NATURAL RESOURCES

Assessment of Sample Selection and Replacement for Eelgrass Monitoring

May 10, 2017

PUGET SOUND ECOSYSTEM MONITORING PROGRAM

WASHINGTON STATE DEPT OF NATURAL RESOURCES

The Submerged Vegetation Monitoring Program (SVMP) is funded by the Washington State Department of Natural Resources as part of the agency's work as steward of public lands to ensure environmental protection (http://www.dnr.wa.gov/). It is a component of the Puget Sound Ecosystem Monitoring Program (PSEMP) (http://sites.google.com/site/pugetsoundmonitoring/).

Assessment of Sample Selection and Replacement for Eelgrass Monitoring

May 10, 2016

Pete Dowty
Bart Christiaen
Helen Berry

Nearshore Habitat Program
Aquatic Resources Division

PUGET SOUND ECOSYSTEM MONITORING PROGRAM

WASHINGTON STATE DEPT OF NATURAL RESOURCES

Acknowledgements

The Nearshore Habitat Program is part of the Washington State Department of Natural Resources' (DNR) Aquatic Resources Division, the steward for state-owned aquatic lands. Program funding is provided through the Aquatic Lands Resource Management Cost Account (RMCA). The Nearshore Habitat Program monitors and evaluates the status and trends of marine vegetation for DNR and the Puget Sound Partnership.

Washington State Department of Natural Resources
Aquatic Resources Division
1111 Washington St. SE
P.O. Box 47027
Olympia, WA 98504-7027

www.dnr.wa.gov

Contents

Executive Summary

The Washington State Department of Natural Resources (DNR) is steward of 2.6 million acres of state-owned aquatic land. DNR manages these aquatic lands for the benefit of current and future citizens of Washington State. As part of this responsibility, DNR's Submerged Vegetation Monitoring Program (SVMP) has conducted annual monitoring of eelgrass (*Zostera marina*) in greater Puget Sound since 2000. Eelgrass is both an important habitat for valued fauna and a sensitive indicator of environmental degradation. Data from the monitoring program form the basis of an ecosystem Vital Sign reported by the Puget Sound Partnership as well as performance measures tracked by Results WA (www.results.wa.gov) and internally by DNR.

The SVMP monitoring effort relies on a detailed and statistically rigorous sampling design in order to produce reliable estimates. The focus of this report is SVMP site sampling with underwater video transects. Since the inception of the program in 2000, new transects have been selected within sites for each annual sampling event using simple random sampling (SRS). This is known as sampling with replacement as the sample is replaced with a new sample for each occasion. It was observed that SRS can produce clumping in the distribution of transects. Clumping can increase the variability in eelgrass area estimates when clumped transects fall in locations with relatively low or high eelgrass cover. Over multiple sampling occasions, this discrepancy can make change detection difficult if newly drawn SRS transects fall in portions of the bed with different eelgrass cover. To address these concerns, the SVMP began to test alternate transect selection and replacement methods in 2012, including:

- Selection approaches that spatially distribute transects. The primary method tested has been stratified random sampling with one unit per stratum (STR). This method places one transect within each of a number of equally sized subsections of the site. Each transect is selected by SRS within its subsection. Systematic sampling (SYS) has also been used in limited cases where the first transect is selected by SRS within the first subsection and the other transects are equally spaced through the other subsections.
- Repeated sampling over time of the same transect sample. This eliminates the original approach of total sample replacement each sampling occasion.

This study evaluates the performance of the alternate transect selection and replacement methods through three linked modelling tasks:

- Develop spatial models of the eelgrass at two contrasting sites based on existing SVMP monitoring data.

- Develop a series of change scenarios for testing the ability of different sampling and analysis methods to detect the change between two sampling occasions.
- Conduct Monte Carlo sampling of the site models to evaluate the performance of the alternative methods.

The key findings from this study include the following.

1. Repeated surveys of the same sample across occasions (sampling without replacement) greatly improved precision of change estimates and the power to detect change as compared to a newly drawn sample each occasion (sampling with replacement).
2. Estimates of site eelgrass area and change in eelgrass area had superior precision when based on sample selection with STR or SYS as compared to SRS.
3. Each selection method studied (SRS, STR, and SYS) produced unbiased estimates of site eelgrass area and change.
4. The power of STR and SYS to detect change in eelgrass area was relatively resilient across a range of change scenarios that differed in spatial heterogeneity. In contrast, power under SRS was strongly degraded for some scenarios.
5. The power of STR and SYS to detect change in eelgrass area between two sampling occasions was highly sensitive to the method used to estimate variance. The two variance estimators studied were inconsistently biased. The consequences of this bias varied, but may include diminished power to detect change, inaccurate confidence intervals and elevated risk of false detections of change (Type I error). The SVMP will need to develop a reliable approach to variance estimation in order to transition to STR and realize the potential it offers.
6. Contrary to what has been stated previously in the literature, STR was found to be weakly sensitive to spatial periodicity in the population sampled. This still represents a strong advantage over SYS which is highly sensitive to periodicity.

In order to develop STR sampling and analysis methods for widespread implementation, several areas were identified for further work.

a) Construct additional site models from contrasting sites. This work has progressed in related projects.
b) Assess if the v8 variance estimator included in this study can be revised to produce consistent results for SVMP site eelgrass populations.
c) Test other variance estimators designed for SYS to assess bias when applied to a range of site models.
d) Investigate the performance of STR with transect–based site trend analysis.

While completing this study, we were struck by the long-term potential of expanding the analysis framework to consider spatial relationships among transects at each site. A model-based perspective could be developed for each site, based on site models similar to those developed in this report. This new paradigm could enrich analysis through considering the spatial characteristics of the eelgrass at each site.

1 Introduction

The Washington State Department of Natural Resources (DNR) is steward of 2.6 million acres of state-owned aquatic land. DNR manages these aquatic lands for the benefit of current and future citizens of Washington State. As part of this responsibility, DNR's Submerged Vegetation Monitoring Program (SVMP) has conducted annual monitoring of eelgrass (*Zostera marina*) in greater Puget Sound since 2000. Eelgrass is both an important habitat for valued fauna and a sensitive indicator of environmental degradation. Annual progress in the monitoring work provides an internal DNR performance measure and the estimated Puget Sound abundance of eelgrass is the basis of ecosystem indicators reported by Results WA (www.results.wa.gov) and the Puget Sound Partnership (www.psp.wa.gov).

The monitoring conducted by the SVMP is extensive and relies on a detailed sampling design in order to produce reliable estimates. The basic approach is to randomly select a sample of sites from the greater Puget Sound study area and then sample the eelgrass at each selected site with underwater video transects that run perpendicular to the shoreline. The SVMP staff continually evaluate the performance of the sampling design and potential improvements that might lead to greater precision and greater ability to detect change occurring in the eelgrass population.

The focus of this report is the SVMP site sampling with transects. Since the inception of the program in 2000, transects have been selected using simple random sampling (SRS). It was observed that SRS can produce clumping in the distribution of transects – an effect that has been well-described in the allocation of sample units to different experimental treatments (Bailey 1987). This can lead to samples with poor representativeness of the site. Starting in 2012, the SVMP began to test an alternate transect selection method that gave better spatial distribution of transects. This alternative selection method is stratified random sampling with one unit per stratum (STR). It can be thought of as an intermediate method between SRS and systematic sampling (SYS). SYS has been implemented at a limited number of sites for gradient sampling related to suspected stressors. The three selection methods are illustrated in Figure 1-1.

The method of sample selection directly affects the performance of the analytical results of a sampling design. As STR was developed and implemented at SVMP special study sites, it was important to ask how STR affects the associated estimates. Especially since the SVMP is evaluating the wholesale replacement of SRS with STR for site sampling across the core monitoring program, it is critical to understand the performance of estimates derived from STR samples. The purpose of this report is to address this issue.

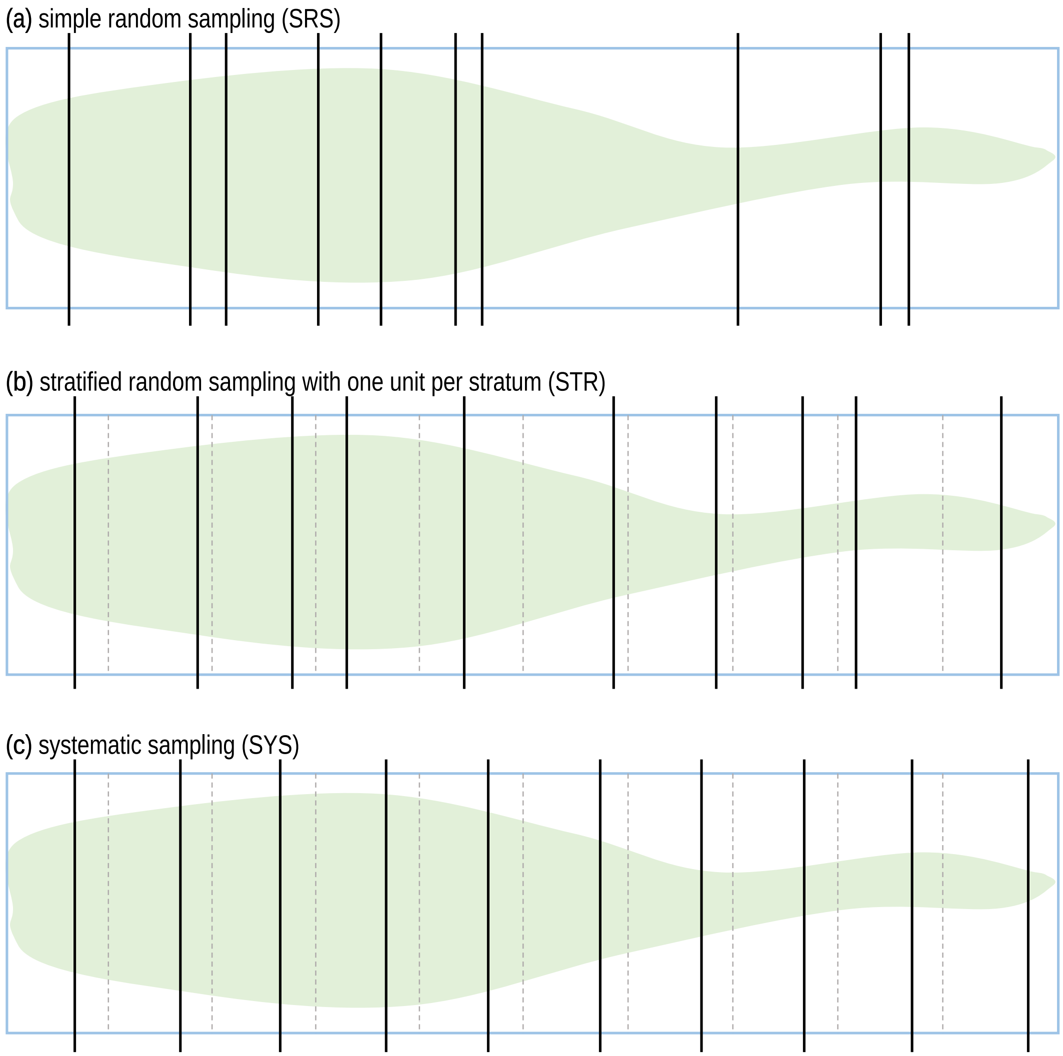


Figure 1-1. Illustrations of the three transect selection methods used in this study for a sample size of 10 transects. The outer rectangle represents the boundary of the site being sampled. The green shaded area represents the eelgrass distribution at the site. The vertical black lines represent the selected transects. The vertical gray dashed lines represent boundaries between the 10 equally sized subsections of the site used for transects selection for STR and SYS. In SRS (a), each transect is randomly selected from across the entire site. In STR (b), each transect is randomly selected from within one site subsection. In SYS (c), the first transect is selected randomly within the first subsection and the other transects have the same location within their subsections leading to equally spaced transects.

1.1 Previous Work

A small number of studies in the literature were consulted to understand the precedent for the use of STR and to provide direction as to analysis options. Cochran (1946) was the earliest reference found that addresses STR, or "stratified random sampling with one element per stratum". In comparing both STR and SYS to SRS, Cochran found that for the correlated population he studied, STR is "always at least as accurate on the average as [SRS]." It is notable that Cochran (1946) did not address the challenge of estimating variance with STR other than noting that neither STR nor SYS "provides the data for an unbiased estimate of the sampling variance of the sample mean." This study did address the sensitivity of SYS to periodicity in the population.

Cochran again addressed STR in his seminal reference text (Cochran 1977). Therein Cochran labelled the method as "stratified random sampling with one unit per stratum". He dealt with this as a special case of conventional stratified random sampling which he described in a separate chapter. He did address the challenge of estimating variance from an STR sample and presented the "collapsed strata" method as an alternative. The collapsed strata method was tested in this study.

Saunders and Robinson (1989) provide a more recent reference to STR. It is invoked only as a contrast to the sampling method that is the focus of their study. They do make reference to geosciences literature that appears to address more comprehensively the issue of variance estimation with STR. A theoretical treatment of variance estimation is referenced that relies on determination of the variogram for the population being sampled.

In addition to the above studies that reference STR, a few studies that focused on SYS were helpful in the work reported here. Wolter (1984) presented a set of eight alternative estimators for SYS (labelled v1 through v8) and evaluated their performance on a number of simulated populations. Although he identified two estimators for use "if all else fails" (v2 and v3), none of the estimators had consistently good performance across the populations studied.

Skalski et al. (1993) compare SYS to conventional stratified random sampling and several variations on these designs. Their general recommendation was to implement conventional stratified random sampling over SYS due to problems with variance estimation and the unpredictable response of performance to sample size.

McGarvey et al. (2016) had an important influence on this study. They found that SYS was consistently more precise than SRS. They studied the v1-v8 estimators of Wolter (1984) and some additional estimators of variance under SYS. McGarvey et al. was particularly relevant here because they modelled more spatially complex populations that were representative of vegetation patterns. Most importantly, McGarvey et al. concluded with a recommendation for Wolter's v8 estimator. Given this recommendation and the shared context with the current study (vegetation spatial pattern), the v8 estimator was selected for evaluation in this study. McGarvey et al. did not address the sensitivity of SYS to periodicities in the population.

1.2 Objectives

The objectives in this study were to address the following questions:

- Has this STR method of sample selection been described in the literature?

- If there is previous work, does it suggest that STR sampling would be preferable to SRS for SVMP transect sampling?
- What are the effects of repeatedly sampling the same transects over time compared to sampling a new draw of transects each occasion?
- Is it valid to treat the STR sample as an SRS sample for the purposes of analysis – i.e., for estimating a population parameter and its variance and conducting statistical testing?
- If standard (SRS) estimators are not appropriate for STR samples, are alternative estimators available?
- Is STR sensitive to periodicity in the population in the same way as SYS?

1.3 Approach

- Limited search to see if STR is described in the literature.
- Create site models from existing SVMP data for two sites with contrasting patterns of the spatial distribution of eelgrass.
- Simulate SRS, STR and SYS sampling from the site models and compare the performance of estimates. SYS is included here to provide an additional reference point for evaluating STR.
    - Compare site eelgrass area estimates on a single sampling occasion
    - Compare the detection of difference between two site eelgrass area estimates based on independent samples (new random draw of transects on the second sampling occasion).
    - Compare the detection of difference in site eelgrass area between two occasions based on paired analysis of repeated sampling of the initial sample on the second occasion.
    - Compare performance with spatially homogeneous change and spatially heterogeneous change.
    - Assess effects of periodicity in the population.

Throughout this report the expression "sampling with replacement" is used interchangeably with the use of "new draw transects" to mean the use of newly drawn transects for each sampling occasion. Similarly, the expression "sampling without replacement" is used interchangeably with the use of "repeat transects" to mean repeated sampling of the same sample of transects on each sampling occasion.

2 Site Area Estimates with Standard Estimators

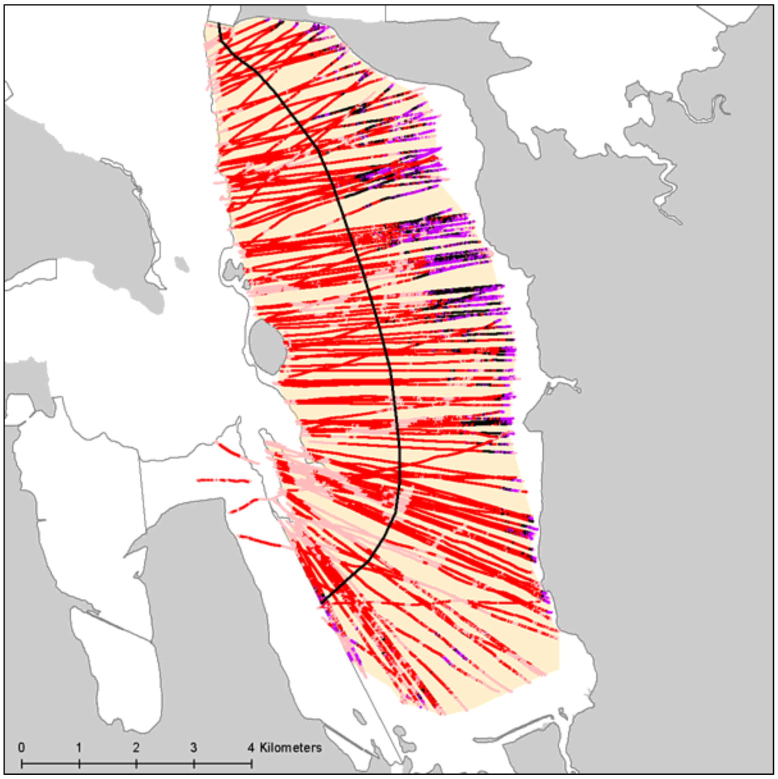
2.1 Methods Overview

- Build two static site models based on existing SVMP transect data – one relatively homogeneous site (core001 – Padilla Bay) and one relatively heterogeneous site (flats26 – Snohomish Delta North).
- Conduct Monte Carlo sampling of the sites using SRS, STR and SYS transect selection with a sample size of $n = 10$ transects.
- Compare accuracy and precision of site area estimates with the different selection methods but with the same standard SVMP estimator for site area, i.e., the estimators developed for SRS sampling (Skalski 2003).
- Compare accuracy and precision of estimates of standard error on the site area estimates. Compare results with the different selection methods but with the same standard SVMP estimator for standard error.

2.2 Site Models

Site models for core001 (Figure 2-1, Figure 2-2) and flats26 (Figure 2-1, Figure 2-3), each defined by 1000 1 meter-wide transects, were generated as follows:

- Transect point data were extracted from the 2000-2014 SVMP database for core001 (2001-14 data) and for flats26 (2005-2009 data). Only random transects were selected and transects 4 and 7 from core001 in 2001 were discarded because of anomalous orientation. This resulted in 145 transects for core001 and 55 transects for flats26.
- A 3-meter buffer around the median lines were used to select enclosed points for each transect. The y-coordinate of these selected points were used to represent the relative linear position of the transects in the longshore dimension.
- Transect fraction and length were joined from the transect results table.
- The transect positions were mapped to a [0,999] interval keeping relative spacing along the median line intact.
- Transect fraction and length values were linearly interpolated between the SVMP transects to produce site models consisting of 1000 1 meter-wide transects.



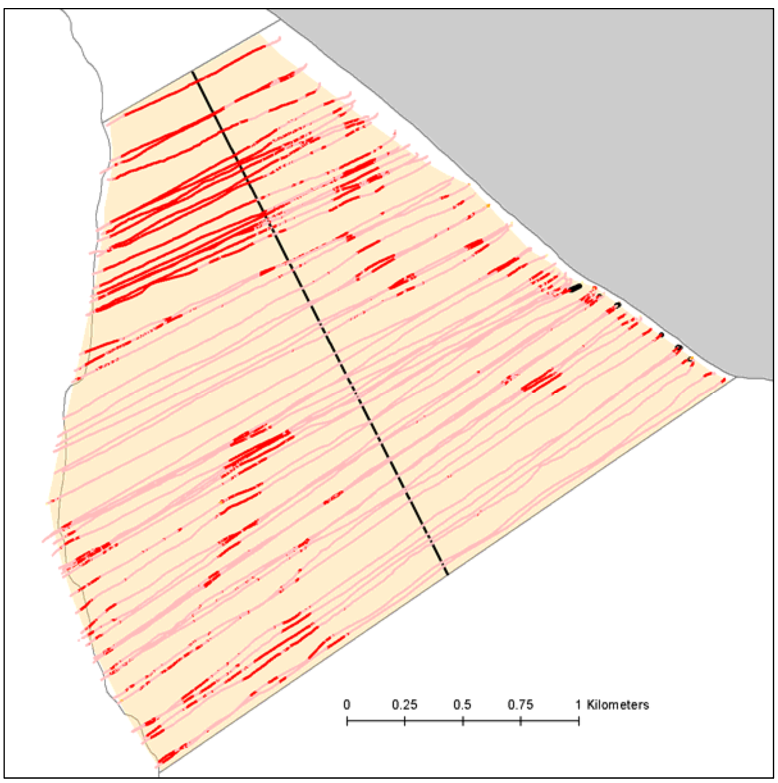


Figure 2-1. SVMP site data for core001 (top; 2001-14) and flats26 (bottom; 2005-09) used to generate site models. Transects show observations of ***Z. marina*** (red), ***Z. marina/japonica*** mixed (black) and ***Z. japonica*** (purple). The black median lines were used to linearize transect position.

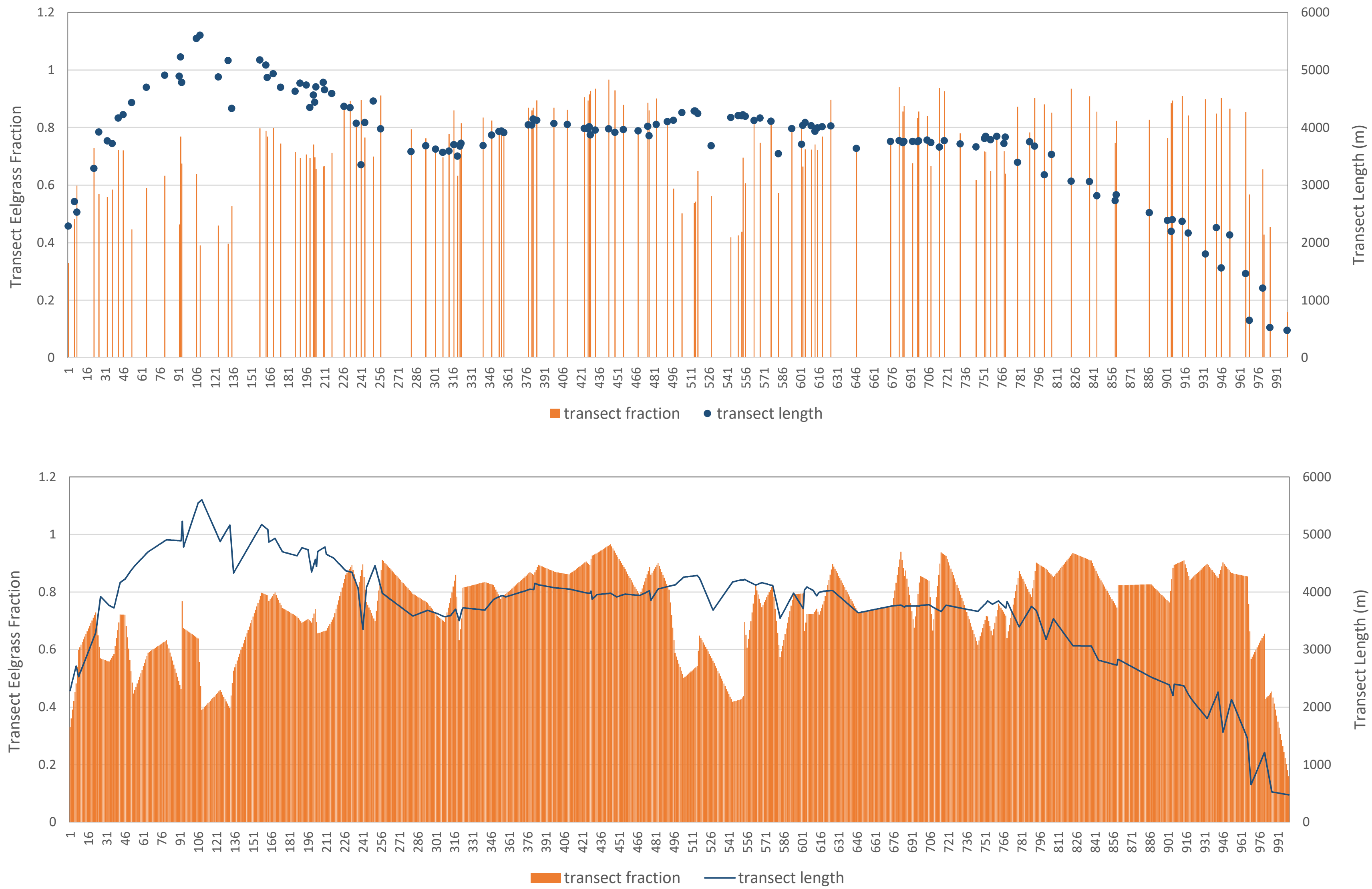
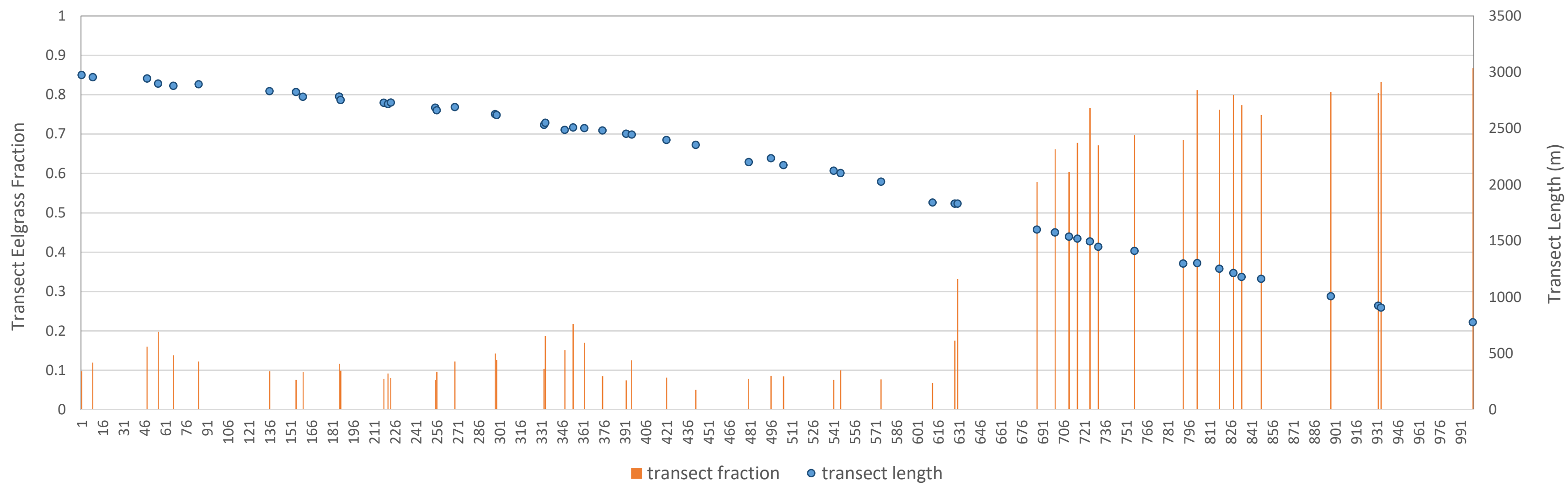


Figure 2-2. Core001 data (top) and site model (bottom). The ***x***-axis represents longshore position with 1000 1 meter –wide transects in the site model.



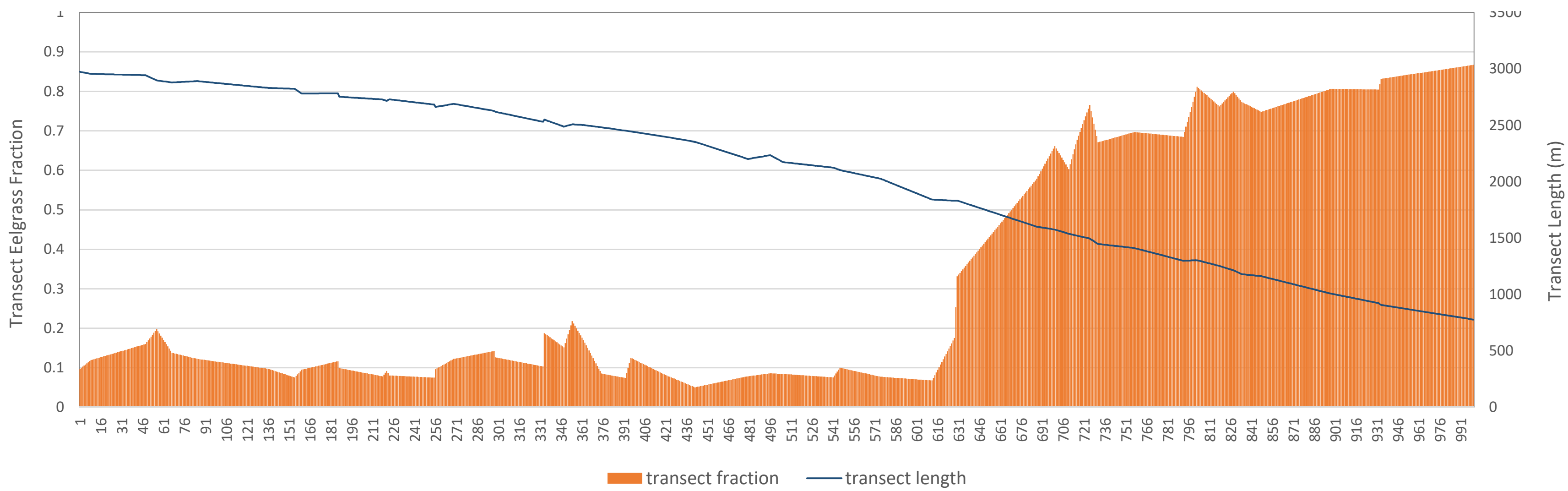


Figure 2-3. Flats26 data (top) and site model (bottom). The ***x***-axis represents longshore position with 1000 1 meter –wide transects in the site model.

2.3 Simulated Sampling

Transect sampling was simulated by Monte Carlo sampling from the site models with a small amount of Gaussian noise added to represent measurement error (video classification error). Each model run included n = 5000 iterations producing 5000 simulated samples. Using these populations of samples, the precision and accuracy of sample estimates of site area and standard error were evaluated.

For each model iteration 10 transects were selected to constitute a sample. The selection methods were implemented as follows:

SRS each transect was randomly selected from the 1000 1 meter-wide transects in the site model (sampling with replacement).

STR each transect i for i = 0, 9 was selected from the set of transects ranging from 0+100 i to 99+100 i.

SYS an initial transect, k, was randomly selected from transects 0 to 99, and the remaining 9 transects (i = 1,9) were selected as $k + 100i$.

The eelgrass fraction values for each transect in a sample were determined by the actual model value for the 1 meter-wide transect selected with an added noise factor drawn from the normal distribution N(0, s.e.=0.0125) to represent video classification error. This level of error essentially gives a 95% confidence interval on each transect fraction observation of total width 0.05 (±0.025).

For each iteration, the sample estimates of site area and standard error were calculated using the standard SVMP calculation and the simulated sample for that iteration.

"True" site area of each model was determined using the standard SVMP calculation (mean fraction × sample polygon area) with all 1000 1 meter-wide transects. The most recent sample polygon area was used for core001 (2014) and flats26 (2009).

"True" standard error for the site area estimator associated with each transect selection method was specified by the standard deviation of the site area sample estimates from the 5000 iterations.

2.4 Results – Precision and Accuracy

First, normality of the distributions of estimates was assessed by inspection of sampling distributions and Quantile-Quantile (QQ) plots, which plot data against data expected under the normal distribution. For the site area estimates, the frequency histograms show clear departures from normality that are greatest for SYS sampling and when sampling from the heterogeneous site model (flats26) (Figure 2-4 and Figure 2-5). The estimates based on STR samples conform most closely to normality.

