EVALUATING THE RELIABILITY AND EQUITABILITY OF AT-SEA OBSERVER RELEASE REPORTS IN THE B.C. OFFSHORE GROUNDFISH TRAWL FISHERY

by

Matthew H. Grinnell
B.Sc., University of Victoria, 2001

Research Project submitted in partial fulfillment
of the requirements for the degree of
Master of Resource Management (Planning)
In the School of
Resource and Environmental Management

Project No. 479

SIMON FRASER UNIVERSITY
Spring 2010

APPROVAL

Name:	Matthew H. Grinnell
Degree:	Master of Resource Management (Planning)
Title of Research Project:	Evaluating the reliability and equitability of at-sea observer release reports in the B.C. offshore groundfish trawl fishery
Project Number:	479

Examining Committee:

Adam R. Batty
Chair
Master of Resource Management Candidate

Dr. Sean P. Cox
Senior Supervisor
Associate Professor
School of Resource and Environmental Management

Dr. Andrew B. Cooper
Supervisor
Associate Professor
School of Resource and Environmental Management

Richard D. Stanley
Supervisor
Research Biologist
Fisheries and Oceans Canada

Date Approved: ______________________________

Abstract

At-sea observer release reports must be reliable for stock assessment accuracy, and equitable for individual transferable quota management programme validity. However, reliability and equitability may be compromised when harvesters benefit economically from under-reported releases. For example, harvesters in the British Columbia offshore groundfish trawl benefit from under-reported marketable released sablefish and dead released halibut. When monitoring programmes provide essential data for management, a review of the programme's veracity is required.

In this analysis, releases are predicted using environmental, social and economic predictors, and then compared with reported releases. Compared to the average individual, some observers report more- or less-than-expected releases, and some skippers have more- or less-than-expected releases deducted from quota. However, these weights appear to be negligible for both species. The analysis does not provide strong reasons to suspect that release data are unreliable or inequitable for their intended purpose.

Keywords: at-sea observer programme (ASOP) · British Columbia offshore groundfish trawl (BC trawl) · classification and regression tree (CART) · discard · equitable · individual transferable quota (ITQ) · linear mixed effect (LME) · misreport · Pacific halibut (*Hippoglossus stenolepis*) · random forest (RF) · release · reliable · sablefish (*Anoplopoma fimbria*)

To my family

The Baron had fallen to the bottom of a deep lake. Just when it looked like all was lost, he thought to pick himself up by his own bootstraps.

— *The adventures of Baron Munchausen*
Rudolph Erich Raspe (1737-1794)

Acknowledgements

I am very thankful to my colleagues, friends, and family for their help and support.

First, I'd like to thank my supervisory committee for guiding me in developing the concept and methodology for the project, interpreting results, and navigating the many writing stages. I also thank my colleagues at Fisheries and Oceans Canada (DFO). Thanks to Norm Olsen and Kate Rutherford for their help extracting data for the analysis. Thanks to Barry Ackerman, Murray Gilchrist, Shelly Joslin, and Frank Snelgrove for their support. Thanks also to Scott Buchanan and Howard McElderry at Archipelago Marine Research Ltd for their suggestions and interpretations. Two people generously provided code for the analysis: Joseph Sexton's code quantifies the uncertainty in univariate marginal dependence, and Adele Cutler's code describes bivariate marginal dependence relationships. An influential conversation with Martin Hall early in the project influenced the methodology with his suggestion to investigate the random forest analysis. The British Columbia trawl harvesters and the at-sea observers both work in challenging conditions, and I thank them for their hard work collecting data.

Thanks to my friends and fellow students in the School of Resource and Environmental Management (REM) for their help and support within REM, and good memories outside. A big thanks to the fishies, especially those in my cohort, for their support. A special thanks to Jessica Finney for feedback on drafts, and endless support all around. Thanks to Cameron MacKenzie for being a great office mate and useful edits. I also thank Aaron Springford for setting up a machine to run the analysis, and, along with Jonathan Martin, for maintaining a stimulating office atmosphere. I thank the REM office staff for their support.

Financial support was provided by DFO's Conservation and Protection unit, National Science and Engineering Research Council of Canada (NSERC) Postgraduate Scholarship, an NSERC Discovery Grant to Sean Cox, Simon Fraser University's (SFU) Graduate Fellowship and SFU's Faculty of Applied Sciences.

I especially thank my family for supporting me financially, emotionally, and in my pursuit of watery distractions during my many years as a student. Despite being so long away from home, the Island never felt out of reach.

Contents

Approval ii

Abstract iii

Dedication iv

Quotation v

Acknowledgements vi

Contents viii

List of Tables xi

List of Figures xii

Glossary xiii

1 Introduction 1

1.1 Impacts of releases 2

1.1.1 Effects of individual transferable quotas on releases 4

1.1.2 Economic incentives to reduce releases 6

1.2 At-sea observer programmes 7

1.3 BC trawl management and releases 8

1.4 Sablefish and halibut releases in the BC trawl 11

1.5 Research goals 15

1.6 Analytical methods to assess releases 17

2 Methods 19
2.1 BC trawl at-sea observer data . . . 20
2.2 Classification and regression trees . . . 26
2.3 Random forest analysis . . . 27
2.3.1 Predictor importance and marginal dependence . . . 28
2.3.2 Random forest event-specific residuals . . . 30
2.4 Linear mixed effects model specification . . . 30
2.4.1 Observer reliability index . . . 32
2.4.2 Skipper equitability index . . . 32
2.4.3 Yearly and total over- or under-reported releases . . . 33
2.5 Proportion dead released halibut . . . 34
2.6 Observer reliability index correlation . . . 34
2.7 Computational details . . . 35

3 Results 36
3.1 Sablefish regression tree . . . 36
3.2 Random forest prediction accuracy . . . 38
3.3 Relative predictor importance . . . 39
3.4 Predictor marginal dependence . . . 40
3.5 Predictor pair correlation . . . 43
3.6 Observer reliability index . . . 44
3.7 Skipper equitability index . . . 47
3.8 Yearly and total misreported releases . . . 50
3.9 Observer reliability index correlation . . . 50

4 Discussion 54
4.1 Reliability and equitability status . . . 55
4.1.1 Observer reliability . . . 56
4.1.2 Magnitude of yearly and total misreported release weight . . . 57
4.1.3 Skipper equitability . . . 58
4.2 Predictors affecting release rates . . . 59
4.2.1 Predictors affecting marketable released sablefish . . . 61
4.2.2 Predictors affecting dead released halibut . . . 62
4.3 Assumptions for assessing observer reliability . . . 63

4.3.1 Alternative analytical approaches 63
4.4 Management considerations and implications 64
4.4.1 Pathological indicators of reliability 66
4.4.2 Alternative applications for random forest 67
4.4.3 Management actions to improve release reliability 67
4.5 Extensions to the current analysis 70
4.5.1 Improving release rate prediction accuracy 71
4.5.2 Using decision analysis to assess management actions 72

5 Conclusion **73**

A Supplemental Material **75**
A.1 Halibut condition factors 75
A.2 Data analysis and summaries 75
A.3 Sablefish random forest sensitivity analysis 77
A.4 Random forest predictions 82
A.5 Bivariate marginal dependence 87

References **90**

List of Tables

1.1 Catch and quota flexibility mechanisms 6
1.2 Sablefish and halibut marketability and mortality rates 14

2.1 Available BC trawl observer data 21
2.2 Predictor variables for the sablefish and halibut analyses 22
2.3 Random forest baseline parameter values and constraints 29

3.1 Random forest tuning parameters and prediction accuracy 39
3.2 Sablefish and halibut predictor pair correlation 44
3.3 Number of observer reliability and skipper equitability indices 45
3.4 Per-skipper over- or under-reported weight of sablefish and halibut 49
3.5 Yearly and total over- or under-reported weight of sablefish and halibut . . . 51

4.1 Inferred reliability status and error types 65

A.1 Halibut conditions, indicators, and mortality rates 76
A.2 Summary of marketable released sablefish and continuous predictors 78
A.3 Summary of categorical sablefish predictors 79
A.4 Summary of dead released halibut and continuous predictors 80
A.5 Summary of categorical halibut predictors 81
A.6 Number of trees and random forest prediction accuracy 82

List of Figures

1.1 BC trawl quota species management areas 9
1.2 BC trawl catch, sorting and proportion estimates 12
1.3 Yearly sablefish catch, halibut by-catch, and quotas 13
1.4 Sablefish mortality and fishing event duration 14

2.1 Flow chart representing the two-step methodology 20
2.2 Proportion marketable released sablefish and dead released halibut 26

3.1 Example sablefish regression tree dendrogram 37
3.2 Sablefish relative predictor importance 40
3.3 Halibut relative predictor importance 41
3.4 Sablefish predictor marginal dependence 42
3.5 Halibut predictor marginal dependence 43
3.6 Sablefish observer reliability indices 46
3.7 Halibut observer reliability indices 48
3.8 Median bootstrapped sablefish and halibut observer effects 53

4.1 Release tax rates and weight released 69

A.1 Sablefish random forest prediction accuracy sensitivity analysis 83
A.2 Sablefish random forest predictor importance sensitivity analysis 84
A.3 Reported and median predicted marketable released sablefish 85
A.4 Reported and median predicted dead released halibut 86
A.5 Sablefish bivariate marginal dependence for *totalCatch* and *totalRetCatch* . . 88
A.6 Halibut bivariate marginal dependence for *totalCatch* and *totalRetCatch* . . . 89

Glossary

Acronym	Meaning
ASOP	At-sea observer program
BC trawl	British Columbia offshore groundfish trawl fishery
CART	Classification and regression trees
DFO	Fisheries and Oceans Canada
DR	Dead released
ITQ	Individual transferable quota
LME	Linear mixed effects
MR	Marketable released
MSE	Mean squared error
NA	Not applicable
OOB	Out-of-bag
PE	Percent error
PVE	Percent variance explained
RF	Random forest
SSQ	Sum of squares
TAB	Total allowable bycatch
TAC	Total allowable catch
TR	Total released
US trawl	United States West coast groundfish trawl fishery

Chapter 1

Introduction

Worldwide, commercial fisheries releases (i.e., discards) warrant serious consideration (Hall and Mainprize, 2005; Matsuoka, 1999; Myers *et al.*, 1997; Rochet *et al.*, 2002; Squires *et al.*, 1998) because they create problems for sustainable fisheries management (Kennelly, 1995; Saila, 1983). Historically, estimated worldwide fisheries releases have been up to 20 million metric tonnes per year (t/yr), or approximately 25% of the total catch (Food and Agriculture Organization, 1997), but have declined to approximately 8% in recent years (Kelleher, 2005). Release-induced mortality rates, which contribute to total extraction, can be high and vary depending on fishing gear, species and other factors (Berghahn *et al.*, 1992; Pascoe, 1997; Saila, 1983). In some cases, under-estimating total extraction can lead to over-exploitation, over-estimates of stock biomass, and reduced stock abundance over time (Hilborn and Walters, 1992).

International agreements and domestic management plans require that Canadian fisheries account for releases. At the international level, Canada signed the Food and Agriculture Organization's *Code of Conduct for Responsible Fisheries* in 1995, indicating that States should work with industry to promote fishing methods that reduce releases and release-induced mortality rates (Food and Agriculture Organization, 1995). Domestically, Fisheries and Oceans Canada (DFO) Pacific Region details the rules and policies for managing the British Columbia offshore groundfish trawl fishery (BC trawl), and indicates *inter alia* the requirement of individual harvester accountability for targeted species and by-catch, as well as the importance of complete at-sea fleet monitoring (Fisheries and Oceans Canada, 2005a). Accurate release estimates for all gear types, typically obtained from at-sea observer program (ASOP) accounts, are essential to determine total extraction and set total allowable catches (TAC) to biologically sustainable and economically efficient levels (Grafton *et al.*, 2003; Hammond and Trenkel, 2005; Kelleher, 2005; Rochet *et al.*, 2002). However, testing the accuracy of ASOP data is challenging because there is no other account of fishing activity to compare with ASOP reports.

ASOP release reports generally have a negative affect on harvester profitability in the BC trawl. Thus, economic incentives may cause skippers to pressure observers into under-reporting releases in some situations, resulting in under-reported total extraction. Further, the economic benefits associated with under-reported releases could be distributed inequitably between skippers.

This project assesses BC trawl ASOP release estimate reliability and equitability using a two-step procedure. First, I predict expected release rates on individual fishing events using environmental, social and economic predictors. In the second step I compare predicted and reported release rates to assess potential tendencies to over- or under-report, which are then used to quantify observer reliability and skipper equitability. The intent is to evaluate the veracity of the BC trawl ASOP, and in doing so, to consider which management changes might improve data utility, if required.

1.1 Impacts of releases

Releases, defined as the proportion of catch weight brought on-board vessels that is returned to the sea, include fish that are above (marketable) as well as below (unmarketable) legal size limits (Alverson *et al.*, 1994; Vestergaard, 1996), and those established for management purposes. The release problem can be categorized in terms of its environmental and economic impacts (Catchpole *et al.*, 2005; Pascoe, 1997). Negative environmental impacts include decreased fish stocks resulting from release-induced mortality. Positive environmental impacts may include providing food for species of high economic value, or those at risk of extinction (Zhou, 2008). However, an additional food source could have negative effects, such as feeding undesirable species (e.g., sea lice) or feeding non-target species that prey on commercial stocks. Reliable estimates of release-induced mortality must be available for stock assessments, regardless of whether releases have positive or negative environmental impacts. Economic impacts of high release rates include foregone income, costs to other harvesters and fisheries, and costs to collect reliable release estimates. Conversely, enforcing low releases can constrain fishing fleets, and may be economically inefficient. In either case, the economic costs or benefits of releases must be accounted for, and distributed equitably among harvesters. For example, a management programme that penalizes harvesters for releasing fish may be unfair if some harvesters release fish without penalty.

Harvesters typically release marketable fish only when size- or quota-based incentives are present. Size-based incentives include high-grading, whereby profit maximizing harvesters release less valuable, but potentially marketable, pieces to conserve quota or vessel hold space for higher value pieces. Such incentives typically exist when there are price premiums for pieces of a preferred size, usually but not necessarily larger pieces. Quota-based incentives include conserving quota for restrictive species in order to continue fishing for other species. For example, catch rates for restrictive species may be higher than available quota. Restrictive species, which may be present in multi-species fisheries, are species caught in disproportionate amount to their quota. By-catch (i.e., non-target species) can be either retained for market, or released when there is no market for the product or the species is prohibited.

Releases impact individual harvesters, other harvesters in the fishery, and even other fisheries. Harvesters who release fish balance private economic benefits against both private and public costs. For example, harvesters who receive extra profits by high-grading may also incur costs via extra crew and increased handling times (Pascoe, 1997). While high-grading may be economically efficient for an individual harvester, high-grading can be an inefficient use of the resource due to release-induced mortality (Squires *et al.*, 1998), which may impose costs on other harvesters. These imposed costs, known as negative externalities, describe costs imposed on individuals not directly involved in economic transactions (Ward, 2006). In the high-grading example above, negative externalities include decreased quota share weight and value for all harvesters in the fishery due to TAC reductions as well as increased management costs required to estimate releases and release-induced mortality rates.

Negative externalities can also occur between fisheries when by-catch of one sector is the target species for another sector, resulting in negative revenue impacts from by-catch release mortality (Pascoe, 1997; Terry, 1998). For example, each metric tonne of halibut by-catch in the Alaska groundfish trawl fishery causes the Alaska fixed-gear halibut fishery to lose about 1.8 t of yield, worth an estimated net benefit of approximately US$ 2,400 based on discounted quota value (Terry, 1998).

1.1.1 Effects of individual transferable quotas on releases

Managing fisheries by individual transferable quota (ITQ) programmes is an increasingly popular mechanism used to divide up the TAC among harvesters (e.g., percentages, Sanchirico *et al.* 2006). However, such rights may have unintended effects on harvester incentives to release fish. Designed to address economic problems (Hannesson, 1996; Ward, 2006), ITQs grant harvesting rights to individuals who can harvest fish, or transfer the quota to other licence holders. Equitable harvester treatment may be necessary for ITQ management program success (Branch, 2004; Stanley *et al.*, 2009). The following section gives a brief overview of some impacts of ITQs on harvester release incentives; however the issue is complex and the reader is referred to Branch (2004) for a more comprehensive discussion.

ITQs may offer solutions to the release problem. Potential linkages between fishery profitability and quota value may encourage a resource husbandry ethic among ITQ holders because economic incentives may favour fishery sustainability (Branch, 2009; Branch and Hilborn, 2008; Costello *et al.*, 2008; Hilborn *et al.*, 2004; McCay *et al.*, 1995; Sanchirico *et al.*, 2006). Economic incentives, such as increased harvester quota shares, places transferable value on the fish population itself, which may increase harvester support to enhance or rebuild depleted stocks as well as to invest in stock assessment research (Beddington *et al.*, 2007). Hilborn *et al.* (2004) describe several examples of apparent fisheries successes following ITQ implementation. For example, the Canadian Sablefish Association, an organization of BC fixed-gear sablefish quota holders, funds stock assessments and develops technology to reduce juvenile sablefish mortality (Hilborn *et al.*, 2004). These sablefish harvesters, presumably acting in their own self-interest in this competitive fishery, are cooperating to fund programs and adopt new technologies which benefit the fishery. These conservation initiatives may only apply to target species, and by-catch regulations (e.g., quotas) may still be required to reduce non-target species capture (Hilborn *et al.*, 2004).

However, the time required to rebuild a depleted fish stock can be long, and economic incentives to rebuild resources are reduced when rebuilding times are longer than a harvester's career, or harvesters do not own quota. For example, quota owned by retired harvesters may be leased to active harvesters (Branch, 2004) when high quota sale value prevents new harvesters from buying quota and high lease costs cause harvesters who lease quota to have a marginally profitable business (Pinkerton and Edwards, 2009).

Although multispecies ITQ programmes can increase release incentives (Branch, 2009;

Squires *et al.*, 1998; Turner, 1997), these same programmes can increase incentives to reduce releases. For example, the release problem can be reduced by either applying quota to total extraction, encouraging harvesters to fish more selectively, or both. Thus actual releases may increase or decrease depending on the specific fishery (Pascoe, 1997), the associated benefits and costs to the harvester who releases fish, and whether quota is applied to landed catch or total extraction (Arnason, 1994).

Because ITQs can increase release incentives, species-specific catchability, productivity (Branch, 2004) and market forces must be considered when setting TACs. Multispecies ITQ programmes motivate harvesters to closely match individual species catch to quota holdings over the fishing year to avoid paying high quota lease prices for species with restrictive quota (Eythorsson, 2000), sometimes called "choke-point" species. For example, vessels that run out of quota for restrictive species (e.g., threatened rockfish) may be forced to cease fishing, resulting in a shorter fishing season and unfished quotas for other species. Skippers may alter fishing behaviour to match catch and quota. For example, skippers may target areas where they expect low abundance of restrictive species, conduct short sets to ascertain species compositions before conducting longer sets (Branch and Hilborn, 2008) and release unwanted fish (Squires *et al.*, 1998).

Catches are also important on a per-trip basis because skippers may begin trips with species-specific lists of desired weights corresponding to favourable market prices or processing plant requirements. Mismatches between encountered and desired species compositions could increase release incentives. Quota pressures can cause harvesters to release larger-than-usual fish when catches approach quotas (Stratoudakis *et al.*, 1998), possibly due to high-grading. For example, ITQs can increase economic incentives for high-grading when larger fish have higher per-weight value (Anderson, 1994; Arnason, 1994; Dewees, 1998; Turner, 1997; Vestergaard, 1996).

Because catches and quotas will not match up exactly in multispecies fisheries based on behavioural changes alone (Sanchirico *et al.*, 2006), harvesters require flexibility mechanisms to balance catches and quotas (Table 1.1). Setting the level of flexibility is a balance between meeting social and economic objectives, reasonably limiting the risk of over-exploitation, and operating within reasonable programme costs (Sanchirico *et al.*, 2006; Squires and Kirkley, 1995). Harvesters who use flexibility mechanisms to correct catch and quota imbalances do so under the assumption that all harvesters are treated equitably (Stanley *et al.*, 2009). If a harvester is able to secretly release quota-limiting species instead of acquiring quota through

Table 1.1: Flexibility mechanisms that facilitate a harvester's ability to match catches and quotas in multispecies individual transferable quota fisheries. Note: adapted from Sanchirico *et al.* (2006).

Mechanism	Definition
Releases	Fish that are not retained for market
Quota markets	Transfer annual quota within one fishing season (e.g., leasing quota) or in perpetuity
Retrospective balancing	Period of time allotted to match catch and quota
Roll-over allowances	Ability to carry-forward unused annual quota to the next fishing year, or borrow a portion of next year's annual quota
Species equivalence	Ability to convert annual quota of one species into annual quota of another species at a pre-specified rate

flexibility mechanisms, the quota trading system may break down. Equitable harvester treatment may also be required with respect to species subject to high-grading. However, the economic impacts of inequitable harvester treatment for quota-limited species may be greater than the value of the species itself. For example, if quota-limiting species have the ability to shorten a harvester's fishing season, or restrict the capture of other species taken in association, then the value of the quota-limited species may actually reflect the opportunity to extend the fishing season or capture other species.

1.1.2 Economic incentives to reduce releases

Some fisheries management programs have implemented economic methods to reduce releases by increasing release costs. For example, Arnason (1994) proposed to solve the release problem by taking externalities into account via a Pigouvian tax, which is a market-based incentive that reduces negative externalities by taxing resource use (Ward, 2006). A Pigouvian tax internalizes a resource's negative externalities by increasing the user's private costs to reduce resource use; thus decreasing imposed public costs (Ward, 2006). For example, a release tax reduces net economic high-grading benefits by increasing marginal private release costs (Arnason, 1994), thereby reducing negative externalities imposed on other harvesters

(Pascoe, 1997; Ward, 2006). For by-catch species and species not subject to high-grading, a tax may increase the incentive to alter fishing behaviour. However, setting an appropriate tax rate can be complicated due to difficulties in quantifying a resource's negative externalities (Ward, 2006).

Some harvesters faced with economic penalties for releasing fish may have incentives to under-reported releases, and the tax may be economically inequitable if some harvesters are able to under-report. Thus, release reduction initiatives must: (1) create penalties for releasing fish; and (2) monitor compliance via observer programmes (Kelleher, 2005; McCay *et al.*, 1995; Squires *et al.*, 1998). Setting an appropriate release tax rate is a trade-off between various goals including appropriate release weights, reliable and equitable release estimates, compliance monitoring costs, and economic efficiency. For example, accurate release reports, which may improve extraction estimates, may be more important than reduced release weights. There is also a distinction between eliminating and reducing releases; eliminating releases may greatly reduce fishing revenues (Pascoe, 1997). When taxes are implemented correctly and accurate release reports are available, release rates may be reduced from individually, to socially optimal levels (Pascoe, 1997).

1.2 At-sea observer programmes

ASOPs are initiated by either government or industry, and are used to provide an unbiased account of at-sea fishing activity. In practice, observers are placed on-board active commercial fishing vessels. Some ASOPs provide partial fleet coverage (i.e., random 20% of trips or catch), and some programs provide complete coverage. ASOPs can be funded by government and/or industry; industry funding is typically obtained from vessels via either per-day-fished fees or yearly fishing licence fees. Observers monitor fishing activity and report regulation infractions, but are not typically mandated to enforce regulations. Two critical assumptions underlying the use of ASOP data in fisheries assessment and ITQ management are that: (1) release reports are reliable for estimating total extraction; and (2) harvester treatment is equitable for catch and quota balancing. Some situations may compromise these two assumptions, even when ASOPs provide complete coverage (Kelleher, 2005).

Harvester pressure may affect reliability when observer reports affect vessel profitability.

For instance, observer willingness to accurately report catches may be compromised if harvesters are hostile, or if catch monitoring duties are obstructed (Kelleher, 2005). Observers must have adequate access to the catch as it is sorted, but this could be compromised at night of if catches are released before the observer has access. Many vessels operate 24 hours per day and observers may miss events when sleeping. Observers who are not on deck to monitor fishing activity obtain catch estimates from skipper log-books and by talking with the crew.

When harvester pressure is present, elements of human nature may cause observers to under-report releases; however both over- and under-estimating releases can have negative consequences on fisheries (Kelleher, 2005). Release over-estimation can restrict fishing and result in forgone economic opportunities and yield (Kelleher, 2005); the consequences of under-estimating releases have already been discussed.

1.3 BC trawl management and releases

Each year, approximately 60 active BC trawl vessels (licence option A) land about 140,000 t of fish valued at approximately C$ 65 million (Fisheries and Oceans Canada, 2009). These vessels are allowed to bottom-trawl in all management areas except 4B, and midwater-trawl in all eight areas (Figure 1.1)[1]. In the 2005/2006 fishing year, the ITQ management programme accounted for 29 targeted species including halibut; because some species have area-specific quotas, there were 60 species-area groups requiring TACs and monitoring. The ITQ establishes TACs for most directed species, but does not include many additional non-directed species. That same year, approximately 74% of the total fleet-wide quota was extracted. Harvesters are financially accountable, through quota shares, for marketable quota species (both retained and marketable released, MR) and dead released (DR) halibut in each species management area. The BC trawl ITQ management programme contains flexibility mechanisms given in Table 1.1 to enable harvesters to match catches and quotas.

In 1978, BC trawl management changed from trip limits to annual quotas. The DFO implemented licence limitations, TACs, closures and trip limits in 1979. Throughout the 1980s, the DFO managed more species with quotas and introduced partial ASOP coverage in 1987 and dockside monitoring programme of all landed catch in 1995. Due to concerns

[1]There is a small amount of bottom-trawling in area 4B by inshore trawl vessels (licence option B). These vessels provide about 10% ASOP monitoring.

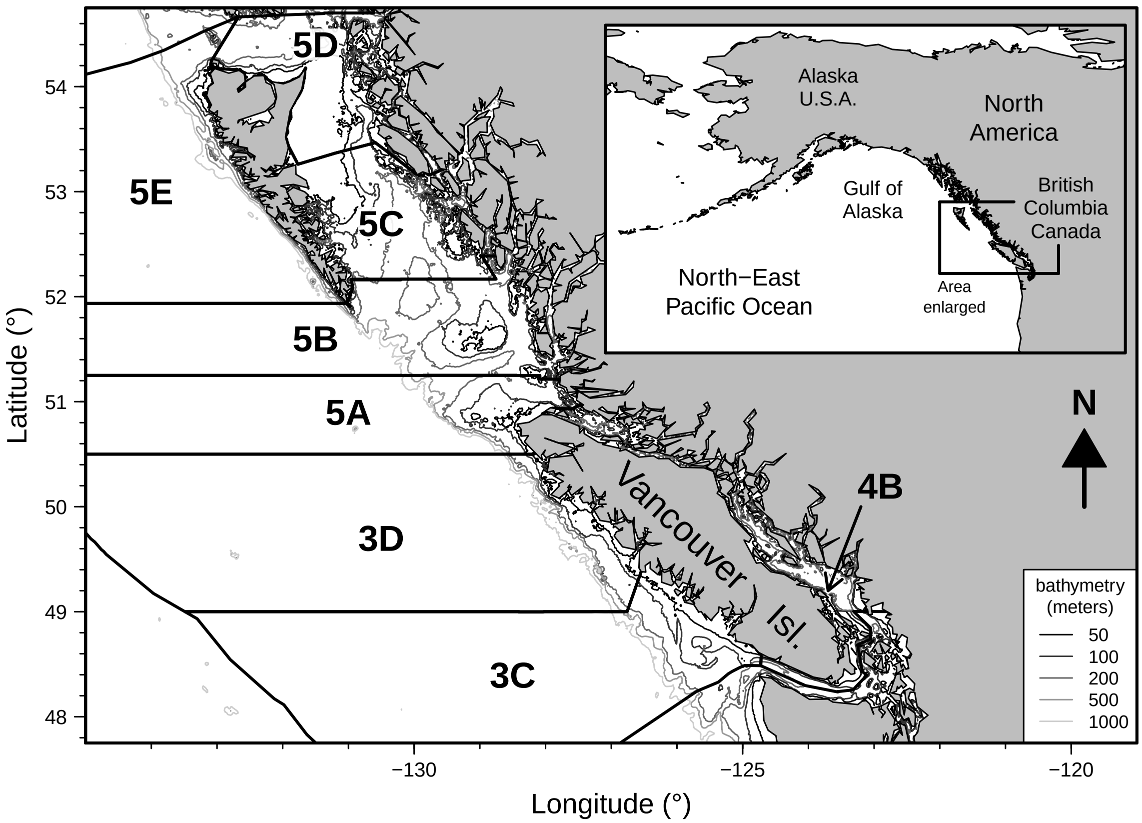

Figure 1.1: Fisheries and Oceans Canada (2005a) areas used to manage area-species quotas for the British Columbia offshore groundfish trawl fishery.

about potential impacts of halibut trawl by-catch on the directed halibut fishery, Canada and the US agreed to reduce the halibut total allowable by-catch (TAB) in 1991, and further reduced the halibut TAB in 1995. Halibut retention is prohibited from BC trawl vessels but allowed by the directed hook and line fishery[2]. That same year, the BC trawl closed mid-season due to concerns that TACs had been exceeded. When the fishery opened in 1996, vessels required 100% ASOP and dockside monitoring program coverage. Complete ASOP monitoring is essential for stock assessment, long-term sustainability, encourages harvesters to fish responsibly, supports Canada's international obligations (Fisheries and

[2]BC trawl vessels are prohibited from retaining halibut due to concerns that they catch fish that are smaller than the size that produces optimal yield (Myhre, 1969). For example, 78% (by weight) of halibut by-catch in the BC trawl is smaller than the directed fishery's marketable size (Stanley, 1984).

