Clam survey redesign, part 2

Principal methods

- Optimal allocation with ≥ 2 tows per stratum
- Measure survey performance using design efficiency statistics $DEF = DEF_{Allocation} + DEF_{Stratification}$.
- Assign building blocks to new strata based on historical mean catch, mean catch + location and depth, predicted mean catch from GAM or tree models.
- Cluster analysis to group building blocks into new 2-10 new strata
  - Univariate (catch only)
  - Multivariate (catch, location and depth)
- Bootstrapping to evaluate strata options
- Stability analysis
- Simulate effects on management advice
- Domain statistics to evaluate historical post-stratification
- Use clam station data (3 methods) to locate bad locations and exclude 1 nm circle (about one long tow, 0.8 nm^2) around each

Questions, methods and answers

Q1) Should clam surveys target surfclams and ocean quahogs separately rather than simultaneously? ***Yes***

Q2) Should sampling of poor habitat areas cease, particularly if the two species are surveyed separately? ***Yes***

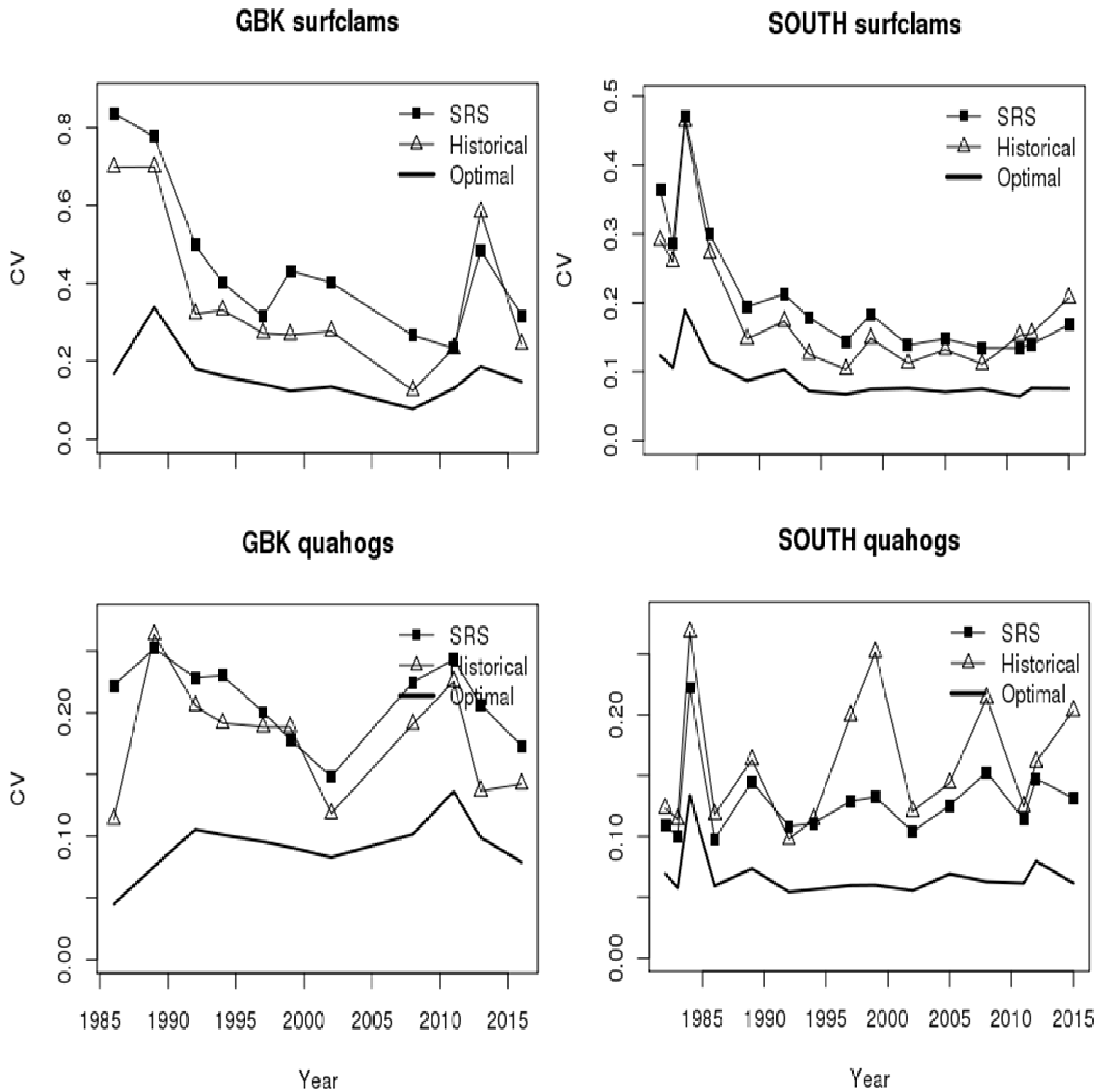
Q3) Should new species-specific stratification schemes be considered if the two species are surveyed separately? ***Yes***

Rationale:

Precision increases, smaller areas include almost all habitat, negligible bias in swept-area estimates, surfclams moving to deep water will be observed in quahog surveys

Historical analysis

Figure 3.9 Historical analysis showing CVs for stratified mean catch density in NEFSC clam surveys based on the current design and CVs expected under random sampling and with optimal allocation. All three scenarios use the current stratification scheme.



Q4) Is it feasible to survey the entire stock (GBK plus south) for surfclams or ocean quahogs in one survey year if the species are separated and sampling area reduced? ***No***

Q5) What scheduling options (number, location and frequency of surveys for both species) should be considered if surveys for the two species are separated? ***Option 5 (see below)***

Rationale:

Survey frequency for surfclams (more variable) is unchanged. High frequency surveys unnecessary for both species (stock biomass high, mortality low, old animals common). Precision more important than frequency for quahogs based on simulations. Rapid changes in stock probably detected during routine review of logbook data. Survey frequency can be adjusted if necessary.

Table 4.1. Six options for scheduling NEFSC clam surveys for surfclams and ocean quahogs over an eighteen year planning horizon. In the second column, “separate” means that surfclams and ocean quahogs are targeted separately (i.e. during separate years), “test in 3rd year” refers to time for gear testing (currently every third year), and “2x surfclams” means that the frequency of surfclam surveys is double the current frequency. “Sn”, “Ss”, “Qs” and “Qs” refer to surveys for surfclams and ocean quahogs in the northern (GBK) and southern assessment areas. Option 5 (target the two species separately, double the survey frequency for surfclams and halve the frequency for ocean quahogs) is recommended.

Option	Description	Yr 1	Yr 2	Yr 3	Yr 4	Yr 5	Yr 6	Yr 7	Yr 8	Yr 9	Yr 10	Yr 11	Yr 12	Yr 13	Yr 14	Yr 15	Yr 16	Yr 17	Yr 18	N surveys per year	
																				Surfclam	Quahogs
1	Current	Ss	Sn	T	Ss	Sn	T	Ss	Sn	T	Ss	Sn	T	Ss	Sn	T				0.33	0.33
		Qs	Qn		Qs	Qn		Qs	Qn		Qs	Qn		Qs	Qn						
2	Separate, test in 3rd year	Ss		T	Qs		T	Ss		T	Qs		T	Ss		T				0.20	0.20
			Sn			Qn			Sn			Qn			Sn						
3	Separate, test in 3rd year, 2x surfclams	Ss		T	Ss		T	Qs		T	Ss		T	Ss		T	Qs		T	0.22	0.11
			Sn			Sn			Qn			Sn			Sn			Qn			
4	Separate, no testing	Ss		Qs		Ss		Qs		Ss		Qs		Ss		Qs				0.25	0.25
			Sn		Qn		Sn		Qn		Sn		Qn		Sn		Qn				
5	**Separate, no testing, 2x surfclams**	**Ss**		**Ss**		**Qs**		**Ss**		**Ss**		**Qs**		**Ss**		**Ss**		**Qs**		**0.33**	**0.17**
			Sn		**Sn**		**Qn**		**Sn**		**Sn**		**Qn**		**Sn**		**Sn**		**Qn**		
6	Separate, no testing, 3x surfclams	Ss		Ss		Ss		Qs		Ss		Ss		Ss		Qs				0.38	0.13
			Sn		Sn		Sn		Qn		Sn		Sn		Sn		Qn				

Q6) Can rough ground with risk to equipment damage be avoided? ***Yes***

Rational: Current survey uses a large and expensive commercial dredge mounted on a steep ramp. Major repairs require return to port, substantial loss of time and reduced data precision.

The database describing untowable grounds should be updated whenever possible based on any available information.

208 unique locations covering 561 km^2 (163 nm^2) identified, most on northern GBK in surfclam habitat

Fraction excluded = 561 / 17,514 = 3.2% GBK surfclam area

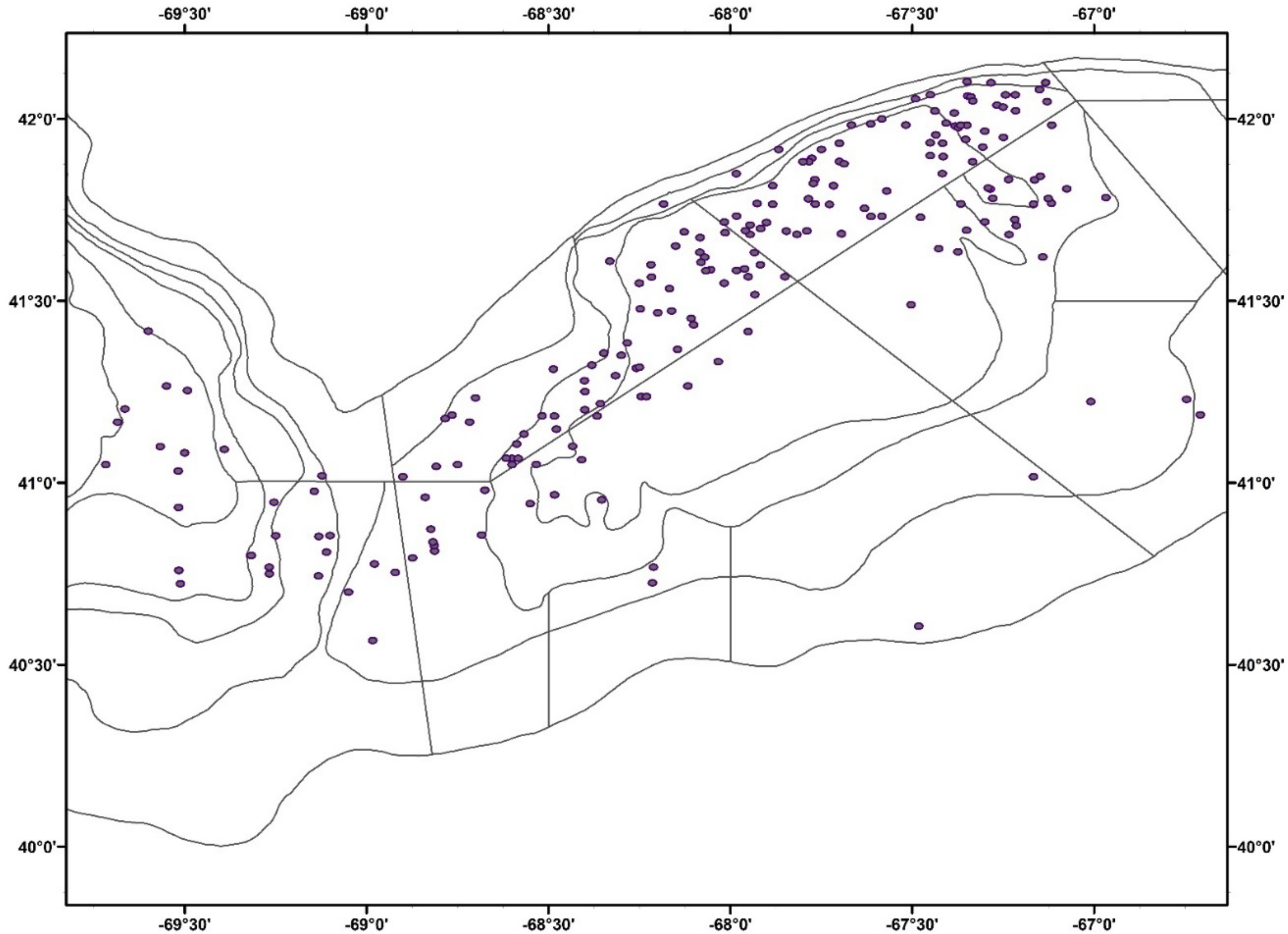


Figure 3.20. NEFSC clam survey stations locations that identify untowable ground. Each location is surrounded by a circle one nautical mile in diameter.

