International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2023; SP-8(5): 682-686

https://www.mathsjournal.com
Received: 11-08-2023
Accepted: 16-09-2023

Tokesh Sahu
Department of Agricultural Statistics, Indira Gandhi Krishi Vishwavidyalaya, Raipur, Chhattisgarh, India

Neelam Chouksey
Department of Agricultural Statistics, Indira Gandhi Krishi Vishwavidyalaya, Raipur, Chhattisgarh, India

Sandeep Kumar
Research Scholar, Department of Agricultural Statistics, College of Agriculture, Acharya Narendra Deva University of Agriculture and Technology (ANDUAT), Kumarganj, Ayodhya, Uttar Pradesh, India

Shiv Kumar Rana
Research Scholar, Department of Agricultural Statistics, College of Agriculture, Acharya Narendra Deva University of Agriculture and Technology (ANDUAT), Kumarganj, Ayodhya, Uttar Pradesh, India

Saurabh Kanauajia
Research Scholar, Department of Agro-Forestry, College of Horticulture & Forestry, Acharya Narendra Deva University of Agriculture and Technology (ANDUAT), Kumarganj, Ayodhya, Uttar Pradesh, India

Randhir Yadav
Research Scholar, Department of Agricultural Economics, College of Agriculture, Acharya Narendra Deva University of Agriculture and Technology (ANDUAT), Kumarganj, Ayodhya, Uttar Pradesh, India

Corresponding Author:
Sandeep Kumar
Research Scholar, Department of Agricultural Statistics, College of Agriculture, Acharya Narendra Deva University of Agriculture and Technology (ANDUAT), Kumarganj, Ayodhya, Uttar Pradesh, India

Forecasting of honey bee population by ARIMAX model using weather variables

Tokesh Sahu, Neelam Chouksey, Sandeep Kumar, Shiv Kumar Rana, Saurabh Kanauajia and Randhir Yadav

Abstract
Historical data has been considered for forecasting of honey bee population. For the purpose, autoregressive integrated moving average with explanatory variables has been applied along with all estimation procedures. There are ARIMAX technique employed for forecasting of honey bee population on time-based data of Surguja district of Chhattisgarh. Data of Weather variables *viz.* maximum temperature, minimum temperature rainfall and Precipitation is taken from meteorological observatory of Rajmohini Devi College of, Agriculture and Research Station Ambikapur, Chhattisgarh as input variables in ARIMAX model. Comparative study of the fitted models is carried out from the viewpoint of mean absolute percentage error (MAPE), root mean squared error (RMSE). After comparison resulted that ARIMAX (1, 1, 1) and ARIMAX (1, 1, 0) model provided a lower RMSE from other models for yield and production respectively.

Keywords: Forecasting, ARIMAX, honey bee population, root mean square error (RMSE)

Introduction
Bee keeping is an art of keeping and manipulating of bees for purpose obtaining honey, a food of high nutritional and medicinal value and bee wax which is used in preparation cosmetic, boot polish and water proof paint. Bee keeping is the one of the oldest traditions in India for collecting the honey. Bee keeping is also play important role to increase agriculture productivity through pollination. About 80% crops are cross pollinated, as they need to receive pollen from other plant of same species with the help of external agents. Yield is increased about 44% in mustard, 32-45% in sunflower & 17-20% in cotton though bee pollinators. among all G-20 countries, India has sixth rank in honey production and first position in all SAARC Countries with total production 1.2 lakh metric tons in 2017-2018 (Source: http://www.indiastat.com). Honey production has increased from 76,150 MTs (2013) to 1,20,000 MTs (2019-20) which is 57.58% increase. In India Export of honey has increased from 28,378.42 MTs (2013-14) to 59536.74 MTs (2019-20) which has 109.80% increased. Punjab is the major state in beekeeping in the nation, with around 35,000 beekeepers delivering around 1500 metric tons honey. This is more than 39% of the nation's total honey production.

