

A Model for Assessing Alternative Sampling Designs For Monitoring Puget Sound Eelgrass

June 2, 2017



PUGET SOUND ECOSYSTEM
MONITORING PROGRAM



WASHINGTON STATE DEPT OF
**NATURAL
RESOURCES**

A Model for Assessing Alternative Sampling Designs For Monitoring Puget Sound Eelgrass

June 2, 2017

Pete Dowty
Nearshore Habitat Program
Aquatic Resources Division



PUGET SOUND ECOSYSTEM
MONITORING PROGRAM



WASHINGTON STATE DEPT OF
**NATURAL
RESOURCES**

Acknowledgements

The Nearshore Habitat Program is part of the 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 and the Puget Sound Partnership as one component of the Puget Sound Ecosystem Monitoring Program (PSEMP).

This work benefited from many discussions with others in the Nearshore Habitat Program including Helen Berry, Bart Christiaen, Jeff Gaeckle and Lisa Ferrier.

We are grateful to Victor Karpovich of DNR's Information Technology Division who made a special effort to provide access to a Redhat Linux machine that was used for the work reported here.

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:

<http://www.dnr.wa.gov/programs-and-services/aquatics/aquatic-science/nearshore-habitat-eelgrass-monitoring>

Contents

<i>Acknowledgements</i>	ii
Introduction	1
<i>The Submerged Vegetation Monitoring Program (SVMP)</i>	1
<i>Purpose of the Model</i>	2
<i>Modelling Approach</i>	2
<i>Distribution of Source Code</i>	3
An Eelgrass Population Model for Puget Sound	5
1 Initial Site Eelgrass Area Values	6
1.1 Core and Persistent Flats	6
1.2 Rotational Flats	6
1.3 Narrow Fringe	8
1.4 Wide Fringe.....	11
2 Initial Transect Sets	14
3 Twenty-Year Change Scenarios	18
3.1 Size Classes.....	18
3.2 Prevalence and Variability of Change across Sites in a Sub-Population	18
3.3 Prevalence and Variability of Change within Sites	20
3.4 Example Change Scenarios	20
3.4.1 Scenarios with contrasting amounts of eelgrass loss	20
3.4.2 Scenarios affecting contrasting size classes	23
3.4.3 Scenarios with contrasting prevalence of change	24
Sampling from the Model Population	27
4 Sampling Approach	28
5 Panel Revisit Designs	29
5.1 The Revisit Description String.....	29
5.2 Panel Groups	30
5.3 Start up plans.....	31
5.4 Split-panel designs	32
5.5 Limitations on allowed panel designs	32
5.6 The SVMP Revisit Design and Alternatives	33
6 Panel Membership Designs (Site Selection)	35
6.1 Random Selection of Eligible Sites	35
6.2 Site Selection Eligibility in Rotating Designs [x-n]	35
7 Transect Selection	37
7.1 Simple Random Selection	37
7.2 Stratified Random and Systematic Selection.....	38

Analysis of Sample Data.....	41
8 Conceptual Framework	42
8.1 Data Streams from Field Sampling	42
8.2 Analysis Pathways	44
8.3 Analyzing for Linear Trend.....	44
9 Analysis Methods	47
9.1 Soundwide Time Series from Annual Abundance	47
9.1.1 Based on Annual Samples (Original SVMP Design).....	47
9.1.2 Based on 3-Year Samples	47
9.2 Soundwide Time Series from Annual Change	48
9.2.1 Change Relative to Previous Occasion.....	48
9.2.2 Change Relative to Initial Occasion	49
9.3 Site Trend Scaled to Soundwide Trend	49
Model Implementation in a Computer Application.....	51
10 Guide for Users	52
10.1 Specification of the Population	52
10.2 Hard Coded Parameters.....	52
10.3 Parameter Files.....	53
10.4 Input Data Files	58
10.5 Output Files.....	59
10.6 Execution	61
11 Guide for Developers	63
11.1 Application Structure	63
11.2 Build Management.....	63
11.3 Potential Future Development.....	65
References	66
Appendix A Frequency Distribution Functions.....	68
Appendix B Application Data Structures.....	69
B.1 **schedule	70
B.2 **selected_sites[]	71
B.3 **selected_trans[]	72
B.4 site_scenarios[][]	73
B.5 sites_all[]	74
B.6 transect_sets[]	75



Introduction

The Washington State Department of Natural Resources (DNR) is steward of 2.6 million acres of state-owned aquatic land. The Aquatic Resources Division of DNR manages these aquatic lands for the benefit of current and future citizens of Washington State.

The Submerged Vegetation Monitoring Program (SVMP)

As part of its stewardship responsibilities, DNR conducts annual monitoring of the Puget Sound eelgrass population through the Submerged Vegetation Monitoring Program (SVMP). This monitoring is the basis for the eelgrass vital sign tracked by the Puget Sound Partnership as part of its efforts for the ecosystem recovery of Puget Sound.

The purpose of the monitoring is to provide an ecological indicator that is a sensitive measure of environmental degradation and provides ongoing assessment of the condition of an important ecological resource. Seagrasses provide valuable habitat for many species and policies exist for their protection. Monitoring tracks the status of the population and provides an assessment of the effectiveness of these policies.

SVMP monitoring and the eelgrass vital sign actually encompass the entire native seagrass population in the marine intertidal and shallow subtidal zone of greater Puget Sound. This native seagrass population is commonly referred to as an eelgrass population because it is made up primarily of eelgrass (*Zostera marina*) but it also includes species of surfgrass (*Phyllospadix* spp.). The distinction between *Zostera marina* and *Phyllospadix* spp. is outside the scope of this report and the terms eelgrass and native seagrass are used interchangeably.

It is challenging to track a plant population that is dispersed over 2400 miles (3800 km) of shoreline, that is submerged, and that reaches underwater depths beyond the reach of airborne remote sensing techniques. In addition, eelgrass is often intermixed with other species that require direct observation in order for them to be distinguished and excluded from the ecological indicator – primarily macroalgae and the introduced seagrass *Zostera japonica*.

There is no viable approach for direct, comprehensive observation of the entire Puget Sound eelgrass population that meets the program's species discrimination and depth range requirements. Consequently, the monitoring approach relies on sampling the population and estimating characteristics of the overall population based on samples. This approach

necessitates a sampling design which specifies exactly how locations are selected for measurement, how measurements are made and how measurements are used to make estimates regarding the eelgrass population. Sampling designs that perform poorly can lead to erroneous results and a flawed understanding of the population. Due to the challenges in sampling the Puget Sound native seagrass population, a fairly complex sampling design is being used.

Due to the complexity of the SVMP sampling design and the geographic scale of the population being monitored, it is very difficult to assess the performance of the sampling design. Without such an assessment critical questions remain unanswered. Are the results unbiased? When tests show change in the population what is the likelihood this is a spurious results? Are confidence intervals on estimates accurate? Are there refinements to the sampling design that, if made, would lead to greater accuracy and precision and greater ability to detect change in the population? Clearly these questions are critical to understanding the effectiveness of the monitoring program and warrant dedicated effort to complete an assessment.

Purpose of the Model

Given the complexity and scale of the monitoring, a modelling approach is particularly useful. The purpose of developing the model presented in this report was to provide the capability to assess the performance of the existing SVMP sampling design and to evaluate a number of alternative designs to see if they might provide significant improvements in performance. The model was completed at a time when specific alternatives were being considered for implementation by the SVMP but there were no objective means to compare the expected performance of these alternatives.

This report describes the model, including the conceptual model and its implementation in a software application. Initial studies that apply the model to answer specific research questions will be presented in a separate report.

Modelling Approach

The overall model presented in this report is an example of a Monte Carlo model. This means that the model explicitly represents an element of chance, or stochasticity, and a large number of model runs when aggregated give an exact representation of how this stochasticity affects the estimates produced.

In practice, a monitoring sample from one sampling occasion is used to estimate a population parameter, say the total area of the Puget Sound eelgrass population. But in recognition that the numerical value of the estimate reflects random effects (e.g., the random selection of specific sampling locations), the sample is also used to estimate a confidence interval that hypothetically encompasses some proportion of the estimates, commonly 95%, that would be produced if the sampling and estimation were to be repeated a large number of times. In Monte Carlo modelling, this concept is no longer hypothetical because the large number of repetitions are actually conducted and the actual

spread in the estimates is known. This is very useful for evaluating many aspects of the performance of the sample design and the estimates it produces.

While we refer here to one overall model, this model is actually a hierarchical set of self-contained components that are models themselves. The two major modelling components include are.

- Puget Sound Eelgrass Population Model – a model population that includes detailed representation of the eelgrass present along each potential transect at each site in the population. This includes a representation for the initial conditions as well as annual representations over 20 years under different models of 20-year change scenarios.
- Sampling Model – a model that simulates the actual field sampling process that includes site selection, transect selection, transect surveying and video classification.

The population model was designed to emulate the actual Puget Sound eelgrass population in terms of key characteristics that have been observed in the 2000-2013 SVMP dataset and are known to affect estimation of eelgrass area and eelgrass area trend for the population. The current SVMP sampling frames were adopted for the model (flats and fringe) (see Christiaen et al. 2016 for details). This fixed the total numbers of flats and fringe sample units (sites) in the population. In addition, the construction of the model relied on the current SVMP stratification within each sampling frame. A different construction approach was used to model each stratum.

This report presents the conceptual details of the model. The model was implemented in a computer application written in the C language in a development process that utilized the GNU C compiler and Make for build automation. Development was conducted in a number of discrete periods spread out over approximately three years which necessitated the use of different systems including a Linux system (Ubuntu 14.04), a Solaris 9 System and a Redhat Enterprise Linux 7 system. Random numbers throughout were generated with the Mersenne Twister algorithm (Matsumoto and Nishimura 1998).

Distribution of Source Code

The source code for the application developed for this study was written almost entirely by a Washington State government agency. As such, this source code is in the public domain and available upon request. Two functions used in the application are exceptions in that they were written by third parties. The code for the Mersenne Twister random number generator was written by Matsumoto and Nishimura (1998) and downloaded from the project web site¹. Initially this code was made available under the GNU Public License but after 2001 it has been distributed without restrictions² and can therefore be included with the distribution of code developed for this study.

¹ <http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/emt.html>. Viewed June 1, 2015.

² <http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/MT2002/elicence.html>. Viewed June 1, 2015.

The function that returns critical values of the Student- t statistic was incorporated from the GNU Scientific Library³ and is distributed under the GNU Public License (GPL). Since the restrictions of the GPL are not compatible with the public domain code, the code for this function will not be distributed by DNR. In the future, it would be straightforward to eliminate dependence on the GNU Scientific Library as the core algorithm has already been included in the application with original code (the acceptance-rejection method of Von Neumann [1951]) and could be readily adapted to producing critical values of the Student- t statistic.

³ <http://www.gnu.org/software/gsl/>. Viewed June 1, 2015.

An Eelgrass Population Model for Puget Sound

1 Initial Site Eelgrass Area Values

The methods used to set the initial (time = 0) eelgrass area values for each site in the Puget Sound model population are presented in this section. The methods differed by stratum.

1.1 Core and Persistent Flats

Since all nine sites in the core and persistent flats strata have been sampled, the latest (2013) estimates of eelgrass area were used to specify initial site eelgrass area for these sites.

1.2 Rotational Flats

The site area for each rotational flats site is known from the SVMP site polygon GIS baselayer. Roughly 80% of these sites have been sampled in the 2000-2011 period: 53 out of 67 sites = 79% sampled. For the 53 sites sampled, the latest eelgrass area estimate was used to specify the initial site eelgrass area.

For the remaining 14 sites, an eelgrass area was assigned by first specifying a site R value – the ratio of site eelgrass area to site area. This follows the approach first used in Dowty (2005, p.20) and updated in Dowty (2006, p.23). First, each unsampled site greater than 425 ha in size was categorized as having a high, medium or low value of R – essentially a measure of eelgrass cover within the site boundary – based on field experience and expert opinion. Sites smaller than 425 ha were simply placed in the medium R category. Then, nominal values of R were set for each category based on reference to the distribution of R values from the sampled sites (Figure 1-1). The nominal values of R selected were 0.1 (low), 0.3 (medium) and 0.5 (high), although none of the 14 unsampled sites were placed in the high- R category. Site eelgrass area was determined by multiplying the nominal R value for the category by the site area (Figure 1-2, Table 1-1).

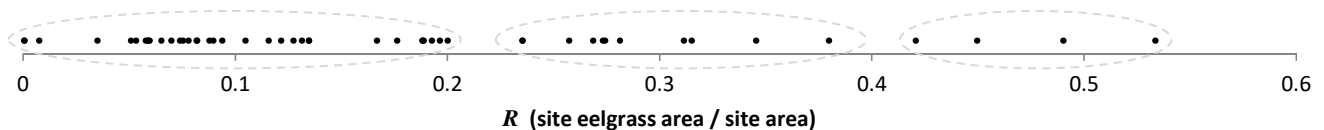


Figure 1-1. The distribution of R values for the 53 rotational flats sites sampled between 2000 and 2011. The dashed-line ovals indicate the groupings used to characterize sites with high, medium and low R values.

Table 1-1. Initial site eelgrass area and R values for the 14 rotational flats sites that had not been sampled by the SVMP at the time of this study.

site code	site name	region	site area (ha)	R	Initial site eelgrass area (ha)
flats04	Lummi Flats N	nps	731.3	0.3	219.4
flats07	Portage Bay S, Seattle	nps	517.9	0.3	155.4
flats22	Port Susan West, Stanwood	swh	2276.9	0.1	227.7
flats23	Port Susan East, Stanwood	swh	2516.0	0.1	251.6
flats27	Snohomish Delta Mid, Everett	swh	1627.5	0.1	162.7
flats29	Cornet Bay, Whidbey	swh	55.9	0.3	16.8
flats30	Cultus Bay, Whidbey	cps	445.6	0.1	44.6
flats31	Oak Harbor, Oak Harbor	swh	363.8	0.3	109.1
flats36	Eagle Harbor, Bainbridge	cps	43.0	0.3	12.9
flats38	Port Madison	cps	79.7	0.3	23.9
flats40	Miller Bay, Suquamish Reservation	cps	110.7	0.3	33.2
flats48	Sequim Bay	sjs	78.4	0.3	23.5
flats72	Boundary Bay	nps	710.3	0.1	71.0
flats74	Wycoff Shoal, McNeil Island	cps	77.4	0.3	23.2

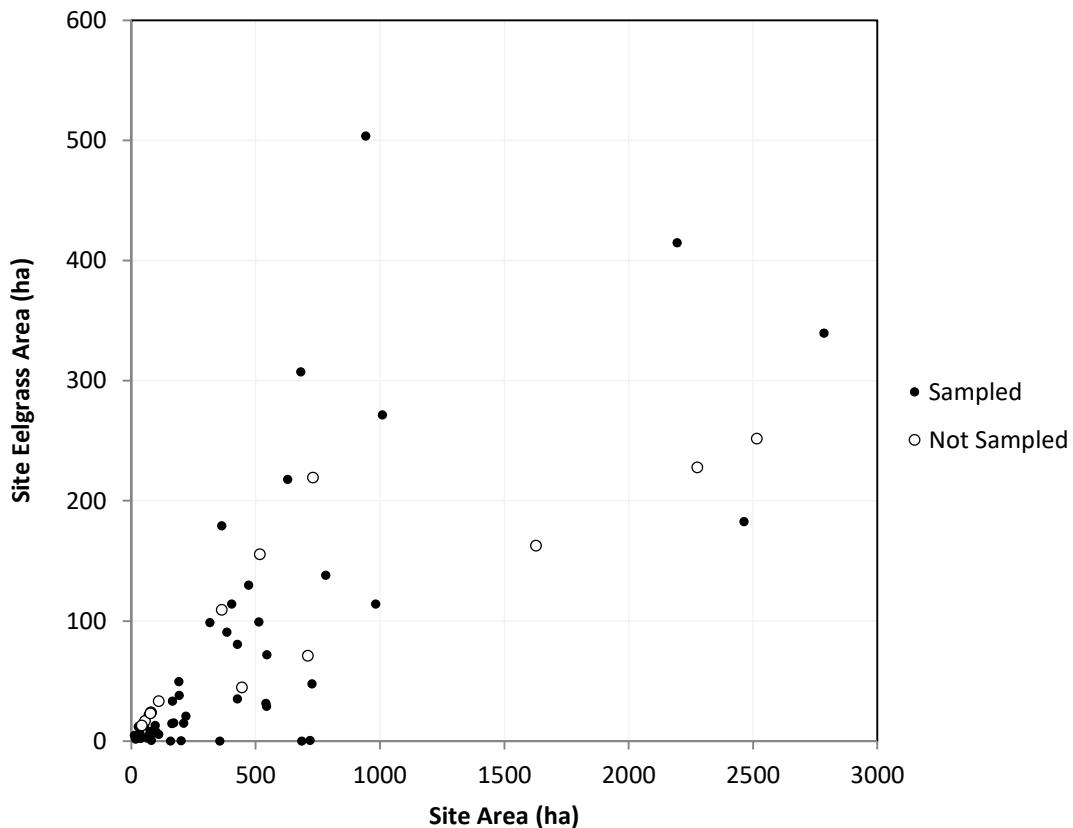


Figure 1-2. The initial values of site eelgrass area plotted against site area for all 67 sites in the rotational flats stratum. The “Sampled” points are based on sample SVMP data while the “Not Sampled” points are simulated data (Table 1-1).

1.3 *Narrow Fringe*

Only 13% of the narrow fringe sites have been sampled over the 2000-2013 period (255 out of 1965 sites). Information from the sampled sites was used to create a simulator that generated site eelgrass area values for the entire stratum (1965 sites). The simulator consisted of two components:

1. Conduct a Bernoulli trial to determine if eelgrass is present or absent for each site in the model. A Bernoulli trial can only have one of two outcomes, such as flipping a coin, but in this case the probability of the outcomes is set to match the frequency of presence/absence in the sample dataset. Of the 255 narrow fringe sites sampled by the SVMP, 51 had no eelgrass giving a probability p for a site having no eelgrass of $p = 0.20$. A Bernoulli trial was conducted for each of the 1965 narrow fringe sites by generating a uniform random number on the interval (0,1) and setting eelgrass absent when the random number fell in the interval (0, 0.20].
2. For each model site with eelgrass present, a random value was drawn from a distribution that mimicked the frequency histogram of site eelgrass area of the 204 sampled sites with eelgrass present. The narrow fringe histogram and several alternative distributions that were investigated are shown in Figure 1-3. The challenge was to find a mathematical distribution that could be parameterized to match the histogram in both the steeply sloped portion at low eelgrass areas as well as the tail at high eelgrass areas. The distributions investigated included lognormal, Weibull, exponential and Pareto. The parameters for each distribution were adjusted to match the histogram shape. This distribution was then used in the simulator to generate 1965 eelgrass area values for the stratum. Inspection of the frequency histogram for the simulated stratum led to further adjustments of the distribution parameters. Ultimately, the Weibull distribution was selected with parameters that generated the curve shown in Figure 1-3.

Random values drawn from the Weibull distribution that exceeded a specified maximum threshold of 25 ha were rejected. This prevented the selection of unrealistically large eelgrass beds (low probability events in the tail of the distribution). Also, random values that were below a minimum threshold of 0.01 ha were adjusted to the threshold value. This avoided the simulation of unrealistically small beds (e.g., fractions of square millimeters) that would have relatively high frequencies under the upward sloping frequency distribution.

The narrow fringe simulator generated eelgrass area values for the 1965 sites in the stratum. The proportion of sites with eelgrass absent was 0.24. The frequency histogram for the sites with eelgrass present is shown in Figure 1-4. The overall shape of the histogram of the simulated narrow fringe population (Figure 1-4) closely tracks that of the sample data (Figure 1-3).

