UC Davis
San Francisco Estuary and Watershed Science

Title
A Framework for Evaluating the Effects of Reduced Spatial or Temporal Monitoring Effort

Permalink
https://escholarship.org/uc/item/5qh251f3

Journal
San Francisco Estuary and Watershed Science, 20(3)

Author
Bashevkin, Samuel M.

Publication Date
2022

DOI
10.15447/sfews.2022v20iss3art5

Peer reviewed

A Framework for Evaluating the Effects of Reduced Spatial or Temporal Monitoring Effort

Samuel M. Bashevkin*[1]

ABSTRACT

Monitoring in the San Francisco Estuary (estuary) has fluctuated in sampling effort over time with changes to resources, objectives, and unforeseen events. I designed an approach to evaluate how reduced sampling would alter our ability to describe the status and trends of key species. This approach can evaluate the sensitivity of the estuary monitoring program to disruptions in sampling, and whether sampling effort could be reduced without compromising the information provided by these surveys. I simulated reduced sampling on top of the historical data record (1985–2018) by selectively removing data and evaluating the effect on model inference. The same model structure is fit to the full data set and several reduced data sets that represent simulations of reduced sampling effort. I then compared model predictions from reduced models to those from the full model to evaluate how reduced sampling may have affected our ability to detect key patterns in the data. In a case study, I applied this approach to Sacramento Splittail abundance trends from the Bay Study and the Suisun Marsh Fish Study otter trawls. Sampling reductions of 10% and 20% had fairly low impacts on the overlap of reduced model predictions with those from the full model. These results demonstrate the utility of my approach, but they are not generalizable beyond our ability to detect trends in Splittail abundance from Bay Study and Suisun Marsh Fish Study otter trawl data. A thorough analysis should run these simulations on multiple species and multiple parameters (e.g., abundance, distribution, length). By simulating sampling reductions on top of historical conditions, this approach could evaluate differential effects in varying environmental or historical conditions (e.g., droughts, species declines, invasions). In addition, this approach can easily be extended to other functional groups (e.g., zooplankton, phytoplankton) as well as physical parameters (e.g., temperature, salinity, Secchi depth).

INTRODUCTION

Ecological monitoring is a requisite for effective environmental management (Callahan 1984; Lindenmayer and Likens 2010). Long-term monitoring data sets are valuable for applied work in environmental management as well as in answering broader, fundamental questions in ecological and evolutionary biology (Callahan 1984; Hobbie et al. 2003; Lindenmayer and Likens 2010). However, as with any long-term program, periodic review is needed to ensure the program is most effectively meeting its stated objectives and using its resources (Vos et al. 2000; Radinger et al. 2019). These periodic reviews are an opportunity to (1) evaluate whether the program is still producing useful information, (2) consider new management options that may change monitoring objectives, and (3) synthesize monitoring data to evaluate the relevance of the produced information and the adequacy and efficiency of the sampling design (Reynolds et al. 2016).

The San Francisco Estuary (estuary) has a long history of ecological monitoring (Tempel et al. 2021), rivalling that of many other large estuarine systems. Much of this monitoring is coordinated under the Interagency Ecological Program for the San Francisco Estuary (IEP), a consortium of state and federal agencies. The IEP conducts numerous seasonal and gear-specific monitoring surveys that measure water quality, fishes, zooplankton, phytoplankton, benthic organisms, and other variables. Sampling effort has fluctuated over time as stations have been added and removed, or events have interfered with normal sampling schedules (Gaeta 2021; Tempel et al. 2021). Previous reviews focused on single surveys or species have qualitatively reviewed sampling designs, resulting in improvements to a subset of the monitoring program (e.g., the review of winter-run Chinook Salmon monitoring [Johnson et al. 2017] and the review of the Delta Juvenile Fish Monitoring Program [IEP SAG 2013]). However, no studies have yet quantitatively investigated the effects of altered sampling effort across multiple estuary monitoring surveys. An optimal sampling program (optimizing cost

and accuracy) may require shifts to the spatio-temporal sampling effort.

With a wealth of available historical data, one can look into the past to evaluate how different sampling design scenarios may have affected the quality of monitoring information. Specifically, I focus on the effects of reduced sampling effort to illustrate an approach that takes advantage of prior data. Understanding these effects is important for determining (1) the sensitivity of the estuary monitoring program to disruptions in sampling, and (2) whether sampling effort could be reduced without compromising the value of the information these surveys provide. Discovering redundancies in the sampling program could also help release resources to address issues of catchability (Huntsman and Mahardja 2021; Huntsman et al. 2022) or redirect monitoring efforts to less-sampled regions, taxa, or habitats. In this study, I developed a framework for data-focused statistical evaluations of the estuary monitoring program. I then demonstrated this framework on the Bay Study and UC Davis Suisun Marsh Study (Suisun Study), by evaluating the ability of their otter trawls to describe the status and trends of a key fish species: the Sacramento Splittail (*Pogonichthys macrolepidotus*) (Figure 1).

I designed the framework around the general question: *how do the abilities of these surveys to monitor the status and trends of the estuary change when sampling is reduced?* Specifically, I evaluated how much and along which axes (time or space) sampling could be reduced before the surveys no longer provided useful information. To address this question, I adopted an approach similar to the non-random sampling of previously-collected monitoring data described by White and Bahlai (2021) and White (2019). In this approach, data points in a long-term monitoring data set are removed to simulate scenarios of altered sampling effort, and the resulting effects on model inference are evaluated. Rather than simulating data for a power analysis as many monitoring program design studies do (Gerrodette 1987; Rhodes and Jonzén 2011; Barry et al. 2017; Christie et al. 2019; Weiser et al. 2019), this approach makes full use of the historical
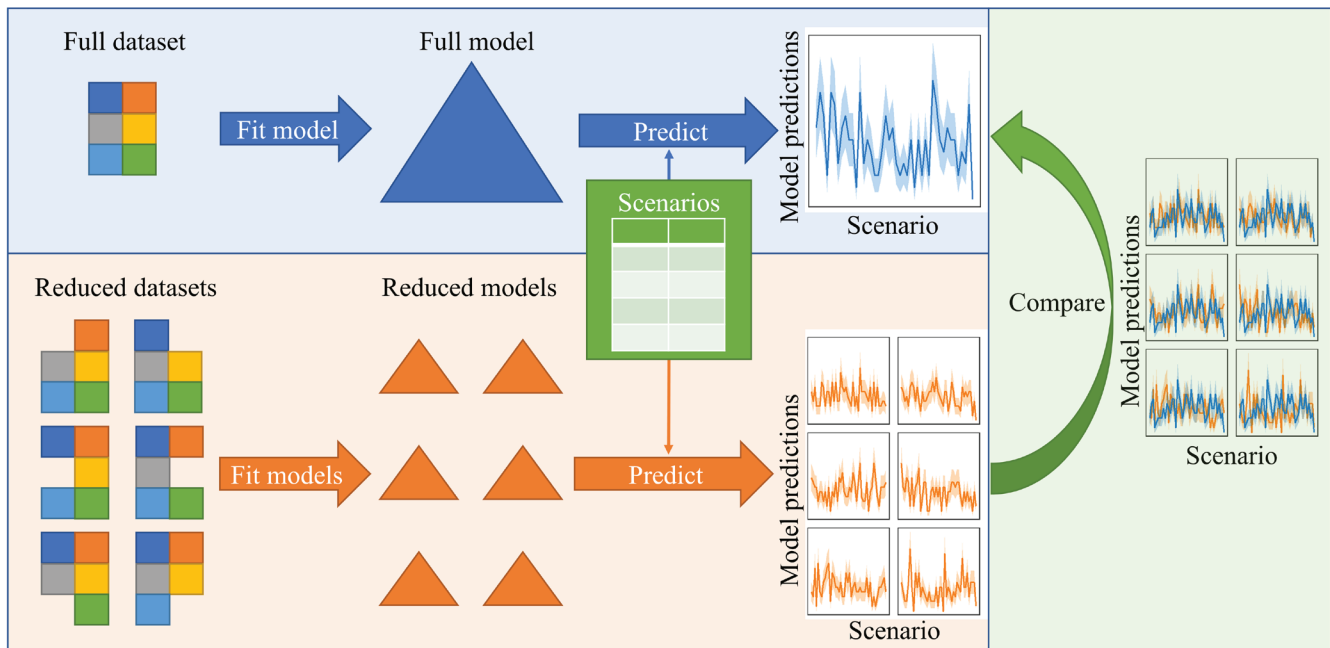
Figure 1 Conceptual model of the simulation framework. The full data set is used to fit a model which is then used to generate model predictions across a range of scenarios. The scenarios in this example were a range of dates, and the model predictions were expected fish abundance. However, the scenarios could also be a range of covariate values such as salinity to evaluate the ability to assess habitat relationships, and the model predictions could be other parameters of interest such as fish length or fecundity. The full data set is then reduced to simulate sampling reductions and generate reduced data sets. The same model structure as used for the full model is then fit on each reduced data set, and each reduced model is used to generate predictions across the same range of scenarios. Model predictions from the reduced models are then compared to those from the full model to examine how sampling reductions may affect inference from monitoring data.

data set. By incorporating prior data, one can more accurately simulate how changes to the monitoring program's design may have influenced past inferences (e.g., abundance trends used to guide management decisions), on top of the measured dynamics of the system as they appear in the historical data record. In contrast, simulation approaches rely on many assumptions that must be grounded in a deep understanding of the population and community dynamics of the system (White and Bahlai 2021).

