Bull. Hiroshima Inst. Tech. Research Vol. 46 (2012) 173–179

Article

Non-parametric Modelling and Its Application on Real-life Data

Prasenjit PAL*, Subhabaha PAL**, Satyabrata PAL*** and Sanpei KAGEYAMA****

(Received Oct. 13, 2011)

Abstract

Some non-parametric models (kernel, loess and spline) have been compared with different parametric models in order to judge their precision by measuring the proximity between the observed and the predicted values based on the data-sets on the pest incidence (measured in terms of the number of pests) in respect of two important crops, namely, brinjal and chilli, with a view to describing the pattern of the incidence of pests over the year based on weekly incidence data (28 weeks). The importance of such studies stems from the fact that the farmers need the specialized scientific information in the form of advisory services in respect of the pattern of pest incidence over the year in advance and creation of pest-models helps us to obtain such important information. This paper is devoted to generate pest-incidence models in cases of brinjal and chilli crops. Such pieces of generated information assist the farmers in their planning and execution of the cultivation of the said crops over the year. It is revealed that the precision levels (as are obtained by calculating the actual values of the well-known R^2 criterion) are much higher when the non-parametric models are applied on the data-sets considered under the purview of the paper, in comparison to other methods exploited to obtaining the above-said information.

Key Words: Kernel, Spline, Loess, Parametric model, Non-parametric model, R^2 criterion

1. Introduction

A model is a representation of the outcomes of a real–life phenomenon presented in terms of analytical rules expressed by means of equations developed on the basis of an observed data-set. As presented in the Abstract that extensive data were collected on the incidence (measured by the number of pests) of different pests on the crops, namely, Brinjal and Chilli, using fixed plot technique, the outputs (predicted values of the incidences related to different pests) obtained after fitting of the different models generate a wealth of information related to the dynamics of the infestation-pattern of insects over a year. The fact is that there exist problems in developing models on the dynamics of pest incidence (pest population) on crops grown in field owing to the scarcity of information on the incidences of pest population over a long period of time (a year), hence, indeed, on their dynamics. Unless and otherwise planned long-term experiments are conducted for the purpose, hardly, anyone is able to collect the chronological data on pest population over a considerable duration of time to facilitate the studies on modeling of pest incidences in respect of the above crops of interest. So, from natural pest control point of view, modelling of incidences (Agarwal and Kumar, 2007) of these predator-insects also has a very important role in their management. The dynamics of the

* NCIPM, LBS Building, IARI Campus, New Delhi-12,India
** Sikkim Manipal University and Consultant, Manipal Universal, Bangalore, Karnataka, India
*** Regent Education and Research Foundation, Barrackpore, West Bengal, India
**** Department of Environmental Design, Hiroshima Institute of Technology, Hiroshima 731–5193, Japan

populations of pests or insect vectors of disease are complex in nature, and consequently the models representing the prediction of their outbreaks are difficult to achieve. To help unravel such complex systems and to understand how pests and vectors interact with the environment and with other organisms as reflected in the dynamics, attempts have been made in this paper to apply different models (parametric, non-parametric and semi-parametric) on the data on pest incidence obtained from the planned year-long field experiments on two important crops, Brinjal and Chilli, in respect of a specific ecological situation. Indeed, the non-parametric and semi-parametric methods have revealed their worth in modelling real-life data on pest infestations mentioned above owing to the degree of closeness of the predicted values (obtained from such models fitted on the data-sets) to the observed values they can produce. Other related references are Markkula (2003), Matis et al. (2005), Kageyama et al. (2006), and Agrawal and Kumar (2007).

2. Materials and methods

For the purpose of formulating models on the data mentioned in the Abstract, the rainy season data were not considered as there were almost no incidences of pests at that time for both the crops. In this study we have used the data on the infestation of the pests (Thrips, Jassids, Whitefly, Borer) on Brinjal, and pests (Whitefly, Yellow Mite, Thrips) on Chilli for the period (September, 2007 to March, 2008) obtained from an experimental station located in Southern part (Gayespur farm in Bidhan Chandra Krishi Viswavidyalaya) of West Bengal, India. A reference to the methods employed in this paper is given below.

