NOT TO BE CITED WITHOUT PRIOR
REFERENCE TO THE AUTHOR(S)

Northwest Atlantic Fisheries Organization

Serial No. N4811 **NAFO SCR Doc. 03/5**

SCIENTIFIC COUNCIL MEETING – JUNE 2003

Testing Methods for Estimating the Factor Power Correction Obtained from the Comparative Fishing Trial: C/V *Playa de Menduíña Versus* R/V *Vizconde de Eza*

by

Diana González Troncoso and Xabier Paz

Instituto Español de Oceanografía,
P.O. Box 1552. Vigo, Spain.

Abstract

In May 2001, a comparative Fishing Trial was conducted by Spain between the old research vessel C/V *Playa de Menduíña* and the new research vessel R/V *Vizconde de Eza* in order to calibrate the new ship. The corresponding Factor Power Correction (FPC) and its confidence interval were calculated by six analytical methods proposed in the fisheries literature for American plaice: ratio of mean CPUE, linear regression model, generalized linear regression model by haul, generalized linear regression model by stratum, Kappenmas's ratio of scale parameters and a length conversion method. We present the results of these calculations and the transformed biomass from the old vessel data by all the methods.

The old vessel catches were in the order of three times more than the new vessel catches. The model proposed by Kappenman gave FPC values with the least variation, although his FPC estimation is lower than the rest of the models, so the transformed biomass is lower, too.

Introduction

When a new vessel and/or gear are used in a survey, changes are made in fishing power (Gulland, 1956). In this case, it was necessary to calculate the factor power correction (FPC), typically estimated by use of the catch-per-unit effort (CPUE) observations for the two vessels. From this calculation, we estimated the transformed CPUE series for the new conditions. That transformed series have at least two components of variation, one steaming from the usual sampling variance in the observations, and the other due to uncertainty in the estimate of FPC (Munro, 1998).

So, a good estimate of the FPC is critical, since to estimate the abundance of a species, a vessel CPUE is adjusted by multiplying by an estimate of a FPC. The abundance estimate is quite sensitive the estimate of the FPC used.

Several procedures for standardization for different vessel classes, estimating the FPC, have been proposed in the fisheries literature (Wilderbuer *et al.*, 1998; Salthaug and Godø, 2001).

This work analyses various conversion models for indices and adjustment procedures in order to observe which best adjust, i.e. which model entails the least variation, to the American plaice (*Hippoglossoides platessoides*) catches obtained from comparative experiments made in May 2001 by Spain between the C/V *Playa de Menduíña* and the R/V *Vizconde de Eza*.

Material and Methods

In May 2002, the R/V *Vizconde de Eza*, with a net type *Campelen*, replaced definitively the C/V *Playa de Menduíña*, with a net type *Pedreira*, as a research vessel to carry out the Spring Spanish Bottom Trawl *Platuxa* made since 1995 on NAFO Div. 3NO. In order to maintain the data series from the C/V *Playa de Menduíña*, in May 2001 a comparative fishing experiment between the two vessel was conducted in the surveyed area. The objective was to calculate the more precise Fishing Power Correction (FPC), which permit transform the C/V *Playa de Menduíña* data series into the R/V *Vizconde de Eza* data series minimizing the associated error.

In total, 92 paired hauls for a period of half an hour was made, two of them no valid. The vessels remained as close each to other as safety considerations permitted. For details, see Paz *et al.*, 2002. To obtain the FPC, six different models were considered, as follows:

The first model (Cochran, 1977) is the simplest. It is simply based on dividing the means of CPUE for each vessel to obtain an idea of how much a ship catches than the other ship. The formula is:

$$\hat{R} = \frac{\dfrac{\sum_{i=1}^{n} y_i}{n}}{\dfrac{\sum_{i=1}^{n} x_i}{n}} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i}$$

where x_i is the C/V *Playa de Menduíña* catch in the haul i, y_i is the R/V *Vizconde de Eza* catch in the haul i, and n is the total number of hauls. We can estimate the variance of this expression with the following formula (Cochran, 1977):

$$Var(\hat{R}) = \frac{\sum_{j=1}^{n} (y_j - \hat{R}x_j)^2}{n\overline{x}^2 (n-1)}$$

where $\overline{x}$ is the mean of the catches of the C/V *Playa de Menduíña*.

