A Guide for the Analysis of Cumulative Effects of Environmental Stressors to Fraser River Sockeye Salmon

by

James Johnson

B. Math., Carleton University, 2006

Research Project Submitted in Partial Fulfillment
of the Requirements for the Degree of
Master of Resource and Environmental Management

in the
School of Resource and Environmental Management
Faculty of the Environment

Report no. 555

SIMON FRASER UNIVERSITY

Summer 2012

Approval

Name: **James Johnson**

Degree: **Master of Resource Management**

Title of Thesis: *A Guide for the Analysis of Cumulative Effects of Environmental Stressors to Fraser River Sockeye Salmon*

Examining Committee:

Chair: Sabrina Garcia, REM Graduate Student

Andrew Cooper
Senior Supervisor
Associate Professor

David Patterson
Supervisor
Research Biologist

Date Defended/Approved: August 10, 2012

Partial Copyright License

The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the right to lend this thesis, project or extended essay to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users.

The author has further granted permission to Simon Fraser University to keep or make a digital copy for use in its circulating collection (currently available to the public at the "Institutional Repository" link of the SFU Library website (www.lib.sfu.ca) at http://summit/sfu.ca and, without changing the content, to translate the thesis/project or extended essays, if technically possible, to any medium or format for the purpose of preservation of the digital work.

The author has further agreed that permission for multiple copying of this work for scholarly purposes may be granted by either the author or the Dean of Graduate Studies.

It is understood that copying or publication of this work for financial gain shall not be allowed without the author's written permission.

Permission for public performance, or limited permission for private scholarly use, of any multimedia materials forming part of this work, may have been granted by the author. This information may be found on the separately catalogued multimedia material and in the signed Partial Copyright Licence.

While licensing SFU to permit the above uses, the author retains copyright in the thesis, project or extended essays, including the right to change the work for subsequent purposes, including editing and publishing the work in whole or in part, and licensing other parties, as the author may desire.

The original Partial Copyright Licence attesting to these terms, and signed by this author, may be found in the original bound copy of this work, retained in the Simon Fraser University Archive.

Simon Fraser University Library
Burnaby, British Columbia, Canada

revised Fall 2011

Abstract

It is often recognized that a quantitative assessment of the cumulative effects, both additive and non-additive, of multiple stressors would provide a more realistic representation of the factors that influence sockeye salmon (*orhynchus nerka)* migration mortality. Towards achieving this type of assessment, this research project first presents a literature review of multivariable methods currently applied in Fraser River sockeye salmon migration studies and in the fields of fisheries, biology, and medicine which could be used to analyze cumulative effects. Papers taken solely from Fraser River sockeye salmon research revealed a limited number of multivariable methods being applied and the sub-optimal reliance on univariable methods for multivariable problems. The review of fisheries and biological science literature identified a number of additional methods for dealing with cumulative effects while the review of medical science literature did not reveal any additional methods. The literature review also presents a guide for how to apply each of these methods to other cumulative effects studies and more specifically how to apply them to study Fraser River sockeye salmon migration survival. The second part of this project presents an application of two of these multivariable methods, regression trees and random forests, to describe and predict the cumulative effects of multiple habitat and stressor variables on Fraser River sockeye salmon prespawn mortality (PSM). The results of this analysis show that although a number of these variables may relate to sockeye salmon PSM, only a few variables representing the timing of entry into the Fraser River, the destination spawning ground, and human population density are required to predict Fraser River sockeye PSM.

Keywords: cumulative-effects; salmon; Fraser River; prespawn mortality; stressor; quantitative

Dedication

To my beautiful, intelligent, supportive and patient wife Akiko. You put just as much work into this as I did.

Acknowledgements

First and foremost, I would like to thank my supervisory committee, Andrew Cooper and David Patterson, who supported me with my research, had the patience to answer all of my questions and provided me with a wealth of knowledge from which to draw upon. I would also like to thank Eduardo Martins (Fisheries and Oceans Canada) for contributing a great deal of time and experience to the publication of my research and Merran Hague (Pacific Salmon Commission) for her expert advice during the initial stages of this project. Thanks to Scott Hinch (University of British Columbia), Steven Cooke (Carleton University), and Carl Swartz (Simon Fraser University) for your much needed assistance during the final stages of publication. Last but not least, thanks to the Fisheries Science and Management Research Group at REM, especially the Fish Stix, for your ongoing support.

Funding for this project was provided by Fisheries and Oceans Canada through the Cooperative Resource Management Institute (CRMI) and the Center of Expertise on Aquatic Habitat Research (CAHR).

.

Table of Contents

Approval ii
Partial Copyright License iii
Abstract iv
Dedication v
Acknowledgements vi
Table of Contents vii
List of Tables ix
List of Figures x
List of Acronyms xi

Chapter 1. General Introduction 1
References 2

Chapter 2. Quantitative methods for analyzing cumulative effects to fish migration success: a review (published as: Johnson, J. E., Patterson, D. A., Martins, E. G., Cooke, S. J. & Hinch, S.G. (2012). Quantitative methods for analyzing cumulative effects on fish migration success: a review. Journal of Fish Biology 81, 600-631.) 4
Introduction 4
Compilation of Multivariable Papers on Fraser River *O. nerka* 6
    Background 6
    Compilation Methods 7
    Compilation Results 7
Review - Quantitative Methods Applied within Fisheries , Biological and Medical Sciences 9
    Review Background 9
    Review Methods 9
        Fisheries Science Literature 9
        Non-Fisheries Literature 10
    Review Results 11
Discussion 12
Conclusion 17
Tables 18
Figures 27
References 28

Chapter 3. A Quantitative Analysis of Cumulative Effects to Fraser River Sockeye Salmon Prespawn Mortality 45
Introduction 45
Methods 47
    PSM Data and Study Populations 47
    Predictor Variables and their Potential Relation to PSM 48

        - Timing of Return River Migration 48
        - Temperature 49
        - Discharge 50
        - Water Quality 51
        - Stored Energy 52
        - Density 53
        - Population Specific Differences 53
        - Physical Habitat Variables 54
            - Forestry & Mountain Pine Beetle Activity 54
            - Large Scale Hydroelectric 55
            - Urbanization 55
            - Agriculture 56
            - Inlet/Outlet Spawners 56
    - Classification and Regression Tree Modelling 57
    - Random Forest Modeling 57
- Results 59
    - CART Model 59
    - Random Forest Model 59
    - Final Reduced Model for Prediction 61
- Discussion 62
- Tables 68
- Figures 70
- References 74

Chapter 4. General Conclusions 80

- References 81

Appendices 82

- Appendix A. List of References 83
- Appendix B. Boxcar Model and Data Cleaning 89

List of Tables

Table 2.1 Classification of all O. nerka papers found within the Fisheries and Oceans Canada (DFO) database. 18

Table 2.2 Model classifications along with descriptions and examples of appropriate routines in R statistical software and references which can act as an initial guide for choosing between different techniques. Models are ordered in the list according to their classifications 22

Table 3.1 List of all predictor variables included in classification and regression tree and random forest modeling procedures. 68

List of Figures

Figure 2.1. Frequency distribution (left) for multivariable methods used within the Fraser River O. nerka literature (black), fisheries literature (white) and other biological science literature (grey) along with mean and standard deviation (right) for number of predictors used in the fisheries and other biological science literature. MFANOVA= Multi-Factor Analysis of Variance, MLR= Multiple Linear Regression, QR= Quantile Regression, SA= Survival Analysis, GLM= Generalized Linear Model, GAM= Generalized Additive Model, NLR= Non-linear Regression, MARS= Multivariate Adaptive Regression Splines, CART=Classification & Regression Trees, BAT= Bagging Tree, BOT=Boosted Tree, RF=Random Forest, CLA= Cluster Analysis, ANN= Artificial Neural Network, GWR= Geographically Weighted Regression, SEM= Structural Equation Modeling. 27

Figure 3.1. Map of the Fraser River watershed with approximate locations of river temperature logger stations (stars) and Sockeye salmon spawning grounds included in our analysis. Associated names for spawning grounds, as well as sockeye salmon Conservation unit and Production unit (italics) are as follows: 1. Gluske Creek, Forfar Creek, Kynock Creek, Bivouac Creek (Takla/Trembleur-ESTU, Early Stuart); 2. Tachie River (Stuart-S, Late Stuart); 3. Middle River (Takla/Trembleur-S, Late Stuart); 4. Stellako River (Fraser-S, Stellako); 5. Nadina River, Nadina Channel (Francois-ES, Nadina); 6. Horsefly River (Quesnel-S, Horsefly); 7. Chilko River (Chilko, Chilko); 8. Raft River (Kamloops-ES, Raft); 9. Adams River, Lower Shuswap River (Shuswap Complex-L, Lower Shuswap); 10. Scotch Creek, Seymour River (Shuswap Complex-ES, Scotch, Seymour); 11. Gates Channel, Gates Creek (Anderson-ES, Gates); 12. Birkenhead River (Lillooet-L, Birkenhead), Weaver Creek, Weaver Channel (Harrison(U/S)-L, Weaver) 13. Harrison River (Harrison (D/S)-L, Harrison); 14. Cultus Lake (Cultus, Cultus). 70

Figure 3.2. Classification and Regression Tree (CART) model output for the 1977-2008 data set. Each of the tree splits is labelled with the variable and its corresponding values that determine the split. The mean value of PSM is displayed at each of the terminal nodes. Length of lines/branches at each split represent the proportion of total sums of squares (SSQ) explained by each split. Variables chosen at each split are those which minimize within group SSQ and maximize between group SSQ. Variable descriptions are given in Table 2.1. 71

Figure 3.3. Variable importance plots for predictor variables used in the (a) 1977-2008 model and the (b) 1991-2008 model. Percentage increase in mean squared error (%IncMSE) indicates a predictor variables effect on prediction accuracy. Higher values of %IncMSE indicate variables of higher importance to model prediction. The percentage of total variance explained by each model is also shown. 72

Figure 3.4. Partial dependence plots for the 9 most important variables described by our 1977 random forest model. The Y-axis indicates mean value of prespawn mortality rate. The X-axis indicates the value of each predictor variable. Dashmarks along the X-axis indicate deciles. 73

List of Acronyms

ANOVA	Analysis of Variance
ANN	Artificial Neural Network
ASFA	Aquatic Sciences and Fisheries Abstracts
BAT	Bagging Trees
BC	British Columbia
BOT	Boosted Trees
CART	Classification and Regression Trees
CLA	Cluster Analysis
Co	Cobalt
Cu	Copper
CU	Conservation unit
DFO	Fisheries and Oceans Canada
GAM	Generalized additive model
GLM	Generalized Linear model
GWR	Geographically weighted regression
km	Kilometres
m	Metres
MARS	Multivariate Adaptive Regression Splines
MFANOVA	Multi-factor Analysis of Variance
MPB	Mountain pine beetle

MSE	Mean squared error
NPI	North Pacific Index
PD	Partial dependence
PDO	Pacific decadal oscillation
PU	Production unit
PSM	Prespawn mortality
QR	Quantile regression
RF	Random forest
SEM	Structural Equation Modeling
SSQ	Sums of squares
SVM	Support Vector Machine
VI	Variable importance
WOK	Web of Knowledge
WOS	Web of Science

Chapter 1.

General Introduction

Over the past few decades, an increasingly large number of Fraser River sockeye salmon have died prematurely during and after return migration to the spawning grounds (Marmorek *et al.,* 2011). This is especially troubling considering these premature mortalities may be partially responsible for the recent declines of Fraser River sockeye abundances that have resulted in fisheries closures, and economic and ecosystem impacts (Marmorek *et al.,* 2011).

Research suggests that premature mortality in migrating salmon can largely be attributed to increasing amounts of stress (Barton, 2002; Gilhoussen, 1990; Marmorek *et al*., 2011). Over the course of return migration, sockeye salmon encounter a range of biological (e.g. pathogens, predators) and physical (e.g. extreme temperatures, flows/currents) stressors. Exposure to these stressors leads to a suite of physiological and behavioural changes that should increase the chance of overcoming the stressor in the short term (Mazeaud *et al.,* 1977). Sockeye salmon, however, depend on fixed energy reserves to complete their return migration. When stressors persist, these changes can cause sockeye salmon to die prematurely from exhaustion (Barton, 2002).

During migration, sockeye salmon may be repeatedly exposed to a single stressor, such as extreme river temperature. Migrating fish can also experience multiple stressors simultaneously, resulting in combined effects, both additive and non-additive to fish (Crain *et al*., 2008). Each of these phenomenon are often referred to as cumulative effects or impacts (e.g. Crain *et al*., 2008) and understanding their role in premature migration mortality is essential if fisheries managers hope to mitigate their effects.

The ability to quantify cumulative effects relies heavily on different multivariable methods available for analyzing the combined effects of two or more predictor variables on a single response. Hence Chapter 2 of this research project presents (1) a literature review of quantitative methods that are currently being applied to Fraser River sockeye migration survival, with an emphasis on the identification of areas where the application of multivariable statistical approaches could be improved; (2) a review of other potentially useful multivariable methods that are available from the fisheries science, biological and medical realms; and (3) recommendations on how to apply multivariable methods to study cumulative effects on fish migration success.

Chapter 3 of this project presents an application of two multivariable methods discussed in Chapter 2, regression trees and random forests, to describe and predict the cumulative effects of multiple stressor variables on premature mortality in Fraser River sockeye who have reached the spawning grounds (prespawn mortality, PSM). Specifically, the random forest model developed in this chapter will be useful for fisheries managers who could use it as a tool to forecast high PSM rates and adjust sockeye salmon harvest limits to compensate for the loss in spawning ground egg deposition. Alternatively, this model could provide estimates of PSM when obtaining visual estimates of PSM rates at the spawning grounds is impossible. In this chapter, the combined effects of multiple stressors and the effects of single stressors applied repeatedly over migration are considered.

The main goals of this research project are to advance the current state of research surrounding cumulative effects to sockeye salmon migration survival and also to provide useful tools for predicting, managing and mitigating these effects. Additionally, much of this study should serve as an excellent guide for anyone seeking to quantify the cumulative effects of multiple stressors to other animals.

References

Barton, B. A. (2002). Stress in fishes: a diversity of responses with particular reference to changes in circulating corticosteroids. Integrative and Comparative Biology 42, 517–525.

Crain, C. M., Kroeker, K. & Halpern, B. S. (2008). Interactive and cumulative effects of multiple human stressors in marine systems. Ecology Letters 11, 1304-1315.

DFO (Fisheries and Oceans Canada). Canada’s Policy for Conservation of Wild Pacific Salmon. (2005). Available at www.pac.dfo-mpo.gc.ca/fm-gp/species-especes/salmon-saumon/wsp-pss/docs/wsp-pss-eng.pdf

Gilhousen, P. (1990). Prespawning mortalities of sockeye salmon in the Fraser River system and possible causal factors. International Pacific Salmon Fisheries Commission Bulletin XXVI.

Mazeaud, M. M., Mazeaud, F. & Donaldson, E. M. (1977). Primary and secondary effects of stress in fish: some new data with a general review. Transactions of the American Fisheries Society 106, 201–212.

Marmorek, D., Pickard, D., Hall, A., Bryan, K., Martell, L., Alexander, C., Wieckowski, K. Greig, L. & Schwarz, C. (2011). Fraser River sockeye salmon: data synthesis and cumulative impacts. ESSA Technologies Ltd. Cohen Commission Tech. Rep. 6. 273p. Vancouver, B.C. www.cohencommission.ca

Chapter 2.

Quantitative methods for analyzing cumulative effects to fish migration success: a review (published as: Johnson, J. E., Patterson, D. A., Martins, E. G., Cooke, S. J. & Hinch, S.G. (2012). Quantitative methods for analyzing cumulative effects on fish migration success: a review. Journal of Fish Biology 81, 600-631.)

Introduction

Migrating fishes are exposed to a myriad of biological (e.g. pathogens, predators) and physical (e.g. extreme temperatures, flows/currents) stressors that can reduce their *en route* survival. Because fish migrations tend to be cyclical and predictable in both timing and location (Lucas & Baras, 2001), fisheries exploit such patterns, which creates additional challenges (e.g. gear/boat avoidance behaviours, stress from capture and release on non-target species) for migratory individuals (Froese & Torres, 1999; McDowall, 1999). When exposed to a stressor, fishes respond in a rather predictable manner with the elevation of circulating glucocorticoid concentrations (the primary response), which leads to a suite of adaptive physiological and behavioural changes (secondary and tertiary) that should increase the chance of overcoming the stressor in the short term (Mazeaud *et al.,* 1977). When stressors persist (i.e., become chronic), the biological responses in fishes become detrimental (Barton, 2002).

Complicating matters is the fact that exposure to multiple stressors sometimes create combined effects that can alter survival during activities such as migration (Crain *et al*., 2008). The combined effect of multiple stressors is commonly referred to as a cumulative effect (e.g. Crain *et al*., 2008) and herein considered synonymous with cumulative impact. Two or more stressors can combine to produce the following three general outcomes: additive – the combined effect is the simple sum of the individual stressors; synergistic – the combined effect is greater than the sum of the individual stressors; and antagonistic – the combined effect is less than the sum of the individual stressors (Folt *et al*., 1999; Crain *et al*., 2008). For migrating fishes, cumulative effects can include both the combined impact of a single stressor repeatedly applied over time (e.g. repeated interactions with fishing gear or predators during a migration), and the combined impact of multiple stressors applied simultaneously or at different points of space or time (e.g. fisheries interactions combined with a high flow event and cold shock). The above is an inclusive description of cumulative effects, but this review will focus on examining different quantitative methods for assessing the combined impact of multiple stressors that are relevant to fish migration studies.

Fraser River sockeye salmon (*Oncorhynchus nerka*, Walbaum 1792) are a well-studied aggregate population of anadromous fish which migrate every year from the ocean to their freshwater spawning grounds to reproduce and then die (Burgner, 1991). Extensive research has demonstrated that exposure to adverse physical conditions (e.g. high water temperature) and biological factors (e.g. pathogens) during spawning migration can lead to premature mortality, either *en route* to the spawning ground or at the spawning grounds (e.g. Cooke *et al*., 2008). However, even within a system with well-documented multiple impacts, the majority of research has focused on relationships between *O. nerka* migration success and exposure to only one or two stressors (e.g. high temperature and discharge). As the number of identified stressors and factors that mediate the stress response grows, the need for adequate methods to assess the cumulative effects of stressors on migration success for *O. nerka* and other migratory fishes becomes more pressing. Addressing this methodology gap for assessing cumulative effects will be crucial for future management and viability of fish populations, especially due to the anticipated changes to the biological and physical stressors associated with further human development and climate change (IPCC, 2007; U.S. EPA, 2007;).

The ability to quantify cumulative effects relies on both a good study and on different multivariable methods (not to be mistaken for multivariate methods-which involves the analysis of multiple outcomes or responses) available for analyzing the combined effects of two or more predictor variables on a single response. However, simply quantifying the cumulative impact is not always sufficient for fish migration research. In practice, researchers may need to better understand how two or more variables combine (e.g. additive, synergistic or antagonistic) to influence survival. Furthermore, multivariable methods are not all equal in dealing with the types of relationships among predictor and response variables. Currently no clear approach or guideline exists on which method to use in order to best quantify cumulative impacts. Towards achieving that goal, this review represents the following: (1) a compilation of what is currently being applied within a single well-studied fish migration system, with an emphasis on the identification of areas where the application of multivariable statistical approaches could be improved or expanded; (2) a review of other potentially useful multivariable methods that are available from the fisheries science, biological, and medical realms; and (3) a synthesis of both the compilation and review with recommendations on how to apply multivariable methods to study cumulative effects on fish migration success. This overview and evaluation of quantitative methods gathered from a diverse range of scientific fields should serve as a primer for anyone seeking to quantify cumulative effects of multiple variables to fish migration success. Indeed, this framework could also be applied to other migratory animals.

Compilation of Multivariable Papers on Fraser River *O. nerka*

Background

Fraser River *O. nerka* are comprised of several hundred distinct spawning populations experiencing differences in migration timing (July to Nov), distances (50-1150 km), elevation (10-1000 m) and thermal conditions (4-22°C) (Burgner, 1991). Therefore, population differences in physiology, behavior and morphology, such as thermal tolerance, energy use, swim strategies and body shape are thought to be local adaptations in response to migratory stressors (Crossin *et al*., 2004*b*; Eliason *et al*., 2011). Even within a population, individual variability in response to changing migratory stressors will likely drive the future evolution of these populations (e.g. Reed *et al*., 2011). Numerous studies have examined the effects of multiple variables on migration success of Fraser River *O. nerka* and this relatively large body of research now offers an excellent opportunity for assessing several, commonly used, multivariable statistical methods. The lessons learned from this comparatively well-studied system (e.g. Hinch *et al*., 2005) on the current ability to quantify the effects of multiple variables are likely exportable to other fish migration studies, or even to studies of other taxa.

Compilation Methods

Research articles were compiled using an internal Fisheries and Oceans Canada (i.e., DFO; the Federal science-based government agency responsible for fisheries research and management) database that has kept track of Fraser River *O. nerka* publications (accessed by D. Patterson in September 2011). Papers published since 1990 that focused on upstream spawning migration and had a minimum of two predictor variables (continuous or categorical) used to describe a migratory stress response were selected. These papers included examples where more than one predictor variable was a migratory stressor (e.g. temperature, pathogens), and examples of a where a single stressor was analyzed along with other non-stress factors (e.g. stock, sex). Both types of examples were assumed to showcase multivariable statistical methods that could be used to study cumulative effects. The response could either be lethal or involve sub-lethal outcomes. Sub-lethal outcomes that could potentially impact fish migration success were categorized as physiological (including a variety of disease states and health conditions) or behavioural responses. The specific focus of this compilation was on the range of quantitative methods currently being applied and to determine the range and type of different predictors and response variables being used.

Compilation Results

A total 56 papers dealing with multiple predictor variables related to spawning migration success in Fraser River *O. nerka* from 1990 year to 2011 were found (Table 2.1). There were a total of five different multivariable statistical methods used (see Table 2.2 for brief description). The majority of papers presenting multivariable analyses applied multi-factor analysis of variance (MFANOVA) (Table 2.1; Figure 2.1). A few articles used multiple linear regression (MLR), generalized linear model (GLM), non-linear regression (NLR) or survival analysis (SA) to address migration success (Table 2.1; Figure 2.1).

Interestingly, not all of the papers used traditional analyses (Table 2.1). In several cases unique simulation type models were built to describe the relationship between two or more variables. This included very system specific simulation models that were constructed based on known relationships and interactions between stressors (Hague *et al*., 2011; Reed *et al*., 2011). Several articles also used ordination methods for both variable reduction (e.g. principle component analysis; Macdonald *et al*., 2010) and visual interpretation of multivariable interactions (non-metric multi-dimensional scaling; Cooperman *et al*., 2010). Another approach included categorical data analysis techniques, such as the contingency chi-square or Fisher's exact test (Roscoe *et al*., 2011). However, ordination methods encountered were not used as multivariable models of cumulative effects, and simulation models were not readily adaptable to other systems, therefore neither approach was included in subsequent discussions.

Over one third of all papers in the database applied non-multivariable approaches (Table 2.1). Many of these papers chose a series of univariable analyses to describe or infer the individual effect of two or more variables separately. These univariable methods included t-tests, one-way ANOVAs, simple linear regressions, and correlation analyses (Table 2.1).

Approximately 60% (n=36) of the papers dealt directly with mortality, 20 papers dealt exclusively with sub-lethal responses, and 19 papers reported on both sub-lethal responses and mortality (Table 2.1). Only 16 papers dealt with sub-lethal effects on behaviour compared to 27 papers dealing with physiological responses. The behavioural endpoints focused on changes in timing (e.g. Cooperman *et al*., 2010), migration rate (e.g. Hanson *et al*., 2008), or holding behaviour (e.g. Mathes *et al*., 2010). The physiological responses were more variable and included energy use, cardio-vascular performance, disease, osmoregulation, reproduction, and general stress response. For most of the physiological response variables used, a biological rationale was provided for how the sub-lethal effect related to overall survival during migration, although in some cases such links are speculative and/or tenuous.

