Volume 3 - Issue 9– September,2023

SABUJEEMA
An International Multidisciplinary e-Magazine

STATISTICAL EVALUATION OF STEPWISE REGRESSION METHOD FOR ROOT GRUB POPULATION IN GROUNDNUT (*Arachis hypogaea* L.)

[Article ID: SIMM0289]

P. Minruhi[1], P. Lavanya Kumari[2], Santosha Rathod[3], B. Ramana Murthy [4] and K. Devaki[5]

1,2ANGRAU, S.V Agricultural College, Tirupati-517502

3 ICAR-Indian Institute of Rice Research, Hyderabad-500030

4 ANGRAU-SMGR Agricultural College, Udayagiri-524226

5. ANGRAU-Regional Agricultural Research Station, Tirupati-517502

ABSTRACT

Root grub or white grub (*Holotrichia serrata*) is one of the most important soil pests affecting groundnut. Secondary data on Root grub (RG) light trap catches during *kharif* was collected from Regional Agricultural Research Station (RARS) Tirupati, for the period of 19 years, from 2004 to 2022 during *Kharif* (26 SMWs from June to 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. In the stepwise regression method, the weekly average values of MAXT, MINT, RHM, RHE, RF, SSH and WS were employed as independent variables, whereas the RG population served as dependent variables. This study revealed the insignificant influence of WS on RG population and the fitness of the stepwise regression model was also found to be weak (low R^2) due to non-linearity and high heterogeneity in the dependent variable.

Key words: Groundnut Root Grub, Stepwise Regression, contributing weather factors

INTRODUCTION

Root grub or white grub (*Holotrichia serrata*) is one of the most important soil pests affecting groundnut. Groundnut crop is known to be infested by more than 360 species of insect pests in different parts of the world (Anitha *et al.,* 2006). The white grubs belonging to Scarabaeidae are the diverse and devastating pests of several crops and assumed national importance due to the high percent loss incurred (Sreedevi and Tyagi, 2015). In endemic areas, the damage to groundnut ranges from 20-100 %. The presence of one grub/m^2 may cause damage up to 39.40% (Umeh *et al.,* 1999),12-60% (Pokhrel *et al.,* 2004). In the majority of crops, pests have a major effect on crop yield as it reduces the quality of crops, the photosynthetic activity of plants, and the attack of pests may result in total crop failure. 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 studies taken up by Rathod *et al.,* 2022 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 centres 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). The stepwise regression analysis was employed to analyse 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). Input variables were selected from the stepwise regression model. (Kumar *et al.,* 2018). The t and p 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). The stepwise multiple regression method is used to forecast the fish landing (Ghani *et al.,* 2010). Stepwise regression indicated the significant influence of air temperatures, rainfall and relative humidity on whitefly population and MBYMV severity (Khan et al., 2006). 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). Basavaraj et al. (2020) used regression models to study the influence of weather parameters on progression of white rust disease of Indian mustard and reported that weather variables have significant impact on the disease progression in studied location. Baswaraj et al. (2023) developed statistical models for quantifying the relationship between weather variables and lepidopteran pest populations in Kalyan Karnataka. Reddy *et al.* (2022) studied the relationship between weather variables and Rice Yellow Stem Borer population and reported that minimum temperature, sunshine hours and evening relative humidity significantly contributing to dynamics of Rice Yellow Stem Borer population in study area. In this article, an effort has been made to model the relationship between weather variables and pest population under consideration.

1. Materials and Methods

2.1. Data Collection

Secondary data on root grub light trap catches during *kharif* were collected from Regional Agricultural Research Station (RARS) Tirupati from 2004 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 Root grubs (RG) 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 root grubs and weekly averages of weather parameters were considered for this study.

2.2 Statistical Models

Statistical modelling started by calculating descriptive statistical parameters, 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 root grub populations and exogenous weather variables, a stepwise multiple regression analysis was conducted. 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 root grub populations, X is the matrix of independent variables comprising the exogenous weather variables, β be the vector of regression coefficients and e is the residuals term assumed to be normally distributed with e ~N (0, σ^2). Correlation analysis and stepwise regression analysis were carried out in SAS version 9.3. available at ICAR-Indian Institute of Rice Research Hyderabad.

