Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Fisheries Research 70 (2004) 239–250

FISHERIES RESEARCH

www.elsevier.com/locate/fishres

Examining the importance of consistency in multi-vessel trawl survey design based on the U.S. west coast groundfish bottom trawl survey

Andrew B. Cooper[a,*], Andrew A. Rosenberg[a], Gunnar Stefánsson[b], Marc Mangel[c]

[a] *Department of Natural Resources, College of Life Sciences and Agriculture, University of New Hampshire, Durham, NH 03824, USA*
[b] *Marine Research Institute, University of Iceland, Reykjavik, Iceland*
[c] *Department of Applied Mathematics and Statistics, Center for Stock Assessment Research, Jack Baskin School of Engineering, University of California, Santa Cruz, CA 95064, USA*

Abstract

Generalized linear mixed-effects models can be used to combine bottom trawl data from multiple vessels, each with a different fishing power, into a single time series of relative abundance. However, how important might it be to have a consistent set of vessels and vessel characteristics from year to year given we can model differences in fishing power among vessels? We demonstrate how changes in the suite of fishing vessels performing the survey can affect the results of the data analysis using sablefish catches in the U.S. west coast groundfish bottom trawl survey from 1998 to 2000. The results do not indicate that one must have a consistent set of vessels over time to provide useful data, but rather that there is benefit to consistency even when the survey data are analyzed using advanced statistical models. Further research should be undertaken to quantify these benefits specifically to aid in contracting and bidding of survey vessels.

Keywords: Generalized linear mixed-effects models; Power; Sablefish; Type I error

---

1. Introduction

Bottom trawl surveys have been conducted on the continental shelf and upper slope off the west coast of the U.S. (Washington, Oregon and California) since 1977. These surveys provide the primary source of abundance and trend information for most stock assessments conducted on west coast groundfish. During 1998–2001, trawl survey effort expanded to include three independent trawl surveys. One survey covers the shelf (30–200 fm) and two cover the slope (100–700 fm). The three surveys differ in vessel size (chartered 75 ft fishing vessel to 225 ft FRV *Miller Freeman*), season (mid-summer to -autumn), net size and footrope, and tow speed (2–3 knots). The shelf survey uses larger chartered fishing vessels and has been conducted triennially since 1977. There have been substantial improvements in net mensuration and some

---

* Corresponding author. Tel.: +1 603 862 4254;
fax: +1 603 862 4976.
E-mail address: andrew.cooper@unh.edu (A.B. Cooper).

doi:10.1016/j.fishres.2004.08.006

changes in trawl gear, towing protocols, and stratification schemes during the 24 years of the shelf survey. One of the two slope surveys utilizes the FRV *Miller Freeman* in the autumn. It started in 1988 but did not achieve annual coast wide coverage until 1997. The other slope survey uses smaller chartered fishing vessels, started in 1998, and is conducted roughly coincident with the shelf survey. Lauth (2001) and Turk et al. (2001) provide additional details of the two slope surveys.

To be most useful, resource assessment surveys must take advantage of advancing technology and knowledge to be as accurate and precise as possible, but also must maintain comparability with historical survey data. Available vessels, staffing and funding will not allow all three surveys for west coast groundfish to continue into the future. All three surveys were conducted in summer-fall 2001, thus providing a good, and probably final, opportunity for comparison and calibration. A goal of the future survey strategy is an annual survey that covers the shelf and slope in an integrated design. In addition, the data from the earlier surveys must be reanalyzed to assure that they are as comparable with the results from the present surveys as possible.

Helser et al. (2004) use generalized linear mixed-effects models (GLMMs) to combine the existing bottom trawl data from multiple survey vessels into time series of relative abundance. The GLMM assumes vessels are independent from one another, they may have different fishing power, and the differences in fishing power among vessels can be modeled as a normal random variable with mean zero. In statistical terminology, the estimates of the fishing power differences (i.e. vessel effects) are called random effects, and are derived, in large part, from the temporal and spatial overlap of sampling by vessels, with increasing overlap leading to improved precision of the random effect estimates. The GLMM is a powerful approach that allows researchers to account for changes in vessel participation over time and improve the precision of the resulting estimates of abundance.

The GLMM approach, however, also raises some challenging questions for multi-vessel survey design: (1) what is probability of detecting a vessel effect when there is not one – Type I error, (2) what is the probability of detecting a vessel effect when there is one – power, and (3) how important is it to have a consistent set of vessels and vessel characteristics from year to year given we can model differences in fishing power? The third question can be examined through three related questions in which the method used to analyze the survey data is that of Helser et al. (2004).

1. How sensitive are the estimates of the biomass index and the change in the biomass index over time to the exclusion of individual vessels from the data set?
2. How do the estimates of the biomass index and the percentage change in the biomass index from 1998 to 2000 vary as we treat existing vessels as being new vessels in the last year of the survey, thus requiring us to estimate additional parameters for the vessel effect?
3. How do the estimates of the biomass index and the percentage change in the biomass index from 1998 to 2000 vary when we treat all vessels as being new each year? This final question presents a more extreme case of the previous question and allows us to better examine the importance of vessel participation over time.

