FORECASTING SUGARCANE YIELD OF INDIA BASED ON ROUGH SET COMBINATION APPROACH

Haresh Kumar Sharma[1], Kriti kumari [2] and Samarjit Kar [3]

[1] Department of Mathematics, Shree Guru Gobind Singh Tricentenary University, Gurugram, India
[2] Department of Mathematics, Banasthali Vidyapith, Jaipur, Rajasthan, India
[3] Department of Mathematics, National Institute of Technology Durgapur, West Bengal, India

Original scientific paper

Abstract: *This study applied a combination approach using rough set approach for forecasting sugarcane production in India. The rough set is a new mathematical approach that can deal with qualitative time series data. The method of combining forecast values based on rough set, the original time series and single forecasts obtained from single models are taken as condition and decision attributes. Finally, the decision table based on actual time series and single forecasting results are used to calculate the weights for the combination of forecasts. Moreover, dependency, importance and weights are also calculated for time series through condition and decision attributes. The paper uses autoregressive integrated moving average (ARIMA), double exponential smoothing (DES) and Grey model (GM) to generate the single forecasts. To validate our proposed analysis, Sugarcane production data from 1950 to 2011 was used for the overall empirical analysis and out-sample forecasts were generated from 2012 to 2021 for the comparative analysis. Also, ARIMA (2, 1, 1) model was found more appropriate for forecasting Sugarcane production.*

Key words: *Sugarcane, Forecast, time series models, Rough set combination.*

1. Introduction

India produces the largest amount of sugarcane and thus lands second on the list of top sugarcane-producing nations just after Brazil according to Foreign Agriculture Service (FSC) 2020. It reported Uttar Pradesh has the largest contribution amounting to 38.61% of the overall sugarcane production in the fiscal year 2020-21(ICAR-Sugarcane Report and Molasses Production, 2019). Then come Maharashtra and

---

* Corresponding author.
E-mail addresses: haresh_fosc@sgtuniversity.org (Haresh. Sharma), kriti.kri89@gmail.com (Kriti. Kumari), kar_s_k@yahoo.com (Samarjit. Kar)

Karnataka as the second and third largest sugarcane producing states. Some other contributors on the list are Bihar, Tamil Nadu, Haryana, Gujarat and Andhra Pradesh. Thus sugarcane being such a precious commodity it becomes really significant for the Indian economy to have highly reliable and accurate forecast of its productions.

There exists an effective relationship between the productivity and the price of the crop. With an unanticipated fall in the production, the market stock of the crop declines, thereby reducing the income of the farmer which is followed by the price rise as its consequence. Oppositely, if the market is flooded with the crop, it leads to a sudden fall in the prices and thus affecting the income of the farmer. Therefore, it can be concluded that these repercussions due to the variations in the prices of the commodity play a significant role in the formulation of the significant economic policies like inflation rate, GDP, wages, salaries etc. Apart from this it also affects the production level of other industries which further processes sugarcane and its by-products thereby affecting their profit margins.

From the last few years, various literature has been applied single time series models in the area of time series forecasting. ARIMA models are very popular to forecast sugarcane production (Bajpai and Venugopalan, 1996; Kumar and Madhu, 2014). For example, Venugopalan and Srinath (1998), Suresh at al. (2011) and Tsitsika et al. (2007) applied ARIMA models for modelling and forecasting of fish catches. Also, Hanson et al. (1999) compare the forecasting efficiency of neural network models with ARIMA models. ARIMA models were used to forecast the production and productivity of a variety of crops of Tamilnadu (Balanagammal et al. 2000). Boken (2000) used ARIMA models for the forecasting of wheat production in Pakistan and Canada. Balasubramanian and Dhanavanthan (2002) have applied ARIMA models to forecast seasonal paddy in Tamilnadu and food grains in India. Ravichandran and Prajneshu (2001) and Prajneshu et al. (2002) were compared the accuracy of structural time series models with ARIMA models. Maccioitta et al. (2002) use ARMA models to forecast milk, fat and protein yields of Italian Simmental cows. State level agricultural production forecasting was also done by applying ARIMA models (Indira and Datta, 2003). Also, Chandran and Prajneshu (2005) compare the forecasting performance of ARIMA models with nonparametric regression approach for the forecasting of oilseed production in India. Forecasting of irrigated crops like Potato, Mustard and Wheat were forecasted by employing ARIMA models (Sahu, 2006). Milk production in India was also predicted using different time series methods (Pal et al. 2007).

