



Empirical Analysis for Crop Yield Forecasting in India

S. Dharmaraja¹ · Vidyottama Jain² · Priyanka Anjoy³ · Hukum Chandra³

Received: 4 July 2017 / Accepted: 27 April 2019 / Published online: 18 May 2019
© NAAS (National Academy of Agricultural Sciences) 2019

Abstract Several factors, including weather vagaries, possess a serious threat to agricultural crop production in India and also are noteworthy risks to the economy. Crop yield depends on nutrition level of soils, fertilizer availability and cost, pest control, agro-meteorological input parameters like temperature, rainfall and other factors. Further, each particular crop needs specific growing weather conditions. Therefore, prognosticating crop yield is a challenging task for every nation. Statistical models are the most commonly used tools to forecast the crop yield, whereas statistical forecasting model for predicting dynamic behavior of crop yield should be able to take advantage not only of historical data of crop yield, but also the impact of various driving forces of the external environment. This paper describes both the linear regression and time-series models to predict crop yield efficiently and precisely. In particular, Bajra yield data for Alwar district of Rajasthan have been considered for empirical fitting of the models. Additionally, the selection of auxiliary variables, based on the knowledge of crop growth stages, has mediated the outperformance of time-series model.

Keywords Crop yield · Crop growth stages · Regression · Time series · Environment · Prediction

Introduction

Agriculture is the key livelihood for the vast majority of population in India. Direct or indirect reliance on a large proportion of farmers, middlemen, private companies and public sectors in the crop venture makes this agriculture

sector a prime stone for development of the country. Agriculture is being seen as profitable by farmers only if a good crop year turns up with high yield, which basically invites remunerative prices. Therefore, beforehand prediction of crop yield is crucial for them for farm-level planning and for making strategic decisions regarding marketing, storage, etc. It is as much important to the public and private sectors in framing various policies with respect to domestic food supply, international food trade, export, imports. Thus, this sets a strong need for reliable and efficient forecasting methodologies for various aspects relating to agriculture and more importantly the crop yield.

Among earlier authors, Fisher [8] has explained the influences of weather on the yield of wheat by considering weekly rainfall; Baier [1] has formulated the crop weather model for yield prediction. Recently, remote sensing (RS)-based approach [2, 13] is continually gaining popularity; satellite data offer significant benefits to assessing agricultural yield and production throughout the cropping season due to their spatial coverage, temporal and spectral resolutions, their availability to users in a timely manner

✉ S. Dharmaraja
dharmar@maths.iitd.ac.in

Vidyottama Jain
vidyottama.jain@curaj.ac.in

Priyanka Anjoy
anjoypriyanka90@gmail.com

Hukum Chandra
hukum.chandra@icar.gov.in

¹ Department of Mathematics, IIT Delhi, New Delhi, India

² Department of Mathematics, Central University of Rajasthan, Ajmer, India

³ ICAR-Indian Agricultural Statistics Research Institute (IASRI), New Delhi, India

and affordability. But they have limitations too; they might not be suitable and profitable in developing countries and in stratified agricultural systems with very small farm sizes. Further, yield forecasting models based on plant characters, e.g., examining the color of the leaf to check plant's health, are also used in some instances where regular system of measuring plant physiological characters exists [12], whereas it is to be noted that statistical models are mostly adopted in prediction or forecasting purposes due to their relative simplicity in usage [14].

The most investigated statistical yield forecasting models are regression models [3, 7, 11] that offer flexibility in incorporating relevant auxiliary information generally through linear modeling framework; such covariates play an important role in explaining the variability of the response variable. A good amount of covariates with significant correlation is therefore crucial for accuracy of such kind models, but incorporating a large number of covariates becomes limited in terms of parsimony. The yield of any crop is mostly affected by technological change and weather variability. It can be assumed that the technological factors may increase yield smoothly through time, and weather is also having volatile effect on yield; therefore, time (in year) or some functions of time can be used to study the overall effect of technology and weather on crop yield. Time-series forecasting models based on historic data may be a suitable alternative for between- or within-year crop yield prediction [4]. When the past data are used as auxiliary, they may take care of the cumulative effect of existing technological factors, weather and all other relevant covariates.

