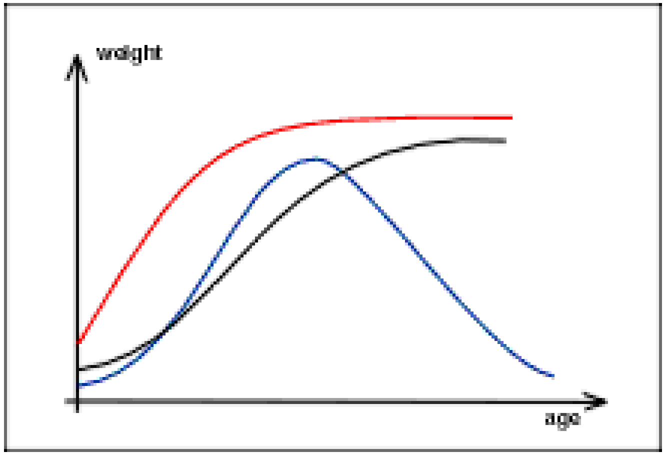
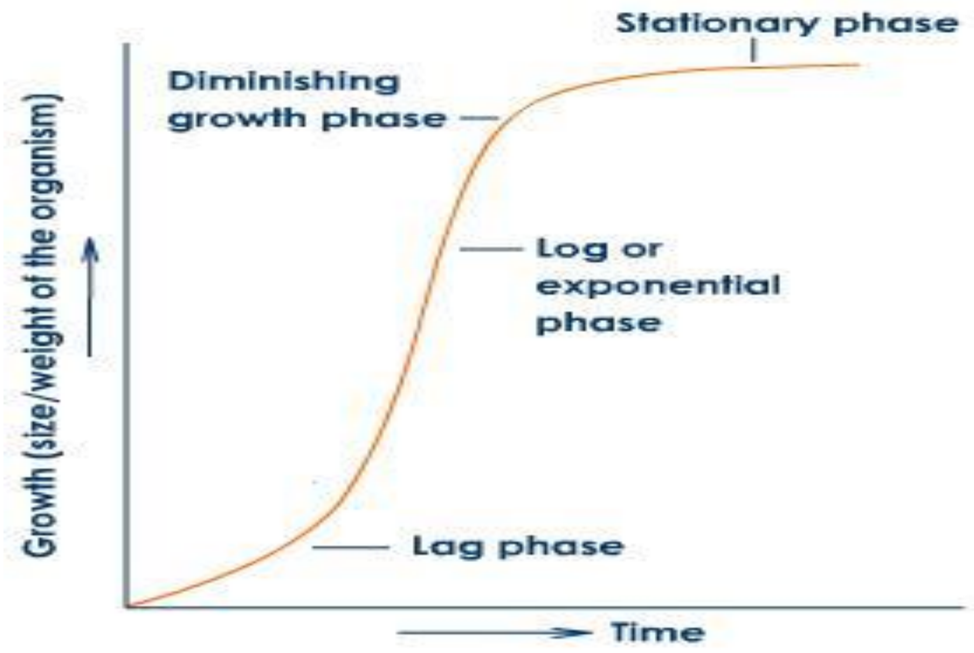
Nonlinear Statistical models for Pest Populations- A case study of Aphid



