IOP Conference Series: Materials Science and Engineering

PAPER • **OPEN ACCESS**

Predication of Dry Rubber Content in Hevea Latex based on SPSS Multiple Regression Model

To cite this article: Fan Zhang *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **563** 022030

Predication of Dry Rubber Content in Hevea Latex based on SPSS Multiple Regression Model

Fan Zhang[1], Lusheng Liao[1], Hui Huang[1*], Fuquan Zhang[1], Bingbing Wang, Huaqian Long[1]

[1] Key Laboratory of Tropical Crop Products Processing of Ministry of Agriculture and Rural Affairs, Agricultural Products Processing Research Institute, Chinese Academy of Tropical Agricultural Sciences, Zhanjiang, Guangdong , 524001, P. R. China

[*]Corresponding author's e-mail: 554244189@qq.com

Abstract. The Dry Rubber content (DRC) in hevea latex is affected by a great many of factors, including producing area, strains, tree age, cutting system, origins and so on. Each factor was discussed and evaluated to confirm the influence degree. A predication model based on SPSS multiple regression models was constructed to analyse the factors influencing the DRC, and the regression equation was used to forecast the DRC in natural rubber latex. The results show that the predication model has a good achievement and the average relative error was 7.4%

1. Introduction

Natural rubber (NR) is an important industrial raw material, which is widely used in aerospace, military project, heavy vehicles, aircraft tyres and rail transit and other fields. Natural rubber latex is a biological product of a complex composition, including dry rubber (about 20-45%), protein（1.5%）, inorganic matter（2%）, carbohydrate（1%）, resinous substances(2%), water（22-48%）and other substances. However, the dry rubber content (DRC) plays a key role in the latex and it determined the price of the latex [1].

The predication of DRC can provide scientific data for both plantation and processing, which may also have a positive impact on the allocation of this material and the sustainable of NR industry [2]. Generally speaking, the predication data can be obtained by empirical analogy such as comprehensive index method and average growth rate method. However, many factors for determination the DRC of hevea latex have been confirmed, such as producing area, strains, tree age, modification and strength of Rubber Cutting, and there are irregular changes in the data of these above factors. As a result, the use of these methods cannot achieve the desired results [3].

In this article, an attempt is made to find an alternative method of predication of DRC of field latex. The multiple regression models were utilized to help to construct the predication of DRC model. Using this method, we need to test the regression coefficient, and to discuss and study the prediction and hypothesis test. After that, we should consider the relationship between the independent variables and to test whether there is a collinearity problem. If there is, variables need to be screened in order to overcome collinearity problem and increase the accuracy of prediction.

2. Univariate analysis

Rubber tree is affected by many nature factors during its growth process, and its composition of fresh latex also varies greatly in latex output and dry rubber content (DRC). As the DRC is determined the price of hevea latex, it is necessary to study its condition of production. According to what we mentioned above, the DRC can be affected by many factors. So in this section, we will analysis the relationship between five factors, including strains, tree age, cutting time, cutting style and producing area, and DRC.

2.1 Relationship between strains and DRC

Five strains of PR107, Reyan 7-33-97, Reyan 7-20-59, RRIM600 and Reyan 88-13 were sampled from the experimental site of the Chinese Academy of Tropical Agricultural Sciences, all this sample was made in May, 2018. The DRC testing standard method was used to analyze and test the rubber processing plant of the experimental site of the Chinese Academy of Tropical Agricultural Sciences. The test results are shown in Table 1. And the result of Table 1 shows that the absolute content of dry latex of different strains can reach 12.9%.

Table1. Change of DRC in different strains.

No.	1	2	3	4	5
Strain	PR107	Reyan7-33-97	Reyan7-20-59	RRIM600	Reyan88-13
DRC(%)	34.1	36.6	23.7	32.8	28.0

2.2 Relationship between cutting time and DRC

The cutting time and month will greatly affect the DRC, as it closely related to the temperature. According to the relevant research reports, the correlation coefficient between DRC and temperature is about 0.852. Based on this, the difference of dry rubber content between different months in 2018 of the 5 teams of the experimental site of the Chinese Academy of Tropical Agricultural Sciences was analyzed. The experimental data were obtained from the rubber processing factory of the experimental site of the Chinese Academy of Tropical Agricultural Sciences. Table 2 shows the change between month, temperature and DRC. As we can see that, the absolute difference of DRC between different months is 6%.

