NATURAL RESOURCES

Evaluating Sampling Designs For a Regional Seagrass Monitoring Program

March 2023

PSEMP
PUGET SOUND ECOSYSTEM
MONITORING PROGRAM

WASHINGTON STATE DEPT OF
NATURAL
RESOURCES

Evaluating Sampling Designs For a Regional Seagrass Monitoring Program

---

March 2023

Pete Dowty
Nearshore Habitat Program
Aquatic Resources Division

PUGET SOUND ECOSYSTEM
MONITORING PROGRAM

WASHINGTON STATE DEPT OF
NATURAL
RESOURCES

Preface

The completion of this report in March 2023 comes eight years after the bulk of the work was completed. At that time, a partial draft of this report had been internally reviewed and the last remaining task was to do a final revision. Then I suddenly took an extended leave of absence. When I finally returned to work my focus was on other projects and eventually, I completely forgot that this report was still unfinished. Bart brought that to my attention sometime around 2020 and now, in 2023, it is finally complete.

In some places, I have made updates to language during the final revision that reflect more recent sampling design changes made in the seagrass monitoring program. In other places, the original perspective remains where focus was on the initial design still in place eight years before, as well as a few design alternatives any of which could potentially have been implemented in the future.

March 1, 2023

Acknowledgements

The Nearshore Habitat Program is part of the Aquatic Resources Division in the Washington Department of Natural Resources (DNR), the steward for state-owned aquatic lands. Program funding is provided through the Aquatic Lands Enhancement Act. The Nearshore Habitat Program monitors and evaluates the status and trends of marine vegetation for DNR in association with the Puget Sound Partnership as one component of the Puget Sound Ecosystem Monitoring Program (PSEMP).

This work benefited from many discussions with Helen Berry, Bart Christiaen, Jeff Gaeckle and Lisa Ferrier in the Nearshore Habitat Program while this work was being conducted. This report benefited from reviews by Bart, Jeff and Lindsay Anderson.

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

Copies of this report may be obtained from:
https://www.dnr.wa.gov/programs-and-services/aquatics/aquatic-science/nearshore-habitat-program

Contents

Executive Summary

Since 2000, the Washington State Department of Natural Resources (DNR) has conducted annual monitoring of the native seagrass population in greater Puget Sound – a population dominated by eelgrass (*Zostera marina*). Eelgrass is useful as an ecosystem indicator species and there are currently two eelgrass indicators included in the Puget Sound Partnership's Vital Signs based on DNR seagrass monitoring data and analysis.

DNR's Submerged Vegetation Monitoring Program (SVMP) conducts the monitoring with the goal of characterizing the entire native seagrass population in greater Puget Sound with estimates of total native seagrass area, annual change and long-term trend. An in situ (as opposed to remote sensing) method is required to isolate native seagrass from other species and to collect data across the species' range (subtidal and intertidal). The SVMP selected the use of boat-towed underwater video for data collection which is used within a sampling design so that the entire population can be characterized from a sample covering only a fraction of the total study area.

The sampling design is critical. A non-optimal design may give unreliable results, or waste effort in the sense that greater sampling effort is needed to achieve the same detection capability of an optimal design. Given the substantial resources required for regional-scale monitoring, it is prudent to assess the performance of the sampling design to guide any adjustments that may be needed. That is the purpose of the work summarized in this report.

The original SVMP sampling design was largely delivered in response to an RFP, although there were many refinements made in collaboration with DNR staff. A key element of the original design was partial sample replacement whereby each year 20% of the sites surveyed the previous year rotate out of the sample to be replaced by new randomly selected sites.

The role of the 20% site rotation was not well understood by DNR staff (e.g., why not 60%?), but early results led to concern that the rotation may introduce spurious change in the data record. In 2010, analysis pointed to an increasing long-term trend in soundwide seagrass area, but given the earlier concerns, this was not considered reliable. Two statistical consultants were retained for independent review of the issue. The reports from these reviews corroborated the concerns but did not suggest a path forward. It was clear that a focused effort would be required to resolve this issue and that was the key motivation that led to this report.

The specific objectives of this study include:

- Quantitatively assess the role and effects of the 20% site rotation.

- Summarize the key findings of the two reviews by independent statisticians.
- Conduct modeling experiments to evaluate the performance of the initial SVMP sampling design and compare to the performance of four alternative designs. Performance was evaluated in terms of the ability of each design to detect changes in the seagrass population ('power' to detect change), as well as the precision and accuracy of population estimates.

The modeling approach taken here utilized a computer application specifically developed in C for this purpose. The details of this application are presented in a companion report (Dowty 2017). This report presents results from the use of that computer application to address the specific objectives above.

The key findings include:

- **Adjusted Estimates**. The impetus for site rotation was to enable what is termed 'adjusted estimates' that improve precision of soundwide seagrass area estimates.
- **Balancing Area and Annual Change**. Site rotation also provided a balancing of the precision of seagrass area estimates against the precision of annual change estimates. The 20% rate of rotation was found to lie within a broad range that provided an optimal balancing. Note that this optimization only considers total seagrass area and annual change estimates. Trend estimates are not considered.
- **Trend Result Confirmed Unreliable**. A modeling experiment corroborated the non-significance of the 2010 trend result. The trend analysis from the SVMP design was demonstrated to be unreliable and was subsequently discontinued.
- **Site Rotation Degrades Trend Estimates**. Site rotation was shown to be detrimental to the precision of trend estimates. The 20% rate of rotation was shown to be particularly detrimental.
- **Comparison of 5 Designs**. A modeling experiment that compared 5 different designs under increasingly severe scenarios of seagrass loss showed that two designs were far superior to the others for trend detection:
  high-performance designs: '3 rotating panels', 'fixed sites'
  low-performance designs: '20% rotation', 'new draw', 'new draw (3 yrs pooled)'
- **Further comparison of 2 Best Designs**. The two high-performance designs were subject to an additional modeling experiment with a sequence of change scenarios with the same total seagrass loss but distributed in the population in a way that is increasingly difficult to detect. The '3 rotating panels' design was found to have superior performance.

Additional modeling studies were identified for future work to advance the understanding of the sampling design performance and inform future adjustments to the SVMP design. The modeling work presented here was conducted in 2014-2015, but completion of this report was delayed until 2023. The findings of this work directly led to a fundamental change in the SVMP sampling design from '20% rotation' to '3 rotating panels'. This change was implemented in 2015.

1 Introduction

1.1 The Submerged Vegetation Monitoring Program (SVMP)

The Washington State Department of Natural Resources (DNR) is steward of 2.6 million acres of state-owned aquatic land. The Aquatic Resources Division and Districts of DNR manage these aquatic lands for the benefit of current and future citizens of Washington State. DNR's responsibilities include the stewardship of native seagrasses, an important marine nearshore habitat in greater Puget Sound. As part of that responsibility, DNR conducts annual monitoring of the native seagrass population – a population dominated by eelgrass (*Zostera marina*). Eelgrass is both ecologically valuable and sensitive to water quality degradation thereby making it an ideal ecosystem indicator species. The annual monitoring has been conducted by DNR's Submerged Vegetation Monitoring Program (SVMP) since 2000. A portion of the SVMP effort is dedicated to more localized surveys, but the majority of effort (~60%) has been dedicated to the "soundwide" study that is our focus here. The study area is the Washington State portion of the Salish Sea, also known as greater Puget Sound, or within the EPA National Estuary Program, just Puget Sound. A small area of southern inlets that are most removed from the Pacific Ocean are excluded from the study area due to only rare occurrence of seagrass (see Figure 1-1, p.8).

Since its inception, the goals of the SVMP have included the generation of the following "soundwide" estimates:

- annual estimates of the total native seagrass area in greater Puget Sound,
- estimates of year-to-year change in total native seagrass area,
- estimates of multi-year trend in total native seagrass area with the ability to detect a 20% change over ten years.

These goals place important constraints on the data collection methods employed for monitoring. First, since the target population encompasses all native seagrass in greater Puget Sound, the methods must be effective in both intertidal and subtidal areas since native seagrass spans both. Second, since the target population is limited to native seagrass, the methods must be capable of distinguishing native seagrass from non-native seagrass, in addition to distinguishing seagrass from macroalgae.

These constraints preclude existing remote sensing approaches and require an in-situ data collection for both species discrimination and to reach subtidal areas inaccessible with remote sensing methods. The SVMP selected the approach of collecting boat-towed underwater video at depth just above the seagrass canopy or substrate surface. This video is later classified for the presence of native seagrass and, separately, non-native seagrass.

While the use of towed underwater video satisfies the depth range and species discrimination requirements, it is a much more intensive data collection method in terms of required effort per unit of nearshore area as compared to remote sensing approaches. Consequently, it is not feasible to survey the entire nearshore area of greater Puget Sound (~2400 miles of shoreline, or ~3800 km) with underwater video due to the magnitude of effort that would be required. The solution is to employ sampling so that in-situ data collection is only required for a fraction of the total nearshore area. Any regional monitoring program of this scale built on in-situ[1] data collection will utilize sampling.

1.2 Sampling Design

Sampling is crucial for the SVMP as it makes an intractable problem tractable, but there is a cost. It introduces the need to dedicate a substantial effort to sampling design and maintenance of the design to ensure that derived results are reliable and responsive to the monitoring goals. More specifically, the sampling design serves to minimize bias and maximize precision in estimates derived from the collected data. The design also aims to maximize the power of statistical tests to detect some targeted pattern. A sampling design that performs poorly will lead to low quality results that may lead to erroneous conclusions. A poorly performing design may also be wasteful in that greater sampling effort is needed to achieve the same detection capability of a more optimal design.

We consider ‘sampling design’ to encompass five elements (Table 1-1). This study is not a comprehensive investigation of all SVMP sampling design elements. Rather, we are particularly interested in how sample selection is handled across time (#3) and the associated estimation methods (#4).

A wide variety of sampling designs are possible and different designs may be optimal to measure different parameters of the population. A monitoring program that is tasked with estimating multiple parameters may have to weigh trade-offs in sampling design.

---

[1] It has been argued that towed underwater videography, later classified for vegetation, should be categorized as ‘remote sensing’ as the human researcher is remote from the population individuals (either a few meters distant on the boat, or up to 100s of km distant reviewing video in the office at a later time). Also, the central role of imaging is shared between the underwater video and traditional remote sensing from air or space platforms. But here we consider ‘in-situ’ to be the better categorization because the instrument collecting data (camera) is in close proximity to the population individuals and allows for derived data that would typically be associated with in-situ methods (e.g. species-specific presence and sub-meter cover) and inaccessible to traditional remote sensing methods. Underwater video is akin to diver-collected photos of quadrats to be used later to extract quantitative data. We would argue that is also an in-situ data collection method.

Table 1-1. Elements of a sampling design. This report is primarily focused on elements #3 and #4 and how they affect the performance of the sampling design.

1. Field data collection protocols.
2. Specification of sampling frame – i.e. the delineation of individual sample units (sites) that are mutually exclusive but comprehensively cover the population to be sampled (all potential native seagrass habitat in the study area).
3. Procedures for selection of spatial locations (sites) across temporal sampling occasions for data collection
4. Procedures for making specific quantitative estimates using equations known as ‘estimators’ that have precision and accuracy attributes within a given sampling design.
5. Procedures for conducting statistical tests using the estimates to address specific hypotheses about the population.

After a data record of some length has been compiled under a given sampling design, it is prudent to evaluate the performance of the design. The purpose of the evaluation is to test whether the precision and accuracy of the estimators, and the errors in statistical testing, are reliable and meet the design criteria. This will depend on whether the assumptions inherent in the design hold or turn out to be poor assumptions in a particular monitoring setting. In the latter case, the evaluation of performance may lead to adjustments of the sampling design.

A sampling design may also need to be adjusted as priorities shift with respect to multiple quantitative monitoring goals. For example, as the data record begins to successfully address some goals, we might prioritize other unaddressed goals, or our evolving understanding of the population and its changing environments might elevate different quantitative monitoring goals over time. Any design changes must be carefully considered and only implemented after adequate study to ensure the long-term integrity of the data record.

1.3 Objectives and Modeling Approach

The purpose of the work summarized in this report was to evaluate the performance of the initial SVMP sampling design and compare this design to selected alternatives. The alternatives vary in the pattern of sample replacement or retention across sampling occasions. Performance was evaluated in terms of the power of a trend test to detect trends and in terms of the bias and precision in trend estimates.

The task of evaluating the performance of a regional sampling design is non-trivial. In essence, a substantial apparatus for testing the design must be constructed. Here, this apparatus is a software application written in C. This application does the following: (1) simulates the entire native seagrass population currently in greater Puget Sound based on existing monitoring data (i.e., generates a seagrass population model); (2) applies different hypothetical 20-year change scenarios to the population; (3) conducts repeated sampling of the population for a given sampling design over the 20-year period in a Monte Carlo framework; (4) summarizes the Monte Carlo results to generate measures of sampling design performance. This application has been previously described (Dowty 2017). Here we use this application in a study to evaluate sampling design performance.

The specific objectives of this study include:

- Analyze the design considerations that led to the initial SVMP sampling design.
- Present monitoring results that led to concerns about the performance of the initial sampling design.
- Summarize the key findings of two reports that were commissioned from professional statisticians to investigate these concerns.
- Develop hypothetical models of 20-year change in native seagrass to serve as the basis for sampling design evaluation.
- Conduct modeling experiments to compare the performance of the initial SVMP sampling design with four alternative designs with respect to the detection of the multi-year trends in the model change scenarios.

1.4 The Initial SVMP Sampling Design

It is helpful to have a grasp of key elements of what we term the initial SVMP sampling design. The fundamental components of the design are due to an initial effort led by Jim Norris (Marine Resources Consultants, Port Townsend), Sandy Wyllie-Echeverria (University of Washington) and John Skalski (University of Washington) with further development done collaboratively with DNR staff.

DNR staff have made several adjustments to the design since monitoring was initiated in 2000. Most notably, adjustments were made after the 2000 field season and after the 2003 field season. These adjustments are mentioned in section 2.2, but the modeling work conducted for this study ignores the effects of those adjustments and treats the design that was operational in the 2004-2014 period to be the ‘initial’ design that we apply to the entire analysis period. Greater detail on the adjustments is presented elsewhere (Dowty et al. 2022, Dowty 2005).

The basic elements of the design include the use of two separate sampling frames, and the use of two-stage sampling with stratified random selection to sample from the greater Puget Sound study area. Partial sample replacement was used between sampling occasions. These elements are now described in more detail.

1.4.1 The Sampling Frames

The nearshore area within greater Puget Sound was manually divided into two geomorphic categories: flats and fringe (roughly, concave embayments and straight linear shoreline). A different procedure was used within each category to subdivide the area into discrete sites that serve as sample units. In each case, the set of sites is termed a sampling frame. For the flats sampling frame, individual embayments were used as a natural division of the total flats area, except for larger embayments that were subdivided into multiple sites.

The fringe sampling frame was constructed by dividing the -20 ft MLLW bathymetric contour along the shoreline into 1000 m segments, each segment delineating a site. There were frame errors which meant that the sampled area (area associated with all 1000 m segments) did not exactly match the target area (all nearshore area in the fringe geomorphic category). These frame errors occurred because the lengths of contour lines being subdivided were not an exact multiple of 1000 m, leaving small remainder segments. Overall, less than 2% of the fringe nearshore area fell outside any sites delineated by the 1000 m segments.

1.4.2 Stratification

Each sampling frame was stratified which simply divided all the sites in the frame into separate groups. The primary purpose was to group sites that are similar in order to improve overall precision. Here stratification was also used as a means to set aside a small number of sites for more intensive survey effort.

The flats frame (n = 74 sites) was stratified into core (n = 4), persistent flats (n = 3) and rotational flats (n = 67) strata. The rotational flats stratum is the only stratum that is sampled and is also just referred to as the ‘flats stratum’. The fringe frame (n = 2393) was stratified into core (n = 2), narrow fringe (n = 1965) and wide fringe (n = 426) strata. While the number of sites in the fringe sampling frame is much greater than in the flats sampling frame, in terms of nearshore area covered the two frames are almost identical (Figure 1-1).

1.4.3 The Two Stages of Sampling

The SVMP sampling design involves two stages of sampling. The first stage of sampling occurs when sites within a stratum are selected by simple random selection (SRS). The selected sites are a sample of the stratum. The rotational flats, narrow fringe and wide fringe strata are subject to this first stage of sampling and the random site selection is conducted independently in each stratum. The core stratum (pooled flats and fringe core strata) and the persistent flats stratum are not subject to this first stage of sampling. Rather, all sites within these strata are surveyed, as in a census.

The second stage of sampling applies to all strata. For each site to be surveyed, a set of transects are randomly selected to represent the site. These transects are a sample of the site. The transects selected for a given sampling occasion are surveyed with underwater videography. In the initial SVMP design, transects were selected by SRS by selecting random points along a line following shoreline for placing transects perpendicular to the shoreline.

This study is exclusively focused on the first stage of sampling (sampling of strata with selected sites). Performance of the second stage of sampling (sampling of sites with selected transects) has been discussed elsewhere (Dowty et al. 2017).

1.4.4 Sample Replacement/Retention Across Sampling Occasions

SVMP monitoring involves data collection over time. The issue of site selection for stage 1 sampling must be addressed at each sampling occasion. After the first occasion, there are three main possibilities for sample selection:

1. Make a new random draw of sites for surveying (replace previous samples with a newly selected sample of sites).
2. Re-survey a previous random selection of sites (retain a previous sample for ‘repeat’ surveying, also termed a ‘fixed sites’ approach).

3. Partial replacement of a previous sample with newly selected sites (also referred to here as sample rotation). The proportion of a sample to be replaced is specified by the design.

