Transactions of the American Fisheries Society 137:720–735, 2008

DOI: 10.1577/T06-272.1

[Article]

An Index-Removal Abundance Estimator That Allows for Seasonal Change in Catchability, with Application to Southern Rock Lobster *Jasus edwardsii*

THOMAS F. IHDE*[1] AND JOHN M. HOENIG

Virginia Institute of Marine Science, College of William and Mary, Post Office Box 1346, Gloucester Point, Virginia 23062, USA

STEWART D. FRUSHER

Tasmanian Aquaculture and Fisheries Institute, University of Tasmania, GPO Box 252-49, Hobart, Tasmania, Australia 7001

Abstract.—The index-removal method provides estimates of abundance, exploitation rate, and catchability coefficient. Estimates from the original method suffer from poor precision. Recent work has improved the precision of model estimates; however, the method still includes the strong assumption of constant survey catchability over years and seasons. This assumption is not tenable in many fisheries. This work introduces a new multiyear model, 2qIR, that allows catchability to differ between surveys of the same year. Simulations were performed to examine the effects of variability in (1) the exploitation rate among years, (2) survey catchability, and (3) the number of years of data on model performance. The 2qIR model estimates were always more accurate and precise than those of the other models examined and other model scenarios in which there was moderate contrast in exploitation among years, regardless of the seasonal difference between survey catchability coefficients. The ratios of survey catchability tested ranged from 0.1 to 10, but the model worked best at catchability ratios greater than 0.3. The 2qIR model performance improved slightly when a third year was added to the data set, but performance was similar with 3 or 5 years of data. In all types of simulations, the 2qIR model estimates were usable (i.e., not negative, infinite, or made with a convergence error) a greater proportion of the time than were annual model estimates. The 2qIR model produced reasonable results when applied to data from a population of southern rock lobster *Jasus edwardsii* in Tasmania, whereas the models that assume constant catchability among surveys sometimes predicted exploitation rates exceeding 100%. The results from both the simulations and the lobster data suggest that the 2qIR model can be reliably applied in more situations than models that assume constant survey catchability.

Index-removal (IR) models estimate the abundance and survey catchability coefficient in a population that experiences a relatively large, known removal. The method requires that a survey index be obtained before and after the removal and assumes that the population is closed except for the known removals (i.e., that there is no recruitment, immigration, or emigration between surveys and the time between surveys is short enough that no natural mortality occurs). Though the original method (hereafter, the "annual model") is attractive (Dawe et al. 1993) and has been known for some time (Petrides 1949), it has received only moderate development (Hoenig and Pollock 1998). This may be because annual model estimates often have poor precision (Routledge 1989; Roseberry and Woolfe 1991; Chen et al. 1998a). Ihde et al. (2008) demonstrated that precision could be improved by simultaneously estimating parameters for multiple years of data. The multiple-year index-removal ("1qIR") model (Ihde et al. 2008) shares the assumptions of the annual model (the closed-population assumption is not required across years) and further assumes that the catchability of the survey gear remains constant across years and seasons. However, in practice, survey catchability may be affected seasonally by a variety of factors, such as changes in water temperature (Paloheimo 1963), life history stage (Ziegler et al. 2002), or fishing gear. If catchability varies seasonally, both the annual model and 1qIR will provide biased results, a decrease in catchability over the season causing a negative bias in the population estimate and a positive bias in the exploitation estimate.

In this paper we develop and test a multiple-year IR model, the 2qIR model, which allows catchability to vary by season. The 2qIR model can be used when (1) pre- and postharvest survey indices of abundance have been obtained in at least 2 years, (2) the exploitation

* Corresponding author: ihde@cbl.umces.edu
[1] Present address: Chesapeake Biological Laboratory, University of Maryland Center for Environmental Science, Post Office Box 38, Solomons, Maryland 20688, USA.

Received December 8, 2006; accepted November 6, 2007
Published online May 5, 2008

rate varies among years, and (3) the seasonal catchability coefficients remain constant across years. We use simulation to evaluate the performance of the 2qIR model. We then apply the model to a fishery for southern rock lobster *Jasus edwardsii* off Tasmania, Australia.

Methods

Model Development

The annual and 1qIR models.—Both the annual and the 1qIR models have been described previously and are only briefly reviewed here. The annual model was described by Petrides (1949), and its performance was evaluated by Eberhardt (1982). The 1qIR model was described and evaluated by Ihde et al. (2008).

For the annual model, assume that catch I_j in survey j (for $j = 1, 2$) is distributed as a Poisson random variable, $I_j \sim \text{Poisson}(\lambda_j)$ and the mean λ_j is modeled as $\lambda_j = qN_j f_j$, where N_j is the population size at the time of survey j, f_j is the sampling effort expended in survey j, and q is the catchability coefficient. That is, survey catch is proportional to abundance and sampling effort.

The assumption of Poisson-distributed survey data is used throughout the work presented here. This distribution was employed for its simplicity of presentation and because it has been applied previously, both in the development of the IR method (Eberhardt 1982; Chen et al. 1998b) and for parameter estimation for the management of the southern rock lobster fishery (Frusher et al. 1998, 2003).

Let $N_2 = N_1 - R$, where R is the removal between surveys. The likelihood function, Λ_{ann}, for the annual model is

$$\Lambda_{\text{ann}} = \prod_{j=1}^{2} \frac{(qN_j f_j)^{I_j} e^{-qN_j f_j}}{I_j!}. \tag{1}$$

For the multiyear 1qIR model, we generalize the notation by adding a second subscript to account for year. Thus, N_{ij} refers to the abundance at the time of survey j in year i, and similarly for f_{ij} and I_{ij}. The likelihood function, Λ_{1qIR}, for the model for n years of data is

$$\Lambda_{\text{1qIR}} = \prod_{i=1}^{n} \prod_{j=1}^{2} \frac{(qN_{ij} f_{ij})^{I_{ij}} e^{-qN_{ij} f_{ij}}}{I_{ij}!}, \tag{2}$$

with $N_{i2} = N_{i1} - R_i$, where R_i = removal in year i.

The 2qIR model.—The development of the seasonal-q model, 2qIR, follows that of the annual and 1qIR models as described by Ihde et al. (2008), where survey catches are assumed to be Poisson random variables. If the pre- and postharvest catchability coefficients differ but are constant over years, the likelihood function, Λ_{2qIR}, for n years of data is

$$\Lambda_{\text{2qIR}} = \prod_{i=1}^{n} \prod_{j=1}^{2} \frac{(q_j N_{ij} f_{ij})^{I_{ij}} e^{-q_j N_{ij} f_{ij}}}{I_{ij}!}, \tag{3}$$

with q_j referring to the catchability coefficient in season j. A more generalized model could incorporate k surveys. However, the corresponding removals must be known for the time period between each pair of successive survey indices. For simplicity, we assume that only two surveys are conducted per year. For this case, 2qIR requires a minimum of 2 years of pre- and postharvest indices of abundance and different exploitation rates in at least 2 years. After 2 years of data collection, we have four survey indices that can be modeled as a system of four equations with four unknown parameters, that is,

$$E(I_{11}) = q_1 f_{11} N_1 \tag{4a}$$

$$E(I_{12}) = q_2 f_{12} (N_1 - R_1) \tag{4b}$$

$$E(I_{21}) = q_1 f_{21} N_2 \tag{4c}$$

$$E(I_{22}) = q_2 f_{22} (N_2 - R_2), \tag{4d}$$

where E denotes expectation. The four expected values can be replaced with observed survey indices and the four equations solved simultaneously to obtain moment estimates of the parameters. Without contrast in exploitation rates between years, the four equations reduce to two sets of replicate observations, which is insufficient to uniquely estimate four parameters. Parameter estimates can be calculated analytically when 2 years of data are available:

$$\hat{N}_1 = \frac{I_{11}(I_{12}R_2 - I_{22}R_1)}{I_{12}I_{21} - I_{11}I_{22}} \tag{5}$$

$$\hat{N}_2 = \frac{I_{21}(I_{12}R_2 - I_{22}R_1)}{I_{12}I_{21} - I_{11}I_{22}} \tag{6}$$

$$\hat{q}_1 = \frac{I_{12}I_{21} - I_{11}I_{22}}{I_{12}R_2 - I_{22}R_1} \tag{7}$$

$$\hat{q}_2 = \frac{I_{12}I_{21} - I_{11}I_{22}}{I_{11}R_2 - I_{21}R_1} \tag{8}$$

where the carets denote estimates, I_{ij} = catch in survey j of year i, and the other symbols are as before. For clarity of exposition, survey effort (f_{ij}) for equations 5–8 was assumed to be constant over all surveys in both years and, for convenience, was set equal to 1 (if survey effort differed from 1, each I_{ij} term would also be divided by its corresponding f_{ij}). When more than 2 years of data are available, nonlinear maximization software is required to

TABLE 1.—Degrees of freedom for three index-removal models that estimate abundance and exploitation—annual, multiple year (1qIR), and a multiple-year model that allows survey catchability to vary seasonally (2qIR). See text for details; n is the number of years for which data are available.

