CDFW/USBR/DWR/USFWS Monitoring Survey Design Team

Evaluation and Analysis of Five Long-Term Biological Monitoring Studies in the Upper San Francisco Estuary

2021 Final Report

Apr 1 2022

Prepared by: Aroon Melwani (Applied Marine Sciences), Michael Tillotson (ICF International), James Hobbs (CDFW), Steve Slater (CDFW), April Hennessy (CDFW), Brian Schreier (DWR), Kristin Arend (USBR), and Jeff McLain (USFWS)

Prepared for: 6 Agency Steering Committee

This page is intentionally blank

State of California – Natural Resources Agency
DEPARTMENT OF FISH AND WILDLIFE
Water Branch
P.O. Box 944209
Sacramento, CA 94244-2090
www.wildlife.ca.gov

GAVIN NEWSOM, Governor
CHARLTON H. BONHAM, Director

April 1, 2022

EVALUATION AND ANALYSIS OF FIVE LONG-TERM BIOLOGICAL MONITORING STUDIES IN THE UPPER SAN FRANCISCO ESTUARY

The attached report provides a detailed technical evaluation of five long-term monitoring surveys of pelagic fishes in the Sacramento-San Joaquin River Delta (Delta) conducted by the California Department of Fish and Wildlife (CDFW) under cooperative agreements with the U.S. Bureau of Reclamation (USBR) and the California Department of Water Resources (DWR). The surveys evaluated are:

- Spring Kodiak Trawl (SKT);
- Smelt Larval Survey (SLS);
- 20mm Survey (20mm);
- Summer Townet Survey (STN); and
- Fall Midwater Trawl (FMWT).

The purpose of the evaluation was to identify opportunities to improve utility, increase efficiency, and reduce redundancy amongst the five surveys. The evaluation was guided by an interagency steering committee including representatives from CDFW, USBR, DWR, U.S. Fish and Wildlife Service (USFWS), National Marine Fisheries Service (NMFS) and California State Water Resources Control Board (SWRCB). The evaluation is part of an ongoing effort to improve monitoring activities in the Delta. Based on the findings of the evaluation, the interagency steering committee recommends the following actions be taken:

1. **Assess Littoral Habitat** – Assess the need to add surveys for shallow, littoral habitats by conducting a special study to evaluate how suitability of shallow water (< 6 foot) habitat has changed as well as abundance relationship relative to deeper waters.

2. **Improve Abundance Estimates** - Implement design-based estimators to provide a more standardized method for estimating abundance across surveys and species.

3. **Improve Spatial Balance** - Balance survey effort spatially (by strata) within and between surveys to increase certainty and improve species detection. This should be considered in context of other surveys (Enhanced Delta Smelt Monitoring, San Francisco Bay Study, UC Davis) and should include:

Conserving California's Wildlife Since 1870

- Determining which regional strata are needed for which species;
- Increasing the number of stations in prioritized areas where population estimates are desired such as San Pablo Bay, Napa River, Suisun Marsh and Cache Slough;
- Decreasing the number of stations where appropriate;
- Standardizing new stations between surveys.

4. **Evaluate use of Random Stations** to understand bias of fixed design, including:
   - Evaluating results from 2021 FMWT Special Study comparing fixed and random sampling;
   - Implementing a special study, building off of the recent FMWT special study for other surveys.

5. **Evaluate Differences in Species Detection between Surveys** - Consider possible gear comparison studies or additional analytical efforts to resolve this issue. Also consider developing adjustment factors to allow catch from one net type to be directly compared to another net type.

6. **Examine Overlapping Sampling** – Conduct special studies to understand efficiencies of overlapping surveys to better understand the timing of transitions between surveys (e.g. STN and FMWT).

If you have questions or comments regarding the attached report, or the above recommendations, please contact Carl Wilcox at Carl.Wilcox@wildlife.ca.gov. Comments received will be considered by CDFW and the interagency steering committee during implementation, including development of future surveys and special studies.

Sincerely,

DocuSigned by:

Josh Grover

Joshua Grover, Chief
Water Branch

List of Acronyms

BO	Biological Opinion
CARI	California Aquatic Resources Inventory
CDFW	California Department of Fish and Wildlife
CESA	California Endangered Species Act
CMWT	Chipps Island Midwater Trawl
CPUE	Catch Per Unit Effort
CVP	Central Valley Project
DSLCM	Delta Smelt Life Cycle Model
DOP	Directed Outflow Program
DWR	California Department of Water Resources
EDSM	Enhanced Delta Smelt Monitoring
FESA	Federal Endangered Species Act
FLOAT	Flow Alteration
FMWT	Fall Midwater Trawl
FL	Fork Length
GRTS	Generalized Random Tessellation Stratified
HAB	Harmful Algal Bloom
IEP	Interagency Ecological Program
ITP	Incidental Take Permit
MAST	Management Analysis and Synthesis Team
NMFS	National Marine Fisheries Service
POD	Pelagic Organism Decline
SAV	Submerged Aquatic Vegetation
SFE	San Francisco Bay-Delta Estuary
SFEI	San Francisco Estuary Institute
SKT	Spring Kodiak Trawl
SLS	Smelt Larval Survey
SMSCG	Suisun Marsh Salinity Control Gate
STN	Summer Tow Net
SWP	State Water Project
SWRCB	California State Water Resources Control Board
USBR	U.S. Bureau of Reclamation
USFWS	U.S. Fish and Wildlife Service

List of Figures

List of Tables

Executive Summary

The purpose of this report is to present a summary of the approach and technical findings from the 2021 CDFW/USBR/DWR/USFWS Monitoring Design Team's evaluation of five long-term CDFW biological monitoring studies conducted in the Upper San Francisco Estuary (SFE). Readers interested in zooplankton monitoring should note that this component of the biological monitoring was not evaluated for this report. A forthcoming workplan will synthesize the Design Team's findings into proposed changes that resulted from this evaluation for redesign of the fish monitoring studies.

The objectives of this expedited review were to address concerns that the studies:

(1) lack sufficient statistical resolution to expand catch data for estimating population abundance
(2) do not provide metrics of uncertainty
(3) only sample a subset of available habitats
(4) target only a subset of species
(5) have overlapping spatial sampling
(6) have data gaps in certain surveys, stations, and times

The Design Team's review encompassed an integration of the five monitoring studies. The sampling stations were disambiguated to standardize locations across studies. A fish species list that comprised 15 of the most commonly caught species by age-class were the basis for evaluations.

The five studies were organized into two monitoring systems for evaluation.

Real Time Monitoring Program comprised of:

(1) the Smelt Larval Survey (SLS)
(2) the 20mm Survey (20mm)
(3) the Spring Kodiak Trawl Survey (SKT)

Status and Trend Monitoring Program comprised of:

(4) the Summer Townet Survey (STN)
(5) the Fall Midwater Trawl Survey (FMWT)
(2) the 20mm Survey.

The monitoring design systems were spatially stratified and evaluated for overlap, data gaps, and redundancy in their spatial and temporal effort. Bias in spatial sampling of fixed vs. randomized stations was evaluated by comparison of SKT and 20mm Surveys to the relevant Enhanced Delta Smelt Monitoring (EDSM) sampling. The well-established, design-based methodology for estimating population abundance and associated uncertainty was employed in monitoring design simulations to understand the sensitivity of the designs to changes in sampling effort.

The main findings from the review are:

1. Statistical Resolution: Status and Trends studies currently conduct sampling effort that is not standardized across regions, varies among studies, and is not balanced with water volumes for estimating regional abundances and uncertainty. **Spatial balance** could be sought with higher frequency of stations than currently sampled in prioritized areas where population estimates are desired. **San Pablo Bay, Napa River, Suisun Marsh**, and **Cache Slough** were identified as areas with the best opportunities for reduced uncertainty and improved species detections based on the existing datasets.

   The redesign of the Status and Trends studies should **evaluate use of random stations** to better understand sampling bias in the fixed design. Comparisons of SKT and 20mm Surveys to relevant EDSM monitoring were inconclusive when simulated with comparable effort. Although neither SKT or 20mm had consistently different abundances or standard errors relative to EDSM, the analyses were caveated by the differences in effort and sampling protocols. Future probabilistic monitoring should be conducted to test whether the sampling of fixed stations has pre-selected abundances to be higher or lower relative to estimates from unbiased, random samples. These design improvement experiments should be coordinated among **20mm, STN, and FMWT studies,** and build off the special study that was recently initiated in San Pablo Bay, Napa River, Suisun Bay / Marsh for the FMWT.

2. Uncertainty: **Design-based estimators** can provide a standardized method for estimating abundance across the studies and species. The Design Team found that the well-established abundance estimation approach was **highly correlated with the traditional indices** with the benefits of being **flexible and efficient** for calculating survey-specific regional estimates of relative abundance and uncertainties. In addition to achieving the important goal of uncertainty estimation, the methodology is also sufficiently generalizable to be readily **applied across surveys, species and regions of the Delta**, and to alternative stratification scales to those used in the review. The methods also have the benefit that they can be applied in **fixed or probabilistic spatial designs**, and thus will provide a consistent analytical framework to incorporate the data from design improvement experiments.

   The simulations of status and trend monitoring designs suggested a fundamental **limit to the size of uncertainty that can be accounted for by increasing sampling effort**. Uncertainty in monthly design-based abundance estimates for most species spanned orders of magnitude, which can only be **reduced by 25% by adding twice the current effort** or more in most regions.

3. Habitats: All five studies sample in **pelagic habitat**, which corresponds with the **majority of water volume** available to the gears. The Design Team considered that biological monitoring of shallow water habitat would require different survey methods that were outside the scope of the review.

4. Species: The **pelagic fish species** caught by the review studies are a function of the differences in **catchability and gear efficiency** associated with the sampling gear types. Regional catch patterns and trends revealed that SLS routinely catches the fewest Age-0 species, while the 20mm Survey gets the most. SKT often catches Age-0 and Age-1 species that are rare to other gear types. These **differences in species detection** can be largely attributed to the sampling **methods** and the **timing** of the studies. Catchability analysis indicated that seasonal overlap in SKT and STN sampling with the FMWT may be warranted. However, limitations of the study design combined with low sample sizes for some comparisons make definitive conclusions challenging. The Design Team identified gear comparison studies as a key theme for future study improvements.

5. Overlapping Sampling: All of the **review studies have overlapping spatial sampling**. Using a **standardized spatial extent** (the Review Regions and Strata) the review studies clearly have substantial overlap that can be associated to the information needed for **real-time vs. status and trends monitoring.** However, due to the different efficiencies of the gears, the Design Team recognized that **none** of the surveys could be **considered redundant**. The overlapping studies are beneficial to provide confidence in the data for understanding presence/absence, spatial patterns, and size: frequency changes.

6. Data Gaps: Design improvement experiments are needed to optimize the transition between surveys and gears. The key information need is to fill in gaps in understanding of **overlapping surveys** based on side-by-side data to inform the **optimal timing to shift from one gear type to another**. Integration of the study designs has shown that spatial extent of the review studies aligns with the regional use of Delta pelagic habitat for recruitment of several young-of-year pelagic fish species. However, the **limited understanding** of **gear efficiencies** prevents integration of these datasets. Understanding catchability on seasonal and regional scales will allow for merging of catch patterns between overlapping and adjacent surveys. For long-term integration of data for understanding seasonal patterns in relative abundance, the design of **gear comparison studies for STN to FMWT should be prioritized.**

Preface

The California Department of Fish and Wildlife (CDFW) and U.S. Bureau of Reclamation (USBR) prompted this expedited review of biological monitoring studies in the Delta to improve utility, increase efficiency, and reduce redundancy amongst the ongoing long-term studies. The CDFW and USBR initiated the design review to address concerns that the studies do not provide sufficient statistical resolution to expand catch data for estimating population abundance, do not provide metrics of uncertainty, only sample a subset of available habitats, target only a subset of species, have overlapping spatial sampling, and have data gaps in certain surveys, stations, and times.

In the fall of 2020, the CDFW and USBR formed a Steering Committee with the California Department of Water Resources (DWR), U.S. Fish and Wildlife Service (USFWS), National Marine Fisheries Service (NMFS) and California State Water Resources Control Board (State Water Board) to provide high-level guidance on survey redesign considerations. The Steering Committee commenced the interagency effort to identify opportunities to improve the design of five existing biological monitoring surveys conducted by CDFW to address current management needs. The surveys administered by CDFW pursuant to a cooperative agreement with USBR, are the Spring Kodiak Trawl (SKT), Smelt Larval Survey (SLS), 20mm Survey (20mm), Summer Townet Survey (STN) and Fall Midwater Trawl (FMWT). San Francisco Bay Study, another CDFW run survey was considered for this review, but given its role in compliance monitoring for the State Water Resource Control Board (SWRCB) Decision 1641 and likely inclusion in survey design efforts for Longfin Smelt monitoring requirement for the 2020 ITP to DWR, the Design Team with guidance from the Steering Committee decided to not include the Bay Study in this effort.

In January 2021, a CDFW/USBR/DWR/USFWS Monitoring Survey Design Team (here-on 'Design Team') was assembled to conduct the expedited review of five long-term fish monitoring program elements. The Design Team is led by Aroon Melwani (Applied Marine Sciences) and supported by Michael Tillotson (ICF International). Agency oversight is being provided by James Hobbs, April Hennessy, and Steve Slater (CDFW), Kristi Arend (USBR), Brian Schreier (DWR), and Jeff McLain (USFWS). Additional investigators were invited to participate as necessary.

The Steering Committee provides management, direction, and accountability for the design review process. The Steering Committee received bi-monthly progress updates from the Design Team on the evaluation and recommendations regarding potential modifications to the existing studies, while promoting cooperative, consensus-driven decisions. The Steering Committee has also managed outreach to ensure transparency. The Steering Committee tasked the Design Team with providing a technical evaluation of the five CDFW study designs, with a strong focus towards optimization, by promoting consistency and integration of existing and future study data, to improving the value and efficiency of the studies.

The Design Team integrated management objectives from the Steering Committee into conceptual models, guiding technical discussions and working with the Technical Support teams to conduct technical analyses to evaluate the monitoring designs. The Design Team's strategy to the review followed a systematic approach, generally aligned with the 'road map for biological monitoring survey design' (Reynolds et al., 2016). The problem definition for the biological monitoring studies under review can be summarized as:

"How to continue to provide a reliable source of clean water to 35+ million Californians and to support a multi-billion-dollar agriculture industry, while balancing water needs for all beneficial uses that protect human health and safety, halt ecological degradation, recovery of native fishes in decline, and minimizing impacts to natural resources?"

The Steering Committee developed three "Fundamental Objectives" that identify the primary management drivers for the biological studies under review.

1. Provide data and analyses that support management decisions intended to promote a healthy estuarine ecosystem.

2. Improve understanding of drivers of ecosystem change (e.g., climate change, habitat modifications and CVP and SWP operations).

3. Provide data and analyses that support management decisions intended to obtain a reasonable balance between fish and wildlife, water supply, and power generation in compliance with applicable laws.

To assess achievement of these fundamental goals, the biological monitoring surveys should facilitate five "Means Objectives":

1. Assess the long-term status and trends of the ecosystem and recruitment patterns for fish and zooplankton assemblages of the Bay-Delta by region, across seasons, and over years in the context of existing data sets.

2. Determine the presence/absence, estimated abundance and spatial distribution of native and non-native fish species.

3. Support quantitative modeling and predictive tools, so efforts can be advanced to predict the outcome of future Delta management scenarios.

4. Avoid duplication, and complement surveys conducted by parties other than CDFW.

5. Provide flexibility to adaptively mange while preserving the utility of long-term data.

Report Organization

Readers interested in an overview of the project scope and the five studies under evaluation should refer to the *Background* (Page 16)

Readers interested in the technical focus of the review should refer to the *Major Needs* (Page 22)

Readers seeking an overview of the analytical outcomes from the technical analysis should refer to the *Conclusions* (Page 28)

Readers interested in a synthesis of the findings from the review effort should refer to the *Major Findings and Design Improvements* (Page 74)

Readers interested in the fish study data integration should refer to *Appendix 1: Data Integration of Pelagic Fish Monitoring* (Page 89)

Readers interested in analyses to support the sample frame and standardization of index calculations should refer to *Appendix 2: Spatial and Temporal Evaluation* (Page 96)

Data users interested in the selected species list, species detection patterns and analyses of catchability should refer to *Appendix 3A: Taxonomic Comparison Among Studies and Appendix 3B: A Comparison of Relative Catchability and Size Selectivity* (Pages 112 and 132)

Readers interested in the technical approaches to evaluate statistical sensitivity of the sampling designs should refer to *Appendix 4 Stratified Sampling Designs for Estimation of Regional Abundance* (Page 150)

A glossary of technical terms has been included on Page 84.

1. Background

The CDFW's Bay-Delta Office in Stockton California has been conducting a series of biological monitoring surveys in collaboration with USBR and DWR to meet permit obligations to the State Water Resource Control Board (SWRCB) and USFWS-NMFS biological opinions for Delta Smelt and salmonids, and for incidental take permits issued by CDFW for operation of Central Valley Project (CVP) and State Water Project (SWP). Five of these long-term biological monitoring surveys are subject to the current analysis described here-in. Specifically, Smelt Larval Survey (SLS), 20mm Survey, Summer Townet (STN), Fall Midwater Trawl (FMWT), and Spring Kodiak Trawl (SKT) surveys.

Together these five studies constitute a substantial portion of CDFW's overall monitoring efforts for juvenile fish and zooplankton assemblages in the Upper San Francisco Estuary (SFE). The STN is the longest running (since 1959) of the five review studies, which was initially implemented to provide a recruitment index for Striped Bass (38.1-mm in length; *Morone saxatilis*) in Suisun Bay and the Delta. Soon after, the FMWT was initiated (since 1967), to better encompass the spatial and temporal distribution of Striped Bass recruits in relation to entrainment by State and Federal water projects. After the listing of Delta Smelt (*Hypomesus transpacificus*) in 1993 under the Federal Endangered Species Act (FESA), the 20mm Survey began in 1995, followed by the Spring Kodiak Trawl Survey in 2002, to document larval and spawning adult Delta Smelt populations and to inform real-time operations at the CVP and SWP in an attempt to avoid/minimize entrainment. In 2009 Longfin Smelt were listed as threatened under the State Endangered Species Act and the Smelt Larval Survey was initiated to track the distribution of larval Longfin Smelt to inform real-time operations . Although each of these studies was originally developed with its own set of methods and objectives, collectively the surveys provide data to reduce risk of individual loss, to track abundance and distribution to identify unhealthy trends and to provide data and information improving our understanding of the ecology of the estuary to support management and regulatory decisions aimed at protecting the estuary. Within this context, the five studies were considered together here to provide an organized approach to the evaluation, especially given that some of the objectives of the individual studies have become more aligned over time.

These monitoring surveys fulfill three broad management needs:

1. **Real-Time Operations** – Provide information for real-time operations decision making, including observations of biotic and abiotic conditions that inform weekly assessments and support evaluations of how different CVP and SWP operations may affect entrainment of listed species and other important indicator species.

2. **Status and Trends** - Characterize the condition and rate of change of condition of species and communities in response to freshwater outflow, water diversions and exports.

3. **Special Management Studies** – Provide Delta-wide and/or regional context for special studies designed to assess the impacts of specific management actions (i.e., Pelagic Organism Decline-POD,-Fall Low Salinity Habitat Study-FLaSH, Directed Outflow Project-DOP, Suisun Marsh Salinity Control Gate operations, habitat restoration, Yolo Bypass food action). More targeted sampling may be required to test efficacy of additional management actions.

Project Description

The CDFW Bay-Delta Office conducts a suite of long-term pelagic fish monitoring studies in the Upper SFE. These include the SLS, 20mm Survey, STN Survey, FMWT, and SKT Survey. Each of these studies deploy towed nets in open water to capture fish in the water column. None of the studies sample in shallow water (< 6 ft. depth), however, due to the size specifications of the nets and risks of entanglement. The primary goal of these studies is to characterize different life stages of young-of-the-year ("Age-0") fish and provide abundance and distribution information to inform annual recruitment success ("status and trends") and also inform entrainment risk ("real-time operations") via export facilities in the north and south Delta. The studies count and identify all fish and macroinvertebrates that are caught, measure length for the first 50 individuals of each fish species, calculate an index of relative abundance for several indicator species, and provide catch-per-unit effort data for all species. Studies also collect environmental data to inform habitat use of pelagic organisms. Environmental data includes water depth, temperature, turbidity, salinity, observation of harmful algal blooms (HABs), and conditions experienced during sampling (e.g., tidal current direction and tow direction). Several studies also collect zooplankton samples to inform food availability for pelagic planktivorous species. Most fish are zooplanktivorous while young, and food use is determined from samples provided to the CDFW Diet and Condition Study (implemented as a result of POD). These studies are part of IEP compliance monitoring but also coordinate data and samples sharing with other ongoing IEP management efforts, such as the USFWS Delta Smelt Life Cycle Model (DSLCM), Flow Alteration (FLOAT) Management Analysis and Synthesis Team (MAST), USBR Directed Outflow Program (DOP), and DWR Suisun Marsh Salinity Control Gate actions (SMSCG). These CDFW studies were initiated over time and independently to address specific objectives with the common task of monitoring fish abundance relative to water project operations, entrainment, freshwater outflow, and habitat restoration. Studies have evolved over time with adaptive management following review. The CDFW Bay-Delta Office also conducted special studies to better inform these studies, such as gear efficiency comparing catchability among gears. The following is a brief description of the five studies describing the gear, methods, sample frequency, geographic range, and key information generated.

The methods used by each survey are a function of the target species and habitats. All but the SKT conduct an oblique tow (sampling technique), where the net (Gear) is towed from near-bottom to the surface of the water column over 10 to 12 minutes (Sample Effort). The SKT samples at the surface to 1.8 m (6 ft.) of depth. The studies use different gear types with different mesh sizes optimized to retain specific life-stages/sizes of fish, but all provide raw counts of individuals caught and volume of water sampled to determine a Catch-Per-Unit-Effort (CPUE) standardized metric of relative abundance at each site. Some studies were implemented to target a single species and life stage (e.g., SLS targets larval Longfin Smelt, SKT targets adult Delta Smelt) or habitat type and are used to inform on a species community important to management (e.g., FMWT and STN). Most studies are important to the management of the endangered Delta Smelt, which is identified as an endangered species by the California Endangered Species Act (CESA) and threatened by the Federal Endangered Species Act (Figure 1), as well as the CESA listed Longfin Smelt. The studies are all described in detail with methods how, where, and when gear is used (https://wildlife.ca.gov/Conservation/Delta).

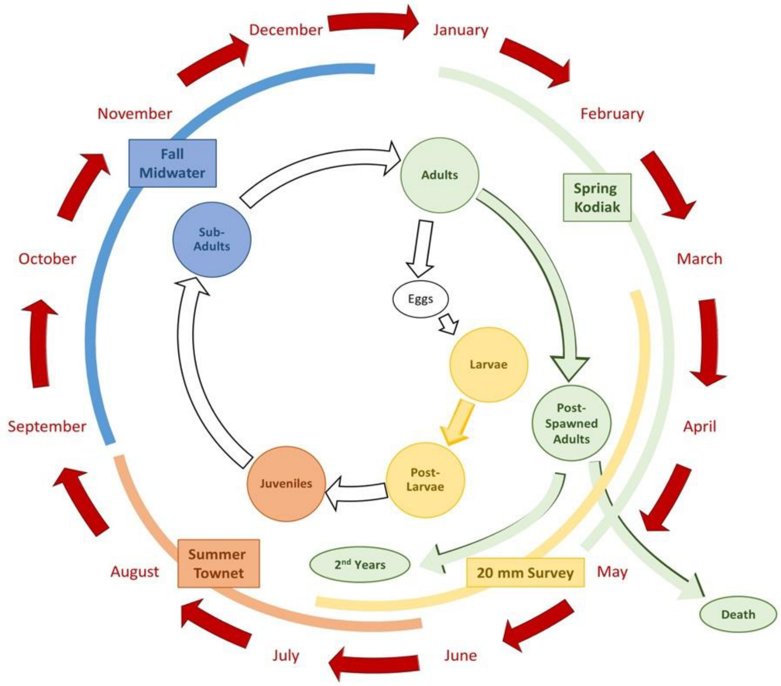


Figure 1. Example of Delta Smelt life cycle with life stages relative to CDFW sampling during the year. Adapted from Moyle et al. (2016)*.*

The timing of each study generally follows the seasonal life-cycles of most fish species occupying the Upper SFE; spawning occurring in late-winter and spring, larval and post-larval stages occurring from late-winter to early-summer, juvenile stages from late-spring to early-summer, and juvenile-sub-adults from late-summer to fall (Table 1). All five CDFW surveys sample geographically "fixed" sites along the axis of the estuary in open water habitat, with the exact geographic location of fixed sites (identified by a 3-digit station code) varying slightly between the five surveys. The number of geographically-fixed sites visited by each study is described in Table 2. Sample frequency (number of times and duration between revisit of fixed sites) varies among studies based on the range in sizes that the target species attain over the year. The 20mm and STN Surveys conduct bi-weekly sampling events and conduct 2-3 "replicate" tows at each fixed site to sample larval and juvenile fish. The mean length per bi-weekly sampling event is used to determine a size-based index of relative abundance. FMWT and SKT conducts monthly sampling events with a single un-replicated tow per station to sample juveniles of larger-longer lived species (e.g., Striped Bass) and adults of smaller-short lives species (e.g., Delta Smelt and Longfin Smelt). Catch from each monthly sampling event is used to produce a monthly index of relative abundance with the sum of monthly indices used to generate a seasonal index. The SLS survey conducts bi-weekly sampling events with a single un-replicated tow per station specifically to inform entrainment risk of Longfin Smelt via catch density at specific locations and was not tasked with producing an annual abundance index. The 20mm and STN studies include a mesozooplankton sample at each station collected concurrently with fish sampling. The FMWT conducts an additional zooplankton tow for meso- and macro-zooplankton at a subset of 32 stations that correspond to STN index stations (Table 2).

The geographic range (sample frame) of the studies represent open water habitats (>6ft depth) from eastern San Pablo Bay through Suisun Bay to the Delta, with the specific stations selected relative to location of the Federal and State water project pumping facilities in the South Delta. The origins of CDFW station locations was influenced by experimental work in the 1940s and 1950s, that led to the STN to sample summer recruitment of Striped Bass in the Delta and Suisun Bay since 1959. FMWT has the largest spatial extent of the five studies including San Pablo Bay and the Delta, which began in 1967 to look at expanded rearing habitat of juvenile Striped Bass, and at one time included stations in Central Bay and South San Francisco Bay. More recent studies have been initiated for Delta Smelt, and thus focus on stations through Suisun Bay and Delta, to evaluate proximity to entrainment in the South Delta. Stations have been added to studies over time to address evolving management needs, such as increased sampling frequency in the eastern Delta in response to drought, in the Napa River to expand Longfin Smelt coverage, and in the north Delta to provide better coverage of Delta Smelt habitat.

Information generated from these studies includes raw catch data, catch-per-unit-effort (CPUE), annual abundance indices, seasonal summaries, index memos, annual status and trends reports, and peer reviewed articles. Several of the studies contribute to weekly entrainment risk monitoring coordination with other agencies to inform water project operations during the spring entrainment period. Files and data visualizations are hosted via the CDFW Bay-Delta website https://wildlife.ca.gov/Conservation/Delta. Fish catch data from these studies is used by USFWS DSLCM modeling efforts to calculate population estimates for Delta Smelt to inform management actions. Recently, data from these studies have been the focus of major synthesis activities on zooplankton in the estuary (https://deltacouncil.ca.gov/pdf/science-program/2020-11-09-iep-93-zooplankton-integrated-dataset-metadata%20.pdf) with visualizations (https://deltascience.shinyapps.io/ZoopSynth/). These studies also have contributed to integrated water quality data sets covering decades of monitoring in the Upper SFE (https://portal.edirepository.org/nis/metadataviewer?packageid=edi.731.1). Long-term monitoring by these studies provides foundational understanding to fish life history and abundance trends in the estuary. Monitoring consistency is important to relate effort over time, but changes in sampling effort and measures collected over time have been made to adaptively manage the studies.

Table 1. CDFW Monitoring Survey Timing. A filled cell indicates a survey event is conducted. All of the surveys take approximately 3-7 boat days to complete a sampling event (SLS 3 boat days, 20mm 7 boat days, SKT 4 boat days, STN 6 boat days), except for FMWT that typically takes 10-14 boat days to complete an event. Sampling events are often conducted using 2 boats to accomplish sampling within a calendar week. Size range notes general target range of fish among the studies.

Surveys	**Size Range**	**Jan**				**Feb**				**Mar**				**Apr**				**May**				**Jun**			
		1	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
SLS	5-20mm	■		■		■		■		■		■													
20mm	10-50mm										■		■		■		■		■		■		■		■
STN	20-55mm																					■		■	
FMWT	30-120mm																								
SKT	40-150mm		■				■				■				■				■						
Zooplankton											■		■		■		■		■		■	■	■	■	■

Surveys	**Size Range**	**Jul**				**Aug**				**Sep**				**Oct**				**Nov**				**Dec**			
		1	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
SLS	5-20mm																						■		■
20mm	10-50mm		■																						
STN	20-55mm	■		■		■		■																	
FMWT	30-120mm									■				■				■				■			
SKT	40-150mm																								
Zooplankton		■	■	■		■		■		■				■				■				■			

Table 2. Projected number of fish and zooplankton samples to be collected by the five CDFW surveys in 2022.

Survey[1]	Fish					Zooplankton[2]					
	SLS	**20mm**	**STN**	**FMWT**	**SKT**	**20mm CB**	**STN CB**	***STN CB[3]***	**FMWT CB**	**FMWT Mysid**	***FMWT CB[3]***
1	44	141	112		40	47	40				
2	44	141	112		40	47	40				
3	44	141	112	122	40	47	40	*3*	32	32	*18*
4	44	141	112	122	40	47	40	*3*	32	32	*18*
5	44	141	112	122	40	47	40	*3*	32	32	
6	44	141	112	122		47	40	*3*	32	32	
7		141				47					
8		141				47					
9		141				47					
10											
11											
12	44										
13	44										
Total Samples	**352**	**1269**	**672**	**488**	**200**	**423**	**240**	***12***	**128**	**128**	***36***

[1] The surveys are bi-weekly (SLS, 20mm, and STN) and monthly (SKT and FMWT). The FMWT currently samples monthly September to December as surveys 3-6, as the study historically started survey 1 in July. Grey fields represent no sampling.

[2] Zooplankton tows collected concurrently with fish sampling includes mesozooplankton collected using a modified Clark-Bumpus ("CB") net or macrozooplankton samples collected using a "mysid" net.

[3] Suisun Marsh Salinity Control Gate (SMSCG) samples. FMWT added biweekly CB only tows (no concurrently fish or mysid net tows) during FMWT surveys 3 and 4 for SMSCG.

2. Major Needs

The five long-term biological monitoring studies collect data on the relative abundance and distribution of sensitive life-stages of fish and zooplankton assemblages in the SFE. The surveys are used to inform impacts of water management (i.e., exports) on fish populations, and in some cases to inform recovery of the species (1995 USFWS). This review and evaluation aimed to identify opportunities to improve the utility, increase the efficiency, and reduce redundancy in the five review study efforts. Due to time constraints, **zooplankton monitoring has not been considered as part of this effort.** At the onset of the review process, the Steering Committee established a Charter that memorialized the goal of evaluating the existing studies to improve biological monitoring data to meet management objectives.

The Design Team's evaluation process began with consideration for several prior reviews and analytical efforts that have identified potential improvements in the studies under review (Table 3). These prior fish study reviews encompass the key topics that the Design Team considered in its evaluation.

Table 3. Rationale, recommendations, trade-offs, and changes to CDFW survey designs based on prior study reviews and research.

Rationale and Scope	Recommendation /Issues Identified	Relevant Studies	Trade-offs to be considered	Changes Implemented	References
Status and trends are monitored properly	Abundance estimation bias from fixed stations	STN, FMWT, SKT, 20mm, SLS	The need to compare the different designs to ensure consistent interpretation (Tuckey and Fabrizio 2013). Dependent on objectives, fixed stations may be more effective (Mahardja et al. 2020).	none	(Peterson and Barajas, 2018; Polansky et al., 2019)
Monitoring of the estuary's fish community	Lack of coverage in non-pelagic habitat, may miss species important to the system or certain life stage	STN, FMWT, SKT, 20mm, SLS	Limited resources, design should be guided by objectives; gears limited to pelagic habitat	none	(Grimaldo et al., 2017; Honey et al., 2004; Stompe et al., 2020)
Key species in the fish community are monitored properly	Incomplete spatial coverage for key species	STN, FMWT, SLS	Limited resources may mean that sufficient coverage for most species may not be feasible	A number of new stations were added since the early 2000s	(Grimaldo et al., 2020, 2017; Lewis et al., 2020; Mahardja et al., 2021; Merz et al., 2011; Rosenfield and Baxter, 2007;

Rationale and Scope	Recommendation /Issues Identified	Relevant Studies	Trade-offs to be considered	Changes Implemented	References
					Stompe et al., 2020)
Address issue of imperfect detection	Incorporate catchability in reporting or design	STN, FMWT, SKT, 20mm, SLS	Accounting for catchability can come at a cost given limited resources (MacKenzie and Royle 2005; Reynolds et al. 2016)	None directly; did result in a suite of gear comparison studies to inform USFWS DSLCM	(Interagency Ecological Program, 2020; Latour, 2016; Mahardja et al., 2017; Mitchell et al., 2017a; Mitchell and Baxter, 2021; Newman, 2008; Peterson and Barajas, 2018; Polansky et al., 2019, 2018; Tamburello et al., 2019)
Species can adapt to changing environment by shifting distribution	Possibly interpreting a shift in fish distribution as change in abundance (due to fixed stations)	STN, FMWT, SKT, 20mm, SLS	See above.	none	(Enwright et al., 2013; Honey et al., 2004; Sommer et al., 2011)
Reporting of status and trends metrics	Abundance index calculation do not portray/quantify variability in data	STN, FMWT, SKT, 20mm, SLS	None? Multiple indices (CPUE, model-based estimates) can be reported at the same time	none	(Honey et al. 2004; Newman 2008; Thomson et al. 2010; Polansky et al. 2019)

The first major topic is whether the approach of sampling fish at **geographically fixed stations has inherently biased the indices** generated by the studies. In sampling theory, an individual site selected for sampling should be representative of the targeted habitat available and relative abundance of target species. However, if sample sites are selected because they represent “hot-spots” of abundance or high-quality habitat, then indices of abundance can be biased in a positive fashion relative to the true abundance of the species in the area. To avoid such biased sampling, a number of random site selection techniques have been developed to aid in selecting unbiased sites. If the targeted area or habitat for sampling is relatively uniform, then a simple “random” selection of spatially arranged sites can be an efficient and effective means to select unbiased sites. In this sense a random sample has an equal probability of being sampled. When species aggregate across a varying landscape in relation to some measure of habitat quality a simple random sample can be biased if selected in an area of high or low abundance. In such cases, the sampling area can be divided (stratified) into an area of known high or low abundance. Whether a spatially stratified probabilistic study design would provide “better” information than fixed station monitoring has been subject to debate (Polansky et al., 2019). In 2016, the USFWS Lodi office expanded its Delta Juvenile Fish Monitoring Program (DJFMP) by initiating the Enhanced Delta Smelt Monitoring Program (EDSM). EDSM implements a spatially balanced, probabilistic sampling program to obtain accurate, and spatially and temporally comprehensive data on Delta Smelt. Evaluating whether similar innovations of the CDFW monitoring design may be warranted, and if so, what study designs are needed to ensure consistent interpretation is of interest.

The second key topic is whether **all of the key species are being monitored appropriately** by the studies (Honey et al., 2004). This issue has arisen because of concerns that the lack of spatial sampling coverage in non-pelagic habitat may miss species important to the Delta ecosystem or at certain life stages (Grimaldo et al., 2017; Stompe et al., 2020). Prior research has suggested that Delta Smelt may migrate to spawning grounds in Cache Slough and in the Sacramento River Deep Water Ship Channel, which resulted in the addition of stations to these regions in 2009 (Baxter et al., 2010, 2008). More recently, expanding the downstream spatial coverage for key species has been shown to be warranted, particularly in the case of Longfin Smelt. Currently, 20-mm does not sample most of San Pablo Bay, and therefore lacks effort where Longfin Smelt larval and early juvenile populations reside during wetter years. CDFW special studies in the late 1990s examined Delta Smelt larval use of shallow waters versus channel waters. Although using different gears to those under review here, Aasen (1999) found no difference in larval use of habitats based on catch densities. The findings led to conclusions of open water channels being representative of densities found in nearby shallow waters. As considerable change in the SFE has occurred since the 1990s this topic warrants re-evaluation.

Two of the more challenging themes from prior reviews and recent literature relate to the need to **account for catchability** and movement patterns when estimating abundance from the fish survey data. Several studies have used SFE fish catch data to demonstrate that catchability (i.e., the probability of retention by a gear type given availability to the gear) varies significantly among species and surveys, indicating the need for incorporating catchability estimates into abundance calculation methods (Interagency Ecological Program, 2020; Peterson and Barajas, 2018). Recent efforts have been undertaken by the USFWS to translate the FMWT catch based data to design-based estimates that incorporate the probability of retention for Delta Smelt (Mitchell et al., 2017a; Mitchell and Baxter, 2021; Polansky et al., 2019). Similar evaluations are warranted to determine the data needs to make similar extrapolations for other CDFW studies and additional species.

Similarly, accounting for **species occupancy** has been a focus topic in recent work (Mahardja et al., 2020, 2017; Peterson and Barajas, 2018). For example, Delta Smelt occupancy has been associated with salinity and temperature, and detection probability driven by body size, sample volume, water clarity, and tidal stage (Hendrix et al., 2021). The **imperfect detection of survey nets** means the true occupancy state of surveyed regions will not always be observed, and therefore creates ambiguity about true changes in occupancy state. Similarly, accounting for the relation between **timing of sampling** and the **tidal cycle** with respect to species observation patterns warrants consideration. Tidal cycles in the SFE influence turbidity, salinity, and the availability of thermal and physical habitat (Enwright et al., 2013), and thus the distribution of pelagic organisms. Currently, sampling occurs exclusively during daylight hours and is conducted monthly or bi-monthly, regardless of tidal cycle. Evaluation of the timing of sampling relative to the tidal cycle, often referred to as tidal aliasing has been suggested as a method to normalize the geographic position of monitoring locations.

Lastly, reporting of status and trends metrics based on **CPUE and design-based estimators at regional scales is a known management desire** that has been identified by several prior reviews (Honey et al. 2004; Newman 2008; Thomson et al. 2010; Polansky et al. 2019). Optimizing sampling designs to provide estimates of variability across the Delta system, identify areas of high uncertainty, and provide additional insight to the variation in sampling methods over time were identified as high priorities for future redesign.

External Considerations

The scope of this review considered long-term monitoring elements outside the five prioritized surveys, including the EDSM and DJFMP, San Francisco Bay Study among others, to allow for a holistic evaluation of the SFE monitoring enterprise. This review and update to the five CDFW survey designs is intended to facilitate these other long-term programs in keeping up with a dynamic and evolving management framework in the SFE. New demands are being placed on the monitoring system, from an increased need to evaluate the effectiveness of management actions, to an increasing emphasis on the fish community in general and other species of concern like Longfin Smelt, to a need to provide more spatially explicit abundance estimates, that require thorough reflection and analysis to keep monitoring programs adapted to the data needs. New regulations, such as those in the 2020 Incidental Take Permit for Long-Term Operation of the State Water Project in the Sacramento-San Joaquin Delta (ITP), have requirements for new and expanded monitoring in the SFE that are being planned concurrently with the efforts detailed in this report. Therefore, early on in the process, the Design Team acknowledged that other inter-related surveys and survey design efforts may need to be incorporated to provide additional perspectives to the current evaluation. Monitoring that was originally designed for a set of management needs and species is often utilized for new needs and different species, which requires adaptation and rigorous consideration of the limitations of implementing such changes.

Review Objectives

The Design Team's overall approach to the review effort was to apply a fish community perspective to evaluate improvements to the studies in support of the management needs presented in the Charter. Two means objectives were used to focus the design review on evaluating the five studies for their ability to:

1. Determine the presence/absence, estimated abundance and spatial distribution of native and non-native fish species.

2. Assess the long-term status and trends of the ecosystem and recruitment patterns for fish assemblages of the Bay-Delta by region, across seasons, and over years in the context of existing data sets.

The review employed a systematic approach to evaluate the existing datasets, using quantitative analytical tools to explore monitoring program design changes. These technical approaches are summarized in the **Appendix Chapters** of this report. A strong focus was placed on catchability, abundance, spatial distributions, and long-term status. The Design Team acknowledged that given the currently low abundances of many species in the Upper SFE, redesign of monitoring for trends in recruitment of juvenile fishes would not be the focus of evaluation efforts. The Design Team considered it most efficient to review the studies grouped together, organized by habitat, seasons, regions, gear-types, and species/age groups (as appropriate) for review. By taking a data-oriented approach, all fish catch data was considered in the evaluation, across all of the sampled habitats, seasons, and years, for any species that were vulnerable to catch by the deployed gear in each of the studies. This broad organization across the surveys provided opportunities to integrate the quantitative elements of the existing methodologies and datasets, and allowed for potential changes to be evaluated across studies.

The Design Team recognized three means objectives in the Charter that were not the focus of quantitative analysis, but will be considered in the subsequent planning and assessment phase to adopt the findings from this review. Briefly, the revised monitoring designs shall:

1. Avoid duplication, and complement surveys conducted by parties other than CDFW.

2. Support quantitative modeling and predictive tools, so efforts can be advanced to predict the outcome of future Delta management scenarios.

3. Provide flexibility to adaptively manage while preserving the utility of long-term data.

In the sections that follow, readers will be presented with a general framework that the Design Team agreed upon to integrate the monitoring designs under review. The Conclusions walk through a summary of the technical outcomes from these evaluations focused around specific components of the monitoring designs:

1) Spatial Sampling – address concerns that the studies only sample a subset of available habitats and have data gaps in certain surveys, stations, and times;
2) Abundance Indices and Uncertainty – address concerns that the studies lack sufficient statistical resolution to expand catch data to abundances and do not provided metrics of uncertainty;
3) Species Detection and Catchability – address concerns that studies only sample a subset of available pelagic species

The Major Findings section then synthesizes the conclusions into a summary of outcomes to support improvements to the sampling designs. Summary tables are provided to facilitate an integration of the findings into potential design changes by study for effort planning and evaluation. A detailed discussion of the statistical methods and technical evaluations that supported these conclusions and findings are detailed in the Appendices.

3. Evaluation Conclusions

Monitoring Framework

Highlights

- Monitoring designs were organized into 'Real Time Monitoring' and 'Status and Trends Monitoring'
- Real Time Monitoring comprised of SLS, 20mm, and SKT
- Status and Trends Monitoring comprised of 20mm, STN, FMWT

The Design Team approached this review effort first, by focusing on an evaluation of the studies as a whole (framework). This was accomplished by implementing a data integration and standardization of the five monitoring survey datasets. This integration required a disambiguation of the individual stations to create a standardized dataset (Appendix 1. Data Integration) of spatial and temporal effort (Appendix 2. Spatial and Temporal Evaluation). The Design Team considered various options for constraining the temporal extent of data considered in this review; ultimately agreeing to focus on the post-POD period (2002-2019). A species list was assembled that could be applied across studies and timeframes based on the top-ranking species catch present among the five studies, supplemented with species of management interest (Appendix 3A. Taxonomic Comparison Among Studies). A comparison of relative catchability and gear selectivity of the review studies was performed based on a prior gear comparison study (Appendix 3B). The Design Team evaluations of abundance indices and statistical resolution was used to inform a holistic sampling design evaluation across studies (Appendix 4).

The Design Team identified two distinct 'systems' of monitoring from this integrated dataset (Figure 2). SLS, 20mm, and SKT provide data on bi-monthly or monthly time steps that support "Real Time Monitoring" at prescribed stations and areas proximate to the SWP and CWP facilities in the South Delta (USFWS, 1996, 1995). These studies provide data on specific species of interest at critical life stages, e.g., Delta Smelt and Longfin Smelt, that is focused spatially in specific subregional areas of the Delta. Samples are converted to data often at short turnaround times (i.e., within 72 hours) to inform water management decisions.

STN and FMWT are conducted bi-monthly or monthly and employ spatial sampling effort over wider spatial scales and with station frequency that could allow for regional inference. STN and FMWT studies are conducted during the summer through fall when many of the pelagic fish community indicators are present at sizes retained by these gears. Survey design includes regional water volumes ("strata") to expand station catch to abundance indices, but strata vary among the surveys. Additionally, through this evaluation, the 20mm Survey was recognized for sampling many of the same species caught by STN and FMWT, and thus could support monitoring needs for status and trends, in addition to the real-time information. For these reasons, STN, FMWT, along with 20mm were organized together to describe the monitoring design system representing "Status and Trends Monitoring". This organizational framework helped the Design Team to segregate study design and statistical questions within the context of the metrics relevant to each system. The redesign evaluations proceeded in a stepwise manner to evaluate the critical elements of the monitoring design. In the sections that follow, readers will be presented with an overview of the evaluations to support potential changes to the monitoring designs organized around the two Program areas.

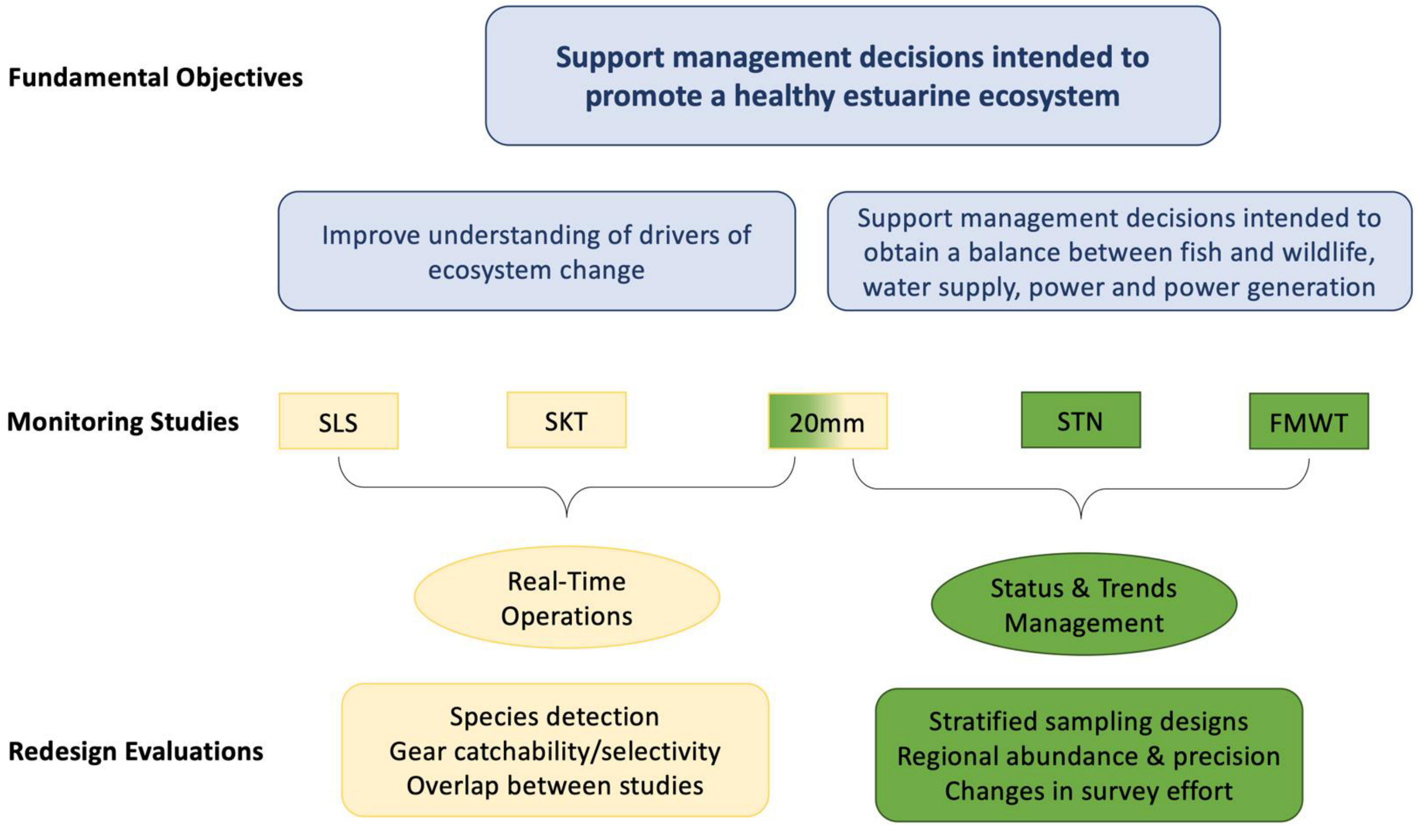


Figure 2. Organizational Framework to the Monitoring Studies and Redesign Evaluations

Spatial Sampling Evaluation

The Design Team evaluated the review studies from the perspective of spatial sampling design. Spatial sampling refers to when studies are undertaken to estimate the geographic distribution of an attribute, such as fish counts, abundance or presence/absence. The Design Team initiated its review of the studies by developing a sample frame that was used to conduct an evaluation of habitats sampled and the spatial and temporal effort.

The key sampling design concepts considered were:

1) regional and habitat scales relevant to the existing survey designs
2) spatial and temporal emphasis of the studies
3) bias associated with fixed-station sampling

Habitats Represented

Highlights

- All study locations align with pelagic habitat
- The majority of water volumes represented in the Review Strata correspond to pelagic waters
- Tidal marsh habitats are significantly under sampled by the pelagic studies

Readers interested in the technical approaches to support this section should refer to Appendix Chapter 2.

Long term biological monitoring designs for spatial sampling must be explicit about where sampling locations are placed in the universe of possible sites. The primary purpose of monitoring designs with a spatial component are that they provide data that allows for valid inferences for all spatial elements of the study. Spatial sampling can be undertaken using random, stratified random, or fixed sampling. Therefore, defining spatial parameters of the study designs retained the flexibility for potential design changes, while being explicit about the definition of strata and regional extent of the evaluation.

The Design Team developed a spatially-stratified sample frame to conduct the evaluation of the spatial sampling designs. The sample frame consisted of all the spatial locations that define the target population (i.e., locations where fish species are surveyed by the review studies) and used this information to evaluate the habitats, locations, and fish species represented in the sample frame. The target population of the monitoring designs was defined as:

- All waters within the extent of the existing monitoring stations that are represented by pelagic habitat (San Francisco Estuary Institute, 2017);
- All waters in 6 feet (1.8 m) depth or more (Wang et al., 2019). This depth criterion was selected to ensure that regional volumes used as the basis for extrapolations were explicit to depths that the currently deployed gear can affectively and safely access, as well as ensuring a consistent basis for integration across study data.

Due to inconsistent use, definition, and application of “strata” amongst the current studies, the Design Team chose to **“post-stratify”** each of the monitoring designs to provide a consistent basis to inform the evaluation. Here, **‘stratification’** refers to the spatial distribution of the sampling locations across the target resources of the Delta, and was used to evaluate whether areas of interest, including reporting scales have adequate effort. The upstream and downstream extent of the sample frame was *a priori* selected to represent the existing extent of monitoring in the review studies. The Design Team employed two stratification scales to evaluate the representativeness of the review studies; referred to throughout this report as ‘Review Regions’ and ‘Review Strata’ (see Maps in Appendix 2). These stratification scales were selected specifically for the purposes of this review in order to have a consistent basis to separate sampling locations across studies. The selected groupings of Regions and Strata were determined based upon on multivariate analyses of catch and environmental datasets starting from a GIS base layer of 30 EDSM Subregions (L. Mitchell, pers. comm March 15, 2021).

Five Review Regions (Figure 3) were used to represent designs using a high-level spatial organization of the existing stations. All of the monitoring designs have stations that correspond with the regions when viewed at this spatial scale. The Review Regions captured the dominant east-west gradient in catch composition and environmental conditions (refer to Appendix 2 for further details).

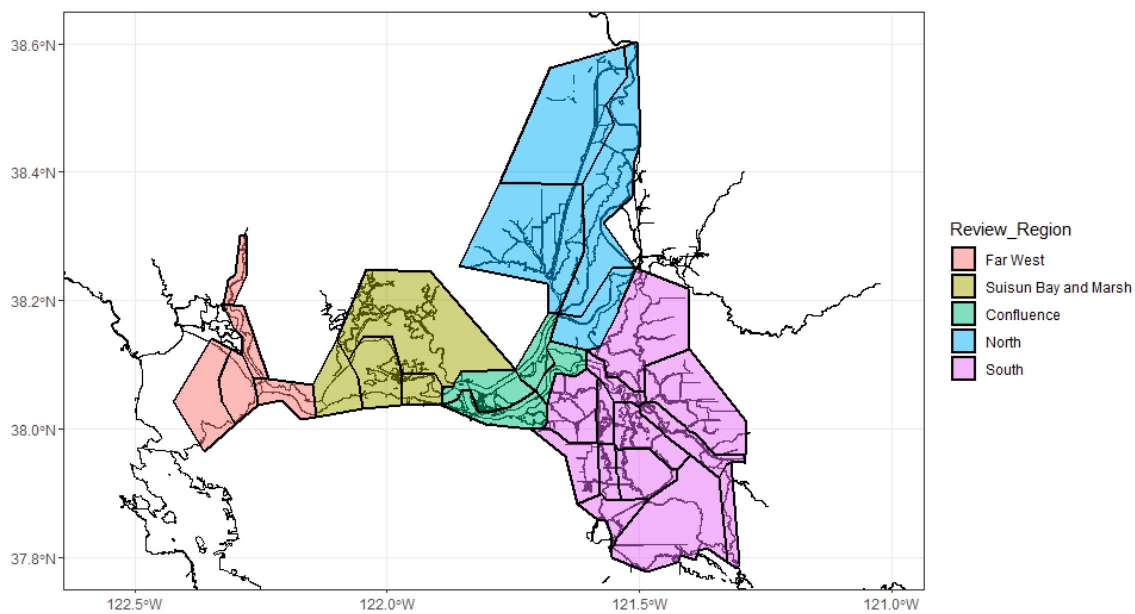


Figure 3. Overview Map of the Sample Frame Represented by the Five Review Regions

Ten Review Strata (Figure 4) were also used to represent designs using an enhanced (finer or smaller) spatial organization of the existing stations. Not all of the monitoring designs have stations that correspond with the strata at this spatial scale. Notably, the Far West, Suisun Bay/Marsh, and the North and South Regions were separated into two or more spatial groups. The Review Strata scale incorporates some of the local differences in catch composition that are likely associated with distinct habitat types (e.g., Suisun Marsh vs. Suisun and Honker Bays).

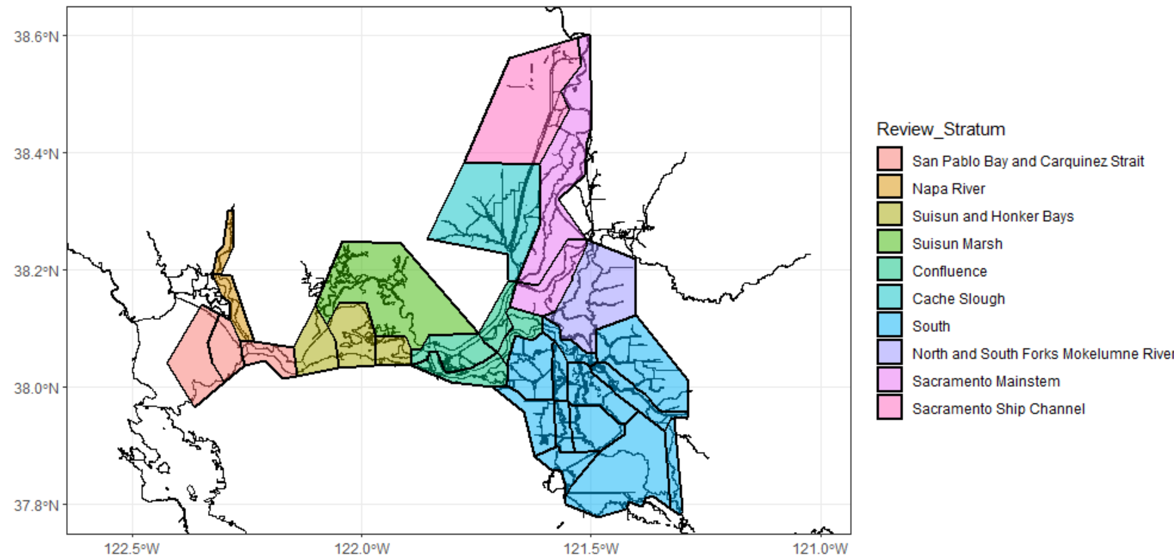


Figure 4. Overview Map of the Sample Frame Represented by the Ten Review Strata

Based upon these stratification scales, the water volumes represented in the sample frame were calculated. Further standardization of the water volumes was undertaken by determining the maximum depths that each gear type is deployed. The gear-specific water volumes were incorporated into calculations of abundances that follow. The key conclusion from this component of the evaluation was that nearly **all of the study locations in the review studies align with pelagic habitat** and that the **majority of water volumes** present in the Review Regions or Strata **correspond to pelagic waters** (Figure 5). The one extreme exception to this was in the South where a significant fraction of habitat is tidal marsh. However, most of this water volume was associated with areas that were qualified in the GIS layer due to the uncertainty in the extent of marsh habitats. The Design Team recognized that despite the five surveys good coverage of pelagic waters, tidal marsh habitats are significantly under sampled by these and other studies and could be habitats that warrant incorporation into long-term compliance monitoring studies.

