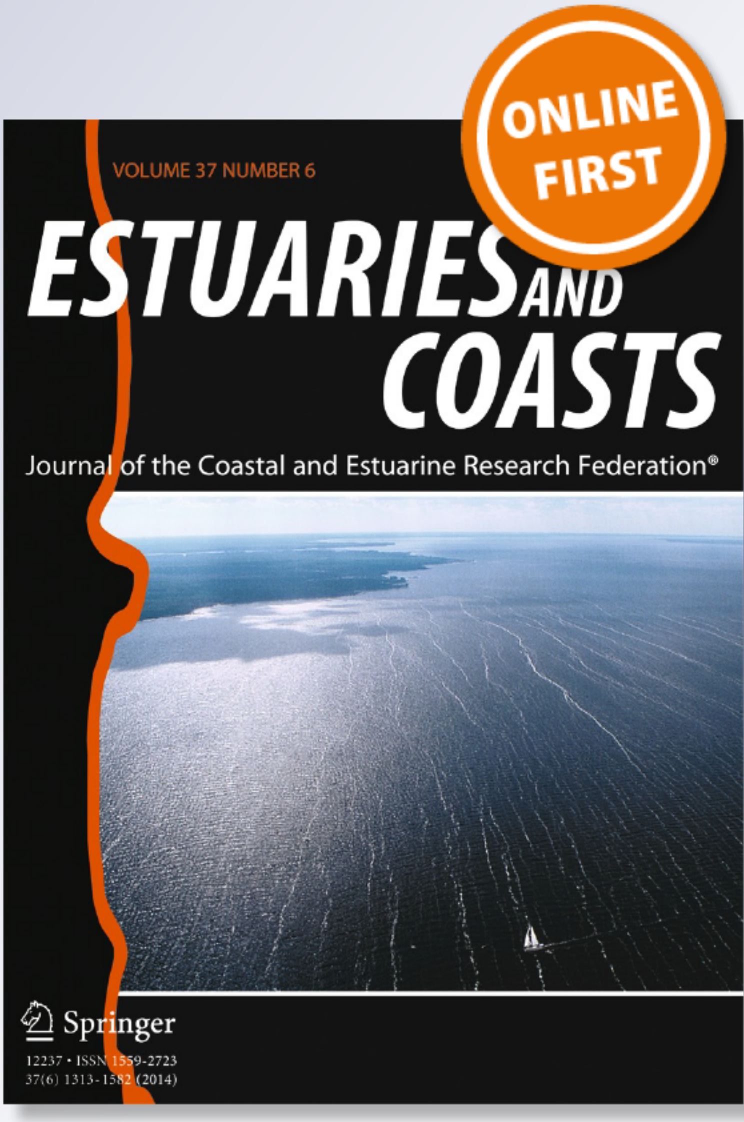
Comments on Feyrer et al. "Modeling the Effects of Future Outflow on the Abiotic Habitat of an Imperiled Estuarine Fish"

B. F. J. Manly, D. Fullerton, Albert Noble Hendrix & K. P. Burnham

Estuaries and Coasts
Journal of the Coastal and Estuarine Research Federation

ISSN 1559 2723

Estuaries and Coasts
DOI 10.1007/s12237 014 9905 3



Estuaries and Coasts
DOI 10.1007/s12237 014 9905 3

TECHNICAL COMMUNICATION

Comments on Feyrer et al. "Modeling the Effects of Future Outflow on the Abiotic Habitat of an Imperiled Estuarine Fish"

