I. AYKSSI TITLE PAGE

Arctic-Yukon-Kuskokwim Sustainable Salmon Initiative
Project Final Product[1]

RUN TIMING FORECAST MODELS FOR KUSKOKWIM RIVER CHINOOK SALMON

by:

Matthew J. Catalano[2] and Benjamin A. Staton[2]

School of Fisheries, Aquaculture, and Aquatic Sciences, Auburn University, Alabama

September 19, 2018

1 Final products of AYK Sustainable Salmon Initiative-sponsored research are made available to the Initiatives Partners and the public in the interest of rapid dissemination of information that may be useful in salmon management, research, or administration. Sponsorship of the project by the AYK SSI does not necessarily imply that the findings or conclusions are endorsed by the AYK SSI.

II. ABSTRACT

Management of fisheries for Pacific salmon Oncorhynchus spp. in western Alaska relies heavily on pre-season run forecasts to gauge the potential for harvest opportunities and schedule potential openers. Typically, run forecasts focus primarily on the anticipated size of the run in terms of numbers of returning fish and are commonly derived from sibling relationships or recent run averages and trends in productivity. However, another important aspect of a salmon run is the phenology (i.e., timing) with which it occurs. The broad-scale phenology of these runs is often fairly predictable in that they typically begin and end within the same several weeks each year. However, substantial uncertainty surrounds finer-scale characteristics of run timing including the date of peak run timing and the duration (i.e., how quickly the run escalates and terminates) as these qualities can vary substantially between years.

Uncertainty in these temporal qualities of the annual run makes in-season management difficult because the proportion of the run that has passed (and has yet to pass) is unknown on any given day of the run. Given a daily index of in-river abundance, such as a test fishery catch-per-unit-effort (CPUE), tower count, or sonar count, it is difficult to discern whether the run is larger or smaller than average without a priori information about year-specific run timing. Thus, the ability to predict run timing could improve in-season management of Pacific salmon by providing an estimate of how much of the run has passed, which would presumably increase the precision of in-season run size predictions. For this reason, a model that reliably forecasts run timing could largely aid efforts to develop in-season run size prediction models.

Runs that are earlier and smaller or later and larger than average are especially problematic to in-season management because both scenarios mimic averaged sized runs with average timing based on early in-season information. If a positive relationship exists between run timing and size, then these small/early and large/late runs will be more common. These scenarios have been found to be especially troublesome for the efficacy of the daily opening/closure management system to both meet (and not overshoot) escapement goals and to allow escapement (and harvest) to be dispersed proportionately throughout the run. These relationships should be investigated for other stocks in western Alaska so fishery managers are aware of their prevalence (or lack thereof).

In the Kuskokwim River, located in western Alaska, Chinook salmon return beginning in late May and continuing through early August. While run timing forecast models have been investigated for other salmon stocks in western Alaska, such as Bristol Bay sockeye salmon and Yukon River Chinook salmon, no published run timing models currently exist for Kuskokwim River Chinook salmon. Given the recent declines in Chinook salmon stocks, particularly in the Kuskokwim River, it is imperative that managers have access to tools for predicting run abundance during the fishing season. Based on the potential utility of run timing forecasts for in-season management, development of such models could prove valuable to Kuskokwim area fishery managers. Having these tools is especially important when run abundance is low so that managers can effectively guard against overharvest while safely allowing for subsistence harvests that are critical to the economy and culture of the region.

We investigated to what extent run timing could be forecasted for the Kuskokwim River Chinook stock using available environmental variables, and what extent incorporating run timing predictions would improve precision of run size estimates. We hypothesized that environmental and hydrologic factors that serve to impede or advance riverine entry in other systems should result in the same responses in the Kuskokwim River Chinook stock and that the value of these models to in-season management is high. Specifically, our objectives were to:

(1) Quantify associations between run timing and environmental variables,
(2) Assess relationships between run timing and run size,
(3) Evaluate the utility of a run timing forecast model to in-season run size predictions,
(4) Increase stakeholder familiarity with how run timing forecasts could be used by managers.

We met these objectives in two manuscripts, one that has been published in Fisheries Research and one that is in its second revision with the Canadian Journal of Fisheries and Aquatic Sciences. The first involved the development of an objective and adaptive statistical approach (using model-averaging and a sliding window algorithm to select predictive time periods, both calibrated annually) to deal with multidimensional selection of four climatic variables to predict run timing (Objective 1) based entirely on predictive performance. Forecast cross-validation was used to evaluate the performance of three forecasting approaches to predicting run timing: the null (i.e., intercept only) model that did not predict run timing from climatic variables, a single variables run timing prediction model with the lowest mean absolute error, and a model-averaged timing model across 16 nested linear models. As of 2016, the null model had the lowest mean absolute error (2.64 days), although the model-averaged forecast performed as well or better than the null model in the majority of retrospective years. The model averaged forecast had a consistent mean absolute error regardless of the type of year (i.e., average or extreme early/late) the forecast was made for, which was not true of the null model. The availability of the run timing forecast was not found to increase overall accuracy of in-season run assessments in relation to the null model, but was found to substantially increase the precision of these assessments, particularly early in the season. IN this analysis we also investigated relationships between run timing and run size (Objective 2) and found none.

In our second manuscript, we assessed the performance of two Bayesian information-updating procedures to predict the run size during the season (Objective 3): one that uses auxiliary run timing information and one that does not, and compared the performance to methods that did not involve updating. We found that in-season Bayesian updating provided more accurate run size estimates during the time when harvest decisions needed to be made, but that the incorporation of run timing forecasts had little utility in terms of providing more accurate run size estimates. The latter finding is conditional on the performance of the run timing forecast model we used; a more accurate timing forecast model may yield a different conclusion. The Bayesian approach we developed provided a probabilistic expression of run size beliefs, which could be useful in a transparent risk-assessment framework for setting and altering harvest targets in-season.

The information from these analyses was disseminated to stakeholders by integrating our work into the broader AYKSSI Capacity-Building Project (Objective 4). We attended these stakeholder meetings where we gave presentations on run timing and in-season management more broadly. We received thoughtful suggestions from stakeholders for improvement of our work and helped to build stakeholder capacity to engage with scientist and managers regarding this type of work.

III. PRESS RELEASE

Salmon return each year to their natal rivers to spawn, putting them within reach of people who rely on the runs for food, income, and cultural enrichment. Responsible management of the harvest of these salmon is an ever-present goal for these fisheries. Management of salmon harvest during the run is a difficult business because managers are never quite sure how many salmon will return to the river to spawn. They often rely on noisy indicators of run strength from test fisheries or verbal reports on what fishers are seeing on the river. Having some idea of the timing of the run is often desired because, for example, if it is known that that run is just getting started, then there should be many more fish on their way that could be harvested. Run timing is affected by environmental conditions such as temperature and ice cover. If these variables can measured before the run starts, then the timing of the run could be predicted, and ultimately the runs managed more effectively. A team of researchers from Auburn University has been studying run timing of Kuskokwim River Chinook salmon since 2016 to better understand whether run timing is predictable from these environmental variables and whether including run timing predictions into management leads to improved performance. The Auburn team build a complex statistical model to select which variables to use to predict timing, as well as when and where to measure those variables. They found that indeed, run timing is predictable from things like water temperature and ice cover, but they also found that when they included run timing into run size estimates and harvest management decisions, there was not much improvement in performance. It could be that the variables don't do a good enough job of predicting run timing when all of the other sources of variability that could affect the fishery are considered. Ultimately, the Auburn team hopes to provide fishery managers and stakeholders with better tools to help them move forward with decisions on when, where, and how many Chinook salmon to harvest in the Kuskokwim and beyond.

IV. PROJECT EVALUATION

The proposed project had four objectives as follows:

Objective 1: Quantify associations between run timing and environmental variables. We completed this objective as planned. In **Appendix A**, we investigated predictive relationships between run timing and environmental variables for Kuskokwim River Chinook salmon. First, we quantified annual run timing at the Bethel gillnet test fishery. Then we investigated predictive relationships by fitting regression models between run timing quantities (median run date and duration) and potential environmental drivers of riverine entry. The environmental variables included air temperature, sea surface temperature, ice cover, and the PDO index. The models were dynamic and annually adjusted the weighting factors on the predictors, the time window for their inclusion, and the locations at which they are measured based on predictive performance. We found relationships between run timing and environmental factors, but the resulting run timing predictions performed no better at predicting end of season test fishery CPUE, on average, than the mean run timing prediction.

Objective 2: Assess relationships between run timing and run size. This objective was completed as planned. In **Appendix A**, we assessed whether run timing was predictable from run size for Kuskokwim River Chinook salmon and found no associations. We did not pursue this issue further in subsequent analyses.

Objective 3: Evaluate the utility of a run timing forecast model to in-season run size predictions. We completed this objective as planned. We used a retrospective analysis to evaluate the degree to which using run timing forecasts would have increased the accuracy of in-season run size predictions for Kuskokwim River Chinook salmon for the years 1995 – 2017. We also broadened the scope of the analysis to assess whether Bayesian updating of run size predictions during the run as new observations are available would perform better than using either a pre-season forecast or in-season test fishery predictions alone. To accomplish this objective we developed a probabilistic Bayesian run prediction tool that we took beyond this objective by introducing it to in-season managers. The tool was used during the 2018 run. We report on this objective in **Appendix B**.

Objective 4: Increase stakeholder familiarity with how run timing forecasts could be used by managers. We shared our findings with area stakeholders and biologists at the November 2017 Capacity Building Workshop in Bethel. At the meeting we presented our findings on the predictive performance of run timing forecast models and their utility for in-season run size predictions via oral presentation. In addition, we incorporated run timing into the development of a probabilistic Bayesian run size updating tool. Specifically, we included an option in which users can specify whether the run timing is expected to be early, average, or late. Selecting these different options affects the model's interpretation of test fishery data and this impacts the magnitude of run size forecasts. For example, expecting an early run would suggest that there is less of the run yet to come and thus smaller overall run than would be predicted from a late arriving run, given identical in-season data. The run prediction tool was made available via an online application that we developed, which allowed stakeholders and managers to access and use the tool throughout the run via a user-friendly interface. The tool can be accessed at https://bstaton.shinyapps.io/BayesTool/. The user manual can be accessed at https://bstaton.shinyapps.io/BayesTool_UserMan/. Technical documentation can be accessed at: https://bstaton.shinyapps.io/BayesTool_TechDoc/.

V. DELIVERABLES

The findings of our project have been and will continue to be disseminated via conference and management meeting presentations and peer-reviewed manuscripts. We have completed five presentations, attended tone meeting, and submitted two manuscripts for peer review publication with one having been published and the other under review. We have also developed an online computer application to allows stakeholders and managers consider run timing in making in-season run size predictions.

Reports:

Semiannual progress reports January 2017, July 2017, and January 2018.

Presentations:

Staton, B. A., M. J. Catalano, T. M. Farmer, A. Abebe, and F. S. Dobson. 2018. Development and evaluation of a Pacific Salmon migration timing forecast model for Kuskokwim River Chinook Salmon. Western Division of the American Fisheries Society Conference. Anchorage, Alaska.

Staton, B. A. and M. J. Catalano. 2018. Evaluation of several approaches to Bayesian updating of pre-season indicators of run strength in Pacific Salmon fisheries. Western Division of the American Fisheries Society Conference. Anchorage, Alaska.

Staton, B. A., M. J. Catalano, and T. Farmer. 2017. Development and evaluation of a run timing forecast model for Kuskokwim River Chinook salmon. AYKSSI Capacity Building Workshop, Bethel, Alaska.

Staton, B. A. and M. J. Catalano. 2017. Evaluation of several approaches to Bayesian updating of pre-season indicators of run strength in Pacific Salmon fisheries. American Fisheries Society Annual Conference. Tampa, Florida.

Staton, B. A., M. J. Catalano, T. Farmer, A. Abebe, S. Dobson. 2017. Climatic variable selection across space and time: development of a Pacific salmon migration timing forecast model. Southern Division of the American Fisheries Society Annual Conference. Oklahoma City, OK.

Manuscripts:

Staton, B. A. and Catalano, M. J. *Under review*. Bayesian information updating procedures for Pacific salmon run size indicators: evaluation in the presence and absence of auxiliary migration timing information. Canadian Journal of Fisheries and Aquatic Sciences. Decision: major revision. Submitted revised manuscript on 7/31/2018.

Staton, B. A., M. J. Catalano, T. M. Farmer, A. Abebe, F. S. Dobson. 2017. Development and evaluation of a migration timing forecast model for Kuskokwim River Chinook salmon. Fisheries Research 194:9-21.

Computer Applications Developed

Staton, B. A. and M. J. Catalano. A Bayesian risk assessment tool for in-season management of Kuskokwim River Chinook Salmon: a shiny app in program R. https://bstaton.shinyapps.io/BayesTool/.

Meetings Participated:

2017 AYKSSI Capacity Building Workshop. Bethel, Alaska.

VI. PROJECT DATA SUMMARY

Our analysis produced simulated data sets and parameter estimates from Bayesian and maximum likelihood assessment models. All model outputs are available upon request from the PI.

VII. APPENDIX: SUBMITTED OR DRAFT MANUSCRIPTS

Appendix A:

Staton, B. A., Catalano, M. J., Farmer, T. M., Abebe, A. and Dobson, F. S. 2017. Development and evaluation of a migration timing forecast model for Kuskokwim River Chinook salmon. Fisheries Research 194: 9-21.

Appendix B:

Staton, B. A. and Catalano, M. J. *Under review*. Bayesian information updating procedures for Pacific salmon run size indicators: evaluation in the presence and absence of auxiliary migration timing information. Canadian Journal of Fisheries and Aquatic Sciences. *Decision: major revision. Submitted revised manuscript on 7/31/2018.*

APPENDIX A

Fisheries Research 194 (2017) 9–21

Contents lists available at ScienceDirect

Fisheries Research

journal homepage: www.elsevier.com/locate/fishres

Development and evaluation of a migration timing forecast model for Kuskokwim River Chinook salmon

Benjamin A. Staton[a,b,*], Matthew J. Catalano[a], Troy M. Farmer[a], Asheber Abebe[c], F. Stephen Dobson[d]

[a] School of Fisheries, Aquaculture, and Aquatic Sciences, Auburn University, 203 Swingle Hall, Auburn, AL 36849, United States
[b] U.S Fish and Wildlife Service, Yukon Delta National Wildlife Refuge, P.O. Box 346, Bethel, AK 99559, United States
[c] Department of Mathematics and Statistics, 221 Parker Hall, Auburn University, Auburn, AL 36849, United States
[d] Department of Biological Sciences, 101 Rouse Hall, Auburn University, Auburn, AL 36849, United States

ARTICLE INFO

Handled by Prof. George A. Rose

Keywords:
Pacific salmon
Phenology
Arrival timing
Forecast
Climate variable selection

ABSTRACT

Annual variation in adult salmon migration timing makes the interpretation of in-season assessment data difficult, leading to much in-season uncertainty in run size. We developed and evaluated a run timing forecast model for the Kuskokwim River Chinook salmon stock, located in western Alaska, intended to aid in reducing this source of uncertainty. An objective and adaptive approach (using model-averaging and a sliding window algorithm to select predictive time periods, both calibrated annually) was adopted to deal with multidimensional selection of four climatic variables and was based entirely on predictive performance. Forecast cross-validation was used to evaluate the performance of three forecasting approaches: the null (i.e., intercept only) model, the single model with the lowest mean absolute error, and a model-averaged forecast across 16 nested linear models. As of 2016, the null model had the lowest mean absolute error (2.64 days), although the model-averaged forecast performed as well or better than the null model in the majority of retrospective years. The model-averaged forecast had a consistent mean absolute error regardless of the type of year (i.e., average or extreme early/late) the forecast was made for, which was not true of the null model. The availability of the run timing forecast was not found to increase overall accuracy of in-season run assessments in relation to the null model, but was found to substantially increase the precision of these assessments, particularly early in the season.

1. Introduction

In-season management strategies for Pacific salmon *Oncorhynchus* spp. fisheries rely heavily on indices of in-river abundance (e.g., test fisheries, sonar counts, etc.) to inform harvest control rules that attempt to attain the balance of meeting pre-determined escapement objectives while allowing adequate opportunity for harvest (Catalano and Jones, 2014). However, because indices of abundance are confounded by the phenology (i.e., timing) of the migration, their interpretation is very difficult in-season. For example, smaller-than-average index values early in the season could be due to either a small run with average timing or by a late large run, when interpreted in the context of historical years (Adkison and Cunningham, 2015). This ultimately leads to great uncertainty about how much of the incoming run has passed, which is a key piece of information that dictates fishery harvest opportunities. There exists no information in the current year's abundance index to inform the manager if (for example) 25% or 75% of the run has passed on any given day. Yet, depending which is true, the optimal management decision could be vastly different. Thus, in-season assessment typically involves some characterization of the variation in historical run timing to formulate a range of possible run size scenarios that could be representative of the current year's run size. However, given the amount of variation in historical run timing, these scenarios are rarely informative during the majority of the migration, when key harvest decisions are being made because the run scenarios may span all possible run sizes. As a result, the pre-season run size forecast remains the most precise piece of information for much of the season. If it were possible to predict the timing of the incoming run (e.g., earlier- or later-than-average) with some level of confidence, it could prove valuable for in-season assessment and decision-making by reducing uncertainty in run size predictions.

