

# Application of machine learning to identify predators of stocked fish in Lake Ontario: using acoustic telemetry predation tags to inform management

Natalie V. Klinard<sup>1</sup>  | Jordan K. Matley<sup>1</sup> | Silviya V. Ivanova<sup>1</sup> | Sarah M. Larocque<sup>1</sup> | Aaron T. Fisk<sup>1</sup> | Timothy B. Johnson<sup>2</sup>

<sup>1</sup>Great Lakes Institute for Environmental Research, University of Windsor, Windsor, Ontario, Canada

<sup>2</sup>Ontario Ministry of Natural Resources and Forestry, Glenora Fisheries Station, Picton, Ontario, Canada

## Correspondence

Natalie V. Klinard, Great Lakes Institute for Environmental Research, University of Windsor, 401 Sunset Ave, Windsor, ON N9B 3P4, Canada.

Email: natalie.klinard@gmail.com, natalie.klinard@dal.ca

## Present address

Natalie V. Klinard, Department of Biology, Dalhousie University, Halifax, Nova Scotia, Canada

## Funding information

Great Lakes Fishery Commission, Grant/Award Number: 2017\_JOH\_44065; Ministry of Natural Resources; Natural Sciences and Engineering Research Council of Canada; Ontario Great Lakes Protection Fund, Grant/Award Numbers: 07-46, 07-50

## Abstract

Understanding predator–prey interactions and food web dynamics is important for ecosystem-based management in aquatic environments, as they experience increasing rates of human-induced changes, such as the addition and removal of fishes. To quantify the post-stocking survival and predation of a prey fish in Lake Ontario, 48 bloater *Coregonus hoyi* were tagged with acoustic telemetry predation tags and were tracked on an array of 105 acoustic receivers from November 2018 to June 2019. Putative predators of tagged bloater were identified by comparing movement patterns of six species of salmonids (*i.e.*, predators) in Lake Ontario with the post-predated movements of bloater (*i.e.*, prey) using a random forests algorithm, a type of supervised machine learning. A total of 25 bloater (53% of all detected) were consumed by predators on average ( $\pm$  S.D.)  $3.1 \pm 2.1$  days after release. Post-predation detections of predators occurred for an average ( $\pm$  S.D.) of  $78.9 \pm 76.9$  days, providing sufficient detection data to classify movement patterns. Tagged lake trout *Salvelinus namaycush* provided the most reliable classification from behavioural predictor variables (89% success rate) and was identified as the main consumer of bloater (consumed 50%). Movement networks between predicted and tagged lake trout were significantly correlated over a 6 month period, supporting the classification of lake trout as a common bloater predator. This study demonstrated the ability of supervised learning techniques to provide greater insight into the fate of stocked fishes and predator–prey dynamics, and this technique is widely applicable to inform future stocking and other management efforts.

## KEYWORDS

bloater, *Coregonus hoyi*, predation, predator–prey interaction, random forests, survival

## 1 | INTRODUCTION

Predator–prey interactions provide fundamental linkages that define food webs and structure ecological communities. Changes in the

abundance, distribution or availability of prey (or predators) can have multiplicative impacts throughout the ecosystem with the potential to alter community dynamics at evolutionary scales (Beauchamp *et al.*, 2007). Humans have played a large part in

altering predator–prey and community relationships in aquatic environments worldwide. For instance, the addition and removal of fishes through stocking, harvesting, extirpations and introductions has reshaped pre-industrial population demographics, biodiversity and ecological interactions in marine and freshwater environments (Eby *et al.*, 2006; Rahel, 2000; Scheffer *et al.*, 2005). There is inherent value in tracking the ecological impacts of changing fish communities from fishery management and conservation perspectives. Nonetheless, it is often difficult to identify and quantify novel trophic relationships, in addition to existing ones (Christensen, 1996; Scheffer *et al.*, 2005). For example, determining which species consume stocked fish is inherently difficult to ascertain, especially if the introduced animals are not pervasive or self-sustaining.

Fish stocking is conducted worldwide to supplement fish populations for both economic and conservation purposes (Brown & Day, 2002; Molony *et al.*, 2003). An influx of stocked fishes, serving as predator or prey, can have large influences on patterns of energy flow, species abundance and distribution and both competitive and non-competitive interactions (Beauchamp *et al.*, 2007; Eby *et al.*, 2006). Predation of hatchery-reared fishes following stocking is often a principal cause of mortality and can contribute to >25% of post-stocking mortality (Howell, 1994; Karpinnen *et al.*, 2014; Thorstad *et al.*, 2012). As fish stocking practices continue to expand (Born *et al.*, 2004) and ecosystems continue to change rapidly from external factors (*e.g.*, large-scale distribution shifts due to ocean/lake warming; Cheung *et al.*, 2013; Perry *et al.*, 2005), understanding how stocking affects community dynamics and ecological interactions is essential for effective adaptive management.

Bloater *Coregonus hoyi* Milner 1874 are a deepwater forage fish that are currently undergoing reintroduction in Lake Ontario (Canada/USA). They were once widespread throughout the Laurentian Great Lakes and comprised an integral part of the native fish community (Eshenroder *et al.*, 2016). Through diel vertical migration (DVM) in the water column, bloater link deep benthic production with higher trophic-level piscivores, serving as an important energetic link within food webs (Favé & Turgeon, 2008; Ives *et al.*, 2019). Survey data and recent research suggest that bloater frequently occupy depths from near surface to 110 m (Brown *et al.*, 1985; Jobes, 1949; Klinard *et al.*, 2020; Koelz, 1929). With an average lifespan of 10 years, bloater typically reach 23 cm in length (Scott & Crossman, 1973). Although the native range of bloater encompasses all the Great Lakes except Lake Erie, local extirpations have restricted bloater to Lakes Superior and Michigan (Eshenroder *et al.*, 2016). In Lake Ontario, bloater became scarce by the 1960s, and the last documented catch occurred in 1983 (Baldwin, 1999; Owens *et al.*, 2003). According to the IUCN, the conservation status of bloater is listed as vulnerable (Gimenez, 1996).

Binational efforts to reintroduce bloater in Lake Ontario have culminated as a result of overarching objectives to restore native fish fauna to achieve a healthy lake ecosystem (Baldwin, 1999; OMNRF, 2015; Zimmerman & Krueger, 2009). The goals of these efforts are to reestablish original energetic pathways linking benthic

production and top predators, improve recruitment of piscivorous predators and increase prey fish diversity (Favé & Turgeon, 2008; Klinard *et al.*, 2020). Nonetheless, bloater and other fishes are often difficult to monitor post-stocking, which creates uncertainty surrounding their survival and new role in the food web.

Machine learning algorithms, which can be categorized into supervised and unsupervised learning, have been widely applied to animal movement data in recent decades to identify and classify behaviours (Valletta *et al.*, 2017; Wang, 2019). Unsupervised learning involves the classification of unlabelled behavioural modes (*e.g.*, clustering algorithms) and can aid in identifying commonalities among individuals based on various observed behavioural metrics. In contrast, supervised learning requires input data with pre-determined behavioural modes to train an algorithm to make predictions from new unknown input data (*e.g.*, random forests). Although machine learning in spatial ecology has primarily been applied to bilogger, remote sensing and satellite telemetry data, it is increasingly being used to analyse acoustic telemetry data sets (Brewster *et al.*, 2018; Brownscombe *et al.*, 2020; Daniels *et al.*, 2018). Acoustic telemetry is used to monitor the movements of fishes, which consists of a transmitter (hereafter tag) implanted in (or on) an animal which is detected and logged on moored receivers (Donaldson *et al.*, 2014). The near-continuous spatial and temporal monitoring of acoustic telemetry and the recent development of predation tags that can detect the occurrence of predation allow for the estimation of the post-stocking survival and predation of bloater (Halfyard *et al.*, 2017).

The predation of bloater and resulting retention of predation tags in the gastrointestinal tract of predators in Lake Ontario for  $\geq 6$  months (Klinard *et al.*, 2019a) enabled the survival of this stocked prey fish to be effectively quantified for the first time. Furthermore, tracking the movements of retained predation tags, in conjunction with acoustically tagged salmonid predators in Lake Ontario, provided a unique opportunity to use supervised machine learning techniques to assess the feasibility of predicting the species of unknown predators of bloater following stocking. The objectives of this study were to (a) evaluate and quantify the survival of bloater equipped with predation tags and (b) demonstrate the potential to classify post-predation movements of tagged bloater relative to six predator species in Lake Ontario. Quantifying aspects of survival and predation of stocked bloater would provide important information relating to the trophic role of reintroduced bloater and contribute to adaptive management and restoration approaches.

## 2 | MATERIALS AND METHODS

The care and use of experimental animals complied with the Canadian animal welfare laws, guidelines and policies as approved by the University of Windsor Animal Care Committee (AUPP 18-13) and the OMNRF (Animal Use Protocol #135).

## 2.1 | Study site and acoustic receiver array

An array of 105 acoustic receivers (69-kHz VR2W receivers, Vemco Inc., Bedford, NS, Canada) spanning c. 375 km<sup>2</sup> was deployed in June 2018 in the St. Lawrence Channel of eastern Lake Ontario (43° 55.307' N, 76° 31.715' W; Figure 1). The St. Lawrence Channel is approximately 4 km wide and 20 km long with depths between 50 and 60 m and bordered by shallower water (c. 20 m). Receivers were spaced c. 1 km apart based on the detection efficiency determined in previous studies (Klinard *et al.*, 2019b). Receiver moorings were constructed following methods described in Klinard *et al.* (2019b), and receivers were suspended with the hydrophone facing upwards c. 2 m above the lake bottom. Additional receivers were concurrently deployed throughout Lake Ontario as part of other ongoing acoustic telemetry projects through the Great Lakes Acoustic Telemetry Observation System (GLATOS) network (Figure 1) (Krueger *et al.*, 2018). The full receiver array was downloaded by 4 June 2019.

