Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 70(2) 2016 139–144

Forecasting of Common Carp Fish Production from Ponds using Nonlinear Growth Models—A Modelling Approach

Sanjeev Panwar[1], Anil Kumar[1], Susheel Kumar Sarkar[1], Ranjit Kumar Paul[1], Bishal Gurung[1] and Abhishek Rathore[2]
[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi
[2]International Crops Research Institute for the Semi Arid Tropics, Hyderabad

Received 22 March 2016; Revised 31 May 2016; Accepted 31 May 2016

SUMMARY

Fish cultivation has been medium to earn livelihood from old ages. In the hilly areas due to low temperature, fish cultivation is not as feasible as in plain areas. An attempt has been made to understand the growth and forecast pattern of Carp fish in hilly regions through reparametrization of the existing non-linear growth models viz. Logistic, Gompertz and Von-Bertalanffy. It has been found that partial re-parameterized version of Gompertz model i.e. Gompertz-I model is the best fit under these conditions. It has been illustrated with average growth datasets of common carp fish growth data in polyponds pond environments. Suitability of the models for two months ahead forecasting of fish yield from various ponds have been reported.

Keywords: Carp fish, Nonlinear growth models, Re-parametrization, Hougaards measure, Box % bias.

1. INTRODUCTION

Fish is the cheapest source of animal protein as well as an alternate source of income generation. Moreover, fish culture in hilly regions encourages conserving the water and indigenous biodiversity. In the uplands, aquatic animals are generally stocked in ponds, fertilizer and feed are used to promote rapid growth. Though the carp fish species in polyculture system is well developed in the plain area also, the same fish species are not preferred as candidate species for culture in ponds of upland region due to its low thermal regime. This is considered as a serious problem for the polyculture of carps in uplands. Timely and accurate forecast of fish yield from ponds of upland in particular will be of immense useful for the farmers to plan for marketing their produce profitably.

Growth of fish is not uniform through the year especially in seasonal and fluctuating or seasonal environment. The need to use a seasonal version of growth models has been discussed by many (Kathuria *et al*. 1992, 1993, Bathla *et al*. 1995, Sarada and Prajneshu 2005, El-Shehawy 2010, Ross *et al*. 2010, Ueda *et al*. 2010, Singh *et al*. 2015). Sigmoid curves are frequently used in biology, agriculture and economy to describe growth. Such curves begin at certain point and increase their rate of growth in monotonic form until reaching an inflexion point, after which the growth rate decreases and the function approaches an asymptotic value (Ratkowsky 1983). The mathematical functions proposed to model these curves are: Logistic, Gompertz, von-Bertalanffy.

Following nonlinear models will provide a reasonable representation of average fish size (say, weight or length) W_t at time t whose model function is of the form

$$W_t = f(t, \beta) + \varepsilon_t$$

Corresponding author: Sanjeev Panwar
E-mail address: scientist1775@gmail.com

Logistic model:

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

Gompertz model:

$$W_t = \beta_1 \exp\left[-\beta_2 \exp(-\beta_3 t)\right] \tag{2}$$

Von-Bertalanffy model:

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

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. For the above growth models, expected-value parameters cannot be obtained for the Gompertz model as parameters β_2 and β_3 cannot be eliminated while β_3 cannot be eliminated from logistic model.

If 'β_2' is likely to be an offensive parameter say, in equation (1), then it can be partially reparameterized by expected-value parameter. The parameter which shows nonlinear behavior or likely responsible for high correlation among the estimated parameters is known as 'offensive parameter'. To obtain an expected-value parameter from above equation (1), we need to choose value t_2 of the regressor variable *t*, within the observed range of *t*. Then, we get the expected value from equation (1) as follows:

$$W_2 = \frac{\beta_1}{1+\beta_2 \exp(-\beta_3 t_2)}$$

Solving this equation for the parameter 'β_2' only, we get

$$\beta_2 = \left(\frac{\beta_1}{W_2}-1\right)\exp(\beta_3 t_2)$$

Substituting back into the original equation (1), we get

$$W_t = \frac{\beta_1}{1+\left(\frac{\beta_1}{W_2}-1\right)\exp\left\{-\beta_3\left(t-t_2\right)\right\}} \tag{4}$$

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_2' is reparameterized by expected-value parameter while the other parameters are not changed. The form of the partial reparameterization of the logistic model given by equation (4) is referred to as 'logistic-I' in the subsequent discussions.

