Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 70(3) 2016 277–285

Development of Non-linear Models for Forecasting

Anil Kumar[1] and Sanjeev Panwar[2]
[1]*ICAR-Indian Agricultural Statistics Research Institute, New Delhi*
[2]*Indian Council of Agricultural Research, New Delhi*

Received 04 June 2016; Accepted 29 November 2016

SUMMARY

An attempt has been made to develop non-linear models for forecasting purpose. In the proposed models, more than one explanatory variables have been used. Two approaches are considered for building up the models. In the first approach, exact relation between the response variable and the explanatory variables are linearly added to get the final model. In the second approach, time variable of a non-linear model is directly replaced by a linear relationship of explanatory variables. Proposed models are illustrated with a real life data of aphid population over time. It has been found that the model based on the first approach gives better forecast results.

Keywords: Non-linear model, Forecast, Aphid, Mustard, Estimation, Linear regression.

1. INTRODUCTION

Forecasting technique is an important tool in any field. It provides important and useful input for proper, foresighted and informed planning, more so, in agriculture which is full of uncertainties. Now-a-days agriculture has become highly input and cost intensive. Without judicious use of fertilisers and plant protection measures agriculture no longer remains as profitable as before. Uncertainties of weather, production, policies, prices etc. lead to mass suicides by farmers. New pests and diseases are emerging as an added threat to the production. Under the changed scenario today, forecasting of various aspects relating to agriculture are becoming essential.

Forecasting in agricultural system is not new. A lot of studies related to various aspects of agriculture clearly indicate its importance. A lot of studies including the latest technologies like GIS and Remote Sensing have been extensively used to forecast crop production, pest/disease infestation etc. However, achievements in this direction are not up to the mark. One of the reasons might be lack of use of proper statistical techniques. Most of the earlier workers have utilised regression models (taking data as such or suitable transformation of data or some indices), discriminant function analysis, agro-meteorological models, etc. for crop yield forecasting (to cite a few, Chakraborty *et al*. 2004, Agrawal and Jain 1996, Prasad and Dudhane 1989, Kumar and Bhar 2005). Off course there are other approaches for forecasting also. For example, Matis *et al*. (1985) proposed a statistical methodology for forecasting crop yields at successive stages of the growing season of any crop using Markov Chain theory. Other work on this subject are due to Matis *et al.* (1989), Jain and Agrawal (1992) and Ramasubramanian and Bhar (2014). Application of Time Series modelling in forecasting is very popular. Some examples in this field are due to Mcleod and Vingilis (2005), Suresh *et al.* (2011), Paul (2015) and Paul *et al.* (2015).

Corresponding author: Anil Kumar
E-mail address: akumarchi72@gmail.com

Survey of existing literature reveals that in most of these studies, simple linear regression models are employed, whereas relationship exists in most of the Agricultural situations is non-linear. For example, disease/pest infestation over a time generally follows a growth curve which is non-linear in nature. The most of the biological phenomena in agriculture generally follow a non-linear pattern over time. Off course some attempts are made to model such phenomena by non-linear modeling techniques. Time Series models also cannot take account of the explanatory variables. But in non-linear models the response, *i.e.*, observed values are modelled with time variable only. However, generally response variable is also governed by many affecting factors. For example, disease/pest infestation over time may also be affected by prevailing weather conditions like temperature, relative humidity etc. But in non-linear model building, incorporation of these weather variables along with the time variable is somewhat impractical; reason being theoretical. Theory of non-linear modelling with more than one explanatory variables is not well established. We are somewhat handicapped in exploiting the information on many explanatory variables while developing a non-linear model. But we are convinced that use of these variables definitely improve the quality of the models and hence the performance of forecasting.

In the present study, an attempt has been made to exploit the information on other explanatory variables like temperature, relative humidity in non-linear models. The approach is empirical, but definitely, it will provide a good framework for building forecasting models using non-linear modelling technique. It will open up an area, where researchers will be able to utilize their resources to the full extent. The developed methodology has been illustrated with a real life data set pertaining to aphid population in mustard over time in Beherampore district of Odisha. Aphid population over time is a typical example of non-linear nature of a variable. In the literature, such type of data is fitted through non-linear modelling technique by taking time as the only explanatory variable. Through this example, we have shown how other important affecting weather variables like temperature, relative humidity can be made used to improve the forecasting ability of a non-linear model.

