Indian J. Fish., 65(3): 99-104, 2018
DOI: 10.21077/ijf.2018.65.3.70601-13

Note

Growth modelling and forecasting of common carp and silver carp in culture ponds: A re-parametrisation approach

SANJEEV PANWAR, ANIL KUMAR, K. N. SINGH, SUSHEEL KUMAR SARKAR, BISHAL GURUNG AND ABHISHEK RATHORE
Coordination Unit, Indian Council of Agricultural Research, Krishi Bhawan, New Delhi - 110 001, India
e-mail: sanjeevp.icar@nic.in

ABSTRACT

The available forecasting models for growth pattern in fish are based on either classical approach or a particular growth model. In the present study, reparamerisation methodologies were attempted for forecasting growth of fish cultured in cemented ponds of plain areas. Forecasting methodology is not readily available for any other types of ponds for uplands of India. So, other appropriate growth curves (Logistic, Gompertz and von-Bertalanffy) were considered while developing the most suitable model for forecasting fish (common carp *Cyprinus carpio* var *communis* and silver carp *Hypophthalmichthys molitrix*) production from cemented ponds. Gompertz-1 and Logistic-1 models gave the best fit as well as fish yield forecasting, two months ahead from various ponds.

Keywords: Common carp, Culture ponds, Forecasting, Growth models, Production, Reparameterisation, Silver carp

Aquaculture encourages conserving the water, indigenous biodiversity of aquatic plants and animals. It also generates employment and serves as an alternate source of income generation especially in hilly region. Moreover, fish is the cheapest source of animal protein available today and people are more aware of the presence of omega-3 fatty acids which helps in prevention of cardiac diseases. In fact, fish and fishery related activity has increased many fold in the recent past as the demand on fish biomass has risen significantly. A good number of aquaculture practices and packages have been developed by many researchers which normally offer benefits to the fish farmers. However, forecasting of fish growth/ production in culture systems has not been taken seriously so far.

This important task has not been seriously taken into consideration in the uplands although a little work has been done for plains of India. The available methodology for forecasting fish production from ponds is based on either the classical assumptions or a particular growth model. While developing such types of forecasting models, the importance of close-to-linear behaviour of parameters and the high correlation among the estimated parameters were never highlighted. The above circumstances are encountered in most of the practical situations with a variety of consequences. The well-known solution to above issue is the reparameterisation of those parameters of the model. Reparameterisation by expected-value parameters is most widely used. However, the concept of reparameterisation is rarely employed in forecasting fish production. Sigmoid curves (Logistic, Gompertz and von-Bertalanffy) 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.

The growth data of a sample of size of 30 per each pond type comprising of 10 fish from fish group (common carp *Cyprinus carpio* var *communis* and silver carp *Hypophthalmichthys molitrix*) 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 each fish species recorded (10 individuals per each type of pond), at monthly intervals was utilised for the present study and thus there were 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 all analyses.

Following nonlinear growth models will provide a reasonable representation of average fish size (say, weight), W_t at time t whose model function is of the form $W_t = f(t,\beta)+\varepsilon_t$:

Logistic model:

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

Gompertz model:

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

Von-Bertalanffy model:

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

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.

β_2 is likely to be an offensive parameter say, in equation (1), it can be partially reparameterised by expected-value parameter. The parameter which shows nonlinear behaviour 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, the expected value can be estimated 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', 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,

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

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

Similarly, β_1 is likely to be an offensive parameter say, in equation (2), it can be partially reparameterised 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: $W_t = \beta_1 \exp[-\beta_2 \exp(-\beta_3 t)]$

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

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

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

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

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 reparameterisation of the Gompertz model given by equation (5) is referred to as 'Gompertz-1' in the subsequent discussions.

