IDENTIFYING MIGRATION FATE AND FACTORS CONTRIBUTING TO MORTALITY OF ATLANTIC SALMON SMOLTS

By

Daniela V. Notte

Submitted in partial fulfillment of the requirements for the degree of Master of Science

At

Dalhousie University

Halifax, Nova Scotia

November 2021

Dalhousie University is located in Mi'kma'ki, the ancestral and unceded territory of the Mi'kmaq. We are all Treaty people.

Table of Contents

List of Tables

List of Figures

Abstract

Atlantic Salmon populations are in decline throughout their native distribution largely due to poor estuarine and marine survival. Predation is a significant source of salmon smolt mortality during migration from freshwater to marine environments. This thesis investigates potential mechanisms of predation and other mortality in a population of Endangered inner Bay of Fundy Atlantic Salmon smolts over three years. Predated smolts were identified through the use of novel acoustic predation tags combined with machine learning algorithms trained to differentiate predator and prey behaviour. From 2017 to 2019, survival rates increased as predation rates decreased. Migration rate was identified as a behavioural mechanism of mortality where slower migrating smolts were more likely to be predated. No physiological mechanism of mortality was identified through analyses of host gene expression and pathogen presence. Predation of salmon smolts in this population is highly variable between years and appears to be more opportunistic than selective.

List of Abbreviations Used

ΔAIC	Delta Akaike information criteria
AIC	Akaike information criteria
ANOVA	Analysis of variance
APC	Artificial positive construct
CA4	Carbonic anhydrase 4
CCL19	C-C motif chemokine 19
CCL4	C-C motif chemokine 4
cDNA	Complementary deoxyribonucleic acid
CJS	Cormack-Jolly-Seber
CLEC4E	C-type lectin receptor A
COSEWIC	Council on the Status of Endangered Wildlife in Canada
Ct	Cycle threshold
DFO	Fisheries and Oceans Canada
DNA	Deoxyribonucleic acid
FKBP5	FK506-binding protein
HSP90	Heat shock protein 90
iBoF	Inner Bay of Fundy
IHNV	Infectious hematopoietic necrosis virus
IL12B	Interleukin-12 beta
LGB	Live Gene Bank
MANOVA	Multivariate analysis of variance
NKAa1.b	Na^+/K^+-ATPase alpha subunits 1b
NKAA1C	Na^+/K^+-ATPase alpha subunits 1c
OOB	Out-of-bag
PC	Principal component
PCA	Principal components analysis
PCR	Polymerase chain reaction
qPCR	Quantitative polymerase chain reaction
RNA	Ribonucleic acid
SARA	Species At Risk Act
SHOP21	Salmon hyperosmotic protein 21
STA	Specific target amplification
VDD	Viral disease development

Statement

Chapter 2: A version of this chapter is in review at a peer-reviewed journal with the following citation: Notte DV, Lennox RJ, Hardie DC, Crossin GT. Application of machine learning and acoustic predation tags to classify migration fate of Atlantic salmon smolts. DVN performed field work, contributed to concept development, performed analyses, and wrote initial manuscript drafts. RJL contributed to concept development and manuscript drafts. DCH performed field work, developed study design, and contributed to manuscript drafts. GTC contributed to manuscript drafts. Ethics Approval: All animal experiments were approved by the Canadian Committee on Animal Care, via permits issued by Fisheries and Oceans Canada (Maritimes Region Animal Care Committee Animal Utilization Protocols 17-16, 18-13, 19-10) and by Dalhousie University (University Committee on Lab Animals permit 18-126).

Acknowledgements

Chapter 1: Introduction

Atlantic Salmon (*Salmo salar*) is a significant species in Canada due to their commercial and recreational value (ASF, 2011), spiritual importance to Indigenous Peoples (DFO, 2018a), and ecological role transferring nutrients between freshwater and marine environments (Jonsson & Jonsson, 2003). However, Atlantic Salmon population numbers have declined throughout their native distribution (Parrish et al., 1998). In Canada, there are 16 Atlantic Salmon designatable units, five of which have been assessed as Endangered under COSEWIC and one is Extinct (SARA, 2021). Main threats to Atlantic Salmon include environmental degradation, habitat fragmentation, and over-exploitation (Parrish et al., 1998; DFO, 2010). Researchers have identified the out-migration and early marine stages of the Atlantic Salmon life cycle as a period of high mortality (LaCroix, 2008; Thorstad et al., 2011; Halfyard, 2014).

Atlantic Salmon are primarily an anadromous species, migrating from freshwater to marine environments as juveniles (smolts) then returning to their natal freshwater habitats 1-3 years later as adults to spawn (McCormick et al., 1998). Mortality is high during smolt out-migration, potential contributing factors include water pollution, rising water temperatures, stress from the smoltification process, high energy expenditure required to migrate, exposure to novel pathogens combined with decreased immune function, and increased predator presence (Parrish et al., 1998; DFO, 2010; Altizer et al., 2011; Furey, 2016).

Predation is a significant contributor to total salmon smolt and post-smolt mortality (Handeland et al., 1996; LaCroix, 2008; Halfyard, 2014; Furey, 2016). However, predation is extremely difficult to observe in the natural environment, limiting the ability to determine the effect predation is having on a population or what factors may be influencing predation rates. One novel approach to studying predation is based on the development of acoustic predation tags

(InnovaSea Systems Inc., Bedford, NS), these tags can be used to track a smolt's migration progress and determine if a predation event has occurred through changes in pH (Halfyard et al., 2017).

This thesis will aim to quantify smolt predation rates in an Endangered Atlantic Salmon population and identify potential behavioural and physiological mechanisms of predation susceptibility by combining predation tag technology with transcriptomic analyses.

1.1 Migration

Migration is a specialized form of movement and a key life history strategy developed by species across animal taxa (Wilcove & Wikelski, 2008; Milner-Gulland et al., 2011). Migration is characterized by longer and more direct movements than ranging or station keeping, initial suppression of responses to stimuli, unique activity during departure and arrival, cues to initiate departure, and specific resource allocation to support movement (Dingle, 2014). Migration has both genetic and adaptive components. An individual is genetically pre-disposed to migration, the cues that initiate migration and migratory behaviour are based in the genome (Dingle, 2014). Additionally, the biotic and abiotic environment to which the individual is adapted to also influences migratory behaviour (Dingle, 2014). For example, the age at which migration occurs may differ between populations based on the productivity of the starting environment and therefore the amount of time it takes to reach a growth threshold (Hendry et al., 2004). Exact timing of migration may also differ between populations and years due to differences in timing of environmental cues, such as temperature or photoperiod.

Migration strategy can vary significantly between species based on the reason for migration, resource needs, and temporal variability in a habitat. A main cause for migration is to

take advantage of resources or avoid threats on different spatiotemporal scales (Wilcove & Wikelski, 2008; Milner-Gulland et al., 2011). Seasonal migrations, based largely on resource availability, typically occur between foraging and breeding grounds (Dingle, 2014). For example, the Arctic tern (*Stern paradisaea*) annually migrates from breeding grounds in the Arctic to highly productive wintering feeding grounds in the Southern Ocean (Egevang et al., 2010). The need for migration can also arise from differences in habitat and food requirements between juvenile and adult life stages (Milner-Gulland et al., 2011). In some amphibians, breeding adults migrate to aquatic habitats because water is required for their young (Dingle, 2014). Predation risk is another cause for migration, some shorebirds migrate from temperate regions to the Arctic where mammalian predators are less abundant (Dingle, 2014). Species also conduct migrations to refugia during vulnerable periods in their life history, this includes waterfowl which molt flight feathers and the many insects, amphibians, reptiles, and mammals that have dormancy periods (Dingle, 2014).

Diverse migration strategies also exist within species, such as partial migration, in which individuals of a population that typically migrate instead remain resident (Chapman et al., 2011; Skov et al., 2011). This migration dimorphism is largely due to the trade-offs associated with migration. Migration is associated with higher costs and risks than more stationary periods, but typically occupies a short part of the life cycle resulting in concentrated periods of mortality (Sergio et al., 2019). Migration is energetically expensive, can leave an organism exposed to novel predators and pathogens, and has the potential to delay reproduction and shorten lifespan (Milner-Gulland et al., 2011). However, residents may face high levels of competition and poor food availability resulting in reduced growth and fecundity. Therefore, migration strategy largely depends on balancing survival and reproductive success.

Worldwide, migratory species are facing population declines due to poor survival during outbound and/or return migrations, habitat fragmentation, and physical barriers to migration (Wilcove & Wilkelski, 2008). Migration poses a challenge for conservation efforts due to difficulties determining areas of high mortality along the migration route, determining causes of mortality, and protecting large geographic areas that often cross national boundaries (Wilcove & Wilkelski, 2008). Additionally, migration ecology can differ significantly between populations or stocks of the same species (Wilcove & Wilkelski, 2008). Understanding migratory behaviour and contributing factors is vital to improving population conservation and management.

Salmonids (family Salmonidae) are among the most widely studied migratory animal taxa due to their impressive migration ecology, recreational and commercial importance, and significant population declines (Milner-Gulland et al., 2011). Anadromous salmonids exhibit the range of diverse migration strategies discussed above. For example, migration distance can vary widely among Canadian Atlantic Salmon populations, most of which migrate offshore to Greenland and the Labrador Sea, while the inner Bay of Fundy population remains coastal (DFO, 2019a). Juvenile salmon (smolts) emigrate from oligotrophic natal freshwater environments when they have reached a body size or growth rate threshold in favour of highly productive marine environments to achieve greater growth (Hendry et al., 2004). The downstream migration undertaken by smolts to reach marine feeding grounds is a major mortality bottleneck in the life cycle of salmonids (Thorstad et al., 2011). Partial migration is also seen in many salmonid species; resident individuals lack the sexual advantages (large size, courtship behaviour, etc.) of anadromous individuals but avoid high rates of migration mortality (Thorstad et al., 2011). Residency is favoured in populations or years where freshwater growth is high (Hendry et al., 2004).

The smoltification process, during which an anadromous salmonid transitions from the parr to smolt stage, involves several energetically expensive and stressful morphological, behavioural, and physiological changes for individuals to become suited for the marine environment. These changes include a more fusiform body shape, countershading colouring with silver sides, pelagic swimming, and shoaling behaviour (Thorstad et al., 2011). Several environmental factors control the timing of smoltification and out-migration, including temperature, photoperiod, and water discharge (Vollset et al., 2021; McCormick et al., 1998). The type and timing of environmental cues may vary among populations due to local adaptation. There is a physiological and environmental smolt window. The probability of marine survival increases when smolts are physiologically prepared for migration, and migration timing coincides with optimal environmental conditions (McCormick et al., 1998). Delays in migration lead to lower survival because migration timing is now outside the optimal environmental window. One of the most crucial aspects of the physiological smolt window involves a restructuring of osmoregulatory systems to prepare individuals for salt water. Principally, this leads to the development of seawater tolerance via the cellular reorganization of gill, gut, and kidney, and is accompanied by an increase in gill Na^+/K^+-ATPase activity which regulates the increase in plasma ions after entering salt water (Zadunaisky, 1996; Bjornsson et al., 2010). The transition from fresh water to salt water is extremely stressful and entering salt water before these osmoregulatory changes have been completed can negatively impact smolt homeostasis, behaviour, and survival (Handeland et al. 1996; Dieperink et al., 2002; Halfyard et al., 2012; Halfyard et al., 2013). The optimal smolt window and smolt survival is also affected by spatiotemporal differences in the abundance of predators, disease, parasites, and food availability (McCormick et al., 1998).

1.2 Predation

Predators within many animal taxa have evolved to take advantage of seasonably predictable prey migrations to varying extents. The coupling of predator and prey migrations, defined as predator movements beyond a home range to track or intercept migrating prey, is a large-scale movement in response to prey influxes (Furey et al., 2018). For example, grey wolves (*Canis lupus*) will track migrating caribou (*R. tarandus groenlandicus*) for up to 500 km (Furey et al., 2018). On a smaller scale, piscine, avian, and mammalian predators are known to aggregate in certain areas, such as spatial bottlenecks, to take advantage of mass migrations (Furey et al., 2015; Daniels et al., 2019; Flavio et al., 2019). Bull Trout (*Salvelinus confluentus*) have been shown to aggregate and binge feed on pulses of Pacific salmon (*Oncorhynchus spp.*) smolts exiting the Chilko River, British Columbia (Furey, 2016). Other predators will time reproductive events to coincide with high prey availability (Furey et al., 2018).

Lima and Dill (1990) reviewed the ability of animals to make behavioural decisions based on perceived predation risk, concluding that not only is predation a strong selective force over evolutionary time but also over the lifetime of an individual. Decisions on what, where, and when to feed are made based on trade-offs between energy gain and predation risk. This decision-making process often leads to diurnal feeding patterns and shifts in habitat choice to avoid predators. Changes in prey behaviours, such as feeding and habitat use, are an indirect impact of predation on individual prey (Bender, 2018).

Similarly, predation risk largely influences migration behaviour. In recent years, barnacle geese (*Branta leucopsis*) have delayed migration from their wintering to breeding grounds to avoid increased predator populations at stopover sites (Jonker et al., 2010). Skov et al. (2011) found that in partially migrating populations of Common Bream (*Abramis brama*), smaller

individuals are more likely to migrate from lakes to streams where predation risk is lower than larger individuals. Salmon smolts primarily migrate at night to avoid visual predators but have been shown to switch to day migrations in turbid waters and towards the end of the smolt run (Ibbotson et al., 2011; Clark et al., 2016; Furey et al., 2016; Flavio et al., 2019). Predation has also been linked to the evolution of sociality (Lima & Dill, 1990). Salmon, which are territorial prior to the smolt stage, will synchronize their migration to travel in groups and swamp predators (Thorstad et al., 2011; Furey et al., 2016). Migratory prey species may also use social cues as a proxy for unfamiliar predator cues as they encounter new predator species along the migration route (Sabal et al., 2021).

For migratory prey species, the predation landscape is spatially and temporally variable (Sabal et al., 2021). Variation in predator communities and habitat features along the migration route impact mortality risk. For example, high predator abundance in a confined habitat confers a high mortality risk. Antipredator response is determined by the perception of this risk and is balanced between the costs and benefits of the response (Sabal et al., 2021). Antipredator responses can have energetic costs, delay migration, and may come at the expense of feeding or navigation opportunities.

Body size is one intrinsic factor that has been shown to influence predator avoidance behaviours and predation risk. Ibbotson et al. (2011) found that smaller Atlantic Salmon smolts are more likely to conduct nocturnal migrations to avoid visual predators than larger bodied smolts. Additionally, it has been shown that smolts of poor body condition and smaller size are more susceptible to predation by avian predators than larger smolts (Hostetter, 2009; Tucker et al., 2016). In contrast, Furey et al. (2016) found that body size was not a significant determinant

of predation susceptibility in Chilko salmon smolts, but predator swamping and nocturnal migration did effectively reduce predation risk.

Predation events occurring during migrations can have a significant, negative impact on salmon population numbers, especially for endangered populations (Grout, 2006; LaCroix, 2008). The rate of predation on salmon smolts can vary between years and rivers, as well as between different areas within a single river (LaCroix, 2008; Thorstad et al., 2012; Halfyard et al., 2013). Predation rates are often highest when down-river migrating smolts reach the head of tide or the estuary, due to increased physiological stress from the change in salinity and predator aggregations (Handeland et al., 1996; Dieperink et al., 2002; LaCroix, 2008; Thorstad et al., 2012; Halfyard et al., 2013; Daniels et al., 2019).

However, predation is not always detrimental to a population. The effect of predation on a prey population is variable depending on the circumstances (Bender, 2018). The act of predation can affect the individual, the population, and the community. The direct effect of predation on the individual, mortality or survival, is often dictated by that individual's predisposition to mortality. For example, an individual with poor body condition may have a limited ability to flee from a predator and is therefore predisposed to mortality. Predisposition is a mechanism of mortality and can be based on the physiological condition of the individual as well as environmental conditions (Bender, 2018). When a predated individual has a predisposition to mortality, predation is the proximate cause of mortality, but the underlying factors are the ultimate cause. Compensatory predation at the population level occurs when the predated individuals have a high predisposition to mortality and were likely to die regardless of predation occurring (Bender, 2018). In this sense predation is not adding to total mortality but substituting other causes of mortality (Creel, 2011; Bender, 2018). In compensatory mortality

when predation rates increase, rates of other mortality sources decrease and vice versa (Bender, 2018; Payton et al., 2020). Predation may benefit a prey population by removing weak or disease carrying individuals from a population and reducing intraspecific competition (Miller et al., 2014; Furey, 2016; Bender, 2018; Sabal et al., 2021). If survival in a population is density-dependent, predation is more likely to be compensatory as the removal of individuals makes more resources available for the surviving individuals in the population (Bender, 2018).

Conversely, additive mortality is when predation adds to total mortality thus decreasing survival rates and leading to declines in population numbers (Bender, 2018). When individuals of high fitness that have a high probability of survival are predated, this is additive mortality (Payton et al., 2020). Additive mortality can still occur when there is predisposition to mortality if predation pressure is high due to an abundant predator population or multiple predatory species (Bender, 2018). There is a continuum from compensatory to additive mortality, the closer predation is to additive mortality, the greater the impact predation has on a population (Bender, 2018). Strong additive mortality that drives down population numbers can destabilize a community (removes part of food web, ecosystem services, etc.) while compensatory mortality tends to bring balance to a community.

The role of predation in applied ecology and conservation is both complex and controversial. Predators are a vital part of an ecosystem, providing top-down trophic control and/or ecosystem services (Ormerod, 2002). Predators may even be introduced to an area as a biological control to limit growth of a pest prey species (Ormerod, 2002). However, predators, either native or introduced, can also be viewed negatively when their prey is valuable to humans. In these cases, human intervention may be considered. A review of predator removal attempts

found that only 26% of cases where prey protection was the desired outcome were successful and typically only in the short-term (Lennox et al., 2018). Less direct methods of anti-predator intervention include habitat management to favour prey, exclusion of predators from certain areas, predator deterrents, and supplementary feeding of predators (Ormerod, 2002). Predator control or intervention is based off the assumption that predation is a source of additive mortality and that the removal of predators will increase prey population numbers (Bender, 2018; Payton et al., 2020). In any case, human intervention on predator-prey interactions must be considered carefully as unexpected outcomes with negative impacts on the ecosystem may occur.

1.3 Gene Expression and Pathogen Presence

Identifying common intrinsic and extrinsic factors that cause mortality or confer a predisposition to predation is vital for informed population management. Examining the expression of genes related to important physiological processes or that are markers of environmental stressors is one method of achieving this. The analysis of transcriptomes has linked gill Na^+/K^+-ATPase activity (osmoregulatory condition) to mortality in Atlantic Salmon smolt out-migrations (Stich et al., 2015) and Pacific salmon adult return-migrations (Miller et al., 2011; Hinch et al., 2012).

Additionally, testing for the presence of pathogens that may impact behaviour, movement ability, or overall health may reveal a source of increased predation susceptibility. Rhinoceros auklets (*Cerorhinca monocerata*) feed on Sockeye Salmon (*Oncorhynchus nerka*) smolts with greater pathogen diversity, pathogen load, and levels of *Parvicapsula* sp. than the population average (Miller et al., 2014). Steelhead trout (*O. mykiss*) in poor condition due to infections were highly consumed by avian predators (Hostetter, 2009). Additionally, sockeye smolts with infectious hematopoietic necrosis virus (IHNV) or infected by *Flavobacterium psychrophilum*

were more likely to be predated upon by Bull Trout (Furey, 2016). Behaviour, osmoregulation, and swim performance of fish can be impacted by pathogens, leaving an individual more susceptible to predation or other causes of mortality (Thorstad et al., 2012; Miller et al. 2014). In sea trout (*Salmo trutta*), parasites such as salmon lice (*Lepeophtheirus salmonis*) and *Icthyobodo* sp. can alter fish behaviour and are correlated to longer times spent in fresh water (Thorstad et al., 2015; Lennox et al., 2020).

Migratory species are exposed to more pathogens as they travel through new and diverse landscapes compared to more resident species (Altizer et al., 2011). Pathogen load has also been found to be positively correlated with distance travelled and accumulates throughout the migratory period (Altizer et al., 2011; Chapman et al., 2020). For migratory birds, stop-over sites are pathogen hotspots due to high densities and multiple species of birds. Immunosuppression is common in migratory species where innate immune responses are reduced in a trade-off for energetic demands of migration, leaving migrants more susceptible to pathogens and associated physiological impacts (Altizer et al., 2011; Miller et al., 2014).

External environmental factors are related to internal state; poor environmental conditions can negatively impact an individual's physiology, which can reduce their ability to escape a predator and have long-term negative impacts on fitness, i.e., carryover effects (Midwood et al., 2015; Furey, 2016). There is increasing concern about climate change and the negative impacts of thermal stress on fitness as well as its effect on changing distributions and growth rates of pathogens and predators (Miller et al., 2014). The cumulative interaction of internal and external stressors, pathogen load, and predation is likely the main contributor to high smolt mortality (Jeffries et al., 2014; Miller et al., 2014). High water temperatures have been related to premature mortality of returning Sockeye Salmon adults (Miller et al., 2011). Tank

studies have shown that the mortality threshold for pathogen load is lower at warmer water temperatures, likely due to decreased host resistance and increased stress at high temperatures (Miller et al., 2014).

Evaluating individual health and physiological state at the start of seaward migration combined with tracking both movement and survival to the early marine environment provides insight into delayed or sublethal effects of detrimental biotic or abiotic factors a smolt may be exposed to in their natal habitat. For example, the impacts of water contaminants may not be seen until entry into saltwater environments where there is increased stress placed on the osmoregulatory system (Thorstad et al., 2012). Additionally, the level of pathogenicity of some microbes varies between fresh and salt water, as is the case with IHNV (Miller et al., 2014). Factors that contribute to early marine mortality of salmon but occur in fresh water are easier to manage than marine threats with more immediate impact on survival. It is therefore crucial to identify and mitigate these factors with delayed effects to reduce mortality during the smolt and post-smolt stages (Thorstad et al., 2012).

1.4 Threats to Atlantic Salmon

Atlantic Salmon population numbers have declined throughout their native distribution (Parrish et al., 1998). Generally, in both Europe and North America, southern populations are at greater risk than northern populations. Most rivers with extirpated populations are in regions of high human population density (Parrish et al., 1998). In the freshwater environment, altered environmental conditions, water contaminants, river acidification, and barriers to fish movement (namely dams) negatively impact survival (Parrish et al., 1998; DFO, 2010). Threats in the marine environment are hypothesized to include increases in predator abundance, decrease in prey abundance, shifting water temperature regimes, aquaculture, and fisheries (Parrish et al.,

1998; DFO, 2010). However, a recent review suggests that illegal, unreported, and unregulated fisheries are the primary cause of low adult salmon returns (Dadswell et al., 2021).

1.5 Study System

The Stewiacke River, Nova Scotia is one of the 50 rivers within the Endangered inner Bay of Fundy (iBoF) Atlantic Salmon designatable unit (DFO, 2019b). In 2010, the Stewiacke was one of ten rivers listed as critical freshwater habitat for Atlantic Salmon and was included as part of the iBoF Salmon Recovery Plan (SARA, 2010). Between the years 1980 and 2000, annual adult returns of iBoF salmon declined from 400,000 to 250 (SARA, 2010). The iBoF Atlantic Salmon commercial fishery was closed in 1985, and recreational and Aboriginal fisheries were closed in 1990 (SARA, 2010). In an effort to restore dwindling salmon abundances, Fisheries and Oceans Canada (DFO) began stocking the Stewiacke River in 2003 with Atlantic Salmon fry produced through the Live Gene Bank (LGB) Program (DFO, 2018b). Cultured salmon transferred from the Coldbrook Biodiversity Facility are released at the fry stage to maximize wild exposure and natural selection prior to the high-risk smolt stage. The LGB maintains detailed pedigree records to ensure genetic diversity is maintained in stocked populations.

The Stewiacke River meets the Shubenacadie River 22 km upstream from the mouth of the Shubenacadie River which drains into the upper Minas Basin. These rivers are heavily influenced by the Bay of Fundy tides, the head-of-tide is located 15 km upstream on the Stewiacke River. iBoF Atlantic Salmon remain within the Bay of Fundy and Gulf of Maine to feed, and most Stewiacke salmon return to natal streams to spawn after one winter at sea (Gibson et al., 2015).

1.5.1 Potential Predators within Study Area

Piscine, avian, and mammalian predatory species of Atlantic Salmon smolts are found within the study area. In the freshwater stretches of the Stewiacke River, invasive Chain Pickerel (*Esox niger*), Smallmouth Bass (*Micropterus dolomieu*), and Brown Trout (*Salmo trutta*) are potential predators of smolts. Chain Pickerel was introduced to Nova Scotia in 1945 and are now present in over 95 water bodies across the province (Mitchell et al., 2010). The novel presence of Chain Pickerel is associated with decreases in fish diversity, fish abundance, and small bodied fishes (Mitchell et al., 2010). Smallmouth Bass was introduced in 1942, the degree to which salmonids make up their diet is unknown (LeBlanc, 2010). Smallmouth Bass and Chain Pickerel were first confirmed in the Stewiacke River in 2010 and 2014, respectively (J. MacMillan, pers. comm., April 2019).

Striped Bass (*Morone saxatilis*) is a common predator of smolts as their upstream spawning migration coincides in space and time with the smolts' out-migration (Gibson et al., 2015; Daniels et al., 2018; Andrews et al., 2019). The tidal portion of the Stewiacke River is annually used as spawning habitat by Striped Bass (Bradford et al., 2015). Striped bass numbers have increased in the Bay of Fundy and Gulf of Maine as Atlantic Salmon populations have declined (Gibson et al., 2015).

Avian predators include kingfishers (*Megaceryle alcyon*), cormorants (*Phalacrocorax auritus*), and mergansers (*Mergus merganser*) (COSEWIC, 2010). Small river mammals such as muskrat and mink, and marine mammals including seals are also potential predators (Cairns & Reddin, 2000).

1.6 Thesis Goals

The goal of this thesis is to assess the role of predation as a threat to juvenile Atlantic Salmon survival in the Stewiacke River, NS, and to explore intrinsic and extrinsic factors that may interact and contribute to smolt mortality via predation or other causes.

Chapter 2 explores technology and modelling based methods to identify predation of tagged salmon smolts. A standardized workflow is developed that other researchers can apply to their own study species in an effort to reduce predation bias in telemetry studies.

Chapter 3 explores mortality rates of out-migrating smolts in the Stewiacke River over three study years and investigates differences in smolt susceptibility to predation based on behavioural, physiological, and temporal factors.

Chapter 2: Application of machine learning and acoustic predation tags to classify migration fate of Atlantic Salmon smolts

2.1 Introduction

A major assumption of animal telemetry studies is that the data collected from tags represent the natural movements of a live individual of the study species, and not an expelled tag, a mortality, or the movements of a predator (Gibson at al., 2015; Klinard et al., 2019). However, the violation of this assumption is often not addressed, despite the negative impact it can have on study results, population management, and conservation efforts (Klinard & Matley, 2020). In the aquatic environment, predation of tagged fish presents a serious challenge to telemetry studies, because acoustic tags can continue to transmit through the body of the predator for as long as six months (Klinard et al., 2019; 2021). Therefore, failure to identify predation events of tagged individuals introduces a "predation bias", such that survival rates are inflated, individual movement patterns (e.g. depth use, rate of travel) are calculated based on both prey and predator movement, and the locations of areas of high mortality are skewed (Gibson et al., 2015; Daniels et al., 2019; Klinard et al., 2019). Even when predation events are identified, it is often on a subjective basis (Perry et al., 2010; Buchanan et al., 2013), dependent on predator and prey behaviour being significantly different and distinguishable (Romine et al., 2014; Gibson et al., 2015; Moxam et al., 2019), and difficult to pinpoint the time and location of mortality, hindering attempts to remove detections of consumed fish (Gibson et al., 2015; Daniels et al., 2018). Therefore, predation may reduce confidence in the conclusions of animal telemetry studies.

Movement ecologists recognize the negative impact of predation on the interpretation of acoustic telemetry data and have been developing methods for identification of predation in order to reduce its bias on study results. Early approaches to classify predation were to gather

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

The random forest and cluster analysis methods described above are classification approaches in the machine learning family, a branch of statistics that is used to predict outcomes from training data to in-sample or out-of-sample data (Thessen, 2016). In supervised machine learning (e.g., random forests), models are trained on data sets with independent and dependent variables, the model learns how the variables are related, and the model is then able to predict the dependent variable on future data sets where only the independent variables are provided (Thessen, 2016). Unsupervised methods (e.g., cluster analysis) find patterns among the independent variables to organize data based on underlying similarities in the data ascertained by the algorithm (Olden, 2008). Machine learning approaches are becoming increasingly used in ecology because they are able to model data that are non-linear, contain interacting variables, and have missing values, all of which are common in ecological data sets (Olden, 2008; Thessen, 2016). Applications of machine learning in ecology include habitat modelling and species distribution (Cutler et al., 2007; Brownscombe et al., 2019), species identification (Tabak et al., 2018), monitoring biodiversity (Cordier et al., 2017), and predicting the conservation status of species (Bland et al., 2014). The ability to make accurate ecological predictions is vital for informed management and decision making (Clark et al., 2001; Olden, 2008; Coreau et al., 2009).

Ideally, a combination of both behavioural and sensor-based methods for determining predation events would do much to increase confidence in fate classification of tagged fish, as tag sensors may sometimes fail and predator behaviours may not always be significantly different than prey behaviour (Weinz et al., 2020). Juvenile Atlantic Salmon (smolts) out-migrating from the Stewiacke River, Nova Scotia present an ideal opportunity to apply this combined approach. Natural mortality of smolts during seaward migration is high, with predation

often accounting for the majority of mortalities and challenging management efforts, especially those that rely on fish tracking data (LaCroix, 2008; Thorstad et al., 2011; Thorstad et al., 2012). The Stewiacke River is dominated by Striped Bass, a common predator of Atlantic Salmon smolts. Salmon smolt behaviour during migration consists largely of short, linear movements directed downstream with some reversals during out-migration, especially when first entering the estuary, likely as a response to osmotic stress (Halfyard et al., 2012; Halfyard et al., 2013). The impact of incoming tides or hydropower backwater may also push smolts back upstream (Beland et al., 2001; Babin, 2019). Except for these occasional path reversals, these movements are distinct from the extensive and tortuous movements with frequent reversals in up and downstream movement exhibited by Striped Bass (Romine et al., 2014; Gibson et al., 2015; Daniels et al., 2018). These differences form the basis for the behavioural metrics with which we can distinguish live and predated smolts, conducive to supervised machine learning approaches to identifying predation based on movements. However, these machine learning methods have not been adequately tested against objective empirical data with which models can be evaluated and best practices developed for a workflow to identify predation of tagged fish. The introduction of predation sensor tags provides a unique opportunity to compare machine learning methods designed to identify predated Atlantic Salmon smolts using models based either solely on behavioural metrics (unsupervised) or informed by data obtained from predation tags (supervised) to determine the best method for fate classification and the value of using predation tags. In this chapter, rates of estimated smolt migration survival and predation are compared among three approaches: modelling of behavioural metrics, tag pH sensors, and a combination of the two.

2.2 Methods

2.2.1 Study System

The Stewiacke River, Nova Scotia is one of 50 rivers within the inner Bay of Fundy (iBoF) Atlantic Salmon designatable unit (DFO, 2019b). The iBoF unit is currently listed as Endangered under Canada's Species at Risk Act (SARA). Low survival during the estuarine and marine stages of the Atlantic Salmon life cycle is preventing population recovery (DFO, 2019a). Reducing adult marine mortality is challenging, therefore, identifying sources of mortality and quantifying predation rates of migrating smolts is vital to informing population management. Smolts migrate down from the Stewiacke River and its tributaries out to the Minas Basin via the Shubenacadie River (Fig. 2.1). The Stewiacke River is the only river in the iBoF unit that is confirmed as an annual spawning site for Striped Bass (Bradford et al., 2015). Striped Bass congregate in the tidal waters of the Stewiacke River to spawn in May-June (Bradford et al., 2015), the same time and location as the smolt out-migration.

2.2.2 Sampling and Tagging Procedures

Sampling of Atlantic Salmon smolts occurred within the Stewiacke River watershed in three years, spanning 2017-2019, during the annual smolt run. Smolts were captured via rotary screw trap just downstream of the Stewiacke River head-of-tide in 2017 and just upstream of the head-of-tide in 2018 (<2 km apart; Fig. 2.1). In 2019, smolts were captured using a barrier fence on the Pembroke River, ~40 km upstream of the head-of-tide (Fig. 2.1). Both types of traps were checked for fish daily. Smolts were transferred from the traps to floating bins in a calm section of the river for holding prior to sampling and surgeries. Fifty smolts were tagged in both 2017 and 2018; 56 smolts were tagged in 2019 (total N=156).

Fish were measured prior to surgery (fork length [mm], mass [g]). Only smolts longer than 12 cm in fork length were chosen for tagging to ensure that the recommended tag-to-body size ratio was not exceeded (<8% for Atlantic Salmon; LaCroix et al., 2004). The average tag-to-body size ratio across all years was 3.0% (range 1.0-5.2%). Smolts were then anaesthetized in a buffered 10 mg/L solution of tricaine methanesulfonate (MS-222), until loss of equilibrium and spinal reflexes. A maintenance solution of buffered 5 mg/L tricaine methanesulfonate was circulated over the gills of the fish during surgeries. V5D-180 kHz predation acoustic transmitters (12.7 x 5.6 mm, 0.68 g in air; Innovasea Systems Inc., Bedford, Nova Scotia) were surgically inserted through a ~8 mm incision in the abdomen of smolts following standard procedure (Cooke et al., 2011). Incisions were closed with two single interrupted sutures. Smolts were returned to the floating river-side bins and held until dusk to recover from surgery before release just downstream from the point of capture. The average duration for the measuring and surgical procedures was 3.3 +/- 0.7 mins, and average recovery times were ~7 +/- 1 hrs.

Fish collection permits were issued by Fisheries and Oceans Canada (DFO 323354). All fish handling and surgical procedures conformed to standards established by the Canadian Committee on Animal Care, via permits issued by Fisheries and Oceans Canada (Maritimes Region Animal Care Committee Animal Utilization Protocols 17-16, 18-13, 19-10) and by Dalhousie University (University Committee on Lab Animals permit 18-126). Field work was done in conjunction with the Mi'kmaw Conservation Group who were operating under the Aboriginal Fund for Species at Risk.

2.2.3 Description of Tags and Receiver Array

The V5D tags (Innovasea Systems Inc.) have a biopolymer coating that triggers a change in transmitter ID (from an even number to the next odd number) when dissolved by the stomach

acids of a predator, thus indicating that a predation event has occurred. It is assumed that only predation events by fishes will be detected using this technology because avian or semi-aquatic predators would more likely remove the tag from the study site (Daniels et al., 2019). The lag time between tag consumption and the activation of the predation signal is ~5.8 hrs at 20°C (S. Smedbol, InnovaSea Systems Inc., pers. comm., January 2020) or 35.4 ± 17.7 hrs at a mean temperature of 11.8°C (Hanssen, 2020). In addition to temperature, lag time is dependent on the species and size of the prey (tagged) fish, and on the species and digestion rate of the predator (Halfyard et al., 2017).

Prior to tagging, an array of VR2W-180 kHz acoustic receivers (Innovasea Systems Inc.) was deployed along the migration route from the release/tagging site to the mouth of the Shubenacadie River (n=16 in 2017, n=15 in 2018, n=24 in 2019; Fig. 2.1). Supplemental detection data were provided by additional receivers (VR2W-180 kHz and HR2; Innovasea Systems Inc.) deployed in the Minas Basin (Fig. 2.1) and maintained by other researchers including the Ocean Tracking Network. Receivers were recovered in mid to late July of each year. Mobile tracking runs were periodically conducted on several river sections throughout the sampling period using a VR100 hydrophone (Innovasea Systems Inc.) to detect predation events that occurred outside of the range of stationary receivers.

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

2.2.4 Data Analyses

All analyses were conducted in R 3.6.2 (R Core Team; https://www.R-project.org). Detections occurring before or after the study period were removed as well as detections of tagged fish belonging to other studies. Detections were filtered using the *false_detections* function from the *glatos* package (Binder & Dini, 2012). This function identifies potentially false detections based on the programmed time interval at which the tags emit the ID signal and the recorded time between detections. Detections were then plotted for each individual smolt and visually assessed; detections identified as potentially false by the filtering function that also looked improbable given the location of receivers were removed from the data set. In the case that a dead smolt or evacuated tag dropped within range of receiver (i.e., resulting in a continuous string of detections for extended periods of time), the detection data were truncated to the first detection at that receiver (n=3).

2.2.5 Fate Classification

Detection data and the V5D pH sensor were used to classify smolts as belonging to one of three fate groups: successful migrant, mortality, or predation. Smolts were considered to have successfully completed migration if the last recorded detection was either at the mouth of the Shubenacadie River or in the Minas Basin. Smolts were presumed to be a mortality if their last recorded detection occurred upstream of the Shubenacadie River mouth. This pattern of detections could also result from tag ejection, failure to be detected when passing receivers, or predation by an animal that removed the tag from the study site. Predated smolts were identified if the pH sensor triggered a change in tag ID. However, preliminary analysis of detection data revealed that some smolts identified to be successful migrants or mortalities displayed movements more similar to predator behaviour than migratory smolt behaviour (several reversals

between up and downstream movement; Fig. 2.2). Consultation with the tag manufacturer confirmed the possibility of undetected predation events (Type II error). Additionally, a previous validation study has shown V5D predation tags to have only 50% accurate detection of predation (Hanssen, 2020). Therefore, machine learning methods were also applied to the detection data to classify smolt fate.

Behavioural metrics for the machine learning models were calculated from detection data of both live and predated tag IDs. Only data collected from stationary receivers were used in the calculations. The metrics were selected based on behaviours that are expected to be significantly different between salmon smolts and a predator such as Striped Bass. Some of these metrics are adapted from Gibson et al. (2015) and Daniels et al. (2018). The chosen metrics were total number of detections, maximum and minimum number of detections at a single receiver, number of days with detections, time between release and last detection, total distance travelled (river km), mean and maximum upstream speed (m/s) between two consecutive receivers, mean and maximum downstream speed (m/s) between two consecutive receivers, total number of reversals in up and downstream movement, total time on Striped Bass spawning grounds, total number of detections above the Stewiacke River and Shubenacadie River confluence, cumulative upstream distance travelled (river km), mean and maximum upstream distance travelled in a single step (river km), migration rate (river km/day), and for 2019, maximum speed in fresh water and tidal water (m/s). Metrics were tested in unsupervised k-means cluster analyses and supervised random forest models to compare fate classification based solely on the behavioural metrics, classification based on behaviour but also trained on individuals with known fate, and classification from detections and tag sensor only. Due to differences in receiver array set-up between years, models were run separately for each year. Attempts to pool years by truncating

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

2.2.6 Fate Classification: K-means Clustering

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

K-means clustering was performed using the *kmeans* function in base R. Behavioural metrics were centered and scaled to remove the effect of variables with larger values. Individual smolts were clustered into three groups (*K*=3) to represent the three fate classes, successful migrant, mortality, and predated. The *fviz_cluster* function was used to visualize cluster results, which plots observations using principal components (Kassambara & Mundt, 2019). Variable importance for clustering was measured by the rate at which individuals were misclassified if that variable was removed from the data set (misclassification rate). A higher misclassification rate means a variable is more important for assigning an individual to the best cluster. ANOVAs and Tukey tests were used to test if variables were significantly different among clusters. Each group was then assigned a fate based on metric means for each cluster and expected behaviour of

out-migrating smolts. Total distance travelled is expected to be longest in successful migrants that reach the Minas Basin and shortest among mortalities that die along the migration route. Total distance should also be long in predated smolts due the distance accumulated by the up and downstream movements made by predatory Striped Bass. It is expected that total time would follow a similar trend, with predated smolts showing less time than successful migrants due to the ejection of tags through the gastrointestinal tract of predators, and mortalities being detected for the least amount of time. Upstream speed should be fastest among predated smolts and very slow among successful migrants and mortalities. Similarly, upstream distance should be longest in predated smolts and shortest in successful migrants and mortalities because Striped Bass are expected to make frequent, extensive reversals in swimming direction while smolts are expected to conduct directed, downstream movements.

