

A Bayesian hierarchical approach to integrating historical and in-season genetic data for real-time assessment of a mixed stock fishery

Lukas B. DeFilippo, Daniel E. Schindler, Kyle Shedd, and Kevin L. Schaberg

Abstract: With advances in molecular genetics, it is becoming increasingly feasible to conduct genetic stock identification (GSI) to inform management actions that occur within a fishing season. While applications of in-season GSI are becoming widespread, such programs seldom integrate data from previous years, underutilizing the full breadth of information available for real-time inference. In this study, we developed a Bayesian hierarchical model that integrates historical and in-season GSI data to estimate temporal changes in the composition of a mixed stock of sockeye salmon (*Oncorhynchus nerka*) returning to Alaska's Chignik watershed across the fishing season. Simulations showed that even after accounting for time constraints of transporting and analyzing genetic samples, a hierarchical approach can rapidly achieve accurate in-season stock allocation, outperforming alternative methods that rely solely on historical or in-season data by themselves. As the distribution and phenology of fish populations becomes more variable and difficult to predict under climate change, in-season management tools will likely be increasingly relied upon to protect biocomplexity while maximizing harvest opportunity in mixed stock fisheries.

Résumé : Les avancées en génétique moléculaire font en sorte qu'il est de plus en plus envisageable d'utiliser l'identification génétique au stock (IGS) pour éclairer le choix des mesures de gestion à utiliser durant une saison de pêche. Si les exemples d'application de l'IGS pour la saison en cours sont de plus en plus répandus, ces programmes intègrent rarement les données d'années précédentes, sous-utilisant donc le vaste éventail d'information disponible pour l'inférence en temps réel. Nous avons développé un modèle hiérarchique bayésien qui intègre des données d'IGS passées et pour la saison en cours pour estimer les changements dans le temps de la composition d'un stock mélangé de saumons sockeyes (*Oncorhynchus nerka*) retournant dans le bassin versant de Chignik, en Alaska, au fil de la saison de pêche. Des simulations montrent que, même en tenant compte de contraintes temporelles associées au transport et à l'analyse des échantillons génétiques, une approche hiérarchique peut permettre de réaliser rapidement et avec exactitude l'affectation au stock au fil de la saison, donnant de meilleurs résultats que d'autres méthodes qui reposent exclusivement sur des données passées ou pour la saison en cours. La répartition et la phénologie des populations de poissons devenant plus variables et difficiles à prédire dans un contexte de changements climatiques, il y aura vraisemblablement un recours croissant à des outils de gestion reposant sur des données pour la saison en cours pour maintenir la biocomplexité tout en maximisant les possibilités de prises dans les pêches de stocks mélangés. [Traduit par la Rédaction]

Introduction

Managing harvest that simultaneously targets multiple populations is a classic and common challenge in fisheries. Failing to account for such demographic heterogeneity can lead to biased estimates of biological reference points, nonstationary stock-recruit relationships, overfishing of less productive stocks, and may have evolutionary consequences deleterious to long-term yield and sustainability (Cadrin et al. 2018; Hard et al. 2008; Hilborn 1985; Hilborn and Walters 1992). Several methods of stock identification have been used to infer the composition of mixed stock fisheries, sometimes relying on morphological or meristic variation among populations (Cadrin 2000; Cadrin et al. 2013; Marshall et al. 1987). In particular, genetic stock identification (GSI) methods have become common, including the use of single nucleotide polymorphisms (SNPs) that are easily surveyed, powerful for discriminating stock structure, and amenable to high throughput and automation (Hauser and Seeb 2008). With continuing advances in molecular genetics, it is becoming increasingly feasible

to conduct GSI in near real time to inform management decisions that occur within the fishing season (Bue et al. 2009; Seeb et al. 2000; Shaklee et al. 1999; Smith et al. 2005). While in-season applications of GSI are becoming common, such programs seldom make use of all available data, ignoring information from previous years that may be leveraged to improve real-time inference on compositional dynamics.

