



Association of delta smelt with other small pelagic fishes suggests the potential for competition and proxy monitoring

Michael Tillotson*, Calvin Y. Lee, Crystal Garcia

ICF, 980 9th Street, Suite 1200, Sacramento, CA 95814, USA

ABSTRACT: Development of habitat occupancy models for protected species with low detection rates can be difficult; however a possible solution is to use proxy species. In the San Francisco Bay-Delta estuary, detection of the endemic and endangered *Hypomesus transpacificus* is extremely rare despite extensive survey efforts. We applied a tree-based machine learning algorithm to evaluate habitat characteristics associated with detection of *H. transpacificus* using paired data from a pelagic fish survey and a lower-trophic monitoring study. Preliminary analysis using only *H. transpacificus* data produced a model with limited predictive ability; models combining *H. transpacificus* with a closely related species, *H. nipponensis*, or a broader set of potential surrogate pelagic fishes were better supported. The catch per unit volume (CPUV) of the small pelagic fish community (SPFC) was the strongest explanatory variable for both *H. transpacificus* and combined *H. transpacificus*–*H. nipponensis* detections, indicating shared habitat use among multiple species in the estuary. Salinity and zooplankton prey density had the highest relative influence on SPFC CPUV. These results indicate that *H. nipponensis* and the SPFC are utilizing similar habitats as *H. transpacificus* and may therefore serve as appropriate proxies for *H. transpacificus* in habitat modeling or other monitoring and analysis. This finding also suggests the potential that *H. transpacificus* may compete with multiple species for limited food resources within small patches of high-quality habitat; the consequences of such dynamics on species recovery is identified as an important topic for future research.

KEY WORDS: Estuaries · *Hypomesus transpacificus* · Machine learning · Native fish · Proxy monitoring · Competition

1. INTRODUCTION

Efforts to establish habitat preferences for endangered and vulnerable species can be difficult given the rarity of detecting the species of interest, as much time and effort can be required for even a few detections (Shea & Mangel 2001, Thompson 2013, Belbachir et al. 2015, Jeliaskov et al. 2022). Even when large-scale efforts for sampling rare species are feasible, the resulting data can be challenging to analyze; large numbers of absences can prevent models from correctly identifying suitable habitat (Cianfrani et al.

2010) or create models with large predictive uncertainties (Hamilton et al. 2015). This may lead to ineffective actions or even failed efforts to bolster populations. Notwithstanding these challenges, understanding relationships between endangered species and their preferred habitat is key to conservation efforts (Rushton et al. 2004).

Delta smelt *Hypomesus transpacificus* have undergone a drastic decrease in abundance over the past several decades and inhabit a heavily altered and degraded environment. *H. transpacificus* is a small, semi-anadromous, planktivorous fish endemic to the

*Corresponding author: michael.tillotson@icf.com

San Francisco Bay Estuary and Sacramento-San Joaquin Delta (the Bay-Delta) with an annual life cycle (Moyle et al. 2016). In the Bay-Delta, *H. transpacificus* spawn during the winter–spring season, larval fish hatch in the spring and the majority of larvae make their way to the low salinity zone (LSZ, 0.5 to 6 ppt) where they transition into juveniles from summer to fall before migrating upstream to freshwater regions to spawn. Various anthropogenic stressors have contributed to the decline of *H. transpacificus* including destruction of habitat, exposure to contaminants, decreased prey availability, introduced invasive predators, competitors and prey, changes in flow regimes and direct entrainment due to water diversions, and increased temperatures during critical life stages (Moyle et al. 2016, Hamilton & Murphy 2018, FLOAT-MAST 2020). Numerous and substantial conservation and restoration efforts have been undertaken since *H. transpacificus* was listed as threatened under the Federal and California Endangered Species Acts in 1993 (the California listing was upgraded to endangered in 2009) (Moyle et al. 2016). Recently, large-scale management actions including flow augmentation during the fall and flow pulses during the summer were implemented to improve LSZ habitat for *H. transpacificus* (USFWS 2019, Sommer et al. 2020, Frantzich et al. 2021). However, *H. transpacificus* abundance has continued to decline in recent years, making detections of *H. transpacificus* even rarer and recovery efforts more urgent (Bacher 2022, USFWS et al. 2022, CDFW 2025). Ongoing efforts to introduce cultured *H. transpacificus* into different regions of its historical range began with experimental releases in 2021 (also known as supplementation, Lessard et al. 2018).

Understanding habitat preferences of *H. transpacificus* using the best available analytical tools and data therefore remains crucial to restoration efforts and management actions (e.g. flow actions and supplementation; Lessard et al. 2018, Hamilton & Murphy 2020); many models have been produced to predict *H. transpacificus* habitat suitability or habitat usage. Previous efforts have identified various combinations of abiotic and biotic variables as important components of *H. transpacificus* habitat (Feyrer et al. 2011, Bever et al. 2016, Hamilton & Murphy 2018, Polansky et al. 2018, Smith & Nobriga 2023). Other studies of *H. transpacificus* have utilized occupancy models to better separate the potentially independent processes of habitat preference and the ability to detect a species, given its presence (Mahardja et al. 2017, Peterson & Barajas 2018, Duarte & Peterson 2021, Davis et al. 2022, Hendrix et al. 2023). While these modeling efforts have collectively advanced the theoretical

understanding of historic *H. transpacificus* habitat preferences, verifying model results through additional sampling has become challenging, given that *H. transpacificus* are now rarely encountered in the long-term monitoring surveys on which many prior analyses were based. Even intensive, targeted *H. transpacificus* monitoring efforts have captured relatively few fish. Further insights are therefore likely to require novel analytical tools, new approaches for detecting rare species and/or use of cultured conspecifics or surrogates — more common taxa with shared habitat preferences — for modeling purposes.

A relatively recent development in modeling habitat preferences has been the use of machine learning algorithms that, in theory, can overcome some of the challenges associated with multiple variables, including potential interactions and non-linear relationships, while relaxing some of the assumptions associated with classical modeling approaches (e.g. linear and generalized linear models). Moreover, automated tree-based algorithms are now commonly used in ecological studies (Stupariu et al. 2022, Tuia et al. 2022), and have been used previously to predict species distribution and identify key habitat associations over different spatial scales for other endangered and threatened species (e.g. Hopkins 2009, Wang et al. 2015). Use of tree-based algorithms including random forest regression, gradient boosting machines (GBMs, also referred to as boosted regression trees [BRTs]) has proliferated in the field in part due to their adaptability and efficiency in generating predictive models. This family of machine learning algorithms are all fundamentally based upon classification and decision trees that employ various ensemble approaches to combine multiple, stochastically generated trees to improve the predictive performance and generalizability of the final ensemble. Additional advantages of using tree-based models are their ability to handle diverse types of predictor variables and accommodate missing data (Elith et al. 2008). When using tree-based models, the form of relationships between response and predictor, and the extent of interaction between covariates is dictated primarily by the data rather than making *a priori* assumptions as with generalized linear and additive models (Abeare 2009). Recent application of both tree-based machine learning and occupancy models to *H. transpacificus* have been met with mixed success. Grimaldo et al. (2021) used GBMs (i.e. BRTs) to evaluate factors that could predict *H. transpacificus* entrainment at water export facilities. However, despite identifying a model capable of explaining a large proportion of the null deviance, the predictive performance of the model

was poor, potentially due to the ongoing decline in species abundance and the resulting low sample size.

The initial objective of our study was to characterize the relative importance of abiotic factors, zooplankton prey, and other potential small pelagic fish competitors on the presence and abundance of *H. transpacificus*. We hypothesized that salinity, turbidity, and calanoid copepod prey, especially *Pseudodiaptomus forbesi* (Slater & Baxter 2014, Slater et al. 2019), would be among the strongest predictors. We sought to achieve this objective by applying the powerful, flexible, and interpretable BRT approach, and by combining 2 relatively new and high resolution datasets; the first resulting from an intensive sampling effort targeting *H. transpacificus* and other small pelagic fishes in the Bay-Delta, and the second from a similarly intensive effort to sample the invertebrate prey-base of the Bay-Delta that was frequently paired in space and time with the fish sampling effort. The models resulting from preliminary analyses using *H. transpacificus* alone as the response variable had limited predictive ability, but did identify the presence of potential competitors, including waka-sagi *H. nipponensis*, American shad *Alosa sapidissima*, threadfin shad *Dorosoma petenense*, and Mississippi silverside *Menidia audens*, as the strongest predictor of *H. transpacificus* presence.

These preliminary results made clear that the continued decline of the species in recent years has left researchers with 2 suboptimal options for analyzing the habitat relationships of *H. transpacificus*: rely on historical data that may have little bearing on current and future conditions due to the fundamental changes that have occurred in the Bay-Delta or rely on more recent data where the low rate of detections challenge even sophisticated models. The use of indicator or proxy species may offer a third, complementary option to draw inferences about habitat preferences and quality. Studies suggest that other non-native potential competitors such as *H. nipponensis* and *D. petenense* may be suitable proxies for *H. transpacificus*, because they have overlapping ecological niches, (Feyrer et al. 2003, Davis et al. 2022), are associated with *H. transpacificus* occupancy (Hendrix et al. 2023), and are caught in the same monitoring gear types (Stompe et al. 2020, McKenzie & Mahardja 2021).

Proxy species and community level analyses have been successfully used for modeling other endangered and low abundance species (Dunn et al. 2015, Silvano et al. 2017, Rodríguez et al. 2021). Dunn et al. (2015) used 2 proxy bird species with similar macro-habitat requirements to the critically endangered

Himalayan quail *Ophrysia superciliosa* to model suitable survey areas. Rodríguez et al. (2021) used community-level models to understand vulnerable fish species responses to environmental stressors through shared species responses to the environment. Prior research has shown that when using proxy species, appropriate precautions need to be considered, such as if the proxy species shares the same key ecological traits that make the endangered species population vulnerable to disturbance or if the appropriate habitat scale is being considered (Caro et al. 2005, Loman et al. 2021). With these precautions in mind, and based on the strong, linear, and positive relationship between the presence of *H. transpacificus* and other members of the small pelagic community, we proceeded to evaluate whether a proxy species approach involving the grouping of *H. transpacificus* with 1 or more of the other Bay-Delta small pelagic species could improve model performance and identification of important habitat features. In addition to potentially increasing power to evaluate *H. transpacificus* habitat associations, this approach may provide broader insight into the drivers of the Bay-Delta pelagic forage fishes and serve as a potential modeling framework for other rare species that are now infrequently detected.