Also, there are different time series models, such as econometric, smoothing models and different combination approaches. In recent years, rough set (RS) approach has been widely used in combine forecasting approach. For example, Bao et al. (2006) employed a combination approach based on rough set theory to determine the weighting coefficient in predicting the future of electric power load from 1994-2000 in Zhejiang. Xiao et al. (2009) examined the forecasting of international trade in the Chongqing Municipality of China using a combined approach based on the rough set, which he goes on to compare with individual models. Ahmed et al. (2009) applied a combination of forecasts based on rough set. Suo et al. (2013) evaluated the weight coefficient by using rough set theory to combine the forecasts of the quadratic curve, Grey, and cubic exponent smooth models for forecasting agriculture machinery total power from the period of 1996 to 2008. They explain that rough set combination approach is higher than the individual forecasting methods. Zhou and Zhang (2013) employed rough set combination method by using support vector machine and neural network to predict the Chinese CO_2 emissions from 1990 to 2011. Sharma et al. (2019) proposed hybrid rough set based forecasting model and applied on tourism demand of air transportation passenger data set in Australia tourism demand. Tang et al.

(2021) applied hybrid fuzzy rough set models in missing traffic data. Patra and Barman (2021) employed rough set based dependency measure to reduce dimensionality of hyperspectral images. Ala'raj et al. (2021), proposed SEIRD dynamic model for forecasting of COVID-19 data and applied ARIMA correction model for validation of data set. For instance, Jahangir (2020) employed rough set based Artificial Neural Networks model to predict multimodal short-term wind speed. Li and Wang (2019) proposed hybridized NMGM-ARIMA and NMGM-BP models to forecast India's dependence on foreign oil. Sharma et al. (2018) applied rough set based forecasting methods in airline data. Wang et al. (2018) used Hybrid ARIMA and Metabolic Nonlinear Grey Model to Forecasting U.S. Shale Gas Monthly Production. Also Sharma et al. (2020) applied rough set theory for forecasting model's ranking. Rough set theory has been successfully applied to various real life decision making problems (Karavidić, and Projović, 2018; Roy et al., 2018; Roy et al., 2019; Sharma et al., 2018a; Stanković et al. 2019). Other soft computing approach use to tackle imprecision and vagueness of a data, which has been successfully applied to various real life problems (Karavidić and Projović, 2018; Žižović and Pamucar, 2019, Vasiljević et al. 2018; Mukhametzyanov and Pamucar, 2018). Moreover, Elgabbanni et al. (2014) applied rough set combination model (RSC) with an appropriate weight coefficient to forecast traffic accident time series data for Washington DC in the US from 1982-2008. They reveal that the combination method outperforms other individual methods. Additionally, the main concern in the combination of forecasts is that how to evaluate some appropriate weight coefficient to combine the forecasts of various single time series models. There have been various ways of determining the weight coefficient in the combination approach such as simple average, the inverse of MAPE, variance-based, the inverse of mean square error etc.

However, previous researchers have not been yet studied the rough set theory in sugarcane production literature, to the best of our knowledge. Hence, the main objective of our study is to forecast sugarcane production in India using a novel rough set combination (RSC) approach. The study aims to apply an appropriate way to combine the different single models to improve the forecasting accuracy of single time series models. ARIMA, DES and GM models have been combined by applying rough set theory to forecast sugarcane production in India for the period 1950 to 2011. We also study the comparative analysis of single time series and rough set combination methods by underlying mean absolute percentage error (MAPE) criterion.

The remaining of the study is organized as follows. A methodology section discusses the rough set theory. The next section illustrates the procedures of rough set combination method to the study of sugarcane production. Data section explains the data. Empirical results section describes the results of the empirical study, which includes time series models and their combination. Performance comparison of different models section presents the different performance criteria used in the forecasting comparison and the last section gives the conclusions.

2. Research Methodology

Rough set is a very useful classification technique for categorical variables like low, average and high. In this method, time series data has been arranged in an information table (decision table) with their objects (data points) by using a dependent and independent time series variables. Then, time series variables are transformed into condition and decision variables (attributes). Table 1 shows the hypothetical example

of decision table based on actual time series and single forecasts. These attributes are categorized into different grades like low, average or high and true or false etc. Hence, the applications of rough set are applied to generate the weights by establishing the relationships between single forecasts and actual time series. The method of combining forecast values based on rough set, the original time series and single forecasts obtained from single models are taken as condition and decision attributes. Finally, the decision table based on actual time series and single forecasting results is used to calculate the weights for the combination of forecasts.

Table 1. Hypothetical example of decision table is a table.