This article attempts to focus on forecasting the yield of 'Bajra' (*Pennisetum glaucum*) crop through implementing appropriate statistical models, in particular regression and time-series models. Bajra or 'pearl millet' is a coarse grain crop and is considered to be the poor man's staple nourishment. Bajra is the fifth most important cereal crop in the country, whereas India is the largest producer of Bajra in the world. In India, the total area under Bajra is 7.89 million hectares, from where the total amount of the cereal produced is about 9.18 million tonnes [6]. The crop survives in adverse climates with high temperatures, drought, low fertility of soil and also low rainfall. Due to these characteristics of its growth requirements, Bajra has been successfully cultivated in areas where other cereal crops cannot survive.

In India, Bajra-growing belt mostly constitutes of Rajasthan, Maharashtra, Gujarat, Haryana and Uttar Pradesh, in which Rajasthan is the top Bajra-growing state. The weather condition in Rajasthan is challenging for vast cultivation of many crops including cereals mostly, but since millet can withstand climatic adversity, millet cultivation has covered a large proportion of the area in

Rajasthan. Rajasthan shares 56.23% (4.43 million hectares) of total area under Bajra with 41.40% (3.80 million tonnes) of all India's production [6]. This also meets the staple food requirement of the state. In the years 2016–2017, Bajra yield (2149 kg/ha) of Alwar district was highest among all Bajra-growing districts of Rajasthan [9]. Hence, this paper focuses on capturing the yield variability of Bajra as well as its prediction, particularly for Alwar district of Rajasthan. The key content of the paper includes comparative assessment of various statistical methodologies within the purview of forecasting Bajra yield in Alwar district in Rajasthan.

The rest of the paper is organized as follows: Data sources and description are illustrated in "Data Collection and Sources" section. In "Results" section, we first describe the models considered in the analysis and it is followed by discussion about the empirical results from the model fitting. Finally, concluding remarks and further research scopes are summarized in "Conclusions" section.

Materials and Methods

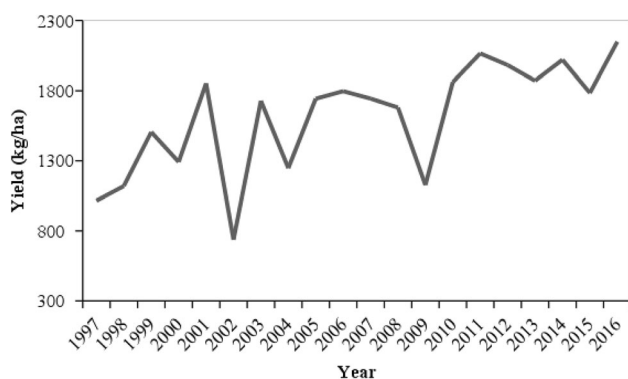
Data Collection and Sources

The most responsible factors for the variation in crop production are uncertainties in the weather conditions, depletion of nutrition level of soils, fertilizer availability and cost, pest control and other factors. It is very difficult to collect old data on soil nutrient levels, pest outbreak and various plant physiological characteristics, as there is no regular system of collecting and managing this information season-wise. Therefore, we restrict to two main weather variables, i.e., rainfall and average temperature, as predictors for crop yield.

In Rajasthan, pearl millet is grown in kharif season only. Alwar district mostly cultivates early- and medium-maturing Bajra varieties (e.g., HHB-67, HHB-94, ICMH-356, MH-169, HHB-60, RHB-30 and ICTP-8203), which take around 65–75 days (early) and 85–90 days (medium) for maturity. Yield data of kharif Bajra crop for Alwar district from 1997 to 2016 have been used in empirical analysis. The data have been collected from agricultural statistics at a glance (<http://eands.dacnet.nic.in>). Table 1 reports the descriptive statistics for these data. From 1997 to 2013, data have been kept for model building purpose and 2014–2016 data will be used for validating the regression and time-series models. For 2016, the fourth advance estimate of yield is available till date, still we have included that year in validation set, as little difference is usually observed between fourth advance estimates and revised final estimates of yield.

