AATCC Review

06 October 2025: Received
16 December 2025: Revised
20 December 2025: Accepted
07 January 2026: Available Online

https://aatcc.peerjournals.net/

Original Research Article  Open Access

# Count time series and machine learning for climate-driven prediction of rice yellow stem borer infestation in andhra pradesh

**P. Lavanya Kumari*[1], I. Paramasiva[2], U. Vineetha[3], A. Veeraiah[4], SK. Shameem[5], P. N. Harathi[6], A.D.V.S.L.P Anand Kumar[7], M. Siva Rama Krishna[8], N. Sambasiva Rao[9], P. Udayababu[10], J. Manjunath[8], N. Kamakshi[9], V. Visalakshmi[11]**

[1]*Department of Statistics & Computer Applications, S.M.G.R. Agricultural College, Udayagiri, Acharya N. G. Ranga Agricultural University (ANGRAU), Andhra Pradesh, India*
[2]*Department of Entomology, Agricultural Research Station, Nellore, Acharya N. G. Ranga Agricultural University (ANGRAU), Andhra Pradesh, India*
[3]*Department of Agronomy, Agricultural Research Station, Nellore, Acharya N. G. Ranga Agricultural University (ANGRAU), Andhra Pradesh, India*
[4]*Krishi Vigyan Kendra (KVK), Kadapa, Acharya N. G. Ranga Agricultural University (ANGRAU), Andhra Pradesh, India*
[5]*Department of Statistics & Computer Applications, S.V. Agricultural College, Tirupati, Acharya N. G. Ranga Agricultural University (ANGRAU), Andhra Pradesh, India*
[6]*Department of Entomology, Regional Agricultural Research Station, Tirupati, Acharya N. G. Ranga Agricultural University (ANGRAU), Andhra Pradesh, India*
[7]*Department of Entomology, Regional Agricultural Research Station, Maruteru, Acharya N. G. Ranga Agricultural University (ANGRAU), Andhra Pradesh, India*
[8]*Department of Entomology, Regional Agricultural Research Station, Nandhyal, Acharya N. G. Ranga Agricultural University (ANGRAU), Andhra Pradesh, India*
[9]*Department of Entomology, Agricultural Research Station, Bapatla, Acharya N. G. Ranga Agricultural University (ANGRAU), Andhra Pradesh, India*
[10]*Department of Entomology, Agricultural Research Station, Ragolu, Acharya N. G. Ranga Agricultural University (ANGRAU), Andhra Pradesh, India*
[11]*Department of Entomology, Regional Agricultural Research Station, LAM, Guntur, Acharya N. G. Ranga Agricultural University (ANGRAU), Andhra Pradesh, India*

## ABSTRACT

*Yellow Stem Borer (YSB) (Scirpophaga incertulas) is one of the most destructive pests affecting rice production in India, causing significant yield losses across different agro-climatic regions. Accurate forecasting of YSB populations is crucial for timely pest management and minimizing crop damage. This study evaluates the performance of statistical and machine learning models for predicting YSB populations using weekly pest incidence data collected from five research stations in Andhra Pradesh (Nellore, Maruteru, Bapatla, Ragolu, and Nandyal) over multiple years. The study employs Integer-Valued Generalized Autoregressive Conditional Heteroskedastic (INGARCH) models along with Artificial Neural Networks (ANN), Support Vector Regression (SVR), Extreme Learning Machines (ELM), and their hybrid counterparts (INGARCH-ANN, INGARCH-SVR, and INGARCH-ELM) to improve forecasting accuracy. Results indicate that hybrid models, particularly NBINGARCH-ELM, consistently outperformed standalone models in terms of Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) across different locations and seasons. The findings reveal that YSB populations are significantly influenced by climatic factors such as temperature, relative humidity, and rainfall, with distinct seasonal patterns. The Box-Pierce test confirmed minimal autocorrelation in residuals for most models, validating their reliability. These results highlight the potential of hybrid statistical machine learning models for pest forecasting, providing valuable insights for integrated pest management (IPM) strategies. Future research can further enhance these models by incorporating additional environmental and agronomic variables for improved precision in pest outbreak predictions.*

***Keywords:*** *Yellow Stem Borer, Machine Learning, INGARCH Model, Hybrid Models, Extreme Learning Machine, Support Vector Regression, Artificial Neural Networks and Rice Pest Forecasting, Box-Pierce Test..*

## Introduction

Rice (*Oryza sativa*) is a staple crop that serves as the primary food source for more than half of the world's population and plays a crucial role in global food security. However, rice cultivation is severely affected by insect pests, among which the Yellow Stem Borer (*Scirpophaga incertulas* Walker) (YSB) is one of the most destructive pests, particularly in South and Southeast Asia.

*Corresponding Author: ***P. Lavanya Kumari***

*DOI: https://doi.org/10.21276/AATCCReview.2026.14.01.61*

In India, YSB is a major pest affecting rice production, leading to significant yield losses across diverse agro-climatic regions.

YSB is a monophagous pest that feeds exclusively on rice plants, attacking crops at all growth stages. The larvae bore into rice stems, disrupting nutrient and water transport, leading to "dead heart" symptoms in the vegetative stage and "white earhead" symptoms in the reproductive stage. These damages significantly reduce yield potential and grain quality. The life cycle of the yellow stem borer consists of four stages: egg, larva, pupa, and adult. The total duration of the life cycle (egg-adult) can vary between 32 to 45 days depending on temperature [9]. Female moths lay eggs in masses of 100 to 200 eggs near the tip of leaf blades. Eggs hatch in 6 to 8 days under warm conditions. The larva is the damaging stage and undergoes six instars in 21 to 29 days. Pupation takes place inside the stem, usually in the lowest node of the plant or just above the water level. The pupal period lasts about 5 to 7.5 days depending on temperature. The female moth is larger than the male with forewings that are bright yellowish brown with a distinct black spot in the center. Moths are active at night; a female can lay up to three egg masses during her 6 to 10-day life span. YSB populations are strongly influenced by climatic factors such as temperature, relative humidity, rainfall, and sunshine hours, along with agronomic practices like continuous rice cropping, absence of crop rotation, and the presence of alternative hosts. Peak infestations typically occur during the monsoon (July–September) and post-monsoon (October–November) seasons, coinciding with high humidity (above 80%) and temperatures ranging from 22°C to 30°C. Persistent infestations often occur in rice fields that experience overlapping crop cycles, facilitating the uninterrupted proliferation of YSB populations.

Despite its economic significance, the ability to predict YSB outbreaks remains a challenge due to the highly variable and nonlinear nature of its population dynamics. Conventional statistical models such as multiple linear regression and autoregressive integrated moving average (ARIMA) models have been widely used for pest forecasting but are often ineffective in capturing autocorrelated, over-dispersed, and nonlinear patterns in insect population data. Count time series models, such as the Integer-Valued Generalized Autoregressive Conditional Heteroskedastic (INGARCH) model, offer a better alternative for modeling YSB population dynamics as they account for integer-valued autocorrelated count data, making them more suitable for insect pest forecasting.

In recent years, machine learning (ML) techniques have gained prominence in agricultural forecasting, demonstrating superior performance in handling complex datasets with nonlinear dependencies. Artificial Neural Networks (ANN), Support Vector Regression (SVR), and Extreme Learning Machines (ELM) have been successfully applied to various domains, including crop yield prediction, disease forecasting, and pest modeling. These ML models can efficiently capture intricate relationships between YSB populations and climatological parameters, offering more accurate forecasting capabilities. Recognizing the limitations of standalone statistical and ML models, this study proposes the integration of INGARCH with ANN, SVR, and ELM i.e INGARCH-ANN, INGARCH-SVR, and INGARCH-ELM to enhance forecasting accuracy by leveraging the strengths of both approaches.

This research was conducted in Bapatla, Nandyal, Nellore, West Godavari (Maruteru), and Srikakulam (Ragolu)districts of Andhra Pradesh, where YSB infestations are frequently observed. Weekly cumulative YSB population data were collected using solar light traps, while meteorological parameters, including temperature, rainfall, relative humidity, and sunshine hours, were recorded from automatic weather stations in and around the study locations. The study evaluates the predictive performance of INGARCH, ANN, SVR, and ELM models and their hybrid versions; INGARCH-ANN, INGARCH-SVR, and INGARCH-ELM using Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) as comparative measures.

By integrating statistical and machine learning-based approaches, this study aims to improve the accuracy of YSB population predictions and provide timely pest management recommendations. The findings will help farmers in Andhra Pradesh adopt effective pest control strategies, reduce crop losses, and ensure sustainable rice production.

## 2. Materials and Methods

### 2.1. Data Collection

Secondary data on light trap catches of major pests in rice was collected from the Nellore, Maruteru, Bapatla, Ragolu, and Nandyal agricultural research stations run under Acharya NG Ranga Agricultural University, Andhra Pradesh. The dataset provides weekly observations from different research stations in Andhra Pradesh, covering multiple years for both the Kharif and Rabi seasons. At Nellore (NLR), data is available from 2009 to 2023, with 285 observations in Kharif (SMW 27–45) and 255 observations in Rabi (SMW 46–10). Ragolu (RGL) has data from 2011 to 2023, with 286 observations in both Kharif (SMW 26–47) and Rabi (SMW 48–17). Maruteru (MTU) has the longest record, with Kharif data from 2002 to 2023 (616 observations, SMW 25–52) and Rabi data from 2003 to 2023 (420 observations, SMW 1–20). Bapatla (BPT) has Kharif data from 2011 to 2023 (364 observations, SMW 32–7), while Nandyal (NDL) has Kharif data from 2014 to 2022 (225 observations, SMW 33–5). The Standard Meteorological Weeks (SMW) indicate the time frames within each season when data was collected. Ten-week observations were used as testing/validation sets, and the remaining observations were used as the training data set.

### 2.2. Statistical Models

Statistical modeling started with descriptive statistical parameters encompassing mean, standard error (SE), skewness, kurtosis, minimum observation, maximum observation, and coefficient of variations (CV), which are important in depicting the nature of the studied data. Apart from the descriptive statistics, data were depicted graphically with time series plots. Pearson's product-moment correlation analysis was carried out to determine the interrelationship among the variables used in the study. Various time series models, machine learning models, and hybrid models were considered as mentioned below.

