ARTICLE

Keeping it classy: classification of live fish and ghost PIT tags detected with a mobile PIT tag interrogation system using an innovative analytical approach

J. Benjamin Stout, Mary Conner, Phaedra Budy, Peter Mackinnon, and Mark McKinstry

Abstract: The ability of passive integrated transponder (PIT) tag data to improve demographic parameter estimates has led to the rapid advancement of PIT tag systems. However, ghost tags create uncertainty about detected tag status (i.e., live fish or ghost tag) when using mobile interrogation systems. We developed a method to differentiate between live fish and ghost tags using a random forest classification model with a novel data input structure based on known fate PIT tag detections in the San Juan River (New Mexico, Colorado, and Utah, USA). We used our model to classify detected tags with an overall error rate of 6.8% (1.6% ghost tags error rate and 21.8% live fish error rate). The important variables for classification were related to distance moved and response to monsoonal flood flows; however, habitat variables did not appear to influence model accuracy. Our results and approach allow the use of mobile detection data with confidence and allow for greater accuracy in movement, distribution, and habitat use studies, potentially helping identify influential management actions that would improve our ability to conserve and recover endangered fish.

Résumé : Les données de transpondeurs passifs intégrés (PIT) ont permis d'améliorer les estimations de paramètres démographiques, ce qui a engendré une progression rapide des systèmes d'étiquettes PIT. Les étiquettes fantômes génèrent toutefois une incertitude quant au statut des étiquettes détectées (c.-à-d. poisson vivant ou étiquette fantôme) quand des systèmes d'interrogation mobiles sont utilisés. Nous avons mis au point une méthode pour distinguer les poissons vivants des étiquettes fantômes qui fait appel à un modèle de classification par arbres décisionnels intégrant une structure d'entrée de données novatrice basée sur les détections d'étiquettes PIT de destin connu dans la rivière San Juan (au Nouveau-Mexique, au Colorado et en Utah, États-Unis). Nous avons utilisé le modèle pour classer les étiquettes détectées avec un taux d'erreurs global de 6,8 % (taux d'erreurs de 1,6 % pour les étiquettes fantômes et de 21,8 % pour les poissons vivants). Les variables importantes pour la classification sont reliées à la distance parcourue et la réaction aux débits de crue de mousson; les variables associées à l'habitat ne semblent cependant pas influencer l'exactitude du modèle. Nos résultats et notre approche permettent l'utilisation fiable de données de détection mobile et une plus grande exactitude dans les études des déplacements, de la répartition et des habitats, ce qui pourrait aider à cerner des mesures de gestion efficaces pour améliorer la conservation et le rétablissement d'espèces de poissons menacées. [Traduit par la Rédaction]

Introduction

Successful management of sport fisheries, conservation of native fish, and endangered species recovery rely on the ability to accurately assess the effectiveness of focused management actions (Parma 1998; Pine et al. 2009; Clark et al. 2018). To assess the impact of management actions, managers often estimate demographic parameters such as survival and abundance (Gibbs et al. 1998; Maxwell and Jennings 2005; Osmundson and White 2017). However, the ability to detect population trends depends on both the accuracy and precision of estimates. Vital rates can be affected by sampling efforts, methods, gear, and data analysis (e.g., Walther and Moore 2005). Owing to a lack of accuracy and precision in some estimates, for many species there is still uncertainty regarding the

processes limiting their viability, probability of persistence or recovery, and the effectiveness of management actions (Al-Chokhachy et al. 2009; Osmundson and White 2017; Clark et al. 2018).

Currently, one of the most common methods to generate estimates of survival and abundance of fishes is to sample with active gear (e.g., electrofishing) and perform a mark–recapture analysis of the collected data (Mesa and Schreck 1989). Active sampling is expensive and gear-intensive (Schramm et al. 2002; Evans et al. 2017), which can limit the amount of survey effort expended per survey. In addition to negative impacts on the sampled fishes when electrofishing (Dwyer and White 1997; Ruppert and Muth 1997; Snyder 2003), netting and handling can have negative effects on fishes, and all of these effects can be exacerbated by water temperatures at time of capture (e.g., higher water temperatures

J.B. Stout. Department of Watershed Science, Utah State University, 5200 Old Main Hill, Logan, UT 84322, USA; The Ecology Center, Utah State University, 5205 Old Main Hill, Logan, UT 84322, USA.

M. Conner. Department of Wildland Science, Utah State University, 5200 Old Main Hill, Logan, UT 84322, USA; The Ecology Center, Utah State University, 5205 Old Main Hill, Logan, UT 84322, USA.

P. Budy. US Geological Survey, Utah Cooperative Fish and Wildlife Research Unit, Department of Watershed Sciences, Utah State University, 5210 Old Main Hill, Logan, UT 84322, USA; The Ecology Center, Utah State University, 5205 Old Main Hill, Logan, UT 84322, USA.

P. Mackinnon. Department of Watershed Science, Utah State University, 5200 Old Main Hill, Logan, UT 84322, USA.

M. McKinstry. US Bureau of Reclamation, 125 State Street #6103, Salt Lake City, UT 84138, USA.

Corresponding author: J. Benjamin Stout (email: ben.stout@aggiemail.usu.edu).

Fig. 1. The San Juan River study area began at the Public Service Company of New Mexico (PNM) diversion at river kilometre (rkm) 268 and ended at Clay Hills at rkm 4.6. Horizontal and vertical lines are state boundaries, and state abbreviations are UT (Utah), CO (Colorado), AZ (Arizona), and NM (New Mexico). Cities are represented by filled circles, and select landmarks are shown with rkm. Map data are from Esri, National Atlas of the United States, and the US Geological Survey (map produced using ArcMap 10.5.1).
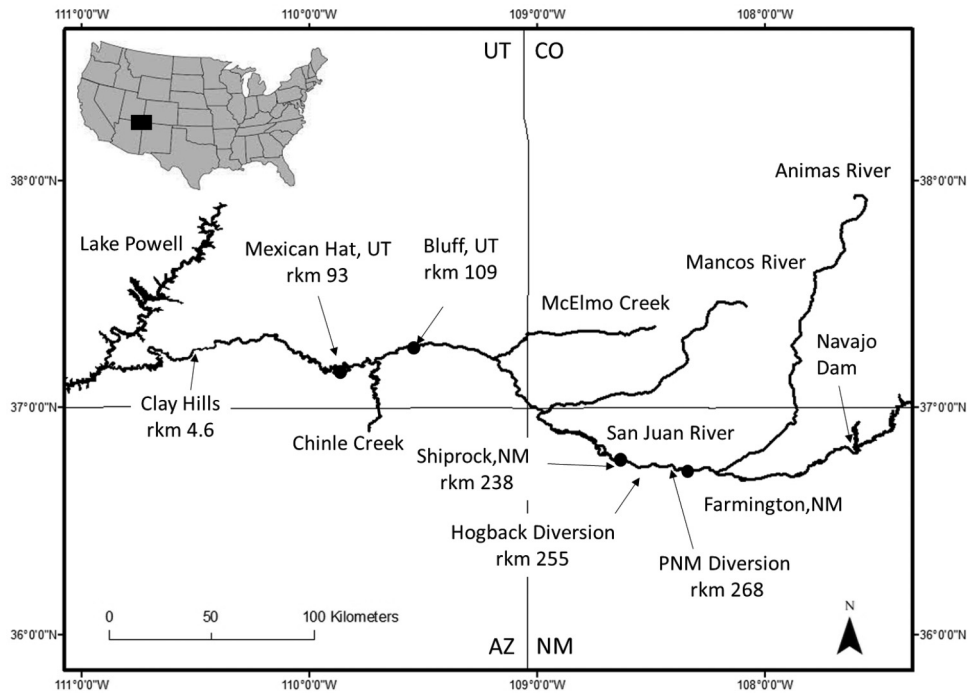
can cause greater impacts; Paukert et al. 2005; Hunt et al. 2012). In addition to the potential negative effects of physically capturing fish, low rates of detection or recapture can be problematic for estimating demographic parameters (Osmundson and White 2017). Endangered fishes, by definition, have low abundances and can be difficult to detect in large rivers in general. Desert rivers in particular also experience high turbidity, which can further reduce the capture probability of fishes (Lyon et al. 2014). In some systems with rare or endangered fish species, low recapture rates have resulted in poor estimates of demographic parameters (Hewitt et al. 2010; Dudgeon et al. 2015), potentially exacerbating factors limiting our conservation efforts.