The paper is organized as follows: In Section 2, proposed models are discussed. Two approaches are made to utilize the full information on explanatory variables. Various aspects of fitting non-linear models are also discussed in this Section. Illustration of the proposed models are considered in Section 3. Real aphid population in mustard in Beharampore district of Odisha has been taken for this purpose. Temperature and relative humidity are two variables that are used as explanatory variables. The paper is concluded with a Section on Discussion.

2. PROPOSED METHODS

Suppose we have p explanatory variables $x_1, x_2,...,x_p$ and y as the response variable. For simplicity, let us consider that we have only three explanatory variables x_1, x_2 and x_3. More clearly, these three explanatory variables may be temperature, relative humidity and rainfall and the response variable y may be aphid population. As a first approach for building up of forecasting model, one may consider a multiple linear regression model as

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + e \quad (1)$$

where α is the intercept and β_1, β_2 and β_3 are three regression coefficients associated with three explanatory variables or regressor variables. The logic behind such model might be that each of the explanatory variables is highly correlated with the response variable and their relationship is linear. That is x_1 is related linearly with y. Similarly, x_2 and x_3 are also related linearly with y. Thus each of these variables multiplied by a constant (regression coefficient) and added linearly to have a multiple linear model as given in (1).

But as mentioned earlier, these relationships may not be linear in practice. Therefore, question may arise why not to exploit exact relationship between an explanatory variable and the response variable. This exact relationship may be non-linear of different kind or even a linear. This may be sigmoidal or curvi-linear and so on. If the exact relationship between an explanatory variable and the response variable is correctly accommodated in the model then it may, perhaps give a better forecast result. The present paper aims to use this idea and provide a better non-linear model building technique for forecasting purpose.

The proposed methods consist of exploiting the exact relationship between the explanatory variables and response variable. We propose two approaches to build up the models.

Approach 1: Once the relationship for each explanatory variable is established, we add them linearly just like multiple linear regression model in (1) to have the final model. We can also add these non-linearly. But for simplicity, we confine ourselves to linear addition. If we have p explanatory variables, then the proposed model is formulated as

$$y = \sum_{i=1}^{p} f_i\left(x_i\right) + e \tag{2}$$

where $f_i(x_i)$ is the exact relationship between i^{th} explanatory variable and the response variable. Whatever relationship is revealed, we as such put in the model (2). Some of the variables may be related linearly with the response variable.

To be specific, once again, we consider the case with three explanatory variables, x_1, x_2, x_3. Suppose relationship of x_1 with y is logistic, *i.e.*,

$$f_1\left(x_1\right) = \frac{\alpha}{l + \beta \exp\left(-\gamma x_1\right)} \tag{3}$$

If we build up a non-linear model with this single variable x_1, then our model becomes

$$y = \frac{\alpha}{l + \beta \exp\left(-\gamma x_1\right)} + e \tag{4}$$

where α, β and γ are the parameters of the model. But we want to exploit the information on x_2 and x_3 as well. Suppose the relationship of x_2 with y is exponential, *i.e.*,

$$f_2\left(x_2\right) = \lambda \exp\left(-\delta x_2\right) \tag{5}$$

and corresponding non-linear model with x_2 becomes

$$y = \lambda \exp\left(-\delta x_2\right) + e \tag{6}$$

For illustration, let us consider that the third variable x_3 has a linear relationship with y, *i.e.*

$$y = \theta x_3 + e \tag{7}$$

Since our objective is to develop a model with all three explanatory variables, we add all three relationships linearly to get the final model as

$$y = \frac{\alpha}{l + \beta \exp\left(-\gamma x_1\right)} + \lambda \exp\left(-\delta x_2\right) + \theta x_3 + e \tag{8}$$

Approach 2: As mentioned earlier, non-linear model generally fitted with a single variable as explanatory variable. Suppose a non-linear model with a single variable is given by

$$y = \lambda \exp\left(-\delta x\right) + e \tag{9}$$

Generally, x is taken as a time variable t. However, explanatory variables are generally related among themselves. For example, temperature may be highly related to relative humidity and so on. So, we can have a linear relationship among the explanatory variables. For example with three variables, this relationship may be depicted as

$$\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \tag{10}$$

Then why not this relationship should be exploited in non-linear model? Our second approach is based on this idea. We replace x in (9) by the exact linear relation in (10), *i.e.*, we get a non-linear model with three explanatory variables as

$$y = \lambda \exp\left\{-\left(\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3\right)\right\} + e \tag{11}$$

This is again a non-linear model with three explanatory variables x_1, x_2 and x_3 and four parameters λ, β_1, β_2 and β_3. Thus all information of explanatory varaiables can be fully exploited through this approach.

