REVISED RESULTS ON THE EFFECT OF OBSERVERS ON BOARD PELAGIC FLEETS

S. Somhlaba, A. Brandão and D.S. Butterworth
MARAM (Marine Resource Assessment and Management Group)
Department of Mathematics and Applied Mathematics
University of Cape Town, Rondebosch 7701,
South Africa

May 2005

INTRODUCTION

This is a revised report on pelagic data analyses which aim to quantify the amount of dumping, or confirm the absence thereof, that is happening in pelagic fisheries. This study includes the analyses of anchovy data which was not included in the previous papers (Somhlaba *et al*, 2004 a,b) and two extra years' data for sardine that were not included previously. Furthermore the effects of additional factors not included in the previous papers are investigated. Some diagnostic testing of residuals is also conducted for the chosen models.

DATA

The data available for this study span the period from 1999 to August 2004. The data were provided by MCM.

METHODOLOGY

Additional factors such as boat type, factory and fish size that were perceived to possibly have some influence on the catch rates of anchovy and sardine were investigated and final models including relevant factors have been chosen. Some outliers were removed before data analyses in order to minimize the impact of inappropriately influential points. Table 1 shows the number of observations for several categories of total number of hauls per trip performed for both species and including the perceived outliers. Table 2 shows similar results for the total number of hours spent per trip. Factors that were investigated are shown in Table 3. The four models detailed below were for both anchovy and sardine investigated in the same manner as in Somhlaba *et al* (2004 a). The models depicted as methods 1 to 4 below take into account all factors available; however the final models chosen do not necessarily include all these factors.

The model designated as "model 1" is for the case when the effort is measured in total number of hours spent per trip and the error distribution is assumed to be lognormal. The second model designated "model 2" is the same as model 1 except that the effort is measured in total number of hauls per trip. The third model designated as "model 3" assumes a Poisson error model with effort in hours as an offset. The fourth model is the same as the third except that effort is measured in hauls.

Model 1

$$\log_e(catch/hours)_{ijklmnpq} = \mu + \beta_i + \alpha H + \sigma T + \theta_j + M_k + Z_l + Y_m + G_n + F_p + S_q + \varepsilon_{ijklmnpq}$$

where:
μ is the intercept,
β_i is an observer factor with two levels { $i =$ observer present, observer not present}
θ_j is a catch category factor { $j =$ directed catch, by catch},
H is the total number of hauls per trip with α the associated estimable parameter,
M_k is the month effect with 12 levels { $k =$ January,…,December},
Z_l is the vessel length effect with twenty four levels ,
Y_m is the year effect with six levels from 1999 to July 2004 { $m = 1,..,6$},
G_n is the boat type effect with three levels { $n =$ bait, ordinary, steel},
T is the trip length in hours with σ the associated estimable parameter,
F_p is the factory or quota holder effect with levels fixed at ten { $p = 1,...,10$ }
S_q is the factor designating fish size with two levels { $q =$ adult, juvenile}, and
$\varepsilon_{ijklmnpq}$ is the error term assumed to be log normally distributed with mean zero and variance σ^2.

Model 2

$$\log_e(catch/hauls)_{ijklmnpq} = \mu + \beta_i + \alpha H + \sigma T + \theta_j + M_k + Z_l + Y_m + G_n + F_p + S_q + \varepsilon_{ijklmnpq}$$

This model is identical to “model 1”, except that the number of hauls replaces the duration of the trip (number of hours) as the effort unit.

Model 3

$$\log_e(catch)_{ijklmnp} = \log_e(T) + \beta_i + \alpha H + \log_e(T) + \theta_j + M_k + Z_l + Y_m + G_n + F_p + S_q + \varepsilon_{ijklmnpq}$$

where:
$\log_e(T)$ is an offset, where T is the duration of the trip in hours and
$\varepsilon_{ijklmnpq}$ is an error term with Poisson distribution.

Thus this model is like one shown in model 2, but the log normal distribution assumed for catch/hauls is replaced by a Poisson distribution for catch.

Model 3

$$\log_e(catch)_{ijklmnpq} = \mu + \beta_i + \log_e(H) + \sigma T + \theta_j + M_k + Z_l + Y_m + G_n + F_p + S_q + \varepsilon_{ijklmnpq}$$

where

$\log_e(H)$ is an offset, where

H is total number of hauls, and

$\varepsilon_{ijklmnpq}$ is an error term assumed to be Poisson distribution.

This is the same as for “model 3”, except that the number of hauls replaces trip duration as the effort unit.

RESULTS

The results are shown in Table 4 for each of the four models for both sardine and anchovy. For each model and both species, various sub-models were explored, with increasing number of explanatory factors included as explained in previous papers (Somhlaba *et al.* 2004a,b). The presence of an Observer factor was kept throughout as this was the aspect of particular interest. Other factors were added in turn by forward selection. In other words, at each step of the process the factor that led to the greatest reduction in deviance (-2 log likelihood) was retained, as long as this reduction was statistically significant (as determined by the 5 % χ^2 value for one degree of freedom). Table 4 lists estimates and standard errors of β, the estimated magnitude of the Observer factor, and provides an indication of whether or not it is significant.

DISCUSSION

The key result from the models is the estimate of β for the selected combination of model and factors included in each of the four cases. In each of these, one wants to know whether, after the effect of the other factors have been taken into account, there remains an indication of a sizable and significant observer effect. For anchovy all four models (Table 4) give consistent estimates for β which are above 10% in size and significant at the 5% level. For sardine the two models (1) and (3) give results above 10% in size and significant at the 5% level; these correspond to when catch rate is measured per hour for the lognormal model and when a Poisson error is assumed with time as an offset (Table 4). However, when effort is measured in hauls for sardine, the β estimate is around 1% for model 1 and -2% for model 4, with both not significant at the 5% level. Generally for sardine, when catch rate is measured per hauls, the observer effect is not significant.

Some diagnostics for all the models are given in Figure 1.1 and Figure 1.2. For both species a plot of standardized residual means versus effort and a plot of standardized residual standard deviations versus effort are given.

Model 1

For both anchovy and sardine the assumption of heteroscedasticity of residuals is satisfied under scenarios with trip duration less than one hundred hours and the mean of the residuals is close to zero (Figure 1.1a & 1.1b). However, this assumption cannot be checked when trip duration exceeds one hundred hours due to too few data.

Model 2

The assumption of heteroscedasticity seems to be a problem for this model for both anchovy and sardine (Figure 1.1d) when the number of hauls exceed 7. However the mean seems close to zero, indicating no model misspecification for hauls less than 7 (Figure 1.1 c).

Model 3

This model seems to be consistent with the heteroscedasticity assumption (Figure 1.2 e and Figure 1.2 f) for both species when hours are less that 200 or less. For more than 100 hours there is indication model misspecification.

Model 4

For anchovy this model's assumptions for both the mean and variance are satisfied for hauls less than 5 (Figure 1.2 g-h). The variability seems to increase thereafter, but this could be an artifact arising from outliers.

CONCLUSION

The presence of an observer on-board a commercial pelagic vessel has a significant effect on catch rate as confirmed by most of the models considered. Catch rate is significantly higher when an observer is on-board vessels. This conclusion seems to be very strong for anchovy based on the results from the models and their statistical significance.

The diagnostics suggest that results based upon hauls rather than hours are more reliable, as mean residual trends for hours indicate some model misspecification, but this aspect needs to be checked more carefully.

These results may constitute an indication that annually 10% or more fish is dumped from both sardine and anchovy directed fishing operations. This study is, however, still in progress; better data refinement and more factors, such as hauls and the duration of trips disaggregated by boat type need to be considered.

REFERENCES

Somhlaba, S., Brandão, A and Butterworth, D.S. 2004 a. A preliminary study of the observer effect on board the commercial pelagic vessels. MCM document WG/AUG2004/PEL/09a.

Somhlaba, S., Brandão, A and Butterworth, D.S. 2004b. Further results on the effect of observers on board the commercial pelagic vessels. MCM document WG/DEC2004/PEL06

TABLES

Table 1: The number of observations in each category of the total number of hauls per trip performed for both sardine and anchovy. Perceived outliers (taken to be values greater than 14 hauls) were removed when fitting the models.

Hauls	**Sardine**	**Anchovy**
	Number of observations	Number of observations
1-2	17767	10335
3-4	6527	6943
5-6	851	1164
7-8	87	111
9-10	2	1
11-12	11	4
13-14	1	2
15-16	0	0
17-18	0	0
19-20	0	0
21-22	5	5
23-24	0	0
25-26	1	0
27-28		0
29-30		1
31-32		1

Table 2: The number of observations in each category of the total number of hours per trip for both anchovy and sardine. Perceived outliers (taken to be values greater than 300 hours) were removed when fitting the models.

Class interval (hrs)	**Sardine**	**Anchovy**
	Number of observations	Number of observations
0-50	24712	18423
51-100	464	126
101-150	51	11
151-200	10	3
201-250	9	6
251-300	2	0
301-350	4	1
351-400	0	0
401-450	1	0
451-500		1

Table 3: Summary descriptions of model factors investigated. Further descriptions of each factor are given in the methodology section.

Factor	Description	Units
μ	Intercept, or the mean	Ton
β	Observer factor with two levels: either 0 or 1	-
θ	Catch category factor with two levels: either directed or by-catch	-
T	Total time taken for a trip	Hours
H	Total hauls made during the trip	-
M	Month factor, with 12 levels	-
Y	Year factor from 1999 to 2004, with 6 levels	-
G	Vessel group factor with three levels: bait, ordinary, steel	-
L	Vessel length with 24 levels	Metres
F	The factory factor with ten levels	-
CPUE	Catch rate, catch in tonnes divided by trip length	T/hour
CPUEST	Catch rate, catch in tonnes divided by total number of hauls	T/haul

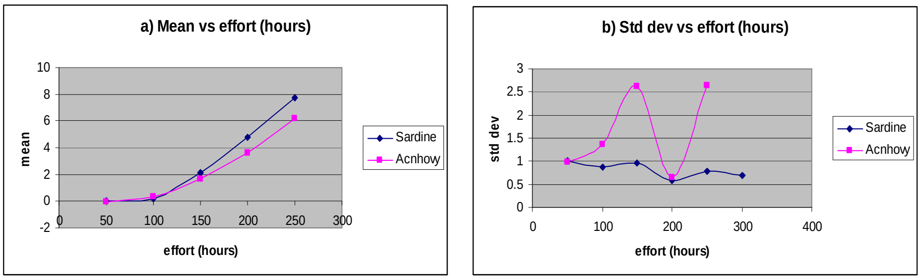
Note: Data for 2004 to August only.

Table 4: Estimates of the observer effect β (together with its standard error and p-value) for the selected model for each of the four models fitted. The results for the model with observer as the only factor are also given.

Species	Model	Error Model	Selected Model	β-Observer	std error β	p-value
Sardine		log(catch/hours)	**B**	-0.39	0.07	0.001
	1	log(catch/hours)	**β+θ+Z+T+F+M+Y**	0.12	0.04	0.001
		log(catch/hauls)	**B**	-0.68	0.07	0.001
	2	log(catch/hauls)	**β+θ+Z+H+S+G+M+Y**	0.01	0.04	0.815
		Catch(hours offset)	**β**	-0.37	0.05	0.001
	3	Catch(hours offset)	**β+θ+Z+H+S+G+Y**	0.12	0.03	0.001
		catch(hauls offset)	**β**	-0.37	0.05	0.001
	4	catch(hauls offset)	**β+θ+Z+H+Y**	-0.02	0.03	0.460
Anchovy		log(catch/hours)	**β**	0.44	0.06	0.001
	1	log(catch/hours)	**β+M+T+F+H+Y+θ**	0.11	0.03	0.006
		log(catch/hauls)	**β**	0.23	0.05	0.001
	2	log(catch/hauls)	**β+G+T+M+H+F+Y+θ**	0.10	0.03	0.006
		Catch(hours offset)	**β**	0.35	0.03	0.001
	3	Catch(hours offset)	**β+Y+S+H+Z+G+T**	0.13	0.02	0.001
		catch(hauls offset)	**β**	0.13	0.03	0.001
	4	catch(hauls offset)	**β+H+Y+S+T+G**	0.12	0.02	0.001

FIGURES

Log (CPUE) model 1



Log(CPUEST) model 2

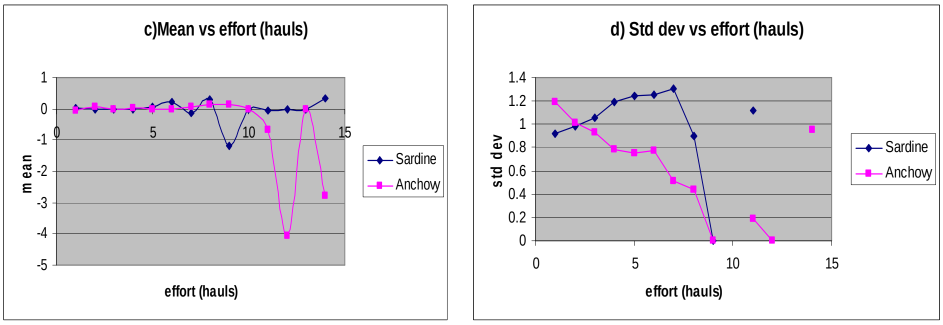
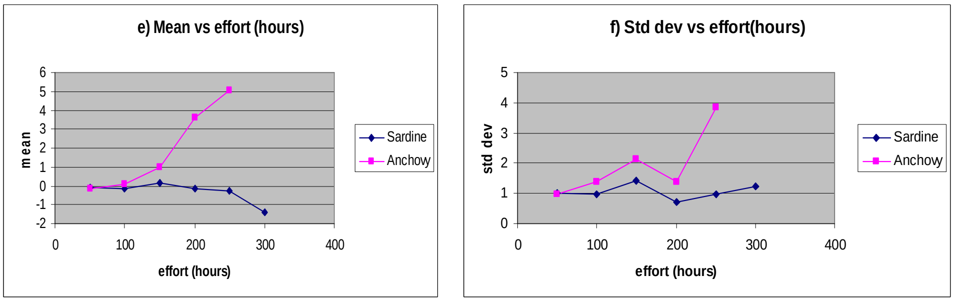


Figure 1.1: Summaries of some model diagnostics for the four models are shown (a-d). The plots give standardized residual means versus effort in hours and in hauls, and standardized residual standard deviations versus effort.

Catch(hours offset) model 3



Catch (hauls offset) model 4

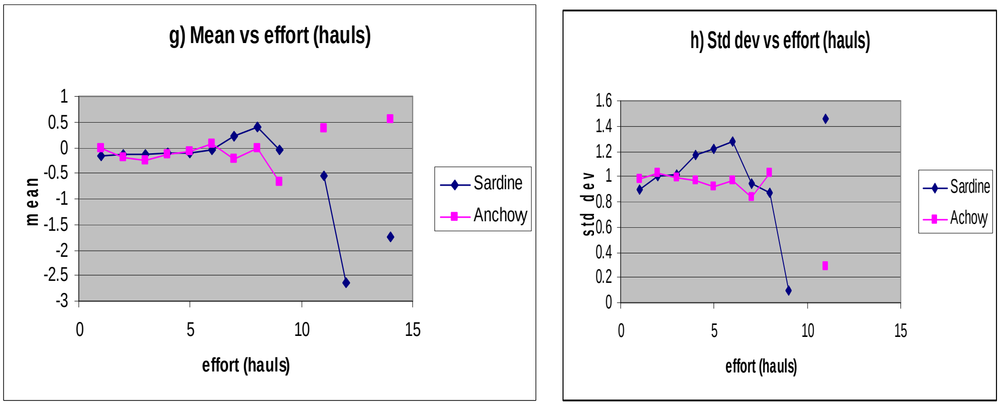


Figure 1.2: Summaries of some model diagnostics for the four models are shown (e-h). The plots give standardized residual means versus effort in hours and in hauls and standardized residual standard deviations versus effort.