Collectively, these issues clearly indicate a need for more effective ways to sample and monitor fish with fewer potential negative consequences and higher recapture–resight rates. The use of passive integrated transponder (PIT) tags and passive integrated antennas (PIAs) can reduce sampling stress because capture and handling of fish is not required after the initial capture and tagging event (which can also cause stress and mortality) and still generate large amounts of individual-based movement data. PIAs can be used singly to answer questions about total numbers of fish detected at a location (Burke and Jepson 2006) or paired to determine direction of movement (Fetherman et al. 2015) to answer questions about fish passage and use of tributaries for spawning or habitat use (Cathcart et al. 2015; Howell et al. 2016). In addition, remote sighting data, such as data from PIAs, have substantially improved estimates of fish vital rates and abundance when added to data collected with other methods (Pine et al. 2003; Webber et al. 2012; Webber and Beers 2014). However, PIAs rely on the fish swimming past a fixed point, which can limit the numbers of fish detected to fish that are inclined to move (Snook et al. 2016). Different forms of mobile PIT tag antennas (Fischer et al. 2012; Hodge et al. 2015; Richer et al. 2017) have been created to address some of the limitations of PIAs.

While the mobile PIT tag antennas can improve detection probability, increase spatial coverage of PIT tag sampling, and answer questions about habitat association, there are limitations and analytical issues to be resolved. In other studies where mobile PIT tag antennas have been used (Fetherman et al. 2015; Richer et al. 2017), an emerging concern is the prevalence of ghost tags. Ghost tags are created when tag loss, predation, and natural mortality leave a PIT tag in the environment (O'Donnell et al. 2010). If not known or properly accounted for analytically, these ghost tags would create bias in estimated vital rates due to an inflated number of fish perceived as alive and detected with mobile passive gear (O'Donnell et al. 2010).

Importantly, the detection of live fish cannot a priori be differentiated from the detection of ghost tags; yet that knowledge is critical in order for the mobile techniques to be used effectively. The objective of this study was to develop a methodology to classify each detected PIT tag as a live fish or a ghost tag based on tag location data and build a set of guidelines or rules that could be adapted for use here and in other systems if the correct data were available. We built two different random forest models using data from known fate tags to determine the best data input structure. The first structure treated all individual movements as independent and unrelated to any other movements. The second structure combined all movements of an individual to account for the relatedness of consecutive movements.

Methods

Study site

Our study site encompassed ~264 river kilometres (rkm) of the San Juan River from the Public Service Company of New Mexico (PNM) diversion near Farmington, New Mexico, at rkm 268.2 to the Clay Hills takeout near Lake Powell at rkm 4.6 (Fig. 1). The San Juan River contains federally designated critical habitat for multiple endangered fish species (US Fish and Wildlife Service 1990,

1566

Can. J. Fish. Aquat. Sci. Vol. 77, 2020

1998), is a part of the upper Colorado River system, and the basin also covers ~99 200 km² in Colorado, New Mexico, Utah, and Arizona. The river is ~616 km long with the headwaters located in the San Juan Mountains of Colorado. The Navajo Dam was completed in 1962 and is the only major impoundment on the river. The Animas River, which is the largest tributary of the San Juan River, is largely unregulated and joins the San Juan River downstream of Navajo Dam resulting in a more natural hydrograph below its confluence. Historically, discharge was snowmelt-driven (during our study spring runoff discharge peaked at ~254 m³·s⁻¹) with potential for flash floods during monsoonal rain events (during our study the monsoonal flood discharge peaked at ~566 m³·s⁻¹). Since 1993, dam operations have also attempted to match flows with the Animas River to mimic natural flow regimes for native fish conservation (Gido and Propst 2012).

The San Juan River supports seven native species: Colorado pikeminnow (*Ptychocheilus lucius*), razorback sucker (*Xyrauchen texanus*), flannelmouth sucker (*Catostomus latipinnis*), bluehead sucker (*Catostomus discobolus*), roundtail chub (*Gila robusta*), speckled dace (*Rhinichthys osculus*), and mottled sculpin (*Cottus bairdii*). In addition, there are over 20 nonnative fishes in the river, with the most abundant being channel catfish (*Ictalurus punctatus*), smallmouth bass (*Micropterus dolomieu*), and common carp (*Cyprinus carpio*). The Colorado pikeminnow and razorback sucker, two large-bodied and long-lived (up to 50 years) fishes, endemic to the Colorado River Basin, are listed as endangered under the US Endangered Species Act. The decline of these two species is attributed to the effects of instream diversions and the major impoundment of Navajo Dam, habitat alteration from nonnative vegetation, and predation by and competition with nonnative species (US Fish and Wildlife Service 1990, 1998). These two species were considered essentially extirpated in the San Juan River in 1992, and stocking of both species began soon after. Since 1994, ~50 000 Colorado pikeminnow and 150 000 razorback sucker have been PIT-tagged in the San Juan River (STReaMS 2018). The flannelmouth sucker, the bluehead sucker, and the roundtail chub, some of which are also PIT-tagged in the San Juan River, are considered to be Species of Special Concern in at least two of the following states: Wyoming, Utah, Colorado, New Mexico, and Arizona (Bezzerides and Bestgen 2002), and in some states they are managed under a range-wide conservation agreement (Budy et al. 2015). Given the amount of tagging in the system and elsewhere, ghost tags are a problem of increasing concern, and the long life-spans of the fishes in the San Juan River eliminates the possibility of classifying tags as ghosts based solely on age.

Sampling methods

In a previous study (Stout et al. 2019), we describe the Passive Integrated Transponder Portable Antenna SystemS (PITPASS, a raft-based PIT tag antenna system with integrated GPS), our sampling method, the distribution and detection of known ghost tags, and the data recorded for each detection in greater detail. We used three PITPASS boats, and each sampling pass varied in length (115–264 rkm) and river discharge (11.3–130.3 m³·s⁻¹). In total, we sampled ~2233 rkm during 13 passes over the course of two field seasons. The data recorded or calculated from our detection data included distance moved between detections, direction moved relative to flow, metres moved per day, which habitat feature a tag was detected in (*i*) riffle — water velocity moderate to rapid, water surface disturbed, and substrate usually cobble and rabble, (*ii*) run — water velocity moderate to rapid, little or no surface disturbance, and substrate sand and (or) silt in low velocity runs and gravel or cobble in high velocity runs, (*iii*) pool — area within channel, flow either barely or not perceptible, substrate varied, and usually deeper than 30 cm, (*iv*) low velocity — all low-velocity habitats not included in pool, such as backwater, eddy, etc., and (*v*) shoal — less than 25 cm deep, laminar flow, and varied substrate), whether a detection occurred in a single thread or an-

abranching channel, whether a detection occurred in a canyon or unconfined alluvial valley, whether a movement occurred during the monsoonal flash flood, and whether a movement occurred during overwinter–spring runoff. Although the area each habitat type occupied varied with flow level, during our sampling the following approximate percentages can be used: run = 80%, riffle = 6%, shoal = 10%, pool = 0.1%, low velocity = 4% (Bliesner and Lamarra 2000).

For this study, we divided the data collected during field sampling into three categories: ghost tag, unknown tag, and live fish. Ghost tag refers to the PIT tags we distributed randomly in the river to simulate ghost tags. We distributed 5000 total known ghost tags into two distinct morphological reaches ~16 km in length (a shallow, unconfined alluvial reach and a deeper canyon reach). Each reach received ~1250 tags per year distributed on the first sampling trip of the season. We randomly distributed one-third of the tags from each boat at a rate of one tag per minute across all habitat types (Stout et al. 2019). Unknown tag refers to any PIT tag implanted in a fish at any time in the San Juan River. We used data from the STReaMS database, which is the repository for all PIT tag data collected in the Upper Colorado River Basin and the San Juan River Basin (STReaMS 2018), to confirm detections as live fish. A fish was considered live if the fish associated with the tag was physically captured or detected at a PIA system in the same field season as our detection or any time after. PIAs in the San Juan River are overwhelmingly in either fish passage structures or tributaries because of the size of the river and its substrate (sand bed river up to 100 m wide). In order for a tag to be detected by a PIA in the San Juan River after being detected in the main channel by the PITPASS, a tag would have to move upstream into a tributary or a fish passage structure. Therefore, we considered it appropriate to consider tags demonstrating these movements as live fish, as it is unlikely a ghost tag would move this way.