2.2.7 Fate Classification: Random Forest

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

The *randomForest* package and function (Liaw & Wiener, 2002) in R was used to create a model with fate as determined by the tag pH sensor and detection data as the response and the behavioural metrics as explanatory variables. Individuals with uncharacteristic smolt behaviour (i.e., upstream movement) were removed from the data set prior to training the algorithm (n=16

in 2017, n=14 in 2018, n=7 in 2019). Training data sets were therefore comprised of the remaining smolts in each year (n=34 in 2017, n=36 in 2018, n=9 in 2019). The migration fates of individuals in the training data sets are referred to as having known fates but because fates were determined through detection data there is still uncertainty associated with these fates due to the assumptions of no tag loss or failure and the potential for imperfect detection efficiency. Small sample size prevented cross validation with training and test data sets, therefore, out-of-bag (OOB) error produced from bootstrapping was used to calculate a confusion matrix and model accuracy. The number of trees made in the model was increased from the default 500 until OOB and class error rate fluctuations stabilized. The number of variables tried at each node was chosen based on minimizing OOB error. Due to class imbalance, the classes were assigned weights to penalize misclassification of underrepresented classes, class weights were chosen to minimize and balance class error rates (Table 2.1). The final model was then run on the individuals removed from the data set to predict their fates using the *predict* function. Variable importance was described by the average decline in model accuracy after permutations of that variable (mean decrease accuracy) and the average decrease in node purity if that variable was not used (mean decrease Gini). Larger values for both mean decrease accuracy and mean decrease Gini indicate greater variable importance.

2.3 Results

2.3.1 Predation Tags

The number of tagged fish determined to be predated based on the predation sensor was 24 (48%), 18 (36%), and 14 (25%) in 2017, 2018, and 2019, respectively. In 2019, two of the predation events occurred after entry into the Minas Basin and were therefore classified as successful migrants rather than predation.

2.3.2 K-means Clustering

For each year, smolts were placed into one of three clusters (Fig. 2.3). The most important variables differed somewhat among years; variables with consistently high misclassification rates included total distance travelled, total time detected, upstream swimming speed, and upstream distance travelled (Figs. A1-3). These variables were significantly different (ANOVAs, Tukey tests) between at least two of three clusters for each year. Therefore, clusters were assigned fate classes based on the differences in these variables and the expected behaviour of live salmon smolts, dead smolts, and predators.

For 2017, cluster 2 (n=9) had faster upstream swimming speeds, longer upstream distances travelled, and farther total distance travelled than clusters 1 and 3 (Fig. A4). These trends are more characteristic of Striped Bass movement than smolt movement, therefore, cluster 2 was determined to represent the predated fate class (Fig. 2.4). Clusters 1 (n=36) and 3 (n=5) were not significantly different from each other (Tukey tests; upstream speed: t=-0.08, p=0.997; total distance: t=-1.16, p=0.476; upstream distance: t=-0.33 p=0.941). Based on the short total distance travelled (Fig. A4), both of these clusters were assigned the fate of mortality. A successful migrant cluster was not identified.

The cluster plot for 2018 revealed some overlap between clusters 2 and 3 when plotted on the first two principal components (Fig. 2.3), however, they were significantly different from each other when examining variables with the highest misclassification rates (Tukey tests; total distance t=-10.5, p<0.001; total time t=-5.40, p<0.001). Cluster 2 (n=24) had the longest total time and farthest total distance (Fig. A5); therefore, it was assigned the successful migrant class (Fig. 2.4). In contrast to cluster 2, cluster 3 (n=23) showed the briefest total time and shortest distance (Fig. A5), which are metrics indicative of mortality. Cluster 1 (n=3) had intermediate

values between clusters 2 and 3, and total distance was greater than total time leading to the assignment of the predated fate to this cluster.

In 2019, cluster 3 (n=10) had a significantly greater number of reversals (Tukey test; cluster 1 t=8.77, p<0.001; cluster 2 t=9.75, p<0.001), longer time on Striped Bass spawning grounds (Tukey test; cluster 1 t=5.29, p<0.001; cluster 2 t=5.30, p<0.001), and longer upstream distance travelled (Tukey test; cluster 1 t=6.48, p<0.001; cluster 2 t=7.18, p<0.001), all of which are behaviours indicative of predation by Striped Bass (Fig. 2.4). Cluster 2 (n=34) had the longest total distance (Fig. A6) and was therefore, assigned the successful migrant fate. Conversely, cluster 1 (n=12) had the shortest distance travelled and was identified as the mortality class.

Model accuracy, calculated by the number of known fates (the fates of smolts in the data sets used to train the random forest models) within a cluster that matched that cluster's assigned fate (Table 2.2), was 38.2%, 52.8%, and 82.4% for 2017, 2018, and 2019, respectively (Fig. 2.5).

2.3.3 Random Forest

In-sample prediction accuracy of random forest algorithms ranged between 81.6 and 94.4% among years (Fig. 2.5). OOB error rates ranged from 5.6-18.4% (Table 2.1). Classification of the successful migrant class had 100% error rate in the 2017 model, but 0% in 2018 and 2019. Mortality class error rates ranged from 25-43% (Table 2.1). Error rates for the predated class were similar in 2017 and 2018 at 4 and 5%, respectively, but was 50% in 2019.

The most important variables in common among all years were time on Striped Bass spawning grounds, total distance travelled, and time detected (Figs. A10-12). Upstream speed, upstream distance, and number of reversals were also important variables in 2017 (Fig. A10).

Partial plots revealed that the probability of being classified as a successful migrant increased with increasing cumulative distance travelled, total time detected, and number of days detected (Figs. A13-18). The probability of being classified as predated increased with number of reversals, upstream distance travelled, upstream speed, and time spent on Striped Bass spawning grounds (Fig. A19-24). The trends in probability of being classified as a mortality were similar to those for the predated class except for time on Striped Bass spawning grounds, time detected, and distance travelled in which cases the trends were opposing.

The 2017 random forest algorithm reclassified all suspect individuals (five successful migrants, 11 mortalities) as predated (Fig. 2.4). In 2018, two of the suspect mortalities were reclassified as successful migrants but these individuals were retained as mortalities in the final fate counts. All other suspect mortalities were reclassified as predated, but only two among eight suspect successful migrants were reclassified as predated (Fig. 2.4). The 2019 algorithm reclassified the two suspect mortalities as predated (Fig. 2.4) and three of five suspect successful migrants as predated.

2.3.4 Comparison of Classification Methods

When comparing the assigned cluster fates to the known fates of individuals within clusters, as determined through detection data, the accuracy was highest in 2019 because the majority class in each cluster was the same as the assigned cluster fate (Table 2.2). The 2018 cluster assignments were also consistent with individual fates; however, two of the clusters were mostly comprised of predated smolts (Table 2.2). Cluster 1, which was assigned the predated fate, contained only two predated individuals while cluster 3 contained the majority of predated individuals (15) but was assigned a mortality fate based on the behavioural metrics. The 2017 clusters were difficult to distinguish based on behaviour and individual fates due to the high

number of predation events that year, predated individuals were spread among all three clusters (Table 2.2). Compared to predation tag data, the cluster analysis reduced predation estimates by 30% in 2017, 30% in 2018, and 3.5% in 2019 (Table 2.3).

Random forest algorithms consistently increased the percent of individuals classified as predated and resulted in a reduction of estimated migration success and mortality classes compared to the numbers obtained from the predation tag sensor and detection data (Table 2.3). Predation rates increased by 32%, 12%, and 9% in 2017, 2018, and 2019, respectively.

Unsupervised clustering methods are capable of fate classification but are less accurate than supervised methods (Fig. 2.5). On average, both machine learning approaches had higher accuracies than pH tag sensors alone (50%; Hanssen, 2020).

2.4 Discussion

This methods comparison built on previous studies to develop a standardized workflow for identifying predated individuals in acoustic telemetry studies (Fig. 2.6). We used tag sensor technology, unsupervised machine learning, and supervised machine learning to address the issue of "predation bias" in the field of telemetry and showed that using data collected from tag sensors to train supervised models provides the greatest accuracy for fate classification of tagged fishes (Fig. 2.5).

Unsupervised k-means clustering had lower classification accuracy due to almost all clusters in all years containing a mixture of individual fates. The mortality cluster (cluster 1) in 2019 contained an even split of mortalities and predated smolts, but the predated smolts in this cluster were identified in fresh water by mobile tracking and therefore the behavioural metrics resembled mortalities more closely than predation events detected in tidal water by stationary

receivers. The nature of mobile tracking downriver allows for only a few detections of a given tag in a single location which is insufficient to pick up distinct behaviour. Additionally, the most likely freshwater predators, Brown Trout (*Salmo trutta*) or Chain Pickerel (*Esox niger*), are relatively stationary species so detection data resemble a dropped tag or dead smolt rather than the active Striped Bass behaviour we were testing for. Similar to the cluster analysis, the 2019 random forest algorithm did not successfully differentiate the six freshwater predation events from mortalities.

Data from 2017 showed the greatest disparity of fate assignments amongst the three different classification methods (Table 2.3). In addition to overall model classification accuracy, balancing accuracy amongst classes is important especially for unbalanced data sets because models will ignore minority classes to achieve greater overall accuracy (Chen et al., 2004; Brownscombe et al., 2020). The small number of successful migrants compared to the number of mortalities and predated smolts in 2017 made it difficult for these individuals to be recognized by either type of machine learning approach. The few successful migrant smolts were masked in the cluster analysis by the behavioural characteristics of the other fate classes (Table 2.2), and despite the addition of class weights, the random forest model was still unable to accurately classify successful migrants. In contrast, the estimated percentage of successful migrants was relatively similar amongst all three methods in 2018 and 2019, while mortality and predation classes had larger disparities, especially for the 2018 cluster analysis (Table 2.3).

The amount of time a tag is retained within a predator and continues to function can impact a model's ability to accurately classify it as a predation. The retention time of tags in the gastrointestinal tract of predatory fishes depends on several factors including water temperature, predator size, prey size, and tag size (Romine et al., 2014; Halfyard et al., 2017; Daniels et al.,

2019; Klinard et al., 2019). The longest known retention time of predation tags is over 6 months observed in an acoustic telemetry study of Bloater (Klinard et al., 2019; 2021). Additionally, acoustically tagged Rainbow Trout (*Oncorhynchus mykiss*) and Yellow Perch were retained in predatory Largemouth Bass (*Micropterus salmoides*) for 1.1 –11.5 days (Halfyard et al., 2017). In species more comparable to this study, gut retention time of tagged juvenile Chinook Salmon (*Oncorhynchus tshawytscha*) consumed by Striped Bass ranged from 1.2-2.7 days, with a negative relationship to water temperature (Schultz et al., 2015). Here, tags triggered as predated were detected for an average of 2.9 days (range 0-32.7 days). After this period, tags were either evacuated through the gastrointestinal tract, the predator left the study area, or the tag ceased signal transmissions. The longer a tag is in a predator, the easier it is to identify it as a predation because there will be more detections tracking predator behaviour (Daniels et al., 2018). Predation events where the tag is ejected quickly and distinct predator movements are not captured are then more likely to appear as mortalities. This is prevalent in the 2018 cluster analysis where 13 of the 15 predated smolts in the mortality cluster (cluster 3) had retention times shorter than the average 2.9 days, the same is true for 15 of the 16 predated smolts in mortality cluster 1 in 2017.

The supervised random forest approach was the most accurate of the three fate classification methods (Fig. 2.5). This method increased estimated predation rates greatly beyond estimates made by the tag pH sensor alone and by the unsupervised cluster analysis, however, total estimated mortality only showed a large increase in 2017 (Table 2.3). The cluster analysis also only increased estimates of total mortality from tag sensor estimates in 2017. Predation accounted for a majority of all smolt mortalities (71-83%) under the random forest estimates while predation tag estimates showed predation as accounting for just above half of all

mortalities (56-67%). Whether mortality was attributable to predation or unknown causes, estimated total mortality did not differ greatly between methods. Both predation rates and total mortality decreased from 2017 to 2019 for both the random forest and tag sensor methods (Table 2.3). The variation in estimated migration mortality rates among years could be due to a number of factors including changes in predator and prey abundance, changes in the timing of the Striped Bass spawning period, or differences in sampling methods.

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

The results presented here show that there is value in using predation tags combined with modelling methods to identify predated individuals (Fig. 2.6). Data including individuals with known fates that have been determined by detection data and a pH or other tag sensor increases confidence in model results and improves model accuracy. The unsupervised cluster analysis had model accuracies ranging from 38.2-83.4%, while the supervised random forest was 81.6-94.4%

accurate at in-sample fate classification (Fig. 2.5). The k-means clustering method was able to cluster individuals based solely on behavioural metrics, but it can be difficult to discern which cluster represents which fate group and the decision is likely to be subjective. Assigning fates to clusters was dependent on distinct and predictable predator and prey behaviour with smolts moving downstream and Striped Bass exhibiting multiple reversals. However, it is possible that smolts could exhibit upstream movement if they were being carried by the tides (Beland et al., 2001) or as a response to osmotic stress (Halfyard et al., 2012). The random forest algorithms were trained on smolts of known fate and classified suspect smolts on an individual basis compared to the cluster analysis where smolts were classified by group, leading to a mixture of fates in each cluster. While random forest models were shown to be the best approach for identifying predation events and predicting the migration fate of smolts, these models were still not 100% accurate and the potential for misclassification remains.

Differences in model results and prediction accuracies among years highlight the importance of having a large sample size not only for greater power in model predictions but also in an attempt of balancing classes for individuals of known fates. Random forests are among the least sensitive classification algorithms to reductions in sample size (Maxwell et al., 2018; Moghaddam et al., 2020), however, issues of class imbalance and potentially unrepresentative data remain when using small training data sets (Chen et al., 2004; Brownscombe et al., 2020). A recommendation for machine learning in general is to have a training sample size ten times the number of predictor variables, but the minimum recommended sample size for classification algorithms specifically is dependent on the type of data and algorithm (Indira et al., 2010; Maxwell et al., 2018).

Other considerations to optimize model performance are receiver configuration and coverage, which are vital to capturing the distinct behaviour needed to differentiate predator and prey species. The distances between receivers in a river system limits the accuracy of distance travelled and speed calculations because the movement of the individual between receiver detection ranges is unknown. It is therefore not ideal to have large gaps between receivers but the number of receivers available is often limited, especially for large study areas. The behavioural metrics required for machine learning approaches are context-specific and must be tailored to the prey and predator species of interest. Deciding on behavioural metrics prior to receiver deployment can aid in array design to ensure receiver coverage is adequate for calculating the necessary metrics. However, it is possible to have multiple or unknown predatory species in a study system, calculating metrics or concentrating receiver coverage for only one species could mask predation by another. Additionally, avian predation typically resembles mortalities in terms of detection data and could therefore not be identified here. Other researchers have identified avian predation by searching colonies or nesting sites for evacuated tags (Evans et al., 2012). For tracking salmon smolts specifically, good up and downstream receiver coverage of the river is important for distinguishing predator and prey movement based on smolt migration behaviour. Predation tags are recommended when tracking smolts due to the high predation pressure from various species during out-migration.

A limitation of the modelling approaches used here is that a timestamp for the moment of predation is not provided. A benefit to using predation tags is that detection histories can be truncated to represent only movements of the live prey based on the change in tag ID and estimated signal lag time (Fig. 2.2). A fine scale or gridded receiver array where the position of the tagged fish can be triangulated allows for more accurate calculations of speed and turning

angle, which can be used for behavioural change point analysis. Behavioural change point analysis identifies significant changes in movement parameters across a time series (Gurarie et al., 2009) so not only can it be used for identifying predated individuals based on behavioural anomalies, but it can also provide a time estimate for when the predation occurred. However, triangulation is difficult to achieve in rivers given their size and shape.

K-means clustering underestimated the number of predation events and due to type II error, the tag sensor did as well. Random forest modelling and the example workflow we provide, allows one to study predation by using predation tags, therefore removing the need to tag predators, while also accounting for sensor malfunctions. Combining acoustic tag sensors with supervised machine learning approaches to identify mortalities and predation events of tagged fishes is recommended to increase confidence in telemetry study results.

Tables

Table 2.1 Random forest model metrics. Number of decision trees made (ntree), number of variables considered at each node (mtry), class weights assigned to mortalities, predated smolts, and successful migrants, respectively (claswt MPS), out-of-bag error rate (OOB error), and class error rate for mortalities, predated smolts, and successful migrants, respectively (class error MPS).

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

Table 2.2 Number of individuals of each fate (predated P, other mortality M, successful migrant S, successful migrant or mortality suspected of being predated U) as determined by predation tag and detection data in each cluster. Cluster fates, in brackets, assigned based on average behavioural metrics of each cluster.

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

Table 2.3 Percent of smolts belonging to each fate (predated P, other mortality M, successful migrant S) as determined by the V5D predation tag sensor and detection data, unsupervised cluster analysis (CA), and supervised random forest (RF).

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

Figures
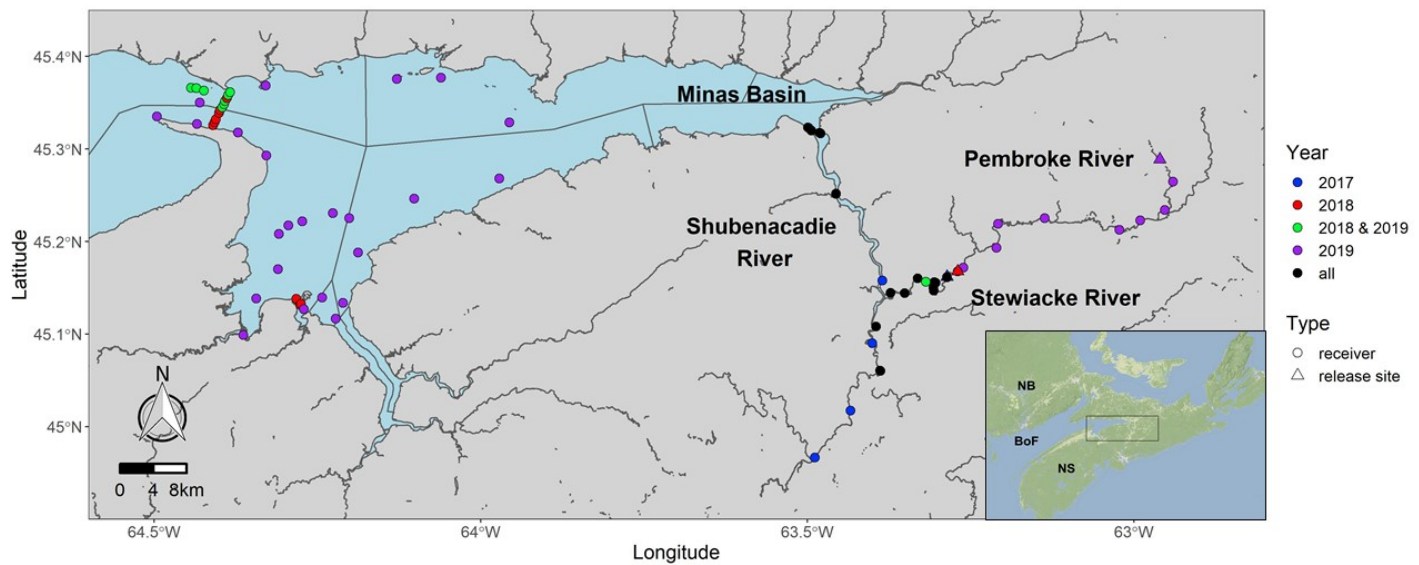
Fig. 2.1 Map showing receiver and release site locations for each study year in the Stewiacke River watershed and Minas Basin, Nova Scotia, Canada. Inset shows location of study area (box) in relation to Nova Scotia (NS), New Brunswick (NB), and the Bay of Fundy (BoF).
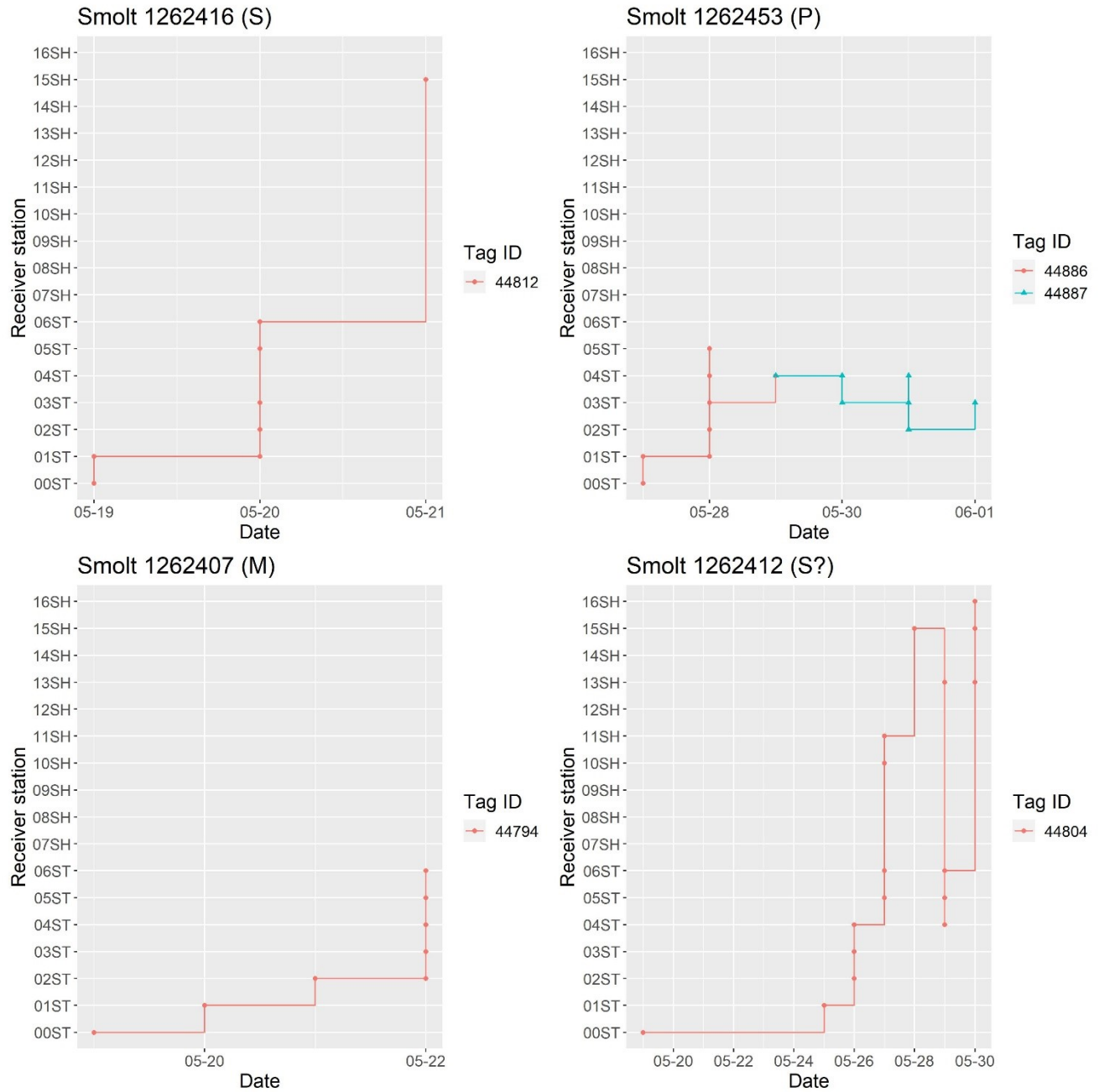
Fig. 2.2 Plots showing detections of four individual Atlantic Salmon (*Salmo salar*) smolts in 2017 representing typical migration paths of a successful migrant (S), a predated smolt (P), a mortality of unknown cause (M), and a smolt indicated to have successfully completed migration but suspected of being predated (S?). Receiver stations listed in order from release on the Stewiacke River (00ST) to the mouth of the Shubenacadie River (14-16SH). Receivers 07-12SH are upstream of the ST-SH confluence.
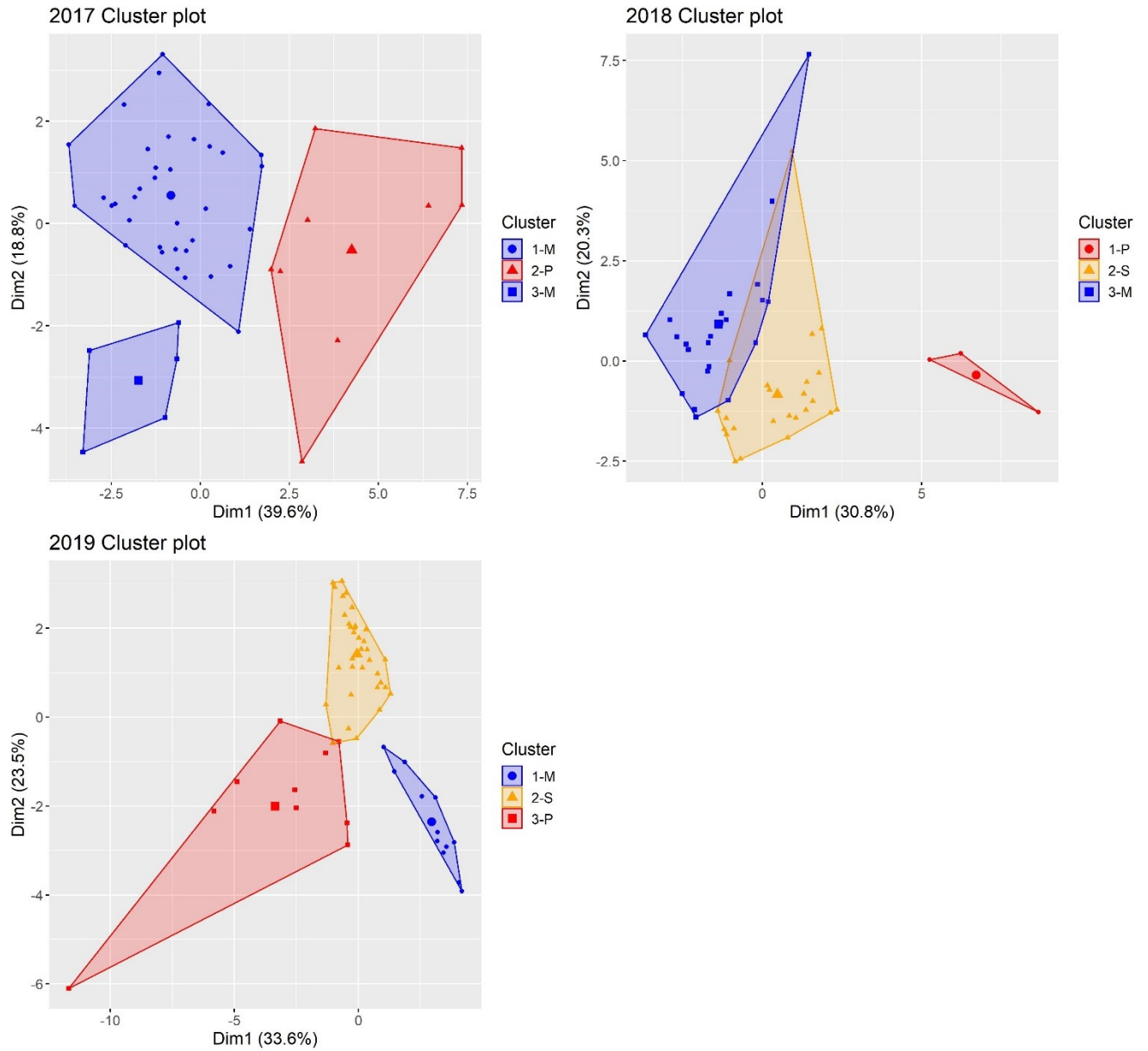
Fig. 2.3 Cluster plots for all study years showing k=3 clusters representing the three fate groups (mortality M, successful migrant S, predation P). Cluster colour corresponds to fate, point shape corresponds to cluster number. Small points are individual smolts, large points are cluster centroids. Clusters plotted on the first two principal components (Dim1, Dim2 [% variance explained]).
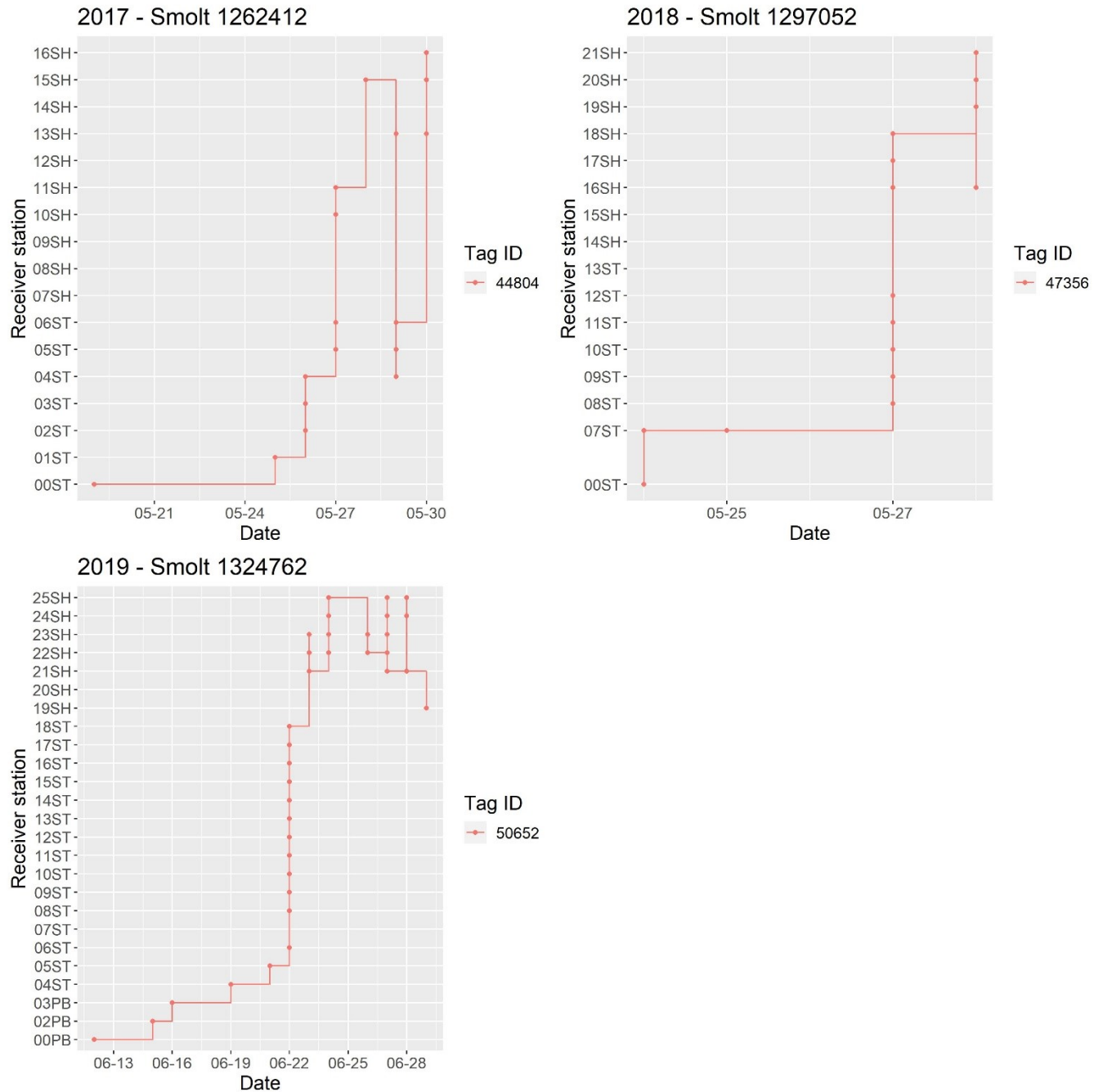
Fig. 2.4 Plots showing detections of three individual smolts in 2017, 2018, and 2019. Receiver stations listed in order from release on the Stewiacke River (00ST) or Pembroke River (00PB) to the mouth of the Shubenacadie River (14-16SH; 17-21SH; 22-25SH). Smolt 1262412 was classified as a successful migrant based on tag detections but classified as predated by both k-means clustering and random forest. Smolt 1297052 was classified as a successful migrant based on tag detections and k-means clustering but classified as predated by random forest. Smolt 1324762 was classified as a mortality based on tag detections but classified as predated by both k-means clustering and random forest.
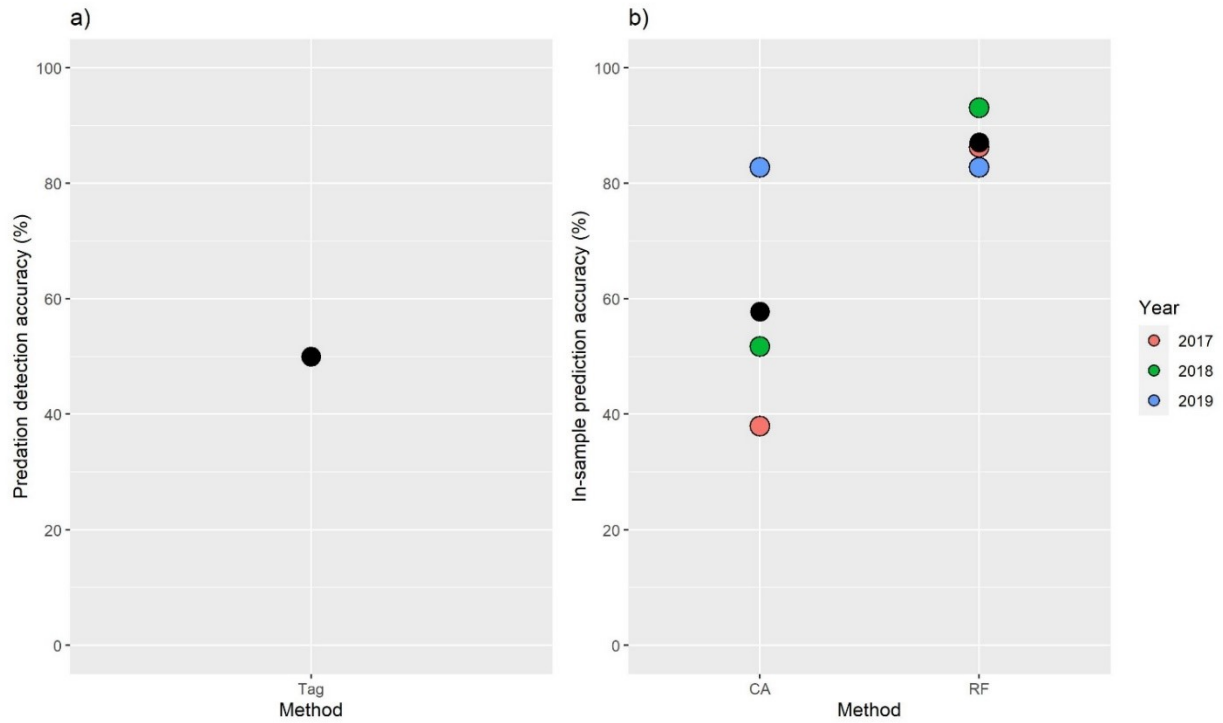
Fig. 2.5 a) Rate at which V5 predation tags accurately identified predation events in Hanssen (2020) tank study. b) Model accuracies for k-means clustering (CA) and random forest (RF) models by year, mean accuracy shown by black circle.

Step 1: Remove false detections and continuous detections of mortalities

Step 2: Determine known fates based on detection history and predation tag sensor data

Step 3: Identify behavioural metrics hypothesized to differentiate predator and prey species of interest (speed, space use, etc.)

Step 4: Split data into training and test data sets

Step 5: Run random forest on training data, tune model parameters to minimize classification error

Step 6: Use final model to predict fate classes of test data

Step 7: Check model accuracy

Step 8: Revise fate estimates from Step 2. Truncate detection history of predated individuals prior to further analyses

Fig. 2.6 Diagram of example workflow to identify predated individuals in telemetry data using predation tags and a machine learning framework.

Chapter 3: Behavioural and physiological factors related to predation susceptibility of Atlantic Salmon smolts

3.1 Introduction

Migration is a vital life history strategy for species across animal taxa (Wilcove & Wikelski, 2008; Milner-Gulland et al., 2011) but is often a period of concentrated mortality for migrants (Sergio et al., 2019). Juvenile anadromous salmon, referred to as smolts, emigrate from less productive natal freshwater environments to highly productive marine environments to achieve greater growth (Hendry et al., 2004). Salmon populations (*Salmo salar* and *Oncorhynchus spp.*) have declined largely due to poor survival during seaward and return migrations (Wilcove & Wilkelski, 2008). Predation of smolts is considered to be a key contributor to out-migration mortality (LaCroix, 2008; Thorstad et al., 2012).

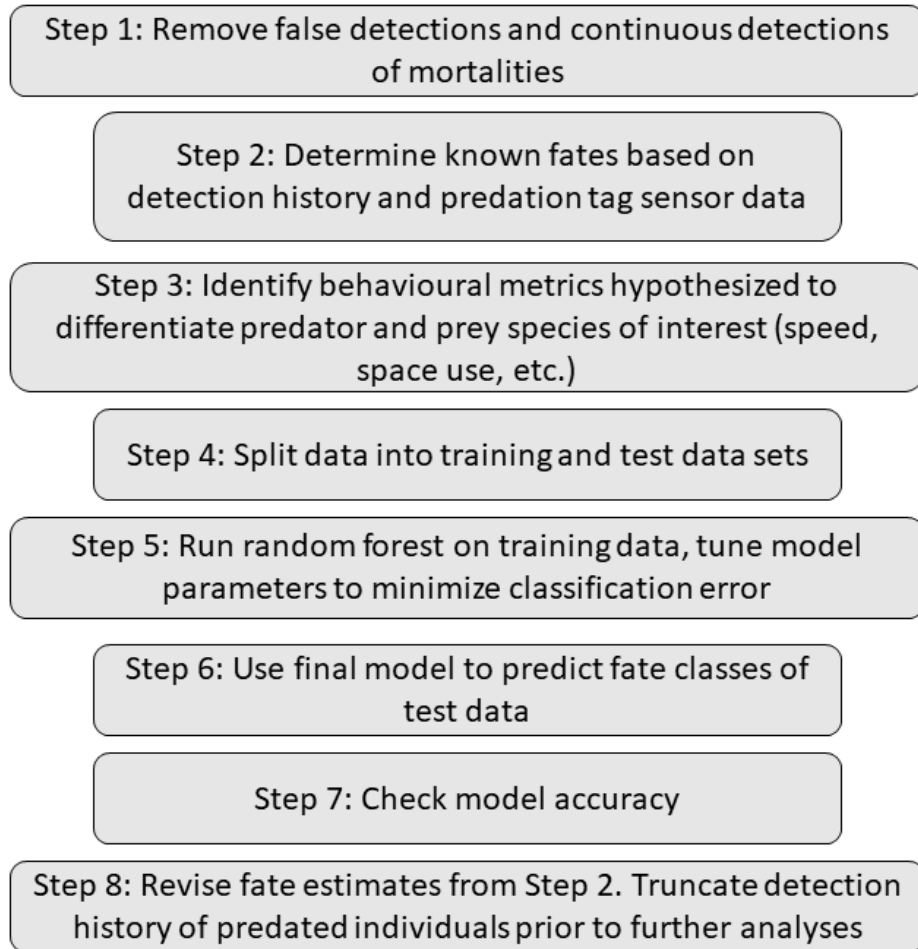Predation events occurring during migrations can have a significant, negative impact on salmon population numbers, especially for endangered populations (Grout, 2006; LaCroix, 2008). The rate of predation on salmon smolts can vary among years and rivers, as well as among different areas within a single river (Halfyard et al., 2012). Piscine, avian, and mammalian predators are known to aggregate in certain areas, such as spatial bottlenecks, to take advantage of predictable, mass salmon migrations (Furey et al., 2015; Daniels et al., 2019; Flavio et al., 2019). Bull Trout (*Salvelinus confluentus*) have been shown to aggregate and binge feed on pulses of Sockeye Salmon (*Oncorhynchus nerka*) smolts exiting the Chilko River, British Columbia (Furey, 2016). Striped Bass (*Morone saxatilis*) aggregate in dam tailraces on the Merrimack River, Massachusetts to feed on Atlantic Salmon (*Salmo salar*) smolts (Blackwell & Juanes, 1998). Smaller populations of migrating salmon are not likely to attract such predator aggregations but are subject to more opportunistic predation (Furey et al., 2020).

45

Salmon smolt migration behaviour is largely influenced by predator avoidance strategies. Smolts primarily migrate at night to avoid visual predators and synchronize their migration to travel in groups and swamp predators (Ibbotson et al., 2011; Thorstad et al., 2011; Clark et al., 2016; Furey et al., 2016; Flavio et al., 2019). Predator swamping is unlikely to be effective with small salmon populations; however, concurrent migrations of other diadromous species such as Alewife (*Alosa psuedoharengus*) may act as a prey buffer (Svenning et al., 2005; Saunders et al., 2006; Furey et al., 2020).

While predation is often the proximate cause of smolt mortality it may not be the ultimate cause because biotic and abiotic stressors can increase susceptibility to predation (Bender, 2018) by reducing swimming abilities and anti-predator behaviours (Handeland et al. 1996; Dieperink et al., 2002; Halfyard et al., 2012). Internal and external stressors that commonly confer a predisposition to mortality in smolts include metabolic exhaustion, physical barriers to migration, poor body size/condition, osmotic stress, water pollution, high water temperatures, and novel pathogen exposure (Wilcove & Wilkelski, 2008; Hostetter, 2009; Ibbotson et al., 2011; Jeffries et al., 2014; Miller et al., 2014; Furey, 2016; Tucker et al., 2016; Sergio et al., 2019). Identifying the causes, locations, timing, and magnitude of migration mortality is vital to improving population conservation and management (Gibson et al., 2015; Daniels et al., 2019).