Q7) Should new strata be constructed from current strata or built from scratch using smaller building blocks? ***Use current strata.***

Q8) How heavily should location and depth information vs. survey catch data be weighted in developing new strata? ***Down-weight catch (wt 0.5).***

Q9) Should new strata schemes with discontiguous strata be considered or should strata be defined traditionally as single contiguous areas? ***Use contiguous strata.***

Q10) What are the recommended stratification options (method, location, shape and number of strata) for each species and area? ***6-7 new strata identified by multivariate cluster analysis based on lat, lon, depth and catch (wt=0.5). Survey branch can make modest changes if needed based on logistic considerationsl.***

Rationale: Not enough data to reliably assign FMSQ to new strata in cluster analysis. Univariate clusters based on catch only and multivariate clusters with higher weights on catch were unstable. Recommended approach performed consistently in out-of-sample bootstraps. GAM and tree procedures did not perform as well. Six strata judged sufficient based on statistical advice, n tows per stratum and DEF bootstrap analysis (seven strata for southern ocean quahogs to avoid discontiguous pattern). Domain sampling shows recommendation work well for historical data.

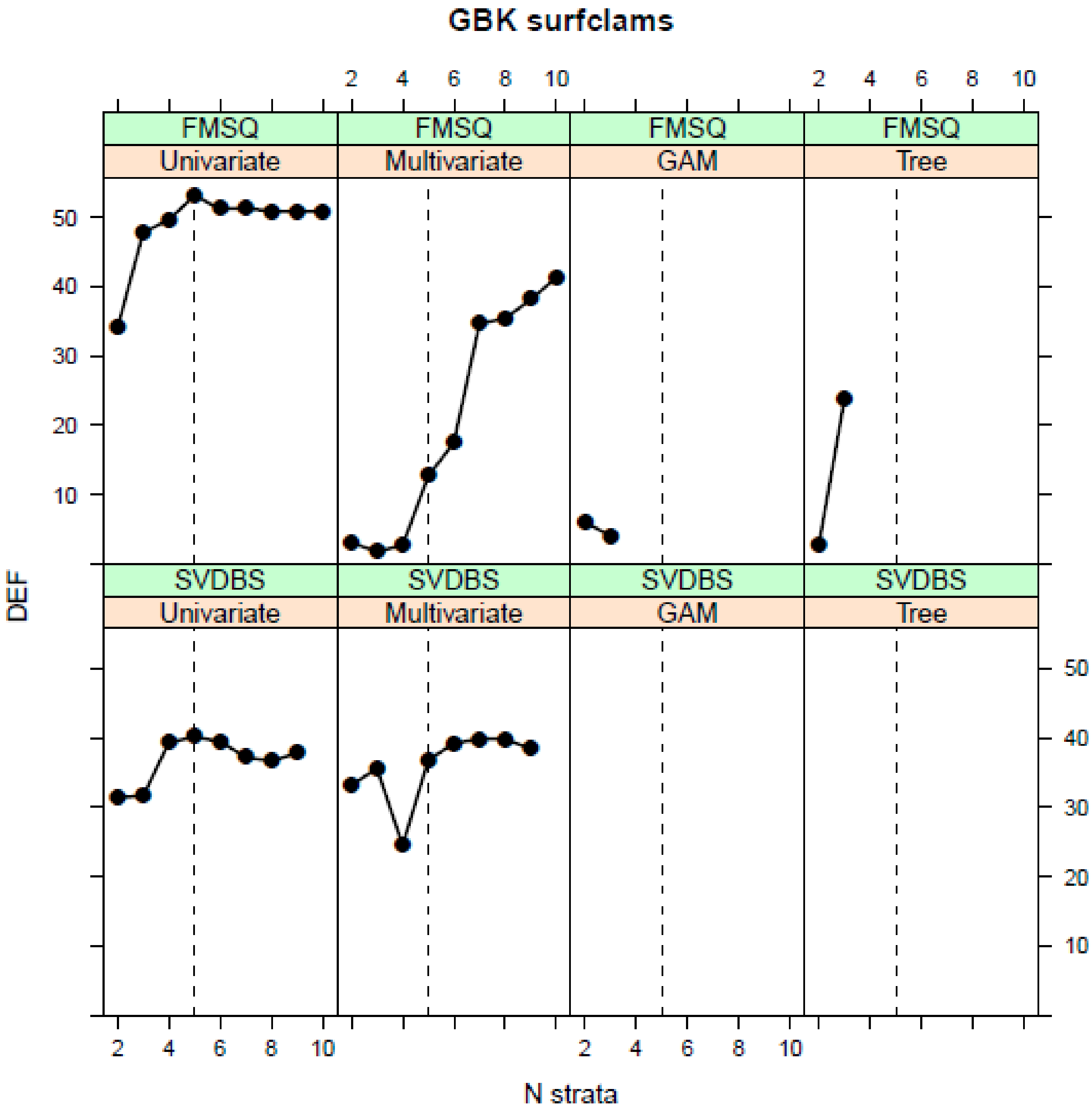


Figure 3.11. Median DEF results for ocean quahogs on GBK from bootstrap analysis. Results are shown for two types of building blocks (FMSQ vs. SVDBS = current strata), four cluster methods and 2-10 potential new strata (clusters).

PRELIMINARY DATA – shows general patterns

New (?) rule of thumb for N strata

Say at least 25 stations per stratum required to estimate means and variances (used in allocation) with adequate precision (Figure 4.9).

If 150 stations can be handled on GBK during one survey, then the number of strata should not exceed 150/25=6. Similarly, the number of strata in the south should not exceed 200/25=8.

If resources for clam surveys were reduced and station numbers fell by 25% then the maximum number of strata would be 150*0.75/25=4.5 on GBK and 200*0.75/25=6 in the south.

Future reductions in sample size are possible.

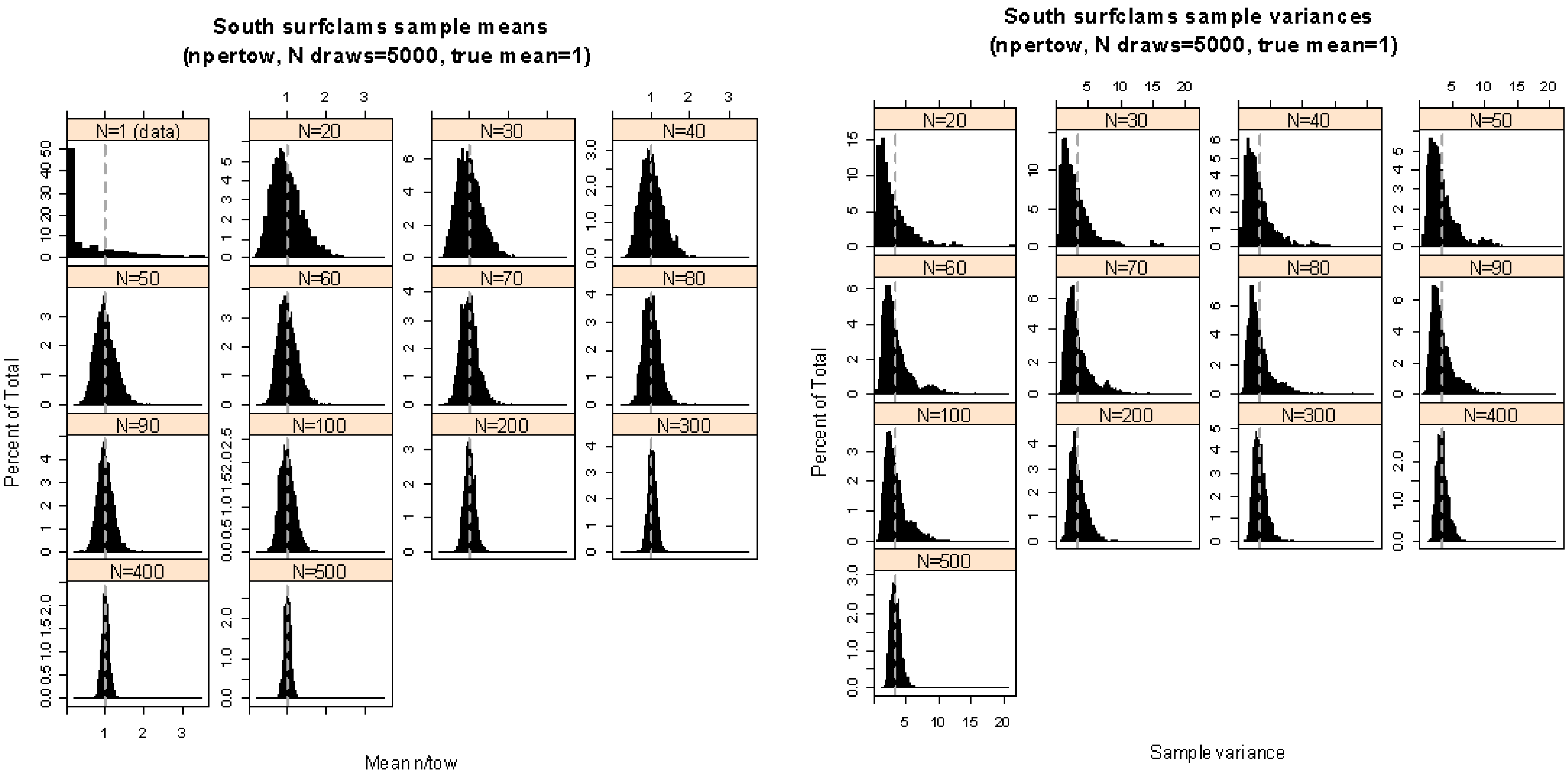


Figure 4.9. Simulation illustrating relationship between sample size and the precision of estimated means and variances in simple random sampling. Each panel shows the distribution of 5000 estimated means and variances for samples of the size indicated at the top of the panel (e.g. N=20 means 20 tows were sampled 5000 times). The data were for surfclams in the southern area during 1997-2016 in building blocks selected based on the 1% rule. The selected data were scaled to a mean on one in each year to avoid exaggerating variance. Vertical lines in each panel show the simulated true population statistic.

Surfclams										
Catch Weight	GBK current strata					South current strata				
N new strata	2	3	4	5	6	2	3	4	5	6
0.25	***74***	***66***	***60***	***55***	39	79	80	81	70	65
0.50	73	65	45	46	40	78	79	80	69	64
1.00	70	62	50	42	***42***	82	84	82	72	65
Univariate	63	54	51	44	36	***93***	***89***	***84***	***78***	***67***
Range	11	12	15	13	6	15	10	4	9	4
Catch Weight	GBK FMSQ					South FMSQ				
0.25	0	8	10	30	7	21	4	19	31	24
0.50	0	20	24	30	12	23	19	27	24	34
1.00	19	18	28	28	23	**48**	45	42	36	44
Univariate	***24***	***42***	***31***	***38***	***32***	30	***66***	***62***	***60***	***73***
Range	24	35	21	10	25	26	62	43	36	49

Ocean quahogs										
Catch Weight	GBK current strata					South current strata				
N new strata	2	3	4	5	6	2	3	4	5	6
0.25	63	***59***	37	39	37	116	***105***	94	***87***	***87***
0.50	61	56	46	39	38	116	***105***	97	***87***	78
1.00	61	55	***52***	39	38	116	107	95	84	77
Univariate	***64***	46	49	***42***	***39***	***120***	101	***99***	84	76
Range	3	13	16	3	1	4	6	5	3	11
Catch Weight	GBK FMSQ					South FMSQ				
0.25	19	10	1	8	20	18	0	22	23	21
0.50	20	16	1	30	32	17	14	24	17	46
1.00	23	26	9	17	25	35	42	47	68	31
Univariate	***36***	***33***	***33***	***32***	***26***	***61***	***63***	***88***	***85***	***77***
Range	17	23	32	24	12	44	63	67	68	57

Table 3.2. Root mean squared error (RMSE) statistics measuring stability of potential stratification options with 2-6 new strata based on univariate cluster analysis and multivariate cluster analysis with preliminary data and a range of weights on catch density. Results for different species, areas, numbers of strata and type of building blocks (current strata vs. FMSQ) are not comparable. The most stable option in each set (lowest RMSE) is grey and the least stable (highest RMSE) are bold and italicized.

