DEVELOPMENT AND VALIDATION OF POTATO LEAF ROLL VIRUS DISEASE PREDICTION MODEL BASED ON ENVIRONMENTAL FACTORS FOR FAISALABAD, PAKISTAN

Ummad-ud-Din Umar*, Muhammad Aslam Khan**, *Ateeq-ur-Rehman*, Abdul Hannan***, Syed Atif Hasan Naqvi*, Azhar Ali Khan****, Muhammad Asif Zulfiqar****

ABSTRACT:- A disease predictive model for Potato leaf roll virus (PLRV) was developed based on a ten year dataset of disease incidence and environmental variables and validated by two years data collected in Faisalabad, Pakistan. Maximum and minimum air temperatures, relative humidity and rainfall were employed as independent variables while PLRV disease incidence served as the response variable. Stepwise regression analysis was performed to select potentially useful predictor variables for model development. In stepwise regression analysis minimum temperature, relative humidity and rainfall emerged as the main contributing environmental variables for disease development. The model performance was evaluated based on coefficient of determination (R^2) and the comparison of observed and predicted disease incidence. Model so developed explained 70% of the variability in disease development and when validated on an independent dataset, the model explained 67.5% of the variability in disease incidence. This prediction model can be used for the management of PLRV in this region.

Key Words: Epidemiology, PLRV, Multiple regression models.

INTRODUCTION

Potato (*Solanum tuberosum* L.) is one of the world's most important crops. Over the years, potato has become an important crop for both farmers and consumers in Pakistan being cultivated over an area of 128 thousand hectares, with an average yield of 23 tonnes ha^{-1} (Anonymous, 2010). Out of many factors responsible for low potato yield in Pakistan, the most important are inferior seed quality and poor management practices to control pests and diseases. In Pakistan, potato is affected by many diseases, among which viral diseases are considered more serious than fungal and bacterial, primarily due to the fact that they are difficult to control. Potato is infected by several viruses, of which potato leaf roll virus (PLRV) is the most important, causing serious losses (Salazar, 1996). It is transmitted by aphid species in a persistent and non-propagative manner. Myzus persicae Sulz. regarded as most efficient vector (Harrison, 1984). PLRV infected seed produce plants with fewer and smaller tubers.

In Pakistan, PLRV is re-emerging disease because of disease incidence is increasing in the fields (Mughal, 2003), which has resulted in high

*Department of Plant Pathology, Bahauddin Zakariya University, Multan.
**Department of Plant Pathology, University of Agriculture, Faisalabad.
***Directorate of Land Reclamation, Irrigation Department Lahore, Pakistan.
****PARC, Research and Training Station, BahauddinZakariya University, Multan.
Coresponding author: ummadumar@hotmail.com

yield losses (Bhutta and Bhatti, 2002). High severity of PLRV and lack of resistance to this disease in the majority of varieties/lines indicated that the inoculum level is building up and may continue to increase in the future (Khan and Obaidullah, 2006). The large-scale spread of PLRV has been attributed mainly to continuous introductions of viruses through imported seeds and through the recurrent occurrence of the aphid vector (Ahmad and Ahmad, 1995). The rational use of pesticides requires a comprehensive understanding and identification of the various environmental factors responsible for the spread of disease and its vector i.e., aphid population. Epidemiology deals with the factors affecting the outbreak and spread of infectious diseases. The detailed knowledge and information regarding the host, pathogen and environmental factors are of fundamental importance to enable us to precisely forecast disease development and initiate effective plant protection measures. The present research includes the development of epidemiological models. The objectives of this study were to develop epidemiological model to predict the PLRV disease incidence by multiple regression ana-lysis of ten years dataset and its validation by new dataset to check model prediction. The validated model will be used in future for PLRV forecasting.