Model	Number of observations	Number of parameters to estimate	df
Annual	$2n$	$2n$	0
1qIR	$2n$	$n + 1$	$n - 1$
2qIR	$2n$	$n + 2$	$n - 2$, for $n \geq 2$

make parameter estimates. Degrees of freedom accumulate as years of data are added (Table 1).

Model Evaluation by Simulation

We performed three types of simulations. In the first, the effect of exploitation rate (u) on model performance was studied. In the second, we compared model performance across a range of values for the catchability coefficient (q) in the second surveys. For the first two simulation types, 2 years of data were analyzed. In a third type of simulation, the effect of increasing the number of years of data on model performance was studied.

Survey data were generated by Monte Carlo simulation. The data used in all comparisons were Poisson random variables, that is,

$$I_{ij} \sim \text{Poisson}(q_j \cdot f_{ij} \cdot N_{ij}), \quad i = 1, 2, \ldots, n, \quad j = 1, 2, \text{ and } N_{i2} = N_{i1} - R_i \tag{9}$$

and were created by application of the "rpois" function in S-PLUS statistical software (MathSoft 2000). Survey effort was assumed to be constant over all surveys in all years and, for convenience, was set equal to 1.

When the pre- and postharvest catch rates are similar in magnitude, extremely large abundance estimates can result for all IR models. When the postharvest survey catch equals or exceeds the preharvest survey catch, annual estimates of abundance are infeasible (i.e., negative or infinite) and multiple-year model estimates may also be infeasible. Additionally, multiple-year models that make parameter estimates by nonlinear maximization may fail to converge on solutions. Simulations with infeasible estimates and simulations with convergence failures were counted, excluded from further analyses, and defined as "unusable." Chen et al. (1998b) concluded that the mean and variance were unreliable indicators of the performance of the estimator because extreme values of the estimates sometimes occur. Consequently, we used the median of the usable estimates (i.e., those that were not "unusable") and the spread of the central 95% of the usable estimates as the performance measures of the estimators.

Model performance was evaluated in terms of (1) the proximity of the median usable estimate to the known abundance (as a measure of accuracy), (2) the spread of the central 95% of the usable estimates (as a measure of precision), and (3) the percentage of unusable simulations (as a measure of the model failure rate).

Exploitation rate variation.—The performance of the 2qIR model was examined over a range of contrasting exploitation rates between years when only 2 years of data were available. The exploitation rate was fixed at 10% for all scenarios in the first year but varied from 10% to 80% in the second year. Each simulation was performed at three different levels of contrast between pre- and postharvest catchability coefficients. The simulation parameters were as follows:

(1) population size prior to removals:

$$N_{11} = N_{21} = 1{,}000{,}000 \text{ animals}$$

(2) catchability:

$$q_1 f = 2 \times 10^{-4} \quad \text{and} \quad q_2 f = 1 \times 10^{-4}, \quad \text{or}$$
$$q_1 f = 1 \times 10^{-4} \quad \text{and} \quad q_2 f = 2 \times 10^{-4},$$
$$q_1 f = q_2 f = 1 \times 10^{-4}$$

(3) survey effort: $f = 1$

(4) removals:

$$R_1 = u_1 \cdot N_{11} = 100{,}000$$
$$R_2 = u_2 \cdot N_{21} \quad \text{for}$$
$$u_2 = 0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.5,\ 0.6,\ 0.7,\ 0.8,$$
where $u_i = R_i / N_{i1}$ is the exploitation rate

(5) the number of years of data: $n = 2$.

Survey catchability, q_j, is multiplied by effort, f, because it is the product qf that is of interest. Survey data were simulated 10,000 times for each comparison.

Catchability variation.—The performance of the 2qIR model was evaluated over a range of contrasts between pre- and postsurvey catchability coefficients when only 2 years of data were available. Catchability for the preharvest surveys was set at 1×10^{-4}. Catchability for the postharvest survey varied from 0.10 to 3 times the preharvest survey catchability but was constant over years. To ensure that the model requirements for contrast in exploitation rates between years were met, in this type of simulation u was set to 0.2 for year 1 and 0.6 for year 2. The simulation parameters were as follows:

(1) population size prior to removals:

$$N_{11} = N_{21} = 1,000,000 \text{ animals}$$

(2) catchability:

$$q_1f = 1 \times 10^{-4}$$
$$q_2f = (1 \times 10^{-5}, 3 \times 10^{-5}, 5 \times 10^{-5}, 7 \times 10^{-5}, 9 \times 10^{-5}, 1 \times 10^{-4}, 1.1 \times 10^{-4}, 1.3 \times 10^{-4}, 1.5 \times 10^{-4}, 1.7 \times 10^{-4}, 2 \times 10^{-4}, 3 \times 10^{-4})$$

(3) survey effort: $f = 1$

(4) exploitation rate:

$$u_1 = 0.2$$
$$u_2 = 0.6$$

(5) removals:

$$R_1 = u_1 \cdot N_{11} = 200{,}000$$
$$R_2 = u_2 \cdot N_{21} = 600{,}000$$

(6) number of years of data: $n = 2$.

Survey data were simulated 10,000 times for each of the twelve catchability scenarios.

Additional years of data.—In the third type of simulation, the performance of the 2qIR model was evaluated for improvement as more years of data were analyzed together and compared with the performance of the annual and 1qIR models. Model estimates were made with all three models over a range of contrasts between pre- and postsurvey catchability coefficients as described above for the second type of simulation, except that the range of catchability variation in the second survey was restricted to one-quarter to two times that of the first survey. The exploitation rate for years 3 to n ($n \geq 3$) was assumed to be moderate ($u = 0.3$). Estimates for multiple-year models were made using the "nlminb" function in S-PLUS. The simulation parameters were as follows:

(1) population size prior to removals:

$$N_{11} = N_{21} = \cdots = N_{n1} = 1{,}000{,}000 \text{ animals}$$

(2) catchability:

$$q_1f = 1 \times 10^{-4}$$
$$q_2f = (2.5 \times 10^{-5}, 5 \times 10^{-5}, 7.5 \times 10^{-5}, 1 \times 10^{-4}, 1.25 \times 10^{-4}, 1.5 \times 10^{-4}, 2 \times 10^{-4})$$

(3) survey effort: $f = 1$

(4) exploitation rate:

$$u_1 = 0.2$$
$$u_2 = 0.6$$
$$u_3 = \cdots = u_n = 0.3, \quad n \geq 3$$

(5) removals:

$$R_1 = u_1 \cdot N_{11} = 200{,}000$$
$$R_2 = u_2 \cdot N_{21} = 600{,}000$$
$$R_i = u_i \cdot N_{i1} = \cdots = 0.3 \times 1{,}000{,}000 = 300{,}000, \quad i \geq 3, \quad 3 \leq i \leq n$$

(6) number of years of data: $n = 2$, 3, or 5.

Survey data were simulated 1,000 times for each of the seven catchability scenarios.

Application to Southern Rock Lobster

Study Site.—Survey and fishery removal data were collected from 1996 to 2000 for a population of southern rock lobster at a study site near Port Davey, Tasmania (43.39°S, 145.88°E), by the Tasmanian Aquaculture and Fisheries Institute in Taroona (Table 2). Both survey and fishery data were collected at similar locations (the 7E2 block of stock assessment area 8) and from similar depths (40–80 m).

Three fishery-independent surveys were performed each year, commercial harvest and effort data being documented for the times between surveys. Surveys were performed during the first week of the fishing season (the "spring" survey [early to mid-November]), in midseason (late February to mid-March), and again near the end of the season (the "fall" survey [mid-July to mid-August]). Model estimates were made with each of the IR models (annual, 1qIR, and 2qIR) using two sets of data. Both data sets incorporated the spring survey but differed as to which survey index was used as the second required survey. One data set incorporated the midseason survey index as the second survey, while the other incorporated the fall survey. The relative performances of the models were compared for each of the data sets used.

Model Choice.—Model performance with these data was evaluated by two methods, a likelihood ratio test and a diagnostic plot. A likelihood ratio test was applied to decide which of the IR models (annual, 1qIR, or 2qIR) was most parsimonious (Miller and Miller 1999) for the southern rock lobster data.

To test whether the annual model or the 1qIR model is more appropriate, both equations (1) and (2) are fitted to the data. For n years of data ($n > 1$),

$$\theta = \frac{\Lambda_{\text{restricted}}}{\Lambda_{\text{unrestricted}}} = \frac{\Lambda_{\text{1qIR}}}{\Lambda_{\text{ann}}}, \tag{10}$$

and $-2 \cdot \log_e \theta$ approximates χ^2_r with $r = n - 1$ degrees of freedom. The null hypothesis is that there is no difference in catchability coefficients across years. If the test fails to reject the hypothesis, the restricted model (1qIR) is appropriate. If the hypothesis is rejected, there is evidence that the catchability coefficients differ among years and thus that the unrestricted annual model is more appropriate.