Honey bee population is influenced by many parameters in which weather parameters *viz*., Temperature, rainfall, relative humidity, sunshine hours and wind speed are play a vital role in determining abundance and distribution of honeybee's population. But even today, a very few theoretical frameworks are available to examine the effect of climate on population dynamics of honey bees. So that continue monitoring of honey bee if essential work for beekeeper for higher productivity of honey. It is also subject of attention for aphidologists and honey bee researchers. (Meikle and Holst 2015) [8]. Honey bees have made behavioural changes before rainfall, which is enormously important for their survival in severe weather conditions (He *et al*. 2016) [6]. Different bee species also prefer to forage at different temperatures, weather directly impacts hygro and thermodynamic processes within the pollinator, impacting survival rate and energy cost of foraging (Vicens and Bosch 2000) [11]. Chouksey *et al*. (2018a) [4] employed ARIMAX models for forecasting of Rice yield for different nutrient combinations of Raipur district of Chhattisgarh.

Chouksey *et al*. (2018b) [3] used combination of AR and MA order ARIMAX model for different nutrient combinations of nitrogen content and organic carbons included as an input variable. They found that ARIMAX model out performed as compared to ARIMA model.

Materials and Methods

Weekly data of honey bee population from 2014 to 2020 of Surguja district of Chhattisgarh was collected from ACRIP honey bee project at RMD College of agriculture and research station Ambikapur. Weather data like Temperature (max. and min), Evaporation, were collected from meteorological station at RMD college of agriculture & research station Ambikapur (C.G.).

Autoregressive Integrated Moving Average with explanatory variable (ARIMAX) – transfer function model: There is not possible to express the multivariate time series changing rules using the ARIMA model with only one time series. The reason is that the only ARIMA model is imperfect. Therefore, it is necessary to create a model with multivariate ARIMAX model. Assume two time series denoted Y_t and X_t which are both stationary. Then, the transfer function model (TFM) can be written as follows:

$$Y_t = C + V(B)X_t + N_t \tag{2}$$

Where: Y_t is the output series (nitrogen uptake); X_t is the input series (nitrogen content and organic carbon); C is constant term and N_t is the stochastic disturbance, i.e. the noise series of the system that is independent of the input series; $V(B)X_t$ is the transfer function (or impulse response function), which allows X to influence Y via a distributed lag and B is backshift operator.

Results and Discussion

The basic step in any statistical analysis is to know the behaviour of the data which are taken into consideration by calculating summary statistics and by drawing plot of time series. The summary table of maximum temperature, minimum temperature, Evaporation and honey bee population indicates that the data under consideration is normal in nature with less skewness and kurtosis as compared to other weather variables. Figure 1 to Figure 4 illustrated that series has been near to stationary in nature.

Table 1: Descriptive statistics of maximum temperature, minimum temperature, Evaporation, honey bee population

,	Honey bee population	Evaporation	Maximum Temperature	Minimum Temperature
Mean	108.14	3.2892	28.422	15.151
SD	83.167	1.4295	4.1515	5.7573
Minimum	15.800	1.0000	18.700	4.7000
Maximum	520.50	8.8000	40.300	23.700
Skew	1.5483	1.1557	0.4938	-0.0565
Kurtosis	2.4531	1.7189	0.3137	-1.3994

