DRAFT | Peer Review Purposes Only | Not for Citation

Forecasting Outmigration Timing and Abundance of Juvenile Spring-run Chinook Salmon at Rotary Screw Traps in Sacramento River Tributaries and the Mainstem to Support a Juvenile Production Estimate

Authors

Josh Korman, Ecometric Research; Liz Stebbins, FlowWest; Ashley Vizek, FlowWest; Noble Hendrix, QEDA Consulting; Brett Harvey, California Department of Water Resources

Acknowledgments

The data used in this modeling effort are rich and extensive—in some cases, data collection has been ongoing since the late 1990s. This work would not be possible without the field staff collecting daily rotary screw trap data and the data stewards that have managed these data over time. We specifically acknowledge the current data stewards, Anna Allison, Nicolas Bauer, Drew Huneycutt, Jeanine Phillips and Corey Fernandez, and the field staff participating in data collection, who assisted the effort to curate and make compatible the data used in this analysis, and provided valuable insights to contribute to the modeling process.

We also thank the Spring-run Juvenile Production Estimate Core Team, Modeling Advisory Team, and Interagency Review Team for their useful comments and advice.

Work by Ecometric Research and FlowWest on this project was supported by California Department of Water Resources. Work by QEDA Consulting was supported by the California State Water Contractors organization.

Executive Summary

The goal of the in-season outmigrant model described in this chapter is to annually forecast the timing and abundance of spring-run Chinook salmon (*Oncorhynchus tshawytscha*) (spring-run) outmigrants expected to pass rotary screw trap (RST) sites in Sacramento River tributaries and the mainstem. The model fits weekly outmigration abundance estimates from multiple years to predict the proportion of outmigrants that will pass a trapping site each week of the outmigration season. The variation in outmigration timing among years is used to forecast outmigration-timing predictions in a future year. Given an observation of outmigrant abundance up to a specific week in a forecast year, the outmigration-timing model can predict the abundance for all weeks from the forecast week to the end of the migration season and therefore the total outmigrant abundance for the year. Forecasts of weekly abundance at an RST site are then routed to the Sacramento–San Joaquin River Delta (Delta)-entry location on the Sacramento River by a separate model that predicts travel time and losses during migration due to natural mortality.

We applied the in-season outmigrant model to weekly estimates of spring-run juvenile outmigrant abundance at six RST sites in tributaries of the Sacramento River: upper Battle Creek, upper Clear Creek, Deer Creek, Mill Creek, Butte Creek, and the Yuba River. In addition, the model was applied to data from the Knights Landing and Tisdale RST sites on the mainstem Sacramento River. The model generally provided good fits to the observed cumulative weekly outmigration abundance data (from the BT-SPAS-X model). Upper Clear Creek had a much earlier median outmigration date than other sites, while Mill and Deer creek had the latest median outmigration dates. Upper Clear Creek had a very steep outmigration-timing curve (i.e., rapid departure), while those for the Mill and Deer RST sites had the lowest-sloped outmigration-timing curves (i.e., gradual departure). There was evidence for a negative relationship between median outmigration date and run-timing steepness across years at six of eight RST sites. However, with the exception of Battle Creek, there was large uncertainty in the magnitude of the correlation, and most estimates were low. There was a weak negative effect of peak flows prior to February on the median outmigration date for seven of eight RST sites. However, with the exception of the Yuba River RST site, flow effects were close to zero and highly uncertain. The analysis presented here is only intended to provide an example of how covariates can be included in the model. Model updates are expected as new covariates are identified and tested.

There was considerable uncertainty in *forecasted* outmigration timing largely driven by the interannual variability in *observed* outmigration timing. Estimates of outmigration timing were most precise for upper Clear Creek and were least precise for Yuba River, Knights Landing, and Tisdale sites. For most sites (i.e., all but upper Clear Creek), historical observations show only a small proportion have typically migrated past RST sites early in the outmigration period (e.g., by the beginning of

the December 28 forecast week). As a result, the forecast model allows the possibility that only a small proportion of the total outmigrant population has passed the RST site by this time, which results in large and highly uncertain estimates of forecasted total abundance. The certainty in total abundance forecast improves for forecast dates made later during the outmigration season, as the proportion of total observed outmigrant population that has passed the site in that current year increases.

In-sample and out-of-sample forecast error of annual outmigrant abundance was lower for later forecast weeks compared to earlier weeks. The only exception was upper Clear Creek, which had the earliest outmigrant run timing and therefore low error for all forecast weeks. In this case the out-of-sample error was driven largely by uncertainty in the cumulative abundance up to each forecast week. Out-of-sample error based on null and covariate models (peak flows prior to February) predicted median outmigration dates that were generally similar. Differences in out-of-sample error only occurred on the earliest two forecast weeks when out-of-sample error was high for both models.

Preliminary evaluation of the in-season outmigrant model presented here indicates it could be a useful tool for designing and implementing protective actions for spring-run juveniles. That said, forecast error on annual abundance from early forecasts were very uncertain. Considering the in-season model is intended to serve in a multi-model framework for forecasting juvenile production estimate (JPE) for spring-run entering the Delta each year, it is unclear if the more-accurate forecasts later in the outmigration season will be more useful for management than less-certain earlier forecasts until decisions are made with regards to how the JPE will be used to guide management. At a minimum the pre-season forecast of outmigration timing provides a more-rigorous definition of the period when spring-run juveniles are most likely to be vulnerable to entrainment. Additional work on covariate effects is needed. The in-season outmigration model will be applied to data from Feather River RST sites when results from ongoing work on the Feather River probabilistic length-at-date (PLAD) model become available.

Contents

Tables

Figures

Appendices

Acronyms and Abbreviations

Term	Definition
°C	degrees Celsius
BT-SPAS	Bayesian Temporally Stratified Population Analysis System
CV	coefficient of variation
Delta	Sacramento–San Joaquin River Delta
JPE	juvenile production estimate
LOOCV	leave-one-out cross validation
PLAD	probabilistic length-at-date
RST	rotary screw trap
spring-run	spring-run Chinook salmon

1 Introduction

The goal of the in-season outmigrant model described in this chapter is to forecast the timing and abundance of spring-run Chinook salmon (spring-run) outmigrants passing rotary screw trap (RST) sites in Sacramento River tributaries and the mainstem. The model fits weekly outmigration abundance estimates from multiple years to predict the proportion of outmigrants that will pass a trapping site for each week of the outmigration season. The variation in outmigration timing among years is used to generate outmigration-timing predictions for a future year where a full set of weekly abundance estimates for the entire year are not available. Given an observation of outmigrant abundance up to a specific date in a forecast year, the outmigration-timing model can predict the abundance for all weeks past this forecast date and therefore the total outmigrant abundance for the year. Predictions of weekly abundance at an RST site will be routed to the Sacramento–San Joaquin River Delta (Delta) by a separate model that predicts travel time and losses due to natural mortality (Cordoleani and Korman 2025). A spring-run juvenile production estimate (JPE) in the Delta can then be derived by summing the predictions from spring-run tributaries with RST data. These predictions can be used by water managers to make operational decisions, or take other protective actions, to minimize impacts on spring-run juveniles.

The in-season model is fit to weekly and annual estimates of outmigrant abundance passing an RST site that are calculated from BT-SPAS-X (refer to Chapters 4 and 5). BT-SPAS-X estimates weekly trap efficiency for each week of the outmigration season using a Bayesian hierarchical model. This allows the model to estimate weekly trap efficiency for all weeks in an outmigration season, including the majority of weeks when no trap efficiency data were collected. The model also includes a Bayesian spline interpolation method to estimate weekly abundance, which allows the model to interpolate abundance in weeks when the RST was not operated. BT-SPAS-X makes no assumptions about the outmigration timing in a year. The in-season model described here treats the weekly BT-SPAS-X abundance estimates as observations (hence our use of “observations”) but propagates uncertainty in the estimates of weekly abundance in the estimation of outmigration timing.

A pre-season estimate of juvenile outmigrant abundance at an RST site or in the Delta based on a spawner-outmigrant stock-recruitment model may be highly uncertain, while predictions from the in-season outmigrant model described here could potentially be more precise because they are based in part on in-season abundance estimates. The model forecasts abundance and timing for the total Sacramento River spring-run production, and the contribution of individual populations to that total, providing the ability to make pre-season decisions to protect a desired segment of the juvenile outmigrant population. Updated in-season predictions of outmigrant abundance using new sampling data after the initial

forecast date could potentially be used to fine-tune these decisions as the outmigration season progresses.

This chapter describes the in-season outmigrant model. We applied the model to weekly estimates of spring-run juvenile outmigrant abundance at six RST sites in tributaries of the Sacramento River: Battle Creek, upper Clear Creek, Deer Creek, Mill Creek, Butte Creek, and Yuba River. In addition, the model was applied to data from the Knights Landing and Tisdale RST sites on the mainstem Sacramento River. Weekly estimates of juvenile spring-run abundance are currently not available for RST sites on the Feather River. Work is ongoing to apply the probabilistic length-at-date (PLAD) model (Chapter 3) to data from the Feather River, which is more challenging than application at other RST sites due to a more complicated run structure and to the release of large numbers of hatchery-origin spring-run juveniles upstream of the RST sites. The challenges of modeling outmigration abundance for the Feather River, and plans for addressing those challenges, are described in more detail in Chapter 3.

2 Methods

2.1 Model Description

In the equations that follow, variables beginning with Greek letters denote parameters that are directly estimated by the model. Bolded Roman letters represent data, and non-bolded Roman letters represent model variables calculated from estimated parameters and sometimes data. Lower case Roman letter subscripts represent indices (e.g., model week 'iwk').

2.1.1 Predicting Outmigration Timing

We predict weekly cumulative outmigrant-timing using a beta distribution (beta). At any RST site, the cumulative proportion of the total outmigrant abundance passing the trap up to week *iwk* in year *iyr* ($p_{iwk,iyr}$) is calculated from,

Equation 1.

$$p_{iwk,iyr} = \sum_{iwk=1}^{iwk=1:53} beta(\boldsymbol{t_{iwk}}, alpha_{iyr}, beta_{iyr}) \cdot e^{\varepsilon_{iwk,iyr}}$$

Where:

$\boldsymbol{t_{iwk}}$ is the proportional week of the year (iwk= 1 represents November 1 through November 7 and has $t_{iwk=1}$ = 1/53 = 0.019)

alpha and beta are year-specific parameters of the beta distribution, and

$\varepsilon_{iwk,iyr}$ are weekly deviations from the beta distribution.

They represent variation in outmigration timing that cannot be explained by the smooth beta function and are thus referred to as extra-beta variation. Weekly values returned from Equation 1 are standardized so they sum to one across the 53 weeks in a year. The summation in Equation 1 indicates that outmigration timing (p) represents a cumulative probability that increases over the outmigration season and will equal one at or before the last week in the year.

We used two alternate methods for modeling extra-beta variation in outmigration timing. The simplest assumes weekly deviates are independent. That is, the deviate in one week does not depend on deviates in adjacent weeks. For this case, deviates are drawn from a normal distribution with a mean of 0 and estimated standard deviation σ_p,

Equation 2a.

$$\varepsilon_{iwk,iyr} \sim normal(0, \sigma_p).$$

The second approach allows weekly deviates to be correlated over time as one might expect if deviations from successive weeks tend to be above or below the outmigration timing predicted by the beta distribution. We use a lag-1 autocorrelation model to capture this pattern,

Equation 2b.

$$\nu_{iwk,iyr} \sim normal(0, \sigma_p)$$

$$\varepsilon_{iwk,iyr} = \rho_p \cdot \varepsilon_{iwk-1,iyr} + (1 - \rho_p) \cdot \nu_{iwk,iyr}.$$

Where:

ρ_p is an estimated autocorrelation coefficient which can range from -1 to 1.

High absolute values of ρ_p result in higher autocorrelation among weekly outmigration timing deviates ε. A high positive value indicates that deviates from a series of weeks will tend to be above or below the mean predicted by the beta distribution. Estimates of ρ_p near zero indicate that weekly deviations are randomly distributed. In this situation predictions from Equation 2b will be very similar to those from Equation 2a.

Rather than estimate the alpha and beta terms in the beta distribution directly, we estimate the more intuitive mean (phi) and sample size (lambda) of the distribution and then convert them using,

Equation 3a.

$$alpha_{iyr} = lambda_{iyr} \cdot phi_{iyr}$$

Equation 3b.

$$beta_{iyr} = lambda_{iyr} \cdot (1 - phi_{iyr}).$$

Phi represents the proportional week of the year when 50% of the outmigrant population has passed the RST site. This is often termed the mean of the beta distribution, and in the context of the outmigration model represents the median outmigration date. Lambda represents the sample size of the beta distribution. A larger lambda indicates the beta distribution has more information (i.e., less variance), which results in a tighter/narrower distribution. In the context of outmigration timing, a larger lambda indicates a steeper rise in the cumulative probability across the weeks when outmigration is predicted to occur. That is, a greater concentration of outmigration occurring around the median date of outmigration. We refer to this parameter as outmigration-timing steepness.