	Condition attributes					Decision attribute
Time	$\widehat{X_t(1)}$	$\widehat{X_t(2)}$	$\widehat{X_t(3)}$	...	$\widehat{X_t(N)}$	$y^{(t)}$
t_1	$\widehat{X_1(1)}$	$\widehat{X_1(2)}$	$\widehat{X_1(3)}$	...	$\widehat{X_1(N}$	$y^{(1)}$
t_2	$\widehat{X_2(1)}$	$\widehat{X_2(2)}$	$\widehat{X_2(3)}$	...	$\widehat{X_2(N}$	$y^{(2)}$
t_3	$\widehat{X_3(1)}$	$\widehat{X_3(2)}$	$\widehat{X_3(3)}$	...	$\widehat{X_3(N)}$	$y^{(3)}$
	...	...	...	...	...	
...	...	...	...	...	...	...
...	$\widehat{X_m(1)}$	$\widehat{X_m(2)}$	$\widehat{X_m(3)}$	...	$\widehat{X_m(N}$	...
t_m						$y^{(m)}$

In Table 1, $\widehat{X_t(1)}$, $\widehat{X_t(2)}$ $\widehat{X_t(3)}$, ... , $\widehat{X_t(N)}$ and $y^{(t)}$ are single forecasts called independent variable sand actual time series called dependent variables, respectively and $\widehat{X_m(1)}$, $\widehat{X_m(2)}$ $\widehat{X_m(3)}$, ... , $\widehat{X_m(N)}$ and $y^{(m)}$ are the objects (data points) of time series variables. Further, these variables are transformed into condition and decision attributes using normalization. The normalization is used to convert quantitative data into qualitative data. The normalization (Nt) technique is defined as: $N_t = \frac{Z_{kt}-Z_{min}}{Z_{max}-Z_{min}}$

where, Z_{kt} is the set of actual and single forecasts time series variables, $Zmin$ and $Zmax$ are the minimum and maximum values of Z_{kt}, k =1,2, ...

Further, actual and single forecasts are transformed into qualitative normalized values (NV) like low, average and high which are defined as; low(L) (IF 0 < NV <0.4), average(A)(IF 0.4 < NV≤0.8), high(H)(IF NV > 0.8).

3. Rough Set Theory (RST)

The rough set theory (RST) is a new mathematical technique to handle imprecision, vagueness, and uncertainty (Pawlak, 1982). For the evaluation of a vague description of the objects, it is the excellent mathematical tool. The adjective vague express the information quality that is uncertainty or ambiguity that chase from information granulation. The main aim of the rough set theory is the approximation of a set by a pair of two crisp sets called the lower and upper approximations of the sets.

Let U be the non-empty finite set of objects referred to as universe and A be a nonempty finite set of attributes, then $S = (U, A, C, D)$ is called an information system, where C and D are condition and decision attribute, respectively. For $S = (U, A, C, D)$ and P ⊆ A, R ⊆ U can be approximated based on the knowledge having in P by assembling the P-lower and P-upper approximation of R, represents by $\underline{P}(R)$ and $\overline{P}(R)$ respectively; where

$$\underline{P}(R) = \{x|[x]_P \subseteq R\} \tag{1}$$

$$\overline{P}(R) = \{x|[x]_P \cap R \neq \emptyset\} \tag{2}$$

The objects in $\underline{P}(R)$ is known as the set of all members of U which can be certainly classified as an object of R in the knowledge P whereas objects in $\overline{P}(R)$ is the set of all elements of U that can be possibly classified as an object of R involving knowledge P. The boundary region of R is expressed as: $BN_p(R) = \overline{P}(R) - \underline{P}(R)$ is the set of a member which cannot decisively classify into R consisting knowledge P. If lower approximation and upper approximation set are similar then boundary region of set is empty set. In the opposing case, if the boundary region having some member (object) than the set R is referred as rough set concerning P.

RST gives an accuracy measure on the quality of classification (Pawlak, 1982; Pawlak and Skowron, 2007). The quality of classification illustrates the ratio of all correctly classified objects of the data set, is calculated in the following manner:

$$\gamma_C(D) = \frac{|POS_C(D)|}{|U|}$$

Where, $|U|$ is the cardinality of the universal set (objects) and $|POS_C(D)|$ is the cardinality of union the of all lower approximation of D on C.

4. Autoregressive Integrated Moving Average (ARIMA)

Box-Jenkins (1976) introduced ARIMA model for modeling a time series with the trend and seasonal component. It is the combination of autoregressive (AR) and moving average (MA) models. ARIMA model for a time series, say ARIMA model for a time series, say $Xt(t = 1, 2. \ldots T)$, is given by $\varphi_p(B)\Delta^d X^t = \theta_q(B)a_t$, where

$\varphi_p(B) = 1 - \varphi_1(B) -, \ldots, -\varphi_p(B^p). and\ \theta_q(B) = 1 + \theta_1(B) + \cdots + \theta_q(B^p)$ are AR and MA models, respectively, B is the backshift operator, $\Delta^d\Delta^D_s X_t = (1 - B)(1 - B^s)X_{t,}\varphi_p < 1, \theta_q < 1$.