Table 1 Descriptive statistics for yield of kharif Bajra in Alwar district of Rajasthan

Measure	Yield (kg/ha)
Minimum	737
Mean	1616
Median	1744
Maximum	2149
Skewness	− 0.77
Kurtosis	− 0.38
SD	394
% Coefficient of variation	24

**Fig. 1** Trend in yield of Bajra in kharif season for Alwar district of Rajasthan during 1997–2016

Selection of auxiliary variables has been driven by the knowledge of crop growth stages. Bajra is having three major growth phases in its life cycle; we mark growth phases as GP-I, GP-II and GP-III which are defined as follows:

1. GP-I: vegetative phase (1–21 DAS),
2. GP-II: reproductive phase (22–42 DAS),
3. GP-III: grain filling and drying phase (43–80 DAS),

where the term DAS is the days after sowing. We have considered 80-day crop for simplifying covariate selection based on growth phases. Taking 1 July as the sowing date, we have collected daily rainfall data (in mm) and mean temperature (°C) data for 81 consecutive days of every year. Such weather data are collected for every year during 1997 to 2016. The data have been obtained from India water portal (http://www.indiawaterportal.org/met_data/). We assume that for a particular year R_j is the average rainfall at j th phase ($j = \text{GP-I, GP-II and GP-III}$) and T_j is the average temperature at j th phase. Thus, we denote six weather variables as R_1, R_2, R_3 and T_1, T_2, T_3 . Moist weather prevailing at early phases of crop growth is congenial to good yield along with the crop and does best under the conditions of light showers followed by bright

sunshine. Further, the inclusion of particular weather variables in the statistical models has been driven by this information on agro-climatic requirement along with valid statistical check. We have also considered ‘year’ as one of the auxiliary variables, because over the year yield data are presumed to follow some trends. We denote this auxiliary variable as X . Trend in yield of kharif Bajra from 1997 to 2016 is depicted in Fig. 1.

Results

Linear Regression Model

Linear regression method tries to establish a linear relationship between the response variable and predictor(s). Let y be the response variable which depends on m independent variables $X_1, \dots, X_m$. Then, the linear regression equation can be mathematically expressed as,

$$y = \beta_0 + \beta_1 X_1 + \dots + \beta_m X_m + \varepsilon \quad (1)$$

where β_i 's ($i = 1, \dots, m$) are unknown parameters and ε is the error term, usually assumed to be identically and independently distributed as $N(0, \sigma^2)$. Regression equation involving one auxiliary is called simple regression model, whereas multiple regression equation involves more than one covariate.

Implementing Linear Regression Model

For fitting simple regression model, we have not considered any of the weather variables as regressor, rather ‘year’ or X has been taken as predictor for simple regression model. As previously defined, the years in the range [1997, 2013] have been used only for model building purpose, and the estimated model is then considered for prediction in the validation set. For fitting multiple regression model, inclusion possibility of all the weather variables (R_1, R_2, R_3, T_1, T_2 and T_3) was also considered along with X . Based on ‘stepwise regression analysis,’ only R_1, R_2 and T_2 have been retained in the final model. The critical climatic requirement for Bajra crop comprises moist weather at early growth stage and bright sunshine during panicle development. Hence, inclusion of R_1, R_2 and T_2 as weather variables is significant from both statistical and agro-climatic point of view. Table 2 shows the parameter estimates of fitted simple and multiple regression models along with standard error as well as significance level. In the table, β_0 is the intercept; for simple regression equation, X is the only auxiliary variable, and for multiple regression equation, R_1, R_2 and T_2 are the auxiliary variables. Hence, β_1, β_2 and β_3 are the concerned slope parameters in the model. Further, we found that simple regression model is having

Table 2 Parameter estimates of fitted regression models

Model	Parameters	Estimate	Standard error	Probability
Simple regression	β_0	− 94,778	30,966	0.008
	β_1	48.04	15.44	0.007
Multiple regression	β_0	− 4894.12	1870.55	0.024
	β_1	188.94	63.91	0.013
	β_2	2.52	0.79	0.009
	β_3	2.01	0.81	0.031

adjusted R^2 value of 35% and Akaike information criterion (AIC) value of 273.74, whereas multiple regression model is having adjusted R^2 value of 71% and AIC value of 248.05.