Analysis: random forest

We used random forest analysis to develop and evaluate models for classifying tags (into two classes: live fish and ghost tags) using the known fate tag data. We used the randomForest function in package randomForest in R (Liaw and Wiener 2002) to create our models using the default parameters (500 trees and four variables tried per split). Random forest is a method for classification and regression and was used, in part, because it makes no assumptions about the distribution of data and its strength as a statistical classifier (Cutler et al. 2007). We used the randomForest function that develops classification trees using a bootstrapping method leaving out roughly one-third of the data to be used for testing. The unevaluated third of data is referred to as out-of-bag and is used to evaluate the accuracy of the constructed model. Since out-of-bag data are not used for developing the trees, out-of-bag error estimates serve as cross-validated accuracy estimates (Cutler et al. 2007). We used the out-of-bag error estimates to assess model performance.

We structured the input data for the random forest analysis in two different ways to achieve the most accurate model. Typically in random forest classification models, each site or individual is only observed one time (Cutler et al. 2007), and sites or individuals are considered independent. In our case, we considered each observed movement independent and unrelated to any other movements even if made by the same tag (despite the obvious fact that a second movement of a tag will begin where the first tag ended, possibly introducing bias regarding habitat use). We refer to the models using this data structure as independent and include an example of the input data structure in the online Supplemental

Table 1. Descriptions of variables used to create the model with the independent data structure for random forest analysis of data from the San Juan River study area in 2016 and 2017.

Variable group	Variable name	Type	Description
Distance moved	distance	Continuous	Distance moved between detections (metres)
	m_per_day	Continuous	Distance divided by number of days between detections (metres)
Direction moved	direction	Categorical	Direction moved relative to flow (up, down, left, and right)
Habitat	hab_start	Categorical	Habitat type of first detection (run, riffle, pool, shoal, low velocity)
	hab_stop	Categorical	Habitat type of second detection (run, riffle, pool, shoal, low velocity)
	channel_start	Categorical	Type of channel of first detection (single thread or anabranching)
	channel_stop	Categorical	Type of channel of second detection (single thread or anabranching)
	habunit_start	Categorical	Habitat unit of first detection (canyon or unconfined alluvial reach)
	habunit_stop	Categorical	Habitat unit of second detection (canyon or unconfined alluvial reach)
Discharge	monsoon	Binary	Whether or not the movement occurred during the monsoonal flood flows
	overwinter	Binary	Whether or not the movement occurred during the spring runoff flows

materials (Table S.1[1]). The variables used for classification in the independent model included two variables for distance moved (total distance and metres moved per day), one variable for direction moved (direction), six variables for habitat (hab_start, hab_stop, channel_start, channel_stop, habunit_start, and habunit_stop), and two variables for flow conditions (monsoon and overwinter; all variables defined in Table 1).

We developed a novel data input structure that we refer to as dependent that condensed all movements of a single tag into one observation to connect all movements of a single PIT tag. In our study, some tags were resighted up to seven times giving us repeated measures of the same individual. An example of this data structure is included in the online Supplemental materials (Table S.2[1]). The new variables we used for classification in the dependent model included the total distance moved in a specific direction relative to flow: upstream, downstream, left, and right. We also included two variables encompassing cumulative distance moved (total and metres moved per day) and two variables for flow conditions observed during movement (monsoon and overwinter). For habitat data, we used a count of how many times a tag was detected in each habitat type (run, riffle, shoal, pool, low velocity, single thread channel, anabranching channel, canyon, and unconfined alluvial; variables defined in Table 2). Of the two data structures, the model with the lowest out-of-bag error rate was chosen as the most accurate.

Our initial model used all live fish detected for classification, regardless of species. Then, we limited our input data to a single species at a time and reanalyzed each data structure to determine differences among the species, because while the species detected exhibit similar movement behaviors, the habitat associations may be different. In each case, the entire ghost tag dataset was used. The only three species detected and subsequently used for analysis were razorback sucker, flannelmouth sucker, and Colorado pikeminnow (listed from most to least abundant in our sample); however, limiting our live fish data to a single species decreased the number of observations used to construct each model.

We used a backwards stepwise variable selection method to evaluate the influence of the predictor variables on the best model (based on the out-of-bag classification error rates; Cutler et al. 2007). We performed this step to determine which variables were necessary to maintain the accuracy of the model while reducing its complexity. Random forest models rank variables in order of importance for correct classification based on the "mean decrease in accuracy", which describes the loss in accuracy of the model when a variable is not used in the analysis (Cutler et al. 2007). Based on the ranking of variables, we removed the worst performing variable (i.e., the variable with the lowest mean decrease in

accuracy) in the model and reran the analysis. These steps were repeated until only one variable remained, and the accuracy of each version of the model was examined. Then, we examined the variables identified as important by our variable selection process (removal from the model led to a decline in accuracy) using partial dependence plots to determine the differences in live and ghost tag dynamics. Partial dependence plots display the logit of probability of belonging to a specific class and all values of the variable of interest. For partial dependence plots, all other variables are held to an average value; therefore, partial dependence plots are a vast simplification of complex data and should be interpreted with caution.

Results

Tag detections

Our detection data allowed us to describe the movement of 899 known ghost tags over two seasons. Of the 5000 known ghost tags distributed in the San Juan River, 899 were detected more than once, for a raw resight rate of 18%. Of those 899 tags, there were 1401 pairs of detections where movement could be measured, because some tags were detected more than two times (Table 3).

The detection of unknown tags allowed us to describe the movement of 302 confirmed live fish over two seasons. We detected a total of 3958 unique unknown tags, but only 847 of those tags were detected a second time, for a raw resight rate of 21%. Of those 847 tags, there were 1190 pairs of detections for which we could measure movement. However, we were only able to confirm 302 of those tags as live fish, with 370 pairs of detections where we could measure movement (Table 3). All of the data from both confirmed live fish, and the known ghost tags were used in the construction of the random forest classification models.

Random forest models

The random forest models using the dependent data structure were more accurate than the independent models. Further, limiting input data by species did not improve either model. With the independent model, we were able to differentiate the live fish from the ghost tags with an overall error rate of 7.6% (out-of-bag estimate of error rate). When examined separately, the two classes exhibited error rates of 28.9% (live fish incorrectly classified as ghost tags) and 1.9% (ghost tags incorrectly classified as live fish; Fig. 2). Our ability to correctly classify was better with the dependent model than with the independent model, lowering the overall error rate to 6.8%. The accuracy of both classes improved and, when examined separately, demonstrated error rates of 21.8% (live fish) and 1.6% (ghost tags; Fig. 2). Separating known live

---

1568

Can. J. Fish. Aquat. Sci. Vol. 77, 2020

Table 2. Descriptions of variables used to create a live versus ghost tag classification model with the dependent data structure for random forest analysis of data from the San Juan River study area in 2016 and 2017.

Variable group	Variable name	Type	Description
Direction	up	Continuous	Sum of movements upstream relative to flow direction (metres)
	down	Continuous	Sum of movements downstream relative to flow direction (metres)
	left	Continuous	Sum of movements towards river left (metres)
	right	Continuous	Sum of movements towards river right (metres)
Cumulative distance	total	Continuous	Sum of all movement regardless of direction (metres)
	m_per_day	Continuous	Total divided by number of days between first and last detection (metres)
Habitat	run	Discrete	Number of times detected in run habitats
	riffle	Discrete	Number of times detected in riffle habitats
	shoal	Discrete	Number of times detected in shoal habitats
	pool	Discrete	Number of times detected in pool habitats
	low_velocity	Discrete	Number of times detected in low velocity habitats
	single	Discrete	Number of times detected in single thread areas of channel
	anabranching	Discrete	Number of times detected in anabranching areas of channel
	canyon	Discrete	Number of times detected in the canyon reach
	unconfined	Discrete	Number of times detected in the unconfined alluvial reach
Discharge	monsoon	Binary	Whether or not the movements observed occurred during the monsoonal flood flows
	overwinter	Binary	Whether or not the movements observed occurred during the spring runoff flows

Table 3. Total number of tags detected, numbers of total movements observed, and numbers of unique tags detected more than one time (unique resights) in the San Juan River study area during 2016 and 2017 for known ghost tags, unknown status tags, and live fish.