2. Results and discussion

2.1. Summary statistics

The methodological framework begins with basic descriptive statistics, correlation analysis, and stepwise regression analysis to understand the causal relationships between the root grub populations during *Kharif* and input variables. Figure 1 displays the time series plots of weekly counts of root grub light trap catches at RARS, Tirupati centre observed over the period from 2004-2022 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 root grub populations at RARS, Tirupati throughout the specified time frame. During *kharif* in the year 2019 high RG pest was predominant in the 40th SMW and in the year 2014 high RG pest was observed in 28th SMW & 29th SMW. The plots provide valuable insights into the long-term changes in RG populations, which are crucial for understanding the dynamics of this insect pest in the study area during the observed 19 years period.

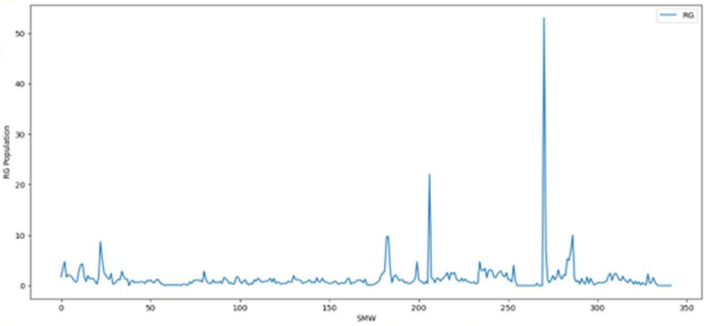

Fig 1 Time series plot of root grub population

Table 1 provides the summary statistics for both the dependent variable, the root grub population, and the exogenous weather variables. The RG population counts exhibit a wide range of oscillation, spanning from 0-53. As a result, the coefficient of variation (CV) is relatively high (232 %), indicating heterogeneous nature of 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 is presented in Table1, revealed their inherent heterogeneity. The summary statistics

	MAXT	MINT	RHM	RHE	RF	SSH	WS	RG
Mean	33.94	24.55	75.86	50.76	10.21	4.65	6.55	2
Median	33.89	24.6	76.66	50.07	3.71	4.6	5.84	1
Mode	33.9	24.5	65	43	0	5.8	0	1
Standard Deviation	1.89	1.69	10.90	9.93	20.88	2.25	3.63	4
Kurtosis	0.52	0.71	21.12	0.05	22.79	13.39	0.18	169
Skewness	-0.21	-0.54	2.13	0.40	4.43	1.91	0.61	12
Minimum	27.1	18.4	50.86	31.57	0	0	0	0
Maximum	39.57	28.34	77.71	85	158.2	23	17.5	53
CV%	5.56	6.90	14.37	19.57	204.43	48.29	55.38	232

(Table 1. Summary statistics of root grub light trapped individual collections at RARS, Tirupati)
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.

2.2.Correlation analysis

Pearson correlation coefficients between gall midge populations and considered weather variables are depicted in Table 2. A very low negative and nonsignificant correlation was observed between RG population and RHM (-0.05), RHE (-0.05) and SSH (-0.03); very low positive and nonsignificant correlations between the RG population and MAXT (+0.06), MINT (0.08), RF (+0.0012) and WS (+0.08) variables.

Table 2. Pearson correlation coefficients between root grub light trapped individual collections and weather variables

	RG	MAXT	MINT				
MAXT	0.06 (0.24)						
MINT	0.08 (0.16)	0.59 (<0.0001)					
RHM	-0.05 (0.35)	-0.54 (<0.0001)	-0.50 (<0.0001)				
RHE	-0.05 (0.33)	-0.86 (<0.0001)	-0.52 (<0.0001)	0.62 (<0.0001)			
RF	0.0012 (0.95)	-0.39 (<0.0001)	-0.21 (<0.0001)	0.18 (<0.0001)	0.35 (<0.0001)		
SSH	-0.03 (0.60)	0.23 (<0.0001)	-0.11 (0.04)	-0.03 (0.60)	-0.22 (<0.0001)	0.04 (0.41)	
WS	0.08 (0.12)	0.43 (<0.0001)	0.58 (<0.0001)	-0.56 (<0.0001)	-0.59 (<0.0001)	-0.04 (0.42)	-0.09 (0.09)

2.3.Stepwise regression analysis

The results of the stepwise regression analysis were depicted in Table 3. For the response variable RG population, explanatory variables like WS are non-significantly contributing and WS have positive impact on the RG population for the data under consideration. Though the listed variables had a nonsignificant influence on the RG populations, the model R^2 value for the fitted regression for RARS, Tirupati was low, indicating that the model was not a strong fit the reason for the same could be due to non-linearity and presence of high heterogeneity in the dependent variable.