2. MATERIALS AND METHODS

We focused our study on the upper portion of the Bay-Delta (Fig. 1), where the USFWS conducts routine monitoring of *Hypomesus transpacificus* through the Enhanced Delta Smelt Monitoring program (EDSM; USFWS et al. 2022). Sampling locations were chosen using a generalized random tessellated stratified (GRTS) sampling design, ensuring spatially well-distributed random sites (Stevens & Olsen 2004, Starcevich et al. 2016). Kodiak trawls were conducted from July through March for juvenile and adult *H. transpacificus* in different regions of the overall sample frame. Environmental parameters (temperature, turbidity, salinity, and chlorophyll *a*) and lower-trophic invertebrate samples (mesozooplankton and macrozooplankton) were collected alongside EDSM sampling from 2017 to 2021 as part of the Directed Outflow Project Lower Trophic Study (DOP; Schultz & Kalmbach 2022). This paired sampling was conducted with a subset of the EDSM monitoring sites; pairing occurred for 3 different sites weekly within a region. The total number of sampling strata used by EDSM has varied over time, and the DOP did not sample all strata visited by EDSM; 19 regions were con-

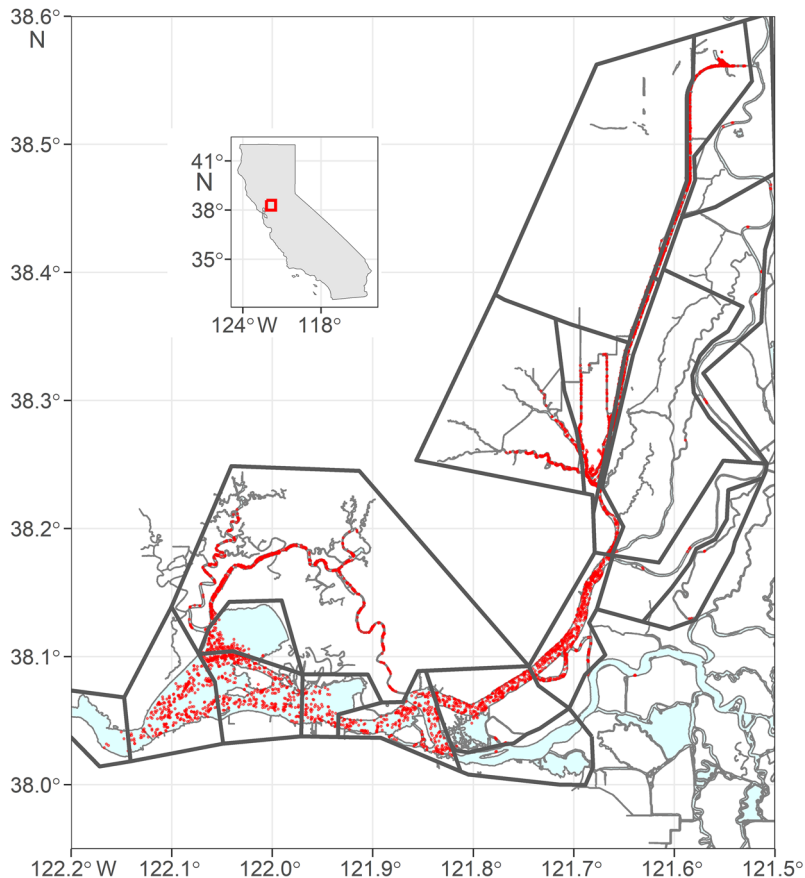


Fig. 1. Sacramento-San Joaquin Delta. Polygons show the sampling strata across which invertebrate data were averaged; (•) location of each fish sampling station

sidered for this analysis with 5 regions typically sampled each week (Fig. 1). While sampling was conducted year-round, this study focuses on the summer through late fall period (July through November) to align with the Delta Smelt Summer-Fall Habitat Action (DWR & USBR 2023) which aims to improve the recruitment, growth, and survival of *H. transpacificus*. Data were examined from 2017 through 2021. Only fall season data were available for 2017 to 2018 from the DOP; summer and fall data were available for 2019 to 2021.

Kodiak trawling was conducted following methods described by USFWS et al. (2022). Between 2 and 10 tows were completed at each site, with a minimum duration of 2 min and a maximum of 10 min per tow. The Kodiak trawl net was towed in the first few meters of the water column against the water current. Mean net mouth size was estimated to be 12.54 m². The distance traveled during each tow was recorded using a calibrated mechanical flow meter (distance = volume/cross-sectional area), which was used to calculate catch per unit volume (CPUV; catch per 10 000 m³).

If a *H. transpacificus* was captured within the first 2 tows at a site, then the site was considered complete and no further tows were conducted; otherwise, tows were repeated until a *H. transpacificus* was caught or the maximum number of tows (typically 4) were completed. A suite of water quality parameters was also collected during fish sampling, including temperature (°C), dissolved oxygen (mg l⁻¹), specific conductance (microsiemens [μS] cm⁻¹ for freshwater and milli-siemens [mS] cm⁻¹ for brackish water), turbidity (nephelometer turbidity units [NTU]), and Secchi depth (cm). Water quality parameter measurements were taken from the top 1 m of the water column, using a YSI Pro2030 meter. Turbidity values were obtained using Hach 2100Q turbidity meters.

Macrozooplankton and mesozooplankton sampling was conducted according to methods described by Schultz & Kalmbach (2022). One tow was conducted for each habitat type (shoal, channel surface, channel deep) present at a site. A shoal was defined as habitat that was roughly 3 m deep or less. Channel habitat was greater than 3 m deep; a surface tow was

conducted in both channel and shoal habitats in the first 1 m of the water column. In channels greater than 6 m deep, an additional tow was conducted in the bottom half to third of the water column. Not all habitat types were present at each site, so between 1 and 3 tows were conducted per site. In 2017, invertebrates were sampled with 2 Clarke-Bumpus nets (50 cm diameter) attached to a bongo net frame; mesozooplankton were targeted with a 150 μm mesh size net and macrozooplankton were targeted with a 50 μm mesh size net. Beginning in 2018, the mesozooplankton net was switched to a 20 cm diameter net. The tow duration was 8 min in 2017 but was shortened to 5 min beginning in 2018 to reduce sample volume and lab processing time. Zooplankton samples were processed according to methods described in the Interagency Ecological Program's Environmental Monitoring Program (Kayfetz et al. 2020), except a maximum of 10 aliquots were used to account for higher sample densities resulting from using larger volume nets as described by Schultz & Kalmbach (2022).

2.1. Data processing

Fish catches were summarized to station-level CPUV by dividing the total catch of each species summed across all tows conducted at a station by the total tow volume. The potential for depletion across tow replicates was assumed to be negligible given the pelagic and dynamic nature of the waters sampled. In addition to *H. transpacificus*, 5 other fish species were incorporated into either dependent or independent variables in this analysis (Fig. S1 in the Supplement at www.int-res.com/articles/suppl/n056p117_supp.pdf). Although many other species may share the same pelagic habitat and food preferences as *H. transpacificus* during portions of their life histories, we selected the most encountered forage fish species from the fish survey to represent potential competitors of *H. transpacificus* (Fig. 2). Northern anchovies were excluded from the potential competitors variable despite a high frequency of capture in some areas of the Bay-Delta because of their primarily marine life-history, preference for much higher salinities than *H. transpacificus*, and very infrequent detection in the regions where the remnant *H. transpacificus* population now occurs. The potential competitors variable (Competitors_All) included *Dorosoma petenense*, *Alosa sapidissima*, *Menidia audens*, *H. nipponensis*, and longfin smelt *Spirinchus thaleichthys*; all

taxa were filtered by fork length (i.e. proxy for life stage) for greater accuracy. Abundances of each of the potential surrogates/competitors showed a positive association with *H. transpacificus* presence (Fig. S2), but *H. nipponensis* were selected as the most suitable single-species surrogate based on multiple characteristics including phylogeny, morphology, and life-history (Davis et al. 2022). When modeling combined *H. transpacificus* and *H. nipponensis* catch as the response variable, *H. nipponensis* were removed from the independent potential competitor variable (Competitors_NoWakasagi).

Fish catch data were zero-inflated, indicating patchy distributions or distinct processes governing presence–absence and abundance. A 2-step approach was therefore used, modeling presence–absence and abundance separately (Smith et al. 2019). Each response variable was first considered as a binary metric of presence and absence, then as a continuous variable (CPUV) for positive detections. Six potential response variables were considered: presence/absence and CPUV for *H. transpacificus* alone, *H. transpacificus* with *H. nipponensis*, and a broader group of small pelagic species (Table 1; the *H. transpacificus* CPUV model was later excluded because of small sample size). Zooplankton were classified into 3 groups: Mysidae, Amphipoda, and mesozooplankton (0.2 to 20 mm total length). This was then

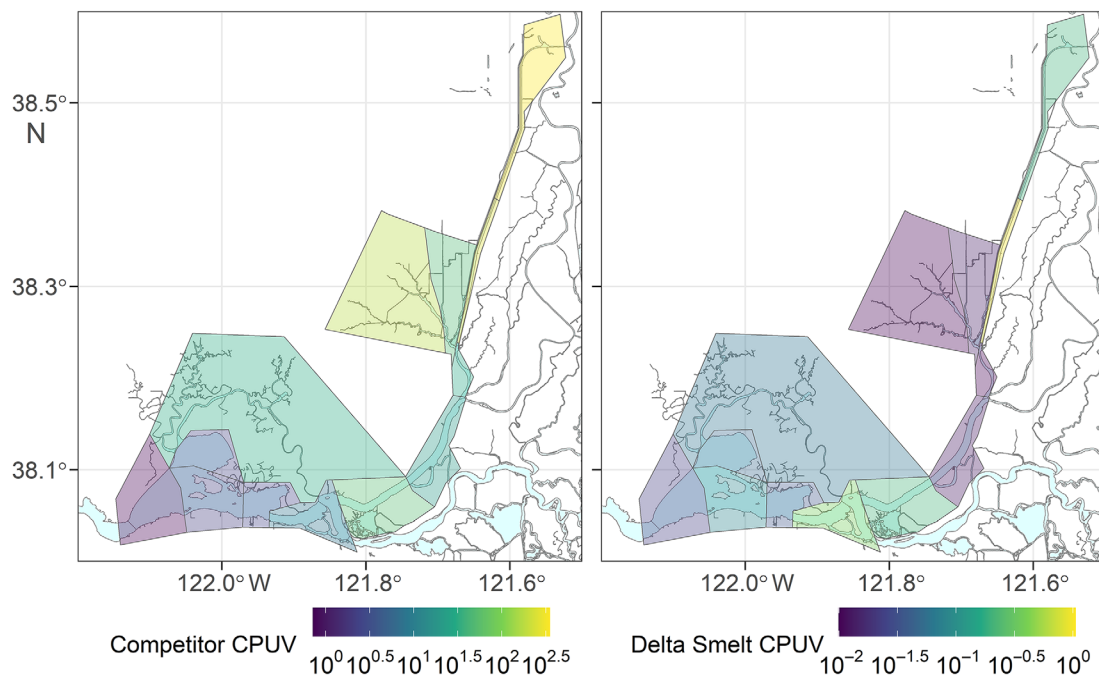


Fig. 2. Comparison of mean catch per unit volume (CPUV; catch per 10 000 m³) by sampling stratum for *Hypomesus transpacificus* and their potential small pelagic competitors