Telemetry is a useful tool for identifying migration mortalities and can reveal spatial and temporal hot spots of mortality (Clark et al., 2016; Furey et al., 2016; Flavio et al., 2019). Tag sensor technologies such as pH sensors can also distinguish between general mortality and predation mortality, further refining the information needed for proper population management (Halfyard et al., 2017, Daniels et al., 2019). Combining telemetry with transcriptomics, a method of studying an individual's physiological condition through gene expression levels, provides

insight into potential direct and indirect (sublethal) causes of mortality (Miller et al., 2011; Jeffries et al., 2014; Bass et al., 2019). The analysis of transcriptomes has linked gill Na^+/K^+-ATPase activity (osmoregulatory condition) to mortality in Atlantic Salmon smolt out-migrations (Stich et al., 2015) and Pacific salmon (*Oncorhynchus spp.*) adult return-migrations (Miller et al., 2011; Hinch et al., 2012). Transcriptomic and histological methods have also linked immune responses and pathogen presence to increased salmon smolt mortality through piscine and avian predation (Hostetter, 2009; Jefferies et al., 2014; Miller et al., 2014; Furey, 2016).

The onset of sublethal, delayed, or carryover effects from stressors experienced in the natal freshwater habitat can cause mortality during out-migration (McCormick et al., 1998; Thorstad et al., 2012; Midwood et al., 2014; 2015; Birnie-Gauvin et al., 2021). For example, the impacts of water contaminants may not be seen until entry into saltwater environments where increased stress is placed on the osmoregulatory system (McCormick et al., 1998; Thorstad et al., 2012). Additionally, the level of pathogenicity of some infectious agents varies between fresh and salt water (Miller et al., 2014). Stressors that occur in fresh water but contribute to early marine mortality of salmon are easier to manage than stressors occurring at sea. It is therefore crucial to identify and mitigate freshwater stressors with delayed effects to reduce mortality during smolt and post-smolt stages in an effort to maintain populations (Thorstad et al., 2012).

In this study, I explore the migration mortality of Atlantic Salmon smolts from the Stewiacke River, Nova Scotia which is part of the Endangered inner Bay of Fundy (iBoF) designatable unit of Atlantic Salmon. The iBoF unit declined more than 99% between the years of 1980 and 2000 due to overfishing, changes in environmental conditions, water contaminants, interactions with aquaculture, barriers to migration, and shifts in prey and predator communities (COSEWIC, 2006; SARA, 2010). Predation risk is particularly high for smolts in the Stewiacke

River primarily due to the large population of spawning Striped Bass, as well as invasive species such as Chain Pickerel (*Esox niger*), Smallmouth Bass (*Micropterus dolomieu*), and Brown Trout (*Salmo trutta*).

To better understand the mechanisms of smolt mortality in the Stewiacke River, three years of acoustic predation tag data and one year of transcriptomic data were used to meet the following objectives. 1) To quantify migration survival and predation rates across three study years and identify areas of high mortality/predation along the migration route. 2) To determine if certain migratory or predator avoidance behaviours are associated with migration fate. 3) To determine if physiological status or pathogen presence is associated with migration fate.

3.2 Methods

3.2.1 Study System

The Stewiacke River shares its confluence with the Shubenacadie River which empties into the Minas Basin (Fig. 2.1). The Stewiacke and Shubenacadie Rivers are heavily influenced by the Bay of Fundy tides. The large tides (>10 m) cause great changes in salinity, temperature, water level, and turbidity over the daily tidal cycle. The head-of-tide is 30 km upstream of the river mouth in the Shubenacadie River and 15 km upstream of the confluence in the Stewiacke River. Salinity at the mouth of the Shubenacadie River ranges from 15-27 ppt over the tidal cycle and can be reduced to as low as 2 ppt by heavy rainfall (Martec Ltd., 2007).

The Stewiacke River is the only river confirmed to be annually used for spawning by the Bay of Fundy population of Striped Bass (Bradford et al., 2015). The Atlantic Salmon smolt migration and Striped Bass spawning migration both occur in May-June in the Stewiacke River (Bradford et al., 2015; DFO, 2019b). Striped Bass spawning occurs primarily in the tidal waters

of the Stewiacke. Striped Bass feed heavily on Atlantic Salmon smolts and other fishes including Rainbow Smelt (*Osmerus mordax*), Alewife (*Alosa pseudoharengus*), Blueback Herring (*Alosa aestivalis*), and American Shad (*Alosa sapidissima*) (Bradford et al., 2015). Striped Bass are highly abundant within the Bay of Fundy and Gulf of Maine while salmon populations have greatly declined (Grout, 2006; Gibson et al., 2015), therefore, predator swamping strategies are not likely to be effective. However, Alewife and Blueback Herring may act as a prey buffer to out-migrating smolts because they are abundant in the area and their spawning migrations also peak in late May/ early June (Gibson et al., 2017). A fishing weir in the Minas Basin captured six Atlantic Salmon (post-smolts and adults), 1388 Striped Bass, and 350 343 Alewife and Blueback Herring from April to July in 2017 (Dadswell et al., 2020). Additionally, Chain Pickerel, Smallmouth Bass, and Brown Trout are potential predators of salmon smolts in the freshwater stretches of the Stewiacke River. There are also several avian predators of smolts in the area (SARA, 2010).

3.2.2 Field Methods

Sampling procedures, acoustic tags, and receiver arrays were as described in Chapter 2 (sections 2.2.2 and 2.2.3) with the addition of gill biopsies in 2019 for transcriptomic work. Immediately after surgical insertion of tags, small non-lethal gill tissue samples (approx. 1 mm, 2-3 gill lamellae) were taken from tagged smolts (n=56). Tissue samples were preserved in RNAlater (Life Technologies, Grand Island, New York) and stored at -80°C until RNA extractions could be performed.

3.2.3 Laboratory Analyses

Preserved gill tissue samples were transported to the Genomics laboratory at DFO's Pacific Biological Station in Nanaimo, BC to be processed. After thawing, RNAlater was poured off and tissue samples were transferred to collection microtubes. Tissue samples were homogenized physically using stainless steel beads and chemically with 200 µL of TRIzol for organic phase separation. Samples were shaken at 30 Hz for 3 min then spun down. 25 µL of 1-bromo-3-chloropropane (BCP) was added to the homogenate followed by 1 min hand shaking, 5 min incubation at room temperature, and 6 min centrifugation to produce an RNA aqueous layer. 100 µL aliquots of the aqueous layer were removed and placed into 96 well plates for extraction. A Biomek FXP liquid handling instrument (Beckman-Coulter, Mississauga, Ontario) was used to clean the aqueous RNA sample by removing DNA with Turbo DNase. Samples were eluted by binding RNA to MagMax beads. RNA concentration was normalized to 50 ng/µL. Reverse transcription protocol recommended by BioMark (Fluidigm Corp., San Francisco, California) was followed to synthesize cDNA using 16 µL of normalized RNA, 4 µL of VILO cDNA MasterMix (Life Technologies), and polymerase chain reaction (PCR) cycling at 25ºC for 10 mins, 42ºC for 60 min and 85ºC for 5 min.

Specific target amplification (STA) of cDNA to pre-amplify genes of interest using a 200 nM STA primer mix and PCR TaqMan Master Mix (Thermo Fisher Scientific, Waltham, Massachusetts) was done following BioMark protocol. The primer mix contained 4 µL of each 100 µM primer (forward and reverse) and 688 µL of DNA suspension buffer. Amplification of 1.3 µL of cDNA combined with 3.8 µL of the primer mix was done at 95ºC for 10 min, 14 cycles of 95ºC for 15 secs, 60ºC for 4 min followed by a 4ºC hold at the end of the cycle sequence. Leftover primers were removed with 2 µL of ExoSAP-IT High-Throughput PCR Product Clean

Up (MJS BioLynx Inc., Brockville, Ontario). 5 µL of sample mix (pre-amplified cDNA, 2X TaqMan Gene Expression MasterMix, and 20X GE Sample Loading Reagent) and 5 µL of assay mix (9µM primer pairs, 2µM probes, and 2X Assay Loading Reagent) were loaded onto Fluidigm 96.96 Dynamic Array chips to be run on a Fluidigm BioMark (Fluidigm Corp.) platform for real-time quantitative PCR (qPCR). Probes for each assay were previously prepared by dilution from 100µM to10µM in DNA suspension buffer. The BioMark platform performed PCRs at conditions of 50°C for 2 min, 95°C for 10 min, then 40 cycles of 95°C for 15 sec and 60°C for 1 min.

In total, across the 56 samples analyzed, 57 host genes were targeted under the following categories: smoltification, osmotic stress, general stress, stress mortality, imminent mortality, mortality related signature, inflammation, immune stimulation, and viral disease development (Table 3.1). These genes, or biomarkers, comprise the Salmon Fit-Chip, a microarray used as a measure of physiological fitness in salmon species (Houde et al., 2019). The presence of a stressor or physiological status is determined through the co-expression of biomarkers within a gene category. Additionally, the presence of RNA from 18 pathogenic viruses, bacteria, and fungal/protozoan parasite species was also tested for in duplicate assays.

3.2.4 Statistical Analyses

All analyses were conducted in R 3.6.2 (R Core Team; https://www.R-project.org).

3.2.5 Statistical Analyses: Telemetry Data

Filtering of telemetry data was as described in Chapter 2 (section 2.2.4).

3.2.6 Statistical Analyses: Transcriptomic Data

Cycle thresholds (Ct) were set in the middle of the log phase on the qPCR amplification curves using BioMark software (Fluidigm Corp., www.fluidigm.com). Assay efficiencies were calculated using the slope of the regression between Ct values and serial dilutions of host and microbial artificial positive construct (APC) DNA. Only genes with assay efficiencies between 0.80 and 1.20 were retained in the final data set, therefore eight of the 57 genes were removed (Table 3.1). Amplification curves were visually assessed for abnormalities or low expression of reference host genes. Host gene expression was normalized using the efficiency corrected delta-delta Ct method (Pfaffl, 2001) using the FluidFish v 2.8 package in R (Bass, 2020). Target gene expression levels were relative to the mean expression of three housekeeping (reference) genes and APC serial dilutions (control) while accounting for assay efficiency. For the pathogens, both duplicate assays must show positive detections for that agent to be considered present in the sample. Pathogen load (RNA copy number) was calculated based on APC serial dilutions and averaged between the two duplicates.

3.2.7 Statistical Analyses: Migration Fate

The migration fate of smolts was estimated through supervised random forest modelling as described in Chapter 2 (sections 2.2.5 and 2.2.7). For all further analyses, detection data were truncated to represent movement of live smolts only. For individuals identified as predated by the predation tags, detections were truncated to the last detection of that individual's even numbered (live) ID. For individuals classified as predated by the random forest algorithm but not the predation tags, detections were truncated by removing the average retention time of tags in predators (70 hrs) from the last known detection of that individual tag.

Cormack-Jolly-Seber (CJS) models were used to estimate survival (Phi) and detection (p) probabilities between and at receiver stations to identify areas of high mortality. CJS models were set so Phi and p were allowed to vary for each receiver station (Phi ~ array, p ~ array). In 2019, the type of water (fresh water or tidal water) the station was located in was also included as a covariate (Phi ~ array + water, p ~ array + water). Additionally, areas of high predation were identified through the last known live location of predated smolts.

3.2.8 Statistical Analyses: Migration Behaviour and Fate

A multinomial logistic regression was used to determine the effect of morphological, behavioural, and tagging factors on the probability of belonging to one of the three fate groups. The variables included in the model were as follows: weight (g), fork length (cm), release date (Julian day), release site (tidal or fresh water), tagging year, migration rate (river km/day), number of reversals in swimming direction, number of pauses (>24 hrs spent at the same receiver or between two consecutive receivers), and percent of receiver stations arrived at and departed from at night. Since migration success is based on detection at the mouth of the Shubenacadie River and behaviour is expected to change once smolts have entered the Minas Basin, only detections from stationary receivers within the river system were used to calculate these variables.

ANOVAs and Tukey tests were used to compare each variable between fate groups. Type III ANCOVAs were used for weight and fork length comparisons between years and fate groups.

Multinomial logistic regressions were conducted using the *multinom* function in the *nnet* package (Venables & Ripley, 2002) in R. Individuals not detected after release (n=8) were removed from the data set for this analysis. The levels of the response variable, fate, were set

relative to successful migrants. Alternate models were tested, swapping out co-linear variables, the model with the highest McFadden's pseudo R^2 was used as the global model for stepwise variable selection using *stepAIC* in the *MASS* package (Venables & Ripley, 2002). P-values for the relationship between each variable and fate level were calculated using the Wald test. The function *allEffects* in the package *effects* (Fox & Weisberg, 2019) was used to plot partial effect plots for each of the variables retained in the final model.

3.2.9 Statistical Analyses: Physiological Status and Fate

A MANOVA was run to determine if gene expression levels were significantly different between the three fate groups.

A principal component analysis (PCA) was used to reduce dimensionality of the gene expression data. The PCA was run with a correlation matrix of the 49 host gene expression levels and the plate number each sample was run on using the *princomp* function in base R. Principal components (PCs) were retained so that cumulative variance explained was $\geq$ 50%. The function *find.clusters* in the *adegenet* package (Jombart, 2008) was used to run successive k-means clustering to identify the optimal number of clusters, k, in the gene expression data based on within sum of squares values and the number of retained components. The maximum number of clusters considered was one tenth of the number of observations rounded (max.n.clust=6). The optimal k value was selected as the value before decreases in within sum of squares values switched from sharp to mild declines where sharp and mild are defined by Ward's clustering method. The function also determined which observations (smolts) belong in which cluster.

Biplots of principal components where points represent individual smolts were plotted and the relationships between clustered smolts, gene expression, and fate was visually assessed. Additional factors such as release and capture date were also assessed.

3.3 Results

3.3.1 Migration Fate

The number of smolts estimated by random forest modelling to have survived out-migration from the Stewiacke River to the Minas Basin was lowest in 2017 (4%) and highest in 2019 (57.1%; Table 3.2). Modeled predation rates decreased each year from 2017 to 2019, while rates of other mortality (avian predation, disease, stress, etc.) were relatively constant.

Cumulative survival for the entire migration route based on CJS estimates were 10.4%, 51.3%, and 62.6% for 2017, 2018, and 2019, respectively. These estimates are 5.5-9.3% higher than the observed percent of successful migrants (Table 3.2). CJS models provided survival estimates for river sections where detection efficiency was less than 1.0 (Table 3.3) and observed rates were limited to ranges. Otherwise, observed survival rates were similar to rates estimated by CJS models with some exceptions in 2019 towards the end of the migration route (Figs. 3.1-3.2). There was low total survival in 2017 and high mortality observed throughout the migration route (steep decline in cumulative survival). 2018 survival declined at greater rates after Kent Farm. There was less mortality on Striped Bass spawning grounds (Eddy Pool to Moxam's) in 2019 than 2018 and 2017. High mortality areas in 2019 were centered around Cloverdale and from Moxam's to the Minas Basin. The area around Ridge Rd and Cloverdale has low water flow and is where the confluence with South Branch, which is known to contain many invasive Chain Pickerel, is located. Areas of predation (Fig. 3.3) were consistent with areas of lower

survival (Figs. 3.1-3.2). The location of greatest predation was around Kent Farm which is the center of Striped Bass spawning grounds.

3.3.2 Migration Behaviour and Fate

The multinomial logistic regression model of best fit as determined by AIC retained only year, migration rate, and number of reversals as explanatory variables (ΔAIC of global model = 7.45). Probability of being a successful migrant was lowest in 2017 and highest in 2019 (Fig. 3.4). Probability of being a predated smolt decreased with increasing migration rate (Fig. 3.4). Migration rate was significantly different between fate groups (p=0.004). Predated smolts had significantly slower migration rates than successful migrants in 2017 (p=0.016) and 2018 (p=0.002). Migration rates were not significantly different between fate groups in 2019 (Fig. 3.5). The probability of predation increased with increasing number of reversals (Fig. 3.4). Number of reversals was not significantly different between fate groups (p=0.156; Fig. 3.6).

Release day, although not retained in the final model, was significantly different between fate groups with all years combined (p=0.043). On average, successful migrants were released later than mortalities and predated smolts. However, it was not significant when testing each year individually. No other variables used in the logistic regression were significantly different between fate groups.

3.3.3 Physiological Status and Fate

There was a low detection rate for the pathogens tested in the smolts sampled in 2019. Only three of 18 pathogens were detected in seven of the 56 smolts (Table 3.4). No smolt had more than one pathogen present and pathogen loads were very low (mean 94.0, range 0.4-506.4; Table 3.4) in most of the infected smolts. The pathogens detected were *Flavobacterium*

psychrophilum, piscine orthoreovirus, and *Piscichlamydia salmonis*. *F. psychrophilum* causes bacterial cold-water disease and can induce ataxia or spiral swimming (Starliper, 2011). Piscine orthoreovirus causes heart and skeletal inflammation (HSMI) (Palacios et al., 2010). *P. salmonis* causes gill epitheliocystis (Draghi et al., 2004). The infected smolts carried a range of pathogen loads and had varying migration fates, but there was no apparent trend in the relationship between pathogen presence and survival, however, the small sample size limits conclusions.

The 49 host genes quantified were not significantly different between migration fate groups (MANOVA: Hotelling-Lawley = 16.64 ~ F(2,53)=0.68, p=0.823). Principal components analysis of overall gene expression data did not reveal grouping of smolts by fate on any of the four PCs retained (Fig. 3.7). Further investigation of smolts that were grouped on PCA biplots revealed that there is some relationship between gene expression and release (sampling) date. Grouping of points by release date was stronger when smolts were categorized by release week (week 1: May 20-27, week 2: May 28- June 3, and week 3: June 4-12; Fig. 3.8).

K-means clustering revealed four clusters when four PCs were retained. There was some overlap between clusters on PCA biplots, however, cluster 4 stood out on the positive end of PC1, cluster 3 extended down the negative end of PC2, and cluster 2 grouped on the negative end of PC3 (Figs. 3.7-3.8). Individuals belonging to cluster 4 were released in week 2 (Fig. 3.8). However, half the smolts from week 2 were placed in cluster 2. Cluster 1 consisted of 22 of the 24 smolts released in week 1. Individuals released in week 3 were in cluster 3. Cluster 2 was predominantly comprised of smolts released in week 2 and some from week 3.

Only genes with loadings > |0.2| on at least one of the first four PCs were considered when analyzing biplots (Figs. 3.7-3.8). C-type lectin receptor A (CLEC4E), a marker of imminent mortality was strongly associated with the positive end of PC1 and appears to be

driving the grouping of cluster 4 (Fig. 3.10). However, quality control revealed a chip effect for plate No. 3722, and all individuals in cluster 4 were the only samples assayed on this plate. The seven smolts from week 2 that were in cluster 2, rather than cluster 4, were not on plate 3722. Plate No. was positively associated with all four PCs (Plate No. range 3718-3723). All other genes were associated with the negative end of PC1 (Fig. 3.9). Genes within the viral disease development (VDD) category were grouped on the negative end of PC2 with cluster 3 (Figs. 3.7-3.9). Three of the seven smolts that had positive detections of pathogens were associated with VDD genes. Smoltification markers (one ion regulation gene and three immunity genes) were associated with the negative end of PC3 with cluster 2 (Figs. 3.7-3.9).

Heat shock protein 90 (HSP90) was highly upregulated in smolts released in week 1 (Fig. 3.10). HSP90 was positively associated with PC2 and PC4 (Fig. 3.9). Salmon hyperosmotic protein 21 (SHOP21), a marker of osmotic stress, was upregulated in cluster 2 (Fig. 3.10) and negatively associated with PC3 (Fig. 3.9).

3.3.4 Behaviour and Physiology

Exploratory analysis revealed that smolts released in the first week of the sampling period in 2019 (May 20-27) paused before entering tidal waters and therefore took a longer time to reach the head-of-tide from release than smolts released in weeks 2 and 3 (Fig. 3.11). There was a quadratic relationship between release day and time to head-of-tide (Fig. 3.12). I hypothesized that this pattern could be related to osmoregulatory ability or smoltification stage, therefore, expression of smoltification biomarkers were compared between the three release weeks (Fig. 3.13). Carbonic anhydrase 4 (CA4), Na^+/K^+-ATPase alpha subunits 1b (NKAa1.b) and 1c (NKAA1C) are genes related to osmoregulation and are expected to be upregulated during smoltification (Seear et al., 2009; Houde et al., 2019). Interleukin-12 beta (IL12B), C-C

motif chemokine 4 (CCL4) and 19 (CCL19) are genes related to immunity and are expected to be downregulated in smolts due to an energetic trade-off between immune function and osmoregulation/homeostasis (Houde et al., 2019). FK506-binding protein 5 (FKBP5) is also an immune function gene but is expected to be upregulated in smolts (Houde et al., 2019). CA4 expression was lower in week 1 than week 2 releases (p=0.002). NKAa1.b and NKAA1C were upregulated in week 1 releases (p=0.015 and p=0.007, respectively). CCL19 and CCL4 expression was lower in week 1 releases (p=2.0^{-5} and p=0.0002, respectively), but IL12B expression was higher in week 1 releases (p=0.0004). Expression of FKBP5 was higher in week 2 than week 1 (p=0.002) and week 3 (p=0.003) releases.

Additionally, the same PCA and clustering methods used to analyze all biomarkers were applied to only the smoltification markers. Two PCs were retained, and three clusters were identified (Fig. 3.14). Week 1 releases were grouped into cluster 2 and loaded with IL12B, NKAA1C, and NKAa1.b. Weeks 2 and 3 releases were split among clusters 1 and 3 and loaded with the remaining genes, FKBP5, CCL19, CCL4, and CA4.

Environmental factors were also compared to timing of entry into tidal waters (Fig. 3.15). It appears that peak water discharge coincides with peak movement into the head-of-tide. Lunar cycle and water temperature were also examined but there was no clear relationship to smolt movement at the head-of-tide.

3.4 Discussion

3.4.1 Migration Fate

In this study we examined the river migration fate of 156 Atlantic Salmon smolts, sampled over three consecutive years. There was an estimated upwards trend in successful

migrants across the three study years that appears to be driven by a reduction in predation. Migration rate is a potential behavioural predictor of fate, with slower moving smolts more likely to be predated than faster moving smolts. There was no physiological predictor of migration success in the 56 fish examined in 2019.

Gibson et al. (2015) estimated survival of smolts in the Stewiacke River to be 41.1% in 2008 and 19% in 2011. These rates are lower than survival estimated in 2018 and 2019 but greater than 2017. There is large variation in estimated survival rates in this river between years. Estimated predation accounted for a lower proportion of total mortality in 2008 and 2011 than 2017-2019, this may reflect increasing predator abundance. Mortalities of unknown cause were still estimated to account for 16-38% of total mortalities in 2017-2019.

Estimated survival was extremely low in 2017; however, migration success or failure was determined via detection at the last river receiver station located at the river mouth, which we know in 2019 was not 100% efficient. In 2019, some smolts passed through the Shubenacadie River mouth without being detected by the receivers stationed there, but smolts were later detected at receivers in the Minas Basin, which were not yet deployed in 2017. Detection efficiency at the river mouth was estimated to be 88% percent, potentially due to the spacing between receivers not covering the whole of the river mouth or the great changes in water flow throughout the tidal cycle. It is therefore possible that estimated survival might have been higher in 2017 if receivers had been placed in the Minas Basin at that time. Alternatively, the 2017 tagging site was closer to the head-of-tide and known Striped Bass spawning grounds than the other two study years, which could have been responsible for the high total mortality estimates that year. A combination of handling and osmotic stressors may have reduced the ability of smolts to detect and evade predators (Handeland et al. 1996; Diepernek et al., 2002; Russell et

al., 2012; Daniels et al., 2019). Migratory behaviour is also often affected by osmotic stress, leading to increased upstream movement and/or increased residency time to acclimate to the change in salinity (Halfyard et al., 2012; Halfyard et al., 2013). It is possible that the placement of the smolt wheel in 2017 and 2018 near the head-of-tide prevented tagged smolts from adapting their behaviour in response to osmotic stress. The first group of smolts tagged in 2019 paused just before the head-of-tide and did not enter tidal waters prior to June 4th, potentially due to underdeveloped osmoregulatory ability.

Further, 2017 smolts were released within the first week of the smolt run (May 19-27), and therefore it is possible that smolts from earlier in the smolt run are not as fit or encounter sub-optimal environmental conditions compared to smolts that begin migration later in the run. Release dates for 2018 and 2019 were May 24-June 15 and May 20-June 12, respectively. Stich et al. (2015) found that Atlantic Salmon smolt survival was highest in the middle of the run and when smolts arrived at the estuary in the middle of the run. Release day had an effect on migration fate when all years were combined but not within each year individually. This is likely due to the very low survival estimates and early release dates in 2017 compared to higher survival estimates and more spread-out release dates in 2018 and 2019.

Additionally, smolts tagged in 2017 and 2018 could have been from a mixture of tributaries, whereas smolts tagged in 2019 were only from the Pembroke River. Environmental differences among tributaries can cause differences in the quality of smolts or the exact timing of migration. Whalen et al. (1999) found that Atlantic Salmon smolts in warmer tributaries migrated earlier and had higher Na^+/K^+-ATPase activity than smolts from cooler tributaries of the West River, Vermont. Distance of the tributary to the marine environment can also affect migration timing (Stich et al., 2015).

The large variation in estimated survival among study years may also be due to differences in environmental factors and migration timing of various species among years. The interaction between the physiological and environmental window of migration timing affects smolt survival (McCormick et al., 1998). Water temperature plays a role in timing smolt migration (Whalen et al., 1999; Jutila et al., 2005; Otero et al., 2014), is related to Na^+/K^+-ATPase activity (Whalen et al., 1999; Strand et al., 2011), and can cause stress and lead to indirect mortality (Miller et al., 2011; Hinch et al., 2012). Water temperature also dictates the timing of Striped Bass spawning. The smolt run typically begins when water temperature is at 10°C (Jutila et al., 2005; Otero et al., 2014) and temperatures must be above 13°C for Striped Bass spawning (Bradford et al., 2015). Higher predation rates are expected when there is a greater temporal overlap between the two species' migrations. The timing of the smolt run appears to be consistent among study years. IBoF salmon have a later smolt run than other Atlantic Salmon populations (SARA, 2010) which increases the odds of direct overlap with Striped Bass spawning. It is likely that the benefits of the environmental conditions encountered at this time outweigh the risks of a high predation landscape (Sabal et al., 2021). The salmon population in this river is small especially compared to the number of Striped Bass and other predators. Therefore, typical predator avoidance strategies such as predator swamping are unlikely to be effective. This makes the migration timing of alternative prey species such as Alewife and Blueback Herring important because they can act as a prey buffer for smolts (Svenning et al., 2005; Saunders et al., 2006; Furey et al., 2020).

In all years of study, the locations of high mortality along the migration route are consistent with locations of high predator abundance. It can thus be assumed that predation is playing a large role in out-migration mortality of smolts in this river system. However, other

rivers in the iBoF unit also experience high out-migration mortality but do not have the elevated predation pressure from spawning Striped Bass (SARA, 2010). Therefore, it was expected that predated smolts in the Stewiacke River had a predisposition to mortality and that predation in this system is substituting other sources of mortality (disease, heat stress, etc.) that would be the primary direct causes of smolt mortality in other rivers but would be indirect causes in this system. However, the results of this thesis do not support this because only one behavioural factor, slower migration rate, and no physiological factors were found to have a strong correlation to increased predation risk. There is also the possibility that given the high concentration of predators and small numbers of migrating smolts, that predators are not differentially selecting for weaker prey but are feeding opportunistically.

3.4.2 Migration Behaviour and Fate

Through multinomial logistic regression, migration rate was identified as the most important migratory behaviour in relation to fate. Predated smolts consistently had slower migration rates than successful migrants. It is possible that these smolts were experiencing stress or had some other underlying cause that resulted in a decreased swimming ability, making these smolts more vulnerable to predation. In 2017 and 2018, where the relationship between fate and migration rate was significant, there was a shorter distance between the release site and predator dominated areas than in 2019. Potentially, the predated smolts were still experiencing handling stress following release resulting in slower movement and a higher probability of being predated. Flavio et al. (2020) similarly found that smolts swam at low speeds after release then increased speed further away from the release site.

Although not statistically significant, the number of reversals that smolts displayed was the only other behavioural variable retained in the final regression model in addition to migration

rate. Effect plots showed that a greater number of reversals increased the probability of a smolt belonging to the predated fate class. A high number of reversals in swimming direction can increase residency in a high predation landscape (McCormick et al., 1998; Halfyard et al., 2012). Reversals are also thought to be a response to osmotic stress, which is known to increase susceptibility to predation (Handeland et al. 1996; Dieperink et al., 2002; Halfyard et al., 2012; Halfyard et al., 2013) or due to upstream water flow from incoming tides (Beland et al., 2001). Additionally, reversals in swimming direction is a behaviour more commonly seen in spawning Striped Bass than salmon smolts (Romine et al., 2014; Gibson et al., 2015; Daniels et al., 2018). Despite the removal of post-predation detections from the dataset prior to analyses, it is possible that some of the reversals included in the regression were made by Striped Bass that had consumed tagged smolts and were detected before predation was registered.

Factors such as smaller body size and decreased nocturnal movement which have been found to increase predation susceptibility in smolts (Hostetter, 2009; Ibbotson et al., 2011; Clark et al., 2016; Furey et al., 2016; Tucker et al., 2016; Flavio et al., 2019), were not found to be significant here. That body size was unrelated to fate could be due to the fact that only smolts longer than 12 cm were tagged to avoid exceeding maximum recommended tag burden. Nocturnal migration, which was not generally observed, may not have been an important factor because the tidal waters of the Stewiacke and Shubenacadie Rivers are very turbid reducing the need to avoid visual predators by travelling at night.

The results indicate that slow migration rate and a great number of reversals in swimming direction increase predation susceptibility for smolts, physiological status and external stressors may have caused these behaviours.

3.4.3 Physiological Status and Fate

Physiological status and pathogen presence were not found to have a direct association with migration fate in the 56 fish assessed in 2019. The low number of detected pathogens limited analyses and conclusions on the relationship between migration fate and pathogen presence. From the information available it does not appear that the presence of pathogenic microbes is indicative of migration fate in the sampled smolts. These results do however show that there is a low occurrence of smolt infection by microbial pathogens in the Pembroke River which is a positive point for salmon in this system.

Time from infection with *F. psychrophilum* to death is 10 days on average for juvenile salmonids (Holt et al. 1989). The infected smolt lived for 16.2 days prior to predation. The five smolts infected with *P. salmonis* were captured on 2019-05-21 and released on 2019-05-22. It is possible that transfer of this bacterium occurred while smolts were being held. The loads of all detected pathogens were presumably too low to have a great impact on host physiology, at least at the time of sampling. While the relationship between pathogen load and physiological impact varies by pathogen, physiological effects are typically not present below 1000 RNA copies (K. Miller, pers. comm., October 2021), and here, copy numbers were less than 507. There may have been other pathogens present that were not tested for or some pathogens may have been more prevalent in other areas of the body. Additionally, smolts were sampled early in the migration route, and pathogens are expected to accumulate throughout migration (Chapman et al., 2020).

The PCA clustering of smolts based on CLEC4E in PC1, which can be upregulated in response to parasites (Robledo et al., 2018), and VDD genes along PC2 implies that there may be pathogens present that were not detected or were not among the 18 pathogenic species tested for. Cluster 3 which shows higher expression of six of the VDD genes contains three smolts that

tested positive for microbial pathogens, it is possible that other smolts in this cluster were also infected by pathogens.

Gene expression, which was applied as a measure of physiological status, was not related to estimated migration fate based on PCA clustering, however, the power to detect this relationship was low with only 56 fish. In addition to the VDD genes (driving PC2), smoltification (driving PC3), thermal stress and osmotic stress markers were the genes most strongly associated with principal components and clusters. A relationship between gene expression and release date was observed and is likely a reflection of changes in environmental conditions and progression through the smoltification process throughout the smolt run.

3.4.4 Behaviour and Physiology

The observed station holding before the head-of-tide in 2019 week 1 releases appears to be partially related to osmoregulatory ability and smolt stage because there were differences in the expression of smoltification biomarkers between release weeks. Week 1 releases appear to be carrying a physiological signal of pre-smolts, and expression patterns in the faster moving week 2 fish are more consistent with full smolts. Week 3 fish showed more variability and may have contained a more mixed group of smolt-ready fish, potentially including de-smolts. De-smolts are smolts that have remained in fresh water for too long during the out-migration period and have reverted to a physiology suited for fresh water rather than salt water (Houde et al., 2019). The smoltification panel of genes is designed to classify pre-smolts, full smolts, and de-smolts (Houde et al., 2019). Through PCA and k-means clustering, the panel successfully identified three clusters of smolts. The cluster comprised of week 1 smolts most likely represents pre-smolts, but the other two clusters are a mixture of week 2 and week 3 smolts and were less easy to classify.

Whalen et al. (1999) found Na^+/K^+-ATPase activity increased from the pre-migratory period until it reached peak activity at which point 97% of smolts had initiated migration. Na^+/K^+-ATPase activity was then measured to have returned to pre-migratory levels in juveniles that did not migrate. There was higher expression of Na^+/K^+-ATPase subunits in week 1 releases which is consistent with expected expression of these genes in smolts that are salt water prepared rather than pre-smolts (Whalen et al., 1999; Strand et al., 2011; Stich et al., 2015; Houde et al., 2019). However, Houde et al. (2019) found that NKAA1C and NKA1.b genes are not as consistent at classifying smoltification stage as other biomarkers (CA4 and immunity genes) because they also respond to temperature and may only significantly increase closer to salt water entry. In contrast, CA4 which is also expected to be upregulated in smolts, shows lower expression in week 1 releases than weeks 2 and 3, which is consistent with pre-smolts. Additionally, IL12B and FKBP5 expression patterns across weeks were consistent with what would be expected as individuals progress through smoltification stages. However, CCL4 and CCL19 expression patterns were not.

Alternative explanations for this station holding are that smolts had paused migration to sync up with conspecifics before entering a more predator dominated field or were waiting for better environmental conditions (temperature, water discharge, food availability, etc.). Peak water discharge measured in a nearby river aligned with peak entry of smolts into tidal waters. Increased water discharge may be beneficial to smolts by decreasing salinity and reducing osmotic stress, increasing turbidity to aid in predator avoidance, and increasing current speed for faster migration. Furey et al. (2020) found that smolt survival was highest in fast-flowing areas with high turbidity. The increased current speed may have also simply pushed smolts downstream.

3.4.5 Conclusions

From 2017 to 2019, Atlantic Salmon smolt survival in the Stewiacke River watershed was estimated to increase with decreasing predation rates. Predation was not the sole cause of mortality but was found to be the primary cause. Migration rate was an important behavioural predictor of fate in this system. No relationship was found between smolt physiology and migration fate; however, conclusions are limited due to small sample size and a single sampling event. There may be other host genes or pathogens that were not tested for here that are related to fate and predation susceptibility. The potential for interactions between external and internal stressors, namely osmotic stress, to be either directly or indirectly causing mortality still exists. Further sampling at different time points and locations along the migration route may help to uncover these underlying mechanisms.

Tables

Table 3.1 Gene abbreviations, names, and categories (MRS mortality related signature; VDD viral disease development). *Removed from final analyses due to poor assay efficiency.

Gene abbreviation	Gene name	Category
78d	Si:dkey-78d16.1 protein (s100v2)	housekeeping
Coil-P84_R2_tm	coilin	housekeeping
MrpL40_F1_tm	39S ribosomal protein L40	housekeeping
CA4_v1	Carbonic anhydrase 4	smoltification
CCL19_v1	C-C motif chemokine 19	smoltification
CCL4_v1	C-C motif chemokine 4	smoltification
FKBP5_v1	FK506-binding protein 5	smoltification
IL12B_v1	Interleukin-12 beta	smoltification
NKAa1.b	Na/K ATPase alpha-1b (seawater)	smoltification
NKAA1C	Na/K ATPase alpha-1c	smoltification
ALD1_chr3	fructose-bisphosphate aldolase A1	stress mortality
hsp90a_15_v2*	heat shock protein 90 alpha	stress mortality
B2M	beta(2)-microglobulin	immune stimulation
C5aR	anaphylatoxin (receptor)	immune stimulation
HEP	hepcidin	immune stimulation
IFNa	IFN-alpha (interferon)	immune stimulation
IGMs*	immunoglobin	immune stimulation
IL15	interleukin 15	immune stimulation
IL1B	interleukin 1 beta	immune stimulation
ILIR	interleukin-1 receptor complex	immune stimulation
IRF1	interferon regulatory factor 1	immune stimulation
RIG1_MGLSYBR_1	RNA helicase RIG-1	immune stimulation
SAA	serum amyloid protein a	immune stimulation
TF*	transferrin	immune stimulation
EPD_2	ependymin-2	inflammation
ES1_1	ES1 protein homolog	inflammation
GILT_1	gamma-interferon inducible lysosomal thiol reductase	inflammation
IL-17D*	interleukin 17D	inflammation
IL11*	interleukin 11	inflammation
MMP13	matrix metalloproteinase-13	inflammation
MMP25	matrix metalloproteinase-25 precursor	inflammation
tgfb_2	transforming growth factor beta	inflammation
ATP5G3-C	ATP synthase lipid-binding protein, mitochondrial	inflammation
C7	complement component C7 precursor	MRS
FYB	FYN-T-binding protein	MRS
HTA	HIV-1 Tat interactive protein	MRS
KRT8*	Cyclokeratin-8	MRS

Table 3.1 (cont'd)

Gene abbreviation	Gene name	Category
PRAS*	*O. mykiss* G-protein(P-ras) mRNA, complete cds	MRS
RGS21	regulator of G-protein signalling 21	osmotic stress
SHOP21	salmon hyperosmotic protein 21	osmotic stress
AARDC_1*	arrestin containing protein 1	imminent mortality
BSG_1 (not Tuba1C)	basigin	imminent mortality
CLEC4E_2	C-type lectin receptor A	imminent mortality
GLUL_1	glutamine synthetase	imminent mortality
H1F0_1	H1 histone family, member 0	imminent mortality
IQGAP1_2	IQ motif containing GTPase activating protein 1	imminent mortality
ODC1_2	Ornithine decarboxylase 1	imminent mortality
TAGLN3_2	transgelin	imminent mortality
CA054694_MGL_1	mitochondrial ribosomal protein	VDD
GAL3_MGL_2	Galectin-3-binding protein precursor	VDD
HERC6_1	probable E3 ubiquitin-protein ligase HERC6	VDD
IFI44A_MGL_2	interferon induced protein 44	VDD
IFIT5_MGL_2	interferon-induced protein with tetratricopeptide repeats 5	VDD
MX_ONTS	antiviral protein	VDD
PXMP2[UBL1]	Ubiqitin-like protein-1, Peroxisomal membrane protein 2	VDD
RSAD_MGB2	radical S-adenosyl methionine domain-containing protein 2	VDD
VHSVIP4_MGL_3	VHSV-inducible protein-4	VDD
HSC70	heat shock cognate 70 protein	general stress
HSP90	heat shock protein 90	general stress
JUNB	transcription factor AP-1	general stress

Table 3.2 Number of smolts tagged and percent of smolts belonging to each fate group (successful migrant S, mortality M, predation P) by year as estimated by random forest models.

	2017	2018	2019
# tagged	50	50	56
S	4%	42%	57.1%
M	16%	10%	12.5%
P	80%	48%	30.4%

Table 3.3 Observed receiver station detection efficiency, PB fence [2019] and ST wheel [2017, 2018] are release sites. See Appendix B for corresponding receiver station locations.

Station name	2019	2018	2017
PB fence	NA	NA	NA
J Graham	1	NA	NA
Gaults	1	NA	NA
Stewart Hill	1	NA	NA
Ridge Rd	1	NA	NA
Cloverdale	1	NA	NA
Brenton Cross	1	NA	NA
Brenton & Hemlock	1	NA	NA
Scout Ground	1	NA	NA
River Park	0.93	NA	NA
ST wheel	NA	NA	NA
Rockpile	1	1	1
Eddy Pool	1	1	0.94
Porter	1	1	1
Trestle	1	1	NA
Kent Farm	1	1	1
Stewiacke 2	1	1	1
Moxam's	1	1	1
Gosse Bridge	0.75	1	1
Shubie mouth	0.88	1	NA
Minas Basin	NA	NA	NA

Table 3.4 Information on the seven smolts where pathogens were detected to be present out of the 56 smolts tested. Smolt ID (acoustic tag SN), pathogen species, pathogen type, smolt fate (successful migrant S, mortality M, predation P), and pathogen load (RNA copy number; calculated based on APC serial dilutions and averaged between the two duplicates).

