



Application of machine learning and acoustic predation tags to classify migration fate of Atlantic salmon smolts

Daniela V. Notte¹ · Robert J. Lennox² · David C. Hardie³ · Glenn T. Crossin¹

Received: 15 May 2021 / Accepted: 15 February 2022 / Published online: 4 March 2022
© Crown 2022

Abstract

Mortality and predation of tagged fishes present a serious challenge to interpreting results of acoustic telemetry studies. There is a need for standardized methods to identify predated individuals and reduce the impacts of “predation bias” on results and conclusions. Here, we use emerging approaches in machine learning and acoustic tag technology to classify out-migrating Atlantic salmon (*Salmo salar*) smolts into different fate categories. We compared three methods of fate classification: predation tag pH sensors and detection data, unsupervised k-means clustering, and supervised random forest combined with tag pH sensor data. Random forest models increased predation estimates by 9–32% compared to relying solely on pH sensor data, while clustering reduced estimates by 3.5–30%. The greatest changes in fate class estimates were seen in years with large class imbalance (one or more fate classes underrepresented compared to the others) or low model accuracy. Both supervised and unsupervised approaches were able to classify smolt fate; however, in-sample model accuracy improved when using tag sensor data to train models, emphasizing the value of incorporating such sensors when studying small fish. Sensor data may not be sufficient to identify predation in isolation due to Type I and Type II error in predation sensor triggering. Combining sensor data with machine learning approaches should be standard practice to more accurately classify fate of tagged fish.

Keywords Telemetry · Random forest · Clustering · Population management · *Salmo salar*

Introduction

A major assumption of animal telemetry studies is that the data collected from tags represent the natural movements of a live individual of the study species, and not an expelled tag, a mortality, or the movements of a predator (Gibson et al. 2015; Klinard et al. 2019). However, the violation of this assumption is often not addressed, despite the negative impact it can have on study results, population management, and conservation efforts (Klinard and Matley 2020). In the aquatic environment, predation of tagged fish presents a

serious challenge to telemetry studies, because acoustic tags can continue to transmit through the body of the predator for as long as 6 months (Klinard et al. 2019, 2021). Therefore, failure to identify predation events of tagged individuals introduces a “predation bias”, such that survival rates are inflated, individual movement patterns (e.g. depth use, rate of travel) are calculated based on both prey and predator movement, and the locations of areas of high mortality are skewed (Gibson et al. 2015; Daniels et al. 2019; Klinard et al. 2019). Even when predation events are identified, it is often on a subjective basis (Perry et al. 2010; Buchanan et al. 2013), dependent on predator and prey behaviour being significantly different and distinguishable (Romine et al. 2014; Gibson et al. 2015; Moxam et al. 2019), and difficult to pinpoint the time and location of mortality, hindering attempts to remove detections of consumed fish (Gibson et al. 2015; Daniels et al. 2018). Predation may therefore reduce confidence in the conclusions of animal telemetry studies (Hal-fyard et al. 2017).

Movement ecologists recognize the negative impact of predation on the interpretation of acoustic telemetry data and have been developing methods for identification of predation

Communicated by Yannis Papastamatiou.

✉ Daniela V. Notte
d.notte@dal.ca

¹ Department of Biology, Dalhousie University, 1355 Oxford Street, Halifax, NS B4H 4R2, Canada

² Laboratory for Freshwater Ecology and Inland Fisheries, NORCE Norwegian Research Centre, Bergen, Norway

³ Fisheries and Oceans Canada, 1 Challenger Drive, Dartmouth, NS B2Y 4A2, Canada

to reduce its bias on study results. Early approaches to classify predation were to gather contextual information from temperature sensors to detect predation by endothermic predators (adult salmonids preyed by seals identified by an increase in temperature; Bendall and Moore 2008) or depth sensors for identification through uncharacteristic swimming patterns (predatory Atlantic cod and saithe swim to significantly greater depths than juvenile salmon; Thorstad et al. 2012). Later, analytical methods emerged that were able to detect predation events of tagged fish using supervised or unsupervised machine learning approaches that identified anomalous movement patterns in the data suggestive of preyed individuals. Researchers have previously tagged both prey, juvenile Atlantic salmon (*Salmo salar*), and predator, striped bass (*Morone saxatilis*), and used either a cluster analysis (Gibson et al. 2015) or random forest (Daniels et al. 2018) approach to identify preyed salmon based on movement metrics. Klinard et al. (2021) used random forest to identify the species of predator responsible for consumption of each individual bloater (*Coregonus hoyi*) by tagging both prey and multiple predatory species. However, in some studies, it may not be logistically feasible to tag non-target species. Moxham et al. (2019) were able to estimate predation events on tagged bonefish using an unsupervised approach that did not include data from predator movements using clustering methods to differentiate habitat space use and speed metrics of consumed bonefish from those that survived following catch-and-release. Now, recent developments in acoustic tag technology have led to the ability to detect predation events via changes in pH that trigger a change in the unique ID of the tag, referred to as predation tags (Halfyard et al. 2017; Lennox et al. 2021). Predation tags have been used in studies on yellow perch (*Perca flavescens*) in the Detroit River (Weinz et al. 2020), bloater in the Great Lakes (Kilnard et al. 2019, 2021), and Atlantic salmon in the Miramichi River, NB (Daniels et al. 2019).

The random forest and cluster analysis methods described above are classification tools in the machine learning family, a branch of statistics that is used to predict outcomes from training data to in-sample or out-of-sample data (Thessen, 2016). In supervised machine learning (e.g., random forests), models are trained on data sets with independent and dependent variables, the model learns how the variables are related, and the model is then able to predict the dependent variable on future data sets where only the independent variables are provided (Thessen 2016). Unsupervised methods (e.g., cluster analysis) find patterns among the independent variables to organize data based on underlying similarities in the data ascertained by the algorithm (Olden 2008). Machine learning approaches are becoming increasingly used in ecology because they are able to model data that are non-linear, contain interacting variables, and have missing values, all of which are common in ecological data sets (Olden 2008;

Thessen 2016). Applications of machine learning in ecology include habitat modelling and species distribution (Cutler et al. 2007; Brownscombe et al. 2019), species identification (Tabak et al. 2018), monitoring biodiversity (Cordier et al. 2017), and predicting the conservation status of species (Bland et al. 2014). The ability to make accurate ecological predictions is vital for informed management and decision making (Clark et al. 2001; Olden 2008; Coreau et al. 2009). Additionally, mortality is a vital metric for fisheries and resource management and there are increasing efforts towards improving the accuracy and precision of mortality estimates using new methodologies and technologies.

Refinements are still needed to operationalize predation classification for telemetry users. Ideally, a combination of both behavioural and sensor-based methods for determining predation events would do much to increase confidence in fate classification of tagged fish, as tag sensors may sometimes fail and predator behaviours may not always be significantly different than prey behaviour (Weinz et al. 2020; Lennox et al. 2021). Juvenile Atlantic salmon (smolts) out-migrating from the Stewiacke River, Nova Scotia present an ideal opportunity to apply this combined approach. Natural mortality of smolts during seaward migration is high, with predation often accounting for the majority of mortalities and challenging management efforts, especially those that rely on fish tracking data (LaCroix 2008; Thorstad et al. 2011, 2012). The Stewiacke River is dominated by striped bass, a common predator of Atlantic salmon smolts. Salmon smolt behaviour during migration consists largely of short, linear movements directed downstream with some reversals during out-migration, especially when first entering the estuary, likely as a response to osmotic stress (Halfyard et al. 2012, 2013) or due to the impact of incoming tides (Beland et al. 2001). Except for these occasional path reversals, these movements are distinct from the extensive and tortuous movements with frequent reversals in up- and downstream movement exhibited by striped bass (Romine et al. 2014; Gibson et al. 2015; Daniels et al. 2018). These differences form the basis for the behavioural metrics with which we can distinguish live and preyed smolts, conducive to supervised machine learning approaches to identifying predation based on movements. However, these machine learning methods have not been adequately tested against objective empirical data with which models can be evaluated and best practices developed for a workflow to identify predation of tagged fish. The introduction of predation sensor tags has revealed limitations and opportunities for refinement of available approaches to identify predation of acoustically tagged fish. In this study, we developed and tested methods for classification of predation using either predation tag sensors, behavioural metrics (unsupervised machine learning), or a combination of the two (supervised machine learning) to determine the best approach for fate classification and

the value of using predation tags. We apply these tagging and analytical tools to compare rates of estimated Atlantic salmon smolt migration survival and predation.