2.1 Parametric methods

Parametric models can broadly be categorized into linear and non-linear. Different parametric models as in Table 1 like, linear model, parabolic model, polynomial model, logarithmic model, Gompertz model, etc. are tried.

2.2 Non-parametric methods

The non-parametric models, Kernel model and Loess model are employed, also used is the semi-parametric model, Spline model. Such models are well known and can be found in Thisted (1988), Simonoff (1996), Härdle et al. (2004) and Kageyama et al. (2006).

Table 1: Forms of different parametric models

Linear	$b_0 + b_1 t$
Quadratic	$b_0 + b_1 t + b_2 t^2$
Cubic	$b_0 + b_1 t + b_2 t^2 + b_3 t^3$
Compound	$b_0\, b_1^{\,t}$
Logarithmic	$b_0 + b_1 \log t$
Inverse	$b_0 + b_1 / t$
S Type	$e^{b_0 + b_1 / t}$
Exponential	$b_0 e^{b_1 t}$
Power	$b_0 t^{b_1}$
Gompertz (GRO)	ab^z, where $z = c^t$

3. Results and discussions

The descriptive statistics in Tables 2 and 3 next concerning the incidences of the different pests in case of the crops, Brinjal and Chilli, for the year 2007–2008 are presented. The salient features, which are revealed, are:

(1) In case of Brinjal, there is maximum variability in the incidences in case of the pest, Whitefly and least variability (almost no variability) in the incidences in case of the pest, Borer. The incidence data are positively skewed in all cases. The incidence diagrams, if prepared from the observed data (on data on incidence of pests), will be seen to exhibit platykurtic (more flatness over the entire range when the incidence data are plotted on the Y axis against months on the X axis) feature in case of all types of pests.

(2) In case of Chilli, there is maximum variability in the incidences in case of the pest, Yellow Mite and least variability in the incidences in case of the pest, Whitefly. The incidence data, however, exhibit positive skewness in all cases. The incidence diagrams, if prepared from the observed data (on incidence of pests), will be seen to exhibit platykurtic (more flatness over the entire range) feature in case of all (excepting the Yellow Mite) pests.

Among the parametric models fitted, it is observed that the best fitted models in Tables 4 and 5 next are found to be cubic (based on R^2 criterion, values of R^2 lying between 0.6 to 0.8) in all the cases (for the two crops and over different pests). Though the obtained R^2 values are satisfactory, a search for models which

Table 2: Descriptive statistics with respect to the data on the incidence (number of pests) of different pests on Brinjal crop

Pest	N	Min	Max	Mean	Variance	Skewness	Kurtosis
Whitefly	28	0.63	15	6.59	22.64	0.34	−1.19
Thrips	28	0.14	11.99	4.72	14.83	0.45	−1.13
Jassid	28	0.10	5.12	1.95	2.03	0.86	−0.44
Borer	28	0.01	0.83	0.22	0.07	0.97	−0.21

Table 3: Descriptive statistics with respect to the data on incidence (number of pests) of different pests on Chilli crop

Pest	N	Min	Max	Mean	Variance	Skewness	Kurtosis
Whitefly	28	0.20	3.93	1.49	1.62	1.01	−0.49
Yellow Mite	28	0.01	18.60	4.96	35.20	1.33	0.47
Thrips	28	0.01	14.00	4.48	20.71	0.75	−0.88

Table 4: R^2 values for different pests of Brinjal (2007–2008)

