we can find the answer from the national standards and relevant literature. So far, more than 400 standards have been found, which involve various factors affecting the shelf life of spare parts. For example, the main storage failure factor of metal spare parts is electrochemical corrosion; and main factor of rubber spare parts is oxidation, and main factor of optical spare parts is mildew corrosion. Induction, the main influencing factors include temperature, humidity, solar radiation, microbial growth, etc. For the second problem, Gram-Schmidt orthogonalization method can be applied to transform the vector space into standard orthogonal subspace to overcome the effect of multicollinearity. For the third problem, we restrict the scope of alternative model in linear model and nonlinear model which can be converted into linear, yet even had the limits, the number of all possible combination of alternative models is quite large, so the exhaustive method is obviously can't work. And gram-Schmidt regression method provides a new way to overcome this difficult.

### 2.1 Similar Linear Regression Model

The nonlinear regression models which can convert to linear model through variable substitution are named similar linear regression model. For example, the growth model is

$$y = \exp(\alpha + \beta x)$$

Take natural logarithm on both sides of the equation, then

$$\ln y = \alpha + \beta x$$

$$\text{Let } y^* = \ln y$$

We can get the similar regression model :

$$y^* = \alpha + \beta x$$

In the study of the shelf life of spare parts, the often used similar linear models are shown in Table 1.

**Table 1. Often used Similar Linear Model in Predicting the Shelf Life of Spare Parts**

Name	Models	Name	Models
Logarithm	$y = \alpha + \beta \ln x$	Logistic	$y = 1 / (1 / \mu + \alpha \beta^x)$
Inverse	$y = \alpha + \beta / x$	Growth	$y = \exp(\alpha + \beta x)$
Polynomial	$y = \alpha + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \dots + \beta_p x^p$	Exponent	$y = \alpha \exp(-\alpha x)$

Power	$y = \alpha x^\beta$	Weibull	$y = \alpha \beta^\alpha x^{\alpha-1} \exp(-\beta x)^\alpha$
Compound	$y = \alpha \beta^x$	Normal	$y = (1/\sqrt{2\pi}\beta) \exp(-(x-\alpha)^2 / (2\beta^2))$
S	$y = \exp(\alpha + \beta/x)$		

### 2.2 Gram-Schmidt Orthogonalization Method

The Gram-Schmidt orthogonalization method is proposed by professor Gram and Schmidt. If a set of vectors in inner product space can form a subspace, this set of vectors called a basis for this subspace. Any basis in this subspace can transform into a orthogonal basis through Gram-Schmidt process, and it will become standard orthogonal basis after unitization. Assuming that space  $V^k$  is a subspace of dimension  $k$  in space  $V^p$ , its standard orthogonal basis is  $\{\eta_1, \eta_2, \dots, \eta_k\}$ . If  $v \in V^p$ , and  $v$  don't exist in  $V^k$ , an orthogonal basis of dimension  $k+1$  can be structured. The projection of vector  $v$  in space  $V^k$  can be written as

$$proj_{V^k} v = \sum_{i=1}^k (proj_{\eta_i} v) \cdot \eta_i$$

The difference between vector  $v$  and vector  $proj_{V^k} v$  can be represented by  $\beta$ , i.e.

$$\beta = v - \sum_{i=1}^k (proj_{\eta_i} v) \cdot \eta_i \tag{1}$$

The vector  $\beta$  and space  $V^k$  are orthogonal to each other.  $\eta_{k+1}$  denotes the unitization of vector  $\beta$ .

$$\eta_{k+1} = \frac{\beta}{\|\beta\|} = \frac{\beta}{\sqrt{\langle \beta, \beta \rangle}} \tag{2}$$

Then,  $\{\eta_1, \eta_2, \dots, \eta_{k+1}\}$  is standard orthogonal basis of subspace  $L(v, \eta_1, \eta_2, \dots, \eta_k)$  which is a extended space from  $V^k$  and vector  $v$ .

