The Pharma Innovation Journal 2023; 12(12): 397-403

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277-7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2023; 12(12): 397-403

www.thepharmajournal.com
Received: 15-10-2023
Accepted: 18-11-2023

P Minruhi
ANGRAU, S.V Agricultural College, Tirupati, Andhra Pradesh, India

P Lavanya Kumari
ANGRAU, S.V Agricultural College, Tirupati, Andhra Pradesh, India

Santosha Rathod
ICAR-Indian Institute of Rice Research, Hyderabad, Telangana, India

B Ramana Murthy
ANGRAU-SMGR Agricultural College, Udayagiri, Andhra Pradesh, India

K Devaki
ANGRAU-Regional Agricultural Research Station, Tirupati, Andhra Pradesh, India

Corresponding Author:
P Minruhi
ANGRAU, S.V Agricultural College, Tirupati, Andhra Pradesh, India

Statistical evaluation of stepwise regression method for earwig population in groundnut (*Arachis hypogaea* L.)

P Minruhi, P Lavanya Kumari, Santosha Rathod, B Ramana Murthy and K Devaki

Abstract

Background: Groundnut (*Arachis hypogaea* L.) is a widely cultivated oilseed crop with significant economic and nutritional value. Among the pests, Earwig (*Anisolabis stali*) are nocturnal insects known to feed on groundnut plants, causing damage that can result in yield losses and reduced crop quality. The matured pods were infested to an extent of 44.1% and the intensity of infestation in the case of immature pods was 52.1% in *Asirya muitunde*, a variety of groundnut. In any crop, weather parameters are one of the major causes of pest infestation. Further, which weather parameter influences the best at higher extent need to be known before prediction of pest as per the weather parameters. To effectively manage earwig populations in groundnut fields, it is essential to develop accurate predictive models that can assist farmers in implementing targeted pest control strategies. Stepwise regression is one of the statistical methods that was commonly employed in agricultural research for building predictive models.

Methods: (i) Data: Secondary data on Earwig (EW) light trap catches during *kharif* was collected from Regional Agricultural Research Station (RARS) Tirupati, for the period of 15 years, from 2008 to 2022 during *Kharif* from 26 SMWs (June) to 43SMWs (October). Weather parameters *viz.*, Maximum temperature (MAXT), Minimum temperature (MINT), Rainfall (RF), Morning Relative Humidity (RHM), Evening Relative Humidity (RHE), Sunshine hours (SSH) and Wind speed (WS) for the respective standard meteorological weeks (SMWs) also collected from Automated Weather Station (AWS) situated at RARS Tirupati.

(ii) Statistical analysis: In the stepwise regression model, the weekly average values of MAXT, MINT, RHM, RHE, RF, SSH and WS were employed as independent variables, whereas the EW population served as dependent variables.

Result: By stepwise regression analysis, the study revealed that the response variable, EW population was significantly influence by WS, MAXT, MINT, RHM whereas SSH and RF showed non-significant influence, the model R2 value for the fitted regression was low, indicating that the model was not a strong fit due to non-linearity and high heterogeneity in insect pests.

Keywords: Groundnut earwig, stepwise regression, contributing weather factors

1. Introduction

Earwigs as pod borers of groundnut were reported by Cherian and Basheer (1940) [1]. They found *Euborcllia stali* Dohrn, feeding on groundnut kernels by boring into the pods at Coimbatore, India. They stated that Burr (1910) [6] observed the association of this earwig with groundnut in Madras and Pondicherry in India. It has been suggested that it may be more widely distributed in southern India (Cherian and Basheer, 1940) [1]. Cherian and Basheer (1940) [1] observed that infestation of pods by earwig *Euborcllia stali* Dohrn, ranged from 2.7% to 19.95%. Counts of the attacked pods taken at the time of harvest of groundnut in different fields in three areas indicated 2.7-6.1% at Palur, 6.2-13.5% at Tindivanam and 9.6-19.95% infestation at Coimbatore. Purushottaman *et al.*, 1970 reported as high as 46.6% pod infestation. The matured pods were infested to an extent of 44.1% and the intensity of infestation in the case of immature pods was 52.1% in *Asirya muitunde*, a variety of groundnut. A total loss of 114 kg of oil per ha was registered. A similar high infestation of *Anisohbis atznulipcs* (= *E. stali*) to the extent of about 40% of bored pods has been observed on the susceptible genotypes in a vertisol field at ICRISAT (1986).

