Indian J. Fish., 65(3): 25-32, 2018
DOI: 10.21077/ijf.2018.65.3.76544-03

Local polynomial regression estimation of trawl size selectivity parameters using genetic algorithm

C. G. JOSHY, N. BALAKRISHNA* AND V. R. MADHU
ICAR-Central Institute of Fisheries Technology, Matsyapuri P. O., Willington Island, Kochi - 682 029, Kerala, India
**Department of Statistics, Cochin University of Science and Technology, Kochi - 682 022, Kerala, India*
e-mail: cgjoshy@gmail.com

ABSTRACT

This study used a local polynomial generalised linear model to estimate the trawl selectivity curve and its parameters. This modeling technique was applied to trawl selectivity data obtained from the codend selectivity studies of the Dussumier's anchovy *Thryssa dussumieri*, an important trawl resource along the Gujarat coast, India, for 40 mm diamond and square mesh codends. The results of this model were compared with the results obtained from the parametric approach and found to have a superior fit based on the model performance statistics. Genetic algorithm was used to estimate the trawl selectivity parameters by minimising the objective function, *i.e.*, estimated squared distance from target. The nonparametric approach was used to estimate the trawl selectivity parameter values (L_{50} and SR) for two species *viz*., *Upeneus moluccensis* and *Trichiurus lepturus* to confirm its superiority over the parametric approach.

Keywords: Genetic algorithm, Parametric and nonparametric regression, Square and diamond mesh, *Thryssa dussumieri*, Trawl selectivity

Introduction

Trawls contribute significantly to the total catch from Indian waters. Estimation of selection properties of trawl is important for stipulating the most effective mesh sizes and is the most widely used fishing gear based management measure for sustainable exploitation of fisheries resources. Accurate estimation of fishing gear selectivity is considered as an important aspect for conservation fisheries. The goal of selectivity in fishing gear is to catch only fishes above the minimum legal size (MLS) and thus to reduce the proportion of bycatch (Pope *et al*., 1975; Millar and Walsh, 1992; Suuronen, 1995; Wileman *et al.,* 1996; Millar and Fryer, 1999; Hall *et al*., 2000; Cook, 2003). This ensures that fishing operations are economically as well as ecologically sensible by maintaining fish stocks at a sustainable level (Collie *et al*., 2000; Thrush and Dayton 2002). Therefore, accurate estimation of gear selectivity also determines the profit/loss of fishing operations in economic terms. Additionally, this would help researchers or fishery managers to formally compare the performance of different fishing gears; and thus, provide convincing evidence to the end-user regarding the benefit of a newly developed fishing gear, which would improve its adoption. Most of the works for optimising mesh size in tropical waters is based on the premise that at least 50% of the population is left out as breeding stock, so the optimum mesh size is fixed at that length which ensure that half of the population is matured.

The main aim of size-selectivity experiments is to investigate the contact-selection curve of the chosen gear. Millar and Fryer (1999) defined fishing gear selectivity as "the (relative) probability that a fish of length 'l' is captured given that it contacted the gear" and it is being quantified by two parameters *viz*., $L_{50,}$ the length of fishes with a 50% retention probability; and SR, the selection range which is obtained by taking the difference between the lengths of fishes with 25% (L_{25}) and 75% (L_{75}) retention probabilities.

To understand selectivity further, a covered-codend experimental study was considered, where we catch more than one fish at a fixed length l_i in the codend and cover. The fishes caught can be classified into 'k' groups of different length classes. When more than one fish is caught at a fixed length $l_{i,}$ in the codend, it is sufficient to take number of fishes caught in codend and number of fishes caught in both codend and cover rather than taking binary responses. Let $Y_{l_1}, Y_{l_2}, ..., Y_{l_i}, ... Y_{l_k}$ be the number of fishes caught, the attribute of interest, in each length class ($l_1, l_2, ..., l_i, ..., l_k$) in the codend and ($n_{l_1}, n_{l_2}, ..., n_{l_i}, ..., n_{l_k}$) be the total number of fishes caught in each length class in both codend and cover. Then, ($Y_{l_1}, Y_{l_2}, ..., Y_{l_i}, ..., Y_{l_k}$) are independent binomial random variables with equal probability $\pi(l_i)$ to catch a fish at length l_i. Now using

The random variables Y_{l_i} can take the values 0, 1, ..., n_{l_i} and n_{l_i} observations in each length class are independent. The total number of fishes caught in the codend is:

$$Y_{l_i} = \sum_{j=1}^{n_{l_i}} Y_{l_{ij}},\ j = 0,1,2...,\ n_{l_i};\ i = 1,2,\ ...,\ k$$

and Y_{l_i} follows binomial distribution with parameters $\pi(l_i)$ and n_{l_i} (Millar and Fryer, 1999).

