Modeling of Delta Smelt Distribution and Food

Technical Memorandum to SDM Technical Work Group

Scott Hamilton, Draft 12-15-22

Introduction

The factors influencing the distribution of delta smelt are numerous and include turbidity, temperature, salinity, food and region (Bever et al 2016, Latour 2016, Mahardja et al. 2017, Petersen and Barajas 2018, Polansky et al 2018, Simonis and Merz 2019, Hamilton and Murphy 2020, Hendrix et al. 2022). To that list we hypothesized that distribution was also influenced by prior distribution. Similarly, the factors influencing the availability of food in a region are influenced by water temperature, salinity, flows, prior abundance in that region and prior abundance in upstream locations (Hamilton et al (2020). Within CSAMP's structured decision making process for delta smelt, being implemented by Compass Resource Management, numerous actions are being considered that could modify one or more of the environmental factors influencing the distribution of the fish or its food. For example, a food action in one location could cause food availability to increase in downstream locations as copepods are disbursed by flows, and in turn that could influence the distribution of delta smelt. To reasonably assess the effects of the proposed actions on delta smelt, the influence of the actions on food availability ad distribution need to be assessed. Towards that end, two models have been developed a "food mode; and a "Distribution Model"

Methods

Food Model

The model estimating food availability by month and region ("food model") is an enhancement of the model of Hamilton et al (2020). It is a series of equations that estimate the adult calanoid biomass in each region for each month from spring through autumn. The 2020 model was expanded to include the months of March and November, salinity was added as an explanatory variable, and the model was refit using 20mm survey data for the months of March through July for the period 1995 to 2014. The use of 20mm data allowed for modeling locations in Yolo Bypass, Rio Vista and Northeast Suisun Bay – locations that could not be modelled using zooplankton data due to lack of surveys in those regions.

Equations were estimated for each month and region and had the general form:

$$A_{r,m} = a + b_1A_{r,m-1} + b_2A_{r-1,m} + b_3F_{r,m} + b_4F^2_{r,m} + b_5T_{r,m} + b_6T^2_{r,m} + b_7E_{r,m} \quad (1)$$

Where: $A_{r,m}$ is the biomass of adult calanoid copepods (μC/m³) in region r in month m, m-1 denotes the prior month, r-1 denotes the upstream region, T is temperature in degrees C and E is electrical conductivity (μS/cm). Covariates where data were not available for the month or the region were excluded. If data on covariates were missing in some years, those years were excluded when estimating coefficients. If two upstream regions could affect biomass, both upstream regions were included. The equations were estimated using data from the period 1995 to 2014 using both 20mm data for March to July and zooplankton survey data for July through November. The preferred equations for July were selected based on goodness of fit and plausibility of coefficients. The equations were estimated by fitting coefficients that produced the minimum residual sum of squares, using the GRG nonlinear routine in Excel (Solver). Large coefficients, typically a result of overfitting, were found to create spurious results. To reduce the likelihood of overfitting, the absolute value of the sum of the coefficients were constrained to be less than twice the values of the mean of the dependent variable. Coefficients for prior and upstream abundances were constrained to be greater than or equal to zero because such influences were believed to only have a positive influence on biomass at a particular place and time.

Equations[1] were estimated for 14 regions using 20mm data: Stockton, Mid San Joaquin River, Lower San Joaquin River, Old River, Franks Tract, Yolo Bypass, Upper Sacramento River (Rio Vista), Lower Sacramento, Confluence, NE Suisun, SE Suisun, NW Suisun (Grizzly Bay), SW Suisun, and Suisun Marsh (Montezuma Slough). Zooplankton data was used to model 11 regions (the same regions but excluding Yolo, Rio Vista and NE Suisun). In total 123 equations (region-month combinations) were estimated.

The resulting equations were incorporated into a model in an excel workbook. In that workbook, the IBMR input sheets are copied and pasted into their own dedicated worksheets, the change in covariates resulting from the action are calculated, and the modified covariate data are used as inputs, as appropriate, into the equations for each location and month. Those equations generate new estimates of food availability in each region and month. By design, prior and upstream abundances influence food availability in a month, along with temperature, salinity and flow. Therefore, changes in prey availability upstream have the potential to influence prey availability downstream in later months. The results are compiled into a new prey -availability table that has a format identical to the IBMR food input data, so the results can be cut and pasted to an IBMR input file.

