ISSN 1175-1584

MINISTRY OF FISHERIES
Te Tautiaki i nga tini a Tangaroa

Differences in otolith measurements and gill raker counts between the two major spawning stocks of hoki *(Macruronus novaezelandiae)* in New Zealand

Allan C. Hicks
Peter J. Smith
Peter L. Horn
David J. Gilbert

New Zealand Fisheries Assessment Report 2003/7
March 2003

Differences in otolith measurements and gill raker counts between the two major spawning stocks of hoki (*Macruronus novaezelandiae*) in New Zealand

Allan C. Hicks[1]
Peter J. Smith[1]
Peter L. Horn[2]
David J. Gilbert[1]

[1]NIWA
Private Bag 14901
Wellington

[2]NIWA
P O Box 893
Nelson

New Zealand Fisheries Assessment Report 2003/7
March 2003

Published by Ministry of Fisheries
Wellington
2003

ISSN 1175-1584

Citation:
Hicks, A.C.; Smith, P.J.; Horn, P.L.; Gilbert, D.J. (2003).
Differences in otolith measurements and gill raker counts between the two major spawning stocks of hoki (*Macruronus novaezelandiae*) in New Zealand.
New Zealand Fisheries Assessment Report 2003/7. 23 p.

This series continues the informal
New Zealand Fisheries Assessment Research Document series
which ceased at the end of 1999.

EXECUTIVE SUMMARY

Hicks, A.C.; Smith, P.J.; Horn, P.L.; Gilbert, D.J. (2003). Differences in otolith measurements and gill raker counts between the two major spawning stocks of hoki (*Macruronus novaezelandiae*) in New Zealand.

New Zealand Fisheries Assessment Report 2003/7. 23 p.

Two pilot studies have been undertaken to determine if there are characteristics in hoki that could be used to differentiate between the two spawning stocks, west coast South Island (WCSI) and Cook Strait (CKST). The pilot studies used otolith ring measurements to predict the proportion of each stock in a sample, and mean number of gill rakers which showed a significant difference between WCSI and CKST for the 1997 year class. Additional year classes were sampled for this study to assess year class variation and to evaluate the preliminary results. Analysis of covariance was used on each year class to determine the effect of each stock on the regression of the first annual ring against the juvenile ring. The findings did not entirely support the preliminary results for otolith ring measurements, with only one year class showing a significant difference between WCSI and CKST. A difference in the mean number of gill rakers was observed for the 1996 and 1997 year classes, and agrees with the earlier findings, but no other year classes showed any significant differences. No definite correlations were observed between sea-surface temperature and the otolith measurements or gill raker counts, although 1996 and 1997 had high temperatures with WCSI greater than CKST. It appears that significant differences in the mean number of gill rakers and otolith measurements are present, but, due to high variation, large sample sizes would be needed to detect them. Even larger samples would be needed to accurately predict the proportion of each stock in a mixed fishery outside the spawning season.

1. INTRODUCTION

The current stock assessment of hoki (*Macruronus novaezelandiae*) in the New Zealand Exclusive Economic Zone (EEZ) uses a two-stock model (Ballara et al. 2000, Annala et al. 2002), differentiating fish which spawn off the west coast of the South Island (WCSI), called the western stock, and those that spawn in Cook Strait (CKST), the eastern stock. Accurately identifying the proportion of eastern and western fish in the Chatham Rise fishery is of interest to improve the stock assessment of hoki, but requires a tool to discriminate the two stocks.

Hicks & Gilbert (2002) showed that spawning hoki in WCSI typically have a larger distance between the juvenile otolith ring and the first annual otolith ring than the spawning hoki in CKST, and that this might be used to discriminate the two stocks. Hoki from the 1997 year class were examined extensively by Hicks & Gilbert (2002), but few otoliths from other year classes were available for the analysis to determine year-to-year variation. Small samples from the 1987, 1988, 1992, and 1994 year classes all showed that WCSI hoki have a significantly larger distance between the juvenile and first annual rings, but the 1991, 1993, and 1995 year classes showed no difference.

Smith et al. (2001) evaluated eight meristic characters to determine if any could be used to discriminate between the WCSI and CKST stocks. The number of gill rakers was the only character that showed a significant difference between eastern and western spawning hoki in the 1997 year class. The difference was small and the distributions from each stock overlapped extensively. It was also uncertain if year class variation exists because only one year class was analysed.

The work described here uses additional otolith ring measurements and gill raker counts from seven year classes (1992–98) sampled in 2001 to test the preliminary results of Hicks & Gilbert (2002) and Smith et al. (2001), to examine year class variation, and to correlate any significant results with environmental variables. The ultimate goal of this research is to estimate the proportion of WCSI and CKST fish in samples from the non-spawning fishery on the Chatham Rise. These estimated proportions could be used in the hoki stock assessment. The potential for otolith ring measurements and gill raker counts for estimating the proportion of each stock in the Chatham Rise fishery is discussed in light of the new data.

2. METHODS

The aim was to collect a minimum of 50 fish in each of the year classes 3–9 (year classes 1998–92) for each sex in both stocks. This number was chosen as a compromise between a sample being large enough to detect a significant difference and the feasibility of reading the total number of otoliths required to fill each year class with at least 100 samples. How the total number of fish to sample was determined is explained below.

2.1 Sampling design

The sampling programme was integrated with the normal collection of otoliths for the hoki stock assessment to make the collection protocol as simple as possible. Using the age and length data from the 1999–2000 fishery (Richard O'Driscoll, NIWA, pers. comm.), the predicted proportion of each year class in the 2000–01 catch was used to calculate the expected number of otoliths that would be collected for each sex and year class under several different sampling designs (Table 1). A completely random sampling design interfered least with the normal collection of otoliths and was sufficient to obtain many different year classes for both sexes.

Table 1: The predicted proportion of each year class for CKST and WCSI males and females as determined from the 1999–2000 age and length data. The oldest age is a "plus" group.

	3	4	5	6	7	8	9	10+
CKST females	0.034	0.092	0.039	0.025	0.081	0.050	0.031	0.120
CKST males	0.062	0.123	0.055	0.042	0.104	0.056	0.040	0.046
WCSI females	0.025	0.029	0.050	0.047	0.127	0.071	0.108	0.104
WCSI males	0.052	0.059	0.062	0.038	0.088	0.045	0.058	0.036

Using the most under-represented year classes between 3 and 9 for each area separated by sex, the total number of otoliths that would need to be collected to sample 50 fish, on average, of that sex and age was determined for each area. Age 3 and 4 females were predicted to be the least represented year classes in WCSI in 2000-01 and age 6 females were predicted to be weakly represented in CKST, given the 1999–2000 age and length data from shed samples collected by NIWA and the MFish Scientific Observer Programme (SOP). A total sample of 2000 otoliths from each area would need to be taken to get a sample size of at least 50 in each sex and year class on average. The sample size of different year classes (n_{Si}) that would result with a given total sample size (N) was determined to see if a threshold sample size of 50 could be obtained for most year classes using a considerably smaller total sample size. Figures 1 and 2 show which year classes would have a sample size of 50 or more for a range of total sample sizes (N).

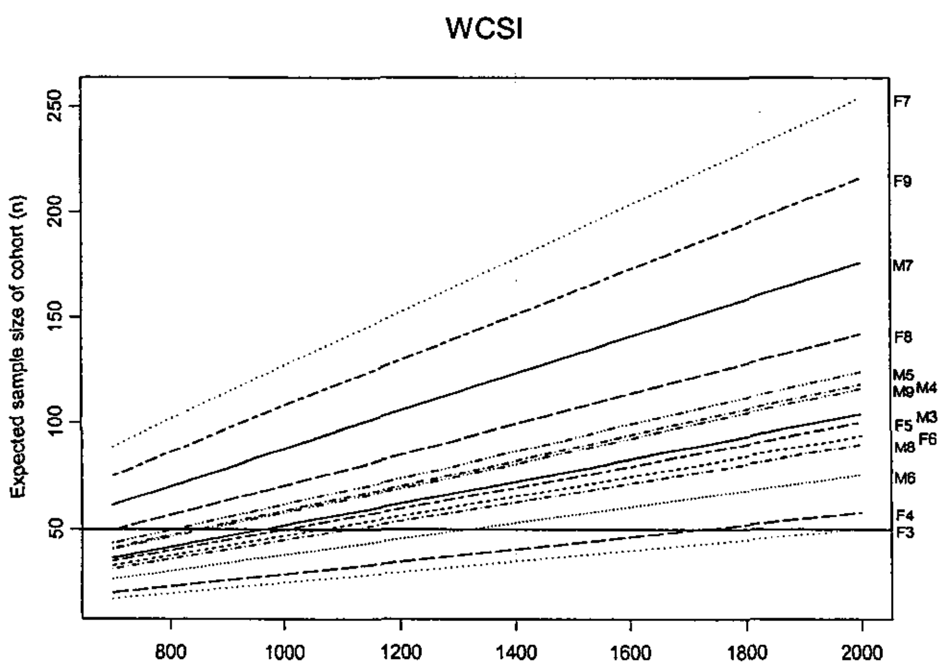

Figure 1: Expected sample size for each year class from the west coast of the South Island (WCSI) given a total sample size of otoliths from the Scientific Observer Programme in the hypothetical 2000–01 fishery. The labels on the right refer to sex and age. The horizontal line is drawn at the sample size threshold of 50.