The approach starts with fitting a statistical model to the full data set for a given species and gear type. The model structure should mirror the objectives of the surveys and review. Next, the data set is split into multiple reduced data sets that represent scenarios of reduced sampling effort. The same model structure is then fit to each reduced data set. The full model and all reduced models are used to generate fitted values

(predictions) over a range of scenarios. Finally, the fitted values from each reduced model are compared to the full model to evaluate how the sampling reduction altered model inference (Figure 1). This approach assumes that the full model is closer to the "truth" than the reduced models. The model structure and scenarios could also be tailored to different time-scales. While the case study used this methodology to evaluate the ability to measure year-to-year trends in abundance, these methods could also be applied to long-term trends with a different model structure and set of scenarios.

Splittail Case Study
To demonstrate this approach, I analyzed Sacramento Splittail (*Pogonichthys macrolepidotus*) otter trawl catch from the Bay Study and Suisun Study (Figure 2). The Bay Study and Suisun Study both collect monthly samples with an otter trawl at a set of fixed stations (Figure 3) (Matern et al.
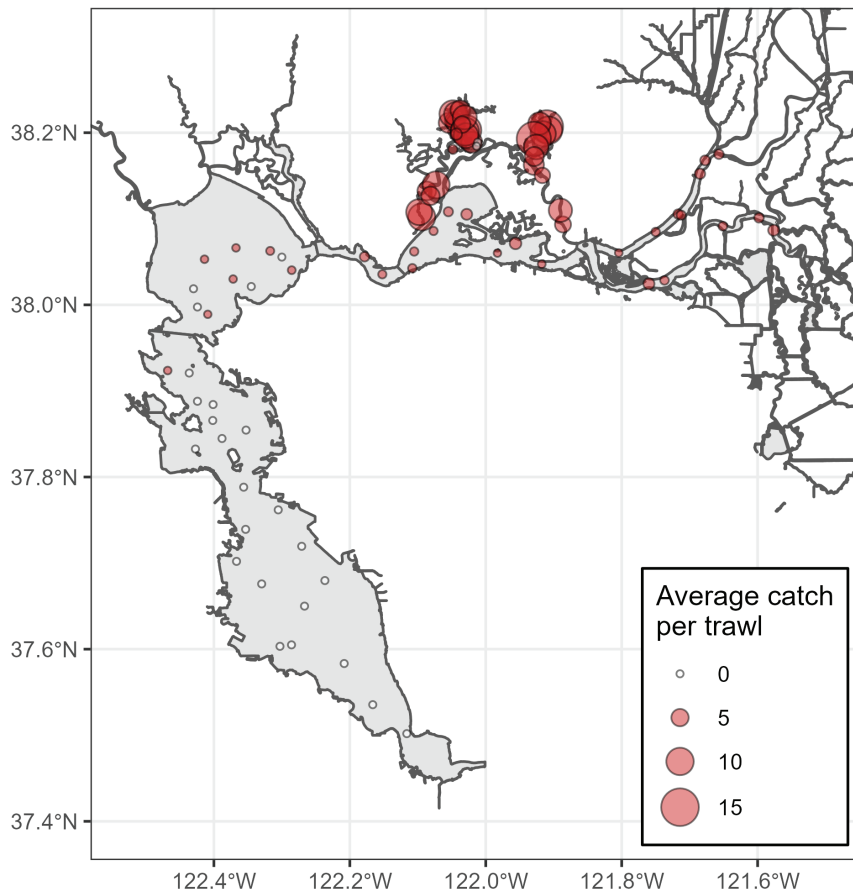
Figure 2 Spatial distribution of Splittail at Bay Study and Suisun Study sampling stations. *Point size* represents average catch per otter trawl at each sampling station, adjusted for the mean sampling effort across all trawls. *White-filled circles* represent stations with no otter trawl Splittail catch.
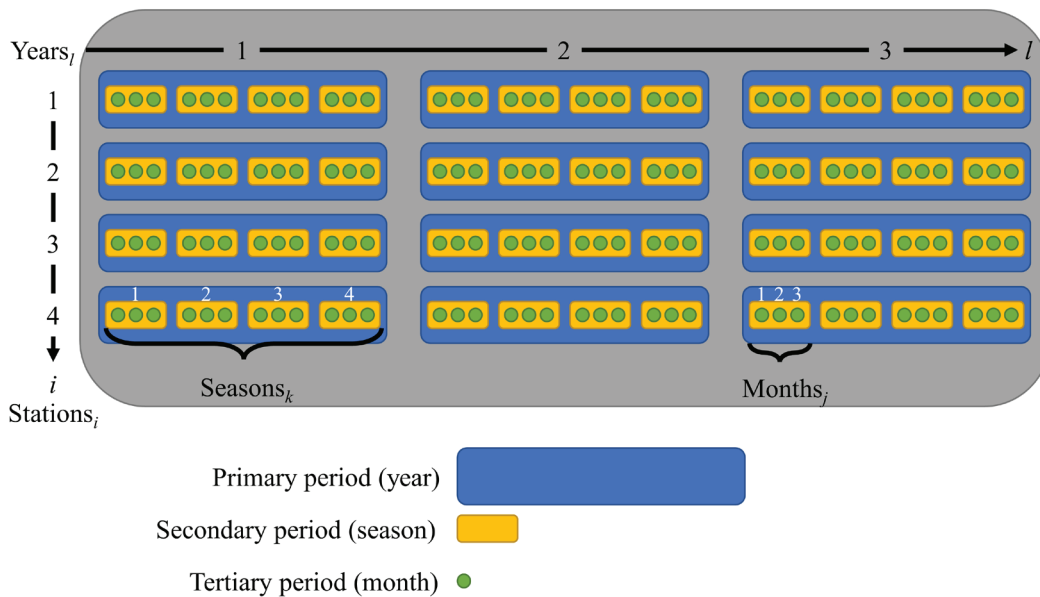
Figure 3 Structure of the data set used in the Splittail case study. Samples were collected monthly at a set of fixed stations. The generalized linear mixed model used in this case study (Equation 1) considered one spatial dimension (stations [i]) and three temporal dimensions (months [j], seasons [k], and years [l]).

2002; CDFW 2021; O'Rear et al. 2021). A complete description of the surveys, their sampling procedures, regulatory mandates, and changes over time can be found in Gaeta (2021). Splittail were identified as a representative species of a group also including Starry Flounder (*Platichthys stellatus*), Tule Perch (*Hysterocarpus traskii*), and Striped Bass (*Morone saxatilis*) from the otter trawl data used in this study (Gaeta and Beakes 2021) and they are a species of management concern in the estuary (Moyle et al. 2004; Sommer, Baxter, et al. 2007). Splittail are most abundant in brackish tidal sloughs, such as Suisun Marsh, but are commonly found in salinities from freshwater up to 18 ppt (Moyle et al. 2004). Splittail are strongly associated with seasonally inundated floodplains, where they spawn and the resulting juveniles rear (Sommer, Baxter, et al. 2007).

Splittail data were integrated as described in Bashevkin et al. (2021), and the integrated data set can be found in the package LTMRdata for the R statistical programming language (Bashevkin, Gaeta, Nguyen, et al. 2022a, 2022b). Briefly, data sets were obtained from the survey principal investigators and combined with their consultation. Fish smaller than 40 mm fork length were removed from the data set since they were not counted in all samples, and then counts were summed for all remaining fish lengths for each sample to obtain the total Splittail catch. This catch was rounded to the nearest integer value since some counts represented estimated values from sub-sampling that resulted in non-integers. I then filtered the resulting data set to remove years before 1985, when the Suisun Study sampling was less consistent and did not encompass all seasons.

It is important to note that the data (catch) used in this case study are a metric of relative abundance, but not an estimate of true abundance, because of unresolved catchability issues. While some elements of the model used in this study—such as the station random intercept or the coefficient for the fixed effect of tow area (see Model Structure, below)—may partially account for catchability, I do not attempt to fully resolve catchability issues in my modeling framework because of the lack of available data. Therefore, biases related to variable catchability may influence the results I present and the results of any further studies that use these methods. However, with the appropriate catchability parameters, these considerations could be incorporated within the model structures used in this framework. Catchability issues are discussed in more detail in Huntsman and Mahardja (2021).

MATERIALS AND METHODS

Model Structure

One important motivation for these sampling programs is describing trends in species abundances over time (Gaeta 2021). To evaluate our ability to identify these trends on a time-scale of management relevance, I focused the case study on year-to-year trends. The model structure was designed to fit this objective. I fit Bayesian generalized linear mixed models with a Poisson error distribution. Models were fit in the statistical programming language R v4.1.2 (R Core Team 2021) with the package brms (Bürkner 2017, 2018), which uses the probabilistic programming language Stan (Carpenter et al. 2017). The response variable was catch, with fixed predictors for tow area (sampling effort for otter trawl), season coded as a factor, year coded as a factor, and the interaction between season and year. I also tested a model structure accounting for sampling effort with an offset of the log of tow area, but it was equivalent to the original model structure by exact ten-fold cross-validation using the 'kfold' function in the R package loo (Vehtari et al. 2017), so I used the original model structure. To capture the fluctuating year-to-year population levels, I coded season and year as factors to enable the model to estimate different values for each unique season and year. I included random intercepts for each station and sample (i.e., each unique monthly tow). I included the sample random intercept as an observation-level random effect to deal with over-dispersion (Harrison 2014). The model that included the sample random intercept was superior to a Poisson model without the sample random intercept, and also superior to a negative binomial model without the sample random intercept by exact ten-fold cross-validation. The model formula was:

$$Count_{ijkl} \sim Poisson\,(\lambda_{ijkl})\,, \qquad \textbf{Eq 1}$$

$$log(\lambda_{ijkl}) = \beta_1 + \beta_2 Tow_{area} + \beta_{3_i} + \beta_{4_k} + \beta_{5_{kl}} + \alpha_i + \alpha_{ijkl} + \varepsilon_{ijkl}$$

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

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

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

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

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

where λ represents the mean and variance of the Poisson distribution, β represents estimated coefficients, α represents varying intercepts, σ represents variance, ε represents the residual error, i represents station, j represents month, k represents season, and l represents year (Figure 3). As mentioned above, since this model does not adjust for catchability issues, it is only modeling catch, not the latent state of true abundance (Huntsman and Mahardja 2021). I used weakly informative priors as recommended by the package authors (Stan Development Team 2021).