While previous research has uncovered several key physiological mechanisms that are involved with natal homing (Hasler and Scholz, 1983) and return migrations of adult salmon to freshwater environments (Cooperman et al., 2010; Cook et al., 2008; Hinch et al., 2012), the exact physiological and behavioral responses of adult salmon to

* Corresponding author at: School of Fisheries, Aquaculture, and Aquatic Sciences, Auburn University, 203 Swingle Hall, Auburn AL 36849, United States.
E-mail address: bas0041@auburn.edu (B.A. Staton).

http://dx.doi.org/10.1016/j.fishres.2017.05.003
Received 8 February 2017; Received in revised form 3 May 2017; Accepted 5 May 2017
Available online 20 May 2017
0165-7836/ Published by Elsevier B.V.

relatively small-scale environmental gradients within estuaries, which are likely the ultimate determinants of freshwater entry timing, are still poorly understood. Despite this uncertainty, several hypotheses have been put forth that are broadly consistent with the observed timing patterns of several species across a large geographic area (i.e., western and southwestern Alaska). Two primary influences have been suggested: genetic (Quinn et al., 2000; Anderson and Beer, 2009; O'Malley et al., 2010) and environmental (Hodgson et al., 2006; Keefer et al., 2008) mechanisms. Substantial evidence exists to suggest that both genetic and environmental controls are involved in determining migration timing, however it is broadly thought that genetic variation influences sub-stock variation (i.e., different tributary spawning groups within the same major river basin) and environmental variation influences the timing of the aggregate (i.e., basin-wide) run (Keefer et al., 2008; Anderson and Beer, 2009). This is consistent with the notion that genetically distinct components of the aggregate run behave differently as a result of their life history strategies and/or the characteristics of their specific spawning grounds (e.g., sub-stocks that must travel farther in-river to reach spawning grounds enter freshwater earlier [Clark et al., 2015]; sub-stocks that spawn in tributaries influenced by warmer lakes enable later spawning; [Burger et al., 1985]) but that certain environmental conditions act on the aggregate run to either hasten or delay freshwater entry. It has also been suggested that run size may have an influence on migration timing, although empirical support for this claim seems to be lacking. If there were indeed relationships between run timing and run size, these need to be quantified as certain combinations are particularly troublesome for managers (e.g., small/early runs and large/late runs appear the same early in-season; Adkison and Cunningham, 2015).

At the aggregate population scale, which is the focus of this paper, it has been observed that migrations occurring in the spring and summer generally occur earlier in years with warmer spring temperatures (Mundy and Evenson, 2011; Hodgson et al., 2006). Mundy and Evenson (2011) suggested that this pattern may be explained by the stability of the estuarine water column where adult salmon stage in preparation for riverine entry (or alternatively, marine exit). High estuarine water column stability was hypothesized to impede riverine entry through two mechanisms: (1) by presenting an osmotic barrier between freshwater riverine discharge and the saline ocean water which prevents osmotically incompetent individuals from crossing and (2) by preventing freshwater competent individuals from receiving olfactory cues essential to the homeward migration. Thus, Mundy and Evenson (2011) hypothesized that years in which the estuarine water column is stable over a longer period of time would be associated with later migration timing. Although water column stability is a difficult variable to measure over large spatial scales, several variables that are known to influence it are available at large scales via remote sensing (e.g., satellite observations). Such variables are sea ice cover which prevents wind-driven mixing, associated local temperature-related variables like land-based air temperature or sea surface temperature (SST), and broader scale indicators such as the Pacific Decadal Oscillation (PDO), an index of temperature anomalies in the northern Pacific Ocean. Observational studies across the North American range of Chinook salmon have found environmental-run timing correlations that are consistent with this hypothesis (Hodgson et al., 2006; Keefer et al., 2008; Mundy and Evenson, 2011). Even if the water column stability hypothesis is incorrect, observed patterns suggest that environmental variables may be used to forecast run timing with some level of accuracy and certainty.

Several efforts have been made at exploiting these environmental-run timing relationships to develop run timing forecast models for Pacific salmon migrations. Mundy and Evenson (2011) developed a model for Yukon River Chinook salmon (*O. tshawytscha*) that used air temperature, sea surface temperature, and ice cover to predict the day at which the 15th and 50th percentiles of the run passed a test fishery index location. Model predictions fit the observed data well (nearly always within seven days, usually within three days), although out-of-sample predictive ability was not presented (Mundy and Evenson, 2011). Keefer et al. (2008) developed a similar framework for Columbia River spring run Chinook salmon and found run timing relationships with river discharge, river temperature, and ocean condition indices (e.g., PDO). Their best model explained 49% of the variation in median run timing with variation in the environmental variables. Anderson and Beer (2009) continued this work on the Columbia River spring Chinook stock, but added genetic components to their analysis based on the arrival timing of precocious males. Their findings revealed that both environmental variables and changes in abundance of genetically distinct populations, which had their own distinct migration timing, affected overall run timing of the spring Chinook salmon run in the Columbia River. These advancements have shown that relationships between migration timing and environmental variables exist and may have utility for use in forecasting efforts.

The Kuskokwim River, located in western Alaska, is the second largest river system in the state and supports culturally and economically important Chinook salmon fisheries. Chinook salmon return beginning in late May and continue through early August, with the median date of passage occurring between June 14th and July 2nd. Fisheries within the region harvest salmon in-river during freshwater migrations using primarily drift gillnet gear. The Kuskokwim River salmon fishery has a distinct cultural importance: nearly all inhabitants are native Alaskans belonging to the Yup'ik group and take salmon for subsistence purposes (Linderman and Bergstrom, 2009). While commercial salmon fisheries operate within the river, these fishers often also participate in subsistence take and revenues from the sale of commercially-harvested salmon often contribute directly to participation in subsistence activities (Wolfe and Spaeder, 2009). To ensure long-term sustainable harvest, the Chinook salmon fishery is managed with a drainage-wide escapement goal derived from an age-structured state-space spawner-recruit analysis (Hamazaki et al., 2012; Staton et al., 2017). To meet these pre-determined escapement goals, in-season management strategies implement time, gear, and area closures based on limited and imprecise information regarding annual run size. The distant locations of the majority of escapement assessment projects makes direct measurement of escapement performance unavailable until late in the season. Thus, the primary sources of run size assessment information are (1) a pre-season run size forecast range (previous year's run size estimate ± approximately 20%) and (2) an in-river drift gillnet test fishery operated in Bethel, AK which has been implemented using consistent methods since 1984. The interpretation of this test fishery index suffers from the same issue of being confounded by run timing described earlier, making management decisions difficult. Without precise in-season indicators of run size, managers must often choose to either (1) trust a pre-season run size forecast for the majority of the season or (2) take their chances at deciding if the run is early or late when interpreting in-season data. Both options could lead to the wrong interpretation of the actual run size, which could have serious consequences for the management of the fishery (i.e., the unwarranted opening or closing the fishery resulting in severe under- or over-escapement). No published run timing forecast models currently exist for Kuskokwim River Chinook salmon but given the potential utility of independent run timing estimates for interpretation of in-season data, the development and evaluation of such a model is needed. The necessity of more accurate and precise in-season perceptions of run size is particularly evident in years with anticipated low runs, such as in recent years (i.e., since 2010), as this may allow managers to more effectively guard against over-exploitation while still allowing for limited harvest opportunities to support the cultural and subsistence needs of the region.

We present an analysis that develops and evaluates the performance of a run timing forecast model for Kuskokwim River Chinook salmon. The objectives were to (1) quantify historical run timing, (2) develop a run timing forecast model using environmental variables selected based

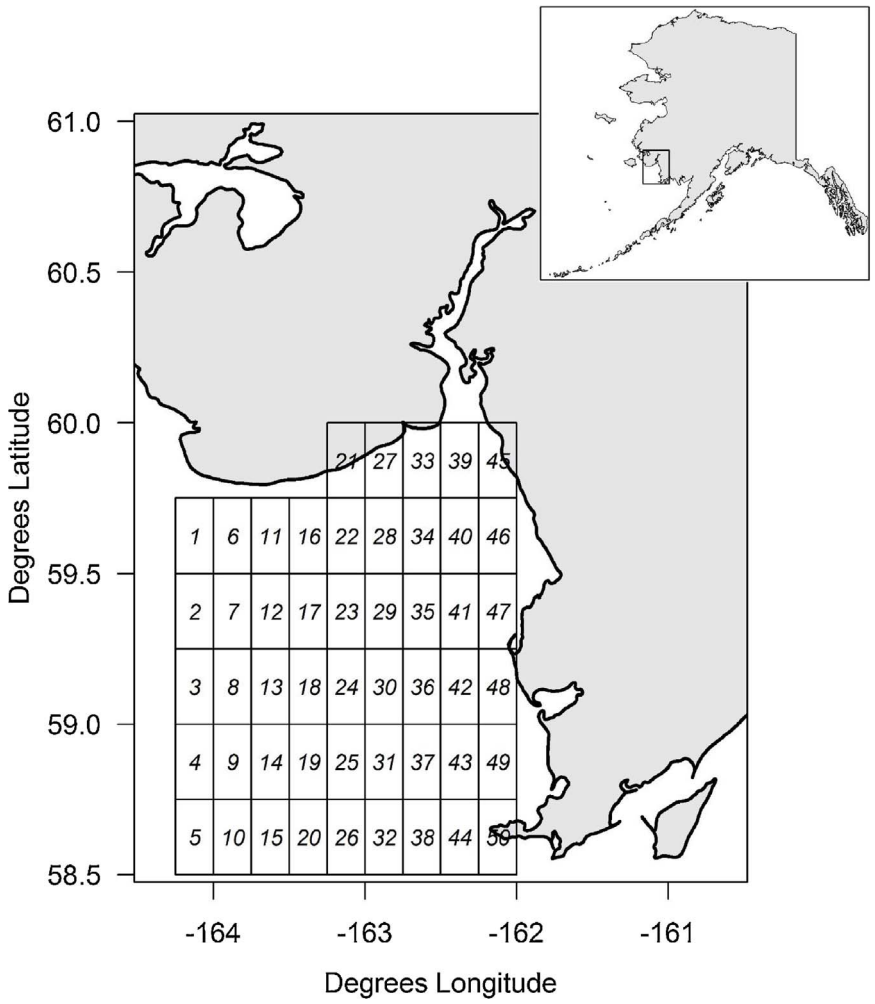


Fig. 1. Map of Kuskokwim Bay where Chinook salmon likely stage for transition to freshwater. Shows grid cells from which daily SST values were used. Daily SIC values came from the same grid cells, though excluding grid cell 45 below due to missing values.

on out-of-sample predictive performance, (3) assess the utility of the forecasting model for improving predictions of end-of-season test fishery indices of run size, and (4) determine if there is a relationship between run size and run timing for the Kuskokwim River Chinook salmon stock.

2. Methods

2.1. Estimates of migration timing

In this analysis, the forecasted quantity that represented migration timing was the day at which 50% of the run passed an index location (hereafter, D_{50}). To inform this quantity for each year in the analysis, we used daily catch-per-unit-effort (CPUE) data from the Bethel Test Fishery (BTF) operated by the Alaska Department of Fish and Game (ADFG), which spans 1984–2016. The raw data were daily CPUE beginning on 1 June and ending 24 August each year. The cumulative sum of these daily CPUE values within a year follows a sigmoidal pattern reflecting the shape of the incoming salmon run which is characterized by relatively few early migrants, a peak where the majority of the fish are running, and relatively few late migrants. To estimate the median day of passage as a continuous variable, a logistic model was fitted to the cumulative proportion of daily CPUE of the form:

$$p_{d,t} = \frac{1}{1+e^{-h_t(d-D_{50,t})}} \quad (1)$$

where $p_{d,t}$ is the predicted cumulative proportion on day of the year d in calendar year t, h_t is the parameter that controls the steepness of the curve (i.e., duration of the run), and $D_{50,t}$ is the day at which 50% of the total annual CPUE was caught in year t. Annual estimates of D_{50} and h were obtained by fitting $p_{d,t}$ to observed daily cumulative proportion by minimizing the sum of squared deviations from the model. Uncertainty in these parameter estimates was not further considered in the analysis as the uncertainty was negligible. Further, by using the BTF daily values to infer the location and shape of year-specific logistic timing curves, we made the assumption that these data provided an accurate representation of daily run strength within a year (influence of weather conditions or harvest on sampling was negligible).

2.2. Environmental variables

Environmental variables to be assessed for forecasting performance were chosen based on (1) previously established association with salmon run timing, (2) availability for the Kuskokwim River during the years for which BTF index observations exist (1984–2016), and (3) availability for use in a pre-season forecast model (i.e., available no later than June 10th in the year for which the forecasted value would be used). Based on these criteria, four environmental variables were chosen for analysis: SST, percent sea ice cover (SIC), PDO, and land-based air temperature taken in Bethel, AK.

2.2.1. PDO data

Data collected for the PDO variable came from one of several indices

produced by the National Oceanic and Atmospheric Administration (NOAA) (Mantua et al., 1997; values available at http://research.jisao.washington.edu/pdo/PDO.latest.txt). The index is produced by taking the first principal component of monthly SST anomalies in the northern Pacific Ocean, after removing any global trends due to any systematic change over time (Mantua et al., 1997). Thus, for each year of the data set, a single monthly value was available for PDO. In the selection of relevant time periods (Section 2.3), we considered monthly PDO values from January to June of each year, as previous studies have found PDO values prior to the initiation of the run have predictive value for Chinook salmon populations (Beer, 2007; Keefer et al., 2008).

2.2.2. Bethel air temperature data

Air temperature data for Bethel, AK were accessed from the Alaska Climate Research Center webpage (http://akclimate.org/acis_data). These data were available as daily means for each day of each year in the 1984–2016 data set. Daily air temperature values from the start of January to mid-June of each year were considered in the selection procedures.

2.2.3. SST and SIC

SST and SIC data were accessed from the NOAA Optimum Interpolation SST V2 High Resolution Dataset (Reynolds et al., 2007; available at: http://www.esrl.noaa.gov/psd/data/gridded/data.noaa.oisst.v2.highres.html). These data were available as daily means for any 0.25° by 0.25° latitude by longitude grid cell on the globe. To limit the search, only grid cells within Kuskokwim Bay were selected for analysis (Fig. 1) as that is the area that Chinook salmon bound for the Kuskokwim River likely aggregate prior to riverine entry. The area with grid cells ranged from 58.5° N to 60° N by 164.25° W to 162° W, which resulted in a total of 54 0.25° latitude by 0.25° longitude grid cells. For SST, four grid cells fell partially over land (resulting in 50 grid cells with daily data) and for SIC, five grid cells were partially over land (49 grid cells with daily data). "Empty" grid cells were excluded and the remaining grid cells were used for prediction. Previous analyses have used a simple average over a wide spatial area (e.g., Mundy and Evenson, 2011) to create a single value for SST or SIC each year. However, this is somewhat arbitrary and does not account for the possibility of certain areas having stronger timing signals than others or that these areas with stronger signals may change over time. Thus, the gridded spatial structure of these variables was retained and the treatment of this structure in the forecast analysis is discussed below in Section 2.4. Because of the time-intensive nature of the selection procedure described in Section 2.3, considered time periods for gridded variables were shortened based on preliminary correlation analyses showing excluded times had little or no correlation with D_{50}. Daily means for SST from mid-April to mid-June and SIC from mid-February to mid-May in each year were considered in the predictive variable selection techniques.

2.3. Selection of predictive time periods

Climatic variables are frequently associated with biological quantities for the purpose of prediction, however, oftentimes the average over an arbitrary time period, such as daily values in the month of February, is used based on *a priori* assumptions of the behavior of important factors (van de Pol et al., 2016). While this approach is simple to implement and explain, it is possible that a better time window (i.e., reliably more accurate) exists but is not considered. Furthermore, the importance of various time windows may change over time and the arbitrary selection of a single window does not allow for such changes to be detected. To avoid these issues, a rigorous temporal selection process, known as the sliding window algorithm (van de Pol et al., 2016), was implemented to determine the best predictive time period for each variable considered in the forecast model. To find the most reliable temporal window for prediction, this method evaluates all possible windows (subject to certain restrictions) over which to average for use in the forecast model. As input constraints, the sliding window algorithm used in this analysis required: (1) the start date of the first window to be evaluated, (2) the end date of the last window to be evaluated, and (3) the minimum size of a candidate window. It then proceeded to evaluate all possible combinations of window averages (using only consecutive days). For example, if the start date of the first possible window was set to 1 May, the minimum window size was set to five days, and the end date of the last possible window was set to 30 May, the window algorithm would start by averaging the daily values between 1 May and 5 May for each year separately and would evaluate the predictive performance of that variable using forecast cross-validation (described below). Next, the algorithm would begin to increase the window size by one day on each iteration, averaging daily values and evaluating the predictive performance for 1 May to 6 May, then 1 May–7 May, and so on, until it reached the largest window starting with 1 May (which would be 1 May–30 May). Following completion of all windows starting with 1 May, the algorithm would proceed to evaluate all possible windows starting with 2 May, then 3 May, and so on, until it reached windows starting with 26 May. At this point, there would only be one window to be evaluated, since the minimum number of days to include in a candidate window was set at five days. In the analysis, five days was used as the minimum window size for all variables except PDO (which had monthly values), in which case the minimum window size was set such that a single month could constitute a window. Example R code for how the sliding window algorithm was implemented is provided in Appendix A in Supplementary material.

All windows were evaluated using a time series forecast cross-validation procedure, which is an out-of-sample technique for data that are collected through time (Arlot and Celisse, 2010). The procedure operated by producing a forecasted value of D_{50} for year $t+1$ trained based on all data available from years 1, …, t. It then continued for all $t = m, \ldots, n\text{-}1$, where m is the minimum number of years necessary to fit the model (set at $m = 10$ in all cases) and n is the number of years of available data. Then, absolute forecast error in D_{50} was calculated based on all forecasted years as $|y_{t+1} - \hat{y}_{t+1}|$, and yearly forecast errors were averaged to obtain mean absolute error ($\overline{AE}$) which was used as the measure of model performance in window selection. The window with the lowest $\overline{AE}$ was selected as the optimal window to average over for prediction. The forecasting cross-validation procedure was used as opposed to other out-of-sample validation procedures, such as k-fold or leave-one-out methods, because the data were collected through time and it does not make sense for the model to need to predict (for example) year 2010 from years 1984–2009 and 2011–2016, but rather it would always need to predict year $t+1$ from all previously-collected data. Example R code for how the forecast cross-validation was conducted is provided in Appendix A in Supplementary material.

