**Advances in Bioresearch**
Adv. Biores., Vol 6 (4) July 2015: 137-141
©2015 Society of Education, India
Print ISSN 0976-4585; Online ISSN 2277-1573
Journal's URL:http://www.soeagra.com/abr.html
CODEN: ABRDC3
ICV 7.20 [Poland]

Advances in Bioresearch

**ORIGINAL ARTICLE**

# Use of Statistical and Crop Growth Simulation Models for Predicting rice yield in Lower Gangetic Plain of India

**Ria Biswas[1*], Saon Banerjee[2] and Banjul Bhattacharyya[3]**
[1]Research Scholar, Department of Agricultural Statistics, Bidhan Chandra Krishi Viswavidyalaya, Faculty of Agriculture, P.O. – Krishi Viswavidyalaya, Mohanpur, Nadia, West Bengal, India,
[2]Associate Professor, AICRP on Agrometeorology, Directorate of Research, BCKV, Kalyani, West Bengal, India
[3]Associate Professor, Department of Agricultural Statistics, Bidhan Chandra Krishi Viswavidyalaya, Faculty of Agriculture, P.O. – Krishi Viswavidyalaya, Mohanpur, Nadia, West Bengal, India,
E-Mail: ria.biswas6@gmail.com

**ABSTRACT**
*Weather variation during crop growth period of rice (Oryza sativa) strongly influences the yield. To predict the crop yield, crop growth simulation models (CGSM) are used in general, which gives reliable results. In the present study, the applicability of statistical model has been evaluated and the performance of statistical and CGSM were compared. First the statistical model was developed considering sowing time as dummy variable. The WOFOST model, which is a popular CGSM, was used to simulate the growth of rice and predict the yield. It was observed that the statistical model performs better than CGSM. The mean predicted value of wet-season rice is 4341.62 kg ha$^{-1}$, whereas the actual average yield is 4250.62 kg ha$^{-1}$. The mean predicted yield value of statistical model is 4239.86 kg ha$^{-1}$ which is closer to the actual value. The mean absolute error value for predicted statistical model (87.28) is much lesser than for CGSM (229.00). $R^2$ and RMSE values are lower than the CGSM. Hence, for region-specific yield prediction, statistical model can be used as an effective tool, if developed properly and other effects can be incorporated through dummy variable.*
***Key words:*** *Rice, Statistical model, Dummy variable, Crop Growth Simulation Model, Yield Prediction*

Received 28/04/2015 Accepted 24/06/2015
©2015 Society of Education, India

**How to cite this article:**
Ria B, Saon B and Banjul B. Use of Statistical and Crop Growth Simulation Models for Predicting rice yield in Lower Gangetic Plain of India. Adv. Biores. Vol 6 [4] July 2015: 137-141. DOI: 10.15515/abr.0976-4585.6.4.137141

## INTRODUCTION

Rice (Oryza sativa) is the most important staple food for a large part of the world's population, especially in South East Asia [9]. It is the most popular cereal with the second-highest worldwide production, after corn [5]. About 90% of the world's rice is produced by farmers from Asian countries [1]. Presently, the variability of rainfall and monsoon shifting along with temperature rise create problems in rice production of Eastern India [2, 3]. To assess the crop-dependence on weather, dynamic crop growth simulation models can be used successfully [10, 8]. Statistical analysis through regression model is also a very good tool to predict crop-weather relationship and statistical model enables to predict the yield [11]. As the input requirements of running crop growth simulation model are plenty and not easily available in many regions, the location specific statistical model plays vital role to provide first-hand information on yield forecast. Considering the background, the present research work aims at to compare the statistical model and dynamic simulation model for rice yield prediction.

## MATERIALS AND METHODS

### Study area

The study was conducted in Nadia District, West Bengal, and Eastern India, which is situated under New Alluvial Agro-climatic Zone. The said district is located almost in central position of Lower Gangetic Plains of India. The climate of the region is characterized by hot summers and mild to moderately cool winters.

The mean annual rainfall varies from 1400 to 1600 mm in the Lower Gangetic Plains. The potential evapotranspiration (PET) ranges from 1100 to 1400 mm and the water deficit is about 400 mm. The crop-growing period is more than 270 days. The region is characterized by alluvium-derived soil.

**Development of statistical model**

To establish the relationship between crop yield and different weather parameters, multiple regression equation can be the easiest option. But one of the serious limitations of multiple regression analysis is that it accommodates only quantitative response as explanatory variables. Quantitative variable can be introduced in the regression model through dummy variable. A Dummy variable is an artificial variable created to represent an attribute with two or more distinct categories or levels.