TABLE 2.—Survey catch rates, survey effort, and commercial removals for a southern rock lobster fishery in Tasmania. The fishing season runs from mid or late October to August or September, so the "fished year" spans two calendar years. Scientific surveys were conducted in the first week of commercial harvest ("preharvest"), in midseason ("midyear" [March]), and again in the last weeks of the season ("postharvest"). Commercial removal data are for time periods between the surveys.

Time of survey or harvest interval	Fished year				
	1996–1997	1997–1998	1998–1999	2000–2001	2001–2002
Catch rate (lobsters/trap haul)					
Preharvest	1.22	2.33	3.50	1.40	1.26
Midyear	0.35	0.61	1.44	0.44	1.06
Postharvest	0.03	0.22	0.36	0.20	0.16
Survey effort (trap hauls)					
Preharvest	49	94	100	100	50
Midyear	99	100	100	100	50
Postharvest	98	150	100	100	50
Removals (kg)					
Start of season to midyear	37,708	30,031	60,894	22,281	19,137
Start of season to year-end	45,476	57,894	81,214	37,654	27,683

When applying a likelihood ratio test to determine whether the 2qIR model is more appropriate than the 1qIR model, 1qIR is again the more restricted model. The null hypothesis is that there is no difference between the pre- and postharvest catchability coefficients. The test statistic is $-2 \cdot \log_e(\Lambda_{1qIR}/\Lambda_{2qIR})$ but now with 1 degree of freedom. If the test fails to reject the hypothesis, 1qIR is again the more appropriate model.

Model choice was also evaluated with diagnostic plots. If a model is effective in estimating biomass, (1) biomass estimates should be strongly related (using R^2 values) to preharvest survey catch rates and (2) a regression of model abundance estimates on preseason survey indices should have an intercept close to the origin. Also, since the slopes of the regression lines calculated for the diagnostic plots estimate the reciprocal of the survey gear catchability ($1/q$), (3) estimates of q calculated from the slope of the regression line should also be consistent with estimates of q from model outputs.

Model Sensitivity to Error in Removals.—The method assumes that removals are known precisely. Since parameter estimates are scaled entirely by the magnitude of the removals, it was expected that any error in removals would bias abundance estimates proportionally. This was tested by adding 10% to each removal before fitting to the model and then comparing the resulting parameter estimates with those of the original model estimates made from the actual data.

TABLE 3.—Model comparison by the likelihood ratio test.

Models	Log_e likelihood objective value	Test statistic $(-2 \cdot \log_e \theta)$	df (r)	χ^2 critical value $(\alpha_r = 0.05)$
Spring and midseason data				
1qIR	−6.8843			
Annual	−6.6929	0.3828	4	9.488
2qIR	−6.8668	0.0350	1	3.841
Spring and fall data				
1qIR	−4.8178			
Annual	−4.7408	0.1540	4	9.488
2qIR	−4.7525	0.1306	1	3.841

Results

Simulation Results

Exploitation rate variation.—The 2qIR model results were accurate, precise, and almost always usable when there was moderate contrast in the range of exploitation rates among years (i.e., $|u_2 - u_1| \geq 0.3$), regardless of the degree of contrast between the catchability coefficients (Figure 1), relative to the results of the annual model. With moderate exploitation contrast and with only 2 years of data, the abundance estimates of the 2qIR model were always more accurate and precise than annual model estimates and were seldom (<5% of the time) unusable. The performance of the annual model, however, was very sensitive to changes in the survey catchability coefficient and model performance varied greatly among scenarios.

The median 2qIR model estimates were always more accurate than those of the annual model when the contrast between the exploitation rates of the 2 years was at least 0.2, regardless of the degree of contrast in catchability coefficients (Figure 1). The annual model estimates were more accurate only when the catchability coefficients were equal or the exploitation rate was extremely high.

The 2qIR model estimates were most variable when the simulated population was lightly ($u = 0.2$) or moderately ($u = 0.3$) exploited, but the range of the

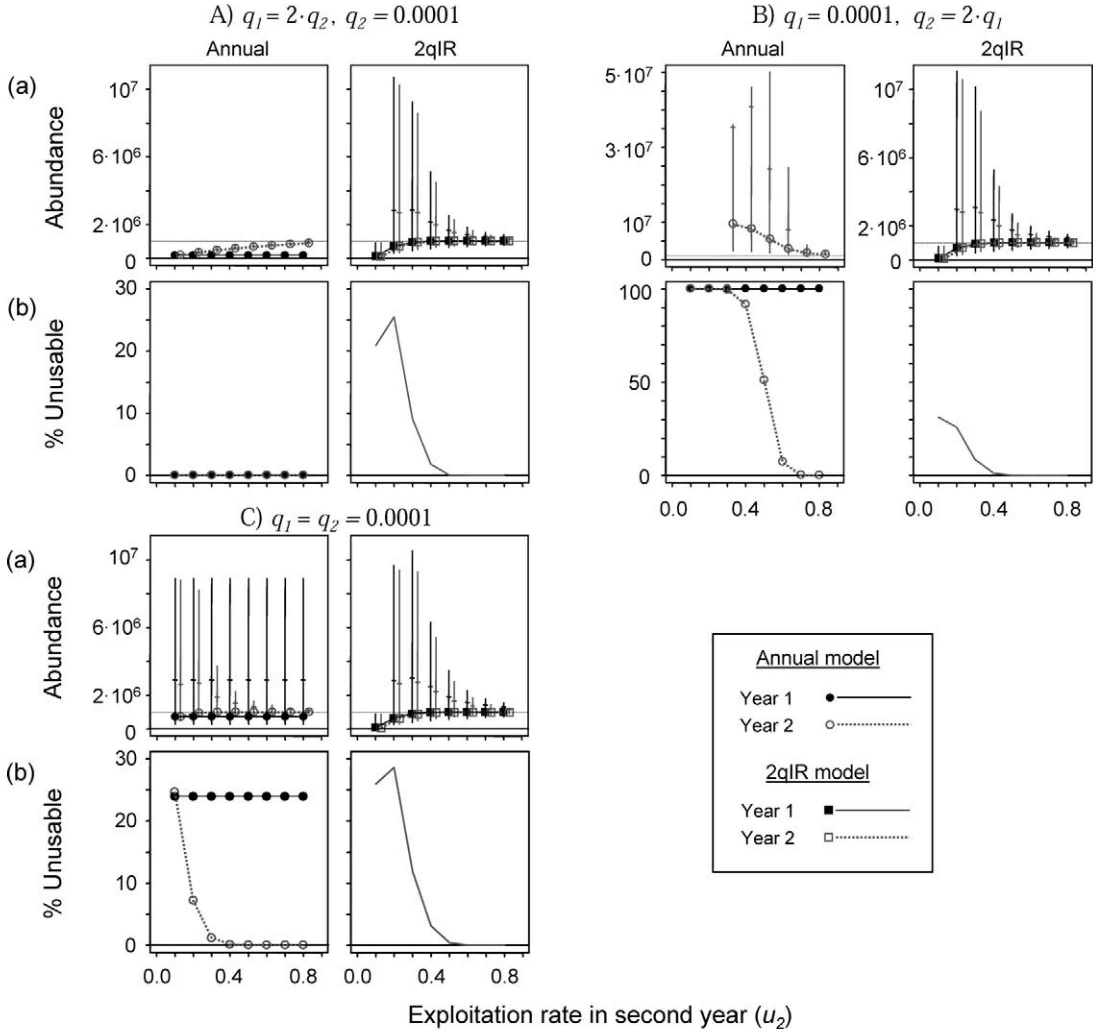

FIGURE 1.—Comparison of the performance of two index-removal models—an annual model and a multiple-year model that allows survey catchability to vary seasonally (2qIR)—for three sets of contrasting values for survey catchability. The symbols show the medians of the estimates. The exploitation rate was fixed at 0.1 for all simulations in year 1 but varied among simulations in year 2. Each vertical bar represents the central 95% of the usable estimates from 10,000 simulations of 2 years of survey data. Ten percent of the usable estimates were above the horizontal hash marks on the bars. Panel groups **(A–C)** differed in catchability between the pre- and postharvest surveys. Row **(a)** depicts abundance estimates. The true abundance is indicated by the horizontal line in each plot; the curves for estimates of years 1 and 2 are slightly offset horizontally so that both are visible. All abundance plots have the same scale except for the annual model plot of scenario (B). Row **(b)** depicts the percentage of unusable estimates for each model. Only one line is drawn for the 2qIR model because the years were estimated simultaneously and a failure for either or both years was counted as an unusable simulation.

central 95% of the 2qIR model estimates always included the true abundance when u was greater than 0.1. Though the 2qIR estimates were variable, the upper tails of the distribution of estimates were long, and when the upper 10% of the estimates were excluded the 2qIR estimates were always more precise than the corresponding estimates of the annual model (Figure 1).

