**Methods**

**Introduction:**

In Bayesian analyses, inference regarding model parameters is drawn from the posterior density. According to Bayes’ rule, the posterior density of an unknown parameter is proportional to the product of the likelihood and the prior density. Multiple Chain Monte Carlo (MCMC) methods can be used to approximate the posterior density after the likelihood and prior density have been formulated.

State space models refer to a class of models that describe the probabilistic dependence between latent (or hidden) state variables and observed measurements. In this modeling framework, if Z is a state variable (or vector of state variables) and X is an observation variable (or vector of observation variables), the joint likelihood can be factored as . The first component of the likelihood, , expresses the interdependencies among the state variables and will be referred to as the state process; whereas, describes how the observation variables relate to the state variables and will be referred to as the observation process.

The state-space model (Appendix XX) assumes a Ricker spawner-recruit relationship and time-varying maturity. It has an age-structured framework, which facilitates an accurate depiction of observation error in inriver abundance, age composition, and harvest. The model is fit to multiple sources of information on historical abundance, age composition, and harvest, which allows the model to simultaneously reconstruct historical abundance and obtain estimates of stock productivity. Uncertainty from the run reconstruction is passed through to the spawner-recruit analysis and subsequent reference points such as MSY and the escapement that provides for maximum sustained yield (*SMSY*). The model accommodates missing data, measurement error in the data, absolute and relative abundance indices, and changes in age at maturity. By constructing an integrated model, all relevant data are considered and weighted by their precision. Sensitivity analyses are conducted to assess robustness of the results to assumptions of the run reconstruction and spawner–recruit analyses and model selection procedures are used to determine which modeling assumptions are most appropriate.

**The Latent Process:**

In developing the mathematical framework of the latent process, let index the year. In this section, if the domain of is unspecified, it is assumed that . The upper bound of this set was chosen so that predictions can be made through the year 2023. Additionally, let be an index representing the river where corresponds to the Chena and corresponds to the Salcha. Next, let and be state variables where is defined to be the number of Chinook salmon that spawn in river in year and is defined to be the number of salmon recruited into the river population in year .

**Recruits given Spawners:**

Spawner-recruit (SR) models assume that a functional relationship exists between the number of spawners and the subsequent number of recruits. The most common SR relationships contain two parameters: describes the productivity of the stock at low abundances whereas governs the degree of density dependence. Different assumptions can be made on the dynamics of the density dependence. When the density dependence is assumed to be compensatory (the per capita production increases with a reduction in the number of spawners) the Beverton-Holt model is used; when the density dependence is over-compensatory (production decreases at high levels of spawning abundance), the Ricker model is used. The Beverton-Holt and Ricker model can be extended to account for process errors, environmental effects, and time varying productivity.

Because the true dynamics of the SR relationship are unknown and these relationships can be overparametrized, ten separate models with different SR relationships were constructed, and model selection techniques were used to determine the SR relationship with the best predictive ability (as measured by DIC). These models are described in more detail and the results of the model selection procedure are given in Appendix XX; however, a Ricker model of the general form

was ultimately selected. In this equation, is a white noise random error and is assumed to be distributed as and a non-informative prior density is placed on .

**Returners given Recruits:**

The number of Chinook salmon that returning to each river each year () was assumed to depend upon the number of recruits old enough to spawn and was modeled using a deterministic relationship (i.e. there was no process error). Let be a vector containing the probabilities that a fish survives to and matures at a given age (for ages 3-8) in year given that the fish belongs to the river population. In river r, the number of Chinook salmon returning to spawn at age-a in year prior to harvest was calculated as the product of the recruitment in year and the associated probability of survival/maturation:

for all . When this relationship cannot be used as is undefined.

The total number of Chinook salmon returning to river in year is then calculated as the sum of abundance at age across ages:

The probability of survival and maturity was modeled hierarchically: i.e., it was allowed to vary among cohorts to a specified extent. Here, was drawn from a where . The model was run when different prior distributions were placed on as a means of sensitivity testing. These results are detailed in appendix XX; however, in the final model, survival and age-at-maturity was assumed to vary over time. In this model, the Dirichlet parameters were expressed in an alternate form where . Here, is the (inverse) dispersion parameter. We assumed that for all , reflecting consistency of survival and age-at-maturity among brood years. A low value of D is reflective of a large amount of variability in survival and age-at-maturity. The location parameters are proportions that sum to one across age classes, reflecting the age-at-maturity central tendencies. To model time varying age at maturity, the location parameters were assumed to trend according to a multivariate logistic relationship, with each and denoting the logistic slope and intercept parameters associated with each river and age :

Weakly informative uniform priors were placed on , , and .

**In-River-Run-Abundance given Returners and Middle Yukon Harvest:**

The in-river-run-abundance was also modeled using a deterministic relationship. Here, the in-river-run-abundance () was assumed to be equal to the number of river specific returners minus the proportion () of midriver harvest () coming from each river. That is,

for all . For , The in-river-run-abundance was modeled hierarchically. Here, it was assumed that and non-informative priors were placed on and .

