Journal of Probability
and Statistical Science
21(1), 50-59, Feb. 2023

Rice crop growth analysis using Auto Regressive Models

Khadar Babu SK [1]
Vellore Institute of Technology.

Christophe Chesneau[2*]
LMNO, Universite de caen-Normandie.

Victor Anthonysmay [1]
Vellore Institute of Technology.

ABSTRACT
Time series play a vital role in predicting and forecasting different types of agricultural applications with respect to different types of problems among successive units of observations. Time series forecasting techniques are applied in all areas of statistics, and one of the most important applications includes backscatter generating time-series data using advanced forecasting techniques. Agriculture is a major food sector in the world, and it is also a major income source for low-income people. In this paper, we present two aspects of the rice crop growing time series process. The first one is to identify different types of rice crop growing stages for backscatter datasets, and the second is to make a mathematical time series model for the generation of different data sets. The different operator techniques (DOT) method was introduced to identify different types of rice crop growing stages in a season. We proposed the DOT method for identification of different phenological stages for a short-term crop and adopted first and second-order auto-regressive models for prediction and forecasting of the generating backscatter time series observations. The measures of the quality fit are mean absolute percent error (MAPE), mean percent error (MPE), and mean absolute error (MAE).

Keywords: Backscatter values, Auto Correlation, Linear regression

1. Introduction

Agriculture is a major food sector in India and around the world, and agricultural work supports most of the population. It provides about 18% of the GDP of India's economy. Rice crop production is a major income source for rural area people. Many farmers in the world usually grow three types of crops in a year. This is the summer (Kharif) from June/July to September, the winter (Rabi) crop from October to February/March (most of this crop season is affected by heavy rainfall) and the spring season from February/March to June/July. In the standard major crop seasons, there are three types of crops: single-cropped, double-cropped, and double-cropped rice, plus additional crops. The crop allotment at a place depends on location and weather conditions. (Devienne, 2006). Among all crops, rice is the major food crop and the major income source for different types of communities in India and the world. The backscatter values are generated by the scatterometer, which was fixed to the different satellites. SCATSAT-I is the Indian satellite, and it continuously monitors ground level observations with a temporal resolution of four days and a spectral resolution of four.45km.

□ Received October 2022, revised December 2022, in final form January 2023.
□ Christophe Chesneau (corresponding author) is affiliated with the Department of Mathematics, LMNO, Universite de caen-Normandie, Campus 2, Science 3, 14032 Caen, France, *christophe. chesneau@unicaen.fr,* Khadar Babu SK and Victor Anthonysamy are affiliated Department of Mathematics, Vellore Institute of Technology (VIT), Vellore, Tamil Nadu, India-632014. *khadar.babu36@gmail.com, avictorarun@gmail.com*

The current work focuses on different rice crop growing stages during the June–July–September (Kharif) season in 2019, in the East and West Godawari districts of Andrapradesh in India.

It is also focused on developing first and second order auto-regressive (AR) models for prediction and forecasting of the backscatter observations. And it is obviously useful to find missing and invalid observations in the dataset.

2. Literature Review

There are various methods existing for forecasting, so it is significant to know their applicability and reliability to make a suitable selection before using them in a specific situation. Time series modeling is an immensely popular method among researchers and specialists for providing exact forecasts. The main task of time series modeling is to study previous observations in order to develop a fitting model for the specific dataset and to use this model for forecasting upcoming values for the series. A time series is a set of observations on an activity collected at regular time intervals, such as daily, weekly, monthly, quarterly, annually, etc.

Brockwell and Davis (2002) used different types of time series models for prediction and forecasting of the data collected to afford specific methods for handling data. Popenda (1987) defined Δ for real valued function k(v) as the difference of two consecutive observation for constructing Alpha Difference tables ($\Delta \geq 1$).

Crop growth analysis and forecasting of crop yield are important tasks in supporting policymaking regarding land use allocation, food security, and environmental issues. Statistical techniques can provide crop forecasts with reasonable things well in advance. Box and Jenkins (1970) developed autoregressive integrated moving average models for the prediction and forecasting of time series data.