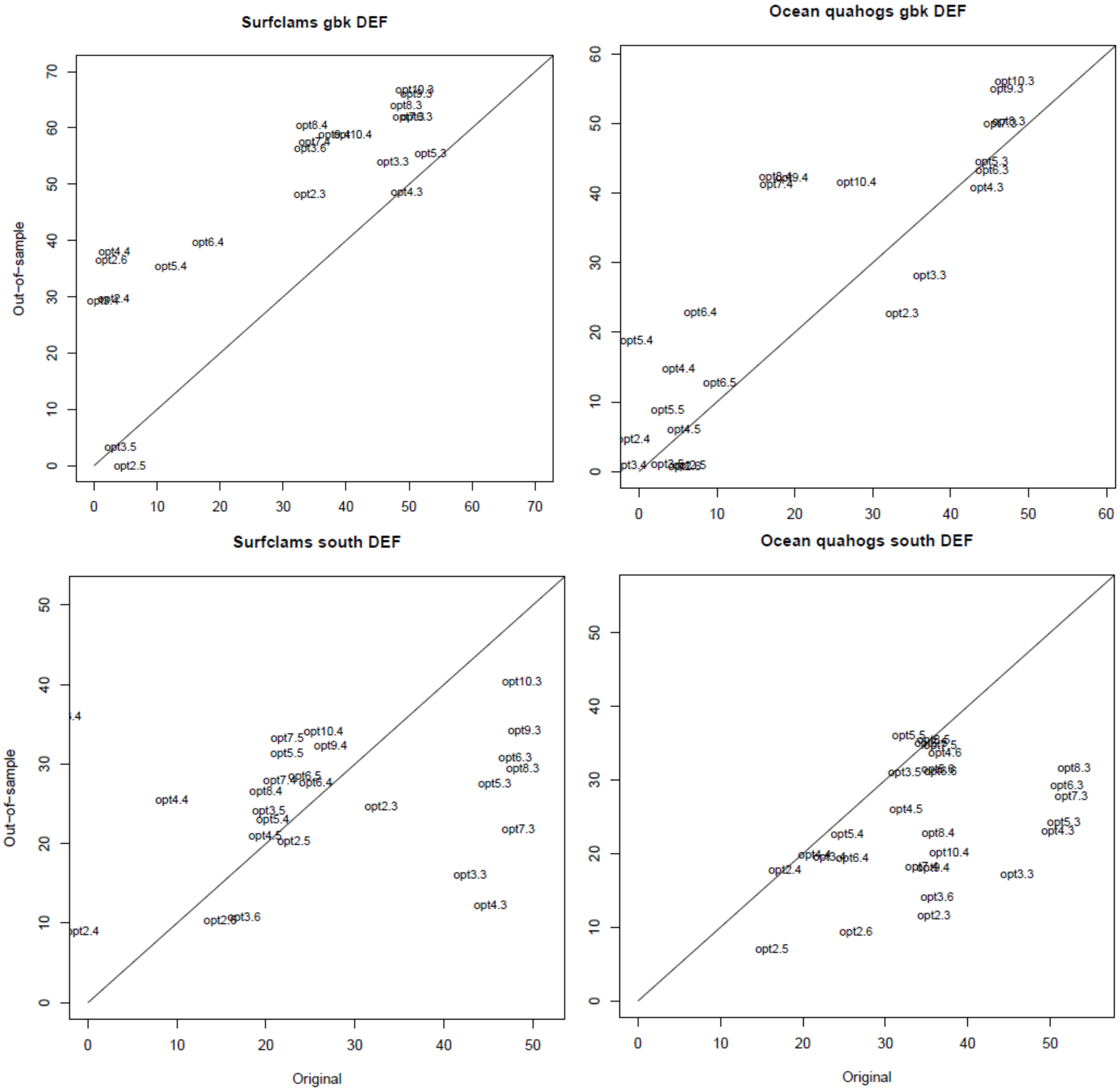


Figure 3.14. Median DEF scores from in-sample (x-axis) and out-of-sample (y-axis) bootstrap analyses using FMSQ as building blocks and preliminary data.

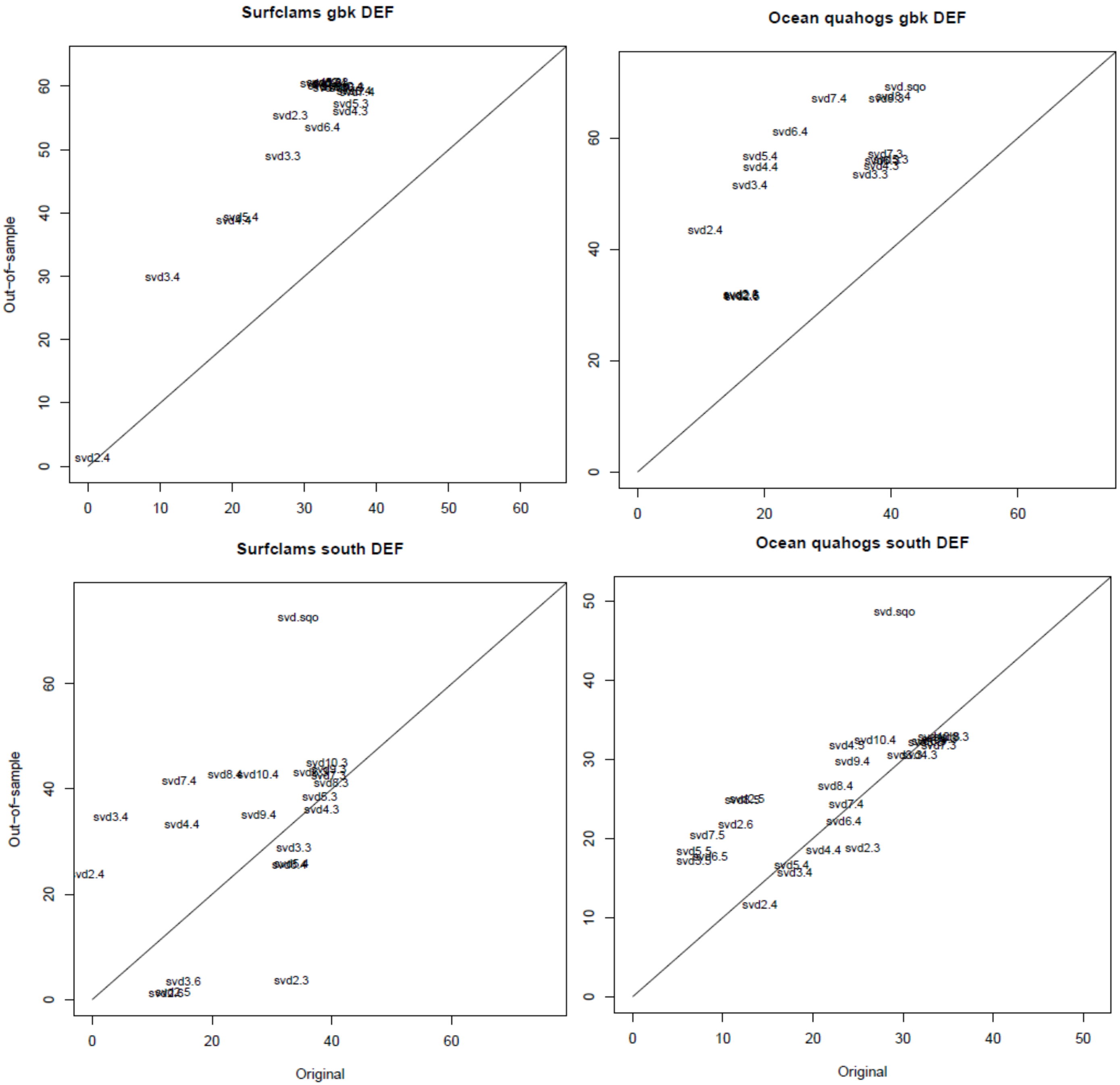


Figure 3.15. Median DEF scores from in-sample (x-axis) and out-of-sample (y-axis) bootstrap analyses using current strata as building blocks and preliminary data.

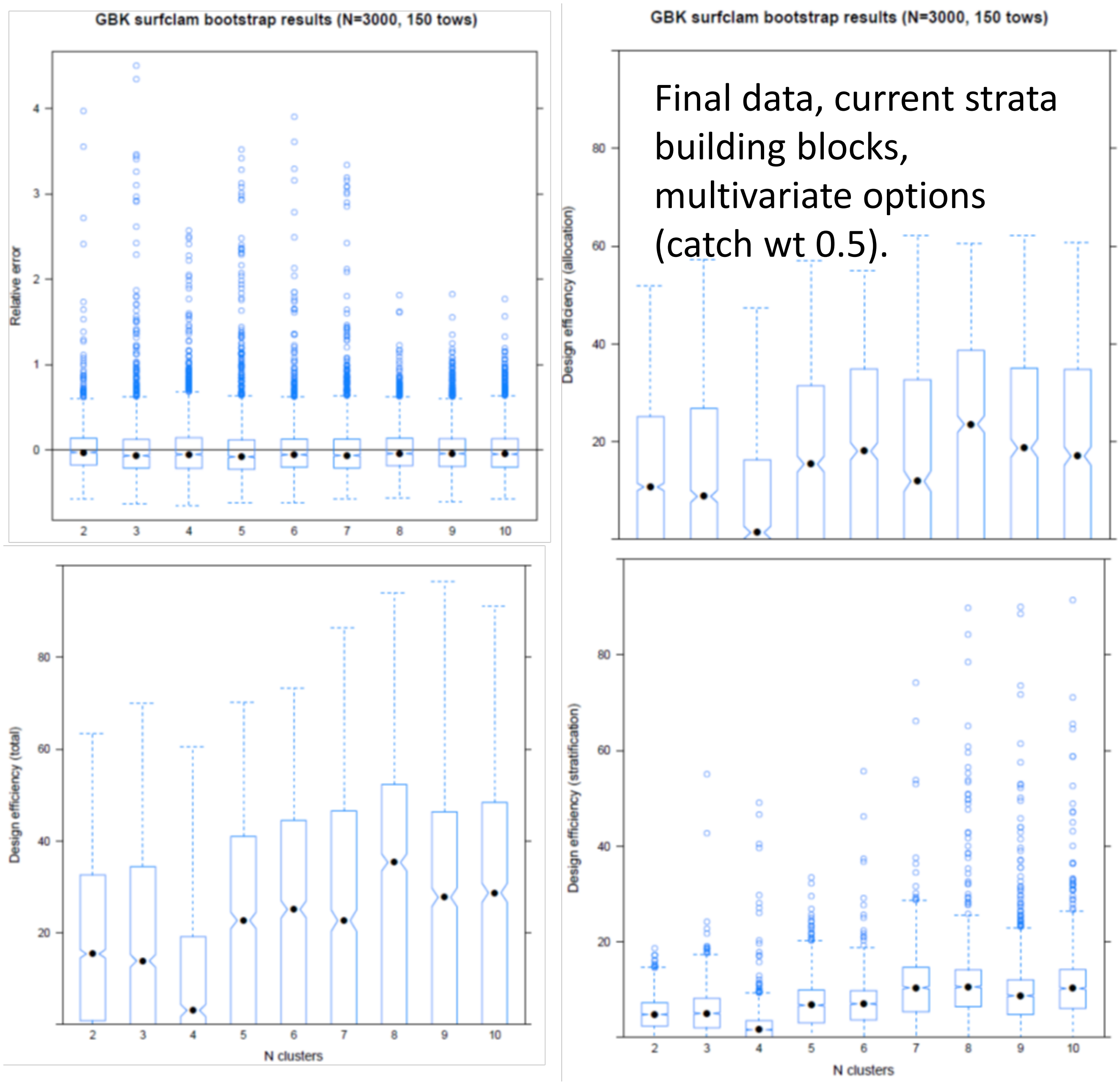


Figure 4.5. Bootstrap results (3000 iterations) for GBK surfclams and 2-10 new strata (clusters) based on final data. Top left: Relative errors. Top right: *DEF* (total). Bottom left: $DEF_{Efficiency}$. Bottom right: $DEF_{Allocation}$. Only positive DEF values are shown for clear presentation. Black dots show medians and notched portion of figures shows approximate 95% confidence intervals for the medians.

