Statistical modeling of delta smelt (*Hypomesus transpacificus*) survey data in the San Francisco-San Joaquin Delta, with reference to temporal and spatial autocorrelation.

Final report submitted to
AECOM

December, 2016

Robert J. Latour, Ph.D
Latour Environmental Consulting
Hayes, VA 23072

1. Background

This project was initiated under the auspices of the Collaborative Science and Adaptive Management Program (CSAMP), which is an overarching entity with the following four-tiered organizational structure:

1.) Policy Group consisting of agency directors and top-level executives from the entities that created CSAMP;
2.) CAMT made up of managers and staff scientists that serve at the direction of the Policy Group;
3.) Scoping Teams created on an as-needed basis to scope specific science studies; and
4.) Investigators contracted to conduct studies.

CSAMP was created following a decision by the United States District Court for the Eastern District of California in April, 2013. The decision, titled Memorandum Decision and Order Regarding Motion to Extend Remand Schedule" (Court Order), was issued in response to a motion filed to extend the court-ordered remand schedule for completing revisions to salmonid and delta smelt Biological Opinions (BiOps). The motion was filed by the U.S. Bureau of Reclamation, U.S. Fish and Wildlife Service, National Marine Fisheries Service, and the California Department of Water Resources, and the Court Order allowed additional time for those parties to develop a proposed robust science and adaptive management program, through collaboration among scientists, experts from the Public Water Agencies (PWAs), and the non-governmental organization (NGO) community with the intent to inform the management actions incorporated into the existing BiOps and consider alternative management actions. In 2015, the Ninth Circuit reversed the Court's decision with respect to the smelt and salmonid BiOps and issued a final judgment, thereby ending the Court Order. Following this action, all parties agreed to continue the CSAMP to promote the collaborative development of scientific information to inform sound decision-making in the future.

In July, 2013, the CAMT mutually agreed on the following mission statement:

> *The Collaborative Adaptive Management Team (CAMT) will work, with a sense of urgency, to develop a robust science and adaptive management program that will inform both the implementation of the current Biological Opinions, including interim operations; and the development of revised Biological Opinions.*

Over the past few years, CAMT has continued to focus on implementation of a work plan that included four areas of scientific investigation:

1.) Improved application of delta smelt survey data;
2.) Old and Middle River (OMR) flow management and entrainment of delta smelt;
3.) Fall outflow management of delta smelt; and
4.) South Delta salmonid survival.

This project falls under the first focal area and is designed to broadly address the following management question:

Are there biases in the existing delta smelt survey data, and if so, how should those survey data be utilized?

The term ***biases*** in the above management question can elicit a wide array of responses and interpretations. However, in the context of this project, that term is viewed in relation to presently unexplored structural aspects of extant survey data, and how they may affect inferences derived from those data about the delta smelt population. It should be noted that the fish surveys in the San Francisco-San Joaquin Delta are some of the longest running sampling programs in the United States and they undoubtedly contain a great deal of valuable information. Accordingly, the core scientific issue of this project pertains more to the notion of what else can be learned from delta smelt survey data (advancing the so-called scientific enterprise) rather than trying to demonstrate that data are somehow poor or lack utility.

Accordingly, the analytical strategy adopted with this project is to formally apply statistical models to extant delta smelt survey data, with particular attention directed at elucidating the effects of key environmental variables on survey catchability, thereby leading to estimates of delta smelt occurrence and relative abundance in relation to those environmental variables. However, the robustness of the analyses underpinning estimation and statistical inference depends on meeting the assumptions of analytical methods, so the characteristics of delta smelt survey data need to be fully explored and, where appropriate and possible, analytical procedures need to be modified to account for structural elements of the underlying data.

2. Introduction

Aquatic fish and invertebrate populations are routinely surveyed by scientists using a variety of sampling techniques such as trawls, dredges, gillnets, and traps (Kimura and Somerton 2006). With all of these sampling methods, the primary goal of a survey is to obtain representative data that allow estimation of key population quantities for use in management. When available, survey data are often considered vital components of analytical investigations into the biology, ecology, and population status of natural aquatic populations.

Fish survey data are often analyzed to estimate indices of relative abundance in relation to a variety of covariates. Indices are commonly expressed over temporal domains such as years or seasons for population trend information and use in stock assessment modeling (Hilborn and Walters 1992, Quinn and Deriso 1999). Relative abundance can be characterized in relation to environmental regimes such as temperature or salinity to understand aspects of habitat preferences (Buchheister et al. 2013). Spatially, indices displayed over longitude/latitude or broader areal definitions aid inferences about community structure and/or species distributions within an ecosystem (Malak et al 2014).

An underlying assumption associated with the interpretation of abundance indices is that they proportionally reflect changes in the true population abundance according to the following expression:

$$C = qEN \quad (1)$$

where E is survey effort, N is total population abundance, and q is the catchability coefficient defined as the fraction of the population captured with one unit of effort (Ricker 1975). Equation (1) can be re-arranged to yield the following key relationship:

$$\frac{C}{E} = CPUE = qN \quad (2)$$

which implies that catch-per-unit-effort (CPUE) is proportional to population abundance provided q remains constant. However, in practice, the catchability coefficient can vary depending on when, where, and how survey operations are conducted. In situations when q is not constant, interpretation of survey data becomes difficult because patterns of relative abundance are confounded by changes in both catchability and population size. As a result, survey data are often analyzed using statistical models to adjust for the effects of factors influencing catchability, which is a process often referred to as 'catch-effort standardization' (Maunder and Punt 2004).

Early attempts to standardize catch and effort data were applied to fishery-dependent, commercial landings information and focused on estimating 'fishing power' coefficients. Such coefficients are defined as fishing efficiencies of the various vessels within a fleet relative to that of a standard fishing vessel (Beverton and Holt 1957). Although numerous approaches for estimating efficiency coefficients have been developed, they cannot be easily generalized to account for multiple factors affecting catchability nor can they handle typical survey situations that rely on a single vessel. More recently, statistical models such as generalized linear (GLM, McCullagh and Nelder 1989) and additive (GAM, Hastie et al. 2001, Wood 2006) models have been used to standardize CPUE data, including those collected by surveys.

GLMs, GAMs, and their extensions have been applied extensively to fishery-independent and -dependent CPUE over the past three decades primarily to develop standardized time-series of yearly relative abundance for inputs to stock assessment models (Maunder et al. 2004). However, when sufficient data are available, these models can also yield estimated relationships of standardized relative abundance with other covariates of interest, particularly those of known or hypothesized ecological significance. Additionally, model-based treatment of survey CPUE data can provide insight into the effects of sampling procedures on catchability, which in turn, can be used to address a variety of survey design questions (Peel et al. 2013).

Fish survey data span several decades in many major freshwater, estuarine, and coastal ecosystems within the United States. Among estuaries, the San Francisco Bay is the largest on the west coast of the United States and systematic fish sampling was initiated in the 1960s. Freshwater is supplied to the bay primarily from the Sacramento and San Joaquin rivers, which converge to form a complex mosaic of tidal freshwater areas known collectively as the Sacramento-San Joaquin Delta (referred herein as the Delta). Over the past century, the Delta has experienced considerable anthropogenic changes in the form of lost wetlands (Atwater et al. 1979), sediment loading resulting from large-scale hydraulic mining (Schoellhamer 2011), introduction of invasive and nonindigenous species (Cohen and Carlton 1998), input of contaminants (Connor et al. 2007), and decreased

chlorophyll-*a* (Alpine and Cloern 1992) and zooplankton densities (Orsi and Mecum 1996). Analyses of fish survey data by the California Department of Fish and Wildlife (CDFW) have shown long-term declining patterns in abundance indices for several fish species, particularly since the early 2000s (Sommer et al. 2007). These patterns along with complementary analyses of other Delta fish and ecosystem attributes have collectively supported the conclusion that overall abundance of several pelagic fishes is currently quite low. However, specific to the available survey data underpinning fish population status determinations, routine analyses and estimation of relative abundance indices have not utilized methods that standardize data for potential changes in catchability resulting from variation in the timing and location of sampling activities.

Despite the wealth of available survey data for the Delta, few studies have analyzed 'raw' survey observations to make inferences about the biology and ecology of resident fish species. Feyrer et al. (2007) applied GAMs to quantify occurrences of three fish species in relation to water quality variables, and Feyrer et al. (2011) updated and extended that analysis specifically for delta smelt (*Hypomesus transpacificus*). These studies provide important insight into how environmental covariates affect survey capture probabilities of several Delta fish species. Latour (2016) applied zero-inflated GLMs to standardize CPUE data and assess patterns with respect to a suite of covariates across multiple temporal scales, and results revealed significant annual, seasonal, areal, and water clarity (Secchi) effects on CPUE for four fish species.

The aforementioned studies applied statistical models formulated to match the data structures being analyzed and overcome the often limiting assumptions of traditional regression analysis. In particular, the use of binomial GAMs to model capture probabilities alleviated the need for normally distributed response variables, variance homogeneity, and linearity among predictors. Application of zero-inflated GLMs also relaxed several classic regression assumptions (although linearity among predictors was required) and provided a synthetic treatment of excess zeros present in the survey data for many Delta fish species. However, the GAMs and GLMs used still required the assumption of independence among survey observations. Dependencies among observations can manifest in time (temporal or serial autocorrelation) or space (spatial autocorrelation) and failure to account for them in statistical modeling frameworks can lead to poor model fit, biased parameter estimates (effects of covariates influencing catchability) and unrealistic estimated standard errors of parameters used for statistical inference (Nishida and Chen 2004, Dormann et al. 2007).

Accordingly, the objectives of this study were three-fold: 1) build on analyses by Feyrer et al. (2007, 2011) and Latour (2016) by standardizing delta smelt CPUE data from two fish surveys by incorporating a broad suite of environmental and sampling covariates, 2) evaluate the presence of temporal and spatial autocorrelation among observations in each data set, and if present, modify modeling frameworks to reflect an appropriate autocorrelative structure, and 3) provide statistically derived relationships among delta smelt relative abundance and measured covariates that both advance the scientific understanding of how delta smelt are affected by environmental attributes and serve future statistical and mechanistic modeling studies.

3. Methods

3.1 Delta Smelt

The delta smelt is a member of the family Osmeridae and genus *Hypomesus*. This species is endemic to the Delta where it inhabits primarily low-salinity and freshwater habitats (Moyle et al. 1992, Bennett 2005). Delta smelt largely feed on various copepods and attain maximum sizes of approximately 60-70mm standard length (SL). Peak delta smelt spawning occurs in river channels of freshwater habitats during spring months and eggs and larvae are transported downstream into the saltwater-freshwater mixing zone. Delta smelt are semi-anadromous and semelparous with a maximum age of approximately one year (Bennett 2005). Relative to related smelt species, overall fecundity of delta smelt is low, which is an unusual life history characteristic for a primarily annual species, and one that leads to low productivity and high vulnerability to population depletion. Following apparent declines in abundance during the 1980s, the delta smelt was listed as threatened under both the U.S. Endangered Species Act (ESA) the California Endangered Species Act (CESA) in 1993, and listed as endangered under the CESA in 2010.

3.2 Data Sources

Datasets from two fishery-independent surveys were analyzed for the project. The first is the longstanding Fall Midwater Trawl (FMWT) survey, which the California Department of Fish and Wildlife (CDFW) has been conducting nearly continuously since 1967 (Stevens and Miller 1983; see http://www.dfg.ca.gov/delta/projects.asp?ProjectID=FMWT for additional details). The FMWT survey was initiated to measure the relative abundance of age-0 striped bass (*Morone saxatilis*), however, survey data have been used to infer patterns in relative abundance of a variety of species, including juvenile and adult delta smelt (Kimmerer 2002; Sommer et al. 2007). Monthly cruises are conducted from September to December and the number of tows each month has increased from approximately 75-80 during the early years of the program to > 100 in more recent years. The survey follows a stratified fixed station design such that sampling occurs at approximately the same location within predefined regional strata (17 areas excluding areas 2, 6, and 9 per the CDFW's protocol). Sampling intensity is related to water volume in each regional stratum such that samples are taken every 10,000 acre feet for areas 1-11 and every 20,000 acre feet for areas 12-17; Fig. 1). At each sampling location, a 12 minute oblique tow is made from near bottom to the surface using a 3.7 m X 3.7 m square midwater trawl with variable mesh in the body and a 1.3 cm stretch mesh cod end. Vessel speed over ground during tows can be variable since sampling procedures are designed to maintain a constant cable angle throughout the tow. Each catch is sorted, enumerated by species, and station-specific measurements of surface water temperature (°C), electrical conductivity (specific conductance, µS/cm), Secchi depth (m), overall depth (ft), tidal stage, tow direction with respect to tide, and water volume sampled (m^3, beginning in 1985) are recorded.

The second survey is the more recently initiated Spring Kodiak Trawl (SKT), which the CDFW has been conducting since 2002 specifically to measure the relative abundance of adult delta smelt during the spawning season (Bennett 2005; see https://www.wildlife.ca.gov/Conservation/Delta/Spring-Kodiak-Trawl for additional details). Monthly cruises are conducted from January to May during which 40 tows are made following a

fixed survey design. At each sampling location, a 10 minute surface tow is made using a 7.6 m X 1.8 m trawl net with variable mesh in the body and a 64 mm stretch mesh cod end. The net is towed by two boats to maintain spread and mouth opening. Each catch is sorted, enumerated by species, and station-specific measurements of surface water temperature (°C), electrical conductivity (specific conductance, μS/cm), Secchi depth (m), overall depth (ft), tidal stage, tow direction with respect to tide, and water volume sampled (m^3) are recorded.

3.3 Data Filtering

Prior to conducting formal analyses, exploration of the delta smelt data from each survey was conducted by generating basic summary statistics and plots of candidate response variables (catch and CPUE defined as catch-per-volume-sampled) over the observed domains of potential temporal, spatial, and environmental covariates. These initial explorations lead to the following decisions.

3.3.1 Fall Midwater Trawl

1.) Several of the covariates presently measured during towing operations, namely, hour of sampling (time-of-day), tow direction with respect to tide, depth, and tidal stage were not recorded during the early years of the survey. The first year in the time-series with measurements of all of these covariates was 1985, and because of hypotheses regarding their potential to affect survey catchability, data spanning the years 1985-2015 were analyzed. Water volume sampled by the trawl net was also not quantified until 1985, so the ability to adjust for variation among tow volumes further supported the chosen start year.

2.) Only data collected from surveys in September-December were analyzed to maintain consistency with index months defined by CDFW.

3.) Given that several areas contained very few nonzero catches, only data from areas 11-16 were analyzed (Fig. 1A, 2, Latour 2016).

4.) Tows with missing values of potential covariates were omitted, which corresponded to 5.1% of the tows remaining after applying the filters specified in points 2, 3 above.

5.) The covariate Tide was collapsed into three levels: slack, ebb, and flood, which amounted to eliminating the distinction between high slack and low slack tide. Relative to flood and ebb tidal conditions, the number of observations made during slack tide was quite low, so defining a single slack tide category aided sample size.

6.) Since the sampling design is based on fixed stations, there is no variation in the latitude/longitude coordinates of sampling locations across months and years. Therefore, to aid analyses aimed at addressing spatial autocorrelation, the spatial coordinates of sampling locations were randomly jittered slightly to make them unique. Operationally in the field, this small added spatial variation is potentially realistic since it can be difficult for the survey vessel to maintain the exact starting location of tows for a given station over months and years.

3.3.2 Spring Kodiak Trawl

1.) Observations associated with supplemental surveys were excluded from analysis because they were designed to intensively sample areas of highest delta smelt abundances (informed from

standard monthly surveys) for estimation of proportions of fish in pre-spawning, spawning, and spent reproductive stages.

2.) Tows with missing values of potential covariates were omitted, which corresponded to 1.3% of the tows remaining after applying the filter specified in point 1 above.
3.) Tow volume was calculated as the water filtered through the net (m^3) and was not corrected by the calibration factor for the flow meter (unknown to author). In cases where tow volume measurements were negative or zero (2.1% of the tows), the overall mean of the positive tow volumes was assigned.
4.) The covariate Tide was collapsed into three levels: slack, ebb, and flood – see point 5 from FMWT section.
5.) Spatial coordinates of sampling locations were randomly jittered slightly – see point 6 from FMWT section.

3.4 Statistical Modeling

3.4.1 Analytical Framework

Following data preparation, it was evident that each filtered survey data set contained high frequencies of zero observations. The proportion of positive observations by year, defined as capture of at least one delta smelt, ranged between 0.02 and 0.41 for the FMWT and 0.14 and 0.53 for the SKT (Figs. 3 A, B). Mean proportion of positive tows across years was 0.17 and 0.29 for the FMWT and SKT surveys, respectively. When applying statistical models that require specification of an underlying probability distribution for the response variable, zero observations can be challenging, particularly if they are common in data sets. Two-part zero-altered and mixture zero-inflated distributions (Zuur et al. 2009, 2012) were explored but ultimately not used because currently available software does not easily permit fitting models that can accommodate autocorrelated response data. Therefore, a 'Δ-distribution' was assumed (referred herein as delta-model), which is a two-part distribution that specifies (i) the probability of obtaining a non-zero observation (encounter or capture probability), and (ii) a conditional distribution for the positive observations (Aitchison 1955). In effect, the encounter probability and the positive observations are modeled separately. Delta-distributions have also been termed hurdle distributions (Cragg 1971) because a hurdle needs to be crossed to obtain a positive observation.

The general form of a delta-model model is:

$$\Pr(Y = y) = \begin{cases} p & y \neq 0 \\ p \cdot f(y) & \text{otherwise} \end{cases} \quad (3)$$

where p is represents the encounter probability and $f(y)$ is probability distribution of positive observations. The parameter p was modeled with the binomial distribution to accommodate the presence/absence of delta smelt in the FMWT and SKT surveys, and diagnostic plots of preliminary model fits suggested that a lognormal distribution was appropriate for $f(y)$. Therefore, the assumed distribution for both survey data sets was a delta-lognormal, which has been used extensively for the analysis of fisheries data (Pennington 1983, Lo et al, 1992, Stéfansson 1996, Dick 2004, Fletcher

2008). All statistical analyses were conducted using the software package R (R Core Development Team 2015).

3.4.2 Covariates

Since both surveys collect information on the same covariates, a single initial candidate set of covariates was defined for analysis of both survey data sets: Year (categorical), Month (categorical), Area (categorical, FMWT only), Tide (categorical), Hour of sampling (continuous), surface Temperature (continuous), surface Conductance (continuous), Secchi depth (continuous), TowDirection (continuous), and overall Depth (continuous). Scatter plots of covariate matrices (SPLOMs), estimated correlations among covariates, and variance inflation factors (VIF) associated with preliminary GLM fits suggested collinearity among Month/Temperature and Tide/TowDirection in both data sets, and Area/EC in the FMWT data. Centering (subtract mean) and standardizing (subtract mean and divide by standard deviation) the covariates did not reduce collinearity, so Month, Area, and TowDirection were eliminated in an effort to retain environmental covariates. Given the finalized set of seven predictor covariates, eight parameterizations were analyzed within the binomial and lognormal components of both data sets (Table 1). All models contained Year because of the need to estimate an annual time-series of relative abundance, Temp and EC as proxies to control for seasonal and spatial patterns, and Secchi because it has been shown to be an important covariate in previous analyses of delta smelt encounter probabilities (Feyrer et al. 2007, 2011) and relative abundance (Latour 2016). Various combinations of the remaining covariates were included to explore their importance. For binomial models, the natural logarithm of TowVolume was used as an offset variable, and CPUE was defined as catch-per-m^3 for the lognormal models. All continuous covariates were standardized prior to analysis to facilitate comparisons among estimated effects (Schielzeth 2010).

3.4.3 Baseline Models

As a first step in the analysis of each data set, binomial and lognormal GLMs and GAMs were fitted as baseline models to provide insight regarding model performance and coherence with required analytical assumptions. These models also importantly facilitate comparisons of parameter estimates and model prediction with those designed to accommodate temporal and spatial autocorrelation. Formally, a GLM specifies a linear relationship between the expected value of the response variable and a set of covariates:

$$g(E(y)) = X\beta \tag{4}$$

where y is the vector of observations of the response variable, X is the fixed effects model matrix containing observations of the covariates, β is the vector of coefficients, and g is the link function which is differentiable and monotonic. GAMs are extensions of GLMs that replace the linear predictor by an additive predictor:

$$g(E(y)) = X\beta + \sum_{j=1}^{p} s_j(x_j) \tag{5}$$

where s_j is a smoothing function such as a spline or a loess smoother that provides a partially non-parametric aspect to the model. For normally distributed data, the error covariance matrix of a GLM or GAM is given by $\varepsilon \sim N(0, \sigma^2 I)$, where I is the identity matrix. The R library 'mgcv' was accessed to fit GAMs.

