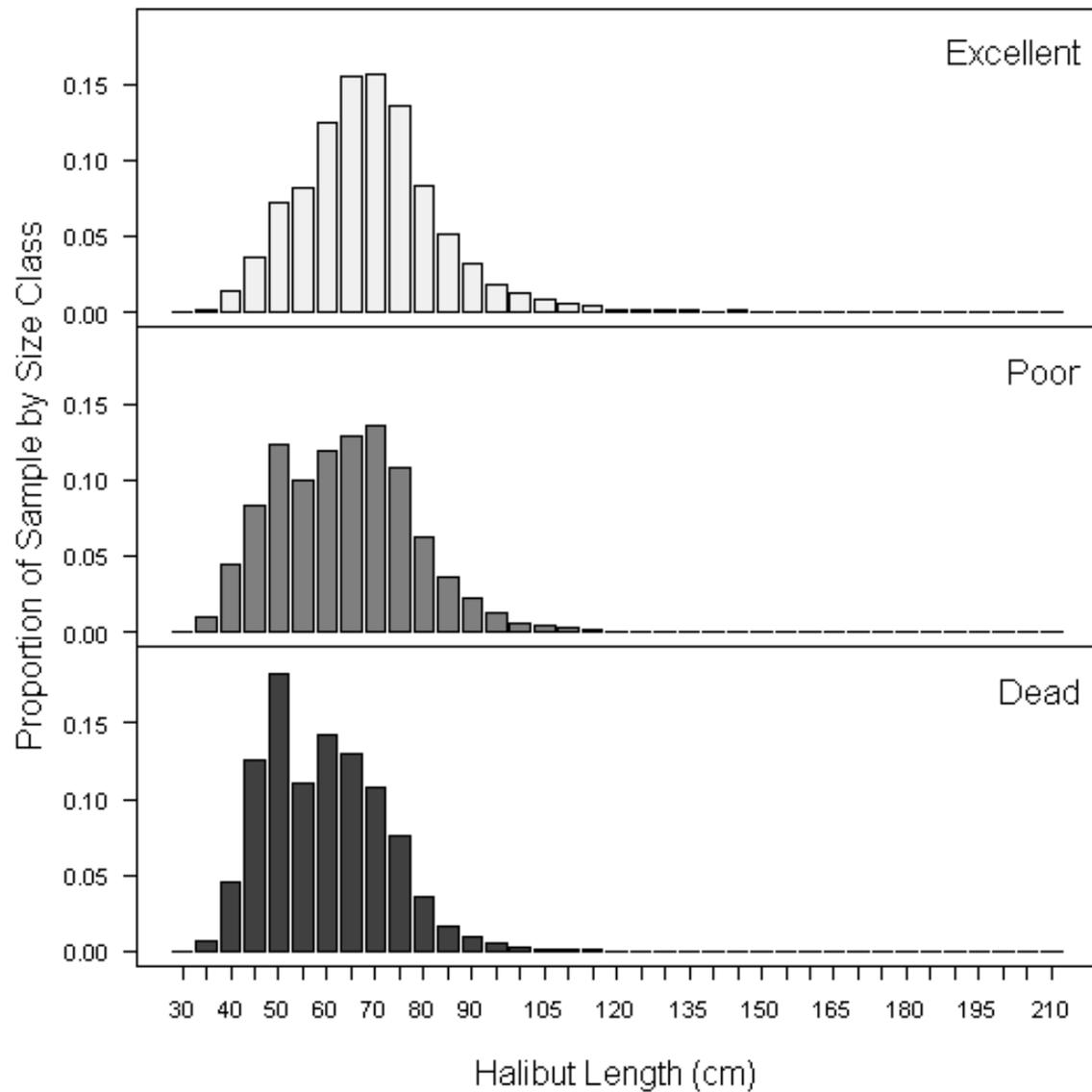
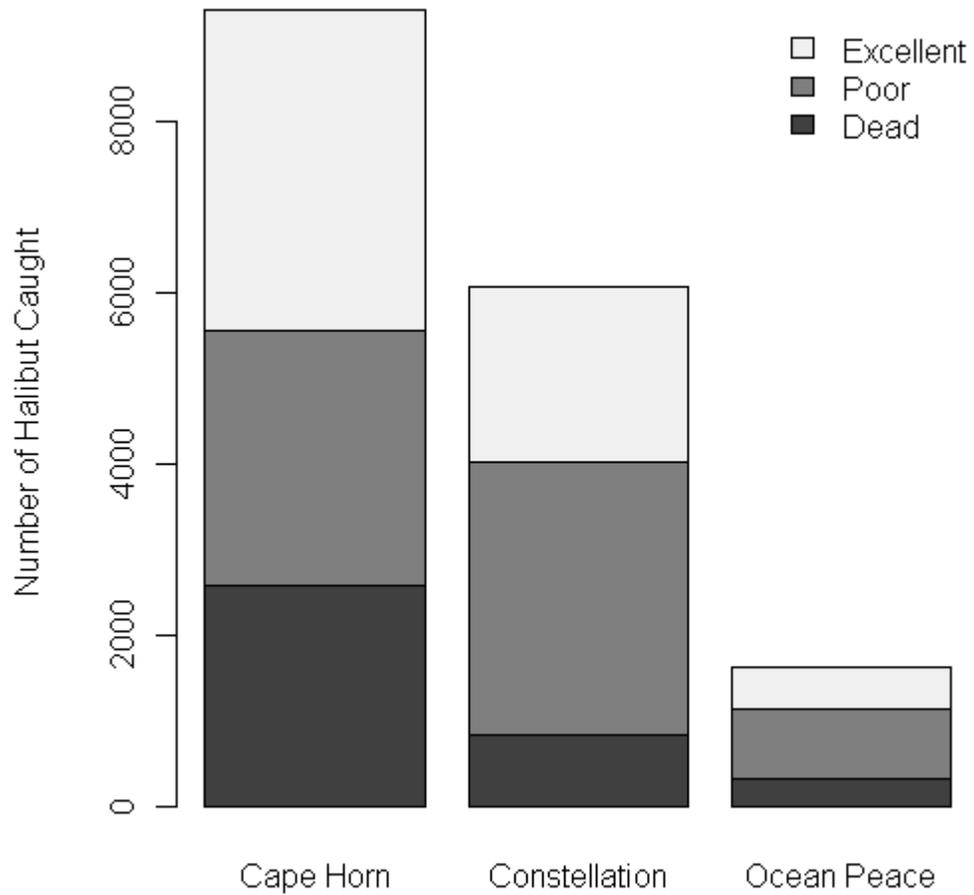


APPENDIX 2

Halibut Size Distribution by Viability



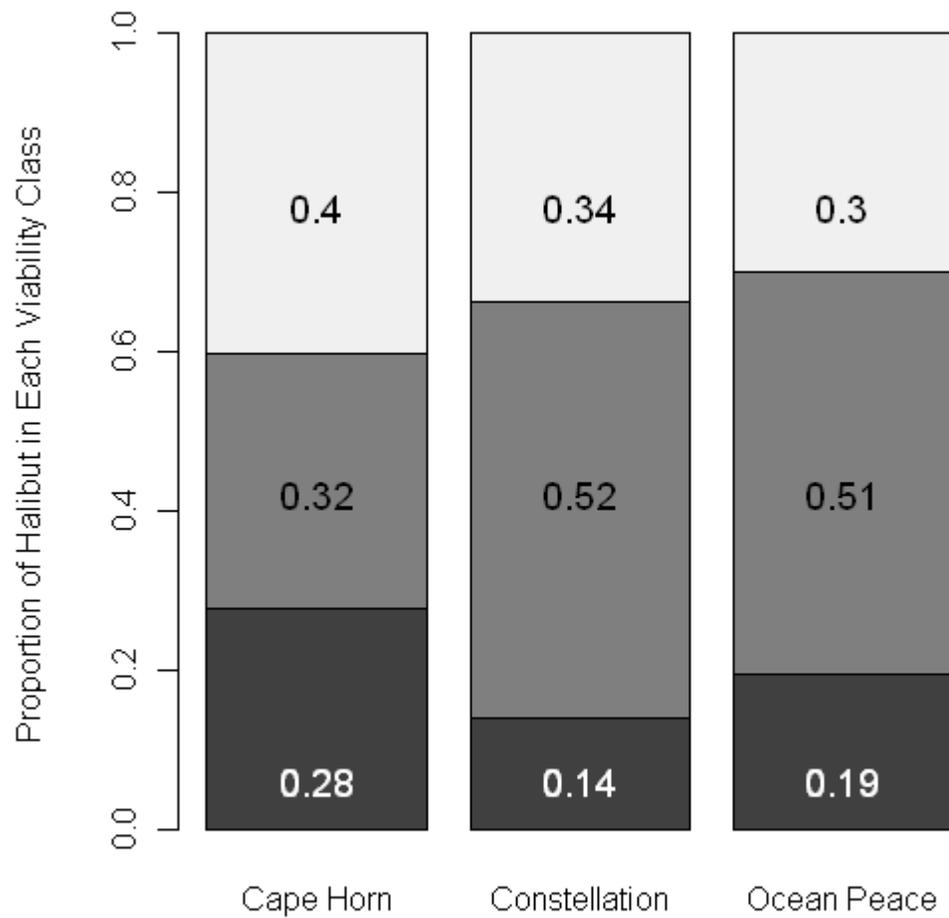


There were:

82 Hauls for Cape Horn,
138 Hauls for Constellation, and
32 Hauls for Ocean Peace.

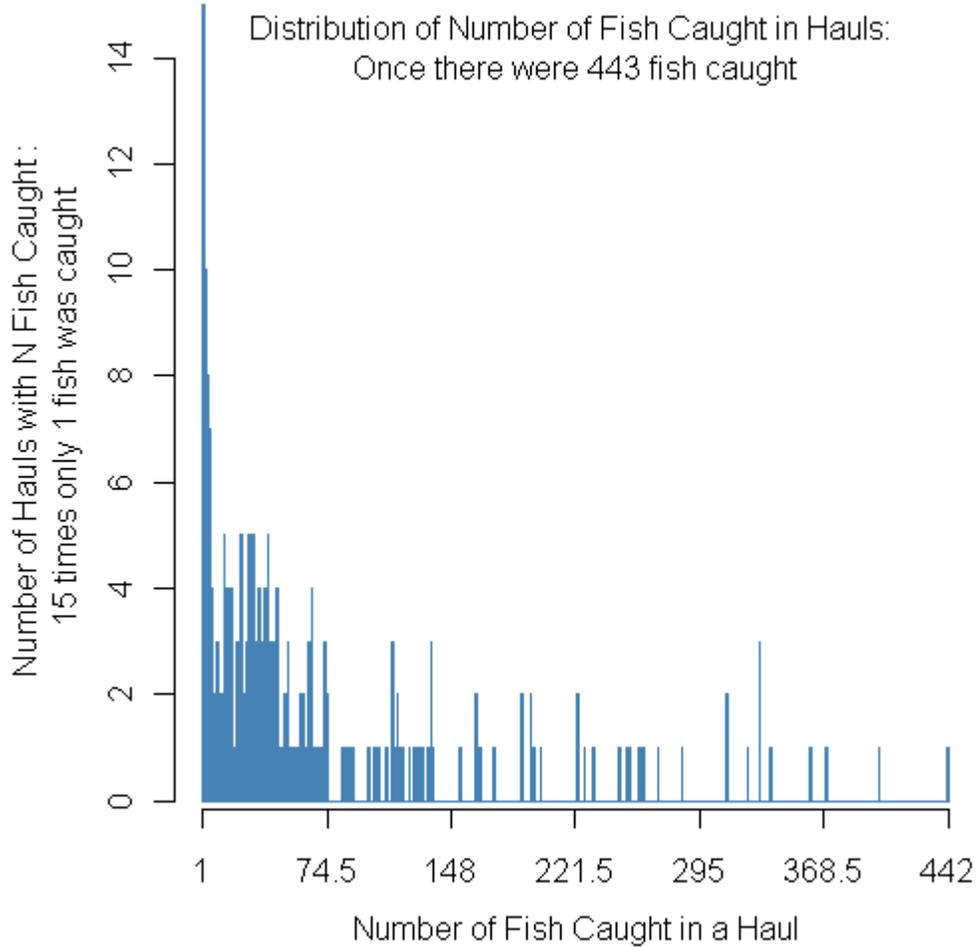
Count by Viability Class and by Vessel:

Viability	Cape Horn	Constellation	Ocean Peace
Excellent	3742	2045	486
Poor	2975	3178	823
Dead	2584	836	314