In addition to calculating the goodness of fit (R^2) for each region-month equation, model validity was assessed by calculating R^2 for the years 2015 to 2020.

Distribution Model

Equations to explain the distribution of delta smelt were generated for each month and region and had the general form:

$$DS_r = R_r + f_{r,}.F + a.P_r^{\,b}.T_r^{\,c}.S_{r,}^{\,d}.C_r^{\,e} + gD_r + h_rOMR \qquad (2)$$

where DS_r is the observed percentage of delta smelt in region *r*, R is a constant specific to region *r* intended to capture influence of the physical landscape features unique to each

[1] The coefficients for the region-month equations were estimated in the file CBA (Calanoid Biomass Analysis). Those coefficients are then used in a separate set of files (SDM *Scenario #*) where a new food distribution is estimated for each scenario.

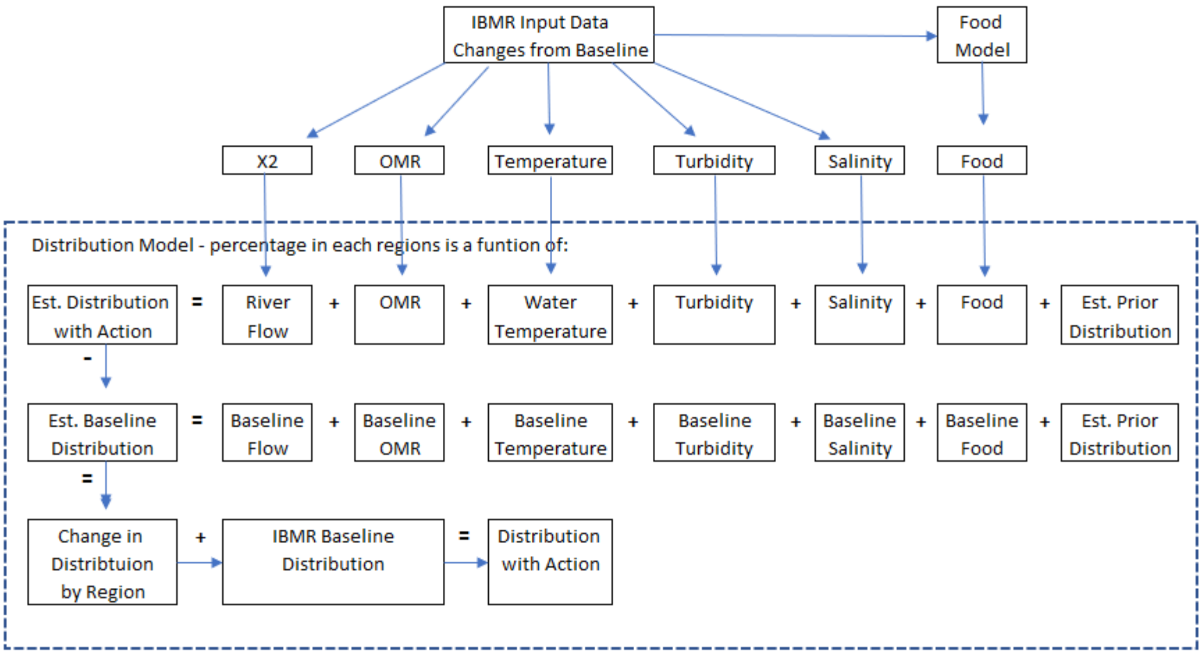
region, F is log of flow (cfs) during the month, P_r is the affinity value for prey (μgC/m^3) in region *r*, T_r is the affinity value for average turbidity (Secchi depth in cm) in region *r*, S_r is the affinity value for salinity (electrical conductivity (μS/cm) in region *r*, C_r is the affinity value for temperature (°C) in region *r*, D is the observed percentage of delta smelt in region *r* in the prior month, and OMR is average flow in Old and Middle rivers during the month (cfs). For simplicity, the subscripts to denote each year and month have been omitted. The letters *a* through *h*, and *R*, are coefficients to be estimated. The monthly models were designed so that OMR flows could potentially influence distribution directly in the Confluence, Lower Rivers, East Delta and South Delta subregions.