5. Grey Model

Grey model developed by Deng (1982). In this model, future trend is estimated using linear differential equation of order one. The parameters involved in the model can be estimated using the ordinary least squares (OLS) method Wang (2004) and Xu et al. (2016)). The Grey model of first order linear differential equation is written as $\frac{dX_t}{dt} + a * X_t = b$.

where X_t is a time series and a & b are the parameters.

6. Combination forecast based on rough set

Because the combination method yields better results than a single method, the modelling, and forecasting approach with high accuracy is adopted in this study. There are three main steps involved in the combined approach, i.e. single forecasts, computation of weight coefficient and forecast combination. Let: $X_t(t = 1, 2, ...,n)$ is an actual time series with time t and $X_{1t}, X_{2t}, X_{3t}, .., X_{jt}(j = 1, 2, .., m)$ respectively, are m_{th}, forecasting value of single forecasts at time t and $W_j(j = 1, 2, ..., m)$ is the m_{th} weight coefficient of single forecasts X_{mt} and then the combined forecasting approach can be written as follows: $\widehat{Z_{c,}}(t) = \frac{\sum_{i=1}^{m} W_i * X_{mt}}{\sum_{i=1}^{m} W_i}$

Where, $\widehat{Z_{c,}}(t)$ rough set combination forecasted values. Also, the overall procedure is described in Figure 1.
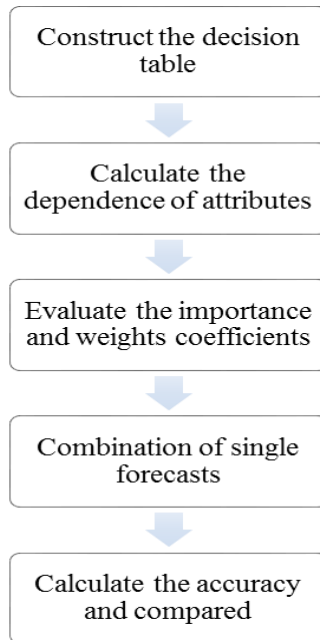
Figure 1. The framework of the proposed work.

6.1. Weight determination based on rough set

Let $S = (U, A, C, D)$ denote the rough set decision table, where U represents the universal set of time points in time series, and $C = X_{1t}, X_{2t}, X_{3t}, .., X_{jt}(j = 1, 2, .., m)$ indicate the set of single forecasts, condition attributes and D is the decision attribute, X_t in order to determine the weight coefficient. The overall procedures of deriving a weight coefficient are as described below1 as the several steps:

Step 1: Input the actual time series $X_t = (t = 1, 2, ..., n)$ and forecasts it with respective single forecasts, $X_{1t}, X_{2t}, X_{3t}, .., X_{jt}(j = 1, 2, .., m)$.

Step 2: To construct the rough set decision table, S by arranging the condition and decision attributes.

Step 3: Compute the dependence (Ahmed et al., 2009) of decision attribute (D) on the condition attributes (C).

Step 4: Evaluate the dependence of each attribute concerning D using the following expression. $\gamma_{C-\{C'\}}(D) = \frac{\left|POS_{C-\{C'\}}(D)\right|}{|U|}$

Where C' is the subset of C.

Step 5: Calculate the importance and weight coefficients of each single forecasts and then combine each single forecasts.

6.2. Data

Our empirical analysis uses yearly Sugarcane production from 1950 to 2011 time series data in India. Time series data are obtained from Sugar and Molasses Production (2019). The R-3.0.3 software is used for the overall empirical analysis of ARIMA, DES and Grey models. Also, the weight coefficient has been calculated for rough set combination method via Rough Set Data Explorer (ROSE2) software (Predki et al. 1998). The whole time series is divided into two periods (1) 1950-2011, in-sample, consists of 61 observations for the modelling process of the several methods; (2) data of 2012-2021 are used to generate the out-of-sample forecasts for different models.

6.3. Comparison criterion

The performance of all respective models for seasonal time series has been evaluated using mean absolute percentage error (MAPE) criterion for measuring level prediction accuracy, as follows:

$$MAPE = \frac{\sum_{t=1}^{n}(|Ac_t - \widehat{Pr_t}|)/Ac_t}{n} \tag{3}$$

Where $Ac_t(t = 1, 2, \dots, n)$ is the actual value, $\widehat{Pr_t}$ (t = 1, 2,..., n) represents the predicted values and n is the total number of observations. Lewis (1982) demonstrates that the value of MAPE being less than 10% indicates the high accuracy of forecasting. Moreover, when it lies between 10-20% forecasting is good, 20-50% is reasonable, and more than 50% denotes inaccuracy in forecasting.