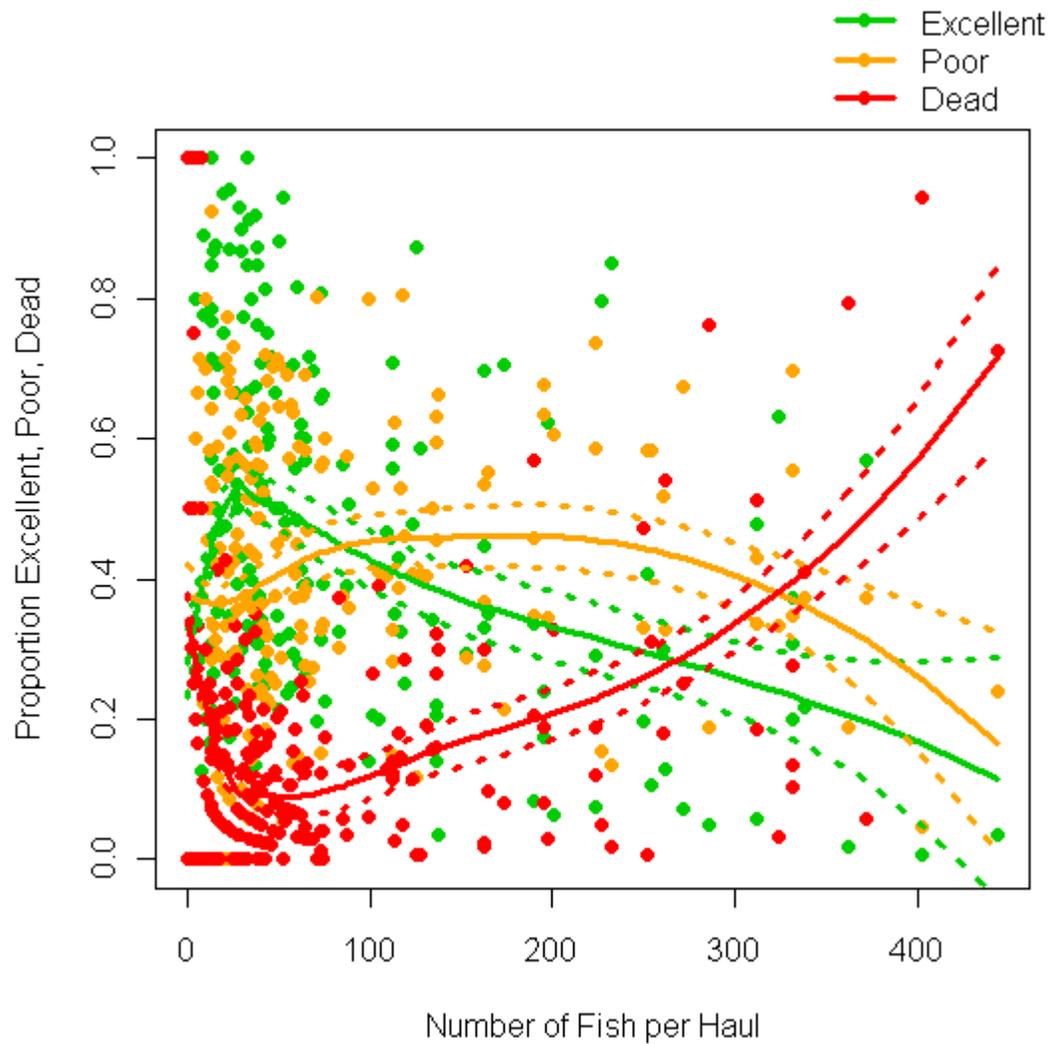
Median Lengths (cm) by Ship and by Viability Class

	Excellent	Poor	Dead
Cape Horn	68	61.0	55
Constellation	67	62.5	62
Ocean Peace	71	71.0	66

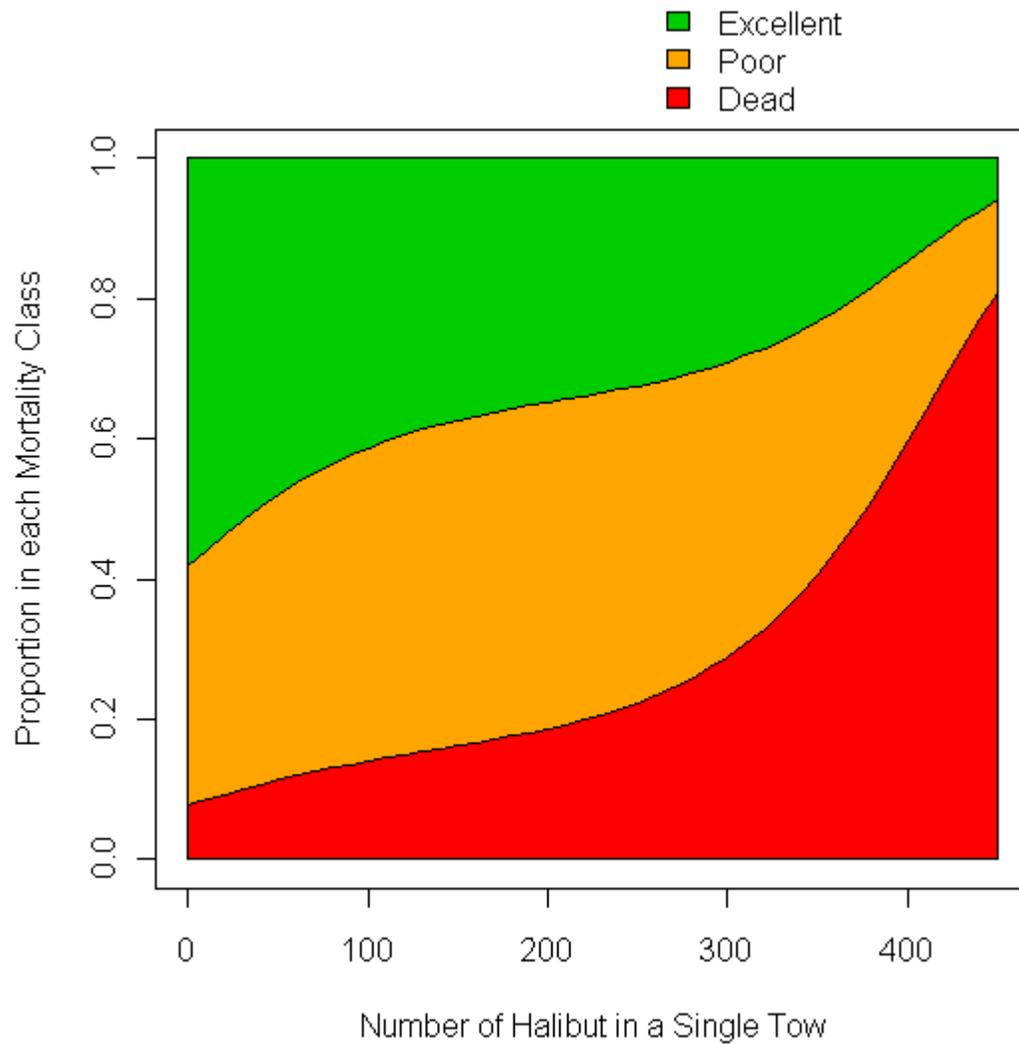


77% of hauls had 75 fish or less.

	number	count	cumul.sum	prop.sum
1	(0,25]	94	94	0.37
2	(25,50]	65	159	0.63
3	(50,75]	34	193	0.77
4	(75,100]	5	198	0.79
5	(100,125]	12	210	0.83
6	(125,150]	8	218	0.87
7	(150,175]	6	224	0.89
8	(175,200]	5	229	0.91
9	(200,225]	3	232	0.92
10	(225,250]	3	235	0.93
11	(250,275]	5	240	0.95
12	(275,300]	1	241	0.96
13	(300,325]	3	244	0.97
14	(325,350]	4	248	0.98
15	(350,375]	2	250	0.99
16	(375,400]	0	250	0.99
17	(400,425]	1	251	1
18	(425,450]	1	252	1



Local Polynomial Regression Fit (Loess):
A quick-and-dirty localised fitting procedure which shows the more halibut in the haul, the higher the proportion with viability rating "Dead", and the less with rating "Excellent".



Multinomial Logit Model using spline-smoothed regression coefficients:
Modeled result is the same as previous page: The more halibut in the haul, the higher the proportion with viability rating "Dead", and the less with rating "Excellent".

Notes on Methods :

I did not include the tows that were not processed on deck immediately. There were 5 tows that were not included because of this, three that were already excluded from the Excel Deck_Hbt worksheets, and 2 additional ones that I selected out (Constellation Haul 71, Ocean Peace Haul 33).

I analysed the halibut mortality class data as multinomial logit models, which are regression models that generalize the more familiar logistic regression by allowing more than two discrete outcomes (for us this allows for outcomes "Excellent", "Poor", "Dead"). This modeling approach is appropriate as it uses the logit transform to transform probabilities, such that the sum of the probabilities of the three mortality classes sums to 1, the confidence intervals are restrained between 0 and 1, and their variances follow a known functional form.

I wanted to allow the fitted regressions to be more flexible than assuming a linear relationship between the independent and dependent variables. I therefore used a regression spline (with 3 degrees of freedom) to model the relationships on the following graphs. This makes it impossible to interpret the coefficients directly, but shows well the relationships between the variables. This helps me think about how to structure the Power Analysis simulation, and I have provided simple interpretations of results are beside the figures in textboxes.

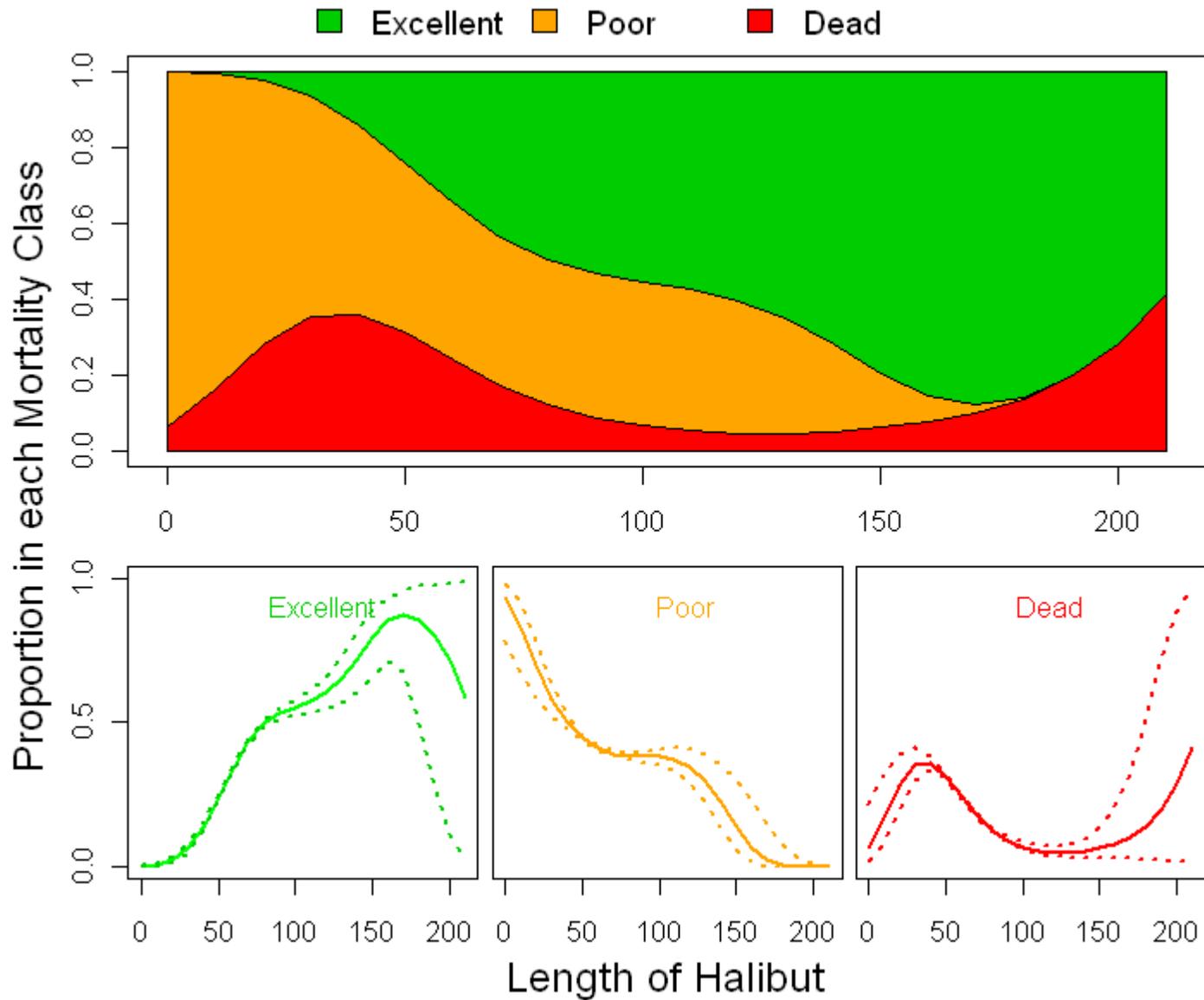
With respect to the power analysis, I am currently thinking there are 2 levels to this problem:

Level 1:

The number of halibut caught in each tow is important in resulting viability of halibut in the tow. So we need to make sure that we have enough tows sampled to capture this variability. As we discussed, this would be achieved through a simple randomized sampling design.