	Total detected	Unique resights	Movements
Ghost tags	5000	900	1405
Unknown tags	3958	847	1190
Live fish	302	302	370
Razorback sucker	205	205	254
Flannelmouth sucker	70	70	87
Colorado pikeminnow	24	24	26

detections by species led to decreases in model accuracy (specifically of the error rate of the live class) when using either independent or dependent data structures. The ranking of models from best to worst (based on error rate of the live class) for both data structures were (i) all species combined, (ii) razorback sucker only, (iii) flannelmouth sucker only, and (iv) Colorado pikeminnow (Fig. 2).

We used the ranking of predictor variables to determine the five most important variables for correct classification of detected tags: (i) distance moved up, (ii) total distance moved, (iii) metres moved per day, (iv) distance moved down, and (v) monsoon (Fig. 3). These variables were determined to be the most important by our backwards stepwise variable selection method. Any other variable, other than these five, could be removed without an effect on the accuracy of the model. When any of these top five variables were removed from the model, the accuracy of the model declined (Fig. 4). All of the variables related to habitat use or association were ranked lower, with regard to importance, as were the variables describing movement perpendicular to the river's flow. Habitat features and measures of river complexity did not appear to influence the accuracy of the models, as the performance of the model did not change with their removal from the analysis.

When the five most important variables were examined individually using partial dependence plots, some patterns emerged (Fig. 5). We reiterate here that partial dependence plots are simplifications of complex data, where all other variables are held constant while examining the variable of interest. Larger total distances and upstream distances moved are more likely to belong to a live fish. The probability of being a live fish increased with larger upstream movements up to ~2000 m, at which point a larger distances moved upstream did not increase the probabil-

ity of classification as a live fish (some ghost tags were observed to move upstream up to 99 m; Stout et al. 2019). Distance moved upstream is an example of the limited interpretability of these plots. It would be expected that any movement upstream should belong to a live fish, and it would also be expected that the probability of being classified as a live fish would be higher than the maximum probability of ~75% this variable achieves when a tag moves more than 2 km in an upstream direction. The probability of being a live fish increased as the total distance moved increased up to ~8000 m, at which point a further increase in distance no longer increased the probability of classification as a live fish. The probability of being a live fish increased as distance moved downstream increased until ~8000 m, when a larger distance no longer increased the probability of classification as a live fish. The highest probability of indicating a live fish, based on the number of metres moved per day, occurs around 650 m. Known ghost tags had a higher probability of having a monsoonal flood affected movement as demonstrated by the negative relationship shown on the graph (Fig. 5).
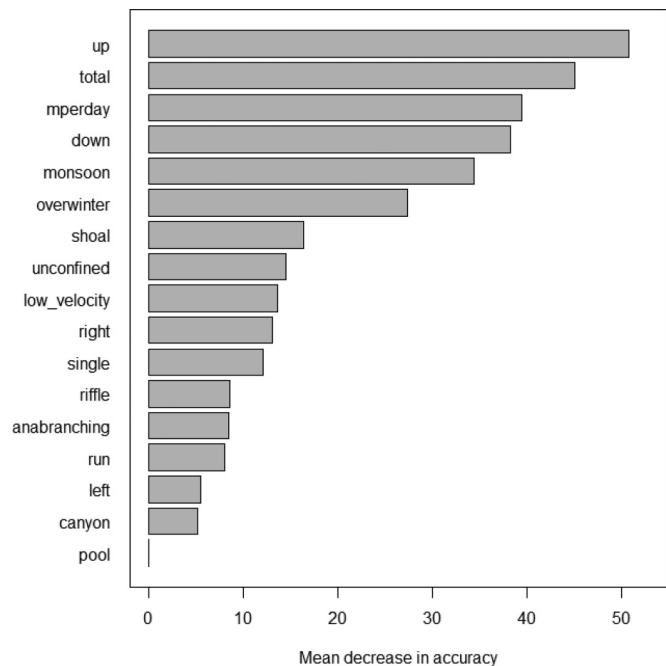
Discussion

Our study is the first to evaluate a classification method for determining PIT tag status (live versus ghost) when using mobile interrogation systems. We were successful with a very high degree of overall accuracy. While we could identify ghost tags with a high degree of accuracy (only 2% were misclassified as live fish), the model was less effective with live fish. Using our current error rates as a guide, ~20% were misclassified as a ghost tag. This degree of error makes our classification of live tags conservative. However, only ~20% of all detected tags were ever resighted, meaning 80% of tags detected were never classified during our 2-year sampling period. Therefore, these unclassified detections could not be used for estimating vital rates without introducing bias. A longer sampling period and more detection passes would increase detections and subsequent resights, potentially allowing greater accuracy in classification of live fish.

Our top model used the dependent data structure and combined the data from all detected species. The dependent structure accounted for the possibility of a tag being detected more than two times and resulted in a better description of an individual tag's movements and a greater ability to classify it correctly. When the input data were limited to a single species, there was a lower number of observations to build the model, and the subsequent decline in accuracy of the model appeared to correspond to

Fig. 2. Change in accuracy of the random forest model, using the dependent (*x* axis) and independent data structure (*y* axes), when all species of live fish are combined and then limited to a single species. The models are (A) all species combined, (B) only razorback sucker, (C) only flannelmouth sucker, and (D) only Colorado pikeminnow. "# Obs" is the number of observations of confirmed live fish used to create the model for each species. The dashed line for the number of observations is drawn only to illustrate the trend of fewer observations as the data are split into species and does not have any statistical meaning. Data are from the San Juan River study area in 2016 and 2017.
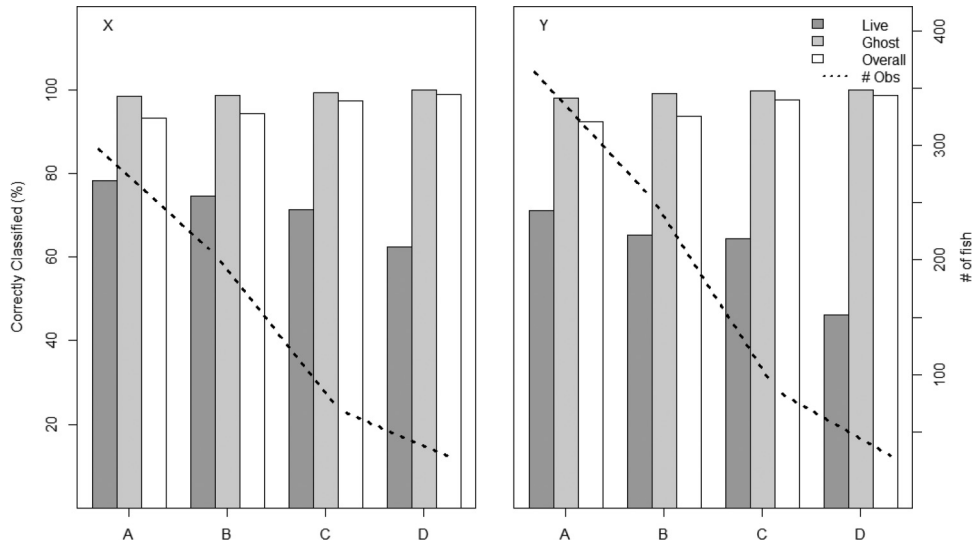
Fig. 3. Variable importance plot of the predictor variables used to create the classification model ranked from top (most important) to bottom (least important). The *x* axis demonstrates the decrease in accuracy of the model when each variable is removed; higher values indicate more important variables for classification. Data are from the San Juan River study area in 2016 and 2017.

sample size imbalance in Appendix A. Since all three species analyzed exhibit similarities in movement behavior (McKinney et al. 1999; Irving and Modde 2000; Zelasko et al. 2010), we believe combining all species for the analysis was appropriate. Combining all species might not work in other cases where the species are less related or exhibit very different movement behaviors.

We expected differences in movement between ghost tags and live fish would be important in classification. Razorback sucker, flannelmouth sucker, and Colorado pikeminnow are all known to move very large distances (McKinney et al. 1999; Irving and Modde 2000; Zelasko et al. 2010), and while ghost tags can move distances up to 4 km (Bond et al. 2019; Stout et al. 2019), we observed a large difference between the two classes. Fish carcasses can also move large distances after death (Havn et al. 2017), but the average time for a white sucker carcass to shed an abdominally located tag was only 73.3 h (Muhametsafina et al. 2014). It is perhaps unsurprising then that, of the five variables identified as influencing the accuracy of the model, four were related to the distance moved by a tag. As expected, live fish generally moved greater distances than the ghost tags. Also, fish and other organisms exhibit behavioral changes to minimize the effect of high flows (Lytle and Poff 2004), and ghost tags respond to higher flows similarly to sediment (Bond et al. 2019; Stout et al. 2019). Therefore, we expected ghost tags and live fish to exhibit different responses to changes in discharge, and this expectation was demonstrated by the fifth variable influencing classification, monsoonal flows.