Similarly, if 'β_1' is likely to be an offensive parameter say, in equation (2), then it can be partially reparameterized by expected-value parameter. As β_1 represents the asymptotic size, which is more important parameter as compared to the scale parameter, β_2 which is not a naturally stable 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:

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

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

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

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

$$W_t = W_1 \frac{\exp\left[-\beta_2 \exp(-\beta_3 t)\right]}{\exp\left[-\beta_2 \exp(-\beta_3 t_1)\right]} \tag{5}$$

The above model (5) is proposed to mitigate both the nonlinear behaviour of parameters and high correlation among the estimated parameters. The form of the partial reparameterization of the Gompertz model given by equation (5) is referred to as 'Gompertz-I' in the subsequent discussions.

1.1 Measures of Model Accuracy

1.1.1 The White Test (White 1980)

When the form of the heteroscedasticity is of unknown, we can apply the white test. Most test for heteroscedasticity specify some functional form relating the error term to a set of explanatory variables in a particular way e.g. multiplicative. The first step is to estimate the regression model using ordinary least square (OLS) method. Secondly, we obtain the residuals from the OLS regression model. We then obtain the statistic $n \times R^2$ from an auxiliary regression of the residuals on the Z-variables (i.e. the subset of the X-variables we believe are involved), the squares, and the cross-products. The White test is implicitly based on a comparison of the sample variance of the least squares estimators under homoscedasticity and heteroscedasticity. If the null hypothesis cannot be rejected, then the following:

1.1.2 Durbin-Watson Test (Durbin and Watson 1950)

The Durbin-Watson test is based on the assumption that the errors follow AR(1) and the test statistic '*d*' is defined as

$$d = \frac{\sum_{t=2}^{n}\left(\varepsilon_t - \varepsilon_{t-1}\right)^2}{\sum_{t=1}^{n}\varepsilon_t^2},\ 0 \le d \le 4$$

A statistic '*d*' value ranges between 0 and 4. A value of '*d*' near 2 indicates little autocorrelation; a value toward 0 indicates positive autocorrelation while a value toward 4 indicates negative auto-correlation.

To examine model performance, a measure of how the predicted and observed variables cover in time is needed. Thus, the coefficient of determination, R^2 is generally used.

$$R^2 = 1 - \frac{\sum_{t=1}^{n}\left(W_t - \hat{W}_t\right)^2}{\sum_{t=1}^{n}\left(W_t - \overline{W}\right)^2},$$

Further, it is desirable to use some other summary statistics like root mean square error (RMSE) and mean absolute error (MAE):

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

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

where $\hat{W}_t$ = Predicted fish weight of t^{th} observation; $\hat{W}$ = Average fish weight; n = number of observations, t = 1, 2, ..., n.. The better model will have the least values of RMSE and MAE while larger value of R^2 is expected for the same. 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 used. However, the normality assumption is not so stringent for selecting nonlinear models because their residuals may not follow normal distribution.

1.2 Intrinsic and Parameter Effects Nonlinearity

Bates and Watts (1980) proposed measures to assess the adequacy of the linear Taylor approximation of the regression function using two measures of nonlinearity, the maximum intrinsic curvature (IN) and the maximum parameter-effects curvature (PE).

1.2.1 Hougaard's Measure of Skewness and Box % Bias (Hougaard 1985)

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

$$E\left[\hat{\beta}_t - E\left(\hat{\beta}_t\right)\right]^3 = -\left(MSE\right)^2 \sum_{jkl}^{np} L^{ij} L^{jk} L^{il} \left(W_{jkl} + W_{kjl} + W_{lkj}\right)$$

Moreover, the bias of *Box reveals* which parameters are responsible for the nonlinear behavior. The bias of Box is calculated in multivariate form as given by Cook *et al*. (1986):

$$\text{Bias} = \left(D2^T D2\right)^{-1}\left(D2^T H2\right)$$

where D2 is the $n \times p$ first derivative matrix; $H2 = -1/2\,\sigma^2 trac\left\{\left(D2^T D2\right)^{-1} F2_i\right\}$, the expected difference between the quadratic and linear components of the Taylor approximation and $F2_t$, $t = 1, 2, ..., n$ are $p \times p$ faces of the second derivative matrix.

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

Here, $\hat{\beta}$ is the estimated parameter. Ratkowsky (1983) suggested using an absolute value of greater than 1% as an indicator of nonlinear behavior.

1.3 Forecast/ Validation of Models

For validation of the forecast models, developed through above approaches was done on the basis of RMSE and MAE.