B. F. J. Manly · D. Fullerton · Albert Noble Hendrix · K. P. Burnham

Received: 8 May 2013 / Revised: 5 September 2014 / Accepted: 20 October 2014

Abstract Feyrer et al. (*Estuaries and Coasts* 34:120–128, 2011) constructed a habitat index for delta smelt (*Hypomesus transpacificus*) as a function of abiotic covariates (specific conductance, Secchi depth, and temperature) to evaluate how future hydrologic conditions in the San Francisco Estuary might affect the habitat of delta smelt. In this article, we identify three methodological issues that pertain to the results of Feyrer et al.: (1) the use of an independent abundance estimate, (2) the detection of spatial bias in the Feyrer et al. habitat index, and (3) the procedure used to link the habitat index to estuarine outflow. Like Feyrer et al. (*Estuaries and Coasts* 34:120–128, 2011), we fit general additive models (GAM) to presence of delta smelt data; however, our models included a region factor. We found that the amount of variability in the presence of delta smelt explained by the conductivity and Secchi terms was reduced relative to Feyrer et al.; conductance dropped from 12.2 to 2.5 % and Secchi dropped from 8.2 to 2.1 %. Furthermore, we found that an annual habitat index based solely on estuarine flow had low predictive ability, but the two-stage process of GAM analysis and subsequent regression modeling on GAM analysis output may mask the detection of low predictive performance. We agree with Feyrer et al. that defining a habitat index for delta smelt is an important contribution to understanding the ecology of the species and to facilitating its recovery. Given our results, the delta smelt habitat index could be improved by including static regional effects, dynamic salinity and turbidity effects, and an independent abundance index.

Keywords Delta smelt · Generalized additive model · Habitat index · Abundance index · Turbidity · Flow

Understanding how habitat may affect the abundance and spatial distribution of a species is an important step in understanding the ecology of the species. Habitats are typically heterogeneous across the landscape and may be influenced by interacting abiotic and biotic conditions that vary spatially and temporally (Wiens 2000). When the species is at low abundances, the habitat may be protected to ensure that habitat impairment does not threaten the existence of the species (USFWS 1994) or jeopardize recovery efforts (USFWS 1996). As a result, developing an understanding of what constitutes habitat is an important step toward understanding population dynamics (Rose 2000).

In a recent article, Feyrer et al. (2011) conducted an analysis of variables in the Sacramento-San Joaquin Delta (delta) with the objective of constructing a habitat index for the threatened estuarine fish, delta smelt (*Hypomesus transpacificus*) (Bennett 2005). Feyrer et al. (2011) used a generalized additive model (GAM) to relate several physical variables (temperature, Secchi depth, and specific conductance) to the probability of occurrence of delta smelt in samples taken among stations, months, and years. They dropped temperature from their GAM model, and hereafter we refer to Secchi depth and specific conductance as Secchi and

Communicated by Wayne S. Gardner

B. F. J. Manly
Western EcoSystems Technology Inc., 200 South 2nd Street, Laramie, WY 82070, USA

D. Fullerton
Metropolitan Water District, 1121 L St # 900, Sacramento, CA 95814, USA

A. N. Hendrix (✉)
QEDA Consulting, LLC, 4007 Densmore Ave N, Seattle, WA 98103, USA
e mail: noblehendrix@gmail.com

K. P. Burnham
Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, CO 80523, USA

Published online: 07 January 2015

conductivity, respectively. An annual habitat index was constructed by averaging the modeled probabilities of occurrence across months and summing across stations (with stations weighted according to station-specific areas). The annual habitat index was then regressed against September to December average *X*2 (distance in km upstream from the Golden Gate Bridge to where mean bottom salinity is 2‰) using locally weighted regression scatterplot smoothing (LOESS) regression. Finally, a series of climate change scenarios were used to predict *X*2, which were then used to predict a habitat index for each of the climate change scenarios.

The results of the Feyrer et al. (2011) analysis and its preceding work (e.g., Feyrer et al. 2007) have been cited as the underlying conceptual model for factors affecting delta smelt in both scientific and regulatory applications. For example, "fall *X*2 affects surface area available for fish through salinity distribution" and "turbidity favors all fish at various life-history stages by offering increased protection from predators" as described by MacNally et al. (2010) in their conceptual model for evaluating the factors responsible for pelagic species declines in the San Francisco Estuary. In addition, the United States Fish and Wildlife Service issued a delta smelt biological opinion in 2008 (USFWS 2008) regarding operation of the Central Valley Project (CVP) and State Water Project (SWP) that export water out of the Sacramento—San Joaquin watershed. One of the reasonable and prudent alternatives (RPA) specified the operation of the CVP-SWP to ensure that *X*2 in the fall be located a specific distance from the Golden Gate Bridge to maintain delta smelt habitat.

