Measurement errors in body size of sea scallops (*Placopecten magellanicus*) and their effect on stock assessment models

Larry D. Jacobson (contact author)[1]
Kevin D. E. Stokesbury[2]
Melissa A. Allard[2]
Antonie Chute[1]
Bradley P. Harris[2]
Deborah Hart[1]
Tom Jaffarian[2]
Michael C. Marino II[2]
Jacob I. Nogueira[2]
Paul Rago[1]

Email address for contact author: Larry.Jacobson@noaa.gov

[1] NOAA Fisheries
Northeast Fisheries Science Center
166 Water Street
Woods Hole, Massachusetts 02543-1026

[2] Department of Fisheries Oceanography
School for Marine Science and Technology
University of Massachusetts School of Marine Sciences
838 South Rodney French Boulevard
New Bedford, Massachusetts 02744-1221

Abstract—Body-size measurement errors are usually ignored in stock assessments, but may be important when body-size data (e.g., from visual surveys) are imprecise. We used experiments and models to quantify measurement errors and their effects on assessment models for sea scallops (*Placopecten magellanicus*). Errors in size data obscured modes from strong year classes and increased frequency and size of the largest and smallest sizes, potentially biasing growth, mortality, and biomass estimates. Modeling techniques for errors in age data proved useful for errors in size data. In terms of a goodness of model fit to the assessment data, it was more important to accommodate variance than bias. Models that accommodated size errors fitted size data substantially better. We recommend experimental quantification of errors along with a modeling approach that accommodates measurement errors because a direct algebraic approach was not robust and because error parameters were difficult to estimate in our assessment model. The importance of measurement errors depends on many factors and should be evaluated on a case by case basis.

Manuscript submitted 22 June 2009.
Manuscript accepted 25 January 2010.
Fish. Bull. 108:233–247 (2010).

Two fishery-independent surveys are important for monitoring Atlantic sea scallop (*Placopecten magellanicus*) abundance and biomass levels off the northeastern coast of the United States because they provide abundance, body size,[1] meat weight (weight of marketable adductor muscles), and other data (NEFSC[2,3]). The National Marine Fisheries Service, Northeast Fisheries Science Center (NEFSC) sea scallop dredge survey has been conducted annually since 1977 (Serchuk et al., 1979; Serchuk and Wigley, 1986). In addition, an underwater video survey for sea scallops and other benthic organisms has been conducted annually since 2003 (Stokesbury, 2002; Stokesbury et al., 2004) by the University of Massachusetts Dartmouth, School for Marine Science and Technology (SMAST). The dredge and video surveys are carried out across the range of sea scallops in U.S. waters.

In this analysis, we used sea scallops to draw attention to errors in body-size data when the data are used in a length-structured stock assessment model. The topic of measurement errors in body-size data has received relatively little attention, although Heery and Berkson (2009) evaluated effects of systematic errors (biased sampling) in fishery size-composition data used in an age-structured model. Our work was motivated by questions that arose from examining video survey shell-height data in sea scallop stock assessments (NEFSC[2,3]). Our experimental and analytical results may be important and useful in other situations where body-size data are imprecise. Body-

[1] Shell height (SH, the distance in mm between the umbo and shell margin) is the body size measurement for sea scallops.

[2] NEFSC (Northeast Fisheries Science Center). 2004. Stock assessment for Atlantic sea scallops. *In* 39th northeast regional stock assessment workshop (39th SAW) assessment summary report and assessment report. Northeast Fisheries Science Center, National Marine Fisheries Service, Woods Hole Laboratory, 166 Water St., Woods Hole, MA 02543. Ref. Doc. 04-10, p. 87–211.

[3] NEFSC (Northeast Fisheries Science Center). 2007. Stock assessment for Atlantic sea scallops. *In* 45th northeast regional stock assessment workshop (45th SAW) assessment summary report and assessment report. Northeast Fisheries Science Center, National Marine Fisheries Service, Woods Hole Laboratory, 166 Water St., Woods Hole, MA 02543. Ref. Doc. 07-16, p. 139-370.

size data may be imprecise, for example, when collected by scuba (St. John et al., 1990; Edgar et al., 2004), remotely operated underwater vehicles (ROV; Butler et al., 2006), camera sleds (Rosenkranz and Byersdorfer, 2004), or in other optical surveys where body-size measurements are obtained without handling individual specimens.

In fishery stock assessment modeling, body-size measurements are almost always assumed to be without error. In contrast, statistical sampling errors that arise from too few are often considered in modeling (Fournier and Archibald, 1982; Pennington et al., 2001). Measurement errors in fishery age data have received substantial attention and are often addressed in stock assessment modeling (Methot, 1989, 1990). Approaches to dealing with measurement error in body-size data have not been explored.

Shell-height composition data for sea scallops are of two types: 1) distributions of shell-height measurements, which include measurement errors and true variability among individuals in size; and 2) distributions of shell-height measurements, which include measurement errors only. It is important to distinguish between these two types of data. In particular, shell-height compositions are sample specific and depend on the underlying distribution of true sizes. In our study measurement errors are the difference between the video or board measurements and the true shell height of individual specimens (i.e., after removing differences in true shell height among individuals). Shell-height composition data are important because they are interpreted in stock assessments to estimate year-class strength, mortality, and other biological characteristics. In our study measurement errors are important because they can be used to quantify the accuracy of the measurement process itself and because they affect shell-height data from all samples.

Two types of measurement errors are considered in this study. The first type is bias that causes individual shell-height measurements and estimated sample means to differ, on average, from their true values (Cochran, 1977). The second type is random errors, which cause variability in shell-height measurements and affect the precision of measurements and estimated mean values (Cochran, 1977).

Figure 1 shows how hypothetical errors in sea scallop shell-height measurements tend to smooth the true underlying distribution of the data. Measurement errors tend to smooth modes in the data (which usually correspond to recruitment events) by moving individuals from size bins with relatively high numbers into adjacent bins with lower numbers. Random measurement errors also tend to expand the range of observed sizes by decreasing the smallest observed size and increasing the largest (Fig. 1). Bias degrades body-size data by making measurements consistently larger or smaller than the true value. Methot (1989, 1990) highlighted these issues in the context of age data from survey and fishery samples. We use Methot's modeling methods in our analysis for shell-height data.

In principal, body-size measurement errors can cause errors in a wide range of important fishery estimates but biomass estimates are of particular importance. In the absence of bias, imprecise body-size data tend to cause positive bias in mean weight and biomass estimates because of the nonlinear relationship between size and biomass and Jensen's inequality (Feller, 1966). For example, according to Jensen's inequality, if body weight is a cubic function of body size, then a –10% error in body size will cause a $0.9^3 - 1 = -27\%$ error in estimated body weight for one individual. In contrast, a +10% error in body size will cause a $1.1^3 - 1 = +33\%$ error in body weight. The combined effect of the two errors for two scallops of the same size would be a positive bias of +6%.

The length-based Beverton-Holt mortality estimator involves equilibrium and other assumptions that may make it inappropriate to use in some cases (Gedamke and Hoenig, 2006), but it clearly demonstrates the potential effects of errors in body-size measurements on stock assessment model mortality estimates:

$$Z = \frac{K\left(L_\infty - \bar{L}\right)}{\bar{L} - L_c}, \quad (1)$$

where Z = the instantaneous rate of mortality from all sources;
L_∞ = asymptotic length;
K = rate parameter from the von Bertalanffy growth equation;
$\bar{L}$ = average length of individuals in a sample from the fishery; and
L_c = the "critical" length at which individuals are fully vulnerable to the fishery (Quinn and Deriso, 1999).

With all other factors held constant, a positive bias in $\bar{L}$ will make the numerator in Equation 1 too small, the denominator too large, and the mortality estimate will be biased low. Conversely, a negative bias in $\bar{L}$ will bias the mortality estimate high.

In this article, we characterize measurement errors in shell-height data for sea scallops in two types of surveys, using experimental data. The experimental results are used to evaluate effects on mean body weight and swept-area biomass estimates, and on biomass and mortality estimates from a modern size-structured stock assessment model. The assessment model demonstrates a promising approach (used originally for age data) for accommodating measurement errors in body-size data. In the appendices, we use numerical and bootstrap techniques to evaluate robustness of the assessment model approach in comparison to an algebraic one. Our purpose is not to evaluate the merits of any particular survey, rather, we use sea scallops as an example for dealing with general problems arising from body-size measurement errors in survey and fishery-dependent data, and for suggesting possible approaches to using such data.