The standard parametric function used to model $\pi(l_i) = \Pr\ \{Y_{l_i} = 1 |\ l_i\}$ as a function of l_i in a trawl size-selectivity study is the Logistic curve (McCullagh and Nelder, 1989). One of the key assumptions in ordinary logistic regression is that, observations are independent of each other. Violations of the assumption of independence of observations may result in incorrect statistical inferences. In a trawl selectivity study, subjects of same species are measured at monotonous length interval which may lead to correlated observations. Generalised estimating equation approach was used by Prentice (1998) for the regression analysis of correlated binary data when each binary observation has its own covariates. Cessie and Houwelingen (1994) modeled correlated binary data using logistic regression by keeping marginal response probabilities as still logistic.

In the present study, in order to fit fishing gear selectivity curve and estimate selectivity parameters in a trawl operation accurately, we considered generalised parametric and nonparametric regression models for fitting selectivity curve. Based on the best fitted model, the selectivity parameters were estimated using a powerful genetic algorithm (GA).

Materials and methods

In this section, we discuss in detail parametric and nonparametric regression models to fit trawl size-selectivity curve and estimation of selectivity parameter values.

Parametric regression for trawl size-selectivity curve

The functional form of the logistic selection curve with regression coefficients β_0 and β_l for a trawl size-selectivity study is defined as:

$$\pi(l_i) = \left(\exp(\beta_0 + \beta_l l_i) \Big/ \{1 + \exp(\beta_0 + \beta_l l_i)\} \right), i = 1, 2, \cdots, k. \qquad (1)$$

The response variable Y_{l_i} is binary and the range of $\pi(l_i)$ on the left hand side of the above model lies between 0 and 1. Therefore, logit link function is used to transform $\pi(l_i)$ to linearise the relationship and it is defined as:

$$\text{logit}(\pi(l_i)) = \log\left(\pi(l_i) \big/ \{1 - \pi(l_i)\}\right) = \beta_0 + \beta_l l_i,\ i = 1, 2, \cdots, k, \qquad (2)$$

Where logit link function equates to the linear predictor $\eta(l_i) = \beta_0 + \beta_l\, l_i =$, i = 1, 2, ..., k, which forms the systematic component of a generalised linear model. So this model does not have the structural problem and the linear probability model can be written as:

$$\text{logit}\,\pi(l_i) = \beta_0 + \beta_l\, l_i + e_i = \eta(l_i) + e_i, i = 1, 2, \cdots, k, \qquad (3)$$

where e_i's are error terms distributed independently and identically with constant variance.

Estimation of parameters

The unknown parameters, $\beta = (\beta_0,\ \beta_l)'$ of Equation (3) are estimated through maximum likelihood method by deriving the maximum likelihood equation from the probability distribution of the dependent variable Y. The details of estimation procedure are given in Agresti (2002). The normal equations of log-likelihood function with respect to β_0 and β_l are obtained as:

$$\frac{\partial}{\partial \beta_0} \ell(\beta \mid y_{l_i}) = \sum_{i=1}^{k} \left\{ y_{l_i} - n_{l_i} \left(\exp(\beta_0 + \beta_l l_i) \Big/ \{1 + \exp(\beta_0 + \beta_l l_i)\} \right) \right\} = 0,$$

$$= \sum_{i=1}^{k} \left\{ y_{l_i} - n_{l_i}\, \pi(l_i) \right\} = 0. \qquad (4)$$

and

$$\frac{\partial}{\partial \beta_l} \ell(\beta \mid y_{l_i}) = \sum_{i=1}^{k} \left\{ y_{l_i} l_i - n_{l_i} \left(\exp(\beta_0 + \beta_l l_i) \Big/ \{1 + \exp(\beta_0 + \beta_l l_i)\} \right) l_i \right\} = 0,$$

$$= \sum_{i=1}^{k} \left\{ y_{l_i} l_i - n_{l_i}\, \pi(l_i) l_i \right\} = 0. \qquad (5)$$

The above equation depends on the binomial counts only through the sufficient statistics, $\sum_{i=1}^{k} y_{l_i} l_i$

The likelihood equations appear in nonlinear form. Therefore, Newton-Raphson iterative method was used for solving nonlinear equations to get the maximum likelihood estimates for β_0 and β_l. Now let $\hat{\beta}_0$ and $\hat{\beta}_l$ be the maximum likelihood estimates of β_0 and β_l, respectively. Then $\hat{\pi}(l_i)$ is the maximum likelihood estimate of $\pi(l_i)$, where,

$$\hat{\pi}(l_i) = \left(\exp(\hat{\beta}_0 + \hat{\beta}_l l_i) \Big/ \{1 + \exp(\hat{\beta}_0 + \hat{\beta}_l l_i)\} \right), i = 1, 2, \cdots, k. \qquad (6)$$

Now, the likelihood equations (4) and (5) have the form:

$$\boldsymbol{l'y} = \boldsymbol{l'\hat{\mu}},\ \text{where}\ \hat{\mu}_{l_i} = n_{l_i}\ \hat{\pi}(l_i), \qquad (7)$$

This equation illustrates that likelihood equations equate the sufficient statistics to the estimates of their expected values. The expected number of fishes in the codend using maximum likelihood estimates of β_0 and β_l equals the observed number of fishes caught in the codend.