3.4.4 Temporal Autocorrelation

Fish surveys that follow fixed station designs, as in the case of the FMWT and SKT surveys, can be generally categorized as longitudinal studies since the same locations (or approximate locations) are sampled repeatedly through time. Longitudinal data have several unique features, most notably that measurements taken at the same sampling locations over time may be related or correlated due to within-site variation. That is, some sites may be located in areas with favorable habitat conditions such that surveys catches are consistently high, while others may be in areas with unfavorable habitat conditions and survey catches are regularly low. If survey measurements through time at the fixed stations are appreciably correlated, then the amount of true information in a survey data set will be misrepresented and analyses of those data will be flawed.

In the context of statistical models such as GLMs and GAMs, the assumption of independence among observations can be translated to imply independence among residuals (defined as $\varepsilon_i, i = 1, \ldots, n$). Formally, this idea is expressed as:

$$\text{cor}(\varepsilon_s, \varepsilon_t) = \begin{cases} 1 & \text{if } s = t \\ 0 & \text{otherwise} \end{cases}$$

which indicates that the correlation among pairs of residuals is zero. However, when observations are not temporally independent, then the correlation among residuals at different time points is nonzero and can be modeled with a correlation function (denoted by h):

$$\text{cor}(\varepsilon_s, \varepsilon_t) = \begin{cases} 1 & \text{if } s = t \\ h(\varepsilon_s, \varepsilon_t, \rho) & \text{otherwise} \end{cases}$$

where h takes on values between -1 and 1 and ρ represents the correlation between residuals s and t that is estimated from the data. The function h can assume different structural forms depending on hypotheses about the nature of autocorrelative process.

3.4.4.1 Binomial Component

Two modeling approaches were applied to FMWT and SKT survey data to explore temporally autocorrelated binomial errors, namely generalized estimating equations (GEE) and generalized linear mixed models (GLMM).

1.) Generalized Estimating Equations – The GEE approach is commonly used to analyze data where responses are measured repeatedly on the same sampling units, often denoted as clusters (Liang and Zeger 1986). To formally describe GEE, first note that equation (4) can be expressed as: $E(y) =$

$\mu = g^{-1}(X\beta)$, and parameter estimates of a traditional GLM are obtained by minimizing the gradient of the log-likelihood function (score equation), which is given by:

$$D^T V^{-1}\left(y - g^{-1}(X\beta)\right) = 0 \tag{6}$$

where D^T is the transposed matrix of partial derivatives, $D = \partial(g^{-1}(X\beta))/\partial\beta$, and V is the variance-covariance matrix. However, as noted above, for GEEs the data are split up into m clusters such that the variance-covariance matrix has block diagonal form which leads to the following transformed score equation:

$$\sum_{j=1}^{m} D_j^T V_j^{-1}\left(y_j - g^{-1}\left(X_j\beta\right)\right) = 0 \tag{7}$$

where j indexes the clusters and the matrix V is augmented to take into account non-independent observations: $V = \phi A^{1/2} R A^{1/2}$, where A contains the GLM variances, ϕ is the dispersion parameter to adjust for overdispersion (set to 1 for binomial data), and R is an additional matrix that incorporates the correlation structure among residuals (Carl and Kuhn 2007, Dormann et al. 2007).

GEEs are sometimes referred to as population average (so called PA interpretation) or marginal models because they provide an average response for observations sharing the same covariates in relation to those covariates. That is, for each one-unit increase in a covariate across the population, GEEs indicate how much the average response would change. It has been noted that GEEs are best suited for parameterization rather than prediction (Augustin et al. 2005).

For the GEE analysis, the clusters were defined as the survey sampling locations (station) and a two-step procedure was adopted. First, model fits of the saturated parameterization (all seven covariates included) with different correlation structures for the residuals were compared. The specific correlation structures considered were:

Exchangeable: $\mathrm{cor}(\varepsilon_s, \varepsilon_t) = \begin{cases} 1 & \text{if } s = t \\ \rho & \text{otherwise} \end{cases}$

Auto-regressive first order (AR1): $\mathrm{cor}(\varepsilon_s, \varepsilon_t) = \begin{cases} 1 & \text{if } s = t \\ \rho^{|s-t|} & \text{otherwise} \end{cases}$

Independence $\mathrm{cor}(\varepsilon_s, \varepsilon_t) = \begin{cases} 1 & \text{if } s = t \\ 0 & \text{otherwise} \end{cases}$

Exchangeable assumes the correlation among residuals is the same regardless of the length of time between observations, AR1 assumes that the correlation among residuals decreases as a function time between observations, and independence assumes there is no correlation among residuals and is typically explored as a reference to determine if more complex structures are warranted (such a model is equivalent to GLM). Once the appropriate correlation structure was identified for each survey data set (note that GEE is robust to modest misspecification of the correlation structure given

large sample sizes), the second step involved fitting the eight aforementioned main effect covariate combinations. The R library 'geepack' was accessed from model fitting.

2.) Generalized Linear Mixed Models – An alternative approach that can be used for temporally correlated data is GLMMs (Bolker et al. 2009). These models are extensions of GLMs in which the linear predictor contains both fixed and random effects. Formally, GLMMs are defined as:

$$g(E(y)) = X\beta + Zb \tag{8}$$

where Z is the model matrix for the random effects, b is vector of random effects, other model terms are the same as defined above, and $\varepsilon \sim N(0, \Sigma)$, $b \sim N(0, \mathrm{D})$, with D being symmetric and positive-definite.

In general, mixed effects models assume characteristics of sampled subjects (defined as random effects) induce correlation among repeated measurements. However, it is important to recognize that mixed model regression parameters reflect the expected response of a subject in relation to the covariates associated with that subject (so called subject-specific (SS) interpretation). As a result, GLMMs can be referred to as conditional mean models, where conditioning is performed with respect to one or more random effects, as opposed to GEEs which are marginal mean models that provide population averages over subjects or dependencies in the data.

For dichotomous response variables, a latent formulation of equation (8) can be written in terms of a continuous response variable, y:

$$y = X\beta + Zb$$

where y is greater than zero for a 'success' response outcome and less than zero for 'failure' response outcome. For GLMMs with a single random effect, it can be shown that the covariance matrix of y is comprised of the variance of the residuals (σ_ε^2), which measures the within-subject variation, and the variance of the random effect (σ_b^2), which measures the between-subject heterogeneity not captured by the fixed effects. Specifically, the diagonal elements of the covariance matrix are given by $\sigma_b^2 + \sigma_\varepsilon^2$ (variance) and the off-diagonal elements are given by σ_b^2 (covariance), which illustrates the induced correlation structure resulting from inclusion of the random effect. For binomial regression using a logit link function, the within-subject variance σ_ε^2 equals $\pi^2/3$ (Hedeker and Gibbons 2006). It follows then that the correlation between two observations is: $\rho = \sigma_b^2/(\sigma_b^2 + \pi^2/3)$, where ρ describes the proportion of the total response variation resulting from between-subject variation and is a useful measure of the degree of within-subject dependence relative to the total variation. Low values of ρ are the result of high within-subject variability (dissimilar repeated observations) while high values of ρ are due to low within-subject variability (similar repeated observations). A secondary use of ρ is in the calculation of the design effect, which represents the extent to which the expected sampling error in a survey differs from the expected sampling error under simple random sampling. Formally, the design effect is: $d_{\text{eff}} = 1 + (n-1)\rho$, where n is the number of observations per subject (or mean number if sampling is unbalanced) and, in turn, aids calculation of effective sample size, defined as: $N_{\text{effective}} = Mn/d_{\text{eff}}$, where M is the

number of subjects (Zuur et al. 2009). Estimates of d_{eff} and $N_{\text{effective}}$ can be helpful when evaluating design characteristics of surveys.

For the GLMM analysis of presence-absence of delta smelt in both survey data sets, the eight covariate fixed effects parameterizations were augmented to allow a random intercept for each station. That is, station was defined as a random effect and the variation around the model intercept for each station was assumed to be normally distributed with a certain variance. If that variance is small, the differences among stations from the perspective of the model intercept are small, whereas a large variance would be indicative of greater station-specific variation. The library 'glmmADMB' was accessed for model fitting (note, functions to fit GLMMs in the R libraries 'lme4' and 'glmmML' were unsuccessful).

3.4.4.2 Lognormal Component

Three modeling approaches were applied to FMWT and SKT survey data to explore temporally autocorrelated nonzero CPUE observations, namely, GEE, GLMMs, and generalized additive mixed models (GAMMs).

1.) Generalized Estimating Equations – The GEE analysis for temporal autocorrelation of the nonzero CPUE data followed the above description for temporal GEEs fitted to the presence-absence data, with the only difference being specification of the normal rather than binomial distribution.

2.) Generalized Linear Mixed Models – Since the underlying CPUE data is normally distributed (in log space), the parameter σ_ε^2 is estimated from the data. Accordingly, the possibility exists for heterogeneous variance across the domain of observed CPUE values. Therefore, using the saturated model parameterization, a variety of heterogeneous variance structures were explored in addition to the standard model of constant variance. Heterogeneous variances were modeled assuming that the heterogeneity depended on the mean, which led to exploration of the power-of-mean: $\text{var}(\varepsilon_i) = \sigma^2(f(X,\beta))^{2\theta}$, exponential of the mean: $\text{var}(\varepsilon_i) = \sigma^2\exp\{2\theta f(X,\beta)\}$, and constant plus power of the mean: $\text{var}(\varepsilon_i) = \sigma^2[\theta_1 + \{f(X,\beta)\}^{2\theta_2}]$ forms. Following determination of the appropriate variance structure, the eight parameterizations were fitted. Maximum likelihood was used for parameter estimation and the R library 'nlme' was accessed for model fitting.

3.) Generalized Additive Mixed Models – The final approach for addressing temporal autocorrelation involved fitting additive mixed models (GAMM; Lin and Zhang 1999). GAMMs are extensions of GAMs and parallel GLMMs fitted to normally distributed data with a known identity link function. The additive predictor in equation (5) is augmented to include both fixed and random effects:

$$E(y) = X\beta + \sum_{j=1}^{p} s_j(x_j) + Zb \qquad (9)$$

with all terms as defined above, including the distributional assumptions about b and ε (residual variance). The GAMM analysis followed the same procedure outlined for the GLMMs where station was the random effect modeled as an intercept and the appropriate structure of σ_ε^2 was first explored using the saturated model parameterization. Following identification of the appropriate variance structure, the eight fixed effect covariate combinations were fitted. The R library 'mgcv' was accessed for model fitting.

3.4.5 Spatial Autocorrelation

Fish surveys are generally designed with explicit considerations of allocation of sampling locations across space. Spatial stratification variables are often defined to ensure sampling is conducted over the full domain of the sampling frame, which in turn, aids population level inferences based on analyses of survey data. However, depending on the habitat characteristics of the study ecosystem and the dispersal tendencies of study species, observations collected different sampling locations may be related (not independent) and thus spatially autocorrelated. Although the causes of spatially correlated data are numerous, three common sources have been identified (Dormann et al. 2007, Legendre and Legendre 2012): i) biological and ecological processes of species are distance-related, ii) model misspecification that creates spatial dependency in, for example, estimated residuals, and iii) failure to account for key, spatially structured covariates in statistical models (Besag 1974). Not accounting for spatial autocorrelation in data or emergent spatial dependencies associated with specific models structures may lead to biased parameter estimates and incorrect statistical inferences. The primary approach used to explore spatial autocorrelation in both the response data and subsequent model residuals from baseline models fitted to both survey datasets was graphical analysis using correlograms (Box et al. 1994).

3.4.5.1 Binomial Component

One modeling approach was applied to FMWT and SKT survey data to explore spatially autocorrelated binomial residuals, namely autocovarite models (AUTO). Spatial GEEs were also explored but ultimately not considered for the following reasons. Often, spatial GEE models are structured to include only one cluster with a fixed n x n spatial correlation matrix, where the elements of the matrix represent distance based correlations between sampling locations (Dormann et al. 2007, Carl and Kuhn 2007). For datasets with many observations, this approach leads to a large spatial correlation matrix and the iterative procedure inherent to fitting spatial GEEs becomes very computationally intensive. Given that the FMWT presence-absence dataset being analyzed spans 30 years, the dimensions of the resultant spatial correlation matrix were very large and efforts to fit spatial GEEs were computationally unsuccessful. Although the SKT presence-absence data spans a much shorter time frame, convergence issues and numerical instability was evident for many fitted spatial GEE models.

1.) Autocovariate Models – The binomial GLM for presence-absence data can be modified to include an explanatory variable that is a distance-weighted function of neighboring values of the response variable. This function captures how much the response variable at a particular location reflects the response at surrounding locations. Such models have been termed AUTO models, since

characteristics of the response data are used in the explanatory component of the model. AUTO models take the form:

$$g(E(y)) = X\beta + \rho A \tag{10}$$

where ρ is the coefficient of the autocovariate matrix A such that $A_i = \frac{\sum_{j \epsilon k_i} w_{ij} y_j}{\sum_{j \epsilon k_i} w_{ij}}$. For the i^{th} sampling location, A_i is a weighted average of the response variable at the j^{th} sampling location among i's set of k_i neighbors, and w_{ij} is the weight given to j's influence over sampling location i (Augustin et al. 1996, Gumpertz et al. 1997; Bardos et al. 2015).

In developing the autocovariate, a variety of neighborhood radiuses was explored, ranging from 1 to 10 km using the saturated fixed effects model. Also, binary weights were used which implied that sampling locations were coded as either present or absent from the neighborhood leading to a symmetric weights matrix (Bardos et al. 2015). The R library 'spdep' was accessed to estimate the autocovariate needed for model fitting.

3.4.5.2 Lognormal Component

Three modeling approaches were applied to FMWT and SKT survey data to explore spatially autocorrelated lognormal errors, namely spatial error models (SAR_{err}), generalized least squares (GLS), and generalized additive models with correlated errors (GAM_{err}; McMullan et al. 2007).

1.) Spatial Error Models – SAR_{err} models assume that the autoregressive process is found only in the error term, which is likely if spatial autocorrelation is not fully explained by the explanatory variables (induced spatial dependence), or if spatial autocorrelation is an inherent property of the response variable itself (inherent spatial autocorrelation). The SAR_{err} model is given by:

$$E(y) = X\beta + \lambda W u \tag{11}$$

where λ is the spatial autoregressive coefficient, W is the spatial weights matrix such that λW represents the spatial structure in the spatially dependent error term u. As in the AUTO analysis, a variety of neighborhood distances were explored for calculating elements of W (Kissling and Carl 2008) and binary weights were used (Bardos et al. 2015). The R library 'spdep' was accessed for model fitting.

2.) Generalized Least Squares – The estimation technique GLS is generalization of the commonly applied method of ordinary least squares (OLS) for regression analysis of normally distributed response data. A key characteristic of GLS is the specification of a general error covariance matrix such that $\varepsilon \sim N(0, \Sigma)$, where Σ is symmetric and positive-definite.

The strategy taken with the spatial GLS analysis followed the approach outlined with the GLMMs structured to account for temporal autocorrelation. In particular, several variance and correlation structures were first explored using the saturated model parameterization, including

parameterizations modeling variance and correlation separately and together. Given two sampling locations r distance apart, the specific correlation structures examined were:

$$\text{Exponential:} \quad \mathrm{C}(r) = e^{\frac{-r}{d}}$$

$$\text{Gaussian:} \quad \mathrm{C}(r) = e^{\left(\frac{-r}{d}\right)^2}$$

$$\text{Spherical:} \quad \mathrm{C}(r) = 1 - 1.5\left(\frac{r}{d}\right) + 0.5\left(\frac{r}{d}\right)^3$$

where d is a parameter estimated from the data (Cressie 1993). Once optimal variance and correlation structure(s) were identified, the eight model parameterizations were fitted. Maximum likelihood was used for parameter estimation (as opposed to restricted maximum likelihood, REML) and the R library 'nlme' was accessed for model fitting.

3.) Generalized Additive Models with Correlated Errors – The final approach for addressing spatial autocorrelation involved fitting GAM_{err} models. These models are extensions of additive models (GAMs) fitted to normally distributed data (equation (5)) that allow specification of a general error covariance matrix, much like the extension provided by GLS for linear models. As with the spatial GLS analysis described above, combinations of variance and correlation structures were explored first with the saturated model, again including variance and correlation structures separately and together. Following selection of the optimal structure(s), the eight fixed effects parameterizations were then fitted. The R library 'mgcv' was accessed to fit all GAM_{err} models.

3.4.6 Model Evaluation, Selection, and Predictions

Model selection (except for GEE models) within each class was based on Akaike's Information Criterion (AIC; Akaike 1974, Burnham and Anderson 2002):

$$\mathrm{AIC} = -2\log(L) + 2p$$

where L is the estimated negative likelihood value and p is the number of estimated parameters. AIC provides insight regarding the trade-off between model complexity and goodness-of-fit, and the model with minimum AIC value is viewed as having received the most empirical support. Since GEEs are based on quasi-likelihood rather than maximum likelihood, there is no formal likelihood function and AIC or related measures cannot be used for model selection. Instead, comparisons of GEE models with different correlation structures were accomplished using the Quasi-likelihood under Independence Criterion (QIC; Pan 2001):

$$\mathrm{QIC} = -2Q(\beta_R, I) + 2\mathrm{trace}(\Omega_I^{-1}, V_R),$$

where the first term is the quasi-likelihood calculated under the independent working correlation assumption, evaluated at the parameter estimates under the specified correlation structure, and the second term involves the variance estimator (Ω_I) under the independence correlation structure and the robust variance estimator (V_R) under the specified correlation structure. $\mathrm{QIC_u}$ (Pan 2001) was used for model selection among the eight fixed effects parameterization, conditioned on the appropriate correlation structure:

$$\mathrm{QIC_u} = -2Q(\beta_R, I) + 2p$$

where p is the number of estimated parameters. The R library 'MuMIn' was accessed to calculate QIC and $\mathrm{QIC_u}$.

Fits of empirically supported models within each class were also evaluated through graphical analyses. For binomial models, plots of residuals against the linear or additive predictor were generated using binned residuals (Gelman et al. 2000) while Pearson residuals in relation to fitted values were examined for lognormal models. Temporal autocorrelation in model residuals was examined using partial autocorrelation functions (PACF), which provide insight about degrees of correlation among of two points lagged τ time units apart after any linear dependence has been removed. Spatial autocorrelation was evaluated with correlograms of Pearson residuals. Where possible, deviance explained by models was also calculated.

Given identification of the most supported model from each class, AIC and a 3-fold cross-validation analysis was used to discrimination between classes (for the CPUE data from the SKT survey, a 2-fold analysis was conducted due to convergence problems for some model types). The cross-validation analysis assesses model fit (training error) and prediction accuracy (testing error; Hastie et al. 2001) and involved randomly dividing the full data set into three sub-datasets of equal size. Each sub-dataset was then used as a test dataset for prediction while the remaining sub-datasets were considered training data for model fitting. Training and test error was calculated as:

$$\text{Training (Testing) error} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

where n is the number of observations, y_i is the i^{th} observation, and $\hat{y}_i$ is the associated estimated value. The model that produces lower training and testing errors is generally preferred.

From the overall selected model structure and parameterization, predictions in relation to each modeled covariate were then generated for the binomial and bias corrected lognormal components separately using marginal means (Searle et al. 1980). These predictions are intended to provide an understanding of how delta smelt presence-absence and density relate to covariates. However, when applying a delta-model to survey data, it is important to combine both model components to provide an integrated estimate of overall relative abundance, defined as the product of the encounter probability and density. Therefore, predictions representing overall estimated mean relative abundance were also generated for the modeled covariates common to both the binomial and lognormal components for each survey data set. Uncertainty estimates (coefficients of variation, CV) in predicted presence-absence and density across the domains of observed covariates were generated using nonparametric bootstrapping resampling (Efron and Tibshirani 1993).

4. Results

4.1 Fall Midwater Trawl

Following data filtering, a total of 7,441 tows remained in the FMWT dataset spanning the years 1985 – 2015. Of those, 6,190 were zeros and 1,251 were nonzero. Although the maximum catch was 115 delta smelt, which occurred in September of 2000 in statistical Area 13, the mean and median catches-per-tow over the full dataset were 0.89 and 0.0, respectively. These latter two statistics broadly indicate that low (near zero) delta smelt catches have been prevalent in the FMWT survey data for many years.

4.1.1 Binomial Component

4.1.1.1 Baseline Models

1.) Generalized Linear Models – For the eight GLM fixed effects models fitted to the presence-absence data, AIC was lowest for model M_6 followed by models M_7 (ΔAIC=1.0), M_8 (ΔAIC=1.2), and M_1 (ΔAIC=1.9). Deviance explained by model M_6 was 23.4%. Beyond the covariates included in all parameterizations, model M_6 also contained the covariate overall water depth, which suggests that the log odds encounter probability of delta smelt in the FMWT survey varies with water depth of sampling locations. The estimated coefficients for all the continuous covariates in model M_6 were negative and statistically significant ($p < 0.001$) suggesting the log odds of encounter probability decreases with increased temperature, specific conductance, Secchi depth, and overall depth. Also, the relative magnitudes of the estimated coefficients were highest for specific conductance and Secchi depth, modest for overall depth, and lowest for temperature.