The midriver harvest and proportion of midriver harvest coming from each river were also modeled hierarchically. For the midriver harvest, it was assumed that , and non-informative hyperpriors were placed on and . For the proportion of midriver harvest coming from each river, it was assumed that is distributed for all . Afterwards, and were re-parameterized as +1 and , and weakly informative hyperpriors were placed on and .

**Spawners given In-River-Run-Abundance and River Specific Sport Harvest:**

Finally, the number of river specific spawners () was calculated by subtracting the river specific sport harvest () from the in-river-run-abundance. Here,

for all and . Again, the harvest was modeled hierarchically. It was assumed that , and non-informative hyperpriors were placed on and for all r.

**The Observation Process:**

Because the observed data is only available for a subset of the years for which the latent process is defined, it is henceforth assumed that the domain of is the set of all years for which the relevant data is available.

**Abundance Estimates:**

In developing the mathematical framework of the observation process, let the observation variables and be defined as annual estimates of river specific abundance obtained through mark-recapture and tower counting methods respectively. These estimates are assumed to be related to the state variables according to the following relationships:

where the multiplicative errors and are assumed to be distributed as

where and are defined in terms of the estimated coefficient of variation associated with the relevant mark-recapture and tower count abundance estimate. Specifically, it is assumed that

One benefit to the state-space modeling approach is that abundance

**Harvest Estimates:**

Next, let be an estimate of the total harvest in the Middle Yukon in year , and let be an estimate of the sport harvest coming from river in year . These estimates are assumed to be related to the true harvest according to the following:

Because coefficients of variation were not available for the harvest estimates, the errors associated with the harvest are assumed to be distributed as

and non-informative prior densities were placed on and .

**Movement Between the Middle Yukon and the Chena and Salcha:**

Results from a Yukon River Chinook salmon radiotelemetry study were used to apportion middle run harvest to the Chena and Salcha river stocks. The number of radiotagged fish in the middle Yukon and the number of radiotagged fish that moved to the Chena and Salcha river are available for three years (2002-2004). Letting denote the total number of radiotagged Chinook in the middle Yukon in year and denote the number of radiotagged Chinook that moved from the middle Yukon to river in year , the data is related to the proportion of fish that moved from the middle Yukon to each river () according to

Because the probability of movement from the Middle Yukon to the Chena and Salcha was modeled hierarchically, the 2002-2004 telemetry data will inform the probability of movement for the years when no data is available.

**Age Counts in the Chena and Salcha:**

Finally, the age counts in the Chena and Salcha are assumed to be related to the survival and maturity probabilities according to

where are the age counts for years 3-8 coming from river in year , and is the total number of fish whose ages were sampled in river and year .

**Biological Reference Points:**

Reference points were calculated for each individual MCMC sample. Spawning abundance providing maximum sustained yield was approximated by (Hilborn 1985):

Sustained yield at a specified level of S was obtained by subtracting spawning escapement from recruitment:

Other relevant quantities include harvest rate leading to maximum sustained yield, approximated by (Hilborn 1985)

Escapement leading to maximum sustained recruitment

And equilibrium spawning abundance, where recruitment exactly replaces spawners:

The quantity

In equations XX, XX, XX, and XX adjusts for the difference between the median and the mean of a right-skewed lognormal error distribution and the AR(1) process.

The probability that a given spawning escapement *S* would produce average yields exceeding X% of *MSY* was obtained by calculating *YS* at incremental values of *S* for each MCMC sample, and then comparing *YS* with X% of the value of *MSY* for that sample. The proportion *PY* of samples in which *YS* exceeded X% of *MSY* is an estimate of the desired probability, and the plot of *PY* versus *S* is termed an optimal yield probability profile (Fleischman et al. 2013).

The probability that yield would be reduced to less than X% of *MSY* by supplying too few spawners *S* was obtained by calculating *YS* at incremental values of *S* and tallying the number of MCMC samples for which *YS* was less than X% of *MSY* and *S* was less than *SMSY*. A plot of the fraction of samples in which this condition occurred versus *S* is termed an overfishing profile (Bernard and Jones III 2010).

The probability that a given *S* would produce average recruitments exceeding X% of maximum sustained recruitment (MSR) was obtained by calculating *R* at incremental values of *S* for each MCMC sample, then comparing *R* with X% of the value of *MSR* for that sample. The proportion *PS* of samples in which *R* exceeded X% of *MSR*, plotted versus escapement, is an optimal recruitment probability profile (Fleischman and Reimer 2017).

**Sampling from the Posterior Distribution**

MCMC samples were drawn from the joint posterior probability distribution of all unknowns in the model. For results presented here, 2 Markov chains were saved. Of these, the first 50,000 samples were discarded, and every 300th sample from 600,000 additional samples were used to estimate the marginal posterior medians, standard deviations, and percentiles. The diagnostic tools of RJAGS (Plummer 2013) within R (R Core Team 2016), including trace plots and the Gelman-Rubin statistic (Gelman and Rubin 1992), were used to assess mixing and convergence. Credibility interval estimates were constructed from the percentiles of the posterior distribution.

**For Model Selection Appendix:**

Simple Ricker:

Ricker with an AR(1) term:

where

Ricker with Time Varying Productivity:

Ricker with An Environmental Covariate:

where is the mean August discharge in river r and year y

Simple Beverton Holt: