

Original Research Article

<https://doi.org/10.20546/ijcmas.2021.1001.337>

Study of Biometrical Characters on Rice using Multiple Linear Regression Models

J. S. Harshavardhini^{1*}, R. Vasanthi², N. Thavaprakash³ and M. Kalpana⁴

¹Agricultural statistics, ³Department of Agronomy, Tamil Nadu Agricultural University, Coimbatore, India

²Department of Physical Sciences & Information Technology, AEC & RI, TNAU, Coimbatore

⁴Computer science, ADAC & RI, Trichy, India

**Corresponding author*

ABSTRACT

Keywords

Preharvest yield, Biometrical characters, Multiple linear regression, Prediction

Article Info

Accepted:
18 December 2020
Available Online:
10 January 2021

Pre harvest yield forecasting models play a vital role in policy planning and decision - making premises. The data on biometrical characters of paddy along with yield was collected to train the model and test for prediction. Three models were developed using the Multiple linear regression method. Model 1 was developed using the original data values, Model 2 was developed using the logarithmic values and Model 3 using the square root of the independent variables. From the validation measures such as RMSE, AIC and BIC values, it was observed that Model 2 was proved to be the best fit for preharvest prediction.

Introduction

Agriculture in India, is highly pluralistic and multidimensional and has an irremarkable improvement in the last three decades. The growth in Indian agriculture is evidential on how it has cope up with the demand of food grains for increasing population and its contribution to the national income is more than any other developed countries in the world. Farming sector occupies two-third of the country's working population.

For undertaking the policy decisions regarding storage, distribution, marketing, exports and imports, information about the yield forecasts for major and commercial crops are necessary. Pre-harvest crop forecasting is of immense help for the farmers for formulating their future plan of action. It also helps in pest and disease management. The advance estimates are provided by the Directorate of Economics and Statistics but, they are subjected to the impressionistic evaluation and therefore a need to develop

statistically sound preharvest forecasting models became a prerequisite. Crop cutting experiments are also being carried out wherein the results can be obtained only after the harvest. Therefore, statistical techniques are being employed to obtain the objective forecasts with well-reasoned precision with prior to the crop harvest. Yield prediction is an operational research technique which provides the timely and reliable information to the decision makers and private actors for making planning premises during the growing phases of the crop. Many forecasting models have been developed using yield and the weather variables. In this research an attempt has been made to develop various forecasting model using the plant biometrical characters

which determines the crop yield.

Materials and Methods

Data description

The data on yield and its related biometrical characters were recorded from each plot of different lines using the standard method of measurements. The experiment was conducted at Tamil Nadu Agricultural University, Coimbatore in the year 2019. The data encompasses the yield components and eight biometrical characters which influences the yield of 64 different varieties. The recognized biometrical characters are listed as follows

| | | |
|---|--|--|
| 1. Leaf length (x ₁) | 4. Panicle length (cm) (x ₄) | 7. Plant height (cm) (x ₇) |
| 2. Leaf breadth (x ₂) | 5. Basal girth (x ₅) | 8. No of panicles/m ² (x ₈) |
| 3. No of tillers/plot (x ₃) | 6. Days to 50% flowering (x ₆) | 9. Yield (kg/plot) (Y) |

Multiple linear regression

Regression analysis is a statistical modelling technique primarily used for prediction and forecasting purposes. Multiple linear regression, which is an extension of linear regression attempts to predict the outcome of the predictor variable based on two or more explanatory variables. The data on yield and the biometrical characters of the rice crop was fitted using the Multiple Linear regression model to determine the linear relationship between the dependent and the independent variables.

Four models were formulated using the regression approach. Model 1 was developed using the biometrical characters measured Model 2 was developed using the logarithmic transformation of the independent variables along with yield. Model 3 was developed using the square root transformation of the independent variables.

The multiple linear regression model fitted to the data on paddy are as follows

Model 1

$$Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_5x_5 + \beta_6x_6 + \beta_7x_7 + \beta_8x_8 + e$$

Model 2

$$Y = \beta_0 + \beta_1\log x_1 + \beta_2\log x_2 + \beta_3\log x_3 + \beta_4\log x_4 + \beta_5\log x_5 + \beta_6\log x_6 + \beta_7\log x_7 + \beta_8\log x_8 + e$$

Model 3

$$Y = \beta_0 + \beta_1\sqrt{x_1} + \beta_2\sqrt{x_2} + \beta_3\sqrt{x_3} + \beta_4\sqrt{x_4} + \beta_5\sqrt{x_5} + \beta_6\sqrt{x_6} + \beta_7\sqrt{x_7} + \beta_8\sqrt{x_8} + e$$

Where,

Y is the crop yield per plot (kg/ha),

x₁, x₂, x₃, x₄, x₅, x₆, x₇, x₈ are the explanatory variables,

β₀ is the y-intercept,

$\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8$ are the slope coefficients of each explanatory variables, e is the model's error term which is independently normally distributed with mean zero and constant variance.

