MINISTRY OF FISHERIES
Te Tautiaki i nga tini a Tangaroa

ISSN 1175-1584

Environmental predictors of hoki year-class strengths: an update

R. I. C. C. Francis
M. G. Hadfield
J. M. Bradford-Grieve
J. A. Renwick
P. J. H. Sutton

New Zealand Fisheries Assessment Report 2005/58
December 2005

Environmental predictors of hoki year-class strengths: an update

R. I. C. C. Francis
M. G. Hadfield
J. M. Bradford-Grieve
J. A. Renwick
P. J. H. Sutton

NIWA
Private Bag 14901
Wellington

New Zealand Fisheries Assessment Report 2005/58
December 2005

Published by Ministry of Fisheries
Wellington
2005

ISSN 1175-1584

©
Ministry of Fisheries
2005

Citation:
Francis, R.I.C.C.; Hadfield, M.G.; Bradford-Grieve, J.M.; Renwick, J.A.; Sutton, P.J.H. (2005).
Environmental predictors of hoki year-class strengths: an update.
New Zealand Fisheries Assessment Report 2005/58. 22 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.; Hadfield, M.G.; Bradford-Grieve, J.M.; Renwick, J.A.; Sutton, P.J.H. (2005). Environmental predictors of hoki year-class strengths: an update

New Zealand Fisheries Assessment Report 2005/58. 22 p.

The relationship between hoki year-class strengths (YCSs) and potential environmental predictors was investigated. The predictors considered were the southern oscillation index, synoptic weather patterns, sea-surface temperatures, wind speeds, and two characteristics of the ocean off the west coast of the South Island: the depth of the mixed layer and nitrate concentration. Four types of predictands were investigated: YCSs for the western, eastern, and combined stocks and the annual proportion of young fish that migrate from the Chatham Rise to the sub-Antarctic. This last proportion is believed to control how much of the total hoki recruitment ends up spawning in each of the two major spawning grounds. Prediction performance was measured by percent variance explained (PVE) and estimated using cross-validation.

Results using simulated data showed the importance of full cross-validation and demonstrated that the mistaken use of a regression predictor which is actually not correlated with our predictand can produce worse results than those from simply using the mean of previous values as a predictor. Also, estimates of PVE for hoki would not be very precise because of the relatively short time series of data (a maximum of 23 years).

Little or no evidence was found that either YCSs or proportions migrating could be predicted from environmental variables. Nitrate concentrations (hypothesised to be important as nutrients for a major prey of hoki larvae) were strongly correlated with western YCSs over the last decade but this correlation was reversed in the previous decade.

1. INTRODUCTION

This report addresses objective 2 of Ministry of Fisheries project HOK200401: *To investigate the prediction of year class strength from environmental variables.* It aims to extend and update earlier work by Livingston (2000) and Bull & Livingston (2001), which found strong correlations between the year-class strength (YCS) of the western hoki stock and various environmental variables. An important reason for updating this work is that Myers (1998) found that many such published correlations were not verified when retested some years later. Also, the time series of predictors and predictands used by Bull & Livingston (2001) have been revised and extended, and new potential predictors and predictands are available. Another innovation in the present study is the use of simulated data to increase our understanding of the problem of measuring prediction performance.

This work is potentially important for the assessment and management of the New Zealand hoki fishery. It is unlikely to have much effect on the estimation of the current status of hoki because that depends mostly on the relative strengths of year classes that have been fully recruited to the fishery for several years, and these are generally already fairly well known from catch-at-age data (Francis 2005). However, the ability to relate YCSs to environmental predictors could have a substantial effect on the short-term projections that are used to infer how the fish stock is likely to respond to different levels of catch in the next few years. The results of these projections depend strongly on information about the strengths of very recent year classes that have not yet fully entered the fishery, and this is related to our understanding of the mechanisms through which the environment might affect survivorship of very young fish. It is in the prediction of these YCSs that this work is likely to have its greatest effect.

Results from previous work on the diet of first feeding hoki larvae (Murdoch 1990), on effect of the environment on the quantity and vertical distribution of larval food (Bradford-Grieve et al. 1996), and on temporal changes in the physical oceanography of the Tasman Sea (Sutton et al. in press) were brought together to form an hypothesis concerning the mechanisms through which environmental variability may affect the survival of first feeding hoki larvae off Westland.

2. MATERIALS

The available materials fall into two classes: predictands (the YCSs, or other quantities, we were trying to predict) and predictors (the environmental variables we used in the prediction).

2.1 Predictands

Both actual and simulated predictands were used, and we start by describing the former, which are summarised in Table 1.

Table 1: Actual (as opposed to simulated) predictands used in this study, and the range of years covered by each. Versions labelled "original" are as used by Bull & Livingston (2001); all others were developed for this study.

Predictand	Version (years covered)	
W YCS	original (1980–1996)	new (1978–2000)
E YCS	original (1980–1996)	new (1978–2000)
Total YCS	survey (1989–2002)	model (1978–2000)
pmigr.w	run 3.4 (1982–2000)	run 3.12 (1982–2000)
pmigr.n	run 3.4 (1982–2000)	run 3.12 (1982–2000)

The "original" predictands, as used by Bull & Livingston (2001), were YCSs estimated using the then-current hoki stock-assessment model (Cordue 2000). Hoki is modelled as two stocks – eastern (E) and western (W) – which are fished in four areas (Figure 1). The stocks have separate spawning and adult-home grounds, but a common nursery ground (Table 2). Estimated YCSs were available for 1978 to 1996 (see table 2 of Bull & Livingston 2001) but the predictands were restricted to 1980 to 1996 (a year class labelled 1980 was spawned in the winter of 1980).

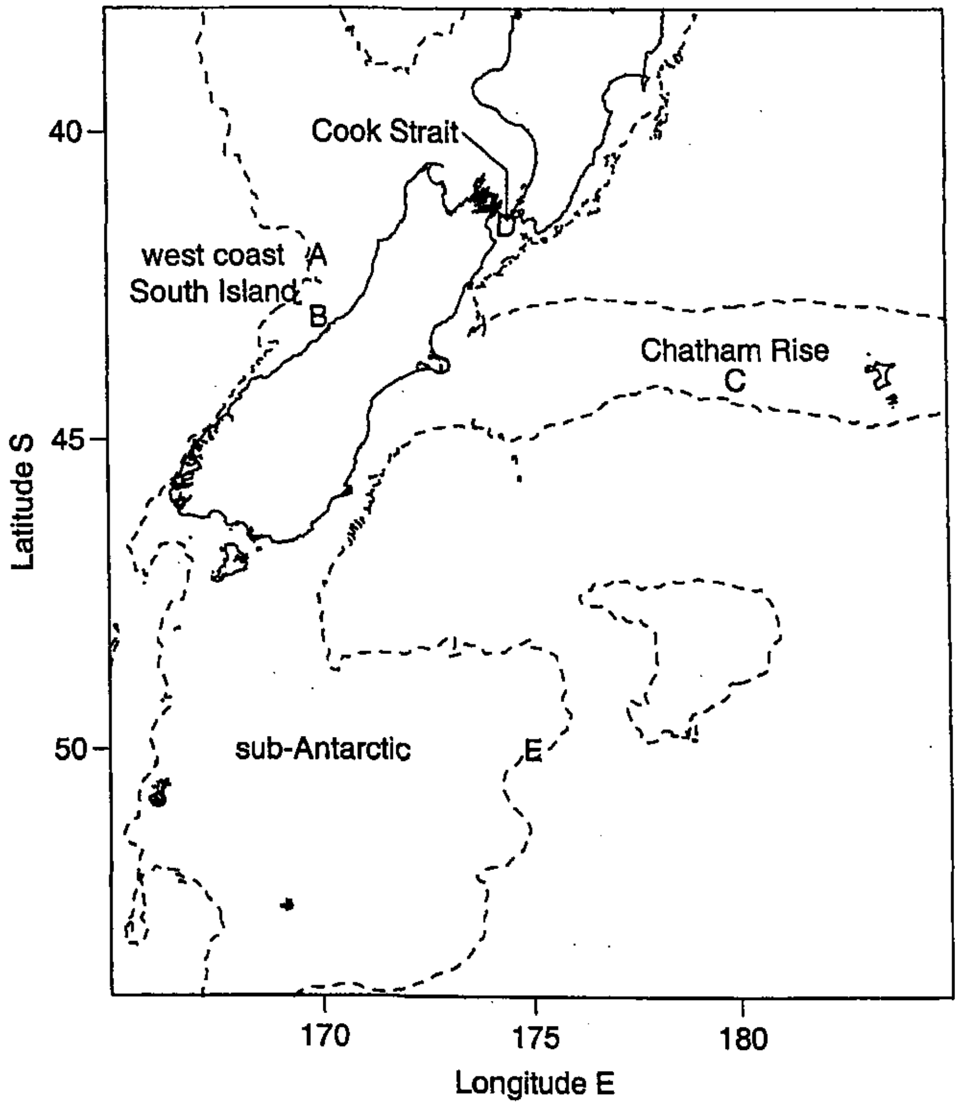

Figure 1: Southern New Zealand, showing the four hoki fishing areas, five locations for environmental variables (A-E), and the 1000 m contour (broken line).