For the two most common physical predictor variables used, water temperature and discharge (Table 2.1), there were examples applied to describe a direct impact on survival, and sub-lethal impacts on behaviour and physiology. The biological predictors spanned the range of biological organization levels and include the following examples; fish abundance (Macdonald *et al*., 2010), aerobic performance (Hague *et al*., 2011), energy density (Rand *et al*., 2006), pathogen presence (Wagner *et al*., 2005), ion levels (Bradford *et al*., 2010*a*; Jeffries *et al*., 2011), and transcriptional gene expression (Miller *et al*., 2011).

Review - Quantitative Methods Applied within Fisheries , Biological and Medical Sciences

Review Background

The three most intriguing results from the search of multivariable papers dealing with Fraser River *O. nerka* migration were the ubiquitous use of MFANOVAs, the limited number of multivariable methods used, and the non-ideal application of univariable techniques to deal with multivariable issues. These areas clearly speak to the need to review and consider adopting alternative and in some cases more appropriate quantitative methods for dealing with multivariable problems. This stimulated the need to review the fisheries, biological and medical literature to look for other multivariable methods being used to investigate cumulative effects.

Review Methods

Fisheries Science Literature

The main purpose of this key-word based, systematic literature search was to find additional multivariable methods that could be used to assess cumulative effects on *O. nerka* and other migratory fishes. In order to reduce the large number of articles available for review and also to ensure that techniques were suitable for analyzing cumulative effects to fish migration success, the search was restricted to articles from fisheries science literature analyzing effects on fishes. Two search engines, the ISI Web of Knowledge (WOK) and Aquatic Sciences and Fisheries Abstracts (ASFA), were used and included articles published from 1900 to August 2011.

A large number of article titles, abstracts or texts contain words like “cumulative” or “fish”, therefore the search was restricted to identify articles which contained at least four keywords from the following three groupings: (fish or fishes) AND (statistic* or quantitative or math* or multivar* or numerical) AND ((cumulative or synergy* or additive or non-additive or antagonistic or interaction) SAME (impact* or effect* or stress* or pressure* or result* or consequence* or outcome* or response*)). The Boolean operator “SAME” denotes that a keyword from each subgroup must be identified in the same sentence within an article’s title, keywords or abstract. Hence, one keyword was required from the first two groups and two keywords were required from the third group. In the ASFA search, the operator “NEAR” was used instead of “SAME” to yield similar results. The ASFA search was further restricted so that words from the second and third groupings were required within the abstracts of papers returned. Furthermore, the WOK search was restricted to only identify articles found within the Web of Science (WOS) online database pertaining to Fisheries Science, Ecology, Oceanography, Marine and Freshwater Biology and Environmental Science categories.

After these criteria were imposed, over 900 articles remained and were added to a database. All duplicate articles and those that did not use quantitative, multivariable techniques to analyze fish responses were removed from the database.

Non-Fisheries Literature

Fisheries-based literature contains a variety of different studies using multivariable techniques. However, other methods not common to fisheries sciences, hereafter described as novel methods, might be found within research journals of other disciplines. Therefore, two searches were conducted using the WOK to find methods uncommon in fisheries research.

The first search looked for methods used in general biology and was restricted to identify scientific journal articles published from January 2007 to August 2011 which contained at least three keywords from the following two groupings: (statistic* or quantitative or math* or multivaria* or numerical) AND ((cumulative or synergy* or additive or non-additive or antagonistic or interaction) SAME (effect* or stress* or pressure* or result* or consequence* or outcome* or response*)). The search was further refined to identify only articles from the WOS in the fields of Ecology and Biology;

however, this restriction still identified over 2,200 different papers. Hence a random sample of 150 papers from the 2,200 was examined.

Another expectation was that the field of medicine would contain a wealth of literature using novel techniques to assess cumulative effects uncommon to any ecological or biological science. A second search was conducted using the MEDLINE database for articles published from January 2007 to August 2011 using the same keyword groupings as in the above search through biological sciences. Again, these search parameters returned thousands of papers, so a random sample of 150 papers was examined from these search results.

Each multivariable method found within the various searches was classified based on the following categories: categorical and numerical (i.e., continuous or discrete) response variables; whether or not a researcher needs to specify or test for non-linearities and interactions in the modeling process; and whether a model type is parametric, non-parametric or semi-parametric (i.e., having both parametric and non-parametric components). Methods that do not require researchers to test for or specify how nonlinearities and interactions are included in a model are especially useful when modeling a large number of variables with potentially complex relationships.

A number of papers identified in the literature searches mentioned the use of novel quantitative methods but failed to describe that method with sufficient detail. Therefore reference sections and statistical textbooks were searched to locate more comprehensive descriptions of each method and examples of usage within fisheries sciences. Several of these references, along with short descriptions of each method are provided to assist in reader comprehension. Additionally, some examples of functions or packages with free R statistical software that can be used to perform each method are provided.

Review Results

The systematic search of fisheries science literature identified 88 articles that used multivariable methods to study fishes (Appendix A) and 14 distinct multivariable methods (Figure 2.1). Only two of these articles using two distinct methods (MLR and GLMs) analyzed fish migration success. Ten of the 14 methods identified were not applied to study Fraser River *O. nerka* (Figure 2.1) and 42 out of 87 articles found in this literature search used at least one of these 10 methods (Appendix A). The majority of articles that applied these 10 techniques, with the exception of those using Generalized Additive Models (GAMs), Artificial Neural Networks (ANNs), and Cluster Analysis (CLA), were published after the year 2005. The most common technique used for analyzing effects of multiple variables within these articles was GAMs (30 papers), followed by MFANOVA (25 papers), MLR (24 papers) and then GLMs (17 papers) (Figure 2.1). The number of predictor variables used in conjunction with GAMs, MLR, GLMs, ANNs NLR and Boosted Trees (BOTs) varied throughout the literature while MFANOVA studies typically used the fewest predictors (Figure 2.1). Other methods were used too infrequently in studies to get any real sense of the range of number of predictor variables used.

The search using a random sample of 150 papers in general biology found an additional three papers that discussed two unique methods not found within the searches of fisheries science literature. These methods were structural equation modeling (SEM) and geographically weighted regression (GWR). No unique methods examining cumulative effects were found in the search of 150 randomly selected medical papers.

Overall, the search through DFO's database, fisheries science and biological literature yielded a total of 17 different multivariable methods that could be used to assess cumulative impacts to fish migration success (Table 2.2). Five techniques are able to model either numerical or categorical response variables, 8 can only model numerical responses and 4 can model only categorical response (Table 2.2). Eight techniques do not require a researcher to specify non-linearities and interactions (Table 2.2). Seven methods were parametric, 7 were non-parametric, two techniques are considered semi-parametric and one (SA) has all 3 forms (Table 2.2).

Discussion

The compilation using the Fraser River *O. nerka* system identified only a few multivariable methods that could be used in analyses of cumulative effects on fish migration. The widespread use of MFANOVAs for multivariable analyses possibly reflects their appealing simplicity and robustness as well as easy implementation within various readily available statistical software packages. MFANOVAs are not suited to dealing with large numbers of predictor variables (Ginot *et al*., 2006) and their extensive use is also likely related to the fact that Fraser River sockeye migration studies have typically analyzed the effects of only a few predictor variables. While the use of MFANOVA may have been justified in these cases by specific study limitations or goals, researchers wishing to study cumulative effects using a much broader set of predictor variables should consider using more suitable techniques.

Only a few studies analyzed survival using GLMs (i.e. logistic regression) or survival analysis techniques, despite the interest in the effects of multiple stressors on migration success of Fraser River *O. nerka*. For example, survival data was usually analyzed with chi-square-type tests (e.g. Mathes *et al*., 2010; Roscoe *et al*., 2011) or used as a categorical "predictor" to compare the physiology of fish that were successful or failed on their migration (e.g. Crossin *et al*., 2009*a*; Donaldson *et al*., 2010). These approaches are useful for comparing characteristics between successful and unsuccessful fishes, but they cannot be used for predicting survival and elucidating how the odds of survival vary with changes in stressors and underlying physiology. Miller *et al*. (2011) provides an example of using survival analysis to both discriminate the physiologies amongst fate groups as well as calculate the odds of survival as a function of the underlying gene expression patterns.

Another interesting observation in the Fraser River *O. nerka* literature was the common application of several univariable tests to problems that could be dealt with using multivariable techniques (e.g. Pon *et al*., 2009*b*; Hruska *et al*., 2010). While the application of univariable procedures may have been justified by limitations or the research goals for each of these studies, these authors missed the opportunity to determine how a predictor variable influences the response in the presence of another explanatory variable and whether these variables interact with one another to produce synergistic or antagonistic effects (Kaplan, 2009). Future research should therefore avoid using univariable methods when multivariable approaches are an option in order to improve understanding of cumulative effects.

The breadth of potential relationships between biophysical factors and migratory success at different levels of biological organization that have been explored in Fraser River *O. nerka* makes it a reasonable surrogate for studying other fish migrations. The predictor variables ranged in variety of aquatic environments from coastal marine, to large rivers and lake environments. The lethal response criteria also covered both immediate and delayed mortality associated with cumulative effects. The range of sub-lethal responses for physiology was considerably more varied than behaviour, which is likely a reflection of the difficulty in assessing fish behavior in the wild as well as a lack of analytical techniques for characterizing fish behaviour and spatial ecology across broad spatial scales. Collectively, this information should be of benefit to other migration studies for examples of what to do and what to avoid in multivariable analysis of cumulative effects.

The search through fisheries science and biology literature yielded information on 12 multivariable techniques that have not been used to study Fraser River *O. nerka* migration success. Several reasons may explain the lack of adoption of these alternative methods within the Fraser River system and potentially elsewhere in the fish migration literature. Foremost, fish migration researchers may be unfamiliar with their use. The majority of these techniques (with the exception of GAMs and ANNs) appear to have only been applied in fields of fisheries research very recently (since 2005) and infrequently if at all. Therefore, newer techniques, such as BOTs, Random Forests (RFs), and Support Vector Machines (SVMs) were possibly unknown to scientists modeling cumulative effects to *O. nerka*. Many methods, such as ANNs, GAMs and SVMs are also fairly complex and computationally intensive and past researchers may have lacked the necessary knowledge, time and computer power to run these models. However, these computational restrictions no longer apply as all methods outlined in this review can be run quickly and easily on a standard laptop with free software packages such as R (Table 2.2).

Finally, even in the case of such a well-studied system like Fraser River *O. nerka*, data can be sparse or limiting. Complex techniques can sometimes require large amounts of data to properly build and parameterize the model. For example, ANNs often require thousands of data cases to be trained properly (Hill & Lewicki, 2007). The adoption of more simple methods, like MFANOVAs, is understandable in situations where data is less available, and where a justification for much more complex method does not exist. However, these limitations are becoming less restrictive with the adoption of new technologies available to researchers that can assist in data collection, information sharing, and data collation. Examples of each include: recent advances in electronic tagging technologies, such as the development of miniaturized and multi-sensor tags (i.e. which records variables related to fish behaviour, physiology, and environmental conditions) which enable researchers to collect data on migrating fishes that was unattainable in the past (Cooke *et al*., 2004*b*, in press); the recent development of the free, online database, Movebank, allows researchers to easily share and track animal movement data (Kranstauber *et al*., 2001); and the push by regional, national and international science communities or governments to collate physical and biological information into large accessible databases (e.g.: work on ecosystem management - e.g. Canadian Aquatic Biomonitoring Network Database, www.ec.gc.ca;). Improvements to data management and accessibility in conjunction with readily available software (i.e. R) should facilitate the adoption of the quantitative multivariable methods techniques described in fisheries science.

The searches through the ecological and biological science literature yielded only two additional multivariable techniques, SEM and GWR, whereas the search through the medical literature did not result in any additional techniques. This lack of results from the medical field was contrary to the original expectation that the medical field would contain information on a number of techniques not used in ecology or biology. A simple explanation could be that the sample of 150 papers may have been too small to capture any novel techniques being used in the field of medicine.

Each of the 12 additional techniques found within the searches are applicable to assessing cumulative effects to fish migration success. Quantile Regression (QR) and GAMs each provide useful extensions of the typical MLR model. QR for example, is an alternative form of MLR that approximates the median and other quantiles of a response variable as opposed to the mean. Estimating a number of regression quantiles can potentially reveal multiple rates of change (slopes) between a predictor and response variable, therefore providing a more complete picture of how variables relate to fish migration success in the presence of ecological stressors or limiting factors (e.g. prey availability) (Cade, 2003). GAMs could be used to predict survival based on a number of variables in as similar manner as a GLM. Their use of non-parametric smoothers can help provide excellent fit to the data in the presence of non-linear relationships (Hill & Lewicki, 2007). However, the relationships modeled by GAMs can be difficult to interpret, making them less useful for researchers interested in exploring potential relationships within a system (Elith *et al*., 2008).

Non-parametric methods, such as Multivariate Adaptive Regression Splines (MARS), Tree learning methods, ANNs and SVMs and CLA all differ in terms of their underlying mechanics. However, each of these methods can inherently model or take into account non-linear relationships and interactions between variables to assess or predict cumulative impacts of multiple variables to fish migration (Prasad *et al*., 2006; De'ath, 2007; Hill & Lewicki, 2007) Hence, they do not require a researcher to specify these types of relationships when building the model and are all excellent choices when a researcher needs to model a large number of predictor variables with complex relationships and interactions. However, SVMs, CLA and Classification and Regression Trees (CART) for are only useful for categorizing a response variable (Moguerza & Munoz, 2006; Hill & Lewicki, 2007). Furthermore, the relationships modeled by some of these methods (e.g., SVMs, CLA, RFs, BATs and ANNs) can, like GAMs, be difficult or impossible to interpret making them more suited for predictive purposes.

The first technique found in the general biology search, GWR, is a relatively new technique that allows model parameters to vary spatially (Austin, 2007) and could be applied to situations where the effects of stressors to fish survival vary across migratory locations. For example, the effects of high water temperatures to Fraser River *O. nerka* are possibly exacerbated in areas where fish cannot find thermal refugia in the form of lakes or cold tributaries (Donaldson *et al*., 2009). However, GWR has difficulty distinguishing between non-linear relationships and spatial non-stationarity (Austin, 2007) and is therefore not suited to modeling complex relationships between predictor and response variables. The second technique, SEM, is a form of pathway analysis that could easily be applied to problems in fisheries science. SEM provides information on the magnitude of direct and indirect relationships between variables (Palmores *et al*., 1998). A possible application would be to help analyze how land use factors like deforestation or agricultural activity indirectly influence fish migration survival by influencing freshwater stressors like extreme water temperatures and pollution. The main problem is that functional relationships modeled in SEM are assumed to be linear and therefore is less suitable to situations involving complex relationships between variables (Austin, 2007).

Finally, a cautionary note regarding searching explicitly for papers that deal with cumulative effects and/or cumulative impacts. A major reason why the WOK and ASFA literature searches found no Fraser River *O. nerka* papers from the DFO database was that none of the *O. nerka* papers reviewed make specific reference to being a cumulative effects assessment. There are likely numerous papers in fisheries science or other disciplines that deal with multivariable approaches that are suitable for assessing multiple stressors in fish migration, but do not necessarily portray themselves as cumulative effects methodology. Therefore it is important in searching for examples of different methods for assessing multivariable techniques that are relevant to fish migration success, to have a sense of the potential methods available. This reinforces the importance of this review paper.

Conclusion

Researchers are becoming aware of the ever-increasing number of biological and environmental stressors (e.g. dissolved oxygen, pathogens, predators, contaminants) that could potentially affect fish migration success. As the number of potential stressors, moderating factors, and interactions grows, so does the need to adopt a wider range of multivariable techniques for analyzing and predicting their effects. As exemplified by this review, a number of different quantitative techniques could be used to model cumulative effects of different variables to fish migration success. With such a wide range of techniques available the best technique to use in each situation will depend on a number of factors like: availability of data, complexity of relationships between variables, and the overarching goal of the research. However, the use of promising new methods outlined in this review (e.g., CART, MARS, RFs or BOTs), can help researchers to move past using MFANOVAs to both predict actual survival and to describe the potential cumulative effects. Therefore, while little has been done in the past to formally quantify cumulative effects to migration success of fishes, the use of newer quantitative methods, in combination with ever increasing computing power will enable researchers to gain much better insight into the cumulative effects of multiple variables to fish migration success.

Tables

Table 2.1 Classification of all O. nerka papers found within the Fisheries and Oceans Canada (DFO) database.

Model	Physical predictor	Biological predictor	Response variables			Reference
MFANOVA	Discharge	Activity, Heart rate, Sex			P	Clark *et al.* (2010)
MFANOVA	Handling	Physiology, Stock, Sex	L	B		Cooke *et al.* (2005)
MFANOVA		Stock, Physiology	L			Cooke *et al.* (2006*b*)
MFANOVA		Timing, Sex, Stock			P	Cooke *et al.* (2008)
MFANOVA		Physiology, Stock, Sex	L	B		Crossin *et al.* (2007)
MFANOVA	Temperature	Sex	L	B	P	Crossin *et al.* (2008)
MFANOVA		Stock, Sex, Physiology	L	B		Crossin *et al.* (2009*a*)
MFANOVA		Stock, Sex, Physiology, Timing	L	B	P	Crossin *et al.* (2009*b*)
MFANOVA		Physiology, Sex, Timing	L	B		Donaldson *et al.* (2010)
MFANOVA	Year, Location	Sex		B		Hinch & Rand (1998)

MFANOVA		Osmoregulation, Timing	L			Jeffries *et al.* (2011)
MFANOVA	Temperature	Sex, Timing, Holding	L		P	Mathes *et al*. (2010)
MFANOVA	Discharge	Handling			P	Pon *et al*. (2009*a*)
MFANOVA	Location, Handling	Sex	L	B	P	Roscoe *et al*. (2011)
MFANOVA	Location, Discharge	Stock			P	Shrimpton *et al.* (2005)
MFANOVA	Temperature	Behaviour	L	B	P	Steinhausen *et al*. (2008)
MFANOVA	Temperature	Pathogens	L		P	Wagner *et al.* (2005)
MFANOVA	Salinity	Activity			P	Wagner *et al.* (2006)

MLR	Temperature, Discharge	Abundance, Timing	L			Gilhousen (1990)
MLR, MFANOVA	Temperature	Stock, Sex, Physiology		B		Hanson *et al.* (2008)
MLR, PCA	Discharge, Temperature	Abundance, Stock, Timing	L			Macdonald *et al.* (2010)
MLR, GLM	Temperature			B		Roscoe *et al.* (2010)

GLM	Temperature, Location, Year		L			Martins *et al.* (2011)
GLM	Temperature, Location	Stock	L			Martins *et al.* (2012)

GLM, NLR	Year	Timing, Sex	L			Bradford *et al.* (2010*b*)
GLM, MFANOVA		Physiology, Pathogen	L		P	Bradford *et al.* (2010*a*)
GLM, MFANOVA	Discharge, Handling	Sex, Physiology	L		P	Patterson *et al.* (2004)
NLR	Temperature, Discharge				P	Rand *et al.* (2006)
SA		Genomics, Stock, Sex	L			Miller *et al.* (2011)
SA	Discharge	Sex	L			Nadeau *et al.* (2010)
Simulation Models	Temperature	Stock	L			Hague *et al.* (2011)
Simulation Models	Location, Discharge	Activity			P	Hinch *et al.* (1996)
Simulation Models	Discharge, Temperature		L	B		Reed *et al.* (2011)
Non Multivariable	Temperature	Energy, Timing, Physiology	L	B		Cooke *et al.* (2004*a*)
Non Multivariable		Physiology, Energetics	L			Cooke *et al.* (2006*a*)
Non Multivariable		Physiology, Energetics	L			Cooke *et al.* (2009)

Non Multivariable		Physiology	L	B	P	Cooperman *et al.* (2010)
Non Multivariable	Year, SST, NPI	Stock, Standard length			P	Crossin *et al*. (2004*a*)
Non Multivariable	Location	Stock			P	Crossin *et al.* (2004*b*)
Non Multivariable	Location, Discharge	Migration Rate, Stock, Timing		B		Donaldson *et al.* (2009)
Non Multivariable		Handling	L	B	P	Donaldson *et al.* (2011)
Non Multivariable	Temperature	Timing, Stock, Physiology			P	Eliason *et al.* (2011)
Non Multivariable	Temperature	Stock, Physiology	L			Farrell *et al.* (2008)
Non Multivariable	Discharge	Swim Behaviour	L			Hinch & Bratty (2000)
Non Multivariable	Discharge	Stock		B		Hinch & Rand (2000)
Non Multivariable		Stress, Sex, Timing			P	Hruska *et al.* (2010)
Non Multivariable		Longevity, Timing, Size	L			Hruska *et al.* (2011)
Non Multivariable	Location	Stock			P	Kelly *et al.* (2011)
Non Multivariable	Temperature	Stock			P	Lee *et al.* (2003*a*)
Non Multivariable	Temperature	Stock			P	Lee *et al.* (2003*b*)
Non Multivariable	Temperature, Discharge		L	B		Macdonald (2000)
Non Multivariable	Temperature, Discharge	Disease	L	B	P	Macdonald *et al.* (2000)

Non Multivariable	Temperature, Discharge		L			Macdonald *et al.* (2007)
Non Multivariable	Temperature	Species			P	MacNutt *et al.* (2006)
Non Multivariable		Behaviour, Physiology	L			Pon *et al.* (2009*b*)
Non Multivariable		Behaviour, Physiology	L			Young *et al.* (2006)

MFANOVA=Multi-Factor Analysis of Variance, MLR=Multiple Linear Regression, PCA=Principal Component Analysis, GLM= Generalized Linear Model, SA=Survival Analysis, NLR= Non-linear Regression, L= Lethal, B= Behavior, P=Physiology, NPI= North Pacific Index, SST= Sea Surface Temperature.