2.1 Estimation of Parameters

Models in (8) and (11) are non-linear models with more than one explanatory variable. In model (8), there are 3 explanatory variables and 6 parameters α, β, γ, λ, δ and θ in model (11), there are again three explanatory variables and four parameters. To estimate these parameters least squares or maximum likelihood method can be employed. By least square method, we minimize the error sum of squares. From (8), the i^{th} error would be

$$e_i = y_i - \frac{\alpha}{l + \beta \exp\left(-\gamma x_{1i}\right)} - \lambda \exp\left(-\delta x_{2i}\right) - \theta x_{3i} \tag{12}$$

and from (11), it is

$$e_i = y_i - \lambda \exp\{\beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}\} \quad (13)$$

If we have n observations, then we minimize a quantity Q given by $Q = \sum_{i=1}^{n} e_i^2$, *i.e.*, we differentiate Q with respect to each of the parameters and equate it to zero. Then we get as many equations as the number of parameters. But like non-linear model with a single variable, also called normal equations will be non-linear in nature. Therefore, iterative least squares method can be applied to obtain the parameter estimates. Therefore, it is not possible to solve nonlinear equations exactly; the next alternative is to obtain approximate analytic solutions by employing iterative procedures. Three main methods of this kind are available in the literature, (i) Linearization (or Taylor Series) method, (ii) Steepest Descent method and (iii) Levenberg-Marquardt's method. The details of these methods along with their merits and demerits are given in Draper and Smith (1998). The linearization method uses the results of linear least square theory in a succession of stages. However, neither this method nor the Steepest descent method, is ideal. The most widely used method of computing nonlinear least squares estimators is the Levenberg-Marquardt's method. This method represents a compromise between the other two methods and combines successfully the best features of both and avoids their serious disadvantages. Does not use derivatives (DUD) procedure is also available in some of the statistical software in which empirically parameters are estimated.

2.2 Goodness of Fit and Diagnostics

For model adequacy checking and diagnostic, usual methods as applied in non-linear model with a single variable can be applied. Model adequacy is generally assessed by the coefficient of determination, R^2. However, as pointed out by Kvalseth (1985), eight different expressions for R^2 appear in the literature. Thus it is confusing that which measure we should take to judge the model. Following Kvalseth (1985) we propose to use R^2 given by

$$R^2 = 1 - \Sigma\left(y_i - \hat{y}_i\right)^2 / \Sigma\left(y_i - \overline{y}_i\right)^2 \quad (14)$$

to test the goodness of fit of the model. However, it is always desirable to use some other summary statistics like Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) along with R^2. These statistics are given by

Mean Absolute Error

$$\text{(MAE)} = \Sigma\left|y_i - \hat{y}_i\right| / n, \quad (15)$$

Mean Squared Error

$$\text{(MSE)} = \Sigma(y_i - \hat{y}_i)^2 / (n-p), \quad (16)$$

where $\hat{y}_i$ is the predicted value of the i^{th} observation and $\bar{y}$ is the grand mean. The better model will have the least values of these statistics while higher value of R^2 is expected.

Uncritical use and sole reliance on the above statistics may fail to reveal important data characteristics and model inadequacies. Additional detailed analysis of the residuals is strongly recommended to decide about the suitability of a model. Two important assumptions made in the model are (i) errors are independent and (ii) errors are normally distributed. These assumptions can be verified by examining the residuals. If the fitted model is correct, the residuals should exhibit tendencies that tend to confirm or at least should not exhibit a denial of the assumptions. The principal ways of plotting the residuals are: (a) in time sequence, (b) against fitted values. To test the independence assumption of residuals run test procedure is available in the literature (Ratkowsky 1990). For testing normality assumption, Kolmogorov-Smirnov test is applied.

