Obtaining a standardised CPUE series for toothfish (*Dissostichus eleginoides*) in the Prince Edward Islands EEZ calibrated to incorporate both longline and trotline data over the period 1997-2013

Anabela Brandão and Doug S. Butterworth

Marine Resource Assessment & Management Group (MARAM)
Department of Mathematics and Applied Mathematics
University of Cape Town
Rondebosch, 7701, South Africa

October 2013

Abstract

The previous GLMM standardisation approach for PEI toothfish (Spanish) longline CPUE data is extended to include data for the 2012 and 2013 seasons and the same approach is applied to trotline CPUE data for the 2008–2013 period. CPUE data from a research program carried out in 2012 and 2013 in which longline and trotline sets were paired to within three nautical miles and a period of two weeks is analysed to obtain a calibration factor between longlines and trotlines. A model is then fitted to combine the two individual standardised CPUE series and the calibration factor to obtain a "calibrated" longline CPUE series (incorporating both longline and trotline information) and an estimate of the calibration factor. This indicates an increase of about 18% in standardised CPUE in 2013 compared to the preceding year, but this remains about 20% below the 2011 value. However it should be noted that data for 2013 are available only until July.

Introduction

The General Linear Mixed Model (GLMM) of Brandão and Butterworth (2011) has been applied to standardise the longline (Spanish) CPUE data for toothfish in the Prince Edward Islands EEZ for which data are now available until July 2013. The same form of GLMM has also been applied to the trotline CPUE data that are available since 2008.

A GLM analysis has been performed on paired longline and trotline CPUE data obtained from a research program carried out in 2012 and 2013 to obtain a calibration factor between the two types of gear. Results from these three analyses are then utilised jointly to obtain a calibrated longline CPUE series for the 1997 to 2013 period.

The Data

Longline commercial catch data (as kg green weight), and effort data (as total number of hooks set) are available for the period 1997 to July 2013, and a total of 7 630 sets are available for analyses (Table 1a). Trotline CPUE data are available for the 2008 to July 2013 period. The effort for a trotline is defined as:

$$\left(\frac{\text{Length of line}}{\text{Spacing of droppers}}\right)\times \text{Number of clusters per dropper.}$$

A total of 1 139 trotline sets (Table 1b) is available for analyses.

In 2012 and 2013 a research program was carried out in which longline and trotline sets had to meet certain criteria. In main, they had to be set within three nautical miles and within a period of two weeks of each other. A total of 127 pairs of such data is available for analyses.

Brandão and Butterworth (2012a) reported on some questions about the accuracy of the commercial CPUE that is available from different sources (such as the data used in previous CPUE analyses, the CCAMLR database and the original C2 forms and observer forms). In particular, a difference that was corrected in the data used in the analyses Brandão and Butterworth (2012a) is that previously some sets were accorded a zero catch but in the CCAMLR database they are recorded “NA” sets indicating that the set has no catch for some reason presumably unrelated to local abundance of toothfish. All these sets were omitted from their analyses, resulting in a marked different CPUE trend to that previously obtained (see Brandão and Butterworth, 2012b). A full data verification exercise has now been undertaken and the correct assignment of zero or “NA” catches to sets has been part of this exercise. Table 2 shows the comparison of total number of data entries per year (1197 to 2010) available for the longline GLMM analyses performed in 2011, 2012 (last year) and in 2013 (this year). The percentage reduction in observations from 2011 is also shown.

Methods

GLMM model to standardise CPUE data

Brandão and Butterworth (2011) proposed that a General Linear Mixed Model (GLMM) be used in the standardisation of the toothfish CPUE in which all interaction terms are considered as random effects because of the low number of longline sets (Table 1a) in the most recent years, as otherwise a large number of interaction terms have to be set using interpolation. In this paper GLMMs have been used to standardise the longline as well as the trotline commercial CPUE data.

