Quantifying annual variation in catchability for commercial and research fishing

R.I.C. Chris Francis
Rosemary J. Hurst
James A. Renwick
National Institute of Water and Atmospheric Research
P.O. Box 14901
Wellington, New Zealand
E-mail address (for R.I.C.C. Francis): c.francis@niwa.cri.nz

Abstract—Biomass indices, from commercial catch per unit of effort (CPUE) or random trawl surveys, are commonly used in fisheries stock assessments. Uncertainty in such indices, often expressed as a coefficient of variation (CV), has two components: observation error, and annual variation in catchability. Only the former can be estimated directly. As a result, the CVs used for these indices either ignore the annual-variation component or assume a value for it (often implicitly). Two types of data for New Zealand stocks were examined: 48 sets of residuals and catchability estimates from stock assessments using either CPUE or trawl survey indices; and biomass estimates from 17 time series of trawl surveys with between 4 and 25 species per time series. These data show clear evidence of significant annual variation in catchability. With the trawl survey data, catchability was detectably extreme for many species in about one year in six. The assessment data suggest that this annual variability typically has a CV of about 0.2. For commercial CPUE the variability is slightly less, and a typical total CV (including both components) of 0.15 to 0.2. This is much less than the values of 0.3 to 0.35 that have commonly been assumed in New Zealand. Some estimates of catchability are shown to be implausible.

Manuscript accepted 17 December 2002.
Manuscript received 31 December 2002 at NMFS Scientific Publications Office.
Fish. Bull. 101: 293–304 (2003).

Catchability is a key parameter that is estimated in many fish stock assessments (Arreguín-Sánchez, 1996). It is the constant of proportionality between biomass indices (either from commercial catch per unit of effort (CPUE) or random trawl surveys) and absolute biomass. Despite its importance it is usually thought of as a "nuisance" parameter: one that is not of intrinsic interest but which needs to be estimated so that other quantities, which are of interest (e.g. biomass), can be estimated. For this reason estimates of catchability are not often reported.

Uncertainty in biomass indices, often expressed as a coefficient of variation (CV), has two components: observation error and annual variation in catchability. Only the former can be estimated directly. As a result, the CVs used for these indices either ignore the annual-variation component or assume a value for it (often implicitly). Our objectives in this study were to estimate the extent to which catchability varies from year to year for New Zealand stocks. We used all available data, including residuals and catchability estimates from stock assessments and biomass estimates from time series of trawl surveys. We show that standard New Zealand practice typically overestimates catchability variation for trawl survey indices and underestimates it for CPUE, and suggest that some catchability estimates are clearly implausible. More details concerning the data and analyses below are given by Francis et al. (2001).

We assume throughout that catchability does not vary systematically with abundance. There is much controversy surrounding this assumption, particularly for CPUE. Since Paloheimo and Dickie (1964) gave theoretical reasons to expect that catchability would increase as biomass declined, many authors have presented confirmatory data (e.g. Schaaf, 1975: Pope, 1980; Winters and Wheeler, 1985; Quinn and Collie, 1990). Nevertheless, many stock assessments are based on the assumption that catchability is independent of abundance. It is data from such assessments that we examine here. We also assume, of necessity, that the role of CVs in stock assessments is to describe the precision of biomass indices, rather than their quality. This issue is discussed further in the final paragraph of this paper. To begin with, we describe more precisely what we mean by catchability.

Definitions

"Catchability" is used in several slightly different ways in the fisheries literature. The use we are concerned with is as a parameter (conventionally denoted by q) in a stock assessment model, defined by the equations

$$I_i = qB_i\varepsilon_i \text{ or } \log(I_i) = \log(qB_i)\,\varepsilon_i' \quad (1)$$

where I_i = the index in year i;
B_i = the corresponding true biomass; and
the error terms, ε_i and ε_i' are random variables with expectation 1 and CV (coefficient of variation) c_i.

The interpretation of q in (Eq. 1) depends on whether the I_i are from CPUE or trawl surveys (other types of biomass index—e.g. from acoustic surveys—are possible but not considered here). In the former case, q may be interpreted as the proportion of the population biomass that is caught by one unit of effort. Often the CPUE is standardized (using methods akin to those of Punt et al., 2000), so that the unit of effort is a standard one (e.g. if nationality and area are factors in the CPUE standardization, then the standard unit of effort will be that for a vessel from the reference nation in the reference area). However, the unit of effort is changed when, as is common in New Zealand, CPUE indices are standardized to have value 1 in a reference year. If the I_i are from a trawl survey series, the interpretation of q is slightly different. Here, it is the product of the survey area and the proportion of the biomass that is caught per unit of area swept (because trawl survey indices are usually scaled up by the survey area, whereas no such scaling is done for CPUE indices).

Trawl survey catchability may also be interpreted as the product of three components: vulnerability, v, vertical availability, u_v, and areal availability, u_a (Francis[1]). These components are defined in the framework of a conceptual model in which the trawl gear is thought of as sweeping a volume of water in the shape of a cuboid of width equal to the distance between the trawl doors, height equal to the headline height, and length equal to the distance trawled. Vulnerability is the average proportion of fish in the swept volume that are caught. Vertical availability is the proportion of fish in the survey area that could be encountered by the trawl gear (i.e. that are close enough to the bottom to be below the trawl headline but not so close as to pass under the footrope). Areal availability is the proportion of fish in the population being surveyed that are in the survey area at the time of the survey (this is important in stock assessment when the full range of the stock being assessed is not covered by a survey).

These three components are usually of more theoretical than practical use. That is, they help us to think about the relationship between a trawl survey biomass index and the actual biomass. In New Zealand the common practice is to calculate survey biomass indices as if all three constants had value 1. This means that the catchability associated with these indices is the product of the three components, i.e. $q = vu_vu_a$. This interpretation restricts the range of plausible values for a trawl survey q. Because all three catchability components are defined as proportions their product should be less than (or equal to) 1. (It is technically possible for v to exceed 1 [if, for instance, fish that are initially above the headline, and thus unavailable to the net, are herded downwards] but it is most unlikely that vu_v would be greater than 1; u_a cannot exceed 1.) Thus, if the default values of the catchability components have been used, we would expect q to be less than 1. Also, very small values of q are implausible for any species that is assessed by using trawl survey biomasses. Although there are species that are not well caught by trawls (e.g. because they are fast-swimming, high above the bottom, or because they burrow in the substrate) and thus have very low values of v or u_v (or both), such species are not, for that reason, assessed with trawl survey indices. Similarly, a very low areal availability (implying that most of a fish stock is outside the survey area) would rule out the use of trawl surveys in assessing a stock.