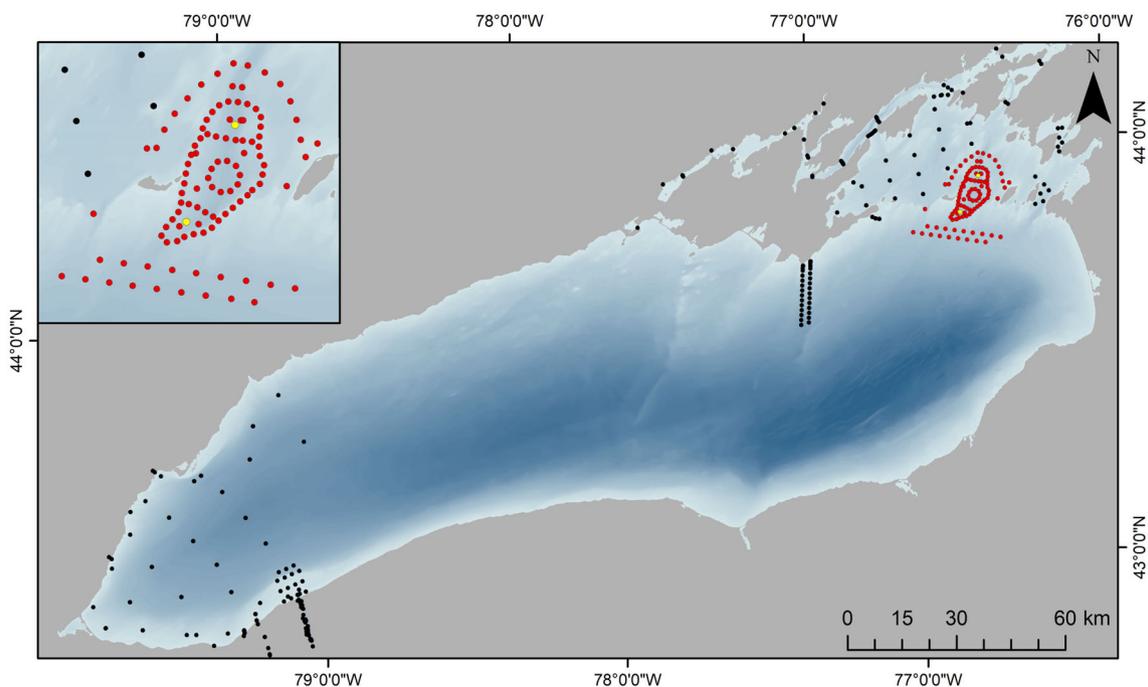
## 2.2 | Bloater tagging and stocking

A total of 48 bloater were tagged on 6 November 2018 with V9DT-2X acoustic tags (31.5 mm length × 9 mm diameter; 3.0 g weight in water; nominal delay 210 s; estimated battery life 360 days; Vemco Inc.) designed to detect predation events. All bloater were reared and tagged at the Ontario Ministry of Natural Resources and Forestry

(OMNRF) White Lake Fish Culture Station (Sharbot Lake, Ontario), following procedures described in Klinard *et al.* (2019a). Surgeries lasted c.120–150 s, and fish were monitored daily for 2 weeks following surgery. Tagged bloater ranged in mass from 61 to 119 g (mean ± S.D.: 89.7 ± 13.6) and in fork length from 172 to 215 mm (mean ± S.D.: 191.6 ± 9.3). There was negligible mortality and no tag loss prior to transportation and stocking (described in Klinard *et al.*, 2020). Bloater were released at two locations near the east and west ends of the receiver array on 19 November 2018 (Figure 1). Underwater videography during transport and release indicated no visible signs of stress and showed tagged bloater swimming to depth.

## 2.3 | Predator tagging

A total of 126 adult individuals across six top predator species in Lake Ontario were tagged from 25 October 2016 to 6 September 2019 (Supporting Information Table S1). Fish tagged included 20 hatchery-raised Atlantic salmon *Salmo salar* L. 1758, 18 brown trout *Salmo trutta* L. 1758, 31 Chinook salmon *Oncorhynchus tshawytscha* (Walbaum 1792), 16 coho salmon *Oncorhynchus kisutch* (Walbaum 1792), 17 rainbow trout *Oncorhynchus mykiss* (Walbaum 1792) and 55 lake trout *Salvelinus namaycush* (Walbaum 1792). Tagging of these individuals occurred at various locations and times, but fish were generally angled from a boat while trolling, although some lake trout were also captured by short-duration gillnet sets near spawning shoals.



**FIGURE 1** Location of acoustic receivers deployed in Lake Ontario from 19 November 2018 to 4 June 2019. Map inset features the project receiver array. Yellow circles denote release locations where bloater were stocked into the lake, red circles represent receivers deployed as part of the study and black circles signify receivers deployed through the Great Lakes Acoustic Telemetry Observation System (GLATOS) network. Note that GLATOS receivers depicted were not always present for the entire study duration but were deployed at some point during the study. (●) release location, (●) project receivers, (●) GLATOS receivers

Once captured, fish were anaesthetized using clove oil or electro-anaesthesia with fish handling gloves, and transmitters were inserted into the body cavity through a 2–3 cm mid-ventral incision. A variety of V9, V13 and V16 69 kHz transmitters (Vemco Inc.) were used. Incisions were closed with two to three interrupted sutures. Fish recovered in a cooler with recirculating water and were released as soon as they were able to maintain normal orientation, as extended holding can increase post-release mortality in salmonids (Donaldson *et al.*, 2011, 2013).

## 2.4 | Data analysis

All bloater detections were assessed for each tag to determine their fate (see Klinard *et al.*, 2019a, and Supporting Information for examples). Bloater were considered dead if there was evidence of predation indicated by the tag signal or detection patterns indicated no horizontal movement for the remainder of the detection period. If bloater stopped being detected with no indication of death in prior detection data, they were considered alive at the time of last detection with final fate unknown. If post-predation detections (*i.e.*, tag switched to a predated signal) were recorded, the bloater was presumed to have been predated and fate was designated following the change in transmission signal. Tags that had post-predation detections were further examined to evaluate the potential occurrence of false positives (*i.e.*, the tag switching from a pre- to post-predation transmission without predation). If the detections exhibited patterns characteristic of a dead fish (*i.e.*, constant detection at the same location for the duration of the study and no change in movement between pre- and post-predation detections) prior to the transmission signal changing, the switch was presumed to be the result of a false positive, and suspect detections were considered to be from a dead bloater. Detections were considered to originate from a dropped tag that was egested by a predator if the detections were from a valid post-predation signal and suggested a sudden cessation of movement (*i.e.*, stationary horizontal movement). For all bloater that were considered predated, the time at which the tag triggered (*i.e.*, switched to a post-predation transmission) was calculated from tag sensor data that transmits the number of hours since triggering. Time since triggering is recorded at varying resolution (1–16 h) that decreases with the passage of time and is presented as a range (minimum to maximum number of hours) based on the resolution. The minimum time since triggering was used to calculate the time when tags triggered (*i.e.*, bloater were consumed/digested) as a conservative approach to ensure that all tags had switched at the reported time. Similarly, all detections of tagged salmonid predators were assessed on an individual basis to remove fish that appeared dead. For all tagged predator and post-predation bloater detections, individuals that had limited detection data (*i.e.*,  $\leq 2$  unique days with detections and  $< 5$  km between detections) were removed from further analyses. All analyses were completed in R version 3.6.1 (R Development Core Team, 2019), and statistical significance was assumed at  $\alpha = 0.05$ .

Ten predictor variables were selected to characterize the movement patterns of predators with the goal of selecting variables that could lead to distinguishing species. These included (a) proportion of days detected during the detection period, (b) maximum range of distance moved, (c) average distance detected from the release site, (d) S.D. of distance detected from the release site, (e) average time between all detections, (f) S.D. of time between detections, (g) count of switches between the east and west sides of the lake, (h) average distance moved between consecutive days, (i) proportion of detection days that were consecutive and (j) average number of days between daily detections. The distance was calculated using the release location of the tagged bloater as a point of reference. To reduce biases that may be associated with the location of tagging and release, the average and S.D. of the distance detected were represented as a proportion of the range of distance moved for individual fishes.