R^2	Lin	Log	Inv	Qua	Cub	Com	Pow	S	GRO	Expo
Whitefly	0.020	0.025	0.110	0.796	0.807	0.040	0.023	0.177	0.040	0.040
Thrips	0.014	0.033	0.117	0.796	0.830	0.010	0.062	0.220	0.010	0.010
Jassids	0.001	0.068	0.141	0.774	0.775	0.007	0.055	0.198	0.007	0.007
Borer	0.509	0.521	0.438	0.509	0.639	0.453	0.647	0.754	0.453	0.453

Table 5: R^2 values for different pests of Chilli (2007–2008)

R^2	Lin	Log	Inv	Qua	Cub	Com	Pow	S	GRO	Expo
Whitefly	0.210	0.047	0.012	0.287	0.622	0.128	0.004	0.082	0.128	0.128
Yellow Mite	0.027	0.008	0.056	0.506	0.602	0.043	0.015	0.120	0.043	0.043
Thrips	0.032	0.137	0.138	0.622	0.637	0.246	0.550	0.588	0.246	0.246

can produce higher precision levels (in terms of higher values of R^2) needs to be carried out.

In this paper, along with the parametric models, two non-parametric and one semi-parametric models have been tried. The R^2 values as presented in the graph plots (in the Annexure) next indubitably confirm the superiority of the representative power (in terms of higher precision) of the non-parametric (and semi-parametric) modelling in comparison to parametric modelling, when called for, in situations to depict the pattern of pest incidence based on weekly pest-incidence data on two important crops, namely, Brinjal and Chilli. As the reality demands that the farmers are keen to be provided with specialized scientific information in the form of advisory services in respect of the pattern of pest incidence over the years in case of the major agricultural crops, models built up on the real-life data on pest infestations reveal the patterns of intensity of pest incidence importantly useful to them. The implication and importance of this study are that the non-parametric model like Kernel and Loess and semi-parametric model like Spline are capable to produce forecasts on the basis of magnitude of incidences (data acquired from field experiments) at any time-point during the year most precisely so as to provide accurate advisory services against possible pest attacks to crop, an advance knowledge of which is of utmost help to the farmers. It is no denying that to model the fluctuations/characteristics/dynamics evidenced in temporal real-life data situations, the non-parametric and semi-parametric models offer better representations almost often.

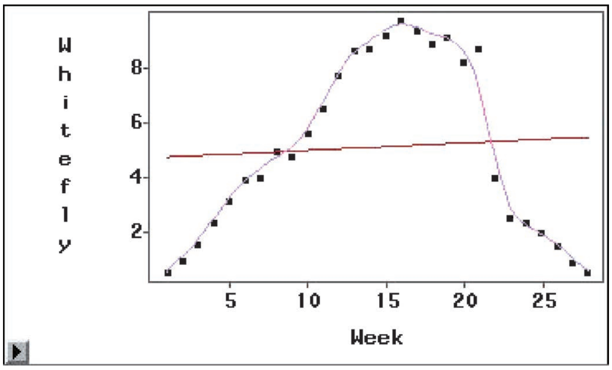
ANNEXURE

Crop – Brinjal

Y Axis - No. of incidences

X Axis – Week numbers

Graph 1: Spline Fit (Whitefly)



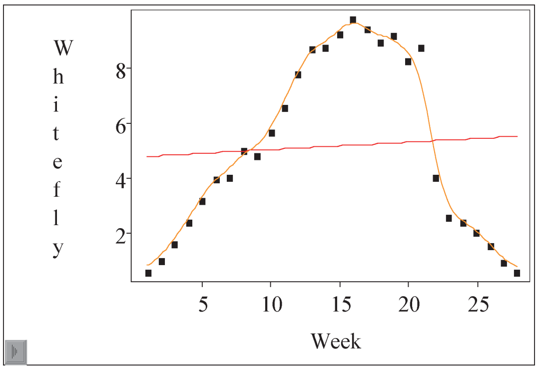
R^2 = 0.9879

Figure demonstrating Spline modelling of Whitefly incidence on Brinjal data

Red Line – Linear Fit

Violet line – Spline fit (good fit – nearing 1.0)