In several studies it is reported weather variable influences the occurrence and dynamics of pest populations. Stepwise regression analysis identifies the climatological factors influencing the incidence of gall midge population. One of the study taken up by Rathod *et al.*, 2022 [16] revealed that MINT, RHE, SSH at Warangal; RHM at Maruteru; MAXT, RHM and SSH at Pattambi, and MINT, RHM, RHE and SSH at Jagdalpur showed significant influence on the gall midge population. The model R^2 value for the fitted regression in all four of the centers is low, indicating that the model is not a strong fit, for which non-linearity and high heterogeneity in dependent variables may be responsible. The poll stepwise regression analysis revealed that all the weather parameters collectively accounted for variability in the major pest population with R^2 values of aphid (0.41), leaf hopper (0.62), Thrips (0.66), and whitefly 0.66 (Kadam *et al.,* 2022) [11]. The stepwise regression analysis was employed to analyze the correlation of six epidemiological variables (minimum temperature, maximum temperature, minimum relative humidity, maximum relative humidity, rainfall and wind speed) with disease severity and yield loss (%). The stepwise regression method gives better results for forecasting groundnut productivity in the Junagadh district of Gujarat. (Kumar *et al.,* 2020) [13]. Input variables were selected from the stepwise regression model. (Kumar *et al.,* 2018) [14]. The test statistics and probability of significance values show the impact of the independent variables on the dependent variable. Only minimum temperature and NDVI showed a significant positive relationship with BPH trap catches. The large t-value (t = 5.140) and corresponding lowest p-value ($p < 0.01$) support the result for NDVI at a one-month lag, which had the highest beta coefficient by using stepwise selection using Multiple linear regression (Skawsang *et al.,* 2019) [20]. The stepwise multiple regression method is used to forecast the fish landing (Ghani *et al.,* 2010) [10]. Stepwise regression indicated the significant influence of air temperatures, rainfall and relative humidity on the whitefly population and MBYMV severity (Khan *et al.,* 2006) [12]. A strong relationship between temperature, relative humidity, rainfall, and sunshine hours with the development of gall midge populations in rice for successive generations using a stepwise regression approach (Samui *et al.,* 2004) [17]. Basavaraj *et al.* (2020) [4] used regression models to study the influence of weather parameters on the progression of white rust disease of Indian mustard and reported that weather variables have a significant impact on the disease progression in the studied location. Baswaraj *et al.* (2023) [5] developed statistical models for quantifying the relationship between weather variables and lepidopteran pest populations in Kalyan Karnataka. Reddy *et al.* (2022) [19] studied the relationship between weather variables and the Rice Yellow Stem Borer population and reported that minimum temperature, sunshine hours and evening relative humidity significantly contributed to the dynamics of the Rice Yellow Stem Borer population in the study area. In this article, an effort has been made to model the relationship between weather variables and the pest population under consideration.

1.1 Objectives

1. Descriptive statistics for earwig insect pests of groundnut along with weather variables.
2. Trend analysis of earwig insect pests of the groundnut population.
3. Correlation and stepwise regression analysis for major insect pests of groundnut along with weather variables.

2. Materials and Methods

2.1. Data Collection

Secondary data on earwig light trap catches during *kharif* were collected from Regional Agricultural Research Station (RARS) Tirupati from 2008 to 2022. Generally, under the light trap method, a bulb is daily illuminated from 6:00 pm to 6:00 am. In the morning, the collected Earwigs (EW) were brought to the laboratory and the number of individuals caught per day was manually counted, summed and presented as cumulative catches. Similarly, weather data during the corresponding crop period on maximum temperature (MAXT), minimum temperature (MINT), morning relative humidity (RHM), evening relative humidity (ERH), rainfall (RF), sunshine hours (SSH) and Wind speed (WS) were also collected from the Automatic Weather Station (AWS) situated at the research site *i.e.,* Regional Agricultural Research Station, Tirupati, Andhra Pradesh. Standard meteorological week (SMW)-wise cumulative catches of earwigs and weekly averages of weather parameters were considered for this study.