Materials and methods

Study system

The Stewiacke River, Nova Scotia is one of 50 rivers within the inner Bay of Fundy (iBoF) Atlantic salmon designatable unit (DFO 2019a). The iBoF unit is currently listed as Endangered under Canada's Species at Risk Act. Low survival during the estuarine and marine stages of the Atlantic salmon life cycle is preventing population recovery (DFO 2019b). Reducing adult marine mortality is challenging, therefore, identifying sources of mortality and quantifying predation rates of migrating smolts is vital to informing population management. Smolts migrate down from the Stewiacke River and its tributaries out to the Minas Basin via the Shubenacadie River (Fig. 1). The Stewiacke River is the only river in the iBoF unit that is confirmed as an annual spawning site for striped bass (Bradford et al. 2015). Striped bass congregate in the tidal waters of the Stewiacke River to spawn in May–June (Bradford et al. 2015), the same time and location as the smolt out-migration.

Field methods

Sampling and tagging procedures

Sampling of Atlantic salmon smolts occurred within the Stewiacke River watershed in three years, spanning

2017–2019, during the annual smolt run. Smolts were captured via rotary screw trap just downstream of the Stewiacke River head-of-tide in 2017 and just upstream of the head-of-tide in 2018 (<2 km apart; Fig. 1). In 2019, smolts were captured using a barrier fence on the Pembroke River, ~40 km upstream of the head-of-tide (Fig. 1). Both types of traps were checked for fish daily. Smolts were transferred from the traps to floating bins in a calm section of the river for holding prior to sampling and surgeries. Fifty smolts were tagged in both 2017 and 2018; 56 smolts were tagged in 2019 (total $N = 156$).

Fish were measured prior to surgery (fork length [mm], mass [g]). Only smolts longer than 12 cm in fork length were chosen for tagging to ensure that the recommended tag-to-body size ratio was not exceeded (<8% for Atlantic salmon; LaCroix et al. 2004). The average tag-to-body size ratio across all years was 2.95% (range 0.95–5.23%). Smolts were then anaesthetized in a buffered 10 mg/L solution of tricaine methanesulfonate (MS-222), until loss of equilibrium and spinal reflexes. A maintenance solution of buffered 5 mg/L tricaine methanesulfonate was circulated over the gills of the fish during surgeries. V5D-180 kHz predation acoustic transmitters (12.7 × 5.6 mm, 0.68 g in air; Innovasea Systems Inc., Bedford, Nova Scotia) were surgically inserted through a ~8 mm incision in the abdomen of smolts following standard procedure (Cooke et al. 2011). Incisions were closed with two single interrupted sutures. Smolts were returned to the floating river-side bins and held until dusk to recover from surgery before release just downstream from the point of capture. The average duration for the measuring and surgical procedures was 3.27 ± 0.74 min, and average recovery times were $\sim 7 \pm 1$ h.

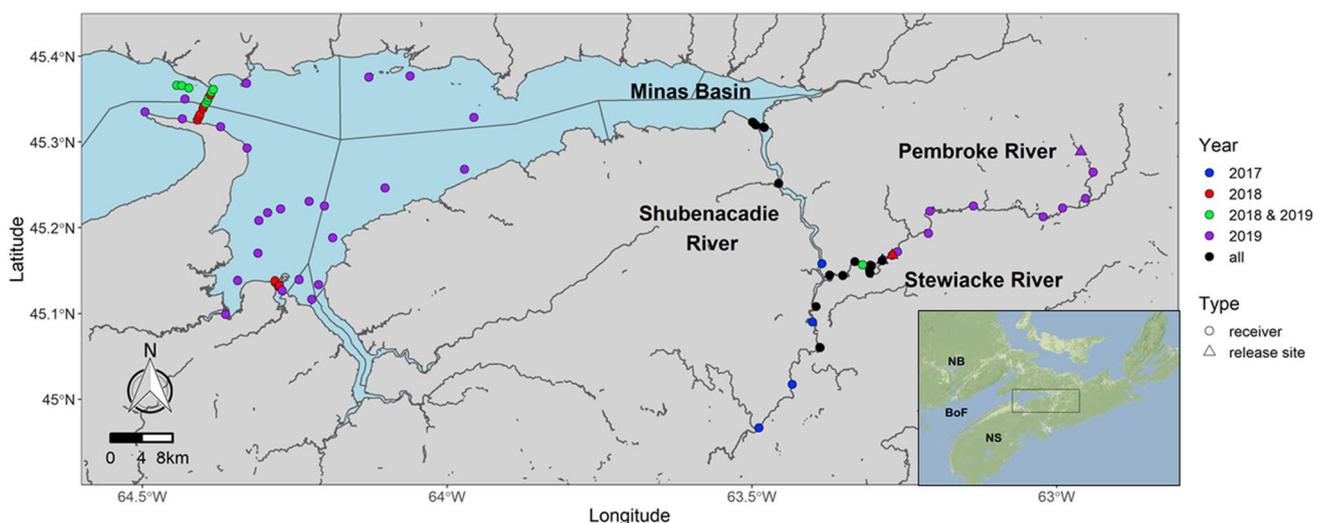


Fig. 1 Map showing receiver and release site locations for each study year in the Stewiacke River watershed and Minas Basin, Nova Scotia, Canada. Inset shows location of study area (box) in relation to Nova Scotia (NS), New Brunswick (NB), and the Bay of Fundy (BoF)

Fish collection permits were issued by Fisheries and Oceans Canada (DFO 323,354). All fish handling and surgical procedures conformed to standards established by the Canadian Committee on Animal Care, via permits issued by Fisheries and Oceans Canada (Maritimes Region Animal Care Committee Animal Utilization Protocols 17–16, 18–13, 19–10) and by Dalhousie University (University Committee on Lab Animals permit 18–126). Field work was done in conjunction with the Mi'kmaw Conservation Group who were operating under the Aboriginal Fund for Species at Risk.

Description of tags and receiver array

The V5D tags (Innovasea Systems Inc.) have a biopolymer coating that triggers a change in transmitter ID (from an even number to the next odd number) when dissolved by the stomach acids of a predator, thus indicating that a predation event has occurred. It is assumed that only predation events by fishes will be detected using this technology because avian or semi-aquatic predators would more likely remove the tag from the study area (Daniels et al. 2019). The lag time between tag consumption and the activation of the predation signal is ~5.8 h at 20 °C (S. Smedbol, Innovasea Systems Inc., pers. comm., January 2020) or 35.4 ± 17.7 h at a mean temperature of 11.8 °C (Hanssen 2020). In addition to temperature, lag time is dependent on the species and size of the prey (tagged) fish and on the species and digestion rate of the predator (Halfyard et al. 2017).

Prior to tagging, an array of VR2W-180 kHz acoustic receivers (Innovasea Systems Inc.) was deployed along the migration route from the release/tagging site to the mouth of the Shubenacadie River ($n = 16$ in 2017, $n = 15$ in 2018, $n = 24$ in 2019; Fig. 1). Supplemental detection data were provided by additional receivers (VR2W-180 kHz and HR2; Innovasea Systems Inc.) deployed in the Minas Basin (Fig. 1) and maintained by other researchers including the Ocean Tracking Network. Receivers were recovered in mid-to late July of each year.

The V5D tags were programmed to transmit individual-specific coded signals every 12–18 s for detection on VR2W receivers in all years, and every 1.9–2.1 s for detection on HR2 receivers in 2018 and 2019. Tags in 2017 had an estimated battery life of 47 days, while tags in 2018 and 2019 had a battery life of approximately 24 days due to the dual programming for both types of receivers.

Data analyses

All analyses were conducted in R 3.6.2 (R Core Team; <https://www.R-project.org>). Detections occurring before or after the study period were removed as well as detections of tagged fish belonging to other studies. Detections were

filtered using the false detections function from the *glatos* package (Binder and Dini 2019; Pincock 2012; Simpfendorfer et al. 2015). This function identifies potentially false detections based on the programmed time interval at which the tags emit the ID signal and the recorded time between detections. Detections were then plotted for each individual smolt and visually assessed; detections identified as potentially false by the filtering function that also looked improbable given the location of receivers were removed from the data set. In the case that a dead smolt or evacuated tag dropped within range of receiver (i.e., resulting in a continuous string of detections for extended periods of time), the detection data were truncated to the first detection at that receiver ($n = 3$).

Fate classification

Detection data and the V5D pH sensor were used to classify smolts as belonging to one of three fate groups: successful migrant, mortality of unknown cause, or predation. Smolts were considered to have successfully completed migration if the last recorded detection was either at the mouth of the Shubenacadie River or in the Minas Basin. Smolts were presumed to be a mortality if their last recorded detection occurred upstream of the Shubenacadie River mouth. This pattern of detections could also result from tag ejection, failure to be detected when passing receivers, or predation by an animal that removed the tag from the study site. Predated smolts were identified if the pH sensor triggered a change in tag ID. However, preliminary analysis of detection data revealed that some smolts identified to be successful migrants or mortalities displayed movements more similar to predator behaviour than migratory smolt behaviour (several reversals between up- and downstream movement; Fig. 2). Consultation with the tag manufacturer confirmed the possibility of undetected predation events (Type II error). Additionally, a previous validation study has shown V5D predation tags to have only 50% accurate detection of predation (Hanssen 2020). Therefore, machine learning methods were also applied to the detection data to classify smolt fate.