Table 2: The spawning, adult-home, and nursery grounds associated with the two hoki spawning stocks.

Ground	Eastern (E) stock	Western (W) stock
Spawning	Cook Strait	west coast South Island
Adult home	Chatham Rise	sub-Antarctic
Nursery	Chatham Rise	Chatham Rise

The new YCSs used in this study came from the 2004 assessment (Francis 2005), where estimates were made, for both stocks, for all years from 1975 to 2002, inclusive. However, the first three and last two years were ignored in this study because the associated estimates were so imprecise (Figure 2) (it's of interest to note that the spikes in this graph – e.g., at 1986 and 1990 – are associated with weak year classes). Three sets of estimates were available from that assessment, coming from model runs with different assumptions. Only the set associated with run 3.4 is used here because these were highly correlated with the other sets (correlations ranged from 0.979 to 0.998). The YCSs were not corrected for spawner abundance because, for the data available, there is no clear relationship between YCS and spawner abundance (Francis, unpublished results).

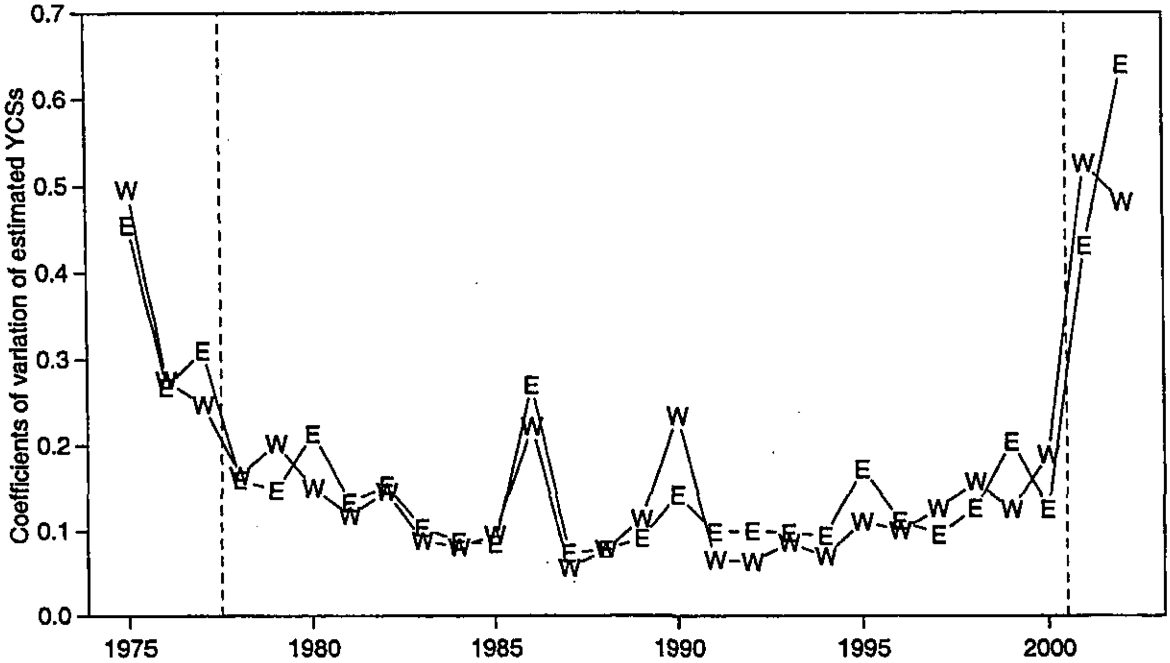

Figure 2: Coefficients of variation (c.v.s) of estimates of E and W YCSs from the 2004 assessment. The years outside the vertical broken lines were ignored in this study because the c.v.s were so high.

These YCSs are broadly similar to those used by Bull & Livingston (2001), but there are some marked differences. The previous E YCSs show wider variation than the present ones and, for the W YCSs, there is a trend in the comparison, with the previous values being lower in early years, and higher in later years, than the present ones (upper panels, Figure 3). Some of these differences will be due to the fact that more data were available to estimate the present YCSs, but others have undoubtedly been caused by differences in the assumptions underlying the two stock assessment models. For example, the earlier assessment model gave more weight to west coast South Island acoustic biomass estimates which showed a strong upward trend in the 1990s. This presumably required the model to estimate stronger W YCSs in the early 1990s to explain this upward trend. The current view is that the acoustic estimates are less certain than previously thought and so should be given less weight (O'Driscoll 2002).

Some additional predictands were used in the present study. These all relate to an alternative stock-structure hypothesis which postulates that there is only one genetic stock for hoki. Under this hypothesis, adult fish may be separated into E and W, according to which spawning and home grounds they use, but there is no between-generation spawning-ground fidelity (so, for example, the offspring of fish that spawn in Cook Strait need not spawn there). As fish move towards adulthood they choose whether to stay on the Chatham Rise and spawn in Cook Strait for the rest of their lives, or to migrate to the sub-Antarctic and spawn in west coast South Island. With this hypothesis we are interested in the total YCS, for which there were two estimates. The first was from the stock assessment model and was formed by adding the estimated recruitments for the two stocks and then rescaling by dividing by the average total recruitment over the period 1989–2002. (This rescaling was done purely for convenience in plotting the YCSs; it has no effect on the results of this study. The W and E YCSs, as output from the stock-assessment model, are also scaled to average to 1 over the years 1989–2002.) As before, we used only the values from run 3.4 because these were highly correlated with those from the other runs ($\rho = 0.99$ and 0.94).

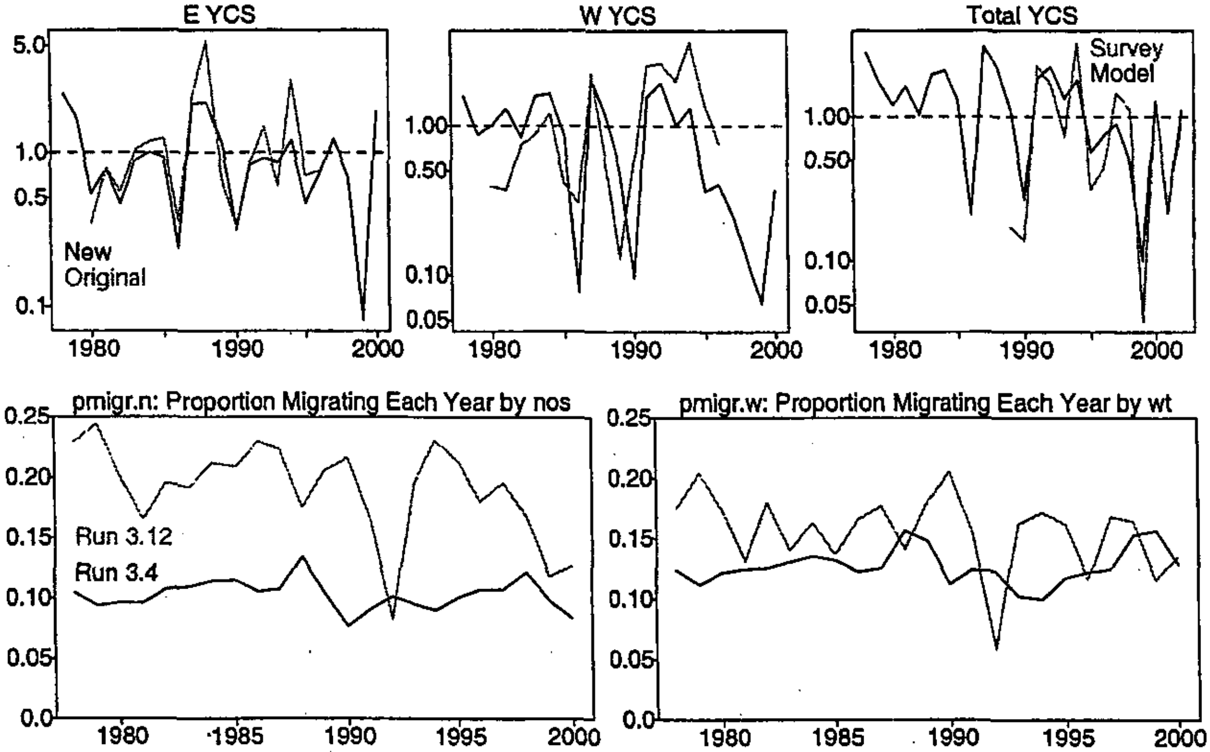

Figure 3: Actual predictands used in this study: YCSs in the upper panels and proportions migrating in the lower panels. Those labelled 'original' are as used by Bull & Livingston (2001). The YCSs are plotted on a log scale, as they are used in this study.