The GLMM applied to the longline (and to trotline) CPUE data is of the form:

$$\ln\left(CPUE+\delta\right)=X\alpha+Z\beta+\varepsilon\,, \tag{2}$$

where

CPUE is the longline/trotline catch per unit effort,

δ is a small constant (10% of the average of all CPUE data values = 0.016 for longline and = 0.151 for trotline) added to the toothfish CPUE to allow for the occurrence of zero CPUE values,

α is the unknown vector of fixed effects parameters which includes: $\mu+\kappa_{vessel}+\omega_{year}+\gamma_{month}+\lambda_{area}$, where

μ is the intercept,

vessel is a factor with 9 levels associated with each of the vessels that have operated in the longline fishery (to an appreciable extent):

Aquatic Pioneer
Arctic Fox
El Shaddai

Eldfisk
Isla Graciosa
Koryo Maru 11
Ross Mar
South Princess
Suidor One

Only two vessels have operated trotlines: the *Koryo Maru 11* and the *El Shaddai.*

year is a factor with 17 levels associated with the years 1997–2013 for longlines or with 6 levels associated with the years 2008–2013 for trotlines,

month is a factor with 12 levels (January– December),

area is a factor with 4 levels associated with the four spatially distinct fishing areas:

A: 43–48°S latitude and 32–37°E longitude,

B: 43–45.3°S latitude and 37–40.3°E longitude,

C: 45.3–48°S latitude and 37–40.3°E longitude,

D: 43–48°S latitude and 40.3–43.3°E longitude,

X is the design matrix for the fixed effects,

β is the unknown vector of random effects parameters which includes the following interaction terms:

$\eta_{year\times area} + \theta_{year\times month} + \phi_{month\times area}$, where

year×area is the interaction between year and area (this allows for the possibility of different changes with time for the different areas),

year×month is the interaction between year and month,

month×area is the interaction between month and area,

Z is the design matrix for the random effects, and

ε is an error term assumed to be normally distributed and independent of the random effects.

It is assumed that both the random effects and the error term have zero mean, i.e. E(β) = E(ε) = 0, so that E($\ln(CPUE+\delta)$) = $\mathbf{X}\alpha$. We denote the variance-covariance matrix for the residual errors (ε) by **R** and the variance-covariance matrix for the random effects (β) by **G**. In the analyses of this paper it is assumed that the residual errors as well as the random effects are homoscedastic and are uncorrelated, so that both **R** and **G** are diagonal matrices given by:

$$\mathbf{R} = \sigma_{\varepsilon}^{2}\mathbf{I}$$
$$\mathbf{G} = \sigma_{\beta}^{2}\mathbf{I}$$

where **I** denotes an identity matrix. Thus, in the mixed model, the variance-covariance matrix (**V**) for the response variable is given by:

$$\mathrm{Cov}(\ln(CPUE + \delta)) = \mathbf{V} = \mathbf{ZGZ}^{\mathrm{T}} + \mathbf{R},$$

where $\mathbf{Z}^{\mathrm{T}}$ denotes the transpose of the matrix **Z**.

The estimation of the variance components (**R** and **G**), the fixed effects (α) and the random effects (β) parameters in GLMM requires two steps. First the variance components are estimated. Once estimates of **R** and **G** have been obtained, estimates for the fixed effects parameters (α) can be obtained as well as predictors for the random effects parameters (β). Variance component estimates are obtained by the method of residual maximum likelihood (REML) which produces unbiased estimates for the variance components as it takes the degrees of freedom used in estimating the fixed effects into account.

To provide additional insight a GLMM analysis was also performed by introducing an extra "gear" fixed factor to incorporate CPUE data from both longlines and trotlines so as to obtain an estimate of a "gear" effect. In this instance we are ignoring the pairing of some of the longline and trotline sets in 2012 and 2013 that were part of a research program for the purposes of getting a calibration factor for longlines and trotlines, so that the information content of these paired sets as regards the calibration factor is underweighted.

GLM to analyse research paired CPUE data from longlines and trotlines

The GLM considered allows for possible differences in "catchability" between the two types of gear (i.e. different multiplicative bias factors *g*) as well as for varying spatial and temporal distribution of toothfish density. The model is thus given by:

$$\ln\left(CPUE + \delta\right) = \mu + \alpha_{g} + \beta_{pair} + \varepsilon,$$

where

CPUE is the catch per unit effort for longlines or trotlines, where the effort for the different gears are described earlier in the paper,

δ is a small constant (10% of the average of the paired CPUE data values = 0.090) added to the toothfish CPUE to allow for the occurrence of zero CPUE values,

μ is the intercept (which incorporates the longline gear factor),

g is a factor with 2 levels associated with the type of gear (longline or trotline),

pair is a factor with 127 levels associated with set pairs between the Spanish longline and the trotline gear (capturing the different areas and times that the experiments took place, for each of which the underlying toothfish density may have been different), and

ε is the error term assumed to be normally distributed.