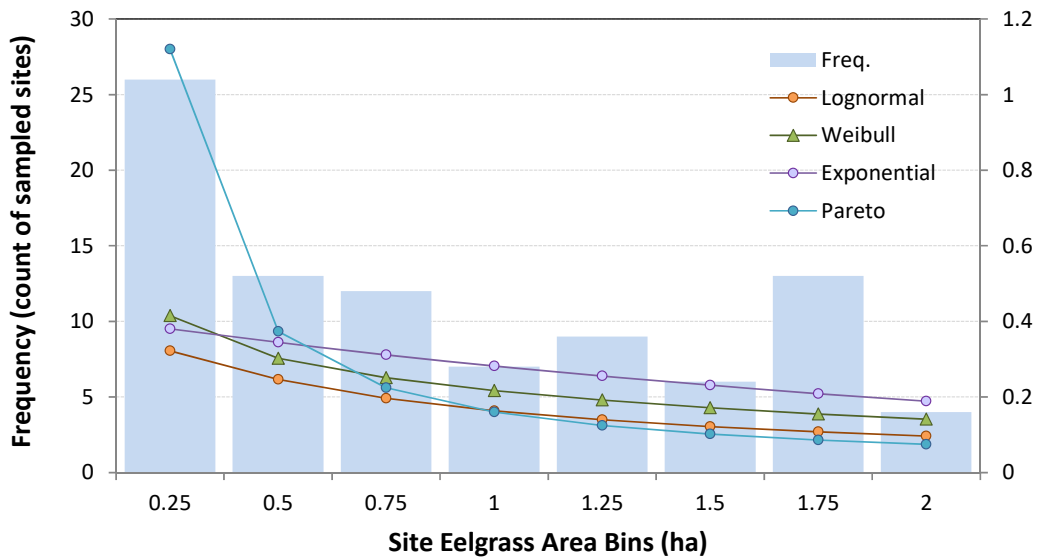
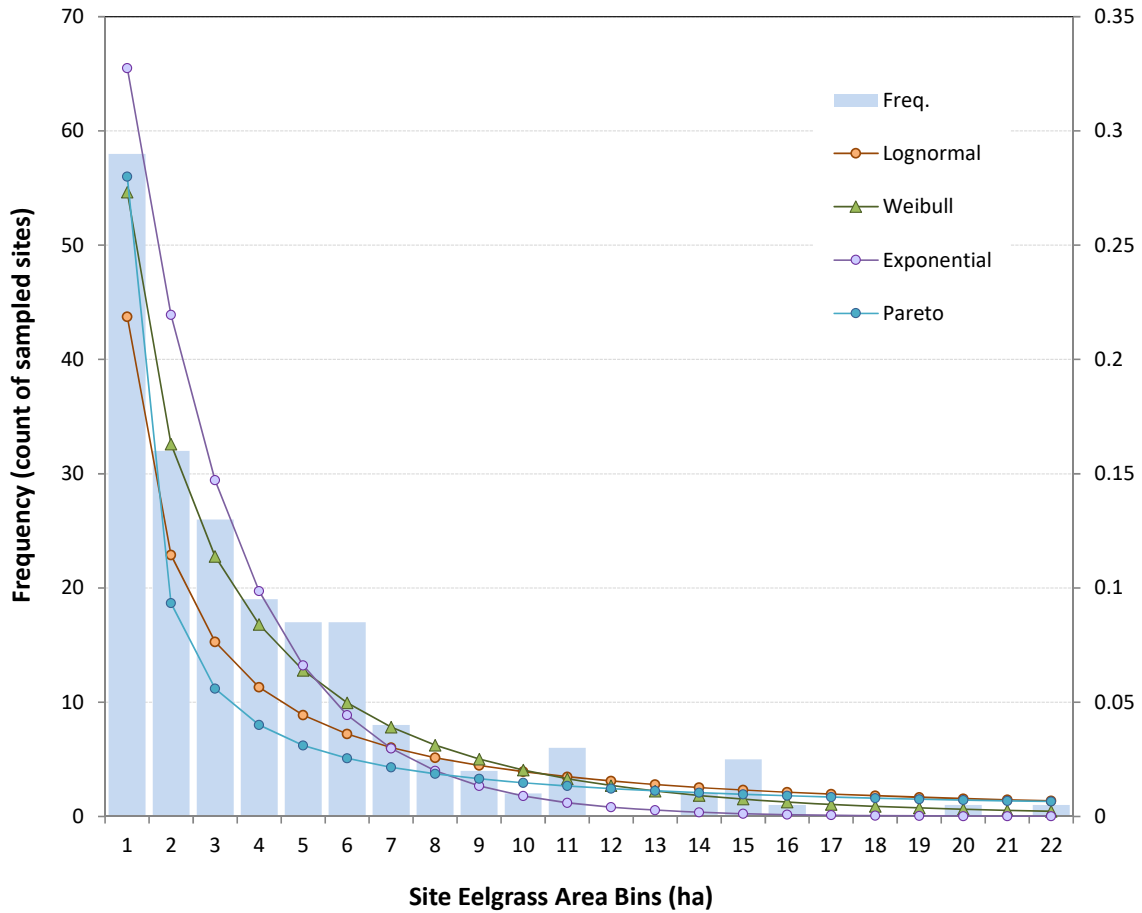


Figure 1-3. Narrow Fringe. Frequency histograms of site eelgrass area for the sampled narrow fringe sites with eelgrass present ($n = 204$). The curves represent different attempts to characterize the shape of the histogram. The Weibull distribution shown was used to construct the narrow fringe model. Top: the entire distribution; Bottom: more detail for sites with small eelgrass areas. Details of distributions are in Appendix A.

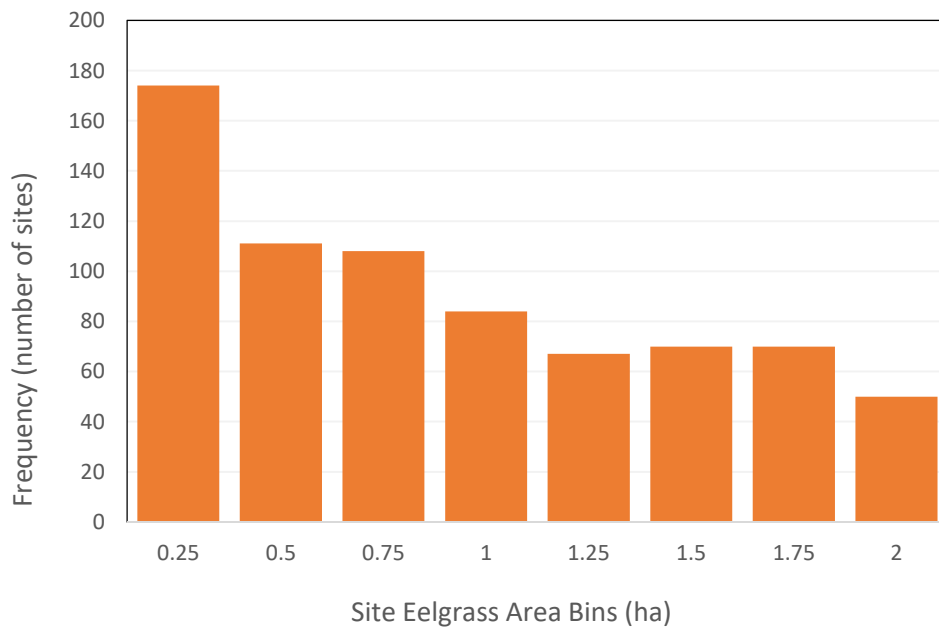
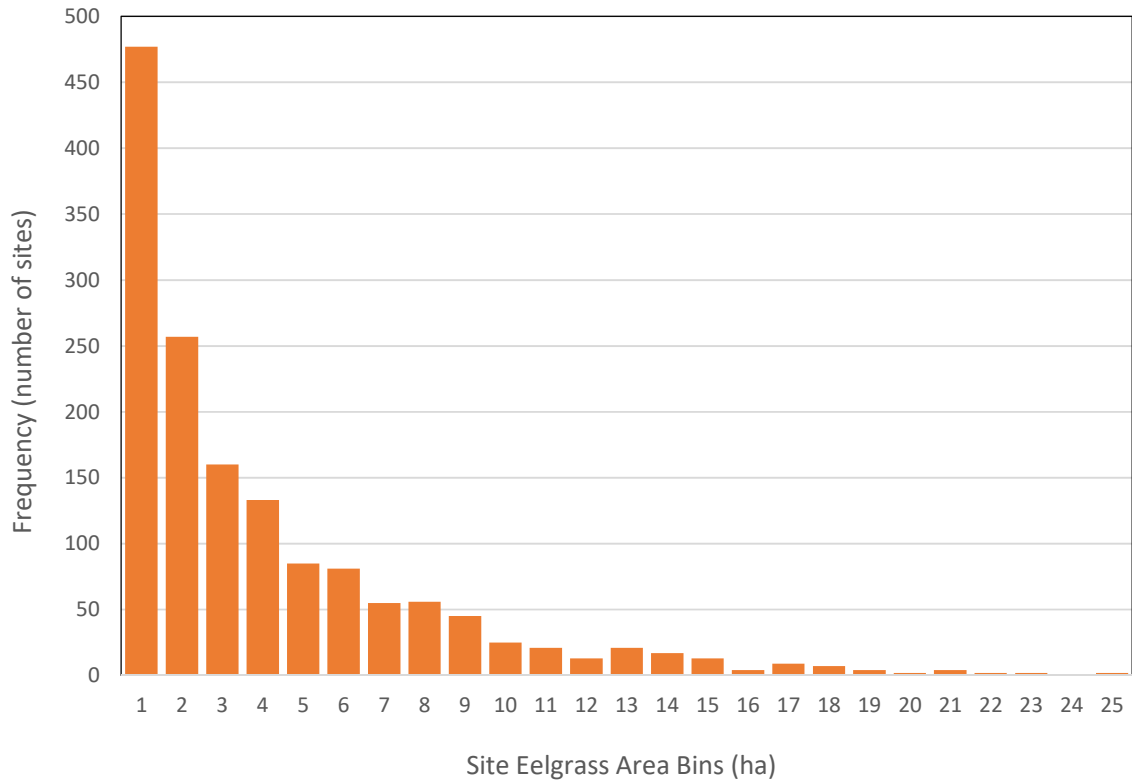


Figure 1-4. Frequency histogram of site eelgrass area simulated for the narrow fringe model ($n = 1965$). Each site eelgrass area value was randomly drawn from the Weibull distribution shown in Figure 3.

1.4 Wide Fringe

Only 16% of the wide fringe sites have been sampled over the 2000-2013 period (70 out of 426 sites). A simulator was used to generate eelgrass area values for the 426 sites in the stratum. As with the narrow fringe stratum, this simulator had two components:

1. Conduct a Bernoulli trial to determine if eelgrass is present or absent for each site in the model with a probability of $p = 0.16$ that eelgrass is absent (11 sites in the sample of 70 had eelgrass absent; 59 had eelgrass present).
2. For each model site with eelgrass present, a random value was drawn from a Weibull distribution fit to the frequency histogram of wide fringe eelgrass area values. The wide fringe histogram and several alternative distributions that were investigated are shown in Figure 1-5 (p.12).

A minimum eelgrass bed size of 0.1 ha was imposed (smaller simulated values adjusted to 0.1) and a maximum bed size of 100 ha (larger simulated values rejected).

The wide fringe simulator generated eelgrass area values for the 426 sites in the stratum. The proportion of sites with eelgrass absent was 0.22. The frequency histogram for the sites with eelgrass present is shown in Figure 1-6. The overall shape of the histogram (Figure 1-6) tracks that of the sample (Figure 1-5) fairly well but with noticeably excessive skew toward sites with small eelgrass area relative to the sample data.

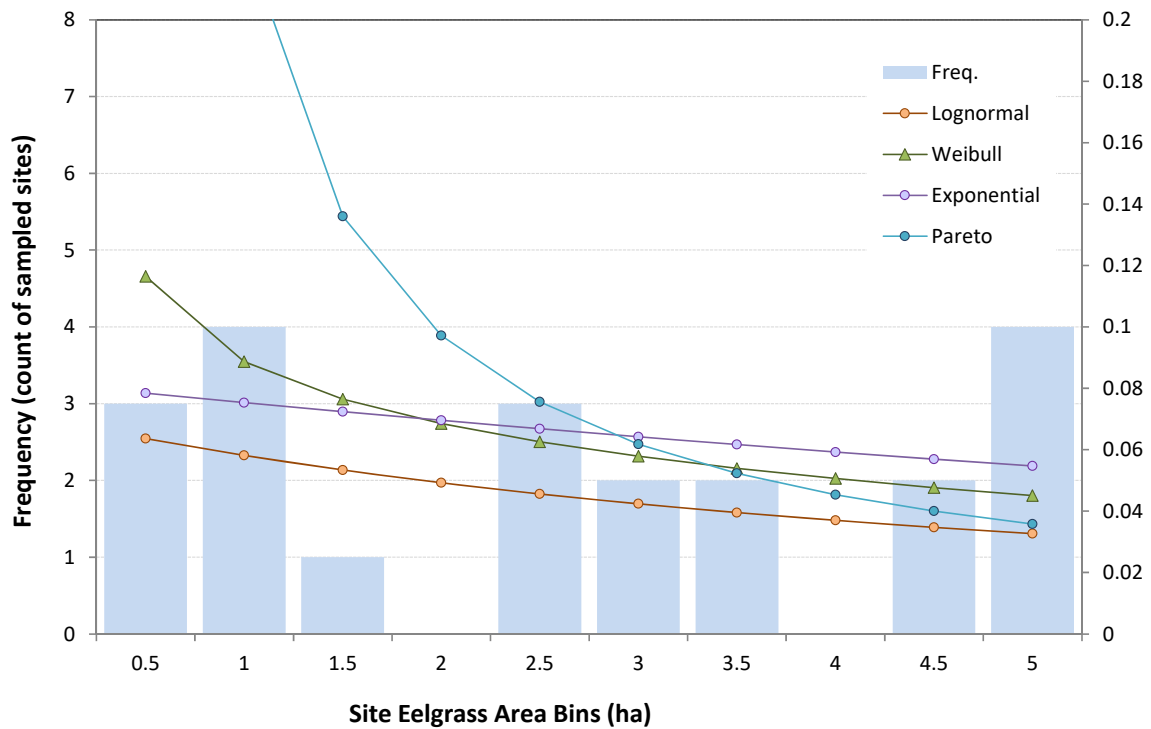
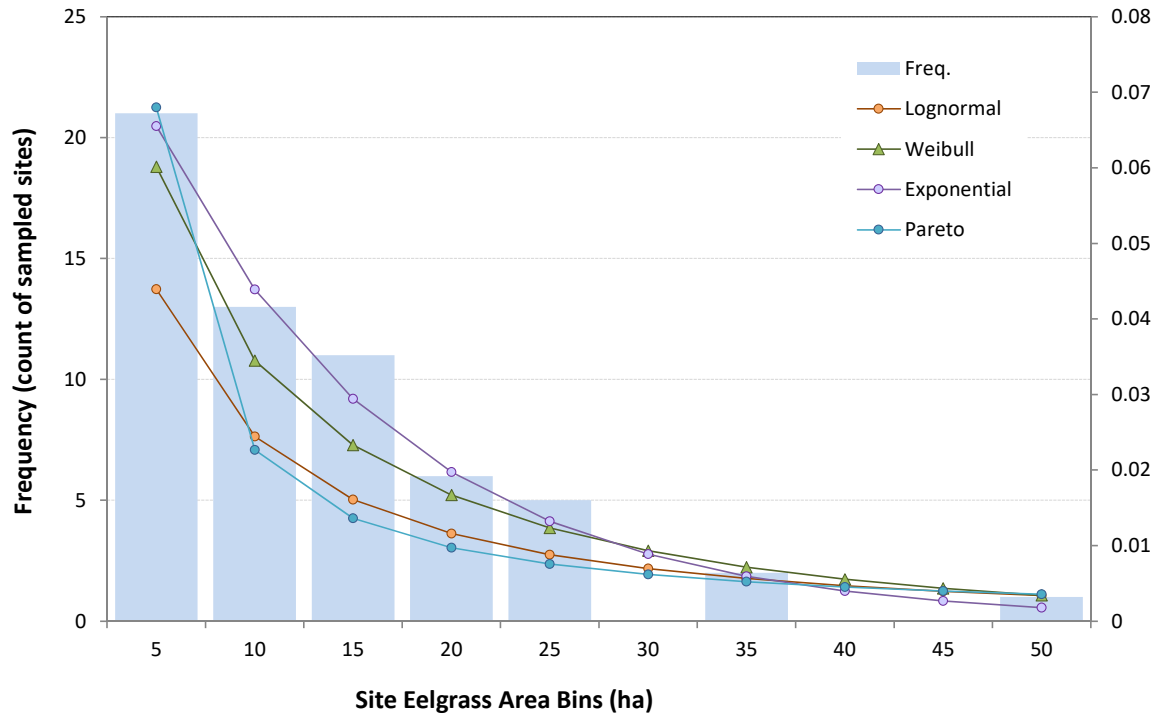


Figure 1-5. Wide Fringe. Frequency histograms of site eelgrass area for the sampled wide fringe sites with eelgrass present ($n = 59$). The curves represent different attempts to characterize the shape of the histogram. The Weibull distribution shown was used to construct the wide fringe model. Top: the entire distribution; Bottom: more detail for sites with small eelgrass areas. Details of distributions are in Appendix A.

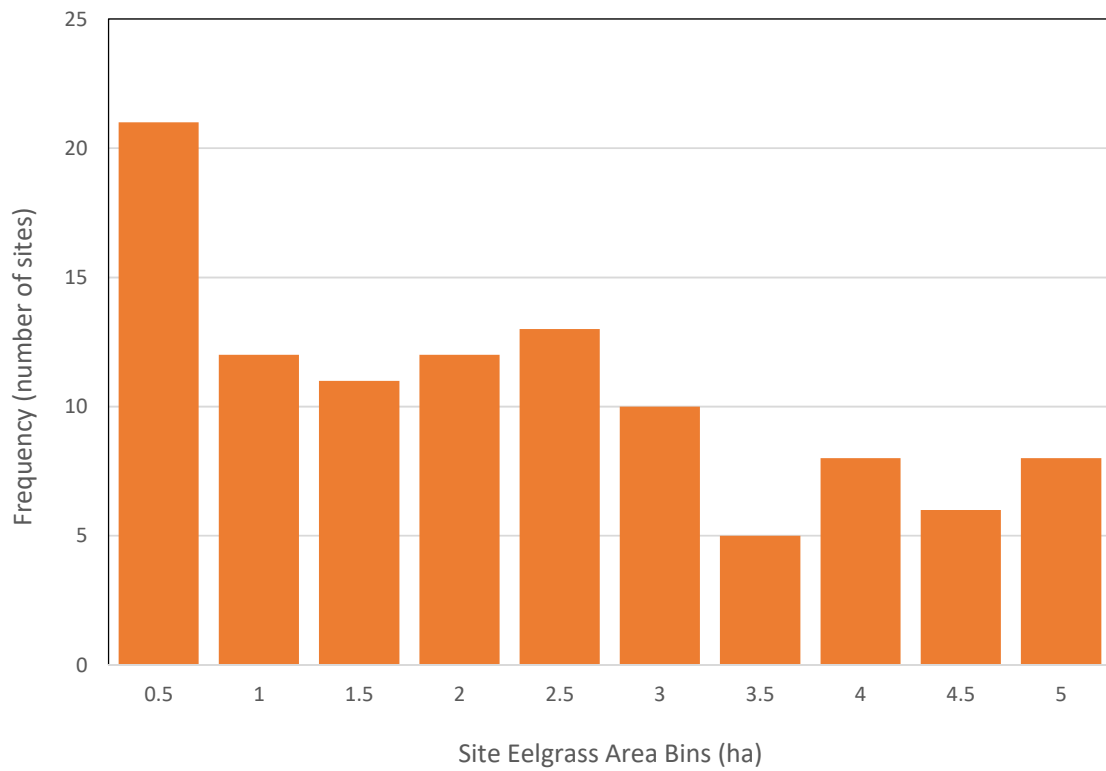
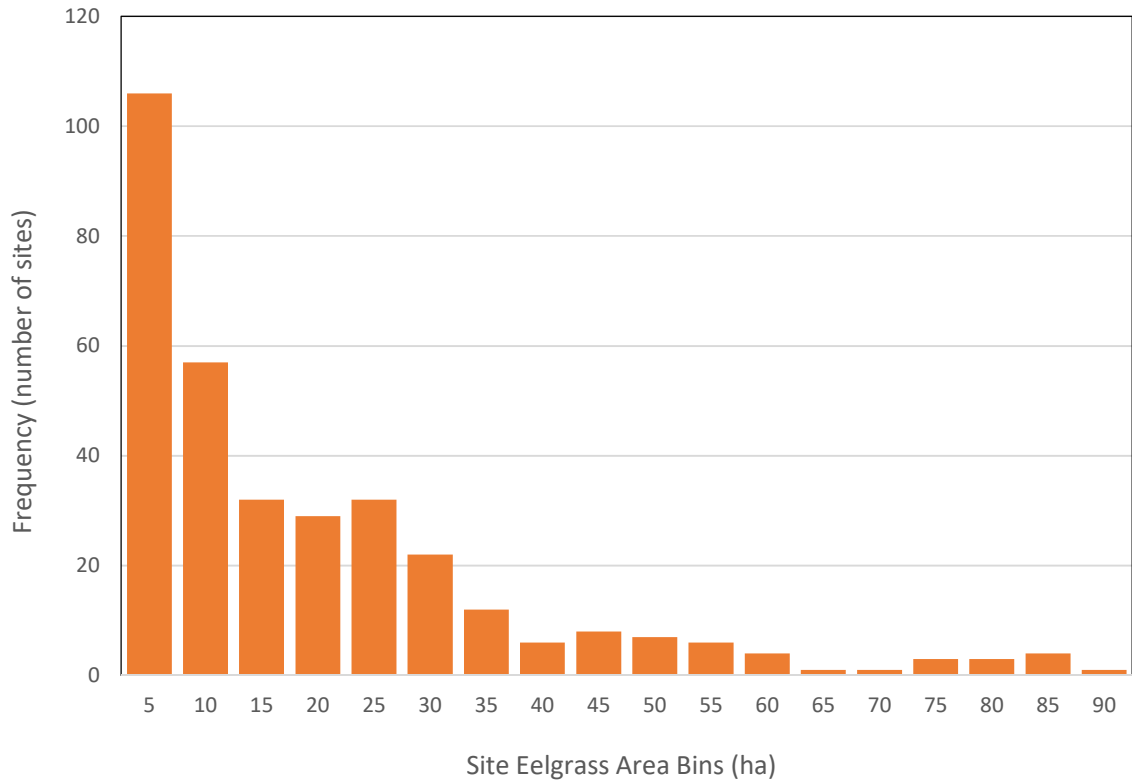


Figure 1-6. Frequency histogram of site eelgrass area for the wide fringe model ($n = 426$). Each site eelgrass area value was randomly drawn from the Weibull distribution shown in Figure 1-5.