Performance Metrics of the model

The accuracy measures used for selection of the best fit model are as follows

Root mean squared error (RMSE)

It is the most frequently used measure of accuracy used to compare different forecasting models. The metric is scale dependent and can be used only for the models having the same dataset. It measures the error values of the actual and the observed values of the prediction model. The model is better with low RMSE values and the model is a perfect fit for zero RMSE value.

The RMSE value is mathematically expressed as follows

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{Y}_i - Y_i)^2}{n}}$$

Akaike's information criterion (AIC)

Akaike's information criterion (AIC) compares the quality of a set of statistical models to each other. The basic formula is defined as:

$$AIC = -2(\log\text{-likelihood}) + 2K$$

Where,

K is the number of model parameters (Number of variables in the model plus the intercept).

Log-likelihood is a measure of model fit. Minimum AIC is the score for the best model.

This is usually obtained from statistical output.

Bayesian Information Criterion (BIC)

It is an index used in Bayesian statistics to choose between two or more alternative models. The BIC is also known as the Schwartz information criterion (SIC) or the Schwartz-Bayesian information criteria.

The Bayesian Information Criterion (BIC) is defined as

$$BIC = k \log(n) - 2\log(L(\theta))$$

Where,

n is the sample size

k is the number of parameters

θ is the set of all parameters

L (θ) represents the likelihood of the model tested

The model with the lowest BIC to be considered as best.

Results and Discussion

Model 1

The regression model was developed using the original variables. The Multiple linear model formed was written as

$$y = -7.606 + 0.016x_1 - 0.262x_2 + 0.059x_3 + 0.063x_4 + 0.125x_5 + 0.047x_6 + 0.012x_7 + 0.001x_8$$

Model 2

The Multiple linear model developed using the logarithmic transformation of the data was written as

$$y = -37.057 + 1.1971x_1 - 0.7792x_2 + 2.3000x_3 + 2.4871x_4 + 3.8884x_5 + 9.2665x_6 + 3.3489x_7 + 1.1769x_8$$

Model 3

The Multiple linear model developed using the square root transformation of the data was written as

$$y = -17.5515 + 0.1895x_1 - 0.5962x_2 + 0.5006x_3 + 0.5298x_4 + 0.9197x_5 + 0.8801x_6 + 0.2691x_7 + 0.0611x_8$$

The regression models developed was statistically significant as the p value of the overall significance of the model is less than 0.05. The value of R² calculated for model 1, 2 and 3 was 0.6965, 0.7126 and 0.705 respectively which is an explained measure of variation of the model.

Table.1 Predicted yield of the regression models

| Actual yield (Kg/ha) | Model 1 | Model 2 | Model 3 |
|----------------------|---------|---------|---------|
| 3.90 | 2.75 | 2.77 | 2.76 |
| 2.74 | 3.49 | 3.52 | 3.51 |
| 3.34 | 2.54 | 2.58 | 2.56 |
| 2.38 | 3.04 | 3.03 | 3.03 |
| 2.09 | 2.53 | 2.54 | 2.53 |
| 3.23 | 3.55 | 3.60 | 3.58 |
| 3.34 | 3.75 | 3.77 | 3.76 |
| 2.40 | 1.70 | 1.74 | 1.72 |
| 2.67 | 2.50 | 2.47 | 2.49 |
| 4.49 | 4.69 | 4.61 | 4.66 |
| 2.79 | 2.80 | 2.78 | 2.79 |
| 4.15 | 3.92 | 3.95 | 3.94 |

Table.2 Performance metrics of the model

| | Model I | Model II | Model III |
|------|---------|----------|-----------|
| RMSE | 0.5790 | 0.5731 | 0.5770 |
| AIC | 91.90 | 89.05 | 90.42 |
| BIC | 111.41 | 108.56 | 109.93 |