Since $\alpha_{longline}$ is incorporated in the intercept, the (log-transformed) calibration factor from this analysis, $K^* = \alpha_{trotline}$, with the analysis providing an estimate of K^* and of its associated variance $\sigma^2_{K^*}$.

During the research sets cetacean predation was observed to a much higher extent by the observers on one vessel than on the other vessel. However this information has not been included in the analyses because the information recorded is only whether cetaceans in the vicinity were observed to be feeding on the toothfish or not. This information is recorded only by the observers on the vessels so not every set has this information. Also, cetaceans could be feeding on the toothfish underwater and therefore not be seen to be feeding by the observer.

Model to calibrate the standardised longline CPUE series given the standardised trotline CPUE

The following negative log-likelihood function is minimised to estimate a calibrated longline CPUE series:

$$\begin{aligned}-\ln L = &\frac{1}{2}\ln\left|\mathbf{V}_L\right| + \frac{1}{2}\left(\ln \mathbf{CPUE}^L - \ln \mathbf{CPUE}^{cal}\right)^T \mathbf{V}_L^{-1}\left(\ln \mathbf{CPUE}^L - \ln \mathbf{CPUE}^{cal}\right) + \\ &\frac{1}{2}\ln\left|\mathbf{V}_T\right| + \frac{1}{2}\left(\ln \mathbf{CPUE}^T - \ln\mathbf{K} - \ln \mathbf{CPUE}^{cal}\right)^T \mathbf{V}_T^{-1}\left(\ln \mathbf{CPUE}^T - \ln\mathbf{K} - \ln \mathbf{CPUE}^{cal}\right) + \\ &\frac{1}{2\sigma^2_{K^*}}\left(\ln K^* - \ln K\right)^2\end{aligned}$$

where

$\mathbf{CPUE}^{L/T}$ is the vector of the predicted longline/trotline CPUE values obtained from fitting the GLMM described earlier to obtain standardised longline/trotline CPUE series,

$\mathbf{V}_{L/T}$ is the variance-covariance matrix of the predicted longline/trotline (log) CPUE series,

$\mathbf{CPUE}^{cal}$ is the vector of estimated longline CPUE which incorporates the calibrated trotline data,

$\mathbf{K}$ is the vector of the estimated calibration factor between longlines and trotlines (this is defined as a vector for the purposes of conducting vector/matrix calculations but it contains only one value, indicated as K in the last part of the equation above),

K^* is the calibration factor obtained from analysing the paired research CPUE data, and

$\sigma^2_{K^*}$ is the variance of the K^* parameter.

It might appear that the data from the paired longline/trotline sets are being used twice in this likelihood. Note however that the GLMM analyses inputs in the first two lines of the RHS take only the trend information in these data into account, whereas their information in regard to the method calibration is taken into account only by the term in the final line.

Results and Discussion

Table 3 and Figure 1 show the relative abundance indices for toothfish provided by the standardised commercial longline and trotline (calibrated to longline) CPUE series for the Prince Edward Islands EEZ as well as the longline CPUE series calibrated by the trotlines. There is a large difference between the 2011 index from the longline GLMM and the calibrated index. It should be noted, however, that there were only two longline sets in 2011 (see Table 1a), so that appropriately the calibrated index is very close to the estimate related to the trotline data for this year. Figure 2 reproduces the CPUE series of Figure 1 individually with their 95% confidence intervals.

Table 4 gives the parameter estimates and their 95% confidence intervals for the three ways to estimate the "gear" factor.

Figure 3 compares the calibrated longline CPUE trends from this year's analyses to that of the GLMM-standardised CPUE trends for the Spanish longline obtained in last year's analyses and that obtained in 2011.

Acknowledgements

Thanks to Deon Durholtz for providing and verifying the data and Rob Leslie of DAFF and Chris Heinecken of CAPFISH, for assistance with queries about the data upon which this paper is based.