Despite our expectation that live fish behavior and ghost tag deposition would result in differences in habitat association, habitat type did not appear to be important in the correct classification of tags. We believe this disconnect between tag location and habitat might not be true for all river systems. The PIT tags we used to simulate ghost tags were an artificial addition to the system randomly placed in the river in all habitat features. However, as time passes and these tags move, their distribution tends to favor riffles over runs (Stout et al. 2019). We believe the signal from this phenomenon could be masked by the overwhelming numbers of tags randomly associated with habitat as per our initial distribution. Additionally, the San Juan River has very few large pools or deepwater habitat where tags, and possibly even fish, could accumulate, further reducing the ability to detect differences in habitat association.

the decline in the number of observations. Even when all species were combined, there was still an imbalance in our classes, which can cause problems with overfitting. However, random forest models have been shown to perform well despite class imbalance at much higher ratios than our situation of ~3:1 (ghost tags : live fish; Khoshgoftaar et al. 2007; Elrahman and Abraham 2013; Muchlinski et al. 2016). We explore the effects of different methods (up-sampling and down-sampling to have equal sample sizes) of accounting for

1570

Can. J. Fish. Aquat. Sci. Vol. 77, 2020

Fig. 4. The change in classification error when predictor variables are removed from the model. The *x* axis lists which variable has been removed sequentially from left to right (and in order from least to most important based on the variable importance plot; Fig. 3), beginning with "none" (no variables have been removed yet; all variables used in the model) and ending with "total". The variable "up" is never removed because one predictor variable is required to build the model. The vertical line delineates the beginning of a large increase in classification error based on our backwards stepwise variable selection process. Data are from the San Juan River study area in 2016 and 2017.
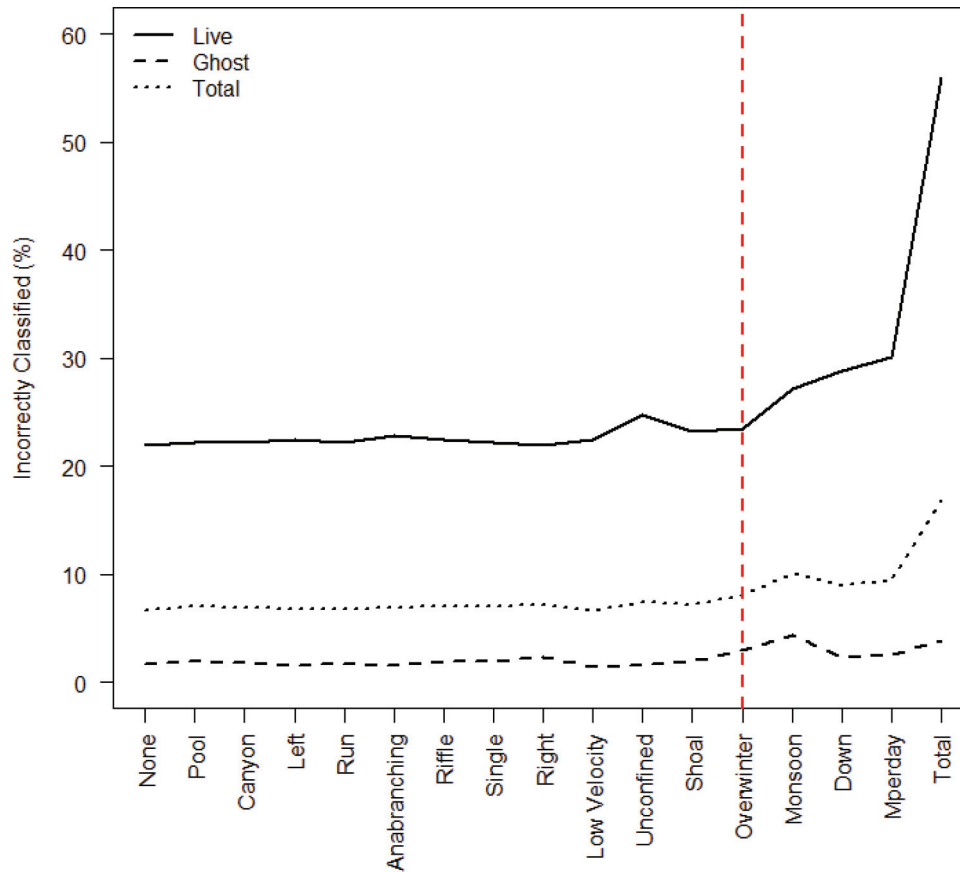
Mobile PIT tag interrogators have been developed to supplement low recapture rates and reduce negative effects of capture and handling of organisms (Paukert et al. 2005; Hunt et al. 2012; Reynolds et al. 2012), but the inability to separate ghost tags from live fish based on detection data are one of the recurring points of discussion in studies using mobile systems (e.g., O'Donnell et al. 2010; Fetherman et al. 2014; Richer et al. 2017). Extensive use of PIT tags has led to large numbers of potential ghost tags in aquatic systems, and in other work, extensive movement of ghost tags has been observed (from 2.0 km in Bond et al. 2019 to 4.1 km in Stout et al. 2019), which highlights the need for a classification system. Our study demonstrates the possibility of classifying detected tags, yet there is uncertainty associated with the method. In particular, our system relies on detecting the tag at least twice, which is not always possible. With limited sampling (i.e., passes), a short study period, or remote areas or complex habitats that are difficult to access, this system may be limited in its applicability.

Despite the capabilities of our system, one limitation is the time required to collect comprehensive data. The small sample size of confirmed live fish (only 30 individuals) from our first year of sampling made it difficult to successfully classify detected tags. However, we were able to achieve a much higher level of model accuracy with the addition of the second year's data (272 additional live fish). However, the low number of redetections for live fish made it impractical to use mobile detection data to estimate survival rates without another source of data (i.e., electrofishing or detections at stationary antenna). We note these limitations to emphasize this process is iterative, and potential users need to

remember classification techniques will require multiple passes, ideally within and across years. Once classification models are devised for a system, they can be continuously improved upon with the addition of subsequent data.

While we were unable to use individual characteristics of tagged fish as predictor variables in our classification model (known ghost tags were never in a fish), this approach could be useful elsewhere. Potential individual variables that could influence movement or habitat used by individuals include (*i*) the age of the fish or how long the tag has been deployed, (*ii*) species of fish tagged, (*iii*) length and (or) weight at last capture, and (*iv*) sex. In the San Juan River, fishes can be extremely long-lived (up to 40–50 years; McCarthy and Minckley 1987; Scoppettone 1988; Osmundson et al. 1997); therefore, no tags could be classified as ghosts based on age. However, in systems with shorter-lived fishes (e.g., Rio Grande silvery minnow (*Hybognathus amarus*), which live <4 years), some ghost tags can be classified definitively based on age and used to model ghost tag movement to calibrate a classification model. This approach to identifying ghost tags has been used successfully in streams with coho salmon (*Oncorhynchus kisutch*), which have a definitive age or time of mortality (Bond et al. 2019).