Time-Series Model

Time-series analysis deals with observations that are sequentially collected over time. One distinguishing feature in this sequence is that values in the future depend on the past records in a quite stochastic manner, which is mandatory for depicting the future. The objective of the time-series analysis is to find out that or those kinds of models which would better reflect the underlying dynamics, and the same can be used for forecasting the series phenomenon. The most pioneered class of models that are frequently used to model linear dynamics structure is the autoregressive moving average (ARMA) models, which render the linear forecasting. The stationarity property of ARMA models reflects certain time-invariant properties of time series and is a necessary condition for making a statistical inference. However, real time-series data often exhibit time trend and/or cyclic features that are beyond the capacity of stationary ARMA structure. A convenient and effective way is to de-trend and de-seasonalize the data to remove those unstable components by taking the difference, after that the differenced data can be modeled with a stationary ARMA component. Because the original series is the integration of the differenced series, this is an autoregressive integrated moving average (ARIMA) process. A univariate process $\{y_t\}$ is said to follow ARIMA (p, d, q) model, if it can be represented as:

$$\phi_p(B)\nabla^d y_t = \theta_q(B)\varepsilon_t \quad (2)$$

where p = autoregressive order, q = moving average order, d = differencing order. $\phi_p(B)$ and $\theta_q(B)$ are the autoregressive (AR) and moving average (MA) polynomials, respectively.

$$\phi_p(B) = 1 - \sum_{i=1}^p \phi_i B^i = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p;$$

$$\theta_q(B) = \sum_{i=1}^q \theta_i B^i = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q;$$

$$\nabla^d = (1 - B)^d \text{ and } B(y_t) = y_{t-1}$$

$\{\varepsilon_t\}$ is assumed to be standard white noise process following normal distribution with mean zero and variance σ^2 . A constant term μ may also be added to the right-hand side of Eq. (2) if y_t is not mean-adjusted. The ARIMA methodology is followed in three stages, viz. identification, estimation and diagnostic checking. Parameters of the tentatively selected ARIMA model at the identification stage are estimated at the estimation stage, and adequacy of tentatively selected model is tested at the diagnostic checking stage. If the model is found to be inadequate, the three stages are repeated until satisfactory ARIMA model is selected for the time series under consideration. The ARIMA with exogenous variable (ARIMAX) model is a generalization of the ARIMA model, which is capable of incorporating an external input variable (X). Given a historic input vector x_t , the ARIMAX model assumes the form,

$$\phi_p(B)\nabla^d y_t = \mu + v(B)x_t + \theta_q(B)\varepsilon_t \quad (3)$$

where μ is the intercept term, p and q are, respectively, the AR and MA order of the model; $\phi_p(B)$ and $\theta_q(B)$ have the form defined for ARIMA model; $v(B)x_t$ is defined as,

$$v(B)x_t = \beta_1 x_t + \beta_2 x_{t-1} + \dots + \beta_m x_{t-m+1}$$

where m is the number of regressor in the model and $\{\varepsilon_t\}$ is assumed to follow standard white noise process with $N(0, \sigma^2)$.

Implementing ARIMA and ARIMAX Models

It is noteworthy that the time-series model fitted in this paper is based on 20 observations. However, in time-series literature the thumb rule is that at least 50 observations are required for fitting ARIMA model. But this thumb rule of availability of at least 50 observations is reasonable if there is seasonal or periodic effect in the data. Many other authors have pointed out the fact that specific thumb of minimum sample sizes is misleading and unsubstantiated in theory or practice. It also depends on the number of model parameters to be estimated and the amount of randomness/noise in the data; see, for example, Hyndman and Kostenko [10], Silver [15, 16]. In case of Bajra yield series used in this paper, there was no question regarding the presence of seasonality or periodic effects because it is annual time-series data.

