

2014

ECOLOGICAL BENEFITS OF THE MARINE STEWARDSHIP COUNCIL ECOLABEL: A GLOBAL ANALYSIS

Sarah Blackstock
University of Rhode Island, sblackstock@my.uri.edu

Follow this and additional works at: <https://digitalcommons.uri.edu/theses>

Terms of Use

All rights reserved under copyright.

Recommended Citation

Blackstock, Sarah, "ECOLOGICAL BENEFITS OF THE MARINE STEWARDSHIP COUNCIL ECOLABEL: A GLOBAL ANALYSIS" (2014). *Open Access Master's Theses*. Paper 396.
<https://digitalcommons.uri.edu/theses/396>

This Thesis is brought to you by the University of Rhode Island. It has been accepted for inclusion in Open Access Master's Theses by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons-group@uri.edu. For permission to reuse copyrighted content, contact the author directly.

ECOLOGICAL BENEFITS OF THE MARINE
STEWARDSHIP COUNCIL ECOLABEL: A GLOBAL
ANALYSIS

BY

SARAH BLACKSTOCK

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
MASTERS OF SCIENCE
IN
OCEANOGRAPHY

UNIVERSITY OF RHODE ISLAND

2014

MASTER OF SCIENCE THESIS

OF

SARAH BLACKSTOCK

APPROVED:

Thesis Committee:

Major Professor Jeremy Collie

Susanne Menden-Deuer

Hirotsugu Uchida

Nasser H. Zawia
DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND
2014

ABSTRACT

Recent reports suggest that many of the world's commercial fisheries are overfished, and therefore require rebuilding. Ecolabels have been introduced as a potential method to mitigate this chronic misuse of natural resources. These labels are intended to indicate to consumers that a product is sustainable. Assuming consumers value the ecolabel, their business provides a market benefit to fishermen in exchange for the sustainable exploitation of fish stocks. Though some fisheries ecolabels have been in operation for over a decade, little research has been done to examine their measurable benefits, economic or ecological. Here, we analyze the ecological effects of the Marine Stewardship Council (MSC) ecolabel on six exploited fish taxa, represented by 41 separate stocks. Within each taxon, we use a case-control design to compare trends in stock status (standardized spawning stock biomass and fishing mortality rates) between similar certified and uncertified fish stocks in the 8-10 year period surrounding certification. Certified stocks of only one taxon showed significantly greater improvements in stock status than uncertified stocks. In the other taxa, there was no detectable difference, or the certified stocks showed significantly more negative trends in status than the uncertified stocks. In a combined meta-analysis, certified stocks were found to be decreasing in stock status, while uncertified stocks increased, indicating that MSC ecolabels may not be actively improving the stock status of their fisheries.

ACKNOWLEDGMENTS

I would like to acknowledge my major professor, Dr. Jeremy Collie, for his invaluable guidance as I completed this study. I also thank my committee members, Dr. Hiro Uchida and Dr. Susanne Menden-Deuer, who have advised me both in the classroom and in the design and writing of this study. For their advice on study design and revisions of previous drafts, I would like to thank Dr. Marie-Joelle Rochet and Dr. Cathy Roheim. Finally, for the provision of data and other important information, I would like to thank Simon DeLestang, Adrian Linanne, Rebecah Rademeyer, Matt Dorn, Jorge Farias, and Bob O'Boyle.

PREFACE

This report is original work by the author, S. Blackstock. It is submitted in partial fulfillment of the requirements of a Master of Oceanography degree at the University of Rhode Island's Graduate School of Oceanography. While the study is currently unpublished, it is in preparation for submission to the Canadian Journal of Fisheries and Aquatic Sciences. Manuscript format is in use.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGMENTS	iii
PREFACE	iv
LIST OF TABLES	vi
MANUSCRIPT 1	1
ABSTRACT	2
INTRODUCTION	3
MATERIALS AND METHODS	7
RESULTS	15
DISCUSSION	22
REFERENCES	27
TABLES	32
FIGURES	35
APPENDICES	47
APPENDIX A. STOCK INPUT INFORMATION	47
APPENDIX B. STOCK DATA SOURCES	52
APPENDIX C. PROPENSITY SCORE METHODOLOGY	57
APPENDIX D. DETAILED MODEL OUTPUT	64
APPENDIX E. ALTERNATIVE LINEAR MODELS	68
BIBLIOGRAPHY	89

LIST OF TABLES

TABLE	PAGE
Table 1. Summary table of model selection results.....	32
Table 2. Summary of model output for each taxonomic group and the global analysis of MSC certification	34
Table A1. Summary of input data for each stock by taxonomic group	47
Table C1. Decision Makers and potential covariates for the propensity score model	61
Table C2. Estimated coefficients of the propensity score model	63
Table C3. Diagram of stock propensity scores by taxonomic group.....	64
Table D1(A-F). Estimated coefficients of the certification analysis by taxonomic group	65
Table E1. Output summary of linear mixed effects model with random slopes and intercepts.....	72
Table E2. Output summary of model with non-linear terms.....	84

LIST OF FIGURES

FIGURE	PAGE
Figure 1. Global distribution of control and treatment fish stocks.	35
Figure 2. Histogram of certified and control propensity scores.	37
Figure 3 (A-F). Observed data and fitted trends in spawning stock biomass and fishing mortality rate for certified and uncertified stocks of each taxonomic groups	38
Figure 4. Forest plot for meta-analysis of certification effects on trends in spawning stock biomass and fishing mortality rate	45
Figure E1 (A-G). Observed data and trends fitted by mixed effects model for certified and uncertified stocks of each taxonomic group.....	74
Figure E2 (A-D). Observed data and trends fitted by the model with non-linear terms for certified and uncertified stocks of each taxonomic groups.	85

MANUSCRIPT 1

In preparation for submission to the Canadian Journal of Fisheries and Aquatic Sciences

Ecological Benefits of the Marine Stewardship Council Ecolabel: A Global Analysis

by

Sarah Blackstock¹, Jeremy Collie², Marie-Joëlle Rochet³, Hirotsugu Uchida⁴,

¹ Masters Candidate, Department of Oceanography, The University of Rhode Island, Narragansett, RI 02882. Email: sblackstock@my.uri.edu

² Professor, Department of Oceanography, The University of Rhode Island, Narragansett, RI 02882. Email: jcollie@mail.uri.edu

³ Laboratoire MAERHA IFREMER, Rue de l'Île d'Yeu, B.P. 21105, 44311, Nantes, France. Email: mjrochet@ifremer.fr

⁴ Assistant Professor, Department of Environmental and Natural Resource Economics, The University of Rhode Island, Kingston, RI 02881. Email: uchida@uri.edu

ABSTRACT

Recent reports suggest that many of the world's commercial fisheries are overfished, and therefore require rebuilding. Ecolabels have been introduced as a potential method to mitigate this chronic misuse of natural resources. These labels are intended to indicate to consumers that a product is sustainable. Assuming consumers value the ecolabel, their business provides a market benefit to fishermen in exchange for the sustainable exploitation of fish stocks. Though some fisheries ecolabels have been in operation for over a decade, little research has been done to examine their measurable benefits, economic or ecological. Here, we analyze the ecological effects of the Marine Stewardship Council (MSC) ecolabel on six exploited fish taxa, represented by 41 separate stocks. Within each taxon, we use a case-control design to compare trends in stock status (standardized spawning stock biomass and fishing mortality rates) between similar certified and uncertified fish stocks in the 8-10 year period surrounding certification. Certified stocks of only one taxon showed significantly greater improvements in stock status than uncertified stocks. In the other taxa, there was no detectable difference, or the certified stocks showed significantly more negative trends in status than the uncertified stocks. In a combined meta-analysis, certified stocks were found to be decreasing in stock status, while uncertified stocks increased, indicating that MSC ecolabels may not be actively improving the stock status of their fisheries.

INTRODUCTION

Persistent overexploitation of fish stocks has become an ecological and economic burden worldwide (Jackson et al. 2001; Worm et al. 2009; Lotze et al. 2006; Pauly et al. 2003). However, the success of various recent efforts to recover fish stocks suggests that this historical misuse can be mitigated, provided the swift enactment of programs that sufficiently lower exploitation rates in the long term (Worm et al. 2009). While management authorities have set goals over the last few decades to constrain overexploitation, progress toward sustainable fishing is hindered by the unwillingness or inability to tolerate the short-term economic and social consequences of reduced fishing (Beddington et al. 2007; Worm et al. 2009). Fisheries ecolabeling programs such as the Marine Stewardship Council (MSC) may offer a socially efficient method to address this dilemma, and are being used increasingly as a market-based approach to fisheries sustainability (Hall and Manprize 2005; Wessels, 2001). The purpose of this labeling effort is to indicate to consumers that the labeled product was harvested sustainably. If consumers trust the MSC to identify sustainable fisheries and are willing to pay a premium for sustainable food, MSC certification will ideally return a market benefit to the fisheries that have earned and invested in the label.

The Marine Stewardship Council is an international non-profit organization founded in 1997 with the goal of transforming the fishing industry to a sustainable standard by minimizing environmental impacts and promoting effective management. Of all existing seafood ecolabels, the MSC label is the longest running and most broadly applied. There are currently 188 fisheries certified by the MSC, representing 7% of the global wild capture (MSC Global Impact Report 2013). The organization

sets its standards for sustainability based on three core principles: 1. the health of the target stock, 2. the impact of the fishery on the environment, and 3. the efficacy of the fishery's management system. Unlike many ecological sustainability programs, the "unit of certification" for the MSC is the fishery, not the species. This distinction acknowledges that different fishing practices and management structures in different areas can be more or less sustainable.

During the certification process, an assessment team independent of the MSC scores a fishery on a scale from 0-100 on 31 performance indicators that are grouped under each of the three core principles described above. To be certified a fishery must have an average score of 80 and a score of at least 60 on each performance indicator. A fishery can still be certified with a performance indicator score between 60 and 80, but only if an improvement plan is implemented. Once certified, fisheries are subject to annual third-party audits and must be recertified every five years. Fisheries involved with the MSC are thus driven to increase the sustainability of their operations before they seek certification, in anticipation of the assessment, and if any performance indicators were found to be lacking, after they are certified (Tlusty 2011; Martin 2012; MSC 2010). The individual fishery is responsible for the cost of certification, including pre-certification, assessments, subsequent audits, and any required improvements. Ranging from \$20,000 to \$300,000 (Goyert et al. 2010) this expense can be prohibitive. A fishery's decision to seek certification is thus based on its ability to absorb the cost of certification, its capacity to make required improvements, and the economic gain it expects from carrying the ecolabel. For the MSC certification, some

of the potential market benefits are a price premium on certified products (Roheim et al. 2011) and accessibility of new markets.

While seafood ecolabels like the MSC have great potential to effect change in global fisheries, it remains uncertain whether they are driving real ecological and economic improvements (Christian et al. 2013). These improvements depend on multiple factors, including the MSC's reliable identification of sustainable fisheries, its ability to communicate the significance of the ecolabel to consumers, and the consumers' willingness to pay higher prices for sustainable seafood. The evaluation of seafood certification has proven difficult because its history is brief; the MSC is the earliest of these programs and certified its first fishery only 15 years ago. Moreover, it is difficult to consistently measure improvements in a fishery, and to ensure that those improvements are directly related to certification (Agnew 2006). Various studies have suggested that MSC certification is associated with measurable environmental improvements (Martin et al. 2013; Gutiérrez et al. 2012; Agnew 2006), while others continue to question the label's ecological efficacy (e.g. Jacquet et al. 2009; Jacquet et al. 2010; Froese and Proelss 2011). In particular, the MSC's ability to detect and prevent declines in stock status, ensure minimal environmental impacts, and promote efficient management has been challenged (for instance, Pacific Hake, Gulf of Alaska Pollock, and Patagonian Toothfish, respectively). Some authors assert that it may still be too early to detect any measurable benefits of certification (Roheim 2009).