In the present study to include the effect of date of transplanting on rice yield, the entire transplanting season has been classified into two categories. First half transplanting (early transplanting date to normal transplanting date) scored as zero (0) and second half transplanting (i.e. late transplanting) scored as one (1) to treat them as dummy variable.

The yield forecasting model has been used in this study was specified as:

$$Y = \beta_0 + \beta_1 MIN + \beta_2 MAX + \beta_3 IR + \beta_4 VP + \beta_5 WS + \beta_6 CR + \beta_7 RD + \beta_8 DMY + \epsilon \quad (1)$$

where, Y= yield of Aman rice (Kg/ha); MIN = Minimum temperature ( $^0$C); MAX = Maximum temperature ( $^0$C); IR= Irradiation (MJ $m^{-2}$ $d^{-1}$); VP = Vapor pressure (hPa);

WS = Wind speed (km $h^{-1}$); CR = Cumulative rainfall for the seasons unto the week (mm); RD = Number of days with rain (d); DMY = Dummy variable (0 = First half transplanting score and 1 = Second half transplanting score); $\epsilon$ = Stochastic term/residual term/error term [$\epsilon \sim NID(0, \sigma^2)$]

Thus, the multiple effects of several weather parameters viz., maximum temperature, minimum temperature, vapor pressure, wind speed, rainfall, relative humidity, bright sun shine hour and evaporation were taken into consideration to justify the impact of them on yield. This multiple linear regression analysis has been performed using SPSS packages.

**Collection of data for developmet of statistical model**

The weather data were collected from meteorological observatory of Bidhan Chandra Krishi Viswavidyalaya (BCKV, State Agricultural University), Kalyani (22.57$^0$N, 88.20$^0$E and 7.8 m above mean sea level), Nadia District. For development of statistical model, the yield data of wet-season rice were collected for ten years (2002 to 2011) along with weather data [7]. The developed model was validated with actual data set for subsequent two years i.e., 2012 and 2013.

**Crop Growth Simulation model for yield prediction**

In the present study, the WOFOST (World Food Studies) model was used to simulate growth and production of rice. WOFOST was originally developed as a crop growth simulation model for the assessment of the yield potential of various annual crops in tropical countries [4]. The WOFOST model considers three growth levels correspond to crop production, namely, potential production, limited production and reduced production. Crop growth is determined by irradiation, temperature, plant characteristics and crop growth management (e.g., nutrient and water management). Atmospheric $CO_2$-concentration is assumed to be constant. In case of estimating potential production, nutrient and water are assumed to be in ample supply. If the supply of water or nutrients is sub-optimal during (parts of) the growing season, this leads to water- limited and/or nutrient-limited production, which is normally lower than potential production. This holds for biomass production, in some cases water limited yield may be higher than potential yield because of more favorable harvest index. At reduced production level, the possible reduction in crop yield occurs mainly due to biotic factors like weeds, pests and diseases.

In the present study the variation of rice production in the Lower Gangetic plain was determined with actual weather data, prevailing soil characteristics and common management practices followed by the farmers in the study region. The pre-calibrated WOFOST model [2] was used here to determine the year to year variation of wet-season rice yield for different sowing dates.

**Validation of WOFOST Model**

Some common statistical parameters, namely, Coefficient of determination ($R^2$), Root Mean Square Error (RMSE), etc., were worked out to evaluate the performance of the model [6]. The following statistical tools were used in the present study:

**(i)** **Bias:** Bias is simply the difference of average predicted value and average observed value.

$$Bias = \frac{1}{N}\sum_{i=1}^{N}(fi - Oi) \quad (2)$$

where, fi is predicted yield, Oi is observed yield and N is the number of observations.

 ©2015 Society of Education, India

**(ii) Mean Absolute error (MAE):** MAE is simply the average of the absolute difference between predicted and observed value.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}(|fi - Oi|) \qquad (3)$$

**(iii) Mean Square error (MSE):** MSE measures the average of squares of the squares of the "errors". The error is the amount by which the value implied by the estimator differs from the quantity to be estimated.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(fi - Oi)^2 \qquad (4)$$

**(iv) Root mean square error (RMSE):** RMSE is the root of the MSE value. It is usually best to report the RMSE rather than MSE, as the RMSE is measured in the same units as the data, rather than in squared units.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(fi - Oi)^2} \qquad (5)$$

