Partial Least Squares Regression for Predicting the Speed of Electromagnetic Fuze Plate

Wenyun Wang*, Huarong Li*
Kunming Shipbuilding Equipment Research and Test Center, Kunming, China
*Corresponding author e-mail: wangwenyun3035@163.com, *4429528030@qq.com

Abstract. The partial least squares regression method was used to establish the regression equation of the speed of the fuze plate after cross validation test. Combined with the measured data, the speed of the fuze plate was analyzed. The results show that the partial least squares regression model effectively overcomes the multiple correlation problem of independent variables, has good interpretability and high fitting accuracy, and is an effective method for analyzing the characteristics of the fuze plate's speed parameters.

1. Introduction
In order to test the electromagnetic fuze function of a certain type of underwater, a massive iron plate is accelerated to a high speed in a short distance on land [1]. In this test, the newly developed test equipment in our range uses a hydraulic method to propel the fuze plate at high speed. Speed control and prediction is difficult, but speed control accuracy is directly related to system test accuracy.

The speed is controlled by open loop through preset input parameters, including hydraulic driving force P, oil circuit solenoid valve control opening voltage V, and hydraulic oil temperature T. As there are serious multiple correlations between these influencing factors, it is very difficult to analyze the speed by using kinematics or dynamic modeling. Partial least squares regression method can simplify the data structure and analyze the correlation between variables. It has great advantages in multivariate data analysis. Based on the partial least squares regression method, the fuze plate speed is analyzed and predicted in the paper.

2. Partial least squares regression model
Partial least squares regression method uses one algorithm to simultaneously implement regression modeling, data structure simplification, and correlation analysis between two sets of variables, which brings great convenience to multivariate data analysis [2-5]. In the paper since the modeling of the speed of fuse plate is single dependent variable regression, the principle of partial least squares regression of single dependent variable is briefly described here.

2.1. Partial least squares regression principle
Given the dependent variable y and p independent variables (x₁, x₂, ..., xₚ), the sample amount n, which forms the matrix X=[x₁, x₂, ..., xₚ]nxp and variable y. First, the components t₁ are extracted from X, and t₁ is a linear combination of x₁, x₂, ..., xₚ. It is required that t₁ carries the mutation signal in X as much as
possible, and it has the greatest correlation with \( y \). In this way, \( t_1 \) represents \( X \) as well as possible, and has the strongest interpretation ability for \( y \) [6].

After the first principal component \( t_1 \) is extracted, the regression of \( y \) and \( X \) on \( t_1 \) is implemented. If the regression equation has reached a satisfactory accuracy at this time, the algorithm stops; otherwise, the second principal component \( t_2 \) will be extracted, and the regression of \( y \) and \( X \) to \( t_1 \) and \( t_2 \) is continued.

To extract the second principal component, the residual information after the first principal component \( t_1 \) is extracted, the regression of \( y \) on \( t_1 \) and the residual information after \( y \) is interpreted by \( t_1 \) will be used. Repeat this process until we can reach a satisfactory accuracy. If \( m \) components \( t_1, t_2, \ldots, t_m \) \((m \leq p)\) are finally extracted for \( X \), the partial least squares regression will perform the regression of \( y \) on \( t_1, t_2, \ldots, t_m \) in order. Since \( t_1, t_2, \ldots, t_m \) are all linear combinations of \( x_1, x_2, \ldots, x_p \), they can be finally expressed as the regression equation of \( y \) on the original dependent variable \( X \) [7-8].

2.2. Partial least squares regression modeling

The modeling method of partial least squares for single dependent variable is as follows:

1) Normalize \( X \) and \( y \) to obtain the normalized independent variable matrix \( E_0 \) and the dependent variable matrix \( F_0 \).

\[
\begin{align*}
F_0 &= (F_{0y})_n \\
F_{0y} &= [y - E(y)]/S_y
\end{align*}
\]

(1)

\[
\begin{align*}
E_0 &= (E_{01}, E_{02}, E_{03}, \ldots, E_{0m})_{n \times m} \\
E_{0i} &= x_i^* = [x_i - E(x_i)]/S_{x_i}, (i = 1, 2, \ldots, m)
\end{align*}
\]

(2)

Where \( F_0 \) and \( E_0 \) are the normalized matrices of \( y \) and \( X \); \( E(y) \) and \( E(x_i) \) \((i \) same as above) are the average values of \( y \) and \( X \); \( x_i^* \) is the vector after \( x_i \) is normalized; \( S_y, S_{x_i} \) are the mean square error of \( y \) and \( X \); \( n \) is the sample amount.