The objective of this study was to examine how MSC certification affects the ecology of fish stocks on a global scale. While any of the MSC's three principles for certification are good candidates for this assessment, we addressed only the first

principle, the status of the target stock. In this sense, this study has a narrow scope; a certified fishery that is not improving in stock status may be improving in other important characteristics such as the mitigation of ecosystem impacts. Nonetheless, stock status is the most comprehensive metric of stock health and improvement, and improvements in other performance indicators and principles will ultimately be reflected in stock status. Moreover, changes in the other principles can be difficult to measure quantitatively. This study presents the first analysis of MSC performance that: 1. Focuses on trends in stock status from pre-certification to post-certification, 2. Uses a treatment-control design with uncertified stocks as the controls, and 3. Analyzes data from independent stock assessments, as opposed to scores generated by the MSC itself.

MATERIALS AND METHODS

Rationale and Definitions

There is substantial evidence that fisheries that earn MSC certification begin making improvements in all three of the MSC's principles before certification is actually obtained (Martin et al. 2012). We therefore assessed whether certification creates a long-term dynamic towards sustainability, starting before the year of certification. Stock status has two components, spawning stock biomass (SSB) and fishing mortality rate (F), both of which are often standardized by reference points defined by a fishery's management. A fishery is considered by management to be in good stock status, or sustainable, when its fishing mortality rate is below its F reference point and its spawning stock biomass is above its SSB reference point. Improvement in stock status is therefore defined here as increases in standardized SSB and/or decreases in standardized F. Given the MSC's mission to increase the sustainability of the fishing industry and its requirement that fish stocks carrying its ecolabel maintain a certain level of sustainability, we expect certified fish stocks to show greater improvements in stock status than similar uncertified stocks over the 10-year period surrounding certification. Comparing stock status trends of certified and uncertified fisheries thus allows us to determine the additive effect of certification on stock status trends.

An important component of this study's rationale is that certified stocks are expected to show greater improvements in stock status than uncertified stocks only if the two groups started at similar stock status levels and share other similarities (for instance, life history traits and economic markets that value sustainability). If a

certified stock and a comparable uncertified stock both begin the time series in a sustainable state, they may both remain sustainable even if their stock status does not change over the ten-year time window. In this case, we do not expect significant differences in stock trends between certified and uncertified stocks. However, if the two stocks started with lower, relatively unsustainable stock status levels, the certified stock is expected to increase in stock status faster than the uncertified one due to pressure to improve from the MSC. We ensure the similarity of treatment and control groups using a propensity scoring methodology described below.

Data

Of the 188 fisheries certified by the MSC worldwide, we selected stocks that met three criteria. First, the stocks must have obtained certification in 2009 or earlier, providing five years of data post-certification. Second, we required a ten-year time series of spawning stock biomass and fishing mortality rate centered at the year of certification. The exception to this rule was North Sea Plaice, for which there were data only three years after certification. We chose to include this stock because it fulfilled the other two criteria satisfactorily. The final criterion for inclusion of certified stocks was the availability of control (uncertified) stocks that belonged to the same taxonomic group and for which similar time series data were available. Taxonomic groups in this study are defined by genus or species for all but the rock lobster group. In this case, the control stocks belong to a separate genus, but the same family (Paniluridae) as the certified stocks. This comparison was justified because these two species, though members of a different genus, share important life history

traits, geographic locations, and similar management structures (Australian Department of Agriculture 1991).

Thirteen certified stocks and 28 control stocks were ultimately selected, and represent a diversity of geographic locations and six distinct taxonomic groups (Fig. 1, Table A1): plaice (*Pleuronectes platessa*), saithe (*Pollachius virens*), herring (*Clupea harengus*), hake (*Merluccius* spp.), pollock (*Theragra chalcogramma*), and rock lobster (*Jasus edwardsii* and *Panulirus cygnus*). Time series of fishing mortality rates, abundance index (usually spawning stock biomass), and available reference points for these two metrics were obtained from published stock assessments or from stock management bodies (Appendix B). Stock assessors must work with existing data to determine the most reliable method to obtain estimates of stock status. The stock assessments used in this project thus were performed on datasets of varying quality using a variety of techniques. Nonetheless, they were all carefully crafted to represent the most complete and reliable description of the included stocks. Several of these control stocks were eventually excluded, as described below.

Propensity Scores

The case-control experimental design implemented in this project is predicated on the control stocks being as similar as possible to certified stocks. It is especially important that stocks have similar status at the beginning of the time series, but other stock characteristics may also affect how the stock will change with time. While controlling for all of these variables is unfeasible, it remains necessary to ensure that the selected uncertified stocks are appropriate controls for the treatment (certified) groups. We accomplished this using a propensity score matching approach. Propensity

scoring attempts to summarize the pre-treatment conditions of each stock into a single value, the propensity score (Rosenbaum 1983, Costello et al. 2008, Angrist et al 1998, Rubin 2008). This allows for the exclusion of unsuitable control groups whose pre-treatment conditions are dissimilar from those of the treatment group.

In this study, the propensity score is the probability that a stock was certified in the year that its corresponding treatment group was certified (5 years after the beginning of the times series). This probability was calculated by linear regression, where the dependent variable was certification status, and the independent variables were covariates that may have affected certification status, and therefore could be potentially confounding. One of these covariates is stock status at the beginning of the time series, which represents the stock's initial sustainability, its need to improve in stock status, and its likelihood of obtaining certification. Another important factor is the geographic location of the stock's primary market, which reflects the economic motivation to seek certification. The greatest demand for the MSC ecolabel is in northern Europe; fisheries with markets in this region are more likely to be certified. These two covariates, initial stock status and primary market location, were thus used to determine each stock's propensity score. Other covariates considered for inclusion in the propensity score model were standardized SSB at the beginning of the time series, stock size at the beginning of the time series, rebuilding status (whether a stock was in a rebuilding program during the time series), fishing gear type (fixed or mobile), and variables for primary market location in North America, South America, and Asia. These variables were excluded from the final model by a stepwise algorithm, due to unavailability of data, or due to a lack of representative stocks.

Once propensity scores were calculated, a range of propensity scores centered on the propensity score(s) of the certified stock(s) of each species was defined. A control stock was excluded if its score fell outside the propensity score range of the certified stock(s) it would be compared to. This propensity score methodology was deemed unnecessary for the rock lobster stocks due to their intrinsic similarity in both their management structures and geographic location. For a detailed description of propensity score calculation and selection of control stocks, readers are referred to Appendix C.

Data Normalization

A significant problem encountered in this analysis was normalization of the stock indicators, fishing mortality rate (F) and standardized spawning stock biomass (SSB). Because stocks vary in size and management goals, stock indicators are more meaningful when standardized by a reference point specific to each stock, and different types of reference points can strongly alter observed trends and conclusions about the stock's status. Several intergovernmental organizations and individual countries require that fish stocks be maintained at a level that supports maximum sustainable yield (MSY) (UNCLOS 1982). This biomass (B_{MSY}) and the corresponding fishing mortality rate (F_{MSY}) can then be used as targets for the fishery. In keeping with this international convention, the MSC has adopted these reference points in its standards of sustainability. We thus normalized fishing mortality rates in this analysis as:

$$\text{Standardized F} = \frac{F - F_{MSY}}{F_{MSY}}$$

When F is equal to F_{MSY} , the standardized value is 0. This approach was feasible for fishing mortality rates, since estimates of F_{MSY} were available for nearly all selected stocks. However, there was a much wider diversity of SSB reference points among these stocks, and in many cases, SSB_{MSY} was not defined. For many stocks, no biomass reference points were defined at all. We therefore performed this part of the analysis on the natural log of SSB, obviating the need for any reference point. This normalization conveniently down-weighted magnitudes and emphasized trends, the focus of the study. One shortcoming of this technique is that, in eliminating magnitude it prevents us from determining whether a stock is sustainable based on SSB relative to a reference point. However, this information can still be gleaned from the values of F relative to F_{MSY} for each stock. If F_{MSY} was not available for a stock, the published management reference points were used (as for Japanese pollock, Irish Sea plaice, Southwest of Ireland plaice, and South Argentine hake, Table A1). If no F reference points were available, the time series average over the ten years included in this study was used as a reference point (as for Iceland plaice and all rock lobster stocks).

Linear Model

A linear model was used to fit the control stocks with a single overall control slope, and the certified stocks with a single treatment slope. The control and treatment group slopes were then compared using ANCOVA to determine if the certified stocks displayed a significantly different trend than the non-certified stocks. The linear model, for the i^{th} stock, where y is the stock status indicator, F or SSB, t is time (years, set to be zero in the year of certification), and C is certification status, is:

$$y_{t,i} = \beta_{1,i} + \beta_2 * t + \beta_3 * t * C + \eta_{t,i} \quad (\text{Equation 1})$$

Where $\eta_{t,i}$ is random error ($\varepsilon_{t,i} \sim N(0, \sigma^2)$). For some analyses, a model with autoregressive autocorrelated errors fit by generalized least squares was used in place of the generalized linear model because it was more strongly supported by AIC. In these cases, the error term, $\eta_{t,i}$ is composed of two sources of variation, random error ($\varepsilon_{t,i} \sim N(0, \sigma^2)$) and autocorrelated error ($\phi\eta_{t-1,i}$ where $0 \leq \phi \leq 1$). For both model specifications, the β_3 parameter describes how certification affects stock trends; if this parameter was statistically significant at a probability level of 5% or less, the certification effect was significant. A statistical significance at a probability level of 10% represented moderate significance, and 1% was considered highly significance.

To determine whether the linear model should be fit with autocorrelated residuals, the data was first fit without autocorrelation by ordinary least squares (Model 1, Table 1). The lag 1 autocorrelation coefficient for this model (ϕ_{LM}) was then determined to provide an estimation of the degree of autocorrelation in each case. The linear model with autocorrelated residuals (Model 2) fit by generalized least squares was then applied to the data. In some cases, the autocorrelation coefficient, ϕ_{GLS} , was estimated to be nearly one. In these cases (where $\phi_{GLS} \geq 0.9$), the estimate from Model 1 (ϕ_{LM}) was used and fixed in the autocorrelated model. In each case, the model that was most strongly supported by AIC (had a lower AIC value) was selected for the final analysis. The statistical power of each final analysis was calculated using the "onewayanova" option in SAS's power procedure.

Meta-analysis

To estimate the overall effect of certification on stock status trends among all taxonomic groups, a meta-analysis was conducted treating each taxonomic group as a

study. Meta-analyses are used to consolidate results from studies that estimate the same phenomenon, giving stronger weight to studies that have more reliable estimates. In this study, each taxonomic group analysis estimated the effect of certification on stock status trends, so each group is thus treated as a study. While most meta-analyses attempt to pool results from studies by many researchers executed with various methods, in this meta-analysis each "study" was performed by the same researchers with an identical methodology, which renders the consolidation of results more reliable. In each group, the standardized mean difference (SMD) between slopes for control and certified groups was determined using the Hedge's adjusted g (the difference between slopes divided by a pooled standard deviation). These SMD's were combined across taxonomic groups using DerSimonian and Laird random effects meta-analysis, which uses inverse variance to pool data (Cooper and Hedges 1994; DerSimonian and Laird 1986). The random effects model anticipates that the true effect of certification varies from taxonomic group to taxonomic group, as opposed to assuming that there is a single true effect. Thus, it accounts for two sources of variance: within group variance and among group variance. This model is appealing because MSC certification is likely to have various effects on different species, depending on characteristics such as life history traits, geographic range, etc. Statistical heterogeneity among taxonomic groups was assessed by inspection of confidence intervals on each group's SMD as well as with a chi-squared test.