Values of phi and lambda are calculated from the mixed effect models,

Equation 4a.

$$phi_{iyr} = inv\_logit(\phi_{iyr} + \beta_1 \cdot \boldsymbol{X_{iyr,1}})$$

Equation 4b.

$$lambda_{iyr} = exp(\lambda_{iyr} + \beta_2 \cdot \boldsymbol{X_{iyr,2}})$$

Where:

ϕ and λ are the estimated year-specific intercepts predicting phi and lambda, respectively (the random effects),

the β's are estimated fixed effect coefficients, and

the X's are standardized annual covariate values (e.g., the peak flow prior to February in each year).

The parameters for calculating phi are estimated in logit space and predictions are transformed by the inverse logit function (inv_logit) to ensure estimates of phi fall between 0 and 1. The parameters for lambda are estimated in log space and predictions are transformed by the exponential function (exp) to ensure lambda is greater than zero. To run a "random effects only" model, β values are not estimated and are instead fixed at zero.

It is plausible that there is a correlation between annual estimates of phi and lambda. For example, if high flows stimulate outmigration, we could expect years with high flows from November through January to have both a lower median outmigration date (low phi) and a steeper outmigration-timing curve (high lambda) relative to the timing in a year with average flows for this period. Conversely, years with lower flows during this period would be expected to have later median outmigration date (high phi) and a slower increase in cumulative outmigrant abundance over the season (small lambda). It is important to account for such potential correlations when forecasting outmigration timing in a future year. If these terms are correlated, independent sampling of posterior distributions of annual values of phi and lambda to predict outmigration timing in a future year (i.e., assuming no covariation between them) would overestimate uncertainty in forecasted outmigration timing, since it would produce a higher frequency of low-phi:low-lambda and high-phi:high-lambda cases than observed in the historical data. To account for this possibility, annual estimates of ϕ and λ are assumed to be random variables drawn from a multivariate normal distribution,

Equation 5.

$$\phi_{iyr}, \lambda_{iyr} \sim multi\_normal(\mu = c(\overline{\phi}, \overline{\lambda}), vcv = c(\sigma_\phi \sigma_\lambda, \rho))$$

where $\overline{\phi}$ and $\overline{\lambda}$ represent the estimated across-year means (μ) of the multivariate normal (MVN) distribution for ϕ and λ in transformed space, the σ's are their

estimated standard deviations representing the extent of interannual variability in these parameters, and ρ is the estimate of the magnitude of the covariation between ϕ and λ across years (i.e., the Pearson correlation coefficient). Estimates of the σ's and ρ are used to calculate a variance-covariance matrix (vcv) that is needed along with the means to define the multivariate normal distribution. This MVN distribution is the joint hyper-distribution governing the extent of interannual variation in ϕ and λ estimates. The distribution captures the potential correlation between ϕ and λ across years. Note that if the data indicate that annual values of ϕ and λ are not correlated, the model will estimate a value of ρ near zero. The MVN approach does not force the model to predict a correlation between ϕ and λ if none exists.

2.1.2 Fitting the Outmigration-Timing Model to Data

With predictions of the proportion of the total annual outmigrant population passing the RST site from the start of outmigration up to each week ($p_{iwk,iyr}$ from Equation 1), the abundance up to each week in a given year ($pN_{iwk,iyr}$, the cumulative weekly abundance) can be calculated from,

Equation 6.

$$pN_{iwk,iyr} = \exp\left(Ntot_{iyr}\right) \cdot p_{iwk,iyr}$$

Where:

$Ntot_{iyr}$ is the annual abundance of the outmigrant population in log space estimated from the BT-SPAS-X model (refer to Chapters 4 and 5).

As annual abundance is uncertain, it is treated as a latent (unobserved) variable and modeled using,

Equation 7.

$$\boldsymbol{Ntot\_mu_{iyr}} \sim normal\left(\log\left(Ntot_{iyr}\right), \boldsymbol{\sigma_{Ntot_{iyr}}}\right)$$

Where:

$Ntot\_mu_{iyr}$ and $\boldsymbol{\sigma_{Ntot_{iyr}}}$ are the mean and standard deviation of the log of annual abundance estimates determined from the posterior distributions estimated by BT-SPAS-X (Korman et al. 2024a, b).

The model will estimate a posterior distribution of log($Ntot\_mu_{iyr}$) for each year with means and standard deviations close to what BT-SPAS-X estimated to maximize the probability returned from Equation 7. However, the model may not predict means of Ntot that are the exactly the same as Ntot_mu values. Some deviations are expected if they improve the probability from the data likelihood described below. The estimation of Ntot is done using the log of abundance estimates and

transformed in Equation 6 so that negative estimates of annual abundance cannot occur.

The model is fit by comparing the predicted abundances from the start of the outmigration season up to each week ($N_{iwk,iyr}$) with the log-transformed values estimated from BT-SPAS-X using the data likelihood,

Equation 8.

$$\boldsymbol{N\_mu_{iwk,iyr}} \sim normal(\log(p_{iwk,iyr}), \boldsymbol{\sigma_{N_{iwk,iyr}}})$$

Where:

N_mu$_{iwk,iyr}$ and $\boldsymbol{\sigma_{N_{iwk,iyr}}}$ are the mean and the standard deviation in the log-transformed cumulative abundance estimates provided by BT-SPAS-X for each week of the outmigration season (refer to Appendix B of Chapters 4 and 5 for weekly estimates of spring-run outmigrant abundance).

The uncertainty in the estimated cumulative weekly abundance estimates will be higher early in the outmigration season when fewer weekly estimates contribute to the cumulative abundance estimate. As a result, these early estimates will have less influence on parameter estimates than later ones that may be more precisely estimated. Estimates of certainty in **N_mu** values also vary across RST sites and years. Sites and years with more trap efficiency trials and higher trap efficiency will have more precise estimates of **N_mu** (lower $\boldsymbol{\sigma_N}$), which will in turn provide more information when fitting outmigration-timing parameters.

2.1.3 Forecasting Outmigration Timing and Annual Abundance

Once the model is fit to data over a set of years from an RST site, a forecast of outmigration timing and annual abundance for a future year can be calculated. First, posterior distributions of parameters in Equation 5 are used to generate correlated random draws for the intercepts of phi and lambda for the forecast year (Equations 4a and 4b). The extent of variation in these draws will depend on the estimated extent of interannual variability in their estimates and covariation as quantified by σ_ϕ, σ_λ, and ρ. Second, if a fixed effect was estimated, the covariate value for the year of the forecast is multiplied by the posterior estimates of β to adjust the estimate of phi or lambda using Equations 4a or 4b. Third, posterior estimates of process error (σ_p) and ρ_p (if the lag-1 model is used) are used to generate deviates for all weeks in the forecasted year (ε_{iwk}) using Equations 2a or 2b, and are included in the forecasted outmigration timing for each week of the year (fp_{iwk}) via Equation 1. Finally, a forecast of annual abundance (fNtot) is calculated from,

Equation 9.

$$fNtot = \exp\left(fN_{fwk}\right) \cdot \frac{1}{fp_{fwk}}$$

Where:

fN_{fwk} is the log-transformed value of the "observed" cumulative weekly abundance for the week used to make the forecast (*fwk*).

As the value is estimated (by BT-SPAS-X) and therefore uncertain, the error in the estimate is simulated using,

Equation 10.

$$fN_{fwk} \sim normal\left(\boldsymbol{fN\_mu_{fwk}}, \boldsymbol{\sigma_{fN_{fwk}}}\right)$$

Where:

fN_mu_{fwk} and $\boldsymbol{\sigma_{fN_{fwk}}}$ are the mean and standard deviation of the log-transformed cumulative weekly abundance estimate for the forecast week generated by BT-SPAS-X.

Cumulative weekly estimates of abundance from the forecast week and weeks after (iwk=(fwk+1):53) are then calculated from,

Equation 11.

$$fN_{iwk} = fNtot \cdot fp_{iwk}$$

We made forecasts for weeks starting December 28, February 1, March 1, and March 29, which are the model weeks closest to the start of January, February, March and April. This range of forecast weeks represent a broad range of information available to make forecasts. Forecasts from the early weeks will be less certain than later ones due to both lower and less-certain estimates of the proportion of the run that has passed the RST (fp_{iwk}, Equation 8). In addition, earlier forecasts will be less certain due to greater uncertainty in the observed abundance passing the RST up to the forecast week (fN_{fwk}). Estimates of observed abundance for later forecast weeks are more certain because they are informed by more data. For example, the cumulative abundance estimate up to the first forecast week analyzed here (i.e., up to the week of December 28) depends only on weekly estimates between November 1 and this date, and many of the estimates can be near zero. Thus, the cumulative abundance may depend on only a few non-zero weekly estimates, leading to high uncertainty in the cumulative abundance estimate relative to the case for later forecast weeks where many weekly estimates make substantive contributions to the cumulative estimate.

2.1.4 Estimation

The model was fit in the Stan Bayesian statistical modeling software (Stan Development Team 2024). Acceptable convergence, as assessed by the Gelman-Rubin convergence statistic (Gelman et. al., 2004, $\hat{r} < 1.05$) was achieved by running three chains for 2,000 iterations. Uninformative uniform priors were used for all parameters except σ_p. We used a minimally informative gamma prior on σ_p with shape and rate parameters of 20 and 10 (σ_p~gamma(20,10)). The lag-1 model required an additional weak prior on the autocorrelation coefficient (ρ_p~dbeta(10,10)). Even with this latter prior it was sometimes difficult to achieve convergence for both σ_p and ρ_p for the lag-1 model.

2.2 Out-of-Sample Error

We used a leave-one-out cross validation (LOOCV) approach to calculate the out-of-sample error in annual forecasts of spring-run juvenile abundance from the in-season outmigration model. The analysis was conducted in the following six steps for a given model at an RST site:

1. We estimated parameters using the same approach described in Section 2.1, but leaving out one year of data from the full set of years available to fit the model (n years). This process was repeated for each year, resulting in n different model fits based on n-1 years of data.
2. For each n-1 fit, we used the posterior distribution of the forecasted cumulative proportion of the run that passed the trapping site up to a forecast week ($fp_{fwk}$from Equation 8) to expand the "observed" estimate of juvenile abundance up to that week for the year not included in the fitting (fN_{fwk}). This leads to a prediction of annual abundance for the year left out of the fitting ($fNtot$). Note the calculation includes the uncertainty in fN_{fwk} (as determined by BT-SPAS-X) in the prediction of $fNtot$.
3. The difference between the forecasted annual abundance ($fNtot$) and the "observed" annual abundance (Ntot) in the left-out year was then calculated. Note uncertainty in the "observed" Ntot from BT-SPAS-X was included in the calculation. Given posterior distributions of both $fNtot$ and *Ntot*, a posterior distribution of differences is derived.
4. Two statistics for each left-out year are computed from the distribution of differences produced in Step 3: a) the median of the absolute values of differences (diff=abs(Ntot - $fNtot$) ; and b) the median of relative differences (reldiff=100*abs(Ntot- $fNtot$)/Ntot).
5. Finally, we summarized the annual predictions of out-of-sample error by taking the across-year medians of median diff and reldiff values from Step 4.

6. Steps 1–5 were repeated for each of the four forecast weeks used in the analysis. This quantified the extent of the decrease in out-of-sample error in forecasted annual abundance as more information on outmigration timing and abundance becomes available over the outmigration period.

3 Results

The amount of information available to estimate outmigration timing varied among the eight RST sites (Table 1). Battle Creek, upper Clear Creek, Butte Creek, and Knights Landing had approximately 20 years of outmigration-timing data, while Yuba River had less than 10 years of data. Some RST sites had relatively precise estimates of abundance (Yuba River, Deer Creek), while mainstem RST sites and Butte Creek had less-precise abundance estimates. The model generally provided good fits to the "observed" (from BT-SPAS-X) cumulative weekly outmigration abundance data (Figure 1). Owing to flexibility in the outmigration-timing model from the weekly deviates (Equation 1), model predictions (gray lines and gray shaded areas in Figure 1) were able to almost perfectly predict the observed run timing (black dots) for most years. There were a few exceptions (e.g., Battle Creek in 2002 and 2024, and upper Clear Creek in 2018), which typically occurred when outmigration timing was unusually late and increased more slowly over the run year, or when there were multiple peaks in the run (identified by a pattern of steeply rising points followed by a flat section, followed by another steep rise). The extent of uncertainty in model predictions of outmigration timing (the shaded area in Figure 1) largely depended on how well the beta distribution described the weekly pattern in outmigrant abundance, but also depended on the uncertainty in observed cumulative abundance estimates (Appendix A).