If the two vessels catch the same, the estimated value is 1. If it is lower than 1, it indicates that the C/V *Playa de Menduíña* catches more than the R/V *Vizconde de Eza*, and *vice versa*.

An asymptotic confidence interval can be estimated with a level of significance $\alpha = 0.05$ as follows:

$$\left(\hat{R} - z_{0.025}\sqrt{Var(\hat{R})}\, , \hat{R} + z_{0.025}\sqrt{Var(\hat{R})} \right)$$

where $z_{0.025} = 1.96$ is the 97.5th quantile of a standard normal distribution.

The second model was proposed by Robson (1966) as a multiplicative model to establish the relationship between the CPUEs of two vessels:

$$CPUE_{ij} = \exp(\mu + t_i + \varepsilon_{ij})$$

where t_i is the vessel effect, $i = 1, 2$, for haul j. Data are transformed by logarithms in order to obtain a linear expression:

$$\ln(CPUE_{ij}) = \mu + t_i + \varepsilon_{ij}$$

giving the following estimation of the FPC (Sissenwine and Bowman, 1978):

$$\hat{F}PC = \frac{\hat{C}PUE_2}{\hat{C}PUE_1} = \exp(2t(1 + 0.5s^2))$$

where parameters are estimated by linear regression, taking into consideration that $t_1 = t = -t_2$ and s^2 is the variance obtained in the estimate of t. In this case, the 95% confidence interval is calculated as follows:

$$\exp(2t - 2*1.96*s, 2t + 2*1.96*s)$$

In this model, we assume that the catch depends only on the ship effect, which is no realist. So, we introduce a third model based in the model of Robson, too, that introduce the haul effect, h_j:

$$CPUE_{ij} = \exp(\mu + t_i + h_j + \varepsilon_{ij})$$

We transform this equation by logarithms. The expressions of the linear model, the FPC estimation and the confidence interval are the same as in the last model. The only difference are that the linear model was adjusted by generalized linear regression assuming that $h_1 = -\sum_{j=2}^{n} h_j$.

A fourth model was proposed by Allen and Punsli (1984) and it is described in Hilborn and Walters (1992). This is a multiplicative model, as the last two models, but by stratum instead of by haul. The model is as follows:

$$U_{ti} = U_{11}\alpha_t \beta_j \varepsilon_{ti}$$

where U_{ti} is the catch of the vessel i in the stratum j (calculated as the sum of the catches of all hauls made in that stratum), α_t is the stratum factor and β_i is the vessel factor. Calculating this equation by generalized linear regression after transforming it with logarithms, the value of $\exp(\beta_2)$ would give us the efficiency of the research vessel in terms of the commercial vessel. In this case, the 95% confidence interval is:

$$\exp(\hat{\beta}_2 - 1.96s,\ \hat{\beta}_2 + 1.96s)$$

This model has fewer pairs of data than the rest as we restrict to strata instead of to hauls, so that the estimate would be more imprecise.

The fifth model was proposed by Kappenman (1992). It assumes that the two CPUE positive random variables have unknown but identical distributions, except possibly for the values of the scale parameters of the distributions. He estimates the FPC as the ratio of those scale parameters, $\hat{r}$. This is the value that minimizes the following function:

$$g(r)=\left(\frac{2n}{2n-1}\right)^2\left[\sum_{i=1}^{n}\left(x_i^d-\frac{\sum_{i=1}^{n}x_i^d+r^d\sum_{j=1}^{n}y_j^d}{2n}\right)^2+r^{2d}\sum_{j=1}^{n}\left(y_j^d-\frac{\sum_{j=1}^{n}y_j^d+\frac{\sum_{i=1}^{n}x_i^d}{r^d}}{2n}\right)^2\right]-$$