Graph 2: Kernel Fit (Whitefly)



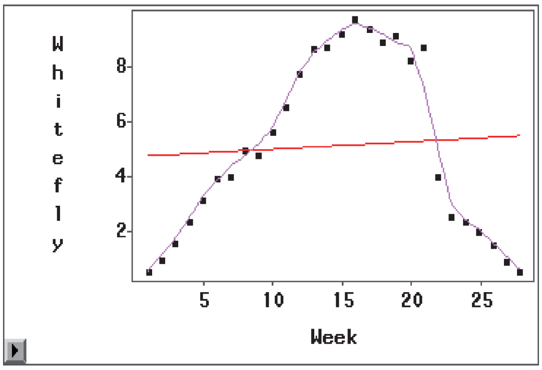
R^2 = 0.9878

Figure demonstrating Kernel modelling of Whitefly incidence on Brinjal data

Red Line – Linear Fit

Orange line – Kernel fit (good fit – nearing 1.0)

Graph 3: Loess Fit (Whitefly)



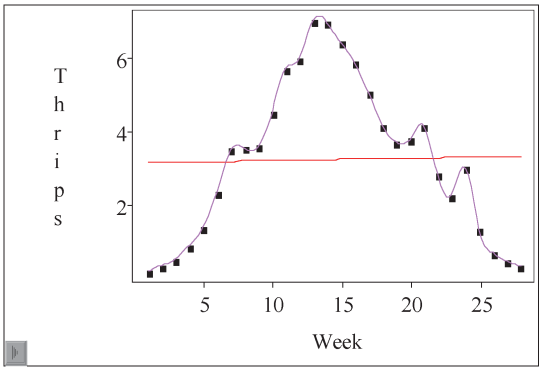
R^2 = 0.9855

Figure demonstrating Loess modelling of Whitefly incidence on Brinjal data

Red Line – Linear Fit

Violet line – Loess fit (good fit – nearing 1.0)

Graph 4: Spline Fit (Thrips)



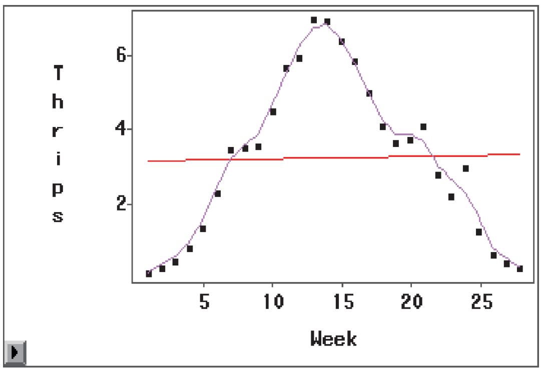
R^2 = 0.9999

Figure demonstrating Spline modelling of Thrips incidence on Brinjal data

Red Line – Linear Fit

Violet line – Spline fit (good fit – nearing 1.0)

Graph 5: Loess Fit (Thrips)



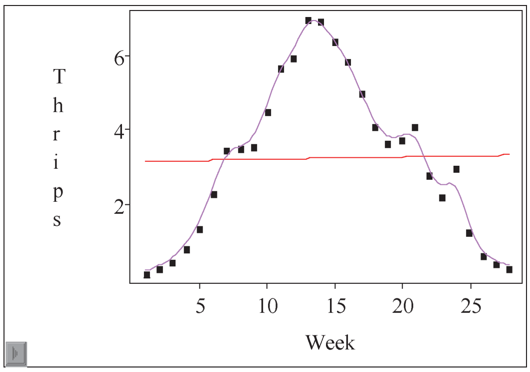
R^2 = 0.9881

Figure demonstrating Loess modelling of Thrips incidence on Brinjal data

Red Line – Linear Fit

Violet line – Loess fit (good fit – nearing 1.0)

Graph 6: Kernel Fit (Thrips)