Behavioural metrics for the machine learning models were calculated from detection data of both live and predated tag IDs. The metrics were selected based on behaviours that are expected to be significantly different between salmon smolts and a predator such as striped bass. Some of these metrics are adapted from Gibson et al. (2015) and Daniels et al. (2018). The chosen metrics were total number of detections, maximum and minimum number of detections at a single receiver, number of days with detections, time between release and last detection, total distance travelled (river km), mean and maximum upstream speed (m/s) between two consecutive receivers, mean and maximum downstream speed (m/s) between two consecutive receivers,

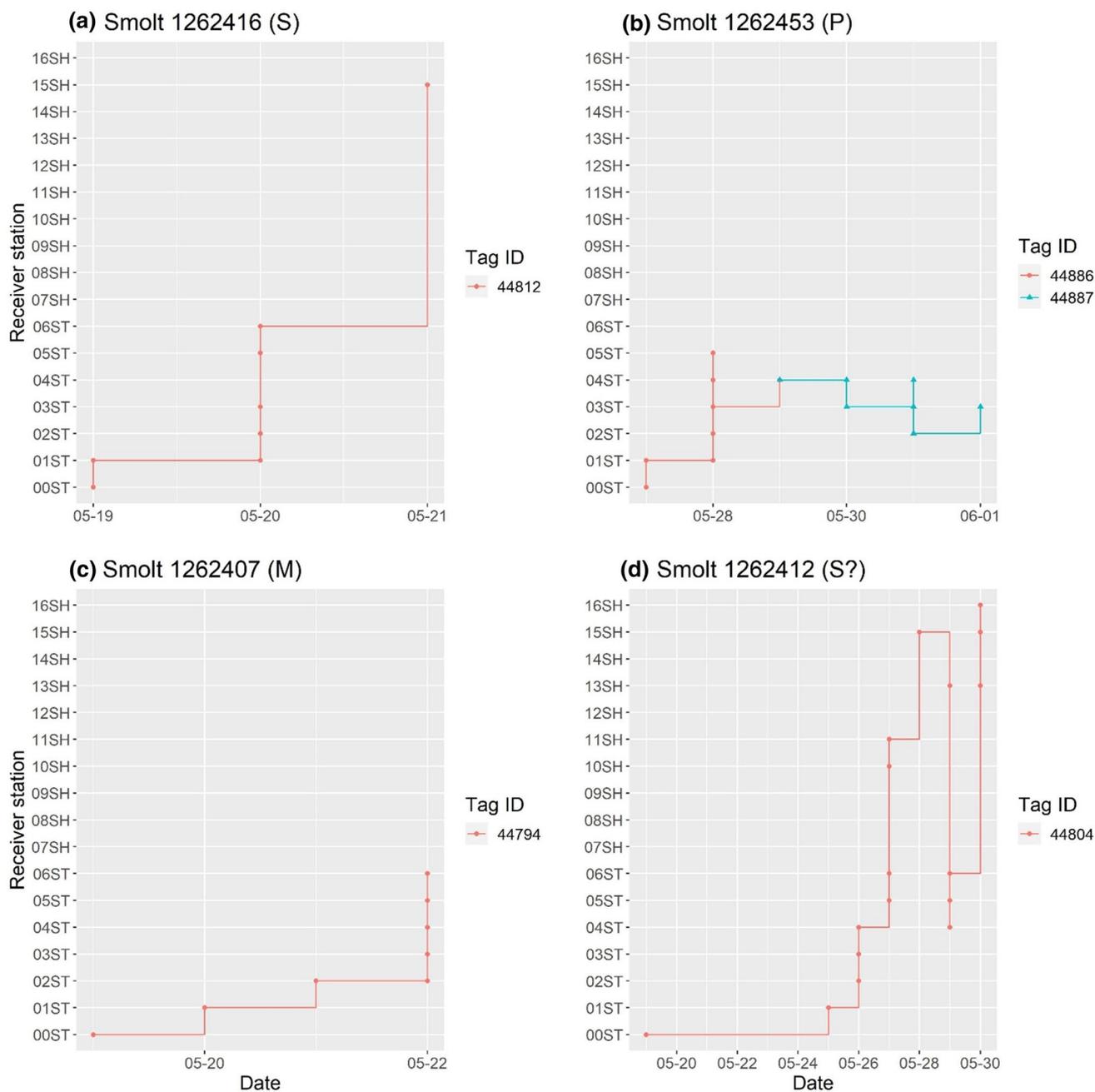


Fig. 2 Plots showing detections of four individual Atlantic salmon (*Salmo salar*) smolts in 2017 representing typical migration paths of a successful migrant (S), a predated smolt (P), a mortality of unknown cause (M), and a smolt indicated to have successfully com-

pleted migration but suspected of being predated (S?). Receiver stations listed in order from release on the Stewiacke River (00ST) to the mouth of the Shubenacadie River (14–16SH). Receivers 07–12SH are upstream of the ST-SH confluence

total number of reversals in up- and downstream movement, total time on striped bass spawning grounds, total number of detections above the Stewiacke River and Shubenacadie River confluence, cumulative upstream distance travelled (river km), mean and maximum upstream distance travelled in a single step (river km), migration rate (river km/day), and for 2019, maximum speed in freshwater and tidal water

(m/s). Metrics were tested into unsupervised k-means cluster analyses and supervised random forest models to compare fate classification based solely on the behavioural metrics, classification based on behaviour but also trained on individuals with known fate, and classification from detections and tag sensor only. Due to differences in receiver array set-up between years, models were run separately for each year.

Attempts to pool years by truncating detection data to the smallest study area among years (2017 array) resulted in the removal of several individuals from the data set and did not increase model accuracy beyond what was generated from individual year models.

K-means clustering

Clustering is a family of unsupervised machine learning where an algorithm is developed to form groups based on similarities in the data without prior identifiers (Jain 2010; Thessen 2016). Therefore, the class of each group is inferred and requires context-specific knowledge to be interpreted. Types of clustering can be categorized as hierarchical or partitional (Jain 2010). Hierarchical methods create nested clusters by either merging data points into clusters (agglomerative) or dividing a single cluster into smaller ones (divisive). Partitional methods, such as k-means clustering, produce all clusters simultaneously. Clusters are formed to maximize similarity within clusters and minimize similarity between clusters. In k-means clustering, the number of clusters K is specified by the user.

K-means clustering was performed using the k-means function in base R. Behavioural metrics were centered and scaled to remove the effect of variables with larger values. Individual smolts were clustered into three groups ($K=3$) to represent the three fate classes, successful migrants, mortalities, and predated smolts. The *fviz_cluster* function was used to visualize cluster results, which plots observations using principal components (Kassambara and Mundt 2019). Variable importance for clustering was measured by the rate at which individuals were misclassified if that variable was removed from the data set (misclassification rate). A higher misclassification rate means a variable is more important for assigning an individual to the best cluster. ANOVAs and Tukey tests were used to test if variables were significantly different between clusters. Each group was then assigned a fate based on metric means for each cluster and expected behaviour of out-migrating smolts. Total distance travelled is expected to be longest in successful migrants who reach the Minas Basin and shortest among mortalities that die along the migration route. Total distance should also be long in predated smolts due the distance accumulated by the up- and downstream movements made by predatory striped bass. It is expected that total time would follow a similar trend, with predated smolts showing less time than successful migrants due to the ejection of tags through the gastrointestinal tract of predators, and mortalities being detected for the least amount of time. Upstream speed should be fastest among predated smolts and very slow among successful migrants and mortalities. Similarly, upstream distance should be longest in predated smolts and shortest in successful migrants and mortalities because striped bass are expected to make

frequent, extensive reversals in swimming direction while smolts are expected to conduct directed, downstream movements.

Random forest

Random forest is a supervised method of machine learning that builds upon classification trees by fitting many trees to a data set to increase the accuracy of classification (Cutler et al. 2007). Each tree is fit to a bootstrapped sample of the original data set with only a subset of the variables considered at each node. Each observation is then classified by majority vote of all the trees. The random forest algorithm is first trained on a data set where the class of each observation is known to learn the relationship between the response and predictors, before being used to predict classes of new observations.