Oceans Canada, 2005a), provides reliable spatially explicit estimates of landed and released catch, as well as other benefits which include biological sampling.

BC trawl release rates used to be quite high but have decreased over time due partly to management programme changes. Prior to ITQ implementation, when the BC trawl had partial observer coverage, approximately 40% of summer trawl catch in BC's Northern coast (i.e., areas 5A, B, C, and D) was released (Stanley, 1985). Considering these statistics, Stanley (1985) suggested that "the impact of bottom-trawling on the fisheries ecology of BC waters is seriously under-represented through the use of landing statistics alone" or harvester logbook release reports. For example, Stanley (1985) estimated sablefish releases to be approximately 1,000 t/yr in 1981 and 1982 for these areas, compared to logbook reports of approximately 150 t/y. Similarly, Stanley (1984) estimated trawl releases of halibut by-catch for these areas to be approximately 720 t in 1981/1982. Extrapolated to the entire BC coast, halibut by-catch could have been approximately 1,180 t, equivalent to approximately 25% of each year's directed halibut hook and line quota (Stanley, 1984).

The BC and United States West Coast groundfish trawl fishery (US trawl) have similar target species, gear and markets, however the BC trawl is managed by ITQs while the US trawl is managed by bimonthly landing limits. BC trawl release rates, which were as high as US trawl release rates prior to ITQ implementation, are currently much reduced (Branch *et al.*, 2006; Stanley, 1985). For example, Branch *et al.* (2006) estimate BC trawl releases at 14 to 19% of the total catch, compared with 31 to 43% in the US trawl[3]. The BC trawl has lower release rates despite predictions that ITQ programs increase release incentives, likely due to 100% ASOP and dockside monitoring coverage, and economic penalties for MR dead fish (Branch, 2004).

A mandatory independent ASOP covers 100% of BC trawl trips, which amounts to approximately 5,000 days each year[4]. ASOP costs are about C$ 560 per day, and are shared between the vessel (70%) and the DFO (30%). Trips range in duration from less than one day to several weeks; observers live and work with the crew in close quarters for the duration of the trip, monitoring various aspects of fishing activity. According to the ASOP briefing

[3]US trawl release estimates may be under-estimated because the partial coverage ASOP monitors less than 20% of the landed weight. When ASOPs provide partial coverage, release activity and reports on observed trips may under-represent release activity on unmonitored trips (i.e., when the at-sea observer is absent), due to changes in harvester behaviour (Benoit and Allard, 2009; Harris, 1998).

[4]An observer is required on every trip, but events are occasionally unmonitored (e.g., when the observer sleeps).

workbook, an observer how-to manual, estimating halibut mortality is one of the ASOP's primary objectives. Other ASOP duties include estimating catch composition (critical to assigning quota to species-areas), quota species weight and marketability (retained and released), collecting biological data, recovering mark-recapture tags and recording event location, depth and time. Observers disembark after the trip and are assigned to another vessel. Due to the relatively small number of vessels and observers, it is not uncommon for observers to be assigned to vessels more than once. Relatively high ASOP turn-over rates provide a consistent influx of observers.

The ASOP is critical to achieving conservation and economic goals in the BC trawl, and the validity of the ITQ programme. For the BC trawl to maintain its credibility as an accountable fishery with conservation groups, other fisheries and the public, observers must act in a professional manner and be perceived to report releases accurately. It is also critical that observers treat harvesters equitably for issues that affect profitability.

1.4 Sablefish and halibut releases in the BC trawl

The intent of this project is to investigate the reliability and equitability of ASOP release reports in the BC trawl. Concerns about release reliability and equitability are due to economic benefits to harvesters from under-reported releases. However, the BC trawl has quota for 29 species, but not all of these species have the same incentives for harvesters to under-report releases. Personal experience as an observer, and discussion with the DFO and ASOP service provider staff indicate that two species are likely candidates of biased reporting, if it exists in the ASOP. These two species are invested so that results can be compared between species.

Sablefish (*Anoplopoma fimbria*) and halibut (*Hippoglossus stenolepis*) are both caught incidentally by bottom-trawl vessels targeting benthic species such as flatfish. Harvesters in the BC trawl must account for MR sablefish mortality and DR halibut via a release tax implemented by the DFO to reduce MR sablefish and DR halibut. Operationally, the reported MR dead and DR weight is deducted from the harvester's quota at the end of the trip. Observers estimate total catch and proportions sorted by the crew using various methods including direct whole weight measurements, visual estimates or extrapolating measured sub-samples to the entire catch (Figure 1.2). Catch size, catch composition, sorting procedures and mortality rates are affected by various factors including environmental conditions,

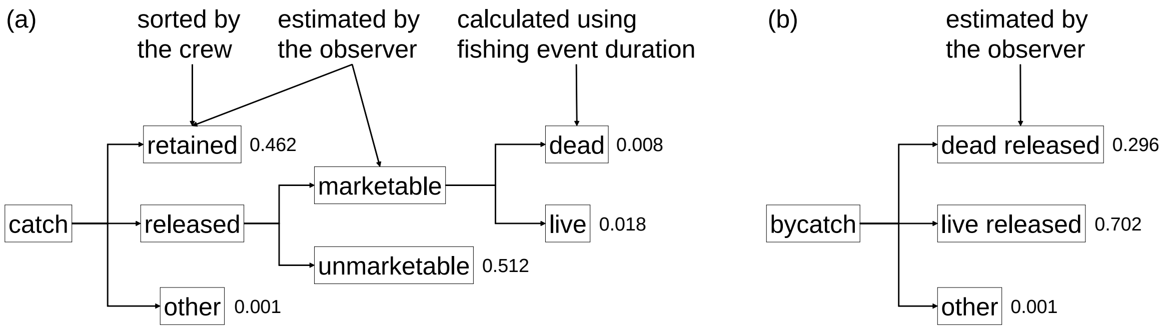

Figure 1.2: BC trawlers sablefish catch (a) and halibut by-catch (b). Sablefish is sorted by the crew into proportions retained and released. For sablefish, observers estimate the weight released, and the proportion of releases that are marketable and unmarketable. Length determines marketability, while fishing event duration determines mortality. All halibut are released, and observers estimate dead and live weights based on fish condition. Numbers indicate proportions with respect to total catch or total by-catch weight for years 1997/1998 through 2005/2006, excluding 2000/2001. *Other* refers to infrequent utilization codes.

economic incentives, legal constraints and harvester ability (Rochet and Trenkel, 2005).

Harvesters typically retain sablefish because of it's high value, but do not target sablefish because of quota constraints. Harvesters who exceed sablefish quota by 30% are restricted from bottom-trawling until they acquire sufficient quota to cover the overage. Individual harvesters are limited to a maximum of 5% (and up to 7% temporarily) of the annual fleet-wide TAC, which varies annually (Fisheries and Oceans Canada, 2005a). Although individual harvesters are limited to 5% of the TAC, the fleet-wide TAC still applies. Annual BC trawl sablefish extraction ranges between 63 and 107% of the TAC (median = 82%, Figure 1.3a). Harvesters are limited by a maximum roll-over allowance of 30% of their sablefish quota (Fisheries and Oceans Canada, 2005a). This flexibility mechanism enables harvesters to carry unused quota forward to the next fishing year, or attribute catch to the next year's quota.

Marketable sablefish (i.e., longer than the minimum legislated length, Table 1.2) may be selectively released because per-weight sablefish value can be more than double for large pieces, which creates a high-grading incentive (Gillis *et al.*, 1995). To conserve sablefish quota, harvesters actively avoid sablefish catch, and areas of high sablefish density throughout the year. However, marketable sablefish may be released early in the year if harvesters believe they will exceed their yearly sablefish quota, or later in the year if the remaining

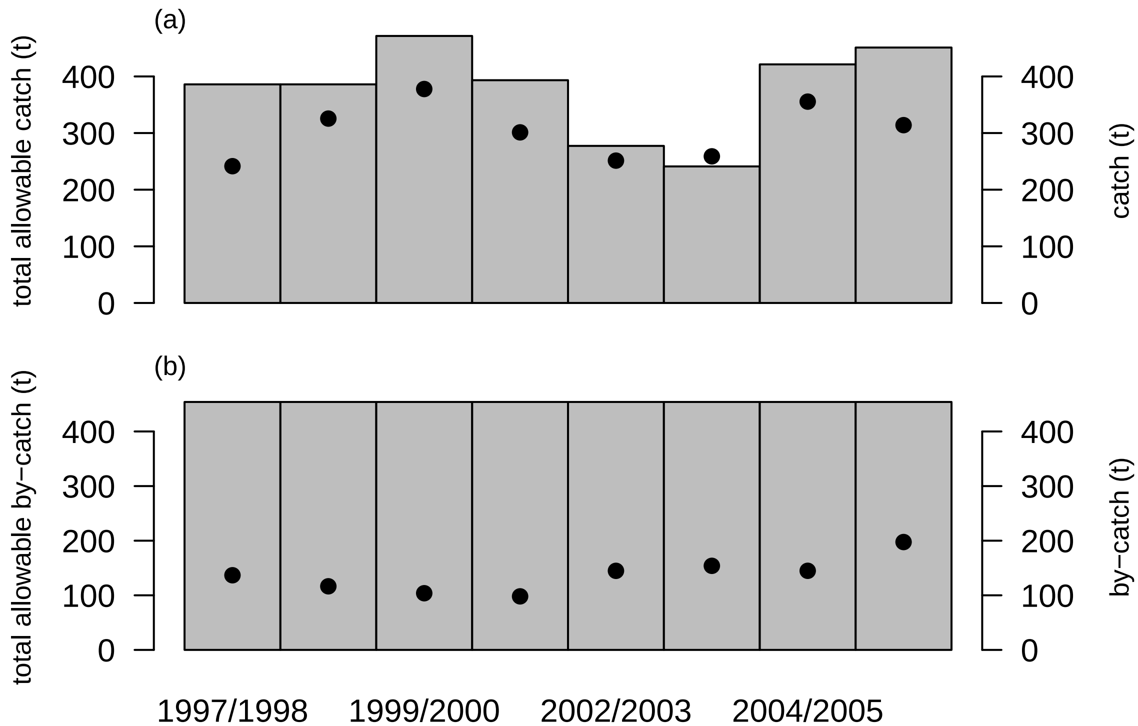

Figure 1.3: Dots show yearly sablefish (sum of retained and marketable released dead, a) and halibut (released dead, b) extraction in metric tonnes (t) for years 1997/1998 through 2005/2006, excluding 2000/2001. Bars show total allowable catch for sablefish (Fisheries and Oceans Canada, 2009) and total allowable by-catch for halibut (Fisheries and Oceans Canada, 2005a).

quota is low (S. Buchanan, pers. comm., 2006). This strategy can free up expensive sablefish quota because only a proportion of MR fish are dead, allowing harvesters to continue fishing for other species (Branch, 2004). Observers calculate MR sablefish mortality using a rule based strictly on event duration (Table 1.2 & Figure 1.4a). Harvesters are aware of the mortality rule and may limit event duration to minimize the MR dead weight (Figure 1.4b). It is unlikely that sablefish are released in order to retain other species because its value is high relative to other species (Gillis *et al.*, 1995).

Harvesters typically try to avoid catching halibut because it is a prohibited species. Halibut by-catch quota allows harvesters to bottom-trawl as long as they have quota to account for DR halibut. Harvesters who exceed their halibut by-catch limit are restricted from bottom-trawling for the rest of the year, or until they acquire more by-catch quota.

Table 1.2: Fisheries and Oceans Canada (2005a,b) specifies the minimum legislated length (i.e., marketable length in meters), retention and mortality rate for sablefish and halibut. Halibut minimum legislated length refers to the directed hook and line fishery because halibut retention is prohibited from the BC trawl.

Species	Minimum legislated length	Retention	Mortality rate
Sablefish	≥ 0.550	Allowed	Mortality is 0.10 for first two hours, and increases by 0.10 each additional hour
Halibut	≥ 0.813	Prohibited	Observers visually assess each piece's condition

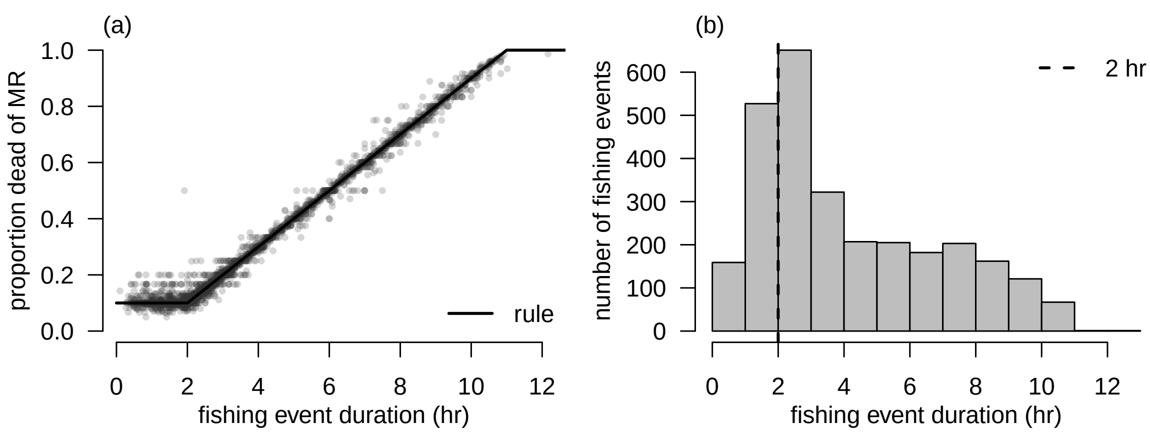

Figure 1.4: The mortality rule (black line) used by observers to calculate marketable released (MR) sablefish mortality is based strictly fishing event duration (a). Dots show the relationship between reported MR sablefish mortality and event duration in hours (hr) for the subset of events with non-zero MR sablefish reports. Deviation from the rule is generally due to rounding: weight is rounded to the nearest 1 pound, and duration is rounded to the nearest 0.1 hr. Dots are semi-transparent to help visualize density. Panel (b) shows the distribution of event duration; the dashed line corresponds to the point at which mortality increases from 0.10 in Panel (a).

Individual harvesters are limited to a maximum of 4% of the annual fleet-wide TAB mortality of 454 t (Fisheries and Oceans Canada, 2005a). As with sablefish, the fleet-wide TAB still applies. Annual BC trawl halibut extraction ranges between 22 and 44% of the TAB (median = 31%, Figure 1.3b). Harvesters are limited to a maximum roll-over allowance of 15% of their halibut quota (Fisheries and Oceans Canada, 2005a). Unlike sablefish, the halibut roll-over allowance enables harvesters to carry unused halibut quota forward, but harvesters are not allowed to apply by-catch to the next year's quota.

Observers estimate total halibut by-catch and determine DR weight by visually assessing each piece's condition (Table 1.2). Observers use predefined indicators adapted from Williams and Wilderbuer (1995) to assess halibut condition (Table A.1, Appendix A). Observers attempt to measure each piece's length, and calculate weight based on predefined length-weight conversion factors. When observers are unable to measure and assess each piece, the proportion assessed is extrapolated to the entire halibut catch. Piece-specific mortality rates are combined into per-event DR weight.

Both sablefish and halibut are targeted by directed hook and line fisheries. The BC trawl's allocation is a relatively small percentage of the entire Canadian TAC for these species, calculated as the sum of the BC trawl's sablefish TAC or halibut TAB, and the respective hook and line fishery's TAC. The BC trawl sablefish TAC is calculated yearly as a fixed percentage of the entire Canadian sablefish TAC, and the BC trawl halibut TAB is maintained at 454 t/yr. In 2005/2006, the BC trawl had about 9% and 8% of the entire Canadian TAC for sablefish and halibut respectively (Fisheries and Oceans Canada, 2005b,c).

The DFO acknowledges that sablefish "mortality rates do not necessarily reflect true mortality rates of fish released at-sea, but are intended to provide incentives for vessel operators to reduce towing time and avoid by-catch wherever possible" (Fisheries and Oceans Canada, 2005a). Although halibut mortality rates may be biologically accurate, I assume that the intention is to reduce by-catch.

1.5 Research goals

It is important to test the assumptions of ASOP release estimate reliability and equitability in the BC trawl. However, because there is no known true account of releases at sea, I use a model to predict release rates on individual events using factors that affect release activity.

This first step will address a knowledge gap for predicting release rates, which are influenced by a variety of complex, weakly-interacting fishery-specific effects (Saila, 1983), for which there is currently a lack of knowledge and possible misconceptions regarding proportionality to catch or effort (Kelleher, 2005; Rochet and Trenkel, 2005).

The second step will quantify over- or under-reporting using indices of observer reliability and skipper equitability. Reliable release reports are essential for stock assessment due to partial release mortality of quota species. Equitable release reports are essential for ensuring that harvesters are treated fairly. Harvesters who have had less-than-expected quota deducted from their licence have an economic advantage over other harvesters. The economic advantage is in terms of the both species in question, and other species that are associated with the species in question. For example, harvesters that account for less MR sablefish than predicted are not only able to retain more sablefish for sale, but also have greater economic benefits resulting from the capture of other species found in association with sablefish when sablefish quota is in short supply. Specifically I aim to quantify:

1. The relative importance of environmental, social and economic variables on predicted release rates.

2. The marginal effect of the most important predictors on release rates.

3. An index of relative reliability for each at-sea observer, comprised of effect and magnitude. The effect indicates each observer's tendency to report more or less MR sablefish or DR halibut with respect to the average observer in similar circumstances. The magnitude is the over- or under-reported weight, calculated as the difference between the reported and the predicted weight for events associated with observers that have positive or negative reliability indices. The over- or under-reported weight is zero if the observer's effect is zero. Because observer reports are typically estimates, I expect that observer reports have some random variation about the true value. Therefore, the observer relative reliability index identifies systematic tendencies to over- or under-report release estimates across multiple events and trips. Observer relative reliability indices will be referred to as "observer reliability indices", knowing that these indices compare each observer to the average, not to an unbiased baseline value.

4. An index of equitability for each skipper, also comprised of effect and magnitude.

The effect indicates each skipper's over- or under-reported weight, calculated by considering fishing events associated with observers with positive or negative observer reliability indices. As with observers, the over- or under-reported weight is defined as the difference between the reported and the predicted weight of MR sablefish or DR halibut. The magnitude is the percent error (PE) of over- or under-reported MR sablefish or DR halibut. The skipper equitability index quantifies inequitable economic benefits resulting from more- or less-than-expected quota being deducted. The over- or under-reported weight is due to the skipper's association with observers with positive or negative reliability indices because observers, not skippers, typically report releases.

5. The yearly and total over- or under-reported weight of MR sablefish and DR halibut.

Observer reliability and skipper equitability indices provide a surrogate for ASOP reliability and equitability. Although these indices do not directly quantify overall observer reliability and skipper equitability, these indices reflect the effect and magnitude of over- or under-reporting estimates. Ensuring the veracity of the BC trawl's ASOP by quantifying indices of reliability and equitability is important for maintaining programme credibility.

If the assumption of reliable and equitable release estimates holds, this project will improve confidence in ASOP data, which underpins stock assessments and the ITQ management system. If the assumption does not hold, I will suggest changes to improve ASOP data utility. Although daily quality-control ensures ASOP data is consistent and provides observer feedback, this is the first formal analysis of BC trawl ASOP release estimates to quantify observer reliability and skipper equitability indices[5].

1.6 Analytical methods to assess releases

Classification and regression trees (CART) are well suited to the analysis of large data sets that cannot be analyzed using traditional statistics due to the presence of many weak predictors (Breiman, 2001), non-linear relationships, and complex and unknown interactions (De'ath and Fabricius, 2000; Maindonald and Braun, 2003). CART are useful for uncovering

[5] All individual identities are encrypted to protect privacy. For example, the observer-skipper combination "ab-CD" indicates observer "ab" and skipper "CD". Fishing locations are grouped by species management areas.

patterns, structure and interactions between predictors due to the ease of interpreting results (De'ath and Fabricius, 2000). Recently, the use of CARTs for understanding ecological data has increased (Cutler *et al.*, 2007; De'ath and Fabricius, 2000; Kolar and Lodge, 2002), and this methodology is an ideal candidate for the analysis of fisheries release data. However, CART have several weaknesses including: (1) continuous predictors are treated as discrete categories which may be inefficient; (2) low order interactions take the same precedence as high order interactions; (3) large trees with many splits may lose predictive power; and (4) CART are better at classification than regression (Maindonald and Braun, 2003).

The random forest (RF) method improves CART's prediction accuracy (Sexton and Laake, 2009). The RF is robust to outliers and noise, has high prediction accuracy (Breiman, 2001), and does not require predictors or response variables to meet distributional assumptions (Cutler *et al.*, 2007). For these reasons, the RF is well suited to predicting the proportion MR sablefish and DR halibut on individual fishing events using a suite of predictor variables. For example, Lennert-Cody and Berk (2007) recently applied RF to high-seas tuna fishery observer reports, in order to identify fishing events where dolphin mortality is expected yet not reported. Dolphin mortality reports have negative financial effects on harvester profitability, and there may be harvester pressure that causes observers to under-report dolphin mortality. Lennert-Cody and Berk (2007) quantified an index of observer reliability as the per-observer probability of being associated with fishing events with expected, but unreported, dead dolphin. Although insufficient to identify over- or under-reporting, results identified specific observers that have suspicious data that should receive further scrutiny (Lennert-Cody and Berk, 2007).

Chapter 2

Methods

My analysis of at-sea observer program (ASOP) release reports in the British Columbia offshore groundfish trawl fishery (BC trawl) involves two main steps. First, the random forest (RF) model predicts the proportion marketable released (MR) sablefish on each event using various predictors (Figure 2.1). Negative residuals, calculated as the difference between reported and predicted releases, suggest events with less-than-expected reported releases. The RF also quantifies the importance and relationship between predictors and the proportion MR.

The second step uses a linear mixed effects (LME) modeling approach to properly quantify observer reliability and skipper equitability indices given the hierarchical structure of the ASOP data. Observer reliability and skipper equitability indices are each composed of an effect and a magnitude, where the "effect" indicates the directional tendency, and the "magnitude" indicates the potential amount. For observer reliability indices, the effect indicates the tendency to over- or under-report MR sablefish and the magnitude indicates the potential total amount over- or under-reported. Note that throughout the remainder of this report, I use the term "misreported" in referring to the residual difference between the reported and the predicted weight. For skipper equitability indices, the effect is defined as the total misreported weight of MR sablefish by all observers fishing with a given skipper, and the magnitude component is represented by the percent error (PE), which is defined as the misreported MR sablefish weight divided by the actual total reported weight.

Because each event's release rate is compared to the average event's release rate in similar circumstances, the analysis will identify residuals that are positive, zero, and negative. Residuals are then grouped by observer to calculate an index of reliability. Therefore,

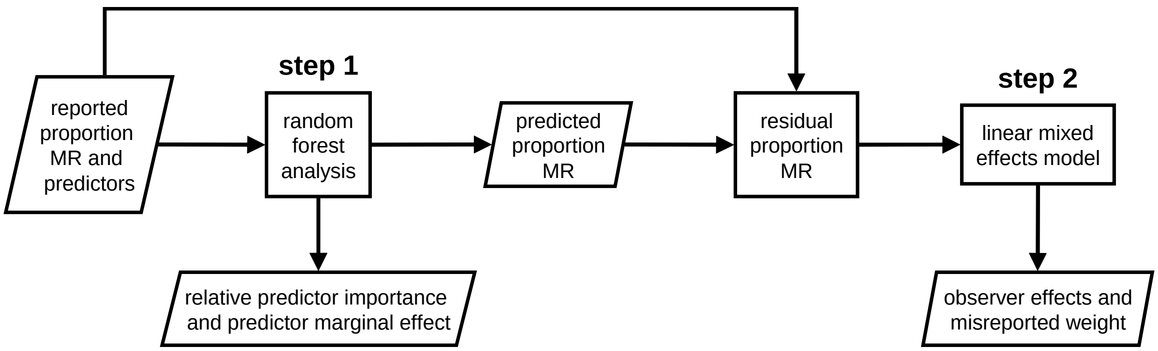

Figure 2.1: Flow chart representing the two-step methodology for investigating the reliability of release reports. First, the random forest analysis predicts the proportion marketable released (MR) sablefish on each fishing event using 26 predictor variables. The random forest also indicates each predictor's relative importance and marginal effect. Residuals are then calculated as the difference between the reported and the predicted proportion MR. Second, linear mixed effects models consider residuals to quantify tendencies of over- or under-reporting (i.e., misreporting). These tendencies are summarized as each observer's effect and misreported weight of MR sablefish. Parallelograms indicate inputs and outputs, and rectangles indicate processing steps.

the analysis will likely, but not necessarily, identify observer reliability indices that are positive, overlap zero, and are negative. Observers with negative reliability indices are not necessarily unreliable, but their index of reliability is significantly lower than the average observer's index. I consider the potential weight misreported by investigating events that are associated with observers that have either positive or negative (i.e., non-zero) reliability indices. Again, the analysis will likely, but not necessarily, identify skipper equitability indices that are positive, overlap zero, and are negative.

This two-step analysis is repeated in a similar fashion to investigate the proportion dead released (DR) halibut. Most observers are common to both sablefish and halibut analyses, which allows the comparison of observer reliability indices between species. For example, between-species correlation of observer reliability indices indicates whether observers who tend to under-report MR sablefish also tend to under-report DR halibut.

2.1 BC trawl at-sea observer data

Release data and predictors are obtained from Fisheries and Oceans Canada (DFO) PacHarv-Trawl database for fishing years from 1997/1998 through 2005/2006, excluding 2000/2001

Table 2.1: Event specific data used for the marketable released (MR) sablefish and the dead released (DR) halibut analyses for years 1997/1998 through 2005/2006, excluding 2000/2001. *Other* refers to infrequent codes (e.g., not sold, crew personal use, discarded at sea). Halibut retention is prohibited in the BC trawl: retained and unmarketable are not applicable (NA). Symbols in parentheses are used in model specification.

Variable or Utilization	Value or Unit	Sablefish	Halibut
Number of predictors (P)	integer	26	24
Number of events (N)	integer	58,315	59,187
Number of trips (J)	integer	6,302	6,694
Number of observers (L)	integer	322	324
Number of skippers (M)	integer	130	137
Catch	metric tonnes	5,064	3,546
Retained	metric tonnes	2,338	NA
Unmarketable released	metric tonnes	2,594	NA
MR dead; DR	metric tonnes	38	1,051
MR live; live released	metric tonnes	89	2,491
Other	metric tonnes	5	4

due to incomplete data (Table 2.1)[1]. Observers report event-specific release rates and record various factors that may affect fishing activity. An event, defined as fishing activity resulting in the capture of the species examined, is considered only if all 26 and 24 predictors are reported for sablefish and halibut, respectively.