Mobile PIT tag detection systems could be extremely useful in the future in habitat use and association studies. Habitat restoration is an increasingly popular option for conservation and recovery of fishes, and due to the high cost of restoration (Bernhardt et al. 2005), plans should be based on accurate response data. Electrofishing and snorkeling are two methods typically used for

Fig. 5. Partial dependence plots for the five variables (A) up, (B) total, (C) m_per_day, (D) down, and (E) flood, identified by variable selection as important to predicting classification of tags as live or ghost. The *y* axis for each panel is the logit-transformed probability of being a live fish. A higher value indicates a higher probability of being alive (0 is equivalent to a 50% probability of being alive and 2 is equivalent to 88% probability). In panel E, a line is drawn to illustrate the decline in probability, but does not indicate that this is continuous variable. Dashed lines are equivalent to a 50% probability of being alive. Data are from the San Juan River study area in 2016 and 2017.
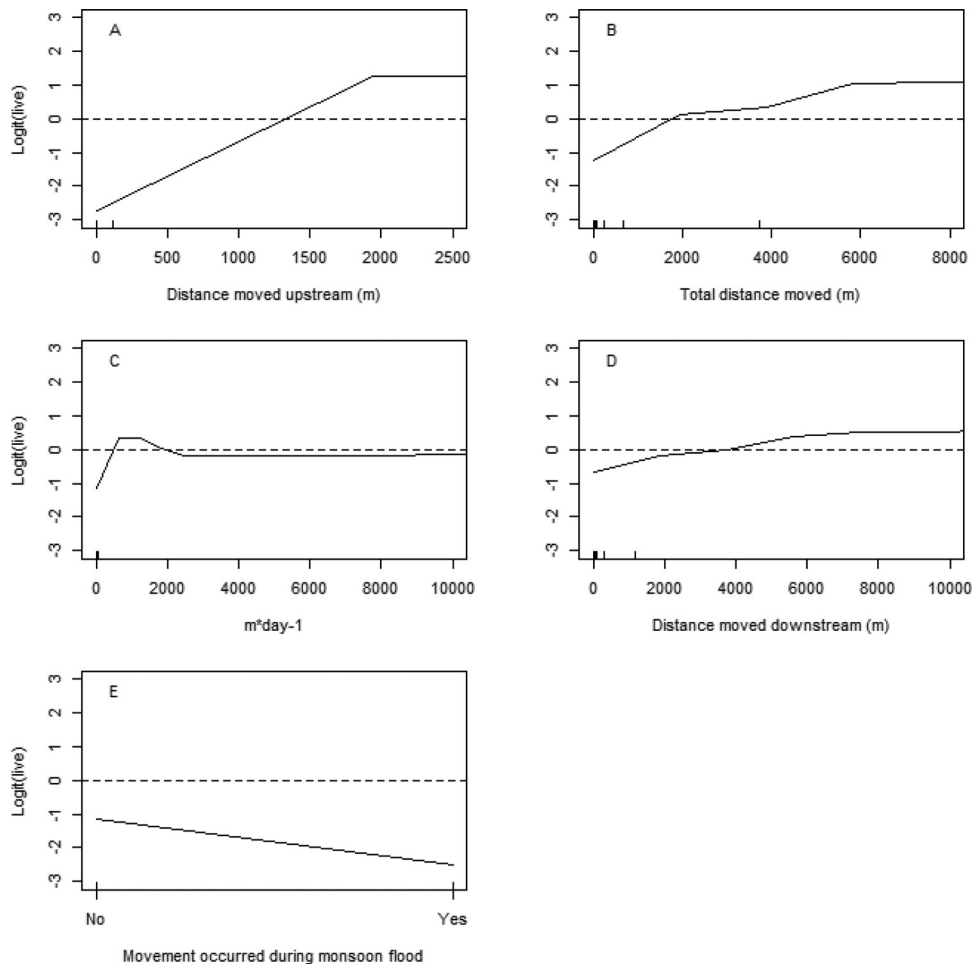
habitat use and selection studies, but they each have issues that can bias the results. Electrofishing can sacrifice some accuracy, as fish are disturbed before capture (Heggenes et al. 1990; Persinger et al. 2004), and snorkeling is limited by turbidity and cover, limiting visibility (Fausch and White 1981; Pert et al. 1997; Persinger et al. 2004). With PITPASS, we demonstrated the ability to cover large areas and detect largely undisturbed fish using mobile PIT tag detection methods. However, this method still requires the researcher to determine habitat type, but classification of habitat type can be both quick and accurate with proper training and calibration (Roper and Scarnecchia 1995).

Our study was the first to use movement and location data to classify PIT tags detected by a mobile sampling method. We were able to build a random forest model with a novel input data structure accounting for the "relatedness" of a detected tag's movements and identified the variables important for correct classification in our system and most likely other large sand bed rivers. The analytical framework we describe herein will be useful in developing similar models for other systems, even though we expect the relative importance of specific predictor variables to vary with location and species of interest. Future applications of this method should examine how the addition of individual characteristics (length, weight, sex, etc.) of the tagged fish affects the model's classification accuracy. While we believe this methodology can be an excellent contributor for habitat use studies, to quantify vital rates

(e.g., survival), other data sources or more intensive sampling will be required.

References

Al-Chokhachy, R., Budy, P., and Conner, M. 2009. Detecting declines in the abundance of a bull trout (*Salvelinus confluentus*) population: understanding

the accuracy, precision, and costs of our efforts. Can. J. Fish. Aquat. Sci. **66**(4): 649–658. doi:10.1139/F09-026.

Bernhardt, E.S., Palmer, M.A., Allan, J.D., Alexander, G., Barnas, K., Brooks, S., et al. 2005. Synthesizing US river restoration efforts. Science, **308**: 636–637. doi:10.1126/science.1109769. PMID:15860611.

Bezzerides, N., and Bestgen, K. 2002. Status review of roundtail chub *Gila robusta*, flannelmouth sucker *Catostomus latipinnis*, and bluehead sucker *Catostomus discobolus* in the Colorado River Basin. Report to the Bureau of Reclamation. Salt Lake City, Utah.

Bliesner, R., and Lamarra, V. 2000. Hydrology, geomorphology, and habitat studies. Report for San Juan River Basin Recovery and Implementation Program, Final Report, Logan, Utah.

Bond, R.M., Nicol, C.L., Kiernan, J.D., and Spence, B.C. 2019. Occurrence, fate, and confounding influence of ghost PIT tags in an intensively monitored watershed. Can. J. Fish. Aquat. Sci. **76**(2): 289–298. doi:10.1139/cjfas-2017-0409.

Budy, P.B., Conner, M.M., Salant, N.L., and Macfarlane, W.W. 2015. An occupancy-based quantification of the highly imperiled status of desert fishes of the southwestern United States. Conserv. Biol. J. **29**: 1142–1152. doi:10.1111/cobi.12513.

Burke, B.J., and Jepson, M.A. 2006. Performance of passive integrated transponder tags and radio tags in determining dam passage behavior of adult chinook salmon and steelhead. N. Am. J. Fish. Manage. **26**: 742–752. doi:10.1577/M05-138.1.

Cathcart, C.N., Gido, K.B., and McKinstry, M.C. 2015. Fish community distributions and movements in two tributaries of the San Juan River, USA. Trans. Am. Fish. Soc. **144**: 1013–1028. doi:10.1080/00028487.2015.1054515.

Clark, S.R., Conner, M.M., Durst, S.L., and Franssen, N.R. 2018. Age-specific estimates indicate potential deleterious capture effects and low survival of stocked juvenile Colorado pikeminnow. N. Am. J. Fish. Manage. **38**: 1059–1074. doi:10.1002/nafm.10214.

Cutler, D.R., Edwards, T.C., Jr, Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., and Lawler, J.J. 2007. Random forests for classification in ecology. Ecology, **88**: 2783–2792. doi:10.1890/07-0539.1. PMID:18051647.

Dudgeon, C.L., Pollock, K.H., Braccini, J.M., Semmens, J.M., and Barnett, A. 2015. Integrating acoustic telemetry into mark–recapture models to improve the precision of apparent survival and abundance estimates. Oecologia, **178**: 761–772. doi:10.1007/s00442-015-3280-z. PMID:25740335.

Dwyer, W.P., and White, R.G. 1997. Effect of electroshock on juvenile artic grayling and Yellowstone cutthroat trout growth, 100 days after treatment. N. Am. J. Fish. Manage. **17**: 174–177. doi:10.1577/1548-8675(1997)017<0174:MBEOEO>2.3.CO;2.

Elrahman, S.M.A., and Abraham, A. 2013. A review of class imbalance problem. J. Net. Innov. Comp. **1**: 332–340.

Evans, N.T., Shirey, P.D., Wieringa, J.G., Mahon, A.R., and Lamberti, G.A. 2017. Comparative cost and effort of fish distribution detection via environmental DNA analysis and electrofishing. Fisheries, **42**: 90–99. doi:10.1080/03632415.2017.1276329.

Fausch, K.D., and White, R.J. 1981. Competition between brook trout (*Salvelinus fontinalis*) and brown trout (*Salmo trutta*) for position in a Michigan stream. Can. J. Fish. Aquat. Sci. **38**(10): 1220–1227. doi:10.1139/f81-164.

Fetherman, E.R., Avila, B.W., and Winkelman, D.L. 2014. Raft and floating radio frequency identification (RFID) antenna systems for detecting and estimating abundance of PIT-tagged fish in rivers. N. Am. J. Fish. Manage. **34**: 1065–1077. doi:10.1080/02755947.2014.943859.

