CSIRO
MARINE RESEARCH

Appendix A

Aerial survey indices of abundance: comparison of estimates from line transect and "unit of spotting effort" survey approach

Jessica Farley
Sophie Bestley
Robert Campbell
Klaas Hartman

Prepared for the CCSBT 5th Meeting of the Stock Assessment Group (SAG5) and the 9th Meeting of the Scientific Committee (SC9)
6-11 and 13-16 September 2004, Seogwipo KAL Hotel in Jeju, Korea

Table of Contents

Appendix A: Preliminary results of standardised SAPUE

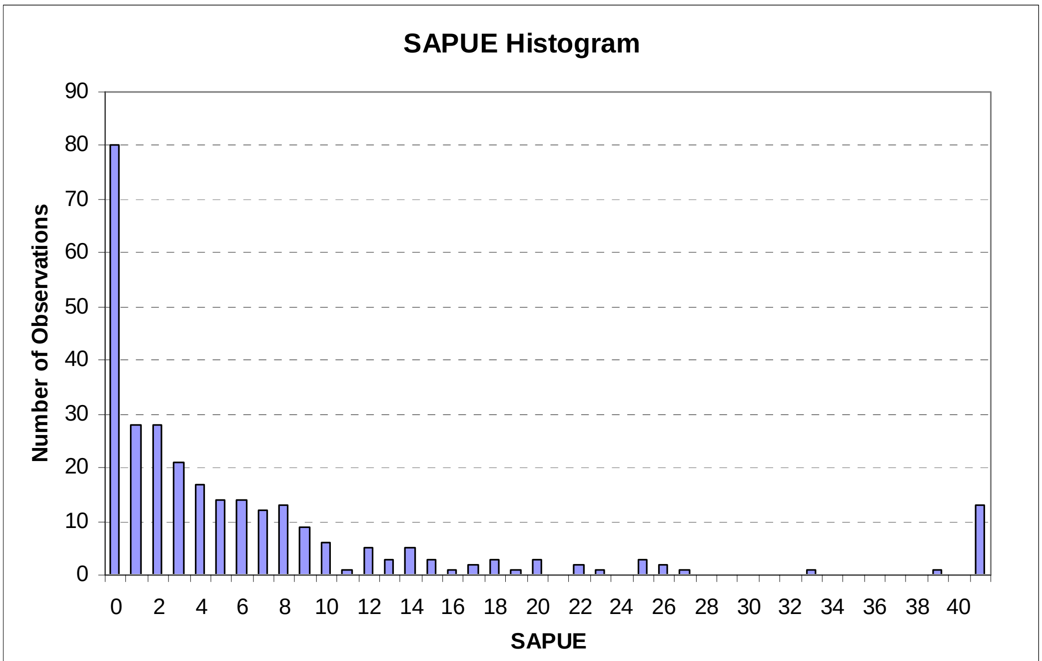
Introduction

This appendix presents preliminary standardisation results for the nominal survey abundance per unit effort (SAPUE) data from the 2002-3004 seasons. Results should be interpreted with caution because, as we show, they are sensitive to model assumptions and full or extensive analyses have not yet been undertaken. There are also assumptions about how the data are processed for use in the analyses. For example, results may be sensitive to how a 'flight' is defined, how fine a scale the data are analysed at, how the environmental variables are summarised for each 'observation. These issues have also not yet been fully explored.

Methods

For the standardisation, nominal SAPUE was calculated as biomass sighted per 10 nautical miles flown (unweighted) for the core fishing area. The histogram of the observed SAPUE for the 293 flights having complete data is shown in Figure A1. Eighty observations (27 percent of the total) have a SAPUE of zero, while a further 162 observations (55 percent) have a SAPUE<10. The remaining observations have a range of values and the distribution has a long tail (13 observations have a SAPUE>40). In summary, the distribution is characterised by a large number of zero –valued observations and a long tail.

Figure A1. Histogram of all SAPUE observations.



Due to the inflated number of zero SAPUE observations, the form of the above distribution suggests that the data might be best modelled as a two stage process: one stage being concerned with the pattern of occurrence of positive sightings, and the other stage with the mean size of the positive sighting. Furthermore, for both stages we can model the means as linear combinations of the factors likely to influence either the probability of a positive sighting and the size of a positive sighting. Once this is done, we can combine the means from the two distributions to give an overall mean sighting abundance.

A small example helps illustrate this approach. Consider a season for which there are n observations of the SAPUE, S_i. The average SAPUE can be expressed as follows:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} S_i = \frac{1}{n_S + n_F}\sum_{i=1}^{n_s} S_i = \frac{n_S}{n_S + n_F}\frac{1}{n_S}\sum_{i=1}^{n_s} S_i = p_S \mu_S \qquad (1)$$

where n_S is the number of positive or successful sightings (S_i>0), n_F is the number of zero or failed sightings (S_i =0), p_S is the proportion of positive sightings and μ_S is the average of the positive sightings. This result shows that the overall mean SAPUE can be expressed as the combination of the parameters from the distributions used to model the probability of a successful sighting and that used to model the non-zero sightings. A similar approach was used in the estimation of egg production based on plankton surveys (Pennington 1983, Pennington and Berrien 1984) and for estimating indices of fish abundance based on aerial spotter surveys (Lo et al 1992).

For the following analyses the data were restricted to the four months December through March when the majority of SBT are sighted. Data were also restricted to the four spotters that participated in all survey seasons. Two 'outlier' flights with very high SAPUE estimates (182 and 139) were also removed (90% of flights had SAPUE estimates <20 within the core area). A total of 269 fights were then available for analysis. Explanatory variables included in the GLM analyses were survey season, month, spotter, wind speed, swell height, cloud cover, spotting conditions and temperature. Actual wind speed and temperature values were recorded while swell height, cloud cover and spotting conditions were recorded as levels. If the values of the environmental variables were recorded more than once during a flight, the average of all recordings for that flight was used.

Stage 1: The Binomial Model for Positive Sightings

We model each observation as either a success (S_i >0) of a failure (S_i =0), with the probability of either expressed as follows:

$$\Pr(S_i > 0) = p_S \qquad \text{and} \qquad \Pr(S_i = 0) = 1 - p_S$$

Associated with each observation is a vector of covariates or explanatory variables X_j thought likely to influence the probability of a positive response. Furthermore, we assume that the dependence of p_S occurs through a linear combination $\eta = \sum \beta_j X_j$ of the explanatory variables. In order to ensure that $0 \le p_S \le 1$ we use the logit link function which takes the following form:

$$\eta = \log\left(\frac{p_S}{1 - p_S}\right)$$

The inverse of this relation gives the probability of a positive sighting as a function of the explanatory variables:

$$p_S = \frac{e^{\eta}}{1+e^{\eta}} = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ...)}{1+\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ...)} \qquad (2)$$

Given the manner in which the environmental variables were recorded (i.e. as continuous or categorical variables), two different models were adopted for the linear function of explanatory variables.

Model Bcov

Each environmental variable was fitted as a continuous covariate with an associated quadratic term. However, the value of each environmental variable for each flight was first normalised using the mean and standard deviation of each associated variable calculated across all flights. The following model was then fitted to the data for the 269 flights:

$$\eta = Intercept + Seaon_i + Month_j + Spotter_k + \sum_{m=1}^{5}\left[E_m + (E_m)^2\right] \qquad (3)$$

where E_m refers to the normalised values of each of the five environmental variables (Temperature, Wind strength, Swell height, Cloud cover and Spotting conditions).

Model Bcat

Swell height, cloud cover and spotting conditions were modelled as categorical variables. The number and definition of the categories used for each variable are described in Annex A1. Wind speed and temperature were modelled as in model *Bcov*. The function form of the linear model therefore takes on the following form:

$$\eta = Intercept + Seaon_i + Month_j + Spotter_k + \sum_{m=1}^{2}\left[E_m + (E_m)^2\right] + \sum_{n=1}^{3} Env_n \qquad (4)$$

where E_m, m=1,2 refers to the two environmental variables fitted as quadratic functions (Temperature and Wind strength) and Env_n, n=1,2,3 refers to the three environmental variables fitted as categorical factors (Swell height, Cloud cover and Spotting conditions).

Each of the above models was fitted using the SAS GENMOD procedure and an iterative backwards fitting procedure was adopted where after fitting the model the least significant term was omitted until only terms significant at the 0.05 level, using the Chi-squared test calculated for the Type 3 statistics, were retained. A standardised probability for a positive sighting was then calculated for each season against a standard set of model factors. This was obtained by first estimating the least-squares means value for each seasonal effect, *LSMEAN(Season$_i$)*, then using equation (2) to obtain the associated probability, p_S. (Note, due to the properties of the inverse link function, the values of p_S can be somewhat sensitive to the reference level used for each of the standardising factors. The LSMEAN estimates give values relative to the average over all factors in the model.)

Stage 2: The log-Normal Model for Positive Sightings

Having fitted the above model to the probability of obtaining a positive sighting, a separate model was fitted to the distribution of positive sightings. For this purpose a log-Normal model was adopted, such that the log(SAPUE) was assumed to have a normal distribution.

The data fitted to the model were limited to positive SAPUE sightings leaving a total of 203 observations. As before, two approaches were used to model the relationship between the dependent and the environmental variables.

Model LNcov

As with model *Bcov,* all variables were fitted as normalised covariates with associated quadratic terms:

$$\log(\mu_S) = Intercept + Seaon_i + Month_j + Spotter_k + \sum_{m=1}^{5}\left[E_m + (E_m)^2\right] \quad (6)$$

Model LNcat

As with model *Bcat*, swell height, cloud cover and spotting conditions were fitted as categorical variables. The number and definition of each level is described in Annex A1. Note, due to the smaller number of observations in the fitted data, the number of levels adopted for one of the variables was fewer than previously used (this was to ensure that the number of observations in each level was greater than 10).

$$\log(\mu_S) = Intercept + Seaon_i + Month_j + Company_k + \sum_{m=1}^{2}\left[E_m + (E_m)^2\right] + \sum_{n=1}^{3} Env_n \quad (7)$$

As before, the model was fitted using the SAS GENMOD procedure and an iterative backwards fitting procedure was adopted where after fitting the model the least significant term was omitted until only terms significant at the 0.05 level, using the associated F-test calculated for the Type 3 statistics, were retained. Finally, a standardised mean sighting rate for those flights which spotted SBT was then calculated for each season against a standard set of model factors. As before, the least-squares means value for each season, *LSM(Season$_i$)*, was first calculated then the following equation was used to convert these LSM estimates back to the nominal scale.

$$\mu_S(Season_i) = \exp(LSM(Season_i)) \quad (8)$$

Stage 3: Combined Annual Index

As explained in the initial section, the overall mean sighting rate can be expressed as a multiple of the probability of successfully sighting SBT and the mean sighting rate on those sets when SBT were seen. Using the standardised annual indices for each model, the standardised annual index for the overall mean sighting rate in the *i-th* season was calculated as follows:

$$Annual\ Mean\ Sighting\ Rate(season_i) = p_S(season_i).\mu_S(season_i) \quad (9)$$

Adopting the terminology used previously, the variance of the mean sighting rate can also be expressed as a combination of the parameters from the two distributions:

$$\sigma^2 = E(S^2) - [E(S)]^2 = \frac{1}{n}\sum_{i=1}^{n} S_i^2 - \mu^2 = \frac{n_S}{n_S + n_F}\frac{1}{n_s}\sum_{i=1}^{n_S} S_i^2 - (p_S\mu_S)^2$$

$$= \left[p_S E\left(S_S^2\right) - p_S \mu_S^2\right] + \left[p_S \mu_S^2 - p_S^2 \mu_S^2\right] = p_S \sigma_S^2 + p_S\left(1 - p_S\right)\mu_S^2$$

where σ_S^2 .= Var(μ_S). An alternative derivation of this expression is given in Annex A2. In order to estimate the variance parameter σ_S^2 for each season, the following Taylor approximation was used:

$$Var\left[\exp(\mu_S)\right] \approx \exp(2\mu_S).Var(\mu_S) \qquad (10)$$

where the corresponding standard errors associated with the LSMEAN estimates of μ_S for each season (c.f. eqn. (8)) were used to estimate σ_S^2.

Nominal Indices

Several nominal indices were calculated for comparison with the indices calculated from the GLM analyses. These indices were defined as follows:

1. $$Nominal\ SAPUE(s) = 10\frac{\sum_{i=1}^{N_s} Biomass\ Sighted_i}{\sum_{i=1}^{N_s} Distance\ Flown_i}$$

2. $$Mean\ SAPUE(s) = 10\frac{1}{N_S}\sum_{i=1}^{N_S}\frac{Biomass\ Sighted}{Distance\ Flown}_i$$

3. $$Delta\ Nominal\ SAPUE(s) = \Pr(Postive\ Sighting)_S .10\frac{\sum_{i=1}^{M_S} Biomass\ Sighted_i}{\sum_{i=1}^{M_S} Distance\ Flown_i}$$

4. $$Delta\ Mean\ SAPUE(s) = \Pr(Positve\ Sighting)_S .10\frac{1}{M_S}\sum_{i=1}^{M_S}\frac{Biomass\ Sighted_I}{Distance\ Flown}_i$$

where N_s is the total number of flights in season, s, M_s is the number of flights in season s with a positive SBT sighting and:

$$\Pr(Positive\ Sighting)_S = \frac{M_S}{N_S}$$

Results

i) Probability of Positive Sightings

The factors that were found to have a significant effect on the probability of a positive sighting of SBT for each of the two Binominal models are listed in Table A1. Apart from the effect of cloud cover, the results for the *Bcov* and the *Bcat* models are seen to be similar, with season, month, spotting condition and wind being significant effects in both models.

The seasonal index of the probability of the positive sighting for each model is shown in Table A2 and displayed in Figure A2. Note that some of the difference in the relative scale of each index is due to differences in the effects each index is standardized against. The *Nominal*, *Bcov* and *Bcat* indices all display similar trends, with the probability of a positive sighting declining each season. The GLM based results indicate a decrease in the probability of a positive sighting of SBT of 13-15% between 2002 and 2003 and of 17-21% between 2002 and 2004.

Finally, the relative effect of each significant model factor on the probability of a positive sighting of SBT for the *Bcov* and *Bcat* models are shown in Figure A3. The results are generally similar between the two models, with the probability of a sighting declining with increases in cloud cover and wind speed. However, the predicted effect of spotting conditions is somewhat different between the two models.

ii) Mean SAPUE for Positive Sightings

The factors that were found to have a significant effect on the mean sighting rate for those flights sighting SBT for the two log-normal models are listed in Table A3. Season, month, spotter, spotting conditions and wind speed were significant in both models, while temperature was significant in the *LNcov* model while both swell height and cloud cover were also significant in the *LNcat* model.

The seasonal index of the mean sighting rate for those flights sighting SBT for each model is shown in Table A4 and displayed in Figure A4. Again, differences in the scale of each index are due to differences in the effects each index is standardized against. All indices display a similar trend between the first and second seasons, though while the two nominal indices indicate a continued decline after this time (with a greater relative decline for the *Mean* index), the indices for the *LNco*v and *LNcat* models remain relatively constant. The GLM results indicate a decrease in the mean sighting rate of 37-48% between 2002 and 2003 and of 34-45% between 2002 and 2004.

Finally, the relative effect of each significant model factor on the mean sighting rate for those flights spotting SBT for the *LNcov* and *LNcat* models are shown in Figure A5. Again, the results are consistent across the two models, indicating an increase in the sighting rate with an increase in spotting conditions and temperature (not shown) and a decrease associated with increased wind speed. However, there is no discernable trend in the influence of either cloud cover or swell height predicted by the *LNcat* model.

Table A1. Summary of SAS Type III statistics for factors included in the models used to estimate the Probability (SBT sighting) on a given flight.

Model	Factor	DF	Chi-Square	Pr > Chi Sq	Sign
Bcov	Season	2	11.21	0.0037	***
	Month	3	8.64	0.0345	**
	Condition	1	2.694	0.1012	
	Condition (sq)	1	4.18	0.0409	**
	Wind speed	1	5.12	0.0236	**
	Cloud cover	1	5.20	0.0226	**
Bcat	Season	2	8.15	0.0170	**
	Month	3	7.86	0.0489	**
	Condition	3	9.62	0.0221	**
	Wind speed	1	4.65	0.0310	**

Table A2.Seasonal indices for the Prob(SBT sighting) based on each model.

Season	*Bcov*	*Bcat*	*Nominal*
2002	0.913	0.907	0.871
2003	0.779	0.792	0.750
2004	0.724	0.749	0.696

Figure A2. Seasonal indices for the Prob(SBT sighting) based on each model.

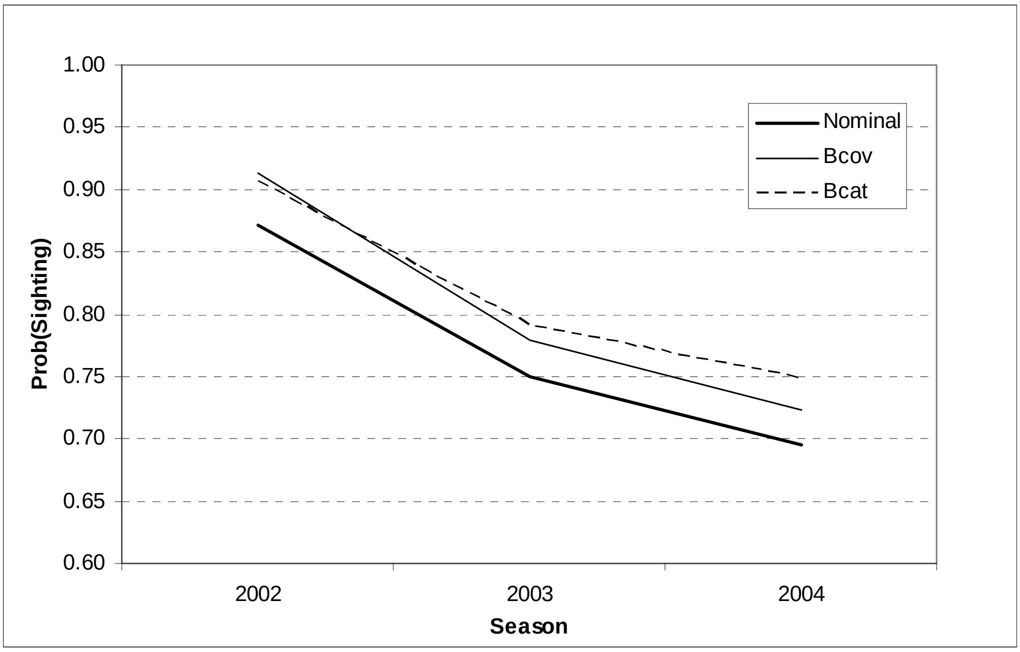


Figure A3. Effect of model factors on Prob(Positive Sighting)