If survey catch is truly a random process, and the differences in fishing power among vessels can be fully accounted for by the model, then the estimate of the biomass index and the percentage change in the biomass index should be insensitive to each of these manipulations. While the percentage change in a biomass index over time is a crude measure of population change compared to that provided by more formal stock assessment methods, it nonetheless provides a summary measure of the behavior of the indices over time, as well as insight into how our perception of a stock's recovery or decline may change based solely on the methods used to derive the biomass index. The last two questions are addressed using simulated data because this allows us to move beyond the specifics of a particular dataset and hence examine the full range of realizations of a catch process where the data meet the specific assumptions of the analysis method (e.g. random catch and normally-distributed vessel effects).

Answers to these questions will become more important as the U.S. National Marine Fisheries Service, NMFS, and other agencies increase the use of commercial vessels for survey work, and they raise additional questions regarding the optimal length of contracts, whether premiums for past participation should be included in contract bids, and the like.

2. Data and general methods

We based the analyses on the sablefish (*Anoplopoma fimbria*) data from 1998 to 2000 from bottom trawl surveys conducted by the Northwest Fisheries Science Center (NWFSC) and Alaska Fisheries Science Center (AFSC) slope surveys (Helser et al., 2004). Strata were defined as in Helser et al. (2004) with Stratum 1 being the most northern shallow stratum, Stratum 5 being the most southern shallow stratum, Stratum 10 being the most northern deep stratum, and Stratum 50 being the most southern deep stratum. The mean and variance of the survey catch by stratum and year are listed in Table 1.

For non-zero tows, we assume the catch within a stratum can be modeled as a gamma random variable, i.e. a mean–variance relationship with the variance proportional to the mean squared. This assumption can be validated by regressing the natural logarithm of the variance of the catch in each stratum on the natural logarithm of the corresponding mean catch (Fig. 1). This regression leads to the equation:

$$\ln(\text{variance}) = -0.09 + 2.16\ \ln(\text{mean}) \tag{1}$$

which transforms to:

$$\text{variance} = 0.91\,\text{mean}^{2.16} \tag{2}$$

The exponent in Eq. (2), is significantly different from zero ($t_{28} = 6.17$, $P < 0.0001$), but it is not significantly different from 2 ($t_{28} = 0.47$, $P = 0.64$). The intercept of Eq. (1) is not significantly different from zero ($t_{28} = -0.09$, $P = 0.92$), which translates to the coefficient of proportionality not being significantly different from one. The variance of catch within a stratum in the NMFS data is therefore approximately proportional to the mean catch within a stratum squared. Occasionally, as described below, we treat the simulated data as being normally distributed, but maintain a mean–variance relationship.

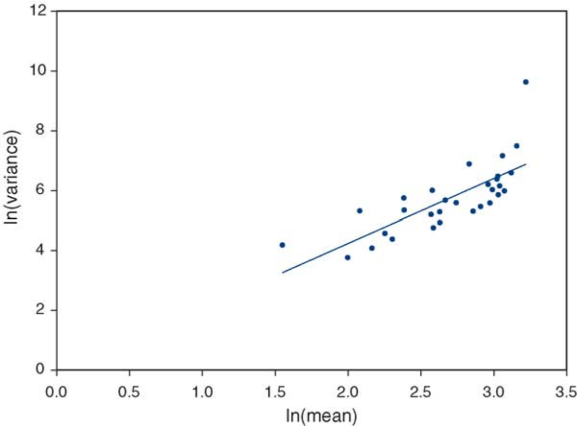

Fig. 1. Regression of ln(variance) on ln(mean) of the catch within a stratum (199–2000).

Table 1
Mean (and variance) of catch by stratum and year

Stratum	Year		
	1998	1999	2000
1	8.0 (205)	8.7 (60)	13.9 (140)
2	4.7 (66)	25.0 (15.284)	13.2 (408)
3	10.9 (211)	10.8 (315)	13.3 (117)
4	9.5 (97)	15.5 (269)	23.5 (1.801)
5	7.4 (43)	10.0 (81)	21.3 (1.293)
10	19.5 (266)	18.3 (236)	20.9 (472)
20	14.4 (293)	19.3 (497)	20.6 (651)
30	13.9 (198)	20.5 (594)	20.7 (350)
40	17.4 (203)	21.6 (399)	22.6 (733)
50	13.1 (182)	17.0 (986)	19.9 (416)

Two types of vessel distributions were used throughout the analyses. The number of observations made by each vessel in each stratum each year was taken directly from the NMFS data for 1998–2000 (Table 2) for one set of analyses. Vessel 1 is the only vessel to participate in all years, vessels 3, 6, and the FRV *Miller Freeman* participate in two of the three years, and all other vessels participate in only a single year (Table 2). For the other set of analyses, nine vessels were distributed in a balanced design so there were an equal number of observations for each vessel in each stratum each year. Regardless of the distribution of vessels among strata and years, the average catch in each stratum each year equals the average catch across vessels for the actual data for sablefish (Table 1).