Given the importance of defining the habitat index for delta smelt, we investigated three components in the analysis by Feyrer et al. (2011). First, we wanted to incorporate an independent estimate of abundance into the GAM analysis. Levels of abundance can have important implications for defining "good" versus "poor" habitat (e.g., Royle and Dorazio 2008). Furthermore, abundance of delta smelt varies among years due to the annual life cycle (Bennett 2005). Feyrer et al. (2011) used an abundance index in their work, but it was constructed from the Fall Midwater Trawl (FMWT) catch data. These catch data were the same source as the presence/absence data used by Feyrer et al. (2011) and thus were not an independent data source. Inclusion of such a predictor variable, one that is derived from the same data source as the dependent variable, can lead to bias issues in regression modeling and potential overconfidence in the relationship between the two variables. Instead, we used the Summer Tow Net Survey (TNS) as the independent index of abundance in the GAM analysis to avoid the possibility of this occurring. Second, we wanted to investigate how well the Feyrer et al. (2011) model could reflect spatial patterns in the probability of delta smelt presence (i.e., capture). Spatial regions were not incorporated into the Feyrer et al. (2011) model; thus, we were interested in knowing whether the dynamic factors of conductivity and Secchi were capable of capturing regional patterns in the probability of delta smelt presence. Third, we were interested in exploring the predictive ability of equation (3) in Feyrer et al. (2011), which links the habitat index (a weighted sum of GAM predictions) to the variable *X*2, which is a proxy for estuarine outflow.

The purpose of these analyses is to use similar data as Feyrer et al. (2011) for fitting models, so that the fit of different models can be directly compared. For that purpose, we started with data provided by Feyrer on February 13, 2013, for the years 1967 to 2008. We then removed some observations with missing values for conductivity, Secchi depth, or temperature and were left with 13,660 observations of delta smelt presence and absence with the station numbers, latitude, longitude, conductivity, Secchi depth, and temperature. Feyrer et al. (2011) state that "there were nearly 14,000 individual samples with complete data for analysis," which would equate to using 100 stations in their GAM analysis. We then added the region (13 regions) for all of the stations (Fig. 1) and values of the Summer Tow Net abundance index (STN), which is based on data collected before the Fall Midwater Trawl sampling, for the years 1969 to 2008 as a measure of the abundance of delta smelt (Contreras et al. 2011). Values of STN are not available for 1967 and 1968, so we set the values for those two years at 0 and made an adjustment for those years to estimate the values that STN would have had, as explained below. Feyrer et al. (2011) subsequently use 73 of the stations for calculating their habitat index; however, the GAM analysis appears to be based on the 100 stations. We believe that we used the same data as Feyrer et al. (2011) (see comparison of variance explained below), but we cannot be sure that it was exactly the same.

The first models estimated the probability of delta smelt presence in the samples as a generalized additive model (GAM) with conductivity only (model 1), Secchi only (model 2), *X*2 only (model 3), and conductivity and Secchi effects (model 4). The percentage of the variation explained by the conductivity and Secchi model (model 4) is 26.1 % (Table 1). Because the models being fitted are for the probability of presence, a binomial link is used in the GAM formulation; thus, the percentage of variation is calculated using the regression and residual deviance (McCullagh and Nelder 1989). The value of 26.1 % is the same as the percentage quoted by Feyrer et al. (2011) for the same model, suggesting that the data sets were essentially the same for the purposes of estimation.

The Feyrer et al. (2011) analysis did not consider the possibility of regional differences in the probability of delta smelt presence, although delta smelt are almost never present in some areas so that regional effects clearly exist (Fig. 1). For this reason, we next estimated a model (model 5) allowing for 13 fixed regional effects that accounted for 21.7 % of the variation (Table 1). It was clear from the estimated parameters that there were considerable systematic differences in the