#### *2.2.1. Integer-Valued Generalized Autoregressive Conditional Heteroscedastic (INGARCH) Model*

The time series following the generalized linear model (GLM) framework was elaborated by [7]. INGARCH models are a class of GLM mentioned in [6] and [2], in which the conditional distribution of the dependent variable is assumed to follow popular discrete distributions like Poisson, negative binomial, generalized Poisson, and double Poisson distributions [17].

Let the count time series be $Yt: t \in N$and the time-varying r-dimensional covariate vector be $Xt: t \in N$ i.e., $Xt = (Xt, 1, \dots, Xt, r)^T$.The conditional mean becomes $E(\, Yt \mid Ft - 1\,)$and Ft is historical data. The generalized model form is expressed as follows:

$$g(\lambda_t) = \beta_0 + \textstyle\sum_{K=1}^{p} \alpha_k\, \tilde{g}\left(Y_{t-i_k}\right) + \sum_{l=1}^{q} \beta_1 g\left(\lambda_{t-j_l}\right) + \eta^T \; \dots (1)$$

Case 1: Consider the situation where g and $\tilde{g}$ are equal to identity, i.e., g(x)=, $\tilde{g}(x) = x$
Further, $Y_t$ follows (Poisson) INGARCH$(p, q)$ model with $p \;>\; 1$ and $q \;\geq\; 0$ if

(a) $Y_t$ conditioned on$Y_{t-1}\,, Y_{t-2}\,, \dots,$ is Poisson distributed
(b) The conditional mean $\lambda_t = E[Y_t \mid Y_{t-1}\,, Y_{t-2}, \dots\,]$ satisfies

$$\lambda_t \;=\; \beta_0 \;+\; \sum_{i=1}^{p} \alpha_i\, Y_{t-i} \;+\; \sum_{j=1}^{q} \beta_j\, \lambda_{t-j}\,,\ ]\text{where}(\beta_0 > 0)$$

and$(\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q \geq 0)$ ... (2)

Assuming further that $Y_t|Y_{t-1}$ is Poisson distributed, then we obtain an INGARCH model of order $p$ and $q$, abbreviated as INGARCH $(p, q)$model. If $q \;=\; 0$, the model can be referred to as the INAGARCH (p) model. These models are also known as autoregressive conditional Poisson (ACP) models.

**Case 2:**
The negative binomial distribution allows for a conditional variance to be larger than the mean $\lambda_t$which is often referred to as over-dispersion (with over-dispersion parameter $\emptyset$) mentioned in [1]. It is assumed that $Y_t \mid F_{t-1} \sim$ NegBinom$(\lambda_t, \emptyset)$. when $\emptyset \to \infty$, the Poisson distribution is a limiting case of the negative binomial distribution by the assumption:

$$Y_t \mid Y_{t-1}, Y_{t-2}, \dots \sim \text{Bin}\left(n, \tfrac{\beta + \alpha Y_{t-1}}{n}\right) \dots (3)$$

Additional details into the estimation of INGARCH models using conditional likelihood methods, particularly regarding their asymptotic properties, can be found in the works of [6] and [3]. The standard INGARCH model is designed to generate forecasts relying solely on past observations of the forecast variable. It operates under the assumption that future outcomes are influenced by both their own lagged values and, when applicable, by the lagged values of explanatory variables. An enhanced version, known as the INGARCHX model, incorporates exogenous variables explicitly into the framework, providing a more flexible structure for modeling time series data with external influences [8].

### *2.2.2. Artificial Neural Network (ANN)*

ANN is the most widely used machine learning technique in the last several decades. In the area of time series modeling, the ANN is commonly referred to as the autoregressive neural network as it considers time lags as inputs. The time series framework for ANN can be mathematically modeled using a neural network with the implicit functional representation of time. The general expression for the final output $Y_t$ of a multi-layer feed-forward autoregressive neural network is expressed as follows:

$$Y_t = \alpha_0 + \textstyle\sum_{j=1}^{q} \alpha_j g\left(\beta_{0j} + \sum_{i=1}^{p} \beta_{ij} Y_{t-p}\right) + \epsilon_t \dots (4)$$

$where, \alpha_j(j \;=\; 0, 1, 2, \dots, q)\ and\ \beta_{ij}(i \;=\; 0, 1, 2, \dots, p, j \;=\; 0, 1, 2, \dots, q)$ are the model parameters, also called the synopsis weights, $p$ is the number of input nodes, $q$ is the number of hidden nodes, and $g$ is the activation function. The training part in ANN minimizes the error function between actual and predicted values. The error function of autoregressive ANN is expressed as follows:

$$E = \frac{1}{N}\sum_{t=1}^{N}(e_t)^2 = \frac{1}{N}\sum_{t=1}^{N}\left(X_t - \left(w_0 + \sum_{J=1}^{Q} w_J g\left(w_{0j} + \sum_{i=1}^{P} w_{ij} X_{t-i}\right)\right)\right)^2$$

Where N is the total number of error terms.
The parameters of the neural network $w_{ij}$ are changed by a number of changes in $\Delta w_{ij}$ as $\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}}$ where $\eta$ is the learning rate [11], [16]. As in INGARCHX and SVRX models, the exogenous variable will also be used To model the pest count, and hence becomes ANNX model.

### *2.2.3. Support Vector Regression (SVR)*

The principal idea involved in SVR is to transform the original input space into high-dimensional variable space and then build the regression or time series model in a transformed high-dimensional feature space. A vector of data set says $Z = \{x_i\ y_i\}_{i=1}^{N}, where\ x_i \in R^n$ is the input vector, $y_i$ is the scalar output, and N is the size of data set. The general equation SVR can be written as follows:

$$f(x) = W^T f(x) + b \dots (5)$$

where, $W$ is the weight vector, $b$ is bias term, and superscript $T$ denotes the transpose. The coefficients $W$ and $b$ are estimated from data by minimizing the following regularized risk function:

$$R(q) = \tfrac{1}{2}||\text{w}||^2 + C\left(\tfrac{1}{N}\textstyle\sum_{i=1}^{N} L_\varepsilon\left(y_i, f(x_i)\right)\right) \dots (6)$$

This regularized risk function minimizes both the empirical error and regularized term simultaneously, which helps in avoiding both under and overfitting of the model. In the above Equation, the first term $\frac{1}{2}||\text{w}||^2$ is called the 'regularized term', which measures the flatness of the function. Minimizing $\frac{1}{2}||\text{w}||^2$ will make a function as flat as possible. The second term $\frac{1}{N}\sum_{i=1}^{N} L_\varepsilon\left(y_i, f(x_i)\right)$ is called the 'empirical error', which was estimated by Vapnik $\varepsilon$ −insensitive loss function as follows:

$$L_\varepsilon\left(y_i, f(x_i)\right) = \; f(x) \; = \begin{cases} |y_i - f(x_i) - \varepsilon|, & |y_i - f(x_i) \geq \varepsilon|, \\ \quad 0 & |y_i - f(x_i) < \varepsilon|, \end{cases}$$

where, $y_i$ is actual value and $f(x_i)$is an estimate value.
The most commonly used kernel function is the radial basis function (RBF) which is given as follows:

$$k\left(x_i, x_j\right) = \exp\{-\gamma||\text{x} - x_i||^2\} \dots (7)$$

The effectiveness of the Radial Basis Function (RBF) kernel relies on the proper tuning of two key hyperparameters: the regularization parameter *C,* which controls the trade-off between model complexity and prediction accuracy, and another parameter related to the kernel width and the Kernel bandwidth parameter, which represents the variance of the RBF kernel function, $g$. In SVR and ANN also, the exogenous variables are used for both modeling and forecasting purposes as in INGARCHX model.

### *2.2.4. Extreme Learning Machine (ELM)*

A single layer feed forward network with $x_1, x_2, \dots, x_m$ input nodes, $h_1, h_2, \dots, h_n$, hidden nodes and $t_i$ be the target node is shown in Fig.1 that was indicated by [14].

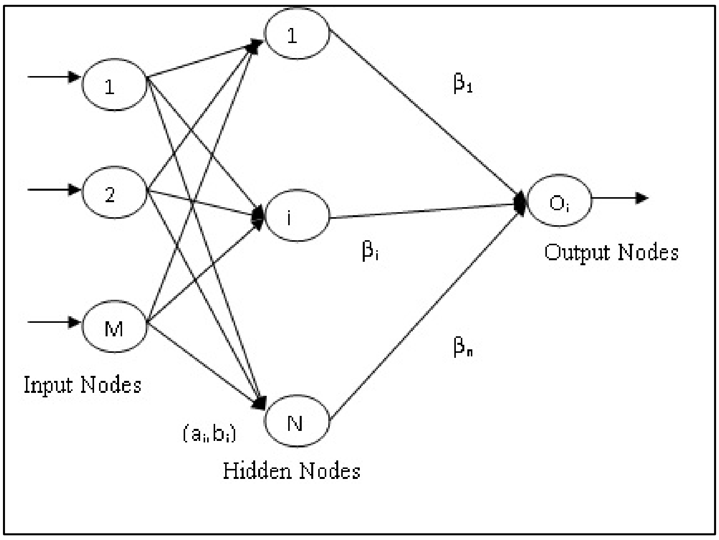

*Figure 1: Single layer feed forward network*

Let $(a_i, b_i)$ be the weights connecting from input layer to hidden layer and $\beta_1, \beta_2, \dots, \beta_n$be the weights of the nodes connecting from hidden layer to the output layer. Let "g" be the piecewise continuous activation function.