### 2.3 Gram-Schmidt Regression Method

Gram-Schmidt regression method is the theoretical basis of automatic modeling process [19] of shelf life. The main idea of Gram-Schmidt regression method is summarized as follows.

$y$  is a dependent variable,  $x_1, x_2, \dots, x_p$  are independent variables. Assuming that the predicted data  $y$  which indicates the degree of storage failure and environment stress variables  $x_j$  collected from Annual technology inspection are centralized, simple regression models of  $y$  with the initial G-S variables  $x_j (j=1, 2, \dots, p)$  are established respectively. The correlation of models should be applied a correlation test, such as t-test, r-test and F-test. The variable whose test statistics of G-S variable corresponding to maximum should enter the high correlated variable sets. Other variables ought to minus the projection component of all the selected variables which belong to .high correlated variable sets, and the results are called the next generation G-S variables. The same steps are repeated until G-S variables cannot meet the requirement of correlation test.

After the above modeling process, those alternative models which have significant influence on storage reliability or cumulative failure rates of spare parts are chosen to form high correlated variable sets, those repeat impact factors between alternative models are cut down through Gram-Schmidt process, and those alternative models which have litter impact on storage failure are deleted for which cannot pass the regression

significance check.. This modeling process is of benefit to overcome the serious defect of multicollinearity.

### 3. Automatic Modeling Process of Nonlinear Regression Model

At first, there is a need to build the alternative function item sets for predicting the shelf life of spare parts. Through the literature, the storage reliability and the cumulate failure rates of spare parts can be expressed by the nonlinear models listed in Table 1. The first step is to select all the possible function items from Table 1 according to the engineering experience and the physical and chemical properties of spare parts.

The second, all the nonlinear function items selected in step 1 should be linearized to become similar linear models, which will be screened by Gram-Schmidt process. In this way, we can choose the most appropriate models out of the similar linear model sets, and these models which have no significant effect on shelf life will be eliminated.

Finally, we can calculate the coefficients of regression model, and check the prediction accuracy of the regression model. Then, coefficient model is modified through the inverse Gram-Schmidt transform.

Taking an automatic regression modeling process describing the relationship between the aging properties of a certain type of nature rubber spare parts and environment stresses as an example, a complete automatic modeling process is given here. For the calculation of rubber shelf life, we measured the rubber breaking elongation under different storage temperature and storage period. The rubber is identified as storage failure when elongation reaches the critical value of 50%.  $y$  is referred to as the logarithm of the percentage elongation.  $T$ ,  $t$  represents aging temperature and aging period respectively. The regression modeling calculation process is as follows.

(1) Informed by working experience and literature about shelf life of rubber, the main factors of rubber breaking elongation are temperature and storage period. Inspired by some models related to rubber shelf life, such as Dakin life equation, P-t model and P-T-t model, the alternative models are chosen as following.

$$u_1 = t^{0.5}, u_2 = t, u_3 = 1/T, u_4 = t^2, u_5 = T.$$

(2) All the models  $u_1, u_2, \dots, u_5$  should be centralized to ensure that mean value of each variable is 0.

(3) Let  $z_j^{(1)} = u_j (j=1,2,\dots,5)$ . The simple regression models of  $y$  with  $z_j^{(1)}$  are established respectively. All these simple regression models must carry on the statistical tests, and those models whose absolute test statistic value is the maximum value should enter the .high correlated variable sets. Suppose that  $u_1$  is chosen, then  $z_1 = u_1$ .

(4) Let

$$z_j^{(2)} = u_j - \frac{u_j^T z_1}{z_1^T z_1} z_1, j = 2, 3, \dots, 5 \quad (3)$$

The bivariate regression models of  $y$  with  $z_1$  and  $z_j^{(2)} (j=2,3,\dots,5)$  should be established, and those models whose absolute test statistic value is the maximum value should enter the high correlated variable sets. We assume that the second model chosen is  $u_2$ , then  $z_2 = z_2^{(2)}$ .