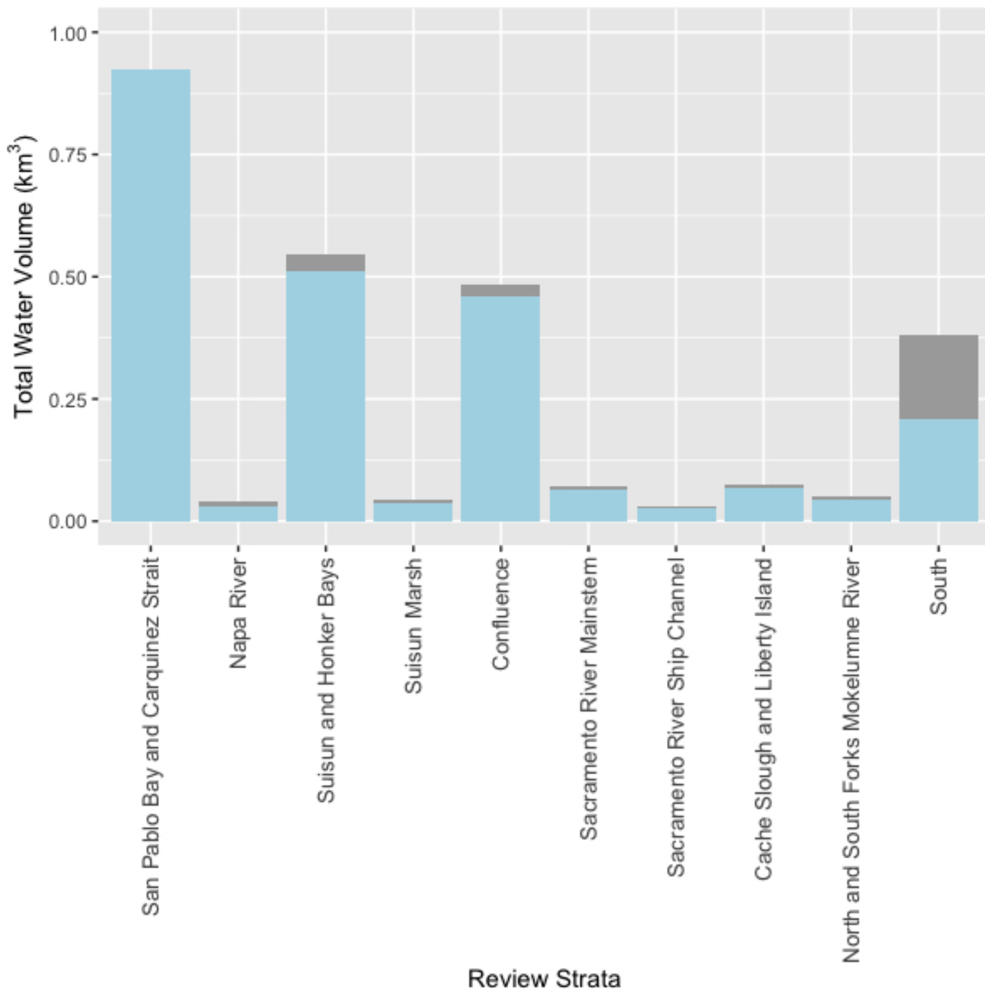


Figure 5. Pelagic (blue) water volume and total (grey) water volume in 10 Review Strata of the Upper SFE. The extent of the sample frame is detailed in Appendix 2.

Spatial and Temporal Overlap

Highlights

- Real-time monitoring studies (SLS, 20mm, SKT) focus sampling effort in the Suisun/Honker Bays, Confluence, and South Review Regions during the spring recruitment period and the most critical periods for entrainment
- FMWT samples the most stations over a broad spatial extent for status and trends
- Status and trends study designs do not have regionally balanced sampling effort with respect to water volume

Readers interested in the technical approaches to support this section should refer to Appendix Chapter 2.

To evaluate the potential for data gaps in the existing surveys, stations, or times, the monitoring designs were integrated into maps to summarize the spatial and temporal overlap in effort.

The Real Time Monitoring Program was defined as comprising the SLS, 20mm, and SKT. The spatial and temporal emphasis was recognized for supporting information needs for presence/absence and spatial patterns in proximity to the SWP and CVP facilities in the southern Delta. Not surprisingly, maps of the existing stations illustrate that the three studies overlap spatially (Figure 6), supporting the approach to organize them together.

All three studies have overlapping sampling effort in the Suisun / Honker Bays, Confluence and the South **with eight or more stations**. Stations are **lacking** in most of **San Pablo Bay**, **Sacramento Mainstem**, and the **Sacramento Deep-water Ship Channel** (Figure 6)**.** The rationale for these gaps in spatial emphasis is that these areas are all **far afield** from the pumping facilities in the South Delta. Additionally, a few areas of variable effort exist among the surveys, where the SKT survey does not survey the **Upper Napa River**, while SLS and 20mm sample throughout the Napa River stratum. Additionally, 20mm and SKT stations extend the upstream range of real time information in **Cache Slough**, while SLS does not.

Heatmaps summarizing the real-time monitoring effort reveal the temporal emphasis of surveys during the spring recruitment period (Figure 7). Recall that evaluations were limited to 2002-2019 to reflect current conditions since the start of the POD decline in pelagic fishes circa 2002. Combining the real-time studies in this way revealed a timeline that spans January through May, when between 1-5 surveys are conducted per month. March through May provides the most intensive sampling effort across survey methods, where 3 – 5 surveys are conducted concurrently, when SKT monitoring overlaps with SLS and/or 20mm. **This temporal emphasis supports real-time information during the most critical periods for entrainment.**

The Status and Trend Monitoring Program was defined as comprising the STN and FMWT, along with 20mm. The Design Team included the 20mm as part of this integration because of the spatial and temporal extent of the study design coupled with frequent catch of several species encountered in the STN and FMWT. The emphasis of status and trends monitoring was recognized for supporting information needs to estimate abundances and evaluate trends (i.e., change in relative abundance over time).

Maps of the existing stations illustrate that the three studies overlap spatially (Figure 6, Figure 8). All of the studies sample five or more stations in each of the Review Regions except for 20mm and STN that only survey **one station in San Pablo Bay**. 20mm focuses effort in Napa River, Suisun / Honker Bays, Confluence, and the South. STN overlaps with 20mm in Suisun Bay and Marsh, Confluence, and the South. The notable differences were the higher number of stations in the 20mm on Napa River and Cache Slough, and the additional effort in the STN study in the Sacramento DWSC.

FMWT provides the most intensive spatial sampling effort with 122 stations. The high frequency of stations along with the broad spatial extent of this study indicated that the status and trends studies could be integrated to build off this large dataset. However, the FMWT sampling effort (# stations and volume sampled) **is imbalanced with the available water volume represented in both the Review Region and Review Strata** compared to 20mm and STN, relative to regional water volume, particularly in the Far West and Suisun Bay/Marsh (Figure 9). **Balancing effort relative to available habitat resources (i.e., water volumes) is an important component of spatial monitoring** that should be considered for integrating and standardizing regional abundance estimates across studies. Sampling design changes to balance effort relative to pelagic water volumes should also consider the areas where uncertainty can be reduced with additional sampling (see "Statistical Resolution").

Sampling effort in the Status and Trend studies (20mm, STN, FMWT) has evolved over time, reflecting the dynamic and novel nature of the scientific information that is generated by these studies. Current monitoring (2002-present) across the three studies begins in March with 20mm, extends through the summer with STN, and ends in December with FMWT (Figure 10), with greatest effortFigure 10 during June and July. Unlike, the real-time studies that conduct monitoring that each overlap between gears, the status and trends studies do not always overlap in the same month. The Design Team identified that the **timing of overlap for status and trends studies requires consideration**. Only 20mm overlaps with STN in June and July, but STN does not overlap with FMWT. This has implications for catchability and integration of abundances that is discussed later.

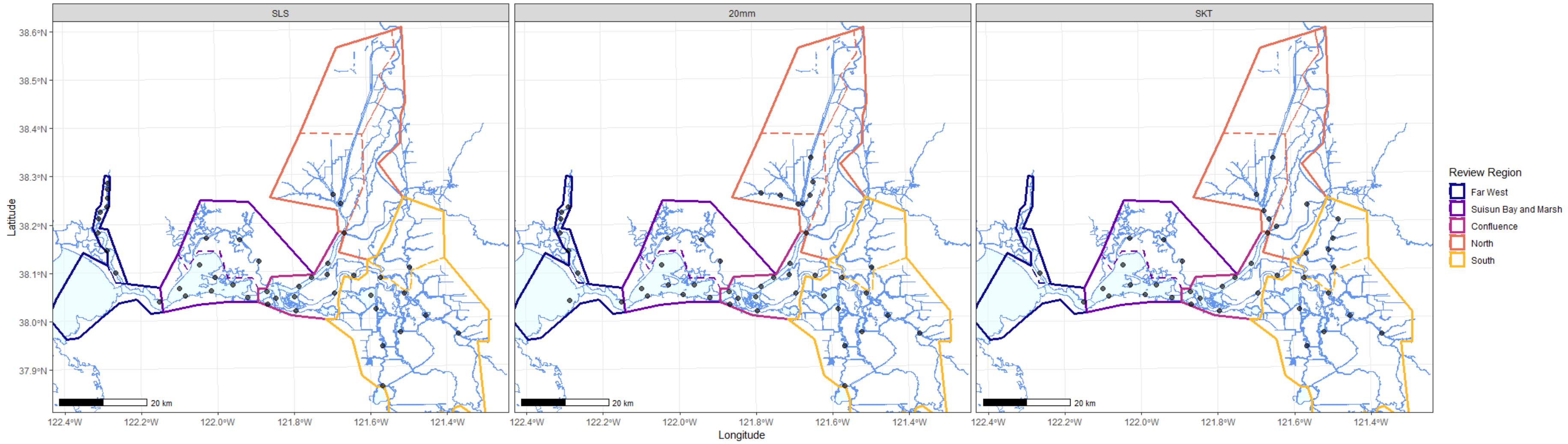


Figure 6. Maps of the SLS, 20mm, and SKT Sampling Locations

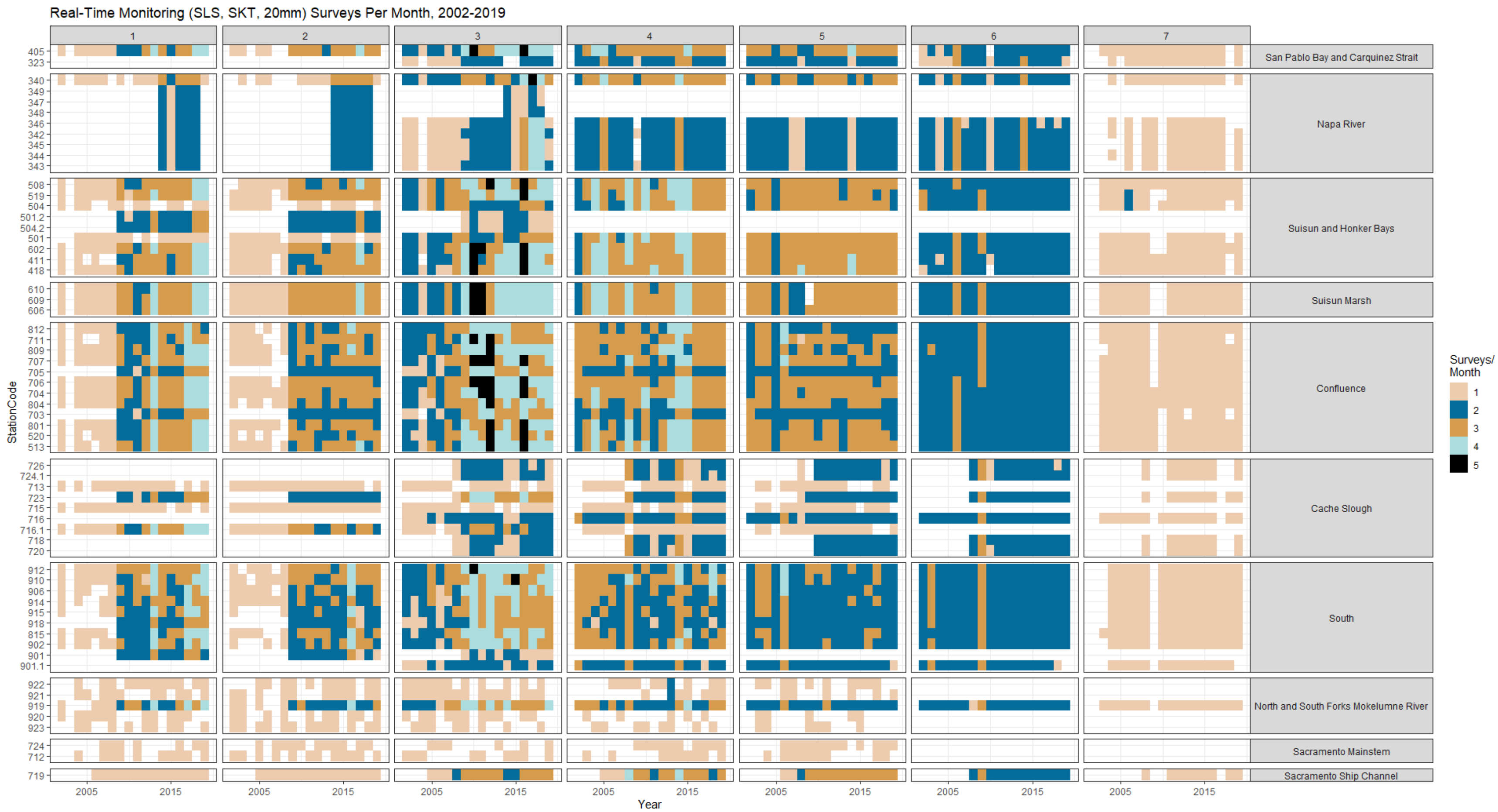


Figure 7. Heatmap of Spatial Sampling Effort in the Real-Time Monitoring Program

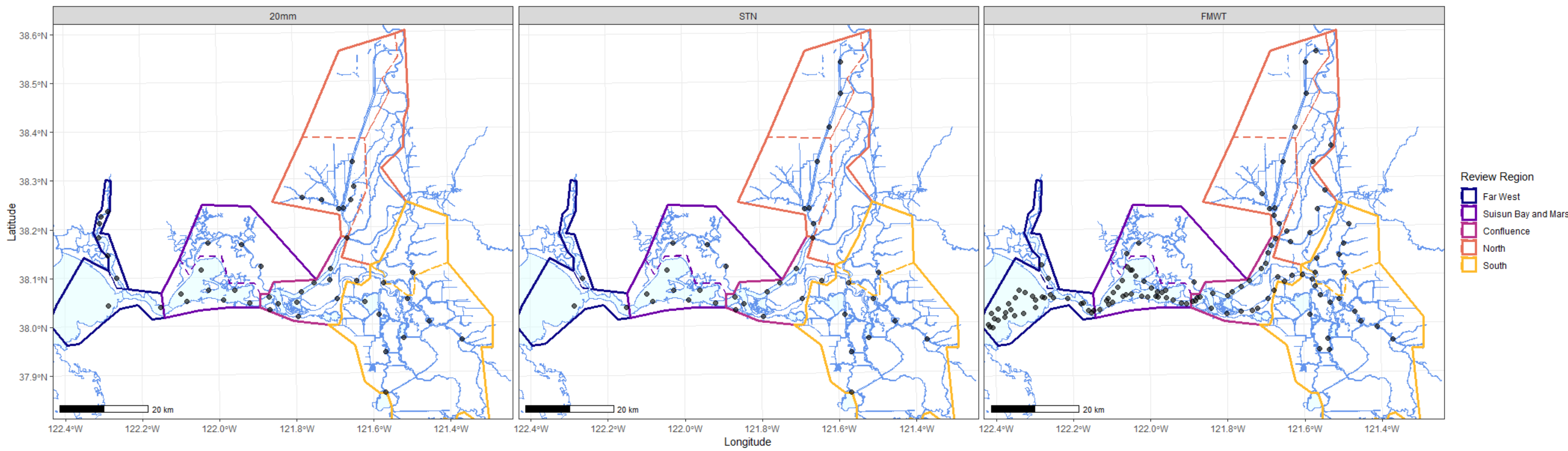


Figure 8. Map of the 20mm, STN, and FMWT Sampling Locations

Figure 9. Average monthly sample volume by study and stratum for the Status and Trends Program.

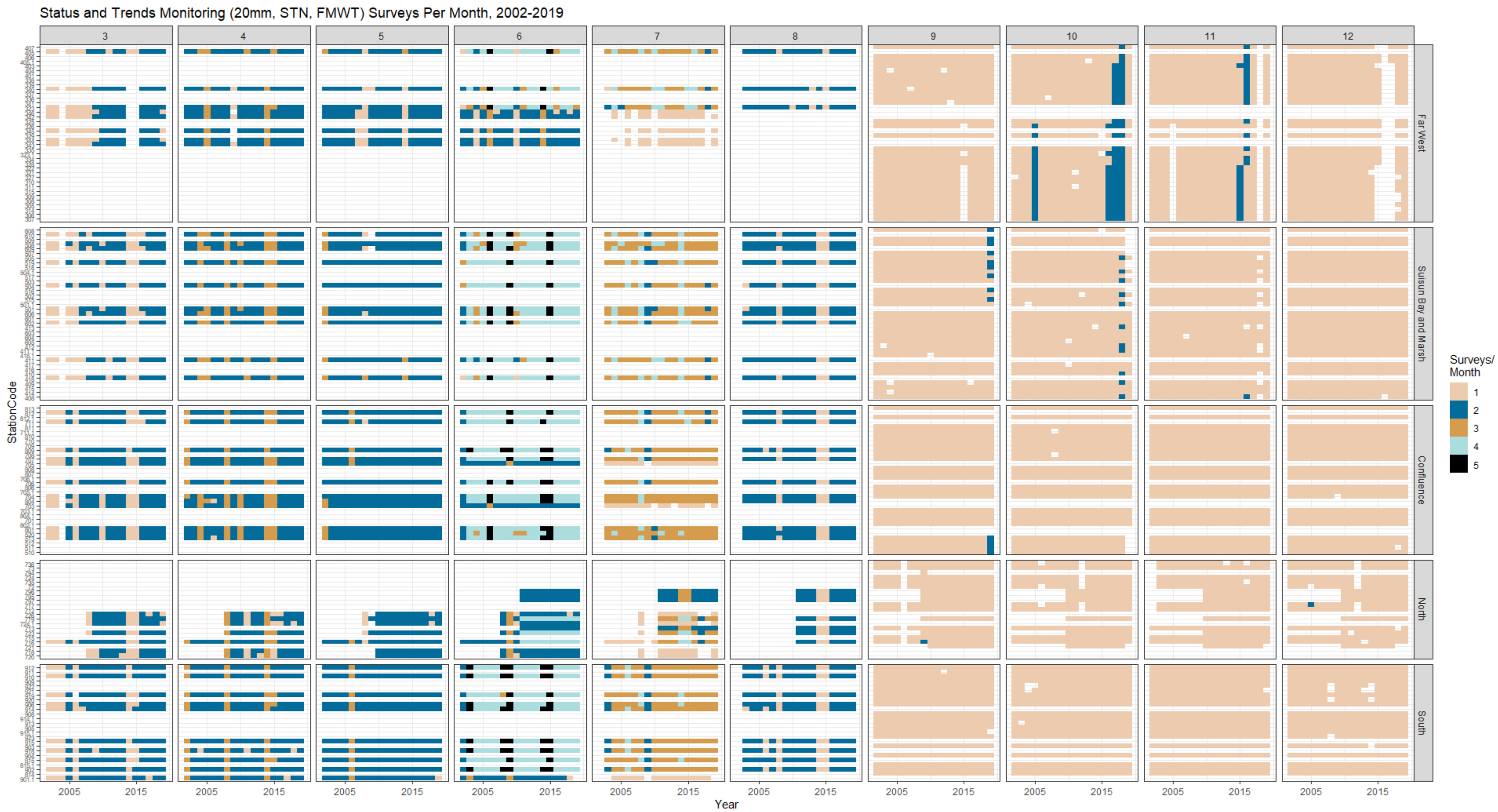


Figure 10. Heatmap of Spatial Sampling Effort in Status and Trend Monitoring Program

Fixed Station Bias: A Case Study Comparing Two CDFW Surveys to EDSM

Highlights

- Comparison of EDSM and CDFW 20mm and Kodiak trawl provided inconclusive evidence that a probabilistic study design could reduce uncertainty of abundance estimates
- The results for EDSM vs. 20mm must be caveated by the differences in surface water vs oblique sampling , respectively
- Increasing the number of SKT trawls did not necessarily reduce the abundance estimate uncertainty
- Differences in uncertainty between fixed and random designs needs further investigation in design improvement experiments

Readers interested in the technical approaches to support this section should refer to Appendix 4.

The Design Team evaluated potential bias in the abundance estimates resulting from the fixed monitoring designs. Potential bias in abundances has been raised in prior reviews due to concerns that fixed stations preferentially sample "hot-spots" of catch or high-quality habitat resulting in higher abundance estimates than would result from an unbiased sample. To address this question, the abundance estimates for the CDFW SKT and 20mm surveys were compared to the estimates generated by the stratified random (GRTS) monitoring design employed in EDSM sampling, which use the same sampling gear. The Design Team aimed primarily to determine **if the probabilistic design would result in** smaller standard errors (i.e., **reduced uncertainty**) or substantially higher or lower mean values (i.e., **bias**) compared to the fixed CDFW monitoring designs. The abundances were calculated using the same standardized methodology used to evaluate design-based estimates in the review studies (discussed in the following section on Abundance Indices and Uncertainty). Only regions, years and months in which EDSM and CDFW surveys occurred together were included the analysis. The resulting set of comparisons was relatively small (matched pairs for 20mm = 42; SKT=75). It must be acknowledged that EDSM monitoring exclusively samples fish in the surface waters, while the CDFW 20mm survey tows nets obliquely throughout the water column.

For the 20mm gear comparison, effort was reasonably similar between the CDFW and EDSM studies in most strata apart from the Napa River and Sacramento Mainstem/ Ship Channel. The comparison of standard errors provided **little evidence that a probabilistic study design reduces uncertainty of abundance estimates** (Table 4). In fact, the opposite appeared to be the case, with the CDFW 20mm survey standard error being lower in the majority of comparisons (i.e., <50%) for all species, except Striped Bass (Figure 11). The smaller uncertainties of the CDFW study was associated with higher abundances for most species, providing an **indication that a probabilistic study design generates lower abundance estimates**. However, these results must be interpreted cautiously because although the surveys use the same net type in the same areas and months, the EDSM study conducts only surface tows in contrast with the oblique tows of the CDFW 20mm survey. Therefore, the lower abundance estimates by EDSM may be associated with fish that move into deeper water during the juvenile life stage. In contrast, the Kodiak trawls conducted by both the CDFW and EDSM studies are surface tows, and so should be more directly comparable.

Table 4. Comparison of EDSM and CDFW 20mm Catch and Design-Based Standard Errors

Species	Total Catch 2017-2020		% of Cases EDSM 20mm SE < CDFW 20mm SE
	EDSM 20mm	CDFW 20mm	
Delta Smelt	42	153	24%
Longfin Smelt	571	3,596	27%
Prickly Sculpin	216	765	36%
Striped Bass	6,987	20,159	67%
Threadfin Shad	27,866	3,211	39%
White Catfish	9	65	30%
White Sturgeon	10	39	6%

For the Kodiak trawl gear comparison, effort was significantly unbalanced between the studies, with EDSM expending up to **20-times more effort than the SKT**. Therefore, in addition to the comparison of the complete SKT and EDSM datasets, a sensitivity analysis was conducted in which random EDSM tows were repeatedly sampled (1,000 iterations) with spatiotemporal effort matched to that of the SKT. The results of the comparison were variable, but generally indicated that **the EDSM survey produced abundance estimates with lower standard errors (**Table 5). This pattern appears to result at least in part from the **randomized sampling locations** (or some other unaccounted-for difference between the studies) and not simply the larger EDSM sample sizes. In many cases, the standard errors of the resampled EDSM estimates were lower than if all EDSM tows were considered. This supports the conclusion that **simply increasing the number of tows conducted will not necessarily reduce the uncertainty of abundance estimates** that can be derived from these trawl survey data.

Table 5. Comparison of EDSM Kodiak and SKT Catch and Design-Based Standard Errors

Species	Total Kodiak Trawl Catch 2017-2020			% of Cases Resampled SE < SKT SE	% of Cases Full EDSM SE < SKT SE
	SKT	EDSM (Simulation Mean)	EDSM (Full)		
Chinook Salmon	313	87	1,202	55%	75%
Delta Smelt	300	21	216	65%	81%
Longfin Smelt	33	31	463	39%	44%
Northern Anchovy	676	185	17,306	44%	6%
Steelhead	87	8	111	58%	69%
Striped Bass	72	12	175	73%	63%
Threadfin Shad	6,816	943	43,526	81%	81%

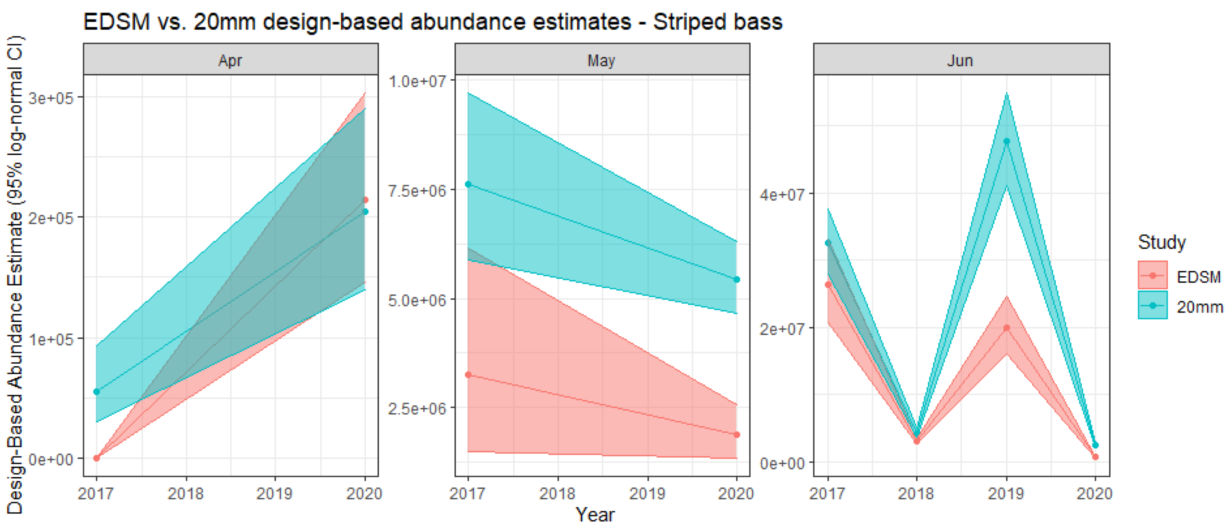


Figure 11. Example of Abundance and Uncertainty Estimates from CDFW and EDSM 20mm Surveys. Note that EDSM conducts surface tows and 20mm conducts oblique tows.

Abundance Indices and Uncertainty

Highlights

- Design-based estimator approach for estimating abundance can be applied consistently across studies and species
- Design-based estimators incorporate sampling design water volumes, survey volumes, and associated uncertainty
- Overlap between status and trends studies can help understanding catch patterns throughout the year and changes in size-frequencies
- Integrating abundances across the status and trends studies seems an achievable long-term goal

Readers interested in the technical approaches to support this section should refer to Appendix Chapter 4.

Fifteen species were selected for use in the current evaluation, which represented many of the highest ranked species by total catch between 2002 and 2019 (Table 6). Age-0 fish are almost exclusively represented in these data. For all species except Delta Smelt and Threadfin Shad, the Age-1 group comprised less than 10% of all individuals. Age-1 Delta Smelt were captured almost exclusively in the SKT; a survey specifically designed to capture adult Delta Smelt during their spawning season. When SKT samples are excluded, less than 1% of fish sampled since 2002 have been Age-1.

Many potential approaches exist for converting raw catch data to relative or absolute measures of abundance. The currently employed methods vary between studies and between species within a study, but in general, an indexing approach is used, where catch-per-tow at each station is expanded based on a water volume represented by that station. Expanded values are then summed across some defined set of stations and dates, depending on the survey and species in question. The limitations of the current approach have been well documented (e.g., Newman, 2008; Polansky et al., 2019). Most critically, the current approaches of relative abundance index reporting **lack specific metrics of uncertainty,** and generally ignore variation in tow volume despite the availability of such data. The Design Team's approach to evaluate abundance metrics in this review reflects the desire to address these limitations. The need to consider a broad range of survey design-options necessitated a computationally efficient approach while the application of the estimator to species with diverse life-histories, distributions and abundances required a general and flexible approach.

Model-based approaches were initially considered because of some beneficial characteristics to abundance estimation, and for consistency with the methods of recent review efforts. However, design-based approaches to abundance estimation are also advantageous due to their relative simplicity, fewer assumptions and the ability to be estimated with sparser data (Newman, 2008). **Trawl catches for many species in the Delta contain a high frequency of zeros – but also rare instances of large catches** – and can thus present a range of challenges for model fitting. In the context of evaluating a broad range of survey designs across multiple species, the relative efficiency of a **design-based approach outweighed the potential benefits**

Table 6. Total frequency of the species selected for evaluation in the review dataset (2002-2019). Three species were included for representativeness that fell outside the top 15 overall rank.

Species Category	Common Name	Scientific Name	Total Catch	Overall Rank
Community	Pacific Herring	*Clupea pallasii*	677,947	1
Community	Tripletooth Gobies	*Tridentiger Spp.*	577,413	2
Listed	Longfin Smelt	*Spirinchus thaleichthys*	303,324	3
Community	Striped Bass	*Morone saxatilis*	181,689	4
Community	Threadfin Shad	*Dorosoma petenense*	171,389	5
Community	Northern Anchovy	*Engraulis mordax*	147,857	6
Community	Prickly Sculpin	*Cottus asper*	132,230	7
Community	Yellowfin Goby	*Acanthogobius flavimanus*	121,728	8
Listed	American Shad	*Alosa sapidissima*	41,541	9
Community	Delta Smelt	*Hypomesus transpacificus*	20,260	11
Listed	Chinook Salmon	*Oncorhynchus tshawytscha*	9,018	13
Community	White Catfish	*Ameiurus catus*	6,244	14
Listed	Steelhead	*Oncorhynchus mykiss*	668	31
Community	Starry Flounder	*Platichthys stellatus*	354	38
Rare	White Sturgeon	*Acipenser transmontanus*	290	42

offered by a model-based approach. Moreover, substantial effort has already been applied toward the development of a design-based estimator for Delta Smelt (Newman, 2008; Polansky et al., 2019), and therefore building from an established methodology with a history of application in the Delta was advantageous. We also found that, although all three methods produced generally comparable estimates, the **design-based estimator approximated the trends evident in the traditional indices very closely** (Figure 12, Figure 13), whereas **model-based approaches did not always correlate as strongly** with historic trends (Figure 14). The model-based approach incorporates the long-term average catch observed at a given station, and therefore appears somewhat less responsive to unusually large catches (e.g., 2006 in Figure 14). This characteristic of the model-based approach may be beneficial in some circumstances, for example in describing long-term patterns of abundance, but may also be problematic if abundance or spatial distribution of a species changes rapidly from year to year. Ultimately, the model-based and design-based methods are complementary, but as an alternative to the current index approach, the design-based estimator provides a balance of familiarity and improved characteristics.

The well-established population estimation approach employed in this review – **stratified random sample ratio expansions** (Polansky et al., 2019) - proved flexible and efficient for calculating survey-specific regional estimates of relative abundance, along with their associated uncertainties. Design-based estimates and associated uncertainty are calculated from the monthly catch per unit effort (CPUE) for different subregions of the Delta, with the estimates extrapolated to the subregion level by applying spatial stratification and weighting of water volumes, while incorporating estimates of gear selectivity. In addition to achieving the important goal of uncertainty estimation and its computational efficiency, the approach is sufficiently generalizable to be **readily applied across surveys, species and regions of the Delta.** This approach therefore allowed the Design Team to rapidly iterate through survey design scenarios and consider a much broader set of options than would have been possible using more computationally intensive methods.

That is not to say that this design-based approach is optimal for all applications or could not be improved upon. A consistent method of estimation is a necessary component for comparison and integration of the separate studies, but failing to correct for differences in the capture efficiencies of the gears can introduce substantial bias. Polansky et al. (2019) demonstrated that with the appropriate comparative data, gear selectivity can be incorporated into both the point and error estimates of this design-based estimator. **Data from appropriate gear comparisons** (Mitchell et al., 2019, 2017a) **that deploy the three status and trends monitoring gears across regions and months are needed to provide the data to derive selectivity-adjusted catches for more species and through space and time.** With the incorporation of such information, use of the design-based approach applied consistently across surveys and would represent a marked improvement upon current methods.

The design-based estimator approach should also be **resilient to future changes in survey design.** Many of the analyses conducted for this review demonstrate the relative stability of point and interval estimates across large changes in effort and the randomization of sampling stations. Moreover, the design-based method is intended to be applied to randomized study designs, and so would not be negatively affected by inclusion of such data if they were collected in the future. The application of the method to the CDFW studies with fixed stations that were not randomly selected may appear concerning, but the use of **post-stratification** of the Delta into relatively small spatial units serves to **minimize the potential bias** introduced by site selection (Polansky et al. 2019). In addition, the **tidal activity** within the Delta likely serves to further **reduce** the influence of selection **bias** because the water present at a given sampling location is transient. Use of this design-based approach as a general monitoring metric does not preclude the use of more sophisticated model-based alternative in specific scenarios. Species of specific concern could still receive targeted special studies and substantial additional analysis (e.g., Delta smelt life-cycle model). The raw catch data will remain available for these purposes, and the design-based estimates should be better suited for direct incorporation into such models than the current indexing approaches.

The Design Team found the **design-based estimator** approach an **improvement** on the index methods currently used, as it can be applied consistently across studies and species, and explicitly incorporates the sampling design water volumes, survey volumes, and associated uncertainty. The estimator can be adjusted for the catchability between gear types, and the uncertainty in those estimates. Given the limited data available for catchability across species and studies under review, the Design Team was unable to incorporate the selectivity into the evaluation, and view this as a next step to improving the integration of the individual studies. Currently, relative catch ratios can only be calculated for a limited set of species (Delta Smelt, Threadfin Shad, Striped Bass) and do not have the sufficient spatial resolution to adjust for regional differences.

The Design Team evaluated 20mm, STN, and FMWT for the ability to integrate abundance estimates for species that overlap among the studies to generate a timeline of **monthly status** throughout the year. **Integrating abundance and distribution trends of species over time based on the design-based approach would appear to be an achievable goal for modernizing the reporting of long-term status and trends monitoring program information (see examples for Age-0 American Shad and Age-0 Longfin Smelt,** Figure 15**).** However, the abundance estimates from these surveys appear influenced by **variability in the catchability of the individual gears**. An example of this divergence of information is provided in the Delta-wide monthly estimates for Age-0 American Shad, where STN abundance estimates are often lower than the corresponding 20mm estimates (Figure 16). **Differences in catchability between 20mm and STN thus needs to be further investigated quantitatively**. Catchability will also need to consider fish behavior and response to avoiding stressful conditions, which could confound interpretation of survey data (i.e., migratory fishes avoiding warm, clear waters during summer influencing low efficiency of STN). Furthermore, the limited catchability information that is available for making such adjustments (see Appendix 3B) does not cover sufficient months and water year types, nor have the studies been regionally stratified to any great extent. The catch ratios for the same species differed by season and gear type that were evident in selectivity patterns of 20mm and STN that will need addressing before adjustments for catchability will be possible. It is not surprising that despite the considerable effort that the surveys have taken to standardize methods, detection efficiency can still vary considerably through space and time (Interagency Ecological Program, 2020).

Of the three studies evaluated for status and trends monitoring, the STN and 20mm surveys use a replicated design with multiple (typically 2-3) tows conducted at each station during each survey event. Several options are available for dealing with replicated tows in calculating estimates of abundance. Given the relatively small size of these nets, replication serves to increase the total volume sampled and increase the probability of catching rare species. Current index methods **sum** catches from replicate tows together, but the tows can also be treated as **independent samples** which **increases the sample size** – and potentially **reduces uncertainty** – of the design-based estimates. The Design Team compared abundance and uncertainty estimates for the STN and 20mm surveys using each tow separately (N = # of stations), calculating station means (N = # of stations), or treating each tow as an independent sample (N = # of tows). This analysis was used to address three questions: 1) is there evidence of depletion between tows (i.e., are abundance estimates from the 2nd or 3rd tows lower than the 1st)? 2) does treating tows independently produce consistently different abundance estimates than combining tows by station? and 3) is uncertainty reduced by treating tows independently (i.e., do design-based estimates have lower standard errors)?

Replication appears to have limited benefit for abundance estimates. Whether combined or treated independently, estimates of abundance from repeated station tows were highly correlated, and where minor differences occurred there was no consistent pattern to which estimate was larger (Figure 17). Our analysis did not show any consistent evidence of depletion suggesting that if it is occurring, it is inconsequential for estimating abundance at the regional scale. Differences in uncertainty between the approaches was largely as predicted, with treatment of replicate tows as independent samples generally resulting in smaller standard errors. However, in regions or months with sporadic catches (i.e., many zeros), use of replicates actually increased the standard errors (e.g., STN longfin smelt in the Far West). **These results suggest that sampling effort would be best allocated over regional scales rather than repeated tows at the same station if the goal is decreasing uncertainty in abundance estimates.**

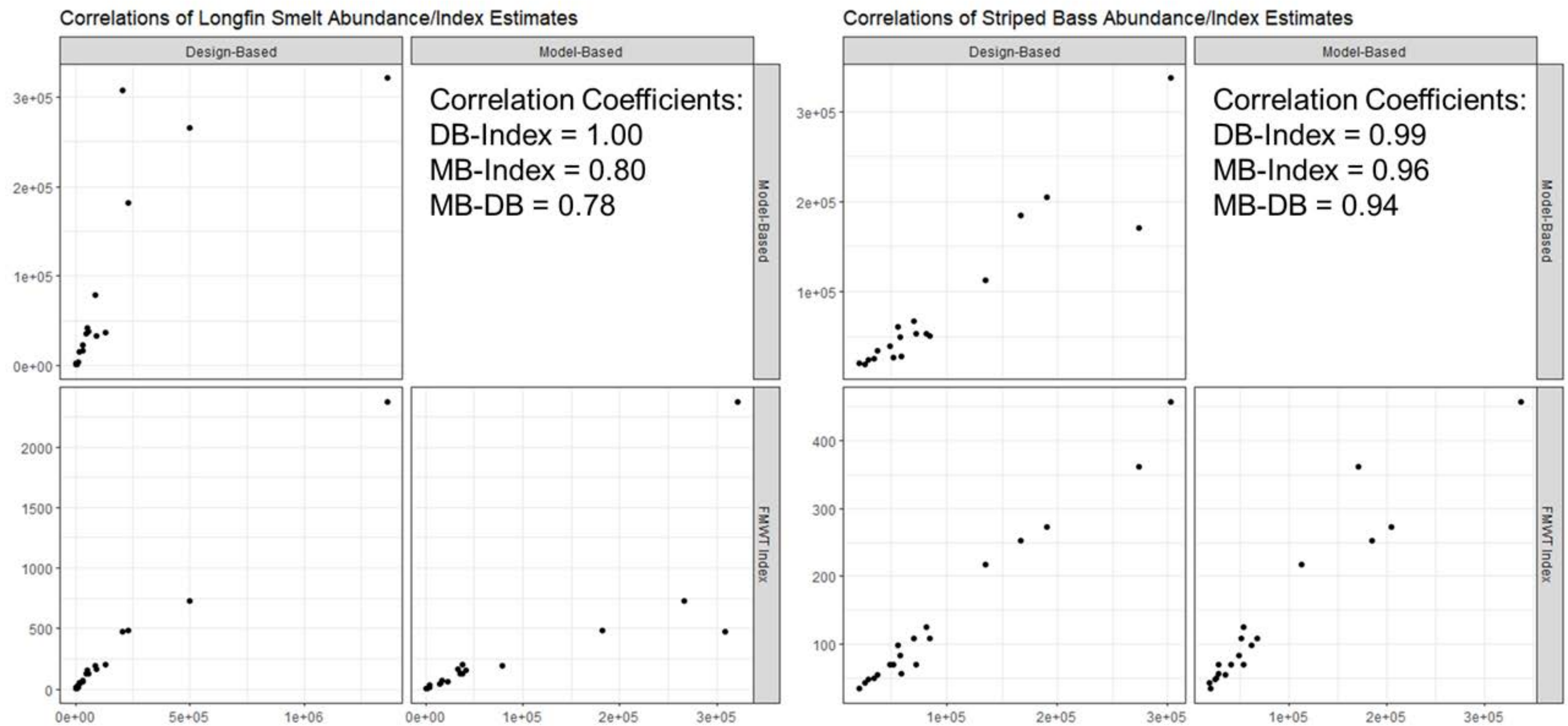


Figure 12. Correlations of Longfin Smelt and Striped Bass Abundance / Index Estimates for the FMWT Using the Traditional Index Method, the Design-Based, and Model-Based methods

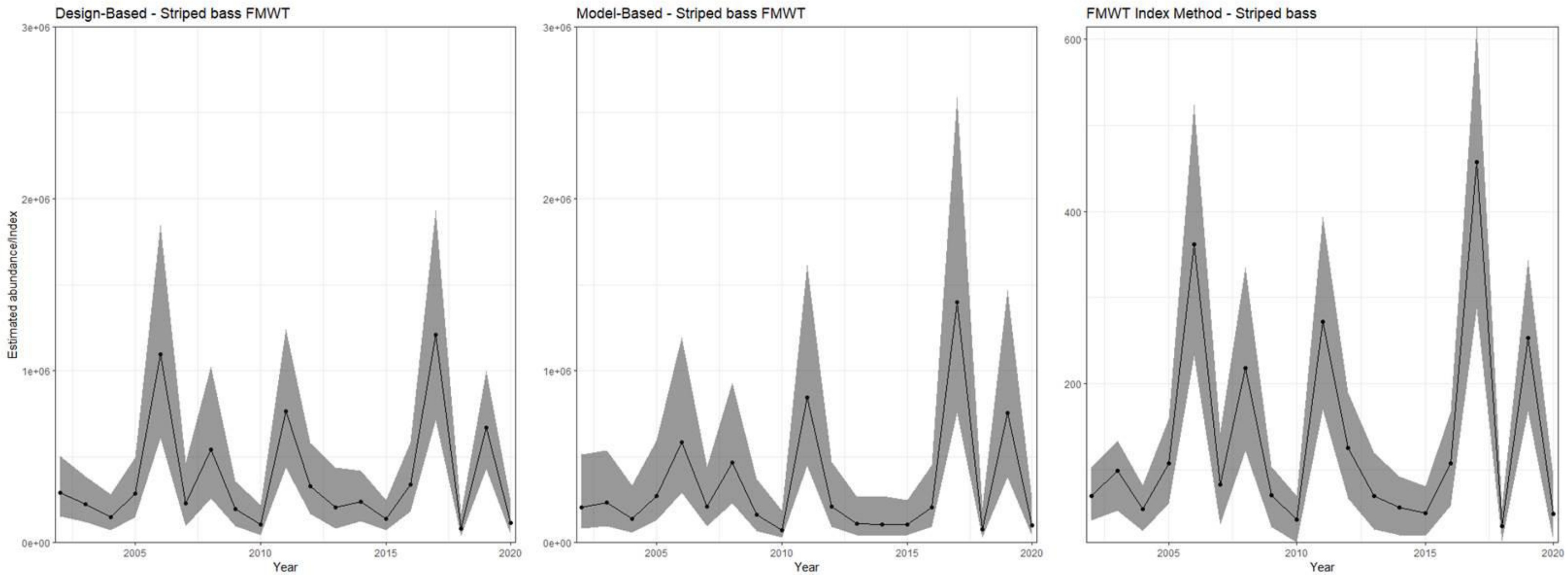


Figure 13. Comparison of three abundance calculation methods for Striped Bass in the Fall Midwater Trawl Study. Black lines are the highest monthly estimate of abundance/index in each year and the shading is the 95% confidence intervals of the estimate.

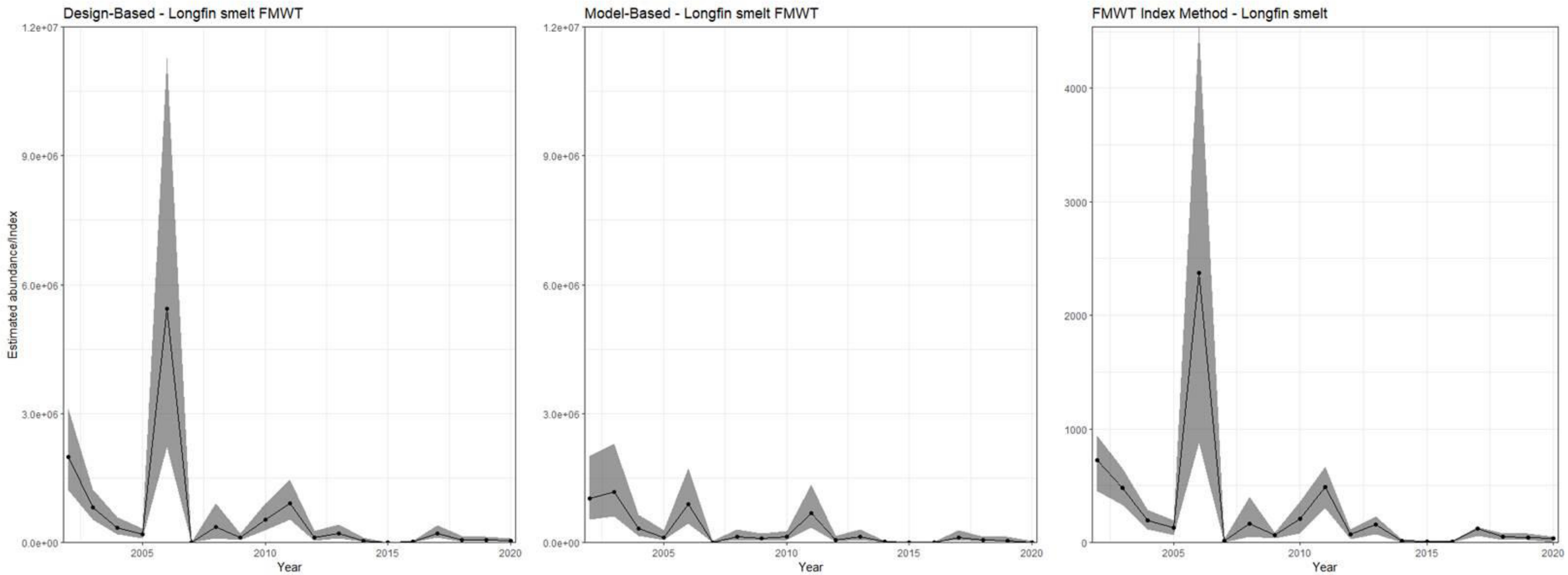
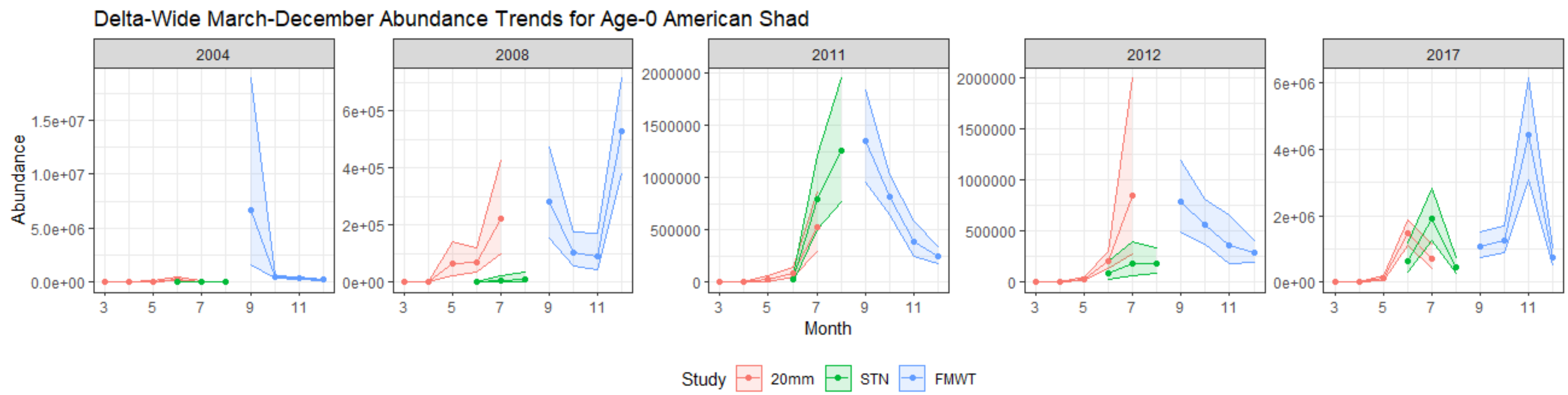


Figure 14. Comparison of three abundance calculation methods for Longfin Smelt in the Fall Midwater Trawl Study. Black lines are the highest monthly estimate of abundance/index in each year and the shading is the 95% confidence intervals of the estimate.

A



B

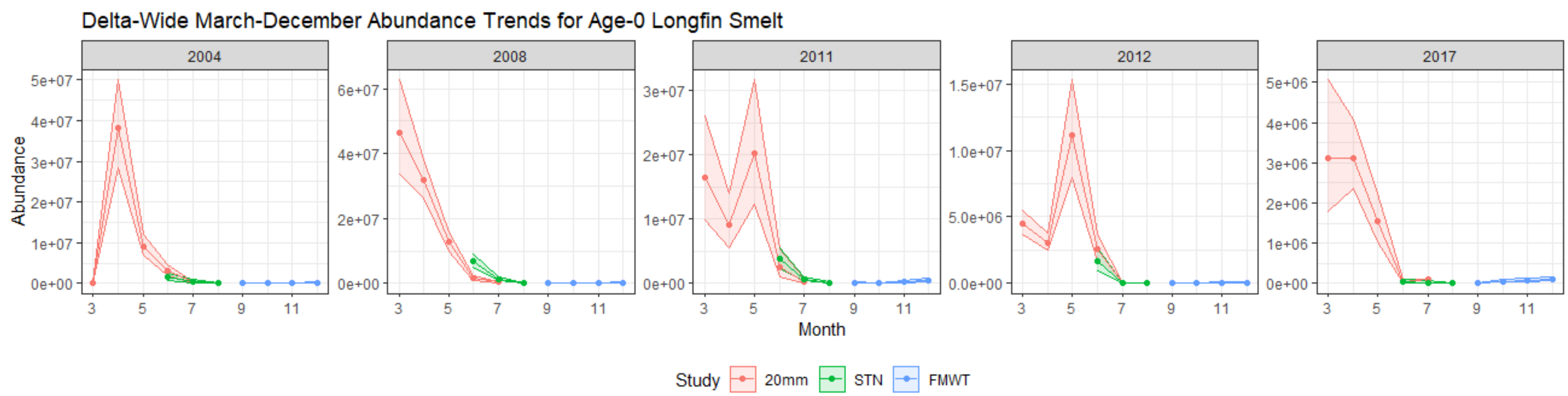


Figure 15. Delta-wide Monthly Abundance Estimates and 95% Confidence Intervals for A) Age-0 American Shad and B) Age-0 Longfin Smelt

Figure 16. Regional Monthly Abundance Trends for Age-0 American Shad in Suisun Bay/Marsh and Confluence, and North Regions. Points are the monthly estimate and the bounds are the 95% Confidence Intervals

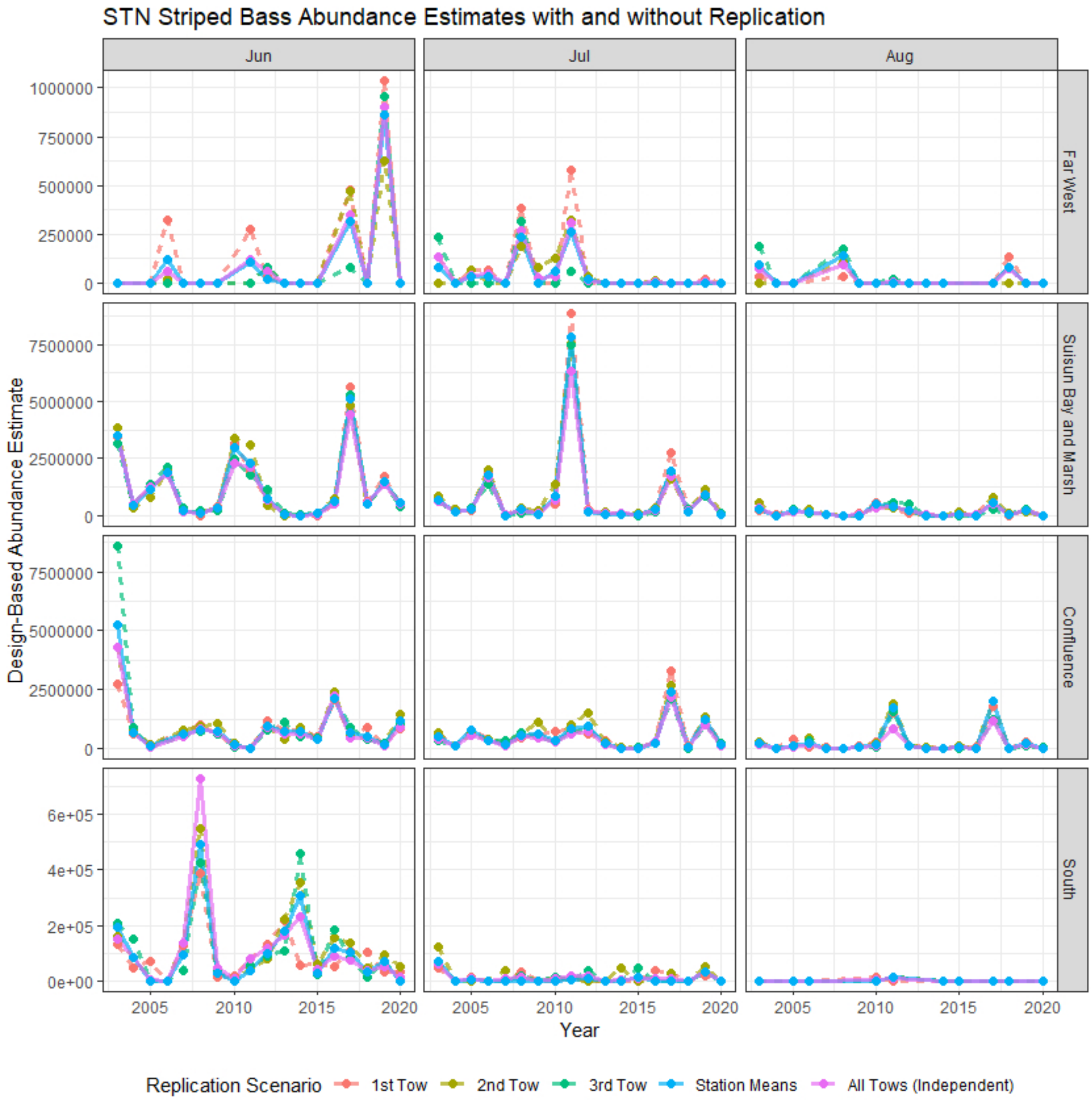


Figure 17. Variability in STN surveys design-based abundances across different treatment of replicate tows. Dashed lines show single tows while solid lines show two methods using all tows

Statistical Resolution for Regional Abundance and Uncertainty

Statistical bias and uncertainty of the existing monitoring designs were evaluated to understand the sensitivity to alterations of effort and to inform the level of effort surveys would require to improve on the statistical resolution of the data. It is important to understand the difference between bias and uncertainty as it refers to sampling theory. Each of the review studies collect samples that can be used to assess the catch abundance of fish species. As each sample is collected, the catch varies by some amount (e.g., 1 fish in the tow 1, 10 in tow 2 and 100 in tow 3), which is referred to as the uncertainty. Uncertainty is defined here as the variability among samples taken in the same region, same month, and same survey for a particular species. Whereas, bias refers to the difference between the center of the sampling distribution (or mean) and the true value of the metric. The sampling distribution is the accumulation of all the samples resulting from the sampling design and effort. A mean or central distribution can be estimated from the samples, in addition to shape and scale of the distribution around some mean or central tendency of the data. Since the true value of the population is unknown, bias must thus be inferred. As discussed earlier in the report, bias in obtaining mean estimates of abundance using the fixed monitoring designs was compared to data generated from stratified, random locations.