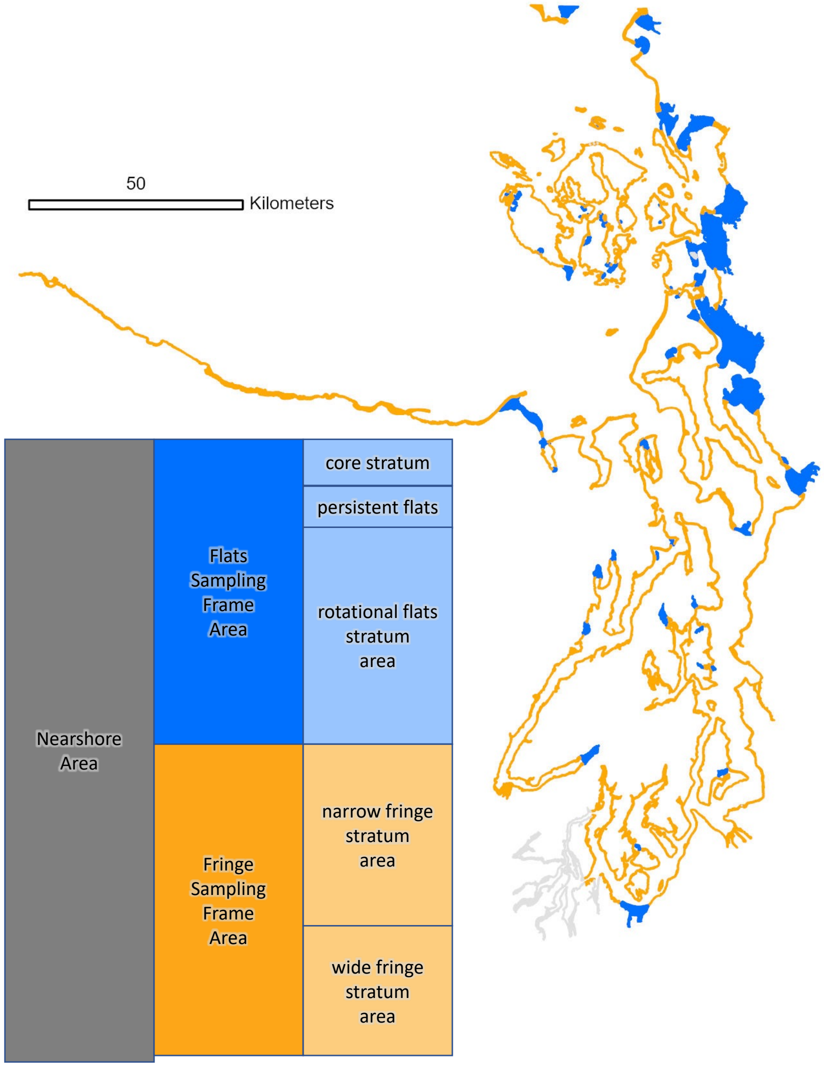


Figure 1-1. Breakdown of all nearshore area (92,710 ha) by the area covered by the flats and fringe sampling frames (map) and the stratification of those frames (diagram). The nearshore area includes all potential seagrass habitat in greater Puget Sound. Note that the flats and fringe frames do not completely cover all nearshore area (noticeable as a lack of alignment along the bottom edge of the diagram). This reflects the frame errors of the fringe sampling frame. The core stratum of the fringe frame covers an area that is too small to be visible. The gray shoreline seen in the map in the vicinity of the southernmost extent of the inland marine waters, is outside of the study area.

The initial SVMP design followed #3 above and employed 20% sample replacement between annual sampling occasions. This was done in a way so that once sites are selected for surveying, they would be surveyed on five consecutive occasions before rotating out of the sample. This is shown visually in Figure 1-2 together with the “new draw” (#1 above) and “fixed sites” (#2 above) options.

This aspect of the sampling design, how sites, or panels of sites, are scheduled for data collection over time, is sometimes referred to as the ‘revisit plan’ (McDonald 2003) and is the focus of this study. The motivating question was how the revisit plan affects the performance of the overall sampling design and how that aligns with the SVMP monitoring goals.

		Sampling Occasions																			
revisit plan	**panel**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
20% rotation	1	x																			
	2	x	x																		
	3	x	x	x																	
	4	x	x	x	x																
	5	x	x	x	x	x															
	6		x	x	x	x	x														
	7			x	x	x	x	x													
	8				x	x	x	x	x												
	9					x	x	x	x	x											
	10						x	x	x	x	x										
	11							x	x	x	x	x									
	12								x	x	x	x	x								
	13									x	x	x	x	x							
	14										x	x	x	x	x						
	15											x	x	x	x	x					
	16												x	x	x	x	x				
	17													x	x	x	x	x			
	18														x	x	x	x	x		
	19															x	x	x	x	x	
	20																x	x	x	x	x
	21																	x	x	x	x
	22																		x	x	x
	23																			x	x
	24																				x

fixed sites	1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

new draw	1	x																			
	2		x																		
	3			x																	
	4				x																
	5					x															
	6						x														
	7							x													
	8								x												
	9									x											
	10										x										
	11											x									
	12												x								
	13													x							
	14														x						
	15															x					
	16																x				
	17																	x			
	18																		x		
	19																			x	
	20																				x

Figure 1-2. Three alternatives for sample replacement/retention across sampling occasions. TOP: 20% sample replacement as implemented in the initial SVMP design. A set of sites are treated as a 'panel'. Note that for each occasion, five panels are indicated for surveying. For example, the sample size for sampling in the narrow fringe stratum is 45 sites. Each panel contains 9 sites (20% of 45 sites) and for each occasion, 9 sites (1 panel) rotate out of the sample and 9 new sites are randomly selected. MIDDLE: For the fixed sites case, there is only one panel that is re-surveyed on each occasion. BOTTOM: each panel of sites is only surveyed on one occasion and then replaced.

2 SVMP Implementation of 20% Site Rotation

This chapter discusses the role of site rotation as implemented in the initial SVMP sampling design. It also presents early monitoring results that led to concern about the effects of rotation and ultimately led to exploration of alternative designs presented in this report.

2.1 The Rationale for Site Rotation

The original SVMP sampling design stipulated "…rotational sampling will be conducted at strata […] where probabilistic sampling occurs" (Skalski 2003, p.9). Here, rotational sampling refers to partial replacement of the sample of sites between sampling occasions. This is only relevant to the strata that are sampled and not strata that are censused such as the core stratum. Skalski (2003) was responsible for the introduction of rotational sampling in the SVMP design and for setting the rate of rotation at 20% per sampling occasion.

DNR staff were aware that the rationale for site rotation involved a balancing of the precision of annual estimates of total seagrass area and the precision of annual estimates of change in seagrass area relative to the previous year. But the nature of the trade-offs underlying this balancing were not understood. For example, there was no understanding of why 20% was selected as the rate of site rotation as opposed to any other value greater than 0% and less than 100%. Consequently, the first objective of this work was to analyze the rationale for rotation (section 1.3, p.5).

Under rotational sampling, annual estimates of soundwide native seagrass area can be adjusted. Skalski (2003) describes a 'retrospective adjustment' that uses sample data from the latest occasion to adjust the estimate for the previous year with an anticipated improvement in precision. This improvement is only possible with partial sample replacement. This technique is unavailable when a sample is retained across occasions (no replacement) or a new sample is drawn for each occasion (total replacement).

2.1.1 Restrospective adjustment

The retrospective adjustment technique is based on the application of double sampling to the repeated sampling of the same population (Cochran 1977, 344pp.). Conventional double sampling is used to estimate a parameter of interest using two separate samples. For example, an estimate of mean tree biomass in a forest stand might be made using a small sample of trees subject to laborious measurements of biomass as well as measurement of an auxiliary variable such as height. A second, much larger sample of trees would be subject only to the relatively easy measurement of tree height. The analysis involves converting the precise estimate of mean tree height into biomass using a relationship developed from the biomass sample. For some scenarios, the loss of precision due to diverting sampling effort from the measurement of biomass is less than the precision gained with the precise average height estimate generated by double sampling.

The retrospective adjustment is a special case of double sampling where the sample from the latest occasion is analogous to the first phase sample (the larger "height" sample) and the matching sites from the previous occasion are analogous to the second phase (the "biomass" sample).

To understand when partial sample replacement and the retrospective adjustment might be beneficial, it is instructive to review Cochran's introduction to sample replacement policy where he considers three alternatives (Cochran 1977, p.345):

1. For estimating change, it is best to retain the same sample throughout all occasions.
2. For estimating the average over all occasions, it is best to draw a new sample on each occasion.
3. For current estimates, equal precision is obtained either by keeping the same sample or by changing it on every occasion. Replacement of part of the sample on each occasion may be better than these alternatives.

It is (3) above that alludes to a rationale for SVMP sample rotation. Note that in Cochran's formulation the sample from the prior year is used to adjust the latest (current) estimate rather than a retrospective adjustment as formulated by Skalski (2003). In other respects, the two formulations are equivalent.

To illustrate when partial replacement can improve the precision of the current estimate (using Cochran's formulation), and the level of improvement to be anticipated, consider a simple example where a sample mean is used as an estimate of a population mean. The sample size, *n*, for the latest occasion can be divided into a number, *m*, of sites matched in the previous sample and a number, *u*, of sites that were just rotated in and are unmatched in the previous sample. The variance of the unadjusted latest sample mean will be S^2/n. The variance of the adjusted mean is expressed as (Cochran 1977, eqn. 12.74, p.347):

$$Var\left(\overline{X'}\right)=\frac{S^2\left(n-u\rho^2\right)}{n^2-u^2\rho^2}$$

Equation 2-1

where

u = the number of replaced (unmatched) sample units,

ρ = the correlation in values of sample units between the two occasions,

$\overline{X'}$ = the adjusted mean.

Note that if u=0 (complete matching) or u=n (no matching), this variance reduces to S^2/n.

To assess the gain in precision associated with the adjusted estimate, the variance of the adjusted estimate can be expressed relative to the variance of the unadjusted sample mean, $\bar{X}$, as

$$\frac{Var\left(\overline{X'}\right)}{Var\left(\overline{X}\right)} = \frac{\dfrac{S^2\left(n-u\rho^2\right)}{n^2-u^2\rho^2}}{S^2/n} = \frac{1-r\rho^2}{1-r^2\rho^2} \qquad \text{Equation 2-2}$$

where r is the fraction of sample rotation and is given by $r = \frac{u}{n}$. Equation 2-2 can be used to assess the gain in precision (reduction in variance) associated with adjustment for different scenarios of correlation and sample replacement (Figure 2-1). Large reductions in variance are possible with adjustment but only when there is high correlation in sample measures (e.g., site eelgrass area) between sample occasions. The optimal rate of rotation, based on the precision of the adjusted estimate of the mean, is always greater than 50% and increases as the correlation increases. For the correlation seen in the SVMP site data, $\rho = 0.99$, the optimal sample rotation from the perspective of the current population estimate is 88% (evaluated with Cochran 1977, Eqn. 12.75, p.347).

While the foregoing discussion demonstrates the value of partial sample replacement for estimating the population mean at a given sampling occasion, it does not consider the estimation of population change between two sampling occasions. Recall that the goal of estimating change calls for retaining the same sample throughout all occasions (p.10). A sampling design with dual goals of estimating both current mean and change must therefore balance the policy of some rate of sample replacement greater than 50% with a policy retaining the same sample across occasions (0% replacement).

Cochran addresses this issue stating that, “If estimates of the change in the population total or mean are of interest, this factor also points toward matching more than half the units from one occasion to the next” (Cochran 1977, p.351). Moreover, Cochran notes that increasing sample retention between occasions has a large benefit to change estimation while having only smaller losses to precision of the current estimates (Cochran 1977, p.354). He summarizes, “The results suggest that retention of 2/3, 3/4, or 4/5 from one occasion to the next may be a good practical policy if current estimates and estimates of change are both important” (Cochran 1977, p.354).

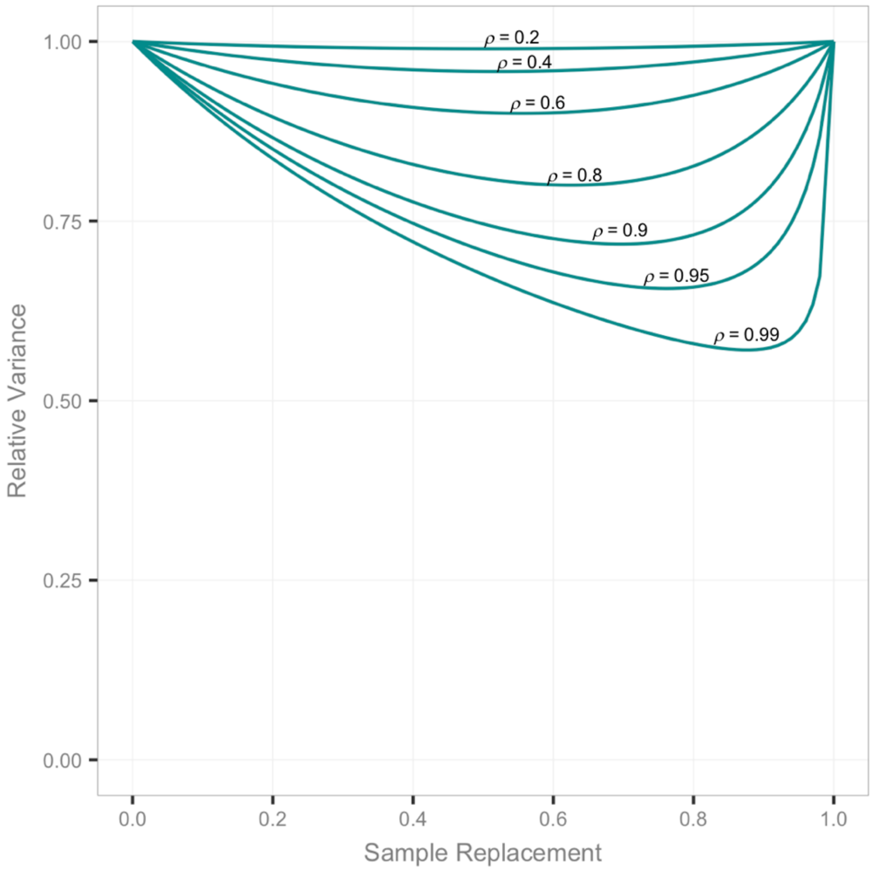


Figure 2-1. The gain in precision associated with adjusting the estimate of the mean as a function of the fraction of sample replaced between occasions. Each curve represents a different level of correlation between sample measures in each of two consecutive occasions. The variance of the adjusted estimate of the mean (which incorporates sample data from a consecutive year) is shown relative to the variance of the unadjusted estimate (i.e., S^2/n). Values less than one indicate a gain in precision with the adjusted estimate.

We now have a range of sample retention (2/3 to 4/5 sample retention or between 33% and 20% sample rotation, respectively) judged to be a good practical policy. The 20% SVMP site rotation implemented in the initial SVMP design is consistent with this guidance and, being on the endpoint of the range with greater matching, appears to weight the change estimate more heavily.

The tradeoff between precision of change and the precision of current estimates can be better understood by visualizing an explicit optimization function. The idea is to find the rate of sample rotation that simultaneously optimizes precision of both change in mean and the current mean estimate, given a specific optimization function. To demonstrate, we specified a simple optimization function and evaluated the function based on Monte Carlo simulations that gave a $\rho = 0.98$ correlation in sample unit measures between the two occasions[2]. A wide variety of optimization functions is possible. Here we use the sum of the variances of change and current estimates expressed relative to the variance obtained with a newly drawn sample each occasion (100% rotation). This function is optimized at the minimum value as a function of sample rotation (Figure 2-2). This particular optimization function for this scenario gives an optimal range of sample rotation between 0 and 70% with little differentiation within this range (Figure 2-2).

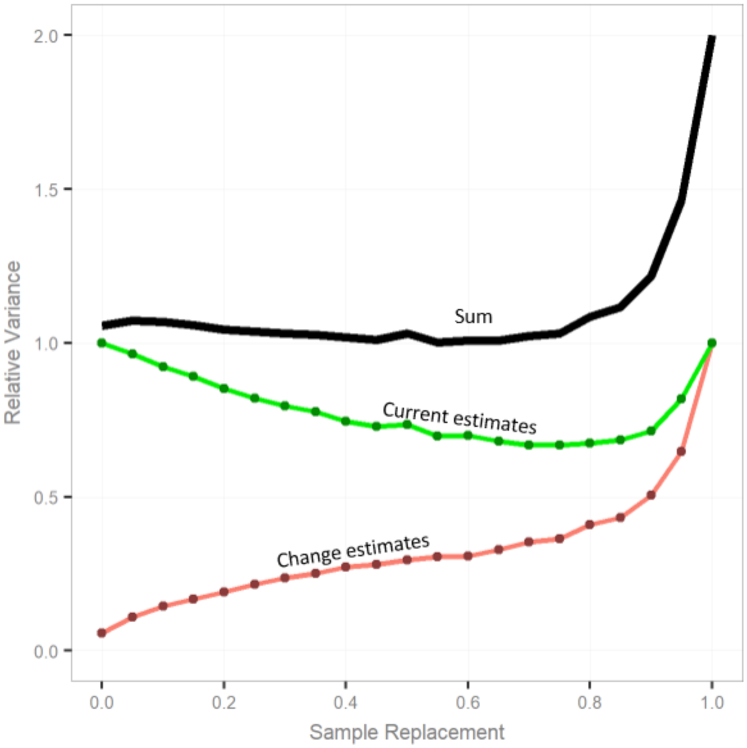


Figure 2-2. The black curve labelled "Sum" is an optimization function for balancing the precision of current estimates (adjusted) and change estimates across different rates of sample replacement between two sampling occasions (*x*-axis). The *y*-axis is variance expressed relative to the variance obtained with a new random draw each sampling occasion. Minimal values of the optimization function (*y* -axis values) indicate optimal values of sample replacement (*x*-axis values) in this example. This optimization function is a simple sum of the relative variances of the current and change estimates which are also shown.

2.1.2 Additional benefits of a rotational design

In addition to optimizing some measure of precision, there are other potential benefits associated with sample rotation. First, continued rotation over many occasions will improve the representation of the overall population – both in its current status and its change characteristics. Unique sub-populations are more likely to be captured in estimates

---

[2] A large (N=100,000) population was generated from a $N(\mu=100, s=10)$ distribution and randomly sampled with sample size n=400 and variances calculated from estimates from 10,000 simulations. Additive change values between sampling occasions were drawn from a $N(5,2)$ distribution and applied to all population units.

with rotation. In the absence of any sample rotation, "Should a new and significant pollution source arise, the success of the monitoring program will depend more upon the chance the point source occurs near one of the earlier established [sites] than upon sampling efforts" (Skalski 1990). Second, rotation spreads the sampling effort over the population. Any effects on sample units from the sampling activity itself will be better controlled with rotation (Rao 1964).

In summary, the 20% SVMP site rotation clearly fits within a paradigm of balancing the precision of current estimates, when adjusted, and change estimates between two occasions. The rationale for sample rotation is primarily tied to improved precision in the status estimate – i.e., the estimate of a population measure at a given sampling occasion. The effect of site rotation on the detection of trend over many sampling occasions was not addressed either in the SVMP framework (Skalski 2003) or in the relevant section of Cochran (1977, Chapter 12, Double Sampling).

It is important to note that the retrospective adjustment was discontinued after 2004 when its SVMP implementation was shown to not be completely reliable (Dowty 2005). Sample rotation, however, was retained through 2014.

2.2 Concern over Effects of Site Rotation

2.2.1 Flats11 Effect

Between 2000 and 2001, the reported estimate of eelgrass area in the sampled flats stratum increased by more than 150%; 3400 ha to 8600 ha (Berry et al. 2003, p.40). This was attributable to the addition in 2001 of a single randomly selected site to the flats sample (sample size of $n = 10$ sites). This site was flats11 in Samish Bay and it contained a large eelgrass bed. This very strong response associated with a change of one site led to concern about the robustness of the stratum estimates in the SVMP design.

