ISSN 1175-1584

MINISTRY OF FISHERIES
Te Tautiaki i nga tini a Tangaroa

Use of biological data in orange roughy stock assessments

R. I. C. C. Francis
D. M. Tracey

New Zealand Fisheries Assessment Report 2000/42
October 2000

Use of biological data in orange roughy stock assessments

R. I. C. C. Francis
D. M. Tracey

NIWA
PO Box 14 901
Wellington

New Zealand Fisheries Assessment Report 2000/42
October 2000

Published by Ministry of Fisheries
Wellington
2000

ISSN 1175-1584

Citation: Francis, R.I.C.C. & Tracey, D.M. 2000:
Use of biological data in orange roughy stock assessments.
New Zealand Fisheries Assessment Report 2000/42. 30 p.

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

EXECUTIVE SUMMARY

Francis, R.I.C.C. & Tracey, D.M. 2000: Use of biological data in orange roughy stock assessments.
***New Zealand Fisheries Assessment Report 2000/42.* 30 p.**

Length frequency, sex ratio, and gonad stage data for orange roughy reside in three databases: *obs_lfs*, for commercial data collected at sea by observers; *market*, for commercial data collected at the point of landing; and *trawl*, for research data. Together they contain length and sex data for 1.5 million fish and gonad stages for 0.5 million fish. The way these data are currently used in stock assessment models is described.

Three overlapping subsets of the data were examined: non-random tow data, market sample data, and that from stratified random trawl surveys. For each subset, annual time series were constructed of mean length, proportion male, proportion mature, and length at 50% maturity. These time series were evaluated against five criteria: duration, comparability, precision, consistency, and relevance.

Two time series not currently used in assessments were identified as of potential use: mean lengths from non-random tows (both commercial and research) in parts of the Spawning Box (north Chatham Rise), and also from random trawl surveys on the Challenger Plateau.

However, most of the data were found not to be useful as model inputs. The main reason is that the data collection systems were not designed for this purpose (except the series of random trawl surveys). That is, they do not ensure that the same areas are sampled every year in the same season, and much sampling (the choice of trips to sample and the location of fishing) is not at random. This is not necessarily a fault because these data are collected for a range of purposes (e.g., compliance, bycatch monitoring, conversion factor estimation).

Despite their lack of usefulness in population modelling, these data are of great use in the informal parts of stock assessments (e.g., the evaluation of model inputs and the interpretation of results). They also have many other research uses, including the planning of trawl and acoustic surveys and as background information for research proposals.

This document is a final report on work carried out under objective 5 of Ministry of Fisheries project ORH199902: *To determine if the length frequency, sex ratio, and reproductive data for orange roughy collected at sea from commercial and research vessels can be incorporated into the orange roughy stock assessment model.*

1. INTRODUCTION

The work fulfils objective 5 of Ministry of Fisheries project ORH199902: *To determine if the length frequency, sex ratio, and reproductive data for orange roughy collected at sea from commercial and research vessels can be incorporated into the orange roughy stock assessment model.*

We first describe the extent of these data (in the three databases in which they reside) and their current use in stock assessments. We then state the criteria they must meet in order to be useful in assessments, and evaluate the data with regard to these criteria.

2. DATA

The data in question come from three sources:

- the Scientific Observer Programme, which collects data at sea from commercial fishing boats;
- so-called "market" samples, collected under the Stock Monitoring Programme from landed commercial catches; and
- research voyages.

In this section we briefly describe the extent of these data, and their current use in stock assessments

2.1 Observer data

The orange roughy data in the observer database, *obs_lfs*, included (on 10 May 2000) 324 200 length measurements by sex, and gonad stages for 129 497 females (males are not staged by observers). These came from 3704 tows from 306 different observer trips spread over 14 areas. In most areas coverage has been patchy over time (Table 1).

2.2 Market sampling data

Market sampling data for ORH (in database *market*) included (on 11 August 2000) 74 263 length measurements by sex, taken from 153 landings, each landing corresponding to a commercial fishing trip. There are no gonad stage data. A cluster sampling scheme was used to select fish for measurement, where each cluster was a fish container: usually a fish box (with about 30 fish to a box), sometimes a fish bin (which holds several hundred fish). A random sample of containers was selected and a sample of fish (usually all of them — *see* Section 4.2.2) from each selected container was sexed and measured. In one landing (which covered QMAs 7A and 7B), the catch was stratified (by QMA) and cluster samples were taken from each stratum).

The data are rather limited, with no times series of any length (Table 2). This is because it was decided in the early 1990s that sampling for this species was better done at sea. More sampling details for pre-1993 samples were given by Fisher & Banks (1991, 1993). Some of the data were analysed in detail by Field (1990) (the first two years' data) and by Field (1991) (the 2A, 2B, and 3A data from 1989–90).

Table 1: Coverage of observer sampling of orange roughy up to 1998–99: number of trips and number of tows, by fishing year and area. Area codes are as used in database *obs_lfs*; most correspond to FMAs, as shown, with the exceptions HOWE = Lord Howe Rise, LOUR = Louisville Ridge, TMAR = South Tasman Rise. Trips may cover more than one area and more than one year, so the trip row 'ALL' is not the sum of the other rows (and similarly for columns). –, no samples

Number of trips

Area:	?*	AKE	AKW	CEE	CEW	CHA	HOWE	KER	LOUR	SEC	SOE	SOI	SOU	SUB	TMAR	ALL
FMA:		1	9	2	8	7		10		3	4	6A	5	6		
Year																
1986–87	7	–	–	–	–	–	–	–	–	–	4	–	–	–	–	11
1987–88	9	–	–	–	–	2	–	–	–	8	–	–	–	–	–	15
1988–89	1	–	–	–	–	10	–	–	–	13	9	–	–	–	–	26
1989–90	–	–	–	7	–	8	–	–	–	6	5	–	–	–	–	21
1990–91	–	–	–	1	–	–	–	–	–	1	9	–	–	–	–	10
1991–92	–	–	1	–	–	5	2	–	–	5	7	–	1	–	–	17
1992–93	–	–	–	2	–	1	–	–	–	18	5	–	1	–	–	26
1993–94	–	1	–	7	–	4	–	–	1	16	12	–	5	–	–	36
1994–95	–	–	–	6	–	2	–	–	8	1	15	1	1	2	–	30
1995–96	–	7	–	1	–	2	–	–	1	5	5	–	2	1	–	20
1996–97	–	3	–	3	–	3	–	–	–	5	5	2	3	3	–	16
1997–98	–	2	2	2	–	4	2	1	1	–	9	4	3	3	4	25
1998–99	–	5	1	3	1	4	5	–	–	7	11	9	11	13	3	46
ALL	17	18	4	32	1	45	9	1	11	85	96	16	27	22	7	294

Number of tows

Area:	?*	AKE	AKW	CEE	CEW	CHA	HOWE	KER	LOUR	SEC	SOE	SOI	SOU	SUB	TMAR	ALL
FMA:		1	9	2	8	7		10		3	4	6A	5	6		
Year																
1986–87	29	–	–	–	–	–	–	–	–	–	22	–	–	–	–	51
1987–88	109	–	–	–	–	6	–	–	–	42	–	–	–	–	–	157
1988–89	1	–	–	–	–	128	–	–	–	156	181	–	–	–	–	466
1989–90	–	–	–	45	–	52	–	–	–	26	142	–	–	–	–	265
1990–91	–	–	–	12	–	–	–	–	–	7	116	–	–	–	–	135
1991–92	–	–	1	–	–	20	6	–	–	22	166	–	3	–	–	218
1992–93	–	–	–	3	–	3	–	–	–	183	92	–	4	–	–	285
1993–94	–	2	–	29	–	26	–	–	6	93	185	–	58	–	–	399
1994–95	–	–	–	74	–	13	–	–	67	6	162	5	2	2	–	331
1995–96	–	68	–	13	–	14	–	–	6	22	44	–	9	3	–	179
1996–97	–	27	–	19	–	48	–	–	–	27	94	2	21	9	–	247
1997–98	–	11	6	9	–	66	5	2	7	–	133	23	4	8	22	296
1998–99	–	44	17	40	1	86	28	–	–	37	100	24	16	27	63	483
ALL	139	152	24	244	1	462	39	2	86	621	1437	54	117	49	85	3512

* area not specified

Table 2: Coverage of market sampling of orange roughy up to August 2000: number of landings sampled by area and fishing year. Some landings covered more than one area so the column 'ALL' is not the sum of each row. Howe = Lord Howe Rise; CET = Challenger Plateau, beyond the EEZ; BPLE = Bay of Plenty; –, no samples

	QMA or area									
Fishing year	7A	7B	Howe	CET	2A	2B	3A	3B	BPLE	ALL
1987–88	18	–	–	–	–	–	–	–	–	18
1988–89	–	–	–	–	21	5	7	–	–	40
1989–90	–	2	–	–	30	4	8	2	–	35
1990–91	3	3	1	–	36	5	7	–	–	54
1991–92	–	–	–	–	1	1	–	–	–	2
1994–95	–	–	–	–	–	–	–	–	2	2
1996–97	–	–	–	2	–	–	–	–	–	2

2.3 Research data

The orange roughy data in the research database, *trawl*, included (on 10 May 2000) 867 981 length measurements, and gonad stages for 354 737 fish (both male and female). These came from 8426 tows in 121 trips (or voyages) since 1980. Of these trips, 116 listed orange roughy as a main species.

Coverage was best in (in descending order) FMA 4 (Chatham Rise), FMA 7 (includes the Challenger fishery), FMA 2 (includes the Ritchie and Wairarapa fisheries), and FMA 3 (including Kaikoura), and was poor elsewhere (Table 3).

Table 3: Number of research trips, by calendar year and area, for which orange roughy was one of the main species. Some trips covered more than one area so the column 'ALL' is not the sum of each row. Classification by area is approximate because *trawl* database areas do not correspond precisely to FMAs; year specified is that for the start of each trip

	Area (numbers relate to FMAs)											
Year	?*	1	2	3	4	5	6	7	8	9	HOWE	ALL
1980	0	0	0	0	1	0	0	0	0	0	0	1
1981	1	1	1	0	3	0	1	2	1	1	0	5
1982	0	0	0	1	2	0	0	1	0	0	0	4
1983	0	0	0	0	0	0	0	4	0	0	0	4
1984	0	0	2	0	1	0	0	4	0	0	0	7
1985	0	3	5	0	1	0	0	1	3	3	0	7
1986	0	2	4	2	1	0	0	3	2	2	0	10
1987	0	0	1	0	2	0	0	4	0	0	0	7
1988	0	0	1	1	5	0	0	2	0	0	0	8
1989	0	0	1	4	6	0	0	6	0	0	1	15
1990	0	0	1	3	3	0	0	1	0	0	0	6
1991	0	0	0	0	1	1	1	1	0	0	0	3
1992	0	0	1	2	3	2	2	0	0	0	0	6
1993	0	0	2	2	1	0	1	1	1	0	0	5
1994	0	1	1	1	1	1	1	0	0	0	0	4
1995	0	1	2	0	4	0	0	0	0	0	0	7
1996	0	0	1	1	2	0	0	0	0	0	0	3
1997	0	0	2	0	3	0	0	0	0	0	0	3
1998	0	1	1	0	4	0	0	0	0	0	0	6
1999	0	2	0	0	3	0	0	0	0	0	0	5
Total	1	11	26	17	47	4	6	30	7	6	1	116

* area not specified

Some of these trips were designed as stratified random trawl surveys, which covered the same area, at the same time of year and using identical (or similar) gear, over at least three years. There are three such series (Table 4).

Table 4: Series of stratified random trawl surveys targetted at ORH

Area	Month(s)	Voyage codes
Spawning Box	July	BUC8401,BUC8501,BUC8601, BUC8701,COR8801, COR8901, COR9002,TAN9206,TAN9406
Challenger Plateau	July	ARR8401,AEX8701,AEX8801, AEX8901,WIL9001
Mid-East Coast	March/April	TAN9203, TAN9303, TAN9403

2.4 Use in current assessments

In NIWA assessments of orange roughy, the only biological data that are currently used as model inputs (i.e., as data to be fitted by the model) are mean lengths for the Mid-East Coast (MEC) and northeast Chatham Rise (NECR) stocks (Table 5). Mean lengths were weighted by catch (and, for the NECR data, by stratum areas). Also, they were calculated as the average of the mean male and mean female lengths, and so are unaffected by changes in the sex ratios of samples. In both assessments the mean lengths were treated as *relative* indices, so that model inference was based only on the trend in mean lengths, and not their actual values.