Data needed to estimate the coefficients in equation 2 were obtained from a variety of sources. Data on delta smelt abundance and abiotic conditions were obtained from fish surveys --Midwater Trawl and Spring Kodiak trawl for January to March, 20mm Survey for April and May, Summer Tow Net survey for June and July supplemented by the 20mm, where values were missing, Summer Tow Net for August, and Fall Midwater Trawl for September through December. Data were converted to affinity values (see Hamilton and Murphy 2020 for data sources and derivation of affinity values). Flow data for Old and Middle rivers were obtained from CDWR's Dayflow and from the California Data Exchange Center (CDEC). The period of analysis was 1990 to 2014; missing data prevented some years from being included in the analysis.

The covariate $H_{r,m}$, being binary, prevents occupancy when environmental conditions make an area uninhabitable. We defined "uninhabitable" conditions to be those where on average, less than 0.5% of delta smelt occurrences were recorded. Arbitrarily, 0.5% was selected as the threshold rather than 0%, understanding that a small percent of the fish may be in uninhabitable conditions because they may be dying or diseased, incapable of escaping, uninhabitable conditions, or are moving through, but not resident in, uninhabitable conditions.

The monthly models were developed in Microsoft Excel using the generalized reduced-gradient, non-linear optimization routine, Solver, to minimize the residual sum of squares between the predicted distribution in each region of each year of the study and the distribution derived from survey data. Employing that approach permits constraints to be placed on coefficients to ensure consistency with ecological theory. The predicted distribution of delta smelt across regions must necessarily sum to 100% so the estimates for each month in each region were adjusted proportionally and the sum across regions was equal to 100%.

Figure 1. *Schematic for modeling changes in distribution of delta smelt as a result of a management action*.



In addition to reporting the explanatory power (R^2) for each month in the preferred set of monthly models, the percentage of times absence was correctly predicted is reported.

The model validity was tested by predicting distribution in the ten years prior to the study period. Only the Summer Townet Survey and the Fall Midwater Trawl had data from 1980 to 1989 limiting the number of months available for validation to the period from July through December.

Results

Food Model

The goodness of fit for the region-month equations varied from an R^2 of 0.14 (November in Suisun Slough) to 0.97 (April in the Lower SJR) with an average across all regions and months of 0.60 (Table 1). Generally, the equations using 20mm data provided better fits that the equations using zooplankton data, and generally better fits were obtained for regions east of the Confluence compared to west of the Confluence.

Table 1. *R^2 Values for region-month models using 1995 to 2014 data. . Values are color coded from red (lowest) to dark green (best). Blanks indicate that data were unavailable to develop a model for the region and month.*

	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Average
Stockton	0.69	0.77	0.60	0.84	0.77	0.66	0.72	0.53	0.53	0.68
Mid SJR	0.78	0.91	0.82	0.60	0.60	0.85	0.82	0.50	0.91	0.75
Old River	0.57	0.83	0.82	0.70	0.41	0.45	0.65	0.57	0.42	0.60
Franks Tract	0.89	0.94	0.69	0.52	0.64	0.55	0.78	0.85	0.87	0.75
Lower SJR	0.83	0.97	0.89	0.81	0.84	0.24	0.27	0.32	0.37	0.62
North Delta (Yolo)	0.78	0.61	0.65	0.16	0.40					0.52
Upper Sacramento	0.56	0.31	0.68	0.59	0.26					0.48
Lower Sacramento	0.75	0.90	0.91	0.83	0.58	0.23	0.60	0.53	0.16	0.61
Confluence	0.89	0.77	0.91	0.79	0.72	0.83	0.57	0.49	0.62	0.73
NE Suisun	0.42	0.70	0.83	0.66	0.81					0.68
SE Suisun	0.17	0.52	0.79	0.51	0.73	0.56	0.31	0.42	0.20	0.47
NW Suisun	0.74	0.59	0.52	0.67	0.54	0.50	0.43	0.44	0.32	0.53
SW Suisun	0.60	0.64	0.64	0.70	0.36	0.42	0.17	0.15	0.43	0.46
Montezuma Slough	0.52	0.94	0.80	0.72	0.91	0.81	0.52	0.42	0.37	0.67
Suisun Slough	0.45	0.27	0.47	0.49	0.77	0.50	0.16	0.23	0.14	0.39
Average	0.64	0.71	0.73	0.64	0.62	0.55	0.50	0.46	0.45	0.60