The addition of flats11 to the flats sample was not associated with planned sample rotation, but the effect nevertheless had important implications for the possible effects of sample rotation. Several changes had been made after the initial monitoring year (2000) to the sampling frames and stratification. For this reason, the planned 20% sample rotation was not implemented until the following year. However, flats11 was added to the flats sample to compensate for the loss of a site due to the site moving to the core stratum as part of the changes after 2000 (Padilla North; Figure 2-3).

Because of the change in the flats stratum between 2000 and 2001, the original 2000 estimate was for a different population than the 2001 estimate (note difference in bar heights in Figure 2-3). For the purposes of reporting, the original 2000 sample was modified to only include sites within the 2001 stratum to make these estimates more comparable. It was comparison of this modified 2000 sample estimate to the 2001 estimate that revealed the outsized effect of flats11 on the 2001 estimate.

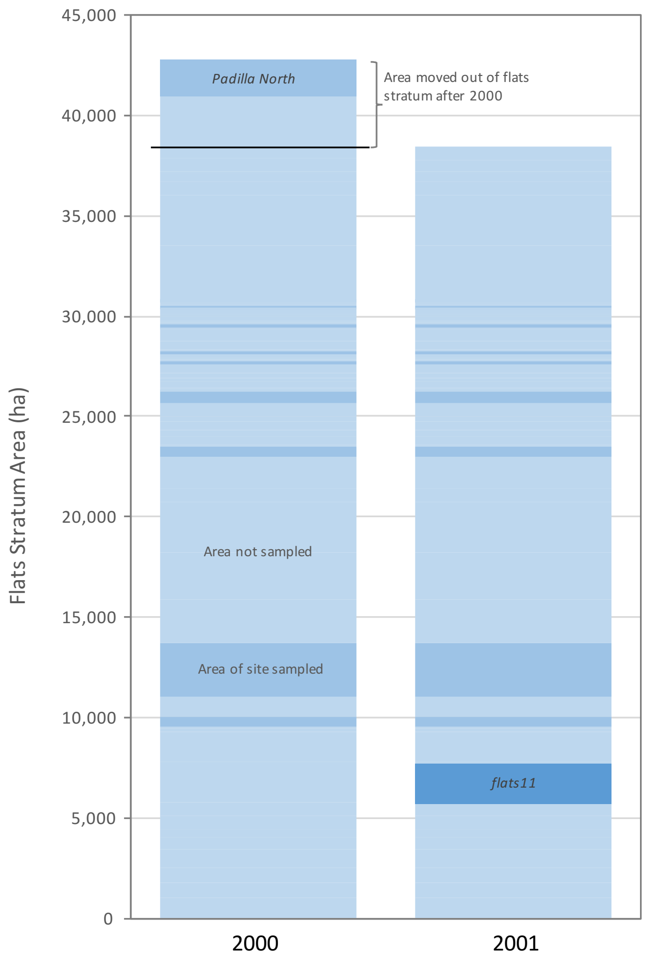


Figure 2-3. The total area of potential habitat within the sampled flats stratum in 2000 and in 2001 (bar height) and areas of the sites randomly selected for sampling in these two years (width of darker bands). Several adjustments were made to the frames and stratification after 2000 that resulted in one of the sites selected for sampling in 2000 (Padilla North) being eliminated from the flats stratum sample. In 2001, a new site was randomly selected (flats11) to bring the sample size back up to n=10 sites.

A similar effect is seen in comparisons of the original and modified 2000 sample estimates where the removal of one site, Padilla North, had a similar outsized effect that is clear in the sample data (Figure 2-4) and the associated estimates (Figure 2-5).

This flats11 effect was explained as an issue of "unusual sites in the flats stratum [that] strongly influence the overall estimate" (Berry et al. 2003, p.41). This sample design was not altered at the time, but it was recognized that "This weakness could be addressed by further partitioning the stratum" (Berry et al 2003, p.41) which echoes the suggestion of Cochran (1977, p.44). A remedy was implemented in 2004 following a systematic attempt to identify the most influential flats sites that were then segregated in a new stratum – the "persistent" flats stratum (Dowty 2005). Flats11 was included in the new stratum in addition to two other influential flat sites.

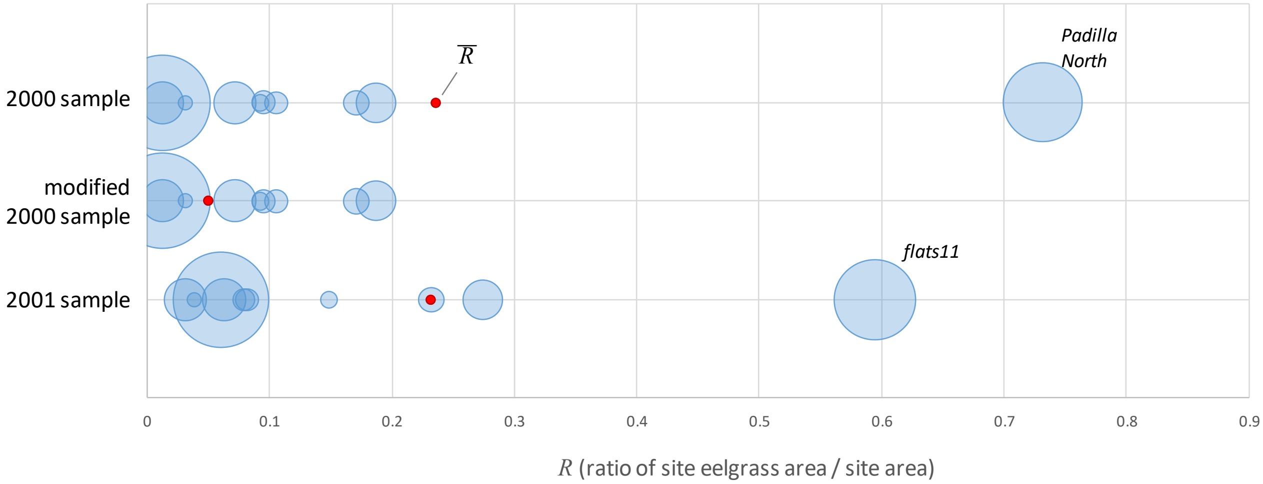


Figure 2-4. Estimates from individual sites and the sample mean value of *R*, the ratio of site eelgrass area to site area (i.e., proportion of site area occupied by eelgrass) in the flats stratum. Values are shown for the samples collected in 2000 and 2001 and a modified 2000 sample that is restricted to sites that fall within the flats stratum as it existed in 2001. The size of each circle indicates the potential habitat area within the site with larger sites having greater weight in the calculation of the mean.

While the creation of the persistent flats stratum was thought to ameliorate instability in flats stratum estimates associated with rotation of outlier sites, the remaining sites in the stratum still exhibited a very skewed distribution (Dowty 2005). There was still concern that there could be residual unreliable statistical behavior associated with site rotation.

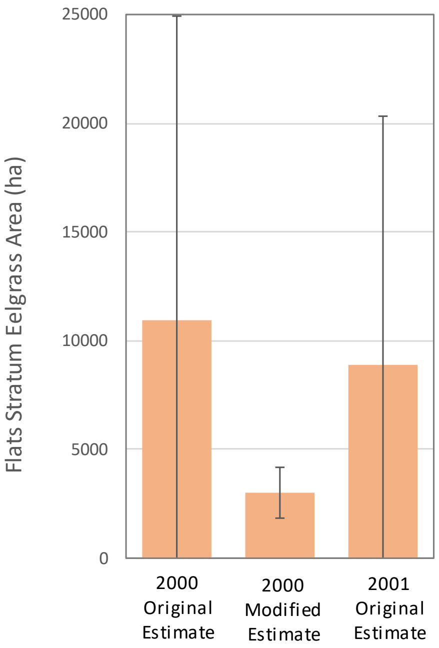


Figure 2-5. The flats stratum eelgrass area estimates for 2000 (original and modified estimates) and 2001. These are based on data shown in Figure 2-4. Error bars are 95% confidence intervals.

2.2.2 *Skepticism About a Significant Trend*

After monitoring data from 2010 was added to the SVMP dataset, for the first time the linear trend analysis gave a statistically significant increasing trend ($p < 0.01$) (Figure 2-6).

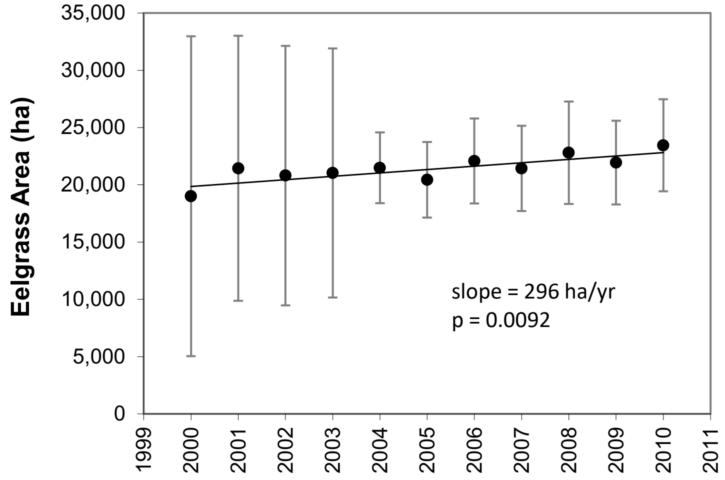


Figure 2-6. 2000-2010 SVMP estimates of soundwide eelgrass area. Error bars are 95% confidence intervals. The line was obtained with an inverse variance weighted regression and had a significant slope ($p < 0.01$). Taken from a March 2012 internal report. The 2000 estimates follow the original estimates ("2000 sample" of Figure 2-4).

There were concerns about the reliability of this trend result based on two considerations. First, the increasing trend was counter to expectations given cases of localized decline that were receiving attention. Second, the performance of the trend statistic was in doubt given that its performance had not been rigorously tested and other statistics in the design had previously been found to not perform as expected (specifically, the flats11 effect on stratum estimates and the unreliable estimates from retrospective adjustment).

These concerns led to the following data exploration. First, examination of stratum-level results indicated that the rotational flats stratum and narrow fringe stratum were responsible for the increasing trend (Figure 2-7). Second, the changes between annual stratum-level estimates were broken down into contributions from changes in matching sites sampled in both years, contributions from sites that rotated out of the sample after the first year, and contributions from sites that rotated into the sample in the second year (Figure 2-8). The contributions from the rotated sites will depend on how anomalous the sites are (how far their mean seagrass area departs from the overall sample mean).

For three of the six annual intervals in the 2004-2010 record for the flats stratum, the direct measure of change from the matching sites was of opposite sign compared to the stratum-level estimate (Figure 2-8). In each of the six intervals, especially in the three with contradictory results in the matching sites, it was clear that the stratum-level results were dominated by the effects of rotating sites. This includes both sites that rotate out of the sample, and new sites that rotate into the sample. These results added to concern about the flats stratum estimates of change – specifically that estimated change reflected changes in the selected sample rather than a change in the seagrass over time.

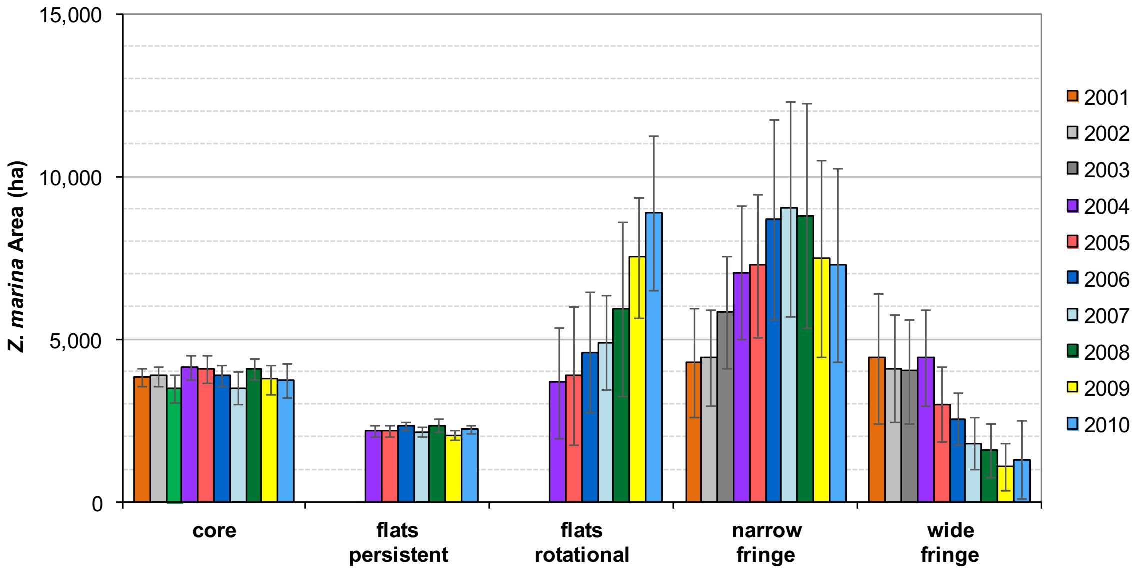


Figure 2-7. Annual estimates of annual soundwide eelgrass area by stratum. Results from early years were removed from the plot where the strata differed from the later years (2000-03 for both flats strata; 2000 for other strata – see Dowty et al. 2022, Dowty 2005 for details). Error bars are 95% confidence intervals.

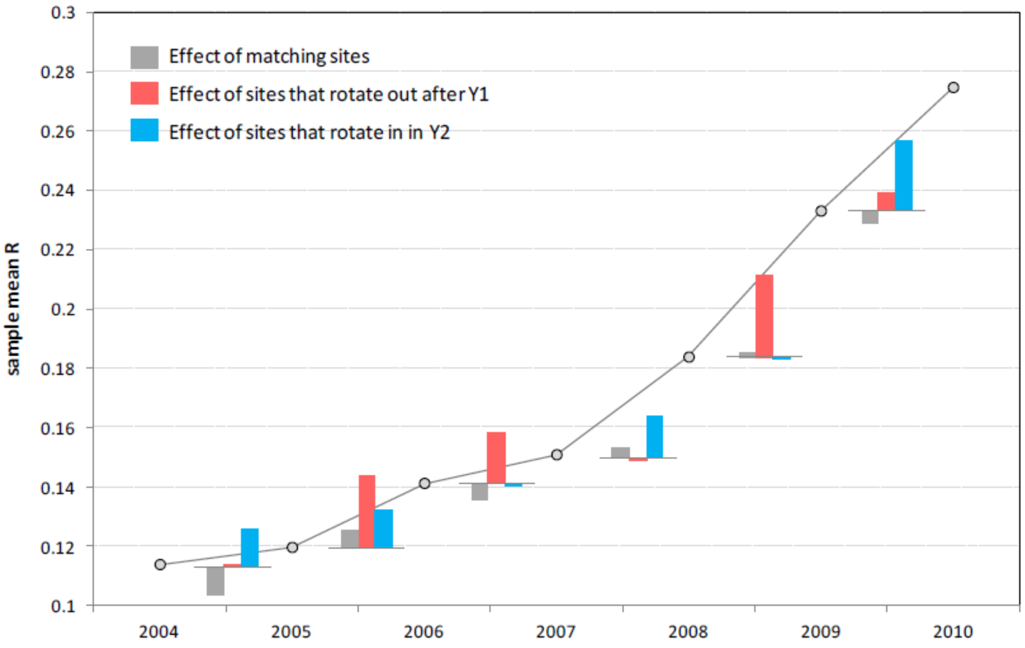


Figure 2-8. Annual estimates of rotational flats stratum parameter R (gray line) with year-to-year changes broken down into contributions from changes in matching sites sampled in both years, sites that rotate out of the sample after the first year and sites that rotate into the sample in the second year.

2.2.3 Statistician Assessments

In 2011-12, the SVMP had for the first time a significant trend in soundwide eelgrass area based on the methods detailed in the SVMP statistical framework (Skalski 2003) and the 2000-2010 data record. This had important implications for the overall assessment of Puget Sound eelgrass as presented both in regular DNR monitoring reports and in the eelgrass indicator used by the Puget Sound Partnership. It was problematic that the significance of the trend was in doubt based on exploratory analysis that suggested this could be an artifact of sample rotation (Figure 2-8).

By 2012, the SVMP had decided to solicit the perspective of professional statisticians. Ultimately, DNR retained two individuals to evaluate the situation and develop a more reliable trend analysis for the existing data: John Van Sickle PhD, a consultant who was formerly a statistician with EPA in Corvallis OR, and Chris Gast PhD who had recently finished his doctorate with John Skalski as his advisor at the University of Washington. These statisticians were aware of each other's involvement and the general approaches pursued, but they worked independently and produced very different perspectives in their final reports. They were provided with summary documents describing the sampling design and details of the implementation as well as all the site results in the 2000-2011 data record.

Van Sickle (2012) identified the assumptions of the significance test for trend with the variance-weighted linear regression as being problematic – specifically, the assumption of normally distributed variation around the trend line and year-to-year independence of annual estimates. The skewness of the distribution of site-level eelgrass areas and the sequential dependence of annual samples represent potentially serious violations of these assumptions. Permutation testing, or randomization testing, was presented as a better significance test for the SVMP dataset as it does not require these assumptions.

The existing SVMP test again resulted in a significant soundwide trend in the 2000-2011 dataset. The positive trend obtained was 219 ha/yr with a significance of $p = 0.017$ (Van Sickle 2012). In contrast, the most relevant permutation test showed the trend to be clearly non-significant. Of 1000 permutations under the null hypothesis of no trend, 683 cases had trends as large or larger than the observed slope (219 ha/yr). This translates to a trend significance of $p = 0.683$ (Van Sickle 2012).

Van Sickle (2012) also presented a linear mixed effects model and a log regression model as alternative methods for trend detection. These approaches estimate an average site-level trend rather than the trend in the total soundwide eelgrass area. Based on the R code used by Van Sickle for the linear mixed effects model (R Core Team 2017), it appears this was a random intercept model (Zuur et al. 2009, p.106). He included only simplified analyses for demonstration purposes that ignored stratification. The two methods produced almost identical trend estimates with the same significance. The demonstration resulted in a significant site-level negative trend of -0.005013 ha/yr ($p = 0.0053$).

Gast presented his work in two separate reports. The first report compared the performance of a linear mixed effects model (random intercept and slope), a simpler fixed effects linear model and other models with both frequentist and Bayesian frameworks used for inference (Gast 2012b). The comparison was based on simulated sample data. Limitations of the work included forcing all data into the latest stratification (stratification changes over time were ignored) and a modest number of simulations (*n* = 200) due to time constraints. This component of Gast's work identified the linear mixed effects model implemented in a Bayesian framework with Markov Chain Monte Carlo (MCMC) methods as the best approach.

In his second report, Gast applied the linear mixed effects model in a Bayesian framework to estimate mean site trend, rather than trend in total soundwide eelgrass area (Gast 2012b). He also estimated mean site-level relative change. In each case, the estimated trend was positive but the associated credible interval encompassed zero so that neither the possibility of a positive or a negative trend could be eliminated (Table 2-1). However, the credible interval falls mostly above zero and, according to Gast (2012b, p.8), the fact that the estimate is positive indicates that the trend is more likely positive than negative.

Gast (2012b) noted two limitations of his study. First, the heteroscedasticity in the data was not addressed (data points drawn from distributions with different variance). Second, "Only the sample (as opposed to the entire population) of sites was simulated" but even if the entire population were simulated he would expect differences to be small. A third potential limitation is that there does not seem to be any differentiation between estimation for censused strata and sampled strata. This seems to ignore a major part of the variance structure, although it is not clear how this might affect the findings.

The key findings of Van Sickle (2012) and Gast (2012b) are presented in Table 2-1. Taken together, the contributions of Van Sickle and Gast were valuable in providing independent perspectives on the SVMP challenge in assessing trend significance. They also introduced new analytical techniques to the SVMP. From a practical standpoint, the permutation test of Van Sickle (2012) was the most useful as it gave a more reliable test for significance of trend in soundwide eelgrass area with the existing dataset.

In contrast, the approaches for estimating mean site-level trend were not as immediately useful. They were certainly valuable in expanding SVMP thinking and, in the long run, may be very influential to the program, but in the short term they fell outside the existing analytical framework. The fact that Van Sickle (2012) and Gast (2012b) gave somewhat contradictory inferences for the mean site-level trend (Table 2-1) also indicated that additional analytical developments would be needed before this approach would be ready for operational use.

Table 2-1. Summary of findings from analyses of independent statisticians. Note that only the first row presents a measure that is directly comparable to the SVMP trend result (trend in total soundwide native seagrass area). The other rows contain mean site level measures. These measures also summarize the soundwide population but are sensitive to how change is distributed across sites.

Source	Parameter / Method	Estimate	Interpretation
Van Sickle (2012)	Trend in total soundwide eelgrass area. Significance from permutation test. Frequentist inference.	219 ha/yr p = 0.683	trend not significant
	Mean site-level trend. Simplified linear mixed effects model (random intercept), frequentist inference.	-0.0050 ha/yr p = 0.0053	significant decreasing trend
Gast (2012b)	Mean site-level trend. Linear mixed effects model (random intercept & slope), Bayesian inference.	0.448 ha/yr [-0.343, 1.268] 95% credible interval	Neither + or – trend ruled out. + trend more likely.
	Mean site-level relative change. Arithmetic mean, Bayesian inference.	0.173%/yr [-0.131, 0.49] 95% credible interval	Neither + or – trend ruled out. + trend more likely.

3 Model Assessment of Rotation and Soundwide Trend Estimates

In this chapter, Monte Carlo sampling of a seagrass population model is used to characterize the effects of site rotation on trend detection. Long-term (12, 20, 40 years) monitoring datasets were simulated by sampling from a static seagrass population model that represents seagrass in greater Puget Sound. A large number of simulated datasets (10,000) were used to generate sampling distributions of trend slope estimates for each scenario (12, 20, 40-year). Each of the simulated datasets differs due to (1) the random selection of sites to be surveyed, (2) the random selection of transects at sites selected for surveying, and (3) random error in video classification. The software application used for this work generated the seagrass population model, sampled the population in a Monte Carlo framework, and summarized the results (Dowty 2017).

We first present the sampling distribution for the trend estimate from a 12-year dataset generated with 20% rotation sampling. It was used to estimate the significance of the 2002-2012 observed SVMP dataset (section 3.1). Then the performance of 20-year trend estimates were assessed for scenarios that vary in the level of site rotation (section 3.2). Finally, the results were compared to those obtained with a 40-year dataset (section 3.3).

3.1 SVMP 2002-2013 Trend Significance with 20% Rotation

At the time the 2000-2013 SVMP data record was being analyzed for trends, there was concern that the uncertainty in the 2000 and 2001 annual estimates may be underestimated. This concern was associated with changes in site sampling protocols in the first two years of monitoring as they were being adjusted. This led to the exploration of trend estimates with the 2000-2001 data removed. This data exclusion is not as severe as it might seem. These two years had lowered weight to begin with because they fell in a period of annual estimates with very high variance as compared to 2004 and later (Figure 3-1). The linear regression used in trend detection uses inverse variance weighting.

The 12-year linear trend over the 2002-2013 period was estimated from SVMP monitoring data to be 210 ha/yr with a slope standard error of 61 ha/yr giving a ±120 ha/yr 95% confidence interval (Figure 3-1). The significance of the trend is $p = 0.0064$ based on the linear regression statistics as specified in the initial SVMP design (Skalski 2003) where significance is estimated with the regression slope standard error.

Under the Monte Carlo approach, significance of the 2002-2013 SVMP trend estimate is determined from the sampling distribution of the linear regression slope (Figure 3-2). The stratification used in the modeling differed slightly from operational conditions in that the most recent stratification is used throughout the 12-year record even though the persistent flats stratum was only created in 2004 (note strong improvement in precision in 2004 in Figure 3-1).

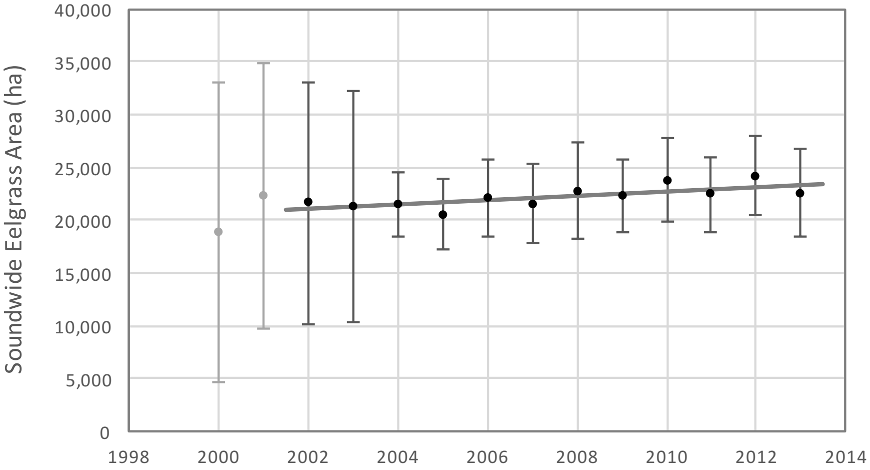


Figure 3-1. SVMP data for 2000-2013 with a regression line fit to the 2002-2013 data. The linear regression statistics indicate that the 2002-2013 trend is 210 ± 61 ha/yr (standard error of slope) and that it is significant (p = 0.0064). The 2000-2001 data are shown but were not included in the regression. The estimates shown for 2000 follow the original flats stratum estimates ("2000 sample" of Figure 2-4) and fringe estimates from the original 2000 fringe stratification (high abundance/low abundance strata).

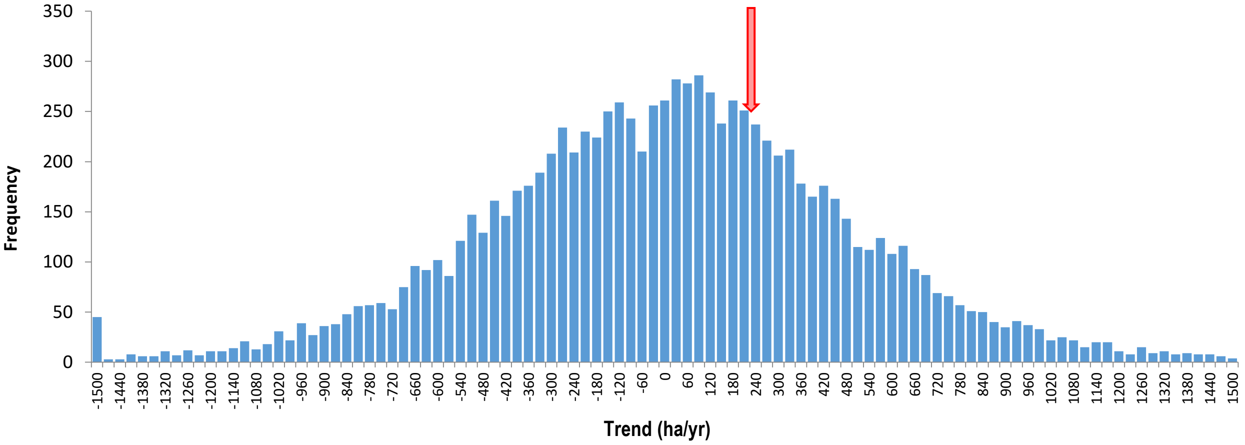


Figure 3-2. Sampling distribution of trend estimates from a 12-year dataset generated with 20% rotation sampling from a static population. The mean trend over the 10,000 estimates is near zero (0.81 ha/yr). The standard deviation of the sampling distribution is 496 ha/yr. The red arrow indicates the position of the SVMP trend estimate from Figure 3-1, 210 ha/yr. The proportion of the sampling distribution with trends of absolute magnitude greater than 210 ha/yr is 64% indicating a significance of p = 0.64 for the SVMP 12-year trend estimate.

The sampling distribution of the 12-year trend estimates is centered around a trend of 0 ha/yr reflecting the static sampled population (Figure 3-2). The mean trend over the 10,000 simulations is a near-zero value of 0.81 ha/yr. The standard deviation of the sampling distribution is 496 ha/yr which indicates a fairly low precision in trend estimation.

The 2002-2013 SVMP trend estimate of 210 ha/yr is well within the central portion of this sampling distribution (Figure 3-2). A total of 64% of the 10,000 simulations had trends of greater absolute magnitude than the SVMP estimate. This translates to a significance of $p = 0.64$ for the SVMP estimate. This means that the 2002-2013 SVMP trend estimate is well within the range expected just from random error when sampling from a static population.

The model-based significance of $p = 0.64$ for the 2002-2013 SVMP trend estimate is similar to the significance of the 2000-2011 SVMP estimate based on a permutation test ($p = 0.683$; Table 2-1). This general agreement of these two different approaches supports the conclusion that the highly significant trend results from the initial SVMP design are not reliable and should not be used.

3.2 Effects of Varying Rotation

To understand the effect of site rotation on trend estimation, simulations were run at different levels of annual site rotation (0, 20, 40, 60, 80, 100%). For each level of rotation assessed, 20-year datasets were simulated by sampling from a static soundwide population model. For each dataset, the 20-year trend was estimated as the linear regression slope. The trend standard deviation was estimated as the slope standard error as specified in the SVMP statistical framework (Skalski 2003). Each rotation scenario was characterized by 10,000 simulations.

If b_i is the trend estimate from the i^{th} simulation, then the true standard deviation of the trend estimator is simply

$$\sigma = \sqrt{\frac{\sum_{i=1}^{10000}\left(b_i - \bar{b}\right)^2}{10000-1}}. \qquad \textbf{Equation 3-1}$$

The mean of the slope standard error estimates $\bar{s}$, is used to calculate the relative bias of slope standard error estimator as

$$\text{Relative Bias} = \frac{\bar{s}-\sigma}{\sigma}. \qquad \textbf{Equation 3-2}$$

Frequency histograms show the strong effects of rate of rotation on the precision of trend estimates (Figure 3-3). There are also strong effects on bias in the estimation of trend standard deviation (Figure 3-4).

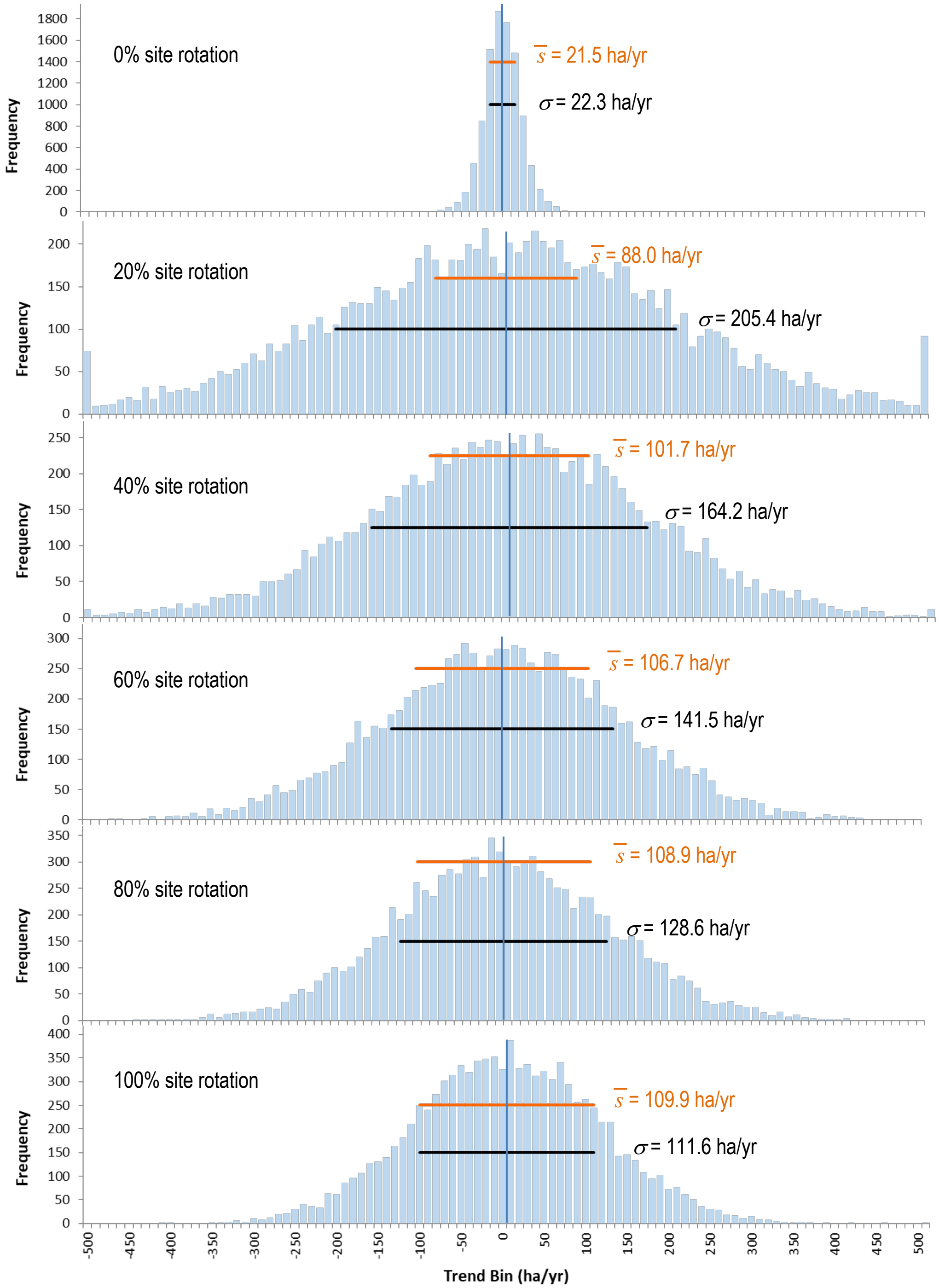


Figure 3-3. Frequency histograms of 20-year trend estimates based on annual sampling from a static soundwide population model with different levels of site rotation between years. Histograms summarize results from 10,000 simulations. Horizontal lines indicate the magnitude of the true standard deviation of the trend sampling distribution (σ) and the mean estimated value ($\bar{s}$). Differences between the lengths of these lines indicate bias in the estimation of trend standard deviation.

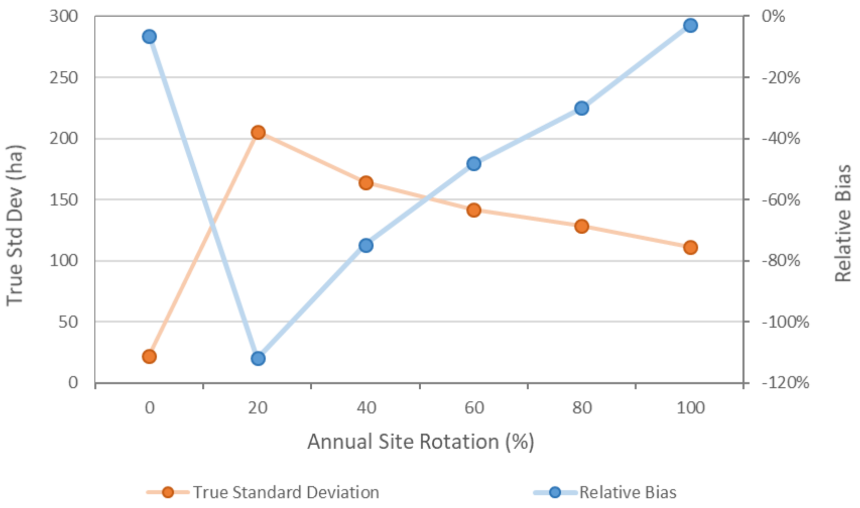


Figure 3-4. The true standard deviation and bias in standard deviation estimation for 20-year trends when sampling from a static population with different rates of site rotation. The true standard deviation is calculated from trend estimates from 10,000 simulated datasets. The relative bias captures the departure of estimates of trend standard deviation from the true standard deviation. Note that the relative bias is negative in each case indicating underestimation of the true standard deviation.

Of the rates of rotation investigated, the 20% rotation used by the SVMP has the worst performance for trend estimation. It had the lowest precision (highest true standard deviation) and the greatest magnitude of bias in estimation of precision (-112% relative bias). The strong negative bias means that with 20% rotation sampling, the estimates of standard error of slope strongly underestimate the true standard error. This is consistent with the overly aggressive determinations of significance seen in the SVMP trend results (Figure 2-6, p.19; Figure 3-1, p.26).

These results suggest that while the 20% site rotation in the SVMP design was designed to balance current estimates and change estimates between two occasions (section 2.1.1), it does not optimize for trend estimates. In fact, 20% rotation appears to be the worst possible choice leading to strong degradation in trend precision and increased bias in precision estimates leading to unreliable trend significance tests.

3.3 Rates of Rotation and Data Record Length

We saw in Figure 3-4 how trend precision is a function of the rate of site rotation. Of the data points shown, 20% rotation led to the greatest degradation in precision. The function, however, is poorly resolved and it is not clear which level of rotation within the (~1%, ~39%) range minimizes precision (maximizes standard deviation). Another limitation in the results in Figure 3-4 is that only 20-year trends are examined. It is possible that the precision curve (as a function of site rotation) may vary as the length of the data record varies.

Additional simulations were conducted to address these issues. A more finely resolved precision curve was generated. Also, results with a 20-year data record were compared to results with a 40-year record. To simplify the task, the scope was limited to one sampling stratum. The narrow fringe stratum was selected since it has the largest sample size (typically, *n*=45) and therefore the greatest possible resolution along the site rotation axis (e.g., 1,2,3… 45 of 45 sites rotate).