Models were run for three chains, each with 5,000 iterations, 1,250 of which were used for the warm-up and discarded. All models were inspected to ensure adequate sampling by verifying the posterior effective sample size (>100 per chain) and Rhat values (<1.05) (McElreath 2015). The full model was further validated by inspecting the trace plots and ensuring model predictions fit the raw data. Lastly, temporal autocorrelation was inspected in the residuals by calculating partial autocorrelations for each sampling station with >10 observations using the "pacf" function in the *stats* package. For a lag of 1 month, 56% of stations exhibited temporal autocorrelation, and for 2 months of lag that dropped to 44%, for 3 months of lag to 28%, and for 4 months of lag to 12%. Since I was very limited in computational time given the data-removal simulations (see Data Reductions, below), I decided not to incorporate an autocorrelation parameter in the models. The uncertainty from the models is likely underestimated because of this autocorrelation,

but the primary objective of this exercise was not an accurate assessment of Splittail populations with appropriate uncertainty. Since my methods are evaluating relative changes in accuracy (see Evaluation of the Effects of Data Reductions, below), I assume that the error caused by autocorrelation remains consistent across the models.

Data Reductions

To assess the effects of reduced sampling effort, I used a combination of systematic and random data reductions. These data reductions were designed to produce meaningful scenarios of sampling reductions without requiring unrealistic computational power. Reduced sampling effort was modeled as reductions in either temporal (number of monthly samples per season) or spatial (number of sampling stations) effort. For the temporal reductions, seasons were defined as follows: Winter = December (of prior year) through February; Spring = March through May; Summer = June through August; and Fall = September through November. Temporal reductions were simulated as the removal of 1 or 2 months per season, corresponding to 1/3 or 2/3 reductions in effort. To ensure sampling months were still regularly spaced and to reduce the required computational time, these reductions were completed systematically by removing the first, second, and/or third month of each season (Figure 4). Spatial reductions were simulated by removing 1/10, 1/5, 1/3, 1/2, or 2/3 of sampling stations. Stations were randomly divided into n groups for each 1/n (n = 10, 5, 3, or 2) cut to sampling stations, then *n* reduced data sets were created with one of those groups removed in each data set (Figure 4). For the two-thirds cut, I randomly split the stations into three groups with one of three station groups present in each of three reduced data sets. Since the 86 stations could not always be evenly divided among groups, the group sizes differed by up to one station in some spatial-reduction scenarios. In total, this resulted in 29 reduced data sets (6 temporal and 23 spatial), each used to fit a reduced model. The full data set had 21,323 observations of Splittail catch, and the smallest reduced model had 6,673 observations.
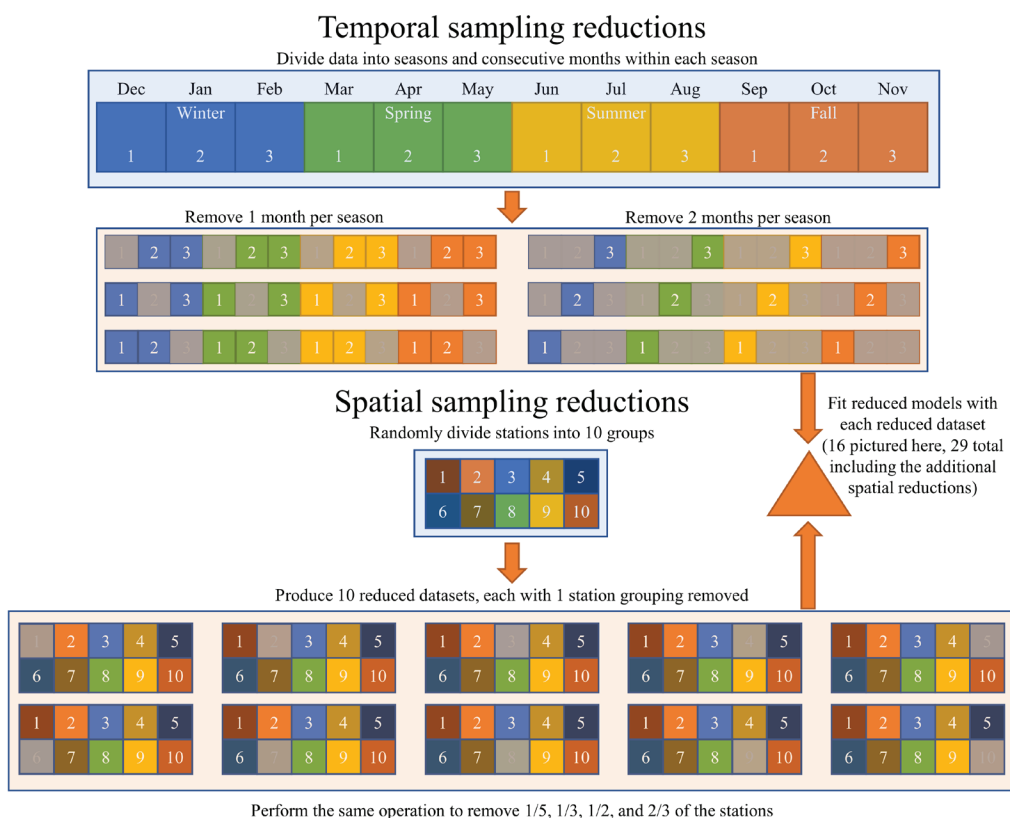
Figure 4 Specific sampling reductions implemented for the Splittail case study. Temporal sampling effort was reduced by removing 1 or 2 months of sampling per season, while spatial sampling effort was reduced by removing 1/10, 1/5, 1/3, 1/2, or 2/3 of sampling stations. Temporal reductions were systematic and based upon the 1st, 2nd, or 3rd month of each season. Spatial reductions were random by randomly assigned station "groups," separately for each spatial cut (1/10, 1/5, 1/3, 1/2, or 2/3). The two-thirds spatial cut was replicated three times, with one of three station groups present in each reduced data set (as opposed to the remainder of the spatial cuts in which 1 of *n* station groups were removed in each reduced data set). *Gray boxes* represent portions of the full data set removed to create each reduced data set.

Evaluation of the Effects of Data Reductions

I focused the evaluations of sampling reductions on the ability of the models to describe trends in abundance over time. Posterior calculations described below were completed with the *tidybayes* R package (Kay 2020). Each model was used to calculate fitted values (expected abundance values) for each season from 1985 to 2018, retaining the full posterior distribution (11,250 post-warm-up model iterations) for each time point (Figures 5, 6A). These fitted values were calculated with the *fitted* method from the *brms* package as:

$$e^{(\beta_1 + \beta_2 Tow_{area} + \beta_{3l} + \beta_{4k} + \beta_{5kl} + \alpha_i + \alpha_{ijkl})}$$ **Eq 2**

with Tow_{area}, α_i, and α_{ijkl} set to 0 (the mean for each variable/parameter set). See Equation 1 for a definition of the model terms.