Table 2. Changes of DRC in different months.

months	5	6	7	8	9	10	11	12
Maximum temperatures	35	35	31	30	33	28	29	28
Minimum temperatures	25	27	24	24	24	17	16	21
Average temperature	30	31	27.5	27	28.5	22.5	22.5	24.5
DRC(%)	34.1	33.9	32.0	31.4	29.7	27.6	28.1	29.9

2.3 Relationship between cutting system and DRC

Studying the influence of different cutting systems on DRC, a series of sampling experiments were designed. Sampling subjects were 12 teams from the experimental site of Chinese Academy of Tropical Agricultural Sciences. The latex strain was PR107. Sampling time was October-December. The cutting system was divided into liquid and gas stimulation. The interval of each sampling was 3-4 days. The results of the influence of different tapping systems on dry rubber content were shown in Table 3. In Table 3, LS2 represents the second cutting after liquid stimulation and GS1 represents the first cutting after gas stimulation.

Table 3. Changes of DRC in different cutting systems

Cutting system	LS2	LS3	LS4	GS1	GS2	GS3
DRC(%)	28.0	24.0	27.8	30.3	30.0	29.0

2.4 Relationship between producing area and DRC

In view of the dry rubber content of latex of the same strain from different origins in China, a comparative experiment was designed in this study. PR107 strains were studied in Danzhou (Hainan), Xishuangbanna (Yunnan) and Gaozhou (Guangdong) respectively. The sampling time was July. The data in Table 4 showed the influence of the average DRC of each producing area.

Table 4. Changes of DRC in different producing areas

Producing area	Danzhou	Xishuangbanna	Gaozhou
DRC(%)	34.1	27.4	32.0

2.5 Relationship between tree age and DRC

This study compared the effects of tree age on dry gum content in PR107 strain. The data were collected from the above-mentioned comparative experiments and the data recorded by rubber processing factories in the experimental site of the Chinese Academy of Tropical Agricultural Sciences over the years, as shown in Table 5.

Table 5. Changes of DRC in different tree ages

age	10 years	15 years	20 years	25 years
DRC(%)	27.3	29.8	31.1	33.9

From Table 5, we can easily see that the DRC presents an increasing trend with the growing of tree age.

3. Predication Model Based on SPSS Multiple Regression Model

3.1 The principle of SPSS

The principle of SPSS multivariate linear regression model is that, when analyzing a dependent variable of statistical analysis, there is often more than one independent variable affecting the dependent variable, then we need to consider the relationship between k independent variables, including *X1, X2, X3*..., and Xk, and dependent variable y, after that the regression equation is established as follows [4].

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \ldots + b_k x_{ik} + u_i$$

Formula: $b_0, b_1, \ldots, b_k$ is the regression coefficient with estimation; $i = 1,2,\ldots N$ (n is sample size) and u_i is random error.

Assuming that the total distribution of random errors N (0, σ^2) is independent of each other, we can estimate $b_0, b_1, \ldots$ *and* b_k with the least square method under the observation samples of X and Y, then the form of matrix can be described as $Y = XB + u$, where:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \end{bmatrix} \qquad X = \begin{bmatrix} 1 & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ 1 & \cdots & x_{ik} \end{bmatrix}$$

$$B = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix} \qquad u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_i \end{bmatrix}$$

The least squares estimator of the total parameters can be obtained by using the least squares method.

$$BLS = (X^T X)^{-1} X^T Y$$

Among them, X^T is the transposition of X matrix.

3.2 Analysis based on SPSS

The SPSS (version 21) used in this study is mainstream statistical analysis software. Its basic functions include descriptive statistics, correlation analysis, regression analysis, clustering analysis, time series analysis and so on. It has the characteristics of simple operation, convenient operation and powerful function.

According to above content, we used SPSS to analyse five factors that affect the DRC in latex. These five factors can be suggested as seven independent variables, including the maximum and minimum temperature of the cutting day and these five factors. From Table 6, X1 stands for strain, X2 stands for months, X3 stands for the highest temperature (Celsius degree) on the day of extracting latex, X4 stands for the lowest temperature (Celsius degree) on the day of extracting latex, X5 stands for cutting system (1 for liquid stimulation, 2 for gas stimulation and 3 for no stimulation), X6 stands for producing area(1 for Hainan, 2 for Yunnan and 3 for Guangdong), X7 stands for tree age (year), Y stands for DRC.