Many Pacific salmon (*Oncorhynchus* spp.) populations in North America are managed under a fixed escapement policy wherein harvest is permitted only after sufficient individuals are projected to have escaped the fishery to achieve a certain target abundance, such as the number of spawners that produces maximum sustainable yield (S_{MSY}) (Hilborn and Walters 1992). Commercial fishing seasons for Pacific salmon are usually limited to the duration of the spawning migration, occurring over 1–3 months in most cases. To maintain life history variation in run timing and intrapopulation diversity, managers often decompose the total escapement goal into a schedule of interim targets across the season. The

Received 20 March 2020. Accepted 4 July 2020.

L.B. DeFilippo and D.E. Schindler. University of Washington School of Aquatic and Fishery Sciences, 1122 NE Boat Street, Seattle, WA 98105, USA.

K. Shedd. Alaska Department of Fish and Game, Commercial Fisheries Division, Gene Conservation Laboratory, 333 Raspberry Road, Anchorage, AK 99518, USA.

K.L. Schaberg. Alaska Department of Fish and Game, Commercial Fisheries Division, 351 Research Court, Kodiak, AK 99615, USA.

Corresponding author: Lukas B. DeFilippo (email: ldefilip@uw.edu).

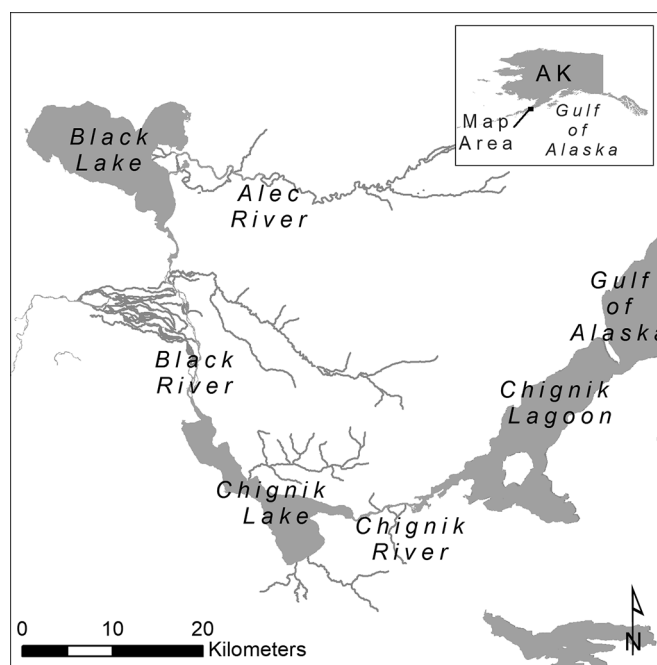
Copyright remains with the author(s) or their institution(s). Permission for reuse (free in most cases) can be obtained from copyright.com.

inferred escapement to a stock can then be compared with its interim escapement targets across the season to determine whether fishing should be allowed at a given point in time. However, during the marine and early freshwater portions of their spawning migration, salmon from multiple stocks may be present in the same location, and an individual's population of origin is not necessarily apparent. In these circumstances, reliable estimation of the escapement to each population can be crucial to regulating harvest on the stock mixture (but see Connors et al. 2020; Freshwater et al. 2020).

Molecular genetics has played an increasingly important role in contemporary fisheries ecology and assessment, identifying cryptic stock structure and management units, informing stock–recruit dynamics, and improving estimates of population viability (Hauser and Seeb 2008; Hauser and Carvalho 2008; Satterthwaite et al. 2014; Shaklee et al. 1999; Spies et al. 2015, 2018; Spies and Punt 2015). In the management of Pacific salmon, genetic information from mixed stock fisheries is often used in run reconstruction and stock–recruit models that form the basis for forecasting future returns and establishing management reference points (Cunningham et al. 2017). Additionally, genetic tools are being used increasingly to inform management decisions that occur within the fishing season (Shaklee et al. 1999; Smith et al. 2005). For example, Dann et al. (2013) used in-season genetic data to detect migratory trends in the stock composition of Bristol Bay sockeye salmon (*Oncorhynchus nerka*) prior to arrival in the fishery, revealing spatial and temporal patterns that were unanticipated from preseason forecasts. With this information, managers were able to re-allocate fishing effort in real time to avoid overfishing less abundant populations while maximizing harvest opportunity on robust stocks. Such successful applications of in-season GSI are encouraging and reflect the broad utility of this approach for addressing the challenges of real-time fisheries management (Beacham et al. 2004; Bednarski et al. 2017; Bue et al. 2009; Seeb et al. 2000; Templin and DeCovich 2009). However, as in-season GSI becomes more common, developing statistical frameworks to make the most efficient use of these data will be essential to maximizing their utility for managers.

Here, we explore a novel application of in-season GSI for sockeye salmon from Alaska's Chignik watershed that incorporates information from other years to improve in-season inference on the stock composition. Fisheries in this region exploit two separate stocks of sockeye salmon that spawn in Black Lake and Chignik Lake. The two stocks differ in their run timing, with Black Lake fish arriving first and the Chignik Lake run occurring later in the season (Dahlberg 1968). However, there is sufficient temporal overlap between the two runs such that both stocks are typically present in appreciable portions for much of the fishing season. Because the two stocks are reproductively isolated (Creelman et al. 2011; Witteveen and Botz 2003), they are managed with separate escapement goals, both of which are factored into real-time harvest control decisions (Wilburn 2019; Wilburn and Stumpf 2018). Throughout the fishing season, managers monitor the inferred escapement to each stock relative to its interim escapement target for that point in time and decide whether or not fishing should proceed. However, escaped fish are counted at a weir on the Chignik River, which both Black Lake and Chignik Lake fish pass through on their spawning migration. As such, methods to distinguish Black Lake fish from Chignik Lake fish in the aggregate escapement are needed to track each stock's escapement status and inform decision-making within the fishing season. Within the past decade, the Alaska Department of Fish and Game (ADF&G) has implemented a program for GSI of Chignik sockeye salmon to estimate the stock composition of the escapement and explore the potential applications of this information for real-time management (Stumpf 2017). In this study, we develop a Bayesian hierarchical modelling framework that integrates historical and in-season GSI data to efficiently estimate the stock

Fig. 1. Map of the Chignik watershed located on the Alaska Peninsula, with associated lakes, rivers, and other relevant locations indicated. Created in Esri ArcMap 10.6.1 using the Alaska Department of Natural Resources 2019 Simplified Alaska Coastline and US Geological Survey (2019) National Hydrography Dataset.