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

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	25.95829	25.95829	2.37 NS	0.1249
Error	340	3729.75113	10.96986		
Corrected Total	341	3755.70943			

Parameters estimation

Variable	Parameter Estimate	Standard Error	F Value	Pr > F	R^2	Model R^2
Intercept	0.92966	0.37012	6.31	0.0125	0.0069	
WS	0.07609	0.04946	2.37	0.1249		

2.4.Discussion

In this particular study, the stepwise regression model showed weak fitness, indicated by a low R^2 value. This can be attributed to the presence of non-linearity and high heterogeneity in the dependent variable, which made it challenging to establish not a strong relationship between the root grub populations and the exogenous weather variables. This fact can also be viewed from fig.2. Results of step wise regression explored the fact that all variables left the model are significant at the 0.1500 level. No other variable except WS, met the 0.1500 significance level for entry into the model. Hence only Wind speed showed significant impact on Root grab 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 root grub 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 root grub populations.

Conclusions

The findings of this study indicate that the stepwise regression model used to examine the relationship between RG populations and exogenous weather variables yielded a weak fitness, as evident from the low R^2 value. The primary reasons behind this weak fitness were the presence of non-linearity and high heterogeneity in the dependent variable, i.e., the RG populations. Consequently, it can be concluded that the relationship between RG 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's highlights the challenges in establishing a strong relationship between RG 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 root grub population and their dynamics.

References

1. Ali, Y., Raza, A., Iqbal, S., Khan, A.A., Aatif, H.M., Hassan, Z., Hanif, C.M.S., Ali, H.M., Mosa, W.F., Mubeen, I and Sas-Paszt, L. 2022. Stepwise Regression Models-Based Prediction for Leaf Rust Severity and Yield Loss in Wheat. Sustainability. 14(21): 13893.
2. Anitha, V., Rogers, D.J., Wightman, J and Ward, A. 2006. Distribution and abundance of white grubs (Coleoptera: Scarabaeidae) in groundnut in southern India. Crop Protection. 25(8): 732-740.
3. Basavaraj, K., Rathi, A. S., Gurav, N. P., Kumar, A., Rathod, S., and Niwas, R. 2020. Influence of weather variables on progression of white rust disease of Indian mustard. J. Agrometeorol., 22(3): 360-364
4. Baswaraj Biradar, Sunil A Kulkarni, Shobharani M, Sidramappa, Gnyanadev Bulla, Santosh Rathod, Naveena K, Gayathri Chitikela, Fakeerappa Arabhanvi. 2023. Enhancing legume crop protection: Machine learning approach for accurate prediction of lepidopteran pest populations in Kalyan Karnataka. Pharma Innovation 2023;12(4):2319-2326.
5. Ghani, I.M.M and Ahmad, S. 2010. Stepwise multiple regression method to forecast fish landing. Procedia-Social and Behavioral Sciences. 8: 549-554.
6. Kadam, D.R., Timke, S.H., Dakhore, K.K and Kadam, Y.E. 2022. Seasonal abundance of major insect pest on Bt cotton vis-à-vis weather parameters in Marathwada region, Maharashtra. Journal of Agrometeorology. 24(4): 430-433.
7. Khan, M.A., Rasheed, S and Ali, S. 2006. Evaluation of multiple regression models based on epidemiological factors to forecast Bemisia tabaci and mungbean yellow mosaic virus. Pakistan Journal of Phytopathology. 18(2): 107-110.
8. Kumar, K.S and Shitap, M. 2020. 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. 9(11): 84-93.
9. Kumar, S.P., Mazumdar, D., Kamei, D and Kamei, A. 2018. Advantages of artificial neural network over regression method in prediction of pest incidence in rice crop. International Journal of Agricultural and Statistical Sciences. 14(1): 357-363.
10. Pokhrel, M.R. 2004. “Field survey of white grubs and laboratory and evaluation of Metarhizium anisopliae (Metsch.) Sorokin for its control with side effects on Bombyx mori Lin. M.Sc.Ag.” Agriculture. Institute of

Agriculture and Animal Science, Tribhuvan University, Rampur, Chitwan, Nepal.