2.2 Statistical Models

Statistical modelling started by calculating descriptive statistics, such as the mean, standard deviation (SD), skewness, kurtosis, minimum observation, maximum observation, and coefficient of variation (CV). These measures are essential in providing insights into the characteristics of the data under study. Additionally, time series plots were used to graphically represent the data over time. To explore the relationships among the variables used in the study, Pearson's correlation coefficient analysis was performed. This analysis helped to determine the degree of interrelationship between different variables. Further, to investigate the cause-and-effect relationship between earwig populations and exogenous weather variables, a stepwise multiple regression analysis model was employed. The regression equation in terms of matrix notation is expressed as follows:

$$Y=X\beta+e \quad ...(1.1)$$

Where Y is the dependent variable representing the earwig populations, X is the matrix of independent variables comprising the exogenous weather variables, β is the vector of regression coefficients and e is the residuals term assumed to be normally distributed with $e \sim N(0, \sigma^2)$. Correlation analysis and stepwise regression analysis were carried out in SAS version 9.3. available at ICAR- Indian Institute of Rice Research Hyderabad.

3. Results and Discussion

3.1 Summary statistics

The methodological framework begins with basic descriptive statistics, correlation analysis, and stepwise regression analysis to understand the causal relationships between the Earwig populations during *Kharif* and weather variables. Figure 1 displays the time series plots of weekly counts of earwigs light trap catches at RARS, Tirupati observed over the period from 2008-2022 during *kharif.* The data is presented on a week-by-week basis (SMW-wise), allowing for a visual representation of the fluctuating patterns and trends in earwig populations at RARS, Tirupati throughout the specified time frame. During *Kharif* in the year 2018 and 2021 high EW pest is observed in 28th SMW and 51thSMW. The plot provides valuable insights into the long-term changes in EW populations, which are crucial for understanding the dynamics of this insect pest in the study area during the observed 15 years period.