7. Results

According to the forecasting results of every single model (see Table 1), we apply the discretization method to discrete data into three grades (0, 1, 2). The transformed discrete value, expressed as an attribute value is exhibited in Table 3. Consequently, decision Table 4 can be a build-up for the evaluation of weights $W_j(j = 1, 2, \dots, m)$ by using rough set in the combination forecast. Further, dependence and importance can be calculated through equation 1 and 2. In next step weights are computed by normalizing the importance of each single model. Finally, the combined forecasting model can be established as:

$$\widehat{X_t} = 0.2683 \; ARIMA + 0.4756 \; DES + 0.2561 \; GREY. \tag{4}$$

The evaluated results of dependence, importance and weights are given in Table 4. Moreover, we applied the actual and forecasts values from 1950 to 2011 for the evaluating the weight coefficient by using rough set. Table 2 gives the sugarcane production and forecasts obtained from ARIMA, DES and Grey models. Further, ARIMA, DES and Grey models are considered the three condition attributes and actual values is the decision attributes in order to establish the rough set theory and then

these attributes have been normalized (Yuan and Xu, 2013), such as shown in Table 1. Where X_{1t} represents the forecasts of the ARIMA model, X_{2t}, indicates the forecasts of DES model and X_{3t}, denote the forecasts of Grey model. Also, 2012-2018 out-of-sample forecasts were generated for different models.

Further simple average and inverse of MAPE combination methods (Bates and Granger 1969, Menezes et al. 2000, Armstrong 2001, Aiolfi and Timmermann 2006, Andrawis et al. 2011) are also employed for the prediction of sugarcane production in India. The forecasting results of ARIMA, DES and Grey models are combined using the weight coefficient $W_j(j = 1, 2, ..., m)$ obtained from simple average and the inverse of MAPE combination methods. According to the forecasting results of each single model weights are computed by the inverse of MAPE obtained from each single model. In the simple average method, each single forecast has equal weight. Finally, the combined forecasting model using simple average and inverse of MAPE methods can be established as:

$$\widehat{X_t} = 0.5\ ARIMA + 0.5\ DES + 0.5\ GREY. \tag{5}$$
$$\widehat{X_t} = 0.2309\ ARIMA + 0.2477\ DES + 0.1663 GREY. \tag{6}$$

Table 2. Actual and forecasts values of models

year	ARIMA	DES	GM	Actual
1952	59.69924	66.21	80.05650072	51
1953	44.34212	55.58	82.16917439	44.41
1954	49.33245	48.99	84.33760107	58.74
1955	68.02704	63.32	86.56325205	60.54
1956	51.24382	65.12	88.84763749	69.05
1957	71.69061	73.63	91.19230735	71.16
1958	66.22566	75.74	93.59885255	73.36
1959	73.1471	77.94	96.06890595	77.82
1960	78.06447	82.4	98.60414353	110
……	…………	………	………………	……
2002	302.09588	303.01	294.450717	281.58
2003	273.44244	286.16	302.2212075	233.87
2004	227.62409	238.45	310.1967595	237.09
2005	269.76906	241.67	318.3827847	281.18
2006	293.25304	285.76	326.7848373	355.52
2007	353.68756	360.1	335.4086182	348.19
2008	297.20754	352.77	344.2599788	285.03
2009	269.79121	289.61	353.3449249	292.31

Forecasting sugarcane yield of India based on rough set combination approach

Table 3. Decision table for rough set

U	ARIMA	DES	GM	Actual values
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0
6	0	0	0	0
7	0	0	0	0
8	0	0	0	0
9	0	0	0	0
……..	……..	……..	……..	……..
51	2	2	2	1
52	2	2	2	2
53	1	2	2	2
54	2	2	2	2
55	2	2	2	2
56	2	2	2	2
57	2	2	2	2
58	2	2	2	2

Table 4. Estimated parameters for models

| Models	Dependency	Importance	Weight
ARIMA	0.6207	0.3793	0.268284057
DES	0.3276	0.6724	0.47559768
GM	0.6379	0.3621	0.256118263

7.1. Discussion

7.1.1. Performance comparison of different models

The forecasting performance of different models has been evaluating using MAPE criterion. Table 5 reports the MAPE results for each of the individual models and rough set combination (RSC) method for the out-of-sample from the period of 2012 to 2018. Regarding MAPE values, the forecasting accuracy of RSC, $w_1 = 0.2685, w_2 = 0.4756$ and $w_3 = 0.2561$ is better than the individual forecasting methods. The combining method inverse of MAPE outperforms the ARIMA, DES, Grey, simple average and rough set combination based on MAPE (Lewis 1982) for out-of-sample forecasts. Also, the forecasting performance of ARIMA model is highly accurate than DES, Grey and combination methods. For better understanding Figure 2 compares the actual and forecasting values for each model based on out-of-sample forecasts. The