The second way of estimating the total YCSs was from annual summer trawl surveys of the Chatham Rise between 1992 and 2004. The survey index used for a given year class was calculated using a combination of two highly correlated survey estimates: n_1, the estimated number of that year class at age 1.6 y in one year, and n_2, the estimated number at age 2.6 y in the following year (Figure 4). The n_2s were scaled up by a factor of 1.47 (the average of n_1/n_2) and then averaged with the n_1s to give the survey YCSs. The values at the two ends of the time series were just based on a single estimate: n_1 from the 2004 survey for the 2002 year class, and the scaled n_2 from the 1992 survey for the 1989 year class. The survey YCSs are reasonably well correlated with those from the model ($\rho = 0.83$) but show some significant differences (top right panel, Figure 3). Each set of indices has its own advantages and disadvantages. The survey indices are direct estimates of YCS but use only one data set and are restricted to just 14 y. The model indices cover more years and use more data sets (including proportions at age from the four fisheries and two times series of trawl surveys) but require a wide range of assumptions.

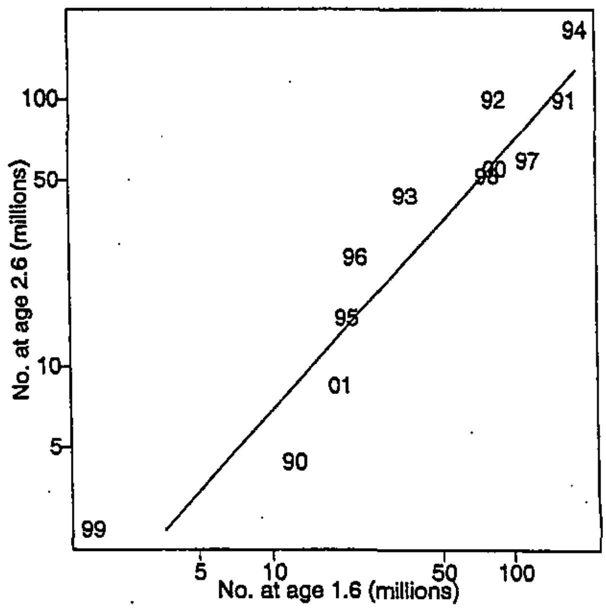

Figure 4: Two trawl survey estimates of YCS: the estimated number at age 1.6 y in one year plotted against the estimated number at age 2.6 y in the following year. The plotting symbols identify the year class (they are the last two digits of the year the year class was spawned) and the diagonal line, $y = x/1.47$, represents the average ratio of these two estimates.

All YCS predictands were used on a log scale. That is, the predictand used was log(YCS), rather than YCS. However, for simplicity, we will refer, for example, to the predictand W YCS, rather than log(W YCS).

The final type of actual predictand used was the annual proportion of young fish on the Chatham Rise that migrate to the sub-Antarctic. With the 1-stock hypothesis it is year-to-year variation in this proportion that causes some year classes to be more abundant in one or the other spawning ground (e.g., the 2000 year class was considered above average in Cook Strait but below average in west coast South Island). Thus an understanding of what drives this variation – in particular, whether it is correlated with any environmental variables – would be useful in managing the hoki fisheries. The proportion migrating was calculated from stock-assessment model outputs as the fraction – by number, pmigr.n, or by weight, pmigr.w – of all fish on the Chatham Rise of age less than 9 y. Indices from runs 3.1 and 3.4 were very similar (ρ = 0.95 and 0.98), but those from run 3.12 were quite different, as were the by-number and by-weight indices. Thus four proportion-migrating predictands were used (lower panels, Figure 3).

2.1.1 Simulated predictands

Simulated predictands are useful because they allow us to determine how well we are likely to be able to estimate prediction performance. Two sets of 20 predictands were simulated. The first set was assumed to be without environmental influence and so was randomly generated from a standard normal distribution. Each predictand was a vector of length 23, nominally covering the years 1978 to 2000.

The second set of simulated predictands was designed to be 80% dependent on environmental variables and 20% random. The environmental contribution to the simulated predictands was assumed to come from the regression equation estimated by Bull & Livingston (2001) for the western stock (see their table 5). This equation was applied to the predictors for years 1978 to 2002, which produced a vector of predicted log(YCS) with variance 0.69. To this was added another vector, of the same length, generated from a normal distribution with mean zero and variance 0.69/4 (this is the random component). This procedure was repeated to generate 20 predictands. Note that the expected total variance for each predictand is 0.86 (= 0.69 + 0.69/4), and 80% of this comes from climate variation.

2.2 Predictors

A total of 84 predictors was used, of which 63 were also used by Bull & Livingston (2001). Each predictor was associated with one of the four seasons (summer (sum) = January–March, autumn (aut) = April–June, winter (win) = July–September, spring (spr) = October–December) (Table 3).

The first two types of predictors do not refer to any specific part of New Zealand. The southern oscillation index (SOI) is the normalised mean sea-level pressure difference between Tahiti and Darwin. Values above 10 indicate La Niña conditions, whereas those below –10 are associated with El Niño. The second type of predictor describes general weather conditions over the New Zealand region using a classification taken from the cluster analysis of Kidson (2000). There are 12 predictors for each season, corresponding to the proportion of days in that season during which the day's weather pattern was assigned to each of the 12 objectively-defined weather classes described by Kidson (2000). These predictors have labels like WP.SW.aut or WP.HSE.spr, where the second element of the label identifies the weather pattern (see figure 1 of Bull & Livingston 2001 for the circulation patterns characteristic of each of the 12 weather classes).

The next two types of predictors were associated with specific locations in each of the four fisheries: west coast South Island (WC; 43°S, 170°E), Chatham Rise (CR; 44°S, 180°E), Cook Strait (CS;

41.5°S, 174.5°E) and sub-Antarctic (SA; 50°S 175°E) (points B–E in Figure 1). Sea-surface temperatures (SSTs) were interpolated from a global 1° x 1° data set (Reynolds et al. 2002). These predictors have labels like SST.WC.win or SST.CR.aut. Near-surface wind speeds were interpolated from the reanalysis fields of Kalnay et al. (1996). The meteorologically standard westerly and southerly components were rotated to derive speeds in the northwesterly and southwesterly directions, which are approximately parallel and perpendicular to the South Island mountain chain, so these predictors have labels like NW.CR.aut or SW.CS.win.

Table 3: Environmental variables used as predictors in this study. Those underlined are new to this study; all others were also used by Bull & Livingston (2001).

		No. of predictors in each season				Seasons used for each predictand			
Type	Area	sum	aut	win	spr	E YCS	W YCS	Tot. YCS	pmigr
Southern oscillation index (SOI)	All	1	1	1	1	aut-spr	aut-spr	aut-spr	sum-aut
Weather patterns (WP)	All	12	12	12	12	aut-spr	aut-spr	aut-spr	sum-aut
Sea-surface temperature (SST)	WC	–	1	1	1	–	aut-spr	aut-spr	–
	CR	1	1	1	–	aut-win	–	aut-win	sum-aut
	CS	–	–	1	1	win-spr	–	win-spr	–
Wind speeds (NW & SE)	WC	–	2	2	2	–	aut-spr	aut-spr	–
	CR	2	2	2	–	aut-win	–	aut-win	sum-aut
	CS	–	–	2	2	win-spr	–	win-spr	–
	SA	2	2	–	–	–	–	–	sum-aut
Mixed-layer depth (MLD6-8)	WC	–	–	1+2	–	–	win	win	–
Nitrate concentration (NIT)	WC	–	–	1	–	–	win	win	–
		Number of predictors:				51	52	64	36

The last two types of predictors derive from a one-dimensional ocean model applied at a point off the west coast of the South Island (42°S, 170°E: point A in Figure 1). They are the mixed-layer depth at 1 June, 1 July, and 1 August (MLD6, MLD7, and MLD8) and the time-integral of surface nitrate concentration between 1 April and 1 August (NIT). The only one of these used by Bull & Livingston (2001) was MLD7. These predictors derive from the hypothesis of Bradford-Grieve et al. (2004) that one of the main factors governing hoki recruitment, derived from the west coast South Island (notionally the western hoki stock), is the availability of dissolved nutrients such as nitrate, to autumn and winter surface waters off Westland. This supply of nutrients is controlled by the underlying physical structure of the interior of the ocean and the timing and extent of the onset of autumn/winter deep mixing to bring nutrients to the surface. These nutrients are needed for the reproduction and growth of copepods of the genus *Calocalanus*, which are a dominant prey item for first-feeding larval hoki in that area (Murdoch 1990). The nutrients are transported up from deeper water by surface mixing, so their concentration is partly dependent on the depth of the mixed layer, which usually increases rapidly, but variably from year to year, at the onset of winter. The ocean model providing these predictors, and the way it was improved for this project are described in the next section.

All predictors were available for all years from 1978 to 2000, except for those from the ocean model which were restricted to 1980 to 2000. Two of the predictor types have been revised since their use by Bull & Livingston (2001): the SSTs and MLD7. The revision to the latter is described in the next section. As to the SSTs, these were revised with new global data (Reynolds et al. 2002). New SSTs from 1982 were blended statistically with those prior to 1982, using a regression procedure. Small changes to the data post-1982, and changes to the regression adjustment coefficients led to differences in some of the SST values used as predictors, especially before 1982.