Oza et al. (2008) studied the performance of a study on rice crops using different types of Scattrometer datasets. Time series forecasting techniques are most suitable for rice crop production in the world and they give the most accurate forecasting discussed by Makridakisa and Hibbon (1979), Granger and Newbold (1983), Pankratz (1983), and Kumari et al. (2014), who applied exponential smoothing techniques to rice crop growth.

Sahu et al. (2015) studied modelling and forecasting of area, production, yield, and total seeds of rice and wheat in SAARC countries and the world towards food security, Mohammed Amir Hamjah et al. (2014) analyzed rice production forecasting in Bangladesh using the Box-Jenkins ARIMA Model. Muhammad Iqbal Ch et al. (2016) investigated forecasting of wheat production in a comparative study of Pakistan and India. Niaz Md. Farhat Rahman et al. (2013) investigated the modeling of pulse production growth and forecasting in Bangladesh. Vishwajith et al. (2014) developed time series modeling and forecasting of pulse production in India, Ashwin Darekar et al. (2017) presented Forecasting oilseed prices in India: The Case of Groundnut.

Pant et al. (2004) have discussed forecasting the wheat production in India using an ARIMA modelling approach. Studying the comparative economics of agro-processing units for groundnuts in Southern Rajasthan and the price behavior of groundnuts in Gujarat, they also applied the ARIMA Model prediction. That study aims to identify the best ARIMA model, which is for fitting and forecasting of rice crop area, yield, and production in the ceded region, respectively, and decisions are drawn to initiate the forecasting for the future.

Miah Mamun (2019) has identified the tentative Autoregressive models that accurately fit and forecast rice output in Bangladesh. Sunandini et al. (2020) discussed to look into the pattern of growth, the degree of instability in the area, production, and productivity of the rice crop in Andhra Pradesh. Abotaleb et al. (2021) used the Holt's Linear Trend (NNAR) model, and the AR and ARIMA models were used to forecast rice output. Wang et al. (2020) discussed the effects of assorted varieties on rice phenology as discovered through satellite-based observations.

Khadar Babu et al. (2022) have discussed the growth of the rice crop using various smoothing techniques. Mahesh et al. (2019) studied the identification of different crop stages using SCATSAT I Scatterometer data. The research gap between Mahesh Palakuru and the present proposed techniques for identification of different rice crop stages is to apply standard scientific difference operator methodology in place of manual identification.

3 Methodology

The difference operator technique (DOT) is a continuous data monitoring process that identifies different inflection points across the entire dataset. DOT has identified a growing increment for rice crop products in different time periods. Usually, the backscatter time series data depends on successive continuous monitored observations given by a Scatterometer, which is fixed to the human satellites. The present technique is useful for identifying different types of rice growing stages by using a DOT-based continuous monitoring system. The growing stages are: 1. the puddling stage, 2. the transplanting stage, 3. the heading stage, and 4. the harvesting stage. In the first stage, there is no vegetation and there is small incremental observation. At the second stage, the incremental values are growing slowly and have reached the heading stage. At the third stage, the growth of leaves and panicles drastic changes for a period and suddenly starts to change the color of the leaves as they reach the harvesting stage. The incremental data set follows a normal distribution with mean μ and standard deviation σ.

In the present research article, it is proposed to develop the procedure for identification of paddy crop stages using the difference operator technique (DOT) method:

$$\Delta_i = \frac{\partial}{\partial x}(\sigma_0),$$

where Δ_i is the difference index, σ_0 the reflective backscatter values.

and

$$\Delta\Delta_i = \frac{\partial}{\partial x}(\sigma_0'),$$

where ${\Delta^2}_i$ $(\Delta\Delta_i)$Second order difference index, σ_0' First derivative values of reflective backscatter values

In $\Delta\Delta_i$, identifies the various concavity shapes for providing a scientific inflection (or reflection) to identify an inflective titrate point in a dataset presented in a graphical approach. In fact, the graphical titrate gives some number of fractional incremental points and chooses an optimal incremental titrate point. The incremental titrate identified different rice crop growing stages in the Kharif season.