Table 2.2 Model classifications along with descriptions and examples of appropriate routines in R statistical software and references which can act as an initial guide for choosing between different techniques. Models are ordered in the list according to their classifications

Model	Numerical or categorical response variable	Non-linear relations must be specified in the modeling process	Interactions must be specified in the modeling process	Parametric, non-parametric or semi-parametric technique	Description	R guide	Ref. for fish migration (**) and fish	Ref. / Texts for methods in detail
BAT	Either	No	No	Non-parametric	Creates multiple bootstrapped classification and regression trees and then averages the results.	'ipred' package	Knudby *et al.* (2010)	Breiman (1996); De'ath (2007); Prasad *et al.* (2006)
RF	Either	No	No	Non-parametric	Similar to BATs except a random set of predictor variables are	'randomForest' package	Knudby *et al.* (2010)	Breiman (2001); De'ath (2007);

					used to build each tree.			Prasad *et al*. (2006)
BOT	Either	No	No	Non-parametric	A sequence of simple classification and regression trees where each tree improves prediction ability	'gbm' package	Elith et al. (2008); Leathwick *et al.* (2006*a*); Leathwick *et al*. (2008)	De'ath (2007); Moisen *et al*. (2006); Sutton (2005)
MARS	Either	No	No	Semi-parametric	Partitions data-space into regions then fits a regression line to each region.	'earth' package	Elith & Leathwick (2007); Leathwick *et al*. (2005)	Friedman (1991); Hill & Lewicki (2007); Prasad *et al*. (2006)
ANN	Either	No	No	Non-parametric	Complex, predictive modeling technique inspired by the neural architecture of the brain.	'neuralnet' package	Olden & Jackson (2001); Palialexis *et al*. (2011)	Carling (1992); Gutierez-estrada *et al*. (2009); Hill & Lewicki (2007)
CART	Either	No	No	Non-parametric	A method for determining a set of decision rules about how different predictor variables influence a response variable.	'tree' & 'rpart' package	**Ostergren *et al*. (2011); Ruppert *et al*. (2010)	De'ath & Fabricious (2000); Harrell (2001); Prasad *et al.* (2006)

CLA	Categorical	No	No	Non-parametric	Can be used to classify a response based on values of predictors.	'stats' or 'cluster' package	Grossman *et al*. (1998); Hinz *et al*. (2009)	Everitt *et al*. (2011); Hill & Lewicki (2007); Rhomsberg (2004)
SVM	Categorical	No	No	Non-parametric	Projects the predictors into higher dimensional space to find a linear classifier.	'e1071' package	Knudby *et al*. (2010)	Hastie *et al*. (2009); Hill & Lewicki (2007); Moguerza & Munoz (2006)
GLM	Either	Yes	Yes	Parametric	Similar to MLR except predictor variables are linearly related to the expected value of a response through a link function.	glm() function	**Bradford *et al.* (2010*a*); Cheng & Gallinat (2004)	Harrell (2001); Hill & Lewicki (2007); McCullagh & Nelder (1989)
GAM	Numerical	Yes	Yes	Semi-parametric	Similar to a GLM except that unspecified functions relate the predictor variables to the expected value of a response.	'mgcv' package	Leathwick *et al.* (2006*b*); Knudby *et al.* (2010);	Hastie & Tibshirani (1990); Hastie *et al.* (2009); Hill & Lewicki (2007)
SA	Numerical	Yes	Yes	All types exist	Suite of methods which model time until an event (e.g. mortality).	'survival' package	**Miller *et al.* (2011); Nadeau *et al* (2010); **	Allison (2010); Harrell (2001); Hill & Lewicki (2007)

NLR	Numerical	Yes	Yes	Parametric	Similar to MLR except predictor variables are non-linearly related to the response variable through a known function.	nls() function	**Bradford *et al.* (2010*b*); Laetz *et al.* (2009)	Hill & Lewicki (2007); Huet *et al.* (1996); Smyth (2002)
GWR	Numerical	Yes	Yes	Parametric	An extension of MLR used when model parameters are not constant over the spatial extent of study.	'spgwr' package		Austin (2007); Foody (2004); Fotheringham *et al.* (2002)
SEM	Numerical	Yes	Yes	Parametric	A form of pathway analysis which analyzes the magnitude of direct and indirect relationships between variables.	'sem' package		Austin (2007); Grace (2008); Palmores *et al.* (1998)
QR	Numerical	Yes	Yes	Parametric	Form of MLR which estimates the median or other quantiles of the response variable.	'quantreg' package	Cade & Noon (2003); Dunham *et al.* (2002)	Austin (2007); Koenker & Basset (1978); Koenker & Hallock (2001)
MLR	Numerical	Yes	Yes	Parametric	Models the relationship between two or more predictor variables and a response variable by fitting a linear equation to observed data.	lm() function	**Hanson *et al.* (2008); **MacDonald (2010)	Harrell (2001); Hill & Lewicki (2007); Zar (1984)

MFANOVA	Numerical	Yes	Yes	Parametric	Assesses the average contribution (main effect) of a few predictors and interactions between variables to the overall mean of a response.	aov() function	Blake & Duffy (2010); Clarieaux & Lagardere (1999);	Hill & Lewicki (2007); Roberts & Russo (1999)

MFANOVA= Multi-Factor Analysis of Variance, MLR= Multiple Linear Regression, QR= Quantile Regression, SA= Survival Analysis, GLM= Generalized Linear Model, GAM= Generalized Additive Model, NLR= Non-linear Regression, MARS= Multivariate Adaptive Regression Splines, CART=Classification & Regression Trees, BAT= Bagging Tree, BOT=Boosted Tree, RF=Random Forest, CLA= Cluster Analysis, ANN= Artificial Neural Network, GWR= Geographically Weighted Regression, SEM= Structural Equation Modeling

Figures

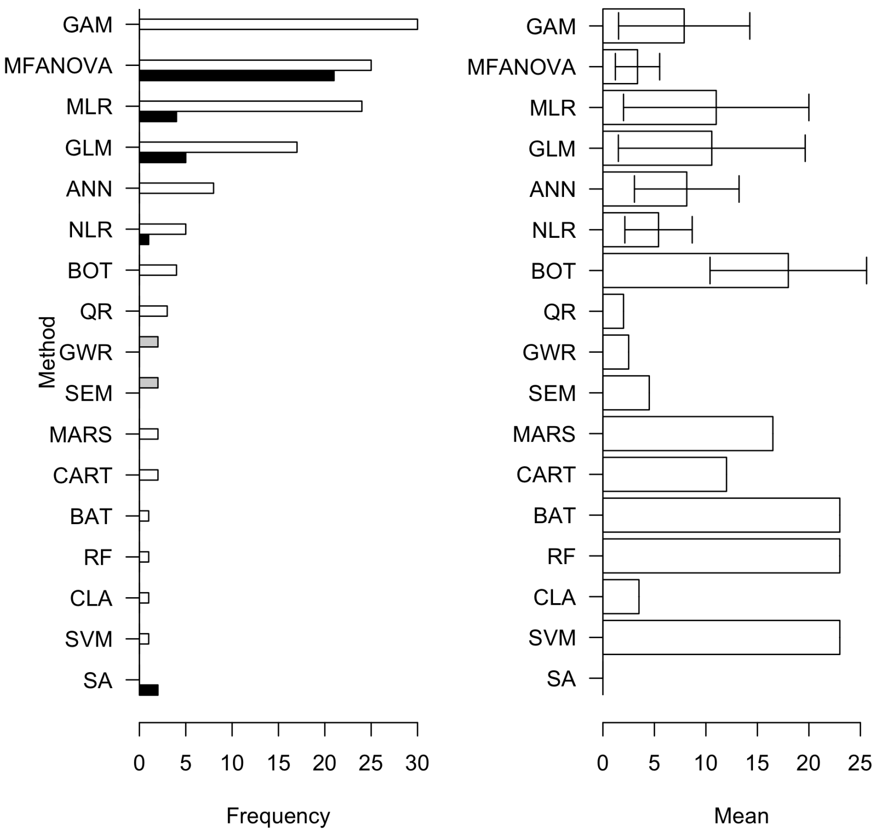

Figure 2.1. *Frequency distribution (left) for multivariable methods used within the Fraser River O. nerka literature (black), fisheries literature (white) and other biological science literature (grey) along with mean and standard deviation (right) for number of predictors used in the fisheries and other biological science literature. MFANOVA= Multi-Factor Analysis of Variance, MLR= Multiple Linear Regression, QR= Quantile Regression, SA= Survival Analysis, GLM= Generalized Linear Model, GAM= Generalized Additive Model, NLR= Non-linear Regression, MARS= Multivariate Adaptive Regression Splines, CART=Classification & Regression Trees, BAT= Bagging Tree, BOT=Boosted Tree, RF=Random Forest, CLA= Cluster Analysis, ANN= Artificial Neural Network, GWR= Geographically Weighted Regression, SEM= Structural Equation Modeling.*

References

Allisson, P. D. (2010). Survival Analysis Using SAS®: A Practical Guide. 2nd edn. Cary, NC: SAS Institute Inc.

Austin, M. (2007). Species distribution models and ecological theory: a critical assessment and some possible new approaches. Ecological Modeling 200, 1–19.

Auth, T. D., Brodeur, R. D., Soulen, H. L., Ciannelli, L. & Peterson, W. T. (2011). The response of fish larvae to decadal changes in environmental forcing factors off the Oregon coast. Fisheries Oceanography 20, 314–328.

Ayllon, D., Almodovar, A., Nicola, G. G. & Elvira, B. (2009). Interactive effects of cover on hydraulics in brown trout habitat selection patterns. River Research and Applications 25, 1051–1065.

Bacheler, N. M., Bailey, K. M., Ciannelli, L., Bartolino, V. & Chan, K. (2009). Densitydependent,landscape, and climate effects on spawning distribution of walleye pollock Theragra chalcogramma. Marine Ecology Progress Series 391, 1–12. doi: 10.3354/

Barton, B. A. (2002). Stress in fishes: a diversity of responses with particular reference to changes in circulating corticosteroids. Integrative and Comparative Biology 42, 517–525.

Blake, R. E. & Duffy, J. E. (2010). Grazer diversity affects resistance to multiple stressors in an experimental seagrass ecosystem. Oikos 119, 1625–1635.

Blanc, J. M. (2005). Contribution of genetic and environmental variance components to increasing body length in juvenile brown trout Salmo trutta. Journal of the World Aquaculture Society 36, 51–58.

Bradford, M. J., Lovy, J. & Patterson, D. A. (2010a). Infection of gill and kidney of Fraser River sockeye salmon, Oncorhynchus nerka (Walbaum), by Parvicapsula minibicornis and its effect on host physiology. Journal of Fish Diseases 33, 769–779.

Bradford, M. J., Lovy, J., Patterson, D. A., Speare, D. A., Bennett, W. R., Stobbart, A. R. & Tovey, C. P. (2010b). Parvicapsula minibicornis infections in gill and kidney and the premature mortality of adult sockeye salmon Oncorhynchus nerka from Cultus Lake, British Columbia. Canadian Journal of Fisheries and Aquatic Sciences 67, 673–683.

Breiman, L. (1996). Bagging predictors. Machine Learning 24, 123–140.

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

Brenden, T. O., Wang, L. Z., Clark, R. D., Seelbach, P. W. & Lyons, J. (2007). Comparison between model-predicted and field-measured stream habitat features for evaluating fish assemblage-habitat relationships. Transactions of the American Fisheries Society 136, 580–592.

Brosse, S., Lek, S. & Dauba, F. (1999). Predicting fish distribution in a mesotrophic lake by hydroacoustic survey and artificial neural networks. Limnology and Oceanography 44, 1293–1303.

Buisson, L., Blanc, L. & Grenouillet, G. (2008). Modelling stream fish species distribution in a river network: the relative effects of temperature versus physical factors. Ecology of Freshwater Fish 17, 244–257.

Burgner, R. L. (1991). Life history of sockeye salmon (Oncorhynchus nerka). In Pacific Salmon Life Histories (Groot C. & Margolis L., eds), pp. 3–117. Vancouver, BC: UBC Press.

Cabral, J. A., Mieiro, C. L. & Marques, J. C. (1998). Environmental and biological factors influence the relationship between a predator fish, Gambusia holbrooki, and its main prey in rice fields of the Lower Mondego River Valley (Portugal). Hydrobiologia 382, 41–51.

Cade, B. S. & Noon, B. R. (2003). A gentle introduction to quantile regression for ecologists. Frontiers in Ecology and the Environement 1, 412–420.

Carling, A. (1992). Introducing Neural Networks. Wilmslow: Sigma Press.

Carol, J., Benejam, L., Alcaraz, C., Vila-Gispert, A., Zamora, L., Navarro, E., Armengol, J. & Garcia-Berthou, E. (2006). The effects of limnological features on fish assemblages of 14 Spanish reservoirs. Ecology of Freshwater Fish 15, 66–77.

Cheng, Y. W. & Gallinat, M. P. (2004). Statistical analysis of the relationship among environmental variables, inter-annual variability and smolt trap efficiency of salmonids in the Tucannon River. Fisheries Research 70, 229–238.

Ciannelli, L., Chan, K. S., Bailey, K. M. & Stenseth, N. C. (2004). Nonadditive effects of the environment on the survival of a large marine fish population. Ecology 85, 3418–3427.

Claireaux, G. & Lagardere, J. P. (1999). Influence of temperature, oxygen and salinity on the metabolism of the European sea bass. Journal of Sea Research 42, 157–168.

Claramunt, R. M. & Wahl, D. H. (2000). The effects of abiotic and biotic factors in determining larval fish growth rates: a comparison across species and reservoirs. Transactions of the American Fisheries Society 129, 835–851.

Clark, T. D., Sandblom, E., Hinch, S. J., Patterson, D. A., Frappell, P. B. & Farrell, A. P. (2010). Simultaneous biologging of heart rates and acceleration, and their relationships with energy expenditure in free-swimming sockeye salmon (Oncorhynchus nerka). Journal of Comparative Physiology B 180, 673–684.

Cooke, S. J., Hinch, S. G., Farrell, A. P., Lapointe, M., Jones, S. R. M., Macdonald, J. S., Patterson, D. A., Healey, M. C. & Van Der Kraak, G. (2004**a**). Abnormal migration timing and high en route mortality of sockeye salmon in the Fraser River, British Columbia. Fisheries 29, 22–33.

Cooke, S. J., Hinch, S. G., Wikelski, M., Andrews, R. D., Kuchel, L. J., Wolcott, T. G. & Butler, P. J. (2004**b**). Biotelemetry: a mechanistic approach to ecology. Trends in Ecology and Evolution 19, 334–343.

Cooke, S. J., Crossin, G. T., Patterson, D. A., English, K., Hinch, S. G., Young, J. L., Alexander, R., Healey, M. C. & Farrell, A. P. (2005). Coupling non-invasive physiological and energetic assessments with telemetry to understand inter-individual variation in behaviour and survivorship of fish: development and validation of a technique. Journal of Fish Biology 67, 1342–1358.

Cooke, S. J., Hinch, S. G., Crossin, G. T., Patterson, D. A., English, K. K., Shrimpton, J. M., Van Der Kraak, G. & Farrell, A. P. (2006**a**). Physiology of individual late-run Fraser River sockeye salmon (Oncorhynchus nerka) sampled in the ocean correlates with fate during spawning migration. Canadian Journal of Fisheries and Aquatic Sciences 63, 1469–1480.

Cooke, S. J., Hinch, S. G., Crossin, G. T., Patterson, D. A., English, K. K., Healey, C. G., Shrimpton, J. M., Van Der Kraak, G. & Farrell, A. P. (2006**b**). Mechanistic basis of individual mortality in Pacific salmon during spawning migrations. Ecology 87, 1575–1586.

Cooke, S. J., Hinch, S. G., Crossin, G. T., Patterson, D. A., English, K. K., Healey, M. C., Macdonald, J. S., Shrimpton, J. M., Young, J. L., Lister, A., Van Der Kraak, G. & Farrell, A. P. (2008). Physiological correlates of coastal arrival and river entry timing in late summer Fraser River sockeye salmon (Oncorhynchus nerka). Behavioral Ecology 19, 747–758.

Cooke, S. J., Donaldson, M. R., Hinch, S. G., Crossin, G. T., Patterson, D. A., Hanson, K. C., English, K. K., Shrimpton, J. M. & Farrell, A. P. (2009). Is fishing selective for physiological and energetic characteristics in migratory adult sockeye salmon? Evolutionary Applications 2, 299–311.

Cooke S. J., Hinch S. G., Lucas M. C. & Lutcavage M. (2012). Biotelemetry and biologging techniques for fishes. In Fisheries Techniques, 3rd edn (Zale A. & Sutton T., eds). Bethesda, MD: American Fisheries Society.

Cooperman, M. S., Hinch, S. G., Crossin, G. T., Cooke, S. J., Patterson, D. A., Olsson, I., Lotto, A. G., Welch, D. W., Shrimpton, J. M., Van Der Kraak, G. & Farrell, A. P. (2010). Effects of experimental manipulations of salinity and maturation status on the physiological condition and mortality of homing adult sockeye salmon held in a laboratory. Physiological and Biochemical Zoology 83, 459–472.

Crain, C. M., Kroeker, K. & Halpern, B. S. (2008). Interactive and cumulative effects of multiple human stressors in marine systems. Ecology Letters 11, 1304–1315. Crossin, G. T., Hinch, S. G., Farrell, A. P., Higgs, D. A., Lotto, A. G., Oakes, J. D. & Healey, M. C. (2004**a**). Energetics and morphology of sockeye salmon: effects of upriver migratory distance and elevation. Journal of Fish Biology 65, 788–810.

Crossin, G. T., Hinch, S. G., Farrell, A. P., Higgs, D. A. & Healey, M. C. (2004**b**). Somatic energy of sockeye salmon at the onset of upriver migration: a comparison among ocean climate regimes. Fisheries Oceanography 29, 22–33.

Crossin, G. T., Hinch, S. G., Cooke, S. J.,Welch, D. W., Batten, S. D., Patterson, D. A., Van Der Kraak, G., Shrimpton, J. M. & Farrell, A. P. (2007). Behaviour and physiology of sockeye salmon homing through coastal waters to a natal river. Marine Biology 152, 905–918.

Crossin, G. T., Hinch, S. G., Cooke, S. J., Welch, D. W., Patterson, D. A., Jones, S. R. M., Lotto, A. G., Leggatt, R. A., Mathes, M. T., Shrimpton, J. M., Van Der Kraak, G. & Farrell, A. P. (2008). Exposure to high temperature influences the behaviour, physiology, and survival of sockeye salmon during spawning migration. Canadian Journal of Zoology 86, 127–140.

Crossin, G. T., Hinch, S. G., Cooke, S. J., Cooperman, M. S., Patterson, D. A., Welch, D.W., Hanson, K. C., Olsson, I., English, K. K. & Farrell, A. P. (2009**a**). Mechanisms influencing the timing and success of reproductive migration in a capital breeding semelparous fish species, the sockeye salmon. Physiological and Biochemical Zoology 82, 635–652.

Crossin, G. T., Hinch, S. G., Welch, D. W., Cooke, S. J., Patterson, D. A., Hills, J. A.,Zohar,Y., Klenke, U., Jacobs, M. C., Pon, L. B., Winchell, P. M. & Farrell, A. P. (2009**b**). Physiological profiles of sockeye salmon in the Northeastern Pacific Ocean and the effects of exogenous GnRH and testosterone on rates of homeward migration. Marine and Freshwater Behavior and Physiology 42, 89–108.

Crozier, L. G., Zabel, R. W., Hockersmith, E. E. & Achord, S. (2010). Interacting effects of density and temperature on body size in multiple populations of Chinook salmon. Journal of Animal Ecology 79, 342–349.

Cury, P., Roy, C. & Faure, V. (1998). Environmental constraints and pelagic fisheries in upwelling areas: the Peruvian puzzle. South African Journal of Marine Science 19,159–167.

De Raedemaecker, F., Miliou, A. & Perkins, R. (2010). Fish community structure on littoral rocky shores in the Eastern Aegean Sea: effects of exposure and substratum. Estuarine Coastal and Shelf Science 90, 35–44.

De Zwart, D., Dyer, S. D., Posthuma, L. & Hawkins, C. P. (2006). Predictive models attribute effects on fish assemblages to toxicity and habitat alteration. Ecological Applications 16, 1295–1310.

De'ath, G. (2007). Boosted trees for ecological modelling and prediction. Ecology 88, 243–251.

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

Del Toro-Silva, F. M., Miller, J. M., Taylor, J. C. & Ellis, T. A. (2008). Influence of oxygen and temperature on growth and metabolic performance of Paralichthys lethostigma (Pleuronectiformes : Paralichthyidae). Journal of Experimental Marine Biology and Ecology 358, 113–123.

Diaz, R. J., Cutter, G. R. & Able, K. W. (2003). The importance of physical and biogenic structure to juvenile fishes on the shallow inner continental shelf. Estuaries 26, 12–20.

Dingsor, G. E., Ciannelli, L., Chan, K. S., Ottersen, G. & Stenseth, N. C. (2007). Density dependence and density independence during the early life stages of four marine fish stocks. Ecology 88, 625–634.

Donaldson, M. R., Cooke, S. J., Patterson, D. A., Hinch, S. G., Robichaud, D., Hanson, K. C., Olsson, I., Crossin, G. T., English, K. K. & Farrell, A. P. (2009). Limited behavioural thermoregulation by adult upriver-migrating sockeye salmon (Oncorhynchus nerka) in the lower Fraser River, British Columbia. Canadian Journal of Zoology 87, 480–490.

Donaldson, M. R., Hinch, S. G., Patterson, D. A., Farrell, A. P., Shrimpton, J. M., Miller-Saunders, K. M., Robichaud, D., Hills, J., Hruska, K. A., Hanson, K. C., English, K. K., Van Der Kraak, G. & Cooke, S. J. (2010). Physiological condition differentially affects the behavior and survival of two populations of sockeye salmon during their freshwater spawning migration. Physiological and Biochemical Zoology 83, 446–458.

Donaldson, M. R., Hinch, S. G., Patterson, D. A., Hills, J., Thomas, J. O., Cooke, S. J., Raby, G. D., Thompson, L. A., Robichaud, D., English, K. K. & Farrell, A. P. (2011). The consequences of angling, beach seining, and confinement on the physiology, postrelease behaviour and survival of adult sockeye salmon during upriver migration. Fisheries Research 108, 133–141.

Dunham, J. B., Cade, B. S. & Terrell, J. W. (2002). Influences of spatial and temporal variation on fish-habitat relationships defined by regression quantiles. Transactions of the American Fisheries Society 131, 86–98.

Eliason, E. J., Clark, T. D., Hague, M. J., Hanson, L. M., Gallagher, Z. S., Jeffries, K. M., Gale, M. K., Patterson, D. A., Hinch, S. G. & Farrell, A. P. (2011). Differences in thermal tolerance among sockeye salmon populations. Science 332, 109–112.

Elith, J. & Leathwick, J. R. (2007). Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. Diversity and Distributions 13, 265–275.

Elith, J., Leathwick, J. R. & Hastie, T. (2008). A working guide to boosted regression trees. Journal of Animal Ecology 77, 802–813.

Erbe, C. & King, A. R. (2009). Modeling cumulative sound exposure around marine seismic surveys. Journal of the Acoustical Society of America 125, 2443–2451.

Esteves, E., Pina, T. & Andrade, J. P. (2009). Diel and seasonal changes in nutritional condition of the anadromous twaite shad Alosa fallax fallax (LacEpede, 1803) larvae. Ecology of Freshwater Fish 18, 132–144.

Evans, M. L. & Neff, B. D. (2009). Non-additive genetic effects contribute to larval spinal deformity in two populations of Chinook salmon (Oncorhynchus tshawytscha). Aquaculture 296, 169–169–173.

Everitt, B., Landau, S., Leese, M. & Stahl, D. (2011). Cluster Analysis. Chichester: Wiley. Farrell, A. P., Hinch, S. G., Cooke, S. J., Patterson, D. A., Crossin, G. T., Lapointe, M. & Mathes, M. T. (2008). Pacific salmon in hot water: applying aerobic scope models and biotelemetry to predict the success of spawning migrations. Physiological and Biochemical Zoology 81, 697–708.

Folt, C. L., Chen, C. Y., Moore, M. V. & Burnaford, J. (1999). Synergism and antagonism among multiple stressors. Limnology and Oceanography 44, 864–867.

Foody, G. M. (2004). Spatial nonstationarity and scale-dependency in the relationship between species richness and environmental determinants for the sub-Saharan endemic avifauna. Global Ecology and Biogeography 13, 315–320.

Fotheringham, A. S., Charlton, M. E. & Brunsdon, C. (1998). Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. Environment and Planning A 30, 1905–1927.

Friedman, J. H. (1991). Multivariate adaptive regression splines. Annals of Statistics 19, 1–67.

Froese, R. & Torres, A. (1999). Fishes under threat: an analysis of the fishes in the 1996 IUCN Red List. In Towards Policies for Conservation and Sustainable Use of Aquatic Genetic Resources (Pullin, R. S., Baitley, D. M., Kooiman, K, eds), pp. 31–144. ICLARM: Conference Proceedings 59.

Galvez, F., Franklin, N. M., Tuttle, R. B. & Wood, C. M. (2007). Interactions of waterborne and dietary cadmium on the expression of calcium transporters in the gills of rainbow trout: Influence of dietary calcium supplementation. Aquatic Toxicology 84, 208–214.

Gilhousen, P. (1990). Prespawning mortalities of sockeye salmon in the Fraser River system and possible causal factors. International Pacific Salmon Fisheries Commission Bulletin XXVI.