There is also a limit to how much we would expect values of q for the same species to vary between surveys. For a given fishing vessel and trawl net, the components v and u_v are determined by fish behavior (e.g. swimming speed, typical height above the bottom, reaction to an approaching net). This means that if the same vessel and gear are used in surveys in different areas we would expect the product vu_v not to vary very much for the same species (except, perhaps, between spawning and nonspawning periods, when there may be substantial behavioral differences). If different vessels, or gear, are used, we might obtain larger differences in vu_v.

Materials and methods

Data

Two types of New Zealand data were examined: those from stock assessments and those from random trawl surveys.

Assessment data We gathered data from all recent stock assessments that used biomass indices from either trawl surveys or CPUE. One data set was constructed for each separate series of biomass indices (so that an assessment using two different series provided two data sets). Each data set consisted of the following variables:

- the biomass indices input to the assessment;
- the years associated with these indices;
- the CV(s) assumed for these indices;
- a description of the assumed error distribution type;
- the model estimates of (absolute) biomass that correspond to each biomass index; and
- the model estimate of catchability, q, for the indices.

For each stock the latest available assessment (usually carried out in 2000) was used. Data sets with fewer than four annual indices were discarded.

A total of 48 such data sets was constructed (30 with CPUE indices and 18 with trawl survey indices), ranging in length from 4 to 40 indices, with CVs between 0.02 and 0.61 (Fig. 1A) (details of the individual assessments are given in Francis et al., 2001). In most data sets (43 of 48) a single CV was assumed for all indices. Two rock lobster assessments used a time step of six months; all other assessments used a one-year time step. Amongst these data sets there were three different error-distribution assumptions; these determine how standardized residuals are calculated (Table 1).

We refer to the CVs for the assessment data sets as "assumed," rather than "estimated," because we can estimate only one component of these CVs, that due to observation

[1] Francis, R. I. C. C. 1989. A standard approach to biomass estimation from bottom trawl surveys. N.Z. Fish. Assess. Res. Doc. 89/3, 4 p. National Institute of Water and Atmospheric Research, P.O. Box 14901, Wellington, New Zealand.

error. By setting a value for one of these CVs we are implicitly making an assumption about the other component: annual variation in catchability. We did not solicit information on how these CVs were set for individual assessments. However, the most common way is based on a subjective assessment of the "reliability" of the associated biomass indices: the less reliable the indices are judged to be, the higher the assumed CV (by "reliability" we mean the combination of two very different properties of an index, its precision, and its "quality"—the extent to which it is likely to be proportional to biomass). This is why almost half of the CPUE series (13 out of 30) have assumed CVs of 0.35, and many of the trawl survey series (13 of 18) have CVs of 0.25. In most cases, the CVs assumed for trawl survey indices differ from the observation error CVs calculated from the trawl survey data (Fig. 1B).

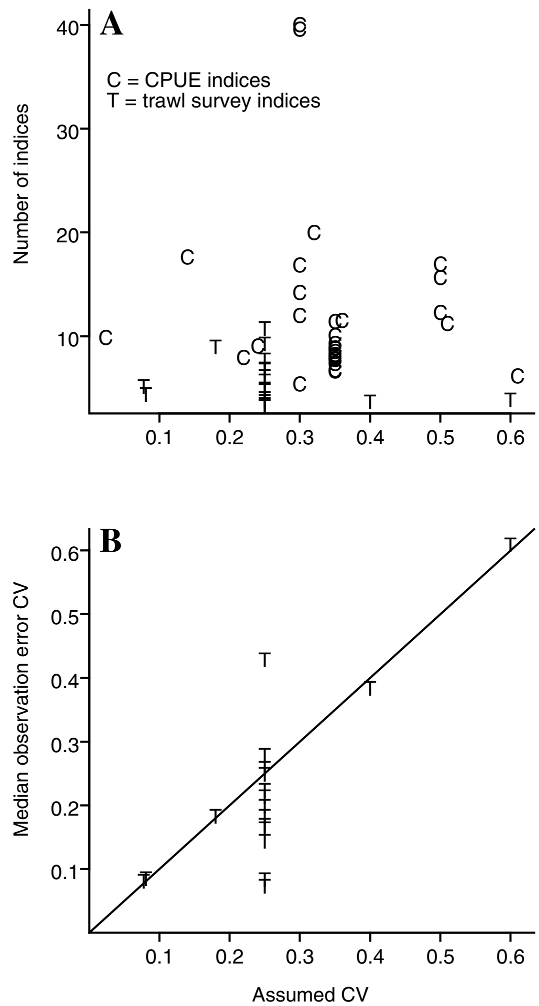

Figure 1

Descriptive statistics for the assessment data sets: assumed CV plotted against (**A**) the number of indices (all data sets) and (**B**) the median observation error CV (data sets with trawl survey indices only). Each point represents one data set; in A the points are jittered slightly to separate coincident points; where different indices in a data set had different assumed CVs the median of these is plotted.

Trawl survey data Data from all New Zealand random trawl surveys were considered. The surveys were grouped into series, each of which contained surveys covering (approximately) the same area at about the same time of year and using the same (or similar) vessel(s) and gear. Some series were split into two, by area, because they were deemed to survey two distinct fish communities. Series with fewer than four surveys were rejected. This left 17 series, with between four and 11 surveys per series.

For each trawl survey series a list of "suitable" species was generated by listing all species caught in the series and then excluding species deemed to be "unsuitable" for any of the following reasons:

- species caught in only a small percentage of tows;
- species caught in small quantities (low mean catch per tow);
- species not well caught by the net because they are too small, too large, too close to the sea floor, or too high in the water column;
- species for which identification was poor, or inconsistent over time;
- species whose range was poorly covered by the surveys (e.g. those occurring mostly on rough ground, or mostly in water shallower or deeper than that covered by the series).

The idea was to include as many species as possible for each series. In considering a particular species in a specific trawl survey series, the following question was useful: "If this were a valuable commercial species would it be appropriate to use this series of trawl surveys to generate biomass indices to put into a stock assessment?" An answer of "yes" (or even "maybe") was a good reason to include this species. The CVs of biomass estimates were not considered in making this decision. For each series the list of suitable species was compiled by people with an intimate knowledge of that series and the associated species. No attempt was made to derive consistent objective criteria (e.g. exclude all species that occurred in fewer than 30% of tows) for all series. The exclusion of a species from one series was no barrier to its inclusion in another. The number of acceptable species in a series varied between 4 and 25.

Biomass indices, and CVs, were calculated for each suitable species in each survey. In all series but two, vulnerability, areal availability, and vertical availability were set to 1. There was a wide range of estimated CVs. Even when the CVs for each species in a series were averaged over all surveys, these averaged values spanned an order of magnitude, from 0.07 to 0.70 (Fig. 2).

Analyses

Our analyses addressed a series of questions, which are given as subheadings in this section.

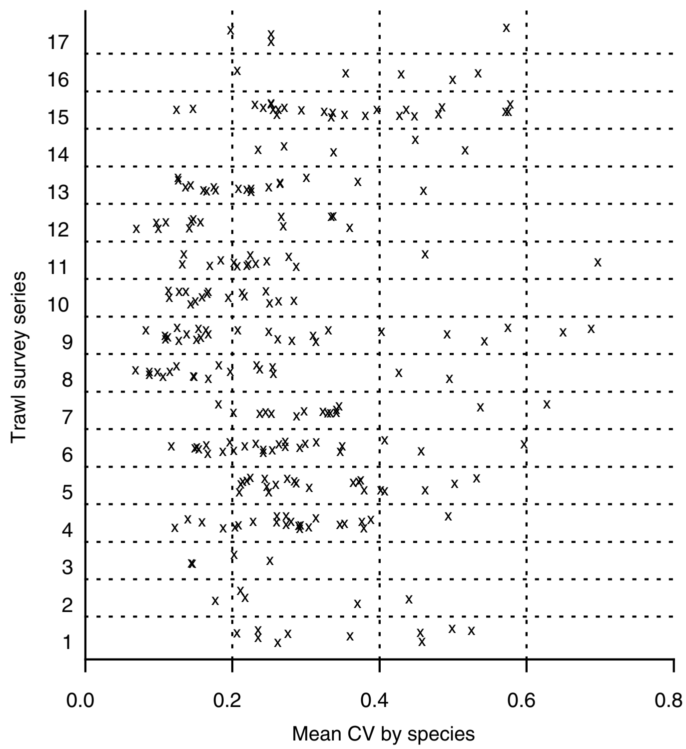

Figure 2

Mean CVs, by species, of biomass indices in the trawl survey data sets. Each plotted point relates to one species in a trawl survey series, and indicates the mean of all estimated CVs for that species in that series. Points are jittered vertically to avoid overlap.

Table 1

Three alternative error-distribution assumptions for biomass indices in stock assessments, and the associated form of the standardized residuals. Notation: I_i is the ith biomass index and B_i is the corresponding model estimate of (absolute) biomass; for assumption lnorm, $\sigma_i^2 = \log(c_i^2 + 1)$.

Label	Description	Standardized residual
norm	I_i is normally distributed with mean qB_i and assumed CV c_i	$\frac{1}{c_i}\left(\frac{I_i}{qB_i} - 1\right)$
lnorm	I_i is lognormally distributed with mean qB_i and assumed CV c_i	$\frac{1}{\sigma_i}\left[\log\left(\frac{I_i}{qB_i}\right) + 0.5\sigma_i^2\right]$
lognorm	$\log(I_i)$ is normally distributed with mean $\log(qB_i)$ and *s.d.* c_i	$\frac{1}{c_i}\log\left(\frac{I_i}{qB_i}\right)$

Are the assessment CVs the right size?

We constructed a residual statistic, κ, that was designed to indicate whether the CVs assumed in the stock assessment (the c_i) were too small or too large in each data set. A positive (or negative) value of κ suggests that the residuals were too large (too small), and thus CVs were too small (too large). The statistic is based on the median absolute standardized residual (MASR), rather than the residual variance, because the latter is not very robust (it is easily inflated by outliers). We defined

$$\kappa = \begin{cases} 2(\mu - \mu_{0.5}) / (\mu_{0.5} - \mu_{0.025}) & \text{if } \mu < \mu_{0.5} \\ 2(\mu - \mu_{0.5}) / (\mu_{0.975} - \mu_{0.5}) & \text{if } \mu > \mu_{0.5} \end{cases},$$

where μ = the MASR from the assessment data; and
μ_r = the rth quantile of the sampling distribution of μ.

To calculate the μ_r we assumed that the standardized residuals follow a Student's t-distribution with n–2 degrees of freedom, where n is the number of indices in the data set. [We assumed n–2 degrees of freedom because, in an assessment with only a single series of relative biomass indices, only two parameters can be estimated, e.g. initial biomass and q (Francis, 1992). When there are many data inputs there may be many more than two parameters estimated.] For each value of the sample size n, the μ_r were estimated by simulating 1000 data sets of size n from a t-distribution with n–2 degrees of freedom, calculating the median absolute value for each simulated data set, and taking the rth quantile of this set of 1000 medians.

We used the κ statistics in two ways. We tested the null hypothesis that the CVs were, on average, of the correct size by using a simple signs test (under this hypothesis we would expect about 50% of the κ's to be of each sign). If significantly more than half are positive (or negative) this shows a tendency to use CVs that are too large (or too small). This test considers all the CVs at once. We also tested each CV separately; a value of κ greater than 2 (less than –2) is statistically significant.

Next, we investigated how much, if at all, we should change the assumed CVs to make their size appropriate. This was done by changing the assumed CVs, recalculating the residual statistic and checking to see whether the new values of κ were evenly distributed about zero. We did this separately for the CPUE and trawl survey indices. For the former we simply set them all to a single default value and searched for the default value that produced an even distribution. For the latter, we assumed that the CV associated with annual variation in catchability was the same for all stocks and "added" this CV to the observation error CVs to obtain assumed CVs for the stock assessments. Note that CVs are "added" as squares, so that when we "add" CVs of 0.2 and 0.3 we get 0.36 [= $(0.2^2+0.3^2)^{0.5}$]. Here we were searching for the value of the catchability CV that produced an even distribution of κ.

Strictly speaking we should rerun each assessment each time we change a CV. However, it is not practical to do this for so many assessments. Thus we have to assume that changing a CV will not change the model estimates too much. Our experience is that this is true for assessments with only one series of biomass indices. It is least likely when there are more than one series and these show markedly different trends.

Can we detect years of extreme trawl survey catchability?

First, as an informal procedure to identify possible years of extreme trawl survey catchability, the trawl survey biomass indices were standardized (by dividing each time series for a particular species by its mean) and plotted by survey. Next, the following more formal procedure was used to identify extreme years. For each species in a trawl survey data set, the survey years were ranked in order of increasing biomass index, and then these ranks were averaged across species to obtain a mean rank for each year. Then the rank deviations, $d_i = |r_i - 0.5(n+1)|$, were calculated, where r_i is the mean rank for year i, n = the number of survey years, and $0.5(n+1)$ is the overall mean of the mean ranks (Table 2).

Table 2

Example, using trawl survey series-3 data, of the stages in the procedure for calculating a mean rank and rank deviation for each survey year in a trawl survey data set.

		Survey year					
	Species	1983	1985	1990	1992	1996	1999
Biomass indices	A	125	482	1565	1141	969	1644
	B	355	47	413	272	320	365
	C	63	48	131	257	118	89
	D	113	111	157	236	191	176
Ranks	A	1	2	5	4	3	6
	B	4	1	6	2	3	5
	C	2	1	5	6	4	3
	D	2	1	3	6	5	4
Mean ranks, r_i		2.25	1.25	4.75	4.5	3.75	4.5
Rank deviations, d_i		1.25	2.25	1.25	1.00	0.25	1.00

The following simulation procedure was used, for each series, to determine which years should be labeled as extreme (i.e. how large the d_i's need to be to be statistically significant).

1. The actual biomass indices were replaced by randomly generated indices (by using a uniform distribution [because our statistic is based on ranks it does not matter what distribution is used to generate the biomass indices]);
2. Mean ranks, and rank deviations, were calculated for each survey year by using these simulated biomass indices;
3. The largest of these rank deviations, $d_{\max,1}$, was stored;
4. Steps 1 to 3 were repeated 999 times, generating $d_{\max,j}$, for j = 2,...,1000;
5. Year i was labeled as extreme if d_i was greater than or equal to at least 95% of the $d_{\max,j}$.

In other words we asked, for each rank deviation d_i, how likely we would be to observe a deviation at least as large as this if there were no between-species correlations. If the probability were less than or equal to 0.05, we would label the year as extreme.

As a diagnostic tool, to examine possible reasons for these extreme years, we calculated between-year changes in biomass indices, expressed as ratios. This was done for all species and for each pair of consecutive surveys that included one extreme year.

Are there consistencies between data sets?

Three types of consistency were sought in the data. First, is the range of estimated trawl survey catchabilities plausible? Second, is there any consistency, between trawl survey series, in the years that are labeled as having high or low catchability? Third, is there consistency between the extreme years in the trawl survey data and the CPUE indices in the assessment data? To address the latter question, each person who provided CPUE data was asked, for each series, which, if any, of the trawl survey data sets were "comparable" in that they related to similar areas, depths, and seasons. For "comparability" it was not necessary that the CPUE species be a target for the trawl survey. We were interested in knowing whether the person thought that the fact that the catchability seemed to be extreme for many species in the trawl survey in some year would be reasonable grounds to believe that this would affect their CPUE index in a similar way (but this person was not told which trawl survey years were considered extreme). For each match that was found between a CPUE index and a trawl survey extreme year we asked whether the two were consistent: that is, whether high (or low) trawl survey catchability corresponded to a positive (or negative) CPUE residual.

Results

Are the assessment CVs the right size?

Results were different for the two types of assessment data (Fig. 3). For those with CPUE indices, there was a tendency for CVs to be too large: κ was negative for 21 of the 30 data sets (this is significantly more than half, P=0.02) and was less than –2 for 9 of them. In contrast, κ was positive for 13 out of the 18 data sets with trawl survey indices (again, significantly more than half, P=0.02) and was greater than 2 for 2 of them. Median CVs for data sets for which the CVs were found to be significantly too large ranged from 0.3 to 0.5; where CVs were significantly too small the median CVs were between 0.02 and 0.24.

If a default CV is to be used for all CPUE series, the best value lies between 0.15 and 0.2; values in this range give approximately equal numbers of positive and negative values of κ (Table 3). The best default value for a trawl survey annual variation CV appears to be about 0.2; this gives approximately equal numbers of positive and negative values of κ (Table 4).

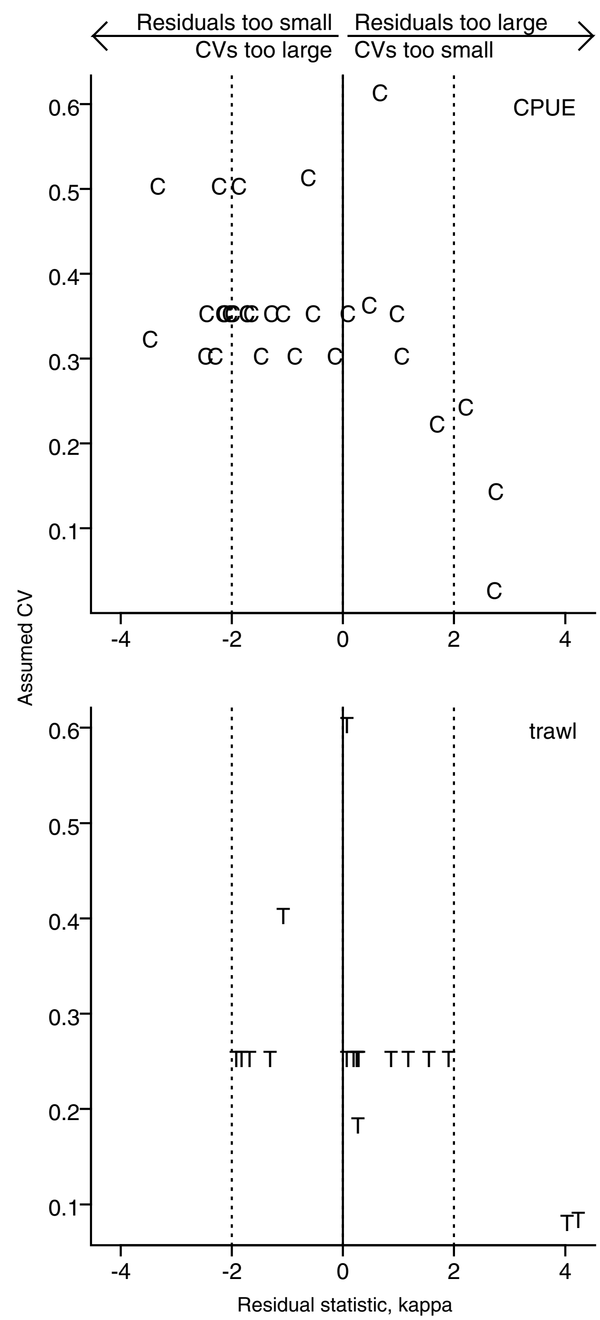

Figure 3

The residual statistic, κ, plotted against assumed CV for each of the assessment data sets: those with CPUE indices in the upper panel, those with trawl survey indices in the lower panel. Each point represents one data set; where different indices in a data set had different assumed CVs, the median of these is plotted.

Can we detect years of extreme trawl survey catchability?

Our informal graphical procedure showed that, for some trawl survey series, the biomass indices for many species fluctuate synchronously, which suggests annual variation in catchability. Two clear examples are shown in Figure 4: for series 5, the biomass indices for many species follow the same up-down-up pattern; for series 6, the opposite pattern (down-up-down) is followed by many species (but not

Table 3

Effect of using different default CVs for the 19 assessment data sets with CPUE indices and assumed CVs of either 0.3 or 0.35. Each line of the table gives the number of these data sets for which κ falls in the given range for the given default CV.

Default CV	Number of data sets			
	$\kappa < -2$	$-2 < \kappa < 0$	$0 < \kappa < 2$	$2 < \kappa$
0.1	0	1	9	9
0.15	0	8	6	5
0.2	0	11	5	3
0.25	3	10	5	1
0.3	6	10	3	0
0.35	7	9	3	0

Table 4

Effect of using different default annual variation CVs for the 18 assessment data sets with trawl survey indices. Each line of the table gives the number of these data sets for which κ falls in the given range when the assumed CV in the assessment is calculated by "adding" the given default CV to the observation error CVs.

Default CV for annual variation	Number of data sets			
	$\kappa < -2$	$-2 < \kappa < 0$	$0 < \kappa < 2$	$2 < \kappa$
0	0	4	10	4
0.1	0	6	9	3
0.15	0	6	11	1
0.2	0	9	8	1
0.25	1	9	8	0
0.3	1	11	6	0

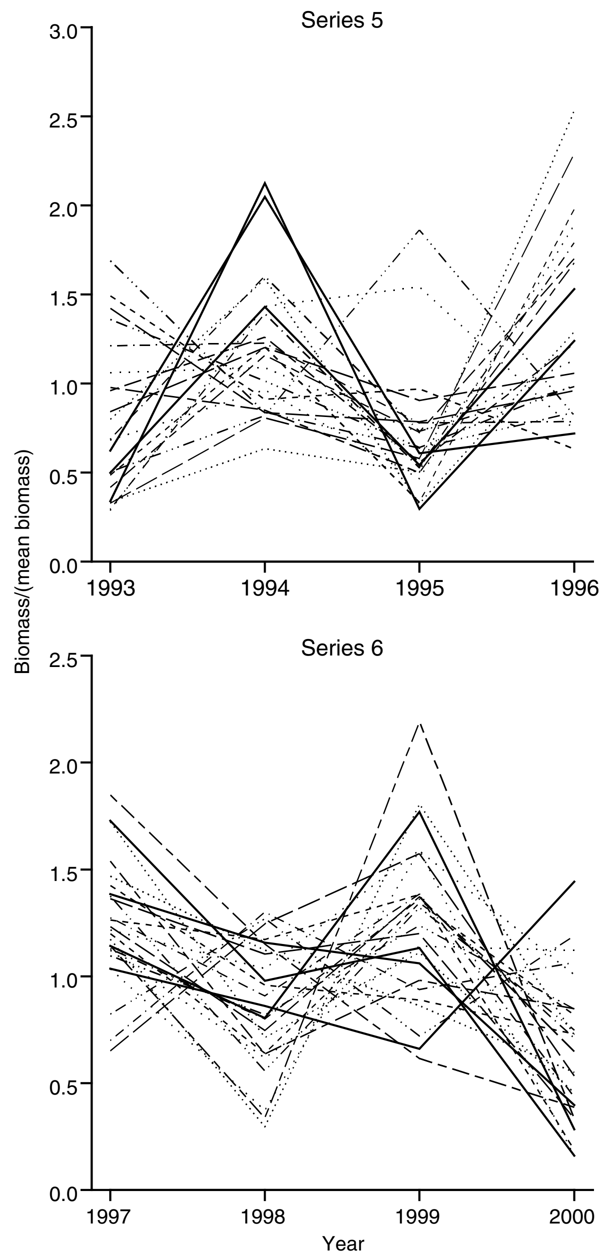

Figure 4

Standardized biomass indices, plotted against year, for trawl survey series 5 and 6. For each species in a series, the biomass indices were standardized by dividing by the mean index for that species.

in the same years). These patterns would be very unlikely to occur by chance alone if there were no between-species correlations. For series 5, 14 of the 22 species had their two highest biomass indices in the same years (1994 and 1996). The probability that an outcome as extreme as this would occur by chance alone (assuming no correlations) is only 6.4×10^{-6}. For series 6, the probability is 9.8×10^{-8} (here 17 of 25 species had their two highest years in 1996 and 1998). These very low probabilities are clear evidence that there are sometimes strong between-species correlations in a survey series. We will argue below that the main cause of these correlations is that catchability was extreme (for many species) in some years.

In our more formal analysis, 16 of a total of 94 survey years were found to be significantly extreme (nine with high catchability, and seven with low), and there were eight survey series for which no years were extreme (Fig. 5). (Note that the vertical distance between the broken lines in Figure 5, which indicates how extreme a mean rank needs to be to be judged significant, decreases with increasing number of species and with decreasing number of survey years.) We also investigated three modifications to the above procedure for identifying extreme years to see whether they might be useful. None was (see Appendix for details).

For some extreme years the biomass ratio statistics (Table 5) are so large that it is unlikely that actual bio-

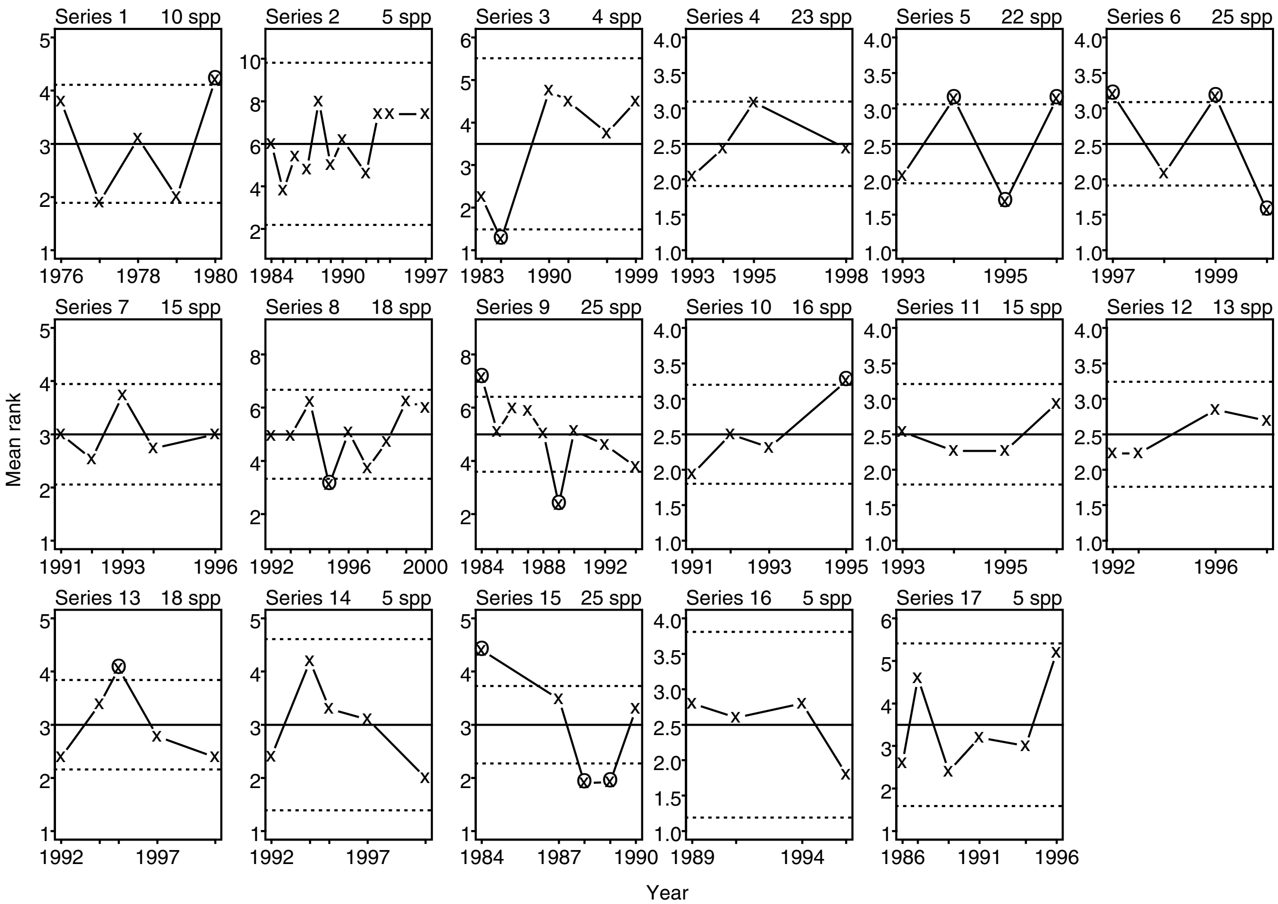

Figure 5

Mean ranks (calculated as illustrated in Table 2) for each trawl survey data set. The solid horizontal line in each panel indicates the overall mean rank; mean ranks outside the broken lines (circled points) indicate extreme survey years.

masses changed by so much. For example, for series 1 the median change in biomass index between 1979 and 1980 (calculated over 10 species) was a factor of 3.4. It is not plausible to say that the biomass of so many species changed by that much in just one year. A second example is years 1988 to 1990 for series 9. Here the median change (over 25 species) was a halving, from 1988 to 1999, followed by a doubling, from 1989 to 1990. Again, it is not plausible to say that the actual biomasses changed by this much.

Are there consistencies between data sets?

The range of estimated trawl survey catchabilities is very wide, covering more than two orders of magnitude, from 0.0035 to 1.6 (Fig. 6). Although the theoretical maximum value for a trawl survey q is 1, the two values that exceed this may not be of concern if we allow for estimation error. However, the lowest values *are* of concern. If these are accurate, then it would seem inappropriate to use trawl surveys to assess these stocks. For example, a q that is less than 0.01 means that more than 99% of the stock is, in some sense, not available to the trawl survey—either because it is outside the survey area (low areal availability), does not encounter the trawl (low vertical availability), or easily avoids it (low vulnerability). For two species the range of values was implausibly wide: for species F the four estimates varied by a factor of 79 (0.0039 to 0.31); for species E the factor was 49 (0.0035 to 0.17) (the next widest range was for species C, a factor of just 2.8).

There is only limited scope for between-series comparisons because the years or seasons covered by different series may not overlap and, in any case, only about one in six years is labeled as extreme. There are three years which were labeled as extreme for more than one series: 1984, 1989, and 1995. In two of these three, the labels are consistent: series 9 and 15 (both of which were deepwater surveys targeting orange roughy, *Hoplostethus atlanticus*, in different areas) agree in finding catchability to be high in winter 1984 and low in winter 1989. For 1995, two surveys found low catchability (series 5 in depths 20–400 m in February and March, and series 8 in 200–800 m in January) and two found high (series 10 in 750–1500 m in October and November, and series 13 in depths 20–400 m in March and April). Given the differences in depth ranges

Table 5

Biomass ratio statistics for "extreme" years (as identified in Fig. 5) in the trawl survey data. For each series, the table has one row for each pair of consecutive surveys that includes one extreme year (extreme years are underlined). A biomass ratio for the two years is calculated for each species; the table presents the median of these ratios, as well as the number of species for which this ratio exceeds 1.5. In each row, the order of the years is such that the expected biomass ratio is greater than 1.

Series	Years	Median ratio	Number of species where ratio exceeds 1.5
1	1980/1979	3.4	8/10
3	1983/1985	1.2	1/4
	1990/1985	3.0	3/4
5	1994/1993	1.8	12/22
	1994/1995	1.6	14/22
	1996/1995	1.9	13/22
6	1997/1998	1.4	11/25
	1999/1998	1.5	14/25
	1999/2000	1.9	18/25
8	1994/1995	1.4	8/18
	1996/1995	1.6	10/18
9	1984/1985	1.7	15/25
	1988/1989	2.0	16/25
	1990/1989	2.1	19/25
10	1995/1993	1.4	7/16
13	1995/1994	1.1	1/18
	1995/1997	1.4	8/18
15	1984/1987	1.4	11/25
	1987/1988	2.3	16/25
	1990/1989	1.9	13/25

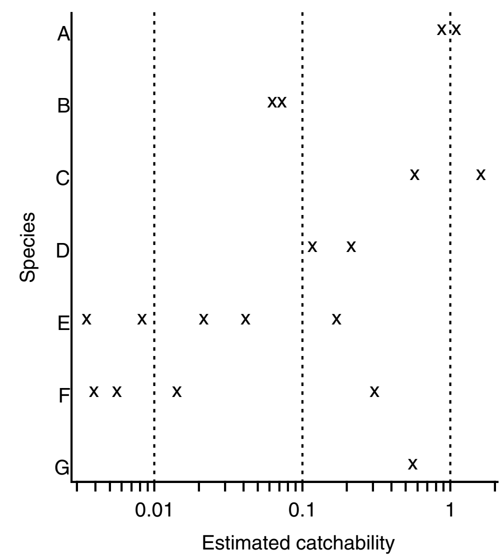

Figure 6

Estimates of trawl survey catchability by species. Each point corresponds to an assessment data set; estimates for different stocks of the same species are plotted in the same horizontal line.

and months for these series, it is unclear how much consistency in catchability could be expected.