2) Extraction of the first component \( t_1 \)

First extract the first component \( t_1 \) from \( E_0 \), \( t_1 = E_0 \omega_1 \), where:

\[
\omega_1 = \frac{E_0^T F_0}{||E_0^T F_0||} = \frac{1}{\sqrt{\sum_{i=1}^{m} r^2(x_i,y)}} \begin{bmatrix} r(x_1,y) \\ r(x_2,y) \\ \vdots \\ r(x_m,y) \end{bmatrix}
\]

(3)

\[
t_1 = E_0 \omega_1 = \frac{1}{\sqrt{\sum_{i=1}^{m} r^2(x_i,y)}} \left[ r(x_1,y)E_{01} + r(x_2,y)E_{02} + \cdots + r(x_m,y)E_{0m} \right]
\]

(4)

Where \( r(x_i,y) \) is the correlation coefficient between \( x_i \) and \( y \).

After \( \omega_1 \) is obtained, component \( t_1 \) can be obtained, and then the regression equations of \( E_0 \) and \( F_0 \) on \( t_1 \) are obtained:

\[
E_0 = t_1 P_1^T + E_1, \quad F_0 = t_1 r_1 + F_1
\]

(5)

Where \( P_1 = \frac{E_0^T t_1}{||t_1||^2} \), \( r_1 = \frac{F_0^T t_1}{||t_1||^2} \) are regression coefficients; \( E_1 \) and \( F_1 \) are the residual matrix of the regression equation, where:

\[
E_1 = E_0 - t_1 P_1^T = [E_{11} + E_{12} + \cdots + E_{1m}], \quad F_1 = F_0 - t_1 r_1
\]

(6)

3) Extraction of the second component \( t_2 \)
Replace the $E_0$ and $F_0$ with the residual matrices $E_1$ and $F_1$, and calculate the second axis $\omega_2$ and the second component $t_2$, then

$$
\omega_2 = \frac{E_1^T F_1}{\|E_1^T F_1\|} = \frac{1}{\sqrt{\sum_{i=1}^{m} \text{Cov}(E_{1i}, F_1)}} \begin{bmatrix} \text{Cov}(E_{11}, F_1) \\ \text{Cov}(E_{12}, F_1) \\ \vdots \\ \text{Cov}(E_{1m}, F_1) \end{bmatrix} \tag{7}
$$

$$
t_2 = E_1 \omega_2 = \frac{1}{\sqrt{\sum_{i=1}^{m} \text{Cov}^2(E_{1i}, F_1)}} \left[ \text{Cov}(E_{11}, F_1)E_{11} + \text{Cov}(E_{12}, F_1)E_{12} + \cdots + \text{Cov}(E_{1m}, F_1)E_{1m} \right] \tag{8}
$$

Where $\text{Cov}(E_{1i}, F_i)$ is the covariance.

Perform the regression of $E_1, F_1$ to $t_2$, and the regression equation is as follows:

$$
\begin{align*}
E_1 &= t_2 P_2^T + E_2 \\
F_1 &= t_2 F_2 + F_2
\end{align*} \tag{9}
$$

where $P_2 = \frac{E_1^T t_2}{\|t_2\|^2} t_2 = \frac{E_1^T t_2}{\|t_2\|^2}$ are the regression coefficient.

4) Extraction of the $h$-th component $t_h$

Similarly, the $h$-th component $t_h$ can be extracted, and $h$ can be identified by the cross-validation principle, and $h$ is not greater than $X$'s rank.