Smolt ID	Pathogen	Pathogen type	Smolt fate	Pathogen load
1324762	*Flavobacterium psychrophilum*	Bacteria	P	13.10
1313439	Piscine orthoreovirus	Virus	S	0.96
1313426	*Piscichlamydia salmonis*	Bacteria	S	133.34
1313419	*Piscichlamydia salmonis*	Bacteria	S	3.35
1313422	*Piscichlamydia salmonis*	Bacteria	P	0.52
1313421	*Piscichlamydia salmonis*	Bacteria	P	0.41
1313427	*Piscichlamydia salmonis*	Bacteria	M	506.36

Figures
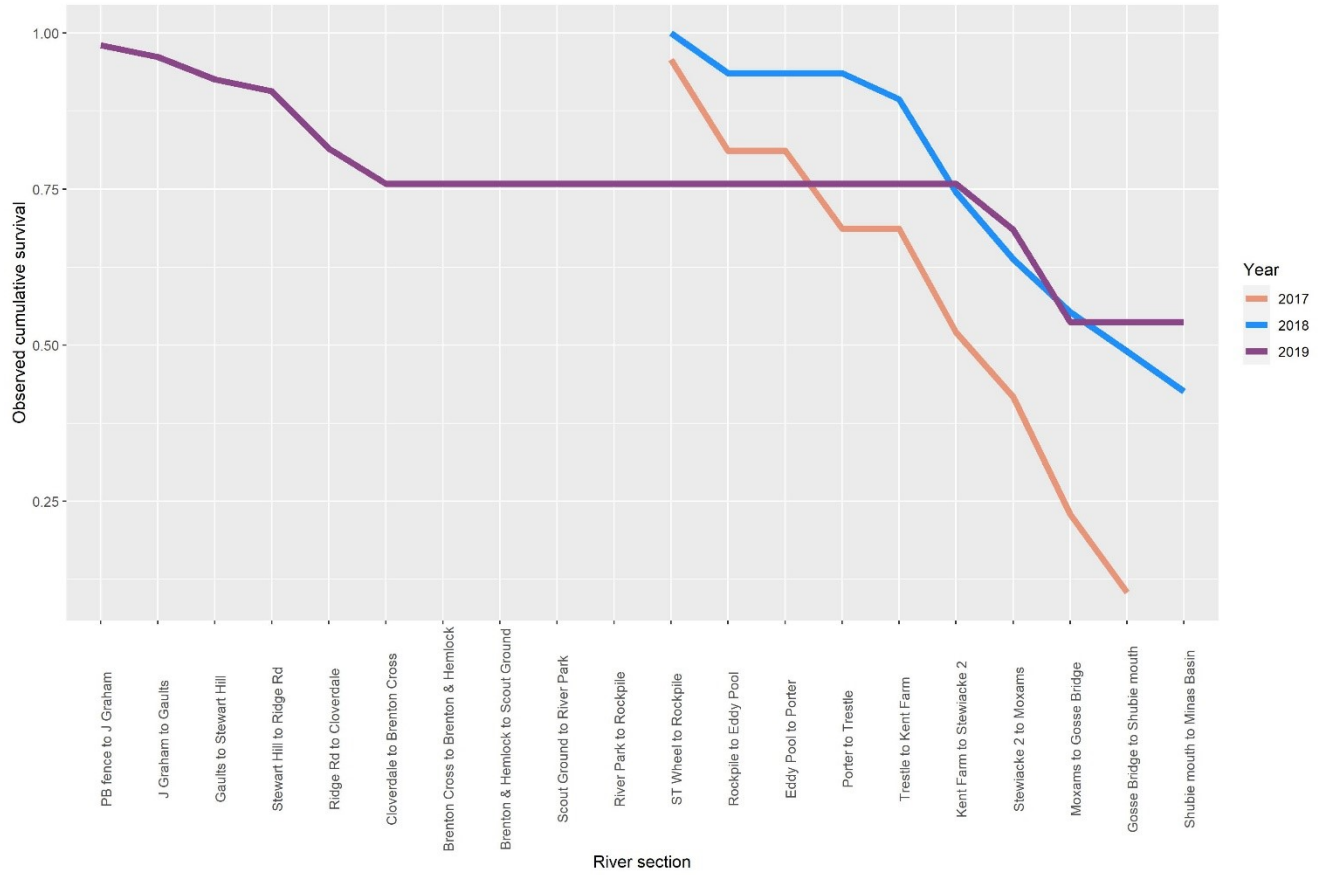
Figure 3.1 Cumulative observed smolt survival rates between receiver stations by each year. See Appendix B for corresponding receiver station locations.
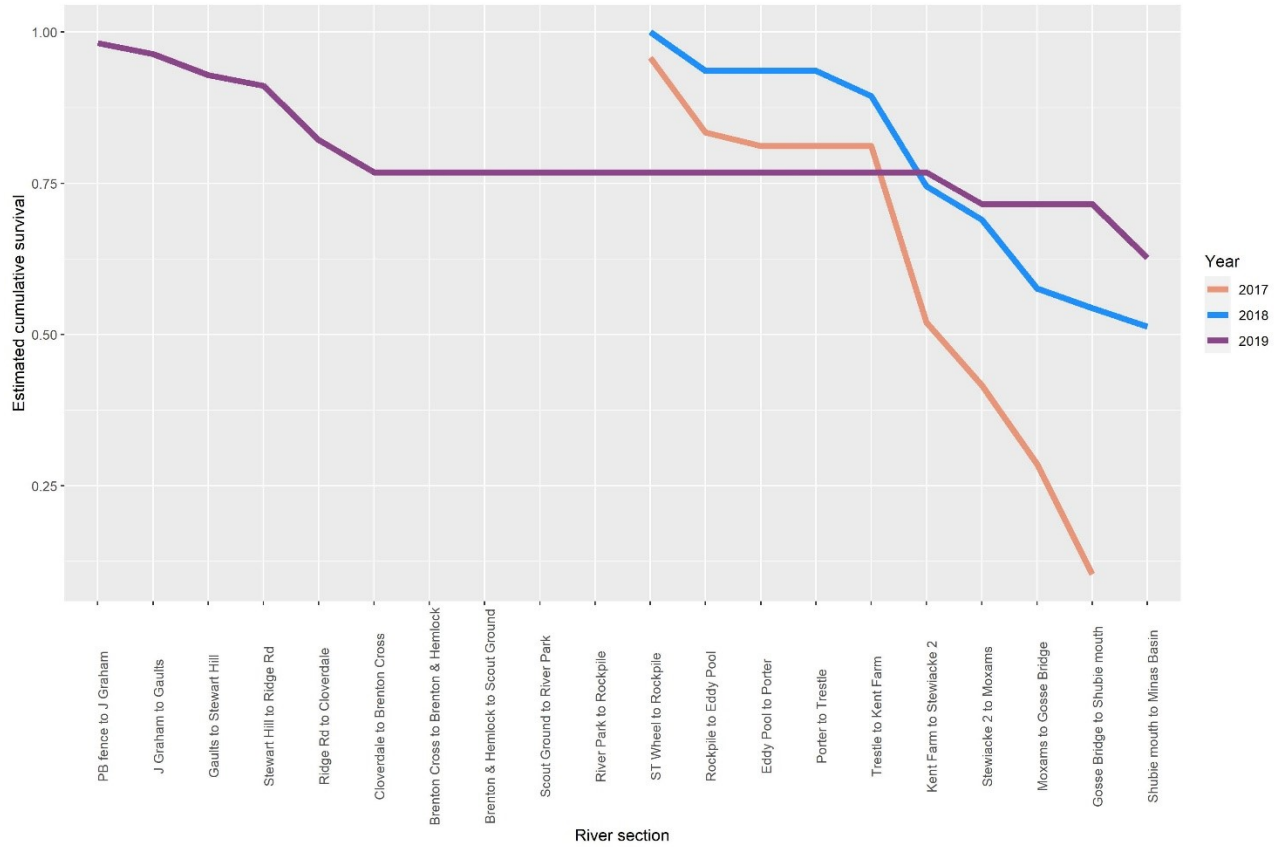
Figure 3.2 Cumulative smolt survival between receiver stations estimated by Cormack-Jolly-Seber models for each year. See Appendix B for corresponding receiver station locations.
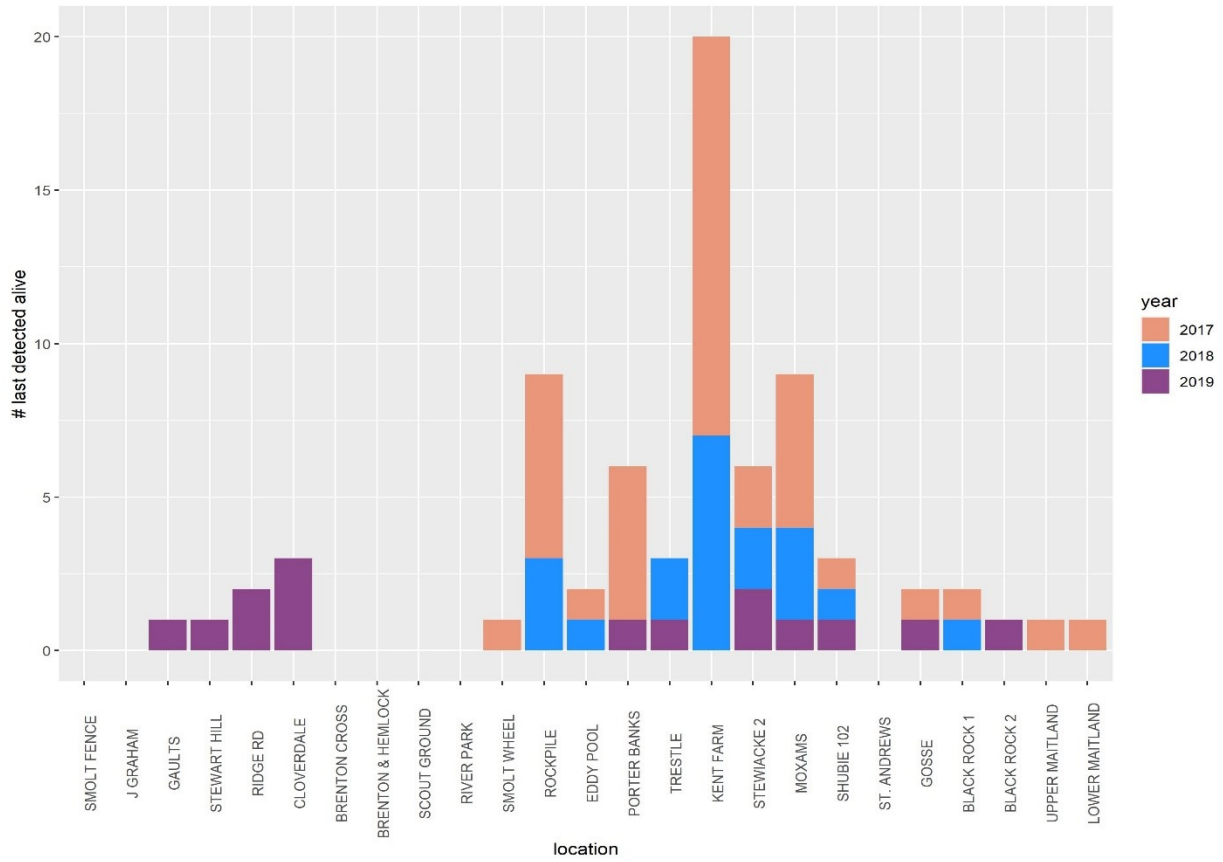
Figure 3.3 Number of smolts last detected alive prior to predation at each receiver location by year. Receiver locations in order from release (smolt fence [2019], smolt wheel [2017, 2018]) to mouth of the Shubenacadie River (Black Rock and Maitland). Striped Bass spawning grounds are from Eddy Pool to Moxam's. Shubie 102 and St. Andrews are above the Stewiacke/Shubenacadie confluence. See Appendix B for corresponding receiver station locations.
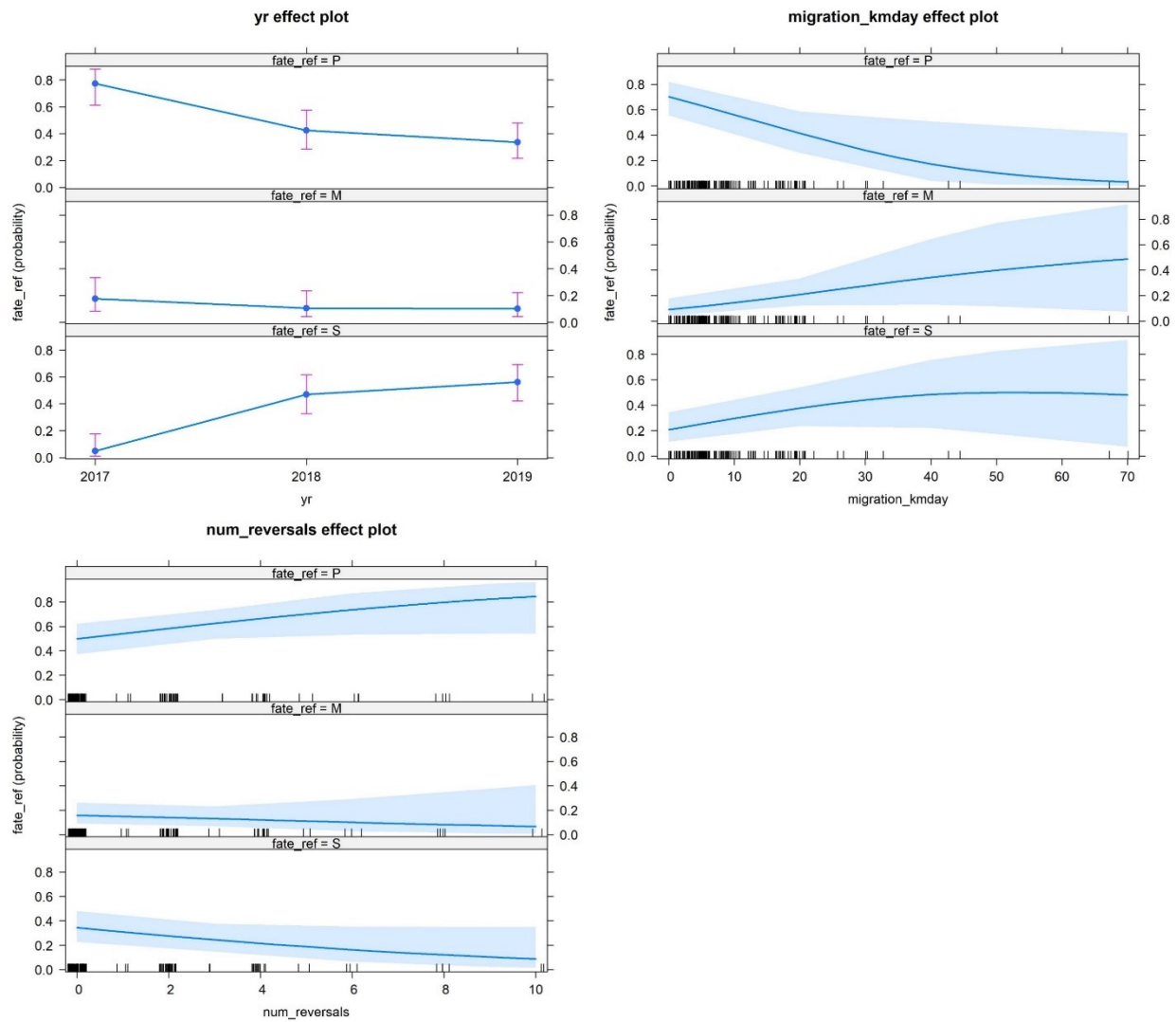
Figure 3.4 Effect plots from multinomial logistic regression: fate ~ year + migration rate + number of reversals.
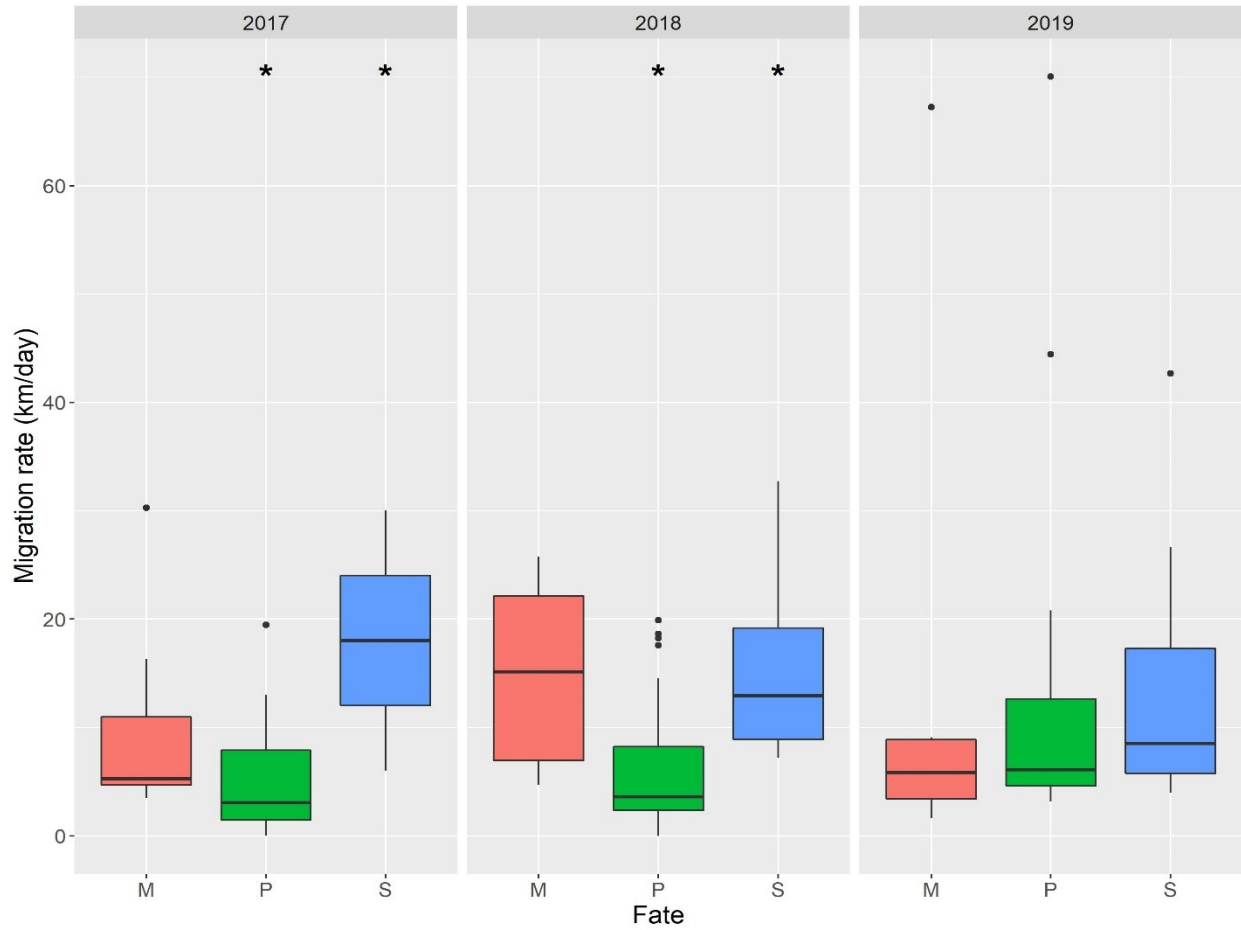
Figure 3.5 Boxplots of smolt migration rate for each fate group (mortality M, predation P, successful migrant S) by year. Asterisks indicate fate groups that are significantly different from each other (p<0.05).
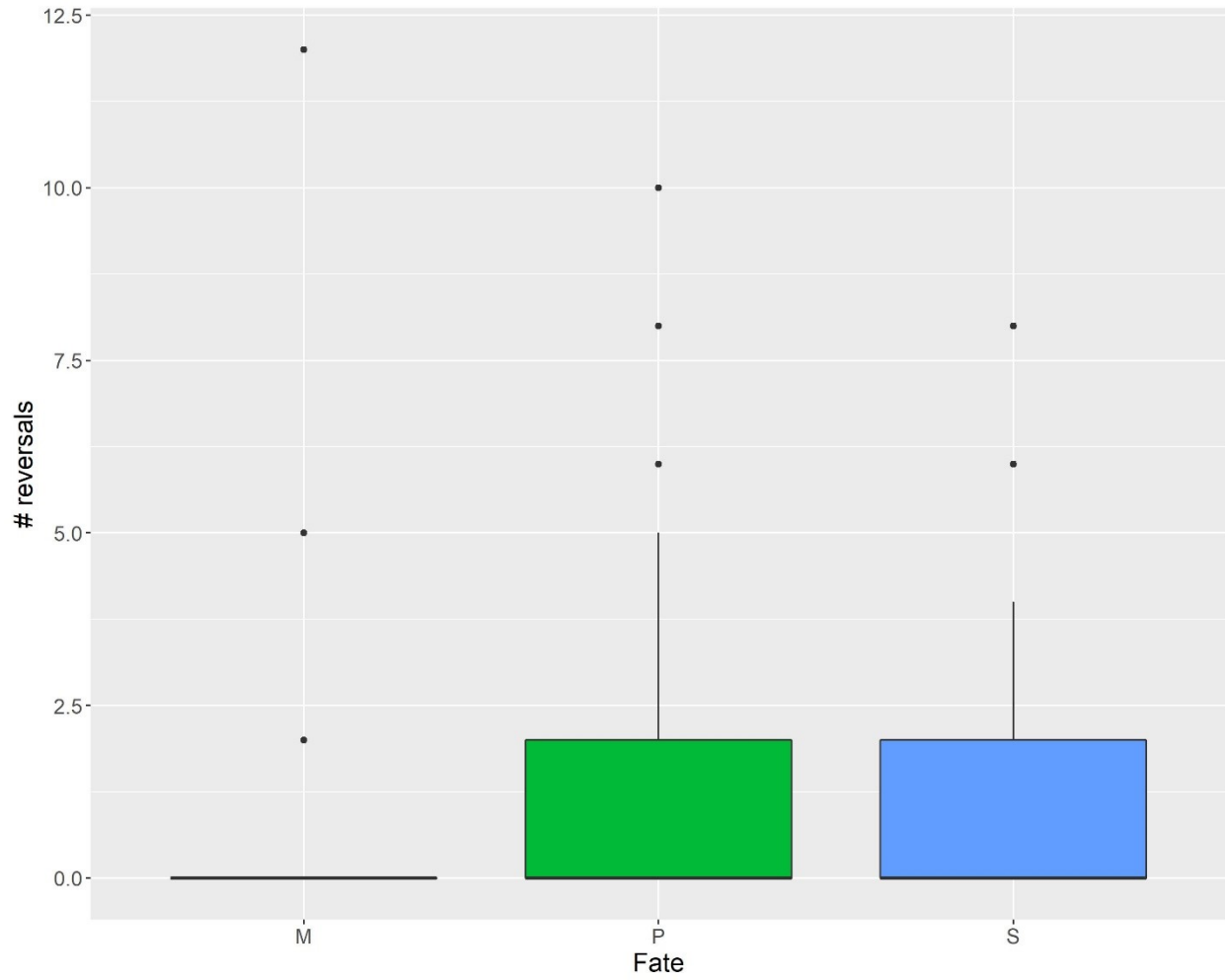
Figure 3.6 Number of reversals in swimming direction for each fate group (mortality M, predation P, successful migrant S) all years combined.
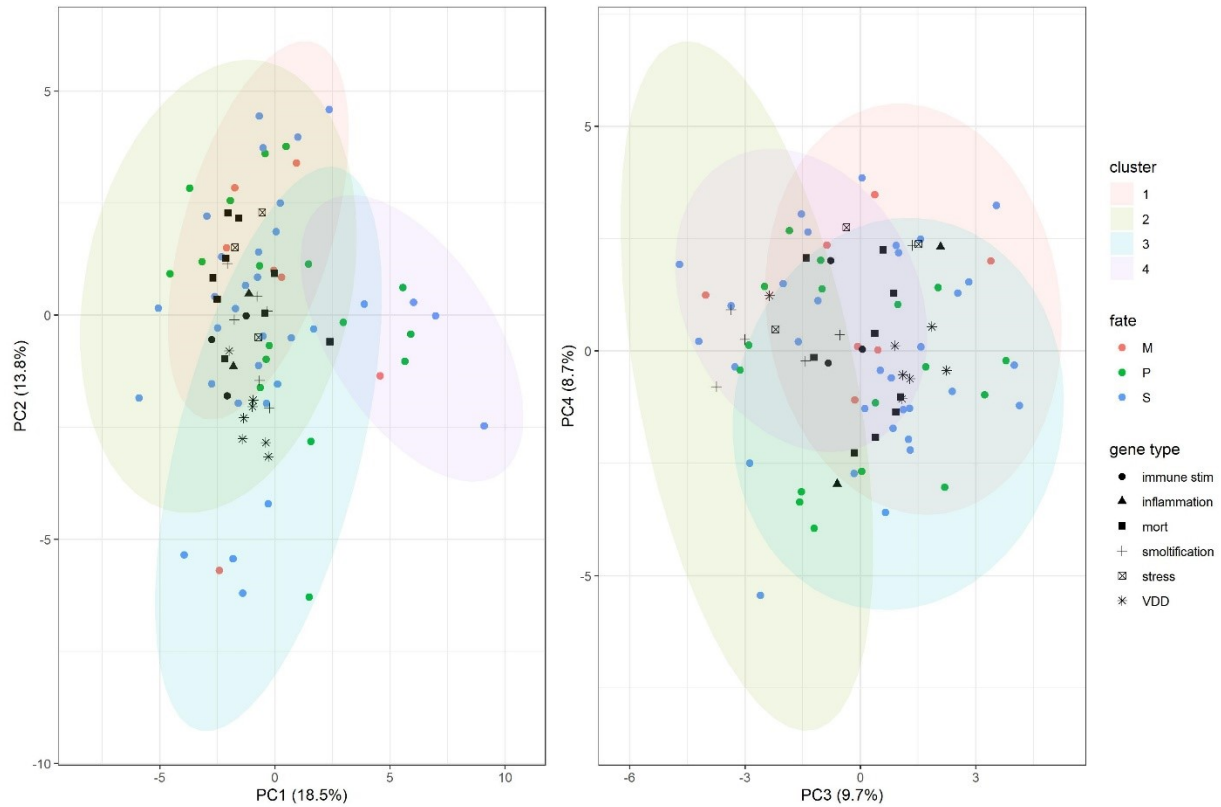
Figure 3.7 Principal components analysis of host gene expression levels in relation to migration fate. Principal components (PC) 1-4 shown (percent variation explained). Points represent individual smolts, colour coded by migration fate (mortality M, predation P, successful migrant S). Ellipses represent cluster as determined by k-means clustering. Black points represent loadings (multiplied by a factor of 10 for easy visualization) of explanatory variables (genes), shaped by gene category. Only variables with loadings > |0.2| were plotted for easy visualization.
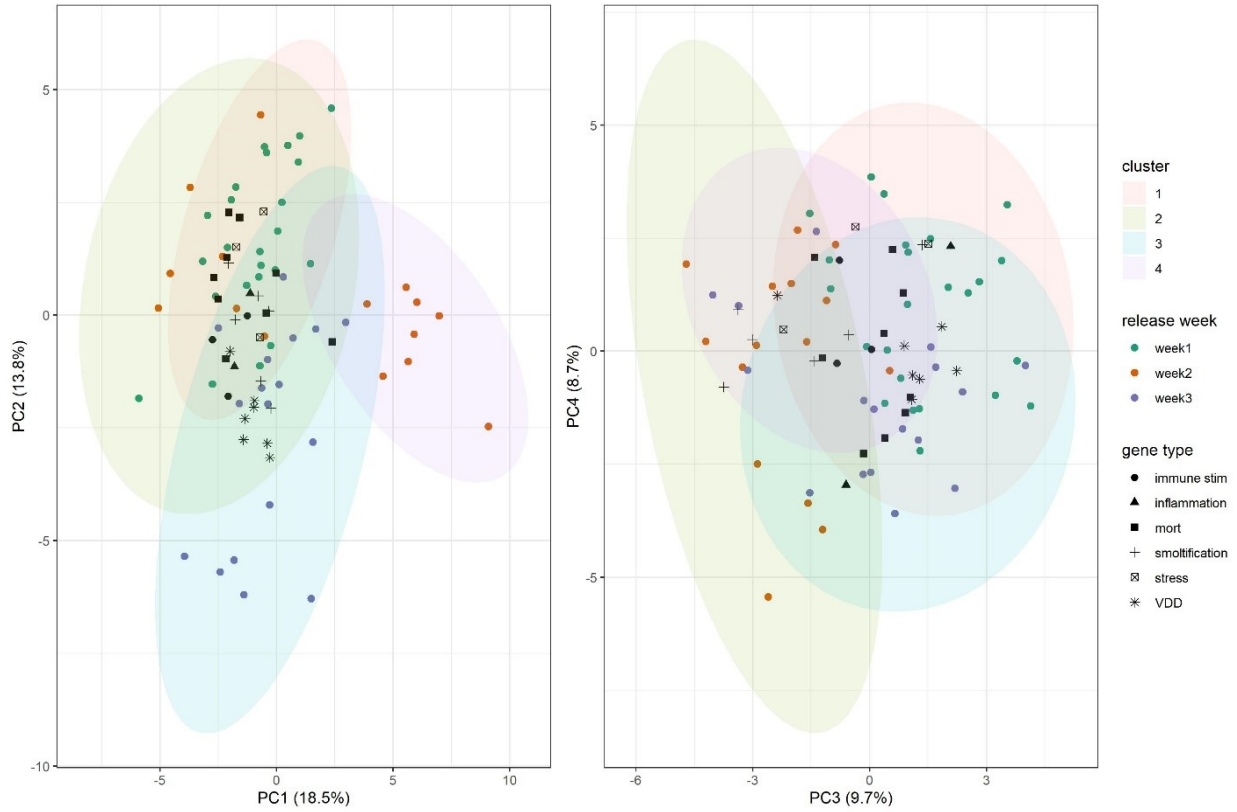
Figure 3.8 Principal components analysis of host gene expression levels in relation to release week. Principal components (PC) 1-4 shown (percent variation explained). Points represent individual, colour coded by release week. Ellipses represent cluster as determined by k-means clustering. Black points represent loadings (multiplied by a factor of 10 for easy visualization) of explanatory variables (genes), shaped by gene category. Only variables with loadings > |0.2| were plotted for easy visualization.
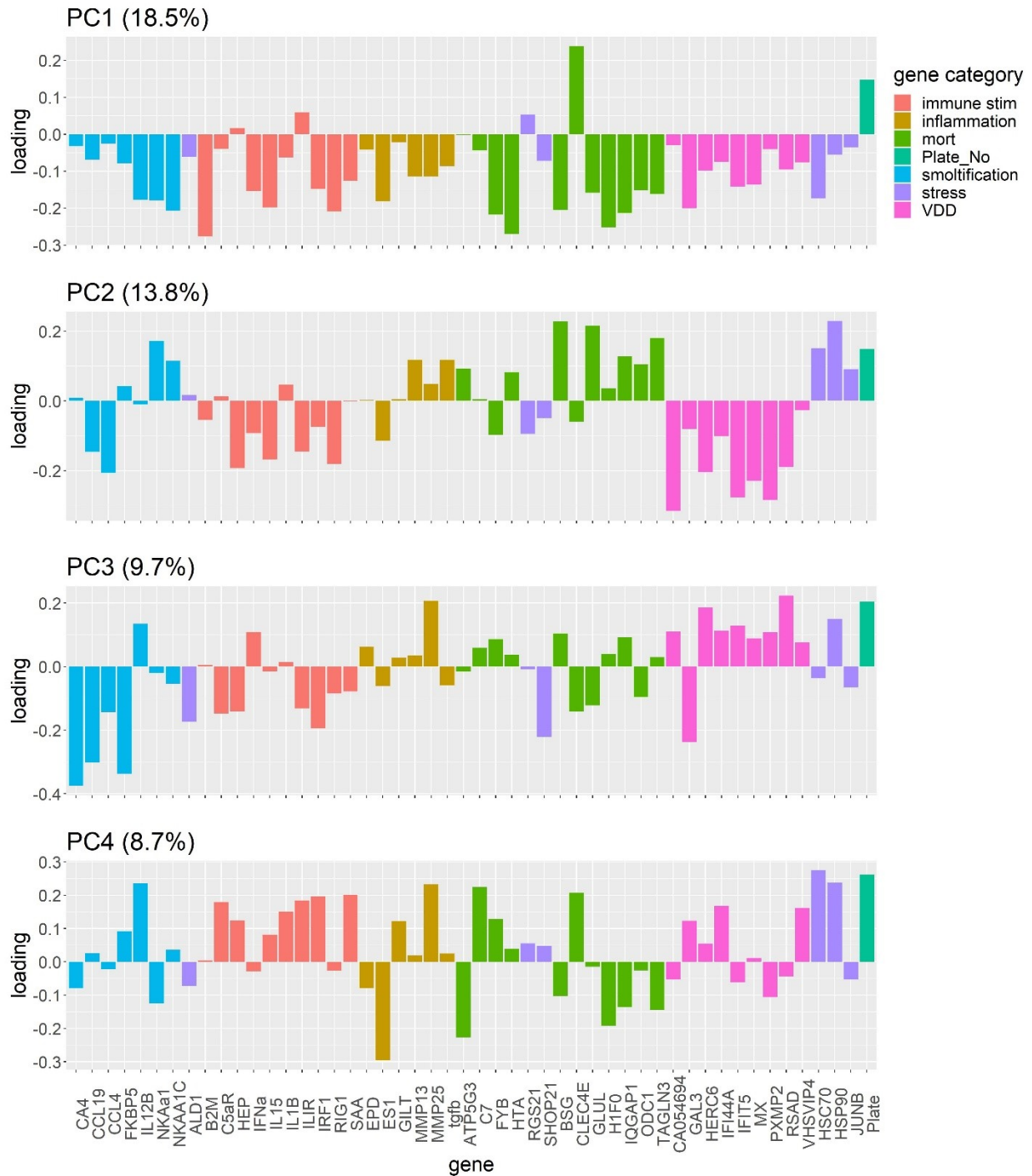
Figure 3.9 Loading values for host gene expression markers on principal components (PC) 1 to 4 (% variance explained). Biomarkers colour coded by category: immune stimulation (immune stim), inflammation, mortality related signature and imminent mortality (mort), qPCR plate number samples were run on (Plate_No), smoltification, stress mortality, osmotic stress, and general stress (stress). See Table 3.1 for gene names.
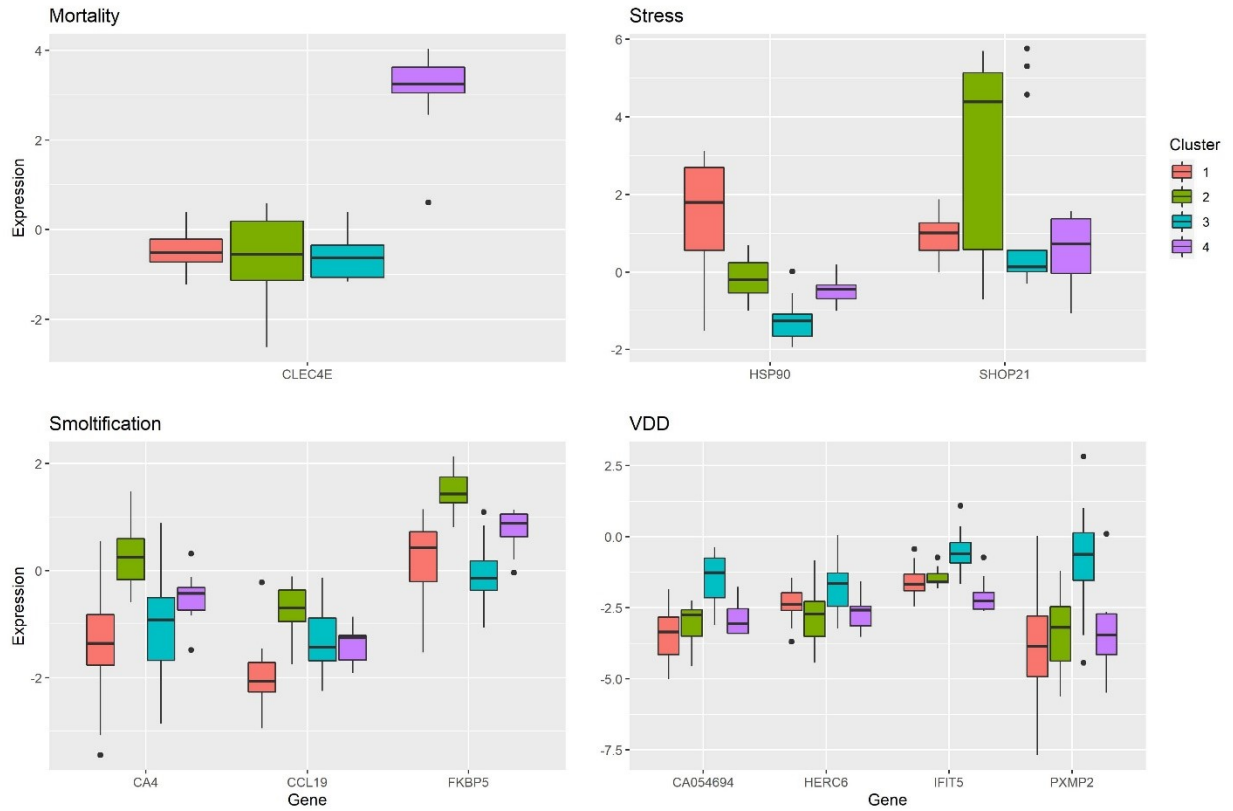
Figure 3.10 Boxplots of expression levels for genes important for clustering on PCA (Figs. 3.7-3.8) separated by cluster and gene category. See Table 3.1 for gene names.
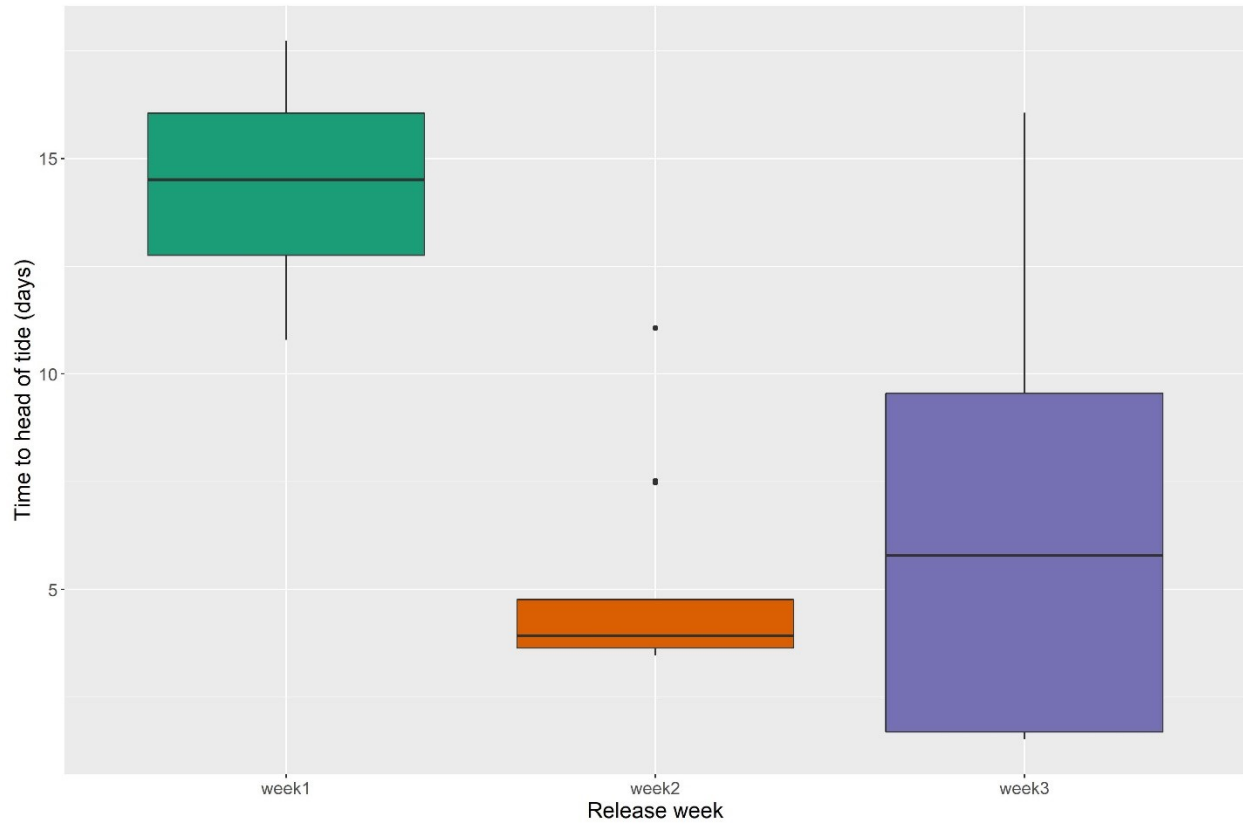
Figure 3.11 Time from release to head-of-tide grouped by release week, only smolts that survived to head-of-tide were included. Release week 1 significantly different from week 2 (p<0.001) and week 3 (p<0.001).
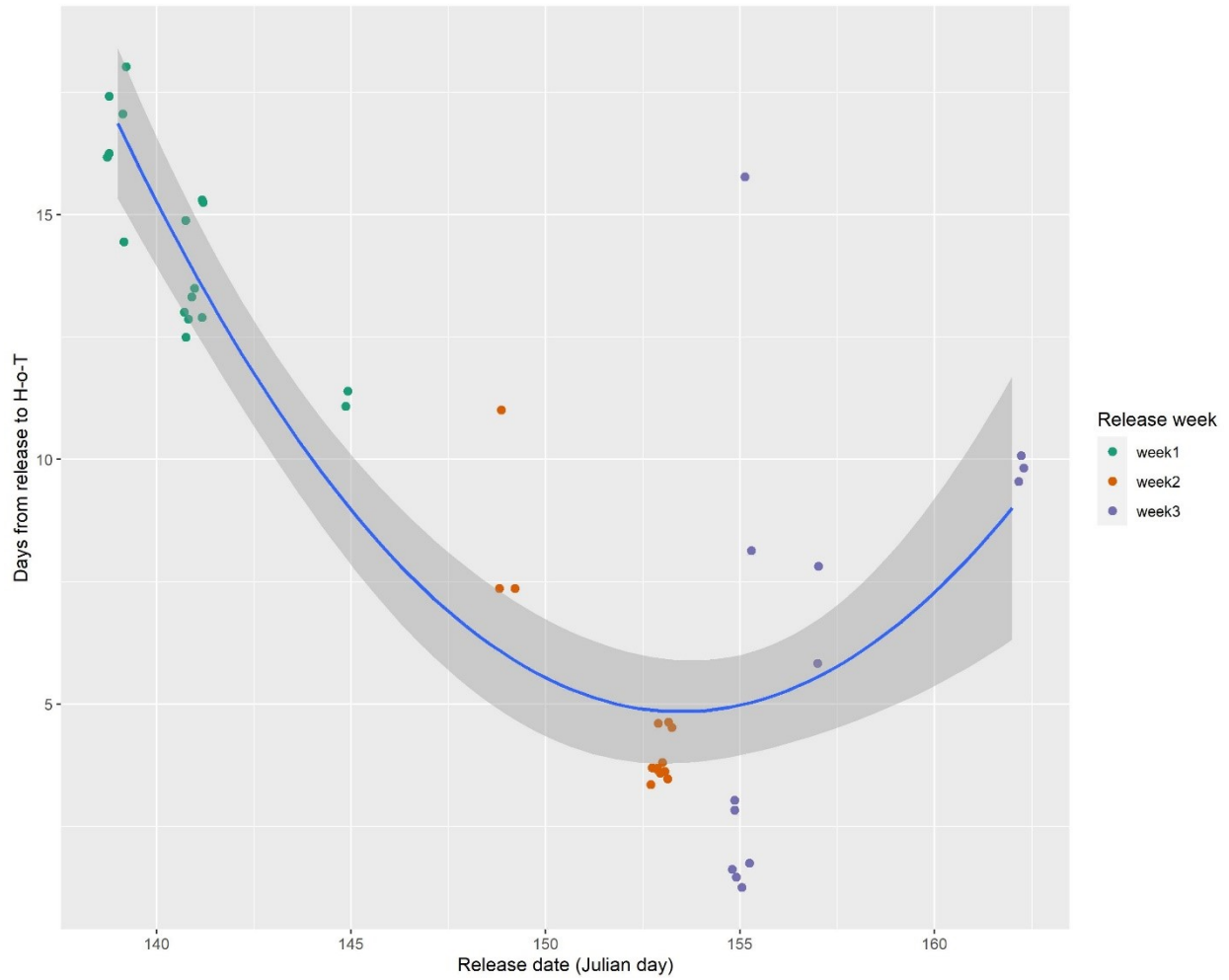
Figure 3.12 The quadratic relationship (y= -17.6x + .057x^2 +1356) between release day of smolts and time from release to the head-of-tide. Each point represents a smolt and is colour coded by release week.
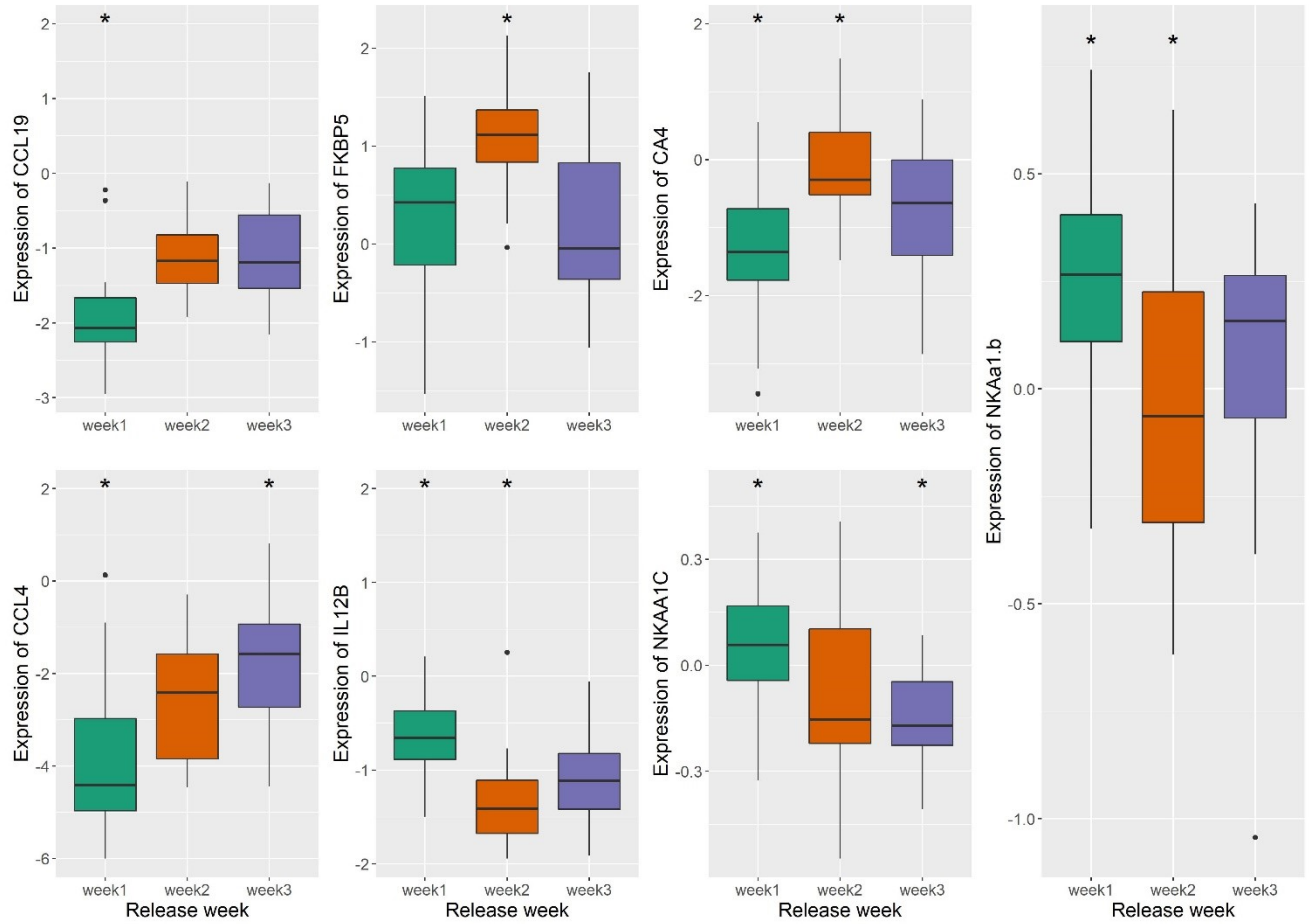
Figure 3.13 Expression of smoltification biomarkers compared between release week groups. Asterisks indicate weeks significantly different from each other (p<0.05). See Table 3.1 for gene names.
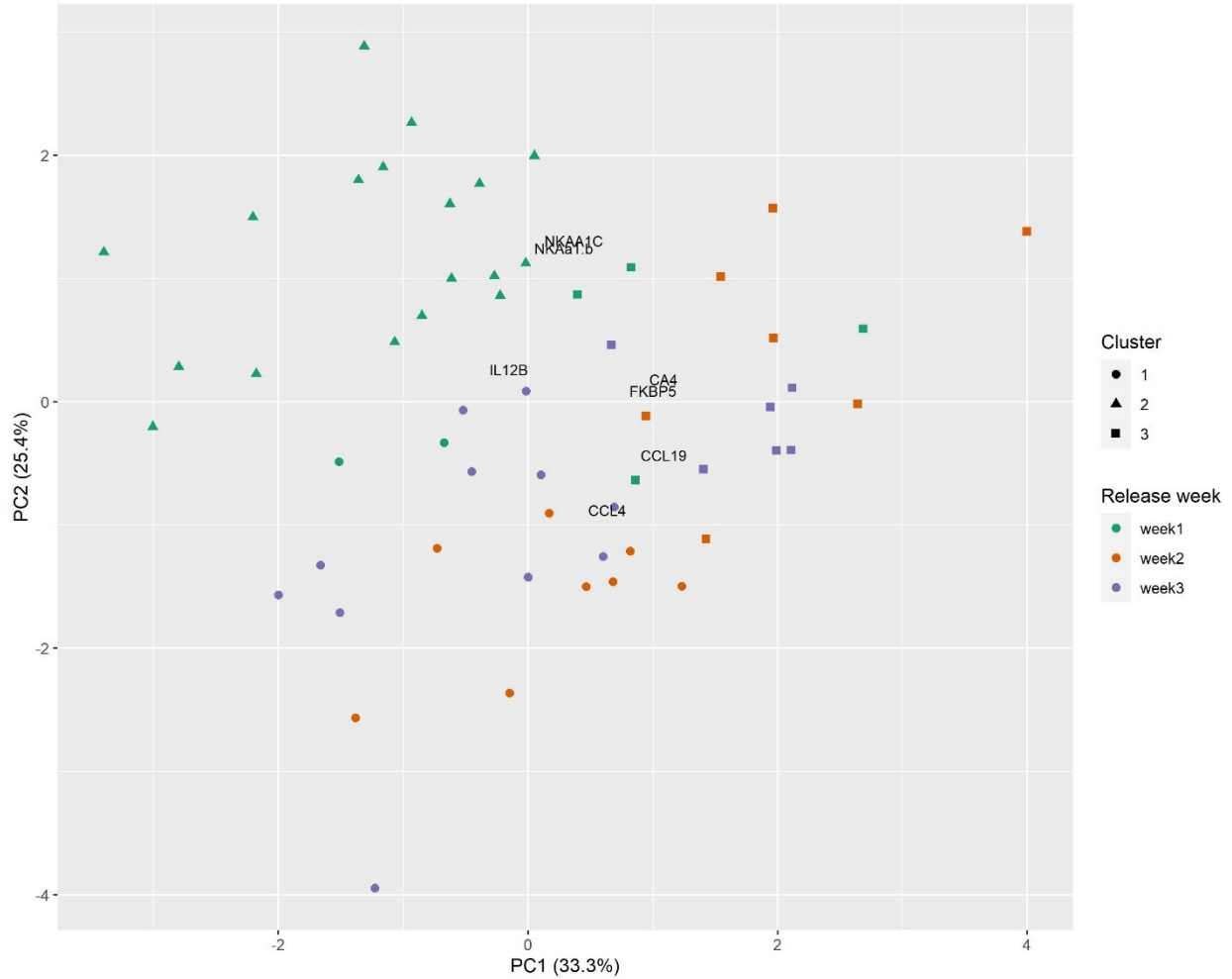
Figure 3.14 Principal components analysis of host gene expression levels of smoltification biomarkers. Principal components (PC) 1-2 shown (percent variation explained). Points represent individual smolt, colour coded by release week and shaped by cluster as determined by k-means clustering. Labels represent loadings (multiplied by a factor of 2 for easy visualization) of explanatory variables (gene expression). See Table 3.1 for gene names.
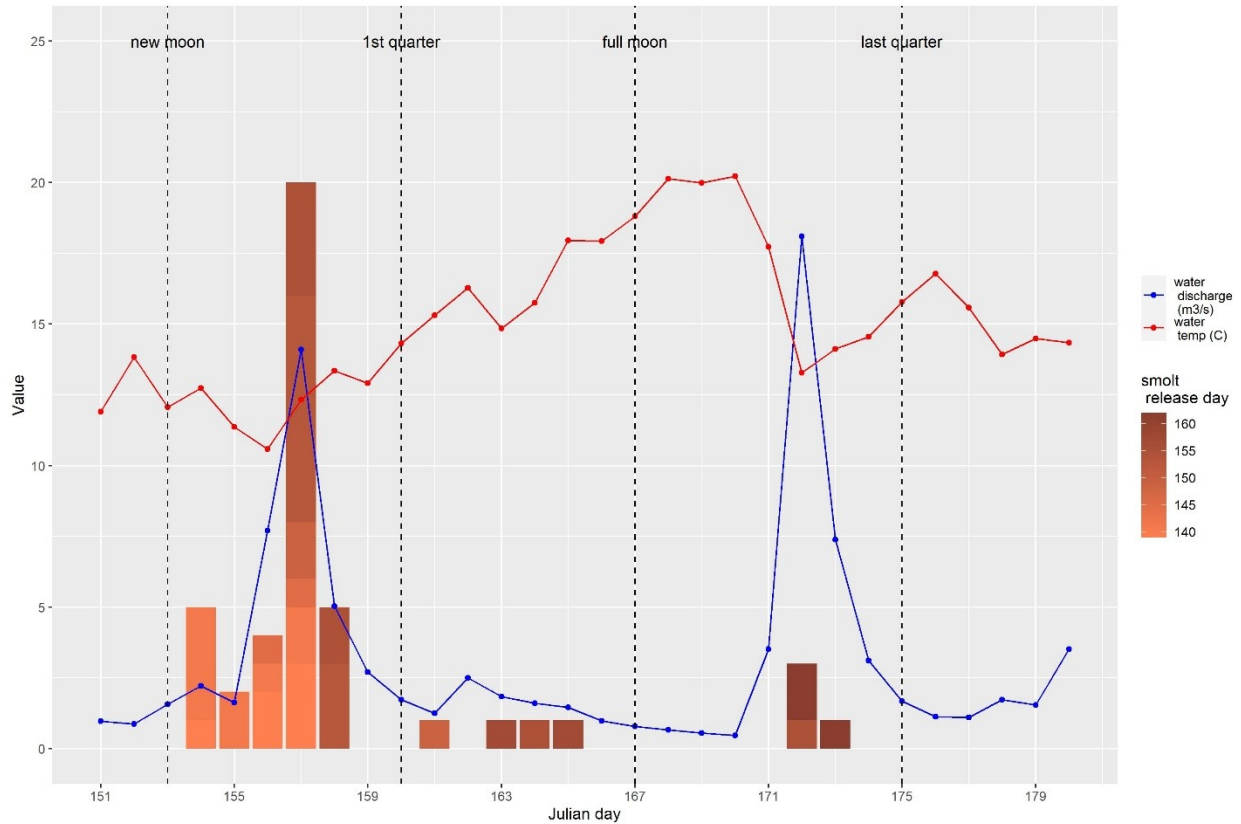
Figure 3.15 Number of smolts entering head of tide by date (bars), smolts colour coded by release date (Julian day). Daily water discharge (m^3/s) measured in St. Andrews River (Data retrieved from https://wateroffice.ec.gc.ca). Average daily water temperature (°C) measured in Little Brook (Data collected by Mi'kmaw Conservation Group). Vertical lines show moon phase.