The binned residuals plot for model M_6 showed no major problems as virtually all points fell within the ± two standard error boundary lines (Fig. 4A). The PACF plot of both the raw presence-absence data and GLM residuals showed evidence of temporal autocorrelation across several time lags (Figs. 4B, C) and correlograms of raw data and residuals showed clear signs of spatial autocorrelation (Fig. 4D). Collectively, the presence of autocorrelation implies the true variance in the data is being underestimated and precision of estimated fixed effect is likely overly optimistic.

2.) Generalized Additive Models – For the baseline GAM models, AIC statistics favored model M_1 followed by model M_8 (ΔAIC=2.4). Relaxation of the linearity assumption in favor of the additive predictor clearly supported an alternative inference about which covariates affect the log odds of FMWT encounter probabilities of delta smelt. In contrast to the results of the GLM analysis, model M_6 was not favored (ΔAIC=10.7) by the GAM analysis. Model M_1 is the saturated parameterization, which suggests that all covariates considered influence the log odds of delta smelt encounter probabilities in the FMWT survey. Thus, in addition to overall depth, hour of sampling and tidal stage influence the log odds of capturing delta smelt. Deviance explained by model M_1 was 27.6%, which was slightly better the most supported binomial GLM.

All of the smoothed terms were statistically significant ($p < 0.001$) and the basis dimensions chosen for each smoother indicated that over-smoothing was not problematic. Overall, the smoothed terms showed modest degrees of nonlinearity. Although model M_1 was favored over M_8,

the coefficients of the levels of the tidal stage covariate (the only difference between the models) were not significantly different from zero ($p > 0.05$).

As in the GLM analysis, the binned residuals plot for model M_8 showed no major problems (Fig. 5A). Temporal autocorrelation was again evident in the PACF plot of the residuals, although to a lesser extent in terms of both magnitude and time lags than the favored GLM model (Fig 5C). The spatial correlogram again provided evidence of spatial autocorrelation, although at a slightly lower magnitude than correlations estimated from residuals of the most supported GLM (Fig 5D). Again, the presence of autocorrelation implies that the precision of estimated parameters is likely too low.

4.1.1.2 Temporal Autocorrelation

1.) Generalized Estimating Equations – Comparison of the three correlation structures considered in the temporal GEE analysis revealed support for the exchangeable form, followed by AR1 ($\Delta QIC=2.3$) and independence (no autocorrelation; $\Delta QIC=6.1$). The stronger empirical support for exchangeable and AR1 correlation structures is consistent with the PACF results from the GLM and GAM models and strengthens the evidence for temporal autocorrelation within the FMWT presence-absence data. Support for the exchangeable correlation structure also indicates a lack of time-dependence among relatedness of within-station observations, which was surprising.

Given selection of the exchangeable correlation structure, comparison of the eight model parameterizations indicated model M_6 received the most empirical support, followed by models M_8 ($\Delta QIC_u=2.3$) and M_7 ($\Delta QIC_u=2.7$). These results are consistent with those from the GLM analysis that the log odds of encountering delta smelt in the FMWT is influenced by overall depth. The coefficients of all continuous covariates were again negative, statistically significant ($p < 0.012$), and the relative magnitudes of those coefficients followed the pattern estimated by the optimal GLM model. The estimate of the within-station correlation parameter was 0.075, which is low and indicative of a relatively weak relatedness among observations from the same station. As expected, accounting for temporal autocorrelation through GEEs did yield slightly different estimated fixed effects and larger standard errors when compared to the GLM results, however, overall statistical inference remained unchanged likely because of the low estimated autocorrelation.

2.) Generalized Linear Mixed Models – Model selection among the eight parameterizations again favored model M_6, followed by models M_7 ($\Delta AIC=0.34$) and M_8 ($\Delta AIC=2.0$). These results generalize the influence of overall depth and, to a slightly lesser degree, tidal cycle on the log odds of delta smelt capture probabilities by the FMWT survey across all model classes considered. Conclusions regarding statistical significance and rank order by magnitude of estimated coefficients from model M_6 were the same as those associated with the supported GLM and GEE models. However, specific to comparisons of the GEE and GLMM model results, it is important to keep in mind the PA versus SS interpretations. That is, GLMMs adhere to the SS interpretation and estimate fundamentally different quantities than marginal mean parameters such that estimated GLMM parameters will be larger in absolute magnitude than their GEE counterparts given a logit link function (Diggle et al. 1994, Fieberg et al. 2009). This was indeed the case when the GEE and GLMM model M_6 parameter estimates were compared.

The estimate of σ_b^2 from model M_6 was 0.76, which leads to σ_b= 0.87. Comparatively, the standard deviation of the random station effect was larger in magnitude than all of the estimated coefficients of the continuous covariates except specific conductance. This implies that the variation among stations is fairly high relative to the magnitude of the estimated environmental fixed effects. The estimate of ρ was 0.19 and indicative of somewhat low similarity of within-station observations. However, since the sampling intensity within- and between-years is high, the estimated d_{eff} was 22.9 which led to an estimate for N_{eff} of 324.6. Although this result suggests that the information contained in the overall sample of 7,441 tows analyzed for presence-absence could be captured by making approximately 325 tows, care needs to be given to the interpretation. By definition, fish surveys must operate annually since a primary goal is to collect data that can be analyzed to yield yearly indices of abundance. For a fixed station survey design that is executed over many years, the number of occasions that each location is sampled will become unavoidably large, thereby driving up the estimated d_{eff} value (recall that the equation for d_{eff} includes a parameter for the number of observations per subject/station). To address the effective sample size issue more directly, the year covariate was removed from model M_6 and the resulting parameterization was fitted to each year of FMWT data separately under the idea that the full survey data set is simply a collection of annually conducted experiments to assess delta smelt relative abundance. From each of the model fits, N_{eff} was estimated and the results were fairly wide ranging. In some years, N_{eff} was considerably lower than the actual number survey tows conducted, while in other years N_{eff} was approximately equal to the number of tows conducted (Fig. 6). However, the average N_{eff} across years was 78% of the actual yearly sample size, which provides some indication of oversampling.

4.1.1.3 Spatial Autocorrelation

1.) Autocovariate Models – Comparison of the saturated model parameterization with an autocovariate explanatory variable based on neighborhood radiuses ranging from 1 to 10 km indicated that a distance of 7 km was optimal. This result implies that the spatial relatedness of delta smelt presence-absence measurements in FMWT collections is best described by allowing neighborhoods surrounding sampling locations to extend outwardly a distance of 7 km.

Model selection statistics associated with fits of the eight parameterizations structured to include the autocovariate favored model M_6, followed by models M_8 (ΔAIC=0.2), model M_1 (ΔAIC =0.6), and model M_7 (ΔAIC=0.9). Such closeness among AIC statistics supports plausibility of all four model fixed effect combinations. Despite addressing an entirely different form of autocorrelation (spatial versus temporal), there is again strong support for the overall depth covariate followed closely by hour of sampling and tidal cycle. The coefficients for the continuous covariates included in model M_6 were again all negative and the relative magnitudes followed the previously described rank order. However, inclusion of the autocovariate in model M_6 led to lower estimated fixed effects when compared to those from the same baseline and temporal autocorrelation models. The deviance explained increased from 23.4% in the GLM model M_6 to 28.3% in the AUTO model M_6.

4.1.1.4 Summary

Evaluation of model selection statistics associated with the most supported baseline GLM, baseline GAM, temporal GLMM, and spatial AUTO models fitted to the FMWT presence-absence data revealed overwhelming empirical support for the autocovariate model. In relative AIC units, the other models did not compare (all ΔAIC values ≥ 32.4), which suggests that spatial autocorrelation is more pervasive than temporal autocorrelation in the FMWT delta smelt presence-absence data. This is related to the relatively low estimated within-station correlations from the temporal GEE and GLMM models. Unfortunately, from the perspective of identifying the most parsimonious description of the full data set, the favored GEE model cannot be compared to its counterparts from other classes because of reliance on quasi-likelihood estimation procedures. However, in the context of prediction accuracy, all models can be compared through cross-validation, and results from the 3-fold analysis showed that the baseline GAM and AUTO models yielded the lowest average training and testing errors (Table 2).

Recommendation: the collective evidence suggests that the optimal AUTO model should be used for statistical inference and generating predicted delta smelt encounter probabilities in relation to modeled covariates. This result leads to the inference that beyond the year, temperature, specific conductance, and Secchi depth covariates, overall water depth also plays an important role in affecting the presence-absence of delta smelt in FMWT survey collections. Model based predictions of capture probabilities over the domains of observed covariates along with estimated uncertainty are provided in Figs. 7A-E.

4.1.2 Lognormal Component

4.1.2.1 Baseline Models

1.) Generalized Linear Models – Examination of AIC statistics associated with GLM fits to the nonzero delta smelt CPUE data collected by the FMWT showed that model M_5 received the most empirical support, followed by models M_8 (ΔAIC=0.7), M_4 (ΔAIC=1.0), and M_6 (ΔAIC=1.7). Deviance explained by model M_5 was 20.7%. In addition to the covariates included in all parameterizations, model M_5 also contained the covariate hour of sampling, which implies there is a signal that CPUE of the FMWT survey varies throughout the sampling day, although formal statistical significance was not detected ($p>0.05$). Comparison of the second and third ranked models, M_8 and M_4, indicated some importance of overall water depth, but again, statistical significance was not detected ($p>0.05$). In decreasing rank order by magnitude, estimated coefficients from model M_5 were negative for specific conductance, Secchi depth, and hour of sampling, indicating an inverse relationship with CPUE. Statistical significance was evident for specific conductance and Secchi depth ($p < 0.001$). The estimated temperature effect was positive but small in magnitude and not statistically significant ($p>0.5$).

The plot of residuals against fitted values showed heterogeneous variance (Fig. 8A), which implies that the true variance in the CPUE data is not well estimated by the GLMs and the above statistical inferences should be taken with caution. The PACF plot of the log of the raw CPUE data showed slight temporal autocorrelation (Fig. 8B), however, the PACF of the GLM residuals showed

no evidence of temporal autocorrelation (Fig. 8C). This suggests that temporal relatedness in CPUE observations was essentially removed by the fixed effects in the GLM. Slight spatial autocorrelation was evident in both the raw data and model residuals (Fig. 8C).

2.) Generalized Additive Models – Model selection statistics of the eight fitted parameterizations overwhelmingly favored model M_8, which again highlights the relationship of CPUE with hour of sampling as well as overall depth. Competing empirical support was detected for model M_1 (ΔAIC=3.4). Deviance explained by model M_8 was 26.4%, which was slightly better the most supported lognormal GLM.

All of the smoothed terms were statistically significant ($p < 0.02$) except temperature, and the basis dimensions used for the smoothed terms indicated no concerns about over-smoothing. The extent of nonlinearity exhibited by the smoothed terms was modest. However, as in the GLM analysis, heterogeneous variance was evident (Fig. 9A), temporal autocorrelation in model residuals was not present (Fig 9C), and slight spatial autocorrelation in residuals was detected (Figs. 9D).

4.1.2.2 Temporal Autocorrelation/Heterogeneous Variance

1.) Generalized Estimating Equations – Given the lack of evidence for temporal autocorrelation from the GLM and GAM analyses, GEEs models were not applied to nonzero CPUE data from the FMWT survey.

2.) Generalized Linear Mixed Models – As noted previously, GLMMs can be structured to model heterogeneous variance, and given the strong evidence of increasing variance with fitted values from the GLM and GAM residuals plots (Figs 8A, 9A), those models were retained in the analytical framework despite no evidence of temporal autocorrelation.

Exploration of model fits with the various heterogeneous variance structures considered supported inclusion of the exponential of the mean form. Model selection statistics associated with eight fitted GLMMs with that variance function indicated support for both models M_4 and M_3 (ΔAIC=0.01), followed by models M_7 (ΔAIC=0.27) and M_6 (ΔAIC=0.66). These results differ from those of the baseline GLM analysis and downplay the effect of hour of sampling. In rank order from model M_4, negative effects were estimated for specific conductance, and Secchi depth while the effect of temperature was positive. Statistical significance was detected for only specific conductance and Secchi depth ($p < 0.001$). The exponential of the mean variance model lead to a considerably improved and much more acceptable residuals plot.

Because of the heterogeneous variance in the data as expressed though the exponential of the mean variance function, the estimate of σ_e^2 was quite large. By comparison, the estimate of σ_b^2 was negligible implying that, relative to the overall heterogeneous variance in the data, the relatedness of observations collected from the same sampling locations over time was approximately zero. As a result, the estimate of ρ was very small such that there was no discernable difference between the estimate of N_{eff} and the actual realized sample size.

3.) Generalized Additive Mixed Models – Investigation of the appropriate variance structure for the nonzero CPUE data with GAMMs fitted reflecting the saturated parameterization again supported the exponential of the mean form. For the subsequent eight fitted parameterizations, model selection statistics indicated model M_5 was favored, although there was comparable support for model M_8 (ΔAIC=1.24) and model M_2 (ΔAIC=1.53). These results differ somewhat from the baseline GAM analysis which draws attention to the effects of accommodating heterogeneous variance. Beyond the required covariates, model M_5 also included hour of sampling as opposed to the favored GAM model which also included both hour of sampling and overall depth. All of the smoothed terms from model M_5 were statistically significant with the exception of temperature, inspection of the basis parameters indicated that there was no evidence of oversmoothing, and the degree of estimated nonlinearity was modest. Although slightly different than the residuals plots from the GLM and GAM models, the exponential of the mean variance function again lead to greatly improved diagnostics.

4.1.2.3 Spatial Autocorrelation

1.) Spatial Error Models – For the SAR_{err} analysis, neighborhood radiuses could only be successfully estimated for distances ranging from 4 to 10 km. Attempts to fit models with smaller radiuses (1-3 km) were unsuccessful due to lack of convergence, which may be the related to the relative weak spatial autocorrelation associated with the baseline GLM and GAM correlograms (Figs. 7C, 8C). Comparisons of models that were able to be fitted supported a spatial radius definition of 4 km, so operationally with the SAR_{err} analysis this neighborhood was implemented. It should be noted that the neighborhood estimated by the AUTO analysis of the FMWT data was larger, which indicates broader spatial domains of relatedness among presence-absence than relative abundance.

Model selection statistics of the eight fitted SAR_{err} parameterizations favored both models M_6, and M_4 (ΔAIC=0.09). The estimated coefficients for the continuous covariates from model M_6 were again negative for specific conductance, Secchi depth, and overall depth, and statistical significance was detected for the first two ($p < 0.001$). The estimated coefficient of temperature was positive but not statistically significant ($p > 0.2$).

2.) Generalized Least Squares – Exploration of various combinations of variance and correlation structures showed that a parameterization with the exponential of the mean variance form alone received the most empirical support, followed by a model with both exponential of the mean variance and exponential correlation structures (ΔAIC=2.0). This result again suggests that the empirical support for spatial autocorrelation of the residuals is comparably not strong and can potentially be ignored in favor of a simpler parameterization designed to model variance heterogeneity alone. Structurally, the GLS analysis only differed from the baseline GLM analysis through the structure of the error covariance matrix.

Among the eight fitted GLS parameterizations, comparison of AIC statistics favored model M_4 followed by model M_5 (ΔAIC=2.2). These results indicate that the simplest model provided the most parsimonious description of the data with some support for the hour of sampling covariate, which differed from the favored GLM model. For the continuous model M_4 covariates, all

estimated coefficients were negative and statistical significance was detected for specific conductance and Secchi depth ($p < 0.001$), with the former being largest in magnitude.

3.) Generalized Additive Model with Correlated Errors – The results of the $\mathbf{GAM}_{err}$ analysis were very similar to those from the GLS analysis. That is, exploration of different combinations of variance and correlation structures revealed the parameterization with exponential of the mean variance form alone was favored based on model selection statistics (ΔAICs of other models ≈ 2.0). Thus, relaxation of the linearity assumption associated with the GLS models did not alter conclusions about the relative importance of heterogeneous variance and lack of spatially correlated residuals. And again, this result is consistent with previous modeling results regarding the relatively low degree of estimated spatial autocorrelation.

Comparisons of the eight fitted parameterizations with the exponential of the mean variance structure indicated that model M_8 received the most empirical support, followed by models M_1 (ΔAIC=0.32), M_5 (ΔAIC=0.37), and M_2 (ΔAIC=0.77). These results lend support for the effects of covariates hour of sampling and depth, and to a lesser degree, tidal cycle. The basis dimensions used for the smoothed terms indicated no concerns about over-smoothing, modest nonlinearity and statistical significance was detected for the smoothed terms associated with specific conductance, Secchi depth, and hour of sampling ($p < 0.001$). The smoothed terms for temperature and depth were both nearly statistically significant ($0.06 < p < 0.08$).

4.1.2.4 Summary

Comparison of model selection statistics estimated from the favored baseline GLM, baseline GAM, GLMM, GAMM, SAR, GLS, and $\mathbf{GAM}_{err}$ models fitted to the nonzero CPUE data from the FMWT survey revealed overwhelming empirical support for the GAMM model. In relative AIC units, the other models did not compare (all ΔAIC values ≥ 23.0). Accommodating the heterogeneous variance was clearly very important, as was the nonlinearity afforded by the additive model structure even though it was low-to-modest across the modeled covariates. Evidence for spatial autocorrelation was not strong amongst multiple model structures, which provides convergent lines of evidence suggesting that this form of dependence among residuals could effectively be ignored. Results of the 3-fold cross-validation analysis showed that the baseline GAM and GAMM models yielded the lowest training and testing errors (Table 3), although the former model has very obvious structure limitations for proper quantification of uncertainty.

Recommendation: the collective evidence suggests that the most supported GAMM model should be used for statistical inference and generating predicted delta smelt density in relation to modeled covariates. This result leads to the inference that beyond the year, temperature, specific conductance, and Secchi depth covariates, hour of sampling also plays an important role in affecting the CPUE of delta smelt in FMWT survey collections. Model based predictions of delta smelt density over the domains of observed covariates along with estimated uncertainty are provided in Figs. 10A-E. Predictions from the combined binomial and lognormal models are given in Figs. 11A-D.

4.2 Spring Kodiak Trawl

Following data filtering, a total of 2,551 tows remained in the SKT dataset spanning the years 2002 – 2015. Of those, 1,846 were zeros and 705 were nonzero. The maximum catch of 375 delta smelt occurred in January of 2009. Although the mean catch-per-tow over the full dataset was 2.8 fish and higher than that of the FMWT, the median was again 0.0 indicating sparse presence of delta smelt in the SKT survey.

4.2.1 Binomial Component

4.2.1.1 Baseline Models

1.) Generalized Linear Models – For the eight GLM fixed effects models fitted to the presence-absence data, AIC was lowest for model M_1 followed by model M_7 (ΔAIC=3.7). Deviance explained by model M_1 was 21.9%. This result implies the fully saturated model was most supported which suggests that, beyond the covariates included in each parameterization, the log odds encounter probability of delta smelt in the SKT survey varies with tidal cycle, hour of sampling, and overall water depth. The estimated coefficients for all the continuous covariates in model M_1 except overall depth were negative and all were statistically significant ($p < 0.01$). The relative magnitudes (absolute value) of the estimated coefficients ranked from largest to smallest were Secchi depth, specific conductance and temperature (indistinguishable), depth, and hour of sampling.

The binned residuals plot for model M_1 showed no major problems as all but one point fell within the ± two standard error boundary lines (Fig. 12A). The PACF plot of raw presence-absence data indicated a high degree of temporal autocorrelation (Fig. 12B) which was less but still evident in the PACF plot of GLM residuals (Fig. 12C). The correlograms of raw data and model residuals both showed signs of spatial autocorrelation (Fig. 12D). Collectively, the presence of autocorrelation implies the true variance in the data is being underestimated and precision of estimated fixed effect is overly optimistic.

2.) Generalized Additive Models – For the baseline GAM models, AIC statistics favored model M_1 followed by model M_7 (ΔAIC=3.3). These results indicate that relaxation of the linearity assumption in favor of the additive predictor did not supported an alternative inference about which covariates affect the log odds of SKT encounter probabilities of delta smelt. Deviance explained by model M_1 was 30.3%, which was notably better the most supported binomial GLM.

All of the smoothed terms were statistically significant ($p < 0.035$) and the basis dimensions chosen for each smoother were appropriate except for the covariate overall depth. More flexibility was required for this covariate and the estimated pattern showed a high degree of nonlinearity. For all other smoothed terms, modest nonlinearity was detected.

As in the GLM analysis, the binned residuals plot for model M_1 showed no major problems (Fig. 13A). Slight temporal autocorrelation was evident in the PACF plot of the residuals at fewer time-lags when compared to the PACF residual plot from the favored GLM model (Fig. 13C). The spatial correlogram provided evidence of slight spatial autocorrelation but weaker in magnitude than correlations estimated from residuals of the most supported GLM (Fig. 13D).

4.2.1.2 Temporal Autocorrelation

1.) Generalized Estimating Equations – Comparison of the three correlation structures considered in the temporal GEE analysis revealed support for the AR1 form, followed by independence (no autocorrelation; ΔQIC=1.2). The stronger empirical support for AR1 correlation structures is consistent with the PACF results from the GLM and GAM models and strengthens the evidence for temporal autocorrelation within the SKT presence-absence data. Support for the AR1 correlation structure also indicates a time-dependence among relatedness of within-station observations.

