New Zealand Fisheries Assessment Research Document 95/11

Rock lobster standardised CPUE analysis

M. N. Maunder and P. J. Starr
New Zealand Fishing Industry Board
Private Bag 24-901
Wellington

August 1995

National Institute of Water and Atmospheric Research Ltd, Wellington

This series documents the scientific basis for stock assessments and fisheries management advice in New Zealand. It addresses the issues of the day in the current legislative context and in the time frames required. The documents it contains are not intended as definitive statements on the subjects addressed but rather as progress reports on ongoing investigations.

Rock lobster standardised CPUE analysis

M. N. Maunder and P. J. Starr

New Zealand Fisheries Assessment Research Document 95/11. 28 p.

Executive Summary

It is suggested from this analysis of New Zealand rock lobster stocks that abundance indices estimated from an unstandardised analysis are essentially the same as those estimated using a standardised one, particularly as the uncertainties in the catch and effort data are considerable.

1 Introduction

The method presented here was developed as an alternative method of calculating yearly abundance indices and for testing the sensitivity of the yearly indices to other sources of variation. The year indices are used in surplus production modelling for stock assessment (Maunder and Starr 1995; Breen & Kendrick 1994). A standardised CPUE (catch per pot-lift) index is calculated for the five rock lobster stocks, NSN (North NSI), NSC (Central NSI), NSS (South NSI), CHI(Chatham Islands) and NSI (North South Island) using a multiple regression technique. The natural log of CPUE [log(CPUE)] is regressed against the explanatory variables year, month, area, number of days fished in the month, total length of vessel, vessel tonnage, vessel kilowatts, and number of crew. These indices are compared with unstandardised indices and the current NIWA/MAF indices (Breen & Kendrick 1994) which are used in stock assessments.

2 Methods

2.1 Standardised CPUE Analysis

The data were checked for errors as described in Appendix 1.

The analysis was carried out using a standardised multiple regression method as described by Doonan (1991) and Vignaux (1992). Log(CPUE) was chosen as the dependent variable because CPUE is usually log-normally distributed. The dependent variable was regressed separately against each of the predictor variables: year, month, statistical area, number of days fished in the month, total length of vessel, vessel tonnage, vessel kilowatts, and number of crew. The categorical variables year, month, and area were included in the regression using dummy (indicator) variables. The remaining variables were included as cubics or the highest polynomial that was significant.

Because the dependent variable is log transformed, records with zero catches were replaced with 0.001 kg per pot-lift.

The predictor variable that explained the most variance (highest R-square) in log(CPUE) was chosen as the first variable to include in the model. The model was then regressed separately against each of the remaining predictor variables with the first predictor variable included. The predictor variable explaining the most variance was chosen as the next variable to include in the model. This stepwise regression technique was repeated until either there were no variables left or the next variable explaining the most variance was not significant at the 5% level.

The log(CPUE) model causes the variables to have a multiplicative effect on the CPUE. A relative year index can be calculated from the regression using the following equations from Vignaux (1992).

$$\hat{A}_j = e^{\hat{Y}_j - \hat{Y}_{base}} \qquad \text{Eq. 1}$$

$$s^2_{\hat{A}_j} = \hat{A}^2_j \, e^{\sigma^2_j}\left(e^{\sigma^2_j} - 1\right) \qquad \text{Eq. 2}$$

$$\sigma^2_j = Var(\hat{Y}_j) + Var(\hat{Y}_{base}) - 2\,Cov(\hat{Y}_j, \hat{Y}_{base}) \qquad \text{Eq. 3}$$

where
$\hat{A}$ is the standardised index value for year j
$\hat{Y}_j$ is the year coefficient from the regression for year j
base indicates the base year
$s^2_{\hat{A}_j}$ is the variance of the year effect for year j

This method compares the year effect of every year to a chosen year (base year). The regression model used in this analysis combines the effect of the base year (and the bases for all other variables) into the regression mean which simplifies these calculations. The coefficients then become differences from the base year, and the variance of the difference (Equation 3) is the square of the standard error of the coefficients. The base year should be chosen as the year with the smallest variance (usually the year with the largest number of records). This creates error bounds that show the greatest difference in relative confidence. The base year used was 1984. The following calculations were used to estimate the year index and its error bounds.

$$\hat{A} = e^{\hat{Y}_j} \qquad \text{Eq. 4}$$

$$s^2_{\hat{A}_j} = \hat{A}^2_j e^{\sigma^2_j}\left(e^{\sigma^2_j} - 1\right) \qquad \text{Eq. 5}$$

$$\sigma^2_j = SE^2_j \qquad \text{Eq. 6}$$

where SE_j is the standard error of the year *j* coefficient from the regression.

This standardised analysis was done for each area group: NSN (CRA1 & CRA2), NSC (CRA3, CRA4, and CRA5), NSS (CRA7 and CRA8), CHI (CRA6), and NSI (all areas excluding CRA6 and CRA9). Because a number of records had no vessel data (Tables 1—5), a model without vessel information (only year, month, and area) is also presented. All models are compared with unstandardised indices and the current NIWA/MAF (unstandardised) indices (Breen & Kendrick 1994). The unstandardised indices were calculated by dividing the total catch by the total effort for each year and normalising for 1984.

2.2 Surplus production modelling

Gilbert's (1992) formulation of the Pella-Tomlinson model as described by Maunder & Starr (1995) was used to investigate the effect of standardising the CPUE for the three stocks, NSN, NSC, and NSS when applied to a population model which has been used to assess these stocks. The only major difference between the model used in this analysis and that used by Breen & Kendrick (1994) was the substitution of B_{msy}/B_0 (the ratio of biomass at maximum sustainable yield to virgin biomass) set to equal 20% rather than using the Fox model (B_{msy}/B_0=37%). This was done because it is believed that this is a more accurate representation of the biology of the stocks (Maunder & Starr 1995). This substitution should not affect the comparison of the effects of the different abundance indices. The model fitting procedure is given in Maunder & Starr (1995) and the catch and effort data are given in Breen & Kendrick (1994).

The model was fitted to three time series of abundance indices derived for each substock from the catch and effort data: (a) the unstandardised CPUE index from 1963 to 1992 (identified as 'Long' in the results and presented in Breen & Kendrick (1994) and also used by Maunder & Starr (1995)); (b) the unstandardised CPUE index from 1979 to 1992 (identified in the results as 'Short' and calculated from the data used in this analysis), and (c) the standardised (Month, Area, and Year model) index from 1979 to 1992 (identified as 'Standard' in the results).

3 Results

3.1 Results of the Standardised CPUE Analysis

In general for the full model, year and area were the variables which had the largest explanatory power for the variation in CPUE (Table 6) when added stepwise. Exceptions were NSC where month explained 5.54% of the variation and tonnage explained 5.84% of the variation, and CHI where month explained most (8.00%) of the variation. All variables (sets of variables: categorical and cubic groups) were significant except tonnage for NSS and CHI. All categorical variables were kept as sets of indicator variables. The other variables were kept in the polynomial form described in Table 7.

All areas showed a general decline in the year index from 1979 to 1992 (Figures 1—5). All areas except NSS and CHI showed an increase in the year index from 1979 to 1980. NSN showed an increase from 1988 to 1990. NSS showed a cyclic decline with increases in the year index from 1983 to 1985 and 1989 to 1991. CHI showed small increases from 1982 to 1983 and 1985 to 1986.

The indices from the full model (including vessel effects) were very similar to the indices using only year, month, and area (Figures 1—5 and Tables 8—12). This indicates that the vessel effect is small. This is also shown by the relative low proportion of the total R^2 attributed to the vessel effects (Table 6). The main differences in the models lie in the early years where vessel data were missing and many records were left out of the analysis (Tables 1—5).

The unstandardised index was very similar to the standardised index for all areas with the greatest differences being in the early years, especially for NSS.

The unstandardised indices are nearly identical to the NIWA/MAF indices indicating that the data used are very similar. Note that the CELR data (1989—92) used in this analysis are estimated not landed (or converted to landed) weight while the NIWA/MAF unstandardised index is based on landed weight.

The month effect for the full model was similar in all areas (Figure 6) with higher CPUE in October to January and a local maximum in June (except NSN). For completeness the statistical area effect is shown in Figure 7.