The same steps are repeated until the rest of alternative models can not meet the requirement of correlation test. The high correlated model will be formed.

$$\hat{y} = \alpha_1 z_1 + \alpha_2 z_2 + \dots + \alpha_m z_m \quad (4)$$

The regression model of  $y$  with  $u_1, u_2, \dots, u_5$  can be derived by inverse Gram-Schmidt transform, and we can get the final model after variable substitution.

#### 4. Case Study

In order to illustrate the effectiveness of this method, we used the elongation data of natural rubber given by Liu Zhenxi [11] to validate this method.

**Table 2. Elongation Data of Nature Rubber**

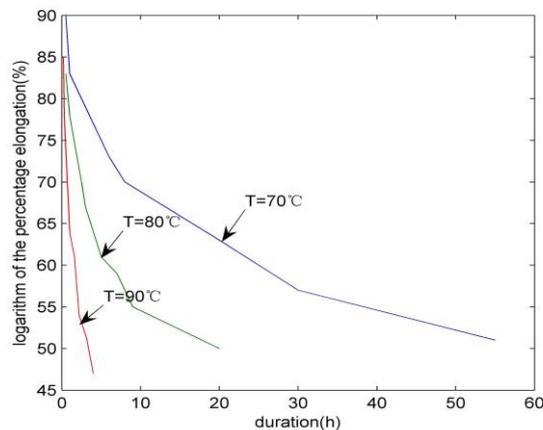
$\{t^{0.5}, t, 1/T, t^2, T\}=70^\circ\text{C}$		$\{t^{0.5}, t, 1/T, t^2, T\}=80^\circ\text{C}$		$\{t^{0.5}, t, 1/T, t^2, T\}=90^\circ\text{C}$	
aging time (h)	percentage (%)	aging time (h)	percentage (%)	aging time (h)	percentage (%)
0.5	90	0.5	83	0.2	85
1	83	1	78	0.3	78
6	73	2.5	70	0.6	72
8	70	3	67	1	64
20	63	5	61	1.6	61
25	60	7	59	2.2	54
30	57	9	55	3.2	51
55	51	20	50	4	47

The alternative model set is  $\{t^{0.5}, t, 1/T, t^2, T\}$ . The objective is selecting the appropriate models from alternative model set to describe the relationship between logarithm of the percentage elongation  $y$ , temperature  $T$  and duration  $t$  using nonlinear automatic modeling process. The elongation data of nature rubber are shown in Table 2. Characteristics of the experimental data under different temperature are shown in Figure 1.

The modeling process is shown as following.

Let  $u_1 = t^{0.5}, u_2 = t, u_3 = 1/T, u_4 = t^2, u_5 = T$ . Then the sample data  $\{u_1, u_2, u_3, u_4, u_5, y\}$  ought to be centralized, and the results are also denotes as  $\{u_1, u_2, u_3, u_4, u_5, y\}$ .

At first, let  $z_j^{(1)} = u_j (j=1, 2, \dots, 5)$ , the simple regression models of  $y$  with  $z_j^{(1)}$  are established respectively, and whether the correlation coefficient  $r$  is larger than 0.6 is the standard to select a model as high correlated model. The values of correlation coefficient  $r$  of simple regression models are listed in Table 3.



**Figure 1. Characteristics of Elongation Data**

In Table 3,  $z_1$  is the only variable whose correlation coefficient value is larger than 0.6, that is  $z_1 = z_1^{(1)} = u_1$ .

**Table 3. The Values of Correlation Coefficient  $r$  of Simple Regression Models**

variable	$z_1^{(1)}$	$z_2^{(1)}$	$z_3^{(1)}$	$z_4^{(1)}$	$z_5^{(1)}$
$r$	0.6021	0.4908	0.1560	0.3750	0.1550

The second, the Gram-Schmidt transformation are conducted on  $u_2, u_3, u_4, u_5$  and  $z_1$ , and bivariate regression models should established with  $y$  and the transformation results together with  $z_1$ . The values of correlation coefficient  $r$  of bivariate regression models are listed in Table 4.