3. ILLUSTRATION

In this section we illustrate the proposed methods through a real-life data. Daily data on aphid population from Behrampur, Odisha, has been collected in aphid growing season from 1st December, 2002 to 10th March, 2003. However, for illustration, data pertaining to the period from 25th December, 2002 to 28th February, 2003 has been used. For explanatory variables, daily weather data corresponding to the said period has been considered. Among the weather variables, maximum temperature (x_1), minimum temperature (x_2) maximum relative humidity (z_1) and minimum relative humidity (z_2) have been considered for model building. Again average of maximum and minimum values of temperature and relative humidity have been taken as the final explanatory variables, *i.e.*, $x = x_1 + x_2$ and $z = z_1 + z_2$. Thus finally we have two explanatory variables. The aphid population has been denoted by y. The aphid population over time has been plotted in Fig. 1.

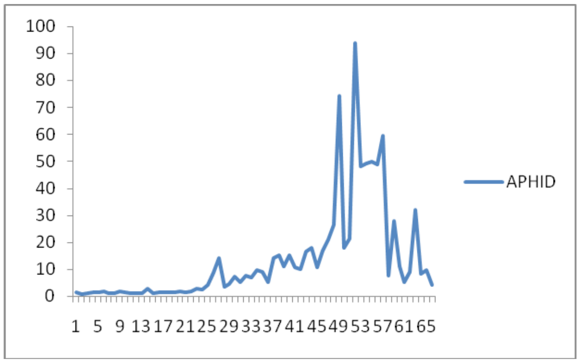

Fig. 1. Aphid population over time

From Fig. 1, it is clear that aphid population over time does not follow any linear pattern and hence a multiple regression model is not at all a choice for the present data. Again as mentioned earlier, fitting a non-linear model with only time variable cannot exploit the information on other explanatory variables. However, for demonstration, we fit a non-linear model of this aphid population against time. The model fitted is a logistic model as follows:

$$y = \frac{\alpha}{\left(1+\beta\ Exp\left(-\gamma t\right)\right)} + \varepsilon, \qquad (17)$$

where α, β and γ are the parameters of the model and t is the time variable. The results obtained by fitting this model is given in Table 1. Table 1(a) clearly shows that the model fit is good (p-value for F-statistic is <0.0001). The parameter estimates of this model along with the standard errors are given in Table 1(b). The final model thus becomes

$$y = \frac{30.29}{1+38671.5\ Exp\left(-0.2653t\right)} \qquad (18)$$

Table 1(a): ANOVA-Logistic Model with *t*

Source	Degree of Freedom	Sum of squares	Mean Sum of Squares	F-value	*p*-value
Model	3	21040.60	7013.50	33.04	< 0.0001
Error	63	13372.60	212.30		
Total	66	34413.20			
R^2	41%				
MAE	0.35				

Table 1(b): Parameter estimate: Logistic model with *t*

Parameter	Estimate	Approx. Std. Error	Approx. 95% Confidence Interval	
α	30.29	4.0374	22.2287	38.3648
β	38671.5	245269	-451459	528802
γ	0.2653	0.1633	-0.0610	0.5917

Table 1(c): Correlation matrix: Logistic model with *t*

	α	β	γ
α	1.000	– 0.446	-0.496
β	-0.446	1.000	0.993
γ	-0.496	0.993	1.000

Though the model fitting is good, yet parameter estimates are unstable which is clear from the correlation table given in Table 1(c). All parameter estimates are highly correlated among themselves. Though calculated value of MAE is low, yet R^2 value is very les which indicates that the model may not give good forecast values. The graph of the observed and predicted values is given in Fig. 2. Fig. 2 clearly indicates that the model could not capture the pattern fully. We, therefore switch to the methodologies as proposed in this paper.

