

Original Research Article

<https://doi.org/10.20546/ijcmas.2020.906.323>

## Prediction of Cocoon Shell Weight of Tasar (*Antheraea mylitta* Drury) Silkworm using LASSO Regression

Rahul Banerjee<sup>1\*</sup>, Debasis Chattopadhyay<sup>2</sup> and ZMS Khan<sup>2</sup>

<sup>1</sup>ICAR-Indian Agricultural Statistics Research Institute, Pusa, New Delhi-110012, India

<sup>2</sup>Central Tasar Research & Training Institute, Piska-Nagri, Ranchi-835303, India

\*Corresponding author

### ABSTRACT

#### Keywords

Tasar Sericulture;  
Shell Weight;  
Cocoon Sorting;  
LASSO Regression;  
RMSE

#### Article Info

Accepted:  
20 May 2020  
Available Online:  
10 June 2020

Tasar Silk is one of the four most important silk produced in India; the other variants being Mulberry, Muga and Eri silk. Uniform quality of tasar silk yarns is required during weaving. The quality of silk yarn depends on the quality of the cocoon accordingly the cocoons are sorted into different categories viz. A, B and C. The present system of grading of Tasar Cocoons is determined by the shell weight of the cocoon. Since the shell weight of the cocoon cannot be obtained without cutting the cocoon and removing the pupae rendering the cocoon unreelable is unrealistic and difficult for larger lots of cocoon. In this paper a study has been done to predict the shell weight of the cocoons of Tasar silkworm (*Antheraea mylitta* Drury) through LASSO Regression based on the other cocoon dimensions, that can provide an alternative option for sorting of the cocoons based on shell weight.

### Introduction

Asia is the top producer of silk in the world contributing 95% of the total global output. There are 40 countries on the world map of silk; bulk of it is produced in China and India, followed by Japan, Brazil and Korea [Nagaraju, 2008]. Sericulture plays a crucial role in the optimal utilization of the natural resources in a most effective manner for socioeconomic upliftment with livelihood and

employment and income generation [Malik MS *et al.*, 2008]. In India there are mainly of 5 types of silk that are being produced viz. Mulberry Silk, Tasar Silk (Tropical and Temperate Tasar).

Tropical Tasar comes third in production in India among all the other silks after Mulberry Silk and Eri silk. Tasar Silk is copperish in colour and coarser than mainly used for furnishings and interiors.

Tasar silk is generated by the silkworm, *Antheraea mylitta* which mainly thrive on the food plants Asan and Arjun. The rearings are conducted in nature on the trees in the open. In India, tasar silk is mainly produced in the states of Jharkhand, Chhattisgarh and Orissa, besides Maharashtra, West Bengal and Andhra Pradesh. Tasar culture is the main stay for many a tribal community in India. Tasar Sericulture is one of the major sources of livelihood among the tribal communities of Jharkhand, Chhattisgarh, Odisha, West Bengal and Telangana.

The existing method of classifying tasar cocoons as category a, b or c is determined by the shell weight of the tasar cocoons. The shell weight is obtained by measuring the weight of the cocoon without the pupae, this involves cutting of the cocoon and removing the pupa from the cocoon. Since it is not possible to cut all the cocoons and measure their shell weight for their effective grading and it also affects the reel ability of the cocoon. Thus development of a method for prediction of the shell weight without cutting open the cocoons is of prime importance. Currently, the grading of tasar cocoons in larger lots is usually done by the visual observation method. This method is an approximate method and leads to erroneous results. A novel method of predicting the shell weight of the tasar cocoons is being illustrated in this paper.

### Materials and Methods

The acronym “LASSO” stands for Least Absolute Shrinkage and Selection Operator. Lasso regression is a type of linear regression that uses shrinkage. Shrinkage is where data values are shrunk towards a central point, like the mean. The lasso procedure encourages simple, sparse models (i.e. models with fewer parameters). This particular type of regression is well-suited for models showing high levels

of multicollinearity or when you want to automate certain parts of model selection, like variable selection/parameter elimination.

The Lasso Regression performs an L1 regularization, in this process a penalty is added the penalty is equal to the absolute value of the magnitude of the coefficients. This type of regularization can result in sparse models with few coefficients; some coefficients can become zero and eliminated from the model. Larger penalties result in coefficient values closer to zero, which is the ideal for producing simpler models. Let us consider a multiple linear regression model.

$$y_i = x_{ij}\beta_j + e_i, \forall i = 1(1)n, j = 1(1)p$$

Lasso solutions are quadratic programming problems, which are best solved with software (like Matlab). The goal of the algorithm is to minimize:

$$\sum_{i=1}^n (y_i - \sum_j x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j|$$

Which is the same as minimizing the sum of squares with constraint  $\sum_j |\beta_j| \leq s$ . Some of the  $\beta$ s are shrunk to exactly zero, resulting in a regression model that’s easier to interpret. A tuning parameter,  $\lambda$  controls the strength of the L1 penalty.  $\lambda$  is basically the amount of shrinkage:

When  $\lambda = 0$ , no parameters are eliminated.

The estimate is equal to the one found with linear regression.