Given selection of the AR1 correlation structure, comparison of the eight model parameterizations indicated model M_1 received the most empirical support, followed by models M_7 (ΔQIC_u=3.8). These results are consistent with those from the GLM and GAM analyses that the log odds of encountering delta smelt in the SKT is influenced by all covariates considered. The coefficients of all continuous covariates except overall depth were again negative, but only Secchi depth, specific conductance, and temperature were statistically significant ($p < 0.001$). The relative magnitudes of those coefficients followed the pattern estimated by the favored GLM model. The estimate of the within-station correlation parameter was 0.10, which is larger than that estimated by the optimal GEE model from the FMWT analysis. As expected, accounting for temporal autocorrelation through GEEs did yield slightly different estimated fixed effects and larger standard errors when compared to the GLM results, and while model selection results remained unchanged, inferences based on traditional hypothesis testing were different.

2.) Generalized Linear Mixed Models – Comparison of the eight parameterizations fitted favored model M_3, followed by models M_7 (ΔAIC=0.70) and M_2 (ΔAIC=1.92). These results highlight the importance of tidal cycle and deemphasize the previously noted importance of hour of sampling and overall depth on the log odds of delta smelt capture probabilities by the SKT survey. The estimated coefficients for all continuous covariates were negative, statistically significant ($p < 0.001$), and the magnitudes in decreasing rank order were Secchi depth, temperature, and specific conductance. Thus, the results of the GLMM analysis indicate a stronger relative effect of temperature than estimated by the optimal GLM and GEE models.

The estimate of σ_b^2 from model M_6 was 2.21, which leads to σ_b= 1.49. Comparatively, the standard deviation of the random station effect was larger in magnitude than all of the estimated coefficients of the continuous covariates. This implies that the variation among stations is very high relative to the magnitude of the estimated environmental fixed effects. The estimate of ρ was 0.40 and indicative of appreciable similarity of within-station observations. The estimated d_{eff} was 26.2 which led to an estimate for N_{eff} of 97.4. These results indicate that the information contained in the overall sample of 2,552 tows analyzed for presence-absence could be captured by making approximately 98 tows, although the same care discussed in the FMWT GLMM analysis is also germane here. Addressing the effective sample size more directly through the annual analysis described in the FMWT GLMM analysis yielded estimates of N_{eff} that were consistently lower than the actual number survey tows conducted (Fig. 14). The average N_{eff} across years was 48% of the actual yearly sample size, which provides fairly strong evidence of oversampling.

4.2.1.3 Spatial Autocorrelation

1.) Autocovariate Models – Exploration of neighborhood radiuses ranging from 1 to 10 km using the saturated model parameterization with an autocovariate explanatory variable revealed that a distance of 2 km was optimal. Thus, the spatial relatedness of delta smelt presence-absence measurements in the SKT survey is best described by allowing neighborhoods surrounding sampling locations to extend outwardly a distance of 2 km.

Model selection statistics associated with fits of the eight parameterizations structured to include the autocovariate supported model M_7, followed by models M_3 (ΔAIC=0.45). Such closeness among AIC statistics supports plausibility of both models. Despite addressing an entirely different form of autocorrelation (spatial versus temporal), there is again support for the tidal cycle covariate and overall depth. The coefficients for the continuous covariates included in model M_7 were all negative and statistically significant ($p < 0.001$), with the exception of the low estimated positive effect for overall depth ($p > 0.1$). As noted from the most supported GLMM model, the estimated effect of temperature ranked second in between Secchi depth and specific conductance. The deviance explained by AUTO model M_7 was 31.5%.

4.1.1.4 Summary

Comparison of AIC model selection statistics associated with the favored baseline GLM, baseline GAM, temporal GLMM, and spatial AUTO models fitted to the SKT presence-absence data revealed overwhelming empirical support for the GLMM model (all ΔAIC values ≥ 24.7). This result suggests that accommodating temporal autocorrelation is more important than accommodating spatial autocorrelation in the SKT delta smelt presence-absence data. Recall that the nonlinearity of the baseline GAM appeared to reduce the strength of both forms of autocorrelation, which suggests that a GAMM may be optimal. However, current software for fitting binomial GAMMs rely on penalized quasi-likelihood estimation which is known to lead to biased parameters estimates, particularly when expected numbers of successes are low (< 5, Bolker et al. 2009). Binomial GAMMs represent an attractive future modeling option for the SKT delta smelt presence-absence data should software packages evolve alternative estimation frameworks for these models. Although results from the 3-fold analysis indicated the baseline GAM and AUTO models yielded the lowest average training and testing error (Table 4), all model types performed similarly such that the improved empirical support for the GLMM outweighs the slight loss in prediction accuracy.

Recommendation: the collective evidence suggests that the most supported GLMM model should be used for statistical inference and generating predicted delta smelt encounter probabilities in relation to modeled covariates. This result leads to the inference that beyond the year, temperature, specific conductance, and Secchi depth covariates, tidal cycle also plays an important role in affecting the presence-absence of delta smelt in SKT survey collections. Model based predictions of capture probabilities over the domains of observed covariates along with estimated uncertainty are provided in Figs. 15A-E.

4.2.2 Lognormal Component

4.2.2.1 Baseline Models

1.) Generalized Linear Models – Model comparisons of the GLM models fitted to the nonzero delta smelt CPUE data collected by the SKT survey showed that model M_6 received the most empirical support, followed by models M_8 (ΔAIC=1.9) and M_7 (ΔAIC=2.3). Deviance explained by model M_6 was 14.9%. In addition to the covariates included in all parameterizations, model M_6 also contained the covariate overall depth. Comparison of the second and third ranked models both include overall depth which strengthens support for that covariate. The estimated coefficients of the continuous covariates from model M_6 were all negative except for overall depth and suggestive of inverse relationships with CPUE. Absolute values of coefficients in decreasing rank order were: Secchi depth, temperature, overall depth, and specific conductance. Statistical significance ($p < 0.006$) was detected for all continuous covariates except specific conductance ($p=0.74$), which indicated little effect of salinity on the CPUE of delta smelt sampled by the SKT survey.

The plot of residuals against fitted values showed heterogeneous variance (Fig. 16A), which implies that the true variance in the CPUE data is likely underestimated by the GLMs and the above statistical inferences should be taken with caution. The PACF plots of both the log of the raw CPUE data and the GLM residuals showed a high degree of temporal autocorrelation (Figs. 16B, C) and the correlograms of the log of the raw CPUE data and model residuals were very similar and both suggestive notable spatial autocorrelation (Fig. 16D).

2.) Generalized Additive Models – Exploration of the model selection statistics of the eight fitted parameterizations favored model M_8, followed by model M_1 (ΔAIC=2.2). Clearly, the use of an additive predictor over the linear predictor altered inference regarding importance of the covariates considered. Deviance explained by model M_8 was 36.1%, which was more than double that explained by the most supported lognormal GLM.

Only the smoothed terms for temperature, Secchi depth, and overall depth were statistically significant ($p < 0.048$) and the basis dimensions chosen for each smoother were appropriate except for the covariate overall depth. More flexibility was required for this covariate and the estimated pattern showed a high degree of nonlinearity. For all other smoothed terms, modest nonlinearity was detected. As in the GLM analysis, heterogeneous variance was evident (Fig. 17A), yet the degree of both temporal and spatial autocorrelation was greatly reduced compared to the favored GLM model (Figs. 17C, D) and likely due to the nonlinearity afforded by the additive predictor.

4.2.2.2 Temporal Autocorrelation/Heterogeneous variance

1.) Generalized Estimating Equations – Comparison of the three correlation structures considered in the temporal autocorrelation GEE analysis overwhelmingly supported the exchangeable structure (all (ΔQIC > 24.1), which indicates no time dependence of the relatedness of samples taken from the same sampling station. Model selection among the eight parameterizations with the exchangeable correlation structure favored model M_6, followed by model M_7 (ΔQICu=1.8). Comparison of output from GEE and GLM models M_6 were very similar in terms of signs, magnitudes, and relative rankings of estimated coefficients. However, statistical significance in the GEE model M_6 was only detected for the temperature covariate ($p < 0.001$) as opposed to the significance found for

temperature, Secchi depth, and overall depth in the favored GLM model. The estimate of the within-station correlation parameter was 0.59 from GEE model M_6, which is fairly high and suggestive of considerable similarity among observations from the same station, at least in the context of a linear predictor.

2.) Generalized Linear Mixed Models – Exploration of the different variance structures supported use of the exponential of the mean form, as was the case in the GLMM analysis of the FMWT data. As noted previously, diagnostic plots associated with fits of the favored GLM and GAM models both showed strong patterns of increasing variance with fitted values (Figs 13A, 14A), so the improved performance of models including a variance function such as the exponential of the mean was not surprising. Evaluation of model selection statistics associated with eight fitted GLMMs including the exponential of the mean variance function indicated support for both models M_4, followed by models M_5 (ΔAIC=1.4) and M_6 (ΔAIC=1.5). These results differ somewhat from those of the baseline GLM analysis in that the effect of the covariate overall depth was considerably lower. Negative effects were estimated for all continuous covariates in model M_4 and statistical significance was associated only with the estimated temperature effect ($p < 0.001$).

Because of the heterogeneous variance in the data as expressed though the exponential of the mean variance function, the estimate of σ_e^2 was quite large. By comparison, the estimate of σ_b^2 was only small implying that, relative to the overall heterogeneous variance in the data, the relatedness of observations collected from the same sampling locations over time was negligible. The estimate of ρ was approximately zero such that there is no discernable difference between the estimate of N_{eff} and the actual realized sample size.

3.) Generalized Additive Mixed Models – Comparisons of different variance structures for the CPUE data with GAMMs fitted to the saturated parameterization again supported the exponential of the mean model. For the eight parameterizations structured to include the heterogeneous variance form, model selection statistics indicated model M_4 was favored, followed by model M_3 (ΔAIC=2.26). These results differ from the baseline GAM analysis which draws attention to the effects of accommodating temporal autocorrelation and heterogeneous variance. The simplest GAMM parameterization was supported as opposed to the optimal GAM model which included the covariates hour of sampling and overall depth. Only the smoothed term for temperature from model M_4 was statistically significant ($p < 0.001$), and inspection of the basis parameters indicated that there was no evidence of oversmoothing and that the degree of estimated nonlinearity was modest. The estimate of σ_e^2 was again very large such that the residual variance far out weighted the estimate of σ_b^2. The parameter ρ was near zero and N_{eff} was approximately the same as the actual realized sample size.

4.2.2.3 Spatial Autocorrelation

1.) Spatial Error Models – Comparison of the saturated SAR_{err} model parameterization with the spatial weights matrix based on neighborhood radiuses ranging from 6 to 10 km indicated that a distance of 7 km was optimal (attempts to fit models with smaller radiuses were unsuccessful). Therefore, the spatial relatedness of delta smelt CPUE measurements in SKT survey samples is best

described by allowing neighborhoods to extend outwardly a distance of 7 km, which is less than the optimal neighborhood estimated by the AUTO analysis of the presence-absence data.

Model selection statistics of the eight fitted parameterizations favored model M_6, followed by models M_8 (ΔAIC=0.3), M_1 (ΔAIC=2.8), and M_7 (ΔAIC=2.8). Addressing an entirely different form of autocorrelation (spatial versus temporal) altered inferences from the simplest parameterization to one that included overall depth. Absolute values of estimated coefficients in decreasing rank order were associated with covariates temperature, Secchi depth, overall depth, and specific conductance. Statistical significance ($p < 0.001$) was detected for all continuous covariates except specific conductance ($p>0.7$), which again indicates little effect of salinity on the CPUE of delta smelt sampled by the SKT survey.

2.) Generalized Least Squares – Model selection statistics associated with various combinations of variance and correlation structures showed that a parameterization with the exponential of the mean variance form alone received the most empirical support, followed by a model with exponential of the mean variance and a spherical correlation structure (ΔAIC=2.0). This result suggests that the empirical support for spatial autocorrelation of the residuals is comparably not strong and can potentially be ignored in favor of a simpler parameterization designed to model variance heterogeneity alone. Although the GLS analysis of the FWMT survey CPUE data yielded similar results, the estimated correlations from the correlogram were considerably lower than those associated with the correlogram from the GLM analysis of the SKT data. Hence, empirical support in favor of ignoring spatial autocorrelation within the SKT was somewhat surprising.

Among the eight fitted parameterizations structured to only accommodate heterogeneous variance, model selection statistics overwhelmingly favored model M_6. No other parameterization received comparable empirical support. These results again highlight the effect of overall depth on the CPUE data collected by the SKT survey. For the continuous model M_6 covariates, the signs, relative magnitudes, and statistical significance was the same as in the favored model from the SAR_{err} analysis.

3.) Generalized Additive Model with Correlated Errors – The results of the GAM_{err} analysis were very similar to those from the GLS analysis. That is, exploration of different combinations of variance and correlation structures revealed the parameterization with exponential of the mean variance form alone was favored based on model selection statistics (ΔAICs of other models $\approx$ 2.0). Thus, relaxation of the linearity assumption associated with the GLS models did not alter conclusions about the relative importance of heterogeneous variance and spatially correlated residuals.

Comparisons of the eight fitted parameterizations with the exponential of the mean variance structure indicated that model M_6 received the most empirical support, followed by model M_8 (ΔAIC=3.3). These results lend support for the effect of covariate overall depth. The basis dimensions used for the smoothed terms followed those used for the baseline GAM analysis and statistical significance ($p < 0.002$) was detected for the smoothed terms associated with all continuous covariates except specific conductance ($p > 0.10$).

4.1.2.4 Summary

Comparison of AIC statistics from the optimal baseline GLM, baseline GAM, GLMM, GAMM, SAR, GLS, and $\mathtt{GAM}_{err}$ models fitted to the CPUE data of the SKT survey indicated empirical support for the temporal GLMM model, followed by the temporal GAMM (ΔAIC=5.1). Clearly, the heterogeneous variance model structure was an important element to incorporate as was accommodating the longitudinal nature of the data through the random effect. Since model M_4 was optimal for both the GLMM and GAMM, and the degree of nonlinearity exhibited by the smoothed terms in the additive model was modest, the additional structure and estimated parameters associated with the GAMM were not necessary. Evidence for spatial autocorrelation in the CPUE data of the SKT survey was not strong suggesting that this form of dependence among residuals could effectively be ignored. Results of the 2-fold cross-validation analysis showed that the GLMM and baseline GAM models yielded the lowest training and testing errors (Table 5). Although the optimal GEE model could not be included in AIC-based model selection comparisons, it did not perform well in the cross-validation analysis.

Recommendation: the collective evidence suggests that the most supported GLMM model should be used for statistical inference and generating predicted delta smelt density in relation to modeled covariates. This result leads to the inference that only the year, temperature, specific conductance, and Secchi depth covariates play important roles in affecting the CPUE of delta smelt in SKT survey collections. Model based predictions of capture probabilities over the domains of observed covariates along with estimated uncertainty are provided in Figs. 18A-D. Predictions from the combined binomial and lognormal models are given in Figs. 19A-D.

5. Discussion

Over the past several decades, a great deal of research has been conducted to advance the scientific enterprise of the Delta with regards to fish life history/life-cycle biology and ecology, including investigations of habitat preferences and how fish populations or components of populations interact with surrounding environments. Two arguably galvanizing studies for the Delta are those by Jassby et al. (1995) and Kimmerer (2002) which reported the importance of variables such as X_2 (defined as the horizontal distance up the axis of the estuary where the tidally averaged near-bottom salinity is 2 psu) and flow within the Delta on the relative abundance (or survival) for several fish species. However, the conclusions of both studies were derived from statistical regressions relating annual indices of relative abundance (or survival) to those environmental variables. Such an approach may be attractive due to its simplicity, but it must be recognized that any index of abundance is a synthesis of a large number of raw field observations. Although the primary purpose of a fish survey is to obtain a collection of measures of fish relative abundance across different time periods and spatial locations, associated measurements of environmental parameters are also routinely recorded for the explicit purpose of providing synoptic representations of how fishes are interacting with the surrounding environment. Therefore, analytical efforts that focus on only annual abundance indices explicitly ignore and lose the wealth of highly informative auxiliary data intentionally collected with each stand-alone survey observation.

In the case of the FMWT, the overall loss of information is substantial since each annual index of relative abundance is based on ~400 individual survey samples. Despite the outstanding efforts of biologists working in the Delta to continually maintain extensive, high quality fish survey databases such as those derived from the FWMT and SKT sampling programs, very little attention has been directed at analyzing raw data for the purposes of gaining insight into the biology/ecology of resident fish species. Feyrer et al. (2007, 2011) and Latour (2016) represent the only published studies that have analyzed raw fish survey data from the Delta, and the former two focused exclusively on presence-absence data as opposed to CPUE or some integration of presence-absence and CPUE. The paucity of such analyses is inconsistent with how fish survey data are treated worldwide (e.g., development of standardized indices of abundance using model based procedures for stock assessments of commercially harvest aquatic resources). This context was a key motivation for the current study.

Statistical analyses of large datasets such as those associated with longstanding fish surveys can be challenging since sound inference science requires careful treatment of the underlying data and explicit consideration of the assumptions inherent to analytical methods. As evidenced by the results of this study, the FMWT and SKT delta smelt survey data each contain several similar characteristics that each should be accommodated within a statistical modeling framework. Most notably are the excessive number of zero observations, longitudinal nature of the data (fixed station designs where measurements are repeatedly taken from the same locations over time), and the relatively high spatial density of samples taken throughout the respective sampling frames. All of these structural features within the FMWT and SKT survey datasets were explicitly accommodated in the current study.

5.1 Presence-Absence

The diagnostic plots associated with the baseline binomial GLM and GAM models indicated evidence of temporal and spatial autocorrelation in the presence-absence data from both surveys. However, the model evaluation, comparison, and selection procedures employed supported a spatial model for the FMWT data and a temporal model for the SKT data. For the FMWT survey data, such strong support for a spatial model implies that there is considerable relatedness amongst delta smelt capture probabilities across space such that an AUTO model with a 7 km spatial neighborhood was favored. A possible explanation of this result is that delta smelt during the fall season are not clustered such that presence-absence differences are relatively homogeneous across modestly large spatial scales. In turn, this raises questions about the need for such highly spatially resolved sampling. For example, within two of the core statistical sampling areas of the FMWT survey (i.e., Areas 12, 13) ten or more trawl tows are conducted during each research cruise. Alternatively, it may be worth exploring the necessity of conducting four monthly cruises each year. Although the temporal GLMM model received considerably less empirical support than the spatial AUTO model, results of the GLMM analysis did indicate that for approximately half of the years analyzed, the effective sample size was notably lower than the actual number of trawl tows conducted.

For the SKT survey data, the strong empirical support for the GLMM offers fairly convincing evidence that there appreciable relatedness among presence-absence observations within

sampling locations over time. This result suggests that there are stations within the SKT sampling frame where delta smelt are routinely and consistently captured and there are other stations where delta smelt are not regularly captured. Since the SKT survey is conducted during winter/spring months, a potential explanation relates to spawning behavior of adult delta smelt and that movements may be restricted during this time period. Annual estimates of effective sample size were consistently below the actual number of trawl tows conducted (less than half in several years).

Delta smelt is a rare species and it is recognized that survey capture probabilities will undoubtedly be low. However, the autocorrelative nature of the data from both surveys can potentially form the basis for discussions surrounding the current designs of both surveys. For example, reallocation of a small portion of current sampling efforts (e.g., 10-15%) within either program would not likely lead to much loss of overall delta smelt information, yet those resources could be directed toward auxiliary studies employing alternative designs such as stratified random sampling or depletion sampling. Discussions of this nature could prove useful moving forward.

Regarding the importance of modeled covariates on the presence-absence of delta smelt in both survey programs, the results of the AUTO and GLMM models draw attention to the variables overall water depth and tidal cycle, respectively. For the FMWT survey, capture probabilities modestly declined with increasing depth which could be the result of either differential water column usage by delta smelt or depth-associated changes in efficiency of trawl gear. It is worth noting that the effect of overall depth on FMWT delta smelt capture probabilities was similar in magnitude as the effect of temperature, which is somewhat surprising given the well-known physiological and behavioral effects of the later in fishes. Delta smelt capture probabilities by the SKT survey varied tidally, with the highest occurring during slack tide and the lowest during ebb tide. Although the estimated effects were not large in magnitude, it is quite plausible that delta smelt behaviorally seek alternative habitat types or locations under conditions of increased water movement, even within relatively localized areas and over short time scales. The differential effects between ebb and flood suggest that those behavioral changes may be exacerbated during conditions of outflowing water.