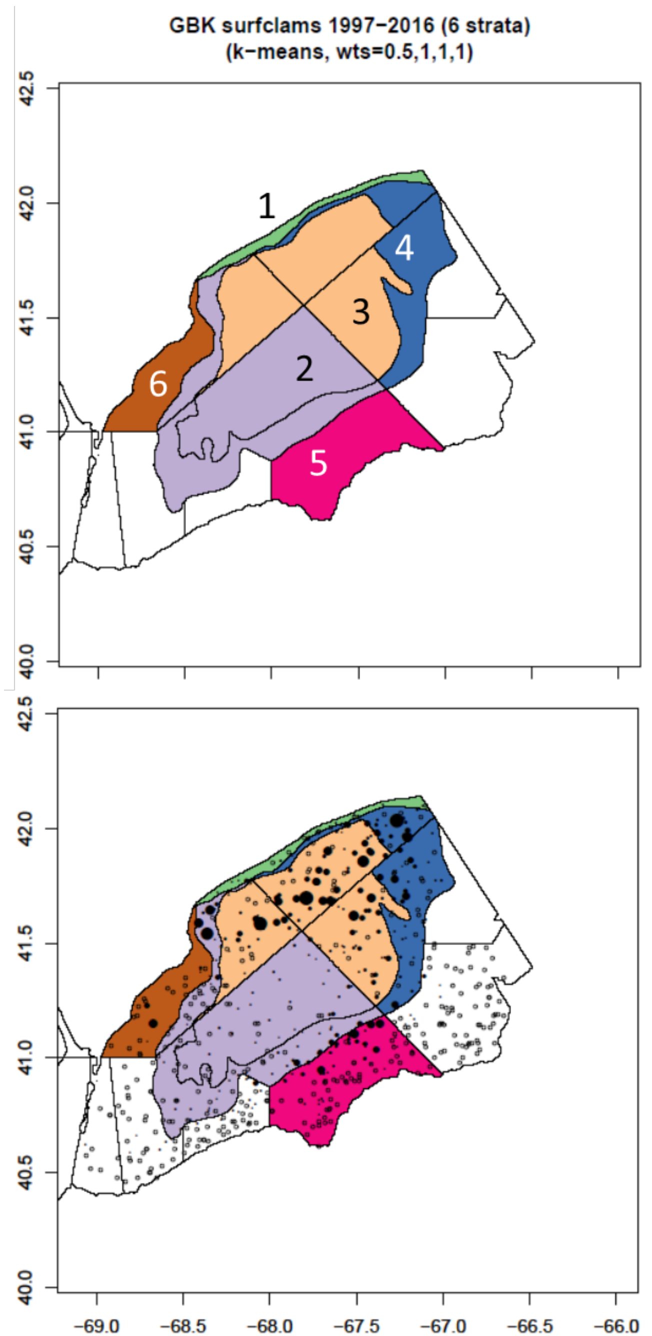


Figure 4.1. Top: Recommended strata for GBK surfclams based on cluster and bootstrap analyses with final data, with new stratum ID numbers. Bottom: Recommended strata with survey data used in analyses. Figures are numbered as in Table 4.3.

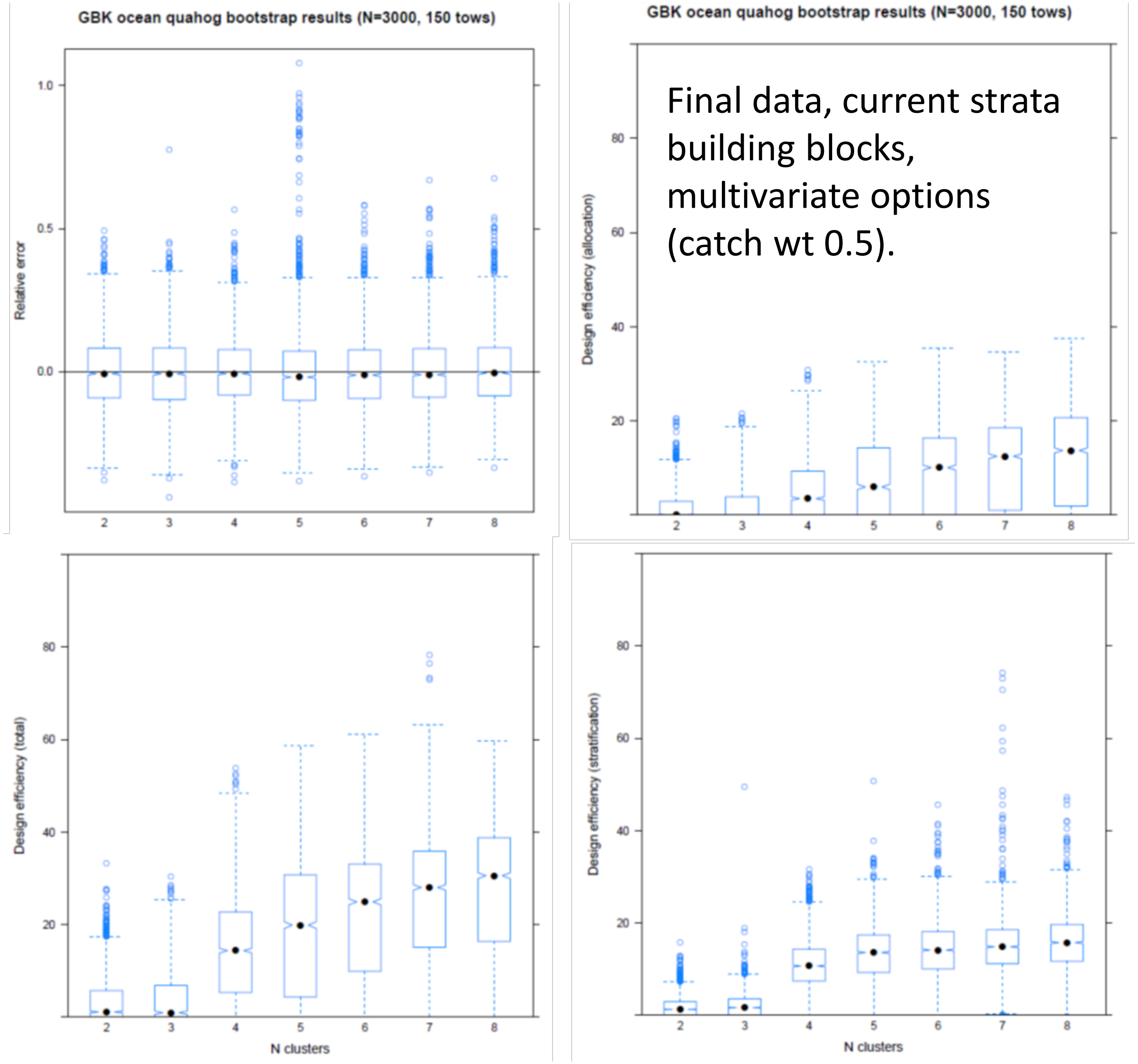


Figure 4.6. Bootstrap results (3000 iterations) for GBK ocean quahogs and 2-10 new strata (clusters) based on final data. Top left: Relative errors. Top right: *DEF* (total). Bottom left: $DEF_{Efficiency}$. Bottom right: $DEF_{Allocation}$. Only positive DEF values are shown for clear presentation. Black dots show medians and notched portion of figures shows approximate 95% confidence intervals for the medians.

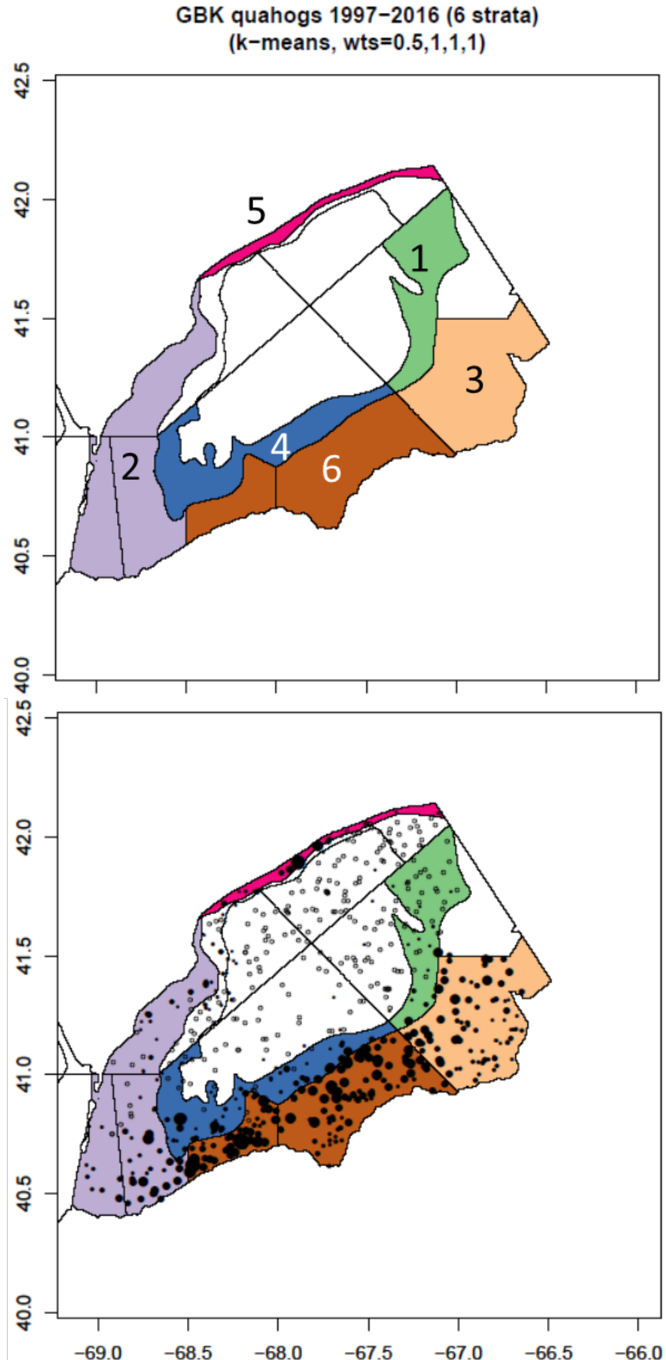


Figure 4.2. Top: Recommended strata for GBK surfclams based on cluster and bootstrap analyses with final data with final data, with new stratum ID numbers. Bottom: Recommended strata with survey data used in analyses. Figures are numbered as in Table 4.3.