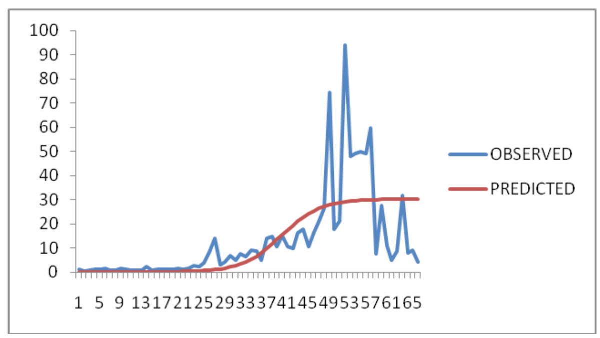

Fig. 2. Observed vs. predicted aphid population by the logistic model with time

In the present attempt, we take aphid population (y) as the response variable and two derived variables as explanatory variables. We first explore all possible relationships between y and x and between y and z. Finally we find that a logistic curve best explains the relation of y with x and between y and z, this relation is exponential. We, therefore, first fit a logistic non-linear model for y by taking x as an explanatory variable and an exponential model for y and z. The models are as follows:

$$\text{Logistic} := \frac{\alpha}{\left(1+\beta\ Exp\left(-\gamma x\right)\right)} + \varepsilon, \qquad (19)$$

$$\text{Exponential: } y = \lambda\ Exp\left(-\mu z\right) + \varepsilon, \qquad (20)$$

where α, β, γ, λ and μ are the parameters of the model and ε is the error term.

We first fit these models separately. For fitting, appropriate programming codes are written in SAS. The results for Logistic model are presented in Table 2(a) to Table 2(c). Results for fitting exponential model with *z* are given in Table 3(a) to Table 3(c). From these tables, it is clear that overall fitting is good but estimates of parameters are not stable as indicated by high correlation coefficients. Moreover, the model adequacy parameter R^2 are also not good to use these models for forecasting purpose. The graphs of observed and predicted values as given Fig. 3 and Fig. 4 respectively indicate that the fitting models with only one explanatory variable is not good enough for forecasting.

Table 2(a): ANOVA-logistic model with *x*

Source	Degree of Freedom	Sum of squares	Mean Sum of Squares	F-value	p-value
Model	3	16598.50	5532.80	19.57	<.0001
Error	63	17814.80	282.80		
Total	66	34413.20			
R^2	46%				
MAE	0.30				

Table 2(b): Parameter estimates-Logistic model with *x*

Parameter	Estimate	Approx. Std. Error	Approx. 95% Confidence Interval	
α	49.29	57.22	-65.06	163.70
β	2655.40	10051.50	-17430.90	22741.70
γ	0.36	0.27	-0.18	0.91

Table 2(c): Correlation matrix-Logistic model with *x*

	α	β	γ
α	1.000	-0.792	-0.909
β	-0.792	1.000	0.972
γ	-0.909	0.972	1.000

Table 3(a): ANOVA-exponential model with *z*

Source	Degree of Freedom	Sum of squares	Mean Sum of Squares	F-value	p-value
Model	2	11718.80	5859.40	16.52	<.0001
Error	64	22694.40	354.60		
Total	66	34413.20			
R^2	49%				
MAE	0.36				

Table 3(b): Parameter estimates: Exponential model with *z*

Parameter	Estimate	Approx. Std. Error	Approx. 95% Confidence Interval	
λ	14.650	23.930	-33.154	62.459
μ	-0.001	0.021	-0.044	0.042

Table 3(c): Correlation matrix-Exponential model with *z*

	λ	μ
λ	1.000	-0.994
μ	-0.994	1.000

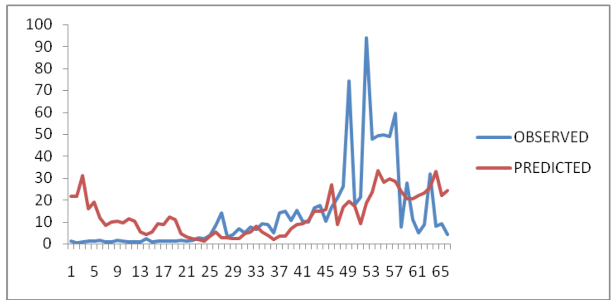

Fig. 3. Observed *vs*. Predicted aphid population by the logistic model with ***x***