The random Forest package and function (Liaw and Wiener 2002) in R was used to create a model with fate as determined by the tag pH sensor and detection data as the response and the behavioural metrics as explanatory variables. Individuals with uncharacteristic smolt behaviour were removed from the data set prior to training the algorithm. Small sample size prevented cross-validation with training and test data sets. Therefore, out-of-bag (OOB) error produced from bootstrapping was used to calculate a confusion matrix and model accuracy. The number of trees made in the model was increased from the default 500 until OOB and class error rate fluctuations stabilized. The number of variables tried at each node was chosen based on minimizing OOB error. Due to class imbalance, the classes were assigned weights to penalize misclassification of under-represented classes, class weights were chosen to minimize and balance class error rates (Table 1). The final model was then used to predict the migration fate of individuals suspected of being predated but where a predation event was not detected by the tag sensor ($n=16$ in 2017, 14 in 2018,

Table 1 Random forest model metrics

Parameter	2017	2018	2019
Ntree	1000	1000	500
Mtry	3	3	2
Classwt c(MPS)	2, 1, 10	N/A	5, 2, 1
OOB error	14.71%	5.56%	18.37%
Class error c(MPS)	0.25, 0.04, 1.00	0.33, 0.05, 0.00	0.43, 0.50, 0.00

Number of decision trees made (ntree), number of variables considered at each node (mtry), class weights assigned to mortalities, predated smolts, and successful migrants, respectively (claswt MPS), out-of-bag error rate (OOB error), and class error rate for mortalities, predated smolts, and successful migrants, respectively (class error MPS)

7 in 2019) using the predict function. Variable importance was described by the average decline in model accuracy after permutations of that variable (mean decrease accuracy) and the average decrease in node purity if that variable was not used (mean decrease Gini). Larger values for both mean decrease accuracy and mean decrease Gini indicate greater variable importance.

Results

Predation tags

The number of tagged fish determined to be predated based on the predation sensor was 24 (48%), 18 (36%), and 14

(25%) in 2017, 2018, and 2019, respectively. In 2019, two of the predation events occurred after entry into the Minas Basin and were therefore classified as successful migrants rather than predated smolts.

K-means clustering

For each year, smolts were placed into one of three clusters (Fig. 3). The most important variables differed somewhat among years; variables with consistently high misclassification rates included total distance travelled, total time detected, upstream swimming speed, and upstream distance travelled (Figs. S1-3). These variables were significantly different (ANOVAs, Tukey tests) between at least two out of three clusters for each year. Therefore, clusters were

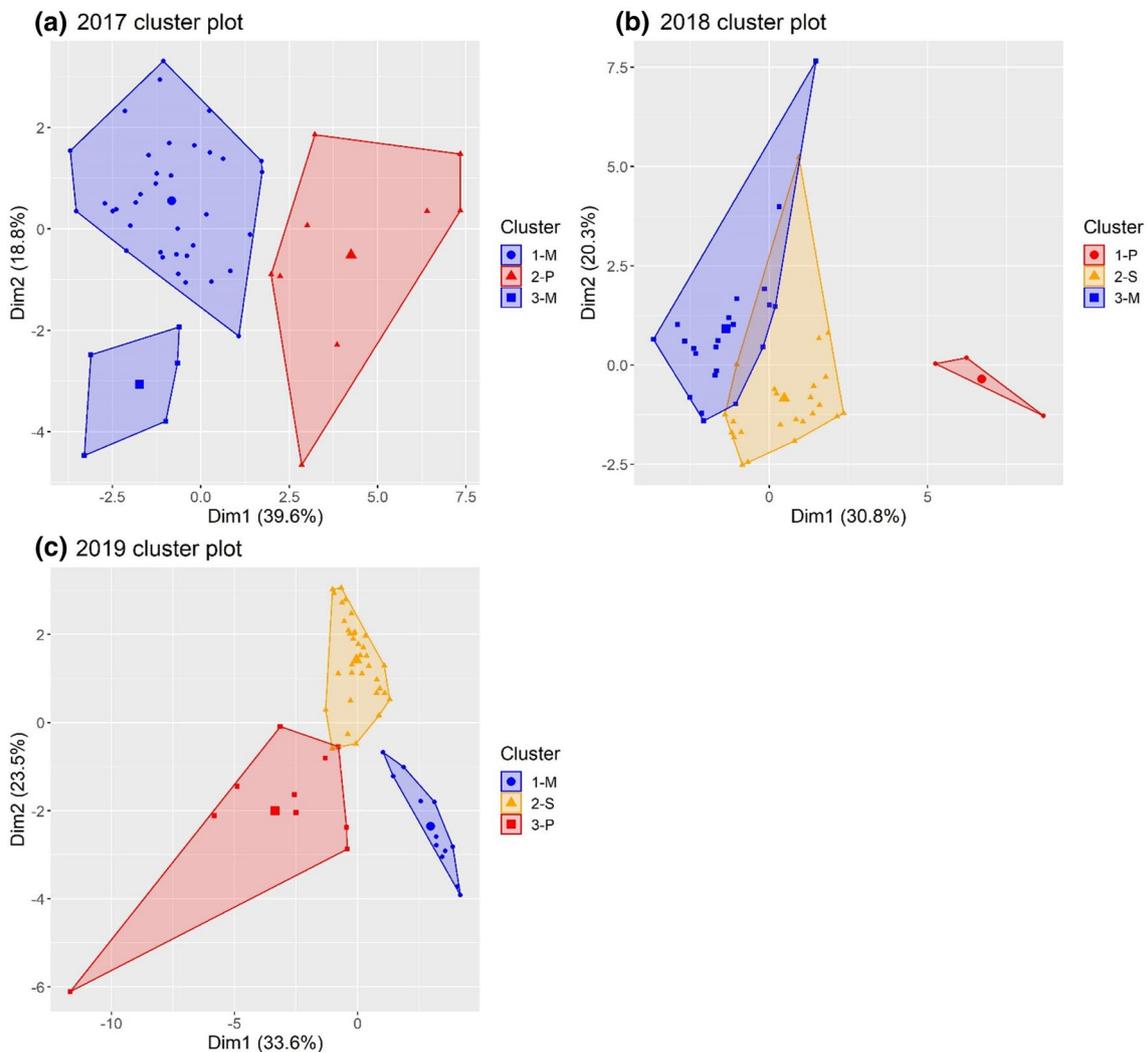


Fig. 3 Cluster plots for all study years showing $k=3$ clusters representing the three fate groups (mortality M, successful migrant S, predation P). Cluster colour corresponds to fate, point shape corresponds to cluster number. Small points are individual smolts (2017 $n=50$,

2018 $n=50$, 2019 $n=56$), large points are cluster centroids. Clusters plotted on the first two principal components (Dim1, Dim2 [% variance explained])

assigned fate classes based on the differences in these variables and the expected behaviour of live salmon smolts, dead smolts, and predators.

For 2017, cluster 2 ($n=9$) had faster upstream swimming speeds, longer upstream distances travelled, and farther total distance travelled than clusters 1 and 3 (Fig. S4). These trends are more characteristic of striped bass movement than smolt movement; therefore, cluster 2 was determined to represent the predated fate class (Fig. 4). Clusters 1 ($n=36$) and 3 ($n=5$) were not significantly different from each other (Tukey tests; upstream speed: $t=-0.08$, $p=0.997$; total

distance: $t=-1.16$, $p=0.476$; upstream distance: $t=-0.33$, $p=0.941$). Based on the short total distance travelled (Fig. S4), both of these clusters were assigned the fate of mortality of unknown cause. A successful migrant cluster was not identified.

The cluster plot for 2018 revealed some overlap between clusters 2 and 3 when plotted on the first two principal components (Fig. 3); however, they were significantly different from each other when examining variables with the highest misclassification rates (Tukey tests; total distance $t=-10.5$, $p<0.001$; total time $t=-5.40$,