2.2.1 The ocean model

An ocean model was used to generate the mixed-layer depth and nitrate predictors (MLD6, MLD7, MLD8, and NIT). This model represents the water column at a single location as a vertical stack of layers, each of which exchanges heat, mass, and momentum by turbulent mixing with its neighbours above and below. Model simulations are forced at the surface by time-varying fluxes of heat and momentum. At each layer in the interior there are variables representing temperature, salinity, nutrient concentration, and phytoplankton biomass (the last of these as chlorophyll *a*). The surface fluxes are calculated from meteorological data; when forced by these fluxes the model simulates ocean properties versus depth and time for several years.

The model was first applied and validated by Hadfield & Sharples (1996) for 1980–1993. They found that the modelled SST agreed well with observations, as did the mixed layer depth and surface nitrate concentrations. They found that the modelled surface chlorophyll concentration agreed with the limited shipboard observations available, but they noted that the model simulates a pronounced phytoplankton bloom each spring (with chlorophyll *a* concentrations typically exceeding 2 mg m^{-3}) and there were not adequate data to confirm or refute this phenomenon. With the benefit of the more recent chlorophyll estimates from satellite-sensed ocean colour (e.g., Murphy et al. 2001) we can now be confident that the modelled spring bloom is excessively large, suggesting that the model does not represent phytoplankton losses due to grazing correctly. However, for the present work the key process to be represented is the deepening of the mixed layer in late autumn and early winter, which mixes nutrient-rich deeper water to the surface. All the evidence suggests that the model does represent this process well.

The same model was applied by Hadfield (2000) for a longer period (1980–97). For this work the nutrient and plankton sub-models were disabled and the emphasis was on the model's ability to simulate SST and mixed layer depth. It did very well in this respect, reproducing the observed seasonal cycles in SST and mixed layer depth as well as several features of the interannual variations in SST. However one significant bias was noted: the model underestimates the SST anomalies in the cooler half of the year (May to October). Hadfield (2000) attributed this bias to the fact that the model does not represent oceanographic processes that generate variability in the temperature below the mixed layer.

The properties of the ocean below the mixed layer are represented in the model via the "base state" profiles, which are estimates of the mean variation of temperature, salinity and nitrate with depth. The model fields are continuously nudged towards the base state. Below a depth of 200 m the nudging time scale is 1 year; in this region the nudging process is a proxy for the various oceanographic processes (horizontal and vertical advection, weak vertical mixing) that maintain the ocean state. Above 200 m the nudging rate is reduced and becomes negligible in comparison with processes represented explicitly in the model.

For the modelling exercises described by Hadfield & Sharples (1996) and Hadfield (2000) the base state profiles were simple algebraic expressions fitted to oceanographic soundings near the modelled location. The base state profiles remained constant during the simulations, as there was no basis for estimating their variation with time. However, recently there have been new measurement programmes to monitor thermal variations in the upper ocean on inter-annual and decadal time scales. These programmes include regular transects in the New Zealand region with expendable bathythermographs (Sutton 2001) and global sea surface height measurements with satellite-based altimeters (Willis et al. 2004). Sutton et al (in press) showed that the temperature of the top several hundred metres in the Tasman Sea has varied significantly over the last decade. Bradford-Grieve et al. (2004) noted that Sutton's data show warm conditions between 1996 and 2002, and deduced that nitrate levels may have been low during this period (given the inverse relationship generally observed between temperature and nitrate concentrations). They hypothesised that "one of the main factors governing recruitment to the western hoki stock is availability of nutrients, such as nitrate, to winter surface waters off Westland". Bradford-Grieve et al. (2004) further suggested that the lack of hoki recruitment since 1996 might be related to the warming of the Tasman Sea in the same period, via the temperature-nitrate relationship.

The present modelling exercise used the Hadfield & Sharples (1996) and Hadfield (2000) model, with the period now extended to 1980–2004. The model's base state profiles of temperature and nitrate varied with time and were calculated as follows.

1. A time series of upper ocean temperature anomaly for each year from 1991 to 2004 was extracted from the Tasman Sea XBT data averaged over depths between 150 and 250 m. (This time series approximates very closely a sinusoidal curve with an amplitude of 0.6 K and a period of 14 years; it has a minimum in 1994 and a maximum in 2001.)

2. To extend the time series before 1991, a *proxy* time series was calculated from the SST dataset of Reynolds et al. (2002). The SST-based time series has a single value for each year, calculated by averaging over several consecutive months of the year and over a rectangular region in the southern Tasman Sea. Both the period and the region were varied iteratively to find values that maximised the correlation between the SST-based and XBT-based time series. The parameters finally selected were area 160–168° E, 42–40° S, months June, July, and August, and the correlation coefficient achieved was $r = 0.93$. (This high correlation between winter SST and subsurface temperature was expected, and results from the fact that the growing mixed layer in winter ingests water from deeper down.)

3. A pair of simulations was run in which the base state profile of temperature was perturbed by the anomaly time series, using both the XBT-based and SST-based anomalies respectively. In both cases, the model's hindcast of SST in winter was improved significantly in that the variability was brought closer to observed levels and the correlation with observed winter SST was increased.

4. A temperature-nitrate relationship was established from historical data (Conkright 1998) from the southeast Tasman Sea between 300 and 800 m depth. This relationship was nearly linear, as expected, with a slope s_{TN} of -3.0 mmol m^{-3} K^{-1}. Accordingly, the final simulation was run with the base state temperature profile perturbed with the SST-based temperature anomaly and the base state nitrate profile perturbed with the same anomaly multiplied by s_{TN}. The mixed layer depth and nitrate predictors were calculated from the output of this model.

Calculation of the mixed layer depth and nitrate predictors was straightforward. Mixed layer depth was evaluated from temperature profiles at 1 June, 1 July, and 1 August each year: the mixed layer depth was defined as the depth at which the temperature first drops below the temperature at a reference level (depth 6 m) near the surface minus 0.5 K. The nitrate predictor is a straightforward time-integral of surface nitrate concentration between 1 April and 1 August each year. The surface nitrate concentration is invariably very small at the beginning of this period and increases abruptly at some time in the next few months. The value of the nitrate predictor thus depends on the time when this increase occurs and the rate at which the nitrate concentration increases thereafter.

It turns out that the mixed layer depth and nitrate predictors are affected very little by the introduction of inter-annual variation into the base state profiles of temperature and nitrate, as described above. While there is significant inter-annual variability in the temperature of the upper ocean, and incorporating this effect into the model does improve hindcasts of winter SST, the predictors chosen here are affected mostly by the deepening of the mixed layer in late autumn and early winter, and this is controlled largely by the meteorological forcing.

3. METHODS

The analyses were carried out, separately for each predictand, in four steps:

1. Select candidate predictors for the predictand
2. Select best predictors from set of candidates using stepwise regression
3. Calculate the regression equation using the best predictors
4. Estimate prediction performance by cross-validation.

The approach is very similar to that used by Bull & Livingston (2001) except for an important modification to the cross-validation procedure at step 4.

The aim of step 1 was to remove, from the full set of 84 predictors, any that were unlikely to be related to the predictand. The candidate predictors for each predictand are shown in the right-hand columns of Table 3. For E YCS and W YCS these are exactly as selected by Bull & Livingston (2001), except for the additional predictors from the ocean model. For Total YCS, it seemed reasonable just to combine the predictors from these two sets (under the 1-stock hypothesis there are still two spawning grounds). The migration of juveniles from the Chatham Rise to sub-Antarctic is believed to occur in summer or autumn, so the predictors selected for the pmigr predictands were from these areas and seasons.

The iterative procedure used at step 2 first finds the candidate predictor that is most highly correlated with the predictand. It then determines which of the remaining predictors would, together with the first one selected, be best in predicting the predictand. This process continues, with one predictor being added at a time, until it is determined that, according to a pre-specified stopping rule, there would be no gain in prediction performance from adding any more. There appears to be no general agreement about the best stopping rule to use in these circumstances. The rule used in this study differed from that of Bull & Livingston (2001) but produced broadly similar results (see Appendix 1).