A random forests model was used to differentiate between the movement patterns of the six species based on the 10 predictors using the “randomForest” R package (Liaw & Wiener, 2002). Random forests construct a series of decision trees that operate as an ensemble by combining outputs from individual trees to improve the prediction of the response values (Breiman *et al.*, 1984). Each tree uses a random sample of the data and a set of randomly selected predictors to choose optimal splits of the trees. To train, tune and evaluate the random forests model, the tagged predator data (*i.e.*, 10 predictor variables) were split into 70% as a training data set and 30% as a test data set. The generalization ability of the model was optimized by determining the optimal number of predictor variables (*i.e.*, *mtry*), maximum nodes and trees using the “caret” R package (Kuhn *et al.*, 2019). Ten-fold cross-validation was used to resample the data, fit random forests and estimate the mean of the Kappa statistic to select the best parameter values. The Kappa statistic is a numerical rating that represents the precision that would be expected by chance (*i.e.*, accounts for imbalanced class sizes; Viera & Garrett, 2005). Kappa values range from  $-1$  to  $1$ , where negative values indicate agreement less than chance,  $0$  is exactly what would be expected by chance and  $1$  is perfect agreement. The best *mtry*, maximum nodes and number of trees were used to refit the final random forests. The test data set then determined the mean prediction accuracy of the model. Variable importance was assessed using the mean decrease in accuracy, which is the mean decrease in prediction accuracy of the model as determined by removing the variable from the model (Han *et al.*, 2016). Finally, the random forests model was applied to the post-predation detections of 20 consumed bloater with sufficient data (*i.e.*,  $> 2$  days of detections needed to calculate variables) to predict the probability that the detections represent the movement patterns of each of the six predator species (*i.e.*, probability of all six classes sum to one for each individual). All available predator detection data from 2016 onwards were opted for use as opposed to selecting only data from the post-predated bloater detection period (*i.e.*, 21 November 2018 to 4 June 2019), despite potential seasonal or alternate patterns, because it allowed for more structure (*i.e.*, patterns) and ranges in behaviour to be identified in the data.

Post-predation detection data using random forests predicted that 10 bloater were consumed by lake trout (see “Results”). Because most post-predation detections were predicted to be lake trout, these data were compared with known lake trout using movement networks (Dale & Fortin, 2010) to evaluate model designations and actual space-use patterns. To account for the large detection range creating simultaneous detections at various receivers (Klinard *et al.*, 2019b), the detections were grouped into 30 min intervals, and the most frequent receiver was selected as the detection location for each interval. A 30 min grouping interval was selected based on a possible minimum 10 detections recorded in 30 min (*i.e.*, predator tags and predation tags had a nominal delay of 90 or 180 and 60 s, respectively). Detection data from 38 tagged lake trout (filtered to the same time period that predated bloater were detected; 21 November 2018 to 4 June 2019) were selected and similarly grouped into 30 min intervals. The selected detection data for predated bloater and tagged lake trout were divided into three time periods to examine the changes in movement over shorter time periods: (a) 21 November 2018 to 31 January 2019; (b) 1 February 2019 to 31 March 2019; and (c) 1 April 2019 to 4 June 2019.

For each of the six groupings of data (*i.e.*, three time periods each for predated bloater data and tagged lake trout data), an inter-receiver movement matrix was created for each individual. The movement matrices counted the relative presence at, and movements between, receivers. Relative movements were defined as the number of times the individual moved between two receivers divided by the total number of movements within its space use (*i.e.*, total edges in the network). Relative presence refers to the number of incoming and outgoing movements of an individual (*i.e.*, edges) at a receiver (*i.e.*, node). Relative metrics were used to reduce biases in the frequency and amount of detections associated with the time and location of post-predation tag expulsion by the predator. The matrices for all individuals within a group were summed and divided by the total number of individuals in the group to standardize for variable numbers of individuals being detected due to tag expulsion and a large number ( $n = 38$ ) of tagged lake trout. The final matrices were used to create standardized weighted directed networks for each group that reflected space use within the array during the respective time period using the “igraph” R package (Csárdi & Nepusz, 2006). To assess the correlation between networks of predated fish and tagged lake trout, a Mantel test was performed on the two matrices for each time period using the “vegan” R package (Oksanen *et al.*, 2019).

### 3 | RESULTS

#### 3.1 | Detection summary and survival

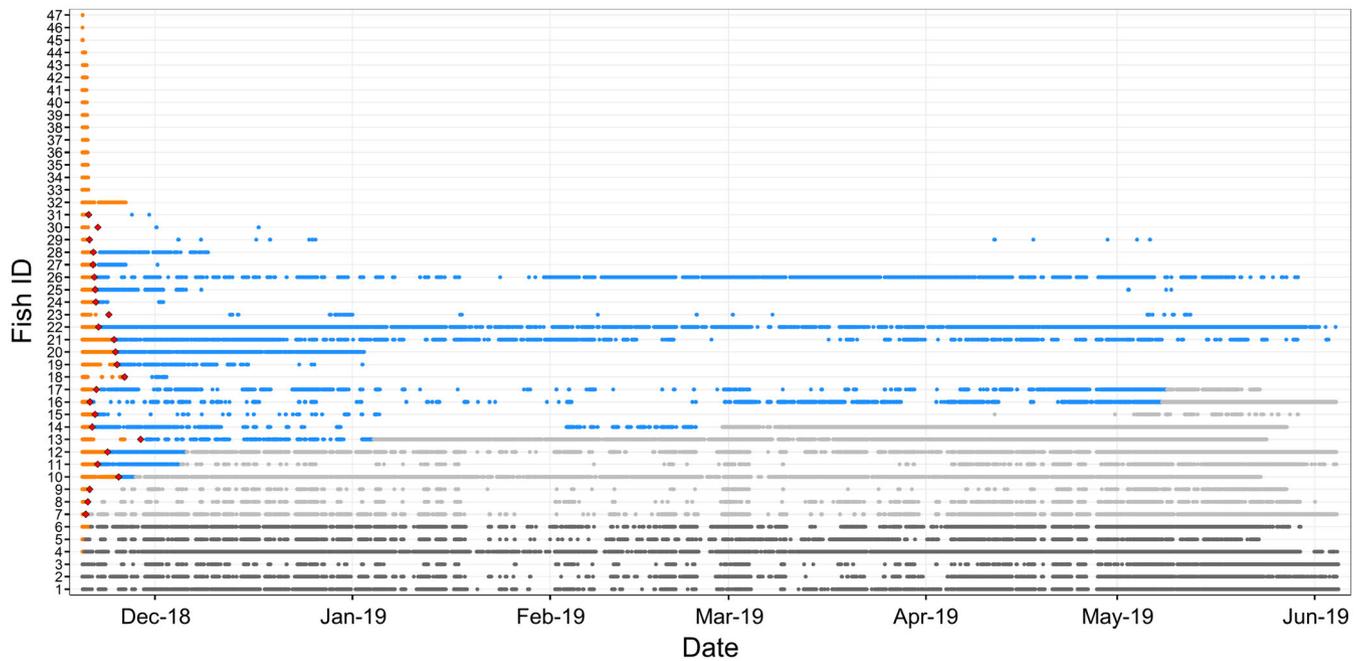
Of the 48 tagged bloater, 47 had recorded detections following their release into eastern Lake Ontario on 19 November 2019 until the array was downloaded on 4 June 2019. Over the course of the 6 month detection period, 30 tags (64% of all detected) had recorded

post-predation detections, suggesting consumption by a predator. A total of 17 tags had only pre-predation detections (34% of all detected), indicating the fish were not consumed by a predator during their detection period. Of the 30 tags that exhibited a switch to the post-predation transmission, five tags appeared to be false positives (IDs 2–6; Supporting Information Figure S1) based on delayed switches to post-predation identification (mean  $\pm$  S.D.: of  $94.6 \pm 51.4$  days) following stationary detections, resulting in 25 predation events (53% of all detected; IDs 7–31; Figures 2 and 3). Note that three of these predation events (IDs 7–9; Figure 2) were inferred, as all post-predation detections indicated no movement and likely already passed through the predator’s gastrointestinal tract (*i.e.*, detections were considered “egested” instead of “predated”).

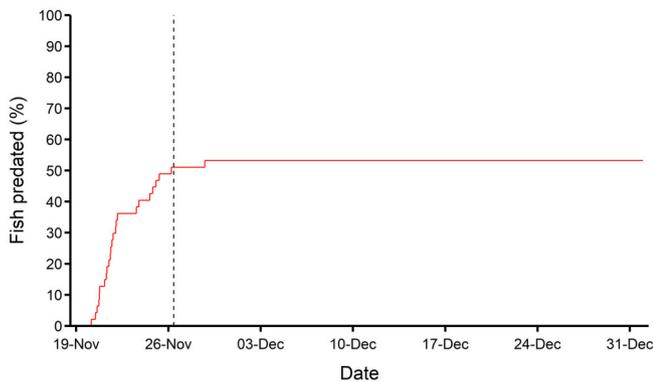
Based on the time at which the tags triggered, the maximum survival of the 25 bloater prior to predation was  $3.1 \pm 2.1$  (mean  $\pm$  S.D.) days (range: 0.7–9.3 days) with detections up to  $9.0 \pm 8.8$  km (mean  $\pm$  S.D.) from the release site (Table 1; Figure 3; Supporting Information Figure S2). Similarly, detection of the 16 tags that were never predated (*i.e.*, had only pre-predation detections) and had no other obvious indication of death occurred for  $1.0 \pm 1.6$  (mean  $\pm$  S.D.) days (range: 0–7 days; IDs 32–47; Figure 2; Table 1) with detections up to  $5.1 \pm 1.9$  km (mean  $\pm$  S.D.) from release. Nonetheless, since these 16 bloater were only detected for up to 7 days following release and then never detected again, it is difficult to determine their fate. Six bloater appeared to die rapidly after release but without being predated based on constant detection at the same location for the rest of the study duration and had detections for only  $0.4 \pm 0.5$  (mean  $\pm$  S.D.) days prior to death (range: 0–2 days; IDs 1–6; Figure 2; Table 1).