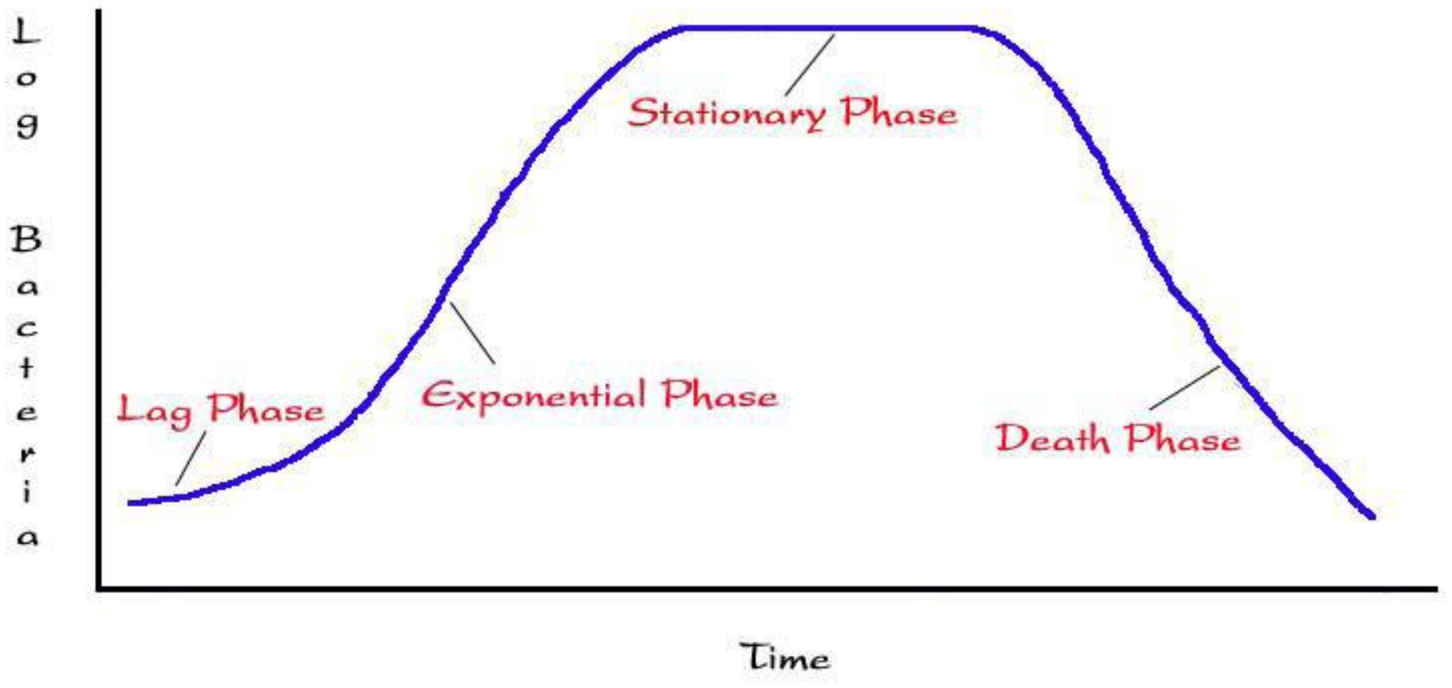
Ch. Sarada
Principal Scientist(Ag. Statistics)
Directorate of Oilseeds Research
Hyderabad
saradac@yhoo.com

ICAR

Plant / Animal Growth Curve

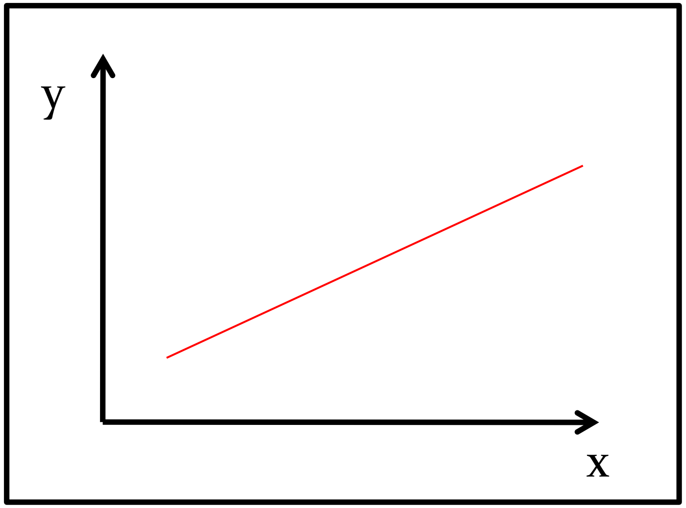


Bacterial Growth Curves

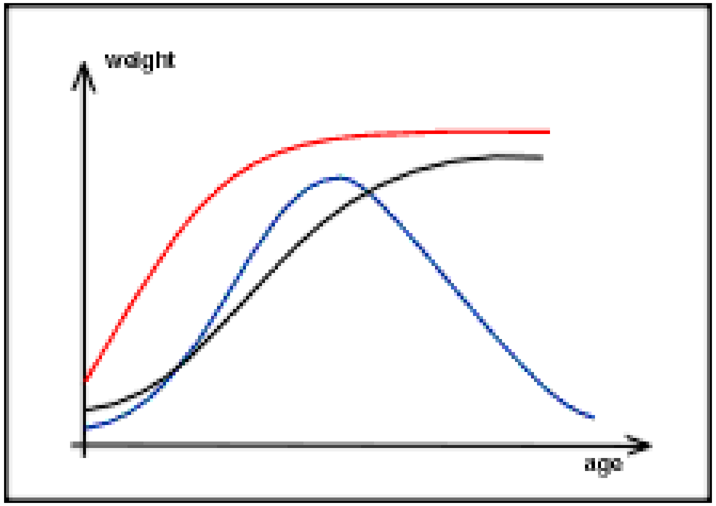


Regression Models

Linear model



Nonlinear model



$$Y = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon$$

$$Y = f(X_1, X_2, \ldots, X_p \mid \theta_1, \theta_2, \ldots, \theta_q) + \varepsilon$$

भाकृअनुप
ICAR

Nonlinear Model

- A non-linear regression model is one in which at least one of the parameters appear non-linearly.
- The derivatives of the model with respect to the model parameters depends on one or more parameters.
- **Linear Model** $y = b_o + b_1x + b_2x^2 + e$
- $dy/db_o = 1$ $dy/db_1 = x$ $dy/db_2 = x^2$