3.1 Parameter Methodology

The primary objective of the probability distribution is to estimate the parameter value using different types of statistical methods. In the present article, we adopt the Maximum Likelihood (ML) methods of estimation for parameters of the normal distribution. The normal distribution.

The normal distribution is written as follows.

Probability density function (PDF):

$$f(\xi) = \frac{1}{\sigma_0\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{\xi-k}{\sigma_0})^2}$$

Cumulative distribution function (CDF):

$$F(\xi) = \phi\left(\frac{\xi-k}{\sigma_0}\right) = \phi(p)$$

where ξ is backscatter, k is the mean, σ_0 is the standard deviation (SD).

Secondly, we implement the first and second order auto-regressive models for predicting and forecasting the data set. The methodology for implementing the model is described as follows.

For stationarity, adopted Auto Correlation and Partial ACF coefficients are introduced. For prediction and forecasting for the time series data consider Auto Regression (AR) models.

3.2 Auto Regression

The considered auto regression has the following form:

$$Z_t = b_0 + b_1 Y_{t-1} + b_2 Y_{t-2} + \cdots + b_k Y_{t-k} \quad . \tag{1}$$

The above model shows that regressing Z_t against Z_{t-1}, Z_{t-2}, ……. Z_{t-k}, which are explanatory variables to the previous value of the forecast variable Z_t. the name auto regression is used to explain the equation 1.

Khadar Babu and Ramanaiah (2013) studied auto-regressive models for discharge level time series datasets. The proposed methodology is adopted to obtain predictions and forecasts of time series data using auto-regressive models.

4 Data Analysis

4.1 Dataset

For statistical analysis, the data collected from MOSDAC site, and it is maintained by Space Application Centre (SAC), Ahmedabad, India, and tool take particularly for Kharif Season in the year 2019 in east Godavari district of Andhra Pradesh, India.

4.2 Proposed System

The prediction and forecasting of the data at any stage using the auto Regression (AR) methodology Actually, the complete data are classified into four stages, and these are:

(i) Puddling Stage, (ii) Transplanting Stage, (iii) Heading Stage, (iv) Harvesting Stage.

But among all the stages, mostly apply prediction and forecasting models to the heading stage only. One of the main objectives is to get more output from the crop. There is a need to estimate the growth of panicles in the rice crop. The growth of yield depends on different parameters, and it shows in the following different forms.

(i) Chlorophyll content, (ii) Canopy water content, (iii) Brassbound of the panicles.

Many of the parameters affected the yield of the rice crop and immensely increased the output also. Here we adopted one of the model methodologies to predict and forecast the growth of the rice crop at the heading stage only.

We have determined the data's slope and intercept from the backscatter value. Additionally, we can determine the auto-regressive model via ACF and PACF using Fig. 4. Hence, AR(1) and AR(2) are applied on this data with parameters.

AR(1) for the data is as follows:

$$Z_{t+1} = -0.09681 + 0.718422\, Z_t \tag{2}$$

AR(2) for the data is as follows:

$$Z_{t+1} = -0.09681 + 0.718422\, Z_t + 0.569599 Z_{t-1} \tag{3}$$

4.3 Identification of Phenological stages

From the methodology, the backscatter approach for the present study area has multiple inflection points, and it gives different stages of the rice crop's growing stages. The data follow the short-term rice crop season in the Godavari region backscatter paradigm (SCATSAT-I satellite data). The methodology shows that the feasible inflection points classify the different rice crop growing stages.