The first post-predation detections of all consumed bloater ( $n = 25$ ) were recorded on average ( $\pm$  S.D.)  $5.6 \pm 5.3$  days after release and within 22 days following release (Supporting Information Figure S2). Due to the time it takes for the transmission signal of the tag to switch from pre-predation to post-predation in the gut of the predator (*i.e.*, signal lag), the precise time and location of predation of tagged bloater cannot be known. Preliminary results from laboratory predation tag trials involving tagged bloater fed to lake trout in a hatchery (10–12°C) indicated a mean signal lag of 77 h (range 24–126 h; unpubl. data). Nonetheless, based on the change in transmission signal and tag sensor data, it can be confidently concluded that at the minimum time since triggering, the tag is inside the predator. Because signal lag is unknown and may vary by species and individual, it is possible that the tagged bloater were already consumed during the final pre-predation detections. As such, the first post-predation detections provide a conservative estimate of when predation occurred, but the final pre-predation detections may be more indicative of the timing of predation. There were no evident patterns in the location of the final pre-predation detections and first post-predation detections (Supporting Information Figure S2).

Post-predation detections of predators occurred for an average ( $\pm$  S.D.) of  $78.9 \pm 76.9$  days before the tag was expelled or the predator ceased to be detected (Figure 2). Eleven predators were considered to



**FIGURE 2** Summary of detections for all bloater (*Coregonus hoyi*) that were detected ( $n = 47$ ) following release into Lake Ontario from 19 November 2019 to 4 June 2019. The colour of the circles denotes fate at the time of detection, and red diamond indicates when the predation sensor triggered. (●) bloater, (●) predator, (●) excreted, (●) dead, (◆) tag triggered



**FIGURE 3** Percentage of tagged and detected bloater (*Coregonus hoyi*;  $n = 47$ ) predated over the first 6 weeks following release into Lake Ontario. The red line represents the time of predation based on the minimum hours since the tag triggered. The dotted line indicates the time of the last detection of all non-predated ( $n = 16$ ) tagged bloater

have expelled the tag based on sudden and constant detections at the same location for the duration of the study. The time that the tags remained in the predators (i.e., tag retention) ranged from <1 day to >6 months (see Klinard *et al.*, 2019a). Predators that consumed tagged bloater were detected on several receiver arrays deployed throughout Lake Ontario through the GLATOS network. The predators exhibited lake-wide movements up to 242.4 km from the release location in northeastern Lake Ontario with average maximum detection distances of  $31.5 \pm 49.2$  km (mean  $\pm$  S.D.; Table 1).

### 3.2 | Predator classification

The final predator detection data set consisted of 3,511,149 detections from five hatchery-raised Atlantic salmon, 16 brown trout, 20 Chinook salmon, 13 coho salmon, 55 lake trout and 12 rainbow trout from 25 October 2016 to 6 September 2019 (Supporting Information Table S1). Pearson correlation coefficients ( $cc$ ) calculated for each pair of predictor variables revealed high correlation of average days between daily detections and S.D. of time between detections ( $cc = 0.88$ ) as well as the range of distance moved and the number of switches between sides of the lake ( $cc = 0.83$ ). All variables were included in the analyses due to the ability of random forests machine learning algorithms to handle correlated variables well because of sampling and bagging (Breiman, 2001). The optimal random forests model selected seven variables at each tree split, had a maximum node value of 6 and consisted of 250 trees. The out-of-bag prediction error for the model was 44.1%, and classification errors ranged from 0.114 to 1.0 (Table 2). The random forests had the highest success rate for predicting lake trout at 89%, followed by success rates of 77% and 50% for Chinook salmon and brown trout, respectively (Table 2). The range of distance moved was the strongest predictor of species, followed by average distance moved between consecutive days (Figure 4a). When applied to the nontraining data set, the mean prediction accuracy of the random forests was 0.649 (95% C.I.: 0.475, 0.798). Of the 20 individuals that had a sufficient amount of post-predation detections following consumption of a tagged bloater, the optimal random forests model predicted that 50% ( $n = 10$ ) were lake trout, 40% ( $n = 8$ ) were brown trout, 5% ( $n = 1$ )

were Chinook salmon and 5% ( $n = 1$ ) were Atlantic salmon (Figure 4b; Table 3).

A total of 38 tagged lake trout had sufficient data from the period of 21 November 2018 to 4 June 2019 that was used to build movement matrices. One network was created for each time period for the post-predation data and tagged lake trout data resulting in a total of

six standardized weighted directed networks (Figure 5). Mantel tests between the networks in time periods 1, 2 and 3 revealed significant correlation in each time period ( $r_{t1} = 0.124$ ,  $r_{t2} = 0.125$ ,  $r_{t3} = 0.393$  at  $P < 0.001$ ). Predicted lake trout and tagged lake trout both exhibited a decrease in network size and extent from the first to second time periods followed by an increase from the second to third time periods

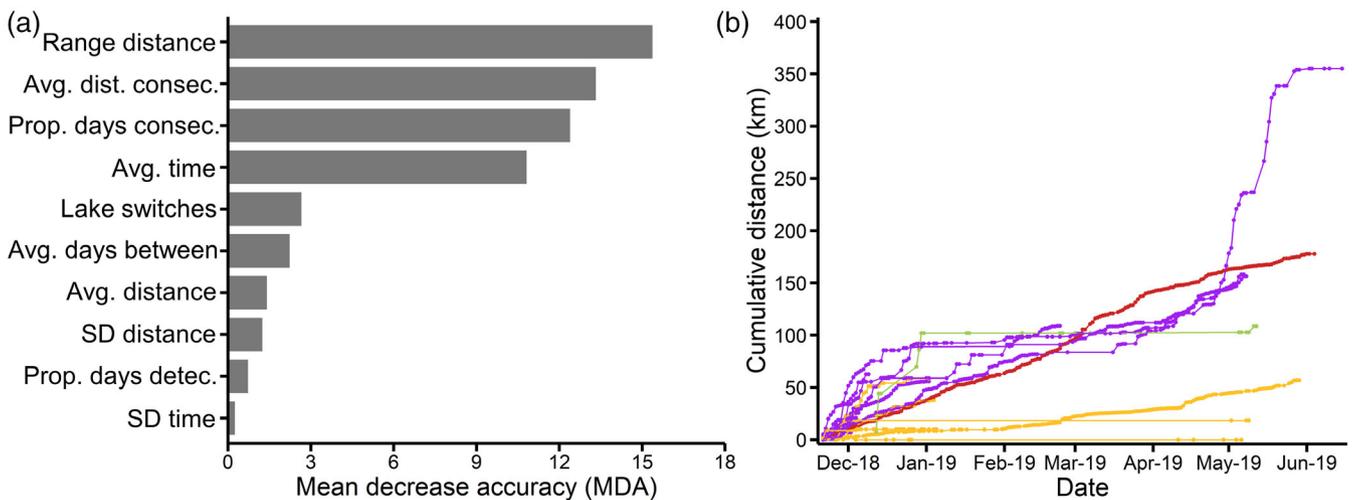
**TABLE 1** Summary of detection data for all detected bloater (*Coregonus hoyi*;  $n = 47$ )

Detection category	n	ID	First detection (days)	Detection period (days)	Max. dist. (km)	Last dist. (km)
Bloater never predated	16	32-47	-	1.0 ± 1.6	5.1 ± 1.9	3.1 ± 2.0
Bloater pre-death	3	1-6	-	0.8 ± 0.6	5.9 ± 2.1	3.1 ± 2.9
Bloater pre-predation	25	7-31	-	2.5 ± 2.0	9.0 ± 8.8	6.3 ± 9.6
Post-predation	22	10-31	5.6 ± 5.3	78.9 ± 76.9	31.5 ± 49.2	24.6 ± 45.2

Note: Detection categories include bloater that were never predated, bloater that were alive at release and then died, bloater prior to predation and post-predation detections. ID refers to the fish IDs (see Figure 2). First detection is the time from release until the first detection, detection period is the time from first to last detection, max. distance is the maximum distance from release detected and last dist. is distance from release of the final detection. Mean ± S.D. is presented for first detection, days detected, max. dist. and last dist.