Model parameters were generally well defined as shown by the relatively low values of the standard deviations of the posterior distributions compared to the means (Table 2). Upper Clear Creek had a much earlier median outmigration date than other RST sites (perhaps because of very limited fry rearing habitat), while Mill and Deer creek had the latest median outmigration dates (inv_logit(ϕ)). Upper Clear Creek had a very steep outmigration-timing curve (exp(λ)), while those for Mill and Deer creek had the lowest slopes. There was evidence for a negative relationship between phi and lambda across years at six of eight RST sites (negative ρ values). However, with the exception of Battle Creek, there was large uncertainty in ρ estimates. In the case of Battle Creek, years with earlier median outmigration dates tended to have steeper outmigration-timing curves, and vice-versa (Figure 2). This pattern among annual phi and lambda estimates likely indicates it would be difficult to estimate covariate effects (e.g., flow) on both phi and lambda in the same model (Equations 4a and 4b) at Battle Creek. To reliably separate covariate effects on these two parameters, interannual patterns in phi and lambda need to show both high and low lambda for a given level of phi, and high and low phi for a given level of lambda. However, there are no cases of high phi and high lambda, or low phi and low lambda at Battle Creek.

The covariate model on phi (Equation 4a) showed a weak negative effect of peak flows prior to February on the median outmigration date for seven of eight RST sites (β[1] in Table 2, Figure 3). However, the magnitude of the flow effects was

close to zero and highly uncertain for all sites except Hallwood. Much more work on covariate models is required to evaluate a range of covariate effects on both phi and lambda. The analysis presented here is only intended to provide an example of how covariates can be included in the model.

Given the limited influence of currently tested flow covariates, uncertainty in forecasted outmigration timing was largely driven by the unexplained interannual variability in observed timing, which was considerable. Estimates of outmigration timing were most precise for upper Clear Creek and were least precise for Yuba River, Knights Landing, and Tisdale (Figure 4). For all RST sites except for upper Clear Creek, only a small proportion of Chinook salmon have migrated past RST sites early in the outmigration period (e.g., by the week of December 28). As a result, the model predicts there is high probability that the proportion of total outmigrant population passing RST sites is very low, which in turn leads to very large and uncertain estimates of forecasted total abundance (i.e., estimated abundance up to the week of December 28 is divided by a very small number in Equation 8). The certainty in total abundance forecast improves for forecast dates later during the outmigration season as the proportion of total outmigrant population passing an RST site increases. However, for sites with later outmigration timing (e.g., Deer and Mill creek) or uncertain timing (e.g., Butte Creek, Yuba River, Knights Landing, and Tisdale), there is still considerable uncertainty in the proportion passing an RST by the week of March 1, which contributed to high uncertainty in forecasted abundance estimates. By the last forecast week of March 29, the majority of the run has passed all RST sites except at Mill and Deer creeks. At this point in the year, most of the outmigration has been observed at six of eight RST sites, so the forecast at these sites should be more precise. However, this late forecast date may not be useful to managers as the majority of the outmigrants would have already been exposed to operational effects of the water projects. On the first forecast week of December 28 within-site uncertainty in outmigration timing and hence forecasts were very high for all RST sites except at upper Clear Creek (Figure 5). Precision of forecasts improved later in the outmigration period owing to higher proportions passing RST sites and more certainty in these estimates (Figure 4). The greater uncertainty in Knights Landing and Tisdale (i.e., the mainstem) outmigration-timing curves compared to tributary sites was driven by the much-higher estimates of process error (σ_p in Table 2), which is evident graphically in the wider credible intervals (gray band) of the year-specific timing curves (Figure 1). Higher levels of process error are required to fit the outmigration pattern at mainstem RST sites because the passage timing is often sporadic, resulting in a non-smooth pattern that cannot be accommodated by the beta distribution. Note that outmigration timing is not the only determinant of forecast error. For example, at upper Clear Creek, the majority of outmigrants have passed the RST site by the week of December 28, and there is near-total confidence that all have passed the RST site by the week of February 1 (Figure 4). Yet forecast abundance at upper Clear Creek does not drop to zero for the week of February 1 and later forecast weeks (Figure 5). This occurs because the model

propagates uncertainty in weekly abundance estimates from PLAD and BT-SPAS-X. Even with perfect knowledge of fp_{fwk}in Equation 8, there is still uncertainty in fN_{fwk}.

There was autocorrelation in weekly deviations in outmigration timing (ρ_p in Table 2). Correlation coefficients (ρ_p) were positive, indicating for example that a deviation above the expected outmigration timing from the beta distribution is more likely to be followed by another positive deviation (Figure 1). Models with lag-1 autocorrelation had difficulty converging for four of eight RST sites based on the minimally informative priors used for σ_p and ρ_p. At the four sites where convergence occurred, forecasted outmigration timing from random and lag-1 autocorrelated error models were almost identical (Figure 6). We therefore only focused on the random error model in this analysis.

As for the within-site analysis of forecast error of annual outmigrant abundance (Figure 5), out-of-sample error in forecasts was lower for later forecast weeks compared to earlier ones (Table 3). The only exception was upper Clear Creek, where the earliest outmigrant run timing occurs, which resulted in low out-of-sample error for all forecast weeks. In this case, the out-of-sample error was driven largely by uncertainty in the cumulative abundance as of each forecast week, and the uncertainty in observed (BT-SPAS-X) annual abundances. Out-of-sample error (with peak flows prior to February) predicting a median outmigration date were generally similar between null and covariate models. The null model had relative errors 10% or lower than the covariate model for six of 32 cases (eight RST sites * four forecast dates). The covariate model had relative errors 10% or lower than the null model for five of 32 cases. These differences only occurred for the earliest two forecast weeks when out-of-sample error was high for both models. Interestingly, the null model had substantively lower out-of-sample error than the covariate model for the Yuba River RST site on the earliest two forecast weeks, even though the magnitude of the covariate effect and the certainty in the effect was higher compared to other sites ($\beta[1]$ in Table 2). This occurred because the covariate effect for the Yuba River RST site was largely driven by one out of the seven available years, which had a high peak flow prior to February (Table 1). Thus, the out-of-sample error for the covariate model was substantively lower than the null model when this year was left out of the fitting (Figure 7). However, in other years with lower covariate values, model accuracy was similar or higher for the null model. Plots of year-specific out-of-sample error for all sites are provided in Appendix B.

4 Discussion

The preliminary evaluation of the in-season outmigrant model presented here indicates it could be a useful tool for designing and implementing protective actions for Sacramento River spring-run juveniles. That said, forecast error on annual abundance from early forecasts in the outmigration season were very uncertain. Until decisions are made about how the JPE will be used, it is unclear if the more-accurate forecasts later in a outmigration season would be useful for management. At a minimum, a forecast of outmigration timing provides a more-rigorous definition of the period when spring-run juveniles are migrating and potentially affected by water operations and flow actions, particularly along the mainstem Sacramento River.

A number of modifications and additional analyses of the in-season outmigration model should be considered or conducted. The simple beta distribution in the current version of the model could not capture more complex outmigration-timing patterns, such as years with both early and late peaks. In such cases, the model tended to split the difference (e.g., Battle Creek in 2012) and increases process error to quantify the lack of fit (i.e., weekly deviations from the beta distribution), which in turn reduced precision of forecasts. While it is possible to employ more complex mixture distributions to better capture more complex patterns of outmigration timing, predicting such patterns based on covariates would be challenging. For example, a mixture of two normal distributions to model outmigration timing for a situation with two defined outmigration peaks would require estimating five parameters (i.e., two means, two standard deviations, and a mixture term). It would challenging to develop covariate models to jointly predict how all five or most of these terms vary with covariate values. Alternate non-parametric approaches to modeling outmigration timing should also be considered for future analyses. For example, historical outmigration-timing patterns could be used to define a library of possible timings. Ideally, covariates could then be used to predict which shape from the library to use in a forecast. Until the requirements of the JPE are better defined by decision-makers, it is difficult to determine if a model that captures subtleties of outmigration timing would be more useful for management.

Additional analysis of covariate effects on median run date and run steepness of the beta need to be conducted. We evaluated a single covariate on one term of the timing model as a demonstration, and there are very likely other flow-based covariates, or biological ones such as spawner abundance, that may be better predictors of median outmigration date. Using covariates to predict run steepness may lead to more-accurate models compared to ones predicting median run date. Improvements to covariate models could result in lower error in forecasts of abundance. However, keep in mind that covariates used in the model for forecasting need to be forecastable. Further work with hydrologic modelers is

needed to determine what aspects of flow and temperature can be forecasted in a manner that will prove useful as a predictor of outmigration timing. We expect a process will be instituted to regularly update this and other models used in the JPE forecast approach as new years of data are collected and new covariates are devised and tested, such as the workshop approach that was used to identify and develop covariates for testing in the stock-recruit analysis.

To date, we have not applied the in-season outmigration model to data from Feather River RST sites because the PLAD model predicting spring-run proportions has not been completed. A number of refinements to the current version of the Feather River PLAD model are required to handle both the large numbers of spring-run hatchery fish captured at RST sites, and the more complex run structure in this river. The challenges of modeling outmigration abundance for the Feather River, and plans for addressing those challenges, are described in more detail in Chapter 3.

5 References

Cordoleani, F, and J Korman. 2025. *Smolt survival submodel V2*.

Gelman, A, JB Carlin, HS Stern, and DB Rubin. 2004. *Bayesian Data Analysis*. Second Edition. Chapman & Hall/CRC, Boca Raton, FL. 668 p.

Stan Development Team. 2024. *Stan Users Guide*. Version 2.35. https://mc-stan.org.

Tables and Figures

Tables

Table 1. Summary Statistics of Precision of Log-transformed Annual Abundance Estimates

Summary statistics of the precision of log-transformed annual abundance estimates derived from the BT-SPAS-X and probabilistic length-at-date (PLAD) models for spring-run Chinook salmon (*Oncorhynchus tshawytscha*) (spring-run) at eight rotary screw trap (RST) sites.

RST Site	Years of Data	CV Min	CV Mean	CV Max
Battle Creek	22	0.11	0.32	0.49
upper Clear Creek	18	0.11	0.33	0.78
Deer Creek	12	0.18	0.24	0.33
Mill Creek	13	0.25	0.35	0.50
Butte Creek	19	0.17	0.41	0.80
Yuba River	7	0.14	0.17	0.21
Knights Landing	27	0.32	0.42	0.57
Tisdale	13	0.29	0.42	0.55

Notes:
CV = coefficient of variation

Table 2. Medians and Standard Deviations of Posterior Distributions of Parameters

Medians and standard deviations (in parentheses) of posterior distributions of parameters that determine annual outmigration timing and variation in outmigration timing across years. Results are shown for the version of the model with random variation in weekly deviates (Equation 4a) without (none) and with a covariate effect (peak flow prior to February) predicting interannual variation in the week when 50% of the total outmigrant population has passed the trapping site. Refer to the text of this chapter for a definition of model parameters.