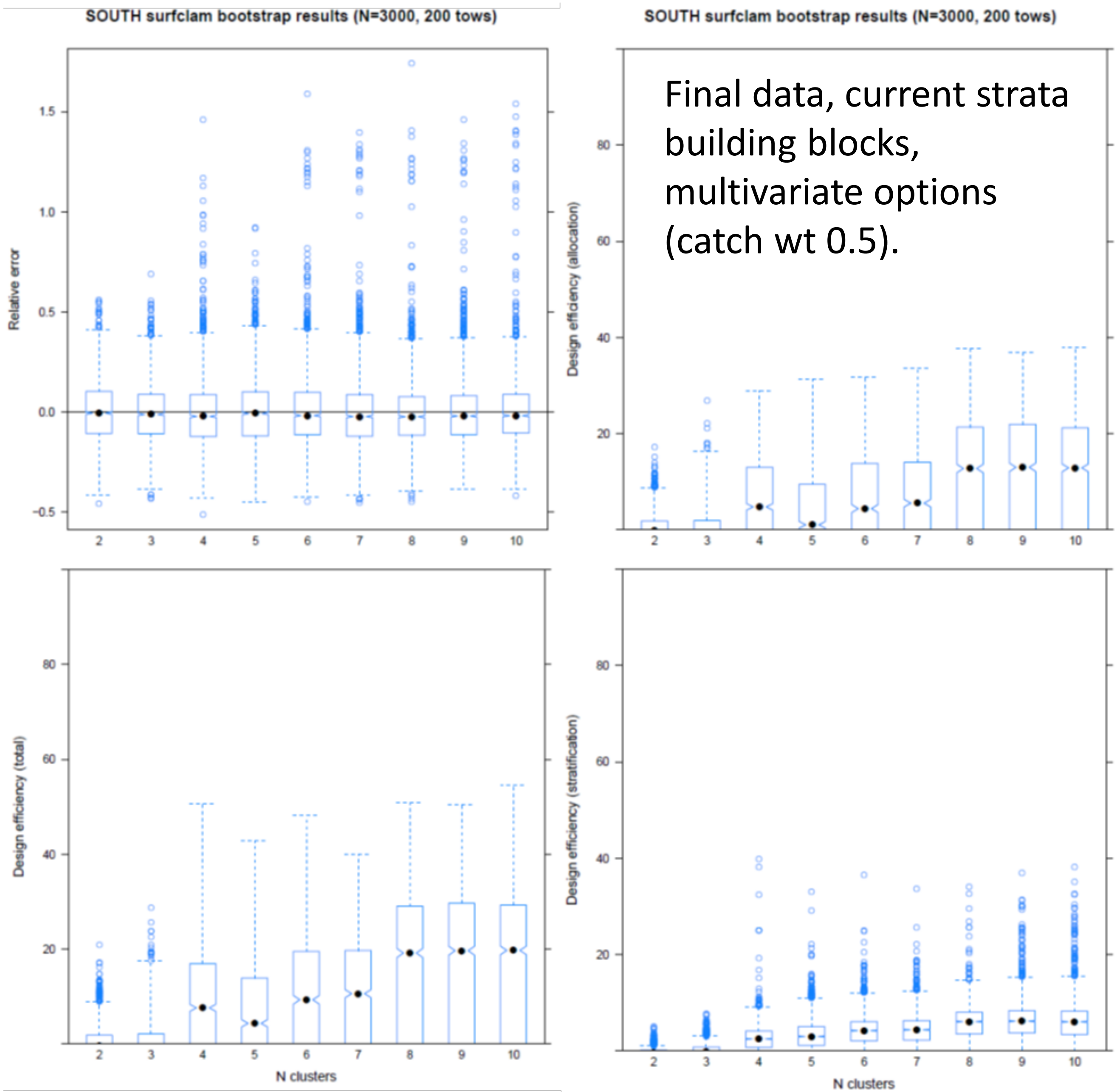


Figure 4.7. Bootstrap results (3000 iterations) for surfclams in the southern area and 2-10 new strata (clusters) based on final data. Top left: Relative errors. Top right: *DEF* (total). Bottom left: $DEF_{Efficiency}$. Bottom right: $DEF_{Allocation}$. Only positive DEF values are shown for clear presentation. Black dots show medians and notched portion of figures shows approximate 95% confidence intervals for the medians.

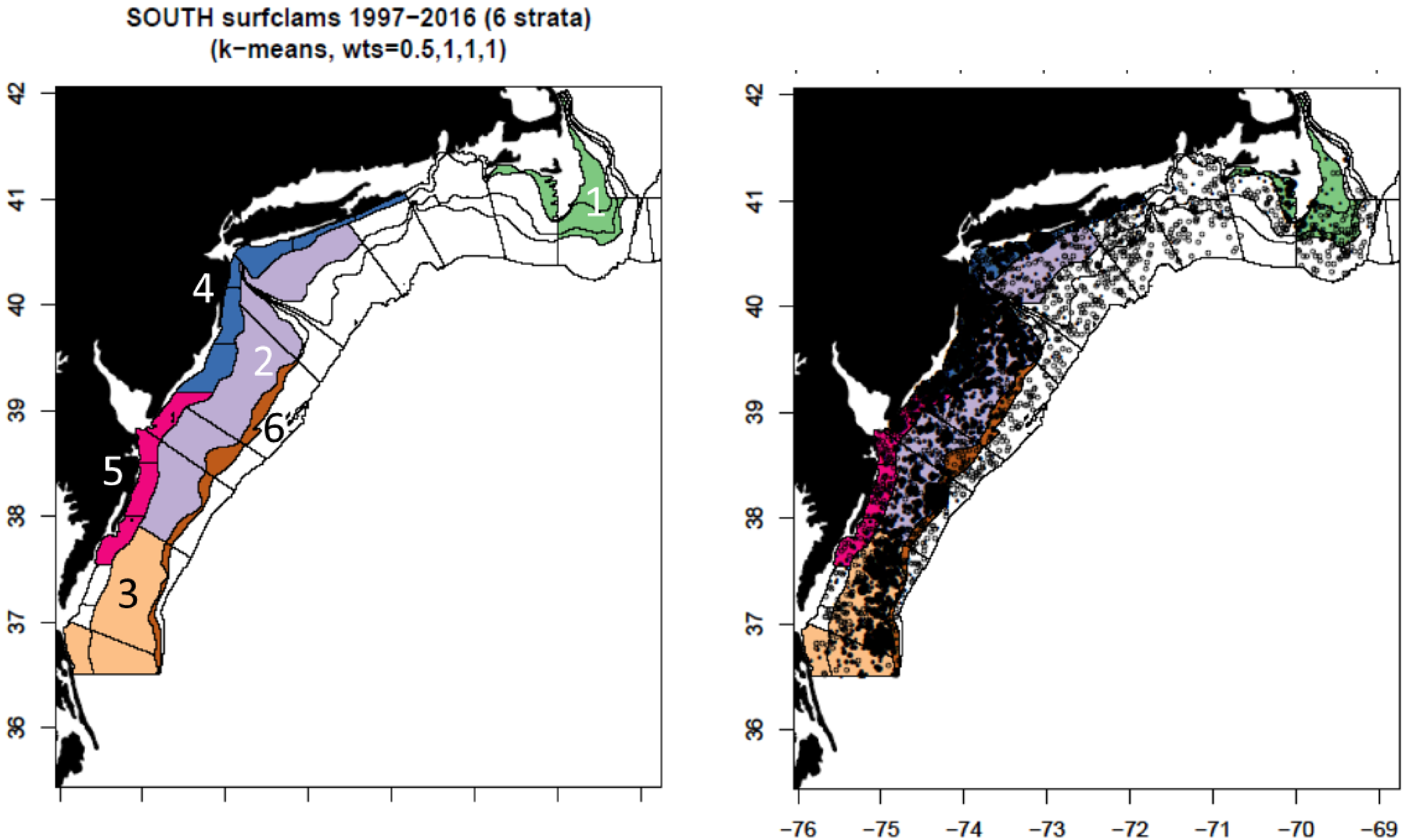


Figure 4.3. Top: Recommended strata for surfclams in the southern area based on cluster and bootstrap analyses with final data, with new stratum ID numbers. Bottom: Recommended strata with survey data used in analyses. Figures are numbered as in Table 4.3.

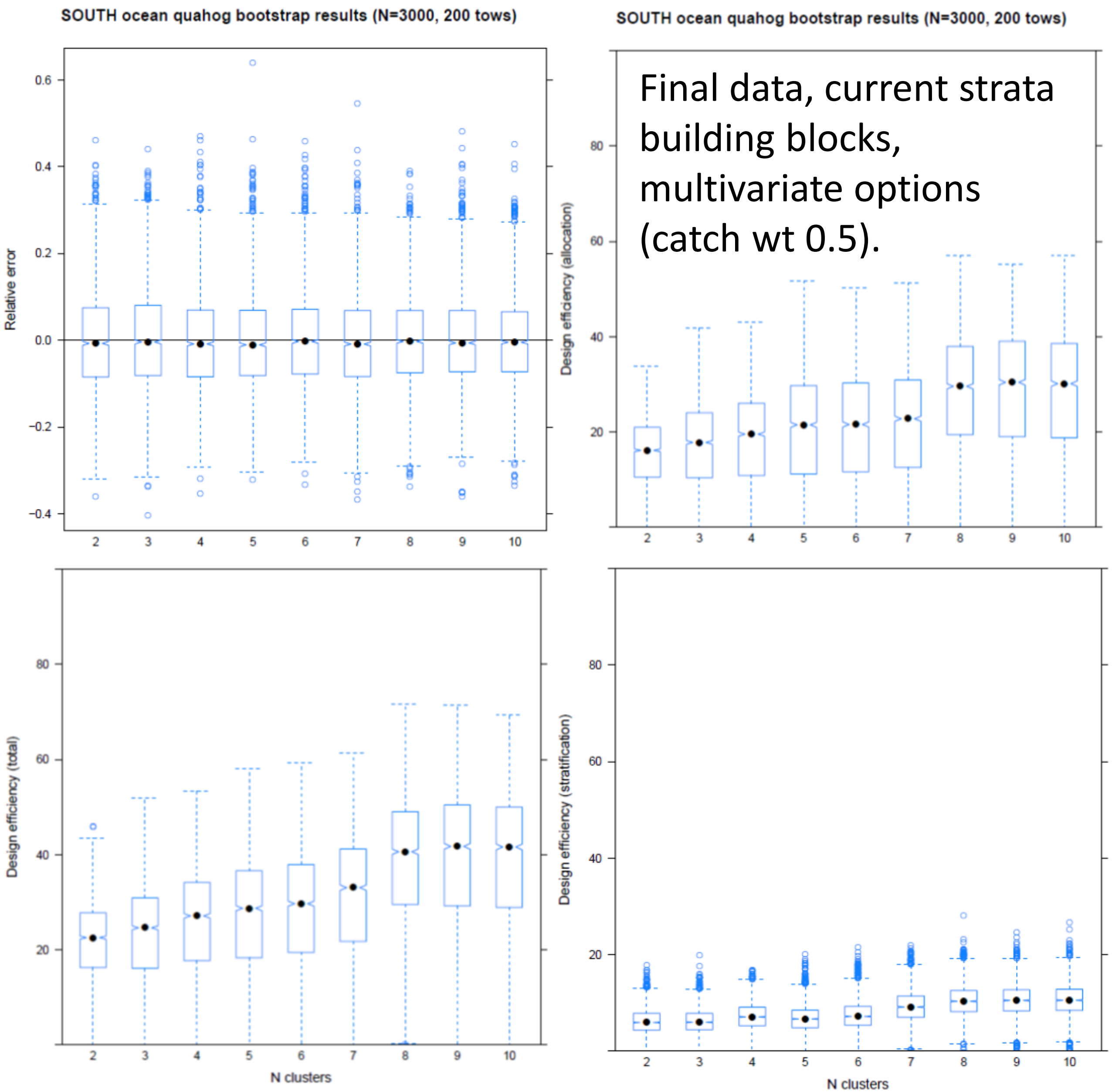


Figure 4.8. Bootstrap results (3000 iterations) for ocean quahogs in the southern area and 2-10 new strata (clusters) based on final data. Top left: Relative errors. Top right: *DEF* (total). Bottom left: $DEF_{Efficiency}$. Bottom right: $DEF_{Allocation}$. Only positive DEF values are shown for clear presentation. Black dots show medians and notched portion of figures shows approximate 95% confidence intervals for the medians.