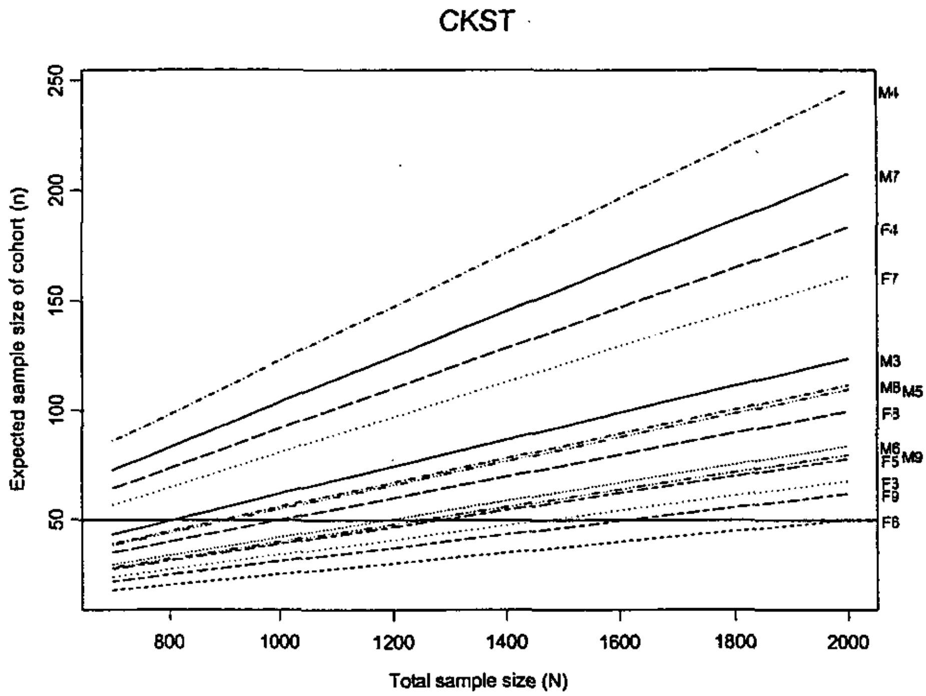

Figure 2: Expected sample size for each year class from Cook Strait (CKST) given a total sample size of otoliths from the shed sampling programme in the hypothetical 2000–01 fishery. The labels on the right refer to sex and age. The horizontal line is drawn at the sample size threshold of 50.

A considerable amount of sampling would have been needed from WCSI to effectively sample the two smallest female year classes. However, in CKST, the age 6 females were uncommon. Total sample sizes of 1700 from each area gave an expected minimum of 50 samples for both sexes and stocks in year classes 5, 7, 8, and 9. The year classes 3 and 4 from WCSI, and 6 from CKST were most likely to be short on females, but they still had expected sample sizes greater than 30, and could be used in the analysis.

WCSI otoliths are sampled by observers instructed to collect otoliths from three fish at random per tow. In 1999–2000, over 500 tows were sampled, resulting in more than 1500 otoliths being collected. NIWA staff in the shed sampling programme sampled CKST otoliths and in the 1999–2000 fishing year over 2000 otoliths were collected.

Not all of these otoliths were read, however. Eight hundred and twenty-eight otoliths from WCSI and 773 from CKST were routinely read in 2000. Using the actual recorded age of each otolith, Tables 2 and 3 show the sample sizes for each sex and age (3–9) that would have been achieved from the two areas (age 2 is not shown because they are most likely not highly selected for). These numbers are lagged one year behind the ages shown in Figures 1 and 2, because they are in reference to the 1999–2000 fishing year. An additional 94 otoliths from WCSI and 213 from CSKT would have to be read to fill each year class with a sample size of 50, assuming otoliths could be chosen in the exact year classes necessary. This may be possible for the younger fish using specific length ranges, but would be more difficult for older ages where the length distributions considerably overlap. Also, it may not be worthwhile to fill in some weaker year classes, as mentioned before, since it would take a significant amount of extra work. For example, excluding age 5 fish, only 59 and 156 extra otoliths would need to be read from WCSI and CKST, respectively.

Table 2: Actual number of otoliths from the 1999-2000 WCSI fishery that were aged for each sex and year class. The ages are in reference to the 1999–2000 fishing year.

Age	3	4	5	6	7	8	9	Total
Male	44	68	36	70	42	74	36	370
Female	42	58	29	60	34	57	43	323
Total	86	126	65	130	76	131	79	693

Table 3: Actual number of otoliths from the 1999-2000 CKST fishery that were aged for each sex and year class. The ages are in reference to the 1999–2000 fishing year.

Age	3	4	5	6	7	8	9	Total
Male	72	29	25	55	35	37	26	279
Female	79	36	18	55	32	27	22	269
Total	151	65	43	110	67	64	48	548

Therefore, to minimise the number of otoliths read and gill arches counted, and to reduce the impact on the normal otolith collecting regime, it was decided to collect and read the otoliths as has been done in the past, but to also collect the first left gill arch from each fish from which an otolith was collected. Otoliths and gill arches were collected via the SOP and NIWA shed sampling from CKST and WCSI during the spawning fisheries from June to September 2001.

First stage samples were taken from the entire collection of otoliths and gill arches of each area by randomly selecting a set number of fish from each of a series of 5 cm length bins covering the bulk of the catch and then systematically selecting additional fish to ensure the tails of the length distribution were represented. The chosen sample sizes approximated those necessary to produce mean weighted c.v.s of less than 20% across all year classes, in each of the spawning areas. These first stage samples were prepared and read using the technique of Horn & Sullivan (1996) as modified by Cordue et al. (2000). The distances to the first three annual rings (if present) and juvenile ring were measured using a micrometer eyepiece in a binocular microscope. All measurements are radial measurements (in mm) from the nucleus to the outer edge of the translucent section of a ring. The juvenile ring is a translucent band occurring inside the first annual zone (Horn & Sullivan 1996), and is likely to be formed at the time of some major life-change event. The distances to the juvenile ring (*RJ*), probably laid down 6 months after birth, and the first annual ring (*R1*) were the only otolith characteristics used in this study. Gill arches were selected from the same fish and the total number of rakers was counted on the first arch taken from the left side of the fish (Smith et al. 2001).

A second-stage sample was selected to fill in year classes that did not have a sample size of at least 50. Since the length and sex are recorded for all fish, these data from the first sample were used to create the second sample. Given the proportions of age at each length (in 1 cm bins), a length range to sample from that would result in a small total sample size, yet fill in the necessary year classes and not be too restricting on the year class of interest (i.e., at least a length range of 5 cm) was found. Second-stage samples were then taken from each length range for each area to supplement the under-represented year classes. Otoliths were read and gill rakers were counted as with the first sample.

2.2 Between-areas analysis

2.2.1 Otoliths

The analysis of the otoliths followed that performed by Hicks & Gilbert (2002). Analysis of covariance (ANCOVA) on the relationship between the distance to the first annual ring from the nucleus (*R1*) and the distance to the juvenile ring from the nucleus (*RJ*) was used to test for an *area* effect within each year class using the following method, and a Bonferroni correction that reduced the significance value to 0.007 (seven year classes being tested). First, the coefficients in Equation (1) were estimated to determine if equal slopes between the two areas could be assumed.

$$RI_i = \alpha + \beta_1 RJ_i + \beta_2 area + \beta_3 (RJ_i : area) + \ldots + \varepsilon_i \tag{1}$$

where "..." refers to other effects such as *sex* and its interactions with other variables.

If β_3 was not significant, then Equation (2) was used to test for a significant *area* effect.

$$RI_i = \alpha + \beta_1 RJ_i + \beta_2 area + \ldots + \varepsilon_i \tag{2}$$

Models with only *RJ* and *area* were analysed first, following Hicks & Gilbert (2002). The addition of *sex* was then tested, along with interactions, to determine if within–year class differences by *sex* occurred.

A multiple regression model incorporating *RJ*, *area*, *sex*, and *year class* as independent variables was also used to test the effects due to each of these variables as well as their interactions with each other.