RESULTS

Stock Selection

The propensity score model indicates that, as predicted, a stock was significantly more likely to be certified in a given year if it had a lower standardized fishing mortality rate in that year (coefficient = -0.22, $p = 0.004$), and if its primary market was in Europe (coefficient = -0.41, $p = 0.04$). These covariates explained a significant proportion of variation in certification status ($R^2 = 0.26$, $p = 0.006$). Propensity scores ranged from 0.44 to 0.64, with scores for the certified and control stocks overlapping from 0.54 to 0.63 (Fig. 2). To obtain balance between the treatment and control distributions, a total of four stocks (Irish Sea herring, Northern European hake, North Argentine hake, and South Argentine hake) were excluded from the analysis (striped region in Fig. 2). Only two taxonomic groups, herring and hake, were affected by stock exclusions based on propensity scores. The reader is referred to Appendix C for a detailed description of the propensity score model results and stock selection.

Model Selection

In most cases, the linear model without autocorrelated residuals (Model 1) was more strongly supported by Akaike's Information Criterion (AIC) than the autocorrelated model (Table 1). The exceptions to this were hake (SSB), saithe (SSB), pollock (SSB), and rock lobster (SSB). Model 1 (which did not account for autocorrelated residuals) had autocorrelation coefficients (ϕ_{LM}) significantly different from zero ($p = 0.01$) in all but three cases (Hake SSB, Saithe SSB, and Pollock SSB). Model 2 resulted in autocorrelation coefficient estimates (ϕ_{GLS}) greater than or equal

to 0.9 in six analyses, necessitating the use of ϕ_{LM} as a fixed autocorrelation coefficient in the model. The failure to account for autocorrelation in time-series analyses can result in underestimated standard errors of coefficient estimates and consequently, overestimated significance levels associated with those coefficients. However, in the current analyses, both specifications of the model ultimately led to the same conclusion in all but one case, Pollock SSB. While Model 1 estimated a certification effect that was significant at the 10% level, Model 2 indicated that this effect was insignificant. Ultimately Model 2 was supported by AIC, so the insignificant certification effect was taken as a final result. It was critical to select models that were appropriate for the data, and in this study the greatest concern was that the model without autocorrelated residuals may have lead to false conclusions of significance. However, this concern was unfounded here because in every case where the model without autocorrelated residuals was selected, the model with autocorrelation corroborated its results.

Taxonomic Groups Analyses

The majority of the 12 analyses indicate that MSC certification did not have a significant effect on trends in stock status (Table 2). This was not the case for 5 analyses: plaice (SSB), saithe (SSB and F), pollock (F), and rock lobster (F). Among plaice stocks, certification had a significant positive effect on stock status trends. However for the other taxonomic groups, the opposite effect was observed; uncertified stocks improved in stock status significantly faster than certified ones.

Plaice stocks showed an overall increase in spawning stock biomass throughout the 8-year time-window, with certified stocks increasing significantly

faster than uncertified stocks (Table 2, Fig. 3A). No significant difference was observed in fishing mortality rate trends between control and treatment groups. Of the four stocks in the plaice analysis, only one stock, Kattegat plaice (a control) remained above its F_{MSY} reference point for the entire time series. The observed data for the control stock in the southwest of Ireland exceeded F_{MSY} for most of the time window, dipped below this threshold in 2010 and 2011, and increased above F_{MSY} levels again in 2012. The predicted linear fit for this stock was never below F_{MSY} . The observed fishing mortality rates for the remaining stocks were below F_{MSY} within a year of certification (2009).

In the saithe group, certification had a negative additive effect on trends in SSB and a positive effect on F , indicating that uncertified stocks are performing significantly better than certified stocks (Table 2, Figure 3B). Certified saithe stocks decreased in SSB and increased in F , while uncertified stocks decreased only slightly in SSB and decreased relatively rapidly in F (Table 2). While the certified stocks were both below their F_{MSY} reference level at the beginning of the time series (2003), both stocks exceeded these thresholds within a few years. The control stocks of this taxonomic group began the time series above their F_{MSY} reference points. While the Icelandic control stock steadily decreased in F , never exceeding its reference point after 2010, the F for Faroe Islands control stock increased sharply throughout the time window. Together, these stocks produce an overall control slope of nearly zero (Table 2). These observations contrast with the certified stocks, which both increase over the time series, producing a relatively sharp upward overall certified trend in F .

Among pollock stocks, there was no significant difference between certified and control groups in spawning stock biomass trends, but certification had a significant effect on trends in fishing mortality rate. While both the control and treatment group were found to decrease in F , control stocks decreased much faster than the treatment group, whose slope was relatively close to zero (Table 2, Fig. 3C). All three certified stocks remained well below their F_{MSY} reference levels for all ten years of the analysis, indicating a level of sustainability that may not have necessitated improvements in stock status. The control Japanese pollock stock also remained below F_{MSY} for the entire time series, and the observed data did not seem to decrease. Thus, the negative overall control slope is driven mostly by the Sea of Okhotsk stock, which displayed high fishing mortality rates at the beginning of the time series, but declined to levels of F below F_{MSY} by 2007.

The final taxonomic group that exhibited significant certification effects on stock trends was rock lobster, whose certified stock increased in fishing mortality rates significantly faster than uncertified stocks (Table 2, Fig. 3D). Note that for the rock lobster F analysis, there was no evidence of separate stock intercepts (at the year of certification) within either group. Each group, certified and uncertified, is thus plotted with a single line. Recall that no fishing mortality rate reference points were available for rock lobster stocks, so F was standardized by its average over the time series, which for each stock was about equal to the value in the middle of the time series. Thus, the observed standardized F data point for each stock was near zero at the year of certification, which guided the model to fit the time series with an intercept of nearly zero. In the spawning stock biomass analysis, the slope of the treatment group

was slightly positive, and the control group's slightly negative. Neither trend was significantly different from zero. The magnitude of $\ln(SSB)$ reflects that the control stocks of this taxonomic groups were larger than the certified stocks.

Among taxonomic groups for which certification was not found to play a significant role in stock status trends (herring and hake, Fig. 3E-F), all certified stocks were below their F_{MSY} reference point at the beginning of the analysis. This was not true for control stocks, whose fishing mortality rates were usually above F_{MSY} and higher than certified stocks in the same taxonomic group at the beginning of the analysis. In the hake taxonomic group, one control stock (the southern stock in the Northeast Atlantic) had higher levels of standardized fishing mortality rates than all other stocks. Nonetheless, this stock was deemed comparable to the certified stock based on its propensity score. If a narrower range of acceptable propensity scores had been defined, this stock could have been excluded from the analysis. However, given that its trends in F mirror the trends of other control stocks, its exclusion likely would not change the results significantly. This observation is true for all taxonomic groups. The propensity scoring methods identified and excluded control stocks that were unacceptably dissimilar from certified ones, based on initial fishing mortality rate and primary market location. Our comparisons of certified and control stocks were valid because the propensity scoring process ensured that the same trends were expected from each group. In summary, certification had a significant positive effect on trends in stock status in only one of 12 analyses (plaice, SSB). In four cases (saithe SSB and F , pollock F , and rock lobster F) certification had a negative effect. The remaining

seven analyses suggested that certified and uncertified stocks show no significant difference in stock trends.

The statistical power of these analyses ranged from 0.050 (Hake SSB, Table 2) to 0.98 (Plaice SSB). Among analyses where a significant certification effect was observed, the power was as low as 0.19 (Rock Lobster F). When there was no evidence of a significant certification effect (the null hypothesis was not rejected), power ranged from 0.050 to 0.39. This suggests that the probability of type II error (not rejecting the null hypothesis when it should have been rejected) is 61% - 95%, which is well above the desired level of 20%. Thus, from both a research and a management perspective, it is important in these cases not to assume that there is no certification effect based solely on this study's failure to reject this null hypothesis (Peterman 1990). Nonetheless, the five cases for which a significant difference between groups was detected still provide firm evidence of MSC certification's effect on fish stock status trends.

Meta-Analysis

Figure 4 shows forest plots of the Standardized Mean Difference (SMD) of stock status slopes in certified stocks relative to uncertified stocks in each of six taxonomic groups. The meta-analysis of MSC certification's effect on spawning stock biomass suggested that in two of six taxonomic groups (plaice and pollock), certification has a positive effect on SSB slopes (Fig. 4). This confirmed the results of the individual species analyses. Effect sizes ranged from -0.7 in the saithe taxonomic group to 2.07 in the plaice group. When these individual studies were pooled, the overall effect of certification was insignificant (0.06; 95%CI -0.53 to 0.66; $p = 0.83$).

Certification had a negative effect on trends in fishing mortality rates in two of six taxonomic groups (hake and plaice). The estimated effect sizes ranged from -0.28 to 1.78. Overall, certification has a positive effect on trends in F, significant at the 10% level (estimate = 0.54; 95% CI -0.08 to 1.16; $p = 0.09$). For both the SSB and F meta-analyses, the I^2 values on the Q-tests for heterogeneity were highly significant (83.9%, and 82.9%, respectively). These values represent the proportion of variation among taxonomic groups that can be attributed to heterogeneity, as opposed to chance. In summary, certification was generally associated with overall decreases in stock status, characterized by decreases in SSB (though these decreases are not significant), and increases in F.

DISCUSSION

The goal of this study was to determine whether MSC certified stocks show greater improvements in stock status when compared to uncertified stocks with similar stock status and economic markets. In other words, we investigated whether certified stocks respond and recover more quickly than uncertified stocks when they require improvements in sustainability (are fished at rates greater than F_{MSY}). The comparison between certified and uncertified stocks allowed us to assess the ability of the MSC to drive necessary improvements in the stocks carrying its ecolabel. Results suggest that MSC certification does not drive significant improvements in fish stock status. In most analyses of stock status indicators within taxonomic groups, there was no detectable difference in stock trends between certified and uncertified stocks. In four cases, the certified stocks actually showed significantly less improvements than uncertified stocks. The meta-analysis of these results indicates that certification has a negative overall effect on trends in stock status.

Efforts were made in this study to identify a statistical model that most parsimoniously addressed the research question. Other candidate models included linear mixed effects models and models with nonlinear terms. While the linear models ultimately selected and described are less complex than these, all successful models led to the same overall conclusion that the MSC certification is not associated with significant improvements in stock status. Even from case to case, the results of each model were almost always in agreement. For a detailed description of the results of other successful models, we refer the reader to Appendix F.

It is important to consider the initial status of both control and certified stocks before assuming that either group has displayed poor performance. For example, certified pollock stocks remained well below their F_{MSY} reference points for the entire time series. Their static stock status relative to control stocks (which improved throughout the time window) thus does not necessarily indicate that they are less sustainably fished. To many fisheries management bodies, intergovernmental organizations, and individual countries, these stocks would still qualify as sustainable, regardless of stock status trends (UNCOLS 1982). To determine the additive effect of MSC certification, ideally the control and treatment stocks would be identical except for their certification status. This is obviously impossible in an observational study such as this one. However, propensity scores allowed us to control for the most important differences (initial stock status and primary economic market), mimicking a randomized control-treatment design. A larger sample size of stocks would have allowed us to be more selective in the choice of control stocks, ensuring even higher levels of similarity between treatment and control groups. This also may have allowed us to account for more sources of variation. Studies like this should thus be repeated as the availability of time series for both certified and uncertified stocks become more available and lengthen.

