ISSN: 2347-5129
IJFAS 2014; 1(3): 171-177

www.fisheriesjournal.com
Received: 22-10-2013
Accepted: 06-11-2013

N. Okendro Singh
Department of Mathematics & Statistics, College of Agriculture, CAU, Iroisemba – 795004, Imphal, Manipur, India
Email: nosingh2000@yahoo.com

N. Gopimohan Singh
Department of Mathematics & Statistics, College of Agriculture, CAU, Iroisemba – 795004, Imphal, Manipur, India
Email: gopi_naorem@yahoo.com

Surinder Kumar
Department of Applied Statistics, Babasaheb Bhimrao Ambedkar University, Lucknow
Email: surinderntls@gmail.com

Prem Kumar
Directorate of Coldwater Fisheries Research, Bhimtal – 263 136, Uttarakhand
Email: prem_nrc@rediffmail.com

Correspondence:
N. Okendro Singh
Department of Mathematics & Statistics, College of Agriculture, CAU, Iroisemba – 795004, Imphal, Manipur, India
Email: nosingh2000@yahoo.com
Tel: 8415991288

Pre-harvest forecasting of common carp yield from polyponds

N. Okendro Singh, N. Gopimohan Singh, Surinder Kumar, Prem Kumar

ABSTRACT
An explicit form of the partially reparameterized version of Gompertz model was used to fit the average growth data of common carp rearing in polyponds of upland region and its suitability for pre-harvest forecasting of fish yield was demonstrated. Two months ahead forecasting of common carp yield was best provided by the above model.

Keywords: Forecasting, reparameterization, nonlinear behavior, high correlation

1. Introduction

In the uplands of India, there is generally a lack of market facility in the nearby locality and a poor infrastructure is available for transportation of the fish produce from the farmers' pond. Reliable and timely forecast of fish yield from ponds of upland in particular will be of immense help for the farmers to plan for marketing their produce profitably. Till today, this important task has not been seriously taken into consideration in the uplands of India. Nonlinear models for forecasting fish yield from cemented ponds based on classical approach were developed [11]. Nonlinear models are appropriate for developing forecasting models since fish weight data are usually collected over time. However, classical approach of nonlinear modeling may not be appropriate. The above work was further extended in the natural ponds considering heteroscedastic error structure [3]. But, the nonlinear models fitted to the data usually resulted in highly nonlinear behavior or highly correlated among the estimated parameters. The importance of reparameterization of the parameters for fitting of Schaefer model in fisheries to reduce large correlations among the parameter estimates was emphasized [6]. Further, many authors highlighted the importance of reparameterization in nonlinear model fitting [5, 9, 10]. However, no one has highlighted yet the importance of close-to-linear behavior of parameters in nonlinear regression models and the high correlation among the estimated parameters while they developed such type of forecasting models although we come across such situations in most of the cases. In fact, a little attention is given to the various reparameterizations and consequently, the parameter estimates hardly satisfy any of the optimum properties. In view of the above, the present study aims to identify the most appropriate forecasting model for fish yield from polyponds of upland region. The identified methodology is illustrated with an example considering the growth data of common carp rearing in polyponds (poly film lined) of upland region during the period March 2009 to February 2010.

2. Methodology

The following nonlinear models will provide a reasonable representation of average fish weight, W_t at time t:

Logistic-I:

$$W_t = \frac{\beta_1}{1+\beta_2 \exp(-\beta_3 t)} \quad \ldots\ldots(1)$$

Gompertz-I:

$$W_t = \beta_1 \exp[-\beta_2 \exp(-\beta_3 t)] \quad \ldots\ldots(2)$$

Von-Bertalanffy-I:

$$W_t = \frac{\beta_1}{[1-\beta_2 \exp(-\beta_3 t)]^3} \quad \ldots\ldots\ldots\ldots(3)$$

Von-Bertalanffy model-II:

$$W_t = [\beta_1 - (\beta_1 - \beta_3)\exp(-\beta_2 t)] \quad \ldots\ldots(4)$$

Von-Bertalanffy model-III:

$$W_t = \beta_1[1 - \exp\{-\beta_3(t-\beta_2)\}] \quad \ldots\ldots(5)$$

Where β_1, β_2 and β_3 are the parameters to be estimated. The parameter β_1 represents the limiting growth value or asymptotic size, β_2 the scaling parameter and β_3, the rate of maturity.