A number of checks were carried out on the assumptions in the regression. The residuals from the regression are normally distributed (Figure 8). The replacement of zero catches with 0.001 kg per pot lift was not sensitive to replacement with 0.1 kg per pot lift. The regression was not sensitive to the estimation of effort (the estimation of effort is described in appendix 1 and Figure 9). The regression was not sensitive to catch being appropriated to incorrect months because of trips overlapping consecutive months.

3.2 Results of Surplus Production Modelling

The results of the surplus production analysis are given in Table 13. Because of the confounding of B_0 (virgin biomass) and r (productivity rate at maximum production), it is suggested that these parameters and B_{msy} (biomass at maximum sustainable yield) are probably not useful in determining the relative effect of using each of the different abundance indices. Two more likely parameters of interest are *msy* (maximum sustainable yield) and B_{cur}/B_{msy} (the ratio of current biomass to biomass at maximum sustainable yield) because they should be less biased by the confounding of B_0 and r.

The results of the comparison between the abundance indices are given in Table 14 as ratios of the parameters being compared. For the parameter/stock comparisons shown, the effect of moving from the longer time series to a shorter time series ('Short/Long') is greater than the effect of moving from an unstandardised time series to a standardised time series ('Standard/Short'—with the clear exception of *msy* for the NSS stock and possibly B_{cur}/B_{msy} for the NSC stock). The percentage difference between 'Standard' and 'Long' confounds the two effects and generally gives the greatest difference between the parameters. This leads to the conclusion that it is more likely that the length of the time series being used is more important in determining the value of the estimated parameters than the procedure used to estimate each time series of abundance indices.

4 Discussion

Between 1979 and 1992, there has been a decreasing trend in the yearly abundance index estimated from catch per potlift data for each of the rock lobster stocks, but NSN, NSS, and CHI stocks each show short periods of increasing yearly abundance indices. Because there are only small differences between the two standardised indices and the two unstandardised indices, this analysis seems to be robust to the data used and to any vessel effects. The biggest difference between the standardised and unstandardised indices occurs in the early years for the NSS analysis.

A large number of records are missing vessel information and were dropped from the Full model analysis (Tables 1—5). Because the Full model and the Year, Month, and Area model results are similar and it appears that the vessel information has only a small amount of explanatory power, it is suggested that the Year, Month, and Area model is superior to the Full model because it includes more data.

Because of the large number of records used in the analysis, the criterion of a 5% level used as a stopping rule allows most variables to be included in the regression (*see* Table 6). It is clear that some of these variables explain very little variation and are probably not relevant to the regression. However, the inclusion of these variables is unlikely to significantly affect the resulting yearly abundance index. This is because the Year, Month, and Area model is only slightly different from the Full model and most of this difference is probably due to the large number of records which were dropped from the analysis for the Full model.

The results from the surplus production modelling indicate that the shortening of the time series of abundance indices has a greater effect on the parameter estimates of *msy* (maximum sustainable yield) and B_{cur}/B_{msy} (the ratio of current biomass to biomass at maximum sustainable yield) than the procedure of standardising the abundance index (except possibly *msy* for NSS stock). Using the standardised index reduces the number of data points from 30 (Breen & Kendrick 1994) to 14 because the detailed data necessary to do the standardised analysis are lacking prior to 1979. Using only 14 years of abundance indices (whether or not they are standardised) can give significantly different results when compared to similar analyses which use 30 abundance indices. However, in the future, it may be necessary to standardise the abundance index analysis in order to account for differences in fishing practices which may have occurred.

It is suggested from this analysis that abundance indices estimated from an unstandardised analysis are essentially the same as those estimated using a standardised one, particularly when the extent of the uncertainties in the catch and effort data are considered. Differences in estimated parameters when using different sets of abundance indices in the same underlying surplus production model are probably due to the nature of the indices and the inadequacies of the surplus production model to represent each rock lobster stock (a classic "one-way" trip—*see* Hilborn & Walters 1992).

References

Breen, P. A. & Kendrick, T. H. 1994: Surplus production and yield-per-recruit analysis for the red rock lobster (*Jasus edwardsii*) fishery. New Zealand Fisheries Assessment Research Document 94/5. 65 p.

Booth, J. D., Robinson, M., & Starr P. J. 1993: Recent research into New Zealand rock lobsters, and a review of rock lobster catch and effort data. New Zealand Fisheries Assessment Research Document 94/7. 22 p.

Doonan, I. 1991: Orange roughy fishery assessment, CPUE analysis-linear regression, NE Chatham Rise 1991. New Zealand Fisheries Assessment Research Document 91/9. 32 p.

Hilborn, R., & Walters, C.J. 1992: Quantitative fisheries stock assessment: choice, dynamics & uncertainty. Chapman & Hall, London and New York. 570 p.

Maunder, M. N. & Starr, P. J. (1995): Sensitivity of management reference points from the rock lobster surplus production model to the ratio of B_{msy}/B_0 determined by the Pella-Tomlinson shape parameter. New Zealand Fisheries Assessment Research Document 95/10. 22 p.

Vignaux, M. 1992: Catch per unit of effort (CPUE) analysis of the hoki fishery. New Zealand Fisheries Assessment Research Document 92/14. 31 p.

Table 1. Number of records (month fished by vessel in a statistical area) used in the standardised analysis for NSN (CRA1 and CRA2). Both models have records dropped which have catch per day >1000, catch per potlift >100 or zero effort. The full model has records dropped that are missing vessel data

Year	Total in data set	Year, month, & area	Full model
1979	1 027	1 027	408
1980	1 158	1 157	622
1981	1 139	1 138	726
1982	1 132	1 131	801
1983	1 087	1 087	843
1984	1 083	1 082	879
1985	1 084	1 084	868
1986	1 009	1 008	817
1987	1 029	1 029	862
1988	858	857	726
1989	585	585	504
1990	635	635	548
1991	792	790	711
1992	687	684	634
Total	13 305	13 294	9 949

Table 2. Number of records (month fished by vessel in a statistical area) used in the standardised analysis for NSC (CRA3, CRA4 and CRA5). Both models have records dropped which have catch per day >1000, catch per potlift >100 or zero effort. The full model has records dropped that are missing vessel data

Year	Total in data set	Year, month, & area	Full model
1979	2 436	2 430	912
1980	2 320	2 315	1 090
1981	2 172	2 167	1 303
1982	2 422	2 418	1 718
1983	2 506	2 503	1 987
1984	2 547	2 546	2 099
1985	2 418	2 417	2 015
1986	2 319	2 319	1 940
1987	2 266	2 265	1 940
1988	1 982	1 978	1 714
1989	1 953	1 950	1 723
1990	2 059	2 056	1 793
1991	2 228	2 223	1 963
1992	2 226	2 221	1 975
Total	31 854	31 808	24 172

Table 3. Number of records (month fished by vessel in a statistical area) used in the standardised analysis for NSS (CRA7 and CRA8). Both models have records dropped which have catch per day >1000, catch per potlift >100 or zero effort. The full model has records dropped that are missing vessel data

Year	Total in data set	Year, month, & area	Full model
1979	2 365	2 264	1 077
1980	2 303	2 234	1 218
1981	1 982	1 936	1 121
1982	1 847	1 807	1 127
1983	1 923	1 899	1 252
1984	2 008	1 988	1 346
1985	1 960	1 937	1 289
1986	1 752	1 715	1 195
1987	1 680	1 662	1 174
1988	1 312	1 291	919
1989	1 578	1 568	1 136
1990	1 515	1 509	1 083
1991	1 543	1 543	1 133
1992	1 575	1 573	1 140
Total	25 343	24 926	16 210

Table 4. Number of records (month fished by vessel in a statistical area) used in the standardised analysis for NSI (all areas except CRA6 and CRA9). Both models have records dropped which have catch per day >1000, catch per potlift >100 or zero effort. The full model has records dropped that are missing vessel data

Year	Total in data set	Year, month, & area	Full model
1979	5 985	5 862	2 444
1980	5 815	5 735	2 939
1981	5 307	5 249	3 158
1982	5 401	5 356	3 646
1983	5 516	5 489	4 082
1984	5 638	5 616	4 324
1985	5 462	5 438	4 172
1986	5 080	5 042	3 952
1987	4 975	4 956	3 976
1988	4 152	4 126	3 359
1989	4 138	4 124	3 376
1990	4 216	4 207	3 427
1991	4 572	4 565	3 808
1992	4 495	4 485	3 749
Total	70 752	70 250	50 412