Table 3 Testing for stationarity

ADF test	Test statistic	Probability
Level series	0.55	0.97
First differenced series	− 26.97	< 0.0001

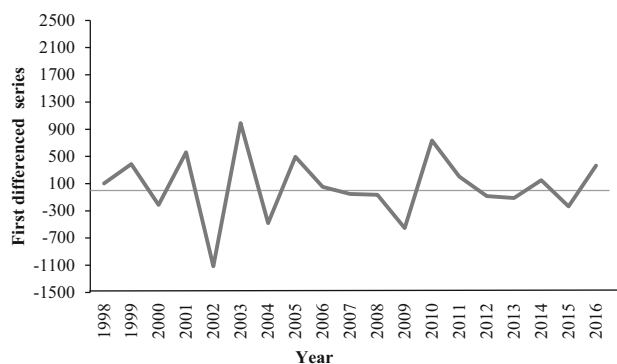
For the implementation of time-series models, it is first necessary to find whether the given time series is stationary or not. If the time series is non-stationary, the foremost task is to make it stationary and then apply usual models. The presence of non-stationarity or stationarity is needed to be confirmed before the application of either ARIMA or ARMA structures. For testing stationarity, several tests can be performed including ADF (augmented Dickey–Fuller), PP (Phillips–Perron), KPSS (Kwiatkowski–Phillips–Schmidt–Shin) test, etc. In this study, ADF test [5] has been implemented for testing the presence of unit root. The null hypothesis here is considered as presence of unit root or non-stationarity against stationarity. The result of ADF test is provided in Table 3 at the sixth lag. The level yield series is found to be non-stationary; after the first difference, the differenced series has become stationary. Now, this first differenced series ($d = 1$) is considered for implementing ARIMA and ARIMAX models. Figure 1 depicts the level yield series, and Fig. 2 presents the first differenced series of Bajra yield. Autocorrelation function (ACF) and partial autocorrelation function (PACF) provide a useful guide for possible order of the ARIMA and ARIMAX models. Minimum AIC value was used to select the best model from the tentatively chosen set of ARIMA models. Finally, ARIMA (1, 1, 0) model with AR order $p = 1$ was found to be the best model with AIC value 237.85. ARIMAX model extends the ARIMA structure with one or more exogenous variable. For ARIMAX model, inclusion of weather variable R_1 was only found to be significant. The best ARIMAX model has been retained based on minimum AIC value (AIC = 203.86) from the initially chosen set of ARIMAX models. The fitted ARIMAX forecast model to forecast the yield is given as:

$$\nabla \hat{y}_t = -0.68y_{t-1} + 2.5R_{1t}$$

where R_{1t} is the mean rainfall (in mm) in GP-I or vegetative phase in the year t and y_{t-1} is the yield in year $(t - 1)$. Table 4 presents the parameter estimates of ARIMA (1, 1, 0) and ARIMAX (1, 1, 0) models along with significance level.

Performance Evaluation of the Fitted Models

The article attempts to predict Bajra yield for Alwar district of Rajasthan implementing appropriate statistical models. Keeping this in view, simple and multiple regression

**Fig. 2** First differenced series of Bajra yield for Alwar district of Rajasthan

model, ARIMA model and ARIMAX models have been fitted on the training data sets (1997–2013); Tables 2 and 4, respectively, provide the parameter estimates of the discussed models. Now, our endeavor is to assess the performance of the models and finally to find out the best forecasting model. To pursue the objective, we use test data set (2014–2016) for validating the above models employing some of the validation criteria as follows,

Relative mean absolute error percentage (RMAE %)

$$= \left(\frac{1}{n} \right) \left[\sum_{h=1}^n |(\hat{y}_{t+h} - y_{t+h}) / y_{t+h}| \right] * 100$$

where $(\hat{y}_{t+h} - y_{t+h})$ is the difference between forecast and actual value for h th period ahead, termed as forecast error, and n is the amount of hold-out or test data set.