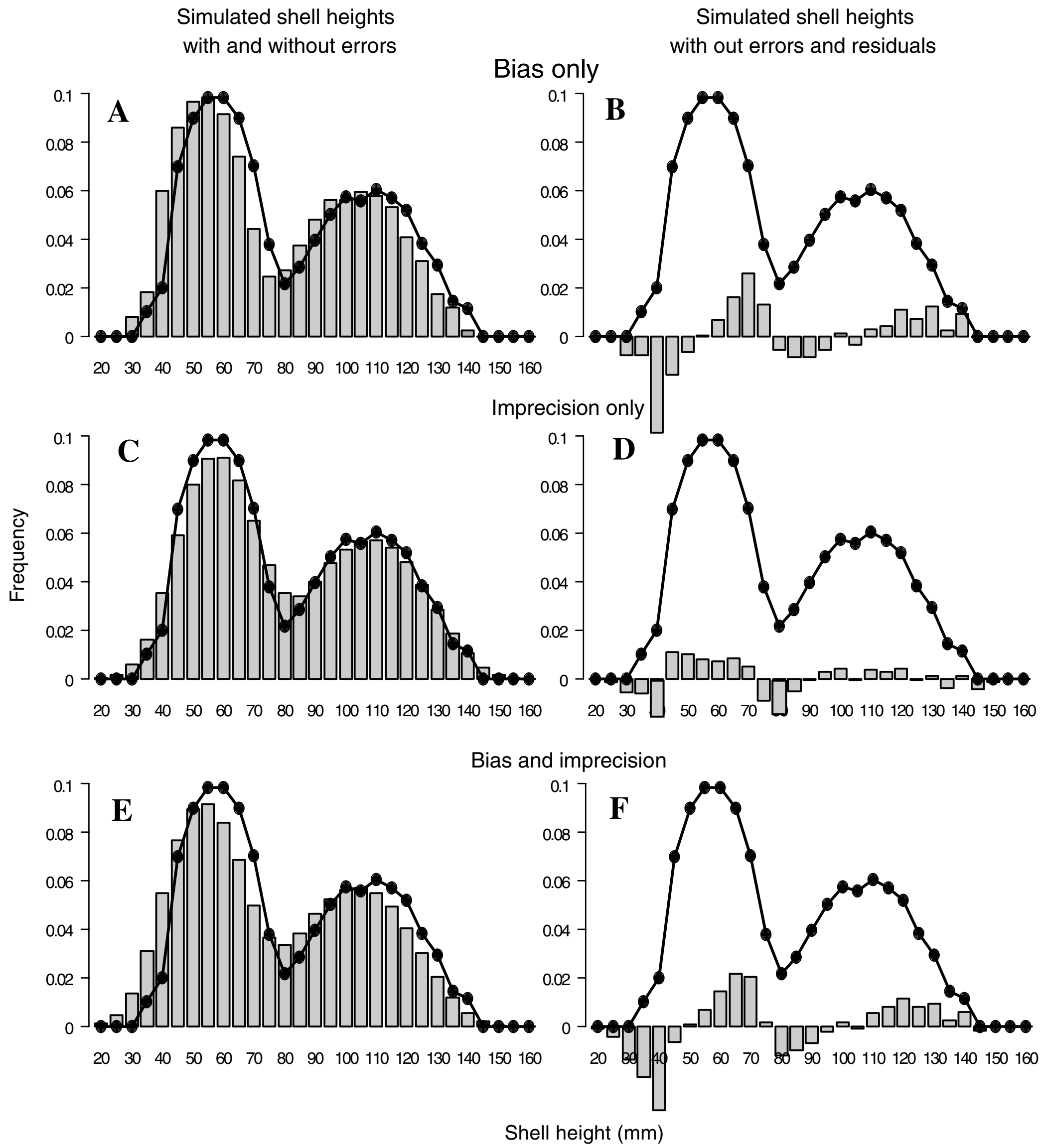


Figure 1

Rootograms (Tukey, 1977) showing hypothetical distributions of Atlantic sea scallop (*Placopecten magellanicus*) shell-height (SH) measurements with and without simulated measurement errors. The black line in each panel shows the distribution of measurements with no errors (5-mm size bins). In the left column, bars show distributions of shell heights with measurement errors. In the right column, bars show residuals (measurement with no errors minus measurements with errors). For the "bias only" scenario (**A** and **B**), precise measurement errors were assumed with a bias of –4.1 mm. For the "imprecision only" scenario (**C** and **D**) unbiased measurement errors were assumed with a standard deviation of 6.1 mm. For the "imprecision and bias" scenario (**E** and **F**), measurement errors were assumed with a bias of –4.1 mm and standard deviation of 6.1 mm.

Materials and methods

The SMAST sea scallop survey is conducted with video cameras mounted on a steel pyramid frame to provide a 3.24-m^2 view of the sea floor and associated macrobenthos (Stokesbury, 2002; Stokesbury et al., 2004). Video images are recorded at sea on high-resolution S-VHS videotape and then replayed in the laboratory where digitized images are created. All sea scallops are counted, and all clearly visible sea scallops (with the hinge and opposite edge visible) within the digitized images are measured to the nearest mm by using Image Pro Plus® software (Media Cybernetics, Inc., Bethesda, MD).

In previous analyses, correction factors were applied to the raw video shell-height measurements to account for distance from the origin (DFO), which is the distance of a specimen from the "origin" (center) of the sampling frame (Stokesbury et al., 2004). Subsequent work during routine stock assessments (unpublished) indicated that adjustments were unnecessary because the distributions of measurement errors were simpler and easier to describe statistically, and data were easier to model without adjustments. Moreover, adjusted data were sometimes less accurate than the unadjusted data. Additional research may result in more accurate adjustments or transformations of body-size data. However, unadjusted video data from the "large" camera on the sampling frame are used in current stock assessments and in this analysis.

NEFSC sea scallop surveys are conducted with a 2.44-m New Bedford sea scallop dredge with a 38-mm liner. The catch is sorted, counted, and measured on the deck of the research vessel. In most cases, the entire catch is counted and measured, but a few large catches were subsampled. During the early 1980s through 2003, sea scallops in the catch were measured to the nearest 5-mm shell-height interval with a standard NEFSC sea scallop measuring board.

Experiments

Two experiments were conducted during 20 and 23 February 2003 when the SMAST video pyramid was placed in a 341,000-L tank filled with seawater in the SMAST laboratory. NEFSC sea scallop measuring boards and SMAST video equipment in the experiments were configured and used in a realistic manner that was similar to use during actual surveys at sea. Accurate measurements used as true shell heights in this analysis were made to the nearest mm by using scientific calipers under laboratory conditions with adequate lighting.

We used the experimental data to evaluate statistical characteristics of shell-height composition data and shell-height measurement errors.

Accuracy, bias, and precision of measurements were quantified by comparing data obtained from the measuring board and video camera with data from the caliper. Accuracy is the closeness to the true underlying value and is measured by mean square error (MSE). For shell-height composition data,

$$MSE = \left(\bar{h} - \bar{H}\right)^2, \tag{2}$$

where $\bar{h}$ = the mean of the measurements; and
$\bar{H}$ = the mean of the true values for the sample (Cochran, 1977).

For measurement errors in our analysis,

$$MSE = \frac{\sum_{j=1}^{n} e_j^2}{n}, \tag{3}$$

where $e_j = h_j - H_j$ = the error for the j^{th} observation (where h_j is the measurement and H_j is the true value).

Bias and variance both contribute to MSE. In fact, MSE = $s^2 + b^2$, where s^2 is the variance and b is bias (Cochran, 1977). In our study, $b=\bar{h}-\bar{H}$ where $\bar{h}$ is the mean of shell-height measurements and $\bar{H}$ is the mean of the true shell heights in the sample. Bias is the same for shell-height composition data and measurement errors as shown below:

$$\sum_{j=1}^{n}\left(h_j - H_j\right)/n = \bar{h} - \bar{H}. \tag{4}$$

Variance (s^2) was computed from shell-height composition data or measurement errors by using the standard formula. Variance of shell-height composition data and measurement errors will generally be different because true shell heights usually differ among specimens in a sample.