Estimation of selectivity parameters

The estimated parameters of the logistic model are then used to compute L_{50}, L_{25}, L_{75} and SR, which are defined as (Millar and Fryer, 1999):

$$\left.\begin{aligned} L_{25} &= \frac{-\hat{\beta}_0 - \log(3)}{\hat{\beta}_l} \\ L_{50} &= \frac{-\hat{\beta}_0}{\hat{\beta}_l} \\ L_{75} &= \frac{-\hat{\beta}_0 + \log(3)}{\hat{\beta}_l} \end{aligned}\right\} \quad (8)$$

and

$$SR = L_{75} - L_{25} = \frac{2\log(3)}{\hat{\beta}_l} \quad (9)$$

Selection factor (Millar and Fryer, 1999) is another term often used in selectivity studies to obtain the optimum mesh size and it can be defined as:

$$\text{Selection Factor} = \frac{\text{50\% retention length}}{\text{Mesh size}} = \frac{L_{50}}{\text{Mesh size}} \quad (10)$$

The optimum mesh size is obtained by dividing length at first sexual maturity (LFM50) by selection factor.

Generalised non-parametric regression for trawl-size selectivity curve

A generalised linear model assumes $\eta(l_i)$ in equation (3) has a parametric linear form, whereas a local polynomial generalised linear model (LPGLM) fits a local polynomial regression model within a smooth window. Therefore, we used LPGLM to approximate logit($\pi(l_i)$) as a function of length (l_i) in the trawl selectivity study. LPGLM was found to produce superior fit and overcome the boundary bias problem compared to other kernel based regression (Tibshirani and Hastie, 1987). The functional form of a LPGLM is defined as:

$$\text{logit}(\pi(l_i)) = \eta(l_i) + e_i, \quad i = 1,2, ..., k. \quad (11)$$

where $\eta(l_i)$ is assumed to be an unknown but smooth function that captures the trend/structure in the data as described by Cleveland (1979) and are error terms distributed independently without assuming any distribution. The non-parametric estimator is obtained as a linear combination of the values of the dependent variable with higher weight to observations closest to the point of prediction l_i.

Estimation of parameters

An approximation of local polynomial regression $\eta(l_i, l_i^*)$ of degree 'p', when the value of the predictor (l_i) is close to (l_i^*) is defined as:

$$\eta(l_i, l_i^*) \approx \beta_0(l_i^*) + \sum_{j=1}^{p} \beta_j(l_i^*)(l_i - l_i^*)^j \quad (12)$$

The local polynomial likelihood function can be written as:

$$L(\beta \mid y_{l_i}) = \prod_{i=1}^{k} \left(\frac{\pi(l_i)}{1-\pi(l_i)} \right)^{y_{l_i}} (1-\pi(l_i))^{n_{l_i}} \text{ for } y_{l_i} = 0,1,\cdots,n_{l_i} \quad (13)$$

Now using equations (1), (2) and (12), we can write equation (13) as:

$$L(\beta \mid y_{l_i}) = \prod_{i=1}^{k} \left(e^{y_{l_i} \eta(l_i, l_i^*)} \right) (1/\{1+e^{\eta(l_i, l_i^*)}\})^{-n_{l_i}} \quad (14)$$

Thus, local polynomial log-likelihood function of equation (14) in the neighborhoods of (l_i) is obtained as:

$$\ell(\beta) = \sum_{i=1}^{k} w(l_i) \left\{ y_{l_i} \eta(l_i, l_i^*) - n_{l_i} \log(1 + e^{\eta(l_i, l_i^*)}) \right\}, \quad (15)$$

where, $w(l_i) = w\left(\frac{l_i - l_i^*}{h}\right)$

is a kernel function and h is the bandwidth, which is based on the values of generalised cross validation (GCV) and we used tri-cube weight function (Takezawa, 2006) defined by:

$$w\left(\frac{l_i - l_i^*}{h}\right) = \begin{cases} \left[1 - \left(\frac{|l_i - l_i^*|}{h}\right)^3\right]^3 & \text{if } \left(\frac{|l_i - l_i^*|}{h}\right)^3 \le 1 \\ 0 & \text{if } \left(\frac{|l_i - l_i^*|}{h}\right)^3 > 1 \end{cases}$$