Since the focused objective was on trends rather than abundance, I then calculated year-to-year expected abundance trends within each season. For each posterior draw (one of the 11,250 iterations), and within each season, I subtracted the expected abundance in the prior year from the expected abundance in the current year (e.g., Winter 1986 draw #1 – Winter 1985 draw #1). I then divided this difference by the sum of both expected abundances to standardize by the temporally local magnitude and obtain the local trend (Figures 5, 6B). The full formula is:
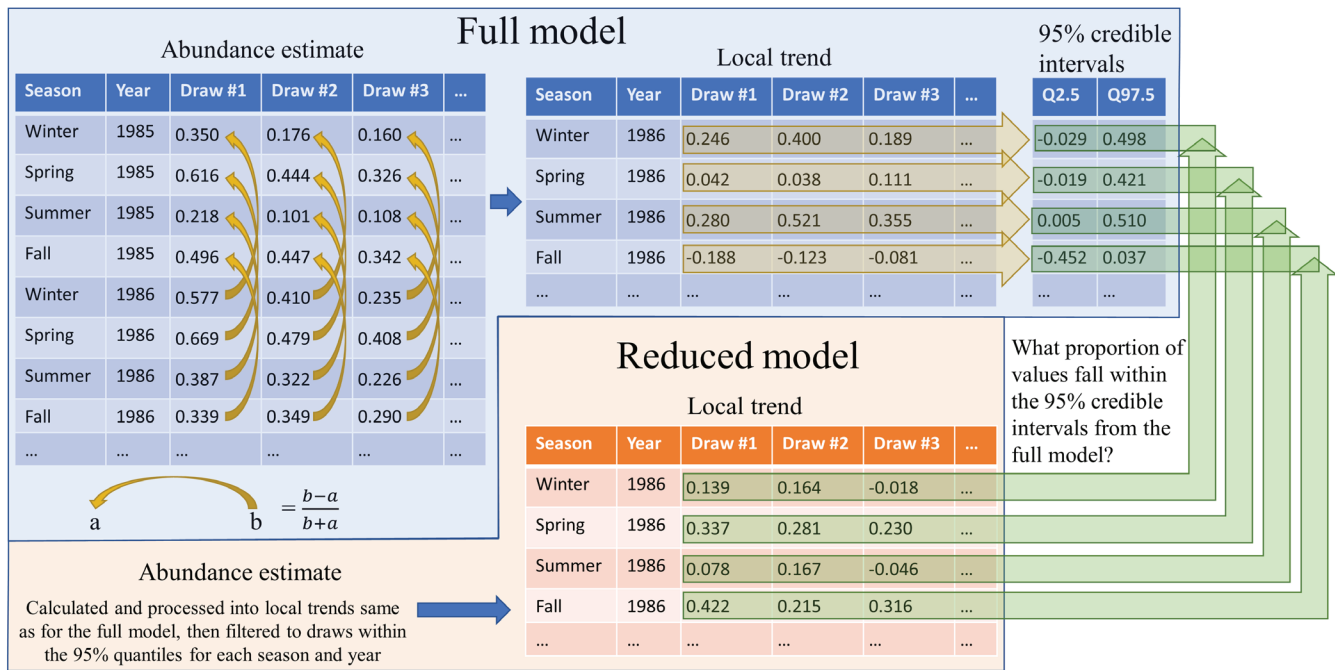
Figure 5 Posterior processing workflow illustrated with example data. The full model and each reduced model were used to generate predictions for each year and season from 1985 to 2018 (the "scenarios" in Figure 1). "Draws" represent posterior samples. The full posterior distribution of those predictions was retained, represented by posterior draws 1–3 (of 11,250 total) in this figure. Within-season trends were then calculated for each posterior draw using Equation 3. Ninety-five percent credible intervals were calculated for the local trend estimates of the full model for each season and year. The proportion of 95% quantile local trend values from the reduced models falling within those intervals were then calculated for the final metric.

$$LT_{s,y,d} = \frac{A_{s,y,d} - A_{s,y-1,d}}{A_{s,y,d} + A_{s,y-1,d}}$$ **Eq 3**

where $LT_{s,y,d}$ = local trend and $A_{s,y,d}$ = expected abundance value for each season (s), year (y), and draw (d). Standardizing by the local magnitude ensured consistency in trend estimates, regardless of the direction of the trend or the magnitude of the expected abundance estimate (e.g., a change from 90 to 10 would be equivalent in magnitude to a change from 10 to 90 and a change from 900 to 100). To compare estimates of the local trend between the full and reduced models, I calculated the overlap between the 95% credible intervals of each reduced model with that of the full model. To do this, I selected the posterior local trend estimates from the reduced model that fell within their 95% quantiles, and calculated the proportion of those values that fell within the 95% credible intervals from the full model, for each season and year (Figure 5). I used the 95% quantiles from both the reduced and full models to ensure that complete overlap

would equate to a proportional overlap value of 1. The proportional overlap was averaged across replicate simulations and across years to create an overall metric of reduced model overlap with the full model. This metric of proportional overlap was used because it captures differences in both certainty (precision) and value (accuracy) between the full and reduced models. Precision and accuracy are both critical determinants of the usefulness of monitoring data.

Variance Analysis

To understand the relative contributions of spatial and temporal factors to the variance in Splittail catch, I fit another Bayesian generalized linear mixed model with a Poisson error distribution. As before, the response variable was catch, and I included a fixed predictor for tow area and a sample-level random intercept to correct for over-dispersion. I also included random intercepts for year, month, and station to estimate the amount of variance described by those three factors.
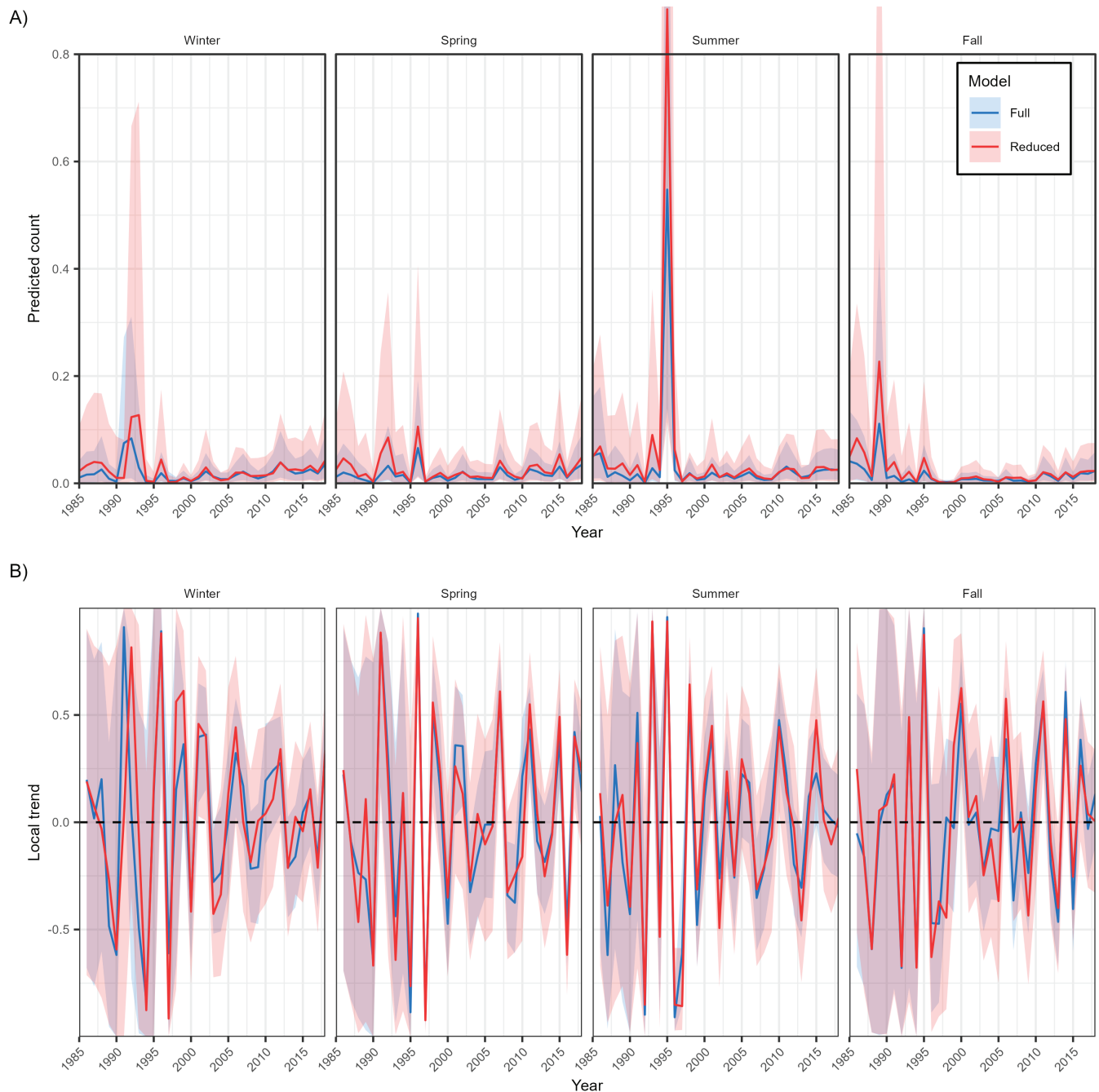
Figure 6   Example model outputs and local trend estimates for the full model and one reduced model (1/2 of sampling stations randomly removed). Point estimates of predicted count were generated for each year and season from 1985 to 2018. The local trend was calculated from those point estimates for each year and season from 1986 to 2018 (the local trend estimate requires a prior-year predicted count so it could not be calculated for 1985). (**A**) Mean fitted values with 95% credible intervals that represent the expected number of fish caught per trawl for the average sampling effort (tow area). The *y*-axis was clipped at 0.8 to prioritize visualizing the majority of values that were < 0.1. (**B**) Mean local trend estimates (calculated with Equation 3) with 95% credible intervals.

Models were fit and evaluated for adequate sampling as described above. I then compared the parameter estimates for the variances of the random intercepts to evaluate their relative influences on the variability in Splittail catch.

Code Availability

R code used in these analyses is available at *https://github.com/sbashevkin/LTMRpilot/tree/ publication/Univariate%20analyses*.

RESULTS

The sampling-reduction scenarios assessed our ability to infer within-season year-to-year trends in Splittail abundance from the Bay Study and Suisun Study otter trawls. These results are not applicable to other species, gears, surveys, or parameters of interest. Throughout this section, I will use "overlap" to refer to the percentage or proportion of 95% quantile local trend estimates from reduced models that were within the 95% credible interval of local trend estimates from the full model.

The percent overlap of reduced model predictions with the full model predictions was fairly consistent across replicate data-reduction simulations. As sampling effort was reduced, the interannual variability in percent overlap tended to increase in the majority of simulations, often leading to larger divergences among replicate simulations (Figure 7). For example, in the 10% sampling station reduction, the simulations had consistently high percent overlap, with only 2% of values below 75% overlap. In contrast, trend estimates from the 2/3 sample station reduction data set covered the full range from 2.88% to 99.6% overlap, with replicates changing asynchronously year to year (Figure 7). Percent overlaps were especially high in the earliest years, before 1995. This was most notable in the station reductions of 10% to 50%, but can also be seen in the monthly reduction of 33% (Figure 7).