Table 1. Overview of Models 1 to 5. Competitor covariates are composed of small pelagic fish that we hypothesize as having similar ecological niches and may cohabitate due to reduced high-quality habitat availability in the Bay-Delta. CPUV: catch per unit volume; DS: delta smelt

Response variable	Competitor covariate	Abiotic covariates	Model no.	Food covariates
<i>Hypomesus transpacificus</i> binary or <i>H. transpacificus</i> – <i>Hypomesus nipponensis</i> binary or Small pelagic binary or <i>H. transpacificus</i> – <i>H. nipponensis</i> CPUV or Small pelagic CPUV	All competitors (<i>Dorosoma petenense</i> , <i>Alosa sapidissima</i> , <i>Spirinchnus thaleichthys</i> , <i>Menidia audens</i> , <i>H. nipponensis</i>) or Competitors without <i>H. nipponensis</i> or None	Delta outflow (14 d moving average), Salinity, Turbidity, Temperature, Chlorophyll <i>a</i> , Week of Year, Coarse Region	1	Total Zooplankton CPUV, Total Mysid CPUV, Total Amphipod CPUV
			2	DS Food Zooplankton CPUV, DS Food Mysid CPUV, DS Food Amphipod CPUV
			3	DS Food Calanoids CPUV, DS Food Cladocerans CPUV, DS Food Cyclopoids CPUV, DS Food Mysid CPUV, DS Food Amphipod CPUV
			4	<i>Pseudodiaptomus forbesi</i> CPUV, <i>Eurytemora affinis</i> CPUV, <i>Sinocalanus doerrii</i> CPUV, <i>Acartiella sinensis</i> CPUV, <i>Tortanus</i> spp. CPUV, <i>Limnoithona tetraspina</i> CPUV, <i>Bosmina longirostris</i> CPUV, <i>Sididae</i> spp. CPUV, <i>Daphniidae</i> _spp. CPUV
			5	<i>P. forbesi</i> adult CPUV, <i>P. forbesi</i> copepodite CPUV, <i>E. affinis</i> adult CPUV, <i>E. affinis</i> copepodite CPUV, <i>S. doerrii</i> adult CPUV, <i>S. doerrii</i> copepodite CPUV, <i>A. sinensis</i> adult CPUV, <i>A. sinensis</i> copepodite CPUV, <i>Tortanus</i> spp. adult CPUV, <i>Tortanus</i> spp. copepodite CPUV, <i>Limnoithona tetraspina</i> CPUV, <i>Bosmina longirostris</i> CPUV, <i>Sididae</i> spp. CPUV, <i>Daphniidae</i> _spp. CPUV

narrowed to taxa verified as *H. transpacificus* prey (Slater & Baxter 2014, Slater et al. 2019; Table 1). Mesozooplankton were further separated into Calanoida, Cyclopoida, and Cladocera, then into species-genus for known or hypothesized *H. transpacificus* prey (Table 1). Selected calanoid and cyclopoid species were differentiated by life stage (adult vs. copepodite). Five models were evaluated for each response variable, with increasingly granular invertebrate data (Table 1).

Invertebrate and fish sampling were not always perfectly paired in time and space, many detections of *H. nipponensis* and *H. transpacificus* occurred when the invertebrate study was not sampling. Invertebrate data were therefore averaged across time and space to increase the number of paired samples. Each fish sample was paired with invertebrate samples collected up to 1 wk prior in the same subregion, summarized to taxon-specific mean CPUV values. Invertebrate sampling occurred in multiple habitats (shoal, channel surface, and channel deep) at each site. Invertebrate CPUV did not vary consistently between habitats; catch was averaged across all habitats sampled at a site. After processing, 2246 unique station–date combinations remained for analysis. Training and testing datasets for each response variable were then created using the 'createDataPartition' function from the R package 'caret' which attempts to balance the classes within each set (Kuhn 2022). For presence–absence models, all 2246 instances were used, while models using CPUV in the response were trained using censored data including only instances when the species or species group was detected (*H. transpacificus*: N = 104, *H. transpacificus*–*H. nipponensis*: N = 284, small pelagic fish: N = 1517). In all cases, 70% of the data was used for training with 30% reserved for testing and cross validation.

2.2. Model fitting

We first developed a range of candidate models incorporating combinations of abiotic habitat features and food variables derived from the invertebrate sampling data based on reported *H. transpacificus* food and habitat preferences from prior studies (Feyrer et al. 2011, Slater & Baxter 2014, Slater et al. 2019). Abiotic covariates included water quality parameters, a 14 d backward-looking moving average of delta outflow (California Data Exchange Center; CDEC), week of year, and region (Fig. 1). Food variables ranged from coarse aggregations (all mysids, amphipods, zooplankton) to individual *H. transpacificus* prey

taxa (Table S1). The influence of these covariates on each response variable was modeled using the GBM algorithm in the 'gbm' and 'dismo' R packages (BRT models).

To limit overfitting, we used a grid search to identify optimal values for learning rate (weight of each new tree) and tree complexity (number of levels per tree). BRT models were fitted using the 'gbm' function with learning rates between 0.1 and 0.0005 and tree complexities of 2 to 5 (Greenwell et al. 2020). Hyperparameter combinations were compared based on root mean square error (RMSE), and the lowest error combination was used for model fitting. Each model was then refit using the 'gbm.step' function from the 'dismo' package, which automates the identification of the optimal number of trees for a given complexity and learning rate (Hijmans et al. 2021).

2.3. Model validation

After fitting, models were evaluated using a variety of cross-validation and diagnostic techniques. The 'gbm' function reports internal cross-validation metrics, including AUC for classification models and cross-validation correlation for regression models. AUC values above 0.8 indicate strong discrimination (Hosmer et al. 2013). Predictions were generated using reserved training data. For presence–absence models, predicted classifications were generated using threshold values between 0.3 and 0.7 and compared with observations using the 'confusionMatrix' function from the R package 'caret', which calculates performance indices like accuracy, no-information rate, Cohen's kappa, sensitivity, and specificity. For CPUV models, observed and predicted values were compared visually using scatterplots and Pearson's correlation coefficients.

The primary objective was to describe, compare, and visualize the influence of biotic and abiotic factors on the presence–absence or CPUV of *H. transpacificus* and their surrogates. The most influential covariates were identified by comparing variable importance, calculated by permuting the model with a single covariate excluded and measuring the change in residual error. Relationships between dependent and independent variables were evaluated using partial dependence plots from the 'plot.gbm' function, showing the effect of a single covariate with others held constant. Shapley values were estimated for each prediction using the 'fastshap' function (Greenwell 2021), indicating the contribution of each covariate to a prediction and providing insight into covariate importance and dependencies (Winter 2002).

3. RESULTS

3.1. Summary

In total, 8875 tows from the fish study and 4526 tows from the invertebrate study were included in the final, paired dataset, which, after summarizing to the station level, resulted in 2246 paired samples. Among these samples, there were 104 positive detections for *Hypomesus transpacificus*, 284 for *H. transpacificus* – *H. nipponensis* and 1517 for small pelagic fish. Prevalence was therefore 4.6% for *H. transpacificus*, 12.6% when combined with *H. nipponensis*, and 67.5% for all pelagic fish. After censoring non-detections, the average CPUV was 3.19×10^{-5} , 9.77×10^{-5} , and 1.24×10^{-3} for *H. transpacificus*, *H. transpacificus* – *H. nipponensis*, and small pelagic fish, respectively. Mean CPUV for the 3 primary invertebrate groupings, amphipods, mysids, and zooplankton, was 9.12×10^{-1} , 3.85×10^1 , and 5.23×10^3 , respectively. In all cases, most of the catch was comprised of potential *H. transpacificus* food items (95, 99, and 79% for amphipods, mysids, and zooplankton, respectively). Summary statistics by region for specific zooplankton, mysid, and amphipod taxa are given in Table S1.

3.2. Delta smelt models

Some validation metrics indicated good fits for the models of *H. transpacificus* presence and absence. For example, regardless of at which level invertebrate data were aggregated, model accuracy was 0.96. However, cross-validation indicated that for each of

the models in this set, the no-information p-value was substantially greater than 0.05, which meant that simply predicting absence in all cases would result in greater model accuracy (Table 2). Given this indication of poor predictive performance and the low prevalence of *H. transpacificus*, we elected not to model censored CPUV for this response variable and interpret the model results cautiously. Examination of variable importance scores and partial dependence plots nevertheless indicated an interesting and unexpected association with hypothesized potential competitors that warranted further evaluation. We hypothesized that *H. transpacificus* would try to avoid other small pelagic fishes to reduce potential competition for food resources, and that this relationship may be difficult to detect in our dataset (i.e. we expected a weak, negative relationship between *H. transpacificus* presence and potential competitor abundance). Contrary to these expectations, potential competitor CPUV was consistently the most important variable in this set of models (Fig. 3), and the relationship with *H. transpacificus* presence was strong, positive, and essentially linear. This association was confirmed visually by comparing boxplots of the CPUV of potential competitor species binned by *H. transpacificus* presence and abundance (Fig. S2). Other variables with high importance rankings were salinity and delta outflow; food variables were generally ranked with lower importance across the various levels of invertebrate aggregation. The relationship between *H. transpacificus* presence and salinity was nonlinear with a low probability of presence predicted at the lowest values, a rapid increase between ~0.1 and 1 ppt, and an intermediate plateau at higher salinities (Fig. 4A).