Covariate Effect on ϕ	**RST Site**	$\bar{\phi}$	$\bar{\lambda}$	σ_{ϕ}	σ_{λ}	ρ	$\beta[1]$	σ_{ρ}	$inverse\ logit(\bar{\phi})$	$exp(\bar{\lambda})$
None	Battle	-0.4 (0.05)	3.84 (0.22)	0.18 (0.05)	0.93 (0.18)	-0.8 (0.14)	n/a	1.04 (0.15)	0.4 (0.01)	46.63 (10.57)
None	Clear	-0.93 (0.03)	5.08 (0.21)	0.1 (0.03)	0.64 (0.26)	0.21 (0.42)	n/a	0.87 (0.18)	0.28 (0.01)	160.94 (33.24)
None	Mill	0.1 (0.07)	2.72 (0.15)	0.12 (0.08)	0.29 (0.16)	0.26 (0.5)	n/a	1.23 (0.19)	0.53 (0.02)	15.25 (2.34)
None	Deer	0.18 (0.08)	2.48 (0.16)	0.1 (0.08)	0.36 (0.17)	-0.19 (0.52)	n/a	1.3 (0.15)	0.54 (0.02)	11.96 (1.93)
None	Butte	-0.23 (0.06)	4.02 (0.15)	0.2 (0.05)	0.51 (0.15)	-0.24 (0.34)	n/a	1.58 (0.15)	0.44 (0.01)	55.89 (8.31)
None	Yuba	-0.49 (0.11)	4.12 (0.36)	0.25 (0.13)	0.78 (0.43)	-0.43 (0.38)	n/a	0.98 (0.15)	0.38 (0.03)	61.3 (25.81)
None	Knights	-0.35 (0.05)	3.77 (0.12)	0.09 (0.05)	0.24 (0.14)	-0.18 (0.58)	n/a	2.26 (0.17)	0.41 (0.01)	43.29 (4.99)
None	Tisdale	-0.34 (0.09)	3.75 (0.24)	0.17 (0.12)	0.58 (0.26)	-0.34 (0.5)	n/a	2.25 (0.25)	0.42 (0.02)	42.61 (10.69)
Flow	Battle	-0.4 (0.05)	3.86 (0.23)	0.19 (0.05)	0.94 (0.18)	-0.84 (0.12)	-0.05 (0.03)	1.03 (0.14)	0.4 (0.01)	47.29 (11.09)
Flow	Clear	-0.93 (0.03)	5.06 (0.22)	0.1 (0.03)	0.63 (0.24)	0.18 (0.42)	-0.03 (0.03)	0.94 (0.16)	0.28 (0.01)	157.59 (33.89)
Flow	Mill	0.1 (0.08)	2.72 (0.15)	0.1 (0.08)	0.29 (0.16)	0.2 (0.51)	-0.08 (0.06)	1.25 (0.18)	0.53 (0.02)	15.24 (2.25)

Covariate Effect on ϕ	**RST Site**	$\bar{\emptyset}$	$\bar{\lambda}$	σ_{ϕ}	σ_{λ}	ρ	$\beta[1]$	σ_{ρ}	$inverse\ logit(\bar{\emptyset})$	$exp(\bar{\lambda})$
Flow	Deer	0.17 (0.08)	2.49 (0.15)	0.09 (0.08)	0.32 (0.17)	-0.24 (0.52)	-0.1 (0.06)	1.31 (0.15)	0.54 (0.02)	12.1 (1.85)
Flow	Butte	-0.23 (0.06)	4.03 (0.15)	0.2 (0.06)	0.5 (0.15)	-0.28 (0.34)	0.04 (0.05)	1.61 (0.16)	0.44 (0.02)	56.08 (8.13)
Flow	Yuba	-0.48 (0.06)	4.1 (0.39)	0.11 (0.1)	0.86 (0.52)	0.25 (0.52)	-0.19 (0.08)	0.97 (0.14)	0.38 (0.02)	60.08 (27.03)
Flow	Knights	-0.33 (0.05)	3.72 (0.13)	0.08 (0.05)	0.32 (0.14)	-0.19 (0.5)	-0.04 (0.05)	2.23 (0.17)	0.42 (0.01)	41.44 (5.6)
Flow	Tisdale	-0.32 (0.09)	3.72 (0.26)	0.18 (0.12)	0.64 (0.27)	-0.33 (0.48)	-0.04 (0.08)	2.22 (0.23)	0.42 (0.02)	41.45 (10.96)

Table 3. Out-of-sample Accuracy of Forecasted Annual Juvenile Outmigrant Abundance Based on Models Without and With Covariate Effect on Median Run Date

Out-of-sample accuracy of forecasted annual juvenile outmigrant abundance of spring-run based on models without (null) and with (pfpf, peak flow prior to February) a covariate effect on median run date. Accuracy statistics show the median of annual out-of-sample error as absolute value of predicted-observed (abs) and relative error as 100*absolute value of predicted-observed)/observed (%). Yellow cells indicate the null model had a relative error that was at least 10% lower than the covariate model. Orange cells indicate that the covariate model had a relative error that was at least 10% lower than the null model.

Site	Week	abs null	abs pfpf	% null	% pfpf
Battle	28-Dec	91	92	436	426
Battle	1-Feb	13	14	55	52
Battle	1-Mar	9	9	35	34
Battle	29-Mar	8	8	31	30
Clear	28-Dec	10	10	34	34
Clear	1-Feb	8	8	29	30
Clear	1-Mar	8	8	29	29
Clear	29-Mar	8	8	29	29
Mill	28-Dec	253	303	797	919
Mill	1-Feb	151	150	316	346
Mill	1-Mar	57	64	139	139
Mill	29-Mar	27	30	61	62
Deer	28-Dec	1,448	2,049	660	710
Deer	1-Feb	950	1,002	304	336
Deer	1-Mar	422	429	147	152
Deer	29-Mar	220	220	72	79
Butte	28-Dec	28,262	18,747	2,730	2,500
Butte	1-Feb	1,888	1,496	173	163
Butte	1-Mar	638	593	48	46
Butte	29-Mar	454	458	37	36
Yuba	28-Dec	19,384	31,318	281	550
Yuba	1-Feb	2,831	3,921	37	60
Yuba	1-Mar	1,594	1,723	20	21
Yuba	29-Mar	1,294	1,251	17	17
Knights	28-Dec	30,964	27,453	905	858
Knights	1-Feb	2,902	2,712	91	90
Knights	1-Mar	1,433	1,330	42	44

Site	Week	abs null	abs pfpf	% null	% pfpf
Knights	29-Mar	1,157	1,129	36	36
Tisdale	28-Dec	16,018	14,112	910	807
Tisdale	1-Feb	1,514	1,649	86	87
Tisdale	1-Mar	804	922	49	48
Tisdale	29-Mar	629	661	40	38

Figures

Figure 1. Comparison of Observed Cumulative Outmigration Timing with Predictions from Model

Comparison of observed (BT-SPAS-X) cumulative outmigration timing (black points) with predictions from the model (random weekly variation in outmigration timing, Equation 2a) at eight RST sites. Each panel shows the data and fit for a single outmigration year (November 1 in yr 't-1' to October 31 in year 't'). The gray line and shaded gray area show the median estimate of outmigration timing and 95% credible interval generated from Equation 1 (labeled "beta+error"). The red line shows predictions of run timing based on the beta distribution only (labeled "beta"). The blue line is the average outmigration timing across years as determined by the estimated by means of the multivariate hyper-distribution of outmigration-timing parameters (labelled "hyper," Equation 5). The value in parentheses as the top of each panel is the average coefficient of variation in cumulative abundance estimates across all weeks from BT-SPAS-X.

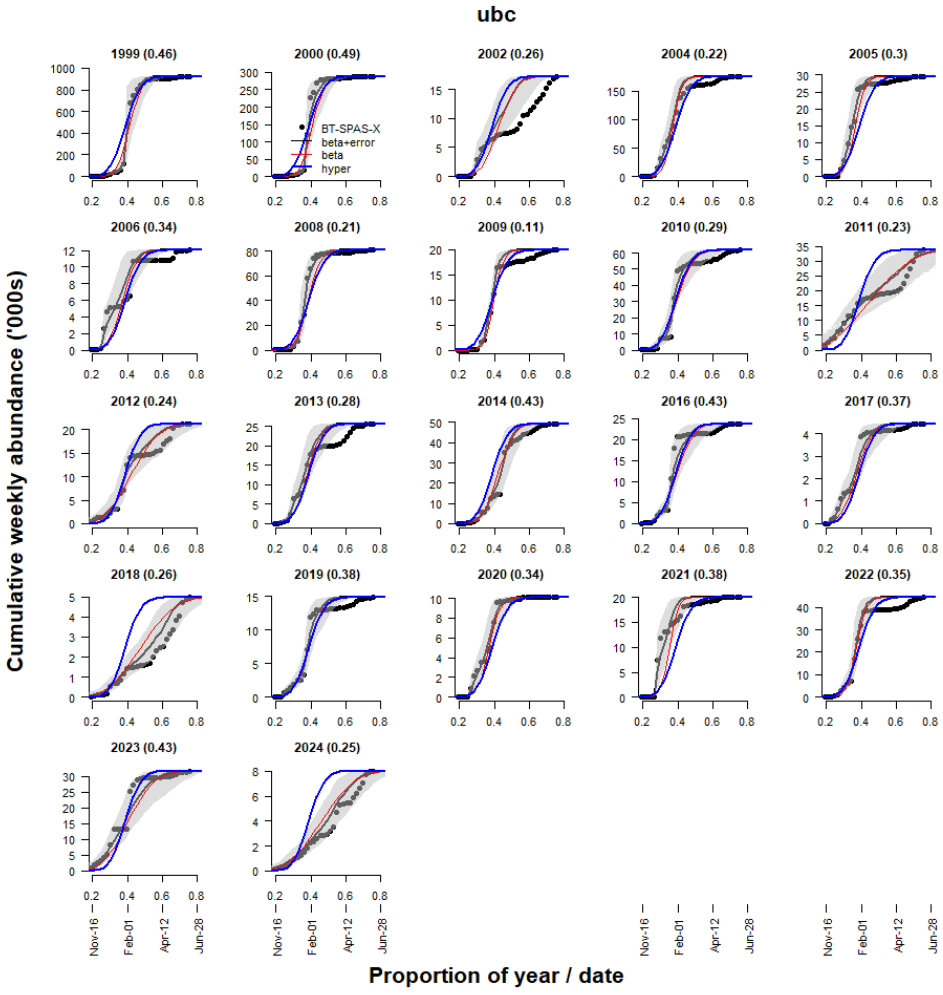


Figure 1, Continued

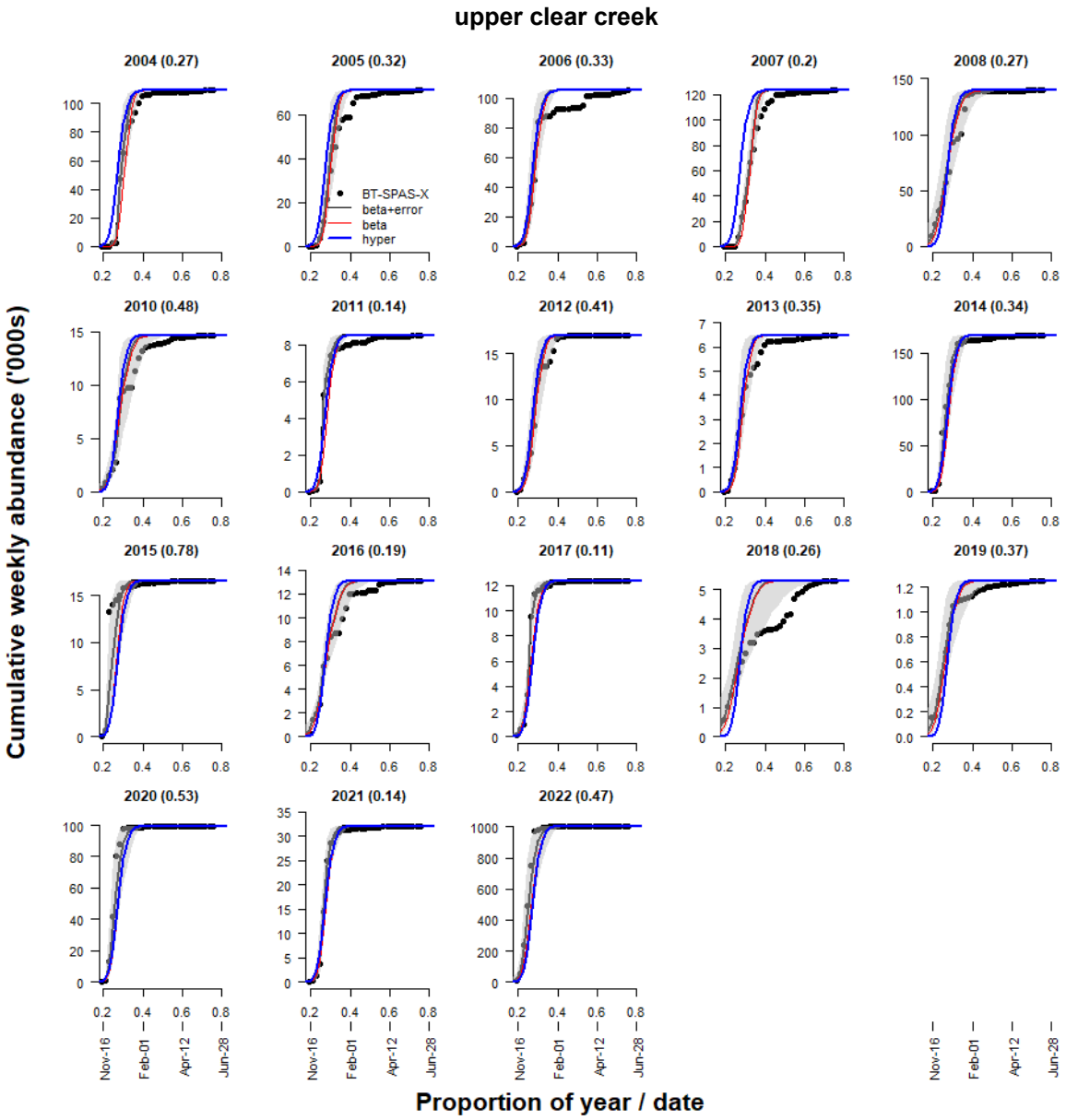


Figure 1, Continued

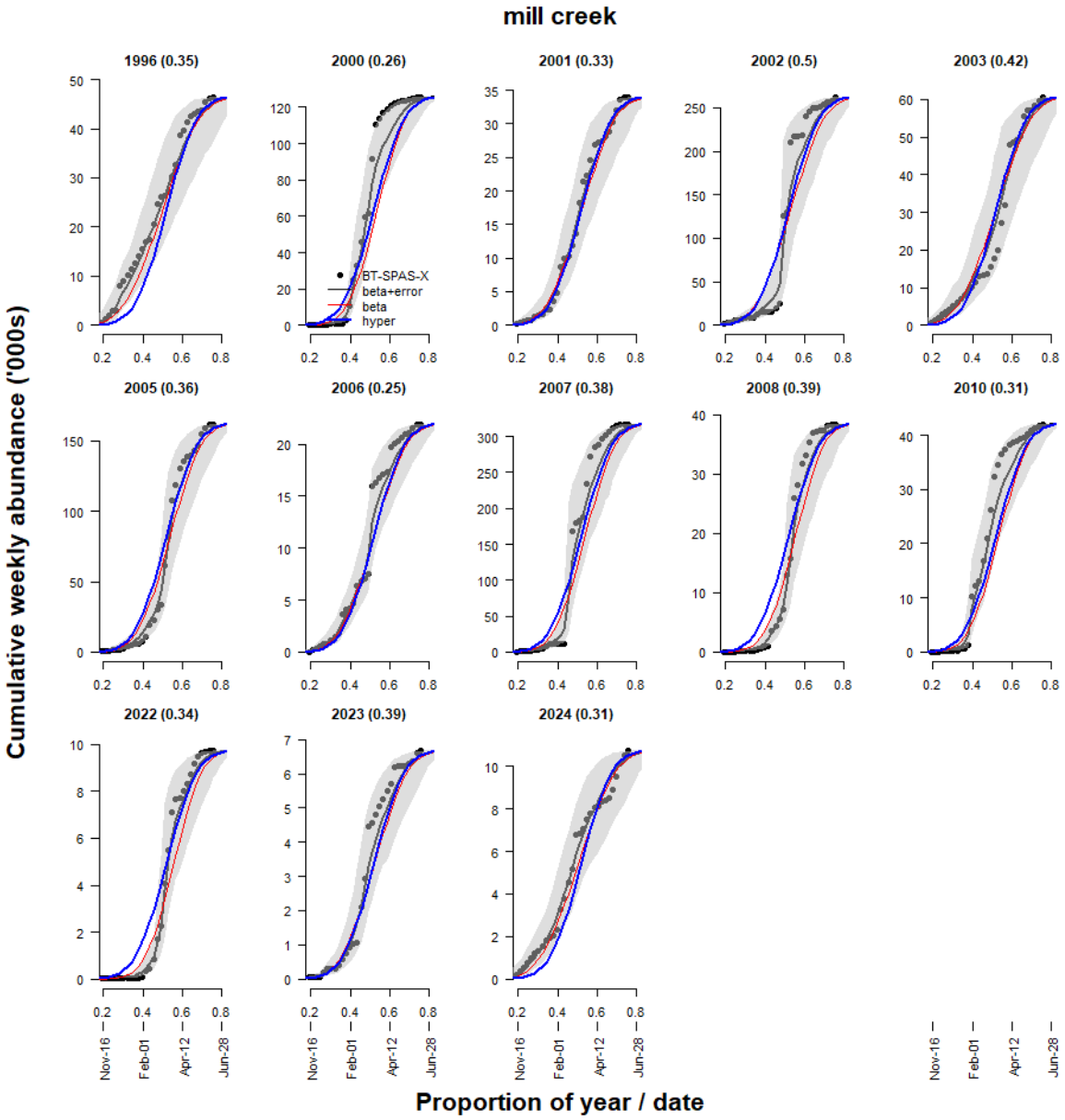


Figure 1, Continued

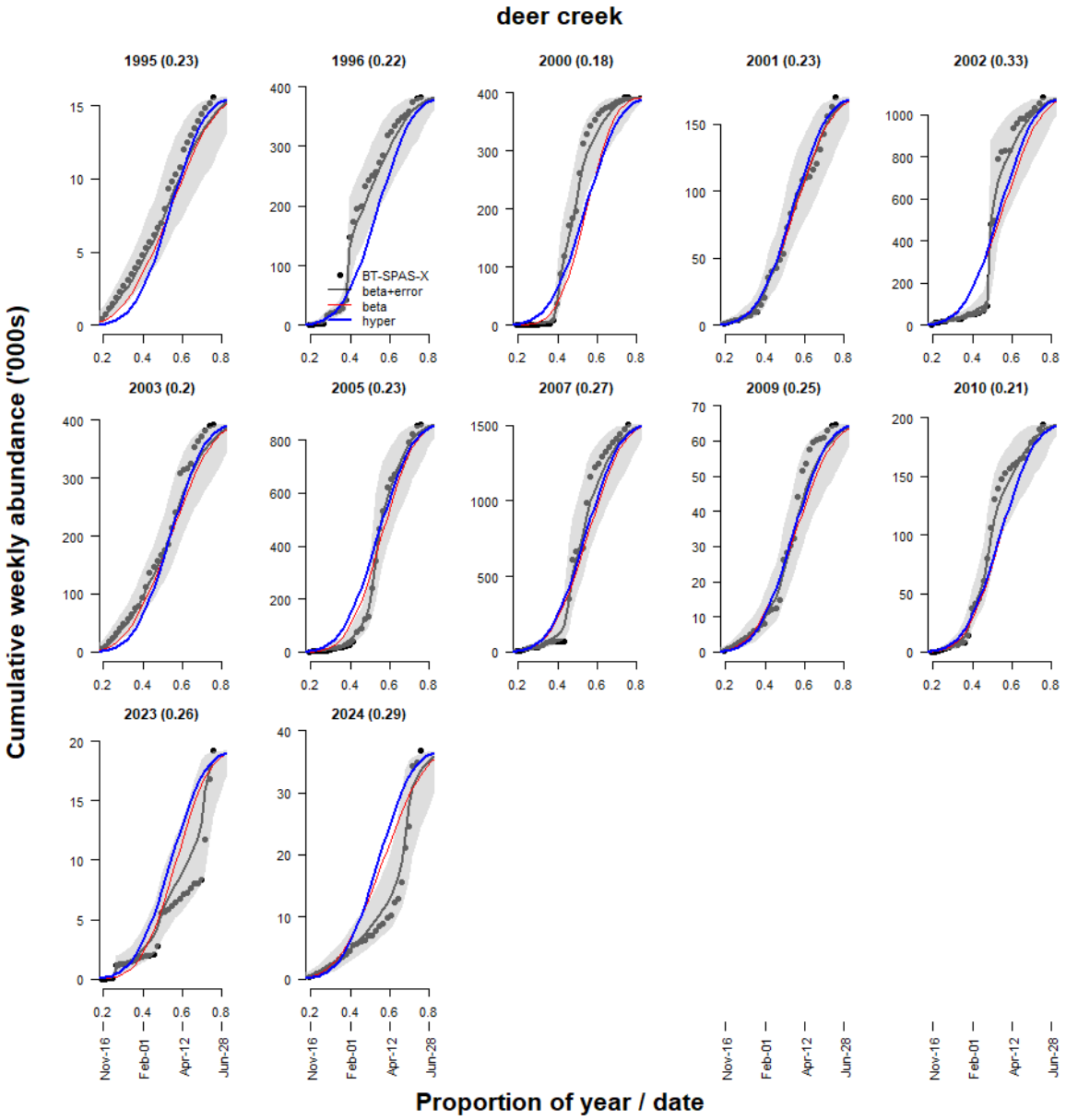


Figure 1, Continued

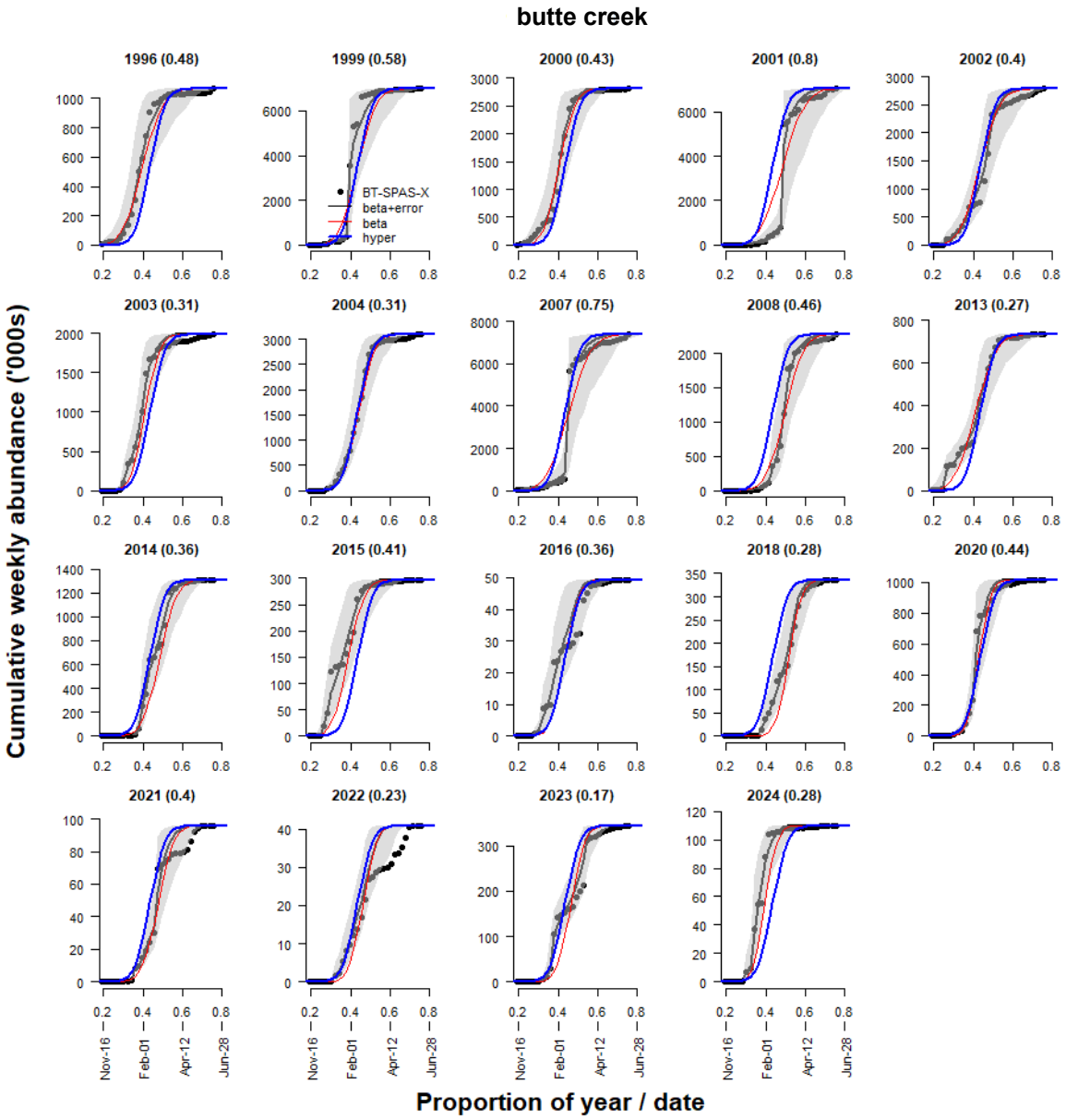


Figure 1, Continued

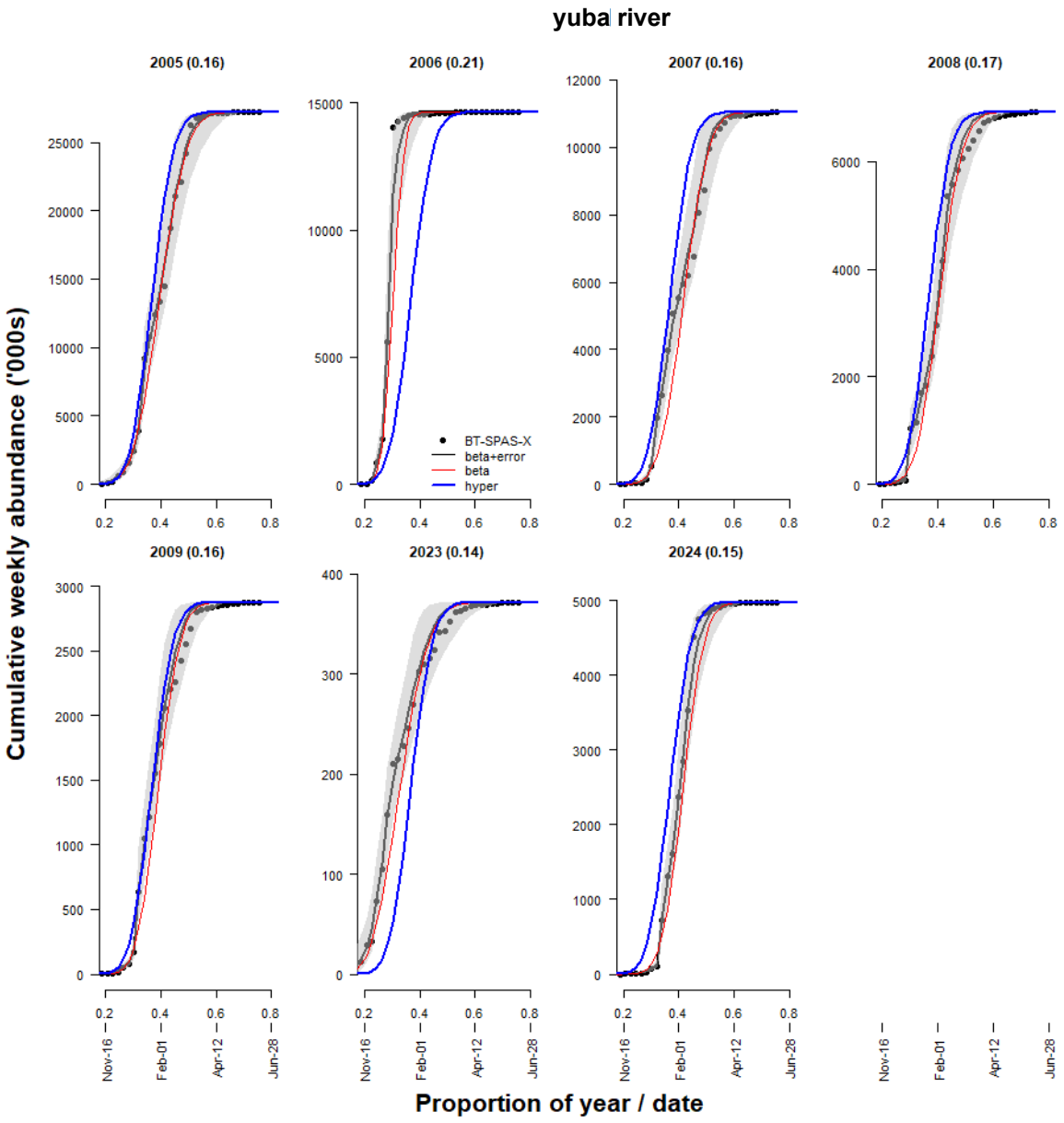


Figure 1, Continued

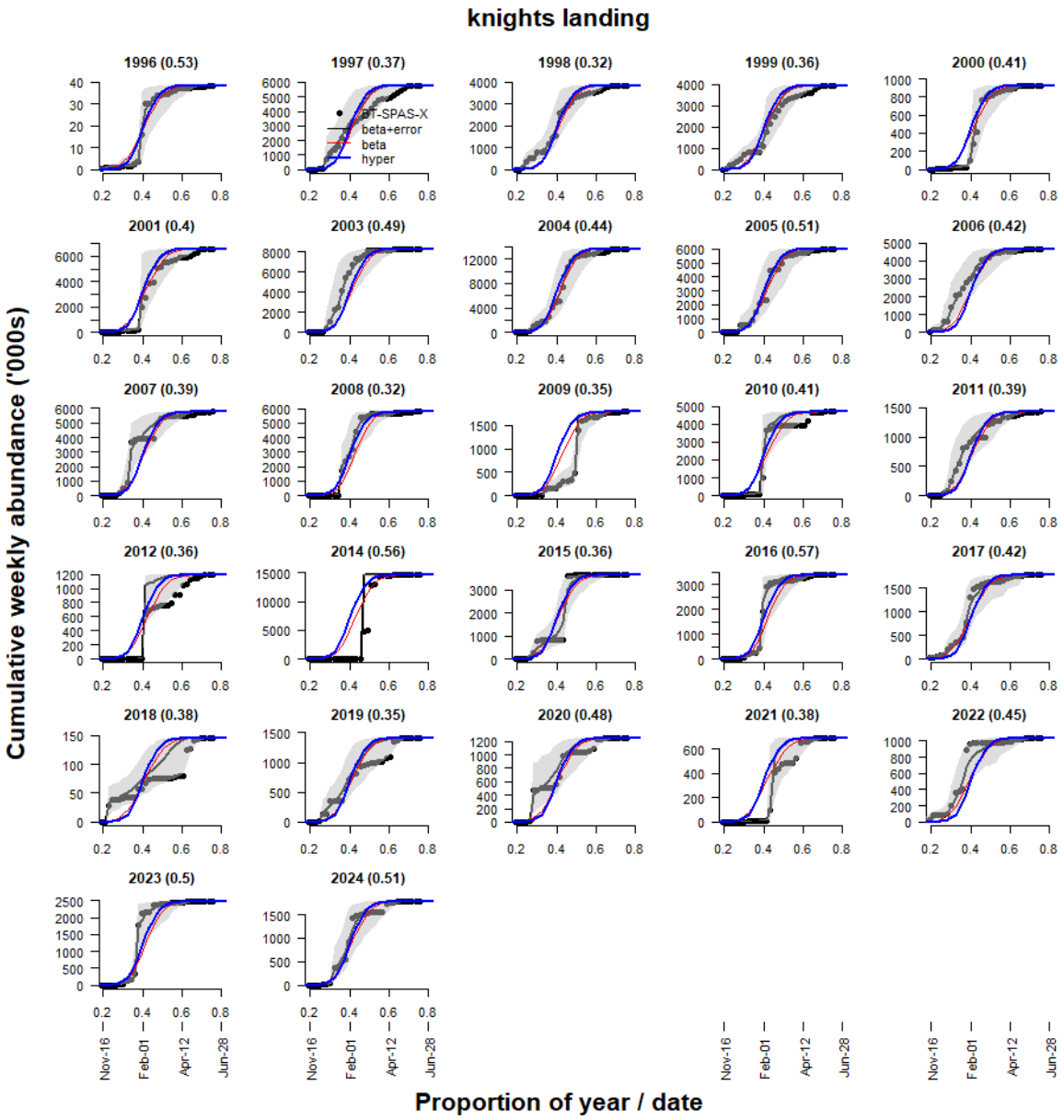


Figure 1, Continued

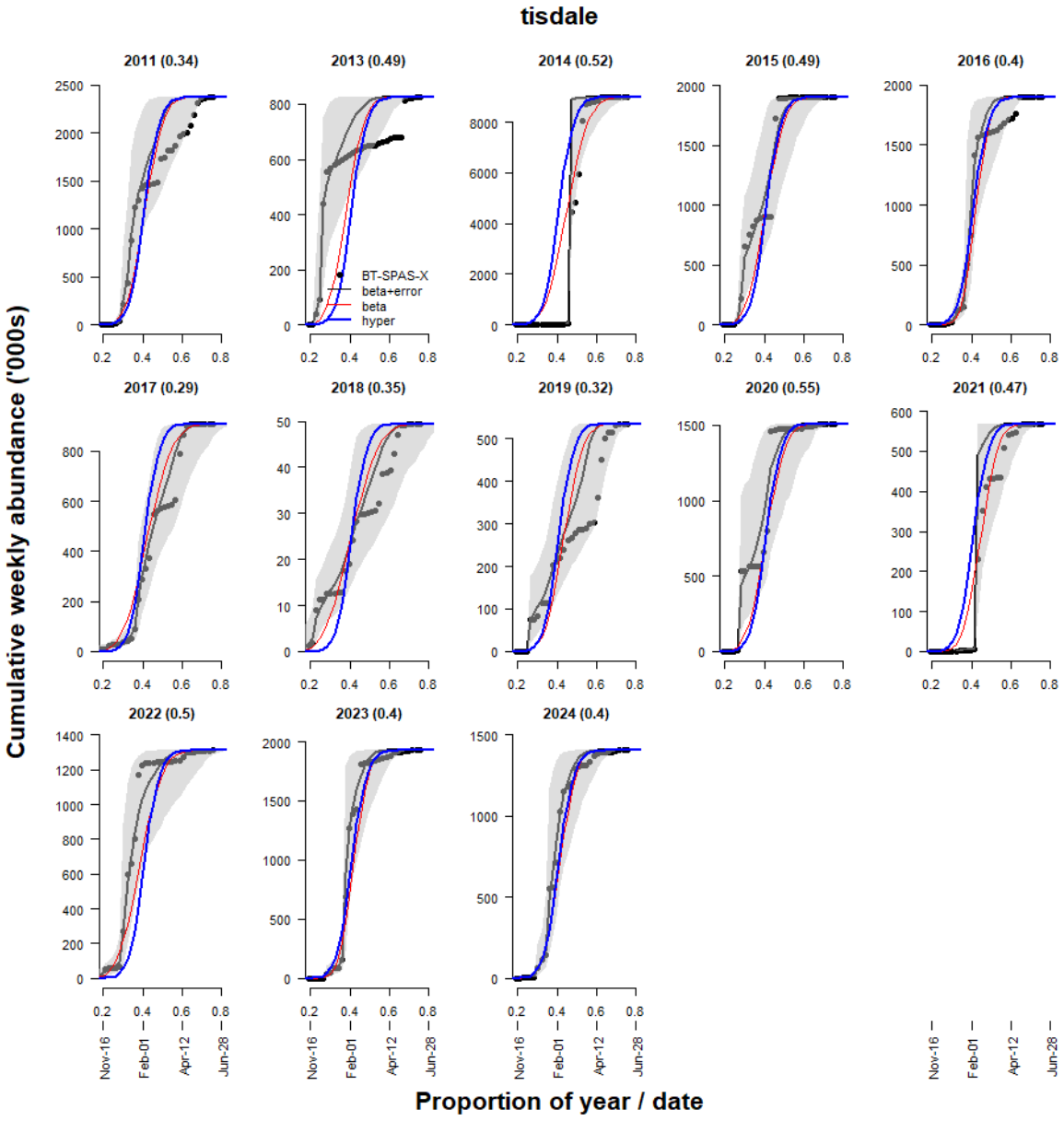


Figure 2. Relationship Between Annual Estimates of Median Outmigration Date and Rate of Abundance Increase During Outmigration

Relationship between annual estimates (points) of the median outmigration date (phi from Equation 4a) and the rate that abundance increases over the course of the outmigration (lambda, outmigration-timing steepness) for eight RST sites. Labels beside each point identify the year of outmigration (November through December in year "t-1" and January through June in year "t"). The red cross shows the across-year means for phi $(\overline{\phi})$and lambda $(\overline{\lambda})$ from Equation 5 (transformed to linear space). The red solid and dashed lines show the 80% and 50% credible intervals of the multivariate distribution for phi and lambda (Equation 5).