It is convenient to express accuracy, bias, and precision in terms of the square root of the MSE (RMSE), bias (b), and standard deviation (s) because all three are absolute measures with the same units (mm for sea scallop shell-height data). Percent RMSE ($RMSE/h_{true}$), percent bias (b/h_{true}), and the CV ($s/\bar{h}$) are useful for making comparisons on a relative basis.

The third and fourth moment statistics, g_1 and g_2, were used to measure skewness (asymmetry) and kurtosis (peakedness) of shell-height composition data and measurement errors, in relation to what would be expected from a normal distribution (Sokal and Rohlf, 1995). Skewness and kurtosis statistics for shell-height composition data and measurement errors from the same sample differ if there is variability in size among specimens. For normally distributed random variables with no skewness, $g_1 = 0$. Negative g_1 values indicate skewness to the left (a distribution with a long left tail and more small values than expected in a normal distribution). Positive g_1 values indicate skewness to the right (long right tail with more large values than expected). Similarly, positive g_2 values indicate distributions more peaked than expected for a normal distribution, and negative g_2 values indicate distributions that are less peaked (flatter) than expected. The two statistics convey information about the shape of any distribution in relation to a normal distribution, but care is required in interpreting g_1 and g_2, particularly for data that are far from normally distributed. The skewness and kurtosis statistics were easier to interpret for measurement errors than for shell-height measurements because the latter were not normally distributed.

We used a test for normally distributed statistics (Sokal and Rohlf, 1995) to evaluate the statistical significance of skewness and kurtosis for distributions of measurement errors that might be otherwise assumed normally distributed. Statistical tests were carried out for distributions of measurement errors because they were closer to normally distributed.

Multiple shell height-measurements were usually made from single specimens in our experiments. We made allowance for repeated sampling when testing skewness and kurtosis by using the number of unique specimens in the experiment as the degrees of freedom instead of the number of measurements (i.e., if n measurements were made on each of k specimens, we used k as the degrees of freedom in statistical tests). The effect of this adjustment was to make the statistical tests more conservative (less likely to reject the null hypothesis of no difference). The number of specimens is a reasonable lower bound estimate of the true effective sample size.

Body weights for sea scallops and other marine organisms are often computed from body size. For sea scallops in this analysis,

$$W = e^{\alpha+\beta \ln(h)}, \tag{5}$$

where W = sea scallop meat weight (g, the weight of the marketable adductor muscle);
h = shell height (mm); and the parameter values α=–12.01 and β=3.22.

Bland-Altman plots (1986, 1995) were used to characterize shell-height measurement errors. In the case of measuring boards, for example, the difference between the measuring board and caliper shell-height measurements for each sea scallop was plotted on the y-axis against the average of the two measures for the same individual on the x-axis. Bland-Altman plots are typically presented as scatter plots with a point for each difference (pair of measurements); however, boxplots may be more useful in some circumstances (see below). Bland-Altman plots are useful because they eliminate spurious correlations when the difference of y–x is plotted against the more precise measure (x) and because patterns are easier to discern along a horizontal line (the x-axis) than along a diagonal line. Spurious correlations occur because the measurement error in x affects the variables plotted on both the x- and y-axes.

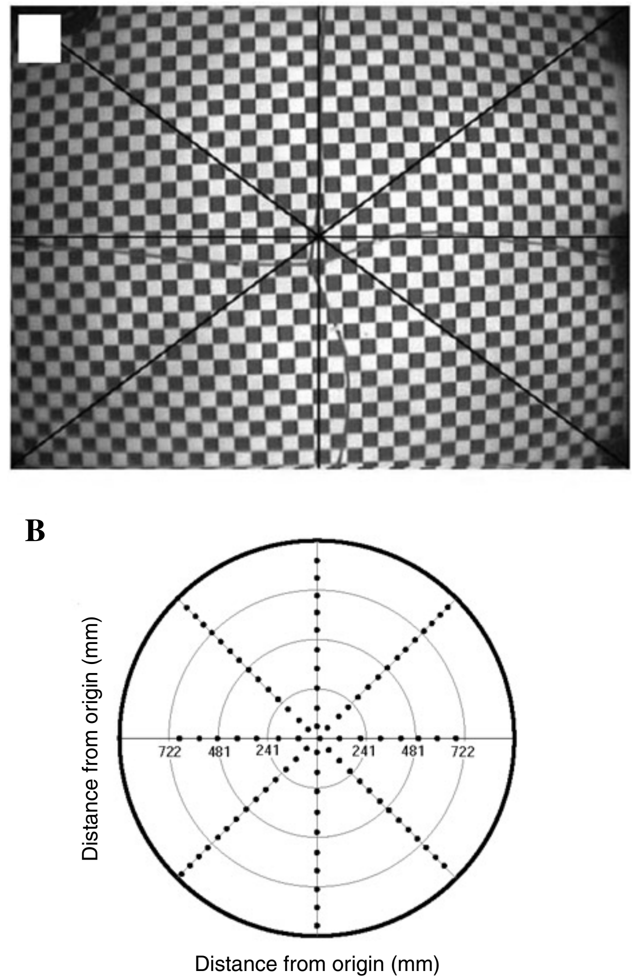


Figure 2

(**A**) A video image showing ceramic tiles under water in the tank where they were placed for experiment 1. (**B**) Dots show the locations of the 91 tiles in the video images that were measured in experiment 1. The height, width, horizontal and vertical distance from the origin (DFO) was recorded for each tile measured. The x- and y-axis are the same in the top and bottom panels.

Experiment 1 was designed to measure the accuracy of video measurements for objects of known size (square ceramic tiles) as a function of position in the video frame as measured by DFO (Fig. 2). Scuba divers in experiment 1 placed black and white ceramic floor tiles (all were 48.5×48.5 mm) in a closely packed square grid on the bottom of the tank, starting at the center of the video pyramid and covering the entire range of view in actual surveys (Fig. 2). The width and height of 91 tiles across the field of view and at various distances and positions from the center of the sampling frame (Fig. 2) were estimated from video images by using the standard video survey procedures described above. Data were recorded in such a way that the length and height measurements from the same tile could be associated with each other and with the particular position of the tile in the video image. The tiles used in experiment 1 (48.5×48.5 mm) corresponded roughly with the size of the smallest scallops fully recruited to the dredge and video surveys (about 40 mm SH) and included in stock assessment analyses. Sea scallops, according to actual survey data, cover a much wider range of shell heights (to about 190 mm SH in experiment 2, see *Discussion* section).

Experiment 2 was designed to measure the accuracy of video shell-height measurements for sea scallop shells of varying sizes (39 to 192 mm SH) placed randomly on a sand-granule-pebble substrate, similar to the random aggregations observed on Georges Bank. All shell-height measurements could be linked with each individual sea scallop in experiment 2 because the right valve of 172 individual sea scallop shells was numbered uniquely. The identification numbers were large and written under the valve with dark indelible ink and clearly visible with video equipment when the sea scallops were turned over so that the labels faced the camera. The numbered sea scallops were assigned randomly to fifteen groups. All members of the same group were stored together in a bag with a unique label for group identification.

In each experimental replicate, a group of shell valves was placed randomly on the bottom of the tank. Two video images were made for each group. The first image (with the valve turned towards the sediment and identification numbers hidden) was used by four technicians to independently measure shell heights. The second image was taken with identification numbers visible after divers turned the shells over and replaced them in their original positions. After video images were recorded, the shell valves were measured with measuring boards by two technicians who could not see the identification numbers and once by a third technician with calipers.

A stock assessment model that incorporates errors from shell-height measurements

Following NEFSC[2,3] procedures, we used results from experiment 2 and a modified version of the CASA (catch-at-size-analysis, Sullivan et al., 1990) stock assessment model (Appendix 1) to investigate potential effects of shell-height measurement errors on model-based biomass and fishing mortality estimates for two sea scallop stocks. Assessment model results in this article should not be used by managers because model runs were tailored to investigate potential effects of shell-height measurement errors and because some types of data were omitted.

As described in Appendix 1, the CASA model that is routinely used for sea scallop stock assessments accommodates both bias and imprecision in shell-height measurements. CASA models were run for sea scallops in the Mid-Atlantic Bight during 1982–2006. In contrast to NEFSC[2], measurement error parameters were obtained from experiments and not estimated in the CASA model itself. The data used in modeling included commercial landings in metric tons (t), survey trend data (numbers per unit of sampling effort) from the camera video and dredge surveys, and shell-height composition data from the commercial fishery, video, and dredge surveys. Survey selectivity patterns were not estimated because the video and dredge surveys have flat selectivity patterns (catch sea scallops equally well) at shell height ≥40 mm, and goodness-of-fit calculations were restricted to this size range (Appendices B7–B8 in NEFSC[3]). Measurement errors in commercial shell-height data were assumed to be the same as those in the dredge survey for lack of better information and because procedures for measuring sea scallops on land in port samples and at-sea in fishery observer samples are similar to procedures followed in surveys.