5.2 CPUE

The diagnostic plots associated with the baseline lognormal GLM and GAM models fitted to the CPUE data of the FMWT survey indicated very little evidence of temporal or spatial autocorrelation. The most supported final model was the heterogeneous GAMM rather than any of the autocorrelation models. This result suggests that there is negligible temporal or spatial relatedness among FMWT collections that successfully capture at least one delta smelt, and that the number of delta smelt captured across those tows is highly variable. This latter concept is consistent with the notion that delta smelt do not form schools due to a lack of synchronized swimming. With respect to the importance of modeled covariates on CPUE data from the FMWT survey, the GAMM results highlight the importance of hour of sampling. The predicted pattern of density with increasing time-of-day was 'U' shaped, which suggests higher observed CPUE levels during crepuscular time periods. This result is typical for fishes that engage in diel vertical behavior where feeding increases during darker portions of the day when zooplankton are more stratified in the water column. However, it should be noted that majority of crepuscular survey observations within

the FMWT survey dataset occurred early in the time-series such that there appears to have been an evolution in the daily timing of survey operations. Crepuscular survey tows are essentially absent from the data in recent years. Another notable result was the lack of a temperature effect on CPUE from the FMWT survey (recall that temperature was retained in all model parameterizations).

Although the extent of temporal and spatial autocorrelation in the CPUE data of the SKT was greater than that in the FMWT data, again the most supported model was one structured with only heterogenous variance. The aforementioned conclusions about the role autocorrelation in the data and the highly variable CPUE catches also apply to the SKT survey data. In contrast to the FMWT results, the estimated effect of the temperature covariate was greatest among the continuous covariates modeled and simplest parameterization was favored, which suggests no effects of overall water depth, tidal cycle, or hour of sampling. Predicted CPUE was highest under the coldest observed temperatures which occur early in the SKT sampling year and may be related to aggregation behavior associated with spawning.

6. Literature Cited

Aitchison, J. 1955. On the distribution of a positive random variable having a discrete probability mass at the origin. Journal of the American Statistical Association 50:901–908.

Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle. In Second international symposium on information theory, ed. B.N. Petrov and F. Csaki, 267–281. Budapest: Akademiai Kiado.

Alpine, A.E., and J.E. Cloern. 1992. Trophic interactions and direct physical effects control phytoplankton biomass and production in an estuary. Limnology and Oceanography 37: 946–955.

Atwater, B.F., S.G. Conrad, J.N. Dowden, C.W. Hedel, R.L. MacDonald, and W. Savage. 1979. History, landforms and vegetation of the estuary's tidal marshes. In San Francisco Bay: The Urbanized Estuary, ed. T.J. Conomos, 347–385. San Francisco: Pacific Division American Association for the Advancement of Science.

Augustin, N. H., M.A. Mugglestone, and S.T. Buckland. 1996. An autologistic model for the spatial distribution of wildlife. Journal of Applied Ecology 33:339-347.

Bardos, D.C., G. Guillera-Arroita, and B.A. White. 2015. Valid auto-model for spatially autocorrelated occupancy and abundance data. Methods in Ecology and Evolution 6:1137-1149.

Besag, J. 1974. Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society B 36:192-236.

Bolker, B.M., M.E. Brooks, C.J. Clark, S.W. Geange, J.R. Poulsen, M.H.H. Stevens, and J.S. White. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. Trends in Ecology and Evolution 24:127-135.

Box, G.E.P., G.M. Jenkins, and G.C. Reinsel. 1994. Time Series Analysis: Forecasting and Control. Englewood Cliffs, N.J.: Prentice-Hall.

Beverton, R.J.H. and S.J. Holt. 1957. On the Dynamics of Exploited Fish Populations. Chapman & Hall.

Bennett, W. 2005. Critical assessment of the delta smelt population in the San Francisco Estuary, California. San Francisco Estuary and Watershed Science 3.

Buchheister, A., C.F. Bonzek, J. Gartland, and R.J. Latour. 2013. Patterns and drivers of the demersal fish community in Chesapeake Bay. Marine Ecology Progress Series 481:161–180.

Burnham, K.P., and D.R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. New York: Springer.

Carl, G. and I. Kuhn. 2007. Analyzing spatial autocorrelation in species distribution using Gaussian and logit models. Ecological Modelling 207: 159-170.

Cohen, A.N., and J.T. Carlton. 1998. Accelerating invasion rate in a highly invaded estuary. Science 279: 555–558.

Connor, M.S., J.A. Davis, J. Leatherbarrow, B.K. Greenfield, A. Gunther, D. Hardin, T. Mumley, J.J. Oram, and C. Werme. 2007. The slow recovery of San Francisco Bay from the legacy of organochlorine pesticides. Environmental Research 105:87–100.

Cragg, J.G. 1971. Some statistical models for limited dependent variables with applications to the demand for durable goods. Econometrica 39:829–844.

Cressie, N.A.C. 1993. Statistics for Spatial Data (revised edition). John Wiley & Sons, NY.

Dick, E.J. 2004. Beyond 'lognormal' versus 'gamma': discrimination among error distributions for generalized linear models. Fisheries Research 70:351-366.

Diggle, P.J., K.Y. Liang, and S.L. Zeger. 1994. Analysis of Longitudinal Data. Oxford University Press, London.

Dormann, C. F. and 16 co-authors. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30:609-628.

Efron, B., and R.J. Tibshirani. 1993. An introduction to the bootstrap, vol.57. London: Chapman and Hall.

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.

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.

Fieberg, J., R.H. Rieger, M.C. Zicus, and J.S. Schildcrout. 2009. Regression modelling of correlated data in ecology: subject-specific and population averaged response patterns. Journal of Applied Ecology 46:1018-1025.

Fletcher, D. 2008. Confidence intervals for the mean of the delta-lognormal distribution. Environmental Ecology Statistics 15:175-189.

Gelman, A. 2000. Diagnostic checks for discrete data regression models using posterior predictive simulations. Applied Statistics 49:247-268.

Gumpertz, M.L., J.M. Graham, and J.B. Ristaino. 1997. Autologistic model of spatial pattern of *Phytophthora* epidemic in bell pepper: effects of soil variables on disease presence. Journal of Agricultural, Biological, and Environmental Statistics 2:131-156.

Hastie, T., R. Tibshirani, and J. Friedman. 2001. The Elements of Statistical Learning: Data Mining, Inference and Prediction, 2nd ed. Springer, New York.

Hilborn, R. and C. Walters. 1992. Quantitative Fisheries Stock Assessment: Choice, Dynamics, & Uncertainty. Chapman & Hall.

Hedeker, D. and R.D. Gibbons. 2006, Longitudinal Data Analysis. John Wiley & Sons.

Jassby, A., W. Kimmerer, S. Monismith, C. Armor, J. Cloern, T. Powell, J. Schubel, and T. Vendlinski. 1995. Isohaline position as a habitat indicator for estuarine populations. Ecological Applications 5:272–289.

Kimmerer, W. 2002. Effects of freshwater flow on abundance of estuarine organisms: physical effects or trophic linkages? Marine Ecological Progress Series 243:39-55.

Kimura, D.K. and D.A. Somerton. 2006. Review of statistical aspects of survey sampling for marine fisheries. Reviews in Fisheries Science 14:245-283.

Kissling, W.D. and G. Carl. 2008. Spatial autocorrelation and the selection of simultaneous autogressive models. Global Ecology and Biogeography 17:59-71.

Latour, R.J. 2016. Explaining patterns of pelagic fish abundance in the Sacramento-San Joaquin Delta. Estuaries and Coasts 39:233-247.

Legendre, P. and L. Legendre. 2012. Numerical ecology, 3nd English edition. Elsevier.

Liang, K.Y. and S.L. Zeger. 1986. Longitudinal data analysis using generalized linear models. Biometrika 73:13-22.

Lin, X. and D. Zhang. 1999. Inference in generalized additive mixed models using smoothing splines. Journal of the Royal Statistical Society, Series B 61:381-400.

Lo, N.C.H., L.D. Jacobson, and J.L. Squire. 1992. Indices of relative abundance from fish spotter data based on delta-lognormal models. Canadian Journal of Fisheries and Aquatic Sciences 49: 2515–2526.

Malak, A.J., J.S. Collie, and J. Gartland. 2014. Fine-scale spatial patterns in the demersal fish and invertebrate community in a northwest Atlantic ecosystem. Estuarine, Coastal, and Shelf Science 147:1-10.

Maunder, M.N. and A.E. Punt. 2004. Standardizing catch and effort data: a review of recent approaches. Fisheries Research 70:141–159.

McCullagh, P., and J.A. Nelder. 1989. Generalized linear models, 2nd ed. Chapman & Hall.

McMullan, A., A.W. Bowman, E.M. Scott. 2007. Water quality in the River Clyde: a case study of additive and interaction models. Environmentrics 18:527-539.

Moran, P.A.P. 1950. Notes on continuous stochastic phenomena. Biometrika. 37:17–23.

Moyle, P.B. 2002. Inland Fishes of California. University of California Press.

Nishida, T. and D. Chen. 2004. Incorporating spatial autocorrelation into the general linear model with an application to the yellowfin tuna (*Thunnus albacares*) longline CPUE data. Fisheries Research 70:265-274.

Orsi, J.J., and W.L. Mecum. 1996. Food limitation as the probable cause of a long-term decline in the abundance of Neomysis mercedis the opossum shrimp in the Sacramento-San Joaquin Estuary. In San Francisco Bay: The Ecosystem, ed. J.T. Hollibaugh, 375–401. San Francisco: Pacific Division American Association for the Advancement of Science.

Pan, W. 2001. Akaike's information criterion in generalized estimating equations. Biometrics 57:120–125.

Peel, D., M.V. Bravington, N. Kelly, S.N.Wood, and I. Knuckey. 2013. A model-based approach to designing a fishery-independent survey. Journal of Agricultural, Biological, and Environmental Statistics 18:1–21.

Pennington, M. 1983. Efficient estimators of abundance, for fish and plankton surveys. Biometrics 39:281-286.

Quinn, T.J. and R.B. Deriso. 1999. Quantitative Fish Dynamics. Oxford University Press.

Ricker, W.E. 1975. Computation and interpretation of biological statistics of fish populations. Bulletin of the Fisheries Research Board of Canada 191.

Searle, S., F. Speed, and G. Milliken. 1980. Population marginal means in the linear model: An alternative to least squares means. American Statistician 34:216-221.

Schielzeth, H. 2010. Simple means to improve the interpretability of regression coefficients. Methods in Ecology and Evolution 1:103-113.

Schoellhamer, D.H. 2011. Sudden clearing of estuarine waters upon crossing the threshold from transport to supply regulation of sediment transport as an erodible sediment pool is depleted: San Francisco Bay, 1999. Estuaries and Coasts 34:885-899.

Sommer, T., C. Armor, R. Baxter, R. Breuer, L. Brown, M. Chotkowski, S. Culberson, F. Feyrer, M. Gingras, B. Herbold, W. Kimmerer, A. Mueller-Solger, M. Nobriga, and K. Souza. 2007. The collapse of pelagic fishes in the upper San Francisco Estuary. Fisheries 32:270–277.

Stéfansson, G. 1996. Analysis of groundfish survey abundance data: combining the GLM and delta approaches. ICES Journal of Marine Science 53:577–588.

Stevens, D.E., and L.W. Miller. 1983. Effects of river flow on abundance of young Chinook salmon, American shad, longfin smelt, and delta smelt in the Sacramento-San Joaquin River system. North American Journal of Fisheries Management 3:425–437.

Wood, S.N. 2006. Generalized Additive Models. Chapman & Hall/CRC.

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

Zuur, A.F., A.A. Saveliev, and E.N. Ieno. 2012. Zero inflated models and generalized linear mixed models with R. Newburgh: Highland Statistics Ltd.

7. Tables

Table 1. Model set fitted to the delta smelt presence-absence and CPUE data for both the FMWT and SKT surveys.

Model	Covariates
M_1	Year, Temp, EC, Secchi, Tide, Hour, Depth
M_2	Year, Temp, EC, Secchi, Tide, Hour
M_3	Year, Temp, EC, Secchi, Tide
M_4	Year, Temp, EC, Secchi
M_5	Year, Temp, EC, Secchi, Hour
M_6	Year, Temp, EC, Secchi, Depth
M_7	Year, Temp, EC, Secchi, Tide, Depth
M_8	Year, Temp, EC, Secchi, Hour, Depth

Table 2. Three-fold cross-validation results for the most supported models fitted to the delta smelt presence-absence data from the FMWT survey. Bold indicates model selected for inference.

Model	Training Error				Testing Error			
	1	2	3	Avg.	1	2	3	Avg.
GLM, model M_6	0.108	0.112	0.109	**0.110**	0.112	0.105	0.110	**0.109**
GAM, model M_1	0.102	0.105	0.102	**0.103**	0.103	0.097	0.102	**0.101**
GEE, model M_6	0.110	0.113	0.110	**0.111**	0.112	0.106	0.110	**0.109**
GLMM, model M_6	0.111	0.114	0.111	**0.112**	0.113	0.107	0.112	**0.111**
Autocovariate, model M_6	0.101	0.104	0.103	**0.103**	0.106	0.099	0.013	**0.102**

Table 3. Three-fold cross-validation results for the most supported models fitted to the delta smelt CPUE data from the FMWT survey. Bold indicates model selected for inference.

Model	Training Error				Testing Error			
	1	2	3	Avg.	1	2	3	Avg.
GLM, model M_5	0.83	0.90	0.86	**0.86**	0.89	0.77	0.81	**0.83**
GAM, model M_8	0.76	0.83	0.80	**0.80**	0.82	0.71	0.75	**0.76**
GEE	n/a	n/a	n/a	**n/a**	n/a	n/a	n/a	**n/a**
GLMM, model M_4	0.75	0.85	0.79	**0.80**	0.89	0.67	0.75	**0.77**
GAMM, model M_5	0.84	0.88	0.86	**0.86**	0.90	0.81	0.85	**0.85**
SAR, model M_6	0.78	0.86	0.84	**0.83**	0.89	0.74	0.79	**0.81**
GLS, model M_4	0.84	0.91	0.87	**0.87**	0.89	0.78	0.84	**0.84**
GAM_{err}, model M_8	0.79	0.87	0.84	**0.83**	0.90	0.76	0.83	**0.83**

Table 4. Three-fold cross-validation results for the most supported models fitted to the delta smelt presence-absence data from the SKT survey. Bold indicates model selected for inference.

Model	Training Error				Testing Error			
	1	2	3	Avg.	1	2	3	Avg.
GLM, model M_1	0.16	0.15	0.16	**0.15**	0.15	0.16	0.15	**0.15**
GAM, model M_1	0.14	0.13	0.13	**0.13**	0.13	0.13	0.13	**0.13**
GEE, model M_1	0.15	0.16	0.15	**0.15**	0.15	0.14	0.16	**0.15**
GLMM, model M_3	0.16	0.17	0.16	**0.16**	0.17	0.14	0.17	**0.16**
Autocovariate, model M_7	0.14	0.14	0.13	**0.13**	0.13	0.13	0.15	**0.14**

Table 5. Two-fold cross-validation results for the most supported models fitted to the delta smelt CPUE data from the SKT survey. Bold indicates model selected for inference.

Model	Training error	Testing error
GLM, model M_6	1.36	1.62
GAM, model M_8	1.05	1.21
GEE, model M_6	1.53	1.92
GLMM, model M_4	1.07	1.09
GAMM, model M_4	1.54	1.96
SAR, model M_6	1.23	1.33
GLS, model M_6	1.38	1.64
GAM_{err}, model M_8	1.19	1.35

8. Figures

Fig. 1. Map of sampling locations (gray dots) for a single cruise of the: (A) the SKT survey and (B) the FMWT survey.

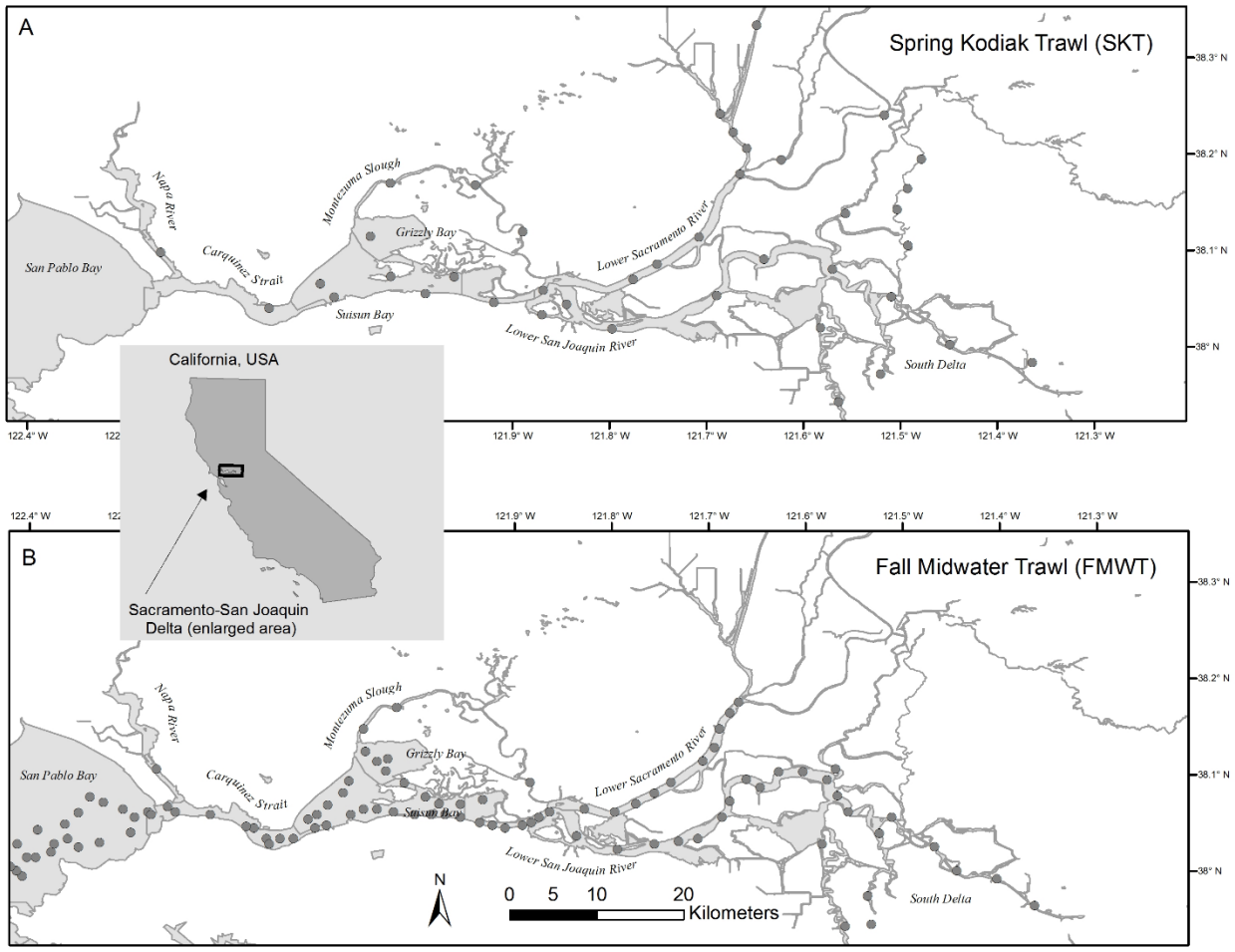

Fig. 2. FMWT catch summary by sampling Area

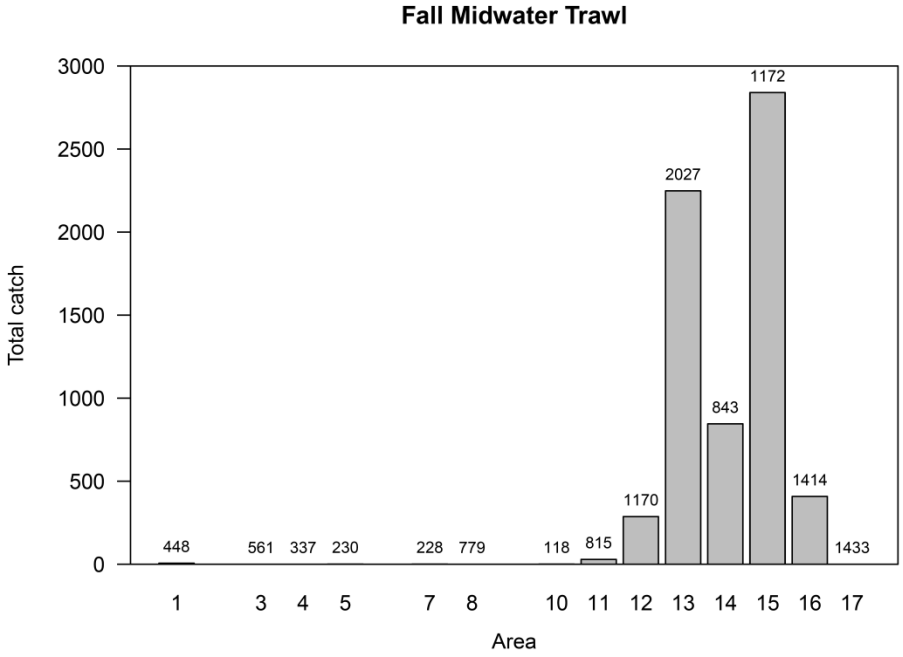

Fig. 3. Proportion of tows conducted annually by the FMWT and SKT where at least one delta smelt was captured. In general, the FMWT conducts approximately 400 tow per year while the SKT conducts about 40. Horizontal black lines represent time-series averages.