If 'β_1' is likely to be an offensive parameter say, in equation (2), then it can be partially reparameterized by expected-value parameter. To obtain an expected-value parameter from above equation (2), we need to choose value t_1 of the regressor variable t, within the observed range of t. Then, we get the expected value from equation (2) as follows:

$$W_1 = \beta_1 \exp[-\beta_2 \exp(-\beta_3 t)]$$

Solving this equation for the parameter 'β_1' only, we get

$$\beta_1 = \frac{W_1}{\exp[-\beta_2 \exp(-\beta_3 t_1)]}$$

Substituting back into the original equation (2), we get

$$W_t = W_1 \frac{\exp[-\beta_2 \exp(-\beta_3 t)]}{\exp[-\beta_2 \exp(-\beta_3 t_1)]} \quad \ldots\ldots(6)$$

The above model is expected to eliminate both the nonlinear behaviour of parameters and high correlation among the estimated parameters. Here, the likely offensive parameter 'β_1' is reparameterized by expected-value parameter while the other parameters are not changed. The form of the partial reparameterization of the Gompertz model given by equation (6) is referred to as 'Gompertz-II' in the subsequent discussions.

2.1 Criteria for Model Selection

To examine model performance, summary statistics like root mean square error (RMSE) and mean absolute error (MAE) are generally used:

$$\text{RMSE} = \left[\sum_{t=1}^{n}\left(W_t - \hat{W}_t\right)^2 / n\right]^{1/2};$$

$$\text{MAE} = \sum_{t=1}^{n}\left|\left(W_t - \hat{W}_t\right)\right| / n, \text{ and}$$

Where,

$\hat{W}_t$ Predicted fish weight of t[th] observation;

$\overline{W}$ Average fish weight;

n Number of observations, $t = 1,2,\ldots,n$.

The better model will have the least values of these statistics. It is, further, recommended for residual analysis to check the model assumptions such as independence or the randomness assumption of the residuals and the normality assumption. To test the independence assumption of residuals run test procedure is available in the literature [8]. Further, Shapiro-Wilk's test was applied to check the normality assumption. For any type of growth data collected over the time period, it is reasonable to check serial correlation among measurements within the same fish species. The presence of autocorrelation can also be checked out by using Durbin-Watson test statistic and Breusch-Godfrey test. Moreover, White's test and Breusch-Pagan test are commonly applied for testing of heteroscedastic.

Moreover, the curvature in a nonlinear model consists of two components: the intrinsic (IN) curvature and parameter effects (PE) curvature. Details of the root mean square (RMS) IN and PE measures of curvature and curvature critical value are given [1, 2]. The IN curvature is typically smaller than the PE curvature, which can be affected by altering the parameterization of the model [7]. Severe curvature effects are indicated by values of IN and PE exceeding the critical value i.e. $1/\sqrt{F_{p,(n-p)}(0.05)}$, p is the number of parameters involved in the model. In usual, PE is computed when IN is within permissible limits and a lower value of PE suggests that the model exhibits close-to-linear behavior [8].

Hougaard's measure of skewness, g_t, can be employed to assess whether a parameter is close to linear or whether it contains considerable nonlinearity. Hougaard's measure is computed as follows:

$$E\left[\hat{\beta}_t - E\left(\hat{\beta}_t\right)\right]^3 = -(\text{MSE})^2 \sum_{jkl}^{p} L_{jk} L_{kl} L_{jl}\left(W_{jkl} + W_{kjl} + W_{lkj}\right)$$

Where, the sum is a triple sum over the number of parameters,

$$\mathbf{L} = \left[\mathbf{J}'\mathbf{J}\right]^{-1},$$

$$W_{jkl} = \sum_{m=1}^{n} J_{mj} H_{mkl},$$

J is the Jacobian matrix, $\mathbf{J}_m$ is the Jacobian vector, **H** is the Hessian matrix, $\mathbf{H}_m$ is its component evaluated at observation m and β_t is the t[th] parameter. This third moment is normalized using the standard error to give Hougaard's measure of skewness as

$$g_t = E\left[\hat{\beta}_t - E\left(\hat{\beta}_t\right)\right]^3 \Big/ \left(MSE * L_{tt}\right)^{3/2}.$$