Statistical evaluations of uncertainty used the existing datasets to conduct sensitivity analysis (simulations) for reduced and additional effort scenarios. The Design Team evaluated the statistical resolution of the existing monitoring data for reducing standard errors in design-based abundance estimates. This analysis was aimed at evaluating the current standard errors in monthly estimates, changes to sampling effort to reduce standard errors, and identify any stations that could be removed without significantly altering abundances or uncertainty. The outcomes of the reduced effort scenarios were evaluated by examining correlations of the point estimates and standard errors with the full-effort results. High correlation coefficients of both these values (i.e., $r \geq 0.95$) indicate that historical patterns and trends would not be substantially affected by the simulated reductions. Simulations of added effort were conducted by first fitting existing catch data to an appropriate probability distribution by region, month and water year type. The resulting distributions were then used to randomly generate "new" tows from the historical data. These new data were used to recalculate the standard errors in abundance estimates and evaluate the proportional change in uncertainty relative to current effort levels.

Changes in Survey Effort for Status and Trends - 20mm, STN, FMWT

Highlights

- Fundamental limit to the amount of uncertainty that can be accounted for by increasing sampling effort
- Most study designs require 2X more sampling effort if standard errors are to be significantly reduced
- The only study that could remove existing stations without significantly altering historic patterns is FMWT

Readers interested in the technical approaches to support this section should refer to Appendix Chapter 4.

The Design Team's evaluation of statistical resolution in the status and trends monitoring designs focused on simulations of design-based abundance estimates using the integrated data from 20mm, STN, and FMWT. These sensitivity analyses clearly indicated that **the precision that is possible** with the current sampling gears has a **fundamental limit**. Consistent with the pattern reported for Delta Smelt by Polansky et at. (2019), for most species in this review the standard errors and confidence bounds produced by the design-based estimator were quite large relative to the point estimates (i.e., coefficients of variation commonly 25-90%). Newman (2008) made this same observation and concluded that even a doubling of FMWT effort would not greatly reduce the variance of Delta smelt abundance estimates. This finding is largely consistent with our analyses across studies and species. Surveying the **low densities** and **patchy distributions** of young pelagic fishes in the Delta with relatively small pelagic trawls results in distributions of catch that are **both zero-inflated and overdispersed.** Under these conditions, **confidence intervals cannot necessarily be reduced by simply adding more tows.** Expectations for the level of confidence in abundance estimates that can be achieved given current conditions and sampling methods should account for this limitation of the data.

Effort Reduction

The sensitivity of the status and trend monitoring designs to **reductions in effort** in the existing fixed monitoring stations identified the potential for **reductions only in the FMWT** (Table 7). This conclusion was reached by removing specific stations from the sampling designs and correlating the revised abundances and confidence bounds to the estimates generated from using all of the data. The stations with the least effect on the sensitivity of monthly design-based estimates in FMWT are shown in Table 7. The majority of these stations occur in embayments with the most existing stations, i.e., San Pablo Bay/Carquinez Strait and the Confluence. Based on the loss of sensitivity between revised and current trends in abundance, **the Design Team concluded that 20mm and STN could not support reduced effort changes without altering the historic trends in abundances or the level of uncertainty in these data.**

Table 7. Fall Midwater Trawl stations considered for removal

Region	Stratum	Stations Considered for Removal
Far West	San Pablo Bay and Carquinez Strait	309, 310, 311, 314, 321, 323
	Napa River	--
Suisun Bay and Marsh	Suisun / Honker Bays	409, 413, 502
	Suisun Marsh	608
Confluence	Confluence	804.1, 809, 810, 811, 812, 813
North	Cache Slough	--
	Sacramento Mainstem	717, 735, 72
	Sacramento DWSC	794
South	N/S Forks Mokelumne R.	--
	South	

Effort Additions

Simulations of sampling designs with additional effort to reduce the standard errors (S.E.) in monthly abundance estimates varied by study and Review Region. Sensitivity analysis indicated that most regions would require **more than twice the current effort to observe a 25% reduction in the S.E.** of abundance estimates. These results point to the limitations of these highly variable fish abundance data that exhibit overdispersed sampling distributions, with a prevalence of zero values.

Results for **20mm** indicated the priority region for increased sampling effort is **San Pablo Bay**, where the abundance estimate S.E. could be reduced by 25% **with 6-15 additional tows per survey** (Table 8)**.** All of the other regions were indicated to require 16+ tows per region and survey.

Results for **STN** indicated the priority region for increased sampling effort is **Napa River**, where abundance estimate S.E. could be reduced by 25% with **6-15 additional tows per survey** (Table 9)**.** All of the other regions were indicated to require 16+ tows per region and survey, suggesting a high degree of variability in monthly catch across species.

Four priority regions (Table 10) where increased sampling effort could reduce the S.E. by 25% were indicated for **FMWT**; **Napa River, Suisun Marsh, Cache Slough, and Sacramento DWSC.** Most notably that 6-15 additional tows per month in the **Napa River** may lead to up to 50% decrease in S.E.. All of the other regions were indicated to require 16+ tows per region and survey.

Overall, the key patterns worth noting across the Status and Trend Monitoring study designs:

- Regions where additional effort was indicated span riverine, sloughs, and embayment areas of the Delta
- 20mm was the only study where additional effort could lead to 25% better standard errors in San Pablo Bay, all the other areas for priority effort are associated with relatively lower water volume strata. This finding indicates the need to stratify sampling effort by water volumes.

- FMWT indicated the best opportunities for reducing uncertainty, where four regions could benefit from the addition of 6-15 tows per survey. Strategies to incorporate this additional effort is further discussed in the next chapter.
- Napa River was indicated in STN and FMWT study design evaluations as being a high priority for lowering uncertainty.
- Suisun Marsh and Napa River were both indicated for additional effort for abundance estimates

Table 8. 20mm Stratified Sampling Effort Evaluation to Reduce Standard Errors in Age-0 Fish Abundances by 25% and 50%

Region	Stratum	Current Number of Stations Per Survey	Current Number of Tows Per Survey	Number of Additional Tows for 25% Reduction in S.E.	Number of Additional Tows for 50% Reduction in S.E.
Far West	San Pablo Bay and Carquinez Strait	2	5	6-15	31-50
	Napa River	6	16	31-50	>50
Suisun Bay and Marsh	Suisun / Honker Bays	7	20	31-50	>50
	Suisun Marsh	3	8	16-30	31-50
Confluence	Confluence	12	33	>50	>50
North	Cache Slough	7	13	16-30	>50
	Sacramento Mainstem	0	--	--	--
	Sacrmento DWSC	0	--	--	--
South	N/S Forks Mokelumne R.	1	3	16-30	31-50
	South	9	25	31-50	>50

Table 9. Summer Townet Stratified Sampling Effort Evaluation to Reduce Standard Errors in Age-0 Fish Abundances by 25% and 50%

Region	Stratum	Current Number of Stations Per Regions	Current Number of Tows Per Survey	Number of Additional Tows for 25% Reduction in S.E.	Number of Additional Tows for 50% Reduction in S.E.
Far West	San Pablo Bay and Carquinez Strait	2	6	16-30	31-50
	Napa River	1	3	6-15	16-30
Suisun Bay and Marsh	Suisun / Honker Bays	7	19	31-50	>50
	Suisun Marsh	3	9	16-30	31-50
Confluence	Confluence	10	26	>50	>50
North	Cache Slough	4	6	16-30	31-50
	Sacramento Mainstem	0	0	--	--
	Sacramento DWSC	3	8	16-30	31-50
South	N/S Forks Mokelumne R.	1	2	16-30	>50
	South	8	19	31-50	>50

Table 10. FMWT Stratified Sampling Effort Evaluation to Reduce Standard Errors in Age-0 Fish Abundances by 25% and 50%

Region	Stratum	Current Number of Stations Per Regions	Current Number of Tows Per Survey	Number of Additional Tows for 25% Reduction in S.E.	Number of Additional Tows for 50% Reduction in S.E.
Far West	San Pablo Bay and Carquinez Strait	30	30	>50	>50
	Napa River	2	2	6-15	6-15
Suisun Bay and Marsh	Suisun / Honker Bays	28	28	>50	>50
	Suisun Marsh	3	3	6-15	16-30
Confluence	Confluence	24	24	31-50	>50
North	Cache Slough	6	6	6-15	16-30
	Sacramento Mainstem	5	5	16-30	31-50
	Sacramento DWSC	4	4	6-15	16-30
South	N/S Forks Mokelumne R.	6	6	16-30	>50
	South	14	14	31-50	>50

Species Detection and Catchability Considerations

The gears used to monitor fish species in the review studies vary considerably in the size, mesh, and maximum depth they can be deployed, and thus the water volume (effort) that the nets filter on each survey. Additionally, the seasonal timing of when the studies are conducted heavily influences the detection patterns. Consequently, the catchability of different gears (i.e., the probability of species retention assuming the availability to the gear) is the key area of uncertainty to evaluate the effectiveness of the studies for sampling fish within a region, season, and size range. **Currently, understanding of catchability and gear effiiciency for the studies reviewed across most of the species is very limited.** Previous studies have adjusted for imperfect detection in the estimation of abundance or occupancy for a few individual species, but this is not yet possible for most species in the pelagic fish community (Mahardja et al. 2017, Peterson and Barajas 2018, Polansky et al. 2019). This issue was identified by the Design Team as a **major theme for design improvements to support integrating the review studies to better understand species detections and abundance patterns.**

Species Detections

Highlights

- SLS detects the fewest species, 20mm detects the most pelagic species
- Status and Trends studies have more overlap in species detections than real-time

Readers interested in the technical approaches to support this section, including maps illustrating the distributions of commonly occurring species, should refer to Appendix Chapter 3A.

The Design Team evaluated the catch patterns to identify the species that are well represented across the review studies. A conservative threshold of 10% was set to evaluate the species detections. For real-time studies, this evaluation identified only three species that are detected in all three real-time studies in more than 10% of tows; **Longfin Smelt, Pacific Herring, and Prickly Sculpin** (Figure 18). This was due primarily to the limited species observed in the SLS survey, which only samples the smallest of larvae and only in January-March period.

The 20mm study contained eight species that have the >10% detections. These species include all three species that are also well represented in SLS, along with **Threadfin Shad, *Tridentiger* spp. (gobies), White Catfish, Starry Flounder, and Striped Bass (**Figure 19**).** These species encompass a variety of life histories of the fish community (e.g., marine, brackish, open water, and demersal).

SKT has exhibited >10% species detections in several species that are not represented by either of the other real-time studies, notably **Chinook Salmon** and **Steelhead**, and several species in the Age-1 size class, such as **American Shad** and historically, **Delta Smelt.** Together, this information indicated that the real-time studies largely provide understanding of recruitment for specific species of concern. Overall, real-time detections have been highest in Napa River and Suisun Bay/Marsh relative to other regions.

The Design Team also considered the **information generated from overlapping surveys** (i.e., surveys conducted in the same month) in real time monitoring. In the context of understanding shifts in the **distribution** of species, SLS was noted for very rarely having no catch of larvae in January, suggesting larvae hatching may occur sooner than the first survey of the year. For this reason, the SLS has recently shifted to conducting two additional surveys in December. SLS and 20mm surveys overlap in March and have shown good correspondence in the size retention of smelt. SLS has been most effective for 5-15 mm larvae, while 20mm Survey averages a wider size frequency from 10mm+ (Figure 20). The size: frequency patterns suggest the overlap between SLS and 20mm surveys is informative for **providing confidence in recruitment patterns** between the two gear types. Depending on the resolution desired on larvae size : frequency detection, **20mm could potentially provide much of the same information**.

The Design Team also performed similar evaluations of species detections for the STN and FMWT, along with 20mm Survey. Several species were identified that are routinely observed in at least 2 of the 3 studies in more than 10% of tows; including **Longfin Smelt, Delta Smelt, Pacific Herring** (not in STN), **Striped Bass, American Shad, and Threadfin Shad** (Figure 21). Several of the evaluated species appear less suitable for regional abundance estimates, however, due to either the lack of sufficient stations across their entire spatial distribution of habitat (e.g., **Northern Anchovy**) or the poor efficiency of the gears for their life-history (e.g., **White Sturgeon, White Catfish, and Starry Flounder**).

Overlap in survey timing between **status and trends studies** was also a key consideration for understanding difference in species detections among the gears. The Design Team considered that overlapping surveys provide understanding of catch patterns that are contiguous with the increasing size of fish during the year and size: frequency changes in the population. Overlap exists in the survey timing (June/July) and targeted size range of the 20mm and STN nets (10-50 mm and 20-55 mm). Overlap is the result of these surveys generating size-based abundance indices, which include variable number of surveys contributing to indices. However, STN and FMWT do not currently overlap in the timing of surveys, and only a narrow overlap in the fish size ranges that can be retained (20-55 mm and 30-120 mm FL, respectively). The lack of overlap could potentially create data gaps in the size distribution timeline without shifts in temporal emphasis of certain gears (e.g., Figure 22). This appears most relevant between STN and FMWT.

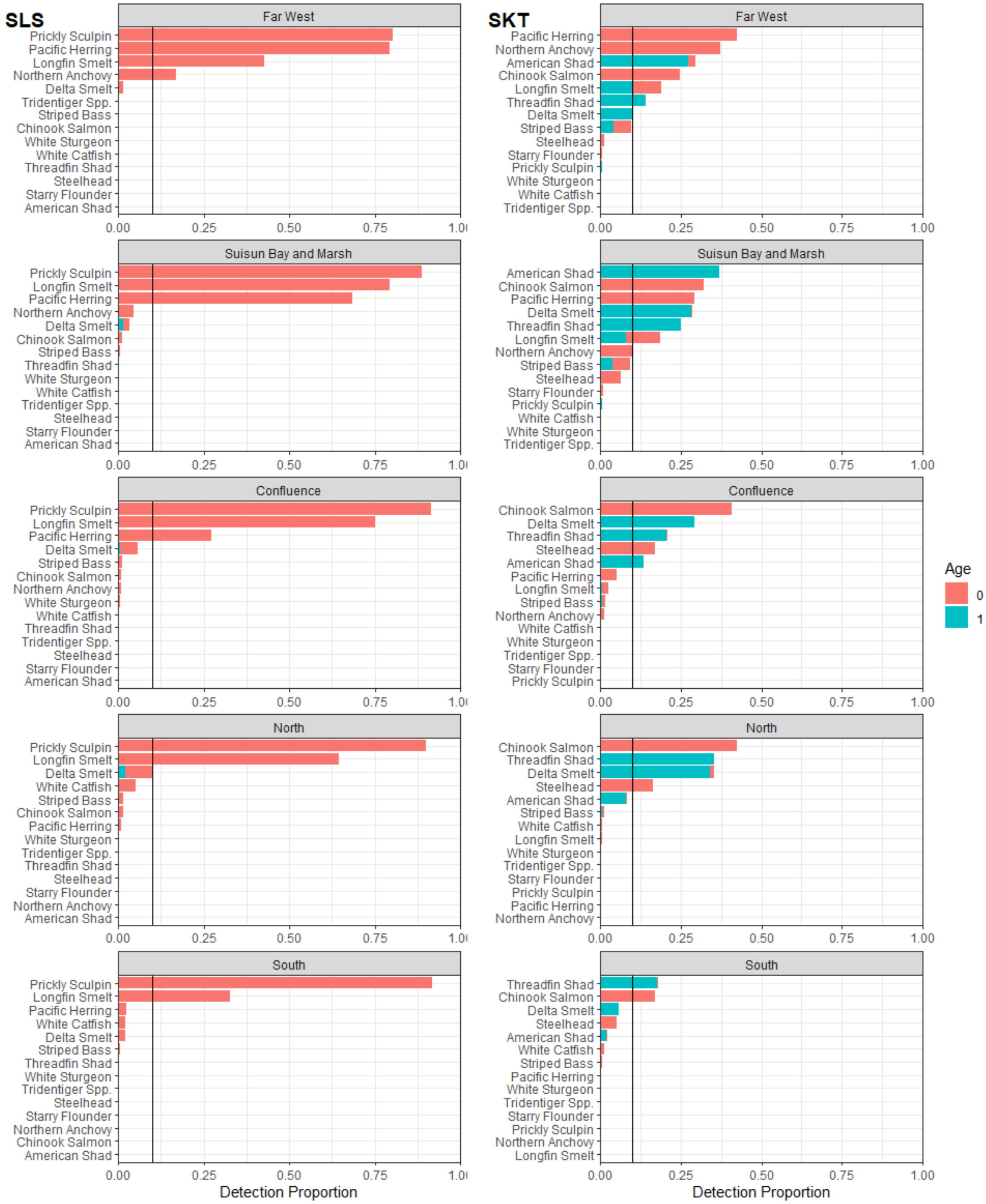


Figure 18. Detection Proportion for SLS and SKT Surveys. Vertical line identifies a 10% detection threshold.

Figure 19. Detection Proportion for 20mm, STN, and FMWT Surveys By Region. Vertical line identifies a 10% detection threshold.

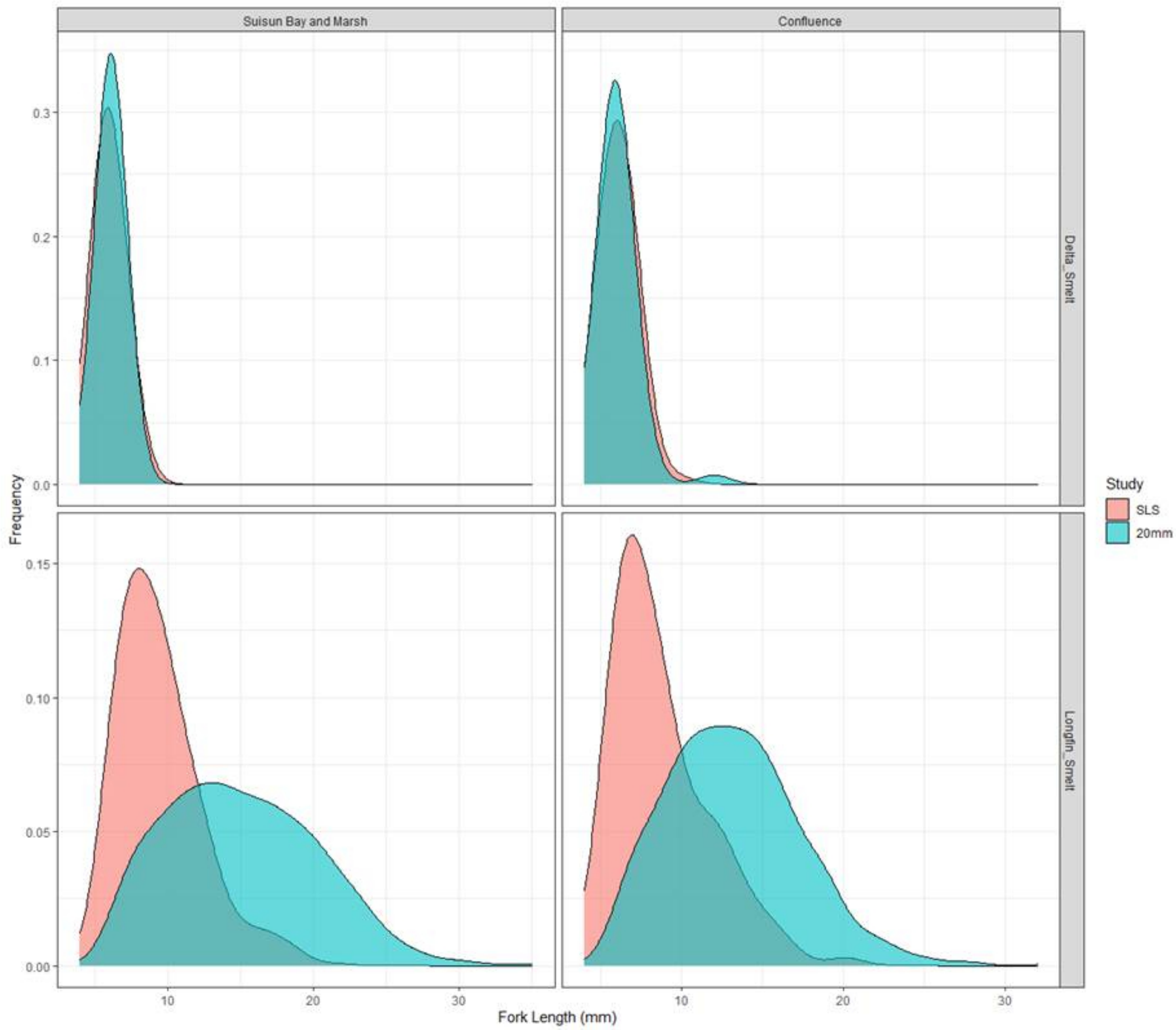


Figure 20. SLS and 20mm - Size: frequency distribution resulting from overlapping surveys for Delta Smelt and Longfin Smelt in Suisun Bay/Marsh and the Confluence

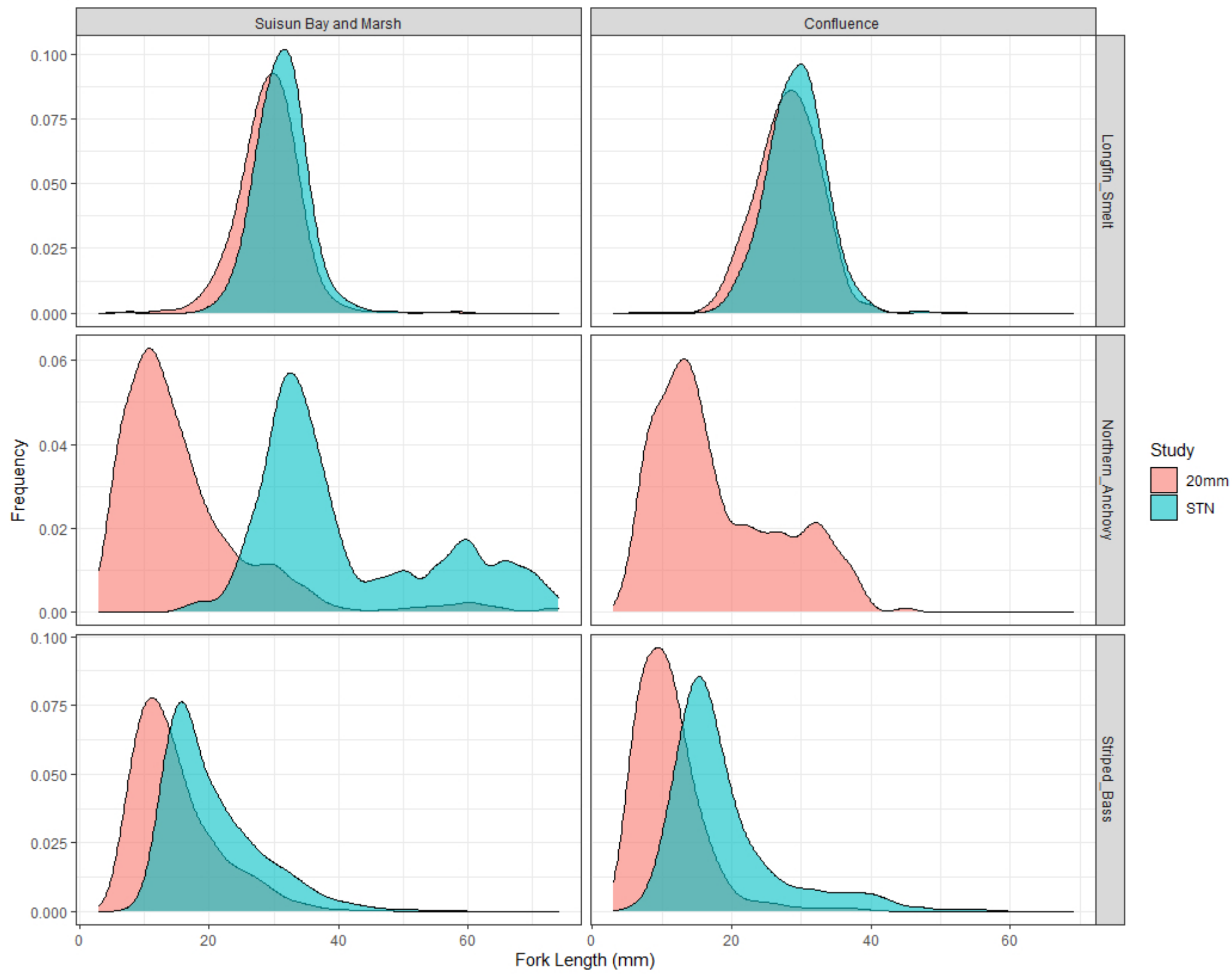
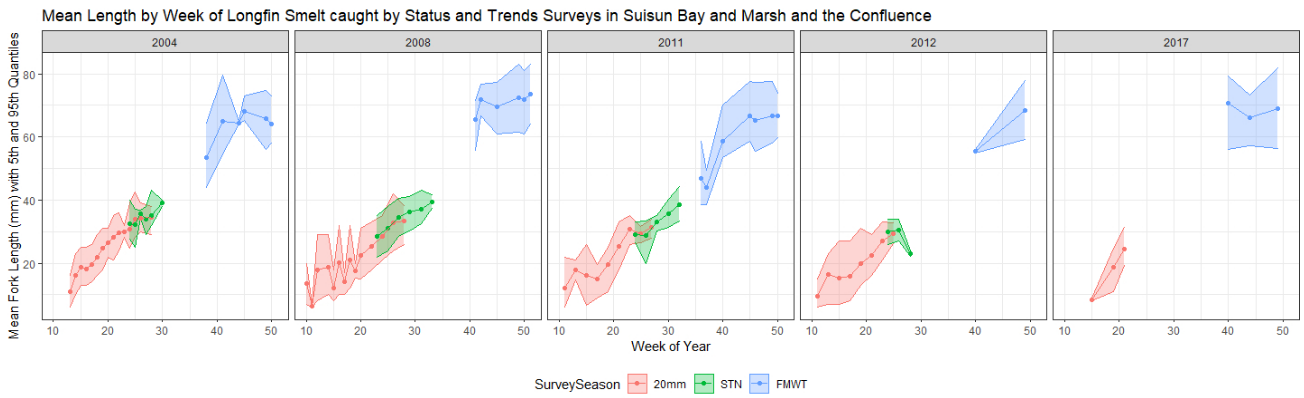


Figure 21. 20mm and STN - Size: frequency distribution resulting from overlapping surveys for Longfin Smelt, Northern Anchovy, and Striped Bass in Suisun Bay/Marsh and the Confluence.

A



B

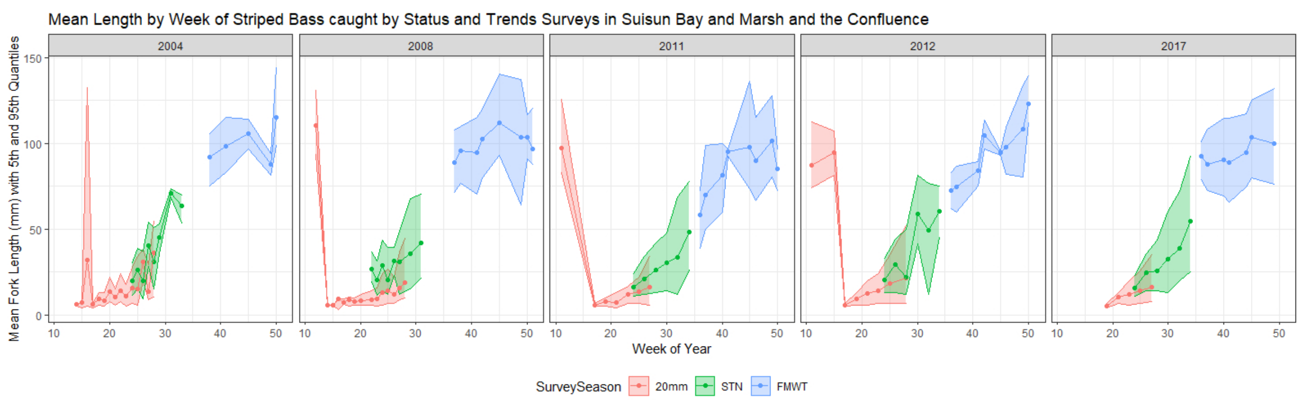


Figure 22. Mean length of A) Longfin Smelt and B) Striped Bass in Suisun Bay/Marsh and Confluence By Week and Year. Points are the weekly mean size and the bounds are the 5th and 95th quantiles

Changes in Survey Effort for Real Time Monitoring -SLS, 20mm, SKT

Highlights

- Real-time monitoring designs with higher effort can only attain high (>95%) species detections for few species, life-stages, and regions
- Additional effort for real-time studies in San Pablo Bay and Suisun Marsh best opportunities to increase species detections

Readers interested in the technical approaches to support this section should refer to Appendix Chapter 3A.

Study design simulations to evaluated changes to the sampling effort in **real-time monitoring designs** focused on **presence / absence** information and **probability of species detection** (i.e., the proportion of tows with 1 or more individuals of a species). The Design Team used the historic patterns of **species detection** to evaluate the statistical resolution of the existing monitoring data associated with changes to sampling effort. Similar to the analysis of design-based estimates, the influence of reduced effort on species detection was assessed by sequentially removing stations from the existing data, and examining correlations with full effort scenarios; in this case correlations were calculated for the proportions of tows with a detection. By treating each tow as a binomial sample in which a target species is either detected or not, it was possible to calculate, using the binomial cumulative distribution function, the probability that a given number of tows would detect one or more individuals. For each species, survey, region, month and water year type, the probability of detection was calculated under status quo effort. The number of tows required to achieve high (95% or greater) detection probability was then calculated (Appendix 3 and 4 contain further details). From these values, the number of additional tows required to reach the 95% detection threshold was calculated.

No clear opportunities were identified for reducing effort in real-time monitoring. Potential loss of information on spatial patterns in species detection was indicated for at least one strata or region in each study-specific comparison. Therefore, simulations for statistical resolution in species detection entirely focused on where **additional effort** could lead to 95% detection in the greatest number of species for each region. Not surprisingly, regional and temporal differences in detection probabilities were apparent across all three studies reflecting changes in the distributional patterns and population abundance over time of individual species. Scenarios where **1-15 additional tows** would be needed to improve detection probabilities were ranked highest (Figure 23).

The **SLS** evaluations indicated **San Pablo Bay** and **Suisun Marsh** were the best opportunities for increasing effort, largely with benefits for Longfin Smelt detection. SLS monitoring has the fewest number species to estimate altered detection probabilities. Current SLS monitoring only exhibits high (>95%) detection probabilities for Longfin Smelt in the Confluence stratum.

In contrast, the **20mm** study can provide understanding of spatial patterns the **largest number pelagic fish species** of the real-time monitoring studies. Current monitoring effort has **>95% probability of detection for seven species**, with the most occurring in the Confluence and Suisun / Honker Bays, including Longfin Smelt, Striped Bass, and Pacific Herring. The 20mm evaluations indicated that **additional effort** in **San Pablo Bay** and **Suisun Marsh** would benefit detection probability for several species. Most species sampled in San Pablo Bay would require **6-15 additional tows** to achieve the 95% probability threshold, while 1-5 additional tows would benefit the majority of species observed in Suisun Marsh.

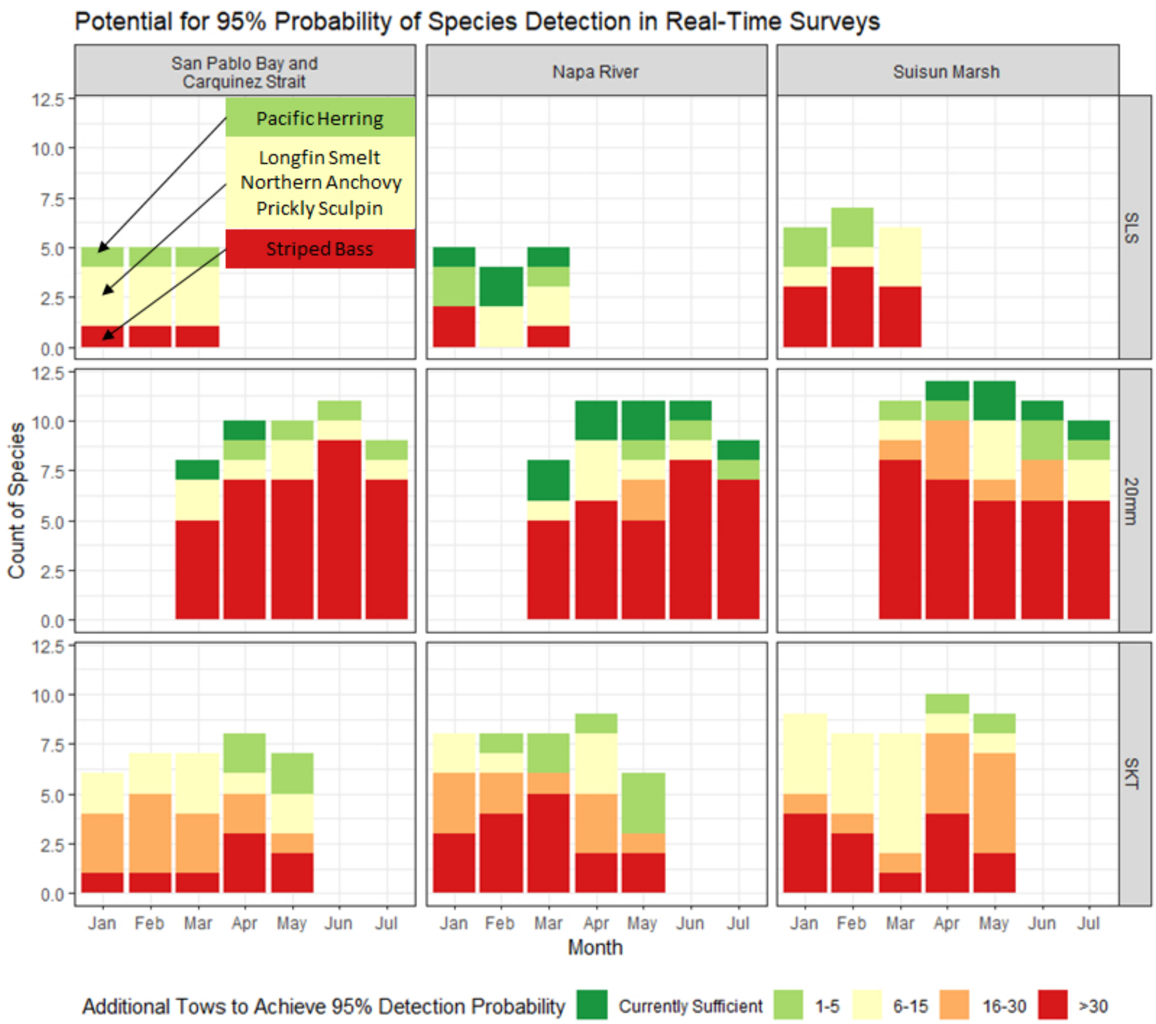


Figure 23. Example of the detection analysis results showing the increases in effort (number of tows) required to achieve a 95% probability of detecting one or more individuals in real-time monitoring surveys. The top left panel provides annotation to highlight that each column represents multiple species, but note that the analyzed species vary between studies.

The SKT Survey could benefit from **additional monitoring** in several of the Review Strata. Both **San Pablo Bay** and **Suisun Marsh** were indicated as areas where species detections could be increased, as well as **Napa River**. Chinook Salmon, Steelhead, and Pacific Herring were the only species that currently have 95% probability of detection in SKT monitoring, and these tended to occur towards the end of SKT monitoring period (April-May) and principally in the Confluence and Suisun/Honker Bays. Much **higher levels of effort (16-30 tows of additional effort)** would be needed for some species regardless of the region or strata. This level of effort was driven by the less frequently caught (rare) species that are unique to SKT monitoring such as Starry Flounder, Threadfin Shad, and White Catfish. Note, the lower species detections for Starry Flounder and White Catfish would be expected, as adults are demersal fishes and SKT is limited to sampling the upper 6 ft. of the water column.

Overall, some key patterns are worth noting across the real time monitoring study designs.

- Regions where additional effort would increase species detections was indicated in riverine, sloughs, and embayment areas of the Delta
- Increasing effort to improve Age-0 species detections would likely provide the greatest benefit for the 20mm and least benefit for the SLS
- San Pablo Bay and Suisun Marsh were indicated in all three of the study design evaluations as being a high priority for additional effort to increase species detections of several species, most notably Longfin Smelt, Pacific Herring, Striped Bass, and American Shad.
- Additional monitoring effort in Napa River was only indicated for the SKT evaluations, which was largely driven by increasing detections for Pacific Herring and Chinook Salmon.

Gear Catchability

Highlights

- Gear comparisons show efficiencies differ by season, but some results heavily caveated by very low catch
- SLS was more effective in spring than either 20mm or STN
- 20mm and STN were equally as effective in summer for a few species
- STN was more effective than FMWT in fall for Delta Smelt, Threadfin Shad, and Striped Bass and should be considered for overlap in sampling with FMWT
- SKT was more effective at capturing Delta Smelt compared to the 20mm in Summer (9 -fold) and FMWT in Fall (80-fold) and thus should be considered for additional overlap in sampling

Readers interested in the technical approaches to support this section should refer to Appendix Chapter 3B.

Overlap in sampling of multiple gears is often used to understand the relative catchability and size selectivity of different gear types. Using previously conducted gear comparison study, the Design Team, supported by Dr. Josh Korman, conducted modeling of these data to understand catchability among the review studies. Three of the gears (SLS, 20mm, FMWT) used in the study were deployed using oblique (surface to bottom) trawls, while the other two gears (SKT and Chipps Island Midwater Trawl Survey - CMWT) were deployed at the surface only. Relative effort in the catchability analysis was adjusted for both the number and volume of tows. Only seven species were sampled well enough to be assessed for relative efficiency (Mitchell et al., 2019; Mitchell and Baxter, 2021). It must be acknowledged that the prior gear comparison sampling was specifically designed to inform Delta Smelt catch and not necessarily to address other species (see Mitchell et al. 2019 for further details). Sampling was conducted between 2012 and 2015, when side by side multiple gear deployments were conducted at one to three locations per season. Surveys were generally limited to a single survey event per season (fall: Sept-Dec; spring: Jan-Apr; summer: May-Aug). With these caveats, results of our catchability and size selectivity modelling provided some limited information on the utility of the various gear types (See Appendix 3B), but left unanswered questions as to the temporal and regional variability in catchability.

Firstly, our modeling of spring side-by-side tows on the Lower San Joaquin River and Deep Water Ship Channel highlighted that the **SLS study was more effective than the 20mm** study for American Shad, Prickly Sculpin, and Threadfin Shad. It is unknown whether SLS and 20mm are similarly effective in Suisun Bay/Marsh and Confluence Regions, where Longfin Smelt and other Age-0 fish that are caught during the spring recruitment period.

Results during summer months provide support for continued use of the STN, but potentially expanding the timing of the STN survey based on the species we assessed. The STN survey during spring tows was less effective compared to the 20mm, but during summer and fall was more effective than 20mm and FMWT for a few species (e.g., Delta Smelt in fall). Conversely, STN was less effective than 20-mm for Gobies, Striped Bass age-0, and Threadfin Shad. Therefore, additional side by side sampling to compare STN and 20mm/FMWT would benefit understanding of STN effectiveness under current conditions.

Along a similar theme, the Design Team identified that the integration of status and trends data would be more effective if the data from different **gears could be combined**. To this end, the gear efficiency of STN vs. FMWT needs to be further evaluated. For some species, our modeling exercise was associated with small sample sizes, and as mentioned above, lacked the spatial resolution to evaluate the gears side-by-side from multiple regions relevant to the individual studies. These conclusions are not influenced by patterns in size selectivity, which were generally similar among gear types when sample sizes were sufficient to reliably estimate proportions-by-size.

Our analysis also provides support for conducting the SKT later into the year. The gear comparison data during summer and fall showed that SKT is more effective than either 20mm or the FMWT for sampling Threadfin Shad (30-fold and 6-fold, respectively) and has equivalent catchability for other species, such as American Shad (Table 11). This may suggest the potential benefit to **increasing the overlap of SKT sampling with summer/fall surveys than is currently done**.

During this evaluation, concern about the **selectivity of the FMWT** for smaller fish (< 60mm FL) was also raised. The Design Team identified that this question would be best addressed through a **covered cod-end experiment**, where the finer mesh from the STN could be attached to cover the cod-end of the FMWT net. Filling data gaps related to gear efficiency are critical to making any substantive changes to the studies in the future, such as shifting the timing of the deployment of gears. The correlations between these catches are a priority area for improving the understanding of the pelagic review studies.

Table 11. Gear Comparison of SKT and STN relative to FMWT in Lower Sacramento River. A value above 1 indicates higher catchability relative to the reference gear. E.G. SKT has ~83X higher catchability for Delta Smelt than the FMWT.

Fall Season Lower Sacramento Reference Gear: FMWT		**Relative Catchability 50%**
Species	**Gear**	
American Shad	SKT	1.08
American Shad	STN	0.02
Delta Smelt	SKT	83.38
Delta Smelt	STN	4.79
Gobies (Unid)	SKT	0.00
Gobies (Unid)	STN	261.95
Striped Bass	SKT	0.00
striped Bass	STN	0.51
Threadfin Shad	SKT	5.94
Threadfin Shad	STN	0.07

4. Major Findings and Study Design Improvements

The Design Team identified several major findings that inform proposed options to improve the pelagic fish monitoring designs for the five review studies.

1. Status and Trends monitoring effort varies regionally, by study, and are not balanced with pelagic water volumes.

The main finding of the Design Team with regards to the statistical resolution for estimating regional abundances and uncertainty is the lack of spatial balance in sampling effort relative to pelagic water volumes among the studies. The Design Team recommends that spatial balance be brought into the coordinated sampling effort among 20mm, STN, and FMWT studies. This redesign should redistribute regional effort that accounts for the water volumes, stratified to emphasize management areas, integrates the data from design improvement experiments, and other relevant ongoing monitoring efforts. This coordinated monitoring design would benefit the long-term integration of the studies for obtaining regional abundances with reduced levels of uncertainty. The existing fixed monitoring design would serve as the basis for this redesign, and should be set up with consideration for the probabilistic design experiments and gear comparisons that are needed. Also, for regional abundance estimation, survey redesigns should consider standardizing the conduct of replicated tows to spread sampling effort across more stations (potentially in a probabilistic approach).

Furthermore, design improvement experiments are needed to test whether probabilistic monitoring can improve the uncertainty (i.e., reduce standard errors) compared to the fixed CDFW monitoring design for generating abundance estimates at the regional scale. Due to dynamic nature of tidal excursion and outflow, the current approach of sampling geographically fixed stations already has an element of randomization to where trawls are conducted between replicate tows and repeated tows over consecutive surveys. Comparisons of EDSM's probabilistic Kodiak trawl data to SKT fixed stations provided some evidence for increased resolution with randomized stations but not in the case for comparisons between EDSM and CDFW 20mm gears. Due to the inherently different monitoring design and effort conducted by EDSM, and the lack of comparable probabilistic data for STN or FMWT, the Design Team emphasize that special studies are needed to test the ability of probabilistic designs to narrow standard errors or provide markedly different design-based estimates of abundance.

2. Design-based estimators provide a standardized method for estimating relative abundance and uncertainty across the studies and species.

The design-based abundance estimation approach employed in this review – stratified random sample ratio expansions – proved flexible and efficient for calculating survey-specific regional estimates of relative abundance, along with the associated uncertainties. The Design Team suggest that application of this approach should be considered across the review studies long-term data. The gear-specific water volumes that have been calculated for the Review Regions and Review Strata may be used, or alternative volumes calculated and applied in the same manner. The methods provide the flexibility to incorporate different volumes or sample frames, selectivity and catch ratios, and estimates of uncertainty in gear selectivity as improvements in understanding is gained for relevant gear comparison studies over time. Furthermore, the design-based estimates were highly correlated with traditional indices and stable to changes to the sampling effort that are prevalent in the historic data, with the added benefit of explicit estimates of uncertainty in abundance.

Simulations of status and trends monitoring designs indicated a fundamental limit to the size of uncertainty that can be accounted for by increasing sampling effort. The results of this review effort help tailor expectations for the level of confidence in abundance estimates that can be achieved with the pelagic gears given current conditions. The coefficients-of-variation estimated in monthly design-based estimators often spanned orders of magnitude. As a result, uncertainty in abundance estimates can only be reduced by 25% with twice the current effort in most regions. These observations seem to suggest there is a limit to the statistical resolution that can be achieved by simply adding effort to the fish surveys. Areas of the Upper SFE that have been identified as the best opportunities for increasing resolution for species detections and abundance estimates should be explored to 'learn and revise' from the long-term datasets.

3. The sample frame and monitoring designs represented by the review studies are explicit for the pelagic habitat of the Delta.

All five studies currently emphasize sampling in pelagic waters of the central and eastern Delta regions of the Upper SFE. The Designs Team's qualitative evaluation of sampling methods employed in the five studies indicated that the methods used for the currently employed gears are appropriate for depths (greater than 6 ft.) and habitats (pelagic, open-water) that are being targeted, and are comparable to techniques used in similar environmental settings elsewhere. The Design Team acknowledged the need to better understand shifts in species distributions into littoral habitat (such as sloughs and marsh area) and the potential for species movements into shallower depths. However, the logistical feasibility of sampling in unknown areas of the Delta and those being too shallow for the gears was identified as the major hindrance to such design changes at this time. Monitoring interest in understanding fish distributions in waters along the margins of the Estuary, shallower than 6 ft. depth, should be evaluated using different sampling methods. Design changes identified below have only considered options that would improve monitoring design in the pelagic waters accessible by vessel as portrayed in the sample frame for monitoring design evaluations.

4. The effectiveness of the gear types for sampling pelagic fish species are a function of the differences in catchability, gear selectivity, and species detection.

Differences in catchability, gear selectivity, and species detection are the main source of uncertainty to understanding effectiveness of the gears and for integrating the catch data from multiple studies. Regional catch patterns and trends revealed that SLS routinely catches the fewest Age-0 species, while the 20mm Survey gets the most. SKT often catches Age-0 and Age-1 species that are rare to other gear types. These differences can be largely attributed to the sampling methods and the timing of the studies. Catchability analysis indicated that seasonal overlap in SKT and STN sampling with the FMWT may be warranted. However, limitations of the study design combined with low sample sizes for some comparisons make definitive conclusions challenging. The Design Team identified gear comparison studies as a key theme for future study improvements.

5. All of the review studies have overlapping spatial sampling

Spatial sampling overlap among the review studies when viewed at broad spatial scales (the Review Regions or Strata). Due to the different efficiencies of the gear, the Design Team recognized that the overlapping spatial surveys provide confidence in understanding patterns in distribution and size: frequency of species over time (e.g., smelt in SLS and 20mm). Therefore. the Design Team viewed this overlap to be an integral component of the spatial designs due to the uncertainty in the efficiencies of the gears across the pelagic fish community. By overlapping over time and space, confidence in the relative differences between catch of multiple gears can be assessed.

6. Data gaps exist in understanding efficiency of multiple gears

Design improvement experiments are needed to optimize the transition between surveys and gears. Integration of the study designs has shown that spatial and temporal extent of the review studies aligns with the regional use of Delta pelagic habitat for recruitment of several young-of-year pelagic fish species and life stages. Real-time studies emphasize the critical periods for entrainment (Jan through May), while status and trends studies provide a broader spatial (sub-regional) and temporal (March through December) emphasis. Yet, the value of these data could be improved through integration of these datasets with better understanding catchability (catch-ratios) during overlapping and adjacent spatial surveys. The limited understanding of gear efficiencies across the studies and for the well represented species, currently prevents integration of these datasets. Therefore, the key data gap identified by the Design Team is in the need to generate scaling factors (catch-ratios) on a regional basis from side-by-side deployments that can adjust the design-based estimators from multiple surveys and spatial samplings. The priorities for this coordinated monitoring are for the status and trend studies that currently do not overlap (STN and FMWT in fall), and to compare oblique and surface sampling (SKT and FMWT in fall).

Review Outcomes to Support the Real-Time Monitoring Program

The Design Team's evaluation of the review studies to support the Real Time Monitoring Program has identified several potential design changes (Table 12 and Table 13) that would likely result in additional resources and permitting requirements. The fixed station design was considered to be appropriate for the real-time information needs and focus of spatial information in specific management areas. Replication was identified as a key component of the sampling methods that should be standardized for real-time monitoring to increase probabilities for species detection.

The evaluation of statistical resolution of the real-time monitoring data for presence/absence and species detections, only identified areas that were far from the pumping facilities, such as Suisun Marsh as priorities for additional monitoring. San Pablo Bay and Napa River were also indicated to be best opportunities to improve species detections. In areas, proximate to facilities, the Design Team suggests that replication should be used to increase probabilities for species detection. Secondly, side by side sampling with the SLS and 20mm gear would have benefits for integrating these data in the future and evaluating the optimal timing for the 20mm Survey. Similarly, to understand if the SKT trawl could provide information on species being missed by the FMWT, sampling of both those gears during the FMWT sampling season will provide needed information on catchability.

A summary of outcomes and considerations to support the real-time monitoring designs for SLS and SKT are outlined in the tables below. Outcomes for 20mm have been integrated into the Status and Trends section that follows.

Table 12. Smelt Larval Survey – Summary of Outcomes

Information Gap	**Design Change**	**Means Objectives**	**Considerations**
Species Detections	Add replication at real-time stations	Presence/Absence, Adaptation, Predictive Tools	Additional resources; Additional take
Larval Detection	December Surveys	Spatial Patterns, Presence/Absence, Adaptation	Additional resources; Additional take
Gear Efficiency	Gear comparison between SLS and 20mm (in Spring)	Redundancy, Adaptation, Predictive Tools	Additional effort to coordinate; Additional resources; Confirm transition of sampling between gears

Table 13. Spring Kodiak Trawl - Summary of Outcomes

Information Gap	**Design Change**	**Means Objectives**	**Considerations**
Species detection	Add replication at real-time stations	Presence/Absence, Adaptation, Predictive Tools	Additional resources; Additional take
Catchability between gears	Gear comparison between SKT and FMWT net (in Fall)	Redundancy, Adaptation, Predictive Tools	Additional effort to coordinate; Additional resources; Additional take; Confirm selectivity between gears
Redundant surface trawling gears	Coordination with EDSM	Redundancy, Adaptation	Additional effort to coordinate;

Review Outcomes to Support the Status and Trend Monitoring Program

The Design Team's evaluation of the review studies to support the Status and Trends Monitoring Program has identified several areas for improvements to the study designs (Table 14, Table 15, Table 16). These design improvements fall into three broad categories.

Firstly, for these studies to be integrated together a re-balance of sampling effort needs to be considered in a stratified sampling design that controls for the number of stations per region and study relative to water volumes. The main observation from the evaluation of 20mm, STN, and FMWT monitoring designs are that they are unbalanced relative to defined regional water volumes, and the amount of effort required to improve the current uncertainty in monthly abundances would require each region being sampled at least twice the current level of effort. A coordinated stratified design across the three studies should be considered that spreads effort regionally for generating abundance estimates and includes consideration for the higher level of effort identified in the sensitivity analyses.

Secondly, to evaluate whether a stratified random design could lead to catch information with less uncertainty than stratified fixed station monitoring, the Design Team suggests adding random stations to both the 20mm, STN, and FMWT survey periods. The probabilistic stations should be planned to maximize comparison to the fixed stations and to provide broader understanding of species detection and regional abundance patterns. This information would allow for time-sensitive comparisons and would ideally focus effort on region(s) where abundance estimates are desired. These design improvement studies could be phased in over time to add value to the long-term routine study design, and be of benefit to understanding statistically based regional sampling distributions.

A FMWT design improvement study was recently initiated that has begun implementing this approach to monitoring and this should be considered for 20mm and STN as well. Furthermore, it is envisaged that this information would need to be collected over multiple years and for several regions of the Delta. FMWT has initiated the special study in the Far West and Suisun Bay/Marsh areas, other regions will likely need similar investigations. The key uncertainty in adopting these experiments are the current drought conditions and whether these additional efforts will yield enough data for comparisons. The recently initiated study can provide the initial testing of this strategy, but these should be carefully planned out for long-term benefit.

Thirdly, gear comparison studies to address the differences in catchability across study elements so that monthly survey data may be integrated over time is the other priority theme for design improvements. Conducting design improvement studies focused on assessing catchability (e.g., Mitchell et al. 2017, Mitchell et al. 2019) are needed to integrate information acquired from concurrent studies into the long-term status and trend monitoring program's abundance estimations (Newman et al. 2008, Polansky et al. 2019). This conclusion has been reached by previous authors tasked with similar review efforts (LTMR, 2021). The Design Team findings have put forth several ideas for where, when, and for which studies and species would such design changes have the most benefit under our simulation scenarios and interpretive assessments. Ideally, this will be regionally stratified and conducted during the seasonal timeframe when these gears are routinely used. Along the same theme, to address concerns that the FMWT may be under-sampling small fish (<~60-mm FL), the Design Team suggests the FMWT net be used at select stations with a cod end covered with finer mesh (the same as used in STN) to increase the retention of small fish. The STN was most effective during summer and fall, and thus could have added value if deployed into September to provide overlap with FMWT. These targeted gear comparisons between STN and FMWT will aid in interpreting seasonal catch patterns and abundances with greater confidence.

The strategies to design improvements for the status and trend studies are further outlined in the tables below.

Table 14. 20mm Study - Summary of Outcomes

Information Gap	**Design Change**	**Means Objectives**	**Considerations**
Statistical resolution for abundance estimates	Stratified monitoring effort using regional volumes and strata	Status and Trends, Abundance	Effort changes coordinated across multiple studies Additional resources; Additional take
	Shift effort from replication to regional effort (maintain replication at real-time stations)	Status and Trends, Abundance, Redundancy, Adaptation	Shift effort to balance regional sampling
Species detection; Statistical resolution for abundance estimates; Bias	Conduct additional probabilistic monitoring (starting with San Pablo Bay and Suisun Marsh)	Spatial Patterns, Presence/Absence, Status and Trends, Abundance	Additional resources; Additional take

Table 15. Summer Townet Study - Summary of Outcomes

Information Gap	Design Change	Means Objectives	Considerations
Statistical resolution for abundance estimates	Stratified monitoring effort using regional volumes and strata	Status and Trends, Abundance	Effort changes coordinated across multiple studies
	Shift effort from replication to regional effort	Status and Trends, Abundance, Redundancy	Shift effort to balance regional sampling
Species detection; Statistical resolution for abundance estimates; Bias	Conduct additional probabilistic monitoring (starting with San Pablo Bay, Napa River, Suisun Marsh)	Spatial Patterns, Presence/Absence, Status and Trends, Abundance	Additional resources; Additional take
Catchability between gears	Gear comparison between STN and FMWT (fall)	Redundancy, Adaptation, Confidence, Predictive Tools	Additional resources; Additional take; Confirm transition of sampling between gears

Table 16. Fall Midwater Trawl Study – Summary of Outcomes

Information Gap	Design Change	Means Objectives	Considerations
Statistical resolution for abundance estimates	Stratified monitoring effort using regional volumes and strata	Status and Trends, Abundance	Effort changes coordinated across multiple studies
Redundancy within regions	Remove up to 20 existing stations that have limited value for abundance estimates	Status and Trends, Redundancy, Adapt	Shift effort to under sampled regions
Species detection; Statistical resolution for abundance estimates; Bias	Conduct additional probabilistic monitoring (starting with Napa River, Suisun Marsh, DWSC, and Cache Slough)	Spatial Patterns, Presence/Absence, Status and Trends, Abundance	Additional resources; Additional take; Already started for Napa River and Suisun Marsh
Size selectivity	Covered Cod End on FMWT net	Confidence, Predictive Tools	Additional resources; Additional take
Catchability between gears	Gear comparison between STN and FMWT (fall)	Redundancy, Adaptation, Confidence, Predictive Tools	Additional resources; Additional take; Confirm transition of sampling between gears

5. Future Work

The Design Team identified several opportunities for improvements that will not disrupt compliance measures, but could include changes to existing studies and the need to identify special studies. These improvements will not likely be feasible to implement all at once. Therefore, the next step in adopting these potential changes to monitoring designs are to map out logistical and resource feasibility of changes to temporal and spatial sampling effort recommended by the Design Team with related special studies. The Design Team has put forward several clear themes that we wish to discuss the implications of for prioritizing the changes to these studies. A forthcoming workplan will present the recommendations that result from evaluation of these technical findings for redesign of the fish monitoring studies.

Another clear area for future work is to begin effort planning relevant to the design improvement studies that will require inter-study coordination and permitting. These effort planning steps are critical to develop coordinated study plans to incorporate the necessary gear comparisons. The Design Team has also suggested the need for design improvement studies for stratified probabilistic sampling, which will also take time to design and plan out. This process was recently initiated for design improvements in the FMWT study, which took a couple months to design, scope, and have approved on an expedited timeline. A coordinated process across the status and trend studies where abundance estimates are desired should be considered in the next steps to plan out these design improvement efforts.

Future areas of work that should also be considered:

- Sampling design evaluations for zooplankton monitoring. Due to time constraints these data were not critically evaluated as part of this review effort.
- Sampling design evaluations that focus specifically on optimizing the stations used for real-time monitoring is needed.
- Spatial autocorrelation of current and proposed survey changes. The spatial autocorrelation of survey designs has an important influence on the relative efficiency of sampling effort. This is an involved topic that was not in the scope of this effort.
- Other areas for future work exist to evaluate introducing additional field techniques into studies, such as mark/recapture methods, occupancy studies, and coordination with other programs for littoral sampling.
- Finally, the data integration effort should not be underestimated, and future efforts to manage and integrate datasets from CDFW studies should consider the steps taken in this review.

The Design Team looks forward to assisting the Steering Committee in implementation of the proposed monitoring design strategies to improve future pelagic fish monitoring efforts as part of the San Francisco Bay's environmental monitoring enterprise.

6. Glossary

Abundance	Population estimate for the number of fish in a given area
Bias	Amount the expected result differs from the true value of the underlying parameter
Catchability	Relationship between resource abundance and the efficiency to capture the resource with a specific sampling gear
Catch-ratio	Relative catchability for one gear type compared to another
Detection Probability	Proportion of tows in which a target species is captured
Demersal	Inhabitants close to the seafloor or bottom of a body of water
Depletion	Decline in catch/abundance with increased fishing effort
Fixed-design	Sampling locations that are predetermined at the same geographic coordinates and have no associated probability
Fundamental Objective	The core outcome that one cares about that represents something to strive for to achieve the program mission
Gear Efficiency	Probability that a fish is retained by sampling gear
Means Objective	A particular way of achieving the core outcome or objective
Oblique Tow	Survey method to collect a vertically stratified sample
Pelagic	Inhabitants of the water column
Presence-Absence	Presence or absence of a species from a collection of samples (*see Detection Probability*)
Probabilistic-design	Sampling locations that have a known probability of being sampled
Real-Time	Monitoring that supports real time decision-making
Relative Abundance	Relative estimate of the number of fish in a given area
Sample frame	All elements of the target population to be used as the basis for a sample
Sensitivity	Smallest amount of change that can be detected
Stratification	Sorting of data and objects into distinct groups or layers
Trend	Change in relative abundance over time
Uncertainty	Amount of error in an estimate of the average value of a population

7. References

Aasen, G.A., 1999. Juvenile delta smelt use of shallow-water and channel habitats in California's Sacramento-San Joaquin Estuary. California Fish and Game 85, 161–169.

Bashevkin, S.M., 2021. Six decades (1959-2020) of water quality in the upper San Francisco Estuary: an integrated database of 11 discrete monitoring surveys in the Sacramento San Joaquin Delta, Suisun Bay, and Suisun Marsh ver 1. [WWW Document]. Environmental Data Initiative. https://doi.org/https://doi.org/10.6073/pasta/1694ea7f9ef9cc8619b01c3588029683

Baxter, R., Breuer, R., Brown, L., Conrad, L., Feyrer, F., Fong, S., Gehrts, K., Grimaldo, L., Herbold, B., Hrodey, P., Mueller-Solger, A., Sommer, T., Souza, K., 2010. Interagency Ecological Program 2010 Pelagic Organism Decline Work Plan and Synthesis of Results.

Baxter, R., Feyrer, F., Nobriga, M., Sommer, T., Estuary, S.F., 2008. Pelagic Organism Decline Progress Report: 2007 Synthesis of Results Interagency Ecological Program for the.

Enwright, C., Culberson, S., Burau, J.R., 2013. Broad timescale forcing and geomorphic mediation of tidal marsh flow and temperature dynamics. Estuaries and Coasts 36, 1319–1339. https://doi.org/10.1007/S12237-013-9639-7

Grimaldo, L., Burns, J., Miller, R.E., Kalmbach, A., Smith, A., Hassrick, J., Brennan, C., 2020. Forage Fish Larvae Distribution and Habitat Use During Contrasting Years of Low and High Freshwater Flow in the San Francisco Estuary. San Francisco Estuary and Watershed Science 18. https://doi.org/10.15447/sfews.2020v18iss3art5

Grimaldo, L., Feyrer, F., Burns, J., Maniscalco, D., 2017. Sampling Uncharted Waters: Examining Rearing Habitat of Larval Longfin Smelt (Spirinchus thaleichthys) in the Upper San Francisco Estuary. Estuaries and Coasts 40, 1771–1784. https://doi.org/10.1007/s12237-017-0255-9

Hendrix, A.N., Fleishman, E., Zillig, M.W., Jennings, E.D., 2021. Relations between abiotic and biotic environmental variables and occupancy of delta smelt (Hypomesus transpacificus) in autumn.

Honey, K., Baxter, R., Hymanson, Z., Sommer, T., Gingras, M., Cadrett, P., 2004. IEP Long-term Fish Monitoring Program Element Review Interagency Ecological Program for the.

Interagency Ecological Program, 2020. Interagency Ecological Program Long-term Monitoring Element Review: Pilot approach and methods development.

Latour, R.J., 2016. Explaining Patterns of Pelagic Fish Abundance in the Sacramento-San Joaquin Delta. Estuaries and Coasts 39, 233–247. https://doi.org/10.1007/s12237-015-9968-9

Lewis, L.S., Willmes, M., Barros, A., Crain, P.K., Hobbs, J.A., 2020. Newly discovered spawning and recruitment of threatened Longfin Smelt in restored and underexplored tidal wetlands. Ecology 101. https://doi.org/10.1002/ecy.2868

Mahardja, B., Mitchell, L., Beakes, M., Johnston, C., Graham, C., Goertler, P., Barnard, D., Castillo, G., Matthias, B., 2020. Leveraging Delta Smelt Data for Juvenile Chinook Salmon Monitoring in the San Francisco Estuary. PrePrints. https://doi.org/10.20944/preprints202006.0362.v2.

Mahardja, B., Tobias, V., Khanna, S., Mitchell, L., Lehman, P., Sommer, T., Brown, L., Culberson, S., Conrad, J.L., 2021. Resistance and resilience of pelagic and littoral fishes to drought in the San Francisco Estuary. Ecological Applications 31. https://doi.org/10.1002/eap.2243
Mahardja, B., Young, M.J., Schreier, B., Sommer, T., 2017. Understanding imperfect detection in a San Francisco Estuary long-term larval and juvenile fish monitoring programme. Fisheries Management and Ecology 24, 488–503. https://doi.org/10.1111/fme.12257
Merz, J.E., HamMilton, S., BergmMan, P.S., Cavallo, B., 2011. Spatial perspective for delta smelt: A summary of contemporary survey data. California Fish and Game.
Mitchell, L., Baxter, R., 2021. Examining retention-at-length for four pelagic fishes caught in the Fall Midwater Trawl Survey. San Francisco Estuary and Watershed Science 19.
Mitchell, L., Newman, K., Baxter, R., 2019. Estimating the size selectivity of fishing trawls for a short-lived fish species. San Francisco Estuary and Watershed Science 17. https://doi.org/10.15447/sfews.2019v17iss1art5
Mitchell, L., Newman, K., Baxter, R., 2017a. A covered cod-end and tow-path evaluation of midwater trawl gear efficiency for catching delta smelt (Hypomesus transpacificus). San Francisco Estuary and Watershed Science 15. https://doi.org/10.15447/sfews.2017v15iss4art3
Mitchell, L., Polansky, L., Newman, K.B., 2017b. DSM TN 2. Design-based Estimates of Delta Smelt Abundance.
Moyle, P.B., Brown, L.R., Durand, J.R., Hobbs, J.A., 2016. Delta smelt: Life history and decline of a once-abundant species in the San Francisco estuary. San Francisco Estuary and Watershed Science 14. https://doi.org/10.15447/sfews.2016v14iss2art6
Newman, K.B., 2008. Sample Design-Based Methodology for Estimating Delta Smelt Abundance. San Francisco Estuary and Watershed Science 6. https://doi.org/10.15447/sfews.2008v6iss3art3
Peterson, J.T., Barajas, M.F., 2018. An evaluation of three fish surveys in the San Francisco Estuary, California, 1995-2015. San Francisco Estuary and Watershed Science 16. https://doi.org/10.15447/sfews.2018v16iss4art2
Polansky, L., Mitchell, L., Newman, K.B., 2019. Using Multistage Design-Based Methods to Construct Abundance Indices and Uncertainty Measures for Delta Smelt. Transactions of the American Fisheries Society 148, 710–724. https://doi.org/10.1002/tafs.10166
Polansky, L., Newman, K.B., Nobriga, M.L., Mitchell, L., 2018. Spatiotemporal Models of an Estuarine Fish Species to Identify Patterns and Factors Impacting Their Distribution and Abundance. Estuaries and Coasts 41, 572–581. https://doi.org/10.1007/s12237-017-0277-3
Polansky, L., Studen, L., 2019. DSM TN 29. Water volumes of the San Francisco Estuary in Areas Occupied by Delta Smelt. DRAFT.
R Core Team, 2020. R: A Language and Environment for Statistical Computing.
Reynolds, J.H., Knutson, M.G., Newman, K.B., Silverman, E.D., Thompson, W.L., 2016. A road map for designing and implementing a biological monitoring program. Environmental Monitoring and Assessment 188. https://doi.org/10.1007/s10661-016-5397-x
Rosenfield, J.A., Baxter, R.D., 2007. Population Dynamics and Distribution Patterns of Longfin Smelt in the San Francisco Estuary. Transactions of the American Fisheries Society 136, 1577–1592. https://doi.org/10.1577/T06-148.1

San Francisco Estuary Institute, 2017. California Aquatic Resource Inventory (CARI) version 0.3 [WWW Document]. URL (accessed 4.20.21).

Sommer, T., Mejia, F., Hieb, K., Baxter, R., Loboschefsky, E., Loge, F., 2011. Long-Term Shifts in the Lateral Distribution of Age-0 Striped Bass in the San Francisco Estuary. Transactions of the American Fisheries Society 140, 1451–1459. https://doi.org/10.1080/00028487.2011.630280

Stompe, D., Moyle, P., Kruger, A., Durand, J., 2020. Comparing and Integrating Fish Surveys in the San Francisco Estuary: Why Diverse Long-Term Monitoring Programs are Important. San Francisco Estuary and Watershed Science 18, 0–16. https://doi.org/10.15447/sfews.2020v18iss2art4

Tamburello, N., Connors, B.M., Fullerton, D., Phillis, C.C., 2019. Durability of environment–recruitment relationships in aquatic ecosystems: insights from long-term monitoring in a highly modified estuary and implications for management. Limnology and Oceanography 64. https://doi.org/10.1002/lno.11037

USFWS, 1996. Recovery Plan for the Sacramento-San Joaquin Delta Native Fishes. Portland, OR, USA.

USFWS, 1995. Formal Consultation and Conference on Effects of Long-term Operation of the Central Valley Project and State Water Project on the Threatened Delta Smelt, Delta Smelt Critical Habitat, and Proposed Threatened Splittail. Sacramento, CA, USA.

Wang, R.-F., Ateljevich, E., Fregoso, T.A., Jaffe, B.E., 2019. A revised continuous surface elevation model for modeling.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the Tidyverse. Journal of Open Source Software 4, 1686. https://doi.org/10.21105/joss.01686

8. Appendices

Appendix 1. Data Integration

Appendix 2. Spatial and Temporal Evaluation

Appendix 3A. Taxonomic Comparison Among Studies

Appendix 3B. A Comparison of Relative Catchability and Size Selectivity of Different Gears Sampling Fish in the San Francisco Bay Delta

Appendix 4. Stratified Sampling Designs for Estimation of Regional Abundance

Appendix 1. Data Integration

Introduction

The Design Team began its evaluation of the review studies by integrating the data from five fish trawl survey databases, resolved discrepancies, and generated summaries of the status quo (i.e., current) sampling effort. Within this task, the CDFW study efforts were evaluated relative to other long-term monitoring programs in the Bay and Delta.

Data Sources

Data for each CDFW study is currently made available through a publicly accessible FTP site. Processed data files that report catch-per-tow or catch-per-station are available for several of the surveys, but these files exclude important information including lengths. Comprehensive survey data are stored as Microsoft Access Databases that typically include separate tables for tow, catch and length data in addition to a variety of necessary lookup tables. For each survey the Access Database file and any associated metadata were downloaded. The Design Team acknowledged several drawbacks to this data management approach including the proprietary nature of the file type (i.e., users must purchase Microsoft Office in order to open Access files) and the substantial amount of preprocessing that is necessary to prepare the data for typical analysis and modeling applications.

Study Data Preparation and Integration

Each database was converted into a comprehensive, long-format flat file by first exporting individual database tables in .csv format and then using each survey's respective relationship diagram to join the tables using the 'tidyverse' family of packages (Wickham et al., 2019) in the programming language R (R Core Team, 2020). At this initial stage all variables were retained. Many variables required renaming to create consistency between surveys and facilitate later data integration. Each survey has a unique protocol for which taxa, and how many of each are measured, but in all cases when large catches of a taxon occur, only a subsample is measured. As such, it was necessary to adjust length-frequencies to account for unmeasured individuals and ensure that any analysis of length is not biased by underrepresenting the lengths of fish in large catches.

In order to expand measured length-frequencies to the total catch the typical CDFW protocol was used, calculating Fa,l, the adjusted frequency of each recorded length as: $Fa,l = Tc(Fm,l/Tm)$ where Tc is the total catch of a taxon, Fm,l is the measured frequency of each recorded length of a taxon, and Tm is the total number of fish measured of a taxon. In early years of the FMWT and STN surveys it was also common to only measure a subsample of target taxa. In cases where no individuals of a species were measured, an adjusted length-frequency could not be calculated, and so an NA was entered for the length of each counted individual. Such unmeasured individuals were retained at this initial stage of data assembly.

Tows that recorded no catch need to be retained in the dataset in order to accurately model detection probabilities and regional abundances. The data joining process introduces NA values for the count and species variables of "no catch" tows. In these instances, the species was renamed to No_Catch and the count replaced with 1 to ensure that these tows were not removed during further data processing and integration.

After including these tows with no catch, adjusted length frequencies were rounded to the nearest integer, and the data tables were then expanded to a 'long-format' so that a unique row exists for each individual captured. The Design Team considered various options for constraining the temporal extent of data considered in this review; ultimately agreeing to focus on the post-POD period (2002-present). As such, the dataset was then reduced to include only samples collected after December 31st, 2001.

Standardization of Effort

The trawl nets used in the five CDFW review studies vary in mouth opening size from less than $0.5m^2$ for the SLS to larger than $13m^2$ for the SKT, and it is necessary to account for these marked differences in comparisons of catches between surveys. During the post-POD period, flowmeters have been consistently deployed along with the nets for all surveys, and it is therefore possible to calculate sample volumes for every tow. Volume was calculated as $V = (Fe - Fs)CM$, where V is the volume sampled, *Fe* and *Fs* are the ending and starting flowmeter readings, C is the flowmeter constant (0.0269 m/revolution, no adjustments or corrections were made to this value) and M is the area of the net mouth. Net mouth areas for each survey are as follows: SLS – 0.37 m^2, STN – 1.49 m^2, 20mm – 1.51 m^2, FMWT – 10.7 m^2 (after accounting for an assumed 80% net opening), SKT – 13.95 m^2.

During the post-POD period, sampling effort has remained relatively consistent across stations for each of the five CDFW surveys. However, a handful of stations have been sampled sporadically or dropped in recent years. In order to appropriately describe current, status quo sampling effort we removed samples from each survey for stations not sampled in each of the five most recent, complete years (2014-2019).

Integration of 'Other' Fish Survey Data

Although the five CDFW surveys were the focus of the review, they exist within a broader system of fish surveys that occur in the SFE monitoring enterprise; several of which overlap substantially in space, time and methodology with the five CDFW studies. In order to gain a more complete understanding of the frequency and intensity of fish sampling throughout a broader set of habitats, we obtained data from three additional long-term fish monitoring studies: USFWS's Delta Juvenile Fish Monitoring Program (DJFMP) and the associated Enhanced Delta Smelt Monitoring (EDSM), as well as CDFW's San Francisco Bay Study (Bay Study). These data presented additional challenges for integration and standardization because they are collected and maintained by several different agencies and use a variety of methods in addition to pelagic trawling (i.e., otter trawling, beach seining). As such, we did not try to fully standardize all aspects of the 'other' survey data, and rather focused only on appropriately characterizing the spatiotemporal distribution of sampling effort.

Data for the Bay Study were extracted from the R package 'LTMRData' that was a product of the recent LTMR effort. These data have been subjected to processing and filtering based on the goals of the LTMR effort and as such contain only selected taxa. They nevertheless retain information on all individual trawls conducted, and so are sufficient for an evaluation of sampling effort. DJFMP and EDSM data were obtained from their publicly available databases and integrated using the same approach described for the CDFW studies. Adjusted length frequencies and tows with no catch were also addressed in the same manner as for the five review studies. Pre-calculated volumes were recorded for most samples in each of these data sets including nearly all EDSM tows. EDSM data were further separated by gear type. Data derived from the 'other' studies were then filtered to exclude samples from before 2002 and joined with the CDFW study data table. All of these other studies, except for EDSM (which began in 2016) have data that overlap with the post-POD period selected for the review study data.

Disambiguation of Sampling Locations

The five pelagic review studies were initiated between the 1950s and 2000s to address questions specific to the ecological and management context of their respective eras. As such, in effect the surveys are independent entities, but many efforts to coordinate between surveys have occurred through the decades resulting in a challenging situation in which some aspects of the surveys are logically related, but others are not. One area that poses a challenge is the naming convention for survey stations. Each survey uses a three-digit identifier for each station, and 60 station identifiers are shared by at least two surveys (2 surveys: 21 stations, 3 surveys: 5 stations, 4 surveys: 5 stations, 5 surveys: 29 stations). In most cases, when the same identifier is used by multiple surveys the geographic location is shared, but there are numerous exceptions to this pattern. Integrated analysis of catch data requires reconciling which station identifiers differ in location between surveys. However, this process is complicated by the dynamic conditions of the Delta. The practical challenge of towing nets through this environment means that the recorded locations of sampling stations must be understood as approximate. Moreover, the relevance of a given geographic discrepancy will vary across regions, with minor variation in location unlikely to be of much importance in open water regions, but potentially more impactful in channel and slough areas. As such, there is some scope for minor variation in the station coordinates between surveys, but such variation must be considered on a case-by-case basis.

In total, 190 station identifiers have been used across the five studies. We first selected only stations for which the identifier was used in at least two CDFW studies (n=60). Data for the 'other' fish studies were not considered. We then calculated the pairwise distances between all stations within and between the five studies. For each station identifier we then calculated the maximum distance between locations (e.g., the distance between FMWT station 323 and STN station 323). We next excluded from further analysis any station identifier for which all locations were within 1km. Examination of trawl start and endpoints indicated that the majority of tow paths fall within ~0.5km of the listed station coordinates, and so a difference of >1km indicates little possibility of overlap. After these exclusions, 25 stations remained with potentially consequential distances between sampling locations. To allow for visual comparison at an appropriate resolution, we divided these stations into four groups and plotted the station coordinates on a map of the Delta (**Figure 1-1**). The design team then considered these geographic discrepancies on a case-by-case basis. During this exercise, it was pointed out that even smaller geographic differences may prove consequential if studies are targeting different depths (e.g., shoal vs. channel). Thus, for each station identifier we calculated the mean recorded bottom depth and 95% confidence interval by study and visualized for all stations where the maximum difference between studies was greater than five feet depth (**Figure 1-2**).

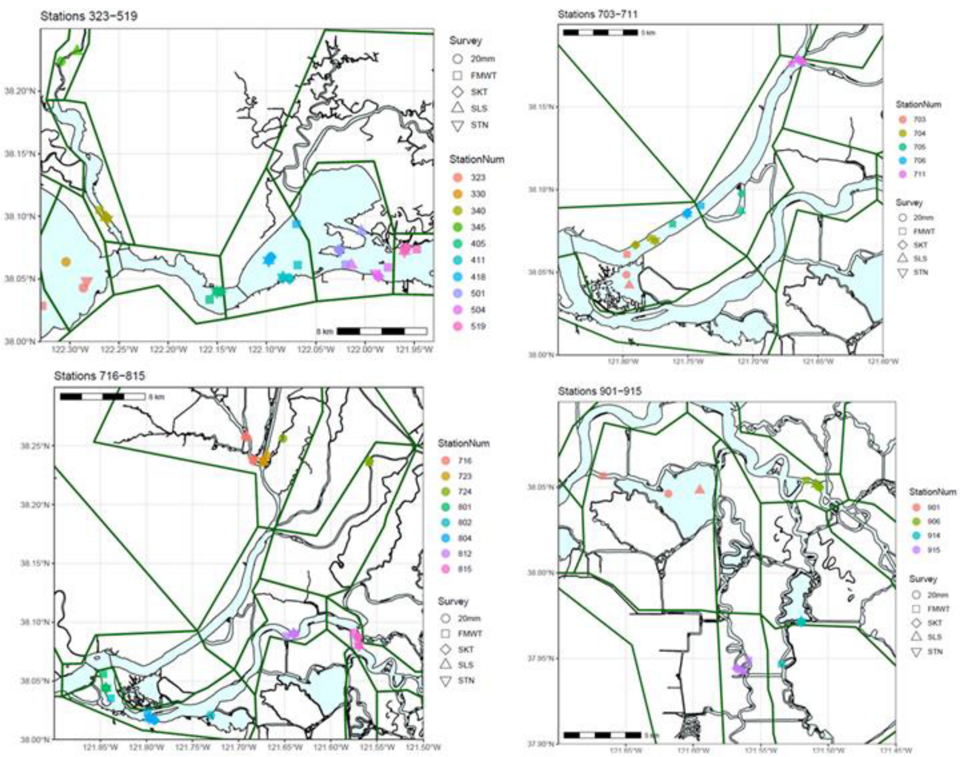


Figure 1-1. Locations of CDFW sampling stations where coordinates differ by 1km or more between CDFW studies.

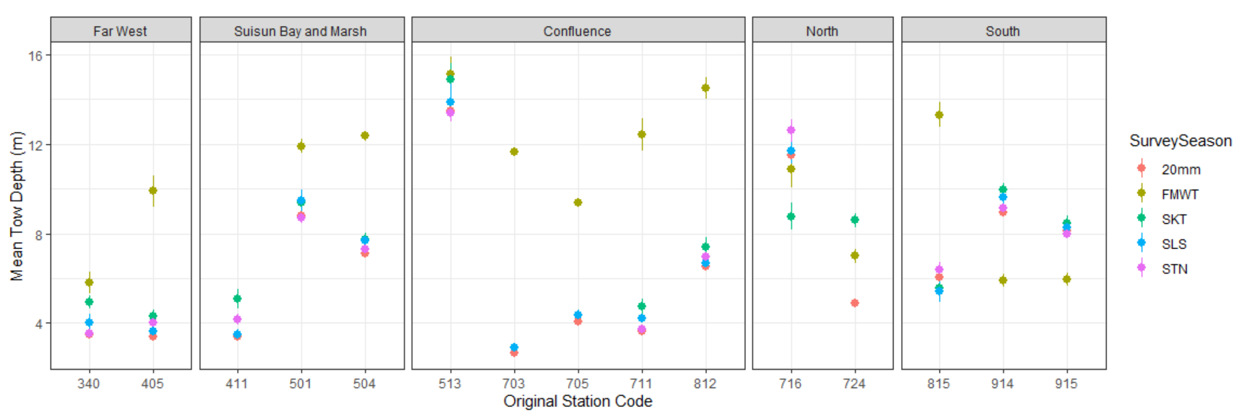


Figure 1-2. Mean bottom depths by study for CDFW trawl sampling stations where mean depth differs by more than 5 ft. between studies.

Through the evaluation of differences in station coordinates and recorded bottom depths, a total of 21 new station identifiers were added to the dataset by adding a decimal of 0.1 – or in the two cases where multiple differences existed for an original identifier, 0.2 – to the original station code (e.g., 716.1, 504.2). After these additions and excluding stations not sampled consistently during the post-POD period, a total of 159 distinct sampling locations were included in the review dataset (**Table 1-1**).

Table 1-1. List of distinct sampling locations.

Station Code	Surveys	Station Code	Surveys	Station Code	Surveys
305	FMWT	418.1	FMWT	719	FMWT,STN,SKT,20mm
306	FMWT	501	STN,SKT,20mm	720	20mm
307	FMWT	501.1	FMWT	721	FMWT
308	FMWT	501.2	SLS	723	FMWT,STN,SLS,20mm
309	FMWT	502	FMWT	724	FMWT,SKT
310	FMWT	503	FMWT	724.1	20mm
311	FMWT	504	STN,SKT,20mm	726	20mm
314	FMWT	504.1	FMWT	735	FMWT
315	FMWT	504.2	SLS	736	FMWT
321	FMWT	505	FMWT	794	FMWT
322	FMWT	507	FMWT	795	FMWT,STN
323	STN,20mm	508	All	796	FMWT,STN
323.1	FMWT	509	FMWT	797	FMWT,STN
325	FMWT	510	FMWT	801	STN,SKT,SLS,20mm
326	FMWT	511	FMWT	802.1	FMWT
327	FMWT	512	FMWT	804	STN,SKT,SLS,20mm
328	FMWT	513	All	804.1	FMWT
329	FMWT	515	FMWT	806	FMWT
334	FMWT	516	FMWT	807	FMWT
335	FMWT	517	FMWT	808	FMWT
336	FMWT	518	FMWT	809	All
337	FMWT	519	All	810	FMWT
338	FMWT	520	STN,SKT,SLS,20mm	811	FMWT
339	FMWT	601	FMWT	812	STN,SKT,SLS,20mm
340	All	602	All	812.1	FMWT
341	FMWT	603	FMWT	813	FMWT
342	SLS,20mm	604	FMWT	814	FMWT
343	SLS,20mm	605	FMWT	815	STN,SKT,SLS,20mm
344	SLS,20mm	606	All	815.1	FMWT
345	SLS,20mm	608	FMWT	901	SLS

Station Code	Surveys	Station Code	Surveys	Station Code	Surveys
346	SLS,20mm	609	STN,SKT,SLS,20mm	901.1	20mm
347	SLS	610	STN,SKT,SLS,20mm	902	All
348	SLS	701	FMWT	903	FMWT
349	SLS	703	SLS,20mm	904	FMWT
401	FMWT	703.1	FMWT	905	FMWT
401.1	FMWT	704	All	906	All
403	FMWT	705	SLS,20mm	908	FMWT
404	FMWT	705.1	FMWT	909	FMWT
405	STN,SKT,SLS,20mm	706	STN,SKT,SLS,20mm	910	All
405.1	FMWT	706.1	FMWT	911	FMWT
406	FMWT	707	All	912	All
407	FMWT	708	FMWT	913	FMWT
408	FMWT	709	FMWT	914	STN,SKT,SLS,20mm
409	FMWT	710	FMWT	914.1	FMWT
410	FMWT	711	STN,SKT,SLS,20mm	915	STN,SKT,SLS,20mm
411	STN,SKT,SLS,20mm	711.1	FMWT	915.1	FMWT
412	FMWT	712	FMWT,SKT	918	STN,SLS,20mm
413	FMWT	713	FMWT,STN,SKT	919	All
414	FMWT	715	FMWT,SKT	920	FMWT,SKT
415	FMWT	716	FMWT,STN,20mm	921	FMWT,SKT
416	FMWT	716.1	SKT,SLS	922	FMWT,SKT
417	FMWT	717	FMWT	923	FMWT,SKT
418	STN,SKT,SLS,20mm	718	20mm		

Catch-per-tow and CPUV
Analyses of catch focused on a selected group of species that were separated into age-classes based on a length-at-date analysis (See Appendix 3A for details). After filtering the long-format dataset to this group of species to focus on for evaluation, calculated ages were appended to common names resulting in 21 species-age combinations, plus the No_Catch designation. This reduced table was then pivoted to a wide format with one row for each trawl tow (or beach seine set, for 'other' fish studies), and 21 columns containing species- and age-specific catches. These catch-per-tow values were then converted to catch-per-unit-volume (CPUV) by dividing each column by the tow volume column. In order to produce more manageable values, volumes were first divided by 1,000, resulting in CPUV units of catch per 1,000m^3.

Summary of Findings
Table 1-2 illustrates the comparison of the current fish sampling effort conducted by the five review studies relative to the 'other' studies on an annual basis. Of the five CDFW studies, the 20-mm Survey conducts the greatest number of pelagic trawls (bi-monthly), while the FMWT

(monthly) samples the largest pelagic volume during the year. In comparison, EDSM, DJFMP, and the Bay Study all sample larger pelagic volumes than the review studies. Additionally, the DJFMP conducts more than 1,700 beach seine hauls and the Bay Study conducts over 500 benthic trawls per year.

Table 1-2. Summary of the status quo of annual sampling effort for CDFW and other fish studies. Volumes are in units of 1,000m^3.

	SLS		20mm		STN		FMWT		SKT	
	Tows	Vol.	Tows	Vol.	Tows	Vol.	Tows	Vol.	Tows	Vol.
Beach Seine	0	0	0	0	0	0	0	0	0	0
Surface Trawl	0	0	0	0	0	0	0	0	196	1,277
Oblique Trawl	256	47	1,120	1,033	589	519	475	2,467		
Benthic Trawl	0	0	0	0	0	0	0	0	0	0
	EDSM-Kodiak		EDSM-20mm		DJFMP		Bay Study			
	Tows	Vol.			Tows	Vol.	Tows	Vol.		
Beach Seine	0	0	0	0	1,718	55	0	0		
Surface Trawl	2,395	8,906	1,420	1,353	3,030 *	43,741*	0	0		
Oblique Trawl	0	0	0	0			516	3,594		
Benthic Trawl	0	0	0	0	0	0	510	NA		

*DFJMP conducts both surface trawls and fixed-depth midwater trawls.

Appendix 2. Spatial and Temporal Evaluation

Introduction

The Design Team evaluated the spatial sampling designs of the five studies to evaluate the habitats sampled and whether data gaps exist in spatial and temporal effort. This task was conducted within the context of a "sample frame" that is explicit about the depths, habitats, and strata represented in the monitoring designs, and the timing and frequency of sampling events and surveys.

Sample Frame

The first step in developing the sample frame began with parsing the Delta into 30 strata overlayed on the combined dataset of CDFW monitoring stations. The basis for the initial stratification was the polygons used by CDFW and EDSM for calculating Delta Smelt indices, which was historically based on where similar environmental conditions and bathymetry occur (Mitchell et al., 2017; Newman, 2008; Polansky and Studen, 2019). This level of regional stratification was deemed to provide sufficiently fine-scale "building blocks" for clustering into coarser spatial groupings based on historical patterns of catch.

The coordinates of all sampling locations were spatially joined with the 30 polygons and bathymetry dataset to facilitate summarization of the distribution of sampling effort. Depths were classified as Shallow (<2m), Intermediate (2-4m), Deep (4-10m) and Very Deep (>10m). Regional classifications included 30 "Subregions" (**Figure 2-1**). The five CDFW review studies sampling locations fall almost entirely within the extent of the 30 subregions. The initial overlay revealed that one or more CDFW sampling locations aligned with 26 of the 30 strata level polygons. However, several of the FMWT stations in San Pablo Bay were located just beyond the western extent of the Far West stratum (stations 305-307). Thus, the Far West region and Mid San Pablo Bay polygons were manually edited in ArcGIS to be inclusive of all stations. A larger number of sampling locations from 'other' fish trawl studies (EDSM, DJFMP, Bay Study) also fell outside the 30 strata. Stations located seaward beyond the extent of the Far West region were given a subregion designation San Francisco and Outer San Pablo Bays. Stations located upstream of the regional extent of the polygons were assigned to either the Upper Sacramento or Upper San Joaquin subregions, as appropriate.

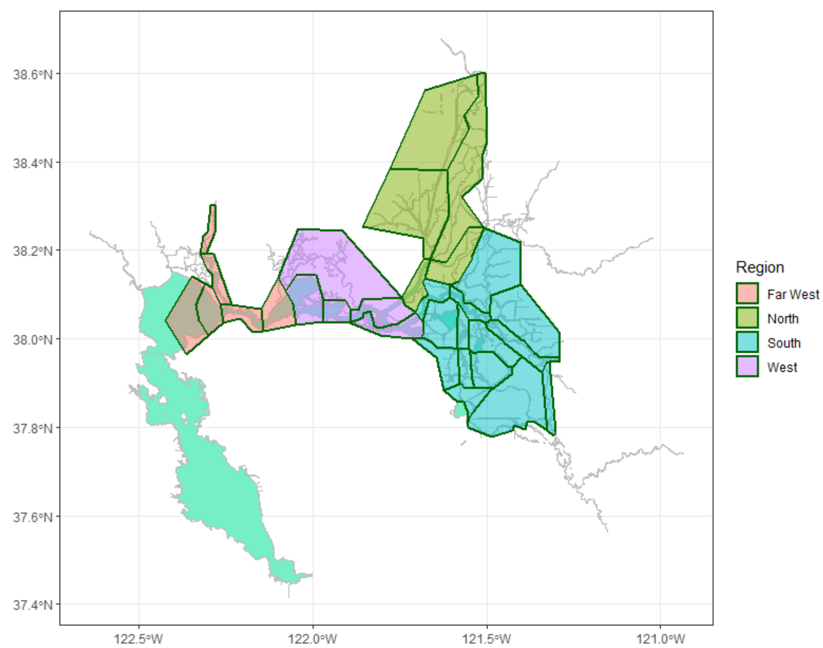


Figure 2-1. Spatial designations used to develop proposed stratification.

Regional Stratification

Given the distribution and intensity of sampling by the five CDFW studies, the Design Team sought to evaluate support for regional stratification of the sample frame based on the environmental conditions of the Delta and historical catches in the CDFW surveys. A multivariate clustering approach was used to identify candidate spatial groupings based on catch and environmental factors.

Although water quality data are collected in association with each tow conducted for the five CDFW studies under review, these samples represent a small fraction of the total discrete water quality data collected throughout the Delta on a regular basis. A prior data integration effort compiled and standardized data from all of the CDFW studies except the SLS with data from eight other monitoring programs (Bashevkin, 2021). These environmental data were obtained from the EDI data portal, appended to the discrete environmental data, and used this much more comprehensive environmental database for the environmental inputs to the clustering analysis.

Surface temperature, secchi depth and salinity/conductivity were by far the most consistently collected environmental variables. Where only conductivity was recorded, it was converted to salinity. Next for each of these variables and in each subregion and year, an annual mean and coefficient of variation (CV) was estimated. Subsequently, the mean CV across all years was determined for each variable, as an indicator of within-year variability (i.e., spatial difference and seasonality), and the CV of the annual means as an indicator of between-year variability (i.e., sensitivity to hydrologic conditions). Finally, the overall subregion means for each variable across all years was calculated (**Figure 2-2**). These metrics – nine in total, three each for salinity, secchi depth and temperature – were used as the input for the clustering analysis. The environment-based clustering was conducted for all months combined, and for subsets of months matched to each of the CDFW studies (e.g., January-March for SLS).

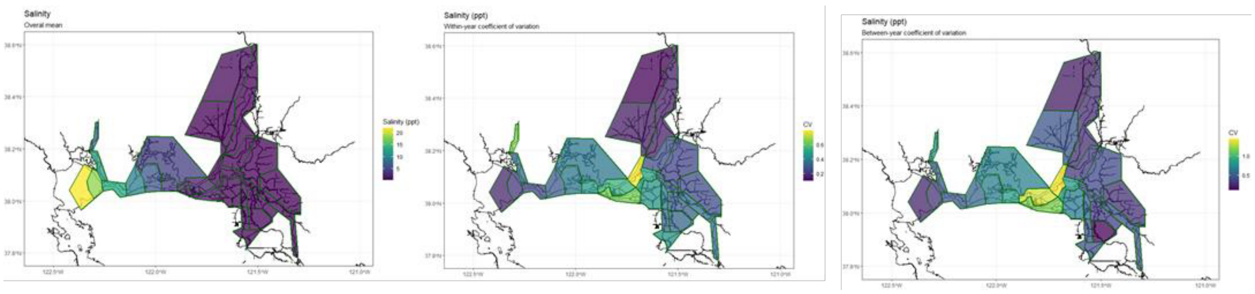


Figure 2-2. Example of environmental inputs to hierarchical clustering analysis.

For catch-based clustering, the tow-level data was separated by survey, station, and subregion to calculate mean CPUVs for each of the 21 species-age groups. Each of the resulting matrices was then used as an input to the hierarchical clustering analysis. Hierarchical agglomerative clustering was then performed separately on the 10 catch matrices (5 station level and 5 subregion level) and six environmental matrices (1 annual and 5 matched to the CDFW survey months) using the 'hclust' function in R. Prior to calculation of the dissimilarity matrix, catch data were cube root transformed and both catch and environmental data were standardized to zero mean and unit variance. Dissimilarity matrices were calculated using Euclidian distance, and then input to the 'hclust' function where Ward's minimum variance algorithm was used to derive the clustering structure.

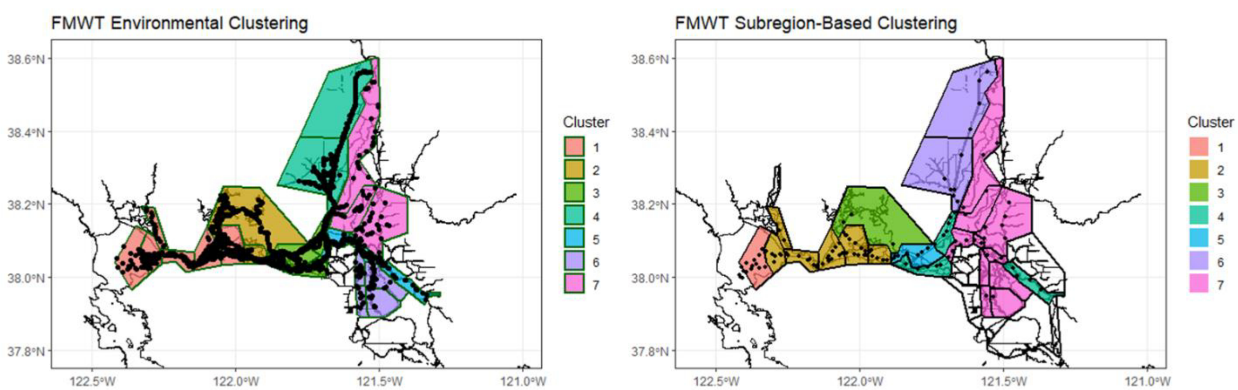


Figure 2-3. Example results of multivariate clustering of 30 subregions based on the September-December multi-survey environmental data (left) and FMWT catch data (right).

Multivariate clustering does not typically prescribe a "correct" number of clusters, but examination of dendrograms provided general support for 5-8 clusters. This range of clusters also generally resulted in groupings of contiguous subregions (**Figure 2-3**).

Clustering results reflected a dominant east-west gradient in both the fish community and environmental conditions, with greater variability in the degree of north-south separation (**Figure 2-4**). In addition to this dominant gradient, several subregions (i.e., the stations within these regions), including Suisun Marsh and Cache Slough commonly clustered separately from adjacent areas (**Figure 2-5**). It then was necessary to reconcile the potential regional clusters across seasons, as the Design Team agreed that a fixed system of regional stratification was preferable to a seasonally variable or environmentally dynamic system.

As an alternative, two-level of stratification was developed. The coarser level has five large regions and captures the dominant east-west gradient in catch composition and environmental conditions, while the finer scale incorporates some of the local differences in catch composition that are likely associated with distinct habitat types (e.g., Suisun Marsh vs. Suisun and Honker Bays). It is not expected that all surveys or species will be analyzed at the finer level of stratification, but in some cases where sampling effort is sufficient it will likely be warranted to separate a larger region. For example, catch patterns and environmental conditions during the SLS appear to justify addressing the Napa River separately from the rest of the Far West Region.

Once proposed strata were assembled in ArcGIS, the Design Team proceeded with defining the habitat and evaluating depths that should be represented in the sample frame, in order to compare "sampled" relative to "total" water volumes. The Design Team felt strongly that the surveys under review should be evaluated in the context that they are explicitly pelagic – as each of the gear types are designed to sample fishes occurring within the water column. Furthermore, each of the gears have limitations on depths where they can be deployed in the pelagic environment, further constraining the sample frame to depths > 6 ft. (1.8m). For this reason, the Design Team worked to refine the sample frame to define the depths and habitats that should be represented.

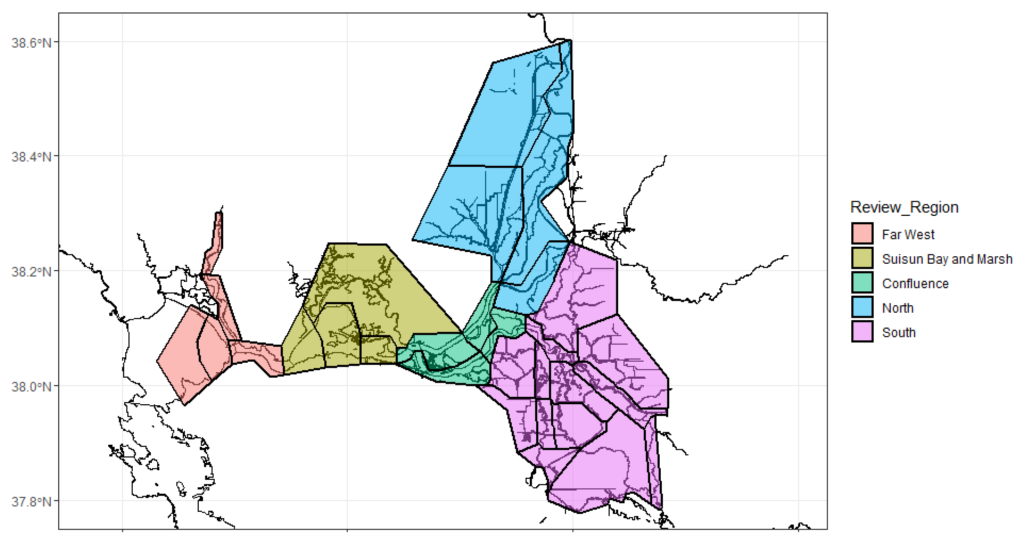


Figure 2-4. Overview of the Upper San Francisco Estuary illustrating the "Region" designations.

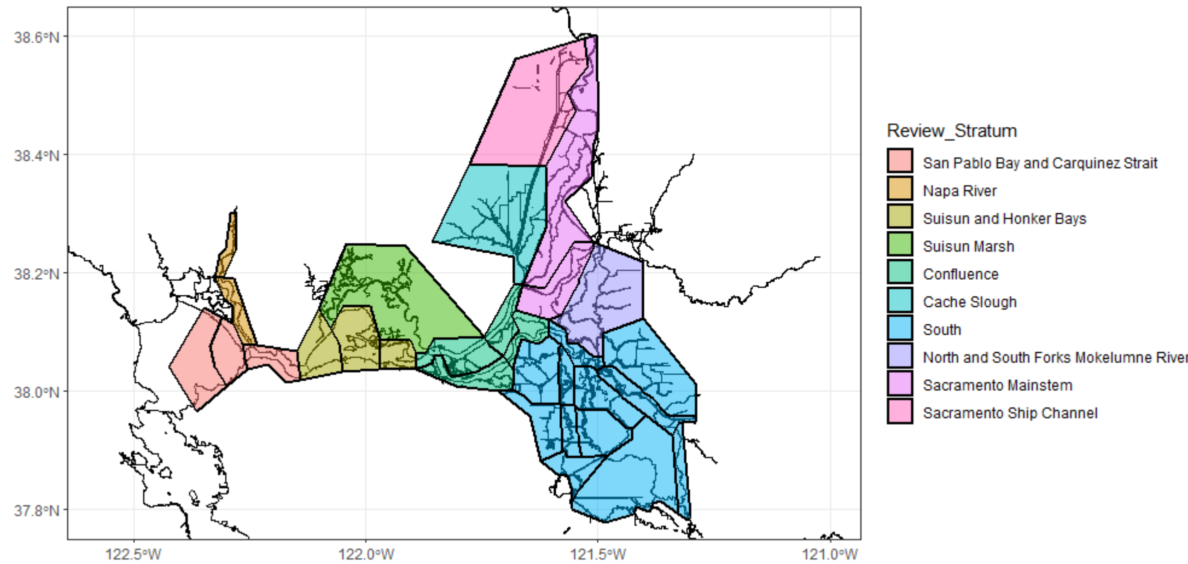


Figure 2-5. Overview of the Upper San Francisco illustrating the "Strata" designations.

Depth Sampled

Through evaluation of the sampled depths by the five gear types over time, each was found to represent a different tow depth resulting in unique volume calculations. For the SKT that is the only surface deployed gear type, a standard volume based on the deployed gear depth of 1.8 m was used. For the four oblique gear types, the depth of each tow was estimated using the angle at which the trawl was deployed, the length of the cable released, and the block height (the height from the water surface to the block from which the cable is released). For the FMWT, the tow depth calculation resulted in the volume between the surface and 10 m depth being used. For the STN, the volume between the surface and 9 m depth was used. For 20-mm, the volume between the surface to 8 m depth was used, and finally, for the SLS, the volume between the surface to 13 m depth was used. Since measurements of depth-based occupancy are not available, it is important to acknowledge that the depth strata are solely based on the depths surveyed and may not reflect the true occupancy depth across species. As a result, the volume of water used to translate a survey catch to an estimate of relative abundance will almost always include portions of the water column that will be unoccupied by one or more species of interest. Note, trawl vessels and gear are limited to operating in depths of 1.8 m (6 foot) or deeper.

Pelagic Habitat

To calculate water volumes, the maximum volume of water likely to be sampled by each gear was calculated from raster files describing the bathymetry of the Delta (Wang et al., 2019). To explicitly define the pelagic habitat, the bathymetry data was additionally constrained to the California Aquatic Resources Inventory (CARI) dataset that includes habitat classifications for the entire Bay-Delta region (San Francisco Estuary Institute, 2017). With acknowledgement that data gaps exist, particularly where extensive tidal marsh restoration are on-going, these data were intersected with the bathymetry and spatial strata to evaluate the habitat types present in the sample frame. This evaluation showed that the majority of the sample frame corresponded to habitat types in the “subtidal water”, “tidal channel”, and “fluvial channel” categories. **Figures 2-6 and 2-7** illustrate the total water volumes in the sample frame corresponding to the 5 regions and 10 strata, respectively, relative to the volumes associated with only the pelagic (> 6ft) sample frame extent. This analysis revealed that the total water volumes of each stratum are predominantly comprised of the three habitat categories, and thus the Design Team agreed to use this as the definition of pelagic habitat. Furthermore, by intersecting the CARI later with the CDFW monitoring locations in ArcGIS revealed that they are already aligned with this definition of pelagic habitat. For this reason, the Design Team agreed upon a refinement of the sample frame to only encompass pelagic waters, and sample volumes were adjusted accordingly (**Table 2.1**).

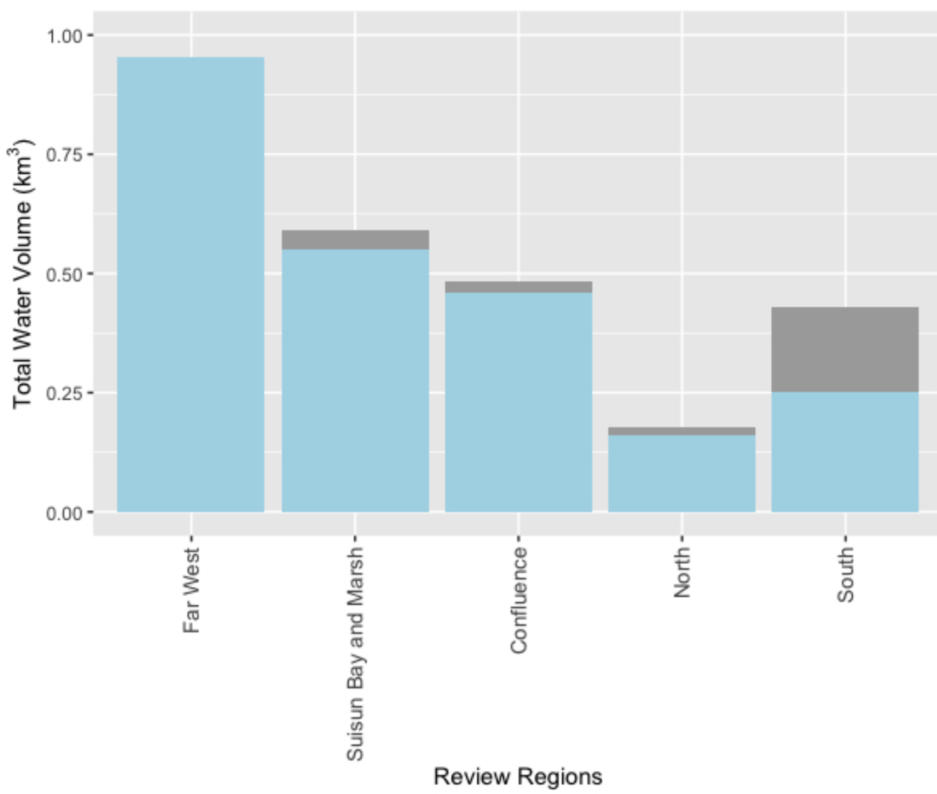


Figure 2-6. Total water volume (grey bars) and pelagic water volume (blue bars) in the Review Regions.

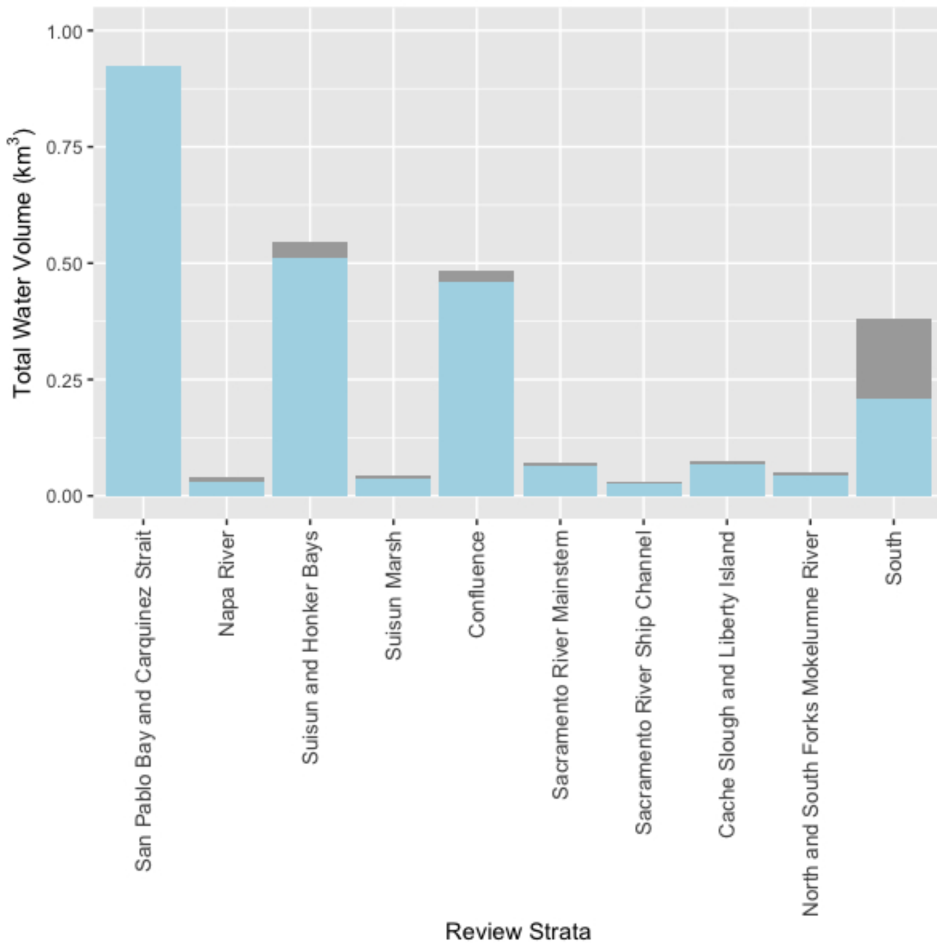


Figure 2-7. Total water volume (grey bars) and pelagic water volume (blue bars) in the Review Strata

Table 2-1. Regional Volumes for Extrapolation of the Pelagic Review Studies. Water volumes were constrained to pelagic habitat in waters 6ft. of deeper.

		Gear-Specific Water Volumes (1000s of m^3)				
Review Region	Review Stratum	SLS	20-mm	SKT	STN	FMWT
Far West	San Pablo Bay and Carquinez Strait	0.843	0.701	0.249	0.739	0.772
	Napa River	0.031	0.029	0.009	0.030	0.031
Suisun Bay and Marsh	Suisun / Honker Bays	0.486	0.409	0.144	0.432	0.451
	Suisun Marsh	0.038	0.036	0.012	0.037	0.038
Confluence	Confluence	0.444	0.363	0.102	0.389	0.409
North	Cache Slough	0.067	0.056	0.017	0.059	0.062
	Sacramento Mainstem	0.066	0.064	0.021	0.065	0.065
	Sacramento DWSC	0.026	0.021	0.005	0.023	0.025
South	North and South Forks Mokelumne River	0.043	0.041	0.014	0.042	0.042
	South	0.192	0.169	0.005	0.176	0.182

Sampled Depths

To address the question of what the depths are sampled by the surveys, the Design Team evaluated the annual effort conducted in the 10 Review Strata categorized by depth (**Figure 2-8**). The five review studies were put in the context of other pelagic, benthic, and beach seine surveys. This analysis provided insight to the relative sample effort with depth across surveys.

Eight of the 10 Subregions are sampled in at least one of the depth categories by each of the studies. The two strata that are not well sampled by the review studies occur in the Sacramento River Deep Water Ship Channel and Mainstem. The Sacramento Deep Water Ship Channel is sampled by FMWT and STN only, entirely at Very Deep (>10m) depths; with STN conducting twice the number of tows than FMWT. EDSM is the only one of the 'other' studies that samples in the Ship Channel with significantly more effort annually (~ 8x as many tows) at both Deep (4-10 m) and Very Deep (>10m) depths. On the Sacramento River mainstem, the SKT and FMWT surveys conduct 10-15 tows per year, while EDSM conduct over 200 tows per year in Deep and Intermediate (2-4m) depths. DJFMP is the only study that samples the Shallow (<2m) depths of this stratum. All of the remaining eight strata are sampled to some degree by each of the review studies. FMWT most frequently samples at Very Deep depths, particularly in the Confluence, Suisun and Honker Bays, and San Pablo Bay. In comparison, STN and 20-mm tend to conduct tows at shallower depths in these areas. Notably of the five review studies, shallow water depths are only frequently sampled in the Napa River stratum by SLS and to a less extent by the 20-mm Survey. None of the 'other' studies are comparable in effort in the shallow water depths of the Napa River, though EDSM does conduct tows in this stratum at deeper depths. The most frequently sampled areas of the SFE by the 'other' studies occur in San Francisco and

Outer Bays for the Bay Study; in the South, and Suisun and Honker Bays for DJFMP; and throughout the South, Cache Slough/Liberty Island, and Confluence for EDSM.