[1]The BC trawl fishing year runs from about April 1 to March 31. At the beginning of each year, TACs are set and ITQs are allocated to licence holders.

Table 2.2: Variables used to predict the proportion marketable released sablefish and dead released halibut. The number of levels for categorical predictors (Cat) are given in parentheses. Value indicates each level's name. Name order for categorical predictors corresponds to name order in the results section. Units of measurement are given for continuous predictors (Cont). Predictors indicate conditions at the start of the event unless stated otherwise.

Predictor	Type	Value or Unit	Description
Environmental predictors			
statArea	Cat (8)	3C, 3D, 4B, 5A, 5B, ..., 5E	Species management area
year	Cat (8)	1997/1998, 1998/1999, ..., 2005/2006	Fishing year, from Apr 1 to Mar 31 (year 2000/2001 is omitted)
quarter	Cat (4)	one, two, ..., four	Fiscal quarter: Jan to Mar, Apr to Jun, ..., Oct to Dec
month	Cat (12)	jan, feb, ..., dec	Month of the year
depth	Cont	meters	Bottom depth (the net is usually, but not always, on the bottom)
Trip predictors			
tripDays	Cont	days	Duration of the fishing trip
totalTows	Cont	integer	Total number of events in the trip
Event predictors			
endTime	Cat (2)	day, night	Time at the end of the event: 5 am to 10 pm; 11 pm to 4 am
method	Cat (2)	observer, skipper	Individual who reported the species composition of the catch

Table 2.2 continued

Predictor	Type	Value or Unit	Description
cumTotalRetCatch	Cont	kilograms	Cumulative weight of all species retained for the trip, updated each event
propFull	Cont	proportion	Proportion of vessel *capacity* that is full of fish (inferred)
relativeTow	Cont	proportion	Event number with respect to *totalTows*
towDuration	Cont	hours	Time interval from when the net is in fishing position until net retrieval begins
capacity	Cont	kilograms	Vessel hold size (inferred from the largest reported *cumTotalRetCatch*)
Social predictors			
obsSkipTows	Cont	integer	Number of events in common for the observer-skipper combination
seaDays	Cont	days	Cumulative number of days that the observer has worked
Economic predictors			
initialQuota	Cont	kilograms	Vessel quota at the start of the fishing year
fleetQuota	Cont	kilograms	Quota available for the species in question, summed over all vessels in the fleet (updated weekly)
quota	Cont	kilograms	Vessel's quota for the species in question (updated weekly)
YTDcatch	Cont	kilograms	Vessel's year-to-date retained catch (updated weekly)
Biomass predictors			
totalCatch	Cont	kilograms	Total caught weight (all species)
totalRetCatch	Cont	kilograms	Total retained weight (all species)

Table 2.2 continued

Predictor	Type	Value or Unit	Description
totalFish	Cont	kilograms	Caught weight of the species examined
retained	Cont	kilograms	Retained weight of the species examined (sablefish only)
totalReleased	Cont	kilograms	Released weight (sum of marketable and unmarketable weight) of the species examined (sablefish only)
thornyhead	Cat (2)	none, present	Both species of thornyhead rockfish (*Sebastolobus alascanus* and *S. altivelis*) are retained (sablefish only)
skateWeight	Cont	kilograms	Encountered skate (family Rajidae) weight (halibut only)

Predictors include environmental, trip, event, social, vessel, economic and biomass variables (Table 2.2). Environmental predictors are included to account for spatial and temporal stock patterns, which may influence harvester release behaviour. Spatial and temporal predictors may also be related to target species. Temporal predictors (i.e., *year*) may also account for changes to the BC trawl management programme and the ASOP. For example, the ASOP training program has changed over time in response to changing management priorities. Trip predictors, such as *tripDays*, may influence harvester behaviour, and long duration trips may also provide opportunities for harvesters and observers to become more familiar. Event predictors may influence observer reporting behaviour and harvester release behaviour. For example, catch that is brought aboard at night may be more difficult to monitor, and releases may be undetected. Unmonitored events, which are reported by the skipper, may have lower release reports than monitored events. Harvesters may alter release activity on events that occur near the end of the trip, or when the vessel is almost full. Social predictors are included to account for potential observer-skipper interactions, and to quantify observer experience. Economic predictors are related to the amount of vessel-specific and fleet-wide quota available, and the amount of quota that has been used to account for retained catch and MR sablefish or DR halibut. Biomass predictors indicate aspects of the catch composition and crew sorting practices. The *thornyhead* predictor indicates fishing events that likely targeted thornyhead rockfish (*Sebastolobus alascanus* and *S. altivelis*), which is a specialized deep-water bottom-trawl fishery. The weight of skate (family Rajidae) encountered may affect halibut mortality because skates can be large, heavy and abrasive fishes. Although some predictors are specific to sablefish or halibut, most are common to both analyses. See Appendix A for extended data summaries showing the range and distribution of data available for the analysis of MR sablefish (Tables A.2 & A.3), and DR halibut (Tables A.4 & A.5).

Some predictor pairs are closely related: for example, *totalRetCatch* is the weight of retained catch, a subset of *totalCatch*. I use Spearman's rank correlation coefficient to determine predictor pair correlation, which indicates predictor pairs that should be investigated with bivariate marginal dependence plots. Spearman's ρ is used when data are not from a bivariate normal distribution, and is more accurate than Kendall's correlation coefficient when N is large (Zar, 1996). Spearman's ρ quantifies the relationship strength between each predictor pair combination, and does not assume dependence (Zar, 1996).

To calculate both observer and skipper effects in the LME model, trips must satisfy at least one of the following conditions: (1) the observer has been with at least 2 different skippers; or (2) the skipper has been with at least 2 different observers. This condition is required to avoid confounding observer and skipper effects.

The reported proportion MR sablefish on the i^{th} of N events is

$$y_i^{\text{Rep}} = \frac{\text{MR}_i}{\text{TR}_i} \text{ for } i = 1,\, 2,\, ...,\, N \tag{2.1}$$

where MR_i is the reported MR weight (kilograms, kg) and TR_i is the sum of reported MR and unmarketable released weights. The superscript "Rep" distinguishes the reported proportion MR from the predicted proportion MR, y_i^{Pred} which appears in later equations. Over 95% of events report zero MR sablefish. Figure 2.2a shows the distribution of MR sablefish reports for the subset of events with non-zero reported MR sablefish (2,808 events). About 56% of observers and 68% of skippers are associated with at least one event with non-zero reported MR sablefish. The other observers and skippers are not associated with any reported MR sablefish.

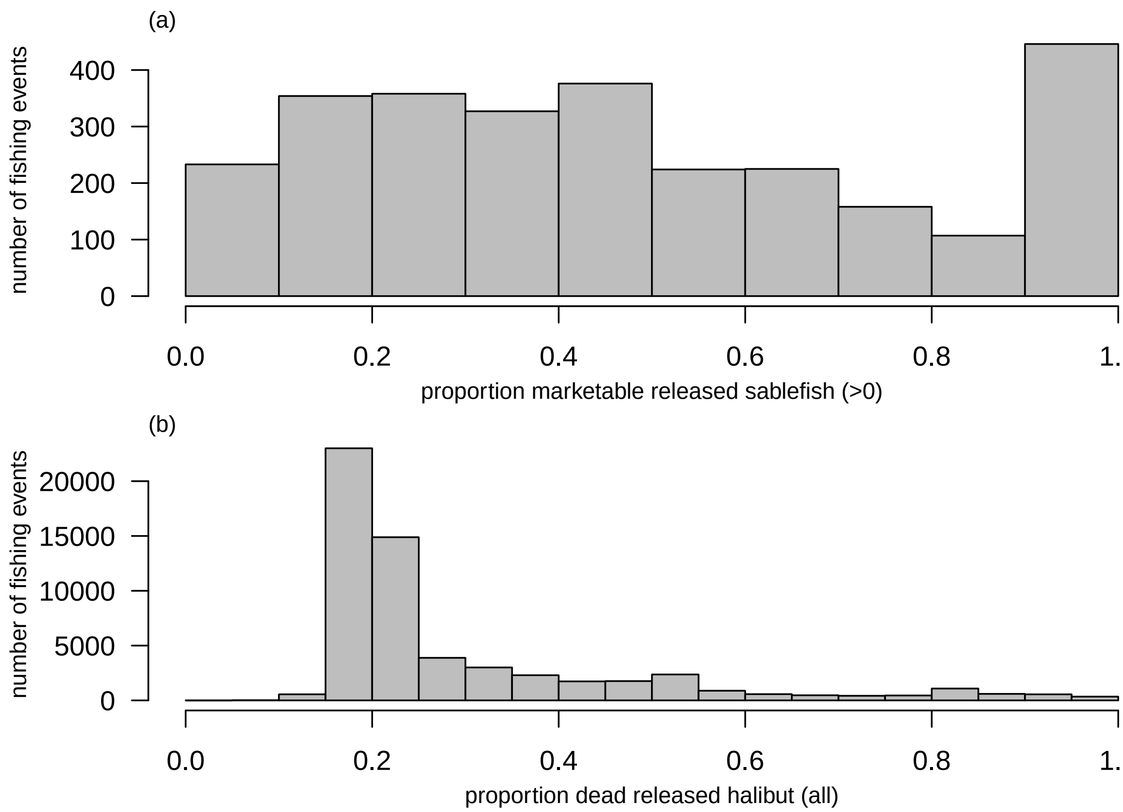

Figure 2.2: Reported proportion marketable released (MR) sablefish for the subset of events with non-zero reported proportion MR sablefish (2,808 events). These events are relatively rare, accounting for less than 5% of the 58,315 events that reported sablefish. Reported proportion dead released halibut for all events that encountered halibut (59,187 events, b).

2.2 Classification and regression trees

Classification and regression trees (CART) sequentially split groups of events into successively smaller subgroups. Splits attempt to minimize the within-subgroup sum of squares (SSQ) while maximizing the between-subgroup SSQ (Maindonald and Braun, 2003). Splits consider all predictors, and conditions split data into two subgroups to maximize deviance reduction. The deviance prior to the first split is given by (Maindonald and Braun, 2003):

$$D = \sum_{i=1}^{N} \left(y_i^{\text{Rep}} - \bar{y} \right)^2 \qquad (2.2)$$

where $\bar{y}$ is the mean y_i^{Rep}. The deviance after the first split is given by

$$D = \overbrace{\sum_{s_1=1}^{S_1} \left(y_{s_1}^{\text{Rep}} - \bar{y}_{s_1}\right)^2 + \sum_{s_2=1}^{S_2} \left(y_{s_2}^{\text{Rep}} - \bar{y}_{s_2}\right)^2}^{\text{within-subgroup SSQ}} + \overbrace{S_1 \left(\bar{y}_s - \bar{y}\right)^2 + S_2 \left(\bar{y}_s - \bar{y}\right)^2}^{\text{between-subgroup SSQ}} \quad (2.3)$$

which shows two subgroups of events: s_1's from $s_1 = 1, 2, ..., S_1$ and s_2's from $s_2 = 1, 2, ..., S_2$, where $y_{s_1}^{\text{Rep}}$ is the reported proportion MR for the s_1^{th} event in sub-group 1, $\bar{y}_{s_1}$ is the mean reported proportion MR in sub-group S_1, and so-on for events in S_2. For example, deep events (i.e., *depth* ≥ 292.6 meters, m) may have a higher mean proportion MR sablefish than shallow events. In this case, the CART will split events into two sub-groups based on whether the reported depth (i.e., predictor) is greater or less than 292.6 m (i.e., condition), if this split maximizes the deviance reduction.

Groups are iteratively split into smaller subgroups until improvements to model fit is compensated by decreasing prediction accuracy, measured by the cross-validation error rate (Maindonald and Braun, 2003). The cross-validation error rate, which includes a penalty for each additional split, initially decreases but then increases as the number of splits increases (Maindonald and Braun, 2003). Terminal nodes, or subgroups that are not split, have a predicted proportion MR sablefish y_S^{Pred} equal to the node's mean proportion MR.

2.3 Random forest analysis

The RF approach improves CART prediction accuracy by building multiple regression trees using bootstrap samples of events and random sub-samples of predictors (Breiman, 2002). Like CARTs, the RF groups similar events (e.g., *depth*, *totalCatch*, *seaDays*) into subgroups (Liaw and Wiener, 2002) according to the following algorithm (Lennert-Cody and Berk, 2007):

1. Draw a bootstrap sample of N events with replacement.

2. Draw a random sample of p predictors without replacement.

3. Determine the first split to maximize the deviance reduction based on data from Steps 1 and 2.

4. Repeat Steps 2 and 3, splitting groups into subgroups. Groups that contain fewer than g events are considered terminal nodes and are not split.

5. Drop data that are not selected in Step 1 down the resulting tree. These data account for approximately 37% of events per tree, and are referred to as out-of-bag (OOB).

6. Save the prediction for each OOB event.

7. Repeat Steps 1 to 6 k times.

The RF model fit is evaluated via the OOB mean squared error (MSE) and percent variance explained (PVE) for all k trees (Liaw and Wiener, 2002):

$$\text{MSE} = N^{-1} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \quad (2.4)$$

$$\text{PVE} = \left(1 - \frac{\text{MSE}}{\hat{\sigma}_y^2}\right) \times 100 \quad (2.5)$$

where $\hat{y}_i$ is the mean predicted OOB response for the i^{th} event. Increasing prediction accuracy is indicated by decreasing MSE and increasing PVE (Liaw and Wiener, 2002). Prediction accuracy is high when trees are grown to maximal depth (i.e., many subgroups) to ensure low individual tree error, and both events and predictors are selected randomly to reduce inter-tree residual correlation (Segal, 2004).

In some cases tuning the RF can improve prediction accuracy, evaluated via decreasing MSE. The RF has three tuning parameters: the number of trees k, the number of predictors p, and the minimum node size g (Table 2.3). Although RF prediction accuracy is not typically sensitive to tuning parameter values, the presence of many weak predictors in this analysis indicates that tuning may improve prediction accuracy (Liaw and Wiener, 2002; Segal, 2004).

2.3.1 Predictor importance and marginal dependence

Because RF creates many trees, it is not feasible to show individual tree structure (Prasad *et al.*, 2006). Instead, I examine relative predictor importance and predictor marginal dependence. There are two measures of predictor importance: permutation accuracy and node purity. Permutation accuracy indicates each predictor's effect on prediction accuracy loss measured via MSE, and takes complex interactions among predictors into account (Liaw and Wiener, 2002). Permutation accuracy is determined by comparing prediction accuracy changes after randomly permuting all the values of the predictor in question on each tree, keeping the other predictors unchanged (Breiman, 2002; Breiman *et al.*, 2005; Liaw and

Table 2.3: Random forest (RF) baseline (i.e., default) parameter values. Tuning these parameters within constraints can improve RF prediction accuracy in some situations. The total number of predictors is P. The constraint on g ($g \geq 5$) is imposed by the analyst.

Effect	Parameter	Constraint	Baseline value
Number of trees	k	$k \geq 1$	500
Number of predictors	p	$1 \leq p \leq P$	$P/3$
Minimum node size	g	$g \geq 5$	5

Wiener, 2002). The difference between MSE before and after permutation is averaged over all trees and normalized by the standard error of the mean difference (Liaw and Wiener, 2002). Important predictors are present on more trees, are closer to the root node, and have larger effects on predictions (Breiman, 2002). Lennert-Cody and Berk (2007) used permutation accuracy importance to determine the most important predictors affecting reported dolphin mortality in the high-seas tuna fishery.

The second measure of predictor importance, node purity, measures the reduction in each individual tree's residual SSQ from splitting on a predictor, averaged over the k trees (Berk, 2006). Permutation accuracy is used to assess predictor importance, but node purity importance is included in the analysis for comparison.

Predictor marginal dependence indicates each predictor's effect on the proportion MR after averaging out the effects of other predictors. Marginal dependence is useful for visualizing results, but can be misleading when characterizing or interpreting high-order interactions, or less important predictors (Cutler *et al.*, 2007). Marginal dependence is characterized by the mean proportion MR sablefish. Determining the combined effect of two predictors can be misleading because of interactions and non-additive effects which may not be reflected in univariate marginal dependence plots. Therefore, I investigate interactions between correlated predictor pairs using bivariate marginal dependence plots (Cutler *et al.*, 2007).

2.3.2 Random forest event-specific residuals

Random forest predictions y_i^{Pred} are compared against reported release rates to calculate residuals. Each tree has a vector of residuals

$$y_i^{\text{Res}} = y_i^{\text{Rep}} - y_i^{\text{Pred}} \text{ for } i = 1, 2, ..., N_{\text{OOB}} \tag{2.6}$$

where N_{OOB} is the number of OOB events. Negative residuals indicate events with a lower reported than predicted proportion MR. Observers who are systematically associated with negative residuals may have tendencies to under-report the proportion MR sablefish. Misreported MR sablefish is calculated as the difference between the reported and the RF predicted weight. The misreported weight of MR sablefish is a component of the observer reliability index, which is quantified using LME models.

2.4 Linear mixed effects model specification

Each of the $k = 500$ RF trees is investigated separately, and the following LME procedure is done 500 times. Fisheries release data typically has a hierarchical structure, in which multiple events take place on a trip, and multiple trips are associated with an observer or skipper. This hierarchical structure must be taken into account when quantifying observer reliability and skipper equitability indices to account for three levels of variation: (1) between observers and between skippers; (2) between trips within individual observers and skippers; and (3) between events within trips within individual observers and skippers. Residuals are more closely correlated between events within trips than between events on different trips (Borges *et al.*, 2004; Kelleher, 2005), possibly due to target species, fishing gear, or other unmeasured trip-specific factors. Also, residuals are likely more correlated between trips within observers and skippers than between trips for different observers and skippers. Thus, events within trips, and trips within observers and skippers are not independent observations.

Although observers ($l = 1, 2, ..., L$) are of primary interest, skippers ($m = 1, 2, ..., M$) are included because fishing activity is likely influenced by skipper-specific effects (Palsson and Durrenberger, 1990), and trips ($j = 1, 2, ..., J$) are included to account for the nested data structure. Observers and skippers are specified as fixed effects, which enables the quantification of the specific effect of each observer and skipper. Trips are specified as random effects to account for the variation among events within trips, and among trips within observers and skippers. The LME predicts the residual proportion MR sablefish for events i on trip j for observer l and skipper m one tree at a time (Zuur *et al.*, 2009):

$$\begin{aligned} \mathrm{Y}_{ijlm} &= \mathrm{O}_{jl} + \mathrm{S}_{jm} + \epsilon_{ijlm} && (2.7) \\ \mathrm{O}_{jl} &\sim \mathcal{N}\left(\mathrm{O}_l,\, \sigma^2_{\mathrm{obs}}\right) && (2.8) \\ \mathrm{S}_{jm} &\sim \mathcal{N}\left(\mathrm{S}_m,\, \sigma^2_{\mathrm{skip}}\right) && (2.9) \\ \epsilon_{ijlm} &\sim \mathcal{N}\left(0,\, \sigma^2_{\mathrm{err}}\right) && (2.10) \end{aligned}$$

where Y_{ijlm} is the residual for event i on trip j for observer l and skipper m. The notation $\mathcal{N}(\mu, \sigma^2)$ indicates a normal distribution with mean μ and variance σ^2. The trip-specific random effects for observer l on trip j, O_{jl} are normally distributed about the observer's mean effect, also called the fixed effect O_l, with variance σ^2_{obs}. All observers' random effects have the same variance. As with the observers, the skipper's trip-specific random effects S_{jm} are normally distributed about the skipper's fixed effect S_m, with variance σ^2_{skip}, which is equal across skippers. Random error terms ϵ_{ijlm} for events i within trip j for observer l and skipper m are normally distributed about zero with variance σ^2_{err} and homogeneous across trips.

The LME model is fit using the restricted maximum likelihood estimation method to estimate unbiased random and fixed effects: (1) random trip effects are the best linear unbiased predictors; and (2) fixed observer and skipper effects are the best linear unbiased estimates. I am only interested in each observer's fixed effect. Iterative tree-by-tree LME analysis calculates a vector of fixed effects for each observer, accounting for the possibility that randomized tree structure may allow y_i^{Pred}'s to be more similar to y_i^{Rep}'s within trees than between trees. Some observers do not have OOB events in every tree and thus have less than k effects.

The observer fixed effects O_l's are not required to sum to zero. Random forest predictions y_i^{Pred} consider only OOB events, which are not used in the tree's construction. Therefore, residuals y_i^{Res} are not required to sum to zero. Because these residuals could have either a positive or a negative tendency, observer effects O_l's could also have either a positive or a negative tendency.

2.4.1 Observer reliability index

Observer reliability indices have effect and magnitude components. The observer effect indicates each observer's tendency to misreport MR sablefish, described by the median of up to k estimates (i.e., one from each LME model), while uncertainty is indicated via the 90^{th} percentile range. These statistics give accurate estimates of central tendency and variability, respectively, when distributions are skewed and outliers are present (Zar, 1996). The observer effect, which is on the same scale as the response (i.e., the residual), indicates the difference between the reported and the predicted proportion MR. For example, an observer effect of -0.05 indicates the observer tends to under-report the proportion MR sablefish by 5% less than the average. Reliability indices are considered significantly different from zero when the 90^{th} percentile range of observer effects does not overlap zero. Thus, observers with negative reliability indices may significantly under-report MR sablefish. Although negative reliability indices do not necessarily identify unreliable observers, individuals with negative reliability indices may be less reliable than the average observer in similar circumstances, if the average observer is reliable. Releases reported on events associated with observers that have non-zero effects (i.e., positive or negative) may warrant further investigation; the misreported weight for these events is

$$\delta_i = \left(y_i^{\text{Rep}} - y_i^{\text{Pred}}\right) \times \text{TR}_i \text{ for } i = 1,\, 2,\, ...,\, N_{\text{OOB}_{\text{NZ}}} \tag{2.11}$$

where $N_{\text{OOB}_{\text{NZ}}}$ is the number of OOB events that are associated with observers that have non-zero effects. If the observer's effect overlaps zero, $\delta_i = 0.00$. For observers with non-zero effects, the observer reliability index magnitude is quantified by the misreported weight of MR sablefish, calculated as $\sum \delta_i$ for each tree. The magnitude is described by the median and 90^{th} percentile range of the 500 estimates of misreported weight. This estimate of misreported weight does not take observer effects into account, but instead relates the observer's reported weight to the RF predicted weight for events associated with observers that have non-zero reliability indices.

2.4.2 Skipper equitability index

As with observers, skipper equitability indices have effect and magnitude components. However, skipper effects are not from the LME model; skipper effects and magnitude are calculated differently than observer effects and magnitude. The effect component is the per-skipper misreported weight, which may indicate an unexpected economic disadvantage (i.e., over-reported MR sablefish) or advantage (i.e., under-reported MR sablefish) due to their association with observers with non-zero $\sum \delta_i$. If all of the skipper's events are associated with observers who have effects that overlap zero, the skipper effect is zero. Skipper effects will also be zero if over- and under-reported weights are of equal magnitude. However, if the skipper has been associated with observers who typically have either over- or under-reported MR sablefish, the skipper effect may be positive or negative, respectively.

The magnitude component is the percent error (PE):

$$\text{PE} = \frac{\sum \delta_i}{\text{MR}^{\text{total}}} \tag{2.12}$$

where MR^{total} is the total reported weight of MR sablefish for all of the skipper's events (i.e., not just events with misreported MR sablefish). Equation 2.12 is repeated once for each tree.

As with observers, the effect and magnitude of skipper equitability indices are quantified by the median and 90^{th} percentile range. Each component of the equitability index is considered significantly different from zero when the 90^{th} percentile range of the effect does not overlap zero. For example, a negative equitability index indicates the skipper may have had less-than-expected MR sablefish deducted from their quota. Therefore, the skipper may have had an economic advantage over other skippers because of their association with observers who typically under-report MR sablefish.

2.4.3 Yearly and total over- or under-reported releases

Again, fishing events associated with observers having non-zero reliability indices may have misreported MR sablefish. Yearly and total misreported MR sablefish is described by effect and magnitude components. These are calculated the same way as skipper equitability indices. The effect is the weight of misreported MR sablefish, calculated as $\sum \delta_i$. The magnitude is the PE, calculated with Equation 2.12. Again, MR^{total} considers all the observers in the analysis (i.e., observers with both non-zero and overlapping zero reliability indices).

2.5 Proportion dead released halibut

The above sablefish procedure quantifies observer reliability and skipper equitability indices based on the predicted proportion MR sablefish. The sablefish procedure is modified for a separate analysis to investigate the proportion dead released (DR) halibut by substituting

$$y_i^{\text{Rep}} = \frac{\text{DR}_i}{\text{TR}_i} \text{ for } i = 1,\ 2,\ ...,\ N \tag{2.13}$$

in Equation 2.1. Most observers use predefined indicators to assess halibut mortality (Table A.1). Although observers are strongly encouraged to follow these guidelines, some use other metrics to assess halibut mortality, which means that the actual reported proportion DR halibut ranges between 0.00 and 1.00 (median = 0.204, Figure 2.2b). All observers and skippers are associated with at least one event with non-zero reported DR halibut.

2.6 Observer reliability index correlation

Because sablefish and halibut are often encountered on the same fishing event, 319 observers are common to both sablefish and halibut analyses. Each observer has about $k = 500$ LME effects, or one from each RF tree. I therefore use Spearman's ρ to quantify whether observers who tend to under-report MR sablefish also tend to under-report halibut, thus providing an index of observer reliability across species. Correlations are computed using the following algorithm:

1. Draw a random sample of 319 observers with replacement.

2. For each observer in Step 1, draw a random sablefish and halibut effect from the observer's $k = 500$ fixed sablefish and halibut effects[2].

3. Compute and save the correlation coefficient ρ between the 319 observer effect pairs. Some bootstrap samples have less than 319 effect pairs because some observers do not have OOB residuals in every LME iteration.

4. Repeat Steps 1 to 3 100,000 times.

[2]Some observers have less than 500 fixed effects because some observers do not have OOB residuals in every LME iteration.

The correlation is quantified by the median and 90th percentile range of the 100,000 bootstrap iterations, and is considered significantly different from zero when the 90th percentile range does not overlap zero.

2.7 Computational details

Data analysis uses the statistical and graphing programme `R-2.8.1` (R Development Core Team, 2009), using the packages `rpart` for the example regression tree (Therneau *et al.*, 2009), `randomForest` for the RF procedure (Breiman *et al.*, 2005), `lme4` for the LME procedure (Bates and Maechler, 2009), and `snow` for parallel processing (Tierney *et al.*, 2009). Uncertainty (i.e., two times the standard error of the mean) in predictor marginal dependence is calculated using a method developed by Sexton and Laake (2009), and bivariate marginal dependence is calculated using a method developed by Cutler *et al.* (2007).

Chapter 3

Results

Results for assessing reported release reliability in the British Columbia offshore groundfish trawl fishery (BC trawl) are divided into two main sections. The first section shows results for the model predicting marketable released (MR) sablefish and dead released (DR) halibut. A regression tree exemplifies the procedure predicting the proportion MR sablefish on fishing events. The random forest (RF) indicates relative predictor importance and marginal dependence, and predicts the proportion MR sablefish and DR halibut on individual events. Supplemental results are included in Appendix A.

The second section shows the observer reliability and skipper equitability indices obtained from the linear mixed effects (LME) model. Observer reliability indices quantify the tendency to over- or under-report (i.e., misreport) releases, and the magnitude of the misreported weight of MR sablefish and DR halibut. Skipper equitability indices quantify the misreported weight and the percent error (PE). Finally, the correlation of observer effects for observers that are common to both sablefish and halibut analyses quantifies trends of misreporting between species.