The variability of the annual model estimates, however, differed greatly for the three different scenarios. The annual model estimates had virtually no variability when preharvest catchability was double that of postharvest catchability; however, the central 95% of the usable estimates never included the true abundance (Figure 1A). When the catchability of the second survey was double that of the first survey

(Figure 1B), the central 95% bands of the annual estimates were roughly an order of magnitude greater than those for the other scenarios and the bands never included the known abundance when the exploitation rate was at or below 60%. When catchability was equal for both surveys, the annual bands were wide at very low exploitation rates but narrow when at least 30% of the population was harvested in the second year (Figure 1C).

The estimates of the 2qIR model were usable more often (92% of the time) than those of the annual model (80% for all simulations of year 2) when all scenarios were combined. Most of the unusable 2qIR estimates (81%) were observed when the exploitation rate contrast was low (i.e., when the difference between exploitation rates was $\leq$10%). When the exploitation rate contrast was at least 20%, the percentage of the 2qIR estimates that were unusable never exceeded 12% for any set of simulations.

Catchability variation.—The 2qIR model worked well over a wide range of contrasts in catchability coefficients between the pre- and postharvest surveys, but the annual model was very sensitive to catchability change (Figure 2). The 2qIR model estimates were more accurate and precise over the entire range of catchability ratios (q_2/q_1) examined than were those of the annual model, and 2qIR model estimates were almost always usable. In contrast, the annual model estimates were more accurate, precise, and usable than 2qIR model estimates only when the catchability ratios were close to unity.

The 2qIR model produced more accurate estimates than the annual model when postharvest catchability differed from preharvest catchability by a factor of 0.3–3.0 (Figure 2). In a preliminary analysis, Ihde (2006) demonstrated that accurate estimates were produced by the 2qIR model for an even broader range of contrast between pre- and postharvest catchability coefficients (from factors of 0.3–10), but because the model estimates stabilized when more than a threefold increase was simulated, the simulation presented here was limited to the threefold increase between survey catchabilities seen in Figure 2. The 2qIR model performed better when postharvest catchability was greater than preharvest catchability than when the opposite was the case. But even the most extreme medians were only slightly below the true abundance. In the worst-case scenario, postharvest catchability was one-tenth that of preharvest catchability, but the median estimates still were within 9% of the true value. In all other cases, the median estimates were within 2% of the true abundance.

In contrast, the median estimates of the annual model were within 10% of the actual abundance only when there was no change (or nearly no change) in catchability between the pre- and postharvest surveys. When the exploitation rate was low ($u = 0.2$ [year 1]), the median annual model estimates differed from the true value by about 30% when the pre- and postharvest catchability coefficients differed. When the exploitation rate was high ($u = 0.6$ [year 2]), the annual model made accurate estimates (within 10% of known abundance) more often, but only if the difference in catchability coefficients was 10% or less.

The range of the central 95% of usable estimates of both models was characterized by upper bounds that were at least three times the magnitude of the lower bounds (Figure 2). Though the precision of the 2qIR model estimates varied somewhat over the range of catchability ratios examined, the central 95% of the estimates always contained the known abundance (Figure 2). The variability of the 2qIR estimates was greatest when preharvest catchability was greater than postharvest catchability. This trend was especially pronounced when the catchability ratio was less than 0.5. However, the variability of the 2qIR estimates was relatively constant when catchability ratios were greater than 1 (Figure 2b). Annual model estimates were more precise than 2qIR estimates when the q-ratios were less than 1. But when the catchability ratios were less than 0.7 or more than 1.3 the central 95% bands of the annual estimates never included the true abundance (Figure 2b).

The estimates of the 2qIR model were almost always usable; in contrast, slightly more than one-half of the annual model estimates were usable (Figure 2c). Of 120,000 simulations analyzed with the 2qIR model (catchability ratios from 0.1 to 3.0), only 1% of the estimates were unusable. Most of the unusable estimates of the 2qIR model (87%) were from the lowest q-ratios examined (0.1 and 0.3). Twice as many simulations were possible for the annual model because each year of data was analyzed separately. Of the 240,000 possible annual model estimates made, 24% were unusable. The year under high exploitation had more usable estimates (92%) than did the year under low exploitation (60%). However, no annual model estimates were feasible for catchability ratios of 2 or more when the exploitation rate was low (0.2) or, in a preliminary analysis, for ratios of 4 or more when the exploitation rate was high (0.6; Ihde 2006).

Additional years of data.—With 5 years of data, the 2qIR model estimates were more accurate and precise than those of the other IR models examined in this scenario; the 2qIR model estimates were almost always usable (Figure 3). Even with 5 years of data, however, the 1qIR model estimates were more accurate, precise,

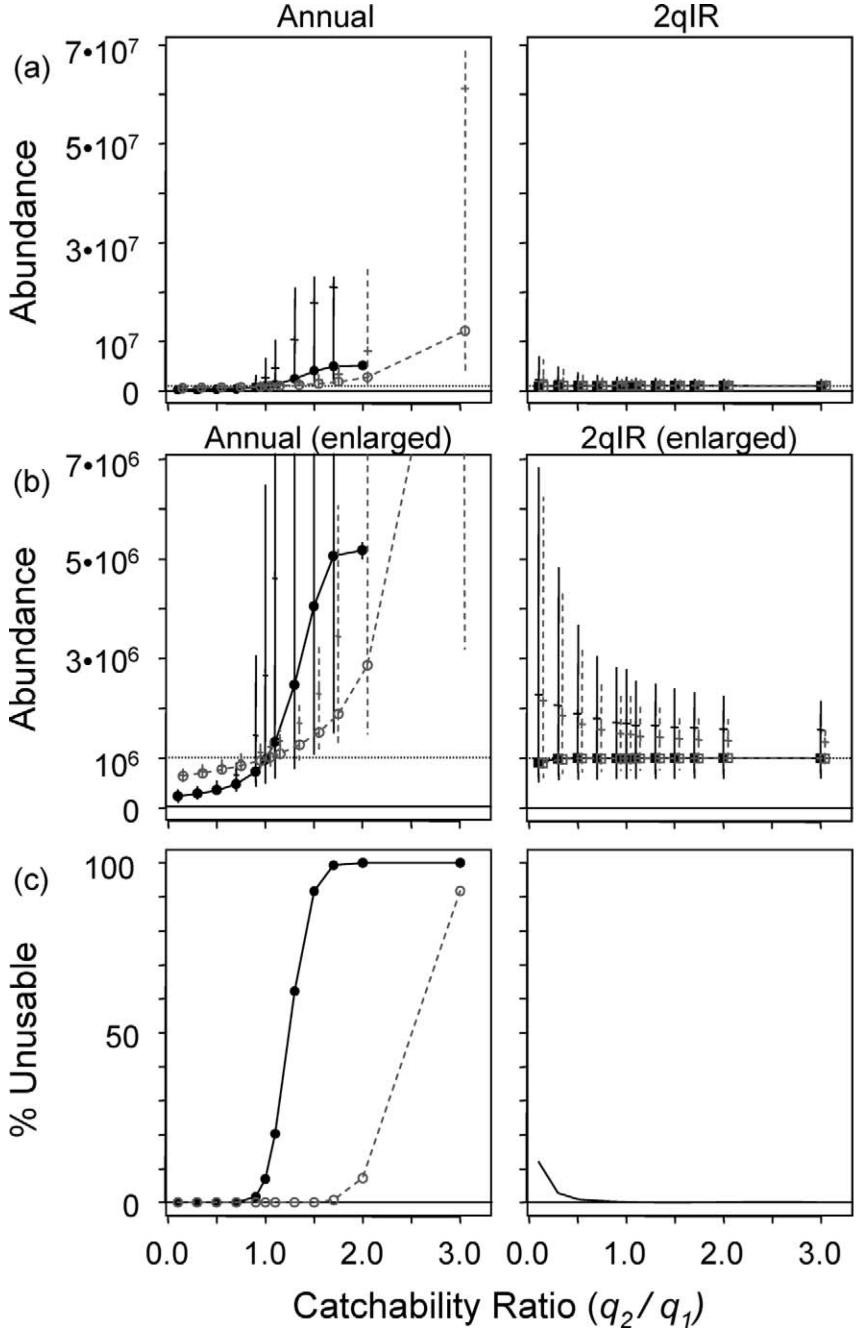

FIGURE 2.—Comparison of the performance of the annual and 2qIR models using 10,000 simulations of 2 years of data. Performance is compared over a range of seasonal change in the survey year catchability coefficient. Catchability in the first survey was 0.0001 in both years; catchability in the second survey was the same for both years in any one simulation but varied from 0.00001 to 0.0003 among scenarios. The panels in row **(b)** present the same results as those in row **(a)** but at a finer scale. In rows (a) and (b) the performance indicators are the median estimates (symbols) and the width of the intervals containing 95% of the usable estimates (vertical lines); in row **(c)** the performance indicator is the percentage of unusable simulations. Ten percent of the usable estimates were above the horizontal hash marks on the vertical lines. The exploitation rates were 20% in year 1 (solid lines, filled symbols) and 60% in year 2 (dashed lines, open symbols). The curves for years 1 and 2 are offset slightly so that both estimates are visible. The horizontal lines in rows (a) and (b) indicate the true abundance.