Finally, the distance to the juvenile ring was studied to determine if differences occur between the areas. An analysis of variance was used with *RJ* as the response and *area*, *sex*, and *year class* as treatments.

$$RJ_{ijk} = \mu + \phi_i + \psi_j + \gamma_k + \phi\psi_{ij} + \phi\gamma_{ik} + \psi\gamma_{jk} + \phi\psi\gamma_{ijk} + \varepsilon_{ijk} \tag{3}$$

where ϕ, ψ, and γ refer to *area*, *sex*, and *year class* effects, respectively.

2.2.2 Gill rakers

Gill raker counts were analysed using analysis of variance with independent terms *year class*, *area*, and *sex*. All interactions were tested and the non-significant higher order interactions were discarded.

$$G_{ijk} = \mu + \phi_i + \psi_j + \gamma_k + \phi\psi_{ij} + \phi\gamma_{ik} + \psi\gamma_{jk} + \phi\psi\gamma_{ijk} + \varepsilon_{ijk} \tag{4}$$

where G_{ijk} is the number of gill rakers for fish from area i, sex j, and year class k. The effects due to *area*, *sex*, and *year class* are notated by ϕ, ψ, and γ, respectively. Each year class was also individually analysed to see what the *area* and *sex* effects were within that year class, which would equate to Equation (4) with all the terms involving γ removed.

2.2.3 Environmental factors

Sea-surface temperatures were studied to determine if any results could be attributed to this feature of the environment. Locations at 43° 0' S, 170° 0' E and 41° 30' S, 174° 30' E (Figure 3) were used to represent WCSI and CKST sea-surface temperatures respectively (obtained from B. Bull, NIWA, pers. comm.). The mean of the average temperatures for each month from June to August was used in each area as an environmental indicator.

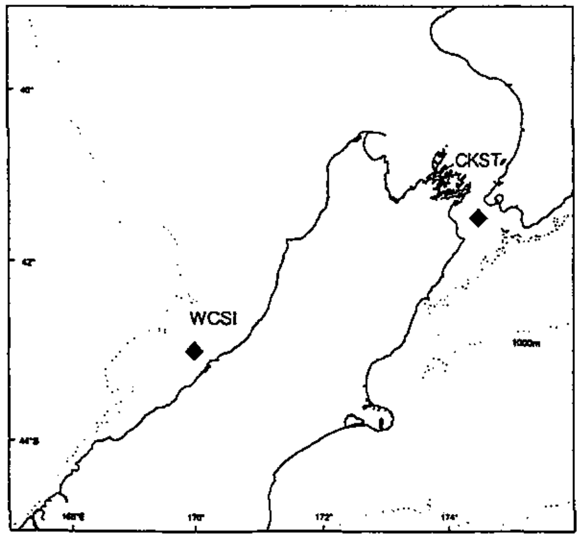

Figure 3: Locations of sea-surface temperature stations (diamonds) in relation to the hoki spawning grounds.

3. RESULTS

3.1 Samples

A total of 2749 hoki otoliths were collected from CKST, spread over 51 shed samples. Twenty-two observer trips and 11 shed samples collected 3170 otoliths from WCSI. First-stage samples of 792 and 763 otoliths from CKST and WCSI, respectively, were selected from 5 cm fish length bins. Gill arches were sampled from most fish that had otoliths sampled, but some gill arches were damaged during removal and gill rakers could not be counted. Therefore, some otoliths that were read did not have a matching gill arch, resulting in smaller sample sizes for the gill raker counts. Under-represented year classes were supplemented based on the number of otoliths read for those year classes.

A second-stage sample was done for both areas to fill in some under-represented year classes. Three year classes were under-represented from WCSI: 1993 (age 8), 1995 (age 6), and 1998 (age 3). All of the possible fish from the 1998 year class were sampled in the first stage sample, so two length ranges for males and females separately were chosen to select the 1993 and 1995 year classes. These length ranges were 53–60 and 76–83 cm for males and 55–59 and 79–86 cm for females. An additional 157 males and 121 females were sampled to supplement WCSI. Five fish from the second-stage WCSI sample were older than 9 years and were not used in this analysis. The 1995 year class (age 6) was the only year class under-represented from CKST and 97 additional males in the length range 72–80 cm and 100 additional females in the length range 72–80 cm were sampled. Two males were older than 9 years in this second-stage sample, thus were not used here. Tables 4 and 5 show the final numbers of otoliths read from each area, summarised by sex and year class

Table 4: Numbers of otoliths read from WCSI in 2001, separated by sex and year class.

Year class	1992	1993	1994	1995	1996	1997	1998	Total
Male	59	57	120	43	70	69	35	453
Female	69	58	108	40	61	54	18	408
Total	128	115	228	83	131	123	53	861

Table 5: Numbers of otoliths read from CKST in 2001, separated by sex and year class.

Year class	1992	1993	1994	1995	1996	1997	1998	Total
Male	35	64	78	48	66	74	66	431
Female	35	54	65	42	74	72	69	411
Total	70	118	143	90	140	146	135	842

Because some gill arches were damaged during collection and could not be counted, additional gill arches were sampled to fill in the under-represented year classes, and matching otoliths were read only for age. These otoliths were not used in the otolith analysis because the ring measurements were not taken since it would have taken a considerable amount of extra time. However, they were used to assign the fish, from which the additional gills were sampled, to the proper year class. Tables 6 and 7 show the numbers of gills sampled in each year class for both areas.

Table 6: Numbers of gill arches read from WCSI in 2001, separated by sex and year class.

Year class	1992	1993	1994	1995	1996	1997	1998	Total
Male	47	47	77	26	56	38	11	302
Female	108	106	134	41	66	29	5	489
Total	155	153	211	67	122	67	16	791

Table 7: Numbers of gill arches read from CKST in 2001, separated by sex and year class.

Year class	1992	1993	1994	1995	1996	1997	1998	Total
Male	26	48	59	30	45	60	58	326
Female	31	43	39	32	47	49	54	295
Total	57	91	98	62	92	109	112	621

3.2 Otolith ring measurements

In the ANCOVA used to analyse the otolith ring measurements, one year class showed unequal slopes ($p = 0.0012$) between areas. Surprisingly, this was the 1997 year class, which was analysed extensively by Hicks & Gilbert (2002) and showed equal slopes as well as a significant difference between the two areas. This difference in slopes makes it difficult to determine if an area effect is present for this year class. The 1997 year class's slope for the WCSI ring measurements was much steeper than that for the CKST ring measurements and the line was not consistently above or below CKST over the range observed. This analysis suggests there is a different relationship in that year between the two areas. Figure 4 shows the distance to the first annual ring plotted against the distance to the juvenile ring for each year class along with estimated regression lines for each area (Equation (1)). WCSI fish typically showed a steeper slope, indicated by a negative β_3 term, although only 1997 showed a significant difference in the slopes (Table 8). A high amount of variability was present in the data.