For the standard error estimates, the departures from normality are stronger (Figure 2-6 and Figure 2-7). Again the estimates from SYS sampling have relatively large departures from normality but surprisingly the estimates from SRS also have large departures because of non-conformance in the tails. The estimates based on STR samples again conform most closely to normality when sampling from the core001 model. All selection methods lead to departures from normality when sampling from the flats26 model.

When these distributions are compared in the form of boxplots against the true values of site eelgrass area and standard error of the mean estimate, then the sampling methods can be readily compared in terms of precision and bias (Figure 2-8).

Note that the box plots used in Figure 2-8 are relatively insensitive to extended tails. This is because the lower and upper whiskers indicate the 10^{th} and 90^{th} percentiles of the sampling distribution rather than other percentiles (e.g., 1^{st} and 99^{th}) that would be more sensitive to extended tails. The standard error, in contrast, is sensitive to extended tails. As a consequence, the extended upper tail in the distribution of standard error estimates at core001 (Figure 2-6 a) elevates the standard error relative to STR and SYS sampling (Figure 2-8 i) but the range spanned by the box is roughly comparable across SRS, STR and SYS (Figure 2-8 g).

Some key findings from these comparisons include:

- There is a marked gain in precision when sampling with STR and SYS as compared to SRS. This applies to both estimation of site area (Figure 2-8 c, d) and of standard error (Figure 2-8 i, j) and is particularly strong when sampling from the flats26 site model.
- STR and SYS gave similar levels of precision but SYS was slightly more precise in some cases (e.g., Figure 2-8 c).
- Site area estimates are unbiased for all selection methods except for a small bias (-3%) with SRS sampling from the flats26 model (Figure 2-8 e, f).
- Standard error estimates are unbiased only for SRS sampling (Figure 2-8 k, l).
- Standard error estimates with STR and SYS are strongly biased especially when sampling from the flats26 site model where bias surpasses +400% (Figure 2-8 l).

These results suggest that STR samples cannot be analyzed with the standard variance estimator to produce reliable estimates. Moreover, the bias that would be introduced cannot be attributed to departures from normality.

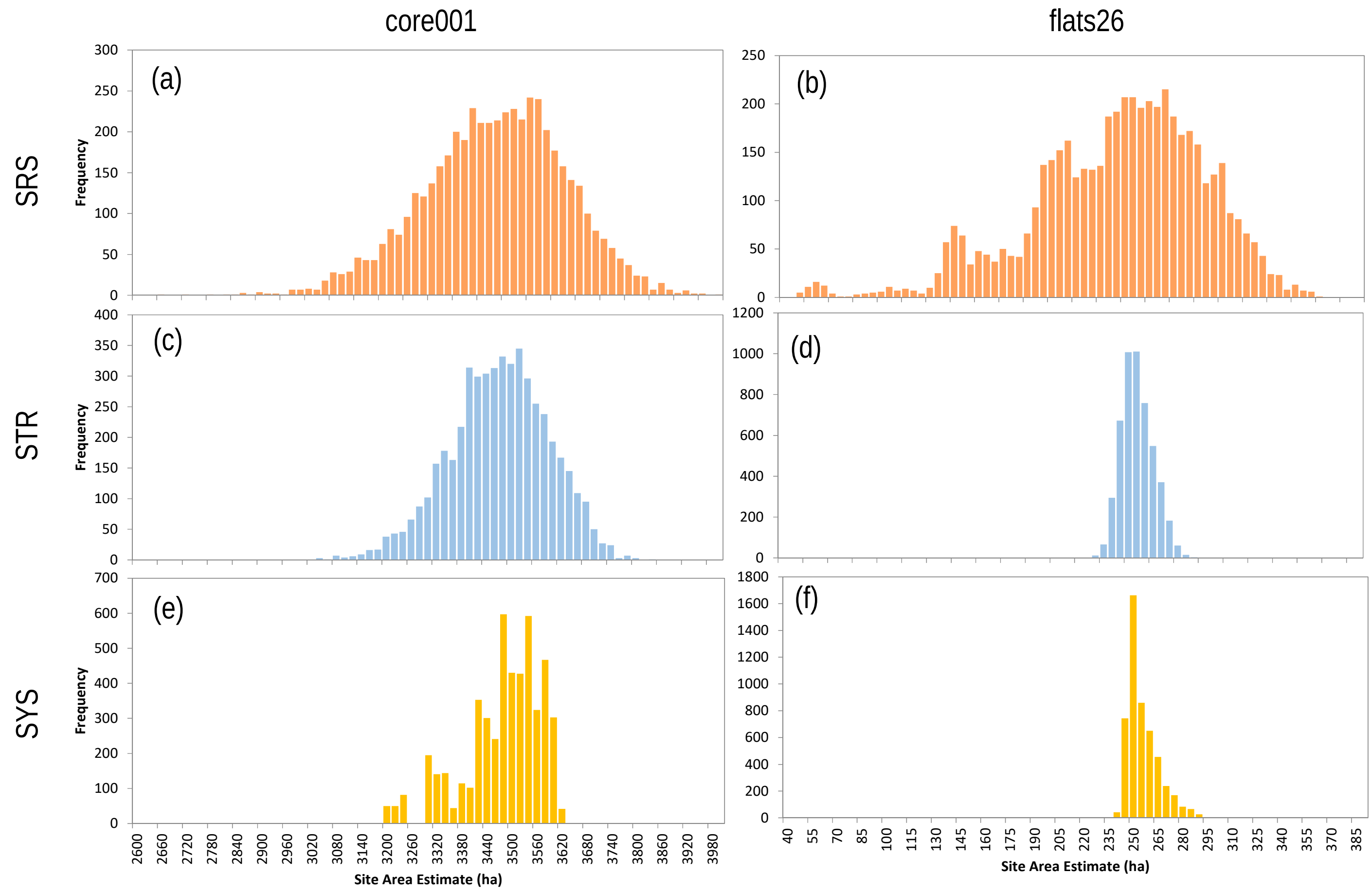


Figure 2-4. Site Area Estimates. Frequency histograms of site area estimates (n=5000) based on sampling from the site models for core001 (left) and flats26 (right) with transect selection methods SRS (top), STR (middle) and SYS (bottom). All estimates were based on the standard SVMP estimator for site area which assumes SRS. The degree of dispersion in these distributions represents precision of the sample estimates of site area. The clear departure of the sampling distribution from normality in the case of SYS is also reflected in the QQ plots (Figure 2-5). The QQ plots for SRS indicate a modest departure from normality.

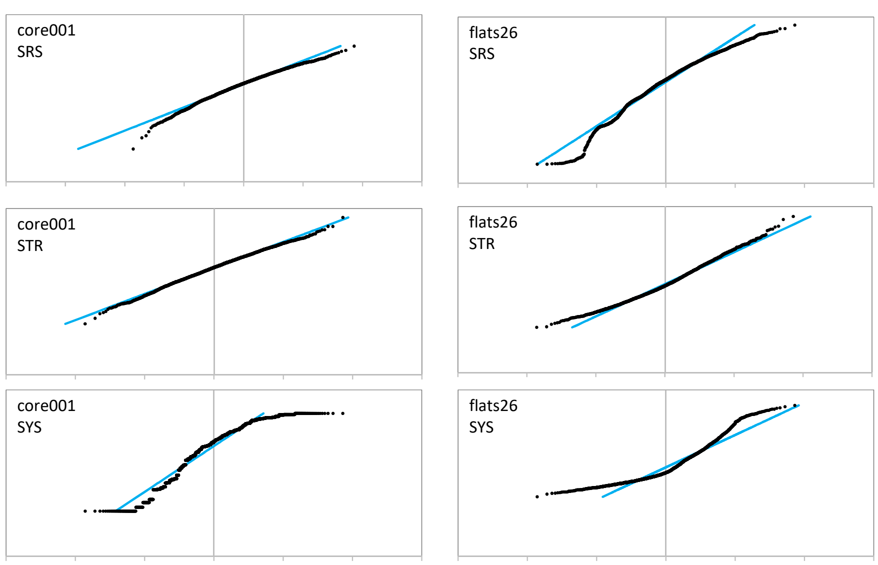


Figure 2-5. Normal QQ plots of the estimates of site eelgrass area shown in Figure 2-4. The distribution of site area estimates (on the vertical axis) are compared to the normal distribution (the horizontal axis). A normally distributed data set would fall on the blue line.

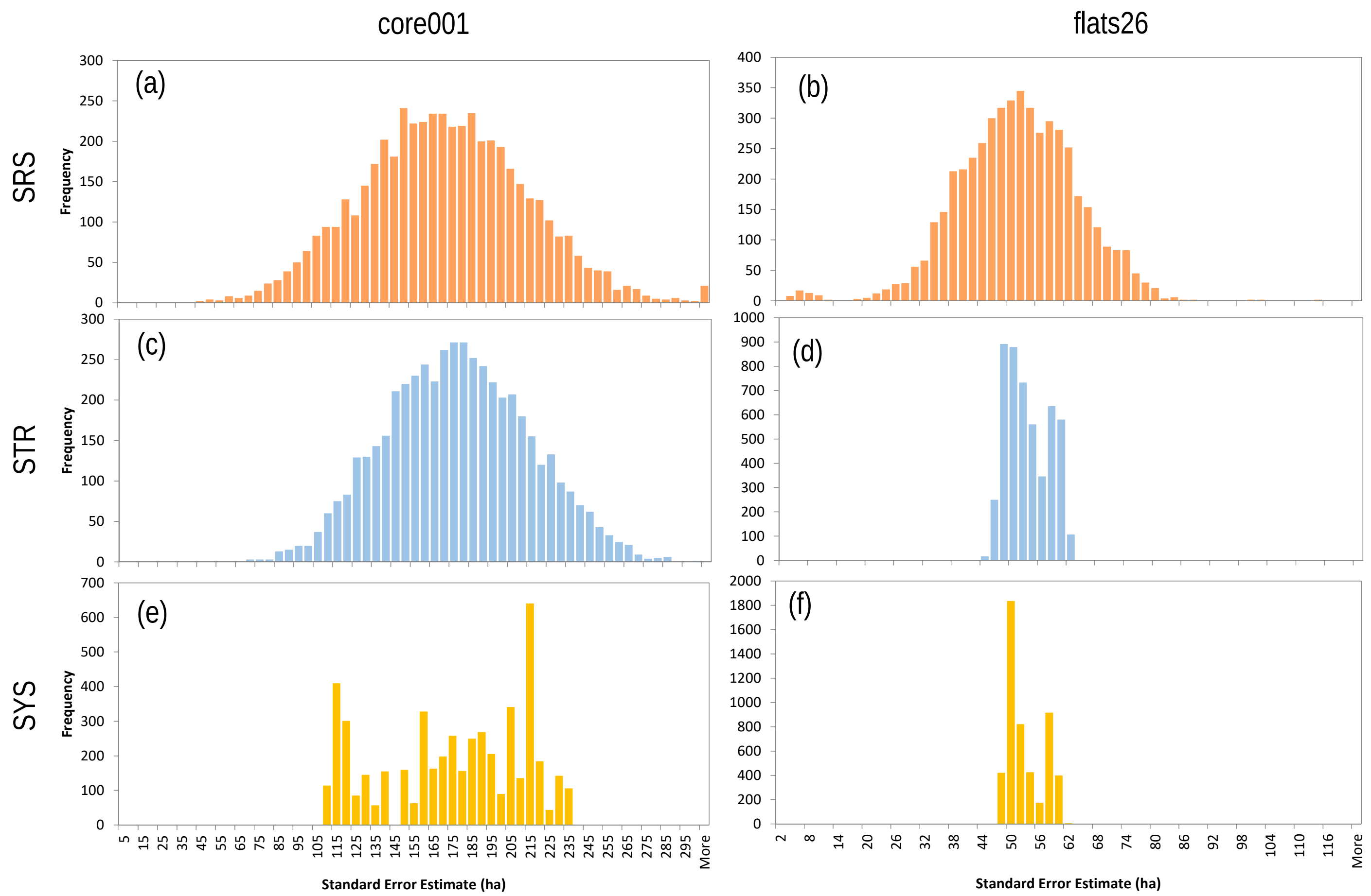


Figure 2-6. Site Standard Error Estimates. Frequency histograms of site standard error estimates (standard error of site area estimate) (***n***=5000) based on sampling from the site models for core001 (left) and flats26 (right) with transect selection methods SRS (top), STR (middle) and SYS (bottom). All estimates were based on the standard SVMP estimator for standard error which assumes SRS. The degree of dispersion in these distributions represents precision of sample estimates of standard error on the site area estimates based on the different sample selection methods.

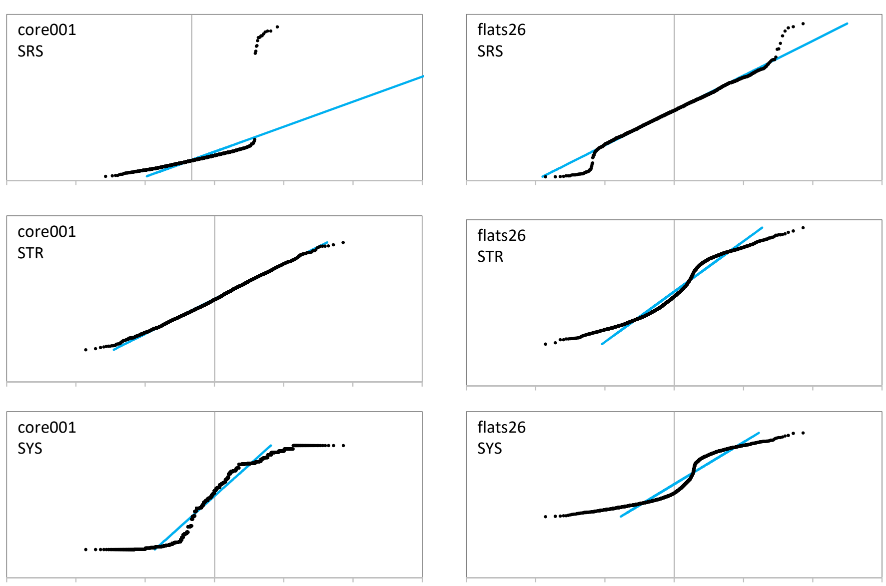


Figure 2-7. Normal QQ plots for the site area standard error estimates shown in Figure 2-6.

Estimation of Eelgrass Area

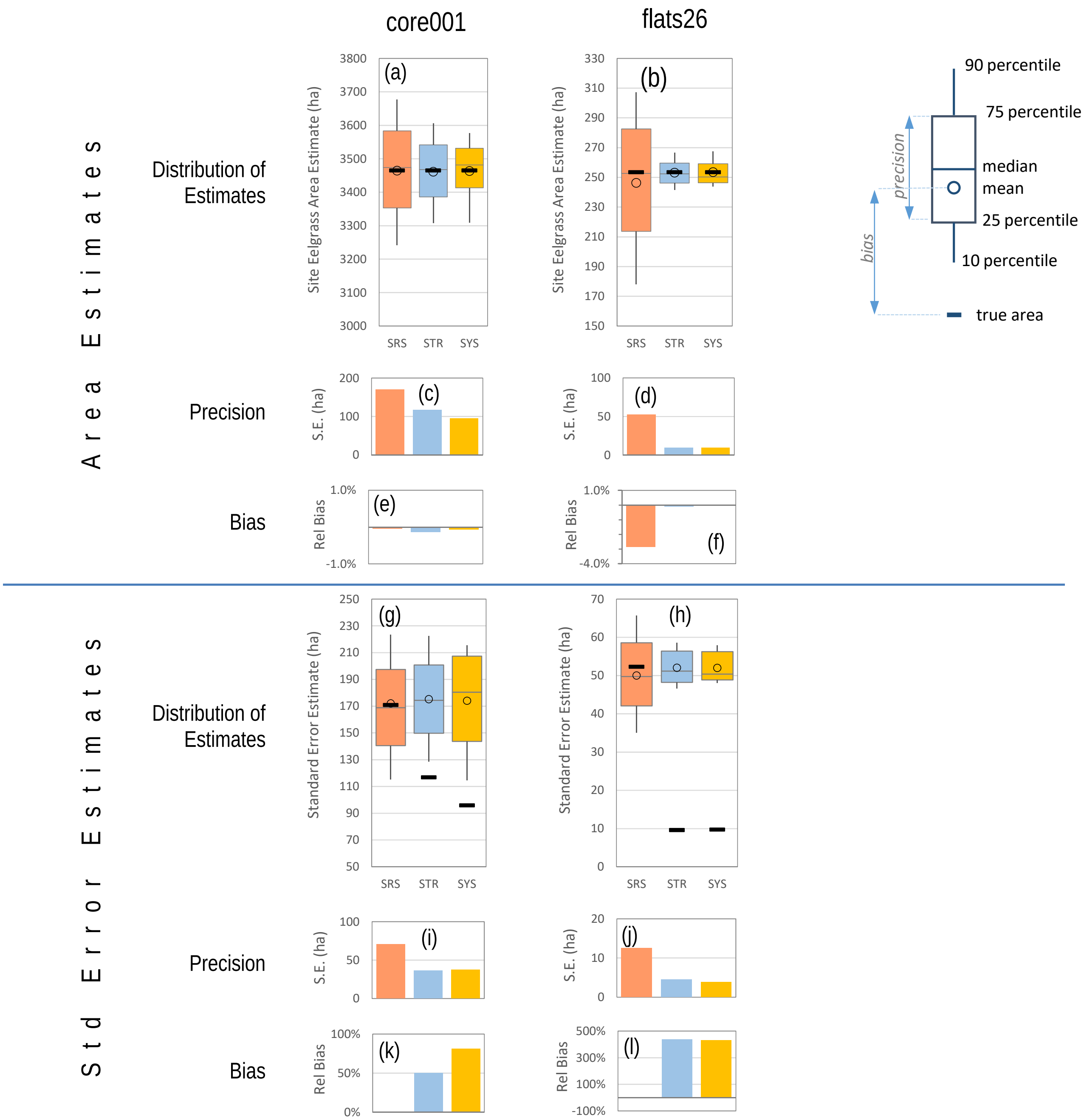


Figure 2-8. The distribution of estimates and the associated precision and bias for estimates of eelgrass area (top half) and for estimates of the standard error of the eelgrass area estimate (bottom half). The estimates are based on samples drawn from the site models for core001 (left) and flats26 (right) and with sample selection by SRS (red), STR (blue) and SYS (yellow).

3 Detecting Homogeneous Change with Standard Estimators

3.1 Site Change Models

A simple model of change in site eelgrass area was developed which applies a spatially homogeneous model of change across the site between two sampling occasions. The model is specified with two parameters – a nominal relative change and a CV representing variability of the change value applied to individual transects of the site model. Model runs were conducted with nominal change values of 0, -0.05, -0.10, -0.15, -0.20 and -0.25. The CV of the transect-level change values about these nominal values was fixed at 0.2. The resulting site change models are shown in Figure 3-1 (core001) and Figure 3-2 (flats26).

3.2 Change Estimators

3.2.1 New Draw Transects – Difference in Site Area Estimates

When a new sample of transects is drawn on the second sampling occasion, change is assessed based on comparing the site eelgrass area estimate from the first occasion to that of the second occasion. The standard estimators of Skalski (2003) are used (Skalski's equation 1 for estimating site area and equation 2 for estimating variance). If X_i is the estimate of site eelgrass area for occasion *i*, then the difference in site area between occasion 1 and occasion 2 is estimated by

$$\Delta X = X_2 - X_1 . \qquad \text{Equation 3-1}$$

If s_{X_i} is the estimate of standard error of the estimate X_i, then the standard error on ΔX is estimated by

$$s_{\Delta X} = \sqrt{s_{X_1}^2 + s_{X_2}^2} \qquad \text{Equation 3-2}$$

3.2.2 Repeat Transects –Difference in Area from Paired Transects

Here change is based on the arithmetic difference in eelgrass fraction between the initial conditions and the second sampling occasion for paired transects. This difference in fraction for a pair is given by

$$\delta_i = p_{2i} - p_{1i} \qquad \text{Equation 3-3}$$

where

δ_i = difference in fraction between transects in pair *i*,

p_{ji} = fraction for the *j*th transect (1st or 2nd) in pair *i*.

The weighted mean difference (weighted by transect length) is estimated by

$$\overline{\delta} = \frac{\sum (L_i \delta_i)}{\sum L_i} \qquad \text{Equation 3-4}$$

and the sample standard deviation of the weighted sample is given by (GNU 2015)

$$s_\delta = \sqrt{\frac{\sum L_i \sum \left[L_i \left(\delta_i - \overline{\delta} \right)^2 \right]}{\left(\sum L_i \right)^2 - \sum L_i^2}} \qquad \text{Equation 3-5}$$

where L_i is the transect length of transect *i*.

The standard error on the mean difference in fraction across the paired transects is given by

$$s_{\overline{\delta}} = \frac{s_\delta}{\sqrt{N}} \qquad \text{Equation 3-6}$$

where *N* is the number of pairs of transects. The difference in site area is estimated by

$$\Delta A = E \cdot \overline{\delta} \qquad \text{Equation 3-7}$$

where *E* is the sample polygon area. The standard error of the estimated difference in site area is itself estimated by

$$s_{\Delta A} = E \cdot s_{\overline{\delta}} \qquad \text{Equation 3-8}$$

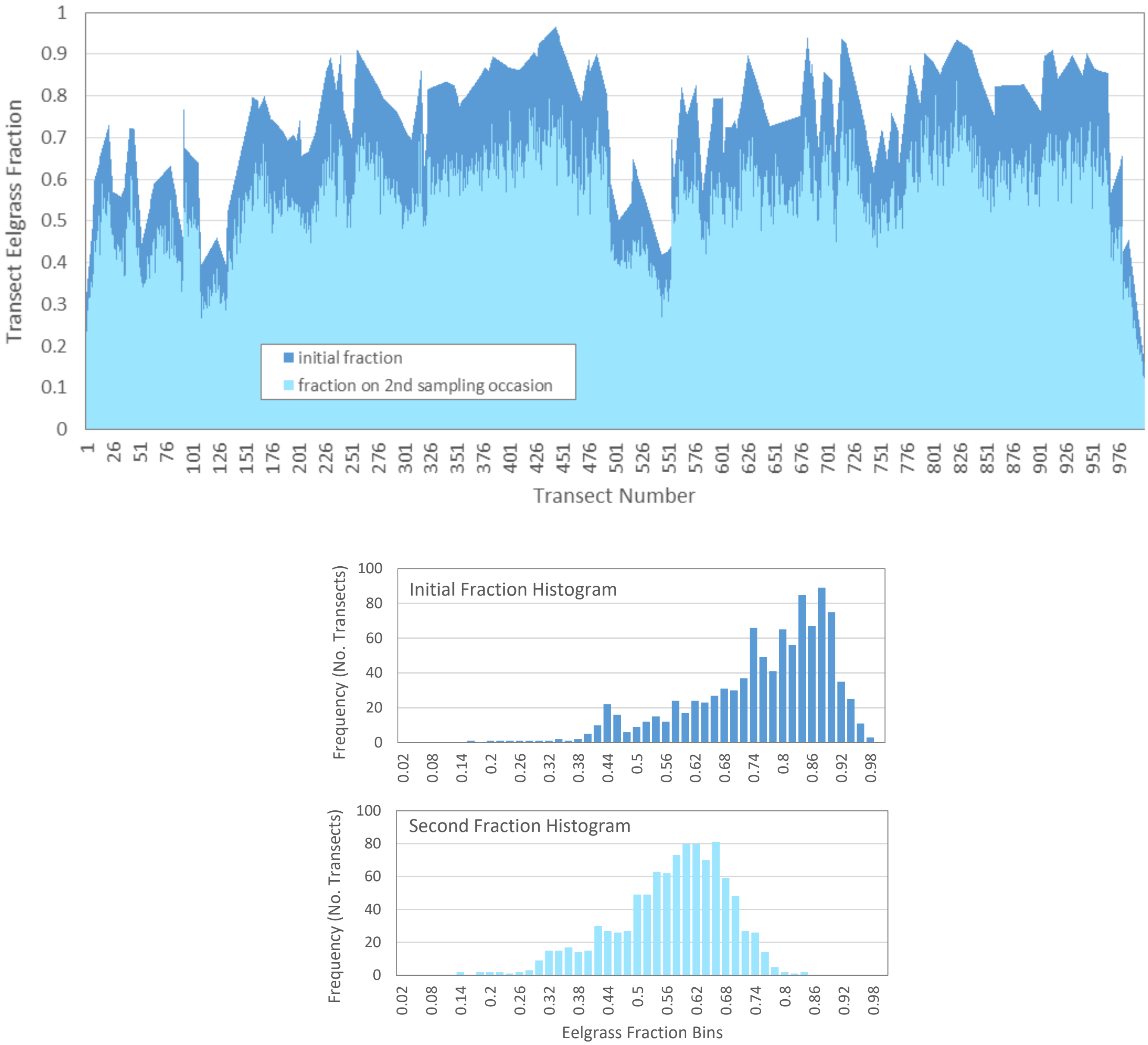


Figure 3-1. Site change model with nominal change of -0.25 applied to the core001 site model (top) and histograms of transect fraction (below).

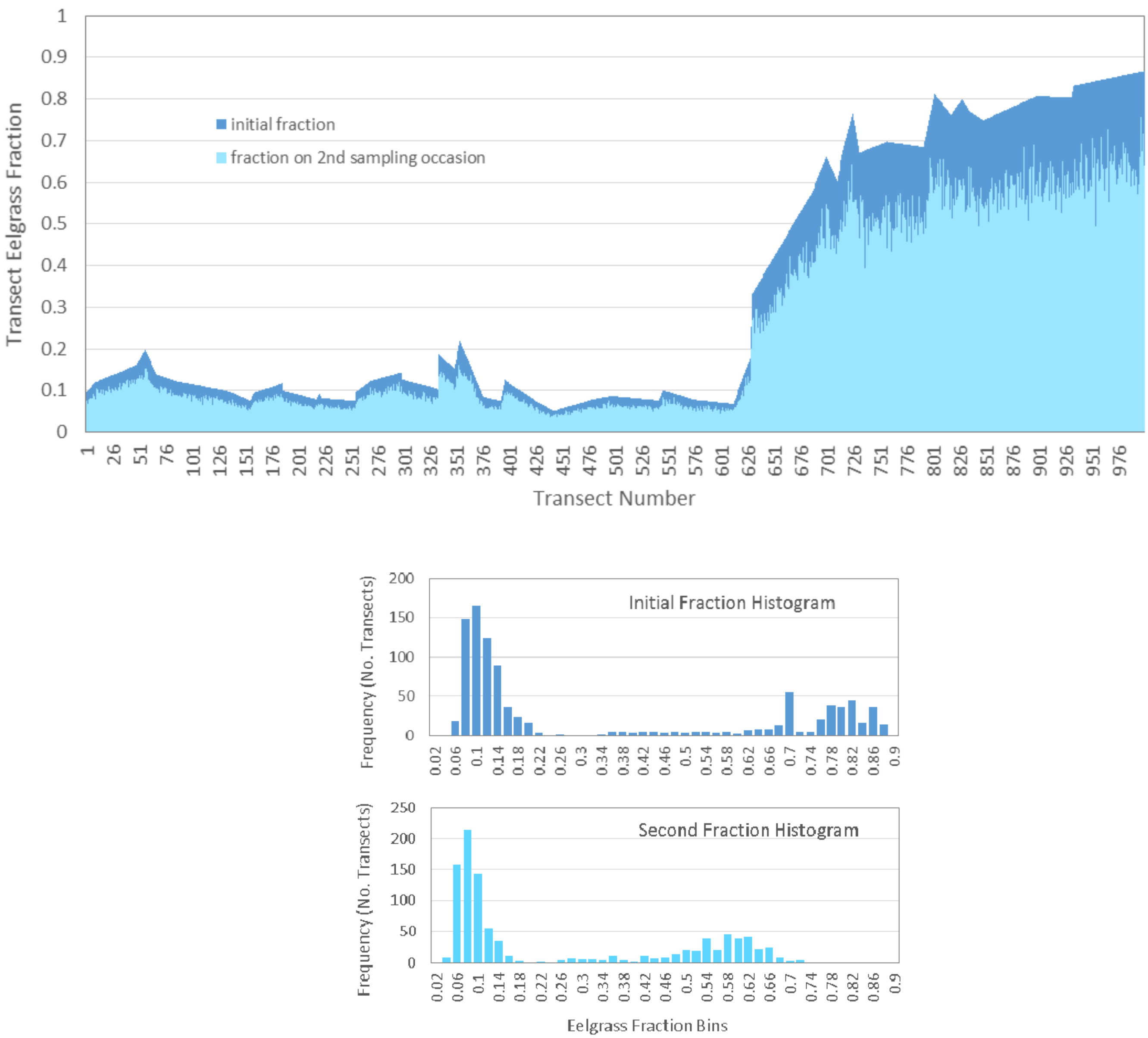


Figure 3-2. Site change model with nominal change of -0.25 applied to the flats26 site model (top) and histograms of transect fraction (below).

3.3 Power to Detect Difference in Site Area with *t*-test

Monte Carlo sampling was conducted of the site models from the initial sampling occasion and the second sampling occasion. This sampling followed the methods used earlier (section 2.3, p.13) except an additional random error was added to the transect fractions from the second sampling occasion to represent positioning error – i.e., the lack of perfect alignment with the initial transect. The standard deviation of this random error was 0.0125, equal in magnitude to the error associated with video processing. The total random error for transects from the second sampling occasion was the square root of the sum of the variances from video classification error and positioning error.

For each site change scenario (sites core001 and flats26; nominal change of 0, -0.05, -0.10, -0.15, -0.20, -0.25) 5000 samples were drawn of size $n = 10$ transects for each of SRS, STR and SYS selection methods for each of the sampling occasions. This resulted in 5000 estimates of difference in site area for each set of conditions. Each difference estimate was subjected to a *t*-test to test for the difference being significantly different from zero. Power was calculated as the proportion of the 5000 tests that were significant. Rates of Type I error were evaluated for the model runs with no change in the site model. The equations for calculating the *t* statistics are presented in the following sections.

3.3.1 New Draw of Transects on Second Occasion – Difference between Two Area Estimates

For a new draw of transects on the second sampling occasion, the test for change consists of testing whether ΔX is significantly different than zero. The *t*-statistic for this test is given by

$$t = \frac{\Delta X}{s_{\Delta X}}. \qquad \text{Equation 3-9}$$

The degrees of freedom for the associated *t* distribution, ν, is estimated by the Smith-Satterthwaite procedure (Milton and Arnold 1990, p.320; attributed to Smith by Zar 1999, p.129) and expressed as

$$\nu = \frac{\left(s_{X_1}^2 + s_{X_2}^2\right)^2}{\frac{\left(s_{X_1}^2\right)^2}{(n_1 - 1)} + \frac{\left(s_{X_2}^2\right)^2}{(n_2 - 1)}} \qquad \text{Equation 3-10}$$

where the sample sizes n_1 and n_2 are 10 for the simulations in this work.

3.3.2 Repeat Transects – Direct Estimate of Difference in Area

Power to detect a significant difference in eelgrass area for paired analysis of repeat transect sampling was based on a *t*-test with the *t* statistic

$$t_{\Delta A} = \frac{\Delta A - 0}{s_{\Delta A}}. \qquad \text{Equation 3-11}$$

3.4 Results – Difference in Site Area Estimates from New Draw Transects

The distributions of difference estimates adhered very closely to a normal distribution for both core001 and flats26 (see histograms and Q-Q plots in Appendix A, p.71). This is in contrast to the clear departures from normality in the distributions of site area estimates presented earlier (Figure 2-4, p.15).

When these distributions are compared in the form of boxplots against the true values of difference in site eelgrass area and standard error of the mean difference, then the sampling methods can be readily compared in terms of precision and bias (Figure 3-3).