**RESULT AND DISCUSSION**

**Development of statistical model**

In the first step the statistical model has been developed by using ten years data-set (2002 to 2011) considering all the weather parameters viz., minimum temperature, maximum temperature, irradiation, vapor pressure, wind speed, rainfall, rainy days in the period and two categories of date of transplanting as dummy variables (0,1). Then multiple regression analysis has been carried out to obtain the equation.
The yield forecasting regression equation model has been specified as:

Yield(Y) = -15731.49 + 204.98 MIN +231.64 MAX +56.62 IR + 1265.73 VP – 134.789 WS + 0.063 CR + 52.18 RD – 848.091 DMY (6)

Since $R^2$ = 0.804, 80.4% of the variance in wet rice yield can be explained by minimum temperature, maximum temperature, irradiation, vapor pressure, wind speed, rainfall, rainy days in the period and two categories for date of transplanting (Table 1). Compared to $R^2$, adjusted $R^2$ provides a less biased estimate (70.70%) of the extent of the relationship between the variables in the population (Table 1).
The ANOVA is significant (F= 9.872, df (regression) = 8, df (residual) = 11, $\rho < 0.05$) which means that the eight predictors collectively account for a statistically significant proportion of the variance in the criterion variable (Table 1).

**Comparison between actual yield and statistical model output**

The yield data of 2012 and 2013 were used to predict the yield. The values of different explanatory variables, namely minimum temperature, maximum temperature, irradiation, vapor pressure, wind speed, rainfall, rainy days during crop growth period of 2012 and 2013 were used in the developed model. Dates of transplanting for these years were also considered to obtain the yield. The Figure 1 shows the variation of actual and predicted yield. It is clear from the figure that, the predicted yield matches very well with actual data. The statistical model slightly over-predicts the yield except for late sown crop of 2013.

**Performance of WOFOST model**

The WOFOST output was compared with actual field data of 2010, 2011 and 2012 (in each year three dates of sowing). Generally the WOFOST model over predicts the yield except two cases for the year 2012 (Figure 2). The Figure 2 shows that the gaps between actual and predicted yields are wider than that of Figure 1. The performance of model is better for early transplanted ($15^{th}$ June transplanted) crops. In actual field condition also the $15^{th}$ June transplanted crop shows better result than later transplanted crops. The WOFOST model simulates yield for the early sown crops in more reliable manner. The effect of change of transplanting date on yield of wet-season rice is correctly simulated by the WOFOST model. The sensitivity analysis for effect of temperature change on crop maturity shows that if temperature is increased by $1^0$C or $2^0$C temperature over normal temperature (NT), the maturity period is delayed by 3 and 7 days (Figure 3). Due to $3^0$ C temperature increase, the crop duration is still reduced. Thus, WOFOST model appears to be highly sensitive to temperature and date of transplanting. However, its sensitivity towards rainfall is not tested and the model is used under irrigated condition.

**Comparison between output of statistical model and CGSM**

For comparison purpose, same transplanting dates (one early and other late transplanting denoted as $d_0$ and $d_1$) were considered for both the CGSM and statistical model. From Table 2 it is observed that the statistical model performs better than CGSM. The CGSM over-predicts the yield (mean predicted value:

4341.62 kg ha-1). The predicted value through statistical model (mean: 4239.86 kg ha-1) is closer to the actual value (mean: 4250.62 kg ha-1). The mean absolute error value for predicted statistical model (87.28) is much lesser than for CGSM (229.00). For the statistical model, $R^2$ value is also higher and RMSE value is lower than the CGSM indicating the statistical model can predict the yield in a better way for this region.

**Table 1.** Summary of developed statistical model and ANOVA[b]

| Model | Sum of Squares | df | Mean Square | F | Sig. | R | R-Square | Adjusted R-Square | Standard Error of the Estimate |
|---|---|---|---|---|---|---|---|---|---|
| Regression | 4.72 E5 | 8 | 589878.25 | 9.87 | .012[a] | .897[a] | 0.804 | 0.707 | 535.709 |
| Residual | 1.07E7 | 11 | 977902.62 | | | | | | |
| Total | 1.55E7 | 19 | | | | | | | |

a. Predictors: (Constant), DMY, VP, IR, CR, WS, RD, MAX, MIN
b. Dependent Variable: Yield