Level 2:

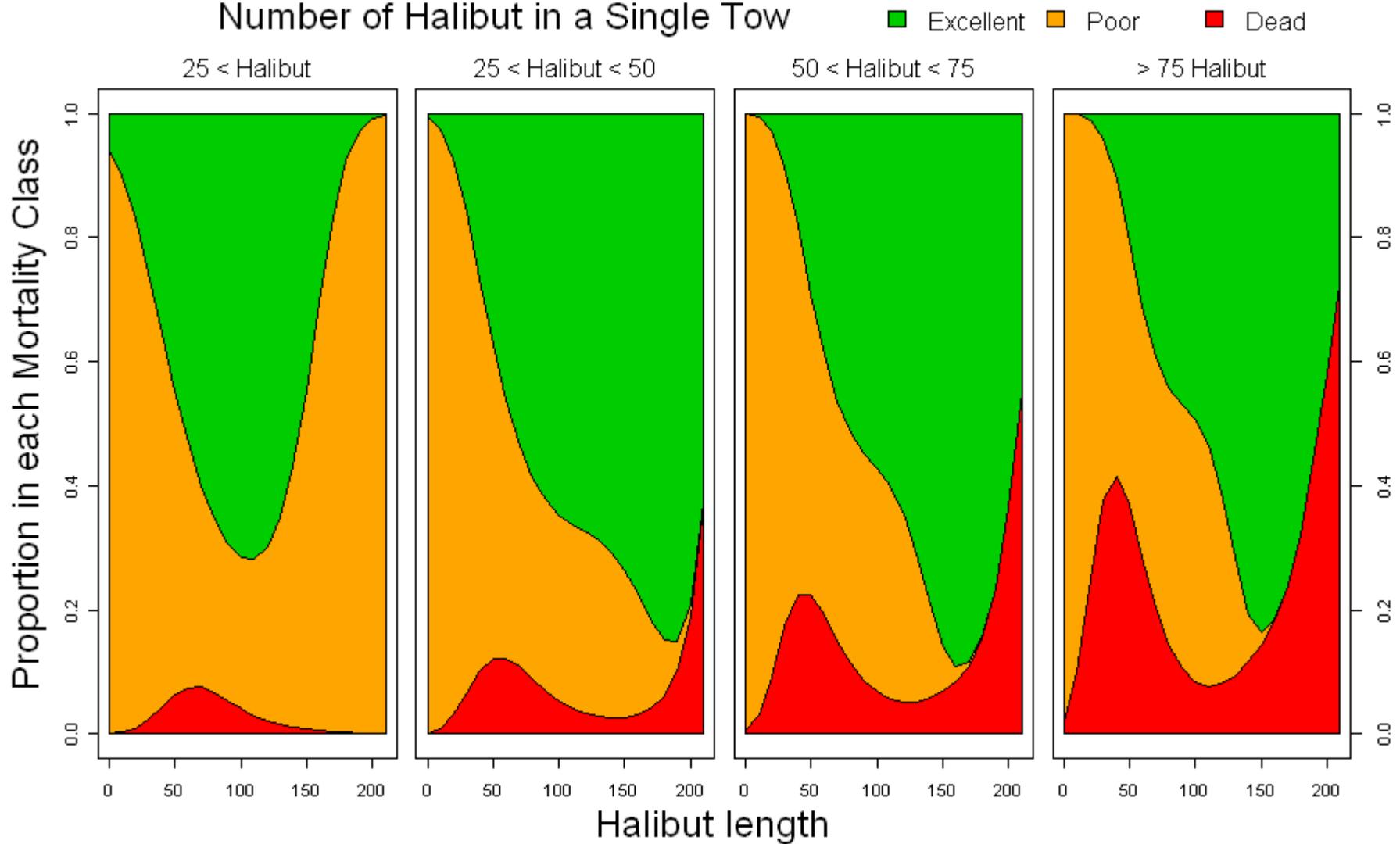
The subsample within a tow is to characterize the probability of the mortality classifications within a certain level of precision. For this we can use the multinomial distribution for the 3 mortality classes : Dead, Poor, Excellent. As we discussed, for this part of the sampling protocol, I think we will want to use a systematic sampling approach, because it will not be known how many halibut are in the net until the contents are processed. Therefore it makes sense to take every one, then every second one, etc. through a simulation, and then i can just quantify the loss of precision by lowering the sampling intensity by one each time. This limits our theoretical ability to get an estimate of error (technically, the sample has to be random to permit estimating errors), but we'll get the empirical estimate of the sampling error through the simulation. I think that will still be good as long as the conditions seen in this years data, repeat themselves next year.



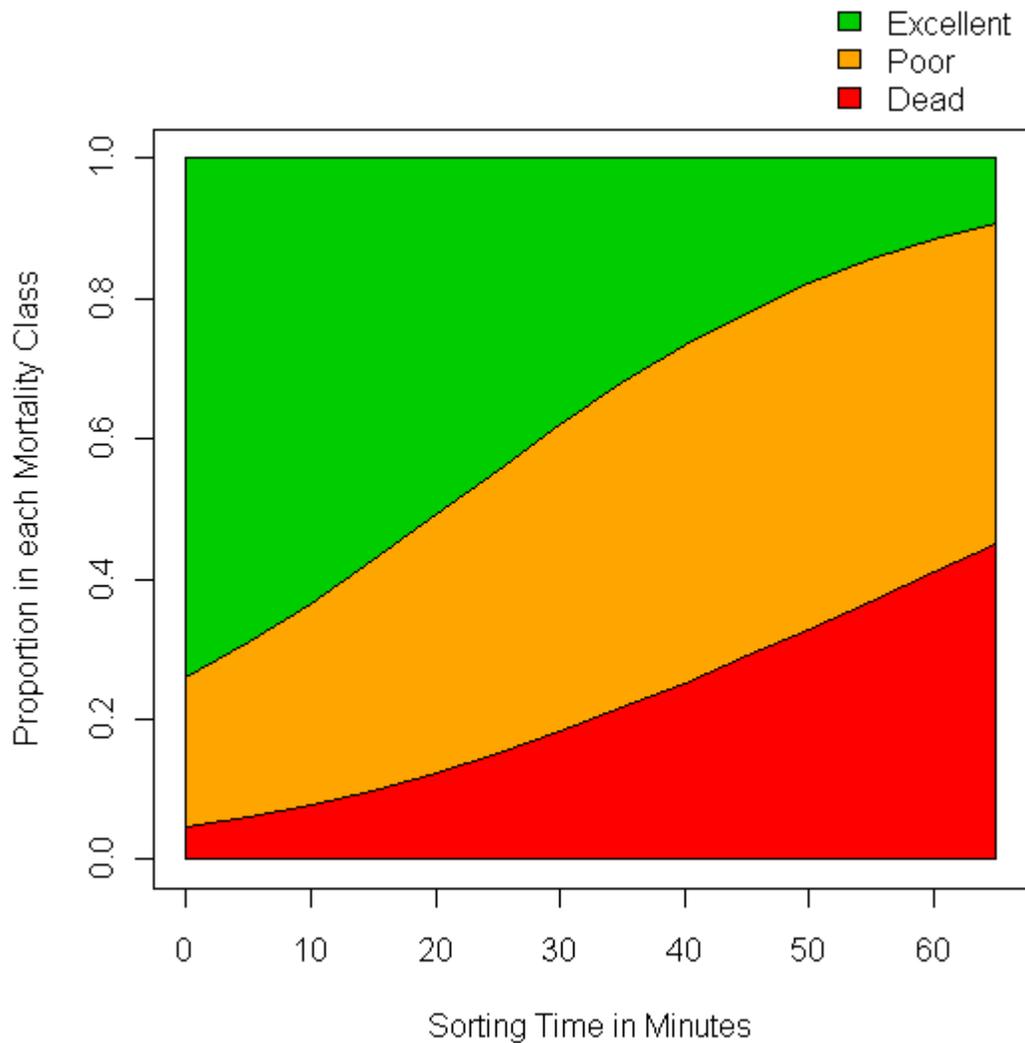
Multinomial Logit Model using spline-smoothed regression coefficients: Modeled result shows the longer the halibut, the higher the probability of having viability rating "Excellent", and the less with rating "Dead". Large confidence intervals at large halibut sizes, make it unreliable to interpret.

The lower 3 panels show the confidence intervals around the probabilities. Note that the top panel is a cumulative probability plot, whereas these show the probability for each category separately.

Number of Halibut in a Single Tow



Multinomial Logit Model using spline-smoothed regression coefficients: Result shows the greater the number of halibut in the tow, the higher the probability of having viability rating "Dead" (the more red there is on the plot). Plot also shows the previous result of higher proportion of "Excellent" ratings with larger halibut.



**Multinomial Logit Model
using spline-smoothed
regression coefficients:**

Modeled result shows the longer the sorting time, the higher the probability of having viability rating "Poor" or "Dead", and the less with rating "Excellent".

SORTING TIME (this page) AND NUMBER OF HALIBUT (pages 5,6) CAUGHT IS SIGNIFICANTLY CORRELATED, and thus are describing the similar phenomenon - if you process quickly, halibut have a better chance of being "Excellent".

Pearson's product-moment correlation

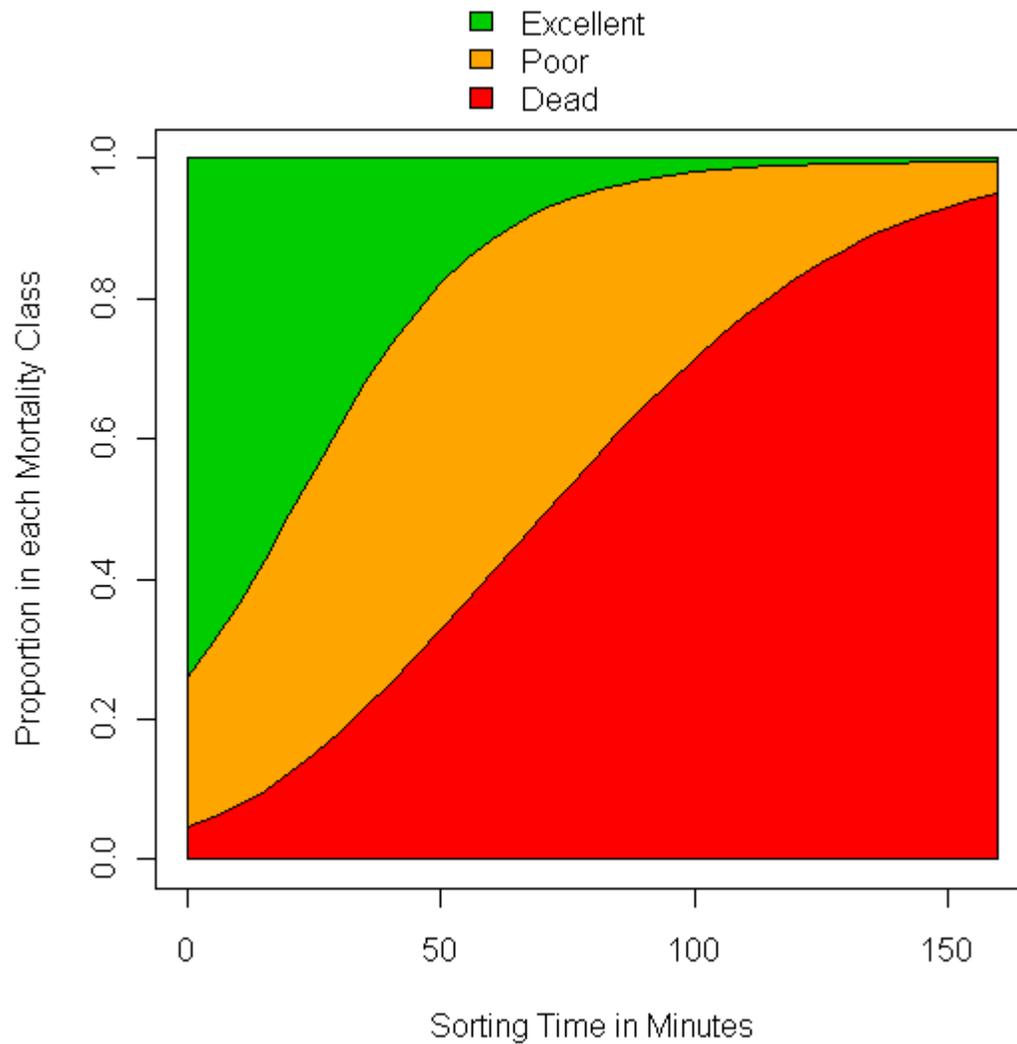
cor
0.5619777
t = 10.6994,
df = 248,
p-value < 2.2e-16

95 percent confidence interval:

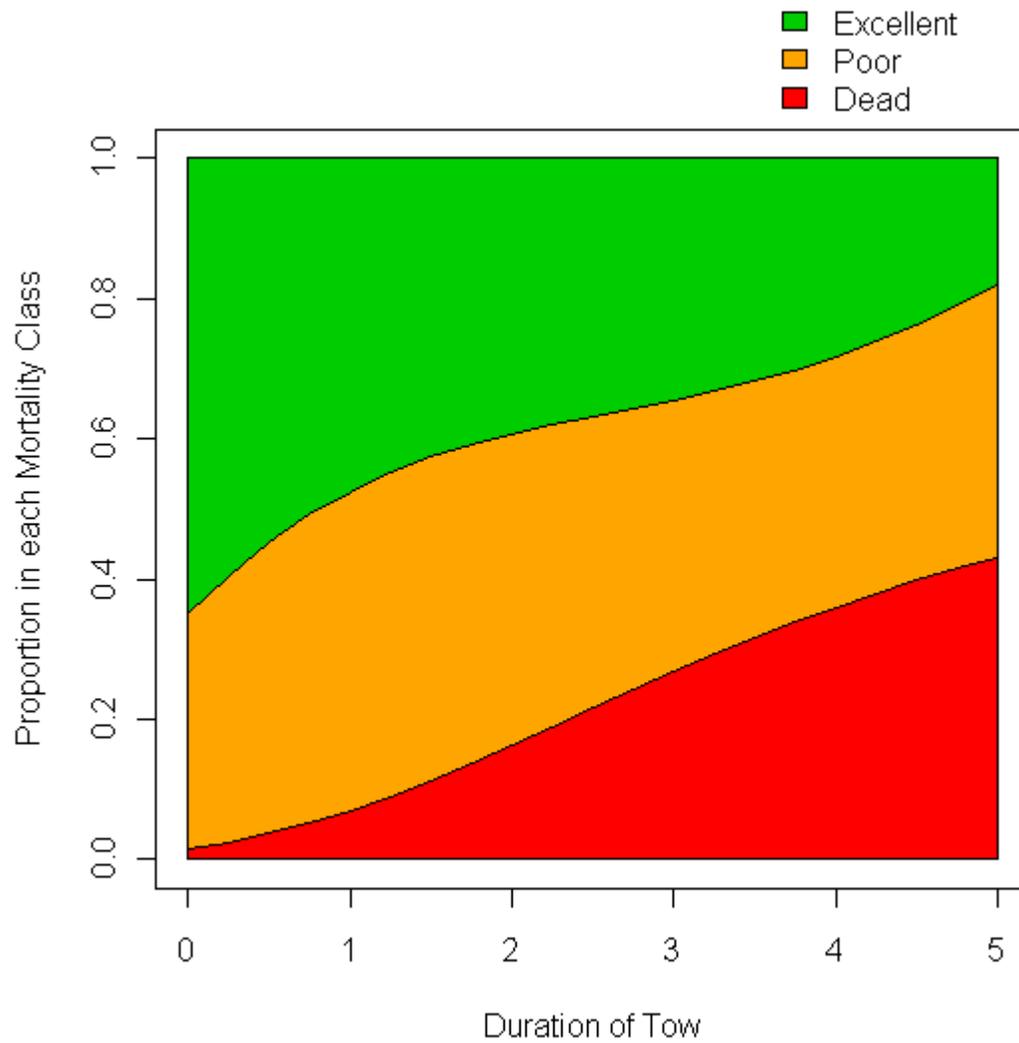
0.4707314 0.6413294

sample estimates:

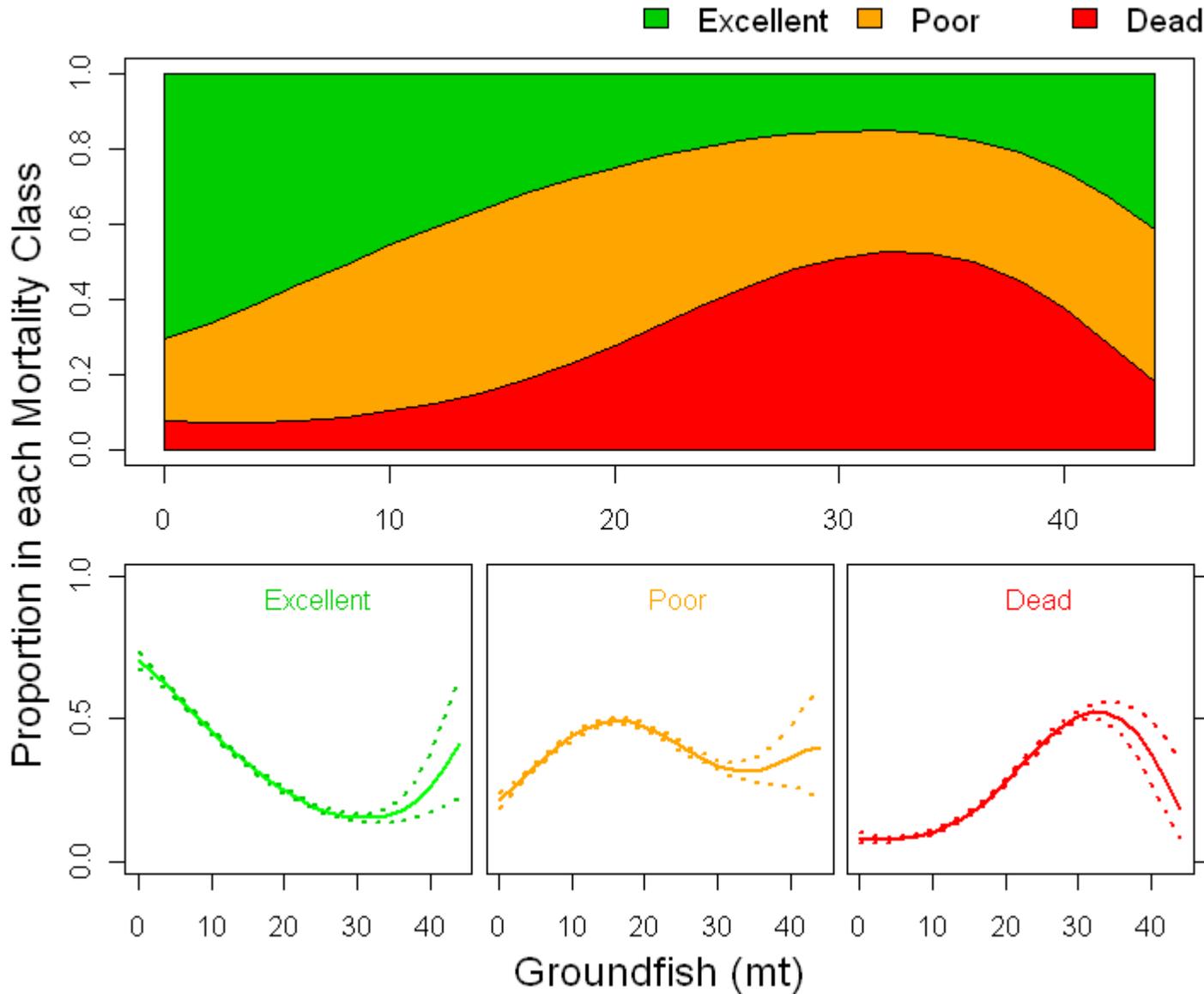
(cor=0.5275942, p < 2.2e-16 if I don't include the 157 minute sorting time)



Same as previous plot,
this one just includes the
outlier of 157 minutes of
sorting time. (All sorting
times except 1 took 65
minutes or less)



Multinomial Logit Model using spline-smoothed regression coefficients:
Modeled result shows the longer the duration of the tow, the higher the probability of having viability rating "Poor" or "Dead", and the less with rating "Excellent".



Multinomial Logit Model using spline-smoothed regression coefficients:
 Suggests an increase in mortality up to a plateau of 35 mt. Lots of error at high biomass of groundfish, so not sure that the region between 35 and 40mt can be interpreted as a decline in "Dead" frequencies. .

Methods for My Investigations into Predictive Error and Bias.

I ran two different power analyses, which were organized around what we talked about - I wanted to highlight how you might have sampled the halibut caught last year differently. The two figures on the following pages, show what the tradeoffs are between the 2 different approaches to minimizing effort.

The first power analysis (and the left hand panel of both plots) looks at the error and bias i.e. if you sampled every single tow, but sampled the halibut at a **systematic** and predetermined frequency. **The second power analysis (and the right hand panel of both plots)** looks at the error and bias if you sampled every single halibut, but **randomly** sampled a subset of the tows.

Percent error on estimates (page 15,16) is defined here as :

the width of the subsample's 95% confidence interval/prediction probabilities from the full data set. In the first power analysis the modeled confidence interval was used, in the second incident a randomization routine was used, and the 95% confidence intervals were derived by taking the 2.5th, and 97.5th ordered predictions. The interpretation of both methods to obtain estimates of the percent error are the same. This interpretation is that if we took 100 samples from the same distribution as ours, 95 of those predictions would fall within this limit. The denominator is to give it a relative error scale with respect to the "true" (full sample) prediction probabilities.

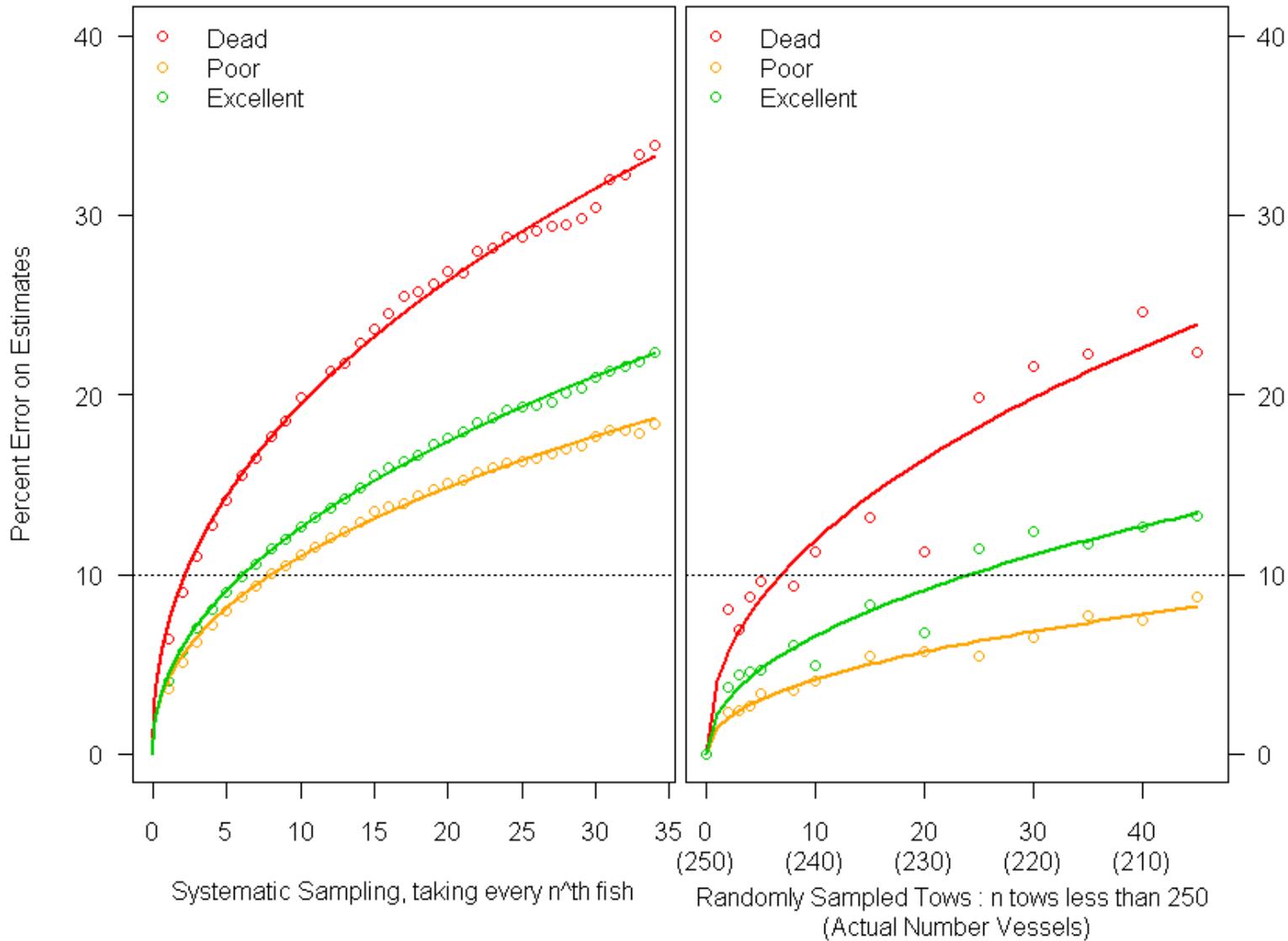
All panels show the 10% error or precision (page 15) and bias (page 16) cutoffs as mentioned in one of our first discussions. These 10% cutoff lines show where the relative error/bias is either > or < 10% of the "true" prediction probability.

Prediction Bias (page 17) is defined as the difference between the subsampled estimate of proportion, and the "true" (full sample) prediction probabilities. Again, bias was scaled to a relative measure by selecting the denominator to be the prediction probabilities from the full data set. This was not rerun as the bias would automatically increase if we took the first halibut rather than systematically sampled ones as the desired benefit of decreasing processing time is to keep more halibut "**excellent**".

A quick note on the "true" precisions against which bias is measured is that it is assumed to be "correct". This means then that the sampling protocol in place last year would have given a bias-free estimate of the proportions in each Viability Category. By the preliminary analyses (pages 8-13), and by the figure on page 16, we know that bigger hauls have higher mortality rates, so there are underlying problems with estimating overall halibut Viability without considering the effect of haul size on these overall estimates of viability.

I have provided simple interpretations of results beside the figures on the next 2 pages.

Percent Prediction Error for Each Viability Category

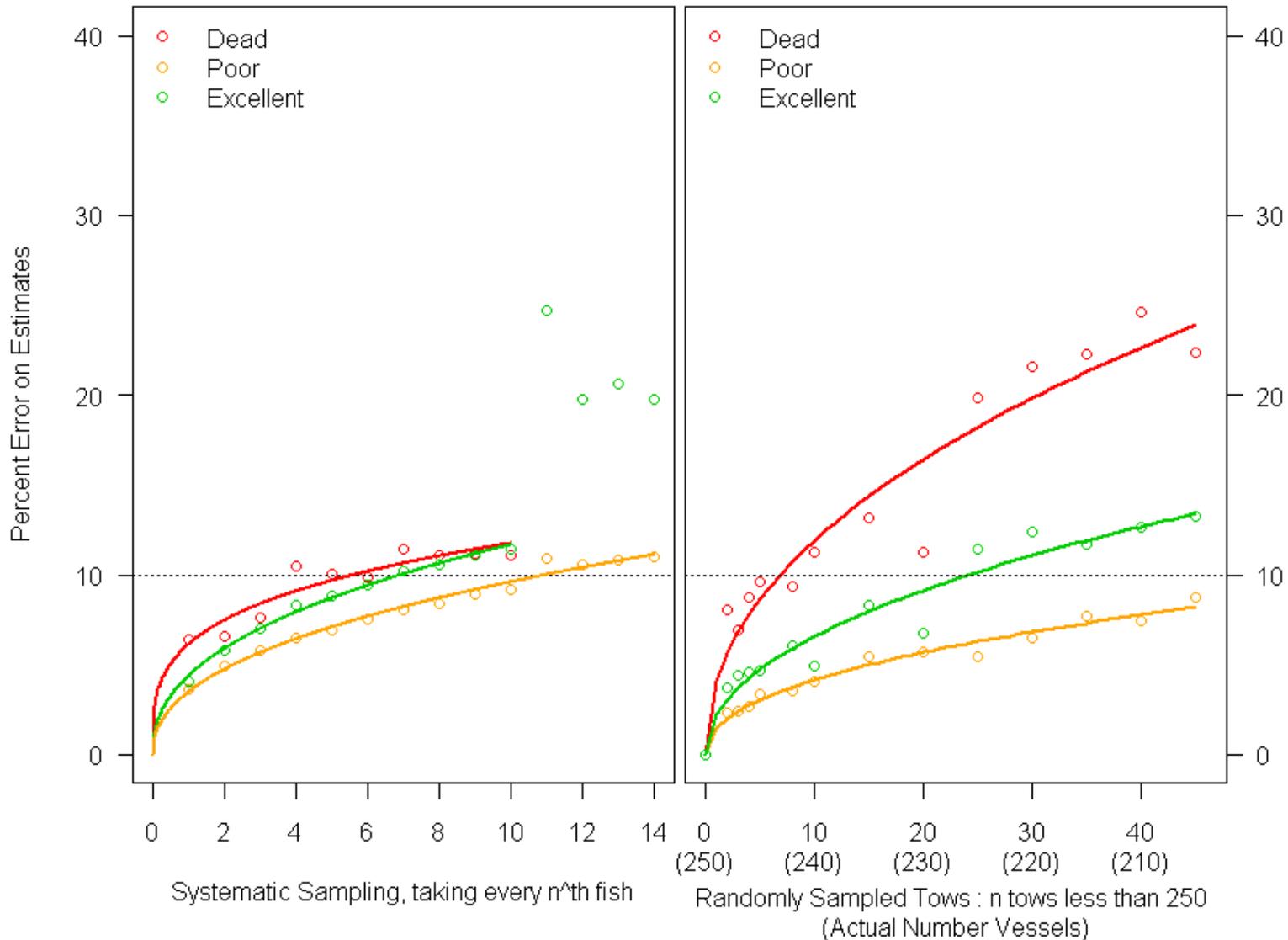


Prediction Errors

To be within 10% error rate of systematically subsetting within a haul (left panel), but sampling all hauls, you can drop the frequency to every 2nd halibut. If the benchmark halibut Viability category was "Excellent", you could safely subsample every 6th halibut.