2 Initial Transect Sets

The previous section described how each sample unit (site) within the soundwide study area was assigned an eelgrass area value in the model. In addition, transect-level detail is required for each of these sites in order for the model to meet the following two objectives:

- Support simulation of measurement error in the estimation of site eelgrass area from sample data.
- Support simulations that contrast site sampling with newly drawn random transects and sampling with repeat transects.

In order to avoid the complexity of explicitly modeling the spatial processes leading to spatial pattern in eelgrass beds and simulating transect sampling of these beds (e.g., as was done by Schultz 2008), the existing transect data in the 2000-2011 dataset were used as the basis for assigning sets of transects to each site in the model. These assigned sets specify a spatial “belt model” of eelgrass within each site as described below.

First, a pool of initial transect sets were prepared from the existing data where each set is a collection of transect data resulting from actual SVMP site sampling. For each site sampled by the SVMP, transect data collected in different years was combined to form one group of transect data per site. For sites that experienced substantial change in the eelgrass bed in the SVMP data record, this essentially shifted real temporal variability to increased spatial variability in the model transect set. Then site transect sets that contained less than 20 transects were removed. This was an arbitrary threshold set to ensure the transect set contained a diversity of transects and a sufficient number to support re-sampling during the simulated sampling. A total of 150 transect sets were removed for an insufficient number of transects and 225 remained for use in the model (Figure 2-1). These 225 transect sets originated from 40 rotational flats sites, 132 narrow fringe sites, 44 wide fringe sites and the nine sites obtained by combining core and persistent flats sites.

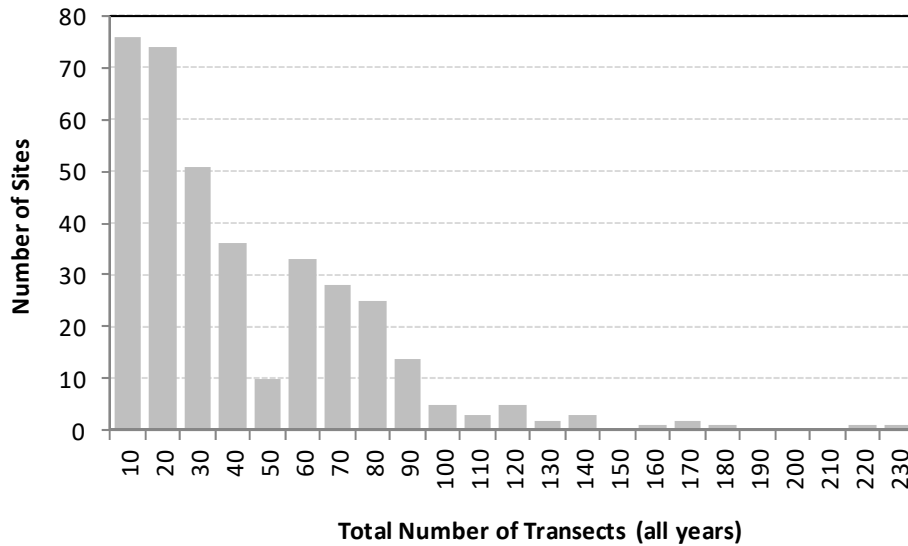


Figure 2-1. Frequency of sites sampled by the SVMP by the total number of transects surveyed across all years.

For each site in the Puget Sound eelgrass model, a transect set was randomly drawn from the available sets for the relevant stratum. The mean transect fraction for the selected set was calculated (Skalski 2003, p.3). Then, the model sample polygon area, E , was calculated as

$$E = \frac{X}{\bar{p}} \quad \text{Equation 2-1}$$

where

X = model site eelgrass area

\bar{p} = mean transect fraction for the selected transect set.

In order to avoid unrealistically large sample polygon areas, a condition was applied during the selection of transect sets so that model sites with large eelgrass areas would not be associated with low-fraction transect sets. Based on inspection of the 2000-2011 narrow fringe data (Figure 2-2), model narrow fringe sites with eelgrass areas greater than 1.5 ha could only draw from transect sets with a mean fraction greater than 0.2. Similarly, based on inspection of wide fringe data (Figure 2-3), model wide fringe sites with eelgrass area greater than 3.5 ha could only draw from transect sets with a mean fraction greater than 0.25.

The combination of the fraction values in a transect set assigned to a site together with the calculated sample polygon area (Equation 2-1) represent the initial spatial model of an eelgrass bed at the site. If the transect set contains N transects then the model bed can be thought of belts arranged side-by-side of width (W/N) meters where W is the longshore length of the site (~ 1000 meters in the case of a fringe site). If under the simulated transect

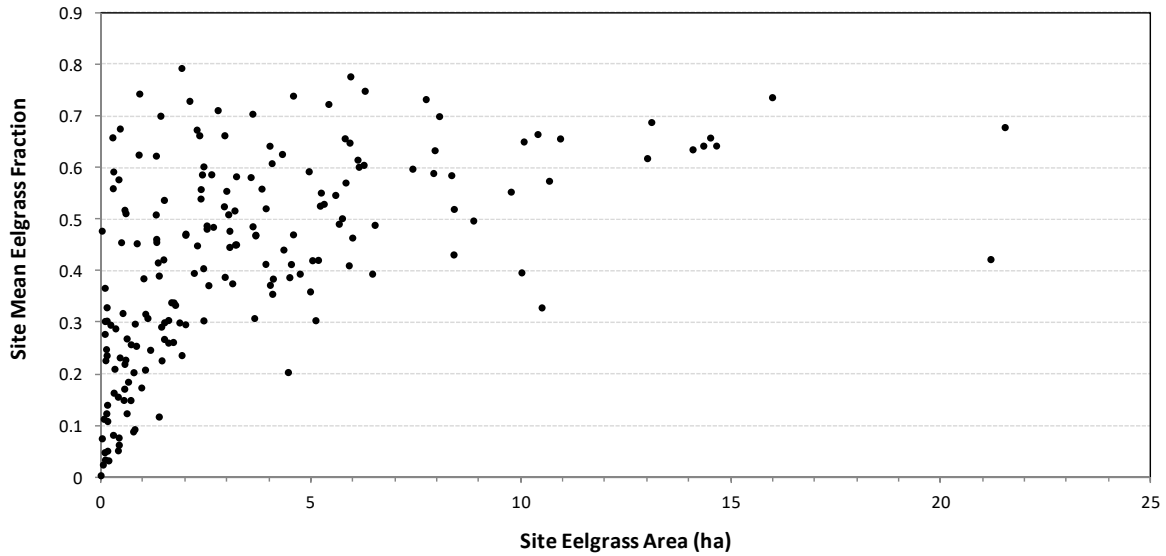


Figure 2-2. Site mean eelgrass fraction for all narrow fringe sites sampled during 2000-2011 that have eelgrass present ($n = 188$) relative to site eelgrass area.

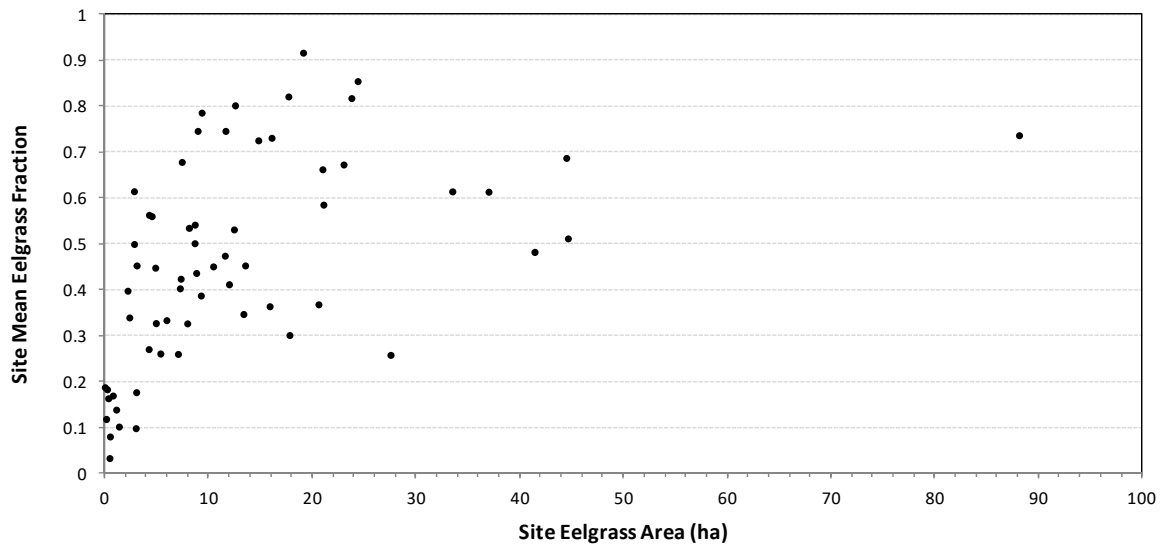


Figure 2-3. Site mean eelgrass fraction for all wide fringe sites sampled during 2000-2011 that have eelgrass present ($n = 62$) relative to site eelgrass area.

sampling a transect is drawn from a particular belt then the resultant transect fraction is simply that assigned to that belt. The cross-shore width of the site belt model is not needed and is not defined since multiple widths with different distributions of local fraction could lead to the same overall transect fraction. The transect length values from the assigned transect set are used in analysis but only for weighting the transect values. Within the model presented in this report, the transect lengths of the initial transect sets become dimensionless weighting factors. This allows a particular transect set that was derived from

an actual SVMP site and reflects the spatial characteristics of that site, to be applied across different model sites that have eelgrass beds of different dimensions.

Once the transect sets are successfully assigned to each site in the model dataset, the initial model is complete. At this point every site in the study area has an assigned initial eelgrass area, sample polygon area and a spatial site belt model based on the assigned transect set.

3 Twenty-Year Change Scenarios

The previous two sections describe how the initial state for the Puget Sound eelgrass population model is generated. This initial state can then be subjected to a scenario of eelgrass change over time to add a temporal component to the population model. This section describes the capabilities of the software application for implementing model change scenarios and the specific scenarios that were selected to generate 20-year model datasets as a demonstration of these capabilities.

3.1 Size Classes

The application allows different eelgrass change scenarios to be applied to different sub-populations within Puget Sound. Currently two sub-populations can be specified. Each site in the model may fall in one or the other sub-population based on the sites' size class – i.e., the area of eelgrass at the site in the initial model state. For example, we can specify that all sites with eelgrass area in the 0-10 ha size class are subject to a specific annual change scenario, such as stable or a specific level of increase or decrease, and all sites in the 10-4000 ha size class can be subjected to a different annual change scenario. It is not necessary that the two size classes comprehensively cover the population or that they be contiguous. If there are sites that are not within a specified size class they are not subject to change in eelgrass area over the 20-year scenario. The cumulative frequency by site area and the cumulative proportion of total eelgrass area by site area are useful in delineating size classes to meet particular study needs (Figure 3-1).

3.2 Prevalence and Variability of Change across Sites in a Sub-Population

The overall mean annual rate of change is specified independently for each sub-population. This rate is the basis for applying change in each of the 20 years in the model. The prevalence of change within a sub-population can range from the entire sub-population (all sites subjected to change) to a specified proportion of the sub-population. This prevalence can be varied to mimic a synoptic, more spatially uniform pattern of change or a pattern of localized change that affects some sites but leaves others unaffected. For each site, a Bernoulli trial with the appropriate probability is used (equal to the specified prevalence) to determine if it is subject to the change scenario.

For sites subject to change within a sub-population, the annual rate of change can be homogeneous (all sites have the same annual rate of change) or can follow a normal distribution about the specified mean for the sub-population based on a specified CV. In the latter heterogeneous case, sites subject to change within a sub-population will have

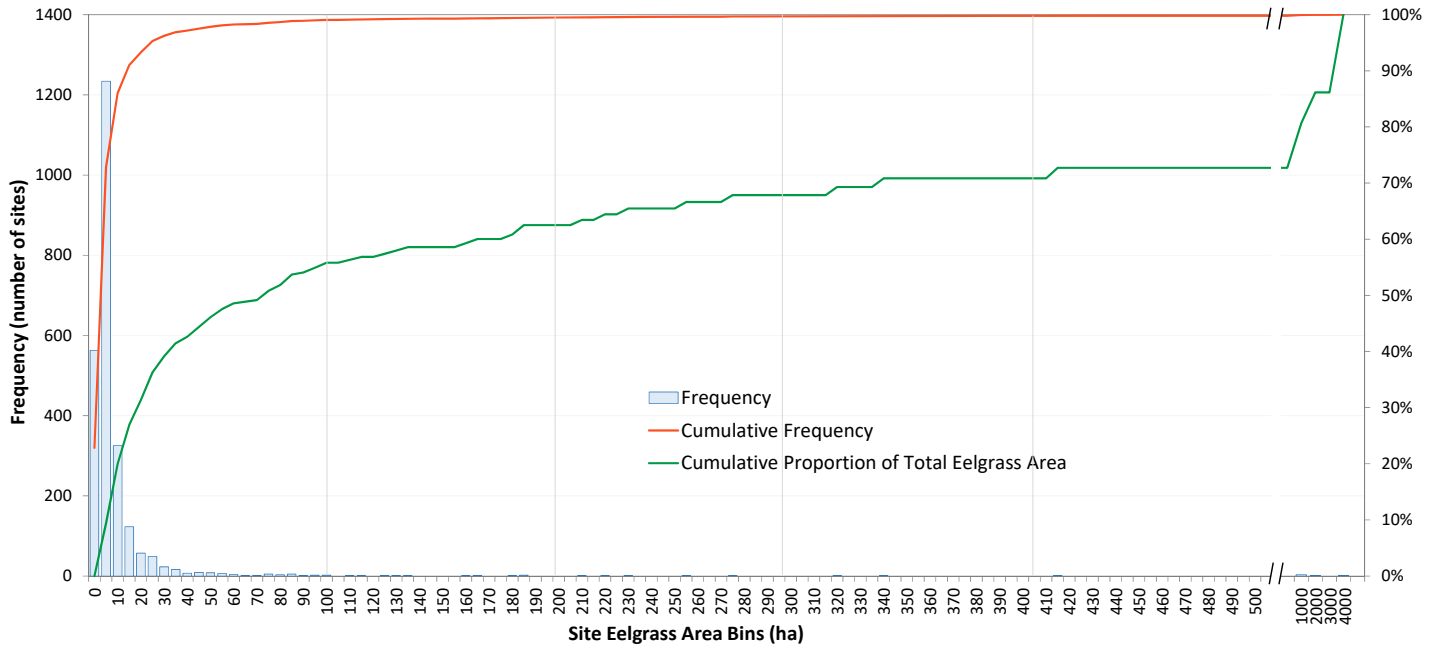


Figure 3-1. Frequency histogram of sites by eelgrass area (bars), the cumulative frequency and the cumulative proportion of total eelgrass area.

varying rates of annual change while maintaining the specified mean rate over all sites subject to change in the sub-population.

For scenarios where annual change for sites follows a normal distribution, the value of annual change for an individual site is calculated as

$$r_i = (z_i CV + 1) \bar{r} \tag{Equation 3-1}$$

where

- r_i = annual change value for site i
- z_i = standard normal random variate
- CV = coefficient of variation of distribution of site change values
- \bar{r} = mean annual site change value specified for the sub-population

Equation 3-1 follows from the definitions of the z -score and CV given by

$$z_i = \frac{(X_i - \bar{X})}{\sigma}, \quad CV = \frac{\sigma}{\bar{X}} \tag{Equation 3-2}$$

where σ is the standard deviation of the variation in the rate of mean site change across sites.

3.3 Prevalence and Variability of Change within Sites

For sites subject to change in eelgrass in the model, change is imposed at the “belt” level. The transect length is the same across all years of the model for transects selecting within a belt, but the eelgrass length (and fraction) varies over time. The prevalence of change within a site can encompass the entire site (all belts in the site model) or only a portion of the site (a subset of the belts). This allows for representation of heterogeneous change at sites. In the latter case, a Bernoulli trial with specified probability equal to the within-site prevalence is used for each belt to determine whether it is within the scope of change at the site. The prevalence is specified as an input parameter for each subpopulation.

For belts that are subject to eelgrass change at a site, the annual rate of change can be homogeneous (all belts have the same annual rate of change) or can follow a normal distribution with mean rate equal to the annual change specified for the site and CV specified as an input parameter for the sub-population. In the latter case, the rate of change for each belt is calculated as

$$b_{ij} = (z_{ij} CV + 1) \overline{X}_i \quad \text{Equation 3-3}$$

where

b_{ij} = annual change value for belt j at site i

z_{ij} = standard normal random variate

CV = coefficient of variation of distribution of change values across belts

\overline{X}_i = mean annual site change value for site i

3.4 Example Change Scenarios

It is envisioned that research studies will typically utilize the model presented in this report with a series of different change scenarios. In this way, the performance of sampling designs can be compared across a range of possible patterns of change. This section presents three different sets of change scenarios that demonstrate this approach.

3.4.1 Scenarios with contrasting amounts of eelgrass loss

As the magnitude of change increases, a sampling design would be expected to have increasing power to detect this change. Here a set of 10 change scenarios are developed reflecting different levels of decline that can be used to develop power curves for sampling designs.

These scenarios have contrasting rates of decline within the same size class. The 0-10 ha size class is used here. This size class includes approximately 86% of all sites but the sites with no eelgrass (0 ha) see no change in these scenarios. The size class includes 63% of sites with eelgrass present and 20% of total eelgrass area in the initial population model (see Figure 3-1). The prevalence of decline across sites in the size class is 100% and the prevalence of decline across belts within each site experiencing decline is 100%.

Each of the 10 scenarios has a different value of \overline{r} , the mean annual site change value specified for the sub-population. The value of r_i for an individual site i is drawn from a normal distribution centered on \overline{r} with a CV of 0.3. Values of b_{ij} for individual belts

within sites subject to change are drawn from normal distributions centered on r_i for site i with a CV of 0.3. Given these parameters, the range of values used for \bar{r} (-0.01 to -0.30) translate to total losses over 20 years for the overall Puget Sound model of -3.5% to -19.9%, respectively (Figure 3-2, Table 3-1). The same number of sites experience decline for each scenario but the mean level of site decline becomes more severe from scenario 1 to 10 (Figure 3-3). At scenario 10 almost all sites within the 0-10 ha size class have experienced total loss.