Table 2. Summary of model cross-validation statistics

Response	Model no.	Accuracy	Kappa	No-information p-value
<i>H. transpacificus</i>	1	0.96	0.124	0.398
<i>H. transpacificus</i>	2	0.96	0.124	0.398
<i>H. transpacificus</i>	3	0.96	0.124	0.398
<i>H. transpacificus</i>	4	0.96	0.124	0.398
<i>H. transpacificus</i>	5	0.96	0.124	0.398
<i>H. transpacificus</i> – <i>H. nipponensis</i>	1	0.92	0.629	<0.001
<i>H. transpacificus</i> – <i>H. nipponensis</i>	2	0.92	0.645	<0.001
<i>H. transpacificus</i> – <i>H. nipponensis</i>	3	0.92	0.631	<0.001
<i>H. transpacificus</i> – <i>H. nipponensis</i>	4	0.92	0.634	<0.001
<i>H. transpacificus</i> – <i>H. nipponensis</i>	5	0.92	0.643	<0.001
Small pelagic fish	1	0.78	0.478	<0.001
Small pelagic fish	2	0.78	0.455	<0.001
Small pelagic fish	3	0.78	0.484	<0.001
Small pelagic fish	4	0.78	0.486	<0.001
Small pelagic fish	5	0.78	0.476	<0.001

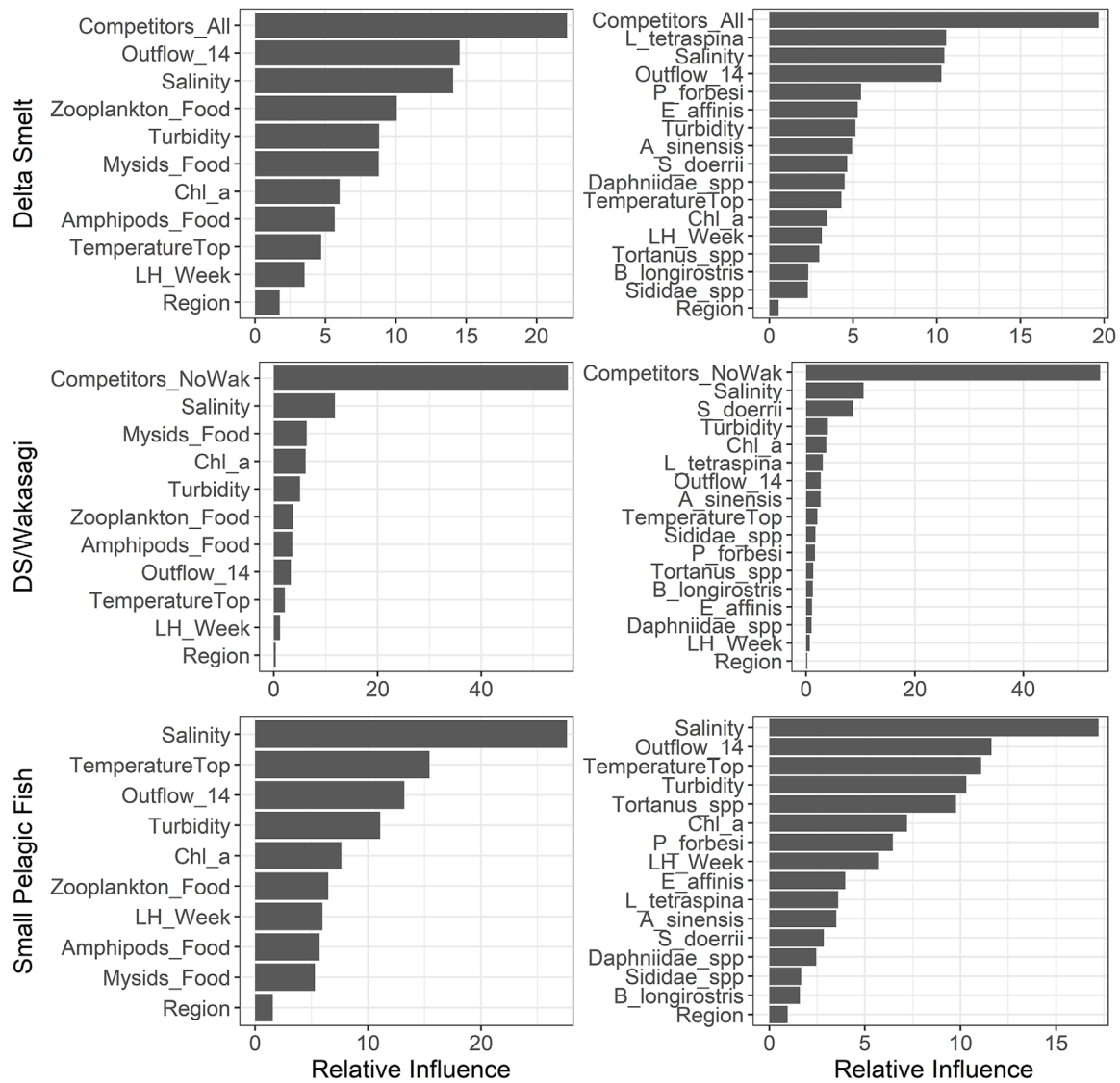


Fig. 3. Variable importance rankings for selected presence–absence models. For each response variable, the first column shows rankings from models with relatively coarse aggregation of food items (Model 2) while the second column shows rankings from models with finer separation of food items (Model 4). Model variables can be found in Table S1. DS: delta smelt

3.3. *Hypomesus transpacificus* and *H. nipponensis* presence–absence

Given the positive and linear association between *H. transpacificus* detection and the abundance of their potential competitors, we evaluated whether combining *H. transpacificus* detections with those of 1 or more proxy species could help to improve model performance while still providing results relevant to threatened species management. Because they are the closest relatives of *H. transpacificus* among the small pelagic fish group, *H. nipponensis* were the most directly analogous proxy species (Davis et al. 2022). As with the *H. transpacificus*-only models,

there was little evidence of any preference for 1 level of invertebrate aggregation over another (Table 2); prediction accuracy was similar across all models in this set, though very slightly higher with intermediate levels of aggregation in the food variables. In contrast with the *H. transpacificus*-only models, the inclusion of *H. nipponensis* in the response resulted in model predictions that were both highly accurate and outperformed a consistent prediction of absence (i.e. no-information rate p -value < 0.05). Despite this improvement in predictive performance, the models of combined *H. transpacificus* and *H. nipponensis* detections nevertheless indicated the same dominant drivers of distribution, with potential competitors and

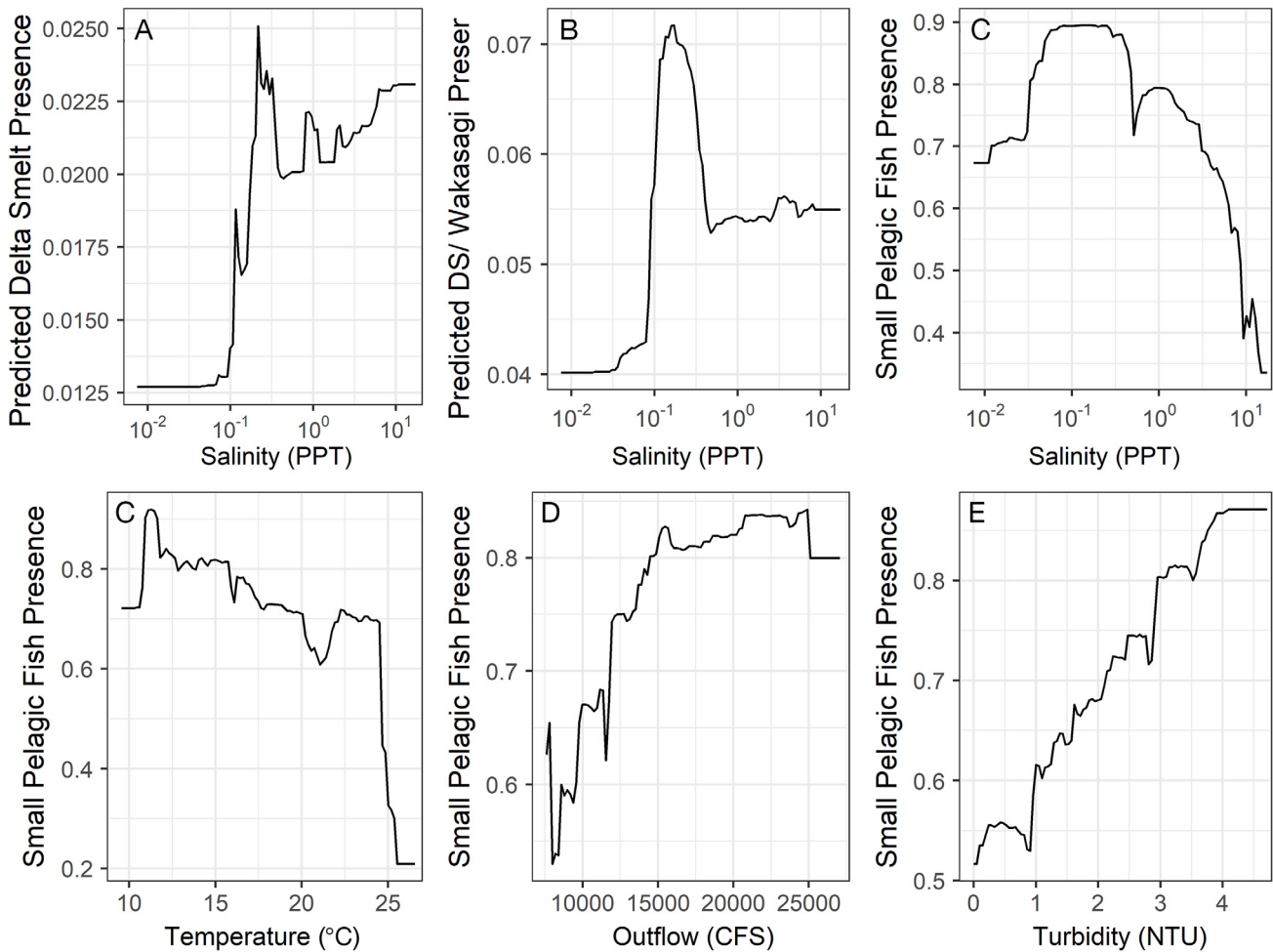


Fig. 4. Conditional effect of important covariates on (A) *Hypomesus transpacificus* presence, (B) combined *H. transpacificus* and *H. nipponensis* presence, and (C–F) small pelagic fish presence. DS: delta smelt; CFS: cubic feet per second

salinity consistently having the highest variable importance rankings. Indeed, the inclusion of *H. nipponensis* resulted in an even stronger influence of potential competitors and further reduced the relative importance of food variables (Fig. 5B); this relationship was again positive and linear. Meanwhile, the influence of salinity differed somewhat from the *H. transpacificus* results with a stronger peak in predicted presence in the 0.1 to 1 ppt range (Fig. 4B). Outflow was ranked lower than in the *H. transpacificus* models, but nevertheless retained a positive and mostly linear relationship with the probability of presence.