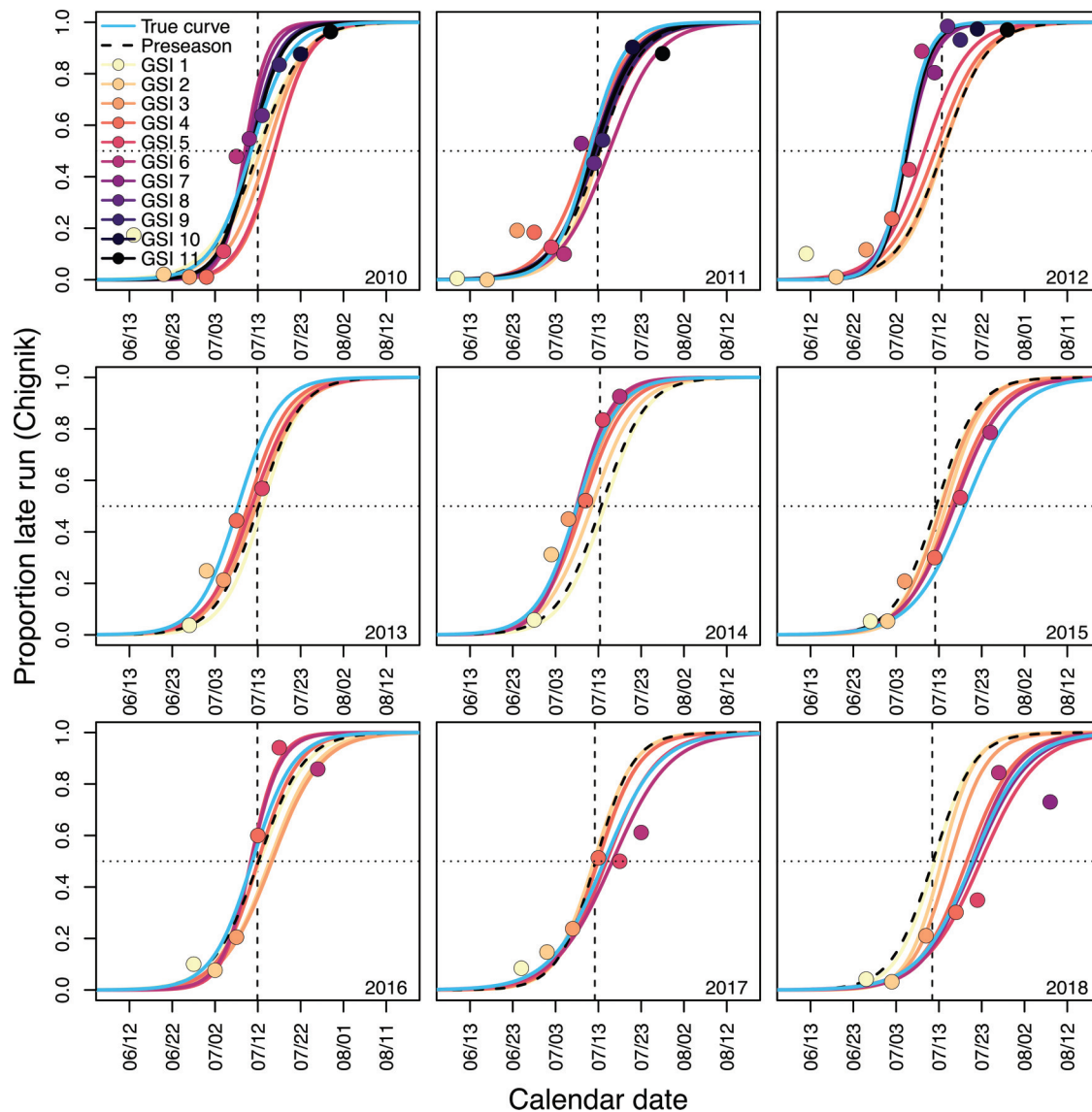
composition within the fishing season. An expectation for the timing and speed of the transition between stocks is developed from historical data and is modified sequentially as new data are accumulated, thereby adjusting for annual deviations. Additionally, we conducted simulation testing to examine the value of this approach for in-season assessment and how its performance varies with alternative implementation scenarios.

Methods

Located on the south side of the Alaska Peninsula, the Chignik watershed is composed of two lakes (Fig. 1), each of which supports genetically distinct populations of sockeye salmon with different migratory timings (Creelman et al. 2011). Black Lake produces sockeye salmon that typically migrate into the watershed in June to July and spawn in late August and early September. A second sockeye salmon run migrates into the watershed in late June through September to spawn after early September in Chignik Lake and its tributaries. The Black and Chignik lake stocks are reproductively isolated (but share rearing habitat) and are managed as two separate runs with their own escapement goals (Schaberg et al. 2019; Wilburn 2019). Both Black Lake and Chignik Lake sockeye salmon are harvested together in the five major fishing districts that target these populations (Dann et al. 2012a). Given the mixed stock nature of the fisheries, reliable estimates of the proportion of early (Black Lake) versus late (Chignik Lake) run fish present in the total escapement at any given point in time are needed to determine whether each stock's interim escapement targets are being met, providing one of several sources of information that can be used in harvest control decisions. Management is then prosecuted through a system of fishery openings and closures that typically span 1–3 days at a time.