In applying the region-month equations to the validation data set, the equations derived from the 20mm data again provided better explanatory power (Table 2). Note that Table 2 presents correlations (R values, not R^2 values).

Table 2. *R Values averaged over months for each region and survey when estimating 2015-2020 adult calanoid biomass (validation data set) for each month and region. Values are color coded from red (lowest) to dark green (best). Blanks indicate that data were unavailable to develop or test an equation for the region and month.*

Region	20mm	Zooplankton
Stockton	0.88	0.49
Mid SJR	0.88	0.39
Old River	0.69	0.35
Franks Tract	0.88	n.a.
Lower SJR	0.92	0.47
North Delta (Yolo)	0.84	
Upper Sacramento	0.77	
Lower Sacramento	0.88	0.23
Confluence	0.89	0.75
NE Suisun	0.72	
SE Suisun	0.72	0.73
NW Suisun	0.73	0.57
SW Suisun	0.65	0.46
Montezuma Slough	0.75	0.68
Suisun Slough		0.76
Average	0.80	0.53

Distribution Model

The explanatory power of the monthly models (R^2) varied between months with the highest value in June (R^2=0.89) and lowest in December (R^2=0.50) with the R^2 for most months falling between 0.5 and 0.65 (Figure 2). Confidence in the model is also influenced by the resuts of the validation tests. Validation results showed lower explanatory power with average R^2 across months decreasing from an average across months of 0.62 for the data set used to estimate the coefficients (1990-2014) to 0.47 when the estimated coefficients were applied to the validation data set (1980-1989) (Table 3). The ability to correctly predict absence did not change on average, remaining at 91%.