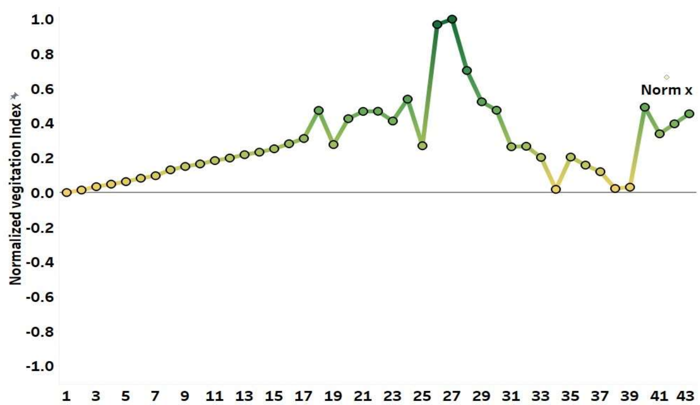

Fig. 1. The Normalized Value of the Backscatter data

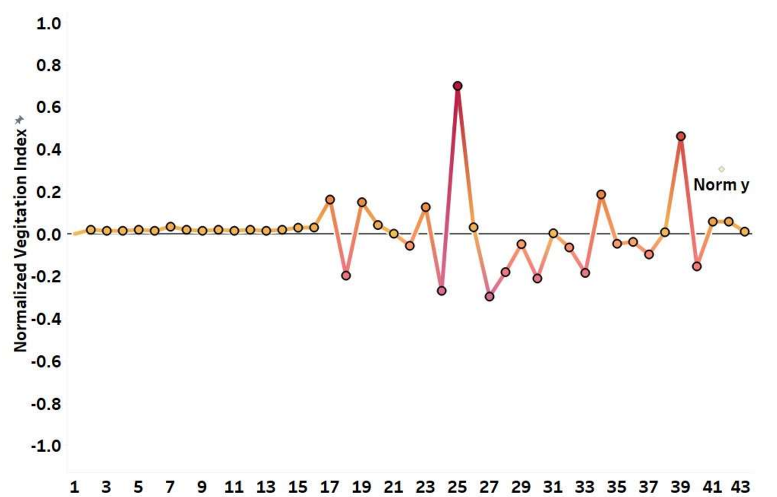

Fig. 2. First Iteration of the data

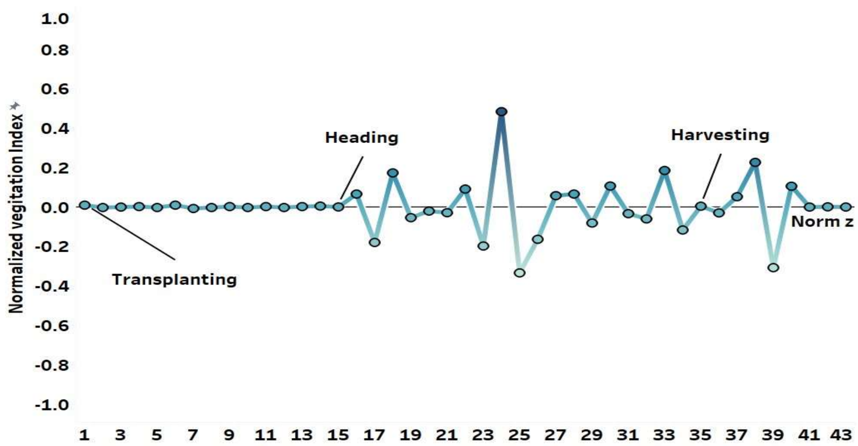

Fig. 3. Second Iteration of the data

In Figures 1, 2 and 3, the concept of concavity starts at the 15th day of the rice crop's growing period, then the heading starts at this point. Similarly, the remains were also identified on the 35th day of the rice crop's growing stage.