5) Solve partial least squares regression model

$F_0$’s least squares regression equation for $t_1, t_2, \ldots, t_h$ is:

$$
\hat{F} = r_1 t_1 + r_2 t_2 + \cdots + r_h t_h \tag{10}
$$

Since $t_1, t_2, \ldots, t_h$ are all linear combinations of $E_0$, we can get:

$$
t_i = E_{i-1} \omega_i = E_0 \omega_i^* (i = 1, 2, \ldots, h) \tag{11}
$$

$$
\omega_i^* = \prod_{k=1}^{i-1} (1 - \omega_k P_k^T) \omega_i \tag{12}
$$

Bring formula(11) into formula (10), then into formula(10):

$$
\hat{F} = r_1 E_0 \omega_1^* + r_2 E_0 \omega_2^* + \cdots + r_h E_0 \omega_h^* = E_0 (r_1 \omega_1^* + r_2 \omega_2^* + \cdots + r_h \omega_h^*) \tag{13}
$$

Note $y^* = F_0 \cdot x_i^* = E_{0i}$, $\alpha_i = \sum_{k=1}^{h} r_k \omega_{ki}^*$ (i = 1, 2, ..., m), the equation (13) can be reduced to a regression equation with standardized variables as:

$$
\hat{y}^* = \alpha_1 x_1^* + \alpha_2 x_2^* + \cdots + \alpha_m x_m^* \tag{14}
$$

Equation (14) can be further written as the partial least squares regression equation of the original variable as:

$$
\hat{y} = \left[ E_0 (y) - \sum_{i=1}^{m} \alpha_i \frac{S_y}{S_{xi}} E(x_i) \right] + \alpha_1 \frac{S_y}{S_{xi}} x_1 + \alpha_2 \frac{S_y}{S_{x2}} x_2 + \cdots + \alpha_m \frac{S_y}{S_{xm}} x_m \tag{15}
$$

2.3. Cross validation analysis

Cross validation analysis can determine the number of components that should be extracted. Let $y_i$ be the original data, and $t_{i1}, t_{i2}, t_{i3}, \ldots, t_{im}$ are the components extracted during the partial least squares regression.
\( \hat{y}_i (h < m) \) is the fitted value of the \( i \)-th sample point after regression modeling using all sample points and taking \( t_1 \sim t_h \) components. \( \hat{y}_{i(-i)} \) is the value of \( y_i \) calculated by deleting the sample point \( i \) during modeling and taking \( t_1 \sim t_h \) components for regression modeling.

Note

\[
\begin{align*}
SS_h &= \sum_{i=1}^{n}(y_i - \hat{y}_h)^2 \\
P_{\text{RESSh}} &= \sum_{i=1}^{n}(y_i - \hat{y}_{i(-i)})^2 \\
Q_h^2 &= 1 - \frac{P_{\text{RESSh}}}{SS_{n-1}}
\end{align*}
\]

Where \( SS_h \) and \( P_{\text{RESSh}} \) are the sum of squared prediction error of \( y \) under two different situations; \( Q_h^2 \) is the cross validation coefficient. When \( Q_h^2 \geq (1-0.95^2)=0.0975 \), the marginal contribution of the \( t_h \) component is considered to be significant, and the introduction of a new principal component \( t_h \) will significantly improve the predictive ability of the model [7].

3. Speed Analysis of electromagnetic fuze Plate

3.1. Basic information

In the paper we collect the system parameter data obtained from the actual measurement during the joint test of the test system, as shown in Table 1. Among them: \( x_1 \) is the hydraulic driving force \( P \) (MPa), \( x_3 \) is the opening control voltage \( V \) (V) of the solenoid valve, \( x_3 \) is the driving hydraulic oil temperature \( T \) (°C), and \( y \) is the speed of the fuze plate (m/s). The data were divided into two groups: from 1 to 46 sets for analysis modeling, and the rest sets for model testing.