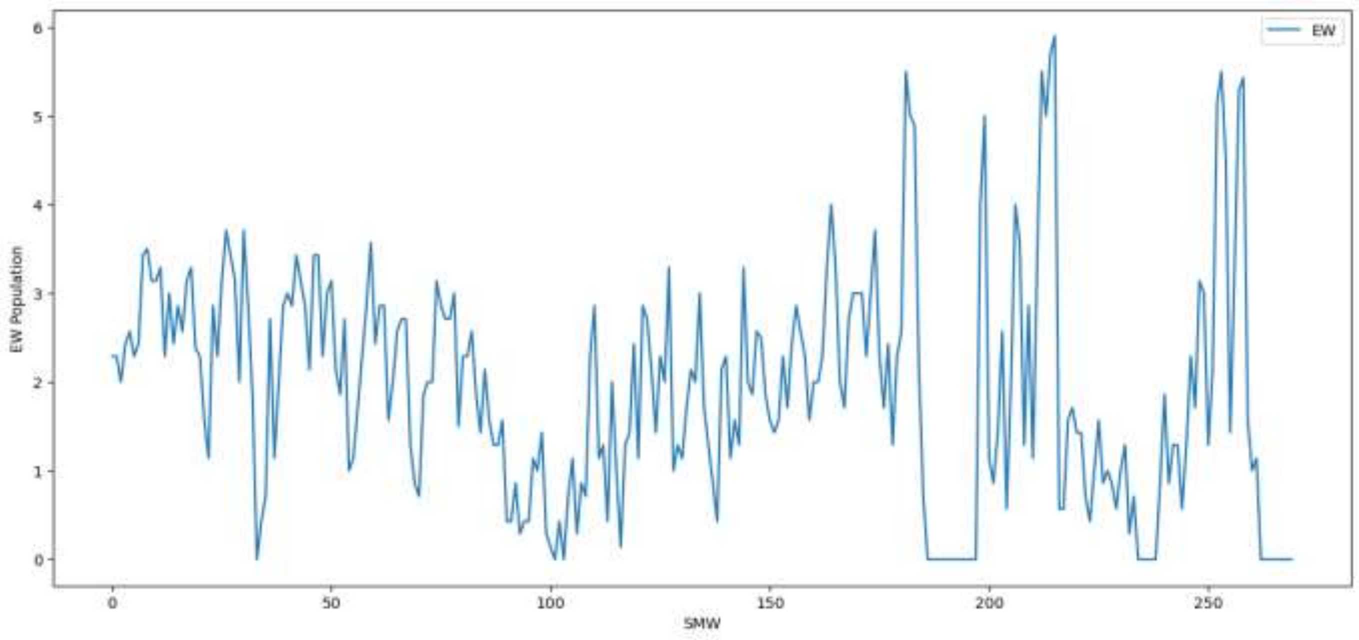

Fig 1: Time series plot of Earwig population during 2008-2022

Table 1 provides the summary statistics for both the dependent variable, the earwig population, and the exogenous weather variables. The EW population counts exhibit a wide range of oscillation, spanning from 0-12. As a result, the coefficient of variation (CV) is relatively high (83.61%), indicating the heterogeneous nature of the data. Moreover, the skewness and kurtosis values deviate from the normal range, suggesting a departure from a normal distribution and highlighting the abnormality of the data. The summary statistics for the weather variables that are presented in Table 1, revealed their inherent heterogeneity. The summary statistics provide a clear understanding of the nature of the data, indicating that the weather variables considered in the study exhibit significant variation and diverse characteristics.

Table 1: Summary statistics of earwig light trapped individual collections at RARS, Tirupati

	MAXT	MINT	RHM	RHE	RF	SSH	WS	EW
Mean	34.04	24.50	76.74	50.81	5.48	4.57	5.79	2
Median	33.84	24.58	77.36	50.43	3.24	4.55	5.28	2
Mode	33	24.5	65	54.14	0	4.24	0	0
Standard Deviation	1.83	1.69	10.68	9.59	8.88	2.06	3.14	2
Kurtosis	0.15	1.04	28.52	-0.20	52.13	2.26	-0.14	9
Skewness	0.07	-0.68	2.91	0.29	5.90	0.62	0.38	2
Range	11.4	9.94	126.86	49.57	98.8	15.8	14.9	12
Minimum	28.17	18.4	50.86	31.57	0	0	0	0
Maximum	39.57	28.34	177.71	81.14	98.8	15.8	14.9	12
CV%	5.37	6.88	13.92	18.88	162.16	45.05	54.17	83.61

3.2 Correlation analysis
Pearson correlation coefficients between Earwig populations and considered weather variables are depicted in Table 2. A very low negative and nonsignificant correlation was observed between the EW population and MAXT (-0.11), RHM (-0.11), SSH (-0.05); very low positive and nonsignificant correlation between the EW population and MINT (+0.15), RHE (+0.05), RF (+0.09) and WS (+0.17) variables.

Table 2: Pearson correlation coefficients between Earwig light trapped individual collections and weather variables

	EW	MAXT	MINT	RHM	RHE	RF	SSH
MAXT	-0.11 (0.51)						
MINT	0.15 (0.69)	0.57 (<0.0001)					
RHM	-0.11 (0.25)	-0.52 (<0.0001)	-0.43 (<0.0001)				
RHE	0.05 (0.07)	-0.85 (<0.0001)	-0.46 (<0.0001)	0.57 (<0.0001)			
RF	0.09 (0.57)	-0.24 (<0.0001)	-0.16 (0.01)	0.14 (0.02)	0.25 (<0.0001)		
SSH	-0.05 (0.40)	0.28 (<0.0001)	-0.13 (0.04)	-0.04 (0.47)	-0.27 (<0.0001)	-0.15 (0.01)	
WS	0.17 (0.17)	0.49 (0.08)	0.56 (<0.0001)	-0.45 (<0.0001)	-0.60 (<0.0001)	-0.16 (0.01)	-0.13 (0.04)

3.3 Stepwise regression analysis
The results of the stepwise regression analysis were depicted in Tables 3 and 4 For the response variable EW population, explanatory variables like MAXT, MINT, RHM and RF are significantly contributing in which MAXT and RHM have a negative impact on the EW population whereas MINT and RF