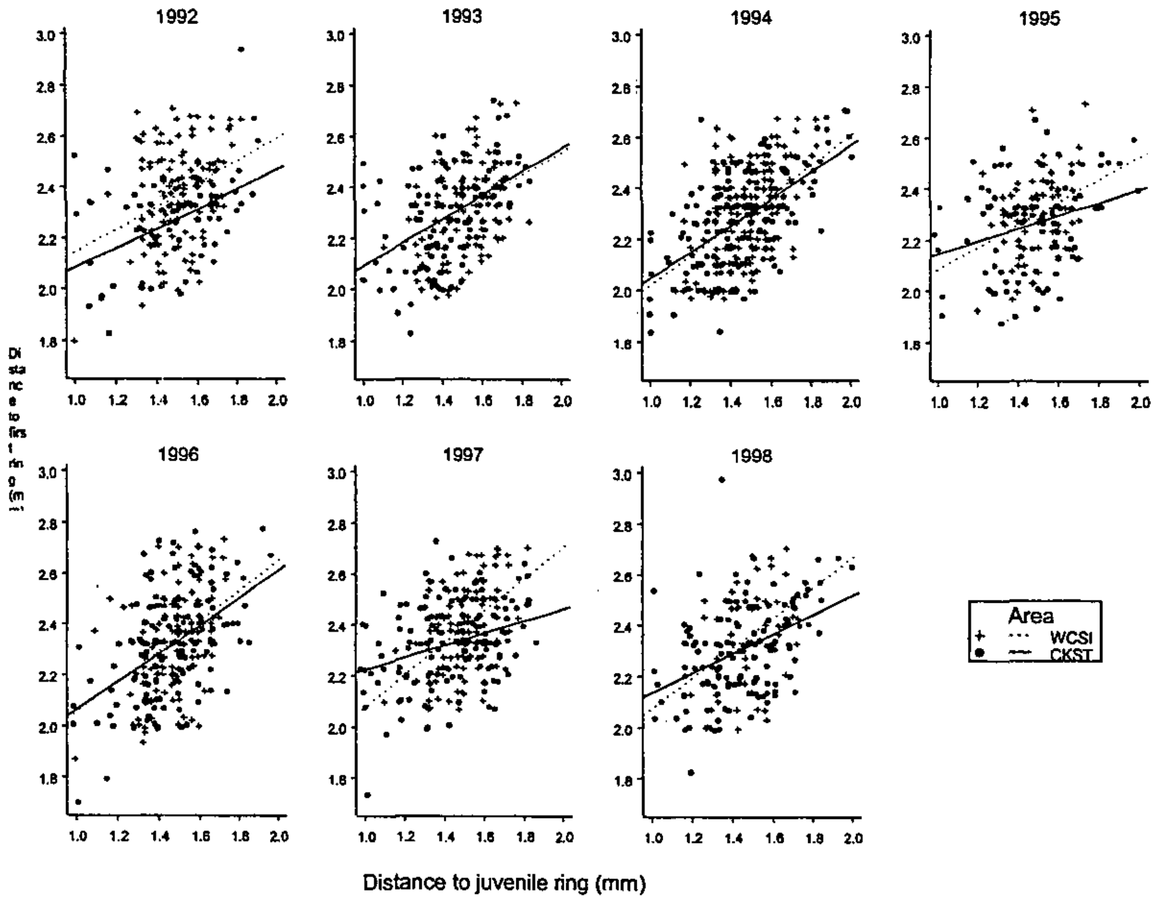

Figure 4: Plot of the distance to the first annual ring against the distance to the juvenile ring for each year class with regression lines for each area drawn in. The two areas are allowed to have different slopes.

Table 8: The regression term indicating a difference in the slope between the two regression lines (see Equation (1)). A negative value indicates that CKST has a less steep slope.

Year class	β_3 term	p-value
1992	-0.063	0.656
1993	0.010	0.941
1994	-0.067	0.479
1995	-0.188	0.260
1996	-0.044	0.762
1997	-0.390	0.001
1998	-0.206	0.275

The year classes other than 1997 did not show unequal slopes and an *area* effect could be tested using Equation (2). The 1992 year class was the only year class with a significant area effect and showed that WCSI fish have a greater distance between the juvenile ring and first annual ring (Figure 5 and Table 9). The slopes were forced equal for the 1997 year class to see what the result would be and WCSI otoliths showed a slightly significantly larger distance between the juvenile and first otolith rings (p-value = 0.0199).

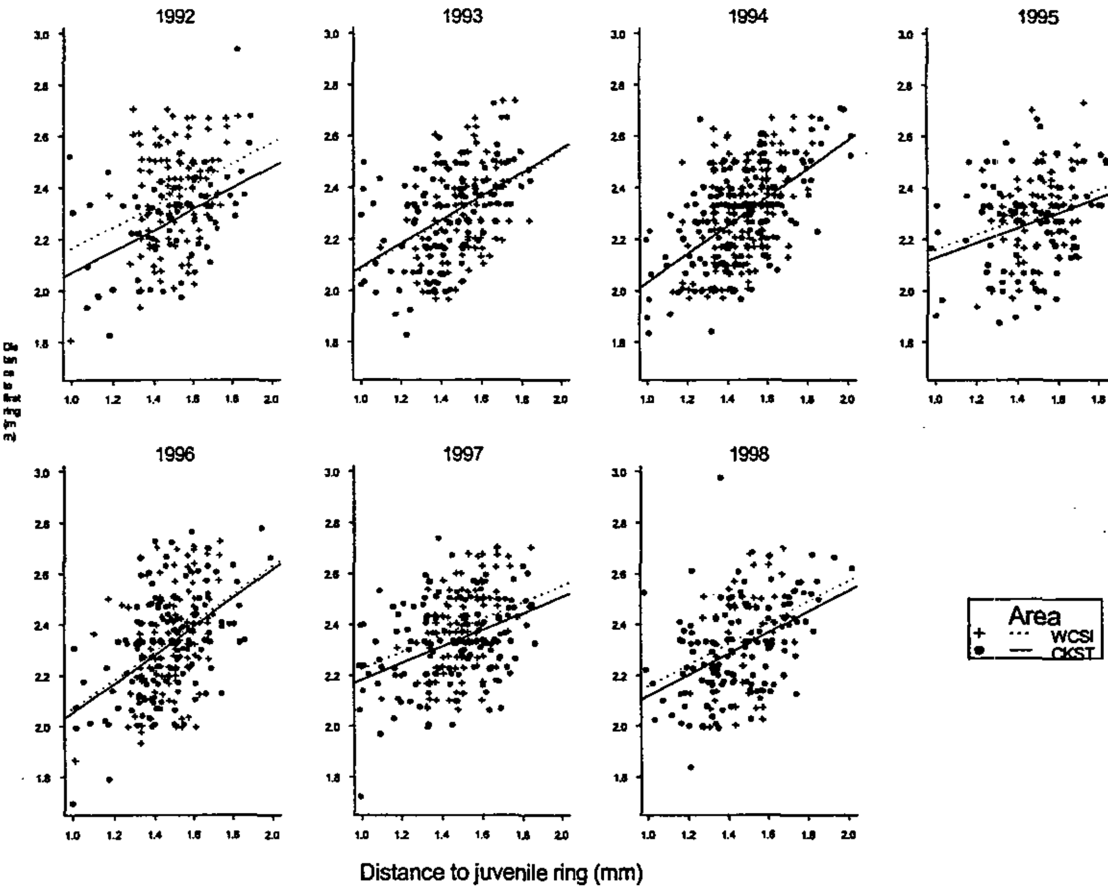

Figure 5: Plot of the distance to the first annual ring against the distance to the juvenile ring for each year class with regression lines for each area drawn in. The two areas were forced to have equal slopes

Table 9: Values for the regression terms when equal slopes are assumed (see Equation (2)). The α term is the intercept for WCSI fish, β_1 is the slope of each line, and β_2 is the difference between the WCSI and CKST intercepts (negative value indicates WCSI has a higher intercept). The p-value refers to the significance of the β_2 term.

Year class	α term	β_1 term	β_2 term	p-value (β_2)
1992	1.755	0.409	-0.095	0.0002
1993	1.622	0.459	0.009	0.6882
1994	1.490	0.548	-0.006	0.7301
1995	1.866	0.292	-0.033	0.1902
1996	1.507	0.561	-0.015	0.4782
1997*	1.899	0.326	-0.044	0.0199
1998	1.745	0.413	-0.040	0.1526

*slopes forced equal

Hicks & Gilbert (2002) analysed only the 1997 year class for sex-specific relationships and did not find a different relationship between sexes. The inclusion of *sex* as a predictor in the year class specific models was significant at the 0.05 significance level for only the 1995 year class. However, the reduced significance level of 0.007 resulted in no significant *area* or *sex* effects for this year class. With WCSI males as the control group, the intercept for every *sex* and *area* combination is shown in Table 10. Because these effects were not truly significant at the proposed alpha level, multiple comparisons were not done, but Table 10 shows that WCSI fish tend to have a larger intercept than CKST fish.

Table 10: Estimated intercepts for each sex within the two areas for the 1995 year class.

Variable	Intercept
WCSI males	1.97
WCSI females	1.89
CKST males	1.78
CKST females	1.81

All the data from the seven year classes were combined and a multiple regression model showed significant effects for all two-way interactions and a three-way interaction between *RJ*, *area*, and *sex*. This means there are differences between the two areas and sexes in terms of their intercepts and slopes. The only significant *year class* interaction was between *sex* and the 1995 *year class*, indicating that the females of the 1995 year class showed a slightly lower intercept than males from the 1995 year class, and a smaller intercept than females from other year classes (p-value=0.013). The R^2 for the model was 0.205, indicating a poor fit to the data and high variability.