2. ILLUSTRATION

2.1 Dataset

The growth data of a sample of size 30 per each pond type comprises of 10 specimens from 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 (Polyponds were created for conducting experiments on integrated fish culture by selecting three clusters of villages in Champawat district of Kumaun region by Directorate of Coldwater Fisheries Research, Bhimtal, Uttarakhand). The average weight of each fish species obtained from 30 (10 individuals per each type of 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 statistical analyses.

3. RESULTS AND DISCUSSION

Table 1. Basic information of the datasets obtained from polyponds

S. No.	Variable	No.	Min.	Max.	Mean	Std. Deviation
1.	Common Carp Weight of Fish (in gm)	12	3.36	196.42	109.56	65.78

Table 2. Summary statistics of fitted models for common carp from polyponds

	Logistic	Gompertz	Gompertz-I
A) Parameter Estimation			
β_1 (or, W_1)	176.60 (6.20)	198.00 (5.35)	161.40 1.33)
β_2	12.43 (1.96)	3.25 (0.13)	3.25 (0.13)
β_3	0.61 (0.05)	0.35 (0.02)	0.35 (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
β_2	0.76 & 2.45	0.30 & 0.24	0.30 & 0.24
β_3	0.21 & 0.44	0.09 & 0.12	0.09 & 0.12
C) Curvature Effects			
RMS IN Curvature	0.05	0.03	0.03
RMS PE Curvature	0.43	0.32	0.14
Critical Value	0.48	0.48	0.48
D) Goodness of Fit			
RMSE	5.14	2.42	2.42
MAE	3.27	1.44	1.44
E) Residual Analysis			
Run test $\lvert Z \rvert$ Value	0.91	0.00	0.00
Shapiro-Wilk's Test p-value	0.32	0.03	0.03
[1]D-W Test Statistic	-	2.24	2.24
[2]B-G Test p-value	-	0.25	0.25
White's Test p-value	-	0.33	0.33

[1]Durbin-Watson test statistic value and [2]Breusch-Godfrey's serial correlation test p-value for order one.

Table 3. Mutual correlations among the estimated parameters for common carp from poplyponds

Correlation Coefficient	Logistic	Gompertz	Gompertz-I
$r_{\beta_{12}}$ (or $r_{w_1\beta_2}$)	-0.41	-0.55	0.05
$r_{\beta_{13}}$ (or $r_{w_1\beta_3}$)	-0.78	-0.91	-0.27
$r_{\beta_{23}}$	0.85	0.81	0.81

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

Month	Observed	Logistic	Gompertz	Gompertz-I
11^{th}	21.57	20.63 (0.94)	21.45 (0.12)	21.45 (0.12)
12^{th}	23.57	20.89 (2.68)	22.10 (1.47)	22.10 (1.47)

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

Month	Observed	Logistic	Gompertz	Gompertz-I
11^{th}	17.26	16.51 (0.75)	17.16 (0.10)	17.16 (0.10)
12^{th}	18.86	16.71 (2.15)	17.68 (1.18)	17.68 (1.18)

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

Month	Observed	Logistic	Gompertz	Gompertz-I
11^{th}	15.10	14.44 (0.66)	15.01 (0.09)	15.01 (0.09)
12^{th}	16.50	14.62 (1.88)	15.47 (1.03)	15.47 (1.03)

The bracketed values are the corresponding forecasting errors.

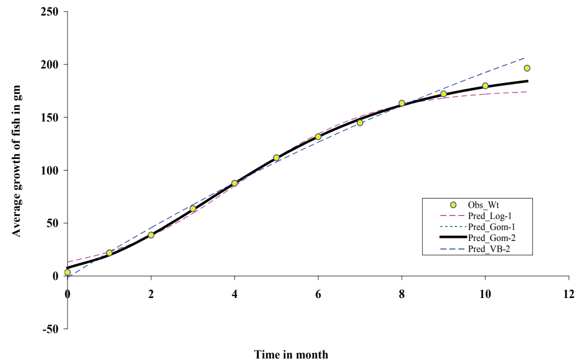

Fig. 1. Actual and predicted common carp weights (in gm) from polyponds provided by different models

3.1 Polyponds

The average growth data in terms of weight (in kg) of common carp obtained from polyponds were considered for illustrations and the basic statistics of the datasets are presented in Table 1. Three different nonlinear models were fitted to the above growth datasets. Von-Bertalanffy model failed to give optimum solution irrespective of the fish species. The summary statistics for fitting of other models are presented in Table 2 for common carp respectively. Gompertz model is found to be the best fitted model based on the criteria of having least values of RMSE and MAE for common carp. The best models identified above are retained for detail analysis as discussed below. The residual analyses showed that the randomness assumption and normality assumption are fulfilled. Durbin-Watson test statistic is ranging between 2.24-2.64 which is approximately 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-value lies between 0.18-0.25 for order one. Further, White's test p-value lies between 0.33-0.36 showed that the assumption of homoscedastic error structure is not violated.