have a positive impact on the EW population for the data under consideration. From Table 4, it was clearly observed that at the beginning, the weather variable WS is entered into the model, resulting in a partial R^2 value of 0.0291. In the second step, in addition to WS, another weather variable MAXT was entered into the model with 0.0472 as a partial R^2 value by which the model R^2 raised from 0.0291 to 0.0762. Similarly, in the third step, the weather variable MINT got a chance to enter into the model after WS and MAXT with a partial R^2 value of 0.0196 by which the model R^2 became 0.0959. Further, in the fourth step, the weather variable RHM was entered into the model in addition to WS, MAXT and MINT with a nominal R^2 of 0.0129 that uplifted the model R^2 up to 0.1088. Later, SSH was entered into the model with 0.0110 as partial R^2 as one of the qualified variables similar to WS, MAXT, MINT, and RHM by which the model R^2 became 0.1198. In the final and sixth step, the weather variable RF was entered into the model along with all the previously mentioned variables, and contributed to the partial R^2 value of 0.0089 and the model R^2 increased up to 0.1287. Stepwise selection stops when there are no other variables to be added to the model as per the selection criterion. The process has resulted in a final model with six variables; each was selected based on their contribution to explain the variation in the response variables of interest. The final model R^2 value is 0.1287 indicates the proportion of the total variance in the response variable that can be explained by the predictor variables presented in the model.

Though the listed variables viz., MAXT, MINT, RHM and RF had a significant influence on the EW populations, the model R^2 value for the fitted regression for RARS, Tirupati was low (0.1287), indicating that the model was not a strong fit due to non-linearity and the presence of high heterogeneity in the Earwig catches.

Table 3: Stepwise regression model for Earwig population during *kharif* and weather variables

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	6	57.805	9.634	6.48***	<0.0001
Error	263	391.275	1.487		
Corrected Total	269	449.080			

Table 4: Parameters estimation of stepwise regression

Variable	Parameter Estimate	Standard Error	F Value	Pr > F	R^2	Model R^2
Intercept	8.504	2.231	14.530	0.000		
WS	-0.303	0.061	8.020	0.005	0.029	0.029
MAXT	0.165	0.061	13.640	0.000	0.047	0.076
MINT	-0.017	0.009	5.770	0.017	0.020	0.096
RHM	0.014	0.009	3.840	0.051	0.013	0.109
SSH	0.083	0.042	3.300	0.071	0.011	0.120
RF	0.093	0.031	2.700	0.102	0.009	0.1287

4. Discussion

In this particular study, the stepwise regression model showed weak fitness, indicated by a low R^2 value (13%). This can be attributed to the presence of non-linearity and high heterogeneity in the dependent variable, which made it challenging to establish a strong relationship between the earwig populations and the exogenous weather variables. This fact can also be viewed in Fig. 2. Results of step-wise regression explored the fact that RHE left the model which is significant at the 0.150 level. Hence, RF, MAXT, MINT, and RHM showed a significant impact on Earwig infestation. These results and Fig. 3 emphasized the need for further exploration and consideration of alternative statistical approaches or more sophisticated models that can handle non-linear relationships and high data heterogeneity. Additionally, it is crucial to consider other potential factors that might influence Earwig populations but were not included in this specific analysis. Future research should focus on investigating different modeling techniques and incorporating a more comprehensive set of explanatory variables to improve the understanding of the factors affecting Earwig populations. One of the similar studies carried out on leafminer by Arunachalan and Kavitha (2012) [2] showed that the incidence of groundnut leaf miner increased with a rise in maximum temperature and decreased with relative humidity, rainfall and leaf wetness. Gadad *et al.* (2013) [9] observed the incidence of groundnut leaf miner and stated that, MAXT(r=-0.53), RHM (r=-0.82) and RHE (r=0.75) showed significant and negative relationships with leaf miner incidence. Groundnut leaf miner was significantly and negatively correlated with rainfall and minimum temperature and positively with sunshine hours under late sown conditions at Anantapur (AICRPAM, 2001). Fenoglio *et al.* (2014) [8] showed that *Doru luteipes* (earwig) density and mean temperature were positively correlated with mortality.