We considered a range of variances in catch and in the vessel effects when generating the simulated data. The observed catch, $R_{i,j,k}$, for vessel i in stratum j and year k was:

$$R_{i,j,k} = C_{j,k} + V_i \tag{3}$$

where $C_{j,k}$ is the random catch for stratum j in year k, and V_i is the effect for vessel i. The value of $C_{j,k}$ was drawn from either a normal distribution or a gamma

Table 2
The number of observations for each vessel by stratum and year

Vessel	Stratum										Total
	1	2	3	4	5	10	20	30	40	50	
1998											
1	2	10	9	10	7	3	16	9	7	6	79
2	4	11	7	13	4	5	11	4	10	5	74
3	2	9	4	12	4	3	12	7	8	7	68
4	3	12	9	11	2	5	18	9	7	3	79
1999											
1	5	15	8	9	5	5	12	11	9	6	85
3	5	12	4	11	6	5	13	9	12	8	85
5	1	13	6	8	6	3	11	10	10	8	76
6	2	9	9	9	6	1	9	10	11	8	74
Miller Freeman	10	19	10	20	13	19	34	21	29	26	201
2000											
1	5	14	9	10	8	5	12	11	11	9	94
6	2	16	6	8	8	2	13	11	11	4	81
7	5	11	6	11	4	5	10	6	12	4	74
8	5	12	8	8	4	5	16	10	8	2	78
Miller Freeman	10	20	10	20	11	19	43	21	28	2	184

distribution with mean $\mu_{j,k}$ equal to the mean catch in the NMFS data for each strata × year combination and variance equal to the square of the mean catch in the NMFS data multiplied by a factor, a, that we varied over a range. For the gamma distribution, the mean of the distribution is shape/rate, the variance is shape/rate^2, and the skewness is $\sqrt[2]{(\text{shape})}$.

$$C_{j,k}\ N(\mu_{j,k}, a(\mu_{j,k})^2) \tag{4}$$

$$C_{j,k} = \text{gamma}\left(\text{rate} = \frac{1}{a\mu_{j,k}},\ \text{shape} = \frac{1}{a}\right) \tag{5}$$

The vessel effect for vessel i, also known as the random effect, was drawn from a normal distribution with mean zero and standard deviation, s, that we varied over a range.

$$V_i \sim N(0, s^2) \tag{6}$$

The vessel effect for each vessel was held constant across all strata and years. Adding annual variability to the vessel effects may increase the realism of the simulations (e.g. Helser et al., 2004), but would have added yet another layer of complexity, and is therefore left for future research.

All simulations were performed using S-PLUS 2000 (Mathsoft, Seattle, Wash.) and the generalized linear mixed-effect extension (GLME) to the non-linear mixed-effects model software within S-PLUS 2000 (NLME, Pinheiro and Bates, 2000). The GLME extension implements the methods in Breslow and Clayton (1993).

3. What is the probability of detecting a vessel effect when there is not one?

3.1. Methods

The number of observations by vessel, stratum, and year is unbalanced (Table 2). This is due in large part to the different sizes of the strata. We therefore wanted to see how this alone might be affecting our ability to detect a vessel effect. However, for a given stratum–year combination, the number of observations taken by each vessel differs, and not all vessels are present in each year (Table 2). Unfortunately, S-PLUS is unable to directly test for the statistical significance of the random effect terms in a GLMM compared to a fixed-effects-only model, when the random effect is modeled simply as a temporally constant vessel effect. The only way to come close to such a test is to examine the confidence intervals for the variance–covariance estimates from the GLMM model and see if the variance estimates get ‘too close’ to zero, for some arbitrary definition of ‘too close’ (J. Pinheiro, *Novartis Pharmaceuticals*, personal communication). We used a different approach in which we assumed that catches were normally distributed and used the S-PLUS linear mixed-effects (LME) code, which can be used to test for the significance of the random effects terms. We fit a linear mixed-effects model with year, stratum, and year × stratum interactions as fixed-effects to the simulated data (described below) and also fit a generalized least squares model with the same fixed-effects. We then compared these two models using likelihood ratio tests to determine whether the inclusion of a random vessel effect in the linear mixed model was significant ($P < 0.05$). We did not take the heterogeneous variance by stratum into account.

This analysis required three scenarios to study the possible impacts of: (1) the patterns in catch by stratum, (2) the distribution of vessels across strata, and (3) the combined effect of catch patterns and vessel distribution.

For Scenario 1, we distributed vessels in a balanced design with each vessel taking five samples from each stratum × year combination. This resulted in 450 samples per year, as compared to an average of 452 samples per year actually taken from 1998 to 2000. We drew catches from a normal distribution with mean u and variance u^2, with u equal to the mean catch in the NMFS data for each stratum × year combination. We used a proportionality coefficient of 1 for the mean–variance relationship as a simplification of the results of the log-linear mean–variance regression model (Eq. (2)). We repeated the analysis 1000 times. For Scenario 2, we distributed vessels as in the NMFS data (Table 2) and drew all catches from a normal distribution with mean u and variance u^2 and varied u from 10 to 100 in steps of 10. We repeated the analysis 100 times for each value of u. For Scenario 3, we distributed vessels as in Table 2 and drew catches from a normal distribution with mean u and variance u^2, with u equal to the mean catch in the NMFS data for each stratum × year combination. We repeated this analysis 1000 times.

3.2. Results

The inclusion of a random vessel effect in the model was significant in only 1.6% of the 1000 replicates for Scenario 1, i.e. with vessels distributed in a balanced design, and each vessel taking five samples from each stratum × year combination and no vessel effects.

The inclusion of a random vessel effect in the model was significant in 2.7% of the 1000 replicates for Scenario 2, i.e. with vessels distributed as in Table 2 and catches drawn from a normal distribution with mean u and variance u^2, with u varying from 10 to 100 in steps of 10. The vessel effect was significant in between 1 and 5% of the replicates when the results were broken down by the levels of u considered.