Some key findings from these comparisons include:

- There is again a consistent gain in precision when sampling with STR and SYS as compared to SRS. This applies to both estimation of difference in area (Figure 3-3 c, d) and of standard error (Figure 3-3 i, j) and is particularly strong when sampling from the flats26 site model. The gain in precision in the standard error estimates when sampling from the core001 model was not as strong but still present.
- STR and SYS gave similar levels of precision but when they differed SYS was generally, but not always, slightly more precise (Figure 3-3 c, d, i, j).
- Estimates of difference in site area have low bias (<3%) for all selection methods (Figure 3-3 e, f) except for one value of +11% for SRS sampling from the flats26 site model. This is likely a random anomaly that would not persist under repeated model runs.
- Standard error estimates are relatively unbiased (bias < 5%) only for SRS sampling (Figure 3-3 k, l).
- Standard error estimates with STR and SYS are strongly positively biased (Figure 3-3 k, l) especially when sampling from the flats26 site model where bias surpasses +400%.

The power achieved in detecting differences in area was modest when sampling from the core001 model (Figure 3-4) and poor when sampling from the flats26 model (Figure 3-5). For core001, a general decline of 20% was needed before power reached 0.8. There was a crossover effect where SRS had greater power than STR and SYS (albeit very low) at low levels of change but lower power at higher levels of change. This has to do with the interplay between precision of difference estimates and bias in the standard error estimates and can be seen in more detailed power diagrams (Appendix C, Figure C-1, p.79). For flats26, power never exceeded 0.2 (Figure 3-5). For STR and SYS sampling, no differences were detected at all (power = 0) which was unexpected but is explained by inspection of the detailed power diagrams (Appendix C, Figure C-2, p.80). This was a consequence of the very large positive bias in the estimation of standard error for STR and SYS sampling from the flats26 site model (Figure 3-3 l).

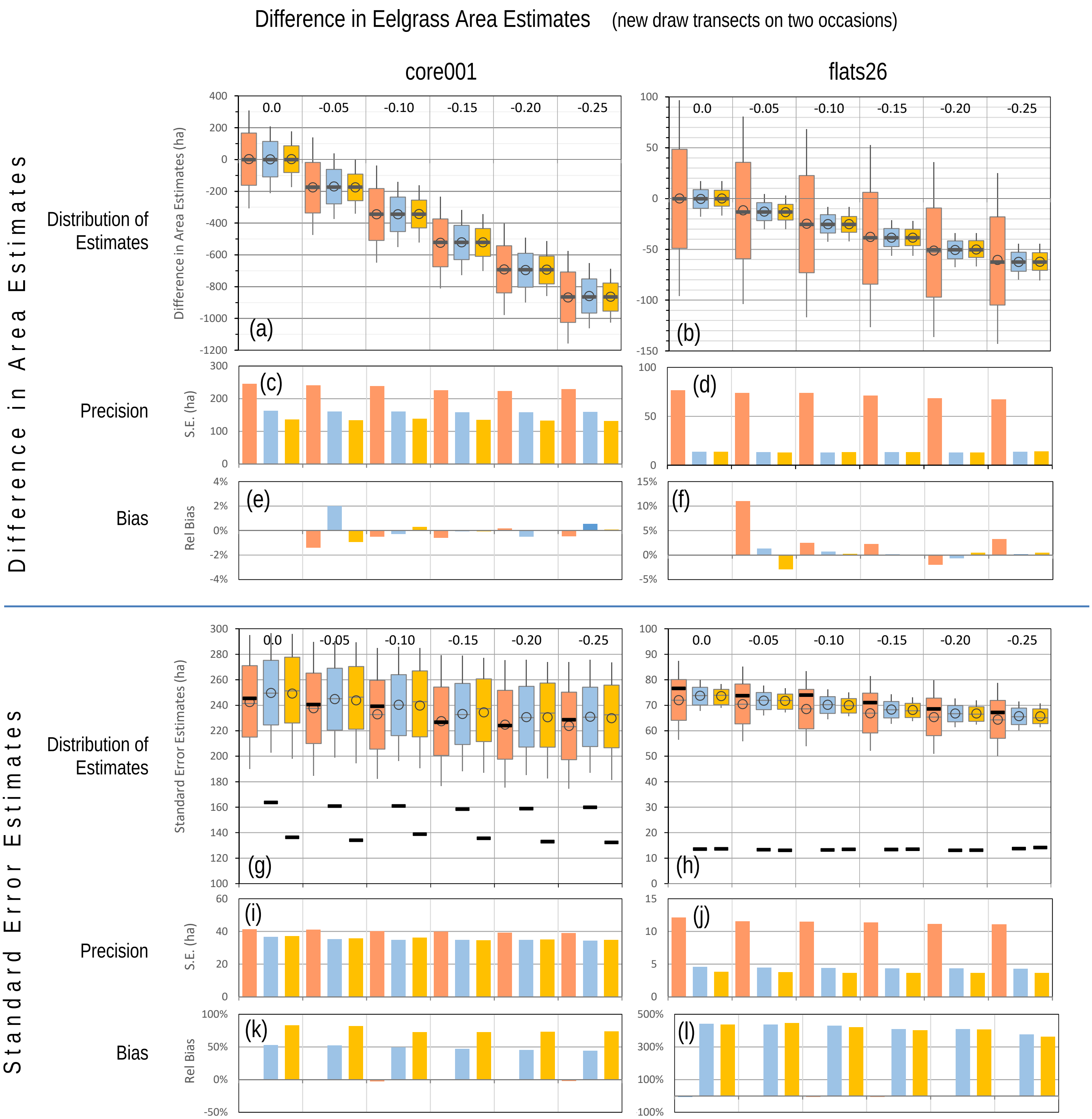


Figure 3-3. The distribution of estimates and the associated precision and bias for estimates of difference in eelgrass area estimates (top half of page) and for estimates of the standard error of the difference estimate (bottom half). The estimates are based on samples drawn from the site models for core001 (left) and flats26 (right) and with sample selection by SRS (red), STR (blue) and SYS (gold). The results are grouped by the change scenario with the nominal change value of the scenario given at the top of the boxplots. Although there was bias present for the scenarios with nominal change of 0 in (e) and (f), the relative bias cannot be calculated (division by zero) so no values are displayed. The template for the boxplots is shown in Figure 2-8 (p.17).

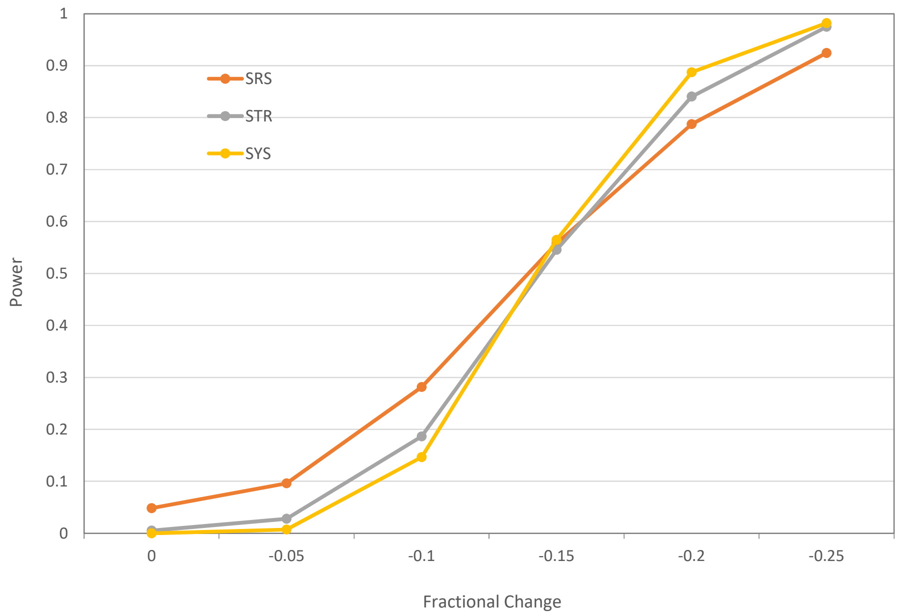


Figure 3-4. Power to detect change in the core001 site model with new draw transects. Power is shown for detecting the different levels of change shown along the x-axis.

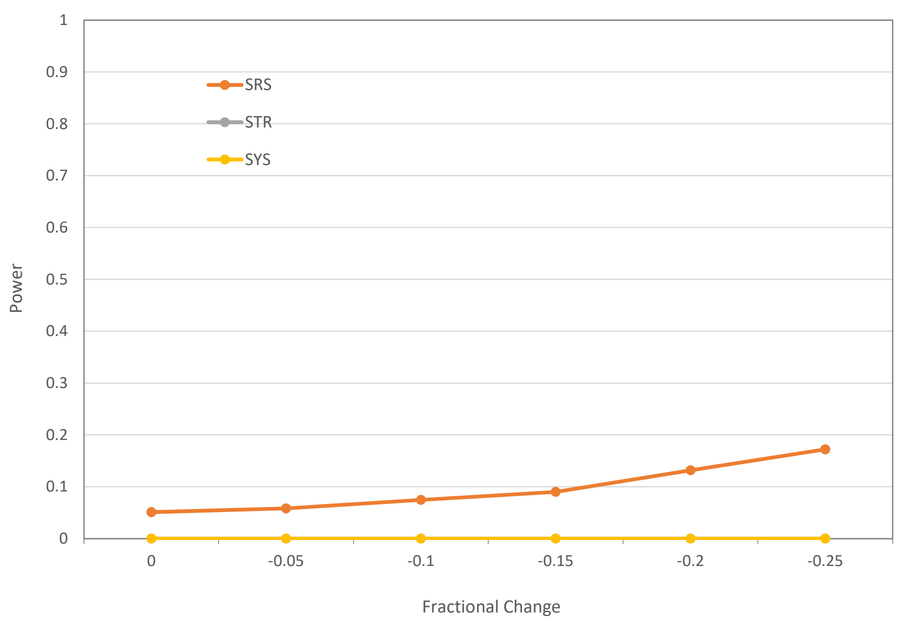


Figure 3-5. Power to detect change in flats26 site model with new draw transects. Power is shown for detecting the different levels of change shown along the x-axis.

3.5 Results – Estimated Difference in Area from Repeat Transects (Paired)

The distributions of difference estimates based on paired transects adhered very closely to a normal distribution for both core001 and flats26 (see histograms and Q-Q plots in Appendix B, p.75).

When these distributions are compared in the form of boxplots against the true values of difference in site eelgrass area and standard error of the mean difference, then the sampling methods can be readily compared in terms of precision and bias (Figure 3-6). In terms of the precision of difference estimates, STR and SYS sampling again give improved precision over SRS sampling. When sampling from the core001 site model this effect is modest (Figure 3-6 c) but when sampling from flats26 it is more prominent (Figure 3-6 d). Overall the standard errors of difference estimates are much lower with repeat transect sampling as compared to sampling with newly drawn transects on the second occasion (Figure 3-3). Bias in estimates of area difference are very low (< 1%) (Figure 3-6 e, f) except for slightly elevated bias when SRS sampling from the flats26 model (Figure 3-6 f).

For the estimation of standard error of the difference estimate, the sample selection method has little effect on precision (Figure 3-6 i, j). There is a modest gain in precision with STR and SYS relative to SRS when sampling from the flats26 model (Figure 3-6 j). There is no consistent effect when sampling from the core001 model. Again, the overall effect of repeat transect sampling is to strongly increase precision compared to sampling with replacement. In general, the standard error estimates increase with higher levels of decrease in site eelgrass area. This reflects the nature of the site change model which introduces stochastic noise into the applied change based on CV. Consequently, as the magnitude of change increases so does the absolute magnitude of the variability in change across the site.

Bias in standard error estimation is modest when sampling from the core001 model (< 10%) (Figure 3-6 k) but gets larger when sampling from the flats26 model (> 50%) (Figure 3-6 l). Estimates based on SRS sampling are consistently negatively biased while estimates based on STR and SYS sampling tend to be positively biased.

The overall increase in precision when shifting to repeat transects has a strong effect on the power achieved when testing for a significant difference in site eelgrass area. When sampling from the core001 model, tests for change have a 100% probability of detecting change for the change scenarios studied for all sample selection methods (Figure 3-7). Even when sampling from the flats26 model, the STR and SYS selection methods achieve power of approximately 0.8 for the lowest level of change (0.05) and power of 1.0 for change of 0.1 or greater (Figure 3-8). Sampling with SRS also has high power but consistently lower than STR and SYS.

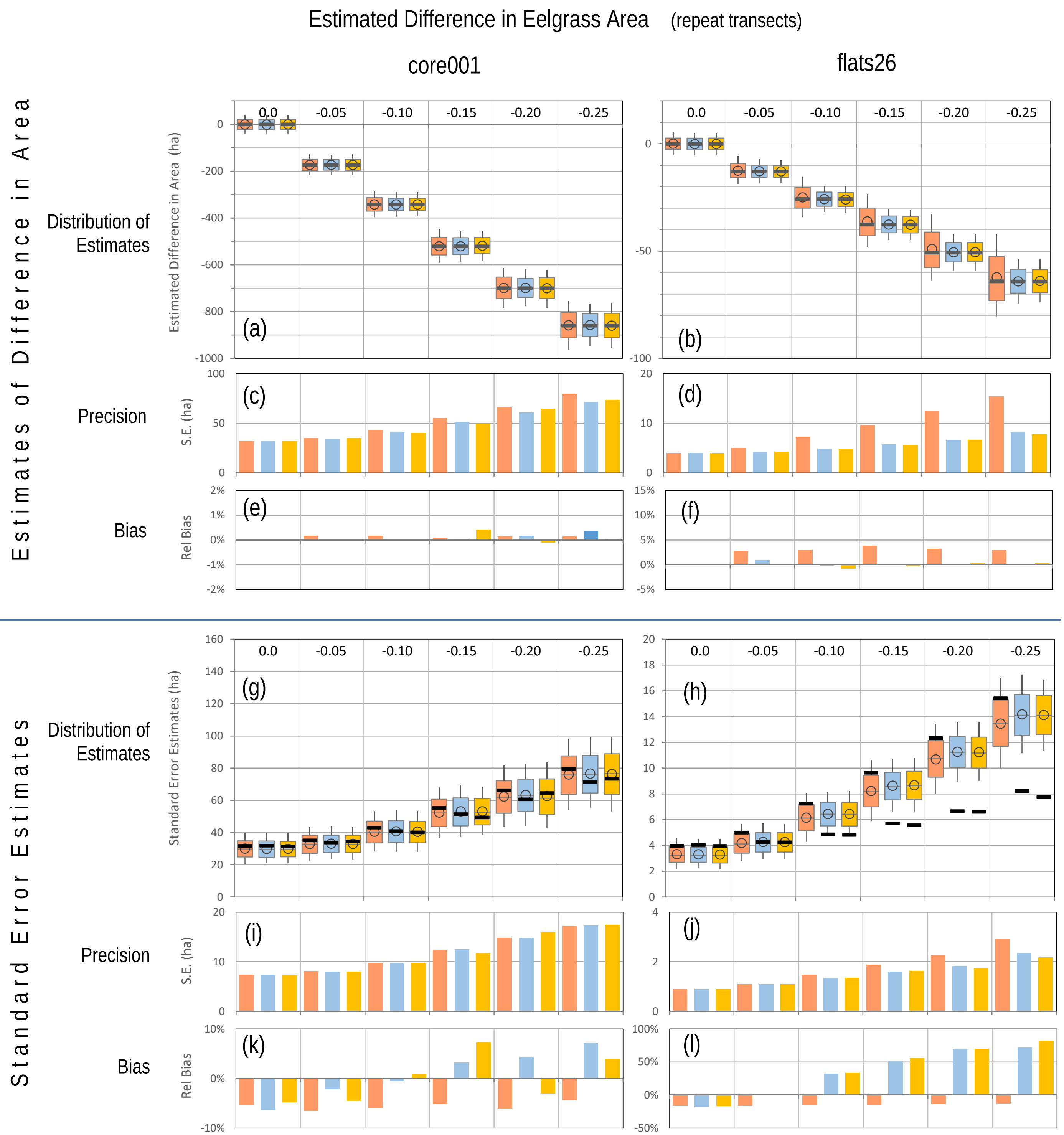


Figure 3-6. The distribution of estimates and the associated precision and bias for estimates of difference in eelgrass area estimates (top half of page) based on paired repeat transects. The bottom half of the page shows results for estimates of the standard error of the mean difference estimate. The estimates are based on samples drawn from the site models for core001 (left) and flats26 (right) and with sample selection by SRS (red), STR (blue) and SYS (gold). The results are grouped by the change scenario with the nominal change value of the scenario given at the top of the boxplots. Although there was bias present for the scenarios with nominal change of 0 in (e) and (f), the relative bias cannot be calculated (division by zero) so no values are displayed. The template for the boxplots is shown in Figure 2-8 (p.17).

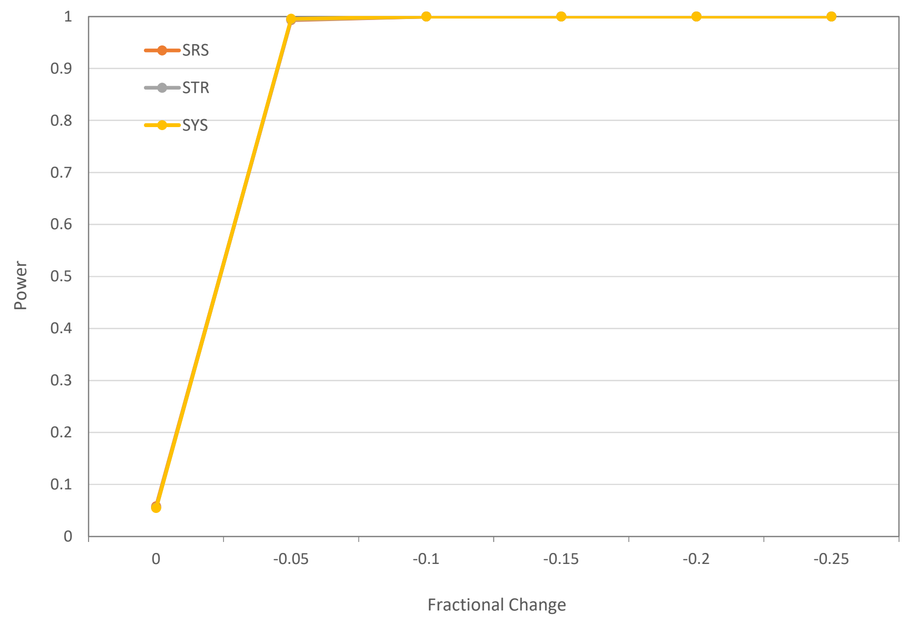


Figure 3-7. Power to detect difference in area in core001 site model with paired repeat transects. Power is shown for detecting the different levels of change shown along the x-axis.

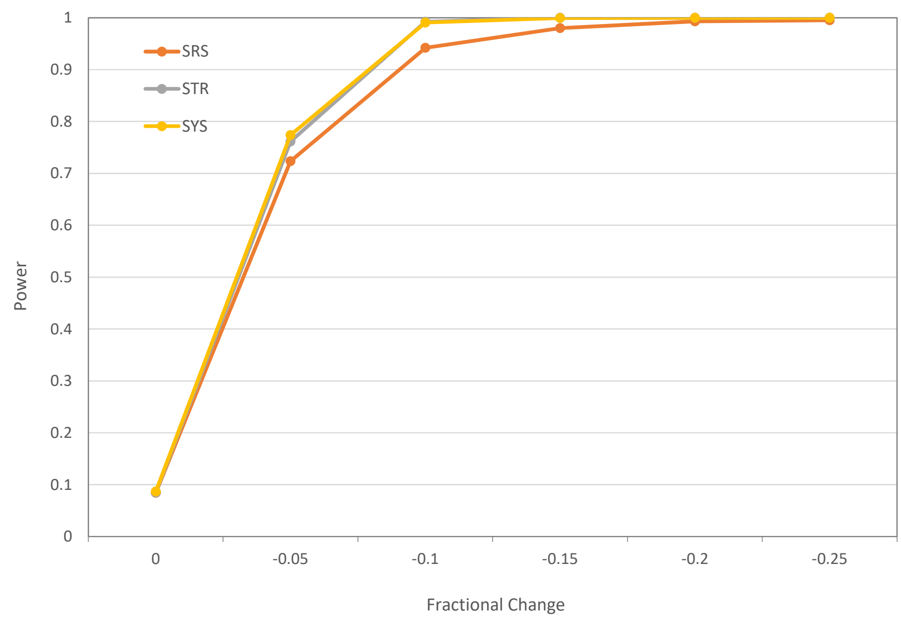


Figure 3-8. Power to detect difference in area in flats26 site model with paired repeat transects. Power is shown for detecting the different levels of change shown along the x-axis.

4 Alternate Estimators for Standard Error Applied to STR and SYS

The potential benefits of STR and SYS sampling relative to SRS are clear. These include:

- greater precision in site area estimates (Figure 2-8 c, d; p.17) and differences between two site estimates (Figure 3-3 c, d; p.25), especially for sites with heterogeneous eelgrass abundance.
- modestly greater precision in difference estimates (repeat transects) for heterogeneous sites (Figure 3-6 c, d; p.28)
- low bias in site area estimates for heterogeneous sites (Figure 2-8, right, p.17)

The important shortcoming of STR and SYS sampling is in the bias in the standard error estimates when the standard estimator is used. This behavior is seen with the standard error estimates of site area (Figure 2-8 k, l), differences in site area estimates (Figure 3-3 k, l) and to a lesser degree with the difference based on paired transects (Figure 3-6). The fact that the observed bias is positive suggests that power to detect change with STR or SYS data would be reduced or, in some cases, eliminated (Figure 3-5).

Three alternative estimators of standard error were explored to see if they might perform better with STR samples. These were initially applied only to the core001 model to first screen their potential. Based on the initial screening, the best performing estimator was studied in more detail with application to both STR and SYS sampling. The standard estimator was retained for SRS samples.

4.1 Collapsed Strata

First, Cochran's method of collapsed strata was examined. In section 5A.12 of Cochran (1977), entitled "Estimation of Variance with One Unit per Stratum", he described this method for cases where the sample size is even (Cochran 1977, p.139) In this case each sample unit (i.e., transect) can be paired with another in the sample. If each transect is used to make an estimate of site eelgrass area (transect fraction × sample polygon area) then each transect pair gives a pair of estimates and the difference in these estimates can be calculated. The collapsed strata estimate is simply based on a sum of these differences. The collapsed strata estimate of variance is given by (Cochran 1977, Eqn 5A.54, p.139)

$$Var\left(\overline{X}\right)=\sum_{j=1}^{N/2}\left(X_{j1}-X_{j2}\right)^2 \qquad \text{Equation 4-1}$$

where

$\overline{X}$ = the estimate of site eelgrass area based on a mean over all transects

X_{j1}, X_{j2} = estimates of site eelgrass area based on the first (1) and second (2) transect of pair *j*

N = the number of transects in the sample.

Cochran notes that this collapsed strata estimate is positively biased to an extent that depends on the pairing of data. He suggests that careful pairing might reduce the variance but this should be based on prior knowledge and not on the sample data. The pairing used here was of adjacent transects – i.e., 1&2, 3&4, …, 9&10.

4.2 Jackknife

Second, a jackknife was applied to the sample data to estimate variance. The jackknife estimate is given by

$$Var\left(\overline{X}\right)=\frac{n-1}{n}\sum_{i=1}^{n}\left(\overline{X}_i-\overline{X}\right)^2 \qquad \text{Equation 4-2}$$

where

$$\overline{X}_i=\sum_{\substack{j=1\\ j\neq i}}^{n}X_j \qquad \text{Equation 4-3}$$

4.3 Wolter's v8 Estimator

Wolter's v8 estimator (Wolter 1984) for systematic samples was selected for testing based on the recommendation of McGarvey et al. (2016) for sampling of vegetation with SYS. While the collapsed strata and jackknife estimators were applied to samples of site area or difference estimates (each estimate derived from one transect), here the v8 estimator was applied to transect fraction (or fraction difference) values and the estimated variance was propagated to site area or difference in site area. While presented by Wolter (1984) as an estimator for SYS samples it is used here for both SYS and STR.

The v8 estimator requires an estimate of correlation in the sample calculated as

$$\rho=\frac{\sum_{i=2}^{n}\left(p_i-\overline{p}\right)\left(p_{i-1}-\overline{p}\right)}{s^2\left(n-1\right)} \qquad \text{Equation 4-4}$$

where s^2 is the estimate of the variance of weighted mean fraction calculated with the variance estimator given by Skalski (2002, p.3).

The v8 estimator for variance is given by

$$v_8 = Var\left(\overline{X}\right) = \begin{cases} \left(\dfrac{s^2}{n}\right)\left[1 + \dfrac{2}{\ln(\rho)} + \dfrac{2}{1/\rho - 1}\right] & if\ \rho > 0 \\ \dfrac{s^2}{n} & if\ \rho \le 0 \end{cases} \qquad \text{Equation 4-5}$$

Strictly speaking a weighted correlation coefficient should be calculated rather than an unweighted coefficient (Equation 4-4) but this issue was neglected. This occasionally led to problems where the estimated correlation was slightly greater than 1. In these cases the value was set to $\rho = 0.99$.

4.4 Initial Comparison

An initial comparison was conducted of standard error estimates for site area for the core001 model. The goal was to screen estimators that did not perform well and only conduct further analysis with strong candidate estimators.

The results showed that the collapsed strata estimator performed very poorly in terms of bias and the v8 estimator was the best (Figure 4-1). Based on these initial results only the v8 estimator was investigated further.

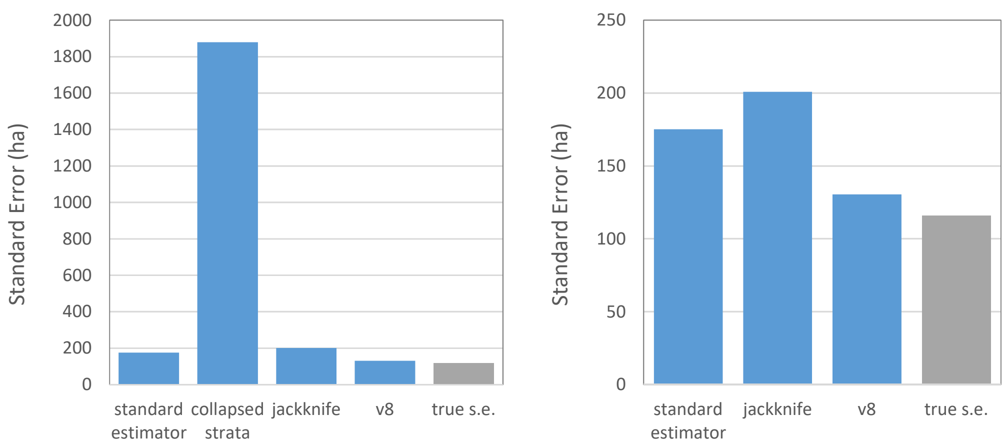


Figure 4-1. Initial comparison of standard error estimates for site area estimates from the core001 site model when sampling with STR. Results with three estimators were compared to the standard estimator and the true standard error. The graph on the right excluded the collapsed strata estimate to better see differences among the remaining estimates.

4.5 Performance of the v8 Variance Estimator

Given that the v8 estimator of Wolter (1984) was the most promising alternative variance estimator (Figure 4-1), it is explored in more detail in this section. The v8 estimator was used for variance estimation in the results for both STR and SYS sampling. The standard estimator was retained for variance estimation with SRS sampling.

4.5.1 Area Estimates

For estimation of standard error on site eelgrass area estimates, the v8 estimator has bias of much lower absolute magnitude than the standard estimator (cf. Figure 4-2 e, f and Figure 2-8 k, l, p.17). This is particularly true in sampling from the flats26 site model where bias was reduced from +440% (Figure 2-8, p.17) with the standard estimator applied to STR samples to -48% (Figure 4-2) with the v8 estimator. When sampling from the core001 site model with STR, bias was reduced from +50% to +11%.

While bias in standard error estimation improved with the v8 estimator, there was only a minor effect on precision. The effect was negligible on precision of these estimates when sampling from the flats26 site model (cf. Figure 4-2 d and Figure 2-8 j, p.17) and a modest loss of precision when sampling from the core001 site model (cf. Figure 4-2 c and Figure 2-8 i).

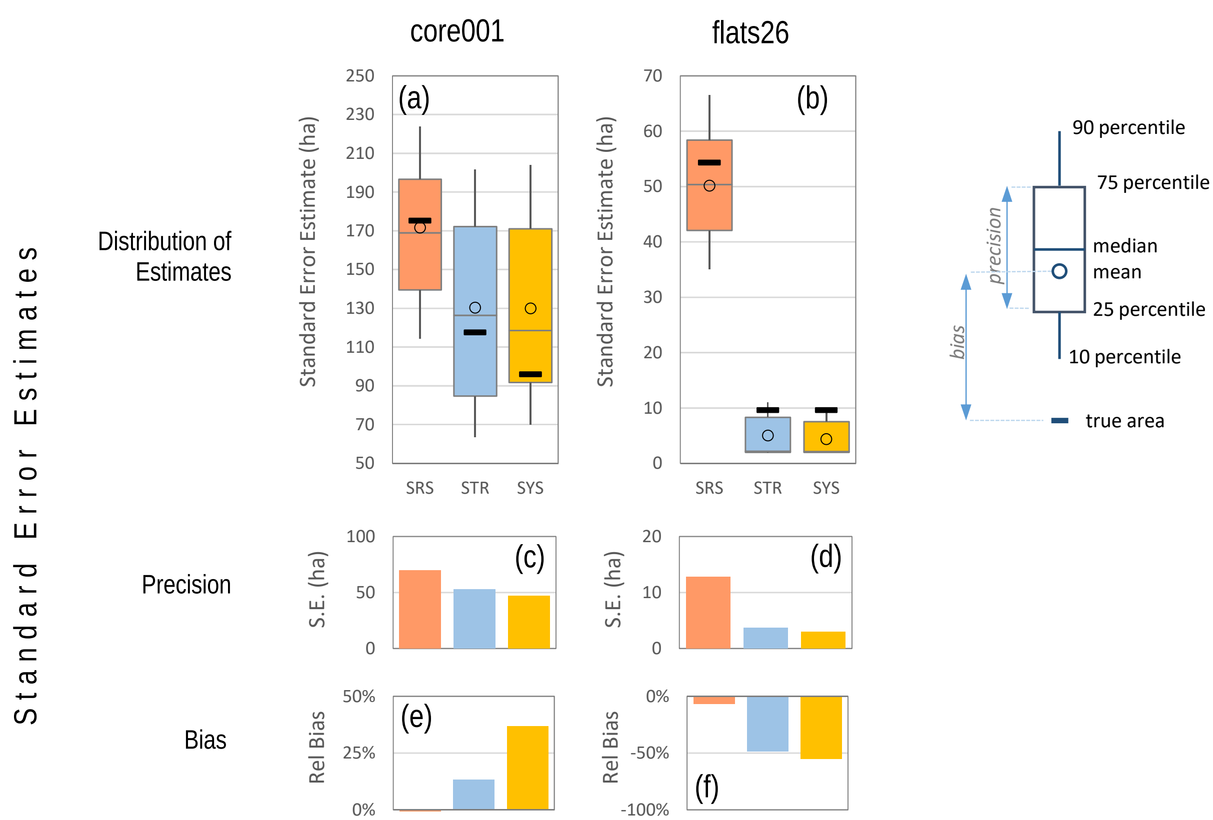


Figure 4-2. The distribution of estimates and the associated precision and bias for estimates of the standard error of site eelgrass area estimates based on the v8 variance estimator for STR and SYS sampling. The estimates are based on samples drawn from the site models for core001 (left) and flats26 (right) and with sample selection by SRS (red), STR (blue) and SYS (yellow). Comparison with results using the standard variance (Figure 2-8, p.17) clearly reveals the lower bias of the v8 variance estimator.

While the bias in standard error estimates with the v8 estimator is substantially reduced in absolute magnitude, it is still great enough in sampling from the flats26 model (-48%) to have implications for results interpretation. This is clear in the coverage probability of estimated 95% confidence intervals (Table 4-1) which is only 63% with STR sampling. The fact that the bias of the v8 estimator varies across sites also complicates interpretation of estimates.

Table 4-1. Coverage probability of 95% confidence intervals for site area estimates. For STR and SYS sampling the v8 variance estimator was used. The coverage probability is the proportion of intervals estimated under repeated sampling that actually encompass the true site area.