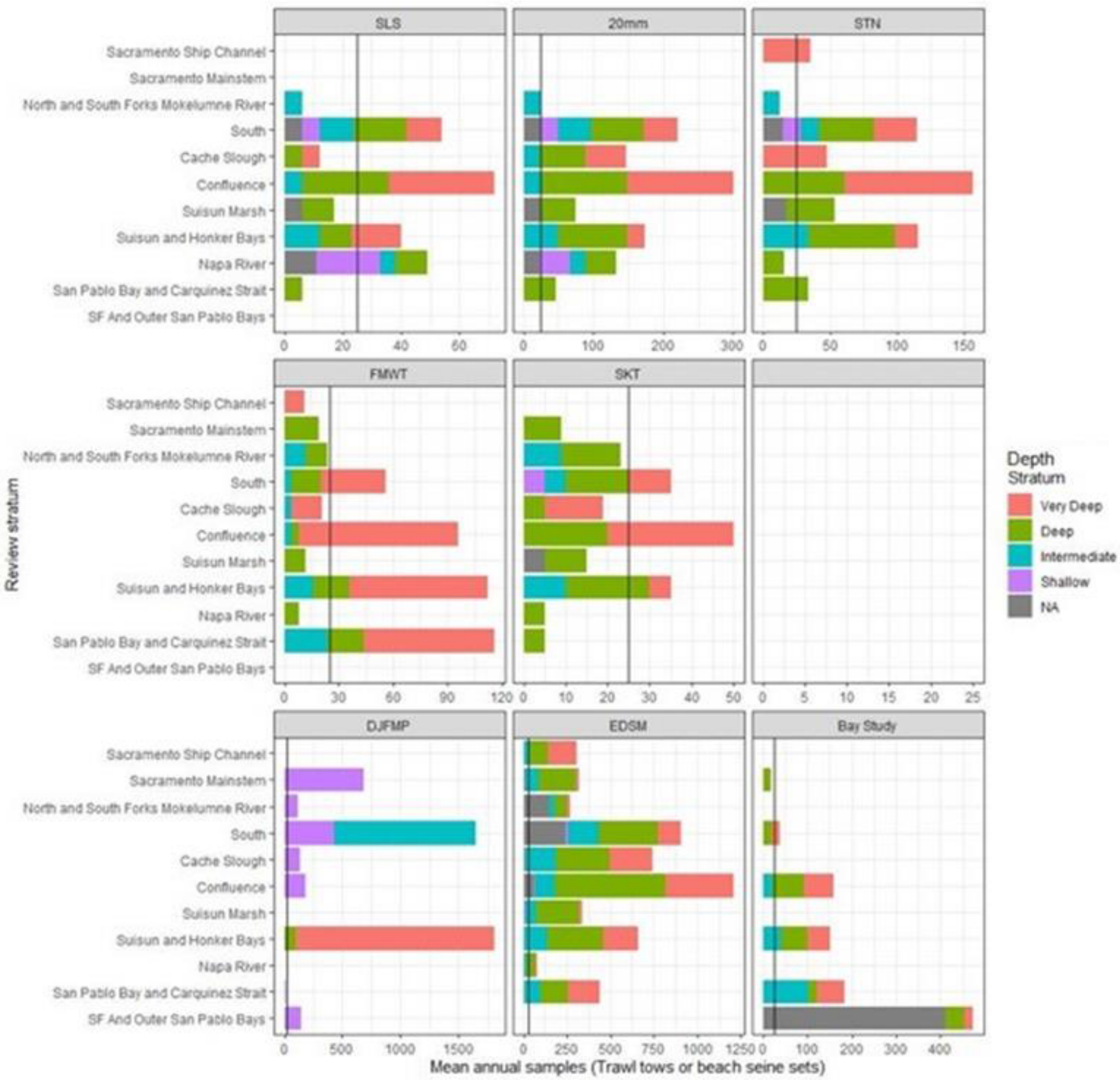


Figure 2-8. Average number of tows per year for the CDFW and Other studies (including pelagic, benthic, and beach seine sets) by Subregion and Depth Stratum (see Sample Frame section for description of categories). Note differences in the scale of the x-axis for each survey.

Regional Sampling Effort

In order to appropriately characterize regional abundances and uncertainty, it is necessary to have sufficient sampling effort across all regions of interest. The Design Team evaluated the distribution of sampling effort as the annual average number of tows, the annual average sampled volume, and the monthly average sampled volume. The spatial and temporal effort was also synthesized into heatmaps to depict the stations and regions that are currently sampled by each of the studies.

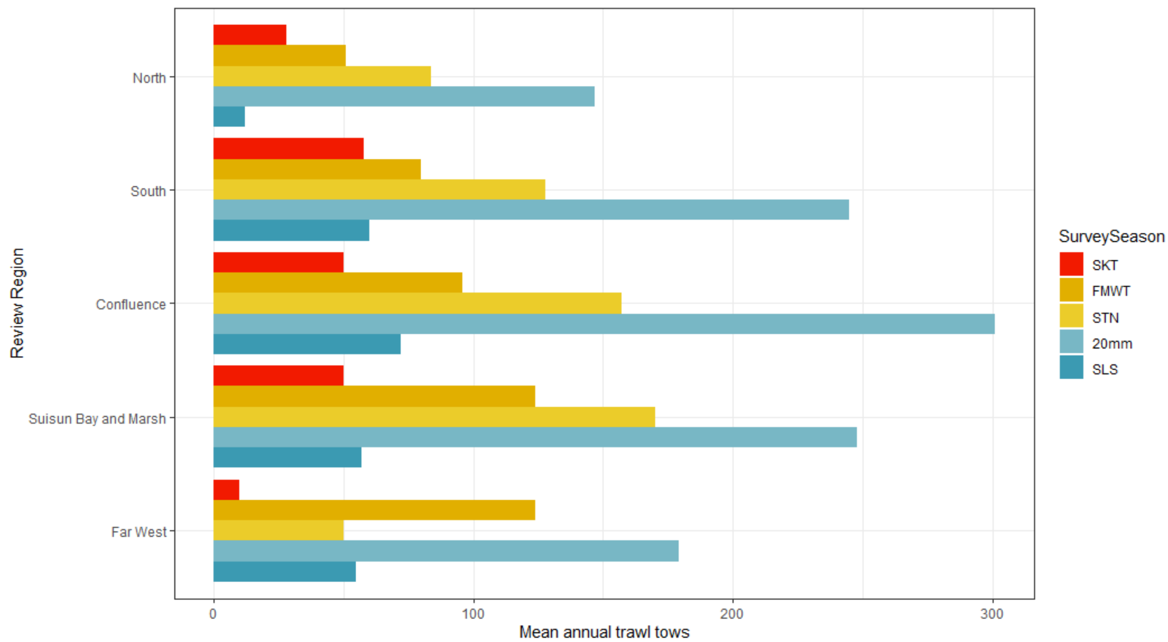


Figure 2-9. Average tows per year by region under status quo sampling effort.

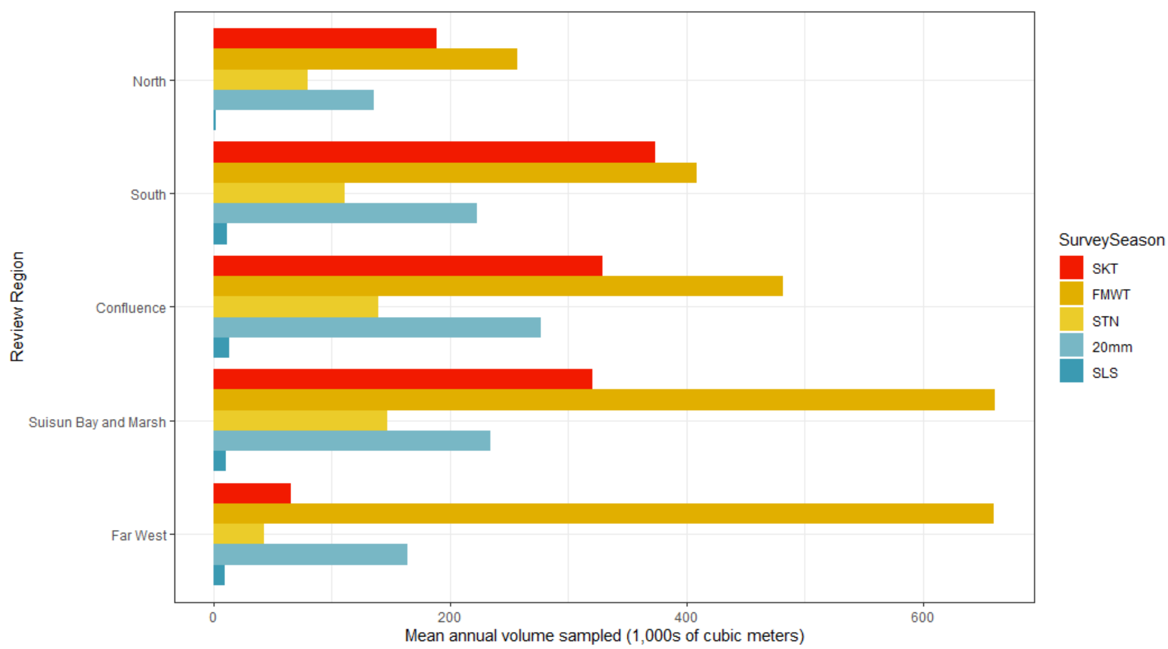


Figure 2-10. Average volume sampled per year by region under status quo sampling effort.

In terms of number of tows, the five studies all sample at varying levels of effort across the Review Regions. SLS exhibits the most balanced effort of the five studies, with 50-75 tows conducted annually in all regions, except for the North. 20mm and STN conduct the highest number tows due to the use of replication, except in the Far West where the STN has much fewer stations relative to the FMWT (**Figure 2-9**). The number of tows in 20mm has the greatest regional variability, with < 150 tows per year in the North to >300 tows per year in the Confluence. SKT conducts the least amount of tows per region (< 75 per year) except in the North, where SLS has less effort. For the FMWT, the Far West and Suisun Bay / Marsh are sampled with significantly more effort (~ 125 tows per year) than the North and South (50 and 75 tows, respectively). These patterns suggest an overall imbalance to the sampling effort.

Evaluating effort as the amount of water volume sampled shows even greater variability across regions. In particular, the SKT is poorly represented in the Far West and the SLS is poorly represented in the North. The large net sizes of the SKT and FMWT result in the largest volumes sampled relative to the number of tows conducted (**Figure 2-10**), and so these surveys sample the most volume across all regional strata. FMWT exhibits the highest amount of volume sampled in the Far West and the least in the North. In contrast, 20mm, STN, and SKT all survey the most volume in the Confluence, Suisun Bay/Marsh and South.

To further evaluate regionally stratified effort, 20mm, STN, and FMWT sample volumes were further constrained by month and Review Strata (**Figure 2-11**). This evaluation provided more clarity on the imbalance to the sampling effort in certain areas and times. Sampling in San Pablo Bay/Carquinez Strait is mostly intensive during September through December by FMWT, while Napa River is sampled most intensively by 20mm during March through June. Suisun / Honker Bays is also heavily weighted towards the FMWT, but Suisun Marsh is sampled at similar volumes among the three studies. The Confluence, Cache Slough and the South show similar patterns in effort where the FMWT samples much higher volumes, followed by 20mm, and least by the STN. The Sacramento mainstem is notable for only being sampled by FMWT. Lastly, both the Sacramento DWSC and North/South Forks of Mokelumne River are overwhelmingly sampled at higher volumes by FMWT compared to the other two studies.

The synthesis of the spatial and temporal effort of the studies was evaluated by the use of heatmaps to depict the stations and regions that are currently sampled by each of the studies. Evaluation of spatial effort (**Figure 2-12**) revealed that only the FMWT currently samples stations in San Pablo Bay with relatively high intensity (31 stations). SKT and STN sample at only 3 or fewer stations in the Far West, and 20-mm and SLS currently sample 7-8 stations, largely in Napa River. Only one station in the Far West (# 340) is sampled in all five surveys. In Suisun and Honker Bays and the Confluence, most of the surveys sample 10-12 overlapping stations. Four stations in both Suisun and Honker Bays and the Confluence are sampled in all five surveys. The North Region has between 2-7 stations for SLS, 20-mm, STN, and SKT surveys. FMWT currently samples 15 stations in this Region. Only one station (# 723) in the North is sampled in all five surveys. Lastly, 10 - 15 stations are monitored in the South by SLS, 20mm and STN. The FMWT Survey currently samples 20 stations in the South. Five stations in the South stratum are sampled in all five surveys. As a result of this evaluation, a clear spatial imbalance of the sampling effort was apparent among the individual studies, as certain regions were found to not be as well as sampled as others. The rationale for these differences was identified to be largely attributable to where management needs for real-time information relative to status and trends.

Figure 2-11. Average monthly sample volume by study and stratum for the Status and Trends Program. Error bars show standard errors and reflect year-to-year variability in sample volumes.

In order to have a consistent evaluation of the temporal sequence of monitoring, the studies were separated into those that support real-time operations (SLS, 20-mm, and SKT; **Figure 2-13)** and status and trends (STN, FMWT, and 20mm; **Figure 2-14**). The rationale for including 20mm in both monitoring designs is further addressed in Appendix 3. Real-time information is generated on a bi-monthly basis by both SLS and 20-mm surveys, and supported by the monthly surface tows by SKT. During the months of March through May the most intensive period of real-time surveys is collected as a result of SKT surveys overlapping with SLS and 20mm. These surveys sample at the same or similar stations providing consistent information on spatial patterns and detections from these studies. The timing of this monitoring supports real-time information during the spring months that are the most critical periods for entrainment.

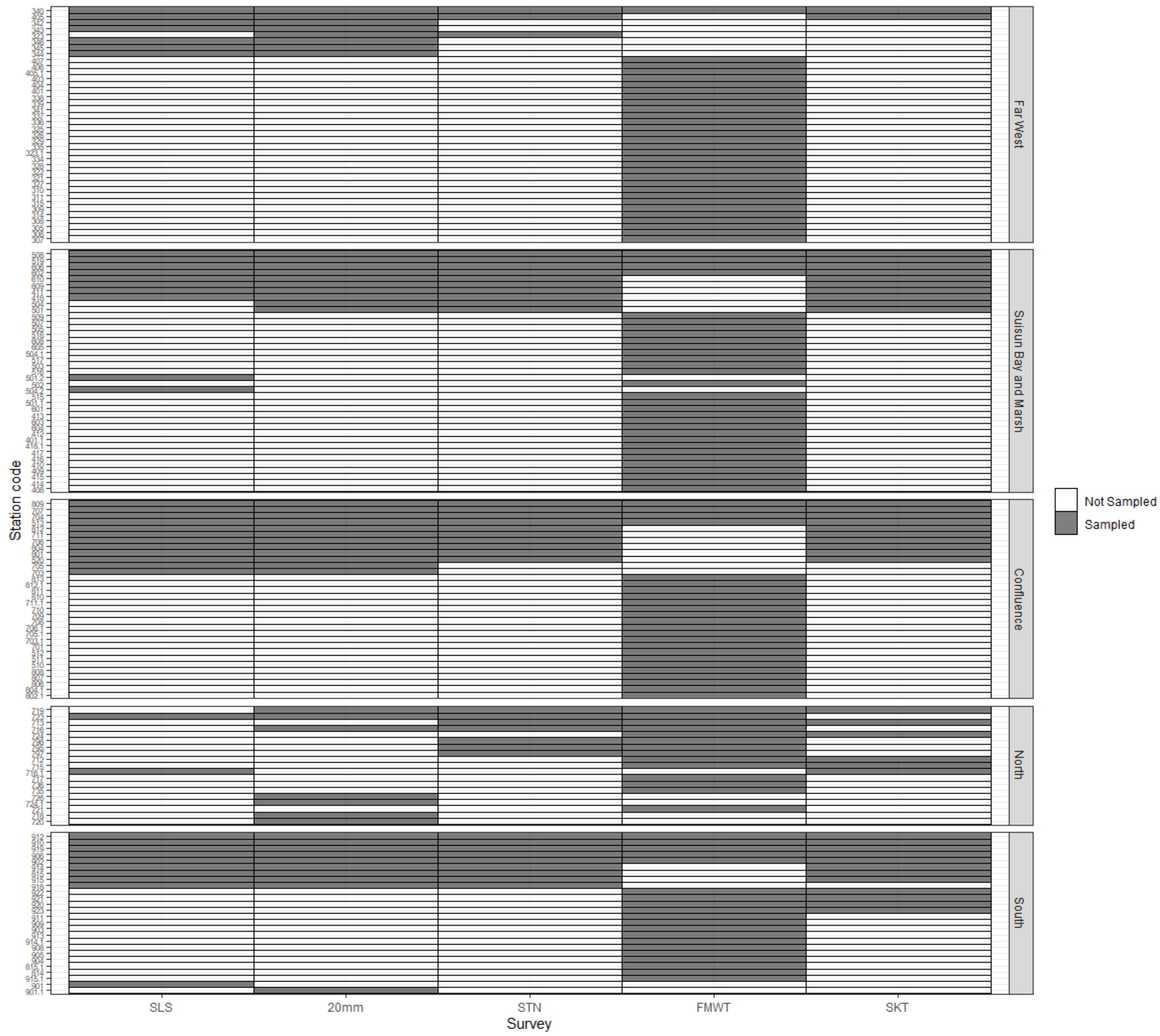


Figure 2-12. Sampling effort across stations and regions for each of the five CDFW studies.

Status and trends information is generated bi-monthly for 20mm and STN, and on a monthly basis for FMWT. FMWT provides the most extensive spatial information, during the months of September through December. One might expect the least extensive monitoring information on status and trends to occur when FMWT is not underway. However, some of the gaps in status and trends information during the March through June period appear to be addressed by the 20-mm survey. As a result of this evaluation and separation of the review studies, no obvious gaps in the temporal sequence were identified. Furthermore, it appears most of the current stations have been sampled relatively consistently over time, providing a basis for informing trends.

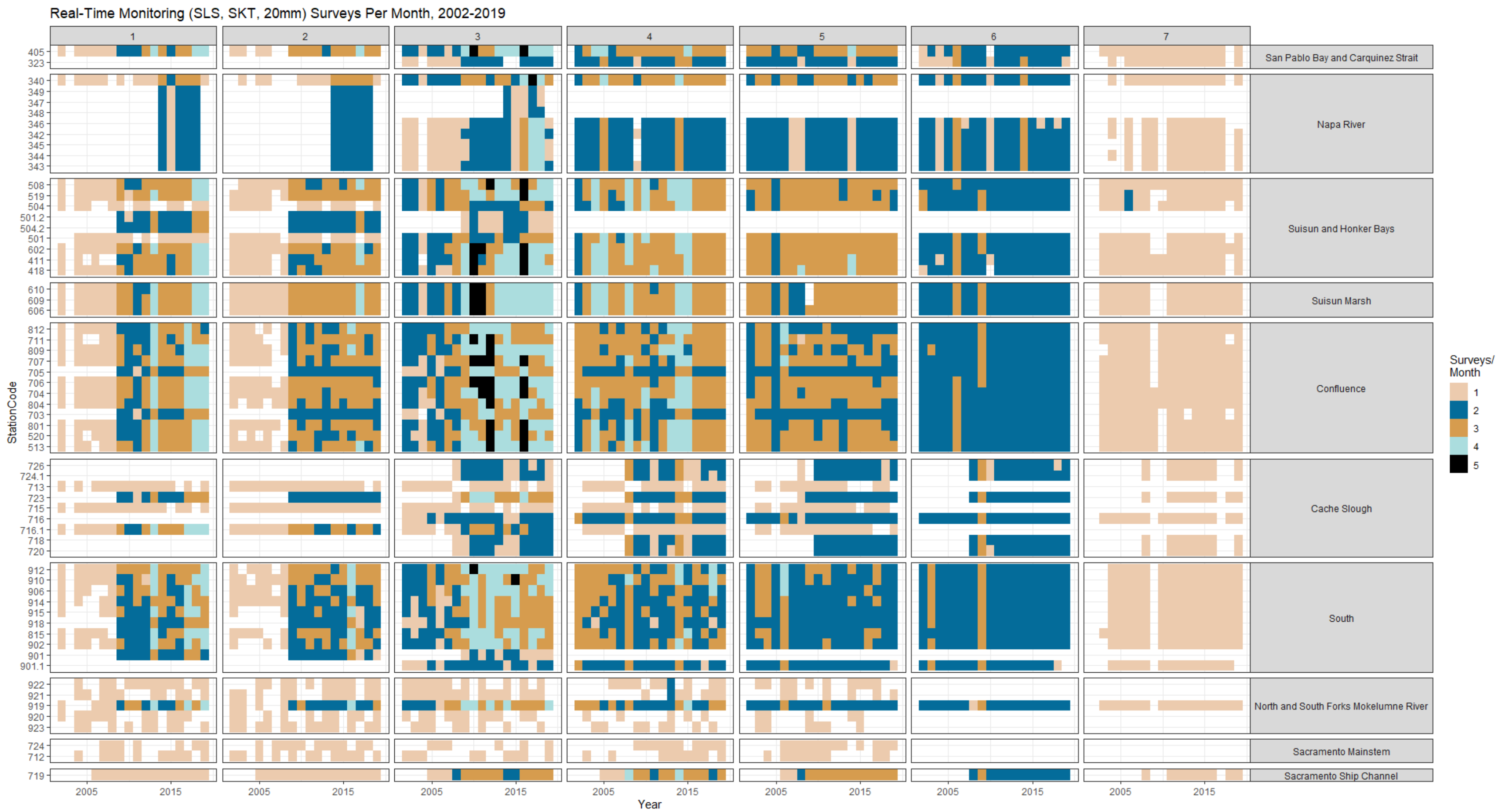


Figure 2-13. Heatmap of Spatial Sampling Effort in Real-Time Monitoring (SLS, 20mm, SKT)

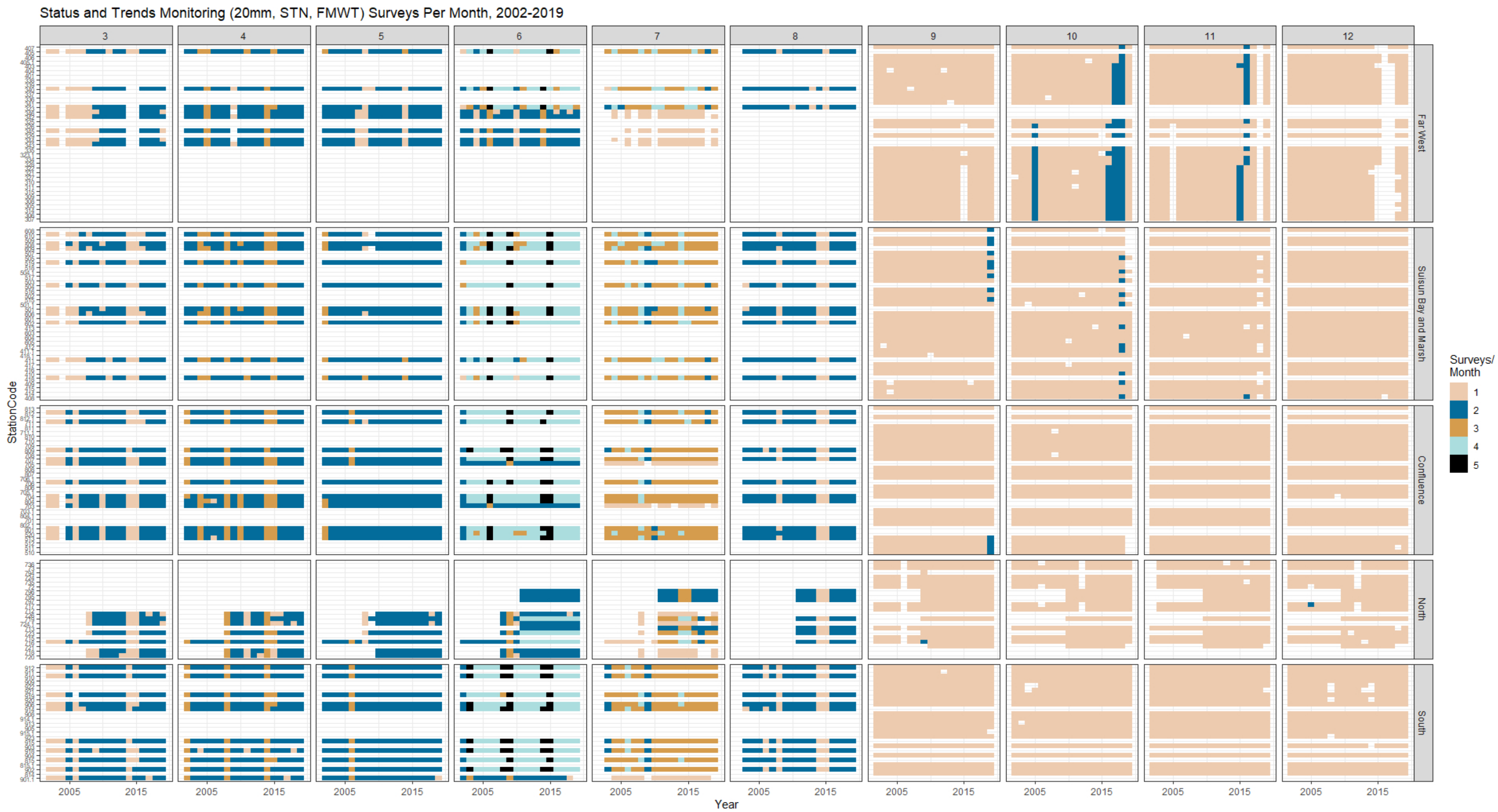


Figure 2-14. Heatmap of Spatial Sampling Effort in Status and Trend Monitoring (20mm, STN, FMWT)

Summary of Findings

The five review studies exclusively sample the pelagic habitat of the Upper SFE. Pelagic habitat represents the majority of available water volume. This observation addressed the concern that the studies only sample a subset of the available water volumes. While non-pelagic habitat does occur in each of the Review Regions, these largely represent waters that are too shallow for the pelagic gears used by the studies.

Two stratification scales were developed for use in review of the monitoring studies. The Review Region scale encompassed the dominant east-west pattern in the fish community and environmental conditions, and the Review Strata provided the means to evaluate sampling effort that emphasized specific habitats of management interest (e.g., Suisun Marsh and Cache Slough).

Real-time monitoring occurs from January to May and is well aligned with the most critical time period for entrainment. Although the priority sampling occurs in the central and south Delta, a number of other regions far from the water facilities are also sampled, particularly Cache Slough, Napa River, and Suisun Marsh.

Status and trends monitoring occurs from March through December and expands spatial sampling with higher effort in the Fall/Winter, including more expansive monitoring in San Pablo Bay and upstream into the Deep Water Ship Channel. Evaluation of spatial and temporal effort at the status and trends studies identified that the three studies have imbalanced sampling of tows and sample volumes at the regional scale. A rebalance of effort relative to regional water volumes as a means to increase effort in undersampled regions, would likely have long-term benefit for interpreting spatial patterns in abundances among the three studies. Areas where the uncertainty in abundances may be reduced by additional monitoring is discussed in Appendix 4.

References

Mitchell, L., Polansky, L., Newman, K.B., 2017. DSM TN 2. Design-based Estimates of Delta Smelt Abundance.

Newman, K.B., 2008. Sample Design-Based Methodology for Estimating Delta Smelt Abundance. San Francisco Estuary and Watershed Science 6. https://doi.org/10.15447/sfews.2008v6iss3art3

Polansky, L., Studen, L., 2019. DSM TN 29. Water volumes of the San Francisco Estuary in Areas Occupied by Delta Smelt. DRAFT.

San Francisco Estuary Institute, 2017. California Aquatic Resource Inventory (CARI) version 0.3 [WWW Document]. URL (accessed 4.20.21).

Appendix 3A. Taxonomic Comparison Among Studies

Introduction
This task evaluated species detection patterns amongst the pelagic gear types to inform understanding about the species common to the review studies. In this task, Design Team sought to address concerns the review studies only sample a subset of available species of the fish community. Secondly, to inform the relative catchability and selectivity of different gears for the commonly occurring species, the catch success of each of the sampling gears was analyzed by Dr. Josh Korman (Ecometric Research) using data from a prior gear comparison study.

Selected Species
When analyzing long-term catch data and using those analyses to inform changes to sampling design, it is beneficial to focus in on a select number species to keep the effort manageable. While a goal of this effort is to design monitoring that provides useful metrics at the fish community level, special status species are still one of the primary monitoring targets, particularly when it comes to producing abundance estimates and providing high resolution spatial distribution data. Within this context, the Design Team initially agreed to take a community perspective to the species evaluation by including species representative of the Upper SFE. Representative species were classified into three categories:

<u>Listed species</u>: Fish species listed as threatened or endangered under either the Federal Endangered Species Act or the California Endangered Species Act. These species are of the greatest management interest and are often the explicit targets for flow actions and habitat restoration. Examples include Delta Smelt, Longfin Smelt, winter- and spring-run Chinook Salmon.

<u>Community species</u>: Species that are representative of a given fish community. Commonly detected species that are known to be of critical ecosystem importance due to their prevalence (e.g., high biomass) in the system. Examples include Striped Bass, Threadfin Shad, Northern Anchovy, and American Shad.

<u>Rare and uncommon species</u>: Species that are of particular interest but do not appear regularly in the SFE, or at least not in current monitoring datasets, to allow for an analytical assessment of their capture probability. This category can also cover early detections of new invaders to the SFE. Examples include Pacific Lamprey, White Sturgeon, and Green Sturgeon.

A total of 15 species (Table 1) were identified for the focus of the analysis and survey design evaluations based upon review of a catch frequency table across surveys. Catch data was limited to 2002-2019 as this period was inclusive of all the surveys and deemed to be representative of a similar ecological period for the SFE, following the regime shift recognized as the Pelagic Organism Decline (POD).

Table 3A-1. Total and survey-specific frequency of the 15 most commonly encountered species in the review dataset (2002-2019). S pecies selected for use in the evaluation are shown in bold. Three species that fell outside the top 15 overall ranks were included for representativeness that each correspond to one of the three species categories (listed, community, rare).

Species Category	Common Name	Scientific Name	20mm	FMWT	SKT	SLS	STN	Total	Overall Rank
Community	**Pacific Herring**	**Clupea pallasii**	192,757	1,944	19,689	463,031	526	677,947	1
Community	**Tripletooth Gobies**	**Tridentiger Spp.**	541,910	133	0	1	35,369	577,413	2
Listed	**Longfin Smelt**	**Spirinchus thaleichthys**	212,042	2,727	3,866	81,644	3,025	303,324	3
Community	**Striped Bass**	**Morone saxatilis**	165,969	2,719	208	48	12,745	181,689	4
Community	**Threadfin Shad**	**Dorosoma petenense**	104,181	30,750	28,365	4	8,098	171,389	5
Community	**Northern Anchovy**	**Engraulis mordax**	13,442	118,500	10,900	407	4,386	147,857	6
Community	**Prickly Sculpin**	**Cottus asper**	13,442	4	10	118,747	27	132,230	7
Community	**Yellowfin Goby**	Acanthogobius flavimanus	49,964	89	27	71,215	433	121,728	8
Listed	**American Shad**	**Alosa sapidissima**	5,256	30,660	2,642	0	2,983	41,541	9
Community	Bay Shrimp	Crangon crangon	0	14,309	696	0	39,828	54,833	10
Community	**Delta Smelt**	**Hypomesus transpacificus**	8,922	1,144	7,457	480	2,257	20,260	11
Community	Inland Silverside	Menidia beryllina	2,080	165	9,025	7	1,200	12,477	12
Listed	**Chinook Salmon**	**Oncorhynchus tshawytscha**	302	122	8,567	18	9	9,018	13
Community	**White Catfish**	**Ameiurus catus**	2,932	215	16	81	3,000	6,244	14
Community	Threespine Stickleback	Gasterosteus aculeatus	3,970	5	1,902	120	364	6,361	15
Listed	**Steelhead**	**Oncorhynchus mykiss**	0	5	662	0	1	668	31
Community	**Starry Flounder**	**Platichthys stellatus**	254	30	11	0	59	354	38
Rare	**White Sturgeon**	**Acipenser transmontanus**	277	5	0	8	0	290	42

Age/Size Separation
The fish sampling gears used by the five CDFW studies are suited for catching primarily the smallest year-classes of the target species. This is by design, since one of the primary goals of the studies is to characterize recruitment and distribution of Age-0 individuals. Despite this focus, the survey gears do capture older year classes with varying degrees of effectiveness, and thus it is useful to separate the ages where possible. For some studies and fish taxa, ages were recorded in the species column (e.g., FMWT reports Age-0 Striped Bass). To allow for consistency between studies, these age designations were first removed from the review dataset and all age-classifications were determined based on breakpoints identified using a length-frequency analysis.

CDFW's Bay Study has developed a length-at-date method for separating Age-0, Age-1 and older year classes for most of the target species. However, because the five CDFW studies under review use a range of mesh sizes and therefore select for different lengths, we did not rely solely on the Bay Study length cutoffs. Instead, monthly length-frequency density figures for each species (example shown in **Figure 3A-1**) were inspected for obvious bimodality. Where multiple ages were detected, there was typically little ambiguity, with the modes well separated. The Bay Study age cutoff lengths were then overlayed in order to confirm that age separation visible in the integrated dataset occurred at comparable lengths. Overall, there was a high level of agreement between the Bay Study conventions and the integrated study data. In several cases, the Bay Study cutoff for a given species and month would intersect the tail of an age classes' length distribution, and in these cases the cutoffs were adjusted slightly.

For each fish species, individuals were classified as Age-0, Age-1 or Age-2+ based on the species- and month-specific age cutoffs. For all species, individuals with no recorded fork length were dropped from the dataset as were any age groups containing fewer than 100 individuals. This excluded all Age-2+ fish and retained Age-1 fish for only five species. In total, the fish included in the review dataset were ~5% Age-1 and ~95% Age-0. For all species except Delta Smelt and Threadfin Shad, the Age-1 group comprised less than 10% of all individuals. Threadfin Shad were ~65% Age-0 and 35% Age-1 while Delta Smelt were only ~18% Age-0 and 82% Age-1. Age-1 Delta Smelt were captured almost exclusively in the SKT; a survey specifically designed to capture adult Delta Smelt during their spawning season. When SKT samples are excluded, only 2,117 of more than ~2 million fish sampled since 2002 were Age-1; emphasizing that the pelagic gears overwhelmingly target Age-0 fish.

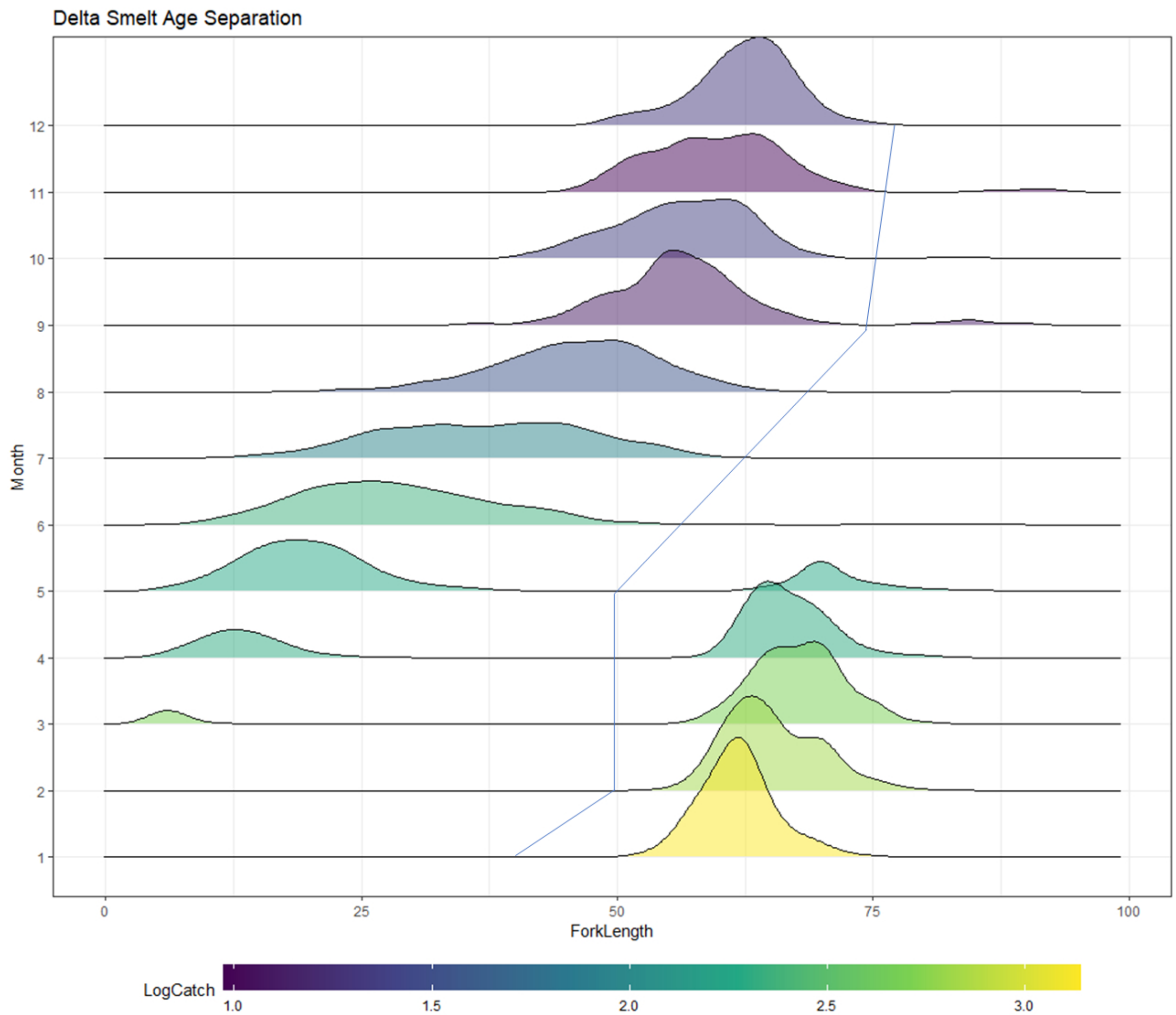


Figure 3A-1. Example of the length-frequency analysis used to distinguish between ages. Blue line shows the approximate Bay Study length cutoff values. Colored shading indicates average monthly catch of Delta Smelt in the review dataset on a logarithmic scale.

Species Detection Patterns

Species detection (i.e., proportion of tows with 1 or more individuals) was evaluated to address concerns that the studies only sample a subset of available species. Catch patterns were grouped by water year type and organized into real-time monitoring and status and trends monitoring, for individual species, the most common species, and for all species combined.

Species detections vary regionally, by study design, and among water year types. Up to 13 of the 15 review species have been detected in a single subregion for both real-time monitoring and status and trends monitoring (**Figure 3A-2, 3A-3**). Further examination of the species detections for the most commonly detected species (> 10% of tows) reduced the number of regionally detected species to a maximum of 8 (**Figure 3A-4, 3A-5**), which undoubtedly represents a subset of the available species in the pelagic community.

For real-time monitoring, the largest number of species has been detected in the 20mm survey and the least in the SLS, irrespective of water year type. Across these surveys, fewer species are detected in downstream areas during dry and critical water years. Overall, the strata associated with the most species in real-time monitoring have been Suisun Marsh, Suisun / Honker Bays, and the Confluence.

Similar evaluations of the status and trends studies, revealed that the 20mm study also has the highest number of species detections compared to STN and FMWT. STN consistently detects about half the number of common species as the 20mm study. Regional species diversity In the STN study is also affected by hydrology, with more species being detected at downstream locations during wet years. This pattern is also seen somewhat for the FMWT study when considering the commonly detected species; likely a result of the relatively large amount of effort of the FMWT in the western portion of the Delta. Several areas sampled in status and trends studies detect very few species in >10% tows, particularly in strata upstream of the Confluence. Overall, the strata associated with the most species in status and trends monitoring have been Suisun Marsh and Suisun / Honker Bays.

Pacific Herring, Chinook Salmon, and Longfin Smelt are three species commonly caught in real-time monitoring. In each of the studies, Pacific Herring (**Figure 3A-6**) is caught in San Pablo Bay upstream to the Confluence, and are most frequently detected during Critical and Dry Years when suitable salinity conditions extend further eastward. The full distribution of this species presumably extends far to the west of the current sample frame, and so reflects the fact that a larger proportion of the population is observed during low-outflow years. Overall, SLS has higher detections for Pacific Herring than 20mm or SKT. Chinook Salmon (**Figure 3A-7**) are only frequently detected in SKT sampling with higher detections in wet years than other water year types. Thirdly, Longfin Smelt (**Figure 3A-8**) are detected in all three real-time studies, with more frequent detections in SLS and 20mm surveys than for SKT. Similar to the spatial pattern for Pacific Herring, the detections of Longfin Smelt appear higher during the drier water years, suggesting that fish distributions may shift out of the sample frame when salinity and water temperature are higher. These results suggest that additional sampling in San Pablo Bay or beyond should be considered if more accurate representation of spatial distributions in Pacific Herring, Longfin Smelt, and fish that occur in higher salinities is desired.

American Shad, Striped Bass, and Northern Anchovy are three species commonly caught in status and trends monitoring. American Shad (**Figure 3A-9**) detection patterns show higher detections in wet years, particularly for FMWT that sample upstream into Cache Slough and the Deepwater Ship Channel. Striped Bass (**Figure 3A-10**) show similarly higher detections upstream in wet water years across the review studies. Suisun Marsh and Confluence are usually associated with the highest detections. Finally, detections of Northern Anchovy (**Figure 3A-11**) appear to exhibit relatively high detections in higher salinities throughout the year. All three studies indicate higher detections in critical and dry years in Suisun Bay and Marsh. However, during wet years the FMWT has shown higher detections in San Pablo Bay than other regions. This spatial pattern was again, considered supportive of sampling more extensively in San Pablo Bay to understand species distributions under critical and dry water years.

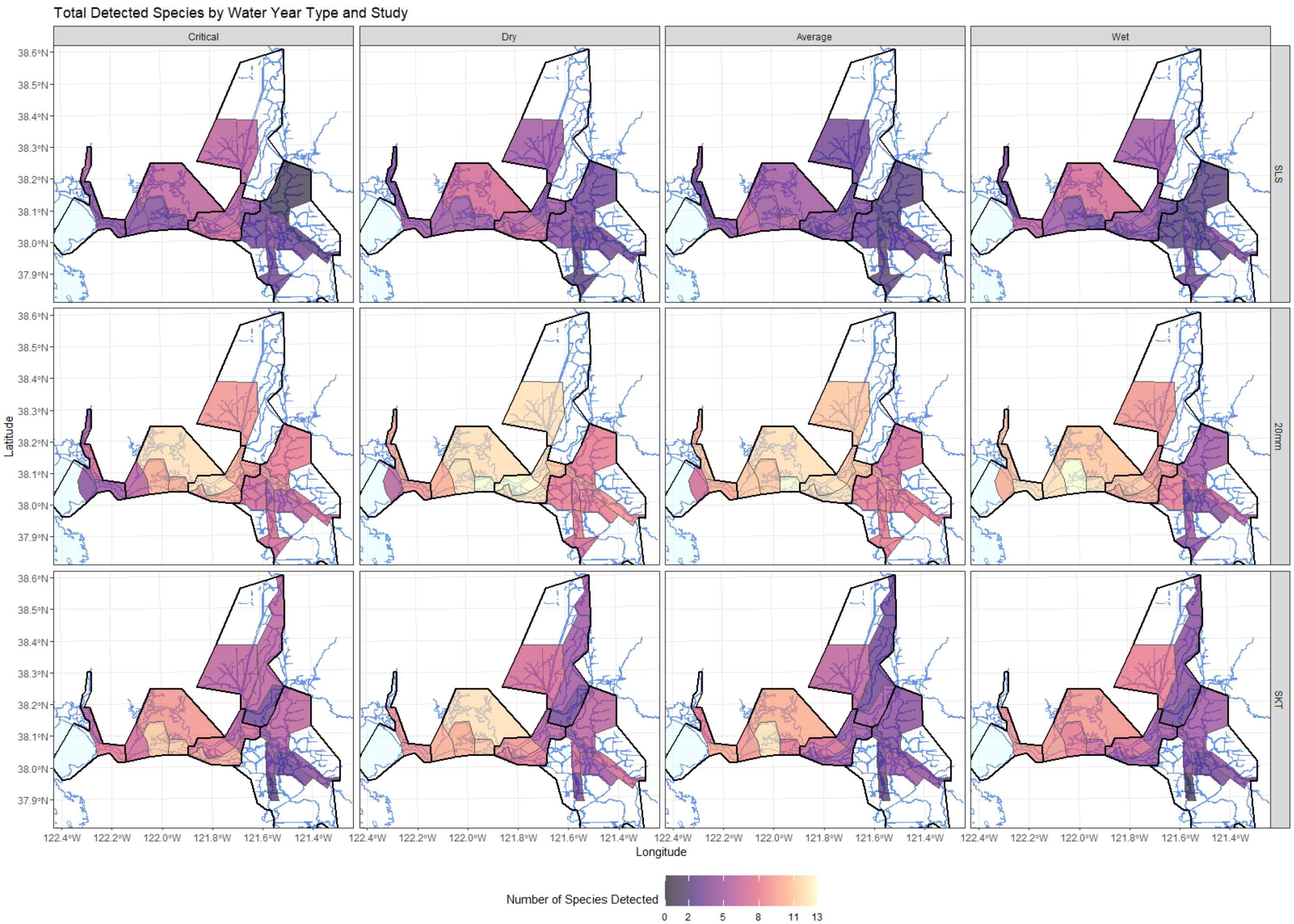


Figure 3A-2. Spatial patterns of Species Detection by Water Year Type for Real-Time Monitoring.

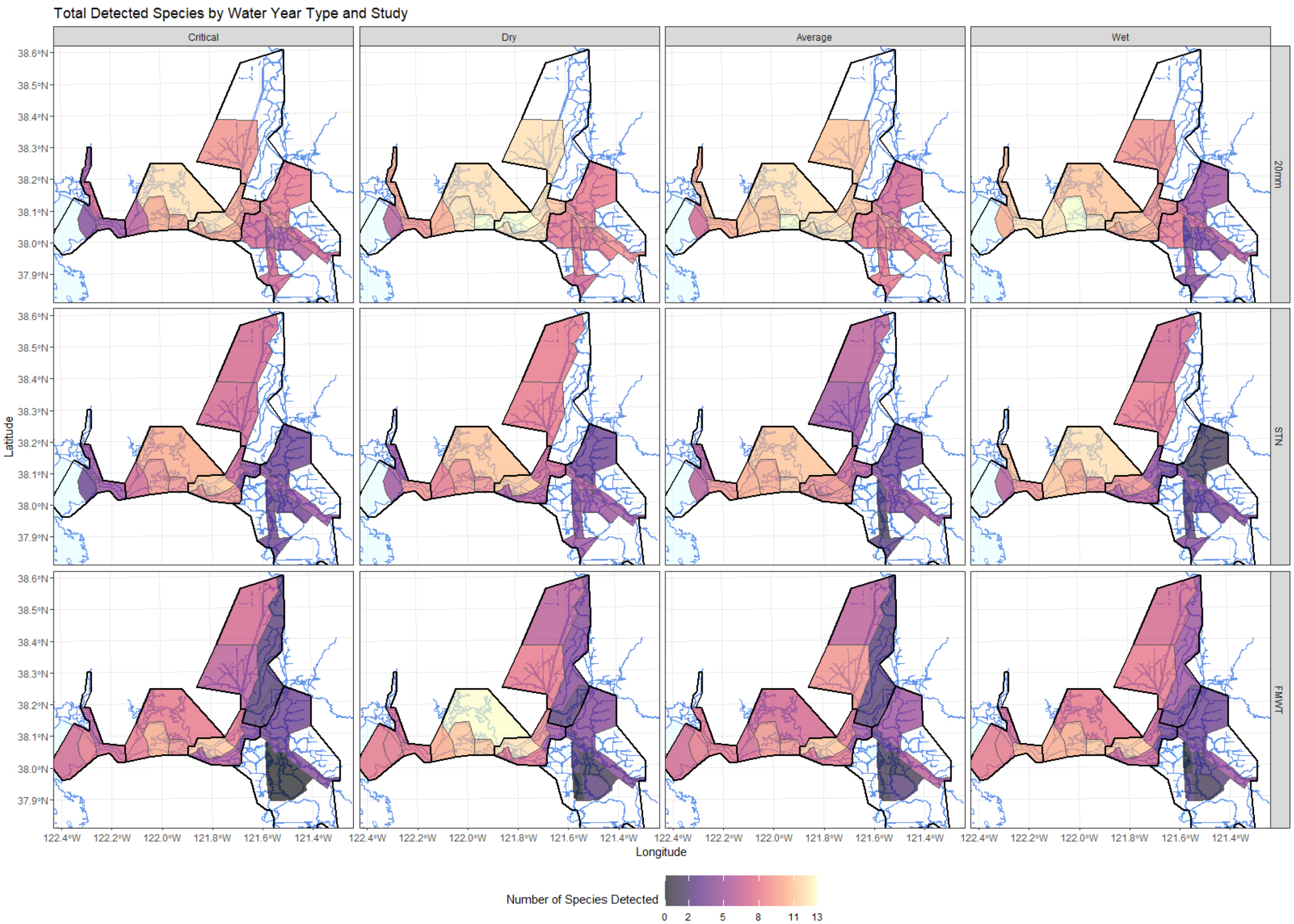


Figure 3A-3. Spatial patterns of Species Detection by Water Year Type for Status and Trends Monitoring.

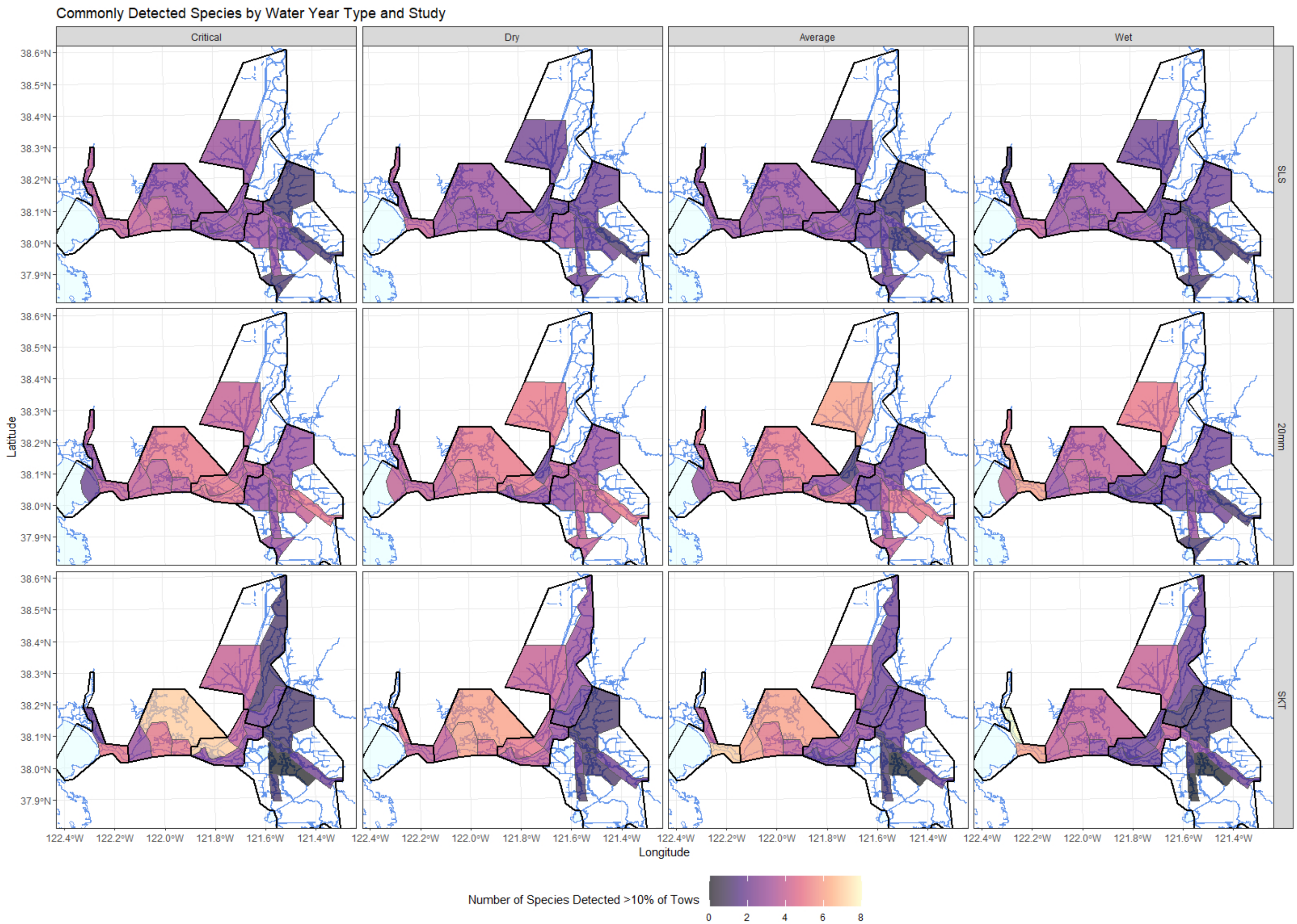


Figure 3A-4. Spatial patterns of Commonly Detected Species by Water Year Type for Real-Time Monitoring.