Roheim (2011) inquired whether there had been allowed enough time to demonstrate detectable improvements in MSC certified stocks. While the results of most taxonomic group analyses were null, several showed significant differences between control and certified groups. Certified plaice stocks, for example, showed significantly greater improvements in stock status than similar uncertified stocks. This

indicates that differences of a certain magnitude are indeed detectable. However, longer time series lead to increased statistical power, allowing for the detection of smaller effect sizes. While we may have detected some of the greatest observable effects of certification, others may have gone unobserved due to lack of data. Generally, the analyses with higher statistical power were the ones in which a significant effect was detected (Table 2). However, some analyses with null results had similar power to those in which a significant effect was detected (for instance, herring SSB). This indicates that an effect size of similar magnitude to the ones observed could have been detected in this group.

Ecological sustainability remains a vaguely defined term. While it generally refers to the endurance of a practice or ecosystem, there is no decided quantitative basis by which to measure it. Many MSC certified stocks may prove sustainable by one definition but not by another. Nonetheless, some certified stocks assessed in this study seem unsustainable by any definition. The certified saithe stocks, for example, were found to have decreased sharply in stock status while being fished above MSY levels for most of the time series. Moreover, these stocks performed significantly worse than uncertified stocks of the same species. This calls into question the MSC's ability to identify sustainable fisheries and to demand continued sustainability post-certification.

Froese and Proelss (2012) estimated that 31% of MSC certified stocks are continuously overfished, and therefore require improvement in stock status. It follows that if the MSC effectively enforces its policy that performance indicators found lacking must be improved, approximately one third of the certified stocks in this study

should be improving in status. However, only one certified stock (North Sea plaice) out of 13 showed significant increases in SSB. While this observed discrepancy between the proportion of certified stocks that require rebuilding and the proportion that are improving in status could be a result of this study's relatively small sample size, the certified stocks in this analysis that are in the greatest need of rebuilding (saithe, for instance) are not the ones showing improvement. Thus the problem may be within the MSC itself, either in the organization's ability to detect weaknesses in certified stocks, or in its capacity to drive improvement in stocks that have low stock status levels.

Gutierrez et al. (2012) found that in the long term (1970 to present), certified stocks have out-performed uncertified stocks, as measured by trends in biomass relative to B_{MSY} . However, the divergence between these two groups occurred in the 1980's, well before the MSC was operating, so the observed difference in performance between certified and uncertified stocks may not be attributable to MSC certification (Gutierrez et al. 2012). By comparing certified stocks to reference uncertified stocks over the period before and after certification, in this study we were able to observe the additive effect of certification on trends in stock status. When results were pooled across all taxonomic groups, certification had an insignificant overall effect on SSB slopes and was associated with higher F slopes, amounting to worse overall performance by certified stocks relative to control uncertified stocks. This observation contradicts the conclusions of Gutierrez et al., suggesting that certification not only fails to drive real improvements in stock status, but overall is associated with declines relative to control stocks. The differing conclusions of these studies may have arisen

from the length of the time series, the method of combining data from different stocks, or the statistical analysis of the time series.

Fisheries that invest in MSC certification expect it to benefit them economically, and may even expect it to help them conform to a sustainable standard. However, these results suggest that MSC certified stocks do not recover from overfished states any faster than uncertified stocks. Thus, if a non-ecolabeled fishery's goal is to increase the sustainability of its operations, these results might discourage it from seeking certification. On the other hand, if a fishery is already sufficiently sustainable and seeks to increase ex-vessel prices or to access new markets, MSC certification may still be economically advantageous. A valuable follow-up study might assess how certification affects economic trends in fisheries.

To attain its goal of contributing to the increased sustainability of the seafood industry, the MSC must accurately inform consumers of the sustainability of seafood products. This study indicates that the MSC ecolabel does not drive ecological improvements in certified stocks when compared to uncertified stocks that share similar stock status and economic markets. Moreover, some MSC certified stocks examined here decreased in status while being fished above sustainable levels. However, other studies suggest that the frequency of overfishing among MSC certified stocks is considerably lower than the global frequency of overfishing, which indicates that the MSC is capable of identifying and certifying sustainable seafood products (Froese and Proelss 2012, Gutierrez et al 2012, Martin et al 2012). Choosing MSC certified products thereby decreases the consumer's likelihood of supporting unsustainable fisheries, but may not guarantee the stock health of its labeled products.

We argue that, while the MSC conveys important and generally reliable information about seafood sustainability, the organization should make greater efforts to identify certified stocks in poor health or overfished states and ensure that they are making improvements in stock status.

REFERENCES

- Agnew, D., Grieve, C., Orr, P., Parkes, G., and Barker, N. 2006. Environmental benefits resulting from certification against MSC's Principles and Criteria for Sustainable Fishing. MRAG UK Ltd. and Marine Stewardship Council. London, UK.
- Angrist, J. D. 1998. Estimating the labor market impact of voluntary military service using social security data on military applicants. *Econometrica* **66**: 249-288.
- Australian Department of Agriculture. 1991. Fisheries Management Act. No. 162.
- Beddington, J.R., Agnew, D.J., and Clark, C.W. 2007. Current problems in the management of marine fisheries. *Science* **316**: 1713-1716.
- Christian, C., Ainley, D., Bailey, M., Dayton, P., Hocevar, J., LeVine, M., Nikoloyuk, J., Nouvian, C., Velarde, E., Werner, R.W., and Jacquet, J. 2013. A review of formal objections to Marine Stewardship Council fisheries certifications. *Biological Conservation* **161**: 10-17.
- Cooper, H., Hedges, L. V. 1994. Fixed effects models. *The handbook of research synthesis*: 285-299.
- Costello, C., Gaines, S. D., and Lynham, J. 2008. Can catch shares prevent fisheries collapse?. *Science* **321**: 1678-1681.
- DerSimonian, R., Laird, N. 1986. Meta-analysis in clinical trials. *Controlled clinical trials* **7**:177-188.
- Froese, R., and Proelß, A. 2012. Evaluation and legal assessment of certified seafood. *Marine Policy* **36**: 1284-1289.

- Goyert, W., Sagarin, R., and Annala, J. 2010. The promise and pitfalls of Marine Stewardship Council certification: Maine lobster as a case study. *Marine Policy* **34**: 1103-1109.
- Gutiérrez, N. L., Valencia, S. R., Branch, T. A., Agnew, D. J., Baum, J. K., Bianchi, P. L., ...and Williams, N. E. 2012. Ecolabel conveys reliable information on fish stock health to seafood consumers. *PloS one* **7**: e43765.
- Hall, S.J., and Mainprize, B.M. 2005. Managing by-catch and discards: how much progress are we making and how can we do better?. *Fish and Fisheries* **6**: 134-155.
- Jackson, J. B., Kirby, M. X., Berger, W. H., Bjorndal, K. A., Botsford, L. W., Bourque, B. J., ... and Warner, R. R. 2001. Historical overfishing and the recent collapse of coastal ecosystems. *Science* **293**: 629-637.
- Jacquet, J., Hocevar, J., Lai, S., Majluf, P., Pelletier, N., Pitcher, T., ... and Pauly, D. 2009. Conserving wild fish in a sea of market-based efforts. *Oryx* **44**: 45-56.
- Jacquet, J., Pauly, D., Ainley, D., Holt, S., Dayton, P., and Jackson, J. 2010. Seafood stewardship in crisis. *Nature* **467**: 28-29.
- Lotze, H. K., Lenihan, H.S., Bourque, B.J., Bradbury, R.H., Cooke, R.G., Kay, M.C., Kidwell, S.M., Kirby, M.X., Peterson, C.H., and Jackson, J.B.C. 2006. Depletion, degradation, and recovery potential of estuaries and coastal seas. *Science* **312**: 1806-1809.
- Martin, S. M., Cambridge, T. A., Grieve, C., Nimmo, F. M., and Agnew, D. J. 2012. An evaluation of environmental changes within fisheries involved in the

- Marine Stewardship Council certification scheme. *Reviews in Fisheries Science* **20**: 61-69.
- MSC. 2010. MSC Theory of Change. MSC, London. http://www.msc.org/documents/msc-brochures/msc-theory-of-change/at_download/file (last accessed December 2012).
- Marine Stewardship Council. 2013. Global Impacts Report.
- Peterman, R. M. 1990. Statistical power analysis can improve fisheries research and management. *Canadian Journal of Fisheries and Aquatic Sciences* **47**: 2-15.
- Pauly, D., Alder, J., Bennett, E., Christensen, V., Tyedmers, P., and Watson, R. 2003. The future for fisheries. *Science* **302**: 1359-1361.
- Roheim, C.A., Asche, F., and Santos, J.I. 2011. The elusive price premium for ecolabelled products: evidence from seafood in the UK market. *Journal of Agricultural Economics* **62**: 655-668.
- Rosenbaum, P.R., and Rubin, D. B. 1997. Estimating causal effects from large data sets using propensity scores. *Annals of internal medicine* **127**: 757-763.
- Rubin, D. B. 2008. For objective causal inference, design trumps analysis. *The Annals of Applied Statistics*: 808-840.
- Tlusty, Michael F. 2010. Environmental improvement of seafood through certification and ecolabelling: theory and analysis. *Fish and Fisheries* **13**: 1-13.
- UNCLOS (1982) United Nations Convention on the Law of the Sea. 1833: 1–186. Available at: <http://treaties.un.org/doc/Publication/UNTS/Volume%201833/volume-1833-A-31363-English.pdf>.

Wessells, Cathy Roheim, ed. Product certification and ecolabelling for fisheries sustainability. Vol. 422. FAO, 2001.

Worm, B., Hilborn, R., Baum, J. K., Branch, T. A., Collie, J. S., Costello, C., ... and Zeller, D. 2009. Rebuilding global fisheries. *Science* **325**: 578-585.

TABLES

Table 1. Summary table of model selection results. A dot (•) indicates that the autocorrelation coefficient estimated by the linear model without autocorrelated residuals, φ_{LM} , was significant at the 1% level. Parentheses around the autocorrelation coefficient associated with the GLS model, φ_{GLS} , indicate that the value was fixed at φ_{LM} due to convergence errors. One asterisk (*) indicates that the certification effect was significant at the 10% level, two asterisks at the 5% level, and three asterisks at the 1% level.

Taxonomic group	Model 1: GLM with No Autocorrelation				Model 2: GLS with autocorrelated residuals			
	Metric	AIC	Certification Effect	Φ_{LM}	AIC	Certification Effect	Φ_{GLS}	Model Selected
Herring	SSB	97	-0.034	0.31•	115	-0.016	0.57	1
Herring	F	61	-0.02	0.25•	93	-0.014	0.46	1
Hake	SSB	59	0.0094	0.33	54	0.0064	-0.33	2
Hake	F	7.6	-0.0049	0.44•	15.4	-0.008	0.76	1
Saithe	SSB	-14	-0.042**	0.57•	-17	-0.048**	-0.57	2
Saithe	F	11	0.098***	0.29	18	0.093**	0.82	1
Plaice	SSB	-29	0.11***	0.63•	-24	0.099**	0.89	1
Plaice	F	50	-0.044	0.44•	64	-0.068	0.8	1
Pollock	SSB	-14	-0.032*	0.65•	-22	-0.027	-0.65	2
Pollock	F	4.8	0.039*	0.34	5.6	0.050*	-0.34	1
Rock Lobster	SSB	-49	-0.0077	0.81•	-88	0.0051	-0.81	2
Rock Lobster	F	-77	0.048***	0.67•	-72	0.037***	-0.67	1

Table 2. Summary of model output for each taxonomic group. The number of observations in each analysis is provided in parentheses after the metric. The Certification effect refers to the estimate of the β_3 parameter in Eq 1. One asterisk (*) indicates significant effect of certification at the 10% level, two asterisks at the 5% level, and three asterisks at the 1% level. For models fit with autocorrelation (Model 2), the estimated phi value is provided in parentheses.