Fetherman, E.R., Winkelman, D.L., Bailey, L.L., Schisler, G.J., and Davies, K. 2015. Brown trout removal effects on short-term survival and movement of *Myxobolus cerebralis*-resistant rainbow trout. Trans. Am. Fish. Soc. **144**: 610–626. doi:10.1080/00028487.2015.1007166.

Fischer, J.R., Neebling, T.E., and Quist, M.C. 2012. Development and evaluation of a boat-mounted RFID antenna for monitoring freshwater mussels. Freshw. Sci. **31**: 148–153. doi:10.1899/11-045.1.

Gibbs, J.P., Droege, S., and Eagle, P. 1998. Monitoring populations of plants and animals. BioScience, **48**: 935–940. doi:10.2307/1313297.

Gido, K.B., and Propst, D.L. 2012. Long-term dynamics of native and nonnative fishes in the San Juan River, New Mexico and Utah, under a partially managed flow regime. Trans. Am. Fish. Soc. **141**: 645–659. doi:10.1080/00028487.2012.683471.

Havn, T.B., Okland, F., Teichert, M.A.K., Heermann, L., Borcherding, J., Saether, S.A., et al. 2017. Movements of dead fish in rivers. Anim. Biotelem. **5**: 7. doi:10.1186/s40317-017-0122-2.

Heggenes, J., Brabrand, A., and Saltveit, S.J. 1990. Comparison of three methods for studies of stream habitat use by young brown trout and Atlantic salmon. Trans. Am. Fish. Soc. **119**: 101–111. doi:10.1577/1548-8659(1990)119<0101:COTMFS>2.3.CO;2.

Hewitt, D.A., Janney, E.C., Hayes, B.S., and Shively, R.S. 2010. Improving inferences from fisheries capture–recapture studies through remote detection of PIT tags. Fisheries, **35**: 217–231. doi:10.1577/1548-8446-35.5.217.

Hodge, B.W., Henderson, R., Rogers, K.B., and Battige, K.D. 2015. Efficacy of portable PIT detectors for tracking long-term movement of Colorado River cutthroat trout in a small montane stream. N. Am. J. Fish. Manage. **35**: 605–610. doi:10.1080/02755947.2015.1012280.

Howell, P.J., Colvin, M.E., Sankovich, P.M., Buchanan, D.V., and Hemmingsen, A.R. 2016. Life histories, demography, and distribution of a fluvial bull trout population. Trans. Am. Fish. Soc. **145**: 173–194. doi:10.1080/00028487.2015.1105870.

Hunt, T.A., Ward, D.L., Propper, C.R., and Gibb, A.C. 2012. Effects of capture by

trammel net on Colorado River native fishes. J. Fish Wild. Manage. **3**: 133–141. doi:10.3996/122011-JFWM-070.

Irving, D.B., and Modde, T. 2000. Home-range fidelity and use of historic habitat by adult Colorado pikeminnow (*Ptychocheilus lucius*) in the White River, Colorado and Utah. West. North Am. Nat. **60**: 16–25.

Khoshgoftaar, T.M., Golawala, M., and Hulse, J.V. 2007. An empirical study of learning from imbalanced data using random forest. *In* The 19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2007), 29–31 Oct. 2007, Patras, Greece, Vol. 2. pp. 310–317. doi:10.1109/ICTAI.2007.46.

Liaw, A., and Wiener, M. 2002. Classification and regression by random Forest. R News, **2**(3): 18–22.

Lyon, J.P., Bird, T., Nicol, S., Kearns, J., O'Mahony, J., Todd, C.R., et al. 2014. Efficiency of electrofishing in turbid lowland rivers: implications for measuring temporal change in fish populations. Can. J. Fish. Aquat. Sci. **71**(6): 878–886. doi:10.1139/cjfas-2013-0287.

Lytle, D.A., and Poff, N.L. 2004. Adaptation to natural flow regimes. Trends Ecol. Evol. **19**: 94–100. doi:10.1016/j.tree.2003.10.002.

Maxwell, D., and Jennings, S. 2005. Power of monitoring programmes to detect decline and recovery of rare and vulnerable fish. J. Appl. Ecol. **42**: 25–37. doi:10.1111/j.1365-2664.2005.01000.x.

McCarthy, M.S., and Minckley, W.L. 1987. Age estimation for razorback sucker (Pisces: Catostomidae) from Lake Mohave, Arizona and Nevada. J. Ariz.-Nev. Acad. Sci. **21**: 87–97.

McKinney, T., Persons, W.R., and Rogers, R.S. 1999. Ecology of flannelmouth sucker in the Lee's Ferry Tailwater, Colorado River, Arizona. Great Basin Nat. **59**: 259–265.

Mesa, M.G., and Schreck, C.B. 1989. Electrofishing mark–recapture and depletion methodologies evoke behavioral and physiological changes in cutthroat trout. Trans. Am. Fish. Soc. **118**: 644–658. doi:10.1577/1548-8659(1989)118<0644:EMADME>2.3.CO;2.

Muchlinski, D., Siroky, D., He, J., and Kocher, M. 2016. Comparing random forest with logistic regression for predicting class-imbalanced civil war onset data. Pol. Anal. **24**: 87–103. doi:10.1093/pan/mpv024.

Muhametsafina, A., Midwood, J.D., Bliss, S.M., Stamplecoskie, K.M., and Cooke, S.J. 2014. The fate of dead fish tagged with biotelemetry transmitters in an urban stream. Aquat. Ecol. **48**: 23–33. doi:10.1007/s10452-013-9463-y.

O'Donnell, M.J., Horton, G.E., and Letcher, B.H. 2010. Use of portable antennas to estimate abundance of PIT-tagged fish in small streams: factors affecting detection probability. N. Am. J. Fish. Manage. **30**: 323–336. doi:10.1577/M09-008.1.

Osmundson, D.B., and White, G.C. 2017. Long-term mark–recapture monitoring of a Colorado pikeminnow *Ptychocheilus lucius* population: assessing recovery progress using demographic trends. Endanger. Species Res. **34**: 131–147. doi:10.3354/esr00842.

Osmundson, D.B., Ryel, R.J., and Mourning, T.E. 1997. Growth and survival of Colorado squawfish in the Upper Colorado River. Trans. Am. Fish. Soc. **126**: 687–698. doi:10.1577/1548-8659(1997)126<0687:GASOCS>2.3.CO;2.

Parma, A.M. 1998. What can adaptive management do for our fish, forests, food, and biodiversity? Integr. Biol. **1**: 16–26. doi:10.1002/(SICI)1520-6602(1998)1:1<16::AID-INBI3>3.0.CO;2-D.

Paukert, C.P., Ward, D.L., Sponholtz, P.J., and Hilwig, K.D. 2005. Effects of repeated hoopnetting and handling on bonytail chub. J. Freshw. Ecol. **20**: 649–653. doi:10.1080/02705060.2005.9664787.

Persinger, J.W., Orth, D.J., and Newcomb, T.J. 2004. A comparison of snorkeling versus throwable anode electrofishing for evaluating stream fish habitat use. J. Freshw. Ecol. **4**: 547–557. doi:10.1080/02705060.2004.9664734.

Pert, E.J., Deinstadt, J.M., and Erman, D.C. 1997. Evidence of differential detectability of lotic adult rainbow and brown trout: implications for habitat use inference from direct-underwater observation. J. Freshw. Ecol. **12**: 359–365. doi:10.1080/02705060.1997.9663546.

Pine, W.E., Pollock, K.H., Hightower, J.E., Kwak, T.J., and Rice, J.A. 2003. A review of tagging methods for estimating fish population size and components of mortality. Fisheries, **28**: 10–23. doi:10.1577/1548-8446(2003)28[10:AROTMF]2.0.CO;2.

Pine, W.E., III, Martell, S.J.D., Walters, C.J., and Kitchell, J.F. 2009. Counterintuitive responses of fish populations to management actions: some common causes and implications for predictions based on ecosystem modeling. Fisheries, **34**: 165–180. doi:10.1577/1548-8446-34.4.165.

Reynolds, J.B., Kolz, A.L., Zale, A.V., Parrish, D.L., and Sutton, T.M. 2012. Electrofishing. *In* Fisheries techniques. 3rd ed. American Fisheries Society. pp. 305–361.