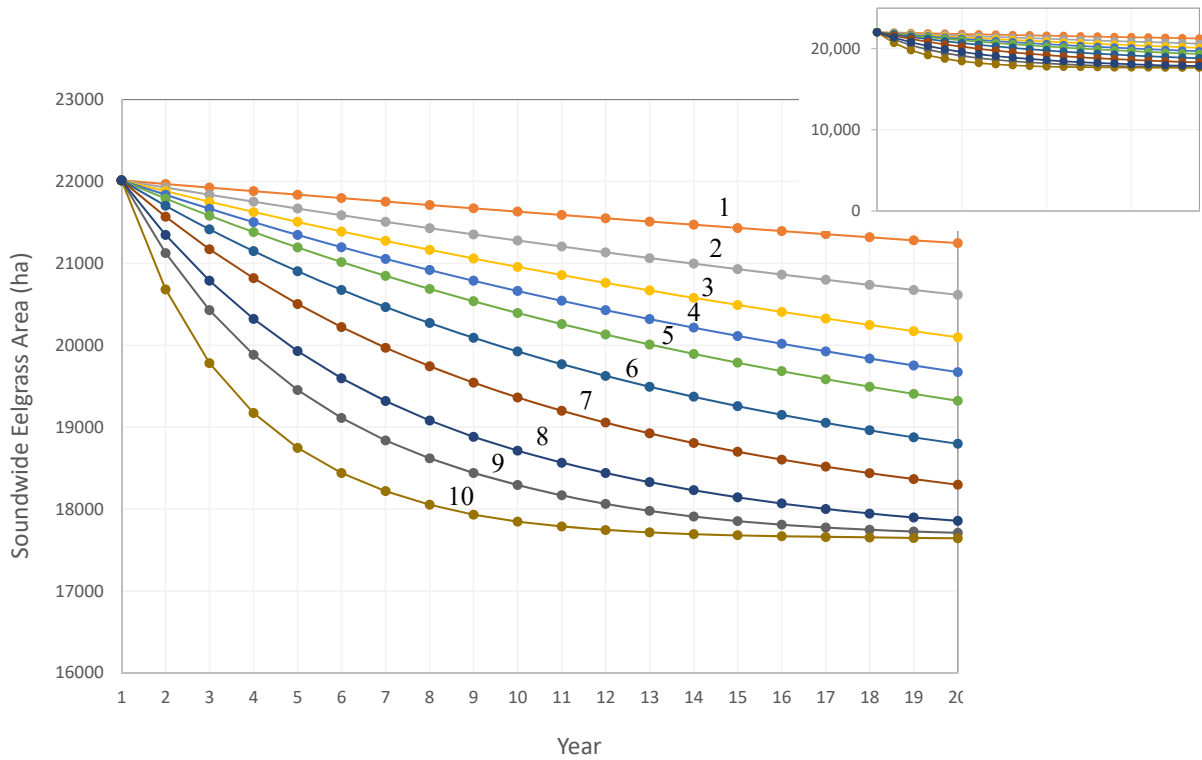


Figure 3-2. Annual soundwise eelgrass area values for the 10 change scenarios that reflect increasing levels of annual decline applied to the 0-10 ha size class. The smaller graph shows a y-axis that includes the origin. In the main graph the scenarios are labelled 1 through 10.

Table 3-1. The mean annual change, \bar{r} , used to generate the 10 change scenarios shown in Figure 3-2 and the resultant 20-year soundwise decline.

scenario	mean annual change \bar{r}	20-year soundwise decline
1	-0.01	-3.5%
2	-0.02	-6.4%
3	-0.03	-8.7%
4	-0.04	-10.6%
5	-0.05	-12.2%
6	-0.07	-14.6%
7	-0.10	-16.9%
8	-0.15	-18.9%
9	-0.20	-19.6%
10	-0.30	-19.9%

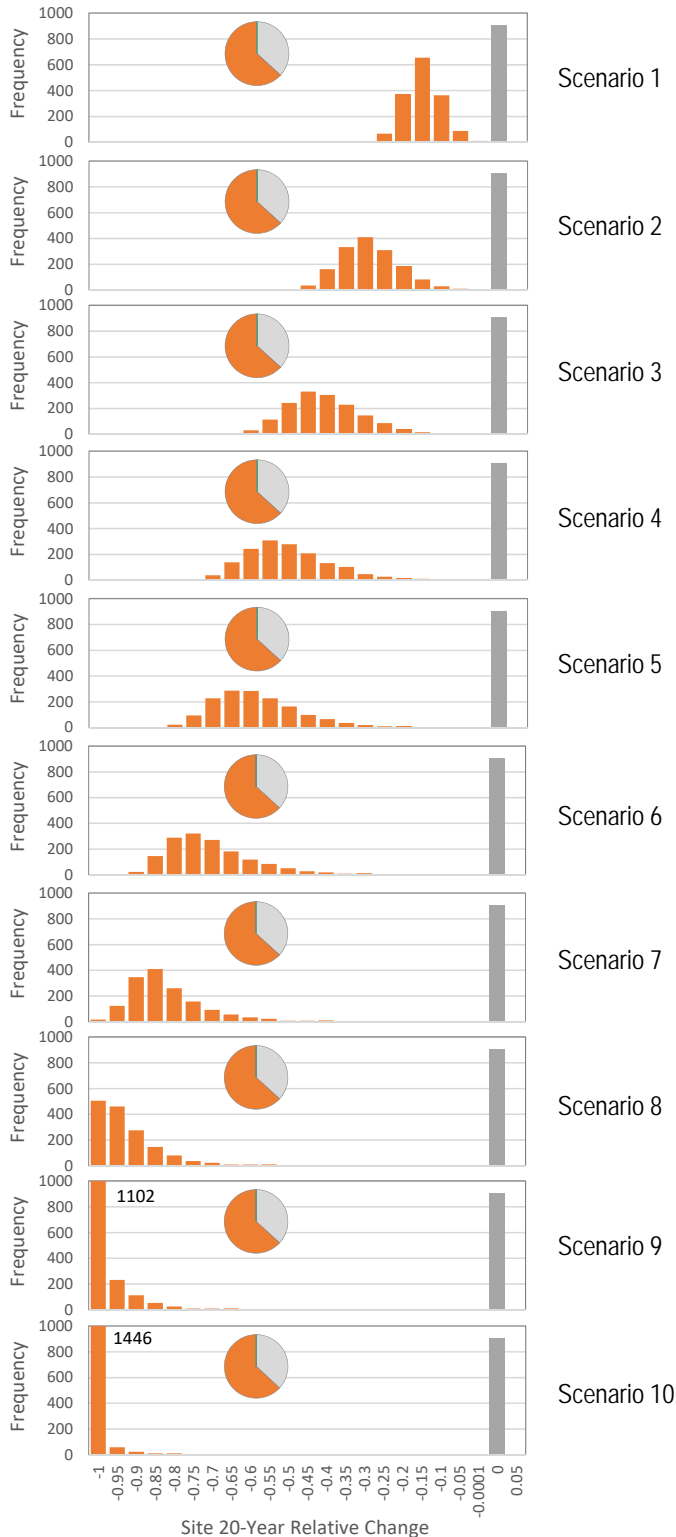


Figure 3-3. Frequency of 20-year site decline across the entire population ($n=2,467$ sites) for each of the 10 scenarios with different levels of loss in the 0-10 ha size class. The pie charts show the prevalence of decline across the entire population of sites with eelgrass (63%) which does not vary across these scenarios. The gray bar indicates the frequency of sites with no change in eelgrass area. Bars that extend beyond the y-axis scale in scenarios 9 and 10 are labelled with the actual frequency.

3.4.2 Scenarios affecting contrasting size classes

A set of five change scenarios was constructed that each reflect the same level of 20-year soundwide loss of eelgrass area (-10%) but with different sized footprints across the population. Change in each scenario is restricted to a different size class (Figure 3-4). Each size class encompasses sites of different sizes and includes different numbers of sites, but the classes were constructed so that the initial total eelgrass area within each class is the same (Figure 3-5).

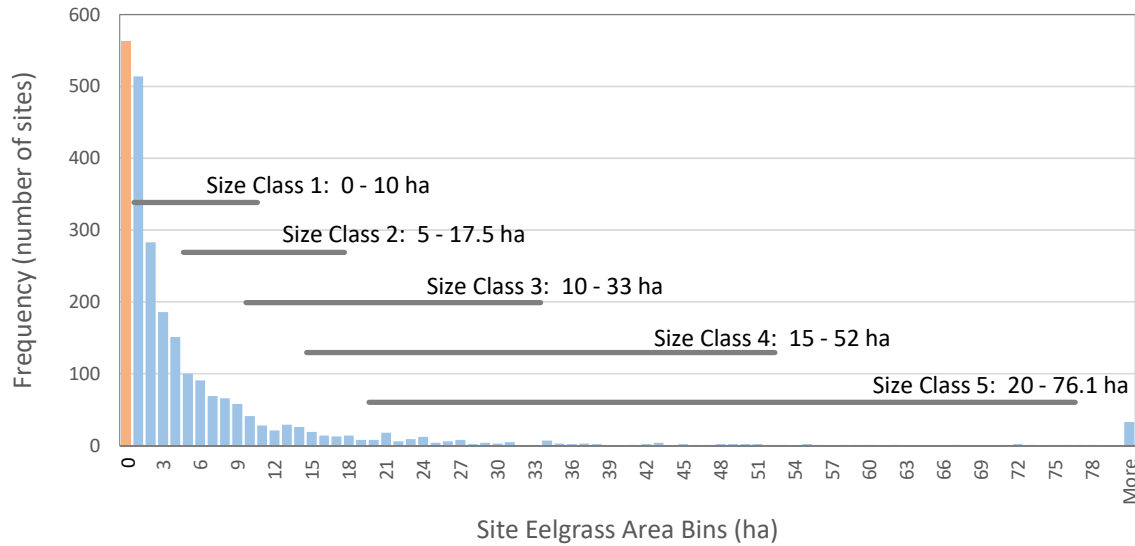


Figure 3-4. Frequency histogram of site eelgrass area under initial conditions with the five different size classes used depicted by horizontal bars that show the site eelgrass areas encompassed within the size class. The frequency of sites with no eelgrass is shown with an orange bar.

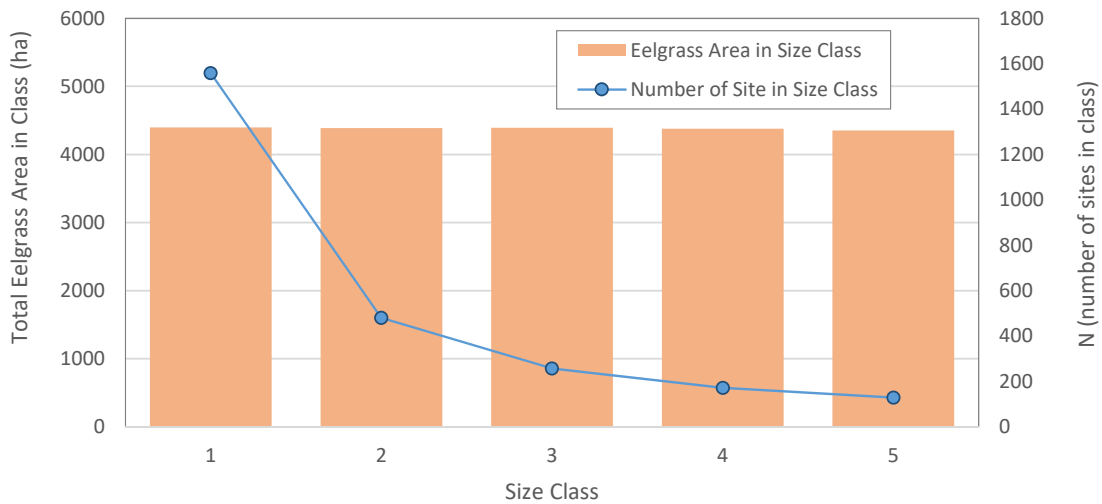


Figure 3-5. The total eelgrass area under initial conditions and the number of sites within each of the five size classes.

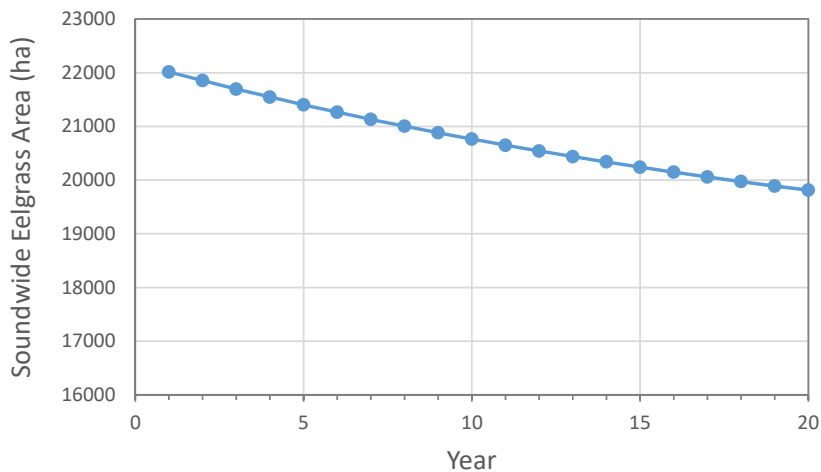


Figure 3-6. The 20-year decline in soundwide eelgrass reflected in each of the five change scenarios that affect different site eelgrass area size classes.

The magnitude of site losses of eelgrass is the same under each of the five size class scenarios when examined on a relative basis (Figure 3-7). There is variation in relative change across sites (multiple orange bars have non-zero values in Figure 3-7) but the range of change values remains the same. The main difference between scenarios that is not apparent in Figure 3-7 is that the sites experiencing decline in scenario 5 are much larger than those in scenario 1. This explains how the low frequency of decline sites in scenario 5 can give the same absolute area of loss as the much higher frequency of decline sites in scenario 1.

3.4.3 Scenarios with contrasting prevalence of change

A set of six change scenarios was constructed that is similar to the size class scenarios of the previous section in two respects. First, the total 20-year soundwide loss of eelgrass area is the same for each scenario and of the same magnitude (-10%). Second, the numbers of sites that experience decline varies across the scenarios. The six change scenarios presented here differ from the previous size class scenarios in that only one size class is involved for all scenarios and there is variation in the intensity of decline at the sites experiencing decline.

All six scenarios involve change constrained to the 0-30 ha size class. This size class includes approximately 73% of sites with eelgrass and 39% of the total eelgrass area in the Puget Sound model. The scenarios were devised by first selecting a sequence of prevalence values reflecting a decreasing proportion of sites in the 0-30 ha size class affected by decline. Then the mean site decline was adjusted for each given prevalence so that each combination gave the same 20-year soundwide loss of eelgrass area (-10%).

The graphical summary of the scenarios (Figure 3-8) clearly shows the sequence of decreasing prevalence of decline coupled with increasing intensity of decline in the scenarios.

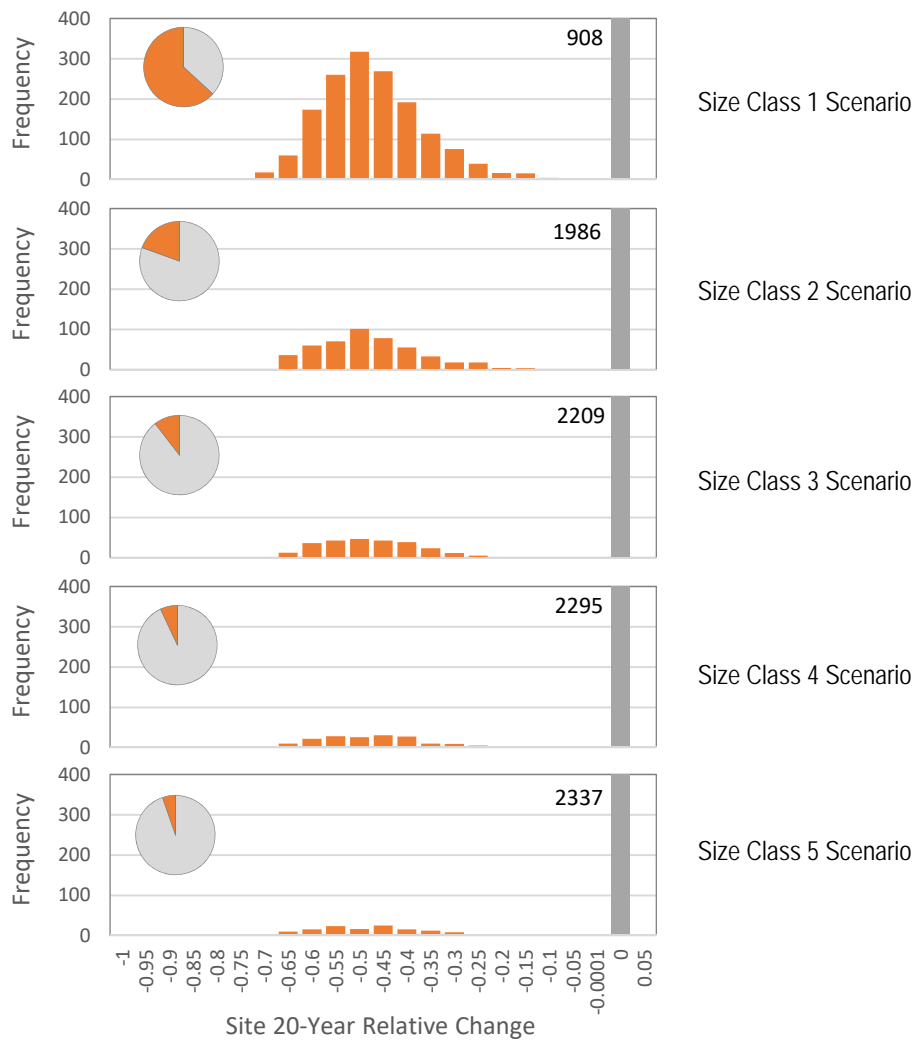


Figure 3-7. Frequency of 20-year site decline across the entire population ($n=2,467$ sites) for each of the five scenarios that differ by size class. The pie charts show the prevalence of decline across the population of sites with eelgrass. The gray bar indicates the frequency of sites with no change in eelgrass area. Bars that extend beyond the y-axis scale are labelled with the actual frequency.

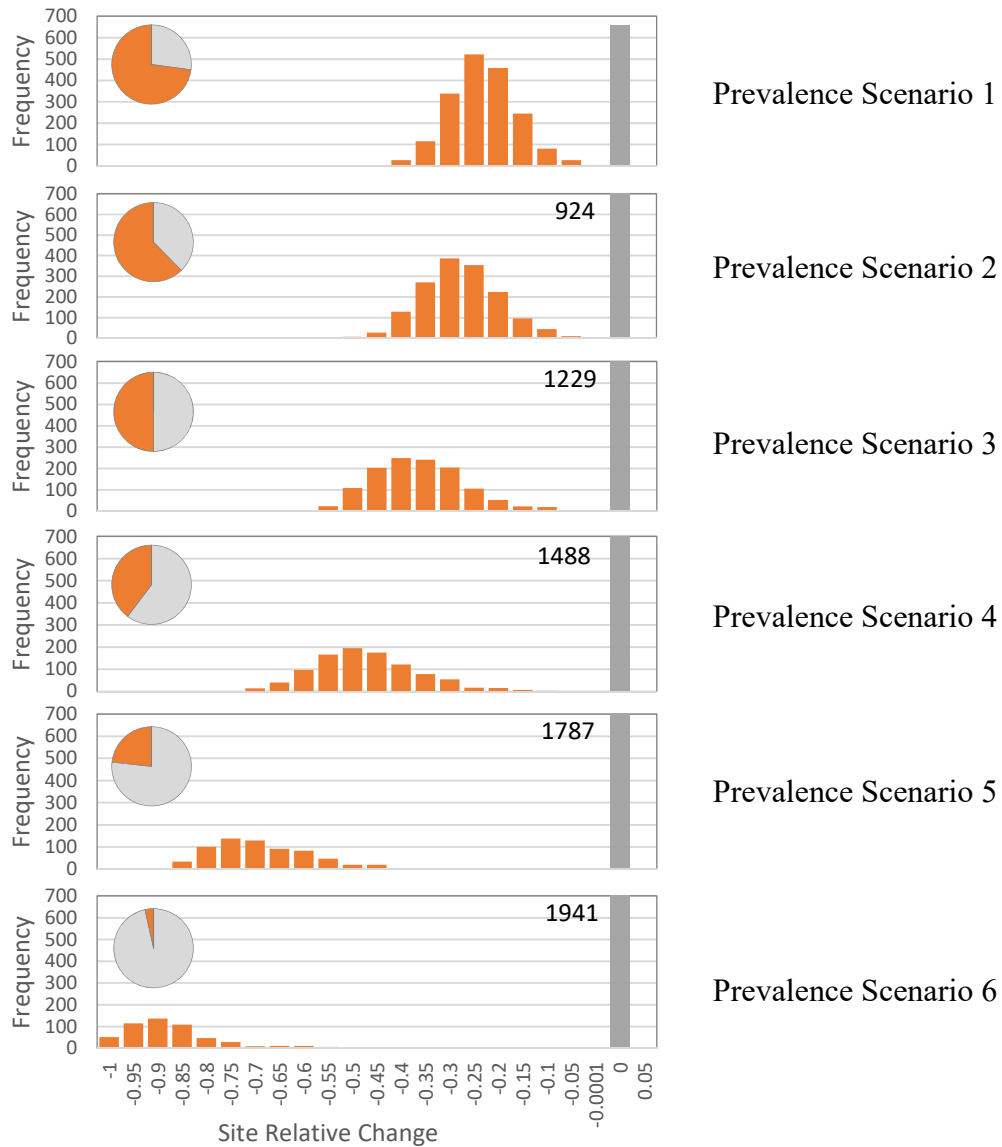


Figure 3-8. Frequency of 20-year site decline across the entire population ($n=2,467$ sites) for each of the six scenarios that differ by prevalence and intensity of decline within the 0-30 ha size class. The pie charts show the prevalence of decline across the entire population of sites with eelgrass. 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

Sampling from the Model Population



4 Sampling Approach

The sampling model simulates sampling of sites from the model population in an automated fashion. The model accommodates any specified stratification of the population but for the purposes of SVMP studies, the stratification described in previous sections is used. Each stratum can either be subject to a census, where eelgrass area is estimated for each site, or a sample, where eelgrass area is estimated at a simple random sample of sites. In order to flexibly handle a variety of sampling designs a systematic framework was needed to represent these designs. This framework, described in the following sections, provides the ability to generate a sampling prescription based on parameters that are varied to prescribe specific designs.

