International Journal of
Statistics and Applied Mathematics

**ISSN:** 2456-1452
Maths 2023; SP-8(5): 166-169

https://www.mathsjournal.com
Received: 19-04-2023
Accepted: 20-05-2023

**Neelam Chouksey**
RMD College of Agriculture & Research Station, Ambikapur, Chhattisgarh, India

**GP Painkra**
RMD College of Agriculture & Research Station, Ambikapur, Chhattisgarh, India

**KL Painkra**
RMD College of Agriculture & Research Station, Ambikapur, Chhattisgarh, India

**PK Bhagat**
RMD College of Agriculture & Research Station, Ambikapur, Chhattisgarh, India

**Corresponding Author:**
**Neelam Chouksey**
RMD College of Agriculture & Research Station, Ambikapur, Chhattisgarh, India

# Forecasting of honey bee population by Arima model in Surguja district of Chhattisgarh

**Neelam Chouksey, GP Painkra, KL Painkra and PK Bhagat**

**Abstract**
Historical data has been leveraged to forecast honey bee population dynamics in the research paper. The study employs the Autoregressive Integrated Moving Average (ARIMA) methodology to predict honey bee population trends using time-based data from the Surguja district of Chhattisgarh. A comparative analysis of various fitted models is conducted, with a focus on assessing their performance through Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE). The findings indicate that the ARIMA (2,0,2) (0,0,2) model outperforms other models in terms of RMSE for forecasting honey bee population. Our approach involves meticulous scrutiny of the time series data, accompanied by a rigorous process of model identification and parameter estimation. This comprehensive methodology establishes a sturdy framework for generating precise predictions. The ramifications of our study are pertinent to enhancing informed decision-making within honey production management strategies.

**Keywords:** Forecasting, ARIMA, honey bee population

## Introduction
Beekeeping in India has a rich legacy, entailing the delicate art of nurturing bees to yield honey and beeswax. In the global landscape, honey bee farming has seized considerable attention due to bees' pivotal role as indispensable crop pollinators. India's ranking as the sixth-largest honey producer among G-20 nations, with Punjab prominently contributing, underscores the economic significance of beekeeping. This industry transcends mere culinary allure, with bees assuming a central role as key pollinators that substantially influence crop yields and agricultural productivity. The relevance of this field has also attracted the attention of aphidologists and honey bee researchers, as seen in the work of Meikle and Holst (20150) [11] and He *et al*. (2016) [8], where honey bees exhibited behavioral adaptations prior to rainfall, essential for their survival in adverse weather conditions.

However, despite its importance, the honey bee population encounters multifaceted influences, with climate conditions taking a lead role. Remarkably, there exists a noticeable scarcity of comprehensive theoretical frameworks delving into the intricate link between climate variables and honey bee population dynamics. This gap in research emphasizes the critical need for continuous monitoring of honey bee populations - a practice vital for beekeepers to optimize honey production. Moreover, this research avenue has piqued the curiosity of entomologists and researchers, unraveling captivating paths for further exploration and study.

Drawing from diverse disciplines, the Autoregressive Integrated Moving Average (ARIMA) model emerges as a potent tool for prognosticating honey bee population dynamics. Aswathi, and Duraisamy, (2018) [4] used ARIMA to predict pest incidence of cotton and Al-Sakkaf and Jones's (2014) [3] applied ARIMA for forecasting campylobacteriosis incidence in New Zealand between 1998 and 2008 attest to the model's adaptability. Boopath *et al.* (2015) [5] studied temporal modeling for prediction of the incidence of lychee insect, *Tessaratoma papillosa* (Hemiptera: Tessaratomidae), using time-series (ARIMA) analysis. Its flexible nature renders it an ideal choice for comprehending and predicting the intricate behaviors within honey bee populations, thereby illuminating their responses under varying climatic scenarios. In the study applying ARIMA model to predict honey bee population in response to future years holds great promise. By shedding light on these dynamics, we can contribute to the preservation of honey bee populations and the crucial pollination services they provide.

## Materials and Methods

Weekly data of honey bee population from 2014 to 2021 of Surguja district of Chhattisgarh was collected from ACRIP honey bee project at RMD college of agriculture and research station Ambikapur.