Figure 2. *R^2 for the preferred monthly models*

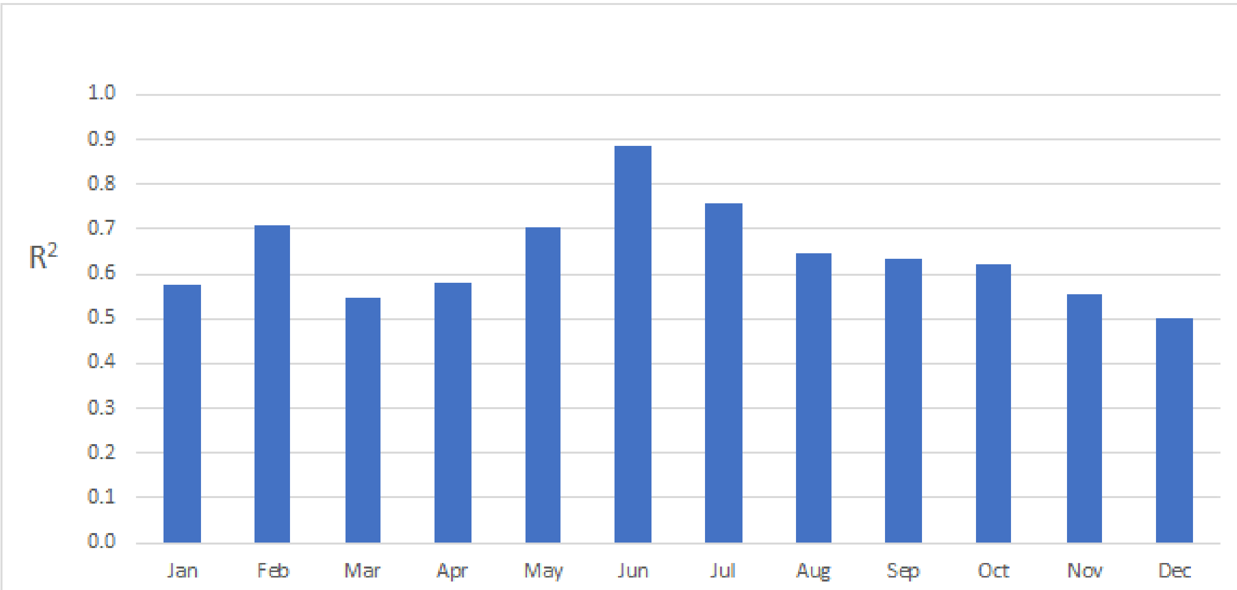


Table 3. *Results of the validation test for data from 1980 to 1989*

	Estimation Set		Validation Set		
	R^2	Absence Accuracy	Obs	R^2	Absence Accuracy
July	0.77	88%	48	0.52	80%
August	0.65	92%	32	0.40	81%
September	0.62	89%	38	0.44	90%
October	0.61	94%	77	0.33	100%
November	0.55	92%	74	0.46	92%
December	0.49	92%	16	0.65	100%
Average	0.61	91%		0.47	91%

An Example

Compass delineated 5 sensitivity runs for a summer X2 action (X2 summer low, X2 summer 1, X2 summer 2, X2 summer 3, and X2 summer high - runs 6.3 to 6.7). Summer_1 and Summer_3 are used here as examples to illustrate model application. The average X2 values for summer (June-August) for each scenario were: 72.1 and 77.3 respectively. In addition to modified X2 locations, the IBMR input sheets for these two scenarios also had salinity values that were modified from the baseline.

Implementing the food model generally indicated that Scenario 1 (the higher flow action) increased food availability in July compared to the baseline, with the increase generally diminishing with time. Only in one instance (in the Lower Sacramento River in September) was food availability lower than the baseline. Generally, Scenario 3 (the lower flow scenario) also showed increase in food above the baseline as a result of higher outflows under this scenario than the (historic) baseline.

The new food values for each region in each month of each year were inserted back into the IBMR input sheets to implement the distribution model. This model estimates distribution based on the factors listed in equation 2. In these examples flows, salinity and food values all differ from the baseline. The changes in distribution by region averaged over all years, are presented in Figure 4. In these scenarios, flows were modified in 13 of 20 years, so the actual changes in distribution in the years when the action was conducted would generally be higher than those shown in Figure 4. Scenario 1 shows a shift of approximately 10% of the population form above the Confluence to below the Confluence.

Figure 3. *Projected changes in food availability along the Sacramento River as projected by the food model for two summer X2 scenarios. Some regions have projections only through July because the models for those locations were derived from 20mm Survey data, and data were not available for later months.*