3.4. Small pelagic fish presence–absence

The outsized importance of potential competitors in explaining the distribution of *H. transpacificus*, both alone and in combination with *H. nipponensis*, was

unexpected, but suggested that a broader suite of small pelagic fish species may reflect to some degree the habitat preferences of *H. transpacificus*. We therefore progressed further in our use of proxy species, next modeling the presence and absence of the full small pelagic fish community. No further potential competitor variable was available for this analysis, so the expectation was that these models would help to clarify the relative importance of abiotic conditions and food availability in determining fish distribution. For each of the small pelagic fish presence-absence models, regardless of the level of taxonomic aggregation in the food covariates, the 4 most important variables were always abiotic and included salinity, temperature, delta outflow, and turbidity (Fig. 3). The probability of presence was positively related to outflow and turbidity, negatively related to temperature, and again, related to salinity in a nonlinear fashion (Fig. 4C–E). In contrast with the *H. transpacificus*–*H. nipponensis*

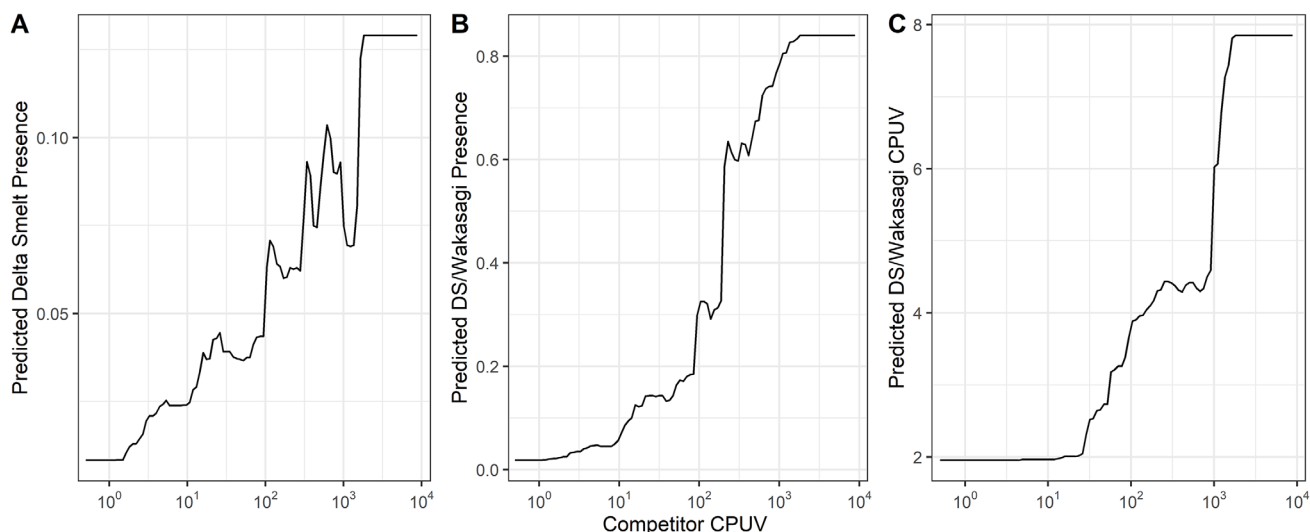


Fig. 5. Conditional effect of competitor density on (A) *Hypomesus transpacificus* detection probability, (B) *H. transpacificus* – *H. nipponensis* detection probability, and (C) *H. transpacificus* – *H. nipponensis* catch per unit volume (CPUV). DS: delta smelt

models, the relationship had a broader peak and with the lowest probability of presence occurring at the highest salinities.

3.5. Abundance

Abundance of *H. transpacificus*–*H. nipponensis* and small pelagic fish was modeled with datasets censored for presence. Cross-validation correlations and their significance varied little across levels of food aggregation for either small pelagic fish ($R^2 = 0.47 - 0.49$, $p < 0.001$) or *H. transpacificus*–*H. nipponensis* ($R^2 = 0.24 - 0.28$, $p < 0.001$). Results of the abundance models shared some similarities with the presence–absence models in terms of variable importance rankings and the general shapes of the relationships (e.g. for salinity and potential competitor density), but also had some interesting differences. Zooplankton, especially *Sinocalanus doerrii*, became one of the most, or the most, important covariate(s) in the small pelagic CPUV models (Fig. S3), with a positive and mostly linear effect of these food covariates (Fig. S4E,F). A similar pattern was observed for the combined *H. transpacificus* and *H. nipponensis* CPUV response variable, with *S. doerrii* density having the highest relative influence (Figs. S3 & S4B,C).

3.6. Regional dynamics

The choice to exclude finer regional classifications from the candidate covariates was intentional, with

the reasoning that allowing the biotic and abiotic characteristics to capture regional differences would produce more generalizable and informative results. However, examination of model predictions across finer-scale regions can provide insight into whether the influence of a given variable is consistent throughout the Bay-Delta, or whether there are localized dynamics. Calculation of Shapley values, which reflect the contribution of each covariate to each model prediction, allowed comparison of model dynamics on a subregional basis, even though subregion was not included as a covariate. Of interest was how the model explained the generally higher frequency of detection and CPUV in the Sacramento River Deep Water Ship Channel (SRDWSC; Fig. 1), which was previously identified as a hotspot for *H. transpacificus* population remnants (Merz et al. 2011). Based on examination of regional patterns in Shapley values, it appears that potential competitor CPUV contributed consistently to model predictions across regions (Fig. 6, Row 1), but that higher potential competitor catch densities were much more common in the SRDWSC. In contrast, the density of *S. doerrii* only contributed appreciably to model predictions for stations within the SRDWSC, and the covariate values had little overlap between regions (Fig. 6, Row 2.); the catch density for *S. doerrii* at SRDWSC stations was consistently higher than in other areas. Lastly, despite substantial overlap in the ranges of salinity values between regions, the contribution of this covariate to model predictions appeared larger within the SRDWSC region and interestingly, the majority of *H. transpacificus* or *H. nipponensis* detections

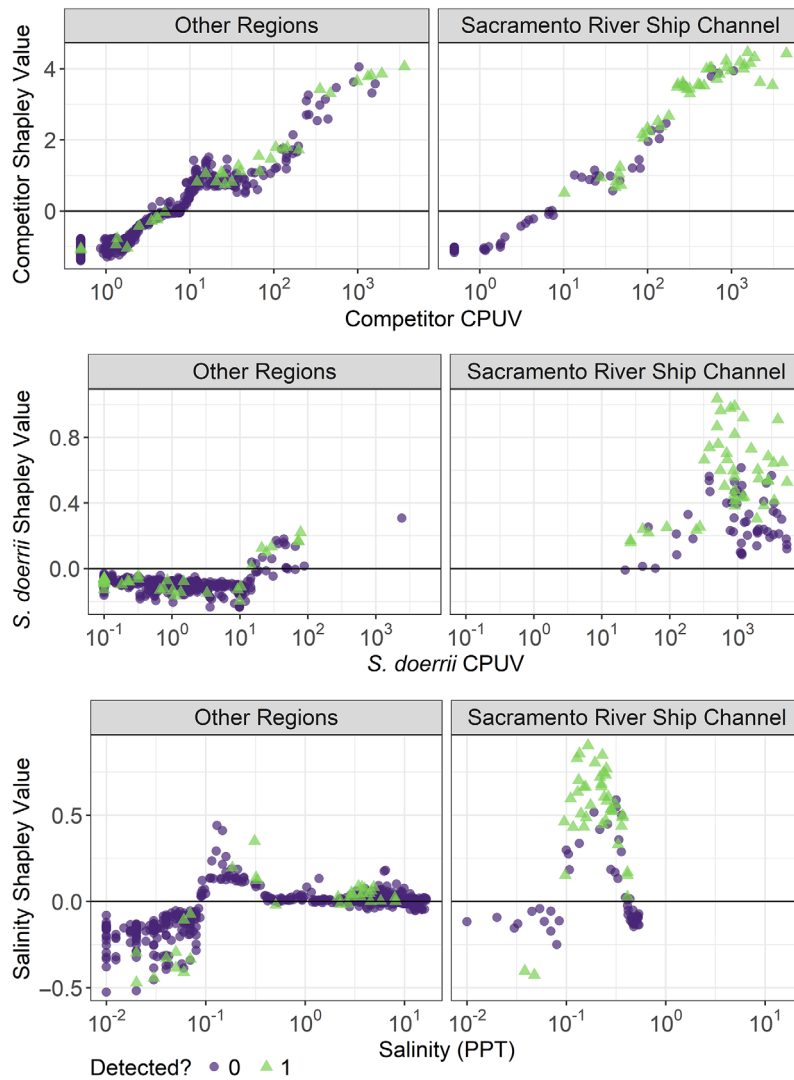


Fig. 6. Shapley values relative to selected covariate values for *Hypomesus transpacificus* – *H. nipponensis* Model 4 with comparison of the Sacramento River Deep Water Ship Channel to other sampled regions. Shapley values reflect the contribution of a covariate to the individual model predictions. Colors show observed presence and absence of *H. transpacificus* or *H. nipponensis*. CPUV: catch per unit volume

occurred within a very narrow salinity range (Fig. 6, Row 3).

4. DISCUSSION

Our study used machine learning to evaluate the importance of abiotic and biotic habitat characteristics in determining the distribution of *Hypomesus transpacificus* in the San Francisco Estuary. Preliminary models were of limited utility because of poor predictive performance, which likely resulted from very sparse detection of the target species. To potentially circumvent some of the issues associated with

sparse detection, we updated the analyses to include a hypothesized surrogate species, *H. nipponensis*, which improved predictive power and generated better-supported models. Contrary to our expectations, both models indicated that the abundance of the small pelagic fish community (which we hypothesized represented potential competitors for the same food resources) had a strong, and positive influence on presence of the target and surrogate species. This pattern was further supported by a simple examination of abundances of *H. transpacificus* and other small pelagic species. Apart from this strong, positive influence of other small pelagic fishes, presence–absence models found abiotic variables to be the most

influential. For CPUV models, food abundance and salinity were typically the strongest predictors. These results suggest abiotic variables may limit habitat occupancy, while biotic variables determine abundance when conditions are suitable. Given the limited predictive performance of models that included only *H. transpacificus* in the response variable, the results of these preliminary models must be interpreted cautiously. The primary conclusion drawn from the initial models was that *H. transpacificus* were more likely to be present and relatively abundant when the abundance of other small pelagic fishes was also high. This pattern was further evaluated by a simple comparison of *H. transpacificus* presence with the abundances of each of the other small pelagic species which found strong agreement with the modeled results (Fig. S2). Remaining conclusions were drawn from models that integrated 1 or more potential surrogate species, which were collectively deemed more reliable given cross-validation results that indicated superior predictive ability.

The relative importance of the potential competitor variable stood out in both *H. transpacificus* and the combined *H. transpacificus* and *H. nipponensis* models, far outweighing all other variables considered. These strong positive relationships are surprising, given previous research of *H. transpacificus* habitat use that evaluated the effect of hypothesized competitors; Hendrix et al. (2023) found a slight negative relationship between *Dorosoma petenense* (a species we included in our competitor variable) and *H. transpacificus* occupancy. However, Hendrix et al. (2023) found a positive relationship between *H. transpacificus* occupancy and predation intensity, which was interpreted as some degree of overlap between *H. transpacificus* and predator habitat. In contrast, comparisons between *H. transpacificus* and *H. nipponensis* have shown overlap in ecological niches, distribution, and environmental tolerances (Swanson et al. 2000, Davis et al. 2022), while analysis of the declines of other pelagic fish species in the Bay-Delta have indicated similar responses to changes in certain environmental variables (such as salinity and turbidity; Feyrer et al. 2007, Thomson et al. 2010) and changes to the prey community (Feyrer et al. 2003). Our results therefore suggest the ecological niche of *H. transpacificus* overlaps to some degree with other small pelagic fish species.