- **Nonlinear Model** $Y = a X^{b} + e$
- $dy/da = X^{b}$

  $dy/db = a X^{b} (\log (X))$

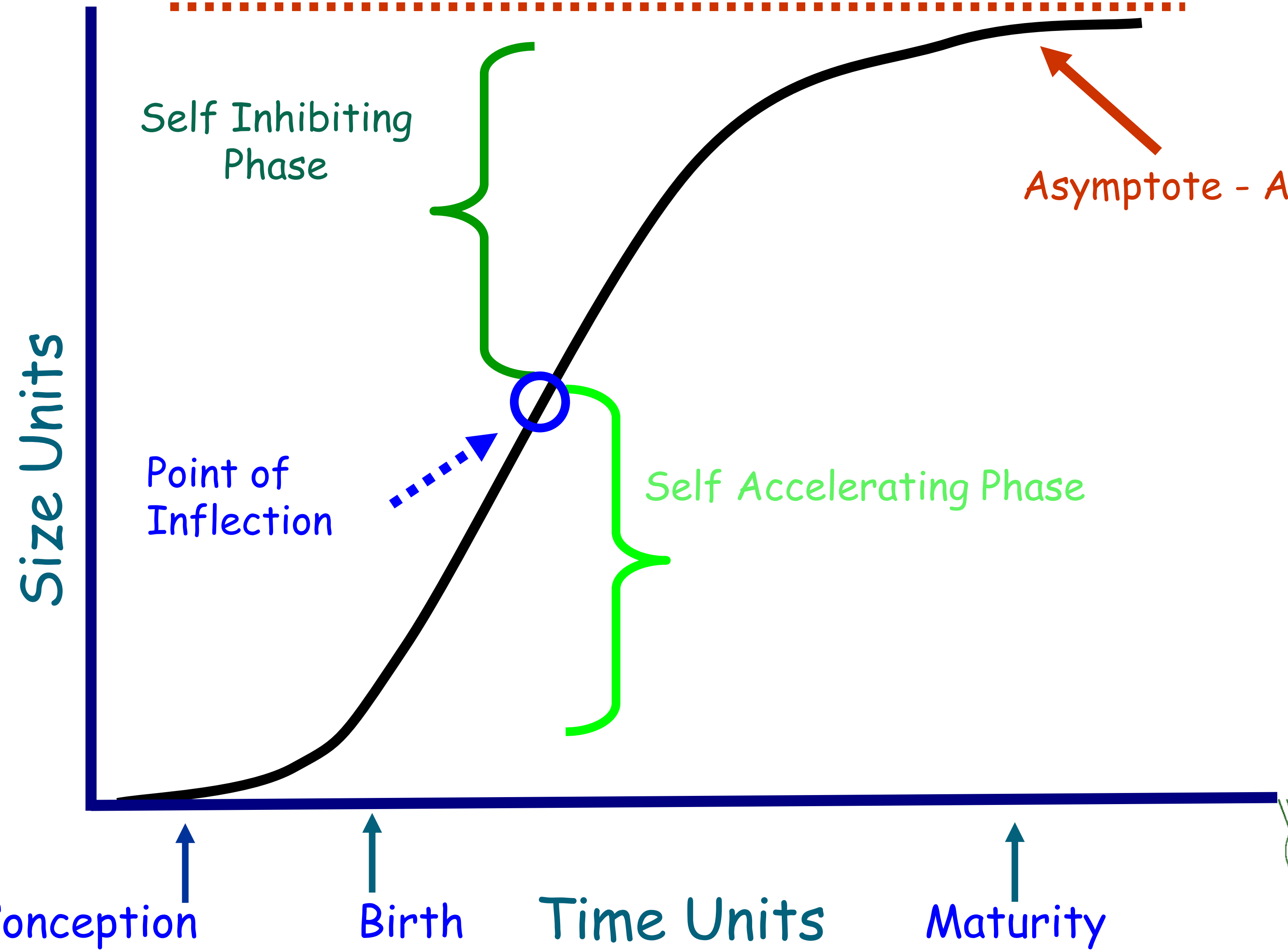
Properties

- Nonlinear models are often derived on the basis of physical and/or biological considerations,.
- The parameters of a nonlinear model usually have direct interpretation in terms of the process under study.
- Constraints can be built into a nonlinear model easily and are harder to enforce for linear models.