The inclusion of a random vessel effect in the model was significant in only 1.8% of the 1000 replicates for Scenario 3, i.e. with vessels distributed as in Table 2 and catches drawn from a normal distribution with mean u and variance u^2, with u equal to the mean catch in the NMFS data for each stratum × year combination.

Based on these analyses, the distribution of vessels and the spatio-temporal patterns in the catch do not appear to affect the probability of committing a Type I error (i.e. concluding that there are significant vessel effects where none exists).

4. What is the probability of detecting a vessel effect given there is one?

4.1. Methods

The bulk of the existing literature on sample size and power calculations for mixed-effects models focuses on the fixed-effects terms, and not on the power of detecting random effects, and is limited to traditional clinical trials study designs (e.g. Snijders and Bosker, 1993; Diggle et al., 1994; Raudenbush and Liu, 2000, 2001). We used a simulation approach to address this question because the programs and/or methods associated with the above articles could not meet our needs; the use of simulation to analyze the power of complex mixed-effects study designs is not new (Mok, 1995; Stoker and Bowers, 2002) and the use of simulation gave us the flexibility to tailor the analysis to our particular problem.

This analysis also required three scenarios; (1) vessels distributed in a balanced design with each vessel taking five samples from each stratum × year combination, (2) vessels distributed in a balanced design with each vessel taking 15 samples from each stratum × year combination, and (3) vessels and numbers of samples

Table 3
The percentage of replicates for which the random vessel effect term was significant ($P < 0.05$) for a balanced vessel design with each vessel taking five samples from each stratum × year combination

Catch variance multiplier, *a*	Vessel effect standard deviation, *s*				
	0.01	0.1	1	10	100
0.5	100	100	100	100	100
1.0	100	100	100	100	100
1.5	98	97	98	100	100
2.0	92	91	90	100	100
2.5	82	79	80	100	100
3.0	70	68	74	100	100
3.5	64	63	65	100	100
4.0	50	51	57	100	100
4.5	45	47	49	99	100
5.0	38	43	43	99	100

This is also the probability of detecting a random vessel effect given there is one.

from each vessel in each stratum × year combination as in Table 2.

The catch for each vessel was simulated as described in Eqs. (3), (4) and (6), with *a* varying from 0.5 to 5 in steps of 0.5, and *s* ranging from 0.01 to 100 in multiples of 10. We fitted a linear mixed-effects model with year, stratum, and year × stratum interactions as fixed-effects and a generalized least squares model with the same fixed-effects to these simulated data and again ignored the heterogeneous variance issue. We compared these two models using likelihood ratio tests to determine whether including a random vessel effect in the linear mixed model was significant ($P < 0.05$). We repeated the analysis 500 times for each combination of *a* and *s* for each scenario. This resulted in 25,000 replicates per scenario (10 levels of catch variance, five levels of vessel effect standard deviation, and 500 replicates per combination).

4.2. Results

We were able to detect even very small vessel effects in all scenarios as long as the ratio of the variance in the catch to the square of the mean catch was approximately one (Tables 3–5). When there are 15 samples per vessel in each stratum × year combination, we could detect vessel effects even when the variance in the catch was many times larger than the square of the mean catch (Table 4). This is not surprising given that this scenario is essentially comparing 450 ‘paired’ observations of the vessels. Given the actual mean–variance relationship in the NMFS data, we are quite likely to identify a vessel effect if one exists.

Table 4
The percentage of replicates for which the random vessel effect term was significant ($P < 0.05$) for a balanced vessel design with each vessel taking 15 samples from each stratum × year combination

Catch variance multiplier, *a*	Vessel effect standard deviation, *s*				
	0.01	0.1	1	10	100
0.5	100	100	100	100	100
1.0	100	100	100	100	100
1.5	100	100	100	100	100
2.0	100	100	100	100	100
2.5	100	99	99	100	100
3.0	99	100	99	100	100
3.5	96	97	96	100	100
4.0	95	94	94	100	100
4.5	90	91	88	100	100
5.0	87	86	87	100	100

This is also the probability of detecting a random vessel effect given there is one.

Table 5
The percentage of replicates for which the random vessel effect term was significant ($P < 0.05$) with vessels and number of samples for each vessel in each stratum × year combination as in Table 2

Catch variance multiplier, *a*	Vessel effect standard deviation, *s*				
	0.01	0.1	1	10	100
0.5	100	100	100	100	100
1.0	99	100	98	100	100
1.5	90	89	89	100	100
2.0	75	73	76	100	100
2.5	61	59	61	100	100
3.0	43	48	50	99	100
3.5	39	42	41	99	100
4.0	28	31	38	99	100
4.5	25	26	32	97	100
5.0	23	23	28	97	100

This is also the probability of detecting a random vessel effect given there is one.

5. How sensitive are the estimates of the biomass index and change in the biomass index over time to the exclusion of individual vessels from the NMFS data set?