The hidden layer outputs are given as

$$\sum_{i=1}^{N}\left[\beta_i g\left(a_i, b_i, x_j\right)\right] = t_j, \quad j = 1, \dots, N$$

This equation can be rewritten as $\beta H \; = \; T$. Here $H$ is called the hidden layer output matrix, which can be expressed as follows,

$$H(a_1, \dots, a_N; b_1, \dots, b_N; x_1, \dots, x_M) = \begin{pmatrix} G(a_1, b_1, x_1) & \dots & G(a_N, b_N, x_1) \\ \vdots & \ddots & \vdots \\ G(a_1, b_1, x_M) & \dots & G(a_N, b_N, x_M) \end{pmatrix}$$

$\beta = [\beta_1 \beta_2 \dots \beta_n]^T$ and $T = [t_1 t_2 \dots t_n]^T$

### *2.2.5. INGARCH-ANN Hybrid Model*

The hybrid model combines INGARCH and ANN, where INGARCH captures linear temporal dependencies while ANN captures nonlinear relationships.

**Hybrid Model Formula:**

$$Y_t = \gamma_1 \lambda_t + \gamma_2 Y_t^{ANN} + \epsilon_t \dots (8)$$

Where, $\lambda_t$ is the INGARCH-predicted mean, $Y_t^{ANN}$ is the ANN-predicted value, $\gamma_1$ and$\gamma_2$ are weighting parameters, $\epsilon_t$ is the error term.

This model leverages the strengths of both INGARCH and ANN to improve predictive performance for count time series data

### *2.2.6. INGARCH-SVR Hybrid Model*

The INGARCH-SVR hybrid model combines INGARCH for linear dependencies and SVR for capturing complex nonlinear patterns.

**Hybrid Model Formula:**

$$Y_t = \gamma_1 \lambda_t + \gamma_2 Y_t^{SVR} + \epsilon_t \dots (9)$$

Where, $\lambda_t$ is the prediction from the INGARCH model, $Y_t^{SVR}$ is the prediction from the SVR model, $\gamma_1$ and$\gamma_2$ are weight coefficients, $\epsilon_t$ is the error term.

### *2.2.7. INGARCH-ELM Hybrid Model*

The hybrid model combines INGARCH to capture linear temporal dependencies and ELM to learn complex nonlinear patterns.

**Hybrid Model Formula:**

$$Y_t = \gamma_1 \lambda_t + \gamma_2 Y_t^{ELM} + \epsilon_t \dots (10)$$

Where, $\lambda_t$is the prediction from the INGARCH model, $Y_t^{ELM}$ is the prediction from the ELM model, $\gamma_1$ and$\gamma_2$are weighting parameters, $\epsilon_t$ is the error term.

### 2.3. Comparison Criteria

Mean Square Error (MSE) and Root Mean Square Error (RMSE) were used as comparison criteria for the model performance. The Mean Square Error (MSE) is the average of the sum of squared error values and given as:

$$MSE = \frac{\sum_{i=1}^{N}(Y_i - \widehat{Y}_i)^2}{N} \dots (11)$$

RMSE is also known as standard error of estimate in regression analysis, and is given as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Y_i - \widehat{Y}_i)^2}{N}} \dots (12)$$

where,$Y_i$ is the actual value, $\widehat{Y}_i$ is the predicted value, and $N$ is the number of observations.

### 2.4 Test for Autocorrelation and Nonlinearity

#### *2.4.1 Box-Pierce Test for Autocorrelation*

The Box-Pierce test is a diagnostic tool used to examine whether the residuals from a time series model are independently distributed. It specifically tests the null hypothesis that there is no autocorrelation (i.e., the residuals are white noise) up to a specified number of lags.

The Box-Pierce test statistic is defined as

$\chi^2 = n \sum_{k=1}^{m} \hat{r}_k^2$ ...(13)

Where:

- χ2 is the Box-Pierce statistic
- n is the sample size (number of residuals)
- m is the number of lags up to which autocorrelation is tested
- $\hat{r}_k$ Is the sample autocorrelation at lag k

The Box-Pierce test essentially accumulates the squared sample autocorrelations up to lag m, scaled by the sample size n, to detect any significant autocorrelation in the residuals.

#### *2.4.2 BDS (Brock-Dechert-Scheinman) test for non-linearity*

The BDS test is a non-parametric test of the null hypothesis that the data is independently and identically distributed (iid) against an unspecified alternative. The test enables one to test for nonlinear dependence because it is not affected by linear dependencies in the data.

### 2.4 Software used

The time series plots, INGARCH, ANN, ELM, SVR, INGARCH-ANN, INGARCH-SVR, and INGARCH-ELM models along with the Correlation analysis were carried out in R software.

## 3. Results and Discussion

### *3.1 Summary statistics*

The summary statistics of Yellow Stem Borer (YSB) count data across five locations: Nellore, Maruteru, Bapatla, Ragolu, and Nandyal during Kharif and Rabi seasons reveal significant spatial and temporal variability in pest populations. The mean YSB count was highest in Maruteru (Kharif: 532.17, Rabi: 1216.63), indicating that this location experiences the most severe infestations, particularly in the Rabi season. In contrast, Bapatla kharif (56.78), Nellore kharif(176), and Ragolu Rabi (25.60) recorded moderate to low mean populations, while Nandyal (kharif 6.76) exhibited the least infestation levels. The median YSB counts across all locations were substantially lower than the mean, signifying a highly skewed distribution where most observations had low YSB counts, but occasional outbreaks caused extreme values. This is further emphasized by the mode (0) for all locations, indicating that YSB was often absent for many weeks, reinforcing the erratic nature of infestations. The standard deviation (SD) and coefficient of variation (CV%) suggest a high degree of fluctuation in YSB populations across locations. Maruteru (Rabi) exhibited the highest variability (SD: 2376.84, CV%: 195.36), followed by Bapatla (Kharif: SD: 123.55, CV%: 217.59), implying that YSB infestations in these regions are unpredictable and subject to extreme spikes. The high CV% (>100%) across all locations indicates substantial population fluctuations, making forecasting challenging. The data also exhibited strong positive skewness, with values exceeding 6.9 in Bapatla and Maruteru, confirming that YSB populations are dominated by periods of low counts with occasional outbreaks of exceptionally high numbers. Furthermore, kurtosis values were extremely high, particularly in Maruteru (Kharif: 76.65) and Bapatla (Kharif: 65.04), signifying a highly peaked distribution with rare but severe infestations. The minimum values (0) across all locations highlight that YSB populations were absent for several weeks, while the quartile values (Q1 and Q3) demonstrate that 25% of observations recorded low counts, with the upper quartile (Q3) and maximum values revealing extreme outbreaks. The highest recorded infestation (20,648 in Maruteru Rabi) underscores the severity of these outbreaks and the need for an effective early warning system.

*Table 1: Descriptive statistics of the YSB data in all the five locations in kharif and rabi*

| Summary Statistics of YSB count data | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Location** | **Nellore** | | **Maruteru** | | **Bapatla** | **Ragolu** | | **Nandyal** |
| **Season** | **Kharif** | **Rabi** | **Kharif** | **Rabi** | **Kharif** | **Kharif** | **Rabi** | **Kharif** |
| Mean | 176 | 48.60 | 532.17 | 1216.63 | 56.78 | 11.16 | 25.60 | 6.76 |
| Median | 67 | 35 | 149.5 | 301 | 22.5 | 7 | 15.5 | 5 |
| Mode | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| SD | 303.55 | 53.48 | 1220.69 | 2376.84 | 123.55 | 13.53 | 26.22 | 7.05 |
| Skewness | 3.80 | 2.74 | 6.95 | 4.04 | 6.91 | 2.09 | 1.12 | 1.80 |
| Kurtosis | 22.89 | 16.67 | 76.65 | 24.75 | 65.04 | 8.72 | 3.56 | 6.83 |
| Minimum | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1$^{ST}$ Q | 22 | 10 | 38 | 91.75 | 8 | 1 | 4.25 | 2 |
| 3$^{rd}$ Q | 175 | 70 | 475 | 1174 | 66.75 | 15.75 | 39 | 9 |
| Maximum | 2635 | 455 | 17428 | 20648 | 1486 | 80 | 132 | 39 |
| CV% | 172.47 | 110.03 | 229.37 | 195.36 | 217.59 | 121.24 | 102.44 | 104.32 |

*Table 2. Pearson correlation coefficients between YSB populations and climatological variables*