3.1 Sablefish regression tree

A regression tree dendrogram shows the first eight conditions predicting the proportion MR sablefish on each event (Figure 3.1). This regression tree is shown to illustrate the procedure used to predict the proportion MR sablefish. The first condition splits events (mean proportion MR $\bar{y} = 0.023$) into two subgroups based on the weight (kilograms, kg) of *totalReleased* sablefish. Events with less than 22.7 kg *totalReleased* on the left branch

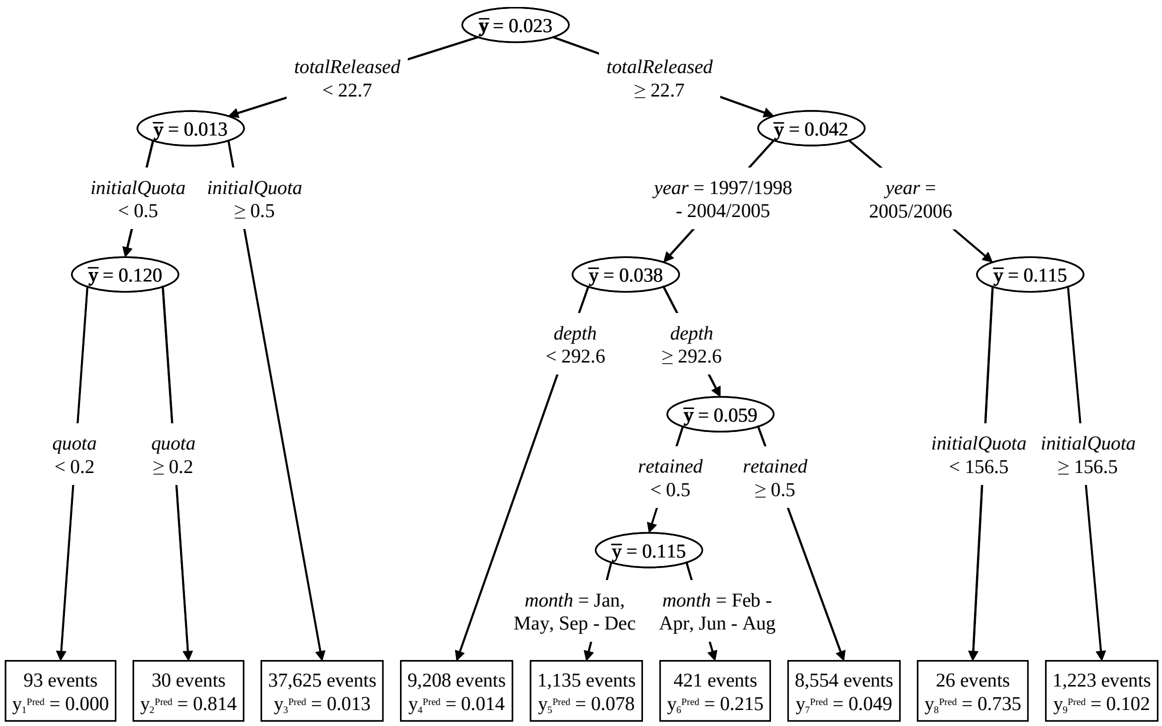

Figure 3.1: An example sablefish regression tree dendrogram showing the first eight conditions predicting the proportion marketable released (MR) sablefish. Ovals indicate split points and show mean proportion MR $\bar{y}$ for events in each subgroup, and conditions below split points specify which branch to follow depending on event characteristics (e.g., *totalReleased*, *year*). Rectangles identify terminal nodes, or subgroups that are not split, and show the number of events and predicted proportion MR sablefish, y^{Pred}.

have a lower proportion MR ($\bar{y} = 0.013$) than events with at least 22.7 kg *totalReleased* on the right branch ($\bar{y} = 0.042$). Thus, events with greater *totalReleased* sablefish also have a higher proportion MR sablefish. This procedure is repeated down to the terminal nodes.

In general, events in year 2005/2006 have relatively high proportion MR sablefish ($\bar{y} = 0.115$). Events with at least 22.7 kg *totalReleased*, in *year* 2005/2006 and with at least 156.5 kg *initialQuota* have predicted proportion MR sablefish $y_9^{\text{Pred}} = 0.102$. As expected, similar events (i.e., at least 22.7 kg *totalReleased* in *year* 2005/2006) but with less than 156.5 kg *initialQuota* have a higher predicted proportion MR ($y_8^{\text{Pred}} = 0.735$). The high proportion MR sablefish suggests that vessels with low *initialQuota* may selectively release marketable sized sablefish in order to conserve quota and continue fishing for the entire year.

The effect of *depth* (meters, m) is important for events with more than 22.7 kg *total-Released* in years 1997/1998 through 2004/2005. Events with *depth* less than 292.6 m have $y_4^{\text{Pred}} = 0.014$. Deeper events ($depth \geq 292.6\,m$) have a higher proportion MR, which is affected by the weight of *retained* sablefish and the *month* of the year ($\bar{y} = 0.059$). Harvesters targeting deep water species may encounter typically larger sablefish because sablefish migrate to deeper water as they age (Maloney and Sigler, 2008). In some cases, these harvesters may become constrained by sablefish quota during the year if they retain all the sablefish that they encounter, and may have a higher proportion MR sablefish to conserve quota.

The actual predictions used in the analysis are obtained from the RF, which uses a modified bootstrap procedure to build multiple regression trees. Because of this randomization, the example regression tree results differ slightly from the final RF results. The RF results presented in the following sections summarize all of the trees in the forest.

3.2 Random forest prediction accuracy

Sablefish RF prediction accuracy, measured via decreasing mean squared error (MSE), is investigated with respect to two tuning parameters: the number of predictors p, and the minimum node size g (Table 3.1). The sablefish tuning procedure uses a reduced number of trees $k = 200$ to decrease computation time without affecting prediction accuracy (Section A.3). The final sablefish RF uses baseline (i.e., default) $k = 500$, tuned $p = 10$, and tuned $g = 5$ to explain 46.9% of MR sablefish variability. Due to negligible sablefish RF prediction accuracy improvements from tuning, halibut is analyzed using baseline values. Halibut RF prediction error approaches the asymptotic minimum well before $k = 500$ and final prediction accuracy is 22.7%, approximately half that of the sablefish RF (Tables 3.1 & A.6).

Unexplained MR sablefish and DR halibut variability may be due to general difficulties in predicting fisheries release rates, or missing predictors that could improve prediction accuracy if included in the analysis (Figures A.3 & A.4, respectively). Low halibut RF prediction accuracy suggests that DR halibut are either more difficult to predict than sablefish, or are affected by factors that are missing from this analysis.

Table 3.1: The sablefish random forest tuning procedure is evaluated via mean squared error (MSE) and percent variance explained (PVE, %) with respect to the number of predictors p and tuned minimum node size g. The sablefish tuning procedure uses a reduced number of trees $k = 200$. The final sablefish analysis uses baseline $k = 500$, tuned p, and tuned g. The halibut analysis uses all baseline parameter values.

	Sablefish tuning		Final analysis	
Parameter	Baseline	Best fit	Sablefish	Halibut
k	200	200	500	500
p	8	10	10	8
g	5	5	5	5
MSE	0.0083	0.0082	0.0081	0.0248
PVE	46.0	46.3	46.9	22.7

3.3 Relative predictor importance

The most important sablefish predictor, assessed via permutation accuracy (Lennert-Cody and Berk, 2007), is *year*, followed by *month*, and then *seaDays* (relative importance of 98, 77, and 72, respectively, Figure 3.2a). The most important halibut predictor is *seaDays*, followed by *depth*, and then *year* (98, 96, and 83, respectively, Figure 3.3a). Observer experience, measured by *seaDays*, is among the three most important predictors in both analyses and indicates a strong effect of human factors on reported release rates in the BC trawl.

The least important predictors are *method* (8, 5) and *endTime* (5, 4) for sablefish and halibut, respectively. The predictor *method* indicates whether the observer or the skipper estimated the catch proportions. This predictor is probably not important because few events are estimated by skippers (1.5 and 0.3% for sablefish and halibut, respectively).

Node purity importance is shown to compare with permutation accuracy importance. Node purity importance indicates that *seaDays* is among the six most important predictors for both sablefish and halibut (Figures 3.2b & 3.3b, respectively). The two least important predictors are again *endTime* and *method* for both sablefish and halibut. Considering the two importance measures together, *initialQuota*, *quota*, *YTDcatch* and *totalReleased* have

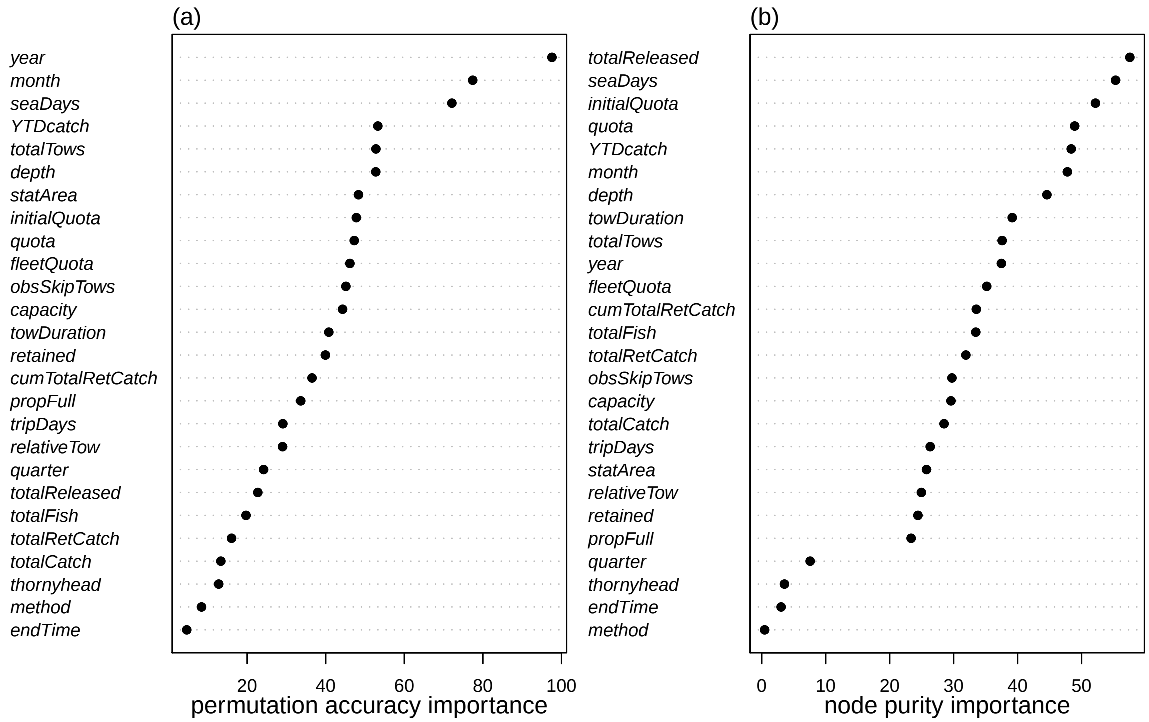

Figure 3.2: Sablefish predictors are ordered from top (most important) to bottom (least important) according to their relative importance measured via permutation accuracy (a) and node purity (b).

strong effects on the proportion MR sablefish, while *totalCatch*, *depth*, and *towDuration* have strong effects on the proportion DR halibut. There is a significant positive correlation between permutation accuracy and node purity importance for both sablefish ($\rho = 0.62$, $p < 0.001$) and halibut ($\rho = 0.64$, $p < 0.001$).

3.4 Predictor marginal dependence

For both sablefish and halibut analyses, observers with more experience report lower release rates than new observers. For example, after the effects of all other predictors are averaged out, new observers report approximately four times higher proportion MR sablefish and 5% higher proportion DR halibut than experienced observers. The predicted proportion MR sablefish y_i^{Pred} decreases from $0.125\,(\pm 0.025)$, where the number in parentheses represents two standard errors of the mean, for observers with less than 27 *seaDays*, to between 0.022 and 0.028 (range) for observers with at least 54 *seaDays* (Figure 3.4c). The predicted

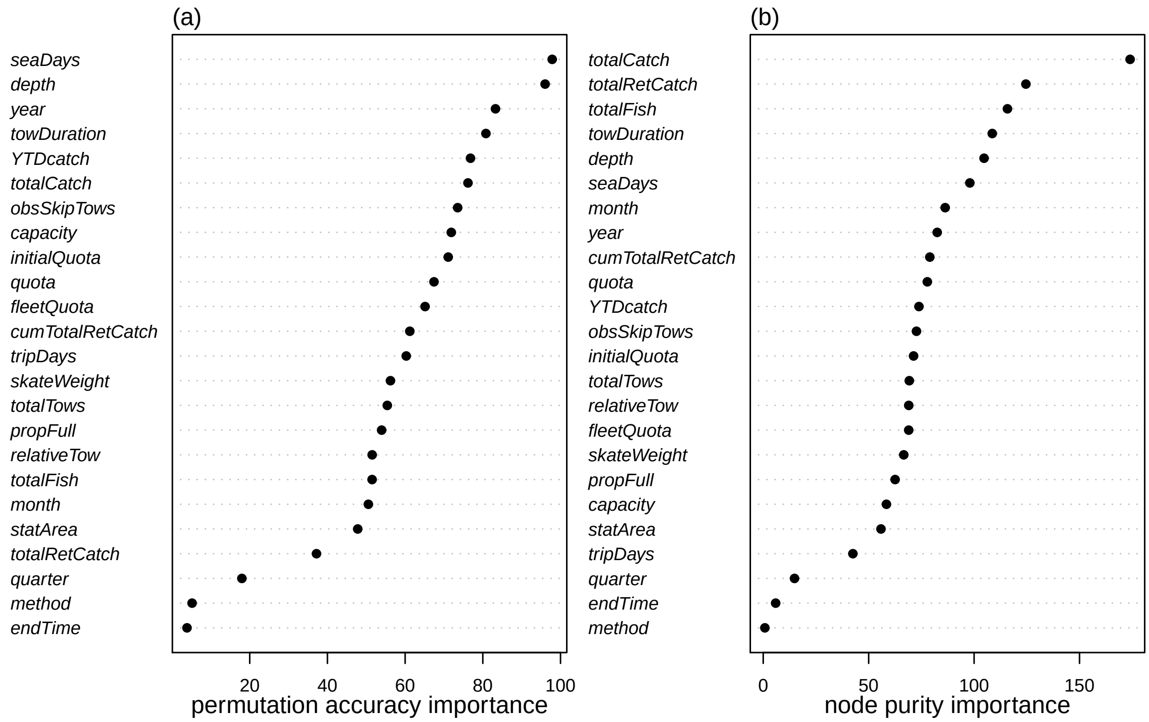

Figure 3.3: Halibut predictors are ordered from top (most important) to bottom (least important) according to their relative importance measured via permutation accuracy (a) and node purity (b).

proportion DR halibut y_i^{Pred} decreases less for halibut than sablefish, from 0.358 ($\pm$0.017) for observers with less than 27 *seaDays*, to between 0.292 and 0.311 (range) for observers with at least 54 *seaDays* (Figure 3.5a).

In both analyses, *seaDays* ranges up to 1,303 days, and a disproportionate number of events have observers with low *seaDays*. For example, more than 10% of events in the BC trawl are monitored by observers with less than 30 *seaDays*. Because many events are monitored by observers with low *seaDays*, and *seaDays* has a strong and consistent effect across species, human factors appear to be important in reported MR sablefish and DR halibut.

The effect of *obsSkipTows*, a measure of observer-skipper familiarity, indicates that higher halibut mortality is generally reported when observers and skippers have more events in common. Although *obsSkipTows* ranges up to 390 events, 50% of events have observers and skippers with less than 19 events in common, and only 10% of events have observers

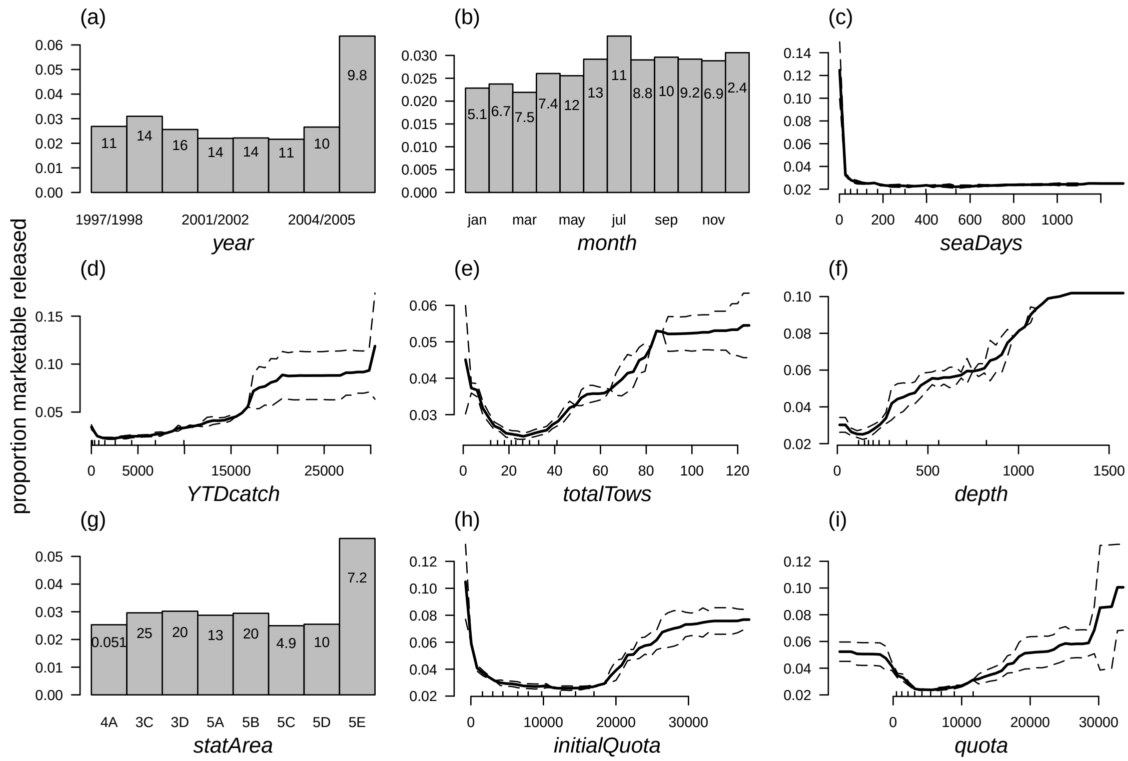

Figure 3.4: The marginal dependence for the nine most important sablefish predictors, measured via permutation accuracy, is indicated by bar plots for categorical predictors and line plots for continuous predictors. The y-axis indicates the mean predicted proportion marketable released sablefish. Numbers in bar plots indicate the percentage of observations per level, and tick marks along the x-axis in line plots indicate deciles. No uncertainty estimates are available for categorical predictors; dashed lines indicate two times the prediction standard error for continuous predictors.

and skippers with more than 63 events in common. The mean proportion DR halibut initially decreases rapidly from 0.312 (±0.001) for events with 1 *obsSkipTows*, to a minimum of 0.298 (±0.001) for events with 25 *obsSkipTows*, and then increases gradually to 0.337 (±0.010) for events with 390 *obsSkipTows*.

Sablefish and halibut release rates are also influenced by temporal, geographical, and event specific predictors, and most predictors have a non-linear relationship to release rates. More MR sablefish is reported for fishing events that occurred in the 2005/2006 fishing *year*, when the sablefish *YTDcatch* is high, at deeper *depth*, and in *statArea* 5E. More DR halibut is reported for fishing events that are at deeper *depth*, have longer *towDuration*, and larger

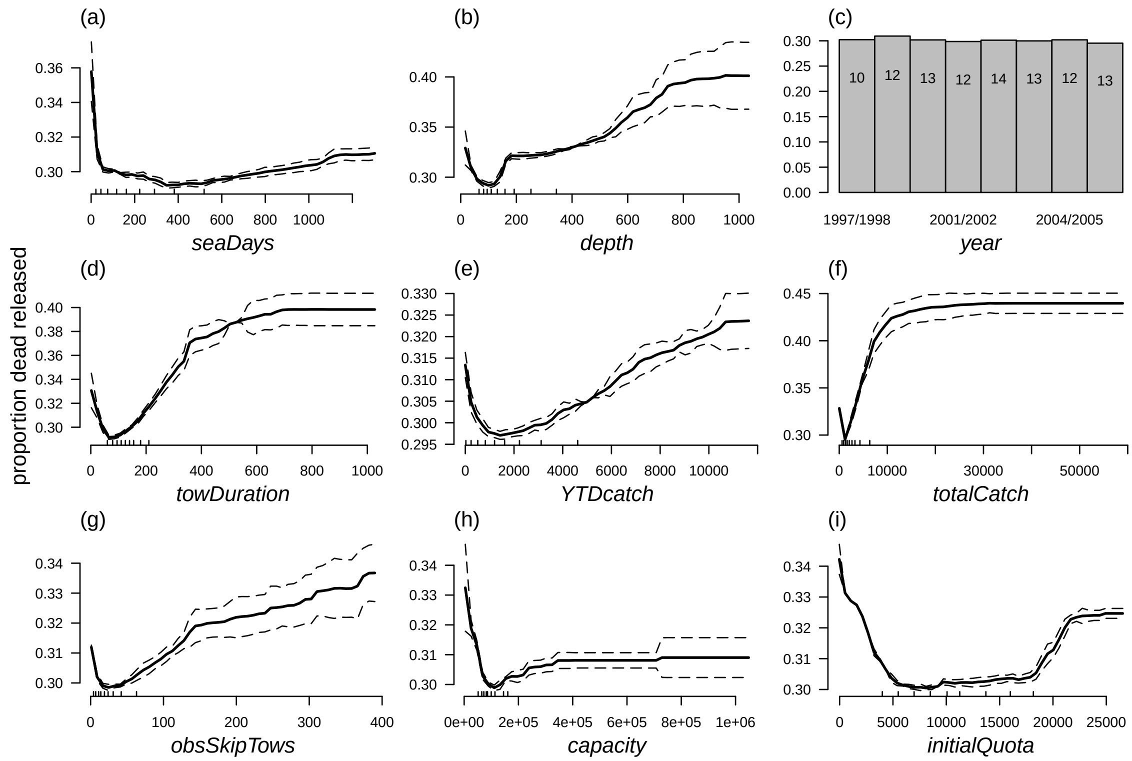

Figure 3.5: The marginal dependence for the nine most important halibut predictors, measured via permutation accuracy, is indicated by bar plots for categorical predictors and line plots for continuous predictors. The y-axis indicates the mean predicted proportion dead released halibut. Numbers in bar plots indicate the percentage of observations per level, and tick marks along the x-axis in line plots indicate deciles. No uncertainty estimates are available for categorical predictors; dashed lines indicate two times the prediction standard error for continuous predictors.

totalCatch.

3.5 Predictor pair correlation

Spearman's ρ indicates that most predictor pairs have low correlation (Table 3.2). The most correlated predictors are *totalCatch* and *totalRetCatch* with $\rho = 0.91$, $p < 0.001$ and $\rho = 0.89$, $p < 0.001$ for sablefish and halibut, respectively. This is expected because *totalRetCatch* is a subset of *totalCatch*. Because these predictors are strongly correlated, I investigate their combined effect on the proportion MR sablefish and DR halibut. The

Table 3.2: Spearman's rank correlation coefficient ρ quantifies predictor pair correlation. Only correlations greater than 0.5 ($|\rho| > 0.50$) are shown and all are significant ($p < 0.001$). Some predictors do not apply to the halibut analysis (NA).

Predictor pair	ρ (sablefish)	ρ (halibut)
totalCatch : *totalRetCatch*	0.91	0.89
propFull : *cumTotalRetCatch*	0.86	0.86
totalFish : *totalReleased*	0.75	NA
relativeTow : *propFull*	0.75	0.75
relativeTow : *cumTotalRetCatch*	0.73	0.73
tripDays : *totalTows*	0.70	0.62
thornyhead : *depth*	0.68	NA
thornyhead : *towDuration*	0.63	NA
totalFish : *retained*	0.60	NA
retained : *depth*	0.58	NA
thornyhead : *tripDays*	0.54	NA
towDuration : *depth*	0.50	$\|\rho\| < 0.50$

bivariate marginal dependence of *totalCatch*:*totalRetCatch* has the same direction as the individual effects for sablefish and halibut (Figures A.5 & A.6, respectively). It does not appear that there is an unexpected interaction between these two predictors.

3.6 Observer reliability index

To get good indices of observer reliability and skipper equitability, each individual should have out-of-bag (OOB) events in most of the $k = 500$ trees. Most observers and skippers appear in most trees (Table 3.3). For example, about 87% of sablefish skippers have 500 estimated effects from the LME analysis. For skippers with less than 500 LME effects, there is one skipper with no LME effects; the median number of effects is 0.92×500.

Although some observers have non-zero reliability indices, the 90$^{\text{th}}$ percentile range of observer effects overlaps zero for most sablefish and halibut observers (Figures 3.6a & 3.7a). Compared to the average observer, observers with relative reliability indices that overlap zero

Table 3.3: Number of observer reliability and skipper equitability indices. For individuals with less than $k = 500$ indices, numbers indicate the minimum, 25$^{\text{th}}$ percentile and median proportion of k indices. For both species, there is one skipper with no equitability indices.

		Proportion of k indices		
Individual	k indices	Minimum	25$^{\text{th}}$	Median
Sablefish observers	0.88	0.36	0.77	0.95
Halibut observers	0.87	0.34	0.87	0.98
Sablefish skippers	0.87	0.00	0.38	0.92
Halibut skippers	0.84	0.00	0.72	0.92

do not have tendencies to consistently over- or under-report (i.e., misreport) MR sablefish or DR halibut. The misreported weight is defined as the difference between the reported and the predicted weight of MR sablefish or DR halibut. There are more observers with non-zero reliability indices for halibut than sablefish, and there are more observers with negative, as opposed to positive, reliability indices for both species. For example, two sablefish observers (0.6%of the 322 sablefish observers, e.g., code *bx*) have positive reliability indices, and eight sablefish observers (2.5%, e.g., code *af*) have negative reliability indices. For halibut, 20 observers (6.2%) have positive reliability indices, and 27 observers (8.3%) have negative reliability indices. Observers with negative reliability indices may have tendencies to under-report MR sablefish. The effect size indicates the degree of under-reporting; a sablefish observer effect of -0.05 indicates the observer tends to under-report the proportion MR sablefish by 5% compared to the RF predictions.