forecasting curves of all models are good fitted with the actual curve but the ARIMA suggest the best fit for the prediction of sugarcane production. All these confirmed that the forecasting results of IMAPE and ARIMA models are higher than the other single and combination time series models. Since ARIMA can forecast more accurately for the sugarcane production according to the out-of-sample forecasts. Consequently, ARIMA is used to predict the sugarcane production for the next three years from 2019 to 2021 with the forecasting results of DES, Grey, a simple average, IMAPE and rough set combination models. Results of forecasting using hybrid model (RSC model) are presented in Table 6. The obtained out-of-sample forecasts results of MAPE for the next three years are shown in Table 5.

Table 5. MAPE of forecasting models

Models	In-sample	Out-of-sample
ARIMA	7.8	0.85
DES	9.2	12.1
GM	13.8	7.85
SA	8.1	1.35
IMAPE	7.6	2.59
RSC	8.2	3.71
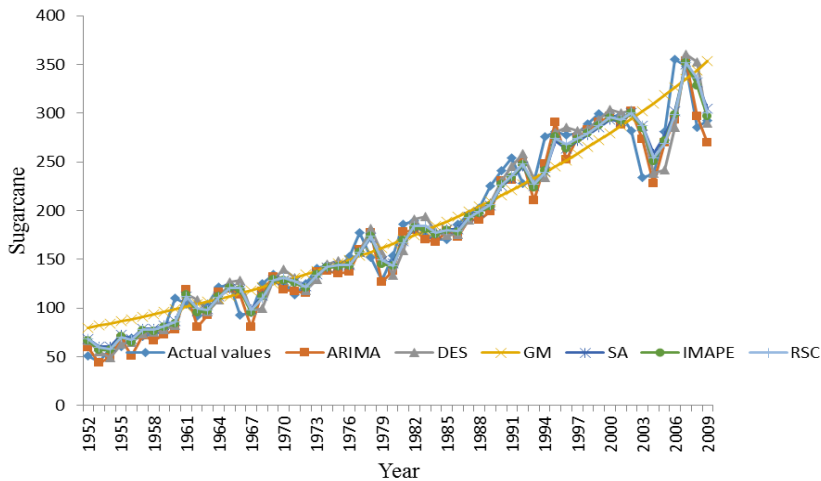
Figure 2. Comparative analysis of different models.

Table 6. Forecasts for the next ten years

Year	ARIMA	DES	GM	SA	IMAPE	RSC
2012	319.5522	306.05	382.0637366	335.888646	2485.14905	329.140
2013	305.7307	310.63	392.1463156	336.169005	2497.903017	330.193
2014	318.7837	315.21	402.4949718	345.496224	2564.058917	338.524
2015	331.7398	319.79	413.1167267	354.882176	2630.777165	346.898
2016	327.1773	324.37	424.0187875	358.522029	2663.934317	350.645
2017	317.3611	328.95	435.2085512	360.50655	2687.587675	353.055
2018	317.3513	333.53	446.6936104	365.858303	2731.273146	358.172
2019	323.7363	338.11	458.4817579	373.442686	2788.326094	365.083
2020	325.6805	342.69	470.5809919	379.650497	2837.53279	370.881
2021	322.1153	347.27	482.9995222	384.128274	2876.82159	375.284

8. Conclusions

Since single forecasting performs always well. Hence, this paper applied a novel combination of forecasts by underlying rough set (RS) approach for the prediction of sugarcane production to India for the period of 1950 to 2011 in order to improve the performance of single models. The forecasting results of autoregressive integrated moving average (ARIMA), DES and Grey models are combined using the weight coefficient obtained from simple average, the inverse of MAPE (IMAPE) and rough set combination methods. Moreover, the performance of several forecasts has been evaluated under mean absolute percentage error (MAPE) criterion.

Our empirical study suggests the following outcomes. First, all of the single forecasting models appeared to provide the accurate and reliable forecasting results according to the less than 10% MAPE values. Secondly, the ARIMA and IMAPE models have better accuracy than the other models according to MAPE values. Further, ARIMA performance is highly accurate among all different approaches. In addition, combination methods are found to be effective for the forecasting of sugarcane production in India.

The contribution of the article is that the combination of the forecast with the rough set approach firstly opts in agriculture empirical study. The obtained results suggest that ARIMA a combination method is an effective way for sugarcane forecasting. It is important to describe the importance of single model and dependency of sugarcane production for better forecasting performance. It is expected that future study would benefit from concentrating on other single methods for agriculture forecasting.

Conflicts of Interest*:* The authors reported no potential conflict of interest.