| LOCATION | SEASON | | YSB | TMAX | TMIN | RF | RHM | RHE |
|---|---|---|---|---|---|---|---|---|
| BAPATLA | KHARIF | TMAX | -0.113*<br>(p=0.0306) | | | | | |
| | | TMIN | -0.184**<br>(p=0.0004) | -0.139**<br>(p=0.0080) | | | | |
| | | RF | -0.164**<br>(p=0.0017) | 0.072$^{NS}$<br>(p=0.1723) | 0.307**<br>(p=0.0000) | | | |
| | | RHM | 0.126*<br>(p=0.0162) | -0.311**<br>(p=0.0000) | -0.134*<br>(p=0.0104) | 0.076$^{NS}$<br>(p=0.1489) - | | |
| | | RHE | -0.048$^{NS}$<br>(p=0.3653) | 0.005$^{NS}$<br>(p=0.9216) | 0.267**<br>(p=0.0000) | 0.307**<br>(p=0.0000) - | -0.067$^{NS}$<br>(p=0.1989) - | |
| | | SSH | 0.250**<br>(p=0.0000) | -0.038$^{NS}$<br>(p=0.4733) | -0.205**<br>(p=0.0001) | -0.073$^{NS}$<br>(p=0.1618) - | -0.031$^{NS}$<br>(p=0.5498) - | -0.088$^{NS}$<br>(p=0.0932) |
| NANDYAL | KHARIF | TMAX | -0.289**<br>(p=0.0000) | | | | | |
| | | TMIN | -0.197**<br>(p=0.0029) | 0.410**<br>(p=0.0000) | | | | |
| | | RF | -0.078$^{NS}$<br>(p=0.2440) | -0.123$^{NS}$<br>(p=0.0656) | 0.317**<br>(p=0.0000) | | | |
| | | RHM | 0.114$^{NS}$<br>(p=0.0867) | -0.445**<br>(p=0.0000) | -0.177**<br>(p=0.0076) | 0.157*<br>(p=0.0181) - | | |
| | | RHE | 0.096$^{NS}$<br>(p=0.1531) | -0.326**<br>(p=0.0000) | 0.528**<br>(p=0.0000) | 0.364**<br>(p=0.0000) - | 0.311<br>(p=0.0000) - | |
| | | SSH | 0.039$^{NS}$<br>(p=0.5564) | 0.263<br>(p=0.0001) | -0.275<br>(p=0.0000) | -0.381**<br>(p=0.0000) - | -0.204**<br>(p=0.0021) - | -0.322 **<br>(p=0.0000) |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| NELLORE | KHARIF | TMAX | -0.045 $^{NS}$<br>(p=0.4508) | | | | | |
| | | TMIN | 0.034 $^{NS}$<br>(p=0.5622) | 0.601**<br>(p=0.0000) | | | | |
| | | RF | -0.096 $^{NS}$<br>(p=0.1043) | -0.147*<br>(p=0.0133) | 0.086 $^{NS}$<br>(p=0.1477) | | | |
| | | RHM | 0.083 $^{NS}$<br>(p=0.1600) | -0.501**<br>(p=0.0000) | -0.423**<br>(p=0.0000) | -0.321**<br>(p=0.0000) | | |
| | | RHE | -0.029 $^{NS}$<br>(p=0.6283) | -0.564**<br>(p=0.0000) | -0.258**<br>(p=0.0000) | 0.560**<br>(p=0.0000) | 0.252**<br>(p=0.0000) | |
| | | SSH | 0.019 $^{NS}$<br>(p=0.7459) | 0.340**<br>(p=0.0000) | 0.043 $^{NS}$<br>(p=0.4707) | -0.077 $^{NS}$<br>(p=0.1976) | -0.173**<br>(p=0.0034) | -0.218**<br>(p=0.0002) |
| | RABI | TMAX | 0.179**<br>(p=0.0042) | | | | | |
| | | TMIN | 0.052 $^{NS}$<br>(p=0.4051) | 0.231**<br>(p=0.0002) | | | | |
| | | RF | 0.112 $^{NS}$<br>(p=0.0740) | -0.162<br>(p=0.0094) | 0.005 $^{NS}$<br>(p=0.9373) | | | |
| | | RHM | 0.098 $^{NS}$<br>(p=0.1177) | -0.220**<br>(p=0.0004) | -0.222**<br>(p=0.0003) | 0.140<br>*(p=0.0252) | | |
| | | RHE | -0.064 $^{NS}$<br>(p=0.3117) | -0.381<br>(p=0.0000) | -0.440<br>(p=0.0000) | 0.411<br>(p=0.0000) | 0.436<br>(p=0.0000) | |
| | | SSH | 0.081 $^{NS}$<br>(p=0.1958) | 0.074 $^{NS}$<br>(p=0.2388) | -0.186**<br>(p=0.0028) | -0.353**<br>(p=0.0000) | 0.131*<br>(p=0.0372) | -0.262**<br>(p=0.0000) |
| MARUTERU | KHARIF | TMAX | -0.028 $^{NS}$<br>(p=0.4943) | | | | | |
| | | TMIN | -0.236**<br>(p=0.0000) | 0.527**<br>(p=0.0000) | | | | |
| | | RF | -0.147**<br>(p=0.0002) | -0.124**<br>(p=0.0021) | 0.143**<br>(p=0.0004) | | | |
| | | RHM | -0.020 $^{NS}$<br>(p=0.6145) | -0.334**<br>(p=0.0000) | -0.196**<br>(p=0.0000) | 0.203**<br>(p=0.0000) | | |
| | | RHE | -0.225**<br>(p=0.0000) | -0.370**<br>(p=0.0000) | 0.150**<br>(p=0.0002) | 0.233**<br>(p=0.0000) | 0.401**<br>(p=0.0000 | |
| | | SSH | 0.190 **<br>(p=0.000) | 0.212**<br>(p=0.0000) | -0.242**<br>(p=0.0000) | -0.342**<br>(p=0.0000) | -0.147**<br>(p=0.0003 | -0.357**<br>(p=0.0000) |
| | RABI | TMAX | 0.331**<br>(p=0.0000) | | | | | |
| | | TMIN | 0.343**<br>(p=0.0000) | 0.811**<br>(p=0.0000) | | | | |
| | | RF | 0.060 $^{NS}$<br>(p=0.2161) | 0.111*<br>(p=0.0235) | 0.162**<br>(p=0.0009) | | | |
| | | RHM | -0.180**<br>(p=0.0002) | -0.138**<br>(p=0.0048) | -0.215**<br>(p=0.0000) | -0.026 $^{NS}$<br>(p=0.5892) - | | |
| | | RHE | -0.339**<br>(p=0.0000) | -0.358**<br>(p=0.0000) | -0.156**<br>(p=0.0013) | 0.046 $^{NS}$<br>(p=0.3513) - | 0.152**<br>(p=0.0018 | |
| | | SSH | 0.147**<br>(p=0.0025) | 0.261**<br>(p=0.0000) | 0.131**<br>(p=0.0071) | -0.147**<br>(p=0.0026 | 0.021*<br>(p=0.6671) | -0.130**<br>(p=0.0075) |
| RAGOLU | KHARIF | TMAX | -0.138*<br>(p=0.0195) | | | | | |
| | | TMIN | -0.276**<br>(p=0.0000) | 0.140*<br>(p=0.0181) | | | | |
| | | RF | -0.091$^{NS}$<br>(p=0.1256) | -0.337**<br>(p=0.0000) | 0.018 $^{NS}$<br>(p=0.7599) | | | |
| | | RHM | -0.128*<br>(p=0.0303) | 0.069 $^{NS}$<br>(p=0.2454) | -0.010 $^{NS}$<br>(p=0.8659) | -0.080 $^{NS}$<br>(p=0.1771) | | |
| | | RHE | -0.120*<br>(p=0.0428) | -0.173**<br>(p=0.0034) | 0.180**<br>(p=0.0023) | 0.355**<br>(p=0.0000) | -0.120 *<br>(p=0.0434) | |
| | | SSH | 0.178 **<br>(p=0.0025) | 0.122*<br>(p=0.0394) | -0.153**<br>(p=0.0094) | -0.138*<br>(p=0.0193) | -0.134*<br>(p=0.0233) | -0.219**<br>(p=0.0002) |
| | RABI | TMAX | -0.093 $^{NS}$<br>(p=0.1164 | | | | | |
| | | TMIN | -0.159**<br>(p=0.0072) | 0.911**<br>(p=0.0000) | | | | |
| | | RF | -0.000 $^{NS}$<br>(p=0.9995) | 0.151<br>(p=0.0107) | 0.026 $^{NS}$<br>(p=0.6625) | | | |
| | | RHM | -0.110 $^{NS}$<br>(p=0.0620 | 0.836**<br>(p=0.0000) | 0.917**<br>(p=0.0000) | -0.008 $^{NS}$<br>(p=0.8893) | | |
| | | RHE | -0.100 $^{NS}$<br>(p=0.0929) | 0.892**<br>(p=0.0000) | 0.952**<br>(p=0.0000) | 0.172**<br>(p=0.0035) | 0.915**<br>(p=0.0000) | |
| | | SSH | -0.056 $^{NS}$<br>(p=0.3447) | -0.040 $^{NS}$<br>(p=0.4957) | 0.055 $^{NS}$<br>(p=0.3542) | -0.334**<br>(p=0.0000) | 0.072 $^{NS}$<br>(p=0.2271) | -0.018 $^{NS}$<br>(p=0.7638) |

**Significant at 5% level, ** significant at 1% level, NS-Not significant*

The below figure 2 shows the year-wise time series plots of the YSB population in five locations during two seasons *i.e* kharif and rabi except for Bapatla and Nandyal as the pest counts were recorded only for kharif season.