If $|g_t| < 0.1$, the estimator $\hat{\beta}_t$ of parameter β_t is very close-to-linear in behavior and, if $0.1 < |g_t| < 0.25$, the estimator is reasonably close-to-linear. If $1 > |g_t| > 0.25$, the skewness is very apparent. For $|g_t| > 1$, the nonlinear behavior is considerable [8].

Moreover, the bias of Box reveals which parameters are responsible for the nonlinear behaviour. The bias of Box is calculated in multivariate form as given below [4]:

$$\text{Bias} = \left(D_2' D_2\right)^{-1}\left(D_2' H_2\right),$$

Where, D_2 is the nxp first derivative matrix;

$H_2 = -1/2\,\sigma^2 \text{trac}\left\{\left(D_2' D_2\right)^{-1} F_{2i}\right\}$, the expected difference between the quadratic and linear components of the Taylor approximation and F_{2t}, t=1,2,…,p are pxp faces of the second derivative matrix.

$$\%\text{Bias} = \frac{\text{Bias}}{\hat{\beta}_t} \times 100$$

Here, $\hat{\beta}_t$ is the estimated value of β_t. An absolute value which is greater than 1% is considered as an indicator of nonlinear behaviour [7].

2.2 Regression Diagnostics

I. **Studentized Residuals:** The studentized residual for the i[th] observation is the square root of an estimate of the variance of the i[th] residual. If the leverage and hence the studentized residual values is large, we consider the case as an outlier. Generally, the studentized residual values ranging between -2 and +2 are acceptable or not extreme and beyond these, they may be considered as outliers or extreme values.

II. **Influential Point or, Cook's D**: Cook's distance measures the influence of each sample observation on the coefficient estimates. Observations that are far from the average of all the independent variable values or that have large residuals tend to have a large Cook's distance value (say, greater than 2). Values >1 require careful checking; those >4 are potentially serious outliers.

III. **Q-Q Plot of Residuals**: Q-Q plot or Quantile plot is made by plotting the residuals against their percentage (empirical cumulative probability) point (0 to 1); if probability holds, this should have an S-shape. More formally, the f quantile is the data value below which approximately a decimal fraction f of the data is found. That data value is denoted by q(f). Each data point can be assigned an f-value. Let a time series t of length n be sorted from smallest to largest values, such that the sorted values have rank i=1, 2,…, n. The f-value for each observation is computed as

$$f_i = \frac{i - 0.5}{n}$$

This equation gives the quantile for any observation.

IV. **Histogram with Superimposed Normal PDF**: The histogram is a bar chart of frequency or number of observations in various data classes. A bell-shaped histogram is indicative of normally distributed data. Visual interpretation of the histogram is aided by over-plotting a properly scaled theoretical probability distribution for normal distribution with the same mean and variance as the sample series.

V. **Autocorrelation Plot of Residuals**: Autocorrelation function (ACF), partial ACF, inverse ACF plots provide information about the presence of autocorrelation in the series. Significant spikes in ACF, PACF and IACF plots at specified times indicate the presence of autocorrelation at the times.

3. Illustrations

As an illustration, the average growth data (weight in gm) of common carp was considered and the dataset is presented in Table 1.

Table 1: Growth data of common carp

Month	Weight (in gm)
Mar-09	3.3648
Apr-09	21.8238
May-09	38.6313
Jun-09	63.5960
Jul-09	87.7013
Aug-09	111.7120
Sep-09	131.6033
Oct-09	144.7017
Nov-09	163.2600
Dec-09	172.1000
Jan-10	179.7627
Feb-10	196.4227