Relative mean squared prediction error (RMSPE)

$$= \left(\frac{1}{n} \right) \left[\sum_{h=1}^n (\hat{y}_{t+h} - y_{t+h})^2 / y_{t+h} \right]$$

Mean absolute error (MAE) = $\left(\frac{1}{n} \right) \left[\sum_{h=1}^n (\hat{y}_{t+h} - y_{t+h}) \right]$

Root mean squared error (RMSE)

$$= \sqrt{\left(\frac{1}{n} \right) \left[\sum_{h=1}^n (\hat{y}_{t+h} - y_{t+h})^2 \right]}$$

Model with the lowest RMAE (%), RMSPE, MAE and RMSE is considered to be the best model. Table 5 presents the actual versus predicted Bajra yield for the years 2014, 2015 and 2016. Table 6 reports the values of RMAE %, RMSPE and MAE, and Table 7 provides the values of RMSE ratio for validation data set. Comparative assessment of the models reveals that time-series forecasting models are having better prediction ability than the traditional regression models. Multiple regression model has proven to be better than simple regression

Table 4 Parameter estimates of fitted ARIMA and ARIMAX models

Model	Parameters	Estimate	Standard error	Probability
ARIMA	ϕ_1	− 0.70	0.19	0.003
ARIMAX	μ	− 426.15	180.03	0.037
	ϕ_1	− 0.68	0.19	0.005
	β_1	2.50	0.98	0.003

Table 5 Actual versus predicted yield value for validation set

Year	Actual	Simple regression	Multiple regression	ARIMA	ARIMAX
2014	2020	1454	1983	2046	2028
2015	1784	2186	2031	2020	1928
2016	2149	2506	2079	2135	2148

Table 6 Values of RMAE %, RMSPE and MAE for the regression and time-series models

Validation criterion	Simple regression	Multiple regression	ARIMA	ARIMAX
RMAE %	22.37	6.31	5.05	2.82
RMSPE	102.63	12.38	10.58	3.86
MAE	441.23	118.03	91.92	50.67

Table 7 Comparative results of the models in terms of RMSE ratio

RMSE ratio	Value	RMSE ratio	Value
Simple regression/ ARIMA	3.27	Multiple regression/ ARIMA	1.09
Simple regression/ ARIMAX	5.42	Multiple regression/ ARIMAX	1.80

model, whereas ARIMAX has been proven to be better than ARIMA. ARIMAX model is similar to a multiple regression model, but allows taking advantage of autocorrelation that may be presented in residuals of the regression to improve the accuracy of a forecast. Inclusion of weather variable as predictor has improved the performance of multiple regression model and ARIMAX model over the simple regression and ARIMA, respectively. However, ARIMA being the time-series model has shown better performance than the multiple regression model. Time-series model takes care of autocorrelation between past records in a quite stochastic manner, which mediates the task of depicting the future quite proficiently. Table 7 illustrates the relative improvement in gain while switching from regression to time-series approach as well as switching to include growth phase-wise significant weather variable(s) as a predictor.

Table 8 ACF, PACF and Box–Ljung test statistic of residuals for ARIMAX model

Lag	ACF	PACF	Box–Ljung test statistic	Probability
1	− 0.19	− 0.19	0.80	0.37
2	− 0.05	− 0.09	0.86	0.65
3	− 0.29	− 0.33	2.94	0.40
4	0.01	− 0.17	2.94	0.57
5	0.03	− 0.09	2.97	0.71
6	0.04	− 0.11	3.02	0.81
7	0.07	0.08	7.66	0.36
8	− 0.13	0.08	8.26	0.41
9	− 0.12	− 0.04	8.81	0.46
10	− 0.23	− 0.07	11.21	0.34

Discussion

The study comes out with ARIMAX as the best forecasting model for predicting kharif Bajra yield of Alwar. In terms of RMAE %, RMSPE, MAE and inference from RMSE ratio suggest that ARIMAX model can further be carried out for out-of-sample forecast. But before proceeding for the further step, it is necessary to check the assumptions of the residuals of the fitted model. The model verification is concerned with checking the residuals of the model to see

whether they contain any systematic pattern which still could be removed to improve the chosen model. This has been done through examining the autocorrelations and partial autocorrelations of the residuals at various lags, to ensure that ACF and PACF of the residuals were computed and Box–Ljung test was performed up to lag 10 as reported in Table 8. A very small value of ACF and PACF confirms the nonsignificant correlation or independence between the residuals of the fitted model. Box–Ljung test under the null hypothesis assumes that error is following white noise process. Nonsignificant result of Box–Ljung test statistics infers the correct specification of the ARIMAX model. Hence, we continue to implement ARIMAX model to forecast for 2017, i.e., one-step-ahead forecast. For the year 2017, predicted Bajra yield is 2152 kg/ha for Alwar district of Rajasthan.