Model Development

- Development
  - Statistical model
  - Assumptions
  - Type of Variables considered
- Validation
  - Goodness of fit
  - Model diagnostics
  - Modification of the Model and revalidation
- Implementation
  - Experimental trials
  - Advisory systems / DSS

Phases of the Growth Curve



Some Important Nonlinear Growth Models

Logistic	$Y_t = A/(1 + Be^{-kt}) + \varepsilon$
Gompertz	$Y_t = A\exp(-Be^{-kt}) + \varepsilon$
Von Bertlanffy	$Y_t = A(1 - Be^{-kt}) + \varepsilon$
Mercer –Foldin	$Y_t = (AB + kt^D/(B + t^D) + \varepsilon$
Wiebull	$Y_t = A - B\exp(-kt^D) + \varepsilon$
Richards	$Y_t = A/(1 + exp(b - kt)^{\frac{1}{D}} + \varepsilon$

Y_t = Yield /growth /length/weight /biomass /Population etc at time period t

A = Asymptote or mature weight at age t approaches infinity/carrying capacity

B = Initial yield/growth/length/weight/population

k = Growth rate

D = Shape parameter

Fitting of Nonlinear Models

- Parameter Estimation
- Choice of Initial Values
- Goodness of Fit of a Model
- Examination of Residuals

Parameter Estimation

- Nonlinear Least Squares
- Normal equations are nonlinear in parameters
- Iterative procedures

  (requires supplying of initial values)

  - Gauss –Newton Method
  - Newton Method
  - Levenberg –Marquardt's Method
  - Steepest Descent (Gradient Method)

Choice of Initial Values

- Making Initial Guess
- Estimates calculated from previous experiments
- Known values for similar systems
- Values computed from theoretical considerations

- Linearization
- Solving system of equations
- Using properties of the model
- Graphical method

Goodness of Fit statistics

Coefficient of Determination

$$R^2 = 1 - \sum (Y_i - \hat{Y}_i)/(Y_i - \bar{Y})^2$$

Mean Square Error (MSE) = $\sum (Y_i - \hat{Y}_i)^2 / (n - p)$

Mean Absolute Error (MAE) = $\sum |Y_i - \hat{Y}_i|^2 / (n)$

भाकृअनुप
ICAR

Examination of Residuals

- Assumptions
  - Errors are independent
    - Plotting of residuals
    - Tests for independence of errors (Runs Test)
  - Errors are normally distributed
    - Plotting Norma probability plot
    - Tests for Normality (Shapiro-Wilk test)

Wald –Wolfowitz Runs Test

- H_0: Errors are independent
- H_1: Errors are not independent

Mean (μ) = $2mn/(m+n)+1$

Variance = $2mn(2mn-m-n)(m+n)^{-2}(m+n-1)^{-1}$

$$Z = r+h-\mu/\sigma$$

Where,

$$h = \begin{cases} 0.5, if\ r < \mu \\ -0.5, if\ r > \mu \end{cases}$$

Test for Normality (Shapiro-Wilk test)

- H_0 : Errors are normally distributed
- H_1: Errors are not normally distributed.

$$S^2 = \sum a(k)\{X_{(n+1-k)} - X_{(k)}\}, b = \sum (X_i - \bar{X})^2$$

Where,

$$K = \begin{cases} 1,2,\ldots,n/2 \text{ when } n \text{ is even} \\ 1,2,..,(n-1)\ /2 \text{ when } n \text{ is odd} \end{cases}$$

SOFTWARE

- SAS – PROC NLIN
- SPSS – NLR
- S-PLUS / R - NLS
- SAS – Enterprise guide
- SPSS –Nonlinear Regression
- JMP

Model Development

- Development
  - Statistical model
  - Assumptions
  - Type of Variables considered
- Validation
  - Goodness of fit
  - Model diagnostics
  - Modification of the Model and revalidation
- Implementation
  - Experimental trials
  - Advisory systems / DSS

Model Diagnostics

$$N(t) = e^{a+bt}\ (1+e^{d+bt})^{-2} + \in$$

- Measure of Nonlinearity of parameters
  - Intrinsic Nonlinearity (IN) - Low value indicates unbiased prediction
  - Parameter –effects nonlinearity (PE)
    - Low value indicates good Confidence region /Symmetric confidence intervals
  - Marginal Curvatures -
    - Properties of each individual parameter
    - Profile t – plots - Close – to – linear
    - Skewness, Kurtosis , Bias
  - Reparameterization of the parameters
    - *Simulation - 1000 data sets - Parameter estimates*
    - Histograms – Long right hand tail – exponential of parameter

      Long left hand tail - logarithm of the parameter

Development of Aphid Population density Model

$$N(t) = ae^{bt}(1+de^{bt})^{-2} + \in$$

Prajneshu, (1998)

Model diagnostics ↓ Simulation studies (1000 data sets)

$$N(t) = e^{a+bt}(1+e^{d+bt})^{-2} + \in$$

Sarada.C and Prajneshu (2005)

Model Properties ↓ Ross.G, Prajneshu and Sarada.C (2010)

Reparametrized Aphid growth model

↓ C.Sarada *et al.* (2012)