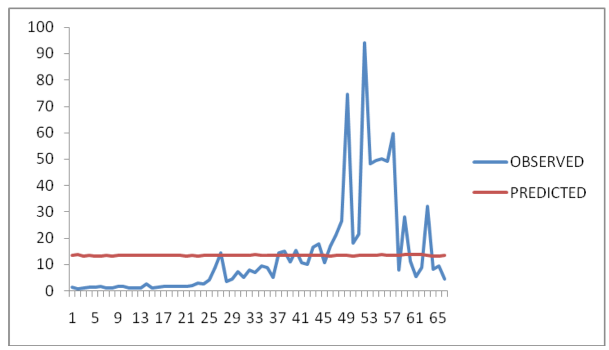

Fig. 4. Observed *vs*. Predicted aphid population by the Exponential model with *z*

We now apply the methodologies as proposed in this paper to this data.

Approach 1: In the first approach, the functional relationship between *y* and each of the explanatory variables are added linearly to get a final model. We have seen that the relation between *y* and *x* is logistic, where as it is exponential in case of *z*. These functional forms are now linearly added and obtained the following model

$$y = \frac{\alpha}{\left(1+\beta\, Exp\left(-\gamma x\right)\right)} + \lambda\ \mathrm{Exp}\left(-\mu z\right) + \varepsilon \qquad (21)$$

The model has now 5 parameters to be estimated. Again SAS programme is written to obtain the parameter estimates. Results are given in Table 4(a) to Table 4(c). The fitted model becomes

$$y = \frac{80.16}{\left(1 + 1032.3\, Exp\left(-0.29x\right)\right)} - 0.00037\, Exp\left(-0.19z\right) + \varepsilon \qquad (22)$$

The R^2 value is quite high and the model has very low MAE value which indicate that overall fitting of this model is very good. The graph between observed and predicted values is depicted in Fig. 5. The graph also shows the closeness between observed and predicted values.

Table 4(a): ANOVA: Approach-1

Source	Degree of Freedom	Sum of squares	Mean Sum of Squares	F-value	p-value
Model	5	17442.80	3488.60	12.54	<.0001
Error	61	16970.50	278.20		
Total	66	34413.20			
R^2	76%				
MAE	0.28				

Table 4(b): Parameter estimates: Approach-1

Parameter	Estimate	Approx. Std. Error	Approx. 95% Confidence Interval	
α	80.16	175.80	-271.30	431.60
β	1032.30	2438.70	-3844.20	5908.90
γ	0.29	0.24	-0.19	0.7762
λ	-0.00037	0.00622	-0.00001	0.000012
μ	0.19	0.18	-0.171	0.5550

Table 4(c): Correlation matrix: Approach-1

	α	β	γ	λ	μ
α	1.000	-0.682	-0.942	-0.356	-0.351
β	-0.682	1.000	0.884	0.521	0.518
γ	-0.942	0.884	1.000	0.439	0.436
λ	-0.356	0.521	0.439	1.000	0.999
μ	-0.351	0.518	0.436	0.999	1.000

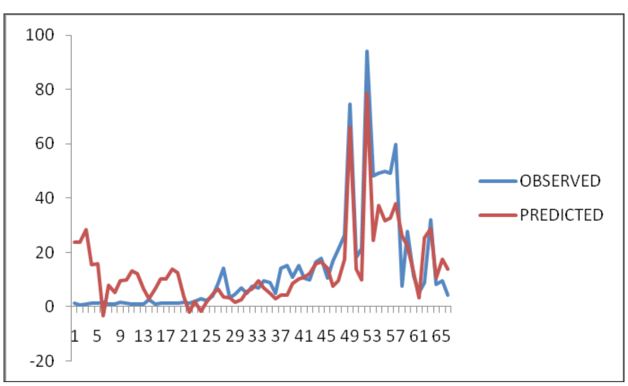

Fig. 5. Observed *vs*. predicted aphid population by a model by approach-1

Approach 2: In the second approach, in the non-linear models the respective variable is replaced by a linear function of all explanatory variables. Two models are developed. In the first model, *i.e.*, in logistic model, the explanatory variable x is replaced by a linear combination of x and z. The final model becomes

$$y = \frac{\alpha}{\left(1 + \beta\, Exp\left(-\gamma x - \lambda z\right)\right)} + \varepsilon \qquad (23)$$

Now this model has four parameters to be estimated. Again SAS programme is written to obtain the parameter estimates. The results are provided in Table 5(a) to Table 5(c). Though the R^2 value and MAE value are quiet good, yet the graph for observed and predicted values as given in Fig. 6 indicates that the model not good enough for forecasting purpose.