As Hougaard's skewness values are less than unity and Box's % bias are also less than 1% except for silver carp, we can say that parameters do not show any extreme nonlinear behavior for common carp. However, RMS PE curvature of Bates and Watts is greater than the corresponding critical value in case of grass carp and it may not be acceptable. Moreover, the correlations among the estimated parameters are also extremely high in some cases which indicate that the parameters are not independently estimated and they may not be reliable estimates. To rectify the above problems of high correlation among the estimated parameters and nonlinear behavior of them, partially reparameterized versions were developed. The Gompertz model was considered for common carp in which the parameter $\hat{\beta}_1$ was taken as an offensive parameter, given in equation (5) and it is referred as 'Gompertz-I'.

A value of t_1 = 8 was chosen and the corresponding value of W_1=163.26 for common carp was taken as an initial values for computation of the final estimate of the parameter W_1, which provided the best result in terms of least correlation coefficient. The reparameterized model was refitted to the datasets and the results are again presented. Further improvements in Hougaard's skewness and Box's % bias are also seen in these refitted models. Moreover, the high correlations among the estimated parameters are almost eliminated except with β_2 in some cases. As a scale parameter, β_2 is not a naturally stable parameter, we do not expect to eliminate this correlation (Ross *et al*. 2010). The graphs of fitted model along with observed fish weight were also depicted in Fig. 1 which shows the appropriateness of the proposed models. 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-I model as 21.45 Kg and 22.10 Kg respectively (shown in Table 4a). Assuming 20% and 30% fish mortality in each pond upto the end of the 10^{th} month. Although logistic model was found to be the best fitted model for silver carp, Gompertz model gives the better results in terms of fish yield forecasting for two months ahead.

3.2 Conclusion

The present study discusses the concept of partial reparameterization by expected-value parameter to tackle the issue of high correlation among the estimated parameters as well as nonlinear behavior of estimated parameters. Consequently, explicit form of the partially reparameterized versions of Gompertz model was developed which were illustrated with average growth datasets of fish species viz. common carp obtained from polyponds pond environments. Suitability of the models for two months ahead forecasting of fish yield from various ponds were also demonstrated.

REFFERENCES

Avanza, M.M., Bramardi, S.J. and Mazza, S.M. (2008). Statistical models to describe the fruit growth pattern in sweet orange 'valencia late'. *Spanish J. Agric. Res*., **6(4)**, 577-785.

Bates, D.M. and Watts, D.G. (1980). Relative curvature measures of nonlinearity. *J. Roy. Statist. Soc.,* B, **42(1)**, 1-25.

Bates, D.M and Watts, D.G. (1998). *Nonlinear Regression Analysis and its Applications*. John Wiley and Sons, New York.

Clarke, G.P.Y. (1987). Marginal curvature and their usefulness in the analysis of nonlinear regression models. *J. Amer. Statist. Assoc.*, **62**, 844-850.

Cook, A.D., Tsai, C.L. and Wei, B.C. (1986). Bias in nonlinear regression. *Biometrika*, **(3)**, 615-623.

Durbin, J. and Watson, G.S. (1950). Testing for serial correlation in least squares regression, I. *Biometrika*, **37(3–4)**, 409-428.

Hougaard, P. (1985). The appropriateness of the asymptotic distribution in a nonlinear regression model in relation to curvature. *J. Roy. Statist. Soc*., Series **B** (Methodological) 103-114.

Ratkowsky, D.A. (1983). *Nonlinear Regression Modelling A Unified Practical Approach*. Marcel Dekker, New York

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

Ross, J.S.G., Prajneshu and Sarada, C. (2010). Reparameterization of nonlinear statistical models: a case study. *J. Appl. Statist.*, **37(12)**, 2015-2026.

Sarada, C. and Prajneshu (2005). On appropriate parameterization of nonlinear statistical model. *J. Ind. Soc. Agril. Statist.,* **59**, 237-242.

Singh, S., Paul, A.K., Paul, R.K., Bhar, L.M., Kumar, A. and Alam, W. (2015) Study of growth pattern of cattle under different error structures. *Model Assist. Statist. Appl.*, **10(2),** 109-115.

Walia, S.S., Jain, R.C. and Sehdev, R.S. (1997). Nonlinear statistical models for pre-harvest forecasting inland fish from ponds. *Project Report*. IASRI, New Delhi.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, **48(4)**, 817-838.