The strong positive associations between *H. transpacificus* and *H. nipponensis* with the small pelagic fish competitor variable may reflect associations between the species at smaller spatial scales and could result from interspecific shoaling behavior. Origi-

nally, *H. transpacificus* were not believed to be a strongly shoaling species (Moyle 2002), but shoaling behavior has been observed in controlled experimental settings (Ferrari et al. 2014, Davis et al. 2019). Shoaling behavior has multiple benefits, ranging from lowering cortisol levels and metabolic rates to increased predator avoidance (Hasenbein et al. 2016, Nadler et al. 2016, Chrétien et al. 2021, reviewed by Pavlov & Kasumyan 2000). Given the decline in the *H. transpacificus* population, individuals may associate with similarly sized and shaped fish that share an ecological niche, since conspecifics are either not present or present at only a low density, which may explain the strong association with other small pelagic species. Other fish species have been shown to readily shoal with morphologically similar species regardless of whether the other species were native or introduced (Camacho-Cervantes et al. 2014).

Larger groups of fish can also find patchy food resources more quickly; however, this may lead to increased food competition in mixed species shoals, unless food preferences differ (reviewed by Pavlov & Kasumyan 2000). There is evidence that some of the other small pelagic fish species (*D. petenense*) prefer similar prey to *H. transpacificus*, mainly pelagic copepods (Feyrer et al. 2003). However, results from our study suggest the small pelagic fish community may have been consuming different copepod species. In our study, *Sinocalanus doerrii* had a high relative influence on the CPUV of small pelagic fish, whereas other studies have shown that *H. transpacificus* feed primarily on *Pseudodiaptomus forbesi* during the summer and fall season. While *H. transpacificus* do consume *S. doerrii*, it is primarily during the adult life stage in the winter (Slater et al. 2019); a period not evaluated here. Given the general decline of zooplankton in the delta, changes in community composition, and seasonal shifts in peak abundances of important prey species (Winder & Jassby 2011, Merz et al. 2016), each fish species could now be targeting different available prey resources (Ingram & Ziebell 1983). Feyrer et al. (2003) observed expanded niche breadth and lower prey overlap in the tidal marsh fish community when the dominant food resource (mysids) greatly declined in abundance, consistent with the niche overlap hypothesis (Pianka 1974). The apparent importance of *S. doerrii* may also be more of a modeling artifact than a reflection of actual ecological dynamics. High densities of this species were found almost exclusively in the SRDWSC (Fig. 6). This suggests that *S. doerrii* might be used by the models as a surrogate for an unobserved characteristic of this hotspot for *H. transpacificus* and *H. nipponensis* (Davis et

al. 2022). Alternatively, the correlation between invertebrate prey and fish abundance might reflect shared habitat requirements rather than tracking preferred prey. Direct comparisons of diets for fish caught concurrently with *H. transpacificus* could help to clarify whether *H. transpacificus* and small pelagic fishes are consuming different species of copepod in the same environment.

Some regions of the Bay-Delta appear to support the small pelagic fish community but not *H. transpacificus*. These include parts of the lower Sacramento River (i.e. below the SRDWSC and Sacramento River confluence near Rio Vista) and Cache Slough (Fig. 3). Historically, Cache Slough supported *H. transpacificus* across multiple life stages (Sommer et al. 2009, Merz et al. 2011). However, the invasion of Brazilian waterweed *Egeria densa* may have reduced habitat by lowering turbidity (Hestir et al. 2016, Lacy et al. 2021) and increasing predation pressure from species like largemouth bass *Micropterus salmoides* (Brown & Michniuk 2007, Conrad et al. 2016). Rising surface water temperatures have also reduced habitat, particularly upstream (Halverson et al. 2022), including Cache Slough and the SRDWSC. Temporary thermal stratification in deep channels may provide refuge for *H. transpacificus*, explaining higher CPUV in the SRDWSC compared to Cache Slough (Lenoch et al. 2021, Mahardja et al. 2022). Shad species have higher thermal tolerance (35°C for juvenile *Alosa sapidissima*, Bayse et al. 2020, 33.3°C for *D. petenense*, Monirian et al. 2010) that allows them to survive higher temperatures in Cache Slough, accounting for the high small pelagic CPUV in this region. Salinity is another key determinant of fish assemblages and distribution across aquatic ecosystems (Thiel et al. 1995, Ostrand & Wilde 2001, Barletta et al. 2005, Shaha et al. 2022, D'Amen et al. 2023), and in our study, it was the abiotic variable with the highest relative influence across all models. For the *H. transpacificus* and *H. nipponensis* model, predicted presence peaked at 0.1 to 1 ppt, while the small pelagic fish community showed a wider peak, likely reflecting species diversity. For example, *D. petenense* and juvenile *A. sapidissima* prefer freshwater (Feyrer et al. 2007, O'Connor et al. 2012), while juvenile *Spirinchus thaleichthys* and *Menidia audens* tolerate higher salinities (Kimmerer et al. 2009, Mahardja et al. 2016).

Across all 3 presence–absence models and the small pelagic fish CPUV model, there was a positive influence of outflow (a 14 d backward moving average of outflow). Freshwater outflow actions have been utilized as part of the management actions aimed at benefitting *H. transpacificus* by improving abiotic

habitat and providing subsidies of zooplankton prey from upstream (Frantzich et al. 2021, Hassrick et al. 2023, Lee et al. 2023). However, results have been mixed: flow actions did not always result in increased *H. transpacificus* abundances (FLOAT-MAST 2020). Our results support previous research and modeling efforts arguing that increased outflow can benefit *H. transpacificus* (Feyrer et al. 2011, Castillo 2019). While the focus has been on increasing and improving habitat for *H. transpacificus*, our results suggest that increased outflow—or perhaps more accurately, the broader set of Bay-Delta conditions associated with wetter years (Mahardja et al. 2022)—benefits other small pelagic fish species as well. Further research is needed to better explain and explore the mechanisms related to outflow and how this benefits the small pelagic fish community, since the effects of increased outflow and the overall better conditions during higher precipitation years may be difficult to disentangle. For example, wetter conditions and higher flows tend to be associated with cooler water temperatures from spring through fall which likely benefits pelagic fishes, independently of habitat and food considerations (Bashevkin & Mahardja 2022).

Collectively, our results demonstrate a strong pattern of association between the threatened *H. transpacificus* and a group of relatively abundant small pelagic fish species, suggesting the potential for use of surrogate species to improve the predictive power of ecological models and facilitate more reliable evaluation of management actions. However, use of surrogate species in endangered species conservation and management is not without challenges. Caro et al. (2005) examined the use of surrogate species using a theoretical framework, concluding that in an ideal scenario, a surrogate species and the target species of interest will have a similar disturbance threshold; that is, the population growth rate for both species will move from positive to negative with a similar magnitude of a given disturbance. The authors then describe the information to evaluate suitability of a surrogate, which include (1) the relationship between the disturbance, the demographic response, and the disturbance threshold, (2) the key traits or set of traits that is linked to the disturbance threshold, and (3) the relationship between the key trait of the target and surrogate species (e.g. thermal tolerance).

Although the ideal scenario is unlikely to be achieved in all cases for *H. transpacificus* and its potential surrogates, because of multiple disturbances and diversity within the small pelagic community, decades of published research (e.g. Mac Nally et al. 2010, Moyle

et al. 2016), and long-term monitoring data (Bashevkin et al. 2022) on the Bay-Delta pelagic fish community can facilitate further investigation of surrogate suitability for evaluating specific disturbances. Caro et al. (2005) identify 3 specific scenarios where using a substitute species may fail: when the target species is under greater disturbance than the surrogate species, when no relationship exists between the disturbance threshold of the surrogate species and the key trait being measured, and when the relationship between the disturbance threshold and key trait no longer holds in the surrogate species or is different compared to the target species. To help avoid these potential pitfalls, we recommend that prior to use of surrogate species in any formal management or decision-making process, analyses like those presented here be extended to include publicly available data from the multiple long-term pelagic fish monitoring programs conducted by the California Department of Fish and Wildlife and others. The durability of these relationships across a longer time period and the generalizability of the relationships across subregions within the Bay-Delta should be evaluated as a logical next step.

Further research should not preclude using surrogate species to inform near-term *H. transpacificus* recovery needs. The impacts of many management actions intended to recover *H. transpacificus* remain uncertain due to infrequent detections. Considering impacts on the broader small pelagic fish community in re-evaluating past actions and monitoring ongoing actions should improve detection of meaningful impacts. An additional application of surrogate species could inform release locations for cultured *H. transpacificus* being introduced into the Bay-Delta. Identifying optimal release locations is challenging due to the dynamic nature of the Bay-Delta and multiple criteria influencing growth and survival. Releasing cultured *H. transpacificus* near high densities of other small pelagic fish could be trialed in an adaptive management framework, comparing performance with alternative criteria.

Using surrogate species introduces challenges and uncertainties, but infrequent detection is a pervasive problem in the study and management of rare species, obscuring effects of recovery actions and leading to management paralysis. Decision-makers may interpret a lack of sensitivity to detect impacts as ineffectiveness, resulting in reluctance to implement alternatives. Demonstrating measurable change is critical to recovery efforts, and surrogate species can provide a viable alternative to infrequent detections and high management uncertainty.

Acknowledgements. This work was conducted as part of the Interagency Ecological Program (IEP) annual work plan. Early reviews of related study plans were supported by the IEP Flow Alteration Project Work Team and the Collaborative Adaptive Management Team (special thanks to Larry Brown [USGS]), and the USFWS, California Department of Water Resources (DWR), CDFW, and US Bureau of Reclamation (USBR). Permitting for this work was facilitated through the CDFW, USFWS, and IEP. This study was supported by the contribution of multiple IEP agency field sampling activities of fish and zooplankton (CDFW, USFWS, DOP [USBR]), as well as informed by DWR DAYFLOW. Funding and contractual support was provided by the USBR (Contract number: GS-00F-010CA), DWR, State Water Contractors (special thanks to Jennifer Pierre), and State and Federal Contractors Water Agency (special thanks to Laura Valoppi). We acknowledge the support of many people from the Delta science community, including support and input from USBR staff Brian Mahardja, Lisa Elliot, Kristen Arend, Michael Beakes, and Andrew Schultz. Contributors from ICF include Colin Brennan, Rob Miller, Athena Maguire, Andrew Winston, Ramona Zeno, Jake Sousa, Angelina Ravani, Tim Carrara, Katherine Maniscalco, Donna Maniscalco, Eric Chapman, Stephanie Owens, Teague Corning, Crystal Garcia, and Thomas Quinn. We thank Catherine Johnston and Lara Mitchell (USFWS) for timely guidance on site selection, coordination, and data; April Hennessy, Steve Slater, and Christina Burdi (CDFW) for assistance with zooplankton and data; and Tom Philippi (National Park Service) for sampling and GRTS advice.