References

Ahmed, E. F., Yang, W. J., & Abdullah, M. Y. (2009). Novel method of the combination of forecasts based on rough sets, Journal of Computer Science, 440-444.

Aiolfi, M., & Timmermann, A. (2006). Persistence in forecasting performance and conditional combination strategies. Journal of Econometrics, 135, 31–53.

Ala'raj, M., Ajdalawieh, M., & Nizamuddin, N. (2021). Modeling and forecasting of COVID-19 using a hybrid dynamic model based on SEIRD with ARIMA corrections. Infectious Disease Modelling, 6, 98-111.

Andrawis, R. R., Atiya, A. F., & Shishiny, H. E. (2011). Combination of long term and short-term forecasts, with application to tourism demand forecasting. International Journal of Forecasting, 27, 870–886.

Armstrong, J. S. (2001). Principles of Forecasting: A Handbook for Researchers and Practitioners. Kluwer Academic Publishers. New York. (Chapter 13).

Bajpai, P. K., & Venugopalan, R. (1996). Forecasting sugarcane production by time series modeling. Indian Journal of Sugarcane Technology, 11(1), 61–65.

Balanagammal, D., Ranganathan, C. R., & Sundaresan, K. (2000). Forecasting of agricultural scenario in Tamilnadu—A time series analysis. Journal of Indian Society of Agricultural Statistics, 53(3), 273–286.

Balasubramanian, P., & Dhanavanthan, P. (2002). Seasonal modeling and forecasting of crop production. Statistics and Applications, 4(2), 107–118.

Bao, Y., Huang, M., Zheyan., Y. H., & Li, X. (2006). Application of combination forecasting based on rough sets theory on electric power system, Proceedings of the 6th Congress on Intelligent Control and Automation. June 21-23. Dallan, China: 1745-1748.

Bates, J. M., & Granger, C. (1969). The combination of forecasts. Operational Research Quarterly, 20, 451-468.

Boken, V. K. (2000). Forecasting spring wheat yield using time series analysis: A case study for the Canadian prairies. Agricultural Journal, 92(6), 1047–1053.

Box, G. E. P., & Jenkins, G. M. (1976). Time Series Analysis: Forecasting and Control, Revised Edition, San Francisco: Holden Day.

Chandran, K. P., & Prajneshu. (2005). Nonparametric regression with jump points methodology for describing country's oilseed yield data. Journal of Indian Society of Agricultural Statistics, 59(2), 126–130.

Deng, J. (1982). Control problems of Grey systems. Systems and Control Letters, 1(1), 288-294.

Elgabbanni, B. O. S., Khozium, M. O., & Ahmed, M. A. (2014). Combination prediction model of traffic accident using Rough Set technology approach. International Journal of Enhanced Research in Science Technology Engineering, 3(1), 47-56.

Hanson, J. V., Macdonald, J. B., & Nelson, R. D. (1999). Time series prediction with genetic algorithm designed neural networks: An empirical comparison with modern statistical models. Computational Intelligence, 15(3), 171–184.

ICAR-Sugarcane Report and Molasses Production (2019). https://sugarcane.icar.gov.in/index.php/en/sugar-stats/sugarcane-statistics. (Accessed 13 November 2020).

Indira, R., & Datta, A. (2003). Univariate forecasting of state-level agricultural production. Economic and Political Weekly, 38, 1800–1803.

Jahangir, H., Masoud Aliakbar G. M., Alhameli, F., Mazouz, A., Ahmadian, A., & Elkamel, A. (2020). Short-term wind speed forecasting framework based on stacked denoising auto-encoders with rough ANN. Sustainable Energy Technologies and Assessments, 38.

Karavidić, Z., & Projović, D. (2018). A multi-criteria decision-making (MCDM) model in the security forces operations based on rough sets. Decision Making: Applications in Management and Engineering, 1(1), 97-120.

Kumar, M., & Madhu, A. (2014). An application of time series ARIMA forecasting model for predicting sugarcane production in India Studies in Business and Economics. 9(1), 81-94.

Lewis, C. D. (1982). International and business forecasting methods. London: Butterworths.

Li, S., & Wang, Q. (2019). India's dependence on foreign oil will exceed 90% around 2025 - The forecasting results based on two hybridized NMGM-ARIMA and NMGM BP Models. Journal of Cleaner Production, 232, 137-153.

Maccioitta, N. P. P., Vicario, D., Pulina, G., & Cappio-Borlino, A. (2002). Test day and lactation yield predictions in Italian simmental cows by ARMA methods. Journal of Dairy Science, 85, 3107–3114.

Menezes, L. M. D., Bunn, D. W., & Taylor, J. W. (2000). Review of guidelines for the use of combined forecasts. European Journal of Operational Research, 120, 190-204.