In the random sampling protocol (right panel), you need to sample all except 7 hauls to be within the 10% error rate of the most error prone Viability category ("Dead"). If you were happy to be within 10% error for a benchmark Viability of "Excellent", you could not sample on 20 of the tows.

Percent Prediction Error for Each Viability Category



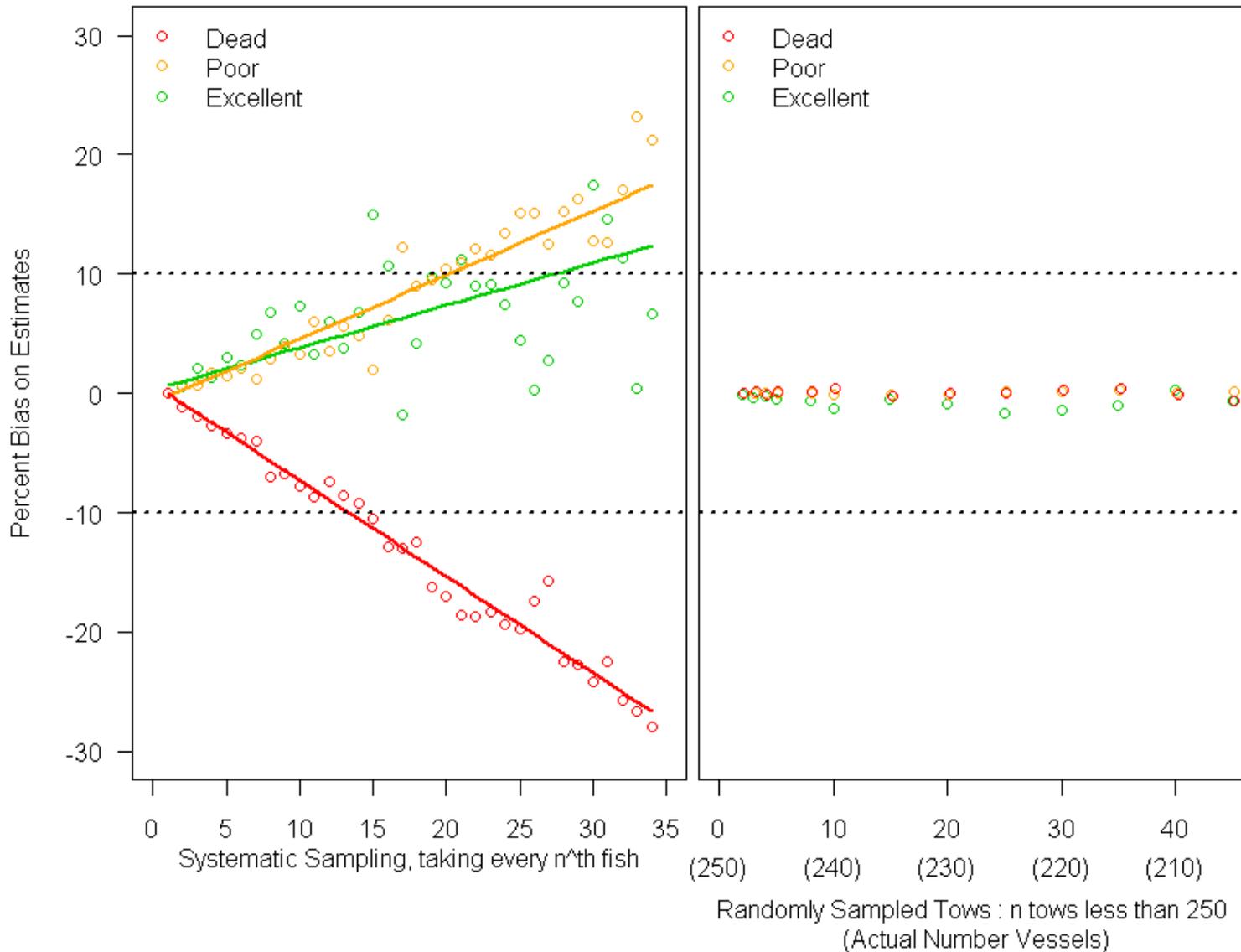
Redoing Power Analysis to Evaluate the Prediction Errors

Redoing the selection of halibut from all hauls, but selected the number of fish as before using the systematic sampling protocol, but took the first fish's viability score, rather than the one associated with the nth halibut. This stabilized the error associated with Viability Category (compare left panels of this page with previous page (pg 15)).

To be within 10% error, you can drop the frequency to every 5th halibut. If you drop below every 10th halibut, the error rate jumps to intolerable levels (see green dots at n>10 for "Excellent", dots are off the upper scale for Category "Dead").

The random sampling protocol (right panel), is the same as previous page, except I reran the simulation.

Percent Prediction Bias for Each Viability Category



Subsampling Bias

Systematic Sampling (left panel) has considerable bias associated with it, Random Sampling (right panel) does not. The reason for this is that as you take less and less fish, you are mostly taking only 1 halibut (the first halibut) per tow.

e.g. see page 4 graph

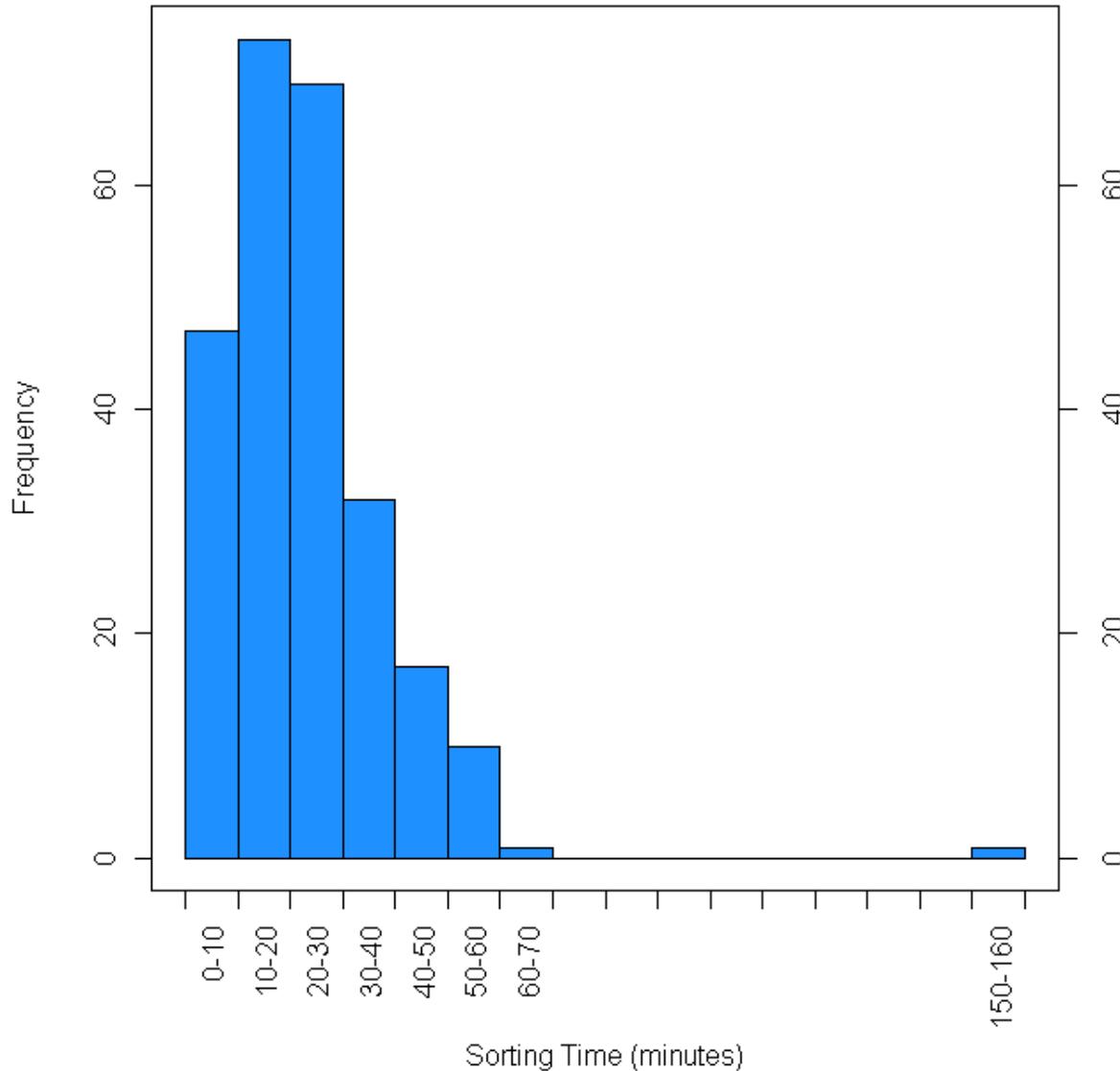
Number halibut Per Haul	Number Vessels
1	15
2	10
3	8
4	7
5	4

Therefore by the time the systematic sampling protocol gets to every n=5th halibut, there are 39/250 hauls that are only selecting the first halibut. (By 13 halibut, 62 tows are only picking the first halibut)

To keep prediction bias < 10%, you must sample at least every 13th.

Below is a figure, and table that show the Number (Count) of Tows that were sorted in each of 10 minute increments, e.g. there were 47 tows processed in 10 minutes or less. (Each bar in the histogram is a row in the table.)

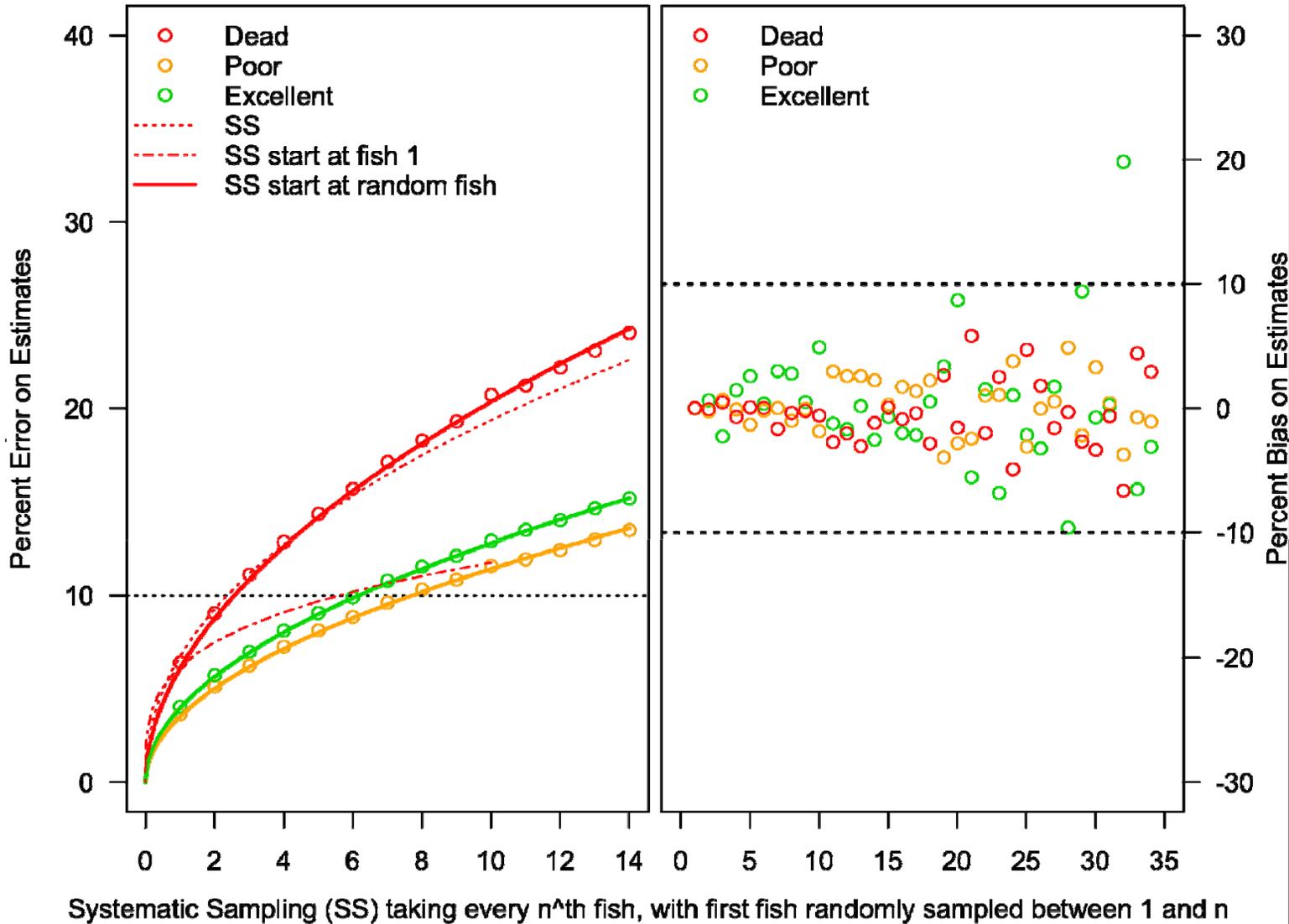
Distribution of Sorting Times



```
> haul.time
  Sorting.Time.Window Count.of.Tows
1      (0,10]           47
2      (10,20]          73
3      (20,30]          69
4      (30,40]          32
5      (40,50]          17
6      (50,60]          10
7      (60,70]           1
8      (70,80]           0
9      (80,90]           0
10     (90,100]          0
11     (100,110]         0
12     (110,120]         0
13     (120,130]         0
14     (130,140]         0
15     (140,150]         0
16     (150,160]         1
```