As described in Appendix 1, bias and precision of shell-height measurements are represented in the CASA model by an error matrix (E) that gives the probability that a sea scallop in each true shell-height bin is assigned to a range of observed shell-height bins (a range that accommodates measurement errors). As described by Methot (1989, 1990) for age data, the error matrix E can be set up to deal with a wide range of situations for bias and variance (e.g., both can vary among shell-height bins or over time).

For the calculation of E for sea scallops in this analysis, shell-height measurement error distributions were assumed to be normally distributed with means and standard deviations from experiment 2. The normal distributions for measurement errors were truncated three standard deviations above and below the mean. In calculating distributions of measurement errors, true shell heights were assumed with or without bias to be uniformly distributed within each true 5-mm SH bin so that, for example, the frequency of sea scallops with true shell heights of 70, 71, 72, 73, and 74 mm (in the 70–74.9 mm SH bin with midpoint 72.5) was the same. Distributions for measurement errors were normalized to sum to one before use in the CASA model.

Results

Height and width measurements from the same tiles in experiment 1 were not significantly different by a paired t-test (t=–0.23, P=0.30, 91 df). Therefore, height and width measurements from 91 tiles in experiment 1 were combined to form a single set of video data (a total of 182 measurements) (Table 1).

The RMSE statistic for video tile-size composition and measurement errors in experiment 1 (Table 1) was 3.5 mm (%RMSE=7%, Table 1). Bias (–2.2 mm) and imprecision (standard deviation 2.7 mm) of video tile measurements were similar. In comparison to the true size of the tiles (48.5 mm), the smallest measurement was 38 mm, and the largest measurement was 50 mm. The video size-composition data and measurement errors were left skewed (g_1=–0.28) and flatter (g_2=–0.53) than expected for a normal distribution. There were gaps in the distribution of the video tile measurements (Fig. 3) due to the resolution of the video images used in digitizing (each pixel≈3×3 mm).

Measurement error increased with DFO for the video tile measurements (Fig. 3). Bias was positive for DFO <400 mm and negative at larger DFO levels.

RMSE for shell-height composition data in experiment 2 was 33 mm (%RMSE 30%) for video and 34 mm (%RMSE=31%) for measuring board data (Table 2). Mean shell height was 106 mm for video and 109 mm for measuring boards, compared to 110 mm for calipers. Minimum shell height was 34 mm for video, 38 mm for measuring boards, and 39 mm for calipers. Maximum shell height was 201 mm for video, 193 mm for measuring boards, and 192 mm for calipers.

Bland-Altman plots for experiment 2 show that measuring board shell heights were more accurate than video measurements, and that bias in video and measuring board data was relatively constant across the range of shell heights in experiment 2 (Fig. 4). However, relatively large outliers sometimes occurred in video measurements at 80–130 mm SH (Fig. 4).

Video and measuring-board shell-height compositions in experiment 2 were similar in terms of skewness with g_1=–0.41 for video measurements and –0.47 for measuring boards compared to –0.46 for calipers (Table 2). The video shell-height distribution was more peaked with g_2=–0.65 compared to g_2=–0.85 for measuring boards, and g_2=–0.84 for calipers (Table 2). Video measurement errors were skewed to the left (g_1=–0.60) compared to measuring-board errors which were nearly symmetrical (g_1=–0.05). The distribution of errors for measuring boards was flatter (g_2=–0.85) and video measurement errors were more peaked (g_2=1.84) than would be expected for normal distribution. The error distribution for measuring boards had a nearly flat mode about 5-mm wide because shell heights are automatically truncated by measuring boards to the next lowest 5-mm shell-height bin.

Table 1

Summary of size-composition data and measurement errors for 182 tile measurements (height and width from 91 tiles, each 48.5×48.5 mm) by video equipment in experiment 1.

Statistic	Video
Measurements and measurement errors	
Bias	–2.2
Standard deviation	2.7
Square root of the mean squared error	3.5
Skewness (g_1)	–0.28
Kurtosis (g_2)	–0.53
Measurements	
Minimum	38.3
5% quantile	41.2
95% quantile	50.1
Maximum	50.1
Average	46.3
Percent bias	–5%
Coefficient of variation	6%
Percent square root of the mean squared error	7%

On a proportional basis, meat weights calculated from shell heights in experiment 2 were much less accurate than the original shell-height measurements. In particular, %RMSE values for meat weights were 71% and 74% for video and measuring boards, respectively (Table 3), compared to 30% and 31% for the original

Table 2

Summary statistics for shell-height composition data and measurement errors (in mm) from 172 uniquely identified Atlantic sea scallop (*Placopecten magellanicus*) shell valves in experiment 2. "NA" means that a statistic is not applicable.

Statistic	True shell height (calipers)	Video	Measuring boards
Shell heights and measurement errors			
n measurements used	172	670	344
n omitted	0	18	0
Bias	NA	–4.5	–0.6
Shell heights			
Minimum	38.5	34.3	37.5
5% quantile	54.8	48.8	52.5
95% quantile	149.6	147.3	147.5
Maximum	192.0	200.6	192.5
Average	109.9	106.5	109.3
Percent bias	NA	–4%	–1%
Standard deviation	33.5	33.1	33.6
Coefficient of variation	30%	31%	31%
Square root of the mean squared error	NA	33.4	33.6
Percent square root of the mean squared error	NA	30%	31%
Skewness (g_1)	–0.46	–0.41	–0.47
Kurtosis (g_2)	–0.84	–0.65	–0.85
Measurement errors			
Standard deviation	NA	6.1	1.7
Square root of the mean squared error	NA	7.6	1.8
Skewness (g_1)	NA	–0.60	–0.044
Kurtosis (g_2)	NA	1.84	–0.85

shell heights (Table 2). The nonlinear shell-height to meat-weight relationship showed exaggerated extremes of the distributions so that the ratio of maximum to mean meat weight was 158/27=5.9 for video data and 138/29=4.8 for measuring boards (Table 3) compared to 201/106=1.9 and 193/109=1.8 for shell heights (Table 2). Variance in meat-weight measurements increases as true meat-weight increases for video data and, to a lesser extent, for measuring boards (Fig. 5).

The meat-weight composition data were more right skewed (g_1=1.53) and flatter (g_2=6.22) than the meat-weight composition data from measuring boards (g_1=0.92 and g_2=2.61) or calipers (g_1=0.99 and g_2=3.00). Errors in meat-weight data were left skewed and not as peaked for video (g_1=–0.80 and g_2=2.48) than measuring board data (g_1=–1.06 and g_2=4.68).

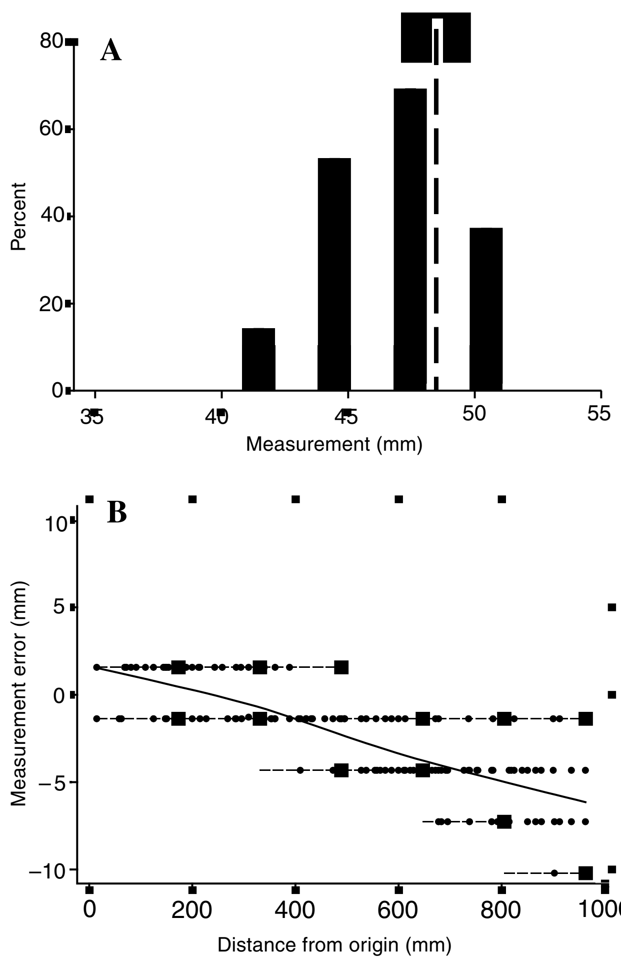


Figure 3

(**A**) Video measurements for tiles in experiment 1. The vertical line shows the true value at 48.5 mm. (**B**) Measurement errors (video measurement minus caliper measurement) for tiles in experiment 1 as a function of distance from the origin (DFO). The nonlinear LOESS regression line shows the overall trend in measurement errors as a function of DFO.

