__Original Research Article__

Development of rice yield forecasting model using linear regression for Imphal west district, Manipur

Abstract: Rice is a staple food crop and also the principal food grain of India. Prevailing weather conditions during the crop growth period determine the yield of Rice. Hence, the crop yield forecasting models based on weather parameters will be an appropriate option for policymakers and researchers to develop sustainable cropping strategies. For the present study, Rice yield data and weather parameter during crop growing period (23rd to 45th SMW) for 23 years (1998-2020) were collected for the study area and 5 (five) models was developed using stepwise multiple linear regression. Analysis was carried out by fixing data from 1998-99 to 2016-17 for calibration and the remaining 2017-18 to 2020-21 data for validation. The models are developed for the Imphal West district of Manipur state. Accuracy of these models were estimated by R^2 (coefficient of determination) and the performance by MSE, RMSE and NRMSE.

Comment [HURS1]: Summary very restricted and unstructured. It would have to be expanded a little more and structured.

Comment [HURS2]: define these terms

Keywords: Rice yield prediction, weather parameters, statistical models, stepwise multiple linear regression,

Introduction

Comment [HURS3]: The introduction is fine. However, there is no background or theoretical framework.

Rice (*Oriza sativa*) is the principal food grain crop of India which occupies about 43.50 m ha area with the production of 104.32 million tons (Government of India, Ministry of Agriculture and Farmers Welfare: Department of Agriculture Cooperation, and Welfare 2016). As per the Directorate of Rice Development, Patna, Government of India, Manipur ranked 8th in the country with 2369 kg/ha in 2006-07 in rice yield,

Crop yield Prediction is important for agricultural planning and resource distribution decision-making. Efficient models were developed, to help reduce the error. Achieving maximum crop yield at minimum cost is one of the goals of agricultural production. Early detection and management of problems associated with crop yield indicators can help increase yield and subsequent profit. Multiple linear regression (MLR) is the standard and simplest approach for development of calibration models. But its application for dataset with independent variables greater sample number is not always successful (Balabin et al. 2011). However, feature selection in the form of stepwise MLR (SMLR) gives good results over large dataset. Stepwise regression procedure was adopted for selection of the best regression variable among many independent variables (Singh et al. 2014). Mallick *et al.* (2007) developed that regression model (linear, exponential and power regression) to forecast rice yield in Punjab, India. Correlations between rice yield and corresponding weather parameters were obtained for identified the sensitive period for rice yield in respect to different weather parameters. Parekh *et al.* (2012) revealed that crop yield is very much dependent on maximum & minimum temperature and relative humidity. The study determined the predominance of various meteorological data on yield of wheat. Crop yield is mostly affected by technological changes and weather variability. It can be assumed that the technological factors will increase yield smoothly through time and therefore, year or other parameters of time can be used to study the overall effect of technology on crop yield. Weather variable affect the crop differently during different stages of development. Thus, there is requirement to quantify the relationship between crop yield and weather variable in order to predict the regional yield, so that it may be useful for the farmers as well as the policymakers. In this present study, rice forecasting models were developed using SMLR for Imphal west district of Manipur, India.

Materials and Method

Manipur, the state of India, is located in the north-eastern part of the country. It is bordered by the Indian states of Nagaland to the north, Assam to the west, Mizoram to the southwest, and Myanmar (Burma) to the south and east. It is located at a longitude of 93°03'E to 94°78'E and a latitude of 23°56'N to 25°68'N. Manipur has a total geographical area of 22, 327 sq. km., out of which ninety percent (20, 089 sq. km.) is covered under hill districts and the remaining (2, 238 sq. km.) under valley districts. Temperature varies from sub 0 to 36 and its annual temperature falls between 20 to 25. The climate is temperate in the valley and cold in the hills. In summer the average high temperature is in the low 90s F (about 32–34 °C), while in the winter temperatures can drop into the mid-30s F (about 1–2 °C). Rainfall is abundant, with about 65 inches (1,650 mm) of precipitation occurring annually. The study was based on the Imphal West district of Manipur, India. (Fig. 1).