As we are dealing with time-series data, it is, therefore, required to check for the validity of the above model by examining the independency assumption of error term. The Durbin-Watson test has been employed for the said purpose and is based on the assumption that the errors (ε_t's) follow autoregressive of order one. To handle a situation when there is an evidence for the presence of autocorrelation, an autoregressive (AR) error term ε_t of order one may be added to the right hand side of above equations: $\varepsilon_t = \Phi\varepsilon_{t-1} + u_t$; $|\Phi|<1$,

where u_t are independently and normally distributed with zero mean and constant variance and Φ denotes the autoregressive parameter. Incorporating an AR(1) additive error structure, the above 'logistic-I' model will become:

$$W_t = \frac{\beta_1}{1+\left(\frac{\beta_1}{W_2} - 1\right) \exp\{-\beta_3 (t - t_2)\}} + \Phi\varepsilon_{t-1} + u_t \quad (6)$$

The above equation (6) will be referred as 'logistic-II' in the subsequent discussions.

When the form of the heteroscedasticity is unknown, we can apply the White test. Most tests for heteroscedasticity specify some functional form relating the error term to a set of explanatory variables in a particular way. 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 x R^2 from an auxiliary regression of the residuals on the Z-variables (*i.e.,* the subset of the X-variables are involved), the squares, and the cross-products. The White test is implicitly based on a comparison of the sample variance of the least square estimators under homoscedasticity and heteroscedasticity.

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

$$d = \frac{\sum_{t=2}^{n} (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{n} \varepsilon_t^2}, \; 0 \leq d \leq 4.$$

The statistic 'd' value ranges between 0 and 4. Value of 'd' near 2 indicates little autocorrelation; a value towards 0 indicates positive autocorrelation, while a value towards 4 indicates negative autocorrelation.

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.

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

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

$$RMSE = \left[\sum_{t=1}^{n} (W_t - W_t)^2 / n\right]^{1/2} \text{ and } MAE = \left[\sum_{t=1}^{n} |(W_t - W_t)| / n\right]$$

where, W_t = predicted fish weight of t^{th} observation; W = average fish weight; n = number of observations and 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 has been used. However, the normality assumption is not so stringent for selecting nonlinear models because their residuals may not follow normal distribution.

Bates and Watts (1980,1988) 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).

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:

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

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 by Cook *et al.* (1986):

$$Bias = (D2^T D2)^{-1} (D2^T H2)$$

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

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

Here, $\hat{\beta}$ is the estimated parameter. Ratkowsky (1983, 1990) suggested using an absolute value of greater than 1% as an indicator of nonlinear behaviour.

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

The average growth data in terms of weight (kg) of common carp and silver carp obtained from cemented ponds were analysed and the basic statistics of the datasets are presented in Table 1. 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 for common carp and silver carp are presented in Table 2a and b, Mutual correlations among the estimated parameters for common carp from cemented ponds are also presented in Table 3a and b respectively. Gompertz model is found to be the best fitted model based on the above criteria for common carp, however, Logistic model was found appropriate for silver carp. The best models identified above are retained for detailed analysis as explained above. The residual analyses showed that the randomness assumption and normality assumption are fulfilled. Further, White's test p-value lies between 0.27-0.29 which indicated that the assumption of homoscedastic error structure is not violated. Durbin-Watson test statistics for common carp is 2.84. But, Breusch-Godfrey's serial correlation test p-values of 0.11 and 0.28 for order one respectively for common carp showed that the presence of autocorrelation

Table 1. Basic information of the datasets obtained from cemented ponds

Variable	n	Minimum	Maximum	Mean	Standard deviation
Common carp					
Weight (g)	12	19.40	135.10	93.89	39.36
Silver carp					
Weight (g)	12	3.50	162.90	77.68	64.10

Table 2a. Summary statistics of fitted models for common carp from cemented ponds