5.1. Methods

We used a simplified version of the approach of Helser et al. (2004) for these calculations. The ap-

Table 6
Estimated biomass indices, percentage difference in the biomass indices compared to having all vessels in the survey, change in the biomass index from 1998 to 2000, and the slope from the regression line: % change from 1998 on year, where year = 1, 2, 3 for 1998, 1999, and 2000 respectively

	Estimated biomass index			Percent change vs. all vessels			1998–2000 Change		Regression line	
	1998	1999	2000	1998	1999	2000	Absolute	Percent	Slope	SE
All vessels	56	85	97				41	74	0.23	0.05
Without vessel 1	61	81	92	9	−5	−5	31	52	0.16	0.03
Without vessel 2	49	85	96	−12	−1	0	47	96	0.31	0.06
Without vessel 3	57	81	95	3	−5	−2	37	65	0.20	0.04
Without vessel 4	51	84	96	−8	−1	−1	45	87	0.28	0.06
Without vessel 5	58	90	97	4	5	1	39	68	0.22	0.05
Without vessel 6	56	89	99	1	4	3	43	77	0.25	0.05
Without vessel 7	56	87	100	1	1	4	44	79	0.25	0.05
Without vessel 8	58	87	103	5	2	7	45	77	0.24	0.05
Without *Miller Freeman*	58	83	89	5	−3	−8	31	53	0.17	0.04

The regression line was fit without an intercept, and the slope can be interpreted as the expected percentage change in the biomass index each year.

proach for estimating the biomass indices was therefore: (a) estimate the probability of zero catches for each year × stratum combination using mixed logistic regression with fixed-effects for year and stratum and a random vessel effect, (b) estimate the average catch rate for each year × stratum combination using a generalized linear mixed-effects model with gamma errors and fixed-effects for year, stratum, and year × stratum interaction and a random vessel effect, and (c) estimate a biomass index for each year as a function of the probabilities of zero catches, the catch rates, and the areas of the strata. The catch was assumed to be normally distributed as in the previous two questions to allow us to test statistically for the significance of the random effects term. We assume the gamma distribution in this and following sections because it better represents the data likely to be observed during actual surveys. We applied this approach to the following 10 scenarios: (1) use data from all vessels – baseline, (2–9) use data from all but one of the eight industry vessels – vessels 1–8, and (10) use data from all the vessels except the FRV *Miller Freeman*.

For each scenario, we estimated the biomass index for each year, the percentage change in the biomass index from 1998 to 2000, and the slope of a regression of the percentage change in the biomass index relative to 1998 on year. The slope of this regression estimates the percentage change in the biomass index per year relative to 1998. The variances of the indices and the change in the indices were not computed and could not be compared.

5.2. Results

The estimates of the biomass indices, the absolute change in the biomass index, and the percentage change in the biomass index each year are sensitive to which vessels are included in the analysis (Table 6). The change in the biomass index is much more sensitive than the biomass indices themselves. The removal of a vessel can affect the estimates in all years even if that vessel participated in the survey for only 1 or 2 years. This is because the random vessel effects are assumed to normally distributed with mean zero, so the removal of a vessel will alter the values of the random effects for all vessels, which will, in turn, alter all the biomass indices.

6. How do the estimates of the biomass index and percentage change in the biomass index from 1998 to 2000 vary as we treat existing vessels as being new vessels in the last year of the survey?

6.1. Methods

This question examines the effect of new, but similar, vessels entering in the final year of the survey. We used two sets of scenarios to explore this question. In the first set, we distributed vessels in a balanced design with each vessel taking five samples from each stratum × year combination. In the second set, we distributed vessels and the number of samples from each vessel in each stratum × year combination as in Table 2.

In each scenario, the observed catch for a given vessel in a given stratum × year equaled the sum of a random catch plus a vessel effect (Eq. (3)). For the first scenario, with balanced vessels, the random catch was drawn from a gamma distribution with mean equal to the mean catch in the NMFS data for each stratum × year combination and variance equal to the square of the mean catch in the NMFS data multiplied by a, which took the values of 2/9, 2/7, 2/5, 2/3, and 2 (Eq. (5)). We chose these values for a to explore a realistic range of variances and for coding simplicity. For the second scenario, with vessels distributed as in Table 2, the random catch was drawn from a gamma distribution with mean equal to the mean catch in the NMFS data for each stratum × year combination and variance equal to the square of the mean catch in the NMFS data multiplied by a, which took the values 2/10, 2/9, 2/8, 2/7,. . ., 2/3, 1, and 2 (Eq. (5)). We again chose values of a to explore a range of variances and for coding simplicity, but the decreased computer time for each analysis in the second scenario allowed for a finer resolution for a. For both scenarios, because generalized linear mixed-effects models assume the random effects are normally distributed even though the data can take on a variety of distributional families, we drew the vessel effect for each vessel (held constant over strata and time) from a normal distribution with mean zero and standard deviation, s, which ranged from 0.01 to 10 in multiples of 10 (Eq. (6)).

We altered the vessel name in the simulated data for the final year to treat existing vessels as 'new' vessels. As such, the 'new' vessel had the same vessel effect as the existing vessel, but the analysis treated the vessel as being new and unrelated to the existing vessel that it replaced. Every time we treat a vessel as new, it adds to the number of parameters that needs to be estimated.

With the total catches generated, we fit a series of generalized linear mixed-effects models with year, stratum, and year × stratum interactions as fixed-effects and vessel as a random effect to the simulated data. For the first scenario, with the balanced vessel design, we fit GLMMs, which assumed 0–9 new vessels in the final year to each realization of the simulated data. For the second scenario, with vessels distributed as in Table 2, we fit GLMMs, which assumed no change in the vessels and treated the FRV *Miller Freeman* as new in the final year. We chose to look at the effect of the FRV *Miller Freeman* because, as will typically be the case when using both research and commercial vessels, the FRV *Miller Freeman* obtained many more observations in each year than its commercial counterparts (Table 2), and, as such, the effect of its loss is of special concern to managers.