The SVMP involves two stages of sampling and these are each represented in the sampling model. The first stage is the sampling of sites from a specified population which in the case of the SVMP is a stratum subject to sampling. This first stage of sampling does not apply to strata subject to a census. The second stage of sampling is the sampling of transects from the population of possible transects at each selected site.

The framework used to represent the sampling of sites (first stage of sampling) was adapted from McDonald (2003). The central concept of this framework is that groups of sites can be selected and these groups can be scheduled for sampling in a variety of ways across the sampling occasions in the study period. These groups of sites are referred to as panels. In this framework, the sample of sites to be surveyed for a particular sampling occasion can be made up of one or more panels of sites. McDonald (2003) draws a distinction between the scheduling of panels for sampling (the “revisit design”) and the selection of sites that are members of each panel (the “membership design”). The revisit designs and the membership designs accommodated by the sampling model are described in section 5 and section 6 respectively.

For the second stage of sampling, the sampling model accommodates different sample selection methods for transect sampling at sites. These include simple random selection, stratified random sampling with one unit per stratum and systematic sampling. Only two revisit designs are accommodated for transect sampling – retention of the same transect sample across occasions or the total replacement of the sample each occasion. Details on the types of transect sampling are presented in section 7 (p.37).

5 Panel Revisit Designs

5.1 The Revisit Description String

A string of characters that conforms to a formal syntax is used to specify a revisit design for each panel. This level of formality allows the sampling model to simulate a wide variety of sampling designs simply by specifying this string as a model parameter. The syntax presented by McDonald (2003) is used here with some additional constraints imposed to simplify the model development.

A revisit string is enclosed within brackets in this text but the brackets are not used in the actual parameter files used with the model. The basic unit of the string is a pair of characters separated by a dash, for example [1-3]. The first character specifies the number of consecutive sampling occasions that a panel is sampled before rotating out of the sample (Table 5-1). The second character specifies the number of consecutive sampling occasions that a panel is not sampled before it rotates back into the sample.

A string can include multiple pairs in which case these pairs are separated by dashes, for example [1-3-2-2] or [1-3-2-2-1-1]. If a string includes multiple pairs, the instructions in the pairs are implemented in the sequence that the pairs appear within the string. If the string includes only a single pair, the instructions in the pair are repeated through the sampling period except when the second character is an “*n*” in which case the panel is never sampled again after it rotates out of the sample. The sampling schedules for some example revisit strings are shown in Table 5-2.

Table 5-1. Possible values for the characters in any character pair in the revisit description string and their meaning.

	Possible values	Meaning
First character	digits 1 through 9	The number of consecutive sampling occasions that a panel is sampled before it can be rotated out of the sample and not visited for some number of subsequent occasions
	digits 1 through 9	The number of consecutive occasions that a panel is not sampled before it is rotated back into the sample
Second character	<i>n</i>	After rotating out of the sample, the panel is never revisited.
	0	The panel never rotates out of the sample.

Table 5-2. Sampling schedules for example revisit description strings when applied to a single panel. An “x” indicates the panel is sampled for the specified occasion.

revisit string	Sampling Occasion									
	1	2	3	4	5	6	7	8	9	10
[2-2]	x	x			x	x			x	x
[2-1]	x	x		x	x		x	x		x
[4-n]	x	x	x	x						
[1-0]	x	x	x	x	x	x	x	x	x	x
[2-1-3-3]	x	x		x	x	x				x

5.2 Panel Groups

While Table 5-2 shows schedules for single panels, each revisit string actually dictates a group of panels. For cases where the revisit string contains only digits (no “n”), the number of panels in the group is equal to the sum of the two digits in the first pair. For example, the revisit string [2-1] dictates a group of three (2+1) panels. This is the number of panels that leads to a stable number of panels per occasion for a one-pair revisit string following the initial ramp up (see Table 5-3).

Table 5-3. Sampling schedules for each panel in the panel groups defined by example revisit strings over 10 sampling occasions. Each example with multiple panels in the group are shown with a staggered panel startup.

revisit string	Panel	Sampling Occasion									
		1	2	3	4	5	6	7	8	9	10
[2-2]	1	x	x			x	x			x	x
	2		x	x			x	x			x
	3			x	x			x	x		
	4				x	x			x	x	
[4-n]	1	x	x	x	x						
	2		x	x	x	x					
	3			x	x	x	x				
	4				x	x	x	x			
	5					x	x	x	x		
	6						x	x	x	x	
	7							x	x	x	x
	8								x	x	x
	9									x	x
	10										x
[1-0]	1	x	x	x	x	x	x	x	x	x	
[2-1-3-3]	1	x	x		x	x	x				x
	2		x	x		x	x	x			
	3			x	x		x	x	x		

For revisit strings with multiple pairs of digits (e.g., [2-1-3-3]), the number of panels sampled per occasion generally varies through the sampling period. For revisit strings with an “n”, the number of panels in the group is not fixed but rather is determined by the number of occasions in the sampling period.

5.3 Start up plans

Panel groups that have multiple panels can have different start-up plans. In Table 5-3 only staggered start up plans are shown where the first sampling occasion for each successive panel is delayed by one sampling occasion. It is possible to have a “block” start up as an alternative. The intent of the block start up is to have a uniform number of panels sampled on each occasion starting from the first occasion. The different start up plans are illustrated in Table 5-4.

The sampling model is limited with respect to the start up plans. For all [x -y] revisit strings where both x and y are digits, a staggered start up is imposed. Of course this only applies to panel groups with more than one panel (e.g., a [1-0] design does not require a start up plan). For [x-n] revisit strings a block start up is imposed.

Table 5-4. Comparison of a staggered start up and a block start up for panel group dictated by a [4-n] revisit string.

revisit string	Start up plan	Panel	Sampling Occasion												
			1	2	3	4	5	6	7	8	9	10			
[4-n]	staggered	1	x	x	x	x									
		2		x	x	x	x								
		3			x	x	x	x							
		4				x	x	x	x						
		5					x	x	x	x					
		6						x	x	x	x				
		7								x	x	x	x		
		8										x	x	x	
		9											x	x	
		10													x
[4-n]	block	1	x												
		2	x	x											
		3	x	x	x										
		4	x	x	x	x									
		5		x	x	x	x								
		6			x	x	x	x							
		7				x	x	x	x						
		8					x	x	x	x					
		9						x	x	x	x				
		10							x	x	x	x			
		11									x	x	x		
		12											x	x	
		13													x

5.4 Split-panel designs

It is possible to include multiple panel groups, each with its own revisit design, in a single sampling design. These are referred to as split-panel designs. In the revisit description string each revisit design is separated by a comma. For example, the revisit string [3-1, 1-0] represents a split panel design with two panel groups: a group of four panels with a [3-1] revisit plan and a panel group made up of a single panel with a [1-0] revisit plan. The sampling schedule for such a plan is illustrated in Table 5-5.

Table 5-5. Sampling schedule for an example split-panel design.

revisit plan	Panel group	Panel	Sampling Occasion									
			1	2	3	4	5	6	7	8	9	10
[3-1, 1-0]	1	1	x	x	x		x	x	x		x	x
		2		x	x	x		x	x	x		x
		3			x	x	x		x	x	x	
		4				x	x	x		x	x	x
	2	1	x	x	x	x	x	x	x	x	x	x

While the split-panel design shown in Table 5-5 bears some resemblance to the current SVMP soundwide sampling with a core stratum visited each sampling occasion, the similarity is only superficial. The sites selected to the panel visited each occasion are selected from the overall population sampled and are therefore representative of the overall population. In fact, for any given occasion, the sites from the panel visited each occasion could be pooled with other panels visited that occasion to make a single sample selected by simple random sampling. In contrast, the SVMP core stratum is censused and the sites in the stratum are only representative of the core stratum. They cannot be pooled from sites from a panel that is sampling a different stratum. In Table 5-5, all panels sampled on a given occasion would be selected from the same stratum.

5.5 Limitations on allowed panel designs

The sampling model has several limitations relative to the more flexible framework presented by McDonald (2003). The greater flexibility was not needed for the sampling designs selected for the modeling studies that were initially planned. By excluding these features the model development effort was simplified. The specific features excluded include the following:

1. Capability to adjust the numbers of panels in a panel group. For example, McDonald (2003) allows for a revisit plan of [3-1] to include a number of panels that deviates from the default (4 panels). Such a plan with, for example, only three panels would be represented as [(3-1)³]. This type of adjustment of panel numbers is not allowed in the model developed for this study.
2. Ability to withhold a panel from sampling in the initial occasions. This would be represented as a leading 0 in a revisit string, e.g. [0-2-3-1] but this is not allowed here.

3. Ability to never revisit panels after a complex revisit scheme. This could be represented, for example, as [3-1-2-*n*] but this is not allowed. The “*n*” option is only allowed in a single pair revisit string, e.g. [3- *n*].
4. Flexibility with respect to start up plan. As noted earlier, any type of [*x* -*y*] revisit plan is only allowed a staggered start up and any [*x*- *n*] revisit plan is only allowed a block start up.

5.6 The SVMP Revisit Design and Alternatives

Four designs are illustrated through sampling schedules. These include the design in the original SVMP framework and three alternatives selected for evaluation for possible implementation by the SVMP. The evaluation of these alternatives will be presented in a separate report but the revisit design strings are presented here in Table 5-6 and 20-year sampling schedules are given in Table 5-7.

Table 5-6. The original SVMP revisit plan and three alternatives identified for evaluation.

label	revisit string	description
20% rotation	[5- <i>n</i>]	The SVMP design of 20% site rotation on each occasion that is applied to the sampled strata (rotational flats, narrow fringe, wide fringe).
fixed sites	[1-0]	One panel of sites sampled on each occasion.
new draw	[1- <i>n</i>]	A new panel is sampled on each occasion.
3 panels	[1-2]	Continual rotation among three panels.

Table 5-7. 20-year sampling schedule for the four revisit designs selected for evaluation by the SVMP.

revisit plan	panel group	panel	Sampling Occasions																				
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
3 panels	1	1	x			x			x			x			x			x			x		
		2		x			x			x			x			x			x			x	
		3			x			x			x			x			x			x			x

Table 5-7 (continued).

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

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

new draw	1	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	



6 Panel Membership Designs (Site Selection)

6.1 Random Selection of Eligible Sites

The panels required for a particular design were populated with sites by simple random selection from the pool of eligible sites. Selection was conducted without replacement so after each site was selected the pool of eligible sites for the next selection was decreased by one.

For $[x-y]$ panel designs (no “n”), there are a fixed number of panels regardless of the duration of the sampling period. For these designs all panels were populated as a group without site replacement. This ensured that site membership in panels was unique. A site could not be a member of more than one panel. It is possible when large sample sizes are specified and panels contain large numbers of sites that there will be an insufficient number of sites to populate all panels. The sampling model does not accommodate such a case and the computer application will issue an error message and terminate.

For $[x-n]$ panel designs, the application returns sampled sites to the pool eligible for re-selection as described in the next section.

6.2 Site Selection Eligibility in Rotating Designs $[x-n]$

The problem of exhausting the sites available for selection before all panels are populated is more likely under $[x-n]$ designs that do not reuse panels through the sampling period. For example, the new draw design in Table 5-7 requires 20 panels to cover 20 years of sampling. In the case of the SVMP flats rotational stratum which has an annual sample size of 10 sites, sampling without replacement would require 200 sites to populate all the panels although the entire population has only 67 sites. Some site replacement is clearly necessary to apply this design to the flats stratum.

The SVMP’s $[5-n]$ design for sampled strata implemented operationally in the period 2001-2014. An additional stipulation was applied a five year waiting period after a previously selected site rotates out of the sample before it can be selected again. This waiting period acts to disperse the sampling effort to new sites. It is not clear if this waiting period was intended as part of the original sampling design. It is not described in Skalski (2003). An alternative is to require a five year waiting period once a site is selected before

it can be re-selected. This is the minimal constraint to ensure that each annual sample will contain unique sites.

This latter replacement policy was implemented in the sampling model in a more general formulation. Specifically, a site selected for a panel as part of a $[x-n]$ design would have a waiting period of x years before it could be selected again to another panel. In the case of a $[1-n]$ design (new draw design of Table 5-7) a site could potentially be selected each year. The SVMP $[5-n]$ design would be implemented in the computer application such that a site could potentially be selected every five years.

7 Transect Selection

The sampling model selects a sample of transects to characterize each site selected on each sampling occasion for which the site is selected. The transects are drawn from the site belt model which is in turn based on the transect set assigned to each site. The exact procedure for determining the transect sample for each sampling occasion can vary and is fully specified by only three parameters (Table 7-1).

The sampling model accommodates three different transect selection methods: simple random, stratified random sampling with one unit per stratum and systematic sampling. Only two revisit plans are accommodated – a new draw of transects each occasion or retention of the initial sample through all occasions. The sample size (number of transects) has a minimal value of one and no hard upper limit. A practical upper limit is imposed by available computer memory and tolerance for impacts on computing performance, but as sample size increases beyond the resolution of the site model the results are less meaningful.

Table 7-1. Transect sampling parameters.

parameter	domain
selection method	SRS = simple random sampling STR = stratified random sampling with one unit per stratum SYS = systematic sampling with a random start
revisit plan	new = new draw of transects for each sampling occasion repeat = retain initial draw of transects on subsequent occasions
sample size	Number of transects sampled with a minimum value of 1 and no set upper limit.

7.1 Simple Random Selection

Simple random transect selection is conducted with replacement from the site belt model for each selected site with the belt fraction values in place for any given year. The site model is conceptualized as having infinitely many transects available for selection. Regardless of how many belts are in the site model, they are stretched onto a scale from 0 to 1 for use with a floating point uniform random number generator on the [0,1] interval. Although this is conceptualized as a model with infinitely small grain, the “numeric” grain

is finite and determined by number of transects in the initial transect set assigned to the site. This resembles a bootstrap in the initial year when actual field data is effectively resampled, but given subsequent years are represented by model data (generated with the model change scenarios) it is more sensible to think of this as sampling from the site belt model that was introduced on p.15 and discussed further in the next section.

7.2 Stratified Random and Systematic Selection

Stratified random sampling with one unit per stratum and systematic sampling each ensure a relatively uniform spatial distribution of transects in comparison to the clustering of samples that can occur with simple random selection. Stratified random sampling, for example, is used by EPA EMAP designs, including implementations with a hexagon-based sample selection, to achieve spatially balanced samples (Stevens and Olsen 2004). Systematic sampling is a more widely known method and has even less randomization and produces regularly spaced samples (Wolter 1984). When systematic sampling is implemented with a random start some level of randomization is retained.

The procedures used here for restricted random and systematic transect selection are similar. The spatial belt model of the sampling site is divided into n sections of equivalent longshore width where n is also the number of transects to be selected for each sample. Let m be the number of belts in the belt model being sampled. Then there will be k strips per section where $k = m/n$. Figure 7-1 illustrates a simple example with $m = 10$ belts and $n = 4$ sections.

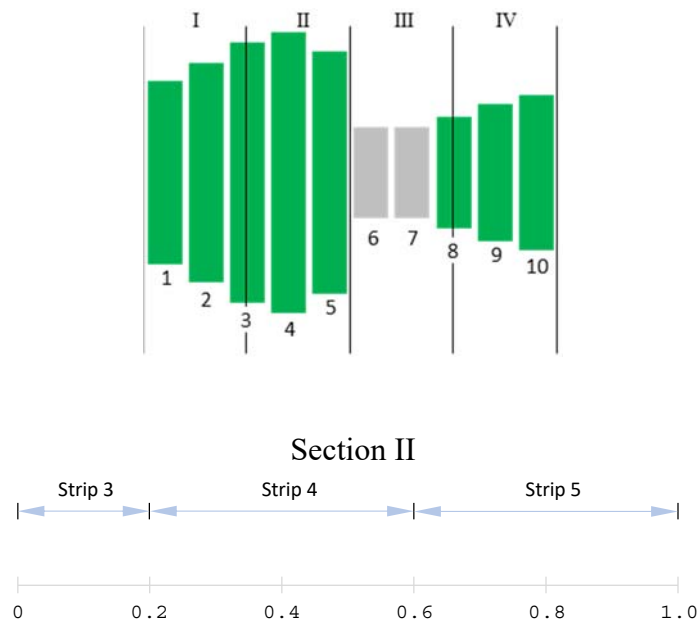


Figure 7-1. Simple belt model of a site represented by 10 belts (numbered 1 through 10) constructed from a transect set of $n=10$. The site has been divided into four sections numbered I through IV. The length of green belts is representative of eelgrass fraction and gray belts indicate the absence of eelgrass. The lower illustration shows how for section II a random number drawn from a uniform distribution on the $[0,1]$ interval is associated with a particular belt which provides the “observed” transect data.

Under stratified random sampling, a random floating point number on the [0, 1] interval is generated to select a transect within the width of each section (giving in concept an infinite population of possible transects) (Figure 7-1). The location of the transect within the section identifies a particular strip whose fraction and length values are taken as the transect observations.

Under systematic sampling, the same approach is taken in the first section to provide a random start to the systematic transects. For all other sections the same random point within each section is used to impose uniform spacing to the selected transects.

There are two important considerations when utilizing either stratified random or systematic in model applications. First, the sequence of transects in each set in the input file is then meaningful as this specifies the spatial pattern in the site model. This would require an intensive GIS effort to properly order the transects. The current sequence of transect data as ordered in the prepared data input file adhere more closely to the temporal sequence of field data collection. Second, the minimum number of transects should be adjusted that is used as a criterion to identify transect sets in the SVMP dataset for use in the modeling. Rather than requiring at least 20 transects in each transect set to qualify for the modelling work (p.14), 60-80 transects may be needed to provide a meaningful distinction between the transect selection methods.

Analysis of Sample Data

8 Conceptual Framework

This section presents a framework for organizing the large number of SVMP analysis possibilities. The purpose is to build an understanding of the variety of sampling and analysis alternatives and help develop a prioritized sequence of particular scenarios for model investigation. Only a subset of the possibilities has been developed in the current version of the model code.

8.1 *Data Streams from Field Sampling*

The SVMP design is a two-stage sampling design (sampling of sites within strata and sampling of transects within sites). Different sampling design choices are made at each stage of sampling. Starting with the sampling of transects, let us represent the domain of all possible transect sampling options as a square region (Figure 8-1).

Within this domain there are different transect selection methods (simple random, stratified random, systematic) and revisit plans (new draw each sampling occasion, repeat initial draw on subsequent occasions). To simplify this discussion we will only consider simple random sample selection. Then the domain really only contains two sub-regions for new draw and repeat transect sampling, each with transects selected by simple random sampling (Figure 8-1).

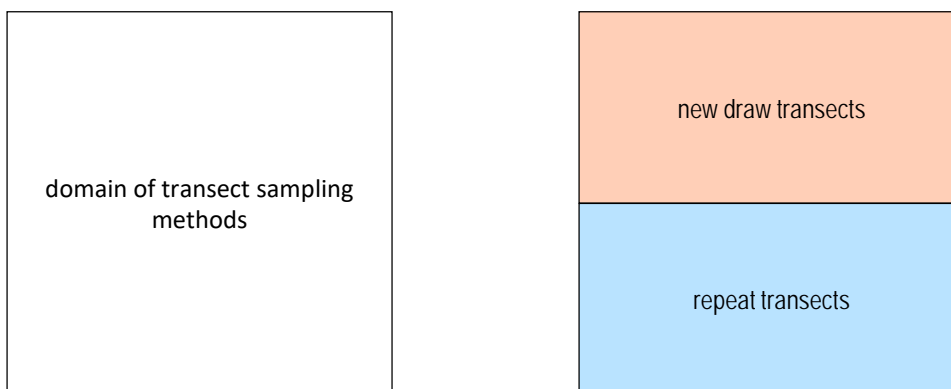


Figure 8-1. The domain of transect sampling options is represented as a square region (left). If we restrict consideration of sampling methods to simple random sampling, then there are only two sub-regions within the domain – new draw and repeat transect sampling (right).