Reference

Brandão, A. and Butterworth, D.S. 2011. General Linear Model and General Linear Mixed Model standardisation of the toothfish (*Dissostichus eleginoides*) CPUE in the Prince Edward Islands vicinity over the period 1997–2010. DAFF Branch Fisheries document: FISHERIES/2011/OCT/SWG-DEM/48.

Brandão, A. and Butterworth, D.S. 2012a. Obtaining a standardised CPUE series for toothfish (*Dissostichus eleginoides*) in the Prince Edward Islands EEZ calibrated to incorporate both longline and trotline data over the period 1997-2012. DAFF Branch Fisheries document: FISHERIES/2012/OCT/SWG-DEM/25.

Brandão, A. and Butterworth, D.S. 2012b. Comparing the current CPUE series to last year's. DAFF Branch Fisheries document: FISHERIES/2012/OCT/SWG-DEM/26.

Table 1a. The number of data entries per month and year available for the GLMM analysis to standardise the commercial Spanish **longline** toothfish CPUE series.

Year	Month												Total
	1	2	3	4	5	6	7	8	9	10	11	12	
1997	32	44		11	72	25	11	38			77	112	422
1998	135	223	215	150	15	35	118	93	89	228	63	81	1 445
1999	48	34	30	69	168	64	13	176	165	124	195	54	1 140
2000	148	183	137	170	134	64	158	139	152	198	100	47	1 630
2001		39	56	14	120	149	47	90	5	28	15	9	572
2002	5	39	71	15	11				34	70	63		308
2003		35	47		17	84	106		39	151	37		516
2004		15	49		45	114	25		5	54	58	19	384
2005		10	45	2					14	48	42		161
2006		20	47					7	43	32			149
2007		38	53	22	135	65	23	87	12	47	39		521
2008		28	40							26	13		107
2009				2						23	33		58
2010	2	41	39							1			83
2011				2									2
2012				7	12		24	5	14	8			70
2013		8	16	21	14		3						62

Table 1b. The number of data entries per month and year available for the GLMM analysis to standardise the commercial **trotline** toothfish CPUE series.

Year	Month												Total
	1	2	3	4	5	6	7	8	9	10	11	12	
2008			6							9	45	2	62
2009										26	29		55
2010		8	19	2					12	71	63	5	180
2011	29	2	50	44	30		13	21	83	16	40	15	343
2012				20	37		33	53	57	49			249
2013		23	48	44	39	35	61						250

Table 2. The total number of sets per year available for the longline GLMM analyses performed in 2011, last year and this year. The percentage reduction in observations (for reasons explained in the main text) from 2011 is also shown.

Year	2011 analyses	Last year's analyses		Current analyses	
		number	% reduction	number	% reduction
1997	488	472	3.3	422	13.5
1998	1455	1386	4.7	1445	0.7
1999	1347	1231	8.6	1140	15.4
2000	1692	1671	1.2	1630	3.7
2001	585	584	0.2	572	2.2
2002	253	319	-26.1	308	-21.7
2003	585	573	2.1	516	11.8
2004	446	417	6.5	384	13.9
2005	181	181	0.0	161	11.0
2006	150	137	8.7	149	0.7
2007	523	509	2.7	521	0.4
2008	113	89	21.2	107	5.3
2009	58	54	6.9	58	0.0
2010	83	73	12.0	83	0.0

Table 3. Relative abundance indices for toothfish provided by the standardised commercial CPUE series for the Prince Edward Islands EEZ for the Spanish longline and for the trotline fisheries, and the calibrated longline CPUE series.

Year	GLMM *CPUE*		Calibrated CPUE
	Longline fishery	Trotline fishery (calibrated to longline)	Longline index incorporating trotline data
1997	0.551		0.595
1998	0.194		0.197
1999	0.201		0.206
2000	0.239		0.241
2001	0.059		0.060
2002	0.129		0.131
2003	0.027		0.028
2004	0.145		0.150
2005	0.134		0.137
2006	0.073		0.074
2007	0.100		0.101
2008	0.128	0.077	0.115
2009	0.096	0.112	0.110
2010	0.100	0.141	0.119
2011	0.054	0.129	0.117
2012	0.055	0.088	0.080
2013	0.050	0.109	0.094

Table 4. Exponentiated "gear" factor estimates from a GLMM that combined Spanish longline and trotline CPUE data, from the paired longline-trotline research data and from the calibration analysis, together with 95% CI's shown in brackets and CVs.