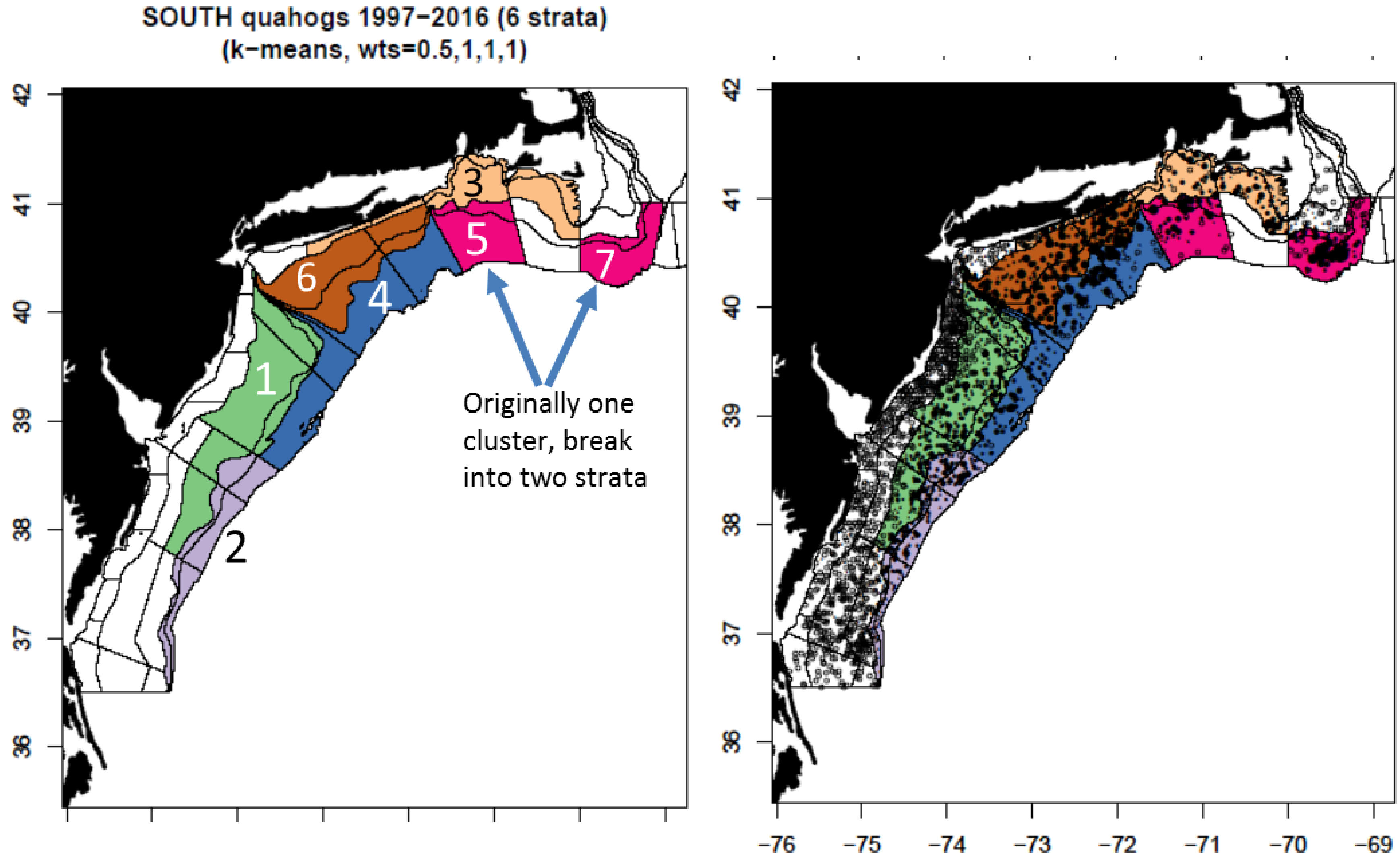


Figure 4.4. Top: Recommended strata for ocean quahogs based on cluster and bootstrap analyses with final data, with new stratum ID numbers. Bottom: Recommended strata with survey data used in analyses. Note recommendation to break one of the strata identified analytically into two new strata based on logistic considerations. Figures are numbered as in Table 4.3.

Table 4.2. Median design efficiency (DEF) statistics for recommended stratification options for the NEFSC clam survey from bootstrap analyses (3000 iterations). DEF measures the percent reduction in variance of stratified random means relative to the variance from a random design. For example, the recommended option for GBK surfclams would be expected to reduce variance by 25% relative to a random design. DEF total = DEF allocation + DEF stratification, where the latter terms are the benefits of optimal allocation and the stratification scheme.

Species	Area	DEF allocation	DEF stratification	DEF total
Surfclams	GBK	18	7	25
	South	6	4	10
Ocean quahogs	GBK	10	14	25
	South	27	9	35

Table 4.3. Bootstrap results for recommended design options in the NEFSC clam survey (stratum level, 3000 iterations). “Mean variance” is the average within-stratum variances across bootstrap samples, “mean allocation” is the average allocation to each stratum and “perfect allocation” gives the optimal Neyman sample sizes for comparison. There were a total of 150 random station on GBK and 200 in the south. “Allocation/100 sq km” is the sampling intensity (number of random stations per area). Strata numbers correspond to Figures 4.1-4.4.

Species	Area	N strata	Statistic	New stratum						
				1	2	3	4	5	6	7
Surfclams	GBK	6	Area (sq km)	467	5632	4583	2239	2310	1231	--
			Mean variance	0.29	4.47	23.74	18.55	4.33	5.64	--
			Mean allocation	2	20	78	31	13	6	--
			Perfect allocation	1	34	65	28	14	8	--
			Allocation/100 sq km	0.44	0.35	1.71	1.38	0.54	0.52	--
Surfclams	South	6	Area (sq km)	4733	18076	10104	5305	4221	3300	--
			Mean variance	1.63	5.67	7.17	5.77	4.13	2.88	--
			Mean allocation	9	94	57	27	6	7	--
			Perfect allocation	12	84	53	25	17	11	--
			Allocation/100 sq km	0.19	0.52	0.57	0.50	0.14	0.21	--
Ocean quahogs	GBK	6	Area (sq km)	1663	3760	2806	1866	467	3091	--
			Mean variance	0.89	1.54	1.99	3.69	10.78	7.54	--
			Mean allocation	6	31	27	24	7	56	--
			Perfect allocation	10	29	25	23	10	54	--
			Allocation/100 sq km	0.34	0.81	0.97	1.28	1.54	1.80	--
Ocean quahogs	South	7	Area (sq km)	13037	5828	6837	11364	4689	8414	3881
			Mean variance	1.08	1.86	4.18	11.97	9.26	8.41	87.96
			Mean allocation	17	10	18	55	19	33	48
			Perfect allocation	18	11	19	52	19	33	49
			Allocation/100 sq km	0.13	0.17	0.27	0.48	0.40	0.40	1.24

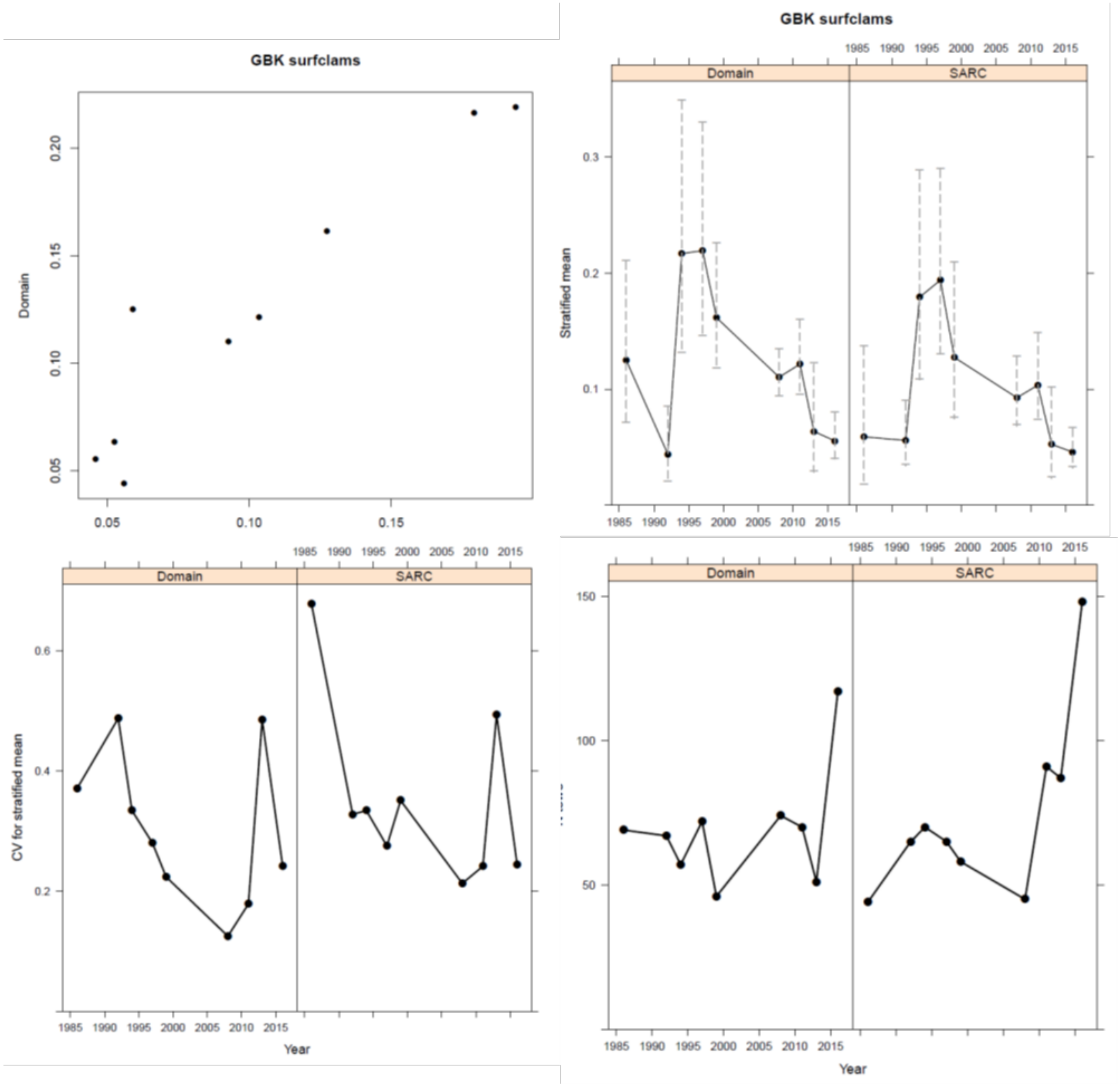


Figure 4.10 Post-stratfified means, 95% confidence intervals and CVs based on domain variance calculations and sample size (N tows) for GBK surfclams calculated using recommended (Domain) and current (SARC) stratification schemes.