11. Rathod, S., Yerram, S., Arya, P., Katti, G., Rani, J., Padmakumari, A.P., Somasekhar, N., Padmavathi, C., Ondrasek, G., Amudan, S and Malathi, S. 2022. Climate-Based Modeling and Prediction of Rice Gall Midge Populations Using Count Time Series and Machine Learning Approaches. Agronomy.12(1): 22.
12. Samui, R.P., Chattopadhyay, N., Sabale, J.P. 2004. Weather based forewarning of gall midge attack on rice and operational crop protection using weather information at Pattambi. *Mausam*. 55: 329–338.
13. SAS Software, Version 9.3, SAS Institute: Cary, NC, USA, 2011.
14. 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. https://doi.org/10.9734/ijecc/2022/v12i111410
15. Skawsang, S., Nagai, M., K. Tripathi, N and Soni, P. 2019. 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. 9(22): 4846.
16. Sreedevi, K., Tyagi, S. 2015. Species diversity of white grubs associated with Sugarcane ecosystem of Western Uttar Pradesh – A Case Study. Current Biotica. 8(4): 404- 410.
17. Umeh, V.C., Waliyar, F., Traore, S and Egwurube E.1999. Soil pests of groundnut in West Africa-Species diversity, damage and estimation of yield losses. International Journal of Tropical Insect Science. 19(2-3): 131-140.