	Logistic	Gompertz	Gompertz-I
A) Parameter estimation			
β_1 (or, W_1)	134.20 (4.17)	144.20 (5.36)	125.40 (1.65)
β_2	4.27 (0.42)	1.85 (0.08)	1.85 (0.08)
β_3	0.51 (0.04)	0.32 (0.03)	0.32 (0.03)
B) Hougaard's Skewness & Box's % Bias			
β_1 (or, W_1)	0.41 & 0.25	0.49 & 0.32	$0.0^{3}2$ & $0.0^{5}1$
β_2	0.46 & 1.17	0.25 & 0.39	0.25 & 0.39
β_3	0.15 & 0.35	0.09 & 0.22	0.09 & 0.22
C) Curvature effects			
RMS IN Curvature	0.05	0.03	0.03
RMS PE Curvature	0.36	0.51	0.16
Critical Value	0.48	0.48	0.48
D) Goodness of fit			
RMSE	3.69	3.02	3.02
MAE	2.64	1.94	1.94
E) Residual analysis			
Run test |Z| value	$0.0^{4}1$	1.01	1.01
Shapiro-Wilk's test p-value	0.81	0.78	0.78
[1]D-W test statistic	-	2.84	2.84
[2]B-G test p-value	-	0.11	0.11
White's test p-value	-	0.27	0.27

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

Table 2b. Summary statistics of fitted models for silver carp from cemented ponds

	Logistic	Logistic-I	Logistic-II	Gompertz
A) Parameter estimation				
β_1	174.20 (4.23)	174.20 (4.23)	169.70 (4.77)	198.20 (16.08)
β_2 (or, W_2)	64.94 (13.48)	133.70 (2.11)	134.80 (2.14)	7.38 (1.68)
β_3	0.67 (0.04)	0.67 (0.58)	0.72 (0.05)	0.36 (0.05)
□	-	-	0.65 (0.47)	-
B) Hougaard's skewness & Box's % bias				
β_1	0.24 & 0.11	0.24 & 0.11	0.23 & 0.13	0.84 & 1.07
β_2 (or, W_2)	0.86 & 3.32	0.01 & $0.0^{5}3$	$0.0^{2}5$ & $0.0^{2}2$	1.24 & 5.21
β_3	0.19 & 0.28	0.19 & 0.28	0.35 & 0.52	0.32 & 1.21
□	-	-	0.07 & 0.50	-
C) Curvature effects				
RMS IN Curvature	0.05	0.05	0.06	0.12
RMS PE Curvature	0.72	0.25	0.40	0.99
Critical value	0.48	0.48	0.49	0.48
D) Goodness of Fit				
RMSE	3.87	3.87	3.67	7.25
MAE	2.99	2.99	2.81	5.79
E) Residual analysis				
Run test |Z| Value	0.84	0.84	0.21	1.27
Shapiro-Wilk's Test p-value	0.18	0.18	0.53	0.09
[1]D-W Test Statistic	1.23	1.23	1.99	-
[2]B-G Test p-value	0.03	0.03	0.83	-
White's Test p-value	0.29	0.29	0.36	-

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

Table 3a. Mutual correlations among the estimated parameters for common carp from cemented ponds

Correlation coefficient	Logistic	Gompertz	Gompertz-I
$r_{\beta 12}$(or, $r_{W1\beta 2}$)	-0.19	-0.19	0.22
$r_{\beta 13}$(or, $r_{W1\beta 2}$)	-0.80	-0.91	-0.40
$r_{\beta 23}$	0.65	0.50	0.50

Table 3b. Mutual correlations among the estimated parameters for silver carp from cemented ponds

Correlation coefficient	Logistic	Logistic-I	Logistic-II	Gompertz
$r_{\beta 12}$(or, $r_{\beta 2\,W2}$)	-0.56	-0.26	-0.37	-0.77
$r_{\beta 13}$	-0.74	-0.74	-0.83	-0.91
$r_{\beta 23}$ (or, $r_{W2\,\beta 3}$)	0.95	0.67	0.71	0.95

$r_{\beta\varphi 1}$=-0.65; $r_{\phi W2}$ = 0.37 and $r_{\beta\varphi 3}$ = 0.64.

is not significant. However, Durbin-Watson test statistic and the Breusch-Godfrey's serial correlation test p-value for order one are respectively 1.23 and 0.03 for silver carp, which indicates the presence of autocorrelation.