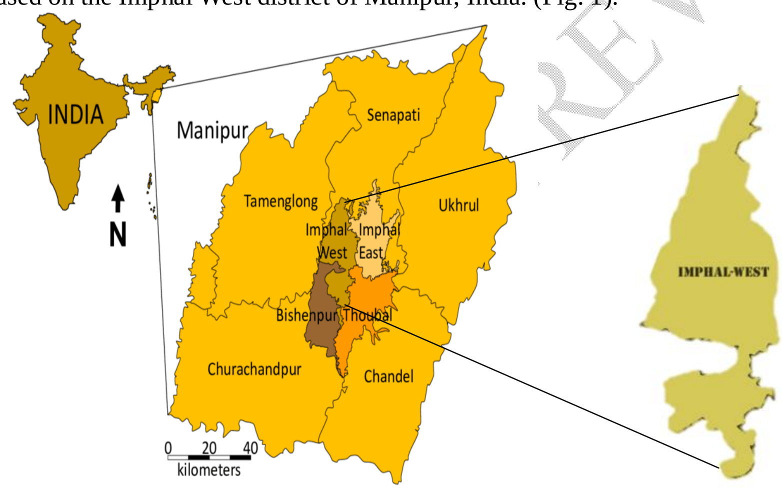

Figure 1. Study map

Development of rice yield forecasting model using Multiple Linear Regression.

Weather data viz maximum temperature, minimum temperature, rainfall, morning and evening relative humidity during crop growing period of rice (23rd SMW- 45th SMW) as well as rice yield data were collected for 23 years (1998-2020) from RMC, Guwahati and Department of Agriculture, Manipur. Analysis was carried out by fixing data from 1998-99 to 2016-17 for calibration and the remaining 2017-18 to 2020-21 data for validation. Stepwise multiple linear regression (SMLR) was used to developed the model with 115th weather variables, five (5) models were developed with the help of SPSS (Statistical Package for the Social Science). Accuracy of the forecasting models developed was estimated by the value of R^2 and performance of the developed models were estimated by calculating MSE, RMSE, and NRMSE, formula for calculating the statistical methods are as follows:

Mean Square Error (MSE) measures the amount of error in statistical models. $MSE = \frac{\sum(y_i - \hat{y}_i)^2}{n}$, where y_i is the i^{th} observed value, $\hat{y}_i$ is the corresponding predicted value and n = the number of observations.

Root Mean Square Error (RMSE) is used as a measure of comparing different models. $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - M_i)^2}$, where O_i is the observed value and M_i is the estimated value.

Normalized root mean square error (NRMSE) relates the RMSE to the observed range of the variable.

$nRMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - M_i)} \times \frac{100}{\bar{O}}$, where, $\bar{O}$ is the average of observation value.

Results and Discussion

Comment [HURS4]: The results shown are limited and some of them are not clear how they were obtained.

Yield prediction models for rice crop have been developed using long term crop yield data as well as long period daily weather data during crop growing period (23^{rd} to 45^{th} standard meteorological week). The coefficient of determination (R^2) was significant at 1% probability level for all the locations. Performance of models was categorized based on value of MSE, RMSE and NRMSE during validation and are presented in Table 1.

Models	Equation	R^2
Model 1	Y= 3.013 + 0.007 * X1	0.271
Model 2	Y= 6.721 + 0.010* X1 – 0.130* X2	0.694
Model 3	Y= 6909+ 0.010* X1-0.131 * X2 -0.006 * X3	0.864
Model 4	Y= 7.963+ 0.011 * X1 – 0.128 * X2 -0.007 * X3 – 0.041 * X4	0.920
Model 5	Y= 6.690 + 0.011 * X1 – 0.139 * X2 – 0.004 * X3 – 0.063 * X4 + 0.076 * X5	0.948

Table 1. Developed yield forecasting models for Imphal west district.

where, X1= rainfall of 43^{rd} week after sowing, X2= maximum temperature of 23^{rd} week after sowing, X3= rainfall of 35^{th} week after sowing, X4= maximum temperat000ure of 33^{rd} week after sowing, X5=maximum temperature of 35^{th} week after sowing, R^2 = Coefficient of determination, SE = Standard Error of the estimate.

Table 2. Performance of the developed models during calibration and validation time.