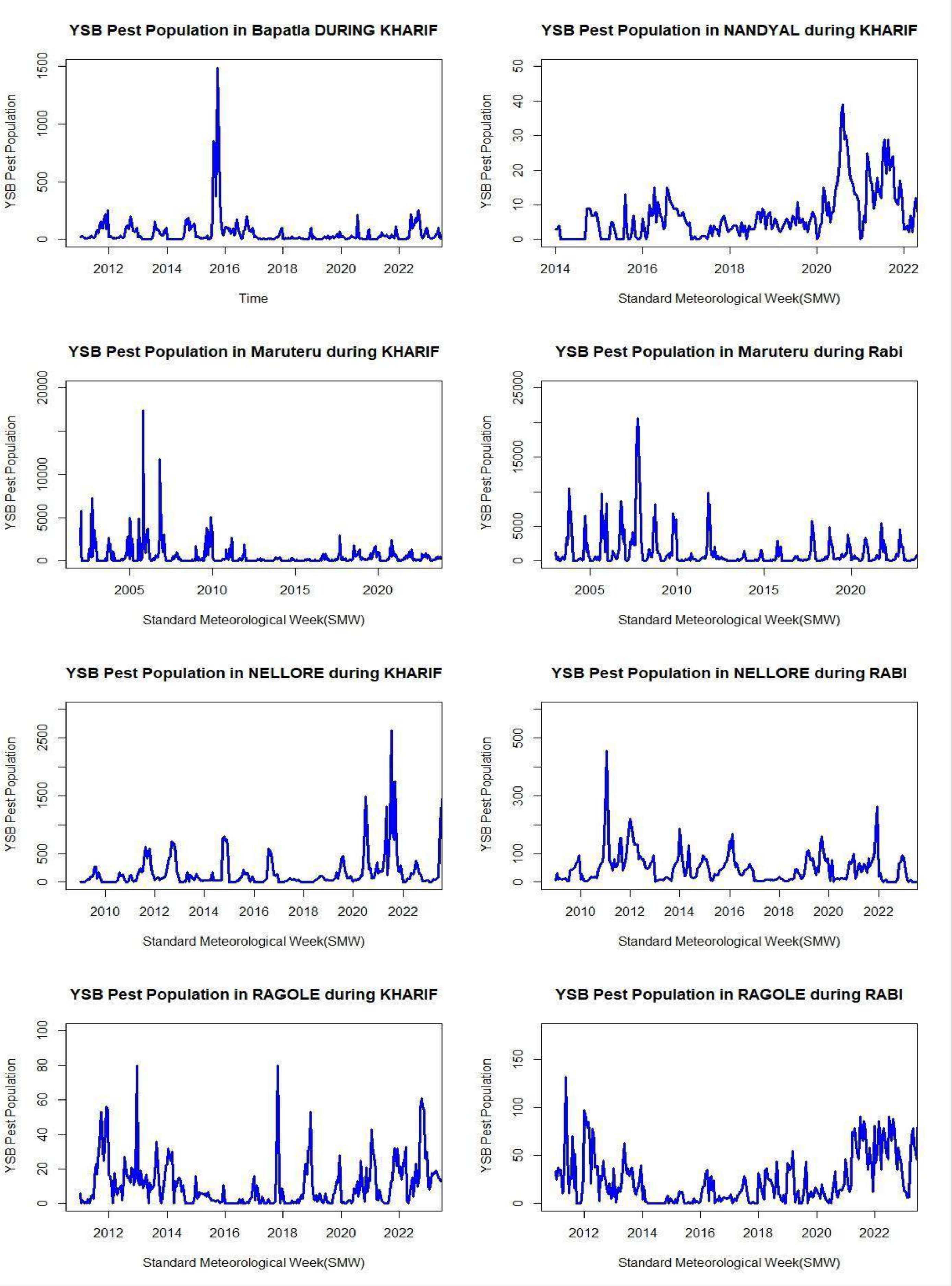

***Figure 2. Time series plots of YSB Population in five locations***

### *3.2 Correlation Analysis*

In Bapatla (BPT) during Kharif, YSB incidence showed a significant negative correlation with TMAX (r = -0.113, p = 0.0306), TMIN (r = -0.184, p = 0.0004), and RF (r = -0.164, p = 0.0017), indicating that higher temperatures and rainfall tend to reduce YSB infestation, while RHM had a significant positive correlation (r = 0.126, p = 0.0162), suggesting that increased morning humidity favors pest development. TMIN had a significant negative correlation with TMAX (r = -0.139, p = 0.0080), while RF was positively correlated with TMIN (r = 0.307, p = 0.0000). RHE showed significant positive correlations with TMIN (r = 0.267, p = 0.0000) and RF (r = 0.307, p = 0.0000), while SSH had a significant positive correlation with YSB (r = 0.250, p = 0.0000) and a negative correlation with TMIN (r = -0.205, p = 0.0001). However, SSH was not significantly correlated with TMAX, RHE, RF, RHM (p > 0.05), and RHE was not significantly correlated with YSB, TMAX, or RHM suggesting limited influence of evening humidity and sunshine hours on these weather parameters. In Nandyal (NDYL) YSB incidence in Kharif was significantly negatively correlated with TMAX (r = -0.289, *p* < 0.0001) and TMIN (r = -0.197, *p* = 0.0029), reinforcing the trend that higher temperatures suppress pest activity. However, TMIN had a strong positive correlation with RHE (r = 0.528, *p* < 0.0001), suggesting that warmer nights combined with high humidity might create favorable conditions for YSB survival. Rainfall did not exhibit a significant correlation (*p* > 0.05). In Nellore (NLR) during Kharif, TMAX did not show a significant correlation with YSB infestation (*p* = 0.4508), while TMIN displayed a strong positive correlation (r = 0.601, *p* < 0.0001) with TMAX, suggesting that warmer nights promote infestation. Rainfall showed inconsistent effects, with one significant negative correlation (r = -0.147, *p* = 0.0133) with TMAX. RHM and RHE are not significantly correlated with YSB incidence. In Rabi, TMAX was positively correlated (r = 0.179, *p* = 0.0042), suggesting that higher temperatures might favor infestation during the dry season. In Maruteru (MTU) during Kharif, TMAX did not significantly influence YSB (*p* = 0.4943), but TMIN showed a negative correlation (r = -0.236, *p* < 0.0001), suggesting that cooler nights reduce infestation. Rainfall displayed mixed effects, with both positive and negative correlations, depending on interaction with other climatic variables. Notably, RHE exhibited a significant negative correlation with YSB (r = -0.225, *p* < 0.0001), indicating that lower evening humidity may favor infestation. In Rabi, TMAX, TMIN, and SSH were positively correlated with YSB, suggesting that higher temperatures along with sunshine hours contribute to increased pest incidence in the dry season. RHM and RHE showed a negative significant correlation with YSB incidence. In Ragolu (RGL) the YSB incidence during Kharif was negatively correlated with TMAX (r = -0.138, *p* = 0.0195) and TMIN (r = -0.276, *p* < 0.0001), similar to other locations. RHM and RHE showed negatively significant correlations with YSB incidence. SSH shows a positive correlation with YSB. Rainfall did not significantly affect infestation (*p* > 0.05), while RHE and SSH exhibited a combination of significant positive and negative correlations, with other weather parameters suggesting an intricate influence of microclimatic factors on pest dynamics. In Rabi, TMIN had a negative correlation (r=-0.159,p = 0.0072) with YSB incidence. TMIN had a strong positive correlation (r = 0.911, *p* < 0.0001) with TMAX, reinforcing the hypothesis that higher nighttime temperatures favor infestation.

The results indicate that temperature, humidity, and rainfall play crucial roles in determining YSB incidence, with distinct seasonal and locational variations. Higher maximum (TMAX) and minimum (TMIN) temperatures are generally associated with reduced infestation during Kharif but contribute to increased infestation in Rabi. Relative humidity, both in the morning (RHM) and evening (RHE) exhibits complex interactions, with morning humidity often showing a positive correlation with YSB incidence. Rainfall has an inconsistent effect, with significant positive correlations in some locations while being negative or nonsignificant in others. Additionally, sunshine hours (SSH) display significant negative correlations with YSB in multiple locations, suggesting that increased solar radiation may suppress infestation levels. These findings highlight the intricate relationships between climatic factors and YSB dynamics, emphasizing the need for location-specific pest management strategies.

### *3.3 INGARCHX model using Negative Binomial distribution*

The INGARCHX model is like a multivariate regression model but allows one to take advantage of autocorrelation that may be present in residuals of the regression to improve the accuracy of a forecast. INGARCHX based on Negative Binomial distribution is performed for all major pests in both the seasons in five locations of Andhra Pradesh and results are summarized in table.3.

*Table 3. Parameter estimation of the INGARCHX model for YSB populations at study locations*