As Hougaard's skewness values are less than unity and we can say that parameters do not show any extreme nonlinear behaviour for common carp and silver carp. However, RMS PE curvature of Bates and Watts (1980) is greater than the corresponding critical value irrespective of the fish species and it may not be acceptable. Moreover, the correlations among the estimated parameters are also extremely high in some cases. To rectify the above problems of high correlation among the estimated parameters and nonlinear behaviour of the parameters, partially reparameterised versions were attempted. 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'. Similarly, logistic model was considered for silver carp and the parameter $\hat{\beta}_2$ was taken as an offensive parameter, given in equation (4) and it is referred as 'logistic-I'. Further, the presence of autocorrelation was detected in case of silver carp data and thus the model 'logistic-I' was modified by incorporating the AR(1) error structure provided by equation (6) and it is referred to as 'logistic-II'.

A value of t_1=8 was chosen and the corresponding value of W_1=124.8 for common carp was taken as initial values for computation of the final estimate of the parameter W_1, which provided the best result in terms of least correlation coefficient. However, a value of t_2=8 and the corresponding value of W_2=134.7 for silver carp was chosen in the similar fashion. The reparameterised 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 the scale parameter β_2 is not a naturally stable parameter, we do not expect to eliminate this correlation. In case of silver carp data, the presence of positive autocorrelation was suspected as Durbin-Watson statistic was 1.23. Thus, logistic-II model was refitted and the summary statistics are presented in Table 2b. Also, residual analyses showed that the randomness assumption and normality assumption are fulfilled. Durbin-Watson test statistic (1.99) which is nearly close to 2 and we can say that presence of autocorrelation is negligible. Further, this is supported by the results of Breusch-Godfrey's serial correlation test p-value 0.83 for order one. Further, White's test p-value 0.36 showed that the assumption of homoscedastic error structure is not violated. Further improvements in other statistics like RMSE, MAE, Hougaard's skewness and Box's % bias are also seen. The correlations among the estimated parameters are not showing any extreme. The graphs of fitted model along with observed fish weight are also depicted in Figs. 1 and 2 which show the appropriateness of the proposed

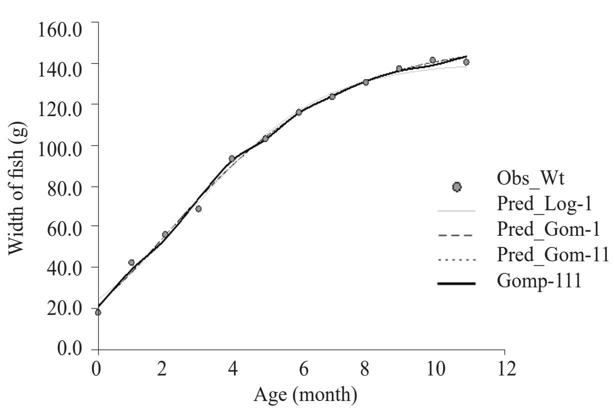

Fig. 1. Actual and predicted weights of common carp (g) from cemented ponds provided by different models

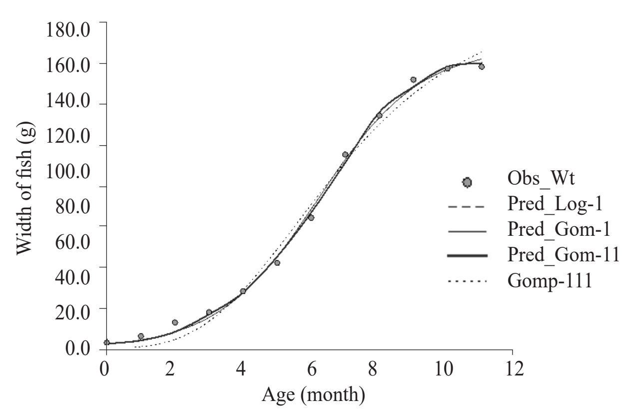

Fig. 2. Actual and predicted weights of silver carp (g) from cemented ponds provided by different models