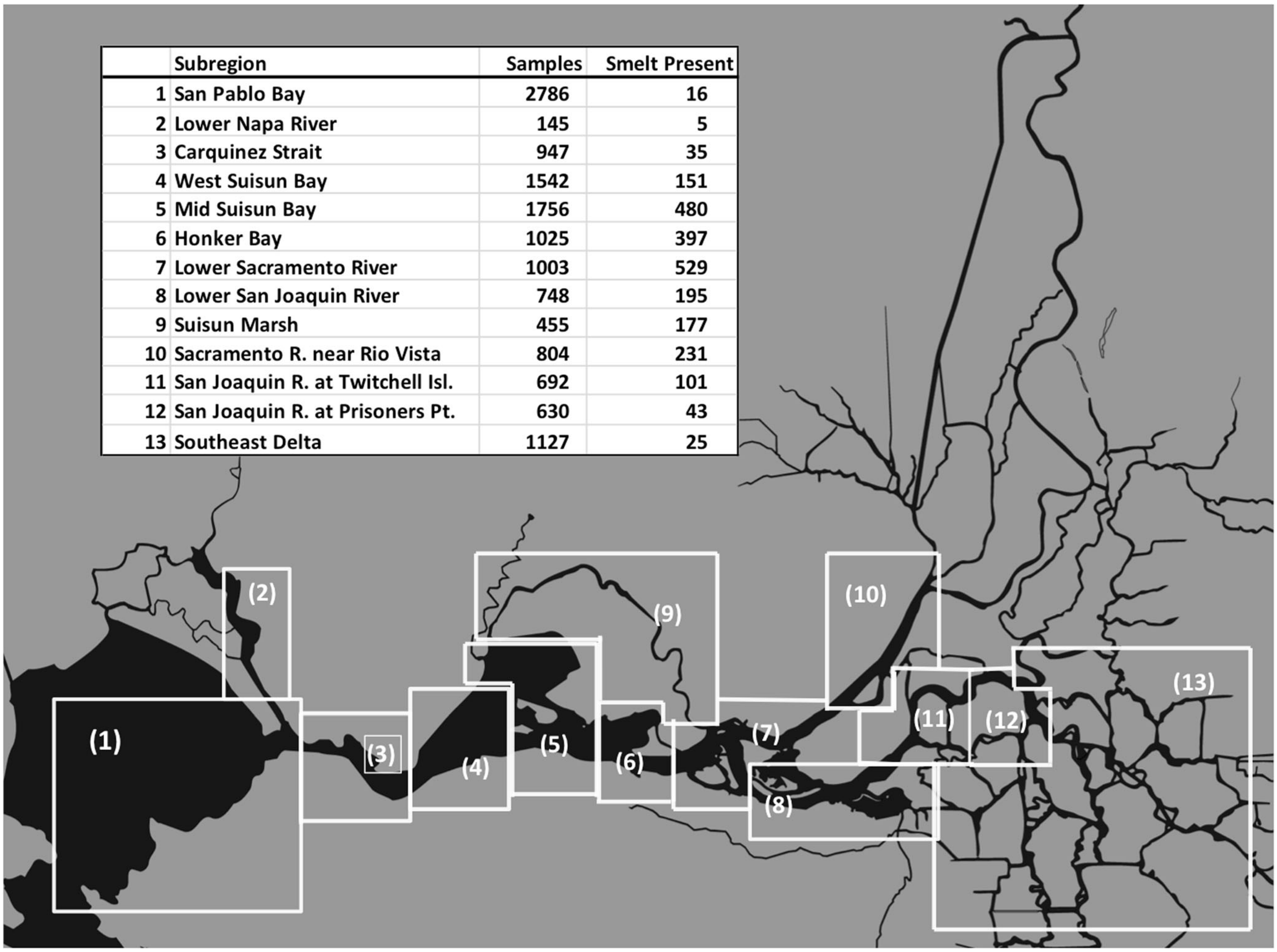


Fig. 1 Geographic subregions used to model delta smelt presence and the number of samples in which delta smelt were present between 1967 and 2008

probability of a presence in different parts of the Delta, with the estimated parameters varying from −5.153 for region 1 (San Pablo Bay) to +0.1098 for region 7 (lower Sacramento River).