Taxonomic group	Metric	Certification Effect	Residual Standard Error	Model Type	Power
Herring	SSB (132)	-0.034	0.33	1	0.39
	F (132)	-0.020	0.29	1	0.11
Hake	SSB (55)	0.0064	0.39	2 (0.33)	0.050
	F (55)	-0.0049	0.24	1	0.054
Saithe	SSB (42)	-0.048**	0.16	2 (0.57)	0.33
	F (42)	0.098***	0.25	1	0.73
Plaice	SSB (44)	0.11***	0.16	1	0.98
	F (44)	-0.044	0.43	1	0.089
Pollock	SSB (55)	-0.027	0.18	2 (0.65)	0.11
	F (55)	0.039*	0.23	1	0.22
Rock Lobster	SSB (66)	0.0051	0.15	2 (0.80)	0.053
	F (44)	0.048***	0.09	1	0.19

FIGURES

Figure 1. Global distribution of control (ovals) and treatment (stars) fish stocks



Figure 2. Histogram of certified and control propensity scores. Striped region indicates stocks that were eventually excluded based on propensity score.

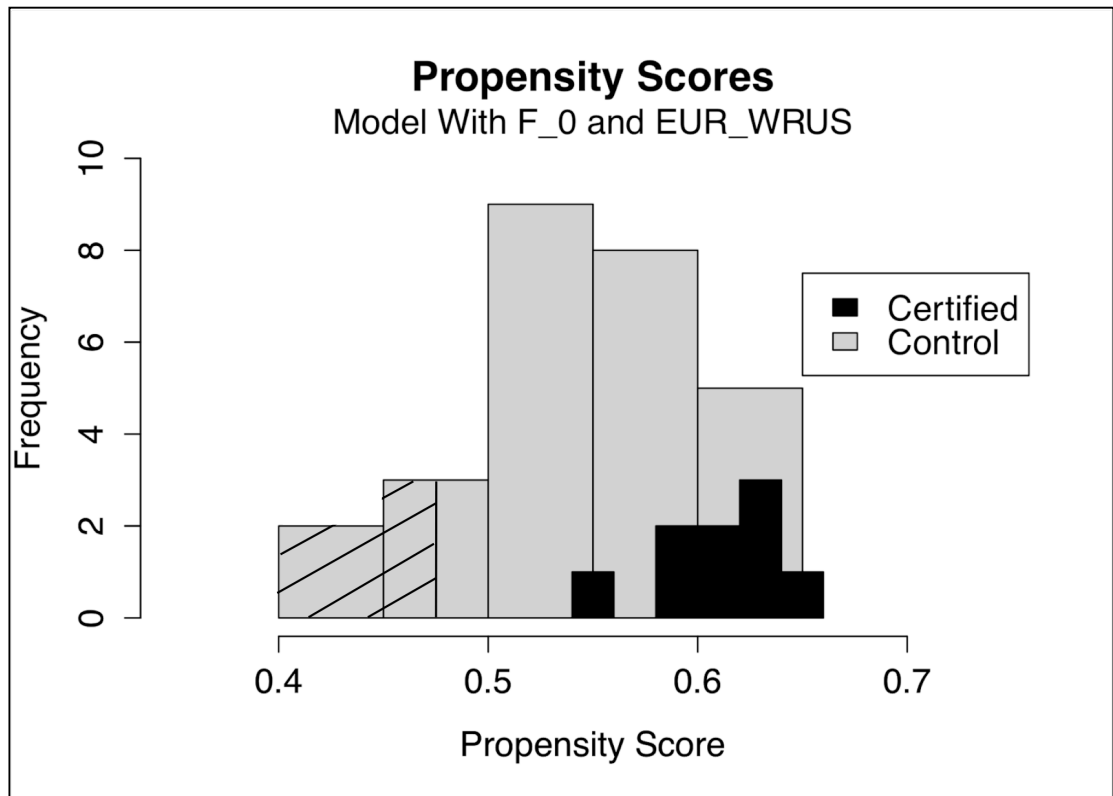
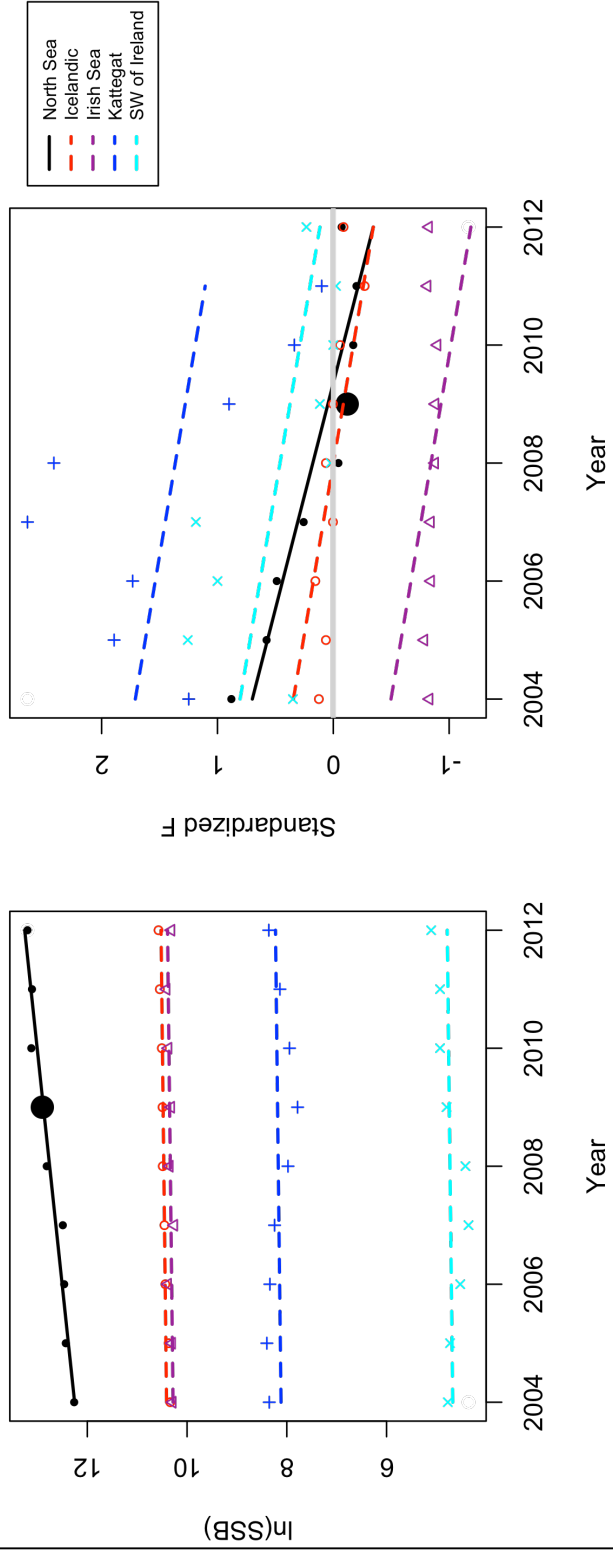


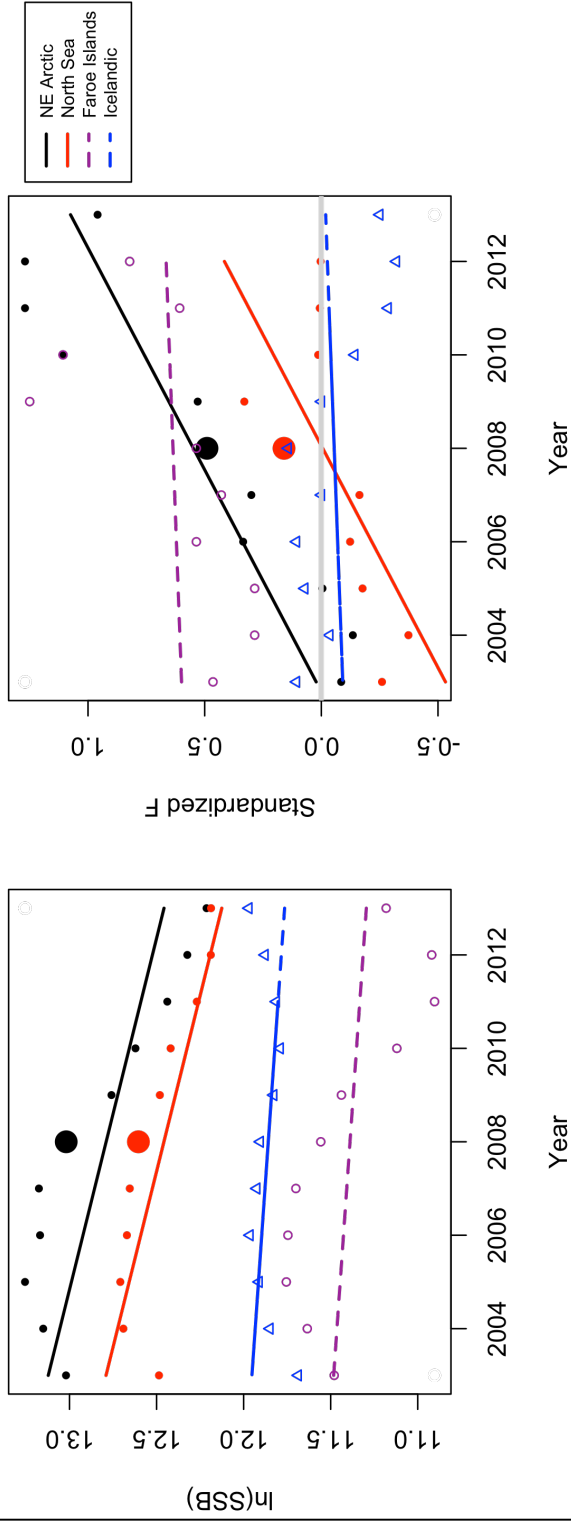
Figure 3. Observed data (small points) and fitted trends (lines) in spawning stock biomass (left) and fishing mortality rates (right) for certified (solid) and uncertified (dashed) stocks of each taxonomic groups. Taxonomic groups in which the effect of certification was significant are marked with an asterisk (*). Large circles represent the observed data point of certified stocks in the year they were certified.