Prediction Error for Each Viability Category

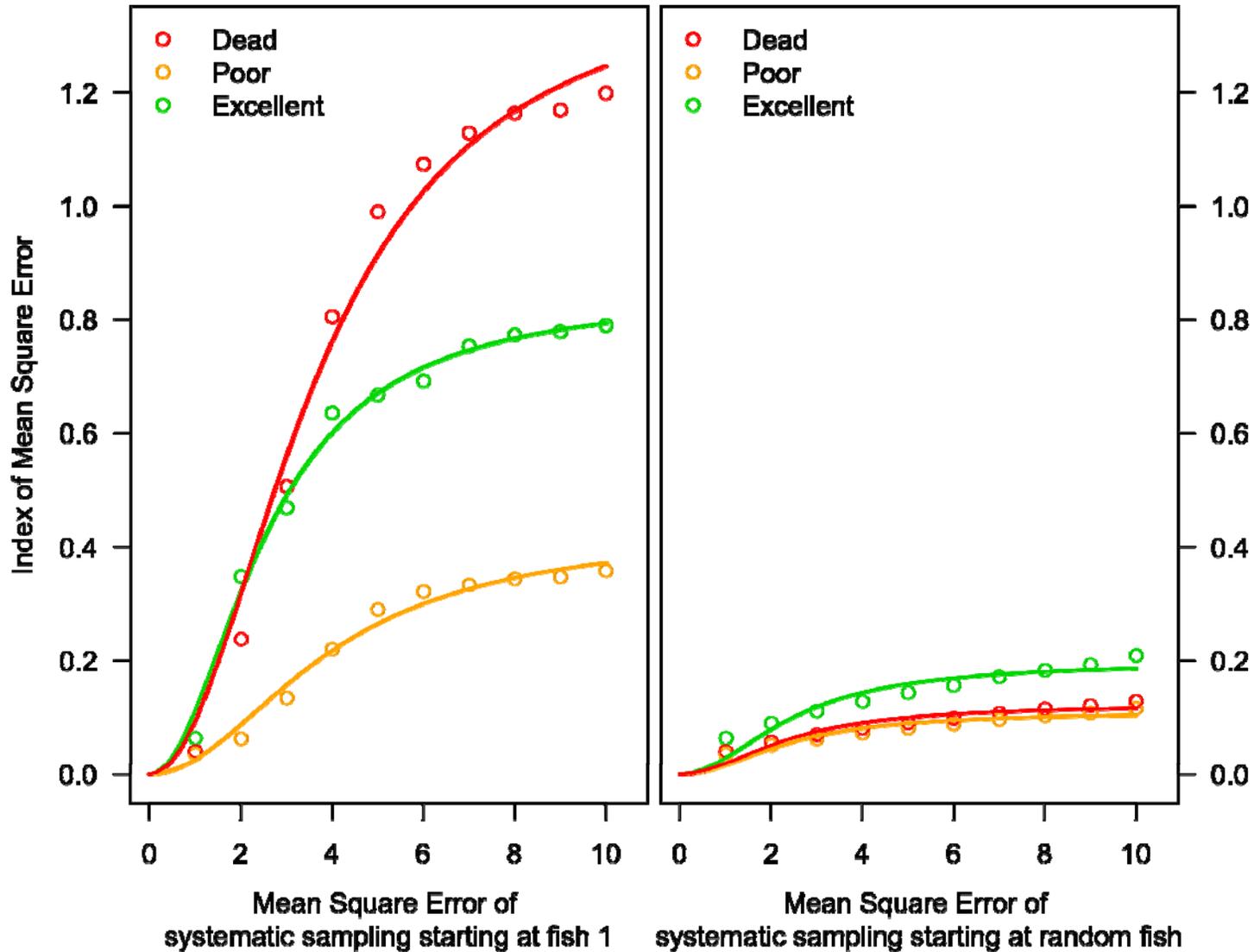
Prediction Bias for Each Viability Category



The left panel shows the percent error in the prediction, the right panel shows percent bias. I have overlaid in the left panel the error associated with the worst category ("Dead") for the two methods already evaluated. This percentage of Ray's method is similar in magnitude to the SS method (left panel, page 16), and higher than the error associated with SS starting with the first fish (left panel, page 17).

This modified design suffers from loss of sample size, similar to the loss in the SS method. The gain in this method is seen in the right panel where there is no systematic pattern to the bias, besides a moderate increase with a decrease in sampling intensity.

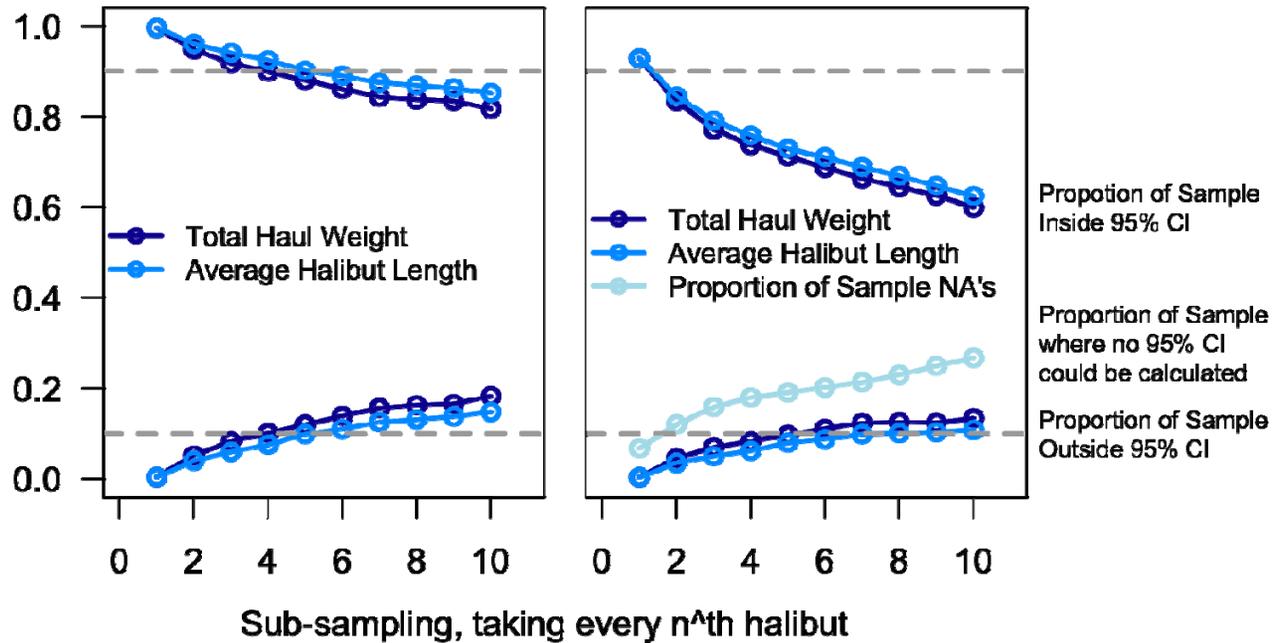
Relative Error of 2 Sampling Methods for Halibut



I used an index of Mean Square Error (MSE) to compare between 2 of the sampling methods for halibut. MSE is usually the variance + bias² of the estimator. I used this formula but used the percent prediction error (left panels, pg 16, 19), and percent prediction bias (left panel, page 16; right panel, pg 19), to give a relative measure of MSE for these 2 methods. This figure suggests that the gain in precision using the SS start at fish 1, is swamped by the magnitude of the bias associated with always selecting the first fish in the series.

It would seem that the systematically sampling that starts with a random fish, has lower overall error.

**Proportion of Time that the Full Dataset's
Total Haul Weight
fell inside/outside of the
Sub-Sample's 95% Confidence Interval**



The left and right panels in the figure above show almost the same thing, the difference is in the denominator used to calculate the proportion Inside and Outside of the 95% CI. On the left, I exclude the NA's from the denominator, on the right, I took the total number of hauls (252) as the denominator, and includes as a significant the counts of NA's in which no estimate of variance was possible (and therefore no CI's either). 10% error is shown as the dashed lines at 0.1, and 0.9 on both graphs.

Of those samples that we can calculate a CI for, it is possible to sample every 4th fish and get no more than 10% of hauls misspecifying the total haul weight (See next page for exact results). Length of halibut has one less source of error associated with the sampling, and every 5th fish will give equivalent confidence.

Brief Methods: I calculated the mean, standard error and confidence intervals for the estimated "total haul weight" of halibut for each of the 252 hauls. This was done as discussed by calculating the mean halibut weight, and then multiplying by the number of halibut in the haul. I set up the simulation to count the number of times the total haul weight of the full dataset fell within the 95% confidence interval of mean haul weight of the subsampled-at-every-nth fish haul sample. This was redone 100 times for each of the 1 through 10 subsampling intensities to ensure the results we obtained were representative of the expected values for these quantities.

```
> temp[,c(1:3)]
Tot.Wt.Out Tot.Wt.In Tot.Wt.NA
1          1.00    234.00    17.00
2         11.51    210.15    30.34
3         17.59    194.45    39.96
4         20.85    185.83    45.32
5         24.45    179.48    48.07
6         27.94    173.12    50.94
7         31.06    167.17    53.77
8         31.59    162.45    57.96
9         31.30    157.77    62.93
10        33.76    150.97    67.27
```

For a vertical slice through 6 on the x-axis of left panel of figure on page 21:

20.85/206.68 = 10.0% ; 20.85/252 = 8.3%
 185.83/206.68 = 90.0% ; 185.83/252 = 73.7%
 45.32/252 = 18.0%

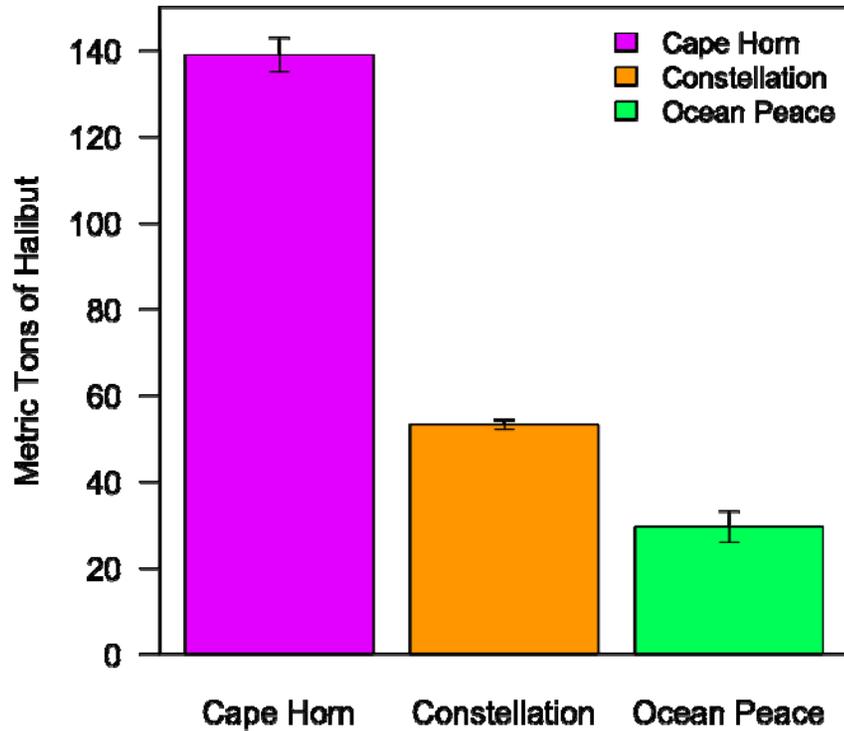
If we were to subsample every 4th halibut, only 10% of all valid trials, would not include the mean of the full dataset from which we've subsampled.

As we subsample and lower and lower frequencies, we get more samples from which no CI can be derived. By sampling every 6th fish, we have over 20% of hauls that have 0 or 1 fish length measured.

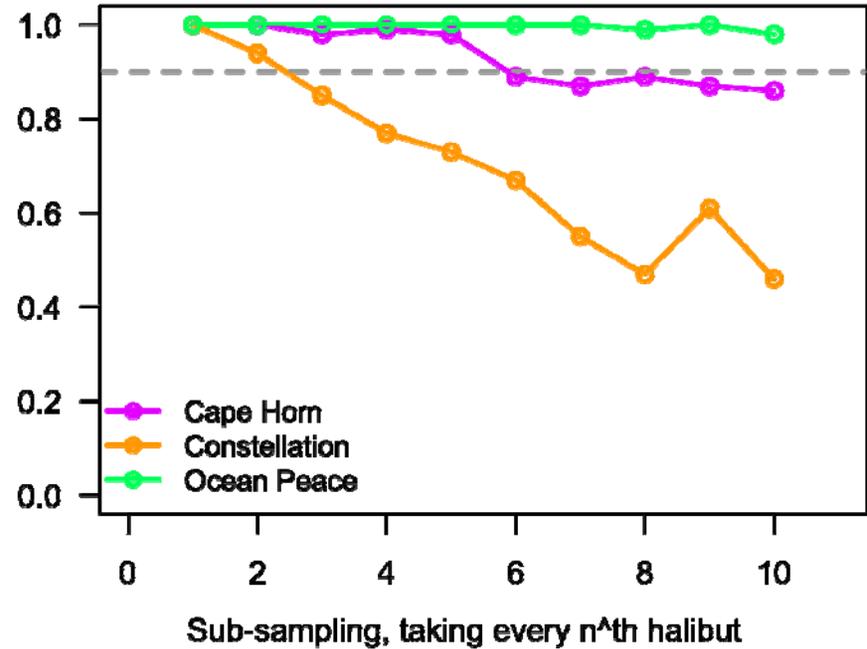
The same protocol was followed for the "average halibut length" analysis.

```
> temp[,c(4:6)]
Leng.Out Leng.In Leng.NA
1          1.00    234.00    17.00
2          8.83    212.83    30.34
3         12.65    199.39    39.96
4         15.62    191.06    45.32
5         20.16    183.77    48.07
6         22.11    178.95    50.94
7         24.86    173.37    53.77
8         25.46    168.58    57.96
9         25.98    163.09    62.93
10        27.33    157.40    67.27
```