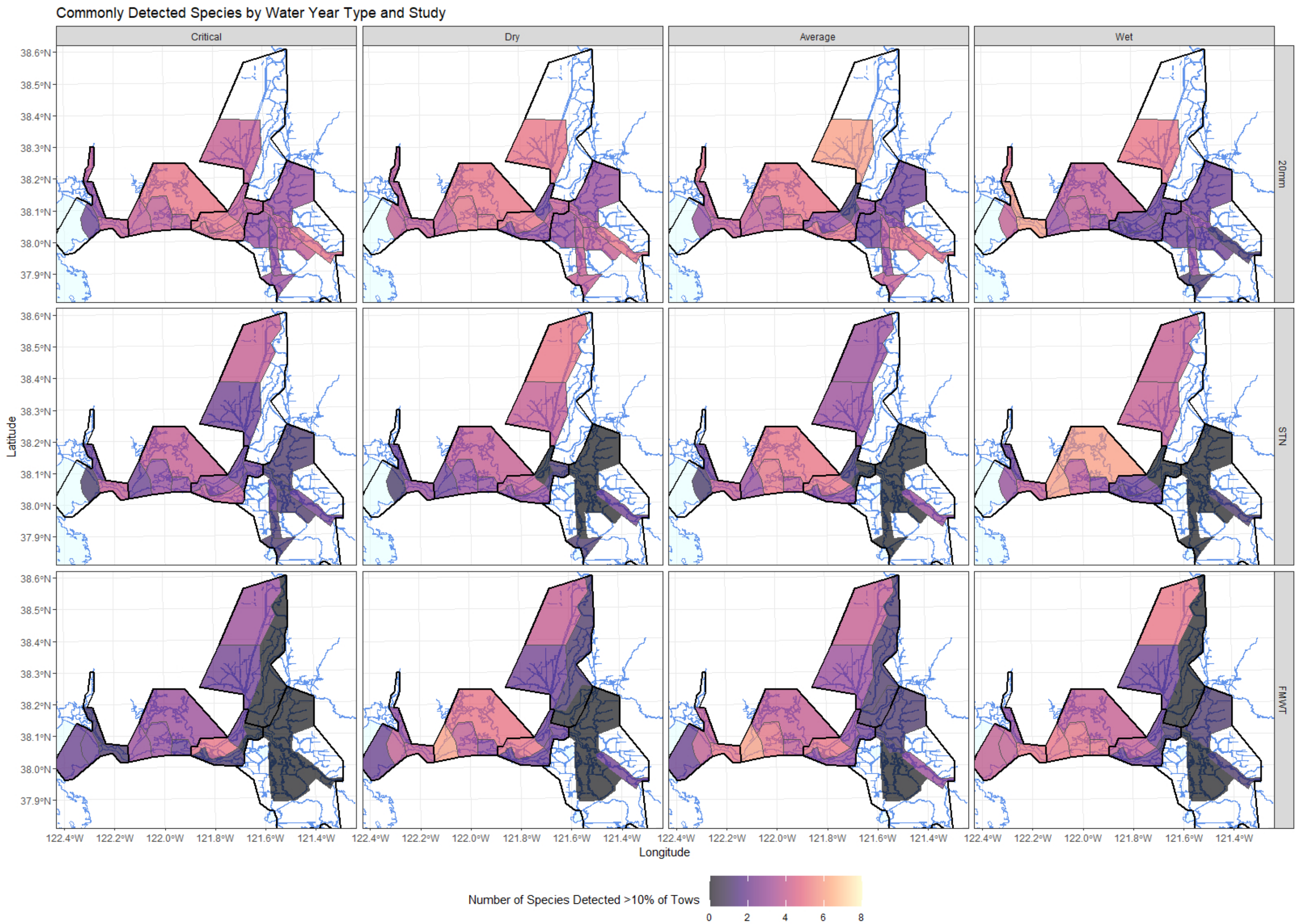


Figure 3A-5. Spatial patterns of Commonly Detected Species by Water Year Type for Status and Trends Monitoring

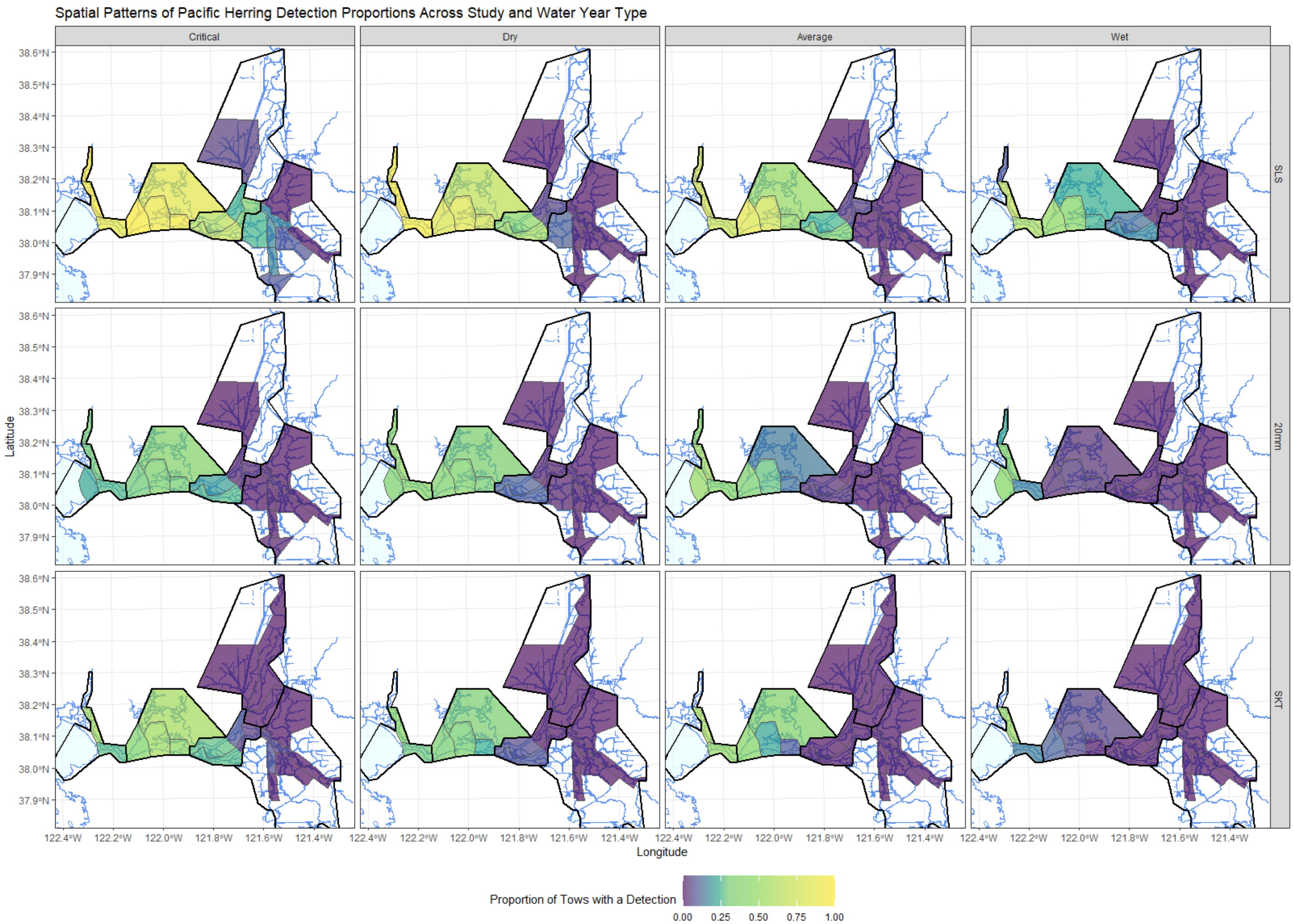


Figure 3A-6. Spatial Patterns of Pacific Herring Detection by Water Year Type for Real-Time Monitoring

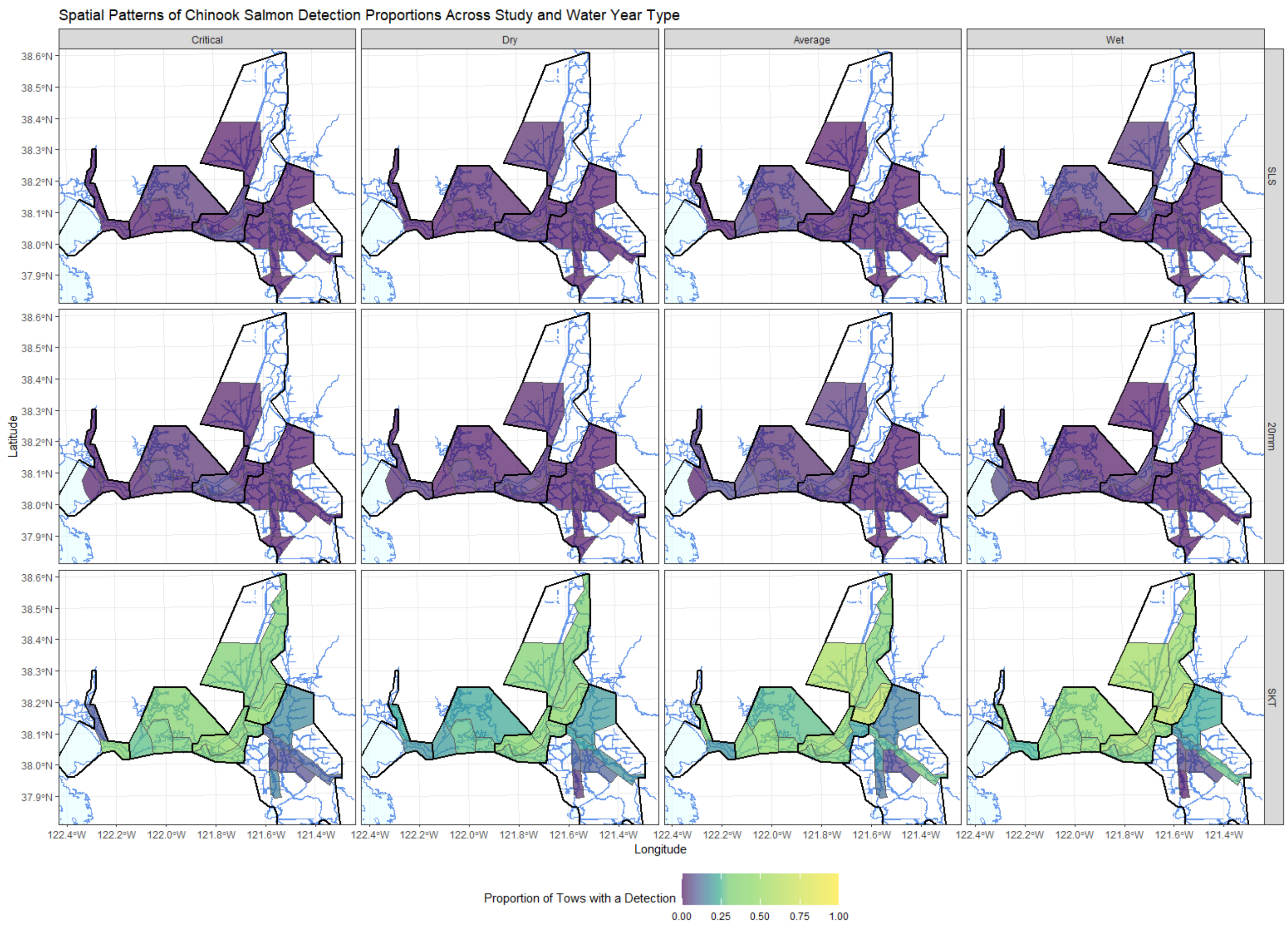


Figure 3A-7. Spatial Patterns of Chinook Salmon Detection by Water Year Type for Real-Time Monitoring

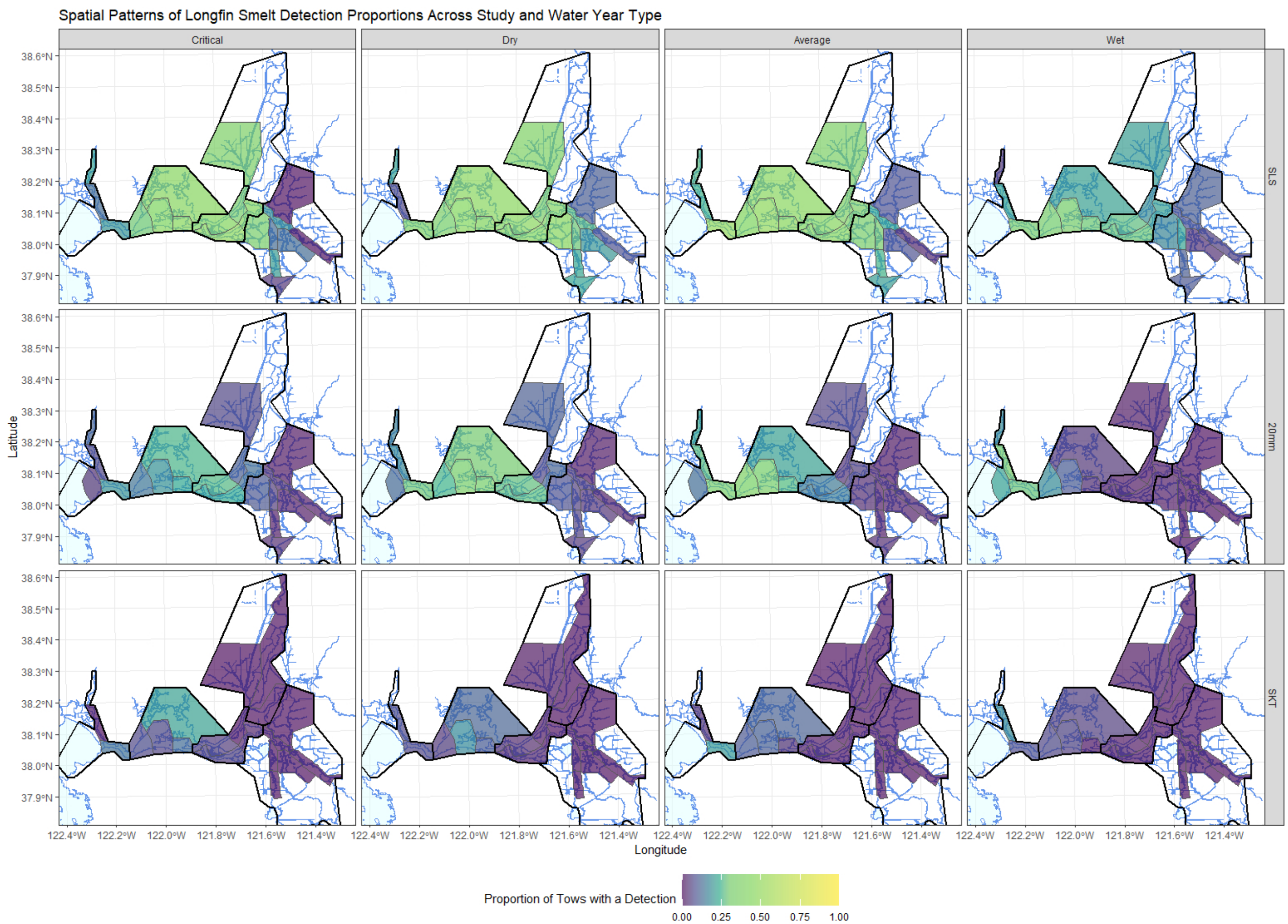


Figure 3A-8. Spatial Patterns of Longfin Smelt Detection by Water Year Type for Real-Time Monitoring

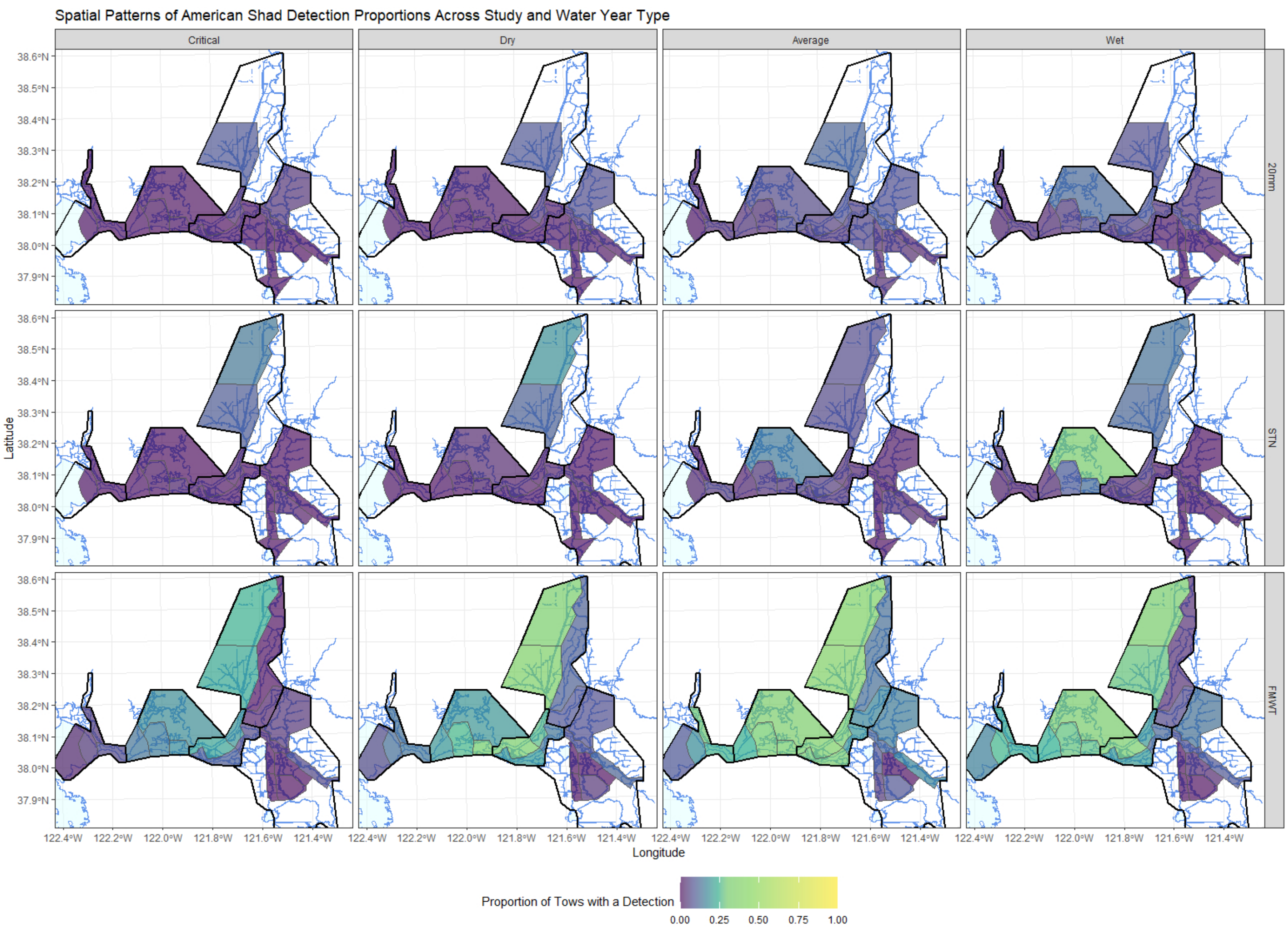


Figure 3A-9. Spatial Patterns of American Shad Detection by Water Year Type for Status and Trends Monitoring

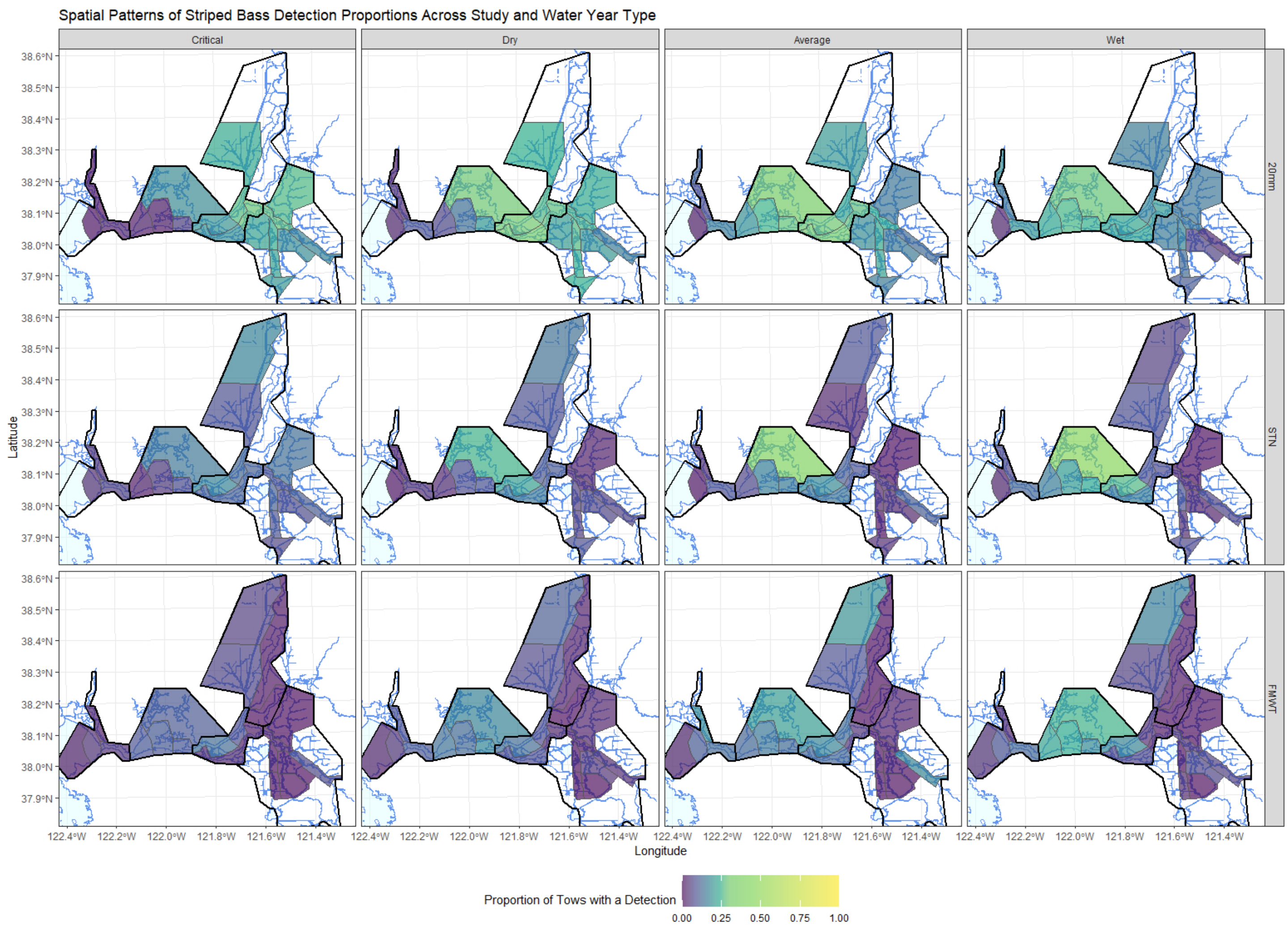


Figure 3A-10. Spatial Patterns of Striped Bass Detection by Water Year Type for Status and Trends Monitoring

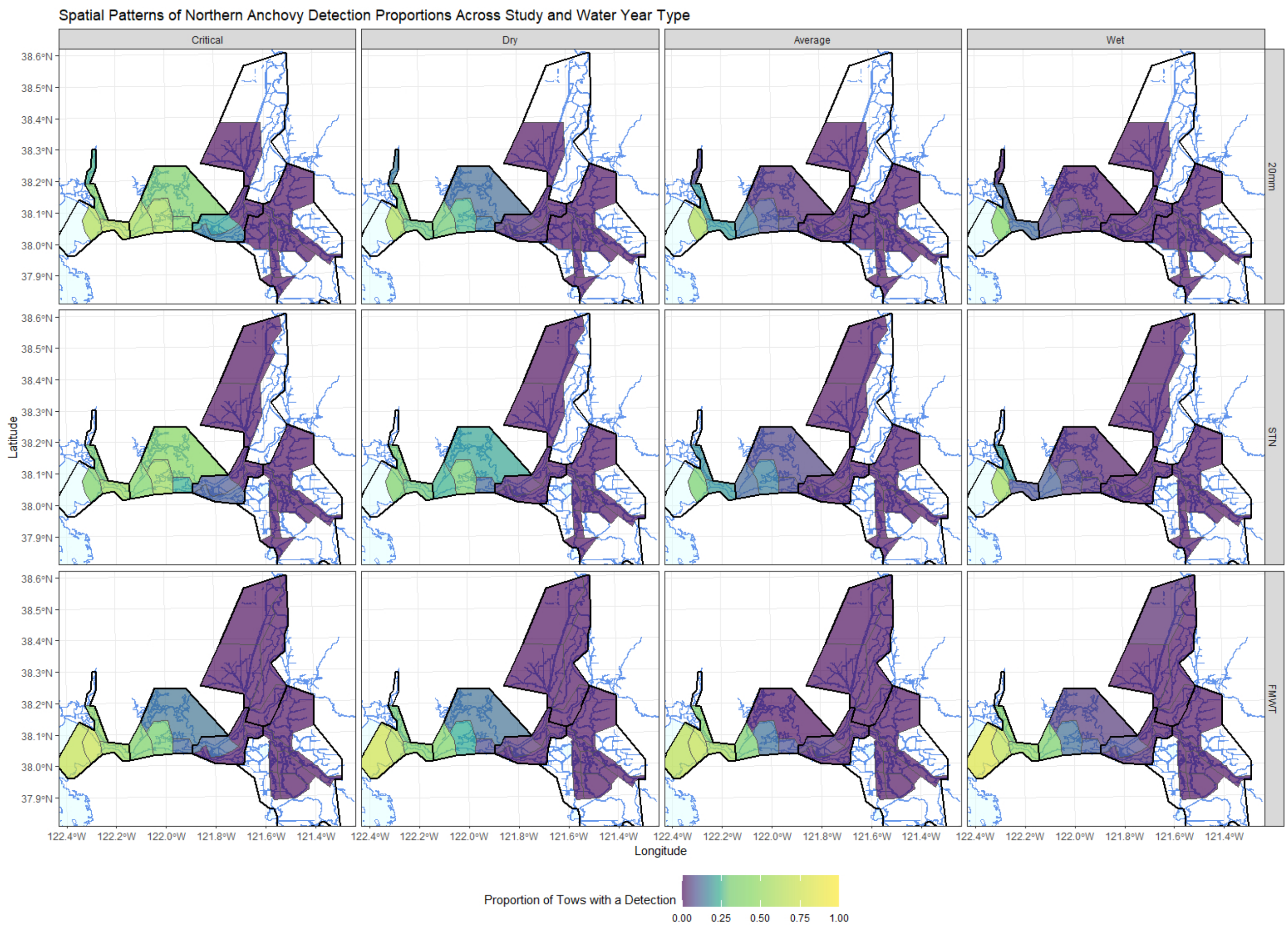


Figure 3A-11. Spatial Patterns of Northern Anchovy Detection by Water Year Type for Status and Trends Monitoring

Identifying Effort Levels Necessary to Achieve High Probability of Species Detection
Simulations of altered sampling designs were conducted to identify the effort levels necessary to achieve high probability of detections in real-time monitoring. Identifying species presence, even at very low levels of abundance is an important goal of real-time decision making. The influence of reduced effort on detection probability was assessed as a component of sequential station exclusion simulations (see Appendix 4 for details). Briefly, this approach involves removing a subset of stations from the historical data, and after each removal calculating the proportion of non-zero tows in each month, region/stratum and year. The reduced-effort scenarios were then compared with full effort through calculation of correlation coefficients at the regional scale. Effort additions were considered using a different approach in order to take advantage of the binomial nature of the data.

By converting catch data to binary, presence/absence format, it is possible to consider tows as binomial random draws. Assuming that each tow in a region, month and study and water year type has some consistent chance of detecting a target species, it is then, using the binomial cumulative distribution function, possible to determine the number of tows required to detect that species with high probability. This calculation was implemented in R using the 'pbinom' function. In addition to the number of samples (i.e., tows) to draw and the target number of detection (in this case 1 or more), the function requires a probability of detection in a single tow. In order to estimate these probabilities, the binary catch table was grouped by species, month, region/stratum and water year type and the proportion of tows with a positive detection was calculated. These proportions of course ignore issues of catchability (i.e., fish may be present that a given gear does not capture) and so the proportion of tows in which a target species is captured does not provide a complete representation of detection probability. Nevertheless, these proportions span nearly the entire range of success possibilities (0-1), and provide important general guidance on the amount of effort required to reliably detect rare and common species.

By utilizing these calculated detection proportions as the probability argument in the 'pbinom' function, the probability of one or more positive detections could be calculated for each species, month, region and water year type grouping (**Figure 3A-12**). The Design Team selected a 95% probability of detecting a species, given that it is present, as a goal for the purposes of this evaluation. In instances where this probability of detection was not achieved with current sampling effort, the number of additional tows required to exceed the threshold was calculated. This involved recalculating the detection probabilities over a range of sample sizes, searching for the number of tows required to exceed the 95% threshold and calculating the difference between the required and current number of tows.

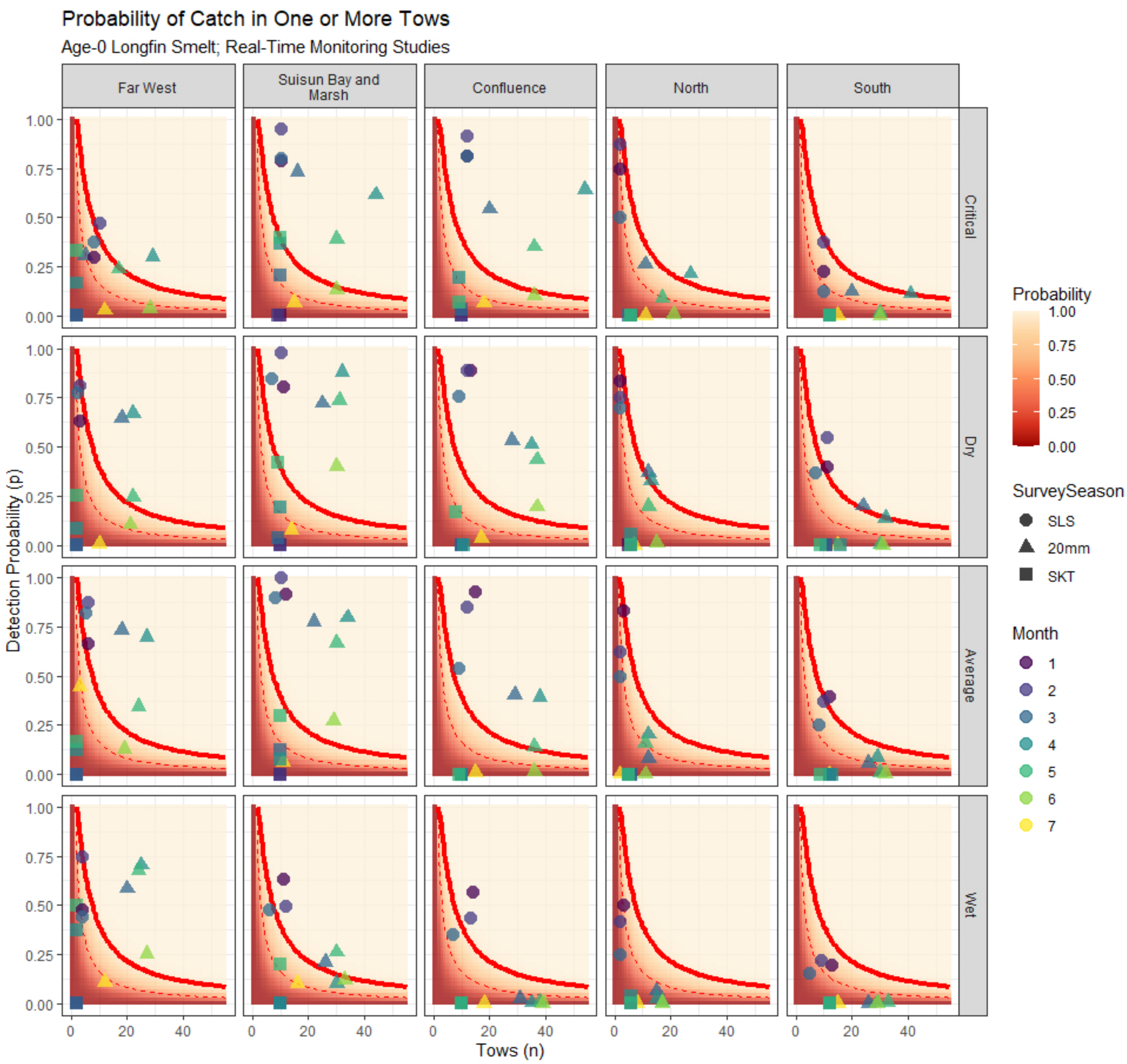


Figure 3A-12. Example of the detection analysis results showing the proportion of tows detecting longfin smelt given current effort levels by study and month (points). Shading shows the probability of detecting one or more individual for all combinations of effort (n) and probability of capture in a single tow (p). Dashed and solid red line show the 50% and 95% probability contours, respectively. Points falling above the solid red line indicate that current effort is sufficient to detect longfin smelt with high probability, given historical patterns of catch.

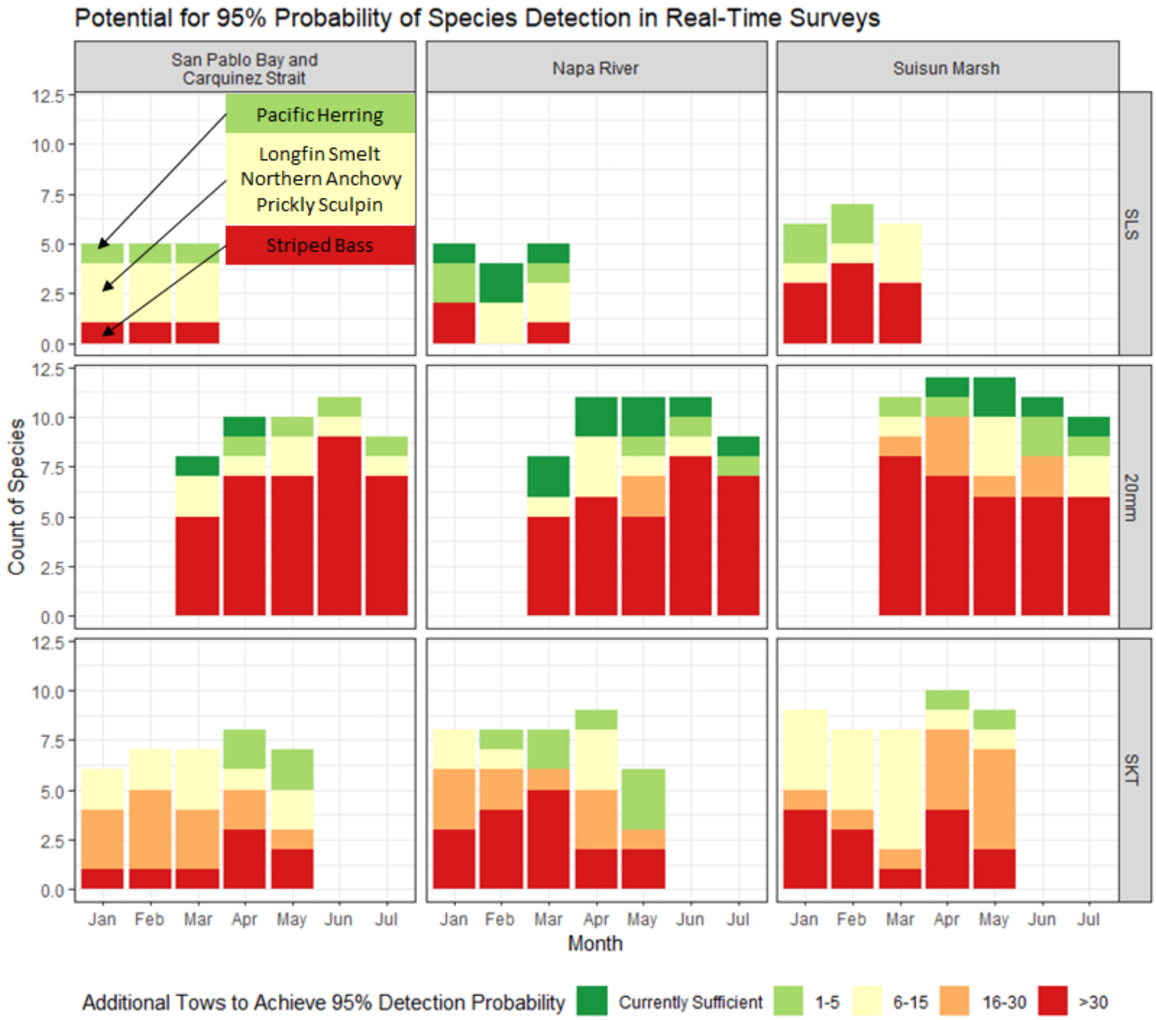


Figure 3A-13. Example of the detection analysis results showing the increases in effort (number of tows) required to achieve a 95% probability of detecting one or more individuals in real-time monitoring surveys. The top left panel provides annotation to highlight that each column represents multiple species, but note that the analyzed species vary between studies.

In all three real-time studies (SLS, 20mm, SKT), each of the regions had species for which conducting 1-15 additional tows was predicted to result in >95% detection. However, given current detection frequencies, there were also species that would not be able to meet this threshold with 30 or more additional samples per survey. Indeed, the amount of effort required to achieve a 95% probability of detection increases dramatically when the target species is captured in less than ~10% of tows (See solid red line in **Figure 3A-12**). Increasing effort would likely benefit species detections for the fewest Age-0 species in SLS and the greatest for 20mm. San Pablo Bay and Suisun Marsh were indicated in all three of the study design evaluations as being a high priority for additional effort to increase species detections of several species, most notably Longfin Smelt, Pacific Herring, Striped Bass, and American Shad (**Figure 3A-13**). Additional monitoring effort in Napa River was only indicated for the SKT evaluations, which was largely driven by increasing detections for Pacific Herring and Chinook Salmon.

Species detection patterns indicate that all of the review studies sample a subset of the pelagic community. Only eight species are commonly occurring (>10% of tows) in the dataset from 2002-2019; Pacific Herring, Northern Anchovy, Longfin Smelt, Delta Smelt, American Shad, Threadfin Shad, Prickly Sculpin, and *Tridentiger* spp. Furthermore, during Critical and Dry water years the eight common species exhibit higher probability of detection in Suisun Marsh and San Pablo Bay, while during wetter years, the species detections are higher upstream.

Summary of Findings

Age-0 fish are almost exclusively represented in the catch data from the five review studies. For all species except Delta Smelt and Threadfin Shad, the Age-1 group comprised less than 10% of all individuals. Age-1 Delta Smelt were captured almost exclusively in the SKT; a survey specifically designed to capture adult Delta Smelt during their spawning season. When SKT samples are excluded, less than 1% of fish sampled since 2002 have been Age-1; emphasizing that the pelagic gears overwhelmingly target Age-0 fish.

Species detection patterns indicate that all of the review studies sample a subset of the pelagic community. On a regional basis, only eight species are commonly detected (>10% of tows) in the dataset from 2002-2019. Furthermore, during Critical and Dry water years fish exhibit higher probability of detection in Suisun Marsh and San Pablo Bay, while during wetter years, species detections are higher upstream.

Simulations to identify the level of effort that would be required to increase species detections in real time monitoring identified the best opportunities for this reside in areas far from the SWP and CVP, notably in San Pablo Bay, Suisun Marsh, and Napa River. This pattern is the result of these areas having more frequent detections, and thus the potential for improving probabilities for the most species.

Appendix 3B. A Comparison of Relative Catchability and Size Selectivity of Different Gears Sampling Fish in the San Francisco Bay Delta

Dr. J. Korman (Ecometric Research)

Introduction

A range of tow net and trawl surveys in the San Francisco Bay Delta provide indices of index spatial and temporal trends in abundance for a wide variety of fish species (e.g., Sommer et al. 2007, MacNally et al. 2010, Thompson et al. 2010, Latour 2016). In some cases, these data are also used to estimate abundance, and even life-stage specific abundance (Maunder and Deriso 2011, Polansky et al. 2019). Inferences on abundance and trends from these surveys rely on the general catch equation,

$$\boldsymbol{C = q \cdot E \cdot N}$$

where C is the catch of a particular species, N is the unknown abundance, E is sampling effort (usually measured as a tow volume), and q is the catchability coefficient, which represents the proportion of the population (N) caught per unit of effort. If q is known or assumed, catch can be used to calculate abundance by re-arranging eqn. 1 to solve for N. There are many analyses of survey data in the Delta where q is calculated as the proportion of the volume sampled relative to the assumed volume that a fish species uses (at Delta or a smaller regional scale). This approach depends on unlikely and at best highly uncertain assumptions. First, this approach assumes that fish are randomly distributed over the available habitat, which is not expected given highly clustered spatial distributions seen for many animal populations (Conroy et al. 2008, Korman et al. 2016, Wyatt 2003). Second, there is considerable uncertainty about the habitat volume that a species is distributed over. In smaller systems, mark-recapture methods can be used to estimate q and avoid the need for such uncertain and tenuous assumptions. However, mark-recapture is nearly impossible to implement effectively in the large area of the Delta, and due to the fragile nature and difficulty in capturing many of the species of interest. Due to these challenges, some researchers in the Delta use the catch equation to only index trends in abundance rather than to estimate the actual abundance. But even here, this approach assumes q is constant over space and time, or that such changes can be predicted as a function of environmental covariates like secchi depth.

Clearly, the catchability term in the basic catch equation (eqn. 1) has important implications for interpreting trends in the data from surveys conducted in the Delta. There are two important elements or sub-components of catchability. At the broadest scale, q reflects the ability of the gear to capture any individual of a species, and gears with higher catchability will be more effective at indexing or estimating abundance than gears with lower catchability. A comparison of relative catchability among gears can therefore be informative for making decisions on which gear to use to index abundance for a particular species. Within a species, fish of different sizes may be less or more vulnerable to capture, an effect often referred to as size selectivity. Size selectivity is also an important consideration for making decisions on sampling gears, especially for life cycle models where abundance estimates are needed for multiple life stages that vary in size.

The Design Team is attempting to optimize data collection to better track long-term status and trends of fish populations in the Delta. An important element of this project is defining the utility of each gear type. This appendix uses data from paired gear comparisons in the Delta (Mitchell et al. 2019) to estimate relative differences in catchability and size selectivity for the 20-mm survey (20-mm), the Chipps Island midwater trawl survey (CMWT), the Fall Midwater Trawl survey (FMWT), the Spring Kodiak Trawl survey (SKT), the Smelt Larval Survey (SLS), and the Summer Townet survey (STN). These comparisons are done for a variety of species that were prioritized in the early stages of the monitoring design project.

Methods

Data

Mitchell et al. (2019) conducted side-by-side comparisons of different tow net and trawl gears used in the Delta between 2012 and 2015. Analyses to date have focused on results for Delta Smelt, though the catch and size of other species were recorded. By sampling with multiple gears at the same times and locations, it is reasonable to assume that the abundance and size structure of a particular species would be the same for all gears deployed at the time and place. Thus, differences in the abundance and size structure among gears would reflect differences in catchability and size selectivity. Mitchell et al. compared SLS, 20-mm, and STN trawls at times of year when Delta Smelt were smaller, and CMWT, FMWT and SKT trawls when fish were larger (fall-winter). Sampling was conducted at three locations which included the Lower San Joaquin River near Jersey Point (SanJoaq), the Lower Sacramento River near Decker Island (LowSac), and the Sacramento Deep Water Ship Channel (SDWSC). Sampling was conducted in fall (CMWT, FMWT, SKT, and STN), spring (20-mm, SLS, and STN), and summer (20-mm, SKT, STN). For additional details see Mitchell et al. (2019).

The Design Team identified 15 species or groups to consider for analysis: Pacific Herring; Tripletooth Gobies; Longfin Smelt; Threadfin Shad; Northern Anchovy; Striped Bass; Prickly Sculpin; Delta Smelt; Yellowfin Goby; American Shad; Chinook Salmon; White Catfish; Steelhead Trout; Starry Flounder; and White Sturgeon. There were no catch records in the Mitchell et al. (2019) paired trawl database for Steelhead Trout and White Sturgeon, and very low total catches (summed across gear types) for Northern Anchovy (29), Yellowfin Goby (10), Pacific Herring (4), Starry Founder (2), and Chinook Salmon (1). Thus, our analysis focused on the remaining seven species or groups, which were American Shad; Delta Smelt; Tripletooth Gobies; Longfin Smelt; Prickly Sculpin; Striped Bass age-0; and Threadfin Shad (**Table 3B-1**).

As mentioned above, not all gear types were compared within each sampling season. Instead, gears were usually only deployed in seasons when they were typically used. Samples sizes for some seasons and species or groups had very low total catches. In our seven species subset of the paired trawl data, we only analyzed location and season strata for species where the total catch per season across gear types was more than 10 fish. This resulted in excluding Longfin Smelt in fall and summer, and Striped Bass age-0 in summer comparisons (**Table 3B-2**).

The abundance of a particular species can potentially vary substantially across locations within a sampling season. As a result, gear comparisons within seasons should only be made over the same sampling locations. If this is not done, differences in catch among gears could reflect differences in abundance across locations rather than differences in catchability among gears. Due to low sample sizes, data from multiple locations were pooled if all gear types were used at the locations in the same survey event. During the fall sampling, four gears (FMWT, CMWT, SKT, STN) were deployed at both the LowSac and San Joaq locations (**Table 3B-3**), where all four gear types were sampled at both locations. Similarly, spring sampling was conducted using three gear types (20mm, SLS, STN) at SanJoaq and SDWSC locations, where all three gear types were sampled at both locations. Summer sampling was conducted in LowSac and SDWSC locations. However, 20-mm, SKT and STN gears were all sampled at LowSac, while only 20-mm and STN was sampled at SDWSC and SanJoaq. In this summer case, the comparison of gears was stratified to account for differences in the gears that were deployed at the three locations. Catchability was compared across the three gears (20-mm, SKT, STN) deployed during summer at LowSac, and for the two gears (20-mm and STN) used at both the SDWSC and SanJoaq locations. Catchability analyses focused on differences between gears, seasons, and species.

<u>Relative Catchability Model</u>

We estimated relative differences in catchability among gear types using a Bayesian model. Using FMWT and SKT gears to illustrate the logic of the model, we begin by defining gear-specific catch equations,

$$\boldsymbol{C_{FMWT} = q_{FMWT} \cdot E_{FMWT} \cdot N, \qquad C_{SKT} = q_{SKT} \cdot E_{SKT} \cdot N}$$

These equations can be re-arranged to solve for abundance (N), and because abundance is assumed to be equal at the location and time where the paired trawls are conducted,

$$\boldsymbol{\frac{C_{FMWT}}{q_{FMWT} \cdot E_{FMWT}} = \frac{C_{SKT}}{q_{SKT} \cdot E_{SKT}}}$$

Relative catchability for one gear type compared to another can then be calculated by re-arranging eqn. 3 to solve for the ratio q's. In this example, the catchability of the SKT survey relative to the FMWT survey (qRelSKT) would be calculated as,

$$\boldsymbol{qRel_{SKT:FMWT} = \frac{q_{SKT}}{q_{FMWT}} = \frac{C_{SKT} \cdot E_{FMWT}}{C_{FMWT} \cdot E_{SKT}}}$$

Note this relatively catchability estimate accounts for differences in both catch and effort among gear types during the paired surveys. For example, if the catch from each of these two gear types was equal, but the effort (tow volume) from the FMWT sampling was double that of the SKT survey, qRelSKT:FMWT would be 2 because the SKT gear caught the same number of fish with half the effort.

We estimated relative catchabilities using a Bayesian approach implemented in the BUGS language so that uncertainty could be quantified (**Table 3B-4**). The model assumes that catch of a species across paired gears are multinomially-distributed random variables that depend on the total catch across gears and the expected proportion of the catch in each gear (P[ig]'s in Table 4). The model estimates the log of catchabilities of each gear type (alpha[ig]'s in Table 4) using uninformative normal prior probabilities (~dnorm()). These estimates are multiplied by effort (Effort[ig]), expressed as the ratio of gear-specific volume to the volume sampled for the first gear type (where effort=1). The products of catchability and effort by gear type is used to calculate the expected proportion of catch from each gear type (P[ig]) for the multinomial likelihood calculation (~dmulti). Relative catchability (qpE[ig]) is a derived variable calculated based on the ratio of the gear-specific q's relative to the estimated q for the first gear type. Note that P's across gear types must sum to one given the multinomial assumption. As a result, the number of q's to be estimated is one less than the number of gears that are compared. As q is estimated in log-space (alpha[ig]), we set alpha[ig=1] to 0 and estimate the remaining values. This is equivalent to calculating the P for the first gear type by subtraction (e.g., P[1] = 1-sum(P[2:3])). Owing to the multinomial likelihood used in the model, paired gear comparisons for a species with large total catch across gears will yield more precise estimates of relative catchability than comparisons based on a smaller catch.

Size Selectivity Model

Size-selectivity is usually incorporated into the basic catch equation using,

$$\boldsymbol{C_{sz} = q \cdot s_{sz} \cdot E \cdot N_{sz}}$$

where C_{sz} and N_{sz} are the catch and abundance for size class sz, q, E, are defined above (eqn. 1), and s_{sz} is the relative catchability of size class sz, expressed on a 0-1 scale. Here q, the overall catchability term, represents the catchability for a fish with a size that is fully vulnerable, that is, has an s_{sz} value of 1. Mitchell et al. (2019) used a two-parameter logistic function to predict s_{sz} as a function of fork length, however this model does not allow for a dome-shaped curve, where selectivity first increases with size as they are more efficiently captured by the gear (i.e., because they are less likely to slip through the mesh of the net) and then decreases with further increases in size resulting from a better ability of larger fish to avoid the gear. Modelling dome-shaped size vulnerability can be accomplished using an exponential-logistic or other more complicated functions, but requires the estimation of an additional parameter (see eqn. 15 in Mitchell et al. 2019). This parameter can be statistically challenging to estimate when the data are sparse or when the data truly indicate logistic-shaped size selectivity, and this was the case when we applied the exponential-logistic model to the paired trawl data in our preliminary efforts. However, retracting to the more estimable 2-parameter logistic size selectivity model, as done by Mitchell et al. (2019), can lead to substantial overestimates of selectivity for larger fish, leading to a negative bias in the index of abundance relative to smaller fish.

Owing to the challenges of defining an estimable and non-biased size selectivity function, we calculated (by gear type) the proportion of catch in each 5 mm fork length class and then fit a spline model through the proportions. The benefit of this simple approach is that it allows visual examination of the shape of the size selectivity function that is not subject to biases from an assumed functional form. The spline fit can be used to interpolate size selectivity for fish of different fork lengths to meet the needs of index or abundance calculations. The limitation of the spline-fitting approach is that it does not account differences in sample size among size classes when fitting the model, and estimates of uncertainty in the size selectivity curve cannot be calculated. However, given the objectives of the monitoring design project, we felt it was more important to provide an unbiased estimate of size selectivity.

Size selectivity estimates determined from the size class proportions of total catch, whether fit by a spline, logistic, or exponential-logistic model, are sensitive to the abundance of each size class. For example, if the abundance of larger size classes is low compared to smaller size classes, the proportion of fish caught in these larger size classes will be low even if their size-selectivity's are high. Across seasons, one would expect size selectivity estimates to change with changes in the relative abundance of different size classes. However, these changes in abundance would affect all gears that are compared. As a result, differences in size selectivity among gears within a season would accurately reflect relative differences in their size selectivity, even if the size selection curves are biased because N_{sz}'s are not equal across size classes.

Results

Catchability

Estimates of relative catchability across gear types were highly variable among species and seasons. In fall (Table 3B-5a), relative catchability of CMWT for American Shad was more than 7-fold higher than for FMWT, while the relative catchability for STN was 50-fold lower than FMWT (qRelSTN:FMWT=0.02). Relative catchability of SKT for American Shad was similar to values for FMWT (~1). Trends were very different for Delta Smelt, where relative catchability was very high for SKT and very low for CMWT. Results among gears were very different for Gobies, where relative catchability was near zero for CMWT and SKT and very high for STN. SKT was ineffective at sampling Striped Bass age-0 relative to the FMWT, but SKT was effective for Threadfin Shad.

Spring comparison of relative catchability were restricted to 20-mm (reference gear), SLS, and STN gear types (**Table 3B-5b**). SLS was considerably more effective than 20-mm for all species except Longfin Smelt, where the catchability for SLS was similar to 20-mm. The STN gear had very low catchability compared to 20-mm and SLS gear.

During summer (**Table 3B-5c**), SKT was much more effective at capturing Delta Smelt compared to 20-mm or STN gears at LowSac, similar to results from the fall (LowSac+SanJoaq results). This was also the case for other species except Gobies, which were more effectively sampled by 20-mm followed by STN. Results for American Shad were uncertain due to the extreme width of credible intervals, which indicated a very high effectiveness by both STN and SKT due to the lack of American Shad in the reference gear (20mm). The summer (SDWSC+SanJoaq) comparison was limited to 20-mm and STN gears. Here 20-mm and STN catchabilities were similar for

American Shad and Delta Smelt. STN was less effective than 20-mm for Gobies, Striped Bass age-0, and Threadfin Shad.

Selectivity

In fall, the SKT showed consistent higher selectivity for smaller size classes for American Shad, Delta Smelt, and Threadfin Shad compared to FMWT and CMWT gears (Figure 3B-1). The SKT was less effective for capturing larger Threadfin Shad. Selectivity for Delta Smelt was similar for SKT and STN surveys. Selectivity for FMWT and CMWT surveys were similar for the species we assessed, except Delta Smelt. Here, selectivity of larger fish from the CMWT was considerably higher than from the FMWT.

During spring, size selectivity's were generally similar among 20-mm and SLS surveys (**Figure 3B-1**). Size selectivity for the STN survey could only be reliably assessed for a few species (Delta Smelt, Longfin Smelt) due to low catch. The STN was more selective of larger Longfin Smelt and perhaps Striped Bass age-0, but even here the sample size was low. Comparisons of size selectivity among gears in summer was limited by low sample size at the LowSac location. Gobies were the only species group with sufficient sample size for more than one gear type, and showed similar size selectivity's for 20-mm and STN surveys. Catches during summer at the SDWSC location were higher, and showed higher selectivity for larger fish in the STN survey compared to the 20-mm survey for American Shad, Delta Smelt, and Striped Bass age-0, but similar selectivity's for Gobies and Threadfin Shad.

Discussion

Results of the catchability and size selectivity modelling provides new information on the utility of various gear types for sampling a variety of fish species in the Delta. Based on relative catchability, the SKT was more effective at capturing Delta Smelt compared to the FMWT as expected, though the extent has not been previously quantified and is perhaps larger than anticipated (>80-fold). It is interesting that the SKT catchability for Delta Smelt was also greater than for 20mm in the summer comparison (>9-fold), when Delta Smelt are smaller. The SKT was also more effective than the FMWT for sampling Threadfin Shad in fall, and had an equivalent catchability for American Shad. The SKT was ineffective at sampling Striped Bass age-0 or Gobies in the fall. This observation is caveated by the fact that the SKT survey conducts surface tows, while the FMWT conducts oblique tows.

Overall, these results provide support for the use of SLS and 20mm gears during spring to summer, STN during summer/fall, and potentially expanding the sampling window for SKT to summer/fall, rather than in only winter as currently done. As STN was more effective than FMWT for select species (Delta Smelt, Gobies), deployment of the STN into September or later months also warrants further consideration. These conclusions are not influenced by patterns in size selectivity estimated in this study, which were generally similar among gear types when sample sizes were sufficient to reliably estimate proportions-by-size. The one exception here is that the CMWT gear appears more effective than the SKT at capturing larger-sized American and Threadfin Shad, and Delta Smelt. However, we do not suggest replacing the SKT gear with CMWT gear, as the SKT gear has considerably higher catchability for Delta Smelt.

Our results on size selectivity of Delta Smelt are helpful for interpreting findings from studies that assume that selectivity does not decline for larger fish (Mitchell et al. 2019, Polansky et al. 2019). Mitchell et al. used a logistic curve to model size selectivity of SKT gear, which resulted in a size selectivity of 1 for all fish with fork lengths ~> 40 mm (their Fig. 7 lower two panels). This prediction is not consistent with our results from fall, which showed a much higher proportion of fish larger than 40 mm in the CMWT gear compared to SKT gear, and a moderately higher proportion of larger fish in the FMWT gear compared to the SKT gear (and see also the Mitchell et al. summary of data in their Fig. 7 panel D). These results indicate that the SKT size selectivity function is likely dome-shaped, with selectivity's dropping for larger sized fish. If this finding is correct, it implies that the abundance of larger Delta Smelt is underestimated based on the current model, which assumes a logistic pattern in size selectivity (Polansky et al. 2019). However, the sample size for CMWT and FMWT proportions is low, which suggests that this interpretation should be viewed with some caution. That said, in support of our interpretation, the SKT caught smaller American and Threadfin Shad compared to CMWT and FMWT gears, where sample sizes for all three gears was much higher. In our view, more work on size selectivity models used to estimate abundance of Delta Smelt is warranted.

References

Conroy, M. J., J. P. Runge, R. J. Barker, M. R. Schofield, and C. J. Fonnesbeck. 2008. Efficient estimation of abundance for patchily distributed populations via two-phase, adaptive sampling. Ecology 89:3362– 3370.

Latour, RJ. 2016. Explaining patterns of pelagic fish abundance in the Sacramento-San Joaquin Delta. Estuaries and Coasts 39:233-247.

Korman, J, Schick, J, and B. Mossop. 2016. Estimating riverwide abundance of juvenile fish populations: How much sampling is enough?. North American Journal of Fisheries Management 36:213-229.

Mac Nally, R, Thomson, JR, Kimmerer, WJ, Feyrer, F, Newman, KB, Sih, A, Bennett, WA, Brown, L, Fleishman, E, Culberson, SD, Castillo, G. 2010. Analysis of pelagic species decline in the upper San Francisco Estuary using multivariate autoregressive modeling (MAR). Ecological Applications 20:1417-1430.

Maunder, MN, Deriso, RB. 2011. A state-space multistage life cycle model to evaluate population impacts on the presence of density dependence: illustrated with application to Delta Smelt (Hypomesus transpacificus). Canadian Journal of Fisheries and Aquatic Sciences 68:1285-1306.

Polansky, L, Mitchell, Newman, KB. 2019. Using multistage design-based methods to construct abundance indices and uncertainty measures for Delta Smelt. Transactions of the American Fisheries Society 148: 710-724.