The measure of prediction performance estimated at step 4 is percent variance explained (PVE). This measures prediction performance relative to that from the default estimator, which is just the average of the other YCSs. To understand the cross-validation procedure suppose we dropped one year, say the ith year, from our data set, repeated steps 2 and 3 with this reduced data set, and then applied the regression equation just obtained at step 3 to the environmental data observed in year i to calculate a estimated value, p_i, for the YCS in that year. Then $p_i - y_i$ is the error in this estimate, where y_i is the actual YCS in that year. Our default estimator would be $d_i = \text{mean}_{i' \neq i}(y_{i'})$, and the error in this would be $d_i - y_i$. If we repeat this exercise for every year in the data set we can estimate the mean-square error for our regression estimator as $\text{MSE}_{\text{regression}} = \text{mean}_i (p_i - y_i)^2$ and that for the default estimator as $\text{MSE}_{\text{default}} = \text{mean}_i (d_i - y_i)^2$ The smaller the MSE the better, and PVE measures how much smaller $\text{MSE}_{\text{regression}}$ is than $\text{MSE}_{\text{default}}$, relatively:

$$\text{PVE} = 100\left(\text{MSE}_{\text{default}} - \text{MSE}_{\text{regression}}\right)/\text{MSE}_{\text{default}}.$$

The cross-validation procedure used by Bull & Livingston (2001) was similar to this, though instead of calculating PVE they just reported $\text{RMSE}_{\text{regression}}$ and $\text{RMSE}_{\text{default}}$ (where RMSE is the square root of MSE). However, there was one important difference: in their procedure they repeated only step 3 with each reduced data set, rather than both steps 2 and 3. In the results below we will call their procedure *partial* cross-validation, and the associated PVE the partial PVE.

4. RESULTS

Three sets of results are presented. The first two derive from our simulated predictands and are intended to help us understand the third set, which is based on the actual predictands in Figure 3.

4.1 Simulated predictands with no environmental influence

The candidate set of predictors for these predictands was that given for Total YCS in Table 3, except that the four ocean-model predictors were dropped (because they are not defined before 1980). Two PVEs were calculated for each predictand using these 60 predictors: one with partial, and one with full, cross-validation. The calculations were then repeated, using just the years 1980 to 1996, to see the effect of having a shorter time series. The partial PVEs were always positive (range 20 to 83) whereas 90% of the full PVEs were negative (range –343 to 38) (Figure 5). In both cases the effect of a shorter time series was to produce estimated PVEs which were more variable and greater on average.

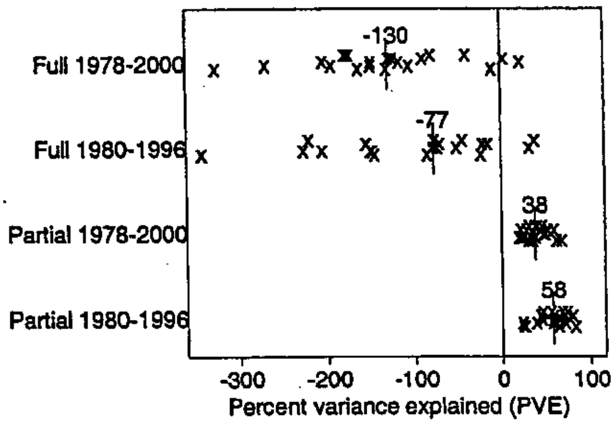

Figure 5: PVEs estimated for 20 purely random predictands, with full or partial cross-validation, and using the whole predictand (from 1978 to 2000), or just the part from 1980 to 1996. The vertical line segments indicate the median estimate for each type of PVE.

It may seem odd to have a negative PVE. However, what this indicates is that the expected error from using a regression estimator is *greater* than that from the default estimator (i.e., $MSE_{regression} > MSE_{default}$). This is illustrated in Figure 6, which shows that although the probability of the regression estimate being closer to the true value than the default estimate is close to 0.5, the former estimate is sometimes very wrong, leading to a mean-square error that is more than double that for the default estimator.

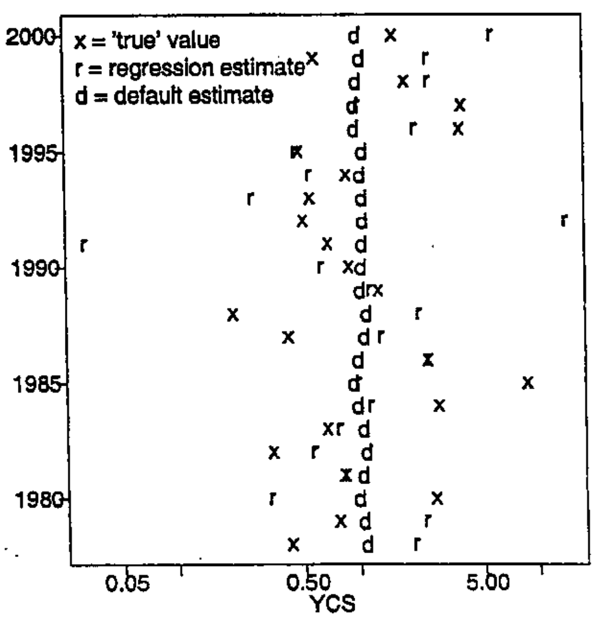

Figure 6: Comparison of the regression ('r') and default ('s') estimates calculated, in the cross-validation procedure, for each year of one of the simulated data sets with no environmental influence. For this data set, $MSE_{regression}$ = 2.11, $MSE_{default}$ = 0.91, and PVE = –132.

4.2 Simulated predictands with environmental influence

For these predictands, the candidate set of predictors was taken as those given for W YCS in Table 3, except that, as before, the four ocean-model predictors were dropped (leaving 48 predictors). The estimated PVEs (with full cross validation) varied quite widely (Figure 7). With these predictands the size of the set of candidate predictors did not have much effect: whether it was reduced to just those 5 predictors in the equation defining the predictands, or increased to the full 80 predictors (all but those from the ocean model), the median PVEs changed only slightly.

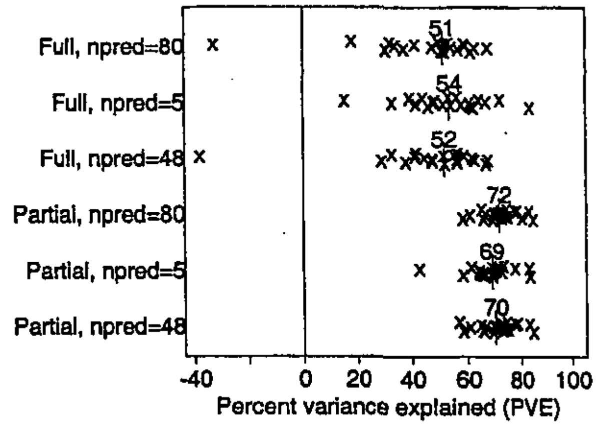

Figure 7: PVEs estimated for 20 simulated predictands that were 80% dependent on environmental variables. PVEs were calculated with full or partial cross-validation, and three alternative sets of candidate predictors (with 5, 48, or 80 predictors).

The two negative PVEs in Figure 7 merit some comment. Because they seemed such extreme outliers the predictand from which they came (the same one in both cases), as well as their calculation, were examined closely for errors. The predictand did not stand out from the others, and no calculation errors were found. We are left with the conclusion that, even when there is a strong climate-recruitment relationship, we may fail to detect it (i.e., calculate a negative PVE) if, by chance, the random non-climate elements affecting recruitment follow some unfortunate pattern for the years for which we have data. To get a better estimate of how likely this would be, a new set of 100 predictands was generated using the same procedure for the original 20, but a different random number seed. With 48 predictors, only two of these 100 produced a negative PVE. Thus, the probability of failing to detect such a strong climate-recruitment relationship with this sort of data is about 0.025 (3 out of 120).

4.3 Actual predictands

PVEs were estimated for 17 different scenarios which differed according to the predictand selected, the version of that predictand, the years used, and the size of the set of candidate predictors (Table 4). We discuss these results separately by type of predictand.

4.3.1 Predicting W YCS

No matter which period was chosen we found no significant ability (as measured by full PVEs) to predict W YCS from our environmental variables.

The partial PVE estimated using the original W YCSs was only 80, markedly lower than the value of 93 that we can calculate from the values of $RMSE_{regression}$ (0.24) and $RMSE_{default}$ (0.89) given by Bull & Livingston (2001). This was due to the change in stopping rule. With the previous stopping rule the partial PVE increased to 93 (the same value, regardless of whether the current or revised SST data were used).

What's of more importance is that the high correlations that existed between the original YCSs and some environmental variables do not exist with the new YCSs. Bull & Livingston (2001) found correlations of –0.84, –0.80, and –0.70 with SOI.aut, SST.WC.spr, and SST.WC.aut, respectively.

The corresponding correlations with the new YCSs were only –0.49, –0.30, –0.17 (for 1980–1996) or –0.34, –0.43, and –0.43 (for 1978–2000, where SST.WC.aut was the most highly correlated of all candidate predictors). Some of this change comes about because the correlation is poor in recent years, but some derives from revisions to earlier YCSs. For example, revisions to YCSs for 1985–86 and 1988–90 substantially degrade the agreement with SOI.aut (Figure 8). The current estimates for these years seem more consistent with the two data sets that are most informative about these YCSs (Figure 9).

Table 4: Estimated prediction performance (percent variance explained) for various scenarios using the predictands of Figure 3. PVEs were calculated with both partial and full cross-validation. For the W YCS and E YCS predictands, the original version was as used by Bull & Livingston (2001) and the new version was as developed for this study.