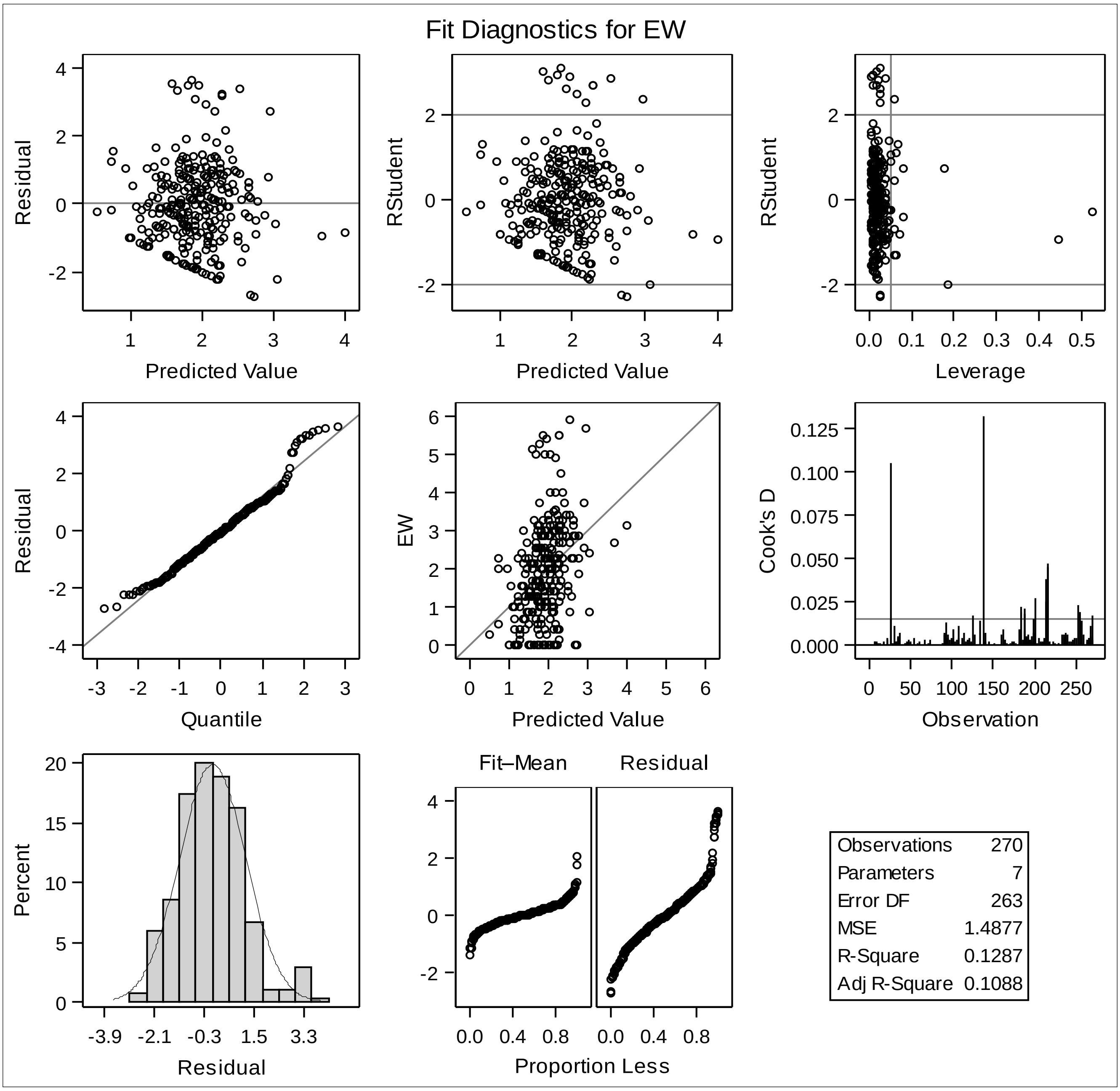

Fig 2: Regression diagnosis of Kharif Earwig population with weather parameters

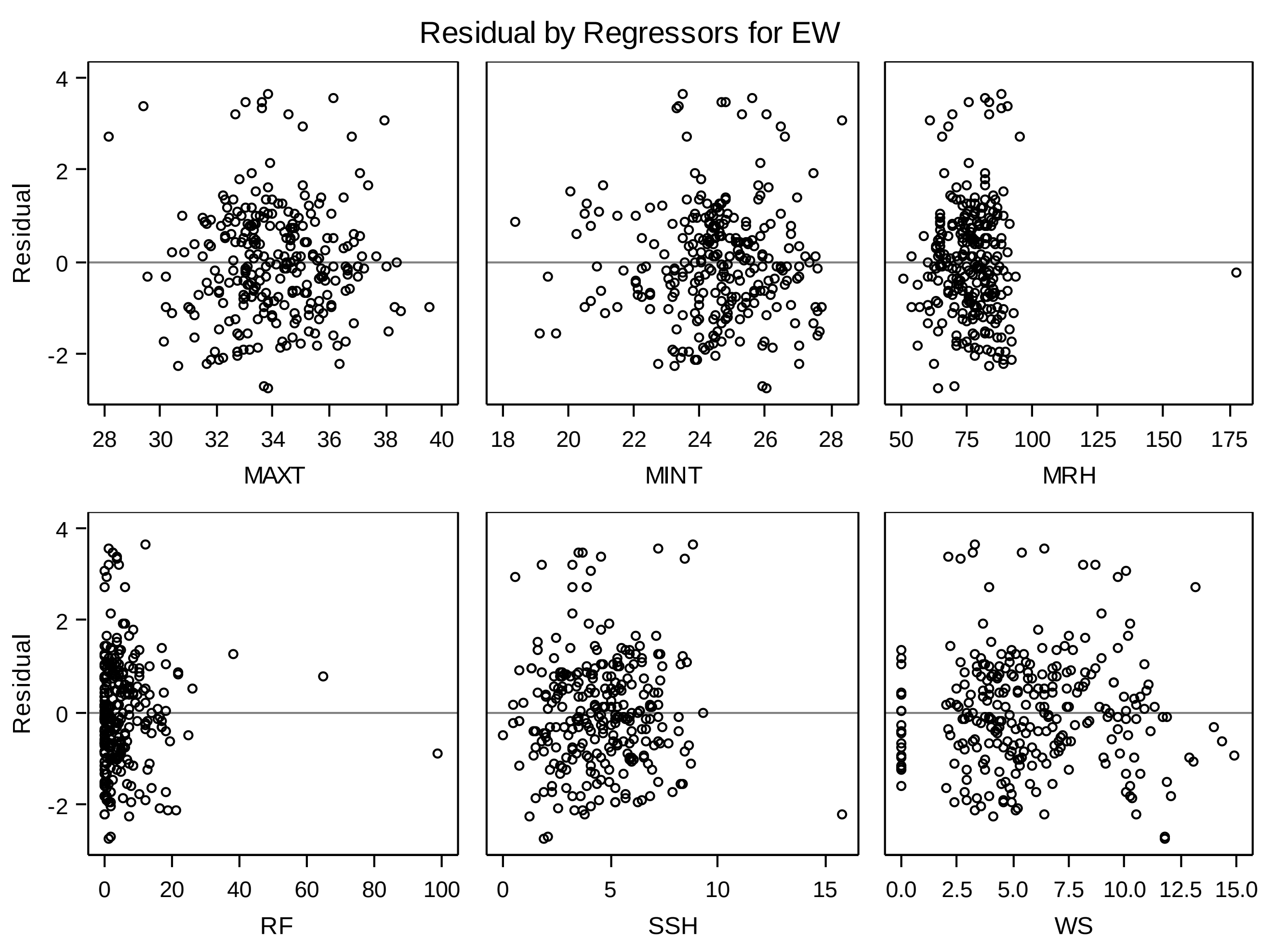

Fig 3: Residual analysis of step wie regression

5. Conclusions

The findings of this study indicate that the stepwise regression model used to examine the relationship between Earwig (EW) populations and exogenous weather variables yielded weak fitness, as evident from the low R^2 value. The primary reasons behind this weak fitness were the presence of nonlinearity and high heterogeneity in the dependent variable, i.e., the EW populations. Consequently, it can be concluded that the relationship between EW populations and the selected exogenous weather variables is not adequately explained by the stepwise regression model. The non-linear and heterogeneous nature of the data likely introduces complexities that the model could not capture effectively.

In conclusion, this study highlights the challenges in establishing a strong relationship between EW populations and exogenous weather variables due to non-linearity and high heterogeneity in the data. These findings underscore the need for future research to improve our understanding of the factors influencing earwig population and their dynamics.

6. Acknowledgements

The author would like to thank the Regional Agricultural Research Station, (RARS) Tirupati for providing the necessary facilities to carry out this work.