The dataset was generated from the NAIP (National Agricultural Innovative Project) Component-III entitled "Enhancement of Livelihood Security through Sustainable Farming Systems and Related Farm Enterprises in North-West Himalaya" conducted at Directorate of Coldwater Fisheries Research, Bhimtal, Uttarakhand. The area or the region under study can be classified as having high potential of water in rainy season but there is marked water shortage during the lean period. In the recent past, emerging technologies like polyponds have been found very suitable for rainwater harvesting and continuous water supply to seasonal agricultural crops in hilly terrains. Polyponds are plastic film lined ponds and they are usually made by spreading a layer of Low Density Polyethylene (LDPE) of 200 gsm (gm/m^2) to the cemented tanks or earthen ponds. Polyponds were created for conducting experiments on integrated fish culture under NAIP by selecting three clusters of villages in Champawat district of Kumaun region, Uttarakhand. The average size and average volume of water for polyponds were 60 m^2 and 100 m^3 respectively. A polypond was selected from each cluster for this experiment. The three different exotic carps were reared in polyculture systems and the species composition was in the pattern of silver carp: 30%, grass carp: 30% and common carp: 40%. The stocking density of the fish was 3 fingerlings per cubic meter of water and thus on an average 300 fish in aggregate (120 specimens of common carp) were reared in each polypond which was replicated in three polyponds. Further, fish mortality rate was reported as much as high ranging from 20% to 30% per pond upto the end of the 10^{th} month during the rearing period. The growth data of a sample of size 30 per polypond comprises of 10 specimens from each fish group was randomly selected and data in terms of length and weight of fish was regularly observed for every month during March 2009 to February 2010. The average weight of common carp obtained from 30 (10 individuals per pond) observations for each month was utilized for present study and thus there are 12 average data points. The first ten data points were used for developing the model and the rest two points were kept for model validation purposes. The SAS 9.3 version was extensively used for various analyses.

4. Results and Discussion

The above nonlinear models were fitted to the available growth data of common carp. Von-Bertalanffy model-I failed to give optimum solution however; the summary statistics for fitting of other models are presented in Table 2.

Table 2: Summary statistics for models fitting

	Logistic-I	Gompertz-I	Gompertz-II	Von-Bertalanffy-II	Von-Bertalanffy-III
A) Parameter Estimation					
β_1 (or, W_1)	176.60 (6.20)	198.00 (5.35)	161.40 (1.33)	484.40 (161.20)	484.40 (161.20)
β_2	12.43 (1.96)	3.25 (0.13)	3.25 (0.13)	0.05 (0.02)	0.06 (0.15)
β_3	0.61 (0.05)	0.35 (0.02)	0.35 (0.02)	-1.53 (3.87)	0.05 (0.02)
B) Hougaard's Skewness & Box's % Bias					
β_1 (or, W_1)	0.39 & 0.26	0.30 & 0.13	$0.0^{2}2$ & 0.00	2.46 & 13.70	2.46 & 13.70
β_2	0.76 & 2.45	0.30 & 0.24	0.30 & 0.24	0.02 & 0.11	-0.39 & -14.80
β_3	0.21 & 0.44	0.09 & 0.12	0.09 & 0.12	-0.02 & 2.33	0.02 & 0.11
C) Curvature Effects					
RMS IN Curvature	0.05	0.03	0.03	0.03	0.03
RMS PE Curvature	0.43	0.32	0.14	15.68	15.68
Critical Value	0.48	0.48	0.48	0.48	0.48
D) Goodness of Fit					
RMSE	5.14	2.42	2.42	4.77	4.77
MAE	3.27	1.44	1.44	3.35	3.35
E) Residual Analysis					
Run test $\lvert Z \rvert$ Value	0.91	0.00	0.00	1.01	1.01
Shapiro-Wilk's Test p-value	0.32	0.03	0.03	0.57	0.57