Table 5: Sources (research voyages or market samples) and sample sizes (*n* = number of landings or tows) for mean lengths used in NIWA stock assessments of the Mid-East Coast (MEC) and northeast Chatham Rise (NECR) stocks. All data were from the spawning season, either in ORH 2A (for MEC) or the Spawning Box (for NECR). For more information *see* Field *et al.* (1994) and Francis *et al.* (1992). –, no data

	MEC		NECR	
Year	Source	*n*	Source	*n*
1984	–	–	BUC8401	112
1985	–	–	BUC8501	109
1986	GAL8603	21	BUC8601	120
1987	ARR8701	36	BUC8701	108
1988	–	–	COR8801	121
1989	market	12	COR8901	112
1990	market	15	COR9002	114
1991	market	19	–	–
1992	–	–	TAN9206	113
1993	TAN9306	50	–	–
1994	–	–	TAN9406	122

There are two non-NIWA assessments that use orange roughy biological data as model inputs. In their assessment of the NECR stock, Hilborn *et al.* (2000a) used length frequencies, by sex, from the NECR research surveys of Table 5, and from observer data; for the East Cape (EC) stock, Hilborn *et al.* (2000b) used the same type of data from a 1995 egg survey (TAN9507, Zeldis *et al.* 1997) and from observer data (Table 6). There are three ways in which these assessments differed from the NIWA assessments in their use of length data: they used full length histograms, rather than just mean lengths; they treated the lengths as absolute, rather than relative; and sex ratio information was passed to the model.

Table 6: Summary of observer data used in assessments by Hilborn *et al.* (2000a, 2000b) of the northeast Chatham Rise (NECR) and East Cape (EC) stocks. "Percentage of catch" was calculated as the total catch of sampled tows (from Hilborn *et al.* 2000a,b) expressed as a percentage of the total catch for the fishing year from either the spawning box or ORH 2A North (as given by Annala *et al.* 2000); for NECR, the number of trips is from Starr (pers. comm.). –, no data used

	NECR			EC		
Fishing year	Number trips	Number tows	Percentage of catch	Number trips	Number tows	Percentage of catch
1986–87	2	16	3	–	–	–
1988–89	7	105	14	–	–	–
1989–90	8	104	11	–	–	–
1990–91	5	58	23	–	–	–
1993–94	–	–	–	5	20	16
1994–95	–	–	–	2	12	7
1995–96	2	24	38	1	12	7
1996–97	3	15	9	3	11	2
1997–98	3	42	38	–	–	–
1998–99	–	–	–	1	5	0

3. DATA REQUIREMENTS

In this section we define five criteria which determine whether a particular subset (or subsets) of these data could be useful in stock assessments: duration, comparability, precision, consistency, and relevance.

A major requirement is that the data must give us information about how some quantity (e.g., mean length or percent male) changes over time (from year to year). Thus the first criterion for usefulness is duration: the data must extend over several (preferably many) years. We must be able to reduce the data to one or more annual time series. It is not usually necessary that the time series be complete (we don't need data for every year of the fishery) but, generally speaking, the longer the time series the better.

The second criterion of usefulness is comparability. To discover whether our data are comparable we must ask, very specifically, what they represent. For example, we may have a time series of mean lengths of females from the commercial catch from ORH 2A in the spawning season. Another series may be of percent male in the recruited population in the Spawning Box in July. The series is comparable if it represents the same quantity in every year. A series of mean lengths from data that were collected in different areas and seasons in each year is of no use to us. This criterion is very important. As will be seen below, it severely restricts the amount of the data sets described in Section 2 that are useful in stock assessments.

A third criterion concerns precision: we must be able to calculate accurate standard errors (*s.e.*s) for our time series. That is, we must know how different our time series is likely to be from the series which it estimates. We need these *s.e.*s to measure the significance of any trends in our series. For example, if a series of mean lengths changes by 0.5 cm over several years then this is clearly significant (i.e., it tells us about a real change in the catch or fish population) if each mean length has an *s.e.* of 0.1 cm, but it is not significant if the *s.e.*s are 1 cm. Strictly speaking, we can calculate *s.e.*s only when the data collection is based on statistical sampling procedures, which means there must be an element of randomness in the sampling. When sampling is not random we must analyse the data as if it were random. A key issue is that we need to infer something about the extent of variability in a large set of numbers (e.g., the lengths of *all fish* caught in a given area over a given time period) from a smaller set (e.g., the lengths of *all sampled fish* from that area in that time). We will say more about the problem of calculating *s.e.*s in the context of specific data sets below.

Our fourth criterion, consistency, applies when we have more than one time series for a given stock. For example, we may have mean lengths from two or more subareas within the stock. If all these series show similar trends (i.e., are consistent) we can have some degree of confidence that the trends are real (i.e., representative of the larger stock), and not just a product of unrepresentative sampling. The various series can then be combined, through some sort of averaging, to make a single series as an input to a stock assessment model. Of course, a lack of consistency does not mean that the series are unrepresentative. It is quite possible that different trends could occur in different parts of a stock. However, we would need a great deal more information than is currently available to allow an extension of our stock assessment models to cope with such disparate trends.

The last criterion is relevance. Our aim in assessing stocks is to determine current stock status (current biomass, relative to virgin or target biomass) and sustainability (is the TACC, or recent catches, likely to move the stock biomass towards, or maintain it near, the target biomass?). Thus the quantities we are primarily interested in are biomass and recruitment (the latter is relevant to sustainability). There is little point in using data series that contain no information about either of these quantities. Their use might make our models somewhat more realistic, but it would not materially change our stock assessments. Mean length data are of interest (i.e., relevant) for what they might tell us about recruitment patterns. For sex ratios to be relevant they must show either a strong imbalance or a marked trend. With proportion mature data, only a marked trend would be useful. If seen in the fish population this might imply changes in recruitment; in catches it could suggest a change in fishing practice that should be modelled.

With regard to the first three criteria it is clear that the ideal data source (amongst those considered here) is stratified random trawl surveys (*see* Table 4). These are designed to cover the same area each year, at about the same time of year, and using the same (or very similar) fishing gear. Also, trawl positions are chosen at random, as are fish that are sampled from the catches. The observer and market databases are more difficult to use because they are not designed primarily for stock assessment purposes. The many factors which influence the decision as to which trips are sampled (e.g., compliance monitoring, marine mammal bycatch, conversion factor estimation, the availability of vessels fishing in a target area) vary from year to year. Within an area and season, the trips that are sampled are not selected at random. Also, the constraints of work schedules and other duties mean that the tows which are sampled within a trip are not selected strictly at random. Thus it is relatively rare to find comparability over many years and there are difficulties in the calculation of *s.e.s* from these data (see below).

4. DATA EVALUATION

In this section we evaluate the usefulness of the data described in Section 2 using the criteria of Section 3. We treat the data in three subsets: the non-random tow data, the market samples, and the trawl survey data. It will be seen that the first and third data sets overlap.

4.1 Non-random tow data

The non-random tow data is by far the largest subset of the data. It includes all the observer and research data, but excludes the market sampling data (because these contain no information on individual tows). Most of the tows were not at randomly selected positions (hence the descriptor "non-random"). Data from tows whose positions were selected at random (e.g., those from stratified random surveys) are here used in ways that ignore their random origin.

Lack of comparability is a major problem with these data. Suppose, for example, we have an annual time series of mean lengths from commercial catches, and that this shows a clear increasing trend. Does this mean that the mean length of the fish population has increased over time? If so, this could be a useful input to a stock assessment model. However, this trend may be just a sampling artifact. This could happen, for example, if spawning aggregations contain larger than average fish, and it happens that more of the samples in the later years came from the spawning period. Or there could be a spatial pattern, with more of the later samples coming from areas known to contain larger fish.

4.1.1 Hotspots

In an attempt to address the comparability problem we considered time series constructed from data gathered from approximately the same area and time of year. A set of 15 small areas ("hotspots") subject to heavy fishing were defined (Figure 1). There was a trade-off between making each area small enough so that geographical variation (in mean length or sex ratio, say) within each area would be minimised, but large enough to contain sufficient data to create a time series. The only temporal restriction was to distinguish between spawning (June and July) and non-spawning months. Within each hotspot and time period we refrained from mixing observer and research data on the grounds that commercial and research fishing practices are very different. This meant that we could, in principle, construct up to four time series for each hotspot (research/spawning, research/non-spawning, observer/spawning, observer/non-spawning).

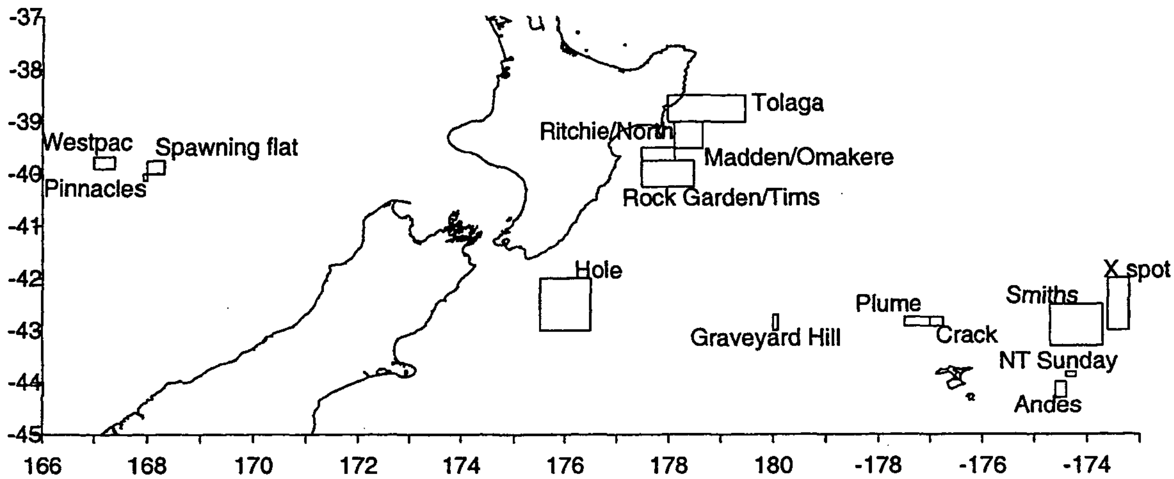

Figure 1: Rectangles defining 15 hotspots used in selecting data series. See Table A1 for rectangle boundaries.

In addition, we applied the following constraints:

- tows for which catch weight or position data were missing were rejected;
- catches of less than 1 t (spawning) or 100 kg (non-spawning) were rejected;
- years with fewer than 10 tows were rejected; and
- time series with fewer than 3 years data were rejected.

The restriction to catches of more than 1 t, for the spawning months, was to ensure that the catch came from the main spawning aggregation; for the non-spawning period, catches of less than 100 kg were considered non-representative of the bulk of the catch. The third constraint relates to our precision criterion (so that we could assess between-tow variability), and the fourth to the duration constraint. Further constraints were applied for specific types of data.

4.1.2 Data selection

Four types of data were selected: mean length, sex ratio (proportion male), proportion mature, and proportion mature by length (i.e., a maturity ogive). Different orange roughy gonad staging systems are used for male or female fish, and by researchers and observers. However, in all systems a fish is deemed immature at stage 1, and mature for stages greater than 1 (classification of large immature fish is difficult so some will be erroneously recorded as stage 2, but this is unlikely to be of major concern in large data sets).

The first two data types used exactly the same data because (overall) mean length was always calculated as the simple average of the mean lengths of males and females. (This method of calculation is now standard practice in orange roughy assessments. The reason for taking a simple average, rather than weighting by the proportions of males and females, is to avoid confusion between changes in length and sex ratio. With a weighted mean, a change in sex ratio, without any change in the mean lengths of either sex, would appear as a change in the overall mean length).