| Number | \( x_1 \) | \( x_2 \) | \( x_3 \) | \( y \) | Number | \( x_1 \) | \( x_2 \) | \( x_3 \) | \( y \) |
|--------|--------|--------|--------|---|--------|--------|--------|--------|---|
| 1      | 9.42   | 2.5    | 31.8   | 9.8 | 27     | 15.38  | 4.2    | 30.41  | 20.23 |
| 2      | 11.28  | 2.5    | 30.84  | 10.33 | 28    | 15.4   | 4.2    | 30.64  | 20.25 |
| 3      | 12.05  | 2.5    | 30.1   | 10.54 | 29    | 15.41  | 4.5    | 29.63  | 21.68 |
| 4      | 12.21  | 2.5    | 30.27  | 10.75 | 30    | 15.42  | 4.5    | 30.96  | 21.9  |
| 5      | 13     | 2.5    | 32     | 10.95 | 31    | 15.46  | 4.5    | 30.95  | 21.89 |
| 6      | 14.85  | 2.5    | 32.32  | 11.19 | 32    | 18.01  | 4.5    | 33     | 21.35 |
| 7      | 18.01  | 2.5    | 32.81  | 12.5  | 33    | 18.01  | 4.5    | 33.3   | 21.54 |
| 8      | 9.96   | 2.7    | 31.74  | 10.85 | 34    | 18.02  | 4.5    | 32.58  | 21.66 |
| 9      | 10.6   | 3      | 30.73  | 13.25 | 35    | 18.02  | 4.5    | 34     | 21.45 |
| 10     | 11.96  | 3      | 29.38  | 13.35 | 36    | 18.02  | 4.5    | 35.3   | 21.65 |
| 11     | 15.38  | 3      | 33.48  | 13.54 | 37    | 18.04  | 4.5    | 34.9   | 21.58 |
| 12     | 24.67  | 3      | 44     | 15.33 | 38    | 20.55  | 4.5    | 29.51  | 22.5  |
| 13     | 15.33  | 4      | 28.56  | 17.85 | 39    | 20.3   | 4.5    | 30.7   | 22.6  |
| 14     | 15.37  | 4      | 34.61  | 18.46 | 40    | 22.06  | 4.5    | 31.11  | 22.56 |
| 15     | 15.41  | 4      | 28.96  | 18.52 | 41    | 23.3   | 4.5    | 32.15  | 22.9  |
| 16     | 17.86  | 4      | 32.1   | 18.6  | 42    | 23.34  | 4.5    | 31.54  | 22.8  |
| 17     | 17.99  | 4      | 35     | 18.8  | 43    | 23.4   | 4.5    | 33     | 22.7  |
| 18     | 18.03  | 4      | 32     | 19.3  | 44    | 23.9   | 4.5    | 48.2   | 23.2  |
| 19     | 18.04  | 4      | 31.6   | 18.4  | 45    | 24.78  | 4.5    | 39.24  | 23.7  |
| 20     | 18.04  | 4      | 34.4   | 18.8  | 46    | 24.14  | 4.5    | 46.5   | 23.6  |
| 21     | 18.04  | 4      | 34.2   | 18.4  | 47    | 13.53  | 3      | 32     | 13.05 |
| 22     | 18.07  | 4      | 32.23  | 18.7  | 48    | 15.35  | 3      | 32.32  | 13.62 |
| 23     | 18.34  | 4      | 29.38  | 19.2  | 49    | 17.49  | 4      | 32.4   | 18.97 |
| 24     | 19.5   | 4      | 27.2   | 20.8  | 50    | 18.03  | 4      | 32     | 19.3  |
| 25     | 20.01  | 4      | 35.13  | 19.6  | 51    | 20.05  | 4.5    | 30.7   | 22.32 |
| 26     | 15.37  | 4.2    | 30.06  | 20.27 | 52    | 23.51  | 4.5    | 32.15  | 22.93 |
3.2. Multiple correlation

Calculate the correlation coefficient between independent and dependent variables. The results are shown in Table 2.

|   | x1  | x2  | x3  | y   |
|---|-----|-----|-----|-----|
| x1| 1   | 0.6364 | 0.5532 | 0.7384 |
| x2| 1   | 0.1503 | 0.9836 |
| x3|     | 1     | 0.2282 |
| y |     |       |       | 1     |

It is known from Table 2: the correlation coefficient between the independent variables is up to \( r(x_1, x_2) = 0.6364 \), which indicates that there is a strong autocorrelation between the independent variables; the correlation coefficient between the dependent variable and the independent variable is the highest up to \( r(x_2, y) = 0.9836 \), indicating that there is a high correlation between the independent variable system and the dependent variable system.

3.3. Modeling regression equation

According to the partial least squares modeling step and the principle of cross validation, the regression equation is obtained as:

\[
\hat{y} = -4.1617 + 0.2140x_1 + 4.9797x_2 - 0.0101x_3 \tag{17}
\]

It can be seen that the speed \( y \) of the fuze plate has a positive correlation with the hydraulic driving force \( P \) of the test device and the opening control voltage \( V \) of the oil circuit solenoid valve, and a weak negative correlation with the hydraulic oil temperature \( T \).