To inform management objectives, the ADF&G operates a weir on the Chignik River where enumeration and sampling of the escapement occur. Fish passing through the weir are counted daily, and up to 240 individuals are sampled for age, sex, and length each week (Wilburn and Stumpf 2018). Prior to genetic

Fig. 4. In-season evolution of the estimated logistic transition function under the hierarchical model for 2010–2018. Preseason curves (dashed) shown in each panel are conditioned on data collected from the other 8 years of the study, and each subsequent curve represents the addition of an in-season GSI sample to the model. Coloured circles in each panel show the simulated compositional data points for each sampling stratum. The logistic curve matching the colour of each circle shows the updated stock transition function fit with all GSI data collected up to and including that point in the season, as well as the data from other years in the study. The light blue curve in each panel depicts the true transition curve for that year from which the simulated compositional data were generated. The horizontal dashed line denotes the 50:50 point in the run, and the vertical dashed line shows the expected preseason 50:50 transition date based on analyzing all other years of data (τ_y^{pred}). Note that different numbers of GSI samples were taken and analyzed in each year of the study. [Colour online.]



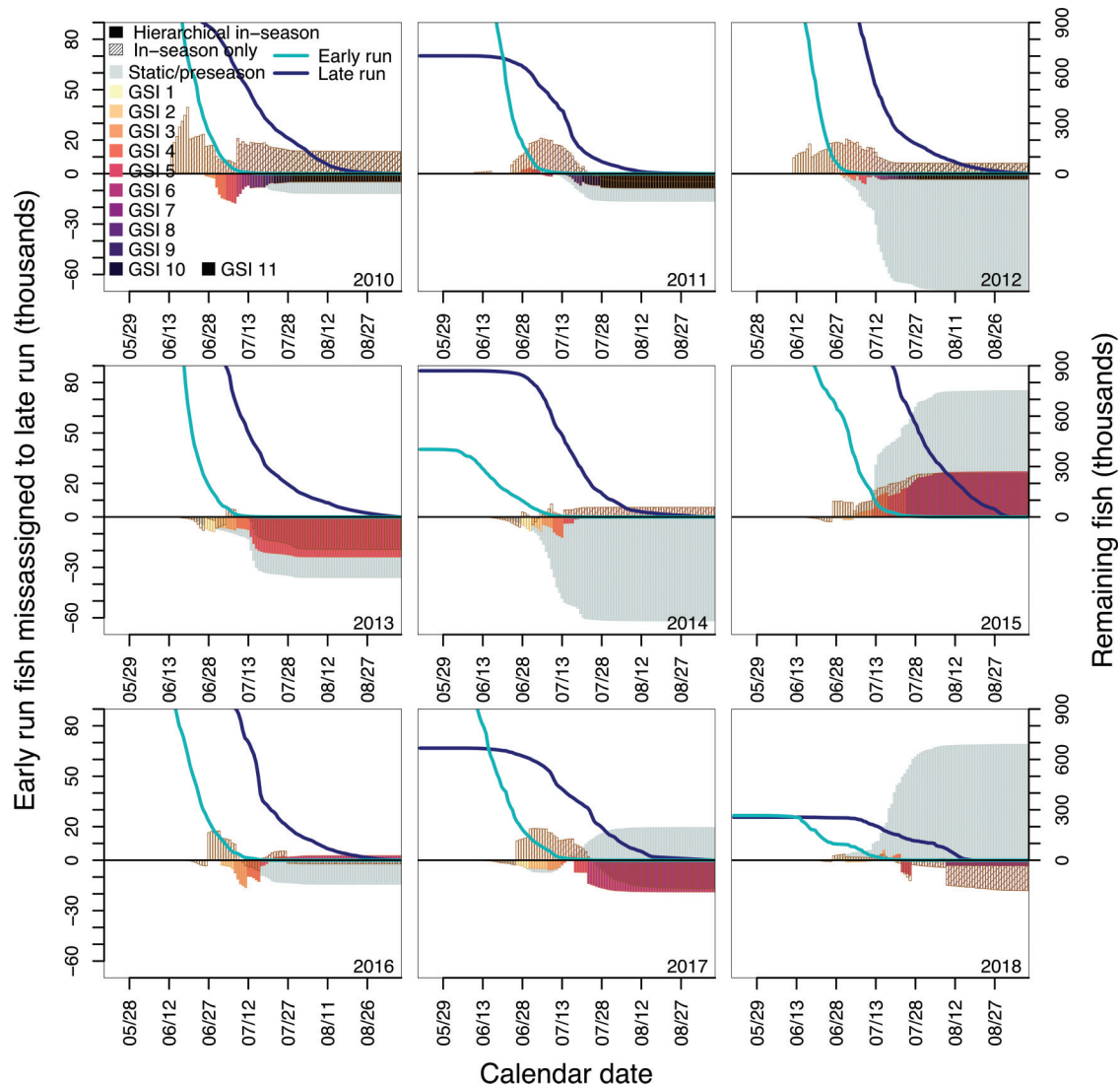
transition during the first two to six sampling strata resulted in larger early season assignment errors (Figs. S3¹, 5, 6).