	SRS	STR	SYS
core001 model	94%	93%	96%
flats26 model	94%	63%	51%

4.5.2 Differences in Area Estimates (New Draw Transects)

For the estimation of standard error on the difference between two site eelgrass area estimates, the performance of the v8 estimator (Figure 4-3) also has strong differences relative to the standard estimator (Figure 3-3, p.25). These differences mimic those seen with estimation of standard error on site area estimates (section 4.5.1). The magnitude of bias is strongly reduced (cf. Figure 4-3 e, f and Figure 3-3 k, l), particularly when sampling from the flats26 model. There is modest effect on precision of the standard error estimates with loss of precision when sampling from the core001 model (cf. Figure 4-3, c, d and Figure 3-3 i, j).

The lower magnitude bias with the v8 estimator now includes negative bias (Figure 4-3 f) which again has implications for the coverage probability of 95% confidence intervals. Across all the change scenarios investigated, the coverage probability for STR sampling ranged from 68% to 71% when sampling from the flats26 model (Table 4-2). The coverage probability when sampling from the core001 model was much more accurate (96-97%).

The reduction in bias in standard error estimation with the v8 estimator leads to improved levels of power to detect change under STR and SYS sampling. For example, when sampling with STR from the core001 model under the -0.15 nominal change scenario, power was raised from around 0.55 (Figure 3-4, p.26) to around 0.75 (Figure 4-4). This represents a substantial advantage relative to the performance of SRS sampling. This is despite a remaining positive bias with the v8 estimator of +16% (Figure 4-3e).

The improvement of power when using the v8 estimator for STR and SYS sampling is more dramatic when sampling from the flats26 model. For example, power improved from zero with the standard variance estimator (Figure 3-5, p.26) to 0.7 with the v8 estimator under STR sampling of the -0.1 nominal change scenario (Figure 4-5). But in this case, the remaining bias was negative (-44%, Figure 4-3 f) and this results in an elevated Type I error when there is no actual change in the population. Whereas we expect a 5% rate of false positives when testing with $\alpha = 0.05$, the actual rate of false positives was 31% (Figure 4-5).

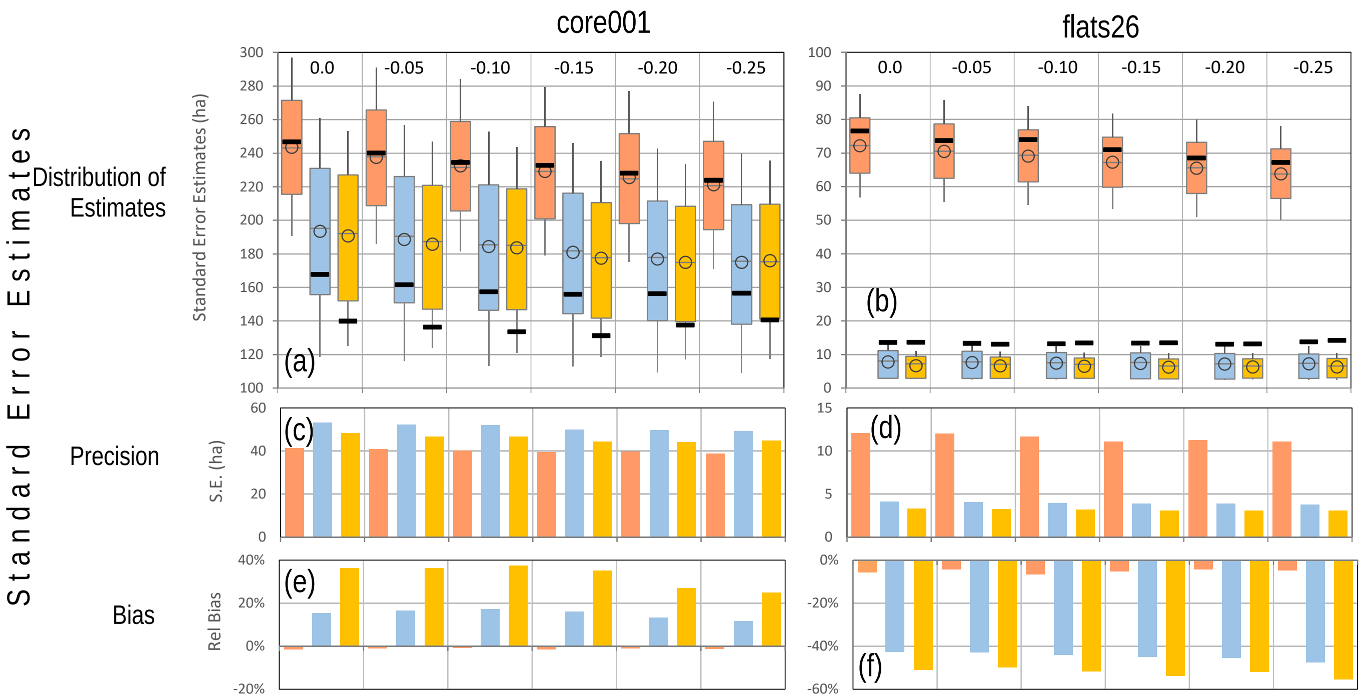


Figure 4-3. The distribution of estimates (a, b) and the precision (c, d) and bias (e, f) of these estimates of standard error in the difference between eelgrass area estimates based on newly drawn transects on the second occasion. The estimates from STR and SYS samples were based on the v8 variance estimator – cf. results with standard estimators in Figure 3-3 (bottom) (p.25). The estimates are based on samples drawn from the site models for core001 (left) and flats26 (right) and with sample selection by SRS (red), STR (blue) and SYS (gold). The results are grouped by the change scenario with the nominal change value of the scenario given at the top of the boxplots. Although there was bias present for the scenarios with nominal change of 0, the relative bias cannot be calculated (division by zero). The template for the boxplots is shown in Figure 2-8 (p.17).

Table 4-2. Coverage probability of 95% confidence intervals for difference between site area estimates based on a new draw of transects on the second sampling occasion. For STR and SYS sampling the v8 variance estimator was used. The coverage probability is the proportion of intervals estimated under repeated sampling that actually encompass the true site area.

site	selection method	0	-0.05	-0.1	-0.15	-0.2	-0.25
core001	SRS	97%	97%	97%	97%	96%	96%
	STR	97%	97%	97%	97%	96%	96%
	SYS	97%	98%	98%	98%	97%	98%
flats26	SRS	95%	95%	95%	95%	96%	95%
	STR	71%	71%	70%	69%	69%	68%
	SYS	70%	70%	69%	67%	67%	65%

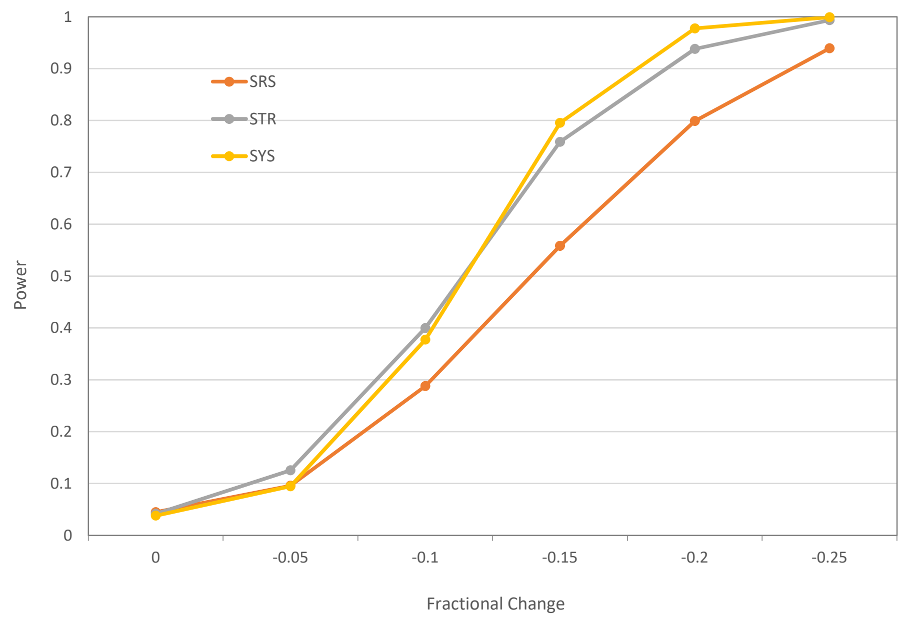


Figure 4-4. Power to detect change in the core001 site model with new draw transects based on use of the v8 variance estimator for STR and SYS sampling – cf. power with standard variance estimators, Figure 3-4 , p.26. Power is shown for detecting the different levels of change shown along the x-axis.

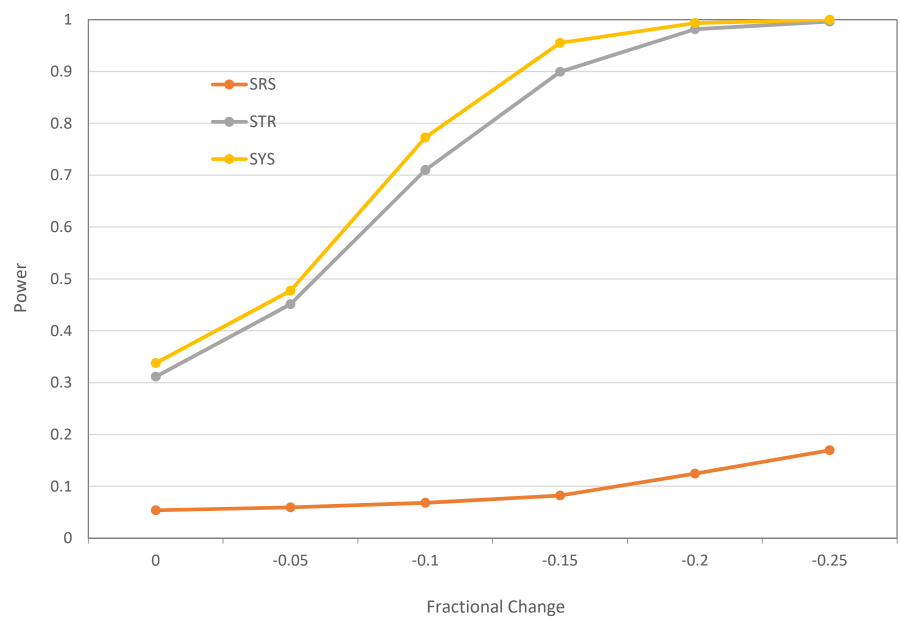


Figure 4-5. Power to detect change in the flats26 site model with new draw transects based on use of the v8 variance estimator for STR and SYS sampling – cf. power with standard variance estimators, Figure 3-5, p.26. Power is shown for detecting the different levels of change shown along the x-axis.

4.5.3 *Estimated Difference in Area (Repeat Transects)*

This section looks at the effect of using the v8 estimator in estimation of the standard error on estimates of difference in site area based on paired repeat transects. For estimates based on sampling of the core001 model, the precision of the estimates (Figure 4-6 c) was slightly degraded relative to variance estimation with the standard estimator (Figure 3-6 i, p.28) for both STR and SYS sampling. Bias in the standard error estimates also became greater in magnitude with a consistent negative bias (-10 to -20%) with the v8 estimator (Figure 4-6 e, compared to Figure 3-6 k).

For estimates based on sampling from the flats26 model, precision with STR sampling improved from s.e.≈ 1-2 ha (i.e., standard error of the mean standard error estimate) with the standard estimator (Figure 3-6 j) to s.e. ≈ 1 ha with the v8 estimator (Figure 4-6 f). In contrast, the positive bias seen with the standard variance estimator (+30 to +70%, Figure 3-6 l) was replaced with consistently negative bias of similar absolute magnitude (-50 to -70%, Figure 4-6 f).

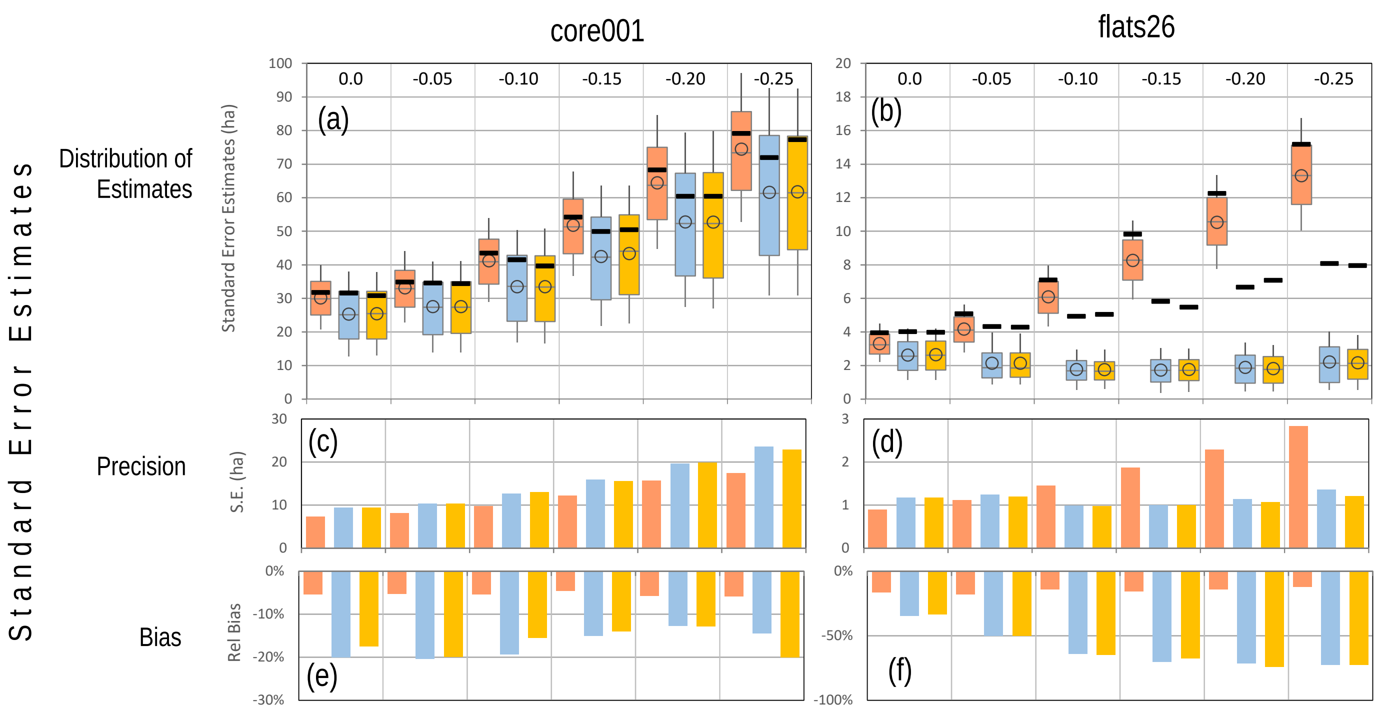


Figure 4-6. The distribution of estimates (a, b) and the precision (c, d) and bias (e, f) of these estimates of standard error in the difference in eelgrass area based on paired repeat transects. The estimates from STR and SYS samples were based on the v8 variance estimator – cf. results with standard estimators in Figure 3-6 (bottom) (p.28). The estimates are based on samples drawn from the site models for core001 (left) and flats26 (right) and with sample selection by SRS (red), STR (blue) and SYS (gold). The results are grouped by the change scenario with the nominal change value of the scenario given at the top of the boxplots. Although there was bias present for the scenarios with nominal change of 0, the relative bias cannot be calculated (division by zero). The template for the boxplots is shown in Figure 2-8 (p.17).

The negative bias in the standard error estimates seen when sampling each of the site models (Figure 4-6 e, f) translates into low coverage probability of the 95% confidence intervals. The coverage probabilities were 87-89% when sampling from the core001 model and 43-78% when sampling from the flats26 model (Table 4-3).

Table 4-3. Coverage probability of 95% confidence intervals for the estimated difference in site area based on paired repeat transects. For STR and SYS sampling the v8 variance estimator was used. The coverage probability is the proportion of intervals estimated under repeated sampling that actually encompass the true site area.

site	selection method	0	-0.05	-0.1	-0.15	-0.2	-0.25
core001	SRS	94%	94%	95%	94%	94%	95%
	STR	87%	87%	87%	88%	89%	88%
	SYS	88%	87%	88%	89%	88%	87%
flats26	SRS	92%	90%	91%	91%	91%	92%
	STR	78%	66%	55%	47%	45%	43%
	SYS	79%	67%	53%	50%	40%	42%

The effect of using the v8 estimator on power to detect change in the core001 site model was not well captured since power was 1.0 at the nominal change values used with the standard variance estimator (Figure 3-7, p.29) and it remained 1.0 with the v8 estimator (Figure 4-7). Presumably the v8 estimator would give different power at lower levels of change than were investigated here (< 5%). At such lower levels of change, power with the v8 estimator would be expected to be higher (than with the standard variance estimator) because of the negatively biased estimates of standard error. However, another effect of this negative bias is elevation in Type I error. When using the standard variance estimator Type I error was consistent with the α used in testing (Figure 3-7, p.29). When using the v8 estimator, Type I error was elevated to 13% when testing at $\alpha = 0.05$ (Figure 4-7).

The v8 estimator has a more noticeable effect on power to detect change in the flats26 model at the nominal change levels investigated. For STR sampling with nominal change of -0.05, power increased from 0.78 with the standard variance estimator (Figure 3-8, p.29) to 0.91 with the v8 estimator (Figure 4-8). This is attributable to the large negative bias in standard error estimation with the v8 estimator (Figure 4-6 f). Again, the negative bias elevates the Type I error – in this case from 9% with the standard variance estimator (Figure 3-8) to 22% with the v8 estimator (Figure 4-8) when testing with $\alpha = 0.05$.

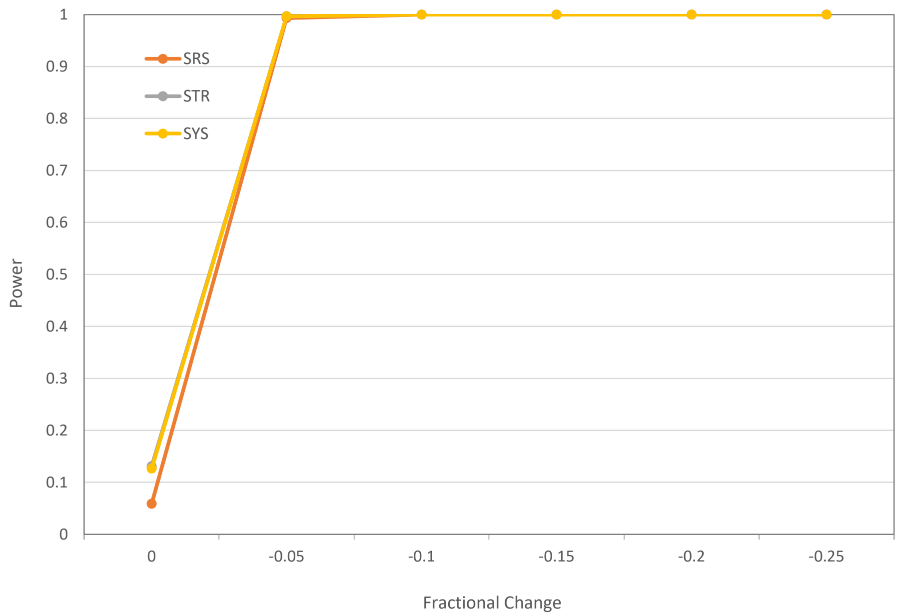


Figure 4-7. Power to detect change in the core001 site model with paired repeat transects based on use of the v8 variance estimator for STR and SYS sampling – cf. power with standard variance estimators, Figure 3-7, p.29. Power is shown for detecting the different levels of change shown along the x-axis.

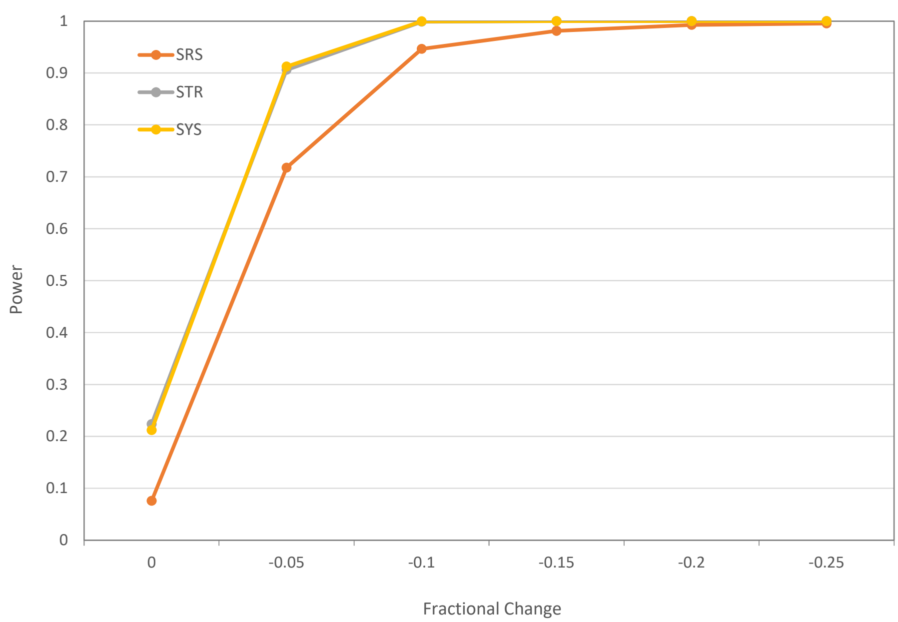


Figure 4-8. Power to detect change in the flats26 site model with paired repeat transects based on use of the v8 variance estimator for STR and SYS sampling – cf. power with standard variance estimators, Figure 3-8, p.29. Power is shown for detecting the different levels of change shown along the x-axis.

5 Detecting Heterogeneous Change

We have seen how spatial heterogeneity in the abundance of eelgrass makes STR and SYS more attractive relative to SRS for sample selection. With strong heterogeneity present, samples selected with STR and SYS produce greater precision in estimates of site area (cf. Figure 2-8 a and b, p.17) and much greater precision in the difference between two area estimates derived from independent samples (newly drawn random transects) (cf. Figure 3-3 a and b, p.25). The decision of which sample selection method to use must weigh this gain in precision against the absence of unbiased estimators for variance and standard error.

Under repeat transect sampling, the trade-off between STR and SRS appears to shift. A more modest gain in precision was seen with STR samples relative to SRS in change estimates based on paired analysis of repeat transects with strong heterogeneity present (cf. Figure 3-6 a and b). This makes sense if we consider STR sampling (as well as SYS) as a way to control for spatial variability when estimating temporal change. Repeating the initial sample survey on the second occasion is an alternate way to control for spatial variability but this approach has a greater benefit for an SRS sample (with no other control for spatial variability) compared to an STR sample (which has some built-in control over spatial variability). The differential performance between the two selection methods is thereby reduced and the relative benefit of STR over SRS is diminished.

It is important to note that the previous results with repeat sampling of transects were based on a spatially homogeneous change scenario (section 3.1, p.19; Figure 3-1, Figure 3-2). It was hypothesized that the differential in performance between STR and SRS under repeat transect sampling would increase under spatially heterogeneous change. This would tend to favor the choice of STR or SYS over SRS, assuming an adequate variance estimator could be identified. The purpose of the work reported in this section was to investigate the relative performance of sample selection methods in terms of power to detect spatially heterogeneous change scenarios. Both the standard variance estimator and the v8 estimator were used for STR and SYS sampling. Sampling was conducted from both core001 and flats26 site models using both sampling with and without replacement.

5.1 Change Scenarios

A series of change scenarios were investigated which included spatial heterogeneity in change. These scenarios were simple in that only one contiguous area of change was allowed but in each scenario this area covered a different portion of the site area.

Within the area where change occurred, the same homogeneous change model was applied that earlier was applied across the entire site (section 3.1). Outside this area of change there was no change in the site model between sampling occasions.

Six scenarios of heterogeneous change were investigated (Figure 5-1). The homogeneous change scenario was included to provide contrast resulting in seven total spatial change scenarios. Within each spatial change scenario different intensities of change were investigated. The nominal change intensities used previously ranged from 0 to -0.25. Here these fixed change intensities were used as a starting point but the intensity was increased further when needed to the point where the total site decline reached -0.25. More intense change was needed because only a portion of the site was subject to change. In several cases complete eelgrass loss within the portion of the site subject to change was not enough to reach a total site decline of -0.25.

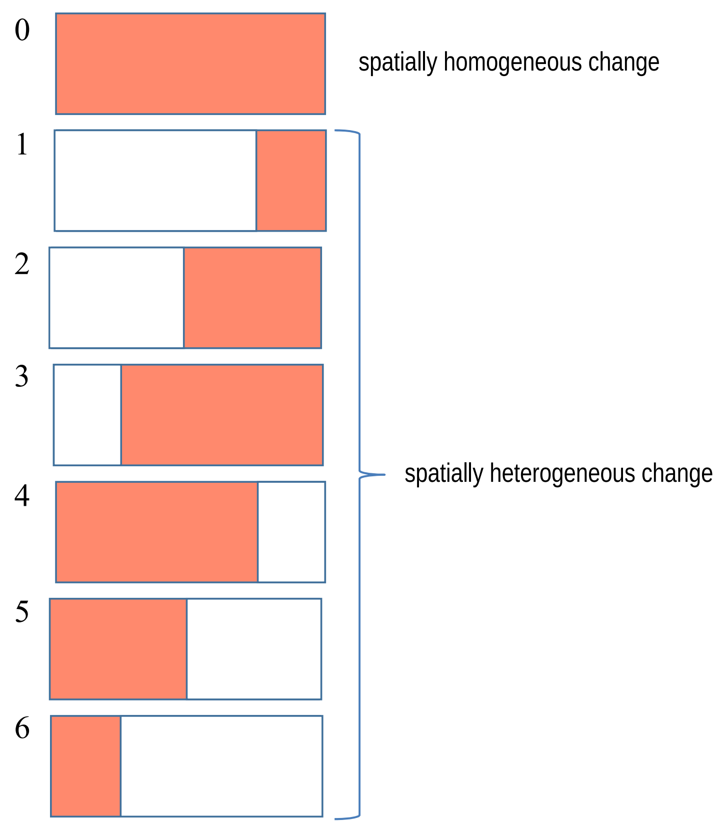


Figure 5-1. Spatial scenarios of change including the spatially homogenous change investigated earlier (section 3, labelled here as scenario 0) and the six spatially heterogeneous scenarios (labeled 1-6). Each rectangle represents the entire site area with the shoreline oriented along one of the horizontal margins. For each scenario change in eelgrass was confined to the red portion of the site.

5.2 Results – Power to Detect Change

Power curves are presented for sampling from the core001 model using the standard variance estimator for STR and SYS (Figure 5-2) and using the v8 estimator (Figure 5-3). Similar curves are presented for sampling from the flats26 model (Figure 5-4 and Figure 5-5). Each figure includes results for a newly drawn sample on the second sampling occasion (the "a" panels) and for a repeat survey of the initial sample (the "b" panels). The spatial change scenario is treated as a random factor in that individual scenarios are not identified. The graphs are designed only to convey the variability in power associated with patterns of spatial heterogeneity in change.

Curves that do not extend to -0.25 along the *x*-axis represent cases where there was complete eelgrass loss within the area of the site where change was applied so no further decline was possible under that spatial scenario.

Several points are clear from inspection of these graphs.

1. In most cases, there is substantial variation in power to detect change across the scenarios of heterogeneity in the spatial pattern of change.
2. Use of the v8 estimator brought large gains in the power of STR and SYS samples to detect change across all spatial scenarios.
   cf Figure 5-2a to Figure 5-3a
   Figure 5-2b to Figure 5-3b
   Figure 5-4a to Figure 5-5a
   Figure 5-4b to Figure 5-5b.
3. Use of the v8 estimator also raised the rate of Type I error in cases where there was a negative bias in variance estimates (variance underestimated). This is clearly seen in
   Figure 5-3b (see bias in Figure 4-6e, p.38)
   Figure 5-5a (see bias in Figure 4-3f, p.36)
   Figure 5-5b (see bias in Figure 4-6f)
4. Even under repeat sampling on the second occasion (which benefits SRS results), the power of STR (and SYS) sampling is more resilient under different patterns of heterogeneous change compared to SRS (Figure 5-3b and Figure 5-5b).

The last point supports the initial hypothesis (p.41) regarding the relative benefit of STR and SYS over SRS in terms of power to detect change with repeat transect sampling under scenarios of spatially heterogeneous change.

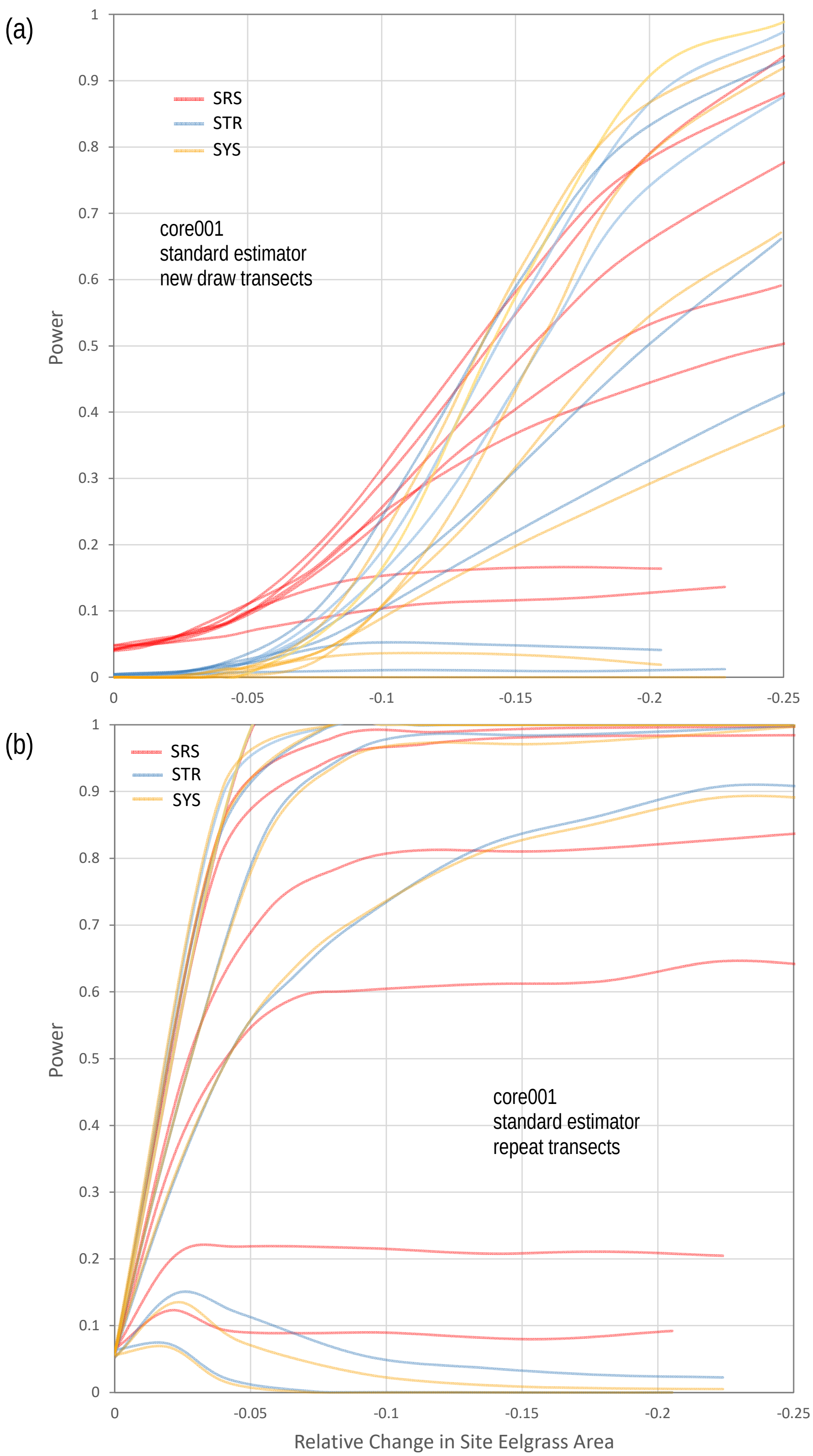


Figure 5-2. Power to detect spatially heterogeneous change in the core001 site model with standard estimators for newly drawn transects on the second occasion (a) and for repeat transects (b). Each curve represents the power achieved with different intensities of change with one spatial pattern of change.