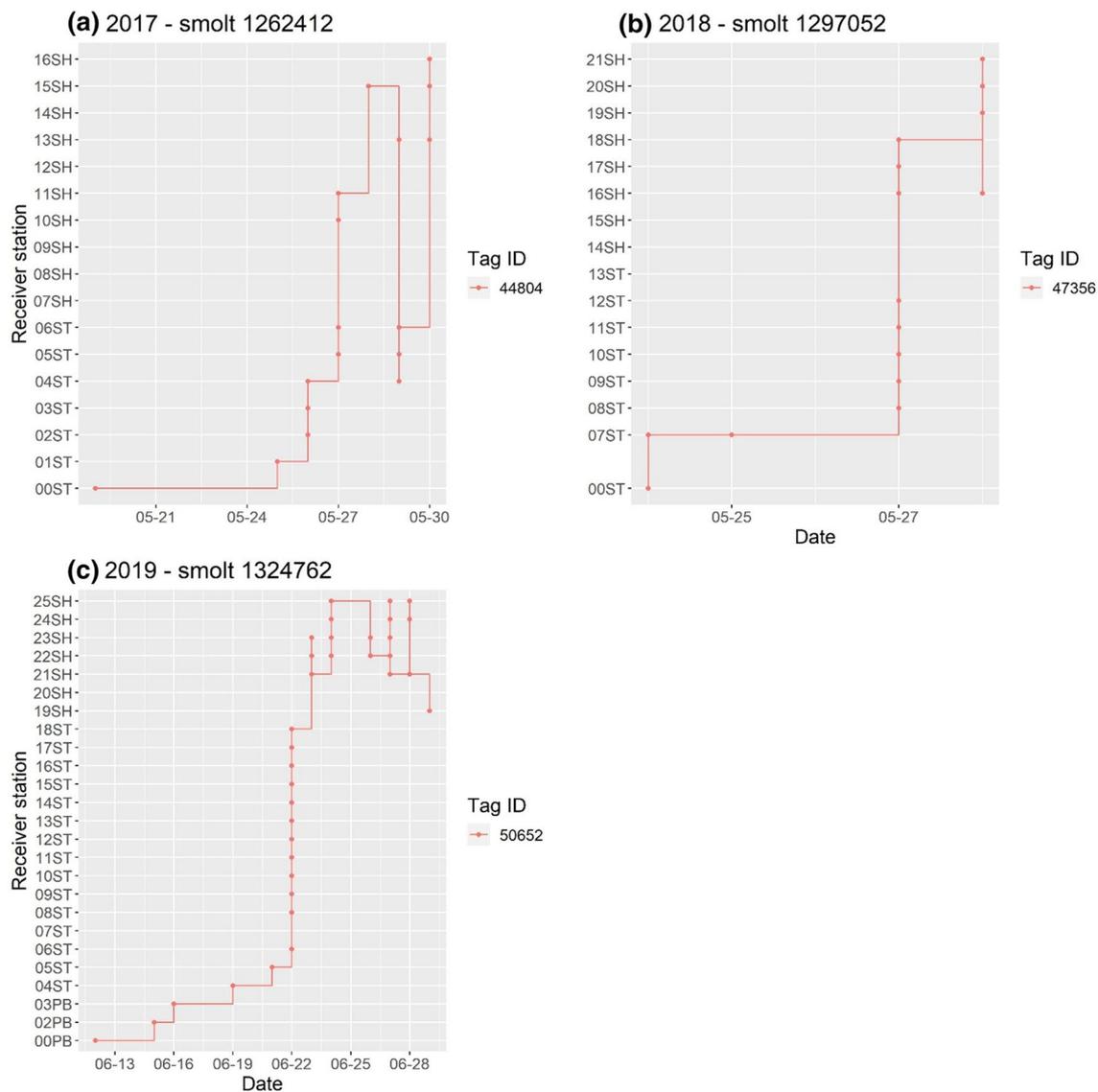


Fig. 4 Plots showing detections of three individual smolts in 2017, 2018, and 2019. Receiver stations listed in order from release on the Stewiacke River (00ST) or Pembroke River (00 PB) to the mouth of the Shubenacadie River (14–16SH; 17–21SH; 22–25SH). **a** 2017-smolt 1262412 was classified as a successful migrant based on tag detections but classified as predated by both k-means clustering and

random forest. **b** 2018-smolt 1297052 was classified as a successful migrant based on tag detections and k-means clustering but classified as predated by random forest. **c** 2019-smolt 1324762 was classified as a mortality based on tag detections but classified as predated by both k-means clustering and random forest

$p < 0.001$). Cluster 2 ($n = 24$) had the longest total time and farthest total distance (Fig. S5); therefore, it was assigned the successful migrant class (Fig. 4). In contrast to cluster 2, cluster 3 ($n = 23$) showed the briefest total time and shortest distance migrated (Fig S5), which are metrics indicative of mortality. Cluster 1 ($n = 3$) had intermediate values between clusters 2 and 3, and total distance was greater than total time leading to the assignment of the predated fate to this cluster.

In 2019, cluster 3 ($n = 10$) had a significantly greater number of reversals (Tukey test; cluster 1 $t = 8.77$, $p < 0.001$; cluster 2 $t = 9.75$, $p < 0.001$), longer time on striped bass spawning grounds (Tukey test; cluster 1 $t = 5.29$, $p < 0.001$; cluster 2 $t = 5.30$, $p < 0.001$), and longer upstream distance travelled (Tukey test; cluster 1 $t = 6.48$, $p < 0.001$; cluster 2 $t = 7.18$, $p < 0.001$), all of which are behaviours indicative of predation by striped bass (Fig. 4). Cluster 2 ($n = 34$) had the longest total distance (Fig. S6) and was therefore, assigned the successful migrant fate. Conversely, cluster 1 ($n = 12$)

had the shortest distance travelled and was identified as the mortality class.

Model accuracy, calculated by the number of known fates within a cluster that matched that cluster’s assigned fate (Table 2), was 38.2, 52.8, and 82.4% for 2017, 2018, and 2019, respectively (Fig. 5).

Random forest

In-sample prediction accuracy of random forest algorithms ranged between 81.6 and 94.4% between years (Fig. 5). OOB error rates ranged from 5.56 to 18.37% (Table 1). In-sample classification of the successful migrant class had 100% error rate in the 2017 model, but 0% in 2018 and 2019. Mortality class error rates ranged from 25 to 43% (Table 1). Error rates for the predated class were similar in 2017 and 2018 at 4 and 5%, respectively, but was 50% in 2019.

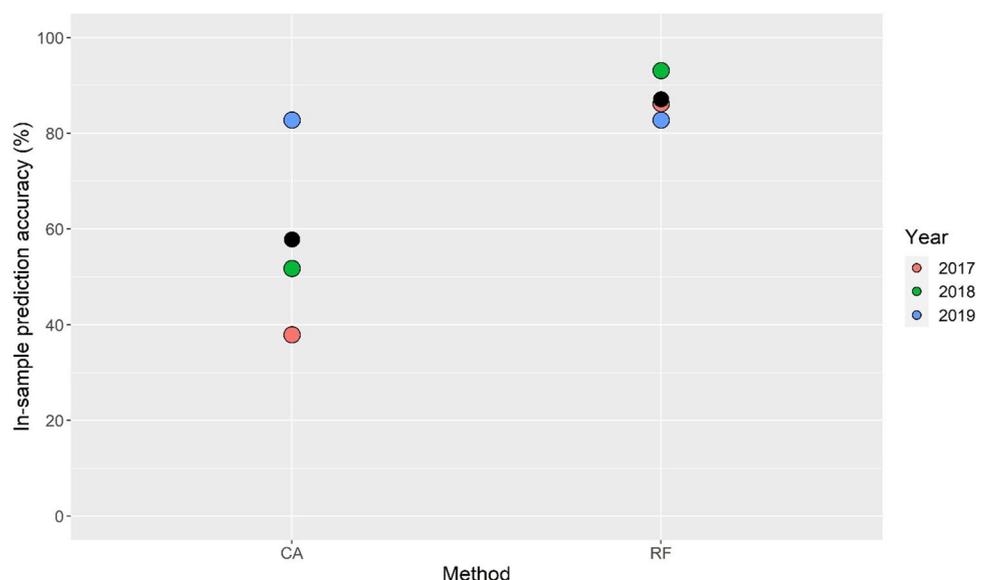
The most important variables in common among all years were time on striped bass spawning grounds, total distance

Table 2 Number of individuals of each fate (predated P, other mortality M, successful migrant S, successful migrant or mortality suspected of being predated U) as determined by predation tag and detection data in each cluster

Fate assigned by tag	2017			2018			2019		
	Cluster 1 (M)	Cluster 2 (P)	Cluster 3 (M)	Cluster 1 (P)	Cluster 2 (S)	Cluster 3 (M)	Cluster 1 (M)	Cluster 2 (S)	Cluster 3 (P)
S	2	0	0	0	14	1	0	30	0
M	9	0	0	0	0	3	6	1	2
P	16	4	4	2	1	15	6	0	6
U	10	5	1	1	9	4	0	3	2
total	36	9	5	3	24	23	12	34	10

Cluster fates, in brackets, assigned based on average behavioural metrics of each cluster

Fig. 5 Model accuracies for k-means clustering (CA) and random forest (RF) models by year, mean accuracy shown by black circle



travelled, and time detected (Figs S10–12). Upstream speed, upstream distance, and number of reversals were also important variables in 2017 (Fig. S10). Partial plots revealed that the probability of being classified as a successful migrant increased with cumulative distance travelled, total time detected, and number of days detected (Figs. S13–18). The probability of being classified as predated increased with number of reversals, upstream distance travelled, upstream speed, and time spent on striped bass spawning grounds (Figs. S19–24). The trends in probability of being classified as a mortality were similar to those for the predated class except for time on striped bass spawning grounds, time detected, and distance travelled in which cases the trends were opposing.