In contrast, the domain of options for the sampling of sites encompasses multiple sub-regions for design alternatives that include a new draw of sites each occasion, fixed sites across all occasions and a large variety of intermediate panel revisit plans. If we restrict our consideration to the panel revisit plans discussed in section 5.6 (p.33) then the domain of site sampling options is divided into four sub-regions (Figure 8-2).

new draw of sites	SVMP 20% rotation revisit schedule [5- <i>n</i>]	3-year rotating panel revisit schedule [1-2]	fixed sites
-------------------------	--	---	----------------

Figure 8-2. The domain of site sampling options represented as a square with the four options considered here represented as four sub-regions.

To enumerate the different possible data streams resulting from site field data collection, the site-level domain is overlaid on the transect-level domain. This produces eight combinations of transect and site sampling but since repeat transect sampling is not possible with new draw site sampling there are only seven viable combinations (Figure 8-3). These are listed as

1. New draw transects – new draw sites
2. New draw transects – 20% site rotation
3. New draw transects – 3-year rotating panels
4. New draw transects – fixed sites
5. Fixed transects – 20% site rotation
6. Fixed transects – 3-year rotating panels
7. Fixed transects – fixed sites

The resulting seven categories are a simplification of the actual variety of possibilities. Not only have we focused only on simple random sampling of transects and excluded stratified random and systematic sampling, we have also excluded all intermediate revisit schedules except the two listed (20% site rotation and 3-year rotating panels).

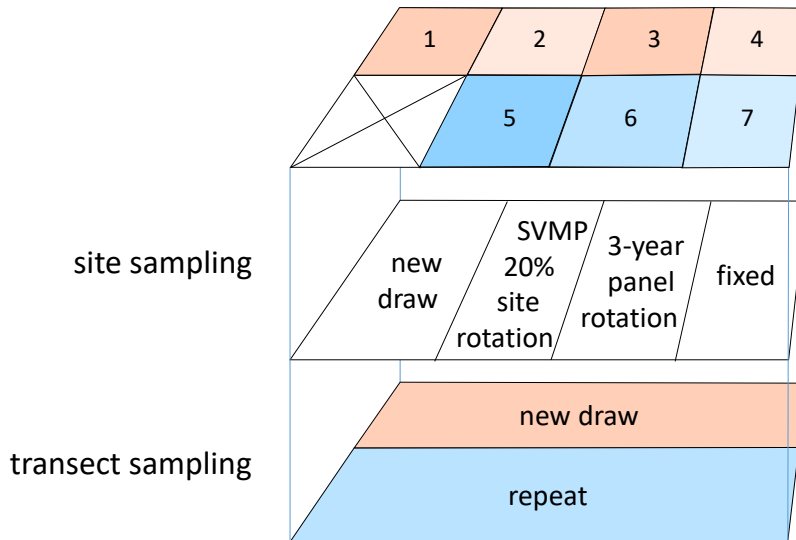


Figure 8-3. The domain of data streams derived from the field sampling (top level) include seven different sub-regions within the constraints considered here. The number of each stream corresponds to the list on the previous page. The sub-regions result from overlaying the four options within the site sampling domain (mid level) with the two options of the transect sampling domain (bottom level). It is not possible to sample a new draw of sites with repeat transects so this option is disallowed in the domain of data streams.

8.2 Analysis Pathways

For each alternative data stream there are alternative analysis options eventually leading to some statistical inference about the Puget Sound eelgrass population. These can be organized into different sequences of estimation that start with transect-level measures and run through site-level measures, soundwide measures and ultimately lead to a statistical inference about the soundwide trend in eelgrass abundance over time. The possible analysis pathways are not discussed in detail but a variety of possible pathways are illustrated in Figure 8-4.

The nature of any given data stream produced from the field data collection (Figure 8-3) will limit the available analysis options. For example, any data stream built on a new draw of transects at each sampling occasion is not compatible with analyses that rely on estimates of transect change or trend (pathways 4-6 in Figure 8-4).

In the current version of the working model, only pathways 1 and 2 have been implemented in the computer application.

8.3 Analyzing for Linear Trend

The approach prescribed in the original SVMP framework is to conduct a linear trend analysis on annual soundwide eelgrass area estimates using linear regression. The regression slope is tested for significance using the variance on the slope estimated by regression. A few issues have been identified with this approach.

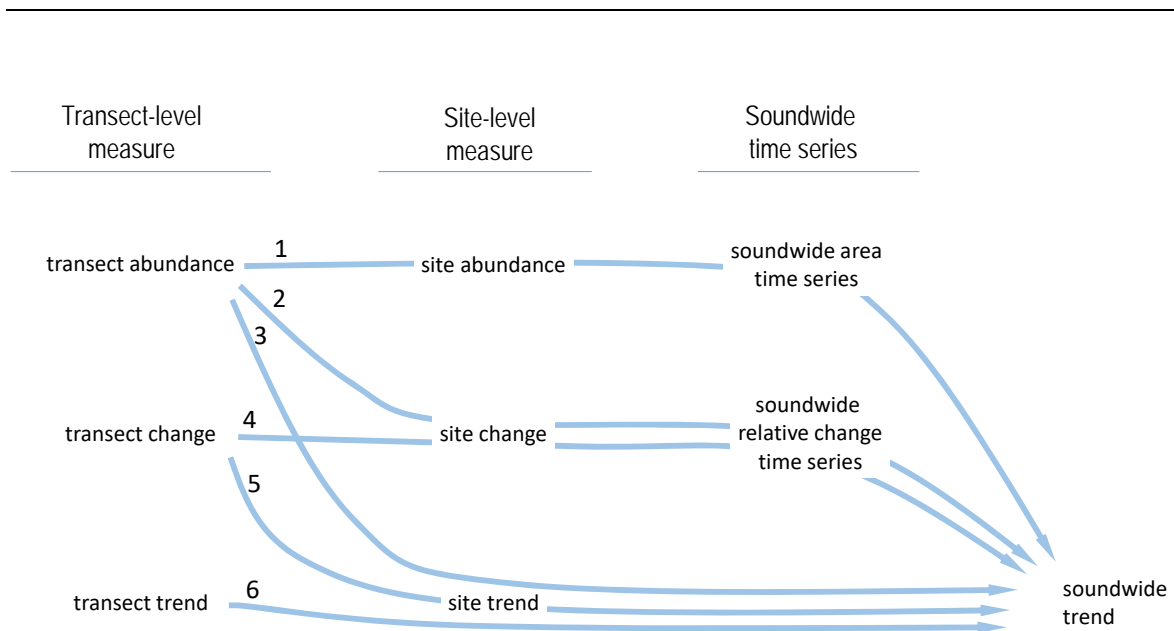


Figure 8-4. Alternative analysis pathways leading from transect-level measures to site-level measures, soundwide measures and finally an assessment of soundwide trend. There are six pathways that have been numbered for referencing within the text.

Site Sample Size

First, the stratum sample sizes (numbers of sites) appear to be inadequate for estimating stratum eelgrass abundance due to the highly skewed distribution of site eelgrass abundance. This problem is manifested in the existing SVMP dataset as apparent trend in soundwide abundance that is actually an artifact associated with the rotation of sites with anomalous eelgrass abundance. This is hardly surprising in the case of the rotational flats stratum with $n=10$ but the same artifacts are observed in the narrow fringe stratum with $n=45$ which demonstrates the severity of the effects of the skewed distribution. This is consistent, however, with the suggested sample size of $n=90$ for a similarly skewed distribution in an example presented by Cochran (1977, p.44).

The solution to the inadequate sample sizes is unclear because of resource constraints but two approaches have been discussed. First, by moving to a fixed sample of sites the artifacts introduced by rotating sites would be eliminated. Appropriate analysis methods would need to be developed for this data stream that accommodate a limited fixed (dependent) sample from a skewed distribution. Second, by moving from abundance-based estimates (soundwide estimates built on transect and site abundance estimates) to change-based estimates (estimates built on transect or site-level change estimates), it is likely that the skewed parent population will be replaced with a parent distribution with a strong central tendency that more closely resembles a normal distribution. The potential benefit there is a lower sample size requirement for a given level of precision or power to detect change. The methods underlying such soundwide estimates have not yet been developed in the computer application.

Data Dependence

A second issue that has been identified with SVMP linear trend analysis is the retention of sites in the sample for five years. This violates the regression assumption of data independence. The quantitative implications of this violation have not been characterized but a separate modelling exercise (to investigate methods for estimating multi-year means of dependent estimates) suggests that true precision can be strongly overestimated with dependent samples. It may be that a permutation test would perform adequately or it may be possible that a new approach such as repeated measured regression or a linear mixed model would be superior.

Adequacy of Linear Model

In many cases, inspection of an SVMP site-level time series of eelgrass area clearly shows a non-linear pattern. A linear model does not fully capture the temporal dynamics in such cases. While the soundwide time series developed to date have not shown obvious departures from linearity, the site-level cases bring up the question of whether a non-linear model should be the SVMP default, perhaps to be replaced by a linear model in cases where this simplification seems appropriate.

A model captures the maximum amount of relevant information when the fit to sample data captures the underlying relationship between variables while avoiding explicitly modeling the noise. In such a case the residuals truly represent noise and there should be no structure apparent in the residuals. In an underfit model there will be some component of the underlying relationship that is not captured by the model. This results in structure appearing in the residuals so they no longer represent only noise. In contrast, in an overfit model not only is the underlying relationship captured by the model but elements of the noise in the sample are explicitly modeled. This reduces the generality of the model and reduces its predictive utility.

When a linear model is used to characterize non-linear data this is a case of underfitting and there will be structure in the residuals. This violates the requirement of independent residuals for the linear regression parameter estimates which prevents, or, at least complicates, the use of regression statistics to assess the statistical significance of the regression slope. Even in such situations, the ecological question of interest may focus on the linear component of the underlying relationship and the linear model may serve a useful purpose. From a management perspective, the question of interest may be whether there is a long-term pattern of improving or worsening ecological conditions and higher frequency variation within the monitoring record is of lesser interest. In such a case the linear component of a time series is of primary interest even if there are clearly higher order terms introducing non-linearity in the population response. In the context of a Taylor series expansion, the task is to isolate the first order term (linear) and neglect higher order terms of lesser interest. From this perspective, rather than being an inappropriate model that violates analytical assumptions (for the purposes of statistical tests on the slope), the linear model represents the component of primary interest of a more complex model. Based on this reasoning, the SVMP has focused trend analysis on the use of a linear model.



9 Analysis Methods

A number of analyses are represented in Figure 8-4 (p.45) arranged sequentially into six “analysis pathways”. Currently only the analyses in the first two pathways of Figure 8-4 have been implemented in the computer application. However, in each case two different analysis variations were implemented so that a total of four alternative analysis pathways are available in the application. Analyses that rely on the repeated sampling of transects have not been implemented at this time.

9.1 Soundwide Time Series from Annual Abundance

The two analysis methods described here have been implemented in the current version of the working computer application.

9.1.1 Based on Annual Samples (Original SVMP Design)

A time series of annual soundwide eelgrass abundance estimates can be based on samples (or a census) of sites within strata and samples of transects for each respective year. This is the original SVMP design. It is described in detail by Skalski (2003) and the details are not presented here. The model as currently implemented is able to conduct this analysis.

9.1.2 Based on 3-Year Samples

This modification to the original SVMP design was proposed in response to the recognized need for large sample sizes (numbers of sites) to adequately represent the highly skewed distributions of site eelgrass area in the sampled strata. The idea is to base the soundwide abundance estimate for each year on a three-year pooled sample that includes sites from the previous year and subsequent year.

The intent was that this analysis method would be coupled with the 3 rotating panels revisit plan ([1-2] revisit description string; see Table 5-6, p.33, and Table 5-7, p.33). This would result in unique sites in the samples from any three consecutive years. If the annual sampling effort is consistent with current SVMP practices this would effectively triple the sample sizes in the samples strata. The cost with this approach is that there would be a loss in temporal resolution as data from three consecutive years are pooled.

Other than the pooling of site data from previous and subsequent years in the sampled strata, other analysis methods, including the specific estimators, follow the original SVMP design (Skalski 2003).

9.2 Soundwide Time Series from Annual Change

The estimation of annual change in soundwide eelgrass abundance relative to the previous year is part of the original SVMP design (Skalski 2003). More recently the possibility of estimating annual change in soundwide eelgrass abundance relative to the initial year under fixed site sampling has been proposed. Each of these approaches is discussed below but neither has yet been developed in the working model.

9.2.1 Change Relative to Previous Occasion

The original SVMP statistical framework included estimators for annual soundwide relative change in eelgrass area and the variance of the relative change estimate (Skalski 2003). The approach relied on a pairing of sites sampled in two consecutive years and a regression estimate of relative change applied on a stratum basis (x -values = year 1 site eelgrass area estimates; y -values = year 2 site eelgrass area estimates).

The use of the estimator for relative change has been retained and is still applied on an operational basis. A previous Monte Carlo study (bootstrap with Gaussian stochastic error) showed that the variance estimator was not reliable (Dowty 2005). Subsequently the variance and confidence intervals on the annual relative change estimates have been estimated using the Monte Carlo approach discussed by Dowty (2005).

The annual relative change values, each with a different baseline (abundance in the previous year), must be transformed into a sequence of change values with a shared baseline in order to support trend analysis. This is done by converted each annual change estimate to be relative to the initial year. The method for this conversion is developed as follows.

Let

B_i = soundwide eelgrass area for year i ,

$r_i = \frac{B_i - B_{i-1}}{B_{i-1}}$ = soundwide annual change in eelgrass area relative to the previous year,

$R_i = \frac{B_i - B_0}{B_0}$ = soundwide annual change in eelgrass area relative to the initial year.

The definition of r_i given above can be rearranged and used to generate a sequence of soundwide area values given by

$$B_0$$

$$B_1 = B_0(1 + r_1)$$

$$B_2 = B_1(1 + r_2) = B_0(1 + r_1)(1 + r_2)$$

$$B_3 = B_2(1 + r_3) = B_0(1 + r_1)(1 + r_2)(1 + r_3)$$

Equation 9-1

and generalized as

$$B_i = B_0 \prod_{j=1}^i (1 + r_j). \quad \text{Equation 9-2}$$

This can be rearranged to give

$$\frac{B_i}{B_0} = \prod_{j=1}^i (1 + r_j). \quad \text{Equation 9-3}$$

Substituting Equation 9-3 into the definition of R_i above gives

$$R_i = \prod_{j=1}^i (1 + r_j) - 1 \quad \text{Equation 9-4}$$

which provides a method for converting estimates of annual change relative to the previous year (r_i) to annual change relative to the initial year (R_i).

The variance of the estimate of R_i is given by

$$\text{Var}(R_i) = \text{Var}\left(\prod_{j=1}^i (1 + r_j) - 1\right) = \text{Var}\left(\prod_{j=1}^i (1 + r_j)\right). \quad \text{Equation 9-5}$$

Using the propagation of error formula for a product and neglecting the covariance term we have

$$\text{Var}(R_i) = \left(\prod_{j=2001}^i (1 + r_j)^2\right) \left(\frac{\text{Var}(r_1)}{(1 + r_1)^2} + \frac{\text{Var}(r_2)}{(1 + r_2)^2} + \dots + \frac{\text{Var}(r_i)}{(1 + r_i)^2}\right). \quad \text{Equation 9-6}$$

where the variance terms, $\text{Var}(r_i)$, are evaluated using Monte Carlo methods (Dowty 2005).

9.2.2 Change Relative to Initial Occasion

This method entails the direct estimation of soundwide change in eelgrass area relative to the initial year using a paired site analysis. This analysis method is only compatible with data streams that have at least a significant portion of site sampling effort for each year dedicated to sites that were also sampled in the first year. If this is the case, then the change relative to the initial year, R_i , is calculated using the relative change methods of Skalski (2003) except that the two years being analyzed are not consecutive except in the case of the first two years. The variance of the estimates of R_i are estimated using the Monte Carlo approach described by Dowty (2005).

9.3 Site Trend Scaled to Soundwide Trend

Another analysis method was suggested by Van Sickle (2012) that scales trends observed at the site scale to the soundwide scale. The approach relies on a random intercept linear mixed model. Conceptually this approach is very appealing but Van Sickle only presented

a limited test application of the approach and it has not yet been further developed nor implemented in the computer application.

Model Implementation in a Computer Application



10 Guide for Users

10.1 Specification of the Population

This overall effort was conducted for the SVMP and the model population discussed in this report (p.5) is specific to the Puget Sound eelgrass population. However, the model and its implementation in a computer application are actually quite flexible in the specification of the population. The population model can be specified as one or more sub-populations. For each sub-population, the total of the population parameter of interest for each sample unit (site eelgrass area for the SVMP) can be specified by one of two mechanisms:

1. External text file. In this case a text file contains a record for each sample unit specifying the total value for the unit and other information in a format specified below (section 10.3). This mechanism was used to specify site information for core and persistent flats sites in the population model presented in this report. There are no distributional constraints on the sub-population.
2. Random variate. Here, the total value for each sample unit is a random variate drawn from a distribution of specified form and parameterization. The distribution can be normal, lognormal, exponential, Pareto or a Weibull distribution.

In the current version of the application, if the model population is made up of a number of sub-populations which have separate specifications, it is necessary to treat sub-populations as separate strata for sampling purposes. When specification is through an external text file, the file can contain sample units from multiple strata.

The within-site variability is always specified by external text file. In the context of the SVMP, each record of the external file specifies a belt in the site belt model, but these records could be conceptualized as quadrats or other measures.

10.2 Hard Coded Parameters

Many model parameters are likely outside the main interest of users. These parameters were hard coded based on the premise that the user would typically have no need to vary them. By hard coding their values, the files containing parameter values to be read in at execution can be streamlined.

These parameter values were centralized in header files but they are still considered hard coded because any changes to header files requires recompiling and rebuilding.

Table 10-1. List of hard-coded parameters and their default values as specified in the *definitions.h* header file.

Parameter	Description	Default Value
NRUNS	Number of Monte Carlo iterations to perform	10,000
MXNSTRATA	Maximum number of sampling strata allowed	5
MXNUMTRANSECTS	Maximum number of transects specified by external data file to construct site belt models	15,000
MXNSTRATTRANSETS	Maximum number of transect sets specified by external data file to construct site belt models	150
MXNSITES	Maximum number of sites in an individual sampling stratum	2,500
MXNUMYEARS	Maximum number of years to be simulated	40
NUMFLATSCORE	Number of sites with population parameter totals specified by external file	76
MXNCHGSCENARIOS	Maximum number of change scenarios	30
MXNPANELGROUPS	Maximum number of panel groups in the specification of sampling design for a stratum	8
MXNPANELS	Maximum number of panels in the sampling design for an individual stratum	750
MXNPAIRS	Maximum number of number pairs in a revisit description string	5

In addition to the parameters listed in Table 10-1, the maximum string length for the names of external parameter files is specified in the *definitions.h* header file with a default value of 100 characters. This string length includes the relative path as a prefix to the actual file name. The “main” parameter file is an exception in that the maximum string length for the name (including the relative path) is specified in the definition of the string (*prmfilename*) in the main program (*focusMC.c*). The default value for the maximum string length for the name of the main parameter file is 50 characters.

10.3 Parameter Files

There is one main parameter file for each model run that must be specified by the user at runtime. This main parameter file in turn points to four additional parameter files as well as input data files and output files (described in sections below). This approach allows different groups of parameters to be varied independently while keeping others constant across model runs.

The functions that read the parameter files require a specific sequence of records so the parameter files must adhere strictly to the form as presented below. In general, some records are considered ‘header’ records, wherever they may be placed within the file, and these are read by the function and then discarded. For records containing parameter values, the functions may expect the parameter value to consist of a number of characters starting with the first character of the record and continuing until white space is encountered (space or tab). Spaces within parameter values are not allowed. After the white space that follows a parameter value any additional characters are ignored by the functions so this is a good

location for explanatory comments. Alternatively, the functions may expect each record to contain a sequence of parameters separated by white space.

The contents of an example main parameter file are shown in Table 10-2. Comments have been included to explain each parameter value. The last parameter (line 15) specifies whether sample analysis is based on an annual sample of sites (section 9.1.1, p.47) or a 3-year pooled sample of sites (section 9.1.2, p.47). Two of the output filenames (lines 10-11) are only needed if the user specifies at runtime that these files are to be written.