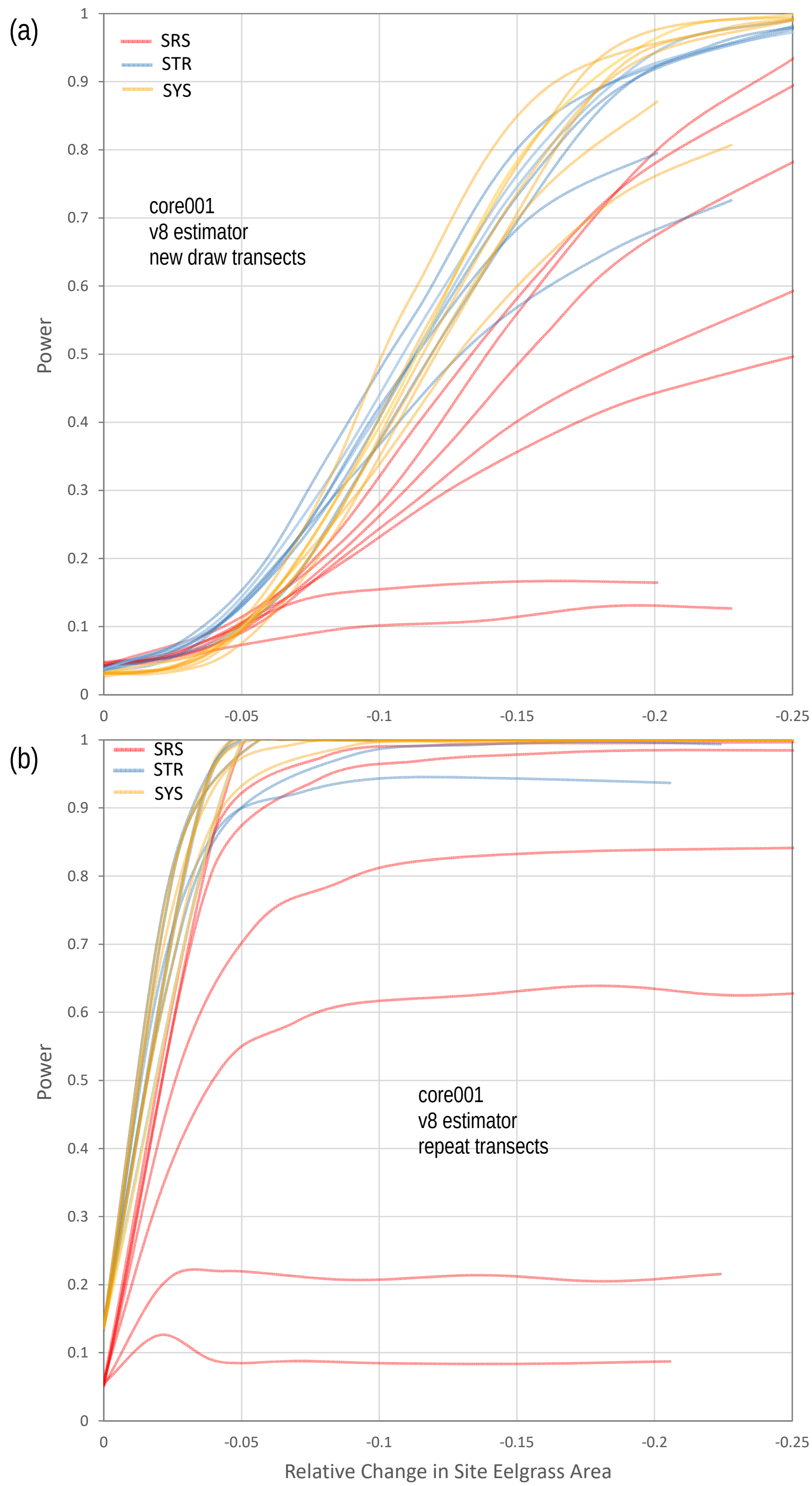


Figure 5-3. Power to detect spatially heterogeneous change in the core001 site model with the v8 estimator used for STR and SYS for newly drawn transects on the second occasion (a) and for repeat transects (b). Each curve represents the power achieved with different intensities of change with one spatial pattern of change.

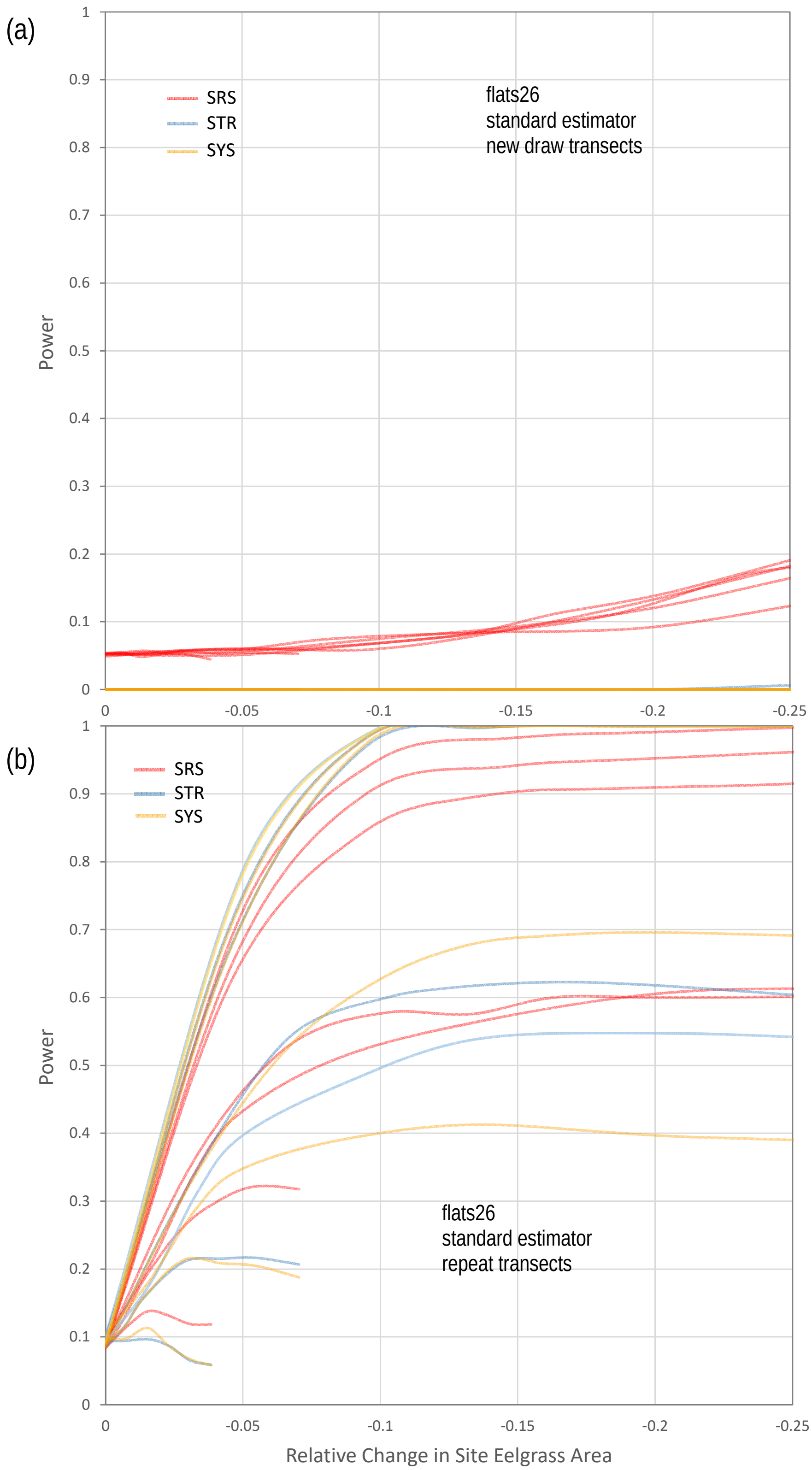


Figure 5-4. Power to detect spatially heterogeneous change in the flats26 site model with standard estimators for newly drawn transects on the second occasion (a) and for repeat transects (b). Each curve represents the power achieved with different intensities of change with one spatial pattern of change.

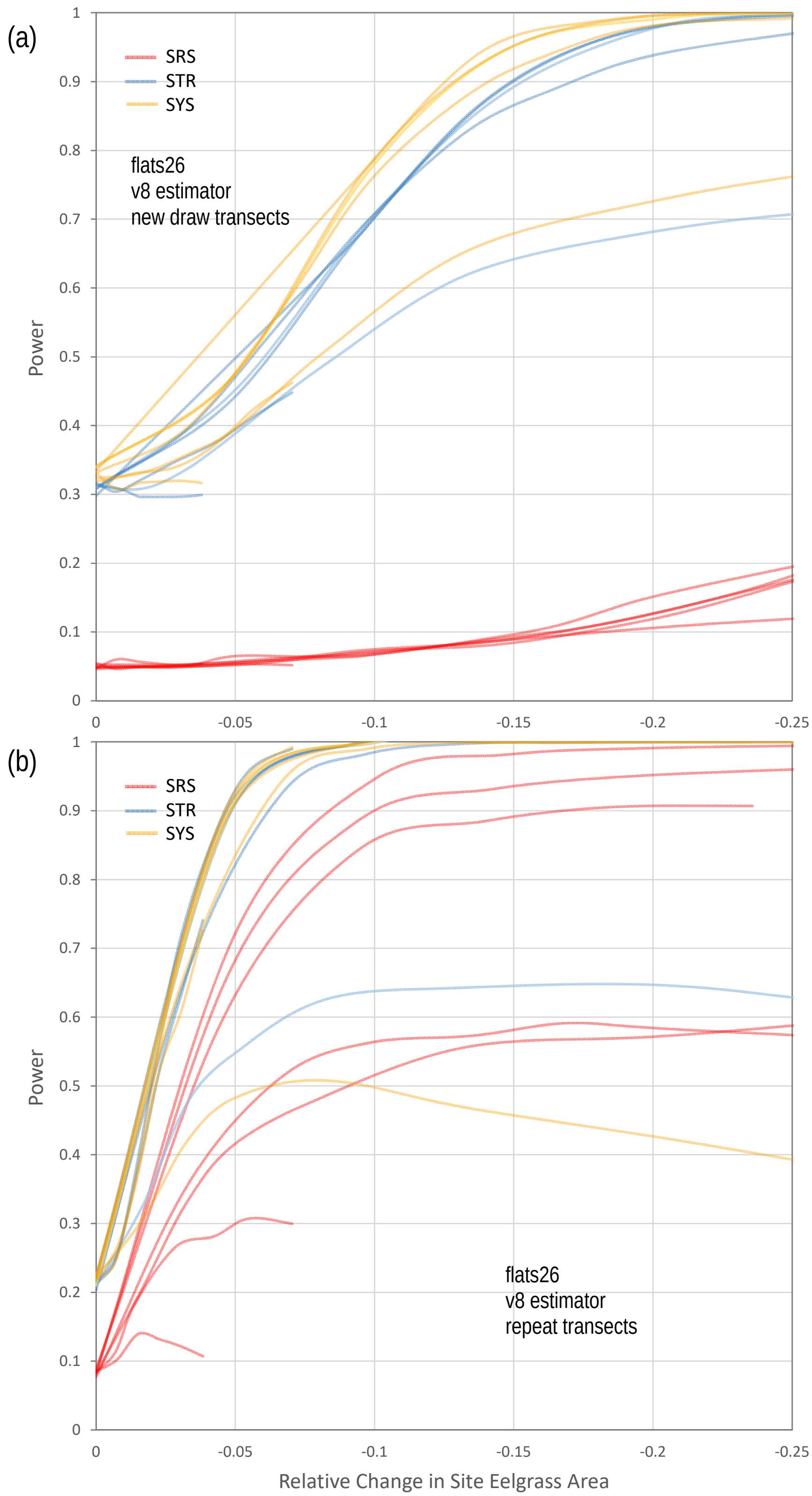


Figure 5-5. Power to detect spatially heterogeneous change in the flats26 site model with the v8 estimator used for STR and SYS for newly drawn transects on the second occasion (a) and for repeat transects (b). Each curve represents the power achieved with different intensities of change with one spatial pattern of change.

6 Effects of Periodicity

It is well known that SYS sampling is sensitive to periodicity in the population sampled (Cochran 1977, p.217). Sampling with STR is expected to be more robust in the presence of periodicity due to an independent random element in the selection of each sample unit. Nevertheless, it is reasonable to ask if STR sampling has some level of sensitivity to periodic variation in the population. This question is addressed in this section.

6.1 Methods

A simple modelling exercise addressed this issue by superimposing periodic functions of incremental transect fraction with different amplitudes on the core001 site model (Figure 6-1). The amplitudes investigated were 0.05, 0.10 and 0.15 (in units of transect fraction) and the results were compared with the site model without a periodic function superimposed (equivalent to superimposing a function with amplitude 0). The period of the functions was held constant at 1/10 of the site width based on the assumption that this period would maximize the effect of the periodicity when sampling with STR and SYS with n = 10. This is essentially looking at a worst case scenario.

In general, other aspects of simulated sampling from these modified core001 site models followed the earlier methods for sampling on one occasion and estimating site eelgrass area and standard error (section 2.3). A minor deviation from the earlier methods is that video classification error was not modelled. Results are presented based on the v8 variance estimator for STR and SYS sampling but the effects of periodicity were essentially the same when the standard variance estimator was used (results not presented).

6.2 Results

At the lowest amplitude of the periodic function (0.05), the effect on the population is not readily apparent upon inspection of the population (cf Figure 6-1 a and b). At the highest amplitude (0.15), the periodicity is easily discernable in the modified site model (Figure 6-1 d).

Results showing the effects of the superimposed periodic functions on estimation of site area and standard error are shown in Figure 6-2. Several specific responses to the superimposed periodic functions are highlighted here.

- Greater amplitude in the periodic function results in a larger standard error in the site eelgrass area estimates (Figure 6-2 b). This is seen for each sample selection method, including SRS, but the magnitude of the effect is clearly greater with SYS sampling.

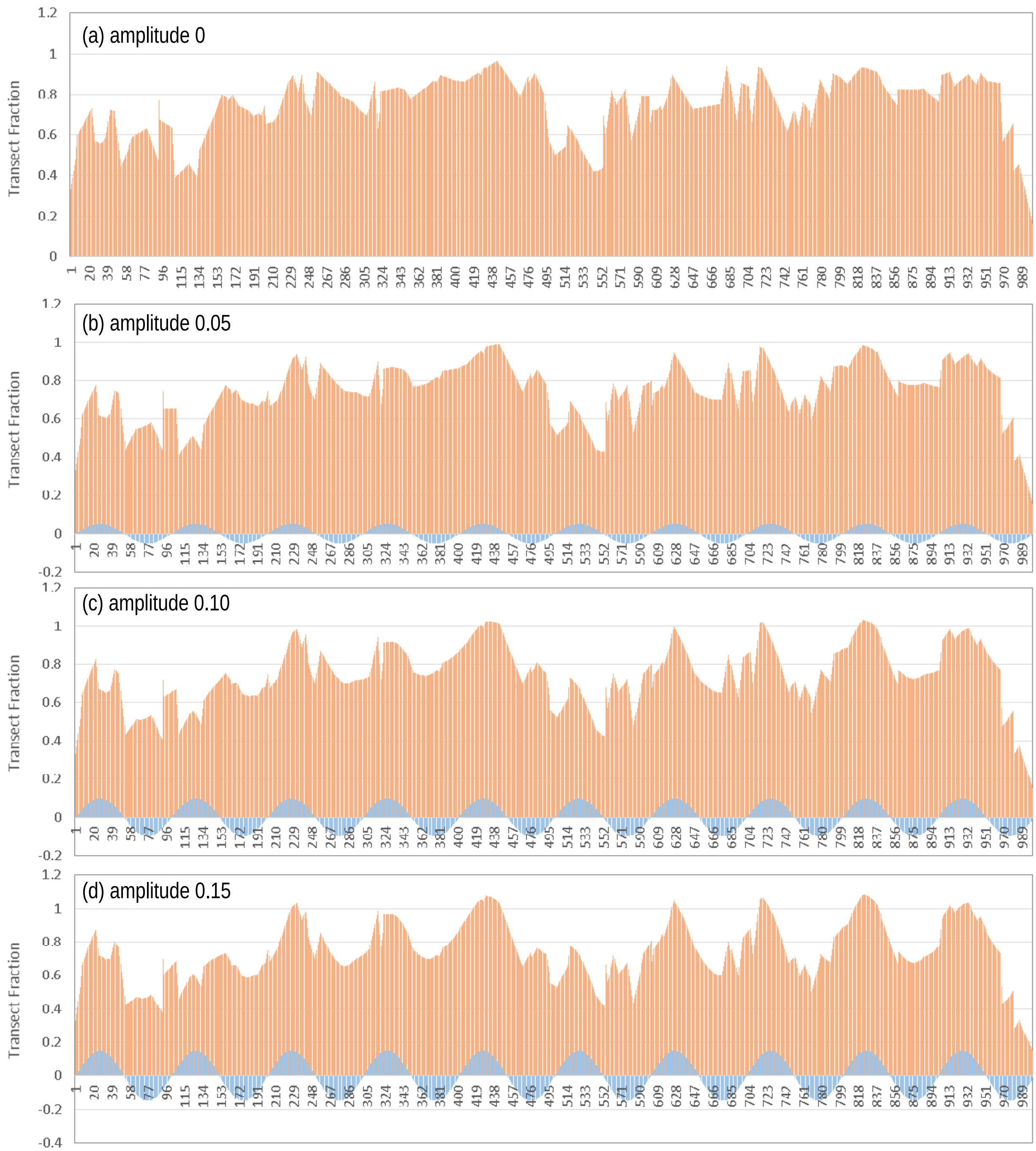


Figure 6-1. Variations on the core001 site model with a periodic function of increasing amplitude superimposed on the original site model. The top graph shows the original model. Below the original model, the blue series show the periodic function with amplitude 0.05 (b), 0.1 (c) and 0.15 (d) and the resulting site model in orange.

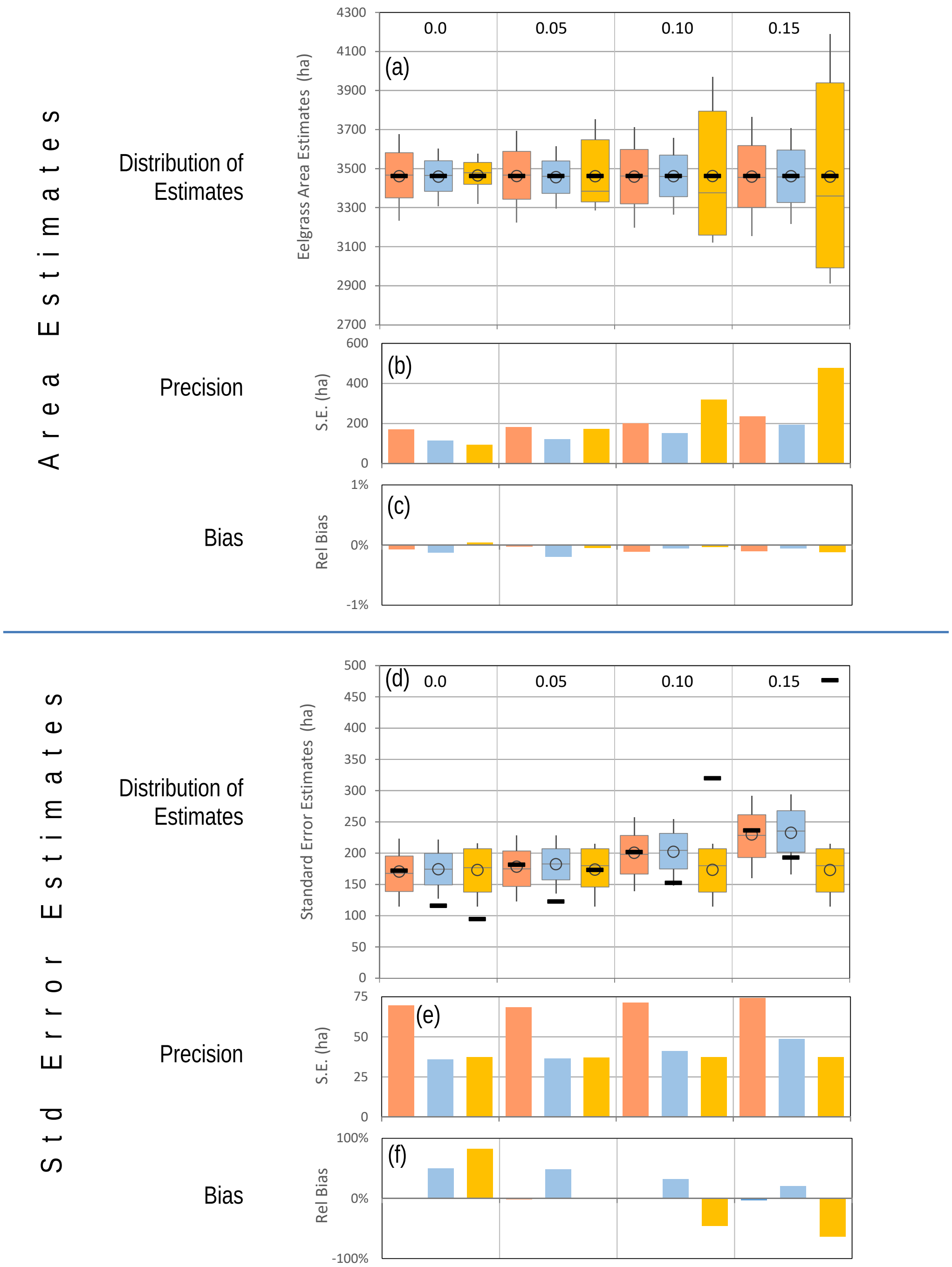


Figure 6-2. The effects of periodicity on the distribution of estimates and the associated precision and bias for estimates of eelgrass area (a, b, c) and for estimates of the standard error of the mean eelgrass area estimate (d, e, f). The estimates are based on samples drawn from the site models for core001 with periodic functions of transect fraction superimposed with amplitudes of 0, 0.05, 0.10 and 0.15 (labelled at top of a and d) and with sample selection by SRS (red), STR (blue) and SYS (yellow).

- Estimates of site eelgrass area remain unbiased for each selection method regardless of amplitude of the periodic function (Figure 6-2 c).
- The response of bias in estimates of standard error varied by sample selection method (Figure 6-2 f):
    - SRS sampling remained unbiased regardless of the amplitude of the periodic function.
    - The bias of SYS sampling was strongly affected by the amplitude of the periodic function. Bias varied from strongly positive (+83%) with no periodic function to strongly negative (-64%) with amplitude 0.15 with a roughly linear response at intermediate amplitudes.
    - The bias of STR was affected by the amplitude of the periodic function but it was much less sensitive than SYS. Bias varied from +50% with no periodic function to +20% with the maximum amplitude.

These results confirm the reported sensitivity of SYS to periodicity. They also show that STR has some sensitivity to periodicity but it is weak – certainly relative to the response of SYS but also given that the range of periodic amplitude investigated varied from 0 to a level resulting in obvious periodicity upon inspection of the population. The sensitivity seen of STR to periodicity appears to contradict the claim of Saunders and Robinson (1989, p.159) that STR is "not susceptible to error due to periodicity". We find here that the statement of Saunders and Robinson is true for the population estimate (Figure 6-2 a, b, c) but not for the estimate of standard error (Figure 6-2 f).

7 Potential Performance of the Sampling Methods

Consider three sets of point estimates that are generated from site sampling data:

- site eelgrass area (Figure 2-8 a,b, p.17)
- change in site eelgrass area under spatially homogeneous change with sample replacement (new draw transects) (Figure 3-3, p.25)
- change in site eelgrass area under spatially homogeneous change without sample replacement (repeat transects) (Figure 3-6, p.28).

In these point estimate results, the precision with STR and SYS is in each case at least as good as the precision with SRS. In most cases there is a strong improvement in precision relative to SRS. This leads us to conclude that STR and SYS have the potential to outperform SRS in two key measures of importance to the SVMP:

- width of confidence intervals on point estimates
- power to detect change

To realize this potential we need to know the variance, or standard error, of the point estimates. In practice, we don't know the standard error and so must rely on estimates of the standard error derived from the sample. This is where we run into trouble because we do not have an unbiased estimator for the standard error with STR and SYS sample selection. We get different results depending on which biased estimator of standard error we use (e.g., the standard estimator or the v8 estimator). This complicates the comparison of the performance of SRS, STR and SYS. In a sense, the inherent performance of the selection methods is confounded with the performance of the standard error estimators. In addition, the bias in the standard error estimates leads to murky comparisons where a gain in power to detect change may have to be weighed against an elevated rate of Type I error.

The purpose of this section is to eliminate the confounding effects of standard error estimation and present a clear, straightforward comparison of the three sample selection methods. The approach is simply to conduct analysis with actual standard errors derived from modeling rather than the biased sample estimates. The cost of this approach is that we ignore practical considerations that are central to implementation. Instead of comparing the performance of the selection methods that we would expect in practice, we are comparing potential performance under a theoretical scenario where an unbiased and perfectly precise standard error estimator is available. To the extent that estimators can be improved for STR and SYS, the operational results would be expected to approach the theoretical potential to varying degrees.

7.1 Methods

The objective is to reproduce the power curves reported earlier in this report but with calculations that are based on the true values of standard error, rather than estimates. While the motivation here was to avoid bias in the standard error estimation for STR and SYS, in order to give meaningful comparisons to SRS, the true standard error must also be used for SRS calculations. While SRS generally gives unbiased standard error estimates with the standard estimator, individual sample estimates will have error reflecting the precision of the standard error estimation. To avoid having these errors only for the SRS results, the true standard error is used for all three sample selection methods.

Power curves are generated for sampling with SRS, STR and SYS from the core001 and flats26 site models under the four scenarios of sample replacement and spatial pattern of change presented earlier (summarized in Table 7-1).

Table 7-1. Matrix of four scenarios for generation of potential power curves.

	Spatially homogeneous change	Spatially heterogeneous change
With sample replacement (new draw transects each occasion)	1. New draw – homogeneous	3. New draw – heterogeneous
Without sample replacement (repeat sampling of transects)	2. Repeat - homogeneous	4. Repeat - heterogeneous

As with the results presented earlier, each power curve is based on a series of power estimates for different change scenarios. A key difference here is that the *t*-test is replaced with a *z*-test since the standard error is a known quantity in these analyses. The sample estimates of standard error from the model iterations are not utilized. Each *z*-test of estimated change in site eelgrass area between two occasions is conducted with $\alpha = 0.05$. Each power estimate is based on 5,000 model iterations that result in 5,000 change estimates and 5,000 test results.

7.2 Potential Power to Detect Change in Site Area with a *z*-test.

7.2.1 New Draw Transects and Spatially Homogeneous Change

Figure 7-1 presents the potential power when sampling with replacement (new draw transects) from the core001 model. These results show greater potential power with STR and SYS relative to SRS with SYS having the highest potential.

A comparison of these power curves (Figure 7-1) with those generated when using the standard (Figure 3-4, p.26) and v8 (Figure 4-4, p.37) variance estimators give examples of how power is diminished when there is positive bias in the variance estimator. Both the standard and v8 estimators were positively biased (Figure 3-3 k and Figure 4-3 e, respectively) with the standard estimator having the greatest bias. For a change scenario of -0.15, the potential power under STR was 0.91. The power achieved when using the standard variance estimator was reduced to 0.55 because of the positive bias. The power achieved with the v8 estimator was closer to the potential (0.76) because of the smaller magnitude of positive bias.

Figure 7-2 presents the potential power to detect spatially homogeneous change when sampling with replacement from the flats26 model. Here the STR and SYS results are nearly identical with a moderate level of potential power (~0.81 with -0.15 change). In contrast, SRS has poor potential power (< 0.1 with -0.15 change).

A comparison of this potential power when sampling from the flats26 model (Figure 7-2) to the power obtained with the standard variance estimator (Figure 3-5, p.26) illustrates again how positive bias in variance estimation diminishes the power achieved. But in this case the magnitude of the bias is so great (> +400%) that the power achieved is actually reduced to zero.

A comparison of Figure 7-2 to the power obtained with the v8 estimator (Figure 4-5, p.37) illustrates the effects of negative bias in variance estimation. The power obtained under STR and SYS when using the v8 estimator is actually greater (0.9 for STR with -0.15 change) than the potential power (0.79 for STR with -0.15 change) but this is only possible because it is accompanied by an elevated Type I error rate of 0.31 (when testing with $\alpha = 0.05$). The potential power results (Figure 7-2) answer the question of what is the maximum power that could theoretically be achieved if Type I error were properly controlled?

7.2.2 Repeat Transects and Spatially Homogeneous Change

The potential power when sampling without replacement (repeat transects) from the core001 model is high (0.99 with -0.05 change) and is virtually identical across the three selection methods (Figure 7-3). This is essentially the same result as that obtained with the standard variance estimator (Figure 3-7, p.29) which is unbiased in this case (Figure 3-6 k, p.28). It is also similar to the result obtained with the v8 estimator (Figure 4-7, p.40) although in this case a modest negative bias of -10 to -20% (Figure 4-6 e, p.38) gave a modest elevation in Type I error for STR and SYS (0.13 when testing with $\alpha = 0.05$).

When sampling from the flats26 model with replacement (repeat transects) (Figure 7-4), STR and SYS have similar potential power and SRS has relatively lower potential power. Of the change scenarios modelled, the greatest separation in power occurred with a change of -0.05 where STR and SYS had power of 0.86 and 0.84, respectively, and SRS had a power of 0.69. A comparison of these results with those obtained when using the standard variance estimator (Figure 3-8, p.29) again reveals the effects of bias in variance estimation. For example, the standard estimator used with STR and SYS in the -0.05 change scenario was unbiased (Figure 3-6 l, p.28) and the power achieved was moderately lower (0.76 and 0.77 respectively) than the potential power (0.86 and 0.84 respectively). Under the no-change scenario, the standard estimator was negatively biased for STR and SYS (~ - 17%) and this was reflected in the elevated Type I error (0.08). For SRS, the standard estimator was negatively biased across all change scenarios (approx.. -15%; Figure 3-6 l) which is consistent with both the elevated Type I error (0.08) and the power elevated (0.72 for -0.05 change) above the potential (0.69 for -0.05 change). A comparison of the STR and SYS results obtained with the v8 estimator (Figure 4-8, p.40) are similar in that negative bias in the v8 estimator is manifested as power elevated above the potential but accompanied by an elevated Type I error.

7.2.3 Spatially Heterogeneous Change

The potential power of the spatially heterogeneous change scenarios are presented for sampling from the core001 site model with replacement (Figure 7-5 a) and without replacement (Figure 7-5 b) and for sampling from the flats26 model with replacement (Figure 7-6 a) and without replacement (Figure 7-6 b). Bias associated with the variance estimators was not examined earlier for the scenarios with spatially heterogeneous change but undoubtedly variations in bias would explain much of the differences seen between results with the standard estimator (Figure 5-2, p.44 and Figure 5-4, p.46), with the v8 estimator (Figure 5-3, p.45 and Figure 5-5, p.47) and the potential power (Figure 7-5, Figure 7-6).

Based on the examination of the potential power results alone, two key points are clear. First, SRS is most sensitive to loss of potential power associated with different patterns of heterogeneous change. Second, SYS has superior performance relative to STR (and SRS). In some cases this is seen as more resilient potential power in the face of the various scenarios of heterogeneity (Figure 7-5 b, Figure 7-6 a and b) but also in overall higher levels of potential power (Figure 7-6 a).