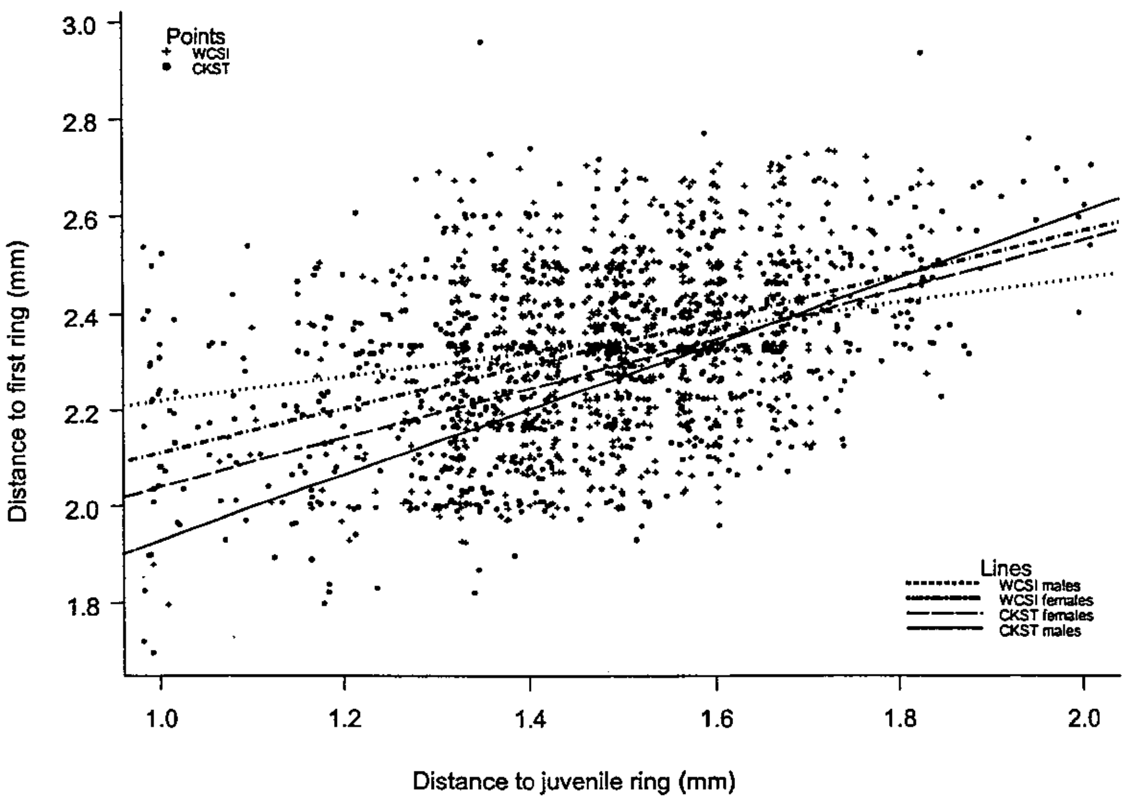

Figure 6: All of the otolith data plotted by stock. The four regression lines are WCSI males and females, and CKST males and females, as determined from the multiple regression analysis.

The interactions between the variables make it difficult to compare the two areas, but if we ignore the slightly significant 1995 *year class:sex* interaction, Figure 6 shows the four regression lines for WCSI males and females and CKST males and females, plotted against all of the data points separated by stock. The females appear to have similar slopes, with WCSI females showing a slightly larger intercept. The CKST and WCSI males show very different slopes. At the smaller distances to the juvenile ring, the males show a considerably larger distance between the two rings than females.

The expected differences to be found using the relationship of the juvenile ring and the first annual ring were not seen, thus just the distance to the juvenile ring was looked at to determine if differences occurred there. Boxplots of the distance to the juvenile ring for each sex and year class separated by each stock are shown in Figure 7. There appears to be heterogeneity in the variances between areas and possibly between sexes. Therefore, an analysis of variance weighted by the variance within an *area* and *sex* combination was used to correct for the differences in variance.

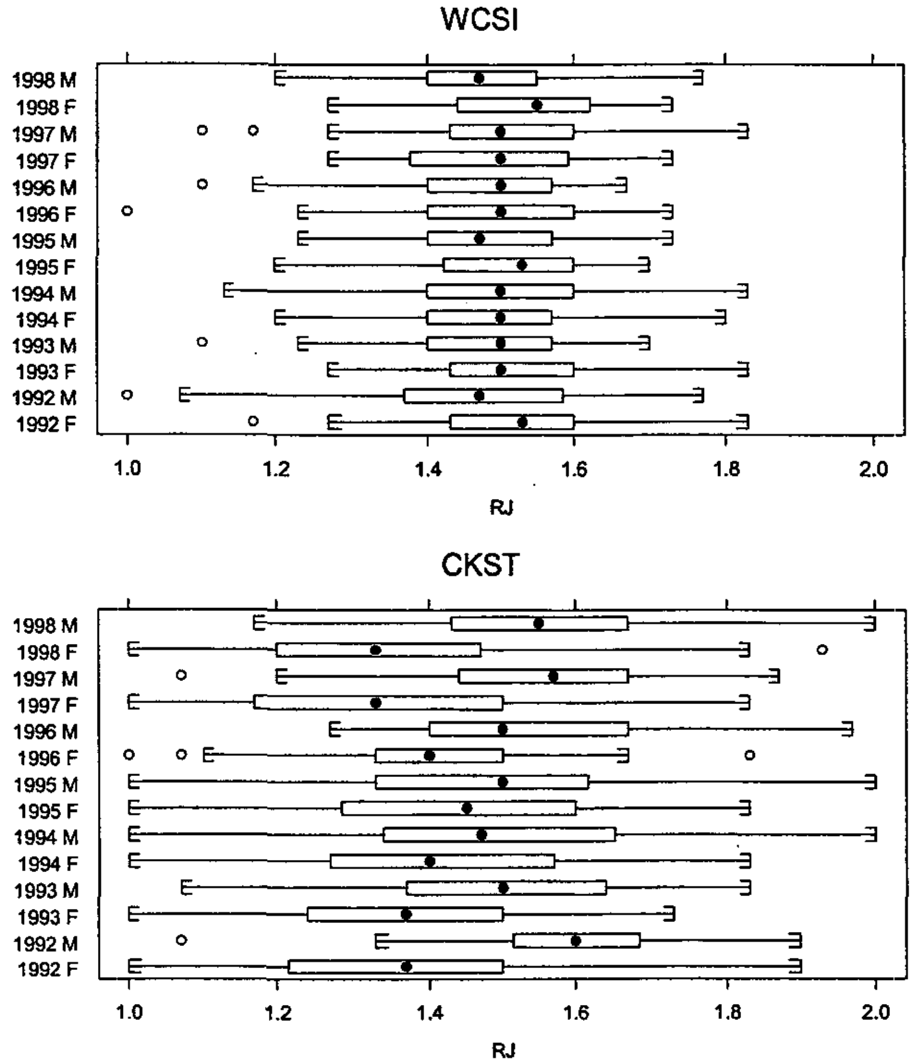

Figure 7: Boxplots of the distance to the juvenile ring (*RJ*) for each year class and sex, separated by stock. The year on the left indicates the year class; M, male; F, female.

The third-order interaction between *area*, *sex*, and *year class* was not significant, but a slight *year class* and *sex* interaction was (p-value = 0.046), suggesting there are slight year class differences. This was again due to the 1995 year class. Ignoring this slight year class interaction, and looking at only the *sex* and *area* treatments, there is a significant interaction between *sex* and *area*. The significance of the two-way interaction can be seen in Figure 7 by the differences of the sexes between the two areas. The WCSI fish do not show a significant effect between the sexes when analysed alone (p-value = 0.12). However, there is a significant difference between the sexes for the CKST fish (p-value < 0.001).

A high amount of variation was observed in the otolith measurements, some of which can be attributed to the sometimes subjective defining of the rings. When reading the otoliths, a clarity grade was given to the juvenile ring and to the first annual ring. The grade ranged from 1 to 3 where 1 is clear and 3 is highly uncertain. Tables 11 and 12 show the number of otoliths of each clarity grade for the juvenile ring and first annual ring separated by the two areas. The juvenile ring was unclear a large proportion of the time but the first annual ring typically showed better clarity.

Table 11: Numbers of otoliths of a clarity grade for the juvenile ring and the first annual ring from WCSI and each year class.

	Juvenile ring			First annual ring		
Year class	1	2	3	1	2	3
1992	25	35	68	110	18	0
1993	15	43	57	96	19	0
1994	31	66	131	184	44	0
1995	9	32	42	66	17	0
1996	11	39	81	110	21	0
1997	4	43	76	98	25	0
1998	4	23	26	33	20	0

Table 12: Numbers of otoliths of a clarity grade for the juvenile ring and the first annual ring from CKST and each year class.

	Juvenile ring			First annual ring		
Year class	1	2	3	1	2	3
1992	26	26	18	56	10	4
1993	45	45	28	89	23	6
1994	37	62	44	107	30	6
1995	21	41	28	64	21	5
1996	18	56	66	91	44	5
1997	15	58	73	92	43	11
1998	19	53	63	77	49	9

Subsets of the data using only clarity grades of 1 or 2 were not used because the samples would be significantly reduced due to the large number of unclear juvenile rings. Because the first annual ring is typically clearer, the distance to this ring was studied in a similar manner to the distance to juvenile ring to see if significant differences occurred between the areas. No effects, including any interactions, were significant.

3.3 Gill raker counts

The mean number of gill rakers per fish for CKST, without regard to year classes, was 30.01. The mean number of gill rakers for WCSI over all year classes was 29.81. The only significant term in the analysis of variance with all year classes was the *area* term (p-value=0.0006) suggesting that CKST fish have significantly more gill rakers than WCSI fish. The year classes and sexes did not show any significant differences (Table 13). The 1995 year class, which was significantly different from other year classes in the otolith analysis, showed a smaller mean number of gill rakers for the sample from CKST when compared to the other year classes (Table 13). There may be very slight year class and sex differences, but the sample sizes used here could not detect them.

