# Edinburgh Research Explorer

# A Bayesian hierarchical model of postlarval delta smelt entrainment: integrating transport, length composition, and sampling efficiency in estimates of loss

# A Bayesian hierarchical model of postlarval delta smelt entrainment: integrating transport, length composition, and sampling efficiency in estimates of loss

SCHOLARONE™
Manuscripts

1    A Bayesian hierarchical model of postlarval delta smelt entrainment: integrating transport, length

2    composition, and sampling efficiency in estimates of loss

3

4

5    William E. Smith[1*], Ken B. Newman[2,3], and Lara Mitchell[2]

6

7

8

9

10    [1]United States Fish and Wildlife Service, Bay-Delta Office, Sacramento, CA, USA

11    [2]United States Fish and Wildlife Service, Lodi Office, Lodi, CA, USA

12    [3]current address: Biomathematics and Statistics of Scotland and School of Mathematics, University of

13    Edinburgh Scotland, UK

14

15    *corresponding author

16    email: wes2316@gmail.com

17    telephone: (915)930-5666

18

19    **Abstract**

20

21    Hydrodynamic models have been used to estimate rates of ichthyoplankton transport across marine and

22    estuarine environments and subsequent geographic isolation of a portion of the population, i.e.,

23    entrainment.  Combining simulated data from hydrodynamic models with data from fish populations can

24    provide more information, including estimates of regional abundance.  Entrainment of postlarval delta

25    smelt, a threatened species endemic to California's Sacramento-San Joaquin Delta, caused by water

26    export operations, was modeled using a Bayesian hierarchical model. The model was fit using data

27    spanning years 1995-2015 from multiple sources: hydrodynamic particle tracking, fish length

28    composition, mark-recapture, and count data from entrainment monitoring.  Estimates of the entrainment

29    of postlarval delta smelt ranged from 10 (SD=23) in May 2006 to 561,791 (SD=246,423) in May 2002. A

30    simulation study indicated that all model parameters were estimable, but errors in transport data led to

31    biased estimates of entrainment. Using only single data sources rather than integration through

32    hierarchical modeling would have underestimated uncertainty in entrainment estimates or resulted in bias

33    if transport, survival, or sampling efficiency were unaccounted.

34    **Introduction**

35

36    Entrainment is a physical process whereby groups are advected and spatially redistributed through the

37    action of moving water, resulting in geographic isolation of the entrained portion of the population.

38    Hydrodynamic variation can have profound effects on entrainment of ichthyoplankton, and models that

39    do not account for these hydrodynamics may result in severely biased estimates of entrainment (Blumberg

40    et al. 2004; White et al. 2010). Water movement may be natural (Lough and Manning 2001) or the result

41    of water diversion for human use, such as power generation (Kelso and Milburn 1979) or water extraction

42    (Grimaldo et al. 2009). While natural entrainment plays a role in early dispersal and recruitment in marine

43    and estuarine systems, the typical fate of organisms entrained by water diversions is mortality or removal

44    and isolation from the population. If population losses are substantial, successful management and

45    conservation of populations entrained by water diversions requires an estimation of losses from the

46    population due to the magnitude and variation of entrainment over time (Dixon et al. 2003). For example,

47    the entrainment of fishes into California's South Sacramento-San Joaquin Delta (Fig. 1a) during the

48    process of water extraction and the subsequent mortality of entrained fish has been implicated in the

49    decline of native Sacramento-San Joaquin fish species (USFWS 2008; NMFS 2009). Management of

50    these population requires periodic assessment of past variation in entrainment and the quality of existing

51    biological and environment thresholds used to management the population and to mitigate entrainment

52    effects.

53

54    Hydrodynamic models coupled with particle tracking models have been used to simulate egg and larval

55    advection or entrainment through marine (White et al. 2010) and estuarine environments (North et al.

56    2008; Blumberg et al. 2004). While many studies have estimated rates of entrainment (Boreman et al.

57    1981) or explored behaviors that modulate entrainment (Culberson et al. 2004), further information can be

58    leveraged by combining transport data with other routinely collected forms of data. For example,

59    population monitoring surveys commonly collect age, length, and abundance information, and mark-

60    recapture studies often yield information about survival and sampling efficiency. However, such data are

61    rarely integrated with transport data in models of entrainment.

62

63    Integrated data analysis is a useful technique for combining information from multiple sources to allow

64    estimation of parameters not identifiable from a single data source (Besbeas et al. 2002), and such

65    analyses are commonly used to model fish population dynamics and assess the status of fish stocks

66    (Methot and Wetzel 2013). Integration of particle tracking data with a population dynamics model of

67    abundance, however, is uncommon. Most studies using particle tracking data to model entrainment do not

68    attempt to estimate abundance but are designed to validate hypothesized interactions between source

69    location, advection, and behavior (Miller 2007); thus, particle tracking data are compared to monitoring

70    data in correlative analyses rather than integrated analyses within a single model (Baumann et al. 2006;

71    Hinckley et al. 2016). Hierarchical modeling is one approach to integrate disparate data types, through use

72    of a multi-level structure (Rochette et al. 2013). Importantly, a hierarchy can facilitate greater statistical

73    power through reduced parameterization and partitioning of errors. A hierarchical approach to modeling

74    abundance of entrained populations can combine independent forms of data in order to model fish

75    population and observation dynamics, such as transport probabilities, survival, length or age composition,

76    and sampling efficiency.

77

78    Previous applications of particle tracking data have generally sought to model spatial structure in two or

79    three dimensions, because the initial distribution, or origin, of fish is a critical factor (Heimbuch et al.

80    2007; Huret et al. 2007) or important hydrodynamic variation occurs in the lateral or vertical dimensions.

81    For example, White et al. (2010) simulated three-dimensional transport of ichthyoplankton in an open

82    coastal system, where water was diverted for power plant cooling. Hydrodynamic complexity in a large

83    open systems creates important 3-dimensional variation in transport, but the volumes of diverted water

84    are very small relative to the volume of the system, and the effect of water diversion on system

85    hydrodynamics is minimal. Conversely, water extraction from smaller riverine systems for irrigation may

86    represent a much higher volume relative to the system and have a greater impact on hydrodynamics,

87    potentially pulling ichthyoplankton along a simplified linear route. For instance, as much as 35% of

88    inflow to the Sacramento-San Joaquin Delta is exported during April–June (CSWR 1999), and  exports

89    rates are often sufficient to reverse the net flows in the two rivers conveying water to the pumps, leading

90    to entrainment events. If the probability of transport from a region of low escapement to sampling

91    locations, e.g., the water diversions, is known, then a simplified 1-dimensional transport model may

92    sufficiently describe net transport without the burden of higher dimensionality. In other words, a model of

93    starting distributions and lateral and vertical dimensions may be unnecessary when transport routes are

94    constrained. Assuming that all fish transported into an entrainment zone are removed from the population,

95    counts at diversions can be used to estimate the number that passed through some downstream region and

96    became entrained. Generally, the abundance of any population being transported or entrained to a set of

97    discrete locations may be estimated from the number arriving at one entrainment location divided by the

98    probability of transport to that location. Alternately, replicated counts of the number arriving at one

99    location can be used to fit a statistical model estimating the number entrained and standard errors. This is,

100   in essence, the N-mixture model, where abundance is estimated from replicate counts and some model of

101   encounter probability (Royle 2004).

102

103   Our primary objective was to estimate the number of postlarval fish that entered an area of high

104   entrainment risk and were retained by hydrodynamic conditions. We take advantage of a linear

105   entrainment routing assumption to estimate regional abundance, where regional abundance is considered

106   to reflect eventual entrainment and geographic isolation of a portion of the population. The reduced

107   spatial dimensions of the model avoid the general problems of modeling a starting spatial distribution and

108   transport over longer times and distances. Integration of information about transport, length structure,

109   sampling efficiency, and abundance was achieved using a Bayesian hierarchical model. The hierarchical

110   model structure facilitated a mechanistic description of the sequence of events leading from entrainment

111   to observation, accounting for population dynamic and observation processes. The general purpose of the

112    model was to explore temporal variation in the number of fish entrained, after accounting for dynamic

113    rates of transport, survival, sampling efficiency and subsampling, and models of each dynamic rate were

114    sufficiently flexible to test covariate effects. The model was applied to estimate regional postlarval

115    abundance of an endangered species, the delta smelt (*Hypomesus transpacificus*), that are entrained

116    during water extraction from the Sacramento-San Joaquin Delta. Water extractions supply more than 25

117    million people with drinking water (CDWR 2011) and support a $45 billion agricultural industry in

118    California (DFA 2016).

119

120    **Materials and Methods**

121    The model described below estimated the number of postlarval delta smelt that passed into a region of the

122    South Sacramento-San Joaquin Delta (South Delta) where postlarvae are unlikely to survive, minus the

123    fraction likely to escape entrainment by being transported downstream to a zone of low entrainment risk

124    (the Low Risk Zone; LRZ), based on hydrodynamic conditions. The sequence of events leading from

125    initial entrainment into the South Delta to observation was modeled as a hierarchy of processes that were

126    informed by independent process-specific data sources. Data to estimate components of delta smelt

127    entrainment were noisy and imperfect measures of the processes they represented. Ongoing and proposed

128    research promises to improve data to estimate delta smelt processes in the future, so a flexible modeling

129    framework was developed to accommodate new information as it becomes available, different

130    assumptions, or completely different species and systems.

131

132    **Study system**

133    Delta smelt are small Osmerid fish endemic to the Sacramento-San Joaquin Delta. Following severe

134    declines in abundance, delta smelt were listed as threatened under the US Endangered Species Act in

135    1993 and as endangered under the California Endangered Species Act in 2009 (Bennett 2005; CDFW

136    2018). Entrainment of delta smelt led to issuance of Biological Opinions by the US Fish and Wildlife

137    Service in 2004, 2005, and 2008 that stipulated water operations to minimize entrainment effects

138  (USFWS 2008). Contentious debate (Kimmerer 2008 and 2011; Miller 2011) and litigation (2010 US

139  District Court ruling) ensued over implementation of the Biological Opinions and management of delta

140  smelt entrainment. Delta smelt entrainment or the population effect of entrainment has been explored by

141  many researchers (Kimmerer 2008; Grimaldo et al. 2009; Maunder and Deriso 2011; Miller et al. 2012;

142  Rose et al. 2013; Kimmerer and Rose 2018) using a variety of statistical methods.

143

144  The Sacramento-San Joaquin Delta forms at the confluence of the Sacramento and San Joaquin rivers

145  (Fig. 1A), draining the California Central Valley. The Delta is a network of freshwater tidal sloughs and

146  rivers that acts as a conveyance system for water exported by the Banks and Jones Pumping Plants,

147  located in the southern portion of the Delta (the South Delta) on Old River. Annual export volumes

148  ranged from 2.5 to 8.5 billion m³ between 1995 and 2014 (ftp://ftp.dfg.ca.gov). High water export rates

149  often reverse the net, or tidally filtered, flows in Old and Middle rivers, and the net upstream flows

150  entrain small fishes into the South Delta. The sum of net Old and Middle River flows (*OMR*) is

151  considered an index of the hydrodynamic forces entraining or drawing delta smelt through the Old and

152  Middle rivers upstream to the water diversions (USFWS 2008). Two fish facilities, operated by the

153  United States Bureau of Reclamation and the California Department of Water Resources, monitor

154  entrainment by screening and removing a fraction of entrained fish (Brown et al. 1996). Large louvers

155  divert fish from intake channels that direct water to the Projects' pumping plants.

156

157  Entrainment of delta smelt may be related to seasonal and ontogenetic variation in population distribution

158  and behavior, with different life stages becoming vulnerable to entrainment at different times of year and

159  possibly by different mechanisms. Delta smelt spawning occurs February through May, and postlarvae

160  recruit to the population April through June (Bennett 2005). December through March entrainment of

161  adults may be related to spawning, and some evidence suggests turbidity is related to individual delta

162  smelt behavior, with greater turbidity leading to higher occupancy of mid-channel habitats (Bennett and

163  Bureau 2015), where vulnerability to advective flow is greater. In adults, this behavior may lead to

164    movement upstream, where delta smelt become more vulnerable to entrainment. The east to west

165    distribution of the population in relation to advective flows into the entrainment region may be related to

166    the geographic position of the zone with 0-2 ppt salinity (the Low Salinity Zone) and critical delta smelt

167    habitat (Grimaldo et al. 2009). The Low Salinity Zone moves downstream during the winter, following

168    seasonal rainfall, and then back upstream during the summer as conditions dry. Larval and postlarval life

169    stages may become vulnerable to entrainment when their geographic distribution overlaps the area of

170    South Delta advective flow. The fate of postlarval delta smelt entrained into the South Delta is mortality

171    via direct export from the Delta (Castillo et al. 2012), predation by the robust non-native piscivore

172    community (Clark et al. 2009), thermally induced mortality during the summer (Swanson et al. 1998), or

173    reproductive isolation if fish or their offspring become entrapped by the net southward flows.

174

175    **Data**

176    ***Delta Simulation Model 2 particle tracking data***

177    Transport data to fit the model of delta smelt entrainment consisted of counts of particles that started in a

178    specific region, were transported by simulated hydrodynamics, and were found in a discrete set of regions

179    after a one month period. The particle transport data were used to provide information about the monthly

180    rate at which small fish passed through various regions in the Delta and then arrived at the fish facilities

181    of the State Water (SWP) and Central Valley Projects (CVP) (Table  A1). Transport was simulated using

182    California Department of Water Resources's one-dimensional coupled hydrodynamic and particle

183    tracking model, the Delta Simulation Model 2 (DSM2) (CDWR 2013). DSM2 calculates the transport of

184    neutrally buoyant particles in the network of channels comprising the Delta. Simulated transport accounts

185    for important dynamics such as river inflows, water exports, and tidal cycles. For a 21 year period (1995–

186    2015) 4,000 particles were released continuously over a 24-hour period on the first day of April, May, and

187    June into each of  three regions of the Delta (Fig. 1B-D). Within each region, particles were inserted at

188    four or five evenly spaced points, and after one month, the individual particle locations were recorded.

189

190    The final locations were spatially aggregated into regions we designated as sinks, and counts of particles

191    were summed for each region (Table A1-A3). Four sinks were designated, SWP, CVP, Low Risk Zone

192    (LRZ), and Indirect Risk Zone (IRZ). Specifically, we defined the intake channels leading to the export

193    pumps to be the sink locations for SWP and CVP. The LRZ and IRZ were dynamically defined

194    depending upon the source region. The LRZ included areas downstream, north and west of the source

195    region, while the IRZ included remaining regions in the South Delta, generally between the source region

196    and the pumps.

197

198    To provide data for a model linking counts at fish facilities to entrainment, for each year-month

199    combination, one of three areas was dynamically selected as an entrainment source region (Fig. 1B-D).

200    Collectively, source regions defined the pathway that entrained fish followed, moving upstream from the

201    San Joaquin River near the mouths of Old and Middle rivers to the SWP and CVP. Source regions

202    represented the downstream edge of a zone of entrainment, which varied as a function of inflow from the

203    Sacramento and San Joaquin rivers and the volume of water exported by the SWP and CVP pumps.  The

204    underlying conceptual model was that as inflow volume increased relative to export volume, the area of

205    the region affected by entrainment decreased (Grimaldo et al. 2009), and source regions further away

206    from the pumps were associated with larger areas affected by entrainment compared to source regions

207    near the pumps. The selection of a single source region for each year-month represented a larger or

208    smaller entrainment area, depending on hydrodynamic conditions. Particles (fish) that ended up within or

209    upstream (towards the pumps) of source regions were considered entrained into the South Delta region

210    which included the Old and Middle rivers, portions of the San Joaquin River, and areas adjacent to the

211    SWP and CVP. The three potential source regions, listed here in decreasing distance from SWP and CVP,

212    were: (1) Prisoners Point on the San Joaquin River, (2) Lower Old and Middle rivers, and (3) Upper Old

213    and Middle rivers (Fig. 1B–D). The source region selected for a given year-month combination was the

214    one where the proportion of particles entrained was closest to 0.50, thus the probability of not being

215  entrained (ending up in the LRZ) was also closest to 0.50.  This assumption could be altered to a more or

216  less risk averse decision about probability of entrainment.

217

218  The suitability of applying transport probabilities of particles to model delta smelt relied on three

219  assumptions:

220

221  1.  Postlarval delta smelt experience the same transport processes as passive particles in the DSM2

222     hydrodynamic model.

223  2.  Fish drawn into the South Delta are effectively removed from the Sacramento-San Joaquin Delta

224     population of delta smelt, and this removal defines entrainment.

225  3. Entrained fish (through the source region) follow a single linear route into the South Delta, from the

226     Lower San Joaquin River through the Old and Middle rivers to the SWP and CVP pumps, and are not

227     primarily routed through smaller sloughs.

228

229  ***Length composition data***

230  Delta smelt length frequency distributions by year and month were summarized from catch in the 20mm

231  Survey conducted by the California Department of Fish and Wildlife. The 20mm Survey is a monitoring

232  program designed to sample postlarval and juvenile Delta Smelt and covers their historical spring range

233  (CDFW 2019). The 20mm Survey length data from all observations in each month-year combination

234  were used to develop an estimate of the population length composition. Lengths considered in the analysis

235  were restricted to 20–45 mm fork length (i.e., snout to end of center caudal fin ray), where 20 mm is the

236  minimum size enumerated in samples at fish facilities, and 45 mm was the mean bioenergetics growth

237  model prediction (Fujiwara et al. 2005) for the oldest possible postlarvae born March 1. The upper limit

238  excluded larger age-1 fish. Lengths were binned in 5-mm size classes .

239

240  ***Counts and length samples from fish facilities***

241    Delta smelt and other fishes entrained by the SWP and CVP have been routinely monitored since 1979 at

242    the Skinner Fish Facility by the California Department of Water Resources and at the Tracy Fish Facility

243    by the United States Bureau of Reclamation, respectively. As mentioned previously, a fraction of the fish

244    directly entrained in water exports are screened by louvers that divert fish for sampling. These diverted

245    fish at fish facilities are termed salvage in the regional jargon. Since 1993 a subset of diverted fish have

246    been enumerated and measured for length during an approximately 30-minute interval every two hours

247    (Karp et al. 1997). Diverted fish include age-0 and age-1 fish, and fish lengths are used to assign fish ages

248    in the samples.

249

250    ***Mark-recapture data***

251    Delta smelt tagging experiments at the Skinner Fish Facility, which samples fish at the SWP, were

252    reported by Castillo et al. (2012). Six batches of adult delta smelt were tagged with visible implant

253    elastomer and released into the intake channel of Banks Pumping Plant; a fraction was recovered by the

254    Skinner Fish Facility, and the remainder were predated in the intake channel or passed through the

255    louvers and were exported (Table A4). Sutphin and Svoboda (2016) reported on delta smelt tagging

256    experiments at the Tracy Fish Facility, which samples fish at the CVP. Sixty-seven batches of 75 to 400

257    adult delta smelt were tagged with visible implant elastomer and released in the intake channel of Jones

258    Pumping Plant, and a fraction were later recovered by the Tracy Fish Facility (Table A5). Recoveries of

259    adult tagged fish were assumed to represent samples when length-based louver selectivity equaled one.

260

261    ***Hydrodynamic data***

262    Monthly average OMR was developed from US Geologic Survey streamflow databases for Old River at

263    Bacon Island and Middle River (https://waterdata.usgs.gov).

264

265    **Model**

266    Motivated by the conceptual model of entrainment described by Kimmerer (2008), we developed a

267    Bayesian hierarchical model of the process leading to the observation of postlarval delta smelt in samples

268    at the SWP and CVP fish facilities between April and June for the years 1995 to 2015. We divided the

269    process into five stochastic sub-processes (Fig. 2): (1) advective transport of fish to the fish facilities, (2)

270    survival of fish during transport, (3) determination of the population length structure, (4) diversion of fish

271    into holding tanks by facility louvers, and (5) subsampling of fish from holding tanks at fish facilities.

272    Sub-processes were modeled as a sequence of probabilities, the product of which was the probability that

273    a fish passing through the source region was observed at the fish facilities. Our primary goal was to

274    estimate the number of postlarval delta smelt at risk of entrainment, either because they entered the

275    facility intake channels or because they were in the South Delta and therefore subject to higher risk of

276    mortality or isolation from the population. A general overview of the model for each sub-process is

277    below, but detailed mathematical descriptions of each sub-process model may be found in Appendix B.

278

279    *Sub-process 1: Transport of fish to facilities.* Transport probabilities $p_{\mathrm{TR}_{qtvs}}$ in year *t* and month *v* from

280    the source region *q* to one of four sinks *s* (Table 1), or final particle locations (labeled SWP, CVP, LRZ,

281    or IRZ), were estimated from observed particle tracking data. We used a Dirichlet-multinomial model to

282    describe the transport sub-process, where the Dirichlet component accounted for state process variability

283    and the multinomial component accounted for observation error in the particle tracking model data. As

284    previously stated, particle tracking data were generated for three potential source regions (Fig. 1), but

285    only one source region was selected to represent delta smelt transport in a single month. Separate models

286    were fit to each source region's particle tracking data, but only a subset of the probabilities estimated for

287    each source region were applied to delta smelt.

288

289    *Sub-process 2: Survival.* During transport from the source to sink regions and before sampling at fish

290    facilities, fish were exposed to predation, starvation and thermally-induced mortality, also known as pre-

291    screen loss (Gingras 1997; Castillo et al. 2012). No estimates of South Delta mortality exist, but we

292    hypothesized that mortality during transport to the SWP and CVP was a function of transport time.

293    Longer transport times resulted in greater exposure to predation and higher mortality. Mortality of fish

294    transported to the LRZ was not modeled, and all fish passing through the source region and transported to

295    the IRZ were assumed to be removed from the population. Survival probability $p_{SV1_{qtv}}$ for the source

296    region selected as the source for a year-month was applied to all fish arriving at the SWP and CVP.

297

298    Additional mortality is experienced by fish going to the SWP. Before entering the SWP intake channel,

299    fish cross a large body of water, Clifton Court Forebay, where they are vulnerable to predation. $p_{SV2_{tv1}}$

300    was the probability of a delta smelt surviving the journey from the Forebay radial gates (where the

301    forebay connects to Old River) to the SWP intake channel in a given year and month. Unlike the other

302    sub-processes in the entrainment model, this sub-process was not directly observed or informed by any

303    data source. Because CVP does not have a forebay, we assumed fish transported to CVP were not subject

304    to this source of mortality and set $p_{SV2_{tv2}} = 1$.

305

306    *Sub-process 3: Population length structure.* We modeled the length structure, or length frequency

307    distribution, of the postlarval delta smelt population so we could use this information in the next sub-

308    process, which addresses size selectivity of the louvers. As in the transport sub-process, we used a

309    Dirichlet-multinomial model, where the Dirichlet component modeled the true population length structure

310    $p_{LN_{tvl}}$, or probability at length $l$, and the multinomial component related delta smelt length frequencies

311    from the 20mm Survey (i.e., the observations) to the population length structure. We divided lengths into

312    five 5mm length classes between 20 and 44mm fork length. We excluded fish 45mm fork length and

313    larger from the 20mm Survey data because delta smelt of this size in spring are likely age 1 and therefore

314    past the postlarval life stage.

315

316     *Sub-process 4: Sampling efficiency at fish facilities.* The sampling efficiency sub-process handled

317     different aspects of how fish facility gear captured entrained fish and was divided into two sampling

318     probabilities. The fish facility louvers elicit an avoidance behavior that cause entrained fish to move into

319     bypass channels that divert them into holding tanks at the fish facilities. Louvers are spaced

320     approximately 20 mm apart. Delta smelt can easily pass through them. We assumed the louvers were size

321     selective, and that the probability of a delta smelt being diverted increases as fish size increases.

322     Conditional probabilities that a fish in the length sample from the fish facilities is of length $l$ (denoted

323     $p_{\mathrm{BYP}_{tvl}}$) depended on population length structure $p_{LN_{tvl}}$ and the length-based selectivity of the louvers $p_{\mathrm{SEL}_l}$.

324     Paralleling the model of observed population length structure, a multinomial model related $p_{\mathrm{BYP}_{tvl}}$ to

325     length frequencies at the fish facilities.

326

327     $p_{\mathrm{BYP}_{tvl}}$ reflected the selectivity of the louvers for one length bin relative to the other bins, but not

328     necessarily the overall efficiency of the louvers for diverting fish from the intake channels to the holding

329     tanks. In particular, $p_{\mathrm{SEL}_l}$ had an upper asymptote at one, while the maximum efficiency of the louvers for

330     delta smelt of any size is likely less than one. Overall length-specific probabilities of fish being diverted

331     to the fish facilities (denoted $p_{\mathrm{EF}_{lf}}$ for length $l$ and fish facility $f$) depended on length-based selectivity and

332     maximum sampling efficiency of the louvers estimated from mark-recapture data.

333

334     *Sub-process 5: Subsampling at fish facilities.* The probability of an entrained postlarval fish reaching the

335     fish facility holding tanks was the product of transport probability, survival probability, and sampling

336     efficiency. We treated the total number of delta smelt enumerated in year $t$ and month $v$ at fish facility $f$,

337     $y_{tvf}$, as a Poisson random variable,

338

339         (1)       $y_{tvf} \sim \mathrm{Poisson}\left(n_{S_{tv}}\, p_{\mathrm{TR}qtvf}\, p_{\mathrm{SV1}_{tv}}\, p_{\mathrm{SV2}_{tvf}} \Sigma_{l=1}^{5}\left(p_{LN_{tvl}}\, p_{\mathrm{EF}_{tvfl}}\right)\frac{\rho_{tvf}}{(1-\omega_{tv})}\right),$

340

341    with expected value equal to the estimated number of delta smelt in the holding tank multiplied by the

342    subsampling rate, $\rho_{tvf}$, and divided by the proportion of fish estimated to be beyond the postlarval stage

343    (namely, 45+mm), $1 - \omega_{tv}$. $n_{S_{tv}}$ represented the number that started downstream in the source region. We

344    calculated year-month-facility specific subsampling rates ($\rho_{tvf}$) as the number of minutes spent sampling

345    fish from the holding tanks divided by the number of minutes during which water was exported; we

346    treated these as fixed values in the model. When subsampling of fish from the holding tanks is carried out,

347    no distinction is made between delta smelt of different life stages. We therefore used the factor $1/$

348    $(1 - \omega_{tv})$ to convert our estimate of the number of postlarval delta smelt in the tanks to an estimate of the

349    total number of delta smelt in the tanks, regardless of age.

350

351    *Number of fish at entrained*: We treated the number of delta smelt passing through the source region $n_{S_{tv}}$

352    for all months and years, as latent variables. We calculated the number of delta smelt entrained $n_E$ as the

353    product of $n_S$ and the proportion of those fish that ended up in IRZ, SWP, or CVP (i.e., the proportion of

354    fish that were not transported to LRZ [s = 3]):

355

356        (2)      $n_{E_{tv}} = n_{S_{tv}}(1 - p_{TR_{qtv3}})$.

357

358    Additional details about the estimation of $\omega_{tv}$ and about $n_{S_{tv}}$ priors may be found in Appendix B.

359

360    ***Simulation model and data***

361    The hierarchical Bayesian model for multiple processes was admittedly complex with many parameters.

362    Two sets of simulations were carried out to determine if, or how well, some of the key parameters could

363    be estimated. One set of simulations focused on the estimability of the number entrained $n_E$ a parameter

364    of primary interest, via estimability of $n_S$. A considerably reduced statistical model treated the estimated

365    number of age-0 fish ending up at the SWP and CVP fish facilities and the probabilities of moving from

366    the source to each of the four sinks as known values.

367

368    A second simulation study examined estimability of key parameters under a much more complex Bayesian

369    hierarchical model that closely paralleled the one applied to the real data and integrated data from multiple

370    sources. The hierarchical model described above was used to define an operating model, based loosely on

371    the parameters estimated for delta smelt. Random sets of true state dynamics were simulated, then new

372    particle tracking, length composition, and fish facility sample data were stochastically simulated from the

373    set of true states. The estimation model was the same as the operating model; it would therefore fit the

374    simulated data perfectly in the absence of observation error, if all parameters were estimable. The model

375    was then fit to each simulated dataset. Additional simulations focused on the effect of errors in the particle

376    tracking model based probabilities of movement from the source region to the four sinks on posterior

377    distributions of key parameters. Further details regarding each of the two simulation studies may be found

378    in Appendix B.

379

380    ***Model fitting and diagnostics***

381    Weakly informative or uninformative prior distribution were used for all parameters. The model was fit

382    using R package R2jags (Su and Yajima 2015) and the Markov Chain Monte Carlo program JAGS

383    (Plummer 2003). R code to run the model is in Appendix C. Model diagnostics were used to assess

384    convergence, posterior correlations between model parameters, model fit to observations, and posterior

385    predictive distributions. Details regarding priors, fitting and diagnostics can be found in Appendix B.

386

387    To explore model robustness, primarily estimation of $n_E$, and to direct future research, data collection,

388    and model refinement, a sensitivity analysis was performed. Sensitivity analyses explored the weighting

389    of transport data, dynamic selection of source regions using empirical values of entrainment probability,

390    Forebay survival estimation (which lacked direct information), the length-based louver selectivity model,

391   and the value of maximum sampling efficiency, respectively. Model sensitivity to each data component

392   was tested by increasing particle tracking data sample size to 4,000, reducing the potential source regions

393   to just the region nearest the water diversions (Upper Old and Middle River), eliminating the survival

394   model by fixing Clifton Court Forebay survival to 1, eliminating the louver selectivity by fixing

395   selectivity at all lengths to a value of 1, and increasing or decreasing maximum sampling efficiency by

396   50%. Sensitivity was measured by proportional change in $n_E$ from the base model.

397

398   **Results**

399

400   ***Simulation***

401   For the simple multinomial model, which assumed known total recoveries at SWP and CVP and known

402   $p_{TR}$ values, maximum likelihood estimates of $n_S$ were unbiased, as expected, and were relatively precise

403   for the given sets of values (Table 2). Assuming that transport was the only source of variation in the

404   number of fish arriving at SWP and CVP, estimability of $n_S$ seems a reasonable assumption. This

405   conclusion was reinforced by the second set of more complicated simulations that is discussed next.

406

407   The second set of simulations explored estimability of entrainment transport, survival, population length,

408   sampling efficiency and selectivity parameters in the more complex Bayesian hierarchical model. All

409   parameters, including the number of entrained fish, appeared to be estimable using the integrated data

410   analysis and Bayesian hierarchical model described here. We concluded that parameters were estimable if

411   z-scores were between -2 and 2, posterior contraction values were greater than 0, and the 95% coverage

412   of the true value was greater than 0.80 (Table 3), but parameters with right skewed prior distributions

413   (lognormal, gamma, or exponential) were expected to be associated with skewed z-scores. Under

414   conditions of low error in transport data as well as the situation of noisy transport data, estimates of $n_E$

415   were relatively unbiased. Though they appeared estimable from most simulated datasets, regression

416 parameters for survival $\beta$ and all errors showed potential for bias, under some conditions, as demonstrated

417 by negative z-scores. The posterior means of the transport random effect $\sigma_{TR}$ and survival standard error

418 $\sigma_{SV2}$ were associated with more error compared to other parameters, as indicated by posterior means

419 versus true values (Fig. 3). $\sigma_{SV2}$ posterior means appeared to be drawn towards the prior means under

420 some conditions, as indicated by a band of posterior $\sigma_{SV2}$ near the prior mean of 0.22. Negative posterior

421 contraction values for length parameters $\alpha_{LN}$, $\sigma_{TR}$, and $\sigma_{SV2}$ showed that even simulated data may contain

422 limited information to estimate these values, and posterior distributions for these parameters should be

423 checked to be sure they moved away from the prior distribution.

424

425 Under the level of error we simulated in transport data, coverage of most parameters was similar to when

426 no additional error was simulated, but some posterior means of transport parameters appeared to be

427 negatively biased (Table 3). The distribution of z-scores for transport parameters included more negative

428 values when additional transport error was simulated; negative bias in transport parameters was sufficient

429 to reduce the 95% coverage of the true values. As lower values of Dirichlet concentration parameters

430 (transport parameters) are associated with greater uncertainty in Dirichlet-distributed probabilities,

431 probabilities were not necessarily biased, but precision was reduced. The relationship between z-scores

432 and posterior contractions for all parameters appeared to be similar regardless of the simulated level of

433 transport noise.

434

435 ***Model fit to delta smelt data***

436 Estimates of entrainment of postlarval delta smelt (posterior means of $n_E$; Eq. 14) during 1995–2015

437 ranged from 561,791 in May 2002 to 10 in May 2006 (Table 4; Fig. 5). Entrainment estimates declined

438 sharply beginning in 2006, and seasonally, the number of postlarvae entrained was greatest in May.

439 Periods with the lowest entrainment estimates were associated with 0s in counts at fish facilities.

440 Coefficients of variation (posterior standard error/posterior mean) for $n_E$ ranged as high as 2.84, reflecting

441 high uncertainty for some year-month combinations. Model diagnostics indicated that all model

442 parameters converged on stationary posterior distributions, had posteriors with more concentrated

443 distributions than priors, and often different means than the prior distributions (Fig. D2). The integrated

444 model presented here facilitated diagnostic checks on each source of information including fits of the

445 transport model to particle tracking data, population length structure model to 20mm Survey length

446 samples, the selectivity model to length samples at the fish facilities, and the observation model to counts

447 at fish facilities. Specific results for each of the 5 processes are presented below, including summaries of

448 posterior distributions (estimates), residual analyses, goodness of fit (Bayesian P-values), correlation

449 among posterior distributions, and sensitivity analyses.

450

451 ***Transport***

452 Estimates of transport probabilities from the source region to SWP ranged from $6 \times 10^{-5}$ to 0.66 and

453 transport to CVP ranged $6 \times 10^{-7}$ to 0.40 (Eq. 1). The effect of OMR on the probability of transport to

454 LRZ (Eq. B2) and the effect of year on transport to the IRZ (Eq. B3) was positive and significantly

455 different from 0 for all months and source regions except Lower Old and Middle rivers (Fig. D3-D5). The

456 year effect on IRZ transport for the Lower Old and Middle rivers source region was not significantly

457 different from 0. Lack of significance was inferred from 95% credible interval coverage of posterior

458 distributions of 0. The model appeared to fit SWP, CVP, and LRZ particle tracking data with little error

459 (Fig. D6), but model fit to IRZ particle tracking data was worse in May and June than in April.  In

460 contrast to SWP and CVP data, IRZ transport data were not anchored by observed counts of fish.

461 Bayesian P-values for particle tracking data were between 0.49 and 0.71 for all sink-month combinations

462 (Table 5). Joint posterior correlation of transport parameters with other model parameters was minimal,

463 but some correlation was evident between LRZ and IRZ intercept and slope parameters ($R^2 < 0.58$).

464

465 Effective multinomial samples size for the model using particle tracking data converged on an estimate of

466 225, using the iterative method of McAllister and Ianelli (1997). All analyses were therefore performed

467    with $N_{PT}$ fixed at 225. Posterior means and coefficients of variation of $n_E$ were sensitive to the value of

468    $N_{PT}$. When $N_{PT}$ was set equal to 4,000, $n_E$ posterior means decreased 18 to 24% on average (Table 6).

469    Sensitivity analyses suggested that the model was less sensitive to the area selected as a source region in a

470    particular year-month combination. Reducing potential source regions to only the region closest to the

471    SWP and CVP water diversions resulted in 1 to 8% higher $n_E$ posterior means.

472

473    ***Survival to SWP***

474    Survival across Clifton Court Forebay to SWP $p_{SV2_1}$ (Eq. 4) appeared to increase over April–June, with

475    mean survival of 0.03 in April, 0.24 in May, and 0.75 in June. As no direct information was available to

476    estimate this process, most diagnostics were not possible; however, survival parameter posteriors did not

477    appear to be highly correlated with the posteriors of other model parameters. We did not expect that the

478    model of delta smelt survival would be highly informed by the available data, as there was no survival

479    data. Simulation testing indicated that survival error $\sigma_{SV2}$ may be poorly informed, even with good

480    correspondence between the model and data. Furthermore, model fit to counts at fish facilities was

481    compromised by inclusion of survival process error $\sigma_{SV2}$, indicated by very large negative residuals. $\sigma_{SV2}$

482    was therefore not estimated, and Eq. 4 was modified to $\text{logit}(p_{SV2_{tv1}}) = \beta_0 + \beta_1 v$. Posterior means of $n_E$

483    were sensitive to survival estimates, e.g., if $p_{SV2_1}$ was fixed at 1 for all months (and years), estimates of $n_E$

484    decreased (Table 6). However, the magnitude of the decrease declined from April to June, which makes

485    sense because the estimated values of $p_{SV2_1}$ increase over this period.

486

487    ***Population and fish facility length structure and sampling efficiency***

488    Consistent with a pattern of growth and increasing mean length, the estimated length distribution in the

489    delta smelt population shifted each month from a mode of 20–24 mm fork length in April to a mode of

490    20–29 mm in May and 30–34 mm in June (Eq. 5). Estimates of the population length distribution

491    accurately fit length distributions from 20mm Survey data; average predicted length compositions

492    appeared to match up with average observed length compositions (Fig. D8; top panels), and most

493    Bayesian P-values indicated adequate fit to length composition data. P-values for larger length classes in

494    April, however, indicated poor fit (Table 5, 20mm length composition sub-table). Although simulation

495    testing suggested that in the hierarchical model we defined, length parameters could become dominated

496    by their prior distributions, as indicated by negative posterior contraction values, this did not appear to be

497    the case for delta smelt length parameters, which had concentrated posterior distributions that moved

498    away from the priors (Fig. D2).

499

500    The model of louver selectivity indicated that the smallest fish, 20-24 mm fork length, had a 0.74

501    probability of being sampled given contact with the louvers (Eq. 9). The slope of the selectivity model ($\eta_0$

502    ) was high, and by 30-34 mm fork length, fish appeared to be fully selected. Model diagnostics indicated

503    somewhat worse fit of the louver selectivity model in June. Mean observed length frequencies at fish

504    facilities were captured within 95% credible intervals (i.e., residuals spanned zero) for April and May but

505    not for the largest length class in June (Fig. D8; bottom panels). High June P-values also indicated poor

506    fit to the largest length classes observed in June at the fish facilities. The selectivity model parameter, $\eta_1$,

507    was not correlated with other model parameters, and posterior means and coefficients of variations of $n_E$

508    did not appear to be sensitive to the louver selectivity model ($p_{SEL}$; Table 6).

509

510    Simulation testing indicated that biases could arise when all transport, survival and efficiency errors were

511    simultaneous estimated, at least when survival data were missing, which is the current situation for

512    postlarval delta smelt.  As was stated previously,  the mark-recapture model was fit externally, and

513    maximum sampling efficiency parameters $\Gamma_f$ and $\sigma_{EF_f}$ were treated as fixed values, with mean SWP

514    efficiency of 0.52 and mean CVP efficiency of 0.17 (Eq. 12) . Estimates of $n_E$ were sensitive to the value

515    assumed for $\Gamma_f$ (Table 6). Lower values resulted in higher estimates of $n_E$, and higher values of $\Gamma_f$ resulted

516    in lower estimates of $n_E$.

517

518     *Subsampling at fish facilities*

519     Overall, the model appeared to accurately predict observed counts at fish facilities. While residuals did

520     indicate a pattern of overestimation of counts at the CVP fish facility (Fig. D7), the medians of all

521     standardized residuals were between 2 and -2. Further, the 95% credible intervals of most standardized

522     fish facility count residuals contained 0. The largest residuals occurred at low values of predicted

523     entrainment, and the highest counts at fish facilities were accurately estimated by the model. Bayesian P-

524     values indicated particularly good and consistent fit to CVP data in all months, while the fit to SWP fish

525     facility counts was worst in April but improved in May and June (Table 5, counts at fish facilities sub-

526     table).

527

528     **Discussion**

529

530     We developed a Bayesian hierarchical model that integrated several sources of information about the

531     entrainment, transport, population length structure and sampling efficiency of postlarval fish. The

532     fundamental property of the hierarchical framework was that some parameters were functions of other

533     parameters. Integration was achieved because parameters were shared among multiple sub-processes. A

534     hierarchy linked data across sub-processes, linked observation models to state process models and

535     facilitated a reduced parameterization when covariates were available. The formulation of process models

536     as linked general linear models (transport and survival) facilitated tests of covariates hypotheses. The

537     hierarchical structure allowed simultaneous accounting of process variation and sampling uncertainty

538     when estimating the abundance of entrained fish. Process variation and observation error can be difficult

539     to disentangle (de Valpine and Hilborn 2005), and this complication may have manifested in the observed

540     tradeoffs between the fit to particle tracking data and counts at fish facilities. Although it is possible to

541     develop a non-hierarchical model that integrates abundance and entrainment estimation, the hierarchical

542    structure provided a clearer accounting of which process uncertainties were represented in estimates of

543    entrainment.

544

545    The inclusion of multiple data sets, in an integrated analysis, meant that information about the abundance

546    of entrained fish came from multiple sources, and integration of mechanistic transport data from

547    hydrodynamics modeling, direct information about sampling efficiency from mark-recapture studies,  and

548    observed counts at fish facilities (sometimes viewed as an index of entrainment) leveraged those data

549    sources against each other. The transport data from particle tracking, for example, provided a measure of

550    movement not contained in other data sets, while counts from SWP and CVP fish facilities adjusted the

551    particle tracking-based transport probabilities and scaled the entrainment estimates. In a sense, the fitted

552    model may be viewed as a calibration of the model provided by particle tracking, and lack of fit, such as

553    the larger May and June particle tracking residuals for the Low and Indirect Risk Zones, may represent

554    instances where particle tracking data was not representative of fish transport. Goodness of fit testing on

555    multiple data sets, within the integrated framework provided a framework to assess those deficiencies and

556    update entrainment estimates should new information become available. Even with hierarchical modeling,

557    some ambiguity remains for delta smelt entrainment because some critical processes were poorly

558    informed by available data (e.g. no survival data). While these data insufficiencies compromised delta

559    smelt estimates, they also demonstrated the benefits of integrating the estimation of several model

560    components and provided a basis to prioritize future data collection.

561

562    ***Modeling the spatial structure of entrainment***

563    Models to estimate entrainment have been applied to three general settings which differ in terms of how

564    spatial structure is handled.  In one setting, individual origins are important, necessitating spatially

565    resolved population models and two- or three-dimensional hydrodynamic modeling (White et al. 2007).

566    In a second setting, entrainment occurs through a complex environment that requires three-dimensional

567    hydrodynamic modeling (Heimbuch et al.2007). Finally, in a third setting and the one presented here,

568    spatial origins are less important and entrainment occurs through a simplified network of channels

569    allowing non-spatial population models and lower hydrodynamic dimensionality (Kimmerer 2008). While

570    complex coastal and estuarine hydrodynamics (the second setting) may require three-dimensional

571    modeling to achieve suitable accuracy, movement of water through linear riverine systems can be viewed

572    as a special case of the third setting and can be accurately modeled in one dimension. For example, the

573    California Department of Water Resources uses the one-dimensional hydrodynamic model DSM2 to

574    evaluate flood risk and manage water transfers (CDWR 2013), and DSM2 has been considered an

575    accurate transport model for early lifestages of delta smelt (Rose et al. 2013; Kimmerer and Rose 2018).

576

577    Transport through a single region per time period was modeled in order to avoid spatially stratifying the

578    sources regions of entrained fish and to reduce the distance over which transport was modeled. Rather

579    than estimating transport from multiple regions, the process of delta smelt transport was assumed to be

580    linearly routed through a single region, extending from the Lower San Joaquin River and up the Old and

581    Middle River corridor to the water diversions. The routing assumption was justified by the fact that only

582    smaller sloughs intersect with the corridor between the SWP and CVP and the confluence with the San

583    Joaquin River.  At the head of Old River or the San Joaquin River upstream of Stockton, delta smelt are

584    rarely captured in monitoring surveys, so it is unlikely that many delta smelt arrive at the SWP and CVP

585    from upstream locations. While restricting the geographic extent of space to just the entrainment routes

586    through the South Delta facilitated a model of transport over relatively short durations, during periods of

587    lesser upstream flow, this advantage was reduced. Based on ad hoc estimates of transport time calculated

588    from the shortest route from the mouth of Old River to the CVP, mean channel area from DSM2

589    documentation (CDWR 2013), and Old and Middle River flow rate during the study period, transport

590    through the Old and Middle River corridor occurred at the scale of days during most of the study period.

591    In situations where entrainment probability was greatest (i.e., high water velocities and strong upstream

592    Old and Middle River flows), transport occurred over a short duration. If errors in a hydrodynamic model

593    of transport increase proportionally to transport time because fish have longer to modify their transport

594      with behaviors, then higher downstream flows, lower entrainment probabilities, and corresponding longer

595      transport times are associated with greater error in a one-dimensional particle tracking model of delta

596      smelt transport. Reducing the subset of potential source regions to areas closer to sampling locations

597      could reduce potential unmodeled interactions between transport time and fish behavior.

598

599      Among the three potential source regions, the region for a particular year-month was selected such that

600      the probability of transport to the Low Risk Zone was close to 0.5, relative to the other potential source

601      regions. Selection of source region influenced estimates of entrainment; sensitivity analysis demonstrated

602      that reduction of source regions to areas near the pumps resulted in lower entrainment estimates during

603      most periods. Transport probabilities to the SWP and CVP are higher in areas closer to the diversions, and

604      probabilities of transport to the Indirect Risk Zone are lower, compared to areas further downstream. The

605      effect of changing source regions was relatively minor, however, compared to the effects of changing

606      effective sample size (weighting of particle tracking data), assumed survival rate, or estimates of sampling

607      efficiency.

608

609      ***Alternate models, data needs, and further work***

610      An important modeling choice was whether to treat transport data as multinomial random observations or

611      as covariates. The distinction between these alternatives was how error was handled. As an observation,

612      transport data are subject to observation error, and this was structured into the state-observation hierarchy

613      of the transport model. As a covariate, transport data could be treated as a predictor in a log-linear

614      Dirichlet regression model, with transport probability as the response; however, linear models generally

615      assume that predictors are measured without error. By treating transport information as an observation,

616      we were able to partially account for the errors in a hydrodynamic model of transport lacking fish

617      behavior.

618

619    Our choice of Dirichlet-multinomial models to describe transport and length observations, given a set of

620    probabilities, introduced a critical uncertainty. Though known quantities of transported particles and fish

621    lengths were sampled, the effective sample sizes were unknown. The multinomial model assumes each

622    observation is independent, but correlations among observations arise from sampling particles released

623    during the same time period or fish lengths measured from groups caught in surveys, resulting in

624    uncertainty in effective multinomial sample sizes. Effective sample sizes were an unknown quantity but

625    can be estimated either externally to the model fitting step, as we did for transport data, (McAllister and

626    Ianelli 1997), or internally as a component of the model likelihood (Thorson et al. 2017). A traditional

627    "rule of thumb" that we applied to length samples, is to set effective sample size equal to the number of

628    independent units from which lengths were sampled, such as number of tows. A final option is to avoid

629    the reweighting associated with choosing an effective sample size by modeling proportions rather than

630    counts (Francis 2017).

631

632    Although we make the argument that one-dimensional hydrodynamic modeling is suitable to model

633    postlarval fish transport, two- or three-dimensional models could be applied to develop transport data and

634    estimate entrainment. During periods of high water export volumes (e.g., as much as 35% of inflow

635    during April–June [CSWR 1999]), advective forces from pumping may dominate transport, but during

636    periods of lower flows, tidal forces dominate the region where delta smelt are entrained. Three-

637    dimensional hydrodynamic modeling may be required to simulate complex tidal mixing (Sridharan et al.

638    2018). Coding particles with hypothesized fish behaviors in coupled hydrodynamics-particle tracking

639    models is a common approach to modeling marine entrainment (Miller 2007). The model presented here

640    is sufficiently flexible to incorporate such information; only numbers of particles reaching a subset of

641    sinks is required for the model to estimate transport rates and entrainment.

642

643    We made simplifying assumptions about the spatiotemporal scale of entrainment, modeling a monthly

644    time interval and transport from only a single source region. A time series approach could be used to

645    model both finer temporal and spatial scales. An entrainment model could be specified to account for the

646    dependence of sequential abundances using a time series in which abundances are linked by survival and

647    regional abundance at time $t$ is equal to abundance times survival from time $t$-1 plus movement from

648    other regions. A time series can be used to leverage entrainment estimates at a finer temporal resolution if

649    only starting abundances need to be estimated. Postlarval delta smelt, however, recruit to the population

650    during the seasonal period we modeled. Postlarval abundance at $t$ does not necessarily depend on

651    abundance at $t$-1, and if abundance must be estimated in each time period, the temporal scale must be

652    coarse in order to reduce the number of abundances to estimate. Important variation in hydrodynamics,

653    related for example to acute storm events, is certainly lost.

654

655    Using a time series approach and more highly resolved transport information, a spatially-explicit model (a

656    special case of the first spatial structure setting mentioned above) could estimate spatially stratified

657    abundance for the entire population, then apply strata-specific transport rates to sampling locations. The

658    drawback to a spatially-explicit approach is that starting population distributions may be highly uncertain

659    (Heimbuch et al. 2007; Huret et al. 2007; Simonis and Merz 2019). Further, any biases in the transport

660    model (e.g., unaccounted fish behavior) compound over both time and space, because fish transport over

661    longer distances must be modeled. For example, few delta smelt are currently captured in monitoring

662    surveys due to low abundance, resulting in imprecise estimates of spatial distributions (USFWS 2018),

663    from which unknown behaviors may modify transport over distances of more than 80 km during the

664    process of entrainment. In the single-region approach we applied, transport assumptions were applied

665    over shorter distances of up to 35 km during low outflow and less than 15 km at high outflow.

666

667    An alternative for modeling transport using particle tracking data would make use of information about

668    fish origins collected from entrained fish observed at fish facilities. Though such information does not

669    currently exist for delta smelt, otolith microchemistry, tagging or genetic data have been used to assign

670    spatial origins to other larval or juvenile fishes. Counts of the number of fish observed at fish facilities,

671    mapped to a set of starting locations, could be used to develop a model paralleling that for transport

672    probabilities and particle tracking data, where Dirichlet-distributed probabilities represent a starting

673    spatial structure times transport probabilities and counts at fish facilities are multinomially distributed into

674    a discrete set of origins.

675

676    We assumed that the group of fish entrained was reset at the beginning of each month, with all fish that

677    entered the South Delta being directly entrained by the SWP and CVP water export pumps, escaping

678    entrainment by being transported back to the Low Risk Zone, or becoming isolated in the South Delta

679    where they eventually died. Future refinements to the model presented here could account for entrained

680    fish that did not die, because they were transported to the Indirect Risk Zone rather than the SWP or CVP,

681    and were then available to be sampled in the subsequent time period, although their entrainment occurred

682    earlier. Transport of Indirect Risk Zone fish can be modeled like entrained fish, using a Dirichlet-

683    multinomial model with the Indirect Risk Zone as the source region of particle tracking data. Numbers

684    available for observation would be the product of the number previously entrained and the joint

685    probability of survival and transport. Particle tracking data for two consecutive one month periods would

686    be required, because the transport of entrained fish from the Indirect Risk Zone at time $t$ depends on their

687    spatial distribution within the Indirect Risk Zone at $t$-1. We were limited in the present study by particle

688    tracking of only one month at a time. Further, available data to model transport were counts of particles

689    released on the first day of each month and tracked until the last day of the month, making them an

690    imperfect index of the continuous process of transport over each month (Culberson et al. 2004; Kimmerer

691    and Nobriga 2008). A preferable strategy would release particles more evenly throughout the month then

692    track them through the end of the subsequent month.

693

694    Sensitivity and simulation testing suggested the model was sensitive to transport, survival, and maximum

695    sampling efficiency but less sensitive to length-based selectivity. This result suggests that more or better

696    delta smelt selectivity information is unlikely to change entrainment estimates but improved delta smelt

697    transport, survival, and maximum sampling efficiency information may be beneficial. In more than ten

698    years since the publication of the first model of delta smelt entrainment (Kimmerer 2008), statistical

699    models have advanced, but little new data to estimate parameters critical to delta smelt entrainment have

700    been collected. Nevertheless, we outline future research to improve the entrainment estimates presented

701    here. The residual pattern of more negative CVP residuals in June suggested a conflict between counts

702    from fish facilities and transport data provided by particle tracking. We expected that as fish progressed to

703    later lifestages, behavioral development would decouple observed transport in counts at fish facilities

704    from hydrodynamics and observed transport in particle tracking data. Further, simulation testing indicated

705    that the precision of entrainment estimates in the Bayesian hierarchical framework were sensitive to errors

706    in transport data; the estimates were only as good as the information provided by particle tracking data.

707    The high standard deviation associated with modeled postlarval delta smelt survival may be symptomatic

708    of this relationship. Research linking ontogenetic development of behaviors in response to environmental

709    cues and transport may help to disentangle these model components in the future.

710

711    Sensitivity analyses revealed that the model was sensitive to the maximum value of sampling efficiency

712    and survival of fish to the SWP (also known as pre-screen loss). Sparse data have been collected to

713    estimate these parameters. Prior research into survival and efficiency was exploratory (Castillo et al.

714    2012), and similar, more focused mark-recapture research is warranted to improve estimates of postlarval

715    delta smelt entrainment. Although we did not model additional mortality for fish transported to the CVP,

716    it is possible that fish experience elevated mortality in the CVP intake channels, analogous to survival in

717    Clifton Court Forebay, before becoming available for observation. Mark-recapture data to estimate

718    survival in Clifton Court Forebay and the intake channels exist for adults and may be used to directly

719    estimate survival of entrained adult delta smelt rather than the indirect methods we used (Smith 2019).

720    Survival during transport to the SWP and CVP was a function of transport time and the daily mortality

721    rate estimated by Bennett (2005), but other models of mortality are possible. Our assumed mortality

722    model was necessary because no estimates of delta smelt mortality in the South Sacramento-San Joaquin

723    Delta exist. We make the strong assumption that all fish entering the Indirect Risk Zone that are not

724    transported back to the Low Risk Zone within one month die. Fish transported back to the Low Risk Zone

725    may experience elevated mortality while in the South Delta, but we cannot account for exposure time to

726    elevated mortality or what that elevated mortality may be. Although particle tracking information

727    accounts for variation in transport due to flow, exports, and tidal cycles, information was not retained

728    about the length of time particle were in the South Delta before returning to the Low Risk Zone.

729

730    Mark-recapture studies could also help to refine the model of length-based louver selectivity. Sensitivity

731    analyses, however, suggested that delta smelt entrainment estimates were not sensitive to selectivity. Low

732    numbers of delta smelt are captured in existing monitoring surveys and fish facility samples, resulting in

733    difficulty estimating population length and selectivity parameters. Model insensitivity to selectivity may

734    be related to the relatively high slope of the fitted selectivity model ($\eta_0$). After the first length class,

735    selectivity to the gear was constant, so changing selectivity only affected values from the first length

736    class. In scenarios with more length classes having lower selectivity, estimates of entrainment may

737    become more sensitive to selectivity estimates.

738

739    **Conclusions**

740    To better guide management decisions made to protect and restore the delta smelt population, estimates of

741    entrainment will be integrated with abundance estimates and errors in the US Fish and Wildlife Service's

742    Delta Smelt Life Cycle Model. Features of the model include a state-space formulation to separate

743    observation errors from stochasticity in survival and reproduction, a time-series approach that smooths

744    over multiple life stages, and formulation using managed environmental quantities. Modeling numbers of

745    entrained fish estimated here as a function of per capita entrainment rates, the hypothesis that past

746    conservation actions changed rates of loss due to entrainment can be tested. Although we estimated large

747    declines in the number of delta smelt postlarvae entrained beginning around the time of implementation

748    of conservation actions by the US Fish and Wildlife Service to reduce entrainment losses, in years 2007

749 and 2008, the overall delta smelt population was also in severe decline at the same time (Moyle et al.

750 2016; Polansky et al. 2019). A population dynamics model that includes survival and reproduction is

751 needed to separate the effects of abundance from entrainment. The accuracy and power of the Delta Smelt

752 Life Cycle Model to detect changes in vital rates depends on the accounting of observation errors

753 described in the current model. Ultimately, this work will aid in the development and assessment of

754 environmental criteria to manage and recover the delta smelt population.

755

756 Entrainment of ichthyoplankton through aquatic systems is a complex function of system hydrodynamics,

757 but sophisticated biophysical models coupling hydrodynamics and particle tracking can provide

758 information about transport rates. Bayesian hierarchical modeling provides a useful tool for integrating

759 models for a sequence of events with disparate datasets that partially inform dynamics of the sequence.

760 With these tools and admittedly strong assumptions about how entrained fish were routed through the

761 system, we demonstrated the possibility of estimating entrainment as the regional abundance of a pelagic

762 postlarval fish while accounting for complex but common sources of variation.

763

764 While past models of entrainment have focused on more complex transport dynamics in marine and

765 estuarine systems, water extraction from a riverine system with more or less linear flow dynamics

766 presents opportunities to make simplifying assumptions about the dimensionality of transport. It is

767 particularly convenient to assume that entrainment follows a linear route through a single region,

768 requiring only a one-dimensional transport model and no assumptions about population starting

769 distributions. With several existing high volume water diversions throughout the U.S. and Canada

770 (Lasserre 2007), many more recently completed or planned in developing areas like

771 Africa (Gereta et al. 2002), China (Wei 2005), and Brazil (Molisani et al. 2006), and growing global

772 demand for freshwater resources, the impacts of water extraction on aquatic populations are likely to

773 increase. Conservation of these aquatic systems and mitigation of impacts from water extraction will

774     require accurate quantification of entrainment losses and a mechanistic model of loss rates as a function

775     of altered hydrodynamics.

776

777     **Acknowledgements**

778

785

786     **References**

787

788     Baumann, H., Hinrichsen, H-H., Möllmann, C., Köster, F.W., Malzahn, A.M., and Temming, A. 2006.

789        Recruitment variability in Baltic Sea sprat (Sprattus sprattus) is tightly coupled to temperature and

790        transport patterns affecting the larval and early juvenile stages. Can. J. Fish. Aquat. Sci. **63**: 2191–

791        2201. doi:10.1139/f06-112.

792

793     Bennett, W.A. 2005. Critical assessment of the delta smelt population in the San Francisco Estuary,

794        California. San Francisco Estuary Watershed Sci, 3. Available from:

795        https://escholarship.org/uc/item/0725n5vk.

796

797     Bennett, W.A. and Burau, J.R. 2015. Riders on the storm: selective tidal movements facilitate the

798          spawning migration of threatened Delta Smelt in the San Francisco Estuary. Estuaries and Coasts

799          **38**(3): 826–835. doi: [10.1007/s12237-014-9877-3](10.1007/s12237-014-9877-3).

800

801     Besbeas, P., Freeman, S.N., Morgan, B.J., and Catchpole, E.A. 2002. Integrating mark–recapture–

802          recovery and census data to estimate animal abundance and demographic parameters. Biometrics **58**:

803          540–547. doi: [10.1111/j.0006-341X.2002.00540.x](10.1111/j.0006-341X.2002.00540.x).

804

805     Blumberg, A. F., Dunnin, D. J., Li, H., Heimbuch, D., and Geyer, W. R. 2004. Use of a particle-tracking

806          model for predicting entrainment at power plants on the Hudson River. Estuaries **27**: 515–526.

807          doi:[10.1007/BF02803543](10.1007/BF02803543).

808

809     Blumberg, A., Goodwin, P., Houde, E., Monismith, S., and Powell, T. M. 2010. Review of IEP and other

810          Bay-Delta modeling focused on hydrodynamics and fish, Interagency Ecological Program. Available

811          from: [http://www.dwr.water.ca.gov/iep/docs/IEPModelWorkshopReview.pdf](http://www.dwr.water.ca.gov/iep/docs/IEPModelWorkshopReview.pdf)

812

813     Boreman, J., Goodyear, C.P., and Christensen, S.W. 1981. An empirical methodology for estimating

814          entrainment losses at power plants sited on estuaries. Trans. Am. Fish. Soc. **110**: 253–260. doi:

815          [10.1577/1548-8659(1981)110<253:AEMFEE>2.0.CO;2](10.1577/1548-8659(1981)110<253:AEMFEE>2.0.CO;2).

816

817     Brown, L. R., Greene, S., Coulston, P., and Barrow, S. 1996. An evaluation of the effectiveness of fish

818          salvage operations at the intake to the California Aqueduct, 1979–1993. *In* San Francisco Bay: the

819          Ecosystem. *Edited by* J. T. Hollibaugh. American Association for the Advancement of Science,

820          Pacific Division, San Francisco, C.A. pp 497–518

821

822    California Department of Agriculture (DFA). 2016. California agricultural production statistics [online].

823        Available from: https://www.cdfa.ca.gov/statistics/

824

825    California Department of Fish and Wildlife (CDFW). 2018. State and federally listed threatened and

826        endangered animals of California.  Available from:

827        https://nrm.dfg.ca.gov/FileHandler.ashx?DocumentID=109405

828

829    California Department of Fish and Wildlife (CDFW). 2019. 20mm survey project overview [online].

830        Available from: https://www.dfg.ca.gov/

831

832    California Department of Water Resources (CDWR). 2011. California State Water Project at a glance

833        [online]. Available from: https://www.water.ca.gov/-/media/DWR-Website/Web-Pages/What-We-

834        Do/Infrastructure/Files/Publications/Californias-State-Water-Project-at-Glance.pdf

835

836    California Department of Water Resources (CDWR). 2013. Delta Simulation Model II [online]. Available

837        from: http://baydeltaoffice.water.ca.gov/modeling/deltamodeling/models/dsm2/dsm2.cfm.

838

839    California State Water Resources Control Board (CSWR). 1999. Water right decision D-1641. California

840        Environmental Protection Agency, Sacramento, CA, USA.

841

842    Castillo, G., Morinaka, J., Lindberg, J., Fujimura, R., Baskerville-Bridges, B., Hobbs, J., Tigan, G. and

843        Ellison, L. 2012. Pre-screen loss and fish facility efficiency for delta smelt at the south Delta's State

844        Water Project, California. San Francisco Estuary Watershed Sci. **10**(4). Available from:

845        https://cloudfront.escholarship.org/dist/prd/content/qt28m595k4/qt28m595k4.pdf.

846

847    Clark, K.W., Bowen, M.D., Mayfield, R.B., Zehfuss, K.P., Taplin, J.D., and Hanson, C.H. 2009.

848        Quantification of pre-screen loss of juvenile steelhead in Clifton Court Forebay, State of California,

849        Sacramento. California Department of Water Resources, Sacramento, CA, USA.

850

851    Culberson, S. D., Harrison, C.B., Enright, C., and Nobriga, M.L. 2004. Sensitivity of larval fish transport

852        to location, timing, and behavior using a particle tracking model in Suisun Marsh, California. Am.

853        Fish. Soc. Symp. **39**: 257–268. Available from:

854        http://www.water.ca.gov/aes/docs/Culberson_et_al_2004.pdf.

855

856    Dixon, D., Veil, J.A., and Wisniewski, J. 2003. Defining and assessing adverse environmental impact

857        from power plant impingement and entrainment of aquatic organisms: symposium in conjunction

858        with the annual meeting of the American Fisheries Society, 2001, in Phoenix, Arizona, USA. CRC

859        Press, Boca Raton, F.L.

860

861    de Valpine, P.D. and Hilborn, R. 2005. State-space likelihoods for nonlinear fisheries time-series. Can. J.

862        Fish. Aquat. Sci. **62**: 1937–1952. doi: 10.1139/f05-116.

863

864    Francis, R. C. (2011). Data weighting in statistical fisheries stock assessment models. Can. J. Fish. Aquat.

865        Sci. **68**: 1124–1138. doi: 10.1139/f2011-025.

866

867    Francis, R.C. 2017. Revisiting data weighting in fisheries stock assessment models. Fish. Res. **192**: 5–15.

868

869    Fujiwara, M., Kendall, B.E., Nisbet, R.M., and Bennett, W.A. 2005. Analysis of size trajectory data using

870        an energetic-based growth model. Ecology **86**(6): 1441–1451. doi: 10.1890/04-1351.

871

872     Gereta, E., Wolanski, E., Borner, M., and Serneels, S. 2002. Use of an ecohydrology model to predict the

873         impact on the Serengeti ecosystem of deforestation, irrigation and the proposed Amala Weir Water

874         Diversion Project in Kenya. Int. J. Ecohydrology Hydrobiology **2**: 135–142.

875

876     Gingras, M. 1997. Mark/recapture experiments at Clifton Court Forebay to estimate pre-screening losses

877         to juvenile fishes 1976–1993. Interagency Ecological Program Technical Report 55. Available from:

878         http://www.water.ca.gov/iep/products/technicalrpts.cfm.

879

880     Grimaldo, L.F., Sommer, T., Van Ark, N., Jones, G., Holland, E., Moyle, P.B., Herbold, B., and Smith, P.

881         2009. Factors affecting fish entrainment into massive water diversions in a tidal freshwater estuary:

882         can fish losses be managed? N. Am. J. Fish. Manage. **29**: 1253–1270. doi:10.1577/M08-062.1.

883

884     Heimbuch, D.G., Dunning, D.J., Ross, Q.E., and Blumberg, A.F. 2007. Assessing potential effects of

885         entrainment and impingement on fish stocks of the New York–New Jersey Harbor estuary and Long

886         Island Sound. Trans. Am. Fish. Soc. **136**: 492–508.

887

888     Hinckley, S., Parada, C., Horne, J.K., Mazur, M. and Woillez, M. 2016. Comparison of individual-based

889         model output to data using a model of walleye pollock early life history in the Gulf of Alaska. Deep-

890         Sea Res. Pt. II **132**: 240–262. doi:10.1016/j.dsr2.2016.04.007.

891

892     Huret, M., Runge, J.A., Chen, C., Cowles, G., Xu, Q. and Pringle, J.M. 2007. Dispersal modeling of fish

893         early life stages: sensitivity with application to Atlantic cod in the western Gulf of Maine. Marine

894         Ecology Progress Series **347**: 261–274. doi: 10.3354/meps06983.

895

896     Karp, C., Hess, L., Lyons, J., and Liston, C. 1997. Evaluation of the sub-sampling procedure to estimate

897         fish salvage at the Tracy Fish Collection Facility, Tracy, California, 1993-1996. Tracy Fish Collection

898    Facility Studies, Volume 8. United States Bureau of Reclamation, Mid-Pacific Region and Denver

899    Technical Service Center, Denver, C.O. Available from: https://www.usbr.gov/mp//TFFIP/technical-

900    reports.html.

901

902    Kelso, J.R. and Milburn, G.S. 1979. Entrainment and impingement of fish by power plants in the Great

903    Lakes which use the once-through cooling process. J. Great Lakes Res. **5**: 182–194.

904    doi:10.1016/S0380-1330(79)72145-9.

905

906    Kimmerer, W.J. 2008. Losses of Sacramento River Chinook salmon and delta smelt to entrainment in

907    water diversions in the Sacramento–San Joaquin Delta. San Francisco Estuary and Watershed Sci.

908    **6**(2).

909

910    Kimmerer, W.J. 2011. Modeling Delta Smelt losses at the south Delta export facilities. San Francisco

911    Estuary and Watershed Sci. **9**(1).

912

913    Kimmerer, W.J. and Nobriga, M.L. 2008. Investigating Particle Transport and Fate in the Sacramento-San

914    Joaquin Delta Using a Particle-Tracking Model. San Francisco Estuary and Watershed Sci. **6**(1).

915    Available from: https://cloudfront.escholarship.org/dist/prd/content/qt547917gn/qt547917gn.pdf.

916

917    Kimmerer, W.J. and Rose, K.A. 2018. Individual-Based Modeling of Delta Smelt Population Dynamics

918    in the Upper San Francisco Estuary III. Effects of Entrainment Mortality and Changes in Prey. Trans.

919    Am. Fish. Soc. **147**(1): 223–243.

920

921    Korman, J., Gross, E.S., Smith, P.E., Saenz, B., and Grimaldo, L. *In prep*. Statistical evaluation of

922    particle-tracking models predicting proportional entrainment loss for adult delta smelt in the

923    Sacramento-San Joaquin Delta.

924

925 Lasserre, F. 2005. Massive water diversion schemes in North America: a solution to water scarcity? WIT

926     Transactions on Ecology and the Environment **80**. Available from:

927     https://www.witpress.com/Secure/elibrary/papers/WRM05/WRM05039FU.pdf

928

929 Lough, R.G. and Manning, J.P. 2001. Tidal-front entrainment and retention of fish larvae on the southern

930     flank of Georges Bank. Deep-Sea Res. Pt. II **48**: 631–644. doi:10.1016/S0967-0645(00)00130-2.

931

932 Mager R., Doroshov, S.I., Van Eenennaam, J.P., and Brown, R.L. 2004. Early life stages of delta smelt.

933     In: Freyer, F., L. R. Brown, R. L. Brown, J. J. Orsi, editors. Early life history of fishes in the San

934     Francisco estuary and watershed. Am. Fish. Soc. Symp. **39**: 169–180.

935

936 Maunder, M.N. and Deriso, R.B. 2011. A state–space multistage life cycle model to evaluate population

937     impacts in the presence of density dependence: illustrated with application to delta smelt

938     (Hyposmesus transpacificus). Can. J. Fish. Aquat. Sci. **68**(7): 1285–1306. doi: 10.1139/f2011-071.

939

940 McAllister, M.K. and Ianelli, J.N. 1997. Bayesian stock assessment using catch-age data and the

941     sampling-importance resampling algorithm. Can. J. Fish. Aquat. Sci. **54**(2): 284–300.

942

943 Miller, T.J. 2007. Contribution of individual-based coupled physical–biological models to understanding

944     recruitment in marine fish populations. Mar. Ecol. Prog. Ser. **347**: 127–138.

945

946 Miller, W J. 2011. Revisiting assumptions that underlie estimates of proportional entrainment of delta

947     smelt by state and federal water diversions from the Sacramento-San Joaquin Delta. San Francisco

948     Estuary and Watershed Sci. **9**(1).

949

950    Miller, W. J., Manly, B. F., Murphy, D. D., Fullerton, D., & Ramey, R. R. (2012). An investigation of

951      factors affecting the decline of delta smelt (Hypomesus transpacificus) in the Sacramento-San

952      Joaquin Estuary. Rev. in Fish. Sci. **20**(1): 1–19.

953

954    Methot, R.D. and Wetzel, C.R. 2013. Stock synthesis: a biological and statistical framework for fish stock

955      assessment and fishery management. Fish. Res. **142**: 86–99. doi:10.1016/j.fishres.2012.10.012.

956

957    Molisani, M.M., Kjerfve, B., Silva, A.P., and Lacerda, L.D. 2006. Water discharge and sediment load to

958      Sepetiba Bay from an anthropogenically-altered drainage basin, SE Brazil. J. Hydrology **331**: 425–

959      433.

960

961    Moyle, P.B., Brown, L.R., Durand, J.R. and Hobbs, J.A. 2016. Delta smelt: life history and decline of a

962      once-abundant species in the San Francisco Estuary. San Francisco Estuary and Watershed Sci. **14**(2).

963

964    Newman, K.B., Buckland, S.T., Morgan, B.J., King, R., Borchers, D.L., Cole, D.J., Besbeas, P., Gimenez,

965      O., and Thomas, L. 2014. Modelling population dynamics: model formulation, fitting and assessment

966      using state-space methods. Springer, New York City, N.Y.

967

968    National Marine Fisheries Service (NMFS). 2009. Biological opinion on long-term operations of

969      the Central Valley Project and State Water Project. Available from:

970      https://www.westcoast.fisheries.noaa.gov/publications/Central_Valley/

971

972    North, E.W., Schlag, Z., Hood, R.R., Li, M., Zhong, L. Gross, T., and Kennedy, V.S. 2008. Vertical

973      swimming behavior influences the dispersal of simulated oyster larvae in a coupled particle-tracking

974    and hydrodynamic model of Chesapeake Bay. Mar. Ecol. Prog. Ser. **359**: 99–115.

975    doi:10.3354/meps07317.

976

977    Plummer, M. 2003. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling.

978    Proceedings of the 3rd International Workshop on Distributed Statistical Computing, Vienna, Austria.

979

980    Polansky, L., L. Mitchell, and K. Newman. 2019. Using multistage design-based methods to construct

981    abundance indices and uncertainty measures for Delta Smelt. Trans. Am. Fish. Soc. **148**(4): 710–724.

982    doi: 10.1002/tafs.10166.

983

984    Rochette, S., O. Le Pape, J. Vigneau, and E. Rivot. 2013. A hierarchical Bayesian model for

985    embedding larval drift and habitat models in integrated life cycles for exploited fish. Eco.

986    Apps. **23**(7): 1659–1676.

987

988    Rose, K.A., Kimmerer, W.J., Edwards, K.P., and Bennett, W.A. 2013. Individual-based modeling of

989    Delta Smelt population dynamics in the upper San Francisco Estuary: II. Alternative baselines and

990    good versus bad years. Trans. Am. Fish. Soc. **142**(5): 1260–1272. doi:

991    10.1080/00028487.2013.799519.

992

993    Royle, J.A. 2004. N-mixture models for estimating population size from spatially replicated counts.

994    Biometrics **60**(1): 108–115.

995

996    Simonis, J.L. and Merz, J.E. 2019. Prey availability, environmental constraints, and aggregation dictate

997    population distribution of an imperiled fish. Ecosphere **10**(3): e02634.

998

999    Sridharan, V.K., Monismith, S.G., Fong, D.A. and Hench, J.L. 2018. One-Dimensional Particle Tracking

1000    with Streamline Preserving Junctions for Flows in Channel Networks. J. Hydraul. Eng. **144**(2).

1001

1002    Smith, W. 2019. Integration of transport, survival, and sampling efficiency in a model of South Delta

1003    entrainment. San Francisco Estuary and Watershed Sci (in press).

1004

1005    Sutphin, Z.A. and Svoboda, C.D. 2016. Effects of Hydraulic Conditions on Salvage Efficiency of Adult

1006    delta smelt at the Tracy Fish Collection Facility. Tracy Series 43. Available from United States

1007    Department of the Interior, Bureau of Reclamation, Washington, D.C. Available from:

1008    https://www.usbr.gov/mp//TFFIP/technical-reports.html.

1009

1010    Swanson, C., Young, P.S., and Cech, J.J. 1998. Swimming performance of delta smelt: maximum

1011    performance, and behavioral and kinematic limitations on swimming at submaximal velocities. J.

1012    Exp. Biol. **201**: 333–345.

1013

1014    Thorson, J.T., Johnson, K.F., Methot, R.D., and Taylor, I.G. 2017. Model-based estimates of effective

1015    sample size in stock assessment models using the Dirichlet-multinomial distribution. Fish. Res. **192**:

1016    84–93.

1017

1018    United States Fish and Wildlife Service (USFWS). 2008. Biological opinion on the long-term operational

1019    criteria and plan for coordination of the Central Valley Project and State Water Project. Available

1020    from: https://www.fws.gov/sfbaydelta/CVP-SWP/index.htm

1021

1022    United States Fish and Wildlife Service (USFWS). 2018. Enhanced delta smelt monitoring 2018 phase 3

1023    sampling preliminary analysis. Available from:

1024    https://www.fws.gov/lodi/juvenile_fish_monitoring_program/jfmp_index.htm

1025

1026    Wei, D. 2005. Beijing water resources and the south to north water diversion project. Can. J. Civ. Eng 32:

1027        159–163. doi: 10.1139/l04-113.

1028

1029    White, J.W., Nickols, K.J., Clarke, L., and Largier, J.L. 2010. Larval entrainment in cooling water

1030        intakes: spatially explicit models reveal effects on benthic metapopulations and shortcomings of

1031        traditional assessments. Can. J. Fish. Aquat. Sci. **67**(12): 2014–2031. doi: 10.1139/F10-108.

**Table 1.** List of parameter indices, state dynamics probabilities and corresponding model parameters.

| Index | Values | Description |
|---|---|---|
| f | f = 1, 2 (SWP, CVP) | fish facility |
| q | q = 1, 2, 3 (Prisoners Point, Lower OMR, Upper OMR) | source region |
| t | t = 1, 2, 3, … 21 (1995 … 2016) | year |
| v | v = 1, 2, 3 (April, May, June) | month |
| s | s = 1, 2, 3, 4 (SWP, CVP, LRZ, IRZ) | sink |
| $l$ | $l$ = 1, 2, 3, 4, 5 | length class |

| Process | State dynamic probability | Parameters | Parameter description |
|---|---|---|---|
| 1. Transport (TR) | $p_{TR_{qtvs}}$ | $n_{S_{tv}}$ | number passing downstream edge of source region, |
| | | $n_{E_{tv}}$ | number entrained |
| | | $\alpha_{TR_{qtvs}}, A_{TR_{qvs}}, \sigma_{TR}$ | Dirichlet concentration ($\alpha$), linear regression parameters ($A$), and standard deviation ($\sigma$) of regression parameters for transport |
| 2. Survival (SV) | $p_{SV1_{qtv}}$ | | survival of transport to SWP and CVP |
| | $p_{SV2_{tvs}}$ | $\beta,$ $\sigma_{SV}$ | logistic regression parameters for survival across Clifton Court Forebay to State Water Project (SWP), standard deviation of survival |
| 3. Population length structure (LN) | $p_{LN_{tvl}}$ | $\alpha_{LN_{v,l}}$ | Dirichlet concentration for monthly mean population proportion at length |
| 4. Sampling efficiency (EF) | $p_{EF_{tvfl}}$ | $\Gamma_f, \gamma_{ftv},$ $\sigma_{EF}$ | maximum sampling efficiency ($\Gamma$), dynamic efficiency ($\gamma$), and standard deviation ($\sigma$) of dynamic efficiencies |
| | | $p_{BYP_{tvl}}$, where $p_{BYP_{tvl}} = f(p_{SEL_{tvl}}|\eta_1, l)$ | length-based louver selectivity and logistic regression parameters ($\eta$) |
| 5. Subsampling | | $\rho_{tvf},$ $\omega_{tv}$ | subsampling rate, proportion age-1 fish in fish facility counts |

**Table 2**. Results of the first simulation, testing the estimability of the number of fish entering the source subregion $n_S$, given assumed transport rates $p_{TR}$ and observed counts at the State Water Project (SWP) and Central Valley Project (CVP). Mean estimated $n_S$ is shown, with associated coefficient of variation in parentheses.

| | $n_S = 500$ | | $n_S = 1{,}000$ | |
| --- | --- | --- | --- | --- |
| | $p_{TR_{SWP}} = 0.1$ | $p_{TR_{SWP}} = 0.4$ | $p_{TR_{SWP}} = 0.1$ | $p_{TR_{SWP}} = 0.4$ |
| $p_{TR_{CVP}} = 0.05$ | 498.2 (10.7) | 489.2 (4.9) | 999.9 (7.5) | 999.3 (3.5) |
| $p_{TR_{CVP}} = 0.3$ | 499.7 (5.6) | 498.8 (2.9) | 997.5 (3.8) | 999.5 (2.1) |

**Table 3**. Comparison of observed and nominal coverage probabilities for selected parameters, among 500 simulations. $\alpha_{TR}$ and $A_{TR}$ represented transport parameters (Eq. B2-B3), $A_{LN}$ represented population length parameters (Eq. B7), $\beta$ represented survival parameters (Eq. B6), $\Gamma$ represented efficiency parameters (Eq. B12), $\eta$ represented louver selectivity parameters ($\eta_1$; Eq. B11), and $n_E$ represented the abundance of entrained fish (Eq. 2).

| Parameter | 95% coverage | | Quantiles of z-scores (0.025, 0.975) | |
|---|---|---|---|---|
| | PTM = $p_{TR}$ | PTM = $p_{TR}$ +$error$ | PTM = $p_{TR}$ | PTM = $p_{TR}$ +$error$ |
| $\alpha_{TR_{SWP}}$ and $\alpha_{TR_{CVP}}$ | 0.96 | 0.89 | (-3.07, 1.03) | (-4.66, 1.08) |
| $A_{TR_0}$ | 0.96 | 0.84 | (-1.85, 1.95) | (-3.30, 1.77) |
| $A_{TR_1}$ | 0.94 | 0.90 | (-1.99, 2.19) | (-2.19, 2.62) |
| $A_{LN}$ | 0.96 | 0.96 | (-2.32, 0.77) | (-2.22, 0.77) |
| $\beta_0$ | 0.95 | 0.94 | (-2.21, 1.71) | (-2.34, 1.60) |
| $\beta_1$ | 0.95 | 0.95 | (-2.09, 1.75) | (-1.90, 1.90) |
| $\Gamma$ | 0.95 | 0.94 | (-2.03, 1.84) | (-2.10, 1.86) |
| $\eta$ | 0.94 | 0.96 | (-1.78, 2.19) | (-1.79, 2.11) |
| $n_E$ | 0.94 | 0.94 | (-2.40, 1.59) | (-2.41, 1.61) |
| $\sigma_{TR}$ | 0.97 | 0.97 | (-2.64, 0.91) | (-2.51, 0.95) |
| $\sigma_{SV}$ | 0.99 | 0.98 | (-1.95, 1.26) | (-2.23, 1.31) |
| $\sigma_{EF}$ | 0.95 | 0.95 | (-2.42, 1.69) | (-2.27, 1.65) |

**Table 4.** Estimated total entrainment of postlarval delta smelt. Standard deviations of estimates are listed in parentheses.

| Year | April | May | June |
|------|-------|-----|------|
| 1995 | 335 (773) | 33 (94) | 111 (125) |
| 1996 | 3,630 (3,394) | 198,095 (132,275) | 42,523 (26,211) |
| 1997 | 19,418 (16,076) | 201,537 (119,769) | 42,246 (20,238) |
| 1998 | 360 (654) | 117 (287) | 526 (1,073) |
| 1999 | 18,047 (13,694) | 304,947 (169,533) | 501,292 (260,490) |
| 2000 | 21,469 (17,439) | 312,135 (151,684) | 168,847 (137,927) |
| 2001 | 50,333 (35,347) | 261,005 (144,519) | 146,483 (108,439) |
| 2002 | 4,765 (3,907) | 561,791 (246,423) | 60,037 (32,188) |
| 2003 | 5,037 (4,066) | 176,258 (88,236) | 45,909 (26,479) |
| 2004 | 4,608 (5,165) | 79,926 (42,366) | 36,725 (14,776) |
| 2005 | 1,672 (1,881) | 7,495 (8,248) | 5,628 (4,568) |
| 2006 | 1,235 (1,608) | 10 (23) | 341 (393) |
| 2007 | 783 (771) | 9,244 (4,970) | 10,477 (4,135) |
| 2008 | 2,929 (3,163) | 6,689 (4,271) | 5,446 (3,131) |
| 2009 | 1,202 (1,295) | 6,701 (5,720) | 6,170 (3,163) |
| 2010 | 744 (927) | 1,035 (1,005) | 200 (205) |
| 2011 | 296 (753) | 3,753 (4,927) | 1,002 (859) |
| 2012 | 13,100 (7,717) | 32,519 (17,604) | 8,899 (4,748) |
| 2013 | 8,058 (6,467) | 28,474 (25,501) | 483 (387) |
| 2014 | 1,072 (1,354) | 3,275 (4,303) | 32 (35) |
| 2015 | 3,881 (4,317) | 1,877 (2,745) | 24 (28) |

**Table 5.** Bayesian P-values for each data component, indicating goodness of fit. P-values represent the proportion of posterior samples in which the model fit observed data better than replicated data, where fit was quantified using the Freeman-Tukey statistic. Particle tracking sinks and fish facilities were the State Water Project (SWP), Central Valley Project (CVP), Low Risk Zone (LRZ) and Indirect Risk Zone (IRZ). Values near 0.5 represent a good fit, with better fit to observed data in exactly half of posterior samples, and values near 0 or 1 indicate poor fit or inability of the model to reproduce observed data.

| | Particle tracking | | | | |
| --- | --- | --- | --- | --- | --- |
| | SWP | CVP | LRZ | IRZ | |
| April | 0.59 | 0.58 | 0.67 | 0.71 | |
| May | 0.63 | 0.61 | 0.61 | 0.55 | |
| June | 0.57 | 0.55 | 0.51 | 0.49 | |
| | 20mm length composition | | | | |
| | 20-24mm | 25-29mm | 30-34mm | 35-39mm | 40-44mm |
| April | 0.82 | 0.86 | 0.98 | 0.98 | 0.96 |
| May | 0.58 | 0.45 | 0.47 | 0.61 | 0.76 |
| June | 0.58 | 0.57 | 0.44 | 0.44 | 0.66 |
| | Length composition at fish facilities | | | | |
| April | 0.56 | 0.82 | 0.73 | 0.76 | 0.78 |
| May | 0.50 | 0.41 | 0.74 | 0.50 | 0.35 |
| June | 0.58 | 0.75 | 0.68 | 0.71 | 0.90 |
| | Counts at fish facilities | | | | |
| | SWP | CVP | | | |
| April | 0.82 | 0.68 | | | |
| May | 0.69 | 0.59 | | | |
| June | 0.66 | 0.69 | | | |

**Table 6.** Mean relative changes in posterior means of $n_E$ (and the relative change in coefficient of variation) from a sensitivity analysis, showing change due to number of particles used ($N_{PT}$) in the multinomial model of transport data, reduction of source regions to just the region closest to the water diversions, the survival of postlarval delta smelt in Clifton Court Forebay ($p_{SV2}$), the inclusion of a length-based selectivity function ($p_{SEL}$), and to decreasing (-50%) or increasing (+50%) the maximum value of sampling efficiency ($\Gamma$). Proportional change was calculated as the difference in posterior mean from the base model divided by posterior mean from the base model (e.g., if difference in $n_E$ from the base model = 50, and the posterior mean of $n_E$ from the base model = 100, then proportional change = 0.5)

| Sensitivity analysis | Month | | |
|---|---|---|---|
| | April | May | June |
| $N_{PT} = 4,000$ | -0.18 (-0.02) | -0.22 (-0.03) | 0.24 (-0.07) |
| source region = Upper Old and Middle River | 0.01 (-0.04) | 0.08 (-0.03) | 0.07 (0.01) |
| $p_{SV2} = 1$ | -0.63 (-0.02) | -0.45 (0.11) | -0.15 (0.01) |
| $p_{SEL} = 1$ | -0.0004 (-0.01) | 0.02 (0.01) | 0.01 (0.01) |
| - 50% $\Gamma$ | 0.42 (-0.01) | 0.58 (-0.04) | 0.42 (0.05) |
| + 50% $\Gamma$ | -0.31 (0.02) | -0.32 (0.01) | -0.21 (-0.06) |

**Figure captions**

**Fig. 1.** Map of the west coast of North America, showing the location of the Sacramento-San Joaquin Delta (panel A) and the three subregions of the Sacramento-San Joaquin Delta used in DSM2 particle tracking. The natural direction of net Delta flow is indicated by the open arrows in panel A, but under high water export and low inflow conditions, the direction of flow in the southern portion of the Delta reverses and flows southwards, along the Old and Middle rivers to the pumps. Source subregions are indicated in red (panels B, C and D), and the corresponding Indirect Risk Zone is indicated in yellow but also includes the source subregions. All unshaded areas in the Delta were considered the Low Risk Zone.

**Fig. 2.** Diagram of the links between the conceptual model of state dynamics, the quantitative model, and data.

**Fig. 3.** Posterior means from the hierarchical model versus true simulated parameter values.

**Fig. 4.** Z-scores versus posterior contractions for all model parameters estimated from simulated data.

**Fig. 5.** Boxplots of posterior distributions of the total numbers of postlarval delta smelt entrained $n_E$ in each year. Medians are indicated by the bolded line; interquartile ranges are indicated by the boxes, and the 95% credible interval is indicated by the dashed whiskers.
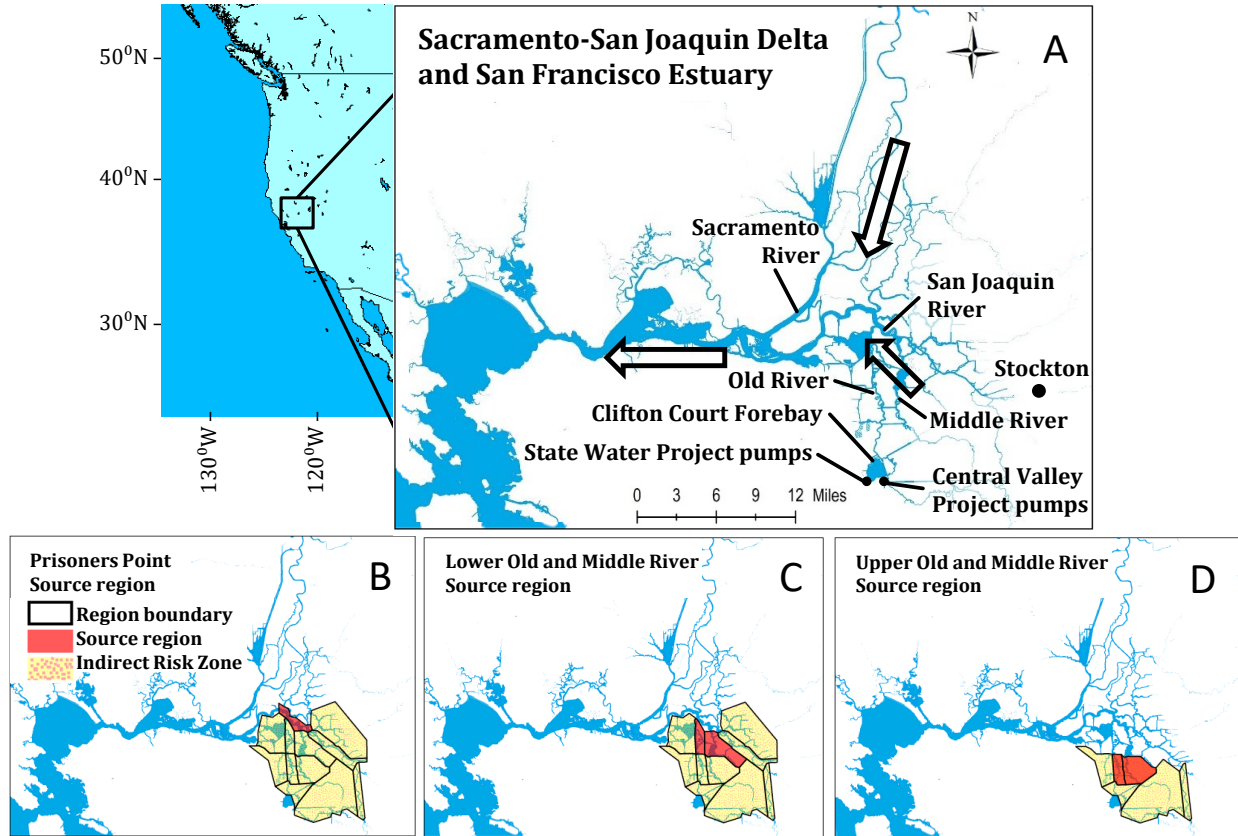
**Figure 1.**

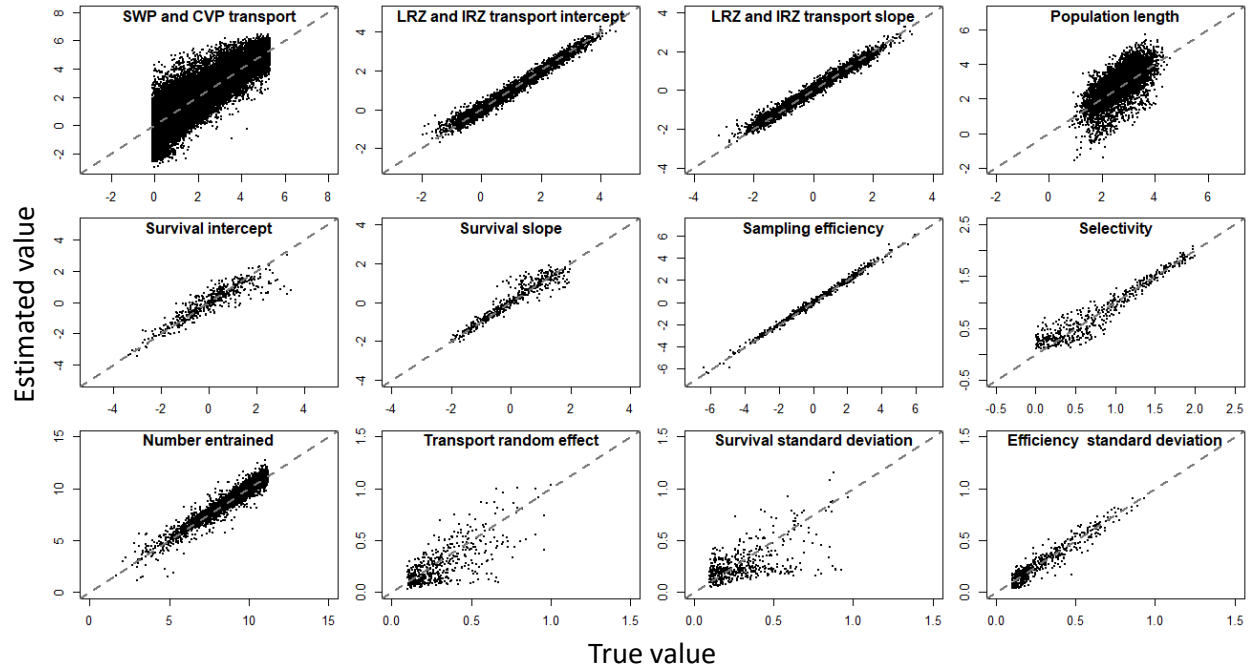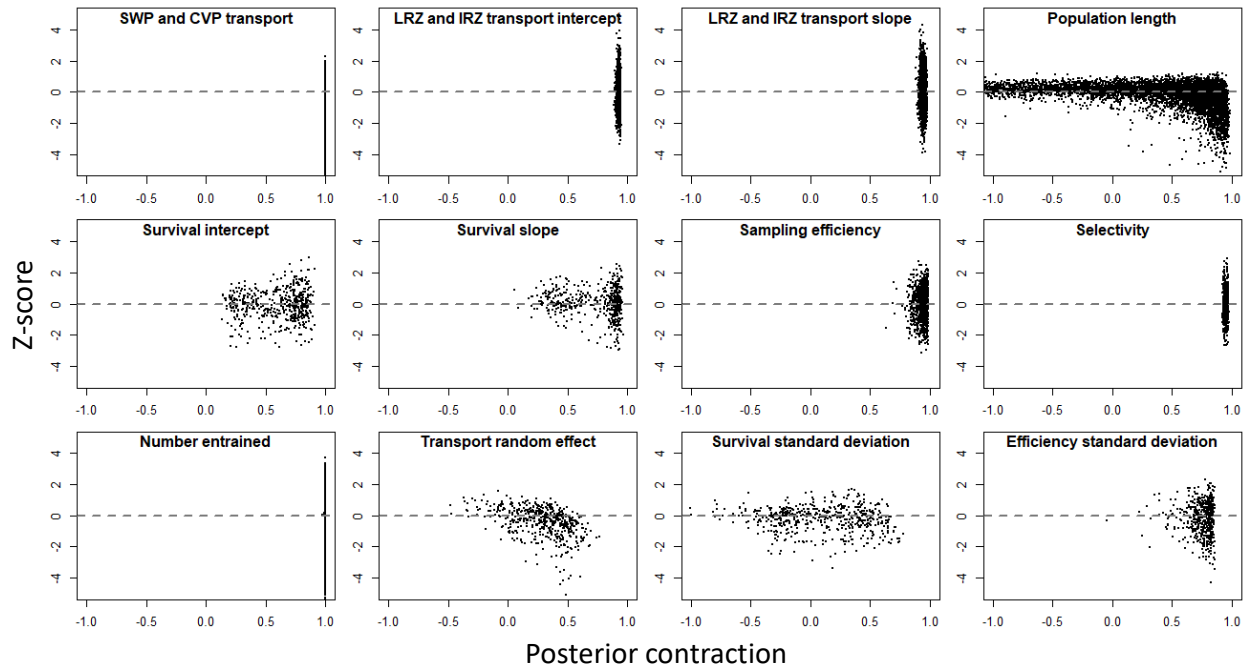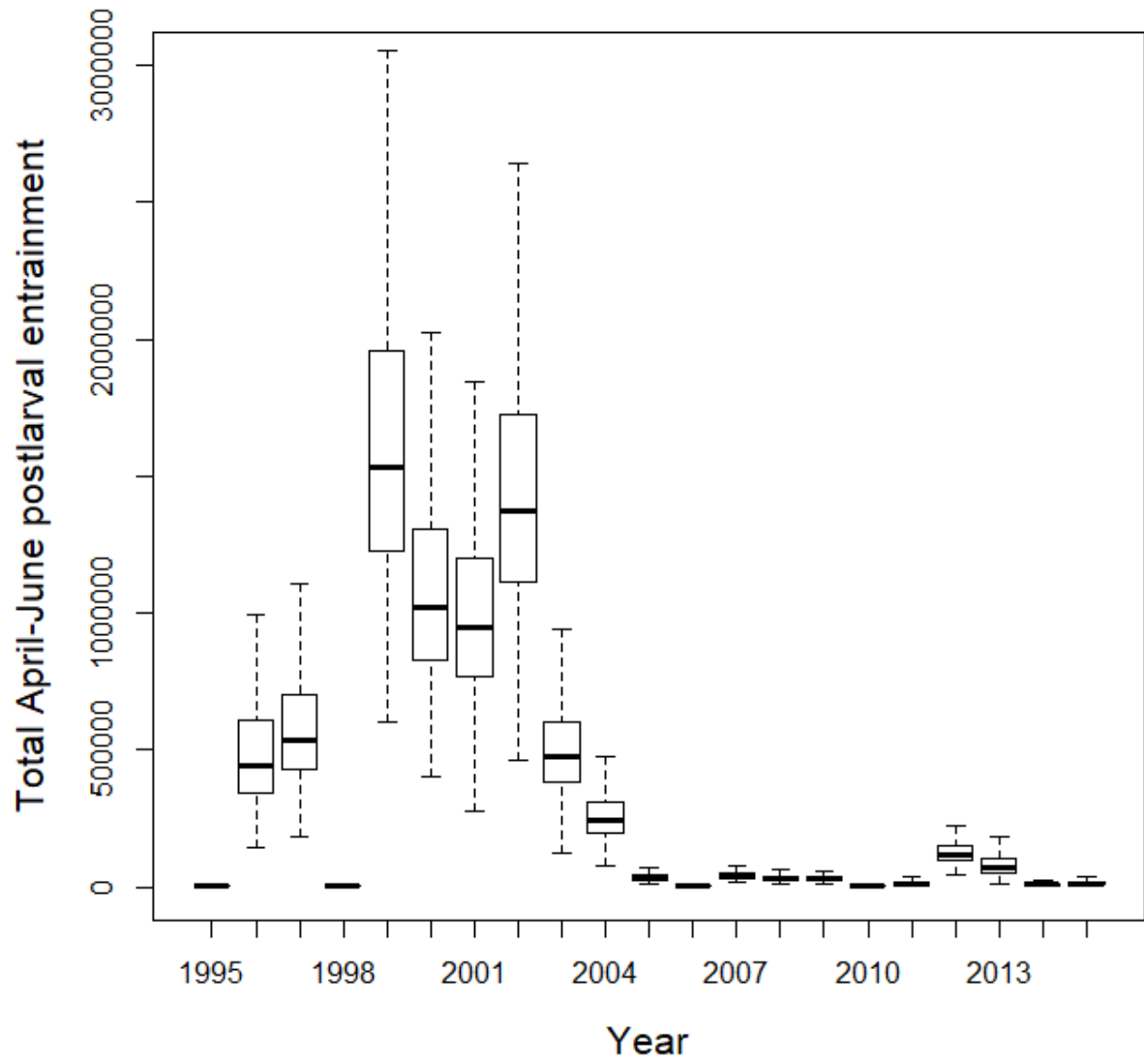| Process | States | State Model | Data | n parameters | n observations |
|---|---|---|---|---|---|
| 1: Transport | $n_{S \to IRZ} \leftarrow n_S \to n_{S \to LRZ}$ <br> $p_{TR}$ <br> $n_{S \to SWP}$    $n_{S \to CVP}$ | estimate $n_{S_{tv}}$ <br><br> $\begin{bmatrix} p_{TR_{qtvSWP}} \\ \vdots \\ p_{TR_{qtvIRZ}} \end{bmatrix} \sim \text{Dirichlet}\left( \begin{bmatrix} \alpha_{TR_{qtvSWP}} \\ \vdots \\ \alpha_{TR_{qtvIRZ}} \end{bmatrix} \right)$ <br> where $\alpha_{TR_{qtvSWP}}$ and $\alpha_{TR_{qtvCVP}}$ are independent among year-months, <br> $\alpha_{TR_{qtvLRZ}} = A_{TR_{0qvLRZ}} + A_{TR_{1qvLRZ}} * OMR_{tv}, \sigma_{TR}$ <br> $\alpha_{TR_{qtvIRZ}} = A_{TR_{0qvIRZ}} + A_{TR_{1qvIRZ}} * t,$ and <br> $A_{TR_{0:1}}$ are random effects with sink-specific means <br> One transport probability $p_{PT}$ was selected for a time period from among multiple source regions $q$ using empirical $p_{PT_{qtvLRZ}}$ | Particle tracking $(m_{PT})$ <br> $\begin{bmatrix} m_{PT_{qtvSWP}} \\ \vdots \\ m_{PT_{qtvIRZ}} \end{bmatrix} \sim \text{Multinomial}\left( PT.N, \begin{bmatrix} p_{TR_{qtvSWP}} \\ \vdots \\ p_{TR_{qtvIRZ}} \end{bmatrix} \right)$ | 131 | 756 |
| 2: Survival | $p_{SV1} * p_{SV2}$ <br> $n_{SV}$    $n_{SV}$ | $p_{SV1_{tv}} = e^{-0.006 * T_{TR_{tv}}}$ <br> $p_{SV2_{tvCVP}} = 1$ <br> $\text{logit}(p_{SV2_{tvSWP}}) = \beta_{tvSWP},$ <br> where $\beta_{tvSWP} \sim \text{Normal}(B_0 * (v - B_1), \sigma_{SV})$ | | 3 | 0 |
| 3. Population length structure | $\begin{pmatrix} n_{LN_{l=1}} \\ \vdots \\ n_{LN_{l=5}} \end{pmatrix}$   $p_{LN}$   $\begin{pmatrix} n_{LN_{l=1}} \\ \vdots \\ n_{LN_{l=5}} \end{pmatrix}$ | $\begin{bmatrix} p_{LN_{tvl=1}} \\ \vdots \\ p_{LN_{tvl=5}} \end{bmatrix} \sim \text{Dirichlet}\left( \begin{bmatrix} \alpha_{LN_{vl=1}} \\ \vdots \\ \alpha_{LN_{vl=5}} \end{bmatrix} \right)$ | TMM length frequency $(m_{20MM})$ <br> $\begin{bmatrix} m_{20MM_{tvl=1}} \\ \vdots \\ m_{20MM_{tvl=5}} \end{bmatrix} \sim \text{Multinomial}\left( \sum_{k=1}^{5} M_{20MM_{tvk}}, \begin{bmatrix} p_{LN_{tvl=1}} \\ \vdots \\ p_{LN_{tvl=5}} \end{bmatrix} \right)$ | 15 | 315 |
| 4. Sampling efficiency | $p_{EF}$ <br> $\begin{pmatrix} n_{EF_{l=1}} \\ \vdots \\ n_{EF_{l=5}} \end{pmatrix}$   $\begin{pmatrix} n_{EF_{l=1}} \\ \vdots \\ n_{EF_{l=5}} \end{pmatrix}$ | $\text{logit}(p_{SEL_l}) = \dfrac{-\ln\left(\frac{0.001}{0.999}\right)}{(5 - \eta_1)}(l - \eta_1)$ <br><br> $p_{BYP_{tvl}} = \dfrac{p_{LN_{tvl}} * p_{SEL_l}}{\sum_{k=1}^{5} p_{LN_{tv5}} * p_{SEL_k}}$ <br> ............................................... <br> $p_{EF_{tvlf}} = \dfrac{1}{(1 + e^{-\gamma_{tvf}})} * p_{BYP_{tvl}},$ <br> where $\gamma_{tvf} \sim \text{Normal}\left( \Gamma_f, \sigma_{EF_f} \right)$ | Salvage length frequency $(m_{FF})$ <br> $\begin{bmatrix} m_{FF_{tvl=1}} \\ \vdots \\ m_{FF_{tvl=5}} \end{bmatrix} \sim \text{Multinomial}\left( \sum_{k=1}^{5} M_{FF_{tvk}}, \begin{bmatrix} p_{BYP_{tvl=1}} \\ \vdots \\ p_{BYP_{tvl=5}} \end{bmatrix} \right)$ <br><br> Mark-recap $(r)$ <br> $r_{fi} \sim \text{Binomial}\left( N_{tagged}, \dfrac{1}{1 + e^{-\gamma_{if}}} \right)$ | 1 <br><br> 4 | 315 <br><br> 73 |
| 5. Subsampling | $\dfrac{\rho}{(1 - \omega)}$ <br> $y_{SWP}$    $y_{CVP}$ | $\lambda_{tvf} = n_{S_{tv}} * p_{TR_{tvf}} * p_{SV1_{tv}} * p_{SV2_{tvf}} *$ <br> $\sum_{k=1}^{5} (p_{LN_{tvk}} * p_{EF_{tvkf}}) * \dfrac{\rho_{tvf}}{(1 - \omega_{tv})}$ <br> where $\rho_{tvf} = \dfrac{sample.time_{tvf}}{export.time_{tvf}}$ | Salvage length frequency $(m_{FF})$ <br> $m_{FF_{tvl=6}} \sim \text{Binomial}(\sum_{k=1}^{6} m_{FF_{tvk}}, \omega_{tv})$ <br><br> Raw salvage count $(y)$ <br> $y_{tvf} \sim \text{Poisson}(\lambda_{tvf})$ | 63 <br><br> + | 63 <br><br> 126 |
| | | | | 217 | 1,722 |

**Figure 2.**

**Figure 3.**

**Figure 4.**

**Figure 5.**

## Appendix A

**Table A1.** Particle tracking data, for the Prisoners Point insertion region, informing sub-process 1 (Transport). Data are represented here as proportions reaching each sink ($\frac{m_{PT}}{N_{PT}}$) from the Prisoners Point insertion region. Proportions are organized by year (rows), month (column subheading) and sink (column main heading).

| | State Water Project | | | Central Valley Project | | | Low Risk Zone | | | Indirect Risk Zone | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | April | May | June | April | May | June | April | May | June | April | May | June |
| 1995 | 0.042 | 0.323 | 0.333 | 0.055 | 0.222 | 0.210 | 0.835 | 0.364 | 0.315 | 0.068 | 0.091 | 0.142 |
| 1996 | 0.001 | 0.000 | 0.096 | 0.000 | 0.000 | 0.056 | 0.992 | 0.972 | 0.797 | 0.114 | 0.050 | 0.050 |
| 1997 | 0.329 | 0.079 | 0.007 | 0.279 | 0.352 | 0.369 | 0.284 | 0.394 | 0.418 | 0.136 | 0.071 | 0.100 |
| 1998 | 0.000 | 0.000 | 0.002 | 0.000 | 0.000 | 0.003 | 1.000 | 0.999 | 0.938 | 0.000 | 0.000 | 0.000 |
| 1999 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 1.000 | 1.000 | 0.100 | 0.036 | 0.021 |
| 2000 | 0.265 | 0.307 | 0.303 | 0.181 | 0.217 | 0.202 | 0.467 | 0.375 | 0.388 | 0.107 | 0.079 | 0.100 |
| 2001 | 0.278 | 0.307 | 0.338 | 0.285 | 0.249 | 0.267 | 0.344 | 0.330 | 0.262 | 0.307 | 0.164 | 0.257 |
| 2002 | 0.000 | 0.043 | 0.115 | 0.000 | 0.012 | 0.062 | 1.000 | 0.904 | 0.732 | 0.271 | 0.214 | 0.171 |
| 2003 | 0.219 | 0.222 | 0.223 | 0.296 | 0.268 | 0.283 | 0.347 | 0.311 | 0.323 | 0.186 | 0.114 | 0.043 |
| 2004 | 0.369 | 0.456 | 0.265 | 0.278 | 0.258 | 0.286 | 0.295 | 0.174 | 0.294 | 0.186 | 0.164 | 0.007 |
| 2005 | 0.572 | 0.303 | 0.226 | 0.282 | 0.191 | 0.296 | 0.107 | 0.362 | 0.323 | 0.171 | 0.000 | 0.029 |
| 2006 | 0.505 | 0.005 | 0.012 | 0.335 | 0.185 | 0.330 | 0.112 | 0.522 | 0.387 | 0.000 | 0.000 | 0.071 |
| 2007 | 0.516 | 0.066 | 0.043 | 0.293 | 0.138 | 0.079 | 0.104 | 0.441 | 0.516 | 0.214 | 0.157 | 0.007 |
| 2008 | 0.029 | 0.014 | 0.036 | 0.117 | 0.077 | 0.069 | 0.406 | 0.494 | 0.541 | 0.393 | 0.021 | 0.121 |
| 2009 | 0.164 | 0.019 | 0.108 | 0.105 | 0.016 | 0.078 | 0.650 | 0.895 | 0.705 | 0.436 | 0.171 | 0.157 |
| 2010 | 0.001 | 0.014 | 0.003 | 0.032 | 0.024 | 0.099 | 0.570 | 0.634 | 0.497 | 0.114 | 0.179 | 0.107 |
| 2011 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 | 1.000 | 1.000 | 0.998 | 0.000 | 0.000 | 0.014 |
| 2012 | 0.015 | 0.004 | 0.399 | 0.006 | 0.001 | 0.242 | 0.949 | 0.982 | 0.308 | 0.121 | 0.236 | 0.257 |
| 2013 | 0.155 | 0.039 | 0.178 | 0.233 | 0.047 | 0.349 | 0.474 | 0.775 | 0.371 | 0.336 | 0.393 | 0.300 |
| 2014 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 1.000 | 0.999 | 0.307 | 0.471 | 0.007 |
| 2015 | 0.089 | 0.003 | 0.038 | 0.022 | 0.006 | 0.146 | 0.792 | 0.939 | 0.737 | 0.536 | 0.493 | 0.021 |

**Table A2.** Particle tracking data, for the Lower Old and Middle river insertion region, informing sub-process 1 (Transport). Data are represented here as proportions reaching each sink ($\frac{m_{PT}}{N_{PT}}$) from the Lower Old and Middle river insertion region. Proportions are organized by year (rows), month (column subheading) and sink (column main heading).

| | State Water Project | | | Central Valley Project | | | Low Risk Zone | | | Indirect Risk Zone | | |
|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | April | May | June | April | May | June | April | May | June | April | May | June |
| 1995 | 0.257 | 0.496 | 0.483 | 0.254 | 0.376 | 0.353 | 0.433 | 0.122 | 0.157 | 0.057 | 0.007 | 0.007 |
| 1996 | 0.008 | 0.001 | 0.430 | 0.015 | 0.001 | 0.177 | 0.813 | 0.775 | 0.369 | 0.114 | 0.050 | 0.050 |
| 1997 | 0.608 | 0.312 | 0.033 | 0.251 | 0.501 | 0.616 | 0.123 | 0.173 | 0.321 | 0.136 | 0.071 | 0.100 |
| 1998 | 0.162 | 0.000 | 0.067 | 0.000 | 0.000 | 0.085 | 0.825 | 0.983 | 0.680 | 0.000 | 0.000 | 0.000 |
| 1999 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 1.000 | 1.000 | 0.100 | 0.036 | 0.021 |
| 2000 | 0.504 | 0.457 | 0.452 | 0.373 | 0.411 | 0.339 | 0.115 | 0.127 | 0.196 | 0.107 | 0.079 | 0.100 |
| 2001 | 0.510 | 0.538 | 0.448 | 0.390 | 0.303 | 0.364 | 0.092 | 0.147 | 0.177 | 0.307 | 0.164 | 0.257 |
| 2002 | 0.000 | 0.375 | 0.492 | 0.000 | 0.075 | 0.193 | 0.998 | 0.488 | 0.288 | 0.271 | 0.214 | 0.171 |
| 2003 | 0.411 | 0.406 | 0.271 | 0.476 | 0.410 | 0.454 | 0.103 | 0.168 | 0.260 | 0.186 | 0.114 | 0.043 |
| 2004 | 0.486 | 0.554 | 0.392 | 0.424 | 0.344 | 0.408 | 0.086 | 0.094 | 0.183 | 0.186 | 0.164 | 0.007 |
| 2005 | 0.642 | 0.401 | 0.367 | 0.310 | 0.356 | 0.422 | 0.046 | 0.221 | 0.196 | 0.171 | 0.000 | 0.029 |
| 2006 | 0.588 | 0.040 | 0.033 | 0.356 | 0.550 | 0.677 | 0.053 | 0.318 | 0.257 | 0.000 | 0.000 | 0.071 |
| 2007 | 0.635 | 0.348 | 0.189 | 0.319 | 0.292 | 0.284 | 0.042 | 0.288 | 0.439 | 0.214 | 0.157 | 0.007 |
| 2008 | 0.331 | 0.137 | 0.212 | 0.506 | 0.303 | 0.177 | 0.114 | 0.461 | 0.490 | 0.393 | 0.021 | 0.121 |
| 2009 | 0.518 | 0.124 | 0.443 | 0.341 | 0.087 | 0.238 | 0.119 | 0.616 | 0.278 | 0.436 | 0.171 | 0.157 |
| 2010 | 0.022 | 0.126 | 0.035 | 0.305 | 0.250 | 0.334 | 0.466 | 0.436 | 0.503 | 0.114 | 0.179 | 0.107 |
| 2011 | 0.000 | 0.000 | 0.026 | 0.000 | 0.000 | 0.023 | 0.997 | 0.997 | 0.905 | 0.000 | 0.000 | 0.014 |
| 2012 | 0.252 | 0.046 | 0.611 | 0.130 | 0.020 | 0.294 | 0.510 | 0.828 | 0.092 | 0.121 | 0.236 | 0.257 |
| 2013 | 0.386 | 0.238 | 0.266 | 0.446 | 0.260 | 0.573 | 0.147 | 0.438 | 0.154 | 0.336 | 0.393 | 0.300 |
| 2014 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 0.987 | 0.977 | 0.307 | 0.471 | 0.007 |
| 2015 | 0.613 | 0.053 | 0.130 | 0.113 | 0.063 | 0.520 | 0.203 | 0.648 | 0.326 | 0.536 | 0.493 | 0.021 |

**Table A3.** Particle tracking data, for the Upper Old and Middle river insertion region, informing sub-process 1 (Transport). Data are represented here as proportions reaching each sink ($\frac{m_{PT}}{N_{PT}}$) from the Upper Old and Middle river insertion region. Proportions are organized by year (rows), month (column subheading) and sink (column main heading).

| | State Water Project | | | Central Valley Project | | | Low Risk Zone | | | Indirect Risk Zone | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | April | May | June | April | May | June | April | May | June | April | May | June |
| 1995 | 0.381 | 0.507 | 0.570 | 0.363 | 0.445 | 0.351 | 0.250 | 0.048 | 0.080 | 0.007 | 0.000 | 0.000 |
| 1996 | 0.085 | 0.036 | 0.671 | 0.077 | 0.012 | 0.181 | 0.752 | 0.832 | 0.147 | 0.114 | 0.050 | 0.050 |
| 1997 | 0.886 | 0.569 | 0.103 | 0.070 | 0.368 | 0.682 | 0.041 | 0.063 | 0.215 | 0.136 | 0.071 | 0.100 |
| 1998 | 0.908 | 0.000 | 0.577 | 0.034 | 0.000 | 0.194 | 0.058 | 0.999 | 0.223 | 0.000 | 0.000 | 0.000 |
| 1999 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 1.000 | 1.000 | 0.100 | 0.036 | 0.021 |
| 2000 | 0.543 | 0.567 | 0.493 | 0.427 | 0.377 | 0.408 | 0.028 | 0.056 | 0.098 | 0.107 | 0.079 | 0.100 |
| 2001 | 0.525 | 0.510 | 0.475 | 0.450 | 0.421 | 0.438 | 0.023 | 0.069 | 0.087 | 0.307 | 0.164 | 0.257 |
| 2002 | 0.000 | 0.786 | 0.629 | 0.000 | 0.067 | 0.279 | 1.000 | 0.142 | 0.091 | 0.271 | 0.214 | 0.171 |
| 2003 | 0.447 | 0.515 | 0.353 | 0.514 | 0.408 | 0.499 | 0.038 | 0.077 | 0.147 | 0.186 | 0.114 | 0.043 |
| 2004 | 0.423 | 0.521 | 0.506 | 0.556 | 0.433 | 0.384 | 0.019 | 0.044 | 0.109 | 0.186 | 0.164 | 0.007 |
| 2005 | 0.670 | 0.331 | 0.325 | 0.318 | 0.581 | 0.573 | 0.012 | 0.084 | 0.101 | 0.171 | 0.000 | 0.029 |
| 2006 | 0.638 | 0.072 | 0.013 | 0.349 | 0.850 | 0.850 | 0.014 | 0.071 | 0.137 | 0.000 | 0.000 | 0.071 |
| 2007 | 0.694 | 0.698 | 0.358 | 0.297 | 0.145 | 0.381 | 0.010 | 0.152 | 0.257 | 0.214 | 0.157 | 0.007 |
| 2008 | 0.625 | 0.345 | 0.299 | 0.357 | 0.346 | 0.337 | 0.017 | 0.302 | 0.357 | 0.393 | 0.021 | 0.121 |
| 2009 | 0.587 | 0.279 | 0.607 | 0.396 | 0.171 | 0.293 | 0.014 | 0.519 | 0.096 | 0.436 | 0.171 | 0.157 |
| 2010 | 0.049 | 0.223 | 0.171 | 0.803 | 0.580 | 0.400 | 0.134 | 0.187 | 0.422 | 0.114 | 0.179 | 0.107 |
| 2011 | 0.000 | 0.000 | 0.077 | 0.000 | 0.000 | 0.056 | 1.000 | 1.000 | 0.827 | 0.000 | 0.000 | 0.014 |
| 2012 | 0.607 | 0.126 | 0.672 | 0.279 | 0.034 | 0.295 | 0.109 | 0.784 | 0.033 | 0.121 | 0.236 | 0.257 |
| 2013 | 0.379 | 0.459 | 0.366 | 0.564 | 0.307 | 0.562 | 0.053 | 0.231 | 0.071 | 0.336 | 0.393 | 0.300 |
| 2014 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 1.000 | 1.000 | 0.307 | 0.471 | 0.007 |
| 2015 | 0.857 | 0.337 | 0.174 | 0.097 | 0.153 | 0.648 | 0.038 | 0.479 | 0.177 | 0.536 | 0.493 | 0.021 |

**Table A4.** Mark-recapture data reported by Castillo et al. (2012) and used to model maximum sampling efficiency (Eq. B13) at the Skinner Fish Facility. $N_{EF}$ delta smelt were tagged and released in each trial (rows), and $r$ were recovered.

| $N_{EF}$ | $r$ |
|---|---|
| 100 | 39 |
| 100 | 36 |
| 100 | 89 |
| 100 | 49 |
| 100 | 43 |
| 100 | 45 |

**Table A5.** Mark-recapture data reported by Sutphin and Svoboda (2016) and used to model maximum sampling efficiency (Eq. B13) at the Tracy Fish Facility. $N_{EF}$ delta smelt were tagged and released in each trial (rows), and $r$ were recovered.

| $N_{EF}$ | $r$ | $N_{EF}$ | $r$ | $N_{EF}$ | $r$ |
|---|---|---|---|---|---|
| 180 | 20 | 100 | 35 | 75 | 31 |
| 179 | 21 | 100 | 28 | 75 | 31 |
| 179 | 52 | 100 | 35 | 130 | 38 |
| 179 | 55 | 100 | 11 | 130 | 70 |
| 180 | 27 | 100 | 6 | 130 | 56 |
| 180 | 29 | 100 | 23 | 130 | 72 |
| 176 | 19 | 75 | 39 | 130 | 36 |
| 178 | 27 | 75 | 28 | 130 | 17 |
| 180 | 16 | 75 | 29 | 130 | 14 |
| 180 | 23 | 75 | 25 | 130 | 36 |
| 400 | 124 | 75 | 24 | 130 | 0 |
| 80 | 26 | 75 | 28 | 130 | 0 |
| 80 | 21 | 75 | 24 | 130 | 1 |
| 80 | 13 | 75 | 26 | 100 | 0 |
| 80 | 38 | 75 | 28 | 100 | 1 |
| 80 | 12 | 75 | 27 | 100 | 0 |
| 78 | 8 | 75 | 24 | 100 | 5 |
| 76 | 9 | 75 | 28 | 100 | 3 |
| 80 | 13 | 75 | 17 | 100 | 4 |
| 80 | 5 | 75 | 27 | 100 | 18 |
| 80 | 7 | 75 | 19 | 100 | 10 |
| 80 | 4 | 75 | 45 | 100 | 16 |
| 79 | 10 | | | | |

**Appendix B**

**Mathematical model**

*Sub-process 1: Transport of fish to facilities.* We modeled transport probabilities $p_{\text{TR}_{qtvs}}$ from source region $q$ (Fig. 1) to sink $s$ in year $t$ and month $v$ (Table 1) using a Dirichlet distribution with parameters $\alpha_{\text{TR}_{qtvs}}$. Hereafter, we denote vectors in bold and the full set of an indexed value with a period. For $s$ in the set (1, 2, 3, 4 [State Water Project {SWP}, Central Valley Project {CVP}, Low Risk Zone {LRZ}, Intermediate Risk Zone {IRZ}]), let the vector

$$(\text{B1}) \quad \mathbf{p}_{\text{TR}_{qtv.}} \sim \text{Dirichlet}(\boldsymbol{\alpha}_{\text{TR}_{qtv.}}).$$

Both particle tracking data and fish facility fish count data were available to inform time-specific estimates of the parameters $\alpha_{\text{TR}_{qtv1}}$ and $\alpha_{\text{TR}_{qtv2}}$ (SWP and CVP); however, $\alpha_{\text{TR}_{qtv3}}$ and $\alpha_{\text{TR}_{qtv4}}$ (LRZ and IRZ) were informed by particle tracking data only. Hence, transport to LRZ (s=3) was modeled as a log-linear function of hydrodynamic covariate *OMR* and month-specific linear regression parameters $A_{\text{TR}}$

$$(\text{B2}) \quad \log(\alpha_{\text{TR}_{qtv3}}) = A_{\text{TR}_{0qv3}} + A_{\text{TR}_{1qv3}} OMR_{\text{tv}},$$

and transport to IRZ (s=4) was modeled as a log-linear function of year,

$$(\text{B3}) \quad \log(\alpha_{\text{TR}_{qtv4}}) = A_{\text{TR}_{0qv4}} + A_{\text{TR}_{1qv4}} t.$$

Intercept and slope $A_{\text{TR}}$ were assumed to be similar among months and source regions, for the same sink; therefore, all intercept and slope $A_{\text{TR}}$ were modeled as normally distributed random effects of a single sink-specific intercept and slope.

Let $m_{\text{PT}_{qtvs}}$ represent the number of particles in the particle tracking model that were released in source region $q$ in year $t$ and month $v$ that reached sink $s$ by the end of the month. We assumed these counts followed a multinomial distribution with total effective number of particles $N_{\text{PT}}$ and probability vector $\mathbf{p}_{\text{TR}_{qtv.}}$

(B4)  $\mathbf{m}_{\text{PT}_{qtv.}} \sim \text{Multinomial}(N_{\text{PT}}, \mathbf{p}_{\text{TR}_{qtv.}})$.

The number of particles released in a source region in the particle tracking model (4,000) was an arbitrarily large number. As multinomial variance scales with sample size, the effect of particle tracking data on parameter estimates could be made arbitrarily large. To remove this arbitrary consequence, we derived a single effective multinomial sample size $N_{\text{PT}}$ using the method described by McAllister and Ianelli (1997).

*Sub-process 2: Survival.* Using OMR flow divided by mean Old and Middle River cross sectional area between the mouths of Old and Middle rivers and the SWP and CVP, as an estimate of mean monthly water velocity, and the distance between the midpoint of each source region and the SWP and CVP, year-month-source specific transport times (days) $T_{\text{TR}_{qtv}}$ were calculated as distance divided by velocity. Zero or downstream OMR flow resulted in infinite transport time using this method, so $T_{\text{TR}}$ was truncated at 15 days to represent mortality at the midpoint of one month. Daily mortality of 0.006 was estimated for larval through subadult lifestages of delta smelt by Bennett (2005) using daily otolith ages and catch

curve analysis. To develop survival probabilities for each year-month combination, transport times were multiplied by the Bennett (2005) daily mortality rate and converted to survival rates $p_{SV1_{qtv}}$ (Fig. D1)

$$(B5) \quad p_{SV1_{qtv}} = e^{-0.006T_{TR_{qtv}}}.$$

$p_{SV1_{qtv}}$ for the source region selected as the source for a year-month was applied to all fish arriving at the SWP and CVP.

The probability of a delta smelt surviving the journey from the Forebay radial gates (where the forebay connects to Old River) to the SWP intake channel in a given year and month $p_{SV2_{tv1}}$ was treated as a logit-normal random variable:

$$(B6) \quad \text{logit}(p_{SV2_{tv1}}) \sim \text{Normal}(\beta_0 + \beta_1 v, \sigma_{SV2});$$

thus, the survival probability increased or decreased with time (and age). CVP does not have a forebay, so set $p_{SV2_{tv2}}$ was set to 1.

*Sub-process 3: Population length structure.* Let $p_{LN_{tvl}}$ represent the population length frequency for length class *l*. We modeled these quantities using a Dirichlet random variable with month-length specific parameters $\alpha_{LN_{vl}}$. For *l* in the set (1, 2, … 5), let the vector

$$(B7) \quad \mathbf{p}_{LN_{tv.}} \sim \text{Dirichlet}(\boldsymbol{\alpha}_{LN_{v.}}).$$

By using this structure, we assumed that length distributions within the same month were similar across years.

We modeled the vector of adjusted counts $\mathbf{m}_{20\mathrm{MM}_{\mathrm{tv.}}}$ with a multinomial distribution

(B8)     $\mathbf{m}_{20\mathrm{MM}_{\mathrm{tv.}}} \sim \mathrm{Multinomial}(N_{20\mathrm{MM}_{\mathrm{tv}}}, \mathbf{p}_{\mathrm{LN}_{\mathrm{tv.}}})$.

We adjusted the length data from the 20mm Survey to account for gear selectivity (Mitchell et al. 2019). Gear selectivity uncertainty was found to be a relatively minor source of variability in 20mm Survey data compared to other sources of variability (Polansky et al. 2019), so it was ignored for this analysis. We made these adjustments by dividing observed delta smelt counts in each of the five length bins by estimates of 20mm Survey gear selectivity at the midpoints of each bin (and rounding to the nearest integer value), resulting in estimates of the number of fish that would have been caught if the gear could catch all available fish with 100% efficiency.

 For the effective sample sizes $N_{20\mathrm{MM}_{\mathrm{tv}}}$ we used the total number of tows, conducted by the 20mm Survey in which the length data were collected. We scaled the adjusted counts so they summed to $N_{20\mathrm{MM}_{\mathrm{tv.}}}$. Length samples of fish populations may exhibit overdispersion due to the tendency of fish to aggregate by length or age. The standard approach to address this overdispersion when estimating population proportion at length, from multinomially distributed length samples, is to set the effective multinomial sample size to the number of tows (McAllister and Ianelli 1997). Note that this approach was different from the approach used to derive $N_{\mathrm{PT}}$, for which information such as number of tows does not exist.

*Sub-process 4: Sampling efficiency at fish facilities.* We first estimated selectivity probability-at-length $p_{\mathrm{SEL}_l}$ with fish facility length data using a logistic function with an asymptotic maximum value of one. Given that the louvers are similar for the SWP and CVP, we used the same length-based louver selectivity model for both facilities, but estimated facility-specific maximum efficiencies. Let $m_{\mathrm{FF}_{\mathrm{tvl}}}$ represent the

number of delta smelt in length class $l$ that were diverted and measured at fish facilities. We assumed

these counts followed a multinomial distribution with effective sample size $N_{\mathrm{FF}_{tv}}$ and probability vector

$$\mathbf{p}_{\mathrm{BYP}_{tv.}}$$

(B9)   $\mathbf{m}_{\mathrm{FF}_{tv.}} \sim \mathrm{Multinomial}\big(N_{\mathrm{FF}_{tv}},\mathbf{p}_{\mathrm{BYP}_{tv.}}\big).$

We set the effective sample size $N_{\mathrm{FF}_{tv}}$ equal to the total number of days (in year $t$ and month $v$) on which

fish were counted and measured, summed across the two fish facilities. $m_{\mathrm{FF}_{tvl}}$ were then rescaled to sum

to $N_{\mathrm{FF}_{tv}}$. The probabilities $p_{\mathrm{BYP}_{tvl}}$ represent the length frequency distribution of fish diverted to the holding

tanks. We calculated these probabilities as the scaled product of the population length structure $p_{\mathrm{LN}_{tvl}}$ and

the length-based louver selectivity $p_{\mathrm{SEL}_l}$:

(B10)  $p_{\mathrm{BYP}_{tvl}} = p_{\mathrm{LN}_{tvl}} p_{\mathrm{SEL}_l} \Big/ \sum_{k=1}^{5} p_{\mathrm{LN}_{tvk}} p_{\mathrm{SEL}_k}.$

We used a logistic function to describe louver selectivity:

(B11)  $p_{\mathrm{SEL}_l} = 1 \Big/ \big(1 + e^{-\eta_0(l-\eta_1)}\big),$

where $l$ takes the values 1, 2, 3, 4, and 5 (representing the five length bins), $\eta_0$ controls the steepness of

the curve, and $\eta_1$ is the $l$ value at which $p_{\mathrm{SEL}_l} = 0.5$. We assumed fish in the fifth length bin had the

highest selectivity probability relative to the other bins (i.e., $p_{\mathrm{SEL}_5} = 0.999$), which allowed us to solve for

$\eta_0$ analytically: $\eta_0 = \dfrac{-\ln\left(\frac{0.001}{0.999}\right)}{(5-\eta_1)}(l-\eta_1)$. The parameter $\eta_1$, was estimated during model fitting.

We defined the length-specific probabilities of fish being diverted to the fish facilities, or overall sampling efficiency, $p_{\text{EF}_{tvfl}}$ of fish facility $f$ for diverting length-$l$ delta smelt as

$$(\text{B12}) \quad p_{\text{EF}_{tvfl}} = \frac{1}{\left(1 + e^{-\gamma_{ftv}}\right)} p_{\text{SEL}_l},$$

where the fraction on the right hand side is a facility-specific scaling factor between 0 and 1 and $\gamma$ were maximum sampling efficiencies. We estimated year-month-facility specific $\gamma_{ftv}$ using a Bayesian model fit to the mark-recapture experiment data sets. These models were fit externally to the hierarchical model described in this article. Note that $\gamma$ for mark-recapture experiments were indexed by trial number, but $\gamma$ for fish facility counts (Eq. 12) were indexed by year and month. Let $N_{\text{EF}_{fi}}$ be the number of delta smelt tagged in experiment number $i$ at fish facility $f$ and let $r_{fi}$ be the number of fish recovered in the holding tanks. We assumed $r_{fi}$ followed a binomial distribution,

$$(\text{B13}) \quad r_{fi} \sim \text{Binomial}\left(N_{\text{EF}_{fi}}, \frac{1}{\left(1 + e^{-\gamma_{fi}}\right)}\right),$$

with $N_{\text{EF}_{fi}}$ trials and success probability $\frac{1}{\left(1 + e^{-\gamma_{fi}}\right)}$. Each $\gamma_{fi}$ (and $\gamma_{ftv}$ for year-months of entrainment) was normally distributed with facility-specific mean $\Gamma_f$ and standard deviation $\sigma_{\text{EF}_f}$

$$(\text{B14}) \quad \gamma_{fi} \sim \text{Normal}\left(\Gamma_f, \sigma_{\text{EF}_f}\right).$$

Because adult delta smelt were used in the experiments, we assumed relative size selectivity (i.e., $p_{\text{SEL}_l}$) was one for all tagged fish.

*Sub-process 5: Subsampling at fish facilities.* We estimated the probability $\omega_{tv}$ with count-at-length data from the fish facilities. Using the same five length bins as in the population length structure sub-process, and defining a sixth bin to represent 45+ mm delta smelt, we modeled the number of delta smelt in the sixth bin $m_{FF_{tv6}}$ using a binomial distribution:

$$\text{(B15)} \quad m_{FF_{tv6}} \sim \text{Binomial}\left(\sum_{k=1}^{6} m_{FF_{tvk}}, \omega_{tv}\right).$$

*Number of fish at entrained*: We used estimates of postlarval delta smelt abundance from the entire Delta for May and June (Polansky et al. 2019) to develop weakly informative lognormal priors for the number of delta smelt passing through the source region $n_{S_{tv}}$ with support, defined as the 99% prior predictive interval, from approximately 0.002 to 4.5 times abundance. Because these abundance estimates were only available for the months of May and June, we set an upper bound for April equal to either the maximum of May abundance or a retrospective extrapolation of May to June abundance. This accounted for periodic years of early recruitment, as evident by declining May to June abundance when April postlarval abundance may have been higher than May.

**Simulation model and data**

In the first set of simulations, independence was assumed among the individual fish comprising $n_S$, and the numbers ending up in each of the four sinks $y_s$ were a multinomial random vector with probability $p_s$, where $s$ was in the set (1, 2, 3, 4 [SWP, CVP, LRZ, IRZ]). Given that $y_3$ and $y_4$ (LRZ and IRZ) were unobserved, the distribution for $y_1$ and $y_2$ (SWP and CVP) alone was a simpler multinomial

$$\text{(B16)} \quad \mathbf{y}. \sim \text{Multinomial}(n_S, \mathbf{p}.), \text{ where } s \in (1, 2).$$

Given known values for $p_1$ and $p_2$, $n_S$ is clearly estimable, e.g., $y_1/p_1$ is a method of moments estimate. Maximum likelihood estimates (MLEs) of $n_S$, which are more efficient than the Bayesian hierarchical model, can be calculated by a grid search for $n_S$ over the integer set $(y_1 + y_2, y_1 + y_2 + 1, \ldots)$. A simulation study was carried out to examine the quality of the MLEs. For each of eight combinations of values of $n_S$ (500 or 1,000), $p_1$ (0.1 or 0.4), and $p_2$ (0.05 or 0.30), 1,000 simulations of $y_1$ and $y_2$ were generated from the simpler multinomial model and MLEs for $n_S$ were calculated and compared to true values.

A second simulation study examined estimability of key parameters under a much more complex Bayesian hierarchical model that closely paralleled the one applied to the real data and integrated data from multiple sources. The hierarchical model described above was used to define an operating model, based loosely on the parameters estimated for delta smelt. New particle tracking, length composition, and fish facility sample data were stochastically simulated to generate 500 replicate datasets. Each model parameter ($n_S$, $\alpha_{TR}$, $\alpha_{LN}$, $\beta$, $\eta$, and $\Gamma$) was drawn from a uniform distribution to simulate a set of true state dynamics (Eqs. B1, B6, B7, and B13). From the set of true states, observed data were then simulated based on the models for observed values (Eqs. B4, B8, B9, B13, B15, and 1). The estimation model was the same as the operating model; it would therefore fit the simulated data perfectly in the absence of observation error, if all parameters were estimable. The model was then fit to each simulated dataset. The estimability of each parameter was evaluated using three metrics that compared estimated to true values (Betancourt 2018): 95% credible interval coverage of the true value, z-scores indicating deviations between true simulated and estimated values,

$$(B17) \quad z_{tv} = \big(\text{mean}(\text{true } n_{S_{tv}}) - \text{mean}(\text{estimated } n_{S_{tv}})\big)/\text{sd}(\text{true } n_{S_{tv}}),$$

and posterior contraction comparing the prior to posterior standard deviation

$$(B18) \quad contraction_{tv} = 1 - \mathrm{sd}(\text{posterior } n_{S_{tv}})/\mathrm{sd}(\text{prior } n_{S_{tv}}).$$

Additional simulations focused on the effect of errors in the particle tracking model based probabilities of movement from the source region to the four sinks on posterior distributions of key parameters. Particle tracking data were assumed to represent observations of the transport process, with some unknown level of additional noise induced by the disconnect between simulated hydrodynamics and true delta smelt transport rates. To assess model performance with noisy transport information, two sets of simulations were performed, one set in which new particle tracking data were simulated from true transport rates and one set in which particle tracking data were simulated from true transport parameters, times random lognormal error. That is, Eq. 1 was modified to

$$(B19) \quad \mathbf{p}_{TR_{qtv.}} \sim \mathrm{Dirichlet}\big(\boldsymbol{\alpha}_{TR_{qtv.}} \times e^{\boldsymbol{error}_{tv.}}\big),$$

where *error* was normally distributed with mean 0 and standard deviation 0.3.

**Model fitting and diagnostics**

While a weakly informative uniform prior distribution was derived for $n_{source}$, uninformative priors were used for all other parameters. Dirichlet parameters $\alpha$ were given Gamma(0.01,0.01) priors, and all logistic regression priors $\beta$, $\eta$, and $\gamma$ were drawn from Normal(0,$\sqrt{\pi^2/6}$) distributions. A Normal(0,$\sqrt{\pi^2/6}$) distribution induced the desired Uniform(0,1) distribution on inverse logistic transformation. Hyper-parameters for the linear models of LRZ and IRZ transport probability (Eq. C1 and C2) were assigned Normal(0, $\sqrt{20}$) priors. Penalized complexity prior distributions, Exponential(4.6), were assigned to all

standard error parameters $\sigma$ (Simpson et al. 2017), assuming a prior probability of 0.01 that $\sigma$ exceeded a value of 1.

The model was fit using R package R2jags (Su and Yajima 2015) and Markov Chain Monte Carlo program JAGS (Plummer 2003). R code to run the model is in Appendix C. A burn-in period of 50,000 was followed by 300,000 samples of posterior distributions among six chains. As preliminary analysis suggested autocorrelation within posterior chains, posterior samples were thinned by 25. Model convergence was assessed by comparing the trace plots of six chains of each model parameter and using Gelman and Rubin's diagnostic (Gelman and Rubin 1992). Model convergence was assumed if trace plots showed that all chains were producing samples of parameters with similar posterior distributions that did not shift with additional samples and if Gelman and Rubin's statistic was less than 1.05 for all parameters. Correlation between parameters was assessed by calculating coefficients of determination ($R^2$) for all pairwise combinations of parameter posteriors. Residual plots of the fish facility counts, population length structure, and fish facility sampling efficiency models were explored for evidence of model fit and bias.

A graphical posterior predictive check was performed and Bayesian P-values were calculated for each data component ($PT$, $m_{20mm}$, $m_{FF}$, and $obs$). Each data component was replicated from the fitted model, and Bayesian P-values were based on a comparison of goodness of fit to replicated and observed data. Replicated and observed goodness of fit were compared using the Freeman-Tukey statistic $FT$, calculated for each observed and replicated data point ($FT = \sum(\sqrt{observed} - \sqrt{expected})^2$; Cressie and Read 1984). $FT$ were summed over dimensions reflecting the hierarchical model structure. $FT_{PT}$ and $FT_{obs}$ were summarized for each fish facility-month combination by summing over years. $FT_{20mm}$ and $FT_{FF}$ were summarized for each length-month combination by summing over years. Graphical analyses were performed by comparing $FT_{replicated}$ and $FT_{observed}$ from all 300,000 Monte Carlo simulations, and P-values were calculated as the number of joint posterior samples of $FT$ where $FT_{replicated} \geq FT_{observed}$, divided

by the total number of samples. When the model is sufficient to replicate observed data, Bayesian P-values are near 0.5. P-values greater than 0.95 and less than 0.05 demonstrated a low probability that the model can reproduce observed data and were considered an indication of poor model fit.

## References

Bennett, W.A. 2005. Critical assessment of the delta smelt population in the San Francisco Estuary, California. San Francisco Estuary Watershed Sci, 3. Available from: https://escholarship.org/uc/item/0725n5vk.

Betancourt, M. 2018. Towards a principled Bayesian workflow. Available from: https://betanalpha.github.io/assets/case_studies/principled_bayesian_workflow.html

Cressie, N., and Read, T.R. 1984. Multinomial goodness-of-fit tests. J. R. Stat. Soc. Series B Stat. Methodol. **46**(3): 440–464.

Gelman, A. and Rubin, D.B. 1992. Inference from iterative simulation using multiple sequences. Stat. Sci. **7**: 457–511.

McAllister, M.K. and Ianelli, J.N. 1997. Bayesian stock assessment using catch-age data and the sampling-importance resampling algorithm. Can. J. Fish. Aquat. Sci. **54**(2): 284–300.

Mitchell, L., Newman, K., and Baxter. R. 2019. Estimating the size selectivity of fishing trawls for short-lived fish species. San Francisco Estuary and Watershed Sci. **17**(1). Available from: https://doi.org/10.15447/sfews.2018v17iss1art5.

Plummer, M. 2003. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. Proceedings of the 3rd International Workshop on Distributed Statistical Computing, Vienna, Austria.

Polansky, L., L. Mitchell, and K. Newman. 2019. Using multistage design-based methods to construct abundance indices and uncertainty measures for Delta Smelt. Trans. Am. Fish. Soc. **148**(4): 710–724. doi: 10.1002/tafs.10166.

Simpson, D., Rue, H., Riebler, A., Martins, T.G., and Sørbye, S.H. 2017. Penalising model component complexity: A principled, practical approach to constructing priors. Stat. Sci. **32**: 1–28.

Su, Y.S. and Yajima, M. 2015. R2jags: Using R to Run `JAGS'. Version 0.5-7. https://CRAN.R-project.org/package=R2jags.

Canadian Journal of Fisheries and Aquatic Sciences

## Appendix B

JAGS code to fit model

```
### Calculate survival of transport (p.SV1) outside of BHM ###
distance=c(NA,33070,25360,16625)
survival.transport=matrix(NA,Y,M)
days=matrix(NA,Y,M)
for (y in startY:Y) {
  for (v in 1:M) {
  if (OMR[t,v] < 0) {
     #T.TR=distance/flow/area/feet/meter/seconds/day
     T.TR[t,v] <- (distance[PTM.map[t,v]]/(abs(OMR[t,v])/7659.96)/0.3048)/(3600*24)
     p.SV1[t,v] <- max(exp(-0.006*15),exp(-0.006*T.TR[y,v]))
     }
  else {
     p.SV1[t,v] <- exp(-0.006*15)
     }
     }
   }

### JAGS model ###
model {

### Priors ###
# Latent states #
for (t in startY:Y) {
  for (v in 1:M) {
     n.source[t,v] ~ dlnorm(log(0.01*nhat[(t-16),v]),(1/log(10))
     }
  }

# Sub-process 1: Transport #
sigma.TR ~ dexp(4.6)
tau.TR <- pow(sigma.TR,-2)
for (i in 3:4) {
  hyper.A.TR[i,1] ~ dnorm(0,0.05)
  hyper.A.TR[i,2] ~ dnorm(0,0.05)
  }
for (q in 1:3) {
for (v in 1:M) {
  for (t in startY:Y) {
     for (i in 1:2) {
        log.alpha.TR[q,t,v,i] ~ dnorm(0,0.05)
        alpha.TR[q,t,v,i] <- exp(log.alpha.TR[q,t,v,i])
        }
     }
  for (i in 3:4) {
     for (j in 1:2) {
        A.random[q,v,i,j] ~ dnorm(0,tau.TR)
        }
     A.TR[q,v,i,1] <- hyper.A.TR[i,1] + A.random[q,v,i,1]
     A.TR[q,v,i,2] <- hyper.A.TR[i,2] + A.random[q,v,i,2]
     }
  for (t in startY:Y) {
     alpha.TR[q,t,v,3] <- A.TR[q,v,3,1]+A.TR[q,v,3,2]*OMR[t,v]
```

```
      alpha.TR[q,t,v,4] <- A.TR[q,v,4,1]+A.TR[q,v,4,2]* (((t-startY+1)-11)/6.2)
      }
  }

# Sub-process 2: Survival #
beta[1] ~ dnorm(0,0.61)
beta[2] ~ dnorm(0,0.61)

# Sub-process 3: Population length structure #
for (l in 1:5) { # population length structure
  for (v in 1:M) {
    alpha.LN[v,l] ~ dgamma(0.001,0.001)
    }
  }

# Sub-process 4: Sampling efficiency
eta[1] <- -log(0.001/0.999)/(5-eta[2]) # selectivity
eta[2] ~ dnorm(0,0.30)
G[1] <- 0.067 # max efficiency
G[2] <- -1.586
sigma.EF[1] <- 1.102
sigma.EF[2] <- 1.259
for (f in 1:2) {
  tau.EF[f] <- pow(sigma.EF[i],-2)
  }

# Sub-process 5: Subsampling #
for (t in startY:Y) {
  for (v in 1:M) {
    omega[t,v] ~ dunif(0,1)

### Model ###
# Sub-process 1: Transport #
# Data (PT) are organized by t=year, m=month, sink (1:4) #
for (q in 1:3) {
      p.PT[q,t,v,(1:4)] ~ ddirch(alpha.TR[q,t,v,(1:4)])
      m.PT[t,v,(1:4),q] ~ dmulti(p.TR[q,t,v,(1:4)],M.PT)
      }
    p.TR[t,v,(1:4)] <- p.PT[PT.map[t,v],t,v,(1:4)]
    n.entrain[t,v] <- n.edge[t,v]*(1-p.TR[t,v,3])

# Sub-process 2: Survival #
    p.survival[t,v,1] <- p.SV1[t,v]*ilogit(beta[1]+beta[2]*v)
    p.survival[t,v,2] <- p.SV1[t,v]

# Sub-process 3: Population length structure #
# Data (m.TMM) are organized by t=year, length (1:5), m=month #
    p.LN[t,v,(1:5)] ~ ddirch(alpha.LN[v,(1:5)])
    m.TMM[t,(1:5),v] ~ dmulti(p.LN[t,v,(1:5)],M.TMM[t,v,1])

# Sub-process 4: Sampling efficiency
# Data (m.FF) are organized by t=year, length (1:5), m=month #
    for (l in 1:5) {
      logit(p.SEL[t,f,l]) <- eta[1]*(l-eta[2])
      iota[t,v,l] <- p.SEL[t,v,l]*p.LN[t,v,l]
      }
```

```
        p.BYP[t,v,(1:5)] <- iota[t,v,(1:5)]/sum(iota[t,v,(1:5)])
        m.FF[t,(1:5),v] ~ dmulti(p.BYP[t,v,(1:5)],M.FF[t,v,1])

        m.FF[t,6,v] ~ dbin(omega[t,v],M.FF[t,v,2])

# Sub-process 5: Subsampling and observation #
# Data (y) are organized by t=year, m=month,f=fish facility #
    for (f in 1:2) {
       gamma[t,v,f] ~ dnorm(G[f],tau.EF[f])
       for (l in 1:5) {
          p.EF[t,v,f] <- p.SEL[t,v,l]*ilogit(gamma[t,v,f])
          n.FF.l[t,v,f,l] <-n.source[t,v]*p.TR[t,v,f]*p.LN[t,v,(1:5)]*p.EF[t,v,f,l]*p.survival[t,v,f]*rho[t,v,f]
          }
       n.FF[t,v,f] <- sum(n.FF.l[t,v,f,(1:5)])

       y[t,v,f] ~ dpois(n.FF[t,v,f]/(1-omega[t,v]))
       }
    }
  }
}
```
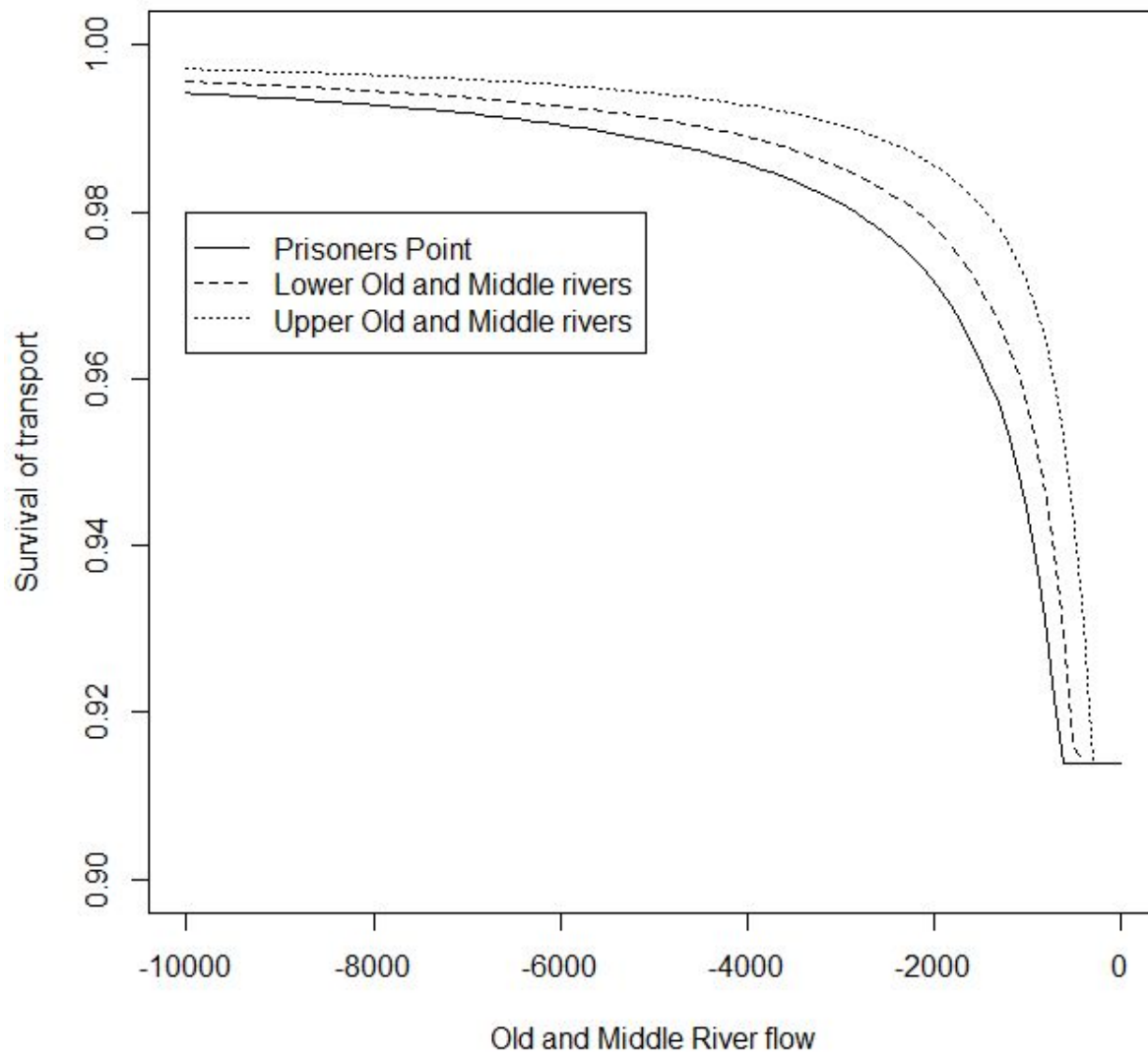
1    **Appendix D**



2

3    **Figure D1.** Assumed survival rates $p_{SV1_q}$ from each source region over a range of Old and Middle River

4    flows (cubic feet per second), representing transport times. Survival was truncated at 0.914.

5

6      **Additional diagnostic plots**

7
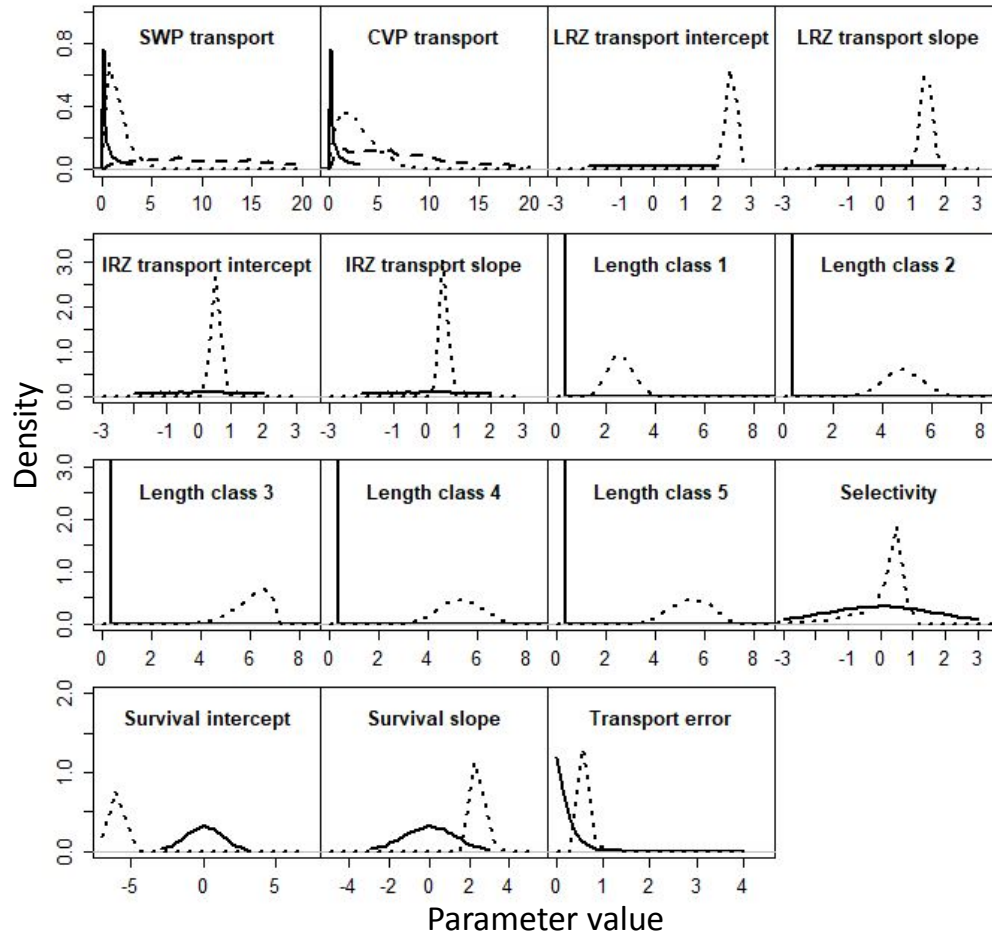


8

9      **Figure D2.** Prior versus posterior distributions for select model parameters. Prior distributions are

10     indicated with the dashed and dotted lines, and posterior distributions are represented with the solid lines.

11     As the model included up to 217 parameters, not all parameters are shown. State Water Project (SWP)

12     and Central Valley Project (CVP) transport parameters for May and June of 2002 and population length

13     parameters for June are depicted as examples. LRZ represents Low Risk Zone parameters, and IRZ

14     represents Intermediate Risk Zone parameters.
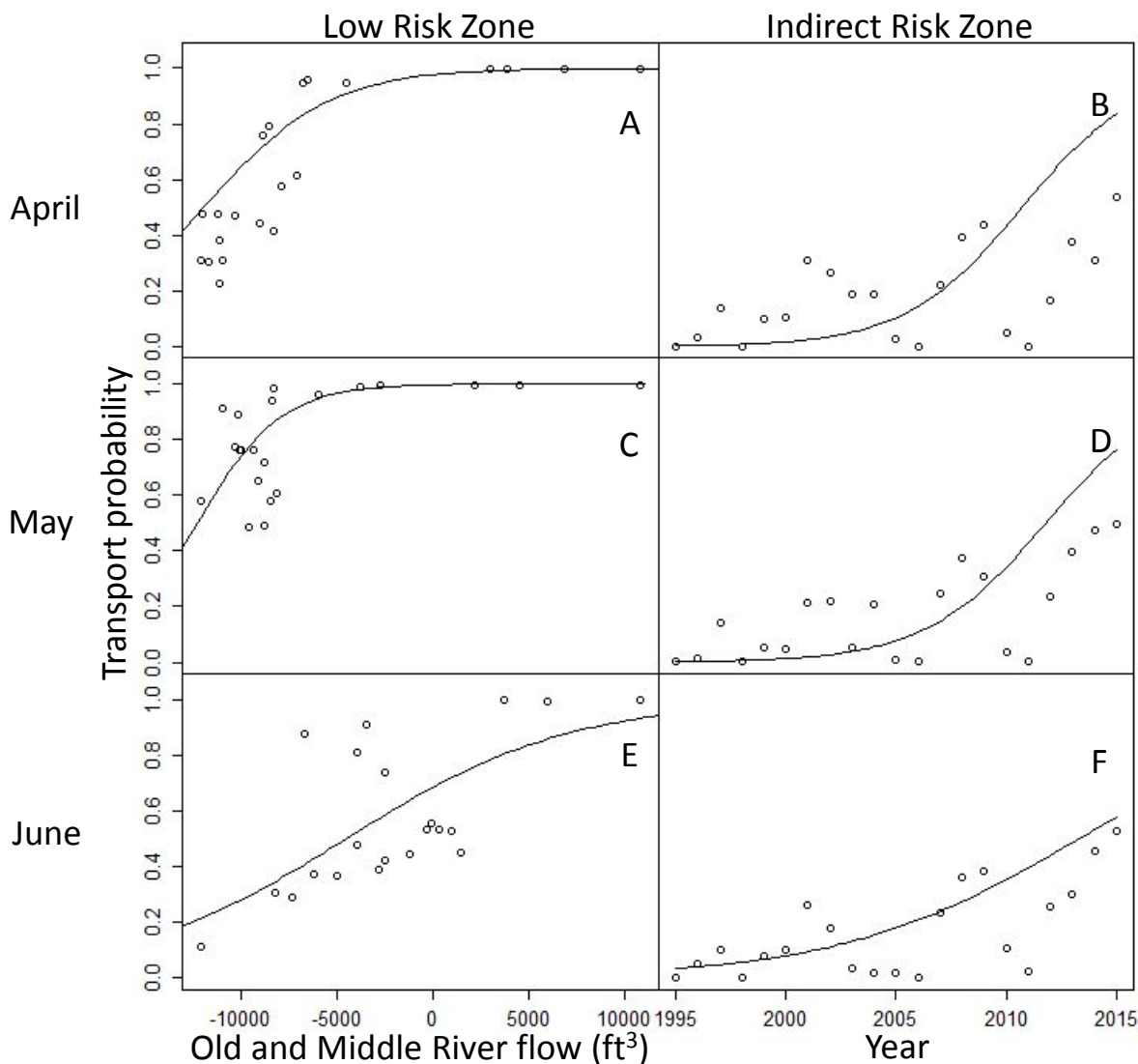
**Figure D3.** Empirical (particle tracking count/total number of particles) transport probabilities (dots) versus model estimated (lines) transport probabilities for the Low (panels A, C, and E) and Indirect Risk Zones (panels B, D, and F) when the source region was Prisoners Point on the San Joaquin River. Low Risk Zone probabilities were estimated as a function of Old and Middle River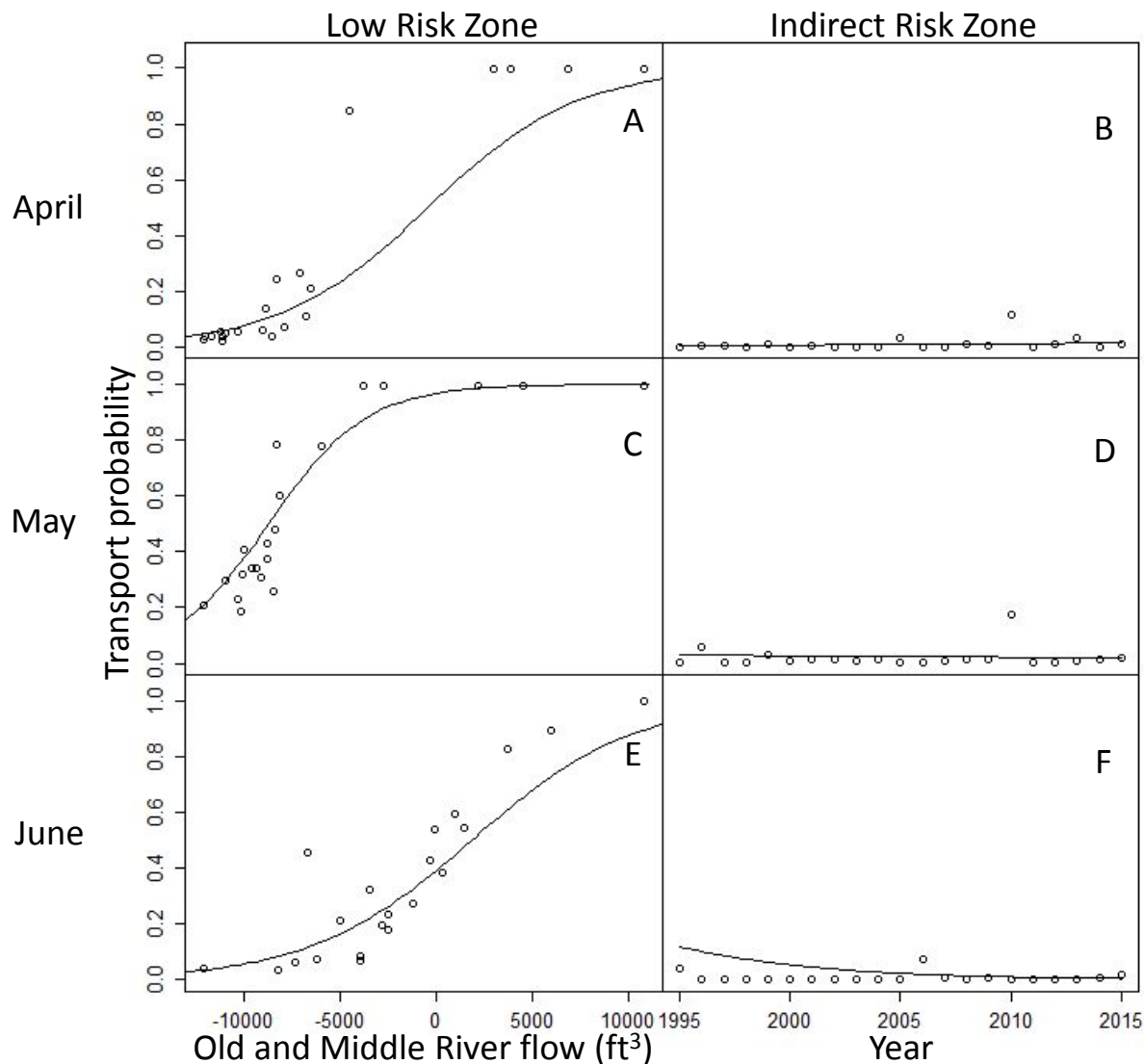 flow (Eq. C1), while holding all other probabilities at the monthly median value, and Indirect Risk Zone probabilities were estimated as a function of Old and Middle River flow (Eq. C2), while holding all other probabilities at the monthly median value.
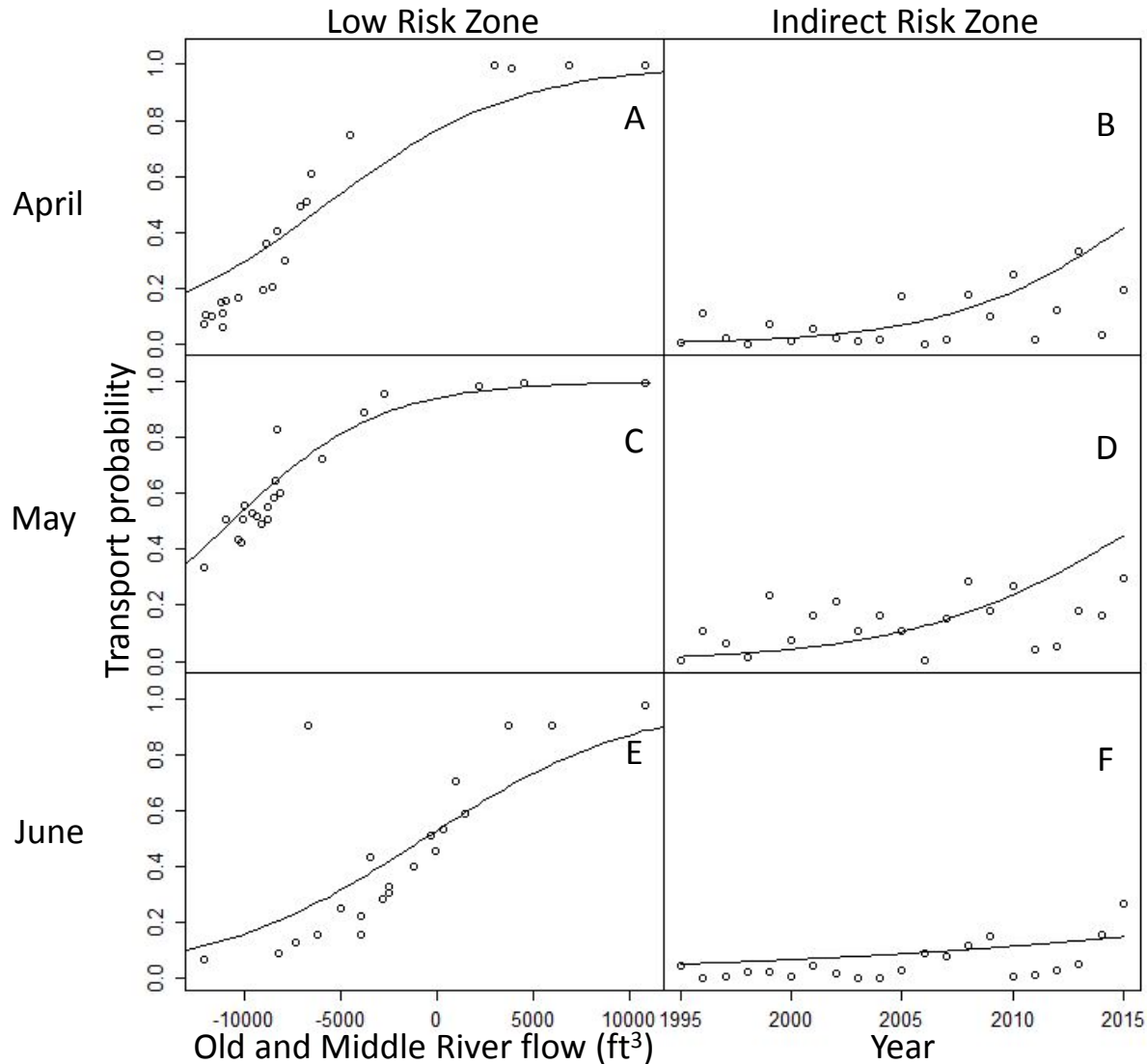
23

24 **Figure D4.** Empirical (particle tracking count/total number of particles) transport probabilities (dots)

25 versus model estimated (lines) transport probabilities for the Low (panels A, C, and E) and Indirect Risk

26 Zones (panels B, D, and F) when the source region was Lower Old and Middle rivers. Low Risk Zone

27 probabilities were estimated as a function of Old and Middle River flow (Eq. C1), while holding all other

28 probabilities at the monthly median value, and Indirect Risk Zone probabilities were estimated as a

29 function of Old and Middle River flow (Eq. C2), while holding all other probabilities at the monthly

30 median value.

31

32

**Figure D5.** Empirical (particle tracking count/total number of particles) transport probabilities (dots)

versus model estimated (lines) transport probabilities for the Low (panels A, C, and E) and Indirect Risk

Zones (panels B, D, and F) when the source region was Upper Old and Middle rivers. Low Risk Zone

probabilities were estimated as a function of Old and Middle River flow (Eq. C1), while holding all other

probabilities at the monthly median value, and Indirect Risk Zone probabilities were estimated as a

function of Old and Middle River flow (Eq. C2), while holding all other probabilities at the monthly
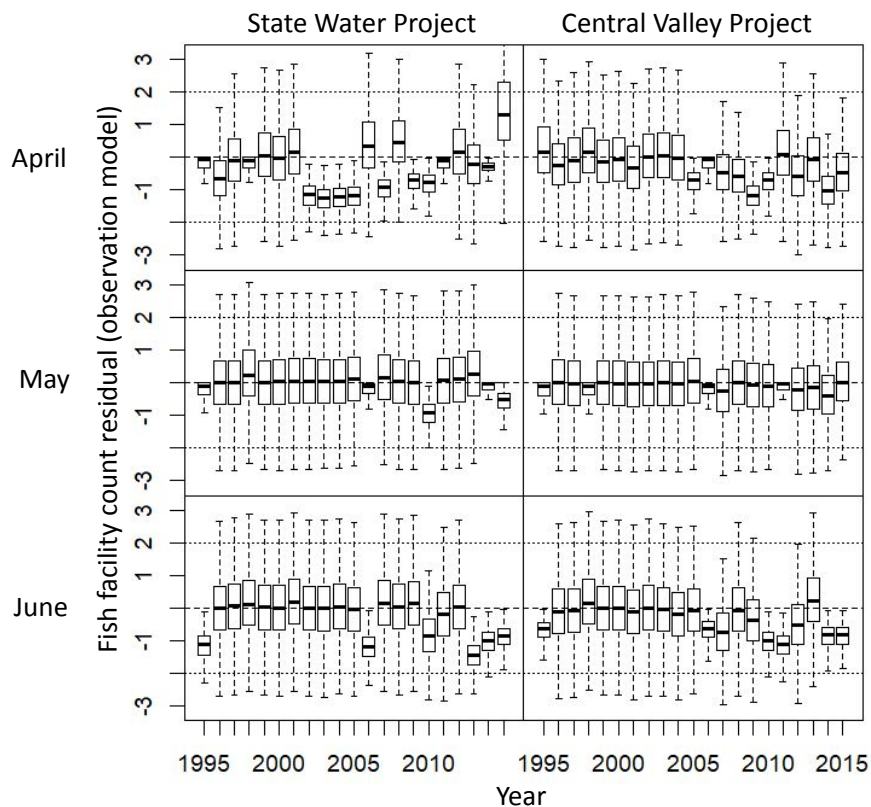
median value.
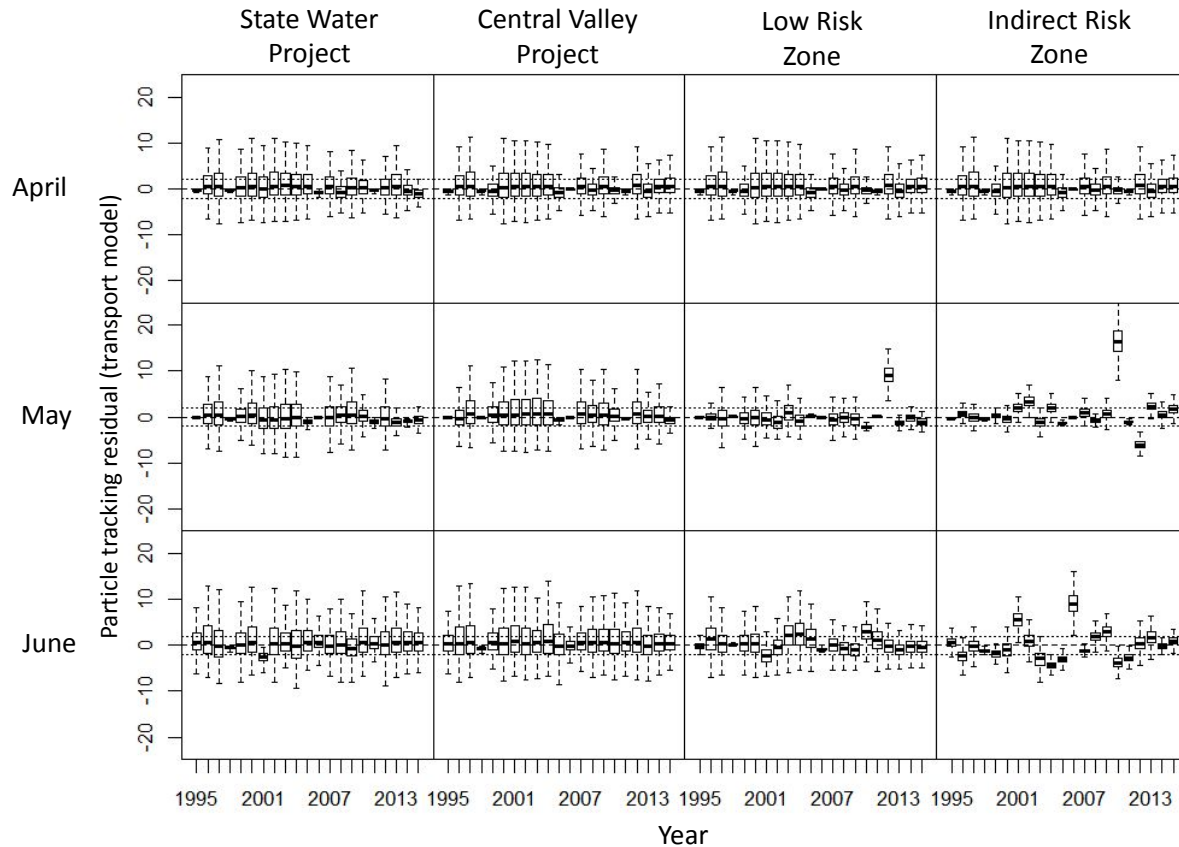
40

41     **Figure D6.** Time series of standardized residuals of the transport model fit to particle tracking data (Eq.

42     2). Boxplots of posterior residual distributions are shown, and the reference lines indicate the reference
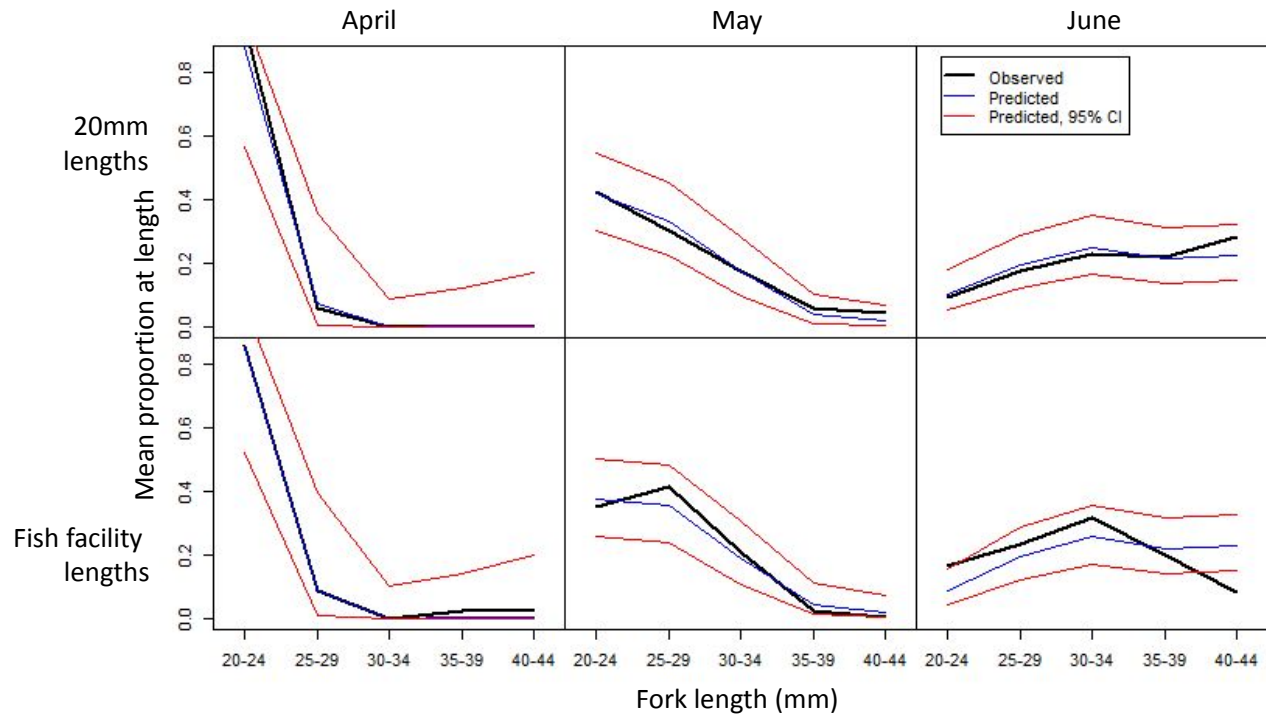
43     values -2, 0, and 2.

44

45    **Figure D7.** Time series of standardized residuals of the observation model fit to counts at fish facilities

46    (Eq. 13). Boxplots of posterior residual distributions are shown, and the reference lines indicate the

47    reference values -2, 0, and 2.

48

49

50

**Figure D8.** Mean observed (black) versus predicted proportion (blue = mean; red = 95% credible

interval) at length from 20mm survey data ($p_{LN}$; top row) and from lengths sampled at fish facilities ($p_{BYP}$

; bottom row).