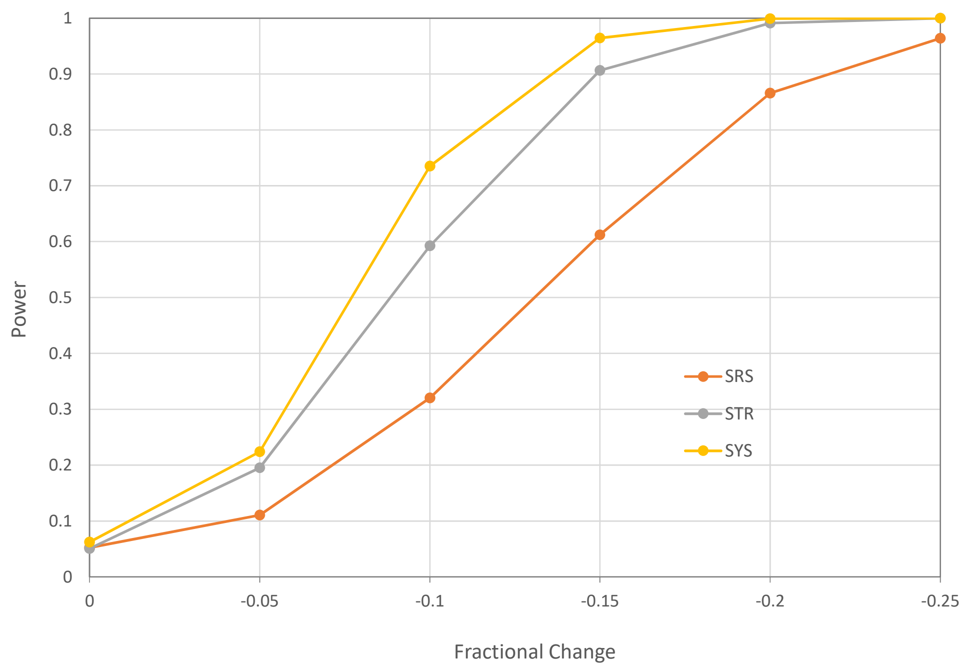


Figure 7-1. Potential power to detect spatially homogeneous change in the core001 site model with new draw transects (assumes standard error known exactly) (cf. power when using standard estimator, Figure 3-4, p. 26, and power when using the v8 estimator, Figure 4-4, p.37).

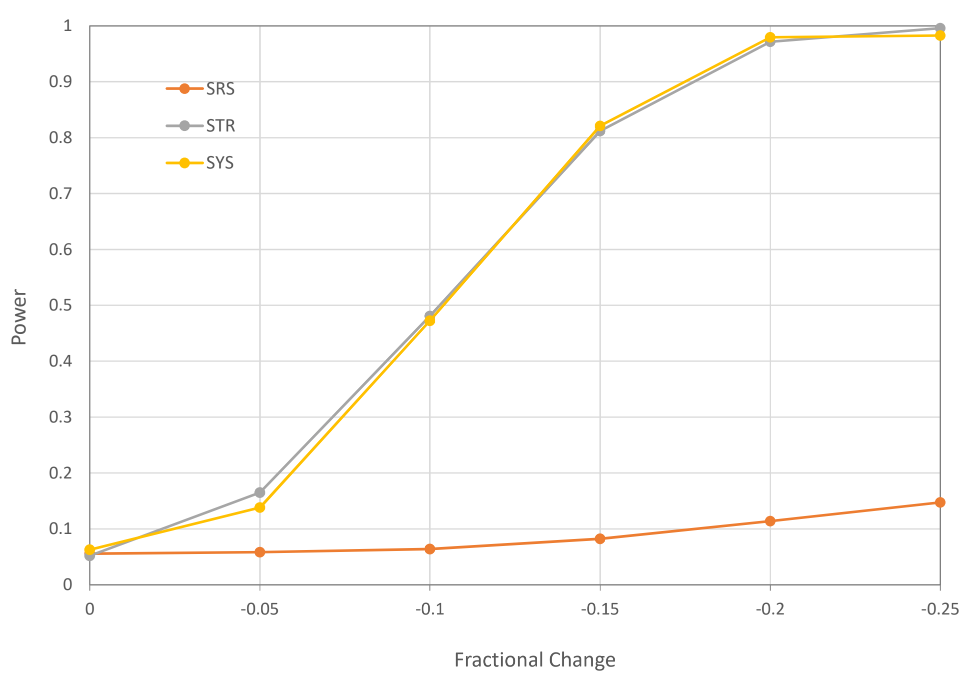


Figure 7-2. Potential power to detect spatially homogeneous change in the flats26 site model with new draw transects (assumes standard error known exactly) (cf. power when using standard estimator, Figure 3-5, p.26, and power when using the v8 estimator, Figure 4-5, p.37).

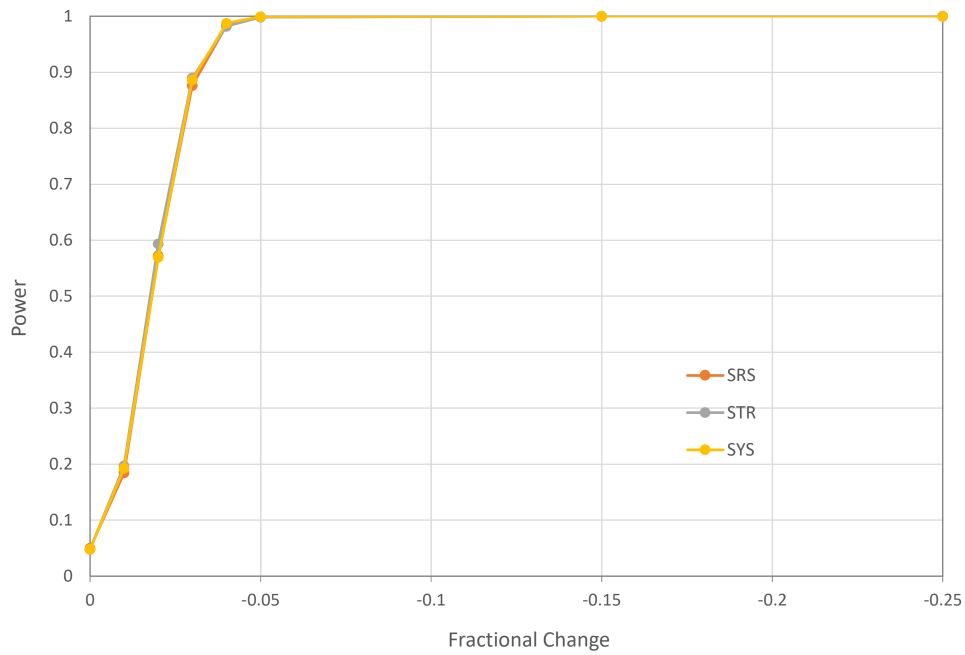


Figure 7-3. Potential power to detect spatially homogeneous change in the core001 site model with paired repeat transects based (assumes standard error known exactly) (cf. power when using the standard estimator, Figure 3-7, p.29, and power when using the v8 estimator, Figure 4-7, p.40).

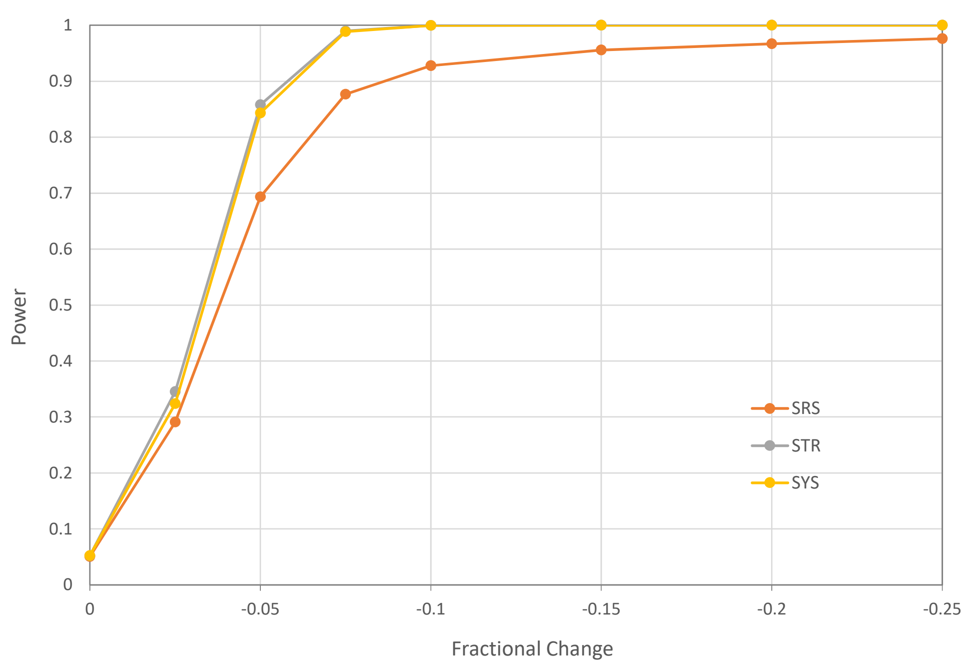


Figure 7-4. Potential power to detect spatially homogeneous change in the flats26 site model with paired repeat transects based (assumes standard error known exactly) (cf. power when using the standard estimator, Figure 3-8, p.29, and power when using the v8 estimator, Figure 4-8, p.40).

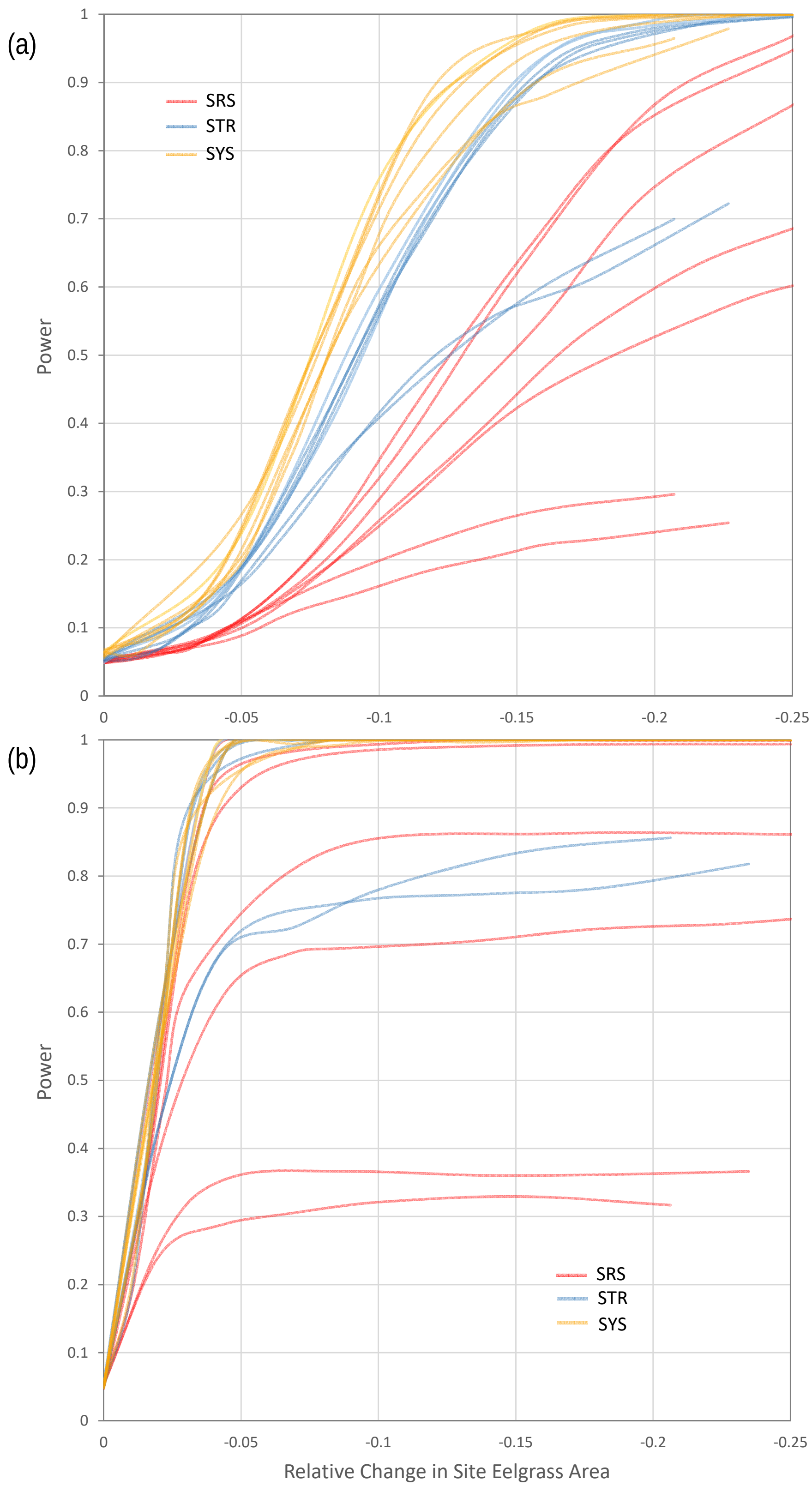


Figure 7-5. Potential power to detect spatially heterogeneous change in the core001 site model for newly drawn transects on the second occasion (a) and for repeat transects (b). Assumes the standard error is known exactly. Each curve represents the power achieved with different intensities of change with one spatial pattern of change.

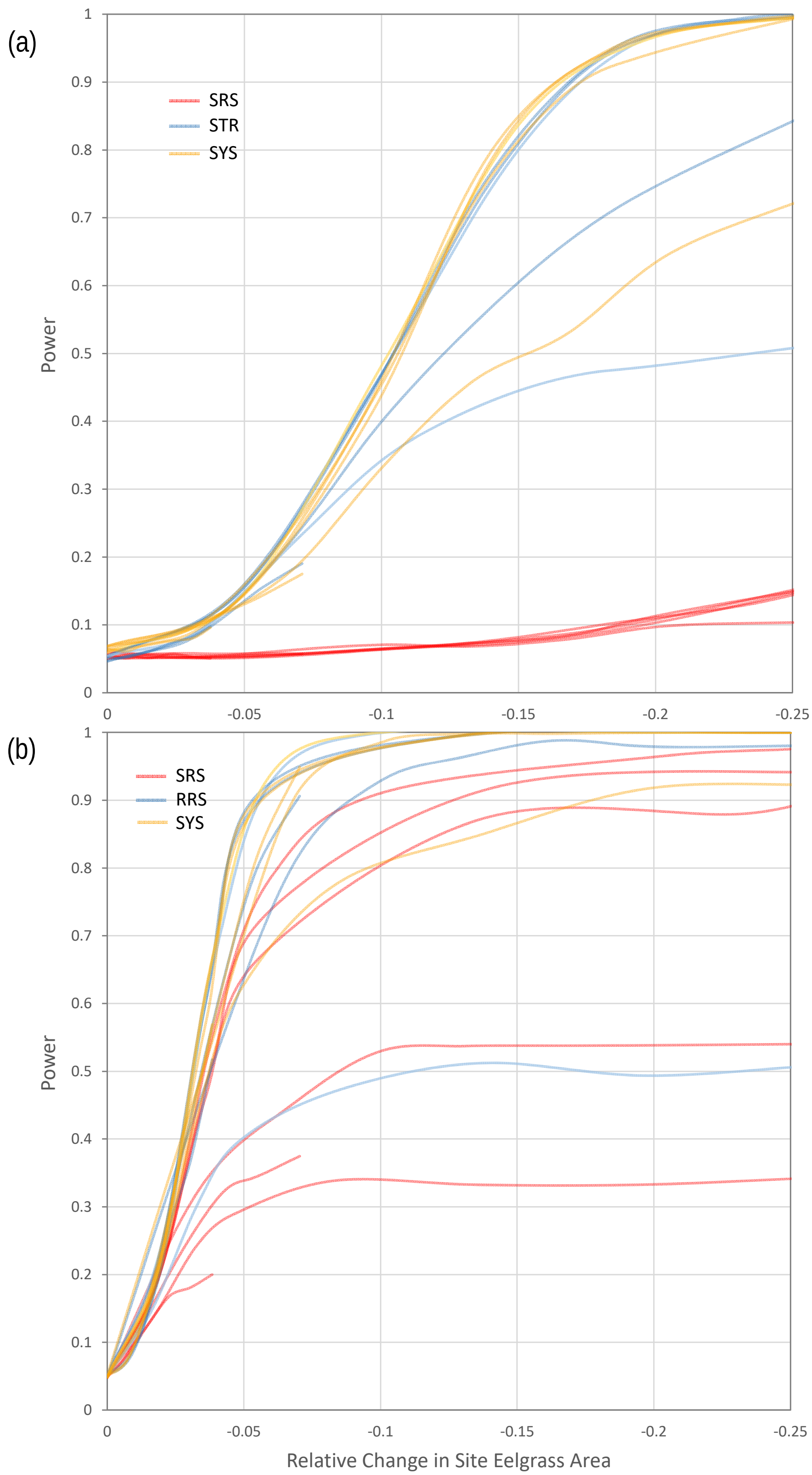


Figure 7-6. Potential power to detect spatially heterogeneous change in the flats26 site model for newly drawn transects on the second occasion (a) and for repeat transects (b). Assumes the standard error is known exactly. Each curve represents the power achieved with different intensities of change with one spatial pattern of change.

8 Discussion

8.1 Synthesis of Results

Let us first consider the performance of the sample selection methods when estimating site eelgrass area from a single sampling occasion. STR and SYS are clearly superior to SRS in terms of precision of the area estimate for the two site models studied (Figure 2-8 a, b; p.17). In the case of sampling from the flats26 site model, the gain in precision with the use of STR or SYS over SRS is very large.

The one advantage that SRS has over STR and SYS is in the sample estimation of the precision, e.g., the standard error. The standard variance estimator is generally unbiased under SRS whereas there is no similarly unbiased estimator for STR and SYS. While we saw that an estimator may be unbiased for a particular scenario under STR and SYS, it will not remain unbiased across a variety of scenarios. For example, the standard estimator was unbiased for the standard error of difference only for a -0.05 change scenario for STR and SYS sampling from flats26 (Figure 3-6 l, p.28). If a reliable variance estimator can be identified for area estimation with STR and SYS at SVMP sites, then STR and SYS would be preferable sample selection methods. SYS slightly outperformed STR when sampling from the core001 model, but further investigation would be needed before this could be considered a general result for area estimation.

One way to explain the contrasting precision of the selection methods is the fact that STR and SYS benefit from producing "spatially balanced" samples in the sense of Stevens and Olsen (2004). They state that "the concept that some degree of spatial regularity should be used for sampling for environmental populations is well established." The essence of this perspective is that natural populations are autocorreleated and sample clumping that occurs with SRS results in:

- over representation of some areas, under representation in others
- lower effective sample size as nearby transects will be correlated
- greater fluctuations in estimates due to lower effective sample size.

To the extent that spatial balance of the sample explains the greater precision of STR and SYS relative to SRS for area estimation, then SYS would be expected to be optimal because the sample is perfectly balanced (i.e., equally spaced) whereas STR samples can reflect some degree of clumping although to a much lesser degree than with SRS.

Next, let us consider the performance of the different sampling methods for estimating the difference in eelgrass area between two sampling occasions. Now the sampling methods compared include differences in sample replacement (new draw vs. repeat sampling) in addition to sample selection. If we consider sampling with replacement (new draw each occasion) with SRS transects as the status quo, then we can compare the gain in power when shifting to each of the other sampling methods investigated. It is simplest to conduct this comparison on the basis of potential power (section 7) since this bypasses issues of performance of the variance estimators. Figure 8-1 presents the gain in potential power for the various sampling methods for spatially homogeneous change scenarios.

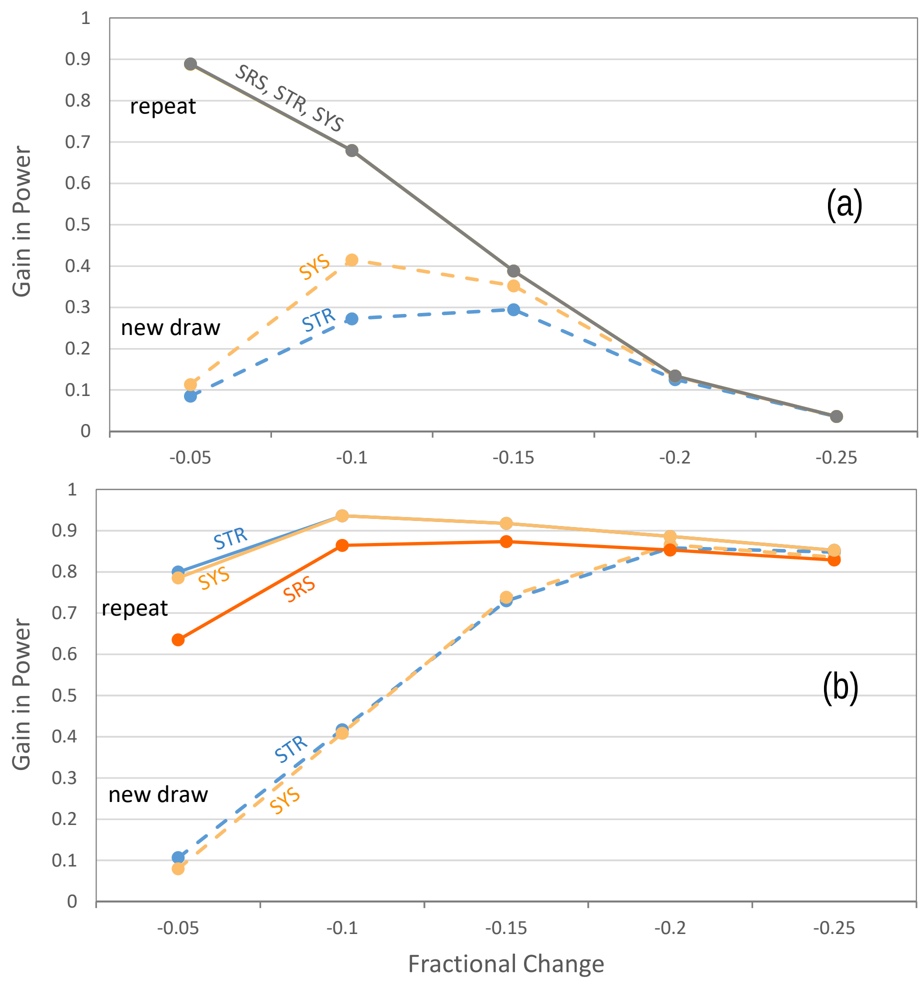


Figure 8-1. The gain in potential power (y-axis) to detect change in site eelgrass area (x-axis) when shifting from SRS sampling with replacement (new draw transects) to other combinations of sample selection and replacement. Gain in potential power is shown for sampling from the core001 site model (a) and the flats26 site model (b). Results from methods that sample with new draw transects are shown with dashed lines and results from methods that sample with repeat transects are shown in solid lines. This figure combines results from Figure 7-1, Figure 7-2, Figure 7-3 and Figure 7-4 (pp.57-58).

It is clear from Figure 8-1 that eliminating sample replacement (i.e., moving to repeat transects) is the single methodological change with the greatest potential to increase the power to detect change in eelgrass area. Secondarily, a move from SRS to STR or SYS also has the potential to increase power. When sampling from the core001 model (Figure 8-1 a), shifting to repeat transect sampling strongly increases the potential power, particularly at lower levels of change, with no differentiation between selection method. In contrast, when sampling from the flats26 model, moving to repeat transects strongly increases the potential power across all levels of change studied and an additional change to STR or SYS selection gives an additional incremental boost to the potential power. When sampling with replacement is retained, shifting to STR or SYS selection still brings a gain in potential power to detect change.

These results are explained as follows. When the difference in area is estimated from a new draw of transects on occasion two, the variance in the difference estimate will include a component due to spatial variability in addition to a component due to temporal variability, which is the component of interest. Paired analysis of repeat transects effectively controls for spatial variability and eliminates this component of the variance in the difference estimate. Depending on the relative variability in space and time, this can lead to large gain in precision. When sampling with replacement, the spatial balance aspect of the STR or SYS sample also controls to some extent the spatial variability, but not as effectively as repeat sampling. The benefit of repeat sampling for change detection is a basic concept. Cochran (1977, p.345) states clearly that “for estimating change, it is best to retain the same sample throughout all occasions.” He is referring to a different analysis scenario (analogous to unpaired analysis of repeat transects) but the same principle holds here.

We must also consider the case where STR sampling does not outperform SRS. We did not see this when sampling from the two site models studied here, but subsequent ongoing work has demonstrated that this does occur. How do we explain this and how do we predict when this will happen? This can be explained if we consider STR as a form of stratified random sampling. Cochran (1977) addresses the relative precision of SRS and stratified random sampling and reduces the question to an ANOVA (Cochran 1977, p.101). Conventional stratified random sampling (and by extension STR) improves on the precision of SRS only if variability within strata is less than that between or among strata. STR will not perform well relative to SRS when the stratum means are similar but there is high variability within strata.

We would expect SYS to outperform SRS except when there is spatial periodicity present in the population (Figure 6-2 b). It also seems that SYS outperforms STR in some cases but this would also not apply in cases with periodicity.

The results presented in Figure 8-1 are limited in two key respects. First, the scenarios were constrained to spatially homogeneous change. When we expand our scope to include spatially heterogeneous change we find that our key results are further supported. Repeat sampling of transects again provides higher potential power generally. Furthermore, the benefits associated with STR and SYS appear even stronger because these selection methods are more resilient to loss of power associated with different patterns of spatial heterogeneity in area change.

The second limitation embodied in Figure 8-1 is that in presenting potential power (which is based on knowing standard error exactly), the complexity of standard error estimation is ignored. This complexity must be addressed before either STR or SYS can be reliably implemented. The two estimators investigated here, the standard estimator and the v8 estimator, had widely varying levels of bias across the model scenarios which can lead to varying levels of power and Type I error. The most extreme example of this was seen when sampling from the flats26 model with replacement. The relevant results are summarized in Figure 8-2 a. In concept the potential power (obtained when the standard error is known) represents a theoretical maximum power that is approached as estimators reduce bias and maximize precision. In this case, however, the standard estimator was so strongly positively biased that power achieved with it is zero regardless of the intensity of change. The v8 estimator for these scenarios was negatively biased which boosted the power but only with an accompanying high level of Type I error. This is the extreme example of how choice of estimator can radically change the results. The results with repeat transects were more consistent (Figure 8-2 b). Nevertheless it is clear that a method of standard error estimation must be developed that has acceptable reliability across all site populations sampled. Only when this has been completed will STR or SYS be ready for reliable implementation by the SVMP.

8.2 Implications for SVMP Site Sampling

Starting with 2016 sampling, the SVMP is implementing repeat transect sampling for all sites sampled as part of the soundwide study (i.e., sites sampled for the purpose of makes estimates of parameters of the Puget Sound eelgrass population). There has been increasing emphasis placed on the estimation of change within the program (as opposed to improving estimates of the total eelgrass area). Given this emphasis, the results presented here add support for the move to repeat transects and provide further rationale for this change in sampling design. These results also improve our understanding of the improvements that might be expected.

The SVMP began sampling with STR transects at experimental or special study sites in 2012 while retaining the use of SRS transects at sites sampled for the soundwide study. Starting with 2016 sampling, soundwide sites will be sampled with two different samples – a sample of SRS-repeat transects and a sample of STR-repeat transects. The general plan is to maintain these two site sampling methods in parallel for a small number of transition years before moving to one method only which is expected to be STR-repeat.

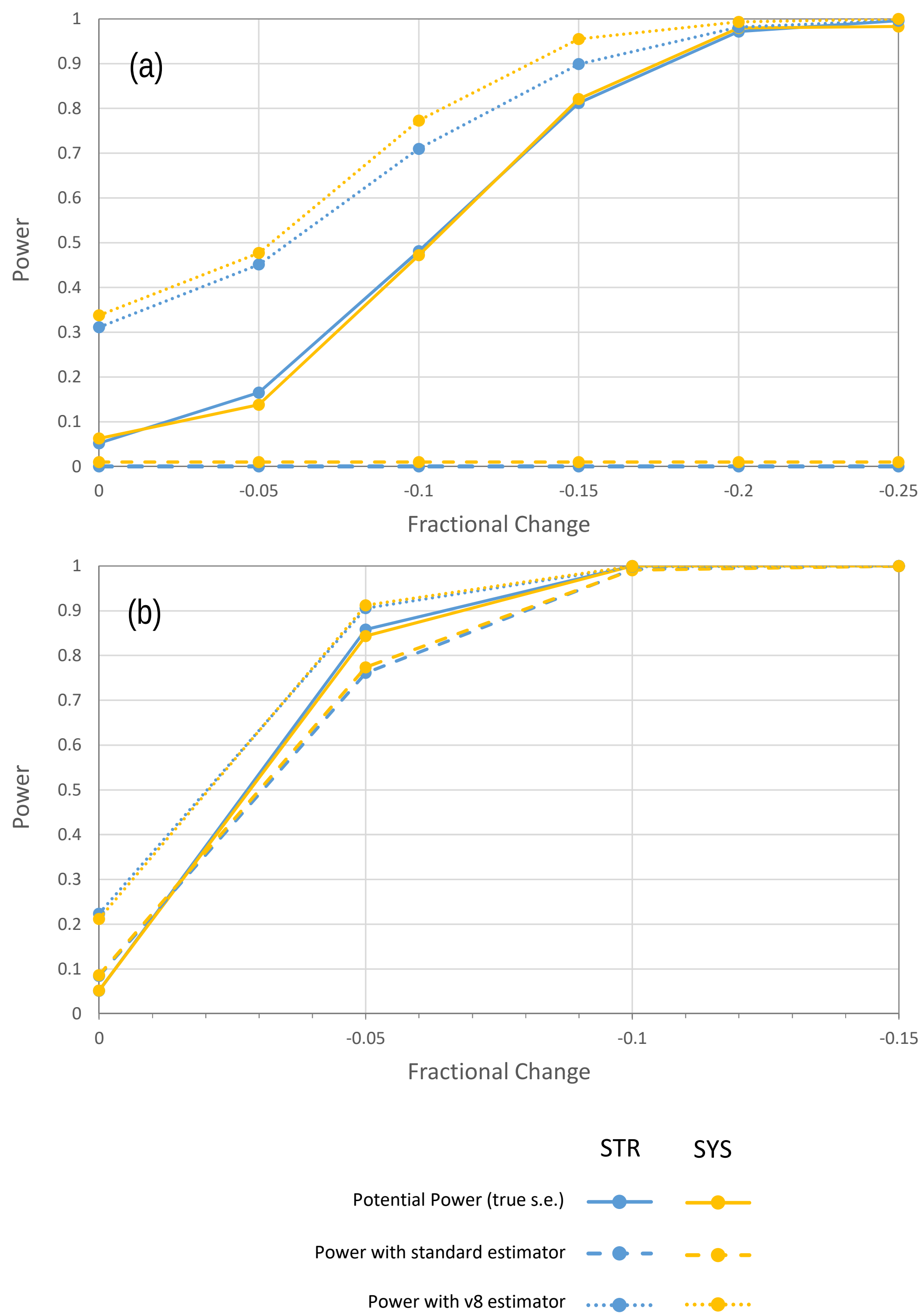


Figure 8-2. Comparison of power achieved to detect change in eelgrass area when sampling from the flats26 model with replacement (a) and without replacement (b). Each graph compares power achieved when the standard error is known (potential power), is estimated with the standard estimator and is estimated with the v8 estimator.

The results presented here add support for this approach and also provide useful information for the planning process for future decision making on site sampling design. Specifically, this study has described the potential of STR sampling for improved precision and power to detect change. Planning for a transition to STR sampling is well justified. Perhaps more importantly, this study has also highlighted the difficulty in estimating variance under STR sampling and the need for this issue to be adequately resolved before the new sampling design is complete and the transition period can be closed. The results have shown the need for an expanded effort to identify an approach to variance estimation that will be adequate across the breadth of sites encountered in Puget Sound. It is critical to use the transition period to build more site models and test more variance estimators. In the unlikely outcome that an adequate solution to variance estimation cannot be identified, the SVMP can fall back on SRS-repeat sampling at the end of the transition period.