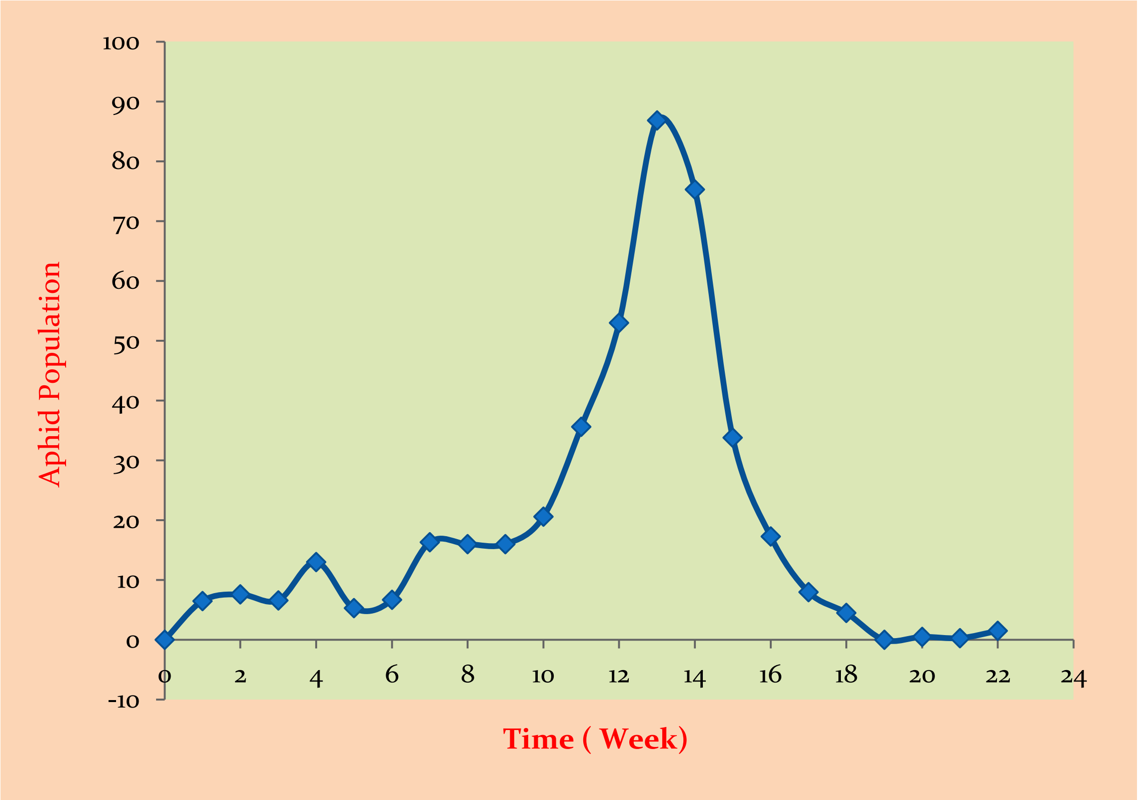
Safflower Aphid population

Development of Aphid Population density Model

Time (weeks)	Aphid Population per 100 leaves
0	0.0
1	6.5
2	7.6
3	6.6
4	13.0
5	5.3
6	6.7
7	16.3
8	16.0
9	16.0
10	20.6
11	35.6
12	53.0
13	86.8
14	75.3
15	33.8
16	17.3
17	8.0
18	4.5
19	0.0
20	0.5
21	0.3
22	1.5

$$N(t) = ae^{bt}(1+de^{bt})^{-2} \quad + \in$$

Prajneshu, (1998)



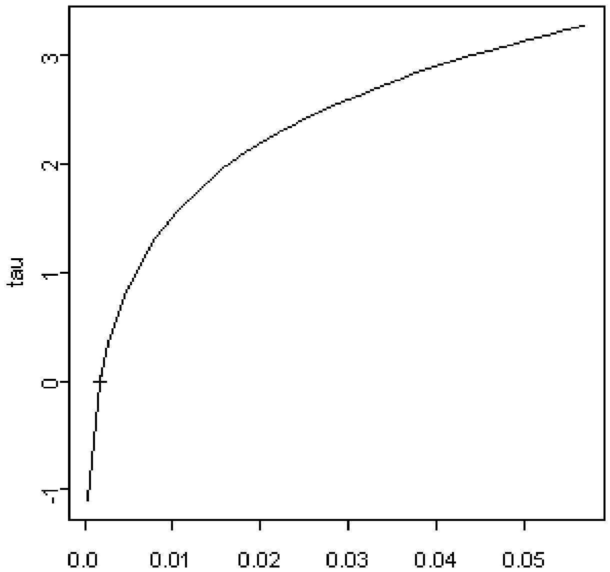
Verma, K.D. and Parihar, S.B.S. (1991). Build up of the vector *Aphis gossypii* glover on potato. *J. Aphidol.*, **5**, 16 - 18.