Chapter 4: Conclusion

The goal of this thesis was to assess the role of predation as a threat to Atlantic Salmon (*Salmo salar*) smolt survival in the Stewiacke River, NS, and to uncover potential behavioural or physiological mechanisms of smolt mortality via predation or other causes. Atlantic Salmon in the Stewiacke River are Endangered and despite fry supplementation through the LGB program, adult return rates are low and there is no sign of population recovery (Jones et al., 2018). Low marine survival is seen as the greatest barrier to population recovery for numerous Atlantic Salmon populations including the iBoF unit (Gibson et al., 2008; Halfyard, 2014). However, because the marine environment poses great challenges both to observing the mechanisms of salmon mortality and in implementing any management/conservation measurements, this thesis focused on another period of high, yet more concentrated, mortality during the salmon life cycle, smolt out-migration. Through a combination of acoustic telemetry, transcriptomic analyses, and other modelling-based approaches, I aimed to determine if piscine predation was the leading cause of smolt mortality and how to predict if a smolt was more or less susceptible to predation.

4.1 Summary

In Chapter 2, I developed a method that combined novel tag technology with machine learning to more accurately classify the migration fate of smolts. I compared three methods of fate classification (pH tag sensor, unsupervised cluster analysis, and supervised random forest) and determined that combining behavioural metrics and pH sensor tag data to train a random forest algorithm was the most accurate method for classifying smolt fate. An algorithm was trained for each of the three study years and then applied to smolts suspected to be predated due to irregular behaviour. Classification by random forest increased estimated predation rates by 9-32% compared to classification by tag sensor only.

In Chapter 3, I used the modeled fate classifications from Chapter 2 to examine trends in smolt survival in the Stewiacke River over three years and found that from 2017 to 2019, smolt survival increased as predation rates decreased. Behaviour, pathogen presence, and gene expression metrics were compared between migration fate groups to determine if there was a behavioural or physiological predictor of fate. I found that a slower migration rate was correlated with higher predation probability. Pathogen presence was too low for in-depth analyses but there was no apparent relationship to fate. Gene expression was not significantly different among fate groups, however, there was a relationship between gene expression and release timing. Further analyses of this relationship revealed that week 1 releases (smolts that migrated earlier) had significantly different expression of smoltification genes than week 2 and 3 releases that was indicative of a pre-smolt physiological signal and lower preparedness for increased salinity. These smolts also took a longer time to enter tidal waters, displaying the relationship between smolt readiness and migration behaviour.

Based on the results presented in this thesis, predation is the leading cause of Atlantic Salmon smolt mortality during out-migration in the Stewiacke River. It appears that predation is not highly selective based on smolt characteristics, behaviour, or physiology but is rather opportunistic due to large predator abundance compared to a small salmon population.

4.2 Management Implications

Based on the results presented here, there are several potential management actions that might be considered to increase smolt survival by reducing predation.

Predator culling is one option, however, past case studies have shown this method to be largely ineffective in the long term and there is a high risk of adverse results (Lennox et al.,

2018). Additionally, culling would not be an option for Striped Bass, the primary predator of salmon smolts in this system, because they are a native species and assessed as Endangered by COSEWIC (COSEWIC, 2012). Sub-lethal predator intervention is also not likely a viable option because Striped Bass are spawning in this system at the time of the smolt run. Predator culling may, however, be considered for invasive predatory species in this system, Smallmouth Bass and Chain Pickerel, but removal of only these species would likely not significantly reduce predation.

Another option is to transport smolts around areas of high predation, however, this can also have negative impacts on smolt physiology and behaviour. Research on the transport of Chinook Salmon smolts through a dam system on the Columbia River, BC has shown that post-release mortality of transported smolts is greater than post-hydropower system mortality of migrating smolts (Muir et al., 2006; Halvorsen et al., 2009). Although mortality during transport is low, post-release mortality is high likely due to a combination of stress reducing predator avoidance behaviour, halted growth during transport, and earlier ocean entry than migrating smolts (Muir et al., 2006; Halvorsen et al., 2009). This approach would also require significant resource investment to capture and transport smolts.

There is no guarantee that reducing smolt predation would translate into greater adult returns, which is ultimately what is needed for population recovery. Depressed population phenomena is one of the leading threats to iBoF salmon where small population numbers are resulting in decreased genetic diversity and fitness, reduced predator avoidance behaviours, and an increase in the negative impact of predation on populations, ultimately contributing to the prevention of population recovery (Amiro et al., 2008). Atlantic Salmon populations in other rivers in the iBoF unit, including those adjacent to the Stewiacke, have also collapsed but do not

have the elevated predation pressure from spawning Striped Bass and invasive predatory species (SARA, 2010). Therefore, additional sources of salmon mortality should be addressed to support smolt survival and population recovery.

Another management approach would be to enhance/restore the ecosystem to facilitate Atlantic Salmon growth and health to provide the best chance of out-migration and marine survival. Despite gene expression profiles indicating that environmental stressors are not related to fate, an ecosystem-based management approach is still valuable and can promote the health of other freshwater and diadromous species in the system (Cowx & Gerdeaux, 2004). Examples include restoring river connectivity, habitat diversity, channel morphology, water flow, and water quality (Cowx & Gerdeaux, 2004). Actions like riparian zone enhancement are becoming increasingly important as water temperatures rise due to climate change (Seavy et al., 2009); water temperatures in the Stewiacke watershed have recently been recorded to be high enough (>20°C) to pose a threat to Atlantic Salmon (N. MacInnis, pers. comm., March 2021). Riparian zones, flood plain habitat, water flow, and river channels have been altered in some areas of the Stewiacke watershed due to human land use activities such as forestry, agriculture, and ATV use (NSFA, 2009; N. MacInnis, pers. comm., March 2021). The Stewiacke watershed is within Colchester County where approximately 54 000 hectares of land are used as farmland (Willcocks-Musselman, 2003). Pesticides and other agricultural runoff can decrease water quality and impair osmoregulation in Atlantic Salmon (Russell et al., 2012). Stream restoration, migration barrier assessments, and water quality monitoring projects are ongoing throughout the Stewiacke watershed by the Mi'kmaw Conservation Group, Nova Scotia Salmon Association, and other organizations to address these issues.

This study, in conjunction with others focused on adult return rates and spawning success, can also inform current management practices. The LGB Program has been stocking the Stewiacke River with salmon fry since 2003 (DFO, 2018b) and without this program salmon would likely be extirpated from this river (Gibson et al., 2008; Jones et al., 2018). However, with 43-96% mortality of smolts each year and very low adult returns (Gibson et al., 2008; Jones et al., 2018), it is unlikely that the population will recover to a point where it is self-sustaining, thus consideration may be given to focusing breeding programs elsewhere. Such a decision would need to carefully consider the ecological and cultural impacts of the loss of salmon in this system.

4.3 Limitations

In this thesis, I was able to successfully apply supervised and unsupervised machine learning methods to this data set and assess the role of behavioural metrics in relationship to fate, however, the small sample size (50-56 smolts/year) limited conclusions in some cases, especially in relation to gene expression and pathogen analyses. This was exacerbated by major changes in sampling and field methods between years which prevented the combination of yearly data sets in some cases and introduced confounding factors into examining mortality trends between years. While sample size is often limited by available funds and other logistics, stronger relationships between variables or physiological mechanisms of mortality may have been uncovered with a greater number of samples. In an effort to achieve this, 2019 field methods and study design were repeated in 2021, and these two data sets will be combined in future analyses.

Although confidence in classifying smolt fate was increased through the use of predation tags and random forest algorithms, true smolt fate is still not known with certainty. The mortality class is associated with the most uncertainty, as smolts in this class could have died via predation

(primarily avian), stress, disease, or other causes. Smolts may have also been placed in this class due to tag loss or malfunction. Making such assumptions about tag function and the fate of a tagged animal is the nature of most telemetry studies, unless recapturing the individual or tag is possible. Additionally, while the migration fates modeled by random forest algorithms were shown to be the most accurate estimates, average in-sample prediction accuracy was still less than 90% and there is the potential that some smolts were misclassified.

4.4 Broader significance

Chapter 2 resulted in the development of a methods framework for analyzing telemetry data that other researchers can apply to their own data sets. While this framework may not be applicable in all cases (many potential predators, similar behaviour between predator and prey), it is a valuable tool for movement ecologists to identify and reduce predation bias, ultimately leading to more accurate survival estimates to inform population management. The use of modelling approaches in addition to predation tags is recommended because even with these tags, predation can be underestimated as was found here. This chapter highlights how advances in technology and modelling-based approaches can provide insight into predator-prey relationships that would otherwise be near impossible to observe.

This thesis focused on a single population of Atlantic Salmon; however, these methods can be applied to other systems and other diadromous fish species. Here, the Salmon Fit-Chip was successfully applied to Atlantic Salmon; whereas previously these biomarker assays were primarily used for Pacific salmon (Miller et al., 2011; Jeffries et al., 2014; Healey et al., 2018; Chapman et al., 2020). The smoltification panel of biomarkers provided the greatest insight into smolt physiology in this study and successfully separated pre-smolts from full smolts despite this

panel being developed through studies on Coho, Sockeye, and Chinook salmon (Houde et al., 2019).

While few direct links were made between smolt behaviour or physiology and migration fate, this thesis highlights the value of combining multiple methods and technologies to examine the interactions between individual fish health, behaviour, and environmental conditions. Through the integration of novel acoustic telemetry and transcriptomic technologies, this thesis was able to provide a more comprehensive analysis of the smoltification process, smolt out-migration, and predator-prey interactions.

References

Altizer S, Bartel R, Han BA. 2011. Animal migration and infectious disease risk. Science 331(6015): 296-302.

Amiro PG, Brazner JC, Giorno JL. 2008. Assessment of the recovery potential for the Atlantic salmon designatable unit inner Bay of Fundy: Threats. DFO Can Sci Advis Sec Rep 2008/059.

Andrews SN, Hirtle SV, Linnansaari T, Curry RA. 2019. Consumption of Atlantic salmon smolt by striped bass: A review of the predator-prey encounter literature and implications for the design of effective sampling strategies. Fishes 4(4), 50.

Atlantic Salmon Federation. 2011. Economic Value of Wild Atlantic Salmon. Retrieved from https://www.asf.ca/assets/files/gardner-pinfold-value-wild-salmon.pdf

Babin A. 2019. Atlantic salmon (*Salmo salar*) migrations in a large hydropower reservoir and the regulated Saint John River. PhD Thesis. Department of Biology, University of New Brunswick, Fredericton, NB.

Beland KF, Kocik JF, vandeSande J, Sheehan TF. 2001. Striped bass predation upon Atlantic salmon smolts in Maine. Northeast Nat 8(3): 267-274.

Bendall B, Moore A. 2008 Temperature-sensing telemetry – possibilities for assessing the feeding ecology of marine mammals and their potential impacts on returning salmonid populations. Fisheries Manag Ecol 15: 339-345.

Bender LC. 2018. Understanding predation. Circular 688, College of Agricultural, Consumer and Environmental Sciences, New Mexico State University.

Binder TR, Dini A. 2012. glatos: An R package for the Great Lakes Acoustic Telemetry Observation System. R package version 0.3.0. https://rdrr.io/github/jsta/glatos/man/glatos.html

Birnie-Gauvin K, Larsen MH, Peiman KS, Midwood JD, Wilson ADM, Cooke SJ, Aarestrup K. 2021. No evidence for long-term carryover effects in a wild salmonid fish. Physiol Biochem Zool 94(5): 319-329.

Bjornsson BT, Stefansson SO, McCormick SD. 2010. Environmental endocrinology of salmon smoltification. Gen Comp Endocr 170: 290-298.

Blackwell BF, Juanes F. 1998. Predation on Atlantic salmon smolts by striped bass after dam passage. N Am J Fish Manage 18:936-939.

Bland LM, Collen B, Orme CDL, Bielby J. 2014. Predicting the conservation status of data-deficient species. Conserv Biol 29(1): 250-259.

Bradford RG, Halfyard EA, Hayman T, LeBlanc P. 2015. Overview of the 2013 Bay of Fundy striped bass biology and general status. DFO Can Sci Advis Sec Res Doc 2015/024. iv + 36 p.

Brownscombe JW, Griffin LP, Gagne TO, Haak CR, Cooke SJ, Finn JT, Danylchuk AJ. 2019. Environmental drivers of habitat use by a marine fish on a heterogenous and dynamic reef flat. Mar Biol https://doi.org/10.1007/s00227-018-3464-2

Brownscombe JW, Griffin LP, Morley D, Acosta A, Hunt J, Lowerre-Barbieri SK, Adams AJ, Danylchuk AJ, Cooke SJ. 2020. Application of machine learning algorithms to identify cryptic reproductive habitats using diverse information sources. Oecologia 194: 283-298.

Buchanan RA, Skalski JR, Brandes PL, Fuller A. 2013. Route use and survival of juvenile chinook salmon through the San Joaquin River Delta. N Am J Fish Manage 33(1): 216-229.

Cairns DK, Reddin DG. 2000. The potential impact of seal and seabird predation on North American Atlantic salmon. DFO Can Sci Advis Sec Res Doc 2000/12. i + 34 p.

Chapman BB, Bronmark C, Nilsson JA, Hansson LA. 2011. The ecology and evolution of partial migration. Oikos 120: 1764-1775.

Chapman JM, Teffer AK, Bass AL, Hinch SG, Patterson DA, Miller KM, Cooke SJ. 2020. Handling, infectious agents and physiological condition influence survival and post-release behaviour in migratory adult coho salmon after experimental displacement. Conserv Physiol 8(1): coaa033.

Chen C, Liaw A, Breiman L. 2004. Using random forest to learn imbalanced data.

Clark JS, Carpenter SR, Barber M, Collins S, Dobson A, Foley JA, Lodge DM, Pascual M, Pielke R, Pizer W, et al. 2001. Ecological forecasts: An emerging Imperative. Science 293(5530): 657-660.

Clark TD, Furey NB, Rechisky EL, Gale MK, Jeffries KM, Porter AD, Casselman MT, Lotto AG, Patterson DA, Cooke SJ, Farrell AP, Welch DW, Hinch SG. 2016. Tracking wild sockeye salmon smolts to the ocean reveals distinct regions of nocturnal movement and high mortality. Ecol Appl 26(4): 959-978.

Cooke SJ, Murchie KJ, McConnachie S, Goldberg T. 2011. Standardized surgical procedure for the implantation of electronic tags in key Great Lakes fishes. Technical Report. Great Lakes Fishery Commission, Ann Arbor, MI.

Cordier T, Esling P, Lejzerowicz F, Visco J, Ouadahi A, Martins C, Cedhagen T, Pawlowski J. 2017. Predicting the ecological quality status of marine environments from eDNA metabarcoding Data using supervised machine learning. Environ Sci Technol 51(16): 9118-9126.

Coreau A, Pinay G, Thompson JD, Cheptou PO, Mermet, L. 2009. The rise of research on futures in ecology: rebalancing scenarios and predictions. Ecol Lett 12: 1277-1286.

COSEWIC. 2010. Assessment and Status Report on the Atlantic salmon *Salmo salar.* Committee on the Status of Endangered Wildlife in Canada, Ottawa, ON. xlvii + 136 p.

COSEWIC. 2012. COSEWIC assessment and status report on the striped bass Morone saxatilis in Canada. Committee on the Status of Endangered Wildlife in Canada. Ottawa. iv + 79 pp.

Cowx IG, Gerdeaux D. 2004. The effects of fisheries management practices on freshwater ecosystems. Fisheries Manag Ecol 11: 145-151.

Creel S. 2011. Toward a predictive theory of risk effects: hypotheses for prey attributes and compensatory mortality. Ecology 92(12): 2190-2195.

Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ. 2007. Random forests for classification in ecology. Ecology 88(11): 2783-2792.

Dadswell MJ, Spares AD, Porter E, Porter D. 2020. Diversity, abundance, and size structure of fishes and invertebrates captured by an intertidal fishing weir at Bramber, Minas Basin, Nova Scotia. Proceedings of the Nova Scotian Institute of Science 50(2): 283-318.

Dadswell MJ, Spares A, Reader J, McLean M, McDermott T, Samways K, Lilly J. 2021. The decline and impending collapse of the Atlantic salmon (*Salmo salar*) population in the North Atlantic Ocean: A review of possible causes. Rev Fish Sci Aquac DOI: 10.1080/23308249.2021.1937044.

Daniels J, Chaput G, Carr J. 2018. Estimating consumption rate of Atlantic salmon smolts (Salmo salar) by striped bass (Morone Saxatilis) in the Miramichi River estuary using acoustic telemetry. Can J Fish Aquat Sci 75: 1811-1822.

Daniels J, Sutton S, Webber D, Carr J. 2019. Extent of predation bias present in migration survival and timing of Atlantic salmon smolt (*Salmo salar*) as suggested by a novel acoustic tag. Anim Biotelemetry DOI: 10.1186/s40317-019-0178-2.

DFO. 2018a. Canada's Wild Atlantic Salmon Conservation Policy. Retrieved from https://www.dfo-mpo.gc.ca/reports-rapports/regs/wildsalmon-atl-saumonsauvage-eng.htm

DFO. 2018b. Review of the science associated with the inner Bay of Fundy Atlantic Salmon Live Gene Bank and supplementation programs. DFO Can Sci Advis Sec Rep 2018/041.

DFO. 2019a. Atlantic Salmon Marine Threats Research. Retrieved from http://www.bio.gc.ca/science/research-recherche/fisheries-pecheries/managed-gere/smtr-rfms-en.php

DFO. 2019b. Atlantic Salmon (Inner Bay of Fundy Population). Retrieved from http://www.dfo-mpo.gc.ca/species-especes/profiles-profils/salmon-atl-saumon-eng.html

Dieperink C, Bak BD, Pedersen LF, Pedersen MI, Pedersen S. 2002. Predation on Atlantic salmon and sea trout during their first day as postsmolts. J Fish Biol 61: 848-852. Dingle H. 2014. Migration: the biology of life on the move, second edition. Oxford, UK: Oxford University Press. 326 p.

Egevang C, Stenhouse IJ, Phillips RA, Petersen A, Fox JW, Silk JRD. 2010. Tracking of Arctic terns *Sterna paradisaea* reveals longest animal migration. PNAS 107(5): 2078-2081.

Evans AF, Hostetter NJ, Roby DD, Collis K, Lyons DE, Sandford BP, Ledgerwood RD, Sebring S. 2012. Systemwide evaluation of avian predation on juvenile salmonids from the Columbia River based on recoveries of passive integrated transponder tags. T Am Fish Soc 141(4): 975-989.

Flavio H, Kennedy R, Ensing D, Jepsen N, Aarestrup K. 2019. Marine moratlity in the river? Atlantic salmon smolts under high predation pressure in the last kilometers of a river monitored for stock assessment. Fish Manag Ecol DOI: 10.1111/fme.12405.

Flavio H, Caballero P, Jepsen N, Aarestrup K. 2020. Atlantic salmon living on the edge: smolt behaviour and survival during seaward migration in River Minho. Ecol Freshw Fish doi: 10.1111/eff.12564.

Fox J, Weisberg S. 2019. An R Companion to Applied Regression, 3rd Edition. Thousand Oaks, CA http://tinyurl.com/carbook

Furey NB, Hinch SG, Lotto AG, Beauchamp DA. 2015. Extensive feeding on sockeye salmon *Oncorhynchus nerka* smolts by bull trout *Salvelinus confluentus* during initial outmigration into a small, unregulated and inland British Columbia River. J Fish Biol 86: 392-401.

Furey, NB. 2016. Migration ecology of juvenile Pacific salmon smolts: the role of fish condition and behaviour across landscapes. PhD Thesis. Department of Forest and Conservation Sciences, University of British Columbia, Vancouver, BC.

Furey NB, Hinch SG, Bass AL, Middleton CT, Minke-Martin V, Lotto AG. 2016. Predator swamping reduces predation risk during nocturnal migration of juvenile salmon in high-mortality landscape. J Anim Ecol 85: 948-959.

Furey NB, Armstrong JB, Beauchamp DA, Hinch SG. 2018. Migratory coupling between predators and prey. Nat Ecol Evol 2: 1846-1853.

Furey NB, Martins EG, Hinch SG. 2020. Migratory salmon smolts exhibit consistent interannual depensatory predator swamping: effects on telemetry-based survival estimates. Ecol Freshw Fish 30: 18-30.

Gibson AJF, Bowlby HD, Bryan JR, Amiro PG. 2008. Population viability analysis of inner Bay of Fundy Atlantic salmon with and without Live Gene Banking Can Sci Advis Sec Res Doc 2008/057.

Gibson AJF, Halfyard EA, Bradford RG, Stokesbury MJW, Redden AM. 2015. Effects of predation on telemetry-based survival estimates: insights from a study on endangered Atlantic salmon smolts. Can J Fish Aquat Sci 72: 728-741.

Gibson AJF, Bowlby HD, Keyser FM. 2017. A framework for the assessment of the status of river herring populations and fisheries in DFO's Maritimes Region. DFO Can Sci Advis Sec Res Doc 2016/105. vi + 69 p.

Grout DE. 2006. Interactions between striped bass (*Morone saxatilis*) rebuilding programmes and the conservation of Atlantic salmon (*Salmo salar*) and other anadromous fish species in the USA. ICES J Mar Sci 63: 1346-1352.

Gurarie E, Andrews RD, Laidre KL. 2009. A novel method for identifying behavioural changes in animal movement data. Ecol Lett 12(5): 395-408.

Halfyard EA, Gibson AJF, Ruzzante DE, Stokesbury MJW, Whoriskey FG. 2012. Estuarine survival and migratory behaviour of Atlantic salmon *Salmo salar* smolts. J Fish Biol 81: 1626-1645.

Halfyard EA, Gibson AJF, Stokesbury MJW, Ruzzante DE, Whoriskey FG. 2013. Correlates of estuarine survival of Atlantic salmon postsmolts from the Southern Upland, Nova Scotia, Canada. Can J Fish Aquat Sci 70:452-460.

Halfyard EA. 2014. The estuarine and early marine survival of Atlantic salmon: estimation, correlates and ecological significance. PhD Thesis. Department of Biology, Dalhousie University, Halifax, NS.

Halfyard EA, Webber D, Del Papa J, Leadley T, Kessel ST, Colborne SF, Fisk AT. 2017. Evaluation of an acoustic telemetry transmitter designed to identify predation events. Methods Ecol Evol 8: 1063-1071.

Halvorsen MB, Wysocki LE, Stehr CM, Baldwin DH, Chicoine DR, Scholz NL, Popper AN. 2009. Barging effects on sensory systems of chinook salmon smolts. T Am Fish Soc 138: 777-789.

Handeland SO, Jarvi T, Ferno A, Stefansson SO. 1996. Osmotic stress, antipredator behaviour, and mortality of Atlantic salmon (Salmo salar) smolts. Can J Fish Aquat Sci 53: 2673-2680.

Hanssen EM. 2020. Novel telemetry predation sensors and mechanistic models reveal the tribulations of Atlantic salmon (*Salmo salar*) smolts migrating through lakes. MSc Thesis, Department of Biological Sciences, University of Bergen, Norway.

Healey SJ, Hinch SG, Bass AL, Furey NB, Welch DW, Rechisky EL, Eliason EJ, Lotto AG, Miller KM. 2018. Transcriptome profiles relate to migration fate in hatchery steelhead (*Oncorhynchus mykiss*) smolts. Can J Fish Aquat Sci 75: 2053-2068.

Hendry AP, Bohlin T, Jonsson B, Berg OK. 2004. To sea or not to sea? Anadromy versus non-anadromy in salmonids. In: Hendry AP, Stearns SC, editors. Evolution illuminated: salmon and their relatives. Oxford, UK: Oxford University Press. pp 92-126.

Hinch SG, Cooke SJ, Farrell AP, Miller KM, LaPointe M, Patterson DA. 2012. Dead fish swimming: a review of research on the early migration and high premature mortality in adult Fraser River sockeye salmon *Oncorhynchus nerka*. J Fish Biol 81: 576-599.

Hostetter NJ. 2009. Susceptibility of juvenile salmonids to avian predation: are Caspian terns and double-crested cormorants only taking the sick and injured? MSc Dissertation. Department of Fisheries and Wildlife, Oregon State University, Corvallis, OR.

Houde ALS, Gunther OP, Strohm J, Ming TJ, Li S, Kaukinen KH, Patterson DA, Farrell AP, Hinch SG, Millr KM. 2019. Discovery and validation of candidate smoltification gene expression biomarkers across multiple species and ecotypes of Pacific salmonids. Conserv Physiol 7(1): coz051.

Ibbotson AT, Beaumont WRC, Pinder AC. 2011. A size-dependent migration strategy in Atlantic salmon smolts: small smolts favour nocturnal migration. Environ Biol Fish 92: 151-157.

Indira V, Vasanthakumari R, Sugumaran V. 2010. Minimum sample size determination of vibration signals in machine learning approach to fault diagnosis using power analysis. Expert Systems with Applications 37(12): 8650-8658.

Jain AK. 2010. Data clustering: 50 years beyond k-means. Pattern Recogn Lett doi:10.1016/j.patrec.2009.09.011

Jeffries KM, Hinch SG, Gale MK, Clark TD, Lotto AG, Casselman MT, Li S, Rechisky EL, Porter AD, Welch D, Miller KM. 2014. Immune response genes and pathogen presence predict migration survival in wild salmon smolts. Mol Ecol 23: 5803-5815.

Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24: 1403-1405. doi: 10.1093/bioinformatics/btn129

Jones RA, Ratelle SM, Tuziak SM, Harvie C, Lenentine B, and O'Reilly PT. 2020. Review of the Inner Bay of Fundy Atlantic Salmon (Salmo salar) Monitoring Activities Associated with the Live Gene Bank. DFO Can. Sci. Advis. Sec. Res. Doc. 2018/043. vii + 128 p.