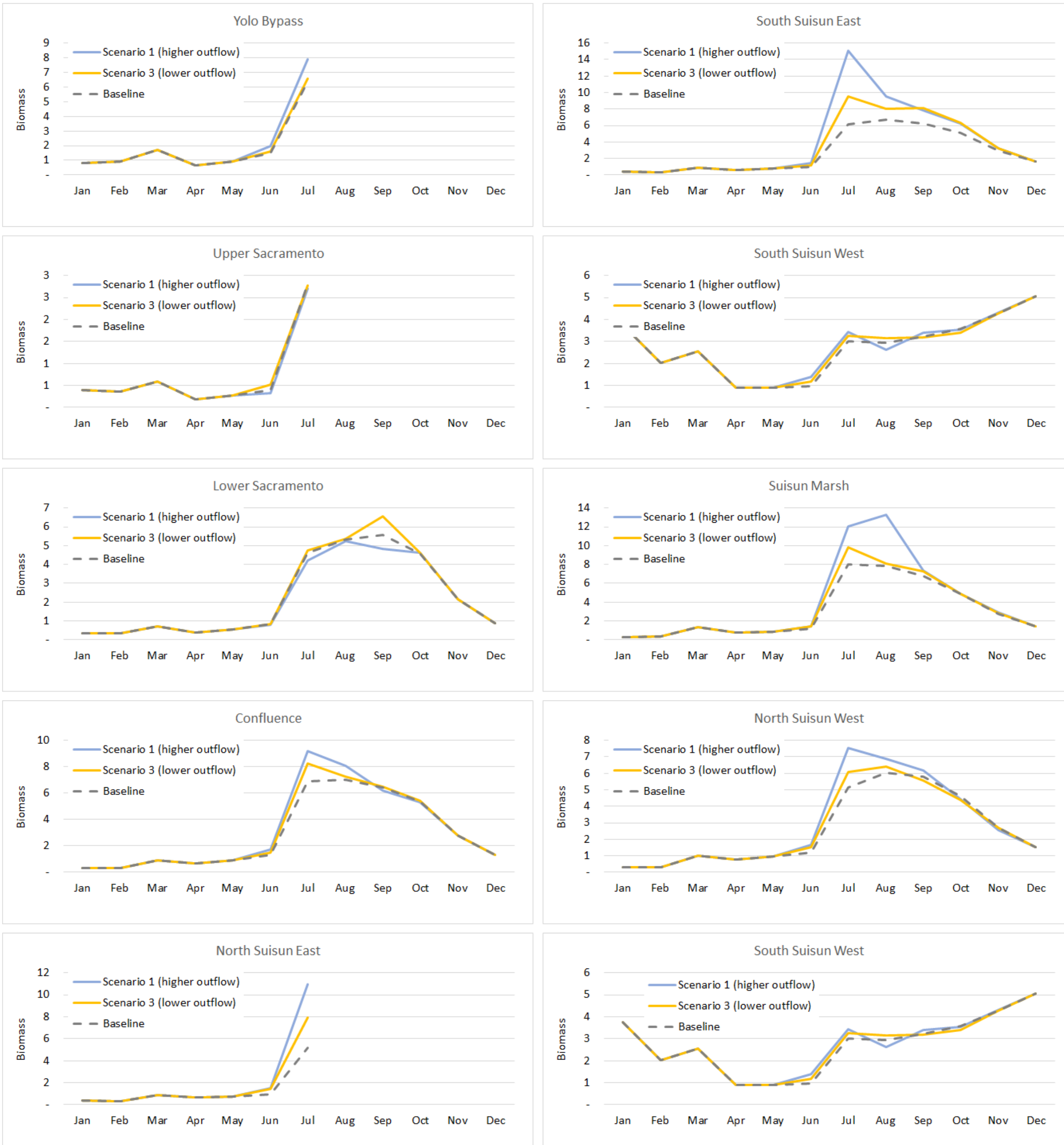
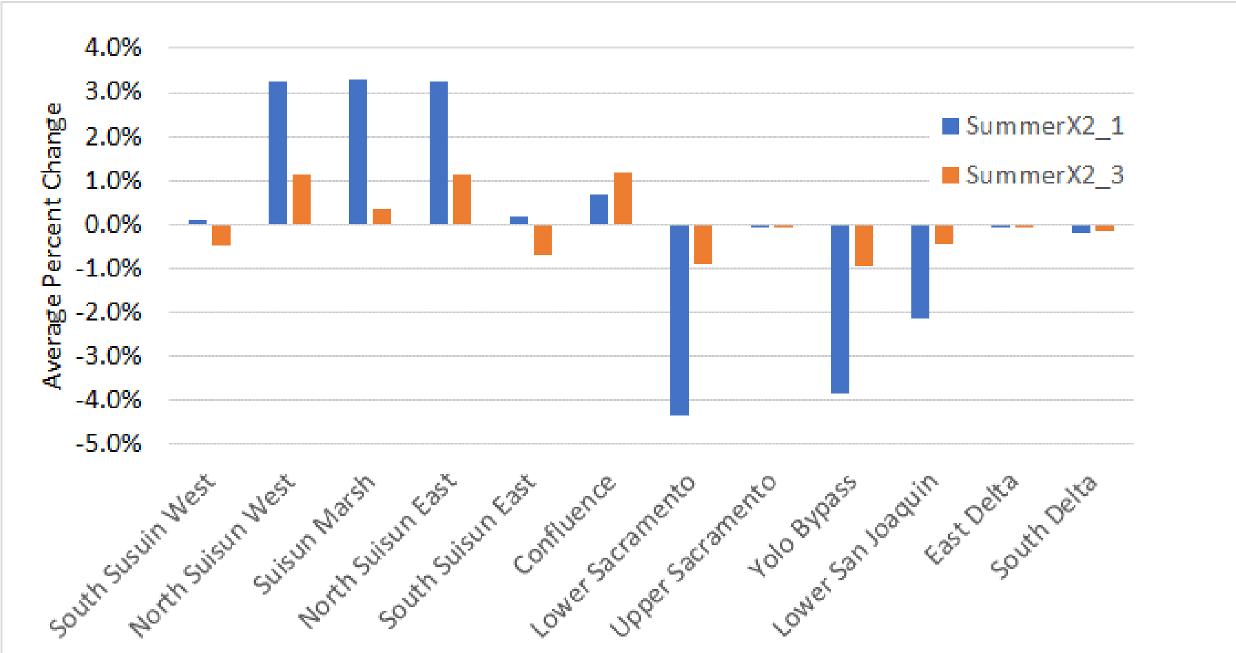