				Percent variance explained (PVE)	
Predictand	Version	Years	No. of predictors	Partial	Full
W YCS	original	1980–1996	52	80	64
	new	1980–1996	52	36	-30
	new	1980–2000	52	64	1
	new	1978–2000	48[1]	31	-89
E YCS	original	1980–1996	51	31	-106
	new	1980–1996	51	46	34
	new	1978–2000	51	58	-175
Total YCS	survey	1989–2002	64	76	35
		1989–2002	39[2]	76	41
	model	1989–2002	64	70	-54
		1989–2002	39[2]	59	-98
	model	1978–2000	60[1]	75	-111
		1978–2000	39[2]	31	-104
pmigr.w	run 3.4	1982–2000	36	51	17
	run 3.12	1982–2000	36	-6	-78
pmigr.n	run 3.4	1982–2000	36	79	36
	run 3.12	1982–2000	36	5	-204

[1]Excludes ocean-model predictors; [2]Just SOI and WP predictors

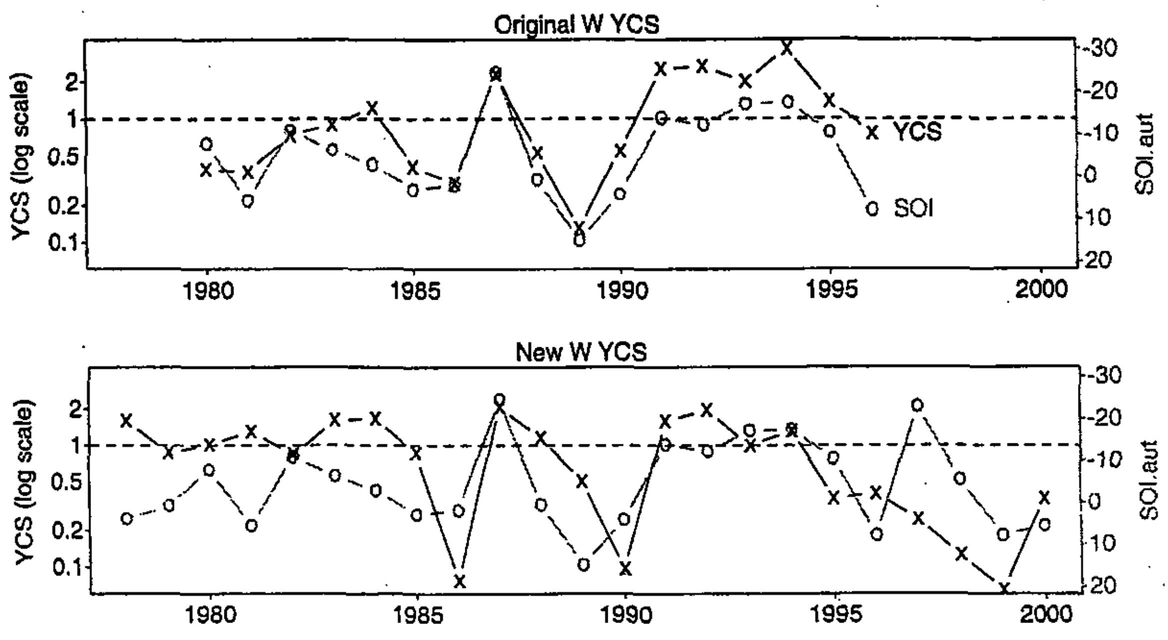

Figure 8: Comparison between two versions (original and new) of the predictand W YCSs (black line, left axis) and the predictor SOI.aut (grey line, right axis).

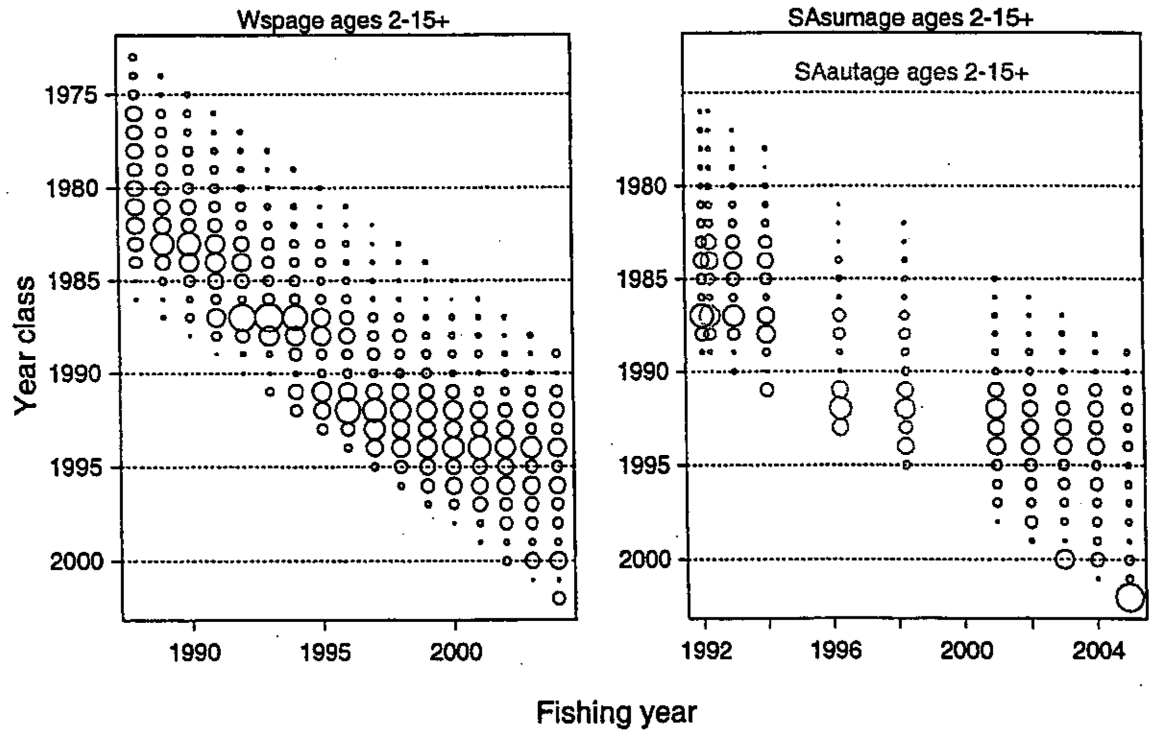

Figure 9: Bubble plots of the three main data sets used in the current stock assessment that contribute to the estimation of W YCS: Wspage (proportions at age in the western spawning fishery) and SAsumage and SAautage (proportions at age in two sub-Antarctic trawl surveys). Each vertical line of circles represents one year's sample, with circle areas proportional to the associated proportion at age (the SAautage data have been offset horizontally to separate them from the SAsumage data). The relative strengths of two year class can be judged by the comparing the sizes of the two corresponding rows of circles.

One surprise, and disappointment, is that none of the ocean-model variables were selected as predictors of W YCS. Correlations for the period 1980–2000 were all less than 0.1 in absolute value. A strong positive correlation with NIT was expected. It's interesting to observe, though hard to explain, that this correlation occurred in the second half of the period but was reversed in the first half (Figure 10).

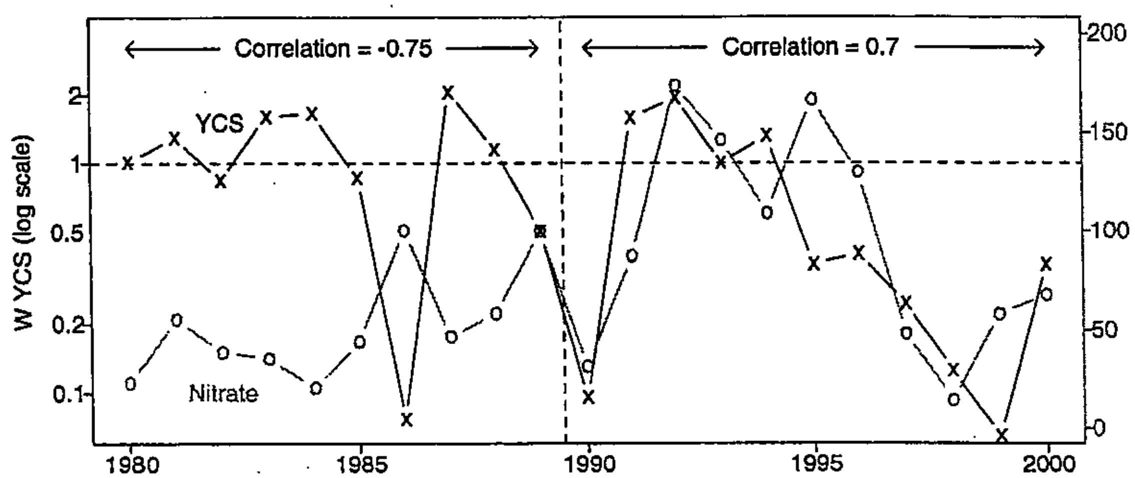

Figure 10: Comparison between the predictand W YCSs (black line, left axis) and the predictor NIT (grey line, right axis).