MATERIALS AND METHOD

Model development

For the development of a disease predictive model, ten year of dataset of PLRV disease incidence from November to April (autumn crop and spring crop) consecutively during 2002 to 2012, recorded on three highly susceptible varieties: Desiree, Diamant and Cardinal, and data of environmental factors (maximum and minimum temperature, rainfall, relative humidity and aphid population) were collected from Plant Virology Section, Ayub Agriculture Research Institute (AARI), Faisalabad. Mathematical relationships between PLRV disease incidence and epidemiological factors were studied through multiple regression analysis. Environmental variables exerting significant influences were determined by regression analysis and the most favorable environmental conditions for PLRV disease development were identified. A predictive model based on ten years data was then developed by multiple regression anlysis.

Model validation

The regression model was validated by 1) Comparison of dependent variable and regression coefficients with physical theory 2) Comparison of observed vs. predicted data 3) Collection of new data to check predictions (Chattefuee and Hadi, 2006; Snee, 1977).

Collection of New Data set

For the collection of new data an experiment was carried out at University of Agriculture, Faisalabad UAF (about 12 Km away from AARI) in the similar way as it has been carrying out for the last ten years. The environmental conditions and the soil type of both the places were almost same. The data of two consecutive seasons from November to April during 2010 to 2012 was recorded. In each season, tubers were sown in winter planting (middle of October) and in spring planting (Middle of January). The planted potato cultivars Desiree, Diamant and Cardinal were highly susceptible to PLRV. The varieties were randomized within blocks with three replications (RCBD design) 25m x 25m plot size having 75 cm row to row (R x R) and 20 cm plant to plant (P x P) distance. Standard dose of fertilizer NPK i.e., at rate of 50-30-30 Kg^{-1} was applied to the trial. The disease free tubers were obtained from Plant Virology Section, Ayub Agriculture Research Institute (AARI), Faisa-labad. All the potato tubers were processed through ELISA for the confirmation of any virus. Normal agricultural practices were applied as recommended and no insecticides were used during the study period. The fungicide metalyxal was used to protect the potato crop from fungal diseases.

Data of aphid (*M. persicae*) population in the area under investigation was recorded by yellow sticky trap method on weekly basis (Gabriel, 1965). The winged potato aphids were sampled by using three yellow plastic sticky traps (30 cm sq). One trap was placed in the middle of the field and one at each end. The traps were held vertically and covered with grease oil and fixed in the soil by a small wooden bar, at the height (one meter) level just above the potato canopy.

The disease incidence of PLRV infected plants showing leaf roll symptoms were determined by visual inspection at every line in every plot (Gabriel, 1965). In each row 10 plants were selected randomly and tagged. Surveys of infected plants were carried out on weekly basis starting from emergence date till the end of the season to collect the disease incidence data. Disease incidence assessments in the field were backed up by DAS-ELISA described by Clark and Adams (1977) Agdia. Inc. U.S.A (www.agdia.com) polyclonal antibodies were used for ELISA.

The new data set of disease incidence on three varieties sown at U.A.F from Nov 2010 to April 2012 was used to check the prediction by running the model with the parameters emerged after stepwise regression ana-lysis of 10 years data set. Data of environmental factors like maximum and minimum temperature, relative humidity and rainfall (from November to April during two years 2010-11 and 2011-12) were collected from the meteorological station (9610-B-1 Orion LX Weather Station) of U.A.F. The meteorological station is situated about 80 meters from the experim-ental field.

The data was subjected to multiple and stepwise regression analysis using forward selection method to determine significant response variables (Myers, 1990). All the data were analyzed by computer software SAS v. 8.0 (SAS Institute, Cary, NC) and Minitab 15 by Minitab Inc. U.S.A.

RESULTS AND DISCUSSION

PLRV disease predictive model based on the ten-year dataset (2002-2012)

The data from ten growing seasons suggested that except maximum temperature all the weather variables and aphid population contributed to PLRV disease development. The model variables were significant at $P < 0.05$. The coefficient of determination R2 = 0.70 showed the significance of overall prediction accuracy (Table. 1 & 2).