Table 5. Number of records (month fished by vessel in a statistical area) used in the standardised analysis for CHI (CRA6). Both models have records dropped which have catch per day >1000, catch per potlift >100 or zero effort. The full model has records dropped that are missing vessel data

Year	Total in data set	Year, month, & area	Full model
1979	343	343	267
1980	294	293	223
1981	361	361	284
1982	436	436	367
1983	447	447	412
1984	442	442	413
1985	451	451	396
1986	402	402	365
1987	363	363	325
1988	393	393	345
1989	320	319	284
1990	336	336	311
1991	370	370	358
1992	447	445	420
Total	5 405	5 401	4 770

Table 6. Increase in R^2 as variables are added in the stepwise regression of the full model for the five areas. • indicates that the variable was not significant at the 5% level

Variable	NSN	NSC	NSS	NSI	CHI
Year	3.88%	13.55%	7.66%	7.06%	6.51%
Month	3.06%	5.54%	1.70%	2.52%	8.00%
Area	9.95%	5.67%	9.51%	13.77%	3.20%
Length	1.10%	0.21%	1.02%	0.32%	1.07%
Tonnage	2.84%	5.84%	*	3.53%	*
Kilowatts	0.03%	1.30%	0.42%	0.98%	0.75%
Number of crew	0.43%	0.26%	2.73%	0.20%	4.15%
Number of days fished	0.32%	1.85%	0.69%	0.50%	0.88%
Total	21.61%	34.22%	23.73%	28.88%	24.56%

Table 7. Polynomial order of variable kept in the model. "-" indicates variable not kept

Variable	NSN	NSC	NSS	NSI	CHI
Length	3	3	1	3	3
Tonnage	3	3	-	3	-
Kilowatts	1	3	3	3	2
Crew number	3	3	2	3	2
Number of days fished	3	3	3	3	3

Table 8. Year Indices for NSN (CRA1 and CRA2)for the standardised full and year, month and area models and the unstandardised and MAF indices

Year	Standardised (Full Model)	Lower 95% Bound	Upper 95% Bound	Standardised (Year, Month and Area)	Unstandar dised	MAF
1979	1.28	1.15	1.40	1.26	1.03	1.02
1980	1.66	1.52	1.80	1.53	1.29	1.28
1981	1.36	1.25	1.46	1.36	1.26	1.26
1982	1.29	1.19	1.39	1.26	1.19	1.19
1983	1.08	1.00	1.17	1.06	1.08	1.08
1984	1.00	1.00	1.00	1.00	1.00	1.00
1985	1.02	0.95	1.10	1.00	1.00	0.99
1986	1.04	0.96	1.12	1.03	0.97	0.97
1987	0.85	0.79	0.92	0.88	0.86	0.85
1988	0.80	0.74	0.87	0.84	0.89	0.89
1989	0.88	0.80	0.96	0.89	0.98	0.97
1990	1.08	0.98	1.17	1.17	1.06	1.06
1991	0.97	0.89	1.05	1.02	0.90	0.90
1992	0.94	0.86	1.02	0.99	0.85	

Table 9. Year indices for NSC (CRA3, CRA4 and CRA5) for the standardised full and year, month and area models and the unstandardised and MAF indices

Year	Standardised (Full Model)	Lower 95% Bound	Upper 95% Bound	Standardised (Year, Month and Area)	Unstandardised	MAF
1979	1.18	1.12	1.24	0.96	1.04	1.05
1980	1.23	1.17	1.28	1.07	1.11	1.11
1981	1.10	1.06	1.15	1.05	1.07	1.07
1982	1.20	1.15	1.25	1.16	1.14	1.15
1983	1.11	1.07	1.16	1.08	1.08	1.08
1984	1.00	1.00	1.00	1.00	1.00	1.00
1985	0.85	0.82	0.88	0.85	0.88	0.88
1986	0.85	0.82	0.89	0.86	0.93	0.93
1987	0.68	0.66	0.71	0.68	0.75	0.75
1988	0.61	0.59	0.64	0.61	0.65	0.65
1989	0.61	0.59	0.64	0.60	0.67	0.68
1990	0.60	0.58	0.62	0.60	0.61	0.61
1991	0.52	0.50	0.54	0.50	0.55	0.54
1992	0.50	0.48	0.52	0.48	0.61	

Table 10. Year indices for NSS (CRA7 and CRA8) for the standardised full and year, month and area models and the unstandardised and MAF indices

Year	Standardised (Full Model)	Lower 95% Bound	Upper 95% Bound	Standardised (Year, Month and Area)	Unstandardised	MAF
1979	1.82	1.71	1.93	1.75	1.31	1.30
1980	1.62	1.52	1.71	1.56	1.28	1.28
1981	1.48	1.39	1.57	1.41	1.24	1.26
1982	1.18	1.11	1.25	1.12	1.15	1.19
1983	0.94	0.88	0.99	0.93	0.96	0.96
1984	1.00	1.00	1.00	1.00	1.00	1.00
1985	1.16	1.10	1.23	1.17	1.07	1.08
1986	1.07	1.00	1.13	1.08	0.92	0.93
1987	1.09	1.02	1.15	1.11	0.96	0.98
1988	0.78	0.73	0.83	0.79	0.69	0.73
1989	0.69	0.65	0.74	0.70	0.74	0.74
1990	0.70	0.66	0.75	0.73	0.65	0.65
1991	0.80	0.75	0.85	0.83	0.70	0.70
1992	0.68	0.64	0.72	0.71	0.61	

Table 11. Year indices for NSI (all areas except CRA6 and CRA9) for the standardised full and year, month and area models and the unstandardised and MAF indices

Year	Standardised (Full Model)	Lower 95% Bound	Upper 95% Bound	Standardised (Year, Month and Area)	Unstandardised	MAF
1979	1.41	1.35	1.46	1.27	1.21	1.19
1980	1.43	1.38	1.49	1.31	1.20	1.20
1981	1.26	1.22	1.31	1.23	1.14	1.16
1982	1.22	1.18	1.27	1.17	1.12	1.14
1983	1.05	1.02	1.09	1.02	1.03	1.03
1984	1.00	1.00	1.00	1.00	1.00	1.00
1985	0.97	0.94	1.01	0.98	0.98	0.98
1986	0.95	0.92	0.98	0.96	0.93	0.93
1987	0.83	0.80	0.86	0.85	0.85	0.86
1988	0.70	0.67	0.72	0.71	0.68	0.70
1989	0.68	0.66	0.71	0.68	0.75	0.75
1990	0.71	0.68	0.73	0.72	0.67	0.66
1991	0.66	0.64	0.69	0.68	0.64	0.64
1992	0.61	0.59	0.64	0.62	0.65	

Table 12. Year indices for CHI (CRA6) for the standardised full and year, month and area models and the unstandardised and MAF indices

Year	Standardised (Full Model)	Lower 95% Bound	Upper 95% Bound	Standardised (Year, Month and Area)	Unstandardised	MAF
1979	1.58	1.40	1.76	1.53	1.52	1.50
1980	1.53	1.34	1.72	1.47	1.51	1.25
1981	1.56	1.38	1.74	1.54	1.47	1.45
1982	1.14	1.02	1.26	1.16	1.19	1.19
1983	1.17	1.05	1.28	1.18	1.18	1.16
1984	1.00	1.00	1.00	1.00	1.00	1.00
1985	0.92	0.83	1.01	0.92	0.93	0.93
1986	0.99	0.89	1.09	1.02	1.03	1.01
1987	0.98	0.87	1.08	0.99	1.05	1.03
1988	0.87	0.78	0.96	0.89	0.97	0.95
1989	0.79	0.70	0.88	0.81	0.94	0.98
1990	0.74	0.66	0.83	0.78	0.84	0.86
1991	0.74	0.66	0.82	0.78	0.77	0.79
1992	0.77	0.69	0.85	0.80	0.83	

Table 13. Results from Gilbert's (1992) formulation of the Pella-Tomlinson model as described in Maunder and Starr (submitted) with B_{msy}/B_0=20% for the three stocks, NSN, NSC and NSS, when fit to the unstandardised CPUE index from 1963 to 1992 (Long, as given in Breen and Kendrick (1994) and used by Maunder and Starr (1995)), the unstandardised CPUE index from 1979 to 1992 (Short) and the standardised (Month, Area and Year model) index from 1979 to 1992 (Standard)