Jonker RM, Eichhorn G, van Langevelde F, Bauer S. 2010. Predation danger can explain changes in timing of migration: the case of the barnacle goose. PLoS One 5(6): e11369. DOI: 10.1371/journal.pone.0011369

Jonsson B, Jonsson N. 2003. Migratory Atlantic salmon as vectors for the transfer of energy and nutrients between freshwater and marine environments. Freshwater Biol 48: 21-27.

Jutila E, Jokikokko E, Julkunen M. 2005. The smolt run and postsmolt survival of Atlantic salmon, Salmo salar L., in relation to early summer water temperatures in the northern Baltic Sea. Ecol Freshw Fish 14: 69-78.

Klinard NV, Matley JK, Fish AT, Johnson TB. 2019. Long-term retention of acoustic telemetry transmitters in temperate predators revealed by predation tags implanted in wild prey fish. J Fish Biol DOI: 10.1111/jfb.14156.

Klinard NV, Matley JK. 2020. Living until proven dead: addressing mortality in acoustic telemetry research. Rev Fish Biol Fisheries 30: 485-499.

Klinard NV, Matley JK, Ivanova SV, Larocque SM, Fisk At, Johnson TB. 2021. Application of machine learning to identify predators of stocked fish in Lake Ontario: using acoustic telemetry predation tags to inform management. J Fish Biol DOI: 10.1111/jfb.14574

LaCroix GL. 2008. Influence of origin on migration and survival of Atlantic salmon (*Salmo salar*) in the Bay of Fundy, Canada. Can J Fish Aquat Sci 65: 2063-2079.

LeBlanc J. 2010. Geographic distribution of smallmouth bass, *Micropterus dolomieu*, in Nova Scotia: history of early introductions and factors affecting current range. DFO Can Sci Advis Sec Res Doc 2010/028. iv + 25 p.

Lennox RJ, Gallagher AJ, Ritchie EG, Cooke SJ. 2018. Evaluating the efficacy of predator removal in a conflict-prone world. Biol Conserv 224: 277-289.

Lennox RJ, Eldoy SH, Vollset KW, Miller KM, Li S, Kaukinen KH, Isaksen TE, Davidsen JG. 2020. Abundance, diversity, and effects on marine habitat use of infectious pathogens among sea trout (*Salmo trutta*) in Norwegian fjord systems. J Fish Dis

Liaw A, Wiener M. 2002. randomForest: Breiman and Cutler's Random Forests for Classification and Regression**.** R package version 4.6-14. https://cran.r-project.org/web/packages/randomForest/index.html

Lima SL, Dill LM. 1990. Behavioural decisions made under the risk of predation: a review and prospectus. Can J Zool 68: 619-640.

Martec Ltd. 2007. Physical description of the Shubenacadie River. Martec Technical Report #TR-07-12. Retrieved from https://novascotia.ca/nse/ea/Alton/AppendixA.pdf

Maxwell AE, Warner TA, Fang F. 2018. Implementation of machine-learning classification in remote sensing an applied review. International Journal of remote sensing 39(9): 2784-2817.

McCormick SD, Hansen LP, Quinn TP, Saunders RL. 1998. Movement, migration, and smolting of Atlantic salmon (Salmo salar). Can J Fish Aquat Sci 55: 77-92.

Midwood JD, Larsen MH, Boel M, Jepsen N, Aarestrup K, Cooke SJ. 2014. Does cortisol manipulation influence outmigration behaviour, survival and growth of sea trout? A field test of carryover effects in wild fish. Mar Ecol Prog Ser 496: 135-144.

Midwood JD, Larsen MH, Boel M, Aarestrup K, Cooke SJ. 2015. An experimental field evaluation of winter carryover effects in semi-anadromous brown trout (Salmo trutta). J Exp Zool 323: 645-654.

Miller KM, Li S, Kaukinen KH, Ginter N, Hammill E, Curtis JMR, Patterson DA, Sierocinski T, Donnison L, Pavlidis P, Hinch SG, Hruska KA, Cooke SJ, English KK, Farrell AP. 2011. Genomic signatures predict migration and spawning failure in wild Canadian salmon. Science 331(6014): 214-217.

Miller KM, Teffer A, Tucker S, Li S, Schulze AD, Trudel M, Juanes F, Tabata A, Kaukinen KH, Ginther NG, Ming TJ, Cooke SJ, Hipfner M, Patterson DA, Hinch SG. 2014. Infectious disease, shifting climates, and opportunistic predators: cumulative factors potentially impacting wild salmon declines. Evol Appl 7: 812-855.

Milner-Gulland EJ, Fryxell JM, Sinclair ARE. 2011. Animal migration: a synthesis. Oxford, UK: Oxford University Press. 216 p.

Mitchell SC, LeBlanc JE, Heggelin AJ. 2010. Impact of introduced chain pickerel (*Esox niger*) on lake fish communities in Nova Scotia, Canada. Report. Waterways Environmental Services, Antigonish, NS.

Moghaddam DD, Rahmati O, Panahi, M, Tiefenbacher J, Darabi H, Haghizadeh A, Haghighi AT, Nalivan OA, Bui DT. 2020. The effect of sample size on different machine learning models for groundwater potential mapping mountain bedrock aquifers. Catena 187: 104421.

Moxam EJ, Cowley PD, Bennett RH, von Brandis RG. 2019. Movement and predation: a catch-and-release study on the acoustic tracking of bonefish in the Indian Ocean. Environ Biol Fish 102: 365-381.

Muir WD, Marsh DM, Sandford BP, Smith SG, Williams JG. 2006. Post-hydropower system delayed mortality of transported Snake River stream-type chinook salmon: unraveling the mystery. T Am Fish Soc 135: 1523-1534.

Nova Scotia Federation of Agriculture. 2009. Riparian Health Assessment Project. Retrieved from https://nsfa-fane.ca/projects/other-projects/riparian-health-assessment-project/

Olden JD, Lawler JJ, Poff NL. 2008. Machine learning methods without tears: A primer for ecologists. Q Rev Biol 83(2): 171-193.

Ormerod SJ. 2002. Applied issues with predators and predation: editor's introduction. J Appl Ecol 39: 181-188.

Otero J, L'Abee-Lund JH, Castro-Santos T, Leonardsson K, Storvik GO, Jonsson B, Dempson B, Russell IC, Jensen AJ, Bagliniere JL, et al. 2014. Basin-scale phenology and effects of climate change variability on global timing of initial seaward migration of Atlantic salmon (Salmo salar). Glob Change Biol 20: 61-75.

Palacios G, Lovoll M, Tengs T, Hornig M, Hutchison S, Hui J, Kongtorp RT, Savji N, Bussetti AV, Solovyov A, et al. 2010. Heart and skeletal muscle inflammation of farmed salmon is associate with infection with a novel reovirus. PLOS One 5(7): e11487.

Parrish DL, Behnke RJ, Gephard SR, McCormick SD, Reeves GH. 1998. Why aren't there more Atlantic salmon (*Salmo salar*)? Can J Fish Aquat Sci. 55: 281-287.

Payton Q, Evans AF, Hostetter NJ, Roby DD, Cramer B, Collis K. 2020. Measuring the additive effects of predation on prey survival across spatial scales.

Perry RW, Skalski JR, Brandes PL, Sandstrom PT, Klimley P, Ammann A, MacFarlane B. 2010. Estimating survival and migration route probabilities of juvenile chinook salmon in the Sacramento-San Joaquin River Delta. N Am J Fish Manage 30: 142-156.

Robledo D, Gutierrez AP, Barria A, Yanez JM, Houston RD. 2018. Gene expression response to sea lice in Atlantic salmon skin: RNA sequencing comparison between resistant and susceptible animals. Front Genet doi: 10.3389/fgene.2018.00287

Romine JG, Perry RW, Johnston SV, Fitzer CW, Pagliughi SW, Blake AR. 2014. Identifying when tagged fishes have been consumed by piscivorous predators: application of multivariate mixture models to movement parameters of telemetered fishes. Anim Biotelemetry DOI: 10.1186/2050-3385-2-3.

Russell IC, Aprahamian MW, Barry J, Davidson IC, Fiske P, Ibbotson AT, Kennedy RJ, Maclean JC, Moore A, Otero J, et al. 2012. The influence of the freshwater environment and the biological characteristics of Atlantic salmon smolts on their subsequent marine survival. ICES J Mar Sci 69(9): 1563-1573.

Sabal MC, Boyce MS, Charpentier CL, Furey NB, Luhring TM, Martin HW, Melnychuk MC, Srygley RB, Wagner CM, Wirsing AJ, et al. 2021. Predation landscapes influence migratory prey ecology and evolution. Trends Ecol Evol doi: 10.1016/j.tree.2021.04.010.

SARA. 2010. Recovery strategy for the Atlantic salmon (*Salmo salar*) inner Bay of Fundy populations. Retrieved from https://wildlife-species.canada.ca/species-risk-registry/document/doc1917f/ind_e.cfm#_2.5

SARA 2021. Species Search: Atlantic salmon. Retrieved from https://species-registry.canada.ca/index-en.html#/species?sortBy=commonNameSort&sortDirection=asc&pageSize=20&keywords=atlantic%20salmon

Saunders R, Hachey MA, Fay CW. 2006. Maine's diadromous fish community: past, present, and implications for Atlantic salmon recovery. Fisheries 31(11): 537-547.

Schultz AA, Kumagai KK, Bridges BB. 2015. Methods to evaluate gut evacuation rates and predation using acoustic telemetry in the Tracy Fish Collection Facility primary channel. Anim Biotelemetry https://doi.org/10.1186/s40317-015-0034-y

Seavy NE, Gardali T, Golet GH, Griggs FT, Howell CA, Kelsey R, Small SL, Viers JH, Weigand JF. 2009. Why climate change makes riparian restoration more important than ever: recommendations for practice and research. Ecol Rest 27(3): 330-338.

Seear PJ, Carmichael SN, Talbot R, Taggart JB, Bron JE, Sweeney GE. 2010. Differential gene expression during smoltification of Atlantic salmon (Salmo salar L.): a first large-scale microarray study. Mar Biotechnol 12: 126-140.

Sergio F, Tavecchia G, Tanferna A, Blas J, Blanco G, Hiraldo F. 2019. When and where mortality occurs throughout the annual cycle changes with age in a migratory bird: individual vs population implications. Nature doi: 10.1038/s41598-019-54026-z.

Skov C, Baktoft H, Brodersen J, Bronmark C, Chapman BB, Hansson LA, Nilsson PA. 2011. Sizing up your enemy: individual predation vulnerability predicts migratory probability. Proc Biol Sci 278(1710): 1414-1418.

Stich DS, Zydlewski GB, Kocik JF, Zydlewski JD. 2015. Linking behaviour, physiology, and survival of Atlantic salmon smolts during estuary migration. Mar Coast Fish 7(1): 68-86.

Strand JET, Davidsen JG, Jorgensen EH, Rikardsen AH. 2011. Seaward migrating Atlantic salmon smolts with low levels of gill Na+,K+-ATPase activity; is sea entry delayed? Envrion Biol Fish 90: 317-321.

Svenning MA, Borgstrom R, Dehli TO, Moen G, Barrett RT, Pedersen T, Vader W. 2005. The impact of marine fish predation on Atlantic salmon smolts (Salmo salar) in the Tana estuary, North Norway, in the presence of an alternative prey, lesser sandeel (Ammodytes marinus). Fish Res 76(3): 466-474.

Tabak MA, Norouzzadeh MS, Wolfson DW, Sweeny SJ, Vercauteren KC, Snow NP, Halseth JM, Di Salvo PA, Lewis JS, White MD, et al. 2018. Machine learning to classify animal species in camera trap images: applications in ecology. Methods Ecol Evol 10(4): 585-590.

Thessen AE. 2016. Adoption of machine learning techniques in ecology and earth science. One Ecosystem doi: 10.3897/oneeco.1.e8621

Thorstad EB, Whoriskey FG, Rickardsen AH, Aarestrup K. 2011. Aquatic nomads: the life and migrations of the Atlantic salmon. In: Aas O, Einum S, Klemetsen A, Skurdal J, editors. Atlantic Salmon Ecology. Chichester, West Sussex, UK: Wiley-Blackwell. pp 1-33.

Thorstad EB, Whoriskey FG, Uglem I, Moore A, Rikardsen AH, Finstad B. 2012. A critical life stage of the Atlantic salmon *Salmo salar:* behaviour and survival during the smolt and initial post-smolt migration. J Fish Biol 81: 500-542.

Thorstad EB, Todd CD, Uglem I, Bjorn PA, Gargan PG, Vollset KW, Halttunen E, Kalas S, Berg M, Finstad B. 2015. Effects of salmon lice *Lepeophtheirus salmonis* on wild sea trout *Salmo trutta* – a literature review. Aquacult Environ Interact 7: 91-113.

Tucker S, Hipfner JM, Trudel M. 2016. Size- and condition-dependent predation: a seabird disproportionately targets substandard individual juvenile salmon. Ecology 97(2): 461-471.

Venables WN, Ripley BD. 2002. Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

Vollset KW, Lennox RJ, Lamberg A, Skaala O, Sandvik AD, Saegrov H, Kvingedal E, Kristensen T, Jensen AJ, Haraldstad T, et al. 2021. Predicting the nationwide outmigration timing of Atlantic salmon (*Salmo salar*) smolts along 12 degrees of latitude in Norway. Divers Distrib DOI: 10.1111/ddi.13285

Weinz AA, Matley JK, Klinard NV, Fisk AT, Colborne SF. 2020. Identification of predation events in wild fish using novel acoustic transmitters. Anim Biotelemetry https://doi.org/10.1186/s40317-020-00215-x

Whalen KG, Parrish DL, McCormick SD. 1999. Migration timing of Atlantic salmon smolts relative to environmental and physiological factors. T Am Fish Soc 128: 289-301.

Wilcove DS, Wikelski M. 2008. Going, going, gone: is animal migration disappearing? PLoS Biol 6(7): e188 DOI: 10.1371/journal.pbio.0060188.

Willcocks-Musselman R. 2003. Minas Basin Watershed Profile. Bay of Fundy Ecosystem Partnership Technical Report #2. Retrieved from http://www.bofep.org/PDFfiles/MB_watershed_profile.pdf

Zadunaisky JA. 1996. Chloride cells and osmoregulation. Kidney Int 49(6): 1563-1567.
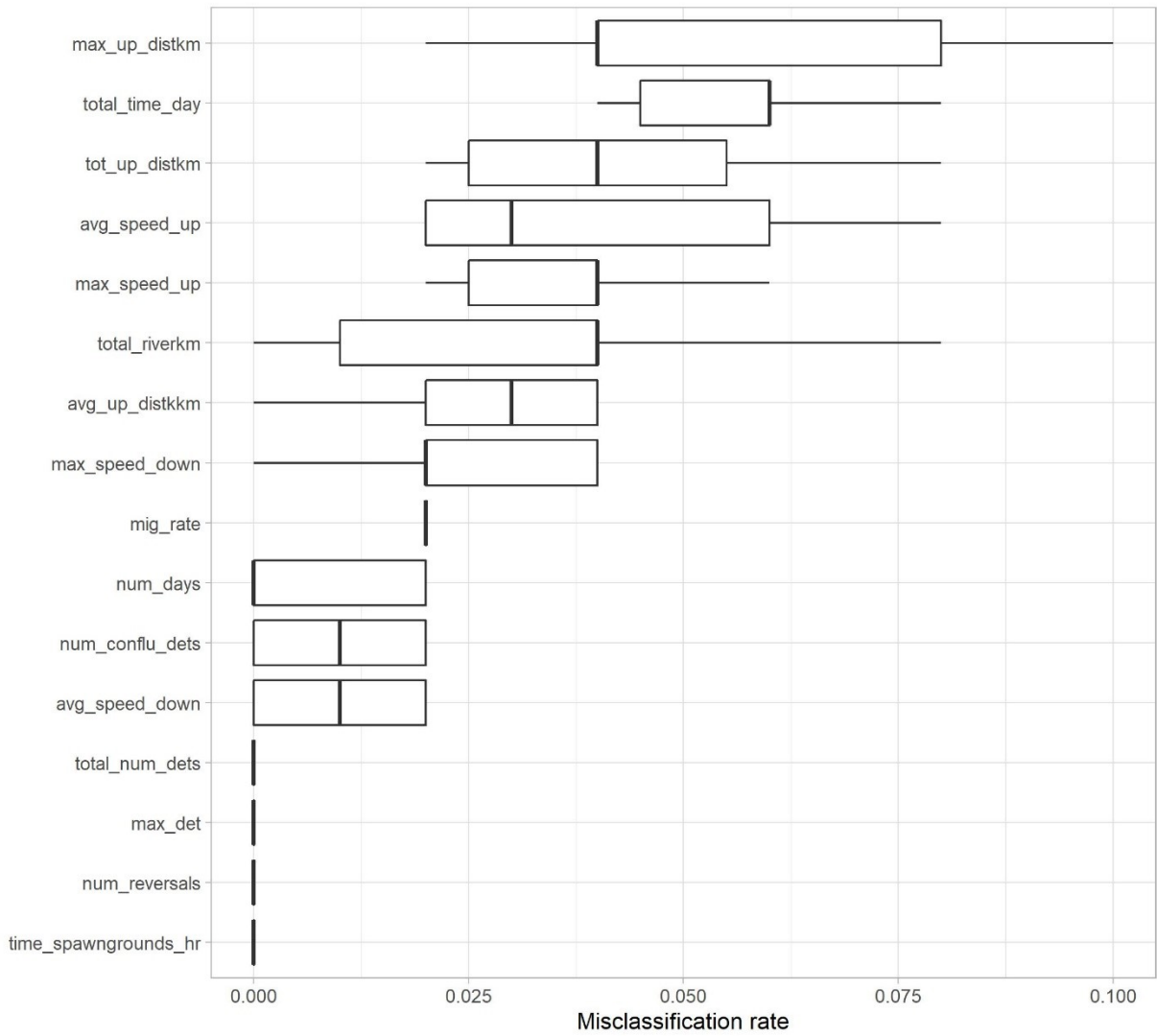
Fig. A1 Variable importance for the 2017 cluster analysis as determined by the rate at which individuals are misclassified if that variable is removed.
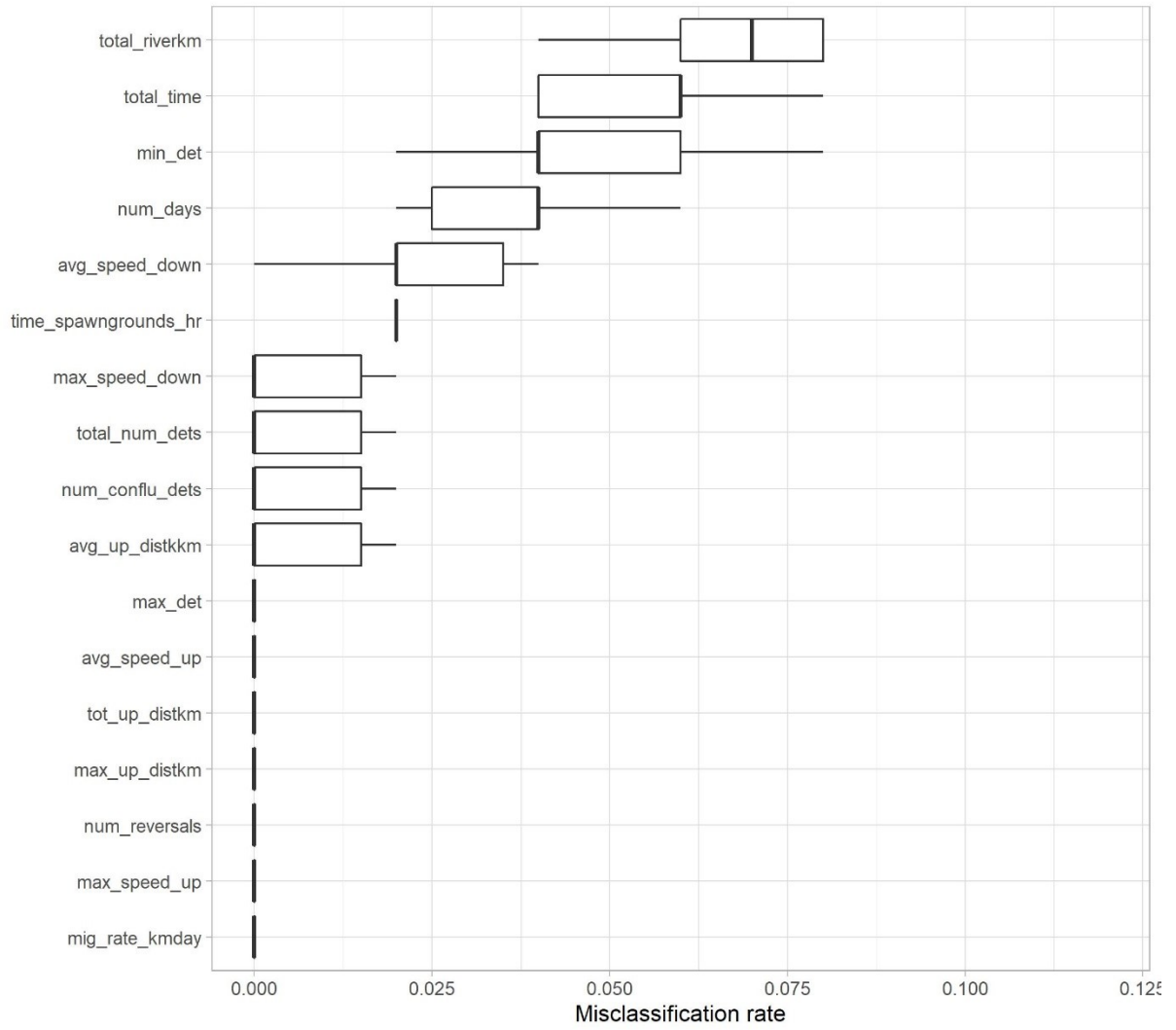
Fig. A2 Variable importance for the 2018 cluster analysis as determined by the rate at which individuals are misclassified if that variable is removed.
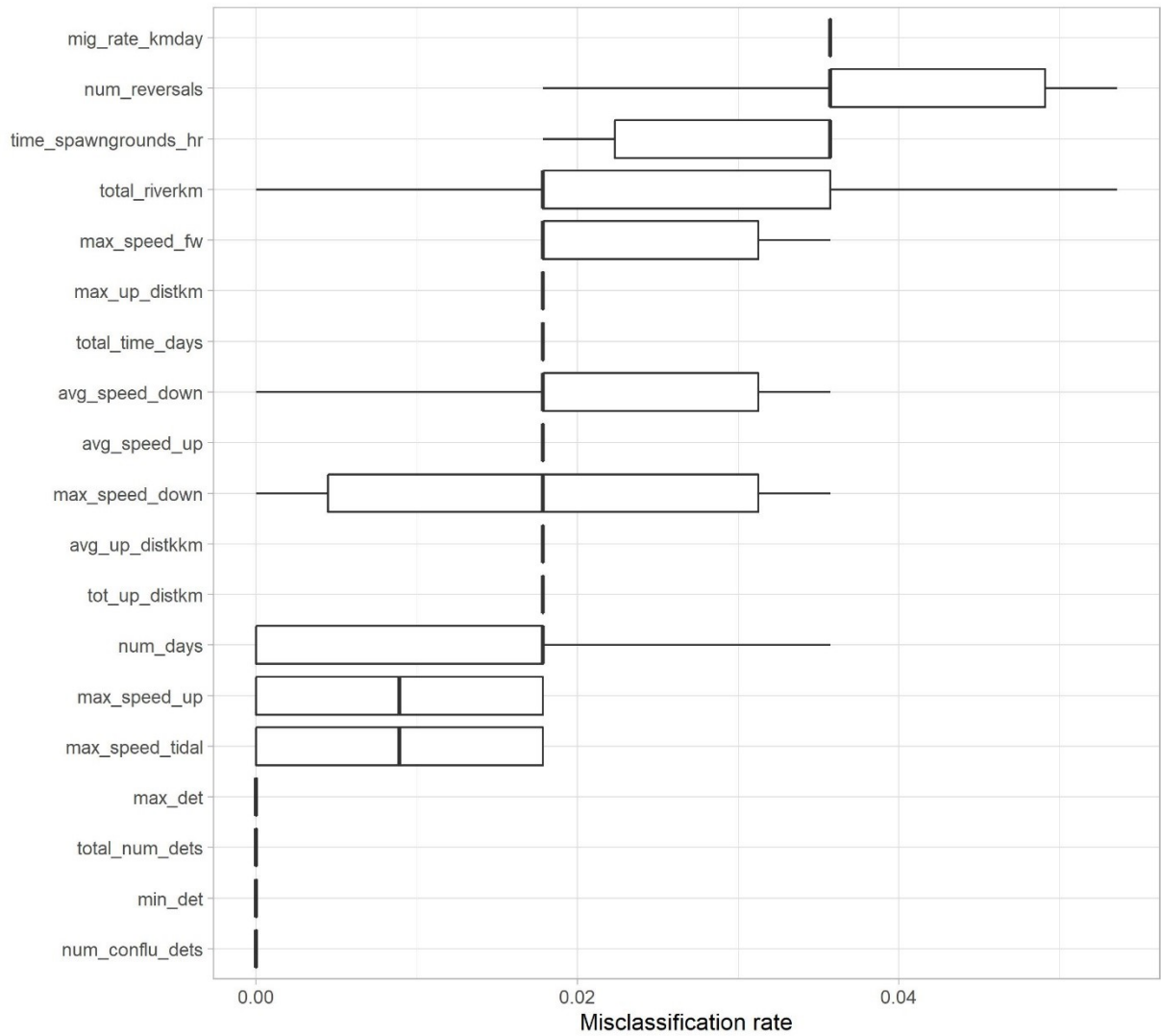
Fig. A3 Variable importance for the 2019 cluster analysis as determined by the rate at which individuals are misclassified if that variable is removed.
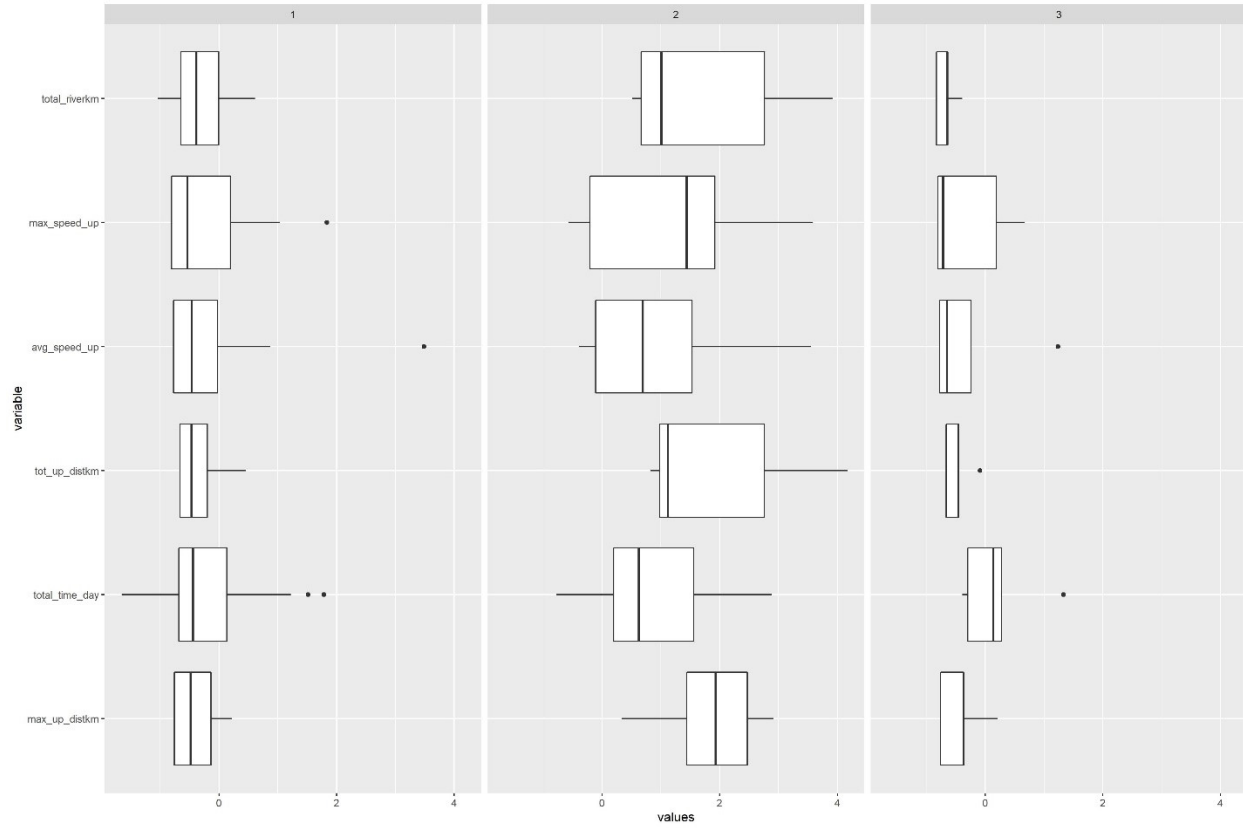
Fig. A4 Box and whisker plots between clusters for the most important variables identified by misclassification rate in 2017. Values are scaled and centered.
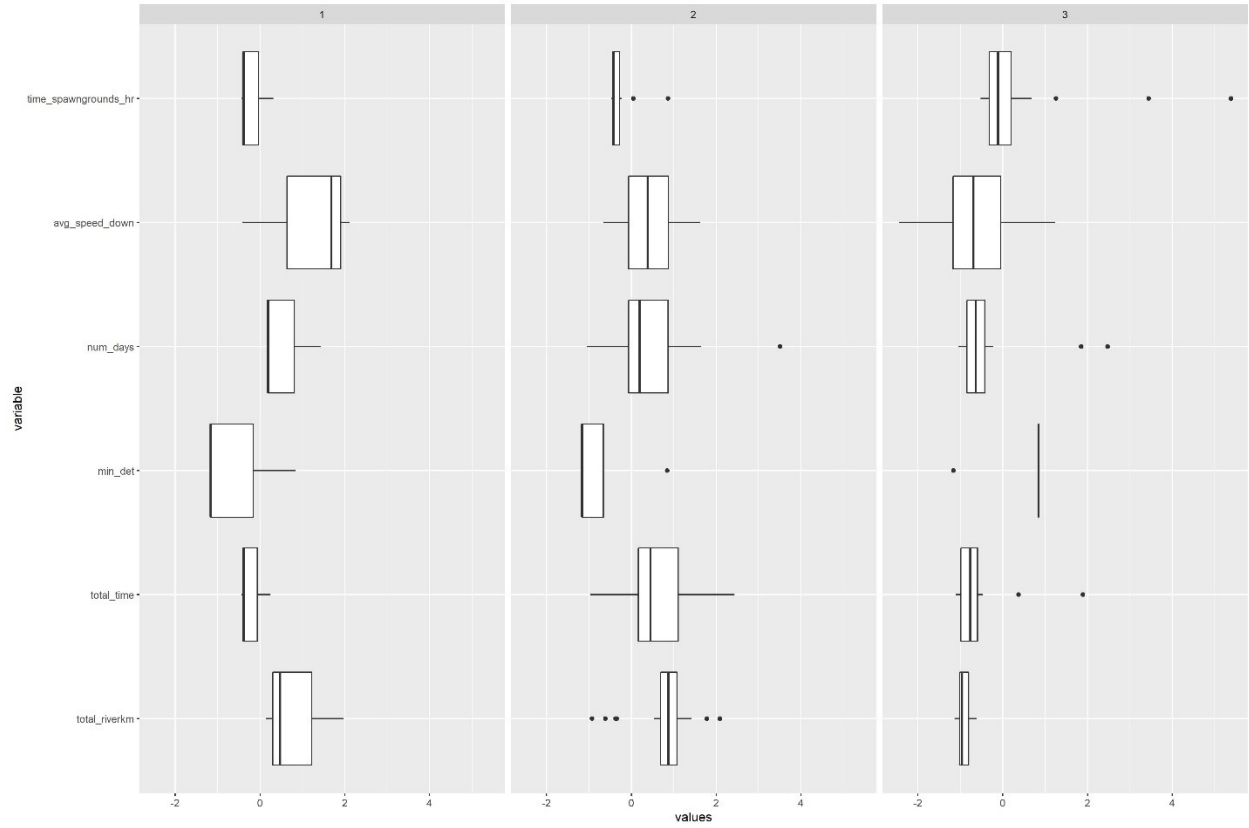
Fig. A5 Box and whisker plots between clusters for the most important variables identified by misclassification rate in 2018. Values are scaled and centered.
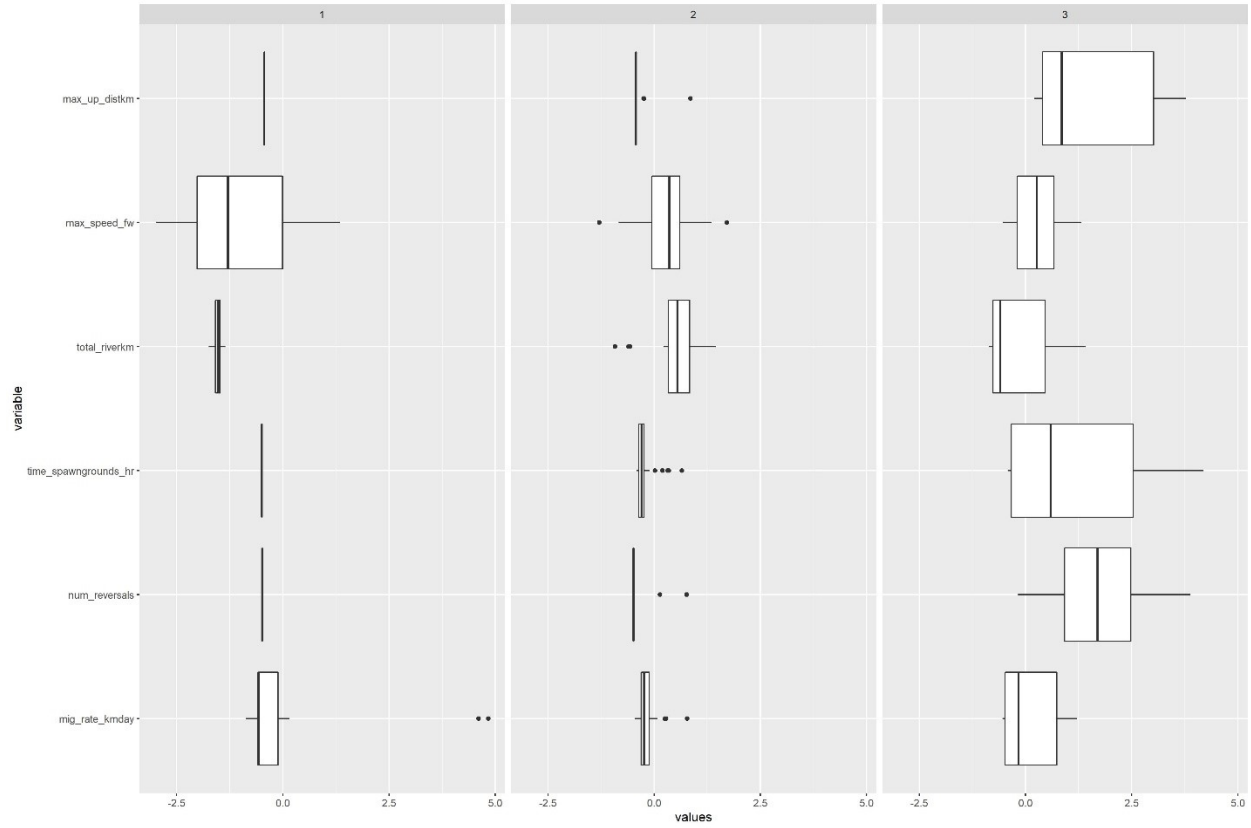
Fig. A6 Box and whisker plots between clusters for the most important variables identified by misclassification rate in 2019. Values are scaled and centered.
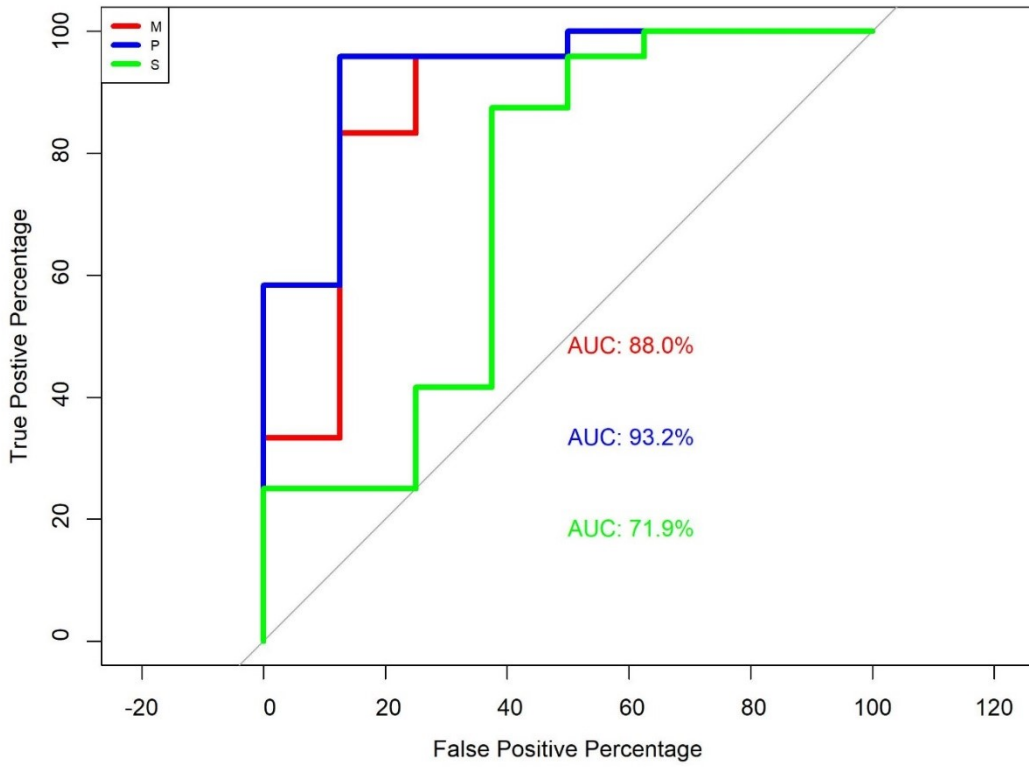
Fig. A7 Receiver operating characteristic (ROC) curve with area under curve (AUC) values for each class; mortality (M), predated (P), and successful migrant (S) for 2017 random forest.
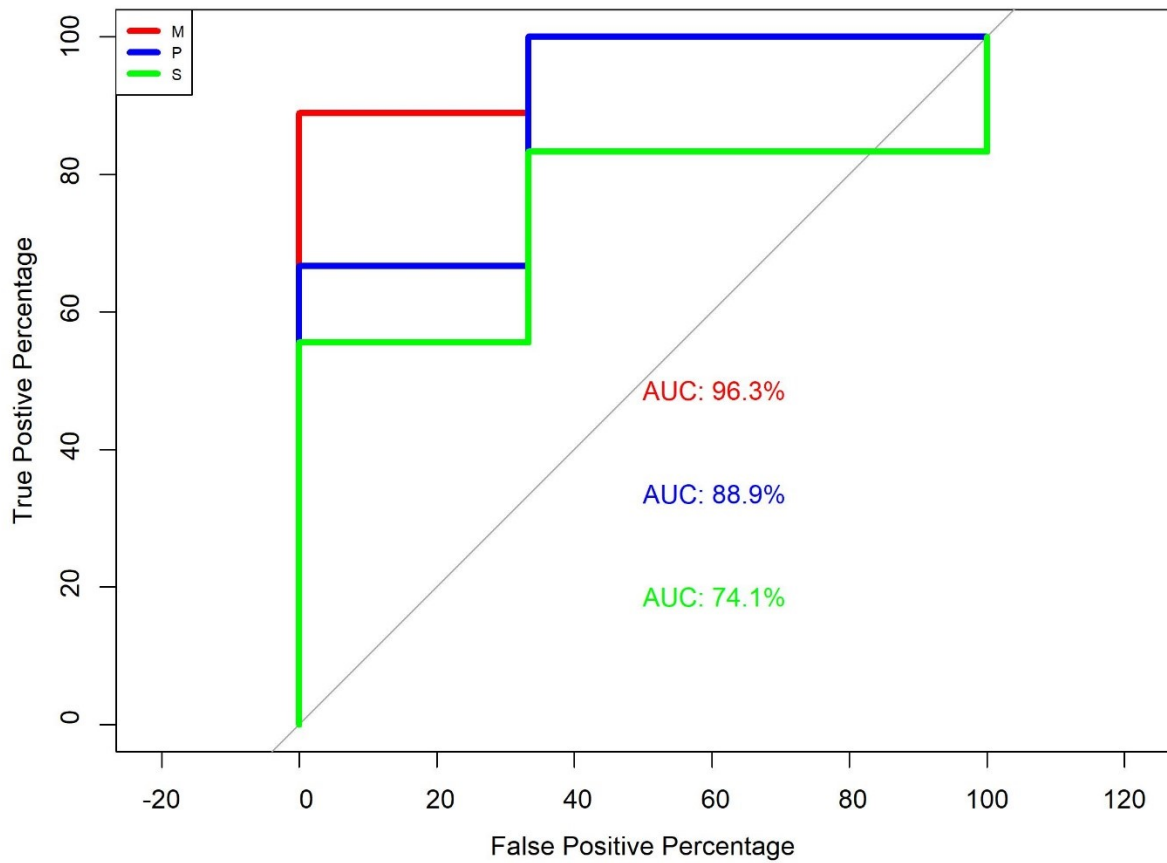
Fig. A8 Receiver operating characteristic (ROC) curve with area under curve (AUC) values for each class; mortality (M), predated (P), and successful migrant (S) for 2018 random forest.
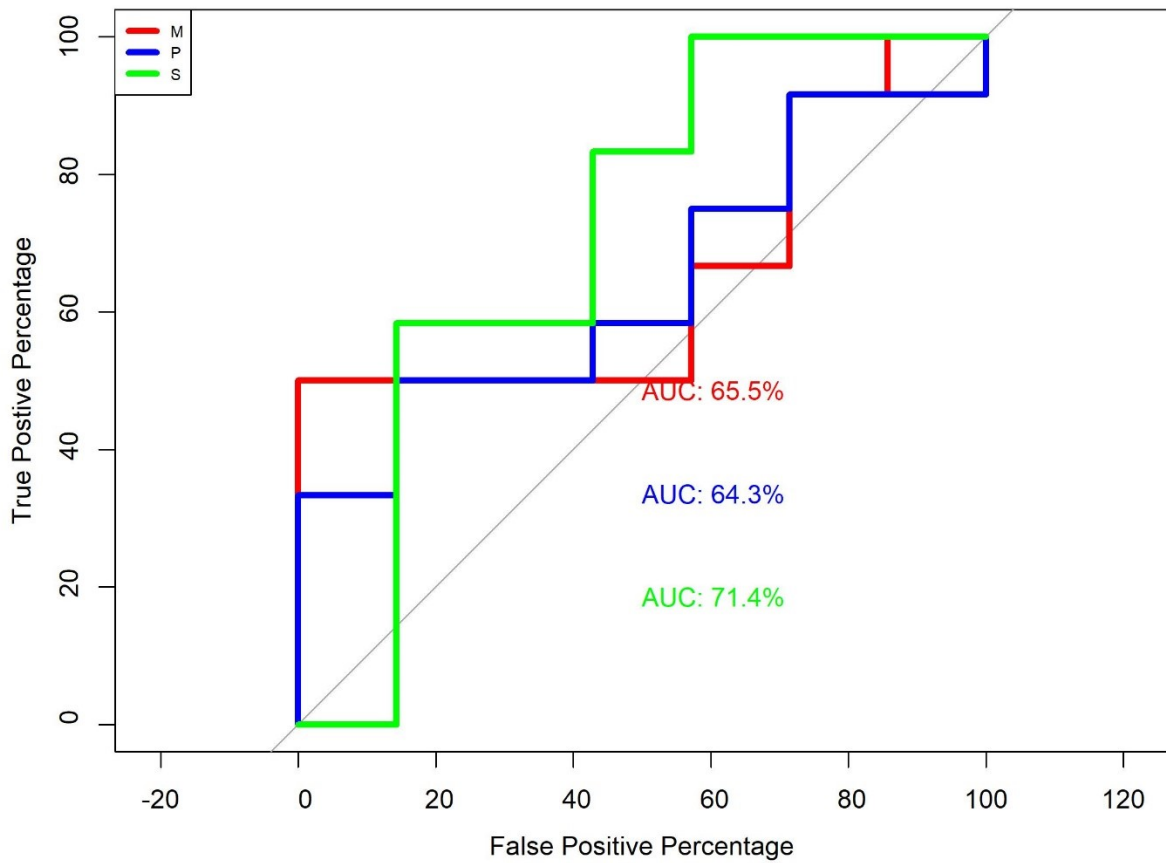
Fig. A9 Receiver operating characteristic (ROC) curve with area under curve (AUC) values for each class; mortality (M), predated (P), and successful migrant (S) for 2019 random forest.
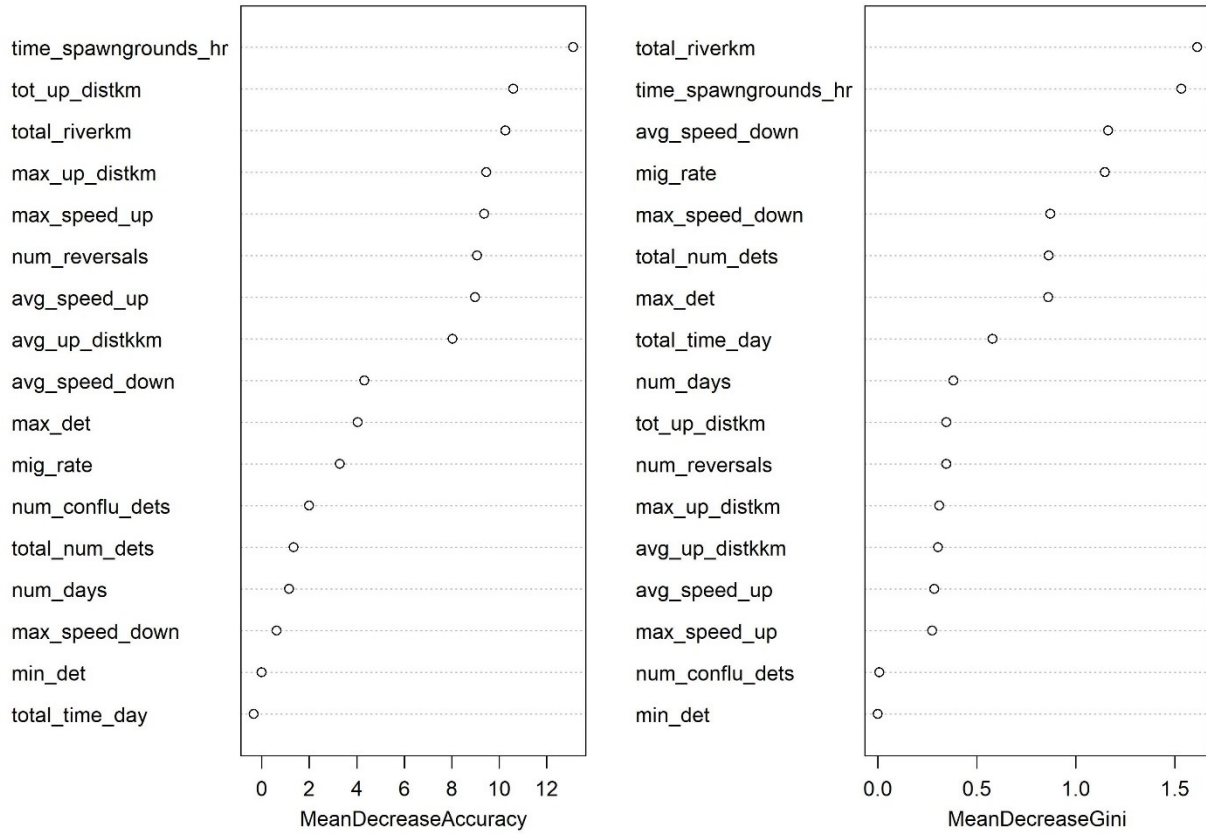
Fig. A10 Variable importance as measured by mean decrease accuracy and mean decrease Gini for 2017 random forest.
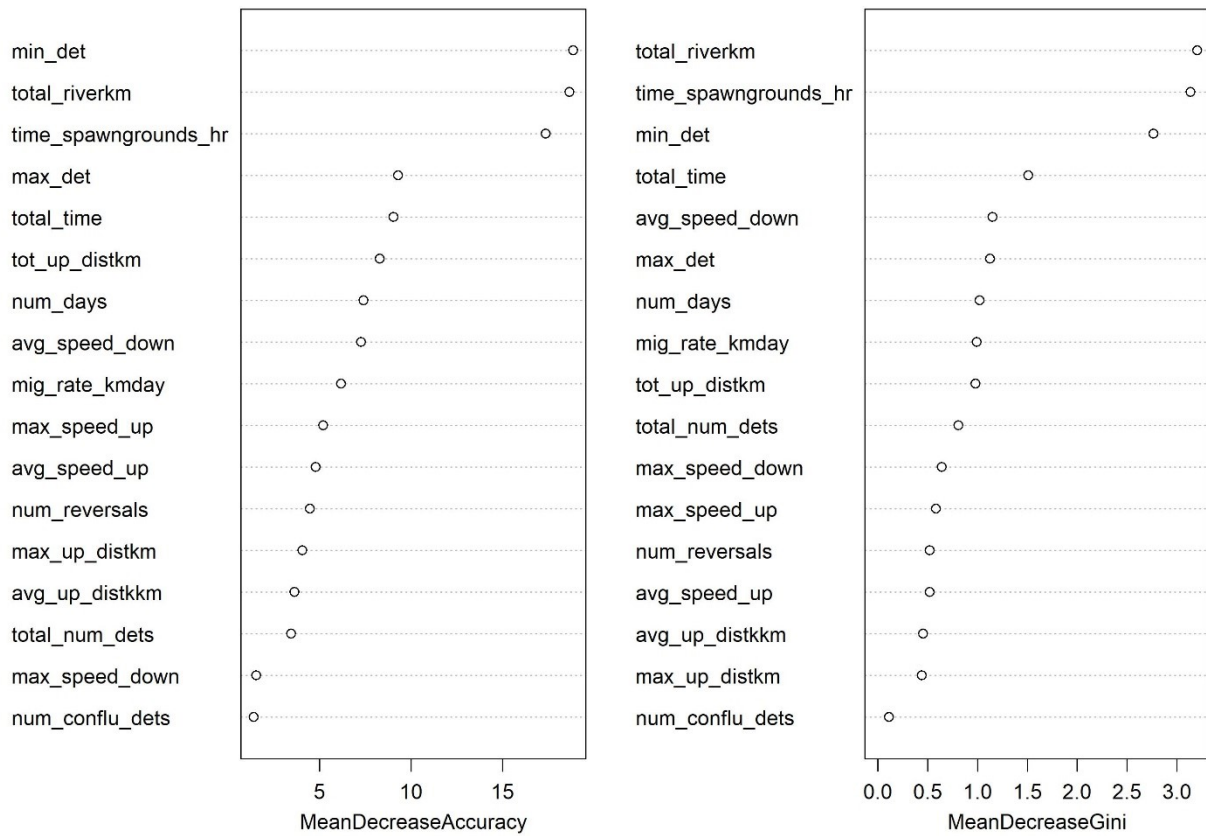
Fig. A11 Variable importance as measured by mean decrease accuracy and mean decrease Gini for 2018 random forest.
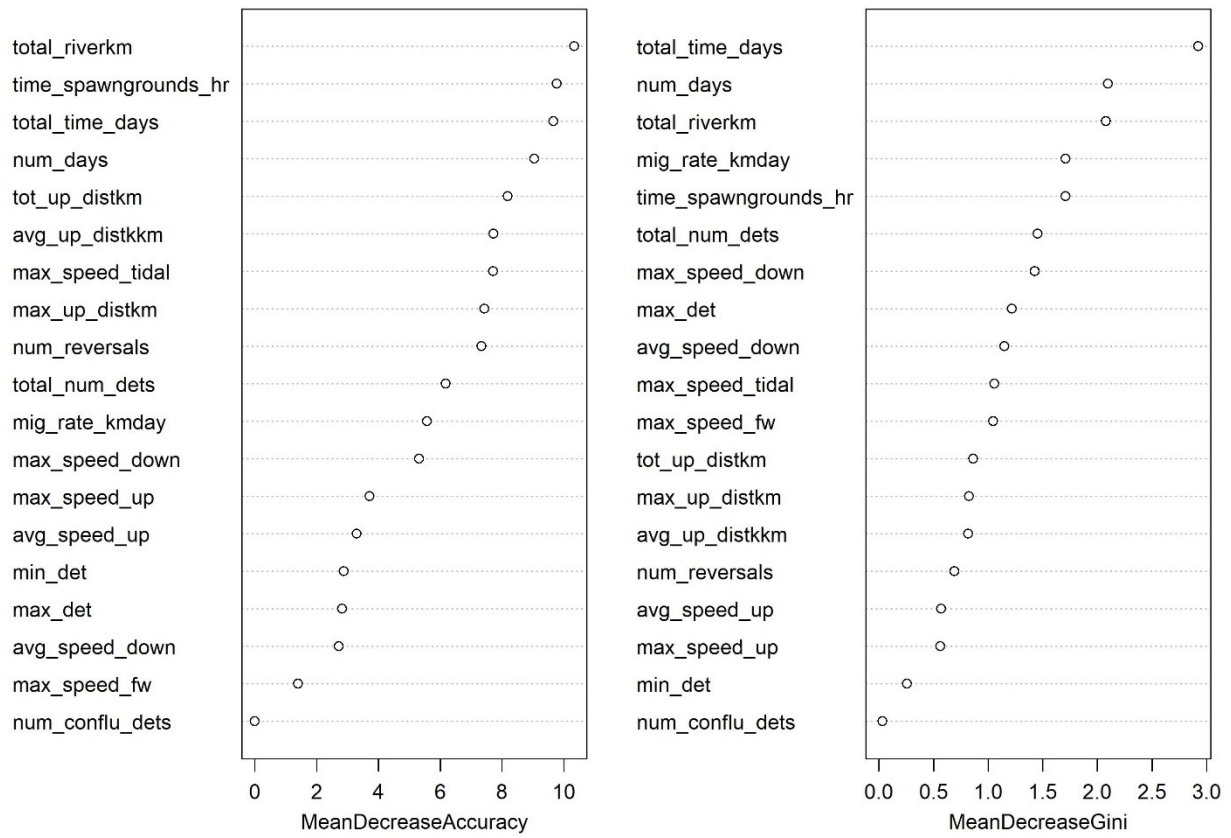
Fig. A12 Variable importance as measured by mean decrease accuracy and mean decrease Gini for 2019 random forest.
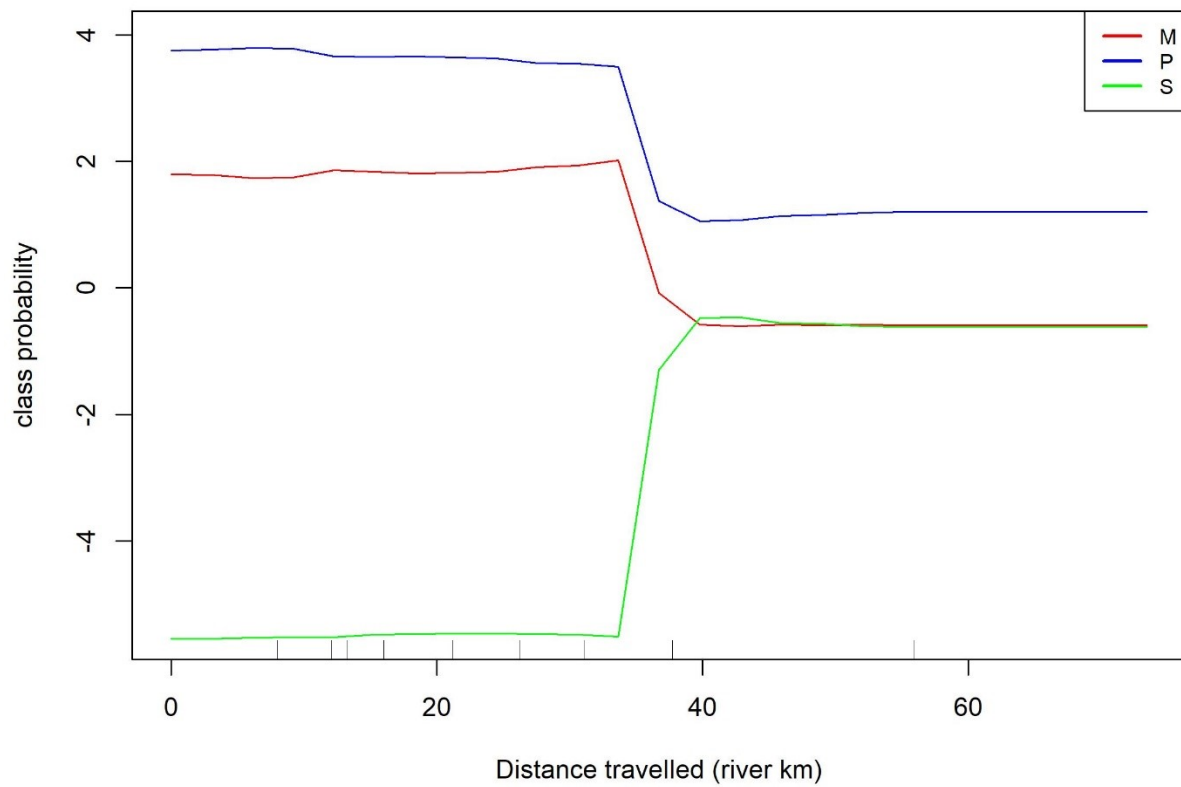
Fig. A13 Partial plot for total distance travelled variable. Probability of being assigned to each class (mortality M, predated P, successful migrant S) by 2017 random forest on a log scale with change in variable value.
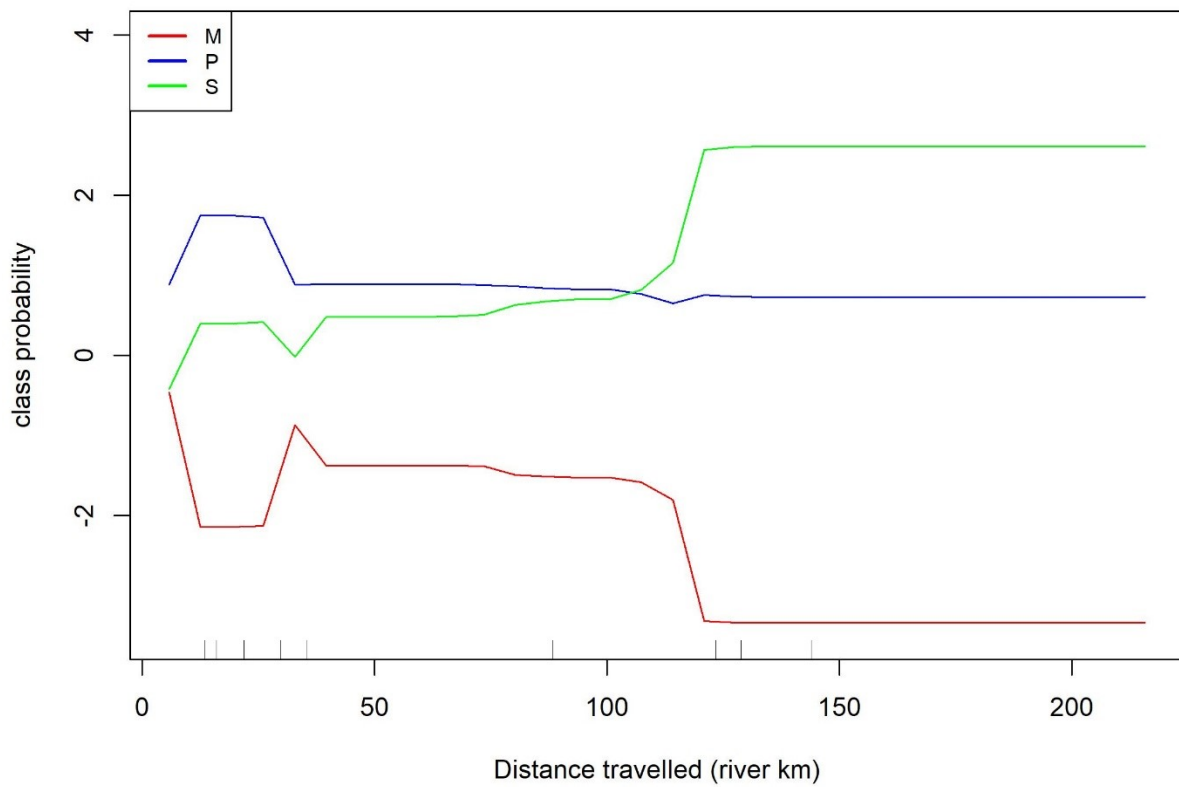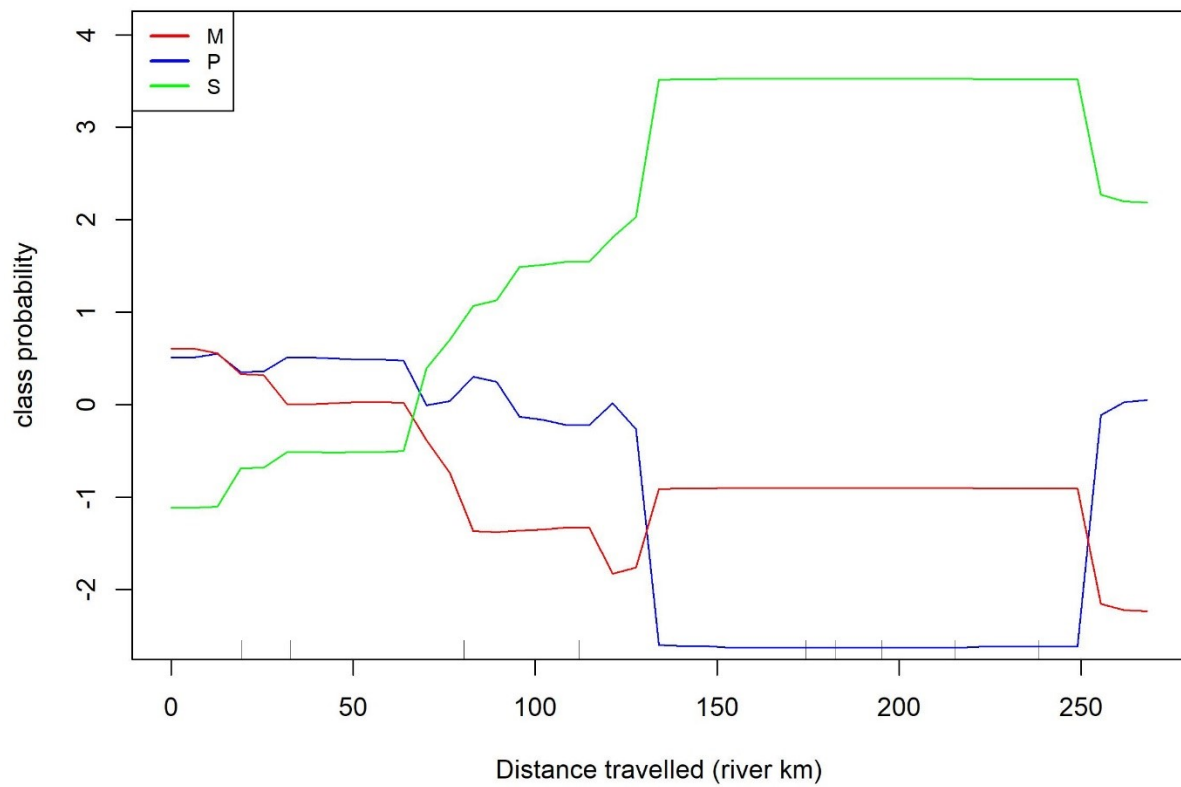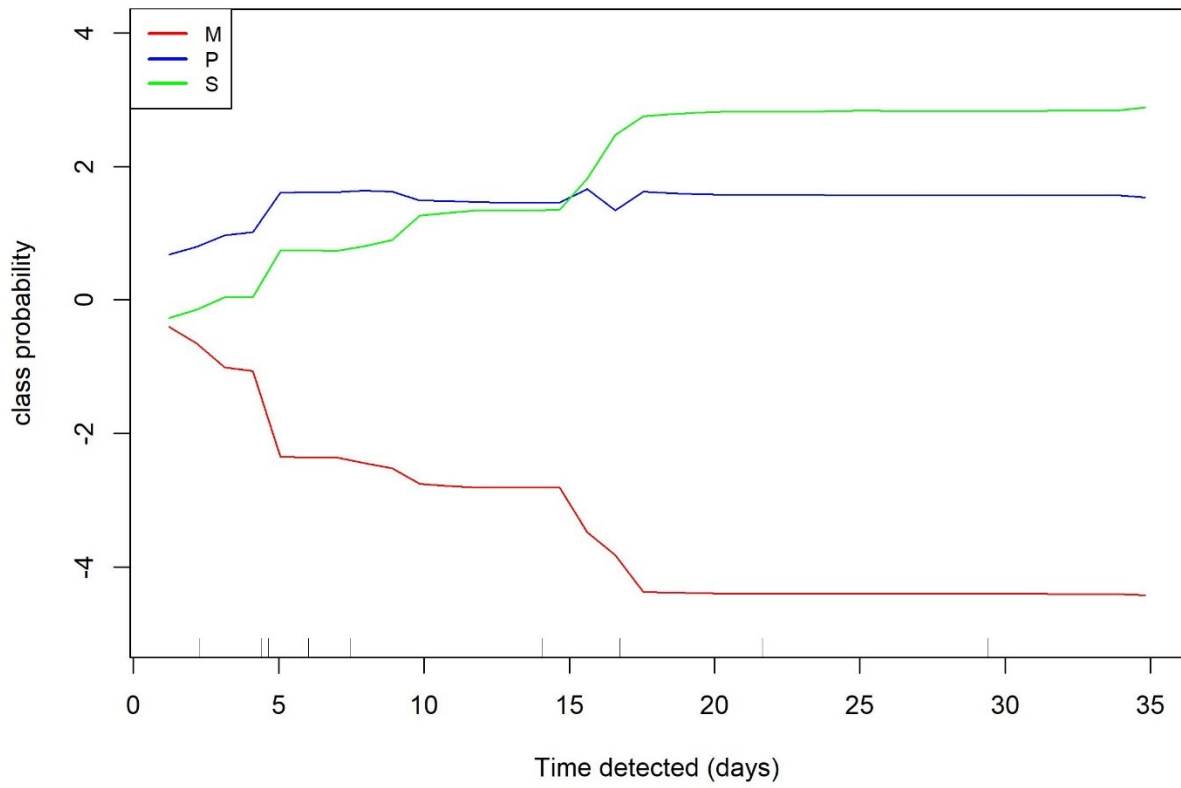
Fig. A14 Partial plot for total distance travelled variable. Probability of being assigned to each class (mortality M, predated P, successful migrant S) by 2018 random forest on a log scale with change in variable value.