Model Diagnostics

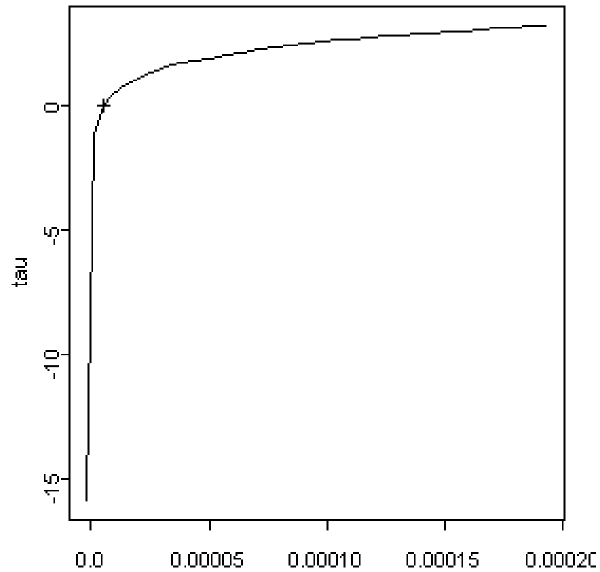
Statistic	Coefficeint	
i. Parameters		
a	0.0018 (0.0017)	
b	0.92 (0.08)	
d	$0.0^{5}55$ ($0.0^{5}59$)	
ii Confidence Intervals	Lower	Upper
a	0.001	0.007
b	0.8	1.07
d	0.000	0.0002
iii Measures of Nonlinearity		
IN	0.21	
PE	12.23	
iii. Marginal Curvatures		
a	0.78	
b	0.07	
d	0.82	

Profile t-plots

Parameter - a

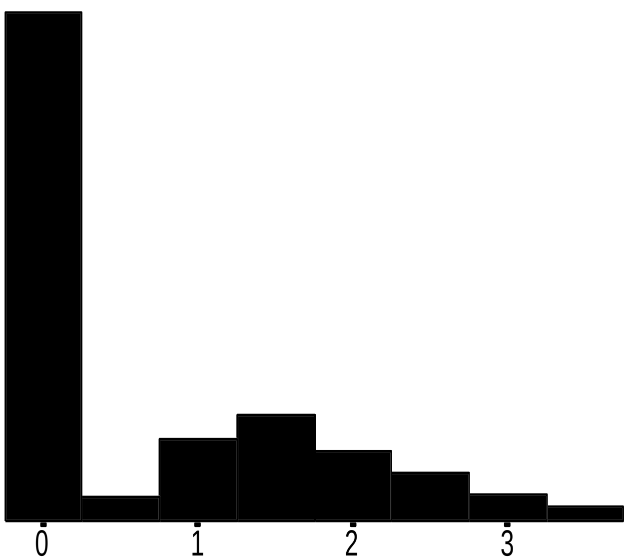


Parameter - d

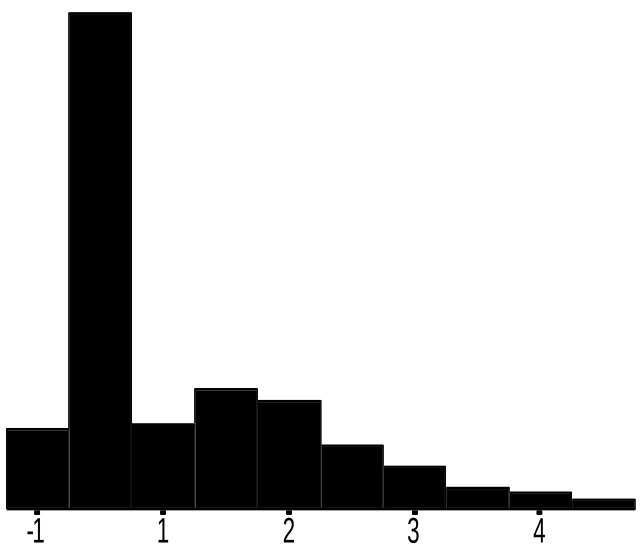


Histograms of 1000 simulated standardized parameter estimates

Parameter - a



Parameter - d



Reparameterization of Aphid Model

Model - I

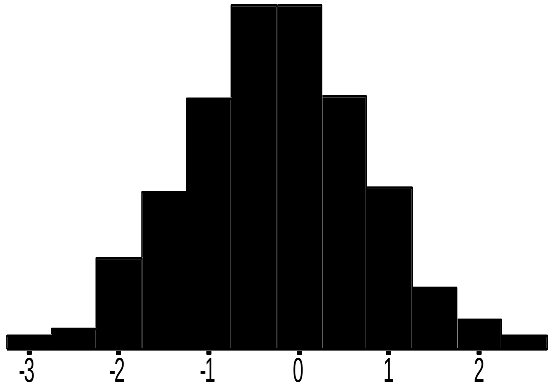
$$N(t) = ae^{bt}(1+de^{bt})^{-2} + \in$$