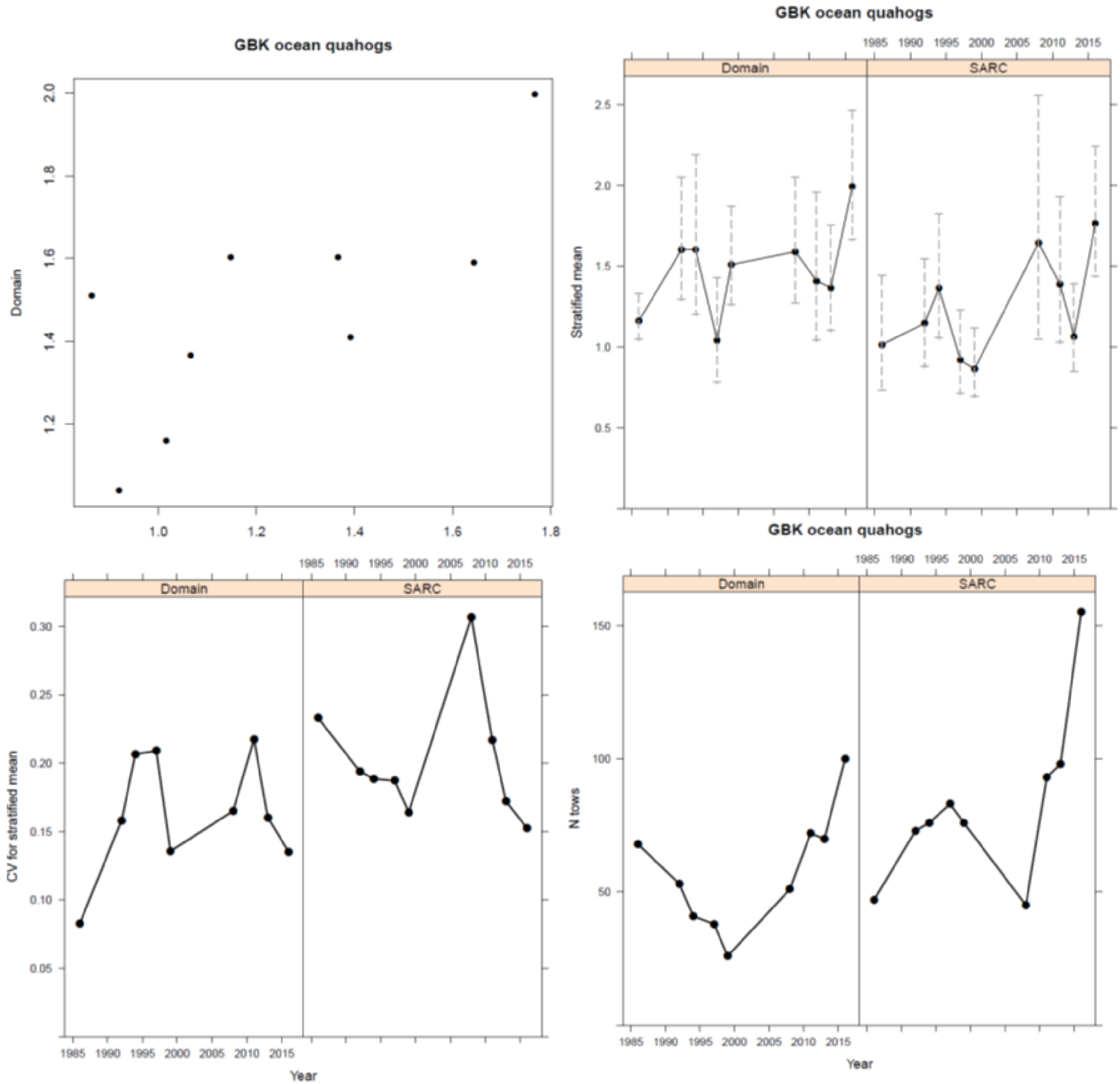


Figure 4.11 Post-stratfified means, 95% confidence intervals and CVs based on domain variance calculations and sample size (N tows) for GBK ocean quahogs calculated using recommended (Domain) and current (SARC) stratification schemes.

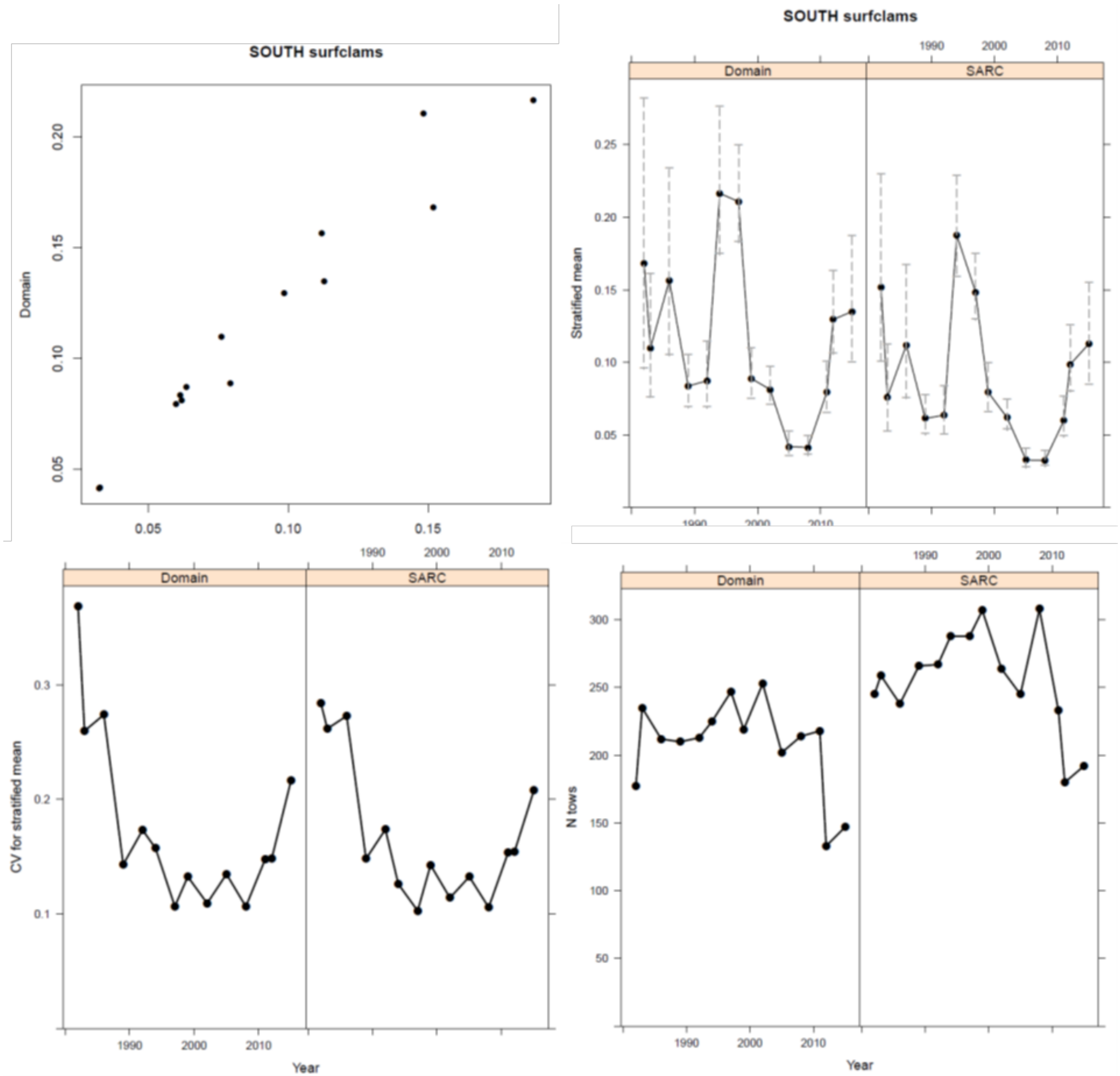


Figure 4.12. Post-stratfified means, 95% confidence intervals and CVs based on domain variance calculations and sample size (N tows) for surfclams in the southern area calculated using recommended (Domain) and current (SARC) stratification schemes.

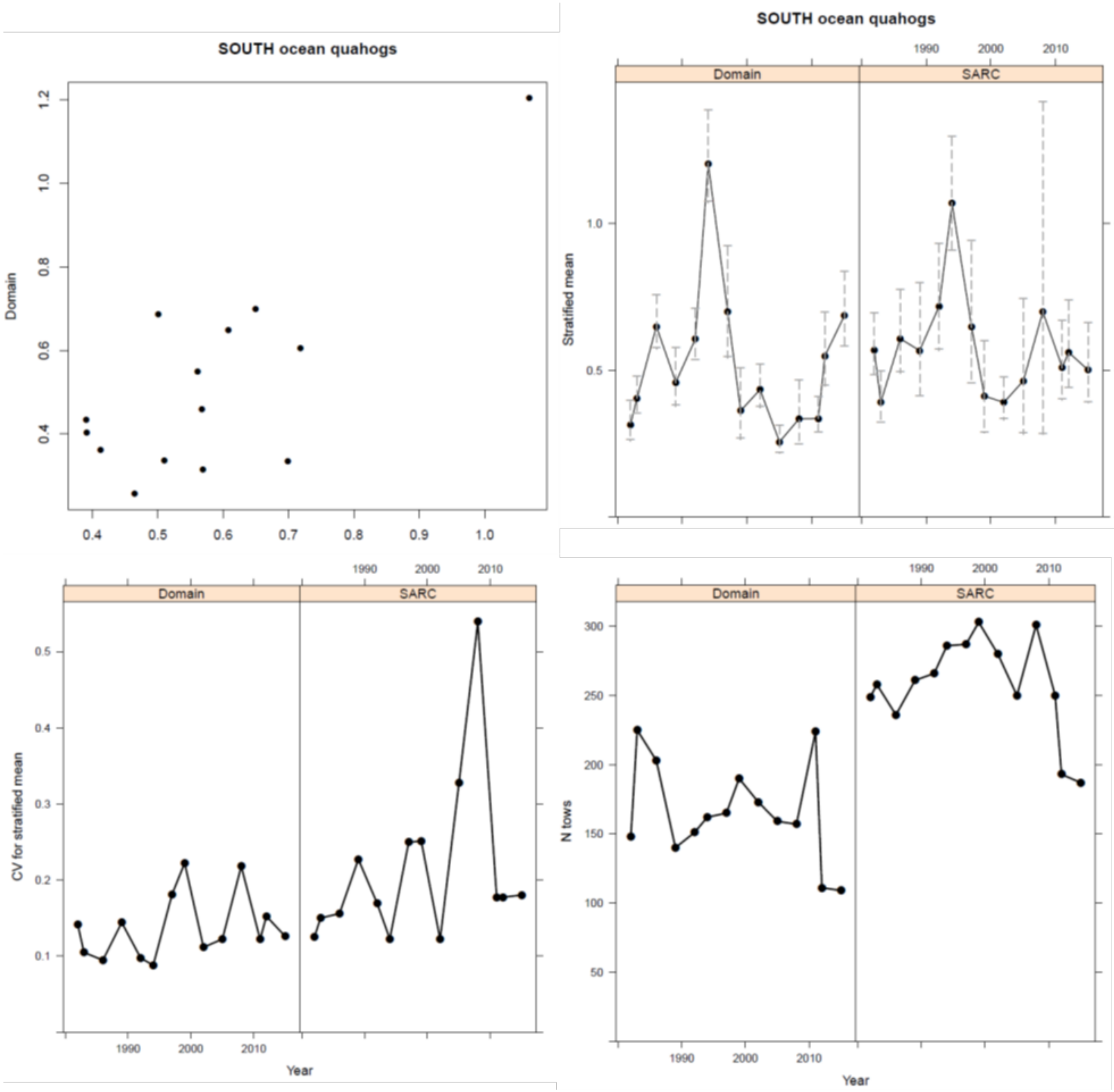


Figure 4.13. Post-stratfified means, 95% confidence intervals and CVs based on domain variance calculations and sample size (N tows) for ocean quahogs in the southern area calculated using recommended (Domain) and current (SARC) stratification schemes.