| Location | Season | Parameter | Estimate | S.E. | Z-value | P-value | Box pierce test of residuals |
|---|---|---|---|---|---|---|---|
| BAPATLA | KHARIF | (Intercept) | 2.55E+01 | 5.55E+01 | 0.4592 | 0.6461 | Chi-squared = 180.33, df = 1, p-value < 2.2e-16 |
| | | beta_2 | 1.03E-01 | 9.35E-02 | 1.1065 | 0.2685 | |
| | | alpha_2 | 3.95E-01 | 4.49E-01 | 0.8801 | 0.3788 | |
| | | TMAX | 4.51E-02 | 9.67E-01 | 0.0466 | 0.9628 | |
| | | TMIN | 1.83E-13 | 9.24E-01 | 0 | 1 | |
| | | RF | 6.11E-08 | 9.77E-02 | 0 | 1 | |
| | | RHM | 4.92E-07 | 4.25E-01 | 0 | 1 | |
| | | RHE | 2.51E-04 | 4.00E-01 | 0.0006 | 0.9995 | |
| | | SSH | 2.22E-02 | 1.85E+00 | 0.012 | 0.9904 | |
| | | sigmasq | 2.47E+00 | | | | |
| NANDYAL | KHARIF | (Intercept) | 3.02E+00 | 8.58E+00 | 0.3516 | 0.72515 | Chi-squared = 104.12, df = 1, p-value < 2.2e-16 |
| | | beta_1 | 2.06E-01 | 1.02E-01 | 2.0219 | 0.04318 | |
| | | alpha_1 | 2.00E-01 | 3.88E-01 | 0.5157 | 0.60606 | |
| | | TMAX | 1.42E-12 | 2.69E-01 | 0 | 1 | |
| | | TMIN | 3.78E-07 | 1.85E-01 | 0 | 1 | |
| | | RF | 9.55E-05 | 1.02E-02 | 0.0094 | 0.99251 | |
| | | RHM | 4.54E-08 | 6.30E-02 | 0 | 1 | |
| | | RHE | 1.10E-08 | 4.74E-02 | 0 | 1 | |
| | | SSH | 5.55E-02 | 1.68E-01 | 0.3295 | 0.74181 | |
| | | sigmasq | 5.55E-01 | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| NELLORE | KHARIF | (Intercept) | 8.22E+00 | 8.71E+01 | 0.0944 | 0.9248 | Chi-squared = 8.132, df = 1, p-value = 0.004349 |
| | | beta_1 | 8.27E-01 | 1.87E-01 | 4.416 | 1.01E-05 | |
| | | alpha_1 | 9.84E-03 | 1.08E-01 | 0.0911 | 0.9274 | |
| | | TMAX | 9.64E-03 | 2.88E+00 | 0.0033 | 0.9973 | |
| | | TMIN | 5.36E-03 | 3.52E+00 | 0.0015 | 0.9988 | |
| | | RF | 5.20E-10 | 7.64E-02 | 0 | 1 | |
| | | RHM | 1.22E-01 | 3.11E-01 | 0.3917 | 0.6953 | |
| | | RHE | 2.39E-05 | 4.69E-01 | 0.0001 | 1 | |
| | | SSH | 1.27E+00 | 2.24E+00 | 0.5693 | 0.5692 | |
| | | sigmasq | 1.33E+00 | | | | |
| | RABI | (Intercept) | 3.85E-05 | 7.85E+01 | 0 | 1 | Chi-squared = 94.251, df = 1, p-value < 2.2e-16 |
| | | beta_4 | 3.33E-01 | 1.12E-01 | 2.9713 | 0.002965 | |
| | | alpha_4 | 5.53E-02 | 1.68E-01 | 0.3289 | 0.742237 | |
| | | TMAX | 5.23E-01 | 1.90E+00 | 0.2758 | 0.782732 | |
| | | TMIN | 2.18E-05 | 3.76E-01 | 0.0001 | 0.999954 | |
| | | RF | 1.34E-01 | 1.05E-01 | 1.2758 | 0.202025 | |
| | | RHM | 2.81E-08 | 5.72E-01 | 0 | 1 | |
| | | RHE | 1.55E-06 | 5.10E-01 | 0 | 0.999998 | |
| | | SSH | 2.11E+00 | 1.20E+00 | 1.7645 | 0.077649 | |
| | | sigmasq | 1.11E+00 | | | | |
| MARUTERU | KHARIF | (Intercept) | 4.04E+02 | 1.31E+03 | 0.3084 | 0.7578 | Chi-squared = 163.84, df = 1, p-value < 2.2e-16 |
| | | beta_4 | 5.84E-02 | 7.16E-02 | 0.8154 | 0.4148 | |
| | | alpha_4 | 1.26E-01 | 7.59E-01 | 0.1664 | 0.8679 | |
| | | TMAX | 2.88E-08 | 2.94E+01 | 0 | 1 | |
| | | TMIN | 9.46E-08 | 2.24E+01 | 0 | 1 | |
| | | RF | 4.46E-08 | 1.00E+00 | 0 | 1 | |
| | | RHM | 9.97E-09 | 1.15E+01 | 0 | 1 | |
| | | RHE | 1.65E-07 | 5.98E+00 | 0 | 1 | |
| | | SSH | 8.84E+00 | 2.70E+01 | 0.328 | 0.7429 | |
| | | sigmasq | 4.55E+00 | | | | |
| | RABI | (Intercept) | 7.07E+02 | 1.97E+03 | 0.3587 | 0.7198214 | Chi-squared = 131.35, df = 1, p-value < 2.2e-16 |
| | | beta_2 | 4.55E-01 | 1.38E-01 | 3.2974 | 0.0009758 | |
| | | alpha_2 | 2.03E-07 | 1.30E-01 | 0 | 0.9999987 | |
| | | TMAX | 4.67E-05 | 5.34E+01 | 0 | 0.9999993 | |
| | | TMIN | 8.00E-01 | 4.63E+01 | 0.0173 | 0.9862125 | |
| | | RF | 2.17E+01 | 1.41E+01 | 1.5344 | 0.1249388 | |
| | | RHM | 4.35E-09 | 1.65E+01 | 0 | 1 | |
| | | RHE | 3.96E-08 | 7.16E+00 | 0 | 1 | |
| | | SSH | 2.61E+00 | 6.51E+01 | 0.04 | 0.9680833 | |
| | | sigmasq | 1.97E+00 | | | | |
| RAGOLU | KHARIF | (Intercept) | 1.09E+01 | 5.49E+03 | 0.002 | 0.9984 | Chi-squared = 110.49, df = 1, p-value < 2.2e-16 |
| | | beta_3 | 8.29E-05 | 6.68E-02 | 0.0012 | 0.999 | |
| | | alpha_3 | 7.45E-05 | 5.02E+02 | 0 | 1 | |
| | | TMAX | 9.34E-08 | 4.63E-01 | 0 | 1 | |
| | | TMIN | 6.12E-13 | 2.19E-01 | 0 | 1 | |
| | | RF | 5.12E-07 | 1.85E-02 | 0 | 1 | |
| | | RHM | 1.82E-08 | 2.94E-02 | 0 | 1 | |
| | | RHE | 3.15E-07 | 7.34E-02 | 0 | 1 | |
| | | SSH | 3.99E-04 | 3.00E-01 | 0.0013 | 0.9989 | |
| | | sigmasq | 1.52E+00 | | | | |
| | RABI | (Intercept) | 2.44E+01 | 5.56E+04 | 4.00E-04 | 0.9996 | Chi-squared = 147.51, df = 1, p-value < 2.2e-16 |
| | | beta_1 | 3.96E-05 | 9.00E-02 | 4.00E-04 | 0.9996 | |
| | | alpha_1 | 8.61E-06 | 2.27E+03 | 0.00E+00 | 1 | |
| | | TMAX | 3.19E-06 | 1.22E-01 | 0.00E+00 | 1 | |
| | | TMIN | 1.72E-11 | 3.25E-01 | 0.00E+00 | 1 | |
| | | RF | 1.57E-05 | 1.46E-01 | 1.00E-04 | 0.9999 | |
| | | RHM | 1.43E-11 | 5.43E-02 | 0.00E+00 | 1 | |
| | | RHE | 6.25E-14 | 1.33E-01 | 0.00E+00 | 1 | |
| | | SSH | 1.28E-07 | 1.22E+00 | 0.00E+00 | 1 | |
| | | sigmas | 1.10E+00 | | | | |

The INGARCHX model was fitted to analyze the time series data incorporating exogenous climatological variables. The estimated model parameters were found to be significant at various locations and seasons, with β coefficients indicating a strong dependence of current values on past observations. However, none of the climatological variables (TMAX, TMIN, RF, RHM, RHE, SSH) were consistently significant across all hot-spot locations, suggesting their minimal contribution to the variation in the dependent variable. The over-dispersion parameters varied across locations and seasons, confirming the heterogeneous and over-dispersed nature of the data, following a Poisson or negative binomial distribution. The estimated dispersion values were 2.47 (Bapatla-Kharif), 0.55 (Nandyal-Kharif), 1.33 (Nellore-Kharif), 1.11 (Nellore-Rabi), 4.55 (Maruteru-Kharif), 1.97 (Maruteru-Rabi), 1.52 (Ragolu-Kharif), and 1.10 (Ragolu-Rabi) (Table 3). Diagnostic checking of residuals using the Box-Pierce test revealed that residuals were significantly autocorrelated ($p < 2.2 \times 10^{-16}$) at most locations, indicating that the model did not fully account for all dependence structures.

Notably, for Kharif at Nellore, the residuals showed relatively lower autocorrelation (p = 0.0043), suggesting a comparatively better model fit in that season at this location. However, in Rabi at Nellore ($p < 2.2 \times 10^{-16}$), Kharif and Rabi at Maruteru ($p < 2.2 \times 10^{-16}$), and Kharif and Rabi at Ragolu ($p < 2.2 \times 10^{-16}$), the residuals exhibited significant autocorrelation, indicating potential model refinements are necessary (Table 3). Overall, while the INGARCHX model successfully captured over-dispersion and temporal dependence in the data, further improvements may be required to eliminate residual autocorrelation, possibly through additional covariates, interaction terms, or model adjustments.

### *3.4 Machine Learning models and Two-stage modeling (Hybrid models)*

In this study, in addition to INGARCH, Artificial neural network (ANN), Support vector regression (SVR), and Extreme learning machine (ELM) models and the two-stage models like INGARCH-ANN, INGARCH-SVR and INGARCH-ELM were developed to forecast pest populations. The two-stage methodology combines both significant original count time series linear and nonlinear significant residual components to provide an aggregate forecast. As explained in the methodology section, the first step was to test the autocorrelation in the residuals by the Box-Pierce test along with conformation of non-linearity by the BDS test. The Box-Pierce tests revealed that residuals obtained by INGRACH models are autocorrelated and are nonlinear also as confirmed by the BDS test. As the residuals were nonlinear, they were predicted using nonlinear and machine-learning models like ANN, SVR, and ELM. The ANN, SVR, and ELM were used for forecasting INGARCH residuals in this study separately. The residuals predicted these models were combined with the predicted values obtained from INGARCH models separately and the residual analysis of the all models was summarized in table.4.

*Table 4. Comparison criteria for different models for YSB populations in training and testing data sets*