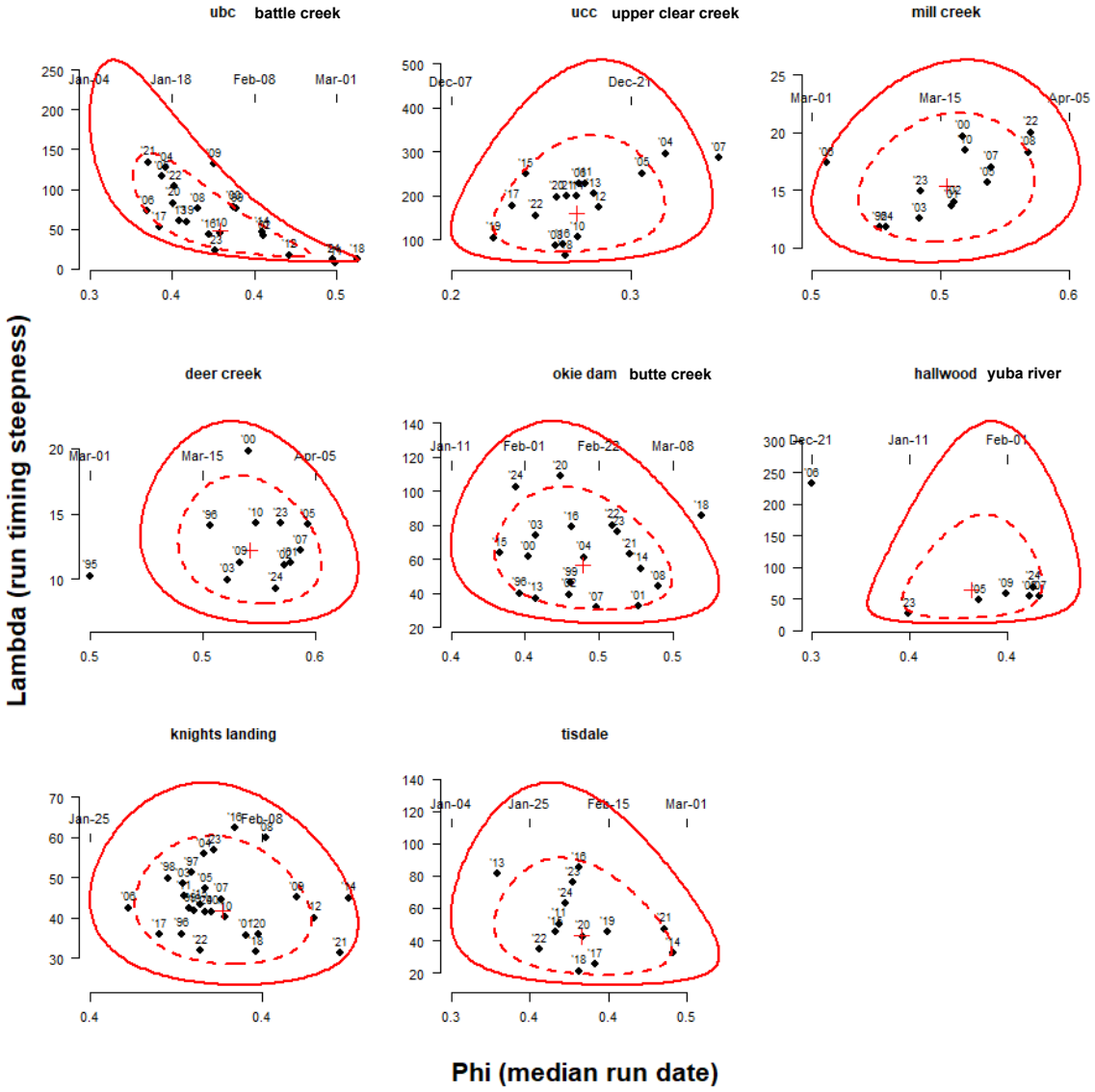


Figure 3. Effect of Peak Flow Prior to February on Median Outmigration Date

The effect of peak flow prior to February on the median outmigration date. The lines and shaded area show the median and 95% credible interval of the relationship.

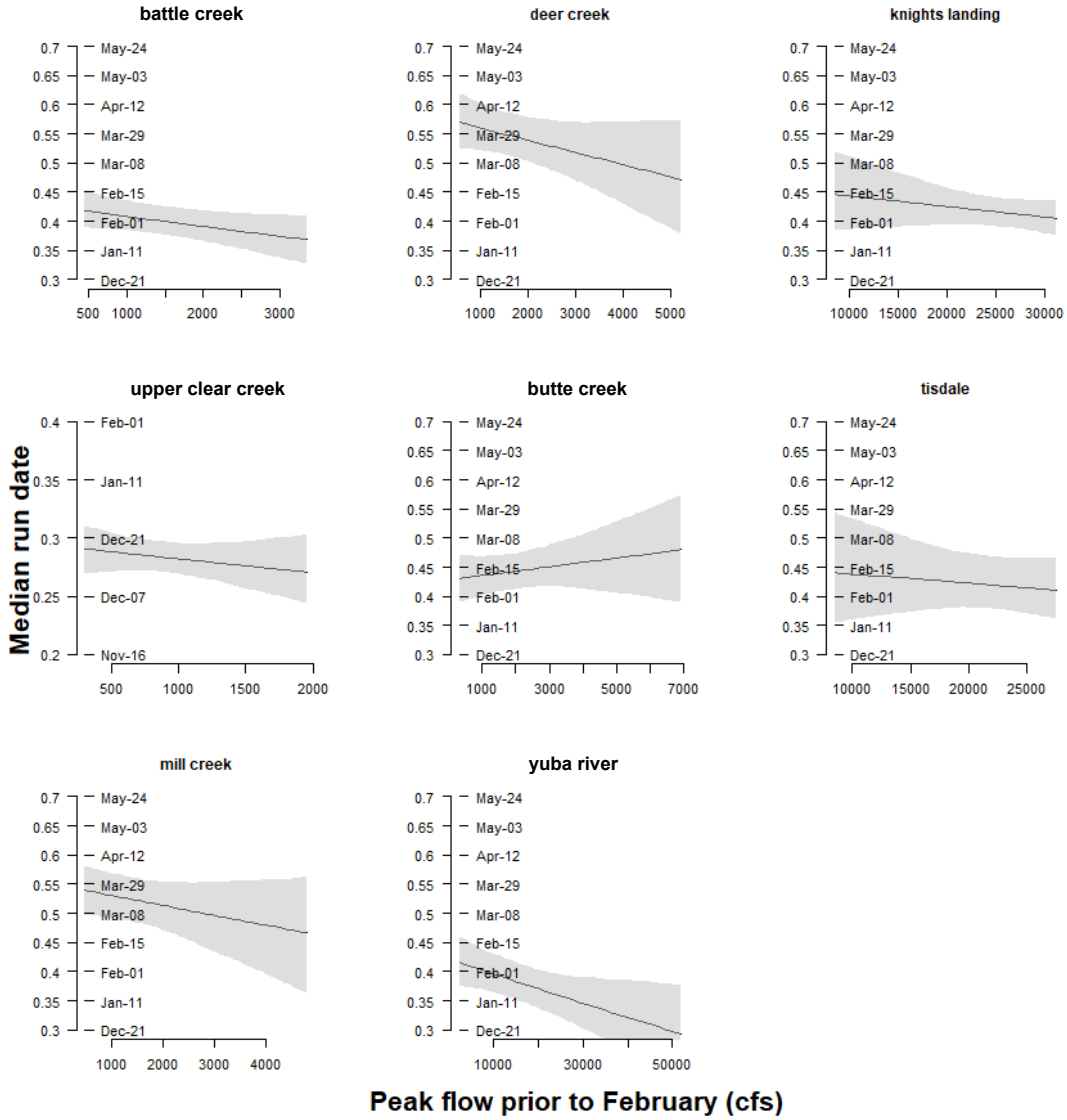


Figure 4. Forecasts of Outmigration Timing from Random Weekly Deviate Model Without Covariate Effects

Forecasts of outmigration timing from the random weekly deviate model (Equation 2a) without covariate effects. The lines and shaded areas show the median and 80% credible intervals. Vertical lines identify the first date of the weeks when abundance is forecast (refer to Figure 6).

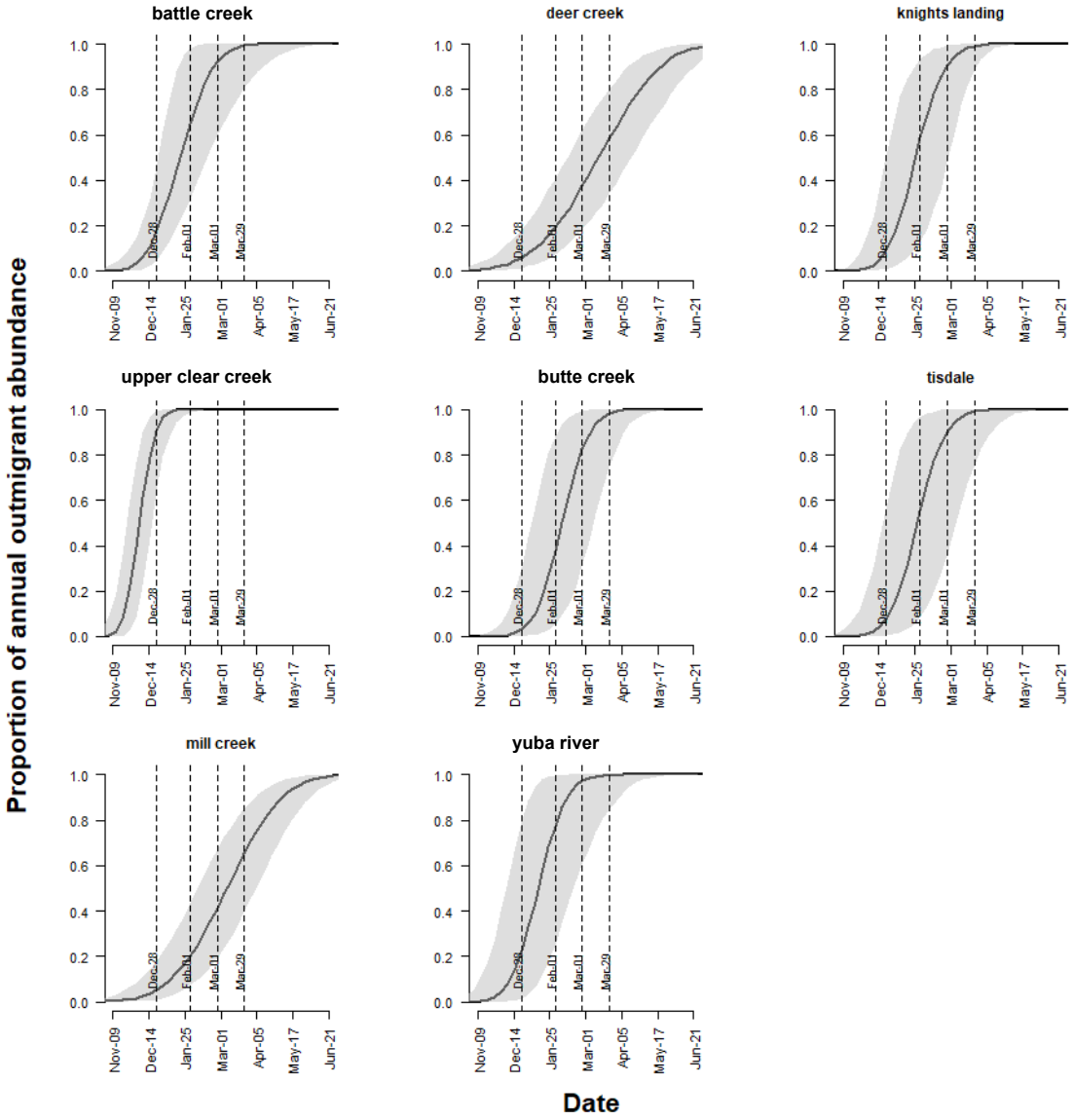


Figure 5. Forecasts of Annual Juvenile Outmigrant Abundance from Random Weekly Deviate Model Without Covariate Effects

Forecasts of annual juvenile outmigrant abundance from the random weekly deviate model (Equation 2a) without covariate effects. The points and error bars show the median and 80% credible intervals, respectively. Annual abundance is forecasted based on the cumulative abundances up to the four separate forecast weeks that are shown on the x-axis. For this example, the historical average cumulative abundance up to each of the four weeks (top row of text) were used as inputs to predict the annual outmigrant abundance forecast.