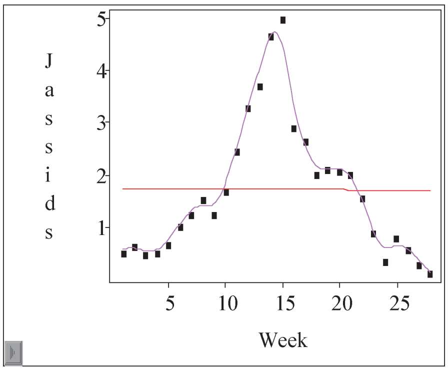
R^2 = 0.9931

Figure demonstrating Kernel modelling of Thrips incidence on Brinjal data

Red Line – Linear Fit

Violet line – Kernel fit (good fit – nearing 1.0)

Graph 7: Spline Fit (Jassid)



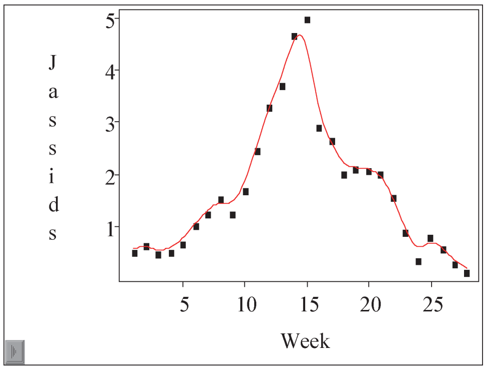
R^2 = 0.9841

Figure demonstrating Spline modelling of Jassid incidence on Brinjal data

Red Line – Linear Fit

Violet line – Spline fit (good fit – nearing 1.0)

Graph 8: Kernel Fit (Jassid)

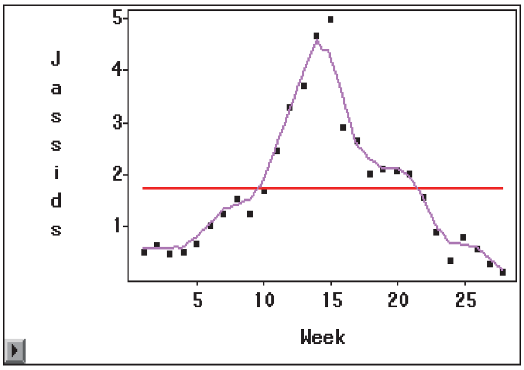


R^2 = 0.984

Figure demonstrating Kernel modelling of Jassid incidence on Brinjal data

Red line – Kernel fit (good fit – nearing 1.0)

Graph 9: Loess Fit (Jassid)



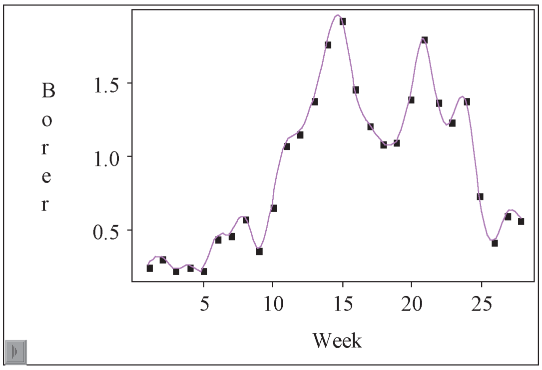
R^2 = 0.9746

Figure demonstrating Loess modelling of Jassid incidence on Brinjal data

Red line – Linear fit

Red line – Loess fit (good fit – nearing 1.0)

Graph 10: Spline Fit (Borer)

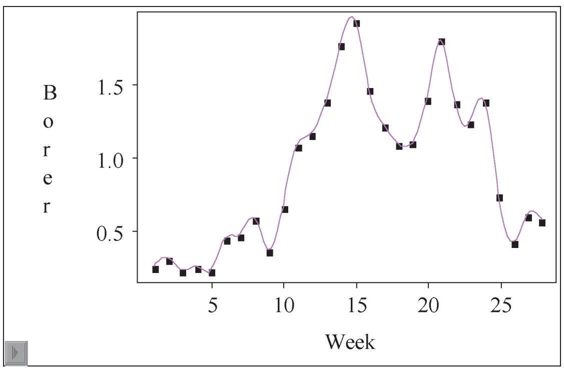


R^2 = 0.9999

Figure demonstrating Spline modelling of Fruit and Shoot Borer incidence on Brinjal data