The 2017 random forest algorithm reclassified all suspect individuals (five successful migrants, 11 mortalities) as predated (Fig. 4). In 2018, two of the suspect mortalities were reclassified as successful migrants but these individuals were retained as mortalities in the final fate counts. All other suspect mortalities were reclassified as predated, but only two among eight suspect successful migrants were reclassified as predated (Fig. 4). The 2019 algorithm reclassified the two suspect mortalities as predated (Fig. 4) and three of five suspect successful migrants as predated.

Comparison of classification methods

When comparing the assigned cluster fates to the known fates of individuals within clusters, as determined through detection data, the accuracy was highest in 2019 because the majority class in each cluster was the same as the assigned cluster fate (Table 2). The 2018 cluster assignments were also consistent with individual fates; however, two of the clusters were mostly comprised of predated smolts (Table 2). Cluster 1, which was assigned the predated fate, contained only two predated individuals while cluster 3 contained the majority of predated individuals (15) but was assigned a mortality fate based on the behavioural metrics. The 2017 clusters were difficult to distinguish based on behaviour and individual fates due to the high number of predation events that year, predated individuals were spread among all three clusters (Table 2). Compared to predation tag data, the cluster analysis reduced predation estimates by 30% in 2017, 30% in 2018, and 3.5% in 2019 (Table 3).

Random forest algorithms consistently increased the percentage of individuals classified as predated and resulted in a reduction of estimated migration success and mortality classes compared to the numbers obtained from the predation tag sensor and detection data (Table 3). Predation rates increased by 32, 12, and 9% in 2017, 2018, and 2019, respectively.

Unsupervised clustering methods were capable of fate classification but were less accurate than supervised methods (Fig. 5).

Discussion

Here, we build on previous studies to develop a standardized workflow for identifying predated individuals in acoustic telemetry studies (Fig. 6). Usership of acoustic predation sensor technology will likely increase in coming years so long as the tools available for users provide accurate fate

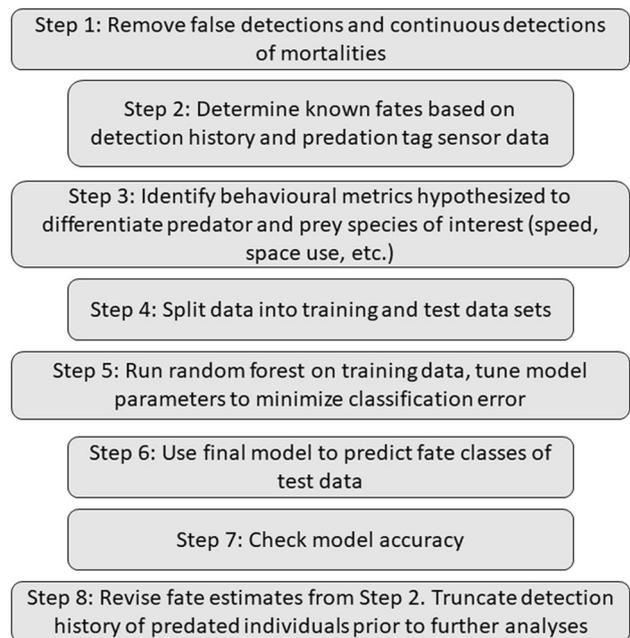


Fig. 6 Diagram of example workflow to identify predated individuals in telemetry data using predation tags and a machine learning framework

Table 3 Percent of smolts belonging to each fate (predated P, other mortality M, successful migrant S) as determined by the V5D predation tag sensor and detection data, unsupervised cluster analysis (CA), and supervised random forest (RF)

	2017			2018			2019		
	Tag sensor	CA	RF	Tag sensor	CA	RF	Tag sensor	CA	RF
S (%)	14	0	4	46	48	42	62.5	60.7	57.1
M (%)	38	82	16	18	46	10	16.1	21.4	12.5
P (%)	48	18	80	36	6	48	21.4	17.9	30.4

classifications; therefore, we compared the results and accuracies of three classification approaches. Unsupervised k-means clustering underestimated the number of predated smolts and due to type II error, the tag sensor may have as well. We demonstrated how tag sensors could be enhanced by supervised machine learning (random forest) in a generalizable workflow that can become standard practice for acoustic telemetry in an effort to address the issue of “predation bias”.

The k-means clustering method was able to group smolts based solely on behavioural metrics, but it can be difficult to discern which cluster represents which fate group and the decision is likely to be subjective. Assigning fates to clusters was dependent on distinct and predictable predator and prey behaviour with smolts moving downstream and striped bass exhibiting multiple reversals. However, it is possible that smolts could exhibit upstream movement if they were being carried by the tides (Beland et al. 2001) or as a response to osmotic stress (Halfyard et al. 2012). Factors such as the amount of time the predator is tracked and location of the predation event impacted clustering of smolts. Predation events where the tag was ejected quickly (under 70 h) from the gastrointestinal tract of the predator led to smolts being placed in mortality clusters rather than predated clusters due to insufficient detections of predator behaviour (2018 cluster 3 $n = 13$, 2017 cluster 1 $n = 15$). Predation events that occurred in freshwater where the most likely predators, brown trout (*Salmo trutta*) or chain pickerel (*Esox niger*), are relatively stationary species were also incorrectly clustered with mortalities because detection data resembled a dropped tag or dead smolt rather than the active striped bass behaviour we were testing for. These predation signals may have also been a result of smolts dying and degradation causing the pH sensor to trigger rather than actual predation (Halfyard et al. 2017). Similar to the cluster analysis, the 2019 random forest algorithm did not successfully differentiate the six freshwater predation events from the mortalities. The unsupervised clustering approach was not well suited for the workflow used here and accuracy would further decrease in systems where less is known about the study system and expected predator-prey interactions.

Data from 2017 showed the greatest disparity of fate assignments amongst the three different classification methods (Table 3). In addition to overall model classification accuracy, balancing accuracy amongst classes is important especially for unbalanced data sets because models will ignore minority classes to achieve greater overall accuracy (Chen et al. 2004; Brownscombe et al. 2020). The small number of successful migrants compared to the number of mortalities and predated smolts in 2017 made it difficult for these individuals to be recognized by either type of machine learning approach. The few successful migrant smolts were masked in the cluster analysis

by the behavioural characteristics of the other fate classes (Table 2), and despite the addition of class weights, the random forest model was still unable to accurately classify successful migrants. In contrast, the percentage of successful migrants was relatively consistent amongst all three methods in 2018 and 2019, while mortality and predation classes had larger disparities, especially for the 2018 cluster analysis (Table 3).

The supervised random forest was the most accurate of the three fate classification methods. This method increased estimated predation rates greatly beyond estimates made by the tag pH sensor alone and by the unsupervised cluster analysis (Table 3). This suggests that the sensor onboard the tag is prone to false negatives, as was shown by Lennox et al. (2021) in a combined laboratory and field study of salmon smolts in Norway. Predation accounted for majority of all smolt mortalities (71–83%) under the random forest estimates while tag sensor estimates showed predation as accounting for just above half of all mortalities (56–67%). Although confidence in classifying smolt fate was increased through the combined use of predation tags and random forest algorithms, true smolt fate is still not known with certainty. The migration fates of individuals in the training data sets are referred to as having known fates but because fates were determined through detection data, there is still uncertainty associated with these fates due to the assumptions of no tag loss or failure and the potential for imperfect detection efficiency. Making such assumptions about tag function and the fate of a tagged animal is the nature of most telemetry studies, unless recapturing the individual or tag is possible. While the migration fates modeled by random forest algorithms were shown to be the most accurate estimates, average in-sample prediction accuracy was still less than 90% and there is the potential that some smolts were misclassified.

The supervised random forest models developed by Daniels et al. (2018), where detection data from tagged salmon smolts ($N = 63$) and striped bass ($N = 228$) were used to train models, had classification error rates of 1.6% for the smolt class and 0% for the striped bass class. The classification error rates of the random forest models developed here ranged from 0 to 100% for the successful migrant smolt class and 4–50% for the predated smolt class. The larger sample size used by Daniels et al. (2018) likely resulted in the reduced classification error rates. Additionally, tagging striped bass provides longer detection history to train models to recognize predator behaviour. This comparison shows that predation tags are not needed if there are detection data available for the predator species and if predator behaviour is distinct from prey behaviour. However, the use of predation tags is likely the best approach for identifying predation events when behaviour is not distinct between predator(s) and prey.

We emphasize the importance of distinguishing predation from other forms of mortality due to the substantial bias it introduces into telemetry study results and interpretation if not addressed (Klinard and Matley 2020). Previous researchers who have used classification algorithms to identify predation of tagged fish found that without these analyses, inferences about the spatial and temporal movement of 81% of bonefish would have been biased (Moxham et al. 2019); mortality rates of salmon smolts in freshwater compared to the estuary were underestimated by 10% (Daniels et al. 2018), and survival estimates of salmon smolts were overestimated by 2.4–13.6% (Gibson et al. 2015). Here, even with the use of predation sensor tags, random forest models suggested that survival estimates were overestimated by 4–10% due to undetected predation events. Therefore, accurate classification of fate in telemetry studies is vital to management not only to investigate sources of mortality in a population but also to ensure accurate conclusions are drawn about the ecology of the study species and population survival rates.