Although a recent R package has been released to conduct sliding climate window analyses ("climwin"; Bailey and van de Pol, 2015), the package can currently only assess candidate windows using Akiake's Information Criterion (AIC) or k-fold cross-validation. Because the more specialized forecasting cross-validation technique was required for this analysis, custom R code was written to select the optimal window for each variable (Appendix A in Supplementary material).

When forecasting $D_{50,t+1}$ from training data from 1, …, t, a single optimal climate window was selected for each variable and that window was used to estimate coefficients based on training data and obtain the environmental variable value for prediction in year $t+1$ to forecast $D_{50,t+1}$. When a new year of data was added to the training data (such as in the retrospective forecast analysis; Section 2.7), the optimal window for each variable was re-assessed using the algorithm again. For PDO and Bethel air temperature, which had no spatial structure, the sliding window algorithm was used to select the range of monthly (PDO) or daily (Bethel air temperature) values to include in the predictive climate window for each year in the analysis. For SST and SIC which contained a series of 50 and 49 grid cells, respectively, each

with unique daily values, the sliding window algorithm was used on each grid cell separately. The result was 50 unique grid cell-specific windows for SST and 49 windows for SIC for each year of the analysis. The treatment of this spatial structure in the forecast analysis is discussed below in Section 2.4.

2.4. Evaluated forecast models

Linear regression was used to assess the forecast performance of each of the variables described above, both in isolation of and in combination with other variables. All possible subsets were evaluated (excluding interactive effects) for predictive ability through time, resulting in a total of 16 models ranging from the null (i.e., intercept only) model to the full (i.e., global) model (all four variables as additive predictors).

For the spatially explicit variables (i.e., SST and SIC), a more complex treatment was required to prevent all grid cell values from being used as predictors in a single model. To handle the spatial structure, grid cell-specific regression models were fitted, then model-averaging based on AIC was used to obtain a single forecast D_{50} for each year (Burnham and Anderson, 2002). Under this approach, each grid cell g received an AIC_c score:

$$AIC_{c,g}=n\log(\hat{\sigma}_g^2)+2K+\frac{2K(K+1)}{n\text{-}K\text{-}1}, \quad (2)$$

where n is the number of data points used in each model, $\hat{\sigma}_g$ is the estimate of the residual standard deviation under grid g, and K is the number of model parameters. The corrected version of AIC (AIC_c) is recommended in cases where the ratio of n to K is small (Burnham and Anderson, 2002). Then, each grid cell received a ΔAIC_c score, representing its relative performance in comparison to the best grid cell:

$$\Delta_g=AIC_{c,g}\text{-}AIC_{c,\min}, \quad (3)$$

where $AIC_{c,min}$ is the minimum AIC_c across all grid cells. Model weights were then calculated as:

$$w_g=\frac{e^{-0.5\Delta_g}}{\sum_j^G e^{-0.5\Delta_j}}, \quad (4)$$

Grid cell-averaged predictions were then obtained as:

$$\hat{y}_{t+1}=\sum_g^G w_g\hat{y}_{g,t+1}, \quad (5)$$

where $\hat{y}_{g,t+1}$ is the forecasted D_{50} for grid cell g, and G is the number of grid cells.

2.5. Forecast uncertainty

In addition to forecast accuracy, forecast precision is also of great importance. For models that did not require AIC_c model-averaging across grid cells, the following equation was used to produce a forecast standard error (SE):

$$SE=\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{(x\text{-}\bar{x})^2}{\sum_i^n (x_i\text{-}\bar{x})^2}}, \quad (6)$$

where n is the number of years the model was fit to, x is the value of the predictor variable used for forecasting, and $\bar{x}$ is the mean of all predictor values excluding the new value used for forecasting. For models that used AIC_c model-averaging (i.e., those including SST and SIC), the following equation was used to produce prediction SE:

$$SE=\sum_g^G w_g\sqrt{SE_g^2+(\hat{y}_{g,t+1}\text{-}\hat{y}_{t+1})^2}, \quad (7)$$

where SE_g is the prediction SE from grid cell g, calculated using Eq. (6). This estimator of unconditional sampling standard error accounts for uncertainty within each model and the uncertainty due to model selection (Burnham and Anderson, 2002). Prediction intervals were calculated using the point estimate of prediction, the prediction SE, and appropriate quantiles from the corresponding t distribution.

2.6. Forecast model selection

Given 16 forecast models, it is impossible to know which will perform the best at forecasting for the current year. Thus, three methods to obtain a forecast for D_{50} were evaluated: (1) the null (i.e., intercept only) model, (2) the single model with the lowest forecast cross-validation score as of the last year, and (3) model-averaging across the ensemble of 16 forecast models based on AIC_c scores. According to Burnham and Anderson (2002), model-averaging should perform better than a single "best model" at prediction when there is a high degree of uncertainty about which model is best. This procedure was performed using Eqs. (2)–(5), by substituting the prediction, prediction SE, and K for forecast model i, in place of grid g. Prediction intervals based on model-averaged predictions and prediction SE present somewhat of a problem when the different models contributing to the average contain differing degrees of freedom as it is unclear how many standard errors the prediction limits should lie from the mean prediction. Thus, the estimator suggested by Burnham and Anderson (2002) of the "adjusted SE" (ASE) was used:

$$ASE=\sum_i^{16} w_i\sqrt{\left(,\frac{t_{df_i,1\text{-}\alpha/2}}{z_{1\text{-}\alpha/2}}\right)^2 SE_i^2+(y_{i,t+1}\text{-}\hat{y}_{t+1})^2}, \quad (8)$$

where $t_{dfi,1\text{-}\alpha/2}$ is the $1–\alpha/2$ quantile of the t distribution with degrees of freedom equal to that of model i and $z_{1\text{-}\alpha/2}$ is the corresponding quantile of the z (i.e., standard normal) distribution. The level $\alpha = 0.05$ was used in all cases.

2.7. Retrospective forecast analysis

The analysis was conducted in a retrospective forecast framework starting in 1994. All data after 1994 were ignored, optimal windows were selected for each of the four variables (and all grids for SST and SIC), all 16 models were fitted, a D_{50} forecast was made for 1995 using the three approaches described in Section 2.6, and each was evaluated for predictive accuracy. This process was repeated annually until the present (i.e., out-of-sample predictions made for 1995–2016), which allowed for the calculation of $\overline{AE}$ through time as if the forecast model would have been available beginning in spring 1995. In addition to $\overline{AE}$, median absolute error ($\widetilde{AE}$) was calculated to validate prediction accuracy of estimates by ignoring the effect of outlying poor predictions.

2.8. Value of forecast to run size assessments

It is important to remember that the purpose of producing a run timing forecast is to aid in the interpretation of in-season indices of run size such as test fisheries. To evaluate the utility of having access to the run timing forecast model, the accuracy and precision of end-of-season cumulative CPUE (*EOS*) predictions were compared using D_{50} forecasts from the model-averaged and the null forecast models. In-season perceptions of run size were produced by predicting *EOS*:

$$\hat{EOS}_{d,t,i}=\frac{CCPUE_{d,t}}{\hat{p}_{d,t,i}}, \quad (9)$$

where $CCPUE_{d,t}$ is the cumulative CPUE caught at the BTF through day d in forecasting year t, $\hat{p}_{d,t,i}$ is the predicted cumulative proportion of the run that had passed the BTF location on day d in year t from model i (i.e., model-averaged or null forecast model) obtained by inserting the forecasted value of D_{50} into the logistic function Eq. (1). Uncertainty in the run timing forecast model was propagated to $EOS_{d,t,i}$ using the delta

method (Bolker, 2008). As the parameter h is also unknown in-season, the mean of all historical h_t was used as the point estimate, and the variance and correlation with $D_{50,t}$ was included in the covariance matrix supplied to the delta method. Accuracy was assessed using proportional bias $[(\widehat{EOS}_{d,t,i} - EOS_t)/EOS_t]$ and precision was assessed using the SE of *EOS* on day *d,* both compared between the null and the model-averaged forecast model. Accuracy and precision of *EOS* predictions were assessed on 15 June, 30 June, 15 July, and 30 July each year. Using the null model to obtain $\hat{p}_d$ is essentially equivalent to what is currently done for predicting *EOS* index of run size in the Kuskokwim River (i.e., in the absence of an environmental variable forecast model for D_{50}).

2.9. Investigation of a run timing versus run size relationship

To test the hypothesis that run timing is related to run size (e.g., small runs are typically early, or vice versa), two models were investigated for their predictive performance using the forecast cross-validation criterion: the null model and a model that included run size as a predictive covariate in place of the environmental variables. Run size was obtained from a maximum likelihood run reconstruction model that compiles all assessment information (i.e., 20 escapement count indices, harvest estimates, drainage-wide mark-recapture estimates, etc.) to estimate the run size that makes the collected data most likely to have been observed (Bue et al., 2012). The forecast absolute errors in each year were then compared using a two-tailed paired *t*-test using $\alpha = 0.05$.

3. Results

3.1. Estimates of run timing

The logistic curve fit the daily cumulative CPUE proportions well in all years of the BTF data set (Table 1), as indicated by an average residual standard error estimate of 0.022, with a maximum estimate of 0.038 in 1992. The majority (95%) of all residuals from all years fell between −0.056 and 0.044. Parameter estimates were quite precise, with D_{50} having a smaller average coefficient of variation (CV) than h, (0.07% and 2.09%, respectively). Given this small degree of parameter uncertainty, it was ignored throughout the rest of the analysis.

3.2. Variable-specific relationships

Looking at each of the environmental variables in isolation of all others, it is clear that there is a distinct relationship between temperature-related environmental variables and Kuskokwim River Chinook salmon migration timing (Fig. 2). For illustration purposes, the figures for the two gridded variables (SST and SIC) were produced by taking an average across all grid cells weighted by the AIC weight for each grid cell. Air temperature, PDO, and SST all had negative relationships with D_{50}, whereas SIC had a positive relationship (Table 2).

3.3. Selected climate windows

It was difficult to generalize on the climate windows selected for each variable based on forecast cross-validation performance, because the selected windows changed with each new year of data and SST and SIC had windows for each grid cell; however, some noteworthy patterns arose. First, the best window for PDO was consistently the value for the month of May for each year the forecasts were produced (not shown). Second, selected windows for air temperature fluctuated from year to year to some extent, but long windows were consistently selected for all years after 1999, and generally spanned early February to late May (Fig. 3a). Third, selected windows through time were substantially more variable for most grid cells for SST and SIC, although many grid cells remained relatively constant or became more "focused" as more years of data were added (Figs. 3b1-4, c1-4). In general, chosen windows for SST began in early to mid-May and ended in late May or early June (Fig. 3b1-4) whereas windows starting in early April and ending in mid to late April were predominately chosen for SIC (Fig. 3c1-4). The selected climate windows in southern-most grid cells appeared more stable for SST (Fig. 3 panels b3 and b4), whereas climate windows in northern grid cells appeared more stable for SIC (Fig. 3 panels c1 and c2; stable in the sense that the optimal windows changed less as new years were added to the training data).

Table 1
Parameter estimates (mean with standard error in parentheses) from logistic curves from Eq. (1) fit to all years. D_{50} is expressed as day of the year; for reference, day 174 is June 22nd in a leap year and June 21st in a non-leap year.

Year	D_{50}	h
1984	174.80 (0.20)	0.15 (0.004)
1985	183.54 (0.11)	0.25 (0.006)
1986	173.19 (0.15)	0.21 (0.006)
1987	172.66 (0.12)	0.17 (0.003)
1988	172.14 (0.18)	0.16 (0.004)
1989	173.70 (0.10)	0.20 (0.004)
1990	175.86 (0.12)	0.17 (0.003)
1991	175.80 (0.10)	0.19 (0.003)
1992	173.29 (0.24)	0.15 (0.005)
1993	168.31 (0.07)	0.22 (0.003)
1994	169.80 (0.12)	0.19 (0.004)
1995	172.44 (0.07)	0.19 (0.002)
1996	166.27 (0.10)	0.21 (0.004)
1997	170.56 (0.13)	0.26 (0.008)
1998	175.07 (0.08)	0.20 (0.003)
1999	180.90 (0.19)	0.13 (0.003)
2000	171.53 (0.18)	0.17 (0.005)
2001	174.13 (0.13)	0.19 (0.004)
2002	169.99 (0.20)	0.17 (0.006)
2003	168.91 (0.16)	0.17 (0.004)
2004	173.98 (0.15)	0.17 (0.004)
2005	173.64 (0.15)	0.16 (0.004)
2006	175.43 (0.11)	0.19 (0.004)
2007	177.82 (0.08)	0.19 (0.003)
2008	176.05 (0.09)	0.19 (0.003)
2009	173.13 (0.07)	0.23 (0.003)
2010	173.14 (0.19)	0.19 (0.006)
2011	173.87 (0.09)	0.16 (0.002)
2012	178.50 (0.11)	0.22 (0.005)
2013	173.53 (0.11)	0.22 (0.005)
2014	166.56 (0.11)	0.17 (0.003)
2015	174.57 (0.18)	0.12 (0.002)
2016	174.13 (0.09)	0.12 (0.001)

3.4. Forecast performance

Of the three investigated forecasting methods (null model, model with lowest forecast cross-validation error up to the forecasting year, and AIC_c model-averaging), the null model had the lowest $\overline{AE}$ from 1995 to 2016 (2.64 days; Fig. 4). AIC_c model-averaging performed better than using the single model with the lowest cross-validation score (3.04 versus 3.34, respectively; Fig. 4). However, these patterns were not consistent across the entire time series. For the period of 1996–2008, the model-averaged forecast had a lower $\overline{AE}$ than the null model, and for the period of 2009–2015, the model-averaged forecast had approximately the same or lower $\overline{AE}$ scores (Fig. 5). It was due in a large part to 2016 that the model-averaged forecast had a slightly higher $\overline{AE}$ than the null model. Each model in the ensemble of 16 models (except the null) predicted an extremely early run in 2016 when in fact the observed run timing in 2016 was close to the historical (1984–2015) average (Fig. 4). A similar case happened in 2015 (Fig. 4). Expressing prediction error in terms of median absolute error ($\widetilde{AE}$) resulted in lower average errors (null = 2.08, single best = 2.56, and model-averaged = 2.35), indicating that extreme prediction errors (i.e., outliers) influenced the value of $\overline{AE}$. However, the relative

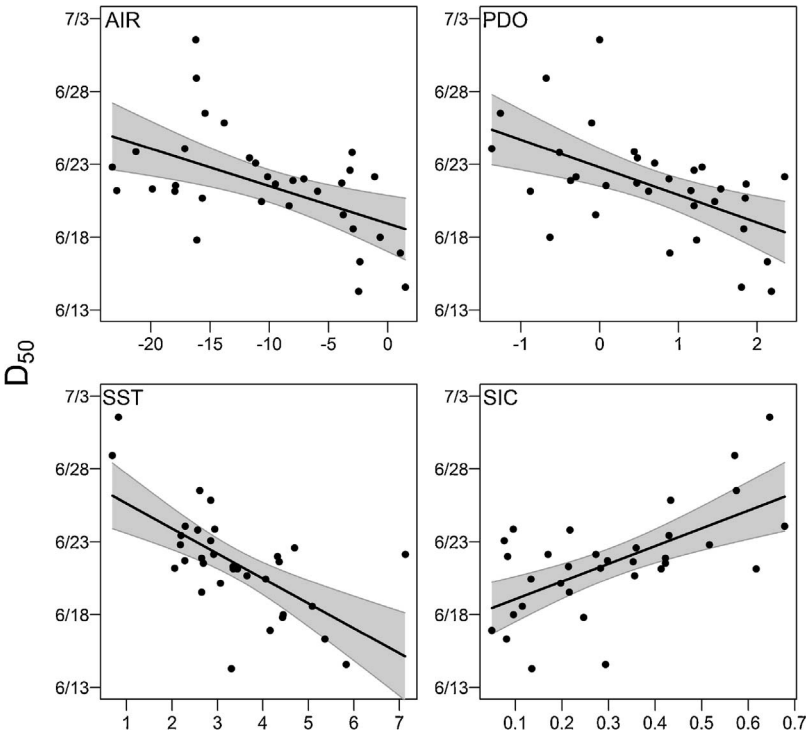


Fig. 2. Relationships between the four single environmental variables and run timing (D_{50}) using data from optimal climate windows when 2016 was added to the training data. For illustration purposes only, gridded variables (SST and SIC) were combined by weighted averaging where the weight of each grid cell was assigned the AIC_c weight of that grid cell when grid cell-specific models were fitted. Grey bands are 95% confidence intervals on the least squares line.

Table 2
Regression coefficient estimates and selected statistics from the single-variable regression models shown in Fig. 2. Intercept and slope estimates are presented as mean (SE) and the critical $F_{1,31}(\alpha = 0.05)$ value in each test was 4.16. Significance codes are * $p \leq 0.001$ and ** $p \leq 0.0001$. Interpretation of the intercept and slope need special attention as the x-variables have specific climate windows and some variables are measured on different scales. Thus, strength of effects should be interpreted using t_{slope} − values farther from zero indicate a stronger positive or negative effect.

Variable	Intercept	Slope	t_{slope}	R^2	F
AIR	170.93 (0.94)	−0.26 (0.08)	−3.41	0.25	11.62*
PDO	174.79 (0.63)	−1.89 (0.52)	−3.63	0.28	13.17*
SST	179.33 (1.34)	−1.71 (0.37)	−4.65	0.39	21.60**
SIC	169.85 (0.99)	12.15 (2.78)	4.36	0.36	19.05**

differences in prediction error between models were approximately the same as for $\overline{AE}$, indicating outliers affected the prediction error scores for each model similarly. Additionally, by comparing the width of the prediction intervals in Fig. 4 across forecasting approaches, it was clear that model-averaging substantially reduced prediction uncertainty (SE) in relation to the null and single best model approaches.

To compare performance in average versus extreme years among forecasting approaches, $\overline{AE}$ was further calculated in a more specific way: based on how similar or dissimilar the included years were to the mean observed run timing across all years. As would be expected, the null model performed well when only years with D_{50} within ± 1 days of the average were in included in the calculation of $\overline{AE}$ (Fig. 6a), but its accuracy became increasingly worse as years with more extreme realized D_{50} values were included in the calculation (increasing x-axis values in Fig. 6a). The two environmental variable forecast approaches (model-averaging or the single "best" model in each year) performed relatively equally well across this continuum and neither $\overline{AE}$ score was sensitive to the overall similarity or dissimilarity the included years had with average run timing (Fig. 6a). The lower panel shows the relative frequency with which these various X scenarios occurred, indicating how much information each scenario contributed to the overall $\overline{AE}$. On the other hand, the null model only performed as well as the model-averaged forecast when years with D_{50} ± 0.5 days outside of the mean were considered (Fig. 6b). As only more extreme years were considered in $\overline{AE}$ (increasing x-axis values in Fig. 6b), the null model rapidly performed worse and the model-averaged forecast remained relatively insensitive to the degree of extremity (Fig. 6b).

3.5. Value to in-season run size assessments

When the model-averaged and null model forecasts for $D_{50,t}$ were retrospectively used to aid in in-season run assessment based on daily cumulative BTF CPUE, it was evident that the range of possible $\widehat{EOS}_{d,t}$ was substantially smaller when the model-averaged forecast was used as opposed to the null forecast. This is evident by the average daily standard errors of $\hat{EOS}_{d,t}$ on 15 June, 30 June, 15 July, and 30 July: 256.97, 62.23, 6.69, 0.59 for the model-averaged forecast model and 337.19, 72.63, 6.90, and 0.57 for the null forecast, respectively. The reduction in the first two evaluated days is of key importance. In terms of accuracy, however, the forecast model did not perform better at informing *EOS* prediction than the null model. On the same days, the average proportional bias [(estimate − observed)/observed] using the model-averaged $D_{50,t}$ forecast was 0.152, 0.006, −0.024, −0.008 as opposed to 0.145, −0.035, −0.027, and −0.008 under the null model forecast. A visual example of two years is provided in Fig. 7. The upper panels show $\hat{EOS}_d$ when the model-averaged and the null model forecast models were used to inform the location of the logistic cumulative timing curve in 2013 and 2014. The horizontal line shows the observed *EOS*. 2013 is an example of when the null model would have been preferable to use (in terms of accuracy) and 2014 shows a case when the model-averaged forecast would have performed better.

3.6. Run timing versus run size relationship

There appeared to be no evidence to lend support for the hypothesis that run timing and run size are related for the Kuskokwim River

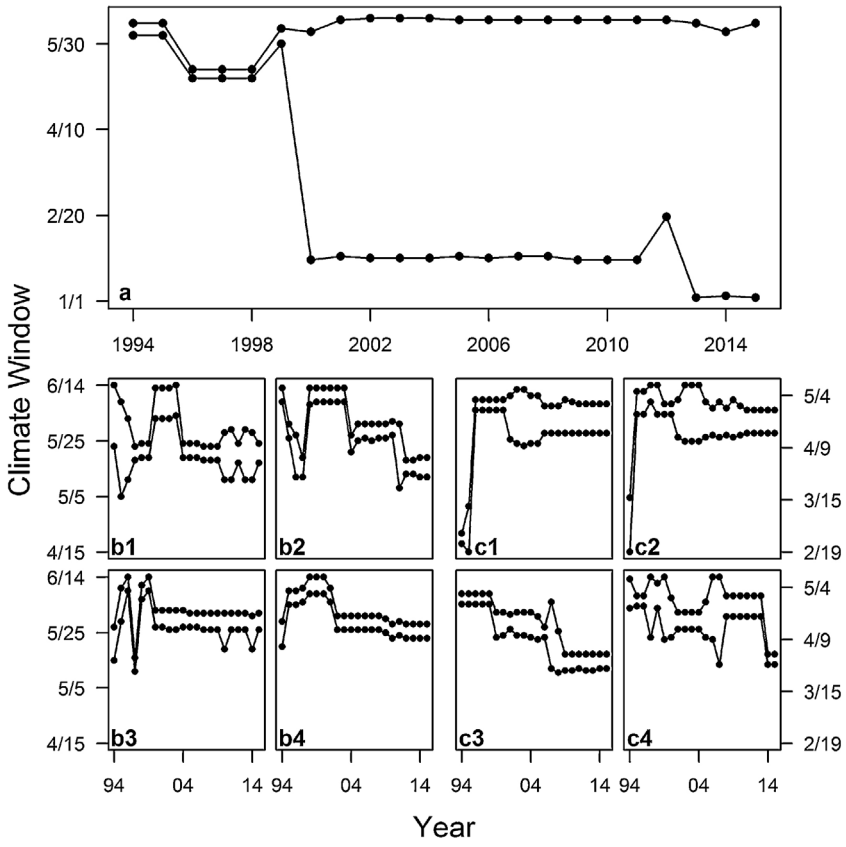


Fig. 3. Changes in selected climate windows as training data were added in the retrospective forecasting analysis. Bottom and top lines show the first and last day of the selected climate window, respectively, as more years were added. The year axis corresponds to the selected window after including environmental and run timing data from that year in the training data. e.g., the windows shown for 2015 were used to produce the forecast for 2016. Panel (a) is Bethel air temperature, panels b1–b4 are SST windows for four sample grid cells and panels c1–c4 are SIC windows for the same four sample grid cells. Sample grid cells from Fig. 1 shown for SST and SIC are as follows: grid cell 8 (b1, c1), grid cell 44 (b2, c2), grid cell 12 (b3, c3), and grid cell 48 (b4, c4). Selected windows for PDO are not shown because the single month of May was selected in all years.

Chinook salmon stock. On average, run timing occurred 0.01 (95% CL; −0.03–0.004) days earlier for each 1000 fish increase in run size, which was not significantly different than no effect of run size on run timing ($p = 0.153$, $R^2 = 0.07$). Additionally, based on forecast cross-validation, the model that included run size did not perform better at prediction than the null model. On average, the null model resulted in an estimated absolute forecast error of 0.2 (95% CL; −1.07–0.68) days less than the run size model ($p = 0.64$).

4. Discussion

The environmental relationships with run timing we detected for the Kuskokwim River Chinook salmon stock are consistent with patterns found elsewhere in the region (e.g., Mundy and Evenson, 2011; Hodgson et al., 2006). Specifically, we found that warmer years were typically associated with earlier-than-average runs as were years with less than average SIC. These findings are consistent with the water column stability hypothesis suggested by Mundy and Evenson (2011). The amount of unexplained variation in the Kuskokwim model appears to be comparable between the Yukon River Chinook salmon stock as well (Mundy and Evenson, 2011). Using the relationships shown in Fig. 3, the correlation with D_{50} was −0.52, −0.64, and 0.62 for air temperature, SST, and SIC, respectively. For the Yukon River Chinook salmon stock, Mundy and Evenson (2011) found correlations of −0.59, −0.72, and 0.66 for the same variables but measured at different spatial and temporal scales and with approximately 10 more years of data included (Table 2 in Mundy and Evenson, 2011). These similar correlations indicate the signals given by environmental variables are of relatively equal strength between these two systems. Given these similarities, future research applying our framework to Yukon River Chinook salmon (and other stocks) may enhance the applicability of our approach, if it could be shown that doing so would improve the accuracy of run timing predictions over other methods.

Given the overall strength of the environmental relationships, it is somewhat surprising that the null model forecast performed better on average than did the model-averaged forecast. This could, potentially, be due to the fact that a variety of biological (size (Bromaghin, 2005) and morphology (Hamon et al., 2000)) and abiotic factors (temperature (Salinger and Anderson, 2006), river discharge (Keefer et al., 2004), and migration distance (Eiler et al., 2015)) may affect migration rate (and subsequently, encounter probability) and catchability, introducing additional variability in our run timing estimates. Future research that accounts for these effects on encounter probability or catchability could offer improved predictions of run timing. Regardless of the underlying drivers, the overall prevalence of years with average run timing likely led to the enhanced performance of the null model. It is not surprising, however, that the model-averaged forecast performed better than the supposedly single “best” model. This finding is consistent with the literature on model-averaging predictions (Burnham and Anderson, 2002). Although the null model performed better in the long-term average (i.e., lower $\overline{AE}$ as of 2016), there are reasons a manager may still justifiably prefer the model-averaged forecast. First, the difference in $\overline{AE}$ between the model-averaged forecast and the null model was 0.4 days, which is small relative to the amount of annual variation in run timing (a 17 day range for D_{50} over 33 years). Second, the model-averaged forecast performed equally well in terms of forecast accuracy regardless of the type of run timing it was used to forecast (i.e., prediction error equal in extreme early/late and average years; Fig. 6b). In contrast, the null model only performed comparably well in years with run timing within ± 3 days from average and error increased

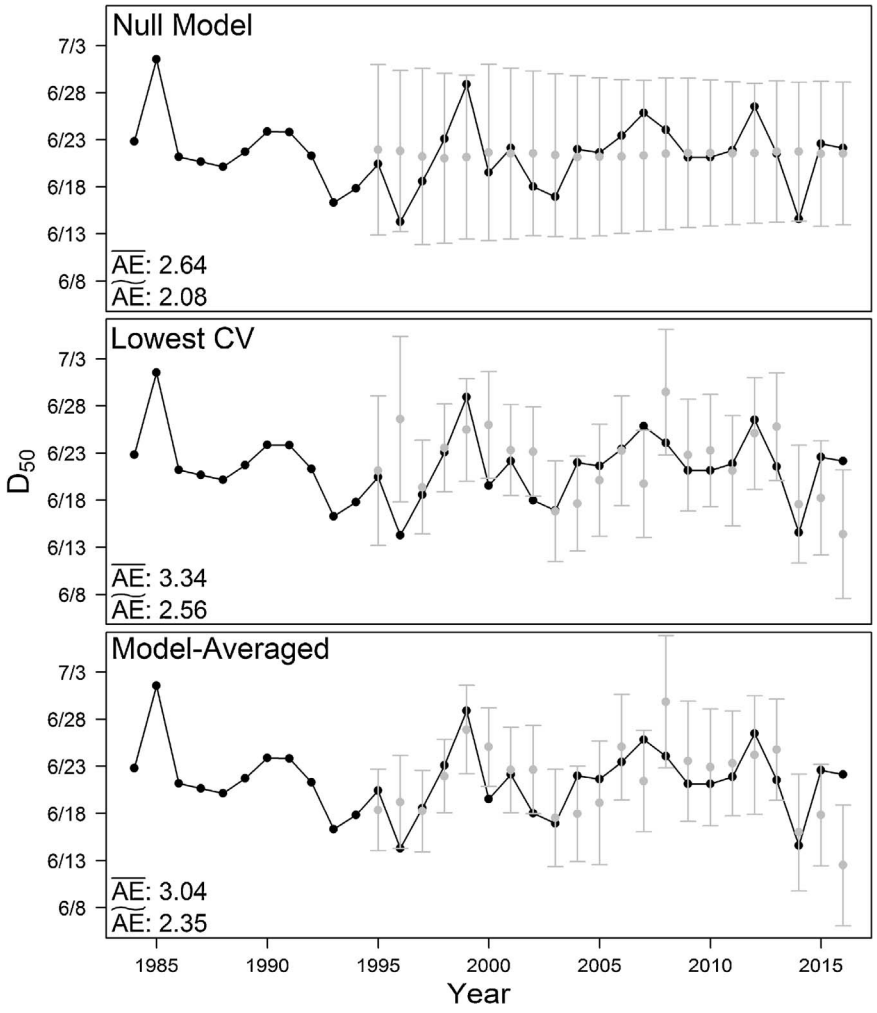


Fig. 4. Produced forecasts under the three approaches. Black points/lines are the time series of D_{50} detected by the BTF. Grey points are out-of-sample forecasts with 95% prediction intervals shown as error bars. $\overline{AE}$ and $\widetilde{AE}$ are the mean and median absolute forecast errors from 1995-2016, respectively.

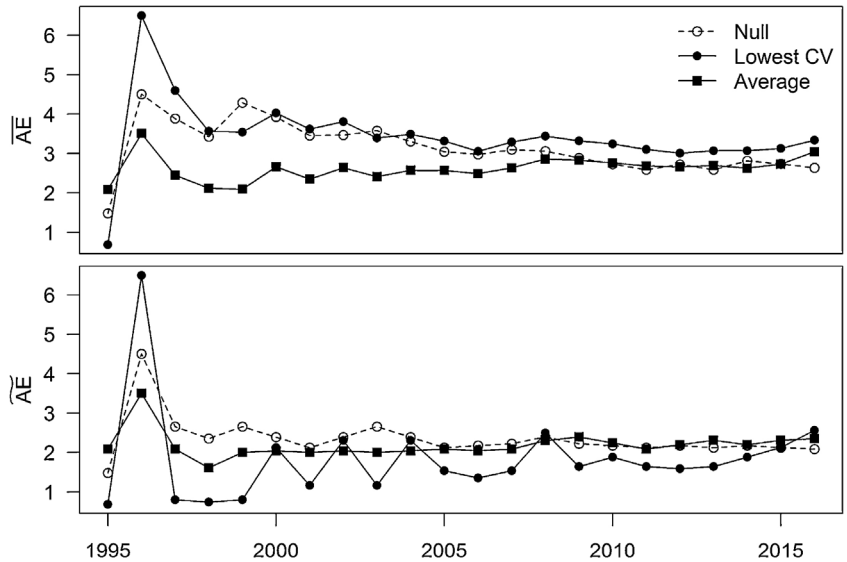


Fig. 5. Evolution of mean ($\overline{AE}$) and median ($\widetilde{AE}$) absolute forecast error under the three investigated forecasting approaches. Each point is the average of absolute errors of all years before and including the corresponding year on the x-axis, starting in 1995.

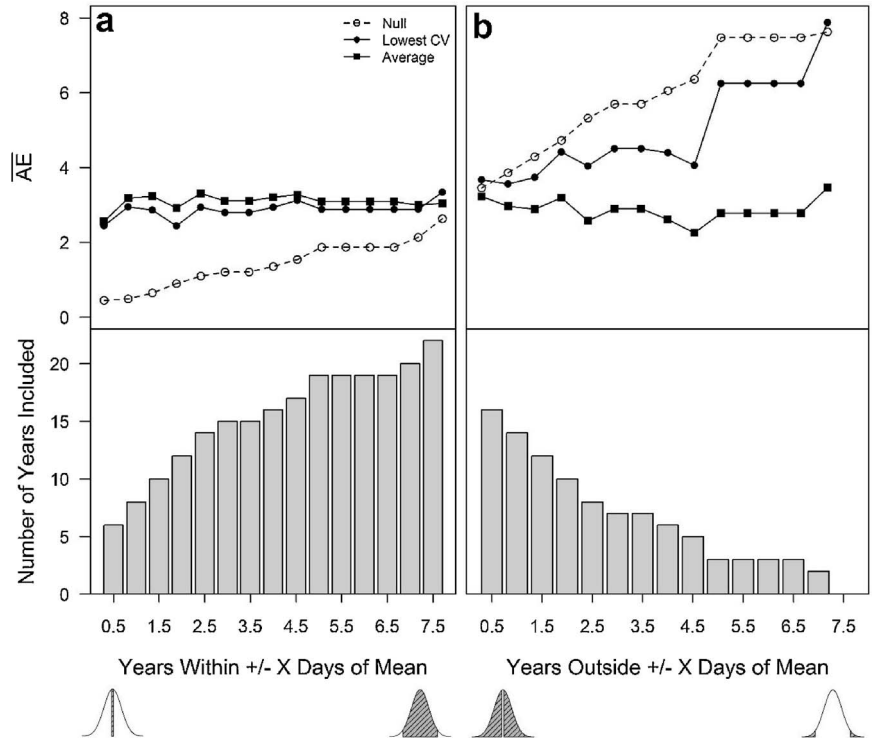


Fig. 6. Mean absolute error ($\overline{AE}$) under three forecast approaches calculated by either (a) including years with a D_{50} value within ± X days of the all-year average or (b) including years with a D_{50} value outside ± X days of average, where X is the number of days indicated on the x-axis. Bottom panels show the number of observed years in which the appropriate ± X days criterion was met. Shaded regions in the hypothetical distributions show the types of D_{50} values that were included in the calculation of $\overline{AE}$. One point that may enrich inference from this figure (and is shown in the shaded normal distributions) is that panel (a) becomes more inclusive from left to right by adding years that are more dissimilar to the average in the calculation of $\overline{AE}$ whereas panel (b) becomes more exclusive from left to right by removing years that are similar to the average.