Results from the assessment models

Based on results from experiment 2 (Table 2) and assumptions listed above, video shell-height measurements for sea scallops with true sizes evenly distributed over 100–104.99 mm SH (i.e., the 100-mm bin with midpoint 102.5 mm) would fall into nine observed shell-height bins with midpoints from 77.5 to 117.5 mm (Table 4). Measuring board shell-height measurements would fall into four observed shell-height bins with midpoints ranging from 92.5 to 107.5 mm (Table 4).

Four model configurations were used. The "no measurement error" model configuration was fitted by assuming no errors in shell-height data. The "bias only" model was fitted by assuming that shell-height data were biased (to the extent measured in experiment 2), but precise (with zero variance). The "imprecision only" model was fitted by assuming that shell-height measurements were imprecise (standard deviations from experiment 2), but not biased. The "imprecision and bias" model was fitted by assuming both types of shell-height measurement errors.

Models which accommodated measurement errors fitted better, with substantially lower negative log likelihoods for both stocks, than models that ignored measurement errors. Differences in negative log likelihood were mostly for shell-height composition data. Mean 2004–06 biomass and fishing mortality rates and coefficients of variation (CV) for biomass and fishing mortality estimates were similar for all model configurations (Table 5).

Discussion

The importance of body-size measurement errors and the need to accommodate them in modeling probably depends on the situation. Biological factors (growth rate, recruitment variability), assessment model type, quality and quantity of fishery and fishery-independent data may be important. Sea scallops may be an atypical case because they are a data-rich species. We suggest that the potential importance of body size measurement errors should be evaluated on a case by case basis, particularly if body-size data may be imprecise or biased. Simulation studies may be useful in determining the importance of experimentally derived body-size measurement errors on stock assessment results.

In the sea scallop case, models that accommodated measurement errors fitted substantially better, but there was little effect on point estimates and variances for recent biomass and fishing mortality. We hypothesize that effects on biomass and mortality estimates would be larger in cases with positive biases in body-size measurements. For both video and measuring boards, the positive bias in meat weights due to the nonlinear relationship between body size and meat weight was mitigated to some extent by the negative bias in shell-height measurements. In contrast, Heery and Berkson (2009) used simulations to evaluate effects of systematic sampling errors (too many small or too many large individuals) in size-composition data from commercial catches and three simulated stocks. The simulated data were used in a forward-projecting age-structured stock assessment and in projection models to estimate stock size and fishing mortality in relation to threshold values, and rebuilding trajectories. Body-size data with too many large individuals biased stock size high and fishing mortality low and tended to support management measures that did not meet management goals, particularly for longer lived and depleted stocks. Body-size data with too many small individuals were less problematic, but tended to support overly restrictive management actions in extreme cases. Heery and Berkson's (2009) results indicate that systematic errors in sampling may be more important than errors in individual measurements of body size.

Variance in calculated meat weights increased rapidly with shell height with both video and measuring board techniques, in contrast to the variance in shell heights (Figs. 4 and 5). This additional source of variability likely increases variance in biomass estimates, particularly for relatively large fishable sea scallops.

In our analysis, assessment models that accommodated shell-height measurement errors fitted better, even though no additional parameters were estimated. The Mid-Atlantic Bight model that accommodated imprecise (but not biased) shell-height measurement errors had a negative log likelihood that was 15 units smaller than the negative log likelihood for the no measurement error model (Table 5). Results for the Georges Bank stock (not shown to conserve space) were similar. In contrast and based on likelihood theory, a difference in negative log likelihoods of just 1.92 units is sufficient to justify an additional parameter in a statistical model at the $P=0.05$ level (Venzon and Moolgavkar, 1988). Comparing results of the "bias only" scenario to results from the "imprecision only" and "imprecision and bias" scenarios, we found that improvements in goodness of fit were mostly due to accommodating imprecision; bias was less important (Table 5).

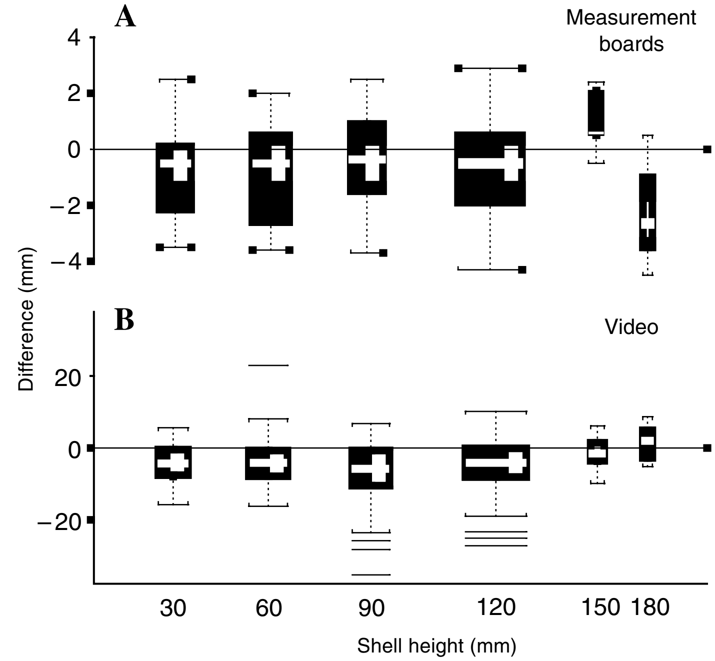


Figure 4

Modified Bland-Altman plots for Atlantic sea scallop (*Placopecten magellanicus*) shell-height (SH) measurements in experiment 2. The *y*-axis shows the difference between the experimental measurement (measuring boards in **A** or video in **B**) and the caliper measurement. The *x*-axis shows the average of the experimental and caliper measurement. Boxplots and 30-mm shell-height bins were used instead of traditional scatter plots for shell height measurements in experiment 2 because the large number of samples between 120 and 150 mm SH gave the impression that variance was higher for those sizes. Boxplots show the interquartile range (a robust variance measure) and are not sensitive to sample size. The width of the boxplots is proportional to the number of observations for the shell-height bin.

Experiments

Our results highlight the value and information that may be gained from evaluating body size measurement errors experimentally. Body-size measurement error experiments should be conducted when survey equipment is changed, particularly if body-size measurements are imprecise. In some cases, frequent "mini-experiments" may be required if the accuracy of the equipment tends to drift over time or change in response to environmental conditions.

Our results indicate the importance of designing measurement error experiments so that individual specimens can be identified and associated with individual measurements; otherwise measurement errors can not be estimated individually and evaluated directly. Data from experiment 2 were most useful because individual sea scallops were numbered and replicate measurements of different types could be linked and analyzed in detail. In addition, the full range of variability for all important factors (i.e., distance from the origin (DFO), shell height, and identity of individual technicians) should be included in the experimental design.

We ignored skewness and kurtosis in measurement errors in calculating measurement error matrices for use in the CASA stock assessment model. In future modeling, it may be better to use the experimental dis-

Table 3

Summary statistics of meat weights and meat weight measurement errors (g) for Atlantic sea scallop (*Placopecten magellanicus*) shell-height measurements in experiment 2 (sample sizes are the same as those for shell-height measurements in Table 2). The original shell heights were obtained with calipers, video camera, and measure boards. "NA" means that a statistic is not applicable.