Overall, the percent overlap of reduced models decreased linearly with the reduction in sampling effort (Figure 8). Removing 10% of sampling stations (spatial sampling effort) had

a small effect on the percent overlap of model predictions; overlap was ≥90% for 92% of time points (Figure 7). However, some replicates of this scenario had a few instances of low percent overlap, especially in earlier years. Even some of the 33% reductions in sampling effort had low effects on the percent overlap of model predictions; overlap was ≥75% for 85% of time points (Figure 7). While there were some outlier simulations at lower reductions, the variability in model prediction percent overlap among years and replicates became much higher at sampling reductions of 33% and above (Figure 7). The overall averaged overlap declined to a low of 61% for simulations of 67% reduced sampling effort. Interestingly, for similar reductions in sampling effort, models with temporally reduced effort had very similar percent overlap to models with spatially reduced effort (Figure 8).

While the reduced sampling effort simulations did not detect a difference between temporal and spatial reductions, the variance component analysis revealed wide differences. In this analysis, the spatial component (represented by station) contributed much more to the variability than the temporal components (represented by year and month). The variance parameter estimated for station was 7.5x greater than that for year, 11.8x greater than that for month, and 3.1x greater than that for the sample-level intercept (Figure 9).

DISCUSSION

I designed and demonstrated a simulation-based approach for evaluating the sensitivity of monitoring programs to changes in the sampling design. The approach worked well for a case study on Splittail sampled by the Bay Study and Suisun Study otter trawls. Below, I will briefly discuss the case study results, then the general approach and further extensions to apply it to the estuary monitoring program

Case Study

The case study evaluated the effects of reduced sampling on our ability to detect accurate trends in Splittail abundance from Bay Study and Suisun
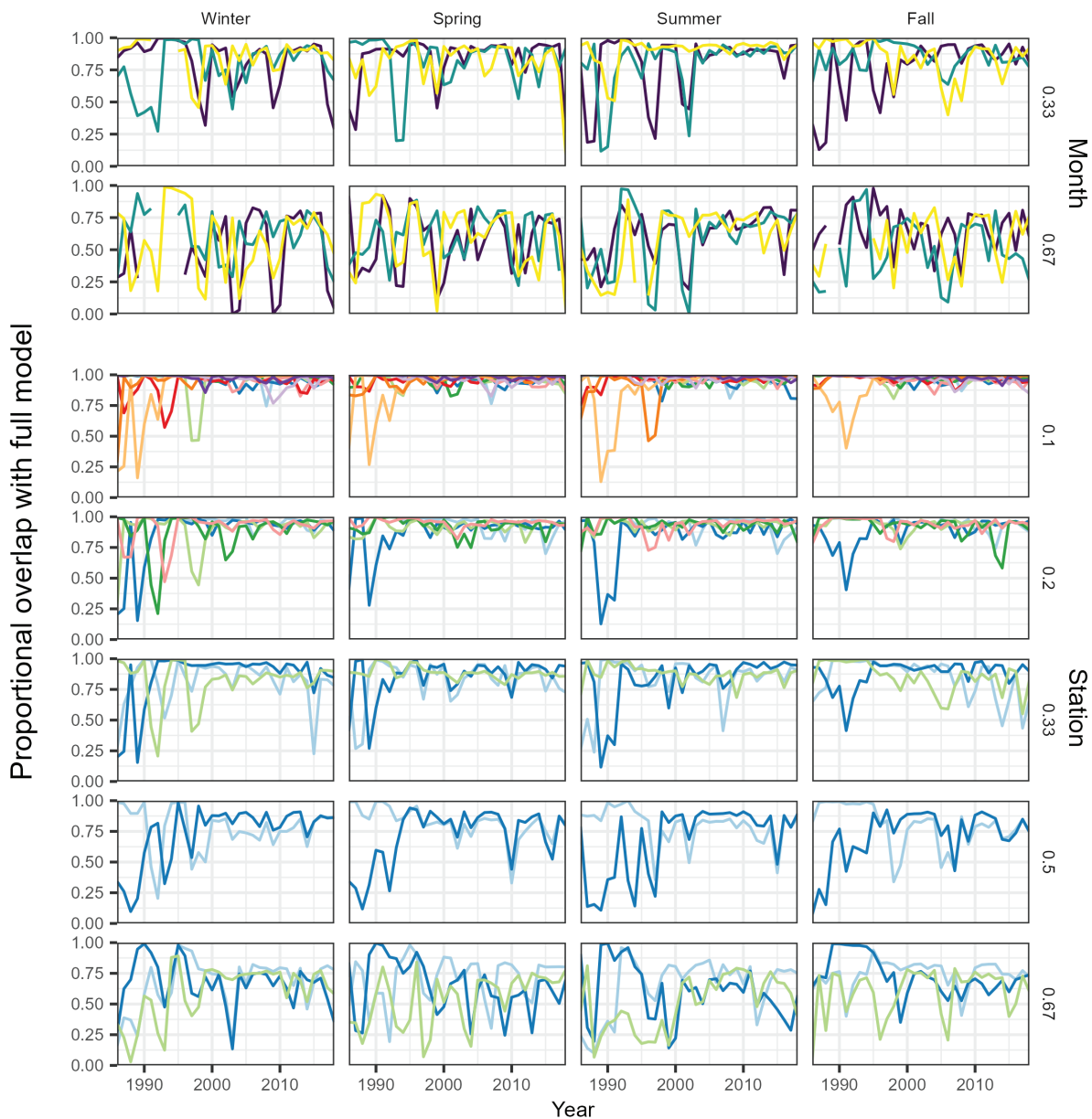
Figure 7   Proportional overlap in local trend estimates between the full model and each of the 29 reduced models. Overlap was quantified as the proportion of 95% quantile posterior values from each reduced model that fell within the 95% credible intervals of the full model. Values were calculated for each year and season from 1986 to 2018. Numbers on the right side of plots indicate the proportion of sampling reduced, with the source of that reduction (removed stations or months) noted on the far right. *Differently colored lines* indicate replicate simulations, but colors are only meaningful within a row of plots. Some values are missing in the month reductions, and these represent seasons not sampled after data from those months were removed.
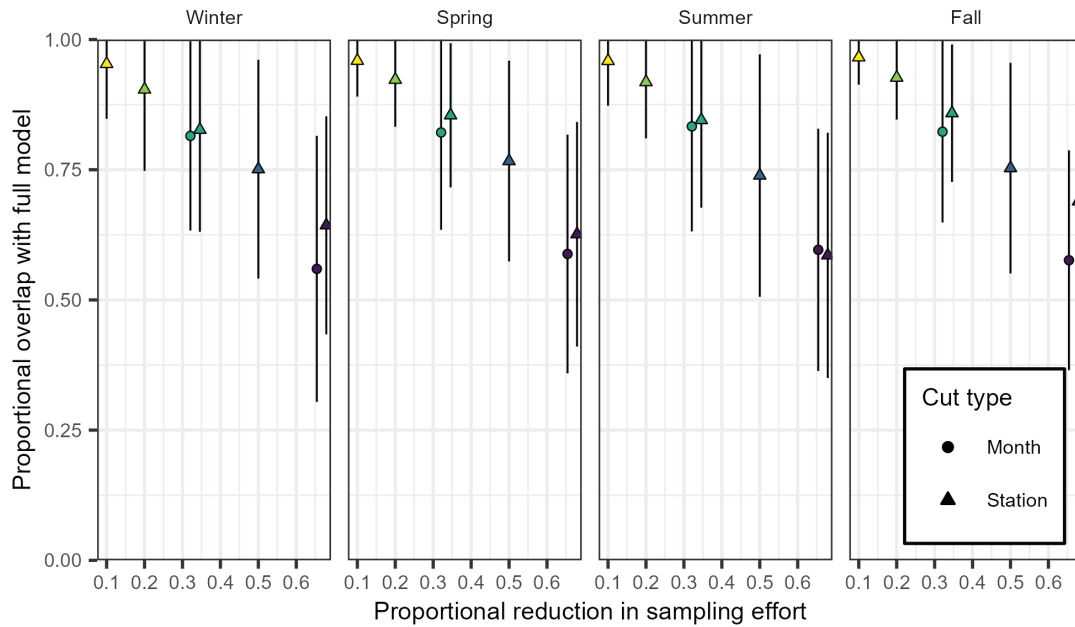
Figure 8 Average proportional overlap (± SD) for each season and sampling reduction scenario. These data represent aggregations of the results in Figure 7 to facilitate the inspection of broad patterns. *Colors* correspond to the proportional reduction in sampling effort (also shown on the *x*-axis) and shapes delineate the source of that reduction (removed months or stations). Points are slightly shifted horizontally to facilitate visualization.
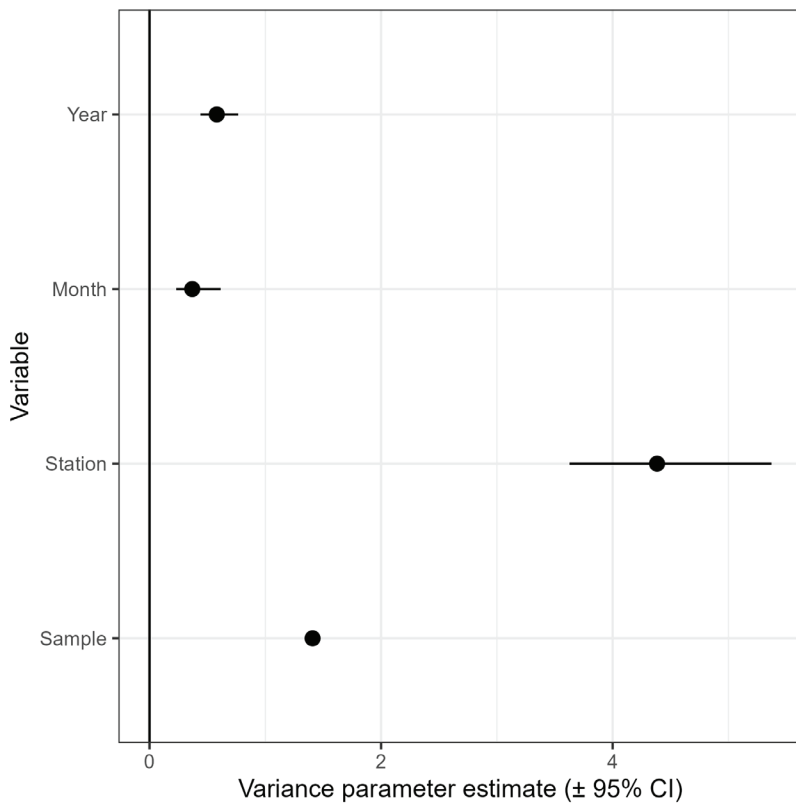
Figure 9 Variance parameter estimates from the variance analysis. This estimates the relative contributions of each variable to the variability in the data. Variance parameters were estimated from a Bayesian generalized linear mixed model with random intercepts for each variable.

