Sep 9, 2023

Hi Flora,

I coded an individual-based Sacramento River mainstem smolt survival model in stan using the acoustic telemetry data you sent. As we discussed, the first version of this model would repeat the reach\*year structure you used (survival and capture probability estimated for each unique reach and year combination = 64 parameters). The idea here was to make sure a Bayesian individual model reproduced the same results you got from the maximum likelihood model implemented in Rmark (with capture history data aggregated by year I think?).

Table 1 demonstrates good alignment between models. I compare results from my Year\_Reach model with those in Table 2 (the second Table 2) of your document SacSurvivalReport.docx. As you can see, we get very strong agreement between expected estimates and standard deviations when the samples sizes (recoveries at each receiver station) are high. The cases when model estimates diverge have low sample sizes. For example, woodson-butte and butte-sac reach survival estimates in 2015. In Table 2 of this document, you will see the detections by release group and station. Note very few detections at Sac and Delta stations in this case. Here, the information is much thinner, so the differences in estimation (Bayesian vs. maximum likelihood and assumed priors) result in different estimates. I may still have a few convergence issues in my model (I’m not chasing this down because I don’t think this structure is a keeper).

In the second part of Table 1 you can see good alignment between the combined woodson-butte and butte-sac survival rate. We see a bit of a difference in 2018 when there were few recoveries at Sac and Delta stations. So again, the differences make sense. 2015, 2018, and 2021 are pretty shaky owing to the really low sample sizes. This could be fixed up by making more restrictive assumptions about detection probability.

The other thing you will notice are the annual estimates of survival for the Sac-Delta reach. You didn’t include this in your results (logically) because the survival and detection probabilities are not statistically separable in this last reach. To some extent, the model can’t tell if survival was high in the reach and detection was low at the Delta station, or visa-versa. That is why my variance estimates are typically much higher for this reach (due to this confounding). However, the variances are not always bad (e.g., 2017 where cv=0.090/0873=0.1). When you get back in the game, we should discuss whether this Sac-Delta reach is included in the mainstem survival component of the JPE model. It is estimable in a few years when you have lots (2017, 2019) and modest (2013, 2016) numbers of detections at Sac and Delta receivers. Or it may be better to just assume that the Butte-Sac survival rate applies downstream to the point of the JPE estimate.

Figure 1 has the full set of results from Year\_Reach.stan. It’s just a graphical version of Table 1, but showing 95% credible intervals and the detection probabilities. You will note big variation in the extent of uncertainty in survival estimates among reaches. It’s all driven by the # of detections (Table 2). We are very certain about survival in the release-woodson reach owing to the large number of detections. But in low survival years, or years with very few releases, the number of detections at more downstream stations gets small, and this is reflected by very wide credible intervals. 2017 and 2019, wet years, we have more precise survival rates in downstream locations because survival was high, so there were more detections at downstream receivers. In years with low survival you will have high uncertainty in survival estimates for downstream reaches even if you release lots of fish because you still end up with low numbers of detections at the downstream receivers. Owing to this issue I think it is going to be challenging to estimate a flow (Or water year type) covariate effect by reach, but I will try.

I think we are in good shape to move on to developing new model structures (in stan) that would be more conducive for making predictions for the JPE model. As mentioned, I will move away from the reach\*year structure for survival and instead predict survival for each reach with a fixed water year effect and year-release group random effects. For now, I will still assume a year\*reach structure for detection probability. When making reach-specific predictions of mainstem survival, we would use the fixed water year effect and a forecast for water year type (or some other flow covariate if forecast are available) and add random error based on the variance term from the random year-release group effect. The detection probabilities are not needed for a forecast (they are really nuisance parameters) so there isn’t a problem using a year\*reach structure here. And I think the structure of the survival model will help with this very flexible detection model structure.

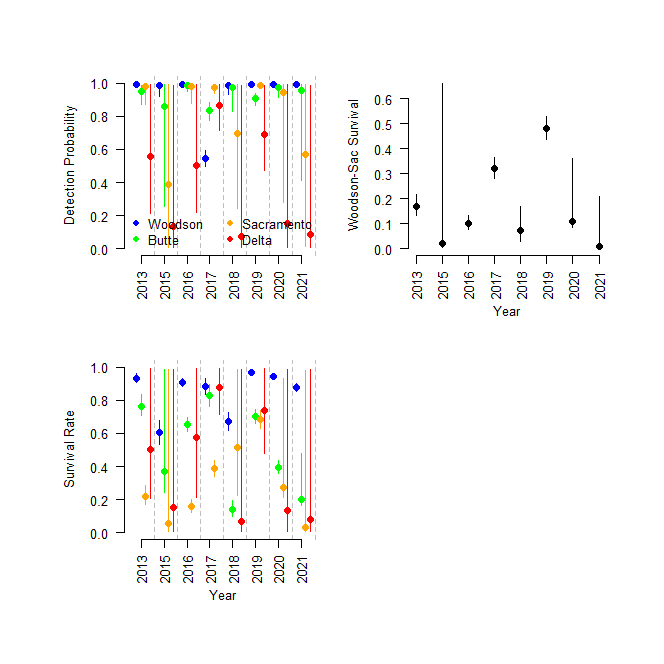
I’ll report on the new model in a separate email or probably a word document. I’ll try a few incarnations. We are all setup to look at the effects of the individual covariates you calculated. It was really efficient for me to get going on this because the data file you prepared had everything I needed and was super-well documented. Thank you!

**Table 1.** Mean and standard deviation in parameter estimates from Year\_Reach.stan individual Bayesian model compared to Flora’s Rmark year\*reach grouped maximum likelihood model. Yellow cells identify cases where estimates differ substantively across models. The rightmost column is the coefficient of variation in survival rate estimates for the Sacramento to Delta reach.



**Table 2.** Number of releases and detection by receiver location for each release group.





**Figure 1.** Reach-specific estimates of detection probability (top-left) and survival rate (bottom-left) by year based on the year\_reach.stan model. Also shown are the Woodson-Sacramento reach survival rates (woodson-butte \* butte-sacramento survival rates). Points show medians from posterior distributions end error bars show 95% credible intervals.