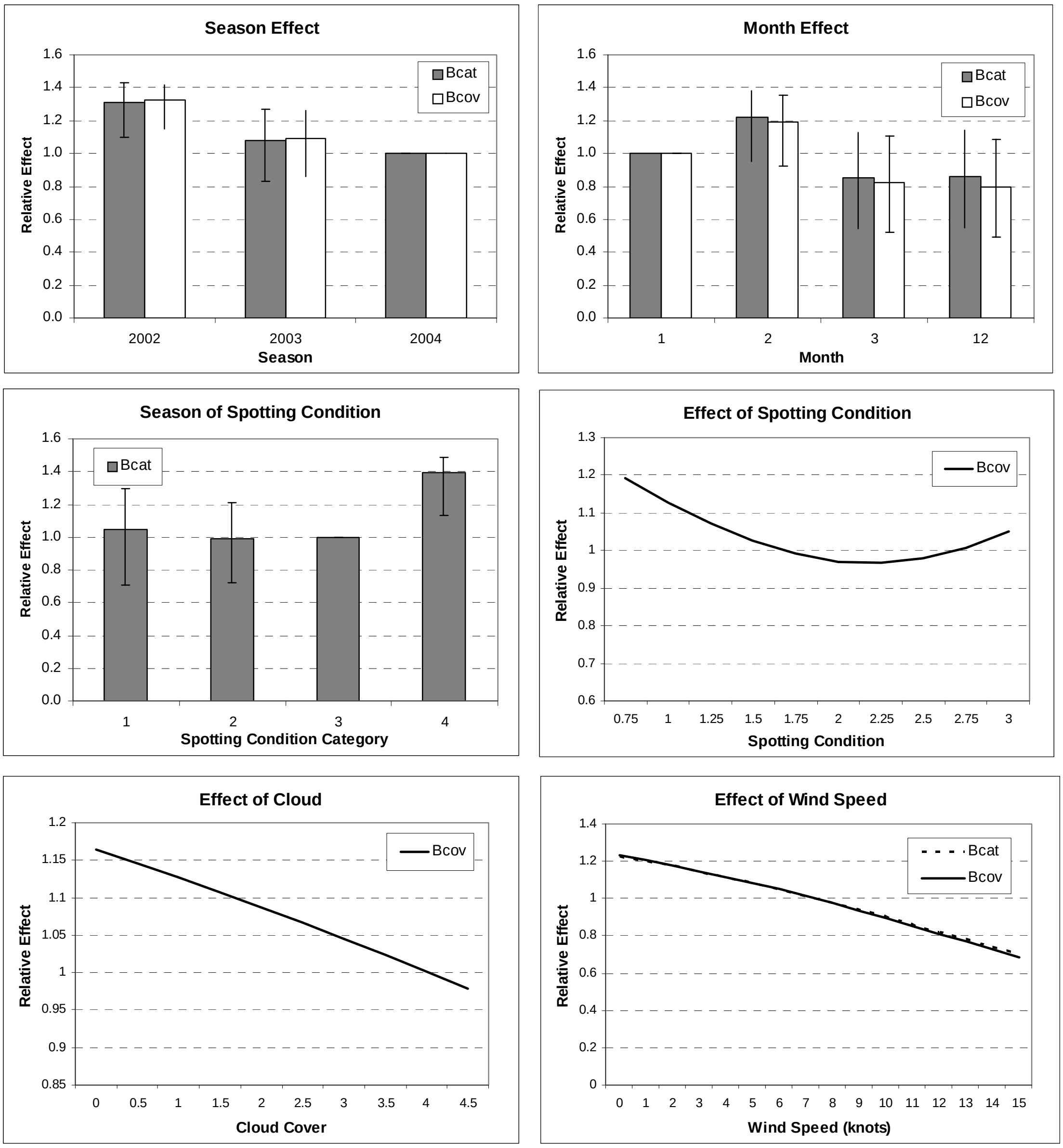


Table A3. Summary of SAS Type III statistics for factors included in the models used to estimate the mean SAPUE for positive sightings.

Model	Factor	DF	F value	Pr > F	Sign
LNcov	Season	2	8.32	0.0003	***
	Month	3	5.09	0.0021	***
	Spotter	3	16.97	<0.0001	***
	Condition	1	7.70	0.0061	***
	Wind speed	1	16.70	<0.0001	***
	Wind speed (sq)	1	12.32	0.0006	***
	Temperature	1	4.42	0.0368	**
LNcat	Season	2	4.35	0.0143	**
	Month	3	4.04	.0082	***
	Spotter	3	13.77	<0.0001	***
	Swell height	1	2.48	0.0454	**
	Condition	1	3.35	0.0203	**
	Cloud	1	2.42	0.0167	**
	Wind speed	1	14.26	0.0002	***
	Wind speed (sq)	1	7.3	0.0076	***

Table A4.Seasonal indices of the positive SAPUE for each model.

Season	*LNcov*	*LNcat*	*Nominall*	*Mean SAPUE*
2002	5.616	5.011	11.559	13.628
2003	2.910	3.134	7.151	8.973
2004	3.047	3.269	5.932	5.615

Figure A4. Seasonal indices of Prob(SBT sighting) based on each model.

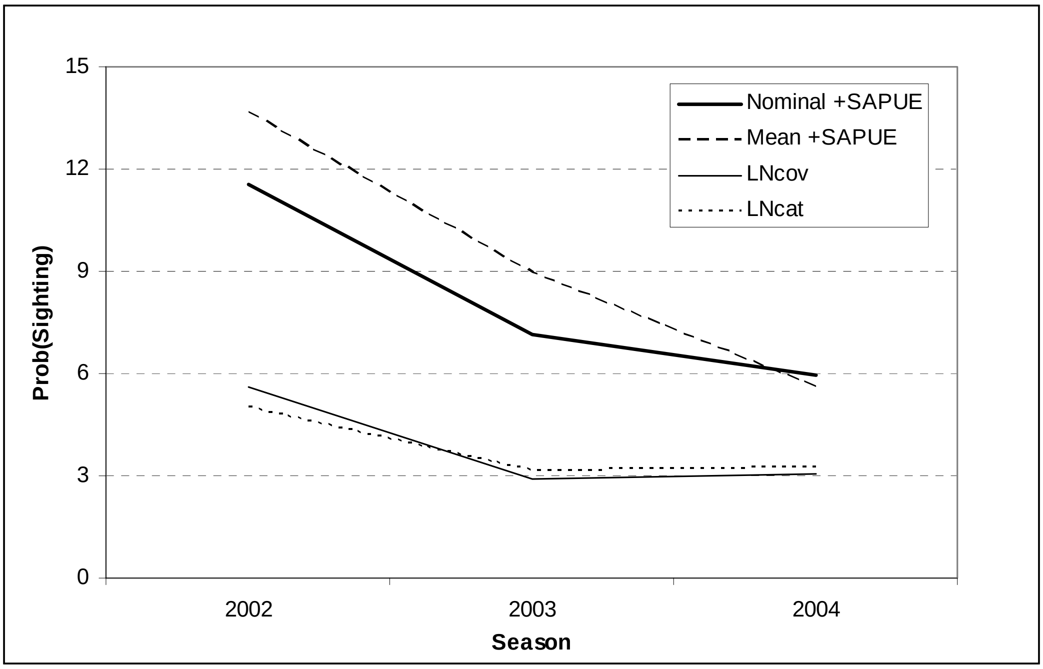


Figure A5. Relative effect of model factors on mean SAPUE.

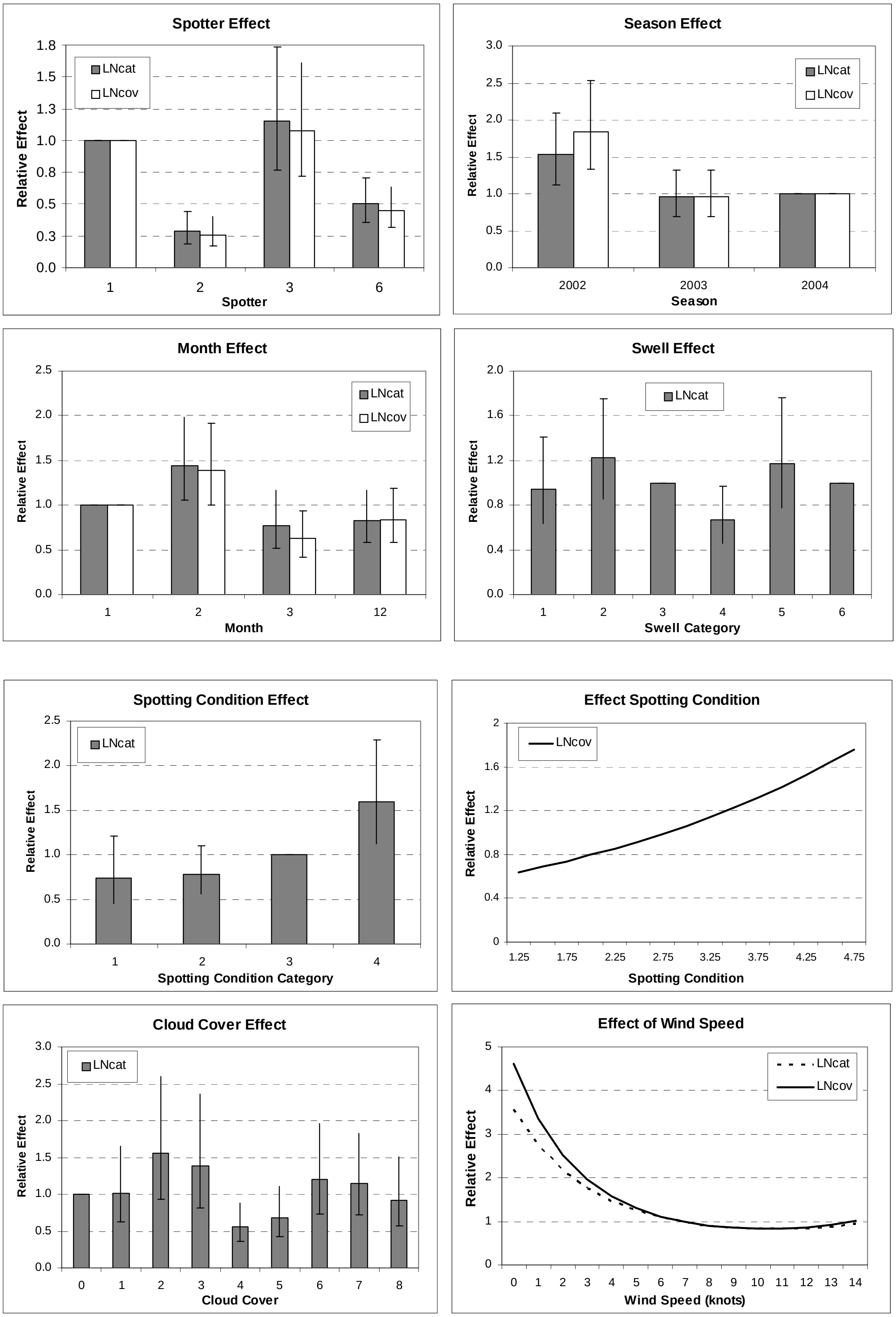
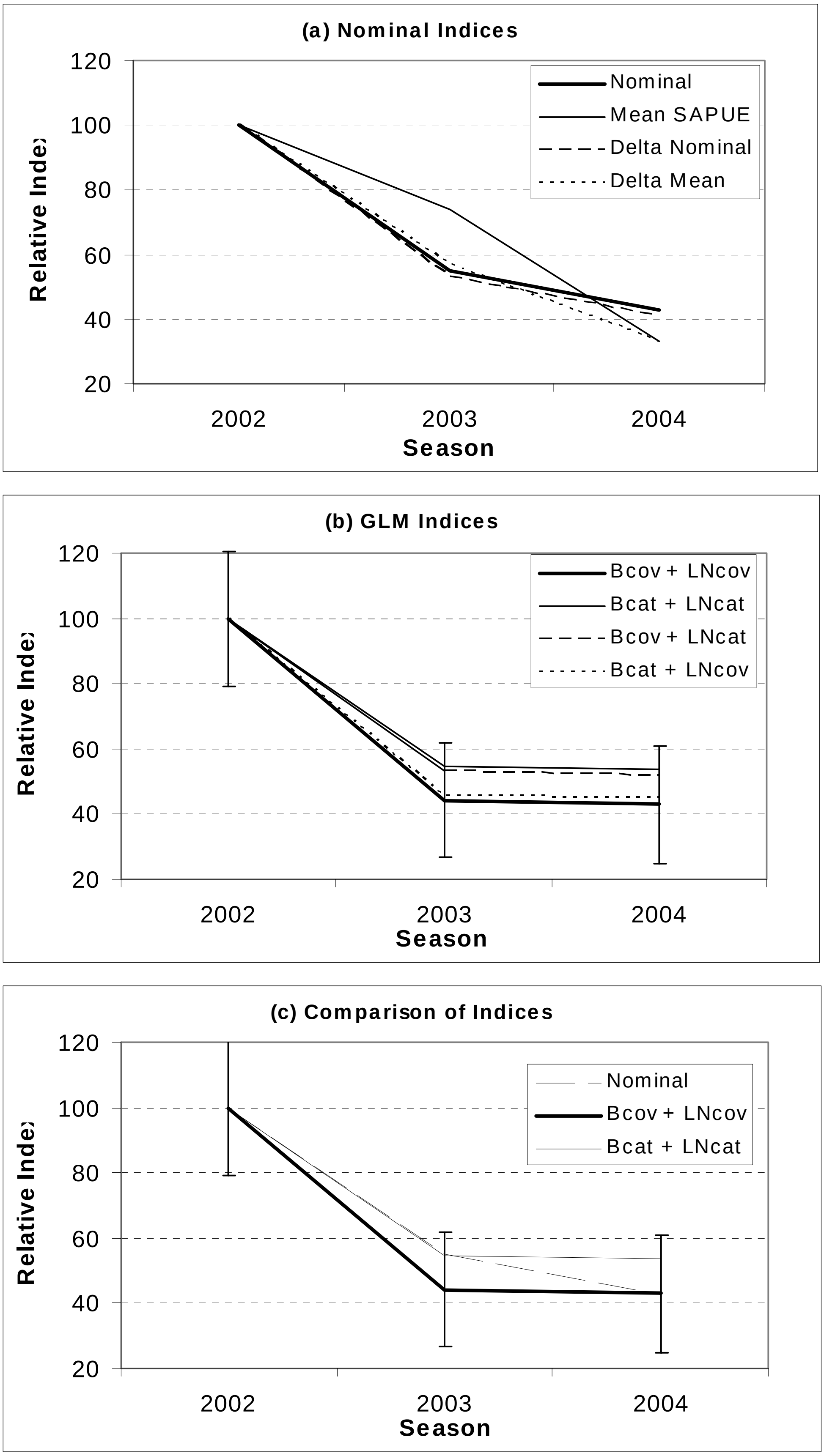


Table A5.Seasonal indices for each combined model.

Model	Season	Relative Index	Standard Error
Bcov	2002	100.0	20.6
+	2003	44.3	17.3
LNcov	2004	43.0	18.0
Bcat	2002	100.0	21.0
+	2003	45.3	17.3
LNcov	2004	44.8	17.8
Bcov	2002	100.0	21.8
+	2003	53.4	20.2
LNcat	2004	51.7	21.0
Bcat	2002	100.0	22.1
+	2003	54.6	20.1
LNcat	2004	53.9	20.8
Nominal	2002	100.0	
SAPUE	2003	55.1	
	2004	42.8	
Mean SAPUE	2002	100.0	
	2003	74.0	
	2004	32.9	
Delta	2002	100.0	
Nominal	2003	53.2	
SAPUE	2004	41.0	
Delta	2002	100.0	
Mean	2003	56.7	
SAPUE	2004	32.9	

Figure A6. Seasonal indices for each combined model.



iii) Combined Models

The seasonal indices of overall mean sighting rate, using equations (9) and (10), were calculated for all four possible combined GLM models together with the four nominal indices defined previously. As the relative values of these indices are dependent on the range of effects each index is standardised against, for ease of comparison a relative index was used such that the index for each season was expressed as a percentage of the index for the first season (2002). Use of this index also allows the relative change in the index between seasons to be easily seen. The relative indices for a range of selected models are listed in Table A5 and displayed in Figure A6.

The *Nominal* and *delta-Nominal* indices display a similar trend over the three years, declining around 45% between 2002 and 2003 and a further 12% between 2003 and 2004 (Figure A6a). On the other hand, the *Mean* and *delta-Mean* nominal indices display quite different trends between 2002 and 2003, though the overall decline between 2002 and 2004 is the same for both (~67%) and greater than that for the previous two nominal indices (~57%).

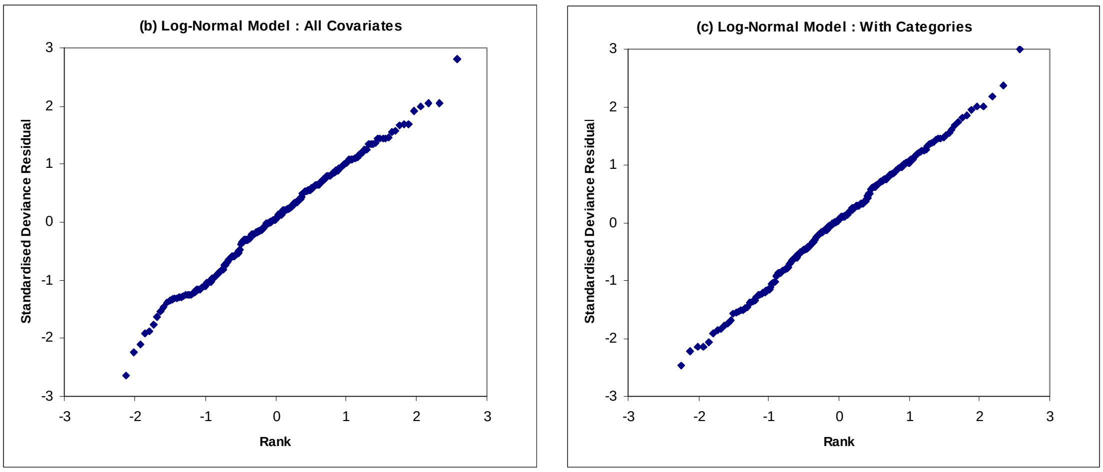
Of the four GLM based indices displayed in Figure A6b, two distinct trends can be seen. Both the *Bcat+LNcat* and the *Bcov+LNcat* models indicate a relative decline of around 46% between the 2002 and 2003 seasons, while the *Bcov+LNcov* and *Bcat+LNcov* indices indicate a respective decline of around 55%. The main difference between these two groups of indices is due to a difference between the positive sighting rates predicted for the *LNcat* and *LNcov* models. On the other hand, all indices shown in Figure A6b indicate that the overall sighting rate remained generally unchanged between the 2003 and 2004 seasons. The standard errors for these four models are generally similar and for clarity are shown for the *Bcov+LNcov* only. The index for each season for the other three models lies within the standard error of this selected model.