We estimated the biomass index each year after we fit the GLMMs. We then compared the estimated biomass index and the estimated percentage change in the biomass index from 1998 to 2000 from the models with no new vessels to those with one or more new vessels in the final year. We repeated the analysis 100 times for each scenario and for each combination of a and s. This resulted in 2000 replicates for the first scenario (five levels of catch variance, four levels of vessel effect standard deviation, 100 replicates per combination) and 4000 replicates for the second scenario (10 levels of catch variance, four levels of vessel effect standard deviation, 100 replicates per combination).

6.2. Results

We present our findings in terms of the median outcome of the simulations rather than the mean outcome because the median is less sensitive to rare, extreme events. For the first scenario, the median effect of new

Table 7
Median percentage difference between the biomass index for 2000 when some of the vessels are treated as new and when none is treated as new

Number of new vessels	Vessel effect standard deviation, s			
	0.01	0.1	1	10
1	−3.71	−3.94	−4.00	−4.59
2	−5.12	−5.14	−5.24	−6.91
3	−6.23	−6.22	−6.01	−6.09
4	−5.78	−6.02	−5.59	−5.65
5	−4.95	−5.27	−4.65	−4.88
6	−3.11	−3.62	−2.87	−5.95
7	−5.49	−5.94	−4.93	−7.08
8	−1.71	−2.02	−1.37	−3.74
9	−3.54	−3.56	−3.32	−4.80

Results are for the balanced vessel distribution and a = 2/3.

Table 8
Median percentage differences between when one vessel is treated as new and when the balanced vessel distribution is unaltered

Catch variance multiplier, *a*	Vessel effect standard deviation, *s*							
	2000 Biomass index				Change in biomass 1998–2000			
	0.01	0.1	1	10	0.01	0.1	1	10
2	−1.85	−1.98	−2.16	−3.46	7.75	6.39	3.42	8.88
2/3	−3.71	−3.94	−4.00	−4.59	26.32	25.83	24.68	26.90
2/5	−4.51	−4.42	−4.56	−1.21	38.65	27.86	35.03	37.01
2/7	−4.68	−4.60	−4.97	−0.33	42.14	41.32	38.4	52.02
2/9	−4.91	−5.07	−4.70	−3.67	50.43	45.85	47.32	50.16

Results are shown for the biomass index for 2000 and change in estimated biomass from 1998 to 2000.

vessels in the final year was for the biomass index for the final year to be less than when there are no new vessels in the final year (e.g. Table 7). For a given number of new vessels in the final year, decreasing the variance in the catch, *a*, leads to an increase in the degree of underestimation of the biomass index in the final year (e.g. Table 8). Increasing the standard deviation of the vessel effects, *s* (i.e. increasing the differences among vessels) also increases the level of underestimation, but there were exceptions to this trend.

The effect of new vessels on the estimated change in the biomass index from 1998 to 2000 was much greater than the effect on each year's biomass index alone, because treating vessels as new in the final year will affect both the 1998 and 2000 biomass estimates (Table 8). The magnitude of the effects in Table 8 is not unexpected given those obtained by removing a vessel from the data analysis (Table 6). For example, removing Vessel 1 increases the 1998 biomass index by 9% but decreases the 2000 biomass index by 5%, resulting in the estimated percentage change from 1998 to 2000 falling from 74 to 52%, a 29.7% decrease in the estimate of percentage change. When examining the effect of new vessels, we calculate the percentage of over- or underestimation as:

$$\frac{\text{median estimated \% change with new vessels---}\atop\text{median estimated \% change with no new vessels}}{\text{median estimated \% change with no new vessels}} \times 100 \qquad (7)$$

The percentage underestimation of the estimated percentage change in the index can be >100% because the median estimated percentage change can be negative (a decrease in the index of abundance) or positive (an increase in the index of abundance). For example, an

Table 9
Median percentage differences between when the *Miller Freeman* is treated as new and when the vessel distribution in Table 2 is unaltered

Catch variance multiplier, *a*	Vessel effect standard deviation, *s*							
	2000 Biomass index				Change in biomass 1998–2000			
	0.01	0.1	1	10	0.01	0.1	1	10
2	4.55	4.65	5.42	7.41	19.17	14.49	21.91	33.43
1	6.93	6.58	5.79	8.38	32.16	26.77	28.06	47.74
2/3	8.16	7.05	7.57	8.76	36.18	33.05	37.08	41.41
2/4	7.71	7.84	7.26	7.55	37.04	38.43	36.62	37.61
2/5	8.56	8.15	7.84	6.82	42.75	42.08	39.19	42.16
2/6	8.35	8.73	8.06	8.39	40.31	43.70	43.58	41.89
2/7	8.40	8.35	8.47	9.42	41.44	42.85	39.54	55.01
2/8	8.39	8.26	8.70	9.59	44.28	45.57	45.80	59.97
2/9	8.81	8.64	8.79	8.34	47.91	44.71	45.01	38.30
2/10	9.08	8.57	8.81	7.95	48.66	44.47	48.38	37.60

Results are shown for the biomass index for 2000 and change in estimated biomass from 1998 to 2000.

estimated change of −2% with new vessels compared to an estimated change of 1% with no new vessels results in a −300% difference or an underestimation of 300%. This is unrelated to the standard deviation of the vessel effect, but does increase as the variance in the catch increases (Table 8).