Statistic	True (calipers)	Video	Measuring boards
Meat weights and measurement errors			
Bias	NA	–3.2	–0.4
Meat weights			
Minimum	0.8	0.5	0.7
5% quantile	2.4	1.7	2.1
95% quantile	61.3	58.3	58.6
Maximum	136.9	157.7	138.0
Average	29.8	27.3	29.4
Percent bias	NA	–10%	–1%
Standard deviation	22.2	21.4	21.8
Coefficient of deviation	74%	78%	74%
Square root of the mean squared error	NA	21.6	21.8
Percent square root of the mean squared error	NA	71%	74%
Skewness (g_1)	0.99	1.53	0.92
Kurtosis (g_2)	3.00	6.22	2.61
Measurement errors			
Standard deviation	NA	5.1	1.5
Square root of the mean squared error	NA	6.0	1.6
Skewness (g_1)	NA	–0.80	–1.06
Kurtosis (g_2)	NA	2.48	4.68

Table 4

Estimated probability distributions for Atlantic sea scallop (*Placopecten magellanicus*) shell-height (SH) measurements based on bias and standard deviations from experiment 2. Condition factors for error matrices used in the catch-at-size-analysis (CASA) stock assessment model scenarios are given also. The shell-height bins are 5-mm wide and identified by their midpoint. For example, sea scallops 80–84.9 mm SH fall into a bin whose midpoint is 82.5 mm.

		Video scenario			Measuring board scenario		
Statistic	Calipers (true shell height)	Bias only	Imprecision only	Imprecision and bias	Bias only	Imprecision only	Imprecision and bias
Condition factor (κ)	NA	3×10^{15}	5457	2638	1.6	2.1	2.3
Bias (mm)	0	–4.5	0	–4.5	–0.6	0	–0.6
Standard deviation (mm)	0	0	6.1	6.1	0	1.7	1.7
Shell height bin (mm)	Probability of observed bins						
72.5							
77.5				0.0009			
82.5			0.0014	0.0167			
87.5			0.0203	0.0820			
92.5			0.0929	0.2158			0.0001
97.5		0.8000	0.2300	0.3101	0.2000	0.1325	0.2008
102.5	1.0000	0.2000	0.3110	0.2436	0.8000	0.7349	0.7181
107.5			0.2300	0.1045		0.1325	0.0810
112.5			0.0929	0.0243			
117.5			0.0203	0.0020			
122.5			0.0014				
127.5							

tributions of measurement errors directly in error matrices, particularly if experimental sample sizes are large.

Drouineau et al. (2008) used simulation analysis to show the importance of alternative assumptions about the distribution of individuals within size groups and the statistical distribution of growth increments in length-structured models like the CASA (catch-at-size-analysis) model. Our experience indicates that the same types of assumptions are important in calculating body-size measurement-error matrices. In particular, it was important to assume that individuals were uniformly distributed within size groups, to make realistic assumptions about the distributions of measurement errors, and to be careful in programming to ensure consistent calculations at the boundaries of length bins for calculating error matrices and for the stock assessment model.

Statistical methods for repeated measurements or random effects may be suitable for analysis of our experimental data. We made allowances for repeated measures in bootstrap calculations (Appendix 2) and in calculating P-values for skewness and kurtosis tests, but not in calculating other statistics (Tables 1–3).

Our experiments were conducted under ideal conditions with tiles and shell valves, rather than live sea scallops. Our results may underestimate the magnitude of errors under more realistic field conditions.

Model results may depend on shell-height bin width such that larger shell height bins would cause measurement errors to have a greater impact on biomass and mortality estimates. We used 5-mm SH bins for sea scallops because 5-mm is the resolution and approximate accuracy for the survey shell-height data. In general, it may be important to consider the magnitude of measurement errors in making decisions about size bins used in stock assessment modeling.

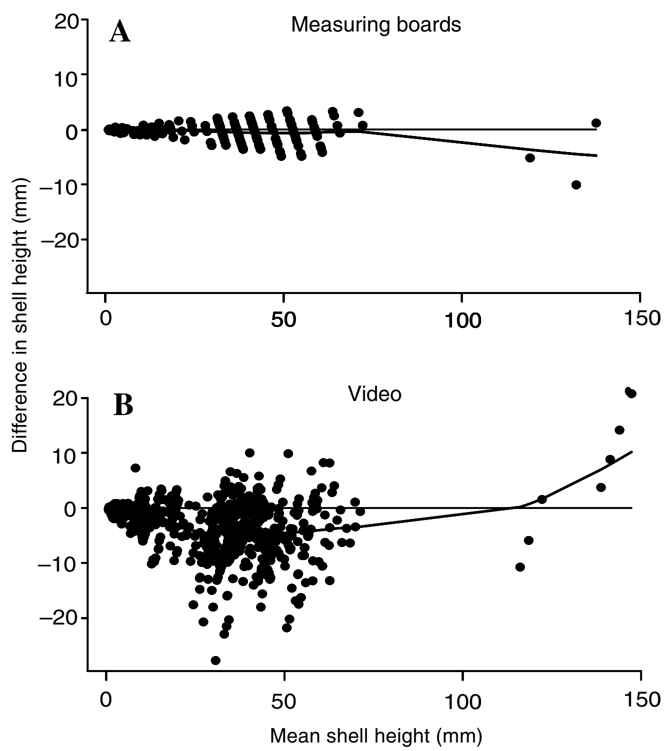


Figure 5

Bland-Altman plots for Atlantic sea scallop (*Placopecten magellanicus*) meat weights calculated from experimental shell-height measurements in experiment 2 (measuring boards in panel **A** and video in panel **B**). The y-axis shows the difference between the meat weights calculated from the experimental (video or measuring board) shell height measurements and the meat weights calculated from caliper measurements. The x-axis shows the average of the experimental and caliper-derived measurements.

Body-size measurement errors

Random measurement errors are unavoidable. One may conclude that it is incumbent on the researcher to search out and correct sources of bias, whatever the source. We suggest that it may be more cost effective to quantify measurement errors experimentally and to accommodate them in modeling. Time series with consistent body-size measurement errors are probably easiest to interpret. Models may become overly complex if multiple sets of assumptions about measurement errors are required to interpret one survey time series. Resources required to quantify measurement errors after each adjustment to survey procedures or equipment may be better spent on more accurately characterizing the measurement errors for survey gear that remains the same for longer periods of time.

Bootstrap results also showed that an algebraic approach to removing errors from the data by using the inverse error matrix E^{-1} gave negative proportions for both video and measuring board data in at least some size groups (Appendix 2). The sampling distribution for algebraically adjusted shell-height data may be difficult to characterize. These results indicate that it may be difficult to remove measurement errors directly from body-size data and we hypothesize that approaches like the one used in the CASA model will generally perform better. Bootstrap results showed that estimates of predicted shell-height composition data with measurement errors as carried out in the CASA model were robust to uncertainties in the measurement-error matrix E (Appendix 2). Models can be designed to be robust to measurement errors. For example, the last size bin in the CASA model is a plus-group that absorbs data for large scallops that may have been strongly affected by measurement errors. Other data in the model may have also contributed to the robustness of biomass and

Table 5

Results from the catch-at-size-analysis (CASA) model for Mid-Atlantic Bight sea scallops (*Placopecten magellanicus*) and four model configurations. The "no measurement error" model configuration does not accommodate shell-height measurement errors. Other model configurations accommodate bias and imprecise measurement errors in various combinations as shown in the table. Lower negative log likelihood (NLL) values indicate better model fit. Coefficients of variation (CV) shown in parenthesis are asymptotic variances calculated by the delta method. For ease of comparison, the "no measurement error" configuration NLL values were subtracted from corresponding NLL statistics for all three configurations. The lowest NLL, biomass or fishing mortality estimates in each row are printed in boldface.

Variable or estimate	No measurement error	Bias only	Imprecision only	Imprecision and bias
Bias and precision (mm) assumed in modeling				
Standard deviation—video survey	0.0	0.0	6.1	6.1
Bias—video survey	0.0	–4.5	0.0	–4.5
Standard deviation—dredge survey	0.0	0.0	1.7	1.7
Bias—survey	0.0	–0.6	0.0	–0.6
Negative log likelihood (NLL)				
Total	0.00	20.92	**–14.62**	–1.16
Commercial fishery shell-height data	0.00	4.99	**–0.34**	2.06
Dredge survey shell-height data	0.00	–4.14	**–10.66**	–6.97
Video survey shell-height data	0.00	19.45	–3.00	4.59
Mean biomass and fishing mortality during 2004–06				
Fishing mortality (*y*–1)	0.45	0.41	0.46	**0.42**
	(8%)	(7%)	(8%)	(8%)
Biomass (t meats)	81,211	84,650	**80,844**	83,602
	(5%)	(5%)	(5%)	(5%)

fishing mortality estimates to assumptions about shell-height measurement errors.

