Applicability of Simple Linear Regression Method on Water Absorption Rate Forecasting of PMMA and Its Composites

Kui Chen, Tianyun Zhang & Xiuge Cao

ABSTRACT: On the basis of researching the steps of single linear regression forecasting method, taking water absorption rate forecasting of PMMA/MMT nanocomposites in H_2SO_4 solution, NaOH solution and H_2O as examples, the applicability of single linear regression method on water absorption rate forecasting of PMMA and its composites was studied. The results show that the maximum water absorption rate relative error of forecasting results of composites in three kinds of medium is 7.02%, which is allowable for industrial application, and single linear regression method can be used in water absorption rate forecasting of PMMA and its composites.

INTRODUCTION

The existing of polar side-chain methyl leads to the fact that water can be absorbed by PMMA and its products, which may cause performances decline and even lead to deformation in severe cases, and affects seriously safe use of PMMA and its products (Wu, 2012). Thus, the data collection of water absorption rate is an important work before the application of PMMA and its products under some special conditions (Chen et al., 2013). Theoretically, this job can be done through experiments under artificial accelerated or actual working conditions. But there are inevitable differences between artificial accelerated condition and actual working condition, while experiments under actual working condition need a long time, may be several years, even decades. It’s difficult to keep up with the rapid development of materials research, let alone the cost of experiments is high (Chen et al., 2014). Recently, according to experimental data related, forecasting properties of materials through mathematical model established is one of the focus of attention and study (Barbero, 2012). For this situation, taking the water absorption rate forecasting of PMMA/MMT composites synthesized by emulsion polymerization as an example, the applicability of simple linear regression method in the water absorption rate forecasting of PMMA and its composites was researched in this paper.

1 The paper was supported by No. LZCU-XZ2014-14 Scientific Research Foundation of Lanzhou City University. Corresponding Author: Tianyun Zhang, Email: lzyunyun@126.com, Mobile phone: 15294185884
1 STEPS OF SIMPLE LINEAR REGRESSION FORECASTING METHOD

1.1 Establishing regression forecasting equation

Regression analysis equation, i.e. regression analysis model, can be set up according to the calculation of historical data of independent variable and dependent variable. Set \( x \) as the independent variable, \( y \) as the dependent variable, and if there is a linear relationship between \( x \) and \( y \), then simple linear regression equation can be expressed as

\[
\hat{y}_i = a + bx_i, \quad i = 1, 2, \ldots, n
\]

where \( \hat{y}_i \) is the estimated value of \( y_i \), \( a \) and \( b \) are regression coefficients. The estimated values of regression coefficient are calculated by Eq. 2 and Eq. 3 according to the least square method.

\[
\hat{b} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}
\]

\[
\hat{a} = \frac{\sum y_i}{n} - \hat{b} \frac{\sum x_i}{n}
\]

1.1 Correlation coefficient calculation

Correlation coefficient \( R \) is an index that is used to measure linear correlation relationship between independent variable \( x \) and dependent variable \( y \). Regression equation established is meaningful only if there is a certain relationship between \( x \) and \( y \). Correlation coefficient \( R \) is calculated by Eq. 4.

\[
R = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}}
\]

where \(-1 \leq R \leq 1\). \( x \) and \( y \) are positive correlation when \( R \) is positive value, otherwise they are negative correlation. Hence, absolute value of \( R \) represents the correlation degree of \( x \) and \( y \). Generally, there are following three kinds of relationships between them. When \( R = 0 \), \( x \) and \( y \) are completely uncorrelated, which means the variation of \( x \) has no influence on \( y \). When \( R = 1 \), \( x \) and \( y \) are completely correlated, which means the variation of \( y \) is determined by that of \( x \), and there is a linear relationship between them. When \( 0 < |R| < 1 \), \( x \) and \( y \) are common correlation, which means the variation of \( x \) can influence \( y \) partially. The correlation degree of \( x \) and \( y \) increases with the increase of \( |R| \). As a general rule, when \( |R| > 0.7 \), there is a high linear
correlation between $x$ and $y$. When $|R| < 0.3$, there is a low linear correlation between them, and when $0.3 \leq |R| \leq 0.7$, they have a medium linear correlation.

1.2 Significance testing

Regression equation can be used to predict on condition that it passes the significance testing. The widely used method of significance testing in linear regression model is the correlation coefficient testing method which can be divided into the following three steps: First, the calculation of correlation coefficient $R$; Second, finding out the critical value $R_{a} (n - 2)$ from the critical value table of correlation coefficient according to freedom degree $n - 2$ of regression model and significance level $\alpha$ given; Third, judgment. If $|R| \geq R_{a} (n - 2)$, then the linear correlativity between two variables is significant, which means that the model can be used to forecast; Conversely, there is no significant linear correlativity between two variables if $|R| < R_{a} (n - 2)$, which means that model established can’t be used to forecast.