Finally, the *Nominal, Bcov+LNcov* and *Bcat+LNcat* indices are compared in Figure A6c. The two GLM indices bracket the range of predicted changes. Again, only the standard error for the *Bcov+LNcov* index is shown. The indices based on the *Nominal* and *Bcov+LNcov* models both indicate an overall decline of around 57% between the 2002 and 2004 seasons, though the decline between the first two seasons is less for the *Nominal* index. On the other hand, the decline in mean sighting rate between 2002 and 2003 is similar for the *Nominal* and *Bcat+LNcat* models.

Model Fits

The Q-Q plots, based on the standardised deviance residuals, for the *LNcov* and *LNcat* GLM models are shown in Figure A7. Similar plots for the *Bcov* and *Bcat* models are not shown as the residuals for these models display two clumps corresponding to the binary nature of the data and are generally not suitable for use as diagnostics of overall model fit.

Figure A7. Q-Q plots for selected fitted models.



Conclusion

Preliminary analyses which attempt to standardise the commercial sighting rate data for prevailing environmental conditions indicate a large decrease (between 46-56%) in sighting rates between the 2002 and 2003 summer seasons, with the lower sighting rate generally being maintained between the 2003 and 2004 seasons. However, the extent of the decline between the first two seasons depends to some extent on the nature of the standardisation model used and further analyses are required to fully explore and understand the nature of these differences.

References

Lo, N.C., Jacobson, L.D. and Squire, J.L. 1992. Indices of relative abundance from fish spotter data based on delta-lognormal models. Can. J. Fish. Aquat.Sci., 49, 2515-2526.

Pennington, M. 1983. Efficient estimators of abundance for fish and plankton surveys. Biometrics, 39, 281-286.

Pennington, M. and Berrien, P. 1984. Measuring the precision of estimates of total egg production based on plankton surveys. J. Plankton Res.., 6, 869-879.

Annex A1: Category Levels (and number of observations) used in models

Season			Month			Spotter		
Level	*Bin*	*LogN*	*Level*	*Bin*	*LogN*	*Level*	*Bin*	*LogN*
2002	70	61	12	61	44	1	96	79
2003	84	62	1	91	69	2	34	26
2004	115	80	2	69	58	3	67	48
			3	48	32	6	72	50

Spotting Condition:

Spotting Condition				
Level	*Bcat*		*LNcat*	
1	< 1.5	35	< 1.5	24
2	[1.5, 2.5)	80	[1.5, 2.5)	56
3	[2.5, 3.5)	93	[2.5, 3.5)	65
4	≥3.5	61	≥3.5	58

Swell Height:

Swell Height				
Level	*Bcat*		*LNcat*	
1	< 0.75	47	< 0.75	36
2	[0.75, 1.25)	60	[0.75, 1.25)	46
3	[1.25, 1.75)	71	[1.25, 1.75)	55
4	[1.75, 2.25)	53	[1.75, 2.25)	39
5	[2.25, 2.75)	26	≥2.25	27
6	≥2.75	12		

Cloud Cover:

Cloud Cover				
Level	*Bcat*		*LNcat*	
0	< 0.5	43	< 0.5	39
1	[0.5, 1.5)	26	[0.5, 1.5)	18
2	[1.5, 2.5)	21	[1.5, 2.5)	17
3	[2.5, 3.5)	21	[2.5, 3.5)	16
4	[3.5, 4.5)	33	[3.5, 4.5)	27
5	[4.5, 5.5)	29	[4.5, 5.5)	22
6	[5.5, 6.5)	27	[5.5, 6.5)	20
7	[6.5, 7.5)	39	[6.5, 7.5)	25
8	≥7.5	30	≥7.5	19

Annex A2: Calculation of variance for combined index

Given the following formula:

$$Var[p\mu_S] = Var[p](\mu_S)^2 + Var[\mu_S]p^2 + Var[\mu_S].Var[p]$$

and substituting Var[p] = p(1-p) we obtain:

$$\begin{aligned} Var[p\mu_S] &= p(1-p)(\mu_S)^2 + Var[\mu_S]p^2 + Var[\mu_S].p(1-p) \\ &= p(1-p)(\mu_S)^2 + Var[\mu_S]p^2 + pVar[\mu_S] - Var[\mu_S]p^2 \\ &= p(1-p)(\mu_S)^2 + pVar[\mu_S] \end{aligned}$$

which is equivalent to the formula given by equation (10) previously.