Model - II

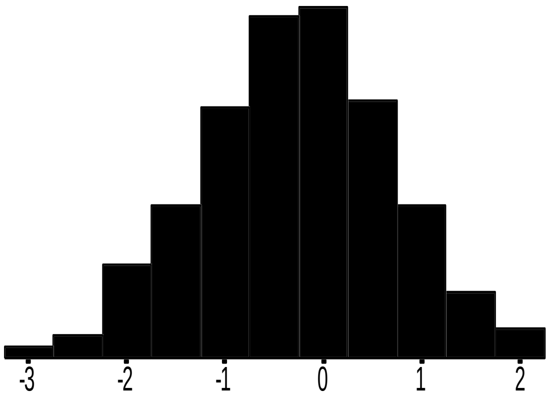
$$N(t) = e^{a+bt}(1+e^{d+bt})^{-2} + \in$$

Histograms of 1000 simulated standardized parameter estimates for Model -II

Parameter - a

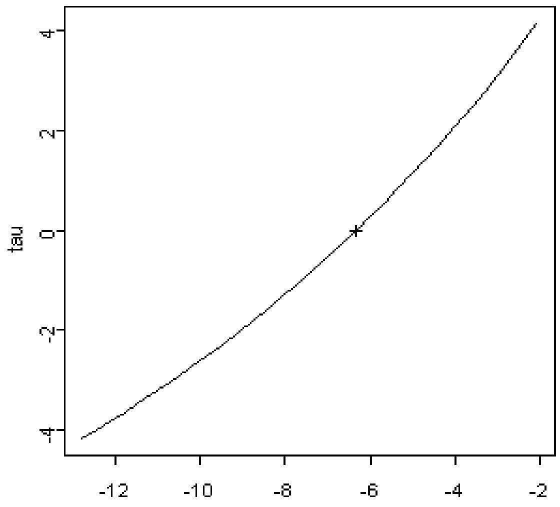


Parameter - d

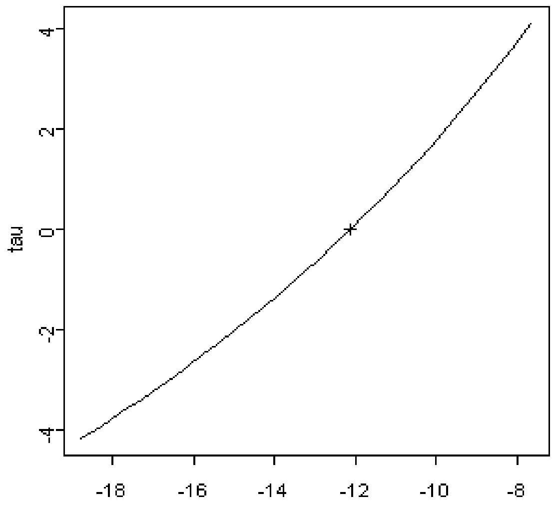


Profile t-plots Model -II

Parameter - a



Parameter - d



Parameter estimation, rms curvatures measures, marginal curvature, measures of nonlinearity and simulation studies

	Model I		Model II	
(i) Parameter estimation:				
a	0.0018 (0.0017)		-6.34 (0.99)	
b	0.92 (0.08)		0.92 (0.08)	
d	$0.0^{5}55$ ($0.0^{5}59$)		-12.13 (1.67)	
(ii) 95% Confidence-intervals:				
	Lower	Upper	Lower	Upper
a	0.001	0.0070	-8.17	-4.76
b	0.800	1.0700	0.80	1.07
d	0.000	0.0002	-14.03	-10.04
(iii) rms Curvature effects:				
(IN) $\sqrt{F_{3,20}(.05)}$	0.21		0.21	
(PE) $\sqrt{F_{3,20}(.05)}$	12.23		0.24	
(iv) (Marginal curvatures) $t_{20}(.05)$:				
a	0.78		0.08	
b	0.07		0.07	
d	0.82		0.07	
(v) Simulation Studies:				
Skewness:				
a	0.61**		-0.18^{NS}	
d	0.61**		-0.10^{NS}	
Kurtosis:				
a	-1.06**		0.29^{NS}	
d	1.05**		0.27^{NS}	
% Bias:				
a	-23.30**		0.02^{NS}	
d	-24.02**		0.01^{NS}	

Development of Safflower Aphid Population Model

$$N(t) = e^{a+bt} \left(1 + e^{d+bt}\right)^{-2} + \in$$

$$C = a(2b^2 d)^{-1}$$

$$N_o = a(1 + d)^{-2}$$

$$r = (b^2 - 2N_o C^{-1})^{1/2}$$

N_0 = Initial population

C = carrying capacity

N(t) = population at time t

r = growth rate

Data

- Safflower variety : CO-1
- Years : 2005-10
- Region : Solapur
- Frequency : Weekly Aphid Population
- Source : Safflower Annual Reports, DOR, Hyderabad