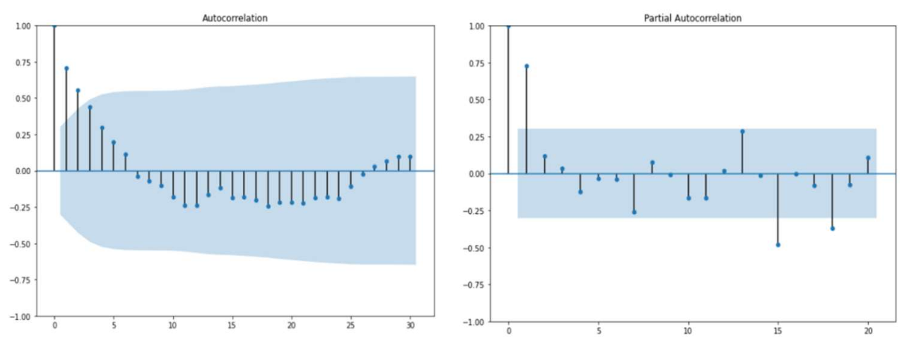

Fig. 4. Auto Correlation Function and Partial Auto Correlation Function

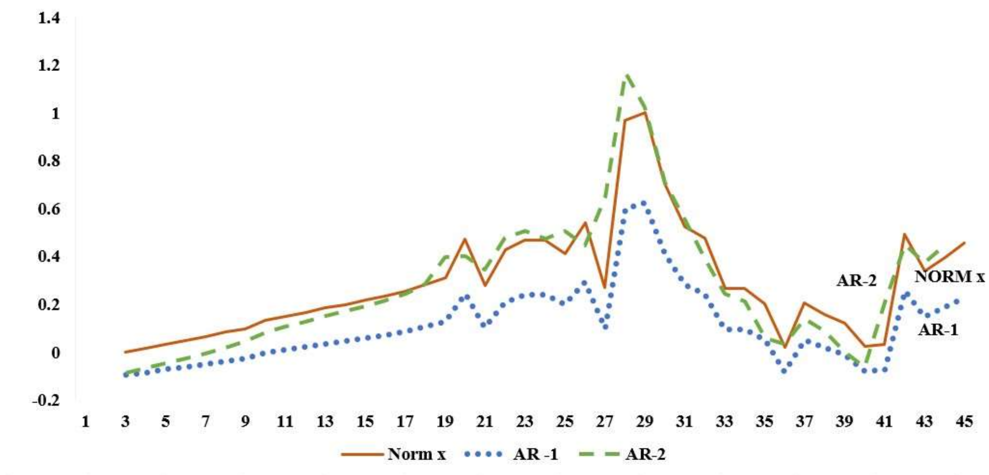

Fig. 5. Comparison curve between actual and AR1 and AR2

Table 1. Error Statistics

Error Statistics	ME	MSE	RMSE	MAE	MPE	MAPE	sMAPE	U_1	U_2
AR(1)	0.1792	0.0363	0.1905	0.1792	1.2353	1.2353	1.2047	0.3328	0.8466
AR(2)	0.0065	0.0088	0.0937	0.0673	0.2678	0.7215	0.5555	0.122	0.5761

Fig. 6. Error Comparison between AR(1) and AR(2)

4.4 Rice Growing Stages

The basic aspect of rice growth development is important for the study of backscattering responses of rice fields in the different rice growing stages. The rice crop's growing cycle usually takes 3-6 months due to different cropping systems and depending on the climate conditions. During the period, the rice growing stages have two major stages like vegetation and heading stages.

The heading stages are mainly subdivided into pre-heading and post-heading stages, known as the ripening period. In the Kharif Season, a 120-day variety, it takes 60 days in vegetation stages, 30 days in the heading stage as reproductive, and 30 days in the harvesting stage (Nguyen Lam-Dao, 2009). In general, the transplanting to the heading stage is called the vegetation stage. The vegetation stage is distinguished by the active increase in plant height and leaf enhancement at regular intervals. The growth of rice crop biomass and height stops after the heading stage and the leaves change their directions. The ripening period refers to the duration from heading to harvesting, which is widely affected by temperature.

5 Results and Discussions

The AR(1) and AR(2) models are good picks for predicting and forecasting data for rice crop backscatter datasets, and in some situations, it is definitely helpful to identify the missing places at any point in the data. For model fitting technology, the present paper demonstrates that for model fitting technology, the following standard statistical measures are useful to judge the fitness of the models.

5.1 Mean Absolute Percent Error (MAPE)

One of the procedures for choosing the best forecasting model with a minimum error value for method accuracy is MAPE. If the MAPE approaches zero, then we conclude that the fitted prediction and forecasting model has no bias. In the present study, the prediction and forecasting models were applied to rice crop backscatter values. MAPE for AR (1) is 1.2353. It shows that the model is the best fitted model for rice crop data. MAPE for AR (2) is 0.7214, and it is a small value compared with AR (1). Futuristic auto-regressive first order and auto-regressive second order models are also feasible to locate missing observations as well as obtain future prediction and forecasting scenarios.