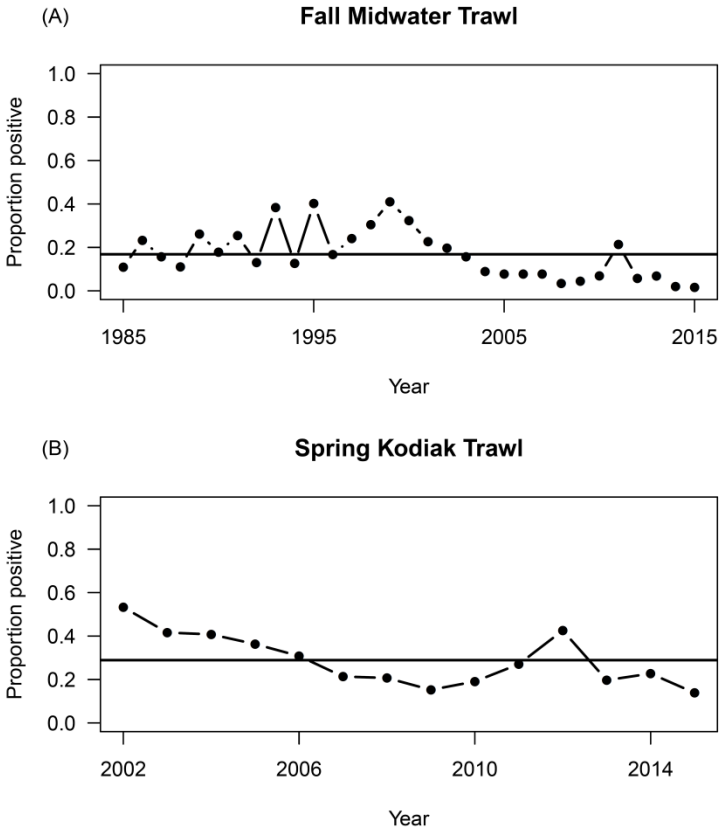

Fig. 4. Diagnostics plots associated with the baseline binomial GLM fitted to the FMWT delta smelt survey data: (A) denotes the binned residuals plot, (B) shows the partial autocorrelation function results of the raw presence-absence data, (C) shows the partial autocorrelation function of the GLM model residuals, and (D) displays the spatial correlograms for the raw presence-absence data (black line) and GLM residuals (gray line). For panels (B) and (C), height of histogram bars above the dotted lines indicates the presence of temporal autocorrelation, while in panel (D), departures from the zero line indicate spatial autocorrelation as a function of distance class. P/A refers to presence-absence.

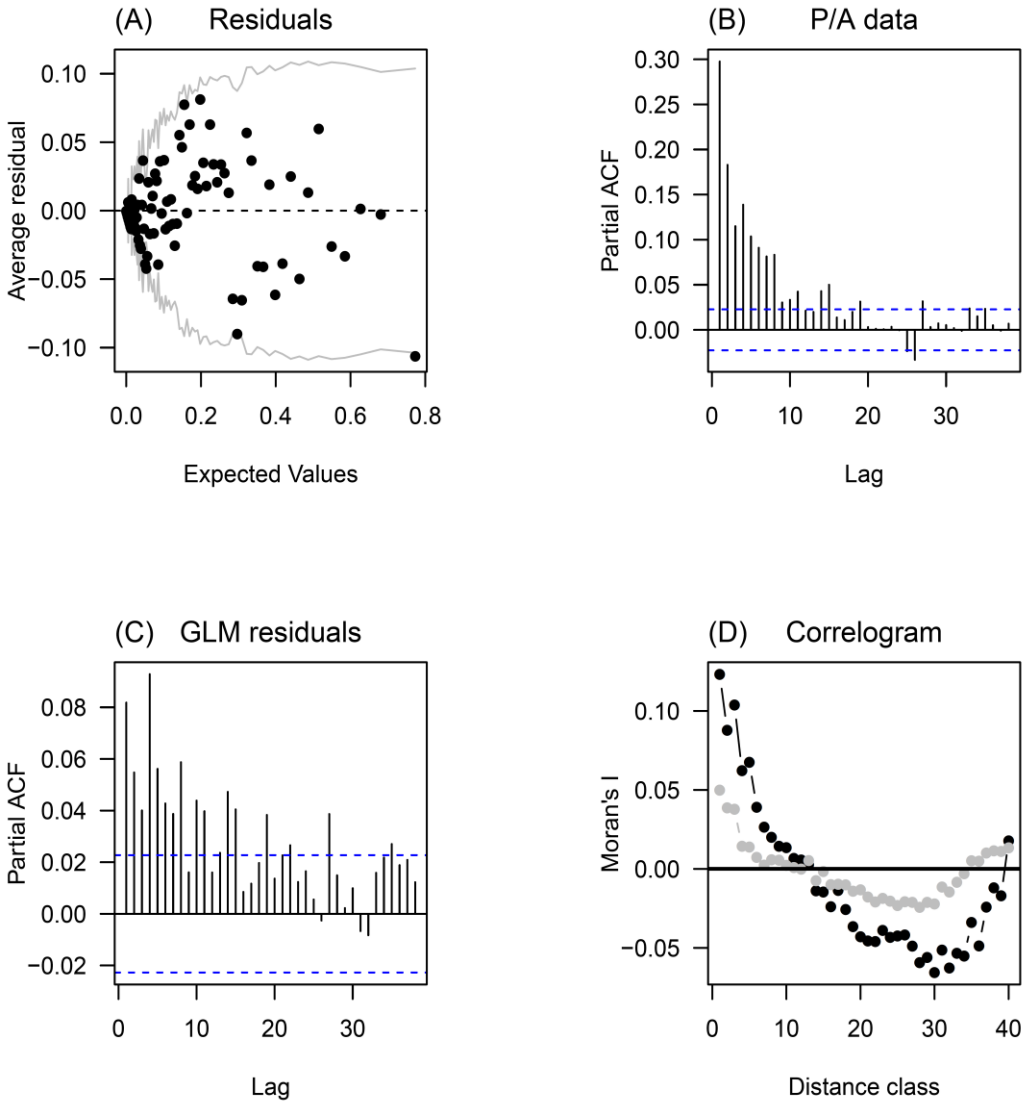

Fig. 5. Diagnostics plots associated with the baseline binomial GAM fitted to the FMWT delta smelt survey data: (A) denotes the binned residuals plot, (B) shows the partial autocorrelation function results of the raw presence-absence data (same as Fig. 4B, displayed for reference), (C) shows the partial autocorrelation function of the GAM model residuals, and (D) displays the spatial correlogram for the raw presence-absence data (black line; same as Fig. 4D, displayed for reference) and GAM residuals (gray line). For panels (B) and (C), height of histogram bars above the dotted lines indicates the presence of temporal autocorrelation, while in panel (D), departures from the zero line indicate spatial autocorrelation as a function of distance class. P/A refers to presence-absence.

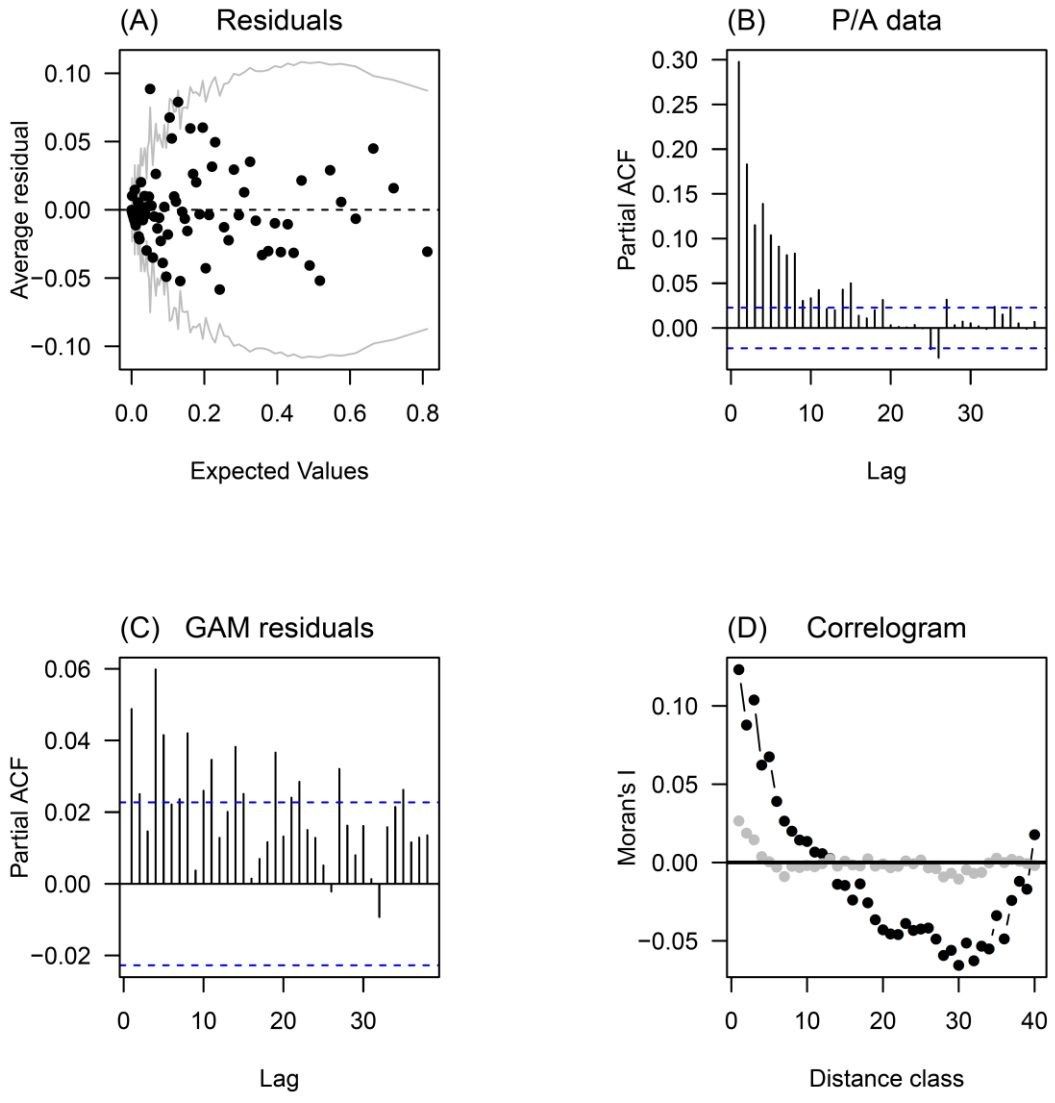

Fig. 6. Plot of annual number of trawl tows conducted (black line) and the annual estimated effective sample size (N_{eff}; gray line) derived from the most supported binomial GLMM fitted to the delta smelt FMWT survey data. Missing years are where there was model convergence issues.

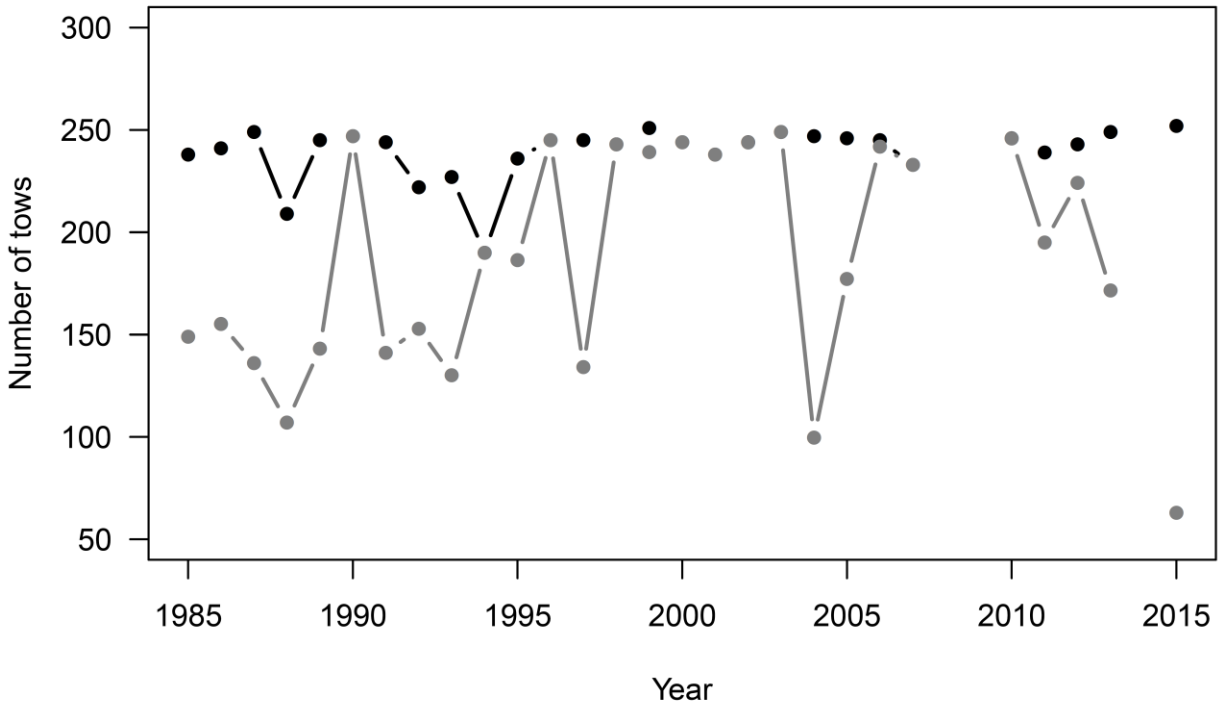

Fig. 7. Predicted capture probabilities (black lines) and associated estimated coefficients of variation (gray lines) of delta smelt for the FMWT survey over: (A) year, (B) temperature, (C) specific conductance, (D) Secchi depth, and (E) overall water depth. Predictions based on the most supported spatial AUTO model.

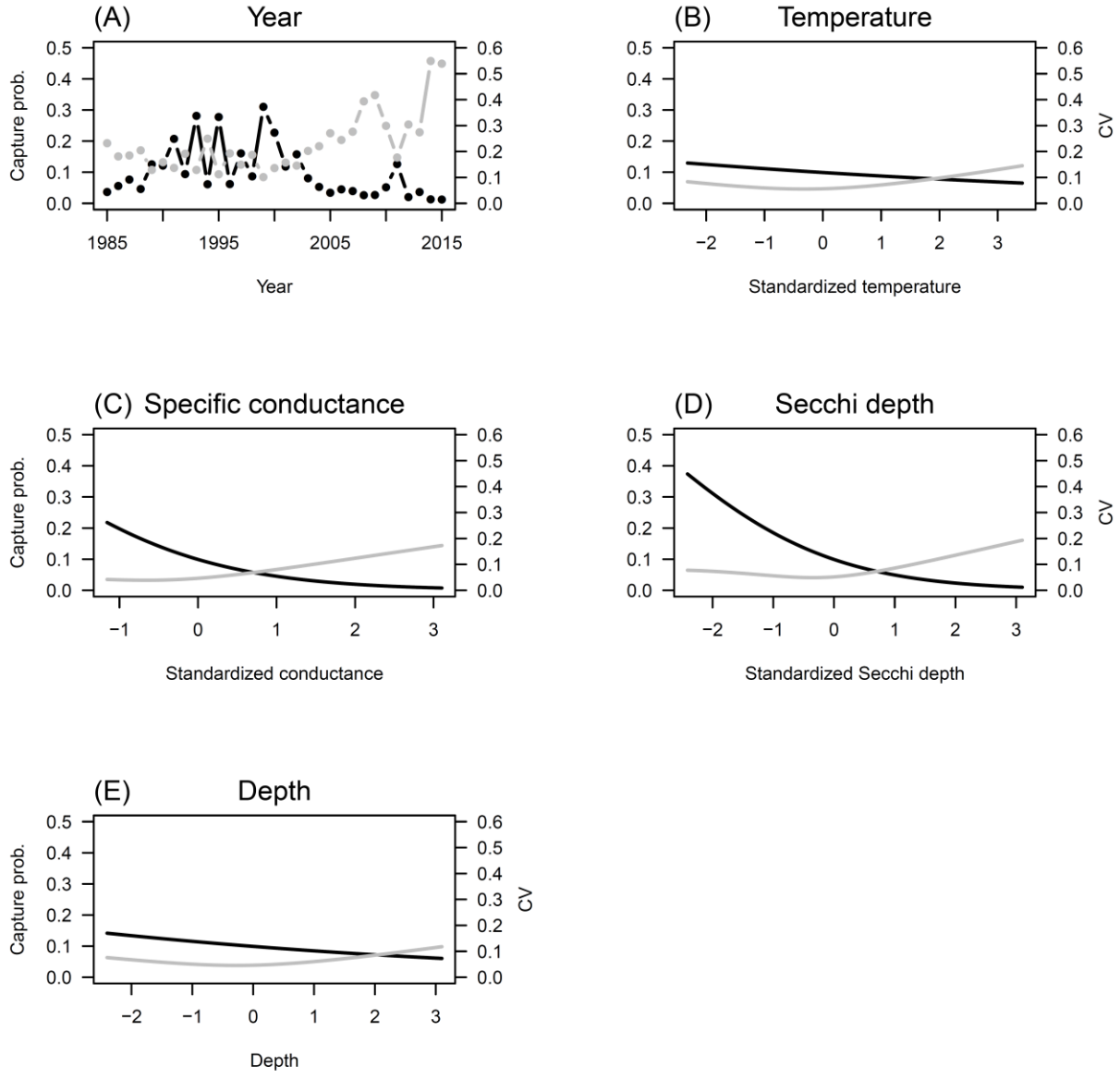

Fig. 8. Diagnostics plots associated with the baseline lognormal GLM fitted to the FMWT delta smelt CPUE survey data: (A) denotes the plot of residuals in relation to fitted values, (B) shows the partial autocorrelation function results of the log of the raw CPUE data, (C) shows the partial autocorrelation function of the GLM model residuals, and (D) displays the spatial correlogram for the log of the raw CPUE data (black line) and GLM residuals (gray line). For panels (B) and (C), height of histogram bars above the dotted lines indicates the presence of temporal autocorrelation, while in panel (D), departures from the zero line indicate spatial autocorrelation as a function of distance class.

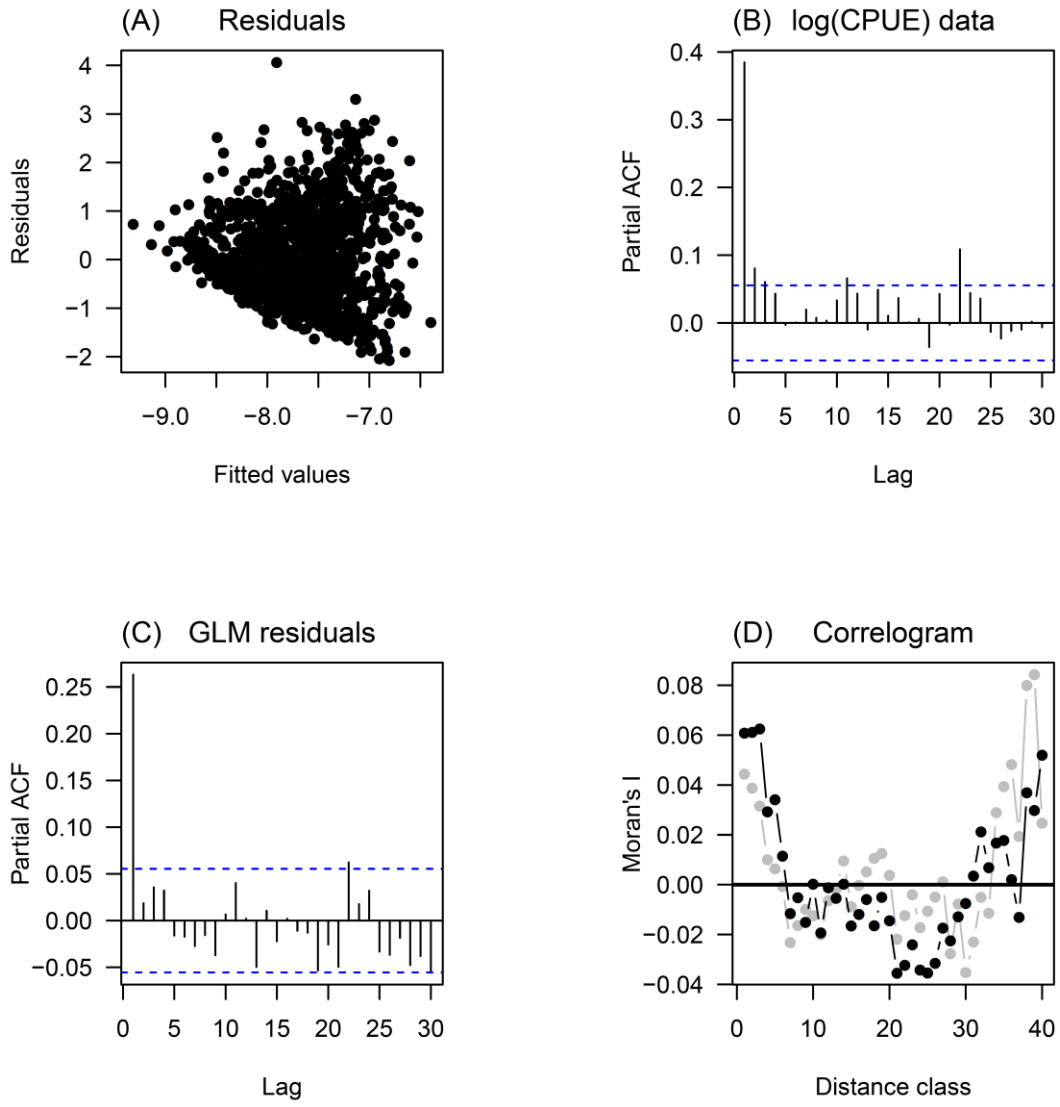

Fig. 9. Diagnostics plots associated with the baseline lognormal GAM fitted to the FMWT delta smelt CPUE survey data: (A) denotes the plot of residuals in relation to fitted values, (B) shows the partial autocorrelation function results of the log of the raw CPUE data (same as Fig. 8B, displayed for reference), (C) shows the partial autocorrelation function of the GAM model residuals, and (D) displays the spatial correlogram for the log of the raw CPUE data (black line; same as Fig. 8D, displayed for reference) and GAM residuals (gray line). For panels (B) and (C), height of histogram bars above the dotted lines indicates the presence of temporal autocorrelation, while in panel (D), departures from the zero line indicate spatial autocorrelation as a function of distance class.