**TABLE 2** Confusion matrix of optimal random forests model using the training data set

Predicted class	Actual class						Class error (%)
	Atlantic salmon	Brown trout	Chinook salmon	Coho salmon	Lake trout	Rainbow trout	
Atlantic salmon	0	1	1	0	2	0	100.0
Brown trout	0	6	1	1	4	0	50.0
Chinook salmon	0	0	10	2	1	0	23.1
Coho salmon	0	2	6	0	4	0	100.0
Lake trout	0	1	1	2	31	0	11.4
Rainbow trout	0	0	5	3	0	0	100.0



**FIGURE 4** Variable importance results from the random forests model: (a) mean decrease in accuracy for all variables used; and (b) cumulative distance moved on consecutive detection days for all post-predation fish included in the model. Variables in (a) are maximum range of distance moved, average distance moved between consecutive days, proportion of detection days that were consecutive, average time between all detections, count of switched between the east and west sides of the lake, average number of days between daily detections, average distance detected from the release site, S.D. of distance detected from the release site, proportion of days detected during detection period and S.D. of time between detections. AS is Atlantic salmon (*Salmo salar*), BT is brown trout (*Salmo trutta*), CS is Chinook salmon (*Oncorhynchus tshawytscha*) and LT is lake trout (*Salvelinus namaycush*) (—●—) AS, (—●—) BT, (—●—) CS, (—●—) LT

**TABLE 3** Random forests prediction of species classes on post-predation detection data

Fish ID	Species class						Final class
	Atlantic salmon	Brown trout	Chinook salmon	Coho salmon	Lake trout	Rainbow trout	
10	0.148	<b>0.376</b>	0.020	0.196	0.244	0.016	BT
11	0.244	0.316	0.000	0.116	<b>0.324</b>	0.000	LT
12	0.236	0.308	0.000	0.116	<b>0.336</b>	0.004	LT
13	0.208	<b>0.488</b>	0.012	0.072	0.204	0.016	BT
14	0.012	0.228	0.008	0.016	<b>0.736</b>	0.000	LT
15	0.008	<b>0.676</b>	0.036	0.152	0.124	0.004	BT
16	0.016	0.120	0.036	0.084	<b>0.728</b>	0.016	LT
17	0.024	0.188	0.000	0.096	<b>0.688</b>	0.004	LT
18	0.004	<b>0.672</b>	0.016	0.180	0.116	0.012	BT
19	0.056	<b>0.440</b>	0.012	0.104	0.380	0.008	BT
20	0.236	0.292	0.000	0.036	<b>0.432</b>	0.004	LT
21	0.028	0.020	0.004	0.024	<b>0.924</b>	0.000	LT
22	<b>0.324</b>	0.316	0.000	0.096	0.264	0.000	AS
23	0.000	0.032	<b>0.652</b>	0.116	0.020	0.180	CS
24	0.012	0.092	0.092	0.292	<b>0.416</b>	0.096	LT
25	0.024	<b>0.628</b>	0.040	0.068	0.236	0.004	BT
26	0.164	<b>0.404</b>	0.000	0.076	0.352	0.004	BT
27	0.044	0.308	0.020	0.200	<b>0.420</b>	0.008	LT
28	0.244	0.320	0.000	0.092	<b>0.340</b>	0.004	LT
29	0.004	<b>0.760</b>	0.016	0.160	0.056	0.004	BT

Note: Highest classification value is bolded for each individual, and final class indicates the respective predicted species class; AS is Atlantic salmon (*Salmo salar*), BT is brown trout (*Salmo trutta*), CS is Chinook salmon (*Oncorhynchus tshawytscha*), LT is lake trout (*Salvelinus namaycush*) and RT is rainbow trout (*Oncorhynchus mykiss*).

(Figure 5). Tagged lake trout were detected on more receivers north-west of the main array than the predicted lake trout across all time periods. As a result, tagged lake trout generally had smaller node sizes and edge densities within the main array. Overall, predicted and tagged lake trout networks displayed similar spatial and temporal changes.

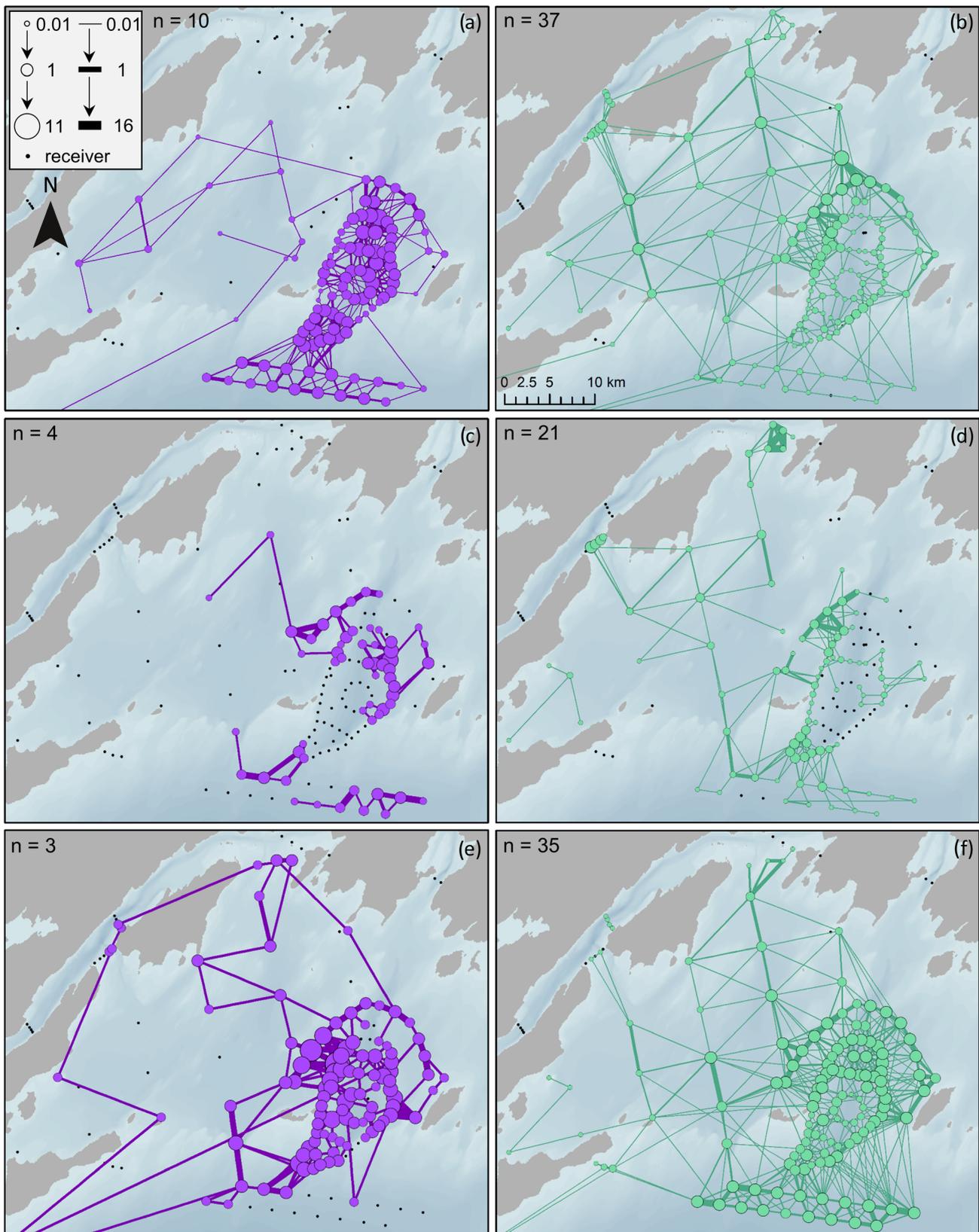
## 4 | DISCUSSION

The post-stocking survival and predation of a hatchery-reared prey fish species undergoing reintroduction in Lake Ontario were inferred, and how novel predation tags and a supervised machine learning algorithm could be used to analyse and predict patterns observed in acoustic telemetry data was demonstrated. Approximately 53% of detected bloater were consumed by predators shortly after release, and nearly all other individuals either died, also shortly after release, or were only detected for a few days with unknown fate. An advanced methodology classifying post-predation detections of tagged bloater into six potential predator species was also tested, demonstrating the potential for predation tags and supervised machine learning algorithms to predict species that are consuming these prey fish. Lake trout had the strongest predictive capabilities in

the optimized random forests classifying predator species, contributing to c. 50% of predated bloater. The identification of predation as a primary source of post-stocking mortality, especially over short time-scales and given the increasingly rapid rate of change in ecosystems, is relevant for the adaptive management of ongoing reintroduction efforts for this and other deepwater fishes. Furthermore, predictive analytical approaches, as demonstrated in this study, provide novel ways to interpret the behaviour and ecology of fishes for applied management.

### 4.1 | Ecology and management

The survival of stocked bloater during the study period was  $\leq 34\%$  based on individuals that were not predated and did not appear to die during detection. However, these 16 bloater were detected for  $< 1$  week and without further detection; therefore, it is difficult to assess their final fate. Although some bloater seem to have perished outside the receiver detection ranges shortly after release (Klinard *et al.*, 2020), some individuals may leave the spatial extent of the receiver array and survive undetected elsewhere in the lake. Given the limited receiver coverage throughout Lake Ontario, especially in deep areas, this is entirely possible. A prior tag effects study (within



**FIGURE 5** Summed networks of post-predation detections from individuals that were predicted to be lake trout (*Salvelinus namaycush*;  $n = 10$ ) and detections of tagged lake trout ( $n = 38$ ) in northeastern Lake Ontario. Networks were created for three time periods: (1) 21 November 2018 to 31 January 2019; (2) 1 February 2019 to 31 March 2019; and (3) 1 April 2019 to 4 June 2019. Networks shown are as follows: (a) predicted lake trout ( $n = 10$ ) and (b) tagged lake trout ( $n = 37$ ) for period 1; (c) predicted lake trout ( $n = 4$ ) and (d) tagged lake trout ( $n = 21$ ) for period 2; and (e) predicted lake trout ( $n = 3$ ) and (f) tagged lake trout ( $n = 35$ ) for period 3. Circles represent nodes, and the circle size corresponds to node degree. Lines indicate edges, and the line thickness corresponds to edge density

the holding facility) that revealed negligible mortality of bloater 6 months after tagging (Klinard *et al.*, 2018) leads us to believe the non-predated mortality in the current study is related to stress associated with stocking or difficulty adjusting to the wild (Klinard *et al.*, 2020). Despite being a deepwater species, these fish are reared in a shallow tank in a hatchery setting and are not exposed to the depths that they would experience in the wild. Similarly, these bloater have never been exposed to predators, new habitats or other aspects of the wild that may result in difficulty acclimatizing. Although some stocked bloater have exhibited regular DVMs and escaped initial predation (Klinard *et al.*, 2020), they may not be fully equipped to handle the transition from hatchery to wild. Hatchery-rearing techniques continue to improve and produce more fishes for stocking, yet post-stocking survival can be highly variable across species and location and low survival (<5%) is not uncommon (Brown & Day, 2002). As technology develops, the ability to track the movements of bloater and other stocked fishes and address survival, either directly (*e.g.*, predation tags) or indirectly (*e.g.*, lack of movement), is an important tool for management, as it provides individual-level assessment not necessarily possible with other approaches (Dudgeon *et al.*, 2015).