The results of the analyses described thus far assume that there is no delay between when genetics samples are collected and when the GSI data are available for updating the hierarchical or in-season only transition functions within the season. However, there are logistical challenges in collecting samples, transporting them to Anchorage for analysis in the GCL, and interpreting the data. Thus, we simulated a range of time lags between sample collection and when GSI data are available for updating the transition curve and assignment of the escapement. For comparison with the results presented above, we executed simulations with lags of 3, 5, and 20 days between sample collection and the application of the GSI-updated transition curve to the escapement data. A 3-day lag between sample collection and updating the transition

curve had a negligible effect on the assignment of fish to the early or late runs over the season for both the hierarchical and in-season only models (Figs. S4, S5¹). At a 5-day lag, there were moderate differences in the run allocations from those observed in the 0-day lag scenario (Figs. S6, S7¹). Not surprisingly, a 20-day lag generated substantial errors in stock assignments for both the hierarchical and in-season only models, effectively negating any benefit to the in-season GSI data (Figs. S8, S9¹).

Across all years, the static approach resulted in a total of over 26.6 million cumulative in-season misassignments, compared with ~9.7 million for the in-season GSI only approach and ~7.4 million for the hierarchical in-season GSI approach. While the in-season only and hierarchical in-season GSI approaches typically reached similar assignment accuracy by the end of the season, the in-season only model exhibited larger early season assignment errors, leading to

Fig. 6. Errors in allocation of sockeye salmon to the early and late runs over the course of the migration season into the Chignik River from 2010 to 2018 under different stock assignment methods. In each panel, the histograms represent the number of early run fish incorrectly assigned to the late run (left y axis). Negative values reflect late run fish misassigned to the early run. Grey histograms show misassignments under the static hierarchical approach (scenario 1), while the solid, coloured histograms show the number of misassigned fish using the hierarchical GSI model (scenario 2), and the bars with diagonal coloured lines show the number of misassigned fish using the in-season GSI only model (scenario 3). For the coloured histograms (solid or with diagonal lines), the different colours correspond to the number of in-season genetic samples that were used at that point in the run to inform the transition curve for both the hierarchical GSI (solid filled) and in-season GSI only (coloured diagonal lines) models. The teal and dark blue lines in each panel show the number of fish belonging to the early run (teal) or late run (dark blue) that have yet to enter the fishery or escape to the river across the season (right y axis). These curves were generated retrospectively using the true transition curve from which the data were generated and the end of season catch and escapement counts. These simulations assume no lag between sampling and when data are available for updating the transition curve. [Colour online.]