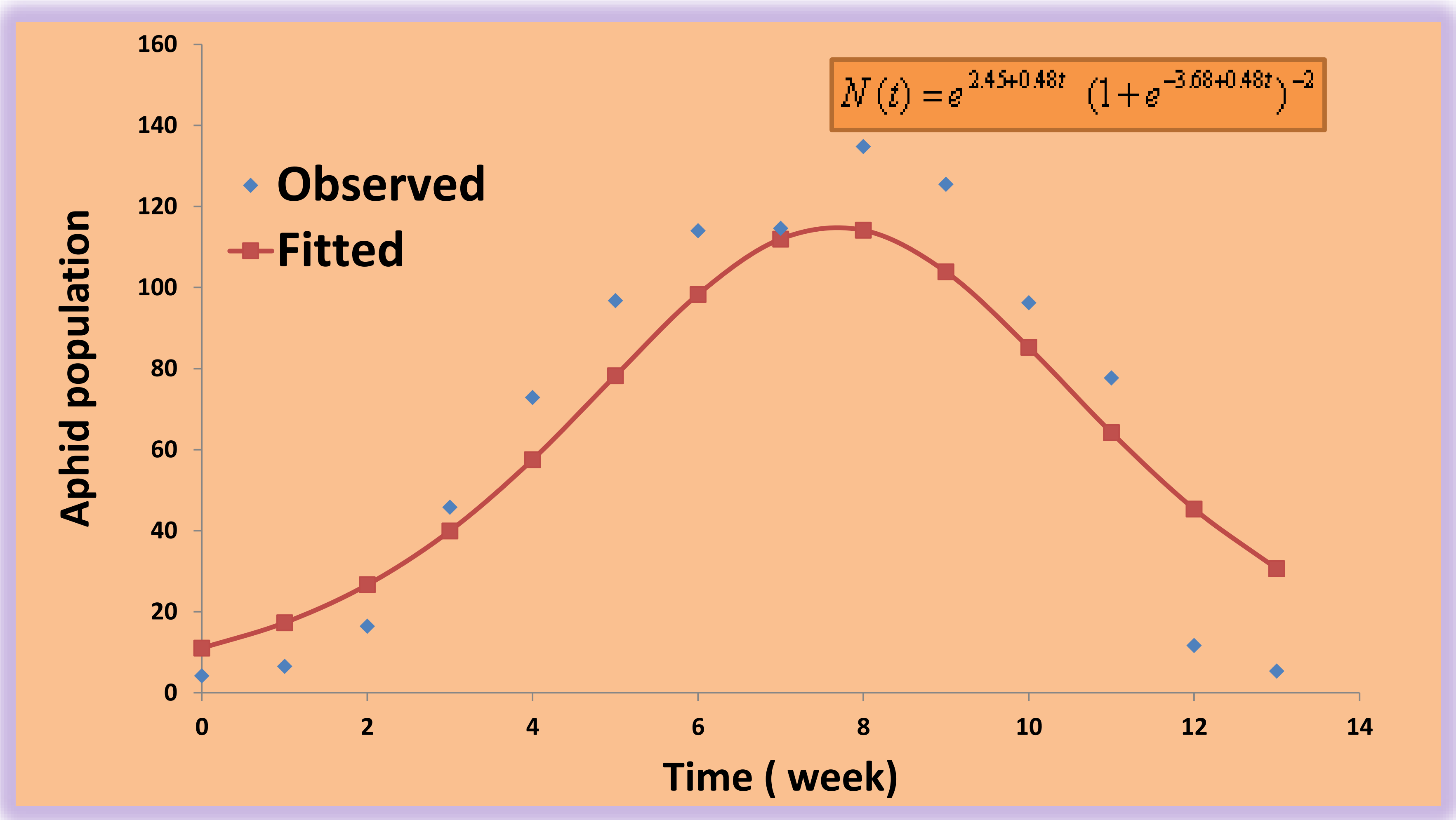
Fitting of Nonlinear Model

$$N(t) = e^{a+bt} \ (1 + e^{d+bt})^{-2} + \in$$

Estimate the parameters	:	a, b, d
Iterative procedure	:	Levenberg-Marquardt
Goodness of fit statistic	:	Root mean square error
Examination of Residuals	:	Run test Shapiro-Wilk test

Software : SAS v 9.2 / Proc : NLIN

Fitted Nonlinear Aphid population model



Results

Statistic	Coefficeint
i. Parameters	
a	2.45(0.35)
b	0.48(0.05)
d	-3.68(0.40)
ii. Examination of Residuals	
Run test | Z|	1.2
Shapiro –Wilk test (W)	0.96
iii. Goodness of fit statistic	
Root mean square error	54.52

Summary

- **Polynomial functions are not nonlinear functions**
- Nonlinear function is one in which the parameters appear nonlinearly
- Biological Interpretation of the data can be understand by Nonlinear models
- Choosing the suitable model and initial guess values are important for obtaining the better/ suitable model for the data under consideration.
- Software : SAS , JMP , R , S -Plus

Thank you