Fig. A15 Partial plot for total distance travelled variable. Probability of being assigned to each class (mortality M, predated P, successful migrant S) by 2019 random forest on a log scale with change in variable value
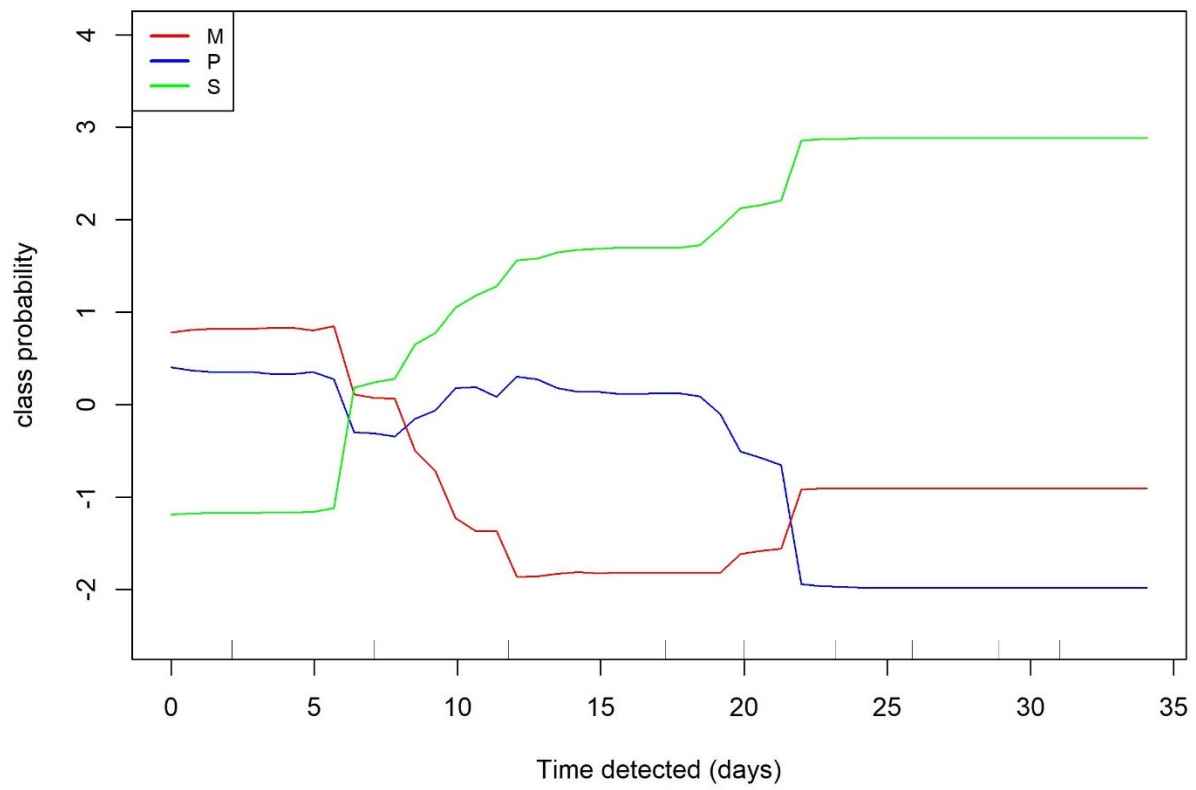
Fig. A16 Partial plot for total time detected variable. Probability of being assigned to each class (mortality M, predated P, successful migrant S) by 2018 random forest on a log scale with change in variable value.

Fig. A17 Partial plot for total time detected variable. Probability of being assigned to each class (mortality M, predated P, successful migrant S) by 2019 random forest on a log scale with change in variable value
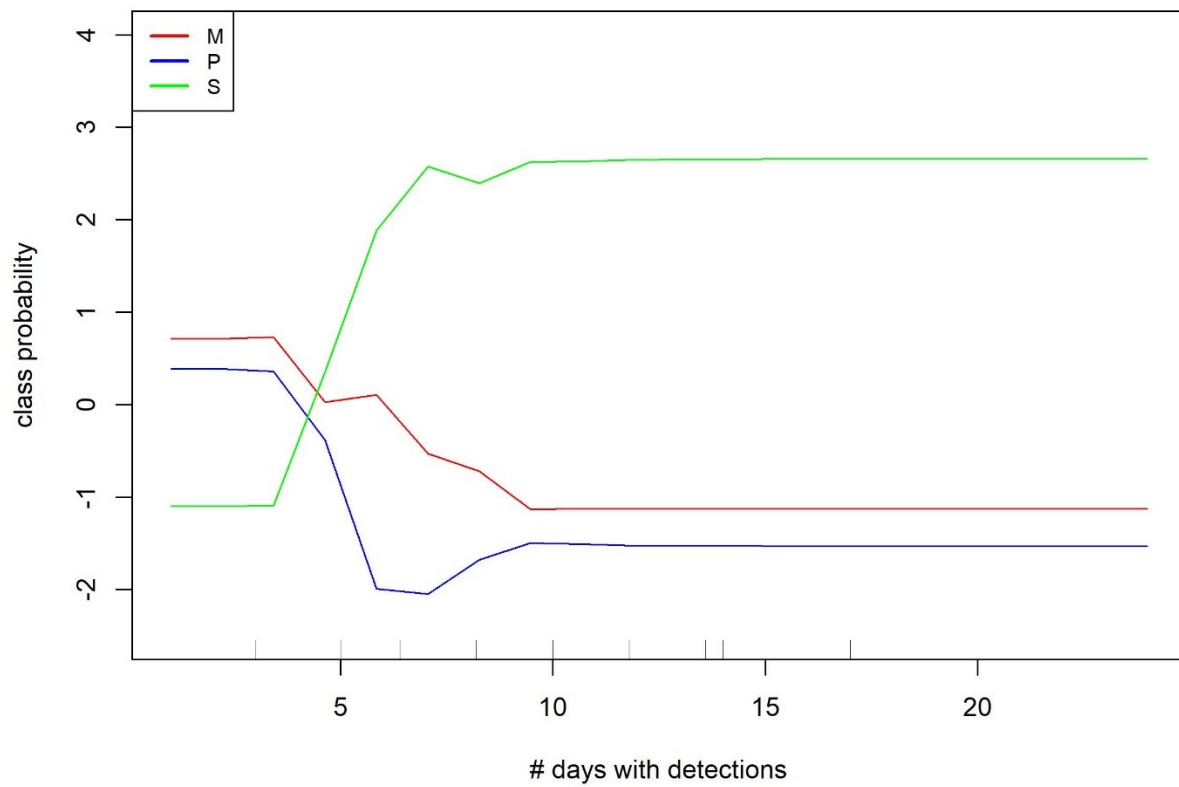
Fig. A18 Partial plot for number of days with detections variable. Probability of being assigned to each class (mortality M, predated P, successful migrant S) by 2019 random forest on a log scale with change in variable value.
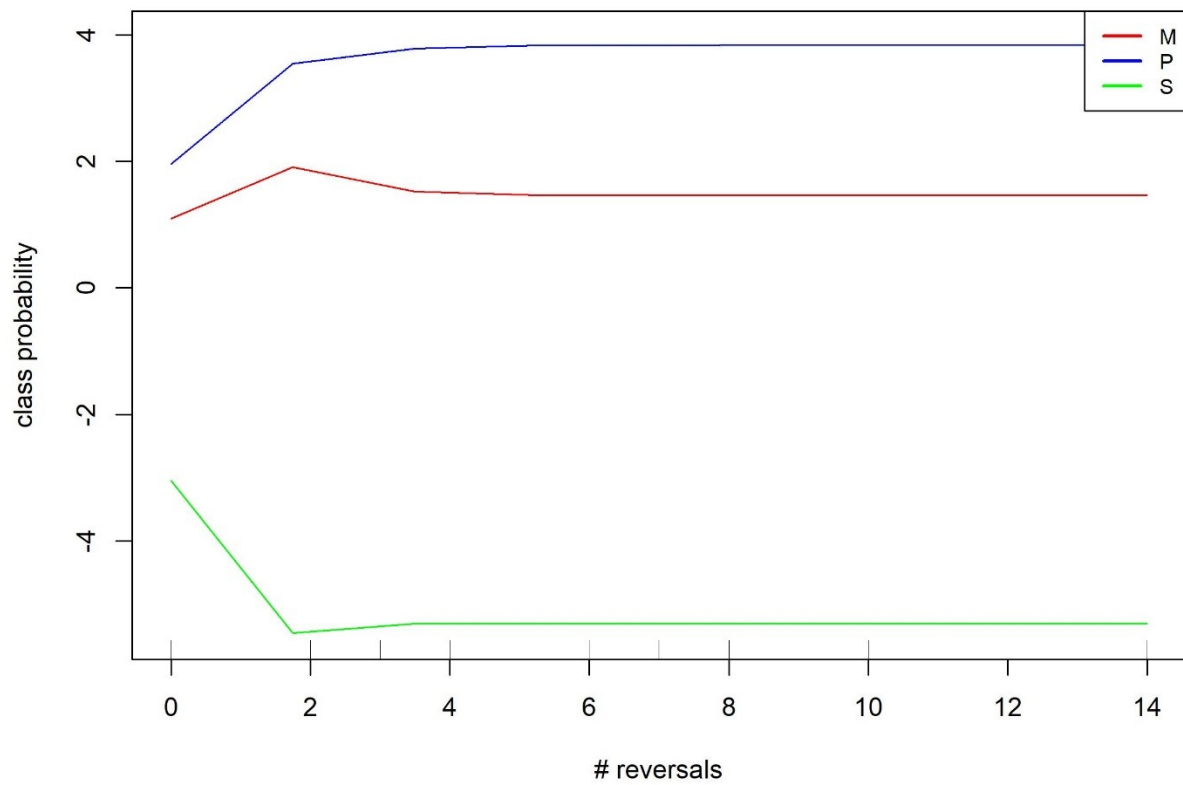
Fig. A19 Partial plot for number of reversals variable. Probability of being assigned to each class (mortality M, predated P, successful migrant S) by 2017 random forest on a log scale with change in variable value.
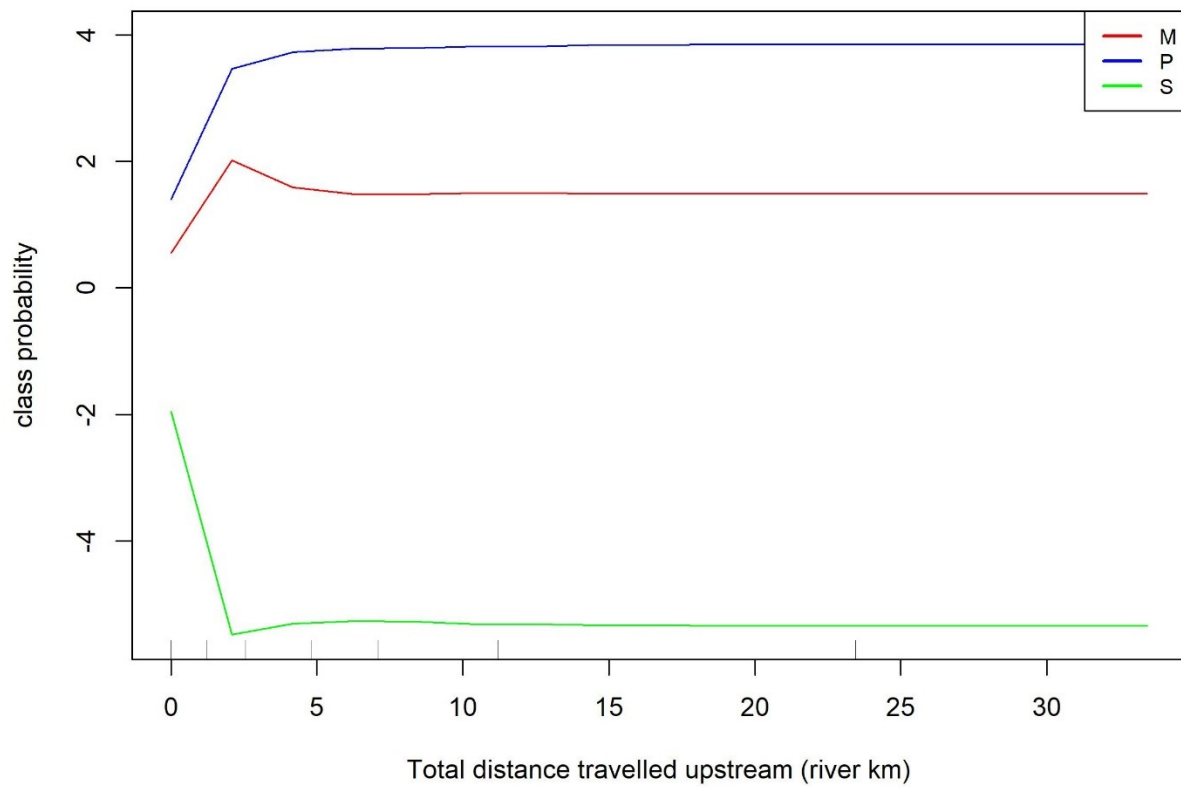
Fig. A20 Partial plot for total upstream distance travelled variable. Probability of being assigned to each class (mortality M, predated P, successful migrant S) by 2017 random forest on a log scale with change in variable value.
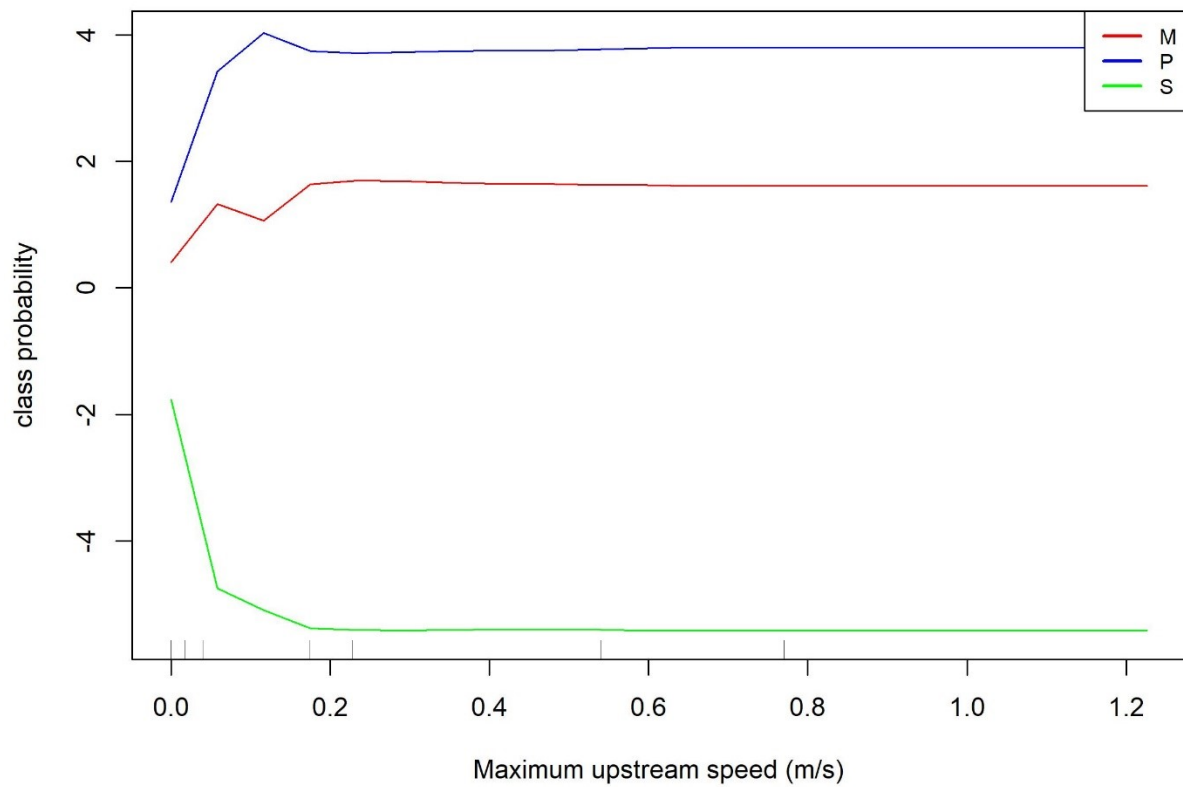
Fig. A21 Partial plot for max upstream speed between two consecutive receivers variable. Probability of being assigned to each class (mortality M, predated P, successful migrant S) by 2017 random forest on a log scale with change in variable value.

