Appendix 2. A general linear model relating an index of proportional entrainment loss to turbidity and Old and Middle River flow

William Smith, 10 April 2019

1 Introduction

This note documents an effort to model an index of proportional entrainment as a function of existing water operations management quantities for adult delta smelt. An *index* of proportional entrainment is used for two reasons. Primarily, there is a mismatch between the timing of abundance estimates and entrainment estimates. Ideally, seasonal entrainment would be expressed as a fraction of abundance at the beginning of the entrainment season, on December 1; however, reliable abundance estimates were not available until January–February. Abundance estimates for November were available but were not considered as accurate as those measured in January–February. Additionally, the simplified model expressed here does not account for competing natural mortality that occurs simultaneously with entrainment mortality.

2 Methods

The objective of the analysis was to develop a regression model of an index of proportional entrainment loss (PEL) for years *y* 1994–2016 (cohorts 1993–2015)

(1) $$PEL_y = \frac{entrainment_y^{estimate}}{abundance_y^{estimate}}.$$

Total December through March entrainment of adult delta smelt $entrainment_y^{estimate}$ were separately estimated using a hierarchical model accounting for survival between entrainment and sampling and the sampling efficiency of fish facilities (Smith *in review*). Adult abundances $abundance_y^{estimate}$ were separately estimated from January–February Spring Midwater Trawl samples for cohorts 1993–2000 (corresponding to years 1994–2001) and from Spring Kodiak Trawl samples for cohorts 2001–2015 using methods described by Polansky et al. 2019. Abundance estimates were design-based stratified mean catch densities, expanded by strata water volumes and accounting for gear contact selectivity at length. Abundance estimates developed from Spring Midwater Trawl samples (1993–2001) were expanded by dividing by Life Cycle Model-estimated bias factor 0.3.

Covariates tested were December–February mean OMR and mean Secchi disk depth (measured throughout the Delta during fish surveys). Three water operations management periods were used to categorize cohort years (calendar year-1) into three management regimes pre-CalFed, CalFed, and BiOp years, corresponding to 1993–1998, 1999–2006, and 2007–2015 and periods of unmanaged OMR flow, management to more negative OMR flow, and management to less negative OMR flow.

The analysis was performed using a weighted generalized linear model, beta regression (betareg in R), and a logistic link

(2) $$PEL_y = \frac{1}{1+e^{-(\beta_0+\beta_1 * OMR_y+\beta_2 * Secchi_y+\beta_3 * OMR_y * Secchi_y+\beta_4 * Regime_y+\varepsilon_y)}},$$

where ε were normally distributed errors with mean 0. For consistency with existing management parameters, only models of OMR, Secchi, and Regime were explored.

Two alternate methods were explored to calculate model weights. In both methods, model weights were set equal to the inverse of Monte Carlo simulated variance of PEL_y (Eq. 1). The first method simulated PEL_y by iteratively resampling random $entrainment'_y$ and $abundance'_y$ values from log-normal distributions with mean and associated errors set equal to the values estimated from Spring Midwater and Kodiak Trawl surveys

(3) $$entrainment'_y \sim \text{Lognormal}\left(\log\left(entrainment_y^{estimate}\right), \sqrt{\left(1+\frac{entrainment_y^{se}}{entrainment_y^{estimate}}\right)}\right)$$ and

(4) $$abundance'_y \sim \text{Lognormal}\left(\log\left(abundance_y^{estimate}\right), \sqrt{\left(1+\frac{abundance_y^{se}}{abundance_y^{estimate}}\right)}\right).$$

PEL'_y was then calculated by dividing $entrainment'_y$ by $abundance'_y$, and variances of calculated PEL'_y were used for model weights.

The second method used an analytical solution for expected *PEL* μ and standard deviation σ, based on the assumption that *PEL* were lognormally distributed, being the product of lognormally distributed values of entrainment and abundance. Rather than simulating *entrainment* and *abundance*, PEL'_y was simulated, then variances of simulated PEL'_y were used for model weights.

(5) $$\mu_y = \log\left(\frac{entrainment_y^{estimate}}{\sqrt{\left(1+\left(\frac{entrainment_y^{se}}{entrainment_y^{estimate}}\right)^2\right)}}\right) - \log\left(\frac{abundance_y^{estimate}}{\sqrt{\left(1+\left(\frac{abundance_y^{se}}{abundance_y^{estimate}}\right)^2\right)}}\right)$$

(6) $$\sigma_y = \log\left(1+\left(\frac{entrainment_y^{se}}{entrainment_y^{estimate}}\right)^2\right) - \log\left(1+\left(\frac{abundance_y^{se}}{abundance_y^{estimate}}\right)^2\right)$$

(7) $$PEL'_y \sim \text{Lognormal}\left(\mu_y, \sigma_y\right)$$

All combinations of models using Secchi, OMR, Regime, and a Secchi-OMR interaction were compared, including an intercept-only model, using AIC.