**Table 2:** Comparison between developed statistical model and WOFOST model

| Treatment | Observed yield (O) | Forecasted Statistical yield(FS) | Forecasted WOFOST yield(FW) | FS-O | FW-O | Abs (FS-O) | Abs (FW-O) | $(FS-O)^2$ | $(FW-O)^2$ | $R^2$ FS | $R^2$ FW | RMSE FS | RMSE FW |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2012 $d_0$ | 4469.5 | 4478.60 | 4360 | 9.10 | -109.5 | 9.10 | 109.5 | 82.85 | 11990.25 | | | | |
| 2012 $d_1$ | 3663 | 3766.96 | 4178 | 103.96 | 515 | 103.96 | 515 | 10808.4 | 265225 | | | | |
| 2013 $d_0$ | 4707.5 | 4747.48 | 4541 | 39.98 | -166.5 | 39.98 | 166.5 | 1598.62 | 27722.25 | | | | |
| 2013 $d_1$ | 4162.5 | 3966.42 | 4287.5 | -196.07 | 125 | 196.07 | 125 | 38446.59 | 15625 | | | | |
| | 4250.62 (Average) | 4239.86 (Average) | 4341.62 (Average) | -10.75 (Bias) | 91 (Bias) | 87.28 (ME) | 229 (ME) | 12734.12 (MSE) | 80140.63 (MSE) | 0.92 | 0.90 | 112.84 | 283.09 |