Table 1 Results of model fitting for a series of models including con ductivity (Cond), Secchi depth (Secchi), abundance (STN), region of the delta smelt range (Region), and the distance (km) upstream from the Golden Gate Bridge to where mean bottom salinity is 2‰ (*X*2)

No.	Model	*df*	Variation explained (%)
1	GAM(Cond)	4	17.9
2	GAM(Secchi)	4	13.9
3	GAM(*X*2)	4	3.4
4	GAM(Cond) + GAM(Secchi)	8	26.1
5	Region	12	21.7
6	Region + GAM(Cond)	16	27.6
7	Region + GAM(Secchi)	16	28.0
8	Region + GAM(Secchi) + GAM(Cond)	20	30.8
9	Region + GAM(Secchi) + GAM(Cond) + GAM(STN)	26	33.4

The next question considered was whether the model including region effects can be improved by adding conductivity and Secchi effects. First, conductivity was added with a GAM model (model 6) followed by a GAM model in which Secchi was added (model 7), and finally a model in which both Secchi and conductivity were added (model 8). This last model accounts for 4.7 % more variation than the Feyrer et al. model with GAM effects for conductivity and Secchi depth (Table 1).

Apart from regional, Secchi, and conductivity effects, it is clear that the probability of delta smelt being captured in an area depends on the overall abundance of delta smelt. For that reason, the STN annual abundance index was added to the region model. This was done by including a GAM model for STN with STN set at 0 in 1967 and 1968, with an indicator variable Y67 added, which is one for samples in 1967 and zero for other years, and also an indicator variable Y68 added, where this is one for samples in 1968 and zero for other years. In this way, the GAM effects in 1967 and 1968 are adjusted based on the data in those years. We fitted a model that included the STN and dummy variables for Y67 and Y68, region, Secchi, and conductivity (model 9), which increased the overall variance explained to 33.4 % (Table 1). Model 9 also accounts much better than the Feyrer et al. (2011) model for the regional patterns in delta smelt presence (Fig. 2). The regional patterns are similar for model 8, which lacks the STN term, because the STN term does not affect the regional patterns. It is seen that the Feyrer et al. (2011) model tends to underestimate the proportion of samples with delta smelt present for stations from about longitude −122.00 to longitude −121.75 but overestimates the proportion for stations further east than about −121.60. This then shows up by the proportion of samples with presences being underestimated in the lower San Joaquin River, but overestimated in the San Joaquin River at Prisoners Point and the southeast Delta; however, the model with region effects (model 9) does not show these biases (Fig. 2).

Using these models, we calculated the amount of variability uniquely attributable to Secchi, uniquely attributable to conductivity, and shared between Secchi and conductivity as follows. In the Feyrer et al. (2011) model, the amount of variability attributable uniquely to conductivity is 12.2 % (calculated as 26.1–13.9 %), whereas the amount of variability attributable uniquely to Secchi is 8.2 % (26.1–17.9 %). The amount of variability that is shared by both Secchi and conductivity was calculated by summing the amount of variability explained in the Secchi only model (17.9 %) and the amount of variability explained in the conductivity only model (13.9 %). This sum equals 31.8 %, which is the amount of variability explained if they were independent (i.e., the shared variability equaled 0). Because they are not independent, the value of 31.8 % is greater than the observed value of 26.1 %; thus, there is 5.7 % of shared variability by Secchi and conductivity.

We decomposed the variance under model 8 in a similar fashion (Table 2). Under model 8, the amount of variability explained uniquely by Secchi is 2.1 %, which is similar to the amount explained uniquely by conductivity of 2.5 %. There is also an approximately equal amount of variability that is shared by both Secchi and conductivity of 2.1 %. These values are lower than the values under the Feyrer et al. (2011) model, which indicates that the static region effect explains much of the variability previously attributed to the dynamic variables of Secchi and conductivity. For example, the region component explained 4.7 % of the variability alone, whereas the shared Secchi and region component explained 5.0 %, and the shared conductivity and region component explained 9.4 % of the total variability (Table 2). This result implies a consistent habitat template, on which annual variability in Secchi and conductivity trigger lesser levels of variability in the probability of delta smelt presence.