Figure 4. *Average Change in Distribution of Delta Smelt in Summer as projected by the distribution model for two summer X2 scenarios.*



References

Bever, A.J., MacWilliams, M.L., Herbold, B., Brown, L.R., and Feyrer, F. (2016). Linking hydrodynamic complexity to delta smelt (*Hypomesus transpacificus*) distribution in the San Francisco Estuary, USA. *San Francisco Estuary and Watershed Science* 14(1).
qt2x91q0fr.pdf (escholarship.org)

Hamilton SA, Murphy DD. (2018) Analysis of limiting factors across the life cycle of delta smelt (Hypomesus transpacificus). Environmental management. 62(2):365-82.
Analysis of Limiting Factors Across the Life Cycle of Delta Smelt (Hypomesus transpacificus) | SpringerLink

Hamilton, S. Bartell, S, Pierson, J, Murphy D. (2020) Factors Controlling Calanoid Copepod Biomass and Distribution in the Upper San Francisco Estuary and Implications for Managing the Imperiled Delta Smelt (*Hypomesus transpacificus*) Environmental Management 65:587–601 https://doi.org/10.1007/s00267-020-01267-8

Hamilton SA, Murphy DD. (2020) Use of affinity analysis to guide habitat restoration and enhancement for the imperiled delta smelt. Endangered Species Research. 24;43:103-20.
Use of affinity analysis to guide habitat restoration and enhancement for the imperiled delta smelt (int-res.com)

Latour, R.J. (2016). Explaining patterns of pelagic fish abundance in the Sacramento-San Joaquin Delta. *Estuaries and Coasts* 39, 233-247.
12237_2015_9968_Article 233..247 (researchgate.net)

Peterson, J.T., Barajas, M.F. (2018). An evaluation of three fish surveys in the San Francisco Estuary, California, 1995-2015. *San Francisco Estuary and Watershed Science* 16 (4).
qt1pv443h2.pdf (escholarship.org)

Polansky, L., Newman, K. B., Nobriga, M. L., & Mitchell, L. (2018). Spatiotemporal models of an estuarine fish species to identify patterns and factors impacting their distribution and abundance. Estuaries and Coasts, 41(2), 572-581.
Spatiotemporal Models of an Estuarine Fish Species to Identify Patterns and Factors Impacting Their Distribution and Abundance (researchgate.net)

Polansky L, Newman KB, Mitchell L. (2021) Improving inference for nonlinear state-space models of animal population dynamics given biased sequential life stage data. Biometrics. 77(1):352-61.
Improving inference for nonlinear state-space models of animal population dynamics given biased sequential life stage data - Polansky - 2021 - Biometrics - Wiley Online Library

Simonis, J.L., Merz, J.E. (2019). Prey availability, environmental constraints, and aggregation dictate population distribution of an imperiled fish. *Ecosphere* 10(3), e02634.
Prey availability, environmental constraints, and aggregation dictate population distribution of an imperiled fish - Simonis - 2019 - Ecosphere - Wiley Online Library