When evaluating the effects of the data record length, it is useful to compare designs with different levels of site rotation in terms of related period of site retention. Under the 20% rotation design, for example, once a site is randomly selected to the sample it is retained for five sampling occasions (five years) before rotating out of the sample (Table 3-1). In some cases, such as with 40% rotation (Table 3-1), different sites may have different retention periods in which case the mean retention period is a useful concept.

Table 3-1. Tabular representation of a 20% rotation design (left) over nine sampling occasions and a 40% rotation design (right) over six sampling occasions for a sample size of *n*=10. An "X" in a cell indicates that the site for that row is surveyed on the occasion represented by that column. Under the 20% design, each site is retained for five sampling occasions before it is rotated out of the sample. Under the 40% design, some sites are retained for two years and some for three years.

	Sampling Occasions								
Site	1	2	3	4	5	6	7	8	9
1	X								
2	X								
3	X	X							
4	X	X							
5	X	X	X						
6	X	X	X						
7	X	X	X	X					
8	X	X	X	X					
9	X	X	X	X	X				
10	X	X	X	X	X				
11		X	X	X	X	X			
12		X	X	X	X	X			
13			X	X	X	X	X		
14			X	X	X	X	X		
15				X	X	X	X	X	
16				X	X	X	X	X	
17					X	X	X	X	X
18					X	X	X	X	X
19						X	X	X	X
20						X	X	X	X
21							X	X	X
22							X	X	X
23								X	X
24								X	X
25									X
26									X

	Sampling Occasions					
Site	1	2	3	4	5	6
1	X					
2	X					
3	X					
4	X					
5	X	X				
6	X	X				
7	X	X				
8	X	X				
9	X	X	X			
10	X	X	X			
11		X	X			
12		X	X			
13		X	X	X		
14		X	X	X		
15			X	X		
16			X	X		
17			X	X	X	
18			X	X	X	
19				X	X	
20				X	X	
21				X	X	X
22				X	X	X
23					X	X
24					X	X
25					X	X
26					X	X
27						X
28						X
29						X
30						X

The results of the additional simulations are presented in Figure 3-5. The finer resolution on the 20-year curve shows that the worst rate of rotation, in terms of trend precision, is not 20% but rather some value slightly less than 10%. The peak standard error shown in Figure 3-5 is 9% but the exact peak is still only coarsely resolved.

The results with a 40-year dataset differ in two main respects. First, precision is much improved with the longer dataset for all levels of site rotation. Second, the worst rate of rotation has shifted lower (longer period of site retention). The peak value of standard error was not well resolved but occurs at approximately 4% site rotation (approximately 20-year site retention).

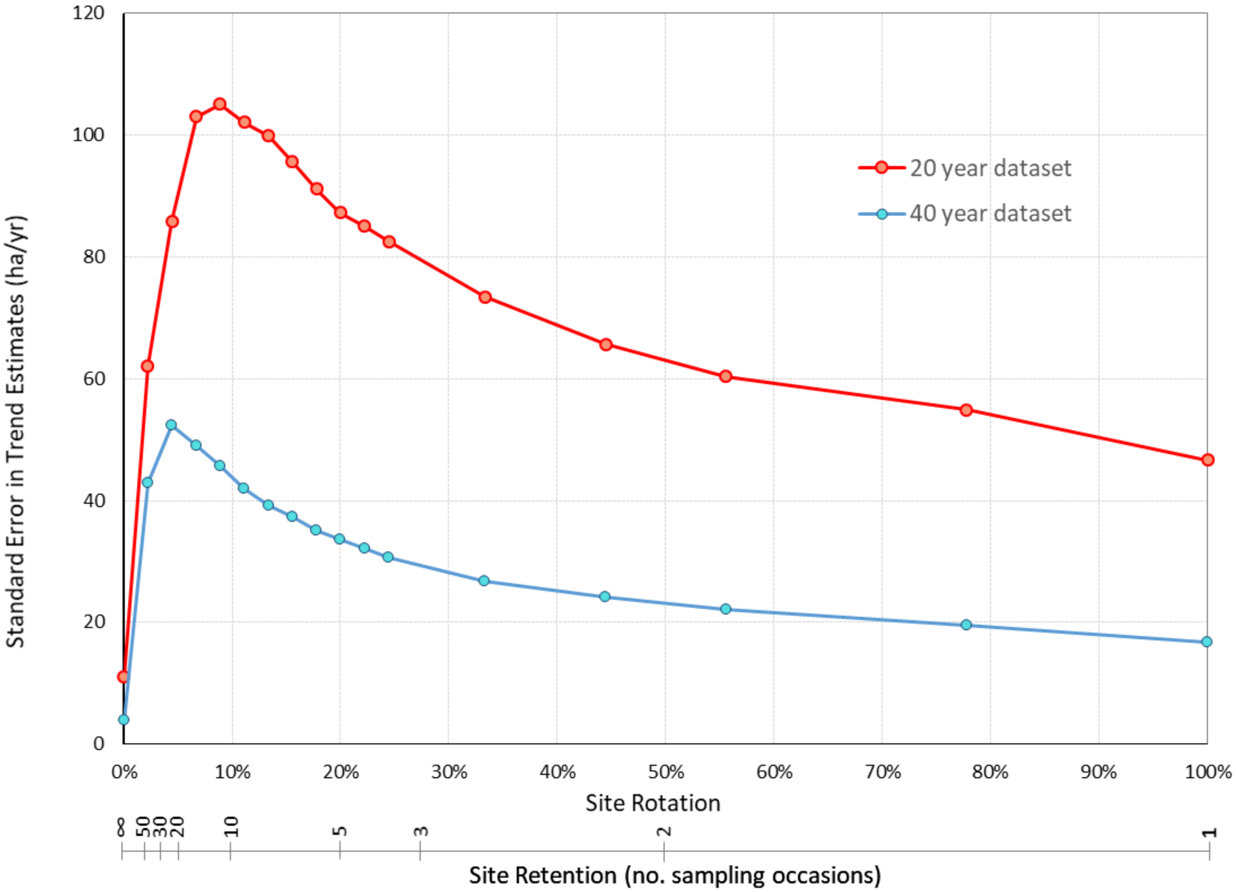


Figure 3-5. Standard error in trend (slope) estimates as a function of the rate of site rotation for a static population sampled over 20 years and over 40 years. Results show true standard error as calculated from 10,000 trend estimates. The secondary *x*-axis is the period that a randomly selected site is retained in the sample before rotating out.

While these results are based on simulations for the narrow fringe stratum alone, it is reasonable to assume that similar patterns would be seen if the simulations encompassed all strata. For SVMP sampling overall, we can then make two main conclusions:

1) Sample rotation, or partial sample replacement, is detrimental to the precision of trend estimation as compared to either retaining one sample over all occasions, or drawing a new sample at each occasion.
2) For analysis of a data record of some given length, a rotational design that gives a period of sample retention of about half that length is the most detrimental to trend precision.

In addition, the results suggest that over time as the record is extended, the magnitude of the loss in trend precision due to site rotation diminishes. For all cases, trend precision is maximized with fixed sites (no rotation).

4 Comparison of Alternative Designs

In 2015, the SVMP was working with a data record of 15 years (2000-2014). Soundwide trend analysis was increasingly important for the program. The current soundwide estimate (total hectares of native seagrass) and most recent annual change estimate were increasingly of lesser relative importance. Given the finding that the 20% site rotation strongly degraded the precision of trend estimates (Figure 3-4, p.29), the question arose of whether an alternative design might be more effective in addressing the program's priorities. The SVMP staff discussed several designs, each with its own anticipated benefits, but there was no straightforward way to quantify and objectively compare the performance of these designs.

We identified simulated sampling from a seagrass population model as a tool to compare sampling design alternatives. Modelling experiments were conducted to determine performance of the alternative designs on the basis of precision of trend estimates, Type I error rates, and power to detect trend under different change scenarios imposed on the model seagrass population.

4.1 Five Alternative Designs

We identified five alternative designs for comparison. These designs vary in sample replacement policy across sampling occasions and in the method of estimation of the annual soundwide native seagrass area that is subsequently used in trend estimation.

Four different sample replacement policies were compared. These include:

- 20% sample rotation (the policy of the original SVMP design)
- same sample for all occasions (0% rotation, or, “fixed” sites)
- newly drawn sample for each occasion (100% rotation)
- three rotating panels

The three rotating panels policy consists of three independent samples of sites that are sampled in three consecutive years and then in a repeating sequence. The four sample replacement policies are represented schematically in Figure 4-1.

Two different methods for estimating annual soundwide eelgrass area were compared. These include:

- annual estimate from annual site sample,
- annual estimate from 3-year pooled site sample (includes year prior and following).

	Sampling Occasions									
Panel	1	2	3	4	5	6	7	8	9	10
1	X									
2	X	X								
3	X	X	X							
4	X	X	X	X						
5	X	X	X	X	X					
6		X	X	X	X	X				
7			X	X	X	X	X			
8				X	X	X	X	X		
9					X	X	X	X	X	
10						X	X	X	X	X
11							X	X	X	X
12								X	X	X
13									X	X
14										X

A. 20% sample rotation.

	Sampling Occasions									
Panel	1	2	3	4	5	6	7	8	9	10
1	X	X	X	X	X	X	X	X	X	X

B. same sample retained for all occasions (0% rotation, “fixed” sites)

	Sampling Occasions									
Panel	1	2	3	4	5	6	7	8	9	10
1	X									
2		X								
3			X							
4				X						
5					X					
6						X				
7							X			
8								X		
9									X	
10										X

C. newly drawn sample for each occasion (100% rotation)

	Sampling Occasions									
Panel	1	2	3	4	5	6	7	8	9	10
1	X			X			X			X
2		X			X			X		
3			X			X			X	

D. three rotating panels

Figure 4-1. Schematic representations of the four sample replacement policies compared. Each row represents a panel – a set of sites selected as a group that are always sampled (or not) as a group – and shows when the panel would be surveyed over a sequence of ten sampling occasions. The number of sites in a panel varies. For example, the number of sites in an annual sample for the narrow fringe stratum is typically n = 45, so the number of sites in each narrow fringe panel would be 9 in A but 45 in B, C and D.

The four different site replacement policies were coupled with the two methods for estimating annual eelgrass area estimates such that there were five unique designs (Figure 4-2). The replacement policy of a newly drawn sample each occasion was coupled with each analysis method giving two distinct designs. These are the five designs that were compared from the perspective of trend analysis.

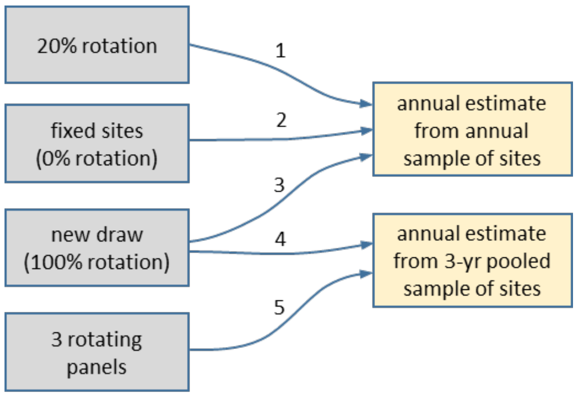


Figure 4-2. The sampling design components that differentiate the five designs compared in this study. The four site replacement policies (left) are coupled with analysis methods for estimating annual soundwide eelgrass area (right). The replacement policy of drawing a new sample each occasion (100% rotation) is coupled with each analysis method to give two different designs for a total of five.

4.2 Basis for Comparisons

For our comparisons, we retain the central role of linear regression from the initial SVMP design. We have seen strong bias in some cases with the regression estimators (Figure 3-4, p.29), and this has important implications for these comparisons. To understand these implications, we need to distinguish between the two independent estimates generated by linear regression.

The first is the estimate of slope of the line fit to the annual seagrass area estimates based on least squares fitting. This estimation of slope from linear regression is used as the trend estimator in all comparisons. We did not see any evidence of bias in slope estimation in the earlier analysis (Figure 3-3, p.28), so we did not anticipate any issues with using these estimates in the Type I error and power assessments associated with statistical testing.

The second estimate from linear regression is the estimate of the standard error of the estimated slope. This estimator did display bias in the earlier analysis (Figure 3-4, p.29), including cases of large magnitude bias (< -100%). We attribute this bias to the violation of independence in the annual seagrass area estimates that are input data for the linear regression, i.e. these estimates are not based on independent random samples. This has important implications for the reliability of analyses that rely on these slope standard error estimates, namely the assessments of Type I error and power associated with statistical testing for slope significance.

The bias in the slope standard error estimate could potentially be avoided by replacing the slope significance test with another test such as the permutation test discussed earlier (p.21) that was proposed by Van Sickle (2012). Alternatively, the underlying linear regression model could be replaced by a mixed effects model that accommodates the dependence in annual seagrass area estimates. Neither of these solutions had been implemented in the computer application (Dowty 2017), so a more expedient approach was taken that utilizes the Monte Carlo simulations.

For the power analyses, the slope standard error estimate from linear regression was replaced with the true standard error evaluated from the slope sampling distribution generated from the Monte Carlo simulations. This insulates the power analyses from any bias due to dependence in the annual seagrass area estimates and provides a fair comparison of the five different designs. These power analyses, then, represent either the performance of sampling designs that incorporate Monte Carlo analysis as a numerical element, or these could be considered the potential power that might be achieved with sampling designs that incorporate alternative testing (permutation test) or an alternative underlying model (mixed effects model).

For the purposes of assessing Type I error of the slope significance test, the regression estimate of slope standard error is retained. This means that any bias will propagate to the estimated Type I error rate. If the magnitude of bias in slope standard error is such that the linear regression model is deemed inappropriate for this purpose, then the Type I error assessments will also reflect this problem.

To summarize the basis for comparing the five designs, design performance is evaluated in three different categories. These categories are:

1) **Precision and accuracy of linear regression estimators.**
   This assesses the performance of the linear regression estimators under the different designs. For the linear regression slope standard error estimator, this quantifies any bias present due to data dependence.

2) **Type I error and its conformance to the α used in hypothesis testing.**
   This assesses a key aspect of performance of the trend slope significance testing. This assessment of Type I error utilizes the linear regression estimators. Consequently, if the slope standard error estimator is biased, this shows that bias transformed to Type I error discrepancy.

3) **Power to detect trend under different change scenarios.**
   This assesses the power of the sampling design to detect different change scenarios imposed on the seagrass population model over a 20-year monitoring period. The power analyses utilize the linear regression estimator slope but slope standard error is determined directly from the slope sampling distribution compiled from the Monte Carlo simulations.

The methods for the comparisons in these three categories are presented in the following sections. The results of these comparison are presented in sections 4.4, 4.5, and 4.6.

4.2.1 Precision and Accuracy of Regression Estimators

The assessment of precision and accuracy of estimates was based more specifically on:

- Precision of the trend estimates from the linear regression slope estimator.
- Accuracy of the trend slope estimates from the linear regression slope estimator.
- Accuracy of the trend slope standard error estimates from the linear regression slope standard error estimator.

These comparisons were based on repeated sampling of a static model seagrass population over a 20-year period (20 sampling occasions). A Monte Carlo approach generated a large number (10,000) of simulated 20-year sample datasets that were analyzed for linear trend using inverse-variance weighted regression. The frequency histogram of the 10,000 trend slope estimates is the sampling distribution of the trend slope estimator, i.e. the parent population from which sample trend slope estimates are drawn.

Trend precision is a measure of the dispersion of this sampling distribution of the linear regression slope estimator $\hat{\beta}$. Given a sample dataset, the estimator $\hat{\beta}$ generates a slope estimate b. The standard deviation of the 10,000 values of b in the sampling distribution is σ_b. It is given by

$$\sigma_b = \sqrt{Var(\hat{\beta})} = \sqrt{\frac{\sum_{i=1}^{10000}(b_i-\mu_b)^2}{10000-1}}$$ **Equation 4-1**

where

b_i= the i^{th} regression estimate of the linear trend slope β, for i = 1, 10,000.

μ_b = the mean trend slope estimate given by $\mu_b = \frac{\sum_{i=1}^{10000} b_i}{10000}$.

$\hat{\beta}$ = the estimator for linear trend slope β, that generates a slope estimate b given data from a sample. This can be denoted $\hat{\beta} = b$, and is simply the linear regression slope estimator.

The **accuracy of the trend slope estimator** was measured by the observed bias – the difference between the mean of the sampling distribution of trend slopes (the expected value of the estimator $\hat{\beta}$, denoted $E(\hat{\beta})$) and the true trend of the model population β. The true trend β in this case has a value $\beta = 0$ since we are sampling from a static model seagrass population.

The bias, B_b, is calculated as

$$\begin{gathered} B_b = E(\hat{\beta}) - \beta \\ B_b = \mu_b - \beta \\ B_b = \mu_b \end{gathered}$$ **Equation 4-2**

The **accuracy of the trend standard error estimator** was similarly measured by observed bias. But in this case, we are concerned with the slope standard error estimator $\hat{\sigma}_b$ which generates a standard error estimate s given a sample dataset. The bias of the estimator $\hat{\sigma}_b$ is the difference between the mean of the sampling distribution of trend standard error (the expected value of the standard error estimator $\hat{\sigma}_b$) and the true slope standard error σ_b. The true standard error σ_b is determined with Equation 4-1 from the sampling distribution of the slope estimator $\hat{\beta}$. Bias in the trend slope standard error estimator is then given by

$$B_s = E(\hat{\sigma}_b) - \sigma_b$$ **Equation 4-3**

where

$$E(\hat{\sigma}_b) = \frac{\sum_{i=1}^{10000} s_i}{10000}.$$

and s_i is the i^{th} estimate of standard error of the estimated slope.

4.2.2 Type I Error

The Type I error of the five designs was compared, also based on the Monte Carlo sampling of a static population over a 20-year period. Beyond the regression estimates themselves, this required a statistical test to detect trend slopes significantly different from zero. A *t*-test was used to detect significant slopes based on the weighted regression estimates of slope and of slope standard error. The *t* statistic was calculated as

$$t_i = \frac{b_i - \beta}{s_i} = \frac{b_i}{s_i}$$ **Equation 4-4**

and the test entails comparing t_i to the critical value t_c determined from a *t* distribution with $\alpha = 0.05$ and $df = 20 - 2$.

The Type I error rate is the proportion of tests of the 10,000 Monte Carlo samples that have statistically significant slopes. Since the true slope is 0 for a static population, the significant results represent false positives. The observed Type I error rate is calculated as

$$\alpha_{obs} = \frac{\sum_{i=1}^{10000} \tau_i}{10000}$$ **Equation 4-5**

where

$$\tau_i = \begin{bmatrix} 0 & |t_i| < t_c \\ 1 & |t_i| \geq t_c \end{bmatrix}$$

and t_c is the critical t value for $\alpha = 0.05$ and $df = 20 - 2$.

4.2.3 Power

The comparison of power was also based on Monte Carlo sampling of model populations over a 20-year period but instead of sampling a static population, different change scenarios were imposed on the population. The actual change scenarios used are described below (section 4.3).