$$-\left(\frac{n}{n-1}\right)^2\sum_{i=1}^{n}\left(x_i^d-\frac{\sum_{i=1}^{n}x_i^d}{n}\right)^2-\left(\frac{n}{n-1}\right)^2\sum_{j=1}^{n}\left(y_j^d-\frac{\sum_{j=1}^{n}y_j^d}{n}\right)^2 r^{2d}$$

where d satisfies the following equation:

$$\frac{2n}{d}+u-\frac{1}{w}\left(\sum_{i=1}^{n}x_i^{2d}\ln x_i-v\sum_{i=1}^{n}x_i^d\ln x_i+q^2\sum_{i=1}^{n}y_i^{2d}\ln y_i-vq\sum_{i=1}^{n}y_i^d\ln y_i\right)=0$$

where:

$$u=\sum_{i=1}^{n}\ln x_i+\sum_{i=1}^{n}\ln y_i$$

$$p=\frac{1}{2}\sum_{i=1}^{n}y_j^{2d}-\frac{1}{4n}\left(\sum_{i=1}^{n}y_i^d\right)^2$$

$$v=\frac{\sum_{i=1}^{n}x_i^d+q\sum_{i=1}^{n}y_i^d}{2n}$$

$$w=\frac{1}{2n}\left(\sum_{i=1}^{n}(x_i^d-v)^2+\sum_{i=1}^{n}(qy_i^d-v)^2\right)$$

$$t=\frac{1}{4n}\left(\sum_{i=1}^{n}x_i^d\right)^2-\frac{1}{2}\sum_{i=1}^{n}x_i^{2d}$$

$$q=\frac{\sqrt{-4pt}}{2p}$$

This was programmed in Visual Basic *regula falsi* in order to be able to calculate these equations. *Regula falsi* is an iterative method estimating roots of functions. To calculate $\hat{d}$, this method was applied directly and, to find $\hat{r}$, the derivate of the function was calculated as 0, this was found to be, in fact, a minimum.

The confidence interval for this case was estimated by Bootstrap (Efron, 1979). Two interval were adjusted, the normal and the percentile. Calculations were developed with a program designed in Visual Basic, performing 100 sub-samples by re-sampling with replacement. The formula of the intervals for a signification level of $\alpha = 0.05$ are as follows:

Normal: $\left(\hat{r}-\frac{s}{\sqrt{m}}1.96,\ \hat{r}+\frac{s}{\sqrt{m}}1.96\right)$

Percentile: $\left(\hat{r}-\frac{s}{\sqrt{m}}y_{0.975},\ \hat{r}+\frac{s}{\sqrt{m}}y_{0.025}\right)$

where m is the number of re-samplings, s is the standard deviation of the re-samplings and y_α is the α^{th} quantile of the Bootstrap statistics distribution.

The last model was proposed by Warren (1997) and applied to convert length distribution from C/V *Playa de Menduíña* series into R/V *Vizconde de Eza* series (Paz *et al.*, 2002). The model is as follows:

$$Ratio = \alpha l^{\beta} e^{\gamma l}$$

where:

$$Ratio = \frac{Campelen\ catch}{Pedreira\ catch}$$

l is the length

$\alpha,\ \beta,\ \gamma$ are the parameters to be estimated

For adjustment, the curve is transformed by logarithms and the parameters are calculated by linear regression. The confidence interval for a 95% confidence level is as follows:

$$\exp\left(\ln\bar{}\,Ratio - t_{0.025,n-p-1}s\sqrt{X^t\left(X^tX\right)^{-1}X},\ \ln\bar{}\,Ratio + t_{0.025,n-p-1}s\sqrt{X^t\left(X^tX\right)^{-1}X})\right)$$

where $t_{0.025,n-p-1}$ is the 97.5^{th} quantile of a Student's t distribution with $n-p-1$ degrees of freedom, n is the total number of data, p is the number of parameters (3 in this case), s is the standard deviation of the fit, and X is the regression matrix formed by the independent variables which, in this case, are $\ln(l)$ and l.