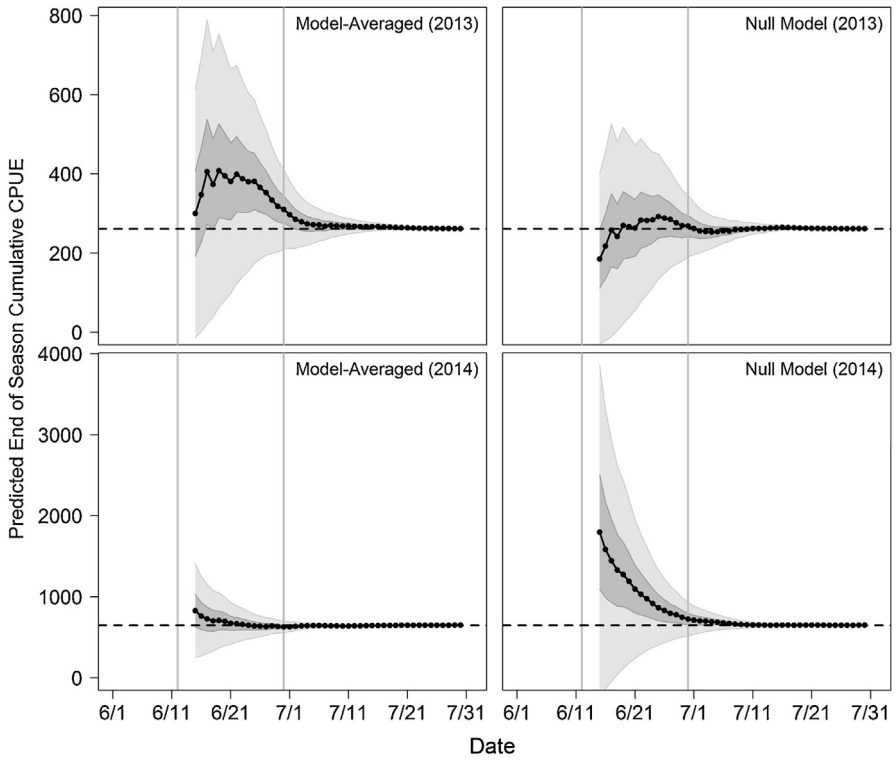


Fig. 7. In-season predictions of end of season cumulative BTF CPUE under the model-averaged forecast using environmental variables and the forecast under the null model in 2013 and 2014. Intended to illustrate cases in which a manager would benefit from having access to the model-averaged run timing forecast model using environmental variables (2014) and when the null model would have performed better (2013). Horizontal lines are the true end of season cumulative BTF CPUE, dark grey regions are 50% confidence intervals, and light grey regions are 95% confidence intervals. Grey vertical lines indicate the period when key harvest decisions are made.

precipitously in more extreme years. Third, the 95% prediction intervals from the null model seemed too wide as 100% of the observations fell within the intervals, whereas 92% of the observations fell within the prediction intervals from the model-averaged forecast (which is closer to the ideal coverage, i.e., 95%). Prediction uncertainty was lower under the model-averaged forecast than the null model, which would ultimately lead to fewer run timing scenarios being considered to explain the observed in-season data (e.g., the earliest or latest scenarios could be excluded earlier in the season) leading to more certain interpretation of in-season indices of run size.

As for the value of having access to a run timing forecast to in-season run size assessments, our findings showed that on average it is preferable to use the environmental forecast over the null forecast. It should be emphasized that we did not evaluate the ability to assess actual run size, only an index of run size (*EOS*). Though because *EOS* can be expressed as a function of actual run size (and vice versa), more precise predictions of *EOS* should presumably result in more precise predictions of actual run size. In terms of prediction error, both approaches led to positive average biases on 15 June (14–15%, model-averaged 1% higher than null model). By 30 June, the bias was reduced to less than 5% and only decreased beyond that for both models. These biases are likely the result of asymmetry in the timing curve detected by the BTF. However, the key difference between approaches was the reduced uncertainty in *EOS* predictions when using the model-averaged forecast due to the exclusion of extreme early or late runs which lead to extreme low and high *EOS* predictions early in the season. The null model was forced to always consider these scenarios, resulting in greater uncertainty in *EOS* predictions, particularly between 15 June and 30 June when key management decisions are made. Due to the large amount of uncertainty under the null model (which is essentially the currently-used method), *EOS* predictions go largely ignored for much of the season and the pre-season run size forecast is trusted instead. If the environmental variable forecast model were to be used, it likely would provide managers with more information when making decisions.

The sliding window algorithm was an objective, adaptive, and data-driven tool for temporal selection of environmental predictors and therefore may be more appropriate than choosing a time window based on *a priori* assumptions, particularly in forecasting applications. The algorithm relied on the data and predictive performance to select the window used for the next year's forecast. This framework allowed for predictor variables to change adaptively as a more accurate window became apparent. This quality of the sliding window algorithm makes it intuitive and potentially preferable in the face of a changing climate. However, the algorithm was computationally intensive. The R code to select windows for all variables/grids for the retrospective forecasting procedure took approximately 1.5 d to complete on a desktop computer with a 3.60 GHz processor with four cores and 32 GB of RAM. Each year took approximately 3.5 h to complete (depending on the number of years the forecast cross-validation procedure was conducted on). The flexibility of the approach also hinders the ability to typify a year as "warm" or "cold", as these criteria may change when an additional year of data is included and a new window is selected. An additional drawback of the sliding window algorithm is that it may be difficult to explain to managers and stakeholders, which may lead to confusion and distrust in the method.

Model-averaging across grid cells for SST and SIC was also an objective, adaptive, and data-driven, (though computationally-intensive) solution for dealing with the spatial nature of predicting run timing from these two variables. Of course, it would be possible to average all daily values across grid cells each year and perform the sliding window algorithm on these means. However, this would ignore the fact that some grid cells inherently have a stronger timing signal and would likely insert more variation into the predictive relationship. Additionally, the model had the flexibility to place more weight on different grid cells as more years of data were added, again adding to the flexibility of our overall approach which may be preferable in the face of a changing climate. However, the inherent complexity of including the spatial structure again makes it more difficult to typify a year as "warm" or "cold" as there are many values each year for SIC and SST and the strength of the run timing signal given by each grid cell varies.

These two complexities to our analysis (sliding window selection and model-averaging across spatial grid cells) made the interpretation of effect sizes and selected windows difficult in a biologically-meaningful way because the best windows and spatial grid cells could change from year to year. We generally see this flexibility based on predictive performance as more important for this particular analysis than biological inference on research topics like determining the most influential variable on run timing variation or determining the migration route of Chinook salmon through Kuskokwim Bay. These examples remain exciting research questions for the future, however our focus was on the prediction of a single quantity, D_{50}, which could aid in in-season decision-making. A separate issue confounding biological interpretation is that each variable had some direct or indirect link to temperature, suggesting that there is strong potential for multicollinearity among predictor variables. It is well known that correlated predictor variables can result in biased coefficient estimates and variance inflation (Neter et al., 1996). This was one reason coefficient estimates were only presented in the single-predictor case (Table 2), as we caution against their interpretation in this particular case. However, our focus was entirely on predictive ability, which is generally thought to be unaffected by multicollinearity (Graham, 2003).

An important caveat of our analysis is that we assumed a negligible influence of downstream harvest on the BTF timing index. This assumption was likely violated in some years but the magnitude of the impact is unknown. From 2008–2015, the average exploitation rate by only those villages downstream of the BTF index was 17% (versus 35% for villages across the whole drainage), thus there is the potential for a downstream harvest bias on perceived run timing. Moderate to high exploitation rates would not necessarily bias the BTF index if the timing of the harvest was similar to the run. However, in the Kuskokwim, subsistence harvest has historically focused on the early portion of the run (Hamazaki, 2011). When coupled with moderate exploitation rates (i.e., 17%), this early nature of the fishery is likely to have resulted in detected timing curves that were biased late (due to early fish being removed before they are sampled by the BTF) to some unknown degree. Historical and future interpretation of the BTF is further complicated by the operation of the fishery, such as a recent regulatory measure which mandates that no directed Chinook fishery may begin on or before 11 June. We suspect that the magnitude of the bias in the index due to the timing of downstream harvest would be small and would not likely affect the general conclusions of our analysis (although residual variation in environmental-run timing relationships would likely be lower if accounted for). We suggest that future studies should attempt to develop methods that remove harvest effects from the BTF index and other similar indices and assess the magnitude of potential bias. Even if harvest bias could be removed from the index, addressing bias in the test fishery index would be unfeasible during the season because spatially and temporally explicit harvest data are often unavailable until the season has concluded, and the data regarding the temporal distribution are fragmentary.

It was unsurprising that no meaningful relationship exists between estimated run size and run timing. Given that small/early and large/late runs are problematic for in-season management (Adkison and Cunningham, 2015), we see the lack of a relationship as beneficial to the management effort. In other words, a small run is no more likely to be early than it is to be late, and the same is true of large runs. For managers, this means that although these small/early or large/late scenarios have occurred in the past, they need not be particularly worried about them due to an overwhelming prevalence over other run size/run timing scenarios.

There is evidence to suggest that, on a population demographic scale, sub-stock structure and relative stock composition may influence the run timing of the aggregate. For example, Clark et al. (2015) showed that Chinook salmon that travel farther in the drainage to spawn (i.e., headwaters) enter the main stem earlier in the season. This point is supported in the Kuskokwim River based on unpublished ADFG radio telemetry data, which show that the date at which 50% of headwaters fish were tagged occurred as many as 10 or 11 days earlier than tagging of fish bound for middle river and lower river tributaries. Thus, it is more appropriate to view the timing curve detected by the BTF index as a mixture distribution made up of several distinct sub-stocks, each entering at different times. The cumulative effect of this is one curve that looks logistic likely because the various stocks overlap to a large extent. However, it is not difficult to see that if in some years the headwaters sub-stocks made up a greater proportion of the aggregate stock than the lower and middle river sub-stocks, the timing curve of the aggregate would be earlier than if other stocks had a greater contribution. Using genetic techniques, Anderson and Beer (2009) found that variations in the relative abundances of the populations composing the spring Chinook salmon run in the Columbia River, USA, explained 62% of the variation in annual run timing. This is a source of variation that was not accounted for in this analysis for at least two reasons. First, the resolution to divide the aggregate curve into its sub-stock components is not available: Kuskokwim River Chinook telemetry studies were conducted from 2003 to 2007 and 2015–2016 and the aggregate timing curve does not deviate enough from the smooth logistic curve to separate the different sub-stock components. Second, information on the relative contribution both in the past and in the forecast year would be necessary to include this complexity in the run timing forecast. This detailed level of sub-stock information is not available for the Kuskokwim River. The telemetry data can shed some light on these issues, but they are confounded by factors like harvest timing (some stocks may be harvested preferentially purely due to the timing of the fishery, which does not mirror that of the aggregate run; Hamazaki, 2011) or the potential of tagging stock components in some proportion other than their true contribution.

Methods exist to incorporate run timing forecasts from our analysis into in-season assessment and management efforts. Our daily predictions of *EOS*, which is an index of run size, could be used to predict the total end-of-season run size on each day using a regression model that relates historical reconstructed total run abundance and *EOS*. These in-season run predictions could be used to update pre-season run size forecasts with in-season data using methods such as inverse variance weighting (e.g., Walters and Buckingham, 1975) or Bayesian inference (e.g., Fried and Hilborn, 1988). Information-updating may be preferable in cases when the pre-season run forecast is biased, because it would allow for the perception of run size to pull away from the forecast when in-season data suggest it is highly unlikely. As we have shown here, uncertainty in *EOS* predictions is a function of the precision in the anticipated proportion of the run completed-to-date ($\hat{p}_d$; Eq. (9)). Our analysis suggests that incorporating run timing forecasts into estimates of $\hat{p}_d$ (and thus *EOS*) may provide managers with more certainty regarding interpretation of in-season abundance indices, which would facilitate updating of pre-season forecasts with data from the run.

Acknowledgements

We would first to thank Lew Coggins (U.S. Fish and Wildlife Service) for his work in advancing the in-season assessment and management efforts in the Kuskokwim River, much of which has inspired the analyses here. We would like to thank all ADFG Kuskokwim River biologists who have ever worked to collect and ensure the accuracy of the hugely important BTF index, without which this analysis would not have been possible. Many thanks are due to the scientists at NOAA and the Alaska Climate Research Center for collecting and compiling the necessary climatic data, in particular Nate Mantua for his work on the PDO index used here and making it readily available in a timely manner. Finally, we would like to thank Kuskokwim Area Fisheries Research Biologist Zachary Liller and three anonymous reviewers for their contributions in revising the manuscript. This work was funded by the Bering Sea Fisherman's Association (administered by the Arctic-Yukon-Kuskokwim Sustainable Salmon Initiative) project AC-1601. The findings and conclusions in this paper are those of the authors and do not necessarily represent the views of the U.S. Fish and Wildlife Service.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fishres.2017.05.003.

References

Adkison, M.D., Cunningham, C.J., 2015. The effects of salmon abundance and run timing on the performance of management by emergency order. Can. J. Aquat. Sci. 72 (10), 1518–1526.

Anderson, J.J., Beer, W.N., 2009. Oceanic, riverine, and genetic influences on spring Chinook salmon migration timing. Ecol. Appl. 19 (8), 1989–2003.

Arlot, S., Celisse, A., 2010. A survey of cross-validation procedures for model selection. Stat. Surv. 4, 40–79.

Bailey, L.D., van de Pol, M., 2015. climwin: Climate Window Analysis. R Vignette available at: https://cran.r-project.org/web/packages/climwin/climwin.pdf.

Beer, W.N., 2007. Run timing of adult Chinook salmon passing Bonneville dam on the Columbia River. Appendix 7 in J.J. Anderson, Columbia River Salmon Passage (CRiSP) Model Monitoring and Evaluating Support Annual Report, October 1, 2006–September 30, 2007. U.S. Department of Energy, Bonneville Power Administration, Division of Fish and Wildlife, Portland, Oregon, USA.

Bolker, B., 2008. Ecological Models and Data. R. Princeton University Press, New Jersey.

Bromaghin, J.F., 2005. A versatile net selectivity model, with application to Pacific salmon and freshwater species of the Yukon River, Alaska. Fish. Res. 74 (1), 157–168.

Bue, B.G., Schaberg, K.L, Liller, Z.W., Molyneaux, D.B., 2012. Estimates of the historic run and escapement for the Chinook salmon stock returning to the Kuskokwim River, 1976–2011. Alaska Department of Fish and Game, Report to the Board of Fisheries, Anchorage, AK. Available at: http://www.adfg.alaska.gov/FedAidPDFs/FDS12-49.pdf.

Burger, C.V., Wilmot, R.L., Wangaard, D.B., 1985. Comparison of spawning areas and times for two runs of Chinook salmon (Oncorhynchus tshawytscha) in the Kenai River, Alaska. Can. J. Aquat. Sci. 42 (4), 693–700.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, second ed. Springer-Verlag, New York, NY, USA (484 pp.).

Catalano, M.J., Jones, M.L., 2014. A simulation-based evaluation of in-season management tactics for anadromous fisheries: accounting for risk in the Yukon River fall chum salmon fishery. N. Am. J. Fish. Manag. 34 (6), 1227–1241.

Clark, S.C., Tanner, T.L., Sethi, S.A., Bentley, K.T., Schindler, D.E., 2015. Migration timing of adult Chinook salmon into the Togiak River, Alaska, watershed: is there evidence for stock structure? Trans. Am. Fish. Soc. 144 (4), 829–836.

Cook, S.J., Hinch, S.G., Farrell, A.P., Patterson, D.A., Miller-Saunders, K., Welch, D.W., Donaldson, M.R., Hanson, K.C., Crossin, G.T., Mathes, M.T., Lotto, A.G., Hruska, K.A., Olsson, I.C., Wagner, G.N., Thomson, R., Hourston, R., English, K.K., Larsson, S., Shrimpton, J.M., Vand der Kraak, G., 2008. Developing a mechanistic understanding of fish migrations by linking telemetry with physiology, behavior, genomics and experimental biology: an interdisciplinary case study on adult Fraser River sockeye salmon. Fish 33 (7), 321–339.

Cooperman, M.S., Hinch, S.G., Crossin, G.T., Cooke, S.J., Patterson, D.A., Olsson, I., Lotto, A.G., Welch, D.W., Shrimpton, J.M., Van der Kraak, G., Farrell, A.P., 2010. Effects of experimental manipulations of salinity and maturation status on the physiological condition and mortality of homing adult sockeye salmon held in a laboratory. Physiol. Biochem. Zool. 83 (3), 459–472.

Eiler, J.H., Evans, A.N., Schreck, C.B., 2015. Migratory patterns of wild Chinook salmon *Oncorhynchus tshawytscha* returning to a large, free-flowing river basin. PLoS One 10 (4), e0123127.

Fried, S.M., Hilborn, R., 1988. Inseason forecasting of Bristol Bay, Alaska, sockeye salmon (Oncorhynchus nerka) abundance using Bayesian probability theory. Can. J. Fish. Aquat. Sci. 45 (5), 850–855.

Graham, M.H., 2003. Confronting multicollinearity in ecological multiple regression. Ecology 84 (11), 2809–2815.

Hamazaki, T., Evenson, M.J., Fleischman, S.J., Schaberg, K.L., 2012. Spawner-recruit Analysis and Escapement Goal Recommendation for Chinook Salmon in the Kuskokwim River Drainage. Alaska Department of Fish and Game, Fish. Man. Ser. No. 12-08, Anchorage, AK. http://www.adfg.alaska.gov/FedAidPDFs/FMS12-08.pdf.

Hamazaki, T., 2011. When people argue about fish the fish disappear: fishery closure windows scheduling as a means of changing the Chinook salmon subsistence fishery pattern: is it an effective management tool? Fish 33 (10), 495–501.

Hamon, T.R., Foote, C.J., Hilborn, R., Rogers, D.E., 2000. Selection on morphology of spawning wild sockeye salmon by a gill-net fishery. Trans. Am. Fish. Soc. 129 (6),

1300–1315.
Hasler, A.D., Scholz, A.T., 1983. Olfactory Imprinting and Homing in Salmon. Springer-Verlag, Berlin (134 pp.).
Hinch, S.G., Cooke, S.J., Farrell, A.P., Miller, K.M., Lapointe, M., Patterson, D.A., 2012. Dead fish swimming: a review of research on the early migration and high premature mortality in adult Fraser River sockeye salmon *Oncorhynchus nerka*. J. Fish. Biol. 81 (2), 576–599.
Hodgson, S., Quinn, T.P., Hilborn, R., Francis, R.C., Rogers, D.E., 2006. Marine and freshwater climatic factors affecting interannual variation in the timing of return migration to fresh water of sockeye salmon (*Oncorhynchus nerka*). Fish. Oceanogr. 15 (1), 1–24.
Keefer, M.L., Peery, C., Jepson, M., Stuehrenberg, L., 2004. Upstream migration rates of radio-tagged adult Chinook salmon in riverine habitats of the Columbia River basin. J. Fish Biol. 65 (4), 1126–1141.
Keefer, M.L., Peery, C.A., Caudill, C.C., 2008. Migration timing of Columbia River spring Chinook salmon: effects of temperature, river discharge, and ocean environment. Trans. Am. Fish. Soc 137 (4), 1120–1133.
Linderman Jr., J.C., Bergstrom, D.J., 2009. Kuskokwim management area: salmon escapement, harvest, and management. In: Krueger, C.C., Zimmerman, C.E. (Eds.), Pacific Salmon: Ecology and Management of Western Alaska's Populations. American Fisheries Society, Symposium 79, Bethesda, Maryland, pp. 541–599.
Mantua, N.J., Hare, S.R., Zhang, Y., Wallace, J.M., Francis, R.C., 1997. A Pacific interdecadal oscillation with impacts on salmon production. Bull. Am. Meterol. Soc. 78, 1069–1079.
Mundy, P.R., Evenson, D.F., 2011. Environmental controls of phenology of high-latitude Chinook salmon populations of the Yukon River, North America, with application to fishery management. ICES J. Mar. Sci. 68 (6), 1155–1164.
Neter, J., Kutner, M.H., Nachtsheim, C.J., Wasserman, W., 1996. Applied Linear Statistical Models. Irwin Chicago, Illinois, USA.
O'Malley, K.G., Ford, M.J., Hard, J.J., 2010. *Clock* polymorphism in Pacific salmon: evidence for variable selection along a latitudinal gradient. Proc. R. Soc. Lond.: Biol. Sci. 277, 3703–3714.
Quinn, T.P., Unwin, M.J., Kinnison, M.T., 2000. Evolution of temporal isolation in the wild: genetic divergence in timing of migration and breeding in introduced populations of chinook salmon. Evolution 54 (4), 1372–1385.
Reynolds, R.W., Smith, T.M., Liu, C., Chelton, D.B., Casey, K.S., Schlax, M.G., 2007. Daily high-resolution blended analyses for sea surface temperature. J. Clim. 20, 5474–5496.
Salinger, D.H., Anderson, J.J., 2006. Effects of water temperature and flow on adult salmon migration swim speed and delay. Trans. Am. Fish. Soc. 135 (1), 188–199.
Staton, B.A., Catalano, M.J., Fleischman, S.J., 2017. From sequential to integrated Bayesian analyses: exploring the continuum with a Pacific salmon spawner-recruit model. Fish. Res. 186 (1), 237–247.
van de Pol, M., Bailey, L.D., McLean, N., Rijsdijk, L., Lawson, C.R., Brouwer, L., 2016. Identifying the best climatic predictors in ecology and evolution. Method Ecol. Evol. 7 (10), 1246–1257.
Walters, C.J., Buckingham, S., 1975. A control system for intraseason salmon management. International Institute for Applied Systems Analysis Working Paper. WP-75-028. 19 pp.
Wolfe, R.J., Spaeder, J., 2009. People and salmon of the Yukon and Kuskokwim drainages and Norton Sound in Alaska: fishery harvests, culture change, and local knowledge system. In: Krueger, C.C., Zimmerman, C.E. (Eds.), Pacific Salmon: Ecology and Management of Western Alaska's Populations. American Fisheries Society, Symposium 79, Bethesda, Maryland, pp. 541–599.

APPENDIX B

Bayesian information updating procedures for Pacific salmon run size indicators: Evaluation in the presence and absence of auxiliary migration timing information

Benjamin A. Staton[1,2]

[1]School of Fisheries, Aquaculture, and Aquatic Sciences, Auburn University, 203 Swingle Hall, Auburn AL, 36849, bas0041@auburn.edu

[2]U.S. Fish and Wildlife Service, Yukon Delta National Wildlife Refuge, P.O. Box 346, Bethel, AK 99559, benjamin_staton@fws.gov

***This author is currently affiliated with both organizations*

Matthew J. Catalano[1]

[1]School of Fisheries, Aquaculture, and Aquatic Sciences, Auburn University, 203 Swingle Hall, Auburn AL, 36849, mjc0028@auburn.edu

Keywords: Chinook salmon, abundance estimation, in-season salmon assessment, Bayesian inference, migration timing

Corresponding Author:

Benjamin A. Staton

203 Swingle Hall, Auburn University, Auburn AL, 36849

Phone: (269) 369-6438

Email: bas0041@auburn.edu

Abstract

Pre-season forecasts of Pacific salmon run strength are notoriously uncertain, and are thus often updated using various abundance indices collected during the run. However, interpretation of these in-season indices is confounded by uncertainty in migration timing. We assessed the performance of two Bayesian information-updating procedures for Kuskokwim River Chinook salmon: one that uses auxiliary run timing information and one that does not, and compared the performance to methods that did not involve updating. We found that in-season Bayesian updating provided more accurate run size estimates during the time when harvest decisions needed to be made, but that the incorporation of run timing forecasts had little utility in terms of providing more accurate run size estimates. The latter finding is conditional on the performance of the run timing forecast model we used; a more accurate timing forecast model may yield a different conclusion. The Bayesian approach we developed provided a probabilistic expression of run size beliefs, which could be useful in a transparent risk-assessment framework for setting and altering harvest targets in-season.

1. Introduction

Management strategies for in-river Pacific salmon *Oncorhynchus* spp. fisheries involve limiting harvest in-season such that some management reference point is achieved. These reference points are typically expressed as either a target escapement abundance or a target exploitation rate, or as ranges of these two quantities. Regardless of which way the management strategy is framed, a reliable measurement of the harvestable surplus is required to successfully implement the strategy on an annual basis. The harvestable surplus varies annually based on the total incoming run size, thus information regarding the total annual run size is often required for management of these fisheries. Run size information can be categorized into two broad classes: pre-season forecasts (i.e., before fish have arrived in fishery areas) and in-season estimates (i.e., once fish can be indexed). Though these terms are often used interchangeably (as are "predictions" and "projections"), for clarity we will refer to methods of the former class as "forecasts" and methods in the latter class as "estimates".

Methods to produce pre-season forecasts range from simple models based only on time series patterns (Haeseker et al. 2005) to complex models that incorporate spawner-recruit relationships (Adkison and Peterman 2000), sibling relationships (Peterman 1982), and/or environmental variables intended to explain variability in survival rates (Adkison and Peterman 2000). Murphy et al. (2017) presented pre-season forecast methodology for Yukon River Chinook salmon *O. tshawytscha* based on trawl surveys targeting juveniles shortly after marine entry, and this model has recently shown promise (K. Howard, pers. comm.). Not surprisingly, it has commonly been found that simpler models that do not require hypotheses about mechanisms driving recruitment variability perform as well or better than more complex forecast models that require such assumptions (Haeseker et al. 2005, 2008; Winship et al. 2015). Still, pre-season forecast models generally perform poorly and have wide uncertainty regions, resulting from incomplete understanding of drivers of survival and recruitment rates (Adkison et al. 1996; Adkison and Peterman 1999). Inaccurate annual forecasts have socioeconomic consequences for the fisheries that rely on them: Bocking and Peterman (1988) found correlations between forecast errors and management performance and Costello et al. (1998) found a high expected value of information for better forecasts

resulting from improved knowledge of the El Niño phase. These findings highlight the need for improved methods to produce pre-season forecasts or otherwise update them with in-season estimates as those data accumulate.

In-season estimators of run size also show quite a range of model complexity. Simple methods may be based purely on catch-per-effort (CPE) indices whereas more complex methods may incorporate observations of size/age structure (Flynn and Hilborn 2004) and substock structure (Hyun et al. 2005). Each of these methods attempt to expand some partially-observed component of the run to the total run size, and thus their predictive performance is tightly linked to uncertainty regarding run timing (i.e., what fraction of the run is complete on any given day of the season). For example, large/late runs and early/small runs have a tendency to create similar CPE indications of an average run early in the season (Adkison and Cunningham 2015), though neither run scenario would likely have the same harvestable surplus as an average run. Put another way, with the observation of indications of an average-sized run, the manager can rarely exclude these other extreme scenarios from consideration, resulting in uncertainty about how to prosecute the fishery to ensure the management strategy is implemented and annual fishery objectives are achieved. For this reason, many efforts have been made at forecasting the run timing pre-season as well as the run size (Staton et al. 2017; Mundy and Evenson 2011; Keefer et al. 2008; Anderson and Beer 2009). However, it is often unclear as to precisely how these run timing forecasts are to be included into run size estimators, or whether it is preferable to do so at all.

In the presence of multiple run size indicators (i.e., pre-season and in-season sources), it is often difficult to decide which information sources to trust at which points in the season for making management decisions when they inevitably disagree. One extreme would be to manage harvests based on the pre-season forecast all season and entirely ignore any indications provided by in-season estimates. The other extreme would be to do the opposite: abandon the pre-season forecast the day the first fish is detected by the in-season index project(s). It is our sense that few managers would feel comfortable taking either of these extremes, which implies that some method of transitioning from a pre-season forecast to in-season estimates is warranted. While some managers may prefer transitional approaches

based on experience and intuition, a logical method to perform such a transition is based on the variance of each information source: sources with less uncertainty should drive management decisions more than those that are more uncertain. The calculations to conduct a formal variance-based transition can be framed in a classical inferential framework (Walters and Buckingham 1975) or as a Bayesian inferential problem (Fried and Hilborn 1988, Hyun et al. 2005). The Bayesian approach has a certain appeal as it provides a full probability model representing uncertainty regarding the truth of all possible run size outcomes (i.e., hypotheses), which can be seamlessly updated as new (i.e., in-season) information is made available. Such a probability model could be useful in formal risk assessments in the context of probabilistic control rules (Catalano and Jones 2014, Prager et al. 2003) used to set harvest targets.

The Kuskokwim River, located in Western Alaska, is a large drainage system that supports large subsistence fisheries for Chinook salmon. Being the species of greatest subsistence interest for this region and coupled with recent low abundances, Chinook salmon have been of primary management concern and is hereafter the focus of this paper. Although the river system is quite large (main-stem > 800 km, drainage area > 50,000 km^2), the majority of the fishery is (in relation) spatially-constricted: 95% of the drainage-wide Chinook salmon harvest is attributable to the 16 villages located in the first 300 km of the main-stem and 70% of the total Chinook salmon harvest is attributable to the 10 villages in the first 125 km (Hamazaki 2011). The fishery is managed with time and area closures implemented by emergency order intended to limit harvest to ensure a drainage-wide fixed escapement goal is met each year. Information sources for in-season management include a pre-season run size forecast and an in-season CPE index of in-river abundance and species composition (the Bethel Test Fishery, BTF, operated annually from June 1-August 24 1984-2017; Bue and Lipka 2016). In recent years, in-season harvest estimates have also been produced (Staton and Coggins 2016, 2017) and have been used to track progress toward the attainment of total allowable harvest. Currently, no formal attempts at producing in-season estimates/updates of run size have been made. Decisions about limiting harvest opportunity have instead been made by qualitatively determining if the BTF index indicates a different run size than that suggested by the pre-season forecast by comparing the accumulation of daily CPE against previous years. This

approach obviously has substantial pitfalls that include: (1) the aforementioned confounding effect of annual variability in run timing, (2) no accounting of annual variability in BTF catchability (i.e., run-per-index; Flynn and Hilborn 2004), (3) no formal consideration of which source provides more information about the true run size at varying points in the season, and (4) no explicit expression of how disagreements between the BTF index and the pre-season forecast should result in alterations to the in-season harvest management strategy (i.e., total allowable harvest).

In this paper, we seek to address these issues by developing a framework to formally update pre-season run size forecasts with in-season estimates of the total run size using Bayesian inference. Using data from the Kuskokwim River Chinook salmon fishery, we evaluated the assessment framework by applying it to previous years as well as determined the potential utility of incorporating auxiliary information from a recently-developed run timing forecast model for this fishery (Staton et al. 2017). Our objectives were to (1) develop two Bayesian updating tools: one that ignores auxiliary run timing information and one that includes it, (2) determine if Bayesian updating provides better (more accurate/precise) inference than using either the forecast or in-season estimates alone, and (3) determine if incorporating the run timing forecast information improves inferential performance.

2. Methods

We developed a Bayesian approach to updating the pre-season perception of run size with in-season data, both in the presence and absence of auxiliary run timing information. The approach proceeds by (1) determining the pre-season run size forecast for each evaluated year to serve as the prior distribution (2) obtaining a likelihood function for observed data (cumulative CPE data through day d of the run) assuming any given run size hypothesis was correct – which involved multiple steps and used a sequential and parametric Monte Carlo algorithm to estimate this function, and (3) the formal combination of the information derived in steps (1) and (2) using Bayes' Theorem to obtain a posterior probability function. The presence or absence of auxiliary run timing information was incorporated into step (2) when interpreting the consistency of the CPE data with any one run size hypothesis. The

reliability of inferences from the prior, likelihood, and posterior were then compared between cases including and ignoring the auxiliary run timing information.

We performed a leave-one-out cross-validation evaluation of Bayesian approach we developed by using all currently available information to measure inferential performance. This approach was not retrospective, as it did not use only information available at the time the tools would have been used in previous years. Our chosen analysis framework emphasizes that we were not interested in how the tools would have performed in the past, but rather how they may perform in the future when presented with runs similar to those that have occurred in the past. The leave-one-out method was used so the model was not trained using the exact year it was attempting to estimate. The years 1995-2017 were evaluated by producing weekly updates on June 10, June 17, June 24, July 1, July 8, and July 15.

2.1 Pre-season run size forecast

Pre-season run size forecasts for Kuskokwim River Chinook salmon are made by assuming the current year's run will be similar in size to the previous year's run, which stems from the observation of high serial auto-correlation in the run abundance time series (Figure 1a). The total run size each year is estimated post-season using a maximum likelihood drainage-wide run reconstruction model that integrates information from 20 escapement indices, fishery CPE data, mark-recapture-based estimates of drainage-wide abundance, and total fishery harvest over the time period of 1976-2017 [originally published in Bue et al. (2012) and most recently presented in Smith and Liller (2018)]. The most recent estimates provided in Smith and Liller (2018) were used in this analysis and we assumed the point estimates represented the true run size in these years. Although the "last-year" rule for producing forecasts has only been used since 2014, we can hindcast its performance over the entire time series to obtain the precision of the forecast rule if it had been used in the past. Errors in the forecast were assumed to be multiplicative:

(1) $$\varepsilon_{F,t} = \log\left(\frac{N_{t\text{-}1}}{N_t}\right)$$

where N_t and $N_{t\text{-}1}$ are the run sizes corresponding to year t and t-1, respectively, and $\varepsilon_{F,t}$ is the natural logarithm of the multiplicative error term in the forecast (values < 0 are underestimates and values > 0 are overestimates). The time series of all such $\varepsilon_{F,t}$ is presented in Figure 1b, and their distribution is shown in Figure 1c. The standard deviation of $\varepsilon_{F,t}$, hereafter denoted σ_F, was used to represent the uncertainty in the forecast in any given year in the analysis, expressed as a bias-corrected lognormal distribution. This lognormal distribution was assumed to represent the prior knowledge regarding the size of the run in the absence of in-season assessment data.

2.2 Pre-season run timing forecast

Run timing forecasts for the Kuskokwim River Chinook salmon stock were produced using the methodology presented in Staton et al. (2017). Briefly, the forecast model predicts the day of the year at which 50% of the total annual cumulative CPE will be observed in the BTF (hereafter $D_{50,t}$) by exploiting linear regression relationships between $D_{50,t}$ and sea surface temperature, sea ice concentration, air temperature in Bethel, AK, and the Pacific Decadal Oscillation index. The forecast model was developed using variable selection criteria to determine the best time periods of these variables to include and model-averaging to handle forecast model uncertainty. We used the Staton et al. (2017) timing forecast model to produce forecasts of $D_{50,t}$ for the years 1995-2017, as well as their associated standard errors of prediction.

2.3 Likelihood Function Construction

Information about run size is contained in the end-of-season cumulative BTF CPE values (hereafter denoted as EOS_t; Figure 2), and thus these data formed the foundation of linking in-season abundance index data to different run size hypotheses in a likelihood framework.. The construction of the likelihood function contained three steps: (1) accounting of the unobserved portion of EOS_t on any given day d of the season (i.e., a run timing expansion), (2) expansion of the EOS_t index to the total run vulnerable to sampling by the BTF (i.e., a catchability expansion) and (3) accounting for harvest that occurred downstream of the BTF index site (i.e., fish that were never vulnerable to sampling by the BTF). Given substantial uncertainty exists in each of these steps that must be accounted for in the likelihood

function, we used parametric Monte Carlo (MC) methods to propagate the uncertainty from one step to the next by drawing $1x10^6$ samples from each respective distribution and performing the respective calculations on each MC sample, hereafter indexed by *b*.