4.3.2 Predicting E YCS

For the E YCS there appeared to be some slight predictive power over the original period (1980–1996, full PVE = 34) but this disappears over the longer period (1978–2000, full PVE = –175). The strongest correlation in this latter period was only –0.41, with WP.HW.spr. The partial PVE of 31 for the original YCSs was higher than the value of 15 that may be inferred from RMSEs of Bull & Livingston (2001).

4.3.3 Predicting Total YCS

The full PVE values for this predictand suggest that there is some predictive power for the survey YCSs, but not for those from the model. This was true whether the set of candidate predictors was as described in Table 3 (with 64 predictors), or contained just the 39 SOI and WP predictors. The strongest correlation for the survey YCSs was –0.81, with WP.HE.spr (the HE weather pattern is characterised by a high-pressure to the east of New Zealand). This correlation may not persist over a longer time period because the model YCSs also show a reasonably strong correlation (–0.70) over the period used for the survey (1989–2002) but the correlation is lower over the longer time period (-0.53 for 1978–2000) (Figure 11).

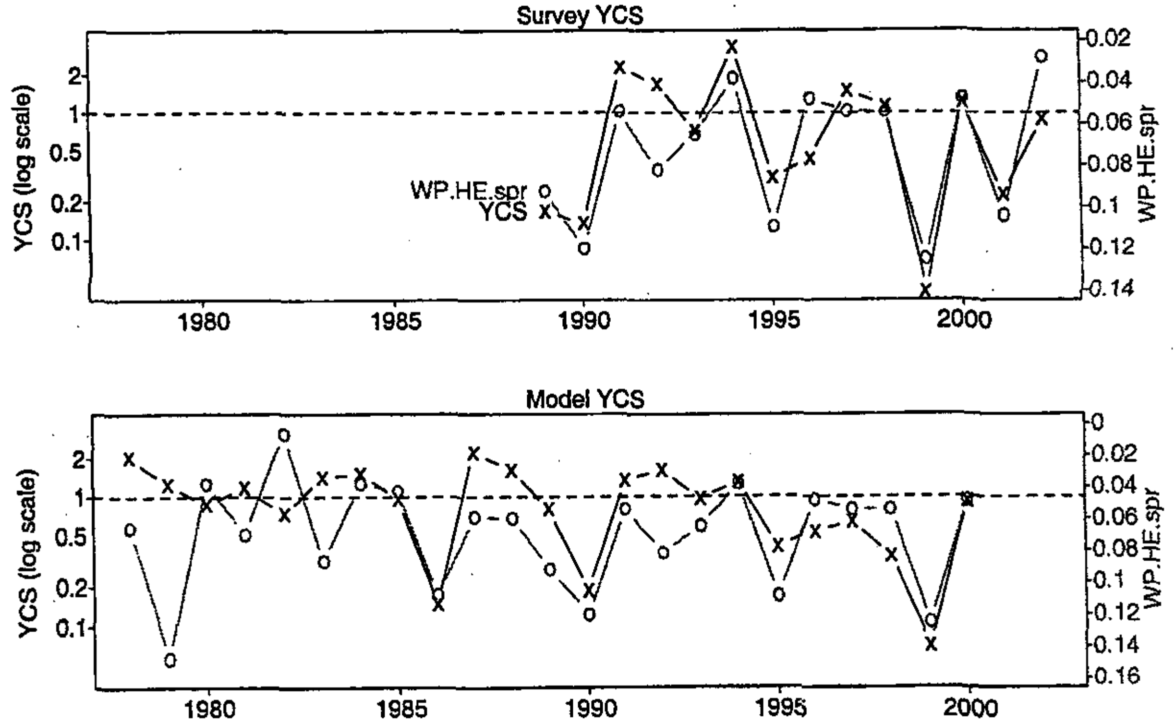

Figure 11: Comparison between two versions (survey and model) of the predictand total YCSs (black line, left axis) and the predictor WP.HE.spr (grey line, right axis).

4.3.4 Predicting pmigr

The full PVE values for these predictands suggest that there is some slight predictive power for the proportions migrating estimated from run 3.4. The highest correlations for the run 3.4 predictors were 0.57 (with SW.SA.aut) for pmigr.w and 0.58 (with WP.HSE.aut) for pmigr.n. However, the fact that no predictive power was found for the predictands from run 3.12 does not increase our confidence in these results.

5. DISCUSSION

We can draw several useful conclusions from the results with simulated data. First, full cross-validation is important because partial cross-validation is likely to suggest that there is some predictive power in environmental variables when, in fact, there isn't (see Figure 5). This is important because the mistaken use of a regression predictor which is actually not correlated with the quantity we wish to predict can produce worse results than those from simply using the mean of previous values as a predictor (see Figure 6). Another point is that, given the length of the time series available for this study (up to 23 years), we cannot expect a very precise estimate of prediction performance (in the form of PVE). If there is a strong causal relationship between some of our predictors and our predictand, the error in our estimate of PVE could easily exceed 20, and there is a small possibility that we might obtain a negative estimate and wrongly conclude that there is no relationship (see Figure 7).

It might be thought that our cross-validation procedure is negatively biased because it produced PVE estimates around 52 in our example in which the simulated predictands were 80% related to our environmental predictors. This is wrong. The point to understand is that we can expect to explain 80% of variance in this case only if we have exact knowledge of which predictor variables are involved and what are the regression coefficients. We achieve less than 80% because we have limited data (only 23 years' worth) with which to identify the predictors and estimate the coefficients. Our simulation experiment shows that, in this particular case, uncertainty about the regression coefficients was more important than uncertainty about which were the right predictors (because the estimates of PVE did not vary substantially as the number of predictors varied).

Our results for hoki are disappointing. We appear to have little or no ability to predict YCSs or pmigr for this population. The apparent reversal of the correlation with nitrate concentrations is tantalising (though puzzling) and may repay further study. Certainly it seems preferable to consider predictors like this, for which there is a clear and plausible supporting hypothesis, than to simply assemble many environmental variables and hope to include some that are causally related to our predictands.

One possible explanation for our failure to find YCS predictors is that our hoki stock hypothesis may be faulty. This hypothesis (summarised in Table 2) is used by the stock-assessment model in estimating the YCSs associated with the two spawning grounds. If this is wrong (e.g., if not all fish spawn where they themselves were spawned) then the model's estimates of YCSs for the two spawning stocks will be wrong, and this will compromise our ability to find links between these YCSs and environmental conditions near the spawning grounds.

In the context of the international literature these results are not surprising. It is common to find that published relationships between YCS and environmental variables are not, on subsequent retesting, found to be valid (Myers 1998). There are at least four explanations for this. The first, but perhaps least likely, is regime shifts. Over large areas the environment is known to undergo major fluctuations on time scales of several decades. Non-linear responses of biological systems to these fluctuations can cause sudden changes in ecosystems, known as regime shifts. One example of a large-scale fluctuation that could be of relevance to hoki is the Pacific Decadal Oscillation (Mantua & Hare 2002), although the timing of recent reversals of this system – in 1977, and possibly near the end of the record, 2001–04 – do not make it a likely explanation for any differences between our results and those of Bull & Livingston (2001). Because regime shifts occur relatively infrequently they are likely to explain only a small proportion of instances of published climate-YCS relationships being subsequently rejected. A second explanation is that authors are insufficiently rigorous in testing the significance of these relationships. For example, a lack of, or only partial, cross-validation is insufficiently rigorous. A third explanation is that many environmental factors contribute to the success or failure of recruitment but, by chance, some will dominate over a decade or more. The last possible explanation is that new information and/or new analytical techniques may substantially change our estimates of the quantity we are trying to predict. This was certainly a factor with the W YCS predictand in comparing the present study with that of Bull & Livingston (2001).

6. ACKNOWLEDGMENTS

We are grateful to Richard O'Driscoll for his characteristically thorough review of an earlier version of this report, and to Mary Livingston for suggesting we investigate the total YCS predictands. This study was funded by the Ministry of Fisheries as part of their project HOK200401.

7. REFERENCES

Bradford-Grieve, J.; Livingston, M.E.; Sutton, P.; Hadfield, M.G. (2004). Ocean variability and hoki decline. *Water and Atmosphere. 12(4)*: 20–21.

Bradford-Grieve, J.M.; Murdoch, R.C.; James, M.R.; Oliver, M.; Hall, J. (1996). Vertical distribution of zooplankton >39 mm in relation to the physical environment off the west coast, South Island, New Zealand. *New Zealand Journal of Marine and Freshwater Research 30*: 285–300.

Bull, B.; Livingston, M.E. (2001). Links between climate variation and year class strength of New Zealand hoki (*Macruronus novaezelandiae*): an update. *New Zealand Journal of Marine and Freshwater Research 35*: 871–880.