For this paper, we considerate American plaice catches in the calibration experience of May 2001. In total, there were 90 pairs of effective hauls. Only in one of this paired hauls the catches of the two vessels was zero, so all the data was included to adjust models 1, 2 and 3. In four hauls, the catch of some of the ships was cero, so it can't be used in Kappenman model. Also, the two ships prospected together 26 hauls; two of them had catch cero for some of the ships, so in model 4 we only included 24 paired data.

Results

The results of the adjustments with their respective confidence intervals are presented in Table 1. Figure 1 illustrates the corresponding values.

Model 6, unlike the other five ones, takes into account the length distributions. Calculations were made in a previous study based on data of 55 length classes (*Paz et al.*, 2002). Since the regression did a poor fit on extreme lengths, the following adjustment for four length classes was made:

For $l \leq 12$: FPC = 9

For $13 \leq l \leq 21$: FPC = 0.63

For $22 \leq l \leq 51$: FPC = exp(13.3892+0.1521 l -5.7222 $L\,\mathrm{n}(l)$)

For $52 \leq l$: FPC = 0.4

In order to compare this model with the other five, it was considerate de FPC corresponding to the modal length of American plaice, that it was 22 cm. The confidence interval was calculated for this length, too.

Kappenman model is the one which have least error which a normal interval, continued closely by the ratio model. Both models provide a very similar FPC estimation (Kappenman = 0.28343453, Ratio = 0.28106522). The following model in terms of precision, however, is the third model (generalized linear regression by haul), and its estimation is bigger than the last ones (0.34181895), although this is approximately in the same scale.

The worst is the fourth model, generalized linear regression by stratum. This is certainly due to the fact that the number of data used was far less, so there is less precision in the calculations.

In this result, it is evident that two models, ratios of CPUEs and Kappenman, obtain a lower estimation of the FPC than the other four models. So, we can do two classes of estimations. To represent each one, we choose an estimation for each class in terms of low variance. So, for represent the first class, the Kappenman's estimation was choosen, and for the second class we select the third model's estimation. With these two estimations, the original C/V *Playa de Menduíña* catches and their transformed estimations were represented in Table 2 and in Fig. 2 for period 1995-2000. For years 2001-2002, the original R/V *Vizconde de Eza* catches were shown.

With all the different models tested, similar values were obtained for American plaice FPC estimation. In all cases, the C/V *Playa de Menduíña* with the *Pedreira* net proved to be much more efficient than the R/V *Vizconde de Eza* with the *Campelen* net, in the order of some 3 times better. These results were foreseeable on the shape of the nets in terms of previous experiences (Warren, 1997; Walsh *et al.*, 2001; Román *et al.*, 2001; Paz *et al.*, 2002).

In this case, the Kappenman model results less imprecise and it permits to calculate the biomass with the lower error.

References

Allen, R. and R. Punsli (1984): Catch rates as indices of abundance of yellowtail tuna, *Thunnus Albaraces*, in the eastern Pacific Ocean. *Inter-Am. Trop. Tuna Comm. Bull.*, **18**(4).

Cochran, W.G. (1977): Sampling techniques. Wiley, New York: 428 p.

Efron, B. (1979): Bootstrap methods: another look at the jack-knife. *Ann. Stat.*, **7**: 1-26

Gulland, J. A.(1956): On the fishing effort in English demersal fisheries. *Minist. Agric. Fish. Food Invest. Ser II*, **20**(5): 41 p.

Hilborn, R. and C. J. Walters (1992): Quantitative fisheries stock assessment. Choice, dynamics and uncertainty. Chapman and Hall, New York: 570 p.

Kappenman, R. F. (1992): Robust estimation of the ratio of scale parameters for positive random variables. *Commun. Statist.-Therory Meth.*, **21**(10): 2983-2996.

Munro, P. (1998): A decision rule based on the mean square error for connecting relative fishing power differences in trawl survey data. *Fish. Bull.*, **96**: 538-546.

Paz, X., D. González Troncoso and P. Durán Muñoz (2002): Comparative exercise between the C/V *Playa de Menduíña* and the R/V *Vizconde de Eza* in the NAFO Divisions 3NO in May 2001. *NAFO SCR Doc.*, No. 5, Serial No. N4603, 25 p.