In the first step (the run timing expansion), the observed cumulative CPE through day *d* in year *t* (hereafter denoted as $CCPE_{t,d}$) was divided by each MC sample of the anticipated proportion of the run complete on day *d* ($p_{d,b}$) to obtain the MC values of $\widehat{EOS}_{d,b}$, which are samples of the index of run size:

$$\widehat{EOS}_{d,t,b} = \frac{\sum_{i=1}^{d} CPE_{t,i}}{p_{d,b}} \quad (2)$$

The way $p_{d,b}$ was obtained was the only difference between the methods in which auxiliary run timing information was either present or absent. For the method that ignored the presence of the run timing forecast, historical daily proportions of EOS_t were calculated for each previously-observed year. Then, daily beta distributions were fitted to these proportions using the method of moment-matching (Bolker 2008), such that the mean and variance of the sample proportions and random variables from the beta distribution were equal. For this method, MC samples from these daily beta distributions were used as $p_{d,b}$. For the method that used run timing forecasts, MC samples of the forecasted quantity ($D_{50,t,b}$) were inserted into the logistic function:

$$p_{d,b} = \frac{1}{1+e^{-h_b(d-D_{50,t,b})}} \quad (3)$$

to obtain $p_{d,b}$. Here, h_b is a MC sample from the historical distribution of steepness parameters and represents the degree of compression or protraction of the run timing curve (i.e., the rate at which the run approaches completion). We have found that D_{50} and *h* are independent for the Kuskokwim River Chinook salmon population, so MC draws capturing their variability were made from independent normal distributions.

The second step in the construction of the likelihood functionused the historical relationship between EOS_t and total vulnerable run size ($N_{vuln,t}$) to expand from the CPE scale to the abundance scale (Figure 2). $N_{vuln,t}$ was calculated as the total run minus the season-wide harvest downstream of the BTF index site. Although BTF data cover the span of 1984-2017, spatially-explicit harvest data were available

only for 1990-2017 which prevented calculation of $N_{vuln,t}$ for the years 1984-1989. Downstream harvest was calculated as the sum of harvests from all villages downstream of Bethel, plus half of the harvest from the village of Bethel, plus total annual harvest in the district W1-B commercial fishery. We tried other values of the assumed amount of Bethel's harvest that has historically occurred downstream of the BTF site, and none affected the results of our analysis, so we opted for the "naïve" value of half. A regression relationship was then fitted of the form:

$$\log(N_{vuln,t})=\beta_{0,q}+\beta_{1,q}\log(EOS_t)+\varepsilon_{t,q} \quad (4)$$

where $\varepsilon_{t,q}$ is normally distributed with mean zero and standard deviation equal to σ_q. Due to a change in the BTF catchability in 2008 resulting from a transition between net-makers (Bue and Lipka 2016), the shape of this relationship changed requiring the use of two regression models (q subscript): one fitted to the years 1990-2007 and one fitted to the years 2008-2017 (Figure 2). To produce in-season estimates of $N_{vuln,t,b}$, joint MC samples of $(\hat{\beta}_{0,q,b},\hat{\beta}_{1,q,b})$ were drawn from a bivariate normal distribution with covariance matrix equal to that estimated from the regression procedure, and bias-corrected MC residual deviates were drawn as $\hat{\varepsilon}_{q,b} \sim N(-0.5\sigma_q^2, \sigma_q)$. These MC regression parameters as well as the MC values of $\widehat{EOS}_{d,t,b}$ were inserted into the predictive relationship:

$$\widehat{N}_{vuln,t,d,b}=e^{\hat{\beta}_{0,q,b}+\hat{\beta}_{1,q,b}\log(\widehat{EOS}_{d,t,b})+\hat{\varepsilon}_{q,b}} \quad (5)$$

to obtain MC samples $\widehat{N}_{vuln,t,d,b}$.

The third step in the construction of the likelihood function involved adding MC samples of cumulative harvest downstream of the BTF index site ($H_{d,t,b}$) to $\widehat{N}_{vuln,t,d,b}$ to obtain MC values of the total run size estimate on day d ($\widehat{N}_{t,d,b}$). Commercial Chinook salmon harvest by day was assumed known without error, and the cumulative sum was calculated on each day and year to obtain $H_{com,d,t}$. A method to reconstruct cumulative subsistence harvest downstream of the BTF was needed, as in-season harvest data have only recently been available (2016 and 2017). The Alaska Department of Fish and Game (ADF&G) has collected subsistence harvest calendars which provide an indication of the timing of harvest from each village (Shelden et al. 2016; Hamazaki 2008). Calendars from villages downstream of the BTF were

combined, then converted to cumulative harvest proportions for each day *d* and year *t*. Multiplying these annual cumulative proportion time series by the total season-wide harvest downstream of the BTF provided estimates of $H_{sub,d,t}$. To generate MC samples ($H_{d,t,b}$), random values of subsistence harvest ($H_{sub,d,t,b}$) were generated from a bias-corrected lognormal distribution with standard deviation equal to 0.15 and the constant $H_{com,d,t}$ was added. These $H_{d,t,b}$ values were added to the $\widehat{N}_{vuln,t,d,b}$ values to obtain MC samples of $\widehat{N}_{t,d,b}$.

Once the MC samples of $\widehat{N}_{t,d,b}$ were obtained (which represented random draws from the likelihood function), the form of the likelihood probability density function (PDF) was estimated using a one-dimensional kernel density estimator fitted to the $1x10^6$ MC samples. The resulting function is hereafter denoted by $\Pr(CCPE_{t,d}|N_{t,i})$.

2.4 Posterior estimation

To provide Bayesian in-season updates of the run size estimate, the lognormal distribution representing uncertainty in the pre-season forecast (Section 2.1) was used as the prior information each day [$\Pr(N_{t,i})$; *i* denotes a continuous run size hypothesis]. Although this was a simple one parameter Bayesian estimation problem, the likelihood PDF $\Pr(CCPE_{t,d}|N_{t,i})$ did not have a well-defined parametric form which could have allowed direct analytical calculation of the posterior PDF [$\Pr(N_{t,i}|CCPE_{t,d})$] using Bayes' Theorem. Instead, a custom random walk Metropolis-Hastings Markov Chain Monte Carlo (MCMC) algorithm (Chib and Greenberg 1995) was written using a lognormal proposal distribution. The lognormal proposal distribution was used as opposed to a symmetrical distribution (like the normal distribution) to prevent negative proposals. The standard deviation of this proposal distribution was tuned such that the acceptance rate of proposals was between 0.2-0.4 (Bédard 2007). Posterior convergence was assessed using two chains with over-dispersed initial values and the Potential Scale Reduction Factor (Brooks and Gelman 1998), and the Raftery-Lewis diagnostic was used to ensure enough effective samples were drawn to make adequate inference (Raftery and Lewis 1992). On each evaluated day and year, $1x10^5$ posterior samples were drawn from each chain with a burn-in period of $1x10^4$. These

specifications resulted in more than enough samples to meet the criteria for convergence and adequate inference in all cases. All analyses were conducted in Program R (R Core Team 2015) and all code and data are provided in the online supplement.

2.5 Metrics of estimator performance

Performance of the estimators was evaluated using four criteria for each evaluated day d and year t: (1) mean absolute proportional error (MAPE) to quantify the size of estimation errors, (2) mean proportional error (MPE) to measure bias, (3) the standard deviation of log-scale multiplicative errors (σ) to measure variability in estimation errors, and (4) the coverage of the 50%, 80%, and 95% confidence/credible regions. For calculation of MAPE, MPE, and σ, the median of the distributions $\Pr(N_{t,i})$, $\Pr(CCPE_{t,d}|N_{t,i})$, and $\Pr(CCPE_{t,d}|N_{t,i})$ were used as point estimates and the reconstructed values of N_t (Smith and Liller 2018) were interpreted as the true run sizes. The purpose of evaluating the performance of inferences from the prior, likelihood, and posterior PDFs was to determine whether Bayesian updating provided better performance than not updating and trusting either pre-season or in-season indicators completely for the duration of the season.

3. Results

3.1 Mean absolute proportional error

Errors in inference from the likelihood distribution alone (i.e., BTF data only) were quite large early in the season, but steadily declined in size as the run approached completion (Figure 3a). On June 10, the median of the likelihood function that used the historical timing frequencies exhibited larger errors on average (MAPE = 0.52) than did the one that used the timing forecast (MAPE = 0.4), but after this date MAPE values were essentially equal (Figure 3a). It was not until July 1 that inference from the likelihood distribution alone exhibited smaller MAPE values than the pre-season forecast (which had a constant MAPE = 0.18 all season). Conversely, posterior inference yielded smaller MAPE values starting from June 10 and remained lower than the pre-season forecast all season long (Figure 3a). Posterior inference when using the timing forecast showed slightly larger errors on average between June 10 and

July 1 (average MAPE = 0.17) than did posterior inference when using historical timing frequencies (average MAPE = 0.15).

3.2 Mean proportional error

In terms of biases, all indicators besides the pre-season forecast showed slight to moderate negative biases for the whole season (Figure 3b), but the magnitude generally decreased as the season progressed. The pre-season forecast was slightly positively biased (MPE = 0.04). Bias based on the likelihood distribution on June 10 differed greatly between the method that used the timing forecast (MPE = -0.22) and the method that did not (MPE = -0.01). After this date, both methods performed essentially equally-well and reached essentially no bias by June 24. For posterior inference, the only notable difference in performance was on June 10, when the method that used historical timing frequencies had less negative bias than the method that used the timing forecast.

3.3 Variability of errors

As would be expected from Figure 3a, the variability in errors was greater for inference from the likelihood distribution than for posterior inference early in the season, and did not become lower than the variability of pre-season forecast errors until July 1. As was found for MAPE, the variability of errors from posterior inference were smaller than the pre-season forecast for the entire season, and was always smaller than inference from the likelihood alone. The only notable difference between the method using run timing forecasts and the one ignoring them was smaller variability in errors in the likelihood distributions on June 10.

3.4 Credible region coverage

With the exception of the 50% region, the pre-season forecast (prior) had appropriate coverage levels (Table 1; appropriate defined here as coverage being within ± five percentage points of optimal coverage). Other indicators tended to have less coverage than appropriate on June 10 for the 50% region. Regions at the 80% and 95% levels were typically more appropriately estimated, with the exception of the posterior 95% regions on July 8 which appeared to be slightly too narrow (Table 1). In general, likelihood

316 coverage was more appropriate than posterior coverage on June 24 and July 8. Coverage values on July
317 15 are not presented as results were the same as for July 8.

318 **4. Discussion**

319 The findings of our analysis suggest that using the Bayesian in-season updating procedures
320 described here would provide more accurate in-season inference regarding the run size than the two
321 aforementioned extremes: trusting either the pre-season forecast or the in-season estimates all season.
322 This is not to say that pre-season forecasts are not useful, in fact they were critical in this analysis by
323 tempering the large errors in the in-season estimates early in the season (Figure 3a). Additionally, pre-
324 season forecasts are required for setting pre-season expectations and anticipated harvest strategies for the
325 season. Instead, we have shown that it is preferable to update the pre-season forecast as the season
326 progresses, which could result in alteration of in-season harvest targets if the posterior distribution is
327 different than the prior distribution.

328 We expect that the finding that updating is preferable to not updating is general to systems like
329 the Kuskokwim River (e.g., the Nushagak and Yukon Rivers located in Western Alaska). These systems
330 have similar in-season run size indicators (both systems have a sonar and the Yukon River has a lower-
331 river test fishery) that suffer from the same problems as the data for the Kuskokwim, namely variability in
332 index catchability and the confounding effect of run timing uncertainty. These two problems, particularly
333 the latter, generally lead to the in-season data providing inaccurate and highly uncertain run size estimates
334 for much of the season (Flynn and Hilborn 2004, Adkison and Cunningham 2015, Walters and
335 Buckingham 1975). Thus, it is logical to expect that the desirability of updating pre-season forecasts
336 rather than fully trusting in-season estimates (i.e., the likelihood function in Bayes' Theorem) alone
337 would be a general finding. The example of the Kuskokwim River assessment tool we developed is a
338 generalization of previously-developed updating methods (e.g., Fried and Hilborn 1988) and could be
339 generalized to other systems for the purpose of arbitrating between the relative information content of
340 pre-season and in-season run size indicators, including those involving mixed stocks if timely data on the
341 substock contribution to the total indicators were available.

We found that posterior inference gave smaller errors and less variability in errors than the pre-season forecast starting on the first day of our analysis: June 10. Under current regulation, directed Chinook salmon fisheries cannot begin prior to June 12 in the Kuskokwim River, which was implemented to allow the early-running headwaters fish to pass through the majority of the lower-river fishery to spawn and be vulnerable to fishers in the upper-river. Thus, updates before this point are of little utility under the current management regime, but updates after are critical. Our work has shown that updates in this latter period are not only reliable, but in fact preferable to not updating at all.

A key finding of our analysis was that incorporating auxiliary run timing information in the assessment provided no real gain in performance. There was even evidence to suggest ignoring the run timing forecast actually performed better than incorporating it (for posterior inference, Figures 3a,b). This is not overly surprising given that Staton et al. (2017) reported that using the mean $D_{50,t}$ values provided slightly more accurate run timing forecasts than the environmental variable forecast. This is likely a result of the large number of years with average timing: although $D_{50,t}$ has exhibited a range of 17 days, 35% of past years have been within ± 1 day of the mean and 53% of past years have been within ± 2 days of the mean (Staton et al. 2017). The conclusion of no gain in performance in the presence of the run timing forecast was conditional on the accuracy and precision of the Staton et al. (2017) forecast model; because we did not evaluate performance for other systems, this finding may not be general. For systems that show greater (and more predictable) annual variability in run timing, it very well may be preferable to incorporate the auxiliary timing information. It is also possible that a better timing forecast model for the Kuskokwim River may become available in the future, in which case this study should be replicated to determine whether and to what degree increased predictive performance regarding run timing is reflected in the performance of in-season run assessments. To our knowledge, our work is the first to formally compare the performance of in-season abundance estimators in the presence and absence of auxiliary run timing information, and it suggests that careful consideration should be made before making the decision to incorporate such information.

Our approach used the only available index of abundance for Kuskokwim River Chinook salmon to provide the information on which to perform the updates: the BTF. Although extensive monitoring activities occur in this drainage, they are used primarily for indexing escapement and given the size of the system, they are not useful for in-season assessment. There are eight weirs operated with some consistency in the system, but fish do not typically arrive until mid-July when much of the run has passed the majority of the harvest area. However, in systems for which this time lag is not so great (like Bristol Bay sockeye salmon in which escapement counting towers are often only several days of travel upstream of the fishery districts), it is possible to include other information into the likelihood component of the Bayesian calculations which may provide better inference. Recently, a lower-river sonar project has been operated in the Kuskokwim River, which could provide such additional information. However, given this project is still very much in its infancy, we suggest waiting until it can be shown that it provides a reliable index of the run. If it is proven to be reliable, a decision of how to incorporate sonar data will need to be made. Two methods are immediately obvious: (1) incorporate it as an additional likelihood term the calculation of the posterior or (2) calculate two posteriors (one for the BTF and one for the sonar) and perform Bayesian model averaging. The latter of the two options would be preferable if placing unequal prior probabilities on the two models is desired.

Although the likelihood component of Bayes' Theorem was derived from parameters without prior distributions (p_d, $\beta_{0,q}$, $\beta_{1,q}$, and σ_q), we still consider the approach to obtaining a daily posterior distribution of total abundance as fully Bayesian. Presumably, an approach could be constructed that incorporated all the MC steps we performed to obtain the likelihood formulation as well as the posterior calculations into a single MCMC algorithm. However, we see this as no more Bayesian than the approach we developed, as in either case the only parameter with a meaningful prior was total run abundance, and no information exists in the current year's CPE data to update any of these other parameters. Other Bayesian approaches (Fried and Hilborn 1988) present the estimation of likelihood functions without priors on the parameters that inform them, which is equivalent to the approach we used.

We sought to address the four previously-identified issues with qualitative salmon run size assessment, which we believe our Bayesian approach does fully. The first issue was inadequate treatment of run timing uncertainty: our method accounts for this uncertainty fully in the MC sampling from the historical (or forecasted) frequency of different run completion proportions on each day. The second issue was the lack of accounting for annual variability in BTF catchability: our method accounts for this in the full propagation of the uncertainty in the regression relationships in Figure 2. The third issue was a lack of the consideration of how much weight to place on pre-season versus in-season run size indicators, which our method handles intuitively using the laws of probability and Bayesian inference. The final issue was the lack of a formal expression for how disagreements in pre-season and in-season indicators should result in alterations to the harvestable surplus. While this last issue is much more about management than assessment, it is not difficult to see that our method provides the information to inform such a decision. On any day of the season, the probability of any escapement outcome of interest [e.g., Pr(escapement < lower bound of the escapement goal)] conditional on a candidate harvest target can be calculated from the posterior. If this probability is deemed unacceptable, additional candidates can be proposed until the probability of the escapement outcome is deemed suitable. A similar approach could easily be extended to salmon fisheries managed with limit exploitation rates, by calculating the posterior exploitation rate if the candidate harvest target were to be taken (candidate harvest divided by posterior samples of total abundance) and determining if the associated probability of falling above the limit rate is acceptable. These calculations are trivial once the posterior total abundance estimate has been obtained, and the overall procedure aligns closely with the probabilistic approach to the use of limit reference points in precautionary fisheries management, which has been gaining popularity in framing sustainable harvest policies for U.S. marine fisheries (Prager et al. 2003; Shertzer et al. 2011). Herein lies what we see as the greatest contribution of this work: it provides an assessment framework that can be used to provide greater transparency for harvest management decisions that are framed in terms of uncertainty and risk.

5. Acknowledgements

We would like to thank all ADF&G personnel involved with collecting the data from the BTF for the past three decades as well as the other assessment projects that made this work possible. Additionally, we would like to thank N. Smith (ADF&G),L. Coggins, and G. Decossas (U.S. Fish and Wildlife Service) for their comments on earlier drafts of this paper and input on the potential utility of the tools we evaluated for in-season management. We would also like to thank two anonymous reviewers for their insightful comments which ultimately improved this work. Funding for this work was provided by the Arctic-Yukon-Sustainable Salmon Initiative and the Bering Sea Fisherman's Association through a grant to M. Catalano. The findings and statements herein are those of the authors and do not reflect the views of the U.S. Fish and Wildlife Service.