Take the first derivative of equation (15) with respect to $\hat{\beta}$ and then setting to zero, we get:

$$\frac{\partial}{\partial \beta_l} \ell(\boldsymbol{\beta} \mid y_{l_i}) = \sum_{i=1}^{k} \left\{ w(l_i) y_{l_i} \sum_{j=1}^{p} (l_i^*)(l_i - l_i^*)^j - w(l_i) n_{l_i} \left(\exp(\eta(l_i, l_i^*)) \Big/ \{1 + \exp(\eta(l_i, l_i^*))\} \right) \sum_{j=1}^{p} (l_i^*)(l_i - l_i^*)^j \right\} = 0,$$

$$= \sum_{i=1}^{k} \left\{ w(l_i) y_{l_i} \sum_{j=1}^{p} (l_i^*)(l_i - l_i^*)^j - w(l_i) n_{l_i} \pi(l_i) \sum_{j=1}^{p} (l_i^*)(l_i - l_i^*)^j \right\} = 0. \quad (16)$$

The parameter β is estimated locally in a smooth window by Newton-Raphson iterative method (Loader, 1999). The local polynomial maximum likelihood estimate of $\eta(l_i, l_i^*)$ is $\hat{\eta}(l_i, l_i^*)$, which gives the maximum likelihood estimate of $\pi(l_i)$ as:

$$\hat{\pi}(l_i) = \left(\exp(\hat{\eta}(l_i)) \big/ 1 + \exp(\hat{\eta}(l_i)) \right) \tag{17}$$

The local likelihood equation in 16 can be written in matrix notation as:

$$\boldsymbol{l}^{*\prime}\mathbf{W}\mathbf{y} = \boldsymbol{l}^{*\prime}\mathbf{W}\hat{\boldsymbol{\mu}} \tag{18}$$

where $\hat{\mu}_{l_i} = n_{l_i}\,\hat{\pi}(l_i)$ and 'W is the diagonal matrix with entries $w(l_i)$.

These likelihood equations equate sufficient statistics to the estimates of their expected values with weight function $w(l_i)$.

Estimation of selectivity parameters using genetic algorithm

Direct computation of selectivity parameters is not possible when selectivity curve is fitted by non-parametric method. Therefore, we have used a powerful optimisation technique known as genetic algorithm (GA) to compute/estimate L_{50} and selection range (SR) (Holland 1975, Goldberg 1989). In a selectivity study, the goal is to obtain the target value of L_{50} and selection range. Here, we take the functional form of objective function as estimated squared distance from target and it is defined as:

$$(\widehat{SDT}) = (\widehat{\pi(l_i)} - T)^2 \tag{19}$$

where T, the target value was taken as 0.25, 0.50 and 0.75 for L_{25}, L_{50} and L_{75}, respectively for estimating the selectivity parameters by minimising the objective function. This was estimated using Genetic Algorithm procedure available in the R package, version 3.0 (R Core Team, 2013) by suitably modifying the objective of the problem (Scrucca, 2013).

Goodness of fit statistics

Akaike information criterion (AIC) and Bayesian information criterion (BIC) were computed (Agresti, 2002). The deviance statistic was also computed for checking the goodness of fit of the model. A measure of discrepancy between observed and fitted values is the deviance statistic, which is given by:

$$D = 2\sum_{i=1}^{k}\left\{ y_{l_i}\log(y_{l_i}/\hat{\mu}_{l_i}) + (n_{l_i} - y_{l_i})\log\left(\frac{n_{l_i} - y_{l_i}}{n_{l_i} - \hat{\mu}_{l_i}}\right)\right\}, \tag{20}$$

where y_{l_i} is the observed counts and $\hat{\mu}_{l_i}$ is the fitted value for the l_i^{th} length classs. In a perfect fit. the ratio observed over expected is one and its logarithm is zero, so the deviance is zero.

Experimental trawl data on size-selectivity

As an illustration, we used a real time experimental data obtained from the codend selectivity study of the Dussumier's anchovy *Thryssa dussumieri*, an important trawl resource along Gujarat coast, India. Selection parameters for 40 mm diamond and square mesh codends by stacked haul method was used to estimate the selectivity parameters assuming error terms are independently distributed. The retention probabilities were derived using cover codend method as described by Pope *et al.* (1975). A cover with 10 mm mesh size and having length and breadth of about 2 times the codend was used to retain the escaped fish. The retention probability was computed as the proportion of fish caught in the codend for a given length class.

Results and discussion

Parametric regression approach

We fitted the logistic model given in the equation (1) to trawl experimental data of *T. dussumieri.* The model fitted well to the data with small values of AIC and BIC for both diamond and square mesh. The model was fitted by GENMOD procedure (Base SAS, 2011) of SAS 9.3. The parameters are estimated using equations (4) and (5) *via* Newton-Raphson iterative method. The estimated parameters along with standard error are given in the Table 1.

The computed values of L_{50} and selection factor for square mesh were 12.69 cm and 3.17 respectively; while the values were 8.89 cm and 2.22 for diamond mesh. The length at first sexual maturity (LFM50) of *T. dussumieri* is 7 cm. The optimum mesh size for square and diamond mesh was 22.06 and 31.51 mm, respectively. The observed *vs*. predicted retention probability is given in Fig.1 and 2, respectively for diamond and square mesh. Akaike information criterion (AIC) and Bayesian information criterion (BIC) were, 96.08 & 97.63 and 154.59 & 153.13 for diamond and square mesh, respectively. The deviance statistic D was 398 and 166 for diamond and square mesh, respectively.