We obtain results similar to those with the balanced design when we distribute vessels as in Table 2 and only the *Miller Freeman* is treated as new, except when the vessel effect standard deviation is large (Table 9). As for the balanced design, the change is greater when the estimated percentage change in biomass rather than the biomass estimates in any one year are considered (Table 9), and the magnitude of change in the estimates is not dissimilar to that seen in the experiments in which one vessel was dropped (Table 6). For example, the percentage underestimation resulting from removing the FRV *Miller Freeman* from the raw data (Table 6) is 28.4% ((53–74)/74), which compares well to the effect of treating the FRV *Miller Freeman* as new in the final year with a catch variance multiplier of 1 (Table 9).

7. How do the estimates of the biomass index and percentage change in the biomass index from 1998 to 2000 vary when we treat all vessels as being new each year?

7.1. Methods

We used two sets of scenarios to explore this question. In the first set of scenarios, we distributed vessels in a balanced design with each vessel taking five samples from each stratum × year combination. In the second set of scenarios, we distributed vessels and numbers of samples from each vessel in each stratum × year combination as in Table 2. For each set of scenarios we generated the observed catch as before (Eqs. (3), (5) and (6)) and used the same values for *a* and *s*.

We again fit a series of generalized linear mixed-effects models with year, stratum, and year × stratum interactions as fixed-effects to the simulated data. For the first scenario, with the balanced vessel design, we fit GLMMs, which assumed either all vessels occurred in all years or that each vessel was new each year. Further refinement could be made by looking at various combinations of 2-year participation or treating only some fraction of the vessels as being new at different points in time, but that is left for future research. For the second scenario, with vessels distributed as in Table 2, we fit GLMMs, which assumed either vessels were distributed as in Table 2 or each vessel was new each year.

We then estimated the biomass index for each year and the percentage change in the biomass index from 1998 to 2000, and computed differences. A 5000 replicates were conducted for the first scenario (five levels of catch variance, four levels of vessel effect standard deviation, 250 replicates per combination) and 10,000 for the second scenario (10 levels of catch variance, four levels of vessel effect standard deviation, 250 replicates per combination).

7.2. Results

With balanced vessels, the median effect of having new vessels each year was to underestimate the biomass index for 2000 (Table 10), with the extent of underestimation increasing with *a*. There was no obvious pattern with increasing values of *s*. The results for the percent-

Table 10
Median percentage differences between when all vessels are treated as new each year and when the balanced vessel distribution is unaltered

Catch variance multiplier, *a*	Vessel effect standard deviation, *s*							
	2000 Biomass index				Change in biomass 1998–2000			
	0.01	0.1	1	10	0.01	0.1	1	10
2	−4.20	−4.34	−3.45	−3.01	−16.16	−17.17	−14.62	55.77
2/3	−5.58	−5.24	−5.47	−4.64	−29.20	−24.55	−29.92	−1.72
2/5	−6.00	−5.86	−5.82	−7.53	−42.23	−37.85	−35.37	−37.24
2/7	−6.02	−6.23	−5.92	−7.47	−48.22	−40.32	−45.13	−86.98
2/9	−6.29	−6.32	−6.19	−9.89	−46.74	−46.72	−44.56	−137.60

Results are shown for the biomass index for 2000 and change in estimated biomass from 1998 to 2000.

Table 11
Median percentage differences between when all vessels are treated as new each year and when the vessel distribution in Table 2 is unaltered

Catch variance multiplier, *a*	Vessel effect standard deviation, *s*							
	2000 Biomass index				Change in biomass 1998–2000			
	0.01	0.1	1	10	0.01	0.1	1	10
2	−0.20	−0.35	0.00	2.22	26.59	20.10	21.67	64.14
1	−0.17	−0.22	0.34	4.53	31.00	30.92	28.75	129.06
2/3	0.00	−0.16	−0.30	5.88	37.56	35.96	38.41	144.49
2/4	−0.29	−0.64	0.02	6.89	35.22	40.16	40.69	176.16
2/5	0.21	−0.52	−0.15	5.88	45.05	38.10	44.01	168.65
2/6	0.13	−0.12	0.24	7.58	46.87	40.01	45.51	234.96
2/7	0.11	0.18	0.12	8.60	47.54	42.96	47.51	210.21
2/8	0.06	−0.08	0.07	4.96	46.39	43.72	45.77	182.49
2/9	0.32	0.13	0.28	4.86	47.61	42.31	47.28	144.60
2/10	0.20	0.23	0.31	8.45	48.36	49.56	52.01	293.79

Results are shown for the biomass index for 2000 and change in estimated biomass from 1998 to 2000.

age change in biomass from 1998 to 2000 (Table 10) are similar, except when variance in catch and standard deviation in vessel effect were high.

With vessels distributed as in Table 2, the median effect of having new vessels each year was to either underestimate or overestimate the biomass index for 2000 (Table 11). However, treating all vessels as new each year led to a median effect of overestimating the percentage change in biomass from 1998 to 2000 (Table 11). As was the case when the FRV *Miller Freeman* was assumed to be new in the final year (Table 9), overestimation increased with decreasing variance in the catch. The overestimation changed little with increasing standard deviation of the vessel effect, except when *s* was 10.