Each of the time series plots generated through first and second order auto-regressive models for backscattered rice crop time series shown in Figure 3 shows that the sensors move more backscattered sensors at the heading stage only, which indicates that chlorophyll content appears at its maximum value on the 27th day of the crop period and in the second order auto-regressive model, almost just above the normal data.

5.2 Mean Percent Error (MPE) and Mean Absolute Error (MAE)

Therefore, if the MPE and MAE accuracy measures are also close to zero, we conclude that the fitted models are ideal for backscatter data sets. According to the results of the prediction and forecasting accuracy methodology, it can be concluded that first and second-order auto-regressive models are the best fitted models for backscattered time series data. The model produces MPE = 1.2353 for AR (1) and 0.2677 for AR (2) and MAE = 0.17915 for AR (1) and 0.06732 for AR (2). Both the accuracy measures are approaching zero, which means that the model has no bias. According to forecasting results, at the 27th day noted maximum backscatter value, it shows that more chlorophyll content is noted at the heading stage.

Results from different operating procedures for continuous monitoring systems suggest a better scientific approach to identify different rice crop growing stages for the Kharif season. Finally, by comparing with the ground truth observation, the classification of the different rice crop growing stages is 15 days, 20 days, and 8 days for the transplanting, heading, and harvesting stages, respectively. Almost ground truth observation also yields transplanting, heading, and harvesting times of 15, 18, and 10 days, respectively.

Acknowledgements

The data were collected as part of the Space Application Centre (SAC), ISRO R&D project titled "Enhanced Vegetation Monitoring Using RepidSCAT and SCATSAT-1 Scatterometer Data."The authors are thankful to the Space Application Centre for providing data and the Vellore Institute of Technology (VIT), Vellore for providing the facility for smooth running of the research.

References

[1]. Abotaleb, Mostafa & Ray, Soumik & Mishra, Pradeep & Karakaya, Kadir & Shoko, Claris & Al Khatib, Abdullah & Ray, Monika & Fernando, W & Lounis, Mohamed & Balloo, Ritisha. (2021). Modelling and forecasting of rice production in south Asian countries. Ama, Agricultural Mechanization in Asia, Africa & Latin America. 51. 1611-1627.

[2]. Ap Patel, G.N., and N.L. Agarwal, (1993), „"Price Behaviour of Groundnut in Gujarat", Indian Journal of Marketing, Vol.7, No.2, pp.50-57.

[3]. Ashwin Darekar, (2017). Forecasting oilseeds prices in India: Case of Groundnut, J.Oilseeds Res..,34(4):235-240.

[4]. Bhola Nath, DS, (2018). Forecasting Wheat production in India: An ARIMA modelling approach, Journal of Pharmacognosy and Phytochemistry, 8(1):2158-2165.

[5]. Box, G.E.P., Jenkins, G.M. and Reinsel, G.C.(1994). Time series analysis. Forecasting and control, Pearson Education, Delhi.

[6]. Brockwell, Peter J, and Richard A. Davis. (2002). Introduction to Time Series and Forecasting. New York: Springer.

[7]. Denning, G. L., & Vo, T. X. (1995). Vietnam and IRRI : a partnership in rice research : proceedings of a conference held in Hanoi, Vietnam, 4-7 May 1994. Manila, Philippines; Hanoi, Vietnam: International Rice Research Institute ; Ministry of Agriculture and Food Industry.

[8]. Devienne, S. (2006). Red River Delta: Fifty Years of Change. Moussons.