Models	During calibration	During validation

	MSE	RMSE	NRMSE	MSE	RMSE	NRMSE
Model 1	0.003	0.06	0.02	0.11	0.32	0.09
Model 2	0.002	0.05	0.02	0.07	0.27	0.07
Model 3	0.003	0.05	0.02	0.07	0.27	0.07
Model 4	0.003	0.05	0.02	0.09	0.30	0.08
Model 5	0.001	0.04	0.01	0.03	0.17	0.05

The weather parameters which were identified to be important through stepwise multiple linear regression were rainfall of 43rd week, maximum temperature of 23rd week, rainfall of 35th week, maximum temperature of 33rd week and maximum temperature of 35th week after sowing. The most accurate model out of all the developed models was Model-5 with R^2 (0.948), during calibration MSE (0.001), RMSE (0.04) and NRMSE (0.01), while the performance of the most accurate developed model during validation was MSE= 0.09, RMSE = 0.17 and NRMSE = 0.05, the least accurate model was Model-1 with only one weather parameter which is rainfall of 43rd week after sowing, and its R^2 value (0.271). Model 2 and 3 have R^2 (0.694; 0.864) respectively and the performance was quite the same with MSE = 0.07, RMSE= 0.27 and NRMSE = 0.07 approximately during validation and during calibration MSE (0.003), RMSE (0.05) and NRMSE (0.02). Model-4 with R^2 (0.920) also had the same values of MSE, RMSE and NRMSE during calibration but during validation it had MSE =0.09, RMSE= 0.3 and NRMSE = 0.08. The graphical representation of the comparison of different models are in Figure 2. During validation time period (2017-18 to 2020-21), the estimated yield for the most accurate model i.e., Model-5 were 2017-18 (3.24t/ha), 2018-19 (3.14 t/ha), 2019-20 (3.69 t/ha) and 2020-21 (3.44 t/ha), Table 3.

Table 3. Rice yields estimated by different models during validation for Imphal west.

Year	Actual yield	Estimated yield 1	Estimated yield 2	Estimated yield 3	Estimated yield 4	Estimated yield 5
2017-18	3.30	3.32	3.53	3.42	3.44	3.24
2018-19	3.19	3.08	3.31	3.20	3.12	3.14
2019-20	3.01	4.25	3.99	4.03	4.13	3.69
2020-21	3.47	3.84	3.85	3.79	3.83	3.44

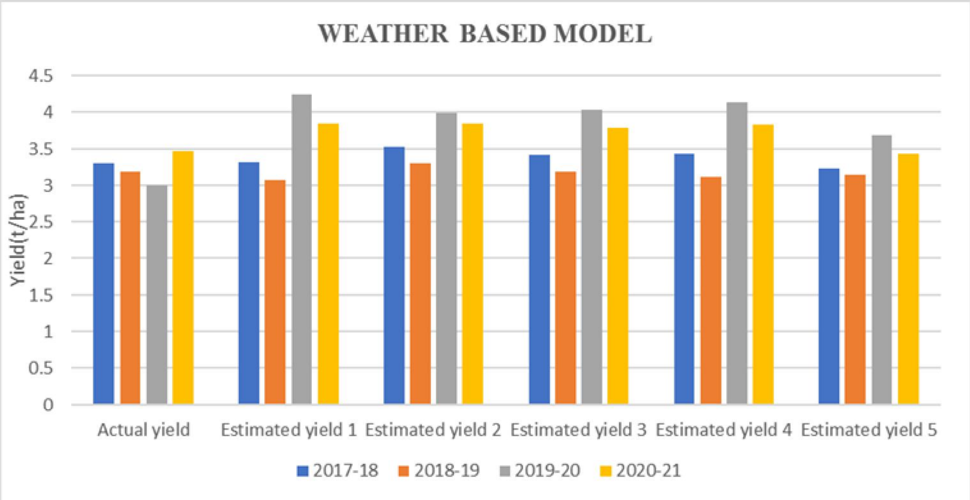

Figure 2. Comparison of different developed models for Imphal west district.

Conclusion

In the present study, five models were developed by using Stepwise multiple linear regression in SPSS software with long term weather data and yield data for 23 years (1998-99 to 2020-21) for Imphal west district, Manipur. The models were compared and the most accurate model was chosen by R^2 value nearest to 1, which was Model-5 which has R^2 value (0.948), the performance of the model during validation were MSE = 0.11, RMSE= 0.33 and NRMSE = 0.09. Hence, SMLR can be used for developing forecasting model for future rice yield prediction.