Table 6. The Prepared Data between factors and DRC

No.	X1	X2	X3	X4	X5	X6	X7	Y
1	1	5.18	36	26	1.1	1.2	15	36.6
2	2	5.21	35	25	1.1	1.1	20	34.1
3	1	12.04	30	21	1.2	1.3	23	30.0
4	1	11.02	22	17	2.1	1.3	23	28.0
5	3	12.01	28	22	2.3	1.3	23	28.0
6	2	7.17	31	24	1.1	1.1	20	33.9
7	1	8.28	30	24	1.3	1.2	15	34.2
8	2	9.25	33	24	1.1	1.2	20	30.0
9	2	7.15	33	27	1.1	2.1	15	32.01
10	2	7.20	34	28	1.3	2.1	13	21.08
11	1	10.12	26	19	1.2	3.1	10	27.4
12	1	11.05	29	15	1.2	3.1	10	26.5
13	3	11.12	27	17	1.3	3.1	15	26
14	1	12.05	30	21	2.1	1.3	23	30.0

After inputting the above data into SPSS, we built the prepared model based on SPSS Multiple Regression Model to analysis it. Figure 1. shows the summary of multiple regression models. And a coefficients table related to Model 1has been showed in Figure.2.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.795[a]	.631	.201	3.67236

a. Predictors: (Constant), Zscore: Tree age, Zscore: Minimum temperature, Zscore: Strain, Zscore: Cutting system, Zscore: Month, Zscore: Maximum temperature, Zscore: Producing area

b. Dependent Variable: DRC

Figure 1. Model Summary

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	29.842	.981		30.405	.000		
	Zscore: Strain	-.170	1.390	-.041	-.122	.907	.537	1.861
	Zscore: Month	-1.686	2.242	-.410	-.752	.480	.206	4.845
	Zscore: Maximum temperature	.371	2.057	.090	.180	.863	.245	4.077
	Zscore: Minimum temperature	-2.280	2.302	-.555	-.991	.360	.196	5.106
	Zscore: Cutting system	-1.068	1.672	-.260	-.639	.547	.371	2.696
	Zscore: Producing area	-3.712	3.423	-.903	-1.084	.320	.089	11.296
	Zscore: Tree age	-.603	3.147	-.147	-.192	.854	.105	9.544

a. Dependent Variable: DRC

Figure 2. The coefficients of Model 1

From Fig.1and Fig.2, we can see that each independent variable is related to the dependent variable. Though the value of R and R square in Fig.1 is low to 0.9, and the value of sig is over 0.1, an approximate optimal model can be acquired from it. And this Stepwise regression equation can be written as follows.

$$Y = 29.842 - 0.170X1 \ - 1.686X2 + 0.371X3 - 2.280X4 - 1.068X5 - 3.712X6 \ - 0.603X7$$

3.3 Error Analysis

According to the above model, we can calculate the predication value of DRC when we know the basic information of the fresh latex. The error analysis result shows in table 7.

Table 7 Error Analysis

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14
DRC(%)	36.6	34.1	30.0	28.0	28.0	33.9	34.2	30.0	32.0	21.1	27.4	26.5	26.0	30.0
Predication DRC(%)	34.7	34.7	30.6	30.5	26.6	33.6	32.6	31.9	28.1	27.3	25.9	27.9	25.2	28.3
Relative error(%)	5.19	1.76	2.00	8.93	5.00	8.85	4.68	6.33	12.21	29.50	5.47	5.25	3.07	5.67

From Table 7, we can see that the predication value of DRC can be obtained approximately. The maximum relative error was Sample 10 at 29.50%，compared with only 1.76% on Sample 2. The average relative error of all these samples was 7.4%.

4. Conclusion

This article has presented five factors that affect the DRC in fresh latex. In order to investigate the degree of influence, the relationship between various factors, including strains, tree age, cutting time, cutting style and producing area, and DRC was analyzed one by one. Then a predication model based on SPSS Regression analysis was built to predict the value of DRC. Error analysis shows this model can approximately acquire the DRC of hevea latex, and the average relative error was 7.4%.

Directions for future work include getting more accurate model by SPSS Curve Regression and Nonlinear Regression model, and constructing a mature predication system.

Acknowledgement

This research was financially supported by the Natural Science Foundation of Hainan Province, China (Grant No. 317248)

References

[1]Zhang F., Li P.W., Lv M.Z., Yang Z.M.. (2015) The Influence Factors of Measurement of Dry Rubber Content in Natural Rubber Latex by Microwave Technique. Guangdong Chemical Industry, 42(313):202-203.

[2] K KAIDA. (1982) Determination of Dry Rubber Content of Hevea Latex by Microwave Techniquel. Pertanika, 5(02): 192-195.

[3]Fang L., Li D.Q., Li H.B.(2008) Application of DH925A Microuwave Set in Determining Dry Rubber Content of Rubber Latex. Journal of Hebei Agricultural Sciences, 12 (6) : 171- 172.
[4] Zhou P.F., Lu Z.Y.. (2018) Prediction of Urban Water Consumption Based on SPSS Multiple Linear Regression Model. Water Conservancy Science and Technology and Economy, 24(5): 6-10.