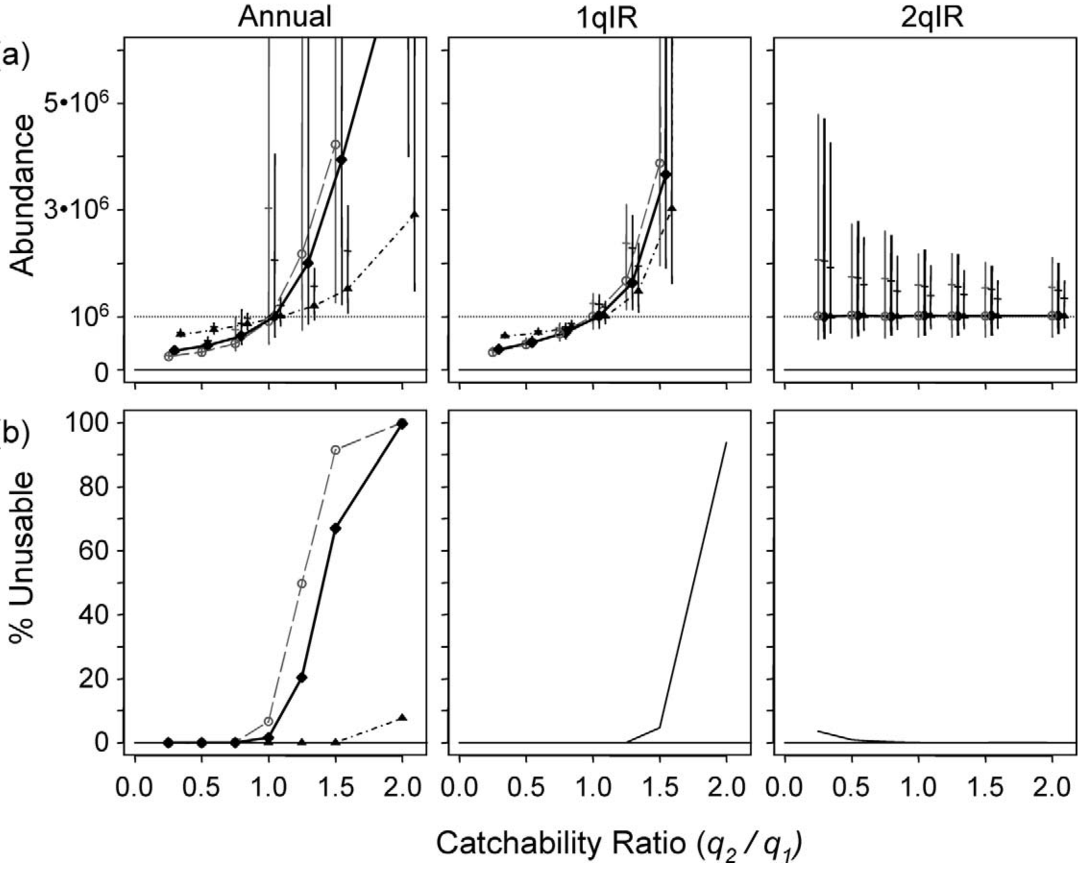

FIGURE 3.—Comparison of annual, 1qIR, and 2qIR model performance with 5 years of simulated data. Seven scenarios are shown that varied in terms of the catchability coefficient of the second survey. Each scenario was simulated 1,000 times. The curves for the different years are offset slightly so that all of them are visible. Row **(a)** depicts abundance estimates for each model. The median estimates for individual years are represented by circles (low-exploitation year [$u = 0.2$]; dashed line) or triangles (high-exploitation year [$u = 0.6$]; dash–dot line). The diamonds represent averages of the medians of the 3 years with moderate exploitation ($u = 0.3$; solid line). The vertical lines extending from the medians represent the central 95% of the usable model estimates. Ten percent of the usable estimates were above the horizontal hash marks on the vertical lines. Row **(b)** depicts the percentage of unusable simulations for each model.

and usable than the 2qIR model estimates only when the catchability ratio was close to unity.

The range of the central 95% of 2qIR estimates was wider when only 2 years of data were available than when 3 or 5 years were available (Figure 4). The improvement gained by using 5 years of data instead of 3 was marginal. The performance of the annual model shown in Figure 3 is representative of the performance of this model with any number of years of data because the estimate is made independently for each year of data. With 2 years of data, however, the 1qIR model estimates were more accurate and usable a greater proportion of the time than the 1qIR results shown in Figure 3, but overall the patterns of the estimates were similar (Figure 5).

Application to Southern Rock Lobster

Parameter estimates.—The 2qIR model predicted lower exploitation rates and catchability coefficients and considerably higher biomass than did the other models (Figure 6). All of the 2qIR model estimates appeared reasonable and were similar, regardless of which data set was fit to the model. However, the patterns of the exploitation estimates of the annual and 1qIR models differed considerably, depending on whether the midseason or fall survey data were used in the analyses. Moreover, the exploitation rate estimates of both the annual and 1qIR models were unreasonably high for estimates based on data that included the midseason surveys. Both the annual and

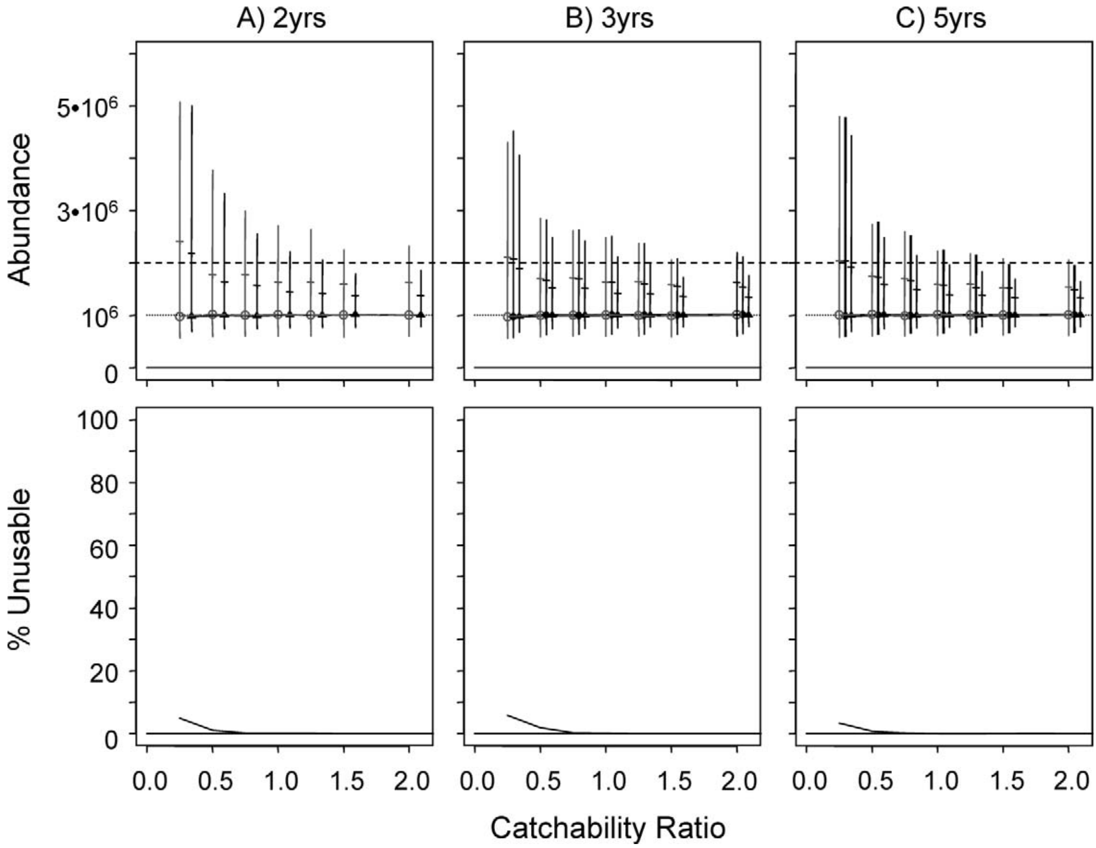

FIGURE 4.—Comparison of 2qIR model performance with **(A)** 2, **(B)** 3, and **(C)** 5 years of data. Model performance improves with a third year of data but is similar with 3 or 5 years. See Figure 3 for additional details. Horizontal dashed line is included to facilitate comparison among panels in top row.