The residual/error terms $e_{l_i} = \pi(l_i) - \hat{\pi}(l_i)$ were computed for the fitted model and are shown in Fig. 3 and 4, respectively for diamond and square mesh. It was observed from the figures that error terms exhibited a definite pattern indicating the dependency among observations. The independence of error terms was examined by run test at 5% level of significance and the null hypothesis was rejected ($p<0.05$). This might have happened due to the biased standard errors. Thus the estimated parameters will not be valid.

Table 1. Estimated parameters of parametric model

Type	Parameter	Estimate	SE	95% Confidence limits		AIC	BIC
Square mesh	β_0	-5.91*	0.510	-6.56	- 4.39	96.08	97.63
	β_1	0.58*	0.045	0.44	0.6467		
Diamond mesh	β_0	-2.40*	0.285	-3.01	-1.7819	154.59	153.13
	β_1	0.27*	0.030	0.21	0.3362		

*Indicates regression coefficients significant at 5% level of significance

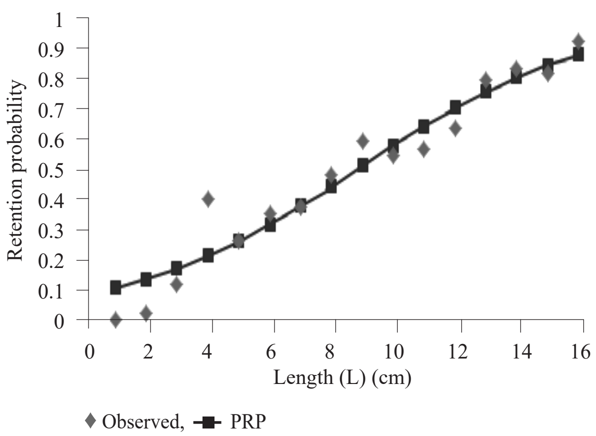

Fig. 1. Observed *vs*. predicted values of parametric model (PRP) of diamond mesh

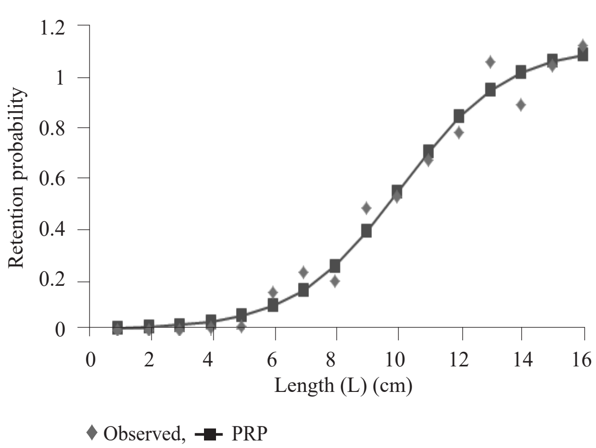

Fig. 2. Observed *vs*. predicted values of parametric model (PRP) of square mesh

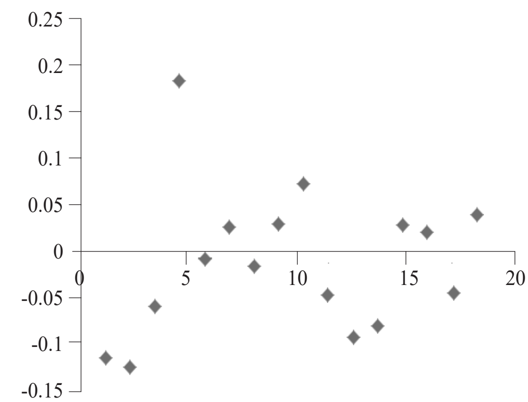

Fig. 3. Error terms of parametric model of diamond mesh

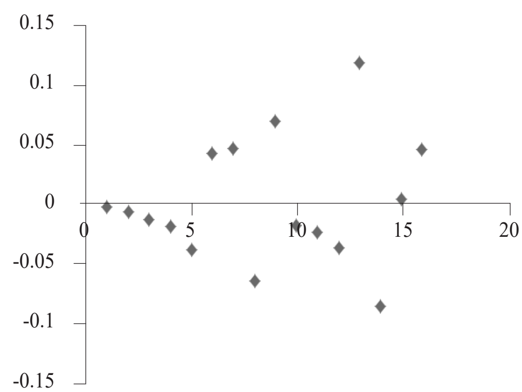

Fig. 4. Error terms of parametric model of square mesh

The dependence among the error terms might have happened due to misspecification of functional form of parametric regression. Thus, to overcome this bias problem, non-parametric regression was used in place of parametric regression to fit trawl selectivity curve and estimate the selectivity parameters.