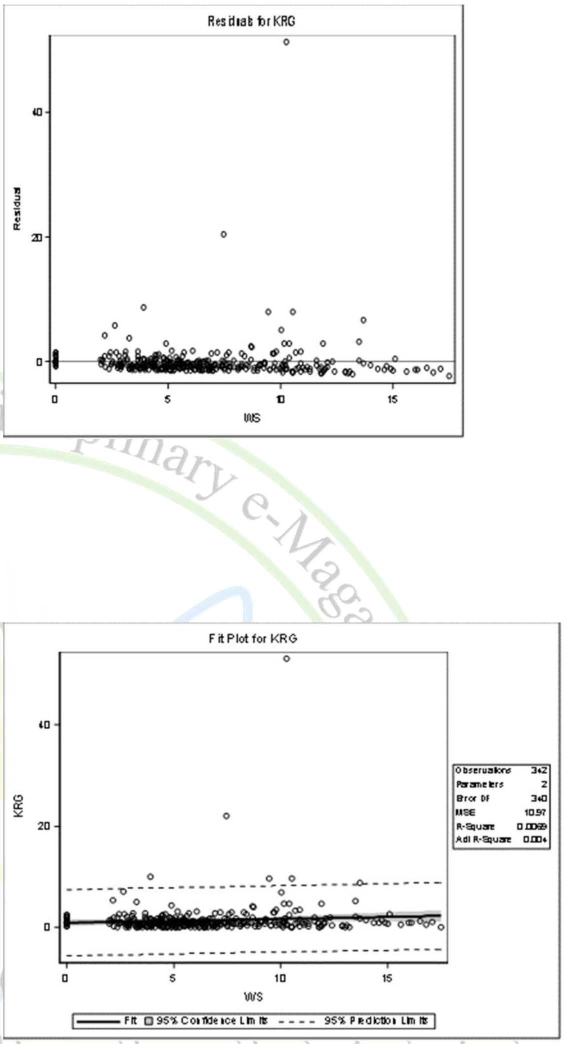

Fig 3: Residual analysis of step wise regression

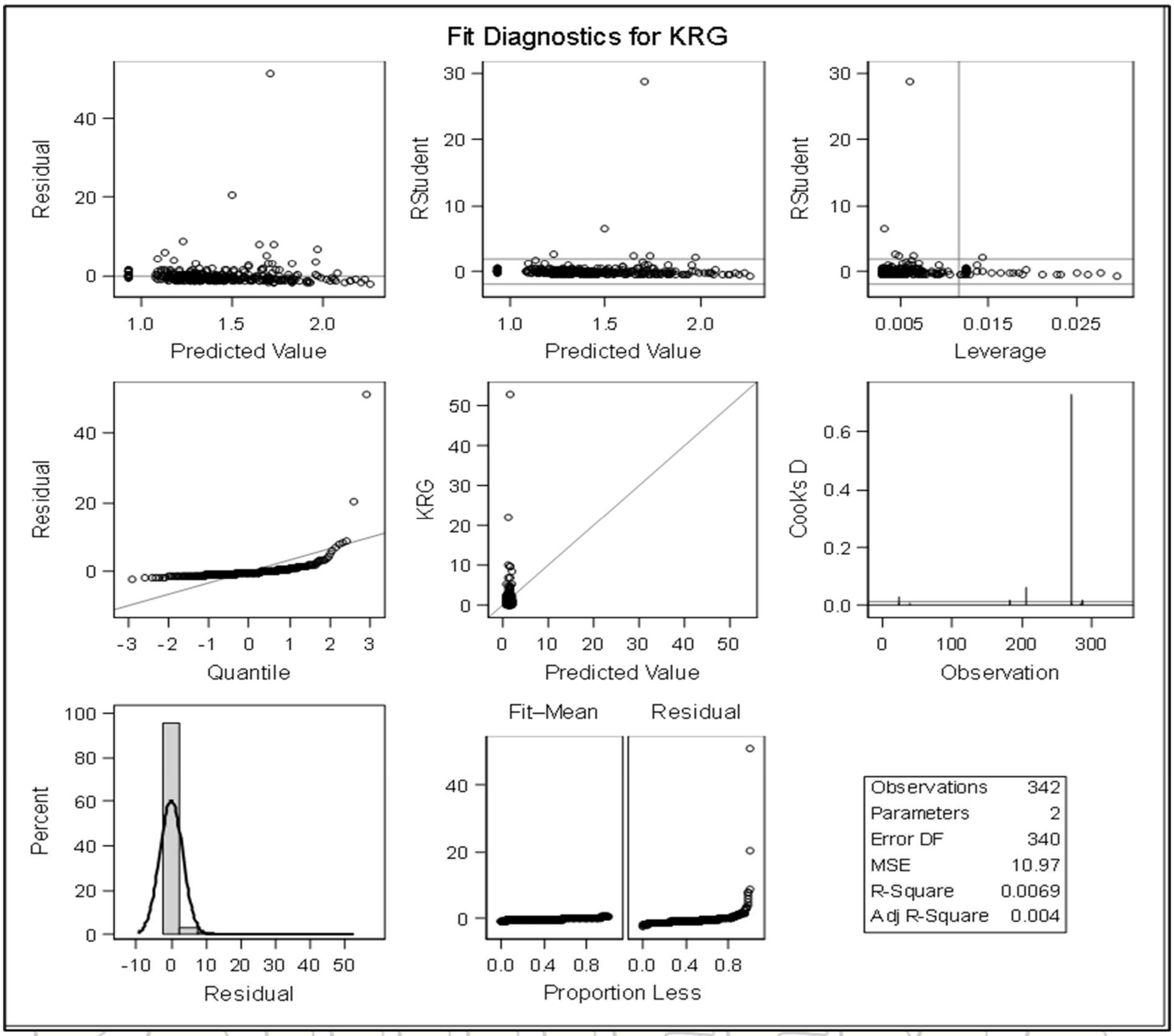

Fig 2: Regression diagnosis of Kharif Root Grub population with weather parameters

Table 2. Pearson correlation coefficients between root grub light trapped individual collections and weather variables

	RG	MAXT	MINT	RHM	RHE	RF	SSH
MAXT	0.06 (0.24)						
MINT	0.08 (0.16)	0.59 (<0.0001)					
RHM	-0.05 (0.35)	-0.54 (<0.0001)	-0.50 (<0.0001)				
RHE	-0.05 (0.33)	-0.86 (<0.0001)	-0.52 (<0.0001)	0.62 (<0.0001)			
RF	0.0012 (0.95)	-0.39 (<0.0001)	-0.21 (<0.0001)	0.18 (<0.0001)	0.35 (<0.0001)		
SSH	-0.03 (0.60)	0.23 (<0.0001)	-0.11 (0.04)	-0.03 (0.60)	-0.22 (<0.0001)	0.04 (0.41)	
WS	0.08 (0.12)	0.43 (<0.0001)	0.58 (<0.0001)	-0.56 (<0.0001)	-0.59 (<0.0001)	-0.04 (0.42)	-0.09 (0.09)