Gompertz-I showed best fitted model based on the criteria of having least values of RMSE and MAE. Also, residual analyses showed that the randomness assumption and normality assumption are fulfilled. Durbin-Watson test statistic (2.24) which is closed to 2 and we can say that presence of autocorrelation is not significant. This was also supported by the results of Breusch-Godfrey's serial correlation test p-values 0.25 and 0.22 for order one and two respectively. Further, White's test and Breusch-Pagan test p-values 0.33 and 0.36 respectively showed that the assumption of homoscedastic error structure is not violated. Regression diagnostic was also done based on various methods as discussed in methodology section which are presented in Fig. 1 and 2. The studentized residuals are ranging between -2 and +2 indicate that there are no extreme values or outliers (in Fig.1). According to Cook's D plot, the first and eight observations are perhaps influential in the residuals. From the Quantile (Q-Q) plot, the approximate S-shape of the curve seems to suggest that normality assumption is not violated. The Fig.2 shows that the histogram of time series with slight positive skew as compared to the smooth line that represents the theoretical probability density function of normal distribution. According to plots of residual ACF, partial ACF and inverse ACF, residuals are white noise and not autocorrelated as shown in Fig. 2. It is, however, cautioned that the above inferences for small sample size may depend on the distribution of residuals, but for moderate or large sample sizes this is not critical.

As Hougaard's skewness values are less than unity and Box's % bias are also less than 1%, we can say that parameters do not show any extreme nonlinear behavior. Also, RMS IN curvature (0.04) and RMS PE curvature (0.32) of Bates and Watts are less than the corresponding critical value 0.48 and they are acceptable. However, the correlations among the estimated parameters are very high which indicate that the parameters are not independently estimated and they may not be reliable estimates. To rectify the above problem, a partially reparameterized Gompertz model was considered in which the parameter $\hat{\beta}_1$ was taken as an offensive parameter, given in equation (6) and it is referred as 'Gompertz-II'. The parameter β_1 is replaced by W_1 in the process of reparameterization as β_1 is considered to be the offensive parameter. A value of $t_1=8$ was chosen and the corresponding value of $W_1=163.26$ is taken as an initial value for computation of the final estimate of the parameter W_1, which gives the best result in terms of least correlation coefficient. The reparameterized model was refitted to the data and the results are again presented in Table 2. Further improvements in Hougaard's skewness and Box's % bias are also seen in this refitted model of Gompertz-II. Moreover, the high correlations among the estimated parameters are almost eliminated except for $r_{\beta_{23}}$ (0.81), given in Table 3. As a scale parameter, β_2 is not a naturally stable parameter, we do not expect to eliminate this correlation. The graph of fitted model along with observed fish weight is also depicted in Fig. 3 which shows the appropriateness of the proposed model. If there is no any fish mortality during the rearing period, the common carp yield in the 11^{th} and 12^{th} months are best forecasted by the proposed Gompertz-II model as 21.45 Kg and 22.10 Kg respectively with forecasting errors less than ½ Kg of fish (shown in Table 4a). Assuming 20% and 30% fish mortality in each pond upto the end of the 10^{th} month, the forecasting of common carp yield for the 11^{th} and 12^{th} month are given in the Table 4b and 4c respectively.

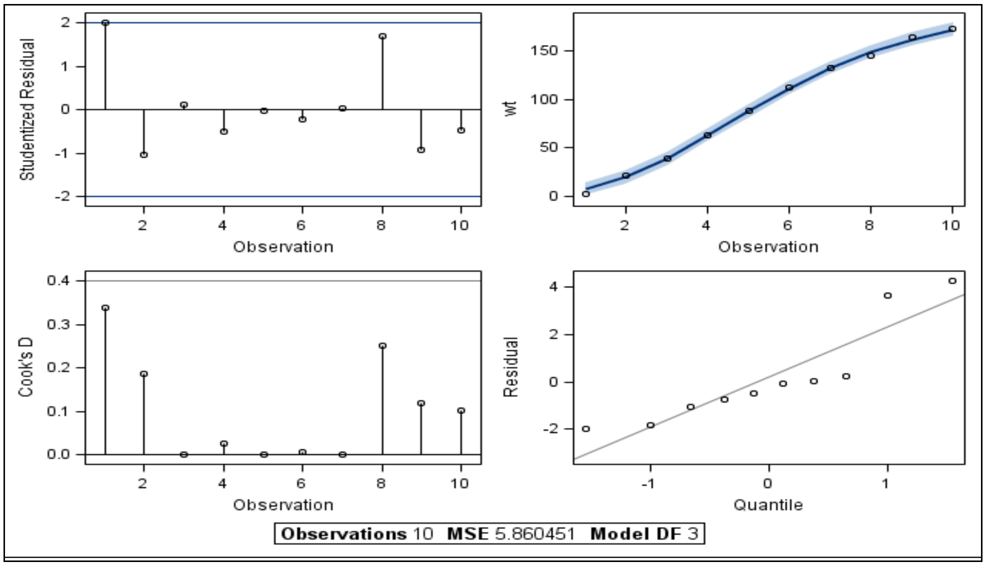

Fig 1: Fit diagnostics for outliers, influential points and normality assumptions.