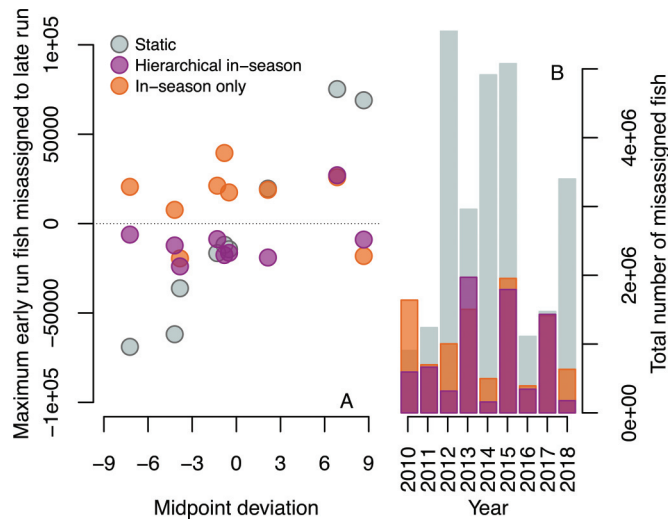


from the early to late run to be particularly abrupt. For instance, in 2012 the midpoint of the transition occurred 1 week earlier than the among-year average and reached 95% late run only 8 days afterwards. As a result, in simulations for 2012, by the time the static transition model indicated that the escapement had reached 50% late run, the stock composition was actually ~95% late run.

For in-season GSI to be useful for management, it must accurately identify the transition function for a given year early enough that there is still opportunity to alter management trajectories while a substantial portion of the run remains subject to fishery management. Our analyses demonstrate that for the 9 years included in this study, GSI would have provided little opportunity for altering in-season management on the early run compared with a static approach. In the case of the in-season only

GSI model, early season assignment errors were often larger than under the static approach, suggesting that management of the early run could be impaired by relying solely on in-season GSI data. While the hierarchical model generally exhibited smaller early season assignment errors compared with the in-season only approach, it typically did not achieve substantially better accuracy than the static approach until much of the early run had arrived at the watershed (Fig. 6). Conversely, by the time the hierarchical model typically achieved improved performance over the static approach, the majority of late run fish had yet to arrive at the watershed, even with realistic analytical time lags (~3 days), thus informing the manager that the transition had occurred and allowing management to be adapted from the middle of the fishing season onwards (Figs. S4, S5¹). Not only was the in-season only

Fig. 7. Transition timing and assignment accuracy under each stock assignment method. In panel A, the x axis shows the median posterior estimates of the midpoint date (τ) for each year centered with respect to the among-year average (μ). The y axis shows the maximum number of early run fish that were misassigned to the late run fish by the static (grey), hierarchical GSI (purple), and in-season only (orange) approaches for each year. Negative values reflect late run fish misassigned to the early run. In panel B, bars show the cumulative total number of fish that were misassigned across the duration of each year's run under the static (grey), hierarchical GSI (purple), and in-season only (orange) approaches. [Colour online.]



model slower to achieve accurate stock allocation than the hierarchical model, but in many years (e.g., 2010, 2012, 2014, 2018) it exhibited larger assignment errors through the end of the season as well. As such, the hierarchical model likely offers some advantages for postseason stock allocation and run reconstruction, in addition to in-season inference. However, it should be noted that in some years (e.g., 2010, 2016), the early season assignment errors produced by the hierarchical GSI model exceeded those from the static transition model, which could potentially have negative consequences for management. Future analyses should explore explicit weighting schedules for treating the preseason and in-season GSI data in the hierarchical model that minimize these errors and optimize overall assignment accuracy. Additionally, other types of data collected in-season that are informative to the stock composition (e.g., age composition) could be included alongside the GSI data in an integrated stock transition model that could improve real-time inference.