For mean length and proportion male, we ignored tows where fewer than 20 orange roughy were measured and sexed, on the grounds that smaller samples would not well represent the whole catch. With this constraint, together with those in Section 4.1.1, we were able to construct 11 time series (Table 7). These ranged in length from 3 years to 10 years but most were short (median 4 years).

Table 7: Descriptions of the 11 time series of mean length and proportion male that were constructed from the non-random tow data. The series names identify the hotspot name (as in Figure 1), the data source (res = research, obs = observer), and the period (sp = spawning, nonsp = non-spawning). For observer time series "% coverage" is the percentage of the total commercial catch in the given hotspot and period that was included in the observer data. Years are fishing years (1984 = 1983–84). – , not available

Series	Description										
Plume res sp											
	Year	1984	1985	1986	1987	1988	1989	1990	1992	1995	1998
	No. of tows	15	12	22	36	16	38	11	10	25	35
	No. of trips	2	1	1	1	1	1	1	1	2	3
Plume obs sp											
	Year	1989	1990	1991	1996	1998					
	No. of tows	51	28	21	20	24					
	No. of trips	3	4	3	1	2					
	% coverage	34	13	38	53	54					
Crack res sp											
	Year	1984	1985	1986	1987	1995					
	No. of tows	19	15	13	17	14					
	No. of trips	2	1	1	1	2					
Crack res nonsp											
	Year	1980	1987	1988							
	No. of tows	10	12	13							
	No. of trips	1	1	1							
Crack obs sp											
	Year	1989	1990	1991	1998						
	No. of tows	32	20	11	10						
	No. of trips	4	4	3	2						
	% coverage	31	17	16	46						
Smiths obs nonsp											
	Year	1989	1991	1992	1994	1995					
	No. of tows	10	15	16	11	28					
	No. of trips	1	2	1	4	8					
	% coverage	–	8	37	39	16					
Andes obs nonsp											
	Year	1992	1993	1994	1995	1997	1998	1999			
	No. of tows	42	13	68	47	13	25	10			
	No. of trips	5	4	9	6	3	2	3			
	% coverage	16	11	34	28	20	17	14			
Ritchie/North res sp											
	Year	1986	1987	1993							
	No. of tows	16	16	11							
	No. of trips	3	1	1							
Spawning flat obs sp											
	Year	1988	1989	1990	1992						
	No. of tows	35	23	25	13						
	No. of trips	4	6	6	4						
	% coverage	8	7	20	27						
Pinnacles obs sp											
	Year	1988	1989	1990							
	No. of tows	17	56	20							
	No. of trips	3	6	1							
	% coverage	14	26	17							
Westpac obs sp											
	Year	1988	1989	1997							
	No. of tows	24	25	12							
	No. of trips	4	2	1							
	% coverage	30	85	51							

The number of tows per year ranged from 10 to 68 (median 17). There were typically very few trips per year (median 2). For series deriving from observer data the coverage of the catch ranged from 7% to 85%, with a median value of 23%.

Proportion mature data were selected separately for males and females. As with the mean length data, we ignored tows with fewer than 20 data. This is a more stringent constraint than for mean length because we required 20 fish by sex, and not all measured and sexed fish are staged. Thus we were able to construct fewer time series for proportion mature: six series, and all were for females. As with those for mean length and proportion male, these series typically contain few years, and there are usually few trips per year (Table 8).

Table 8: Descriptions of the six time series of proportion mature that were constructed from the non-random tow data. The series names identify the hotspot name (as in Figure 1), the data source (res = research, obs = observer), and period (sp = spawning, nonsp = non-spawning), and sex (M = male, F = female). For observer time series, "% coverage" is the percentage of the total commercial catch in the given hotspot and period that was included in the observer data. Years are fishing years (1984 = 1983–84). –, not available

Series	Description						
Plume res sp F							
	Year	1992	1995	1998			
	No. of tows	10	25	35			
	No. of trips	1	2	3			
Plume obs sp F							
	Year	1989	1990	1991	1996	1998	
	No. of tows	39	16	21	20	24	
	No. of trips	3	3	3	1	2	
	% coverage	27	7	38	53	54	
Crack obs sp F							
	Year	1989	1990	1991	1998		
	No. of tows	26	18	11	10		
	No. of trips	4	3	3	2		
	% coverage	25	15	16	46		
Smiths obs nonsp F							
	Year	1989	1991	1992	1995		
	No. of tows	10	14	15	26		
	No. of trips	1	2	1	8		
	% coverage	–	8	34	16		
Andes obs nonsp F							
	Year	1992	1993	1994	1995	1997	1998
	No. of tows	41	13	65	47	13	25
	No. of trips	5	4	9	6	3	2
	% coverage	16	11	32	28	20	17
Spawning flat obs sp F							
	Year	1989	1990	1992			
	No. of tows	14	25	13			
	No. of trips	3	6	4			
	% coverage	3	20	27			

The maturity-at-length data were also selected separately by sex. In selecting these data we did not apply any constraint on the minimum sample size by tow. Instead, we required that there be data for least 50 mature and 50 immature fish (of the appropriate sex) in each year of each series. The reason for this different selection criterion is that we did not weight these data by the catches. Francis & Clark (1998) found, in fitting spawning ogives, that this weighting can lead to stability problems, particularly where there are one or two outstandingly large catches in a sample. The constraint relating to 50 fish was imposed because samples in which all but a few fish are mature (or immature) are often inadequate to define a maturity ogive.

With the above constraint (and those in Section 4.1.1) we were able to construct just three time series. These range in length from 3 years to 6 years and are all for females from observer data in the non-spawning months (Table 9).

Table 9: Descriptions of the three time series of maturity ogives that were constructed from the non-random tow data. The series names identify the hotspot name (as in Figure 1), the data source (res = research, obs = observer), and period (sp = spawning, nonsp = non-spawning), and sex (M = male, F = female). For observer time series, "% coverage" is the percentage of the total commercial catch in the given hotspot and period that was included in the observer data. Years are fishing years (1984 = 1983–84). – , not available

Series	Description						
Smiths obs nonsp F							
	Year	1991	1992	1994	1995		
	No. of tows	15	16	10	27		
	No. of trips	2	1	4	8		
	% coverage	8	37	39	16		
Andes obs nonsp F							
	Year	1992	1993	1994	1995	1997	1998
	No. of tows	39	13	52	46	13	24
	No. of trips	5	4	9	6	3	2
	% coverage	16	11	32	28	20	17
Graveyard Hill obs nonsp F							
	Year	1994	1995	2000			
	No. of tows	13	10	11			
	No. of trips	6	4	2			
	% coverage	13	19	–			

4.1.3 Simple statistics

The following simple statistics were calculated for the data just selected.

For each tow in the Table 7 series we calculated the mean length by sex and the proportion male (by number). Standard errors for these statistics were calculated in the usual ways (the standard deviation divided by the square root of the sample size, for the mean lengths; and using the binomial formula for proportion male). Mean lengths by sex, and proportions male, were then calculated, for each year of each of these series, as weighted means across all tows in each year, where the weights used were the catch weights for each tow, either for the whole catch (for mean proportion male) or for the estimated weight by sex (for the mean lengths). The estimated weight by sex was taken as the total catch weight multiplied by the proportion (by number) of that sex.

Analogous calculations were made, for the Table 8 series, of the proportion mature by sex for each tow (and its *s.e.*) and for each year.

The overall mean length for each Table 7 tow (or year) was calculated as the simple mean of the associated mean lengths by sex, and its *s.e.* as the square root of half the sum of squares of the *s.e.*s for the two sexes.

Logistic maturity ogives were fitted (by maximum likelihood) to the combined data for each year of each series in Table 9, and L_{mat}, the length at which 50% of fish are mature, was calculated from each ogive.

We have shown how to calculate *s.e.*s for the statistics for each tow, but not for the statistics for each year of a series. Before we address the latter it is instructive to examine the nature of heterogeneity in the selected data.

4.1.4 Heterogeneity in the selected data

It would be useful if we could think of all tows within the same trip as simple random samples from the same population. A simple informal way of investigating whether this is a reasonable assumption is to pick a trip at random and compare a given statistic (mean length, proportion male, or proportion mature) from two randomly selected tows from that trip. If our assumption were correct we would expect that, on average, the absolute difference between the two statistics would be about equal to the average of the two associated *s.e.*s. However, when we tried this experiment we found that the median absolute difference was more than twice that big for mean length and proportion male, and more than 1.5 times that big for proportion mature. [In doing this experiment we ignored tows in which all fish were of the same sex or maturity and trips with fewer than 10 such tows]. Thus, we cannot simply lump together all data from the same trip because fish from different tows are, on average, more different than those from the same tow.

We next considered between-trip variation. When the tow statistics are plotted by trip, it is apparent that there is, for some years and series, substantial between-trip variation (Figures 2 & 3). Consider, for example, the mean lengths for 1994 in the series "Smiths obs nonsp". Although two of the four trips that were sampled contributed only one tow each, it seems that each trip had its own characteristic mean length. There are many other examples of obvious between-trip heterogeneity amongst the tow statistics. Consider, for example, years 1990 and 1998 for proportion male in the "Plume obs sp" series (Figure 2) or almost every year for proportion mature in the "Andes obs sp" series (Figure 3).

We evaluated the statistical significance of this sort of heterogeneity using one-way analyses of variance (ANOVA). A separate ANOVA was carried out for each year in each series for which there was more than one trip (trips with only one tow were ignored for these tests). The data used in the ANOVA were the tow statistics, and the factor we were testing for significance was trip. For mean length, the test was significant ($P < 0.05$) in 16 of the 30 combinations of year and series for which a test is possible. The corresponding results for proportion male and proportion mature were 10 significant out of 30, and 8 significant out of 21, respectively (these significant results are marked as asterisks after the year in Figures 2 & 3). We conclude that tows from different trips are, on average, more different than those from the same trip. To view these data as coming from a single collection of tows, without any reference to the trips they came from, is to miss an important source of heterogeneity.

It is not possible to use the preceding techniques to examine between- and within-trip heterogeneity in the maturity ogive data because there were usually not sufficient data to fit an ogive for each tow. We will show (in Section 4.1.5.2) that the ogives are not well determined because of considerable between-trip heterogeneity in these data. For the moment we present a single example of this heterogeneity, in which the proportion mature varied from about 0.1 to 0.9 over four trips, without substantial variation in the length frequencies (Figure 4).

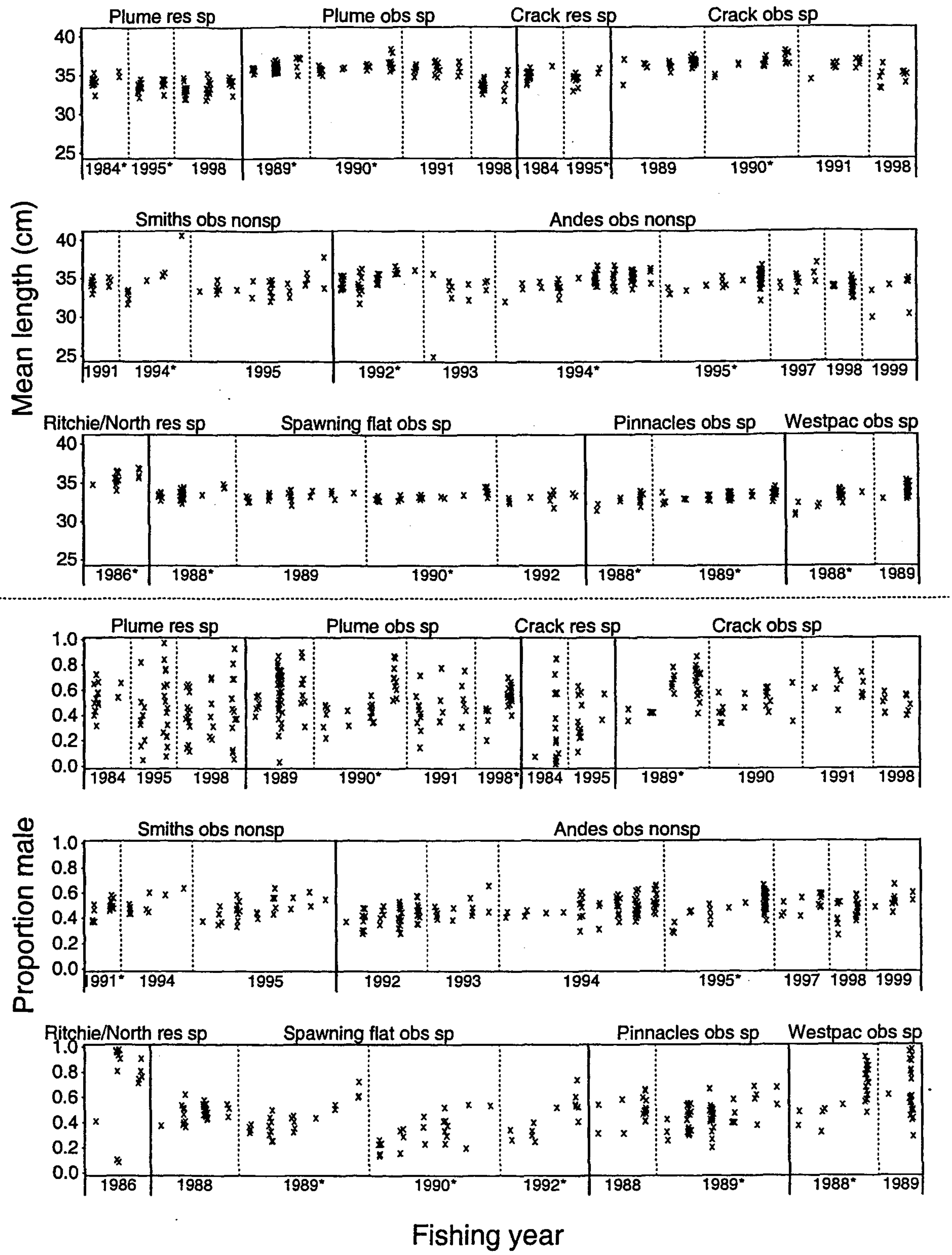

Figure 2: Mean length (upper panels) and proportion male (lower panels) by tow, plotted by trip for each series in Table 7. Each 'x' represents one tow, and tows from the same trip are plotted on the same vertical line. Years with only one trip in a series are omitted; an asterisk after a year indicates significant between-trip heterogeneity (*see* text). Vertical broken lines separate trips by fishing year; vertical solid lines separate trips from different series.

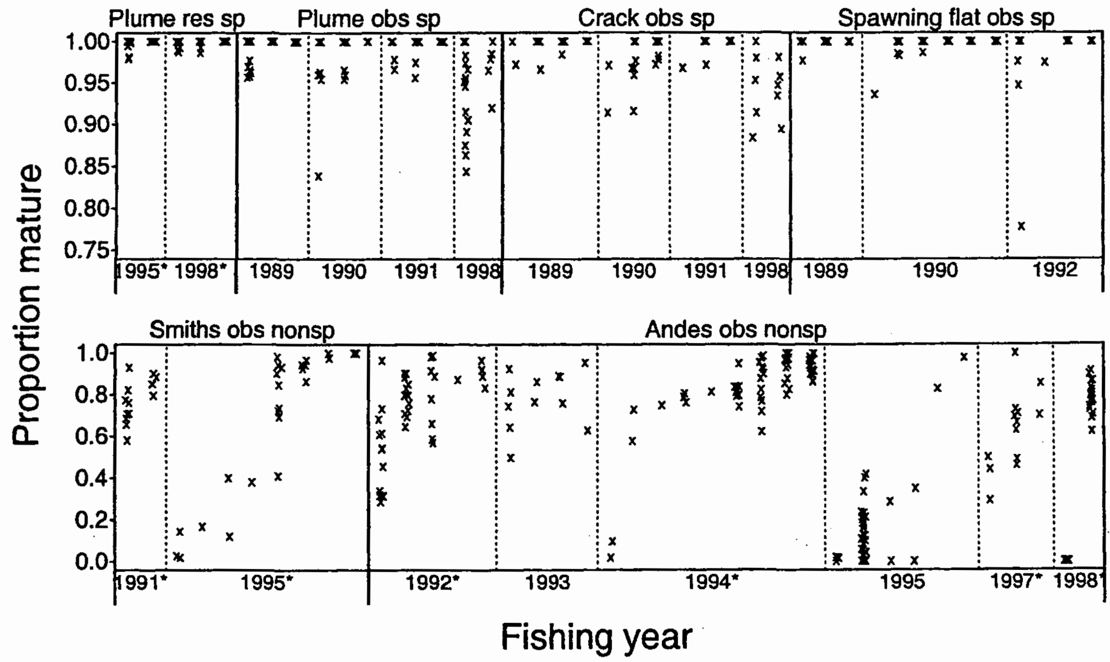

Figure 3: Proportion mature by tow, plotted by trip for each series in Table 8. Each 'x' represents one tow, and tows from the same trip are plotted on the same vertical line. Years with only one trip in a series are omitted; an asterisk after a year indicates significant between-trip heterogeneity (*see* text). Vertical broken lines separate trips by fishing year; vertical solid lines separate trips from different series.

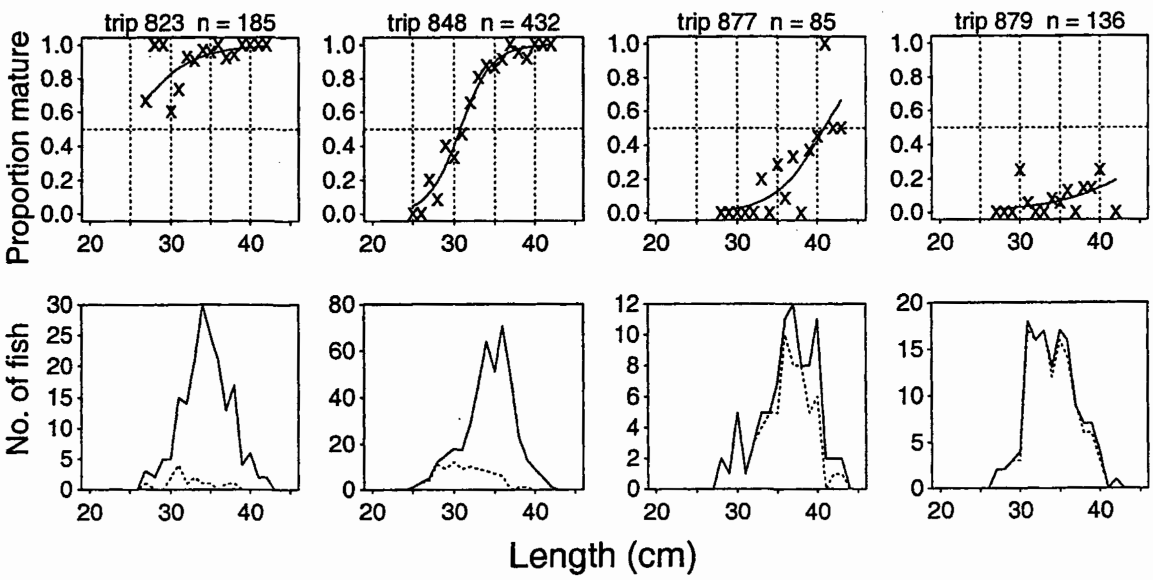

Figure 4: Maturity data for females from four 1994–95 observer trips in hotspot Smiths in the non-spawning season (these constitute part of the data series "Smiths obs nonsp F" of Table 9): upper panels show proportion mature by length, and fitted ogives; lower panels show length frequencies for all females (solid lines) and immature females (broken lines).

4.1.5 Calculation of standard errors for year statistics

4.1.5.1 Standard errors for non-ogive data

It is not immediately obvious how to calculate *s.e.*s for the year statistics for mean length, proportion male, and proportion mature. The sample structure is complex, with three levels — fish, tows, and trips — and, as we have seen in the previous section, there is heterogeneity at all levels. Ideally, we would estimate three components of variance — within-tow, within-trip, and between-trip — and combine these appropriately. However, this estimation is not straightforward, particularly when there are few trips per year or tows per trip. Instead, we calculated three alternative *s.e.*s (which we label $se1$, $se2$, and $se3$) and then used a bootstrapping approach to decide which of these approaches is best.

We need some notation to describe these alternative *s.e.*s. Let the tow statistics be denoted l_{siy} (mean length), p_{iy} (proportion male), and q_{iy} (proportion mature, for females) and let their *s.e.*s be ${}_le_{siy}$, ${}_pe_{iy}$, and ${}_qe_{iy}$ respectively, where the subscripts s, i, and y denote sex, tow, and year, respectively, and s takes values m (male) and f (female). Let C_{siy} be the catch weight (for the given sex, tow, and year) and write the total catch weight for a tow as C_{iy} (= $C_{miy} + C_{fiy}$).

Recall that each year statistic, X_y, was calculated as a weighted mean, so that

$$X_y = \left[\sum_i w_{iy} x_{iy}\right] / \left[\sum_i w_{iy}\right]$$

For the mean lengths, $x_{iy} = l_{siy}$ and $w_{iy} = C_{siy}$; for the proportions male, $x_{iy} = p_{iy}$ and $w_{iy} = C_{iy}$; and for the proportions mature, $x_{iy} = q_{iy}$ and $w_{iy} = C_{fiy}$.

Our first approach to calculating *s.e.*s for year statistics assumes that the within-trip and between-trip variability is negligible. We use the formula

$$se1 = \sqrt{\sum_i \left(w_{iy}\sigma_{iy}\right)^2} / \left[\sum_i w_{iy}\right] \qquad (1)$$

where the weights, w_{iy}, are as above, and σ_{iy} is the appropriate tow *s.e.* (${}_le_{siy}$, ${}_pe_{iy}$, or ${}_qe_{iy}$).

For the second approach we again use equation (1), with the same weights, but define

$$\sigma_{iy}^2 = \sigma_2^2 = \sqrt{\sum_{iy} \left(x_{iy} - x_{\bullet y}\right)^2} / \sum_y \left(m_y - 1\right) \qquad (2)$$

where the x_{iy} are as above, $x_{\bullet y}$ is the mean of the x_{iy} over i, and m_y is the number of tows in year y. This approach treats all tows within a year as equal, ignoring both their grouping by trip, and the fact that the number of measured fish varies between tows.

In the third approach we ignore all the sample structure below the level of a trip (i.e., the fact that some trips have more tows than others, and that more fish were measured in some tows than in others). The formulae are the same as (1) and (2) except that the subscript i, for tows, is replaced by t, for trips:

$$se3 = \sqrt{\sum_t \left(w_{ty}\sigma_{ty}\right)^2} / \left[\sum_t w_{ty}\right] \qquad (3)$$

$$\sigma_{ty}^2 = \sigma_3^2 = \sqrt{\sum_{ty} \left(x_{ty} - x_{\bullet y}\right)^2} / \sum_y \left(M_y - 1\right) \qquad (4)$$

where x_{ty} and w_{ty} take the obvious values ($w_{ty} = \Sigma_i w_{ity}$ and $x_{ty} = [\Sigma_i w_{ity} x_{ity}]/w_{ty}$, where these summations are over all tows within trip t) and M_y is the number of trips in year y. In (4) the summations are made over all years for which $M_y > 1$. Thus we cannot use this procedure for series "Crack res sp", where there was only one trip per year. For this series we estimated *se*3 as 1.8 times *se*2 (1.8 is the median ratio over all other series).

These three ways of calculating *s.e.*s produce very different results. For mean length, the median *s.e.*s for the three approaches were 0.07, 0.18, and 0.42 cm respectively. The corresponding values for proportion male were 0.012, 0.039 and 0.073, and for proportion mature they were 0.0060, 0.0084, and 0.012. Thus it is of great importance to know which approach is best.

Bootstrap estimates of *s.e.*s were calculated as follows for all years for which an ANOVA was carried out (denoting the number of tows in the tth trip for the year as m'_t):

1. Select M_y numbers, $t_1, t_2, ..., t_{M_y}$, at random, with replacement, from the set $1,2,...,M_y$;
2. From trip t_1 select m'_1 tows at random;
3. Repeat step 2 for $t_2, ..., t_{M_y}$;
4. Using the data from the tows selected in steps 2 and 3, calculate the appropriate year statistic (mean length, proportion male, or proportion mature) using the same procedure as for the real data;
5. Repeat steps 1–4, 999 times; and
6. Calculate standard deviations for the resulting set of 1000 statistics.

We would expect these bootstrap *s.e.*s to be, on average, similar in size to the true *s.e.*s. However, they would tend to be a bit larger for those years where the ANOVA was significant, and smaller where they were not. (This is based on the idea that we are more likely to get a significant test result when the between-trip variability in the sample is greater than that in the whole catch, and more likely to get a non-significant result when it is less.) On that basis, *se*2 seems to produce the best estimates for mean length and proportion mature, and *se*3 for proportion male (Figure 5).

It might seem odd that we do not just use the bootstrap *s.e.*s directly, rather than using them to decide which of *se*1, *se*2, and *se*3 is best. The reason is that the bootstrap estimates will be very imprecise when there are few trips per year. Estimates from *se*2 and *se*3 will be more precise because they use data from all years. In doing so they make the assumption that between-tow (for *se*2) or between-trip (for *se*3) variability is the same in all years. Though this may not be strictly true it seems a reasonable approximation and, anyway, we would rarely have sufficient data to prove otherwise.

4.1.5.2 Standard errors for ogive data

Because there are rarely sufficient data to calculate a maturity ogive (and thus estimate the length at maturity, L_{mat}) for each tow, we were not able to calculate something analogous to *se*1, *se*2, and *se*3 for the maturity ogive data. However, we could use a bootstrap procedure analogous to that above. The resulting bootstrap estimates show that L_{mat} is usually very poorly determined (Figure 6). Where this is not true (e.g., 1991 and 1992 in area Smiths) there were usually very few trips sampled. It seems quite likely that the scatter of the bootstrap would be much greater in these years if more trips had been sampled. We conclude that the maturity ogive time series of Table 9 are too imprecise to be useful.

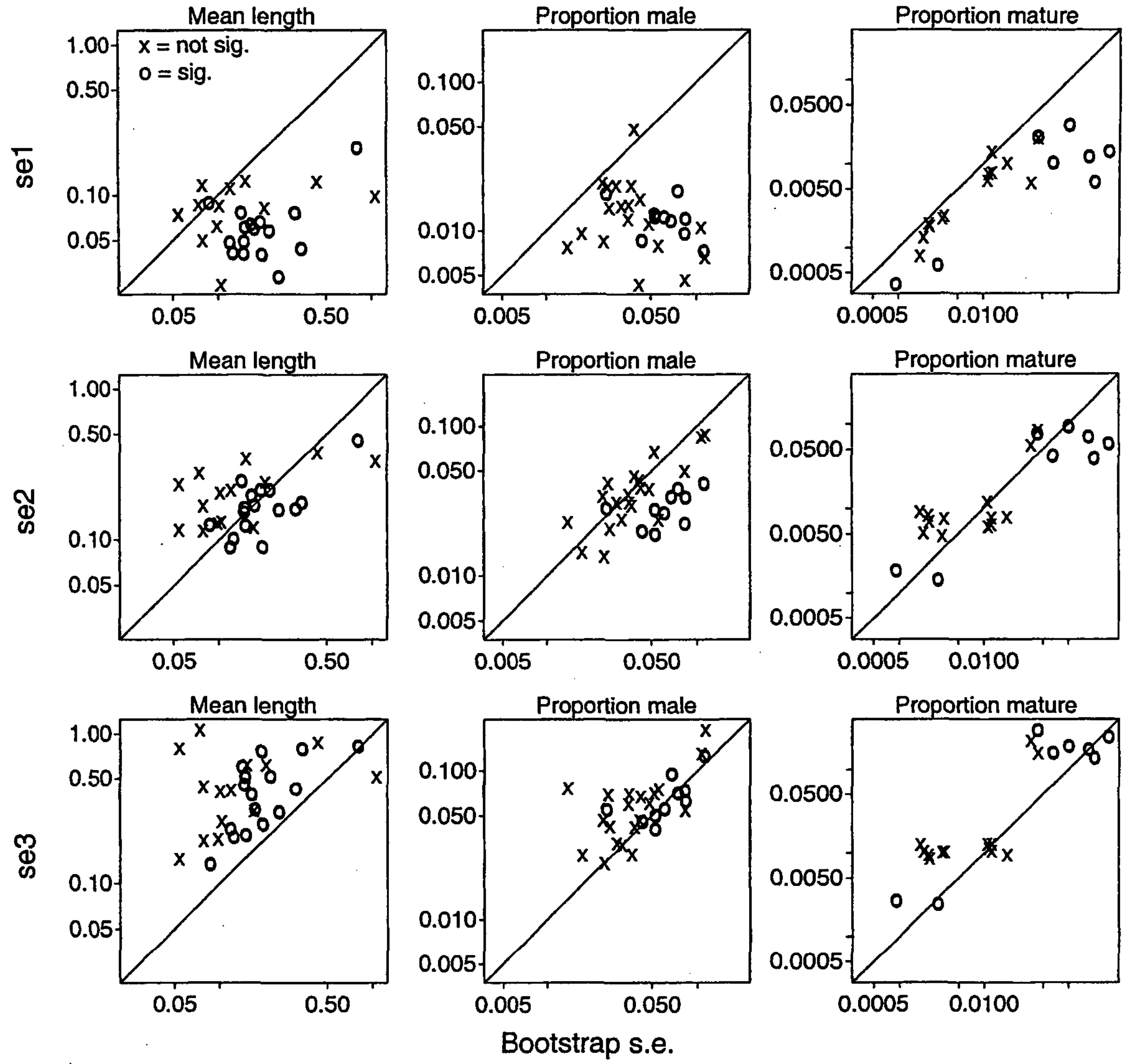

Figure 5: Comparison of bootstrap estimates of *s.e.s* with *se*1, *se*2, and *se*3 for mean length (upper panels) and proportion male (lower panels). Each plotted point represents one year from one of the series in Table 7, with the plotting symbol indicating whether the associated ANOVA was significant ('o') or not ('x').

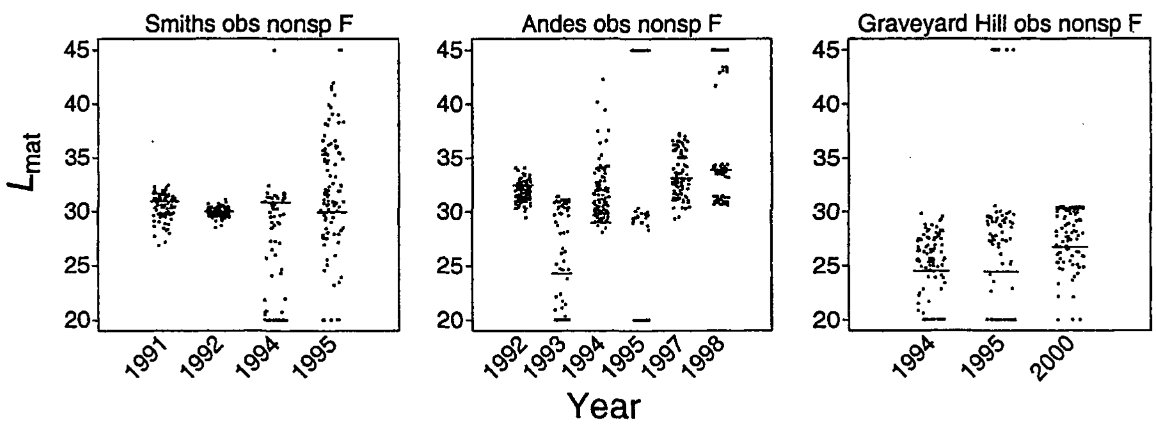

Figure 6: Bootstrap estimates of length at maturity, L_{mat}, for each year of the maturity series of Table 9. Horizontal bars show the estimate of L_{mat} for the original data; each dot represents a bootstrap estimate; estimates outside the range from 20 cm to 45 cm are plotted at the appropriate extreme.

4.1.6 Consistency and relevance

The data series we have examined so far fall into four regions: the Spawning Box on the north Chatham Rise (including our hotspots Plume and Crack); the east Chatham Rise (Smiths and Andes); Ritchie/North; and the Challenger Plateau (Spawning Flat, Pinnacles, and Westpac). The series are plotted in Figures 7 and 8, with confidence intervals from the *s.e.*s derived in Section 4.1.5 (i.e., *se*2 for mean length and proportion mature, and *se*3 for proportion male). We can now assess their consistency and relevance.

The most promising of these plots are the mean lengths in the Spawning Box (top left panel of Figure 8). There are consistent differences between mean lengths in the two subareas. Where comparisons are possible, mean lengths are always higher in Crack than in Plume by an average of 0.7 cm (compare "Plume res sp" with "Crack res sp"; and "Plume obs sp" with "Crack obs sp"). There is also a fairly consistent drop in mean lengths between 1984–1992 and 1995–98: 1.3 cm, 2.3 cm, 0.4 cm and 1.6 cm, respectively for the series labelled 1, 2, 3, and 5 in Figure 8.

The other mean length series are less satisfactory. In the east Chatham Rise there are two outliers ("Smiths obs nonsp" in 1994 and "Andes obs nonsp" in 1993) and no consistent patterns. It is unclear whether these outliers represent real changes in the mean lengths of the catches in these areas, or result from unrepresentative sampling (which would mean that the *s.e.*s are underestimated). It seems unlikely that they represent changes in the fish populations in these areas. The Ritchie/North series is very short and we have no means to judge its reliability. The reliability of the Challenger Plateau series is questionable because of internal inconsistency: the mean length in Westpac was very similar to that in Pinnacles and Spawning flat in 1988, but very different in 1989.

All of the proportion male and proportion mature series fail the relevance criterion (*see* Section 3). None of them shows a marked trend, and there is no strong imbalance in the proportion male series.

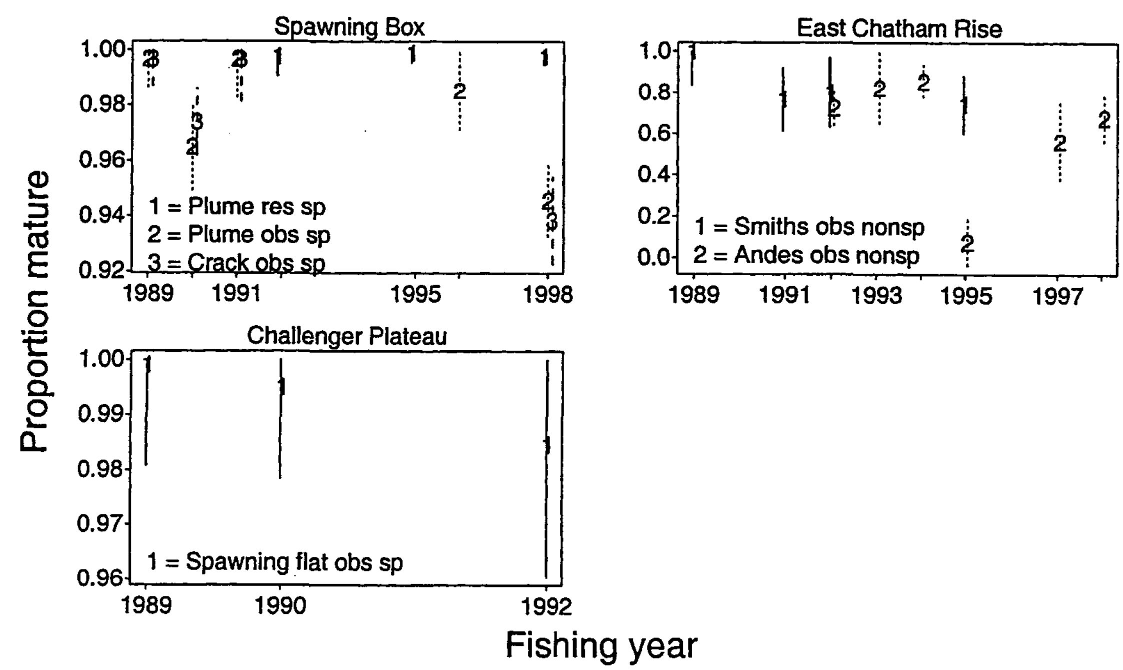

Figure 7: Estimated proportions mature for each year of each of the time series of Table 8, with the series grouped into three areas: Spawning Box, east Chatham Rise, and Challenger Plateau. Vertical bars are approximate 95% confidence intervals (± 2 *s.e.*s).

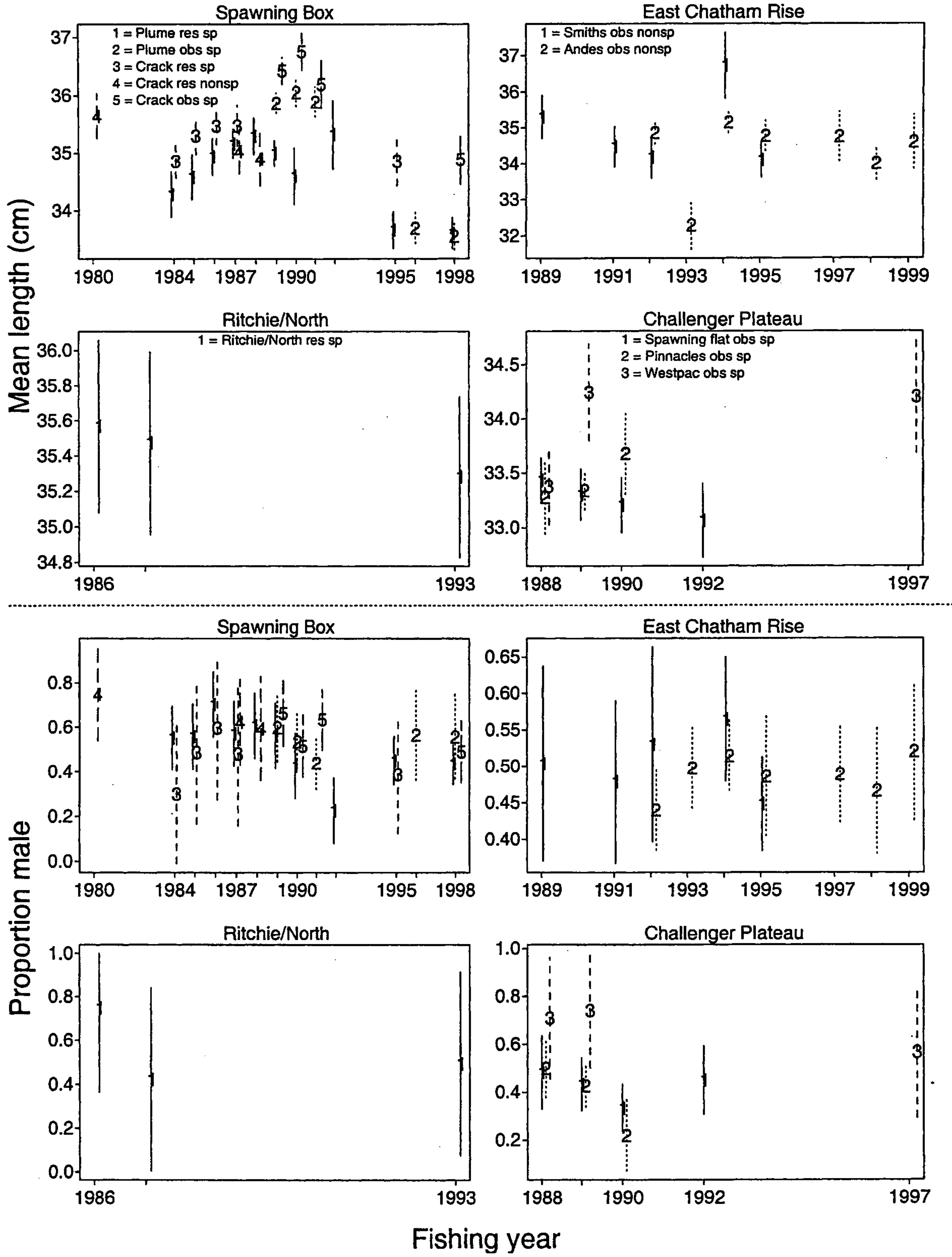

Figure 8: Estimated mean lengths (upper panels) and proportions male (lower panels) for each year of each of the time series of Table 7, with the series grouped into four areas: Spawning Box, east Chatham Rise, Ritchie/North, and Challenger Plateau. Vertical bars are approximate 95% confidence intervals (± 2 *s.e.s*).

4.2 Market data

4.2.1 Data selection

As with the non-random tow data, we attempted to overcome comparability problems in the market samples by selecting time series of data from the same location and season. Location of catches is known only by QMA, and their timing only by landing dates and (sometimes) trip start and end dates. We used landing dates to classify trips into spawning (June and July) and non-spawning seasons. With the constraints

- landings with fewer than two clusters sampled were rejected;
- years with fewer than three landings were rejected; and
- time series with fewer than 3 years data were rejected

we were able to construct four time series, all covering the same three years (Table 10). From now on we will use "trip" in place of "landing" for these data to make the terminology more similar to that of Section 4.1.

Table 10: Descriptions of the four time series of mean length and proportion male that were constructed from the market sample data. The series names identify the QMA and period (sp = spawning, nonsp = non-spawning); "% coverage" is the estimated percentage of the total commercial catch in the given QMA and period that was included in the market samples. Years are fishing years (1989 = 1988–89)

Series	Description			
	Year	1989	1990	1991
2A sp	No. of trips	13	15	21
	% coverage	14	15	29
2A nonsp	No. of trips	7	14	13
	% coverage	24	10	8
2B nonsp	No. of trips	3	3	4
	% coverage	10	20	11
3A nonsp	No. of trips	4	7	7
	% coverage	9	11	9

4.2.2 Simple statistics

For each selected trip we calculated mean lengths by sex and proportion male, together with their *s.e.*s, using standard cluster sample formulae (equations (75) and (78) of Chapter VI of Sukhatme & Sukhatme 1970). The finite sample correction in the latter equation was ignored because the total number of clusters in each catch was not always recorded and, when it was, the sampling fraction was usually low (median = 0.06). These equations assume that all fish in each sampled cluster were measured. This is believed to have been true for the vast majority of clusters (the percent sampled was recorded as 100 for 2343 clusters, less than 100 [between 40 and 83] for 4 clusters, and not recorded for 88 clusters).

Mean lengths by sex, and proportions male, were then calculated, for each year of each series in Table 10, as weighted means across all trips in each year, where the weights used were the catch weights for each trip, either for the whole catch (for mean proportion male) or for the estimated weight by sex (for the mean lengths). The estimated weight by sex was taken as the total catch weight multiplied by the proportion (by number) of that sex.

4.2.3 Heterogeneity in the selected data

It is clear from a plot of the trip statistics that the trip *s.e.*s underestimate the between-trip variability (Figure 9). (This was confirmed by a simple simulation experiment analogous to that described in the first paragraph of Section 4.1.4, except that we are investigating variability between trips in the same year, rather than tows in the same trip. The median absolute differences were more than seven times as big as the average *s.e.*). Thus, we cannot simply lump together all data from the same year in the same area because fish from different trips are, on average, more different than those from the same trip.

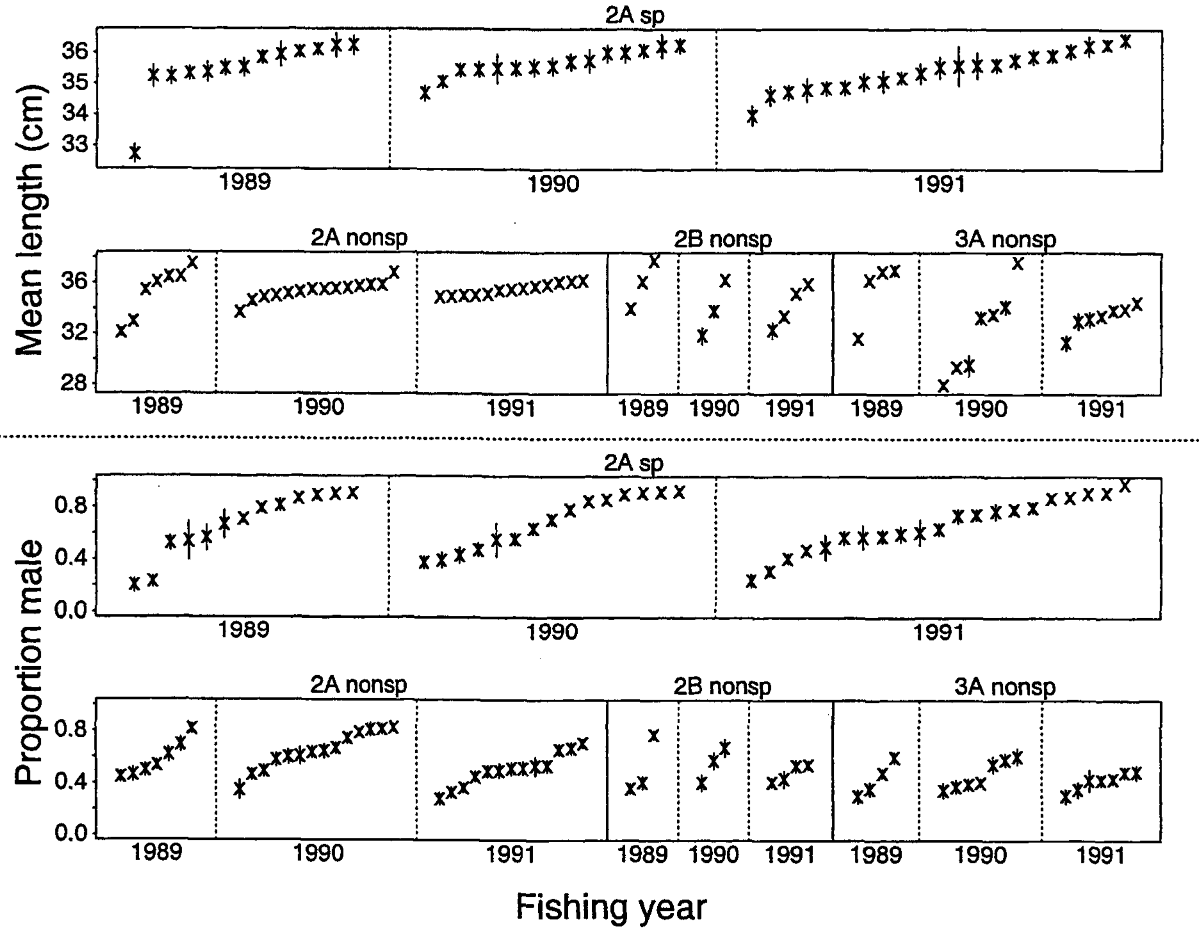

Figure 9: Mean length (upper panels) and proportion male (lower panels) for each trip for each series in Table 10. Each 'x' represents one trip and the vertical line through it is an approximate 95% confidence interval (± 2 *s.e.*s). Vertical broken lines separate trips by fishing year (within each year the trips are ordered by increasing mean length, or proportion male); vertical solid lines separate trips from different series.

4.2.4 Calculation of standard errors for year statistics

In contrast to the random tow data, we have little choice, with the market data, of method for calculating *s.e.*s for the year statistics. The within-year heterogeneity demonstrated in the preceding section means that something analogous to *se*1 (*see* Section 4.1.5) is clearly inappropriate, and the *se*2 formulae do not make sense for these data because we have no tow by tow information. Thus we are left with *se*3. The use of *se*3 is supported by a comparison with the bootstrap estimates of Field *et al.* (1994). The former produces *c.v.*s of 0.005, 0.005, and 0.004 for mean lengths in years 1989 to 1991 in the "2A sp" series. The latter, for slightly different data sets, estimated *c.v.*s of 0.004 for all years.

4.2.5 Consistency and relevance

The market sample time series of mean length and proportion male (Figure 10) seem reasonably consistent (although the mean lengths from ORH 2B and ORH 3A are so imprecisely determined as to be of little value) but are not of great relevance. The mean lengths could perhaps be used but, with only three years' data, will make little contribution to a stock assessment; the proportion male show neither a strong imbalance nor a marked trend.

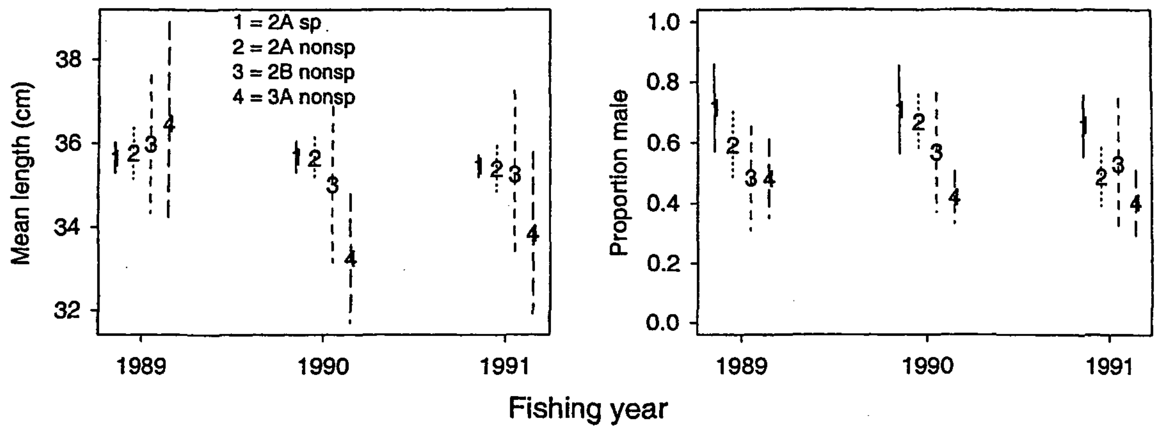

Figure 10: Estimated mean lengths (left) and proportions male (right) for each year of each of the time series of Table 10. Vertical bars are approximate 95% confidence intervals (± 2 *s.e.s*).

4.3 Random survey data

A new program was written to estimate, for each of the random surveys in Table 4, the population length frequencies, and the proportion mature at length, by sex (existing software did not allow calculation of proportion mature). Some strata were dropped from some surveys so that all surveys in the same series covered the same area. Very few immature fish were caught in the Spawning Box and Challenger Plateau surveys (Figure 11) because these surveys were restricted to the main spawning areas during the spawning season. By contrast, the Mid-East Coast surveys, which were carried out over a large area outside the spawning season, caught many immature fish.

Time series calculated from the frequencies in Figure 11 are plotted in Figure 12. Mean lengths were calculated for mature fish only because the assumption (in NIWA assessments) is that recruitment occurs at the onset of maturity (so recruited = mature) and it is the recruited population that is of primary interest in assessments. For the Spawning Box these mean lengths are very similar to those previously used in assessments, where the mean length of recruited fish was approximated as the mean for all fish not shorter than 25 cm (Table 11). Lengths at maturity are not presented for the Challenger Plateau series because there were, in several surveys, too few immature females to allow a maturity ogive to be fitted.

Table 11: Two sets of estimated mean lengths (cm) by year for the Spawning Box: from a recent stock assessment (Francis 1999), and from Figure 12

Source	Mean lengths									Method of
	1984	1985	1986	1987	1988	1989	1990	1992	1994	calculation
Stock assessment	34.72	34.78	34.65	35.27	34.5	34.84	34.35	34.95	33.87	all fish ≥ 25 cm
Figure 12	34.66	34.91	34.86	35.3	34.73	34.97	34.4	35.07	34.09	mature fish

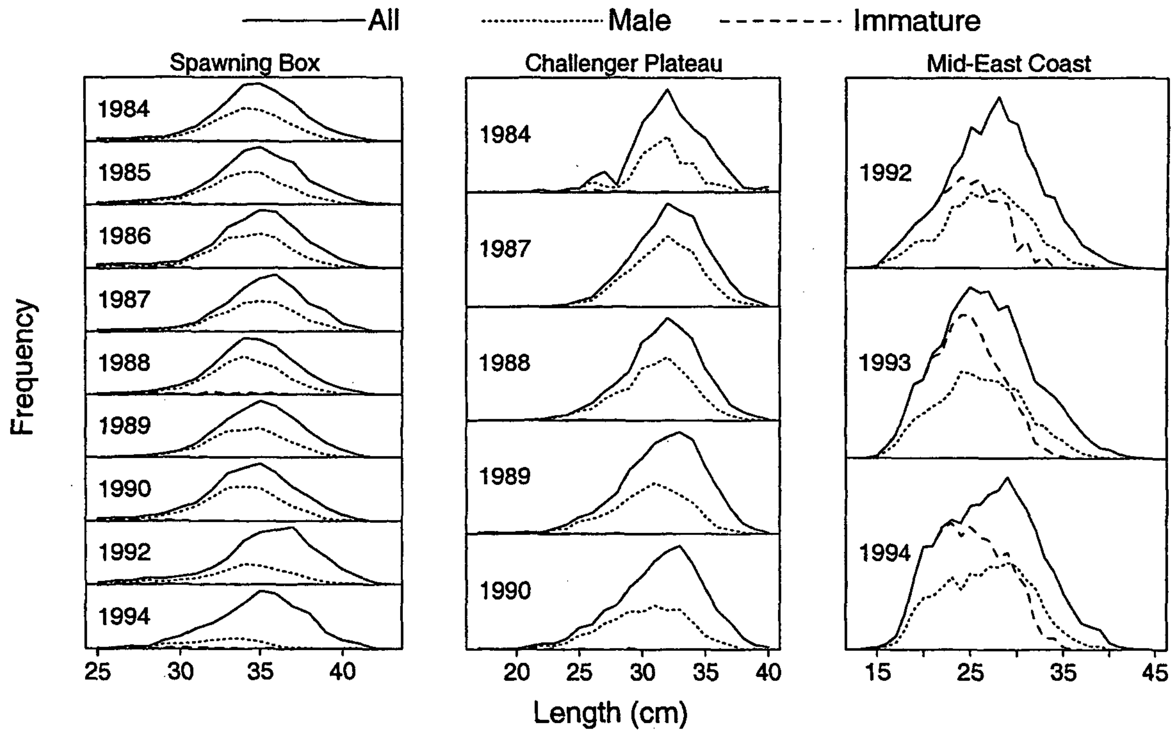

Figure 11: Estimated length frequencies (of all fish, males, and immature fish) for the populations surveyed by each of the stratified random trawl surveys in Table 4.

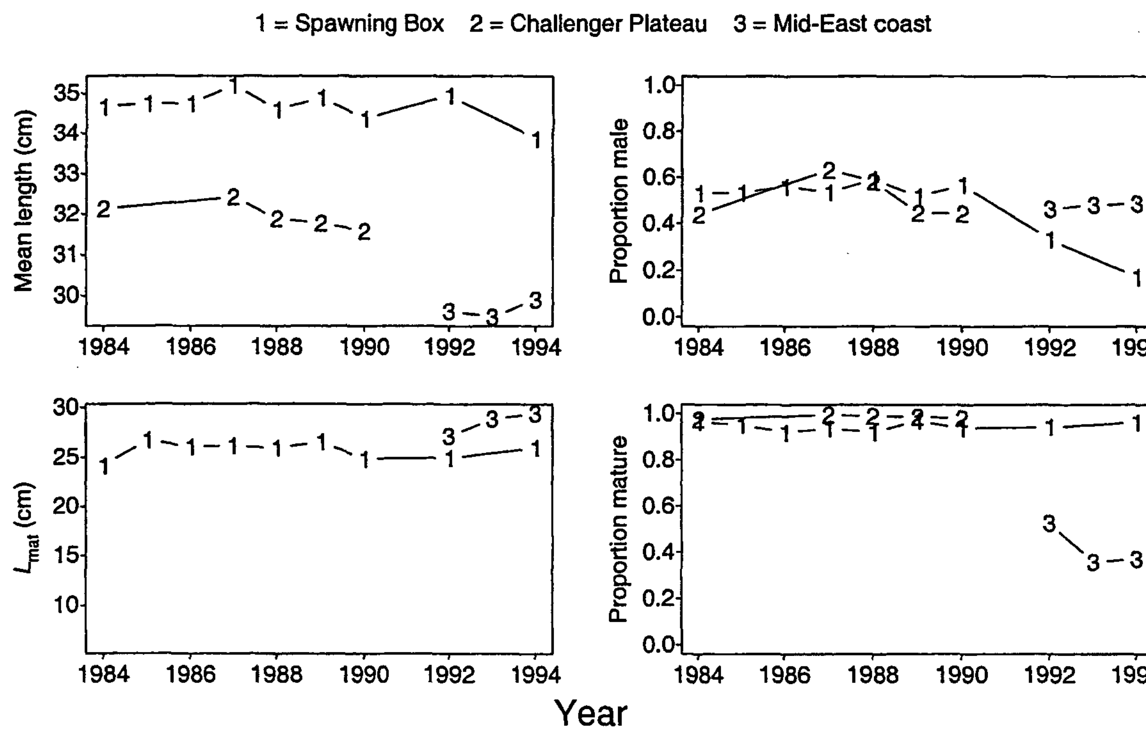

Figure 12: Estimated statistics — mean lengths (mature fish only), proportions male, lengths at maturity (L_{mat}, calculated as the mean of the L_{mat}s for males and females) and proportions mature — for each year of each of the time series of the stratified random trawl surveys in Table 4.

The evaluation of these time series is simpler than those from the non-random tow data (Section 4.1) or the market sample data (Section 4.2). Comparability and consistency are guaranteed by the survey design. Precision is not a major issue because a bootstrap technique already exists for calculating *s.e.*s for mean length (Francis *et al.* 1992) and this could easily be adapted for the proportions male and mature. Either short duration or (for the proportions and L_{mat}) a lack of marked trend rule out most of the series. The trend in proportion male for the Spawning Box appears of interest but the apparent dominance of females in the last two years has been shown to be uncertain (Francis 1996). This leaves just two series that appear useful: mean lengths for the Spawning Box and Challenger Plateau.

5. DISCUSSION

Stock assessments may be thought of as consisting of two parts: a formal part, involving mathematical modelling; and an informal part, in which ancillary information is used to make decisions about data inputs and interpret the modelling results. This report focuses on the formal part and points out (in Section 3) that for biological data to be useful as inputs to stock assessment models they must satisfy stringent requirements concerning duration, comparability, precision, consistency, and relevance. Because of these requirements only a small percentage of the large collection of biological data (length and sex for 1.5 million fish; gonad stages for 0.5 million fish) from commercial and research vessels appears to be suitable for formal use: mean lengths from non-random tows in the Spawning Box, and from stratified random trawl surveys in both the Spawning Box and the Challenger Plateau.

Much of the data was excluded for reasons of comparability and duration. We need to have samples from the same area and season for as many years as possible (four years is probably a sensible minimum). Because there are many factors which affect the decision as to which commercial trips (or landings) are sampled in any year it is hard to maintain the long-term commitment that is necessary to meet these requirements. The same problem occurs with research voyages which are not part of a planned series of surveys.

The sex ratio and gonad stage data were usually excluded on the grounds of irrelevance: they rarely tell us anything about the quantities that are of primary interest in stock assessments: biomass and recruitment.

Another major reason for excluding data is that commercial vessels do not fish at random. One consequence of this is that, although we may have very good data from the trips that were sampled, these data are not representative of the whole catch from an area because different vessels fish differently.

It is often difficult to make definitive statements about the usefulness of these biological data. The procedures we applied (e.g., the hotspot boundaries, constraints on data selection, and the suggestion that a 3-year time series is too short to be useful) are, to some extent, arbitrary. Someone else examining the same data might come to slightly different conclusions.