models. If there is no fish mortality during the rearing period, the common carp yield in the 11th and 12th months are forecasted by the proposed Gompertz-I model as 11.78 and 11.89 kg respectively (Table 4a). Assuming 20 and 30% fish mortality in each pond upto the end of the 10th month, the forecasting of common carp yield for the 11th and 12th month are given in Table 4b and c respectively. Similarly, silver carp yield forecasted by appropriate models are provided in Table 5a, b and c.

Table 4a. Actual and forecast of common carp yield (weight in kg) from cemented ponds (if there is no fish mortality upto the end of the 10th month)

Month	Observed	Logistic	Gompertz	Gompertz-I
11th	12.15	11.78 (0.37)	12.06 (0.09)	12.06 (0.09)
12th	12.07	11.89 (0.18)	12.31 (0.24)	12.31 (0.24)

Table 4b. Actual and forecast of common carp yield (weight in kg) from cemented ponds (if there is 20% fish mortality per pond upto the end of 10th month)

Month	Observed	Logistic	Gompertz	Gompertz-I
11th	9.72	9.42 (0.30)	9.65 (0.07)	9.65 (0.07)
12th	9.65	9.52 (0.13)	9.85 (0.20)	9.85 (0.20)

Table 4c. Actual and forecast of common carp yield (weight in kg) from cemented ponds (if there is 30% fish mortality per pond upto the end of 10th month)

Month	Observed	Logistic	Gompertz	Gompertz-I
11th	8.51	8.24 (0.27)	8.44 (0.07)	8.44 (0.07)
12th	8.45	8.33 (0.12)	8.62 (0.17)	8.62 (0.17)

Values in parenthesis correspond to forecasting errors

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

Table 5a. Actual and forecast of silver carp yield (weight in kg) from cemented ponds (if there is no fish mortality upto the end of the 10th month)

Month	Observed	Logistic	Logistic-I	Logistic-II	Gompertz
11th	14.56	14.53 (0.03)	14.53 (0.03)	14.66 (0.10)	14.49 (0.07)
12th	14.66	15.10 (0.44)	15.10 (0.44)	14.85 (0.19)	15.42 (0.76)

Table 5b. Actual and forecast of silver carp yield (weight in kg) from cemented ponds (if there is 20% fish mortality per pond upto the end of 10th month)

Month	Observed	Logistic	Logistic-I	Logistic-II	Gompertz
11th	11.65	11.62 (0.03)	11.62 (0.03)	11.73 (0.08)	11.59 (0.06)
12th	11.73	12.05 (0.32)	12.05 (0.32)	11.88 (0.15)	12.34 (0.61)

Table 5c. Actual and forecast of silver carp yield (weight in kg) from cemented ponds (if there is 30% fish mortality per pond upto the end of 10th month)

Month	Observed	Logistic	Logistic-I	Logistic-II	Gompertz
11th	10.19	10.17 (0.02)	10.17 (0.02)	10.26 (0.07)	10.14 (0.05)
12th	10.27	10.55 (0.28)	10.55 (0.28)	10.40 (0.13)	10.80 (0.53)

Values in parenthesis correspond to forecasting errors

References

Bates, D. M. and Watts, D. G. 1980. Relative curvature measures of nonlinearity. *J. R. Stat. 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.

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

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.

Walia, S. S., Jain, R. C. and Sehdev, R. S. 1997. Nonlinear statistical models for pre-harvest forecasting inland fish from ponds. *Project Report*. *Indian Agricultural Statistics Research Institute,* Delhi.

Date of Receipt : 24.05.2017
Date of Acceptance : 13.07.2018