In principal, measurement-error parameters could be estimated directly in stock assessment models without resorting to experiments. Measurement-error parameters in the CASA model were estimated in the NEFSC study,[2] but the estimates proved to be unstable (NEFSC[3]). Without at least one source of accurate body-size data, there may be too little information about measurement errors to estimate parameters. In addition, there may be strong correlations between estimated measurement errors and estimates of other factors that affect interpretation of body-size data, such as survey and fishery selectivity, natural mortality, and recruitment variability.

Acknowledgements

We thank F. Serchuk (Northeast Fisheries Science Center, Woods Hole, MA), S. Correia (Massachusetts Division of Marine Fisheries, New Bedford, MA), C. O'Keefe and C. Adams (SMAST, New Bedford, MA), and five anonymous reviewers for useful technical and editorial suggestions. We are grateful for support from the School of Marine Science and Technology, the Massachusetts Division of Marine Fisheries, and NOAA awards: NA04NMF4720332, NA05NMF4721131, and NA06NMF4720097. We are grateful to the crews and scientific staff who collected and measured sea scallops in NEFSC and SMAST surveys. Live sea scallops used in the experiments were provided by commercial sea scallop vessels from New Bedford and Fairhaven, MA.

Literature cited

Bland, J. M., and D. G. Altman.
1986. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1:307–310.
1995. Comparing methods of measurement: why plotting difference against standard method is misleading. Lancet 346:1085–1087.

Butler, J., M. Neuman, D. Pinkard, R. Kvitek, and G. Cochrane.
2006. The use of sonar mapping techniques to refine population estimates of the endangered white abalone (*Haliotis sorenseni*). Fish. Bull. 104:521–532.

Cochran, W. G.
1977. Sampling techniques, 428 p. John Wiley and Sons, Inc., New York.

Drouineau, H., S. Mahevas, M. Bertignac, and A. Fertin.
2008. Assessing the impact of discretisation assumptions in a length-structured population growth model. Fish. Res. 91:160–167.

Edgar, G. J., N. S. Barrett, and A. J. Morton.
2004. Bias associated with the use of underwater visual census techniques to quantify the density and

size-structure of fish populations. J. Exp. Biol. Ecol. 308:269–290.
Feller. W.
1966. An introduction to probability theory and its applications. Volume II. 626 p. John Wiley and Sons, Inc., New York.
Fournier, D., and C .P. Archibald.
1982. General theory for analyzing catch at age data. Can. J. Fish. Aquat. Sci. 39:1195–1207.
Gedamke, T., and J. M. Hoenig.
2006. Estimating mortality from mean length data in nonequilibrium situations, with application to the assessment of goosefish. Trans. Am. Fish. Soc. 135:476–487.
Heery, E. C., and J. Berkson.
2009. Systematic errors in length frequency data and their effect on age-structured stock assessment models and management. Trans. Am. Fish. Soc. 138:218–232.
Horn, R. A., and C. R. Johnson.
1985. Matrix analysis, 561 p. Cambridge Univ. Press, New York.
Methot, R. D.
1989. Synthetic estimates of historical abundance and mortality for northern anchovy. Am. Fish. Soc. Symp. 6:66–82.
1990. Synthesis model: an adaptive framework for analysis of diverse stock assessment data. Int. N. Pacific Fish. Comm. Bull. 50:259–277.
Pennington, M., L.-M. Burmeister, and V. Hjellvik.
2001. Assessing the precision of frequency distributions estimated from trawl-survey samples. Fish. Bull. 100:74–80.
Quinn, T., and R. B. Deriso.
1999. Quantitative fish dynamics, 570 p. Oxford Univ. Press, Oxford, U.K.
Ricker, W. E.
1975. Comparison and interpretation of biological statistics of fish populations. Fish. Res. Board Can. Bull. 191:1–382.
Rosenkranz, G. E., and S. C. Byersdorfer.
2004. Video scallop survey in the eastern Gulf of Alaska, USA. Fish. Res. 69:131–140.
Serchuk, F. M, and S. E. Wigley.
1986. Evaluation of USA and Canadian research vessel surveys for sea scallops, (*Placopecten magellanicus*) on Georges Bank. J. Northw. Atl. Fish. Sci. 7:1–13.
Serchuk, F. M., P. W. Wood Jr., J. A. Posgay, and B. E. Brown.
1979. Assessment and status of sea scallop (*Placopecten magellanicus*), populations off the Northeast coast of the United States. Proc. Natl. Shellfish. Assoc. 69:161–191.
Sokal, R. R., and F. J. Rohlf.
1995. Biometry, 859 p. Freeman, New York.
St. John, J., G. R. Russ, and W. Gladstone.
1990. Accuracy and bias of visual estimates of numbers, size structure, and biomass of a coral reef fish. Mar. Ecol. Prog. Ser. 64:253–262.
Stokesbury, K. D. E.
2002. Estimation of sea scallop abundance in closed areas of Georges Bank, USA. Trans. Am. Fish. Soc. 131:1081–1092.
Stokesbury, K. D. E., B. P. Harris, M. C. Marino II, and J. I. Nogueira.
2004. Estimation of sea scallop abundance using a video survey in off-shore US waters. J. Shellfish Res. 23:33–44.
Sullivan, P. J., H.-L. Lai, and V. F. Gallucci.
1990. A catch-at-length analysis that incorporates a stochastic model of growth. Can. J. Fish. Aquat. Sci. 47:184–198.
Venzon, D. J., and S. H. Moolgavkar.
1988. A method for computing profile-likelihood based confidence intervals. Appl. Stat. 37:87–94.
Tukey, J. W.
1977. Exploratory data analysis, 688 p. Addison-Wesley Pub. Co., Reading, MA.

Appendix 1

Following the approach of the Northeast Fisheries Science Center (NEFSC,[2,3]) we used a likelihood approach to fitting the CASA model to sea scallop stock assessment data. The best estimates from the model minimized the combined negative log likelihood of all the data. Relevant details are described below. Appendix B10 in the NEFSC report (NEFSC[3]) is a complete technical description of the CASA model for sea scallops. Appendix B12 in that same report (NEFSC[3]) describes CASA model performance with simulated stock assessment data.

Estimates of population abundance and survey size selectivity are available for each shell height and year as the CASA model is fitted. In a single year, for example, we calculated the number of sea scallops in the population that were available or selected by the video gear with the following equation:

$$n_h = q_h N_h, \qquad \text{(A1)}$$

where N_h = the predicted number of sea scallops in the population for shell height bin *h;*
q_h = the size-specific probability of detection (selectivity) in the video survey (on a scale of 0 to 1 and relative to the bin with maximum probability of detection); and
n_h = the estimated number of sea scallops in the population that are available to the video survey gear.

In the absence of measurement error, the predicted shell-height composition π_h for the survey is

$$\pi_h = \frac{n_h}{\sum_{i=1}^{L} n_i}, \qquad \text{(A2)}$$

where L = the number of shell-height bins in the model.

If $\vec{\pi}$ is a row vector of length L containing the predicted proportions (before measurement errors) for each length group in the survey, then

$$\vec{p} = \vec{\pi} E, \qquad \text{(A3)}$$

where $\vec{p}$ the row vector of predicted proportions (including measurement errors).

In Equation A3, E is a square measurement error matrix with L rows and columns that distributes numbers at true shell height into observed shell heights bins that are larger and smaller than the true shell height. For example, the first row of E sums to one and gives the probability of observed shell heights for sea scallops in the first true shell height bin. The last row of E sums to one and gives the probabilities that sea scallops in each shell height bin would be assigned to the "plus group" because of measurement error. As described in the text, we estimated E for sea scallops using results from experiment 2.

Appendix 2

Equation A3 in Appendix 1 indicates the possibility of correcting shell-height data measurement algebraically, without resorting to an approach like the CASA model. In particular, if the matrix E is invertible, then it may be possible to estimate the true sample proportions $\hat{\hat{\pi}}$ by multiplying both sides of Equation A3 by the inverse matrix E^{-1}:

$$\hat{\hat{\pi}} = \vec{p}E^{-1}. \qquad \text{(A4)}$$

However, the inverse calculation in Equation A4 will be unreliable if the estimated error matrix E is poorly conditioned. If the error matrix is poorly conditioned, then small inaccuracies in the estimate of E will propagate into larger errors in the inverse E^{-1} and the predicted proportions $\hat{\hat{\pi}}$.