Table 5(a): ANOVA-Approach – 2: Model 1

Source	Degree of Freedom	Sum of squares	Mean Sum of Squares	F-value	p-value
Model	4	16602.5	4150.6	14.45	<.0001
Error	62	17810.7	287.3		
Total	66	34413.2			
R2	63%				
MAE	0.32				

Table 5(b): Parameter Estimates-Approach – 2: Model 1

Parameter	Estimate	Approx. Std. Error	Approx. 95% Confidence Interval	
α	63.8987	129.8	-195.6	323.4
β	1110.0	3598.8	-6083.8	8303.9
γ	0.3159	0.2694	-0.2226	0.8545
λ	-0.00418	0.0215	-0.0471	0.0388

Table 5(c): Correlation matrix-Approach – 2: Model 1

	α	β	γ	λ
α	1.000	-0.677	-0.934	-0.0732
β	-0.677	1.000	0.826	0.4642
γ	-0.934	0.826	1.000	0.0216
λ	-0.073	0.464	0.0216	1.0000

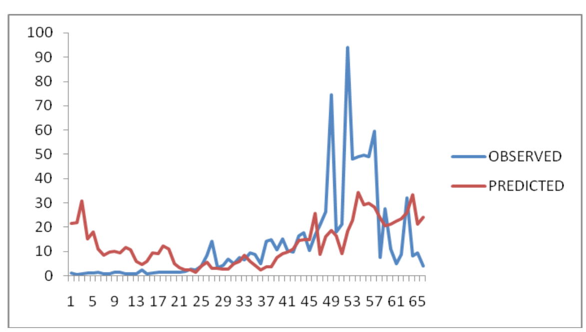

Fig. 6. Observed vs. Predicted Aphid Population by a Model by Approach–2: Model 1

In the second model the explanatory variable z is replaced by a linear combination of other explanatory variables. The model considered is as follows:

$$y = \alpha \; Exp\left(-\beta x - \gamma z\right) + \varepsilon \tag{24}$$

This model has three parameters. SAS is again used to estimate the parameters. Results are presented in Table 6(a) to Table 6(c). Fig. 7 depicts the curve between observed and predicted values.

Table 6(a): ANOVA-Approach – 2: Model 2

Source	Degree of Freedom	Sum of squares	Mean Sum of Squares	F-value	p-value
Model	3	16593.4	5531.1	19.55	<.0001
Error	63	17819.9	282.9		
Total	66	34413.2			
R^2	69%				
MAE	0.36				

Table 6(b): Parameter estimates-Approach – 2: Model 2

Parameter	Estimate	Approx. Std. Error	Approx. 95% Confidence Interval	
α	0.2984	0.4558	-0.6124	1.2093
β	-0.00657	0.0137	-0.0339	0.0208
γ	0.2248	0.0653	0.0943	0.3552

Table 6(c): Correlation matrix-Approach – 2: Model 2

	α	β	γ
α	1.0000	-0.463	-0.7506
β	-0.4630	1.000	-0.2320
γ	-0.7506	-0.232	1.0000

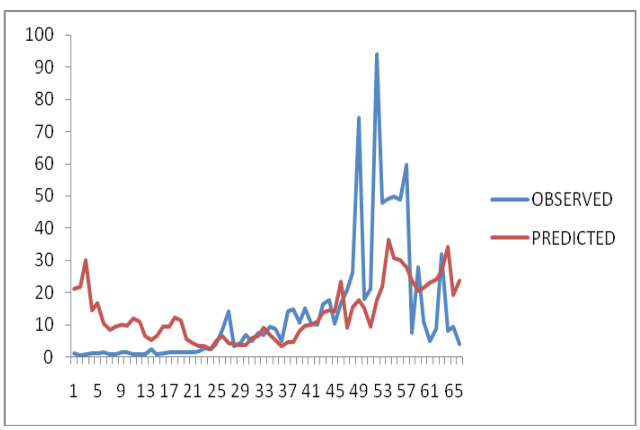

Fig. 7: Observed vs. Predicted Aphid Population by a Model by Approach–2: Model 2

Here also we see that though the R^2 value and MAE value are quiet good, yet the graph for observed and predicted values as given in Fig. 7 clearly indicates that the model not good enough for forecasting purpose. Thus it is expected that the model developed by Approach 1 can be a good approach for building a non-linear model.