8. Discussion

It is clear that the suite of vessels used for the survey and how often they are replaced can have major impacts on the accuracy of the survey indices and their trends. The statistical power to detect differences among the vessels is quite high. If there is a vessel effect, we are likely to detect it, and we are not likely to detect a vessel effect when one does not exist, assuming a vessel effect that is constant across years. Adding inter-annual variability in the vessel effects in the simulations may decrease our ability to detect such effects, but the impact of such variability on the estimates of the biomass index and the percentage change in biomass index is uncertain and left for future research. The fact that Helser et al. (2004) were able to estimate such inter-annual variability using GLMMs suggests that questions regarding detection and sensitivity are worthy of investigation.

Interestingly, it is not the standard deviation, and hence size, of the vessel effects that seems to matter, it is simply that there are differences among vessels, and the estimation method is very sensitive to them. This can be seen clearly from the experiments in which one or another vessel is left out of the analysis in which the estimated percentage change in abundance varies from 52 to 96%. Even when smoothed with a regression analysis by fitting a straight line to the data, the percentage change over a 3-year period varies from 16 to 31% (Table 6). The estimate of this percentage change altered by 30–50% in the experiments that treat a vessel as new in the final year. This shows that the estimates of resource recovery could become highly variable simply due to the change in vessels used if vessels were to drop in and out of the survey frequently. The impact of changing vessels to some degree will be diminished if biologically linked stock assessment models, rather than simply the percentage change in the index of biomass, is used, but examination of this is beyond the scope of the present study.

Though the simulation studies are ostensibly for ‘vessels’, this really is a proxy for any substantive change in the survey instrument, be it the captain, the gear, or the vessel itself. This is not at all surprising given examples from other parts of the USA, such as the northeast U.S., where concerns have been raised about the impacts of gear changes over time and the need for calibration on dedicated research survey vessels (NEFSC (Northeast Fisheries Science Center), 2002). Such calibration is not feasible for commercial vessels, or would be prohibitively expensive or difficult to schedule, given the vessels are primarily dedicated to commercial fishing are only available for a restricted period of time.

The GLMM approach is an innovative and appropriate means to try to reconcile vessel differences. However, the results presented here indicate that this should not be taken to mean that the effects of vessel differences can be ‘modeled away’ or that the specific choice of vessels will not have major impacts. This, in turn, implies that having as consistent a set of vessels and vessel characteristics as possible from year to year is essential to monitoring recovering stocks. Always having the same vessels year after year is an unrealistic goal. However, the results of this study suggest that vessels willing to make longer-term commitments to participation in the survey are of more value than vessels with shorter-term commitments. What would an agency gain by moving from a 2-year contract to a 5-year contract? How much more would an agency be willing to pay for such contracts? How would this compare to the costs of a dedicated research vessel? Future research should examine the information value associated with long-term survey contracts, the potential cost-benefit trade-offs of these contracts, and how inter-annual variability in vessel performance may alter this assessment.

Acknowledgements

Thanks go to Elizabeth Clarke, Richard Methot, Thomas Helser, André Punt, Teresa Turk, and other members of the NMFS groundfish assessment program for initiating this project and guiding us through the intricacies of the data. We would also like to thank two anonymous reviewers for their helpful comments. Funding for this project was provided by the Northwest Fishery Science Center, National Marine Fisheries Service.

References

Breslow, N.E., Clayton, D.G., 1993. Approximate inference in generalized linear mixed models. J. Am. Stat. Assoc. 88, 9–25.

Diggle, P.J., Liang, K., Zeger, S.L., 1994. Analysis of Longitudinal Data. Clarendon Press, Oxford.

Helser, T.E., Punt, A.E., Methot, R.D., 2004. A generalized linear mixed model analysis of a multi-vessel fishery resource survey. Fish. Res. 70, 261–264.

Lauth, R.R., 2001. The 2000 Pacific west coast upper continental slope trawl survey of groundfish resources off Washington, Oregon, and California: estimates of distribution, abundance, and length composition. NOAA Tech. Memo. NMFS-AFSC-120, 292 pp.

Mok, M., 1995. Sample size requirements for 2-level designs in educational research. Multilevel Model. Newsletter 7 (2), 11–15.

NEFSC (Northeast Fisheries Science Center), 2002. Report of the workshop on trawl warp effects on fishing gear performance. Marine Biological Laboratory, Woods Hole, Massachusetts, October 2–3, 2002. Northeast Fish. Sci. Cent. Ref. Doc. 02-15, 90 pp.

Pinheiro, J.C., Bates, D.M., 2000. Mixed-Effects Models in S and S-PLUS. Springer-Verlag, New York.

Raudenbush, S.W., Liu, X., 2000. Statistical power and optimal design for multisite randomized trials. Psychol. Meth. 5, 199–213.

Raudenbush, S.W., Liu, X., 2001. Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. Psychol. Meth. 6, 387–401.

Snijders, T.A.B., Bosker, R.J., 1993. Standard errors and sample sizes for two-level research. J. Ed. Stat. 18, 237–259.

Stoker, L., Bowers, J., 2002. Designing multi-level studies: sampling voters and electoral contexts. Elect. Stud. 21, 235–267.

Turk, T.A., Builder, T., West, C.W., Kamikawa, D.J., Wallace, J.R., Methot, R.D., 2001. The 1998 Northwest Fisheries Science Center Pacific west coast upper continental slope trawl survey of groundfish resources off Washington, Oregon, and California: estimates of distribution, abundance, and length composition. NOAA Tech. Memo. NMFS-NWFSC-50, 122 pp.