As described by Horn and Johnson (1985), the condition factor for an invertible matrix E is

$$\kappa = \|E\| \, \|E^{-1}\|, \qquad \text{(A5)}$$

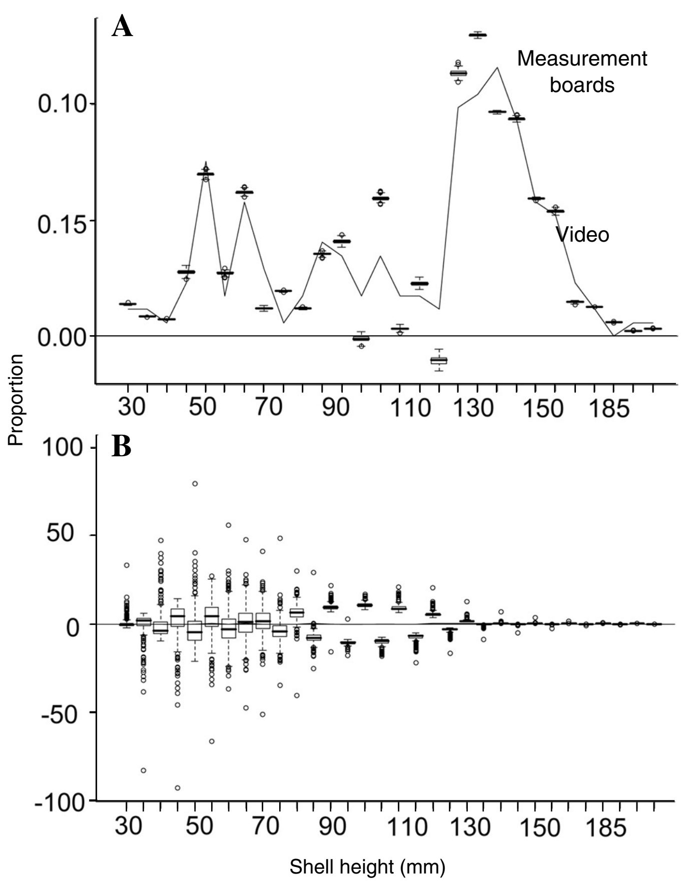
where $\|E\|$ = the matrix norm of E.

The condition factor κ is always at least one and is an upper bound measure of the extent to which errors in the original error matrix E (ignoring errors in $\vec{p}$) will propagate to its inverse. If κ is slightly larger than one, then uncertainty in E^{-1} and $\hat{\hat{\pi}}$ from Equation A4 will be at most slightly greater then uncertainty in E. If κ is large, then uncertainty in E^{-1} and $\hat{\hat{\pi}}$ may be much larger than uncertainty in E.

The measurement-error matrices that included both bias and imprecision are the most realistic according to results from experiment 2. The condition factors for these error matrices were 2638 for video and 2.3 for measuring boards (Table 4). These condition factors indicate that uncertainty in E^{-1} and "corrected" shell-height composition data could be much higher than uncertainty in the original error matrix E for video and at most 2.3 times higher for measuring boards.

Bootstrap analyses show the practical significance of condition factors for video and measuring board data in our study. For example, for the video shell-height measurements in experiment 2, the first step was to resample n data records (including one video measurement and the corresponding caliper measurement) with replacement from the data in experiment 2.

Sample sizes (n=670 for video and n=344 for measuring boards) were the same as the number of experimental measurements and constituted an upper bound on the true effective sample size because they ignore repeated measurements on the same specimens (Table 2). The effect of using an upper bound estimate for effective sample size was to understate effects of uncertainty in error matrices. Our interest was, however, in a "best case" scenario with relatively large sample sizes. Next, the measurement errors (e.g. video or measuring board minus caliper measure-



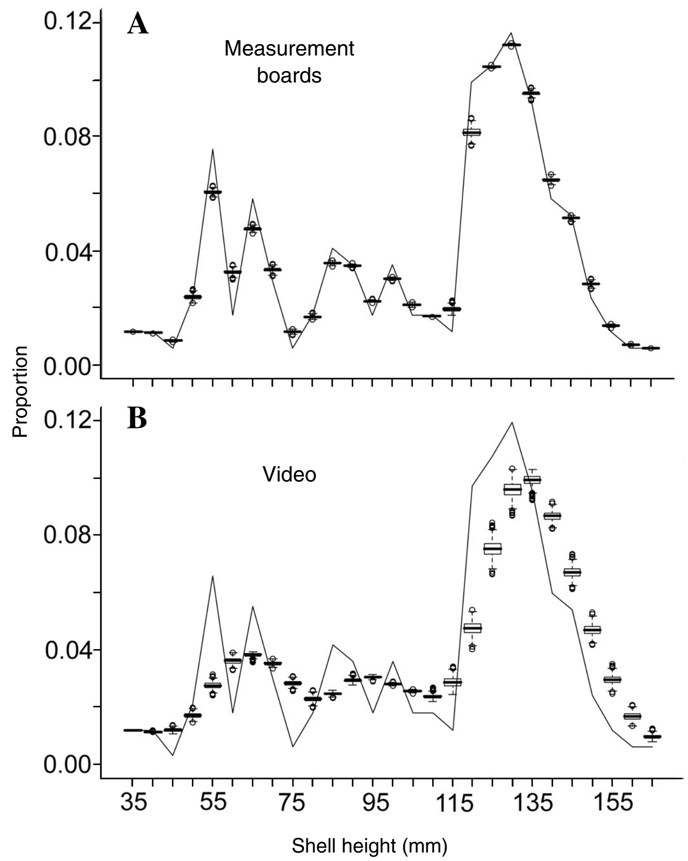
Appendix Figure 1

Boxplots showing bootstrap distributions (1000 iterations) of estimated true shell-height (SH) composition for Atlantic sea scallops (*Placopecten magellanicus*) in experiment 2, based on measurement boards (**A**) and video (**B**) shell-height data. True shell-height compositions were estimated by using bootstrap estimates of the inverse of the measurement error matrix E and Equation A4. The solid line in (**A**) shows the actual caliper-derived shell-height data in the experiment. The solid line is not visible in (**B**) because of the scale of the y-axis.

ments), their mean (bias), and variance were used to calculate the bootstrap measurement error matrix and its inverse. Finally, the original video shell-height composition data used in experiment 2 (expressed as proportions) were then multiplied by the bootstrap inverse matrix (Eq. A4) to remove measurement errors and obtain a bootstrap estimate of the true shell-height composition. There were 1000 bootstrap iterations for both the video and measurement board data. The variability among bootstrap estimates of the true shell-height composition was due entirely to errors in the measurement error matrix E and its inverse E^{-1}.

As expected, based on condition factors (see above) and measurement error statistics (Table 2), bootstrap estimates of true caliper shell-height composition data from video data were highly variable and predicted proportions ranged from –188 to 195 (i.e., outside the feasible range for proportions). Bootstrap estimates from measurement board data resembled the corresponding true caliper measurements. However, the estimated proportions for both measurement methods were often negative and infeasible (Appdx. Fig. 1).

We used a similar bootstrap procedure to evaluate effects of uncertainty in predicted length compositions with measurement errors (Eq. A3 in Appdx. 1), which is the approach used in the CASA model. In this bootstrap analysis, the caliper shell height composition data from experiment 2 were assumed to be true and error matrices were generated by bootstrapping the experimental and video and measuring board data as described above. The sample size was n=172 for both video and measuring boards and the same as the number of individual specimens in experiment 2. This lower bound estimate of the effective sample size was used in order to overstate effects of uncertainty in error matrices. Results indicated that the calculations used in the CASA model for measurement errors were robust to uncertainty about the error matrices and the magnitude of the errors because variability in predicted shell height compositions was relatively minor (Appdx. Fig. 2).



Appendix Figure 2

Bootstrap distributions (1000 iterations) for Atlantic sea scallop (*Placopecten magellanicus*) shell-height data obtained from measurement boards (**A**) and video (**B**), with measurement errors. The solid line shows the actual caliper-derived shell-height data in experiment 2.