Table 10-2. Structure of the main parameter file with example parameter values. Lines 1-3 and 13 are considered header records and are ignored. The number sign (#) is used to indicate the start of a comment although it is not necessary – this and all following text are ignored. For parameters that are filenames, the relative path is included in the parameter value if files are stored in sub-directories (e.g., “prm_files/” and “runs/”).

line no.	record contents
1	# master parameter file for model run
2	
3	# parameter file names
4	prm_files/stratum_prms.txt # stratum prm file name
5	prm_files/flats_and_core_model.txt # flats & core model (input data) file name
6	prm_files/transect_data_all.txt # transect input data file name
7	prm_files/change_scenario_prms.txt # change scenario prm file name
8	prm_files/panel_prms_fixed.txt # panel design prm file name
9	prm_files/tran_SRS_N_13.txt # transect sampling design prm file name
10	runs/eelgrass_model_mp0087.txt # pop model output file name (can be dummy/optional)
11	runs/panel_schedule_mp0087.txt # panel schedule output file name (can be dummy/optional)
12	runs/trend_output_mp0087.txt # trend output from Monte Carlo simulations
13	
14	20 # number of years to be modeled
15	1a # analysis method (1-yr samples=1a; 3-yr samples=1b)

The stratum parameter file is specified by filename in the main parameter file (Table 10-2, line 4). The parameters in this file specify the number and naming of the stratification used both for generation of the model population and for sampling of the population. An example stratum parameter file is presented in Table 10-3 and the parameters are described in Table 10-4.

The change scenario parameter file is specified by filename in the main parameter file (Table 10-2, line 7). The parameters in this file specify the change scenario to be applied to the initial population model over the model period. An example change scenario parameter file is presented in Table 10-5 and the parameters are described in Table 10-6. Typically the model is run with only one change scenario but it is possible to specify multiple change scenarios by adding multiple records to the parameter file. The model will run through

these sequentially and the output files will include the results from all scenarios appended together.

Table 10-3. Structure of the stratum parameter file with example parameter values. Line 1 is a header record and is ignored. The text in the header was rotated here only because of space limitations. The first column with line numbers is not part of the actual parameter file. White space must be used as a delimiter between values in a record (either tab or space). -99.99 indicates a meaningless value that is not used.

line 1	area_name	stratum_name	N	extrap_method	extrap_prm	ZmA_min	ZmA_max	p_absent	ZmA_curve_type	ZmA_curve_prm1	ZmA_curve_prm2	ZmA_curve_prm3	A_threshold	min_frac
line 2	soundwide	core	9	censused	-99.99	-99.99	-99.99	-99.99	none	-99.99	-99.99	-99.99	-99.99	-99.99
line 3	soundwide	flr	67	area	32363.2	-99.99	-99.99	-99.99	none	-99.99	-99.99	-99.99	-99.99	-99.99
line 4	soundwide	frn	1965	linear	2024	0.01	25	0.23	weibull	0.8	3.5	-99.99	1.5	0.2
line 5	soundwide	frw	426	linear	433.6	0.1	100	0.21	weibull	0.8	15	-99.99	3.5	0.25

Table 10-4. Descriptions of parameters in the stratum parameter file. The parameter names correspond with the example parameter file in Table 10-3.

parameter name	description
area_name	Not used. Reflects element of original design calling for the model to make estimates within sub-regions. This has not been fully implemented.
stratum_name	An arbitrary label for each stratum.
N	Total number of sample units (sites) in each stratum.
extrap_method	The extrapolation method used for strata that are sampled or "censused" for strata that are not sampled and are not subject to extrapolation. Domain = {area, linear, censused}.
ZmA_min	The minimum site eelgrass area allowed when area values are random variates drawn from a distribution. A value of -99.99 indicates a stratum where values are read in from file.
ZmA_max	The maximum site eelgrass area allowed when area values are random variates drawn from a distribution. A value of -99.99 indicates a stratum where values are read in from file.
p_absent	The proportion of sites in the stratum where eelgrass is absent. A value of -99.99 indicates a stratum where values are read in from file.
ZmA_curve_type	The type of distribution used when site eelgrass area in the population model is a random variate. A value of "none" indicates a stratum where values are read in from file. Domain = {normal, lognormal, exponential, Pareto, Weibull, none}.
ZmA_curve_prm1	The first parameter used to specify the exact distribution used when site eelgrass area in the population model is a random variate. See Appendix A, p.68.
ZmA_curve_prm2	The second parameter used to specify the exact distribution used when site eelgrass area in the population model is a random variate. See Appendix A, p.68.
ZmA_curve_prm3	The third parameter used to specify the exact distribution used when site eelgrass area in the population model is a random variate. Not used at this time.
A_threshold	Site eelgrass area threshold for assigning initial transect sets. If the site eelgrass area is above this threshold then a minimum mean eelgrass fraction criterion (next parameter) is applied. See discussion on p.15.
min_frac	Minimum mean eelgrass fraction of a transect set for assignment to a site in the population model. See discussion on p.15.

Table 10-5. Structure of the change scenario parameter file with example parameter values. The first record is a header and is ignored. The single record following the header indicates that the model run would apply one change scenario only.

scenario_name	subpop1-minZmA	subpop1-maxZmA	subpop1-prop-subpop-affected	subpop1-annual-decline-rate	subpop1-prop-transects-affected	subpop2-minZmA	subpop2-maxZmA	subpop2-prop-subpop-affected	subpop2-annual-decline-rate	subpop2-prop-transects-affected	site_change_CV	trans_change_CV
all_small_0.5%	0	10	1	-0.01	1	10	10000	0	0	0	0.3	0.3

Table 10-6. Descriptions of parameters in the change scenario parameter file. The parameter names correspond with the example parameter file in Table 10-5.

parameter name	description
scenario_name	Arbitrary label for the change scenario.
subpop1-minZmA	Minimum site eelgrass area delineating size class for sub-population 1.
subpop1-maxZmA	Maximum site eelgrass area delineating size class for sub-population 1.
subpop1-prop-subpop-affected	Prevalence of change among sites in sub-population 1.
subpop1-annual-decline-rate	Mean annual relative change rate for sites subject to change in sub-population 1.
subpop1-prop-transects-affected	Prevalence of change among belts of site belt model in sub-population 1.
subpop2-minZmA	Minimum site eelgrass area delineating size class for sub-population 2.
subpop2-maxZmA	Maximum site eelgrass area delineating size class for sub-population 2.
subpop2-prop-subpop-affected	Prevalence of change among sites in sub-population 2.
subpop2-annual-decline-rate	Mean annual relative change rate for sites subject to change in sub-population 2.
subpop2-prop-transects-affected	Prevalence of change among belts of site belt model in sub-population 2.
site_change_CV	Coefficient of variation of change values across sites subject to change.
trans_change_CV	Coefficient of variation of change values across transects subject to change.

The panel revisit design parameter file is specified by filename in the main parameter file (Table 10-2, line 8). The parameters in this file specify how panels of sites are revisited for re-sampling during the model period. Each stratum in the sampling design must have a panel revisit design specified. Strata subject to census rather than sampling are not represented in this parameter file. Two example parameter files are presented for contrasting designs (Table 10-7, Table 10-8).

Table 10-7. Structure of the panel revisit design parameter file with example parameter values for the original SVMP design with 20% annual site rotation. The first record is a header and is ignored. The total number of records will vary depending on the number of sampled strata and the number of panel groups in the design for each stratum. Split-panel designs (section 5.4, p.32) would have multiple panel groups per stratum. Only single digits are allowed in the revisit description strings ($n \leq 9$). For [X-n] designs, each panel must have the same number of sites, e.g., 2 for flr, 9 for frn and 3 for frw.

Line No.	Record Contents	Description
1	# Sampling Panel Parameters	Header record – ignored.
2	stratum flr 1	Start of parameters for the “flr” stratum which has 1 panel group.
3	flr 1 5-n 2	For the “flr” stratum, the first panel group has a revisit string of “5-n” with 1 site per panel.
4	stratum frn 1	Start of parameters for the “frn” stratum which has 1 panel group.
5	frn 1 5-n 9	For the “frn” stratum, the first panel group has a revisit string of “5-n” with 9 sites per panel.
6	stratum frw 1	Start of parameters for the “frw” stratum which has 1 panel group.
7	frw 1 5-n 3	For the “frw” stratum, the first panel group has a revisit string of “5-n” with 3 sites per panel.
8	END	End of the parameter file records to be read.

Table 10-8. An alternate panel revisit design parameter file with example parameter values for a 3-rotating panels design. In contrast to the design in Table 10-7, here the [1-2] design involves a fixed number of panels for each stratum (3) and the sample size (number of sites) is indicated for each panel. For example, 10-10-10 for the flr stratum (line 3) indicates that each of the three panels will contain 10 sites.

Line No.	Record Contents	Description
1	# Sampling Panel Parameters	Header record – ignored.
2	stratum flr 1	Start of parameters for the “flr” stratum which has 1 panel group.
3	flr 1 1-2 10-10-10	For the “flr” stratum, the first panel group has a revisit string of “5-n” with 1 site per panel.
4	stratum frn 1	Start of parameters for the “frn” stratum which has 1 panel group.
5	frn 1 1-2 45-45-45	For the “frn” stratum, the first panel group has a revisit string of “5-n” with 9 sites per panel.
6	stratum frw 1	Start of parameters for the “frw” stratum which has 1 panel group.
7	frw 1 1-2 15-15-15	For the “frw” stratum, the first panel group has a revisit string of “5-n” with 3 sites per panel.
8	END	End of the parameter file records to be read.

The transect sampling design parameter file is specified by filename in the main parameter file (Table 10-2, line 9). This is a simple parameter file with only three parameters specifying transect selection method, revisit plan and sample size (Table 10-9).

Table 10-9. An example transect sampling design parameter file.

Line No.	Record Contents
1	# parameters to guide transect sampling of sites – HEADER
2	SRS # type of transect selection; domain = {SRS, STR, SYS}
3	new # revisit plan; domain = {new, repeat}
4	13 # number of transects surveyed per site

10.4 Input Data Files

The application requires an input data file containing transect data for constructing the site belt models. The intent was that this input data would be generated from actual field observations but simulated data could also serve this purpose. A second input data file that specifies site eelgrass area values for one or more strata is optional. The values in each record of the input data files must be separated by white space (spaces or tabs).

A portion of a transect data input file is shown in Table 10-10. Each record after the header record contains a field that is the transect set number and the name of the stratum that the transect set should be associated with. This input data file needs to be structured so that each stratum has independently numbered transect sets with numbering starting at 1. The application assumes that the numbering within each stratum in the input data starts at 1 and is consecutive (incremented by 1) up to the last transect set whose set number is then equal to the count of transect sets for that stratum. The transect sets in the input data file do not need to be arranged in sequential order, but each set in the sequence must be present in the input data file (i.e., consecutive numbering is required). In addition, the application assumes that the transects associated with each stratum are grouped together.

Table 10-10. Structure of the transect data input file. The first record is a header that is ignored.

stratum_code	site_code	date_samp_start	transect_date	tran_num	tran_len_ft	zm_len_ft	tran_zm_frac	strat_trans_set_number
core	core001	8/15/2001	8/15/2001	1	12379.1	11638.9	0.94	1
core	core001	8/15/2001	8/15/2001	2	12008.7	11255.8	0.93	1
core	core001	8/15/2001	8/15/2001	3	12209.1	8765.7	0.71	1
core	core001	8/15/2001	8/15/2001	4	17941.0	13825.6	0.77	1
core	core001	8/15/2001	8/16/2001	5	12392.2	9282.4	0.74	1
core	core001	8/15/2001	8/16/2001	6	16102.6	10179.6	0.63	1
core	core001	8/15/2001	8/16/2001	7	15396.2	11949.8	0.77	1
core	core001	8/15/2001	8/16/2001	8	13367.4	10902.1	0.81	1
core	core001	8/15/2001	8/17/2001	9	14629.2	10225.1	0.69	1
core	core001	8/15/2001	8/18/2001	10	15418.9	11480.8	0.74	1

A portion of a site data input file is shown in Table 10-11. This file specified site eelgrass area for each site (*Zm_area_ha*) in strata that are specified by input data rather than random variates from specified distributions. This file also contains site area values (*site_area_ha*) for strata where extrapolations are made on an area basis (such as the SVMP flats stratum).

Table 10-11. Structure of the site data input file for strata specified by input data rather than random variates. The first record is a header that is ignored.

area_name	stratum_name	site_code	Zm_area_ha	site_area_ha
soundwide	flr	flats01	271.386912	1009.780058
soundwide	flr	flats03	132.2189321	782.5901258
soundwide	flr	flats05	503.4068514	943.0313971
soundwide	flr	flats08	35.18607366	428.0577638
soundwide	flr	flats09	1.179876764	685.8945384
soundwide	flr	flats10	0.484109079	719.8991156
soundwide	flr	flats14	317.5555229	682.788976
soundwide	flr	flats15	206.4995614	629.8213638

10.5 Output Files

The application can generate three different output files from each model run. One file is always generated while the other two are optional. The file containing the study area trend estimates from each model iteration is always generated. Additional output files containing the panel revisit schedule and the population model may be generated if the user opts for them at runtime. Each output file is written as a tab-delimited text file.

The format of the study area trend output file is illustrated in Table 10-12. The format of the eelgrass population model output file is shown in Table 10-13 and the format of the panel revisit schedule output file is shown in Table 10-14.

Table 10-12. Example of the study area trend estimates output file. Each record represents the results of one model iteration so this file would typically have thousands of records. The values in each record are, in order, the iteration number, the estimated trend in ha/yr, the estimated variance of the trend estimate in (ha/yr)² and the variance of the annual estimates of Puget Sound eelgrass area (ha²).

0	-156.234616	500.430931	25802.81828
1	-137.260699	858.091457	22850.86462
2	-193.85219	248.112653	21838.85143
3	-210.064337	348.491447	19536.34036
4	-185.403667	955.082803	20227.5662
5	-153.114887	705.930692	23184.21647
6	-209.142003	552.888389	17672.40342
7	-170.839599	371.204955	21445.5147
8	-127.697348	240.410717	17006.11652
9	-133.884812	1335.260461	21279.43799
10	-202.839664	449.31911	16527.04175
11	-166.958934	279.133101	25499.1528
12	-165.248674	124.761928	17403.5722

Table 10-13. Example of the eelgrass population model output file for a 20-year model run. Each record represents one site in one year. The file is sorted by site and then by year. This example table includes only the 20 records for each model year for the first site (flats01) followed by the first six records for the second site (flats03). The scenario_num and scenario_name fields are useful when a model run includes multiple change scenarios. The ZmA_ha field contains the site eelgrass area in hectares and the subpop field indicates which sub-population (size class) the site was assigned. The site_change_flag field indicates whether the site was subject to change.

scenario_num	scenario_name	site	year	area	stratum	ZmA_ha	subpop	site_change_flag	sample_poly_area
0	5_17.5ha_50%	flats01	0	soundwide	flr	271.3869	2	0	814.7079
0	5_17.5ha_50%	flats01	1	soundwide	flr	271.3869	2	0	814.7079
0	5_17.5ha_50%	flats01	2	soundwide	flr	271.3869	2	0	814.7079
0	5_17.5ha_50%	flats01	3	soundwide	flr	271.3869	2	0	814.7079
0	5_17.5ha_50%	flats01	4	soundwide	flr	271.3869	2	0	814.7079
0	5_17.5ha_50%	flats01	5	soundwide	flr	271.3869	2	0	814.7079
0	5_17.5ha_50%	flats01	6	soundwide	flr	271.3869	2	0	814.7079
0	5_17.5ha_50%	flats01	7	soundwide	flr	271.3869	2	0	814.7079
0	5_17.5ha_50%	flats01	8	soundwide	flr	271.3869	2	0	814.7079
0	5_17.5ha_50%	flats01	9	soundwide	flr	271.3869	2	0	814.7079
0	5_17.5ha_50%	flats01	10	soundwide	flr	271.3869	2	0	814.7079
0	5_17.5ha_50%	flats01	11	soundwide	flr	271.3869	2	0	814.7079
0	5_17.5ha_50%	flats01	12	soundwide	flr	271.3869	2	0	814.7079
0	5_17.5ha_50%	flats01	13	soundwide	flr	271.3869	2	0	814.7079
0	5_17.5ha_50%	flats01	14	soundwide	flr	271.3869	2	0	814.7079
0	5_17.5ha_50%	flats01	15	soundwide	flr	271.3869	2	0	814.7079
0	5_17.5ha_50%	flats01	16	soundwide	flr	271.3869	2	0	814.7079
0	5_17.5ha_50%	flats01	17	soundwide	flr	271.3869	2	0	814.7079
0	5_17.5ha_50%	flats01	18	soundwide	flr	271.3869	2	0	814.7079
0	5_17.5ha_50%	flats01	19	soundwide	flr	271.3869	2	0	814.7079
0	5_17.5ha_50%	flats03	0	soundwide	flr	132.2189	2	0	323.1299
0	5_17.5ha_50%	flats03	1	soundwide	flr	132.2189	2	0	323.1299
0	5_17.5ha_50%	flats03	2	soundwide	flr	132.2189	2	0	323.1299
0	5_17.5ha_50%	flats03	3	soundwide	flr	132.2189	2	0	323.1299
0	5_17.5ha_50%	flats03	4	soundwide	flr	132.2189	2	0	323.1299
0	5_17.5ha_50%	flats03	5	soundwide	flr	132.2189	2	0	323.1299

Table 10-14. Example of the panel revisit schedule output file for a 20-year model run. The schedule includes one panel for the core stratum that is scheduled every year and a set of 24 panels for the flr stratum scheduled to give a 20% annual sample rotation. Only a portion of the panel schedule is shown for the frn stratum at the bottom of the table.

panel	stratum	group	strat_panel	Nsites	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	core	0	0	9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	flr	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	flr	0	1	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	flr	0	2	2	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	flr	0	3	2	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	flr	0	4	2	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	flr	0	5	2	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	flr	0	6	2	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
8	flr	0	7	2	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
9	flr	0	8	2	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
10	flr	0	9	2	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
11	flr	0	10	2	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0
12	flr	0	11	2	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0
13	flr	0	12	2	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0
14	flr	0	13	2	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0
15	flr	0	14	2	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0
16	flr	0	15	2	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0
17	flr	0	16	2	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0
18	flr	0	17	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0
19	flr	0	18	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0
20	flr	0	19	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1
21	flr	0	20	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
22	flr	0	21	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
23	flr	0	22	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
24	flr	0	23	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
25	frn	0	0	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	frn	0	1	9	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	frn	0	2	9	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	frn	0	3	9	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

10.6 Execution

The model is a command line application. It can be executed from a Unix-like shell (e.g., Linux or Mac OS terminal) on a system compatible with the built executable file. When the application is started the user must specify the filename for the main parameter file to be used for the model run. The application will accept this filename as a command line argument. If the filename is not supplied as a command line argument then the user is prompted for the filename. The main parameter file must be located in a sub-directory of the current working directory named “prm_files”.

The executable file is named focusMC. If the executable is in the current working directory, it can be invoked as

```
$ ./focusMC
```

where the “./” prefix tells the system to look in the current directory “.” for the executable. If the current working directory is in the search path then this prefix isn’t necessary. If the

main parameter file to be used is named “main_prm.txt”, then the filename is supplied on the command line as

```
$ ./focusMC main_prm.txt
```

If the filename for the main parameter file is not supplied on the command line then the user is prompted for the file name as follows.

```
$ ./focusMC
$
$ Enter the name of the main parameter file:
```

The application proceeds only after the user supplies the filename and the file is successfully located. If this step is successfully completed, then the application proceeds to write status messages to the terminal as various processing steps are completed. There are two more prompts where user intervention is required to complete the model run. These are prompts asking whether the two optional output files should be written. The first prompt asks if the eelgrass population model is to be written to file.

```
$ Write population model to file? (y/n)
```

The response can be upper or lower case. If the user specifies that the file is to be written then a file with the name specified in the main parameter file (Table 10-2, line 10) is created. The next prompt asks whether the eelgrass population model is to be written to file.

```
$ Write panel schedule to file? (y/n)
```

If the user specifies that the file is to be written then a file with the name specified in the main parameter file (Table 10-2, line 11) is created. Regardless of the user responses about writing these two optional output files, the trend estimates from each model iteration will be written to a file with the name specified in the main parameter file (Table 10-2, line 12).



11 Guide for Developers

This section provides information intended to aid developers that need to work with the C code to eliminate bugs, modify the existing model or extend the model with new capabilities.

11.1 Application Structure

The application that implements the model is of moderate size (about 4,200 lines of code) but of sufficient complexity (27 functions) that a flowchart of the control flow is useful. Figure 11-1 gives a high-level view of the control flow that is managed in the main function which resides in a file named focusMC.c. Other than main(), the other functions reside in separate files named with the function name, e.g., read_main_prms() resides in the file read_main_prms.c. There are additional functions not shown in Figure 11-1 that are mostly involved with analysis and called from the analysis function.