Study otter trawl data. I found that relatively low (10% to 20%) reductions in spatial sampling effort had fairly small effects on the percent overlap of reduced model predictions with the full model predictions. Greater reductions in temporal or spatial sampling effort reduced the percent overlap of model predictions, as well as the variability in this overlap among years and among replicate data-reduction scenarios.

These results indicate that sampling effort reductions at a level of 10% to 20% may not have a large effect on our ability to monitor Splittail abundance trends with Bay Study and Suisun Study otter trawl data, within the spatial footprints of those surveys. However, given the variability among replicate simulations, the choice of stations to remove appears to be important. These results are not applicable to other species, gears, surveys, or parameters of interest. Within the footprints of the Bay Study and Suisun Study, Splittail are concentrated in the Suisun region, with very low to zero catch throughout much of San Francisco Bay (Figure 2). This may have contributed to the negligible effects of removing some stations, since the removal of stations with consistently low catch should have very little effect on model inference of trends in Splittail abundance. This may also explain the results of the variance analysis, in which the spatial component contributed much more to the variability in the data than the temporal or sample-level components. Consistently high catches in some stations (Suisun Marsh) and consistently low catches in other stations (San Francisco Bay) would have driven large differences in the individual random intercepts for stations in those two regions, and thus the large variance associated with the distribution of that random intercept. I chose to retain the full set of stations from each study despite the spatial concentration of Splittail catch because this would be necessary for a comprehensive monitoring evaluation that encompassed additional species which might be concentrated in different geographic areas (see below).

In some of the simulations, there was a pattern of decreasing percent overlap over time. This was apparent in the sampling station reductions of 10% to 50%, and to some extent in the monthly reduction of 33%. A likely cause of this pattern is the increase in number of sampling stations over time, which could have increased redundancy in the sampling program in later years. Of the 86 total stations among the two surveys, 53 were sampled before 1995 and 85 were sampled in 1995 and later.

The Approach

The framework I describe was based on earlier work by White (2019) and White and Bahlai (2021). They describe the approach and benefits of experimenting with historical data to inform experimental design, and how this method compares with power simulations, field experiments, and comparative analyses (White and Bahlai 2021). This approach has been applied to identify the number of years of monitoring data needed to (1) detect population trends (White 2019; Bahlai et al. 2021; Cusser et al. 2021), (2) quantify the reliability of population trends (Wauchope et al. 2019), and (3) design optimal sampling to detect ecosystem shifts (Bruel and White 2020).

My simulation-based framework for evaluating survey designs was effective for the case study, but it has both strengths and weaknesses that should be considered before application. The ability to simulate sampling design changes using historical data on top of the past variability of the system is a major strength that grounds the results in an accurate representation of the environment. It allows us to examine how reduced sampling may have affected our ability to understand the system during droughts, climate cycles such as the El Niño–Southern Oscillation, or historical phenomena such as the pelagic organism decline (Sommer, Armor, et al. 2007). However, this advantage is also a limitation of retrospective analyses. By experimenting with the past, I was unable to evaluate where additional sampling might be needed. There may have been historical trends undetected by my methods because the historical sampling needed expansion

in ways these methods are unable to detect. I was also limited in my ability to evaluate the effects of emerging issues not present in the historical data set, such as further climate change and emerging contaminant issues. For this reason, the framework presented here should ideally be paired with a complementary analysis of sampling gaps.

I chose to base my analyses on statistical models of management-relevant parameters to ensure the results would be most useful. This grounded my evaluation of the sampling program in the same methods used to analyze the resulting data, which is a recommended approach in the design of monitoring programs (Radinger et al. 2019).

By using Bayesian models, I was able to easily propagate uncertainty through the calculation of the local trend metric. This enabled comparisons of both point estimates and uncertainty magnitudes to capture the full effect of sampling design changes. It is also especially useful for analyses of irregular monitoring data such as the data I used, with its many changes to sampling sites (Radinger et al. 2019). However, Bayesian modeling (and any hierarchical modeling of large data sets) comes with a high computational cost, which can slow down project progress, increase monetary costs, or limit the number of simulations that can be performed. Nevertheless, the Stan modeling language is generally faster and more efficient than other Bayesian modeling languages such as BUGS or JAGS (Carpenter et al. 2017), and recent improvements—including within-chain parallelization—have enabled even greater speed-ups (Stan Development Team 2021).

Extensions and Applications

My case study demonstrates the utility of the general approach, but much more must be done for a thorough analysis leading to changes in sampling effort. Starting with the randomized and stratified removal of sampling effort, as I have done, is an important first step to identify possible redundancies across multiple surveys. Next, more targeted analyses should identify precisely which months and stations contribute the least useful information (Figure 10). The same

overall framework (Figure 1) would be used for these targeted analyses, but specific stations or months would be removed in simulations to evaluate the effects of new monitoring designs. The specific stations or months to be removed in the simulations could be chosen with a clustering approach as in Gaeta and Beakes (2021), based on geographic distance or temporal frequency, based on logistical concerns, or randomly. My framework would then be used to evaluate the effects of each proposed monitoring design on the accuracy of model inference, on top of historical conditions.

An important first step to monitoring program design is a collaborative discussion between scientists and managers to delineate the objectives of the monitoring program (NRC 1990). In particular, before redundancies can be identified, accuracy targets must be defined. Optimal accuracy will depend on the management questions as well as on practical considerations of sampling cost and complexity. Accuracy targets should be chosen in consultation with the monitoring staff, managers, and statisticians. This step is critical to define the goals of the monitoring program and ensure a tight coupling between these goals and the design of the review, so that the results directly inform improving the monitoring program's ability to achieve its goals.

The estuary monitoring program is composed of multiple intertwined sampling modalities, objectives, and mandates (Gaeta 2021; Tempel et al. 2021). The boat-based surveys monitor demersal, pelagic, and littoral fish communities; zooplankton; phytoplankton; benthic invertebrates; water quality; and contaminants. All of these sampling modalities are important and have been leveraged to better understand the estuary and improve management (e.g., Jassby et al. 1995; Cloern et al. 2007; Feyrer et al. 2007; Sommer, Armor, et al. 2007; Winder et al. 2011; Winder and Jassby 2011; Munsch et al. 2019). They are intertwined because multiple parameters are often collected by each survey (Tempel et al. 2021) and even the non-focal parameters can be important data sources for understanding the
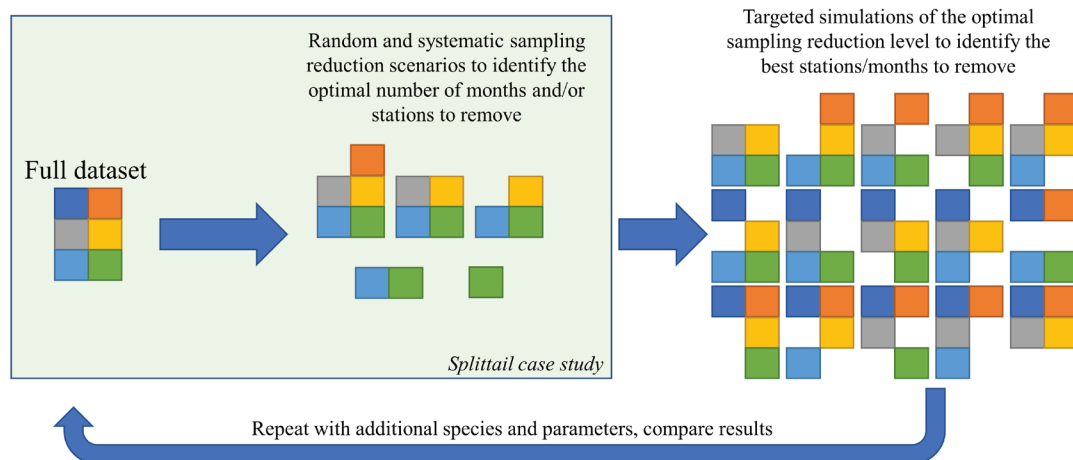
Figure 10 Framework for a thorough analysis of sampling reductions. First, the optimal number of sampling stations or months to be removed should be identified with random and stratified sampling reductions as in the Splittail case study (Figure 4). After the optimal level of sampling reduction has been identified, that level of reduction should be applied in a targeted manner to the full data set by removing as many combinations of months/stations as are computationally feasible to narrow down the best candidates for removal. The entire process should then be repeated for additional species and parameters (e.g., distribution, size, or habitat associations). The final results from each species and parameter must then be compared to identify how sampling effort could be reduced with the least effect on our understanding of the system.