Ginot, V., Gaba, S., Beaudouin, R., Aries, F. & Monod, H. (2006). Combined use of local and ANOVA-based global sensitivity analyses for the investigation of a stochastic dynamic model: Application to the case study of an individual-based model of a fish population. Ecological Modelling 193, 479–491.

Goncalves, R., Scholze, M., Ferreira, A. M., Martins, M. & Correia, A. D. (2008). The joint effect of polycyclic aromatic hydrocarbons on fish behavior. Environmental Research 108, 205–213.

Grace, J. B. (2008). Structural equation modelling for observational studies. The Journal of Wildlife Management 71, 14–22.

Grossman, G. D., Ratajczak, R. E., Crawford, M. & Freeman, M. C. (1998). Assemblage organization in stream fishes: effects of environmental variation and interspecific interactions. Ecological Monographs 68, 395–420.

Growns, I., Astles, K. & Gehrke, P. (2006). Multiscale spatial and small-scale temporal variation in the composition of riverine fish communities. Environmental Monitoring and Assessment 114, 553–571.

Gutierrez-Estrada, J. C., Yanez, E., Pulido-Calvo, I., Silva, C., Plaza, F. & Borquez, C. (2009). Pacific sardine (Sardinops sagax, Jenyns 1842) landings prediction. A neural network ecosystemic approach. Fisheries Research 100, 116–125.

Hague, M. J., Ferrari, M. R., Miller, J. R., Patterson, D. A., Russell, G. L., Farrell, A. P. & Hinch, S. G. (2011). Modelling the future hydroclimatology of the lower Fraser River and its impacts on the spawning migration survival of sockeye salmon. Global Change Biology 17, 87–98.

Hanson, K. C., Cooke, S. J., Hinch, S. G., Crossin, G. T., Patterson, D. A., English, K. K., Donaldson, M. R., Shrimpton, J. M., Van Der Kraak, G. & Farrell, A. P. (2008). Individual variation in migration speed of upriver-migrating sockeye salmon in the Fraser River in relation to their physiological and energetic status. Physiological and Biochemical Zoology 81, 255–268.

Harrell, F. E., Jr (2001). Regression Modelling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. New York, NY: Springer-Verlag.

Hastie, T. J. & Tibshirani, R. J. (1990). Generalized Additive Models. London: Chapman & Hall.

Hastie, T. J., Tibshirani, R. J. & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd edn. New York, NY: Springer.

Hernandez, F. J., Hare, J. A. & Fey, D. P. (2009). Evaluating diel, ontogenetic and environmental effects on larval fish vertical distribution using Generalized Additive Models for Location, Scale and Shape. Fisheries Oceanography 18, 224–236.

Hill, T. & Lewicki, P. (2007). Statistics: Methods and Applications. Tulsa, OK: Statsoft.

Hinch, S. G. & Bratty, J. M. (2000). Effects of swim speed and activity pattern on success of adult sockeye salmon migration through an area of difficult passage. Transactions of the American Fisheries Society 129, 604–612.

Hinch, S. G. & Rand, P. S. (1998). Swim speeds and energy use of river migrating adult sockeye salmon: role of local environment and fish characteristics. Canadian Journal of Fisheries and Aquatic Sciences 55, 1821–1831.

Hinch, S. G. & Rand, P. S. (2000). Optimal swim speeds and forward assisted propulsion: energy conserving behaviours of up-river migrating salmon. Canadian Journal of Fisheries and Aquatic Sciences 57, 2470–2478.

Hinch, S. G., Diewert, R. E., Lissimore, T. J., Prince, A. M. J., Healey, M. C. & Henderson, M. A. (1996). Use of electromyogram telemetry to assess difficult passage areas for river migrating adult sockeye salmon. Transactions of the American Fisheries Society 125, 253–260.

Hinch, S. G., Cooke, S. J., Healey, M. C. & Farrell, A. P. (2005). Behavioural physiology of fish migrations: salmon as a model approach. In Fish Physiology, Vol. 24 (Sloman, K. A., Wilson, R. W. & Balshine, S., eds), pp. 239–295. San Diego, CA: Academic Press.

Hinz, H., Prieto, V. & Kaiser, M. J. (2009). Trawl disturbance on benthic communities: chronic effects and experimental predictions. Ecological Applications 19, 761–773.

Holbrook, S. J. & Schmitt, R. J. (2003). Spatial and temporal variation in mortality of newly settled damselfish: patterns, causes and co-variation with settlement. Oecologia 135, 532–541.

Hruska, K. A., Hinch, S. G., Healey, M. C., Patterson, D. A., Larsson, S. & Farrell, A. P. (2010). Influences of sex and activity level on physiological changes in individual adult sockeye salmon during rapid senescence. Physiological and Biochemical Zoology 83, 663–676.

Hruska, K. A., Hinch, S. G., Patterson, D. A. & Healey, M. C. (2011). Egg retention in relation to arrival timing and reproductive longevity in female sockeye salmon (Oncorhynchus nerka). Canadian Journal of Fisheries and Aquatic Sciences 68, 250–259.

Huet, S., Bouvier, A., Gruet, M.-A. & Jolivet, E. (1996). Statistical Tools for Nonlinear Regression: A Practical Guide with S-Plus Examples. New York, NY: Springer-Verlag.

IPCC (2007). Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M. & Miller, H. L. , eds) Cambridge: Cambridge University

Jeffries, K. M., Hinch, S. G., Donaldson, M. R., Gale, M. K., Burt, J. M., Thompson, L. A., Farrell, A. P., Patterson, D. A. & Miller, K. M. (2011). Temporal changes in blood variables during final maturation and senescence in male sockeye salmon Oncorhynchus nerka: reduced osmoregulatory ability can predict mortality. Journal of Fish Biology 79, 449–465.

Jeschke, J. M. & Strayer, D. L. (2006). Determinants of vertebrate invasion success in Europe and North America. Global Change Biology 12, 1608–1619.

Jobling, S., Burn, R. W., Thorpe, K., Williams, R. & Tyler, C. (2009). Statistical modeling suggests that antiandrogens in effluents from wastewater treatment works contribute to widespread sexual disruption in fish living in English rivers. Environmental Health Perspectives 117, 797–802.

Jowett, I. G. & Davey, A. J. H. (2007). A comparison of composite habitat suitability indices and generalized additive models of invertebrate abundance and fish presence-habitat availability. Transactions of the American Fisheries Society 136, 428–444.

Kai, E. T. & Marsac, F. (2010). Influence of mesoscale eddies on spatial structuring of top predators' communities in the Mozambique Channel. Progress in Oceanography 86, 214–223.

Kaplan, D. T. (2009). Statistical Modeling: A Fresh Approach. Seattle, WA: CreateSpace. Katara, I., Pierce, G. J., Illian, J. & Scott, B. E. (2011). Environmental drivers of the anchovy/ sardine complex in the Eastern Mediterranean. Hydrobiologia 670, 49–65.

Kelly, B. C., Ikonomou, M. G., MacPherson, N., Sampson, T., Patterson, D. A. & Dubetz, C. (2011). Tissue residue concentrations of organohalogens and trace elements in adult Pacific salmon returning to the Fraser River, British Columbia, Canada. Environmental Toxicology and Chemistry 30, 367–376.

Knudby, A., Brenning, A. & LeDrew, E. (2010). New approaches to modelling fish-habitat relationships. Ecological Modelling 221, 503–511.

Koenker, R. & Bassett, G. (1978). Regression Quantiles. Econometrica 46, 33–50.

Koenker, R. & Hallock, K. F. (2001). Quantile regression. Journal of Economic Perspectives 15, 143–156.

Kranstauber, B., Cameron, A., Weinzerl, R., Fountain, T., Tilak, S., Wikelski, M. & Kays, R. (2011). The movebank data model for animal tracking. Environmental Modelling and Software 26, 834–835.

Kupschus, S. (2003). Development and evaluation of statistical habitat suitability models: an example based on juvenile spotted seatrout Cynoscion nebulosus. Marine Ecology Progress Series 265, 197–212.

Laetz, C., Baldwin, D., Collier, T., Hebert, V., Stark, J. & Scholz, N. (2009). The synergistic toxicity of pesticide mixtures: implications for risk assessment and the conservation of endangered Pacific salmon. Environmental Health Perspectives 117, 348–353.

Leathwick, J. R., Rowe, D., Richardson, J., Elith, J. & Hastie, T. (2005). Using multivariate adaptive regression splines to predict the distributions of New Zealand's freshwater diadromous fish. Freshwater Biology 50, 2034–2052.

Leathwick, J. R., Elith, J., Francis, M. P., Hastie, T. & Taylor, P. (2006**a**). Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. Marine Ecology Progress Series 321, 267–281.

Leathwick, J. R., Elith, J. & Hastie, T. (2006**b**). Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. Ecological Modeling 199, 188–196.

Leathwick, J. R., Elith, J., Chadderton, W. L., Rowe, D. & Hastie, T. (2008). Dispersal, disturbance and the contrasting biogeographies of New Zealand's diadromous and nondiadromous fish species. Journal of Biogeography 35, 1481–1497.

Lee, C. G., Farrell, A. P., Lotto, A., Hinch, S. G. & Healey, M. C. (2003**a**). Excess postexercise oxygen consumption in adult sockeye (Oncorhynchus nerka) and coho (O. kisutch) salmon following critical speed swimming. Journal of Experimental Biology 206, 3253–3260.

Lee, C. G., Farrell, A. P., Lotto, A., MacNutt, M. J., Hinch, S. G. & Healey, M. C. (2003**b**). The effect of temperature on swimming performance and oxygen consumption in adult sockeye (Oncorhynchus nerka) and coho (O. kisutch) salmon stocks. Journal of Experimental Biology 206, 3239–3251.

Lee, Y. W., Megrey, B. A. & Macklin, S. A. (2009). Evaluating the performance of Gulf of Alaska walleye pollock (Theragra chalcogramma) recruitment forecasting models using a Monte Carlo resampling strategy. Canadian Journal of Fisheries and Aquatic Sciences 66, 367–381.

Lindegren, M., Ostman, O. & Gardmark, A. (2011). Interacting trophic forcing and the population dynamics of herring. Ecology 92, 1407–1413.

Locham, G. A., Kaunda-Arara, B. & Mlewa, C. M. (2010). The influence of reef type and seasonality on population structure of coral-reef fishes within Malindi Marine Park, Kenya. Marine Ecology 31, 494–505.

Lorance, P., Pawlowski, L. & Trenkel, V. M. (2010). Standardizing blue ling landings per unit effort from industry haul-by-haul data using generalized additive models. Ices Journal of Marine Science 67, 1650–1658.

Lucas, M. C. & Baras, E. (2001). Migration of Freshwater Fishes. Oxford: Blackwell Science.

Macdonald, J. S. (2000). Mortality during the migration of Fraser River sockeye salmon (Oncorhynchus nerka): a study of the effect of ocean and river environmental conditions in 1997. Canadian Technical Report of Fisheries and Aquatic Sciences 2315, 1–120.

Macdonald, J. S., Foreman, M. G. G., Farrell, T., Williams, I. V., Grout, J., Cass, A., Woodey, J. C., Enzenhofer, H., Clarke, W. C., Houtman, R., Donaldson, E. M. & Barnes, D. (2000). The influence of extreme water temperatures on migrating Fraser River sockeye salmon (Oncorhynchus nerka) during the 1998 spawning season. Canadian Technical Report of Fisheries and Aquatic Sciences 2326, 1–117.

Macdonald, J. S., Morrison, J., Patterson, D. A., Heinonen, J. & Foreman, M. (2007). Examination of factors influencing Nechako River discharge, temperature, and aquatic habitats. Canadian Technical Report of Fisheries and Aquatic Sciences 2773, 1–32.

Macdonald, J. S., Patterson, D. A., Hague, M. J. & Guthrie, I. C. (2010). Modeling the influence of environmental factors on spawning migration mortality for sockeye salmon fisheries management in the Fraser River, British Columbia. Transactions of the American Fisheries Society 139, 768–782.

Maceina, M. (1992). A simple regression model to assess environmental effects on fish growth. Journal of Fish Biology 41, 557–565.

MacNutt, M. J., Hinch, S. G., Lee, C. G., Phibbs, J. R., Lotto, A. G., Healey, M. C. & Farrell, A. P. (2006). Temperature effects on swimming performance, energetics, and aerobic capacities of mature adult pink salmon (Oncorhynchus gorbuscha) compared with those of sockeye salmon (Oncorhynchus nerka). Canadian Journal Zoology 84, 88–97.

Magaud, H., Migeon, B., Morfin, P., Garric, J. & Vindimian, E. (1997). Modelling fish mortality due to urban storm run-off: interacting effects of hypoxia and un-ionized ammonia. Water Research 31, 211–218.

Magnan, P., Rodriguez, M. A., Legendre, P. & Lacasse, S. (1994). Dietary variation in a fresh-water fish species-relative contributions of biotic interactions, abiotic factors and spatial structure. Canadian Journal of Fisheries and Aquatic Sciences 51, 2856–2865.

Mairesse, G., Thomas, M., Gardeur, J. N. & Brun-Bellut, J. (2007). Effects of dietary factors, stocking biomass and domestication on the nutritional and technological quality of the Eurasian perch Perca fluviatilis. Aquaculture 262, 86–94.

Malins, D. C., Stegeman, J. J., Anderson, J. W., Johnson, P. M., Gold, J. & Anderson, K. M. (2004). Structural changes in gill DNA reveal the effects of contaminants on Puget sound fish. Environmental Health Perspectives 112, 511–515.

Marr, J. C. A., Hansen, J. A., Meyer, J. S., Cacela, D., Podrabsky, T., Lipton, J. & Bergman, H. L. (1998). Toxicity of cobalt and copper to rainbow trout: application of a mechanistic model for predicting survival. Aquatic Toxicology 43, 225–238.

Martins, E. G., Hinch, S. G., Patterson, D. A., Hague, M. J., Cooke, S. J., Miller, K. M., Lapointe, M. F., English, K. K. & Farrell, A. P. (2011). Effects of river temperature and climate warming on stock-specific survival of adult migrating Fraser River sockeye salmon (Oncorhynchus nerka). Global Change Biology 17, 99–114.

Martins, E. G., Hinch, S. G., Patterson, D. A., Hague, M. J., Cooke, S. J., Miller, K. M., Robichaud, D., English, K. K. & Farrell, A. P. (2012). High river temperature reduces survival of sockeye salmon approaching spawning grounds and exacerbates female mortality. Canadian Journal of Fisheries and Aquatic Sciences 69, 330–342.

Mathes, M. T., Hinch, S. G., Cooke, S. J., Crossin, G. T., Patterson, D. A., Lotto, A. G. & Farrell, A. P. (2010). Effect of water temperature, timing, physiological condition, and lake thermal refugia on migrating adult Weaver Creek sockeye salmon (Oncorhynchus nerka). Canadian Journal of Fisheries and Aquatic Sciences 67, 70–84.

Mazeaud, M. M., Mazeaud, F. & Donaldson, E. M. (1977). Primary and secondary effects of stress in fish: some new data with a general review. Transactions of the American Fisheries Society 106, 201–212.

McCullagh, P. & Nelder, J. (1989). Generalized Linear Models, 2nd edn. Boca Raton, FL: CRC Press.

McDowall, R. M. (1999). Different kinds of diadromy: different kinds of conservation problems. ICES Journal of Marine Science 56, 410–413.

Megrey, B. A., Lee, Y. W. & Macklin, S. A. (2005). Comparative analysis of statistical tools to identify recruitment-environment relationships and forecast recruitment strength. Ices Journal of Marine Science 62, 1256–1269.

Miller, K. M., Li, S., Kaukinen, K. H., Ginther, N., Hammill, E., Curtis, J. M. R., Patterson, D. A., Sierocinski, T., Donnison, L., Pavlidis, P., Hinch, S. G., Hruska, K. A.,Cooke, S. J., English, K. K. & Farrell, A. P. (2011). Genomic signatures predict migration and spawning failure in wild Canadian salmon. Science 331, 214–217.

Moguerza, J. M. & Munoz, A. (2006). Support vector machines with applications. Statistical Science 21, 322–336.

Moisen, G. G., Freeman, E. A., Blackard, J. A., Frescino, T. S., Zimmermann, N. E. & Edwards, T. C. (2006). Predicting tree species presence and basal area in Utah: A comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. Ecological Modelling 199, 176–187.

Nadeau, P. S., Hinch, S. G., Hruska, K. A., Pon, L. B. & Patterson, D. A. (2010). The effects of experimental energy depletion on the physiological condition and survival of adult sockeye salmon (Oncorhynchus nerka) during spawning migration. Environmental Biology of Fishes 88, 241–251.

Olden, J. D. & Jackson, D. A. (2001). Fish-habitat relationships in lakes: Gaining predictive and explanatory insight by using artificial neural networks. Transactions of the American Fisheries Society 130, 878–897.

Ostergren, J., Lundqvist, H. & Nilsson, J. (2011). High variability in spawning migration of sea trout, Salmo trutta, in two northern Swedish rivers. Fisheries Management and Ecology 18, 72–82.

Palialexis, A., Georgakarakos, S., Karakassis, I., Lika, K. & Valavanis, V. D. (2011). Fish distribution predictions from different points of view: comparing associative neural networks, geostatistics and regression models. Hydrobiologia 670, 165–188.

Palmores, P., Ferreras, P., Travaini, A. & Delibes, M. (1998). Co-existence between Iberian Lynx and Egyptian mongooses: estimating interaction strength by structural equation modeling and testing by an observational study. Ecology 67, 967–978.

Patterson, D. A., Macdonald, J. S., Hinch, S. G., Healey, M. C. & Farrell, A. P. (2004). The effect of exercise and captivity on energy partitioning, reproductive maturation and fertilization success in adult sockeye salmon. Journal of Fish Biology 64, 1039–1059.

Planque, B. & Buffaz, L. (2008). Quantile regression models for fish recruitment environment relationships: four case studies. Marine Ecology Progress Series 357, 213–223.

Pon, L. B., Hinch, S. G., Cooke, S. J., Patterson, D. A. & Farrell, A. P. (2009**a**). A comparison of the physiological condition, and fishway passage time and success of migrant adult sockeye salmon at Seton River Dam, British Columbia, under three operational water discharge rates. North American Journal of Fisheries Management 29, 1195–1205.

Pon, L. B., Hinch, S. G., Cooke, S. J., Patterson, D. A. & Farrell, A. P. (2009**b**). Physiological, energetic and behavioural correlates of successful fishway passage of adult sockeye salmon Oncorhynchus nerka in the Seton River. Journal of Fish Biology 74, 1323–1336.

Prasad, A. M., Iverson, L. R. & Liaw, A. (2006). Tree techniques: bagging and random forests for ecological prediction. Ecosystems 9, 181–199.

Pyle, G. G., Rajotte, J. W. & Couture, P. (2005). Effects of industrial metals on wild fish populations along a metal contamination gradient. Ecotoxicology and Environmental Safety 61, 287–312.

Raffenberg, M. J. & Parrish, D. L. (2003). Interactions of Atlantic salmon (Salmo salar) and trout (Salvelinus fontinalis and Oncorhynchus mykiss) in Vermont tributaries of the Connecticut River. Canadian Journal of Fisheries and Aquatic Sciences 60, 279–285.

Rand, P. S., Hinch, S. G., Morrison, J., Foreman, M. G. C., MacNutt, M. J., Macdonald, J. S., Healey, M. C., Farrell, A. P. & Higgs, D. A. (2006). Effect of river discharge, temperature, and future climates on energetics and mortality of adult migrating Fraser River sockeye salmon. Transactions of the American Fisheries Society 135, 655–667.

Reed, T. E., Schindler, D. E., Hague, M. J., Patterson, D. A., Meir, E., Waples, R. S. & Hinch, S. G. (2011). Time to evolve? Potential evolutionary responses of Fraser River sockeye salmon to climate change and effects on persistence. Public Library of Science One 6, 1–13.

Reilly, C. A., Echeverria, T. W. & Ralston, S. (1992). Interannual variation and overlap in the diets of pelagic juvenile rockfish (genus, Sebastes) off central California. Fishery

Riginos, C. & Nachman, M. W. (2001). Population subdivision in marine environments: the contributions of biogeography, geographical distance and discontinuous habitat to genetic differentiation in a blennioid fish, Axoclinus nigricaudus. Molecular Ecology 10, 1439–1453.

Roberts, M. J. & Russo, R. (1999). A Student's Guide to Analysis of Variance. London: Routledge.

Rodd, F. H. & Reznick, D. N. (1997). Variation in the demography of guppy populations: the importance of predation and life histories. Ecology 78, 405–418.

Romesburg, C. (2004). Cluster Analysis for Researchers. Raleigh, NC: Lulu Press.

Rorvik, K. A., Dehli, A., Thomassen, M., Ruyter, B., Steien, S. H. & Salte, R. (2003). Synergistic effects of dietary iron and omega-3 fatty acid levels on survival of farmed Atlantic salmon, Salmo salar L., during natural outbreaks of furunculosis and cold water vibriosis. Journal of Fish Diseases 26, 477–485.

Roscoe, D. W., Hinch, S. G., Cooke, S. J. & Patterson, D. A. (2010). Behaviour and thermal experience of adult sockeye salmon migrating through stratified lakes near spawning grounds: the roles of reproductive and energetic states. Ecology of Freshwater Fish 19, 51–62.

Roscoe, D. W., Hinch, S. G., Cooke, S. J. & Patterson, D. A. (2011). Fishway passage and post-passage mortality of up-river migrating sockeye salmon in the Seton River, British Columbia. River Research and Applications 27, 693–705.

Ruppert, J. L. W., Fortin, M. J., Rose, G. A. & Devillers, R. (2010). Environmental mediation of Atlantic cod on fish community composition: an application of multivariate regression tree analysis to exploited marine ecosystems. Marine Ecology Progress Series 411, 189–238.

Seenappa, D. & Devaraj, K. V. (1995). Effect of different levels of protein, fat and carbohydrate on gowth, feed-utilization and body carcass composition of fingerlings in catla-catla (ham). Aquaculture 129, 243–249.

Shrimpton, J. M., Patterson, D. A., Richards, J. G., Cooke, S. J., Schulte, P. M., Hinch, S. G. & Farrell, A. P. (2005). Ionoregulatory changes in different populations of maturing sockeye salmon Oncorhynchus nerka during ocean and river migration. The Journal of Experimental Biology 208, 4069–4078.

Slawski, T. M., Veraldi, F. M., Pescitelli, S. M. & Pauers, M. J. (2008). Effects of tributary spatial position, urbanization, and multiple low- head dams on warmwater fish community structure in a Midwestern stream. North American Journal of Fisheries Management 28, 1020–1035.

Smyth. G. K. (2002). Nonlinear regression. In Encyclopedia of Environmetrics, Vol. 3(El-Shaarawi, A. H. & Piegorsch, W. W. eds) pp. 1405–1411. Chichester: John Wiley & Sons, Ltd.

Steele, M. A., Forrester, G. E. & Almany, G. R. (1998). Influences of predators and conspecifics on recruitment of a tropical and a temperate reef fish. Marine Ecology Progress Series 172, 115–125.

Steinhausen, M. F., Sandblom, E., Eliason, E. J., Verhille, C. & Farrell, A. P. (2008). The effect of acute temperature increases on the cardiorespiratory performance of resting and swimming sockeye salmon (Oncorhynchus nerka). Journal Experimental Biology 211, 3915–3926.

Stevens, C. E., Council, T. & Sullivan, M. G. (2010). Influences of human stressors on fish based metrics for assessing river condition in central Alberta. Water Quality Research Journal of Canada 45, 35–46.

Stoner, A., Manderson, J. & Pessutti, J. (2001). Spatially explicit analysis of estuarine habitat for juvenile winter flounder: combining generalized additive models and geographic information systems. Marine Ecology Progress Series 213, 253–271.

Strecker, A. L., Casselman, J. M., Fortin, M. J., Jackson, D. A., Ridgway, M. S., Abrams, P. A. & Shuter, B. J. (2011). A multi-scale comparison of trait linkages to environmental and spatial variables in fish communities across a large freshwater lake. Oecologia 166, 819–831.