An important component of the complexity in the application is the data structures used. The data structures can be multi-dimensional arrays combined with a nested use of pointers. In some cases the data structures became incrementally more complex during the development and some complexity reflects development expediency rather than optimal design. To aid working with these data structures, a guide to the more complex structures is presented in Appendix B (p.69).

11.2 Build Management

This section describes the specific tools and procedures used for code development and building of executables. All code development was done with VIM rather than with an IDE. To manage efficient compilation of the many code and header files involved, the make utility was used to manage compilation, linking and building. A makefile accompanies the code to be used with the make utility.

Depending on the system used for building, the makefile may need to be modified to ensure correct linkage with a math library. The existing makefile was used successfully on a number of systems during the development process. These systems included:

1. Sun C 5.11 compiler and Sun's included Make utility
Solaris 10 (SunOS 5.10) on Sun UltraSPARC III.
four SPARC v9 processors (64 bit)
2. Redhat Linux Enterprise 7

Intel Xeon CPU (64 bit) E5-2690
 GNU C compiler (gcc) version 4.8.3
 GNU Make 3.82

3. Ubuntu Linux 14.04
 Pentium M (32 bit)
 GNU C compiler (gcc) version 4.8.2
 GNU Make 3.81

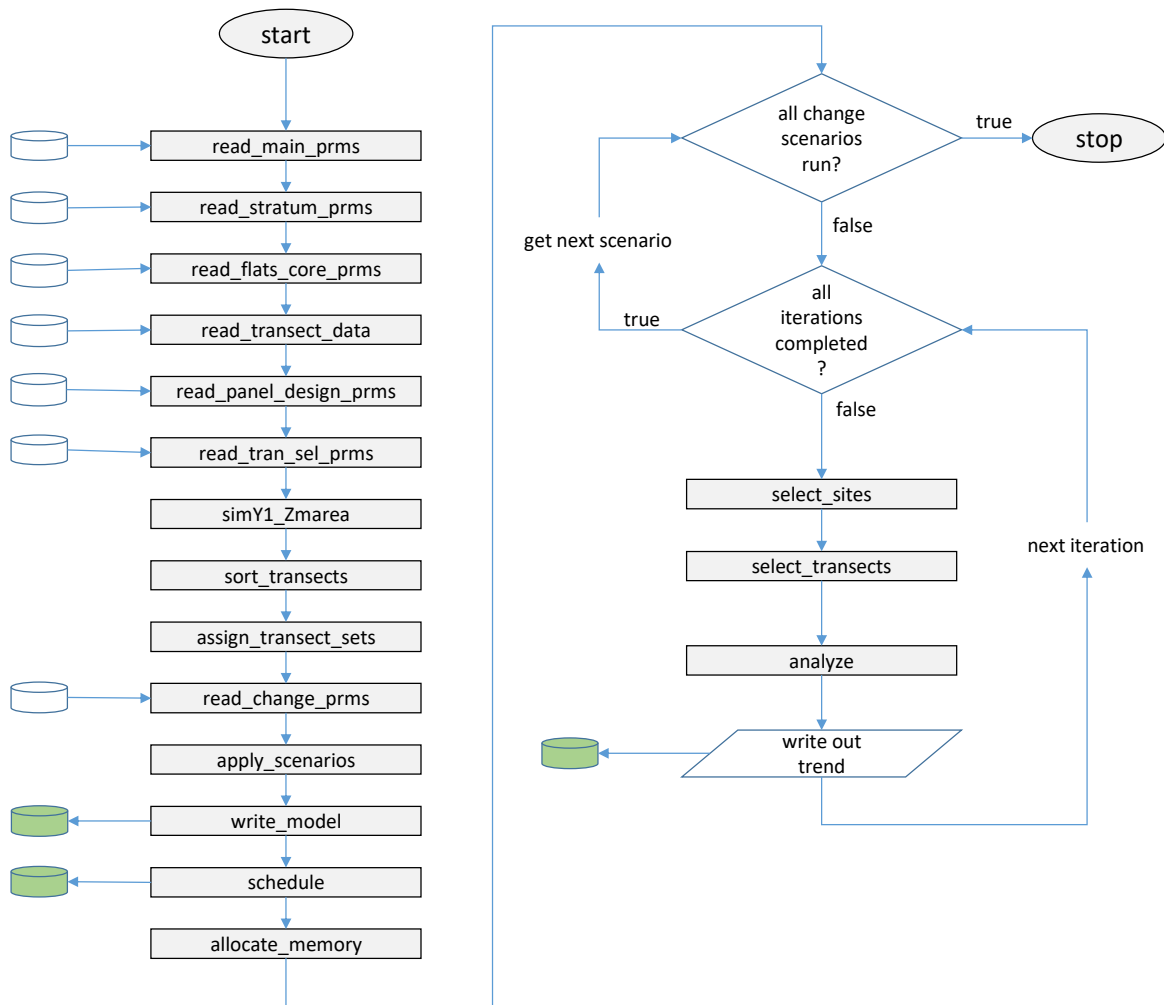


Figure 11-1. A high-level view of control flow in the application. Rectangles represent functions or subroutines and are labelled with the name of the function in the code. Cylinders represent external data – hollow for input data and shaded green for output data. The control flow is fully contained within the main function which passes control to each of the functions shown after which control returns to the main function. There are a number of lower level functions that are not shown. The simY1_Zmarea function simulates site eelgrass area values for strata where these values are determined as random variates. The sort_transects function takes the transect data read in from file and populates a data structure for easy manipulation of transect sets. The schedule function schedules panels for sampling through the study period. The other function names are self-explanatory.

11.3 Potential Future Development

There are a number of potential avenues for further development that could make the application more widely accessible or expand its capabilities. The application is currently available on a Unix-like platform which limits users to those familiar with such a platform. Two different approaches could make the application accessible to either those working on other platforms (e.g., Windows or Mac OS graphical user interface).

One approach would be to convert the application into an R library. Its not clear how practical this approach would be but it would have the advantage of making the code available to any platform that can run R. It would also have the benefit of being able to wrap the core model into other scripts with access to advanced statistical functions.

Another approach would be to port the code to Visual Studio on a Windows platform. Presumably the code could readily be converted to a Windows command line application. With additional effort to create a user interface the command line could be completely avoided in the use of the model.

A number of model capabilities were discussed in this report but were not implemented in the application (Figure 11-1). Certainly the application could be further developed to provide other capabilities not envisioned here. The level of development in the current version of the application reflects the present needs of the SVMP. Future SVMP design questions or the needs of other programs may provide the motivation for further development.

Table 11-1. Capabilities of the model described in this report that have been implemented in the current application (black font) and those capabilities that were described but not implemented (grey font). The STR and SYS transect selection methods have actually been implemented in code but these elements are shown in grey font to represent the effort needed to prepare transect input data to support these methods (i.e., spatially arranged transect data).

Model Puget Sound Population	Sampling Parameters				Analysis Parameters		
	Transect-level		Site-level		transect estimates	site estimates	soundwide estimates
	sample selection	revisit plan	sample selection	revisit plan			
Fixed change scenarios	SRS			new	abundance	abundance	annual abundance → trend
Dynamic change scenarios	STR	new	SRS	repeat	change	change	annual change (rel. previous) → trend
	SYS	repeat		revisit schedules	trend	trend	annual change (rel. initial) → trend trend (direct from site trend)



References

- Bailey, R.A. 1987. Restricted randomization: A practical example. *Journal of the American Statistical Association*. **82(399)**:712-719.
- Christiaen, B., P. Dowty, L. Ferrier, J. Gaeckle, H. Berry, J. Stowe and E. Sutton. 2016. Puget Sound Submerged Vegetation Monitoring Program: 2014 Report. Washington Department of Natural Resources, Olympia WA. http://file.dnr.wa.gov/publications/aqr_nrsh_svmp_report_2014.pdf
- Dowty, P. 2005. Assessment of Sound-Wide Change Estimates from Paired-Site Analysis: Submerged Vegetation Monitoring Project. Internal Report. Nearshore Habitat Program. Washington Department of Natural Resources.
- Dowty, P. 2006. *Reconstruction of Original 2000 SVMP Results, Revised 2000-2005 Results and New Bonus Material*. Internal report. Nearshore Habitat Program. Washington Department of Natural Resources.
- Dowty, P. 2005. *A Study of Sampling and Analysis Methods: Submerged Vegetation Monitoring Project at Year 4*. http://www.dnr.wa.gov/Publications/aqr_nrsh_samp_analysis.pdf
- Matsumoto, M. and T. Nishimura. 1998. Mersenne Twister: A 623-Dimensionally Equidistributed Uniform Pseudo-Random Number Generator. *ACM Transactions on Modeling and Computer Simulation* **8(1)**:3-30.
- McDonald, T.L. 2003. Review of environmental monitoring methods: Survey designs. *Environmental Monitoring and Assessment*. **85**:277-292.
- Schultz, S.T. 2008. Seagrass monitoring by underwater videography: Disturbance regimes, sampling design, and statistical power. *Aquatic Botany* **88**:228-238.
- 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.

-
- Stevens, D.L., Jr. and A.R. Olsen. 2004. Spatially balanced sampling of natural resources. *Journal of the American Statistical Association*. **99(465)**:262-278.
- Urquhart, N.S. and T.M. Kincaid. 1999. Designs for detecting trends from repeated surveys of ecological resources. *Journal of Agricultural, Biological, and Environmental Statistics*. **4(4)**:404-414.
- Urquhart N.S., S.G. Paulsen, D.P. Larsen. 1998. Monitoring for policy-relevant regional trends over time. *Ecological Applications*. **8(2)**:246-257.
- Van Sickle, J. 2012. Testing for a trend in total eelgrass area within Puget Sound. Final Report to Department of Natural Resources. Department of Natural Resources, Olympia WA.
- Von Neumann, J. 1951. Various techniques used in connection with random digits. Monte Carlo Methods. *National Bureau of Standards*. **12**:36-38.
- Wolter, K.M. 1984. An investigation of some estimators of variance for systematic sampling. *Journal of the American Statistical Association*. **79(388)**:781-790.

Appendix A Frequency Distribution Functions

The specific functions and parameters used to generate the frequency distribution curves shown in Figure 1-3 (p.6) and Figure 1-5 (p.12) are shown in the table below.

Distribution Name	Function	Stratum	Parameters
Lognormal	$f(x) = \left(\frac{1}{x}\right) \left(\frac{1}{\sqrt{2\pi\sigma}}\right) e^{-\left(\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right)}$	narrow fringe (frn)	$\mu = 1.5$ $\sigma = 2$
		wide fringe (frw)	$\mu = 2.5$ $\sigma = 1.7$
Weibull	$f(x) = \left(\frac{k}{\lambda}\right) \left(\frac{x}{\lambda}\right)^{(k-1)} e^{-\left(\frac{x}{\lambda}\right)^k}$	narrow fringe (frn)	$k = 0.8$ $\lambda = 3.5$
		wide fringe (frw)	$k = 0.8$ $\lambda = 15$
Exponential	$f(x) = \lambda e^{-\lambda x}$	narrow fringe (frn)	$\lambda = 0.4$
		wide fringe (frw)	$\lambda = 0.08$
Pareto	$f(x) = \frac{\alpha x^\alpha}{x^{\alpha+1}}$	narrow fringe (frn)	$\alpha = 0.14$
		wide fringe (frw)	$\alpha = 0.17$



Appendix B Application Data Structures

This appendix includes reference information on the more complex data structures used in the C application. Some of these data structures are complex enough that regular review of this information was helpful during code development. This information is included here to aid in future code development or debugging.

B.1 **schedule

Declaration: struct panel_prms_tag panel_prms[MXNSTRATA]

```
struct panel_prms_tag {  
    ....  
    short ** schedule;  
    ....  
};
```

Description: schedule contains the sampling schedule by sampling occasion for each panel. It is a member of each struct in the array of structs panel_prms (one array element for each stratum). It is defined in the struct tag panel_prms_tag (specified in header file definitions.h) as a pointer to a pointer to short (short **schedule). Memory is allocated in the schedule() function so that schedule can be utilized as a two-dimensional array schedule[ipanel][iyear] where each array element indicates a panel is sampled (1) or not sampled (0) for each year.

If the panel design includes multiple panel groups, the panels (first index) are ordered so that panels within a group are numbered consecutively and the groups are ordered as in the panel design parameter input file.

B.3 ****selected_trans[]**

Declaration: struct panel_prms_tag panel_prms[MXNSTRATA]

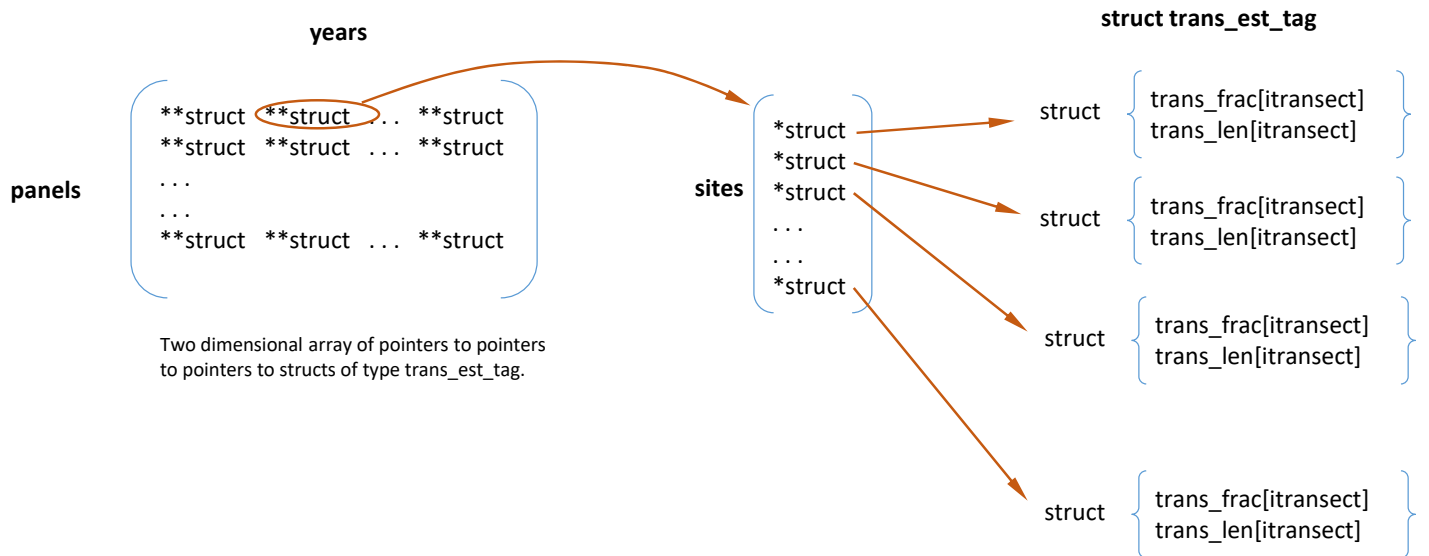
```

struct panel_prms_tag {
    ....
    struct trans_est_tag **selected_trans [MXNPANELS][MXNUMYEARS];
    ....
};

```

Description: selected_trans contains the transect “observations” for each site sampled on each sampling occasion. It is a member of each struct in the array of structures panel_prms (one array element for each stratum). selected_trans is a two-dimensional array of pointers to pointers to float. Memory is allocated for selected_trans itself at the declaration in the main program using the symbolic constants MXNSTRATA and MXNUMYEARS. Additional components of the data structure have memory allocated dynamically in two steps in the function select_trans(). First, memory is allocated for an array of pointers to pointers to float (“site pointers”) with one element for each site in the given panel. Each element of selected_trans is set to point to such a memory block. Second, memory is allocated for an array of float (“transect observations”) with one element for each transect in the transect set for a given site. Each element of the site pointer array is set to point to a transect memory block. Each transect array element is set to the appropriate transect time series value saved as part of the site_scenarios data structure.

A panel may not be sampled (i.e., have a sampling event) on each sampling occasion (year). In this case the element of the 2D array is set to NULL.

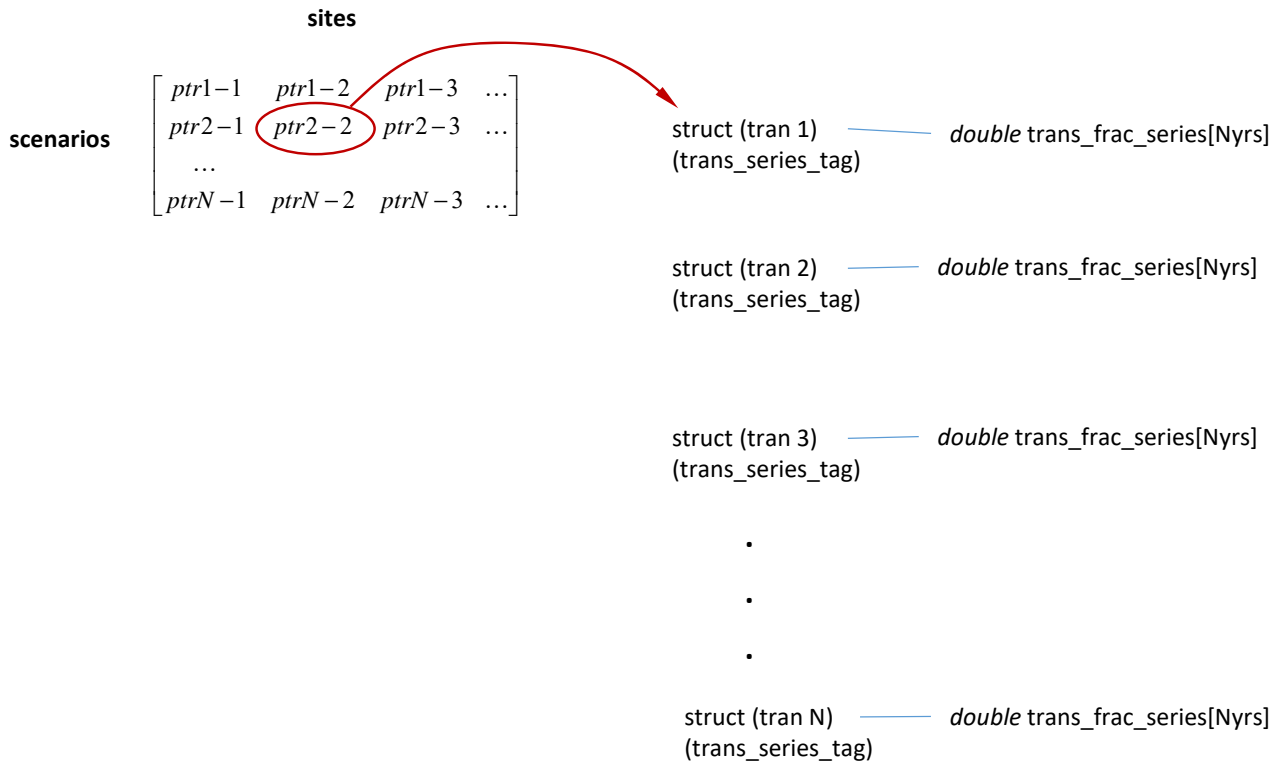


B.4 site_scenarios[][]

Declaration: struct transect_series_tag
*site_scenarios[MXNCHGSCENARIOS][MXNSITES]

Description: Organizes all transect data for each scenario, each site, each transect for pointers to structures of type transect_series_tag (specified in header file definitions.h). Rows represent scenarios and columns represent sites. Each array element is a pointer to a memory block containing a sequence of structures – one structure for each transect in the set for the given scenario and given site. Each transect structure contains an array of doubles representing transect fraction for each simulated year.

The symbolic constants MXNCHGSCENARIOS and MXNSITES are set in the definitions.h headers file will values set in the initial modeling work to be 30 and 2500 respectively.



B.5 sites_all[]

Declaration: struct site_data_tag *sites_all[MXNSITES]

Description: An array of pointers to structures of type site_data_tag (specified in header file definitions.h). Each array element (pointer) represents initial information for a single site in the study area – i.e., it points to a structure containing initial site information. Sites from all strata are represented. The symbolic constant MXNSITES is defined in the definitions.h header file and is initially defined to be 2500.

B.6 transect_sets[]

Declaration: struct transect_set_tag transect_sets[MXNSTRATA]

Description: Organizes the transect data read in from file into sets for assignment to model sites. This is an array of structures. Each element is a structure representing one stratum. In essence, transect_sets[] contains an arrangement of pointers that point to the array of transect data read in from file.

The structure element trans_sort[] is an array of pointers to pointers. Each element of trans_sort[] represents a different transect set. Memory blocks of different sizes are allocated to each pointer in the array depending on how many transects are in each set. This resembles a two-dimensional array, e.g. **trans_sort [iset][itrans]**, where each element is a pointer to a structure in transects[], an array of structures where each array element (structure) represents one transect record read in from file. The array trans-sort[] cannot however be addressed as a two-dimensional array because of the variable row size.