Feyrer et al. (2011) used their equation (3) ($H_y = f_2(X2) + \varepsilon_y$, where H_y is the habitat index each year constructed as a weighted average of predicted probabilities of delta smelt presence) to relate $X2$ to the habitat index (see Figure 2B in Feyrer et al. (2011)). We found two issues with this approach that may limit its utility in predicting future delta smelt habitat. First, $X2$ is a good proxy for the salinity (conductivity) field in the estuary as a whole and is largely determined by outflow from the estuary. Equation (3) is the link between the modeled habitat index and possible future hydrological scenarios and thus is the foundation for the conclusions about habitat under possible future hydrology patterns. However, the habitat index in Feyrer et al. (2011), as well as the habitat index that we

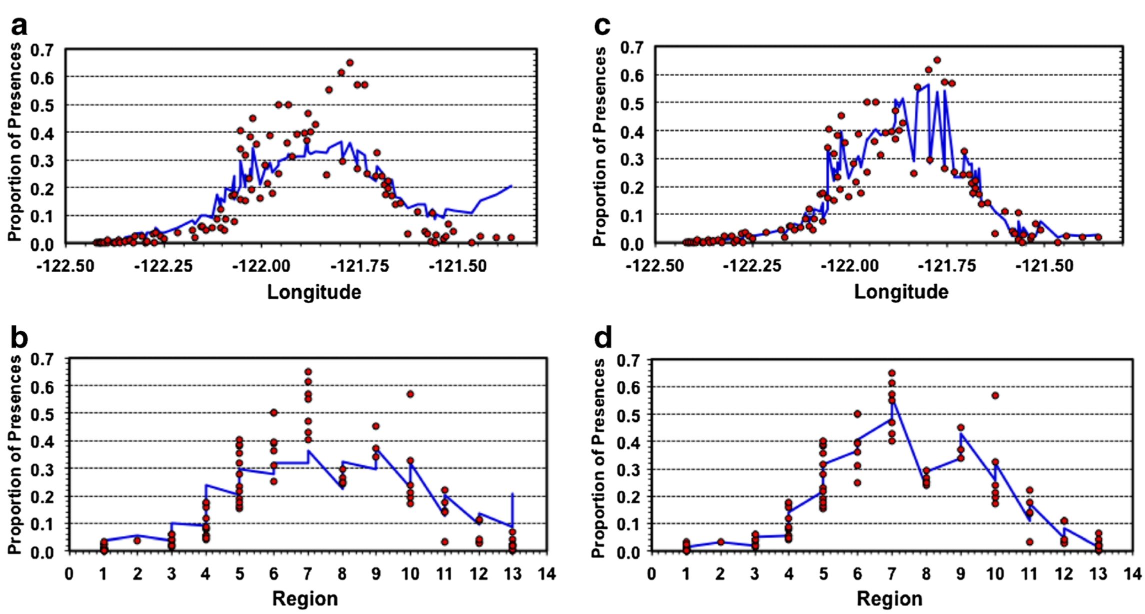


Fig. 2 Plots of the observed proportions of samples with delta smelt present (*points*) for samples taken from 1967 to 2008 at 100 Fall Midwater Trawl stations with the predicted mean proportions from the Feyrer et al. model (model 4) by station versus longitude (**a**), with the predicted mean proportions from the Feyrer et al. model (model 4) by station versus region (**b**), with the predicted mean proportions from model 9 by station versus longitude (**c**), and with the predicted mean proportions from model 9 by station versus region (**d**)

Table 2 Variance explained in probability of presence by conductivity (Cond), region (Region), and Secchi under model 4 and model 8. Model 4 is equivalent to the model used in Feyrer et al. (2011)

Factors	Model 4 (%)	Model 8 (%)
Region only	NA	4.7
Secchi only	8.2	3.2
Cond only	12.2	2.8
Shared Region and Secchi	NA	5.0
Shared Region and Cond	NA	9.4
Shared Cond and Secchi	5.7	3.1
Shared Region and Secchi and Cond	NA	2.6
Total	26.1	30.8

created using region effects, show that Secchi depth is roughly as important as conductivity in defining changes in the annual habitat index. Moreover, there appear to be interactions between the two variables. Thus, any effort to estimate the habitat impact from future changes in flow and $X2$ patterns should also include estimated future Secchi depth patterns as well or should at least include various possible future Secchi depths for sensitivity.