Index	B_0 (t)	r	msy (t)	B_{msy} (t)	B_{cur}/B_{msy}
NSN Long	9 377	0.2891	540	1 869	190%
Short	9 334	0.3326	621	1 867	231%
Standard	3 722	0.8721	649	744	228%
NSC Long	36 663	0.2321	1 696	7 308	163%
Short	59 161	0.0842	996	11 832	140%
Standard	61 122	0.0658	804	12 224	119%
NSS Long	72 764	0.0943	1 369	14 508	96%
Short	106 724	0.0173	368	21 345	69%
Standard	86 520	0.0468	810	17 304	52%

Table 14. Comparison (percentage difference) of msy and B_{cur}/B_{msy} estimates for the three stocks, NSN, NSC and NSS, when fit to the unstandardised CPUE index from 1963 to 1992 ('Long', from Breen and Kendrick (1994)), the unstandardised CPUE index from 1979 to 1992 ('Short') and the standardised (Month, Area and Year model) index from 1979 to 1992 ('Standard')

	Comparison	msy (t)	B_{cur}/B_{msy}
NSN	Short/Long	15%	22%
	Standard/Short	5%	-1%
	Standard/Long	20%	20%
NSC	Short/Long	-41%	-14%
	Standard/Short	-19%	-15%
	Standard/Long	-53%	-27%
NSS	Short/Long	-73%	-28%
	Standard/Short	120%	-25%
	Standard/Long	-41%	-46%

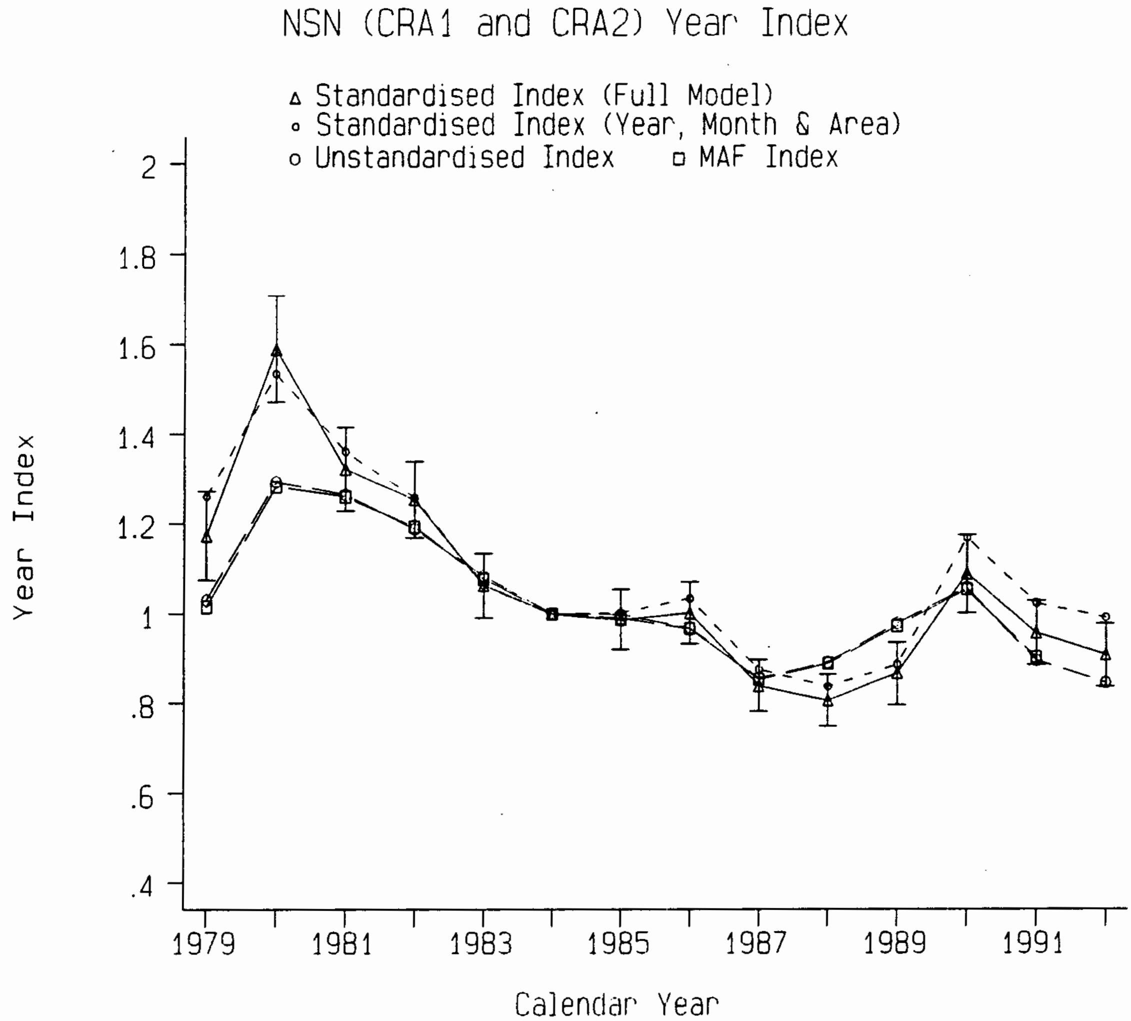

Figure 1. Year indices for NSS (CRA1 and CRA2). The standardised index using the full model is shown with 95% confidence bounds.

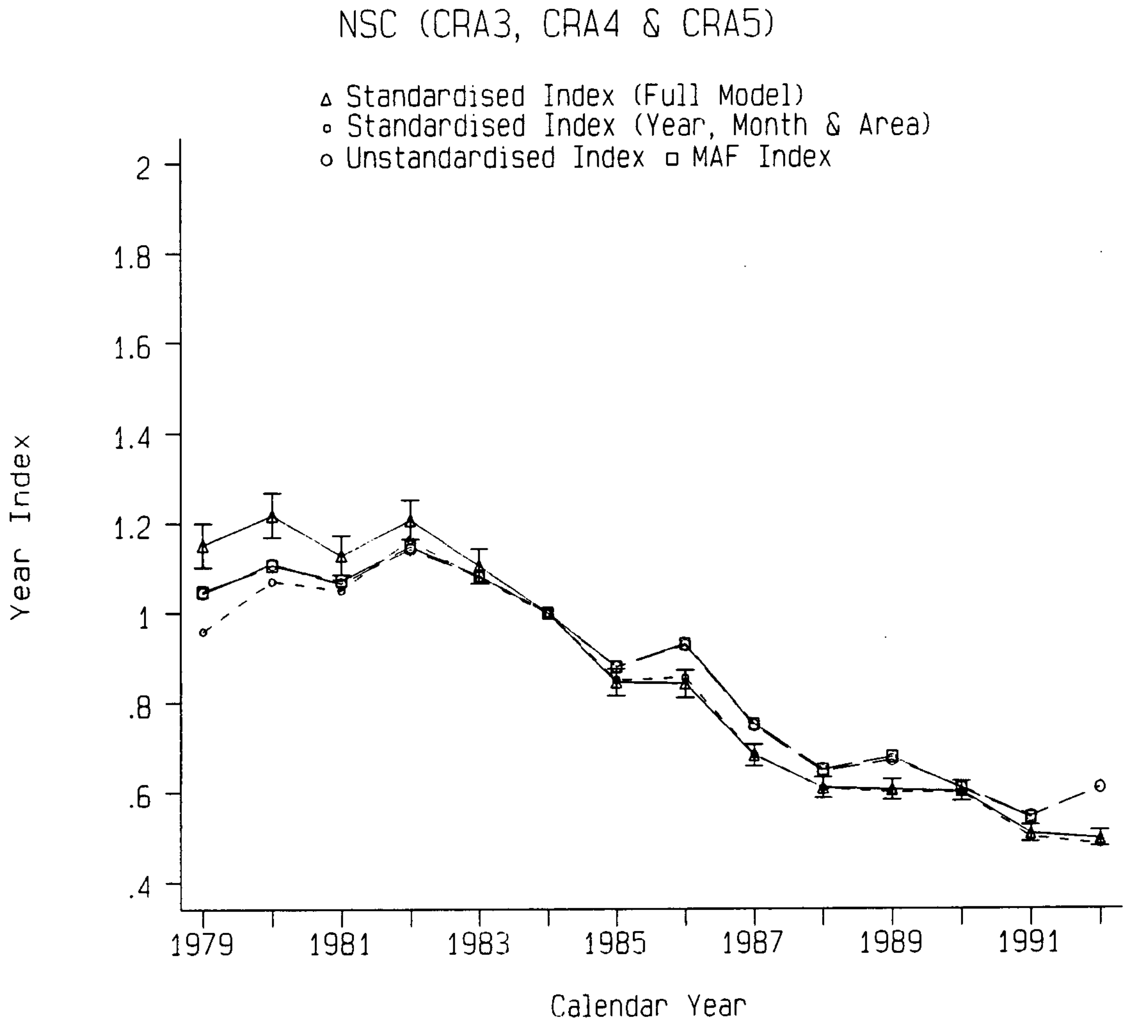

Figure 2. Year indices for NSC (CRA3, CRA4 and CRA5). The standardised index using the full model is shown with 95% confidence bounds.