The results presented here can be viewed as the first step in a management strategy evaluation (MSE; Punt et al. 2016) that uses closed-loop simulations to understand how alternative in-season stock assignment methods affect long-term production and yield given (i) uncertainties in the data and implementation of management procedures, (ii) variation in the timing and relative strength of runs, and (iii) feedbacks among population dynamics, management decision-making, and harvester behavior. Importantly, our simulation analysis did not model management actions in response to in-season stock assignments, nor did it consider key uncertainties in their implementation (Dorner et al. 2009; Holt and Peterman 2006, 2008) or model specification (Punt et al. 2016). A previous MSE for sockeye salmon has shown that when outcome error is considered, better information may not necessarily improve management performance (Dorner et al. 2009). Importantly, the realized effects of using any of the stock assignment approaches considered here will ultimately be determined by the decision-making process used by managers to update their in-season assessments and adjust harvest control, which we did not

evaluate. Future analyses might also explore the performance of in-season GSI across a varying range of assignment accuracies associated with different SNP design protocols in a genetically based MSE (Anderson et al. 2008; Waples et al. 2008).

Despite marked potential for improving the assessment and conservation of living aquatic resources, genetic data can be challenging to integrate formally into fisheries assessment and management (reviewed in Waples et al. 2008). For Pacific salmon, the logistical constraints of real-time harvest control pose arguably one of the greatest challenges to full utilization of genetic information in management (Cunningham et al. 2019; Dorner et al. 2009). However, our analysis shows that a hierarchical modelling approach can improve the performance of in-season GSI for inferring compositional dynamics in real time. As such, similar approaches may be useful in other scenarios to facilitate the incorporation of in-season data into real-time assessment and decision-making. Preseason forecast models to predict salmon returns remain highly uncertain, and MSE has shown that even if forecast accuracy can be improved, there may be no realized benefit to the management of the stock due to outcome error (Dorner et al. 2009). Conversely, in-season inference offers managers the opportunity to adaptively correct for preseason uncertainties in real time, potentially reducing the magnitude of implementation errors (Dann et al. 2013; Walters 1989). As the timing and strength of salmon runs becomes more variable and difficult to predict due to climate change (Crozier et al. 2008, 2011), in-season stock identification methods may become increasingly important to ensuring the sustainable management of exploited populations. While future analyses that consider the full suite of relevant uncertainties are needed to comprehensively assess the value of alternative in-season stock assignment methods, our results indicate that a hierarchical approach may be a productive avenue for improving management outcomes. With growing recognition of the importance of population diversity for maintaining the resilience of fish stocks to environmental change (Connors et al. 2020; Schindler et al. 2010), in-season management tools are likely to become increasingly relied upon to conserve biocomplexity while simultaneously maximizing harvest opportunities.

Acknowledgements

This research was funded by the Chignik Regional Aquaculture Association, and we thank Chuck McCallum, Bruce Barrett, George Anderson, and Matt Siemion for their guidance and feedback on the analyses. We also thank colleagues of the Alaska Department of Fish & Game, Westward Region, for providing data and feedback on the development and interpretation of the analyses and results presented here. In particular, Birch Foster, Dawn Wilburn, and Heather Finkle were instrumental in this regard. Much of the genetic baseline was collected and analyzed by Elizabeth Creelman Fox, with funding from the Alaska Sustainable Salmon Fund under award 63-5096 from the National Oceanic and Atmospheric Administration, administered by ADF&G. We also acknowledge the efforts of the ADF&G Gene Conservation Lab, including Tyler Dann and Chris Habicht, for generating the GSI data used in this study. We also thank André E. Punt and Barry A. Berejikian for helpful comments on earlier drafts and David French for creation of the map for Fig. 1.

References

- Alaska Department of Natural Resources. 2019. Simplified Alaska Coastline [online]. Alaska Department of Natural Resources, Information Resource Management. Available from <http://asgdc.alaska.gov#67>.
- Anderson, E.C., Waples, R.S., and Kalinowski, S.T. 2008. An improved method for predicting the accuracy of genetic stock identification. *Can. J. Fish. Aquat. Sci.* 65(7): 1475–1486. doi:10.1139/F08-049.
- Beacham, T.D., Lapointe, M., Candy, J.R., Miller, K.M., and Withler, R.E. 2004. DNA in action: rapid application of DNA variation to sockeye salmon fisheries management. *Conserv. Genet.* 5: 411–416. doi:10.1023/B:COGE.0000031140.41379.73.