Sommer, T, Armor, C, Baxter, R, Breuer, R, Brown, L, Chotkowski, M, Culberson, S, Feyrer, F, Gingras, M, Herbold, B, Kimmerer, W, Mueller-Solger, A, Nobriga, M, Souza, K. 2007. The collapse of pelagic fishes in the upper San Francisco Estuary. Fisheries 32: 270-277.

Thomson, JR, Kimmerer, WJ, Brown, LR, Newman, KB, Mac Nally, R, Bennett, WA, Feyrer, F, Fleishman, E. 2010. Bayesian change point analysis of abundance trends for pelagic fishes in the upper San Francisco Estuary. Ecological Applications 20 (5),1431–1448.

Wyatt, RJ. 2003. Mapping the abundance of riverine fish populations:integrating hierarchical Bayesian models with a geographic information system (GIS). Canadian Journal of Fisheries and Aquatic Sciences 60:997–1006.

Table 3B-1. Total catch by species (or family/genus for Gobies) and gear type from the Mitchell et al. (2019) paired trawl dataset. Blank values denote zero catch.

Species or	**Gear Type**						
Group	**20-mm**	**FMWT**	**CMWT**	**SKT**	**SLS**	**STN**	**Total**
American Shad	273	130	2,093	122	48	247	2,913
Delta Smelt	357	34	23	1,688	79	190	2,371
Gobies (Unid)	5,920	8		2	6,016	1,152	13,098
Longfin Smelt	152	5			31	16	204
Prickly Sculpin	14				1,182		1,196
Striped Bass (age 0)	497	19	60		1,367	60	2,003
Threadfin Shad	1,258	144	857	814	526	125	3,724
Total	8,471	340	3,033	2,626	9,249	1,790	25,509

Table 3B-2. Catch by species (or family/genus for Gobies) and gear type by season from the Mitchell et al. (2019) paired trawl dataset. Blank values denote zero catch.

Species or		**Gear Type**						
Group	**Season**	**20-mm**	**FMWT**	**CMWT**	**SKT**	**SLS**	**STN**	**Total**
American Shad	fall		130	2,093	69		1	2,293
	spring	33				48	2	83
	summer	240			53		244	537
Delta Smelt	fall		34	23	1,402		45	1,504
	spring	282				79	68	429
	summer	75			286		77	438
Gobies (Unid)	fall		8				553	561
	spring	263				6,016		6,279
	summer	5,657			2		599	6,258
Longfin Smelt	fall		5				1	6
	spring	152				31	14	197
	summer						1	1
Prickly Sculpin	fall							
	spring	14				1,182		1,196
	summer							
Striped Bass age 0	fall		19	60			3	82
	spring	302				1,367	15	1,684
	summer	195					42	236
Threadfin Shad	fall		144	857	427		3	1,431
	spring	105				526		631
	summer	1,153			387		122	1,662
Total		8,471	340	3,033	2,626	9,249	1,790	25,509

Table 3B-3. Total number of records for seven selected species in the Mitchell et al. (2019) paired trawl dataset used for catchability analysis, summarized by sampling location and gear type. Only data for paired sampling events conducted at the same location were included, where both fish lengths and tow-volumes were recorded. Blank values denote that sampling was not conducted.

		Gear Type						
Season	**Location**	**20-mm**	**FMWT**	**CMWT**	**SKT**	**SLS**	**STN**	**Total**
fall	LowSac		166	1,409	901		440	2,916
	SanJoaq		174	1,624	997		166	2,961
spring	SanJoaq	315				2,441	9	2,765
	SDWSC	835				6,808	90	7,733
summer	LowSac	1,128			728		531	2,387
	SDWSC	4,865					266	5,131
	SanJoaq	1,326					235	1,561
Total		8,469	340	3,033	2,626	9,249	1,737	25,454

Table 3B-4. BUGS code for catchability model. Text in bold defines model terms that are data, and green text are comments. Blue text represents prior (dnorm()) and likelihood (dmulti()) distributions.

```
alpha[1]<-0                                   #due to summation constraint only have to
estimate Ngears-1 alpha's
exp_alpha[1]<-exp(alpha[1])*Effort[1]         #Effort[1] will always be one because it is
                                      calculated as a ratio #of volumes or  tows relative
                                      to first gear type

for(ig in 2:Ngears){
        alpha[ig]~dnorm(0,1.0E-3)             #Uninformative normal prior on log of q
        exp_alpha[ig]<-exp(alpha[ig])*Effort[ig]
}


for(ig in 1:Ngears){
P[ig]<-exp_alpha[ig]/sum(exp_alpha[1:Ngears])   #Estimate proportion of N (total catch
across gear #types) for each gear

        q[ig]<-exp(alpha[ig])                 #relatively catchability per unit of effort
        qpE[ig]<-q[ig]/q[1]                   #expressed relative to first gear type
}

NbyGear[1:Ngears]~dmulti(P[1:Ngears],N)       #multinomial likelihood
```

Table 3B-5. Statistics of the posterior distributions of relative catchability by season, location, and species (or family/genus for Gobies). Relative catchability is calculated for each gear type shown in the 'Gear' column relative to the gear shown in the 'Reference Gear' column. Colors highlight high (green), moderate (yellow), and low (red) median estimates. Estimates for each row in the Gear column are relative to the reference gear. (e.g., CMWT:FMWT for first row).

Fall Season

	Reference Gear			Relative Catchability		
Location		**Species**	**Gear**	**2.5%**	**50%**	**97.5%**
LowSac	FMWT	American shad	CMWT	6.10	7.25	8.68
+SanJoaq		American shad	SKT	0.80	1.08	1.42
		American shad	STN	0.00	0.02	0.10
		delta smelt	CMWT	0.17	0.30	0.50
		delta smelt	SKT	59.67	83.38	119.20
		delta smelt	STN	3.01	4.79	7.55
		Gobies (Unid)	CMWT	0.00	0.00	0.02
		Gobies (Unid)	SKT	0.00	0.00	0.08
		Gobies (Unid)	STN	139.69	261.95	587.62
		striped bass age 0	CMWT	0.81	1.34	2.35
		striped bass age 0	SKT	0.00	0.00	0.02
		striped bass age 0	STN	0.11	0.51	1.53
		threadfin shad	CMWT	2.20	2.63	3.13
		threadfin shad	SKT	4.91	5.94	7.21
		threadfin shad	STN	0.02	0.07	0.18

Table 3B-5. Con't.

Spring Season

	Reference Gear			**Relative Catchability**		
Location		**Species**	**Gear**	**2.5%**	**50%**	**97.5%**
SDWSC	20-mm	American shad	SLS	3.96	6.21	9.87
+ SanJoaq		American shad	STN	0.01	0.05	0.17
		delta smelt	SLS	0.93	1.19	1.52
		delta smelt	STN	0.17	0.22	0.29
		Gobies (Unid)	SLS	17.31	19.88	22.87
		Gobies (Unid)	STN	0.00	0.00	0.00
		longfin smelt	SLS	0.58	0.86	1.26
		longfin smelt	STN	0.05	0.08	0.14
		prickly sculpin	SLS	221.49	351.80	630.31
		prickly sculpin	STN	0.00	0.00	0.02
		striped bass age 0	SLS	16.40	18.57	21.10
		striped bass age 0	STN	0.03	0.05	0.07
		threadfin shad	SLS	17.34	21.33	26.58
		threadfin shad	STN	0.00	0.00	0.00

Table 3B-5. Con't.

Summer Season

Location	Reference Gear	Species	Gear	Relative Catchability		
				2.5%	**50%**	**97.5%**
LowSac	20-mm	American Shad	SKT	52.50	9.27E+07	1.70E+19
		American Shad	STN	1.36	3.31E+06	5.97E+17
		Delta Smelt	SKT	5.73	9.27	16.73
		Delta Smelt	STN	0.02	0.15	0.57
		Gobies (Unid)	SKT	0.00	0.00	0.00
		Gobies (Unid)	STN	0.52	0.58	0.64
		Threadfin Shad	SKT	14.80	30.78	79.43
		Threadfin Shad	STN	0.20	0.79	2.84
SDWSC	20-mm	American Shad	STN	0.80	0.96	1.15
+ SanJoaq						
		Delta Smelt	STN	0.83	1.16	1.66
		Gobies (Unid)	STN	0.00	0.01	0.01
		Striped Bass age 0	STN	0.14	0.20	0.28
		Threadfin Shad	STN	0.08	0.10	0.12

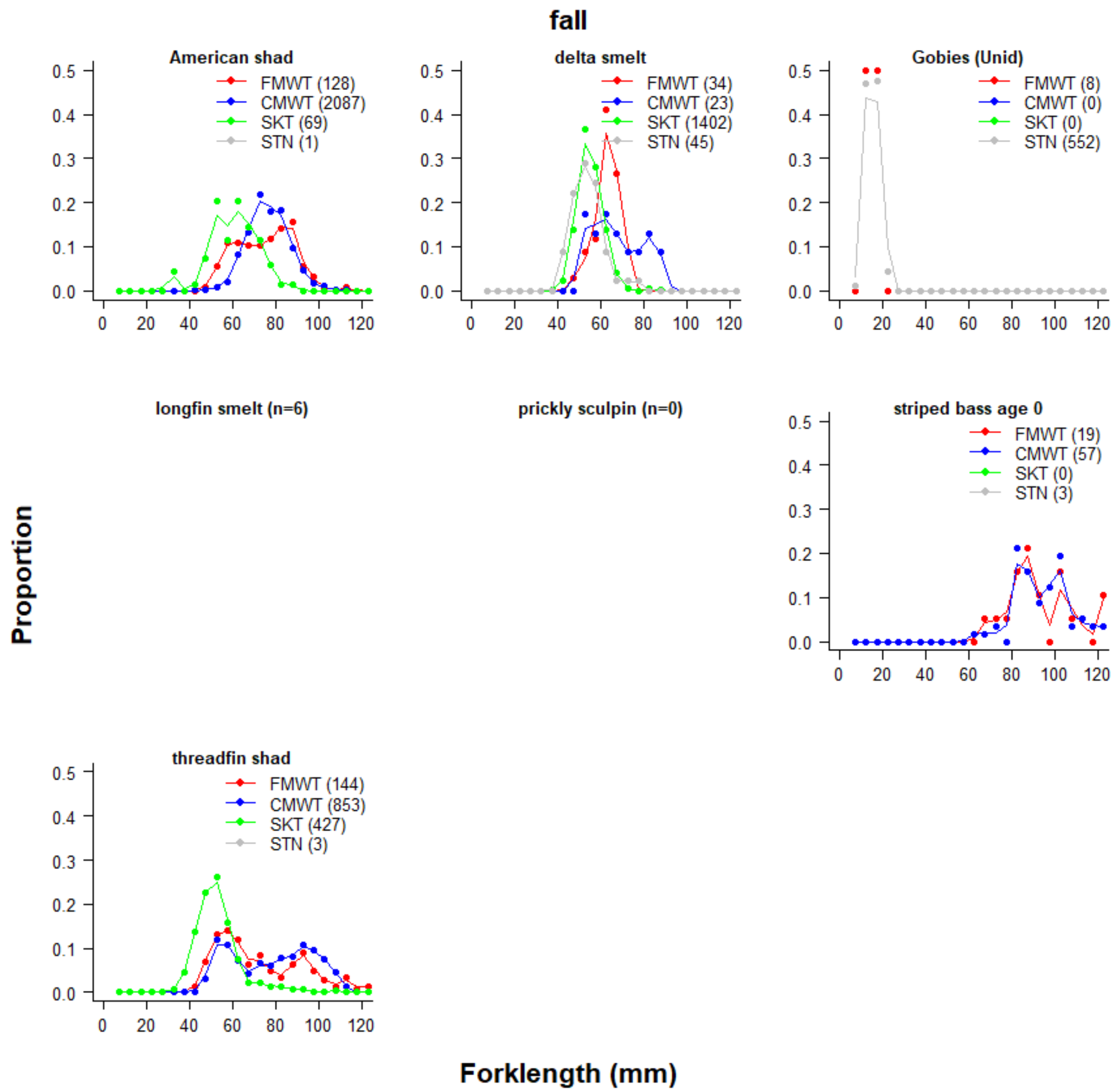


Figure 3B-1. Proportion of catch by 5 mm size class (points) and spline fits (lines) by species (panels). The numbers in parentheses denote the sample size. Selectivity estimates for some gears are not shown in cases when the total catch for a gear was less than 10 individuals. Panels without any data (e.g., longfin smelt) indicate that the total sample size across gears (in parentheses at the top of the panel) was less than 10 individuals.

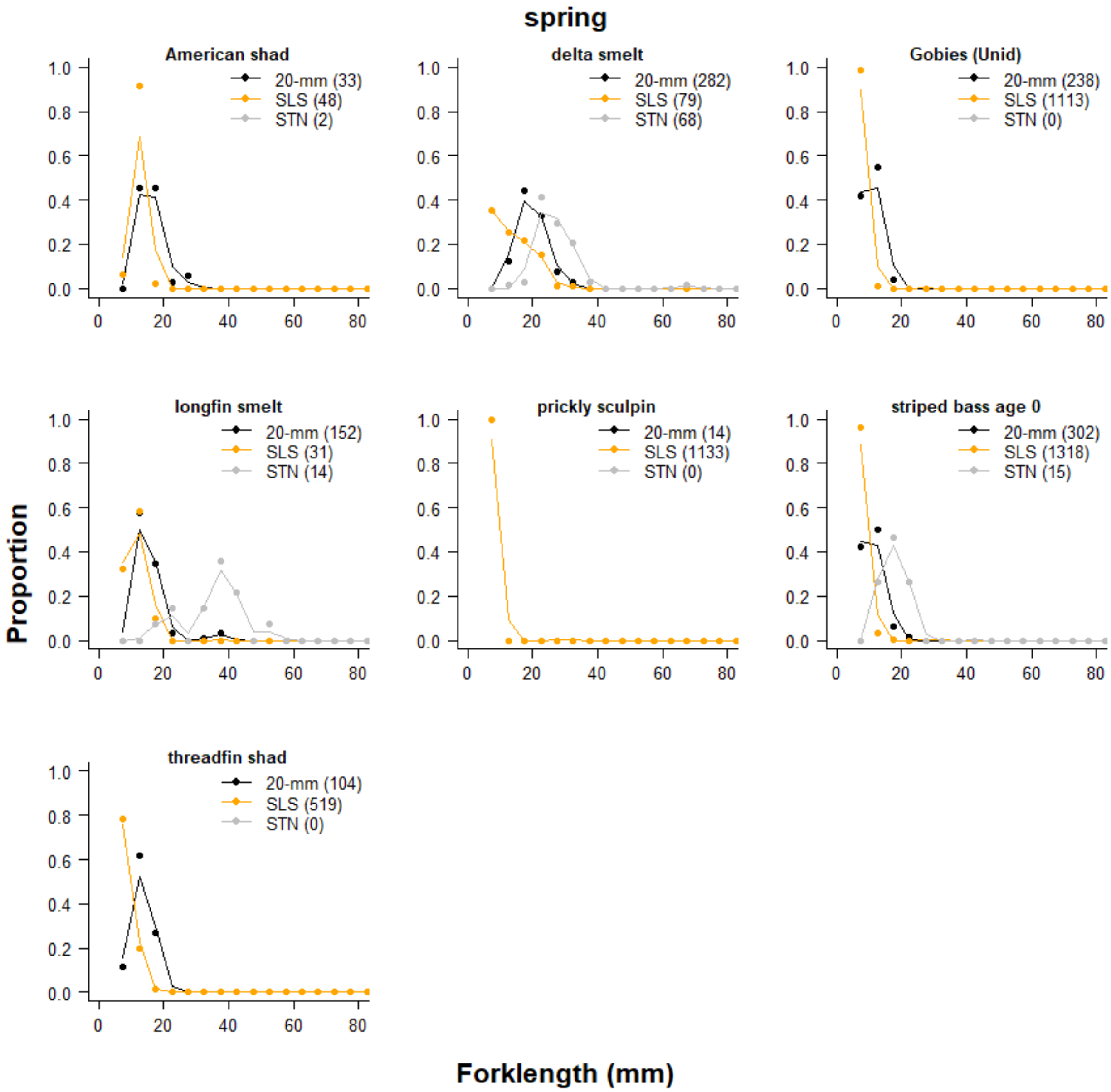


Figure 3B.1. Con't.

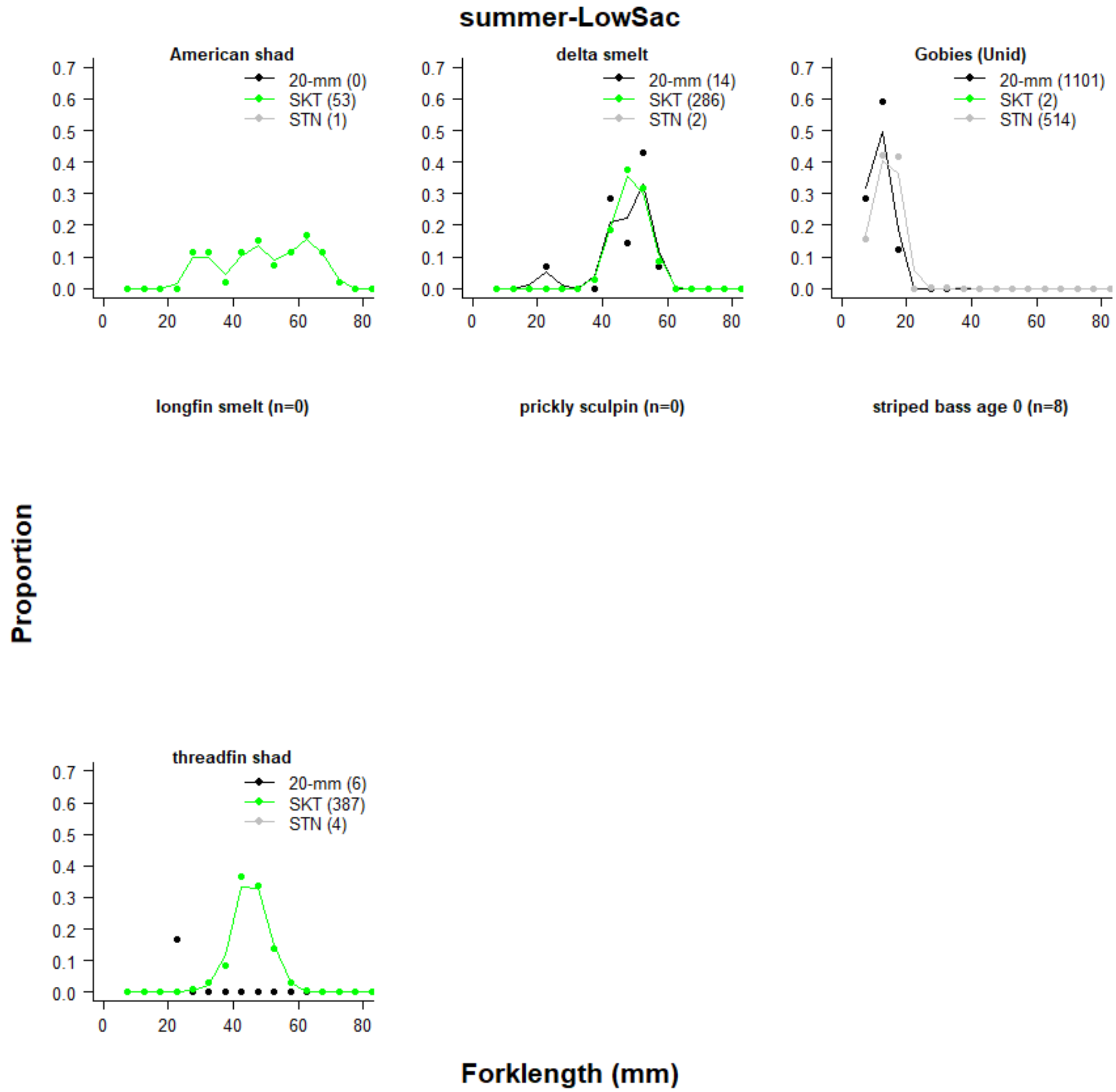


Figure 3B-1. Con't.

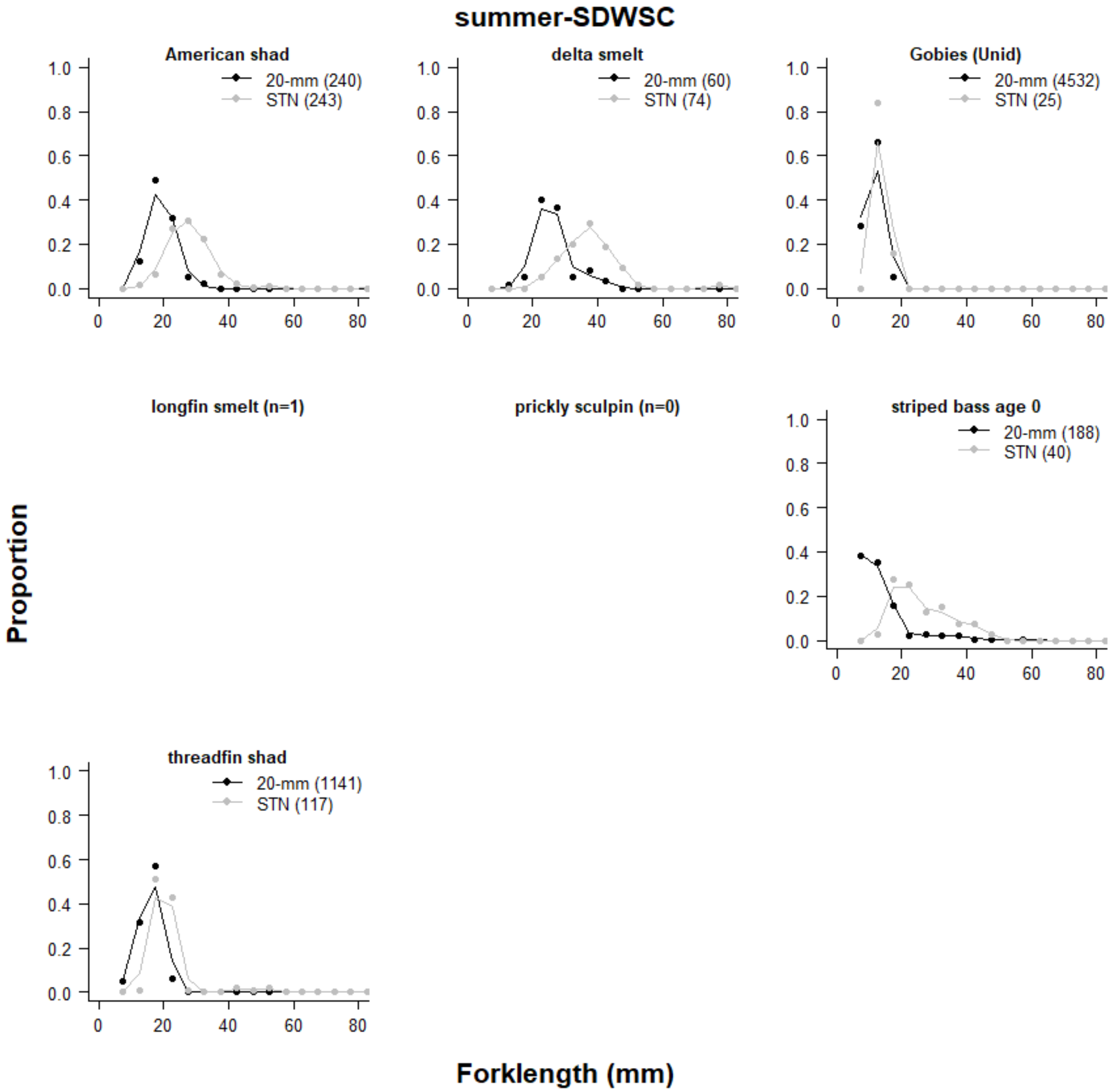


Figure 3B-1. Con't.

Appendix 4. Stratified Sampling Designs for Estimation of Regional Abundance

Introduction

The Design Team evaluated strategies to optimize a stratified monitoring design that can obtain accurate and precise estimates of relative abundance for species across regional strata of the Delta. Relative abundance indices are needed for assessing trends over time and to conduct population-based modeling exercises (i.e., Delta Smelt Life Cycle Model; Maunder and Deriso, 2011; Polansky et al., 2019) using the CDFW survey data. Currently, these indices are simply calculated using the sum or mean of catch per tow calculated for different subregions of the Delta, with the level of effort and weighting of subregion water volumes varying between studies. This method does not determine associated measures of uncertainty in the indices or the methods of estimation and produces unitless values that cannot be readily compared between studies. As a result, it is currently not possible to integrate the index estimates from different studies or to assess the uncertainties of the indices for trend analyses or modeling. The Design Team identified and tested several suitable methods for deriving abundance estimates that would remedy these limitations of the current indexing methods. A design-based approach to abundance estimation was then used to evaluate the effect of various changes in study design on abundance estimates and their precision, such as increasing or decreasing regional sampling effort or reducing station-level replication. In addition, an evaluation of sampling bias for generating abundances and standard errors between fixed and random designs was conducted by comparing 20mm and SKT to the relevant EDSM data.

Comparison of Abundance Estimation Methods

Initially, the Design Team intended to estimate regional abundance and precision by adapting the LTMR model (Chapter 6; Interagency Ecological Program, 2020). This approach employs a Bayesian generalized linear mixed model with a Poisson error distribution, fit in R (R Core Team 2020) with the package brms (Bürkner, 2017); brms employs the probabilistic programming language Stan (Carpenter et al., 2017). In general terms, this model-based approach involves first fitting a catch estimation model for each species and study using all observations from all stations and tows.

This and other model-based approaches to abundance estimation have a number of favorable characteristics including an ability to better accommodate spatial variation in catches within regional strata (Thorson et al. 2015) and the potential to account for environmental or hydrologic drivers of species distribution within a stratum. The Bayesian approach applied in the LTMR may also provide a more complete representation of uncertainty than likelihood-based modeling or design-based estimates since it integrates various sources of uncertainty in the resulting catch-rate estimates through time in a probabilistic framework, and the resulting probability intervals are relatively straightforward to compare in terms of the estimated overlap in abundance indices under different sampling schemes (Bashevkin, 2020). However, with regard to this review effort, this model-based approach proved to have several notable limitations. First, given the need to apply an estimation approach consistently across multiple species and surveys, it was impractical to select and test the inclusion of covariates, thus precluding one of the most notable benefits of the model-based approach. Second, and more critically, model run times during preliminary analysis were on the order of eight or more hours. Proceeding with this method would severely constrain the number of scenarios that could have been evaluated, and so more computationally efficient options were considered.

In addition to the speed at which they can be calculated, design-based approaches to abundance estimation have several other advantages including relative simplicity, fewer assumptions and the ability to be estimated with sparser data (Newman 2008). As with models, many design-based options exist, but substantial effort has already been applied toward the development and evaluation of an estimator for Delta smelt (Newmann, 2008; Polansky et al., 2019), and it was therefore possible to build from an established methodology with a history of application in the Delta. This well-established population estimation approach – stratified random sample ratio expansions (Polansky et al., 2019), offers flexibility and efficiency for calculating survey-specific regional estimates of relative abundance, along with their associated uncertainties.

For the purposes of this review, the benefits of a design-based estimator were clear, but the cost of this approach would be a potentially less refined treatment of uncertainty relative to Bayesian methods. The Design Team therefore conducted an initial, qualitative comparison of abundance estimates and their uncertainty across three methods for the FMWT: the traditional index approach with bootstrapped confidence intervals, the Bayesian model-based approach, and a simplified version of the Polansky et al. (2019) design-based approach.

These three methods were applied to FMWT catch data for each species under review in order to generate abundance/index time-series with associated metrics of uncertainty. Although the precise nature of the interval estimates produced by each method vary, a consistent 95% interval was applied to each. The resulting time series were visually compared for overall trend and magnitude of uncertainty, and pairwise correlations were also calculated.

Model-Based Calculations

Following the approach of LTMR, Bayesian methods were used to fit a generalized linear hierarchical (mixed effects) model with a Poisson error distribution to estimate regional catch rates through time for several species of interest. The models that were fit differed somewhat from those employed by Bashevkin (2020), which estimated catch rates for each year in each *season* (winter, spring, summer and fall). Here the catch rates were estimated for each year in each *region*, where the regional boundaries follow those used in the Enhanced Delta Smelt Monitoring survey (i.e., the five Review Regions identified in Appendix 2: Far West; Suisun Bay and Marsh; Confluence; South; and, North).

The model was fit to data from the FMWT survey during 2002—2020, which corresponds to the current Pelagic Organism Decline regime. In contrast, Bashevkin (2020) fit models to trawl data from multiple surveys, which combined, include catch data across all four seasons. Hence, given the more limited temporal sampling effort from the single survey data used in these analyses, season was not included as a covariate in the model. Instead, an interaction term was used in the model formula on *year*region*, to estimate the annual catch rate in each region. Other than these modifications, the model formula and model fitting methods were the same, including random effects on the intercepts for (1) individual sampling stations, and (2) individual tows. Modeling random effects on individual tows allowed the model to account for overdispersion in the Poisson residual error distribution.

The model formula was:

$Count_{ijk} \sim Poisson(\lambda_{ijk})$,

$log(\lambda) = \beta_1 + \beta_2\, Tow_{\text{volume}} + \beta_3\, \text{Region} + \beta_4\, \text{Year} -- + \beta_5\, \text{Year*Region} + \alpha_{\text{i}} + \alpha_{ijk} + \varepsilon_{ijk}$

$\beta_{[1\text{–}5]} \sim Normal(0, 5)$

$\alpha_i \sim Normal\ (0, \sigma_i)$

$\alpha_{ijk} \sim Normal\ (0, \sigma_{ijk})$,

$\sigma_i \sim HalfCauchy\ (0, 5)$

$\sigma_{ijk} \sim HalfCauchy\ (0, 5)$

Where:

$Count_{ijk}$ is the number of fish for a given species and age-class caught in a tow at station *i*, in region *j* during year *k*; λ represents the mean and variance of the catch rate; β represents the estimated regression coefficients; α represents the intercepts that are modeled as random effects; □ represents the modeled variability around those random effects; and, □ represents the residual error. Parameters were estimated using the *brms* package (Burkner 2017; 2018) in R (R Core Team 2021), which provides a wrapper around the probabilistic programming language, Stan (Carpenter et al. 2017). Weakly informative priors were used, following Bashevkin (2020) and the recommendations of the package authors (Stan Development Team 2020). Models for each species were run with three chains, each with 5,000 iterations sampling from the posterior, of which the first 25% of iterations (1,250) were discarded as a warm-up for the sampling algorithm. Model convergence was diagnosed through Bayesian R-hat values, which for all estimated parameters had values near 1.0. This is consistent with estimates across chains having converged on the joint posterior distribution (e.g., Vehtari et al. 2020).

Design-Based Calculations

In this review, we refer to the design- and model-based estimates as abundance estimates, but for several reasons these approaches are better understood as improved indexing methods. First, the absolute selectivity of the trawl gears remains unknown and although correcting for the relative efficiency of gears can facilitate integration of data, it does not overcome this issue. Second, instead of attempting to estimate the vertical distributions of all species considered in this review and extrapolate trawl catches to unsampled depths, the Design Team has chosen to expand the ratio estimates (i.e., CPUVs) to the water volume that falls within the depth range sampled by each gear type. The design-based estimate for a given taxon, t, in survey s, year y, month m and geographic area a is the product of a region- and survey-specific habitat volume, $V_{s,r}$, and the $CPUV_{t,y,m,s,a}$ of the taxon across tows in that year, month, area and survey; $CPUV_{t,y,m,s,a}$ is calculated as the total catch across tows divided by total volume sampled. These area estimates are then summed to produce a Delta-wide estimate. It is important to note that a can represent any level of geographic aggregation (e.g., Review Region, Review Stratum, EDSM Subregion or any relevant grouping of stations) if the corresponding habitat volumes have been calculated. Finer-scale stratification can reduce the influence of selectivity bias, because the effect of a non-representative station impacts only a smaller proportion of the overall Delta estimate (Polansky et al., 2019). As such, the Design Team chose to utilize the finest-scale geographic stratification for the purposes of this review, and so a represents one of the EDSM subregions while A is the total number subregions (30).

$$I_{t,y,m,s} = \sum_{a}^{A} V_{s,r}\, CPUV_{t,y,m,s,a}$$

with

$$CPUV_{t,y,m,s,a} = \frac{\sum c_{t,y,m,s,a}}{\sum v_{y,m,s,a}}$$

As currently calculated, the variance of $I_{t,y,m,s}$ captures the variability in catch between tows.

$$\widehat{Var}(I_{t,y,m,s}) = \sum_{a}^{A} \frac{{V_{t,s,a}}^2(\hat{s}^2{}_{t,y,m,s,a})}{n_{y,m,s,a}(\bar{v}_{y,m,s,a})^2}$$

with

$$\hat{s}^2{}_{t,y,m,s,a} = \sum_{j=1}^{n_{y,m,s,a}} \frac{(c_{t,y,m,s,a} - CPUV_{t,y,m,s,a} \times v_{y,m,s,a})^2}{n_{y,m,s,a} - 1}$$

where $n_{y,m,s,a}$ is the number of tows conducted, $c_{t,y,m,s,a}$ and $v_{y,m,s,a}$ are the catch and volume of a single tow j, $\bar{v}_{y,m,s,a}$ is the mean volume of tows in a year, month, survey and area and $n_{y,m,s,a}$ is the total number of tows. Polansky et al. (2019) add to this variance formula to incorporate uncertainty in the size-selectivity, and so this approach could be expanded with such information in the future.

Following from the comparison of methods for deriving confidence intervals presented in Polansky et al. (2019), we calculated confidence intervals assuming a log-normal distribution of abundances which served to constrain lower bounds to positive values. The distribution of abundance estimates is then described by the log-mean $\mu = \ln\left(\frac{1}{\sqrt{1+\widehat{CV}^2}}\right)$ with log S.D. $s^2 = \ln\left(1 + \widehat{CV}^2\right)$, and the confidence interval is calculated as the α/2 and 1 – α/2 quantiles of this distribution.

Current Index Calculations

The 20mm, STN and FMWT study each calculate and report an index for at least one species. The number of annual surveys conducted, and the exact calculations vary between studies, but the general steps for a single survey are as follows:

1) Calculate the total catch at each station (i.e., sum across replicate tows if applicable) or area (FMWT divides the Delta in to 17 areas, 14 of which are currently sampled; area means are calculated to account for varying numbers of stations between areas).
2) Multiply the resulting catch values by study- and station/area-specific weighting values. Values reflect the habitat volume represented by each station or area.
3) Sum the expanded catches to a single, Delta-wide estimate.

These survey-specific (i.e., monthly or biweekly) indices are then aggregated to an annual level using different methods depending on the study and species of interest. The FMWT simply calculates the sum across surveys for all indexed species, while the surveys included in the STN and 20mm indices can vary based on the average length of the target species. Because the FMWT uses a consistent methodology and reports indices for six species, comparisons with the model- and design-based approaches were conducted using only FMWT data.

Estimates of uncertainty are not currently calculated or reported for survey catch indices. An approach to estimate the uncertainty of these methods was therefore developed and involved bootstrapping the data, or randomly sampling tows with replacement from the 2002-2019 FMWT data set. The bootstrap sampling was conditional on grouping by year, month, and region. For each bootstrap sample, a value for the design-based index of abundance was calculated. This process was repeated to produce 1,000 bootstrapped datasets, and the corresponding 0.025 and 0.975 percentiles of the resulting values calculated for the indices of abundance represented the 95% bootstrap confidence limits.

Abundance Method Comparison Findings

The three-way comparison of abundance estimation/ index approaches indicated substantial agreement in results across methods (**Figure 4-1, 4-2**). In particular, the estimates produced by the design-based approach were very highly correlated with current indexing approach. Correlations between these methods and the model-based approach were also typically high, though in some cases were reduced by one or more outlier years (e.g., 2006 for Longfin Smelt; **Figure 4-3**). These outliers appear to reflect a greater sensitivity of the design-based and index methods to anomalously high catches in a small number of tows. Such a difference is not unexpected as the modeling approach draws information from other years (through the influence of region and station effects) in generating it's estimates, and so reduces the influence of such outliers. It is not immediately clear whether such smoothing is a beneficial feature of a model-based approach in the context of the Delta, where large and unexpected changes in abundance from one year to the next appear to be common.

A notable and generally consistent difference between the approaches was modestly larger confidence intervals for the model-based estimates (**Figure 4-4**). This may reflect more refined approach to uncertainty inherent in the Bayesian methods. Nevertheless, the differences between interval estimates are quite modest, and given the fundamentally different methods used to derive these metrics of uncertainty, the similarity in their magnitude builds confidence that the model- and design-based approaches produce similar results. Given the high overall level of agreement between methods, the Design Team was confident in proceeding with the more efficient design-based approach.

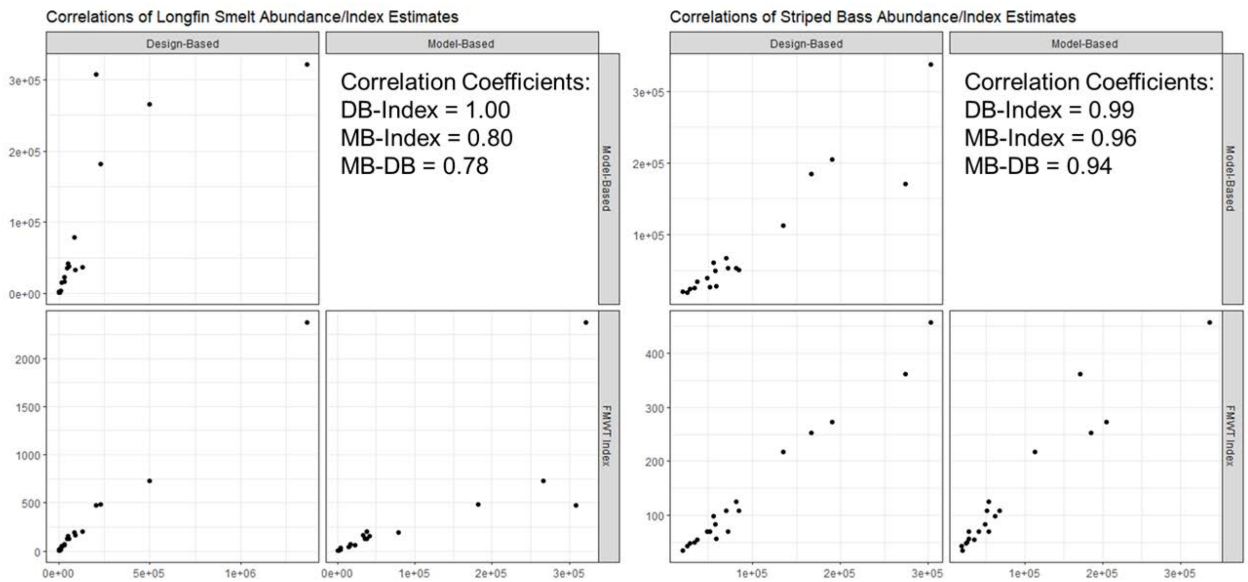


Figure 4-1. Correlations of Longfin Smelt and Striped Bass Abundance / Index Estimates for the FMWT Using the Traditional Index Method, the Design-Based, and Model-Based methods

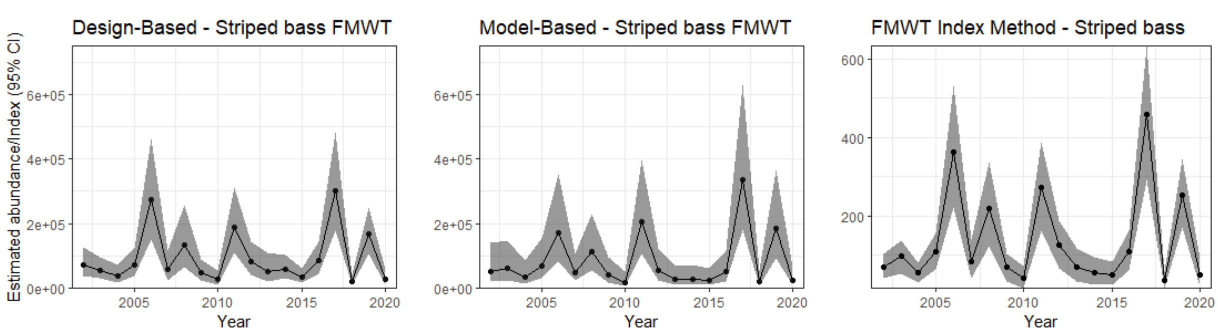


Figure 4-2. Comparison of three abundance calculation methods for Striped Bass in the Fall Midwater Trawl Study. Black lines are the highest estimate of abundance/index in each year and the shading is the 95% confidence intervals of the estimate.

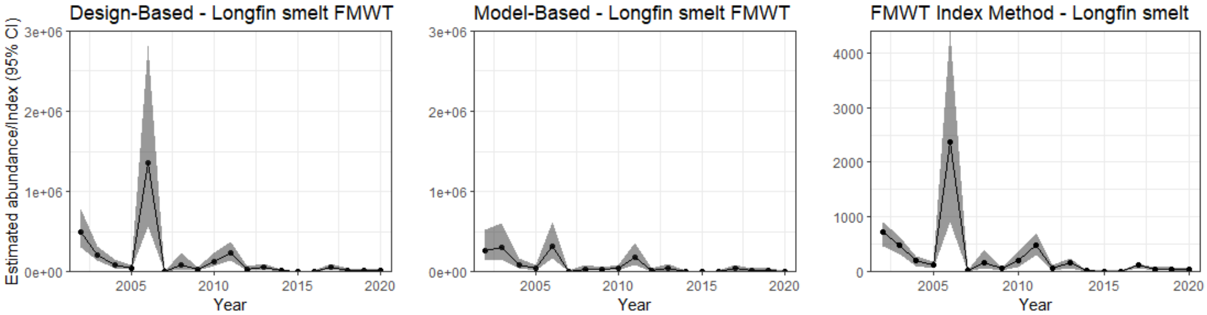


Figure 4-3. Comparison of three abundance calculation methods for Longfin Smelt in the Fall Midwater Trawl Study. Black lines are the highest estimate of abundance/index in each year and the shading is the 95% confidence intervals of the estimate.

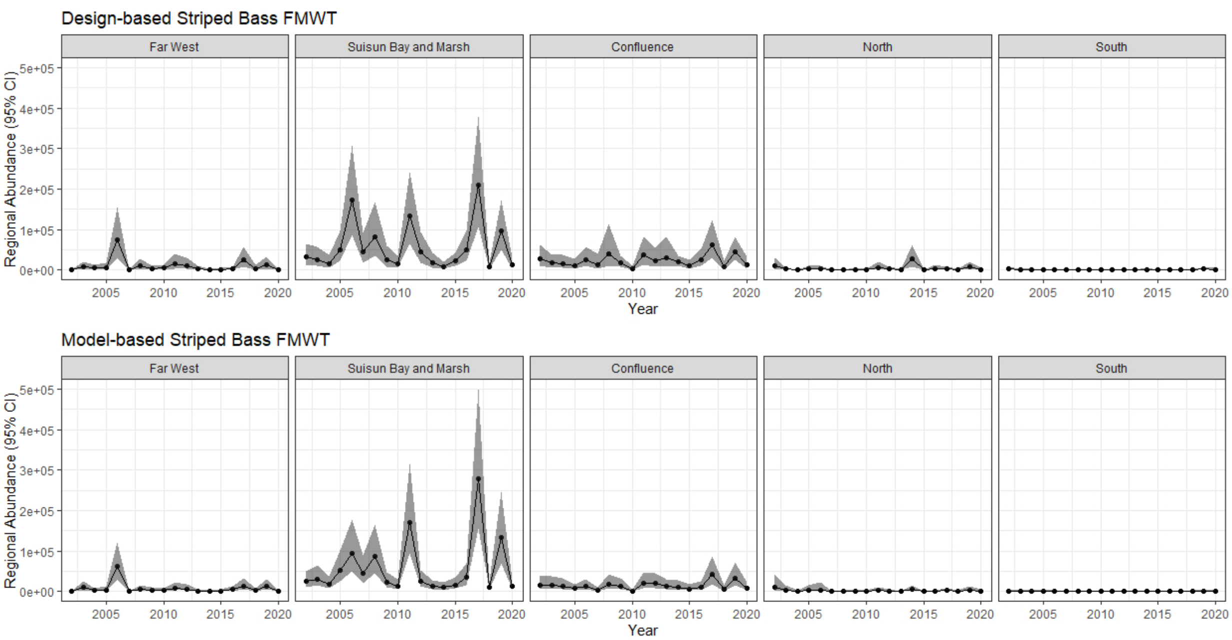


Figure 4-4. Comparison of model- and design-based estimates of regional Striped Bass abundance in the Fall Midwater Trawl Study. Black lines are the mean monthly estimate of abundance/index in each year and the shading is the 95% confidence intervals of the estimate.

Simulations of Altered Sampling Design

The Design Team employed several simulation approaches to understand the sensitivity of changes in sampling effort for status and trends monitoring. Simulations of altered sampling designs for status and trends monitoring focused on metrics of abundance and uncertainty. Effort reduction simulations involved sequential removal of randomly or systematically selected stations, while effort additions involved defining the distributions of historical catch and creating synthetic data through re-sampling. Design-based abundance estimates and their standard errors were then calculated with these modified datasets, and the results compared to status quo effort. The influence of replication on design-based estimates derived from STN and 20mm study data (i.e., the studies for which station replicates are currently collected) was also evaluated through the simulated removal of replicate tows.

Sensitivity of Design-Based Abundance Estimates and Uncertainty to Station Removal
Effort reductions were evaluated at both the region and strata scales and the studies were evaluated separately. The simulation involved sequential exclusion of stations with one additional station from each region or stratum being removed in each iteration. After each removal, design-based estimates of abundance and uncertainty were calculated from the reduced dataset (**Figure 4-5**)**.** Abundance estimates, upper and lower 95% confidence bounds were then compared with full effort through calculation of correlation coefficients (e.g., **Figure 4-6**). High correlations between the full and reduced effort estimates indicate minimal information loss. Stations were removed from each region or stratum until only three stations remained. The number of unique simulations therefore varied between studies and regions/strata since the initial number of stations varies substantially. A correlation coefficient of r = 0.95 was selected as a conservative threshold to minimize information loss resulting from effort reduction.

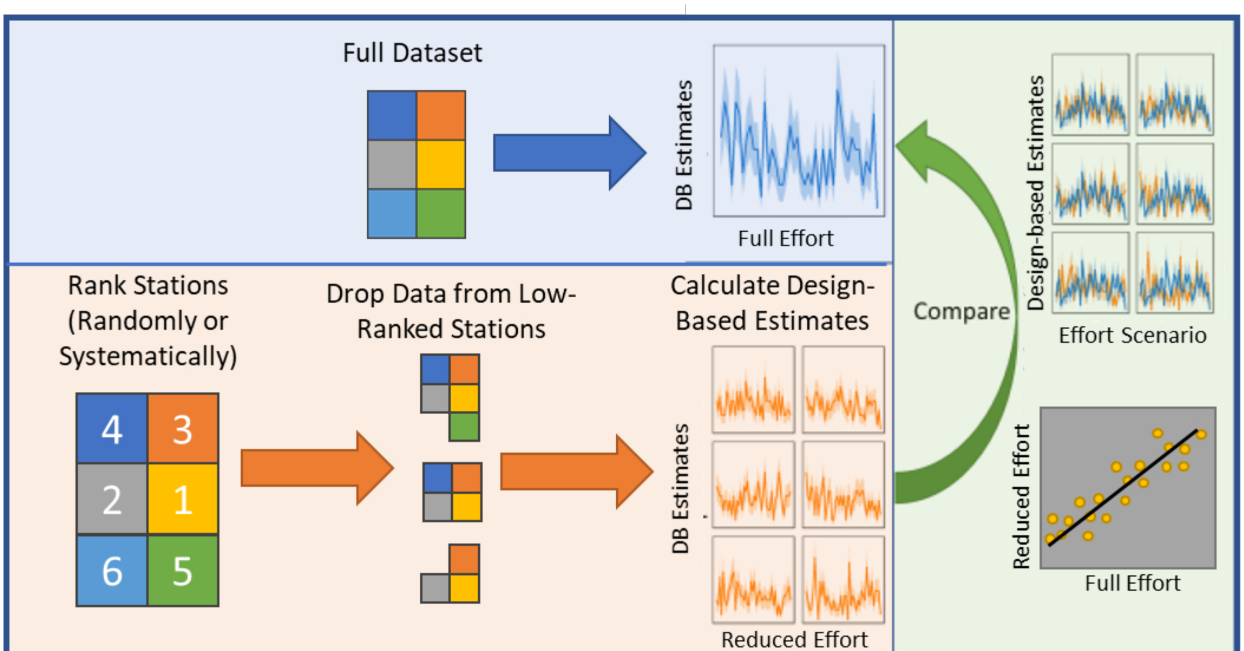


Figure 4-5. Schematic representation of effort reduction simulations. Adapted from Bashevkin et al. (2020).

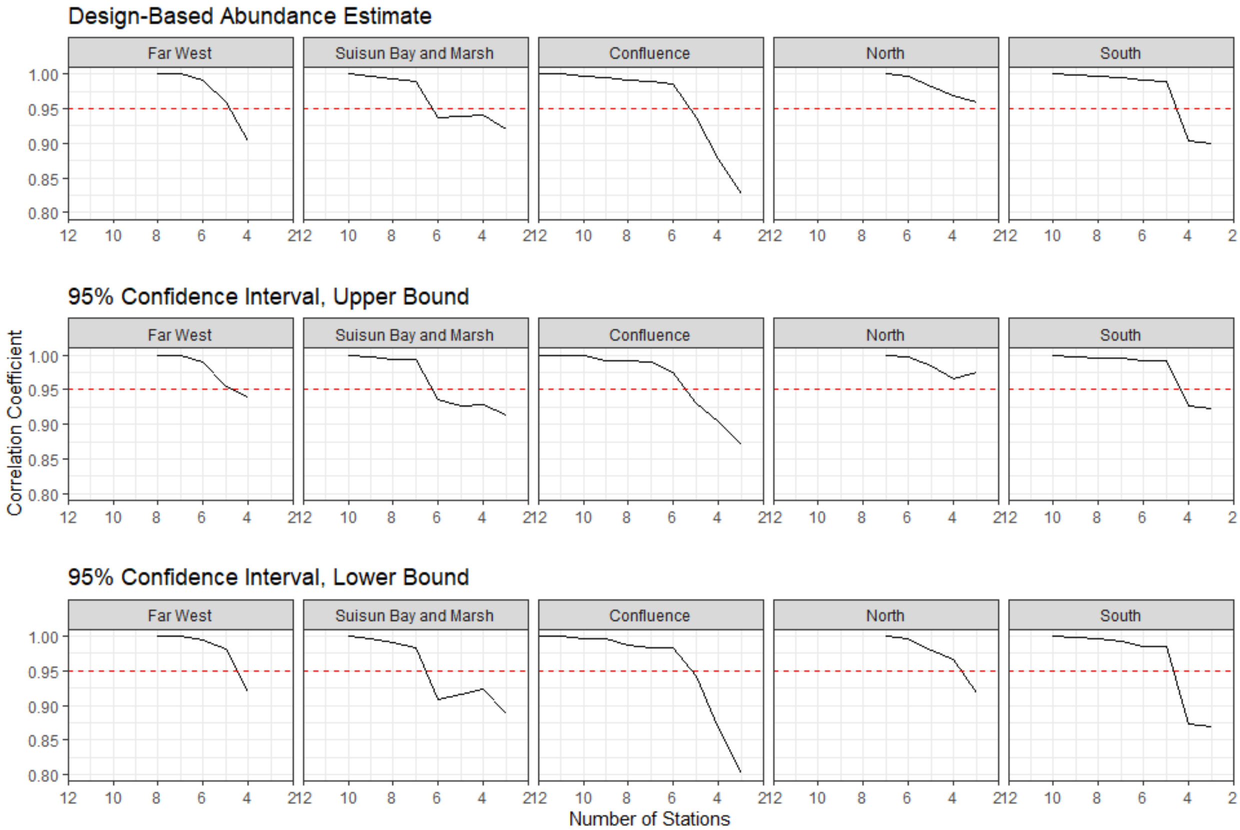


Figure 4-6. Example output of effort reduction analysis showing the change in correlations between full and reduced effort design-based abundance and confidence bound estimates across sequential removal of lowest-ranked stations based on mean CPUV. Results are shown for age-0 striped bass in the 20mm survey.

The order in which stations are removed during these simulations may have important consequences for the outcomes and interpretation of the effort reduction analysis. An exploratory analysis that used FMWT data and randomly removed 25% of stations many times indicated that, on average, the resulting design-based estimates were largely consistent with the full effort results. It was, however, impractical to repeat this analysis for all studies and species. Moreover, some understanding of the variability in information provided by each station is useful, and so the Design Team instead sought to rank the stations in each survey. Any number of ranking criteria could be used, but for the purposes of this review, a relatively straightforward approach was selected. For each study, species and review region the mean CPUV was calculated across all tows in the review dataset and then converted to ranks with the lowest values receiving a rank of 1. The resulting ranks were then added across species within each region to produce a composite ranking (**Figure 4-7**). Under this approach, stations at which many species are regularly encountered in high abundance will be retained while stations at which few species are captured, or catches are generally low will be the first removed during the simulation process.

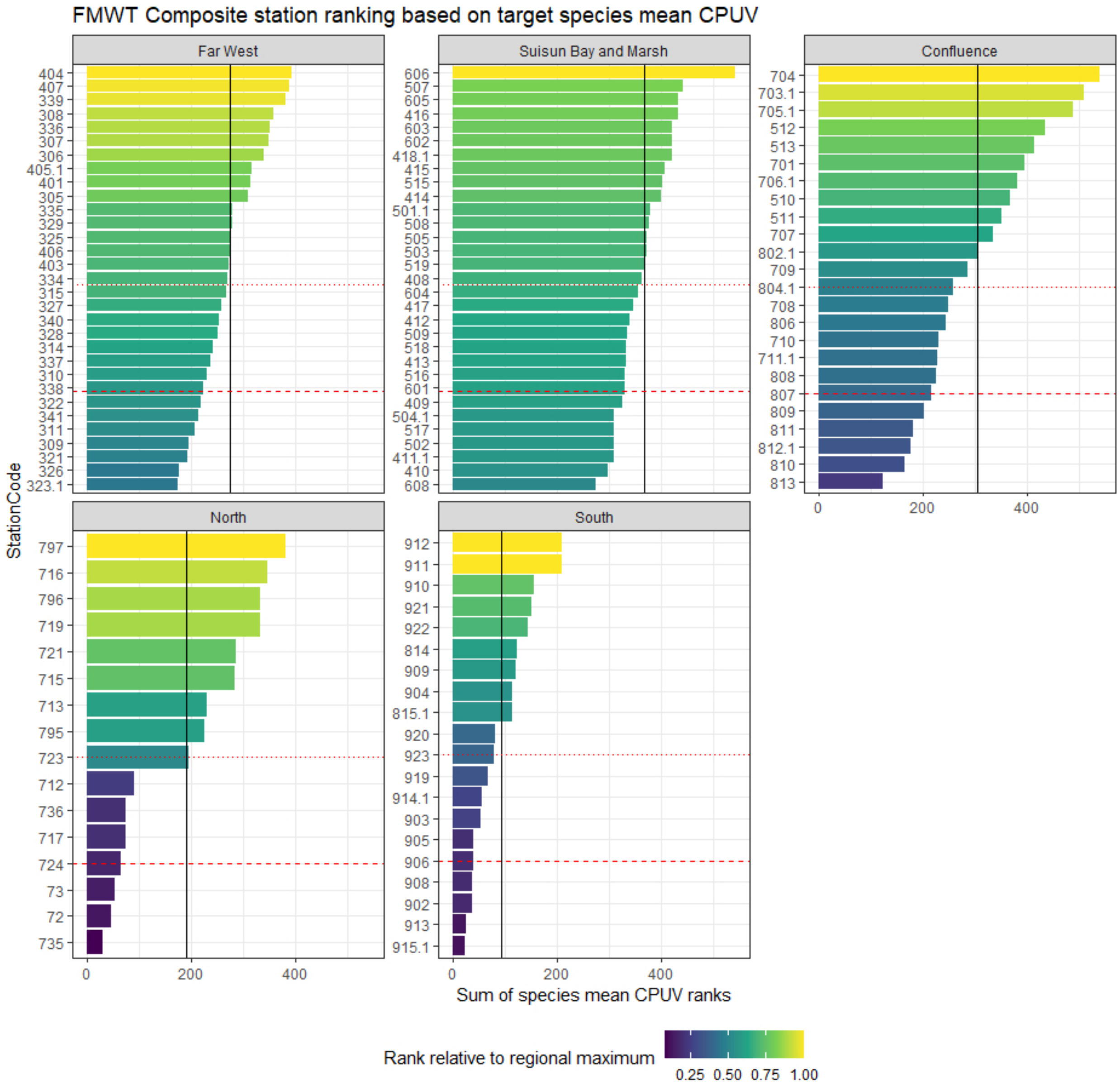


Figure 4-7. Example of station rankings based on CPUV. Bar colors show the ranking relative to the regional maximum in order to highlight within-region variability. Dashed and dotted red lines show the cutoffs for 25% and 50% effort reductions, respectively. Vertical lines show the regional mean of composite rankings.