## Autoregressive Integrated Moving Average Process (ARIMA)

The Autoregressive Integrated Moving Average (ARIMA) Model (Box-Jenkins, 1970) stands as the most commonly employed linear statistical model. ARIMA serves as a conventional technique for analyzing non-stationary time series data. In contrast to regression models, the ARIMA model elucidates the evolution of a variable, denoted as "$r_t$," through its historical values (lags) and stochastic error components. These models are frequently denoted as "mixed models." Although this approach adds intricacy to forecasting methods, it potentially emulates the series' structure more faithfully, resulting in heightened forecast accuracy.

The notion of pure models implies a structure composed solely of autoregressive (AR) or moving average (MA) parameters, excluding the presence of both. Models stemming from this methodology are typically referred to as ARIMA models due to their amalgamation of autoregressive (AR), integration (I) – pertaining to the inverse process of differencing for forecasting – and moving average (MA) operations. An ARIMA model is conventionally denoted as ARIMA (p, d, q). The formulation of an autoregressive integrated moving average model takes the following form:

$$w_t = \nabla^d r_t = (1-B)^d r_t$$

$$w_t = \phi_1 w_{t-1} + \phi_2 w_{t-2} + \dots + \phi_p w_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

If $\{W_t\}$ follows the ARMA (p, q) model, and $\{r_t\}$ is an ARIMA (p, d, q) process. For practical purposes, we can take is usually d = 1 or 2 at most. Above equation is also written as:

$$\phi(B) w^t = \theta_0 + \theta(B)\varepsilon_t$$

Where $\phi(B)$ is a stationary autoregressive operator, $\theta(B)$ is a stationary moving average operator and $\varepsilon_t$ is white noise and $\theta_0$ is a constant.

## Results and Discussion

In this section, we delve into the analysis of the time series data related to honey bee populations and the subsequent construction of an appropriate ARIMA model for accurate forecasting. Weekly data of honey bee population from 2014 to 2021 of Surguja district of Chhattisgarh was collected from ACRIP honey bee project at RMD college of agriculture and research station Ambikapur.

### Status of time series

Before proceeding with the analysis, a fundamental step involved understanding the nature of the data. This was achieved by computing summary statistics and generating time series plots. The summary statistics, detailed in table 1, confirm the normal distribution of the dataset, exhibiting minimal skewness and kurtosis. Moreover, as depicted in figure 1, the time series exhibited characteristics of near-stationarity.

**Table 1:** Descriptive statistics of honey bee population.

| , | Mean | Median | SD | Minimum | Maximum | Skew | Kurtosis |
|---|---|---|---|---|---|---|---|
| honey bee population | 108.48 | 80 | 80.9 | 15.8 | 520.5 | 1.50 | 2.44 |

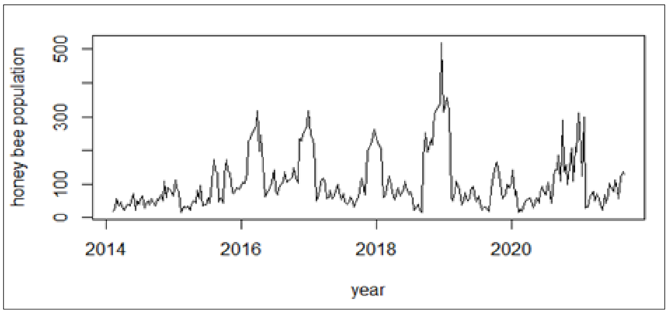


**Fig 1:** Time series plot of honey bee population.

The visual inspection of the weekly time series plots spanning 2014 to 2021 revealed a linear trend, suggesting the series' stationary nature. Corroborating this, the probability-based significant level and the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots, as shown in figure 2, confirmed the presence of autocorrelation within the series. Augmented Dickey-Fuller (ADF) unit root test statistics, outlined in table 2, reinforced the stationary nature of the series, thus negating the requirement for differencing.

### ARIMA model for Honey bee population

The cornerstone of constructing an ARIMA model through the Box-Jenkins methodology is the identification of the most suitable model order. This entails evaluating various orders of Autoregressive (AR) and Moving Average (MA) parameters, denoted as p and q respectively. The final model order is chosen based on the combination yielding the highest log-likelihood while minimizing the Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC), as delineated in table 3. Notably, the ARIMA (2, 0, 2) (0, 0, 2) configuration emerged as the optimal model for the honey bee population series.