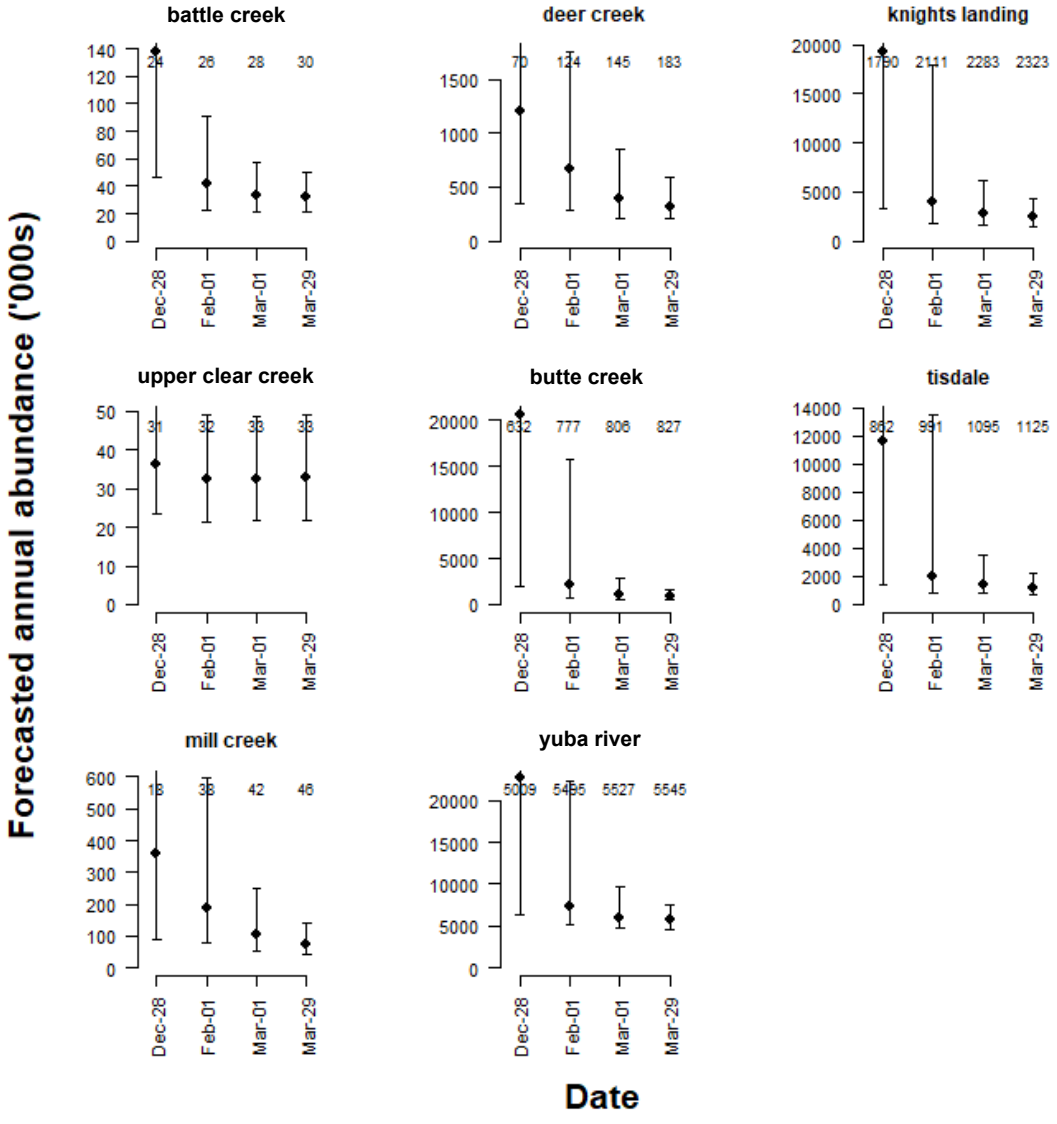


Figure 6. Comparison of Forecasted Run-timing for Models Without Covariate Effect and Random Weekly Errors or Lag-1 Autocorrelated Weekly Errors

Comparison of forecasted outmigration run-timing for models without a covariate effect and random weekly errors (solid blue lines and blue shaded areas) or lag-1 autocorrelated weekly errors (dashed red lines and red shaded areas) in outmigration timing. Results are only shown for RST sites without convergence problems for the lag-1 model.

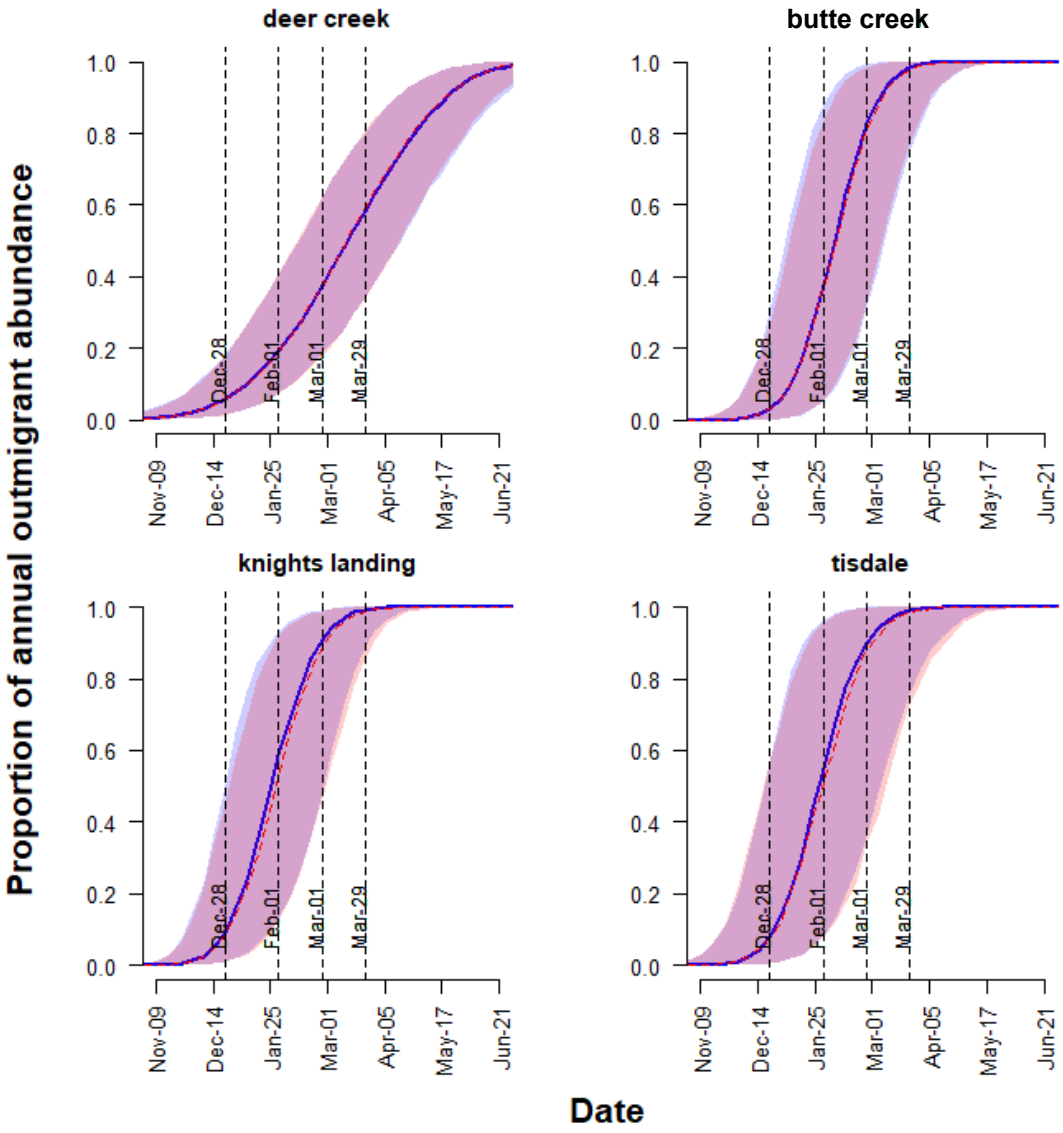
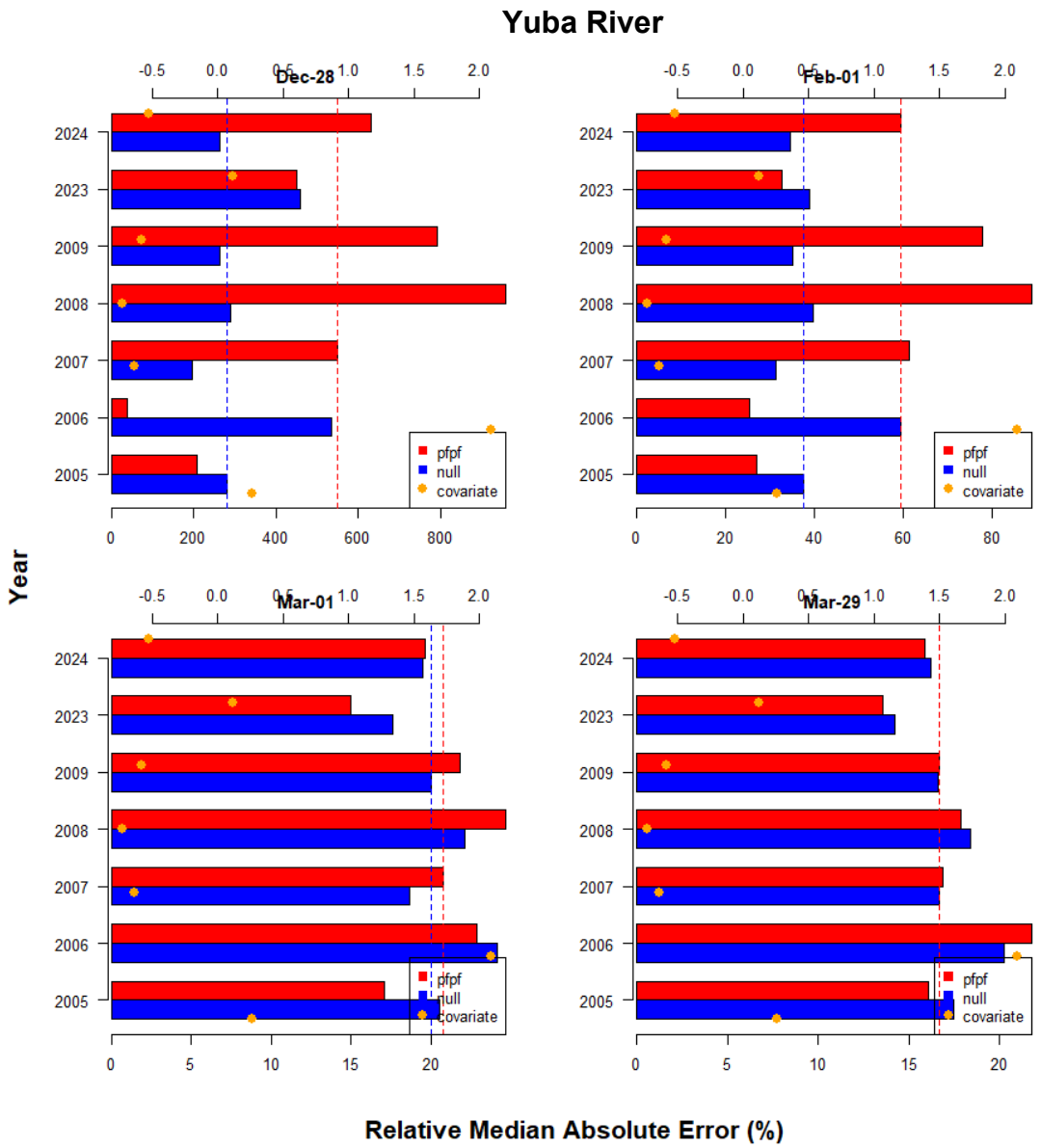


Figure 7. Out-of-Sample Relative Accuracy of Forecasted Annual Juvenile Abundance

Out-of-sample relative accuracy (100*abs(predicted-observed)/observed) of forecasted annual juvenile outmigrant abundance of spring-run for the Yuba River RST site based on models without (null) and with a covariate effect (pfpf, peak flow prior to February) on median run date. Horizontal bars show the annual values of relative error for the years left out of the fitting, and dashed vertical lines show the across-year median relative errors (shown in Table 3). Orange points and the upper x-axis show the covariate values (standardized peak flow prior to February). Plots of year-specific out-of-sample error for all sites are provided in Appendix B.



Appendices

A. Plots of Uncertainty in Cumulative Weekly Abundance of Outmigrating Juvenile Spring-run Chinook Salmon

Figure A-1. Estimated Median Weekly Outmigrant Cumulative Timing and Uncertainty

Estimated median weekly outmigrant cumulative timing (black points) and uncertainty (95% credible intervals) determined from BT-SPAS-X. Each panel shows the cumulative weekly abundance for a single outmigration year (November 1 in yr “t-1” to October 31 in year “t”). Values in parentheses in the title for each panel show the coefficient of variation of the annual outmigration abundance estimate.

Refer to “OutRun_Uncertainty.pdf.”

B. Plots of Out-of-sample Prediction Error in Forecasts of Annual Abundance of Outmigrating Juvenile Spring-run Chinook Salmon

Figure B-1. Annual Out-of-Sample Relative Error in Forecasts

Annual out-of-sample relative error (100*abs(pred-obs)/obs) in forecasts of abundance of outmigrating juvenile spring-run Chinook salmon by forecast week (panels). Forecasts were based on models without (null) and with (pfpf, peak flows prior to February) covariate effects on median run date. Vertical dashed lines show the across-year medians of relative error.

Refer to “OutSampleError.pdf.”