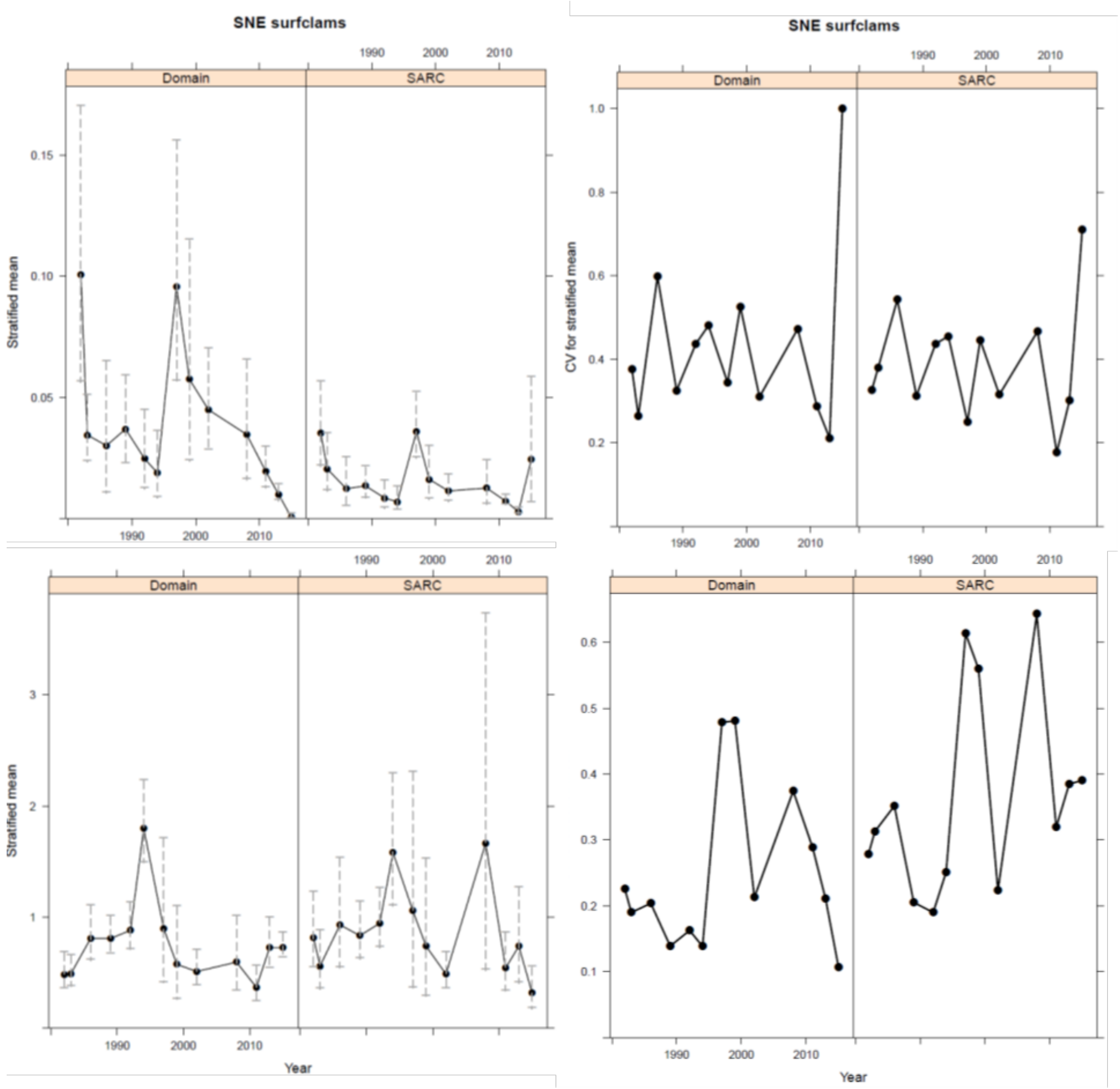


Figure 4.14. Post-stratfified means with 95% confidence intervals and CVs based on domain variance calculations for surflclams (top) and ocean quahogs (bottom) in the SNE region calculated using recommended (Domain) and current (SARC) stratification schemes.

SIMILAR RESULTS FOR LI, NJ, AND DMVSVA REGIONS

Q11) Will the recommended changes affect observation and estimation of biological characteristics, such as, length to weight relationships and growth rates?

- Will improve collection of biological data because more tows will catch target species for sampling.
- Need to adjust sampling protocols that call for a fixed number of samples per stratum because number of recommended strata much smaller than currently.

Q12) How would potential changes in the clam survey (e.g. lower survey frequency and increased precision) affect management advice and stock assessment modeling?

- Dan…

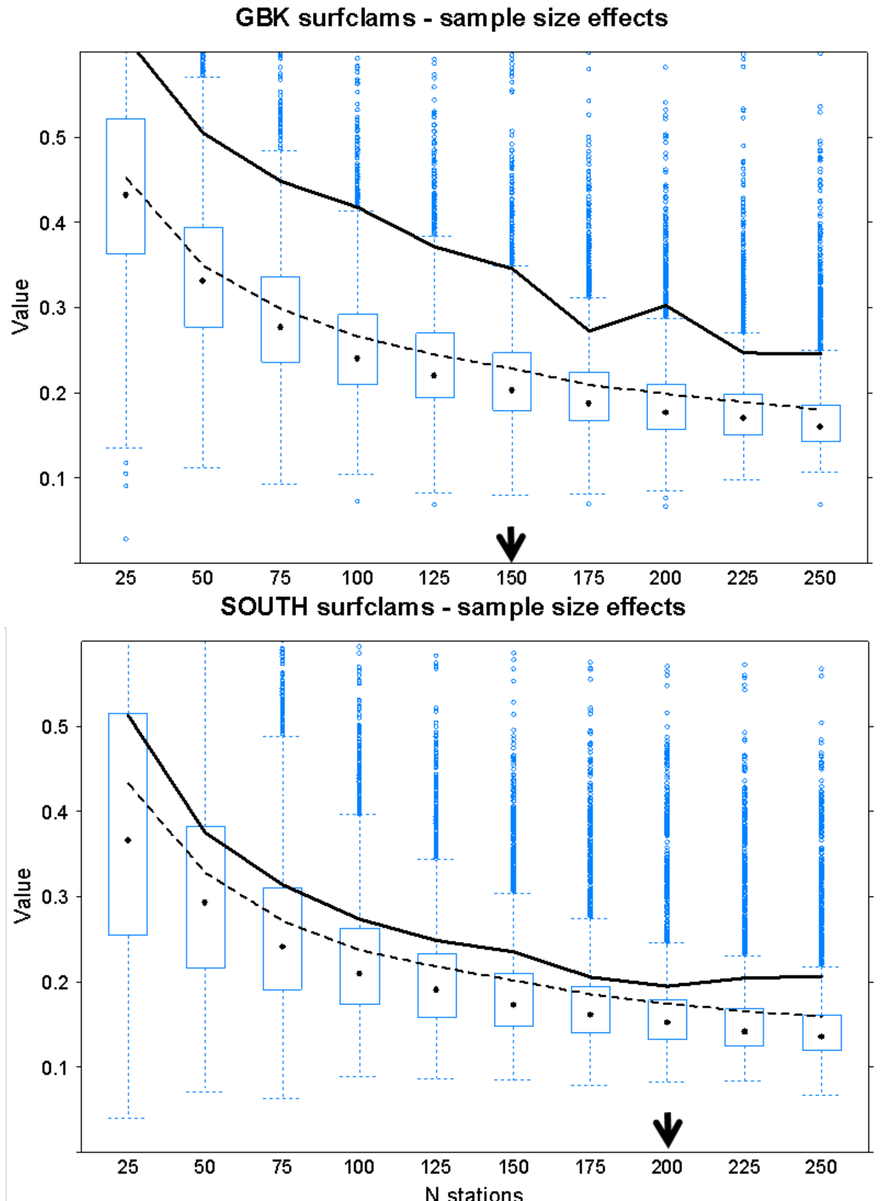
Q13) How often should future changes in stratification be considered? ***Reallocate prior to every survey. Consider stratification after ten years (three surveys for surfclams and two for ocean quahogs) but avoid changing unless benefits are clear.***

Rational: Changes to stratification likely marginal unless system changes dramatically or substantially more data or new analytical methods available. New data accumulate slowly, particularly for quahogs. Benefits from recommendations come from omitting poor habitat areas and optimal allocation—stratification less important (Smith et al. 2017). Changes are expensive.

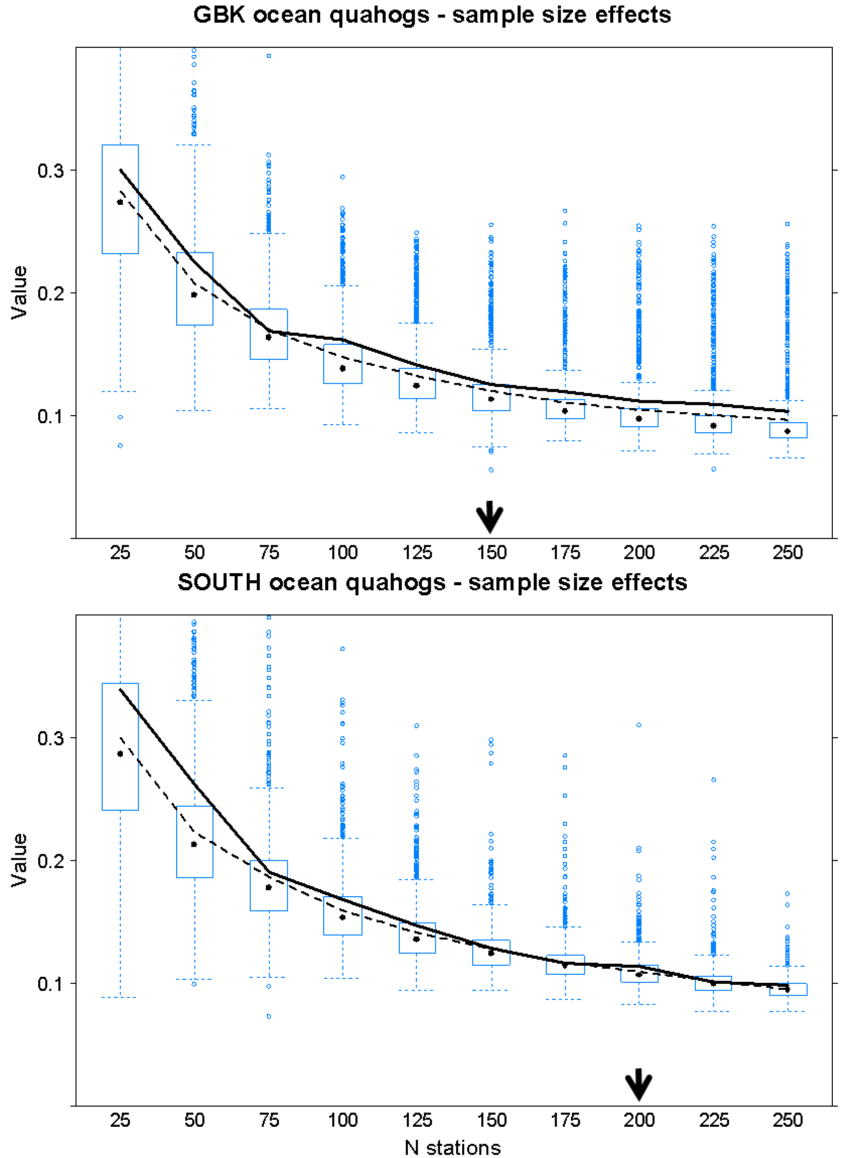
Q14) What types of additional research would benefit the clam survey?

- Augment and refine database for untrawalable ground using data from other surveys (e.g. scallops), equipment (e.g. from multi-beam sonar), etc.
- Use database to avoid gear damage in other surveys.
- Simulation studies to shorten cruise tracks and increase number of stations occupied.
- Data limitations precluded use of small building blocks (FMSQ) to identify strata. Options based on tree and GAM models with location, environmental and climatological data did not perform well but model and spatial seem promising.

What if sample sizes are reduced (for Wendy)?



Appendix Figure A2-1. Bootstrap results (2000 iterations for GBK and 1000 iterations for the southern area) showing relationships between accuracy (relative root mean squared error, solid line), average survey CV (dash line) and the distribution of survey CVs for surfclams in the GBK (top) and southern assessment areas (bottom). The symbol in the middle of the boxplots is at the median and the block shows the underlying spread. The number of tows assumed in evaluating stratification options (based on historical performance and indicated by arrows) was 150 for GBK and 200 for the southern assessment area. Note y-axis scale is different in the next figure.


GBK ocean quahogs - sample size effects
Value
0.3
0.2
0.1
25
50
75
100
125
150
175
200
225
250
SOUTH ocean quahogs - sample size effects
Value
0.3
0.2
0.1
25
50
75
100
125
150
175
200
225
250
N stations

End here….