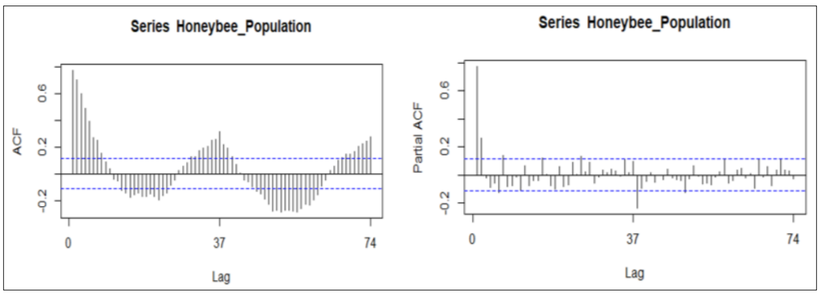


**Fig 2:** ACF and PACF plot

**Table 2:** Stationary test of honeybee population time series

| Coefficients | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 18.25916 | 4.96898 | 3.675 | 0.000283 *** |
| z.lag.1 | -0.16074 | 0.03759 | -4.276 | 2.58e-05 *** |
| z.diff.lag | -0.26125 | 0.05710 | -4.575 | 7.04e-06 *** |
| | test-statistic | Critical value | | |
| | | 5 % | 1 % | |
| tau2 | -4.2763 | -2.87 | -3.44 | |
| phi1 | 9.1908 | 4.61 | 6.47 | |

**Table 3:** Identification parameter for different model of ARIMA for honey bee population

| Models | AIC | BIC | RMSE | MAPE |
|---|---|---|---|---|
| ARIMA (0, 0, 0) (0, 0, 2) | 2967.33 | 2981.88 | 46.60 | 39.73 |
| ARIMA (0, 0, 1) (0, 0, 2) | 3087.53 | 3105.72 | 56.94 | 4.89 |
| ARIMA (2, 0, 0) (0, 0, 2) | 3115.59 | 3145.11 | 44.89 | 38.47 |
| ARIMA (1, 0, 0) (0, 0, 2) | 2962.24 | 2980.43 | 45.82 | 37.90 |
| ARIMA (2, 0, 1) (0, 0, 2) | 2960.12 | 2978.27 | 45.85 | 39.85 |
| ARIMA (2, 0, 2) (0, 0, 2) | 2950.32 | 2974.43 | 44.46 | 38.35 |
| ARIMA (1, 0, 2) (0, 0, 2) | 2953.56 | 2979.03 | 44.70 | 37.70 |
| ARIMA (2, 0, 0) (1, 0, 1) | 2951.75 | 2973.58 | 44.73 | 37.75 |
| ARIMA (2, 0, 1) (0, 0, 2) | 2953.56 | 2979.03 | 44.70 | 37.70 |
| ARIMA (2, 0, 2) (0, 0, 1) | 2960.35 | 2985.82 | 45.54 | 41.01 |

The third and final step in ARIMA model building is diagnostic checking of the model. Based on the residual ACF and PACF plots (figure 3). One can infer that the residuals are non-autocorrelated. Table 3 depicted a comparative analysis of various fitted models was conducted, with a focus on assessing their performance through minimum Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE). After model building, next step is to go for model fitting based on obtained parameters *i.e.* performance of model. The observed and fitted plot and forecasting trend of the time series under consideration is also given in figure 4.

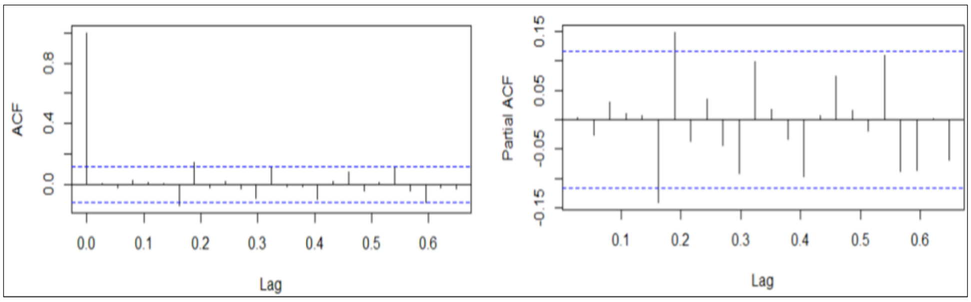


**Fig 3:** ACF and PACF of residual of model ARIMA (2,0,2) (0,0,2)

**Parameter Estimation and Model Adequacy**

Having determined the model order, the subsequent step encompassed parameter estimation using the maximum likelihood estimation method. This procedure, as a part of the Box-Jenkins ARIMA building process, helps refine the model's parameters for enhanced forecasting precision. In conclusion, our analysis demonstrates the applicability of the ARIMA (2, 0, 2) (0, 0, 2) model to forecast honey bee population trends. The observed and fitted plot and forecasting trend of the time series under consideration is also given in figure 10. Other researchers, such as Abac *et al.* (2020) [1], and Clarke and Robert (2018) [7], were also reported similar results in their studies.