estuary (Bashevkin, Mahardja, and Brown 2022; Bashevkin and Mahardja 2022). Each survey measures water quality in some manner (although variables like nutrient concentrations and chlorophyll are more limited), zooplankton are often collected alongside fish samples (Bashevkin, Hartman, et al. 2022), and many of the sampling surveys share boats and crew. Thus, changes to one survey or set of surveys based on a single parameter have the potential to interfere with critical long-term monitoring data sets of other parameters equally important for estuarine management. Therefore, before changes are made to the monitoring program or any individual survey, simulations as described here must be performed for each affected parameter. Monitoring program changes based on optimizing the sampling of a narrowly-focused parameter like Splittail abundance could greatly impede our ability to monitor other fishes such as Longfin Smelt, or our valuable zooplankton monitoring program (Hartman et al. 2021).

To evaluate the ability of these surveys to monitor the whole fish community, these analyses must be conducted on multiple species. Other species with

different geographic distributions, seasonal or annual abundance trends, life-history strategies, or depth preferences would likely show different effects of reduced sampling. Representative or indicator species are commonly used in conservation to manage ecosystems without requiring analyses on every species (Poiani et al. 2000). These representative species could be selected with clustering approaches as in Gaeta and Beakes (2021). Results must then be compared across species to identify sampling stations or months of minimal value across all representative species (Figure 10).

While this case study focused on our ability to detect trends in species abundance, other parameters should also be examined for a thorough evaluation. Models more focused on geographic distribution, fish size, fecundity, and the effects of environmental variables on any of these should also be considered (Figure 10). For example, to evaluate the ability of the monitoring surveys to detect the relationship between salinity and fish abundance, a model of fish abundance by salinity would be constructed, and the ability of reduced data sets to reproduce the relationship

obtained from a full model would be evaluated. These parameters are key for advancing our mechanistic understanding of the system, which can be leveraged to solve ecological problems (Radinger et al. 2019). The focal parameters should be chosen with consideration for the goals of the monitoring program, as well as the management objectives for the system.

This framework can easily be extended to other types of monitoring beyond fishes. The basic approach requires fitting the same model structure to the full data set and a set of reduced data sets that represent scenarios of reduced sampling effort, then comparing model outputs from the reduced models to the full model (Figure 1). This could be applied to other functional groups (e.g., zooplankton and phytoplankton) as well as physical parameters (e.g., temperature, salinity, and Secchi depth).

Once the sensitivity of all relevant parameters has been assessed, common redundancies can be identified for further action. This approach will minimize unexpected effects to important sampling modalities while identifying the best candidate options for increased efficiency. The long-term monitoring data set is a critically important resource for management and for furthering our scientific understanding of this estuary and aquatic systems more broadly (Callahan 1984; Hobbie et al. 2003; Lindenmayer and Likens 2010). Thus, any reductions in sampling effort must be undertaken carefully and with an informed understanding of the consequences for the monitoring program as a whole.

ACKNOWLEDGEMENTS

REFERENCES

Bahlai CA, White ER, Perrone JD, Cusser S, Stack Whitney K. 2021. The broken window: an algorithm for quantifying and characterizing misleading trajectories in ecological processes. Ecol Inform. 64:101336. *https://doi.org/10.1016/j.ecoinf.2021.101336*

Barry J, Maxwell D, Jennings S, Walker D, Murray J. 2017. *Emon*: an R-package to support the design of marine ecological and environmental studies, surveys and monitoring programmes. Methods Ecol Evol. 8(10):1342–1346. *https://doi.org/10.1111/2041-210X.12748*

Bashevkin SM, Beakes MP, Gaeta J, Mahardja B. 2021. Data integration. Chapter 3 In: Gaeta JW, Bashevkin SM, Feyrer F, Huntsman BM, Mahardja B, Culberson, SD, Beakes, MP, Fong S, Louie S. Interagency Ecological Program Long-term Monitoring Element Review: pilot approach and methods development. IEP Technical Report #96. [accessed 2022 May 11]. Sacramento (CA): California Department of Water Resources. p. 77–85. Available from: *https://cadwr.app.box.com/s/5y13dz3wg04t5abkprvjfj9lmwz29a96*

Bashevkin SM, Gaeta JW, Nguyen TX, Mitchell L, Khanna S. 2022a. LTMRdata: an R package to integrate data from 9 fish surveys in the San Francisco Estuary. v2.2.0. Zenodo. [accessed 2022 May 11]. *https://doi.org/10.5281/zenodo.6048977*

Bashevkin SM, Gaeta JW, Nguyen TX, Mitchell L, Khanna S. 2022b. Fish abundance in the San Francisco Estuary (1959–2021), an integration of 9 monitoring surveys. ver 1. Environmental Data Initiative. [accessed 2022 Feb 15]. Available from: *https://doi.org/10.6073/PASTA/0CDF7E5E954BE1798AB9BF4F23816E83*

Bashevkin SM, Hartman R, Thomas M, Barros A, Burdi CE, Hennessy A, Tempel T, Kayfetz K. 2022. Five decades (1972–2020) of zooplankton monitoring in the upper San Francisco Estuary. PLoS ONE. 17(3):e0265402. https://doi.org/10.1371/journal.pone.0265402

Bashevkin SM, Mahardja B. 2022. Seasonally variable relationships between surface water temperature and inflow in the upper San Francisco Estuary. Limnol Oceanogr. 67(3):684–702. https://doi.org/10.1002/lno.12027

Bashevkin SM, Mahardja B, Brown LR. 2022. Warming in the upper San Francisco Estuary: patterns of water temperature change from five decades of data. Limnol Oceanogr. 67(5):1065–1080. https://doi.org/10.1002/lno.12057

Bruel R, White ER. 2020. Sampling requirements and approaches to detect ecosystem shifts. Ecol Indic. 107096. https://doi.org/10.1016/j.ecolind.2020.107096

Bürkner P–C. 2017. brms: an R package for Bayesian multilevel models using Stan. J Stat Softw. 80(1):1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner P–C. 2018. Advanced Bayesian multilevel modeling with the R package brms. R J. 10(1):395–411. https://doi.org/10.32614/RJ-2018-017

Callahan JT. 1984. Long-term ecological research. BioScience. 34(6):363–367. https://doi.org/10.2307/1309727

Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A. 2017. Stan: a probabilistic programming language. J Stat Softw. 76(1):1–32. https://doi.org/10.18637/jss.v076.i01

[CDFW] California Department of Fish and Wildlife. 2021. Bay study data. [accessed 2022 May 11]. https://filelib.wildlife.ca.gov/Public/BayStudy/

Christie AP, Amano T, Martin PA, Shackelford GE, Simmons BI, Sutherland WJ. 2019. Simple study designs in ecology produce inaccurate estimates of biodiversity responses. J Applied Ecol. 56(12):2742–2754. https://doi.org/10.1111/1365-2664.13499

Cloern JE, Jassby AD, Thompson JK, Hieb KA. 2007. A cold phase of the East Pacific triggers new phytoplankton blooms in San Francisco Bay. PNAS. 104(47):18561–18565. https://doi.org/10.1073/pnas.0706151104

Cusser S, Helms IV J, Bahlai CA, Haddad NM. 2021. How long do population level field experiments need to be? utilising data from the 40-year-old LTER network. Ecol Lett. 24(5):1103–1111. https://doi.org/10.1111/ele.13710

Feyrer F, Nobriga ML, Sommer TR. 2007. Multidecadal trends for three declining fish species: habitat patterns and mechanisms in the San Francisco Estuary, California, USA. Can J Fish Aquat Sci. 64(4):723–734. https://doi.org/10.1139/f07-048

Gaeta J. 2021. Survey history, regulatory mandates, and deep metadata. Chapter 2 In: Gaeta JW, Bashevkin SM, Feyrer F, Huntsman BM, Mahardja B, Culberson, SD, Beakes, MP, Fong S, Louie S. Interagency Ecological Program Long-term Monitoring Element Review: pilot approach and methods development. IEP Technical Report #96. [accessed 2022 May 11]. Sacramento (CA): California Department of Water Resources. p. 24–76. Available from: https://cadwr.app.box.com/s/5y13dz3wg04t5abkprvjfj9lmwz29a96

Gaeta J, Beakes MP. 2021. Identifying sub-community assemblages and representative species. Chapter 5 In: Gaeta JW, Bashevkin SM, Feyrer F, Huntsman BM, Mahardja B, Culberson, SD, Beakes, MP, Fong S, Louie S. Interagency Ecological Program Long-term Monitoring Element Review: pilot approach and methods development. IEP Technical Report #96. [accessed 2022 May 11]. Sacramento (CA): California Department of Water Resources. p. 121–153. Available from: https://cadwr.app.box.com/s/5y13dz3wg04t5abkprvjfj9lmwz29a96