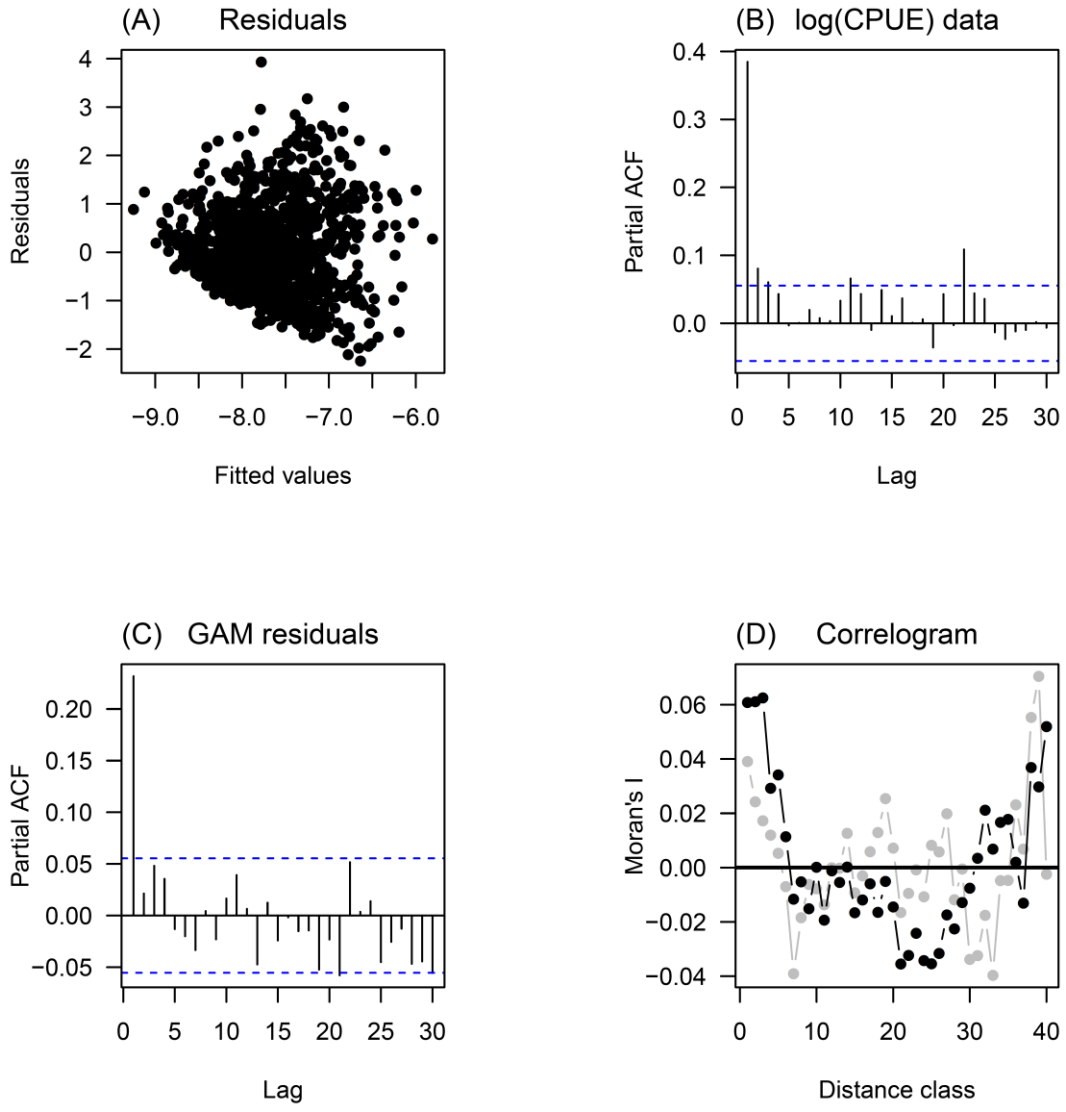

Fig. 10. Predicted CPUE (black lines) and associated estimated coefficients of variation (gray lines) of delta smelt for the FMWT survey over: (A) year, (B) temperature, (C) specific conductance, (D) Secchi depth, and (E) Hour of sampling. Predictions based on the most supported heterogeneous variance GAMM model.

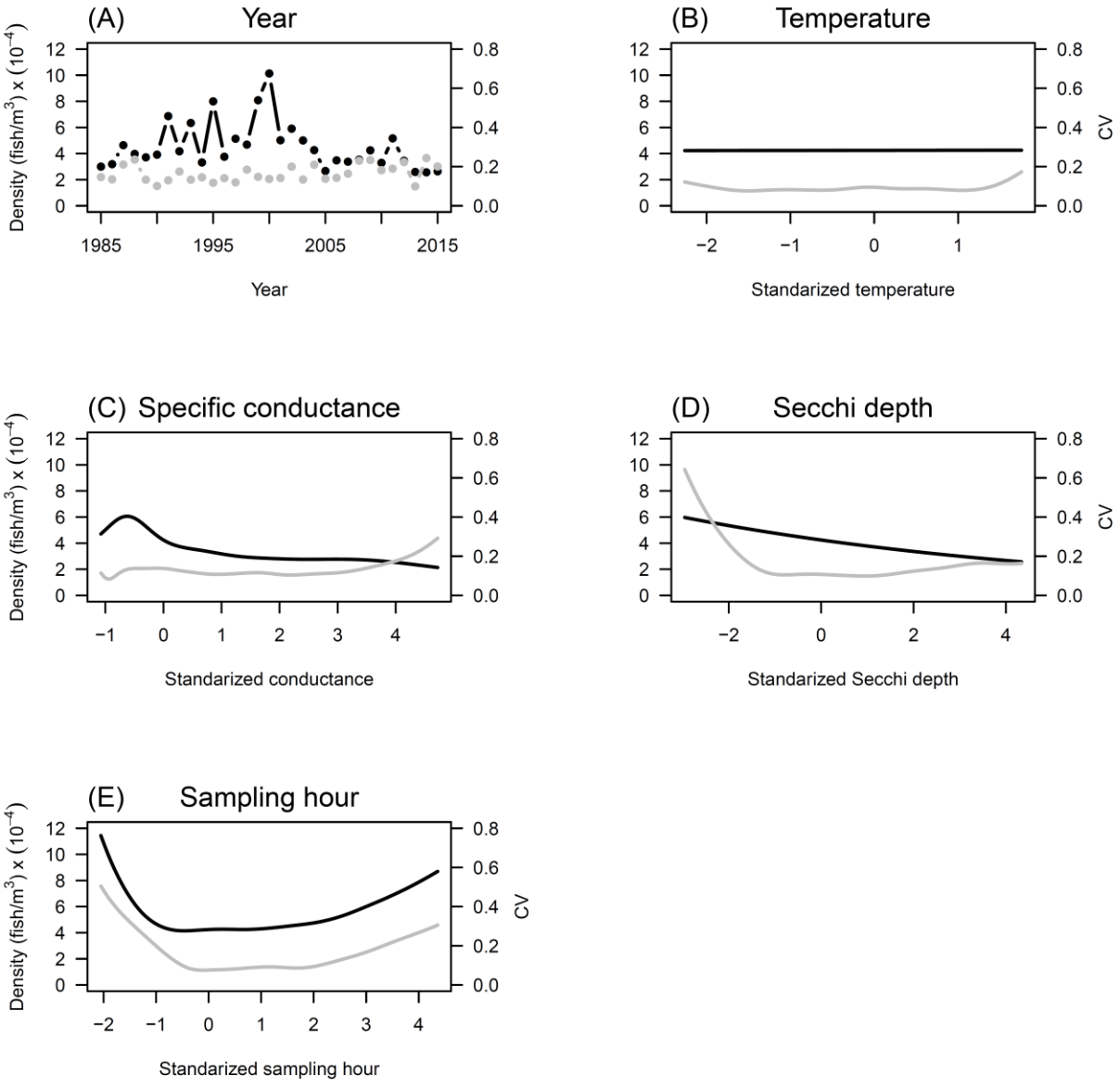

Fig. 11. Combined predictions (product of binomial and lognormal model components; black lines) and associated estimated coefficients of variation (gray lines) of delta smelt for the FMWT survey over: (A) year, (B) temperature, (C) specific conductance, and (D) Secchi depth.

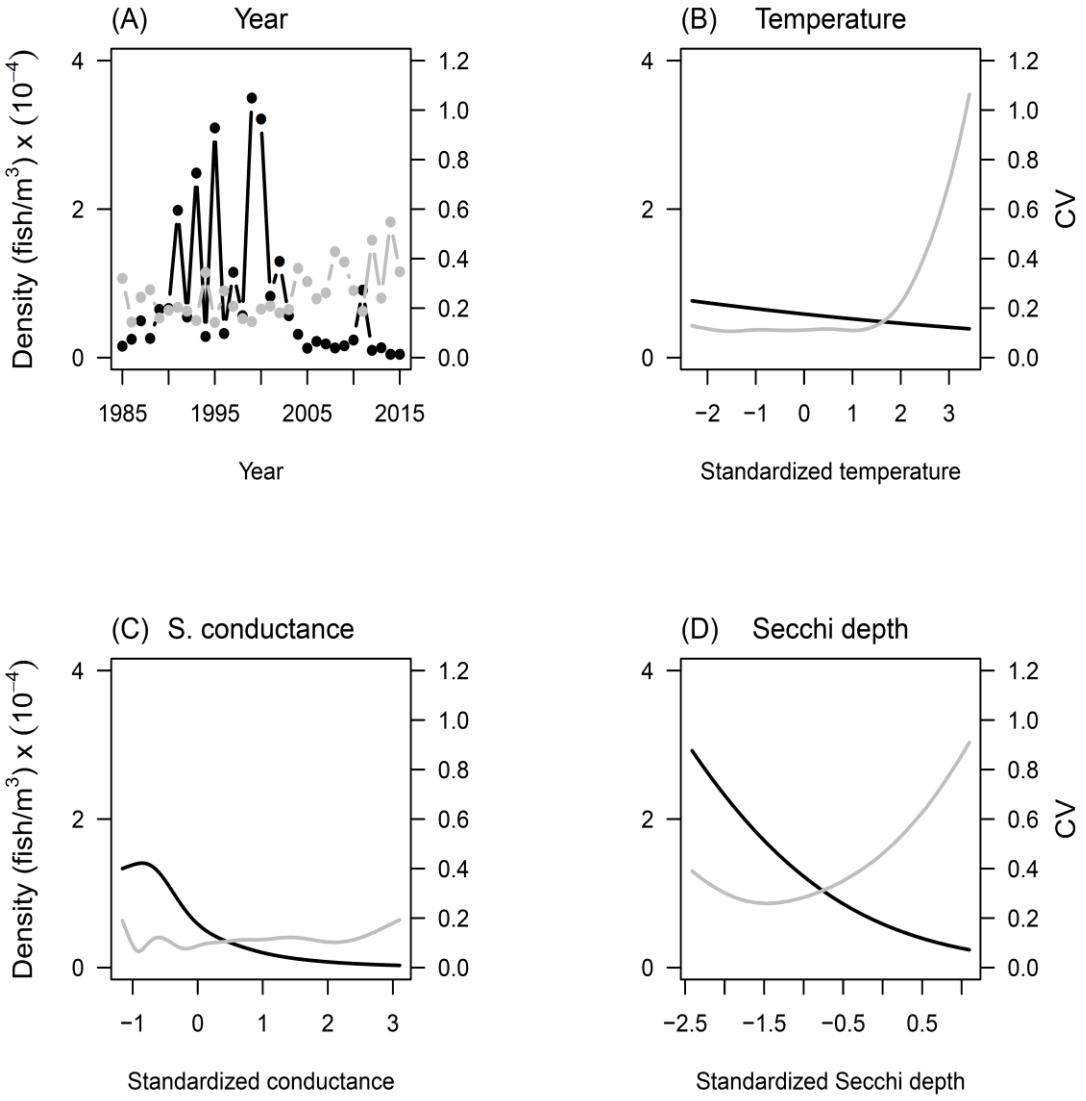

Fig. 12. Diagnostics plots associated with the baseline binomial GLM fitted to the SKT delta smelt survey data: (A) denotes the binned residuals plot, (B) shows the partial autocorrelation function results of the raw presence-absence data, (C) shows the partial autocorrelation function of the GLM model residuals, and (D) displays the spatial correlograms for the raw presence-absence data (black line) and GLM residuals (gray line). For panels (B) and (C), height of histogram bars above the dotted lines indicates the presence of temporal autocorrelation, while in panel (D), departures from the zero line indicate spatial autocorrelation as a function of distance class. P/A refers to presence-absence.

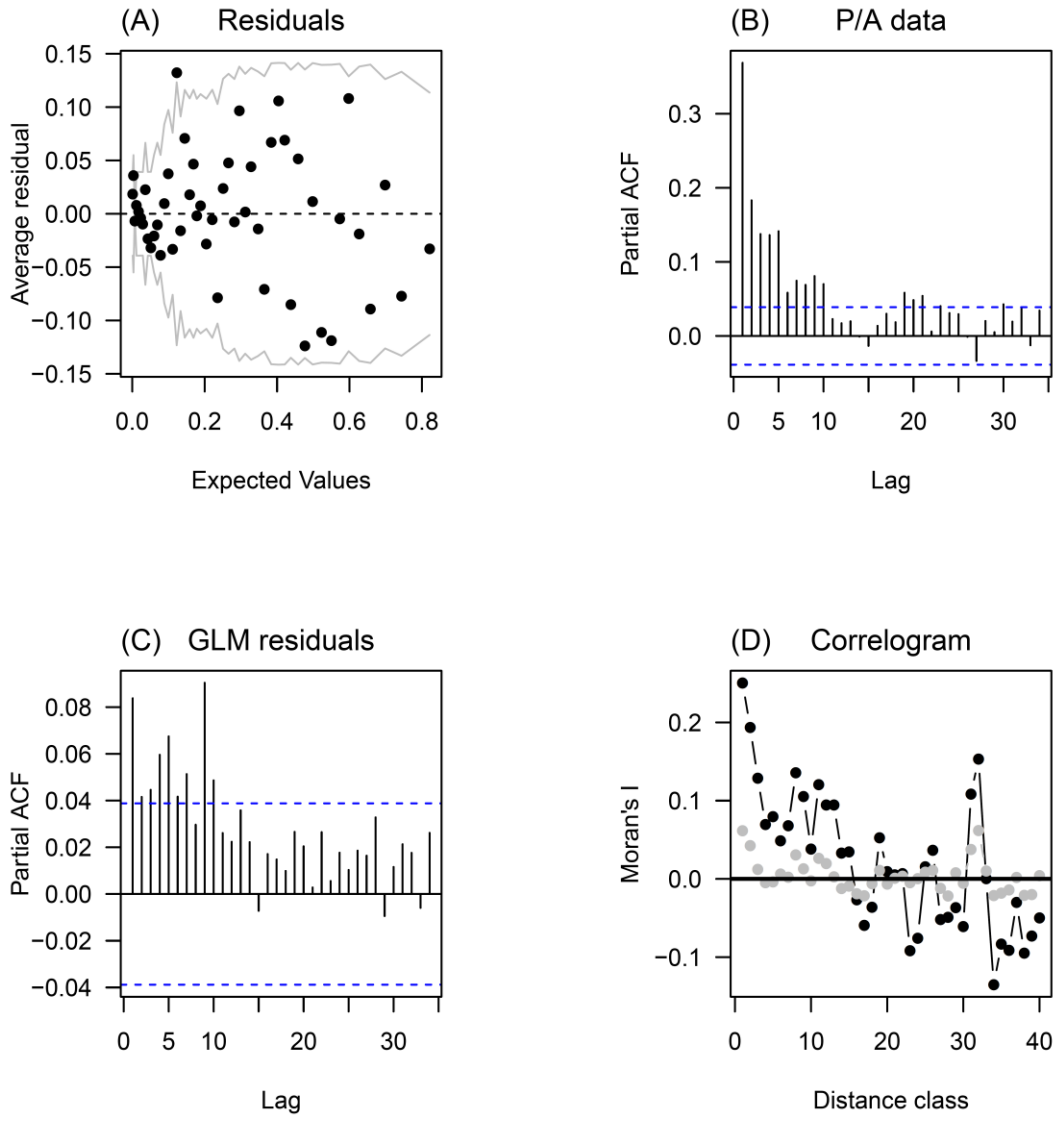

Fig. 13. Diagnostics plots associated with the baseline binomial GAM fitted to the SKT delta smelt survey data: (A) denotes the binned residuals plot, (B) shows the partial autocorrelation function results of the raw presence-absence data (same as Fig. 12B, displayed for reference), (C) shows the partial autocorrelation function of the GAM model residuals, and (D) displays the spatial correlogram for the raw presence-absence data (black line; same as Fig. 12D, displayed for reference) and GAM residuals (gray line). For panels (B) and (C), height of histogram bars above the dotted lines indicates the presence of temporal autocorrelation, while in panel (D), departures from the zero line indicate spatial autocorrelation as a function of distance class. P/A refers to presence-absence.

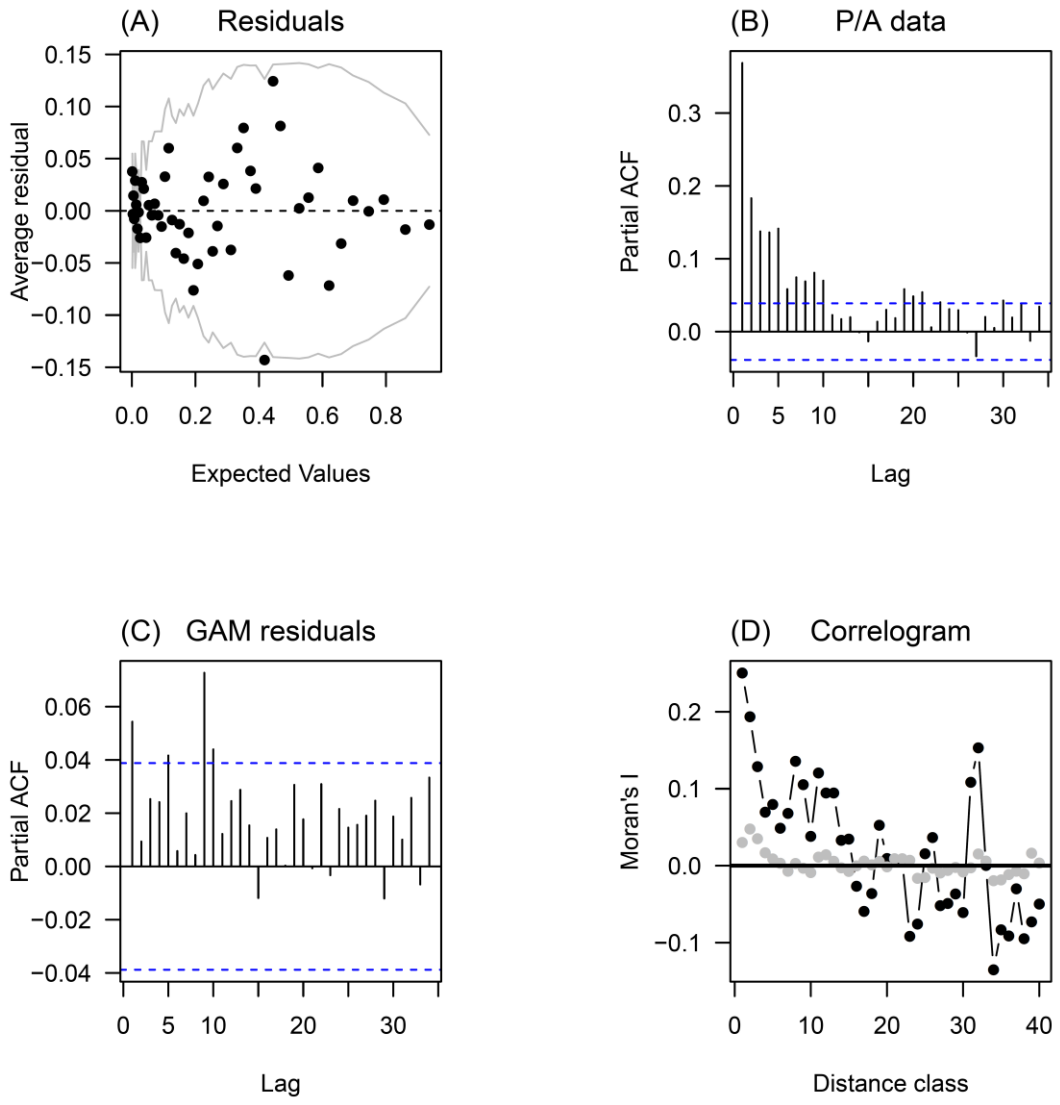

Fig. 14. Plot of annual number of trawl tows conducted (black line) and the annual estimated effective sample size (N_{eff}; gray line) derived from the most supported binomial GLMM fitted to the delta smelt SKT survey data.