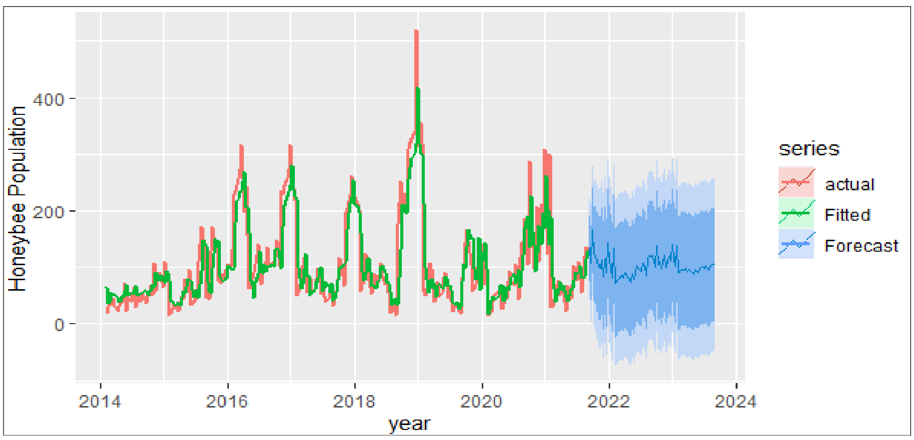


**Fig 4:** Actual v/s ARIMA fitted plot of honey bee population time series with forecasting.

## Conclusion

The Box-Jenkins methodology was meticulously applied to the honey bee population series, resulting in the identification of the ARIMA (2, 0, 2) (0, 0, 2) model as the fitting choice across all stages of analysis. Subsequently, for the purpose of forecasting the honey bee population in Surguja district, Chhattisgarh, the ARIMA (2, 0, 2) (0, 0, 2) model was adopted. The careful examination of the time series data, along with rigorous model identification and parameter estimation, provides a robust framework for accurate predictions. The implications of this study extend to better-informed decision-making in honey production management.

## 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. Agrawal R, Mehta SC. Weather based forecasting of crop yield, pest and diseases-IASRI Models. Journal of the Indian Society of Agricultural Statistics. 2007;61(2):255-263.
3. Al-Sakkaf A, Jones G. Comparison of time series models for predicting campylobacteriosis risk in New Zealand. Zoonoses Public Health. 2014;61(3):167-74. DOI:10.1111/zph.12046. Epub 2013 Apr 2. PMID: 23551848.
4. 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.
5. Boopathi T, Singh SB, Manju T, Ramakrishna Y, Akoijam RS, Chowdhury S, *et al*. Development of temporal modeling for forecasting and prediction of the incidence of lychee, *Tessaratoma papillosa* (Hemiptera: Tessaratomidae), using time-series (ARIMA) analysis. Journal of Insect Science. 2015;15(1):55.
6. Neelam C, Tokesh K, Rajesh C. Honey bee forecasting by arimax model using weather variables. Frontiers in Crop Improvement. (Special Issue-I). 2022;10:490-494.
7. Clarke D, Robert D. Predictive modelling of honey bee foraging activity using local weather conditions. Apidologie. 2018;49:386-396. https://doi.org/10.1007/s13592-018-0565-3.
8. 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.
9. Koi B, Term M, Ceylan M, Dagistan E. General Situation of Beekeeping in the Eastern Anatolian Region of Turkey and ARIMA Model with the Help of Long-term Analysis, Asian Journal of Aural and Veterinary Advances. 2010;5(8):537-546.
10. 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.
11. Meikle WG, Holst N. Application of continuous monitoring of honeybee colonies. Apidologie. 2015;46(1):10-22.
12. 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.
13. Kalia V, Yadav B. Morphometrics of honey bee (*Apis mellifera*) populations from Northern India, Indian Journal of Emtomology. 2012;74(2):148-153.