The probability plots are frequently recommended for assessing the goodness of fit of a hypothesized distribution and are often used as an informal means of assessing the non

Table 1. Analysis of variance for the multiple regression model analysis

Source	Df	SS	MS	*F*-ratio	*P*-value
Regression	5	219094	43819	344.62	0.0001*
Residual Error	715	90785	127		
Total	720	309879			

normality of a set of data (Johnson and Vichem, 1982). The normal probability plot for the ten years model showed that most of the data points were around the reference line, whereas only few data points both at the lower side and at the higher side deviated from the reference line affecting the normal distribution of

Table 2. Coefficients of estimates for multiple regression model during (2002-2012)

	Unstandardized Coefficient		Standardized Coefficients				
	Parameter estimate	Standard error	Beta	Variance inflation	Tolerance	*t*-ratio	*P*-value
Intercept	6.896	3.463				1.992	0.047*
Tmin	2.498	0.109	0.672	2.099	0.476	22.928	0.001*
RH	-0.265	0.041	-0.188	2.046	0.489	6.486	0.001*
RF	-1.326	0.307	-0.089	1.036	0.965	4.323	0.001*

** = Significant at P≤0.05, Tmin.= Minimum Temperature, RH= Relative humidity, RF= Rainfall*

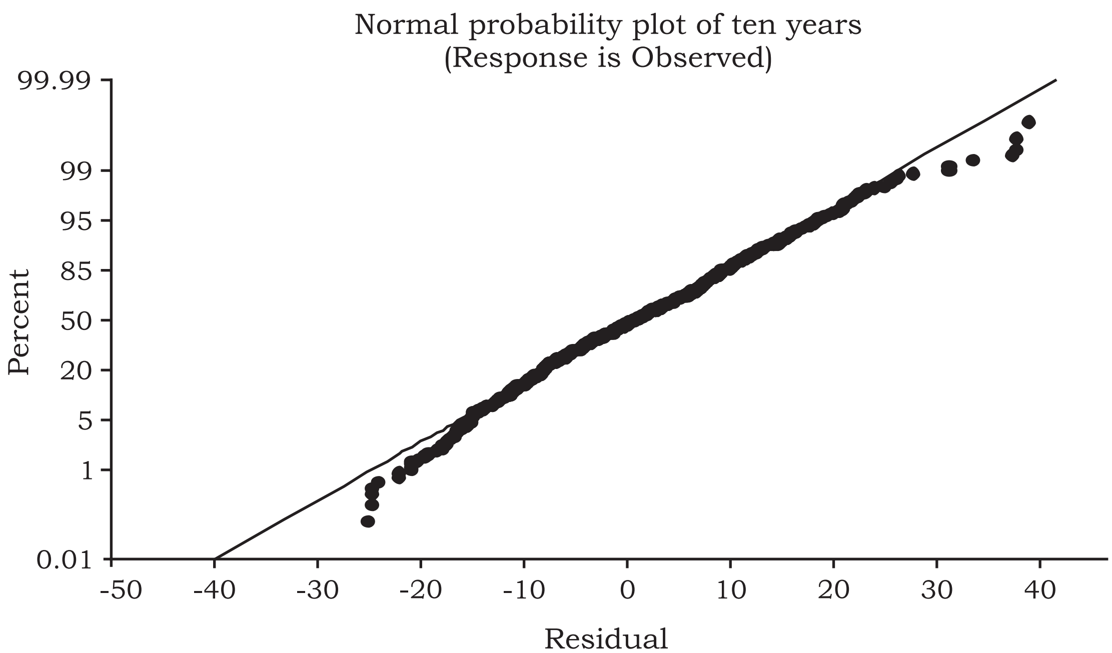

Figure 1. Normal probability plot for 10 years model of PLRV disease incidence.

data which could be the cause of an error in the regression model (Fig. 1). The model was further evaluated by the comparison of observed and predicted data. The R2 of 70% showed that there was a fair degree of matching between observed and predicted values (Fig. 2).