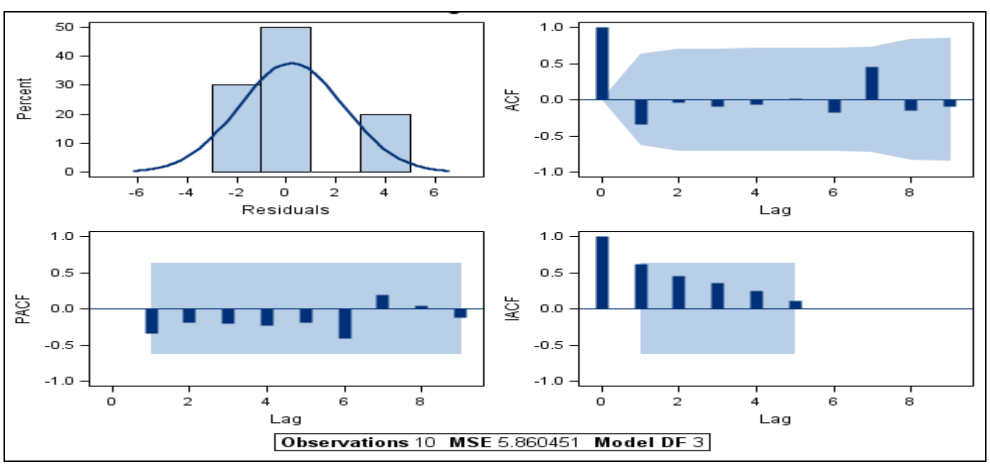

Fig 2: Fit diagnostics for normality and presence of autocorrelation

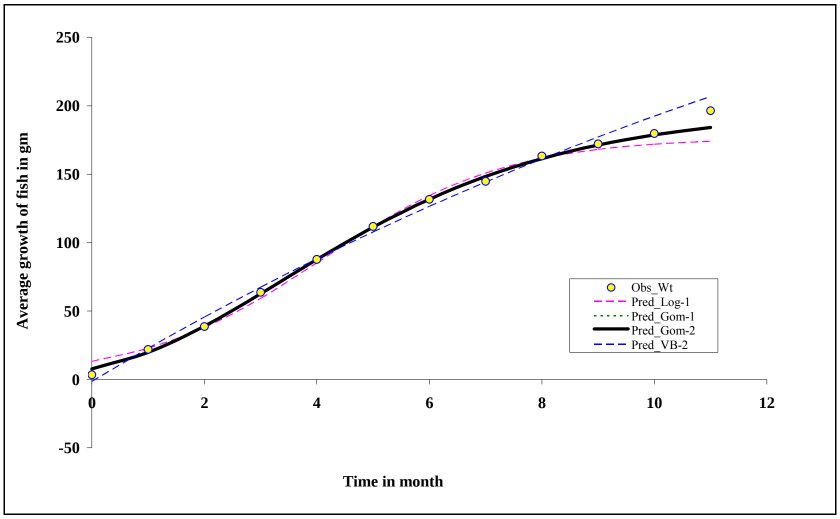

Fig 3: Actual and predicted common carp weights given by different models

Table 3: Mutual correlations among the estimated parameters

Correlation Coefficient	Logistic-I	Gompertz-I	Gompertz-II	Von-Bertalanffy-II	Von-Bertalanffy-III
$r_{\beta_{12}}$ (or $r_{W_1\beta_2}$)	-0.41	-0.55	0.05	-0.99	-0.59
$r_{\beta_{13}}$ (or $r_{W_1\beta_3}$)	-0.78	-0.91	-0.27	0.61	-0.99
$r_{\beta_{23}}$	0.85	0.81	0.81	-0.65	0.63