General considerations and applications

Model performance can be improved by optimizing receiver configuration and coverage, which are vital to capturing the distinct behaviours needed to differentiate predator and prey species. Future users investigating predation in different systems should consider how the spacing and number of receivers in an array will affect the ability to calculate behavioural metrics for both supervised and unsupervised methods. We recommend spacing receivers in a way that allows analysts to calculate accurate migration speeds and distances within a study area. The distance between receivers in a river system affects the accuracy of distance travelled and speed calculations because the movement of the individual between receiver detection ranges is unknown. Different retention times of tags in predators and the amount of time a predator remains within detection range of receivers will also affect the accuracy of behavioural metric calculations and model predictions.

Deciding on behavioural metrics prior to receiver deployment can aid in array design to ensure receiver coverage is adequate for calculating the necessary metrics. However, because it is possible to have multiple or unknown predatory species in a study system, calculating metrics or concentrating receiver coverage for only one species could mask predation by another. Additionally, avian predation typically resembles mortalities in terms of detection data and could therefore not be identified using the presently available tools, although other researchers have identified avian predation by searching colonies or nesting sites for evacuated tags (Evans et al. 2012). For tracking migratory species specifically, good up- and down-stream receiver coverage of an area is important for distinguishing predator and prey.

The behavioural metrics required for machine learning approaches are context-specific and must be tailored to the prey and predator species of interest. Swimming speed is a good metric because predators are generally larger than prey and are therefore able to swim faster. Distance travelled is also a useful metric if one species is more stationary and the other is more mobile. Space use metrics should be used when the species of interest prefer different habitat types. Additional metrics, such as time of day or season, when individuals are most active may be considered if for example one species is expected to feed and move less during daylight hours or winter months and the other species is not.

K-means clustering requires the specification of K prior to running the model. We chose $K=3$ to represent the three fate classes we wanted to differentiate, but in other cases (multiple or unknown predator species), it may be difficult to select K . In this situation, multiple iterations of k-means can be run with different values of K and the best model can then be chosen based on a selection criterion, such as AIC, BIC, within sum of squares, gap statistics, minimum message length, or the Dirichlet Process (Jain 2010). When prey and one or more predator species have similar behaviour, this will likely result in mixed clusters containing more than one species and alternative clustering approaches may also be used. Fuzzy k-means assigns a probability of belonging to each cluster to each observation, this is known as “soft” assignment in contrast to the standard k-means “hard” assignment approach (Jain 2010). Soft assignment allows for uncertainty in clustering results and a predefined cut-off probability value can be used to accept an observation as truly belonging to a cluster. Hierarchical clustering both removes the necessity to pre-define K and produces nested clusters providing a view of the underlying structure within clusters and which individuals are most similar (Whittingham and Ashenden 2021).

As a supervised branch of machine learning, random forest requires training the algorithm with observations where the outcome or classification are known. Here, we used detection data and the predation tag sensor to class the individuals used to train the models. This approach is still somewhat subjective and in cases where there are multiple predatory species or predator behaviour is not distinct, classifying the fate of individuals in this way for the training data set would likely decrease accuracy. Advances in predation sensor technology including improvements to the pH sensors since 2019 and the ongoing development of orientation sensors (Lennox et al. 2021) can improve supervised machine learning predictions and overall classification accuracy in predation studies.

Differences in model results and prediction accuracies among years highlight the importance of having a large sample size not only for greater power in model predictions but also in an attempt of balancing classes for individuals

of known fates. Random forests are among the least sensitive classification algorithms to reductions in sample size (Maxwell et al. 2018; Moghaddam et al. 2020); however, issues of class imbalance and potentially unrepresentative data remain when using small training data sets (Chen et al. 2004; Brownscombe et al. 2020). A recommendation for machine learning in general is to have a training sample size ten times the number of predictor variables, but the minimum recommended sample size for classification algorithms specifically is dependent on the type of data and algorithm (Indira et al. 2010; Maxwell et al. 2018).

Here, we show that there is value in using predation tags combined with modelling methods to identify predated individuals (Fig. 6). Modelling data that include individuals with known fates that have been determined by detection data and a pH or other tag sensor increases confidence in model results and improves model accuracy. Random forest modelling and the example workflow we provide, allows one to study predation using predation tags, therefore removing the need to tag predators, while also accounting for sensor malfunctions. We recommend combining acoustic tag sensors with supervised machine learning approaches to identify mortalities and predation of tagged fishes, thereby increasing confidence in telemetry study results.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00442-022-05138-3>.

Acknowledgements We would like to acknowledge Cindy Hawthorne, Jeff Reader, Alana Ransome, George Nau, and other members of Fisheries and Oceans Canada and the Mi'kmaw Conservation Group for assistance in the field. As well as Darren and Erica Porter for receiver deployment and retrieval in the Minas Basin. We also thank Jake Brownscombe for answering questions about machine learning model tuning and members of the Ocean Tracking Network Data Centre for data acquisition and coding help.

Author contribution statement DVN: performed field work, contributed to concept development, performed analyses, and wrote initial manuscript drafts. R.J.L.: contributed to concept development and manuscript drafts. DCH: performed field work, developed study design, and contributed to manuscript drafts. GTC: contributed to manuscript drafts.

Funding Funding provided by NSERC Strategic Partnership Grant No. 521256. R.J. Lennox was supported by the NFR project LaKES (#320726).

Data availability The data that support the findings of this study will be made publicly available on the Ocean Tracking Network database following publication of the data.

Code availability Associated R code for analyses is publicly available on GitHub: <https://github.com/danielanotte>

Declarations

Conflict of interest The authors declare no competing interests.

Ethical approval All animal experiments were approved by the Canadian Committee on Animal Care, via permits issued by Fisheries and Oceans Canada (Maritimes Region Animal Care Committee Animal Utilization Protocols 17–16, 18–13, 19–10) and by Dalhousie University (University Committee on Lab Animals permit 18–126).

Consent to participate Not applicable.

Consent for publication Not applicable.

References

- Beland KF, Kocik JF, VandeSande J, Sheehan TF (2001) Striped bass predation upon Atlantic salmon smolts in maine. *Northeast Nat* 8(3):267–274
- Bendall B, Moore A (2008) Temperature-sensing telemetry – possibilities for assessing the feeding ecology of marine mammals and their potential impacts on returning salmonid populations. *Fisheries Manag Ecol* 15:339–345
- Binder TR, Dini A (2019) *glatos*: An R package for the Great Lakes Acoustic Telemetry Observation System. R package version 0.5.1. <https://rdrr.io/github/jsta/glatos/man/glatos.html>
- Bland LM, Collen B, Orme CDL, Bielby J (2014) Predicting the conservation status of data-deficient species. *Conserv Biol* 29(1):250–259
- Bradford RG, Halfyard EA, Hayman T, LeBlanc P (2015) Overview of the 2013 Bay of Fundy striped bass biology and general status. DFO Can Sci Advis Sec Res Doc 2015/024, pp.iv + 36
- Brownscombe JW, Griffin LP, Gagne TO, Haak CR, Cooke SJ, Finn JT, Danylchuk AJ (2019) Environmental drivers of habitat use by a marine fish on a heterogeneous and dynamic reef flat. *Mar Biol*. <https://doi.org/10.1007/s00227-018-3464-2>
- Brownscombe JW, Griffin LP, Morley D, Acosta A, Hunt J, Lowerre-Barbieri SK, Adams AJ, Danylchuk AJ, Cooke SJ (2020) Application of machine learning algorithms to identify cryptic reproductive habitats using diverse information sources. *Oecologia* 194:283–298
- Buchanan RA, Skalski JR, Brandes PL, Fuller A (2013) Route use and survival of juvenile chinook salmon through the San Joaquin River Delta. *N Am J Fish Manage* 33(1):216–229
- Chen C, Liaw A, Breiman L (2004) Using random forest to learn imbalanced data. Report number 666, Journal University of California, Berkeley, vol 110, pp 1–12
- Clark JS, Carpenter SR, Barber M, Collins S, Dobson A, Foley JA, Lodge DM, Pascual M, Pielke R, Pizer W et al (2001) Ecological forecasts: an emerging imperative. *Science* 293(5530):657–660
- Cordier T, Esling P, Lejzerowicz F, Visco J, Ouadahi A, Martins C, Cedhagen T, Pawlowski J (2017) Predicting the ecological quality status of marine environments from eDNA metabarcoding data using supervised machine learning. *Environ Sci Technol* 51(16):9118–9126
- Coreau A, Pinay G, Thompson JD, Cheptou PO, Mermet L (2009) The rise of research on futures in ecology: rebalancing scenarios and predictions. *Ecol Lett* 12:1277–1286
- Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ (2007) Random forests for classification in ecology. *Ecology* 88(11):2783–2792
- Daniels J, Chaput G, Carr J (2018) Estimating consumption rate of Atlantic salmon smolts (*Salmo salar*) by striped bass (*Morone saxatilis*) in the Miramichi River estuary using acoustic telemetry. *Can J Fish Aquat Sci* 75:1811–1822