Table 13: Mean number of gill rakers for WCSI and CKST fish, summarised by year class and sex, and for both sexes combined.

	WCSI			CKST		
Year class	Male	Female	Mean	Male	Female	Mean
1992	29.83	29.63	29.69	29.88	30.10	30.00
1993	30.02	29.75	29.84	29.94	30.07	30.00
1994	28.96	29.99	29.98	29.85	30.10	29.95
1995	28.85	29.85	29.85	29.73	29.94	29.84
1996	29.77	29.65	29.70	30.27	30.17	30.22
1997	29.55	29.72	29.63	30.14	30.00	30.08
1998	29.72	29.40	29.63	29.74	30.19	29.96
All year classes	29.84	29.78	29.81	29.95	30.09	30.01

A significant difference was observed between areas when using all the year classes because of the large number of gills counted from each area, giving the analysis high power. Equating the ANOVA to a two-sample t-test, the power of the test for a difference of 0.20 gill rakers was about 0.91. With these large sample sizes the test can correctly reject the null hypothesis that there is no difference in the number of gill rakers 91% of the time, if the true difference is 0.2. However, this test combined all the year classes, and even though no year class effect was found, it may be of interest to analyse individual year classes.

An ANOVA using *area* and *sex* was performed on each year class. An interaction between *area* and *sex* was not significant for any year class, and was eliminated. A *sex* effect was not significant for any of the year classes either, but a significant *area* effect was observed for the 1996 and 1997 year classes. The 1997 year class is the same year class in which Smith et al. (2001) observed a significant *area* effect with an observed difference of 0.775 in the mean number of gill rakers. The largest difference observed in this study was 0.52 in 1996, and the highest power of all the individual year class tests, assuming a true difference in means of 0.5, was 0.96, for the 1994 year class. The 1996 and 1997 year classes had powers of 0.87 and 0.91, respectively. The sample size from the WCSI 1998 year class was small, resulting in a small power (0.37 with a hypothetical difference in means of 0.5), but similar mean numbers of gill rakers for the 1997 and 1998 year classes were seen in both areas suggesting that a larger sample size may detect a significant difference between the two areas for the 1998 year class. Figure 8 shows side-by-side barcharts of the proportion of number of gill rakers for each year class of the two stocks. The overlapping of the distributions can be seen in this figure.

To verify that the significant differences observed for the 1996 and 1997 year classes was not due to the normal distribution assumptions in analysis of variance, a difference between the two areas for each year class was tested using a nonparametric Wilcoxon rank sum test. Both year classes again showed a significant difference between the areas using this test (1996: p-value = 0.0008; 1997: p-value = 0.0054). No other year classes showed significant differences between the areas using this test.

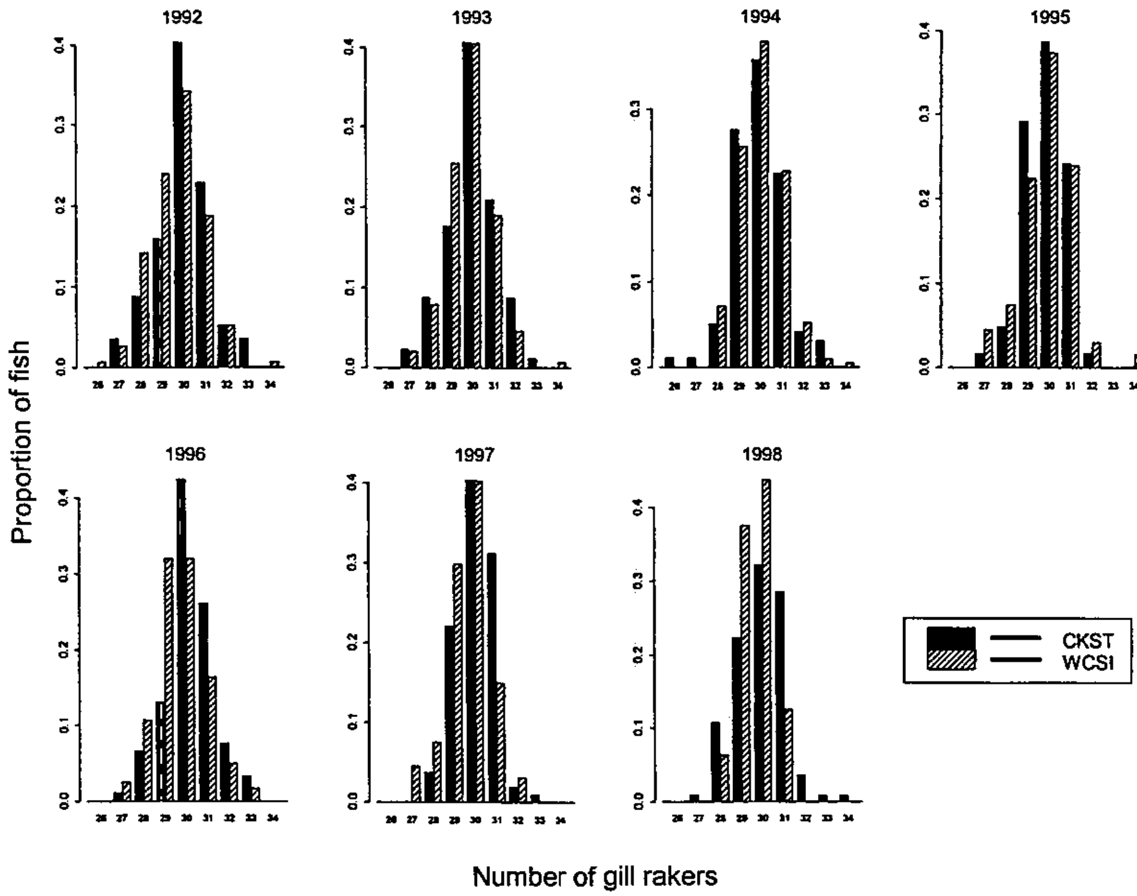

Figure 8: Proportion of the number of gill rakers for each year class of the two stocks.

3.4 Sea surface temperatures

Sea surface temperature data were analysed to test if the differences in otolith measurements and gill raker counts were driven by environmental conditions in specific years. The 1992 year class was the only year class that showed a valid significant difference between the two spawning areas using the otolith ring measurements and the 1996 and 1997 year classes were the only year classes to show a significant difference between the mean numbers of gill rakers. The WCSI sea surface temperature in 1992 showed the lowest temperature for this time series, while the CKST temperature was somewhat high, resulting in a large difference between the two spawning area temperatures (Figure 9 and Table 14). The sea surface temperatures in 1997 were high in both spawning areas and 1996 showed a large difference between the two areas.

The 1995 year class was different from the other year classes when using otolith ring measurements and showed the smallest mean number of gill rakers in CKST. The sea surface temperatures observed in this year marginally showed the largest difference between the two spawning areas and CKST had the largest temperature in its series from 1992 to 1998 (Table 14). However, 1992 also showed a large difference in temperatures and the CKST sea-surface temperature was only slightly less than that in 1995, but this year class did not appear different than any other year classes, besides 1995. In fact, the 1992 year class was the control year class to which all others were compared when testing the significance of each year class individually, for which the 1995 year class typically showed a significant difference.

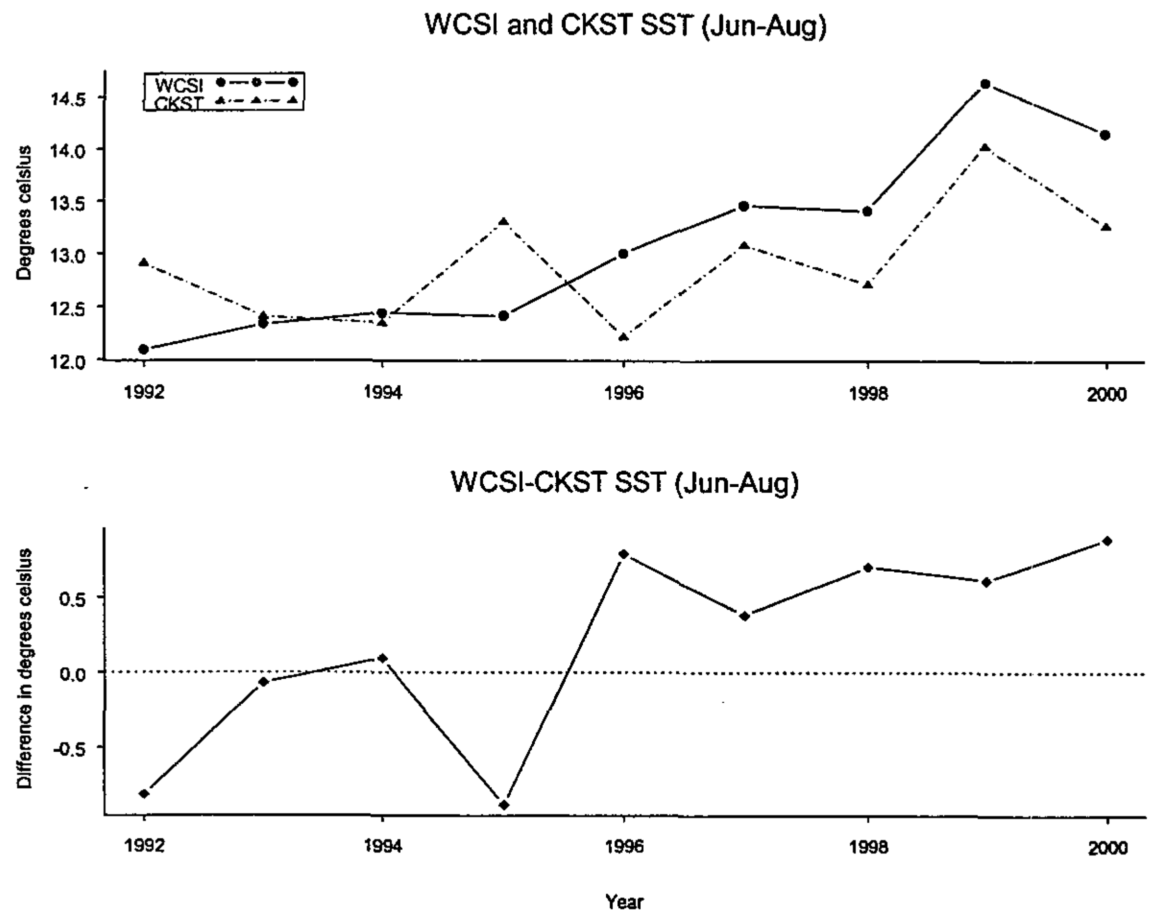

Figure 9: Sea surface temperatures for WCSI and CKST averaged from June to August. The top graph shows the temperatures from each area and the bottom graph shows the difference between the WCSI temperature and the CKST temperature.

Table 14: Sea surface temperatures for the two hoki spawning areas. Each index is the mean of the average sea-surface temperatures for June, July, and August.

Year	CKST	WCSI
1992	12.9	12.1
1993	12.4	12.3
1994	12.3	12.4
1995	13.3	12.4
1996	12.2	13.0
1997	13.1	13.5
1998	12.7	13.4

4. CONCLUSIONS

These results did not completely validate the preliminary otolith findings of Hicks & Gilbert (2002), but did agree with the gill raker results of Smith et al. (2001). Hicks & Gilbert observed significant differences between CKST and WCSI in three year classes (1992, 1994, and 1997), but the 1992 year class was the only year class to show a significant difference between the areas in this study. The 1997 year class produced contradictory results. It is possible that Hicks & Gilbert (2002) would have reached different conclusions if their sample sizes had been larger. The other year classes in this analysis which did not show a significant difference between stocks (1993–98) showed a much smaller effect between the two areas than was observed by Hicks & Gilbert (2002). Smith et al.

(2001) reported a significant gill raker difference for the 1997 year class, which was also observed here, but the 1996 year class was the only other year class to show significant differences between the areas. These results indicate that there are meristic differences between the hoki spawning in the two different areas for specific year classes.

The multiple regression analysis on the otolith ring measurements from all year classes showed a slight difference between the 1995 year classes and the other year classes. *Sex* and *area* differences were observed as significant. At the smaller distances to the juvenile ring, WCSI fish had a larger distance between the two otolith rings. However, the CKST males had a larger distance between the otolith rings than WCSI fish of either sex. Given the sex differences observed here and the possibility that year classes may be different, the techniques shown by Hicks & Gilbert (2002) to discriminate between stocks seem unlikely to be satisfactory. Both year class and sex differences would have to be removed, and a large number of fish would be required to obtain a reliable estimate of the stock effect.

The results from the gill raker counts agreed with those of Smith et al. (2001) for the 1997 year class, but did not show as large a difference between CKST and WCSI in the mean number of gill rakers (0.45 observed here compared to 0.775 observed by Smith et al. (2001)). Fewer fish were analysed for the 1997 year class here and a significant difference was still observed. The 1996 year class was the only other year class to show a significant difference between areas. It is possible that 1996 and 1997 were unusual years or that the sample sizes were too small to accurately detect a significant difference for the other year classes, although the powers of the tests were relatively high for most of the year classes given a hypothetical difference of 0.5, except 1998 which had a small number of gills sampled from WCSI fish.

The otolith characteristics and gill raker counts are likely to be influenced by environmental conditions experienced by individual fish, and may not be consistent between years. Therefore differences in sea surface temperatures between the two areas were examined. Similar differences in sea surface temperatures were observed in 1992 and 1995, with the WCSI colder than CKST (see Figure 9). The 1992 and 1995 year classes showed significant differences from each other in the otolith analysis, and only the 1992 year class showed a significant difference in the otolith ring measurements between the areas suggesting that factors other than sea surface temperature are affecting the otolith development.

Sea surface temperatures were higher in WCSI than CKST in 1996, 1997, and 1998 (see Figure 9). The 1996 and 1997 year classes showed significant differences in the mean number of gill rakers between areas, but the sample sizes from the 1998 year class in WCSI were small and it is unlikely that the test could detect a significant difference (although the mean number of gill rakers was similar for the 1998 and 1997 year classes in each area). It is possible that differences in gill raker counts can be detected when WCSI sea surface temperatures are greater than the CKST sea surface temperatures, in which case the 1996–2000 year classes would also be expected to show a difference between areas (see Figure 9).

The previous analyses of Hicks & Gilbert (2002) and Smith et al. (2001) focused on the 1997 year class as 3-year olds, which were sampled from a length distribution that accurately identified them as belonging to that year class. This study relied on accurate ageing to determine the year class to which the hoki belonged. However, Horn & Sullivan (1996) used a within-reader test to show that about 70% of their hoki otoliths were aged identically in two readings, with most of the remainder differing by +/- 1 year. Also, Francis (2003) has estimated that 3.3% of hoki otolith rings are misinterpreted (either ignored or counted as two rings and this compounds with increasing age). Ageing error would cause some fish to be incorrectly assigned to a year class, resulting in an increase in variance and a loss in the power of the test. Since year class differences were observed for the otolith readings, ageing error may have resulted in the non-significant results seen. Year class variation was not as prevalent in the gill rakers counts, thus ageing error may not have as much effect.

The clarity of the rings is a source of variation that should not be overlooked and may partially account for the non-significant results seen in the otolith analysis. The juvenile ring is the most likely ring to show differences between the two areas, but is also the ring with the highest number of poor clarities (see Tables 11 and 12). Using only otoliths with better clarity ratings would substantially reduce the already small sample sizes, and if discrimination were possible for high clarity otoliths only, the method could only be used to discriminate between fish if it were assumed that the low clarity fish had the same proportions by stock as the high clarity fish. This may not be a reasonable assumption. Gill raker counts, on the other hand, are not as subjective, and the only errors that could occur were if some rakers were missing or some rakers were small and not counted.

Spatial heterogeneity seems to be a possible explanation of some of the results described. If the juvenile ring locations and the gill raker counts of fish from the same catch were correlated they would therefore not be equivalent to a random sample of the same size from the whole population and the significance levels would be incorrect. However, few fish were sampled from each tow in an attempt to alleviate this problem and Hicks & Gilbert (2002) performed a nonparametric test to see if the observed effect for the 1997 year class was an artefact of the tows or landings sampled, and determined that the *area* effect was real and not a result of sampling certain tows. Appendix A describes a test that was used to determine if significant variation is present between tows and should be accounted for. No cases showed any significant tow-by-tow variation and the above significance levels are not influenced by tow-by-tow variability.

There are three related hypotheses that can be tested with the current data, in relation to the current two-stock model (Ballara et al. 2000, Annala et al. 2002). Firstly, that there are no stock differences in the characters measured here, and the differences observed occurred by chance. This hypothesis is unlikely because large sample sizes in previous analyses (Smith et al. 2001, Hicks & Gilbert 2002) showed significant differences in gill raker counts for the 1997 year class, which was also observed in the gill raker analysis here, and the 1992 year class showed a significant area effect in the previous (Hicks & Gilbert 2002) and current otolith analyses. In addition, the analyses using the entire datasets of otoliths or gill rakers showed significant differences between the two areas. It could have been that performing multiple tests increased the error rate and some tests were incorrectly declared as significant. However, the gill raker counts from the 1996 and 1997 year classes would still be declared significant even when lowering the overall level of significance to 0.007, based on seven comparisons. We reject this hypothesis because of the consistency of the gill raker analyses, the agreement of the 1992 year class in the otolith analysis, and the observance of multiple significant results below the chosen statistical error rate of 0.05.

The second hypothesis is that there is infidelity between the two stocks. If hoki do not return to the spawning ground where they were born, any environmentally determined characters that differ between the fish born in different areas would be mixed, making it difficult, or even impossible, to distinguish between two samples of age 3+ hoki taken in different areas. Genetic differences would not be expected to be observed if mixing occurred. Similar arguments to those used to reject the first hypothesis, along with correlations to sea surface temperatures, reduce the likelihood of this hypothesis. Because significant differences were seen in gill raker counts for the 1996 and 1997 year classes, when sea surface temperatures were higher on the west coast of the South Island than in Cook Strait, it may be that the meristic difference is due, at least partly, to environmental factors. This suggests that infidelity is low, at least for the 1996 and 1997 years. Unfortunately, few gills were available from the WCSI 1998 year class for the current analysis, making it difficult to detect a significant difference between the areas in that year, when sea-surface temperatures were similar to 1997. Higher sea surface temperatures for WCSI have also been observed in 1999 and 2000 (see Figure 9), two year-classes that were not studied here. The analysis of additional fish from the 1998 year class, and new samples from the 1999, and 2000 year classes would allow further testing of the infidelity hypothesis. Nevertheless, the current results show that infidelity is unlikely, at least in some years.

The third hypothesis is that sample sizes of 100 fish per year class and area will satisfactorily detect differences between the two stocks using either otoliths or gill raker counts. This hypothesis assumes that differences do occur, but does not state how large those differences are, and does not imply any usefulness from the differences. The large sample sizes used by Hicks & Gilbert (2002) and Smith et al. (2001) showed significant differences and led to the collection of the data for this study. Sample sizes of 50 for each year class, sex, and area were chosen because it would have been unreasonable to collect and read more. The smaller sample sizes for the 1997 year class did not detect the significant otolith difference observed by Hicks & Gilbert (2002) between the two areas, and the significant difference observed by Hicks & Gilbert (2002) for the 1994 year class was not observed with the larger sample sizes in this study. However, a significant difference in the mean number of gill rakers was seen for the 1997 year class, as was reported by Smith et al. (2001), and the powers of the tests for area differences were reasonably high for year classes with sample sizes near 100. Using the variability observed in the 1996 and 1997 year classes, and a hypothetical area difference of 0.5 between the mean number of gill rakers, sample sizes of 100 from each area would result in powers of 0.86 and 0.94. The sample sizes used here are not large enough to satisfactorily detect significant differences between the two stocks using otolith measurements because consistent results were not seen and a large amount of variablilty was present, which may be partly due to ageing error. If otolith ring measurements were to be used to discriminate between stocks, much more effort would be needed to collect and process larger samples from the spawning sites. On the other hand, sample sizes of 100 were adequate to detect significant differences in the mean number of gill rakers because consistent results were seen, the year classes with significant differences corresponded to years where WCSI sea surface temperatures were higher than CKST, and the powers of the test were reasonably high given the variability observed and an expected true difference of 0.5. This hypothesis refers to detecting a significant difference between the spawning areas. Trying to predict the proportion of each stock in a non-spawning area sample would require larger sample sizes (Hicks & Gilbert 2002), which were not determined.

We interpret the results above to indicate that there are differences in the mean number of gill rakers and in otolith ring measurements between the two spawning stocks of hoki. However, these differences are small and vary among year classes. The results also suggest that a high degree of fidelity is maintained for each stock, at least in some years. Furthermore, the differences in gill raker counts between the two stocks appear to be related to the differences in water temperatures and therefore may not occur each year. Further collection and analysis of hoki from non-spawning areas to predict the proportion of each stock in the samples was not done because few valid training data sets could be formed from the spawning area samples. This does not mean that predicting the proportion of each stock in a sample, as described by Hicks & Gilbert (2002), is not possible, but large multi-year class samples would be required to detect a difference and the efficacy of using an overall mean to discriminate fish from a particular year class would be low. Given the limited area differences found with both otolith ring measurements and gill raker counts, these tools are currently not practical for estimating the proportions of CKST and WCSI fish on the Chatham Rise.

5. ACKNOWLEDGMENTS

We thank Chris Francis for advice on the sampling design, Richard O'Driscoll for supplying the age and length data from the 1999–2000 fishing year, and Brian Bull for quickly sending us sea surface temperatures. The efforts of Margaret McVeagh counting gill rakers were essential to this study and are appreciated, as was the help and patience that Susan Kim provided during the analyses.

We also thank Malcolm Francis for reviewing a draft of this report and supplying comments for its improvement. The Ministry of Fisheries is appreciated for funding this research under project code HOK 2001/05.

6. REFERENCES

Annala, J.H.; Sullivan, K.J.; O'Brien, C.J.; Smith N.W.McL.; Varian, S.J.A. (comps.) (2002). Report from the Fishery Assessment Plenary, May 2002: stock assessments and yield estimates. 640 p. (Unpublished report held in NIWA library, Wellington.)

Ballara, S.L.; Cordue, P.L.; Livingston, M.E. (2000). A review of the 1997–98 hoki fishery and assessment of hoki stocks for 1999. *New Zealand Fisheries Assessment Report 2000/8*. 65 p.

Cordue, P.L.; Ballara, S.L.; Horn, P.L. (2000). Hoki ageing: recommendation of which data to routinely record for hoki otoliths. Final Research Report to the Ministry of Fisheries for Research Project MOF1999/01C (Hoki ageing). 24 p. (Unpublished report held by Ministry of Fisheries, Wellington.)

Francis R.I.C.C (2003). Analyses supporting the 2002 stock assessment of hoki. *New Zealand Fisheries Assessment Report 2003/5* 34 p.

Hicks A.C.; Gilbert D.J. (2002). Stock discrimination of hoki (*Macruronus novaezelandiae*) based on otolith ring measurements. *New Zealand Fisheries Assessment Report 2002/2*. 31 p.

Horn, P.L.; Sullivan, K.J. (1996). Validated aging methodology using otoliths and growth parameters for hoki (*Macruronus novaezelandiae*) in New Zealand waters. *New Zealand Journal of Marine and Freshwater Research 30*: 161–174.

Kirk R. (1995). Experimental design: procedures for the behavioral sciences 3rd ed.. Brooks/Cole Publishing Company, Pacific Grove. 919 p.

Smith P.; Bull B.; McVeagh M. (2001). Evaluation of meristic characters for determining hoki stock relationships. Final Research Report for Ministry of Fisheries Research Project HOK 1999/05 Objective one. 10 p. (Unpublished report held by Ministry of Fisheries, Wellington.)

Venables W.N; Ripley B.D. (1997). Modern applied statistics with S-PLUS. Second edition. Springer, New York. 548 p.

APPENDIX A: HIERARCHICAL DESIGN USING TOWS AS THE NESTED VARIABLE

To correctly analyse the gill rakers and otoliths using linear models, it should be taken into account that the design can be interpreted as hierarchical with tows nested within each area. With this design, it is possible to test if there is a high amount of variability between tows. If that is the case, then using a design that does not account for this variation is biased and will reject the null hypothesis more often than expected.

The otolith analysis used a linear mixed effects model because of the covariate *RJ* (Venables & Ripley 1997). Here the data are seen as clustered within *tows*, which is a random effect, as opposed to the fixed *area* effect. A significant variability among tows was tested with all the data and a model incorporating *RJ*, *area*, *sex*, and *year class*, then tested for each year class individually (correcting for multiple comparisons) with *RJ*, *sex*, and *area* effects. An *F*-test with expected mean squares of between-tow and within-tow variability was used to test if a hierarchical model was appropriate or the more parsimonious models reported above are sufficient. The expected mean squared error for the between-tow variability is the sum of the estimated within-tow variance and the estimated between-tow variance. The *F*-test is then simply the ratio of the two values with degrees of freedom n_{tows}-2 and N-n_{tows}.

$$F_{n_{tows}-2,N-n_{tows}} = \frac{\sigma_{\varepsilon}^2 + \sigma_{tows}^2}{\sigma_{\varepsilon}^2}$$

A σ^2_{tows} near zero indicates that the between-tow variability is negligible.

The gill rakers were analysed in a more straightforward manner following conventional and hierarchical analysis of variance techniques (Kirk 1995). A significant between-tow variation requires the mean squared error from the tow effect to be used in the test. If the between-tow variation is negligible, then the mean squared error term is equivalent and a standard ANOVA can be used. All the data were used with, *area*, *sex*, and *year class* effects, then each year class was tested separately with *area* and *sex* effects (correcting for the 7 comparisons). Only the significance of the between-tow variation was tested.

In all the analyses, the tow-by-tow variation was very small and never significant. A hierarchical model did not improve the analysis and taking tows into account is not necessary. Therefore, the results reported above can be taken as unbiased, at least with respect to tow-by-tow variation.