For the purposes of power analyses, we use the true slope standard error (Equation 4-1), rather than the linear regression sample estimate, so we can now use a *z* test, rather than the *t* test used earlier (Equation 4-4). The *z* statistic is calculated as

$$z_i = \frac{b_i - \beta}{\sigma_b}$$ **Equation 4-6**

and the test entails comparing z_i with the critical value determined from the normal distribution with $\alpha = 0.05$.

4.3 Change Scenarios

Three sets of change scenarios were investigated. The scenarios in each set were intended to be compared as a group to highlight the effects of a particular pattern of change. The scenarios are summarized here but are described in greater detail in Dowty (2017).

Each scenario is specified with by size class. A size class represents all the sites in the seagrass population model that have a native seagrass area within a specified interval of size values. Size classes are best understood with reference to the frequency distribution of site native seagrass areas (Figure 4-3). The distribution is highly skewed with many sites with small eelgrass beds and few sites with very large eelgrass beds. For example, in the initial state of the seagrass population model, the largest site has an eelgrass area over 3000 ha but the 0 – 10 ha size class contains 86% of all the sites.

The first set of change scenarios is a set of ten ‘increasing soundwide loss’ scenarios that are a sequence of increasing soundwide seagrass loss over the 20-year model period. The other two sets of scenarios all reflect the same level of 20-year eelgrass loss but the pattern of loss within the population varies. The ‘increasing size class’ and ‘decreasing prevalence’ sets of scenarios are each a sequence with loss restricted to an increasingly small sub-population of sites. These are described in more detail below.

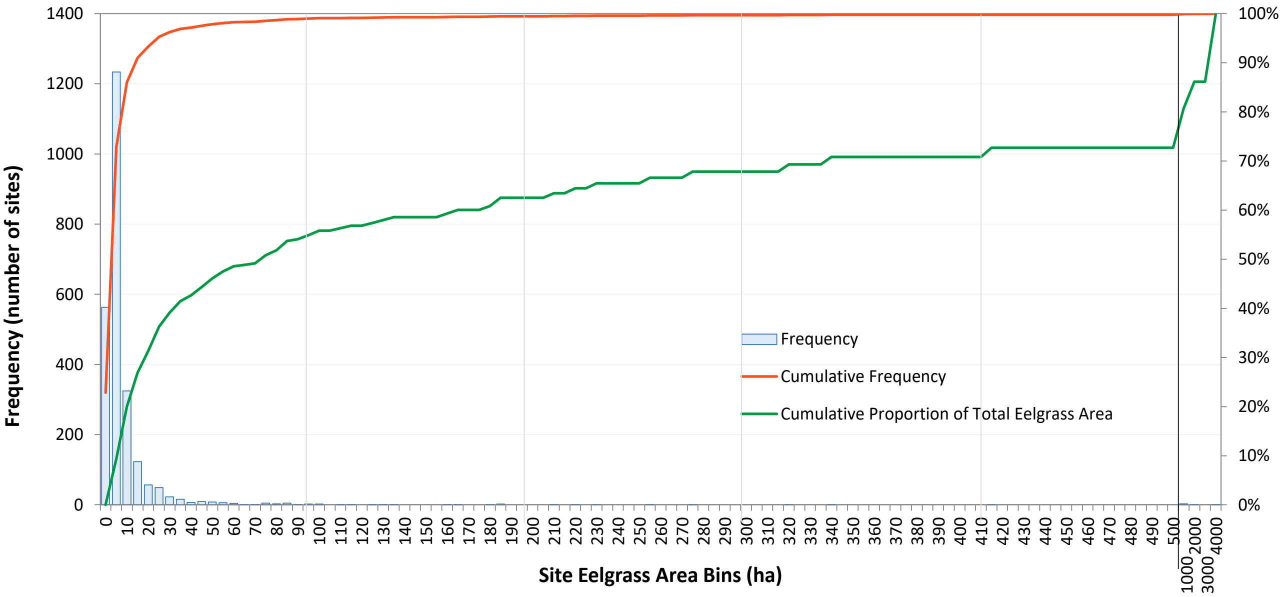


Figure 4-3. Frequency histogram of sites by eelgrass area (bars) with the cumulative frequency (red) and the cumulative proportion of total eelgrass area (green).

4.3.2 Increasing Soundwide Loss Scenarios

There are ten scenarios in this set that reflect 20-year soundwide eelgrass loss ranging from -3.5% to -19.9% (Figure 4-4). Each scenario only affects sites in the 0 – 10 ha size class and within this size class the prevalence of decline is 100% – i.e., every site in the size class with eelgrass present experiences decline. While the 0 – 10 ha size class contains 86% of all sites in the population, it only contains 63% of the sites with eelgrass present. The difference is made up of unvegetated sites in the 0 – 10 ha size class that do not change under these scenarios. The increasing loss across the scenarios is achieved by an increasing intensity of loss applied across all sites affected (see Figure A-1, p.56).

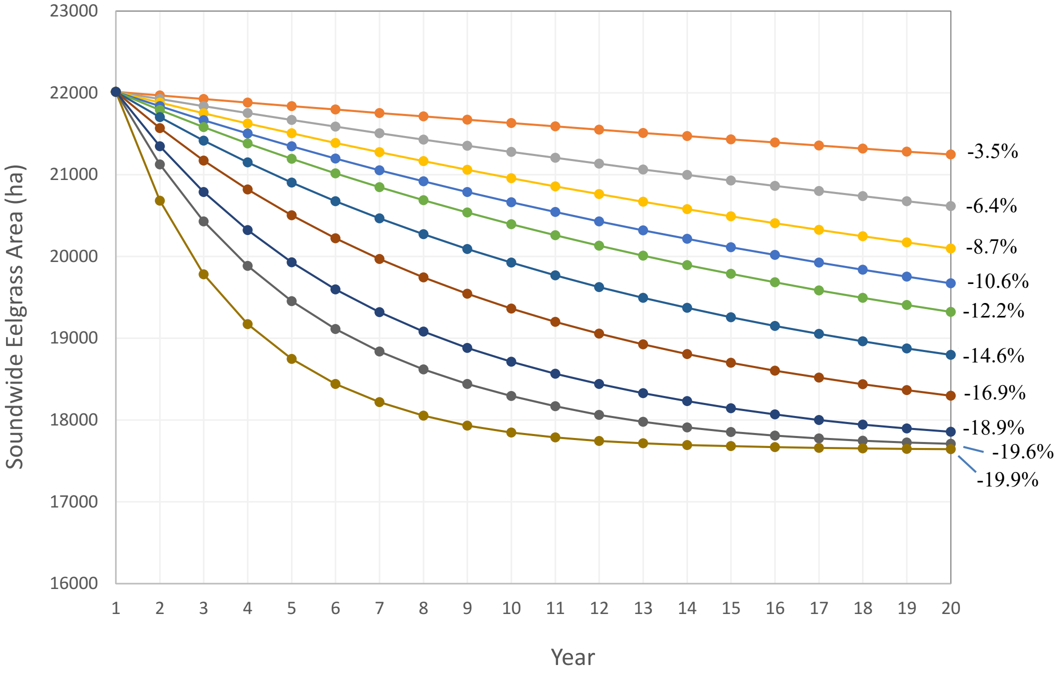


Figure 4-4. The 'increasing soundwide loss' scenarios. Annual soundwide eelgrass area values are shown for the ten change scenarios that reflect increasing levels of annual decline applied to the 0-10 ha size class. Each scenario is labelled with the 20-year eelgrass loss as a percentage of the initial value.

4.3.3 Increasing Size Class Scenarios

There are five scenarios in this set that all reflect the same 20-year soundwide seagrass loss (-10%; Figure 4-5) but the scenarios restrict the declines to a sequence of increasing size classes (Figure 4-6). The sequence of scenarios also has decreasing numbers of sites exhibiting decline, but the sites are larger so that the total 20-year decline is constant. In addition, the five site classes were constructed so that the initial eelgrass area within each size class is the same (Figure 4-7). The intensity of decline at the sites affected by decline is the same across all scenarios in the set (see Figure A-2, p.57).

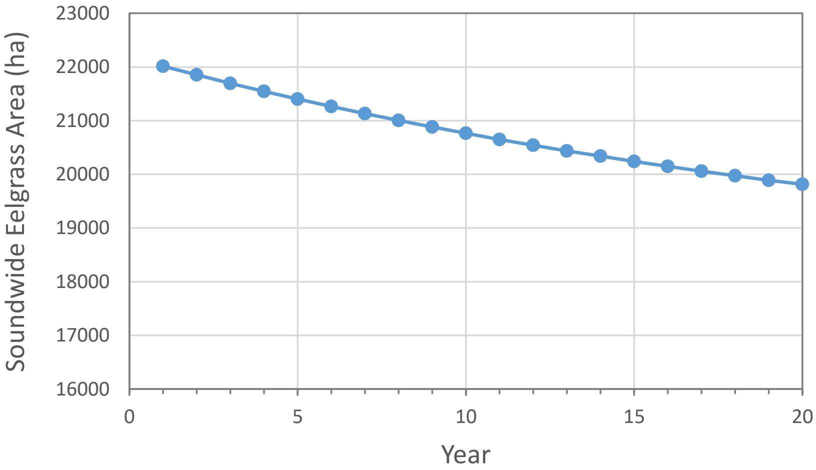


Figure 4-5. The 20-year decline in soundwide eelgrass reflected in each of the five 'increasing size class' scenarios. This same declining pattern is reflected in the six 'decreasing prevalence' scenarios.

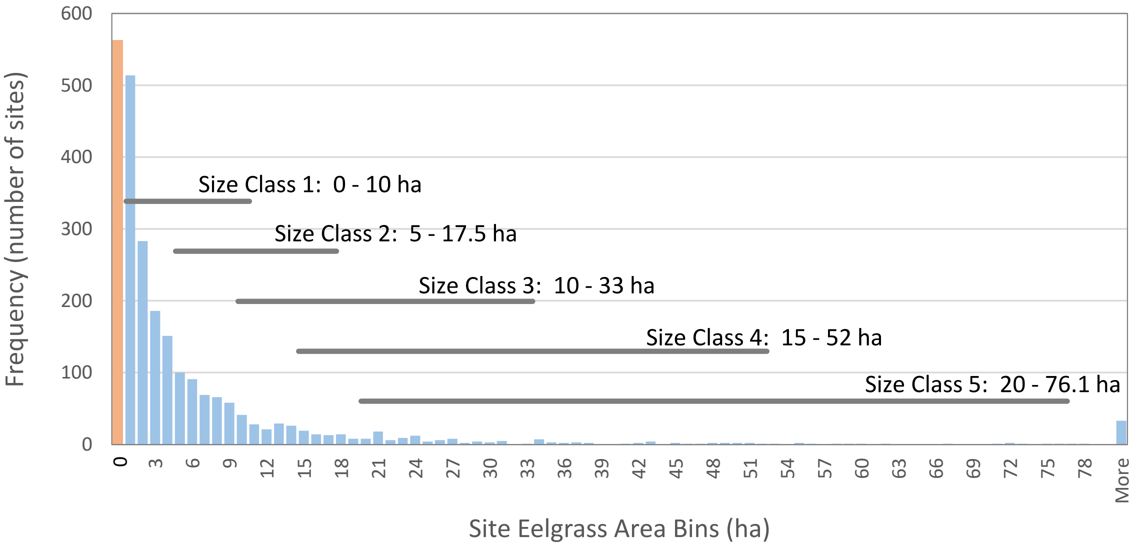


Figure 4-6. Frequency histogram of site seagrass area under initial conditions with the five size classes used in the 'increasing size class' scenarios. The size classes are depicted by horizontal bars that show the site seagrass areas encompassed within the size class. The frequency of sites with no eelgrass is shown with an orange bar. These size classes were constructed so that the total initial eelgrass area within each size class is the same (see Figure 4-7).

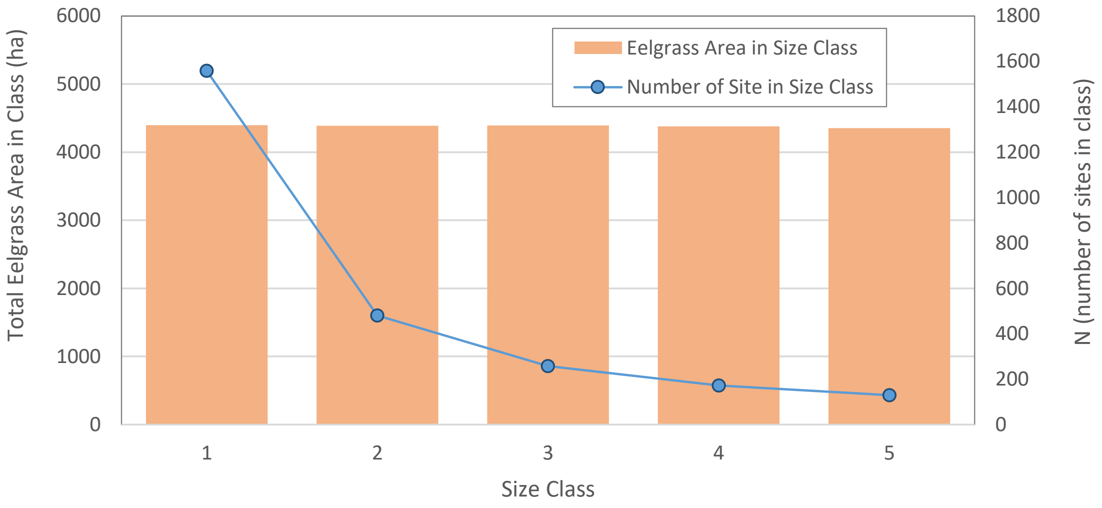


Figure 4-7. The total seagrass area under initial conditions and the number of sites within each of the five size classes.

4.3.4 Decreasing Prevalence Scenarios

There are six scenarios in this set that all reflect the same 20-year soundwide seagrass loss (-10%; Figure 4-5). In each case, site declines are restricted to the 0 – 30 ha size class. This size class includes approximately 73% of sites with seagrass and 39% of the total native seagrass area in the seagrass population model. The scenarios represent a sequence of decreasing prevalence of decline but with increasing intensity (see Figure A-3, p.58). The scenarios were devised by first selecting a sequence of decreasing prevalence values and then adjusting the mean site decline for each given prevalence so that each combination gave the same 20-year soundwide loss of eelgrass area (-10%).

4.4 Precision and Accuracy

The trend slope estimates derived from the simulated 20-year datasets are summarized in frequency histograms in Figure 4-8 for sampling from a static population using the five alternative designs. The dispersion in these sampling distributions represents the precision in trend estimation associated with the five designs. The precision of the trend slope estimates as measured by standard deviation is shown as σ in the labels of Figure 4-8, the bars in Figure 4-9 and the 'true s.d.' listed in Table 4-1. The five designs can be placed into three categories on the basis of trend precision:

low trend precision: 20% rotation
moderate trend precision: new draw (1yr) and new draw (3yr)
high trend precision: fixed sites and 3 rotating panels

The accuracy of trend slope estimation is represented by displacement of the distribution means in Figure 4-8 from the true slope of zero (the model population is static). The displacement is less than the bin size of the histograms and, therefore, not visible in the figure. This suggests that all five designs have high accuracy in trend slope estimation. The actual displacement values are shown as 'mean slope' in Table 4-1 and they indicate bias is low in all cases but the 20% rotation design has the highest bias (-0.24 ha/yr) with magnitude seven times greater than the next highest value.

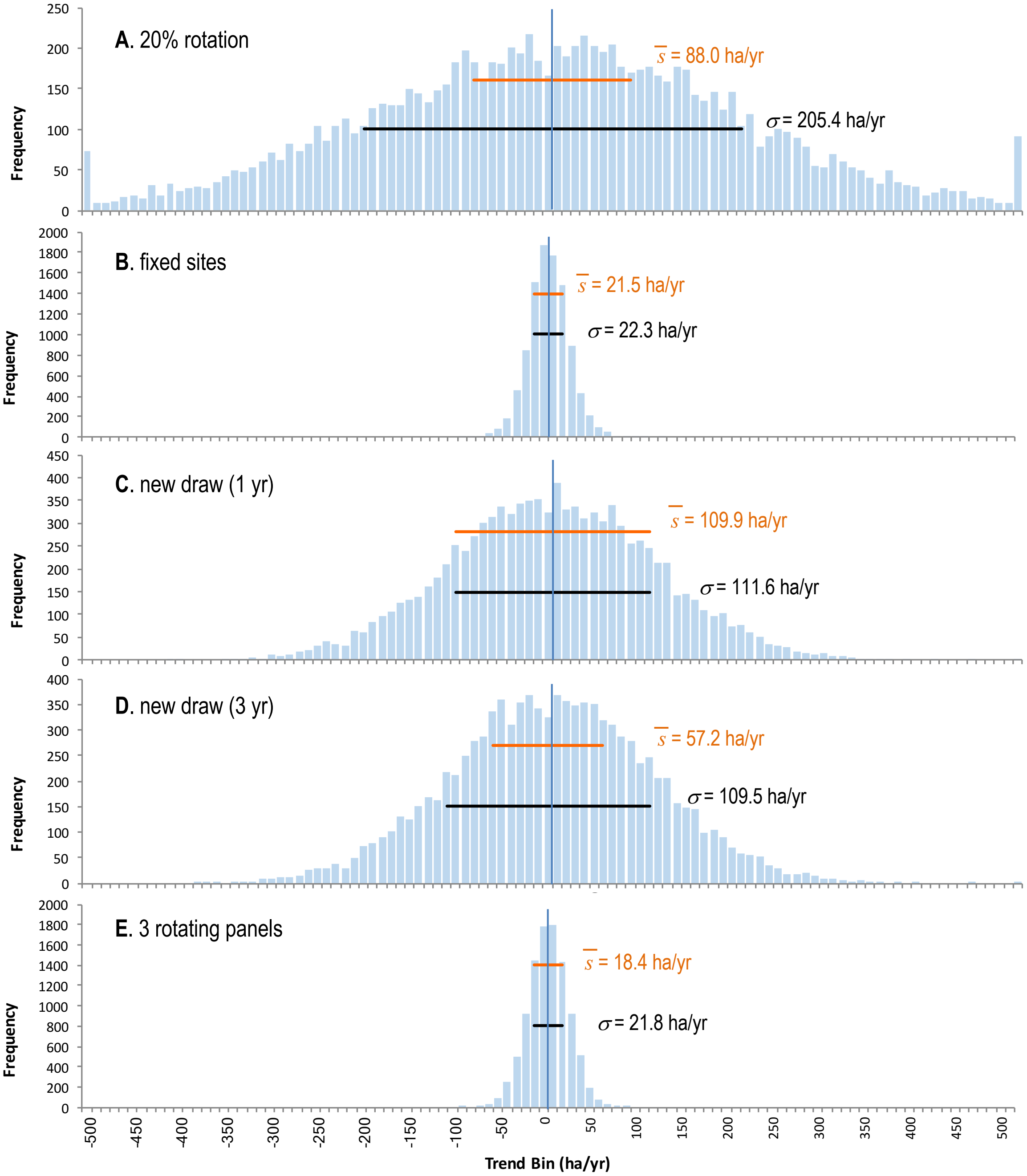


Figure 4-8. Frequency histograms of 20-year trend estimates based on annual sampling from a static soundwide seagrass opulation model for the five alternative designs (Figure 4-2, p.35). Histograms summarize results from 10,000 simulations. Horizontal lines indicate the magnitude of the true standard deviation of the trend sampling distribution (σ) and the mean estimated value ($\bar{s}$). Differences between the lengths of these lines indicate bias in the estimation of trend slope standard deviation.

The accuracy of the trend standard error measurements is indicated by the agreement between the horizontal black lines (true standard error) and the horizontal red lines (mean standard error estimate) in Figure 4-8. Discrepancies in these lines represent bias in trend standard error estimation. This bias is also shown graphically in Figure 4-9 and as 'bias in s.d. estimate' in Table 4-1. The five designs can be placed into three categories on the basis of this bias:

low trend standard error accuracy (high bias): 20% rotation and new draw (3yr)
moderate trend standard error accuracy: 3 rotating panels
high trend standard error accuracy (low bias): fixed sites and new draw (1yr)

Table 4-1. Precision and accuracy results for the five alternative designs based on 20-year trends when sampling from a static population (true trend slope is zero). The departures of the mean slope values from zero reflect negligible levels of bias in the trend slope estimates. The true s.d. values are the standard deviations of the trend slope sampling distributions (Figure 4-8) – these represent the true precision of the trend slope estimates. The mean s.d. estimates are the mean of the standard deviation estimates. The last column contains the relative bias in the standard deviation estimates.

design	mean slope (ha/yr)	true s.d. (ha/yr)	mean s.d. estimate (ha/yr)	bias in s.d. estimate
20% rotation	1.82	205.4	88.0	-57%
fixed	0.19	22.3	21.5	-3%
new (1 yr)	0.22	111.6	109.9	-1%
new (3 yr)	0.09	109.5	57.2	-48%
3 rotating panels	-0.24	21.8	18.4	-16%

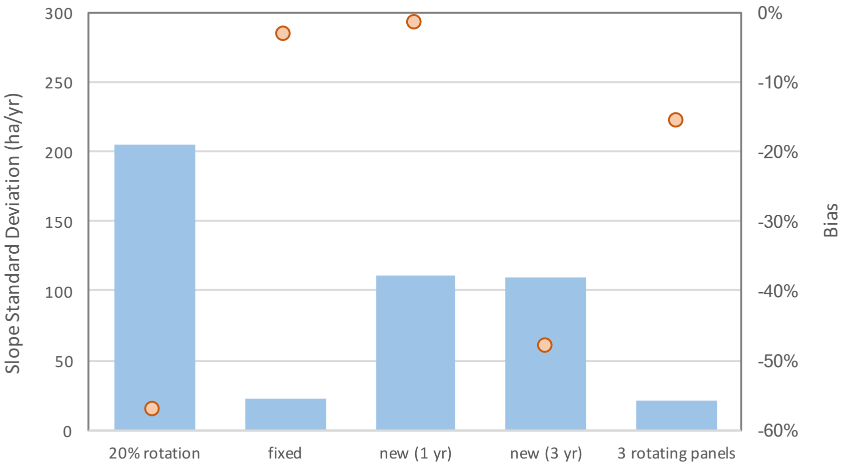


Figure 4-9. The precision of the 20-year trend slope estimates for the five designs as measured by the standard deviation of the sampling distributions when sampling from a static population (blue bars). The relative bias in the trend standard error estimates is shown by the orange points (secondary y-axis).

4.5 Type I Error

Type I error in this context is the rate of falsely concluding the seagrass population has a trend based on a statistically significant sample test, bearing in mind that this is based on the use of linear regression estimators. The Type I error varies widely across the five designs when sampling over 20-years from a static population and testing for significant trend with a *t*-test and $\alpha = 0.05$ (Figure 4-10).

The pattern of Type I error across the designs closely follows the pattern of bias in the trend standard error estimates (points in Figure 4-9). In fact, these two performance measures of the designs are directly related. In that sense, the measures of Type I error show a transform of the bias onto a different scale. On the scale of Type I error, the fixed sites and new draw (1yr) designs have Type I error that closely adheres to the α level used in testing (0.05). There is excessive Type I error with the 3 rotating panels, the new draw (3yr) and the 20% rotation designs. This indicates that trend significance tests based on regression statistics (slope and standard error) should not be used with these designs to avoid an excessive rate of false positives (significant trends).

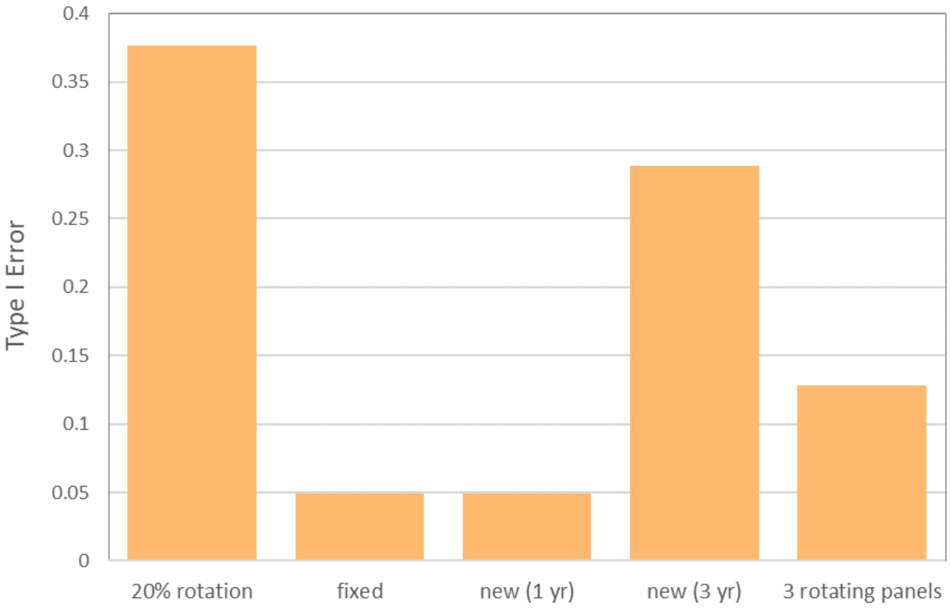


Figure 4-10. Type I error rates when testing for significant 20-year trend with $\alpha = 0.05$ when sampling from a static population. These tests rely on sample estimates of trend slope and standard error from linear regression.

4.6 Power

The power achieved with the ten ‘increasing soundwide loss’ scenarios shows a clear separation among the five designs (Figure 4-11). The 3 rotating panels and the fixed sites designs perform much better than the new draw (1yr), new draw (3yr) and 20% rotation designs.

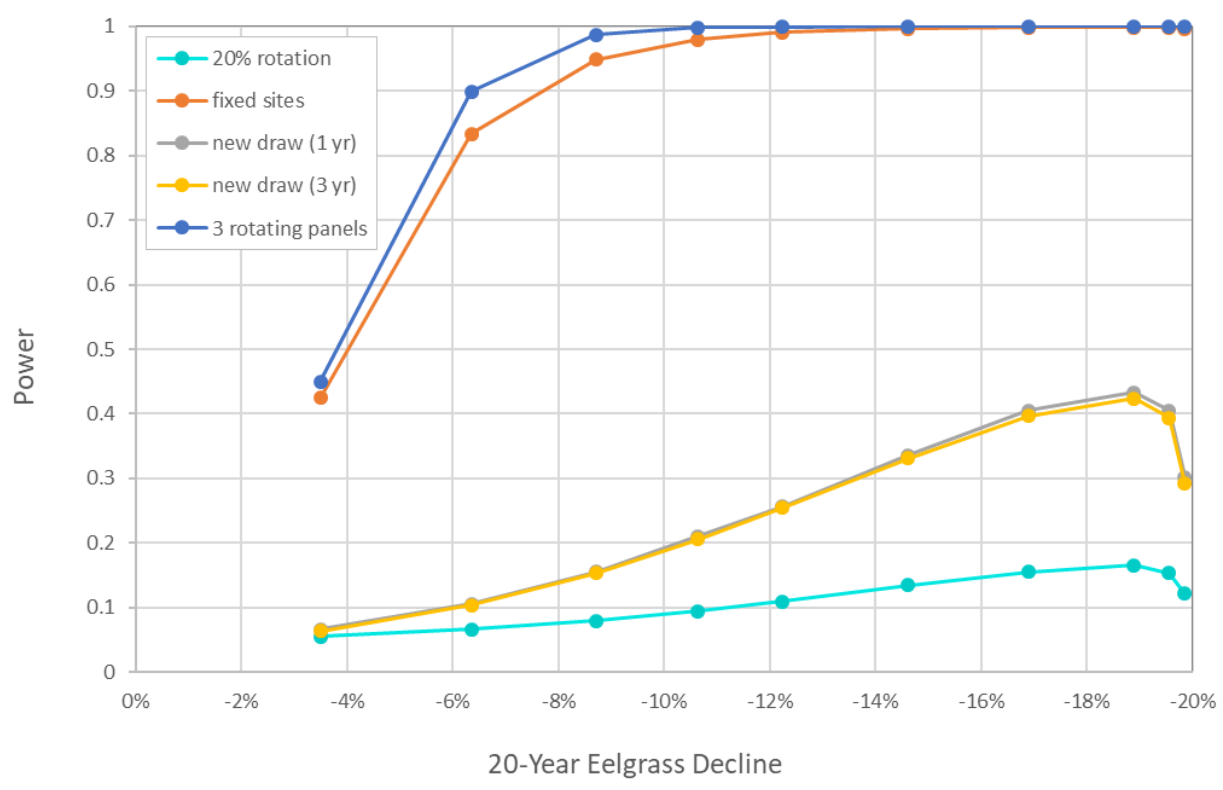


Figure 4-11. Power of each of the five designs to detect the trends in the ten ‘increasing soundwide loss’ scenarios. Each curve represents the observed power for a particular design and each point on the curve represents results from a single change scenario. Power was determined as the proportion of the 10,000 trend tests that were significant based on testing the 10,000 simulated 20-year datasets. These results do not rely on the slope standard error estimate from linear regression which was shown to be subject to bias in some cases.

Other key points from Figure 4-11:

- The power for all designs generally increases as the magnitude of the 20-year decline increases, although the shapes of the response curves are very different.
- There is an unexpected decline in power with the most extreme declines for the new draw (1yr), new draw (3yr) and 20% rotation designs. This is attributable to the strong departures from linearity in these change scenarios (see Figure 4-4, p.40).
- The 3 rotating panels and fixed sites designs reach a power of 0.8 with a 20-year decline of about -6%. The other designs never even reach a power of 0.5.
- A comparison of the new draw (1yr) and new draw (3yr) curves suggest that the effect of pooling three consecutive samples by itself is negligible.

- The 3 rotating panels design had the best results with power slightly exceeding that of the fixed sites design.

The power results from the ‘increasing soundwide loss’ scenarios were used to screen the designs. Subsequent power analyses with the other scenarios only focused on the two best-performing designs: the 3 rotating panels and fixed sites designs.

The power results with the ‘increasing size class’ scenarios display the expected decline in power as the number of sites exhibiting decline diminishes (Figure 4-12). The power of the 3 rotating panel design is more resilient as the footprint of decline in the population diminishes. This design maintains a power of 0.9 even when the number of declining sites in the population model is 130. The power of the fixed sites design is reduced to 0.53 under this same scenario.

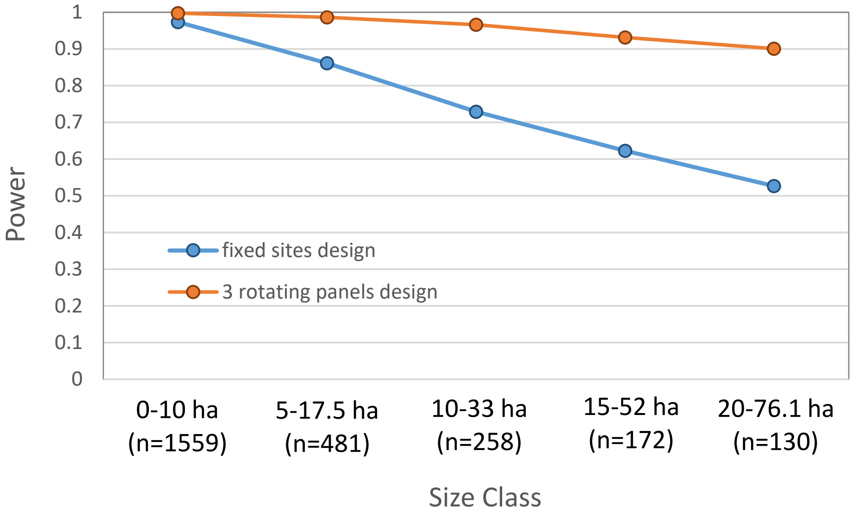


Figure 4-12. Power of the fixed sites and 3 rotating panels designs to detect trends in the ‘increasing size class’ scenarios. Each curve represents the observed power for a particular design and each point on the curve represents results from a single change scenario. Power was determined as the proportion of the 10,000 trend tests that were significant based on testing the 10,000 simulated 20-year datasets.

The power results under the ‘decreasing prevalence’ scenarios reflect a similar drop in power to detect trend as the numbers of sites declining is reduced (Figure 4-13). The drop in the numbers of sites declining is not as strong as with the ‘increasing size class’ scenarios and the drop in power is correspondingly not as dramatic. Again, the power of the 3 rotating panels design shows greater resilience to a reduction in the numbers of declining sites as compared to the fixed sites design.

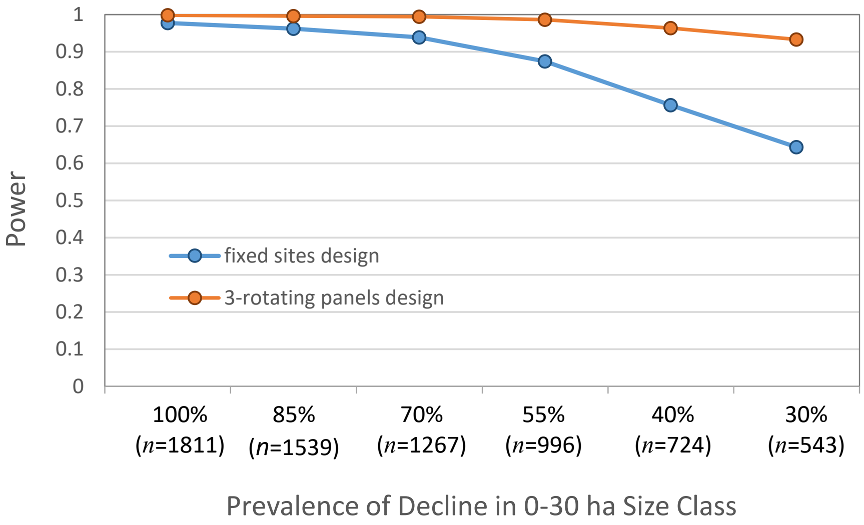


Figure 4-13. Power of the fixed sites and 3 rotating panels designs to detect trends in the 'decreasing prevalence' scenarios. Each curve represents the observed power for a particular design and each point on the curve represents results from a single change scenario. Power was determined as the proportion of the 10,000 trend tests that were significant based on testing the 10,000 simulated 20-year datasets.

5 Discussion

This study was motivated by concern about the sample replacement policy in the regional sampling design used by the SVMP. A 20% sample replacement policy was a component of the design from the second year of monitoring in 2001 until it was ultimately changed in 2015. The 20% sample replacement appeared to degrade the reliability of estimates by making them vulnerable to the effects of anomalous sites rotating into or out of the sample. This effect was noticed as early as 2001 but concerns culminated after the 2010 results were first available and the 2000-2010 trend in soundwide eelgrass area was statistically significant. The earlier concerns led to skepticism about the true significance of the trend.

This skepticism was substantiated first by exploratory analysis and then validated by the work of Van Sickle (2012) and Gast (2012a, 2012b). Van Sickle (2012) introduced an alternative trend significance test (permutation test) that eliminated reliance on untenable assumptions and confirmed that the observed SVMP trend was not significant.

While the permutation test provided for more reliable analysis of soundwide trends in eelgrass area, it did not address the question of whether an alternative sample replacement policy would improve the SVMP trend detection capability. This is the information that was needed to indicate whether changes to the sampling design were warranted. This issue led to the model development reported in Dowty (2017) and the modelling studies summarized in this report.

The first component of the modelling work confirmed that the observed SVMP soundwide trend was not significant. This corroborated the results with the permutation test. In addition, different rates of sample replacement were investigated to determine sensitivity of trend estimates to the replacement rate. The key results from this were as follows:

1. Precision of trend estimates is degraded with partial sample replacement relative to either retaining the same sample (0% replacement) or drawing new samples (100% replacement) on each sampling occasion.
2. The 20% rotation rate gave near the lowest precision of the rates along the 0 – 100% interval for the SVMP record that was analyzed. It also gave near the worst accuracy in the estimate of precision (i.e., the estimate of slope standard error).
3. A comparison of trend precision based on analysis of a 20-year data record and a 40-year data record suggests that the worst rate of rotation for a data record of *Y* years, in terms of precision, is that rotation rate that results in a site retention period that is approximately *Y*/2.

The second component of the modelling work compared the performance of five designs which included 20% rotation and four alternative designs. These comparisons were based on the trend slope and slope standard error estimates from linear regression. Since the assumptions for linear regression are not met under each design (namely, independence of the annual seagrass area estimates used in regression), these comparisons are intended to gain insight as to the behavior of the design elements, but not necessarily compare the designs in an operational setting where analysis methods would be modified to avoid poor performance of the linear regression estimator for slope standard error.

Results based on sampling from a static, unchanging population over 20 years showed separation of the designs in terms of the precision of trend slope estimates and the accuracy of the estimates of trend slope standard error (Figure 5-1). The two designs with the most precise trend slope estimates were fixed sites and 3 rotating panels. The two designs with the most accurate trend slope standard error estimates were fixed sites and new draw (1yr) sites.

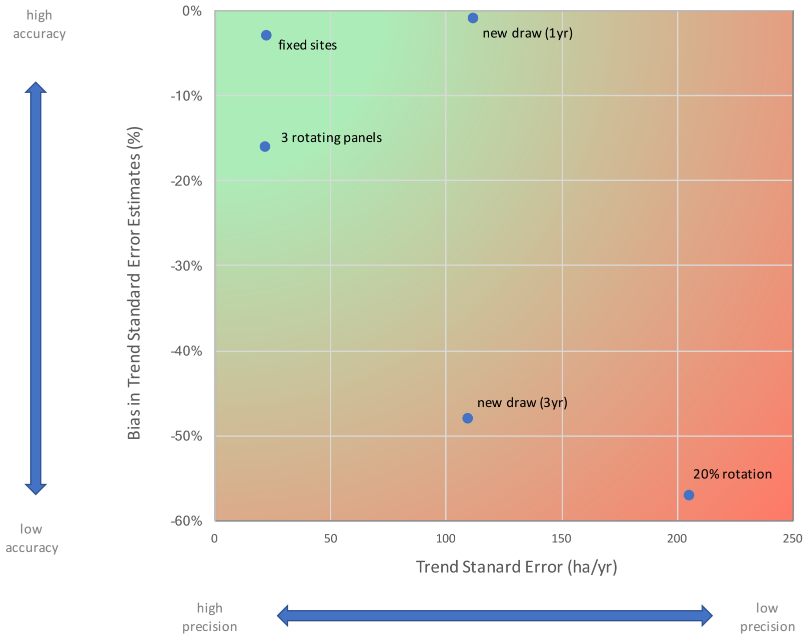


Figure 5-1. The five alternative designs based on true standard error of the trend slope estimate (*x*-axis) and bias in the trend standard error estimates from the linear regression estimator (*y*-axis). Note that this comparison of design performance is limited by the performance of the linear regression estimators.

The most important comparison of the five designs was based on power to detect trend. Power was assessed for each design in a way that utilized the trend precision derived from Monte Carlo simulations. This approach avoids issues with bias in the slope standard error estimator from linear regression. The results represent the design performance that would be expected in an operational setting that incorporated estimation from Monte Carlo simulations. Alternatively, the results represent the relative design performance that could potentially be achieved with other appropriate methods (e.g., permutation test, mixed effects model).

Power was assessed for different levels of 20-year decline spread broadly over the population. There was strong separation in these results with the best performance from the fixed sites and 3 rotational panels designs (Figure 4-11, p.46). The other three designs were eliminated from a further set of power analyses. These subsequent analyses assessed power for a fixed level of 20-year decline but with this change distributed differently within the population, but generally restricted to a smaller footprint within the population. In these scenarios, the 3 rotating panels had the best performance (Figure 4-12 and Figure 4-13, pp.47-48). This result influenced the decision for the SVMP to shift to a 3 rotating panels design in 2015.

An important limitation of this study is that, outside of the assessment of the initial SVMP design (Chapter 2), soundwide trend performance was evaluated in isolation. A second objective of the SVMP is to assess site-level trends at the sites monitored. An important benefit to the 3 rotating panels design over the fixed sites design is that the footprint across sites in the population is three times as large thereby expanding the pool of site-level results. On the other hand, the fixed sites design gives annual data at the site scale but the 3 rotating panels design only gives data every third year which will diminish trend detection capability at the site scale relative to a fixed sites design. The trade-offs here were not explored in this study.

Another important limitation was that only selected elements of the sampling design were selected for study, namely the site revisit policy and the associated estimators. There are other categories of design (Table 1-1, p.5), each with nested sub-elements, that were not investigated.

5.1 Future work

There are several additional outstanding methodological questions that may be addressed in the future to further improve the SVMP. Findings from investigations of these questions would either clarify aspects of the performance of the monitoring program or lead to further optimization of the monitoring sampling design.

- What is the power achieved to detect trend when the test for significance is based on:
    - a permutation test?
    - a linear mixed effects model?
- Following the work presented in this report, the site revisit policy was changed from 20% site rotation to 3 rotating panels. What is the power achieved over this segmented monitoring record, i.e. 2000-2014 with 20% rotation and 2015 forward with 3 rotating panels?
- How is power affected when the change scenario is a pulse or step perturbation?

- How does power of trend detection compare to soundwide indicators based on the categorical frequency within a site change classification (e.g., numbers of increasing/decreasing/stable sites)?
- How does power to detect trend compare to power to detect change between two intervals such as a 2000-2008 baseline and a 2017-2019 current mean?
- Is the current survey effort allocation across strata optimal?
- Is the current allocation of effort to site surveys optimal, e.g., as opposed to surveying more sites at lower intensity, or fewer sites at higher intensity?

6 References

Dowty, P., L. Ferrier, B. Christiaen, J. Gaeckle and H. Berry. 2022. *Submerged Vegetation Monitoring Program: 2000-2020 Geospatial Database User Manual*. Nearshore Habitat Program. Washington Department of Natural Resources, Olympia WA. https://www.dnr.wa.gov/publications/aqr_nrsh_svmp_databse_user_manual.pdf

Dowty, P. 2017. *A Model for Assessing Alternative Sampling Designs for Monitoring Puget Sound Eelgrass*. Nearshore Habitat Program. Washington Department of Natural Resources. Olympia WA.

Dowty, P., B. Christiaen and H. Berry. 2017. *Assessment of Sample Selection and Replacement for Eelgrass Monitoring*. Nearshore Habitat Program. Washington Department of Natural Resources. Olympia WA. https://www.dnr.wa.gov/publications/aqr_nrsh_site_samp_assess.pdf

Dowty, P. 2005. *A Study of Sampling and Analysis Methods: Submerged Vegetation Monitoring Project at Year 4*. Nearshore Habitat Program. Washington Department of Natural Resources. Olympia WA. http://www.dnr.wa.gov/Publications/aqr_nrsh_samp_analysis.pdf

Gast, C.M. 2012a. *An Estimate of Trend in Puget Sound Eelgrass from 2000 to 2011 Based on a Linear Mixed-Effects Model*. Report to DNR under contract 12-266. Nearshore Habitat Program. Washington Department of Natural Resources. Olympia WA.

Gast, C.M. 2012b. *Results from a Monte Carlo Study in Investigate Capability of Model-Based Estimates of Trend in Puget Sound Eelgrass*. Report to DNR under contract 12-266. Nearshore Habitat Program. Washington Department of Natural Resources. Olympia WA.

McDonald, T.L. 2003. Review of environmental monitoring methods: Survey designs. *Environmental Monitoring and Assessment*. **85**:277-292.

R Core Team. 2017. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. URL https://www.R-project.org/.

Rao, J.N.K. and J.E. Graham. 1964. Rotation designs for sampling on repeated occasions. *Journal of the American Statistical Association*. 59:492-509.

Skalski, J.R. 2003. Statistical Framework for Monitoring Zostera marina (Eelgrass) Area in Puget Sound. Appendix L in Berry, H.D., A.T. Sewell, S. Wyllie-Echeverria, B.R. Reeves, T.F. Mumford, Jr., J.R. Skalski, R.C. Zimmerman and J. Archer. 2003. *Puget Sound Submerged Vegetation Monitoring Project: 2000-2002 Monitoring Report*. Washington Department of Natural Resources. Olympia WA.

Skalski, J.R. 1990. A Design for Long-Term Status and Trends Monitoring. *Journal of Environmental Management*. 30:139-144.

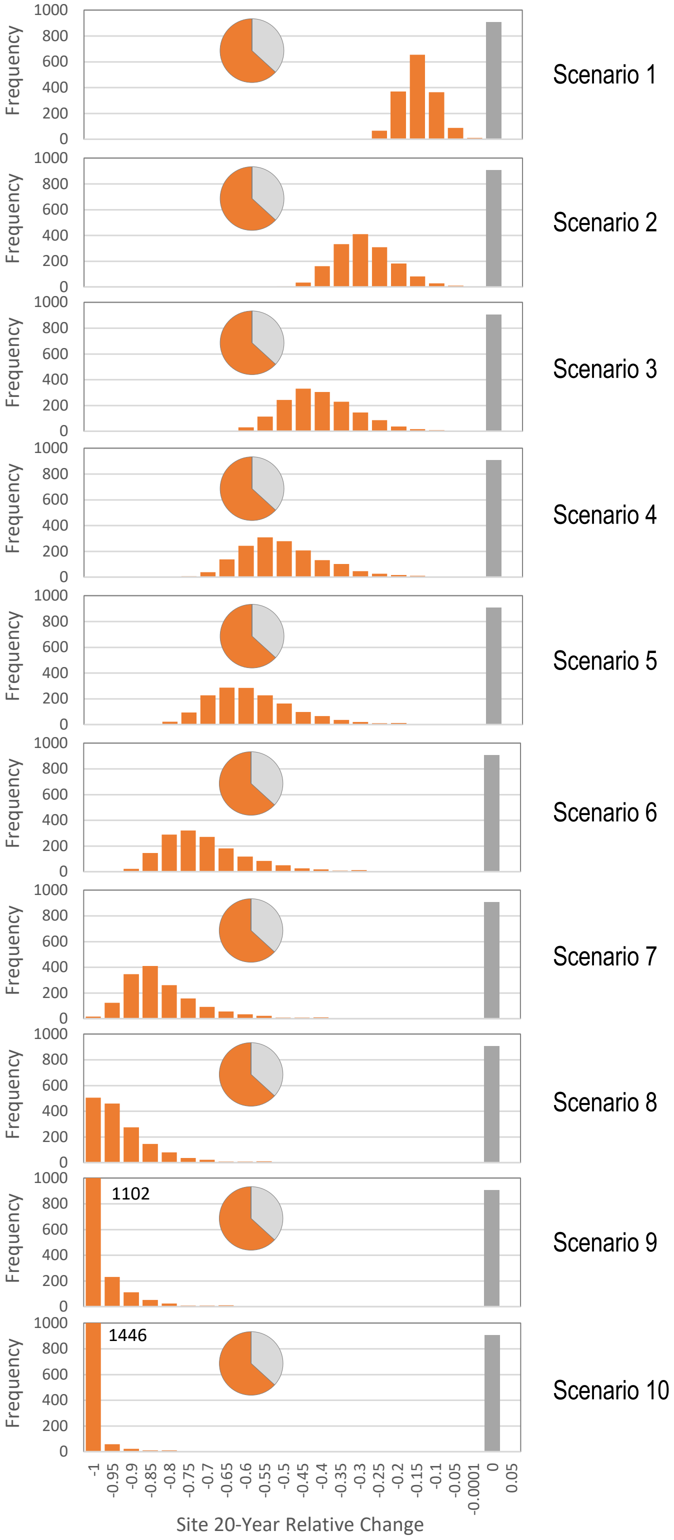
Van Sickle, J. 2012. *Testing for a trend in total eelgrass area within Puget Sound*. Final report to DNR under contract 12-281. Nearshore Habitat Program. Washington Department of Natural Resources. Olympia WA.

Zuur, A.F., E.N. Ieno, N. Walker and A. Saveliev. 2009. *Mixed Effects Models and Extensions in Ecology with R*, 1st edition. Springer.

Appendix A Change Scenarios

Additional details of the 20-year change scenarios presented in section 4.3 (p.39) are presented in a series of graphs below.

Graphs for the ‘increasing soundwide loss’ scenarios are presented in **Error! Reference source not found.**.
Graphs for the ‘increasing size class’ scenarios are presented in Figure A-2.
Graphs for the ‘decreasing prevalence’ scenarios are presented in Figure A-3.



All site losses occur in the 0-10 ha size class. The pie charts show the prevalence of decline across the entire population of sites with eelgrass (63%) which does not vary across these scenarios. The gray bar indicates the frequency of sites with no change in eelgrass area – this includes sites outside the 0 – 10 ha size class and unvegetated sites within the 0 -10 ha size class. Bars that extend beyond the *y*-axis scale in scenarios 9 and 10 are labelled with the actual frequency.

Figure A-1. Distribution of 20-year site decline across the entire population (*n*=2,467 sites) for each of the 10 'increasing soundwide loss' scenarios. The pie charts show the prevalence of decline across the population of sites with eelgrass. The gray bar indicates the frequency of sites with no change in eelgrass area.

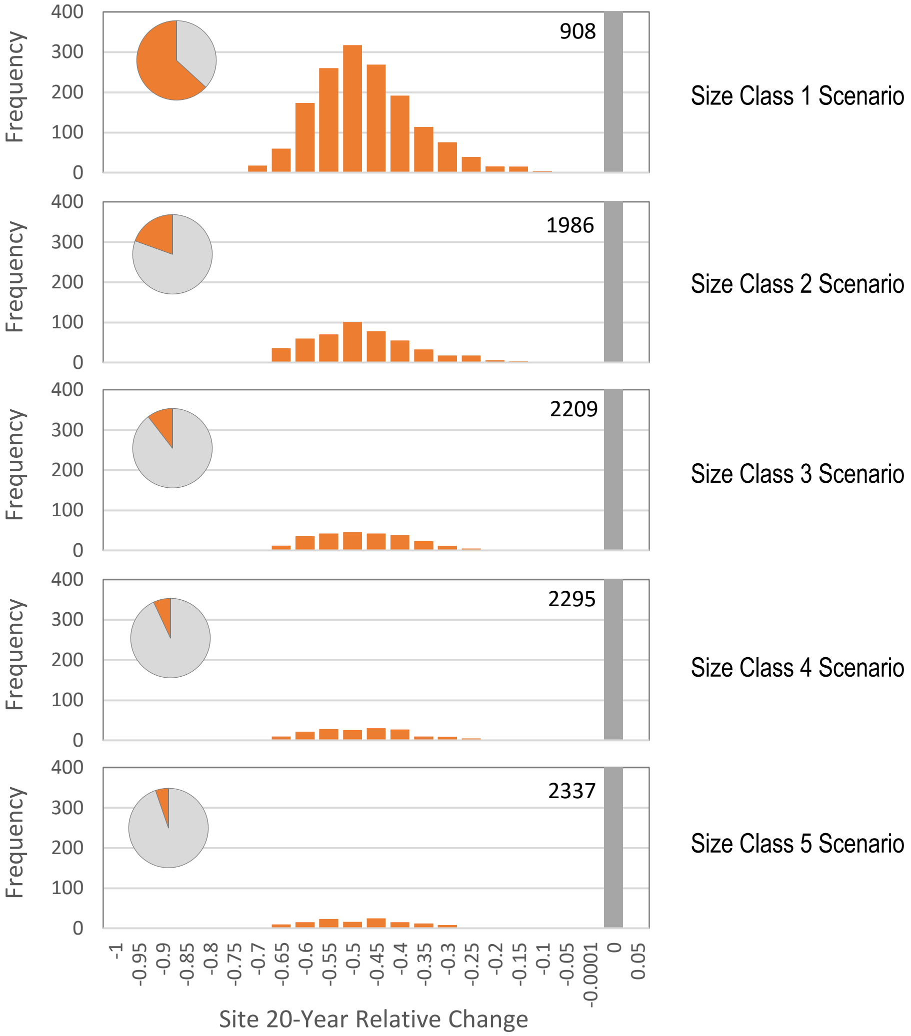


Figure A-2. Distribution of 20-year site decline across the entire population (*n*=2,467 sites) for each of the five 'increasing size class' scenarios. Each scenario represents the same area of seagrass loss, but with different distributions within the population. The pie charts show the prevalence of decline across the population of sites with seagrass. The gray bar indicates the frequency of sites with no change in seagrass area. Each gray bar extends beyond the *y*-axis scale are labelled with the actual frequency.

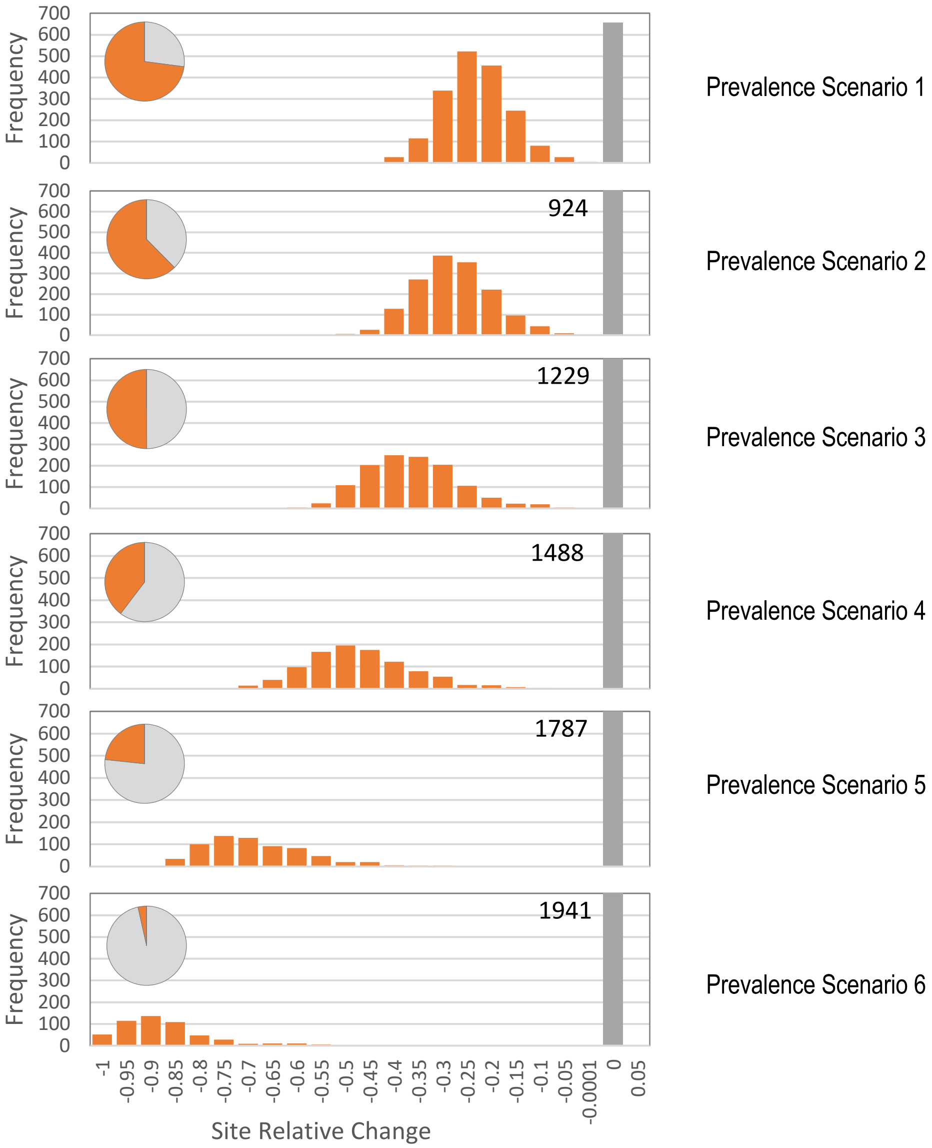


Figure A-3. Distribution of 20-year site decline across the entire population (*n*=2,467 sites) for each of the six 'decreasing prevalence' scenarios that differ by prevalence and intensity of decline within the 0-30 ha size class. Each scenario represents the same area of seagrass loss, but with different distributions within the population. The pie charts show the prevalence of decline across the entire population of sites with eelgrass. The gray bar indicates the frequency of sites with no change in eelgrass area. Bars that extend beyond the *y*-axis scale are labelled with the actual frequency.