While the SVMP has focused on transect selection with SRS and STR, the work in this report raises the question of why SYS has not been equally considered. In terms of benefiting from spatially balanced samples, SYS would be preferable to STR. In terms of the need to resolve the variance estimation problem, this effects both STR and SYS and it is not clear that either selection method has advantage on this issue. One possible explanation is that the variance estimation drawback of SYS was recognized early on. This led to interest in STR as an alternative under the assumption that the variance estimation problem would be alleviated. The work presented here suggests this problem is not alleviated but also suggests there is not a strong argument for either STR or SYS over the other. The presence of periodicity in the population was shown to be a problem for SYS but in the absence of more detailed analysis of site populations it is not clear to what extent this is a curiosity rather than something of practical importance.

8.3 Next Steps

The intent of this report was to contribute to efforts to improve SVMP site sampling – particularly with respect to incorporating repeat transect sampling and STR selection into the site sampling methodology. This report has shown that there is a large potential for improved monitoring with repeat sampling and STR. There is also an associated risk of producing misleading results due to bias in the variance estimators. Some specific steps are suggested here to move toward specific analysis procedures for operational use. The general theme is to further the approach taken in this report and follow the advice of Wolter (1984) to “construct plausible models of the population and … try out different reasonable variance estimators”.

- Construct additional site models from contrasting sites. This work has already progressed in related work with the objective of assessing the effect of varying sample size with STR.
- Assess if adjustment of the v8 estimator based on performance with a number of site models can improve bias. Specifically, if enough site models are

constructed and tested, it may be possible to replace the correction factor (Equation 4-5) with a curve fit to SVMP site model results.

- Test other variance estimates designed for SYS to assess bias when applied to a range of site models (e.g., estimators in Wolter 1984).
- Investigate the performance of STR with transect–based site trend analysis. If a site trend analysis is based on transect-level regression then explicit estimation of variance on the site area or change estimates would not be necessary. This analysis could potentially gain importance over time as site STR data records become more extensive.
- Develop site sampling protocols and analysis procedures for cases where eelgrass is found outside the sampling polygon within which repeat transects were established.

8.4 A Potential Alternative Analysis Paradigm

In the course of completing this report, we were struck that the spatial relationships between transects play no role in the quantitative analysis. The analysis is throwing out valuable information. A new analysis paradigm is needed that incorporates spatial relationships in the sample data. Gregoire (1998) suggests that a model-based framework for SVMP site data analysis could consider spatial context. This framework could augment or replace our current approach of design-based inference.

A new paradigm could build on the site model approach developed in this report as the basis of estimation of site eelgrass parameters (rather than the sample). Based on the effort associated with building site models for this study, it would be possible to generate site models for the 240 sites included in the SVMP soundwide study.

A site analysis method that incorporates spatial pattern could expand the current analysis into the domain addressed in recent years by the detailed inspection of transects by experts. This detailed inspection has served both to interpret spatial patterns at a site and to confirm that the statistical results were not due to transect clumping in unrepresentative areas. This detailed inspection of transects seemed at the same time to be a powerful way to bring human pattern recognition to bear on a complex problem and a sometimes discomfiting reversal of an objective numerical result by a qualitative and subjective assessment. Analysis based on a paradigm that incorporates spatial pattern would reduce the need for manual inspection of transects while benefiting from the additional information.

A third characteristic of a new paradigm is that the sample in hand is not considered just one of an infinite (or large) number of possible samples whose unique individual characteristics are irrelevant. Under the current paradigm the sample is considered only one instance randomly drawn from a population of equally likely samples that are functionally equivalent – completely interchangeable from the statistical perspective. Under a new paradigm, the unique features of the sample in hand are important. In reality, we have one specific sample in hand that is the product of substantial field and analysis resources. It is in our interests to understand as completely as possible what is unique about this particular sample. It may be legitimate to treat the results from this sample as one instance drawn from a sampling distribution but with an understanding of this unique sample we may be able to place it for example precisely within this distribution.

References

Bailey, R.A. 1987. Restricted Randomization: A Practical Example. *Journal of the American Statistical Association*. 82(399):712-719.

Cochran, W.G. 1946. Relative Accuracy of Systematic and Stratified Random Samples for a Certain Class of Populations. *Annals of Mathematical Statistics*. 17:164-177.

Cochran, W.G. 1977. *Sampling Techniques*. John Wiley and Sons, New York.

GNU. 2015. *GNU Scientific Library – Reference Manual* (online). Retrieved 2/3/2016 from
http://www.gnu.org/software/gsl/manual/html_node/Weighted-Samples.html#Weighted-Samples

Gregoire, T.G. 1998. Design-based and model-based inference in survey smapling: appreciating the difference. *Canadian Journal of Forest Research*. 28(10):1429-1447

McGarvey, R., R. Burch and J.M. Matthews. 2016. Precision of systematic and random sampling in clustered populations: habitat patches and aggregating organisms. *Ecological Applications*. 26(1):233-248.

Milton, J.S. and J.C. Arnold. 1990. *Introduction to Probability and Statistics*, McGraw-Hill.

Skalski, J.R., A. Hoffman, B.H. Ransom, and T.W. Steig. 1993. Fixed-Location Hydroacoustic Monitoring Designs for Estimating Fish Passage Using Stratified Random and Systematic Sampling. *Canadian Journal of Fisheries and Aquatic Sciences*. 50:1208-1221.

Skalski, J.R.. 2003. Statistical Framework for Monitoring *Zostera marina* (Eelgrass) Area in Puget Sound. Appendix L in Berry et al. 2003. *Puget Sound Submerged Vegetation Monitoring Project: 2000-2002 Monitoring Report*. Washington Department of Natural Resources. Olympia WA.

Stevens, D.L., Jr., and A.R. Olsen. 2004. Spatially Balanced Sampling of Natural Resources. *Journal of the American Statistical Association*. 99(465):262-278.

Thompson, S. 2012. *Sampling*. Wiley.

Wolter, K.M. 1984. An Investigation of Some Estimators of Variance for Systematic Sampling. *Journal of the American Statistical Association*. 79(388):781-790.

Zar, J.H. 1999. *Biostatistical Analysis*, 4th edition, Prentice-Hall.

Appendix A Normality of Differences in Area Estimates (New Draw Transects)

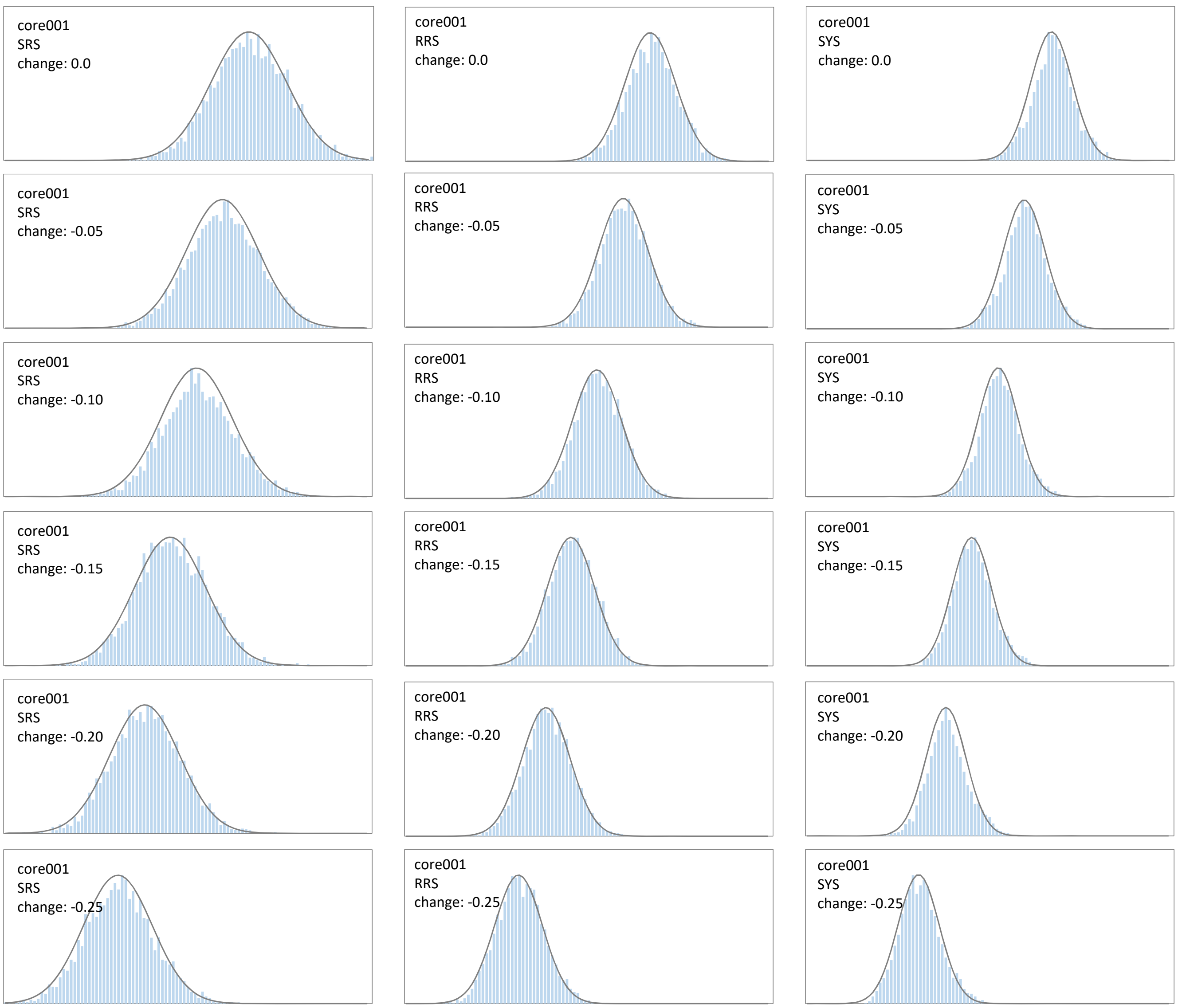


Figure A-1. Histograms of estimates of difference in site area (*n*=5000) based on sampling with *n*=10 new draw transects from the core001 site model using SRS (left column), STR (middle column) and SYS (right column) transect selection. Each row represents a different nominal change value (indicated in graph labels) applied across the initial site model to derive the model for the second sampling occasion. The curve on each graph shows the normal distribution with the same mean and standard deviation as the sample difference data. The *x*-axes for all graphs are identical. The corresponding Q-Q plots to better assess normality are shown with the same layout of scenarios in Figure A-2.

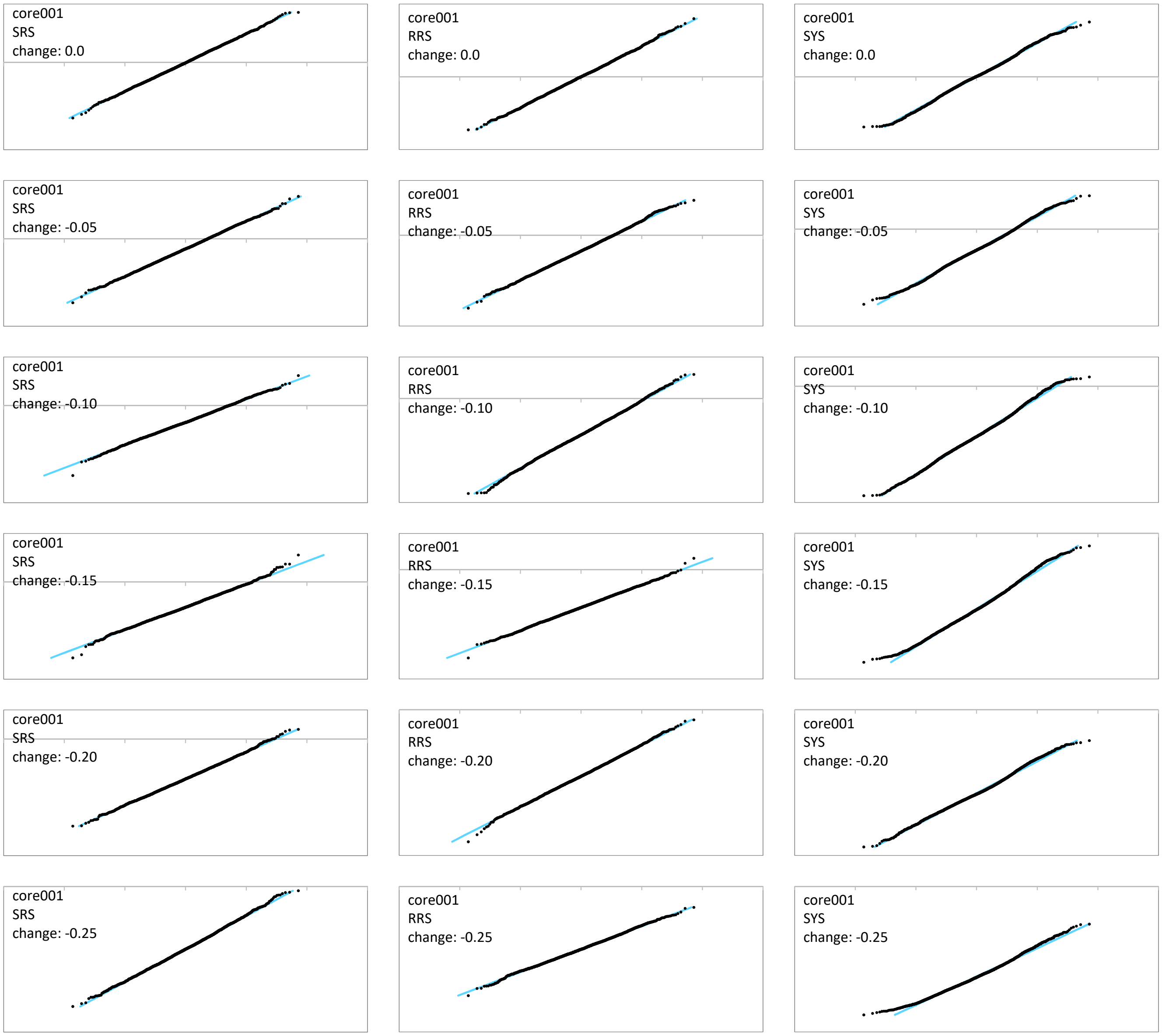


Figure A-2. Normal Q-Q plots to assess normality of site area estimates for the core001 site model. The graphs have the same layout of scenarios as the histograms shown in Figure A-1.

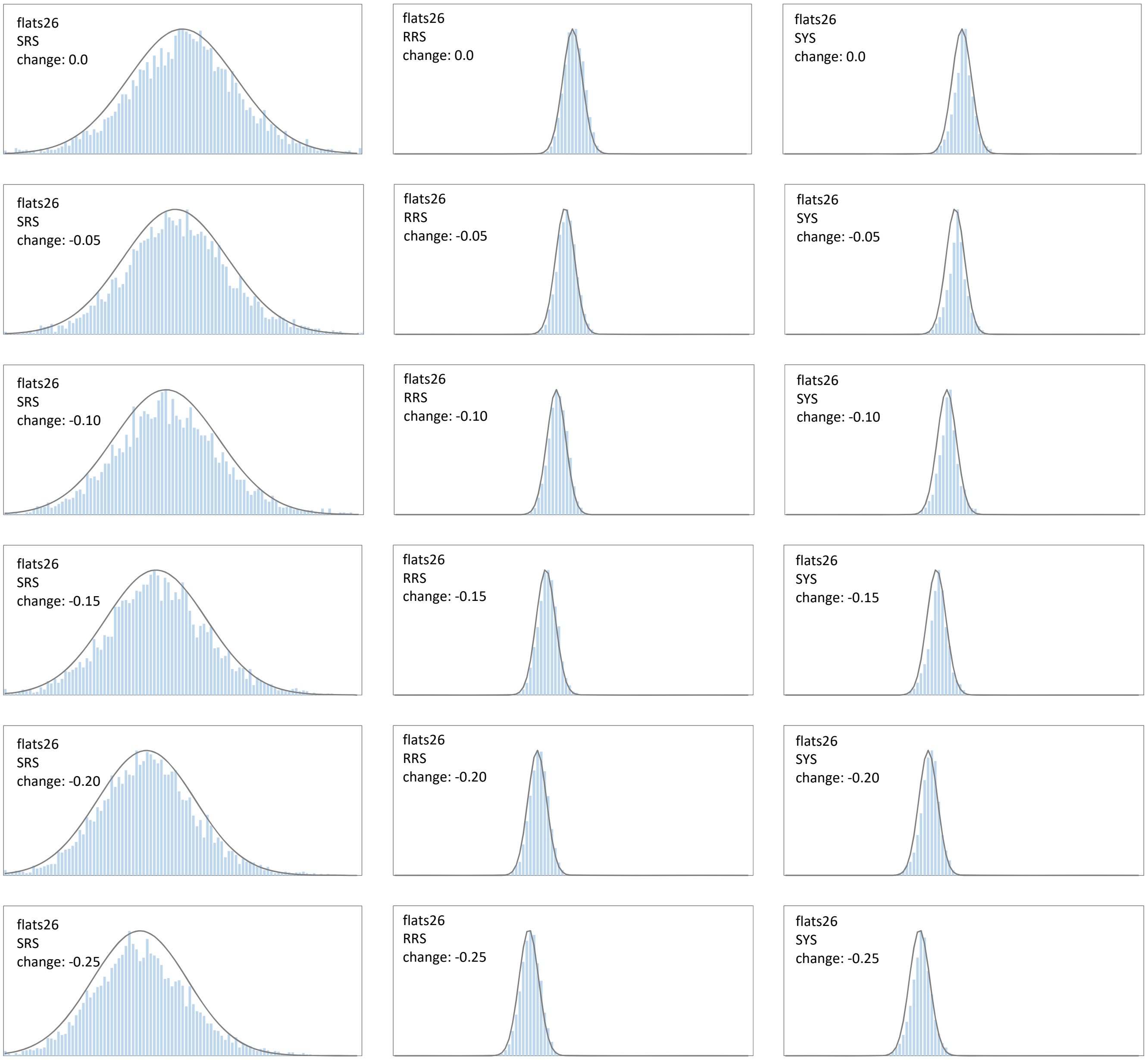


Figure A-3. Histograms of estimates of difference in site area (*n*=5000) based on sampling with *n*=10 new draw transects from the flats26 site model using SRS (left column), STR (middle column) and SYS (right column) transect selection. Each row represents a different nominal change value (indicated in graph labels) applied homogeneously across the initial site model to derive the model for the second sampling occasion. The curve on each graph shows the normal distribution with the same mean and standard deviation as the sample difference data. The *x*-axes for all graphs are identical. The corresponding Q-Q plots to better assess normality are shown with the same layout of scenarios in Figure A-4.

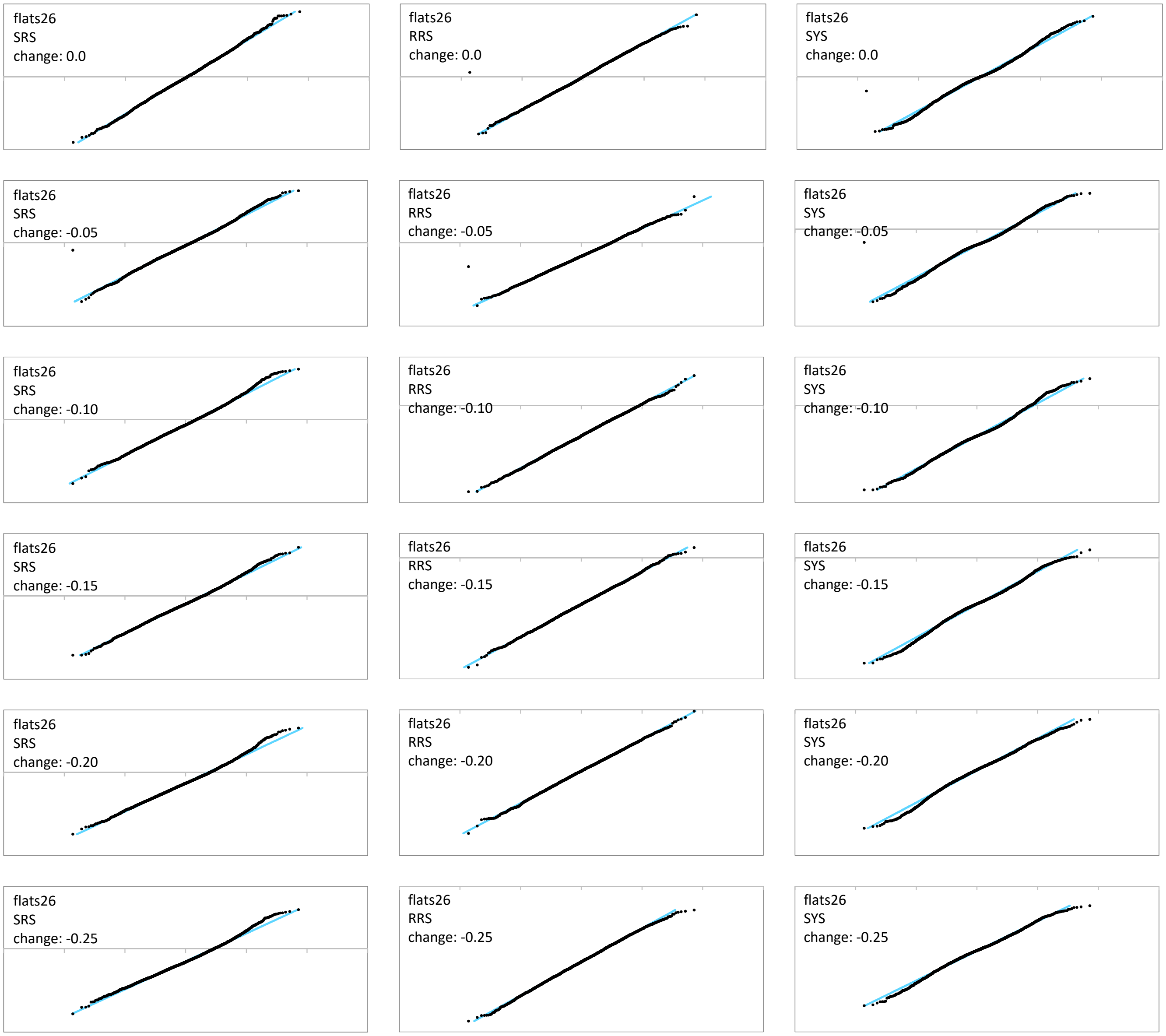


Figure A-4. Normal Q-Q plots to assess normality of the site area estimates for the flats26 site model. The graphs have the same layout of scenarios as the histograms shown in Figure A-3.

Appendix B Normality of Estimates of Difference in Area (Repeat Transects)

Figure B-1. Histograms of estimates of difference in site area (***n***=5000) based on repeat sampling of ***n***=10 transects from the core001 site model using SRS (left column), STR (middle column) and SYS (right column) transect selection. Each row represents a different homogeneous nominal change value (indicated in graph labels) applied homogenously across the initial site model to derive the model for the second sampling occasion. The curve on each graph shows the normal distribution with the same mean and standard deviation as the sample difference data. The *x*-axes for all graphs are identical. The corresponding Q-Q plots to better assess normality are shown with the same layout of scenarios in Figure B-2.

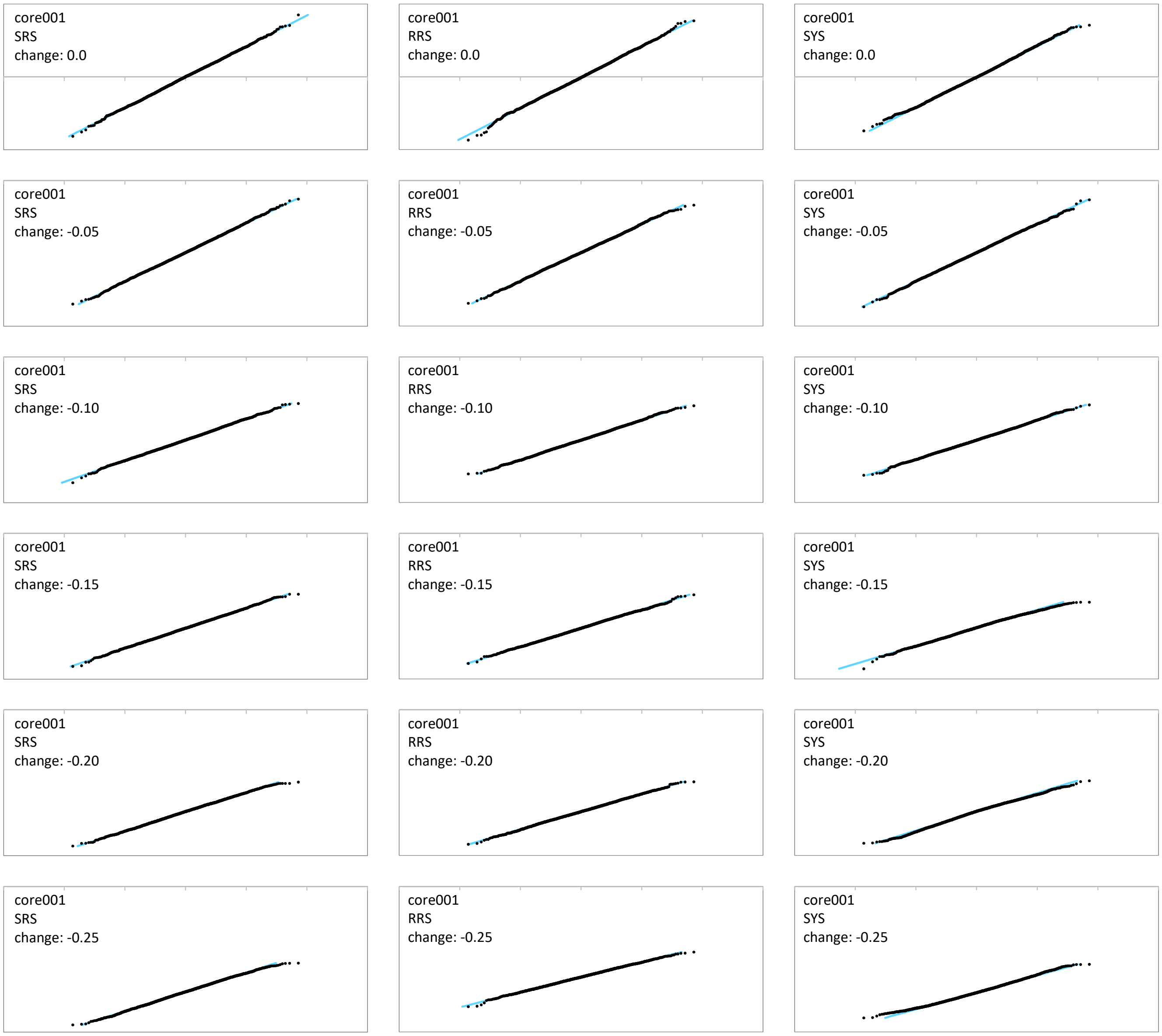


Figure B-2. Normal Q-Q plots to assess normality of the estimates of difference in site area for the core001 site model. The graphs have the same layout of scenarios as the histograms shown in Figure B-1.

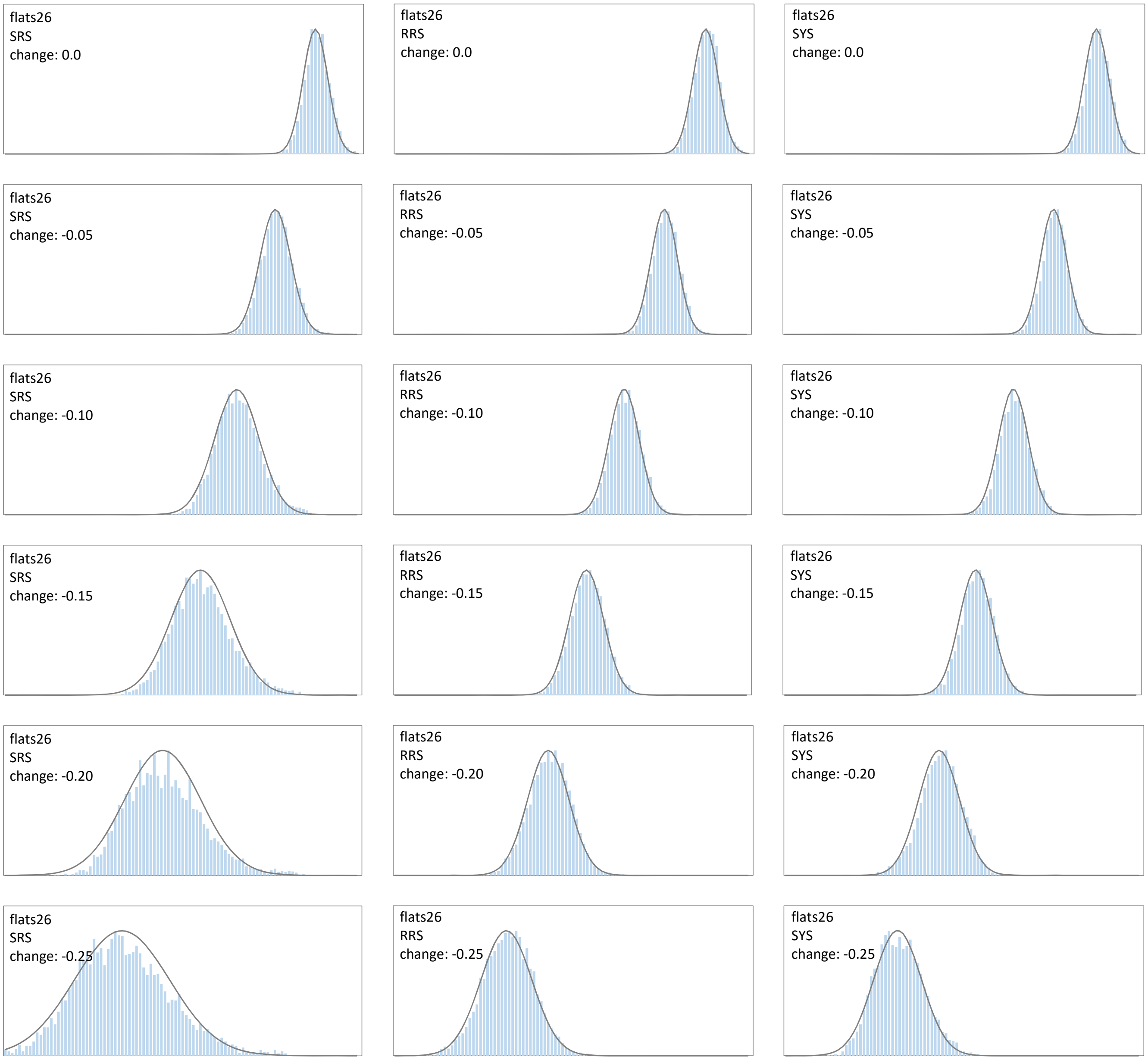


Figure B-3. Histograms of estimates of difference in site area (***n***=5000) based on repeat sampling of n=10 transects from the flats26 site model using SRS (left column), STR (middle column) and SYS (right column) transect selection. Each row represents a different homogeneous nominal change value (indicated in graph labels) applied uniformly across the initial site model to derive the model for the second sampling occasion. The curve on each graph shows the normal distribution with the same mean and standard deviation as the sample difference data. The ***x***-axes for all graphs are identical. The corresponding Q-Q plots to better assess normality are shown with the same layout of scenarios in Figure B-4.