$d_0$ = Early to normal transplanting
$d_1$ = Late transplanting

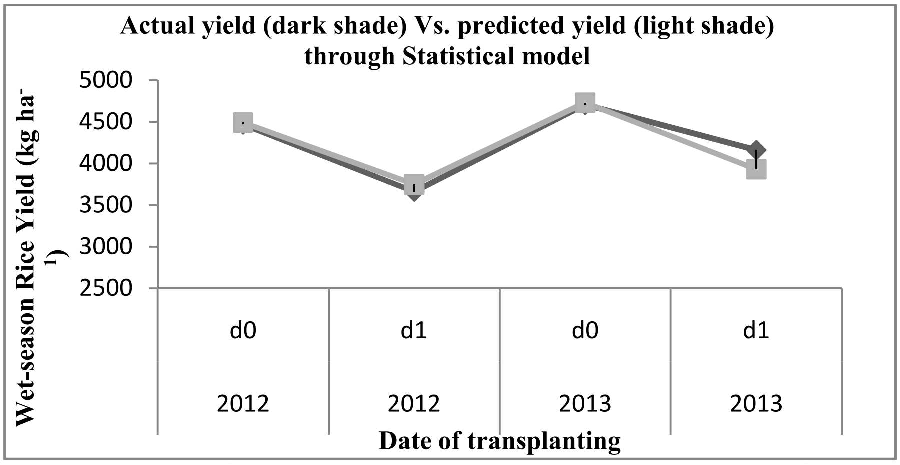


**Fig. 1.** Wet-season rice yield predict through statistical model

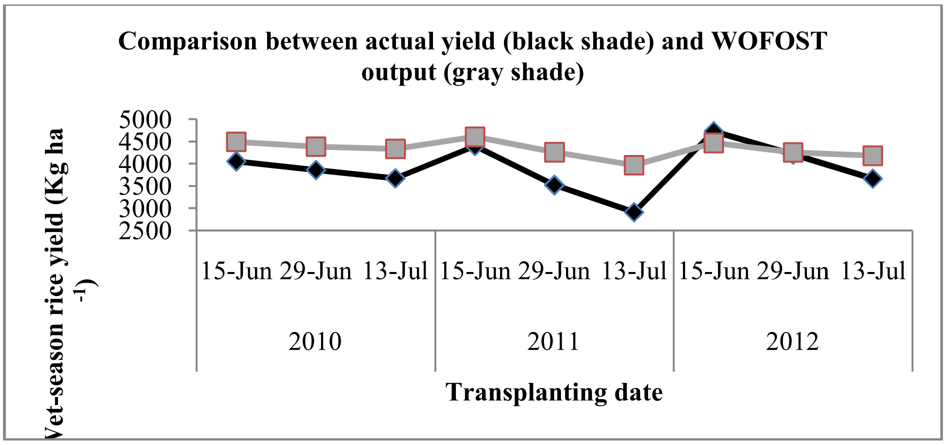


**Fig. 2.** Comparison of WOFOST output with actual yield data

 **©2015 Society of Education, India**

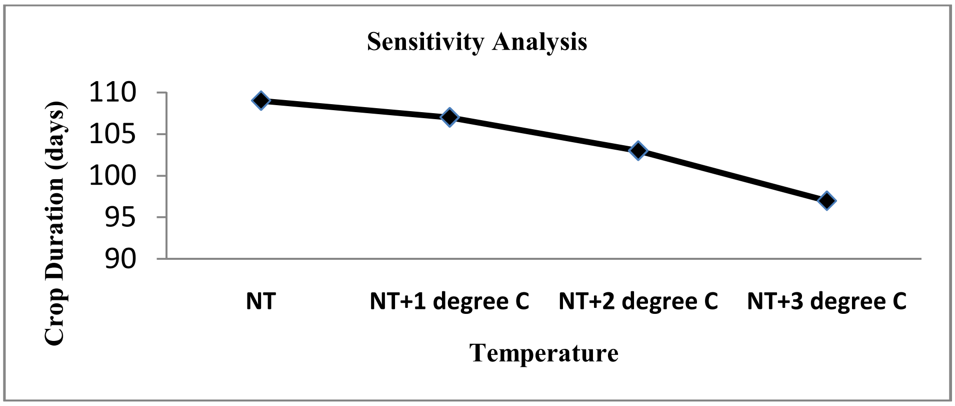


**Fig. 3.** Effect of temperature increase on crop maturity through WOFOST model (Sensitivity analysis)

## CONCLUSION

For regional yield forecasting, both statistical and crop growth simulation models can be used safely. Although the CGSM has wider applicability and requires more number of input-parameters, for a particular region, the statistical model can be used safely. For preparation of statistical model, dummy variable must be used to incorporate qualitative factors. To run a CGSM, the required minimum data set are huge, which are not easily available in Indian sub-continent. Hence, for region-specific yield prediction, statistical model can be used as an alternative one.

## REFERANCES

1. Backpacker-South East Asia (2010): available at: http://www.southeastasiabackpacker.com /south-east-asia-a-rice-ingrained-culture Aug 7. Accessed 20 July 2014.
2. Banerjee, S. (2008). Possible impact of climate change on rice production in the Gangetic West Bengal, India, In: Unkovich MJ (ed) Global Issues Paddock Action, Proceedings of the 14th Australian Agronomy Conference, 21-25 Septembe, Adelaide, South Australia.
3. Banerjee, S., Das, S., Mukherjee, A., Mukherjee, A., Saikia, B. (2014). Adaption strategies to combat climate change effect on rice and mustard in Eastern India. Mitig Adapt Strateg Glob Change, doi: 10.1007/s11027-014-9595-y.
4. Boogaard, H. L., van Diepen, C.A., Rötter, R.P., Cabrera, J.M.C.A. and van Laar, H.H. (1998). WOFOST 7.1; user's guide for the WOFOST 7.1 crop growth simulation model and WOFOST Control Center 1.5.Wageningen (Netherlands), DLO Winand Staring Centre, and Los Banos (Philippines), Intrenational Rice Research Institute. Technical Document 52.
5. FAOSTAT (2005). FAOSTAT Agricultural Production (available at: www.faostat.fao.org/)
6. Fox, D.G. (1981) Judjing air quality model performance: a summary of the AMS workshop on dispension model performance. Bill Am Meteorol Soc, 62: 599-609. doi: http://dx.doi.org/10.1175/1520-0477.
7. Govt. of West Bengal (2002-2010). District Statistical Handbook (Nadia), Bureau of Applied Economics and Statistics, Government of West Bengal.
8. Hoogenboom, G., Wilkens, P. W., Thornton, P.K., Jones, J.W., Hunt, L.A., Imamura, D.T. (1999). Decision support system for agrotechnology transfer v3.5. In: Hoogenboom, G., Wilkens, P.W., Tsuji, G.Y. (Eds.), DSSAT version 3, vol. 4 (ISBN 1-886684-04-9). University of Hawaii, Honolulu, HI, pp. 1-36.
9. IRRI (2006). Bringing hope, improving lives: Strategic Plan 2007-2015, Manila 61 p.
10. Jones, J.W., Tsuji, G.Y., Hoogenboom, G., Hunt, L.A., Thornton, P.K., Wilkens, P.W., Imamura, D.T., Bowen, W.T., Singh, U. (1998). Decision support system for agrotechnology transfer; DSSAT v3. In: Tsuji GY, Hoogenboom G, Thornton PK. (Eds.), Understanding Options for Agricultural Production. Kluwer Academic Publishers, Dordrecht, the Netherlands, pp. 157-177.
11. Rahman, S.M., Md. Huq, M., Sumi, A., Mostafa, G.M., Azad, R.M. (2005). Statistical Analysis of Crop-Weather Regression Model for Forecasting Production Impact of Aus Rice in Bangladesh. International Journal of Statistical Sciences, 4: 57-77.

 ©2015 Society of Education, India