1qIR models predicted that more than 100% of the population was harvested in 2 of the 5 years of data (Figure 6A). Though the 2qIR model estimates were also high at the beginning of the data set (80% when the midseason survey was used, 91% when the fall data were used), the 2qIR model estimates from both data sets predicted a steady decline in the exploitation rate during the next 4 years (Figure 6). The range of contrast in exploitation rates among years was approximately 0.4 when estimates of the 2qIR model were made with either data set. According to simulation results (Figure 1), this was more than enough contrast for the 2qIR model to work well. The 2qIR model almost always estimated lower catchability coefficients than the other models and predicted a more than 70% decrease in catchability between the spring and fall surveys. Correspondingly, the 2qIR model estimates of abundance were much higher than those predicted by the other models. In 1998, the 2qIR model abundance estimates from both data sets were about 50% higher than those of the annual and 1qIR models.

Model choice.—A likelihood ratio test found that the most parsimonious model was 1qIR regardless of which data set was analyzed (Table 3). Diagnostic plots and the occurrence of infeasible estimates (exploitation rate estimates >1.0) with the 1qIR model, however, suggest that the 2qIR model performed best (Figure 7). In addition, an examination of likelihood ratio test (LRT) statistics calculated from simulation results presented earlier (from Figure 3) show that the test was relatively insensitive to changes in the catchability coefficient between surveys (Figure 8), even when the performance of the 1qIR model was poor relative to that of the 2qIR model (Figure 3).

Diagnostic plots of the estimates made from the midseason data (Figure 7A) showed that although both the 1qIR and 2qIR model estimates of abundance had strong relationships with the preharvest survey catch rate (R^2 values were 0.88, and 0.92, respectively), the

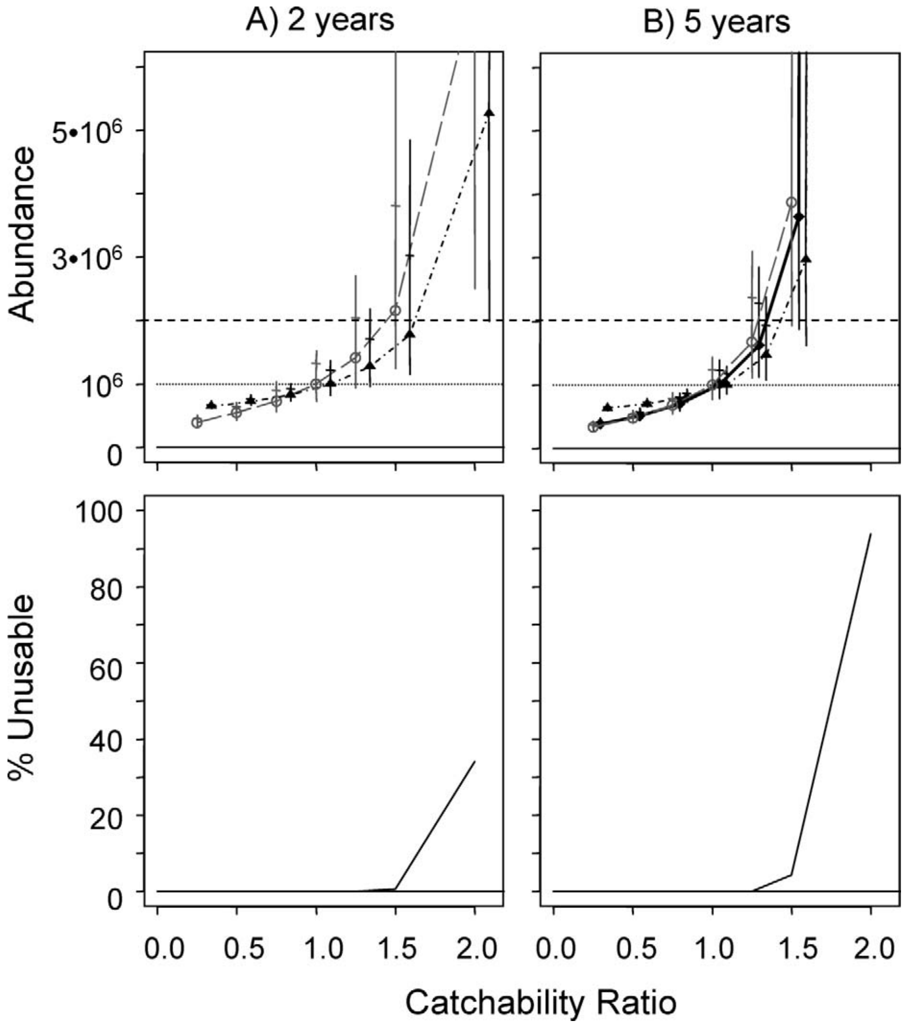

FIGURE 5.—Comparison of 1qIR model performance with **(A)** 2 years and **(B)** 5 years of data. The performance of the model deteriorates somewhat as more years of data are added. See Figure 3 for additional details. Horizontal dashed line is included in top row to facilitate comparison between panels.

annual model estimates were only weakly related to the preharvest survey catch rate ($R^2 = 0.06$). The intercepts for both the 1qIR (8,121) and 2qIR models (8,093) were similar and very close to the origin for the midseason data, but the annual model intercept was over 50,000 kg.

When fall survey data were fit to each of the models, all model estimates of abundance (Figure 7B) had strong relationships with the preharvest survey catch rate (R^2 values were 0.94, 0.95, and 1.00 for the annual, 1qIR, and 2qIR model estimates, respectively), but the 2qIR model demonstrated the strongest relationship. The intercepts were similar for the annual (11,679) and 1qIR models (10,415), but that of the 2qIR model (3,698) was the closest to the origin.

The equations for the regression lines were as follows:

$$\hat{N}_{\text{ann-mid}} = 10{,}099 \cdot I_1 + 50{,}403, \tag{11}$$

$$\hat{N}_{\text{1qIR-mid}} = 25{,}647 \cdot I_1 + 8{,}121, \tag{12}$$

and

$$\hat{N}_{\text{2qIR-mid}} = 38{,}584 \cdot I_1 + 8{,}093 \tag{13}$$

for the midyear data and

$$\hat{N}_{\text{ann-fall}} = 22{,}484 \cdot I_1 + 11{,}679, \tag{14}$$

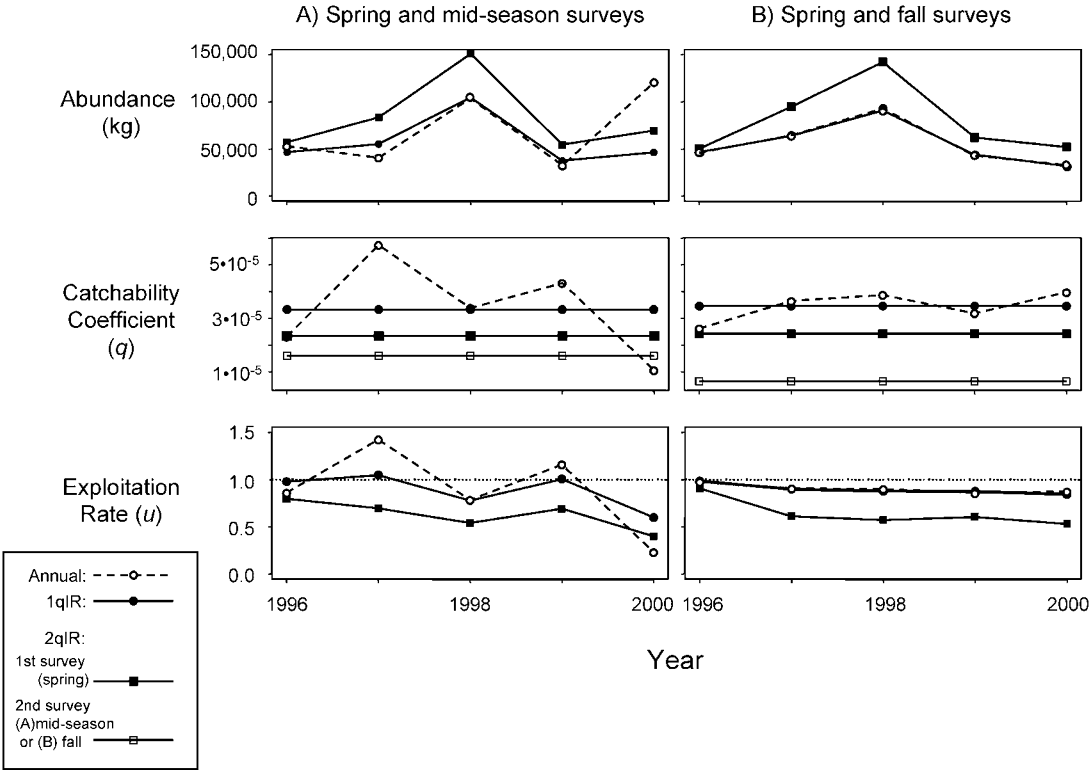

FIGURE 6.—Estimates of abundance, catchability coefficient, and exploitation rate for two sets of southern rock lobster fishery data. The dotted line in plots of exploitation rate, which indicate 100% exploitation, are included for reference.

$$\hat{N}_{\text{1qIR-fall}} = 23{,}410 \cdot I_1 + 10{,}415, \tag{15}$$

and

$$\hat{N}_{\text{2qIR-fall}} = 39{,}442 \cdot I_1 + 3{,}698 \tag{16}$$

for the fall data.