| Location | Season | Models | Train | | Test | | Box pierce test for residuals |
|---|---|---|---|---|---|---|---|
| | | | MSE | RMSE | MSE | RMSE | |
| BAPATLA | KHARIF | INGARCH | 13616.55 | 116.69 | 521.92 | 22.85 | p-value =2.2e-16 |
| | | ANN | 967.2 | 31.1 | 2218.23 | 47.1 | p-value = 0.1568 |
| | | **SVR** | **1.54** | **1.24** | **15427.08** | **124.21** | **p-value = 0.3441** |
| | | ELM | 14281.97 | 119.51 | 3297.55 | 57.42 | p-value = 2.2e-16 |
| | | INGARCH-ANN | 1105.02 | 33.24 | 396.94 | 19.92 | p-value = 0.07092 |
| | | INGARCH-SVR | 1.13 | 1.06 | 0.51 | 0.71 | p-value = 1.012e-10 |
| | | INGARCH ELM | 0.06 | 0.24 | 0.01 | 0.1 | p-value = 5.889e-09 |
| NANDYAL | KHARIF | INGARCH | 34.11 | 5.84 | 14.52 | 3.81 | p-value = 2.2e-16 |
| | | ANN | 3.46 | 1.86 | 155.71 | 12.48 | p-value = 0.7647 |
| | | SVR | 6.23 | 2.5 | 47.98 | 6.93 | p-value = 6.151e-14 |
| | | ELM | 43.07 | 6.56 | 15.58 | 3.95 | p-value =2.2e-16 |
| | | INGARCH-ANN | 8.62 | 2.94 | 17.15 | 4.14 | p-value = 0.2263 |
| | | INGARCH-SVR | 0 | 0.05 | 0 | 0.04 | p-value = 0.02407 |
| | | **INGARCH ELM** | **0** | **0.01** | **0** | **0.01** | **p-value = 0.2315** |
| MARUTERU | KHARIF | INGARCH | 1468591 | 1211.86 | 41123.54 | 202.79 | p-value = 2.2e-16 |
| | | ANN | 336206.8 | 579.83 | 301257.4 | 548.87 | p-value = 0.7341 |
| | | **SVR** | **163.42** | **12.78** | **1412822** | **1188.62** | **p-value = 0.08894** |
| | | ELM | 1354884 | 1164 | 442851.8 | 665.47 | p-value = 2.2e-16 |
| | | INGARCH-ANN | 221917.5 | 471.08 | 324462.1 | 569.62 | p-value = 0.7335 |
| | | INGARCH-SVR | 6628.13 | 81.41 | 33.4 | 5.78 | p-value = 0.7871 |
| | | INGARCH ELM | 306461.7 | 553.59 | 3050.86 | 55.23 | p-value = 0.002149 |
| | RABI | INGARCH | 4274548 | 2067.5 | 535736.4 | 731.94 | p-value = 2.2e-16 |
| | | ANN | 536039.7 | 732.15 | 8690250 | 2947.92 | p-value = 2.685e-06 |
| | | SVR | 3249.9 | 57.01 | 93169405 | 9652.43 | p-value = 0.9665 |
| | | ELM | 4053621 | 2013.36 | 4044693 | 2011.14 | p-value = 2.2e-16 |
| | | INGARCH-ANN | 405145.9 | 636.51 | 1744343 | 1320.74 | p-value = 0.4223 |
| | | **INGARCH-SVR** | **335.39** | **18.31** | **80.08** | **8.95** | **p-value = 0.05765** |
| | | INGARCH ELM | 727278.1 | 852.81 | 2029.85 | 45.05 | p-value =2.2e-16 |
| NELLORE | KHARIF | INGARCH | 48530.82 | 220.3 | 494889.6 | 703.48 | p-value = 0.004349 |
| | | ANN | 3162.55 | 56.24 | 303051 | 550.5 | p-value = 0.9506 |
| | | SVR | 458.38 | 21.41 | 495224.4 | 703.72 | p-value = 5.551e-16 |
| | | ELM | 74915.07 | 273.71 | 446256.2 | 668.02 | p-value =2.2e-16 |
| | | INGARCH-ANN | 10287.82 | 101.43 | 491463.3 | 701.04 | p-value = 0.5367 |
| | | INGARCH-SVR | 57.89 | 7.61 | 52912.65 | 230.03 | p-value = 0.8481 |
| | | **INGARCH ELM** | **21.13** | **4.6** | **65.42** | **8.09** | **p-value = 0.5914** |
| | RABI | INGARCH | 2718.35 | 52.14 | 580.72 | 24.1 | p-value = 2.2e-16 |
| | | ANN | 116.13 | 10.78 | 714.65 | 26.73 | p-value = 0.09557 |
| | | SVR | 62.59 | 7.91 | 947.7 | 30.78 | p-value = 2.2e-16 |
| | | ELM | 2614.09 | 51.13 | 1267.09 | 35.6 | p-value = 2.2e-16 |
| | | INGARCH-ANN | 245.93 | 15.68 | 807.5 | 28.42 | p-value = 0.2165 |
| | | INGARCH-SVR | 0.79 | 0.89 | 0.25 | 0.5 | p-value = 0.9744 |
| | | **INGARCH ELM** | **0.15** | **0.38** | **0.09** | **0.31** | **p-value = 0.7076** |
| RAGOLU | KHARIF | INGARCH | 185.86 | 13.63 | 89.87 | 9.48 | p-value = 2.2e-16 |
| | | ANN | 73.6 | 8.58 | 199.22 | 14.11 | p-value = 0.2917 |
| | | SVR | 0.02 | 0.14 | 743.49 | 27.27 | p-value = 0.1899 |
| | | ELM | 128.87 | 11.35 | 44.75 | 6.69 | p-value = 1.414e-13 |
| | | INGARCH-ANN | 99 | 9.95 | 38.42 | 6.2 | p-value = 0.6544 |
| | | INGARCH-SVR | 0.01 | 0.11 | 0.01 | 0.09 | p-value = 0.08588 |
| | | **INGARCH ELM** | **0** | **0.02** | **0** | **0.02** | **p-value = 0.8702** |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| RAGOLU | RABI | INGARCH | 660.27 | 25.7 | 1422.27 | 37.71 | p-value = 2.2e-16 |
| | | **ANN** | **195.92** | **13.99** | **21.81** | **4.67** | **p-value = 0.8302** |
| | | SVR | 216.23 | 14.7 | 103.53 | 10.18 | p-value = 2.692e-08 |
| | | ELM | 586.3 | 24.21 | 1406.72 | 37.51 | p-value = 2.2e-16 |
| | | INGARCH-ANN | 252.28 | 15.88 | 224.46 | 14.98 | p-value = 0.8827 |
| | | INGARCH-SVR | 0.04 | 0.2 | 0.03 | 0.18 | p-value = 0.8827 |
| | | INGARCH ELM | 0 | 0.05 | 0 | 0.07 | p-value = 0.03749 |

The results of modeling and predicting yellow stem borer (YSB) populations at different study locations were evaluated using MSE and RMSE for both training and testing datasets. Among the various models that were tested in INGARCH-ELM consistently demonstrated superior predictive performance across most of the locations and seasons, achieving the lowest error values and effectively capturing the complex temporal dependencies of YSB population dynamics. In Bapatla (Kharif ) the SVR model outperformed the other models as it has the lowest performance metrics along with no autocorrelation among the residuals. Whereas, in Nandyal (Kharif), Nellore (Kharif, Rabi), and Ragolu (Kharif) INGARCH-ELM outperformed the other models with no significant autocorrelation among the residuals and lowest RMSE and MSE values. In Ragolu (Rabi) among the models with no autocorrelation in the residuals, ANN performed better with the lowest error metrics(RMSE, MSE). In Maruteru (kharif) among the models with no autocorrelation in the residuals SVR model had the lowest performance metrics. In contrast, for Maruteru (Rabi) INGARCH-SVR was the best model having the lowest error metrics and no autocorrelation among the residuals. Overall, these findings suggest that hybrid approaches integrating INGARCH with machine learning techniques, particularly extreme learning machines (ELM), significantly enhance predictive accuracy, supporting more effective pest management strategies and data-driven decision-making in integrated pest control programs. Further, only those models with no autocorrelation among residuals and lowest performance metrics were considered as the best model which are mentioned in table 5.

*Table 5: Best models along with Box Pierce Test values on their residuals*

| Research Station | Season | Years of data availability | SMW | Frequency of SMW | Total no of observations | Training data set | Testing data set | Best model | Training data set | | Testing data set | | Box-Pierce test for autocorrelation on residuals | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | RMSE | MSE | RMSE | MSE | Chi-square | p-value |
| Nellore(NLR) | Kharif | 2009-2023 | 27 to 45 | 19 | 285 | 0.232639 | 276:285 | NBINGARCH-ELM | 4.6 | 21.13 | 8.09 | 65.42 | 0.288 | 0.591 |
| | Rabi | 2009-2023 | 46 to 10 | 17 | 255 | 0.211806 | 246:255 | NBINGARCH-ELM | 0.38 | 0.15 | 0.31 | 0.09 | 0.141 | 0.707 |
| Ragolu(RGL) | Kharif | 2011-2023 | 26 to 47 | 22 | 286 | 0.233333 | 277:286 | NBINGARCH-ELM | 0.02 | 0 | 0.02 | 0 | 0.027 | 0.870 |
| | Rabi | 2011-2023 | 48 to 17 | 22 | 286 | 0.233333 | 277:286 | ANN | 13.99 | 195.92 | 4.67 | 21.81 | 0.15 | 0.690 |
| Maruteru(MTU) | Kharif | 2002-2023 | 25 to 52 | 28 | 616 | 0.4625 | 607:616 | SVR | 12.78 | 163.42 | 1188.62 | 1412822 | 2.893 | 0.089 |
| | Rabi | 2003-2023 | 1 to 20 | 20 | 420 | 0.326389 | 411:420 | NBINGARCH-SVR | 18.31 | 335.39 | 8.95 | 80.08 | 3.6037 | 0.057 |
| Bapatla(BPT) | Kharif | 2011-2023 | 32 to 7 | 28 | 364 | 0.2875 | 355:364 | SVR | 1.24 | 1.54 | 124.21 | 15427.08 | 0.895 | 0.344 |
| Nandyal(NDL) | Kharif | 2014-2022 | 33 to 5 | 25 | 225 | 0.190972 | 216:225 | NBINGARCH-ELM | 0.01 | 0 | 0.01 | 0 | 1.432 | 0.232 |

The results presented in table 5 highlight the best-performing models for forecasting Yellow Stem Borer (YSB) populations at various research stations and seasons, along with their corresponding error metrics and residual autocorrelation tests. The study considered multiple years of data from different locations, including Nellore (NLR), Ragolu (RGL), Maruteru (MTU), Bapatla (BPT), and Nandyal (NDL) for both Kharif and Rabi seasons. In Nellore (NLR), the NBINGARCH-ELM model was the most effective in both seasons. For Kharif, it achieved a training RMSE of 4.60 and MSE of 21.13, while in testing, the RMSE and MSE increased to 8.09 and 65.42, respectively, indicating a moderate rise in error. In Rabi, the model performed exceptionally well, with a training RMSE of 0.38 and MSE of 0.15, and a testing RMSE of 0.31 and MSE of 0.09, demonstrating strong predictive accuracy. The Box-Pierce test for residual autocorrelation confirmed no significant autocorrelation in both seasons, as indicated by p-values of 0.60 (Kharif) and 0.71 (Rabi).

For Ragolu (RGL), the NBINGARCH-ELM model, ANN emerged as the best model in kharif and rabi season respectively. In Kharif, it achieved an RMSE of 0.02 and MSE of 0.00 for training, and an RMSE of 0.02 and MSE of 0.00 for testing, reflecting minimal error. In Rabi, the model's performance was slightly lower but still effective, with a training RMSE of 13.99 and MSE of 195.92, and a testing RMSE of 4.67 and MSE of 21.81. The Box-Pierce test results suggested no significant autocorrelation in the Kharif season (p = 0.870), but mild autocorrelation was detected in Rabi (p = 0.83). At Maruteru (MTU), the results varied by season. In Kharif, the SVR model was the best performer in training, achieving an RMSE of 12.78 and MSE of 163.42. However, in testing, the RMSE drastically increased to 1188.62 and MSE to 1,412,822, suggesting potential overfitting. In Rabi, the NBINGARCH-SVR model was the best model, achieving a training RMSE of 18.31 and MSE of 335.39, and a testing RMSE of 8.95and MSE of 80.08. The Box-Pierce test indicated no residual autocorrelation in Rabi (p = 0.057), whereas in Kharif (p = 0.089), the residuals were mostly uncorrelated. For Bapatla (BPT), the SVR model was the best-performing model in Kharif, with a training RMSE of 1.24 and MSE of 1.54, and a testing RMSE of 124.21 and MSE of 15427.08. However, the Box-Pierce test revealed no significant autocorrelation in residuals (p = 0.344).