The relatively high predation of bloater (53% of tagged bloater detected) by a diversity of predators indicates that bloater could contribute as intended to a diversified forage fish prey base and ecosystem rehabilitation efforts in Lake Ontario. Post-stocking predation of hatchery-reared fishes is a principal cause of mortality and is expected for a stocked prey fish (Howell, 1994; Karpinnen *et al.*, 2014; Thorstad *et al.*, 2012). For example, Thorstad *et al.* (2012) observed rapid mortality of 37% of hatchery-reared Atlantic salmon smolts following release into a river, with at least 25% of mortality because of predation from marine fishes. Similarly, Karam *et al.* (2008) examined the post-stocking mortality of razorback sucker (*Xyrauchen texanus*) stocked in a lake and found a 2 week survival of 75% and 6 month survival of 16%, with the main source of mortality believed to be predation. The main goals of binational bloater stocking efforts are to achieve a self-sustaining population that will diversify the offshore prey fish community, reestablish historic ecological structure and function and increase ecological integrity and resilience (Connerton & Stewart, 2015). As discussed earlier, current stocking practices may need to be reevaluated, especially to reduce such high predation of bloater immediately following release. Other approaches, such as staggering releases either spatially or temporally, pre-conditioning bloater to stressors (*e.g.*, predators) during rearing (Brown *et al.*, 2003; D'Anna *et al.*, 2012; Hutchinson *et al.*, 2012) or releasing more (or larger) bloater each stocking event (Grausgruber & Weber, 2020), may ameliorate post-release survival and remain to be tested. Considering the past and ongoing rehabilitation efforts for stocking of lake trout throughout the Great Lakes and particularly in Lake Ontario (Muir *et al.*, 2012; Stewart *et al.*, 2017), it is unlikely that changes to predator abundance will mitigate bloater predation. Therefore, moving forward, the short-term goals of bloater stocking should prioritize reducing predation to help bolster the long-term goals, such as establishing a self-sustaining population.

Tagged lake trout provided the most reliable classification based on behavioural predictor variables (*i.e.*, 89% success rate) and was identified as the main consumer of bloater (*i.e.*, 50%) during this study. Lake trout and Chinook salmon make up the bulk (70%–80%) of prey consumption in Lake Ontario (Bence *et al.*, 2008; Gaden *et al.*, 2019). As a species native to Lake Ontario, lake trout diet has historically consisted of ciscoes (*Coregonus* spp.) and deepwater sculpin *Myoxocephalus quadricornis* (Girard 1851; Christie *et al.*, 1987). Following the collapse of cisco and deepwater sculpin populations in the 1940s and 1950s, lake trout shifted to consumption of primarily alewife *Alosa pseudoharengus* Mitchill 1815 and rainbow smelt *Osmerus mordax* Mitchill 1814 (Rand & Stewart, 1998). More recently, lake trout consumption has included the invasive round goby *Neogobius melanostomus* (Pallas 1814; Colborne *et al.*, 2016; Dietrich *et al.*, 2006; Rush *et al.*, 2012). Lake trout have been stocked in Lake Ontario since the 1950s, but one of the primary impediments to lake trout restoration is the quantity and quality of prey, which hinders the growth and survival of lake trout at all life stages, as well as their reproductive success (Stewart *et al.*, 2017). Consumption of alewife, a pelagic forage fish, can lead to thiamine deficiency and reduced reproductive success in lake trout and thus raises significant concerns for sustaining the current population (Fitzsimons *et al.*, 2009; Lantry *et al.*, 2015). Bloater likely would provide a better prey alternative than alewife because they efficiently transfer energy between the benthic and pelagic food webs, do not present the same health issues caused by a diet containing alewife and would help shift lake trout feeding ecology back towards historical origins (Connerton & Stewart, 2015; Ives *et al.*, 2019; Zimmerman & Krueger, 2009).

## 4.2 | Machine learning performance and limitations

Network analysis revealed significant correlation between networks during all three time periods, suggesting that similarities existed in the space use of predicted lake trout and tagged lake trout. Movement patterns of predicted lake trout and tagged lake trout exhibited the highest degree of correlation during April and May with smaller but significant correlation from November to March. Low to intermediate correlation between networks may be a result of tag expulsion occurring during the time periods for predicted lake trout. Furthermore, tagged lake trout were equipped with large tags that have higher detection range (Klinard *et al.*, 2019b), which could result in inflated network sizes. Nonetheless, networks followed similar changes in location and size through time, and the classification error for lake trout was low, supporting the idea that post-predation detections were correctly identified as lake trout.

Predictive capacity of the random forests model was less reliable for species other than lake trout, highlighting potential biases concerning the analysis used in this study. The out-of-bag error of the model, which indicates the overall prediction error rate, was 44.1%. Although the classification errors for lake trout, Chinook salmon and brown trout were low (11.4%, 23.1% and 50%, respectively), the

remaining species exhibited high classification errors of 100%, reducing the overall predictive capacity of the model. The poor ability of the random forests to predict predation by Atlantic salmon, Coho salmon and rainbow trout is likely a result of a limited number of tagged individuals and limited structure and variability in the data due to receiver coverage. More detection data for certain species would allow the model to be trained using a wider range of observed data with more variability, thus leading to more robust estimates. Similarly, classes in the training data set were unbalanced due to limited available tagging data; nevertheless, random forests are robust to overfitting, and decision trees frequently perform well on imbalanced data (Breiman, 2001).

Other aspects of the study design and inherent limitations associated with acoustic telemetry may have also affected model performance. The imbalanced receiver coverage in Lake Ontario limited the ability to continuously track tagged fishes and identify more distinct movement patterns between species. Tagging location may have also influenced results, as fishes tagged in or near dense receiver arrays would have an increased frequency of detections. For example, tagged lake trout were captured and released in and around the primary study array, which may have led to more direct comparisons with predated bloater because of similar array configuration availability. Conversely, brown trout, a comparatively short-ranging predator (Raby *et al.*, 2017), were generally tagged on the western end of Lake Ontario and did not directly overlap with the main study array; therefore, their movements may not have been as effectively tracked as consumed bloater. Furthermore, variation in behaviour or environmental preferences within a species (*e.g.*, spawning differences between sexes or ontogenetic shifts; Raby *et al.*, 2020) could also lead to difficulty classifying the species movement patterns. Although this study incorporated telemetry data from six predator species in Lake Ontario, it did not encompass all potential predators (*e.g.*, walleye *Sander vitreus*) in the lake, which likely increases the probability of incorrectly classifying predator species. These deficiencies in data availability were unavoidable. Therefore, this study is used as an opportunity to demonstrate the application of random forests to acoustic telemetry data to identify predator-prey interactions and provide preliminary work into identifying consumers of stocked bloater.

Previous studies have used a few similar methods to identify and estimate the occurrence of predation events using acoustic telemetry prior to the advent of predation tags. Several studies have used unsupervised learning algorithms (*i.e.*, clustering) to hierarchically divide data into groups based on similarity but were unable to validate prey classifications due to a lack of baseline movement data (Gibson *et al.*, 2015; Romine *et al.*, 2014). The unsupervised learning methods implemented in several telemetry studies are also limited in that they are aimed at differentiating tagged fish from predated fish but cannot classify predator species. Daniels *et al.* (2018) conducted a study using a random forests model to identify acoustically tagged Atlantic salmon smolts that may have been consumed by striped bass *Morone saxatilis* Walbaum 1792 in the Miramichi River. The optimal random forests in Daniels *et al.* (2018) had a low misclassification error rate of 1.6%;

nonetheless, there was more distinction between the movement patterns of predator and prey species than in the present study. Furthermore, a binary classification system was used that did not include alternate predatory species present in the system, which could lead to additional misclassification. Overall, as telemetry technology advances, this study and others that implement machine learning techniques provide a framework for identifying and classifying predation and behavioural patterns.

## 5 | CONCLUSION

Predation of tagged bloater by lake trout, brown trout or other salmonid species in Lake Ontario indicates that stocked bloater have the potential to contribute to rehabilitation efforts through reestablishing historic energetic pathways and increasing prey fish diversity. Although the proportion of bloater that currently survive stocking may not be sufficient to create a self-sustaining population, identifying the drivers of post-stocking mortality provides valuable information that can be applied to improve rearing and stocking practices. Furthermore, understanding the role of bloater as a prey species to salmonid predators allows one to determine how they can contribute to the improved recruitment of these species. As receiver coverage in Lake Ontario continues to expand, more opportunities will arise to build upon the existing comprehensive data set and further evaluate bloater survival and predation, which is important for evaluating stocking practices and maintaining effective management. This study also demonstrated the potential and value of classifying predators of stocked fish to help understand food web dynamics and ecosystem functioning. These techniques could also be applied to studies that do not use predation tags given that there is a collection of definitive detection data of the prey species. Random forests are best suited for a study system with comprehensive receiver coverage, a manageable number of potential predators, detection data for all possible predators and data of predators and prey for the same or comparable time periods. Overall, there is a large potential for future studies to utilize acoustic telemetry and supervised machine learning algorithms to uncover dynamic food web relationships and provide information valuable to management efforts.