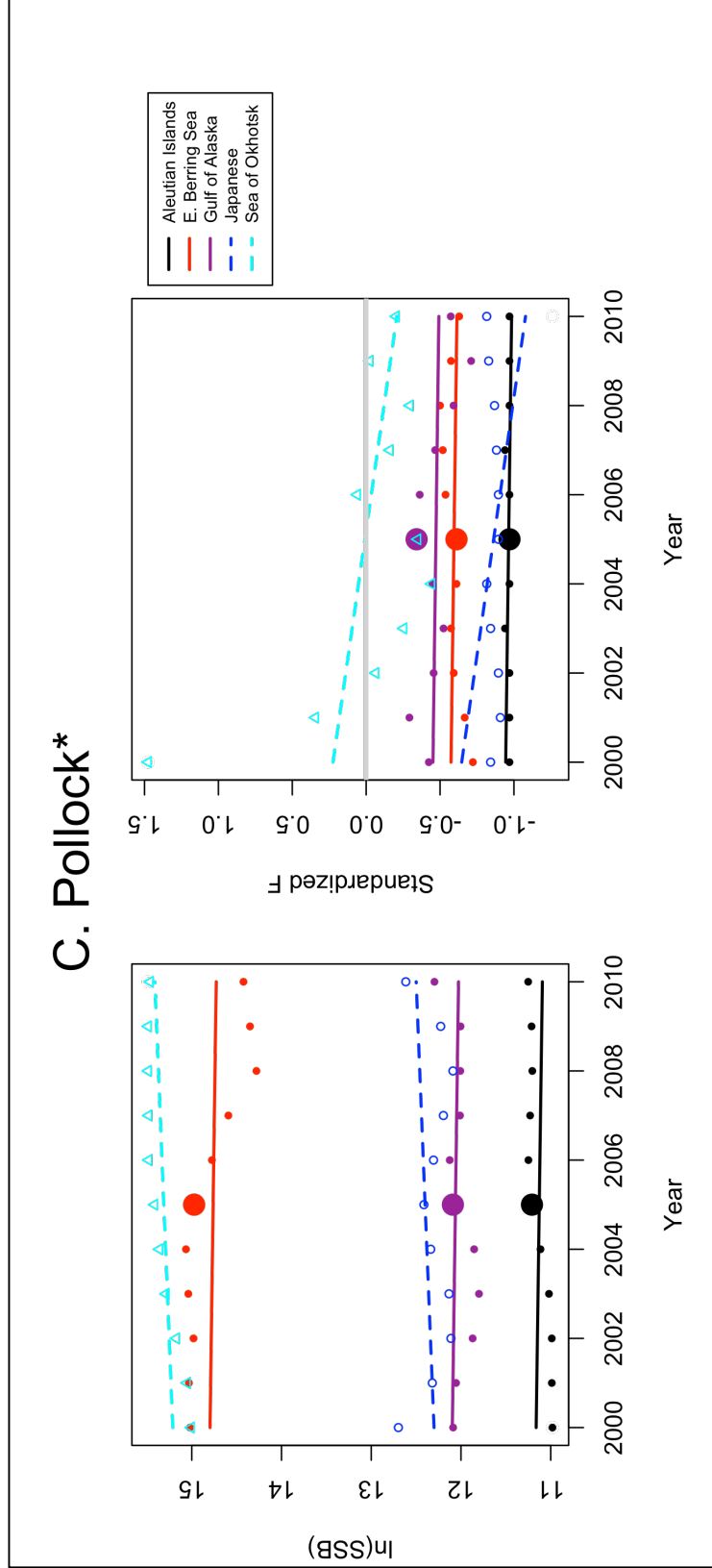
† The certified Western Australian rock lobster stock is divided into 4 egg production areas (Northern, Southern, Central, and Abrulhas) for the assessment of SSB, and only 2 areas (Northern and Southern) for the estimation of harvest rate.

A. Plaice*

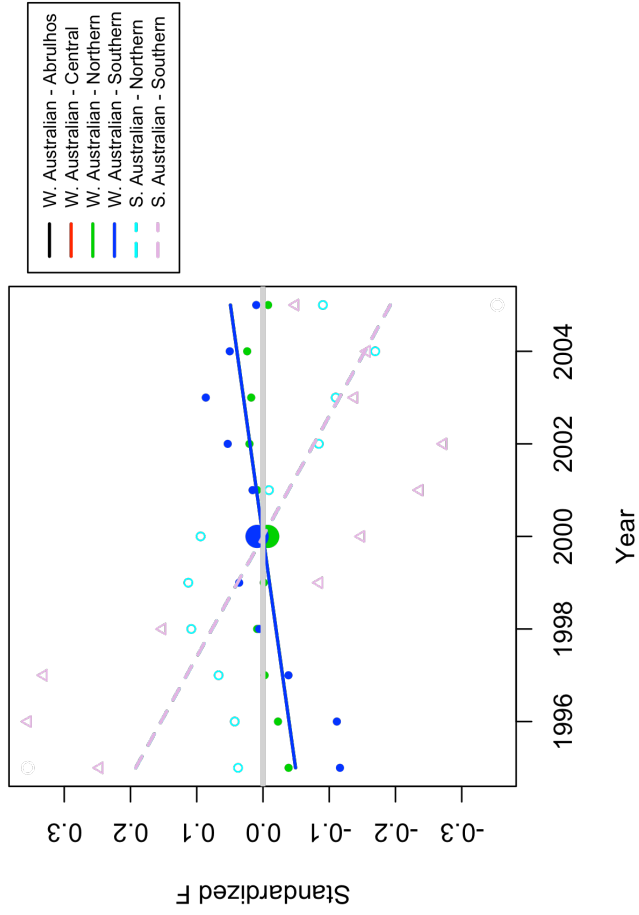
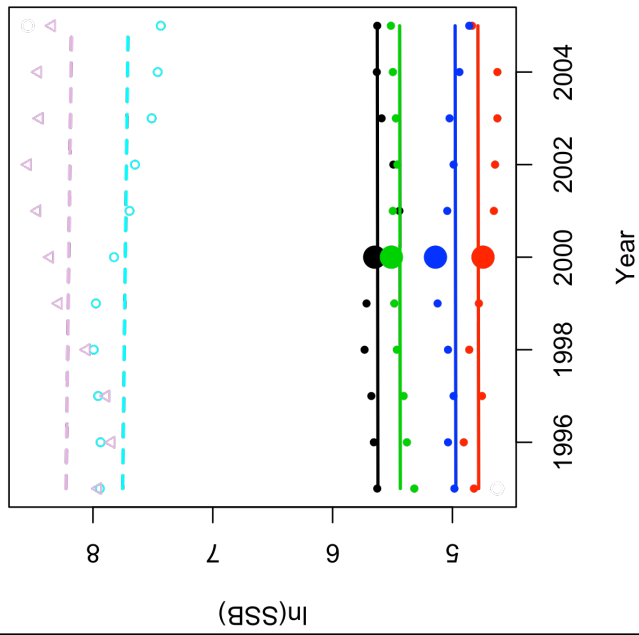


B. Saithe*

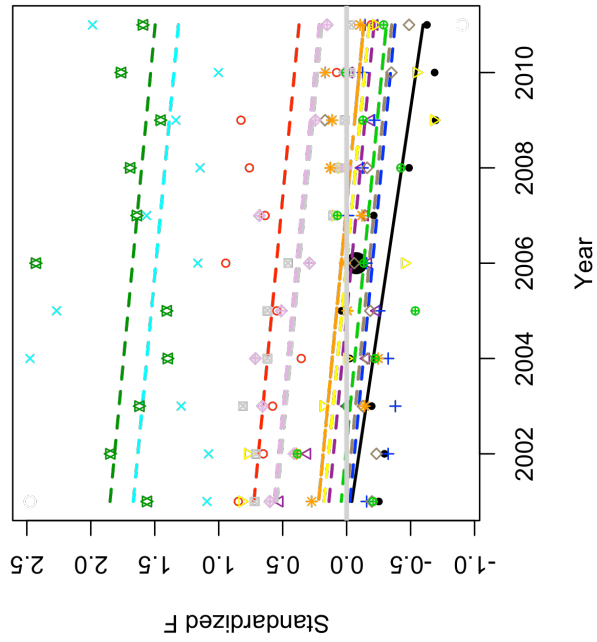
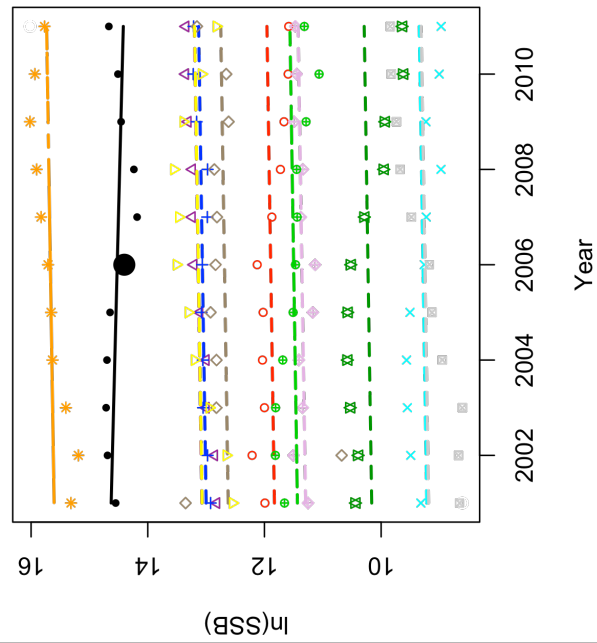




D. Rock Lobster*



E. Herring



- North Sea
- - Baltic Spring Spawners
- - E. Baltic
- - Bothnian Sea (area 30)
- - Bothnian Sea (area 31)
- - Gulf of Maine/Georges Bank
- - Icelandic
- - Irish Sea
- - Norwegian
- - Gulf of Riga
- - West of Scotland
- - Celtic Seas