Violet line – Spline fit (good fit – nearing 1.0)

Graph 11: Kernel Fit (Borer)

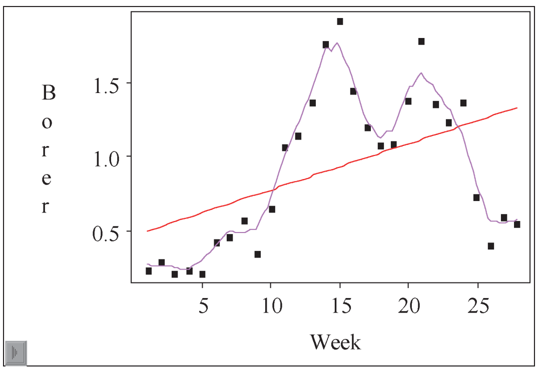


R^2 = 0.9815

Figure demonstrating Kernel modelling of Fruit and Shoot Borer incidence on Brinjal data

Violet line – Kernel fit (good fit – nearing 1.0)

Graph 12: Loess Fit (Borer)



R^2 = 0.9649

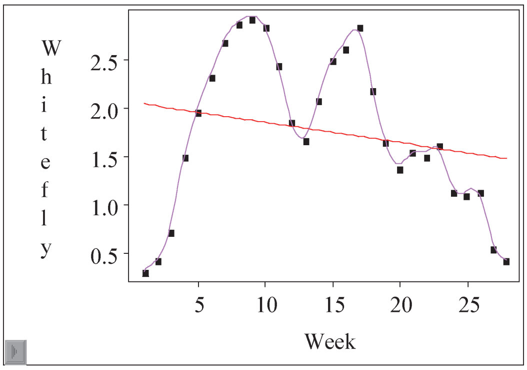
Figure demonstrating Loess modelling of Fruit and Shoot Borer incidence on Brinjal data

Red line – Linear fit

Violet line – Loess fit (good fit – nearing 1.0)

Crop – Chilli

Graph 13: Spline Fit (Whitefly)



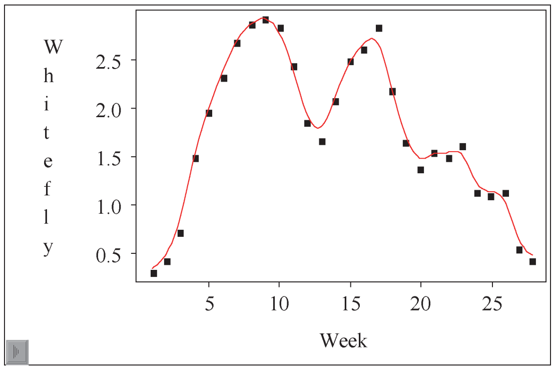
R^2 = 0.9984

Figure demonstrating Spline modelling of Whitefly incidence on chilli data

Red line – Linear fit

Violet line – Spline fit (good fit – nearing 1.0)

Graph 14: Kernel Fit (Whitefly)

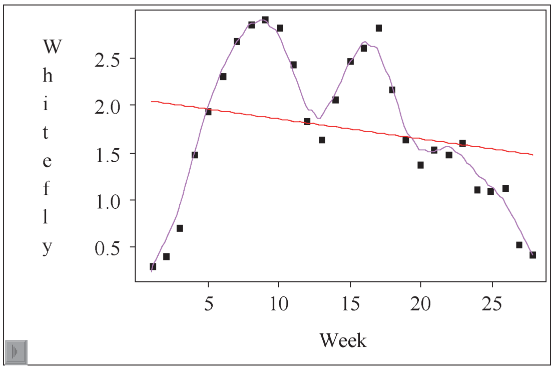


R^2 = 0.9922

Figure demonstrating Kernel modelling of Whitefly incidence on chilli data

Red line – Kernel fit (good fit – nearing 1.0)