7. References

1. Ali Y, Raza A, Iqbal S, Khan AA, Aatif HM, Hassan Z, *et al*. Stepwise Regression Models-Based Prediction for Leaf Rust Severity and Yield Loss in Wheat. *Sustainability*. 2022;14(21):13893.
2. Arunachalam P, Kavitha Z. Response of groundnut genotypes to leaf miner (*Aproaerema modicella* Deventer) under field screening. Journal of Agriculture and Veterinary Science. 2012;1(2):20-23.
3. AICRP on Groundnut, DGR, Junagadh-Annual report; c2021.
4. Basavaraj K, Rathi AS, Gurav NP, Kumar A, Rathod S, Niwas R. Influence of weather variables on progression of white rust disease of Indian mustard. Journal of Agrometeorology. 2020;22(3):360-364.
5. Biradar B, Kulkarni SA, Shobharani M, Sidramappa GB, Rathod S, Naveena K, *et al*. Enhancing legume crop protection: Machine learning approach for accurate prediction of lepidopteran pest populations in Kalyan Karnataka. The Pharma Innovation. 2023;12(4):2319-2326.
6. Burr M. Dermaptera. Fauna of British India, 1910, 217.
7. Cherian MC, Basheer M. *Euborellia stali* Dohrn (Forficulidae) as a pest of groundnut in South India. Indian Journal of Entomology. 1940;2:155-158.
8. Fenoglio MS, Trumper EV. Influence of weather conditions and density of *Doru luteipes* (Dermaptera: Forficulidae) on *Diatraea saccharalis* (Lepidoptera: Crambidae) egg mortality. Environmental entomology. 2014;36(5):1159-1165.
9. Gadad H, Hegde M, Balikai RA. Seasonal incidence of

Spodoptera litura and leaf miner in rabi/summer groundnut. Journal of Experimental Zoology. 2013;16(2):619-622.
10. Ghani IMM, Ahmad S. Stepwise multiple regression method to forecast fish landing. Procedia-Social and Behavioral Sciences. 2010;8:549-554.
11. Kadam DR, Timke SH, Dakhore KK, Kadam YE. Seasonal abundance of major insect pest on Bt cotton vis-à-vis weather parameters in Marathwada region, Maharashtra. Journal of Agrometeorology. 2022;24(4):430-433.
12. Khan MA, Rasheed S, Ali S. Evaluation of multiple regression models based on epidemiological factors to forecast Bemisia tabaci and mungbean yellow mosaic virus. Pakistan Journal of Phytopathology. 2006;18(2):107-110.
13. Kumar KS, Shitap M. Statistical evaluation of stepwise regression method and autoregressive integrated moving average method for forecasting of groundnut (*Arachis hypogaea* L.) productivity in Junagadh district of Gujarat. International Journal of Current Microbiology and Applied Sciences. 2020;9(11):84-93.
14. Kumar SP, Mazumdar D, Kamei D, Kamei A. Advantages of artificial neural network over regression method in prediction of pest incidence in rice crop. International Journal of Agricultural and Statistical Sciences. 2018;14(1):357-363.
15. Purushottnman G, Das GG, Venugopal S. A note on *Euborellia stali* Dohrn. as a pest *Asirya mwitunde*. Madras Agricultural Journal. 1970;57(7):44-45.
16. Rathod S, Yerram S, Arya P, Katti G, Rani J, Padmakumari AP, *et al*. Climate-Based Modeling and Prediction of Rice Gall Midge Populations Using Count Time Series and Machine Learning Approaches. Agronomy. 2022;12(1):22.
17. Samui RP, Chattopadhyay N, Sabale JP. Weather based forewarning of gall midge attack on rice and operational crop protection using weather information at Pattambi. Mausam. 2004;55(2):329-338.
18. SAS Software, Version 9.3, SAS Institute: Cary, NC, USA; c2011.
19. Reddy BNK, Rathod S, Kallakuri S, Sridhar Y, Admala M, Malathi S, *et al*. Modelling the Relationship between Weather Variables and Rice Yellow Stem Borer Population: A Count Data Modelling Approach. International Journal of Environment and Climate Change. 2022;12(11):3623–3632. https://doi.org/10.9734/ijecc/2022/v12i111410
20. Skawsang S, Nagai M, Tripathi KN, Soni P. Predicting rice pest population occurrence with satellite-derived crop phenology, ground meteorological observation, and machine learning: a case study for the Central Plain of Thailand. Applied Sciences. 2019 Nov 12;9(22):4846.