F. Hake

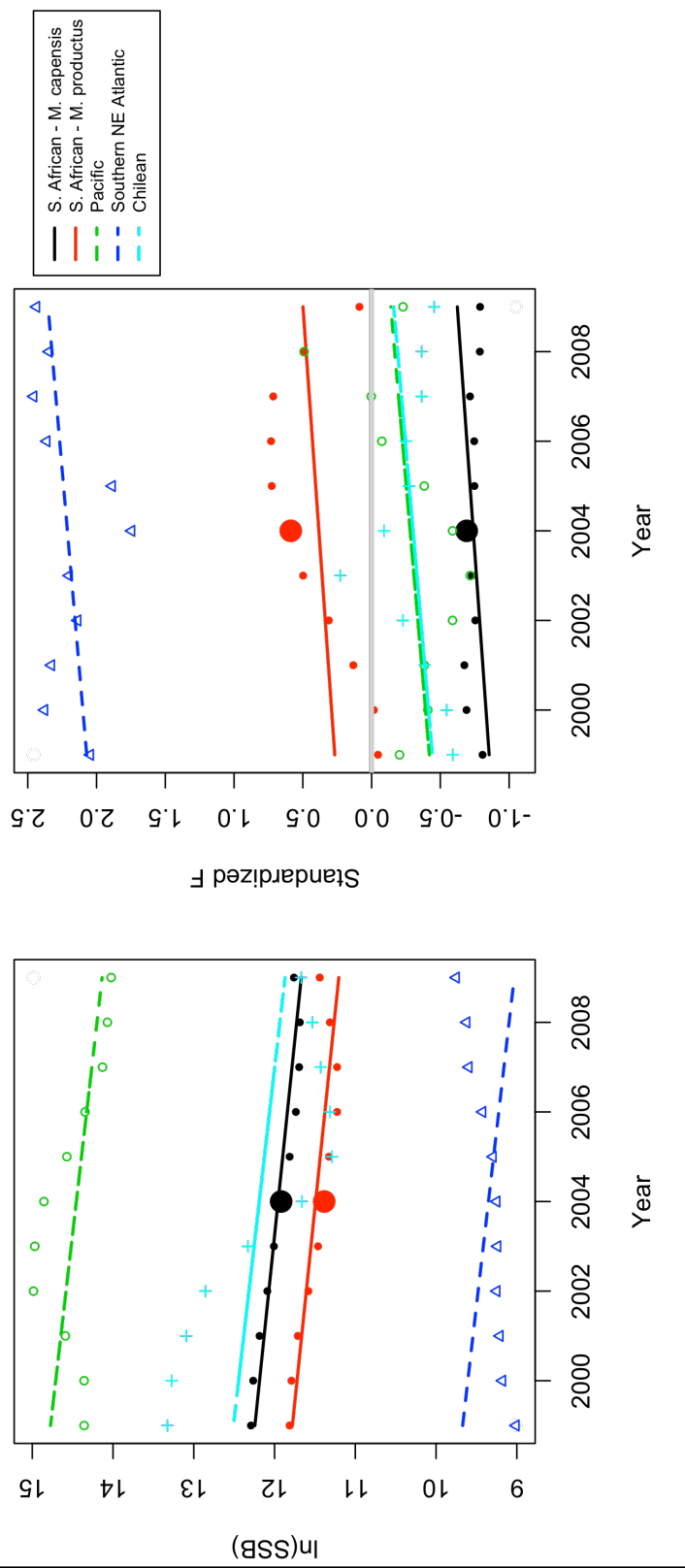
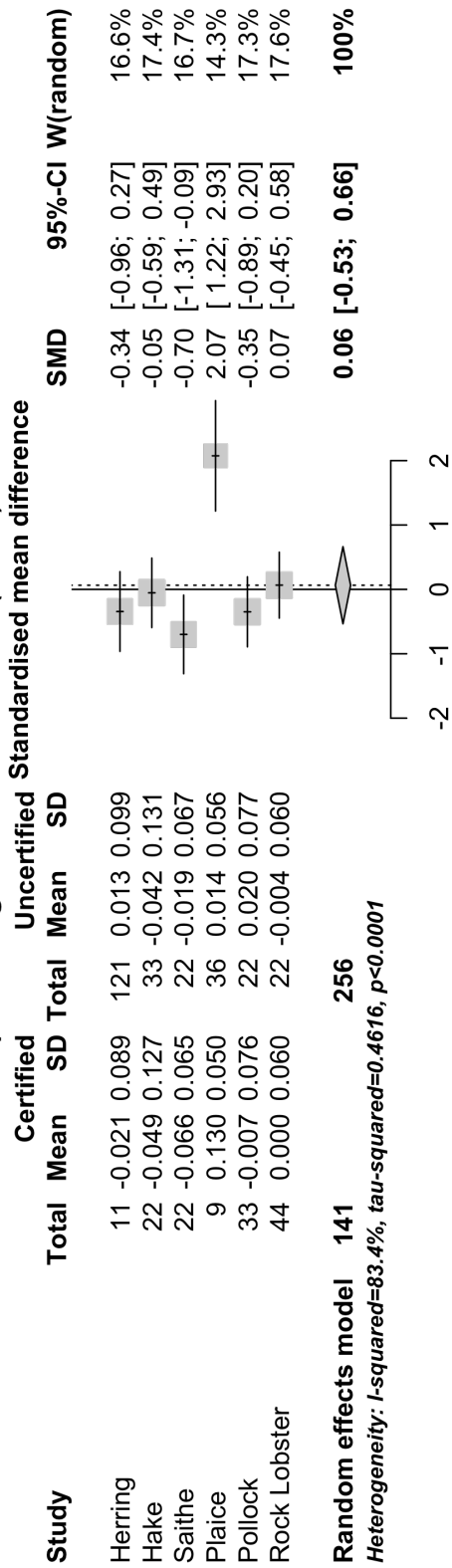
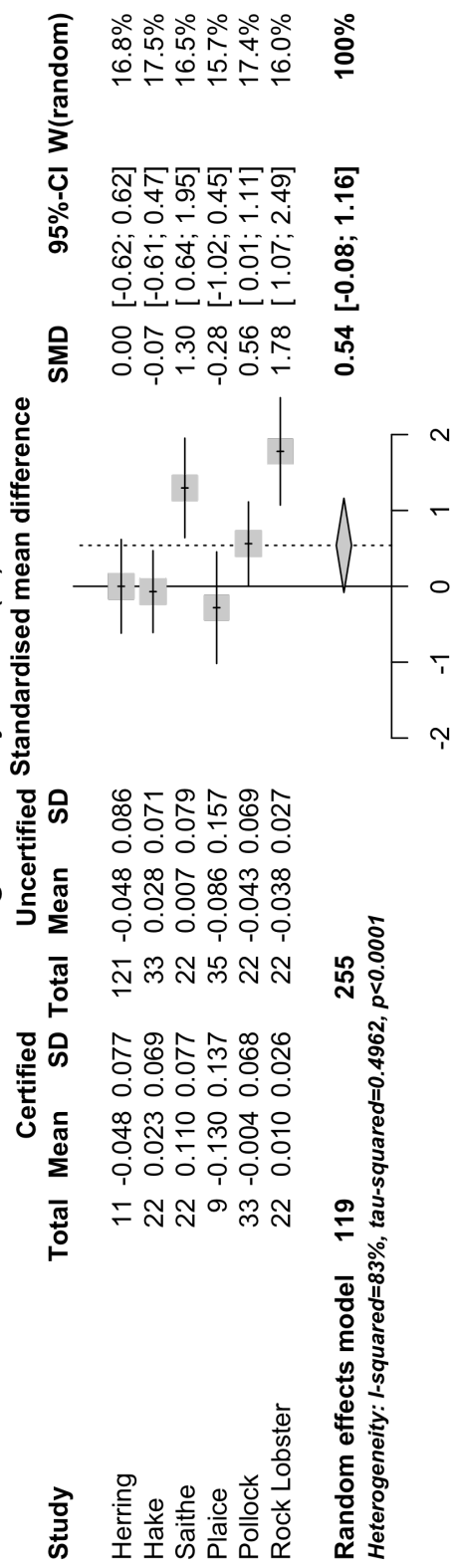


Figure 4. Forest plot for meta-analysis of certification effects on trends in SSB and F, along with input data for sample size ("Total"), estimated slope ("Mean"), and standard deviation ("SD") for certified and uncertified taxonomic groups. Gray boxes represent taxonomic group effect sizes ("SMD") with 95% confidence intervals. Parallelograms represent pooled effect sizes. The weight of each taxonomic group is reported as "W(random)."

Spawning Stock Biomass (SSB)



Fishing Mortality Rate (F)



APPENDICES

APPENDIX A. STOCK INPUT INFORMATION

Table A1. Summary of input data for each stock by taxonomic group. Certified stocks are shaded and sources refer to Appendix B.

Taxonomic group	Stock (species name)	Year of Certification	Times series used	F _{ref} type	Notes	Source
Herring	North Sea Autumn Spawners (<i>Clupea harengus</i>)	2006	2001 - 2011	F _{MSY}	Certified fishery is the Pelagic Freezer-trawler association.	B1
	Gulf of Maine and George's Bank Herring (<i>Clupea harengus</i>)		2001 - 2011	F _{MSY}		B2
	West of Scotland - ICES Division VIa North (<i>Clupea harengus</i>)	2012	2001 - 2011	F _{MSY}		B1
	West of Ireland and Porcupine Bank - ICES Division VIa South and VIIbc (<i>Clupea harengus</i>)		2001 - 2011	F _{MSY}		B1
	Celtic Sea and South of Ireland - ICES division VIIa South, and VIIghjk (<i>Clupea harengus</i>)	2012	2001 - 2011	F _{MSY}		B1
	Irish Sea (<i>Clupea harengus</i>)		2001 - 2011	F _{MSY}	Stock entered MSC full assessment in 2013. Excluded by propensity score.	B1
	Gulf of Riga (<i>Clupea harengus</i>)		2001 - 2011	F _{MSY}		B3
	Norwegian Spring Spawners (<i>Clupea harengus</i>)	2009	2001 - 2011	F _{MSY}		B4
	Bothnian Sea - Subdivision 30 (<i>Clupea harengus</i>)		2001 - 2011	F _{MSY}		B3

	Bothnian Sea - Subdivision 31 (<i>Clupea harengus</i>)		2001 - 2011	F_{MSY}		B3
	Baltic Sea Spring Spawners - ICES Subdivisions 22-24 (<i>Clupea harengus</i>)		2001 - 2011	F_{MSY}		B6
	Iceland Grounds (<i>Clupea harengus</i>)		2001 - 2011	F_{MSY}		B5
	Eastern Baltic Herring - ICES are 25-32 (<i>Clupea harengus</i>)		2001 - 2011	F_{MSY}		B3
Hake	South African (<i>Merluccius capensis</i>)	2004	1999 - 2009	F_{MSY}	F_{MSY} estimated as MSY/B_{MSY} . Both South African stocks were certified together as a single certification unit.	B30
	South African (<i>Merluccius paradoxus</i>)	2004	1999 - 2009	F_{MSY}	F_{MSY} estimated as MSY/B_{MSY}	B30
	Northern Stock in the Northeast Atlantic (<i>Merluccius merluccius</i>)		1999 - 2009	F_{MSY}	Excluded by propensity score	B7
	Southern Stock in the Northeast Atlantic (<i>Merluccius merluccius</i>)		1999 - 2009	F_{MSY}		B7
	Pacific Hake (<i>Merluccius productus</i>)	2009	1999 - 2009	F_{MSY}		B8
	South Argentine Hake (<i>Merluccius hubbsi</i>)		1999 - 2009	F_{sub}	Excluded by propensity score	B11
	North Argentine Hake (<i>Merluccius hubbsi</i>)		1999 - 2009	$F_{0.1} \sim F_{MSY}$	Excluded by propensity score	B9 B10
	Chilean Hake (<i>Merluccius gavi gavi</i>)		1999 - 2009	F_{MSY}		B12 B13

Plaice	North Sea (<i>Pleuronectes platessa</i>)	2009	2004 - 2012	F_{MSY}	Certified fishery is the Ekofish group twin-rigged otter trawl. Three other plaice fisheries in the North Sea have been certified since 2009. Time series shortened due to data availability	B23
	Iceland (<i>Pleuronectes platessa</i>)		2004 - 2012	--	F reference point is time series average from 2004 - 2012.	B25 B26
	Irish Sea (<i>Pleuronectes platessa</i>)		2004 - 2012	$F_{0.95}$		B24
	Southwest of Ireland (<i>Pleuronectes platessa</i>)		2004 - 2012	$F_{0.95}$		B24
	Kattegat and Belts Sound (<i>Pleuronectes platessa</i>)		2004 - 2012	F_{MSY}		B3
Saithe	Northeast Arctic (<i>Pollachius virens</i>)	2008	2003 - 2013	F_{MSY}	The Norway and Germany fisheries are certified. 2013 values of F and SSB are predicted estimates from a 2013 assessment.	B19
	North Sea - ICES subareas IV, VI, and Division IIIa (<i>Pollachius virens</i>)	2008	2003 - 2013	F_{MSY}	Only the Norway fishery is certified. 2013 SSB value is a predicted estimate from 2013 assessment. No estimate for F in 2013.	B21
	Faroe (<i>Pollachius virens</i>)	2013	F: 2003 - 2012 SSB: 2003 - 2013	F_{MSY}	No estimate for F in 2013.	B20
	Icelandic (<i>Pollachius virens</i>)		2003 - 2012	F_{MSY}	Entered full MSC assessment in 2013	B20
Pollock	Easem Berring Sea (<i>Theragra chalcogramma</i>)	2005	2000 - 2010	$F_{OFL} \sim F_{MSY}$		B14
	Alcutian Islands (<i>Theragra</i>)	2005	2000 - 2010	$F_{OFL} \sim F_{MSY}$		B15

	<i>chalcogramma</i>) Gulf of Alaska (<i>Theragra</i> <i>chalcogramma</i>)	2005	2000 - 2010	$F_{35\%} \sim F_{MSY}$	F_{MSY} is time-varying.	B16
	Sea of Okhotsk (<i>Theragra</i> <i>chalcogramma</i>)	2013	2000 - 2010	F_{MSY}	Time series data extracted from figure using plot digitizer.	B17
	Japanese (<i>Theragra</i> <i>chalcogramma</i>)		2000 - 2010	F_{exp}		B18
Rock Lobster	Western Australia Abrolhos (<i>Panulirus cygnus</i>)	2000	1995-2005	--	Egg production index is used in place of SSB. There is no estimated F times series for this stock.	B27
	Western Australia Northern (<i>Panulirus cygnus</i>)	2000	1995-2005	--	Egg production index is used in place of SSB. F reference point is time series average from 1995 - 2005.	B27
	Western Australia Central (<i>Panulirus cygnus</i>)	2000	1995-2005	--	Egg production index is used in place of SSB. There is no estimated F times series for this stock.	B27
	Western Australia Southern (<i>Panulirus cygnus</i>)	2000	1995-2005	--	Egg production index is used in place of SSB. F reference point is time series average from 1995 - 2005.	B27
	Southern Australian Northern Zone (<i>Jasus edwardsii</i>)		1995-2005	--	F reference point is time series average from 1995 - 2005.	B28 B29
	Southern Australian Southern Zone (<i>Jasus edwardsii</i>)		1995-2005	--	F reference point is time series average from 1995 - 2005.	B28 B29