References

Adkison, M.D., Peterman, R.M., Lapointe, M.F., Gillis, D.M., Korman, J., 1996. Alternative models of climatic effects on sockeye salmon, *Oncorhynchus nerka*, productivity in Bristol Bay, Alaska, and the Fraser River, British Columbia. Fisheries Oceanogr. 5:137-152.

Adkison, M.D., Peterman, R.M., 2000. Predictability of Bristol Bay, Alaska, sockeye salmon returns one to four years in the future. N. Am. J. Fish. Manage. 20: 69-80.

Adkison, M.D., Cunningham, C.J., 2015. The effects of salmon abundance and run timing on the performance of management by emergency order. Can. J. Fish. Aquat. Sci. 72: 1518-1526.

Anderson, J.J., Beer, W.N., 2009. Oceanic, riverine, and genetic influences on spring Chinook salmon migration timing. Ecol. App. 19: 1989-2003.

Bédard, M., 2007. Weak convergence of Metropolis algorithms for non-i.i.d. target distributions. The Ann. App. Probab.. 17: 1222-1244.

Bolker, B.M., 2008. Ecological Models and Data in R. Princeton University Press, 396 pp.

Brooks, S.P., Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. J. Comp. Gr. Stat. 7: 434-455.

Bue, D.G., Lipka, C.G., 2016. Characterization of the 2011 salmon run in the Kuksokwim River based on the test fishery at Bethel. Alaska Department of Fish and Game, Fishery Data Series No. 16-05, Anchorage, AK. Available at: http://www.adfg.alaska.gov/FedAidPDFs/FDS16-05.pdf [last accessed 4.12.2018].

Bue, B.G., Schaberg, K.L., Liller, Z.W., Molyneaux, D.B., 2012. Estimates of the historic run and escapement for the Chinook salmon stock returning to the Kuskokwim River, 1976-2011. Alaska Department of Fish and Game, Fishery Data Series No. 12-49, Anchorage, AK. Available at: http://www.adfg.alaska.gov/FedAidPDFs/FDS12-49.pdf [last accessed 4.12.2018].

Catalano, M.J., Jones, M.L., 2014. A simulation-based evaluation of in-season management tactics for anadromous fisheries: Accounting for risk in the Yukon River fall chum salmon fishery. N. Am. J. Fish. Manage. 34:1227-1241.

Chib, S., Greenberg, E., 1995. Understanding the Metropolis-Hastings Algorithm. Am. Stat. 49: 327-335.

Costello, C.J., Adams, R.M., Polasky, S., 1998. The value of El Niño forecasts in the management of salmon: a stochastic dynamic assessment. Am. J. of Agri. Econ. 80: 765-777.

Flynn, L., Hilborn, R., 2004. Test fishery indices for sockeye salmon (*Oncorhynchus nerka*) as affected by age composition and environmental variables. Can. J. Fish. Aquat. Sci. 61: 80-92.

Fried, S.M., Hilborn, R. 1988. Inseason forecasting of Bristol Bay, Alaska, sockeye salmon (*Oncorhynchus nerka*) abundance using Bayesian probability theory. Can. J. Fish. Aquat. Sci. 45: 850-855.

Haeseker, S.L., Peterman, R.M., Zhenming, S., Wood, C.C., 2005. Retrospective evaluation of preseason forecasting models for pink salmon. N. Am. J. Fish. Manage. 25: 897-918.

Haeseker, S.L., Peterman, R.M., Zhenming, S., Wood, C.C., 2008. Retrospective evaluation of preseason forecasting models for sockeye and chum salmon. N. Am. J. Fish. Manage. 28: 12-29.

Hamazaki, T. 2008. “When people argue about fish, the fish disappear.” Fish. 33: 495-501.

Hamazaki, T. 2011. Reconstruction of subsistence salmon harvests in the Kuskokwim Area, 1990-2009. Alaska Department of Fish and Game, Fishery Manuscript Series No. 11-09, Anchorage, AK. Available at: http://www.adfg.alaska.gov/FedAidPDFs/FMS11-09.pdf [last accessed 4.12.2018].

Hyun, S.Y., Hilborn, R., Anderson, J.J., Ernst, B. A statistical model for in-season forecasts of sockeye salmon (*Oncorhynchus nerka*) returns to the Bristol Bay districts of Alaska. Canadian Journal of Fisheries and Aquatic Sciences. 1665-1680.

Keefer, M.L., Peery, C.A., Caudill, C.C., 2008. Migration timing of Columbia River spring Chinook salmon: effects of temperature, river discharge, and ocean environment. Trans. Am. Fish. Soc. 137: 1120-1133.

Murphy, J.M., Howard, K.G., Gann, J.C., Cieciel, K.C., Templin, W.D., Gutherie, C.M., 2017. Juvenile Chinook salmon abundance in the northern Bering Sea: Implications for future returns and fisheries in the Yukon River. Deep Sea Res. Pt. II: Top. Stud. Oceanogr. 135: 156-167.

Mundy, P.R., Evenson, D.F., 2011. Environmental controls of phenology of high-latitude Chinook salmon populations of the Yukon River, North American, with application to fishery management. ICES J. Mar. Sci. 68: 1155-1164.

Peterman, R.M., 1982. Model of salmon age structure and its use in preseason forecasting and studies of marine survival. Can. J. Fish. Aquat. Sci. 39: 1444-1452.

Prager, M.J., Porch, C.E., Shertzer, K.W., Caddy, J.F., 2003. Targets and limits for management of fisheries: a simple probability-based approach. N. Am. J. Fish. Manage. 23:349-361.

R Core Team, 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Raftery, A.E., Lewis, S.M., 1992. How many iterations in the Gibbs sampler? J. Bernardo et al. (Eds.), Bayesian Statistics, University Press Oxford, pp. 765-776.

Shelden, C.A., Hamazaki, T., Horne-Brine, M., and Roczicka, G., 2016. Subsistence salmon harvests in the Kuskokwim area, 2015. Alaska Department of Fish and Game, Fishery Data Series No. 16-55, Anchorage, AK. Available at: http://www.adfg.alaska.gov/FedAidPDFs/FDS16-55.pdf [last accessed 4.12.2018].

Shertzer, K.W., Prager, M.H., Williams, E.H., 2011. Probabilistic approaches to setting acceptable biological catch and annual catch targets for multiple years: reconciling methodology with national standards guidelines. Mar. Coast. Fish. 2: 451-458.

Smith, N.J., Liller, Z.W., 2018. 2017 Kuskokwim River Chinook salmon run reconstruction and 2018 forecast. Alaska Department of Fish and Game, Division of Commercial Fisheries, Regional Information Report 3A18-02, Anchorage, AK. Available at: http://www.adfg.alaska.gov/FedAidPDFs/RIR.3A.2018.02.pdf [last accessed 4.12.2018].

Staton, B.A., Coggins, L.G., 2016. In-season harvest and effort estimates for 2016 Kuskokwim River subsistence salmon fisheries during block openers. U.S. Department of Interior, Fish and Wildlife Service, Yukon Delta National Wildlife Refuge, Bethel, AK. Accessible at:

https://www.fws.gov/uploadedFiles/2016KuskokwimSubsistenceSalmonHarvest.pdf [last accessed 4.12.2018].

Staton, B.A., Coggins, L.G., 2017. In-season harvest and effort estimates for the 2017 Kuskokwim River subsistence salmon fisheries during block openers. U.S. Department of Interior, Fish and Wildlife Service, Yukon Delta National Wildlife Refuge, Bethel, AK. Accessible at: https://www.fws.gov/uploadedFiles/2017KuskokwimSubsistenceSalmonHarvest.pdf [last accessed 4.12.2018].

Staton, B.A., Catalano, M.J., Farmer, T.M., Abebe, A., Dobson, F.S., 2017. Development and evaluation of a migration timing forecast model for Kuskokwim River Chinook salmon. Fish. Res. 194: 9-21.

Walters, C.J., Buckingham, S. 1975. A control system for intraseason salmon management. International Institute for Applied Systems Analysis Working Paper. WP-75-028. 19 pp.

Winship, A.J., O'Farrell, M.R., Satterhwaite, W.H., Wells, B.K., Mohr, M.S. 2015. Expected future performance of salmon abundance forecast models with varying complexity. Can. J. Fish. Aquat. Sci. 72: 558-569.

Table 1. Estimated coverage of various regions in the prior, likelihood, and posterior distributions.

Information Source	June 10			June 24			July 8		
	50%	**80%**	**95%**	**50%**	**80%**	**95%**	**50%**	**80%**	**95%**
Prior	70	78	96	70	78	96	70	78	96
Likelihood (Fcst. Timing)	30	52	87	39	78	96	48	83	96
Likelihood (Hist. Timing)	39	74	100	61	83	96	48	78	96
Posterior (Fcst. Timing)	43	70	91	52	65	91	48	70	87
Posterior (Hist. Timing)	57	87	96	57	78	96	43	70	87

FIGURE CAPTIONS

Figure 1. (a) Estimated run size time series from 1976-2017 presented in Smith and Liller (2018), (b) time series of log scale multiplicative pre-season forecast errors ($\varepsilon_{F,t}$), and (c) distribution of the $\varepsilon_{F,t}$ values.

Figure 2. Fitted regression relationships between the run size vulnerable to sampling ($N_{vulnt,t}$) by the Bethel Test Fishery (BTF) and the end-of-season cumulative BTF catch-per-effort (EOS_t) as described by Eq. (4) and used to predict run size in-season using Eq. (5). Numbers represent the year observations were made. Grey years and lines represent the first catchability period and black years and lines represent the second catchability period, as described in the text.

Figure 3. (a) Mean absolute proportional error of median run estimates, (b) mean proportional error of median run estimates, and (c) standard deviation of log scale multiplicative errors [log(estimate/true)] throughout the season. Triangles represent errors from the likelihood distribution and circles represent errors from the posterior distribution. Filled symbols represent the method that does not use the run timing forecast and empty symbols represent the method that does. Grey lines are the errors from the pre-season run size forecast (i.e., the prior).

Figure 1.

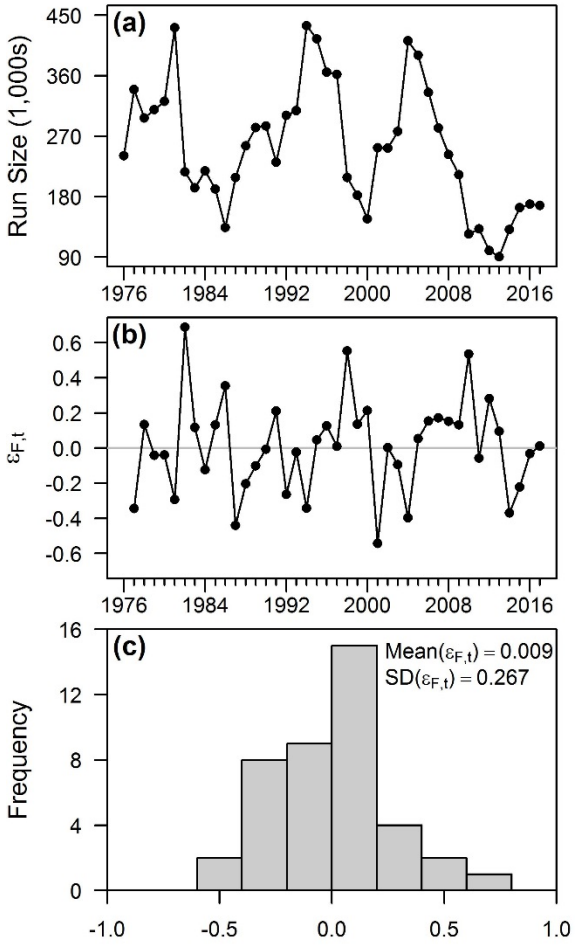

(a)
Run Size (1,000s)
450
360
270
180
90
1976
1984
1992
2000
2008
2016
(b)
$\varepsilon_{F,t}$
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
(c)
Mean($\varepsilon_{F,t}$) = 0.009
SD($\varepsilon_{F,t}$) = 0.267
Frequency
16
12
8
4
0
-1.0
-0.5
0.0
0.5
1.0

Figure 2.

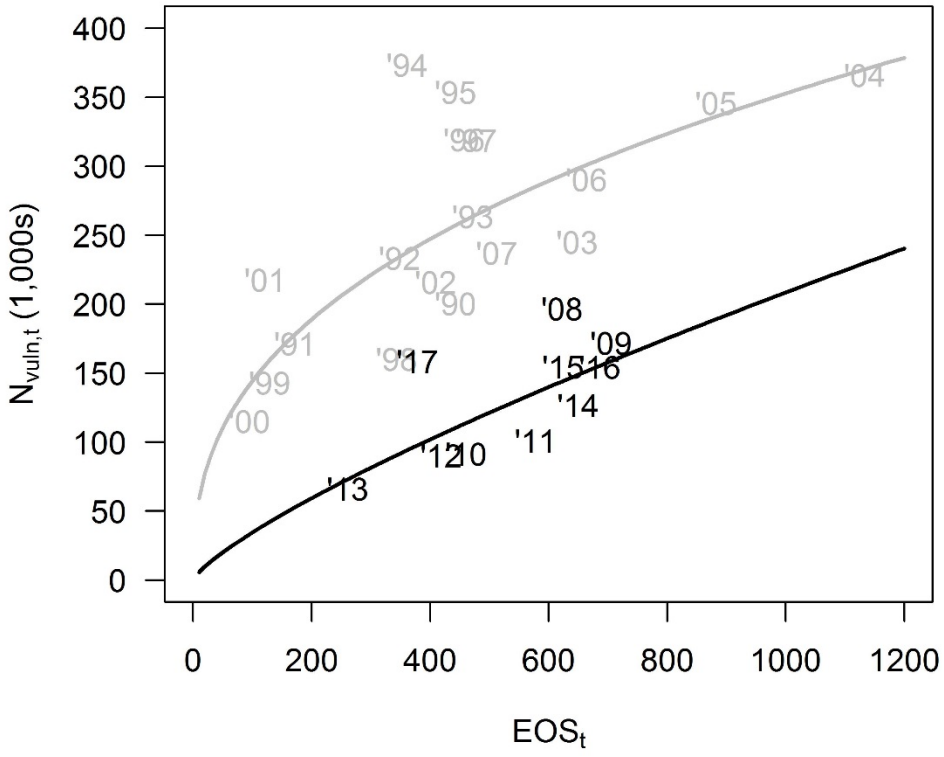
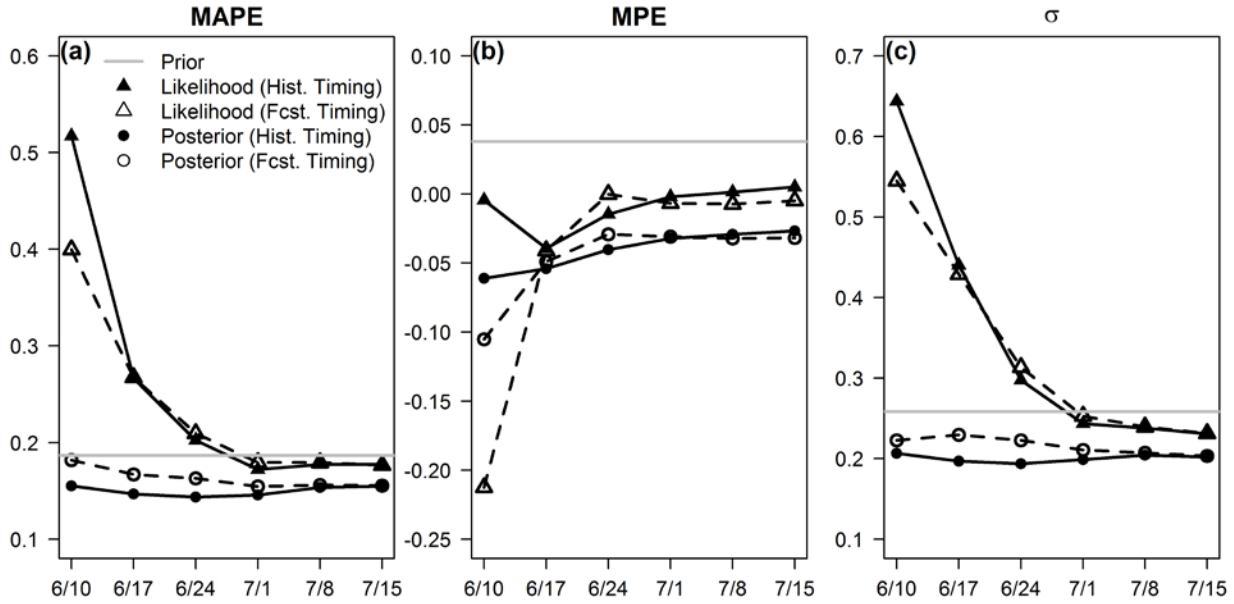

400
350
300
250
200
150
100
50
0
$N_{vuln,t}$ (1,000s)
0
200
400
600
800
1000
1200
EOS_t
'94
'95
'05
'04
'06
'93
'07
'03
'01
'92
'02
'90
'08
'91
'09
'99
'17
'15
'16
'14
00
'11
'13
'10

Figure 3.


MAPE
MPE
σ
(a)
(b)
(c)
Prior
Likelihood (Hist. Timing)
Likelihood (Fcst. Timing)
Posterior (Hist. Timing)
Posterior (Fcst. Timing)
0.6
0.5
0.4
0.3
0.2
0.1
0.10
0.05
0.00
-0.05
-0.10
-0.15
-0.20
-0.25
0.7
6/10
6/17
6/24
7/1
7/8
7/15