In comparing the trawl survey and CPUE series, we found only 12 matches, and the data were consistent at 8 of these (67%). This is not significantly different from the value of 50% that we would expect if the data were uncorrelated (P=0.39). Another place one might look for consistency is between biomass trends for the same species in different survey series. We made plots of biomass trends for every instance where there were at least three survey series with that species and at least three years in common. A few of the 16 such plots showed strong consistency but it was difficult to judge the significance of this because of the possibility of obtaining agreement by chance.

Discussion

It is difficult to make inferences about catchability because we cannot measure it directly. Instead, we must estimate it indirectly with stock assessment models. These estimates are compromised by the weakness of our models, which provide only crude representations of population dynamics (because the data to develop more complex models are not available). With trawl survey data alone we cannot estimate catchability; we can only detect years when catchability was extreme for many species. Nevertheless, the large data sets we have assembled do allow us to draw some conclusions about New Zealand catchabilities.

Are the assessment CVs the right size?

Our results imply that, on average, the CVs used for CPUE in New Zealand are too large, and those used for trawl surveys are too small. For CPUE, the common (but usually tacit) assumption that catchability varies from year to year is supported. It is clear from Figure 3 and Table 3 that, had the CPUE CVs been set equal to the observational error (typically less than 0.1, Francis[2]), the associated stock assessment residuals would have been much too large. However, too much allowance for annual variability seems to have been made: the CVs that are used in stock assessments are, more often that not, too large. In other words, the annual variability in CPUE catchability is not as large as is implied by these CVs. Where the use of a default CV is appropriate, it would seem that values around 0.15–0.2 would be better than the values of 0.3–0.35 that are currently used. This implies that annual variability in CPUE catchability is less that 0.2. For trawl survey indices, the

[2] Francis, R. I. C. C. 1999. The impact of correlations in standardised CPUE indices. N.Z. Fish. Assess. Res. Doc. 99/42, 30 p. National Institute of Water and Atmospheric Research, P.O. Box 14901, Wellington, New Zealand.

results of Table 4 suggest that 0.2 is a reasonable default CV for annual variability in survey catchability. This CV should be "added" to the observation error CVs to obtain a CV for use in stock assessments.

The blanket use of default CVs is clearly undesirable. It is obviously wrong to assume that all CPUE indices have the same CV, regardless of which species or fishery they describe, or the quality and quantity of data from which they are calculated. Similarly, we should expect that annual variability in trawl survey catchability will vary from stock to stock. However, we have little choice in this matter. In most stock assessments we do not have the information to depart from a default value (although there is sometimes evidence that CPUE data sets were unusually weak, Doonan et al.[3]). The above default values imply smaller CVs for CPUE than for trawl surveys. This is surprising and contrary to the prevailing view that trawl surveys are more "reliable" than the CPUE (in the sense defined in the above section on the assessment data). Nevertheless, it is clearly indicated by the data sets examined here.

Can we detect years of extreme trawl survey catchability?

There is clear evidence of extreme years in New Zealand trawl surveys, i.e. years in which the biomass indices for many species are extreme (all low, or all high). However, can we be confident that these extreme years are caused by extremes in catchability? There are two other factors that could cause these extremes.

The first is sampling error, which is associated with the element of chance involved in whether there happen to be many fish at a randomly chosen location at the time it is sampled by the trawl. Because some pairs of species co-occur, we can expect that if we are "lucky" with one species (i.e. we happen to hit dense concentrations of it), then we will tend to be "lucky" with its co-occurring species. Thus, the sampling errors of co-occurring species will be correlated. It seems unlikely that the extreme mean ranks shown in Figure 5 (or the biomass ratios in Table 5) were caused solely by correlated sampling errors. In principle, we should be able to quantify this likelihood. From the survey tow-by-tow data we could infer the extent of between-species correlations at the level of individual stations, from which we could calculate correlations for whole surveys (we would expect more correlations in surveys covering a wider range of species). This information could then be used to calculate the probability of generating biomass ratios as large as those in Table 5. However, to do so would be a major multilevel simulation exercise which is beyond the scope of the present work. What we do know, from other studies, is that between-species correlations, when they exist, are not large. Values of 0.2 to 0.4 seem to be typical (for square-root–transformed catch rates in the same depth stratum, Bull[4]). It does not seem at all likely that such small correlations would cause the very substantial synchronous fluctuations we see in Figure 5 and Table 5.

A second interpretation of the extreme years is that they occur because changes in abundance of co-occurring species are correlated (because fishing that reduces the abundance of one species is likely to do the same for co-occurring species). Table 5 allows a subjective evaluation of the likelihood that biomasses in the extreme years changed by as much as the survey biomass indices suggest. This evaluation is complex because the likelihood depends on the magnitude of the ratios, the number of years between surveys, the number of species involved, and any "adjacent" changes (e.g. for series 9, a large drop in biomass in 1989 is less plausible because it appears to be followed by a large rise in the next year). Thus it is not easy to provide a threshold and say that some changes are plausible but others are not; but there is a clear range of plausibility. At one extreme are the changes associated with 1980 in series 1 and 1989 in series 9; we have argued above that these changes are clearly implausible. At the other extreme the changes for 1995 in series 13 are not as implausible, but it is a matter of judgment as to whether one could call them plausible.

Another point to bear in mind is that if we use observation error CVs (as routinely calculated from trawl survey data) in stock assessments, we obtain residuals that are, more often than not, larger than they ought to be.

We are left with the conclusion that the trawl survey data contain clear evidence that research-vessel catchability does vary significantly from year to year. In most, if not all, of the circled years in Figure 5 the catchability of many species appears to have been either much higher or much lower than normal. This finding is consistent with those of Myers and Cadigan (1995), who expressed this variation in terms of between-age within-year correlations in trawl-survey estimates of numbers at age. Also, Millar and Methot (2002) found evidence of significant departures from mean catchability in four of eight years in the triennial series of trawl surveys carried out on the Pacific coast of the United States. This variation in catchability may be environmentally driven. It would not be difficult to find plausible environmental variables that were extreme in the right years. However, because most of our trawl-survey time series were short we could have little confidence that this correlation was indicative of causation. Another possible cause of variation in catchability is between-survey changes in gear and fishing practice (although care is taken to avoid such changes).

Are there consistencies between data sets?

Our only important result under this heading is that some estimates of trawl survey catchability are not credible. For two species, we found that some estimates were implausi-

[3] Doonan, I. J., P. J. McMillan, R. P. Coburn, and A. C. Hart. 1999. Assessment of OEO 3A black oreo for 1999–2000. N.Z. Fish. Assess. Res. Doc. 99/52, 30 p. National Institute of Water and Atmospheric Research, P.O. Box 14901, Wellington, New Zealand.