Gerrodette T. 1987. A power analysis for detecting trends. Ecology. 68(5):1364–1372. https://doi.org/10.2307/1939220

Harrison XA. 2014. Using observation-level random effects to model overdispersion in count data in ecology and evolution. PeerJ. 2:e616. https://doi.org/10.7717/peerj.616

Hartman RK, Bashevkin SM, Barros A, Burdi CE, Patel C, Sommer T. 2021. Food for thought: connecting zooplankton science to management in the San Francisco Estuary. San Franc Estuary Watershed Sci. [accessed 2021 Sep 28];19(3). https://doi.org/10.15447/sfews.2021v19iss3art1

Hobbie JE, Carpenter SR, Grimm NB, Gosz JR, Seastedt TR. 2003. The US long term ecological research program. BioScience. 53(1):21–32. https://doi.org/10.1641/0006-3568(2003)053[0021:TULTER]2.0.CO;2

Huntsman B, Mahardja B. 2021. Relative catchability bias among sampling gears and species. Chapter 4 In: Gaeta JW, Bashevkin SM, Feyrer F, Huntsman BM, Mahardja B, Culberson, SD, Beakes, MP, Fong S, Louie S. Interagency Ecological Program long-term monitoring element review: pilot approach and methods development. IEP Technical Report #96. [accessed 2022 May 11]. Sacramento (CA): California Department of Water Resources. p. 86–120. Available from: https://cadwr.app.box.com/s/5y13dz3wg04t5abkprvjfj9lmwz29a96

Huntsman BM, Mahardja B, Bashevkin SM. 2022. Relative bias in catch among long-term fish monitoring surveys within the San Francisco Estuary. San Franc Estuary Watershed Sci. [accessed 2022 Mar 11];20(1). https://doi.org/10.15447/sfews.2022v20iss1art3

[IEP SAG] Interagency Ecological Program Science Advisory Group. 2013. Review of the IEP Delta juvenile fishes monitoring program and Delta juvenile salmonid survival studies. Sacramento (CA): IEP SAG. 27 p.

Jassby AD, Kimmerer WJ, Monismith SG, Armor C, Cloern JE, Powell TM, Schubel JR, Vendlinski TJ. 1995. Isohaline position as a habitat indicator for estuarine populations. Ecol Appl. 5(1):272–289. https://doi.org/10.2307/1942069

Johnson RC, Windell S, Brandes PL, Conrad JL, Ferguson J, Goertler PAL, Harvey BN, Heublein J, Israel JA, Kratville DW, et al. 2017. Science advancements key to increasing management value of life stage monitoring networks for endangered Sacramento River winter-run Chinook Salmon in California. San Franc Estuary Watershed Sci. [accessed 2020 Jun 17];15(3). https://doi.org/10.15447/sfews.2017v15iss3art1

Kay M. 2020. tidybayes: tidy data and geoms for Bayesian models. Zenodo. Version 2.1.1. [accessed 2022 May 11]. https://doi.org/10.5281/zenodo.1308151

Lindenmayer DB, Likens GE. 2010. The science and application of ecological monitoring. Biol Conserv. 143(6):1317–1328. https://doi.org/10.1016/j.biocon.2010.02.013

Matern SA, Moyle PB, Pierce LC. 2002. Native and alien fishes in a California estuarine marsh: twenty-one years of changing assemblages. Trans Am Fisheries Soc. 131(5):797–816. https://doi.org/10.1577/1548-8659(2002)131<0797:NAAFIA>2.0.CO;2

McElreath R. 2015. Statistical rethinking: a Bayesian course with examples in R and Stan. Boca Raton (FL): CRC Press. 469 p.

Moyle PB, Baxter RD, Sommer T, Foin TC, Matern SA. 2004. Biology and population dynamics of Sacramento Splittail (*Pogonichthys macrolepidotus*) in the San Francisco Estuary: a review. San Franc Estuary Watershed Sci. [accessed 2020 Jul 17];2(2). https://doi.org/10.15447/sfews.2004v2iss2art3

Munsch SH, Greene CM, Johnson RC, Satterthwaite WH, Imaki H, Brandes PL. 2019. Warm, dry winters truncate timing and size distribution of seaward-migrating salmon across a large, regulated watershed. Ecol Appl. 29(4):e01880. https://doi.org/10.1002/eap.1880

[NRC] National Research Council. 1990. Managing troubled waters: the role of marine environmental monitoring. Washington (DC): The National Academies Press. [accessed 2021 Oct 11]. Available from: https://www.nap.edu/catalog/1439/managing-troubled-waters-the-role-of-marine-environmental-monitoring

O'Rear T, Durand J, Moyle P. 2021. UC Davis Suisun Marsh Fish Study. [accessed 2022 May 11]. Available from: https://watershed.ucdavis.edu/project/suisun-marsh-fish-study

Poiani KA, Richter BD, Anderson MG, Richter HE. 2000. Biodiversity conservation at multiple scales: functional sites, landscapes, and networks. BioScience. 50(2):133–146. https://doi.org/10.1641/0006-3568(2000)050[0133:BCAMSF]2.3.CO;2

R Core Team. 2021. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. [accessed 2022 May 11]. Available from: https://www.R-project.org/

Radinger J, Britton JR, Carlson SM, Magurran AE, Alcaraz–Hernández JD, Almodóvar A, Benejam L, Fernández–Delgado C, Nicola GG, Oliva–Paterna FJ, et al. 2019. Effective monitoring of freshwater fish. Fish Fish. 20(4):729–747. *https://doi.org/10.1111/faf.12373*

Reynolds JH, Knutson MG, Newman KB, Silverman ED, Thompson WL. 2016. A road map for designing and implementing a biological monitoring program. Environ Monit Assess. 188(7):399. *https://doi.org/10.1007/s10661-016-5397-x*

Rhodes JR, Jonzén N. 2011. Monitoring temporal trends in spatially structured populations: how should sampling effort be allocated between space and time? Ecography. 34(6):1040–1048. *https://doi.org/10.1111/j.1600-0587.2011.06370.x*

Sommer TR, Armor C, Baxter R, Breuer R, Brown L, Chotkowski M, Culberson S, Feyrer F, Gingras M, Herbold B, et al. 2007. The collapse of pelagic fishes in the upper San Francisco Estuary: el colapso de los peces pelagicos en la cabecera Del Estuario San Francisco. Fisheries. 32(6):270–277. *https://doi.org/10.1577/1548-8446(2007)32[270:TCOPFI]2.0.CO;2*

Sommer TR, Baxter RD, Feyrer F. 2007. Splittail "delisting": a review of recent population trends and restoration activities. Chapter 3 In: 2007. Brouder MJ, Scheurer JA, editors. Status, distribution, and conservation of freshwater fishes of western North America: a symposium proceedings. AFS Symposium 53. Bethesda (MD): American Fisheries Society. p. 25–38. *https://doi.org/10.47886/9781888569896.ch3*

Stan Development Team. 2021. Stan user's guide. Version 2.27. [accessed 2020 Oct 14]. Available from: *https://mc-stan.org/docs/2_24/stan-users-guide/index.html*

Tempel TL, Malinich TD, Burns J, Barros A, Burdi CE, Hobbs JA. 2021. The value of long-term monitoring of the San Francisco Estuary for Delta Smelt and Longfin Smelt. Sacramento (CA): California Department of Fish and Wildlife. California Fish and Wildlife Special CESA Issue. p. 148–171. Available from: *https://doi.org/10.51492/cfwj.cesasi.7*

Vehtari A, Gelman A, Gabry J. 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat Comput. 27(5):1413–1432. *https://doi.org/10.1007/s11222-016-9696-4*

Vos P, Meelis E, Ter Keurs WJ. 2000. A framework for the design of ecological monitoring programs as a tool for environmental and nature management. Environ Monit Assess. 61(3):317–344. *https://doi.org/10.1023/A:1006139412372*

Wauchope HS, Amano T, Sutherland WJ, Johnston A. 2019. When can we trust population trends? a method for quantifying the effects of sampling interval and duration. Methods Ecol Evol. 10(12):2067–2078. *https://doi.org/10.1111/2041-210X.13302*

Weiser EL, Diffendorfer JE, Grundel R, López–Hoffman L, Pecoraro S, Semmens D, Thogmartin WE. 2019. Balancing sampling intensity against spatial coverage for a community science monitoring programme. J Appl Ecol. 56(10):2252–2263. *https://doi.org/10.1111/1365-2664.13491*

White ER. 2019. Minimum time required to detect population trends: the need for long-term monitoring programs. BioScience. 69(1):40–46. *https://doi.org/10.1093/biosci/biy144*

White ER, Bahlai CA. 2021. Experimenting with the past to improve environmental monitoring. Front Ecol Evol. 8. [accessed 2021 Aug 10]. *https://doi.org/10.3389/fevo.2020.572979*

Winder M, Jassby AD. 2011. Shifts in zooplankton community structure: implications for food web processes in the upper San Francisco Estuary. Estuaries Coasts. 34(4):675–690. *https://doi.org/10.1007/s12237-010-9342-x*

Winder M, Jassby AD, Nally RM. 2011. Synergies between climate anomalies and hydrological modifications facilitate estuarine biotic invasions. Ecol Lett. 14(8):749–757. *https://doi.org/10.1111/j.1461-0248.2011.01635.x*