Mukhametzyanov, I., & Pamucar, D. (2018). A sensitivity analysis in MCDM problems: A statistical approach. Decision Making: Applications in Management and Engineering, 1(2), 51-80.

Patra, S., & Barman, B. (2021). A novel dependency definition exploiting boundary samples in rough set theory for hyperspectral band selection. Applied Soft Computing, 99. 106944.

Pawlak, Z. & Skowron, A. (2007). Rudiments of rough sets, Information Sciences. An International Journal, 177(1), 3-27.

Pawlak, Z. (1982). Rough sets. International Journal of Computer and Information Science, 11, 341-356.

Predki, B., Wong, S. K. M., Stefanowski, J., Susmaga, R., & Wilk. S. (1998). ROSE-software implementation of the rough set theory. In L. Pollkowski, A. Skowron (Eds.).

Rough Sets and Current Trends in Computing. Lecture Notes in Artificial Intelligence. Berlin: Springer, 605-608.

Roy, J., Adhikary, K., Kar, S., & Pamucar, D. (2018). A rough strength relational DEMATEL model for analysing the key success factors of hospital service quality. Decision Making: Applications in Management and Engineering, 1(1), 121-142.

Roy, J., Sharma, H., Kar, S., Kazimieras, Z. E., & Saparauskas, J. (2019). An extended COPRAS model for multi-criteria decision-making problems and its application in web-based hotel evaluation and selection. Economic research – Ekonomska istraživanja, 32 (1), 219-253.

Sahu, P. K. (2006). Forecasting yield behavior of potato, mustard, rice, and wheat under irrigation. Journal of Vegetable Science, 12(1), 81–99.

Sharma, H. K., Kumari, K., & Kar, S. (2019). Short-term Forecasting of Air Passengers based on Hybrid Rough Set and Double Exponential Smoothing Models, Intelligent Automation and Soft Computing, 25(1), 1-13.

Sharma, H. K., Kumari, K., & Kar, S. (2020). A rough set approach for forecasting models. Decision Making: Applications in Management and Engineering, 3(1), 1-21.

Sharma, H. K., Kumari, K., Kar, S. (2018). Air passengers forecasting for Australian airline based on hybrid rough set approach. Journal of Applied Mathematics, Statistics and Informatics, 14(1), 5–18

Sharma, H., Roy, J., Kar, S. & Prentkovskis, O. (2018a). Multi-Criteria Evaluation Framework for Prioritizing Indian Railway Stations Using Modified Rough AHP-Mabac Method. Transport and Telecommunication Journal, 19(2), 113-127.

Suo, R., Huang, M., & Liu. Y. (2013). The application of combination forecasting method in total power of agriculture machinery based on RS. Advanced Materials Research, 601, 476 – 483.

Suresh, K. K., & Krishna, S. R. (2011). Forecasting Sugarcane Yield of Tamilnadu using ARIMA Models. Sugar Tech, 13(1), 23–26

Tang, J., Zhang, X., Yu, T., & Liu, F. (2021). Missing traffic data imputation considering approximate intervals: A hybrid structure integrating adaptive network-based inference and fuzzy rough set, Physica A: Statistical Mechanics and its Applications, In Press.

Vasiljević, M., Fazlollahtabar, H., Stević, Željko, & Vesković, S. (2018). A rough multicriteria approach for evaluation of the supplier criteria in automotive industry. Decision Making: Applications in Management and Engineering, 1(1), 82-96.

Wang, C. H. (2004). Predicting tourism demand using fuzzy time series and hybrid Grey theory. Tourism Management, 25 (3), 367-374.

Wang, Q., Li, S., Li, R., & Ma, M. (2018). Forecasting U.S. shale gas monthly production using a hybrid ARIMA and metabolic nonlinear grey model. Energy, 160, 378-387.

Xiao, Z., Gong, K., & Zoy, Y. (2009). A combined forecasting approach based on fuzzy soft sets. Journal of Computational and Applied Mathematics, 228, 326-333.

Xu, S., Wangshu, S., Jianzhou, W., Yixin, Z. & Yining, G. (2016). Using a Grey-Markov model optimized by Cuckoo search algorithm to forecast the annual foreign tourist arrivals to China. Tourism Management, 52, 369-379.

Yuan, L., & Xu, F. (2013). Research on the multiple combination weight based on rough set and clustering analysis, Procedia Computer Science, 17, 274 – 281.

Zhou, J., & Zhang, X. (2013). Combined forecasting model based on the rough set to predict the Chinese Co2 emissions, Advanced materials Research, 773, 831– 836.

Žižović, M., & Pamucar, D. (2019). New model for determining criteria weights: Level Based Weight Assessment (LBWA) model. Decision Making: Applications in Management and Engineering, 2(2), 126-137.