Sutton, C. D. (2005) Classification and regression trees, bagging, and boosting. In Handbook of Statistics: Data Mining and Data Visualization (Rao C. R., Wegman, E. J. & Solka, J. L. (eds.) pp. 303–330. Amsterdam: Elsevier

Swartzman, G., Huang, C. H. & Kaluzny, S. (1992). Spatial analysis of Bering Sea groundfish survey data using generalized-additive models. Canadian Journal of Fisheries and Aquatic Sciences 49, 1366–1378.

Theriault, V., Garant, D., Bernatchez, L. & Dodson, J. J. (2007). Heritability of life-history tactics and genetic correlation with body size in a natural population of brook charr (Salvelinus fontinalis). Journal of Evolutionary Biology 20, 2266–2277.

Trabelsi, A., Gardeur, J. N., Teletchea, F. & Fontaine, P. (2011). Effects of 12 factors on burbot Lota lota (L., 1758) weaning performances using fractional factorial design experiment. Aquaculture 316, 104–110.

Tsitsika, E. V. & Maravelias, C. D. (2006). Factors affecting purse seine catches: an observerbased analysis. Mediterranean Marine Science 7, 27–39.

USEPA (2007). Climate change and interacting stressors: implications for coral reef management in American Samoa. Global Change Research Program, National Center for Environmental Assessment, Washington, DC; EPA/600/R-07/069. Available at National Technical Information Service, Springfield, VA, and online at http://www.epa.gov/ncea (last accessed 9 October 2011).

Vadas, R. L. & Orth, D. J. (2001). Formulation of habitat suitability models for stream fish guilds: Do the standard methods work? Transactions of the American Fisheries Society 130, 217–235.

Vignon, M. & Sasal, P. (2010). Multiscale determinants of parasite abundance: a quantitative hierarchical approach for coral reef fishes. International Journal for Parasitology 40, 443–451.

Wagner, G. N., Hinch, S. G., Kuchel, L. J., Lotto, A. G., Jones, S. R. M., Patterson, D. A., Macdonald, J. S., Van Der Kraak, G., Shrimpton, J. M., English, K. K., Larsson, S.,

Wagner, G. N., Kuchel, L. J., Lotto, A. G., Patterson, D. A., Shrimpton, J. M., Hinch, S. G. & Farrell, A. P. (2006). Routine and active metabolic rates of migrating adult wild sockeye salmon (Oncorhynchus nerka Walbaum) in seawater and freshwater. Physiological and Biochemical Zoology 79, 100–108.

Wille, K., McLean, E., Goddard, J. S. & Byatt, J. C. (2002). Dietary lipid level and growth hormone alter growth and body conformation of blue tilapia, Oreochromis aureus. Aquaculture 209, 219–232.

Williams, L. R., Taylor, C. M., Warren, M. L. & Clingenpeel, J. A. (2002). Large-scale effects of timber harvesting on stream systems in the Ouachita Mountains, Arkansas, USA. Environmental Management 29, 76–87.

Yanez, E., Plaza, F., Gutierrez-Estrada, J. C., Rodriguez, N., Barbieri, M., Pulido-Calvo, I. & Borquez, C. (2010). Anchovy (Engraulis ringens) and sardine (Sardinops sagax) abundance forecast off northern Chile: a multivariate ecosystemic neural network approach. Progress in Oceanography 87, 242–250.

Yee, T. W. (2010). VGLMs and VGAMs: an overview for applications in fisheries research. Fisheries Research 101, 116–126.

Young, J. L., Hinch, S. G., Cooke, S. J., Crossin, G. T., Patterson, D. A., Farrell, A. P., Van Der Kraak, G., Lotto, A. G., Lister, A., Healey, M. C. & English, K. K. (2006). Physiological and energetic correlates of en route mortality for abnormally early migrating adult sockeye salmon (Oncorhynchus nerka) in the Thompson River, British Columbia. Canadian Journal of Fisheries and Aquatic Sciences 63, 1067–1077.

Young, M. A., Iampietro, P. J., Kvitek, R. G. & Garza, C. D. (2010). Multivariate bathymetry-derived generalized linear model accurately predicts rockfish distribution on Cordell Bank, California, USA. Marine Ecology Progress Series 415, 247–261.

Zabel, R. W., Scheuerell, M. D., McClure, M. M. & Williams, J. G. (2006). The interplay between climate variability and density dependence in the population viability of Chinook salmon. Conservation Biology 20, 190–200.

Zagaglia, C. R., Lorenzzetti, J. A. & Stech, J. L. (2004). Remote sensing data and longline catches of yellowfin tuna (Thunnus albacares) in the equatorial Atlantic. Remote Sensing of Environment 93, 267–281.

Zar, J. H. (1984). Biostatistical Analysis. 2nd edn. Englewood Cliffs, NJ: Prentice Hall.

Zhang, T. Y., Bailey, K. M. & Chan, K. S. (2010). Recruitment forecast models for walleye pollock Theragra chalcogramma fine-tuned from juvenile survey data, predator abundance and environmental phase shifts. Marine Ecology Progress Series 417, 237–248.

Chapter 3.

A Quantitative Analysis of Cumulative Effects to Fraser River Sockeye Salmon Prespawn Mortality

Introduction

Fraser River sockeye salmon abundances have experienced major declines over the past decade, resulting in fisheries closures, and economic and ecosystem impacts (Christensen & Trites, 2011).One factor which could impact future Fraser River sockeye abundances is prespawn mortality (PSM); the variable proportion of female Fraser River sockeye that die at the spawning grounds before they have completely spawned (Gilhousen, 1990; Macdonald, 2000; Marmorek *et al.,* 2012). PSM affects the total number of sockeye salmon eggs deposited into the rivers and streams. Annual PSM rates (the proportion of undeposited eggs) for the various spawning ground populations that make up the total Fraser River population range between 0 and 90%, and extreme PSM events can have major implications to future recruitment of progeny (Gilhousen, 1990). Furthermore, annual PSM events in some Fraser River populations have become more severe (Hinch & Martins 2011). Therefore, the ability to forecast and mitigate large PSM events would be beneficial not only from a conservation standpoint but also from a harvest management perspective.

The main cause for elevated PSM rates in Fraser River sockeye is high amounts of stress experienced while migrating upriver towards the spawning grounds or at the spawning grounds (Gilhousen, 1990). Sockeye salmon stop feeding before they enter the mouth of the Fraser River and depend on fixed reserves of energy for migration and spawning (Young *et al*., 2006). By the time sockeye salmon reach the spawning grounds, only about 10% of their stored energy remains for the spawning process (Young *et al*., 2006). Any factor that causes physical stress on these salmon (often referred to as “stressors”) is likely to further reduce these energy reserves and result in premature mortality from exhaustion (Gilhousen, 1990).

Multiple stressors experienced throughout migration can sometimes interact with each other to affect PSM. For example, both high water temperature and bacterial infection are thought to cause elevated PSM rates but warm temperatures also provide productive environments for bacterial growth (Schreck *et al.,* 2001; St. Hilaire *et al.*, 2001). Hence sockeye salmon that encounter both stressors are likely to be at much higher risk for early mortality. The combined effect of multiple stressors is often referred to as a cumulative effect (e.g., Crain *et al.,* 2008) or impact, and two or more stressors can combine to produce 1) additive effects –the combined effect is the sum of the individual stressors; 2) synergistic effects –the combined effect is greater than the sum of the individual stressors; and 3) antagonistic effects – the combined effect is less than the sum of the individual stressors (Folt *et al.,* 1999; Crain *et al.,* 2008). Cumulative impacts can also refer to the effect of a single stressor applied repeatedly over time. All definitions presented here are incorporated into this study.

Over the past few decades, a number of studies analyzed the complex relations between stressors (e.g., high temperatures, river discharge, pathogen levels in salmon kidneys) and the survival of migrating sockeye salmon (e.g.,Wagner *et al.,* 2005; Rand *et al.,* 2006; Young *et al.,* 2006; Macdonald *et al.,* 2010; Martins *et al.,* 2011). Although similar processes are thought to affect the survival of salmon that have reached the spawning grounds (Gilhousen, 1990), comparatively few studies have analyzed the relation between stressors and PSM rates. Furthermore, there is currently no tool that can accurately predict Fraser River sockeye salmon PSM rates as a function of these stressors or any other factors. Such a tool would be especially useful for fisheries managers who could use it to forecast high PSM rates and adjust sockeye salmon harvest limits to compensate for the loss in spawning ground egg deposition. Alternatively, this tool could provide estimates of PSM rates in situations where time or budget constraints limit researchers from obtaining visual estimates of PSM rates from spawning grounds.

The first goal of this study is to improve the understanding of how a multitude of factors and stressors relate to PSM rates in Fraser River sockeye salmon. The second is to develop a tool for predicting PSM rates that incorporates the cumulative effects of stressors and is useful for fisheries management purposes. In order to achieve these goals we will 1) identify numerous factors and stressors which are known correlates of sockeye salmon survival during or after migration and briefly describe how they could affect PSM; 2) use Classification and Regression Trees (CART) and Random Forest (RF) statistical techniques to describe and model the relationships between these same factors and PSM rates and 3) create a RF model which uses a subset of these factors as predictor variables to predict PSM rates. The results of this study will be useful not only for fisheries management purposes, but will also advance the current understanding of how a broad list of stressors and other factors are related to PSM rates in Fraser River sockeye salmon populations.

Methods

PSM Data and Study Populations

To ensure that our results were applicable across the Fraser River watershed, we included yearly PSM time series data for 24 major Fraser River spawning ground populations (Figure 3.1) in our analysis. PSM time series varied in length and completeness across all populations. However, for each of the populations included in our analysis, time series consisted of yearly PSM estimates from at least as far back as 1977 until 2008. These 24 populations were the only populations for which suitable time series data were available for every predictor variable used in our analysis.

All PSM data was obtained from the Fisheries and Oceans Canada (DFO) database. We used yearly egg retention estimates for each of the populations as an index of yearly PSM as per Gilhousen (1990). Egg retention estimates are obtained through visual examination of dead female body cavities at each of the spawning grounds. If more than 75% of a female’s eggs are still retained after mortality has occurred, the female is considered to be “unspawned”. A female is considered “50%-spawned” if 25-75% of the eggs remain and “spent” if less than 25% of eggs remain. Samples are taken every 1-3 days and at least 100 carcasses are examined each time (unless fewer are available). Daily estimates of egg retention at each location equal the total percentage of “unspawned” females plus half the percentage of “50%-spawned” females. Yearly egg retention for a population is the average of all daily egg retention estimates for that population (Gilhousen, 1990).

Predictor Variables and their Potential Relation to PSM

Timing of Return River Migration

Timing of the onset of return river migration for Fraser River sockeye populations is fairly consistent from year to year and many spawning ground populations will initiate migration on similar dates with one another (English *et al.,* 2005). As such, fisheries management has categorized each of the spawning ground populations into into 4 run-timing groups: Early Stuart, Early Summer, Summer, and Late (Hague & Patterson 2007) which reflect the general time period in which they enter the Fraser River (English *et al.,* 2005) (Figure 2.1). The timing of entry can affect the severity of environmental stressors encountered. For example, the earliest migrants, Early Stuarts, who typically migrate in June-July, tend to encounter higher discharge levels than their Late-run counterparts who migrate in the fall because of seasonal differences in precipitation and the effects of spring snowmelt (Macdonald *et al.,* 2010). Early Summer and Summer groups who migrate in July or August may experience higher temperatures than Early Stuart or Late Run groups (Patterson *et al.,* 2007; Macdonald, 2010). Furthermore, in some years, such as with the Late-run fish in 1995, fish will initiate migration earlier than usual and spend longer amounts of time at the spawning grounds thereby increasing their exposure to adverse river conditions to which they are not adapted to (Macdonald et al., 2010).

We included a predictor variable for yearly date at which 50% of each spawning ground population is estimated to have initiated return river migration (median entry date) as well as a categorical predictor variable for run-timing group. Median entry dates for each run-timing group were estimated using daily counts measured at the Mission, B.C., Hydroacoustic facility (Hague & Patterson, 2007). We also included year as a predictor variable since factors like ecological regime shifts (Crossin *et al.,* 2004) and climate change (Morrison *et al.,* 2002) may alter environmental conditions experienced by migrants from year to year.

Temperature

Extreme temperature exposure elicits stress responses in sockeye salmon (Macdonald *et al.,* 2000), leading to immuno-suppression and development of disease (Schreck *et al.,* 2001) or parasitic infection (Wagner *et al.,* 2005; Crossin *et al.,* 2008) and can also cause thermal shock and mortality (Servizi & Jensen 1977). Research also suggests that differences in migration behaviour (Hodgson & Quinn 2002; Hyatt *et al.,* 2003; Keefer *et al.,* 2008), swimming performance (Lee *et al.,* 2003; Farrell *et al.,* 2008) and adaptation to temperature effects (Zi & Jensen, 1977; reviewed by Richter & Kolmes, 2005) may create species and population-specific differences in thermal threshold tolerances (Eliason *et al*., 2012) .

To account for the potential effects of high water temperatures experienced by each of the populations en-route, we included variables for average and maximum mean daily river temperature recorded over simulated migration using a boxcar model (Appendix B). We also included the total number of days and average number of consecutive days when a boxcar model records temperature above the run-time-specific temperature threshold tolerances in Macdonald *et al. (*2010): 17.5 ^{o}C for Late; 18.5 ^{o}C for Early Stuart; and 19.5 ^{o}C for Early Summer and Summer.

We hypothesized that since daily activity of spawning salmon is different than migrating salmon, the effects of variables such as temperature on sockeye PSM during both stages should be analyzed separately. In addition, including separate variables for temperature experienced during peak of spawn and migration allowed us to avoid making any assumption about the salmons' behaviour and life expectancy once they arrive at the spawning grounds. We included the effects of temperature experienced on the spawning grounds by including variables representing average and maximum mean daily temperatures during the peak spawn period. The peak spawn period represents the time when the majority of salmon are actively spawning. The starting and ending dates for peak spawn are based on visual estimates taken at the spawning grounds. Temperature logger stations at each of the spawning grounds provided daily mean temperature readings. If starting and ending dates for peak spawn were missing, we used the average of historical starting and ending dates because peak spawn periods are fairly consistent over time (David Patterson, DFO, pers. comm, Nov 14, 2011). Time series are a mixture of thermograph readings (1977-1987) and digital logger station readings (1988-2008). DFO updated their data collection stations from thermographs to digital logger stations in 1988 because digital temperature readings are thought to be more accurate than thermograph readings (Lisa Thompson, DFO, pers. comm, May 15, 2012).

Given the strong relationship between water and air temperature (Cooke *et al.,* 2008; Voss *et al.,* 2008), we also included mean ambient air temperature for the summer (June, July, Aug.) and fall (Sept., Oct., Nov.) in the South BC mountains area as a proxy for water temperature because it is publicly available online. Summer values were used for Early Stuart, Early Summer and Summer migrants while Fall values were used for Late migrants. Time Series for these data were provided by Environment Canada and can be publicly accessed at their website (www.ec.gc.ca).

Finally, we included the difference in the number of days between the first date of arrival at the spawning grounds predicted by our boxcar model and the first day of peak spawn as an index of freshwater residency time at the spawning grounds. Freshwater residency time affects the length of exposure to temperature (and other stressor variables) at the spawning grounds.

Discharge

High river discharge (m^3/s) correlates with high water velocities (m/s) and therefore increases the amount of energy salmon need in order to complete migration (Hinch & Rand, 1998). Extreme discharge events could also delay arrival at the spawning grounds (Macdonald *et al.,* 2000), creating prolonged en-route exposure to other environmental stressors. Low discharges correlate with less available habitat (stream depth and width) for sockeye salmon migration and spawning, leading to possible density-dependant effects or increased exposure to higher temperatures or predators (Isaak *et al.,* 2011; Mantua *et al.,* 2010). Hence discharge levels may provide a good indication of habitat suitability and quantity. We included the maximum and average daily mean river discharge experienced during migration as predictor variables. Data for the en-route river discharge variables were collected using the Boxcar model in a similar fashion to en-route temperature variables (Appendix B).

We included the effects of discharges experienced on the spawning grounds by including variables representing average and maximum mean daily discharge during the peak of the spawning period. Data for daily mean river discharge for all spawning locations included in our analysis (Figure 2.1) is collected using discharge gauging stations run by Environment Canada, and time series of this data were taken from their publicly-available Water Survey of Canada online database (www.wsc.ec.gc.ca). Discharge gauging stations are not present on several spawning grounds (e.g., Gluske, Forfar, and Kynock Creek). In these cases, we used surrogate stations from nearby rivers and streams with similar hydrological characteristics (e.g., same river source) that we felt provided reasonable representations of flow experienced at the spawning ground (Appendix B).

We also included percent departure (Patterson & Hague, 2007) from mean precipitation levels in the Southern BC mountains area for the summer (June, July, August) and fall (Sept., Oct., Nov.) seasons as a proxy for discharge. Percentage departures were computed by subtracting the 1961-1990 average seasonal total precipitation from the actual seasonal total precipitation, then dividing this difference by the 1961-1990 average and multiplying by 100 to get the value in percent. Summer values were used for Early Stuart, Early Summer and Summer migrants; while fall values were used for Late migrants. These time series data were provided by Environment Canada and can be publicly accessed at their website (www.ec.gc.ca).

Water Quality

An overabundance of heavy metals in the environment can interfere with physiological processes like metabolism, development, and fecundity, in addition to causing tissue or cytoplasmic abnormalities and behavioural alterations in fish (Brungs, 1969; Pickering & Gast, 1972; Billinski & Jonas, 1973; McIntyre, 1973; Bengtsson, 1974; Anderson, 1978). Additionally, sockeye salmon cease or alter their migration at lower pH levels, thereby altering their timing of arrival at the spawning grounds (Ikuta et al 2001).

For this study, we used water quality variables that reflect conditions experienced by migrants at Hope, B.C., even though factors affecting levels of pollutants or pH (e.g., stormwater or industrial discharge) differ depending on location throughout the Fraser River watershed, because such data are not available at fine spatial scales. Average monthly readings for pH, concentrations (mg/L) of Cobalt (Co) and Copper (Cu), and water hardness (which relates to concentrations of dissolved minerals) were available in the Environment Canada database. Complete time series of monthly readings were only available from 1991 to 2008. We did not consider using an overall mean of average monthly readings to extend time series back to 1977 because average monthly readings for each of pH, Co, Cu, and hardness were highly variable between years. Each day of the month was assumed to have the same water quality readings as the monthly average, and we included variables for 31-day average of pH, Co, Cu, and hardness centered around the median date of entry recorded at Hope.

Stored Energy

Pacific salmon stop feeding just before they begin their return migration and must rely on stored energy to swim upriver and spawn (Crossin *et al.,* 2004). However, changes in oceanic conditions such as sea surface temperature (SST) and the North Pacific Index (NPI, a surrogate index for winter weather patterns in the subarctic Pacific) may affect the ability of Fraser River sockeye salmon to access potential sources of energy (Crossin *et al.,* 2004). For example, in years of relatively high SST, zooplankton production (an index of prey abundance) tends to be relatively low (Crossin *et al.,* 2004).

We included SST and NPI values, averaged from January to June of each year, as sockeye acquire more than 50% of their final mature mass during this time (Brett, 1983). SST readings were taken at the Langara Island lighthouse station in the Queen Charlotte Islands, BC, as this site is within the general region where Fraser River sockeye spend their last 6 months of ocean residency (Crossin *et al.,* 2004). BC Lighthouse SST data is publicly available from the DFO website (www.pac.dfo-mpo.gc/ca/sci/osap/data/default_e.htm). NPI values are available from the National Center for Atmospheric Research's Climate and Global Dynamics website (www.cgd.ucar.edu/~jhurrell/np.html). Furthermore, since the size of sockeye salmon is related to the amount of energy available for return migration (Rand *et al.,* 2006), we included average standard fork length of the fish measured at each of the spawning grounds as an index of energy supply.

Density

Fraser River Sockeye salmon return on a 4 year cyclical basis, meaning that the largest number of fish (or dominant cycle) returns every 4-years. In years with large numbers of returns, density dependant factors may play a role in premature exhaustion of Fraser River sockeye salmon as high densities force fish into suboptimal sections of a river (Macdonald, 2000) thereby exposing them to higher discharge or temperature (Macdonald, 2010).

We included cycle number (1, 2, 3 or 4) for each year as a categorical variable in addition to the average number of salmon counted in-river at Mission, B.C., over the course of the 31-day period centered around the median date of entry. We also considered density dependant effects at the spawning grounds by including estimates of spawning ground escapement and the yearly proportion of males to females at the spawning grounds. The proportion of males to females might influence PSM rates through reduced mating opportunities. DFO estimates escapement using a variety of methods including fish wheels, aerial surveys, and sonar readings. The proportion of males to females is based on sample estimates taken by observers at the spawning grounds.

Finally, we considered incorporating estimates of pink salmon run sizes since they often co-migrate with sockeye but time series for such data are often incomplete and only extend back to the mid-1990s. Effects of other co-migrating salmon species were considered negligible due to their small population sizes. We included each year's pink salmon cycle number (1 or 2) as a predictor because Fraser River sockeye co-migrate with a large number of Pink salmon on alternating years throughout much of their return migration.

Population Specific Differences

Sockeye salmon populations are organized into different groups known as conservation units (CUs) (Figure 2.1) that reflect their isolation from other groups and their genetic and geographic diversity (DFO, 2005). Similarly, sockeye salmon may be organized by Production Unit (PU) (Figure 2.1), which also represents diversity among sockeye salmon groups, but are primarily used for stock assessment purposes. Diversity among CUs (or PUs) may allow salmon in some CUs (or PUs) to better cope with certain environmental stressors (DFO, 2005). Furthermore, since CUs or PUs often cover large areas and a diverse range of ecosystems, impacts from different environmental stressors may vary in intensity across spawning grounds within a CU or PU. Therefore we included sockeye salmon CU, PU, and spawning ground destination as categorical predictor variables for each of the populations in our analysis.

Physical Habitat Variables

A large number of physical habitat characteristics from different migration routes and spawning grounds were also included in our analysis based on information provided in Nelitz *et al.* (2011). Nelitz *et al.* (2011) organizes information for habitat characteristics by CU so we applied these characteristics to the appropriate spawning ground populations. The majority of data was collected using a combination of readily available GIS data provided by Fisheries and Oceans Canada and expert opinion.

Forestry & Mountain Pine Beetle Activity

Construction of roads and stream crossings for forestry can cause increased sedimentation into rivers and spawning beds and/or block passage through streams (Nelitz *et al.,* 2010). Timber harvesting activities can alter the watershed hydrology and remove shade cover which increases nearby stream temperatures (Nelitz *et al.,* 2010). A 15-year, cumulative total for percentage of total area harvested along migration routes and near the spawning grounds were included. The 15-year total was selected to account for forest regrowth (Nelitz *et al.,* 2010). The time series used for the forestry variable began in 1994; hence, we assumed the totals for 1991,1992 and 1993 were identical to those of 1994 so that we could have time series going back to at least 1991 to coincide with the water quality time series.

Recent outbreaks of Mountain Pine beetle (MPB) in B.C.'s interior also prompted us to include a cumulative total forested area destroyed by mountain pine beetle. The effects of mountain pine beetle on forested area differ from that of forestry activity (standing timber and understory vegetation is retained), however MPB disturbance in timber harvested areas are still thought to impact the hydrology of surrounding ecosystems and could potentially exacerbate issues like shade reduction from reduced canopy (Nelitz *et al.,* 2011). MPB disturbance began in 1999 so we include a time series of cumulative total percent of area disturbed by mountain pine beetle in each year from 1999 to 2008. We considered years before 1999 to have no MPB disturbance.