Sensitivity of Design-Based Uncertainty to Increased Effort

Simulating the addition of effort required a different approach than effort reductions. Simply resampling the available data available in each year, month and region for each study was originally considered. This approach would have in essence duplicated randomly selected tows, thus growing the dataset. The Design Team was concerned that this approach would lead to unrealistic distributions in the resulting data given that the number of tows available for resampling would often be quite small (i.e., <10 in many cases). Thus, an alternative approach was pursued in which a discrete probability distribution was fit to data subsets, and then sampled to generate a fully synthetic dataset (i.e., the original catch data were not included when design-based estimates were subsequently calculated). A zero-inflated negative binomial was selected for this analysis given the overdispersion and high prevalence of zeros common in the review data. This approach involved curve fitting, and again, when data were aggregated to the level of year, month and region the resulting distributions were often very sparse and not likely to produce stable estimates of the distributional parameter. Aggregation by water year type was identified as potential solution to this issue. This approach produces larger data groupings and tended to group years with similar species abundances, which should result in more reliable curve fitting. However, this approach may also inflate estimates of dispersion because although water year type accounts for much of the interannual variability in species abundance, large differences also occur independent of hydrology. The Design Team ultimately decided to proceed with the hydrological aggregation approach in order to increase the number of species and regions for which curve fitting could be successfully applied.

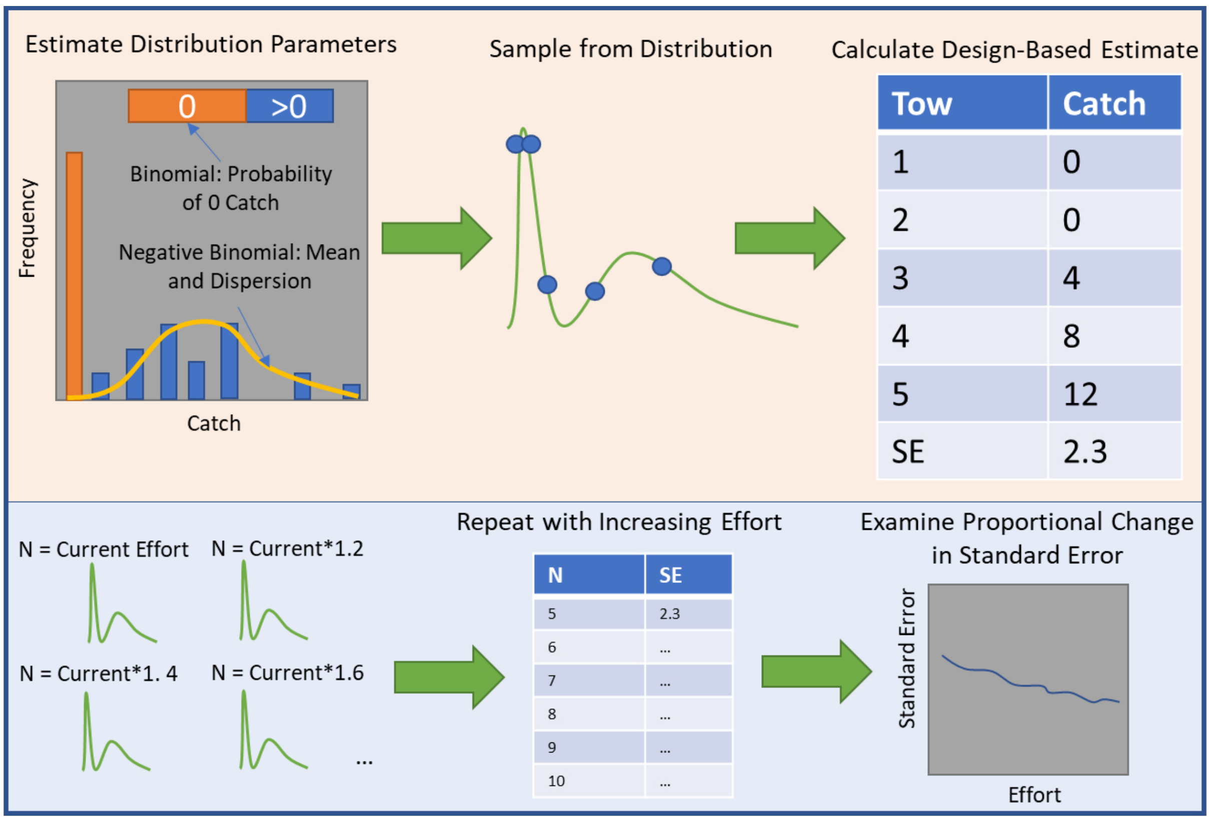


Figure 4-8. Schematic representation of effort addition simulations.

For each combination of species, region/strata, month and water year type, zero-inflated negative binomial data were simulated by first calculating the proportion of tows with zero catch. Next, the parameters of a negative binomial distribution were estimated using the non-zero data and the 'fitdistr' function in the 'MASS' R package (Venables and Ripley, 2002). The three resulting parameters were then used within the R functions 'rbinom' and 'rnbinom' to generate random samples from a binomial and negative binomial distribution with the product of these two samples taken as simulated catch from a single tow. Starting from the current, status quo number of tows conducted during a single survey in each study, month and region, synthetic datasets were generated across an increasing range of effort. The design-based estimates of abundance and uncertainty were then calculated for each resulting dataset (**Figure 4-8**). The focus of this analysis was the ability of increased effort to reduce uncertainty in abundance estimates, and so the change in standard error as effort increased was examined. (Examples shown in **Figures 4-9; 4-10**).

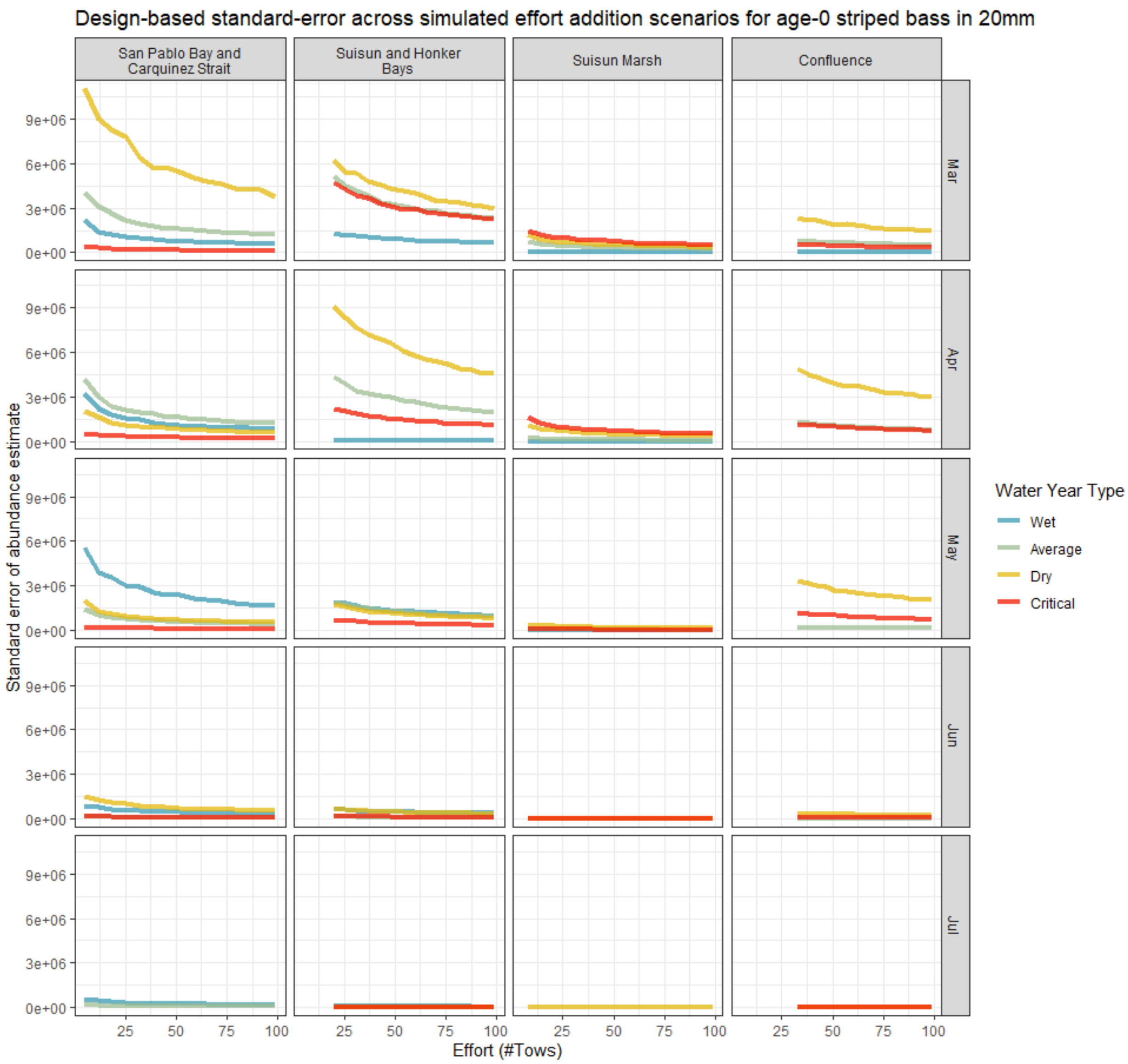


Figure 4-9. Example of effort addition simulation results showing the change in standard error across effort levels and hydrology in selected strata. Simulated data are based on distributions of age-0 striped bass catches in the 20mm study.

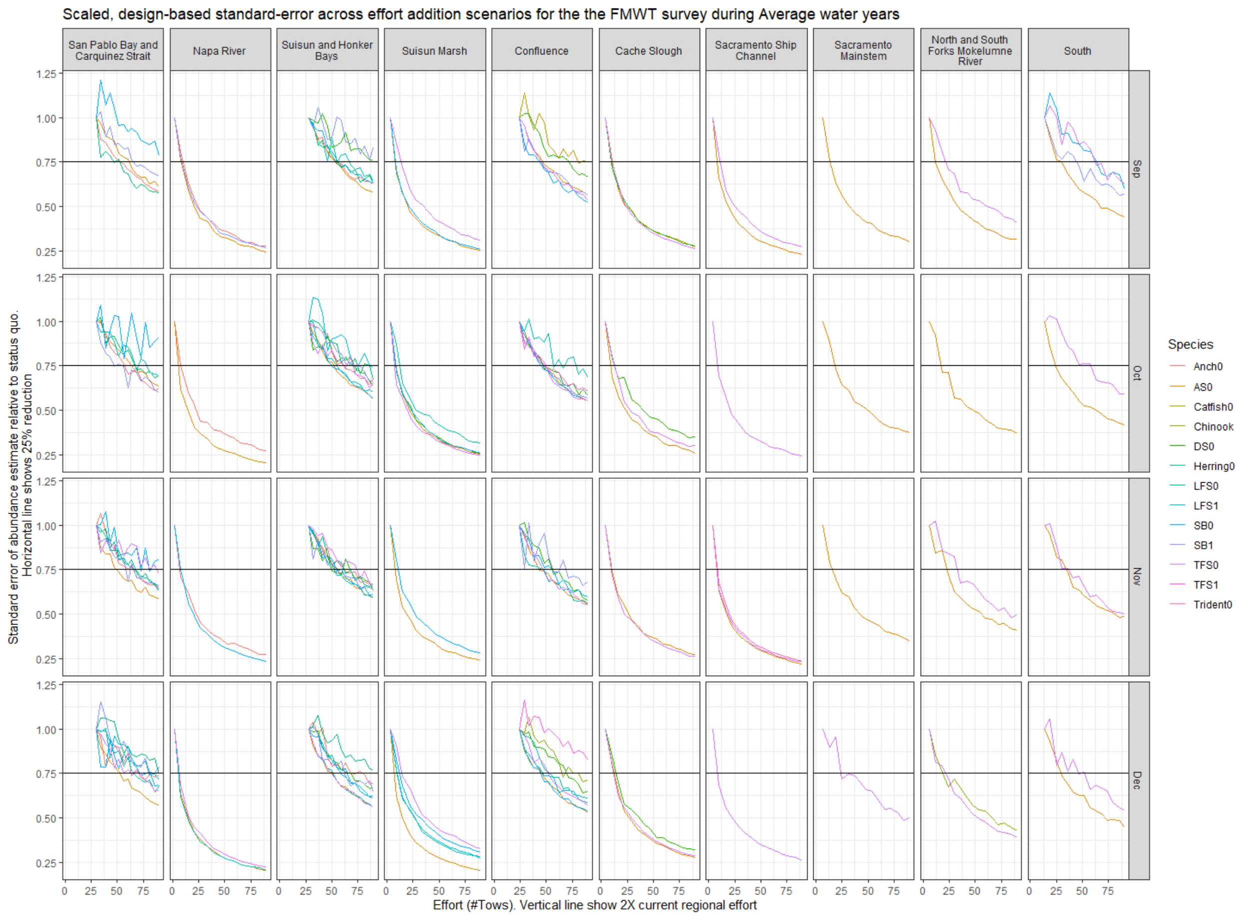


Figure 4-10. Example of effort addition simulation results showing the proportional change in standard error across effort levels for multiple species. Simulated data are based on distributions of catches in the FMWT study. Lines are only shown where sufficient data were available for curve fitting. Solid horizontal line indicates a 25% reduction in SE relative to the status quo.

Evaluating the Influence of Station-Level Replication on Design-Based Estimates

Of the three studies suitable for status and trends monitoring, the STN and 20mm surveys use a replicated design with multiple (typically 2-3) tows conducted at each station during each survey event. Given the relatively small size of these nets, replication serves to increase the total volume sampled and increase the probability of catching rare species. From the perspective of estimating abundance, several options are available for dealing with replicated tows. Current index methods add catches from replicate tows together, but the tows can also be treated as independent which increases the sample size – and potentially reduces uncertainty – of the design-based estimates. This analysis was used to address three questions: 1) is there evidence of depletion between tows (i.e., are abundance estimates from the 2^{nd} or 3^{rd} tows lower than the 1^{st})? 2) Does treating tows independently produce consistently different abundance estimates that combining tows by station? and 3) is uncertainty reduced by treating tows independently (i.e., to design-based estimates have lower standard errors)? The Design Team compared abundance and uncertainty estimates for the STN and 20mm surveys using each tow separately (N = # of stations), calculating station means (N = # of stations), or treating each tow as independent sample (N = # of tows). Catch data from the 20mm and STN studies were first filtered to exclude stations at which fewer than three tows were conducted during a survey, which excluded only a small portion of the dataset. Next, several new subsets of the data were created for each study. Three modified datasets consisted of either the first, second or third tow, and a fourth contained the calculated station mean catches. Design-based estimates were calculated using each of these modified datasets in addition to the full catch data, resulting in five total scenarios. The resulting annual time-series and distributions of standard errors were then visually compared (Examples shown in **Figures 4-11; 4-12; 4-13**).

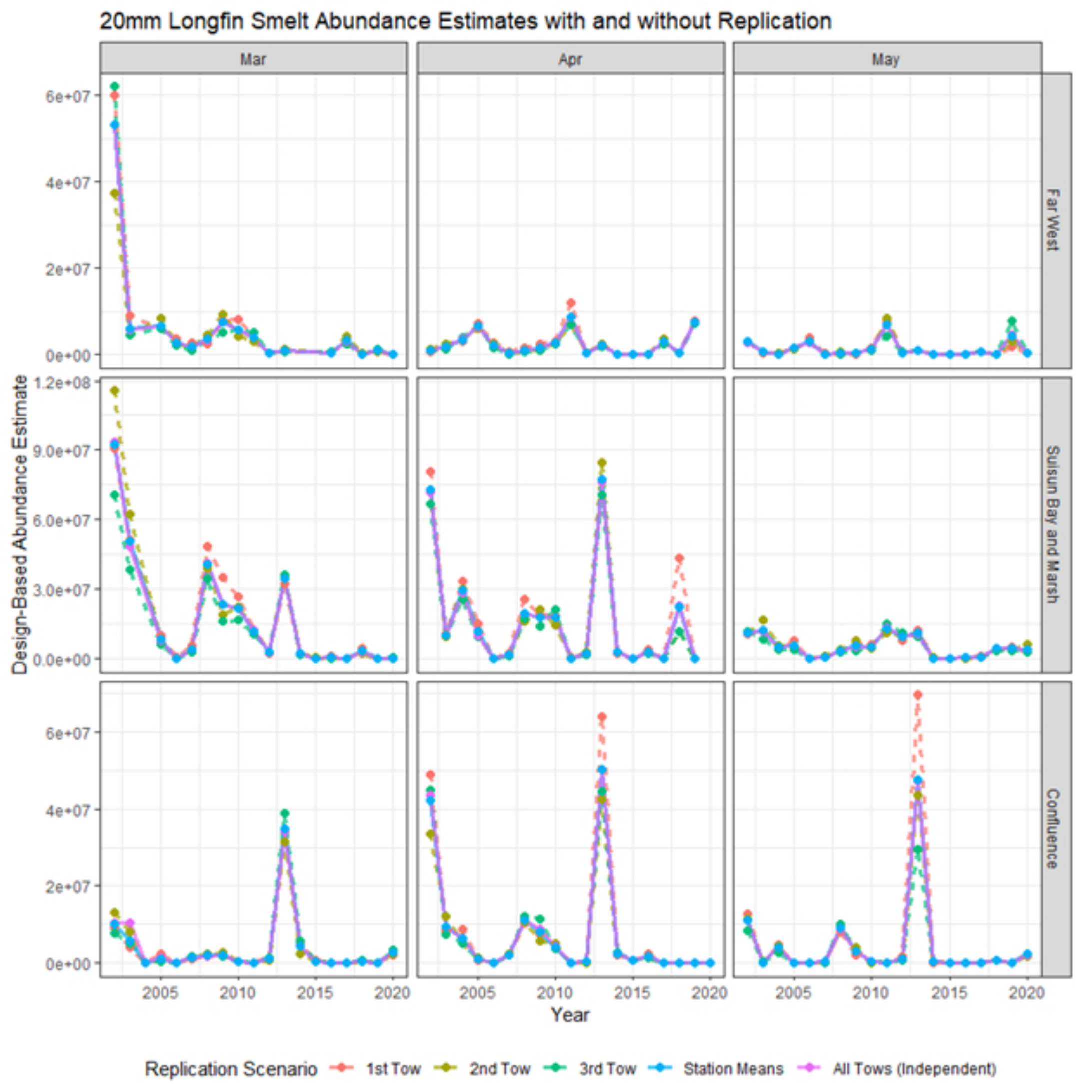


Figure 4-11. Example of variability in 20mm survey design-based abundances across different treatment of replicate tows. Dashed lines show single tows while solid lines show two methods using all tows.

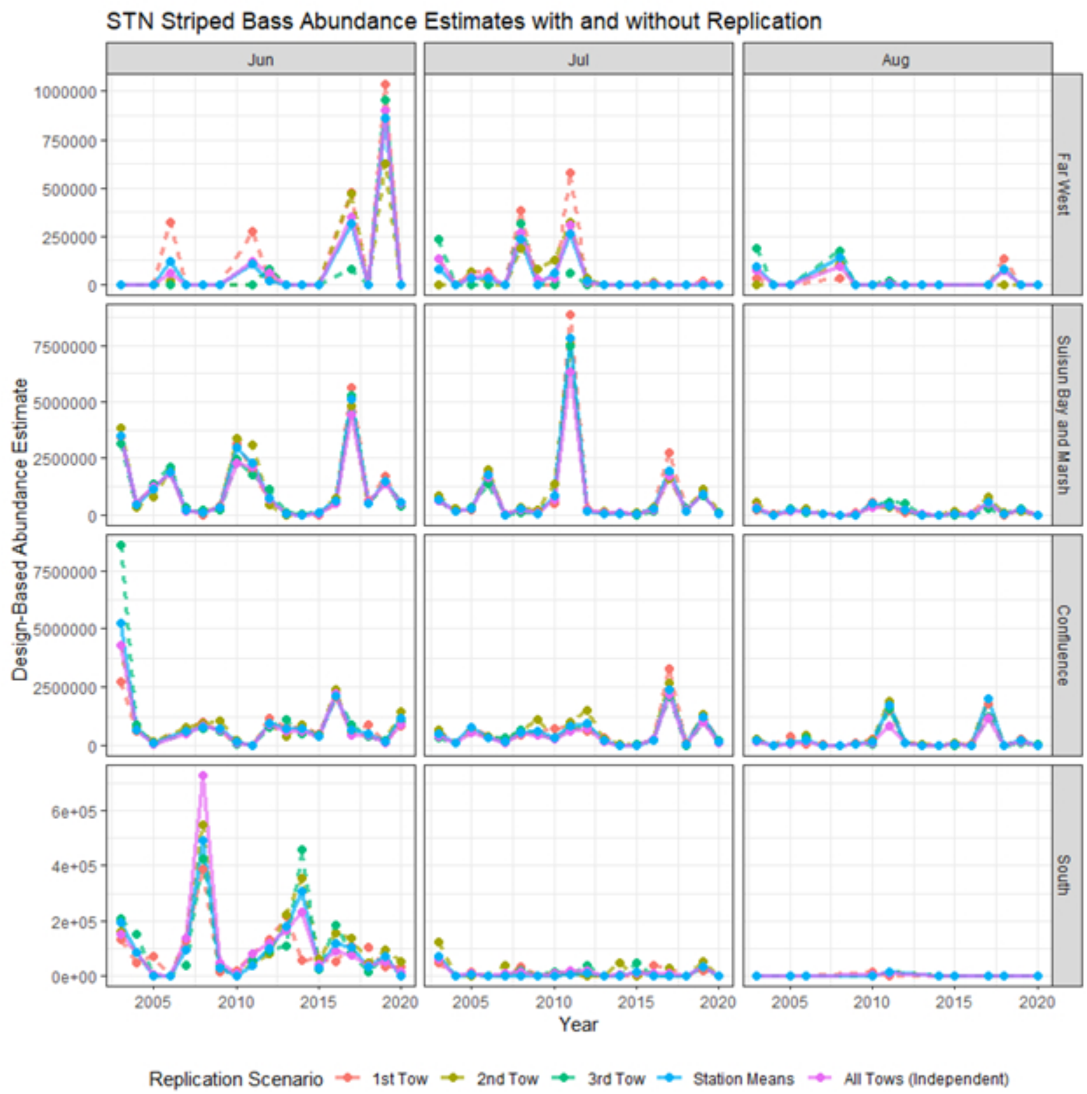


Figure 4-12. Example of variability in STN design-based abundances across different treatment of replicate tows. Dashed lines show single tows while solid lines show two methods using all tows.

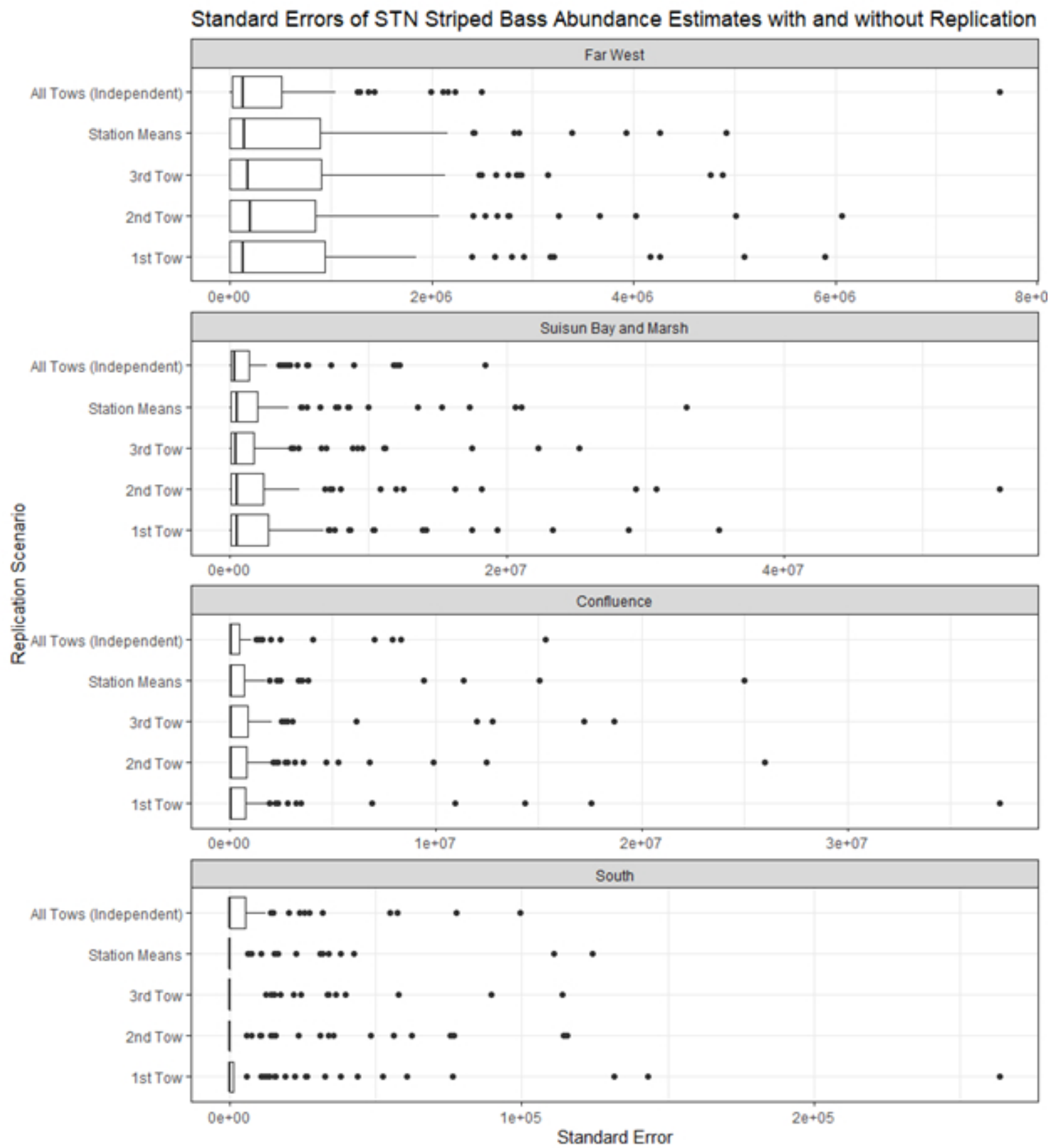


Figure 4-13. Example of variability in the uncertainty of STN abundance estimates across replicate tows and aggregation methods.

Simulation of Altered Sampling Design Findings

The overall results of the effort reduction simulations were fairly intuitive; regions in which current sampling effort is high relative to their habitat volume are more likely to be candidates for dropping stations without significantly changing abundance estimates. As such, only the FMWT, with its large number of stations showed opportunity for effort reduction without data loss. It should be noted, however, that the Design Team imposed a rather conservative limit for defining unacceptable data loss. Moreover, equal weight was given to all species, and a single species could limit the potential for effort reduction. For many individual species, high correlations between full- and reduced-effort abundance estimates could be maintained with very substantial reductions in effort (e.g., **Figure 4-6**), and overall, the resilience of abundance estimates to effort reduction was surprisingly high (e.g., **Figure 4-14**; many correlations remain >0.95 with a 50% or more reduction in effort). However, because these surveys are intended to monitor a fish community with diverse life-history and spatiotemporal distribution, higher levels of effort appear necessary.

The effort addition simulations overall indicated, as would be expected, that increasing effort will generally tend to reduce the standard error of abundance estimates. However, it is also clear that uncertainty in abundance estimates will not decline continuously in response to increased sampling effort. Rather, uncertainty will decline to an asymptote as additional tows are added. For well-sampled areas, further increases in effort will provide limited returns as the underlying, and typically large variability in catch between tows overwhelms the effect of sample size in the standard error calculation (e.g., **Figure 4-9** Confluence; **Figure 4-10** Confluence, Suisun and San Pablo Bays). On the other hand, in regions that are currently sampled with fewer stations (15 or less), standard errors could be reduced efficiently (i.e., with the addition of fewer tows) than in regions where current effort is already relatively high (**Figure 4-15**). Therefore, it is anticipated that spatially balancing effort is likely to have the greatest benefits for reducing uncertainty in regions that are currently under-sampled.

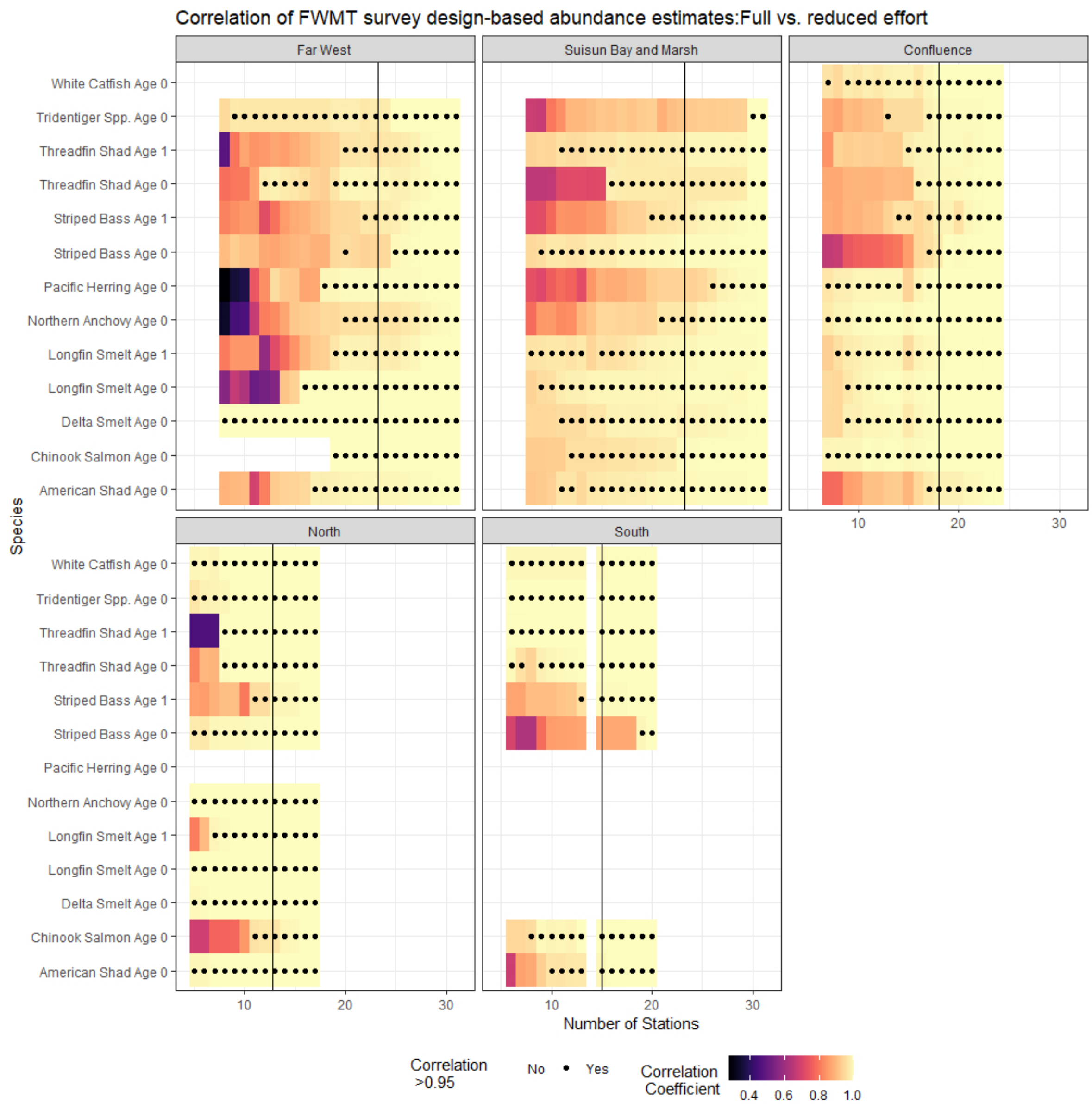


Figure 4-14. Example output of effort reduction analysis showing the change in correlations between full- and reduced-effort scenarios for the FMWT. Colors show the Pearson correlation coefficient, *r*, and black dots indicate that the correlation is greater than 0.95. Vertical line shows the number of stations in each region equivalent to a 25% reduction in effort.

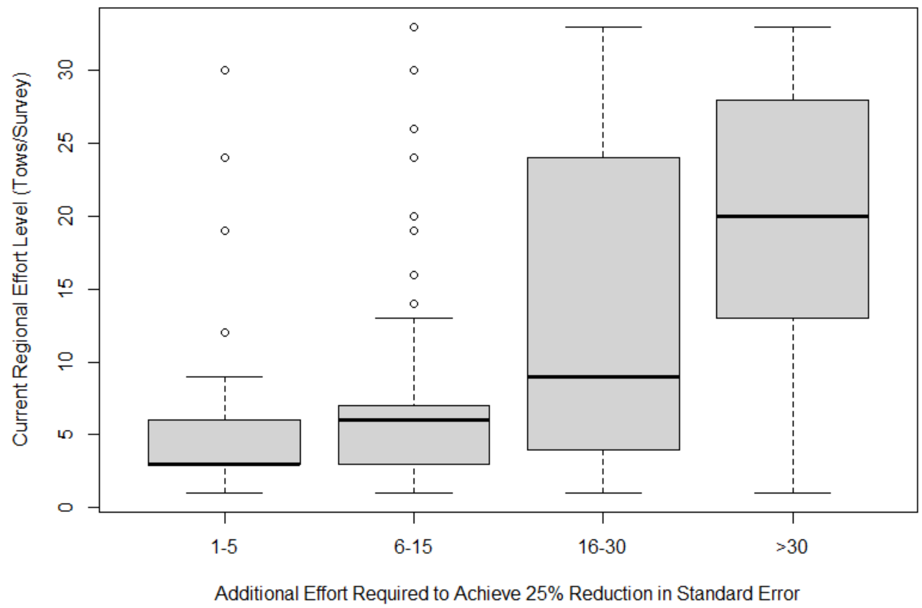


Figure 4-15. Summary of additional tows needed for 25% reduction in standard error.

Replication appears to have limited benefit for abundance estimates. Whether combined or treated independently estimates of abundance from repeated station tows were highly correlated, and where minor differences occurred there was no consistent pattern to which estimate was larger (**Figures 4-11; 4-12**). Our analysis did not show any consistent evidence of depletion suggesting that if it is occurring, it is inconsequential for estimating abundance at the regional scale. Differences in uncertainty between the approaches was largely as predicted, with treatment of replicate tows as independent samples generally resulting in smaller standard errors (**Figure 4-13**). However, in regions or months with sporadic catches (i.e., many zeros), use of replicates actually increased the standard errors (e.g., STN longfin smelt in the Far West). These results suggest that sampling effort would be best allocated over regional scales rather than repeated tows at the same station if the goal is decreasing uncertainty in abundance estimates.

Evaluation of Fixed vs Random Sampling Designs

The Design Team compared the abundance estimates derived from fixed and random monitoring designs by comparing results of the SKT and 20mm survey to the estimates generated by the probabilistic monitoring design employed in EDSM sampling with the same gears. It is assumed that the data generated from random sampling by EDSM represents unbiased estimates against which the fixed station monitoring could be compared. However, it must be acknowledged that EDSM and 20mm differ in their protocols for sampling, with EDSM monitoring exclusively in the surface waters, while 20mm samples obliquely throughout the water column. The abundances from EDSM SKT and 20mm were calculated using the same standardized methodology used to evaluate design-based estimates in the CDFW studies. The Design Team aimed primarily to determine if the randomized study design would result in smaller standard errors (i.e., reduced uncertainty) or consistent higher or lower mean values (i.e., bias) compared to the fixed CDFW monitoring designs. The studies were of course not designed with comparison in mind, and so it was necessary to account for spatiotemporal differences in effort in order to reduce potential bias in the results. To that end, only regions, years and months in which EDSM and CDFW surveys occurred were included the analysis. The resulting set of comparisons was relatively small (Matched pairs for 20mm = 42; SKT=75). For the 20mm comparison, effort was reasonably similar for the CDFW and EDSM studies in most strata apart from the Napa River and Sacramento Mainstem/ Ship Channel (**Figure 4-16**).

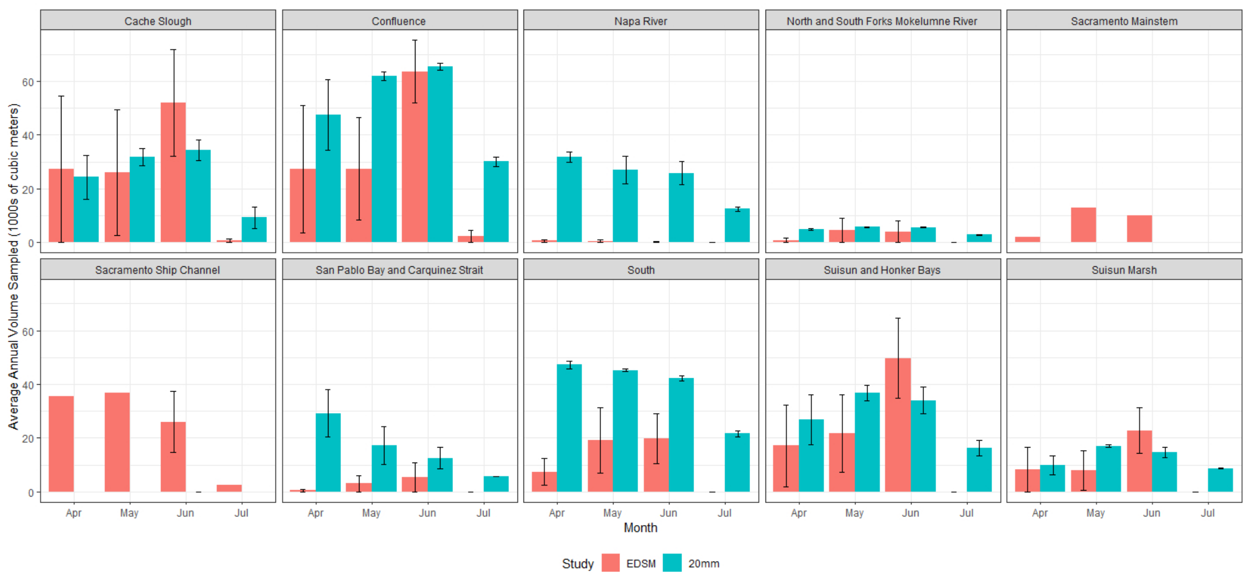


Figure 4-16. Annual 20mm Net Effort by EDSM and CDFW. Error bars show between-year standard error.

For the Kodiak trawl comparison, effort was substantially unbalanced between the studies with EDSM expending up to 20 times more effort than the SKT (**Figure 4-17**). Sample size differences of this magnitude could easily bias the resulting design-based standard error estimates, and so in addition to the comparison of the complete SKT and EDSM datasets, an additional analysis was conducted in which random EDSM tows were repeatedly sampled (1,000 iterations) with spatiotemporal effort matched to that of the SKT.

The comparison of the CDFW and EDSM 20mm surveys provided little evidence that a randomized study design reduces uncertainty of abundance estimates. In fact, the opposite appeared to be the case, with the CDFW 20mm survey standard error being lower in the majority of comparisons for all species except striped bass (**Table 4-1, Figure 4-18**). However, this result must be interpreted cautiously because although the surveys use the same net type in the same areas and months, the EDSM study conducts only surface tows in contrast with the oblique tows of the CDFW 20mm survey. In contrast, the Kodiak trawls conducted by both the CDFW and EDSM studies are surface tows, and so should be more directly comparable.

The results of the Kodiak trawl comparison were variable, but generally indicated that the EDSM survey produced abundance estimates with lower standard errors (**Table 4-2, Figure 4-19**). This pattern appears to result at least in part from the randomized study design (or some other unaccounted-for difference between the studies) and not simply the larger EDSM sample sizes. Some component of the lower abundance estimates by EDSM may be associated with fish that do not reside in surface waters throughout the juvenile life stage. Both the full EDSM and resampled EDSM standard error estimates were lower than the SKT estimates in the majority of comparisons for most species. In many cases, the standard errors of the resampled EDSM estimates were lower than if all EDSM tows were considered. This reinforces the conclusion that simply increasing the number of tows conducted will not necessarily reduce the uncertainty of abundance estimates that can be derived from these trawl survey data.

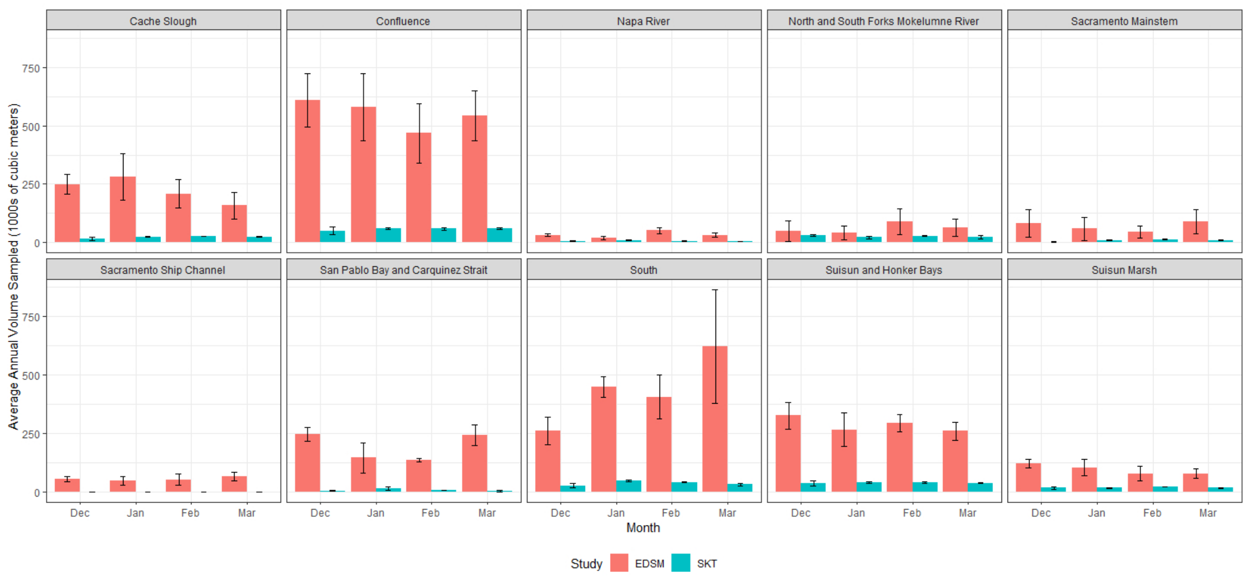


Figure 4-17. Annual 20mm Net Effort by EDSM and CDFW. Error bars show between-year standard error.

Table 4-1. Comparison of EDSM and CDFW 20mm Catch and Design-Based Standard Errors

Species	Total Catch 2017-2020		% of Cases EDSM 20mm SE < CDFW 20mm SE
	EDSM 20mm	CDFW 20mm	
Delta Smelt	42	153	24%
Longfin Smelt	571	3,596	27%
Prickly Sculpin	216	765	36%
Striped Bass	6,987	20,159	64%
Threadfin Shad	27,866	3,211	36%
White Catfish	9	65	30%
White Sturgeon	10	39	6%

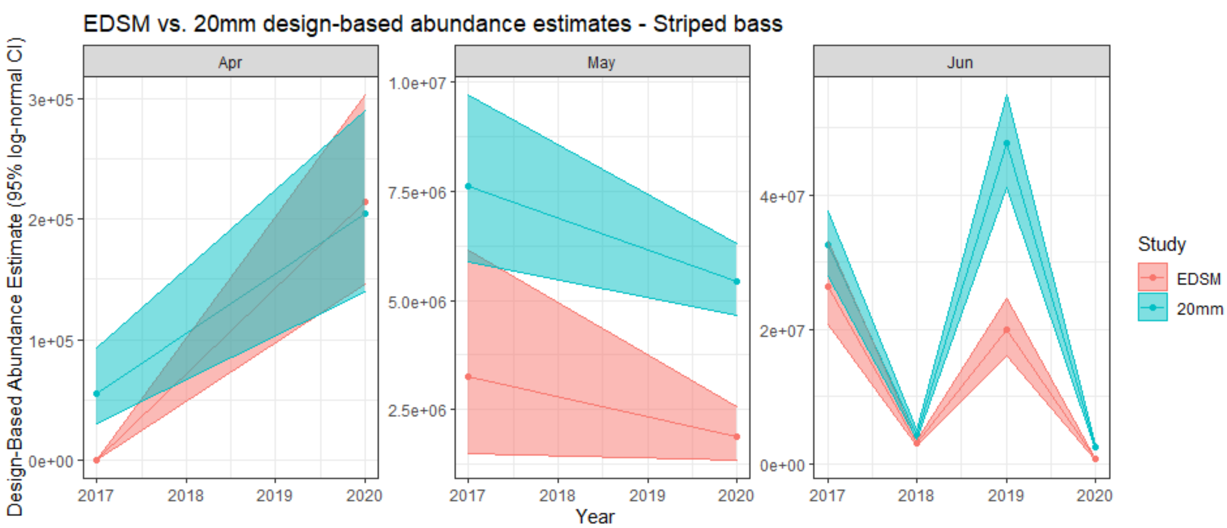


Figure 4-18. Example of Abundance and Uncertainty Estimates from CDFW and EDSM 20mm Surveys

Table 4-2. Comparison of EDSM Kodiak and SKT Catch and Design-Based Standard Errors

Species	Total Kodiak Trawl Catch 2017-2020			% of Cases where Resampled EDSM Kodiak SE < SKT SE	% of Cases where Full EDSM Kodiak SE < SKT SE
	SKT	EDSM (Simulation Mean)	EDSM (Full)		
Chinook Salmon	313	87	1,202	55%	75%
Delta Smelt	300	21	216	65%	81%
Longfin Smelt	33	31	463	39%	44%
Northern Anchovy	676	185	17,306	44%	6%
Steelhead	87	8	111	58%	69%
Striped Bass	72	12	175	73%	63%
Threadfin Shad	6,819	943	43,526	81%	81%

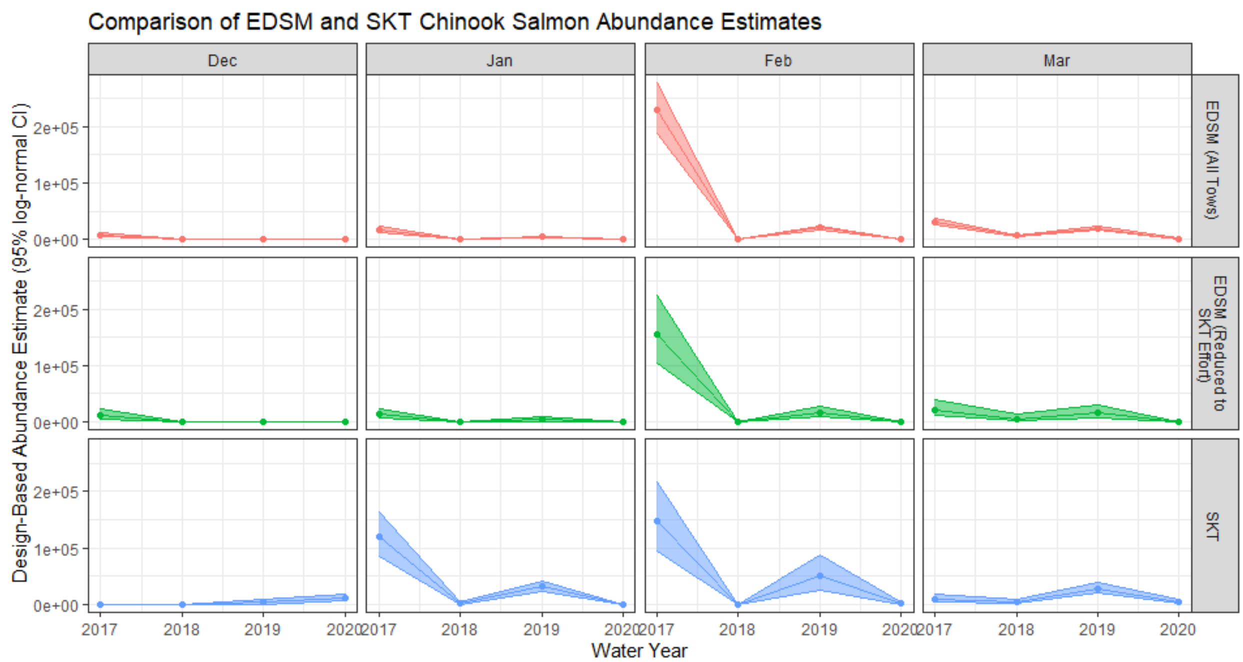


Figure 4-19. Example of Abundance and Uncertainty Estimates from CDFW and EDSM Kodiak Trawl Surveys

This limited analysis suggests that some level of reduction in uncertainty may be possible with a randomized study design, though improvement potential appears to vary between species. Examination of the total catches by the two Kodiak surveys suggests that in many cases the STK survey is more efficient. Perhaps most notably, during this period of comparison, the SKT caught more Delta smelt than the EDSM survey despite conducting only a fraction of the tows. This is not necessarily surprising given that the SKT was designed specifically to target adult and sub-adult Delta smelt, and so it is understandable that selection bias (i.e., non-random selection of sampling locations) would be strongest for this species. In spite of efforts to standardize across space and time, there are still many caveats associated with the comparison of these existing datasets. First, a monthly time scale is likely too coarse for proper comparison and samples from the two studies may have been collected several weeks apart. Similarly, the regional level may be too coarse a level of comparison, but the relatively small number of tows conducted by the SKT precludes analysis at a finer spatial scale. These results ultimately suggest that additional targeted comparisons of fixed and randomized study designs are warranted.

Summary of Findings

Using the data currently generated by the status and trends monitoring studies, it is possible to generate spatially stratified estimates of relative abundance and their associated uncertainties for many of the species considered. A variety of methods are available for this purpose, and overall, there is a high level of consistency in the results regardless of the method selected. Model-based and design-based methods both produced abundance estimates that were generally consistent with the patterns observed using the current indexing approach. However, in contrast to the current indices, these new approaches utilized CPUV instead of catch-per-tow and they can therefore be more reliably expanded based on habitat volumes. Moreover, each of the new methods provides a direct estimate of uncertainty. The efficiency and relative simplicity of the design-based approach led to its use in this review, and these same features could make it a useful replacement for the current indices. Model-based approaches can and should continue to be utilized for specific species in order to incorporate other types of information such as potential hydrologic and environmental covariates of abundance, and the use of a design-based estimator for documenting long-term patterns does not preclude this.

An open question is the appropriate spatiotemporal scale at which these estimates should be considered. The approach used here generates monthly estimates at a range of possible scales of spatial stratification. However, coarser scales (e.g., Annual, Delta-Wide) are often also of management interest. Current methods calculate sums or means across both surveys (i.e., months) and regions to achieve these coarser scales with variation in methods between studies and species. Appropriate application of new abundance estimation methods will similarly require tailoring to species-specific seasonal dynamics. The exact locations and periods for which abundance estimates can be credibly made, given the current spatial extent of sampling, will vary markedly based on the life history characteristics of the species in question. For species with distributions constrained primarily to the Delta as juveniles (e.g., striped bass) or their entire lives (e.g., Delta Smelt), these trawl data should support estimates across a range of scales, including life-stage specific abundances or recruitment indices. On the other hand, estimates for species in which only a portion of the population may be observable in the Delta (e.g., Northern Anchovy) or anadromous species that transit the Delta seasonally must be interpreted much more cautiously. Thus, although the design-based approach to abundance estimation provides a standardized method that can be applied across surveys and species, the application of the method and interpretation of results must account for this important context.

Examination or regional abundances for simulated changes to sampling design indicated that only the FMWT, with its large number of stations relative to habitat volumes, would be a candidate for effort reduction without significant information loss. Conversely, the effort addition simulations indicated that increasing effort in each of the studies will generally reduce the standard error (uncertainty) of abundance estimates for some regions and species. However, it is also clear that the potential for reducing uncertainty in abundance estimates has a fundamental limit, and as such standard errors will not decline continuously in response to increased sampling effort. Achieving 25% reductions in standard error can in some cases be expected with a doubling of effort, but such increases are likely to only be practical in regions where sampling effort is currently low. Although there is certainly scope for fine-tuning, collectively these sensitivity analyses indicate that the design-based abundance estimates are surprisingly robust to changes in effort. As such, balancing effort regionally and across surveys should be prioritized for the status and trends monitoring program.

Replication appears to have limited benefit for abundance estimates. Our analysis did not show any consistent evidence of depletion across replicates. Differences in uncertainty between the approaches was largely as predicted, with treatment of replicate tows as independent samples generally resulting in smaller standard errors. In regions or months with sporadic catches (i.e., many zeros), use of replicates can actually increase the standard error. Our analyses therefore suggest that sampling effort would be best allocated over regional scales rather than repeated tows conducted at the same station if the goal is decreasing uncertainty in abundance estimates. This approach would reduce potential issues associated with pseudo-replication that may occur when replicate tows are treated as independent in the calculation of abundances. This is in contrast to the recommendation for the real-time monitoring studies where replication increases the probability of species detection, and should be maintained or increased at key stations.

Finally, to address concerns about the bias in sampling fish from fixed stations, additional targeted comparisons of fixed and randomized study designs are warranted for status and trends monitoring. The analysis reported here was substantially limited by the available data. The 20mm net comparison is difficult to interpret given the different deployment methods of the CDFW and EDSM surveys (i.e., oblique vs. surface). Although the comparison of Kodiak trawl catches was more robust and suggested that some level of reduction in uncertainty may be possible with a randomized study design, these results cannot be readily generalized across studies. Furthermore, the CDFW and EDSM Kodiak surveys were not conducted with any comparison in mind, and so the actual sampling dates could have occurred on any day in a given month. Targeted comparative studies for the STN, FMWT and 20mm gears should therefore be prioritized in order to better understand potential selectivity bias in the fixed-station designs.

References

Bashevkin, S.M. 2020. Chapter 6: A Framework for Evaluating the Effects of Reduced Spatial or Temporal Monitoring Effort. In: Interagency Ecological Program Long-term Monitoring Element Review: Pilot approach and methods development (2020). IEP Technical Report. 201 pp.

Bürkner, P.-C. 2017. brms: An R Package for Bayesian Multilevel Models Using Stan. Journal of Statistical Software 80: 1–28. doi:10.18637/jss.v080.i01

Bürkner, P.-C. 2018. Advanced Bayesian Multilevel Modeling with the R Package brms. The R Journal 10: 395–411.

Carpenter, B., A. Gelman, M. D. Hoffman, and others. 2017. Stan: A Probabilistic Programming Language. Journal of Statistical Software 76: 1–32. doi:10.18637/jss.v076.i01

Newman, K.B., 2008. Sample Design-Based Methodology for Estimating Delta Smelt Abundance. San Francisco Estuary and Watershed Science 6. https://doi.org/10.15447/sfews.2008v6iss3art3

Polansky, L., Mitchell, L. and Newman, K.B. 2019. Using Multistage Design-Based Methods to Construct Abundance Indices and Uncertainty Measures for Delta Smelt. Trans Am Fish Soc. 148:710-724

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Stan Development Team. 2020. Stan User’s Guide, Version 2.24.

Thorson, J. T., A. O. Shelton, E. J. Ward, and H. J. Skaug. 2015. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. ICES Journal of Marine Science 72:1297–1310.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B. and Bürkner, P.-C. 2020. Rank-normalization, folding, and localization: An improved R-hat for assessing convergence of MCMC. Bayesian Analysis. 1–28. dpi:10.1214/20-BA1221

Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0