With commercial (and any other non-random) data it is important that comparability is established (as much as is possible) before a time series is used in a stock assessment. We have seen that the characteristics of catches can change significantly over very short distances (e.g., the difference of 0.7 cm in mean length between Plume and Crack in the Spawning Box, a distance of only about 30 km). They can also change over short periods of time (Figure 13). Thus, slight between-year changes in the location of timing of fishing can cause an apparent change in, say, mean length which is not representative of the fish population. Of course, we do have the choice, when we use a time series in a population model, of whether it is to represent the catch, or the population. However, our primary interest is always the population.

Although they are of limited use in stock assessment modelling, the biological data considered here are widely used elsewhere. Francis (1996) provided a good example of their use in the informal part of stock assessments. Observer sex ratio data from all areas were examined to help interpret a puzzling trend in the proportion of males in surveys in the Spawning Box. They are also regularly used in planning acoustic and trawl surveys (to identify the timing and location of spawning, for example) and providing background information on research proposals. Appendix 2 includes references to many other uses of these data.

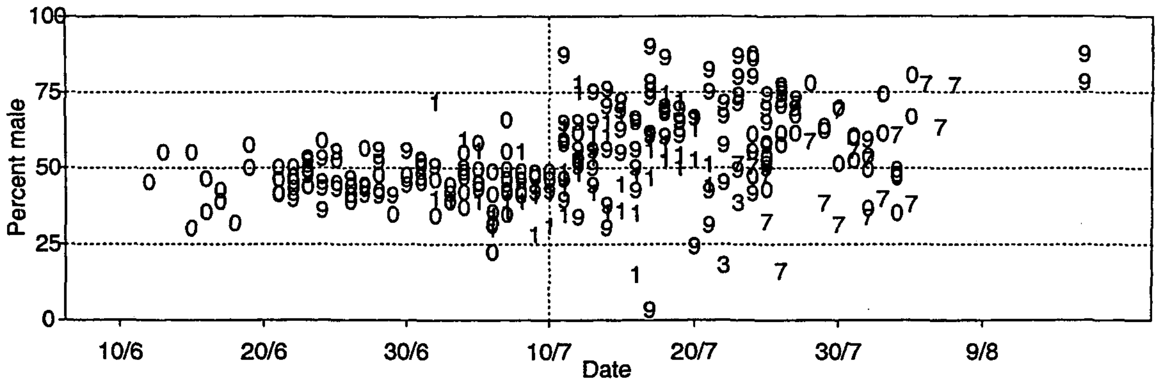

Figure 13: Percent male plotted against date for scientific observer data from commercial tows in the Spawning Box. Each point represents one tow, with the plotting symbol being the last digit of the year (samples sizes: *1987, 19*; *1989, 106*; *1990, 108*; *1991, 57*; *1993, 2*). More information on sex ratios within the Spawning Box was given by Francis (1996), from which this plot is taken.

6. ACKNOWLEDGMENTS

We are grateful to Paul Grimes for help in interpreting the Mid-East Coast trawl survey data and to Malcolm Clark for helpful comments. This work was funded by the Ministry of Fisheries under project code ORH1999/02.

7. REFERENCES

Annala. J.H., Sullivan, K.J., & O'Brien, C.J. 2000: Report from the fishery assessment plenary, May 2000: stock assessments and yield estimates. 495 p. (Unpublished report held in NIWA library, Wellington.)

Field, K.D. 1990: Market sampling of orange roughy from the Challenger Plateau during 1987-88 and east coast of New Zealand during 1988-89. MAF Fisheries Greta Point Internal report 157. 28 p. (Unpublished report held in the NIWA library, Wellington.)

Field, K. 1991: Market sampling of orange roughy from the east coast of New Zealand during 1989-90. MAF Fisheries Greta Point Internal Report 165. 25 p. (Unpublished report held in the NIWA library, Wellington.)

Field, K.D., Francis, R.I.C.C., Zeldis, J.R., & Annala, J.H. 1994. Assessment of the Cape Runaway to Banks Peninsula (ORH 2A, 2B, and 3A) orange roughy fishery for the 1994-1995 fishing year. New Zealand Fisheries Assessment Research Document 94/20. 24 p. (Unpublished report held in the NIWA library, Wellington.)

Fisher, D.O. & Banks, D.A. 1991: Stock Monitoring Programme market sampling data held by the Stock Monitoring Group at MAF Fisheries Greta Point : as at 31 August 1991. MAF Fisheries Greta Point Internal Report 169. 28 p. (Unpublished report held in the NIWA library, Wellington.)

Fisher, D.O. & Banks, D.A. 1993: Market sampling data held by the Stock Monitoring Group on the market database as at 30 April 1993. MAF Fisheries Greta Point Internal Report 206. 30 p. (Unpublished report held in the NIWA library, Wellington.)

Francis, R.I.C.C. 1996. Orange roughy sex ratios and catchrate distributions in the Chatham Rise Spawning Box. New Zealand Fisheries Assessment Research Document 96/13. (Unpublished report held in the NIWA library, Wellington.)

Francis, R.I.C.C. 1999: Stock assessment of orange roughy on the northeast Chatham Rise. New Zealand Fisheries Assessment Research Document 99/41. 16 p. (Unpublished report held in the NIWA library, Wellington.)

Francis, R.I.C.C. and Clark, M.R. 1998. Inferring spawning migrations of orange roughy (*Hoplostethus atlanticus*) from spawning ogives. *Marine and Freshwater Research 49(2)*: 103–108.

Francis, R.I.C.C., Robertson, D.A., Clark, M.R., & Coburn, R.P. 1992. Assessment of the ORH 3B orange roughy fishery for the 1992/93 fishing year. New Zealand Fisheries Assessment Research Document 92/4. 45 p. (Unpublished report held in the NIWA library, Wellington.)

Hilborn, R., Starr, P.J., & Ernst, B. 2000a: Stock assessment of the northeast Chatham Rise orange roughy. WG-Deepwater-00/35. (Unpublished report held by Ministry of Fisheries, Wellington.)

Hilborn, R., Starr, P.J., & Ernst, B. 2000b: Stock assessment of the East Cape stock (2A north) orange roughy. WG-Deepwater-00/44. (Unpublished report held by Ministry of Fisheries, Wellington.)

Sukhatme, P.V. & Sukhatme, B.V. 1970: Sampling theory of surveys with applications. Iowa State University Press, Ames, Iowa. 452 p.

Zeldis, J.R., Francis, R.I.C.C., Field, K.D., Clark, M.R., & Grimes, P.J. 1997. Description and analyses of the 1995 orange roughy egg surveys at East Cape and Ritchie Bank (TAN9507), and reanalyses of the 1993 Ritchie Bank egg survey. New Zealand Fisheries Assessment Research Document 97/28. 34 p. (Unpublished report held in the NIWA library, Wellington.)

Appendix 1

Table A1: Longitude and latitude ranges for the 15 hotspots plotted in Figure 1

Name	Longitudes E		Latitudes S	
Plume	182.5	183	-42.92	-42.75
Crack	183	183.25	-42.92	-42.75
Smiths etc	185.3	186.3	-43.3	-42.5
X spot	186.4	186.8	-43	-42
Andes	185.4	185.6	-44.3	-44
NT Sunday	185.6	185.8	-43.9	-43.8
Hole	175.5	176.5	-43	-42
Graveyard Hill	180	180.1	-43	-42.7
Tolaga	178	179.5	-39	-38.5
Ritchie/North	178.13	178.67	-39.52	-39
Rock Garden/Tims	177.5	178.5	-40.25	-39.75
Madden/Omakere	177.5	178.13	-39.75	-39.5
Spawning flat	168	168.33	-40	-39.75
Pinnacles	167.92	168	-40.13	-40
Westpac	167	167.4	-39.9	-39.67

Appendix 2: Reports and publications using orange roughy research and observer data

Bull, B., Doonan, I., Tracey, D., & Coombs, R. 2000: An acoustic estimate of orange roughy abundance on the Northwest Hills, Chatham Rise, June-July 1999 *New Zealand Fisheries Assessment Report* 2000/20. 25 p.

Bull, B., Doonan, I.J., & Coombs, R.F. 1999: Acoustic estimation of biomass of spawning orange roughy on the Northwest Chatham Rise hills – June-July 1997. Final Research Report for Ministry of Fisheries Research Project MOF804K. 16 p. (Unpublished report held by Ministry of Fisheries, Wellington.)

Clark, M. R. & Tracey, D. M. 1994: Changes in the population of orange roughy (*Hoplostethus atlanticus*) with commercial exploitation on the Challenger Plateau, New Zealand. *Fishery Bulletin (US) 92*: 236–253.

Clark, M.R. & Tilzey, R. 1996: A summary of stock assessment information for orange roughy fisheries on the Lord Howe Rise: 1996. Bureau of Resource Sciences, Canberra. 23 p.

Clark, M.R., Anderson, O.F., & Langley, A. 2000: The ORH 1 fishery in 1997–98 and 1998–99: description of catch-effort, biological, and hydrographic data, and industry comments on further potential. Report prepared for The ORH 1 Exploratory Fishing Company. 30 p. (Unpublished report held in the NIWA library, Wellington.)

Clark, M.R., Anderson, O.F., Francis, R.I.C.C., & Tracey, D.M. 2000: The effects of commercial exploitation on orange roughy (*Hoplostethus atlanticus*) from the continental slope of the Chatham Rise, New Zealand, from 1979 to 1997. *Fisheries Research 45(3)*: 217–238.

Doonan, I. J. 1994: Life history parameters of orange roughy: estimates for 1994. New Zealand Fisheries Assessment Research Document 94/19. 13 p. (Unpublished report held in the NIWA library, Wellington.)

Doonan, I. J., Francis, R. I. C. C., Horn, P., & Tracey, D. M. 1998: Update of biological parameters for orange roughy. WG-Deepwater-98/16. 9 p. (Unpublished report held by Ministry of Fisheries, Wellington.)

Doonan, I., Coombs, R., Barr, R., McClatchie, S., Grimes, P., Hart, A., Tracey, D., & McMillan, P. 1999: Estimation of the absolute abundance of orange roughy on the Chatham Rise. Final Research Report for Ministry of Fisheries Project ORH9701. 23 p. (Unpublished report held by Ministry of Fisheries, Wellington.)

Francis, R.I.C.C. 1996: Orange roughy sex ratios and catchrate distributions in the Chatham Rise Spawning Box. New Zealand Fisheries Assessment Research Document 96/13. 28 p. (Unpublished report held in the NIWA library, Wellington.)

Francis, R.I.C.C. 1999: Stock assessment of orange roughy on the northeast Chatham Rise. New Zealand Fisheries Assessment Research Document 99/41. 16 p. (Unpublished report held in the NIWA library, Wellington.)

Francis, R. I. C. C. & Horn, P. L. 1997: The transition zone in otoliths of orange roughy (*Hoplostethus atlanticus*) and its relationship to age at maturity. *Marine Biology 129*: 681–687.

Francis, R.I.C.C., Clark, M.R., Coburn, R.P., Field, K.D., & Grimes, P.J.: 1995 Assessment of the ORH 3B orange roughy fishery for the 1994–95 fishing year. New Zealand Fisheries Assessment Research Document 95/4. (Unpublished report held in the NIWA library, Wellington.)

Francis, R.I.C.C. & Smith, D.C. 1995: Mean length, age, and otolith weight as potential indicators of biomass depletion for Chatham Rise orange roughy. New Zealand Fisheries Assessment Research Document 95/3. 8 p. (Unpublished report held in the NIWA library, Wellington.)

Francis, R.I.C.C., Clark, M.R., & Grimes, P.J. 1997: Calculation of the recruited biomass of orange roughy on the northwest Chatham Rise using the 1996 Graveyard Hill egg survey (TAN9608). New Zealand Fisheries Assessment Research Document 97/29. 18 p. (Unpublished report held in the NIWA library, Wellington.)

Horn, P.L., Tracey, D.M., & Clark, M.R. 1998: Between-area differences in age and length at first maturity of the orange roughy *Hoplostethus atlanticus*. *Marine Biology* 132:187-194.

Tracey, D.M. & Grimes, P.J. (1997): Trend analysis of the East Coast (ORH 2A, 2B, and 3A) orange roughy research data series. Report for Ministry of Fisheries Project DOR704. 12 p. (Unpublished report held by Ministry of Fisheries, Wellington.)