Robson, D. S. (1966): Estimation of the relative fishing power of individual ships. *Res. Bull. Int. Comm. N. W. Atl. Fish.*, **3**: 5-14

Román, E., X. Paz and p. Durán Muñoz (2001): Comparative fishing between the R/V *Wilfred Templeman* and the C/V *Playa de Menduíña* in the NAFO Divisions 3NO in May 2001. *NAFO SCR Doc.*, No. 69, Serial No. N4447, 14 p.

Salthaug, A. and O. R. Godø (2001): Standardization of commercial *CPUE Fish. Res*., **49**: 271-281.
Sissenwine, M. P. and E. W. Bowman (1978): An analysis of some factors affecting the catchability pf fish by bottom trawls. *ICNAF Res. Bull*., **13**: 81-97
Walsh, S. J., Xabier Paz and Pablo Durán (2001): A preliminary investigation of the efficiency of Canadian and Spanish survey trawls on the Southern Grand Bank. *NAFO SCR Doc*., No. 74, Serial No. N4453, 16 p.
Warren, W. G. (1997): Report on the comparative fishing trial between the *Gadus Atlantica* and *Teleost*. *NAFO Sci. Coun. Studies*, **29**: 81-92
Wilderbuer, T. K., R. F. Kappenman and D. R. Gunderson (1998): Analysis of fishing factor estimates from a trawl comparison experiment. *N. Am. J. Fish. Manage*, **18**: 11-18

Table 1. Results of the different models applied for American plaice. Corresponding correction factor values and 95% confidence intervals. N is the number of paired hauls from the comparative trawling experiment.

Method	N	FPC	CI (95%)	1/FPC	CI Length
Ratio of mean CPUE	90	0.28106522	0.2496-0.3122	3.5579	0.0622
LM	90	0.33824967	0.1924-0.6086	2.9564	0.4162
GLM by haul	90	0.34181895	0.2884-0.4059	2.9255	0.1175
GLM by stratum	24	0.37084652	0.1754-0.7839	2.6965	0.6085
Kappenman (normal)	86	0.28343453	0.2637-0.3105	3.5282	0.0468
Kappenman (percentile)	86	0.28343453	0.2065-0.4577	3.5282	0.2512
Length model	55	0.38614400	0.3295-0.4525	2.8681	0.1230

Table 2. Original and transformed biomass (tons) by the different methods for the C/V *Playa de Menduíña*. Data for the year 2002 are the original for the R/V *Vizconde de Eza*. Spanish Spring Survey on NAFO Div. 3NO: 1995-2002.

Year	1995	1996	1997	1998	1999	2000	2001	2002
Original	54183.18	119437.99	70198.54	227166.64	353800.26	492052.59	387012.38	-
Means Ratio	15229.01	33569.87	19730.37	63848.64	99440.95	138298.87	108775.72	69497.72
LM	18327.44	40399.86	23744.63	76839.04	119672.82	166436.63	130906.81	69497.72
GLM by haul	18520.84	40826.17	13995.19	77649.86	120935.63	168192.90	132288.16	69497.72
GLM by stratum	20093.64	44293.16	26032.88	84243.96	131205.59	182475.99	143522.19	69497.72
Kappenman	15357.38	33852.85	19896.69	64386.87	100279.21	139464.69	109692.67	69497.72
Length model	20922.51	46120.26	27106.74	87719.03	136617.85	190003.15	149442.51	69497.72