Fig.1 Actual Vs Predicted yield of paddy for Model 1

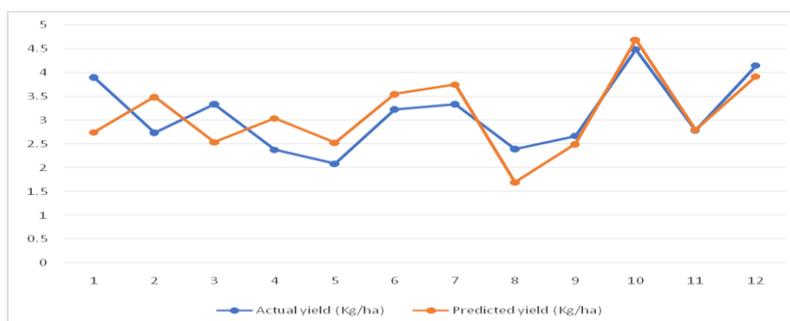


Fig.2 Actual Vs Predicted yield of paddy for Model 2

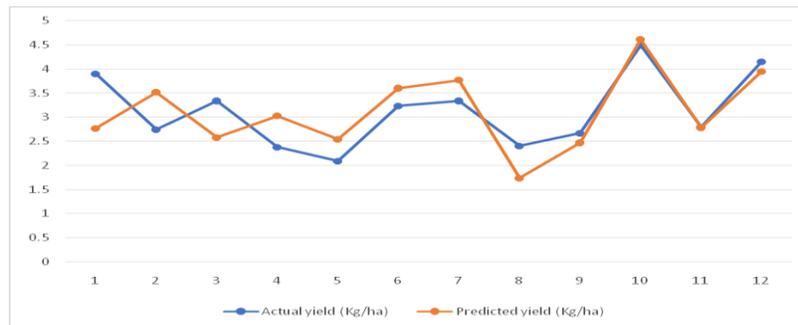
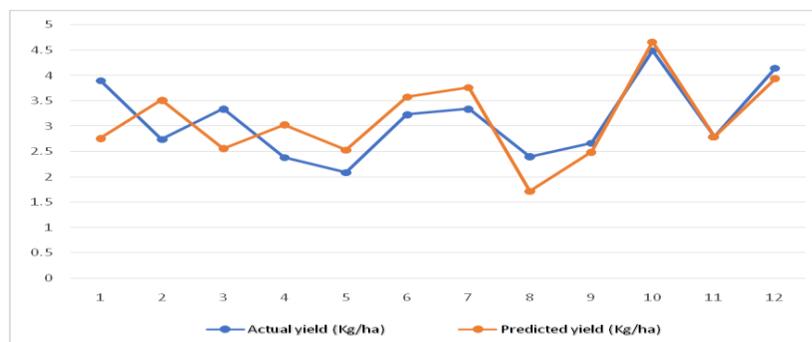


Fig.3 Actual Vs Predicted yield of paddy for Model 3



Since R^2 value affects the significance of the model due to the increase of independent variables, Adjusted R^2 values are considered as the modified and improved form of R^2 value. The adjusted R^2 value for model 1, 2 and 3 was 0.64, 0.6592 and 0.6501 respectively. The actual yield of the test data was compared with the predicted yield of the three models developed and tabulated in table 1.

The overall performance of the model was validated using the accuracy measures such as RMSE, AIC and BIC values. The model with the lowest measure was observed in model 2 and was considered as the best fit model.

In conclusion the Multiple linear regression model was developed using the biometrical characters of paddy. Regression models developed for pre harvest yield prediction

were compared using the validate measures and it was proved that the model 2 showed lesser error proportions and better prediction.

References

- Agrawal, R., 2006. Forecasting techniques in crops. Indian Agricultural Statistics Research Institute.
- Agrawal, R., Jain, R.C., and Singh, D., (1980). Forecasting of rice yield using climatic variables. Indian Journal of Agricultural Sciences. 50(9), 680-684.
- Annu et al., (2016). An application of principal component analysis for pre-harvest forecast model for rice crop based on biometrical characters. J. Appl. & Nat. Sci. 8 (3): 1164– 1167.
- Annu, Sisodia, B.V.S. and Rai, V.N. (2017). An application of principal component analysis for pre- harvest forecast model

- for wheat crop based on biometrical characters. *Internat. Res. J. Agric. Eco. & Stat.*, 8 (1): 83-87.
- Annu, Sisodia, B.V.S. and Kumar, Sunil (2015). Pre-harvest forecast models for wheat yield based on biometrical characters. *Economic Affairs*, 60(1):89-93.
- Banakara K. B., Garde Y. A., Pisal R. R. and Bhatt B. K. (2018). Pre-Harvest Forecasting of Rice Yield for Effective Crop Planning Decision in Surat District of South Gujarat. *Int.J.Curr.Microbiol.App.Sci.*, 7(06): 3410-3422.
- Jain, R.C., Sridharan, H and Agarwal, R. (1984). Principal component technique for forecasting of sorghum yield. *Indian J. Agric. Sci.* 54 (6): 467-70.
- Kumar, M., Singh, M.M. and Kumar, S., (2017). Preharvest Forecasting of Rice Yield Using Biometrical Characters Along with Farmers Appraisal in Muzaffarpur District of Bihar, *Int. J. Pure App. Biosci.* 5(5): 1553-1558.
- Kumar M, Singh MM, Singh SP, Bharat RC, Sanjeev Kumar, Reddy Ravindra B., (2013). Pre-harvest Forecasting of Rice Yield using Biometrical Characters along with Farmers Appraisal in Samastipur District of Bihar. *R.A.U. Journal of Research.* 23(1-2):41-45.
- Pandey, K.K., (2013). Pre -Harvest Forecast Models Based on Weather variable and weather indices for Eastern U.P. *Adv.Biores.* 4(2):118-122.

How to cite this article:

Harshavardhini, J. S., R. Vasanthi, N. Thavaprakash and Kalpana, M. 2021. Study of Biometrical Characters on Rice using Multiple Linear Regression Models. *Int.J.Curr.Microbiol.App.Sci.* 10(01): 2910-2915.
doi: <https://doi.org/10.20546/ijcmas.2021.1001.337>