Fig. A22 Partial plot for time on Striped Bass spawning grounds variable. Probability of being assigned to each class (mortality M, predated P, successful migrant S) by 2017 random forest on a log scale with change in variable value.

 Fig. A23 Partial plot for time on Striped Bass spawning grounds variable. Probability of being assigned to each class (mortality M, predated P, successful migrant S) by 2018 random forest on a log scale with change in variable value.
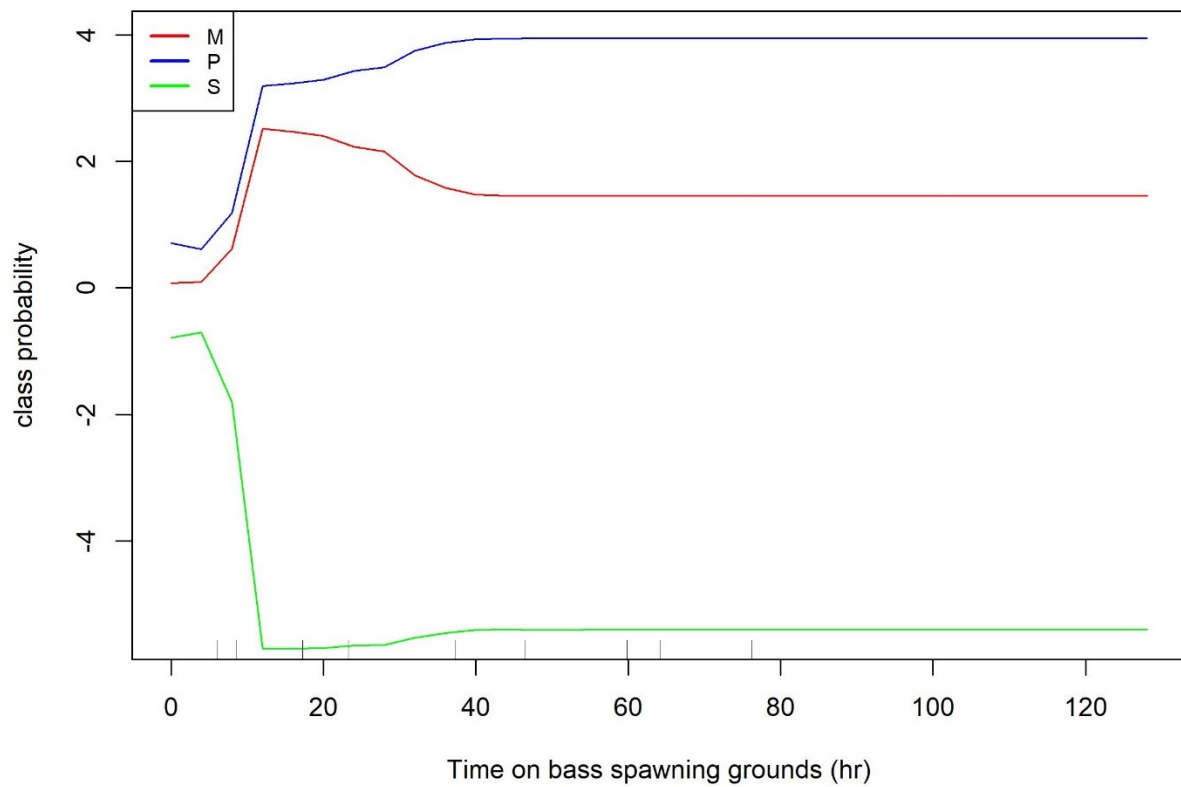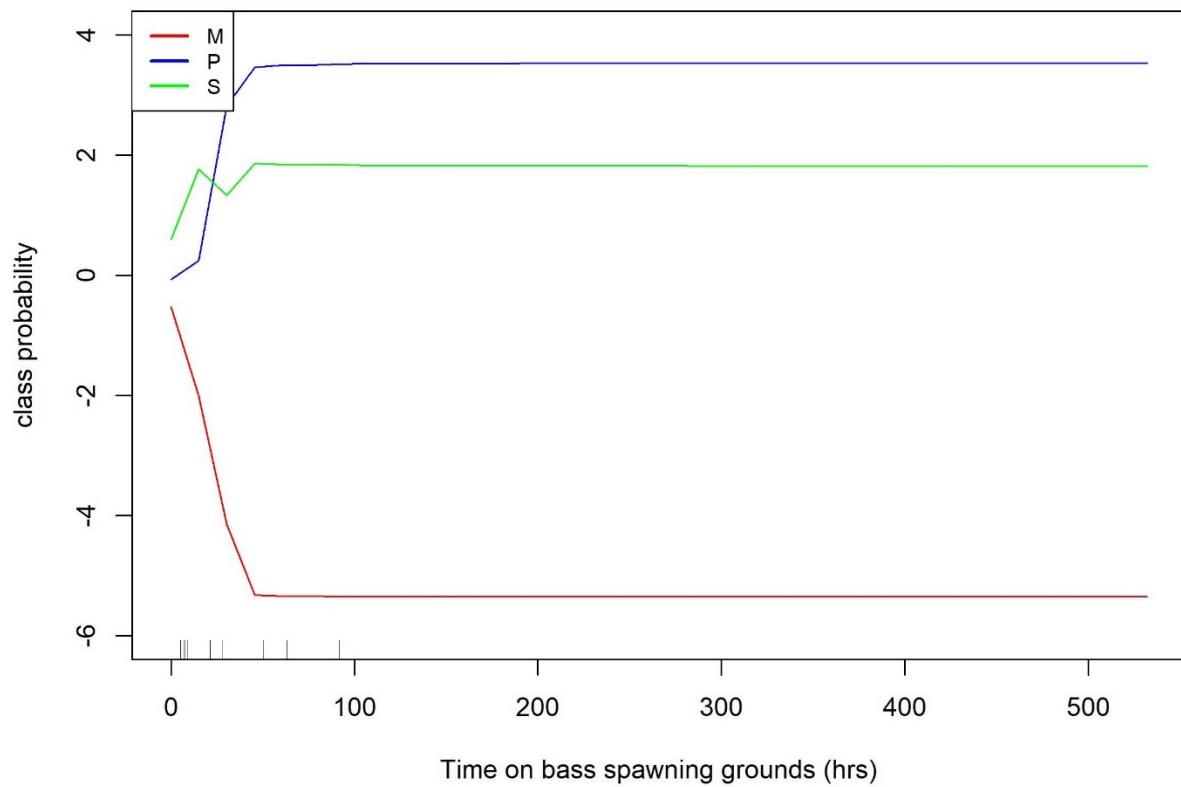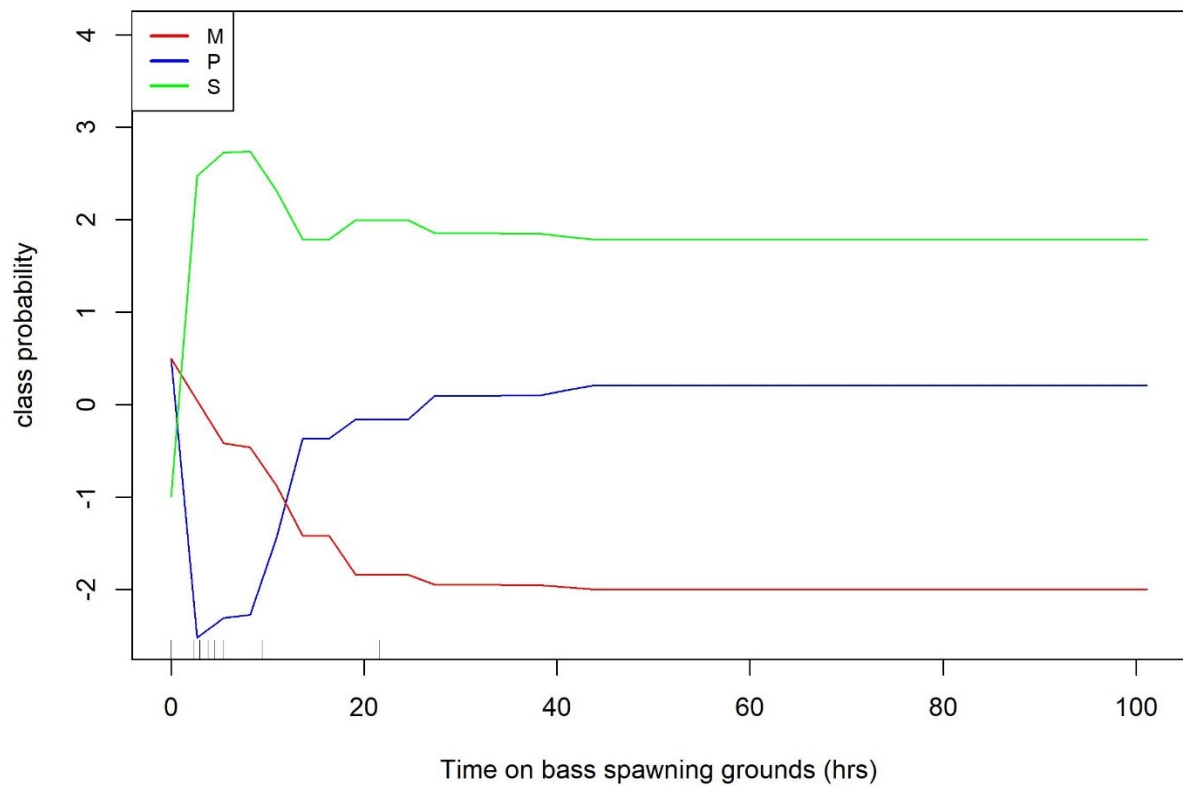
Fig. A24 Partial plot for time on Striped Bass spawning grounds variable. Probability of being assigned to each class (mortality M, predated P, successful migrant S) by 2019 random forest on a log scale with change in variable value.

Appendix B: Receiver Locations

Table B1 Description of receiver stations: station name/location, corresponding number on maps (Figs. B1 and B2), number of receivers at each station, and which study years they were deployed. Receiver stations upstream of the Stewiacke/Shubenacadie confluence are not listed because they are not considered part of the smolt migration route.

Receiver station	Number on map	Number of receivers	Years deployed
J Graham	01	1	2019
Gaults	02	1	2019
Stewart Hill	03	1	2019
Ridge Rd	04	1	2019
Cloverdale	05	1	2019
Brenton Cross	06	1	2019
Brenton & Hemlock	07	1	2019
Scout Ground	08	1	2019
River Park	09	1	2019
Rockpile	10	1	all
Eddy Pool	11	1	all
Porter	12	1	all
Trestle	13	1	2018, 2019
Kent Farm	14	1	all
Stewiacke 2	15	1	all
Moxam's	16	1	all
Gosse Bridge	17	1	all
Shubie mouth	18	4	all
Minas Basin	NA	41	2018, 2019
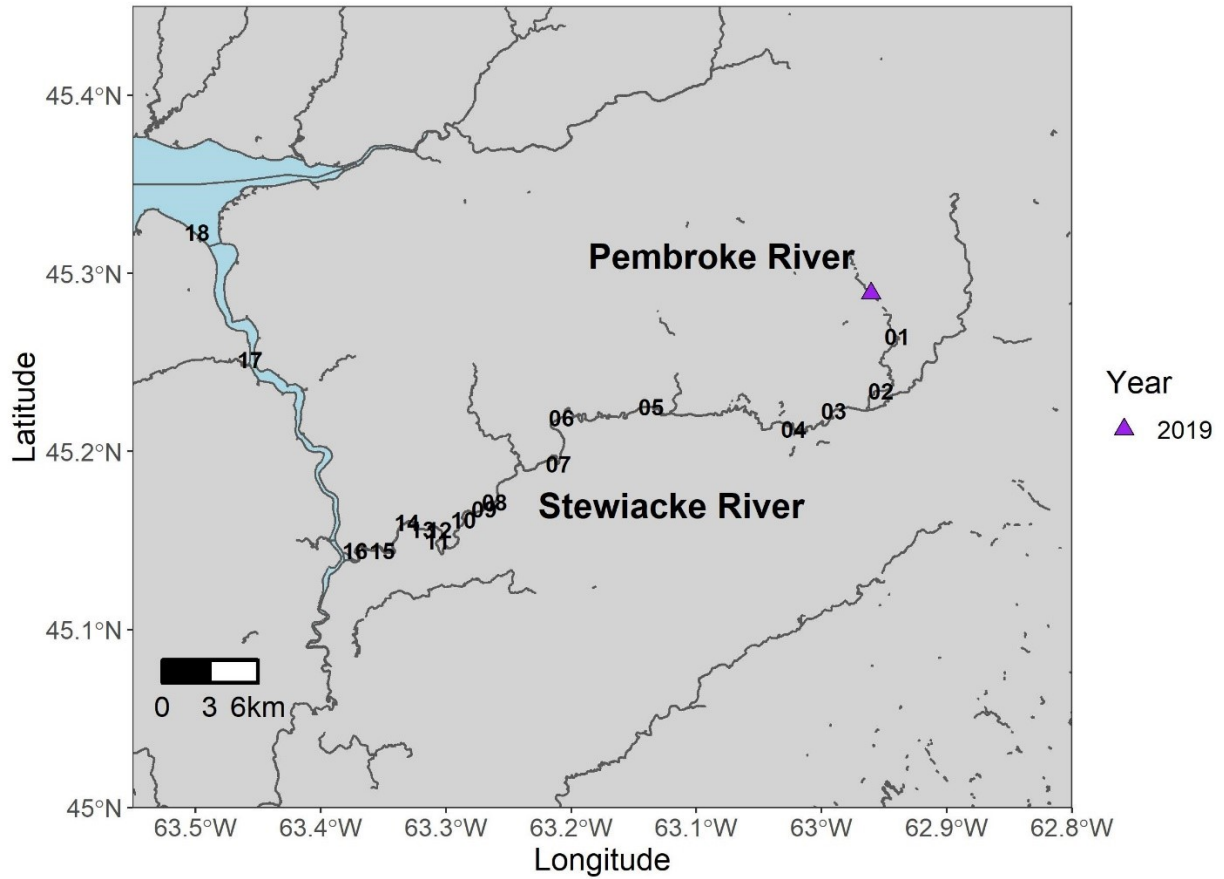
Fig. B1 Close up of receiver stations 01-18 on the Pembroke, Stewiacke, and Shubenacadie Rivers. 2019 release site shown (triangle). See Table B1 for station names. Receivers in the Minas Basin and upstream of the Stewiacke/Shubenacadie confluence not shown, see Fig. 2.1 for view of whole study area.
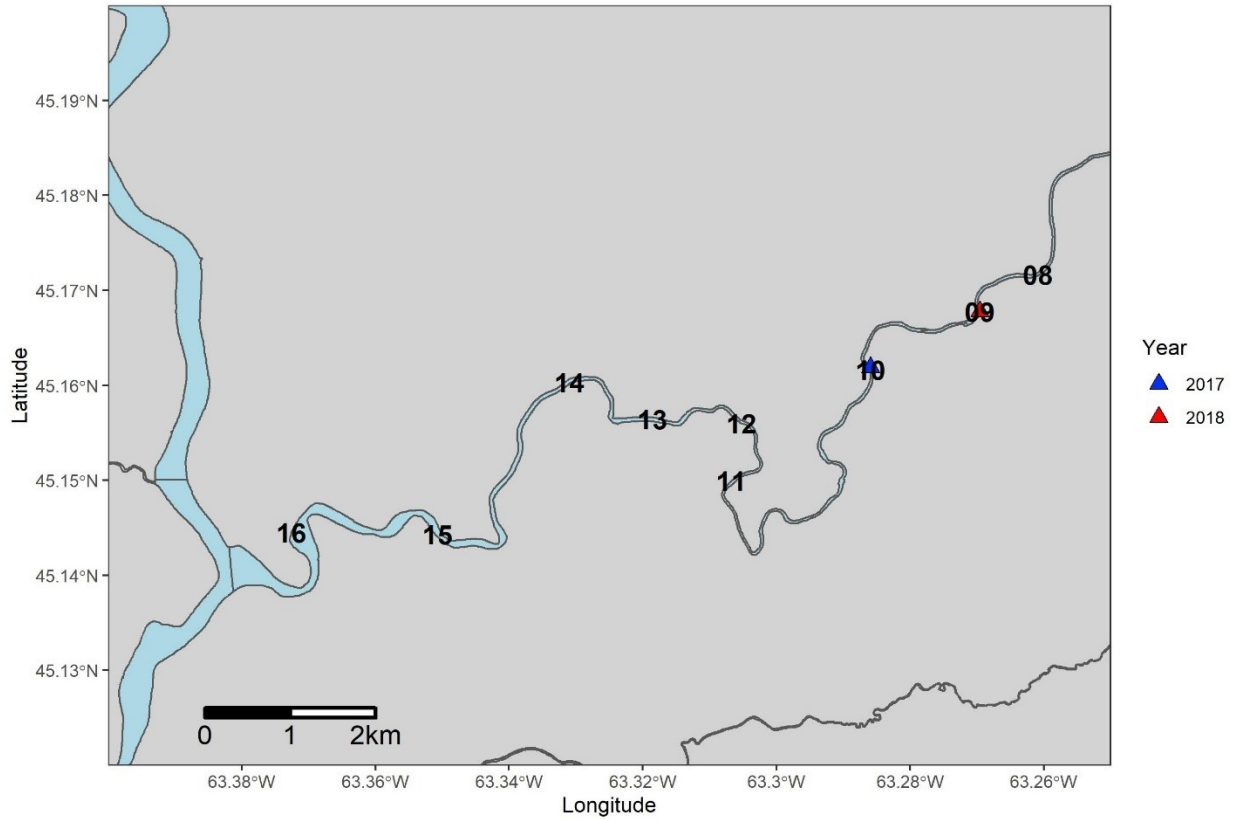
Fig. B2 Close up of receiver stations 08-16 in the tidal waters of the Stewiacke River. 2017 and 2018 release sites shown (triangles). See Table B1 for station names.

Appendix C: Example R Code for Chapter 2 Analyses

```
##### Example code for k-means clustering #####

library(tidyverse)

library(cluster)

library(factoextra)

library(ggplot2)

library(remotes)

devtools::install_github("o1iv3r/FeatureImpCluster")

library(FeatureImpCluster)

library(flexclust)


#read in behavioural metrics

metrics <- read.csv("2017 behavioural metrics cut and no VR.csv")


#set seed to get same results with each run

set.seed(41)


#make tag SN row name

metrics2 <- metrics %>% remove_rownames %>% column_to_rownames(var="tag_sn")


#centers (by mean) and scales (by sd) metrics

metrics2 <- scale(metrics2)


#run k means cluster analysis

k <- kmeans(metrics2, centers=3) #centers (k) = 3 for the 3 fate groups
```

```
#look at results

k


#visualize clusters

?fviz_cluster

fviz_cluster(k, data=metrics2, geom="point", main="2017 Cluster plot") #cluster plot


#variable importance plots

res <- kcca(metrics2,k=3)

FeatureImp_res <- FeatureImpCluster(res,as.data.table(metrics2))

plot(FeatureImp_res)


#make df with cluster number and tag SN

df <- data.frame(k$cluster)

df <- rownames_to_column(df, "tag_sn")

#add cluster numbers to metrics df

metrics$tag_sn <- as.character(metrics$tag_sn)

metrics <- left_join(metrics, df)

#save df



##### Example code for random forest #####

#code adapted from https://github.com/StatQuest/random_forest_demo

library(ggplot2)

library(tidyverse)

#install.packages("randomForest")

library(randomForest)

#install.packages("caret")
```

```
library(caret)

#install.packages("ROCR")

library(ROCR)

#install.packages("pROC")

library(pROC)


#read in data

data <- read.csv("2017 behavioural metrics cut and no VR.csv")


#add fate column to metrics df

f <- read.csv("2017 detections cut.csv")

f <- f %>% select(tag_sn, fate_1) %>% distinct()

data <- left_join(data, f)


#set seed to get same results

set.seed(41)


#remove individuals with suspect fate from model training df and add to a new df

unknown <- data %>% filter(tag_sn %in% c(1262412, 1262413, 1262414, 1262418, 1262423,
1262432, 1262433, 1262435, 1262436, 1262439, 1262440, 1262441, 1262450, 1262455,
1262444, 1262446))

data2 <- data %>% filter(!tag_sn %in% c(1262412, 1262413, 1262414, 1262418, 1262423,
1262432, 1262433, 1262435, 1262436, 1262439, 1262440, 1262441, 1262450, 1262455,
1262444, 1262446))


#make tag SN row name

data2 <- data2 %>% remove_rownames %>% column_to_rownames(var="tag_sn")

unknown <- unknown %>% remove_rownames %>% column_to_rownames(var="tag_sn")


#run initial model
```

```
m1 <- randomForest(fate_1 ~ ., data2, proximity=T, importance=T) #default ntree=500
m1 #look at OOB and class error


#tune number of trees
#try model with ntree=1000
m2 <- randomForest(fate_1 ~ ., data2, proximity=T, importance=T, ntree=1000)
m2


#make df for changes in OOB and class error rates with change in ntree
m2_oob_error <- data.frame(
  Trees=rep(1:nrow(m2$err.rate), times=4),
  Type=rep(c("OOB", "M", "P", "S"), each=nrow(m2$err.rate)),
  Error=c(m2$err.rate[,"OOB"],
      m2$err.rate[,"M"],
      m2$err.rate[,"P"],
      m2$err.rate[,"S"]))


#plot df look for when error rates stabilize - if not stable at 1000 increase ntree and check again
ggplot(data=m2_oob_error, aes(x=Trees, y=Error)) +
  geom_line(aes(color=Type))
#unclear if stable at 1000, check 1500


#try ntree=1500
m3 <- randomForest(fate_1 ~ ., data2, proximity=T, importance=T, ntree=1500)
m3


m3_oob_error <- data.frame(
  Trees=rep(1:nrow(m3$err.rate), times=4),
```

```
    Type=rep(c("OOB", "M", "P", "S"), each=nrow(m3$err.rate)),
    Error=c(m3$err.rate[,"OOB"],
        m3$err.rate[,"M"],
        m3$err.rate[,"P"],
        m3$err.rate[,"S"]))
```

```
ggplot(data=m3_oob_error, aes(x=Trees, y=Error)) +
  geom_line(aes(color=Type))
#this shows stable at 500
```

```
#so use 1000
```

```
#now tune mtry (number of variables tried at each node) at ntree=1000
#default mtry is the square root of the number of variables
```

```
#calculate OOB error for mtry 1 to 10
oob.values <- vector(length=10)
for(i in 1:10) {
  temp.model <- randomForest(fate_1~., data=data2, mtry=i, ntree=1000)
  oob.values[i] <- temp.model$err.rate[nrow(temp.model$err.rate),1]
}
## find the optimal value for mtry (gives lowest OOB error)
which(oob.values == min(oob.values))
#2,3 both have min OOB error
#use 3
```

```
#have inbalanced classes (fate groups) so need to tune class weights
```

```
#run model with ntree=1000 and mtry=3 to check class error rates without class weights

m5 <- randomForest(fate_1 ~ ., data2, proximity=T, importance=T, ntree=1000, mtry=3)

m5


#now try different combinations of class weights - assigned to classes in alphabetical order
(M,P,S)

#try to minimize and balance class error rates

m6 <- randomForest(fate_1 ~ ., data2, proximity=T, importance=T, ntree=1000, mtry=3,
classwt=c(2,1,5))

m6

#same as m5


m7 <- randomForest(fate_1 ~ ., data2, proximity=T, importance=T, ntree=1000, mtry=3,
classwt=c(2,1,10))

m7

#better


m8 <- randomForest(fate_1 ~ ., data2, proximity=T, importance=T, ntree=1000, mtry=3,
classwt=c(5,2,20))

m8

#same as m7

m9 <- randomForest(fate_1 ~ ., data2, proximity=T, importance=T, ntree=1000, mtry=3,
classwt=c(5,2,25))

m9

#same as m7


m10 <- randomForest(fate_1 ~ ., data2, proximity=T, importance=T, ntree=1000, mtry=3,
classwt=c(10,1,100))

m10

#same
```

```r
#m7 is best were getting
#use as FINAL MODEL

#ROC and AUC

#make df with number of votes for each fate group by smolt
m_v <- as.data.frame(m7$votes)
#col 1= M, 2=P, 3=S

#plot ROC curve for each class in layers
roc(data2$fate_1, m_v$M, plot=T, legacy.axes=TRUE, percent=TRUE,
    xlab="False Positive Percentage", ylab="True Postive Percentage", col="red",
    lwd=4, print.auc=TRUE)
plot.roc(data2$fate_1, m_v$P, percent=TRUE, col="blue", lwd=4,
        print.auc=TRUE, add=TRUE, print.auc.y=35)
plot.roc(data2$fate_1, m_v$S, percent=TRUE, col="green", lwd=4,
        print.auc=TRUE, add=TRUE, print.auc.y=20)
legend("topleft", legend=c("M", "P", "S"), col=c("red", "blue", "green"), lwd=4, cex=0.6)
#use final model to predict fate class of the suspect smolts
m7predict <- predict(m7, unknown[,-18]) #remove original fate column

#add predicted fate to df
unknown$predict_f <- m7predict

#variable importance plots
varImpPlot(m7)
```

#partial dependence plots for most important variables

#pot probability of belonging to each fate class with change in variable

#layers for each class

partialPlot(m7, data2, total_riverkm, "P", col="blue", main="",

       xlab="Distance travelled (river km)", ylab="class probability", ylim=c(-5.5,4))

partialPlot(m7, data2, total_riverkm, "M", col="red", add=T)

partialPlot(m7, data2, total_riverkm, "S", col="green", add=T)

legend("topright", legend=c("M", "P", "S"), col=c("red", "blue", "green"), lwd=2, cex=0.8)

Appendix D: Example R Code for Chapter 3 Analyses

CJS Example Code

#step 1: turn detection df into presence/absence df

#step 2: make capture history column (series of 1s (detected) and 0s (not detected))

#step 3: add covariates to df with ch

#step 4: process data (process.data(df, model="CJS"))

#step 5: make design data (make.design.data(proc_data))and add more covariates

#(array, water, etc.) (use PIM index)

#step 6: check assumptions/overdispersion by calculating c-hat (release.gof(proc_data))

#if c-hat > 1, you have overdispersion ---> use QAICc

#step 7: set formulas for Phi and p (ex: Phi.dot=list(formula=~1))

#step 8: make models (ex: mark(proc_data, data.ddl, model.parameters = list(Phi=Phidot, p=pdot)))

#step 9: compare models (collect.models())

#step 10*: adjust c-hat for QAIC (adjust.chat(chat=, model.list=))

#step 11*: if AIC/QAIC values are similar for models ---> model averaging (model.average(model.list=))

#*if necessary

#read in data

data <- read.csv("2017 detections fate_rf truncated 70.csv")

#STEP 1:

#remove predated detections (odd Tag ID)

dets <- data[data$TAG_ID_CODE %% 2 == 0,]

#function adapted from Barbosa, 2014
https://modtools.wordpress.com/2013/04/30/splist2presabs/

```r
presabs_fxn <- function(data, sites.col, sp.col, keep.n = FALSE) {

 # version 1.1 (7 May 2013)

 # data: a matrix or data frame with your localities and species (each in a different column)

 # sites.col: the name or index number of the column containing the localities

 # sp.col: the name or index number of the column containing the species names or codes

 # keep.n: logical, whether to get in the resulting table the number of times each species appears
in each locality; if false (the default), only the presence (1) or absence (0) are recorded


 stopifnot(

  length(sites.col) == 1,

  length(sp.col) == 1,

  sites.col != sp.col,

  sites.col %in% 1 : ncol(data) | sites.col %in% names(data),

  sp.col %in% 1 : ncol(data) | sp.col %in% names(data),

  is.logical(keep.n)

 )


 presabs <- table(data[ , c(sp.col, sites.col)])

 presabs <- as.data.frame(unclass(presabs))

 if (!keep.n)  presabs[presabs > 1] <- 1

 presabs <- data.frame(row.names(presabs), presabs)

 names(presabs)[1] <- names(subset(data, select = sp.col))

 rownames(presabs) <- NULL

 return(presabs)

}


#apply pres abs function

#use station (receiver) number as sites.col and tag SN as sp.col

colnames(dets)
```

```
presabs_df <- presabs_fxn(dets, sites.col=20, sp.col=17)

presabs_df <- presabs_df %>% select(-X00VR) #remove mobile

presabs_df <- presabs_df %>% select(-c( X10SH, X11SH)) #remove up confluence


#make 1 col for SH mouth from the 3 recs
  #turning single receivers in same location into station
presabs_df <- presabs_df %>% mutate(SHmouth=X14SH+X15SH+X16SH)

presabs_df$SHmouth <- ifelse(presabs_df$SHmouth > 0, 1, 0)

presabs_df <- presabs_df %>% select(-c(X14SH, X15SH, X16SH))


#STEP 2:

colnames(presabs_df)

#paste all station detections columns into one column - capture history string

presabs_df$ch <- do.call(paste0, presabs_df[2:10])

ch_df <- presabs_df %>% select(ch)


#STEP 3:

#no covariates used here (done in logistic regression) - see Step 6


#STEP 4:

library(RMark) #need to download program MARK first
https://sites.warnercnr.colostate.edu/gwhite/program-mark/


#process data for running through MARK

data_proc <- process.data(ch_df, model="CJS")


#STEP 5:

#if c.hat > 1 -> overdispersion

release.gof(data_proc)
```

```
#c.hat = (Test2 X2 + Test3 X2)/(Test2 df + Test3 df)


#STEP 6:
#create design data file (ddl)
data.ddl <- make.design.data(data_proc)


base_mod <- mark(data_proc, data.ddl)


#gives parameter index, model index, group, time and Time, and two other pairs, cohort and age.
data.ddl$Phi
data.ddl$p


#add covariates to phi and p - array
#use PIMS index
#Gives index for each individual and specific surv interval or recapture occassion
#assign each array its own number


?PIMS
PIMS(base_mod, "Phi", simplified = FALSE) #each col is survival interval
data.ddl$Phi$par.index
data.ddl$Phi$model.index #same as par.index but use model.index


data.ddl$Phi$array[data.ddl$Phi$model.index %in% 8:36] <- 8 #surv from gosse to sh mouth
data.ddl$Phi$array[data.ddl$Phi$model.index %in% 1] <- 1 #from release to 1st rec
data.ddl$Phi$array[data.ddl$Phi$model.index %in% c(2,9)] <- 2
data.ddl$Phi$array[data.ddl$Phi$model.index %in% c(3,10,16)] <- 3
data.ddl$Phi$array[data.ddl$Phi$model.index %in% c(4,11,17,22)] <- 4
data.ddl$Phi$array[data.ddl$Phi$model.index %in% c(5,12,18,23,27)] <- 5
```

```
data.ddl$Phi$array[data.ddl$Phi$model.index %in% c(6,13,19,24,28,31)] <- 6
data.ddl$Phi$array[data.ddl$Phi$model.index %in% c(7,14,20,25,29,32,34)] <- 7


data.ddl$Phi$array <- as.character(data.ddl$Phi$array)


PIMS(base_mod, "p", simplified = FALSE) #each col is recapture occassion
data.ddl$p$model.index


#added to design data but not processed data


#now make parameter to let p vary by each array (think i will have all 4 at SH mouth as 1 array)
data.ddl$p$array[data.ddl$p$model.index %in% 44:72] <- 9 #sh mouth
data.ddl$p$array[data.ddl$p$model.index %in% 37] <- 2 #1st rec
data.ddl$p$array[data.ddl$p$model.index %in% c(38,45)] <- 3
data.ddl$p$array[data.ddl$p$model.index %in% c(39,46,52)] <- 4
data.ddl$p$array[data.ddl$p$model.index %in% c(40,47,53,58)] <- 5
data.ddl$p$array[data.ddl$p$model.index %in% c(41,48,54,59,63)] <- 6
data.ddl$p$array[data.ddl$p$model.index %in% c(42,49,55,60,64,67)] <- 7
data.ddl$p$array[data.ddl$p$model.index %in% c(43,50,56,61,65,68,70)] <- 8


data.ddl$p$array <- as.character(data.ddl$p$array)


#STEP 7:
#define formulas for survival (phi)
Phi.dot=list(formula=~1) #constant survival
Phi.array=list(formula=~array)


#define models for detection (p)
```

```
p.dot=list(formula=~1)

p.array=list(formula=~array)


#STEP 8:

#run models

#have to make models manually with ddl to get covariates for p (not in proc data)

base_mod <- mark(data_proc, data.ddl, model.parameters = list(Phi=Phi.dot, p=p.dot))

mod1 <- mark(data_proc, data.ddl, model.parameters = list(Phi=Phi.dot, p=p.array))

mod2 <- mark(data_proc, data.ddl, model.parameters = list(Phi=Phi.array, p=p.array))


#STEP 9:

CJS_results <-collect.models() #gives AIC summary for model results

CJS_results



##### Multinomial Logistic Regression example code #####


#read in data - metrics

d <- read.csv("all years live fate_rf trunc 70 metrics for MLR.csv")


#make all NAs 0

d[is.na(d)] <- 0


d %>% filter(migration_kmday == 0)

d2 <- d %>% filter(migration_kmday != 0) #remove individuals with no detections n=8


str(d2)

#fate is factor
```

d2$yr <- as.factor(d2$yr) #year as factor

#check if vars colinear

abs(cor(d2[c(5:7,9:13)]))

#W-FL

#arrive-d2epart night

library(mlogit)

?hmftest #Test the IIA hypothesis (independence of irrelevant alternatives) for a multinomial logit model.

#need to prepare data

d2m <- mlogit.data(d2, varying=NULL, choice="fate_rf", shape="wide")

#now run model

ml1 <- mlogit(fate_rf ~ "S" | FL_cm + release_jd2ay + release_site + yr + migration_kmd2ay + num_reversals +

     d2eparted2_night + tot_num_pause,

    data = d2m, reflevel = "S")

library(nnet)

d2$fate_ref <- relevel(d2$fate_rf, ref = "S") #set response to compare others to

#leave out W and2 arrived night for first model

m1r <- multinom(fate_ref ~ FL_cm + release_jday + release_site + yr + migration_kmday + num_reversals +

     departed_night + tot_num_pause,

    data = d2) #set formula

summary(m1r)

#AIC 263.14

#alternative models:

#swap in W and2 arrived2

m2r <- multinom(fate_ref ~ W_g + release_jday + release_site + yr + migration_kmday + num_reversals +

      arrived_night + tot_num_pause,

    data = d2)

summary(m2r)

#264.03

#now swap one at a time

m3r <- multinom(fate_ref ~ FL_cm + release_jday + release_site + yr + migration_kmday + num_reversals +

      arrived_night + tot_num_pause,

    data = d2)

summary(m3r)

#AIC 264.53

m4r <- multinom(fate_ref ~ W_g + release_jday + release_site + yr + migration_kmday + num_reversals +

      departed_night + tot_num_pause,

    data = d2)

summary(m4r)

#AIC 262.88

##m4r lowest AIC

#calculate McFaddens Pseudo R2 - model fit

```
#create null model
m0r <- multinom(fate_ref ~ 1, d2)
m0r_ll <- logLik(m0r)


#for best model
as.numeric(1 - logLik(m4r)/m0r_ll) #0.209


#compare to other models
as.numeric(1 - logLik(m1r)/m0r_ll) #0.208


as.numeric(1 - logLik(m3r)/m0r_ll) #0.203


as.numeric(1 - logLik(m2r)/m0r_ll) #0.205


# AIC stepwise selection - forward and backward
library(MASS)
summary(stepAIC(m4r, direction = "both")) #use best model


#run final mod
mfr <- multinom(formula = fate_ref ~ yr + migration_kmday + num_reversals, data = d2)


#r2
as.numeric(1 - logLik(mfr)/m0r_ll) #0.179


#look at which vars sig
#calc z score
z <- summary(mfr)$coefficients/summary(mfr)$standard.errors
#find p value
```

```r
(1 - pnorm(abs(z), 0, 1))*2 #2-tailed z test, normal dist (Wald test)


#plot effect plots for regression
library(effects)
plot(allEffects(mfr))



##### Example Code for PCA #####
library(dplyr)
library(ggplot2)

#read in data
  #contains normalized gene expression and Plate No. (cols 5,8:56), smolt ID, migration fate
data_g <- read.csv("2019 Transcriptomic Data 2.csv")

#check assumptions of PCA
  #independence - not a problem
  #normal - not essential
  #0s -> correspondence analysis
  #outliers -> remove if error
boxplot(data_g[,c(8:56)])
  #mising values -> delete or interpolate
anyNA(data_g)

#pca with correlation matrix
pca  <- princomp(data_g[,c(5,8:56)], cor=T, scores=T)
summary(pca) #importance of PCs - variance explained
```

```r
#plot first two PCs

plot(pca$scores[,1], pca$scores[,2], col=c("green", "red", "blue")[data$fate_rf],

    xlab="PC1 (18.5%)", ylab="PC2 (13.8%)") #M=green, P=red, S=blue

legend("topright", legend=c("M", "P", "S"), col=c("green", "red", "blue"), lwd=2, cex=0.8)


#save pca scores as data frame

scores_df <- as.data.frame(pca$scores[,1:4])

#add colomn for fate (grouping factor)

scores_df <- cbind(scores_df, data_g %>% dplyr::select(fate_rf))

#save pca loadings as data frame

loads_df <- as.data.frame(pca$loadings[,1:4])

loads_df <- rownames_to_column(loads_df)


#different way to plot

ggplot()+

  geom_point(scores_df, mapping=aes(Comp.1, Comp.2, col=fate_rf))+ #colour scores by fate

  geom_point(loads_df, mapping=aes(Comp.1, Comp.2, shape=gene_cat)) #shape loadings by
gene type

  xlab("PC1 (18.5%)")+

  ylab("PC2 (13.8%)")+

  labs(col="fate", shape="gene type")



##### run k-means clustering on gene data and PCA results ####

library(adegenet)

library(tidyverse)


?find.clusters
```

clust_4pc <- find.clusters(data_g[,c(5,8:56)], #gene data matrix

                   stat="WSS", #choose number of K clusters based within sum of squares

                   n.pca = 4, #4 PCs retained

                   choose.n.clust = F, #auto choose clusters

                   criterion = "diffNgroup") #The retained K is the one before the first group switch from sharp to mild decreases in WSS

#4 clusters were chosen

#save results as data frame

clusters <- as.data.frame(clust_4pc$grp)

clusters <- rownames_to_column(clusters)

clusters$rowname <- as.numeric(clusters$rowname)

#join to PCA scores data frame

scores_df <- left_join(scores_df, clusters %>% rename(tag_sn=rowname, clust_4pc=`clust_4pc$grp`))

#plot

ggplot()+

  geom_point(scores_df, mapping=aes(Comp.1, Comp.2, col=fate_rf), size=2)+

  labs(x="PC1 (18.5%)", y="PC2 (13.8%)", col="fate", shape="gene type", fill="cluster")+

  geom_point(lds, mapping=aes(Comp.1, Comp.2, shape=gene_cat), size=2)+

  stat_ellipse(scores_df, mapping=aes(Comp.1, Comp.2, fill=clust_4pc), geom="polygon", alpha=0.1, type="norm")+

  theme_bw()