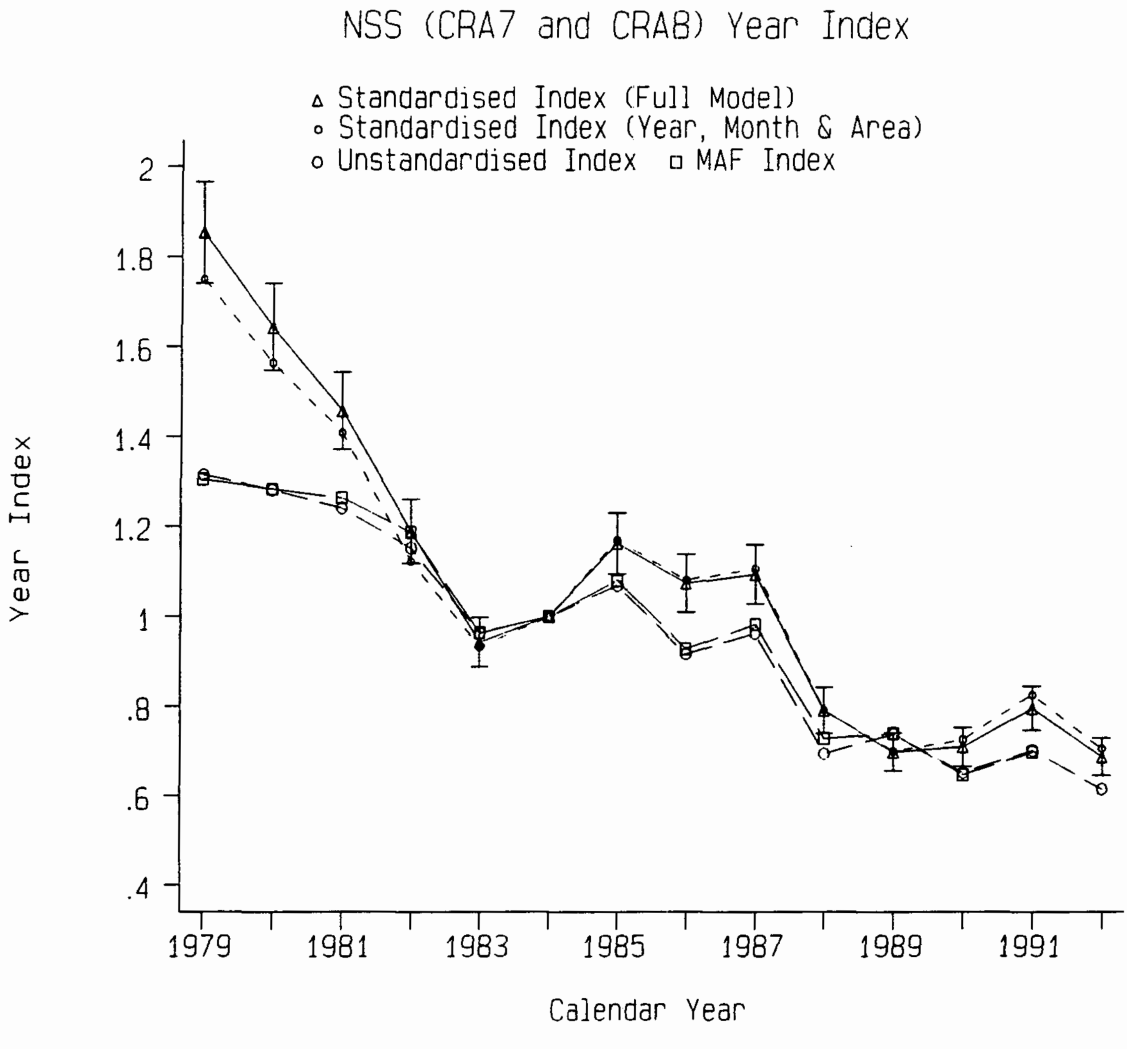

Figure 3. Year indices for NSS (CRA7 and CRA8). The standardised index using the full model is shown with 95% confidence bounds.

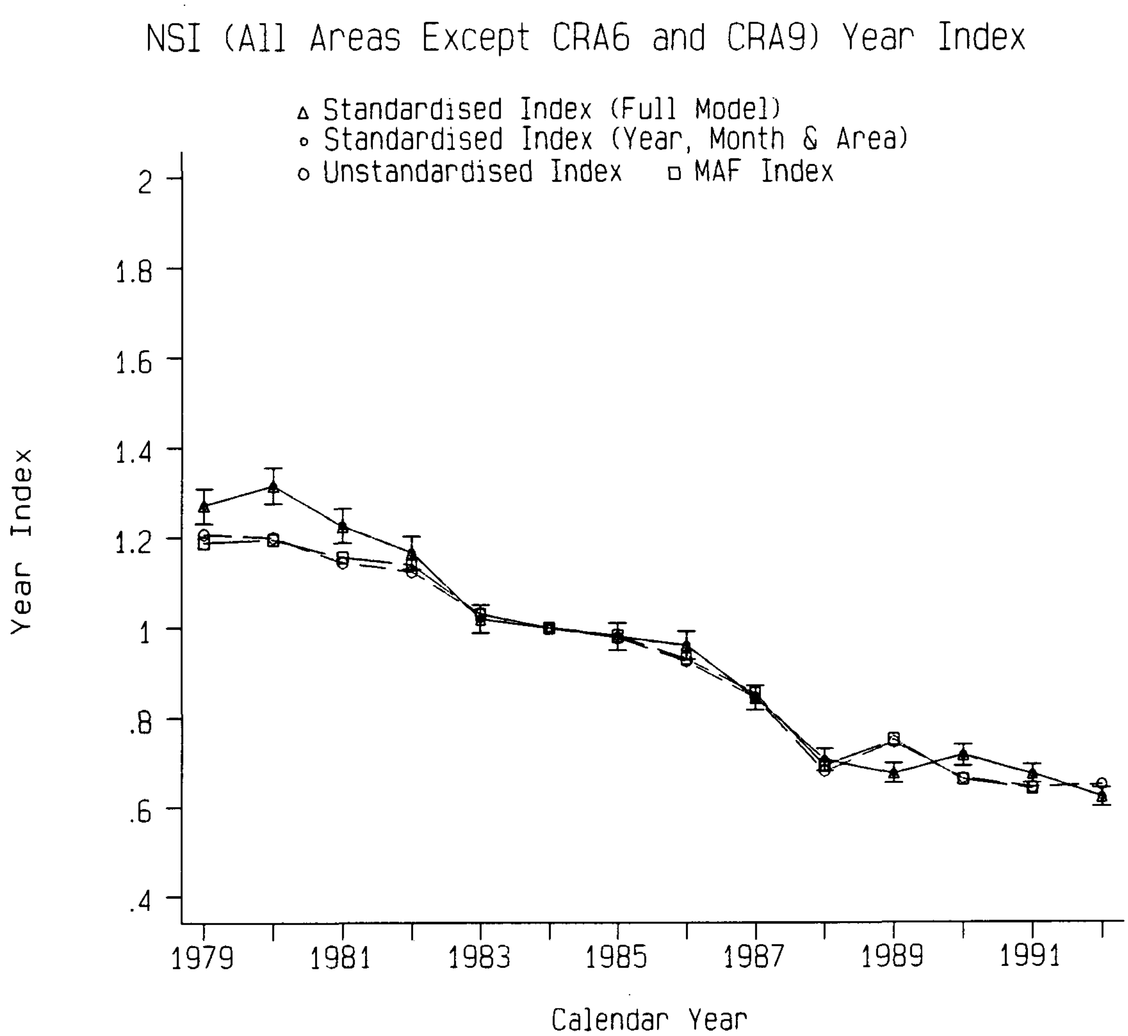

Figure 4. Year indices for NSI (all areas except CRA6 and CRA9). The standardised index using the full model is shown with 95% confidence bounds.