- Daniels J, Sutton S, Webber D, Carr J (2019) Extent of predation bias present in migration survival and timing of Atlantic salmon smolt (*Salmo salar*) as suggested by a novel acoustic tag. *Anim Biotelemetry*. <https://doi.org/10.1186/s40317-019-0178-2>
- DFO (2019a) Atlantic Salmon (Inner Bay of Fundy Population). <http://www.dfo-mpo.gc.ca/species-especes/profiles-profils/salmon-atl-saumon-eng.html>
- DFO (2019b) Atlantic Salmon Marine Threats Research. <https://www.bio.gc.ca/science/research-recherche/fisheries-pecheries/management-gere/smr-rfms-en.php>
- Evans AF, Hostetter NJ, Roby DD, Collis K, Lyons DE, Sandford BP, Ledgerwood RD, Sebring S (2012) Systemwide evaluation of avian predation on juvenile salmonids from the Columbia River based on recoveries of passive integrated transponder tags. *T Am Fish Soc* 141(4):975–989
- Gibson AJF, Halfyard EA, Bradford RG, Stokesbury MJW, Redden AM (2015) Effects of predation on telemetry-based survival estimates: insights from a study on endangered Atlantic salmon smolts. *Can J Fish Aquat Sci* 72:728–741
- Halfyard EA, Gibson AJF, Ruzzante DE, Stokesbury MJW, Whoriskey FG (2012) Estuarine survival and migratory behaviour of Atlantic salmon *Salmo salar* smolts. *J Fish Biol* 81:1626–1645
- Halfyard EA, Gibson AJF, Stokesbury MJW, Ruzzante DE, Whoriskey FG (2013) Correlates of estuarine survival of Atlantic salmon postsmolts from the Southern Upland, Nova Scotia, Canada. *Can J Fish Aquat Sci* 70:452–460
- Halfyard EA, Webber D, Del Papa J, Leadley T, Kessel ST, Colborne SF, Fisk AT (2017) Evaluation of an acoustic telemetry transmitter designed to identify predation events. *Methods Ecol Evol* 8:1063–1071
- Hanssen EM (2020) Novel telemetry predation sensors and mechanistic models reveal the tribulations of Atlantic salmon (*Salmo salar*) smolts migrating through lakes. MSc thesis, Department of Biological Sciences, University of Bergen, Norway, vol 1, pp 1–59
- Indira V, Vasanthakumari R, Sugumaran V (2010) Minimum sample size determination of vibration signals in machine learning approach to fault diagnosis using power analysis. *Expert Syst Appl* 37(12):8650–8658
- Jain AK (2010) Data clustering: 50 years beyond k-means. *Pattern Recogn Lett*. <https://doi.org/10.1016/j.patrec.2009.09.011>
- Kassambara A, Mundt F (2019) factoextra: Extract and Visualise the Results of Multivariate Data Analyses. R package version 1.0.6. <https://CRAN.R-project.org/package=factoextra>
- Klinard NV, Matley JK (2020) Living until proven dead: addressing mortality in acoustic telemetry research. *Rev Fish Biol Fisheries* 30:485–499
- Klinard NV, Matley JK, Fish AT, Johnson TB (2019) Long-term retention of acoustic telemetry transmitters in temperate predators revealed by predation tags implanted in wild prey fish. *J Fish Biol*. <https://doi.org/10.1111/jfb.14156>
- Klinard NV, Matley JK, Ivanova SV, Larocque SM, At F, Johnson TB (2021) Application of machine learning to identify predators of stocked fish in Lake Ontario: using acoustic telemetry predation tags to inform management. *J Fish Biol*. <https://doi.org/10.1111/jfb.14574>
- LaCroix GL (2008) Influence of origin on migration and survival of Atlantic salmon (*Salmo salar*) in the Bay of Fundy, Canada. *Can J Fish Aquat Sci* 65:2063–2079
- Lennox RJ, Nilsen CI, Nash A, Hanssen EM, Johannesen HL, Berhe S, Barlaup B, Wiik VK (2021) Laboratory and field experimental validation of two different predation sensors for instrumenting acoustic transmitters in fisheries research. *Fisheries* 46(11):565–573
- Liaw A, Wiener M (2002) randomForest: Breiman and Cutler's Random Forests for Classification and Regression. R package version 4.6–14. <https://cran.r-project.org/web/packages/randomForest/index.html>
- Maxwell AE, Warner TA, Fang F (2018) Implementation of machine-learning classification in remote sensing an applied review. *Int J Remote Sens* 39(9):2784–2817
- Moghaddam DD, Rahmati O, Panahi M, Tiefenbacher J, Darabi H, Haghizadeh A, Haghghi AT, Nalivan OA, Bui DT (2020) The effect of sample size on different machine learning models for groundwater potential mapping mountain bedrock aquifers. *CATENA* 187:104421
- Moxam EJ, Cowley PD, Bennett RH, von Brandis RG (2019) Movement and predation: a catch-and-release study on the acoustic tracking of bonefish in the Indian Ocean. *Environ Biol Fish* 102:365–381
- Olden JD, Lawler JJ, Poff NL (2008) Machine learning methods without tears: a primer for ecologists. *Q Rev Biol* 83(2):171–193
- Perry RW, Skalski JR, Brandes PL, Sandstrom PT, Klimley P, Ammann A, MacFarlane B (2010) Estimating survival and migration route probabilities of juvenile chinook salmon in the Sacramento-San Joaquin River Delta. *N Am J Fish Manage* 30:142–156
- Pincock DG (2012) False detections: what they are and how to remove them from detection data. *Amirix Document DOC-004691 Version 3*, pp 1–11
- Romine JG, Perry RW, Johnston SV, Fitzer CW, Pagliughi SW, Blake AR (2014) Identifying when tagged fishes have been consumed by piscivorous predators: application of multivariate mixture models to movement parameters of telemetered fishes. *Anim Biotelemetry*. <https://doi.org/10.1186/2050-3385-2-3>
- Schultz AA, Kumagai KK, Bridges BB (2015) Methods to evaluate gut evacuation rates and predation using acoustic telemetry in the tracy fish collection facility primary channel. *Anim Biotelemetry*. <https://doi.org/10.1186/s40317-015-0034-y>
- Simpfendorfer CA, Huveneers C, Steckenreuter A, Tattersall K, Hoener X, Harcourt R, Heupel MR (2015) Ghosts in the data: false detections in VEMCO pulse position modulation acoustic telemetry monitoring equipment. *Anim Biotelemetry* 3:55. <https://doi.org/10.1186/s40317-015-0094-z>
- Tabak MA, Norouzzadeh MS, Wolfson DW, Sweeny SJ, Vercauteren KC, Snow NP, Halseth JM, Di Salvo PA, Lewis JS, White MD et al (2018) Machine learning to classify animal species in camera trap images: applications in ecology. *Methods Ecol Evol* 10(4):585–590
- Thessen AE (2016) Adoption of machine learning techniques in ecology and earth science. *One Ecosystem*. <https://doi.org/10.3897/oneeco.1.e8621>
- Thorstad EB, Whoriskey FG, Rickardsen AH, Aarestrup K (2011) Aquatic nomads: the life and migrations of the Atlantic salmon. In: Aas O, Einum S, Klemetsen A, Skurdal J (eds) *Atlantic salmon ecology* Chichester, West Sussex, vol 1. Wiley-Blackwell, UK, pp 1–33
- Thorstad EB, Whoriskey FG, Uglem I, Moore A, Rikardsen AH, Finstad B (2012) A critical life stage of the Atlantic salmon *Salmo salar*: behaviour and survival during the smolt and initial post-smolt migration. *J Fish Biol* 81:500–542
- Weinz AA, Matley JK, Klinard NV, Fisk AT, Colborne SF (2020) Identification of predation events in wild fish using novel acoustic transmitters. *Anim Biotelemetry*. <https://doi.org/10.1186/s40317-020-00215-x>
- Whittingham H, Ashenden SK (2021) Hit discovery. In: Ashenden SK (ed) *The era of artificial intelligence, machine learning, and data science in the pharmaceutical industry*. Elsevier, Amsterdam, Netherlands, pp 81–102