With respect to the relative importance of weather variables, minimum temperature had recorded highest percentile contribution (68%) followed by relative humidity (1.2%), rainfall 0.9% and aphid population 0.2% towards the disease development (Table 3). The minimum temper-ature out of four studied weather parameters had explained a substantive amount of total variation explained by all four weather parameters. Based on the Mallow's (Cp) values, models 3 had the best combinations of input variables (Table 3). The contribution of significant variables was assessed by stepwise regression and evaluated to predict PLRV disease incidence during ten years. Out of five variables entered, three of them i.e., minimum temperature, relative humidity and rainfall influenced significantly in the development of disease. The aphid population had a significant effect but its contribution was negligibly small so it was dropped and model 3 (Table 3) was selected, containing three variables explained 70% variability in disease development. After stepwise regression the multiple regression model was refined as:

$$y = `6.89 + 2.5x1 - 0.265x2 - 1.326x3 \quad R^2 = 0.70 \qquad (1)$$

where x1, x2 and x3 indicate minimum temperature, relative humidity and rainfall, respectively. The associated residual statistics are given in (Table 4)

Validation of the model with new dataset (2010-2012)

The third step of validation was the collection of new data for model prediction by fitting the model (Equation. 1) to the dataset of two extra years. The new dataset was

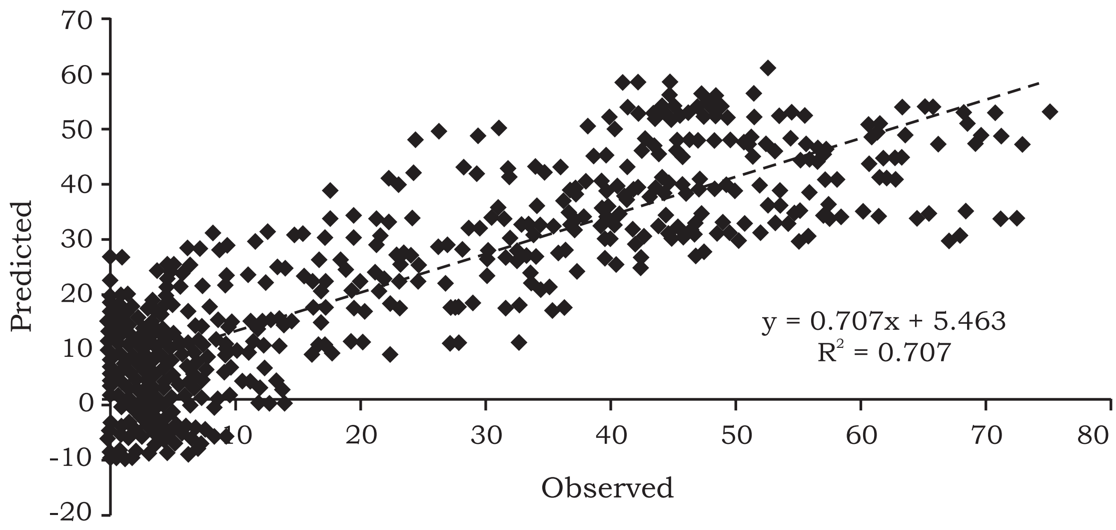

Figure 2. Comparison of observed & predicted data points of PLRV Disease incidence during years 2002 to 2012 in Faisalabad

Table 3. Models summary with input variables for multiple regression model

Model	Input variables	R square	Adj. R Square	Std. Error of Estimate	R Square Change	F Change	C*p*	*Pr* <*F*
1	Tmin	0.680	0.68	11.749	0.68	1526.71	64	0.001
2	Tmin, RH	0.692	0.691	11.535	0.012	27.99	36	0.001
3	Tmin, RH, RF	0.701	0.698	11.408	0.009	17.02	21	0.001
4	Tmin, RH, RF, APH	0.704	0.705	11.269	0.002	18.69	19	0.001

Table 4. Residual Statistics for multiple regression model

	Min.	Max.	Mean	Standard Deviation
Predicted Value	-10.59	46.720	14.180	14.920
Std. Predicted Value	-1.659	2.180	0.000	1.000
Std. Error of Predicted Value	0.892	4.529	1.635	0.581
Residual	-16.156	17.627	0.000	8.347
Std. Residual	-1.901	2.075	0.000	0.982
Stud. Residual	-1.989	2.168	0.001	1.004
Cook's Distance	0.000	0.072	0.008	0.011

used to run the model I with most contributing factors i.e., minimum temperature, relative humidity and rainfall, influencing the PLRV disease development. The fitted line plot for the disease incidence observed during 2 years dataset and predicted disease incidence based on the model (y = `6.89 + 2.5x1- 0.265x2 + 1.326x3) developed on ten- year dataset. At 95 % PI, there is a good match between observed data with that predicted by model with 67.5 % variability in disease development (Fig. 3), showing that model may be used for prediction scenarios under similar set of environmental conditions.

Our findings were in accordance with the work of Khan and Abbas (2006) who determined that there is significant influence of air temperature and relative humidity on PLRV disease incidence with 60% of the variability in disease development. Under field conditions the PLRV disease predictive model was adequate for forecasting. Stepwise regression analysis was used in identifying the most influential variables, towards developing a partial model. Stepwise model-building techniques have been employed extensively by plant pathologists to quantitatively describe the relationship of weather variables to disease intensity and severity whereas accurate regression models were successfully derived in several studies (Coakley et al., 1985; Jacobi et al., 1983). Minimum temperature, relative humidity and rainfall appeared to be the main contributing environmental variables in the stepwise regression analysis. Maximum temperature had no significant effect, so it was dropped during regression analysis. In forward selection of stepwise regression analysis aphid population was also dropped, although the excluded variable had a significant effect, stepwise analysis only retained the most influential variables. Aphid is the carrier of PLRV and had significant effect but in stepwise analysis it was dropped as its contribution was negligible. Transmission of PLRV through aphid plays a crucial role in the spread of

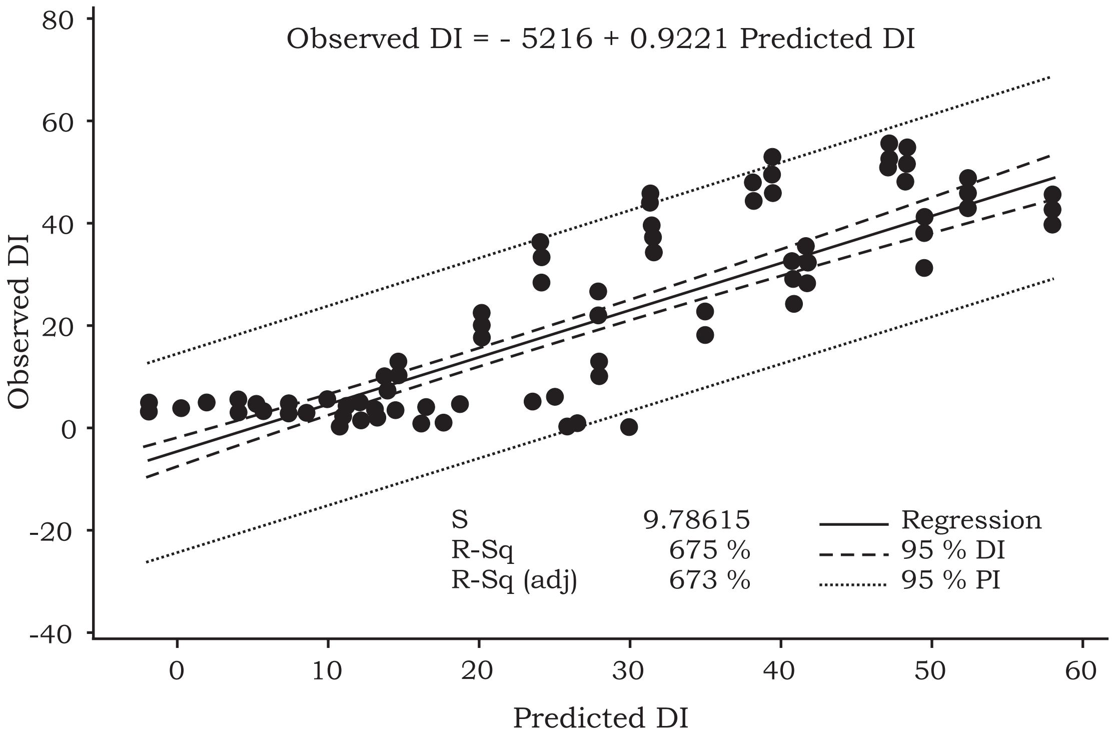

Figure 3. A fitted line plot for PLRV disease incidence with observed and predicted data points at 95% CI and PI

the disease whereas the temperature can affect the development and transmission efficiency of the virus (Syller, 1996). The influence of temperature on the aphid transmission of PLRV is very complex. Transmission is influenced by environmental factors, particularly temperature that may change aphid behavior during acquisition or inoculation (Syller, 1996). There are high correlations of PLRV and PVY with accumulated air temperatures between 1st January and 30th April or 1st July, in different areas (Gabriel, 1965). High correlation coefficients of tempe-rature, degree-days and solar radiation with PLRV have also been observed (Gabriel, 1965; Kemp and Troup, 1978; Harrington et al., 1989; Werker et al, 1998). The aphid population may also be influenced by rainfall and heavy rainfall may damage the winged forms before their first flight (Bonnemaison, 1951). Although amount of rainfall was very low yet it has pronounced effect in disease development as this variable was retained after stepwise regression. There is very low precipitation about 125mm annually in the study area though it has great effect on the aphid population. Where these environmental factors directly affect PLRV disease development, they also affect aphid population. Hayat Zada and colleagues in 2004 developed multiple regression models for aphid population under different weather conditions in Sawat, Pakistan and described the correlation of potato aphid population with temperature, relative humidity and rainfall. It supports our findings that decreased minimum temperature and increased relative humidity and rainfall decreased the aphid population, therefore, it is possible that these environmental factors reduce aphid populations and therefore affect viruliferous aphids and thus the virus transmission rates. Consequently, these factors indirectly affect PLRV disease development. Therefore only the three variables shown in table 2 were evaluated for PLRV disease development using the ten-year dataset.

New two years dataset of PLRV disease incidence, environmental variables and aphid population was used to run the the model based on ten years of dataset. There was a good fit in the predicted model with 67.5 % variability in the disease development showing that model may be used for prediction under the field conditions.

Taking inference from the above cited models it is clear that two environmental variables i.e., minimum temperature and relative humidity are the important factors for the development of PLRV disease in the Faisalabad region of Pakistan. As environmental conditions play crucial role in PLRV disease development, the predictive models would be helpful for farmers in taking decisions of timely spray to control the vector for the better management of disease.

ACKNOWLEDGEMENT

We are highly obliged to Dr. Anwar S. Rizvi for his guidance and help in getting antibodies. Special thanks to the staff of Plant Virology Section, Ayub Agriculture Research Institute, Faisalabad for cooperating in providing the 10 years data. We acknowledge Agdia Inc. USA for providing antibodies for research. This manuscript is a part of Ph.D. dissertation of Ummadud-Din Umar, submitted to the University of Agriculture, Faisalabad, Pakistan.

LITERATURE CITED

Ahmad, M. and W. Ahmad.1995. Detection of major potato viruses from different potato growing localities of Punjab. Nat. Sem. Res. Dev. Potato Prod in akistan, PSPDP/ PARC, Islamabad. p. 39-43.

Anonymous. 2010. Pakistan statistical year Book. Fed Bur Stat Statistical division, Karachi, Pakistan.

Bhutta, A. R. and M. F. J. Bhatti. 2002. Seed potato certification in Pakistan. Federal Seed Certification and Registration Department, Ministry of Food Agriculture and Livestock, Islamabad. Pakistan. p. 66.