**Table 4. The Values of Correlation Coefficient  $r$  of Bivariate Regression Models**

variable	$z_2^{(2)}$	$z_3^{(2)}$	$z_4^{(2)}$	$z_5^{(2)}$
$r$	0.2969	0.6159	0.1914	0.6108

In Table 4,  $z_2$  is the variable whose correlation coefficient value is larger than 0.6 and the absolute value is the maximum value. Then

$$z_2 = z_3^{(2)} = u_3 - \frac{u_3^T z_1}{z_1^T z_1} z_1. \quad (5)$$

The next, the Gram-Schmidt transformation is conducted on  $u_2, u_4, u_5$  and  $z_1, z_2$ , and ternary regression models should established with  $y$  and the transformation results together with  $z_1, z_2$ . The values of correlation coefficient  $r$  of ternary regression models are listed in Table 5.

In Table 5, all the variable values do not reach 0.6, the modeling process is over. The regression model established by high correlated model is

$$\hat{y} = -0.0613z_1 + 2311.5z_2 \quad (6)$$

**Table 5. The Values of Correlation Coefficient  $r$  of Ternary Regression Models**

variable	$z_2^{(3)}$	$z_4^{(3)}$	$z_5^{(3)}$
$r$	0.0014	0.0610	0.5912

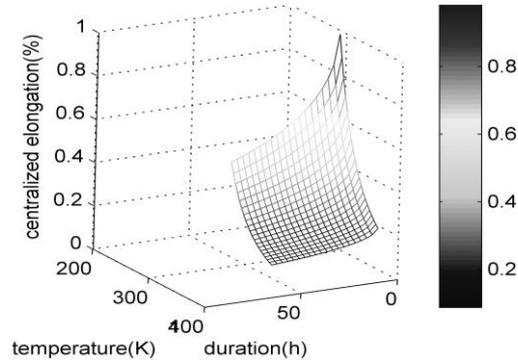
At last, through inverse Gram-Schmidt transform, the regression model of  $y$  and alternative models will be derived.

$$\hat{y} = -0.1061u_1 + 2311.5u_3 \quad (7)$$

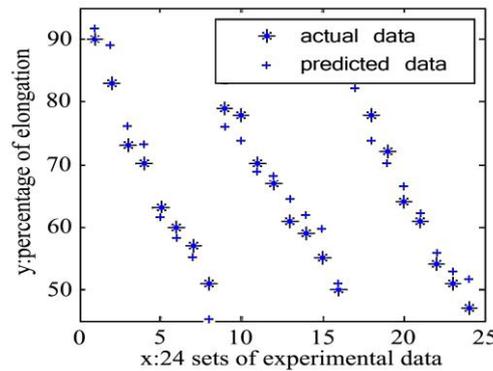
Considering the error of centralization, we can obtain

$$\hat{y} = \exp(-0.1061\bar{t}^{0.5} + 2311.5/\bar{T}) \quad (8)$$

where,  $\bar{y}$ ,  $\bar{t}$  and  $1/\bar{T}$  denote the centralized variable. The trend chart of equation (8) is shown in Figure 2.



**Figure 2. Trend Chart of Regression Equation**



**Figure 3. Control Chart of Actual Data and Predicted Data**

Obviously, the shelf life of nature rubber is 8.7159 in the condition that the elongation critical value is 50% and the temperature is 25 °C.

The relationship between the predicted data of the regression model and the actual data are shown in Figure 3.

The nonlinear variables  $t^{0.5}$  and  $1/T$  are chosen by automatic modeling process, and the final model is a good match for Dakin life equation. The predicted shelf life is 8.7159 which belongs to the scope of period 5-10 years given by manufacturing plant.

## 5. Conclusions

The Automatic modeling process using Gram-Schmidt process for nonlinear regression models can select the proper function items from alternative models sets to solve the prediction problem of the shelf life of spare parts, and the prediction errors are within an acceptable range. However, the function items derived from predictive process are standardized in that the data need to be standardized in the step of initialization. Thus, the forecast errors caused by the standardization of environment stress values are difficult to eliminate. We will improve this automatic modeling process continuously.