Table 4a: Actual and forecasting of common carp production (weight in Kg.) from polyponds (if there is no any fish mortality upto the end of the 10th month)

Month	Observed	Logistic-I	Gompertz-I	Gompertz-II	Von-Bertalanffy-II	Von-Bertalanffy-III
11th	21.57	20.63 (0.94)	21.45 (0.12)	21.45 (0.12)	23.09	23.09
12th	23.57	20.89 (2.68)	22.10 (1.47)	22.10 (1.47)	24.83	24.83

Table 4b: Actual and forecasting of common carp production (weight in Kg.) from polyponds (if there is 20% fish mortality per pond upto the end of 10th month)

Month	Observed	Logistic-I	Gompertz-I	Gompertz-II	Von-Bertalanffy-II	Von-Bertalanffy-III
11th	17.26	16.51 (0.75)	17.16 (0.10)	17.16 (0.10)	18.47	18.47
12th	18.86	16.71 (2.15)	17.68 (1.18)	17.68 (1.18)	19.86	19.86

Table 4c: Actual and forecasting of common carp production (weight in Kg.) from polyponds (if there is 30% fish mortality per pond upto the end of 10th month)

Month	Observed	Logistic-I	Gompertz-I	Gompertz-II	Von-Bertalanffy-II	Von-Bertalanffy-III
11th	15.10	14.44 (0.66)	15.01 (0.09)	15.01 (0.09)	16.16	16.16
12th	16.50	14.62 (1.88)	15.47 (1.03)	15.47 (1.03)	17.38	17.38

The bracketed values are the corresponding forecasting errors.

5. Conclusion

Nonlinear regression models are extensively used to study the population growth of many organisms in agriculture. In most of the cases the parameter estimates behave nonlinearly. As a consequence, the parameter estimates generally do not satisfy any optimum properties like that of unbiasedness, minimum variance, and normality. It is also common to observe high correlation among the estimated parameters, which means they are not independently estimated. Interdependence of parameters can cause a slow convergence rate and nonuniqueness and increase parameter uncertainty. The above consequences of high correlation among estimated parameters and nonlinear behaviour of parameters are generally rectified by various reparameterization methods. The present study discusses the concept of partial reparameterization by expected-value parameters to tackle the issue of high correlation among the estimated parameters. Consequently, an explicit form of the Gompertz model was used to illustrate with an average growth data observed on common carp rearing in polyponds under polyculture system. Suitability of the model for two months ahead forecasting of fish yield is also demonstrated.

6. Reference

1. Bates DM, Watts DG. Relative curvature measures of nonlinearity. Journal of the Royal Statistical Society 1980; B-42(1):1-25.
2. Bates DM, Watts DG. Nonlinear regression analysis and its applications. John Wiley and Sons, New York, 1998.
3. Bhar LM, Walia SS, Roy AK. Forecasting fish production from ponds. Project Report, IASRI, New Delhi, 2002.
4. Cook AD, Tsai CL, Wei BC. Bias in nonlinear regression. Biometrika 1986; 3:615-623.
5. Prajneshu. Fitting of Cobb-Douglas production functions: revisited. Agricultural Economics Research Review 2008; 21:289-292.
6. Prajneshu, Kandala VM. Fitting of nonlinear Schaefer model. Indian Journal of Fisheries 2005; 52(1):33-36.
7. Ratkowsky DA. Nonlinear regression modelling - a unified practical approach. New York, Marcel Dekker, 1983.
8. Ratkowsky DA. Handbook of nonlinear regression models. New York, Marcel Dekker, 1990.
9. Ross JSG, Prajneshu, Sarada C. Reparameterization of nonlinear statistical models: a case study. Journal of Applied Statistics 2010; 37(12):2015-26.
10. Singh NO. A Method for fitting of Schaefer model with autoregressive of order one. Asian Fisheries Science 2011; 24(3):197-208.
11. Walia SS, Jain RC, Sehdev RS. Nonlinear statistical models for pre-harvest forecasting inland fish from ponds. Project Report, IASRI, New Delhi, 1997.