Conclusions

The paper emphasizes on regression and time-series models to predict crop yield using historical data and attempts to determine the predictive accuracy in forecasting crop yield particularly when auxiliary variables are chosen based on the information of plant growth phases. Each crop has its particular agro-climatic requirement for producing optimum yield; therefore, incorporating this knowledge of climatic variability may definitely improve the predictive power of the model. Basically, time-series model has due capability of including past phenomenon in a stochastic manner, such approach has proven to be proficient enough in the present study. It is believed that based on predicted yield data farmers can take the necessary decisions on farm practices and improve their crop productivity in the future. The government may also have the idea for framing strategic decisions for the next year. Since the discussed regression and time-series methods are based on linear modeling framework, scope remains to explore nonlinear dynamics to predict crop yield more accurately.

Acknowledgments Authors are thankful to the editor and two anonymous reviewers for their valuable suggestions and comments which helped improve the manuscript to a great extent.

References

1. Baier W (1977) Cropweather models and their use in yield assessments. WMO Technical Note No. 151. World Meteorological Organization, Geneva, p 48
2. Basso B, Cammarano D, Carfagna E (2013) Review of crop yield forecasting methods and early warning systems. In: Proceedings of the first meeting of the scientific advisory committee of the global strategy to improve agricultural and rural statistics, FAO Headquarters, Rome, Italy, pp 18–19
3. Birthal PS, Khan MT, Negi DS, Aggarwal S (2014) Impact of climate change on yields of major food crops in India: implications for food security. *Agric Econ Res Rev* 27(2):145–155
4. Chattopadhyay C, Agrawal R, Kumar A et al (2011) Epidemiology and development of forecasting models for White rust of *Brassica juncea* in India. *Arch Phytopathol Plant Protect* 44:751–763
5. Dickey D, Fuller W (1979) Distribution of the estimators for autoregressive time series with a unit root. *J Am Stat Assoc* 74:427–431
6. Directorate of Economics and Statistics (2014) Agricultural Statistics at a Glance—2014. Ministry of Agriculture and Farmers Welfare, Government of India. <http://eands.dacnet.nic.in/PDF/Glance-2014.pdf>
7. Dkhar DK, Feroze SM, Singh R, Ray L (2017) Effect of rainfall variability on rice yield in north eastern hills of India: a case study. *Agric Res* 6(4):341–346
8. Fisher RA (1925) The influence of rainfall on the yield of wheat at Rothamsted. *Philos Trans R Soc Lond, Ser B* 213:89–142
9. Government of Rajasthan (2016) Crop-wise and district-wise fourth advanced estimates. Government of Rajasthan. <http://www.agriculture.rajasthan.gov.in/content/agriculture/hi/Agriculture/statistics.html>
10. Hyndman RJ, Kostenko AV (2007) Minimum sample size requirements for seasonal forecasting models. *Foresight Int J Appl Forecast* 6:12–15
11. Kumar SN, Aggarwal PK, Rani S, Jain S, Saxena R, Chauhan N (2011) Impact of climate change on crops productivity in Western Ghats, Coastal and North Eastern Regions of India. *Curr Sci* 101(3):332–341
12. Lobell DB, Gourdji SM (2012) The influence of climate change on global crop productivity. *Plant Physiol* 160(4):1686–1697
13. Prasada AK, Chai L, Singha RP, Kafatos M (2006) Crop yield estimation model for IOWA using remote sensing and surface parameters. *Int J Appl Earth Obs Geoinf* 8:26–33
14. Sellam V, Poovammal E (2016) Prediction of crop yield using regression analysis. *Indian J Sci Technol* 9(38):1–5
15. Silver N (2012) The signal and the noise: why most predictions fail—but some don't. HYPERLINK. Penguin Group, US. https://en.wikipedia.org/wiki/Penguin_Group
16. Silver, N. (2013). The signal and the noise: the art and science of prediction. HYPERLINK. Penguin Group, US. https://en.wikipedia.org/wiki/Penguin_Group

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.