## References

- [1] J. Aitchison, "The statistical analysis of compositional data (with discussion)", *Journal of the Royal Statistical Society*, vol. 44, (1982), pp. 139-177.
- [2] P. L. Priddy, G. P. Bly and Jackson, "Evaluation of Precast Panels for Airfield Pavement Repair", Phase 2: Results of Accelerated Pavement Testing. ADA585856, September, (2013).
- [3] B. E. Barrowes, D. George and F. Shubitidze, "Portable Electromagnetic Induction Sensor with Integrated Positioning", ADA578929, (2012) April.
- [4] S. Chen, S. A. Billing and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification", *International Journal of Control*, vol. 50, (1989) May, pp.1873-1898.
- [5] S. Wold, H. Martens and H. Wold, "The multivariate calibration problem in chemistry solved by the PLS method", *Proc. conf. matrix pencils, Lectures notes in mathematics*. Heidelberg: Springer-Verlag, (1983).
- [6] S. Chen, C. F. N. Cowan and P. M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks", *IEEE Transaction on Neural Networks*, vol. 2, (1991) February, pp. 302-309.
- [7] P. Vincent and Y. Bengio, "Kernel matching pursuit", *Machine Learning*, vol. 48, (2002) September, pp. 165-187.
- [8] F. Soulie-Fogelman, "L'industrialisation des Analyses-Besoins, Outils & Application", *Conference de Knowledge Extraction Engines*, vol. 11, (2006).
- [9] H. Stoppiglia, G. Dreyfus, R. Dubois and Y. Oussar, "Ranking a random feature for variable and feature selection", *The Journal of Machine Learning Research*, vol. 3, no. 7-8, (2003) November, pp. 1399-1414.
- [10] M. Ye, G. Saporta and H. W. Wang, "Nonlinear Regression Automatic Modeling Process", *Systems Engineering*, (2009) July, pp. 81-84.
- [11] Z. X. Liu, "Research on the Application of Mathematical Statistics in the Forecast Period of Rubber Stored Term", *Logistics and purchasing*, (2008) December, pp. 77-79.
- [12] M. Ye, H. W. Wang and L. H. Wang, "Application of improved hierarchical clustering method to classification of curves", In: Xia G, Osaki H. *Proceedings of the Ninth International Conference on Industrial Management*, (2008) September, pp. 325-330, Osaka, Japan.
- [13] H. Wang, Q. Liu, H. M. Mok, L. Fu and W. M. Tse, "A hyperspherical-transformation forecasting method with compositional data", *European Journal of Operation Research*, vol. 179, (2006), pp. 459-468.
- [14] J. Wang, H. W. Wang and S. Gilbert, "Discussion on importance of variable selection in PLS1 modeling", *5th International Symposium on PLS and Related Methods*, Oslo, vol. 9, (2007).
- [15] H. W. Wang, B. Xia and J. Meng, "Fast algorithm of Gram-Schmidt regression method", *Beijing University of Aeronautics and Astronautics*, vol. 39, (2013) September, pp. 1259-1262.
- [16] Hirokawa, Mariko, Kuroki and Yoshimitsu, "A fast implementation of PCA-L1 using gram-schmidt orthogonalization", *IEICE Transactions on Information and Systems*, vol. 96, (2013) March, pp. 559-561.
- [17] B. James and Wilson, "Optimal algorithms of Gram-Schmidt type", *Linear Algebra and Its Applications*, vol. 438, (2013) December, pp. 4573-4583.
- [18] Q. H. Liu and F. D. Zhang, "Incomplete hyperbolic Gram-Schmidt-based preconditioners for the solution of large indefinite least squares problems", *Journal of Computational and Applied Mathematics*, vol. 250, (2013), pp. 210-216.
- [19] B. Yagoubi, "Linear filtering and modelling based on Gram-Schmidt orthogonalization concept", *WSEAS Transactions on Signal Processing*, vol. 10, (2014) January, pp. 75-85.