Total Haul Weights and Confidence Intervals of Each of the Three Vessels

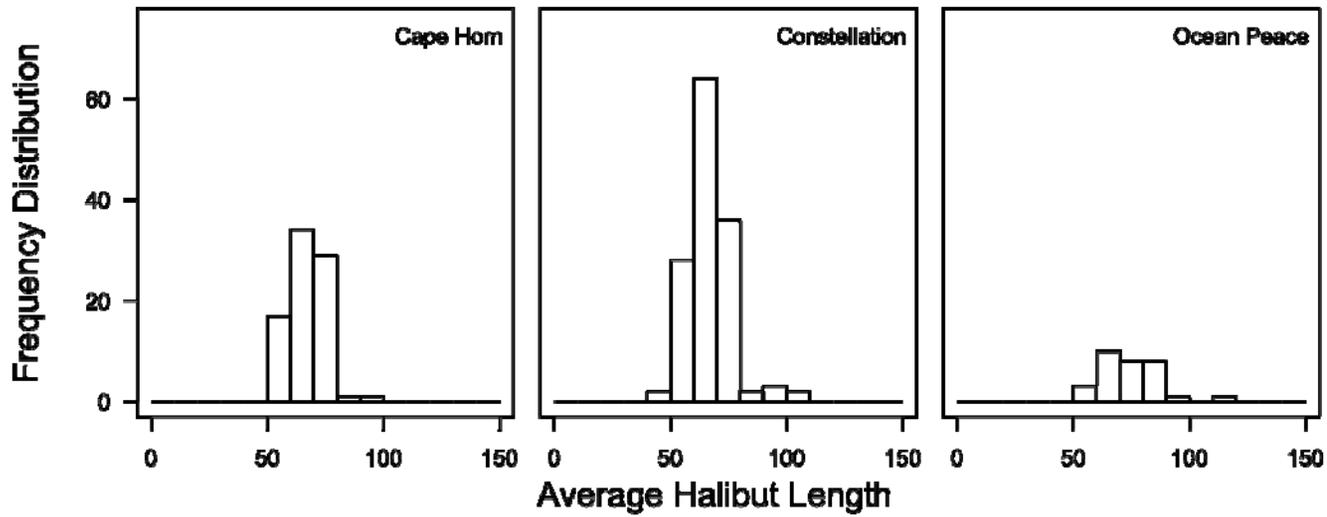


Proportion of Time that the Sub-Sampled Dataset's Total Haul Weight by Vessel fell inside/outside of the Total Haul Weight's 95% Confidence Interval

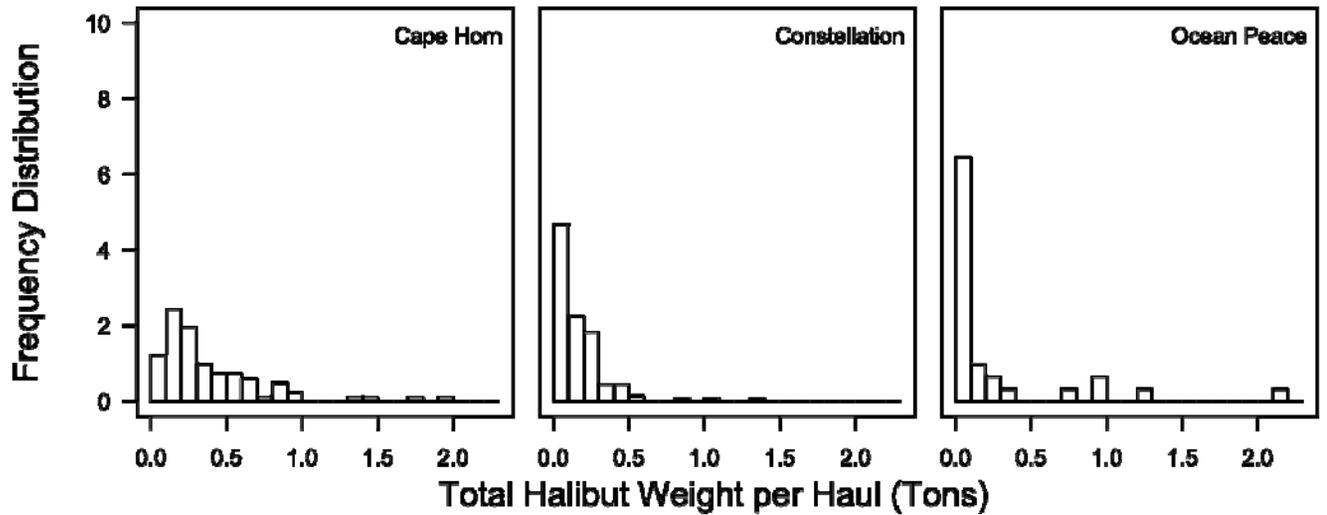


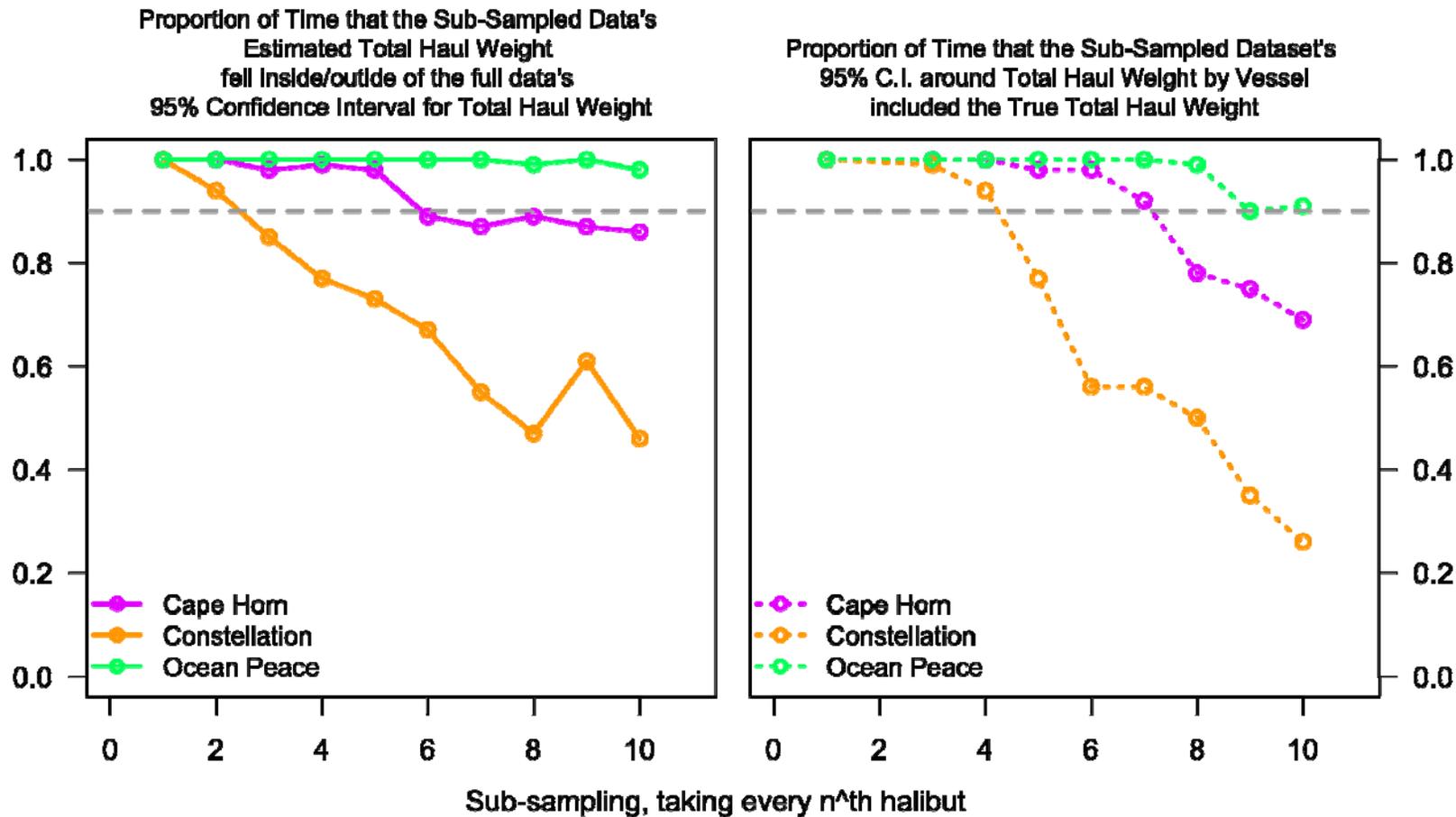
Brief Methods: To determine what the effect of subsampling the hauls would be on the accuracy of the estimate of total haul weight by vessel, I ran a simulation in which I counted the proportion of subsampled total haul size estimates, that were inside the confidence interval of the full dataset. The results are shown in the righthand figure above.

Frequency Distribution (Counts) by Vessel of Average Lengths



Frequency Distribution (Counts) by Vessel of Total Weights





Left Panel the same as Page 23 (=“Plan A”) where we looked at the proportion of time that the sub-sampled estimated total haul weight fell outside of the full data’s 95% C.I. for total haul weight. If you sample every single halibut (i.e. all ~17000), you have a very good estimate of the total catch of halibut in the study for each vessel, so the confidence intervals will be small no matter what you do. If I look at the halibut that the Constellation caught -- the three biggest halibut of all were caught by this vessel, including the 2.08 metre (136kg) monster-halibut. Constellation had, by luck, high within-vessel variance (and low between vessel variance as we talked about before), and sub-sampling in this vessel gives higher uncertainty in estimating the total haul weight from the subsample, than for the other vessels. The Right Panel (=“Plan B”), calculates 95% C.I. on the subsampled data, and shows the proportion of time that these intervals include the true total halibut catch weight by

vessel. Similar issues affect the uncertainty for Constellation, as were observed with previous method.

Another approach was to calculate the proportion of time that the "true" (full dataset's) haul weight fell within the sub-sample's confidence intervals. In other words, as we subsample every 2nd, 3rd, 4th, 5th, ... halibut, the variance (and CI's) on the total haul estimate will get wider and wider, but we hope will still include the "true" total haul weight. If it doesn't then the subsampling is not estimating total haul weight well. The difference in this approach from plan A is that plan A sets a CI on the entire dataset that are thin, and it's hard to hit that place on the dart board if you have high within-vessel variability. The latter approach calculates CI on subsampled data where these wider CI are moving around the dart board and if the sampling is good, including the bullseye ("truth").

This figure below compares the amount of uncertainty associated with subsampling to every 5th fish, and the uncertainty of estimating total haul weight from this (green lines), compared to the more "usual" method of collecting three 100kg samples (white histogram, blue lines). Both methods' estimates of total haul weights appear to be unbiased (median's more or less overlap the yellow "truth"), but we have a much more precise measure by taking every 5th halibut than by the other method (green CI are much narrower than the blue ones, width of the range of estimates shown above each plot). Range of X axis represents the "truth" +/- 7mt, Y axis is between 0,1 for comparison.

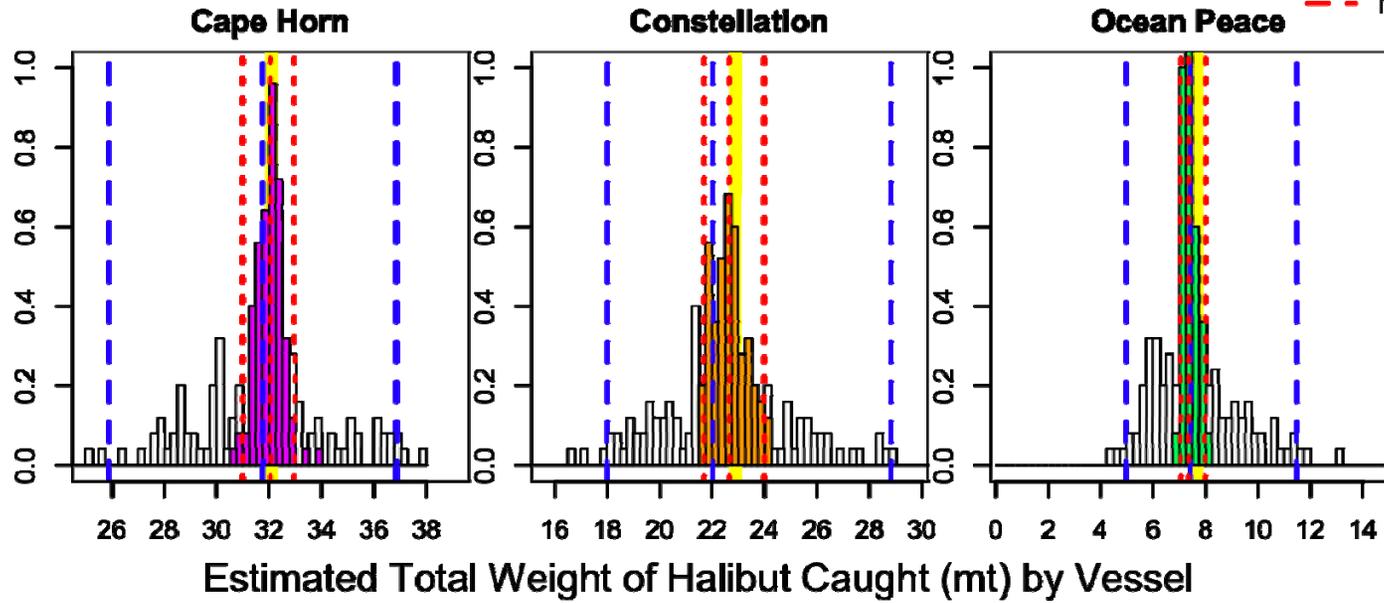
Range of Estimates

Every 5th Fish 3.191
 3 x 100kg 13.667

2.642
 14.111

1.164
 8.875

- █ True Total Weight of Halibut caught
- - - Median (95% CI) from 3 x 100kg sample
- - - Median (95% CI) from measuring every 5th fish