Conkright, M.E., et al. (1998). World Ocean Database 1998: documentation and quality control. Ocean Climate Laboratory, National Oceanographic Data Center Internal Report 14

Cordue, P.L. (2000). MIAEL estimation of biomass and fishery indicators for the 1999 assessment of hoki stocks. *New Zealand Fisheries Assessment Report 2000/10*. 69 p.

Draper, N.R.; Smith, H. (1981). Applied regression analysis. 2nd edition. John Wiley and Sons, New York.

Francis, R.I.C.C. (2005). Assessment of hoki (*Macruronus novaezelandiae*) in 2004. *New Zealand Fisheries Assessment Report 2005/35*. 97 p.

Hadfield, M.G. (2000). Atmospheric effects on upper-ocean temperature in the southeast Tasman Sea. *Journal of Physical Oceanography 30(12)*: 3239–3248.

Hadfield, M.G.; Sharples, J. (1996). Modelling mixed layer depth and plankton biomass of the west coast of the South Island. *Journal of Marine Systems 8*: 1–29.

Kalnay, E.; Kanamitsu, M.; Kistler, R.; Collins, W.; Deaven, D.; Gandin, L.; Iredell, M.; Saha, S.; White, G.; Woollen, J.; Zhu, Y.; Chelliah, M.; Ebisuzaki, W.; Higgins, W.; Janowiak, J.; Mo, K.C.; Ropelewski, C.; Wang, J.; Leetmaa, A.; Reynolds, R.; Jenne, R.; Joseph, D. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society 77(3)*: 437–471.

Kidson, J.W. (2000). An analysis of New Zealand synoptic types and their use in defining weather regimes. *International Journal of Climatology 20*: 299–316.

Livingston, M.E. (2000). Links between climate variation and the year class strength of New Zealand hoki (*Macruronus novaezelandiae* Hector). *New Zealand Journal of Marine and Freshwater Research 34(1)*: 55–69.

Mallows, C.L. (1973). Some comments on C_p. *Technometrics 15*: 661–675.

Mantua, N.J.; Hare, S.R. (2002). The Pacific decadal oscillation. *Journal of Oceanography 58*: 25–44.

Murdoch, R.C. (1990). Diet of hoki larvae (*Macruronus novaezelandiae*) off Westland, New Zealand. *New Zealand Journal of Marine and Freshwater Research 24*: 519–527.

Murphy, R.J.; Pinkerton, M.H.; Richardson, K.M.; Bradford-Grieve, J.M.; Boyd, P.W. (2001). Phytoplankton distributions around New Zealand derived from SeaWIFS remotely-sensed ocean colour data. *New Zealand Journal of Marine and Freshwater Research 35*: 343–362.

Myers, R.A. (1998). When do environment-recruitment correlations work? *Reviews in Fish Biology and Fisheries 8*: 283–305.

O'Driscoll, R.L. (2002). Review of acoustic data inputs for the 2002 hoki stock assessment. *New Zealand Fisheries Assessment Report 2002/36*. 66 p.

Reynolds, J.D.; Rayner, N.A.; Smith, T.M.; Stokes, D.C.; Wang, W. (2002). An improved in situ and satellite SST analysis for climate. *Journal of Climate 15*: 1609–1625.

Sutton, P.J.H. (2001). Hours to decades: the nature of variability and change. *Water and Atmosphere 9(4)*: 22–23.

Sutton, P.J.H.; Bowen, M.M.; Roemmich, D. (in press). Decadal temperature changes in the Tasman Sea. *New Zealand Journal of Marine and Freshwater Research.*

Willis, J.K.; Roemmich, D.; Cornuelle, B.D. (2004). Interannual variability in upper ocean heat content, temperature, and thermosteric expansion on global scales. *Journal of Geophysical Research 109*: C12036, doi:12010.11029/12003JC002260.

Appendix 1: Stopping rules for stepwise regression

In this Appendix we compare the stopping rule used in this report with that used by Bull & Livingston (2001) and show some of the effects of the difference between them.

The stopping rule describes how we answer the second of two questions that are asked at each step of a forward stepwise regression. The first question is, which is the best of the predictors not yet included in the regression? The second is, would the addition of this predictor increase the performance of the regression sufficiently to merit its inclusion? The stepwise procedure stops when the answer to the second question is 'no'

First, some notation. At the pth step we already have a regression with p parameters (and thus $p-1$ predictors) and are deciding whether to add an additional predictor. Let RSS_p be the residual sum of squares for the existing regression, and RSS_{p+1} that for the regression containing the next best predictor.

Both stopping rules can be expressed in terms of these quantities. The rule used in this report is the default one in version 3.4 of Splus, and is based on the C_p statistic (Mallows 1973). The procedure stops if $RSS_{p+1} > RSS_p - 2\ RSS_1/(n-1)$, where n is the number of observations (i.e., the length of the predictand). The rule used by Bull & Livingston (2001) applies what is sometimes called the partial F test (Draper & Smith 1981) of the significance of the additional predictor. The procedure stops if $RSS_{p+1} > RSS_p/[1+F_{1,n-p-1,\alpha}/(n-p-1)]$, where α is the significance level of the test and $F_{1,\nu,\alpha}$ is the $(1-\alpha)$ quantile of an F distribution with 1 and ν degrees of freedom.

There appears to be no consensus as to which of these rules is better. What is apparent from the equations in the preceding paragraph is that neither rule is consistently more conservative than the other. When the calculations for Table 4 were repeated using the other stopping rule it was found that the PVE values for full cross-validation were broadly similar to those in Table 4 (Figure A1), despite the fact that the two methods can lead to quite different sets of predictors (Table A1).

Table A1: Comparison of predictors selected in the calculations of Table 4 (using the 'new' stopping rule) with those selected when the calculations were repeated with the stopping rule of Bull & Livingston (2001) (the 'old' rule). The 17 lines in this table correspond with those in Table 4.

Predictors in both regressions	Additional predictors using new rule	Additional predictors with old rule
SOI.aut, WP.TNW.win	-	SST.WC.win, WP.H.spr
WP.TNW.aut	-	-
SST.WC.aut, WP.HE.spr, WP.NE.aut	-	-
SST.WC.aut, WP.NE.aut, WP.HE.spr	-	-
WP.TNW.aut	WP.SW.win	-
WP.TNW.aut	-	-
-	WP.HW.spr, WP.W.aut, WP.HNW.aut, SST.CS.win	-
WP.HE.spr, SST.WC.aut	-	WP.TSW.aut
WP.HE.spr, WP.TSW.aut	-	WP.T.aut
SST.CR.aut, WP.HE.spr	-	NW.WC.aut, WP.TNW.win
WP.HE.spr, WP.TSW.win	-	WP.R.aut, WP.TSW.aut
SST.WC.aut, WP.SW.aut, WP.T.spr, SW.CR.win, NW.CS.win	-	WP.R.spr, WP.W.win, WP.HE.spr
WP.TNW.aut, WP.H.win	WP.TSW.aut, WP.TNW.spr, WP.HNW.spr	-
SW.SA.aut, SOI.aut	-	WP.T.aut
-	WP.W.aut	-
WP.HSE.aut, WP.HW.sum	-	WP.R.sum, NW.SA.sum
-	NW.SA.sum, WP.H.aut, NW.CR.aut, WP.R.sum	-

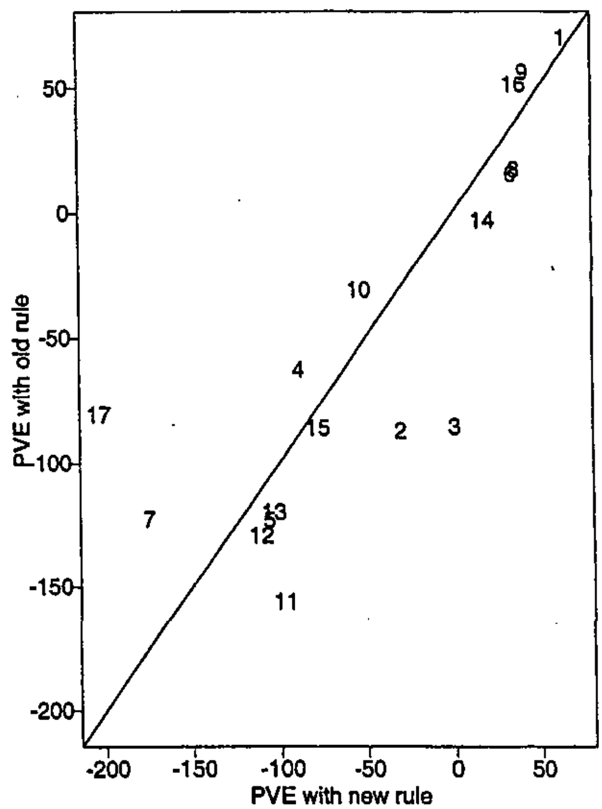

Figure A1: Comparison of the full-validation PVE values in Table 4 (calculated with the 'new' stopping rule) with those calculated using the stopping rule of Bull & Livingston (2001) (the 'old' rule). The plotting symbols correspond to line numbers in Table 4.