As  $\lambda$  increases, more and more coefficients are set to zero and eliminated (theoretically, when  $\lambda = \infty$ , all coefficients are eliminated).

As  $\lambda$  increases, bias increases.

As  $\lambda$  decreases, variance increases.

The structure of a tasar cocoon is ellipsoid in

nature it's a solid ellipse.

**Data description**

The following observations were taken for measuring the cocoons namely; weight of the

cocoon, length of the cocoon two way width of the cocoon and the volume of the cocoon. The following is a sample data description of the 500 data measurements that were being taken for the study

**Table.1**

Physical Characteristics of Tsar Cocoons (Ecorace: Daba, Bivoltine)							
#	Weight of Cocoons (g)	Shell Weight (g)	Shell Ratio (%)	Width of Cocoon (cm)		Length of Cocoon (cm)	Volume of Cocoon (cc)
				a	b		
1	15.46	1.92	12.42	3.10	3.10	5.30	30.00
2	19.82	2.52	12.71	3.40	3.30	5.70	35.00
3	16.38	2.08	12.70	3.20	3.10	5.40	30.00
4	12.48	1.86	14.90	3.10	3.10	4.90	27.00
5	12.05	1.73	14.36	3.10	3.10	5.10	30.00

The dependent variable for the model was shell weight, whereas the independent variables being cocoon weight, Length of Cocoon, Width of Cocoon, Volume of Cocoon.

The observations were taken for 500 randomly selected bivoltinetasar cocoons of DabaEcorace in Post Cocoon Section of CTR&TI, Ranchi.

**Results and Discussion**

Shell weight was taken as the independent variable and the dependent variables being, Length of Cocoon, Width of Cocoon, Volume of Cocoon and Cocoon weight. Pearson's product moment Correlation analysis was performed to find out the correlation between the different variables. The correlation matrix is as follows:

**Table.2**

	Weight of Cocoon	Shell Weight	Length of Cocoon	Width a	Width b	Volume
Weight of Cocoon	1					
Shell Weight	0.157	1				
Length of Cocoon	0.724	0.463	1			
Width a	0.625	0.652	0.868	1		
Width b	0.576	0.632	0.830	0.938	1	
Volume	0.721	-0.00864	0.611	0.491	0.426	1

The LASSO based model was run on the R software using the following model:

['weight\_cocoon','length\_cocoon','width\_cocoon\_a','width\_cocoon\_b','volume\_cocoon']  
 Target variable: 'Shell Weight'  
 ML model used: Lasso regression  
 The results obtained are as follows:

**Cocoon shell weight prediction**

Input features for the model:

**Table.3**

<b>Actual shell weight</b>	<b>Predicted shell weight</b>
1.63	1.733
1.70	1.806
2.62	2.451
2.62	2.595
2.16	2.080
2.02	2.124
2.19	2.169
1.50	1.622
1.86	2.001
2.44	2.269
2.77	2.634
1.94	1.942
2.01	2.026
2.12	2.319
2.02	2.094
1.86	1.550
1.85	1.583
1.95	2.026
1.92	2.051
2.23	2.568
2.07	2.055
1.83	1.895
2.02	2.233
2.17	2.050
1.97	1.895



Fig.1 Tasar Cocoons

Root Mean Square Error (RMSE) is a standard way to measure the error of a model in predicting quantitative data. Formally it is defined as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i^p - y_i)^2}$$

where,  $y_i$  's are the observed values,  $y_i^p$  's are the predicted values and n is the number of observation,  $i=1(1)n$ .

The RMSE value for the current study came out to be 0.15, thereby indicating that LASSO is one of the best methods of accurately predicting the shell weight of tasar cocoons without the need of cutting the cocoon and affecting its reelability. Thus, LASSO can provide a good alternative of sorting of tasar cocoons in different grades.

## References

- Dewangan S. K. *et al.* [2011]: Sericulture - A Tool of Eco-System Checking Through Tribal. Journal of Environmental Research and Development, Vol.6 No.1, July-Sep
- Hans, C. [2009]: Bayesian Lasso Regression, *Biometrika*, 96, 4, pp. 835–845.
- Jolly, MS (1976). Package of practices for tropical tasar culture, Ranchi. Central Tasar Research Station, (Central Silk Board, Bombay), pp 32.
- Malik, M.S. *et al.*, [2008]: Socio-economic upliftment of tribal communities in Jharkhand through Agroforestry based farming system.
- Nagaraju, J [2008]: Silk of India, grace and luster. *Biotechnology News*, vol.3 [5] pp: 4
- Pal, Aet. *al.* [2013]: Development of Machine Vision Solution for Grading Of Tasar Silk Yarn. Proceedings of the 2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013).
- Rai, S *et al.*, [2006]: Tasar sericulture, an emerging discipline for conservation and sustainable utilization of natural resources. The vision review.

## How to cite this article:

Rahul Banerjee, Debasis Chattopadhyay and ZMS Khan. 2020. Prediction of Cocoon Shell Weight of Tasar (*Antheraea mylitta Drury*) Silkworm using LASSO Regression. *Int.J.Curr.Microbiol.App.Sci.* 9(06): 2656-2660. doi: <https://doi.org/10.20546/ijemas.2020.906.323>