At Nandyal (NDL), the NBINGARCH-ELM model achieved the best results in Kharif, with a training RMSE of 0.01 and MSE of 0.00, and a testing RMSE of 0.01 and MSE of 0.00. The Box-Pierce test showed no significant autocorrelation (p = 0.232), confirming the model's reliability in this location. Overall, the findings indicate that the NBINGARCH-ELM model consistently outperformed other models, particularly in Nellore, Ragolu, Bapatla, and Nandyal, while the NBINGARCH-SVR model performed well in Maruteru (Rabi season). The SVR model, despite its low training error in Maruteru (Kharif season), exhibited a large gap between training and testing errors, suggesting overfitting. The Box-Pierce test results across locations, showed no significant autocorrelation. Ultimately, these findings reinforce that hybrid NBINGARCH-based models effectively capture YSB population dynamics, making them valuable tools for pest management and forecasting in different agricultural environments

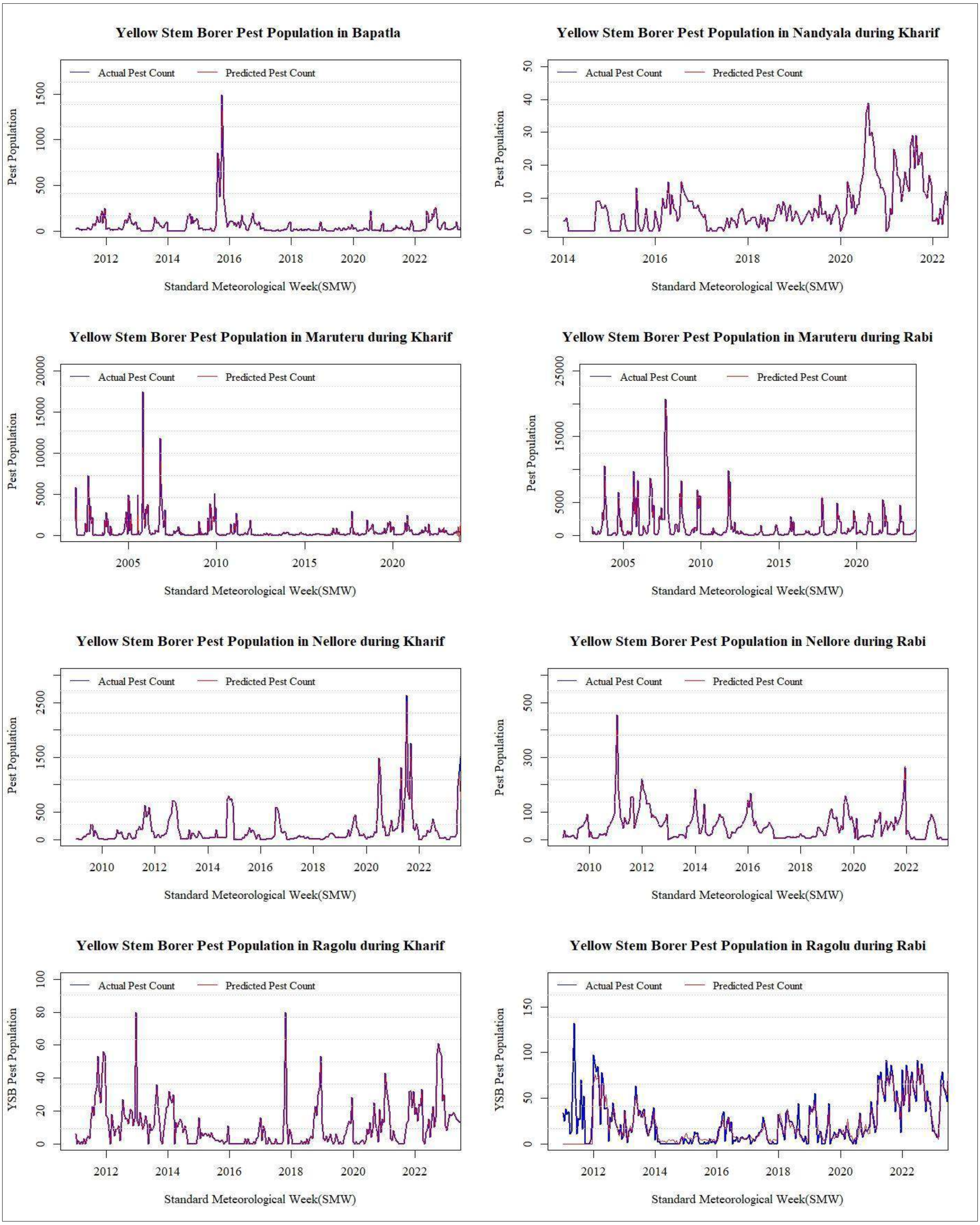

*Figure 3. Actual vs. fitted plots of YSB population*

## Conclusion

The study successfully integrated statistical and machine learning approaches to forecast Yellow Stem Borer (YSB) populations across various rice-growing locations in Andhra Pradesh. The results demonstrated that hybrid models (NBINGARCH-ELM, NBINGARCH-SVR), particularly NBINGARCH-ELM, consistently outperformed standalone models in terms of predictive accuracy. The analysis highlighted the strong influence of climatic factors such as temperature, humidity, and rainfall on YSB dynamics, with notable variations across seasons and locations. While the SVR model showed promising performance in specific cases, it also exhibited signs of overfitting, reinforcing the need for hybrid approaches. The Box-Pierce test confirmed no residual autocorrelation, validating the reliability of the selected models. These findings emphasize the potential of hybrid statistical-ML models in improving pest forecasting, which can aid in timely and effective pest management strategies.

## Future Scope:

Future research can enhance model precision by integrating additional environmental variables such as wind speed and soil conditions, as well as agronomic factors like sowing dates, crop stages, and varietal resistance. Further, the development of real-time forecasting systems and mobile-based advisory tools can bridge the gap between model predictions and farmer-level decision-making.

## Conflict of Interest:

The authors declare no conflict of interest.

## Acknowledgements

The authors are indebted to the Sree Padmavathi Venkateswara Foundation, Vijayawada, Andhra Pradesh for their financial support through Sri Ramakrishna Paramahamsa Research Grants for Agriculture in conducting this research.

## References

1. Christou, V.; Fokianos, K. (2014) Quasi-Likelihood Inference for Negative Binomial Time Series Models. *J. Time Ser. Anal.* , 35, 55–78.

2. Ferland, R., Latour, A., & Oraichi, D. (2006). Integer-valued GARCH process. *Journal of Time Series Analysis, 27*(6), 923–942.

3. Fokianos, K. (2011) Some Recent Progress in Count Time Ser. *Statistics* , 45, 49–58.

4. Fokianos, K.; Rahbek, A.; Tjøstheim, D. (2009) Poisson autoregression. *J. Am. Stat. Assoc.*, 104, 1430–1439

5. Ferland, R.; Latour, A.; Oraichi, D.(2006) Integer-valued GARCH process. *J. Time Ser. Anal,*27, 923–942.

6. Heinen, A. (2003). Modelling time series count data: An autoregressive conditional Poisson model (*MPRA Paper No. 8113*). University Library of Munich. Retrieved from https://mpra.ub.uni-muenchen.de/8113/

7. Kedem, B., & Fokianos, K. (2002). *Regression models for time series analysis*. Wiley-Interscience.

8. Liboschik, T., Fried, R., Fokianos, K., & Probst, P. (2020). *tscount: An R package for analysis of count time series following generalized linear models (Version 1.4.3)*. Retrieved from https://CRAN.R-project.org/package=tscount

9. Manikandan, N., Kennedy, J. S., & Geethalakshmi, V. (2013). Effect of elevated temperature on development time of rice yellow stem borer. *Indian Journal of Science and Technology, 6*(12), 5563–5566.

10. P Minruhi, P Lavanya Kumari, Santosha Rathod, B Ramana Murthy, K Devaki. Statistical evaluation of stepwise regression method for earwig population in groundnut (*Arachis hypogaea* L.). *Pharma Innovation* 2023;12(12):397-403.

11. Rathod, S., & Mishra, G. C. (2018). Statistical models for forecasting mango and banana yield of Karnataka. *Indian Journal of Agricultural Science and Technology, 20*, 803–816.

12. Rathod, S., Yerram, S., Arya, P., Katti, G., Rani, J., Padmakumari, A. P., Somasekhar, N., Padmavathi, C., Ondrasek, G., Amudan, S., Malathi, S., Rao, N. M., Karthikeyan, K., Mandawi, N., Muthuraman, P., & Sundaram, R. M. (2022). Climate-based modeling and prediction of rice gall midge populations using count time series and machine learning approaches. *Agronomy, 12*(1), 22.

13. Reddy, B. N. K., Rathod, S., Kallakuri, S., Sridhar, Y., Admala, M., Malathi, S., Pandit, P., & Jyostna, B. (2022). Modelling the relationship between weather variables and rice yellow stem borer population: A count data modelling approach. *International Journal of Environment and Climate Change, 12*(11), 3623–3632.

14. S, K. C., Mahendran, S., & Natarajan, S. (2016). Forecasting gold prices based on extreme learning machine. *International Journal of Computers Communications & Control, 11*, 372.

15. Vapnik, V. N. (1995). *The nature of statistical learning theory*. Springer.

16. Zhang, G. P. (2003). Time-series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing, 50*, 159–175.

17. Zhu, F. (2012). Modeling time series of counts with COM-Poisson INGARCH models. *Mathematical and Computer Modelling, 56*(1-2), 191–203.