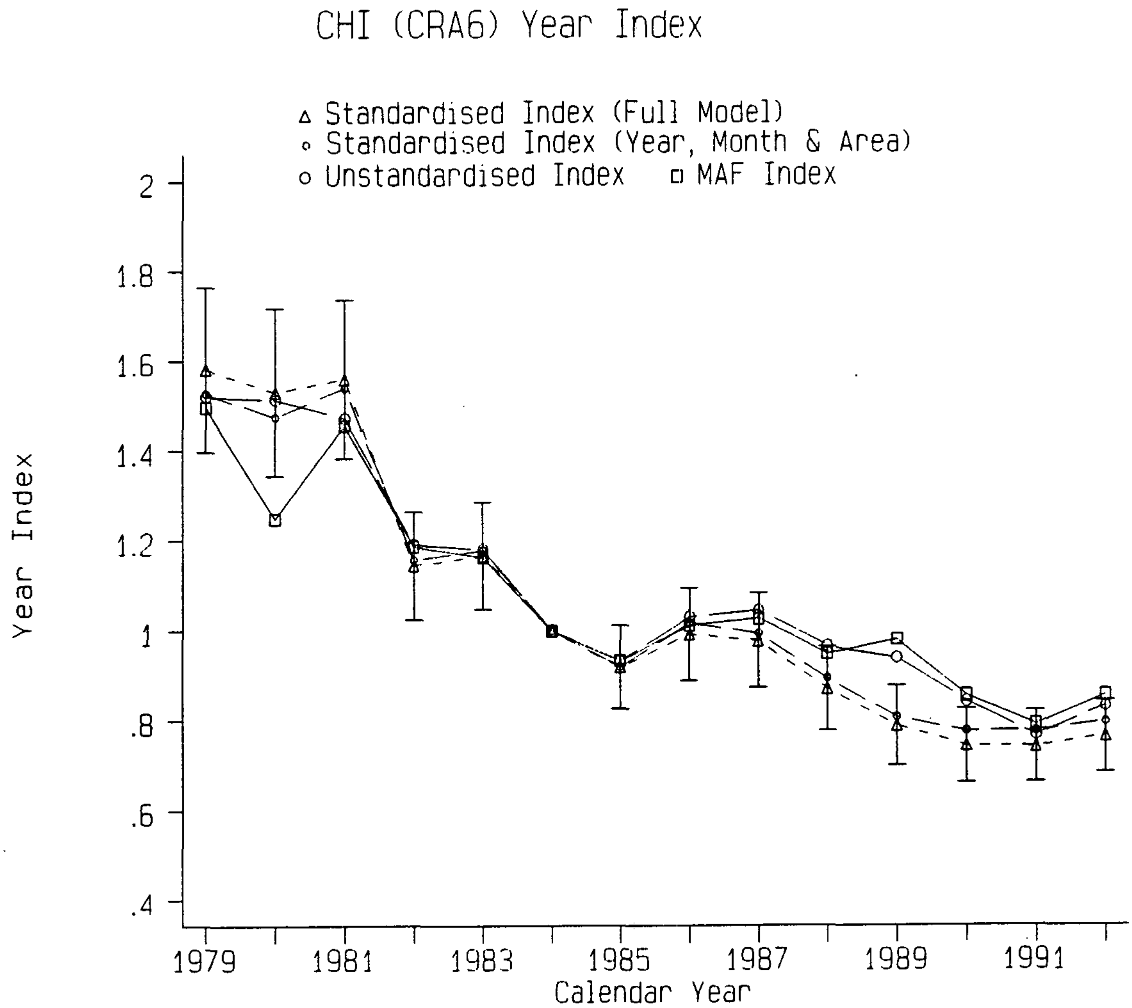

Figure 5 Year indices for CHI (CRA6). The standardised index using the full model is shown with 95% confidence bounds.

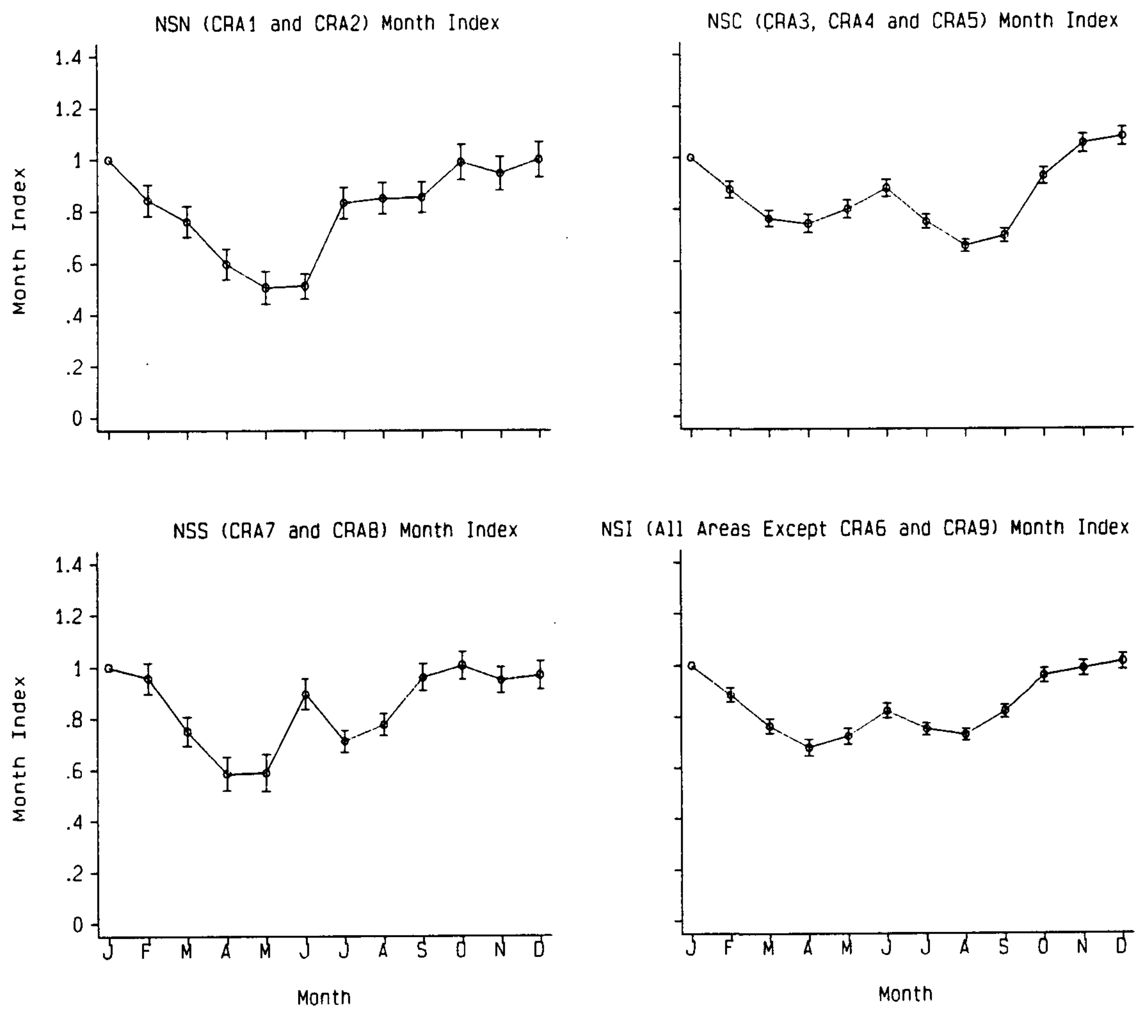

Figure 6 Month effect and 95% confidence bounds from the full model for the four areas NSN, NSC, NSS, and NSI.