Bonnemaison, L. 1951. Contribution 'a l' etude facteursprovoquant l' apparition des forms ailees et sexuees chez les Aphidinae. Paris : Dunod.

Chattefuee, S and A. S. Hadi. 2006. Regression analysis by example. Jhon Wiley and Sons, New Jersey.

Clark, M. F and A. N. Adams. 1977. Characteristics of the microplate method of enzymelinkedimm-unosorbent assay for the detection of plant viruses. J. General Virology. 34: 475-483.

Coakley, S. M., L. R. McDaniel and G. Shaner. 1985. Model for predicting severity of septoria tritici blotch on winter wheat. Phytopathology. 75:1245-1251.

Gabriel, W. 1965. The influence of temperature on the spread of aphidborne potato virus diseases. Annals of Applied Biology. 56: 461-475.

Harrington, R., A. M. Dewar and B. George.1989. Forecasting the incidence of virus yellows in sugar beet in England. Annals of Applied Biology.114: 459-469.

Harrison, B. D.1984. Potato leafroll virus. CMI / AAB Descriptions of Plant Viruses. p. 291.

Hayat-Zada, I. K., S. K. Khalil, S. Ahmed and M. Naeem. 2004. Modeling population dynamics of potato aphids in Swat. Pakistan Entomology. 26: 111-117.

Jacobi, W. R., C. E. Main and N. T. Powell. 1983. Influence of temperature and rainfall on the development of tobacco black shank. Phytopathology. 73:139-143.

Johnson, R. A and D. W. Wichem. 1982. Applied Multivariate statistical analysis. Englewood Cliffs, Prentice-Hall, New Jersey p. 152-156.

Kemp, W. G. and P. A. Troup. 1978. A weather index to forecast potential incidence of aphidtransmitted virus diseases of peer in the Niagara Peninsula. Canadian Journalof Plant Sciences. 58: 1025-1028.

Khan, M. A. and J. ObaidullahI. 2006. Identification of resistant sources against potato leaf roll virus and M. persicae by biological tests and ELISA. Pakistan J. Phyto. 8: 191-198.

Khan, M. A. and W. Abbas. 2008. Evaluation of multiple regression models based on epidemiological factors to predict M. persicae population and PLRV disease incidence. Int Nat Conf Plant Scientists. 21-24 April, 2008. ICPS. p .155-168.

Mughal, S. M. 2003. Some threatening and emerging plant viral diseases in Pakistan. Proce. 4th Nat Conf Plant Pathology. p. 8-16.

Myers, R. H. 1990. Classical and modern regression with Applications. PWS-KENT publishing company, Boston, USA. 488.p.

Salazar, L. F. 1996. Potato Viruses and Their Control. The International Potato Center, Peru. 214.p.

Snee, R. D. 1977. Validation of regression models: methods and examples. Technometrics.19: 415-428.

Syller, J. 1996. Potato leafroll virus (PLRV): its transmission and control. Integrated Pest Management Review. l: 2l7-227.

Werker, A. R., A. M. Dewar and R. Harrington. 1998. Modeling the incidence of virus yellows in sugar beet in the UK in relation to numbers of migrating, Myzus persicae. J. Appl. Eco. 35: 811-818.

AUTHORSHIP AND CONTRIBUTION DECLARATION

S.No	Author Name	Contribution to the paper
1.	Mr. Ummad-Ud-Din Umar	Concived the idea, Methodology, Result and Discussion, wrote abstract, Introduction
2.	Mr. Muhammad Aslam Khan	Technical input at every step
3.	Mr. Ateeq-ur-Rehman	Overall management of the article
4.	Mr. Abdul Hannan	Did SPSS analysis, Conclusion
5.	Mr. Syed Atif Hasan Naqvi	Data entry in SPSS and analysis, References
6.	Mr. Azhar Ali Khan	Data collection
7.	Mr. Muhammad Asif Zulfiqar	Technical input at every step

(Received March 2016 and Accepted july 2016)