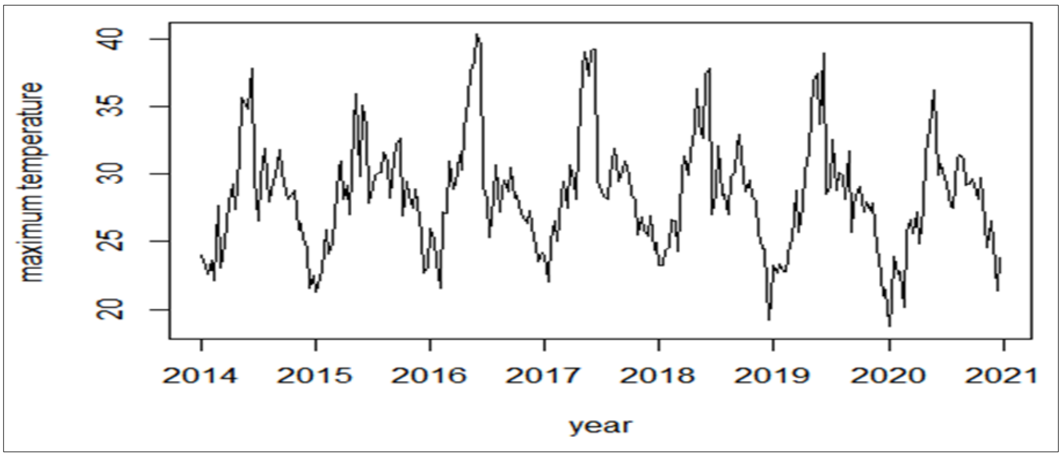


Fig 1: Time series plot of maximum temperature

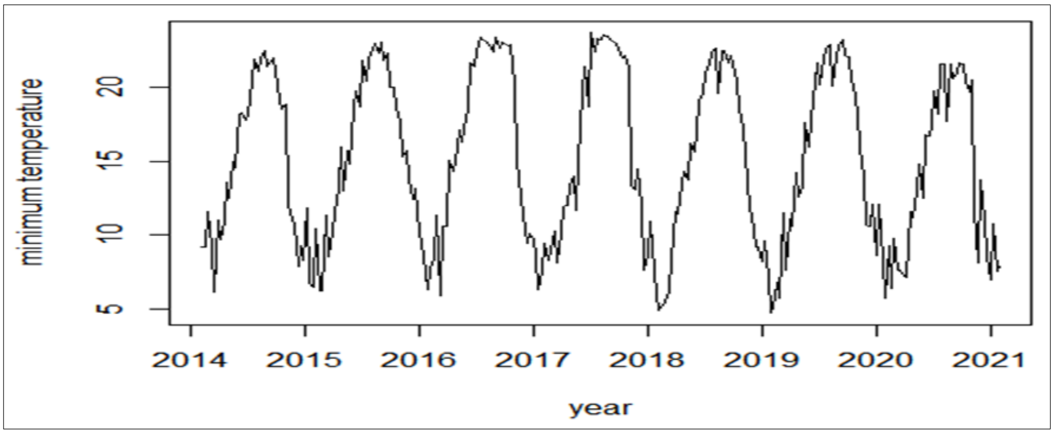


Fig 2: Time series plot of minimum temperature

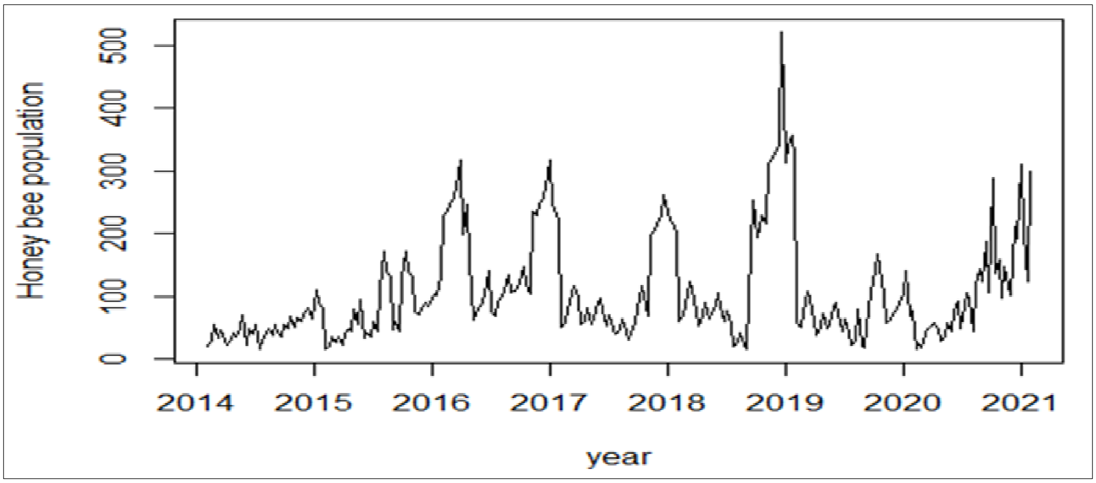


Fig 3: Time series plot of honey bee population

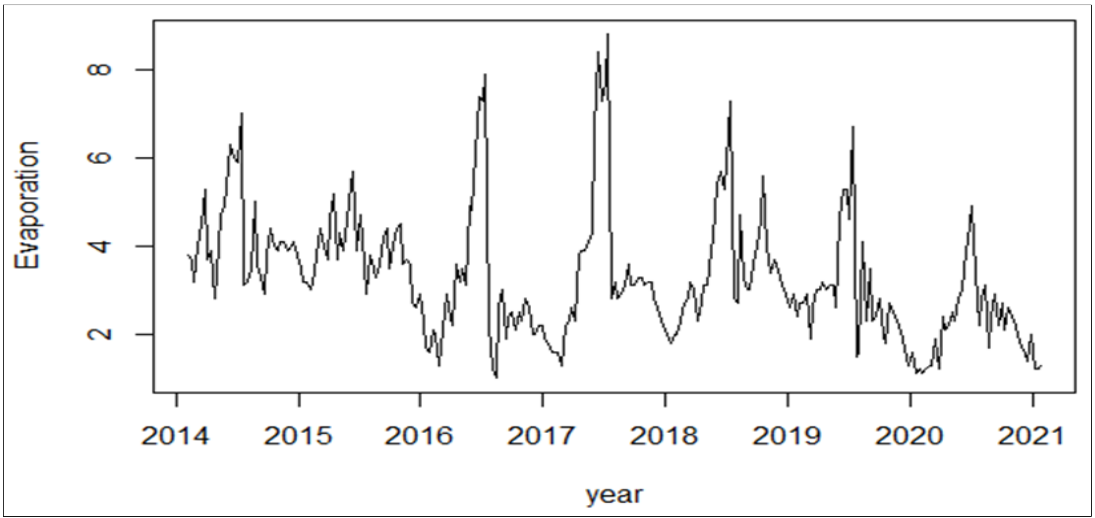


Fig 4: Time series plot of Evaporation

The above time plots showed the weekly time series of maximum temperature, minimum temperature, Evaporation and honey bee population from 2014 to 2020. A perusal of figure 1 to 3 revealed a linear trend over time which indicates the all three-time series under consideration were stationary in nature and Figure 4 illustrated that honey bee population series was near to stationarity. Based on the probability of chi-square value obtained in table 2 to table 5 confirms the all series were auto correlated. Once the series is auto correlated then, the next step is to go for stationarity testing of the series. Which is further confirmed by the results of Augmented Dickey-Fuller (ADF) unit root test and Phillips-Peron unit root test (PP test) statistics given in Table 6.

Table 2: Autocorrelation check for white noise of maximum temptature

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr>ChiSq	Autocorrelations					
6	430.3	6	< 2.2e-16	0.791	0.650	0.560	0.420	0.299	0.185