Effect magnitude for observers with non-zero effects is quantified by the misreported weight of MR sablefish or DR halibut (Figures 3.6b & 3.7b). For example, observer *af* has a median effect of -0.014, indicating he or she tends to report the proportion MR sablefish 1.4% less than expected. The effect's magnitude is quantified by the median weight under-reported, which is 415 kg. In some cases, one single event can influence the observer's reliability index. For example, observer *bx* has reported a total of 6,119 kg of MR sablefish, and has a median over-reported weight of MR sablefish equal to 1,472 kg. Although observer *bx* has been on 24 trips that encountered sablefish, about 43% of the reported 6,119 kg of

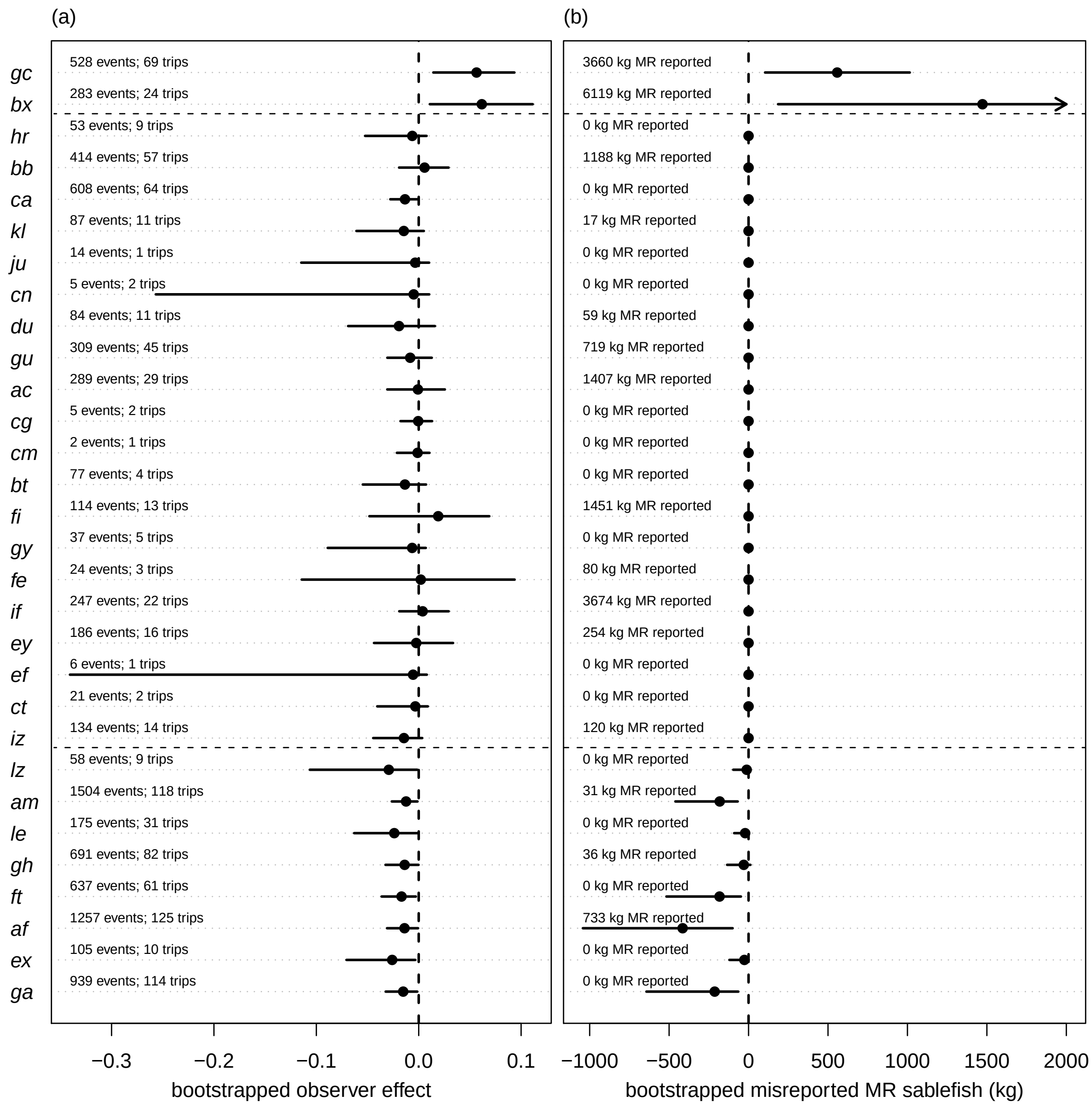

Figure 3.6: Observer effect (a) and magnitude (b) for selected sablefish observers (e.g., code *af*). Effect quantifies reporting tendency, and magnitude quantifies the over- or under-reported weight (i.e., misreported in kilograms, kg) of marketable released (MR) sablefish. The misreported weight is the difference between the reported and the predicted weight. Dots show median estimates and horizontal lines show uncertainty (90^{th} percentile range). The number of events, trips, and total reported MR sablefish is indicated for each observer. Horizontal dashed lines stratify graphs into three types of observer effects: positive (top), overlapping zero (middle), and negative (bottom). A random selection of observers are shown for the middle strata. The x-axis in Panel (b) is restricted to $\leq 2{,}000$ kg even though observer bx's upper 90^{th} percentile range is 3,832 kg. Some observers (e.g., codes *af*, bx) are also in Figure 3.7.

MR sablefish are from one single trip, and about 37% of the 6,119 kg is from one single event.

Some observers have consistent tendencies for both sablefish and halibut, while other observers have inconsistent tendencies across species. For example, observer *bx* has a positive effect for both the reported proportion MR sablefish and the proportion DR halibut. On the other hand, observer *af* has a negative effect for sablefish, and a positive effect for halibut.

Statistical theory predicts that 10% of observers should have significant effects when a 90$^{\text{th}}$ percentile range quantifies significance, as is the case for this analysis. For sablefish, only 3.1% of observers have statistically significant effects, which is less than expected by chance. For halibut, 14.5% of observers have statistically significant effects, which is almost 50% more individuals than expected by chance alone.

3.7 Skipper equitability index

The skipper equitability index quantifies unexpected economic costs or benefits that skippers have received; for example, the weight of MR sablefish that was not accounted for with quota. These economic costs or benefits are due to misreported weight, calculated as the difference between reported and predicted releases for events associated with observers who have non-zero reliability indices. Since observers, not skippers, typically report releases, misreporting typically indicates that events associated with the skipper have higher- or lower-than-expected releases reported, not that the skipper personally misreports the released weight.

Similar to observers, the 90$^{\text{th}}$ percentile range of misreported weight overlaps zero for most sablefish and halibut skippers. There are no sablefish skippers with positive equitability indices, but one sablefish skipper (0.8% of the 130 sablefish skippers) has a negative equitability index, indicating he may have had less-than-expected MR sablefish deducted from his quota (Table 3.4). One halibut skipper (0.7%) has a positive equitability index, and five halibut skippers (3.6%) have negative equitability indices.

Per-skipper PE indicates that misreported MR sablefish and DR halibut weights are negligible compared to the reported weight that is deducted from their quota. For example, skipper *AI* may have had less-than-expected weight MR sablefish deducted from his quota; the median under-reported weight is -30.1 kg, a PE of -0.44% compared to the total weight

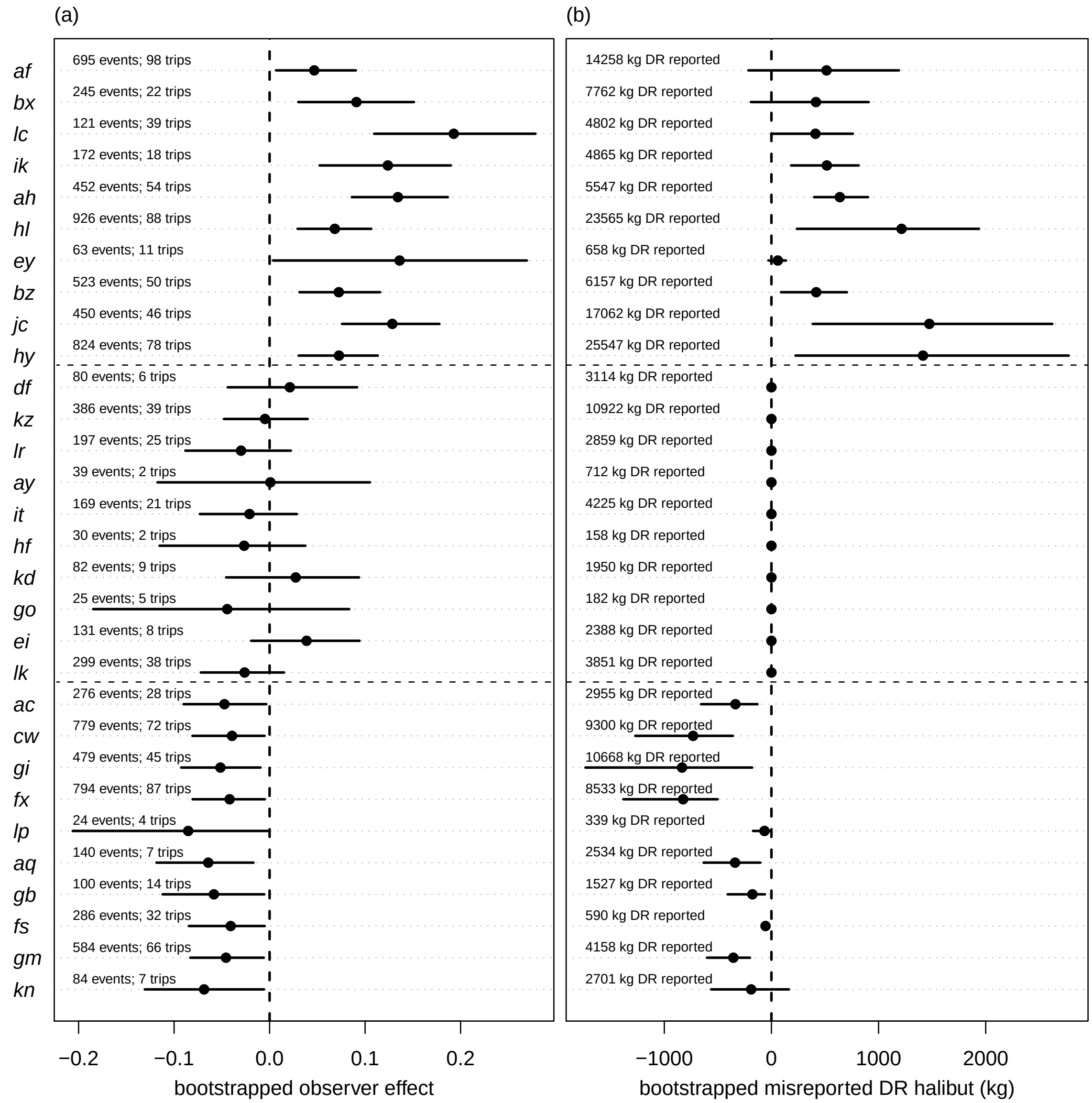

Figure 3.7: Observer effect (a) and magnitude (b) for selected halibut observers (e.g., code *af*). Effect quantifies reporting tendency, and magnitude quantifies the over- or under-reported weight (i.e., misreported in kilograms, kg) of dead released (DR) halibut. The misreported weight is the difference between the reported and the predicted weight. Dots show median estimates and horizontal lines show uncertainty (90^{th} percentile range). The number of events, trips, and total reported DR halibut is indicated for each observer. Horizontal dashed lines stratify graphs into three types of observer effects: positive (top), overlapping zero (middle), and negative (bottom). A random selection of observers are shown for each strata. Some observers (e.g., codes *af*, *bx*) are also in Figure 3.6.

Table 3.4: Per-skipper over- or under-reported (i.e., misreported) weight (kilograms, kg) and percent error (PE, %) of marketable released (MR) sablefish and dead released (DR) halibut, for skippers with non-zero equitability indices. The misreported weight is defined as the difference between the reported and the predicted weight of MR sablefish or DR halibut. The total number of events trips is shown for each skipper (e.g., code *AI*). Reported weight total released (TR) includes unmarketable sablefish and live halibut. Misreported MR or DR weight is calculated using random forest predicted release rates for observers with non-zero reliability indices. The PE compares the weight misreported to the weight reported on all events. Estimated weight and PE is indicated via the median, with 90^{th} percentile range indicating uncertainty.

			Reported weight (kg)		Misreported weight (kg)			Misreported PE (%)		
Code	Events	Trips	TR	MR or DR	5^{th}	Median	95^{th}	5^{th}	Median	95^{th}
Under-reported sablefish										
AI	1,131	82	62,187	6,883	-169.7	-30.1	-0.6	-2.47	-0.44	-0.01
Over-reported halibut										
BB	120	14	3,814	1,179	1.1	19.7	46.2	0.09	1.67	3.91
Under-reported halibut										
BK	1,557	157	94,528	24,713	-1,101.6	-361.9	-0.5	-4.46	-1.46	0.00
CA	899	88	54,951	14,038	-891.8	-435.6	-114.2	-6.35	-3.10	-0.81
BR	959	136	55,495	16,392	-2,111.7	-395.1	-86.4	-12.88	-2.41	-0.53
BM	1,598	150	116,044	33,441	-1,613.8	-613.8	-174.8	-4.83	-1.84	-0.52
AV	1,435	142	109,731	32,116	-859.4	-415.3	-35.4	-2.68	-1.29	-0.11

of MR sablefish reported on all fishing events, which is 6,883 kg. Compared to the total released (TR) weight, which include unmarketable sablefish or live halibut, misreported weights are slight.

3.8 Yearly and total misreported releases

Observers having non-zero reliability indices may have misreported MR sablefish and DR halibut. The median yearly misreported weight of MR sablefish tends to be around zero, and is typically less than 0.75 metric tonnes (t, Figure 3.5a). However, two years (e.g., 1998/1999 and 1999/2000) have higher-than-normal over-estimated weight of MR sablefish. For example, the upper 90th percentile of over-reported weight in 1999/2000 is 2.4 t, or 14.89% of the total reported weight of MR sablefish, which is 16 t. In general, more halibut than sablefish are misreported, but halibut misreported PE is lower because observers report much more DR halibut than MR sablefish. There does not appear to be a temporal trend of misreporting for either species. With the exception of sablefish in 2001/2002, the 90th percentile range overlaps zero each year for both species, indicating that yearly misreported weight is not significantly different from zero.

Total misreported MR sablefish and DR halibut overlaps zero, indicating that it is not significantly different from zero. The PE misreported indicates that the lowest estimate of total under-reported weight is negligible compared to the total reported weight of MR sablefish and DR halibut by all observers. For example, a total of 127 t of MR sablefish was reported by all observers. The 5th percentile under-reported, which is the lowest estimate, is -1.0 t; this is a PE of -0.76% compared to the reported weight by all observers. Again, compared to TR sablefish and halibut weights, misreported weights are negligible, and not significantly different from zero, for both species.

3.9 Observer reliability index correlation

Spearman's ρ determines sablefish and halibut observer effect correlation for the 319 observers common to both analyses. Although each bootstrap sample draws 319 effect pairs, bootstrap sample sizes range as low as 296 because some observers are not present in every LME model (median = 311). Bootstrap ρ standard deviation stabilizes at approximately

Table 3.5: Yearly and total over- or under-reported (i.e., misreported) weight (metric tonnes, t) and percent error (PE, %) of marketable released (MR) sablefish and dead released (DR) halibut. The misreported weight is defined as the difference between the reported and the predicted weight of MR sablefish or DR halibut. Reported refers to the total reported MR or DR weight for all observers for all years considered. Reported total released (TR) weight includes unmarketable sablefish and live halibut. Misreported MR or DR weight is calculated using random forest predicted release rates for observers with non-zero reliability indices. PE compares the misreported weight to the reported weight by all observers. Estimated weight and PE is indicated via the median, with the 90th percentile range indicating uncertainty.

	Reported weight (t)		Misreported weight (t)			Misreported PE (%)		
Year	TR	MR or DR	5th	Median	95th	5th	Median	95th
Sablefish								
1997/1998	332	10	-0.2	0.0	0.2	-2.45	0.16	1.90
1998/1999	353	19	-0.1	0.6	1.8	-0.80	3.28	9.52
1999/2000	424	16	-0.2	0.2	2.4	-1.15	1.22	14.89
2001/2002	351	9	-0.5	-0.1	0.0	-5.26	-1.50	-0.50
2002/2003	528	20	-0.6	-0.1	0.0	-3.04	-0.65	0.26
2003/2004	334	9	-0.6	-0.1	0.1	-6.95	-1.01	1.63
2004/2005	246	21	-0.4	-0.1	0.2	-1.76	-0.26	0.97
2005/2006	154	24	-0.3	0.1	0.5	-1.43	0.23	2.23
All years	2,722	127	-1.0	0.8	3.1	-0.76	0.65	2.44
Halibut								
1997/1998	331	99	-1.3	-0.3	0.6	-1.34	-0.28	0.64
1998/1999	362	116	-0.7	0.3	1.1	-0.63	0.26	0.98
1999/2000	370	106	-0.9	0.2	0.9	-0.82	0.15	0.82
2001/2002	349	99	-1.3	-0.1	0.8	-1.34	-0.11	0.81
2002/2003	477	142	-1.8	-0.4	1.1	-1.28	-0.30	0.74
2003/2004	519	152	-1.9	-0.3	0.9	-1.26	-0.21	0.58
2004/2005	474	143	-1.0	0.0	1.1	-0.72	-0.02	0.74
2005/2006	660	194	-3.4	-1.3	0.7	-1.77	-0.65	0.37
All years	3,542	1,051	-5.8	-2.2	1.5	-0.55	-0.21	0.14

0.06 after 30,000 iterations. The 90th percentile range of 100,000 bootstrap correlation coefficients overlaps zero ($\rho_{0.05}$ = -0.03, $\rho_{0.95}$ = 0.16) and indicates that sablefish and halibut observer effect correlation is not significantly different from zero (ρ_{median} = 0.07, Figure 3.8a). Although there is no statistically significant tendency for observers who under-report MR sablefish to also under-report DR halibut, 88.0% of the 100,000 bootstrap ρ values are greater than 0.00. About 12% of the 319 observers have positive median effects for both the proportion MR sablefish and the proportion DR halibut (i.e., dots in the upper-right quadrant). Conversely, about 40% of observers have negative median sablefish and halibut effects (i.e., dots in the lower-left quadrant).

Considering only the median of each observer's 500 effects, sablefish and halibut observers have different tendencies. About 79% of sablefish observers and 49% of halibut observers have negative median effects (Figure 3.8b & c, respectively). Most sablefish observers have negative LME effects (i.e., tendency to reduce the proportion MR sablefish) more than 50% of the time, even though the 90th percentile range straddles zero for all but eight sablefish observers. Median sablefish observer effects range between -0.03 and 0.17. For halibut, median observer effects have a larger range (between -0.18 and 0.38) and are distributed relatively symmetrically about zero.

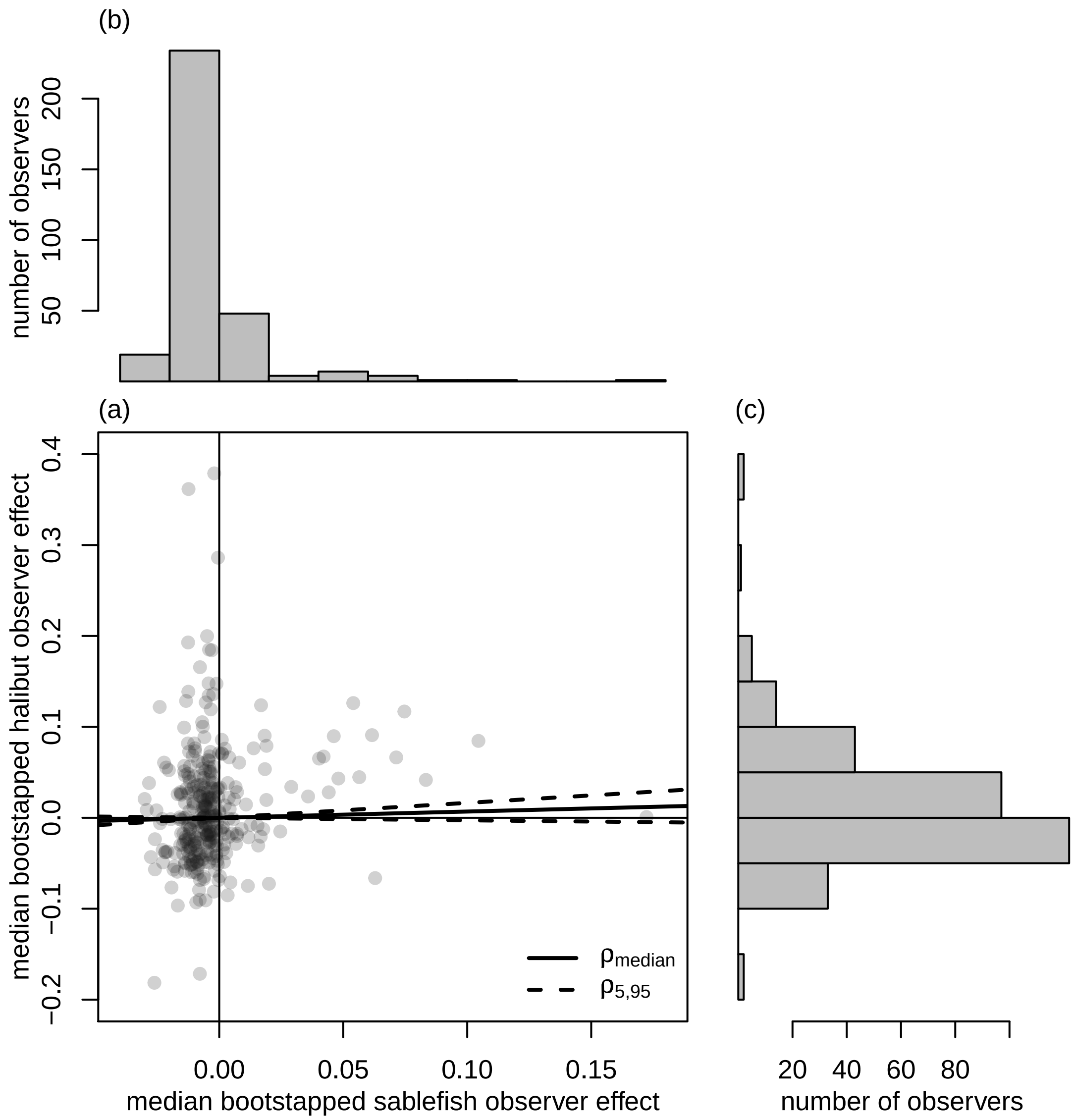

Figure 3.8: Median bootstrapped sablefish and halibut observer effects for the 319 observers common to both analyses (a). Dots are semi-transparent to help visualize density. The thick solid line is the median of the 100,000 bootstrapped Spearman rank correlation coefficients ρ, and thick dashed lines indicate the 5^{th} and 95^{th} percentiles. Thin solid lines delineate quadrants. Histograms show marginal distributions of median sablefish and halibut observer effects (b & c, respectively).

Chapter 4

Discussion

It is essential to evaluate the reliability and equitability of releases reported by at-sea observer programs (ASOP). Reliable observer release reports are important for accurate estimates of total extraction, while equitable harvester treatment is important for individual transferable quota (ITQ) management programmes. The assumptions of reliable and equitable release reports have been tested using the British Columbia offshore groundfish trawl fishery (BC trawl) as a case study. Reliability is investigated by quantifying each observer's tendency to over- or under-report (i.e., misreport). Misreported weight is quantified for each observer, and for the entire fleet by year and over all years. Equitability is investigated by quantifying the unexpected weight that each skipper has had to account for with quota. These results are interpreted by discussing important predictors, and their marginal effects on reported releases. The analysis does not find strong evidence to suggest that these assumptions are incorrect for the BC trawl.

Observer reliability is quantified by the tendency and the weight of misreported releases. Again, the misreported weight is defined as the difference between the reported and the predicted weight of MR sablefish or DR halibut. Although I am unable to confirm that observers with negative reliability indices actually under-report releases, these observers may have a tendency to under-report releases compared to the average observer. Therefore the observer reliability index is a relative metric of reliability, and assumes that the average observer is reliable. In this study, only a small proportion of observers have significant reliability indices, and even for theses cases, the misreported weight is negligible. Overall reliability is also quantified by the yearly and total misreported releases, which are typically not significantly different from zero. Therefore, the analysis does not find strong reasons to suspect that ASOP release data are unreliable.

Equitable treatment of each skipper is important in ITQ management systems. Each skipper's equitability index is quantified by the misreported weight and percent error (PE) of MR sablefish or DR halibut associated with their events. The equitability index indicates unexpected costs or benefits from accounting for more- or less-than-expected releases with quota, respectively. As with observers, only a small proportion of skippers have significant equitability indices, and the misreported weight is negligible. Again, the analysis does not find strong reasons to suspect that ASOP release data are inequitable.

Release rates in the BC trawl are affected by various predictors, which have been quantified by their relative importance and the direction of their effect. Among other predictors, the proportion marketable released (MR) sablefish is affected by temporal and area effects, observer experience, the year-to-date sablefish catch, and sablefish quota. For halibut, the proportion dead released (DR) is affected by observer experience, the depth and duration of the event, temporal effects, the weight of the catch, and the degree of observer-skipper familiarity. Further research may be required to understand the reason for the importance and direction of effect for these predictors (e.g., observer experience), which could influence the conclusion regarding the assumption of ASOP reliability and equitability.

4.1 Reliability and equitability status

Most observers and skippers are associated with events that have either: (1) accurate release reports; or (2) considerable variability in predicted release rates with no tendency to over- or under-report. Accurate release reports are events where reported release rates agree with predicted release rates. However, a few individuals are associated with events that have consistently either over- or under-reported releases. An observer who consistently under-reports releases may under-estimate total extraction, which could affect yearly and total estimates of total extraction. Skippers associated with under-reported MR sablefish or DR halibut may have an economic advantage over other skippers. These events may be of concern, however there are alternative interpretations of observer reliability and skipper equitability.

4.1.1 Observer reliability

One component of ASOP release reports is their reliability, assessed as the tendency and weight of misreported MR sablefish and DR halibut. A small number of observers report either significantly more or less MR sablefish or DR halibut than the average observer in similar circumstances. However, there is no significant tendency of misreporting across species for observers that have reported both sablefish and halibut. Thus, observers that tend to under-report MR sablefish are not more or less likely to under-report DR halibut. Observers who tend to misreport may have actually influenced harvester behaviour, or may provide less reliable data. Both explanations assume a financial benefit to harvesters from lower reports of MR sablefish and DR halibut, which could be due to either a change in harvester behaviour, or a change in observer behaviour.

A change in harvester behaviour could occur when observers influence harvesters to retain, as opposed to release, marketable sablefish (S. Buchanan, pers. comm., 2006). For example, an observer may immediately inform the harvester that marketable-sized sablefish are being released and that these releases will be deducted from their quota. This early and clear communication may alter the harvester's behaviour to retain these marketable sized pieces. For halibut, an observer who informs the harvester of high DR halibut may alter sorting and handling procedures, thereby increasing halibut survival. Other observers may be less self-confident and more likely to avoid confrontation, simply recording releases without giving the harvester the opportunity to change his behaviour. In this case, the uninformed harvester does not have an opportunity to change behaviour. Observers altering harvester behaviour is not a management concern because harvester behaviour to reduce releases supports Fisheries and Oceans Canada's (DFO) goals.

Observer behavioural changes include observers who actively help the crew sort sablefish into marketable and unmarketable portions, and help the crew return halibut to the water quickly. Observers who help the crew may decrease MR sablefish and DR halibut. A change in observer behaviour could also occur because of harvester pressure to avoid financially damaging release reports. This second explanation has serious management implications regarding reliability. Harvesters who alter observer behaviour are a concern because release reliability is compromised and possibly under-estimated. Also, the equitability of the ITQ programme could be compromised if some harvesters alter observer behaviour more than others, and harvesters perceive the ITQ as being unfair.

4.1.2 Magnitude of yearly and total misreported release weight

It is important that yearly and total misreported weight is reliable, otherwise misreporting could influence stock assessments, with potential long-term negative impacts on fish stocks. Yearly and total misreported weight is typically not significantly different from zero because the over- and under-reported weights of MR sablefish and DR halibut are of similar magnitude. The only exception is 2001/2002, when the under-reported weight of MR sablefish is significantly less than zero. However, the magnitude of under-reported MR sablefish for 2001/2002 is negligible compared to the total reported weight of MR sablefish (i.e., median under-reported weight of -0.1 metric tonnes, t compared to 9 t of reported MR sablefish).

Harvesters release over 50% of the total sablefish catch. Less than 5% of these released fish are of marketable size, and only a proportion of these MR sablefish are considered to be dead. Stock assessments do not account for unmarketable released sablefish mortality, and harvesters are not required to account for these unmarketable fish with quota. Even though the DFO handles unmarketable sablefish mortality and quota in this manner, these fish likely have partial release mortality and thus contribute to total extraction. Therefore, a small misreported weight of MR sablefish likely has negligible effects on estimates of total extraction compared to the ignored mortality of unmarketable releases.

As with sablefish, yearly and total over- and under-reported weights of DR halibut are of similar magnitude, and are not significantly different from zero. However, the misreported weight is slightly skewed towards under-reporting: the median total misreported weight is -2.2 t (90^{th} percentile range is between -5.8 and 1.5 t). Unlike misreported MR sablefish, misreported DR halibut is a direct estimate of unaccounted mortality imposed by the BC trawl. However, this weight is slight compared to total reported DR halibut of 1,051 t, and is unlikely to impact stock assessments or skipper equitability.

For some fish species, unmarketable and marketable sized fish may contribute differently to the impact on total extraction because of their different release-induced mortality rates, life-history and population parameters, and relative proportion of available biomass (Kennelly *et al.*, 1998). Also, there are more unmarketable fish than marketable fish in a given weight because of their smaller size. These factors must be considered when assessing the relative impact of misreported releases. For example, releases may have a small impact on total extraction if release induced mortality rates are low, natural mortality rates are high, and a small proportion of the biomass is vulnerable to fishing. Because the BC trawl primarily captures small halibut that would be considered unmarketable by the directed hook and line fishery (Myhre, 1969; Stanley, 1984), mortality caused by the BC trawl may have a different impact on the halibut stock than mortality caused by the hook and line fishery.