## ACKNOWLEDGEMENTS

We thank T. Drew, G. Bluett and the Ontario Ministry of Natural Resources and Forestry (OMNRF) White Lake Fish Culture Station staff for their support and A. Rupnik, M. Hanley and B. Metcalfe for their assistance during surgery and monitoring. We also thank J. Chicoine, T. Dale, B. Perry and T. Schulz of the *Ontario Explorer* vessel crew. This paper is contribution 85 of the Great Lakes Acoustic Telemetry Observation System (GLATOS).

## AUTHOR CONTRIBUTIONS

N.V.K. and J.K.M. conceived the ideas and designed the methodology, all authors collected the data, N.V.K. and J.K.M. analysed the data and

N.V.K. led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

## ORCID

Natalie V. Klinard  <https://orcid.org/0000-0001-7809-9360>

## REFERENCES

- Baldwin, B. (1999). Native prey fish re-introduction into Lake Ontario: Bloater (*Coregonus hoyi*). Lake Ontario Technical Committee Discussion Paper. London, Ontario, Canada: Great Lakes Fishery Commission.
- Beauchamp, D. A., Wahl, D., & Johnson, B. M. (2007). Predator-prey interactions. In C. S. Guy & M. L. Brown (Eds.), *Analysis and interpretation of freshwater fisheries data* (pp. 765–842). Bethesda, MD: American Fisheries Society.
- Bence, J. R., Dobiesz, N. E., Madenjian, C. P., Argyle, R., Barbiero, R., Bowlby, J. N., Claramunt, R. M., O’Gorman, R., & Schaner, T. (2008). *Top-down effects of open-water salmonine predators in the Great Lakes* (Quantitative Fisheries Center Technical Report T2009-07). East Lansing, MI: Michigan State University.
- Born, A. F., Immink, A. J., & Bartley, D. M. (2004). Marine and coastal stocking: Global status and information needs. In D. M. Bartley & K. M. Leber (Eds.), *Marine ranching* (pp. 1–12). Rome, Italy: Food and Agriculture Organization of the United Nations.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.
- Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). Classification algorithms and regression trees. In L. Breiman, J. Friedman, C. J. Stone, & R. A. Olshen (Eds.), *Classification and regression trees* (pp. 246–280). New York, NY: Chapman and Hall CRC.
- Brewster, L. R., Dale, J. J., Guttridge, T. L., Gruber, S. H., Hansell, A. C., Elliott, M., ... Gleiss, A. C. (2018). Development and application of a machine learning algorithm for classification of elasmobranch behaviour from accelerometry data. *Marine Biology*, 165, 62.
- Brown, C., & Day, R. L. (2002). The future of stock enhancements: Lessons for hatchery practice from conservation biology. *Fish and Fisheries*, 3, 79–94.
- Brown, C., Davidson, T., & Laland, K. (2003). Environmental enrichment and prior experience of live prey improve foraging behaviour in hatchery-reared Atlantic salmon. *Journal of Fish Biology*, 63, 187–196.
- Brown, E. R., Jr. Rynicki, R. W., & Poff, R. J. (1985). *Population dynamics and interagency management of the bloater (Coregonus hoyi) in Lake Michigan, 1967–1982*. (Great Lakes Fishery Commission Technical Report No. 44). Ann Arbor, MI: Great Lakes Fishery Commission.
- Brownscombe, J. W., Griffin, L. P., Chapman, J. M., Morley, D., Acosta, A., Crossin, G. T., ... Danylchuk, A. J. (2020). A practical method to account for variation in detection range in acoustic telemetry arrays to accurately quantify the spatial ecology of aquatic animals. *Methods in Ecology and Evolution*, 11, 82–94.
- Cheung, W. W. L., Watson, R., & Pauly, D. (2013). Signature of ocean warming in global fisheries catch. *Nature*, 497, 365–369.
- Christensen, V. (1996). Managing fisheries involving predator and prey species. *Reviews in Fish Biology and Fisheries*, 6, 417–442.
- Christie, W. J., Scott, K. A., Sly, P. G., & Strus, R. H. (1987). Recent changes in the aquatic food web of eastern Lake Ontario. *Canadian Journal of Fisheries and Aquatic Sciences*, 44, s37–s52.
- Colborne, S. F., Rush, S. A., Paterson, G., Johnson, T. B., Lantry, B. F., & Fisk, A. T. (2016). Estimates of lake trout (*Salvelinus namaycush*) diet in Lake Ontario using two and three isotope mixing models. *Journal of Great Lakes Research*, 42, 695–702.
- Connerton, M., & Stewart, T. J. (2015). *Strategies for the reestablishment of native deepwater ciscoes in Lake Ontario. Draft document prepared for the Lake Ontario Committee*. Ann Arbor, MI: Great Lakes Fishery Commission.
- Csárdi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems* 1695. Retrieved from <http://igraph.org>
- D’Anna, G., Giacalone, V. M., Fernández, T. V., Vaccaro, A. M., Pipitone, C., Mirto, S., ... Badalamenti, F. (2012). Effects of predator and shelter conditioning on hatchery-reared white seabream *Diplodus sargus* (L., 1758) released at sea. *Aquaculture*, 356–357, 91–97.
- Dale, M. R. T., & Fortin, M.-J. (2010). From graphs to spatial graphs. *Annual Review of Ecology, Evolution, and Systematics*, 41, 21–38.
- 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. *Canadian Journal of Fisheries and Aquatic Sciences*, 75, 1811–1822.
- Dietrich, J. P., Fvllorison, B. J., & Hoyle, J. A. (2006). Alternative ecological pathways in the eastern Lake Ontario food web – Round goby in the diet of lake trout. *Journal of Great Lakes Research*, 32, 395–400.
- Donaldson, M. R., Hinch, S. G., Patterson, D. A., Hills, J., Thomas, J. O., Cooke, S. J., ... Farrell, A. P. (2011). The consequences of angling, beach seining and confinement on the physiology, post-release behaviour and survival of adult sockeye salmon during upriver migration. *Fisheries Research*, 108, 133–141.
- Donaldson, M. R., Raby, G. D., Nguyen, V. N., Hinch, S. G., Patterson, D. A., Farrell, A. P., ... Cooke, S. J. (2013). Evaluation of a simple technique for recovering fish from capture stress: integrating physiology, biotelemetry and social science to solve a conservation problem. *Canadian Journal of Fisheries and Aquatic Sciences*, 70, 90–100.
- Donaldson, M. R., Hinch, S. G., Suski, C. D., Fisk, A. T., Heupel, M. R., & Cooke, S. J. (2014). Making connections in aquatic ecosystems with acoustic telemetry monitoring. *Frontiers in Ecology and the Environment*, 12, 565–573.
- Dudgeon, C. L., Pollock, K. H., Braccini, J. T., Semmens, J. M., & Barnett, A. (2015). Integrating acoustic telemetry into mark-recapture models to improve the precision of apparent survival and abundance estimates. *Oecologia*, 178, 761–772.
- Eby, L. A., Roach, W. J., Crowder, L. B., & Stanford, J. A. (2006). Effects of stocking-up freshwater food webs. *Trends in Ecology and Evolution*, 21, 576–584.
- Eshenroder, R. L., Vecsei, P., Gorman, O. T., Yule, D. L., Pratt, T. C., Mandrak, N. E., Bunnell, D. B., & Muir, A. M. (2016). *Ciscoes (Coregonus, Subgenus Leucichthys) of the Laurentian Great Lakes and Lake Nipigon*. Great Lakes Fishery Commission Miscellaneous Publication 216-01. Ann Arbor, MI: Great Lakes Fishery Commission.
- Favé, M.-J., & Turgeon, J. (2008). Patterns of genetic diversity in Great Lakes bloaters (*Coregonus hoyi*) with a view to future reintroduction in Lake Ontario. *Conservation Genetics*, 9, 281–293.
- Fitzsimons, J. D., Brown, S. B., Williston, B., Williston, G., Brown, L. R., Moore, K., ... Tillitt, D. E. (2009). Influence of thiamine deficiency on lake trout larval growth, foraging, and predator avoidance. *Journal of Aquatic Animal Health*, 21, 302–214.
- Gaden, M., LaPan, S., & Todd, A. (2019). *Lake Ontario fishery agencies further adjust lakewide stocking of Chinook salmon and lake trout to maintain the world class fishery*. Lake Ontario Committee representing the fishery management agencies of Lake Ontario. Ann Arbor, MI: Great Lakes Fishery Commission.
- Gibson, A. J. F., Halfyard, E. A., Bradford, R. G., Stokesbury, M. J. W., & Redden, A. M. (2015). Effects of predation on telemetry-based survival estimates: insights from a study on endangered Atlantic salmon smolts. *Canadian Journal of Fisheries and Aquatic Sciences*, 72, 728–741.
- Gimenez, D. M. (1996). *Coregonus hoyi*. The IUCN Red List of Threatened Species 1996: e.T5366A11123739. <https://dx.doi.org/10.2305/IUCN.UK.1996.RLTS.T5366A11123739.en>. Downloaded on 15 October 2020.
- Grausgruber, E. E., & Weber, M. J. (2020). Is bigger better? Evaluation of size-selective predation on age-0 walleye. *North American Journal of Fisheries Management*, 40, 726–732.