[4] Bull, B. 2000. Personal commun. National Institute of Water and Atmospheric Research, P.O. Box 14901, Wellington, New Zealand.

bly low and the variation amongst stocks of the same species was implausibly high. Where possible, this variability should also be examined for CPUE catchabilities (as long as they are comparable—note that we should not compare trawl and long-line catchabilities). We did not make this comparison in our study because it involves adjusting for different reference units of of effort in different CPUE series, which requires specialist knowledge about the individual fisheries and data sets.

Concluding comment

Our analyses have not been able to take account of the practice, in some stock assessments, of using CVs as a measure of the "quality" of a biomass index, rather than its precision. This practice happens when a high CV is assigned to a series (often of CPUE) that is believed not to index biomass well. The intention is to lessen the contribution of the series to the assessment. A problem with this practice is that the judgment of quality is subjective, as is the decision as to how high a CV to assign to represent poor quality. It would be very rare that we had sufficient information to determine whether the judgment of poor quality was justified, and whether the assigned CVs were appropriate. It may be that some of the assessments analyzed above produced a biomass trajectory that was a very good fit to a CPUE series (suggesting that a low CV should have been used) but that the trajectory was wrong because, in this case, CPUE was not proportional to abundance. We cannot distinguish such an outcome from one in which a precise CPUE series indexed abundance well. The practice of assigning CVs subjectively is not desirable. Ideally, we should change the model assumption of proportionality between biomass and index rather than inflate CVs. However, we acknowledge that stock assessment is a very pragmatic discipline in which many compromises are necessary, and we hope that the above results will provide practitioners with empirical evidence to support some of their subjective decisions.

Acknowledgments

We are indebted to the many people who provided trawl survey and assessment data for this work. Dave Gilbert, Larry Paul, Paul Breen, and Stuart Hanchet offered useful comments on an earlier version of this paper, and Richard O'Driscoll detected what would have been an embarrassing data error. We are grateful to Steve Cadrin and an anonymous referee for useful suggestions. This project was partially funded by the New Zealand Ministry of Fisheries under project SAM1999/01.

Literature cited

Arreguín-Sánchez, F.
1996. Catchability: a key parameter for fish stock assessment. Rev. Fish Biol. Fisheries 6(2):221–242.

Francis, R. I. C. C.
1992. Use of risk analysis to assess fishery management strategies: a case study using orange roughy (*Hoplostethus atlanticus*) on the Chatham Rise, New Zealand. Can. J. Fish. Aquat. Sci. 49:922–930.

Francis, R. I. C. C., R. J. Hurst, and J. A. Renwick.
2001. An evaluation of catchability assumptions in New Zealand stock assessments. N.Z. Fish. Assess. Rep. 2001/1, 37 p.

Millar, R. B., and R. D. Methot.
2002. Age-structured meta-analysis of U.S. West Coast rockfish (Scorpaenidae) populations and hierarchical modeling of trawl survey catchabilities. Can. J. Fish. Aquat. Sci. 59:383–392.

Myers, R. A., and N. G. Cadigan.
1995. Statistical analysis of catch-at-age data with correlated errors. Can. J. Fish. Aquat. Sci. 52:1265–1273.

Paloheimo, J. E., and L. M. Dickie.
1964. Abundance and fishing success. Rapp. P-V. Réun. Cons. int. Explor. Mer 155:152–163.

Pope, J. G.
1980. Some consequences for fisheries management of aspects of the behaviour of pelagic fish. Rapp. P-V. Réun. Cons. Int. Explor. Mer 177:466–476.

Punt, A.E., T. I. Walker, B. L. Taylor, and F. Pribac.
2000. Standardization of catch and effort data in a spatially-structured shark fishery. Fish. Res. 45:129–145.

Quinn, T. J., II, and J. S. Collie.
1990. Alternative population models for eastern Bering Sea pollock. Int. North Pac. Fish. Comm. Bull. 50:243–257.

Schaaf, W. E.
1975. Fish population models: potential and actual links to ecological models. *In* Ecological modeling in a resource management framework (C. S. Russell, ed.), p. 211–239. Resources for the Future, Inc., Washington, D.C.

Winters, G. H., and J. P. Wheeler.
1985. Interaction between stock area, stock abundance, and catchability coefficient. Can. J. Fish. Aquat. Sci. 42(5):989–998.

Appendix

Alternative mean rank calculations

In this Appendix we describe some alternative (but unfruitful) mean rank calculations referred to in the "Results" section under the subheading "Can we detect years of extreme trawl survey catchability?"

We tried three variations on the above procedure for identifying extreme years. In each case we were evaluating an alternative hypothesis about the nature of between-species correlations. Each hypothesis leads to a different method of calculating mean ranks (or alternative statistics), and we applied the new method to both the survey data, and to simulated data (to calculate threshold values for the new statistics). If the hypothesis were true we would expect to see more extreme years. In fact, we saw fewer extreme years for all of these alternatives.

First, we repeated the above calculations after omitting species for which the mean CV (see Fig. 2) exceeded 0.4. The idea here is that, for species with high CVs, there is little information in the year-to-year changes in their biomass indices. Thus these indices may mask the synchronous fluctuations in the other species, so that omitting them would produce more extreme years. In some cases it did make the most extreme rank deviations more extreme. However, it also had the effect of increasing the threshold (because the number of species decreased). The net effect was to produce slightly fewer extreme years. There was one additional extreme year—1993 for series 7. However, the following years were no longer deemed extreme: 1980 for series 1; 1994 and 1996 for series 5; and 1988 for series 15.

Our second alternative was based on the idea that environmental changes may affect different species differently. That is, an environmental extreme produces extreme catchability, but this may be high for some species and low for others. To test this we calculated rank deviations for each species and then averaged the rank deviations (rather than averaging the ranks and then calculating deviations). This method identified only four extreme years—three were as in Figure 5 (1984 and 1989 for series 9 and 1995 for series 10) and one was new (1979 for series 1).

The third alternative was a variant on the second. We assumed that the species for each series fall into two groups: one group whose catchabilities are all affected in the same way by environmental changes, and a second group for which the effect is opposite. That is, when catchability is high for the first group it will be low for the second, and vice versa. We calculated the mean ranks as above and then determined, for each species, the Euclidean distance between these mean ranks and 1) the species ranks, and 2) the "inverse" of the species ranks (if a species ranks are, say, 1, 4, 2, 3, then the inverse ranks are 4, 1, 3, 2). When the latter distance was smaller, the species was said to fall into the second group. The ranks for all group-two species were replaced by their inverse ranks and the mean ranks (and thus rank deviations) were recalculated. With this method only two extreme years were found, both of which are extreme in Figure 5 (1984 and 1989 for series 9). Often there was no clear separation between groups one and two. Sometimes (but only when there were few species) group two was empty. We also tried a cluster analysis approach to the identification of groups one and two but this produced no better results.