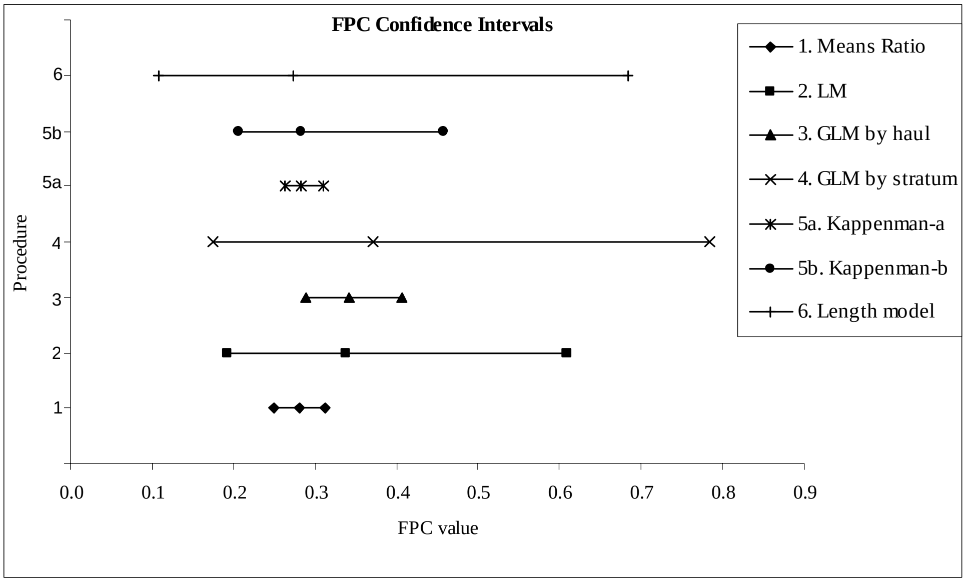

Fig. 1. Estimated correction factor value and confidence interval for American plaice obtained by six methods: Method 1: Ratio of mean CPUE; Method 2: Multiplicative model solved by linear model; Method 3: Multiplicative model solved by generalized linear model by haul, Method 4: Multiplicative model solved by generalized linear model by stratum; Method 5a: Normal Kappenman model; Method 5b: Percentile Kappenman model; Method 6: Length conversion model.

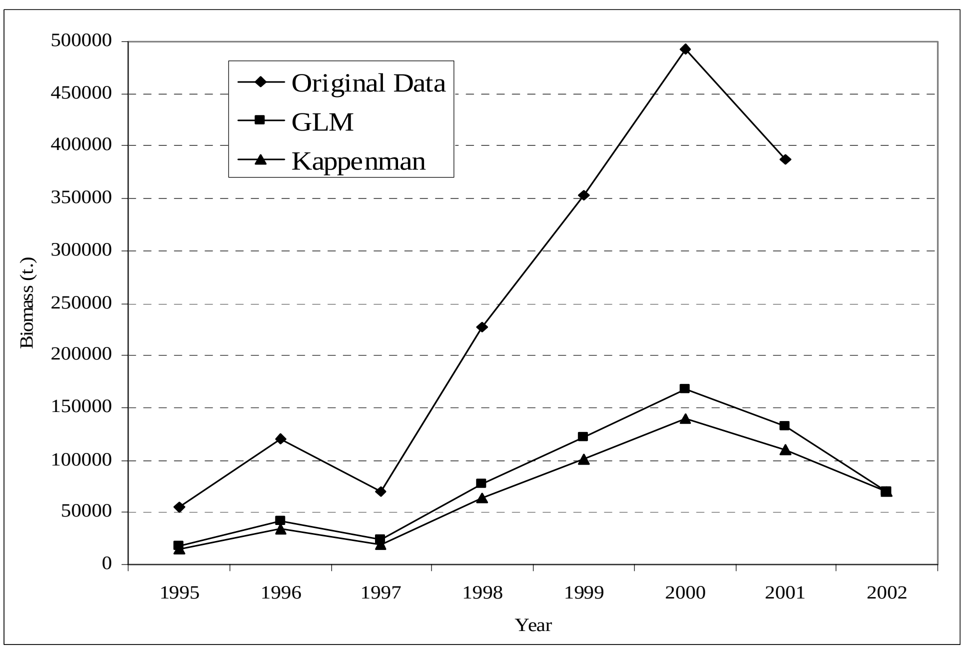

Fig. 2. American plaice biomass (tons) for converted (Kappenman model, GLM by haul model) and original data (from C/V *Playa de Menduíña*). Data for the year 2002 are the original for the R/V *Vizconde de Eza*. Spanish Spring Survey on NAFO Div. 3NO: 1995-2002.