1.3 Calculation of predicted value and prediction interval

Regression model can be used to forecast after passing significance testing. Substituting a given independent variable $x_{0}$ into the model, corresponding predicted value $\hat{y}_{0}$ can be obtained. In practice, predicted value can’t just equal measured value, a certain deviation appears accordingly. Thus the conclusion is questionable if it contains only predicted value. As a matter of fact, both predicted value and its prediction interval should be given. For this purpose, standard error $S_{y}$ is calculated by Eq. 5 first.

$$S_{y} = \sqrt{\frac{\sum y_{i}^2 - \hat{a} \sum y_{i} - \hat{b} \sum x_{i} y_{i}}{n - 2}}$$  \hspace{1cm} (5)$$

Then, t testing is performed. The critical value $t_{n/2} (n - 2)$ is obtained by t distribution table according to freedom degree $n - 2$ of regression model and significance level $\alpha$ given. Forecasting is performed at last. Suppose predicted position is $(x_{0}, y_{0})$, the predicted value can be expressed as

$$\hat{y}_{0} = \hat{a} + \hat{b} x_{0}$$  \hspace{1cm} (6)$$

When significance level is $\alpha$, prediction interval $D$ of predicted value $\hat{y}_{0}$ is

$$D = \hat{y}_{0} \pm t_{n/2} (n - 2) S_{y}$$  \hspace{1cm} (7)$$
where \( S_0 = S \sqrt{\frac{1}{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}}} \).

2 APPLICATION

When PMMA/MMT nanocomposites containing 3wt% MMT synthesized by using emulsion polymerization was soaked in 0.1 mol/L H\(_2\)SO\(_4\) solution, 0.1 mol/L NaOH solution or deionized water, a small amount of inorganic MMT dispersed evenly in the PMMA matrix in nanoscale lamella disorderly, which increases effective way of water diffusion movement on PMMA substrate, and improves liquid barrier property of substrate. Accordingly, water absorption rate of composites reduces obviously. Table 1 is the effect of soaking time on water absorption rate of PMMA/MMT nanocomposites (Chen et al., 2013).

<table>
<thead>
<tr>
<th>Soaking time /min</th>
<th>Water absorption rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H(_2)SO(_4) solution</td>
</tr>
<tr>
<td>7</td>
<td>0.5036</td>
</tr>
<tr>
<td>14</td>
<td>0.6224</td>
</tr>
<tr>
<td>21</td>
<td>0.7128</td>
</tr>
<tr>
<td>28</td>
<td>0.7623</td>
</tr>
<tr>
<td>35</td>
<td>0.8019</td>
</tr>
<tr>
<td>42</td>
<td>0.8473</td>
</tr>
<tr>
<td>49</td>
<td>0.8762</td>
</tr>
<tr>
<td>56</td>
<td>0.9019</td>
</tr>
<tr>
<td>63</td>
<td>0.9125</td>
</tr>
</tbody>
</table>

First, water absorption rate of composites after soaking 64 minutes in 0.1 mol/L H\(_2\)SO\(_4\) solution was predicted according to that after soaking 7 to 63 minutes showed in Table 1. Estimated values \( \hat{\alpha} \) and \( \hat{\beta} \) of regression coefficient are obtained by Eq. 2 and Eq. 3. Thus, regression forecasting equation is \( \hat{y}_i = 0.5307 + 0.0069 x_i \). Correlation coefficient \( R \) between soaking time and water absorption rate equals 0.9555 by Eq. 4. When significance level \( \alpha \) equals 0.05, and freedom degree is \( 9 - 2 = 7 \), \( R_{0.05}(7) \) is equal to 0.6664 by critical value table of correlation.
coefficient. Since $R = 0.9555 > 0.6664 = R_\alpha (7)$, test passes for $\alpha = 0.05$, i.e., the linear correlativity between two variables is significant. The standard error $S$ of evaluated value is 0.0435 by Eq. 5. When significance level $\alpha$ equals 0.05, and freedom degree is 7, critical value $t_{\alpha/2} (7) = 2.365$ is obtained by $t$ distribution table. Substituting soaking time 64 minutes into Eq. 6, predicted value and relative error are obtained. When significance level $\alpha$ is 0.05, prediction interval of predicted value is calculated by Eq. 7. All the results were given in Table 2. Table 2 also gives the predicted values and prediction intervals of water absorption rate after soaking 64 minutes in 0.1 mol/L NaOH solution and deionized water. The detailed calculating process was omitted here.

<table>
<thead>
<tr>
<th>Soaking medium</th>
<th>Actual value</th>
<th>Predicted value</th>
<th>Relative error</th>
<th>Prediction interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>H$_2$SO$_4$ solution</td>
<td>0.9128</td>
<td>0.9705</td>
<td>6.32%</td>
<td>0.8489~1.0921</td>
</tr>
<tr>
<td>NaOH solution</td>
<td>1.0335</td>
<td>1.0909</td>
<td>5.55%</td>
<td>0.9713~1.2105</td>
</tr>
<tr>
<td>Deionized water</td>
<td>0.8113</td>
<td>0.8686</td>
<td>7.02%</td>
<td>0.7491~0.9881</td>
</tr>
</tbody>
</table>

It can be seen by Table 2 that the maximum water absorption rate relative error of forecasting results of composites in three kinds of medium is 7.02%, which is allowable for industrial application. The results above shows that water absorption rates obtained are reasonable and reliable, and single linear regression method can be used in water absorption rate prediction of PMMA and its composites.

3 SUMMARY

According to simple linear regression forecasting method, water absorption rates of PMMA/MMT nanocomposites in H$_2$SO$_4$ solution, NaOH solution and deionized water were predicted respectively. Forecasting results demonstrate that simple linear regression method works very well in forecasting water absorption rate of PMMA and its composites. Simple linear regression method needs less experimental data and is easy to test, is a simple and reliable way for water absorption rate prediction of PMMA and its composites.

REFERENCES