Large Scale Hydroelectric

Large hydroelectric dams can block or delay migration to spawning areas, affect the quality (e.g., changes to sedimentation or temperatures downstream) and quantity of salmon habitats, increase susceptibility to predators, and cause injury and even mortality to migrating adults that swim through hydro turbines or over spillways (Roos, 1991; Marmulla, 2001). The Bridge/Seton River power project and Alcan's Kemano Project on the Nechako River are the two large-scale hydro facilities located in the Fraser River basin that could impact spawning success of Fraser sockeye salmon populations included in our analysis (Roos, 1991). We included a categorical variable that describes the absence or presence of either the Kemano project or the Bridge/Seton River project along each migratory route.

Urbanization

Urbanization in areas adjacent to sockeye salmon habitat can affect salmon health in three ways (Roseneau & Angelo, 2009). First, road construction along with residential and industrial development can increase the amount of impervious surface and affects patterns of runoff, which then alters timing and magnitude of nearby stream discharge. Second, construction of roads or buildings can lead to stream blockage or reduced habitat quality for salmon. Third, increased amounts of runoff and municipal and industrial effluent can affect water quality of nearby rivers and streams by altering sedimentation, nutrient and contaminant levels (Birtwell *et al.,* 1988; Dorcey & Griggs, 1991).

We included the proportion of total migration and spawning ground area that is currently considered "urban" (i.e., proportion of total area developed for residential, business, and industrial purposes) along with human population density (#persons/km^2) and road density for each area. Only 2008 levels for urban area and road density were available and these levels are likely different than those of the past. Therefore, we assumed that areas with high levels of urbanization and road densities in 1977 still had high levels when compared with other areas in 2008. Hence 2008 levels of urban area and road density were included as an index of which areas were highly urbanized and which were not from 1977 to 2008. Time series for human population densities in each area were available back to 1986. We extended these time series back to 1977 by using the average, yearly rate of population growth from 1986-1991 in each area of question.

Agriculture

Agricultural practices such as livestock grazing and cattle crossings can alter habitat, sedimentation, and hydrology in addition to increasing the amounts nutrients and pollutants (e.g., fertilizers, pesticides and manure) in nearby waterways (reviewed Platts, 1991; Rosenau & Angelo, 2009). We included the total area of both migration and spawning habitat that is currently being used for agricultural purposes. Again, only 2008 levels for agricultural area were available so we included this variable as an index of which areas had high agricultural activity and which did not from 1977 to 2008.

Inlet/Outlet Spawners

Most Fraser River spawning ground streams either flow into lakes (inlet) or flow out from lakes (outlet). Outlet spawners tend to encounter more consistent levels of discharge events at the spawning grounds than inlet/tributary spawners since the lake typically acts as a flow-buffer (Nelitz *et al.,* 2011). Hence we included a categorical variable for outlet or inlet spawner as a measure of vulnerability to flow.

A final note with respect to the predictor variables-many are expensive, impractical to collect, and not useful for predicting PSM rates in time for fisheries managers to adjust harvest limits. Furthermore, many of the variables in our study are likely correlated with one another and may even measure similar phenomenon or events. Hence our approach was to initially model all of the candidate predictors in order to improve understanding of how they relate to PSM and then reduce the number to a more practical subset for prediction using properties of our modeling procedure.

Classification and Regression Tree Modelling

We first chose to model the relationships between the predictor variables and PSM rates using classification and regression trees (CART) because understanding CART is fundamental to understanding random forests (RFs). Although we included CART mainly to assist reader interpretation of the RF procedure, we also used CART to help describe the relations between predictor variables and PSM.

CART is a statistical modeling technique that uses predictor variables to sequentially split groups of response variable data (e.g., PSM rates) into successively smaller subgroups (Figure 3.2) (De'ath & Fabricious, 2000). Each of the splits attempts to minimize the within-subgroup sum of squares (SSQ) while maximizing the between-subgroup SSQ (Maindonald & Braun, 2003). At each of the splits in the tree, the predictor variable which explains the largest proportion of sums of squares (SSQ) is shown (Figure 3.2). CART is well suited to the analysis of large data sets with many correlated variables, non-linear relationships, and complex and unknown interactions (De'ath & Fabricius, 2000; Maindonald & Braun, 2003). CART output is easily interpretable and can help uncover patterns, structure and interactions between multiple predictors–an important trait for our analysis due to the potential for many synergistic relationships (De'ath & Fabricius, 2000). We created the CART model using the 'tree' package version 1.0-31 (Ripley, 2011) available in R statistical software (R Development Core Team, 2011). We used only the predictor variables with time series dating back to 1977 create the CART model.

Random Forest Modeling

The next step was to use RFs to help describe relationships between the predictor variables and PSM and also to create our model for predicting PSM rates. RFs are a relatively new type of modeling technique that share many of the same advantages of CART while having much improved predictive performance (Breiman, 2002). Specifically, RFs build multiple CART models, each constructed using different bootstrapped sub-samples of the data set and random subsets of predictor variables at each split (Breiman, 2002). Data not used for the bootstrapped samples are used to assess model accuracy, essentially a form of cross-validation (Breiman, 2002). Predictions from each of the trees are then averaged to create an overall model prediction (Breiman, 2002). RFs are typically built using hundreds or thousands of CART models, thus interpreting how predictions are created or how variables relate to each other is almost impossible (Prasad *et al.,* 2006). RFs do, however, allow researchers to assess predictor variable importance (VI). VI provides insight on how accurately variables can predict a response (such as PSM rates) and can help researchers decide which variables are useful for model prediction (Breiman, 2002). For this experiment, we measured VI using loss in prediction accuracy (or increases in model mean squared error (MSE)) when data for individual predictors are permuted randomly.

We constructed two sets of RF models, one that uses variables with time series from 1977 to 2008 (hereafter referred to as the 1977 model) and another that uses all available variables with time series from 1991 to 2008 (hereafter referred to as the 1991 model). This was done because time series for water quality and forestry variables were only available as far back as 1991. Both of the models were built using 1000 individual trees using R's 'randomForest' package version 4.6-6 (Liaw & Wiener, 2002). We also included partial dependence (PD) plots to examine the individual effects of the most important variables (as dictated by the VI plots) on PSM. PD plots show the relation (linear or non-linear) between a single predictor variable and a response after averaging over the observed values of all other predictors in the model (Elith *et al.,* 2008).

Last, to create our final model for predicting PSM rates, we used the VI estimates from the VI plots to help eliminate variables which were less important and/or not useful for prediction purposes. We first chose to reduce the model to only the top 9 most important variables as dictated by the VI plots and then further reduced the model by examining the relative importance of the remaining 9 variables. We also took into consideration the effects of variable reduction on model prediction accuracy (measured using total model explained variance and MSE) when deciding whether or not to remove certain variables These steps were taken because many predictors are likely correlated and including all variables would be time consuming and costly.

Results

CART Model

The largest split (which explains the most SSQ) is determined by the destination spawning ground (SPAWNGROUND) (Figure 3.2). Fish migrating to Weaver or Gates systems will experience an average PSM rate of almost 29%, while fish migrating to all other spawning grounds will have an average PSM rate of about 7%. Subsequent major splits on median entry date (TIMEentdate) both suggest that earlier migrants all tend to have higher PSM rates than later migrants. For example, migrants to the Weaver or Gates System whose median entry date into the Fraser River is before the 261st day of the year (Sept. 18th) will experience average PSM of approximately 34%. Whereas those fish whose median entry date is on or after the 261st day will experience an average PSM rate of around 6%. The tree also shows that other spawning ground populations who whose median entry date is before the 213th day (Aug. 1st) experience an average PSM rate of approximately 12%. Fish from these populations who migrate on or after this date will only experience an average PSM rate of approximately 5%. Furthermore, the major split on TEMPmaxspawn suggests that early migrants from Weaver or Gates systems who encounter colder temperatures at the spawning grounds tend to have higher average PSM rates. Splits on all other variables were all relatively unimportant, as they explain very little SSQ.

Random Forest Model

Variable importance plots for the 1977 model indicated the top 3 most important variables (by a relatively large margin) were median entry date (TIMEentdate), destination spawning ground (SPAWNGROUND) and year (TIMEyear) (Figure 3.3). The next six most important variables, in order of decreasing importance, were human population density at the spawning ground (HPOPspwn), PU, CU, average and maximum discharge experienced over the course of migration (DISavgmig and DISmaxmig) and finally the average temperature experienced during migration (TEMPavgmig). The model had mean squared error (MSE) value of 0.00999 and explained approximately 55% of the total variance.

Variable importance plots for the 1991 model indicated that the most important variable was destination spawning ground (SPAWNGROUND) followed closely by CU and PU (Figure 3.3). The model had mean squared error (MSE) value of 0.0111 and explained approximately 55% of the total variance. The 1991 model's variable importance plots indicated that water quality variables (WQcu, WQco, WQph) and forestry variables (FORESTmig, FORESTspwn) were not particularly important for model prediction. Since adding the water quality and forestry variables did not provide any real advantage for predicting PSM we chose to exclude the 1991 model from further analysis.

The PD plots give some insight in to how individual predictor variables relate to PSM. According to the PD plots (Figure 3.4), after the effects of all other variables in the model have been accounted for, migrants who's median entry date into the Fraser River is prior to the 220th day of the year (Aug. 8th) will experience a large spike in average PSM rate. The PD plot for TIMEyear suggests that an event occurred in the year 2008 which caused average PSM rates to increase by approximately 8% when compared with previous years. PSM rates also appear to increase with higher human population densities near spawning grounds. However, careful examination of the PD plot for HPOPspwn reveals a disproportionate number of spawning grounds with low population densities as most of the spawning grounds included in our analyses are in more rural areas. The PD plots for the 3 variables measured by our boxcar model, DISavgmig, DIStempmig, TEMPavgmig all showed populations that experience higher values of each of these variables also experience higher average PSM rates. However, while these increases may appear rather drastic in the plots, they each correspond to only about a 1-2% change in average PSM rate.

Finally, SPAWNGROUND, CU and PU all display similar results. The Gates Creek, Gates Channel and Weaver Creek populations each experience about 5% higher average PSM rates than other spawning grounds. This phenomenon is reflected in the PD plots for CU and PU as the Gates and Weaver populations' CUs (Anderson- ES and Harrison (U/S)-L) and PUs (Gates, Weaver) also have increased PSM rates.

Final Reduced Model for Prediction

The RF model built using only the top 9 most important variables from the 1977 model explained approximately 55% of the variance and had a MSE value of 0.01-nearly the same explained variance and MSE values as the model using all candidate predictors. Similarly, the RF model built using only the top 3 most important variables (TIMEentdate, SPAWNGROUND and TIMEyear) explained 51% of the variance and had an MSE value of 0.011-only a slight decrease in predictive accuracy. Values for TIMEyear and SPAWNGROUND are easily inputted into the model for prediction. The median entry date of a run timing group into the Fraser River can also be assessed in time for adjusting sockeye salmon escapement targets (David Patterson-DFO, pers. comm.). However, the problem with this model is that the values of TIMEyear only range from 1977 to 2008. Since individual classification or regression trees in the RF model are constructed using splitting rules, and there is no data post 2008, the trees cannot create any splits which would represent the relation between post 2008 years and PSM rates. Hence the model will assume that years 2009 onwards will relate to PSM rates in the same way as earlier years and this may be an inappropriate assumption. Replacing TIMEyear with the variable for human population density at the spawning ground (HPOPspwn), the variable with the next highest VI relative to TIMEyear, gives a model that explains approximately 48% of the variance and has an MSE value of approximately 0.11. While the replacement of TIMEyear creates slight loss in model prediction accuracy, human populations densities in areas near spawning grounds are fairly consistent from year to year and are easily predicted from one year to the next. Furthermore, a wide range of values of HPOPspwn (from less than one person/ km^2 to 20 persons/ km^2) (Figure 3.4) were included in our model and therefore HPOPspwn was considered a more useful variable for prediction. Using any two of HPOPspwn, TIMEentdate or SPAWNGROUND drops the amount of model explained variance considerably to less than 39%; hence the final model was selected to include all three as predictor variables.

Discussion

The first goal of our study was to advance the current state of knowledge of how different factors and stressors relate to PSM rates in Fraser River Sockeye salmon. While the main focus of our analysis was not on the CART model, the CART output shares some similarities with the output of our 1977 RF model. For instance, in both RF and CART analyses, the spawning ground and median entry date into the Fraser River were important predictor variables. Furthermore as both the CART output and PD plots imply, early migrants and Weaver and Gates populations all experience higher PSM rates when compared to other populations. Hence, one could surmise our single CART shares similarities with many of the trees created inside the RF model and may give good indication as to how some variables interact with one another to affect PSM.

A major drawback of using CART however, is that at each of the splits in the tree, only the predictor variable that explains the largest proportion of SSQ is shown in the output. Any other predictor variable that could produce similar splits are not shown and their influence on the predictor variables is not assessed. This could explain why TIMEyear, produces a relatively small split in the CART model while the RF model indicates TIMEyear as one of top 3 most important variables. Furthermore it is important to have a good understanding of the biological processes being modelled when interpreting CART results. For example, the CART model indicates that Gates and Weaver populations whose median date of entry into the Fraser River is after the 261st day of the year (Sept. 18th) will experience lower PSM rates. The Gates Creek population, however, does not migrate in September and hence this result is only applicable to the Weaver Creek population.

The results of the 1977 model VI plot showed that year, median entry date in to the Fraser River, and destination spawning ground (TIMEyear, TIMEentdate and SPAWNGROUND respectively) were by far the most important variables in predicting PSM. Each of these variables is a surrogate for a number of other variables. For example, the timing of entry into the river affects the type and level of environmental stressors experienced during return migration. Similarly, the spawning ground population (along with PU and CU) is related to timing of entry as well and geographic, environmental and genetic differences between fish. Year may be a proxy for ecological regime shifts and changing climate conditions over time. Therefore perhaps the reason why TIMEyear, TIMEentdate and SPAWNGROUND are so important is that they are doing a good job of representing the overall combined effects of several processes affecting PSM. Alternatively, these variables could just be an excellent surrogate for a single, highly influential process or factor affecting PSM.

The PD plots of TIMEentrydate and SPAWNGROUND show that earlier migrants and Weaver and Gates populations have higher average PSM rates even after the effects of all other predictor variables used in our model were accounted for. Similarly, there seems to be some factor causing PSM in the year 2008 to be much higher than in previous years. Therefore, these 3 PD plots suggest that we have not accounted for all of the processes or factors affecting PSM or alternatively, some of the variables we did include in the model may not accurately represent the phenomenon for which they were intended. When the PD plot of a predictor variable is “flat” or resembles a horizontal line, then the variable in question has no individual effect on the response after accounting for the effects of all other predictors in the model (De’ath, 2007). Our PD plots are not flat, meaning that even after accounting for the effects of all other predictor variables in our model, TIMEyear, TIMEentdate and SPAWNGROUND are still influencing PSM rates. Since TIMEyear, TIMEentdate and SPAWNGROUND are surrogate variables representing other processes affecting PSM and cannot themselves affect PSM rate, we can infer that some processes relating to PSM are still missing from the model.

Our 1977 model VI plot also indicated that migration temperature and discharge variables were relatively important in predicting PSM. This was not surprising considering these variables are two of the most often studied in literature as potential causes of early mortality in Fraser River salmonids (Hinch & Martins, 2011, Johnson *et al.,* 2012). Contrary to our expectations, the PD plots for these variables suggested that the individual contributions of migration temperature and discharge to PSM rates were not substantial. These PD plots, however, do not give any information on the synergistic or interaction effects between variables (Freidman, 2002). Since these variables were some of the most important ones seen in the VI plots, one can surmise that their effect on PSM is largely due to an interaction or synergistic effect with some other variable. For example, if warmer migration temperatures had a major influence on PSM rates in only a few spawning ground populations or only when migration discharge levels were high, then this variable might be considered very important in determining model splits. However, none of these interactions or synergies would appear in the PD plots since the PD plots only show the relation between temperature and PSM after averaging over all spawning ground populations and all levels of average migration discharge. In other words, because of the averaging, the PD plots show very little effect of migration temperature or discharge on PSM.

Although our final model was based on the 1977 model, it should be noted that the1991 model VI plot differed slightly from the 1977 VI plot. One major difference was the drastic increase in importance of TIMEyear from the 1991 model to the 1977 model. This may be explained by a change in the variation of PSM rates in years prior to 1991. Similar phenomenon may explain between-model differences in the VI estimates of other variables as well. Another interesting result was that, contrary to our expectations, the 1991 model VI plots showed water quality variables and several habitat variables (e.g., forestry variables, MPB influence) were not important predictors of PSM. However, this does not necessarily mean that the actual processes for which these variables measure have little effect on PSM. Time series for these variables may contain more measurement error than others leading to lower VI scores. These variables may also have low VI scores because they are less accurate representations of conditions experienced by salmon during return migration (e.g., average monthly water quality indices measured at Hope compared to boxcar measurements of migration temperature). Furthermore, some variables may also be considered unimportant by the RF model if the range of variation in their data was too small to capture any effects to PSM. If, for example, pH levels above 8.5 at Hope had a major influence on sockeye PSM rates but the data used to build the RF model only contained pH readings from 7.5 to 8, then pH would appear unimportant in the VI plots.

The final predictive model built in this study can be used for a number of fisheries management purposes. First, the model can estimate in-season PSM rates which could then be used to set escapement targets for each of the populations included in our analyses. In a given year, if a manager has an estimate of the median date in which a run timing group began migration, they can input that date, along with human population density at each of the spawning grounds to get an estimate of PSM rates for each spawning ground in that run-timing group. For example, if a manager believes that the median date of entry for Late run sockeye is the 220th day of the year and the approximate human population density near Weaver Creek is 20 persons/km2, then the model would predict an average PSM rate of approximately 57% (±11%.) for Weaver Creek sockeye salmon. Similar estimates could be obtained for other Late run populations and managers could then choose to allow for extra escapement to the Late run spawning grounds to compensate for the loss of egg deposition. Managers could use this process along with median entry dates from previous years to set preliminary escapement targets at the beginning of the season. These targets could then be updated once the actual median run date is available.

The final model could also assist with estimating spawner-recruit relationships and productivity when observed PSM data are missing. Observer estimates of PSM are taken from spawning grounds located throughout the Fraser River watershed, and access to each of these spawning grounds may be limited due to time or budgetary constraints. A fisheries scientist could use our model to provide a replacement PSM estimate when observer estimates are unfeasible. These PSM estimates could then be used to better estimate spawner-recruit relationships, productivity, and trends in PSM rates over time at spawning grounds that could not be sampled.

Fisheries managers or scientists could use our model (or similar models) to answer specific questions about the sensitivity of Fraser River sockeye salmon PSM rates to changes in individual predictor variables when other variables are held constant. For example, a manager may be interested to know how sensitive Weaver Creek sockeye are to changes in median date of entry. The manager could input a range of different values for median date of entry, while holding human population density at Weaver Creek constant, to get a better idea of how Weaver Creek PSM will change with median entry date. This information could assist managers or scientists in identifying levels or threshold values of variables which could severely impact PSM rates and require mitigation.

Future studies could develop similar models that are better suited to specific fisheries management purposes. For example, at the beginning of this study, we included as many predictor variables as possible in order to better understand their effects on PSM and guide future research. However, in doing so we omitted several major spawning ground populations that did not have adequate time series data for these variables (e.g., the Bowron lakes population). These populations are important for sockeye salmon fisheries management and stock assessment purposes. Researchers could broaden general applicability of our model by examining the relations between a reduced set of predictor variables and a larger number of spawning ground populations. Furthermore, research could investigate the use of PSM rates from earlier run timing groups (i.e. Early Stuart) as predictor variables to help model PSM rates of later run groups. Within a single year, there may be a relationship between PSM rates experienced by earlier migrants and rates experienced by later migrants. While a model of this sort would no longer be able to predict Early Stuart PSM rates, this type of model would be better suited to adjusting harvest limits since the Early Stuart populations are currently not being targeted commercially.

Future studies may also wish to examine using additional or alternative predictor variables. For example, aside from the variables representing SST and NPI, the majority of our variables focus on freshwater conditions experienced during return migration. A number of other environmental or physiological factors at different life stages (e.g., during ocean residency or juvenile stages) may affect PSM rates. These factors and their influence on the survival of specific salmon stocks are poorly understood. This is particularly true during the ocean life phase, which is often considered to be a “black box” since researchers have relatively little understanding of salmon activity and behavior during this time (Griffiths *et al.,* 2010). Therefore, future models might benefit from incorporating additional factors affecting sockeye during earlier life phases. Future models may also benefit from incorporating a more advanced version of the boxcar model (Appendix B) to measure en route temperature and flow variables. In this study we assumed a constant migration rate for each of the salmon run timing groups. This assumption is unrealistic as salmon migration speeds likely change throughout sections of the Fraser River (i.e. slower near Hell’s Gate and faster near the spawning grounds) (Hague *et al.* 2008). Hence researchers using boxcar models to measure en-route conditions may wish to examine how changes to assumptions like this improves their results.

At present, there is little published literature which analyses the effects of different factors to PSM in Fraser River sockeye. Gilhousen (1990) is typically cited in fisheries literature as the most recent example to how multiple factors relate to PSM (e.g., in Hinch & Martins, 2011, Macdonald *et al.,* 2010 and Johnson *et al.,* 2012). Many of the results in our study are consistent with those found in Gilhousen (1990), namely the importance of timing of river entry timing and temperature variables to PSM. However, Gilhousen (1990) only examined relationships between PSM and a few types of predictors (temperature, discharge, abundance and run timing) using linear regression techniques. In addition, Gilhousen (1990) did not include as many populations as presented in this study. The results of our study extend the work in Gilhousen (1990) to a broader scale while using advanced statistical techniques and data that are more up to date.

This study represents a major step forward for predicting and understanding the complex, cumulative relations between multiple factors and PSM rates in Fraser River sockeye salmon. While more work can still be done to help better understand how different factors relate to PSM, our results should help fisheries management mitigate a potentially large contributor to declining Fraser River sockeye salmon abundances. Moreover, we hope the results and modeling processes described in this study will prove useful for anyone studying cumulative effects of multiple stressors to fish and other animals.

Tables

Table 3.1 ***List of all predictor variables included in classification and regression tree and random forest modeling procedures.***

Predictor	Description
SPWNGRND	Salmon spawning ground.
CU	Salmon Conservation Unit.
PU	Salmon Production Unit.
TIMEyear	Year of return migration.
TIMErun	Run timing group for each spawning ground population.
TIMEentdate	Median entry date of for a spawning ground population.
TIMEholding	Total number of days between the estimated median date of arrival on the spawning grounds and the first day of peak spawn.
TEMPavgmig	Average temperature (^oC) recorded by the boxcar model during simulated migration.
TEMPmaxmig	Maximum temperature recorded by the boxcar model during simulated migration.
TEMPthresh	Total number of days for which the boxcar model records temperatures above stock specific thresholds.
TEMPconsec	Average number of consecutive days for which the boxcar model records temperatures above stock specific thresholds.
TEMPavgspwn	Average temperature (^oC) experienced during peak of spawn.
TEMPmaxspwn	Maximum temperature (^oC) experienced during peak of spawn.
TEMPair	Average summer (Early Stuart, Early Summer, Summer) or fall (Late) air temperature in the lower mainland BC.
DISavgmig	Average Fraser River discharge (m^3/s) recorded by the boxcar model over the course of simulated migration.
DISmaxmig	Maximum Fraser River discharge recorded by the boxcar model over the course of simulated migration.
DISavgspwn	Average river discharge experienced during peak of spawn.
DISmaxspwn	Maximum river discharge experienced during peak of spawn.
DISprecip	Deviance from average summer or fall precipitation in the" South Mountain" area of BC.
WQph	Average river pH at Hope over a 31-day period.
WQco	Average Co concentrations (g/ml) at Hope over a 31-day period.
WQcu	Average Cu concentrations (g/ml) at Hope over a 31-day period.
WQhard	Average CaCo3 concentrations (g/ml) at Hope over a 31-day period.