Misreported weights of MR sablefish and DR halibut are slight, and generally not statistically different from zero. Compared to their respective directed hook and line sablefish and halibut fisheries, the BC trawl's sablefish total allowable catch (TAC) and halibut total allowable by-catch (TAB) are less than 10% of the entire Canadian TACs. Because the BC trawl's quota is a small proportion of the entire Canadian TAC, small over- or under-estimates of total extraction likely have little impact on these stocks. However, simulation studies may be required to quantify the effect of small misreported weights on stock assessments.

4.1.3 Skipper equitability

Another component of ASOP release reports is their equitability, assessed as the unexpected cost or benefit incurred by skippers. These costs or benefits are equal to the weight of unexpected MR sablefish and DR halibut that they have accounted for with quota. Although misreported weights and PEs are small on the fishery scale, small misreported weights could have larger impacts for individual skippers in multi-species fisheries such as the BC trawl that are managed by ITQs. Two examples demonstrate the importance of equitable harvester treatment, and show that small misreported weights have the potential for larger consequences than simply the value of misreported fish.

First, accurate release reports are important in multi-species fisheries when harvester activity can be constrained by mismatches between catches and quotas. Harvesters with under-reported MR sablefish may be able to continue fishing for other species over the fishing year, while those with over-reported MR sablefish may be forced to buy additional quota or stop fishing. For halibut, by-catch quota is essential for all bottom-trawl vessels targeting benthic species. Under-reported DR halibut could enable harvesters to continue to target benthic species without approaching halibut by-catch quota limits. Vessels are restricted to a maximum proportion of the yearly halibut TAB and sablefish TAC, and could therefore potentially enter a situation in which they are not allowed to purchase more quota due to licence restrictions. However, most harvesters typically have surplus sablefish and halibut quota at the end of the year, and carry sablefish quota forward to the next year (unused halibut quota can not be carried forward). Because of these quota surpluses, it is likely rare that harvesters enter situations where they run out of quota.

Second, ITQ management program success may be influenced by whether harvesters perceive the system as being equitable (Branch, 2004). Quota trading value for ITQ species may rely on harvester's perception that they are unable to secretly release restrictive quota species without penalty (Stanley *et al.*, 2009). If some harvesters are forced to purchase quota for restrictive species, while other harvesters are able (or even perceived as being able) to secretly dump these species, the quota trading system and quota value could deteriorate. This scenario may occur in the trawl fishery, as harvesters who run out of quota for these species are either restricted to midwater-trawling, or required to purchase quota from other harvesters. As previously mentioned, this scenario is likely rare, but could have large consequences if it occurs. The analysis does not suggest that the assumption of harvester equitability is incorrect.

Misreporting does not appear to be a problem in the BC trawl because: (1) a small proportion of observers misreport; (2) misreported weights are negligible; (3) over- and under-reported weights are of similar magnitude; and (4) a small proportion of skippers have misreported MR sablefish and DR halibut, and the misreported weight is negligible. Also, harvester incentives to under-report may be low because harvesters are generally able to stay within yearly sablefish and halibut quota constraints. For sablefish, less than 5% of observers misreport, which is less than expected by chance alone. For halibut, about 15% of observers misreport, which is more than expected by chance alone. The analysis does not provide strong evidence that misreporting compromises the reliability of yearly and total extraction estimates, or skipper equitability. However, because even small discrepancies in equitability could have large impacts in multispecies ITQ fisheries, the conclusion regarding harvester equitability must take these considerations into account.

4.2 Predictors affecting release rates

Some predictors that affect release rates in the BC trawl are expected, and some are unexpected. Generally, releases are affected by temporal effects, observer experience, and fishing event depth, duration, and area. Other predictors that influence BC trawl release rates are catches and quotas, and the degree of observer-skipper familiarity. Observer experience is one of the most important predictors affecting reported rates of both MR sablefish and DR halibut. For both species, observers with more experience report lower release rates than new observers. The consistent importance and direction of effect of observer experience for both species indicates that human factors cannot be ruled out as an influence on reported release rates. Determining the reason for the importance and direction of effect of observer experience is essential to conclusions about observer reliability.

There are several interpretations of the "observer experience" effect. New observers may over-estimate smaller MR and DR weights more than larger total released (TR) weights. This bias could be unconscious, perhaps due to observers being new and unused to estimating proportions of marketable and unmarketable, or dead and live. Experienced observers may report less MR and DR fish as they gain experience. On the other hand, the bias could result from harvester pressure to under-report financially damaging release reports. Over time, observers could report lower release rates to facilitate their career among a relatively small network of harvesters. Yet another explanation is that experienced observer may be more likely to communicate with harvesters and influence sorting procedures (S. Buchanan, pers. comm., 2006) than new observers.

The reasons for the direction of the observer experience effect have different consequences on observer reliability. The first two interpretations, in which observer reporting changes, are problematic for the ASOP. ASOPs with high observer turn-over rates could be affected significantly if new observers are less accurate. Even if new observers have less reliable release estimates for a short period of time, a large proportion of events are monitored by relatively new observers. If experienced observers report lower releases due to skipper pressure, the perception of unreliable or inequitable releases may influence quota trading and quota value in ITQ fisheries (Stanley *et al.*, 2009). The third interpretation is not a concern because observers who alter harvester behaviour to reduce releases is one of the benefits of an ASOP.

Skippers have economic incentives to under-report releases, and many BC trawl vessels operate 24 hours per day. For these reasons, skeptic's of the veracity of the BC trawl's ASOP have questioned the reliability of release estimates for events that are unmonitored (i.e., when the observer sleeps) and for events at night (S. Cox, pers. comm., 2009). Despite these concerns, these two predictors have the lowest importance of the predictors examined, possibly because observers monitor over 98.5% of events.

4.2.1 Predictors affecting marketable released sablefish

Higher proportions of MR sablefish are reported for 2005/2006, possibly because of increased market demand for arrowtooth flounder (*Atheresthes stomias*) that same year. Between 1997/1998 and 2004/2005, the BC trawl landed between 2,557 t and 6,643 t of arrowtooth flounder per year (range, median = 4,328 t). However, in 2005/2006, the BC trawl targeted arrowtooth flounder, and increased landings to 16,836 t. Because sablefish is frequently encountered in association with arrowtooth flounder, increased arrowtooth flounder catch likely resulted in increased sablefish bycatch. However, there was no corresponding increase in sablefish quota, so harvesters may not have had sufficient yearly sablefish quota to retain the entire sablefish catch. The discrepancy between sablefish catch and quota may have caused harvesters to selectively release more marketable-sized sablefish to conserve sablefish quota. This argument is supported by the trend of increased MR sablefish with increasing cumulative year-to-date sablefish catch.

Month of the year is also an important predictor for the proportion MR sablefish. Sablefish release rates are high early in the fishing year (i.e., April through December), possibly because harvesters are pro-active in their strategy to avoid low sablefish quota situations. This strategy is opposite to what is intuitive, whereby one might expect harvesters to release more MR sablefish later in the year as a reaction to a low quota situation. Although harvesters actively avoid areas of high sablefish density, harvesters may release more sablefish early in the year to avoid entering a situation where they run out of sablefish quota later in the year (S. Buchanan, pers. comm., 2006). This strategy may allow harvesters to retain sablefish later in the year (i.e., January through March) when they are more confident that they will not run out of sablefish quota. Harvesters with unused sablefish quota can carry quota forward to the next fishing year.

Observers also report more MR sablefish at deeper depths and in management area 5E. Area 5E does not have a wide, shallow shelf that is present in other management areas. This area is also where harvesters target deep-water thornyhead rockfish[1]. These harvesters may be forced to release more marketable sablefish in order to avoid sablefish quota constraints. Sablefish migration patterns from shallow in-shore water to deep off-shore waters (Maloney and Sigler, 2008) may contribute to the pattern of more MR sablefish with depth. Area 5E

[1]The presence of thornyhead rockfish is not an important predictor, but this effect may be confounded with other predictors. There is a significant correlation between the presence of thornyhead rockfish with deeper depths and longer event duration.

has high sablefish catch rates by the directed hook and line fishery (Springford, 2008).

4.2.2 Predictors affecting dead released halibut

Higher halibut mortality is associated with heavier catch weight, deeper depth, longer event duration, and observer and skipper familiarity. The effects of catch weight and event duration support Neilson *et al.* (1989) and Richards *et al.* (1995, 1994). However, halibut mortality is not typically affected by depth (Hoag, 1975; Neilson *et al.*, 1989; Richards *et al.*, 1994).

Halibut mortality may be high when catches are large because of increased sorting times and halibut damage in the net. It may take crew longer to sort through the catch, and release halibut back to the water when catches are large. Even a small increased time on deck increases halibut mortality (Richards *et al.*, 1995). Sorting time is not included in this analysis, but catch weight may serve as a proxy for sorting time because the two measures have a significant positive correlation (Richards *et al.*, 1995). Heavy catches could also increased damage to halibut if they are crushed in the net, or suffocated.

Long duration events may cause higher halibut mortality because halibut may tire and lose their ability to swim ahead of the cod-end, or back of the net. Tired halibut may get swept to the back of the net and become damaged by the rest of the catch, or smothered by inorganic debris such as mud. Because event duration only considers the time interval from when the net is in fishing position until net retrieval begins, the time required to lift the net from the bottom to the deck is not accounted for. Deeper events require additional time to lift the net, and the extended duration may increase halibut mortality. The association of higher halibut mortality with deeper depth observed in this analysis may also be due to other unmeasured predictors.

Generally, higher halibut mortality is found when observers and skippers have more familiarity, contrary to the relationship expected if observers and skippers formed beneficial relationships. For example, a relationship might influence the observer to communicate early to change harvester behaviour. This analysis indicates that reported DR halibut decreases initially by about 1% as observers and skippers have more events in common, and then increases by about 3.5% when they have more than 25 events in common.

4.3 Assumptions for assessing observer reliability

A major challenge to investigating the reliability of fisheries monitoring data is the lack of a true account of release activity. Typically, monitoring data are assumed to be accurate, and this assumption likely holds most of the time. However, there may be events for which monitoring data are not accurate. For these events, another account of release activity is required. Some models are well suited to predicting release rates, and these predictions can then be compared to reported releases. This analysis, in which I develop and apply a novel approach to this problem, makes several assumptions.

The most important assumption is that the random forest's (RF) predicted releases are accurate. I also assume that observers have unbiased and precise reports of the predictors (e.g., depth, month, area) used in the RF analysis, and the weight total released (TR) used in Equation 2.11 on page 32. These assumptions likely hold because there is no economic incentive to misreport the predictors or weight TR. Also, the ASOP service provider checks data for inconsistencies after each trip. Data inconsistencies are addressed by contacting the observer to ensure data accuracy, and making changes if required.

The observer reliability index compares each observer's reliability status to the average observer's reliability status in similar circumstances. However, there is no baseline with which to compare the indices of observer reliability. I compare indices to zero, and assume that a negative or positive index identifies an individual that is significantly different from the average. However, setting the baseline at zero is arbitrary, which makes it impossible to conclude that observers with negative indices are actually unreliable. For example, if all the observers are generally reliable but have a certain amount of bias, the analysis could identify observers that have slightly below average release reports with negative reliability indices. Alternately, if all the observers are completely biased, and tend to under-report release rates, the analysis may again identify observers that have slightly below average release reports in the same way. I assume that most observers are generally reliable, but have some variability and may have some bias.

4.3.1 Alternative analytical approaches

I assume that the RF analysis for predicting releases, and the LME analysis for quantifying observer effects make the best use of the available data. I flag observers that are statistically different, either above or below the average observer's reliability index, as possibly having misreported MR sablefish or DR halibut. Releases are considered reliable if the observer is not statistically different from the average observer in similar circumstances.

However, alternative methods could also be used to investigate monitoring programme data. Misreported weight could be calculated only for events associated with observers that have negative reliability indices. As with Lennert-Cody and Berk (2007), one could assume that misreporting is only under-reporting. Observers with positive reliability indices are actually reliable, but are at an extreme end of a spectrum of reliability indices. This method ignores observers who could willfully or accidentally over-report releases, which could have harmful effects on the fishery. Alternatively, misreported weight could be calculated for all events, regardless of the observer's reliability status.

Another approach to quantifying the misreported weight could use the observer's LME fixed effect. Thus, Equation 2.11 on page 32 would be replaced by

$$\delta_i \quad = \quad \left(y_i^{\mathrm{Rep}} + \mathrm{O}_l\right) \times \mathrm{TR}_i \text{ for } i = 1,\ 2,\ ...,\ N_{\mathrm{OOB_{NZ}}} \tag{4.1}$$

which gives a direct estimate of the misreported weight due to the observer's effect O_l. An advantage of the current method is that the misreported weight is estimated using the predicted weight, ignoring any potential observer or skipper effects.

4.4 Management considerations and implications

There does not appear to be strong evidence to suggest that the BC trawl's ASOP release data are unreliable or inequitable for their intended purpose. However, as with any statistical analyses, uncertainty in the inferred status could mask the true state. For example, research results can infer that ASOP release estimates are either reliable or unreliable (Table 4.1)[2]. Inferences have a probability of being either correct or incorrect given an unknown true state. If releases are correctly inferred to be unreliable (power), managements actions could be taken to improve reliability. Alternatively, if releases are incorrectly inferred to be unreliable (false positive), corrective action may be taken when none are required, possibly wasting financial resources. If releases are correctly inferred to be reliable, no action needs to be taken. Alternatively, if releases are incorrectly inferred to be reliable (false negative), no

[2]In this example, releases will be referred to as either reliable or unreliable. In reality, inferring the degree of release reliability may be more appropriate.

Table 4.1: Research results may infer that release reports are either reliable or unreliable. Because the true reliability state is unknown, the inferred status has a probability of being either correct or incorrect. The costs of false positive and false negative inferences may be unequal. Probabilities are given in parentheses. Note: adapted from Gotelli and Ellison (2004).

	Inferred status	
True state	Releases are reliable	Releases are unreliable
Releases are reliable	Correct ($1 - \alpha$)	False positive (α)
Releases are unreliable	False negative (β)	Correct, power ($1 - \beta$)

action will be taken, but unreliable release estimates could affect total extraction and skipper equitability.

Traditional statistical tests often set $\alpha = 0.05$ to reduce the probability of making a false positive inference (Gotelli and Ellison, 2004; Peterman and M'Gonigle, 1992). However, reducing α increases the probability of making a false negative inference, β (Gotelli and Ellison, 2004). The costs of false negative inferences can be very high in environmental decision making, and thus the trade off between α or β requires serious consideration (Brosi and Biber, 2009; Field *et al.*, 2004; Gotelli and Ellison, 2004; Kriebel *et al.*, 2001; Peterman and M'Gonigle, 1992). For example, if a false negative inference is made, under-estimated releases and biased stock assessments could allow unsustainable catches, resulting in reduced stocks over time. Sacrificing conservation in exchange for maximizing short-term profits can result in long term biological and economic losses (Agardy, 2000; Costello *et al.*, 2008), and unsuccessful stock rebuilding strategies (Kelleher, 2005). For this analysis, the power to detect unreliable releases (i.e., $1 - \beta$) should be quantified via retrospective power analysis before definitive conclusions are made regarding the reliability of the BC trawl's ASOP.

In addition to considering the financial and ecological costs and benefits of reliability inferences, harvester utility should also be considered. Harvesters may place a high value on the perception of an honest and reliable ASOP, or being treated fairly by observers. The perception of an honest ASOP may benefit fisheries if harvesters can assure seafood buyers and other fisheries that they are committed to having honest and accessible monitoring. In this case, a false positive inference could damage the ASOP's reputation. By definition, harvesters may also appreciate equitable treatment because they are assured that all harvesters are treated the same. Regardless of which management action, if any, is taken, the relative probabilities, costs, and benefits of correct and incorrect inferences must be considered.

4.4.1 Pathological indicators of reliability

Because observer reliability indices compare each observer to the average, they do not quantify underlying bias. In this case, it is important to discuss the presence or absence of symptoms that could indicate deliberate misreporting. First, about 44% of sablefish observers and 32% of sablefish skippers have never reported MR sablefish. Considering that harvesters may release marketable-sized sablefish early in the year as a strategy to avoid quota constraints later in the year, it is surprising that MR sablefish reports are so rare (i.e., less than 5% of events). These events may be rare because harvesters are able to avoid sablefish, or because harvesters are not typically constrained by sablefish quota, or both. Conversely, these events may be rare because harvester pressure dissuades observers from reporting MR sablefish.

Second, one might expect unreliable observers to be unreliable for both MR sablefish and DR halibut. For example, observers may under-report both species when personal tendencies or behaviours affect release reports. Although about 52% of observers have either both positive or both negative median effects for sablefish and halibut (e.g., negative median sablefish effect, and negative median halibut effect), the majority of these effects are not statistically significant. Contrarily, about 48% of observers have opposing tendencies for sablefish and halibut. A more detailed investigation of these observers may be useful to other projects that quantify ASOP release reliability.

Finally, results from the RF analysis indicate that reported proportion MR sablefish and DR halibut are not completely influenced by observer reporting. The RF analysis identifies several predictors that are important in determining the proportion MR sablefish and DR halibut, and the majority of these predictors have marginal effects that are consistent with expected relationships. For example, there are more MR sablefish reported at deeper depths, and more DR halibut reported for longer tow duration. If observers misreport due to harvester pressure, one might expect misreporting to be highest in the early years of the monitoring program when harvesters were not accustomed to accommodating human observers at sea. However, the important effect of *year* indicates relatively consistent observer reporting from the ASOP's inception to more recent years. In fact, *year* is important because of the higher MR sablefish in 2005/2006, which likely occurred because of harvesters targeting arrowtooth flounder, as described above.

Further qualitative investigation of the BC trawl ASOP reliability may be useful. For example, formal observer interviews or surveys may identify circumstances that cause pressure to under-report releases, if present. Likewise, skippers may identify the reason why new observers report more MR sablefish and DR halibut than experienced observers. However, interviews and surveys must account for various issues, such as the vested interest of individuals which may bias results.

4.4.2 Alternative applications for random forest

This study demonstrates several strengths of the RF analysis as a broadly applicable tool for fisheries management. For example, the RF analysis could be used to quantify the effect of management actions on angler activity in recreational fisheries. In the BC trawl, the RF may be used to identify conditions (e.g., environmental, spatial, temporal) associated with high catches of unmarketable sablefish or halibut. By weight, the BC trawl catches and releases more unmarketable- than marketable-sized sablefish, and these fish are likely subject to partial release-induced mortality. Annually, BC trawl releases amount to between 130 and 508 t (range) of unmarketable sablefish (median = 330 t). These unmarketable fish may impose direct costs on BC trawl harvesters from increased crew size and handling time required to sort the catch, as well as external costs on harvesters in the directed hook and line sablefish fishery from reduced stock size. Similarly, the yearly weight of dead released halibut by-catch in the BC trawl ranges between 99 and 194 t (median = 129 t). As with sablefish, these dead halibut may impose direct costs on BC trawl harvesters, as well as external costs on harvesters in the directed hook and line halibut fishery.

4.4.3 Management actions to improve release reliability

One management action to ensure that release estimates are reliable is complete at-sea monitoring. Hook and line fisheries, in which catch comes aboard one piece at a time, may achieve accurate release reports using electronic monitoring. Trawl fisheries on the other hand, in which catch comes aboard in large quantities composed of multiple species and sizes, likely require human observers (i.e., ASOPs). When ASOPs are used to estimate releases, elements of human nature can potentially affect the reliability of release reports. If release reports have the potential to be unreliable, periodic reviews using the methodology presented here are essential to ensure their reliability.

When results suggest that releases are generally reliable, but a small proportion of observers appear to be statistically different, groundtruthing may increase the credibility of the analysis. Lennert-Cody and Berk (2007) compared predicted observer reliability to a problem group of observers (i.e., observers who were suspected of under-reporting prior to the analysis) by working closely with the ASOP service provider. They found a relationship between predicted low reliability status and problem group observers, which added credibility to their results. In this case, correction factors could be developed for these observers to account for possible over- or under-reporting and ensure that harvesters are treated equitably. However, developing and applying correction factors could be difficult. For example, it may be difficult to convince harvesters of an expensive and inconvenient ASOP if release reports are inaccurate and likely to be modified after the fishing trip. When results suggest that releases are generally unreliable, or applying correction factors is infeasible, other management actions may be required to increase reliability.

Potential actions to improve the reliability of release estimates could either affect release reporting reliability via the observer, or affect the incentive to release fish via the harvester. Actions that may affect release reporting reliability include increased ASOP training and periodic refresher courses, multiple observers per vessel (e.g., on the first trip), and the addition of electronic monitoring. Increased ASOP training and refresher courses could focus on the importance of reliable release estimates, and how to react when harvester pressure influences release estimates, if present. Refresher courses also have the benefit of providing observer feedback based on recent fishing trips, and the observer's reliability status[3]. All of these options increase ASOP costs, and would have to be compared to the potential and actual cost of releases, which may be higher (Squires *et al.*, 1998).

Actions that may affect harvester release incentives include value-based ITQs, release tax rates, and selective gear (Kennelly, 1995). Value-based ITQ management programmes reduce release incentives by allowing a dollar amount of fish to be landed, regardless of

[3]Currently, observers in the BC trawl receive feedback on the accuracy of their retained weight estimates after each trip, and are occasionally debriefed to review their progress in more detail. These debriefings allow supervisors and observers the opportunity to resolve data reporting, data collection and species identification issues.

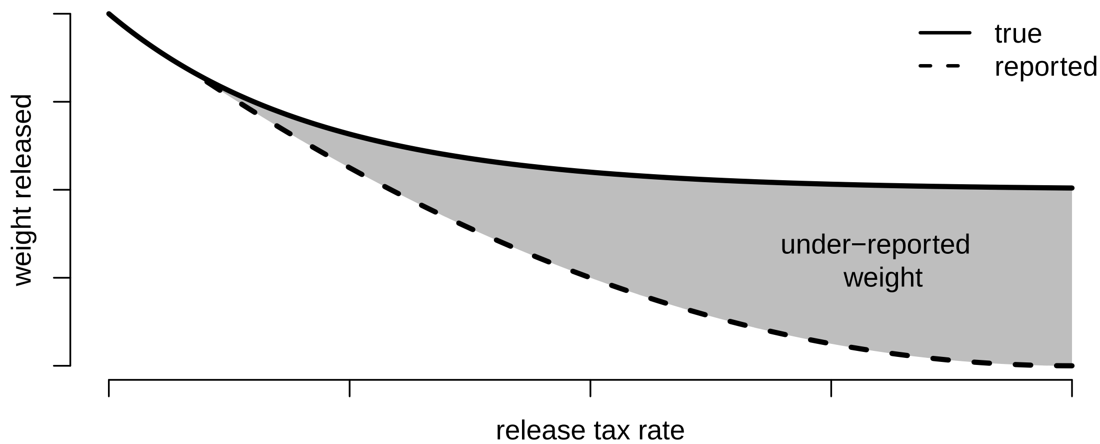

Figure 4.1: Possible relationship between release tax rate (e.g., Pigouvian tax), and the true and reported weight released. Low tax rates are associated with higher weights released. Higher tax rates reduce releases, possibly at the expense of under-reported releases.

the weight required to meet the specified dollar amount (Turner, 1997). A problem of value-based ITQs is that TACs are typically set using a target biomass, however, a value-based ITQ does not allow managers to set landings with certainty because prices may change (Turner, 1997). A benefit of value-based ITQs is increased protection for rare or endangered species which may increase in price due to scarcity value. Scarcity value refers to a rare item's increased value due to increase demand for the item (Ward, 2006). As the value of a species increases, less biomass is required to meet the value-based TAC.

If harvester pressure is the cause of misreported releases, the incentive to reduce releases may be due to a release tax. For example, the BC trawl sablefish release tax forces harvesters to account for a proportion of MR sablefish with quota. In this case, setting the release tax rate may be a trade-off between reducing the weight of released fish, and having reliable release estimates (Figure 4.1). A high weight of released fish (i.e., when the tax is low) may cause biologically inefficient resource use due to partial release mortality. At higher tax rates, the weight released is reduced, but there may be a discrepancy between the true and the reported weight released. Determining the appropriate tax rate can be complicated, and should consider a target release weight, and the potential misreported weight. In some fisheries, 100% retention may also be a viable option to quantify by-catch and total extraction.

Selective fishing gear may reduce the capture of unwanted species, reducing the need to release fish. For trawl vessels, larger mesh sizes may allow small unmarketable pieces and species to escape through the net while retaining larger, more marketable pieces. Extruder grids or panels can prevent large species from entering the net, and can be used sort species based on their physiology. For example, wide flat species can be forced to exit the net while round species are retained. Net design can also affect which species are captured or avoided by investigating fish behaviour. For example, some species may swim up when the net approaches, while other species may swim down.

4.5 Extensions to the current analysis

This is the first formal analysis of the BC trawl's ASOP, and demonstrates a methodology for evaluating the reliability of release reports. The first extension is to establish the credibility of the results by groundtruthing observer reliability indices. The BC trawl provides data from the 100% dockside monitoring program, which could be compared to observer estimates of total retained species for each trip. At the end of the trip when the vessel unloads the retained catch, species are sorted by trained dockside monitors to measure the precise weight for each species. Comparing ASOP estimates to dockside measured weights could determine, for example, whether new observers over-estimate small weights, and under-estimate large weights. Dockside length-frequency distributions of retained sablefish could also be used to detect misreported MR sablefish. For example, consider a trip which retains a substantial weight of sablefish, yet has no reported MR sablefish. Assuming that both unmarketable- and marketable-sized sablefish are encountered on the trip, the minimum length of landed sablefish should be only slightly longer than the minimum legislated length of 0.55 m (i.e., lengths between 0.56 and 0.60 m should be present). If these small but marketable-sized sablefish are not present (i.e., the smallest length is 0.65 m), MR sablefish may have been under-reported.

Observers also provide detailed trip reports for each fishing trip, which include information on working conditions, such as safety and regulatory issues. These reports provide observers the opportunity to report pressure situations which may influence release reports. Properly digitized, accounts of pressure situations could be used to identify harvesters who pressure observers to under-report releases. These strategies could indicate the reasons behind the direction of effect for observer experience, and add credibility of the index of reliability.

4.5.1 Improving release rate prediction accuracy

Fishing activity and reported release rates are likely influenced by predictors that are not included in this analysis, such as market conditions (e.g., gas price, market value, exchange rate), geographical area (e.g., "fishing opportunities" identified by Branch, 2004), stock characteristics (e.g., abundance, size classes), fishing technology, ASOP training, management and regulation changes, and other restrictive species (e.g., canary (*Sebastes pinniger*), bocaccio (*S. paucispinis*), and yelloweye (*S. ruberrimus*) rockfish). To account for possible changes in release behaviour over time I include *year* as a predictor, however this reduces future release prediction accuracy since the underlying effects are not explicit. For example, because the relationship between predictors and the proportion released may be different in the future, the current RF predictions may not be accurate for future events.

Compared to sablefish, the halibut RF has low prediction accuracy. Halibut predictions may have low accuracy because predicting mortality may be more difficult than predicting marketability, and because important halibut predictors are missing. The proportion MR sablefish is determined by harvester behaviour during the event, sorting procedures (i.e., retained vs. released) after the event, as well as the ability of observers to estimate proportions (e.g., marketable and unmarketable) accurately. Harvester release behaviour is typically influenced by expected economic outcomes (Pascoe, 1997). The proportion DR halibut (i.e., mortality rate) is affected by the predictors mentioned previously, but additionally affected by fish physiology, fish handling, and environmental conditions that can affect mortality. For example, fish handling can affect mortality if fish are punctured with picks, or picked up by the tail which can dislocate the vertebrae on larger pieces.

Some important predictors of DR halibut are not included in this analysis. Mortality is higher when halibut spend more time on deck, and are smaller; these predictors are more important than total catch, depth and duration (Richards *et al.*, 1995, 1994). Although the BC trawl ASOP collects data on halibut length and time on deck, these data are not available for this analysis. Consider an event in which halibut are returned to the sea very quickly, and thus have lower-than-expected mortality. Because time on deck is not included in this analysis, the RF may predict high halibut mortality, which is an incorrect prediction for this event. This has repercussions on the assessment of observer reliability; this event would be flagged as having under-reported DR halibut. If time on deck were incorporated in the analysis, the RF model might correctly predict lower mortality. Other predictors may contribute to mortality, such as events that take place on muddy bottom (Berghahn *et al.*, 1992; Saila, 1983), vessel designs such as conveyor belts and salt-water holding ponds, and halibut treatment by individual crew members.