Non-parametric regression approach

LPGLM of degree two was fitted to the experimental data described in Section 2.3 for both diamond and square mesh using 'locfit' method available in the statistical package R, version 3.0. The smoothing parameter for diamond mesh was 0.375 with 6 points in the local neighborhood and 0.37 with 5 points in the local neighborhood for square mesh. The AIC value for diamond mesh was -1.25 and -1.05 for square mesh. The deviance statistic D was 364 and 146 for diamond and square mesh, respectively. The non-parametric fit of retention probability of *T. dussumieri* in diamond and square mesh are given in Fig. 5 and 6, respectively.

The error terms of nonparametric fit of retention probability of *T. dussumieri* in diamond and square mesh are given in Fig. 7 and 8, respectively. There was some definite pattern observed in the residuals of the parametric

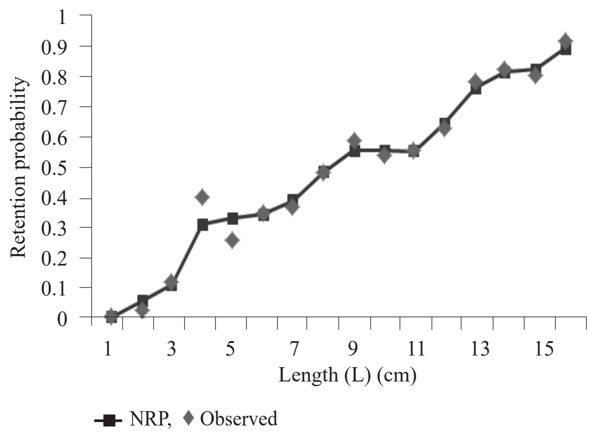

Fig. 5. Observed *vs* predicted values of non-parametric model (NPRP) of diamond mesh

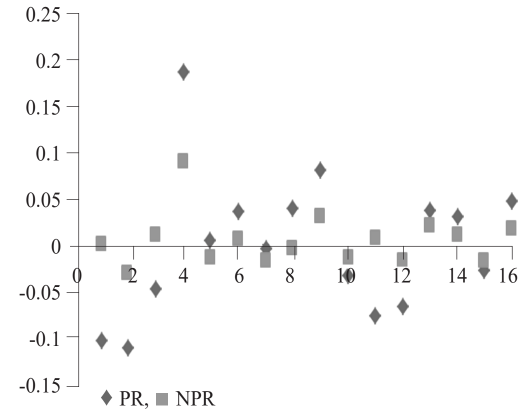

Fig. 7. Residual terms of parametric (PR) and non-parametric (NPR) model (diamond mesh)

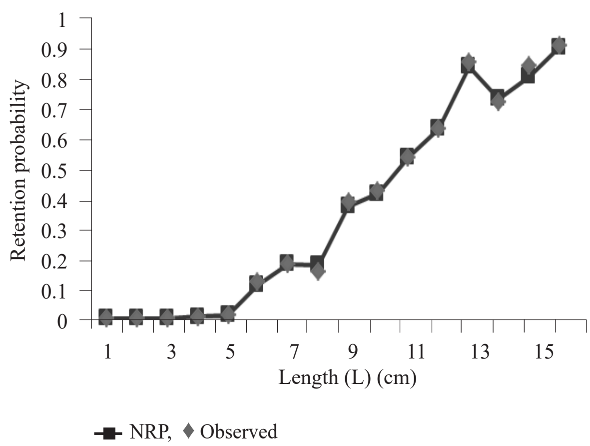

Fig. 6. Observed *vs* predicted values of non-parametric model (NPRP) of square mesh

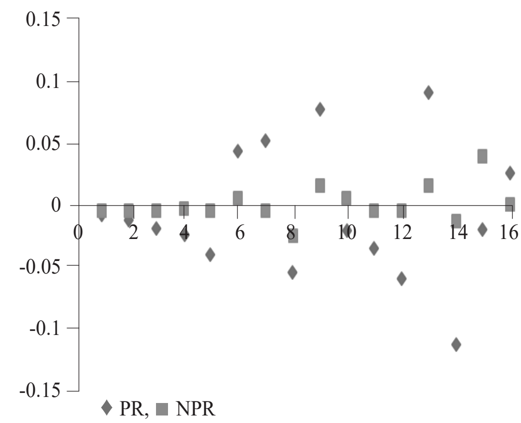

Fig. 8. Residual terms of parametric (PR) and non-parametric (NPR) model (square mesh)

model, whereas residuals obtained from the LPGLM fit do not exhibit any definite pattern and deviated very less from the ideal value of zero. The results of run test revealed that error terms of LPGLM were independently distributed ($p>0.05$) for both types of mesh. The computed deviance statistic was smaller in LPGLM fit than parametric fit for both diamond and square mesh. Thus, the non-parametric approach was found better than the parametric approach in trawl selectivity curve estimation.