Richer, E.E., Fetherman, E.R., Kondratieff, M.C., and Barnes, T.A. 2017. Incorporating GPS and mobile radio frequency identification to detect PIT-tagged fish and evaluate habitat utilization in streams. N. Am. J. Fish. Manage. **37**: 1249–1264. doi:10.1080/02755947.2017.1374312.

Roper, B.B., and Scarnecchia, D.L. 1995. Observer variability in classifying habitat types in stream surveys. N. Am. J. Fish. Manage. **15**: 49–53. doi:10.1577/1548-8675(1995)015<0049:OVICHT>2.3.CO;2.

Ruppert, J.B., and Muth, R.T. 1997. Effects of electrofishing fields on captive juveniles of two endangered cyprinids. N. Am. J. Fish. Manage. **17**: 314–320. doi:10.1577/1548-8675(1997)017<0314:EOEFOC>2.3.CO;2.

Schramm, H.L., Jr Grado, S.C., and Pugh, L.L. 2002. The costs of sampling fishes in riverine habitats of a large river. Fish. Res. **56**: 51–57. doi:10.1016/S0165-7836(01)00316-2.

Scoppettone, G.G. 1988. Growth and longevity of the cui-ui and longevity of other catostomids and cyprinids in western North America. Trans. Am. Fish. Soc. **117**: 301–307. doi:10.1577/1548-8659(1988)117<0301:GALOTC>2.3.CO;2.

Snook, E.L., Letcher, B.H., Dubreuil, T.L., Zydlewski, J., O'Donnell, M.J., Whitely, A.R., et al. 2016. Movement patterns of Brook Trout in a restored coastal stream system in southern Massachusetts. Ecol. Freshw. Fish, **25**: 360–375. doi:10.1111/eff.12216.

Snyder, D.E. 2003. Invited overview: conclusions from a review of electrofishing and its harmful effects on fish. Rev. Fish Biol. Fish. **13**: 445–453. doi:10.1007/s11160-004-1095-9.

STReaMS. 2018. STReaMS — The Species Tagging, Research and Monitoring System [online]. Species Tagging, Research and Monitoring System. Available from https://streamsystem.org [accessed 28 April 2018].

Stout, J.B., Conner, M.M., Budy, P., Mackinnon, P.D., and McKinstry, M.M. 2019. We ain't afraid of no ghosts: Tracking habitat interactions and movement dynamics of ghost PIT tags under differing flow conditions in a sand bed river. N. Am. J. Fish. Manage. **39**(6): 1337–1347. doi:10.1002/nafm.10371.

US Fish and Wildlife Service. 1990. Colorado squawfish recovery plan. US Fish and Wildlife Service, Denver, Colorado.

US Fish and Wildlife Service. 1998. Razorback sucker (*Xyrauchen texanus*) recovery plan. Denver, Colorado.

Walther, B.A., and Moore, J.L. 2005. The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. Ecography, **28**: 815–829. doi:10.1111/j.2005.0906-7590.04112.x.

Webber, P.A., and Beers, D. 2014. Detecting razorback suckers using passive integrated transponder tag antennas in the Green River, Utah. J. Fish Wildl. Manage. **5**: 191–196. doi:10.3996/122012-JFWM-104.

Webber, P.A., Thompson, P.D., and Budy, P. 2012. Status and structure of two populations of the bluehead sucker (*Catostomus discobolus*) in the Weber River, Utah. Southwest. Nat. **57**: 267–276. doi:10.1894/0038-4909-57.3.267.

Zelasko, K.A., Bestgen, K.R., and White, G.C. 2010. Survival rates and movement of hatchery-reared razorback suckers in the upper Colorado River basin, Utah and Colorado. Trans. Am. Fish. Soc. **139**: 1478–1499. doi:10.1577/T09-107.1.

Appendix A

Introduction

Imbalance in class size refers to a condition in the data where one class is represented by a large number of examples and another class is represented by relatively few. Imbalance in class sizes can cause model overfitting, which results in poor model generalizability and an inability to predict future observations accurately. However, random forest models have been shown to perform well despite imbalance in class sizes (Khoshgoftaar et al. 2007; Elrahman and Abraham 2013; Muchlinski et al. 2016). However, a few methods have been developed to deal with class size imbalance. Up-sampling and down-sampling to obtain equal sample sizes has been used as a solution to mitigate imbalances (Japkowicz 2000). Down-sampling consists of sampling a subset of the majority class until it matches the size of the minority class. Up-sampling consists of sampling with replacement until the minority class has as many samples as the majority class (Japkowicz 2000). In this appendix, we use our existing dataset and use our model structure to both up- and down-sample to determine the effect of an unequal sample size.

Methods

To down-sample, we used our original dataset and the dependent model structure described in the manuscript. We chose a random sample with replacement for both classes (live and ghost) to generate a subset with 300 observations for testing our model using the sample function in R. We used the randomForest function in package randomForest in R (Liaw and Wiener 2002) to create our model using the default parameters (500 trees and four variables tried per split). We used all of the predictor variables to build our model and then examined the classification error rates. We repeated the process 100 times and report the mean error rates and variance of all the model runs.

To up-sample, we used the original dataset and the dependent model structure described in the manuscript. We used a random sample with replacement for both classes to generate a subset with 900 observations in each class for building our model using the sample function in R. We used the randomForest function in package randomForest in R (Liaw and Wiener 2002) to create our

Table A1. Mean error rates and variance for up- and down-sampled models for the two classes (live and ghost) and the overall rates.

Sample	Mean ghost error (%)	Ghost variance	Mean live error (%)	Live variance	Mean overall error (%)	Overall variance
Down	4.9	1.6	7.9	1.8	6.7	0.8
Up	2.1	0.2	2.1	0.3	2.3	0.1
Unequal	1.6	NA	21.8	NA	6.8	NA

Note: The unequal sample is the original dependent model in the main manuscript. In the unequal model, the error rates are not a mean and no variance is reported, as it is a single model.

models using the default parameters (500 trees and four variables tried per split). Our model included all of the predictor variables and was used to determine our classification error rates. We repeated the process 100 times and report the mean error rates and variance of all the model runs.

Results

Down-sampling reduced the mean error rate of the live class and increased the mean error rate of the ghost class (Table A1) compared with the original model. The mean overall error rate of the down-sampled model was very similar to the error rate of the model with unequal class sizes. Up-sampling also reduced the mean error rate of the live class and increased the mean error rate of the ghost class compared with the error rate of the model with unequal class sizes. However, the mean overall error rate of the up-sampled models was reduced compared with the error rate of the model with unequal class sizes. For all comparisons, variance was higher in the down-sampled models when compared with the up-sampled models.

Discussion

While error rates changed as a result of up- and down-sampling, there is still error in the model. Down-sampling resulted in a similar overall error rate to the original model, suggesting the error was shifted from the live class to the ghost class, but without any substantial overall improvement. Up-sampling also shifted error from the live class to the ghost class, but demonstrated improvement in the overall error rate. The variance of the up-sampled models was lower than that of the down-sampled models, but we believe this is an artifact of a much larger sample size used to create the models.

Shifting the error between classes can also be done through the manipulation of the cutoff value in the random forest if it is more important to correctly classify one specific class. Random forest assigns a class to specific observations based on a vote tally from all of the trees created in the forest. The default cutoff value for classification of an observation is 1 divided by the number of classes (Liaw and Wiener 2002), meaning the cutoff is 50% if there are two classes. The cutoff value can be changed to the user's specifications, resulting in greater certainty about correct classification of a specific class, but since one class was not more important than another in our case, we left the cutoff value at the default.

References

Elrahman, S.M.A., and Abraham, A. 2013. A review of class imbalance problem. J. Net. Innov. Comp. **1**: 332–340.

Japkowicz, N. 2000. The class imbalance problem: Significance and strategies. *In* Proceedings of the 2000 International Conference on Artificial Intelligence (ICAI), June 2000, Las Vegas, Nevada.

Khoshgoftaar, T.M., Golawala, M., and Hulse, J.V. 2007. An empirical study of learning from imbalanced data using random forest. *In* The 19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2007), 29–31 Oct. 2007, Patras, Greece, Vol. 2. pp. 310–317. doi:10.1109/ICTAI.2007.46.

Liaw, A., and Wiener, M. 2002. Classification and Regression by random Forest. R News, **2**(3): 18–22.

Muchlinski, D., Siroky, D., He, J., and Kocher, M. 2016. Comparing random forest with logistic regression for predicting class-imbalanced civil war onset data. Pol. Anal. **24**: 87–103. doi:10.1093/pan/mpv024.