Currently, categorical RF predictors are restricted to less than 33 levels, which prevents incorporating skipper and vessel effects directly in the RF. Although the LME accounts for skipper effects, vessel effects can not be included because they would be confounded with skipper effects. Future versions of the RF analysis will likely allow more levels for categorical predictors. Including skipper and vessel predictors might improve RF prediction accuracy by accounting for these effects directly.

4.5.2 Using decision analysis to assess management actions

Decision analysis is a method to compare the costs and benefits of possible management actions, which explicitly accounts for uncertainty (Morgan and Henrion, 1990; Peterman and Anderson, 1999). This strategy could facilitate decision making when investigating release reliability by quantifying scenarios and outcomes associated with possible management actions. Another benefit of decision analysis is that each component of the analysis must be stated and quantified explicitly (e.g., alternative management actions, uncertain states of nature, performance measures). For example, reliable release data may influence the probability of accurate stock assessment, the probability of having optimal stock exploitation rates and the probability that stock biomass is at an acceptable level. Outcomes from the possible management actions can be assessed using various performance measures for value (e.g., catch value, management costs), or for utility (e.g., skipper equitability, perception of honesty, consistent catch rates).

Decision analysis may be a useful management tool when release estimates are inferred to be unreliable, or when an analysis has low power to detect unreliable release estimates. Deciding among various actions to address these issues could be difficult and controversial, and a quantitative method to compare each expected outcome may be useful. For example, investigating the expected outcomes may allow the comparison of ASOP costs and benefits for unmonitored fisheries. For monitored fisheries that require an improved ASOP, decision analysis may facilitate the comparison among management actions to update the monitoring program.

Chapter 5

Conclusion

Fisheries monitoring programs must be evaluated periodically to ensure that observers report release activity reliably, and that observers treat harvesters equitably. A mandatory at-sea observer program (ASOP) has monitored the British Columbia offshore groundfish trawl fishery's (BC trawl) for more than 10 years, and provides data that are essential for achieving conservation and economic goals. This is the first statistical analysis to test the assumption of reliable and equitable release data in the BC trawl.

The current analysis does not find strong evidence to suggest that ASOP release data are unreliable for estimates of total extraction, or that skippers are treated inequitably. A small proportion of observers may have less reliable release data than others, but misreported weights are negligible on a yearly timescale, and overall. As with observers, a small proportion of skippers may have had inequitable economic costs or benefits resulting from misreported releases, but misreported weights are negligible. Groundtruthing will increase the credibility of these results, and retrospective power analysis will quantify the probability of detecting unreliable release estimates if they exist. Fisheries with undetected unreliable or inequitable release data could have large negative economic and social impacts on the fishery, the management programme, and environmental impacts. On the other hand, analyses that incorrectly infer releases to be unreliable could damage harvester and ASOP reputations.

Observer experience has an important effect on release rates, and indicates that experienced observers report less released fish than new observers. The consistent importance and direction of effect of this human factor for both sablefish and halibut indicates that it cannot be ignored as having an influence on release rates in the BC trawl. More research is required to determine whether the effect of observer experience is a concern for the reliability and equitability of ASOP data.

This analysis has helped to give confidence in the reliability and equitability of the BC trawl's ASOP, and has identified a few areas that would benefit from further investigation. It provides a framework in which to evaluate other fisheries, and a baseline to compare with future studies on the BC trawl's ASOP. This type of analysis will help inform future management actions, and ensure that the BC trawl is managed in a sustainable manner that is equitable among harvesters.

Appendix A

Supplemental Material

Appendix A contains material that supplements the main document. First, I describe the indicators that observers in the at-sea observer program (ASOP) use to visually assess each halibut's condition and mortality rate. I then show summaries of the data used to predict the proportion marketable released (MR) sablefish and dead released (DR) halibut for each fishing event in the British Columbia offshore commercial groundfish trawl fishery (BC trawl). I describe the tuning procedure that improves random forest (RF) prediction accuracy for three parameters. Next, I show the relation ship between the reported and the median predicted proportion MR sablefish and DR halibut for each event. Finally, bivariate marginal dependence plots show the combined effect of the correlated predictor pair *totalCatch* and *totalRetCatch* on the proportion MR sablefish and DR halibut.

A.1 Halibut condition factors

Observers use indicators to assess each halibut piece's condition and mortality rate (Table A.1). When observers are unable to assess each piece, observers extrapolate the assessed halibut weight to the entire halibut catch. Individual halibut mortality rates are combined into an overall weight of DR halibut for the event.

A.2 Data analysis and summaries

Data are obtained from the Fisheries and Oceans Canada (DFO) PacHarvTrawl database. Data summaries show the range and distribution of data available for the analysis of MR

Table A.1: Indicators used to assess halibut condition (i.e., good, poor, or dead) and mortality rate based on each piece's physical condition and length (meters). Note: adapted from the 2007 at-sea observer briefing workbook, which is based on Williams and Wilderbuer (1995).

Description	Indicators	Mortality rate
Good: no sign of stress	· Vigorous body movement before or after release · Fish closes the operculum (gill cover) tightly for at least 5 to 10 seconds · Jaw may be tightly clenched · Muscle tone or physical activity is strong · Minor external injuries, may be slight haemorrhaging of blind side or minor fin fraying, superficial nicks or cuts · Gills are deep red	0.20
Poor: alive, but showing signs of stress	· Fish closes operculum weakly and without sustained pressure · Muscle tone or physical activity is weak: intermittent movement; may respond if stimulated; body appears limp · Moderate injuries or bleeding may be present: pronounced haemorrhaging on blind side; slight bleeding from fin edges; moderate abrasions or cuts · Gills are deep to bright red	0.53
Dead: no sign on life or, if alive, likely to die from severe injuries or suffocation	· No body or operculum movement; fish does not close operculum, jaw may be open · Physical activity limited to fin ripples or twitches; little or no response to stimuli · Vital organs may be damaged: body cavity may be ripped open; severe skin lacerations; sediment in mouth; haemorrhaging on blind side 50% or more · Severe bleeding may be occurring · Gills may be red, pink or white	0.95 if length is less than 0.80, 0.83 if length is at least 0.80

sablefish (Tables A.2 & A.3), and DR halibut (Tables A.4 & A.5). A fast powerful machine is require to analyse the large data set ($\approx 35 \times 60{,}000$ matrix) in a timely manner. For example, approximately 12 days are required to analyse the sablefish data using a 64-bit quad-core Mac Pro running 2.66 gigahertz under Fedora 11 Linux 2.6 with 32 gigabytes of random access memory. The RF tuning procedure and iterative LME analysis use parallel processing to decrease computation time. Since the RF tuning procedure uses about ten of the 12 days, the halibut analysis, which uses baseline (i.e., default) RF parameters, requires approximately 3.5 days. Spearman's ρ is calculated on a 32-bit Windows machine due to correlation errors on 64-bit Linux (R project for statistical computing, 2009).

Two adjustments are made to avoid arithmetic errors. First, if total released (TR) is reported as zero on the i^{th} event, I set TR_i to 1.00×10^{-6} in Equation 2.1 on page 25, and Equation 2.13 on page 34. Second, if the per-skipper or per-observer total reported MR or DR weight is zero, I set MR^{total} or DR^{total} to 1.00×10^{-6} in Equation 2.12 on page 33.

A.3 Sablefish random forest sensitivity analysis

Sablefish RF prediction accuracy is improved via sensitivity analysis. Prediction accuracy, measured via decreasing mean squared error (MSE) and increasing percent variance explained (PVE), changes with respect to three tuning parameters: the number of trees k, the number of predictors p, and the minimum node size g.

The number of trees k necessary for good performance increases with the number of predictors (Liaw and Wiener, 2002). Prediction accuracy and computation time increase as k increases, however prediction accuracy approaches as asymptote. The RF does not over-fit as more trees are added (Breiman, 2001), but since model error eventually approaches an asymptotic minimum (Peters *et al.*, 2007), k must be set sufficiently close to the limiting value. I use baseline values for p and g, and determine the number of trees required for accurate prediction (Table A.6). The RF MSE decreased rapidly from 0.0453 ($k = 1$) to 0.0082 ($k = 200$), which is negligibly different from the asymptotic minimum error of approximately 0.0081.

The number of trees for tuning ($k = 200$) is sufficient to allow MSE to approach the asymptotic minimum for every combination of $p = 2$, 14, 25 and $g = 3$, 14, 26 investigated, and decreases computation time. Prediction accuracy ranges considerably, between 34.8 and 46.5%, for the nine different combinations of p and g (median=42.3%). Tree accuracy

Table A.2: Reported proportion marketable released sablefish y_i^{Rep} and continuous sablefish predictors.

Variable	Value or Unit	Minimum	25^{th}	Median	Mean	75^{th}	Maximum
y_i^{Rep}	proportion	0.000	0.000	0.000	0.023	0.000	1.000
$depth$	meters	12.8	160.9	230.4	346.6	455.4	1,578.2
$tripDays$	days	0	5	7	9	9	99
$totalTows$	integer	1	17	23	26	31	125
$cumTotalRetCatch$	metric tonnes	0.0	6.8	15.9	22.6	30.3	856.5
$propFull$	proportion	0.000	0.060	0.160	0.213	0.320	1.000
$relativeTow$	proportion	0.000	0.270	0.520	0.519	0.770	1.000
$towDuration$	hours	0	2	2	3	4	21
$obsSkipTows$	integer	1	8	18	27	34	390
$seaDays$	days	1	65	175	239	344	1,303
$capacity$	metric tonnes	3.1	68.0	91.6	117.9	145.1	1,049
$initialQuota$	metric tonnes	-0.7	3.7	7.9	8.8	13.3	38.4
$fleetQuota$	metric tonnes	108.2	372.5	532.2	536.0	702.2	955.1
$quota$	metric tonnes	-7.8	1.7	4.2	5.3	7.8	33.6
$YTDcatch$	metric tonnes	0.0	0.2	1.4	3.5	5.6	30.5
$totalCatch$	metric tonnes	0.0	1.0	1.9	2.8	3.6	66.1
$totalRetCatch$	metric tonnes	0.0	0.6	1.3	2.2	2.7	66.1
$totalFish$	kilograms	0	9	27	87	83	9,049
$retained$	kilograms	0	0	4	40	27	6,804
$totalReleased$	kilograms	0	4	14	47	42	7,802

Table A.3: Reported categorical sablefish predictors. The number of events is indicated for each level (i.e., level : events).

statArea	*year*	*quarter*	*month*	*endTime*	*method*	*thornyhead*
3C : 14,391	1997/1998 : 6,423	one : 11,262	jan : 2,993	day : 49,566	observer : 57,459	none : 46,622
3D : 11,733	1998/1999 : 8,355	two : 18,819	feb : 3,919	night : 8,749	skipper : 856	present : 11,693
4B : 30	1999/2000 : 9,356	three : 17,462	mar : 4,350	–: –	–: –	–: –
5A : 7,341	2001/2002 : 8,210	four : 10,772	apr : 4,324	–: –	–: –	–: –
5B : 11,845	2002/2003 : 8,100	–: –	may : 6,961	–: –	–: –	–: –
5C : 2,876	2003/2004 : 6,189	–: –	jun : 7,534	–: –	–: –	–: –
5D : 5,899	2004/2005 : 5,974	–: –	jul : 6,249	–: –	–: –	–: –
5E : 4,200	2005/2006 : 5,708	–: –	aug : 5,156	–: –	–: –	–: –
–: –	–: –	–: –	sep : 6,057	–: –	–: –	–: –
–: –	–: –	–: –	oct : 5,371	–: –	–: –	–: –
–: –	–: –	–: –	nov : 4,017	–: –	–: –	–: –
–: –	–: –	–: –	dec : 1,384	–: –	–: –	–: –

Table A.4: Reported proportion dead released halibut y_i^{Rep} and continuous halibut predictors.

Variable	Value or Unit	Minimum	25^{th}	Median	Mean	75^{th}	Maximum
y_i^{Rep}	proportion	0.000	0.199	0.204	0.296	0.328	1.000
$depth$	meters	16.5	89.6	131.7	170.5	217.6	1,035
$tripDays$	days	0	5	6	7	7	99
$totalTows$	integer	1	17	23	24	29	125
$cumTotalRetCatch$	metric tonnes	0.0	7.7	18.9	24.3	33.5	856.5
$propFull$	proportion	0.000	0.080	0.200	0.239	0.360	1.000
$relativeTow$	proportion	0.000	0.290	0.540	0.534	0.780	1.000
$towDuration$	hours	0	1	2	2	3	17
$obsSkipTows$	integer	1	9	19	28	36	390
$seaDays$	days	1	59	162	229	332	1,303
$capacity$	metric tonnes	4.3	67.6	86.2	108.9	118.1	1,048.8
$initialQuota$	metric tonnes	0.0	6.5	10.1	10.5	14.7	26.5
$fleetQuota$	metric tonnes	687.0	879.9	944.2	935.8	993.2	1,443.2
$quota$	metric tonnes	-1.5	5.0	8.3	8.7	12.1	26.5
$YTDcatch$	metric tonnes	0.0	0.4	1.2	1.8	2.6	11.6
$totalCatch$	metric tonnes	0.0	1.1	2.1	2.9	3.7	59.0
$totalRetCatch$	metric tonnes	0.0	0.6	1.3	2.2	2.7	58.9
$totalFish$	kilograms	0	12	27	60	60	5,915
$skateWeight$	kilograms	0	0	29	147	109	11,794

Table A.5: Reported categorical halibut predictors. The number of events is indicated for each level (i.e., level : events).

statArea	*year*	*quarter*	*month*	*endTime*	*method*
3C : 9,227	1997/1998 : 6,030	one : 11,798	jan : 2,979	day : 51,854	observer : 59,004
3D : 5,627	1998/1999 : 7,231	two : 19,935	feb : 4,476	night : 7,333	skipper : 183
4B : 7	1999/2000 : 7,860	three : 18,509	mar : 4,343	– : –	– : –
5A : 7,863	2001/2002 : 7,010	four : 8,945	apr : 4,694	– : –	– : –
5B : 16,032	2002/2003 : 8,307	– : –	may : 7,297	– : –	– : –
5C : 4,585	2003/2004 : 7,714	– : –	jun : 7,944	– : –	– : –
5D : 13,485	2004/2005 : 7,131	– : –	jul : 7,146	– : –	– : –
5E : 2,361	2005/2006 : 7,904	– : –	aug : 5,345	– : –	– : –
– : –	– : –	– : –	sep : 6,018	– : –	– : –
– : –	– : –	– : –	oct : 4,240	– : –	– : –
– : –	– : –	– : –	nov : 3,315	– : –	– : –
– : –	– : –	– : –	dec : 1,390	– : –	– : –

Table A.6: Relationship between number of trees k and random forest prediction accuracy, measured via mean squared error (MSE) for sablefish and halibut. Other parameters are kept at baseline values (number of predictors $p = 8$, and minimum node size $g = 5$).

k	MSE (sablefish)	MSE (halibut)
1	0.0453	0.2544
10	0.0118	0.0344
100	0.0083	0.0254
200	0.0082	0.0250
500	0.0081	0.0248

and inter-tree correlation decrease as p decreases (Breiman, 2001; Peters *et al.*, 2007). The RF has low prediction error when individual trees give accurate predictions and between-tree correlation is low (i.e., individual trees are different from one another, Breiman, 2001). Values of p greater than baseline may improve model performance when P is large and most predictors have only weak effects (Liaw and Wiener, 2002; Segal, 2004), as is the case for the current application. The RF shows a slight improvement in prediction accuracy by tuning over a range of p and g parameter values (best fit $p = 10$, $g = 5$, MSE = 0.0082, PVE = 46.3%, Figure A.1a & b). Although prediction accuracy is high when g is less than 5, I constrain $g \geq 5$ to prevent over-fitting. Because the improvement in sablefish RF prediction accuracy is modest, the halibut RF uses baseline parameter values.

Relative predictor importance for the six most important predictors, evaluated via permutation accuracy, is similar for most combinations of $p = 2, 14, 25$ and $g = 3, 14, 26$, but the order varies for different parameter combinations (Figure A.2). For example, *year*, *month*, and *seaDays* are among the 6 most important predictors for each combination. The predictor *year* is consistently the most important, and *seaDays* is consistently among the three most important predictors.

A.4 Random forest predictions

As expected, an average of 37% (±0.04) of events are out-of-bag (OOB) in each tree for both the sablefish and halibut RFs, where the number in parentheses represents two standard

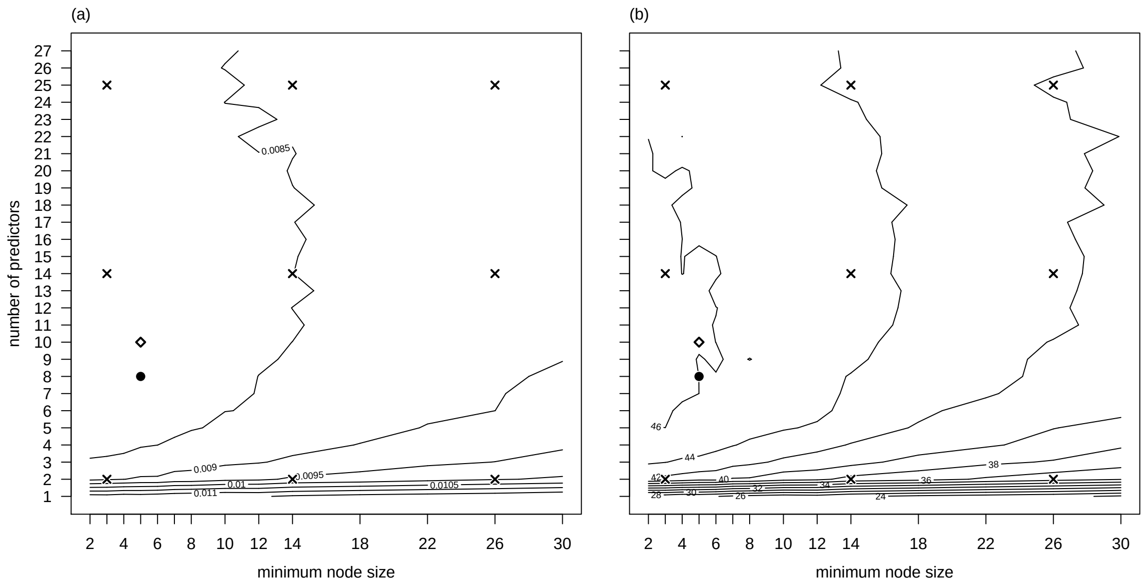

Figure A.1: Sablefish random forest prediction accuracy sensitivity analysis evaluated via mean squared error (MSE, a) and percent variance explained (PVE, b) for combinations of number of predictors p and minimum node size g, keeping the number of trees $k = 200$. Symbols indicate baseline (•: $p = 8$; $g = 5$; MSE = 0.0083; PVE = 46.0%), best fit (◇: $p = 10$; $g = 5$; MSE = 0.0082; PVE = 46.3%) and combinations investigated in greater detail (×: $p = 2$, 14, 25, and $g = 3$, 14, 26) in Figure A.2.

deviations (Liaw and Wiener, 2002). I show the relationship between the reported proportion MR sablefish and DR halibut, and the median RF prediction for each event (Figures A.3a & A.4a, respectively). Residuals are calculated as the difference between reported and predicted releases. Negative residuals (i.e., dots above the 1:1 line) indicate events with a high median predicted proportion and a low reported proportion.

Histograms show the marginal distribution for reported and median predicted proportions MR sablefish and DR halibut (Figure A.3b & c, Figure A.4b & c, respectively). The majority of sablefish events have zero reported and zero median predicted proportion MR sablefish. Both the reported and the median predicted proportion MR sablefish range between 0.00 and 1.00. The reported proportion DR halibut has distinct modes at 0.20, 0.53, 0.83 and 0.95, which correspond to halibut mortality rates (Table A.1). These modes are not evident in the median predicted proportion DR halibut. The reported proportion DR halibut ranges between 0.00 and 1.00, while the median predicted proportion ranges between 0.00 and 0.95.

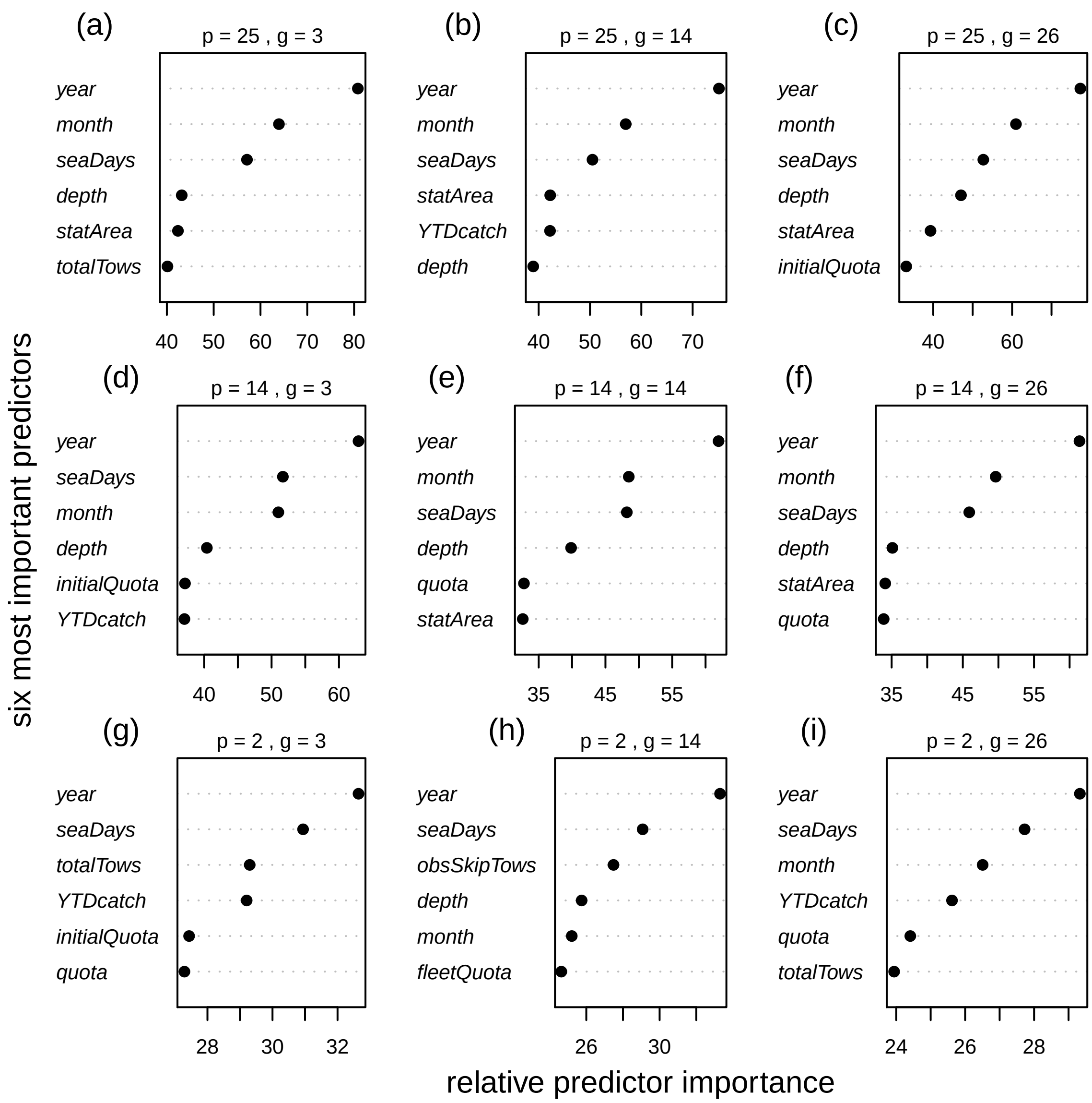

Figure A.2: Sablefish random forest relative predictor importance sensitivity. Permutation accuracy importance shows the six most important sablefish random forest predictors with the number of trees $k = 200$ for various parameter combinations (number of predictors $p = 2$, 14, 25; minimum node size $g = 3$, 14, 26).

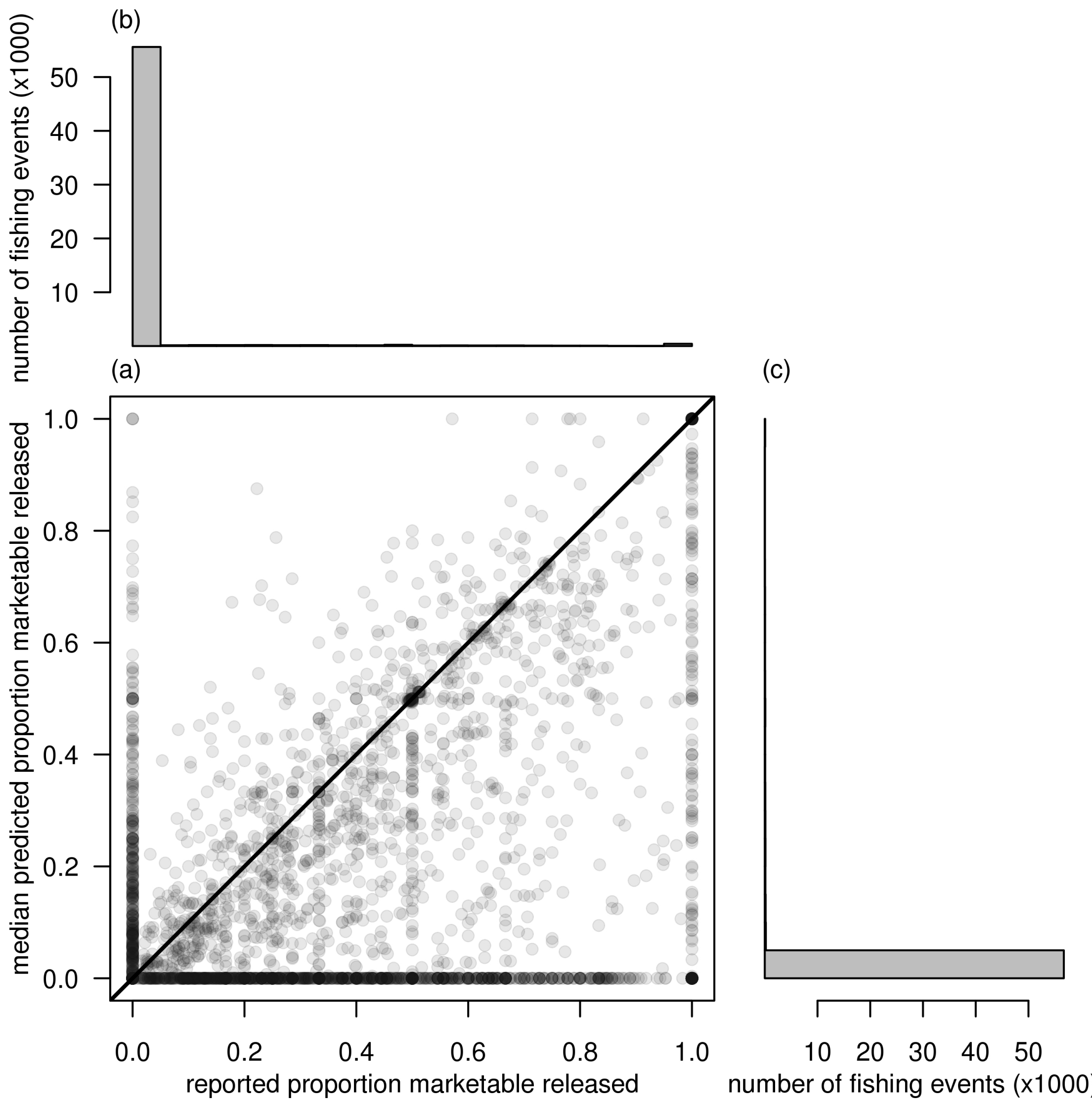

Figure A.3: Reported and median predicted proportion marketable released (MR) sablefish (a). Dots are semi-transparent to help visualize density. The solid line is the 1:1 line. Histograms show marginal distributions of reported and median predicted MR sablefish (b & c, respectively).