	From GLMM with all data	From paired longline-trotline research data	Estimated from calibration analysis
Gear			
Longline	1	1	1
Trotline	13.82 (12.43; 15.36) CV = 0.055	7.198 (6.334; 8.179) CV = 0.067	7.539 (6.689; 8.498) CV = 0.063

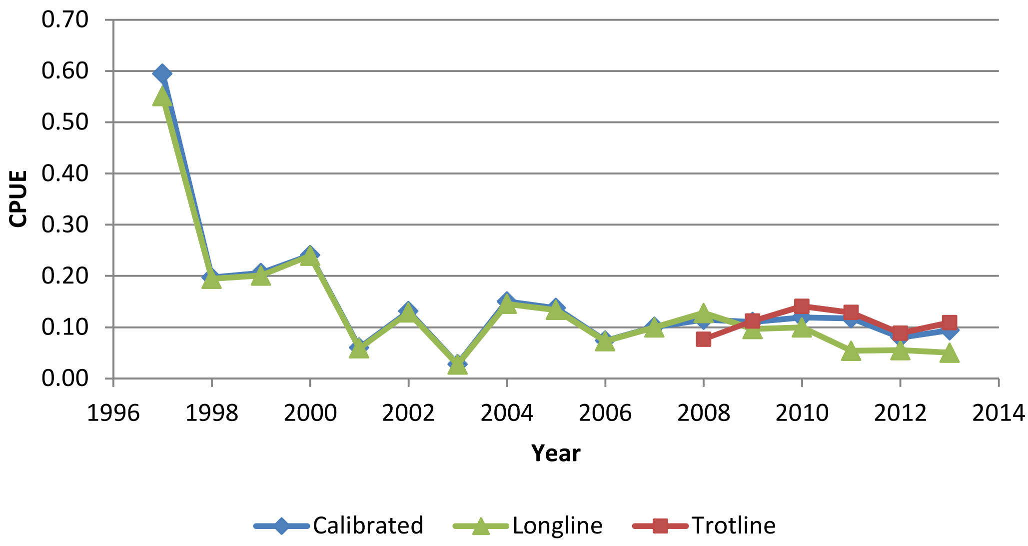

Figure 1. Calibrated longline CPUE trends as well as GLMM-standardised CPUE trends for the Spanish longline and trotline (calibrated to longline) toothfish fisheries for the Prince Edward Islands EEZ.

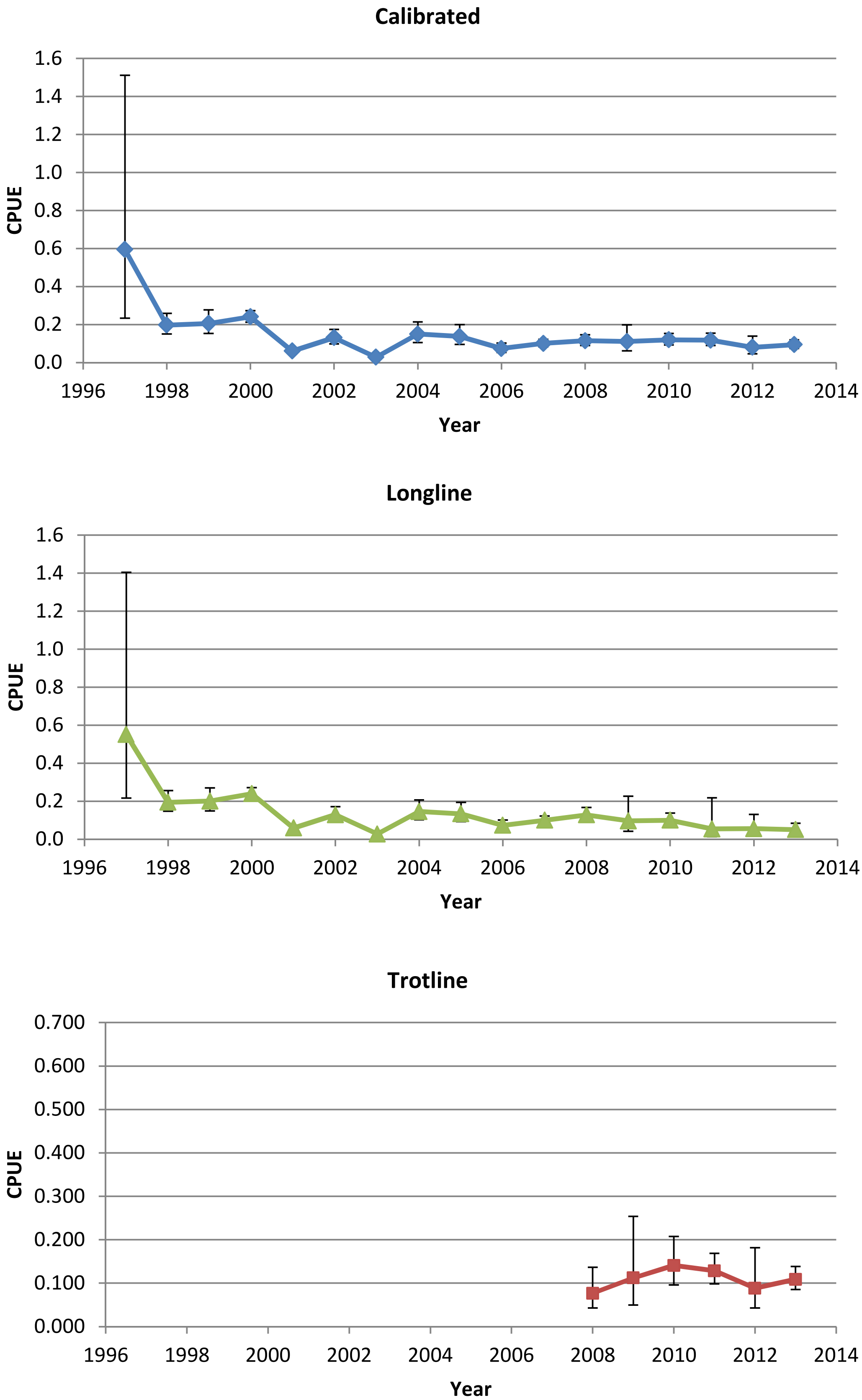

Figure 2. CPUE series of Figure 1 plotted individually with 95% CIs shown.

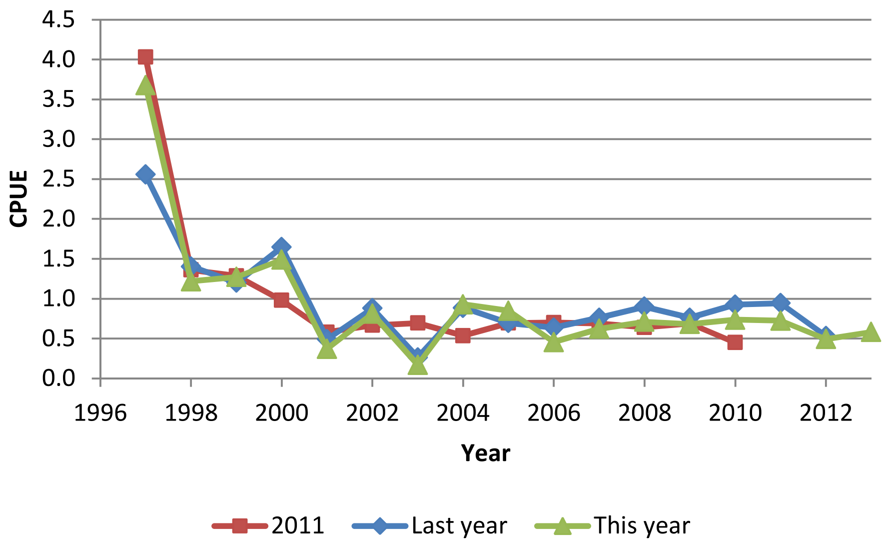

Figure 3. Calibrated longline CPUE trends from this year's analyses compared to the GLMM-standardised CPUE trends for the Spanish longline obtained in last year's analyses and that of 2011 (all are normalised to their mean over the 1997 to 2010 period).