Graph 15: Loess Fit (Whitefly)



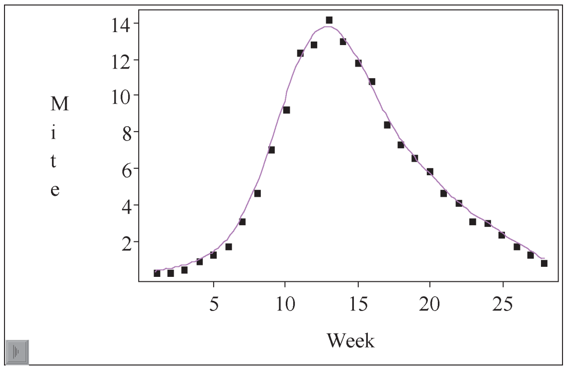
R^2 = 0.9828

Figure demonstrating Loess modelling of Whitefly incidence on chilli data

Red line – Linear fit

Violet line – Loess fit (good fit – nearing 1.0)

Graph 16: Spline Fit (Yellow Mite)

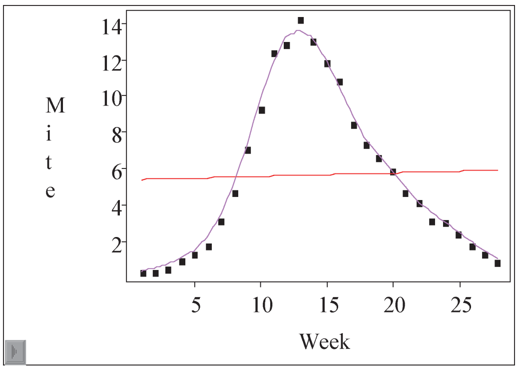


R^2 = 0.9973

Figure demonstrating Spline modelling of Yellow Mite incidence on chilli data

Violet line – Spline fit (good fit – nearing 1.0)

Graph 17: Kernel Fit (Yellow Mite)



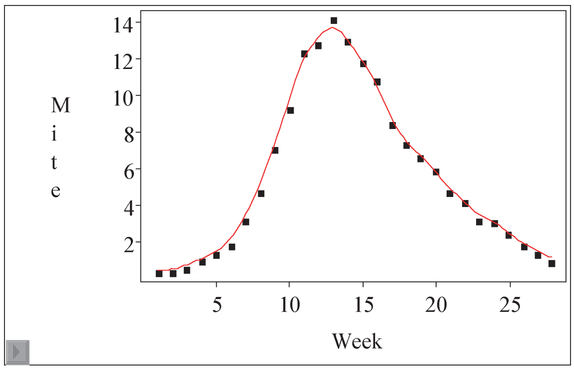
R^2 = 0.9978

Figure demonstrating Kernel modelling of Yellow Mite incidence on chilli data

Red line – Linear fit

Violet line – Kernel fit (good fit – nearing 1.0)

Graph 18: Loess Fit (Yellow Mite)

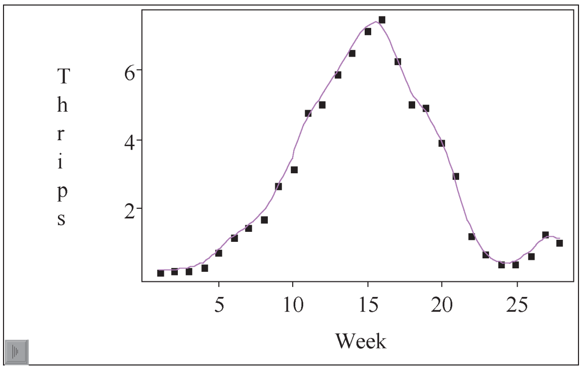


R^2 = 0.9965

Figure demonstrating Loess modelling of Yellow Mite incidence on chilli data

Red line –Loess fit (good fit – nearing 1.0)