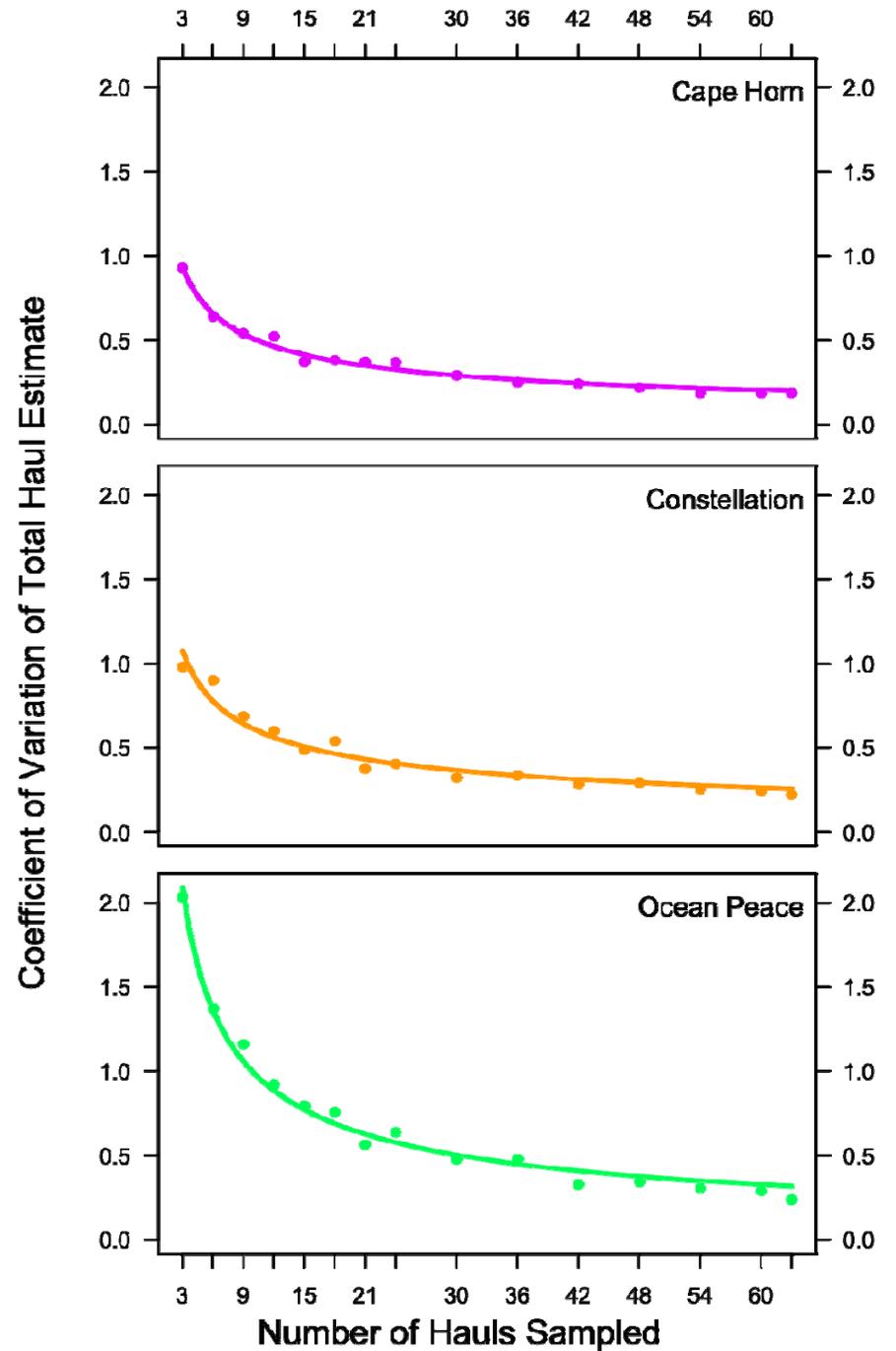
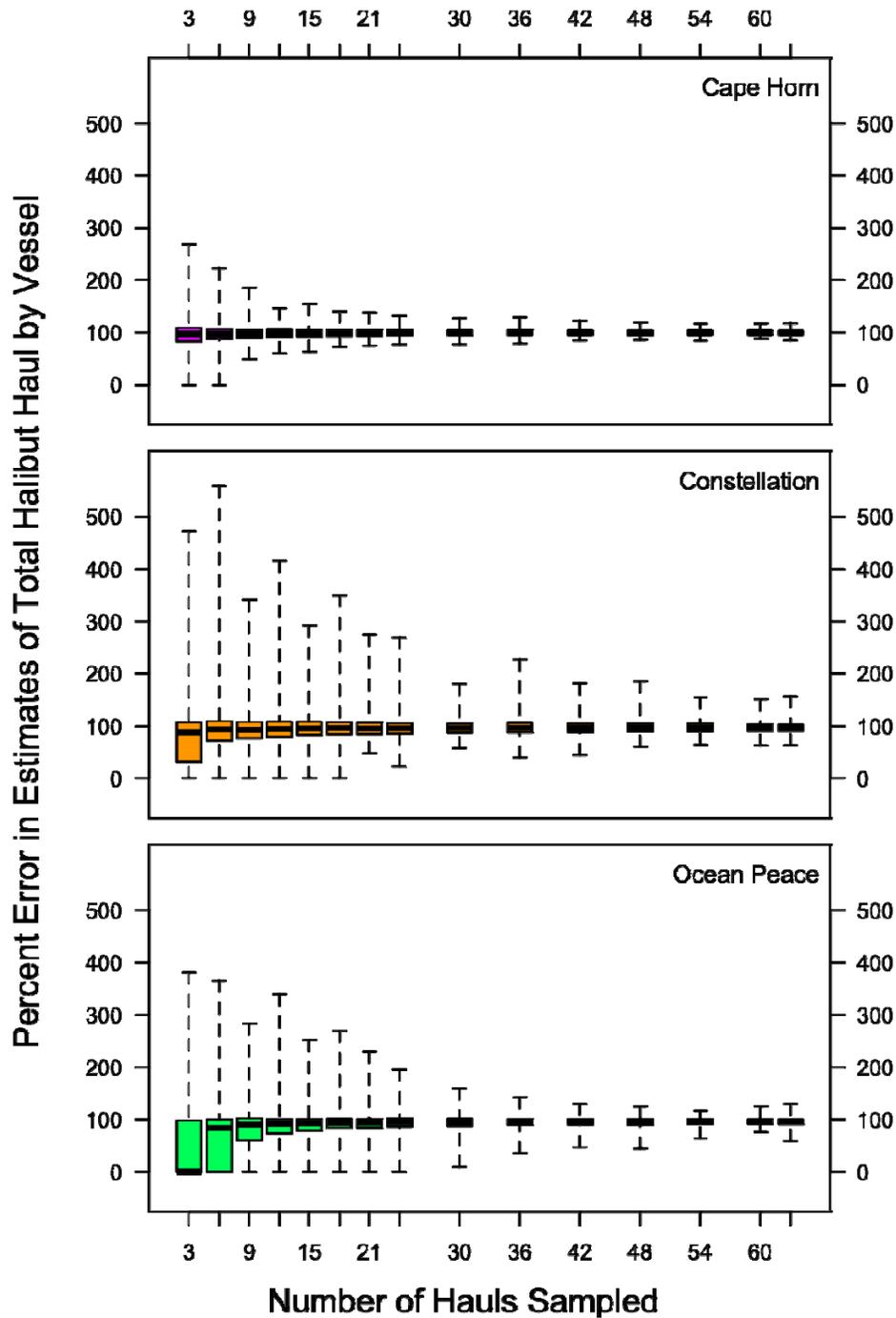
Empirical Confidence Intervals based on a Simulation of Size of 100. For both methods I took the 2.5 and 97.5th limits of our 100 simulated estimates. This is the same as ordering the 100 estimates, and interpolating where the 2.5th and 97.5th value would lie.

```
[1] "Cape Horn Widths of C.I."
[1] "Every 5th Halibut Estimate"
1.962
[1] "3 x 100kg Estimate"
10.959
[1] "Constellation Widths of C.I."
[1] "Every 5th Halibut Estimate"
2.305
[1] "3 x 100kg Estimate"
10.841
[1] "Ocean Peace Widths of C.I."
[1] "Every 5th Halibut Estimate"
0.933
[1] "3 x 100kg Estimate"
6.493
```

The Next Page contains two plots of how the error decreases with the number of hauls. In this simulation, I randomly selected hauls by vessel, with the same number of hauls for each vessel. In other words, I started estimating total catch of halibut per haul with just 3 hauls - one for each vessel, then I increased to 6 hauls - 2 for each vessel, 9 hauls - 3 per vessel, etc. out to 1/3 of 250 hauls or 63 hauls - 21 per vessel; this is reflected in the X axis which goes up by 3's. (X axis = 3,6,9 means 1,2,3 hauls per vessel).

I repeated this 100 times, and plotted the boxplots of the relative errors by vessel (left plot). I also looked at how the coefficient of variation decreased with increased sample size (i.e. square root of the variance of the 100 estimates divided by the total haul size by vessel (or "Truth").

These two plots show that the variability starts to attenuate at +/- 30 hauls. The boxplot interpretation is the whiskers mark the range of estimates, the colored "box" (rectangles) mark where 75% of the estimates are, and the black line in the middle is the median estimate. So one important result seen in the left plot is that Ocean Peace underestimates amount of halibut catch per haul until at least 7 hauls from this vessel are sampled.



Overall Methods:

A simulation study was designed to assess the variability associated with the 2 different sub-sample based estimates of total weight of halibut caught. Each vessel was kept separate in the simulation, as there was significant between vessel variability both in the number of hauls, and the numbers of halibut in those hauls. The number of hauls, and the characteristics of those hauls (number and weights of halibut) were faithful to the original dataset; thus the random procedure was applied at the individual halibut level via the halibut that were selected for the subsample. In this way, the simulated catch captures the observed within and between haul variability, from which the two subsampling-based estimates of the total weight of halibut caught were calculated. The assumption behind this approach assumes that the original data of 250 hauls (Cape Horn=82, Constellation=137, Ocean Peace=31) captures the full variability in hauls for this area, and that if similar conditions were to be observed, this dataset is a good representative of any future year of fishing under similar conditions.

Every 5th fish (sample based) estimate:**Methods:**

In this simulation study we systematically subsampled every 5th halibut, and used the characteristics of the subsample along with the known number of halibut per haul to estimate total weight of halibut.

The simulation proceeded by assuming there was no pattern or bias in halibut size associated with the order in which they come out of the net for processing. The simulation begins by generating a random start number between 1 and 5, and then selecting that and every 5th fish thereafter. When all halibut caught in that haul have their processing order randomized, then measuring every 5th halibut amounts to the same as measuring 20% of the halibut caught, as long as there were at least 5 halibut in the haul. If the random start number was greater than the number of fish in the haul, then no halibut were selected for processing for that haul.

For small hauls of halibut, the error enters the estimate in two ways. If no halibut were selected for weighing, then the haul estimate would equal zero. For example, in the original dataset there were 15 hauls containing only 1 halibut, and the probability of selecting 1 as the random start number for subsampling is 20%. In other words there is an 80% chance that no halibut were sampled, and the haul size estimate would be zero. If 80% of these 15 haul estimates were not subsampled, then 12 out of these 15 vessels would be biased (under) estimates of 0.0 kg halibut. The second source of error for small hauls is the usual estimate error of using a subsample of halibut to represent the whole sample. A small subsample size of halibut on which the estimate is based, is more prone to the vagaries of chance large or small individuals, than a large subsample would be. Larger hauls had larger numbers of halibut to average to dampen the effect of any single outlier.

3 x 100 kg (sample based) estimate:

Methods:

For this simulation study, we selected 100 kg of fish three times from each haul and averaged the proportional weight of halibut in those three samples, to infer total halibut weight in each haul. Because halibut are a large fish and contributed a relatively small weight to the total groundfish catch, a single sample of 100 kg would on average not hold many halibut, and in many cases not hold any. We investigate the error of this subsampling method in the following simulation study.

Fish were randomly assigned to a sample until the cumulative weight of fish was at least 100kg, thus the number of fish in each 100kg sample varied. In addition, since only whole fish were included in the sample, the total weight of the simulated sample varied. Therefore, the 100kg sample would be larger than 100kg by the amount of weight that last fish contributed past 100kg. Each individual in the 100kg sample was either a halibut or not, and thus whether or not a fish is a halibut can be thought of as a Bernoulli trial where the probability of the fish is halibut is equal to the known relative proportion of halibut in the entire haul (total weight halibut/total haul weight (OTC)). Sample weights of halibut were divided by the total sample weight to obtain proportion estimates for each of the three 100 kg samples per haul, and used to calculate the weight of halibut as the proportion of weight in the entire haul. For example, if there was an average of 10 kg of halibut in the three 100kg samples, then the estimate of halibut for the haul would be 10% of the total haul weight.

Census-based estimate:

In each of the simulations described above, a census-based estimate in which the weight of all halibut in each haul was simultaneously calculated. In this way, the simulation mimicks the job of the on-board observers measuring every halibut, and for each run of the simulation, the "true" haul weight of halibut is available for comparing relative error.

Number of Hauls to conduct census and sample based estimation

Methods:

This simulation study was designed to examine how the estimation error of halibut weight decreased as the number of subsampled hauls increased. The simulations were run as before by calculating the estimates of halibut weight from both the every-5th-halibut subsample and the fully censused haul. For one run of the simulation, we selected a fixed number of hauls and calculated both subsample and census estimates of halibut catch weight. The simulation randomly selected hauls by vessel, with the same number of hauls selected for each vessel. In other words, we started estimating total haul with just 3 hauls - one for each vessel, then we increased to 6 hauls, which is 2 hauls for each vessel, 9 hauls - 3 per vessel, etc. out to 1/3 of 250 hauls or 63 hauls (21 per vessel). This was repeated 1000 times, and the relative error and the coefficients of variation were calculated for each increase in number of hauls.