We estimated the values of L_{50} and selection range using genetic algorithm based on the objective function given in equation (20). The estimated selectivity parameters are given in Table 2. Here, we found that the performance of nonparametric approach was superior than parametric approach. Thus, LPGLM provides a better fit to the selectivity curve and more accurate estimation of selectivity parameters. The estimated L_{50} and selection range values for diamond mesh were 8.4 and 9.0 cm, respectively and 10.4 and 3.85 cm, respectively for square mesh. The selection factor was 2.1 and 2.6 for diamond and square mesh, respectively. Optimum mesh size based on LFM50 computed from the nonparametric method was 33.3 and 27.0 mm for diamond and square mesh, respectively.

Validation of LPGLM approach

To confirm the superiority of non-parametric approach, we fitted LPGLM to another trawl selectivity data of *Upeneus moluccensis* and *Trichiurus lepturus* in 40 mm diamond mesh codends. The estimated selectivity parameters of the fitted models are given in Table 3.

The estimated L_{50} and selection range values for *U. moluccensis* were 7.8 and 4.4 cm, respectively and 49 and 25 cm, respectively for *T. lepturus*. The residuals obtained from the parametric and non-parametric fit are plotted and given in Fig. 9 and 10 for *U. moluccensis* and *T. lepturus*, respectively. The selection factor for *U. moluccensis* was 1.95 and was 12.25 for *T. lepturus*.

Table 2. Selectivity estimates of *T. dussumieri* in diamond and square mesh codends

Selectivity parameters	Estimated parameters	
Codend type	Diamond 40 mm	Square 40 mm
Total fishes caught in the codend	1239	518
Total fishes escaped to cover	1802	1301
L_{50} (cm)	8.4	10.4
Selection range (cm)	9.0	3.85
Selection factor	2.1	2.6
Length at first sexual maturity (LFM50) (mm)	70	70
Optimum mesh size based on LFM50 (mm)	33.3	27

Table 3. Selectivity estimates of *U. moluccensis* and *T. lepturus* in diamond mesh codend

	Species	
Selectivity parameters	*U. moluccensis*	*T. lepturus*
Codend type	Diamond 40 mm	Diamond 40 mm
Total individuals in the codend	261	220
Total individuals escaped to cover	322	260
L_{50} (cm)	7.8 cm	49 cm
Selection range (cm)	4.4cm	25 cm
Selection factor	1.95	12.25
Length at first sexual maturity (LFM50) (mm)	105 mm	560 mm
Optimum mesh size based on LFM50 (mm)	53.8 mm	45.7 mm

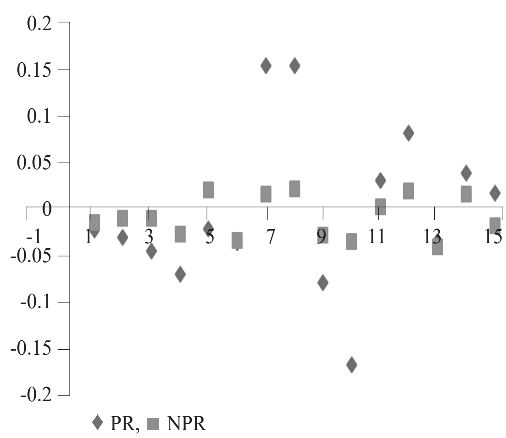

Fig. 9. Residual terms of parametric (PR) and non-parametric model (NPR) *(U. molucensis)*

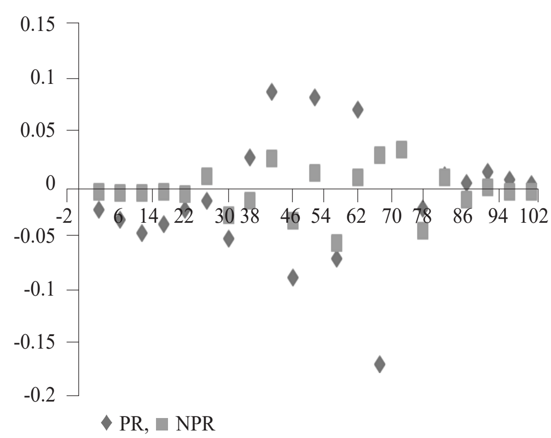

Fig. 10. Residual terms of parametric (PR) and non-parametric (NPR) model *(T. lepturus)*

The optimum mesh size based on LFM50 for *U. moluccensis* was 53.8 mm and 45.7 mm for *T. lepturus*. It is observed from the residual plots of LPGLM that non-parametric approach of trawl selectivity curve fitting is best when compared to parametric fit. The results of run test revealed that error terms are independently distributed ($p > 0.05$). The computed values of deviance statistic D for *U. moluccensis* and *T. lepturus* in diamond mesh codends were 58 and 282, respectively for parametric fit and 50 and 260, respectively for LPGLM fit. The deviance statistic of LPGLM fit was found to be small compared to parametric fit. Thus, nonparametric fit was found to be best compared to parametric fit.