- Halfyard, E. A., Webber, D., Del Papa, J., Leadley, T., Kessel, S. T., Colborne, S. F., & Fisk, A. T. (2017). Evaluation of an acoustic telemetry transmitter designed to identify predation events. *Methods in Ecology and Evolution*, 8, 1063–1071.
- Han, H., Guo, X., Yu, H. (2016). Variable selection using mean decrease accuracy and mean decrease gini based on random forest. pp. 219–224. In *2016 7th IEEE International Conference on Software Engineering and Service Science (ISCESS), Beijing, China, 26–28 August 2016*. Piscataway, NJ: Institute of Electrical and Electronics Engineers.
- Howell, B. R. (1994). Fitness of hatchery-reared fish for survival in the sea. *Aquaculture and Fisheries Management*, 25, 3–17.
- Hutchinson, M., Stewart, D., Chilcott, K., Butcher, A., Henderson, A., McLennan, M., & Smith, P. (2012). Strategies to improve post release survival of hatchery-reared threatened fish species. Murray-Darling Basin Authority, Publication No. 135/11, Canberra, Australia.
- Ives, J. T., McMeans, B. C., McCann, K. S., Fisk, A. T., Johnson, T. B., Bunnell, D. B., ... Muir, A. M. (2019). Food-web structure and ecosystem function in the Laurentian Great Lakes – toward a conceptual model. *Freshwater Biology*, 64, 1–23.
- Jobes, F. W. (1949). The age, growth, and bathymetric distribution of the bloater, *Leucichthys hoyi* (Gill) in Lake Michigan. *Papers of the Michigan Academy of Science, Arts, and Letters*, 33, 135–172.
- Karam, A. P., Kesner, B. R., & Marsh, P. C. (2008). Acoustic telemetry to assess post-stocking dispersal and mortality of razorback sucker *Xyrauchen texanus*. *Journal of Fish Biology*, 73, 719–727.
- Karpinnen, P., Jounela, P., Huusko, R., & Erkinaro, J. (2014). Effects of release timing on migration behaviour and survival of hatchery-reared Atlantic salmon smolts in a regulated river. *Ecology of Freshwater Fish*, 23, 438–452.
- Klinard, N. V., Halfyard, E. A., Fisk, A. T., Stewart, T. J., & Johnson, T. B. (2018). Effects of surgically implanted acoustic tags on body condition, growth, and survival in a small, laterally compressed forage fish. *Transactions of the American Fisheries Society*, 147, 749–757.
- Klinard, N. V., Matley, J. K., Fisk, A. T., & Johnson, T. B. (2019a). Long-term retention of acoustic telemetry transmitters in temperate predators revealed by predation tags implanted in wild prey fish. *Journal of Fish Biology*, 95, 1512–1516.
- Klinard, N. V., Halfyard, E. A., Matley, J. K., Fisk, A. T., & Johnson, T. B. (2019b). The influence of dynamic environmental interactions on detection efficiency of acoustic transmitters in a large, deep, freshwater lake. *Animal Biotelemetry*, 7, 17.
- Klinard, N. V., Matley, J. K., Halfyard, E. A., Connerton, M., Johnson, T. B., & Fisk, A. T. (2020). Post-stocking movement and survival of hatchery-reared bloater (*Coregonus hoyi*) reintroduced to Lake Ontario. *Freshwater Biology*, 65, 1073–1085.
- Koelz, W. (1929). Coregonid fishes of the Great Lakes. *Bulletin of the United States Bureau of Fisheries*, 43, 297–643.
- Krueger, C. C., Holbrook, C. M., Binder, T. R., Vandergoot, C. S., Hayden, T. A., Hondorp, D. W., ... Cooke, S. J. (2018). Acoustic telemetry observation systems: challenges encountered and overcome in the Laurentian Great Lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, 75, 1755–1763.
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., & Engelhardt, A. (2019). caret: classification and regression training. R package version 6.0-84. Retrieved from: <https://CRAN.R-project.org/package=caret>
- Lantry, J., Schaner, T., & Copeland, T. (2015). A management strategy for the restoration of lake trout in Lake Ontario, 2014 update. Ann Arbor, MI: Lake Ontario Committee of the Great Lakes Fishery Commission Document. Great Lakes Fishery Commission.
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2, 18–22.
- Molony, B. W., Lenanton, R., Jackson, G., & Norriss, J. (2003). Stock enhancement as a fisheries management tool. *Reviews in Fish Biology and Fisheries*, 13, 409–432.
- Muir, A. M., Krueger, C. C., & Hansen, M. J. (2012). Re-establishing lake trout in the Laurentian Great Lakes: Past, present, and future. In W. Taylor, A. Lynch, & N. Leonard (Eds.), *Great Lakes fisheries policy and management: A binational perspective* (pp. 533–588). East Lansing, MI: Michigan State University Press.
- Oksanen, J., Guillaume Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R.B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2019). vegan: Community ecology package. R package version 2.5-6. Retrieved from: <https://CRAN.R-project.org/package=vegan>
- Ontario Ministry of Natural Resources & Forestry. (2015). *Stocking strategy for the Canadian waters of Lake Ontario*. Picton, Canada: Lake Ontario Management Unit Document. Ontario Ministry of Natural Resources and Forestry.
- Owens, R. W., O'Gorman, R., Eckert, T. H., & Lantry, B. F. (2003). The Off-shore Fish Community in southern Lake Ontario, 1972-1998. In M. Munawar (Ed.), *State of Lake Ontario – Past, present, and future* (pp. 407–441). Burlington, Ontario, Canada: Ecovision World Monograph Series, Aquatic Ecosystem Health and Management Society.
- Perry, A. L., Low, P. J., Ellis, J. R., & Reynolds, J. D. (2005). Climate change and distribution shifts in marine fishes. *Science*, 308, 1912–1915.
- R Development Core Team. (2019). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from: <https://www.R-project.org/>
- Raby, G. D., Johnson, T. B., Kessel, S. T., Stewart, T. J., & Fisk, A. T. (2017). A field test of the use of pop-off data storage tags in freshwater fishes. *Journal of Fish Biology*, 91, 1623–1641.
- Raby, G. D., Johnson, T. B., Kessel, S. T., Stewart, T. J., & Fisk, A. T. (2020). Pop-off data storage tags reveal niche partitioning between native and non-native predators in a novel ecosystem. *Journal of Applied Ecology*, 57, 181–191.
- Rahel, F. J. (2000). Homogenization of fish faunas across the United States. *Science*, 288, 854–856.
- Rand, P. S., & Stewart, D. J. (1998). Dynamics of salmonine diets and foraging in Lake Ontario, 1983-1993: a test of a bioenergetic model prediction. *Canadian Journal of Fisheries and Aquatic Sciences*, 55, 307–317.
- Romine, J. G., Perry, R. W., Johnston, S. V., Fitzner, C. W., Pagliughi, S. W., & Blake, A. R. (2014). Identifying when tagged fishes have been consumed by piscivorous predators: application of multivariate mixture models to movement patterns of telemetered fishes. *Animal Biotelemetry*, 2, 3.
- Rush, S. A., Paterson, G., Johnson, T. B., Drouillard, K. G., Haffner, G. D., Hebert, C. E., ... Fisk, A. T. (2012). Long-term impacts of invasive species on a native top predator in a large lake system. *Freshwater Biology*, 57, 2342–2355.
- Scheffer, M., Carpenter, S., & de Young, B. (2005). Cascading effects of overfishing marine systems. *Trends in Ecology and Evolution*, 20, 579–581.
- Scott, W. B., & Crossman, E. J. (1973). *Freshwater fishes of Canada*. In *Bulletin* (Vol. 184). Ottawa, Canada: Fisheries Research Board of Canada.
- Stewart, T. J., Todd, A., & LaPan, S. R. (2017). *Fish-community objectives for Lake Ontario*. Fishery Management Document 2017-01. Ann Arbor, MI: Great Lakes Fishery Commission.
- Thorstad, E. B., Uglem, I., Finstad, B., Chittenden, C. M., Nilsen, R., Økland, F., & Bjørn, P. A. (2012). Stocking location and predation by marine fishes affect survival of hatchery-reared Atlantic salmon smolts. *Fisheries Management and Ecology*, 19, 400–409.
- Valletta, J. J., Torney, C., Kings, M., Thornton, A., & Madden, J. (2017). Applications of machine learning in animal behaviour studies. *Animal Behaviour*, 124, 203–220.
- Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Family Medicine*, 37, 360–363.
- Wang, G. M. (2019). Machine learning for inferring animal behavior from location and movement data. *Ecological Informatics*, 49, 69–76.

Zimmerman, M. S., & Krueger, C. C. (2009). An ecosystem perspective on re-establishing native deepwater fishes in the Laurentian Great Lakes. *North American Journal of Fisheries Management*, 29, 1352–1371.

#### SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Klinard NV, Matley JK, Ivanova SV, Larocque SM, Fisk AT, Johnson TB. Application of machine learning to identify predators of stocked fish in Lake Ontario: using acoustic telemetry predation tags to inform management. *J Fish Biol.* 2021;98:237–250. <https://doi.org/10.1111/jfb.14574>