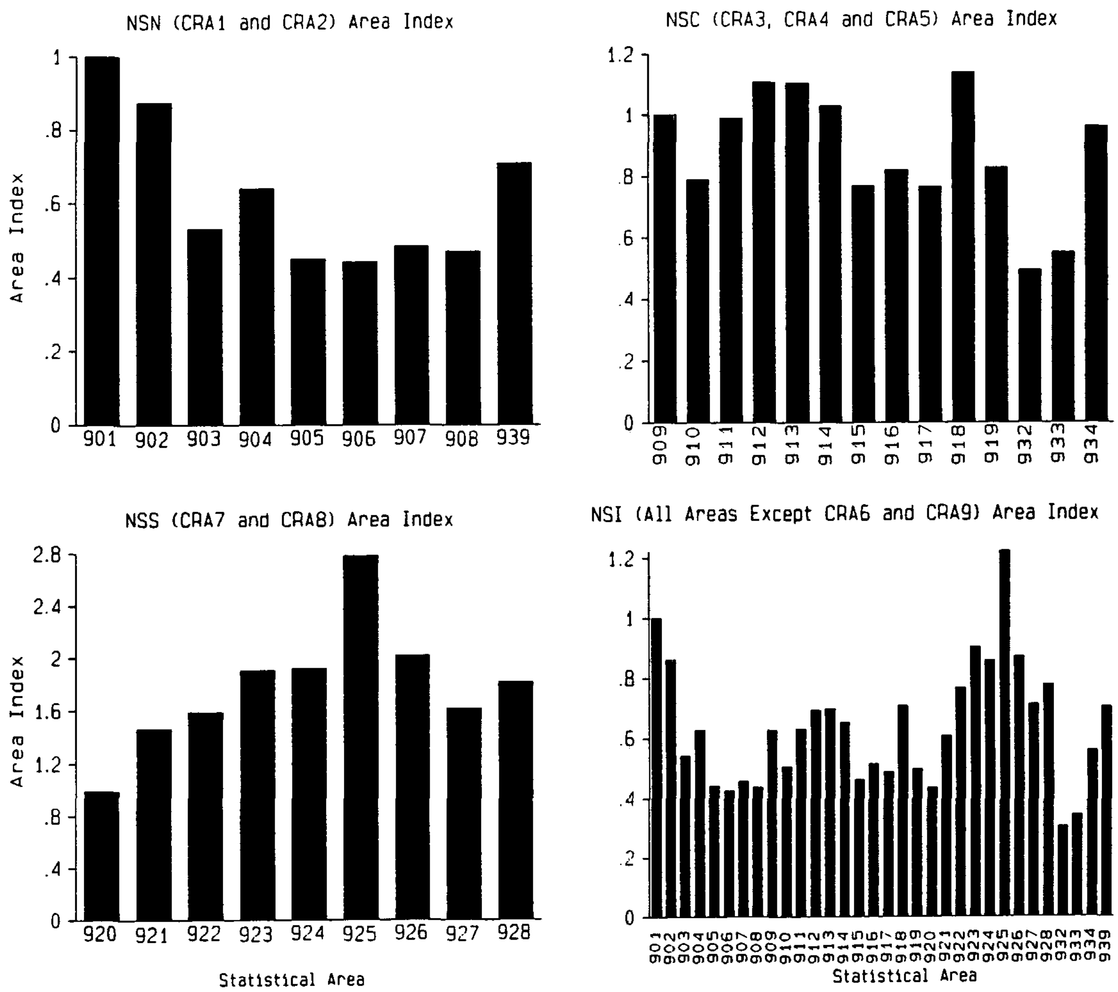

Figure 7 Statistical area effect from the full model for the four areas NSN, NSC, NSS, and NSI.

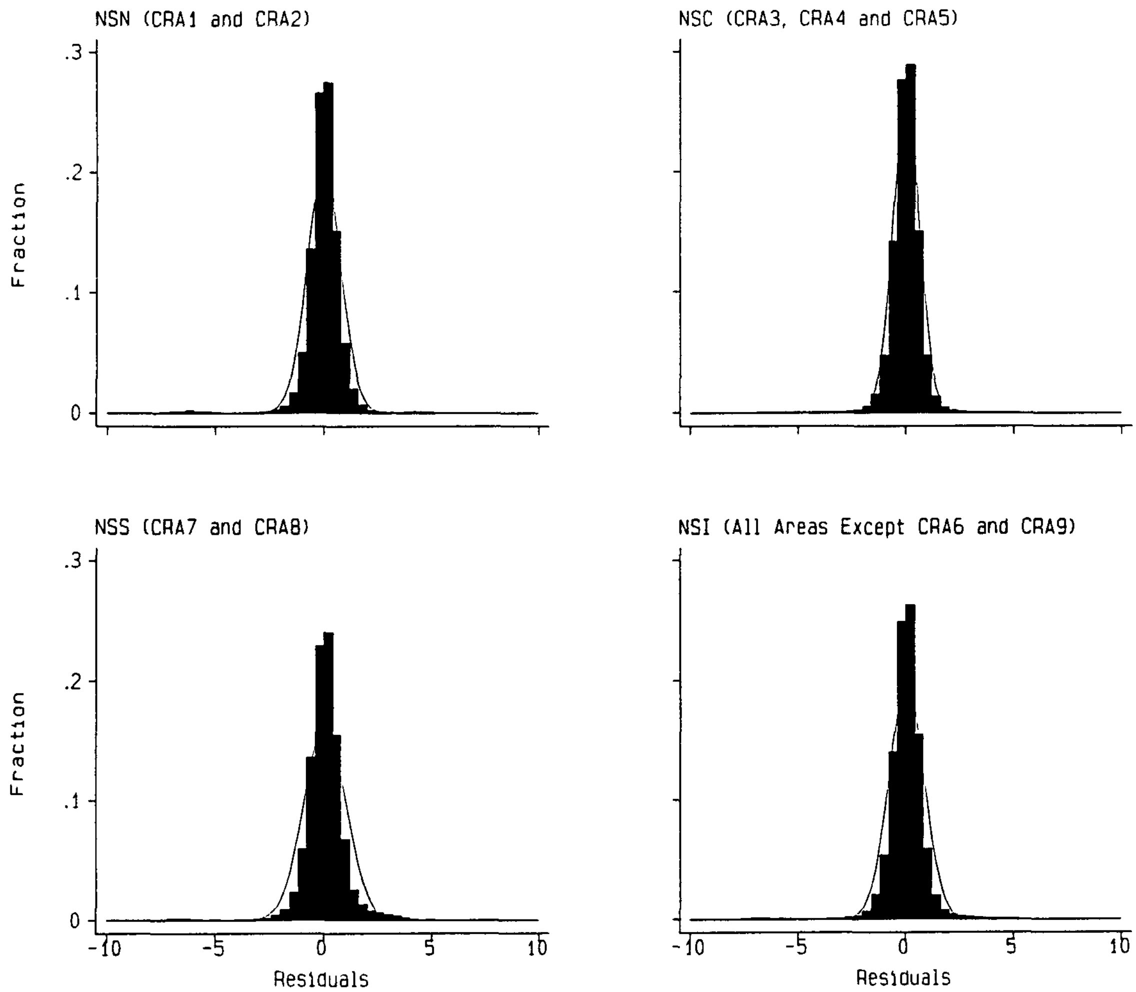

Figure 8 Residual plots of the full model fitted with a normal curve for the four areas NSN, NSC, NSS, and NSI.