In order to more intuitively and quickly observe the marginal role of each independent variable in explaining \( y \), a regression coefficient graph can be drawn, as shown in Figure 1.

![Histogram of regression coefficients](image)

Figure 1: Histogram of regression coefficients

It can be immediately known from the histogram of the regression coefficient that the opening control voltage \( V \) variable of the oil circuit solenoid valve plays an extremely important role in explaining the regression equation.
3.4. Model evaluation

3.4.1. Cumulative interpretation ability analysis. The cumulative interpretation of $t_1$, $t_2$, and $t_3$ are shown in Table 3. As can be seen, the division of the principal component $t_1$ comprehensively explains the 61.38% variation information of the original independent variable system, which is a very good representative of the original independent variable system. At the same time, it comprehensively explains 89.27% of the information of the dependent variable system, which contributes a lot to the dependent variable system. The second and third principal components $t_2$ and $t_3$ have a lower ability to explain the information variation of the original (dependent) variable system. The calculation formula for the cumulative interpretation ability is as follows:

$$R_d(x_i; t_1, t_2, ..., t_h) = \sum_{j=1}^{h} r^2(x_i, t_j)$$

(18)

Table 3 Cumulative interpretation of $t_1$, $t_2$, and $t_3$

|   | $t_1$ | $t_2$ | $t_3$ |
|---|------|------|------|
| 0.8100 | 0.0737 | 0.1163 |
| 0.8070 | 0.1574 | 0.0356 |
| 0.2244 | 0.6915 | 0.0841 |
| 0.6138 | 0.3075 | 0.0841 |
| 0.8927 | 0.00876 | 0.0086 |

3.4.2. Correlation analysis. Figure 2 shows the correlation of $t_1$, $t_2$, $t_3$ and $u_1$ ($u_1 = \frac{F_0}{r_1}$). As can be seen from the figure, there is a clear linear relationship between $t_1$ and $u_1$, indicating that the hydraulic driving force $P$ of the test device, the control voltage $V$ of the solenoid valve opening, and the temperature $T$ of the hydraulic oil have significant correlations with the speed of fuse plate $y$. Residual information $t_2$, $t_3$ and $u_1$ also have a certain linear relationship, but it is already very weak.

![Figure 2](image-url)
3.4.3. Fitted results of the fuze plate speed \( y \). Table 4 shows the comparison between the \( y \)-fitting results of the fuze plate speed and the measured values.

| Section  | Number | Measured values | Fitted values | Relative error /% | Number | Measured values | Fitted values | Relative error /% |
|----------|--------|-----------------|---------------|-------------------|--------|-----------------|---------------|-------------------|
| Fitted section 1 | 9.8 | 9.9823 | -1.8262 | 24 | 20.8 | 19.6554 | 5.8233 |
| 2 | 10.33 | 10.39 | -0.5775 | 25 | 19.6 | 19.6844 | -0.4288 |
| 3 | 10.54 | 10.5622 | -0.2102 | 26 | 20.27 | 19.6554 | 5.8233 |
| 4 | 10.75 | 10.5947 | 1.4649 | 27 | 20.23 | 19.7372 | 2.4966 |
| 5 | 10.95 | 10.7464 | 1.8946 | 28 | 20.25 | 19.7392 | 2.5877 |
| 6 | 11.19 | 11.139 | 0.4579 | 29 | 21.68 | 21.2454 | 2.0456 |
| 7 | 12.5 | 11.8103 | 5.8398 | 30 | 21.9 | 21.2341 | 3.136 |
| 8 | 10.85 | 11.0944 | -2.2029 | 31 | 21.89 | 21.2428 | 3.0467 |
| 9 | 13.25 | 12.7354 | 4.0407 | 32 | 21.65 | 21.7678 | -1.9193 |
| 10 | 13.35 | 13.0401 | 2.3765 | 33 | 21.54 | 21.7648 | -1.0329 |
| 11 | 13.54 | 13.7306 | -1.8811 | 34 | 21.66 | 21.7742 | -0.5245 |
| 12 | 15.33 | 15.6124 | -1.8088 | 35 | 21.65 | 21.7598 | -1.4237 |
| 13 | 17.85 | 18.7513 | -4.8066 | 36 | 21.65 | 21.7467 | -0.4447 |
| 14 | 18.46 | 18.9667 | -1.266 | 37 | 21.58 | 21.755 | -0.8044 |
| 15 | 18.52 | 18.7623 | -1.2914 | 38 | 22.5 | 22.3466 | 0.6865 |
| 16 | 18.6 | 19.2549 | -3.4012 | 39 | 22.6 | 22.2811 | 1.4313 |
| 17 | 18.8 | 19.2535 | -2.3554 | 40 | 22.56 | 22.6536 | -0.4132 |
| 18 | 19.3 | 19.2923 | 0.0399 | 41 | 22.9 | 22.9084 | -0.0367 |
| 19 | 18.4 | 19.2985 | -4.6558 | 42 | 22.8 | 22.9232 | -0.5374 |
| 20 | 18.8 | 19.2702 | -2.44 | 43 | 22.7 | 22.9213 | -0.9655 |
| 21 | 18.4 | 19.2722 | -4.5257 | 44 | 23.2 | 22.8747 | 1.4221 |
| 22 | 18.7 | 19.2986 | -3.1018 | 45 | 23.7 | 23.1535 | 2.3603 |
| 23 | 19.2 | 19.3851 | -0.9549 | 46 | 23.6 | 22.9433 | 2.8623 |