DENSmig	Average abundance of sockeye salmon in the river over a 31-day period.
DENSsockcyc	Sockeye salmon cycle year (1, 2, 3 or 4).
DENSpinkcyc	Pink salmon cycle year (1 or 2).
DENSspwnesc	Total spawning ground escapement abundances.
DENSmperf	Ratio of males to females at the spawning ground.
PHYSlen	Average standard length of all females measured at the spawning ground.
PHYSsst	Average of Jan.- June sea surface temperature recoded at Langara lighthouse.
PHYSnpi	Average North Pacific Index value from Jan. – June.
HPOPmig	Human population density (# persons/km^2) surrounding the migration corridor.
HPOPspwn	Human population density (# of person/km^2) near the spawning ground.
AGRICmig	Percentage of the migration corridor devoted to agricultural activity.
AGRICspwn	Percentage of the spawning ground devoted to agricultural activity.
URBANmig	Percentage of the migration corridor which is considered to be urban.
URBANspwn	Percentage of the spawning ground devoted to agricultural activity.
FORESTmig	Cumulative percentage of forest area harvested along the migration corridor over 15-years.
FORESTspwn	Cumulative percentage of forest area harvested along the spawning grounds over a 15-year period.
MPBmig	Cumulative percentage of area destroyed by mountain pine beetle (MPB) along the migration corridor.
MPBspwn	Cumulative percentage of amount of area destroyed by mountain pine beetle along the migration corridor.
ROADmig	Road density (km of road/km^2) surrounding the migration corridor.
ROADspwn	Road density surrounding the spawning grounds.
HYDRO	Presence or absence of a large hydroelectric facility along the migration route.
INvsOut	Whether or not a spawning population is an inlet or outlet spawner.

Figures

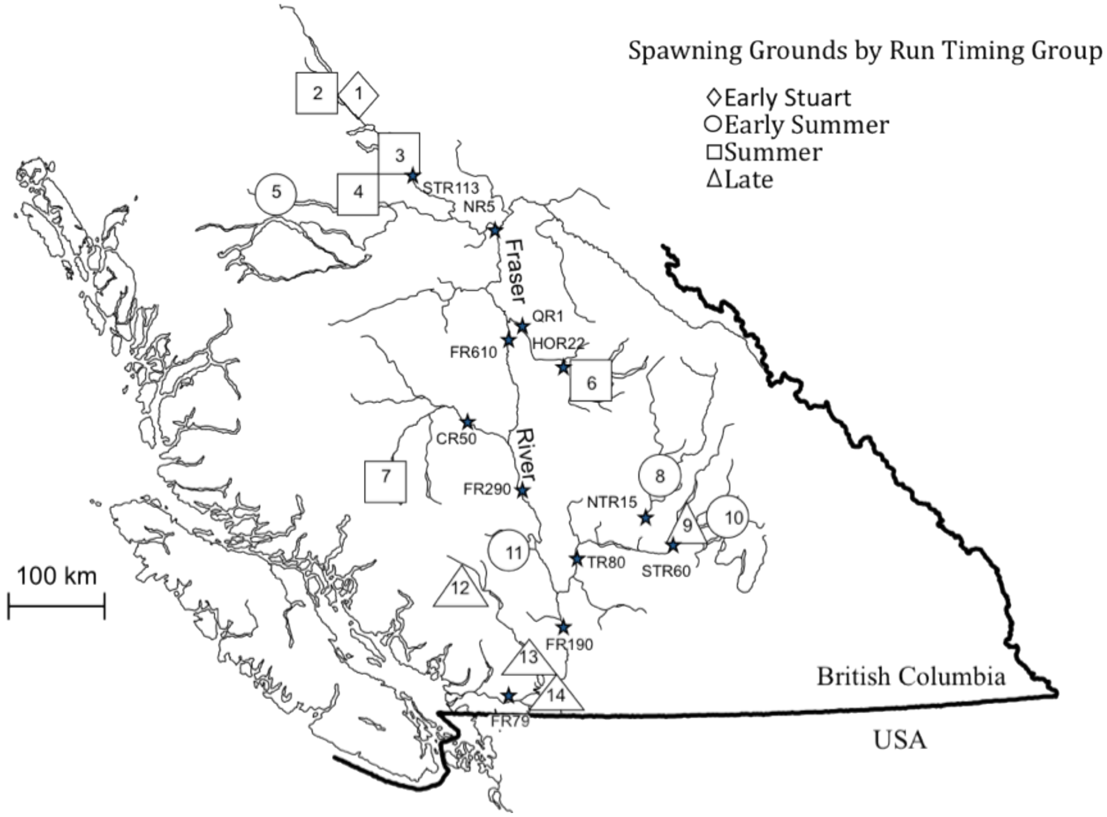

Figure 3.1. *Map of the Fraser River watershed with approximate locations of river temperature logger stations (stars) and Sockeye salmon spawning grounds included in our analysis. Associated names for spawning grounds, as well as sockeye salmon Conservation unit and Production unit (italics) are as follows: 1. Gluske Creek, Forfar Creek, Kynock Creek, Bivouac Creek (Takla/Trembleur-ESTU, Early Stuart); 2. Tachie River (Stuart-S, Late Stuart); 3. Middle River (Takla/Trembleur-S, Late Stuart); 4. Stellako River (Fraser-S, Stellako); 5. Nadina River, Nadina Channel (Francois-ES, Nadina); 6. Horsefly River (Quesnel-S, Horsefly); 7. Chilko River (Chilko, Chilko); 8. Raft River (Kamloops-ES, Raft); 9. Adams River, Lower Shuswap River (Shuswap Complex-L, Lower Shuswap); 10. Scotch Creek, Seymour River (Shuswap Complex-ES, Scotch, Seymour); 11. Gates Channel, Gates Creek (Anderson-ES, Gates); 12. Birkenhead River (Lillooet-L, Birkenhead), Weaver Creek, Weaver Channel (Harrison(U/S)-L, Weaver) 13. Harrison River (Harrison (D/S)-L, Harrison); 14. Cultus Lake (Cultus, Cultus).*

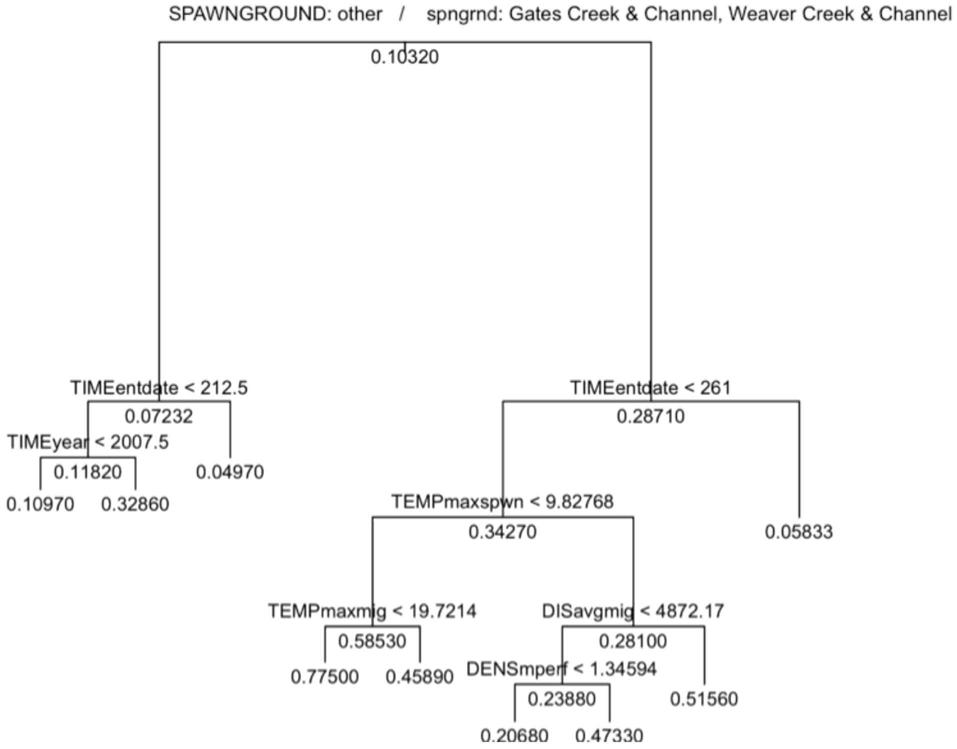

Figure 3.2. Classification and Regression Tree (CART) model output for the 1977-2008 data set. Each of the tree splits is labelled with the variable and its corresponding values that determine the split. The mean value of PSM is displayed at each of the terminal nodes. Length of lines/branches at each split represent the proportion of total sums of squares (SSQ) explained by each split. Variables chosen at each split are those which minimize within group SSQ and maximize between group SSQ. Variable descriptions are given in Table 2.1.

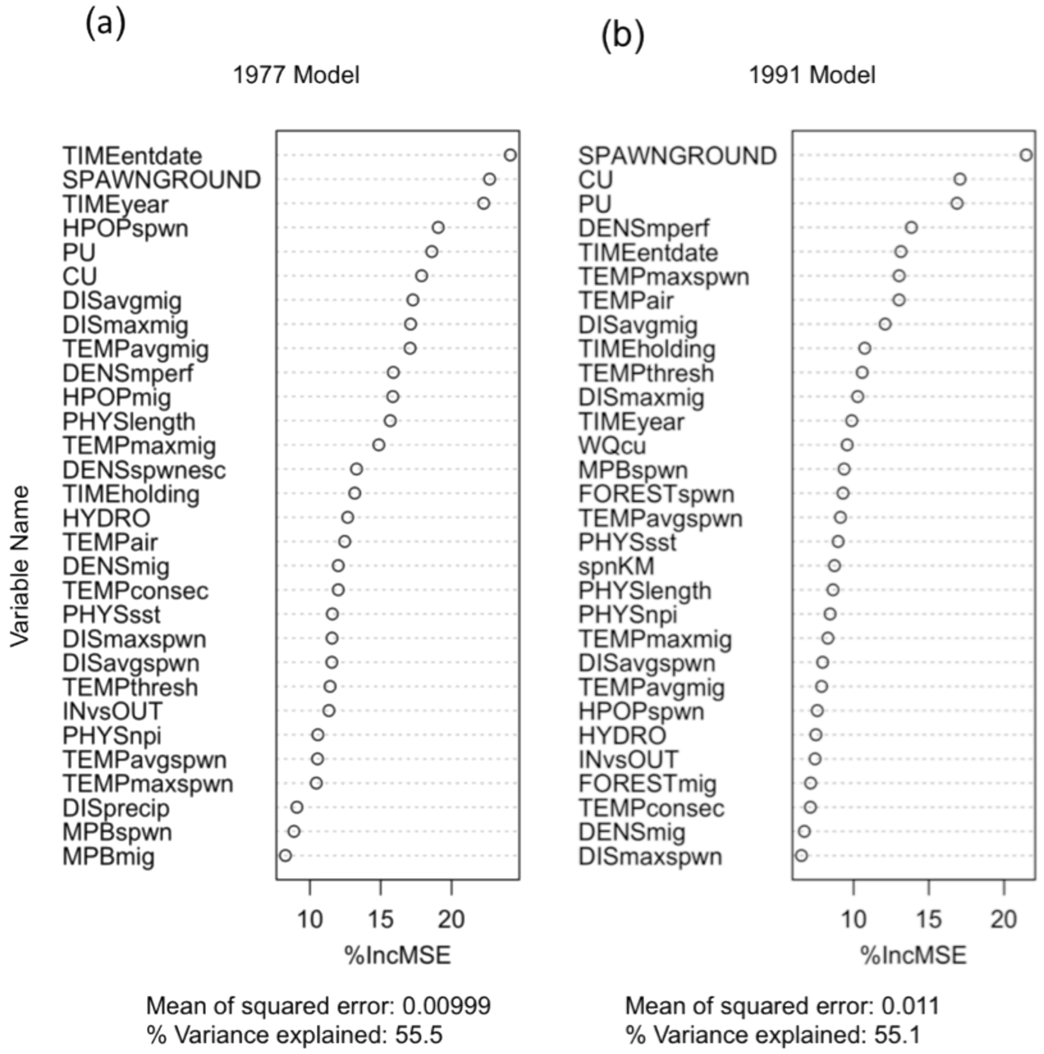

Figure 3.3. *Variable importance plots for predictor variables used in the (a) 1977-2008 model and the (b) 1991-2008 model. Percentage increase in mean squared error (%IncMSE) indicates a predictor variables effect on prediction accuracy. Higher values of %IncMSE indicate variables of higher importance to model prediction. The percentage of total variance explained by each model is also shown.*

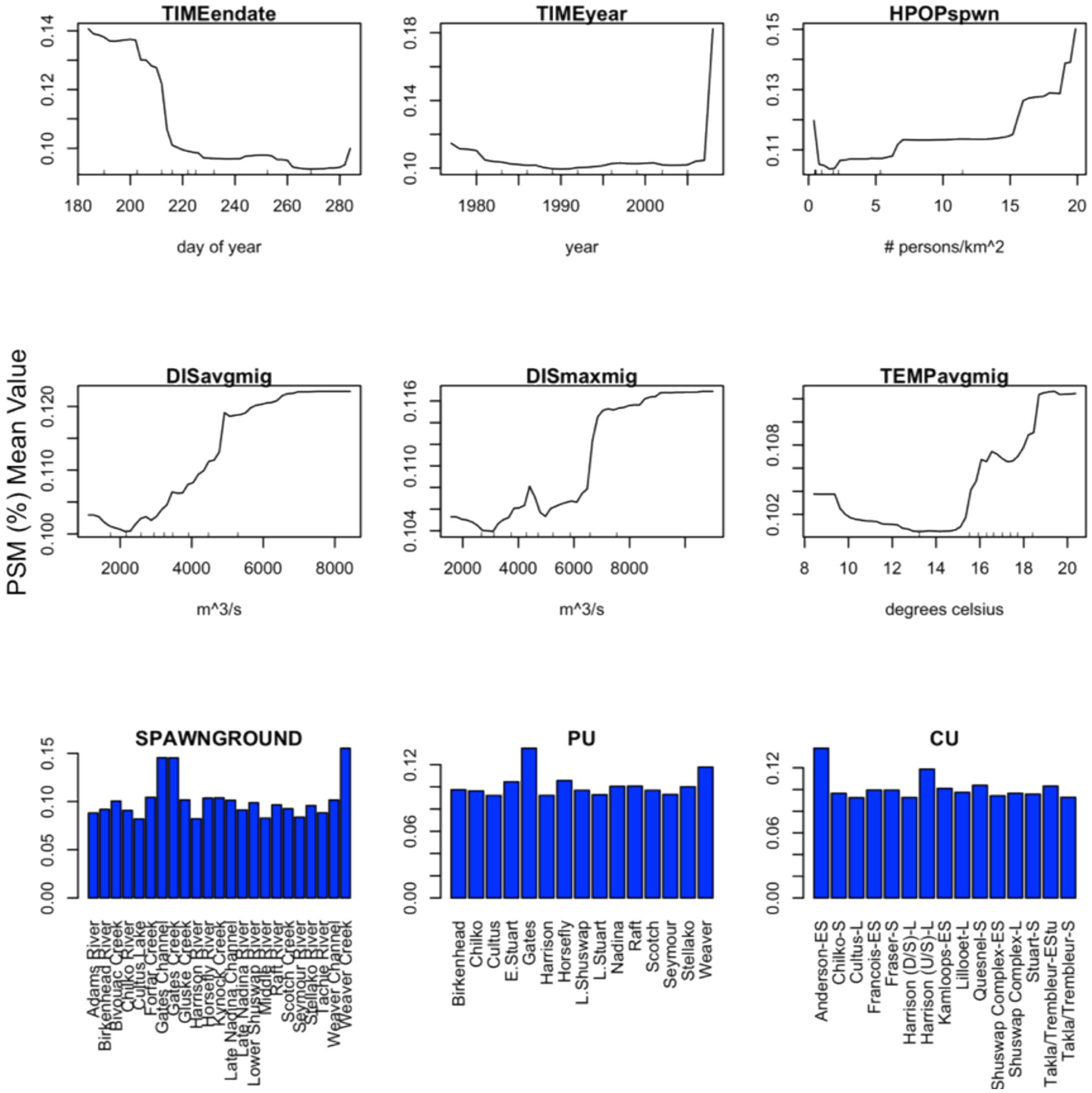

Figure 3.4. *Partial dependence plots for the 9 most important variables described by our 1977 random forest model. The Y-axis indicates mean value of prespawn mortality rate. The X-axis indicates the value of each predictor variable. Dashmarks along the X-axis indicate deciles.*

References

Anderson, R. V. (1978). The effects of lead on oxygen uptake in the crayfish *Orconectes virilis* (Hagan). Bulletin of environmental contamination and toxicology.

Bengtsson, B. E., (1974). Effects of zinc on the minnow Phoxinus phoxinus. Oikos 25, 370-373.

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

Bilinski, E. & R. E. E. Jonas. (1973). Effects of cadmium and copper on the oxidation of lactate by rainbow trout (*Salmo gairdneri*) gills. Journal of the Fisheries Resource Board of Canada 30, 1553-1558.

Brungs, W. A. (1969). Chronic Toxicity of zinc to the fathead minnow Pimephales promelas Rafinesque. Transactions of the American Fisheries Society 98, 272-279.

Christensen, V. & Trites, A. W. (2011). Predation on Fraser River sockeye salmon. Cohen Commission Tech. Rept. 8:129p. Vancouver, B.C. www.cohencommission.ca

Cooke, S. J., Hinch, S. G., Farrell, A. P., Patterson, D. A., Miller-Saunders, K., Welch, M. R., Donaldson, D. W., Hanson, K. C., Crossin, G. T., Mathes, M. T., Lotto, A. G., Hruska, K. A., Olsson, I. C., Wagner, G. N., Thomson, R., Hourston, R., English, K. K., Larsson, S., Shrimpton, J. M. & Van der Kraak., G., (2008). Developing a mechanistic understanding of fish migrations by linking telemetry with physiology, behavior, genomics and experimental biology: an interdisciplinary case study on adult Fraser River sockeye salmon. Fisheries 33, 321-338.

Crain, C. M., Kroeker, K. & Halpern, B. S. (2008). Interactive and cumulative effects of multiple human stressors in marine systems. Ecology Letters 11, 1304-1315.

Crossin, G. T., Hinch, S. G., Farrell, A. P., Higgs, D. A. & Healey, M. C. (2004). Somatic energy of sockeye salmon *Oncorhynchus nerka* at the onset of upriver migration: a comparison among ocean climate regimes. Fisheries Oceanography 13, 345-349.

De'ath ,G. (2007). Boosted Trees for ecological modeling and prediction. Ecology 88, 243-251.

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

Dorcey, A. H. J. & Griggs, J. R. (eds). (1991). Water in Sustainable Development: Exploring Our Common Future in the Fraser River Basin. Research Program on Water in Sustainable Development: Volume II. Westwater Research Centre, Faculty of Graduate Studies.

DFO (Fisheries and Oceans Canada). Canada's Policy for Conservation of Wild Pacific Salmon. (2005). Available at www.pac.dfo-mpo.gc.ca/fm-gp/species-especes/salmon-saumon/wsp-pss/docs/wsp-pss-eng.pdf

Eliason, E. J., Clark, T. D., Hague, M. J., Hanson, L. M., Gallagher, Z. S., Jeffries, K. M., Gale, M. K., Patterson, D. A., Hinch, S. G. & Farrell, A. P. (2011). Differences in thermal tolerance among sockeye salmon populations. Science 332, 109-112

Elith, J., Leathwick, J. R. & Hastie, T. (2008). A working guide to boosted regression trees. Journal of Animal Ecology 77, 802–813.

English, K. K., Kowski, W. R. , Sliwinski, C., Blakely, A., Cass A. & Woody, J. (2003). Migration timing and in-river survival of Late-run Fraser River sockeye using radio-telemetry techniques. Prepared by LGL Limited Environmental Research Assosciates (Sidney, BC) for Fisheries and Oceans Canada (Nanaimo, BC).

English, K. K., Sliwinski, C., Labelle, M., Koski, W. R., Alexander, R., Cass, A. & Woodey, J. C. (2004). Migration timing and in-river survival of Late-run Fraser River sockeye salmon in 2003 using radio-telemetry techniques. Prepared by LGL Limited Environmental Research Assosciates (Sidney, BC) for Fisheries and Oceans Canada (Nanaimo, BC).

English, K. K., Koski, W. R., Sliwinski C. & Blakley A. (2005). Migration timing and in-river survival of Late-run Fraser River sockeye using radio-telemetry techniques. Transactions of the American Fisheries Society134,1342-1365.

Farrell, A. P., Hinch, S. G., Cooke, S. J., Patterson, D. A., Crossin, G. T., Lapointe, M. & Mathes, M. T. (2008). Pacific salmon in hot water: applying aerobic scope models and biotelemetry to predict the success of spawning migrations. Physiological and Biochemical Zoology 81, 697-708.

Folt, C. L., Chen, C. Y., Moore M. V. & Burnaford, J. (1999). Synergism and antagonism among multiple stressors. Limnology and Oceanography 44, 864-867.

Freidman, J. (2002). Tutorial: Getting Started with MART in R. Available at www-stat.stanford.edu/~jhf/r-mart/tutorial/tutorial.pdf

Gilhousen, P. (1990). Prespawning mortalities of sockeye salmon in the Fraser River system and possible causal factors. International Pacific Salmon Fisheries Commission XXVI. Vancouver, BC.

Griffiths, A., Machado-Schiaffino, G., Dillane, E., Coughlan, J., Horreo, J., Bowkett, A., Minting, P., Toms, S., Roche, W., Gargan, P., McGinnity, P., Cross, T., Bright, D., Garcia-Vazquez, E. & Stevens, J.(2010). Genetic stock identification of Atlantic salmon (*Salmo salar*) populations in the southern part of the European range**.** BMC Genetics 11**,** 31.

Hague, M.J. & Patterson, D.A. (2007). Quantifying the sensitivity of Fraser River sockeye salmon (*Oncorhynchus nerka*) management adjustment models to uncertainties in run timing, run shape and profile. Canadian Technical Report of Fisheries and Aquatic Sciences 2776.

Hague, M. J., Patterson, D. A. & Macdonald, J. S. (2008). Exploratory correlation analysis of multi-site summer temperature and flow data in the Fraser River basin. Canadian Technical Report of Fisheries and Aquatic Sciences 2797.

Hinch, S. G. & Rand. P. S. (1998). Swim speeds and energy use of upriver-migrating sockeye salmon (Oncorhynchus nerka): role of local environment and fish characteristics. Canadian Journal of Fisheries and Aquatic Sciences 55, 1821–1831.

Hinch, S. G. & Martins, E. G. (2011). A review of potential climate change effects on survival of Fraser River sockeye salmon and an analysis of interannual trends in en route loss and pre-spawn mortality.Cohen Commission Technical Report 9, 134p. Vancouver, B.C. Available at www.cohencommission.ca

Hodgson, S. & Quinn, T. P. (2002). The timing of adult sockeye salmon migration into fresh water: adaptations by populations to prevailing thermal regimes. CanadianJournal of Zoology 80, 542–555.

Hyatt, K. D., Stockwell, M. M. & Rankin, D. P. (2003). Impact and adaptation responses of Okanagan River sockeye salmon (*Oncorhynchus nerka*) to climate variation and change effects during freshwater migration: stock restoration and fisheries management implications. Canadian Water Resources Journal 28, 689–713.

Isaak, D. J., Wollrab, S., Horan D. & Chandler. G. (2011). Climate change effects on stream and river temperatures across the northwest U.S. from 1980–2009 and implications for salmonid fishes. Climatic Change. doi: 10.1007/s10584- 011-0326-z.