REFERENCES

Balabin, R. M., Lomakina, E. I., & Safieva, R. Z. (2011). Neural network (ANN) approach to biodiesel analysis: analysis of biodiesel density, kinematic viscosity, methanol and water contents using near infrared (NIR) spectroscopy. *Fuel*, *90*(5), 2007-2015.

Bhattacharyya, B., Biswas, R., Sujatha, K., & Chiphang, D. Y. (2021). Linear Regression Model to Study the Effects of Weather Variables on Crop Yield in Manipur State. *Int. J. Agricult. Stat. Sci*, *17*(1), 317-320.

Cai, Q., Wang, W. and Wang, S. (2015). Multiple regression model based on weather factors for predicting the heat load of a district heating system in Dalian, China a case study. *The Open Cybernetics & Systemics Journal*, 9, 2755-2773.

Das, B., Nair, B., Reddy, V. K., & Venkatesh, P. (2018). Evaluation of multiple linear, neural network and penalised regression models for prediction of rice yield based on weather parameters for west coast of India. *International journal of biometeorology*, *62*(10), 1809-1822.

Han X, Chang L, Wang N, Kong W, Wang C (2023). Effects of Meteorological Factors on Apple Yield Based on Multilinear Regression Analysis: A Case Study of Yantai Area, China. *Atmosphere*. 2023; 14(1):183. https://doi.org/10.3390/atmos14010183

Jain, R. C., Agrawal, R., & Jha, M. P. (1980). Effect of climatic variables on rice yield and its forecast. *Mausam*, *31*(4), 591-596.

K.S. Aravind, Ananta Vashisth, Krishnan and B. Das (2022). Wheat yield prediction based on weather parameters using multiple linear, neural network and penalized regression model. *Journal of Agrometeorology*., (2022) 24(1): 18-25.

Mallick, K., Mukherjee, J., Bal, S. K., Bhalla, S. S., & Hundal, S. S. (2007). Real time rice yield forecasting over Central Punjab region using crop weather regression model. *Journal of Agrometeorology*, *9*(2), 158-166.

Matsumura, K., Gaitan, C. F., Sugimoto, K., Cannon, A. J., & Hsieh, W. W. (2015). Maize yield forecasting by linear regression and artificial neural networks in Jilin, China. *The Journal of Agricultural Science*, *153*(3), 399-410.

Nain, G., Bhardwaj, N., Jaslam, P. M., & Dagar, C. S. (2021). Rice yield forecasting using agro-meteorological variables: A multivariate approach. *Journal of Agrometeorology*, *23*(1), 100-105.

Parekh, F. P., & Suryanarayana, T. M. V. (2012). Impact of climatological parameters on yield of wheat using neural network fitting. *International Journal of Modern Engineering Research*, *2*(5), 3534-3537

Piekutowska M, Niedbała G, Piskier T, Lenartowicz T, Pilarski K, Wojciechowski T, Pilarska AA, Czechowska-Kosacka A. (2021). The Application of Multiple Linear Regression and Artificial Neural Network Models for Yield Prediction of Very Early Potato Cultivars before Harvest. *Agronomy*. 2021; 11(5):885

Sridhara, S., Ramesh, N., Gopakkali, P., Das, B., Venkatappa, S. D., Sanjivaiah, S. H., ... & Elansary, H. O. (2020). Weather-based neural network, stepwise linear and sparse regression approach for rabi sorghum yield forecasting of Karnataka, India. *Agronomy*, *10*(11), 1645.

Rutkoski, J. E., Poland, J. A., Singh, R. P., Huerta□Espino, J., Bhavani, S., Barbier, H., ... & Sorrells, M. E. (2014). Genomic selection for quantitative adult plant stem rust resistance in wheat. *The plant genome*, *7*(3), plantgenome2014-02.

Vanitha, G., Kennedy, J. S., Prabhu, R., & Rajkishore, S. K. (2021). Trained neural network to predict paddy yield for various input parameters in Tamil Nadu, India. *Journal of Applied and Natural Science*, *13*(SI), 135-141.