Second, the Feyrer et al. equation (3) implements a regression of a variable on the results of a GAM model. This is the second step in a two-step process, where the first step is to fit explanatory variables to a dependent variable of interest, and the second step is to use the predicted fits from step 1 to create a new dependent variable. This new dependent variable is then regressed against an independent variable. Yet, this two-step process shifts the emphasis away from the initial dependent variable, which is presumably the variable of interest (e.g., presence of delta smelt). Further, it can lead to overconfidence in the predictive ability of the second regression model on the dependent variable of interest. For example, the dependent variable of interest here is the probability of delta smelt presence. The ability to explain the probability of delta smelt presence as a function of $X2$ can be assessed by fitting a GAM($X2$). This step was completed as model 3 (Table 1), and the amount of variability explained was 3.4 %. Although $X2$ was regressed against the H_y, which is a weighted average of model output that was responsible for explaining 26.1 % of the delta smelt presence data, any function of $X2$ cannot explain more than 3.4 % of the variability. This result occurs by definition of the GAM model because any better fit would have been obtained through the GAM fitting process.

This two-step modeling approach may have arisen from the need to obtain coefficient values relating the habitat index to outputs of climate change models. There is a downside to the use of GAMs for fitting non-linear relationships among variables of interest: while they do not require the user to specify the mathematical form of the non-linear model, they also do not provide the mathematical form of the model nor the coefficients for prediction under new levels of the predictor variables (Zuur et al. 2009). In our own model fitting, the shape of the relationships between Secchi, conductivity, abundance, and the probability of delta smelt presence could likely be captured with polynomial models (also see Figure 1 of Feyrer et al. 2011), which would provide the ability to extract the coefficients directly. In most cases, the relationship between the covariate and the dependent variable is quadratic, which requires squaring the covariate and including it as an additional factor in the regression to reflect this relationship. If additional flexibility is needed, a cubed covariate can also be added as a factor in a similar manner. Using polynomial models in future efforts would allow the modeling to occur in a single step and likely avoid the problems identified here with the two-step process of fitting a GAM and subsequently regressing another covariate on the GAM output.

In summary, we evaluated regional patterns in the presence of delta smelt and found that they were not captured in the model presented in Feyrer et al. (2011). We constructed several additional models incorporating a regional factor as well as an independent abundance estimate to address these shortcomings. The inclusion of the independent estimate of abundance was moderate (unique variability explained was 2.6 %, 33.5 % variation explained under model 9, 30.8 % explained under model 8 in Table 1) which was similar to the amount of unique variation attributable to Secchi and conductivity (Table 2). Moreover, we found that the regional effects were substantial, particularly for understanding the dynamic effects of conductivity and Secchi compared with the static effect of region. The effects of conductivity and Secchi are then mostly allowed for by the permanent region effects that cannot be changed, and the only thing that can possibly be managed (e.g., by $X2$) are the short-term local effects of conductivity. These regional effects may reflect geographic variation in biotic variables such as food availability (Bennett 2005; Miller et al. 2012) or abiotic features such as bathymetry and substrate, and the development of habitat indices will need to take these regional differences into account. Furthermore, the relationship of habitat variables to presence may also vary by region in the delta (Merz et al. 2011; Nobriga et al. 2008) suggesting that interaction effects of Secchi and conductivity by region may provide a richer suite of models for defining delta smelt habitat. Finally, the amount of variability explained by $X2$ directly was relatively low (Table 2), yet this result does not necessarily discredit $X2$ as a general habitat indicator. Because $X2$ is a synoptic variable that reflects general outflow, it may be acting via indirect pathways to affect habitat. In light of our results on the role of static regional effects, the evaluation of direct and indirect pathways by which physical and biological processes can be connected at multiple levels (e.g., Wells et al. 2008) could provide an important area of future research.

In general, we feel that continued model development is an important area of research and advocate a larger suite of hypothesized covariates and their interactions to define the delta smelt habitat index. Another area of modeling development includes the role of covariates on the detection of animals given they are present (Royle and Dorazio 2008). Whether in Feyrer et al. (2011) or in our own reanalysis, the development of a habitat index implicitly assumes that differences in catch probabilities represent differences in actual presence and absence. An alternative hypothesis is that the sampled presence and absence data reflect differences in catchability in addition to true presence (Royle and Dorazio 2008). These approaches provide the means to develop multiple competing models that can be compared and evaluated (e.g., Burnham and Anderson 2002) on their ability to accurately predict the probability of delta smelt presence.