Table 3: Autocorrelation check for white noise of minimum temptature

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr>ChiSq	Autocorrelations					
6	137.61	6	< 2.2e-16	0.917	0.863	0.799	0.695	0.575	0.444

Table 4: Autocorrelation check for white noise of Evaporation

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr>ChiSq	Autocorrelations					
6	368.74	6	< 2.2e-16	0.753	0.601	0.498	0.366	0.264	0.206

Table 5: Autocorrelation check for white noise of honey bee population

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr>ChiSq	Autocorrelations					
6	536.85	6	< 2.2e-16	0.813	0.716	0.617	0.515	0.398	0.306

Table 6: Stationary test of maximum temperature time series

Series Name	ADF test statistic				PP test statistic			
	Single mean	With trend	Probability		Single mean	With trend	Probability	
			Single mean	With trend			Single mean	With trend
Maximum temperature	-4.6076	2.4409	0.0012	0.0154	-4.7923	2.3586	0.0006	0.0191
Minimum temperature	-4.7923	-0.4362	0.0006	0.6631	-5.3976	-2.468	0.0000	0.0388
Evaporation	-8.7653	-2.3074	0.0000	0.0492	-3.0966	-2.5700	0.0022	0.0369
Honey bee population	-6.4828	-0.2014	0.0000	0.8405	-9.9985	-0.3874	0.0000	0.6988

Table 6 illustrated that all-time series under consideration is found to be stationary then there is no need for differencing of the series to make it stationary.