Parametric and nonparametric approaches have been employed to fit trawl selectivity curve. Generalised linear models (GLM) and local polynomial generalised linear model (LPGLM) of order 2 was fitted to two types of real time trawl selectivity data obtained from the codend selectivity study of *T. dussumieri*, an important trawl resource along Gujarat coast for 40 mm diamond and square mesh codends by covered codend method. The model performance statistics were computed and compared with the parametric approach, and it was found that nonparametric approach is superior to parametric approach. Genetic algorithm was used to estimate the trawl selectivity parameters by minimising the objective function, the estimated squared distance from target. The estimated L_{50} and selection range values for diamond mesh were 8.4 and 9.0 cm, respectively and 10.4 and 3.85 cm, respectively for square mesh. The selection factor was 2.1 and 2.6 for diamond and square mesh, respectively. Optimum mesh size based on LFM50 computed from the nonparametric method was 33.3 and 27.0 mm respectively for diamond and square mesh. Trawl selectivity data of *U. moluccensis* and *T. lepturus* were also fitted by nonparametric method to confirm the superiority over the parametric method and found to be best compared to parametric fit.

Acknowledgements

The authors thank the Director, ICAR-CIFT, Kochi for granting permission to carry out the work. The technical assistance rendered by the vessel crew of Veraval Research Centre of ICAR-CIFT, Gujarat, India is also acknowledged.

References

Agresti, A. 2002. *Categorical data analysis*. John Wiley and Sons. Inc., New Jersey, 721 pp.

BaseSAS 2011. *BaseSAS® 9.3 Procedures Guide*. SAS Institute Inc., Cary, NC, USA.

Cessie, S. Le and Houwelingen, J. C. V. 1994. *Logistic regression for correlated binary data. Appl. Stat.,* 43(1): 95-103.

Cleveland, W. S.1979. Robust locally weighted regression and smoothing scatterplots. *J. Am. Stat. Ass.,* 74: 829-836.

Collie, J. S., Hall, S. J., Kaiser, M. J. and Poiner, I. R. 2000. A quantitative analysis of fishing impacts on shelf-sea benthos. *J. Anim. Ecol.,* 69(5): 785-798.

Cook, R. 2003. The magnitude and impact of bycatch mortality by fishing gear. In: *Responsible fisheries in the marine ecosystem*. FAO and CABI Publishing, p. 219-233.

Goldberg, D. E.1989. *Genetic algorithms in search optimisation and machine learning*. Addison-Wesley, Reading, MA, 403 pp.

Hall, M. A., Alverson, D. L. and Metuzals, K. I. 2000. Bycatch: problems and solutions. *Mar. Poll. Bull.,* 41(1-6): 204-219.

Holland, J. 1975. *Adaptation in natural and artificial systems*. University of Michigan Press, Ann Arbor.

Loader, C. 1999. *Local regression and likelihood*. Springer, New York, 305 pp.

McCullagh, P. and Nelder, J. A. 1989. *Generalised linear models,* 2nd edn. Chapman and Hall, London.

Millar, R. B. and Fryer, R. J. 1999. Estimating the size-selection curves of towed gears, traps, nets and hooks. *Rev. Fish Biol. Fish.,* 9(1): 89-116.

Millar, R. B. and Walsh, S. J. 1992. Analysis of trawl selectivity studies with an application to trouser trawls. *Fish. Res.,* 13: 205-220.

Pope, J. A., Margetts, A. R., Hamley, J. M. and Akyuz, E. F. 1975. Manual of methods for fish stock assessment, part III. Selectivity of fishing gear. *FAO Fisheries Technical Paper,* 41.

Prentice, R. L. 1998. Correlated binary regression with covariates specific to each binary observation. *Biometrics*, 44: 1033-1048.

R Core Team 2013. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Scrucca, L. 2013. GA: A package for genetic algorithm in R. *J. Stat. Softw*., 53(4): 1-37. doi:10.18637/jss.v053.i04.

Suuronen, P. 1995. Conservation of young fish by management of trawl selectivity. *Finnish Fish. Res.,* 15: 97-116.

Takezawa, K. 2006. *Introduction to nonparametric regression*. John Wiley and Sons Inc., New Jersey, 532 pp.

Thrush, S. F. and Dayton, P. K. 2002. Disturbance to marine benthic habitats by trawling and dredging: implications for marine biodiversity. *Ann. Rev. Ecol. Syst.,* 33(1): 449-473.

Tibshirani, R. J. and Hastie, T. J. 1987. Local likelihood estimation. *J. Am. Stat. Ass.,* 82: 559-567.

Wileman, D. A., Ferro, R. S. T., Fonteyne, R. and Millar, R. B. 1996. Manual of methods of measuring the selectivity of towed fishing gears. *ICES Cooperative Research Report,* 215. ICES, Copenhagen, 126 pp.

Date of receipt : 28.02.2018
Date of acceptance : 13.07.2018