References

Bennett, W.A. 2005. Critical assessment of the delta smelt population in the San Francisco Estuary, California. *San Francisco Estuary and Watershed Science* 3: 1 71. Available at http://repositories.cdlib.org/jmie/sfews/vol3/iss2/art1.

Burnham, K.P., and D.R. Anderson. 2002. *Model selection and multi model inference: a practical information theoretic approach*. New York: Springer Verlag.

Contreras, D., K. Hieb, R. Baxter, S. Slater, and J. Afentoulis. 2011. 2010 Status and trends report for pelagic fishes of the upper San Francisco Estuary, IEP Newsletter. The California Department of Fish and Game. Available at: https://nrm.dfg.ca.gov/FileHandler.ashx?DocumentId=44331.

Feyrer, F., K. Newman, M. Nobriga, and T. Sommer. 2011. Modeling the effects of future outflow on the abiotic habitat of an imperiled estuarine fish. *Estuaries and Coasts* 34: 120 128. doi:10.1007/s12237 010 9343 9.

Feyrer, F., M. L. Nobriga, and T. Sommer. 2007. Multidecadal trends for three declining fish species: habitat patterns and mechanisms in the San Francisco Estuary, California, USA. *Canadian Journal of Fisheries and Aquatic Sciences* 64: 723 734.

MacNally, R., J.R. Thomson, W.J. Kimmerer, F. Feyrer, K.B. Newman, A. Sih, W.A. Bennett, L. Brown, E. Fleishman, S.D. Culberson, and G. Castillo. 2010. An analysis of pelagic species decline in the upper San Francisco Estuary using multivariate autoregressive modeling (MAR). *Ecological Applications* 20: 167 180.

McCullagh, P., and J. Nelder. 1989. *Generalized linear models*. London: Chapman & Hall.

Merz, J., S. Hamilton, P.S. Bergman, and B. Cavallo. 2011. Spatial perspective for delta smelt: a summary of contemporary survey data. *California Fish and Game* 97: 164 189.

Miller, B.J., B.F. Manly, D.D. Murphy, D. Fullerton, and R.R. Ramey. 2012. An investigation of factors affecting the decline of delta smelt (*Hypomesus transpacificus*) in the Sacramento San Joaquin estuary. *Reviews in Fisheries Science* 20: 1 19.

Nobriga, M. L , T. R. Sommer; F. Feyrer, and K. Fleming. 2008. Long term trends in summertime habitat suitability for delta smelt (*Hypomesus transpacificus*). *San Francisco Estuary and Watershed Science* 6(Issue 1) (February), Article 1. Available at http://repositories.cdlib.org/jmie/sfews/vol6/iss1/art1.

Rose, K.A. 2000. Why are quantitative relationships between environ mental quality and fish populations so elusive? *Ecological Applications* 10: 367 385.

Royle, A., and D.M. Dorazio. 2008. *Hierarchical modeling and inference in ecology*. Burlington: Academic. 444 pp.

U.S. Fish and Wildlife Service (USFWS). 1994. Endangered and threatened wildlife and plants; critical habitat determination for the delta smelt. Federal Register 50 CFR Part 17:65256 65277.

U.S. Fish and Wildlife Service (USFWS). 1996. *Sacramento San Joaquin Delta native fishes recovery plan*. Portland: U.S. Fish and Wildlife Service.

U.S. Fish and Wildlife Service (USFWS). 2008. Formal endangered species act consultation on the proposed coordinated operations of the Central Valley Project (CVP) and the State Water Project (SWP) Sacramento, California. 396 pp.

Wells, B.K., J.C. Field, J.A. Thayer, C.B. Grimes, S.J. Bograd, W.J. Sydeman, F.B. Schwing, and R. Hewitt. 2008. Untangling the relationships among climate, prey and top predators in an ocean ecosystem. *Marine Ecology Progress Series* 364: 15 29.

Wiens, J.A. 2000. Ecological heterogeneity: an ontogeny of concepts and approaches. In *The ecological consequences of environmental heterogeneity*, ed. M.J. Hutchings, E.A. John, and A.J.A. Stewart. Oxford: Blackwell.

Zuur, A., E. N. Ieno, N. Walker, A. A Saveliev, and G.M. Smith. 2009. Mixed effects models and extensions in ecology with R. Springer.