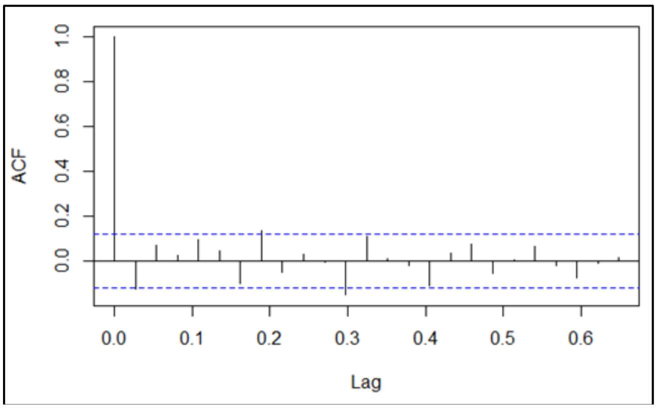
ARIMAX model for Honey bee population

The principal step in Box-Jenkins ARIMA model building is identification of the model. Different orders of Autoregressive (AR) and Moving Average (MA) parameters p and q are considered and combination of the order which yields maximum log-likelihood and lowest values of Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) are considered as final model orders. For honey bee population with weather variables (Maximum temperature, Minimum temperature, Evaporation) time series based on lowest BIC values obtained, ARIMAX (1,0,0) (0,0,2) was found adequate for series honey bee population. Once the model order was determined then, next step is to go for parameter estimation of the model by maximum likelihood estimation method which is the second step in Box-Jenkins ARIMA model building procedure. The results of parameter estimation of ARIMAX (1 0 0) (0,0,2) are given in table 7.

Table 7: Parameter estimation of ARIMAX (1 0 0) (0, 0, 2) by maximum likelihood estimation method for honey bee population

	Estimate	Std. Error	z value	Pr(>|z|)
AR1	0.8509	0.0332	25.60	< 2.2e-16 ***
SMA1	0.2436	0.0675	3.605	0.000311 ***
SMA2	0.1817	0.0753	2.413	0.015810 *
Intercept	154.79	40.46	3.825	0.000130 ***
Temperature Maximum	-1.6949	1.609	-1.053	0.292313
Temperature Minimum	-2.003	1.421	-1.409	0.158730
Evoporation	6.0365	4.216	1.431	0.152241

The third and final step in ARIMA model building is diagnostic checking of the model. Based on the ACF and PACF plots (Figure 5) and probability of autocorrelation of the residuals obtained in table 8 one can infer that the residuals are non-auto correlated.



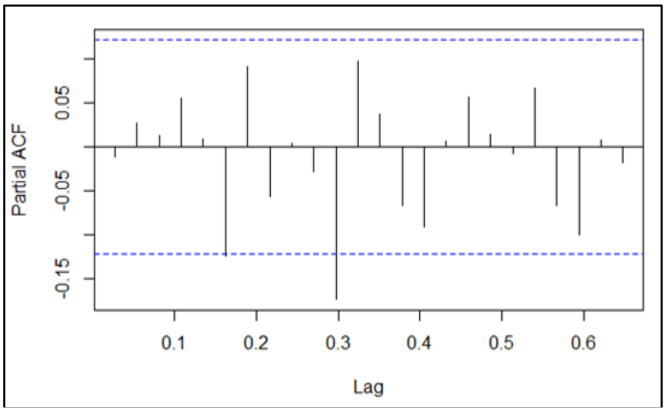


Fig 5: ACF and PACF of residual of model ARIMAX (1, 0, 0) (0, 0, 2) of series honey bee population

Table 8: Autocorrelation of residual of model ARIMAX (1, 0, 0) (0, 0, 2) for honey bee population

To Lag	Chi-Square	DF	Pr>ChiSq
6	2.1759	6	0.9028
12	3.5473	12	0.9903
18	23.871	18	0.1593
24	27.835	24	0.2671

After model building, next step is to go for model fitting called as training and based on obtained parameters; in-sample forecasting or holdout forecasting *i.e.,* performance of model under testing data set is carried out. For time series under consideration 3 years weekly data (from 2018-2020) as holdout sample (testing) is used for comparing model performance based on MASE, RMSE and MAPE. The MASE, RMSE and MAPE values for both training and testing data set is given in table 9 and table 10 respectively. The observed and fitted plot and forecasting trend of the time series under consideration is also given in Figure 6.

Table 9: Performance of ARIMAX (1, 0, 0) (0, 0, 2) for honey bee population time series in training data set

Criteria	ARIMAX
MASE	0.3915
RMSE	40.009
MAPE	34.2254
R SQUARE	0.7586

Table 10: Performance of ARIMAX (1, 0, 0) (0, 0, 2) for honey bee population time series in testing data set