[9]. Drozdowicz, A., & Popenda, J. (1987). Asymptotic Behavior Of The Solutions Of The Second Order Difference Equation. DOI:10.1090/S0002-9939-1987-0866443-0

[10]. Kumari, P., Mishra, G., Pant, A., Shukla, G and Kujur, S. (2014). Comparison of Forecasting Ability of Different Statistical Models for Productivity of Rice (Oryza Sativa L.) in India. The Ecoscan., 8, 193-198

[11]. Makridakis S, Hibbon M. (1979). Accuracy of forecasting an empirical investigation, J Roy. Statist. Soc. A. 41(2):97-145

[12]. Miah, Mamun. (2019). Modeling and Forecasting Rice Production in Bangladesh: An Econometric Analysis. 2348-7909.

[13]. Mohammed Amir Hamjah (2014) for Rice Production Forecasting in Bangladesh: An Application of Box-Jenkins ARIMA Model, Mathematical Theory and Modelling, Vol. 4, No 4, 2014.

[14]. Muhammad Iqbal Ch (2016) for Forecasting of wheat production: A comparative study of Pakistan and India, IJAR 4(12), 698-709.

[15]. Newbold, P. (1983), ARIMA model building and the time series analysis approach to forecasting. J. Forecast., 2: 23-35. https://doi.org/10.1002/for.3980020104

[16]. Nguyen Lam-Dao, Thuy Le Toan, Armando Apan, Alexandre Bouvet, Frank Young, (2009). Effects of changing cultural practices on C-band SAR backscatter using Envisat ASAR data in the Mekong River Delta. Journal of applied remote sensing, Bellingham, WA : SPIE, 2009, 3, pp.033563. ⟨10.1117/1.3271046⟩. ⟨hal-00531685⟩

[17]. Niaz Md. Farhat Rahman et al. (2013), Modeling for Growth and Forecasting of pulse production in Bangladesh,

[18]. Oza, S.R., Panigrahy, S.andParihar, J.S.(2008). Concurrent use of active and passive microwave remote sensing data for monitoring of rice crop. Int. J. Appl. Earth Obs.Geoinform., 10(3): 296-304.

[19]. P.K. Sahu (2015) for Modelling and forecasting of area, production, yield and total seeds of Rice and Wheat in SAARC Countries and the World towards Food Security, SCIEP, American Journal of Applied Mathematics and Statistics, Vol.3, No.1,34-38.

[20]. Palakuru, M., Yarrakula, K., Chaube, N. R., Sk, K. B., & Rao, Y. S. (2019). Identification of paddy crop phenological parameters using dual polarized SCATSAT-1 (ISRO, India) scatterometer data. Environmental Science and Pollution Research, 26(2), 1565-1575.

[21]. Pankratz A. (1983). Forecasting with univariate box Jenkins models concept and cases, J.Wiley and sons, New York.

[22]. Pant, D.C. and Pradeep Pal, (2004). Comparative Economics of Agro-processing units for Groundnut in Southern Rajasthan, Indian Journal of Agricultural Marketing, Vol, 18(1).

[23]. SK, Khadar Babu, Christophe Chesneau, A. Victor, Suganya R, and V. Shakila, C. (2022). Efficacy of the Rice Crop Growth Using Different Smoothing Methods. Journal of Applied Analysis & Computation: 2022, 12(6): 2593-2599. doi: 10.11948/20220116

[24]. SK.Khadar Babu, M.V.Ramanaiah and V.Satish kumar (2013), Adaptable Ridge Regression Estimation for Sure model, Applied mathematical sciences, Vol.7(143), 7137-7141

[25]. Sunandini, Gurajapu & Paul, K. & Devi, Irugu. (2020). Analysis of Trends, Growth, and Instability in Rice Production in Andhra Pradesh. Current Journal of Applied Science and Technology. 39. 40-46. 10.9734/CJAST/2020/v39i4231129.

[26]. Vishwajith K..P. (2014). Timeseries Modeling and forecasting of pulses production in India, Journal crop and weed, 10 (2):147-154.

[27]. Wang, Hongfei & Ghosh, Aniruddha & Linquist, Bruce & Hijmans, Robert. (2020). Satellite-Based Observations Reveal Effects of Weather Variation on Rice Phenology. Remote Sensing. 12. 1522. 10.3390/rs12091522.