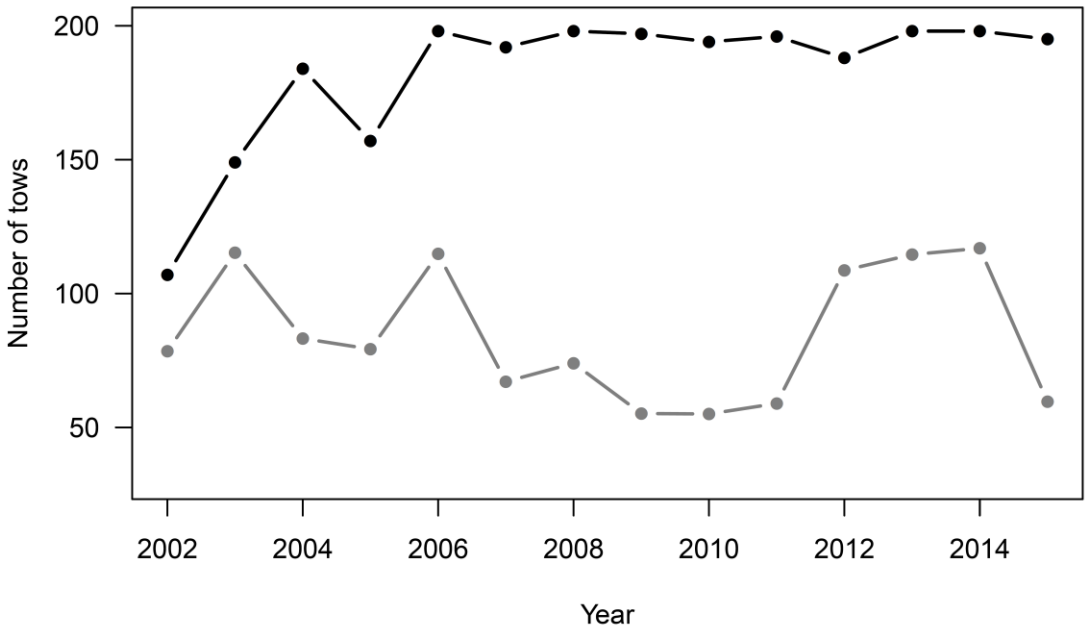

Fig. 15. Predicted capture probabilities (black lines) and associated estimated coefficients of variation (gray lines) of delta smelt for the SKT survey over: (A) year, (B) temperature, (C) specific conductance, (D) Secchi depth, and (E) tidal cycle. Predictions based on the most supported temporal GLMM model.

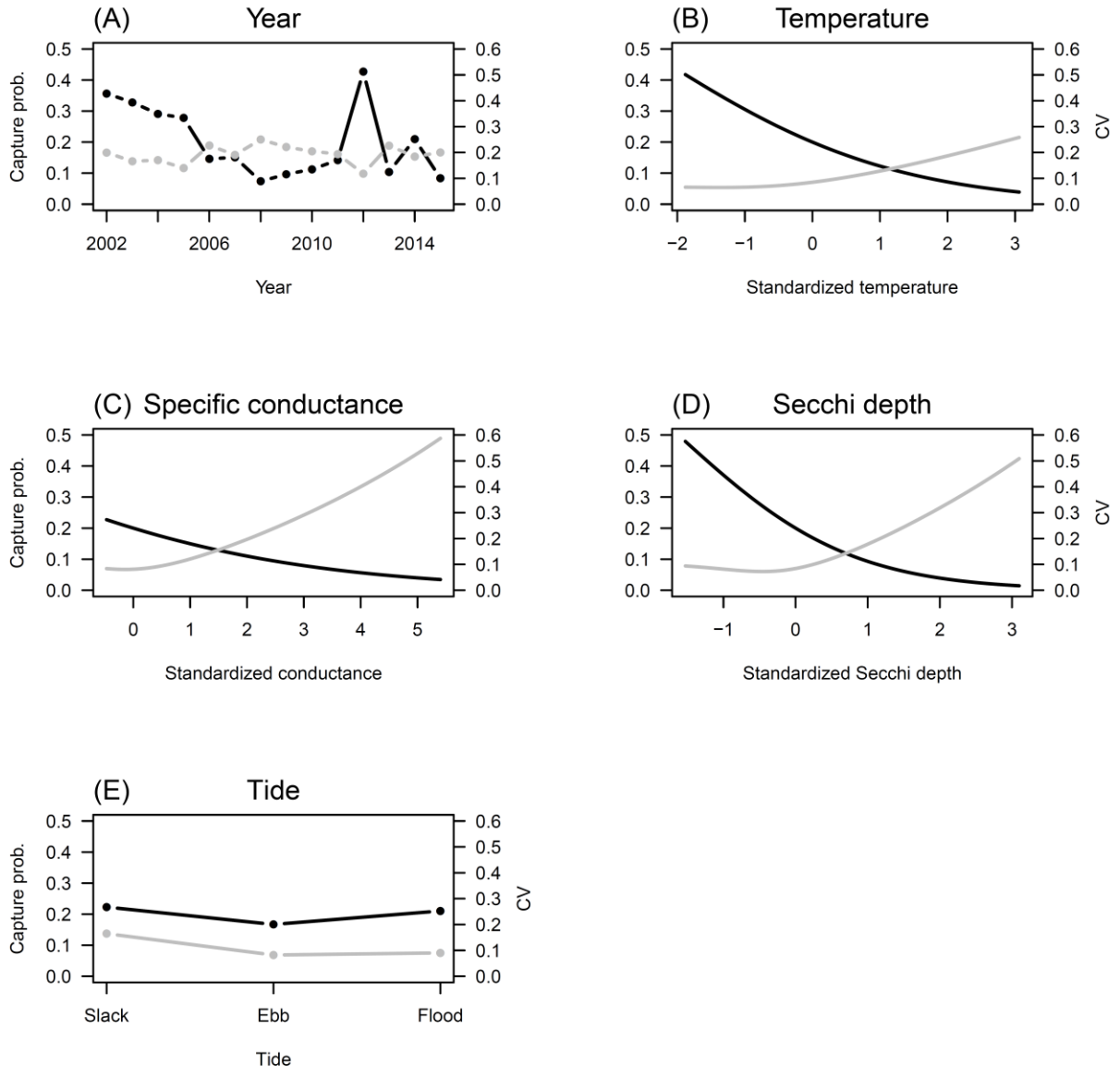

Fig. 16. Diagnostics plots associated with the baseline lognormal GLM fitted to the SKT delta smelt CPUE survey data: (A) denotes the plot of residuals in relation to fitted values, (B) shows the partial autocorrelation function results of the log of the raw CPUE data, (C) shows the partial autocorrelation function of the GLM model residuals, and (D) displays the spatial correlogram for the log of the raw CPUE data (black line) and GLM residuals (gray line). For panels (B) and (C), height of histogram bars above the dotted lines indicates the presence of temporal autocorrelation, while in panel (D), departures from the zero line indicate spatial autocorrelation as a function of distance class.

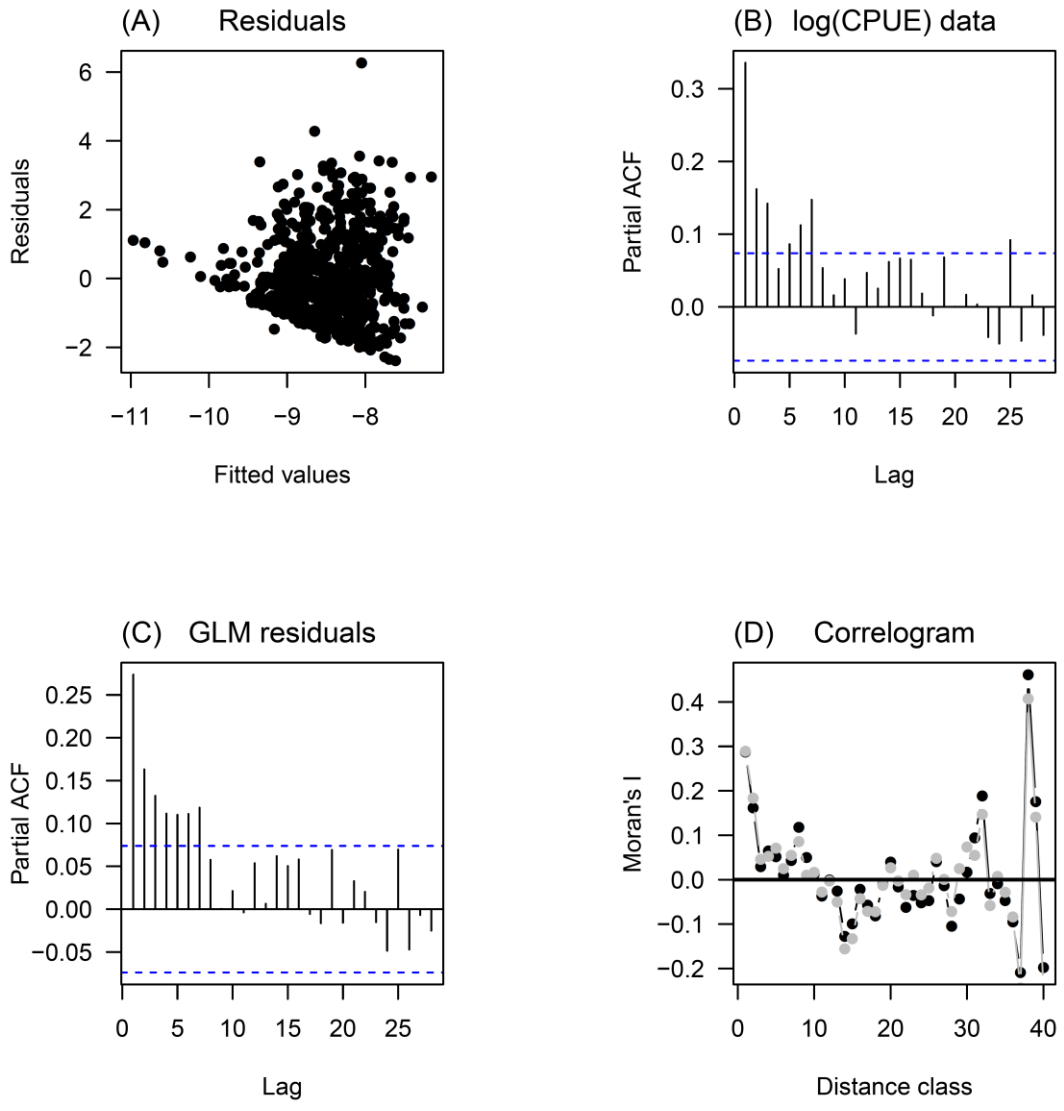

Fig. 17. Diagnostics plots associated with the baseline lognormal GAM fitted to the SKT delta smelt CPUE survey data: (A) denotes the plot of residuals in relation to fitted values, (B) shows the partial autocorrelation function results of the log of the raw CPUE data (same as Fig. 12B, displayed for reference), (C) shows the partial autocorrelation function of the GAM model residuals, and (D) displays the spatial correlogram for the log of the raw CPUE data (black line; same as Fig. 12D, displayed for reference) and GAM residuals (gray line). For panels (B) and (C), height of histogram bars above the dotted lines indicates the presence of temporal autocorrelation, while in panel (D), departures from the zero line indicate spatial autocorrelation as a function of distance class.

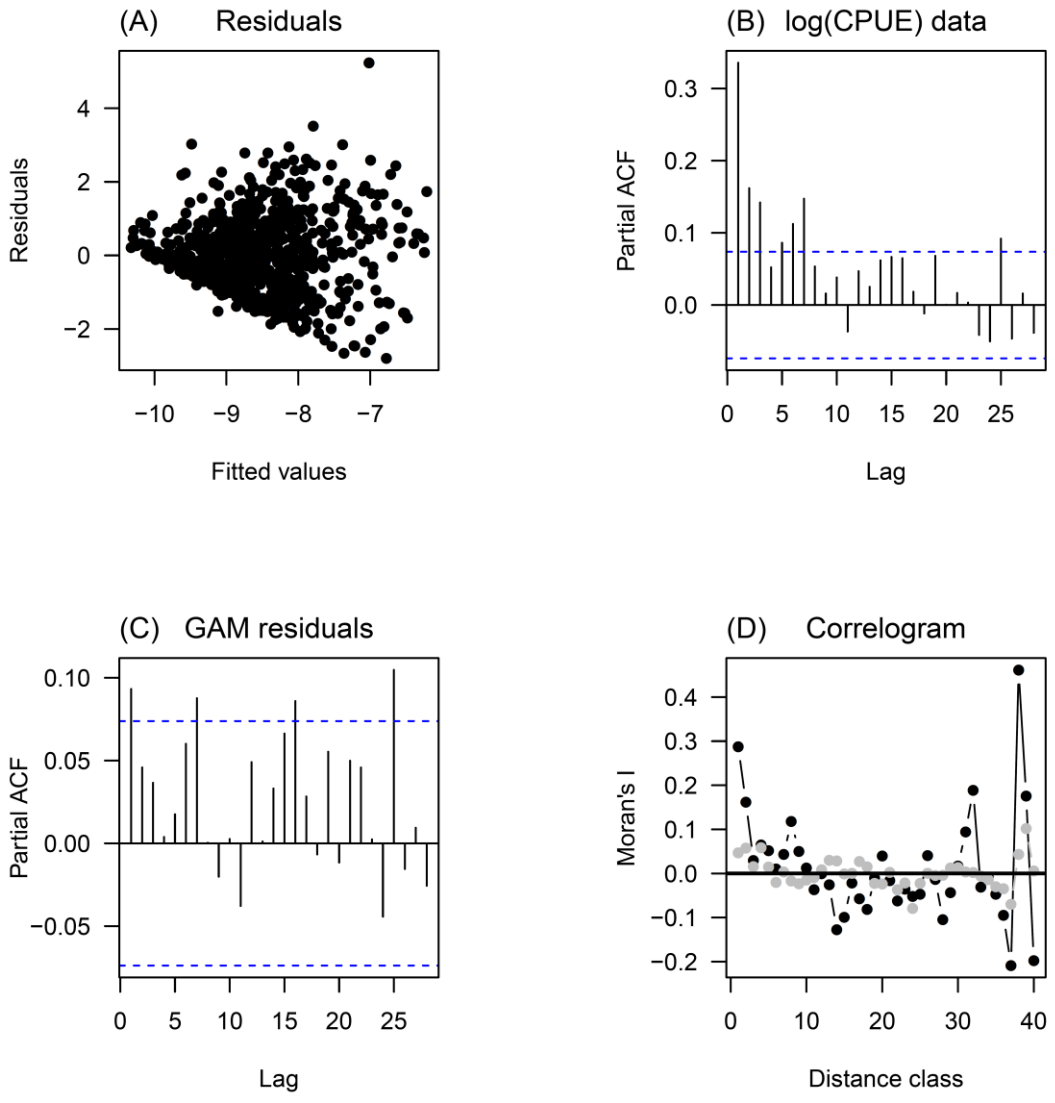

Fig. 18. Predicted CPUE (black lines) and associated estimated coefficients of variation (gray lines) of delta smelt for the SKT survey over: (A) year, (B) temperature, (C) specific conductance, and (D) Secchi depth. Predictions based on the most supported heterogeneous GLMM model.

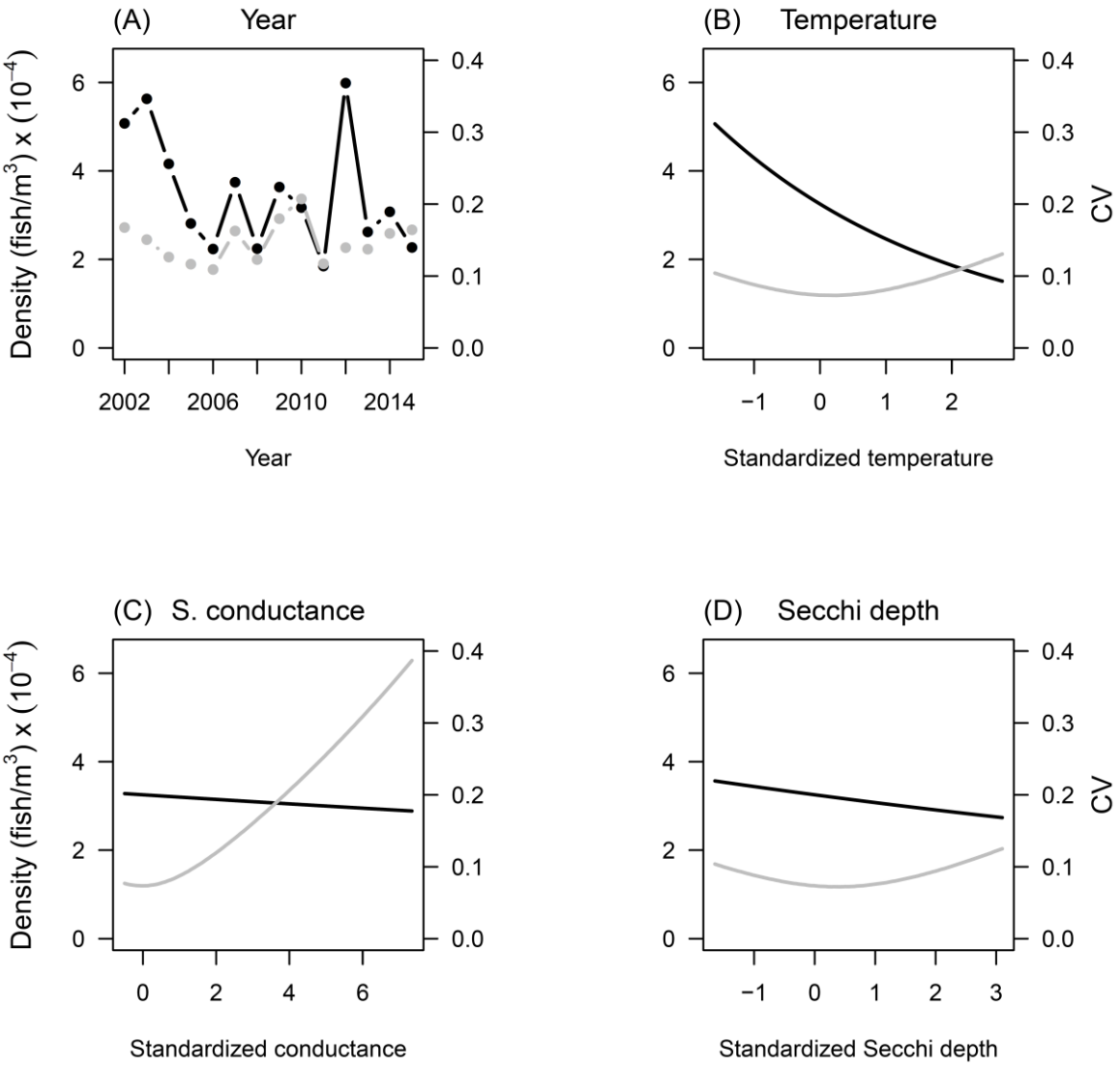

Fig. 19. Combined predictions (product of binomial and lognormal model components; black lines) and associated estimated coefficients of variation (gray lines) of delta smelt for the SKT survey over: (A) year, (B) temperature, (C) specific conductance, and (D) Secchi depth.