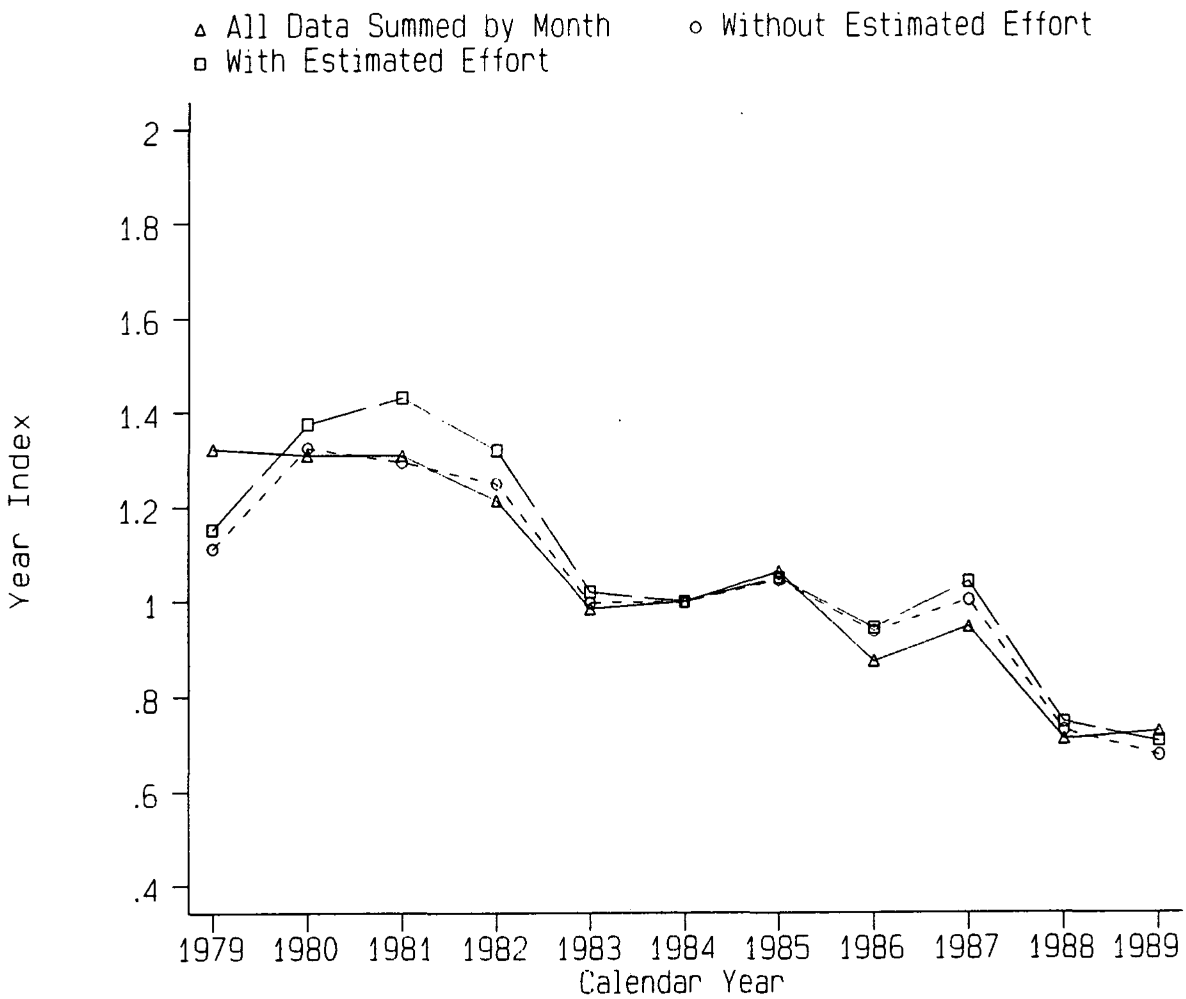

Figure 9 Sensitivity of the full model regression to estimation of effort. QMA8 is used as the example because this is where most of the effort data is estimated.

Appendix 1

Data Error Checking

FSU Data

Data for years 1979 to 1989 (up to 31 July) were modified as described in Booth *et al.* (1993). These data were originally obtained from the QMS held version of the FSU data base (*see* Booth et al. 1993 for a discussion of this distinction).

1) Many FSU records had missing effort data. Any effort data that were missing from the FSU data were estimated using the procedure described in Booth et al. (1993) appendix 1. Missing effort was estimated as the average effort in a boat/statistical-area/month/year stratum. If there were no records in that stratum, then the average for all boats in the statistical-area/month/year stratum was used. Effort for some records could not be estimated and these were dropped from the analysis (Appendix Table 1).

CELR Data

CELR data for years 1989 (1 August on) to 1992 were obtained from the NIWA/MAF database and checked for errors as follows (Appendix Table 2):

1) Only data from rock lobster potting were used in the analysis.

2) The year of capture is not given in the data records. Therefore, the year of capture was calculated from the trip start and end dates and the month was checked if it was in the range of the trip (note: monthly averages were used in the analysis and incorrect days will not affect the analysis). If the year could not be calculated from the data or the month was incorrect, then the record was left out of the analysis.

3) The effort data were checked for duplication of records and no exact duplicates were found. A number of records had duplicates for all fields except for the page number. The record with the highest page number was kept and all records with lower page numbers were dropped from the analysis. These are records that were assumed to have been entered twice with different page numbers. The sequence appears to be that incorrect records were entered and also sent back for correction. When a correct report was sent back, a new record was entered with a new (higher) page number, (M. Robinson, NIWA/MAF Greta Point. *pers. comm.*). For this analysis, all records with an incorrect (lower) page number were dropped. These types of duplicates were also found in the landing data and were treated in the same way.

4) Records with the same catch date and area on the same page were summed into one record and the extra records were dropped from the analysis.

5) Any CELR catch record greater than 1000 kg was dropped from the analysis.

6) CELR effort data were corrected as described by Booth *et al.* (1993). If "efforta"

(the number of pots in the 24 hours) was greater than 499 it was assumed to be incorrect. Missing "efforta" data were replaced with "effortb" (the number of pot lifts in the water at midnight) if "effortb" was less than 500 and greater than 10. Any records with missing effort data were dropped.

7) Note that the CELR data is estimated catch and is not converted to landed catch.

All Data

1) The catch and effort data were summed for each boat/statistical-area/month/year stratum. This was done because, in the earlier years, catch for a whole trip was commonly recorded in the last record and daily catch data were not available (note: this reduces the size of the confidence intervals). The monthly catch per potlift for each stratum was calculated by dividing the total catch for each boat/area/month/year stratum by the total effort for that stratum. Any CPUE estimates that exceeded 100 kg/pot-lift were excluded from the analysis (Appendix Table 3).

2) Any records with average catch per day for the stratum greater than 1000 kg/day or zero effort were dropped from the analysis (*see* Appendix Table 3).

Vessel Data

The vessel data included information on total length of vessel, vessel tonnage, vessel kilowatts, and number of crew. Many vessels had no data and much of the data available had unrealistic values. Values were chosen as cutoff points to eliminate outliers after looking at biplots and discussion with T. Craig (Federation of Commercial Fishermen). The number of records used in the regression are given in Tables 1—5 in the main part of this document.

The record was dropped if any of the following were true.

1) All vessel variables were equal to zero

2) Vessel tonnage or kilowatts was equal to zero and the total length was greater than 5 m

3) Vessel length was less than 3 m or greater than 20 m

4) Vessel tonnage was greater than 65 t

5) Vessel kilowatts was greater than 600

6) The number of crew was zero or greater than 10

Appendix Table 1. The number of FSU records with effort estimated from other records. Records where effort could not be estimated from the available data were dropped

Year	Total Number of Records	Number of Records with Effort Estimated	Number of Records Dropped	Year	Total Number of Records	Number of Records with Effort Estimated	Number of Records Dropped
1979	74 059	1 833	59	1985	78 113	5 924	3
1980	71 748	2 233	70	1986	73 156	4 757	8
1981	69 616	5 655	53	1987	66 609	6 166	32
1982	71 500	8 384	26	1988	55 355	5 800	104
1983	76 003	7 616	19	1989	20 336	2 444	38
1984	80 950	8 005	15				

Appendix Table 2. The number of records dropped from the analysis due to errors in the CELR data. Errors documented include the number of CELR records with efforta (number of pot-lifts in 24 hours) replaced with effortb (number of pots in the water at midnight) and the number of records dropped from the analysis because effort could not be determined

Year	Original Number of Records	Incorrect Date	Duplicates in the Effort Data Base	Duplicates in the Landing Data Base	Catch > 1000 kg	Records Combined into Another Record	Efforta replace Effortb	Missing Effort
1989	34 185	0	462	33	2	81	98	40
1990	55 110	9	523	83	3	162	148	70
1991	58 761	52	181	27	24	205	153	66
1992	57 574	63	198	111	99	179	80	130

Appendix Table 3. The number of records (summed by stratum) dropped from the analysis because catch per day > 1000 kg or catch per pot-lift > 100 and the number of records where catch = 0 and are replaced with 0.001 kg/pot-lift

Year	Catch/Day >1000	Catch/Pot-lift >100	Catch=0	Year	Catch/Day >1000	Catch/Pot-lift >100	Catch=0
1979	119	5	18	1986	36	3	17
1980	76	7	18	1987	19	3	18
1981	48	11	27	1988	13	13	38
1982	40	5	29	1989	9	7	26
1983	22	5	19	1990	9	0	10
1984	21	3	27	1991	7	0	16
1985	25	1	19	1992	10	2	15