Graph 19: Spline Fit (Thrips)

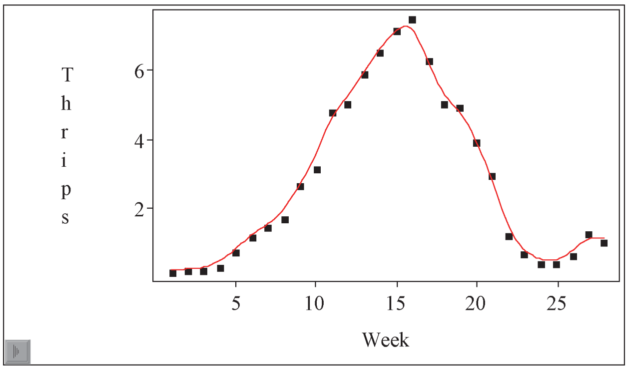


***R^2* = 0.9966**

Figure demonstrating Spline modelling of Thrips incidence on chilli data

Violet line – Spline fit (good fit – nearing 1.0)

Graph 20: Kernel Fit (Thrips)

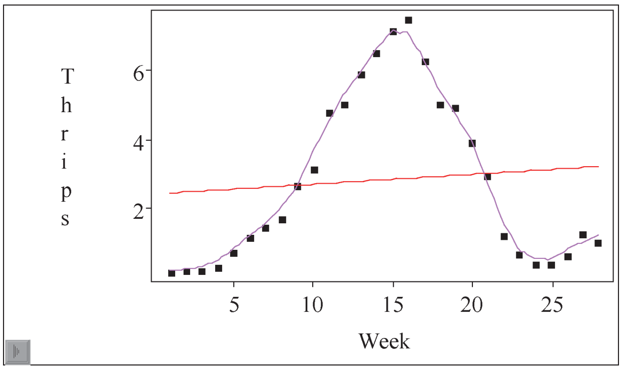


***R^2* = 0.9959**

Figure demonstrating Kernel modelling of Thrips incidence on chilli data

Red line – Kernel fit (good fit – nearing 1.0)

Graph 21: Loess Fit (Thrips)



***R^2* = 0.9932**

Figure demonstrating Loess modelling of Thrips incidence on chilli data

Red line – Linear fit

Violet line – Loess fit (good fit – nearing 1.0)

References

Agrawal, R. and Kumar, A. (2007). Forecasting of crop yield using weather parameters and agricultural inputs. *J. Indian Soc. Agric. Statist.* **61** (3): 317.

Härdle, W., Müller, M., Sperlich, S. and Werwatz, A. (2004). *Nonparametric and Sem-parametric Models*. Springer Series in Statistics, Springer-Verlag. http://private.igf.edu.pl/~jnn/Literatura_tematu/Nonparametric%20and%20Semiparametric%20Models-2004.pdf

Kageyama, S., Pal, S., Sahu, P., Pal, S. and Ghosh, S. (2006). Modelling pest incidence – A nonparametric approach. *Jour. Combin. Infor. and Sys. Sci.* **31**, *Issue* 1–4, 113–126.

Markkula, I. (2003). Pest databases, monitoring and forecasting of pests and diseases in Finland. Nordic Association of Agricultural Scientists 22nd Congress, July 1–4, 2003, Turku, Finland.

Matis, J. H., Kiffe, T. R., Matis, T. I. and Stevension, D. E. (2005). Application of population growth models based on cumulative size to pecan aphids. (Unpublished manuscript)

Simonoff, J. (1996). *Smoothing Methods in Statistics*. Springer, New York.

Thisted, R. A. (1988). *Elements of Statistical Computing*. Chapman and Hall, New York.