It can be seen from Table 4 that the maximum relative error between the measured value and the fitted value is 5.8393%. For the design index of the test device system, the model has a higher fitting accuracy, indicating that the partial least squares regression model can perform well. It reflects the relationship between the fuze plate's speed and the matching parameters, such as the hydraulic driving force \( P \) of the test device, the opening control voltage \( V \) of the oil solenoid valve, and the hydraulic oil temperature \( T \).

4. Conclusion
In this paper, a partial least squares regression method is used, and after cross validation tests, a relationship model is established between the speed of the fuze plate \( v \) and the hydraulic driving force \( P \) of the test device, the opening control voltage \( V \) of the oil circuit solenoid valve, and the hydraulic oil temperature \( T \). Through the test, the model has higher explanatory ability of independent and dependent variables, especially the opening control voltage \( V \) of the solenoid valve. By comparing with the measured values, the model has a high fitting accuracy, which provides a theoretical analysis basis for the reasonable selection of system parameters of the test device and the pre-test prediction of the fuse plate speed. It also plays an important guiding role in exploring how to improve the safety and service life of the system.
Acknowledgments
Corresponding author: Wang Wenyun, wangwenyun3035@163.com, engaged in research and development of mechatronics products.

References
[1] GJB 1995-94, Test methods for electromagnetic fuze and magnetic induction fuze [S]. (In Chinese).
[2] Wang Ju, Meng He, et al. Partial least squares regression for predicting economic loss of vegetables caused by acid rain [J]. Journal of Chongqing University (English Edition), 2009, 8 (1): 10-16.
[3] Liu Lei, Du Jianqiang, et al. Feature selection method based on correlation based feature selection and partial least square [J]. Journal of Jiangxi University of TCM, 2019, 31 (2): 88-91. (In Chinese)
[4] Zhou XP, Wang WS, Zeng HJ. Partial least square method and BP neural network coupled model and its application to prediction of pH value of acid rains [J]. Advances in Science an Technology of Water Resources, 2006, 26 (4): 50-52. (In Chinese).
[5] Wang HW. Partial least-squares regression – method and applications [M]. Beijing: National Defence Industry Press, 1999. (In Chinese).
[6] Wang Wen sheng, Ding Jing, et al. Study on the Long Term Prediction of Annual Electricity Consumption Using Partial Least Squares Regressive Model [J]. Proceedings of the CSEE, 2003, 23(10):17-21. (In Chinese).
[7] Ren Ruoen, Wang Huiwen. Analysis of Multivariate Statistical Data—Theories, Methods, and Examples [M] . Beijing: National Defense Industry Press, 1998. (In Chinese).
[8] Richard A. Johnson, Dean W. Wichern. Practical Multivariate Statistical Analysis (Fourth Edition) [M]. Translated by Lu Xuan. Beijing: Tsinghua University Press, 2001. (In Chinese).