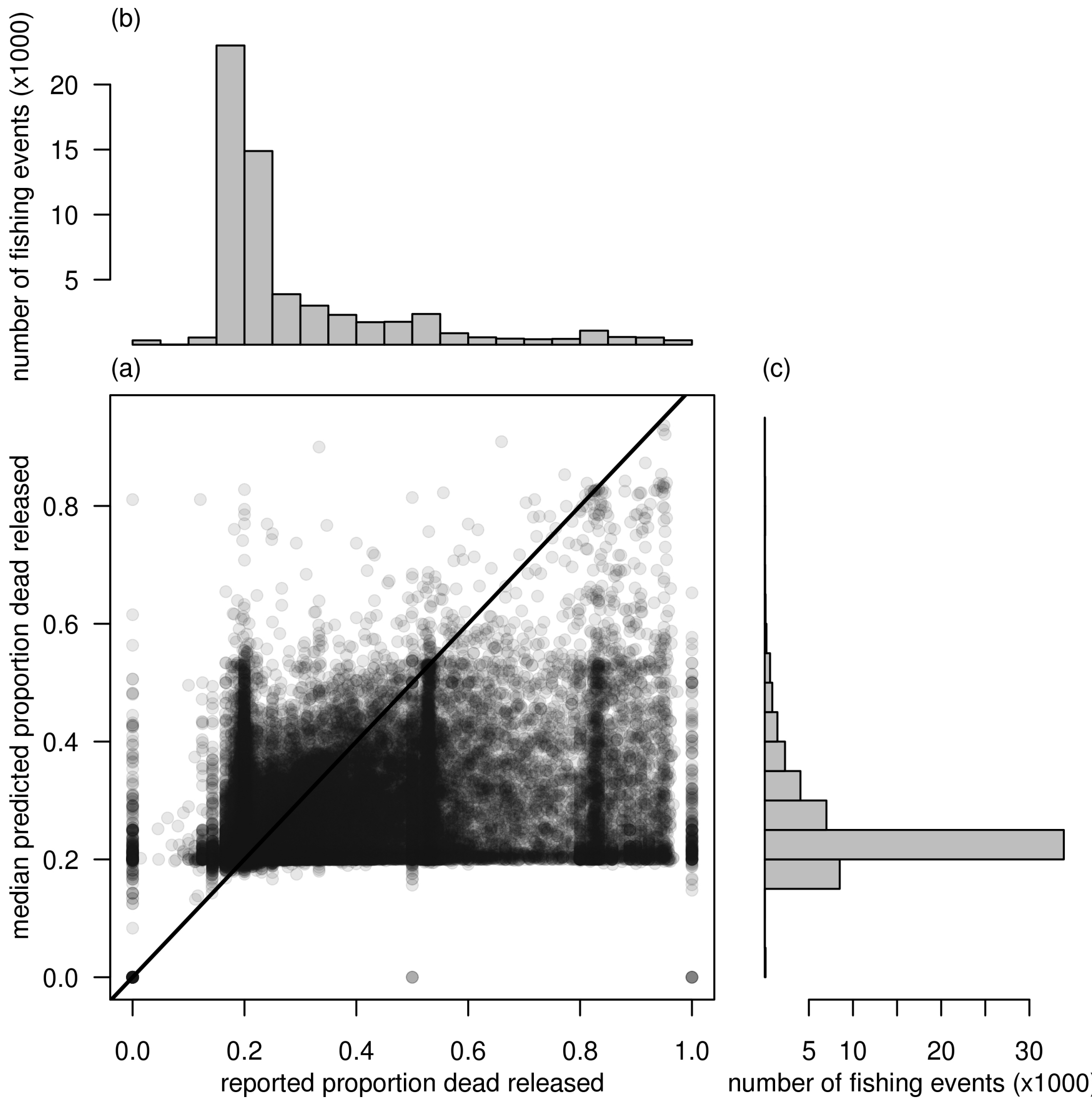

Figure A.4: Reported and median predicted proportion dead released (DR) halibut (a). Dots are semi-transparent to help visualize density. The solid line is the 1:1 line. Histograms show marginal distributions of reported and median predicted DR halibut (b & c, respectively).

A.5 Bivariate marginal dependence

Spearman's rank correlation coefficient ρ determines predictor pair correlation. The highest correlation is between *totalCatch* and *totalRetCatch*, with $\rho = 0.91$, $p < 0.001$ for sablefish, and $\rho = 0.89$, $p < 0.001$ for halibut. Bivariate marginal dependence plots show the relationship between *totalCatch* and *totalRetCatch*, and the predicted proportion MR sablefish and DR halibut (Figures A.5 & A.6, respectively). The bivariate marginal dependence indicates that the combined effect has the same trend as the univariate marginal dependence plots for both species.

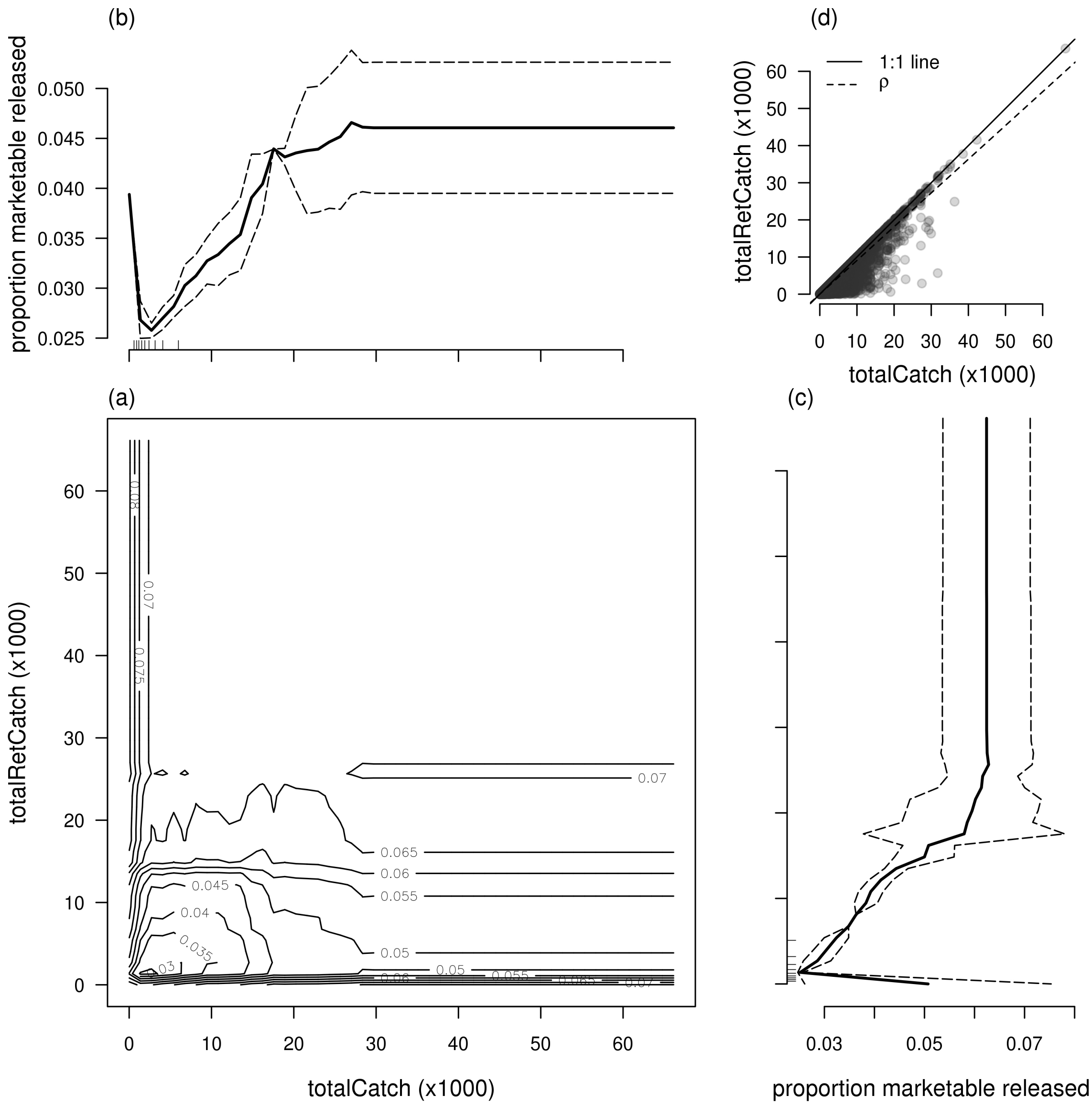

Figure A.5: Bivariate marginal dependence for *totalCatch* and *totalRetCatch* (Spearman's rank correlation coefficient $\rho = 0.91$, $p < 0.001$). Contour lines show predicted proportion marketable released sablefish (a). Univariate marginal dependence for *totalCatch* and *totalRetCatch* show mean prediction, deciles with tick mark and uncertainty as two times the standard error of the mean with dashed lines (b & c, respectively). Panel (d) shows the relationship, the correlation, and the 1:1 line. Dots are semi-transparent to help visualize density. Weight is in kilograms.

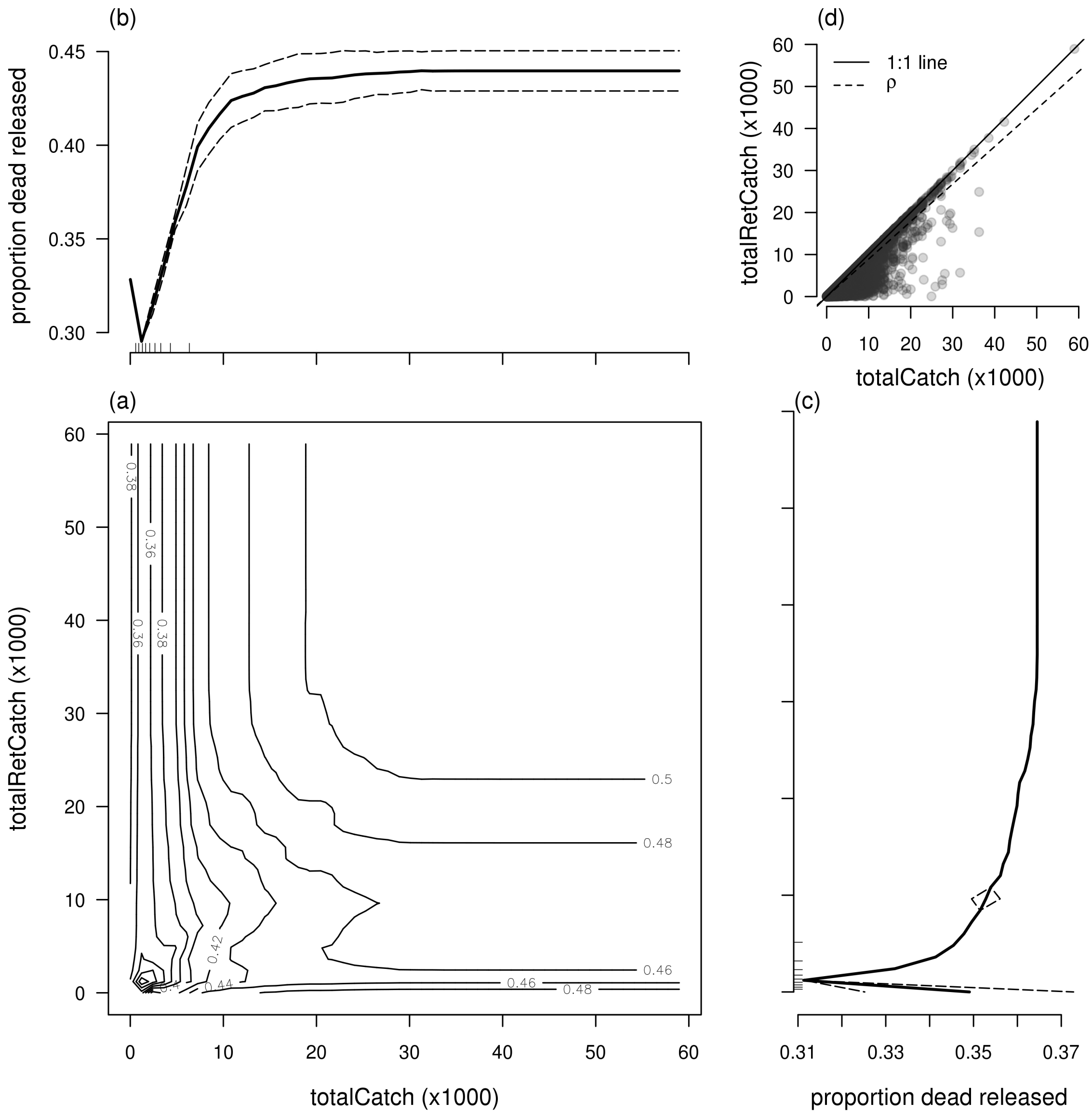

Figure A.6: Bivariate marginal dependence for *totalCatch* and *totalRetCatch* (Spearman's rank correlation coefficient $\rho = 0.89$, $p < 0.001$). Contour lines show predicted proportion dead released halibut (a). Univariate marginal dependence for *totalCatch* and *totalRetCatch* show mean prediction, deciles with tick marks and uncertainty as two times the standard error of the mean with dashed lines (b & c, respectively). Panel (d) shows the relationship, the correlation, and the 1:1 line. Dots are semi-transparent to help visualize density. Weight is in kilograms.

References

Agardy, T. (2000). Effects of fisheries on marine ecosystems: A conservationist's perspective. *ICES Journal of Marine Science*, 57(3):761–765.

Alverson, D., Freeberg, M., Pope, J., and Murawski, S. (1994). A global assessment of fisheries bycatch and discards. *Food and Agriculture Organization, Fisheries Technical Paper*, 339:233. Rome.

Anderson, I. G. (1994). An economic analysis of highgrading in ITQ fisheries regulation programs. *Marine Resource Economics*, 9(3):209–226.

Arnason, R. (1994). On catch discarding in fisheries. *Marine Resource Economics*, 9:189–207.

Bates, D. and Maechler, M. (2009). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-32.

Beddington, J. R., Agnew, D. J., and Clark, C. W. (2007). Current problems in the management of marine fisheries. *Science*, 316(5832):1713–1716.

Benoit, H. P. and Allard, J. (2009). Can the data from at-sea observer surveys be used to make general inferences about catch composition and discards? *Canadian Journal of Fisheries and Aquatic Sciences*, 66:2025–2039.

Berghahn, R., Waltemath, M., and Rijnsdorp, A. D. (1992). Mortality of fish from the by-catch of shrimp vessels in the North Sea. *Journal of Applied Ichthyology*, 8(1-4):293–306.

Berk, R. A. (2006). An introduction to ensemble methods for data analysis. *Sociological Methods Research*, 34(3):263–295.

Borges, L., Zuur, A. F., Rogan, E., and Officer, R. (2004). Optimum sampling levels in discard sampling programs. *Canadian Journal of Fisheries and Aquatic Sciences*, 61:1918–1928.

Branch, T. A. (2004). *The influence of individual transferable quotas on discarding and fishing behaviour in multispecies fisheries*. PhD thesis, School of Aquatic and Fishery Sciences, University of Washington.

Branch, T. A. (2009). How do individual transferable quotas affect marine ecosystems? *Fish and Fisheries*, 10(1):39–57.

Branch, T. A. and Hilborn, R. (2008). Matching catches to quotas in a multispecies trawl fishery: Targeting and avoidance behavior under individual transferable quotas. *Canadian Journal of Fisheries and Aquatic Sciences*, 65:1435–1446.

Branch, T. A., Rutherford, K., and Hilborn, R. (2006). Replacing trip limits with individual transferable quotas: Implications for discarding. *Marine Policy*, 30(3):281–292.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Breiman, L. (2002). Manual on setting up, using and understanding random forests version 3.1. Unpublished manuscript.

Breiman, L., Liaw, A. C. A., and Wiener, M. (2005). randomForest: Breiman and Cutler's random forests for classification and regression. R package version 4.5-15.

Brosi, B. J. and Biber, E. G. (2009). Statistical inference, type II error, and decision making under the US endangered species act. *Frontiers in Ecology and the Environment*, 7(9):487–494.

Catchpole, T. L., Frid, C. L. J., and Gray, T. S. (2005). Discards in North Sea fisheries: Causes, consequences and solutions. *Marine Policy*, 29(5):421–430.

Costello, C., Gaines, S. D., and Lynham, J. (2008). Can catch shares prevent fisheries collapse? *Science*, 321(5896):1678–1681.

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., and Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, 88(11):2783–2792.

De'ath, G. and Fabricius, K. E. (2000). Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology*, 81(11):3178–3192.

Dewees, C. M. (1998). Effects of individual quota systems on New Zealand and British Columbia fisheries. *Ecological Applications*, 8(sp1):S133–S138.

Eythorsson, E. (2000). A decade of ITQ-management in Icelandic fisheries: Consolidation without consensus. *Marine Policy*, 24:483–492.

Field, S. A., Tyre, A. J., Jonz, N., Rhodes, J. R., and Possingham, H. P. (2004). Minimizing the cost of environmental management decisions by optimizing statistical thresholds. *Ecology Letters*, 7:669–675.

Fisheries and Oceans Canada (2005a). Pacific region integrated fisheries management plan: Groundfish trawl. April 1, 2005 to March 31, 2006.

Fisheries and Oceans Canada (2005b). Pacific region integrated fisheries management plan: Halibut. February 27, 2005 to February 26, 2006.

Fisheries and Oceans Canada (2005c). Pacific region integrated fisheries management plan: Sablefish. August 1, 2005 to July 31, 2006.

Fisheries and Oceans Canada (2009). Summary of historic catch vs available weight. Retrieved 20 June, 2009 from the DFO Pacific region extranet website http://www-ops2.pac.dfo-mpo.gc.ca/xnet/content/groundfish/GFTrawl/GfTrawlInfo.htm.

Food and Agriculture Organization of the United Nations (1995). Code of conduct for responsible fisheries.

Food and Agriculture Organization of the United Nations (1997). Fisheries bycatch and discards. March 17-20, 1997.

Gillis, D. M., Pikitch, E. K., and Peterman, R. M. (1995). Dynamic discarding decisions: Foraging theory for high-grading in a trawl fishery. *Behavioral Ecolology*, 6(2):146–154.

Gotelli, N. J. and Ellison, A. M. (2004). *A Primer of Ecological Statistics*. Sinauer Associates, Inc, Sunderland, Massachusetts.

Grafton, Q. R., Kirkley, J., Squires, D., and Weninger, Q. (2003). A guide to economic evaluation of individual transferable quota fisheries. Technical report, Iowa State University, Department of Economics. Staff General Research Papers.

Hall, S. J. and Mainprize, B. M. (2005). Managing by-catch and discards: How much progress are we making and how can we do better? *Fish and Fisheries*, 6:134–155.

Hammond, T. R. and Trenkel, V. M. (2005). Censored catch data in fisheries stock assessment. *ICES Journal of Marine Science*, 62(6):1118–1130.

Hannesson, R. (1996). On ITQs: An essay for the special issue of reviews in fish biology and fisheries. *Reviews in Fish Biology and Fisheries*, 6(1):91–96.

Harris, L.-W. E. (1998). The quantification of perceptual and behavioural variation in commercial fishing logbooks. Master's thesis, Department of Zoology, University of Manitoba.

Hilborn, R., Punt, A., and Orensanz, J. (2004). Beyond band-aids in fisheries management: Fixing world fisheries. *Bulletin of Marine Science*, 74:493–507.

Hilborn, R. and Walters, C. J. (1992). *Quantitative fisheries stock assessment: Choice, dynamics and uncertainty*. Kluwer Academic Publishers, Boston.

Hoag, S. A. (1975). Survival of halibut released after capture by trawls. Technical report, International Pacific Halibut Commission. Scientific Report No. 57.

Kelleher, K. (2005). Discards in the world's marine fisheries: An update. *Food and Agriculture Organization Fisheries Technical Paper*, 470:131.

Kennelly, S. J. (1995). The issue of bycatch in Australia's demersal trawl fisheries. *Reviews in Fish Biology and Fisheries*, 5(2):213–234.

Kennelly, S. J., Liggins, G. W., and Broadhurst, M. K. (1998). Retained and discarded bycatch from oceanic prawn trawling in New South Wales, Australia. *Fisheries Research*, 36:217–236.

Kolar, C. S. and Lodge, D. M. (2002). Ecological predictions and risk assessment for alien fishes in North America. *Science*, 298(5596):1233–1236.

Kriebel, D., Tickner, J., Epstein, P., Lemons, J., Levins, R., Loechler, E. L., Quinn, M., Rudel, R., Schettler, T., and Stoto, M. (2001). The precautionary principle in environmental science. *Environmental Health Perspectives*, 109(9):871–876.

Lennert-Cody, C. E. and Berk, R. A. (2007). Statistical learning procedures for monitoring regulatory compliance: An application to fisheries data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(3):671–689.

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

Maindonald, J. and Braun, W. J. (2003). *Data analysis and graphics using R: An example-based approach*. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, United Kingdom.

Maloney, N. E. and Sigler, M. F. (2008). Age-specific movement patterns of sablefish (*Anoplopoma fimbria*) in Alaska. *Fisheries Bulletin*, 106(3):305–316.

Matsuoka, T. (1999). Sampling and estimation of discards in multi-species fisheries. In *International Conference on Integrated Fisheries Monitoring*, Rome, Italy. Food and Agriculture Organization of the United Nations.

McCay, B. J., Creed, C. F., Finlayson, A. C., Apostle, R., and Mikalsen, K. (1995). Individual transferable quotas (ITQs) in Canadian and US fisheries. *Ocean and Coastal Management*, 28(1-3):85–115.

Morgan, M. G. and Henrion, M. (1990). *Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge University Press.

Myers, R. A., Hutchings, J. A., and Barrowman, N. J. (1997). Why do fish stocks collapse? The example of cod in Atlantic Canada. *Ecological Applications*, 7(1):91–106.

Myhre, J. D. (1969). Gear selection and Pacific halibut. Technical Report 51, Report of the International Pacific Halibut Commission.

Neilson, J. D., Waiwood, K. G., and Smith, S. J. (1989). Survival of Atlantic halibut (*Hippoglossus hippoglossus*) caught by longline and otter trawl gear. *Canadian Journal of Fisheries and Aquatic Sciences*, 46:887–897.

Palsson, G. and Durrenberger, E. P. (1990). Systems of production and social discourse: The skipper effect revisited. *American Anthropologist*, 92(1):130–141.

Pascoe, S. (1997). Bycatch management and the economics of discarding. *Food and Agriculture Organization Fisheries Technical Paper*, 370:137.

Peterman, R. M. and Anderson, J. L. (1999). Decision analysis: A method for taking uncertainties into account in risk-based decision making. *Human and Ecological Risk Assessment*, 5(2):231–244.

Peterman, R. M. and M'Gonigle, M. (1992). Statistical power analysis and the precautionary principle. *Marine Pollution Bulletin*, 24(5):231–234.

Peters, J., Baets, B. D., Verhoest, N. E. C., Samson, R., Degroeve, S., Becker, P. D., and Huybrechts, W. (2007). Random forests as a tool for ecohydrological distribution modelling. *Ecological Modelling*, 207(2-4):304–318.

Pinkerton, E. and Edwards, D. N. (2009). The elephant in the room: The hidden costs of leasing individual transferable fishing quotas. *Marine Policy*, 33(4):707–713.

Prasad, A., Iverson, L., and Liaw, A. (2006). Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems*, 9(2):181–199.

R Development Core Team (2009). R: A language and environment for statistical computing. ISBN 3-900051-07-0.

R project for statistical computing (2009). PR numbers 9568 and 9570: Incorrect matrix of Spearman correlations. Retrieved 10 August, 2009 from the R mailing lists archive website http://www.r-project.org/.

Richards, L. J., Fargo, J., and Schnute, J. T. (1995). Factors influencing bycatch mortality of trawl-caught Pacific halibut. *North American Journal of Fisheries Management*, 15(2):266–276.

Richards, L. J., Schnute, J. T., and Fargo, J. (1994). Application of a generalized logit model to condition data for trawl-caught Pacific halibut, *Hippoglossus stenolepis*. *Canadian Journal of Fisheries and Aquatic Sciences*, 51:357–364.

Rochet, M.-J., Peronnet, I., and Trenkel, V. M. (2002). An analysis of discards from the French trawler fleet in the Celtic Sea. *ICES Journal of Marine Science*, 59(3):538–552.

Rochet, M.-J. and Trenkel, V. M. (2005). Factors for the variability of discards: Assumptions and field evidence. *Canadian Journal of Fisheries and Aquatic Sciences*, 62:224–235.

Saila, S. (1983). Importance and assessment of discards in commercial fisheries. *Food and Agriculture Organization Fisheries Circular*, 765:623.

Sanchirico, J. N., Holland, D., Quigley, K., and Fina, M. (2006). Catch-quota balancing in multispecies individual fishing quotas. *Marine Policy*, 30(6):767–785.

Segal, M. R. (2004). Machine learning benchmarks and random forest regression. Retrieved from: http://escholarship.org/uc/item/35x3v9t4.

Sexton, J. and Laake, P. (2009). Standard errors for bagged and random forest estimators. *Computational Statistics and Data Analysis*, 53(3):801–811.

Springford, A. (2008). A novel bayesian method for making the most of spatial fishery catch and effort data. Master's thesis, School of Resource and Environmental Management, Simon Fraser University.

Squires, D., Campbell, H., Cunningham, S., Dewees, C., Grafton, R. Q., Herrick, S. F., Kirkley, J., Pascoe, S., Salvanes, K., Shallard, B., Turris, B., and Vestergaard, N. (1998). Individual transferable quotas in multispecies fisheries. *Marine Policy*, 22(2):135–159.

Squires, D. and Kirkley, J. (1995). Resource rents from single and multispecies individual transferable quota programs. *ICES Journal of Marine Science*, 52(2):153–164.

Stanley, R. D. (1984). Incidental catch of halibut (*Hippoglossus stenolepis*) by domestic trawlers in Queen Charlotte Sound and Hecate Strait, in 1981 and 1982. Technical Report 1318, Canadian Technical Report of Fisheries and Aquatic Sciences.

Stanley, R. D. (1985). A study of bottom-trawl discards in Queen Charlotte Sound and Hecate Strait, 1981 and 1982. Technical Report 1840, Canadian Manuscript Report of Fisheries and Aquatic Sciences.

Stanley, R. D., Olsen, N., and Fedoruk, A. (2009). Independent validation of the accuracy of yelloweye rockfish catch estimates from the Canadian groundfish integration pilot project.

Marine and Coastal Fisheries: Dynamics, Management, and Ecosystem Science, 1:354–362. Special section: Data-poor fisheries.

Stratoudakis, Y., Fryer, R. J., and Cook, R. M. (1998). Discarding practices for commerical gadoids in the North Sea. *Canadian Journal of Fisheries and Aquatic Sciences*, 55(7):1632.

Terry, J. (1998). Halibut bycatch in the Alaska groundfish fishery. In *Managing the nation's bycatch: Priorities, programs and actions for the National Marine Fisheries Service*, pages 163–166. National Oceanic and Atmospheric Administration.

Therneau, T. M., Atkinson, B., and Ripley, B. (2009). rpart: Recursive partitioning. R package version 3.1-45.

Tierney, L., Rossini, A. J., Li, N., and Sevcikova, H. (2009). snow: Simple network of workstations. R package version 0.3-3.

Turner, M. A. (1997). Quota-induced discarding in heterogeneous fisheries. *Journal of Environmental Economics and Management*, 33(2):186–195.

Vestergaard, N. (1996). Discard behaviour, highgrading and regulation: The case of the Greenland shrimp fisherery. *Marine Resource Economics*, 11(4):247–266.

Ward, F. A. (2006). *Environmental and natural resource economics*. Pearson Prentice Hall, Upper Saddle River, New Jersey.

Williams, G. H. and Wilderbuer, T. K. (1995). Discard mortality rates of Pacific halibut bycatch: Fishery differences and trends during 1990-1993. In *International symposium on North Pacific flatfish*, volume 95, pages 611–622. Alaska Sea Grant College Program.

Zar, J. H. (1996). *Biostatistical Analysis*. Prentice-Hall, Toronto, 3rd edition.

Zhou, S. (2008). Fishery by-catch and discards: A positive perspective from ecosystem-based fishery management. *Fish and Fisheries*, 9(3):308–315.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., and Smith, G. M. (2009). *Mixed Effects Models and Extensions in Ecology with R*. Statistics for Biology and Health. Springer, New York.