The slope of the regression line estimates the reciprocal of survey gear catchability ($1/q$) for the annual and 1qIR models. The estimates for q, then, using the midseason data were 9.9×10^{-5}, 3.9×10^{-5}, and 2.6×10^{-5} for the annual, 1qIR, and 2qIR models, respectively. When the fall data were used instead, the annual, 1qIR, and 2qIR model estimates of survey gear catchability were 4.4×10^{-5}, 4.3×10^{-5}, and 2.5×10^{-5}, respectively.

Model sensitivity to error in removals.—Error in removals resulted in a proportional, added error in abundance estimates that was similar for the annual, 1qIR, and 2qIR model estimates (10% in all cases for all models). The added error to catchability coefficients was negative and nearly proportional, but the error was slightly dampened for estimates of this parameter (–9.1% in all cases for all models). Estimates of the exploitation rate were not biased because these can be made without removal data, requiring only survey indices of the population abundance before and after the removals take place (Hoenig and Pollock 1998). The relevant expression for the exploitation estimate is

$$\hat{u} = \frac{R}{\hat{N}} = \frac{(I_1/f_1) - (I_2/f_2)}{I_1/f_1} = \frac{I_1 - I_2(f_1/f_2)}{I_1}. \tag{17}$$

Discussion

Model Evaluation by Simulation

The 2qIR model was more accurate and precise than the annual model, and the results were usable with almost all simulated data when the exploitation rate differed by at least 0.3 between two years (not necessarily consecutive), regardless of the contrast in catchability coefficient between the pre- and postharvest surveys (Figure 1). In contrast, the erratic performance of the annual model with the same data is a consequence of the violation of the annual model assumption of constant catchability. When this assumption is met (Figure 1C), the annual model performs relatively well. When catchability is lowered

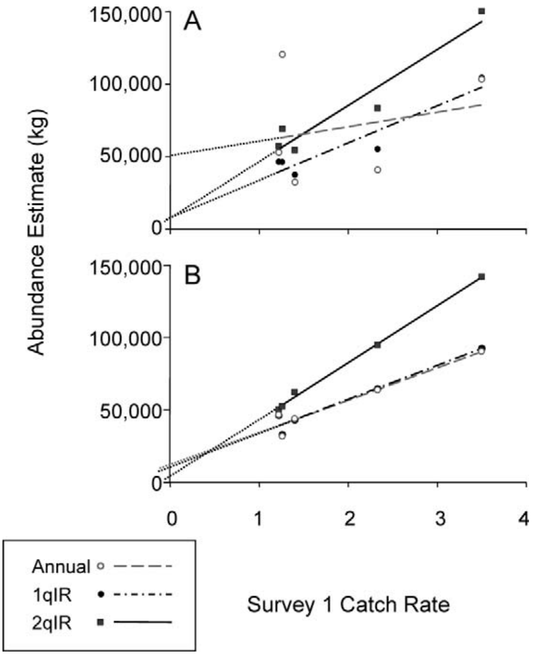

FIGURE 7.—Regressions of model abundance estimates on the preharvest survey catch rates of the corresponding years. The dotted lines indicate the *y*-intercepts for the three models. The regression curves shown in panel **(A)** were based on data collected in spring and midseason surveys, those in panel **(B)** on data collected in spring and fall surveys.

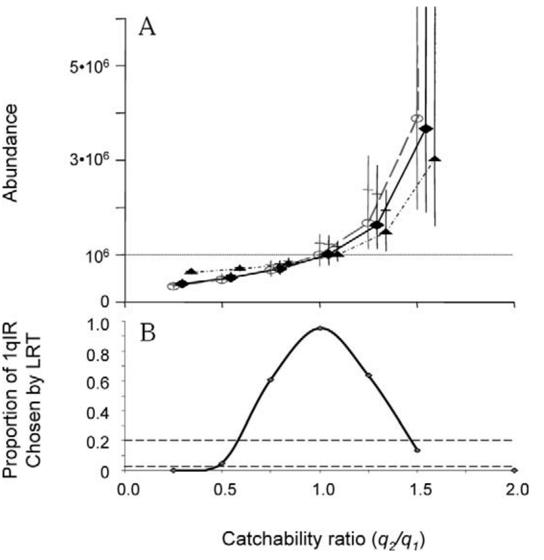

FIGURE 8.—**(A)** Performance of the 1qIR model (from row [a], column 2 of Figure 3) as a function of the catchability ratio and **(B)** the proportion of simulations for which the likelihood ratio test (LRT) failed to reject the 1qIR model. The median model estimate was only accurate when the model assumptions were met and the catchability ratio was 1.0 in this simulation. The LRT failed to select against the 1qIR model the majority of the time until there was a 50% change between the catchability coefficients of the two surveys. In (A) the horizontal line indicates the true abundance; in (B) horizontal lines indicate the proportions 0.20 and 0.025, at which an LRT determined that the 1qIR model was more parsimonious than the 2qIR model 40% and 5% of the time, respectively.

in the second survey, the model also appears to perform well (i.e., estimates are highly precise and 100% of the simulations are usable) because the lower catches observed due to lower catchability are incorrectly accounted for by the model as lower abundance (thus, the negative bias seen in Figure 1A). However, when survey catchability is lower in the first survey (Figure 1B), catches in the first survey may be smaller than or similar in magnitude to those of the second survey, a situation that results in a high percentage of unusable simulations. In this scenario, the annual model cannot produce estimates as accurate and precise as those of the 2qIR model until 70% or more of the population is harvested.

Because the 2qIR model estimates were usable a greater portion of the time, more 2qIR model estimates were made using problematic data (i.e., survey catches that were close in magnitude or simulations in which the catch of the second survey exceeded the catch of the first survey). As a result, the 2qIR model was somewhat disadvantaged when the accuracy and precision of model estimates were compared directly with those of the models that excluded more of the problematic data, and the percentage of unusable estimates was an important factor in assessing overall model performance. Except in the most extreme scenarios (e.g., the difference between exploitation rates between two years was <0.3 [Figure 1] or the catchability ratio was <0.3 [Figure 2]), the 2qIR model performed well in spite of the inclusion of these problematic data.

Though the performance of the 2qIR model was best when postharvest catchability was greater than preharvest catchability in the simulations presented here (Figures 2–4), the difference is thought to be due to simulation design rather than being a characteristic of model performance. Ihde (2006) demonstrated that the performance of IR models improves substantially with a higher *qf* product and thus higher survey catches. The value of *qf* in the preharvest surveys was constant for all simulations. Thus, the performance of the 2qIR model at higher postharvest catchabilities (and consequently at higher *qf* values) was improved. Similarly, the performance of the model was poorer at lower postharvest catchabilities (i,e., catchability ratios <1) owing to the lower *qf* values at these catchabilities, which also imply lower survey catches.

The performance indicators for the 2qIR model improved when a third year of data was added to the data set. However, model performance with 5 years of data was similar to that with 3 years. In contrast, the performance of the 1qIR model deteriorated somewhat when years were added (Figure 5). The 1qIR model is based on the assumption that catchability is constant. So, as more data that violate the assumptions of the model are added to the data set, the chances of getting problematic data (i.e., a higher catch rate in the second survey or one similar in magnitude to that of the first survey) increase.

The only instances when the annual or 1qIR models outperformed the 2qIR model were when catchability was constant (or nearly constant) between surveys and the exploitation rate was very high (60% or more). In such situations, use of the simpler models is appropriate and the 2qIR model suffers a penalty in variability for unnecessarily estimating an extra catchability parameter.

An assumption of the 2qIR model is that the year-to-year change in catchability coefficients is unimportant, but this is not always the case. Ihde et al. (2008) examined the effect of a strong temporal trend on the estimates of the 1qIR model. They found that with even an extreme trend in catchability (a geometric increase of 15% each year over 10 years of simulated data), the error in the median model estimates did not exceed 60%. It seems likely that similar trends in catchability would affect the estimates of the 2qIR model similarly, but this still requires investigation.

Application to Southern Rock Lobster

Though the likelihood ratio test suggests that the 1qIR model was the most parsimonious in this application, diagnostic plots and the patterns of the model estimates suggest that the 2qIR model performed best for this population of southern rock lobsters. Moreover, an examination of the relative insensitivity of the LRT in simulation (Figure 8) suggests that the LRT was not useful in determining the most parsimonious model in this application.

An examination of the LRT statistics calculated from the simulation results (from Figure 3) shows that the test was relatively insensitive to changes in the catchability coefficient between surveys (Figure 8), even though the 1qIR model was shown to perform poorly (i.e., the estimates never included the true abundance) with only a small change of catchability in simulation. Likelihood ratio test statistics were calculated from objective values resulting from 1qIR and 2qIR model runs for data from the third simulation, which considered model performance with additional years of data.