Ikuta, K., Munakata, A., Aida, K., Amano M. & Kitamura, S. (2001). Effects low pH on upstream migratory behavior in land-locked cockeye salmon *Oncorhynchus nerka.* Water Air Soil Pollution 130, 99-106.

Johnson, J. E., Patterson, D. A., Martins, E. G., Cooke, S. J. & Hinch, S.G. (2012). Quantitative methods for analyzing cumulative effects on fish migration success: a review. Journal of Fish Biology 81, 600-631.

Keefer, M. L., Peery, C. A. & Heinrich, M. J. (2008). Temperature-mediated spawning migration mortality and travel rates of endangered Snake River sockeye salmon. Ecology of Freshwater Fish 17, 136–145.

Killick, S. R. (1955). The chronological order of Fraser River sokeye salmon during migration, spawning and death. International Pacific Fisheries Commission Bulletin VII.

Lee, C. G., Farrell, A. P., Lotto, A. G., MacNutt, M. J., Hinch, S. G. & Healey, M. C. (2003). The effect of temperature on swimming performance and oxygen consumption in adult sockeye (Oncorhynchus nerka) and coho (O. kisutch) salmon stocks. Journal of Experimental Biology 206, 3239–3251.

Liaw, A. & Wiener M. (2002). Classification and Regression by randomForest. R News 2(3), 18-22.

Macdonald, J. S., Foreman, M. G. G., Farrell, A. P., Wiliams, I. V., Grout, J., Cass, A. Woody, J. C., Enzenhofer, H., Clarke, W. C., Houtman, R., Donaldson E. M. & Barnes, D. (2000). The influence of extreme water temperatures on migrating Fraser River sockeye salmon during the 1998 spawning season. Canadian Technical Report on Fisheries and Aquatic Sciences 2326.

Macdonald, J. S., Patterson, D. A., Hague, M. J. & Guthrie, I.C. (2010). Modeling the influence of environmental factors on spawning migration mortality for sockeye salmon fisheries management in the Fraser River, British Columbia. Transactions of the American Fisheries Society 139, 768-782.

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

Mantua, N., Tohver, I. & Hamlet., A. (2010). Climate change impacts on stream flow extremes and summertime stream temperature and their possible consequences for freshwater salmon habitat in Washington State. Climatic Change 102, 187-223. doi: 10.1007/s10584-010-9845-2.

Marmorek, D., Pickard, D., Hall, A., Bryan, K., Martell, L., Alexander, C., Wieckowski, K. Greig, L. & Schwarz, C. (2011). Fraser River sockeye salmon: data synthesis and cumulative impacts. ESSA Technologies Ltd. Cohen Commission Tech. Rep. 6. 273p. Vancouver, B.C. www.cohencommission.ca.

Marmulla, G. (eds). 2001. Dams, fish and fisheries: Opportunities, challenges and conflict resolution. FAO Fisheries Technical Paper 419. Rome, FAO.

Martins, E. G., Hinch, S. G., Patterson, D. A., Hague, M. J., Cooke, S. J., Miller, K. M. & Lapointe, M. F. (2011). Effects of river temperature and climate warming on stock specific survival of adult migrating Fraser River sockeye salmon (*Oncorhynchus nerka*). Global Change Biology 17, 99-114.

McIntrye, J. D. (1973). Toxicity of methyl mercury on steelhead trout sperm. Bulletin of Environmental Contaminants and Toxicology 9, 98-99.

Morrison, J., Quick, M. C. & Foreman, M. G. G. (2002). Climate change in the Fraser River watershed: flow and temperature predictions. Journal of Hydrology 263, 230-244.

Nelitz, M., Porter, M., Parkinson, E., Wieckowski, K., Marmorek, D., Bryan, K., Hall, A. & Abraham, D. (2011). Evaluating the status of Fraser River sockeye salmon and role of freshwater ecology in their decline. ESSA Technologies Ltd. Cohen Commission Technical Report 3: 222p. Vancouver, BC.

Patterson, D. A. & Hague, M. A. (2007). Evaluation of long range summer forecasts of lower Fraser River discharge and temperature conditions. Canadian Technical Report on Fisheries and Aquatic Sciences 2754

Pickering, Q.H. & Gast M. H. (1972). Acute and Chronic toxicity of cadmium to the fathead minnow (Pimephales Promelas). Journal of the Fisheries Resource Board of Canada. 29, 1099-1106.

Platts, W. S. (1991). Livestock grazing. In Influence of Forest and Rangeland Management on Salmonid Fishes and Their Habitats (W.R. Meehan, eds). American Fisheries Society. Bethesda, Md. 751pp

Prasad, A., Iverson, L. & Liaw, A. (2006). Newer classifcation and regression tree techniques: Bagging and random forests for ecological prediction. Ecosystems 9, 181-199.

Rand, P. S., Hinch, S. G., Morrison, J., Foreman, M. G. C., MacNutt, M. J., Macdonald, J. S., Healey, M. C., Farrell, A. P. & Higgs, D. A. (2006). Effect of river discharge, temperature, and future climates on energetics and mortality of adult migrating Fraser River sockeye salmon. Transactions of the American Fisheries Society 135, 655-667.

Ripley, B. (2011). tree: Classification and regression trees. R package version 1.0-29. http://CRAN.R-project.org/package=tree

R Development Core Team. (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing,Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Richter, A. & Kolmes, S. A. (2005). Maximum temperature limits for Chinook, coho, and chum salmon, and steelhead trout in the Pacific Northwest. Reviews in Fisheries Science 13, 23–49.

Robichaud, D. & English, K. K. (2006). Assessing in-river migration behaviour and survival of Summer-run sockeye salmon caught and released in the Lower (Sidney, BC) for the Pacific Salmon Commission (Vancouver, BC).

Roos, J. F. (1991). Restoring Fraser River salmon: A history of the International Pacific Salmon Fisheries Commission 1937-1985. Pacific Salmon Commission. Vancouver, Canada.

Rosenau, M. L. & Angelo M. (2009). Landscape-level impacts to salmon and steelhead stream habitats in British Columbia. Prepared for Pacific Fisheries Resource Conservation Council, Vancouver, B.C. Available at www.fish.bc.ca/files/LandscapeReport_2009_0_Complete.pdf

Schreck, J. A., Contreras-Sanchez, W. & Fitzpatrick M. S. (2001). Effects of stress on fish reproduction, gamete quality, and progeny. Aquaculture 197, 3-34.

Servizi, J. A. &. Jensen, J. O. T (1977). Resistance of adult sockeye salmon to acute thermal shock. International Pacific Salmon Fisheries Commission, progress report 34, New Westminster, BC.

St. Hilaire, S., Boichuk, M., Barnes, D., Higgins, M., Devlin, R., Withler, R., Khattra, J., Jones S. & D. Kaiser. (2001). Epizoology of Parvicapsula minibicornis in Fraser River sockeye salmon, *Oncorhynchus nerka* (Walbaum). Journal of Fish Diseases 25, 107-120.

Voss, F. D., Curran, C. A. & Mastin, M. C. (2008). Modeling water temperature in the Yakima River, Washington, from Roza Diversion Dam to Prosser dam, 2005–06: U.S. Geological Survey Scientific Investigations Report 2008-5070, 42 p.

Wagner, G. N., Hinch, S. G., Kuchel, L. J. , Lotto, A. G., Jones, S. R. M. Patterson, D. A., Macdonald, J. S., Van Der Kraak, G., Shrimpton, J. M., English, K. K., Larsson, S., Cooke, S. J., Healey, M. C. & Farrell, A. P. (2005). Metabolic rates and swimming performance of adult Fraser River sockeye salmon (*Oncorhynchus nerka*) after a controlled infection with *Parvicapsula minibicornis*. Canadian Journal of Fisheries and Aquatic Sciences 62, 2124-2133.

Young, J. L., Hinch, S. G., Cooke, S. J., Crossin, G. T., Patterson, D. A., Farrell, A. P., Van Der Kraak, G., Lotto, A. G., Lister, A., Healey, M. C. & English, K. K. (2006). Physiological and energetic correlates of en route mortality for abnormally early migrating adult sockeye salmon (*Oncorhynchus nerka*) in the Thompson River, British Columbia. Canadian Journal of Fisheries and Aquatic Sciences 63, 1067-1077.

Chapter 4.

General Conclusions

The research project presented herein should provide a wealth of useful information and tools that fisheries researchers and managers can use to model, analyze, and mitigate cumulative effects of stressors to Fraser River Sockeye salmon.

In Chapter 2, ways in which to improve the current state of cumulative effects analyses in Fraser River sockeye salmon migration studies were identified. A list of information on numerous multivariable methods, both well established and novel, from fisheries and other biological sciences that are all free to use in R statistical packages (R Development Core Team, 2011) was also provided. Many of the methods discussed in this chapter are extremely powerful for describing and predicting the cumulative effects of stressors or other variables and could prove useful to any future cumulative effects studies.

In Chapter 3, classification and regression trees and random forests were used to advance the current understanding of how different processes affect Fraser River sockeye salmon prespawn mortality (PSM). The results of this study are important for a number of reasons. First, they point to the need to further examine potential causes of PSM in Weaver and Gates populations, as there seems to be some unaccounted for factor causing those populations to have higher than average PSM rates. This study also suggests that earlier migrants tend to experience higher than average PSM rates. This result is especially important considering that Late-run sockeye have migrated much earlier than usual over the past 10 years (Macdonald *et al.,* 2010); and reaffirms the need to further examine potential causes of this phenomenon. Furthermore, our study reaffirms previous hypotheses that river temperature and discharge may play an important role in sockeye PSM, though perhaps only in a synergistic fashion with other variables. Finally, the random forest Model constructed in this chapter provides the most up-to-date and advanced form of predictive model for PSM rates in sockeye salmon. This model requires the input of only a few variables (timing of entry, destination spawning ground and human population density near the spawning ground) and can assist with PSM data collection when observer estimates are not possible. Furthermore, this model can help fisheries managers mitigate future extreme PSM events by providing yearly estimates of PSM rates in multiple stocks in time to adjust escapement targets.

Finally, while the main focus of this research project is the analysis cumulative effects to Fraser River sockeye salmon migration, the methodology presented herein is applicable to all types of migratory fishes and other animals. Indeed, this research project should serve as an excellent guide for anyone who wishes to model the cumulative effects of multiple stressor variables.

References

Macdonald, J. S., Patterson, D. A., Hague, M. J. & Guthrie, I.C. (2010). Modeling the influence of environmental factors on spawning migration mortality for sockeye salmon fisheries management in the Fraser River, British Columbia. Transactions of the American Fisheries Society 139, 768-782.

R Development Core Team. (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Appendices

Appendix A. List of References

List of all references found using the Web of Knowledge and Aquatic Science and Fisheries Abstracts search engines classified by multivariable method used and general class of predictor and response variables.

Reference	Method(s)	Response class(es)	Predictor class(es)
De Raedemacker *et al*. (2010)	MFANOVA	Habitat preference/ distribution	Habitat characteristics
Del Toro-Silva *et al*. (2008)	MFANOVA	Physiology	Environmental stressors
Evans & Neff (2009)	MFANOVA	Physiology	Genetics
Ginot *et al.* (2006)	MFANOVA	Survival, Demographics	Biological stressors
Growns *et al.* (2006)	MFANOVA	Assemblage/ community structure	Spatial and temporal variables
Locham *et al.* (2010)	MFANOVA	Assemblage/ community structure	Habitat characteristics
Mairesse *et al.* (2007)	MFANOVA	Physiology	Behaviour, Stocking biomass
Reilly *et al.* (1992)	MFANOVA	Behaviour	Habitat characteristics, Physiology
Rodd & Reznick (1997)	MFANOVA	Demographics	Environmental stressors, Physiology, Demographics
Slawski *et al.* (2008)	MFANOVA	Assemblage/ community structure	Environmental stressors, Habitat characteristics
Steele *et al.* (1998)	MFANOVA	Survival, Demographics	Environmental stressors
Strecker *et al.* (2011)	MFANOVA	Physiology	Habitat characteristics, Temporal and spatial variables

Trabelsi *et al*. (2011)	MFANOVA	Behaviour, Survival, Physiology	Environmental stressors, Behaviour
Williams *et al.* (2002)	MFANOVA	Assemblage/ community structure	Environmental stressors, Habitat characteristics, Temporal variables
Cabral *et al.* (1998)	MLR	Behaviour	Environmental stressors, Habitat characteristics
Claramunt & Wahl (2000)	MLR	Physiology	Environmental stressors, Habitat characteristics, Temporal variables
Crozier *et al*. (2010)	MLR	Physiology	Environmental stressors
Lindegren *et al.* (2011)	MLR	Survival, demographics	Environmental stressors, Survival
Maceina (1992)	MLR	Physiology	Habitat characteristics, Demographics
Magnan *et al.* (1994)	MLR	Behaviour	Environmental stressors, Habitat characteristics, physiology, Temporal variables
Pyle *et al.* (2005)	MLR	Assemblage/ community structure, Physiology	Environmental stressors
Raffenberg & Parrish (2003)	MLR	Survival, Physiology	Environmental stressors
Rorvik *et al*. (2003)	MLR	Survival	Biological stressors, Behaviour
Stevens *et al.* (2010)	MLR	Assemblage/ community structure	Environmental stressors, Habitat characteristics
Vadas & Orth (2001)	MLR	Habitat preference/ distribution	Environmental stressors, Habitat characteristics
Holbrook & Schmitt (2003)	MLR, MFANOVA	Survival	Habitat characteristics
Riginos & Nachman (2001)	MLR, MFANOVA	Genetics	Habitat characteristics, Spatial variables

Seenapa & Devaraj (1995)	MLR, MFANOVA	Physiology	Physiology
Wille *et al*. (2002)	MLR, MFANOVA	Physiology	Environmental stressors, Physiology
Dunham *et al*. (2002)	QR	Survival, demographics	Habitat characteristics
Planque & Buffaz (2008)	QR	Survival, demographics	Environmental stressors, Habitat characteristics
De Zwart *et al*. (2006)	GLM	Survival, demographics	Environmental stressors, Habitat characteristics
Goncalvez *et al*. (2008)	GLM	Behaviour	Environmental stressors
Magaud *et al*. (1997)	GLM	Survival	Environmental stressors
Malins *et al*. (2004)	GLM	Genetics	Environmental stressors
Marr *et al*. (1998)	GLM	Survival	Environmental stressors
Theriault *et al*. (2007)	GLM	Behaviour	Physiology, Genetics
Tsitsika & Maravelias (2006)	GLM	Catch	Fishing characteristics, Temporal variables
Young *et al*. (2010)	GLM	Habitat preference/ distribution	Habitat characteristics
Galvez *et al*. (2007)	GLM, MFANOVA	Physiology	Environmental stressors
Jeschke & Strayer (2006)	GLM, MLR	Habitat preference/ distribution	Human affiliation, Propagule Pressure, Hunting, Physiology
Brenden *et al*. (2007)	GAM	Assemblage/ community structure	Habitat characteristics
Buission *et al*. (2008)	GAM	Habitat preference/ distribution	Environmental stressors, Habitat characteristics
Carol *et al*. (2006)	GAM	Assemblage/ community structure	Environmental stressors, Habitat characteristics

Ciannelli *et al.* (2004)	GAM	Survival	Environmental stressors, Habitat characteristics
Dingsor *et al.* (2007)	GAM	Survival, Demographics	Environmental stressors
Hernandez *et al*. (2009)	GAM	Habitat preference/ distribution	Environmental stressors, Habitat characteristics, Genetics
Jowett *et al.* (2007)	GAM	Habitat preference/ distribution	Habitat characteristics
Kai & Marsac (2010)	GAM	Assemblage/ community structure	Habitat characteristics
Katara *et al.* (2011)	GAM	Assemblage/ community structure	Environmental stressors, Habitat characteristics
Kupschus (2003)	GAM	Habitat preference/ distribution	Environmental stressors, Habitat characteristics, Temporal variables
Lorance *et al*. (2010)	GAM	Catch	Fishing characteristics, Temporal variables
Swartzman *et al*. (1992)	GAM	Habitat preference/ distribution	Environmental stressors, Habitat characteristics
Zagaglia *et al*. (2004)	GAM	Catch	Environmental stressors, Habitat characteristics
Zhang *et al.* (2010)	GAM	Survival	Environmental stressors, Habitat characteristics
Auth *et al.* (2011)	GAM	Assemblage/ community structure	Environmental stressors, Habitat characteristics
Esteves *et al*. (2009)	GAM, MFANOVA	Physiology	Environmental stressors, Habitat characteristics, Physiology
Cury *et al.* (1998)	GAM, MLR	Catch	Environmental stressors
Bacheler *et al*. (2009)	GAM, GLM	Habitat preference/ distribution	Environmental stressors
Cheng & Gallinat (2004)	GAM, GLM	Catch	Environmental stressors, Habitat characteristics

Jobling *et al*. (2009)	GAM, GLM	Physiology	Environmental stressors
Stoner *et al*. (2001)	GAM, GLM	Habitat preference/ distribution	Environmental stressors, Habitat characteristics
Yee (2010)	GAM, GLM, QR	Physiology, Catch	Fishing characteristics, Temporal and spatial variables
Claireaux & Lagardere (1999)	NLR, MFANOVA	Physiology	Environmental stressors, Habitat characteristics
Laetz *et al.* (2009)	NLR, MFANOVA	Physiology	Environmental stressors
Ayllon *et al*. (2009)	NLR, MLR	Habitat preference/ distribution	Environmental stressors, Habitat characteristics
Blanc (2005)	NLR, MLR, MFANOVA	Physiology	Environmental stressors, genetics
Leathwick *et al.* (2005)	MARS, GLM	Habitat preference/ distribution	Environmental stressors, Habitat characteristics
Leathwick *et al*. (2006*b*)	MARS, GAM	Habitat preference/ distribution	Environmental stressors, Habitat characteristics
Ruppert *et al*. (2010)	CART	Assemblage/ community structure	Environmental stressors, Habitat characteristics
Vignon & Sasal (2010)	CART	Physiology	Environmental stressors, Habitat characteristics, Physiology
Elith *et al*. (2008)	BOT	Habitat preference/ distribution	Environmental stressors, Habitat characteristics
Leathwick *et al*. (2008)	BOT	Habitat preference/ distribution	Environmental stressors, Habitat characteristics
Leathwick *et al.* (2006*a*)	BOT, GAM	Assemblage/ community structure	Environmental stressors, Habitat characteristics, Fishing characteristics
Grossman *et al.* (1998)	CLA, MLR	Assemblage/ community structure	Environmental stressors, Habitat characteristics
Diaz *et al*. (2003)	CLA, GLM	Survival, Demographics	Environmental stressors, Habitat characteristics, Physiology

Knudby *et al*. (2010)	SVM, BOT, RF, BAT, GAM	Assemblage/ community Structure, Survival, Demographics	Environmental stressors, Habitat characteristics
Brosse *et al.* (2009)	ANN	Habitat preference/ distribution	Environmental stressors, Habitat characteristics
Erbe & King (2009)	ANN	Behaviour	Environmental stressors
Olden & Jackson. (2001)	ANN	Habitat preference/ distribution, Survival, Demographics	Environmental stressors, Habitat characteristics
Palialexis *et al.* (2011)	ANN	Habitat preference/ distribution	Environmental stressors, Habitat characteristics
Yanez *et al*. (2010)	ANN	Survival	Environmental stressors, Habitat characteristics
Gutierrez-Estrada *et al.* (2009)	ANN, GAM, MLR	Catch	Environmental stressors, Habitat characteristics
Lee *et al.* (2009)	ANN, GAM, MLR	Survival, Demographics	Environmental stressors, Habitat characteristics
Megrey *et al.* (2005)	ANN, GAM, NLR, MLR	Survival, Demographics	Environmental stressors, Habitat characteristics
Foody (2004)	GWR	Assemblage/ community structure	Environmental stressors
Austin (2007)	SEM, GWR	Case studies	Case studies
Palmores *et al.* (1998)	SEM	Survival	Environmental stressors, Habitat characteristics

MFANOVA= Multi-Factor Analysis of Variance, MLR= Multiple Linear Regression, QR= Quantile Regression, SA= Survival Analysis, GLM= Generalized Linear Model, GAM= Generalized Additive Model, NLR= Non-linear Regression, MARS= Multivariate Adaptive Regression Splines, CART=Classification & Regression Trees, BAT= Bagging Tree, BOT=Boosted Tree, RF=Random Forest, CLA= Cluster Analysis, ANN= Artificial Neural Network, GWR= Geographically Weighted Regression, SEM= Structural Equation Modeling.

Appendix B. Boxcar Model and Data Cleaning

Boxcar Model

The boxcar model functions much like a train with linked boxcars moving along a simulated track. For our study, the track is the Fraser River watershed, a train is one of the 24 spawning ground populations, and each of the boxcars in a train represents a portion of fish from that population. In order to capture temperatures experienced by an entire spawning ground population, each train in our analysis contained a succession of 31 boxcars. 31 boxcars were chosen to mimic the approximate 31-day period it takes for each run-timing group to enter the Fraser River (Hague et al., 2008). The first boxcar in each train begins the trip up the Fraser River 15 days prior to the population's estimated median entry date with each successive boxcar beginning the trip up the Fraser River on successive days. While migration speed likely varies for different spawning ground populations and within different reaches of the river (Hague *et al.,* 2008), for simplicity each of the boxcars in the train moved up the Fraser River at a constant speed based on the average of *median* migration speeds for each run-timing group (45, 42, 42 and 23 km/day for Early Stuart, Early Summer, Summer and Late run groups respectively; Killick, 1955, English *et al.,* 2003; English *et al.;* 2004, English *et al.,* 2005, Robichaud & English, 2006). Order of the boxcars in each train was maintained throughout migration and each boxcar was assumed to represent an equivalent proportion fish.

Boxcars moving along the Fraser River watershed recorded daily mean temperature readings from numerous real-time temperature logger stations (Figure 2.1). These logger stations were thermographs from1977-1987 and then digital temperature logger stations from 1988-2008 (Lisa Thompson, DFO, pers. comm., May 15, 2012). Temperature readings from one station are recorded by a boxcar until that boxcar is halfway to the next station along the route. At this point, the boxcar begins recording temperature readings from the new site. If a boxcar moves into a river tributary, the boxcar begins recording temperature readings from the closest site in the new tributary. A boxcar stops recording temperature readings once it arrives at the spawning ground.

For discharge variables representing conditions experienced during migration, data was collected from only a single station at Hope instead of from multiple stations throughout the Fraser River. This is because discharge was used as an index for water volume and velocity and is measured in m^3/sec. Two different sections of the same river may have equal discharges but very different water velocities or volume of available habitat. Discharge at Hope is known to correlate highly with discharge levels upstream (Hague et al., 2008) hence it was assumed that discharge at Hope would provide a good index of velocities and volumes elsewhere in the Fraser River. Data for daily mean river discharge at Hope (Figure 2.1) is collected using a discharge gauging station run by Environment Canada and time series of this data were taken from their publicly-available, Water Survey of Canada (WSC) online database (www.wsc.ec.gc.ca).

Temperature and Discharge Site Data Cleaning

Time series of daily temperature readings vary in terms of length (some extend back to 1990 while other extend as far back as 1940) and consistency across sites as these stations may go offline for several days. If a particular site was missing data, we used data from other, highly correlated temperature sites to fill in gaps and extend shorter time series. For example, if site FR190 was missing several weeks of data for the year 1992, we found the site used in our analysis for which the daily temperature readings correlated highest with those from FR190 in1992 and then used a simple linear regression relationship to fill in those missing days. We repeated this process using the subsequent most highly correlated site and continued to do so long as the correlation coefficient was above 0.6. In order to fill in missing years and any remaining gaps in data, we found the most correlated site over all years and then used a linear regression relationship to fill in all of the missing years of data. Again, if data was still missing, we then used the subsequent most highly correlated site and continued to do so long as the correlation coefficient was above 0.6. The above process ensured that all temperature stations had complete time series dating back to 1977. The same process was used to fill all gaps in time series for the WSC discharge stations.