LITERATURE CITED

- Abeare S (2009) Comparisons of boosted regression tree, GLM and GAM performance in the standardization of yellowfin tuna catch-rate data from the Gulf of Mexico online [sic] fishery. MSc thesis, Louisiana State University and Agricultural & Mechanical College, Baton Rouge, LA
- ✦ Bacher D (2022) Extinction in the wild? Zero delta smelt found in survey for seventh September in a row. Sacramento News & Review, 8 November 8 2022. <https://sacramento.newsreview.com/2022/11/08/extinction-in-the-wild-zero-delta-smelt-found-in-survey-for-seventh-september-in-a-row/>
- ✦ Barletta M, Barletta-Bergan A, Saint-Paul U, Hubold G (2005) The role of salinity in structuring the fish assemblages in a tropical estuary. *J Fish Biol* 66:45–72
- ✦ Bashevkin SM, Mahardja B (2022) Seasonally variable relationships between surface water temperature and inflow in the upper San Francisco Estuary. *Limnol Oceanogr* 67: 684–702
- ✦ Bashevkin SM, Gaeta JW, Nguyen TX, Mitchell L, Khanna S (2022) Fish abundance in the San Francisco Estuary (1959-2021), an integration of 9 monitoring surveys, ver 1. Environmental Data Initiative, doi:10.6073/pasta/0cdf7e5e954be1798ab9bf4f23816e83 (Accessed 31 Jan 2024)
- ✦ Bayse SM, Shaughnessy CA, Regish AM, McCormick SD (2020) Upper thermal tolerance and heat shock protein response of juvenile American shad (*Alosa sapidissima*). *Estuaries Coasts* 43:182–188
- ✦ Belbachir F, Pettorelli N, Wachter T, Belbachir-Bazi A, Durant SM (2015) Monitoring rarity: the critically endangered Saharan cheetah as a flagship species for a threatened ecosystem. *PLOS ONE* 10:e0115136

- Bever AJ, MacWilliams ML, Herbold B, Brown LR, Feyrer FV (2016) Linking hydrodynamic complexity to delta smelt (*Hypomesus transpacificus*) distribution in the San Francisco Estuary, USA. *San Franc Estuary Watershed Sci* 14(1):3
- Brown LR, Michniuk D (2007) Littoral fish assemblages of the alien-dominated Sacramento-San Joaquin Delta, California, 1980–1983 and 2001–2003. *Estuaries Coasts* 30:186–200
- CDFW (California Department of Fish and Wildlife) (2025) Delta smelt indices. <https://apps.wildlife.ca.gov/Townet/Main/DeltaSmeltIndices> (Accessed 30 Jan 2025)
- Camacho-Cervantes M, Ojanguren AF, Deacon AE, Ramnarine IW, Magurran AE (2014) Association tendency and preference for heterospecifics in an invasive species. *Behaviour* 151:769–780
- Caro TIM, Eadie J, Sih A (2005) Use of substitute species in conservation biology. *Conserv Biol* 19:1821–1826
- Castillo GC (2019) Modeling the influence of outflow and community structure on an endangered fish population in the upper San Francisco Estuary. *Water* 11:1162
- Chrétien E, Boisclair D, Cooke SJ, Killen SS (2021) Social group size and shelter availability influence individual metabolic traits in a social fish. *Integr Org Biol* 3:obab032
- Cianfrani C, Le Lay G, Hirzel AH, Lo A (2010) Do habitat suitability models reliably predict the recovery areas of threatened species? *J Appl Ecol* 47:421–430
- Conrad JL, Bibian AJ, Weinersmith KL, De Carion D and others (2016) Novel species interactions in a highly modified estuary: association of largemouth bass with Brazilian waterweed *Egeria densa*. *Trans Am Fish Soc* 145: 249–263
- D'Amen M, Smeraldo S, Azzurro E (2023) Salinity, not only temperature, drives tropical fish invasions in the Mediterranean Sea, and surface-only variables explain it better. *Coral Reefs* 42:467–472
- Davis BE, Hansen MJ, Cocherell DE, Nguyen TX and others (2019) Consequences of temperature and temperature variability on swimming activity, group structure, and predation of endangered delta smelt. *Freshw Biol* 64: 2156–2175
- Davis BE, Adams JB, Lewis LS, Hobbs JA and others (2022) Wakasagi in the San Francisco Bay–Delta watershed: comparative trends in distribution and life-history traits with native delta smelt. *San Franc Estuary Watershed Sci* 20(3):2
- Duarte A, Peterson JT (2021) Space-for-time is not necessarily a substitution when monitoring the distribution of pelagic fishes in the San Francisco Bay-Delta. *Ecol Evol* 11:16727–16744
- Dunn JC, Buchanan GM, Cuthbert RJ, Whittingham MJ, McGowan PJ (2015) Mapping the potential distribution of the Critically Endangered Himalayan quail *Ophrysia superciliosa* using proxy species and species distribution modelling. *Bird Conserv Int* 25:466–478
- DWR, USBR (California Department of Water Resources, Bureau of Reclamation) (2023) Delta Smelt Summer-Fall Habitat Action. <https://deltacouncil.ca.gov/pdf/science-program/background-documents/2024-02-28-2023-summer-fall-habitat-action-plan.pdf>
- Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. *J Anim Ecol* 77:802–813
- Ferrari MC, Ranåker L, Weinersmith KL, Young MJ, Sih A, Conrad JL (2014) Effects of turbidity and an invasive waterweed on predation by introduced largemouth bass. *Environ Biol Fishes* 97:79–90
- Feyrer F, Herbold B, Matern SA, Moyle PB (2003) Dietary shifts in a stressed fish assemblage: consequences of a bivalve invasion in the San Francisco Estuary. *Environ Biol Fishes* 67:277–288
- 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* 64:723–734
- Feyrer F, Newman K, Nobriga M, Sommer T (2011) Modeling the effects of future outflow on the abiotic habitat of an imperiled estuarine fish. *Estuaries Coasts* 34:120–128
- FLOAT-MAST (Flow Alteration - Management, Analysis, and Synthesis Team) (2020) Synthesis of data and studies relating to delta smelt biology in the San Francisco Estuary, emphasizing water year 2017. IEP Technical Report 95. Interagency Ecological Program, Sacramento, CA
- Frantzich J, Davis BE, MacWilliams M, Bever A, Sommer T (2021) Use of a managed flow pulse as food web support for estuarine habitat. *San Franc Estuary Watershed Sci* 19(3):3
- Greenwell B (2021) fastshap: fast approximate Shapley values. R package version 0.0.7. <https://CRAN.R-project.org/package=fastshap>
- Greenwell B, Boehmke B, Cunningham J, Developers G (2020) gbm: generalized boosted regression models. R package version 2.1.8. <https://CRAN.R-project.org/package=gbm>
- Grimaldo LF, Smith WE, Nobriga ML (2021) Re-examining factors that affect delta smelt (*Hypomesus transpacificus*) entrainment at the State Water Project and Central Valley Project in the Sacramento–San Joaquin Delta. *San Franc Estuary Watershed Sci* 19(1):5
- Halverson GH, Lee CM, Hestir EL, Hulley GC and others (2022) Decline in thermal habitat conditions for the endangered delta smelt as seen from Landsat satellites (1985–2019). *Environ Sci Technol* 56:185–193
- Hamilton SA, Murphy DD (2018) Analysis of limiting factors across the life cycle of delta smelt (*Hypomesus transpacificus*). *Environ Manage* 62:365–382
- Hamilton SA, Murphy DD (2020) Use of affinity analysis to guide habitat restoration and enhancement for the imperiled delta smelt. *Endang Species Res* 43:103–120
- Hamilton SH, Pollino CA, Jakeman AJ (2015) Habitat suitability modelling of rare species using Bayesian networks: model evaluation under limited data. *Ecol Modell* 299:64–78
- Hasenbein M, Fangué NA, Geist JP, Komoroske LM, Connon RE (2016) Physiological stress biomarkers reveal stocking density effects in late larval delta smelt (*Hypomesus transpacificus*). *Aquaculture* 450:108–115
- Hassrick JL, Korman J, Kimmerer WJ, Gross ES, Grimaldo LF, Lee C, Schultz AA (2023) Freshwater flow affects subsidies of a copepod (*Pseudodiaptomus forbesi*) to low-salinity food webs in the upper San Francisco Estuary. *Estuaries Coasts* 46:450–462
- Hendrix AN, Fleishman E, Zillig MW, Jennings ED (2023) Relations between abiotic and biotic environmental variables and occupancy of delta smelt (*Hypomesus transpacificus*) in autumn. *Estuaries Coasts* 46:149–165
- Hestir EL, Schoellhamer DH, Greenberg J, Morgan-King T, Ustin SL (2016) The effect of submerged aquatic vegetation expansion on a declining turbidity trend in the Sacramento-San Joaquin River Delta. *Estuaries Coasts* 39: 1100–1112
- Hijmans RJ, Phillips S, Leathwick J, Elith J (2021) dismo:

- species distribution modeling. R package version 1.3-5. <https://CRAN.R-project.org/package=dismo>
- Hopkins RL (2009) Use of landscape pattern metrics and multiscale data in aquatic species distribution models: a case study of a freshwater mussel. *Landsc Ecol* 24: 943–955
- Hosmer DW Jr, Lemeshow S, Sturdivant RX (2013) *Applied logistic regression*. John Wiley & Sons, Hoboken, NJ
- Ingram W, Ziebell CD (1983) Diet shifts to benthic feeding by threadfin shad. *Trans Am Fish Soc* 112:554–556
- Jeliaskov A, Gavish Y, Marsh CJ, Geschke J and others (2022) Sampling and modelling rare species: conceptual guidelines for the neglected majority. *Glob Change Biol* 28:3754–3777
- Kayfetz KS, Bashevkin SM, Thomas M, Hartman R and others (2020) Zooplankton integrated dataset metadata report. IEP Technical Report 93. California Department of Water Resources, Sacramento, CA
- Kimmerer WJ, Gross ES, MacWilliams ML (2009) Is the response of estuarine nekton to freshwater flow in the San Francisco Estuary explained by variation in habitat volume? *Estuaries Coasts* 32:375–389
- Kuhn M (2022) caret: classification and regression training. R package version 6.0-93. <https://CRAN.R-project.org/package=caret>
- Lacy JR, Foster-Martinez MR, Allen RM, Drexler JZ (2021) Influence of invasive submerged aquatic vegetation (*E. densa*) on currents and sediment transport in a freshwater tidal system. *Water Resour Res* 57:p.e2020WR 028789
- Lee CY, Smith AG, Hassrick JL, Kalmbach AJ and others (2023) Flow augmentations modify an estuarine prey field. *San Franc Estuary Watershed Sci* 21(2):1
- Lenoach LK, Stumpner PR, Burau JR, Loken LC, Sadro S (2021) Dispersion and stratification dynamics in the upper Sacramento River Deep Water Ship Channel. *San Franc Estuary Watershed Sci* 19(4):5
- Lessard J, Cavallo B, Anders P, Sommer T and others (2018) Considerations for the use of captive-reared delta smelt for species recovery and research. *San Franc Estuary Watershed Sci* 16(3):3
- Loman ZG, Deluca WV, Harrison DJ, Loftin CS, Schwenk WS, Wood PB (2021) How well do proxy species models inform conservation of surrogate species? *Landsc Ecol* 36:2863–2877
- Mac Nally R, Thomson JR, Kimmerer WJ, Feyrer F and others (2010) Analysis of pelagic species decline in the upper San Francisco Estuary using multivariate autoregressive modeling (MAR). *Ecol Appl* 20:1417–1430
- Mahardja B, Conrad JL, Lusher L, Schreier B (2016) Abundance trends, distribution, and habitat associations of the invasive Mississippi silverside (*Menidia audens*) in the Sacramento–San Joaquin Delta, California, USA. *San Franc Estuary Watershed Sci* 14(1):2
- Mahardja B, Young MJ, Schreier B, Sommer T (2017) Understanding imperfect detection in a San Francisco Estuary long-term larval and juvenile fish monitoring programme. *Fish Manag Ecol* 24:488–503
- Mahardja B, Bashevkin SM, Pien C, Nelson M, Davis BE, Hartman R (2022) Escape from the heat: thermal stratification in a well-mixed estuary and implications for fish species facing a changing climate. *Hydrobiologia* 849: 2895–2918
- McKenzie R, Mahardja B (2021) Evaluating the role of boat electrofishing in fish monitoring of the Sacramento–San Joaquin Delta. *San Franc Estuary Watershed Sci* 19(1):4
- Merz JE, Hamilton S, Bergman PS, Cavallo B (2011) Spatial perspective for delta smelt: a summary of contemporary survey data. *Calif Fish Game* 97:164–189
- Merz JE, Bergman PS, Simonis JL, Delaney D, Pierson J, Anders P (2016) Long-term seasonal trends in the prey community of delta smelt (*Hypomesus transpacificus*) within the Sacramento–San Joaquin Delta, California. *Estuaries Coasts* 39:1526–1536
- Monirian J, Sutphin Z, Myrick C (2010) Effects of holding temperature and handling stress on the upper thermal tolerance of threadfin shad *Dorosoma petenense*. *J Fish Biol* 76:1329–1342
- Moyle PB (2002) *Inland fishes of California: revised and expanded*. University of California Press, Berkeley, CA
- Moyle PB, Brown LR, Durand JR, Hobbs JA (2016) Delta smelt: life history and decline of a once-abundant species in the San Francisco Estuary. *San Franc Estuary Watershed Sci* 14(2):6
- Nadler LE, Killen SS, McClure EC, Munday PL, McCormick MI (2016) Shoaling reduces metabolic rate in a gregarious coral reef fish species. *J Exp Biol* 219:2802–2805
- O'Connor MP, Juanes F, McGarigal K, Caris J (2012) Describing juvenile American shad and striped bass habitat use in the Hudson River Estuary using species distribution models. *Ecol Eng* 48:101–108
- Ostrand KG, Wilde GR (2001) Temperature, dissolved oxygen, and salinity tolerances of five prairie stream fishes and their role in explaining fish assemblage patterns. *Trans Am Fish Soc* 130:742–749
- Pavlov DS, Kasumyan AO (2000) Patterns and mechanisms of schooling behavior in fish: a review. *J Ichthyol* 40(Suppl 2):S163–S231
- Peterson JT, Barajas MF (2018) An evaluation of three fish surveys in the San Francisco Estuary, 1995–2015. *San Franc Estuary Watershed Sci* 16(4):2
- Pianka ER (1974) Niche overlap and diffuse competition. *Proc Natl Acad Sci USA* 71:2141–2145
- Rodríguez MA, Marselli G, Mandrak NE (2021) Responses of vulnerable fishes to environmental stressors in the Canadian Great Lakes basin. *Can J Fish Aquat Sci* 78:1278–1292
- Rushton SP, Ormerod SJ, Kerby G (2004) New paradigms for modelling species distributions? *J Appl Ecol* 41:193–200
- Schultz AA, Kalmbach AJ (2022) Directed outflow project lower trophic study, ver 1. Environmental Data Initiative, doi:10.6073/pasta/36508317ffdd2a5a9b961177bed9cea7 (Accessed 2 Dec 2022)
- Shaha DC, Ahmed S, Hasan J, Kundu SR and others (2022) Fish diversity in relation to salinity gradient in the Meghna River Estuary, Bangladesh. *Conservation* 2:414–434
- Shea K, Mangel M (2001) Detection of population trends in threatened coho salmon (*Oncorhynchus kisutch*). *Can J Fish Aquat Sci* 58:375–385
- Silvano AL, Guyer C, Steury TD, Grand JB (2017) Selecting focal species as surrogates for imperiled species using relative sensitivities derived from occupancy analysis. *Ecol Indic* 73:302–311
- Slater SB, Baxter RD (2014) Diet, prey selection, and body condition of age-0 delta smelt, *Hypomesus transpacificus*, in the Upper San Francisco Estuary. *San Franc Estuary Watershed Sci* 12(3):1
- Slater S, Schultz AA, Hammock B, Hennessy A, Burdi C (2019) Patterns of zooplankton consumption by juvenile and adult delta smelt (*Hypomesus transpacificus*). In:

- Schultz AA (ed) Directed Outflow Project: Technical Report 1. US Bureau of Reclamation, Bay-Delta Office, Mid-Pacific Region, Sacramento, CA, p 8–54
- Smith WE, Nobriga ML (2023) A bioenergetics-based index of habitat suitability: spatial dynamics of foraging constraints and food limitation for a rare estuarine fish. *Trans Am Fish Soc* 152:650–671
- Smith A, Hofner B, Lamb JS, Osenkowski J and others (2019) Modeling spatiotemporal abundance of mobile wildlife in highly variable environments using boosted GAMLSS hurdle models. *Ecol Evol* 9:2346–2364
- Sommer T, Reece K, Mejia F, Nobriga M (2009) Delta smelt life-history contingents: a possible upstream rearing strategy. *IEP Newsl* 22:11–13
- Sommer T, Hartman R, Koller M, Koohafkan M and others (2020) Evaluation of a large-scale flow manipulation to the upper San Francisco Estuary: response of habitat conditions for an endangered native fish. *PLOS ONE* 15: e0234673
- Starceovich LA, DiDonato G, McDonald T, Mitchell J (2016) A GRTS user's manual for the SDrawNPS package: a graphical user interface for generalized random tessellation stratified (GRTS) sampling and estimation. *Natural Resource Report NPS/PWRO/NRR-2016/1233*. National Park Service, Fort Collins, CO
- Stevens DL Jr, Olsen AR (2004) Spatially balanced sampling of natural resources. *J Am Stat Assoc* 99:262–278
- Stompe DK, Moyle PB, Kruger A, Durand JR (2020) Comparing and integrating fish surveys in the San Francisco Estuary: why diverse long-term monitoring programs are important. *San Franc Estuary Watershed Sci* 18(2):4
- Stupariu MS, Cushman SA, Pleşoianu AI, Pătru-Stupariu I, Fuerst C (2022) Machine learning in landscape ecological analysis: a review of recent approaches. *Landsc Ecol* 37:1227–1250
- Swanson C, Reid T, Young PS, Cech JJ Jr (2000) Comparative environmental tolerances of threatened delta smelt (*Hypomesus transpacificus*) and introduced wakasagi (*H. nipponensis*) in an altered California estuary. *Oecologia* 123:384–390
- Thiel R, Sepulveda A, Kafemann R, Nellen W (1995) Environmental factors as forces structuring the fish community of the Elbe Estuary. *J Fish Biol* 46:47–69
- Thompson W (ed) (2013). *Sampling rare or elusive species: concepts, designs, and techniques for estimating population parameters*. Island Press, Washington, DC
- Thomson JR, Kimmerer WJ, Brown LR, Newman KB and others (2010) Bayesian change point analysis of abundance trends for pelagic fishes in the upper San Francisco Estuary. *Ecol Appl* 20:1431–1448
- Tuia D, Kellenberger B, Beery S, Costelloe BR and others (2022) Perspectives in machine learning for wildlife conservation. *Nat Commun* 13:792
- USFWS (US Fish and Wildlife Service) (2019) *Biological opinion for the reinitiation of cConsultation on the coordinated operations of the Central Valley Project and State Water Project*. US Fish and Wildlife Service, Sacramento, CA
- USFWS, Senegal T, Mckenzie R, Speegle J and others (2022) *Interagency Ecological Program and US Fish and Wildlife Service: San Francisco Estuary Enhanced Delta Smelt Monitoring Program data, 2016-2021, ver 8*. Environmental Data Initiative, doi:10.6073/pasta/e1a540c161b7be56b941df50fd7b44c5 (Accessed 17 Feb 2023)
- Wang HH, Wonkka CL, Treglia ML, Grant WE, Smeins FE, Rogers WE (2015) Species distribution modelling for conservation of an endangered endemic orchid. *AoB Plants* 7:plv039
- Winder M, Jassby AD (2011) Shifts in zooplankton community structure: implications for food web processes in the upper San Francisco Estuary. *Estuaries Coasts* 34:675–690
- Winter E (2002) The Shapley value. In: Aumann R, Hart S (eds) *Handbook of game theory with economic applications*, Vol 3. Elsevier, Amsterdam, p 2025–2054

*Editorial responsibility: Eric Gilman,
Honolulu, Hawaii, USA*

Reviewed by: 2 anonymous referees

Submitted: August 9, 2024; Accepted: December 20, 2024

Proofs received from author(s): February 17, 2025

This article is Open Access under the Creative Commons by Attribution (CC-BY) 4.0 License, <https://creativecommons.org/licenses/by/4.0/deed.en>. Use, distribution and reproduction are unrestricted provided the authors and original publication are credited, and indicate if changes were made