4. DISCUSSION

The methodologies as proposed in the present paper are a noble idea to develop non-linear models with more than one explanatory variable. So far we used to develop non-linear models with time variable only which sometime do not clearly exploit the relation between response variable and time. Also in some situations, it is not at all possible to develop a non-linear model with time. Thus with this new approach, one can now be able to use other explanatory variables. Among the two approaches, the first approach gives better results. However, this is true for data set we have used in the example. In some other data sets, second approach may come out to be useful. Thus one has to exploit all possible kind of models to arrive at the final model. The approaches are empirical, but they have strong power to build up models according to situation. We have also forecast some observations beyond 28th Feb., 2003 up to which data are used to build up models. We have not presented these results in the paper. But among all the models presented in the paper, one developed by Approach 1, gives better forecast values as compared to the observed actual values. As mentioned earlier, fitting of models was carried out by SAS software. Programme codes can be obtained from the author on request.

REFERENCES

Agrawal, R. and Jain, R.C. (1996). Forecast of sugarcane yield using eye estimate along with plant characters. *Biom. J.*, **38**, 731-739.

Chakraborty, S., Ghosh, R., Ghosh, M., Fernandes, C.D. and Charchar, M.J. (2004). Weather- based prediction of anthracnose severity using artificial neural network models. *Plant Pathology*, **53**, 375-386.

Draper, N.R. and Smith, H. (1998): *Applied Regression Analysis*. 3rd ed. John Wiley.

Jain, R.C. and Agrawal, R. (1992). Probability model for crop yield forecasting. *Biom. J.*, **34**, 501-511.

Jain, R.C. and Ramasubramanian, V. (1998). Forecasting of crop yields using second order Markov chains. *Jour. Ind. Soc. Agril. Statist*., **51**, 61-72.

Kumar, Amrender and Bhar, L.M. (2005). Forecasting model for yield of Indian mustard (*Brassica junea*) using weather parameter. *Ind. j. Agril. Sci.*, **75**(**10**), 688-690.

Kvalseth, T. O. (1985). Cautionary note about R^2. *Amer. Statistician*, **39**, 279-285.

Matis, J.H., Saito, T., Grant, W.F., Iwig, W.C. and Ritchie, J.T. (1985). A Markov chain approach to crop yield forecasting. *Agril. Sys.*, **18**, 171-187.

Matis, J.H., Birkett, T. and Boudreaux, D. (1989). An application of the Markov chain Approach to forecasting cotton yield from surveys. *Agril. Sys.*, **29**, 357-370.

Mcleod, A.I. and Vingilis, E.R. (2005). Power computations for intervention analysis. *Technometrics*, **47**, 174-181.

Paul, R.K. (2015). ARIMAX-GARCH-WAVELET Model for forecasting volatile data. *Model Assist. Statist. Appl.*, **10(3)**, 243-252.

Paul, R.K., Gurung, B., Samanta, S. and Paul, A.K. (2015). Modeling long memory in volatility for spot price of lentil with multi-step ahead Out-of-sample forecast using AR-FIGARCH Model. *Eco. Affairs-Quart. J. Econ.*, **60(3),** 457-466.

Prasad, R. and Dudhane, S.N. (1989). Forecasting mustard yield in gangetic West Bengal using rainfall and agricultural technology. *Mausam*, **40(4)**, 441-446.

Ramasubramanian, V. and Bhar, L.M. (2014). Crop Yield Forecasting by Multiple Markov Chain Models and Simulation. *Statist. Appl.*, **12**, 1-13.

Ratkowsky, D.A. (1990). *Handbook of Nonlinear Regression Models*. Marcel Dekker. New York.

Suresh, K.K. and Priya, S.R.K. (2011). Forecasting sugarcane yield of Tamilnadu using ARIMA models. *Sugar Tech*., **13(1),** 23-26.