Diagnostic plots and the patterns of the model estimates suggest that the 2qIR model performed best for this population of southern rock lobsters. When the midseason data were used, diagnostic plots showed only a slight improvement in the performance of the 2qIR model over that of the 1qIR model (Figure 7A). However, an examination of the patterns of parameter estimates demonstrated a distinct improvement in the performance of the 2qIR model over that of the other models (Figure 6). The unreasonable exploitation estimates of the annual and 1qIR models suggest that both of these models performed poorly with the midseason data. When the fall data were incorporated instead of midseason data (Figure 7B), the intercept of the 2qIR model was closer to the origin than those of the other models. And, though all model abundance estimates had strong relationships with the survey catch rate, the 2qIR model estimates were directly related ($R^2 = 1.0$) to it. Additionally, about 5% more of the variation in the 2qIR abundance estimates was explained by the catch rate than was explained for the other models. The estimate of the survey gear catchability coefficient for the 2qIR model, which is predicted by the reciprocal of the slope of the regression line, corresponded closely to model estimates for preharvest catchability regardless of which data set was analyzed, as one would expect if the model were performing well. However, the slope estimates of survey catchability for the annual model were nearly three times greater than that of the mean annual model estimate made from the midseason data and more than 25% greater than the corresponding estimate from the fall data. Slope estimates made with the 1qIR model were more consistent with model estimates than were those of the annual model, but these estimates were still 18% higher for the midseason data and 23% higher for the fall data. The inconsistencies of the annual and 1qIR model estimates of survey catchability suggest poorer model performance relative to the 2qIR model.

The closed-population assumption for southern rock lobsters in southern Tasmania appears reasonable because the animals here move little (Gardner et al. 2003), they recruit once a year (by moulting) prior to the onset of the fishing season (Frusher 1997), and their natural mortality is estimated to be quite low (0.10–0.12/year; Punt and Kennedy 1997; Frusher and Hoenig 2003).

The contrast in exploitation rates required by the 2qIR model seems likely to be met for this data set. Independent estimates of exploitation rates are not available for this population of southern rock lobsters, so the minimum exploitation contrast in this data set cannot be verified. However, a substantial management change occurred during the 5 years of data analyzed here (1996–2000) that seems likely to have resulted in a corresponding change in the exploitation rate. In 1998, an individual transferable quota (ITQ) system was implemented in Tasmania to reduce the catch of southern rock lobsters. After only 2 years of the ITQ system, a substantial (29%) decrease in effort was documented (Ford 2001). Thus, it seems reasonable to expect that this population experienced at least a 0.2 difference in exploitation rates (the minimum contrast suggested by the simulations; Figure 1) between at least 2 of the 5 years of data analyzed here.

Recent work indicates that use of either the annual or 1qIR model may be inappropriate for the data set that included the fall surveys. Ziegler et al. (2003) predicted that the relative catchability of southern rock lobsters in this region decreases markedly after the midseason survey is conducted and that catchability at the time of the fall survey is distinctly lower than that of the spring and midseason surveys. The 2qIR model results presented here appear to support the conclusions of Ziegler et al. (2003). The 2qIR model estimated that catchability declined by more than 70% between the spring survey in November and the fall survey in August. If this predicted change in catchability is real, it is important to take it into account. Our simulation results suggest that when the catchability ratio (q_2/q_1) is less than 1 the single-q models will substantially underestimate abundance (Figures 2, 3). When the IR models were applied to the data set that included the fall data, the 2qIR model predicted a catchability ratio of 0.27. Accordingly, abundance estimated with the 2qIR model was 44% higher and exploitation was 28% lower, on average, than corresponding estimates of the annual and 1qIR models because those models could not accommodate the catchability change. Previous studies of different populations of this species have also documented the importance of accounting for seasonal catchability change (Ziegler et al. 2002; Frusher and Hoenig 2003). Thus, it appears likely that the 2qIR model is the most appropriate model to use in the stock assessment of this population, regardless of which data set is used.

If there is any doubt as to which model to apply, our results suggest that the 2qIR model will probably give the most accurate estimate. Though the estimates may suffer slightly in precision if a simpler model is truly appropriate, the 2qIR model estimates will not have the serious biases that result from applying either the annual or the 1qIR model when their assumption of constant catchability is not met.

Acknowledgments

This work was supported by a National Marine Fisheries Service—Sea Grant Joint Graduate Fellowship in Population Dynamics, the Virginia Marine Resources Commission, and the Virginia Institute of Marine Science. We gratefully acknowledge the Tasmanian Aquaculture and Fisheries Institute for use of the southern rock lobster data presented here. We also thank L. Jacobson, J. Musick, M. Prager, J. Shields, and three anonymous reviewers, whose thoughtful suggestions have improved this manuscript. This is VIMS contribution 2874.

References

Chen, C.-L., J. M. Hoenig, E. G. Dawe, C. Brownie, and K. H. Pollock. 1998a. New developments in change-in-ratio and index-removal methods, with application to snow crab (*Chionoecetes opilio*). Canadian Special Publication of Fisheries and Aquatic Sciences 125:49–61.

Chen, C.-L., K. H. Pollock, and J. M. Hoenig. 1998b. Combining change-in-ratio, index-removal, and removal models for estimating population size. Biometrics 54:815–827.

Dawe, E. G., J. M. Hoenig, and X. Xu. 1993. Change-in-ratio and index-removal methods for population assessment and their application to snow crab (*Chionoecetes opilio*). Canadian Journal of Fisheries and Aquatic Sciences 50:1467–1476.

Eberhardt, L. L. 1982. Calibrating an index using removal data. Journal of Wildlife Management 46:734–740.

Ford, W. 2001. Restructuring the Tasmanian rock-lobster fishery—the effect of two years of management under individual transferable quotas. Marine and Freshwater Research 52:1641–1648.

Frusher, S. 1997. Stock assessment report, rock lobster. Department of Primary Industry and Fisheries, Marine Resources Division, Hobart, Australia.

Frusher, S., C. Gardner, C. Mackinnon, and M. Haddon. 2003. Tasmanian rock lobster fishery, 2001/2002. Tasmanian Aquaculture and Fisheries Institute, University of Tasmania, Hobart, Australia.

Frusher, S. D., and J. M. Hoenig. 2003. Recent developments in estimating fishing and natural mortality and tag reporting rate of lobsters using multi-year tagging models. Fisheries Research 65:379–390.

Frusher, S. D., R. B. Kennedy, and I. D. Gibson. 1998. Preliminary estimates of exploitation rates in the Tasmanian rock lobster (*Jasus edwardsii*) fishery using the change-in-ratio and index-removal techniques with tag–recapture data. Canadian Special Publication of Fisheries and Aquatic Sciences 125:63–71.

Gardner, C., S. Frusher, M. Haddon, and C. Buxton. 2008. Movements of the southern rock lobster *Jasus edwardsii* in Tasmania, Australia. Bulletin of Marine Science 73:653–671.

Hoenig, J. M., and K. H. Pollock. 1998. Index-removal methods. Pages 342–346 *in* S. Kotz, C. B. Read, and D. L. Banks, editors. Encyclopedia of statistical sciences, volume 2. Wiley, New York.

Ihde, T. F. 2006. Development of generalized index-removal models, with particular attention to catchability issues. Doctoral dissertation. College of William and Mary, Williamsburg, Virginia.

Ihde, T. F., J. M. Hoenig, and S. D. Frusher. 2008. Evaluation of a multiyear index-removal abundance estimator, with application to a Tasmanian rock lobster fishery. Fisheries Research 89:26–36.

MathSoft. 2000. S-Plus professional release 3, professional edition for Windows. MathSoft, Seattle.

Miller, I., and M. Miller. 1999. John E. Freund's mathematical statistics, 6th edition. Prentice-Hall, Toronto.

Paloheimo, J. E. 1963. Estimation of catchabilities and population sizes of lobsters. Journal of the Fisheries Research Board of Canada 20:59–88.

Petrides, G. A. 1949. Viewpoints on the analysis of open season sex and age ratios. Transactions of the North American Wildlife Conference 14:391–410.

Punt, A. E., and R. B. Kennedy. 1997. Population modelling of Tasmanian rock lobster, *Jasus edwardsii*, resources. Marine and Freshwater Research 48:967–980.

Roseberry, J. L., and A. Woolfe. 1991. A comparative evaluation of techniques for analyzing white-tailed deer harvest. Wildlife Monographs 117.

Routledge, R. 1989. The removal method for estimating natural populations: incorporating auxiliary information. Biometrics 45:111–121.

Ziegler, P. E., S. Frusher, C. R. Johnson, and C. Gardner. 2002. Catchability of the southern rock lobster *Jasus edwardsii*. I. Effects of sex, season, and catch history. Marine and Freshwater Research 53:1143–1148.

Ziegler, P. E., S. D. Frusher, and C. R. Johnson. 2003. Space–time variation in catchability of southern rock lobster *Jasus edwardsii* in Tasmania explained by environmental, physiological, and density-dependent processes. Fisheries Research 61:107–123.