Models were fit using the betareg library in R. The betareg library does not include a prediction function, so a parametric boostrap was used to simulate variance in regression model parameters and the resulting variance in predicted PEL.

3 Results

Graphical representations of *PEL* versus OMR and Secchi indicate a negative relationship with each (Fig. 1). The three management regime periods were evident in both the time series of *PEL* and OMR, with moderate *PEL* and highly variable OMR during the (cohort year) 1993–1998 pre-CalFed period, higher *PEL* and more negative OMR during the 1999–2007 CalFed period, and lower, less variable *PEL* and less negative, less variable OMR during the 2008–2015 BiOp period. PEL estimates were always low when Secchi averaged more than about 55 cm. Secchi depths were generally higher than this during the BiOp period.

Model selection, fit, and diagnostics

The two different methodologies for calculating model weights resulted in higher weights using Method 2 (simulating PEL'_y), but the relative weights among the different years was consistent between Method 1 (simulating $entrainment'_y$ and $abundance'_y$) and Method 2. The effect of greater model weights was to reduce the standard errors of regression coefficients, but no other change was noted. Model selection and estimated effect sizes of the best model were the same using either weighting method. Acknowledging uncertainty in estimates of entrainment and abundance, the lower model weights from Method 1 were used for all inference.

AIC model selection indicated the best model was the full model including Secchi, OMR, management regime, and secchi-OMR interaction effects (Table 1). Although the full model was selected, the regression coefficient associated with the BiOp period was not significant (P-value = 0.994). Mean PEL was significantly greater during the CalFed regime after accounting for OMR and Secchi depth, but mean PEL was not different between pre-CalFed and BiOp regimes. Overfitting was a concern, because 23 observations were used to estimate 6 regression parameters, but AIC is considered robust to overfitting because it penalizes model complexity. The 2nd best model identified by AIC had no management regime effect and could be considered an alternative model to address potential overfitting.

Residuals of the full model indicated errors were normally distributed, with no concerning patterns (Fig. 2), but standardized residuals were larger than expected and regression parameter standard errors were smaller than expected. Both were the consequence of regression weights; model weights allowed the model to attribute error to *PEL* rather than variation in the effects of OMR and Secchi depth. Removing all weights resulted in identical model selection and similar parameter estimates, but smaller residuals and higher regression parameter standard errors. This demonstrated that the weighted regression technique allowed greater precision in model estimates and predictions. Only *PEL* for cohort year 1996 was associated with a large residual and high leverage, as indicated by Cook's distance.

Model application to predict index of proportional entrainment loss

A table of predicted indices of predicted proportional entrainment losses was developed under two turbidity conditions (muddy [secchi = mean-sd] and clear [mean+sd]), a range of OMR conditions, and the three management regimes (Fig. 3). A negative relationship between PEL and OMR was evident, even at high values of Secchi depth. At values of OMR < -5,000 ft^3, there was no significant difference in predicted PEL between muddy and clear conditions, but at OMR > -5,000 ft^3, model predictions of PEL were significantly lower in clear conditions than in muddy conditions. Three-month average (December–February) OMR was only more negative than -5,000 cfs in cohort year 2006 (Figure 1).

4 Literature Cited

Polansky, L., Mitchell, L., and Newman, K.B. 2019. Using multistage design-based methods to construct abundance indices and uncertainty measures for delta smelt. Transactions of the American Fisheries Society. *In press*.

Smith, W.E. *In review*. Integration of transport, survival, and sampling efficiency in a model of South Delta entrainment.

Table 1. Beta regression model results. Covariates were standardized, so parameter estimates may be interpertted as effect sizes. The best model is indicated by lowest AIC (ΔAIC = 0).

	AIC	ΔAIC	% null deviance explained	intercept	Secchi	OMR	CalFed Regime	BiOp Regime	Secchi* OMR
intercept	-119.7	37.1	0.00	-2.39 (<0.001)	--	--	--	--	--
Secchi	-129.1	27.7	0.09	-2.5 (<0.001)	-0.55 (0.001)	--	--	--	--
OMR	-121.9	34.9	0.04	-2.65 (<0.001)	--	-0.88 (<0.001)	--	--	--
Regime	-131.9	24.9	0.13	-2.65 (<0.001)	--	--	0.88 (0.015)	-0.68 (0.09)	--
Secchi+OMR	-139.1	17.7	0.19	-2.89 (<0.001)	-0.68 (<0.001)	-1.14 (<0.001)	--	--	--
Secchi*OMR	-153.5	3.3	0.31	-3.81 (<0.001)	-1.3 (<0.001)	-3.05 (<0.001)	--	--	-1.62 (<0.001)
Secchi+Regime	-142.2	14.6	0.23	-3.15 (<0.001)	-0.6 (0.001)	--	1.26 (<0.001)	-0.11 (0.81)	--
OMR+Regime	-131	25.8	0.14	-2.66 (<0.001)	--	-0.41 (0.076)	0.67 (0.113)	-0.75 (0.066)	--
Secchi+OMR+ Regime	-144.6	12.2	0.26	-3.13 (<0.001)	-0.63 (<0.001)	-0.68 (0.002)	0.83 (0.025)	-0.27 (0.521)	--
Secchi*OMR+ Regime	-156.8	0	0.37	-4.02 (<0.001)	-1.22 (<0.001)	-2.46 (<0.001)	0.72 (0.027)	0.003 (0.994)	-1.34 (<0.001)

Figure 1. Time series of the index of proportional entrainment loss (PEL), mean December–February Old and Middle River flow (OMR), and mean secchi depth in the Delta, and the relationships among the index of PEL, OMR, and secchi depth.

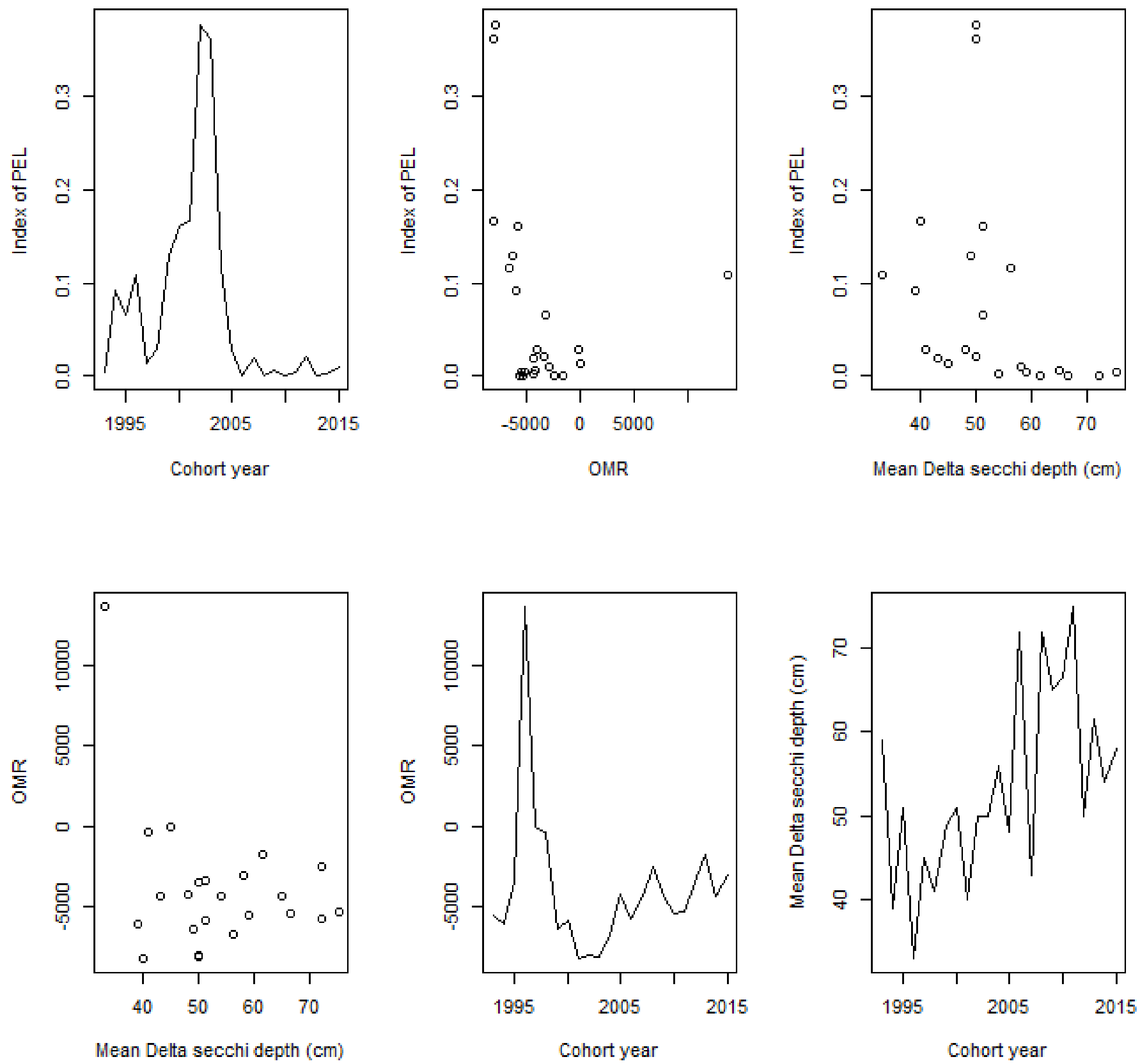

Figure 2. Residual plots for the model $\text{expit}(PEL_y) = \beta_0 + \beta_1 * Secchi_y + \beta_2 * OMR_y + \beta_3 * Secchi_y * OMR_y + \beta_4 * Regime_y$.

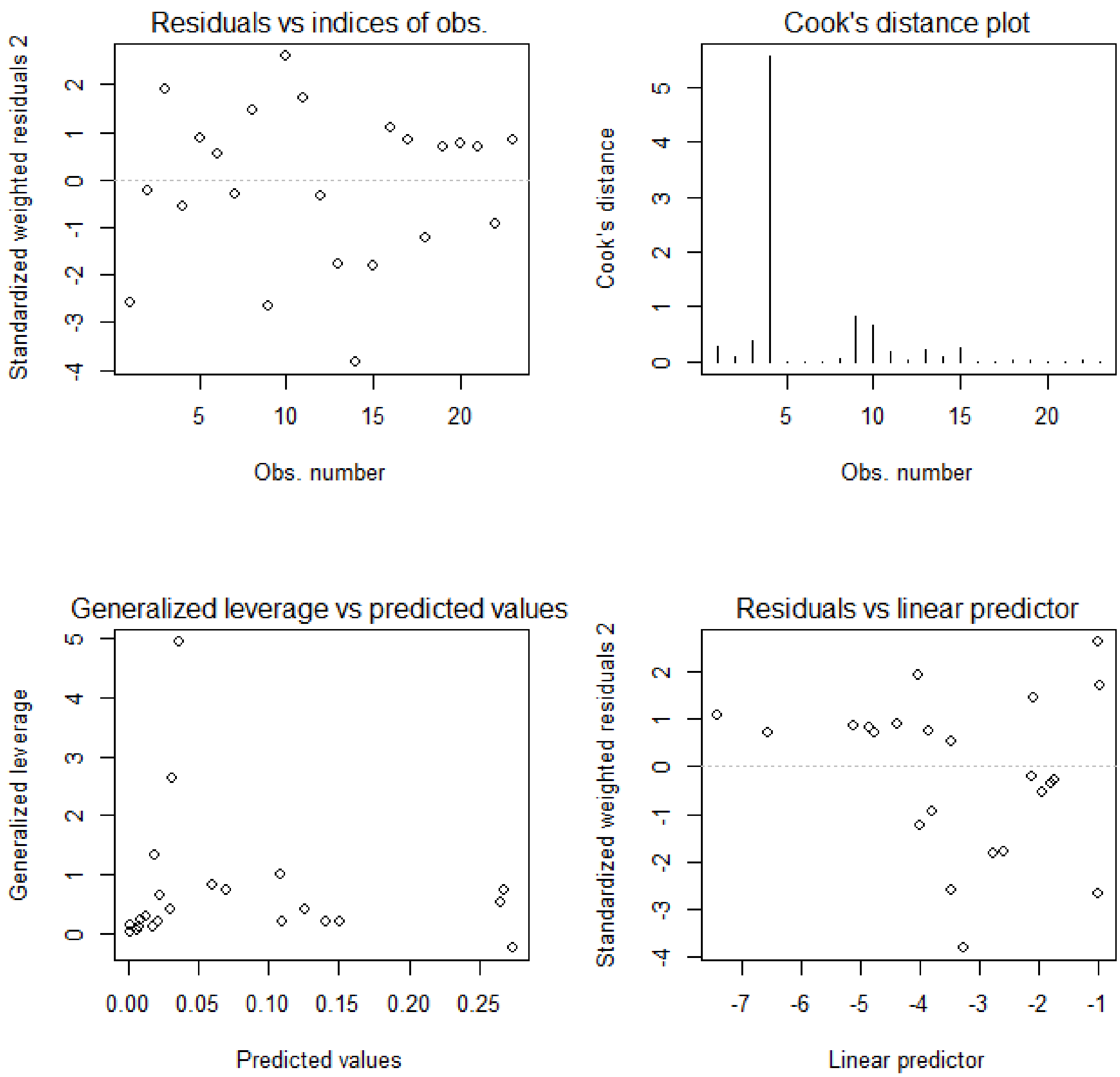

Figure 3. Model predictions from the best PEL model identified using AIC under muddy (secchi depth = 42 cm) and clear muddy (secchi depth = 66 cm) conditions. Black lines indicate mean predictions, and red lines indicate 95% prediction intervals.