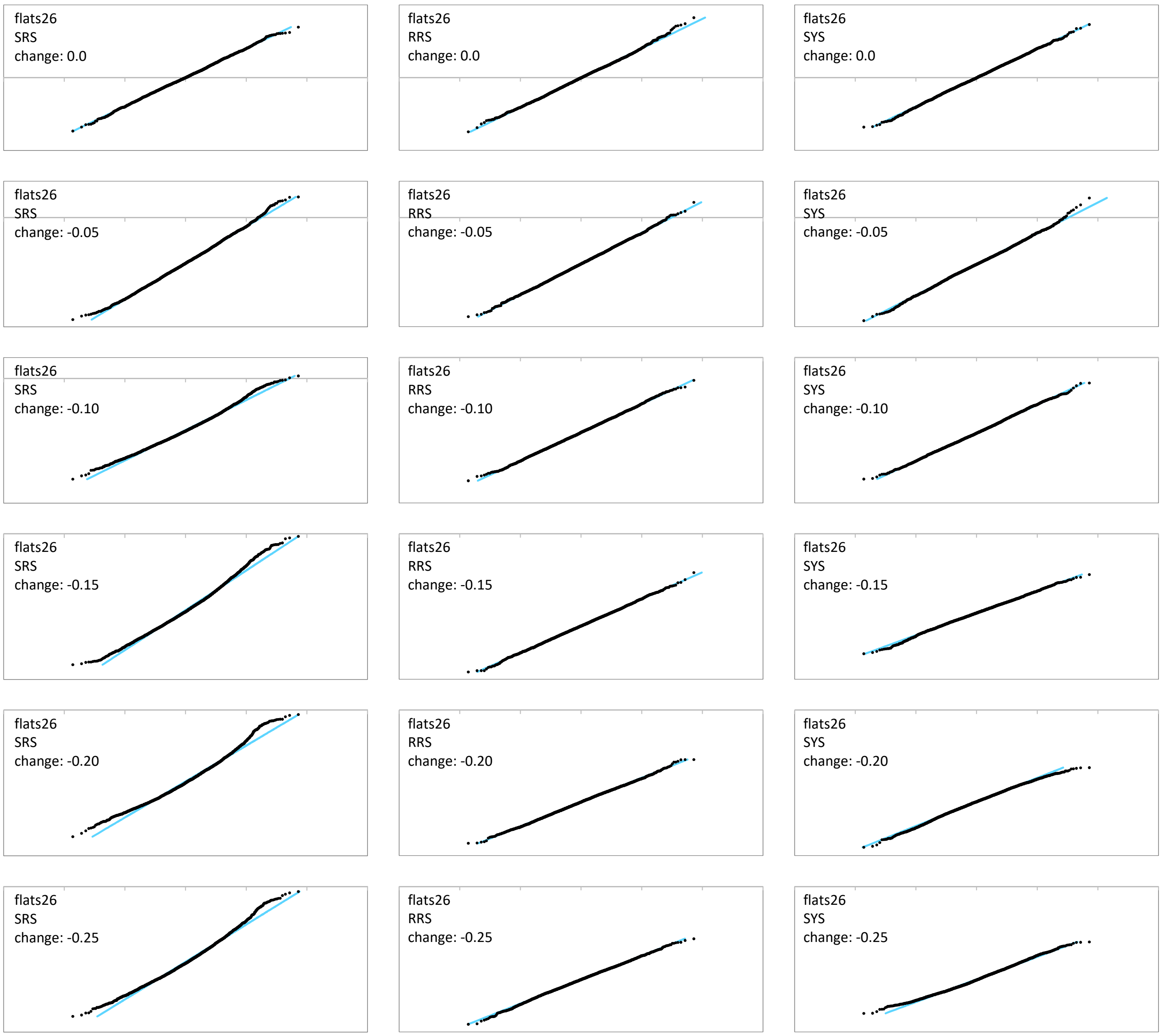


Figure B-4. Normal Q-Q plots to assess normality of the estimates of difference in site area for the flats26 site model. The graphs have the same layout of scenarios as the histograms shown in Figure B-3.

Appendix C Power Diagrams – Difference in Area Estimates (New Draw Transects)

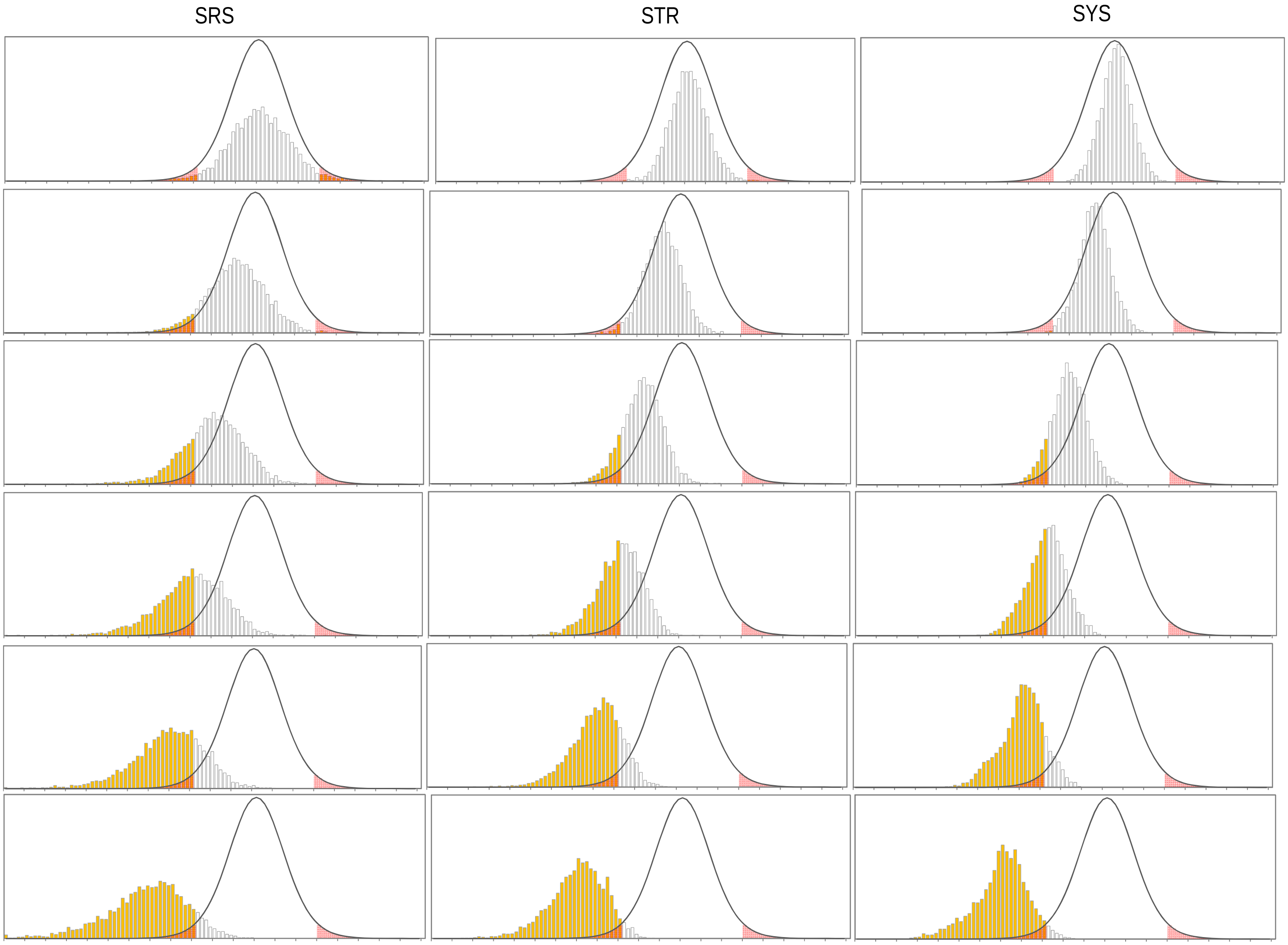


Figure C-1. Power diagrams for tests of significant difference between two area estimates when sampling from the core001 site model with new draw transects and the standard variance estimator. The diagrams show results when sampling with SRS (left column), STR (middle column) and SYS (right column) for scenarios with homogeneous nominal relative change of 0.0, -0.05, -0.10, -0.15, -0.20 and -0.25 (top row to bottom row respectively). The curve is the t-distribution assumed to be true under the null hypothesis when testing for change. The red zones in the tails represents the area associated with t values that would lead to rejection of the null hypothesis – this area represents the Type I error of $\alpha = 0.05$. The histogram is the frequency distribution of the t values obtained from 5000 simulations with the unfilled bars showing cases where the null hypothesis was not rejected. The area of unfilled bars represents Type II error. The gold bars show the frequency of cases where the null hypothesis was rejected. The area of the gold bars represents power.

The t distributions shown are based on d.f. = 15 (they are identical) which was the mean estimate of d.f. of the 5000 simulations based on Equation 3-10 (p.23). The y-axes have the same scale for each histogram.

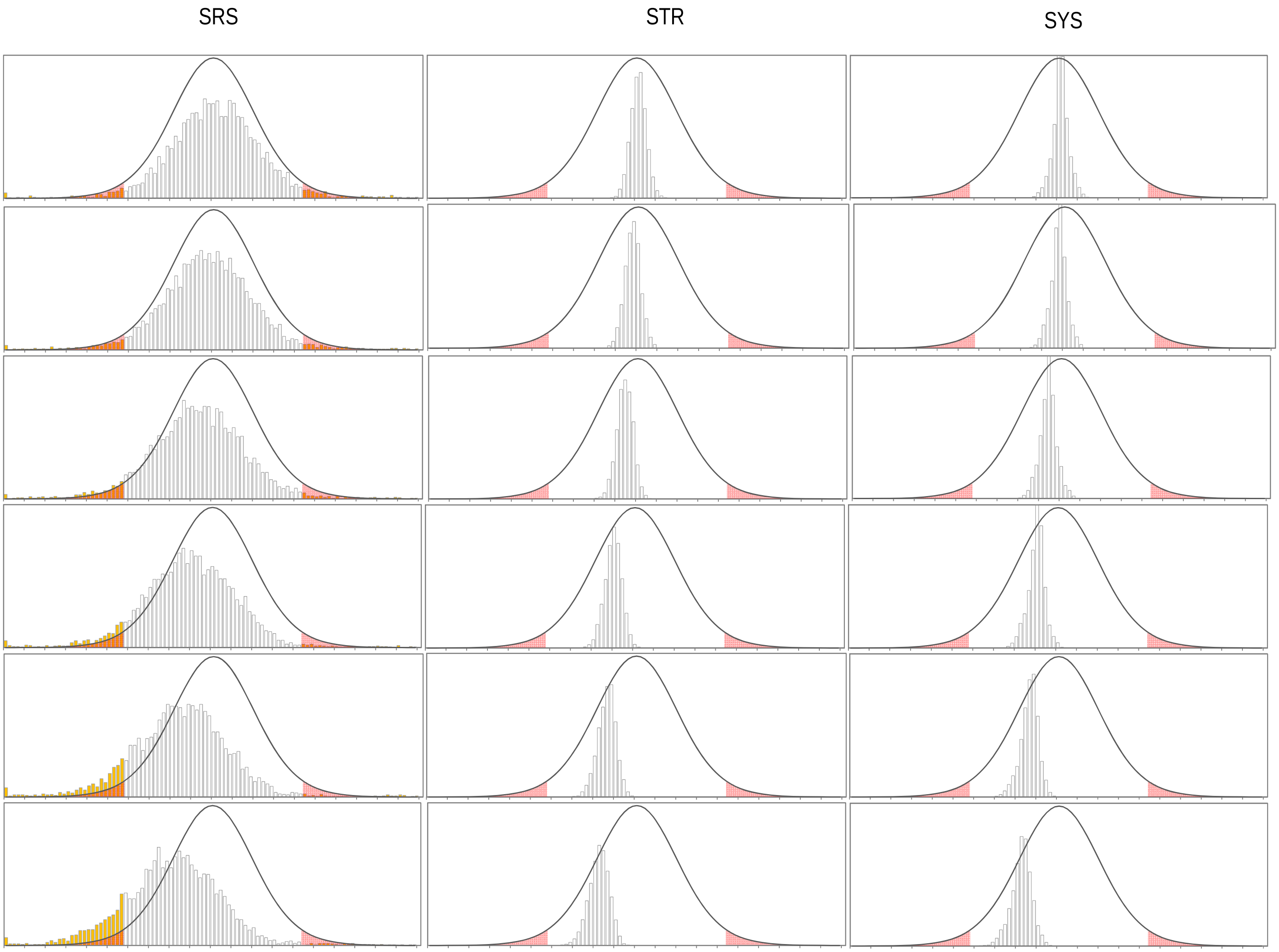


Figure C-2. Power diagrams for tests of significant difference between two area estimates when sampling from the flats26 site model with new draw transects and the standard variance estimator. The diagrams show results when sampling with SRS (left column), STR (middle column) and SYS (right column) for scenarios with homogeneous nominal relative change of 0.0, -0.05, -0.10, -0.15, -0.20 and -0.25 (top row to bottom row respectively). The curve is the t-distribution assumed to be true under the null hypothesis when testing for change. The red zones in the tails represents the area associated with t values that would lead to rejection of the null hypothesis – this area represents the Type I error of $\alpha = 0.05$ if there is truly no change between sampling occasions and the actual sampling distribution adheres to the t distribution. The histogram is the frequency distribution of the t values obtained from 5000 simulations with the unfilled bars showing cases where the null hypothesis was not rejected. The area of unfilled bars represents Type II error. The gold bars show the frequency of cases where the null hypothesis was rejected. The area of the gold bars represents power.

The t distributions shown are based on d.f. = 15 (they are identical) which was the mean estimate of d.f. of the 5000 simulations based on Equation 3-10 (p.23). The y-axes for the SRS graphs are contracted relative to the STR and SYS graphs in order to expand the histogram.

Appendix D Power Diagrams – Estimated Difference in Area (Repeat Transects)

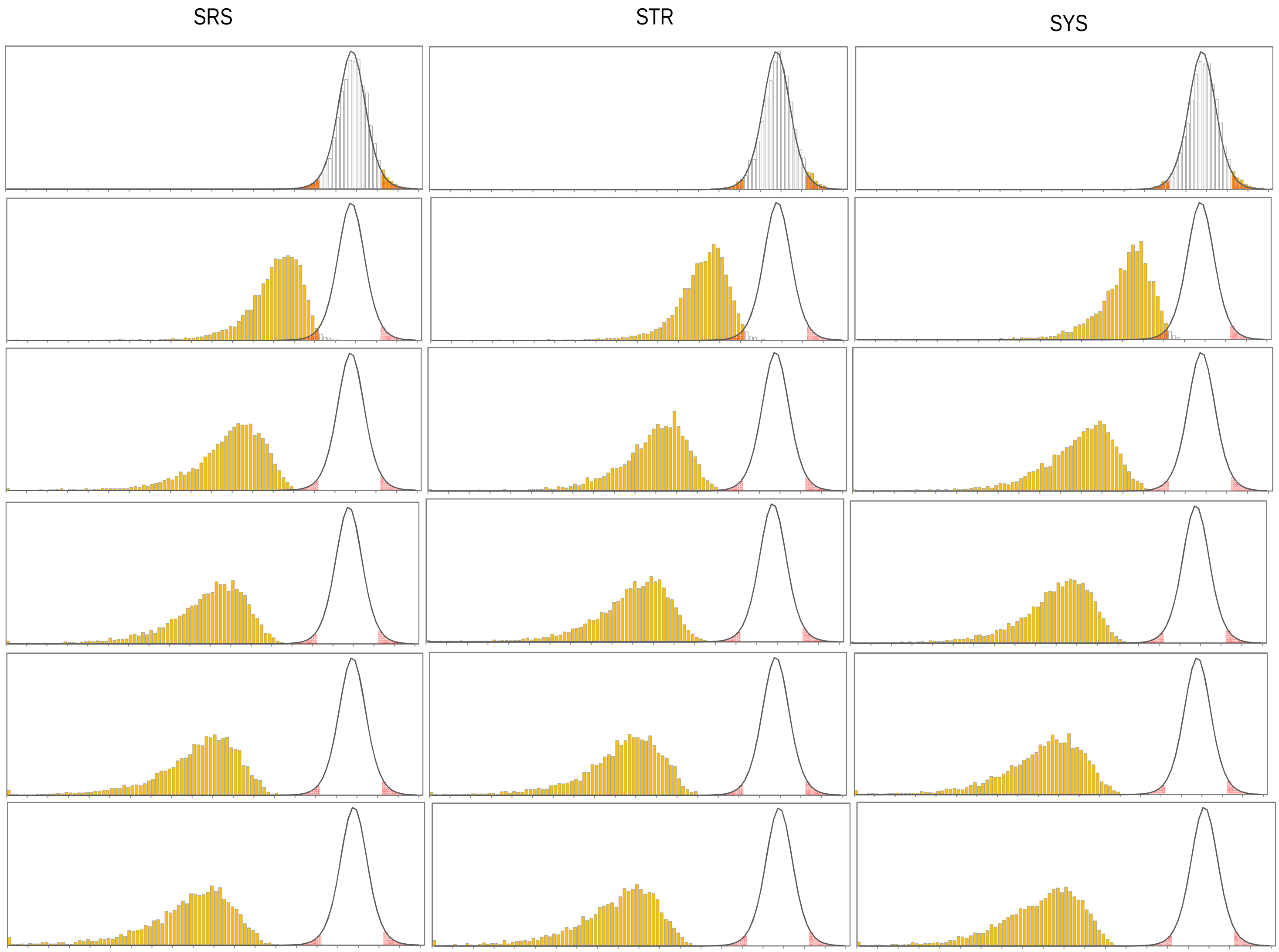


Figure D-1. Power diagrams for tests of significant difference between two area estimates when sampling from the core001 site model with repeat transects and the standard variance estimator. The diagrams show results when sampling with SRS (left column), STR (middle column) and SYS (right column) for scenarios with homogeneous nominal relative change of 0.0, -0.05, -0.10, -0.15, -0.20 and -0.25 (top row to bottom row respectively). The curve is the t-distribution assumed to be true under the null hypothesis when testing for change. The red zones in the tails represents the area associated with t values that would lead to rejection of the null hypothesis – this area represents the Type I error of $\alpha = 0.05$. The histogram is the frequency distribution of the t values obtained from 5000 simulations with the unfilled bars showing cases where the null hypothesis was not rejected. The area of unfilled bars represents Type II error. The gold bars show the frequency of cases where the null hypothesis was rejected. The area of the gold bars represents power.

The t distributions shown are identical and based on d.f. = 9. The y-axes have the same scale for each histogram.

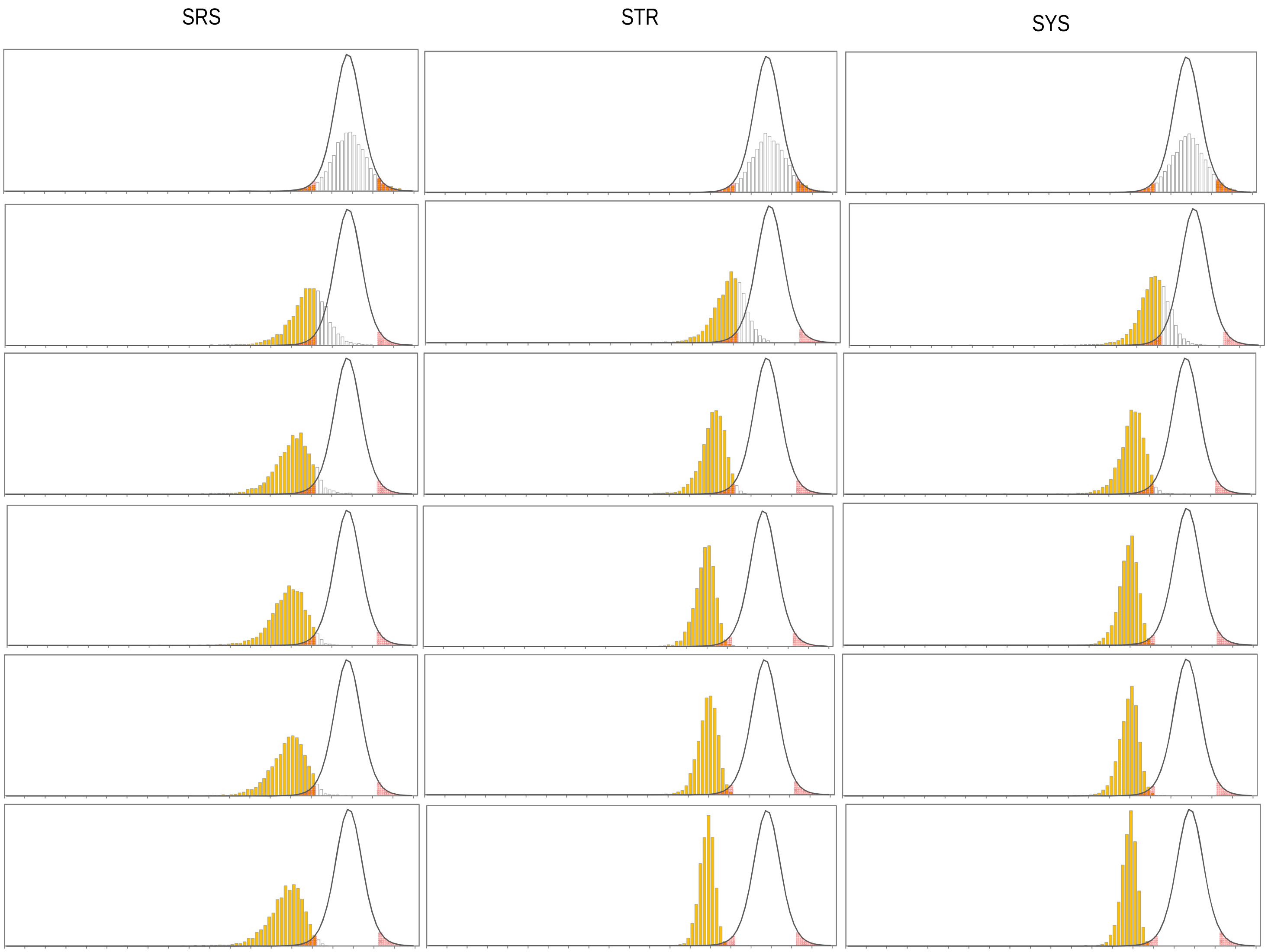


Figure D-2. Power diagrams for tests of significant difference between two area estimates when sampling from the flats26 site model with repeat transects and the standard variance estimator. The diagrams show results when sampling with SRS (left column), STR (middle column) and SYS (right column) for scenarios with homogeneous nominal relative change of 0.0, -0.05, -0.10, -0.15, -0.20 and -0.25 (top row to bottom row respectively). The curve is the t-distribution assumed to be true under the null hypothesis when testing for change. The red zones in the tails represents the area associated with t values that would lead to rejection of the null hypothesis – this area represents the Type I error of $\alpha = 0.05$. The histogram is the frequency distribution of the t values obtained from 5000 simulations with the unfilled bars showing cases where the null hypothesis was not rejected. The area of unfilled bars represents Type II error. The gold bars show the frequency of cases where the null hypothesis was rejected. The area of the gold bars represents power.

The t distributions shown are identical and based on d.f. = 9. The y-axes have the same scale for each histogram

Appendix E Power Diagrams – Estimated Difference in Area (Repeat Transects with v8)

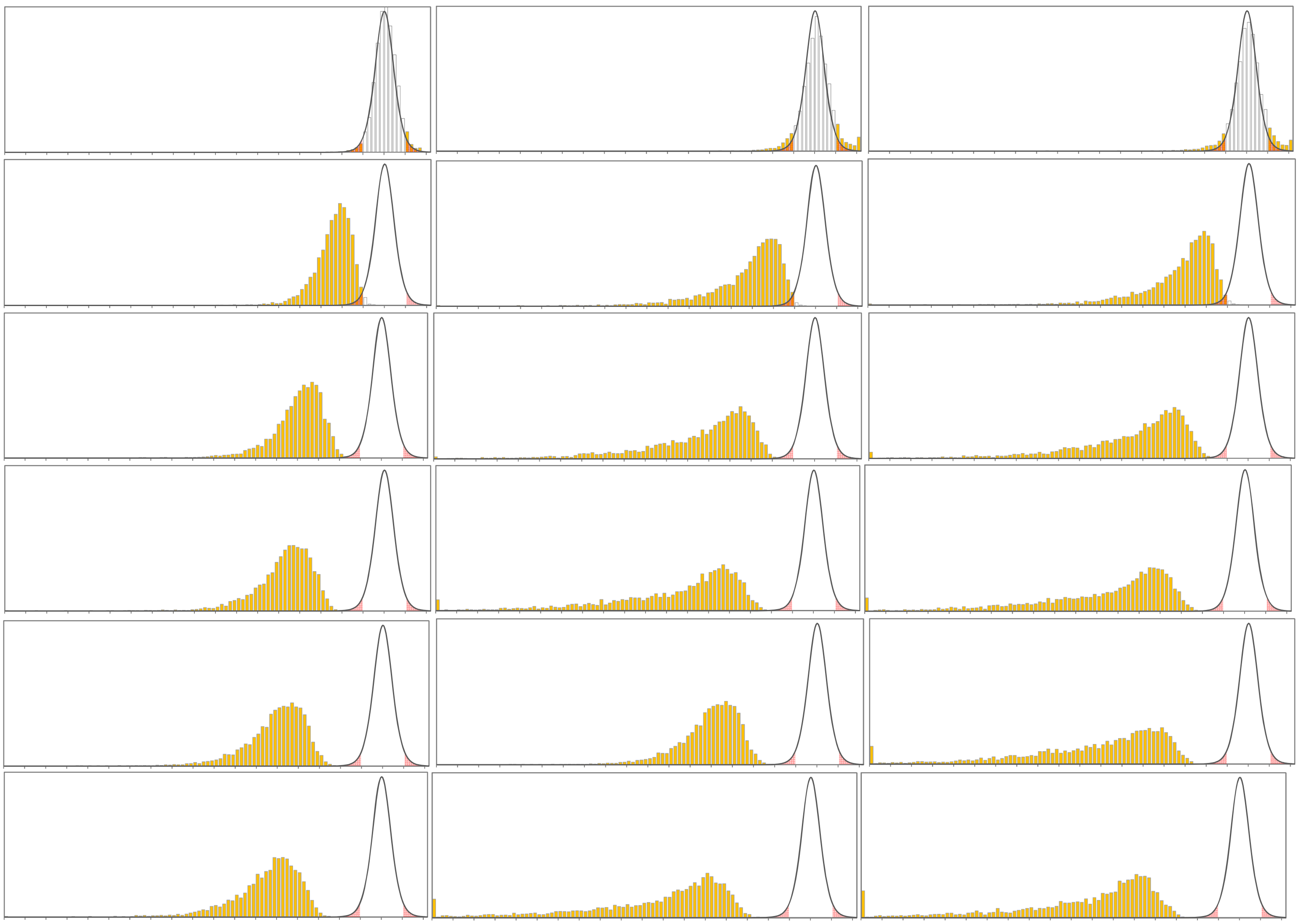


Figure E-1. Power diagrams for tests of significant difference in area when sampling from the core001 site model with repeat transects (paired) and the v8 variance estimator. The diagrams show results when sampling with SRS (left column), STR (middle column) and SYS (right column) for scenarios with nominal relative change of 0.0, -0.05, -0.10, -0.15, -0.20 and -0.25 (top row to bottom row respectively). The curve is the t-distribution assumed to be true under the null hypothesis when testing for change. The red zones in the tails represents the area associated with t values that would lead to rejection of the null hypothesis – this area represents the Type I error of $\alpha = 0.05$. The histogram is the frequency distribution of the t values obtained from 5000 simulations with the unfilled bars showing cases where the null hypothesis was not rejected. The area of unfilled bars represents Type II error. The gold bars show the frequency of cases where the null hypothesis was rejected. The area of the gold bars represents power.

The t distributions shown are based on d.f. = 9. The range of the y-axes is the same for each graph.

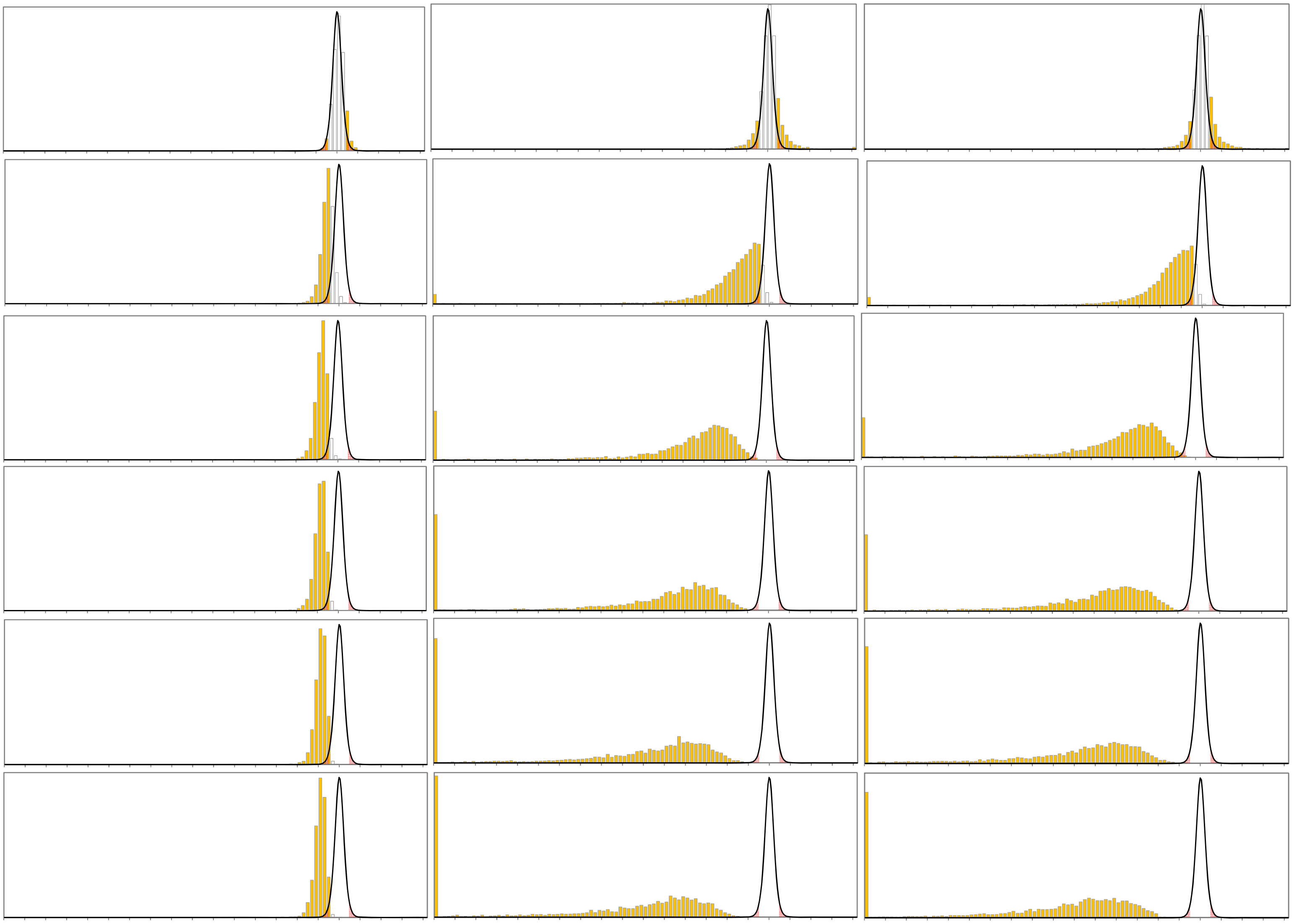


Figure E-2. Power diagrams for tests of significant difference in area when sampling from the flats26 site model with repeat transects (paired) and the v8 variance estimator. The diagrams show results when sampling with SRS (left column), STR (middle column) and SYS (right column) for scenarios with nominal relative change of 0.0, -0.05, -0.10, -0.15, -0.20 and -0.25 (top row to bottom row respectively). The curve is the ***t***-distribution assumed to be true under the null hypothesis when testing for change. The red zones in the tails represents the area associated with ***t*** values that would lead to rejection of the null hypothesis – this area represents the Type I error of $\alpha = 0.05$. The histogram is the frequency distribution of the ***t*** values obtained from 5000 simulations with the unfilled bars showing cases where the null hypothesis was not rejected. The area of unfilled bars represents Type II error. The gold bars show the frequency of cases where the null hypothesis was rejected. The area of the gold bars represents power.

The ***t*** distributions shown are based on d.f. = 9. The ***y***-axes of the SRS histograms span a larger range (0 – 1600 frequency of ***t*** estimates) than the STR and SYS histograms (0 – 1200 frequency).