Criteria	MASE	0.55599
	RMSE	57.6069
	MAPE	29.7258
	R SQUARE	0.4040

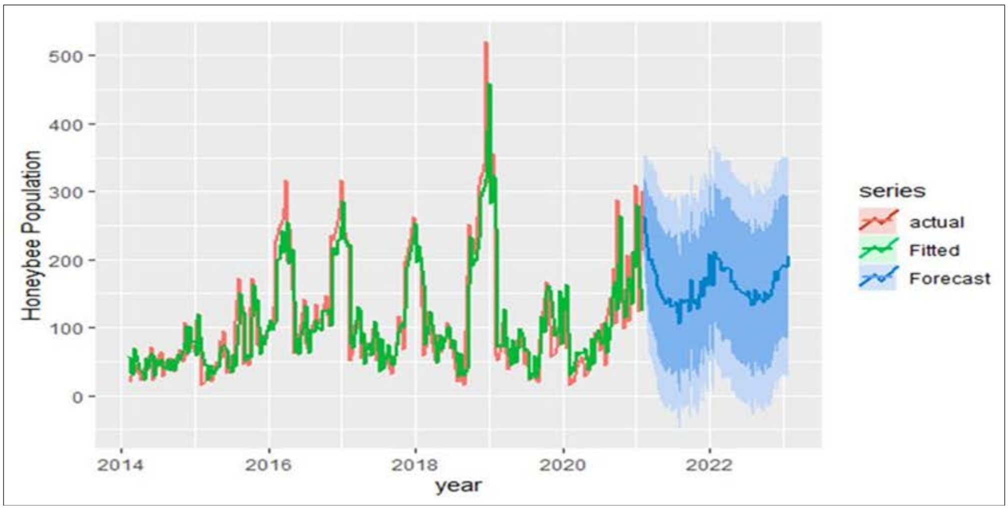


Fig 6: Actual v/s ARIMA fitted plot of honey bee population time series with forecasting.

Conclusion

Honey bee population weekly data of Raipur district of Chhattisgarh have been analysed and considering most important weather variables (maximum temperature, minimum temperature & evaporation) as input variables in ARIMAX model. Prediction for future value of Honey bee population has been done by the best selected ARIMAX model. Forecasted values of honey bee population showed less fluctuation for upcoming years. After study we can say that ARIMAX model may be helpful for researcher and policy makers of honey bee for forecasting and decision making.

References

1. Abac NI, Abac SH, Mustache S. Forecast for the number of colonies and honey yield in Turkey, Turkish Journal of Agriculture - Food Science and Technology. 2020;8(2):464-470.
2. Aswathi VS, Duraisamy MR. Comparison of Prediction Accuracy of Multiple Linear Regression, ARIMA and ARIMAX Model for Pest Incidence of Cotton with Weather Factors. Madras Agricultural Journal. 2018 Sept;105(7-9):313-316.
3. Chouksey N, Mishra GC, Singh SK, Singh M, Chouksey R. Autoregressive integrated moving average with explanatory variable (ARIMAX) model for rice yield on different fertilizer combinations, International Journal of Agriculture, Environment and Biotechnology. 2018b;11:1029-1033.
4. Chouksey N, Mishra GC, Singh SK, Singh M, Chouksey R. Forecasting of nitrogen content in the soil by hybrid time series model, International Journal of Current Microbiology and Applied Sciences. 2018a;7:07.
5. Dominic Clarke, Daniel Robert. Predictive modelling of honey bee foraging activity using local weather conditions, Apidologie. 2018;49:386-396.
6. He XJ, Tian LQ, Wu XB, Zeng ZJ. RFID monitoring indicates honeybees work harder before a rainy day. Insect Science. 2016;23(1):157-159.
7. Ling ASC, Darmesah G, Chong KP, Ho CM. Application of ARIMAX Model to Forecast Weekly Cocoa Black Pod Disease Incidence. Mathematics and Statistics. 2019;7(4A):29-40.
8. Meikle WG, Holst N. Application of continuous monitoring of honeybee colonies. Apidologie. 2015;46(1):10-22.
9. Paul RK, Prajneshu, H Ghosh. Modelling and forecasting of wheat yield data based on weather variables. Indian Journal of Agricultural Science. 2013;83(2):180-183.
10. Paul RK, Vennila S, Singh N, Chandra P, Yadav SK, Sharma OP, *et al*. Seasonal dynamics of sterility mosaic of pigeonpea and its prediction using statistical models for Banaskantha region of Gujarat, India. Journal of the Indian Society of Agricultural Statistics. 2018;72(3): 213-233.
11. Vicens N, Bosch J. Weather-dependent pollinator activity in an apple orchard, with special reference to *Osmia cornuta* and *Apis mellifera* (Hymenoptera: Megachilidae and Apidae). Environmental Entomology. 2000;29(3): 413-420.