APPENDIX B. STOCK DATA SOURCES

- B1. ICES. 2012. Report of the Herring Assessment Working Group for the Area South of 62 N (HAWG). ICES CM 2012/ACOM:06.
- B2. Northeast Fisheries Science Center. 2012. Stock Assessment of Atlantic Herring - Gulf of Maine/George's Bank. 54th Northeast Regional Stock Assessment Workshop (54th SAW) Assessment Summary Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc.
- B3. ICES. 2012. Report of the Baltic Fisheries Assessment Working Group (WGBFAS). ICES CM 2012/ACOM:10.
- B4. ICES. 2013. Report of the Working Group on Widely Distributed Stocks (WGWIDE). ICES CM 2013/ACOM:15.
- B5. ICES. 2012. Report of the North-Western Working Group. ICES CM 2012/ACOM:07.
- B6. ICES. 2012. Report of the ICES Advisory Committee: Herring in Division IIIa and Subdivisions 22-24 (western Baltic spring spawners). ICES Advice. Book 6.
- B7. ICES. 2012. Report of the Working Group on the Assessment of Southern Shelf Stocks of Hake, Monk and Megrin (WGHMM). ICES CM 2012/ACOM:11. 617 pp.
- B8. Martell, S. J. D. 2010. Assessment and management advice for Pacific hake in U.S. and Canadian waters in 2010. Pacific Fishery Management Council. Portland, Oregon.

- B9. Bezzi, S.; Cañete, G.; Pérez, M.; Renzi, M. and Lassen, H., 1994. Report of the INIDEP working group on assessment of hake (*Merluccius hubbsi*) North of 48° S (Southwest Atlantic Ocean). INIDEP Scientific Document No. 3.
- B10. Irusta, C.G., and D'Atri, L.L. 2012. Evaluacion del estado del efectivo norte de 41° S de la Merluza (*Merluccius hubbsi*) y estimación de la captura biológicamente aceptable para el año 2013. Instituto Nacional de Investigación Desarrollo Pesquero (INIDEP).
- B11. Villarino, M.F., Santos, B.A., and Renzi, M.A. 2012. Evaluacion del estado del efectivo sur de 41° S de la Merluza (*Merluccius hubbsi*) y estimación de la captura biológicamente aceptable para 2013. Instituto Nacional de Investigación Desarrollo Pesquero (INIDEP).
- B12. Subsecretaría de Pesca (Subpesca). 2011. Cuota global anual de captura de merluza común (*Merluccius gayi gayi*), año 2012. Inf. Tec. (R.Pesq.) N° 117/2011
- B13. Instituto De Fomento Pesquero (IFOP). Estatus y posibilidades de explotación biologicamente sostenables de los principales recursos pesqueros nacionales año 2012: Merluza Común, 2012. Subpesca.
- B14. Ianelli, J.N., Honkalehto, T., Barbeaux, S., Kotwicky, S., Aydin, K., and Williamson, N. 2012. Assessment of the walleye pollock stock in the Eastern Bering Sea. *In* Stock assessment and fishery evaluation report for the groundfish resources of the Bering Sea/Aleutian Islands Regions. Nrth. Pac. Mgmt. Coun.

- B15. Barbeaux, S., Ianelli, J., and Palsson, W. 2012. Assessment of the Pollock stock in the Aleutian Islands. *In* Stock assessment and fishery evaluation report for the groundfish resources of the Bering Sea/Aleutian Islands Region. Nrth. Pac. Mgmt. Coun.
- B16. Dorn, M., Aydin, K., Barbeaux, S., Darin, J., and Palsson, W., 2012. Assessment of the Walleye Pollock Stock in the Gulf of Alaska. *In* Stock assessment and fishery evaluation report for the groundfish resources of the Gulf of Alaska. Nrth. Pac. Mgmt. Coun.
- B17. O'Boyle, R., Japp, D., Payne, A., and Devitt, S. 2013. Russian Sea of Okhotsk Mid-water Trawl Walleye Pollock (*Theragra chalcogramma*) Public Certification Report. Intertek Moody Marine.
- B18. Fisheries Research Agency of Japan (FRAJ). 2012. Summary Resource Assessment of Alaska Pollock in the Pacific Systems Group, Hokkaido National Fisheries Research Institute.
- B19. ICES. 2013. Report of the Arctic Fisheries Working Group (AFWG). ICES CM 2013/ACOM:05.
- B20. ICES, 2012. Report of the North-Western Working Group (NWWG). ICES CM 2012/ACOM:07. 1425 pp.
- B21. ICES. 2012. Report of the Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak (WGNSSK). ICES CM 2012/ACOM:13.

- B22. Northeast Fisheries Science Center. 2010. 50th Northeast Regional Stock Assessment Workshop (50th SAW) Assessment Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 10-17.
- B23. ICES. 2013. Plaice in Subarea IV (North Sea). ICES advice 2013. Book 6, 6.4.18.
- B24. ICES. 2013. Report of the Working Group for Celtic Seas Ecoregion (WGCSE). ICES CM 2013/ACOM.
- B25. Marine Research Institute. 2012. State of stocks 2011/2012, Prospects 2012/2013.
- B26. Fish Source. <http://www.fishsource.com>.
- B27. West Coast Rock Lobster Harvest Strategy and Control Rules 2015 - 2019. 2013. Government of Western Australia Department of Fisheries. Fisheries Management Paper no. 263.
- B28. Linnane, A., McGarvey, R., Feenstra, J. and Hoare, M. 2013. Northern Zone Rock Lobster (*Jasus edwardsii*) Fishery Status Report 2012/13. Status Report to PIRSA Fisheries and Aquaculture. South Australian Research and Development Institute (Aquatic Sciences), Adelaide. No. 750.
- B29. Linnane, A., McGarvey, R., Feenstra, J. and Hawthorne, P. 2013. Southern Zone Rock Lobster (*Jasus edwardsii*) Fishery Status Report 2012/13. Status Report to PIRSA Fisheries and Aquaculture. South Australian Research and Development Institute (Aquatic Sciences), Adelaide. No. 749.
- B30. Rademeyer, Rebecca. Marine Resource Assessment and Management Group, Department of Mathematics and Applied Mathematics, University of Cape

Town, Rondebosch 7701, South Africa. Personal Communication.

APPENDIX C. PROPENSITY SCORE METHODOLOGY

The case-control design implemented in this project is predicated on the control stocks being as similar as possible to certified stocks. There are many characteristics of a fish stock that can affect how it will change with time. Thus there will never be a control stock that perfectly matches a certified stock. Propensity scoring is one option for evaluating the similarity among stocks across potentially confounding factors. This method attempts to summarize the pre-treatment conditions of each stock into a single value, the propensity score (Rosenbaum, 1983; Costello et al., 2008; Angrist et al, 1998). These scores can then be used to match the treatment groups with the most similar control group, or in our case, to exclude unsuitable controls. This propensity score methodology was deemed unnecessary for the rock lobster stocks due to their intrinsic similarity, in their management structures and geographic location.

Propensity Score Model

To calculate propensity scores, we determine the probability that a fish stock is certified (in the year that the treatment stock it will be compared to was certified). In our propensity score model, the dependent variable is a dummy variable for certification. The independent variables are factors that may have effects on stock trends. The probability that a stock is certified (PrCert) is estimated using the “lm” function in R. The model estimates the parameter values for each independent variable, and the model's fitted values are the propensity scores.

$$\text{PrCert} = \beta_0 + \beta_1 * \text{COV}_1 + \beta_2 * \text{COV}_2 + \dots + \beta_n * \text{COV}_n + u$$

where (COV_1, \dots, COV_n) is the list of n covariates.

Because we assume a normal distribution, but are calculating a probability constrained between 0 and 1, we perform a logit transformation. Once we have the fitted values of this model (PrCert), we calculate the final propensity score for each stock as:

$$PS = \frac{1}{1 + e^{-PrCert}}$$

Covariates

To determine which covariates are important in establishing control and treatment groups, Rubin recommends first identifying the "decision makers" for seeking and obtaining certification (2008). For this study, the decision makers are the MSC and the individual fishery (Table C1). The MSC's decision is guided by its three principles of sustainability: target stock health, management efficiency, and ecosystem health. We thus include two numeric covariates that relate to the stock status: Standardized F and standardized SSB at the beginning of the time series. These initial stock status values are standardized by published management reference points. Other covariates relating to stock sustainability included in the propensity score model initially were stock size, rebuilding status (whether a stock was known to be in an official rebuilding plan), and fishing method (fixed or mobile gear).

The other decision maker, the individual fishery, decides whether to seek certification based on the expected market benefit from the ecolabel. This market benefit, in turn, depends on the fishery's consumers and primary market. According to the MSC, "Europe offers the most developed market for MSC-labeled food, characterized by consumers who are concerned about seafood sustainability and an

active, engaged retail sector offering a wide choice of MSC products" (MSC around the world). Thus, we include in our propensity score model a covariate that indicates whether each stock's primary market is in Europe. Covariates for each primary market location in Asia, North America, and South America were also included in the initial propensity score model, but were ultimately excluded. A stepwise algorithm was used to identify covariates that did not contribute significantly to the fit of the propensity score model. Based on this algorithm, initial standardized SSB, stock size, fishing methods, and North America as a primary market location were eliminated (Table C1). The rebuilding status covariate was also eventually excluded based on unavailability or unreliability of data. Many stock management bodies do not publish information about rebuilding plans, even if one is underway. It is especially difficult to determine whether a stock was in a rebuilding program 5-15 years ago, which was the information required. Finally, the variables representing South America and Asia as primary market locations were excluded because there were too few stocks representing the regions to reliably estimate these parameters. This left only initial standardized F (F_0) and Europe as a primary market location (EUR) in the final propensity score model.

Exclusion of Controls

Once calculated, the propensity scores can be used to find "subgroups (subclasses, or matched pairs) of treated and control units such that within a subgroup, the treated and control units appear to be balanced with respect to their distributions of key covariates" (Rubin 2008). Controls were excluded if they were outside of a

certain range of the propensity score of the certified stock of the same species. For this study a range of ± 0.12 was chosen because it was narrow enough that the least suitable stocks are excluded, but wide enough that no taxonomic group is left without a control stock. For taxonomic groups with more than one certified stock, the certified stock with the lowest propensity score will be used to set the lower limit, and the certified stock with the highest propensity score to set the upper limit.

Results of Propensity Score Model

The propensity score model indicates that stocks are more likely to be certified if they have a lower F_0 and their primary market is in Europe/Western Russia (Table C2). Propensity scores ranged from 0.44 to 0.64 (Fig. 2 in main text). The certified and control stocks overlapped from 0.54 to 0.63. A total of four stocks were excluded, all from the hake and herring taxonomic groups (Figure C1). In the herring group, the lowest propensity scores was assigned to Irish Sea Herring (0.44), which was excluded, and the highest score to the certified North Sea stock (0.59). Among hake stocks, the northern stock in the Northeast Atlantic was assigned the lowest score (0.49), while the South African (*M. paradoxus*) stock had the highest score (0.63). The other certified South African stock (*M. capensis*) had a score of 0.55. In this group, the northern stock in the Northeast Atlantic, the North Argentine stock (PS = 0.52), and the South Argentine stock (PS = 0.51) were excluded.

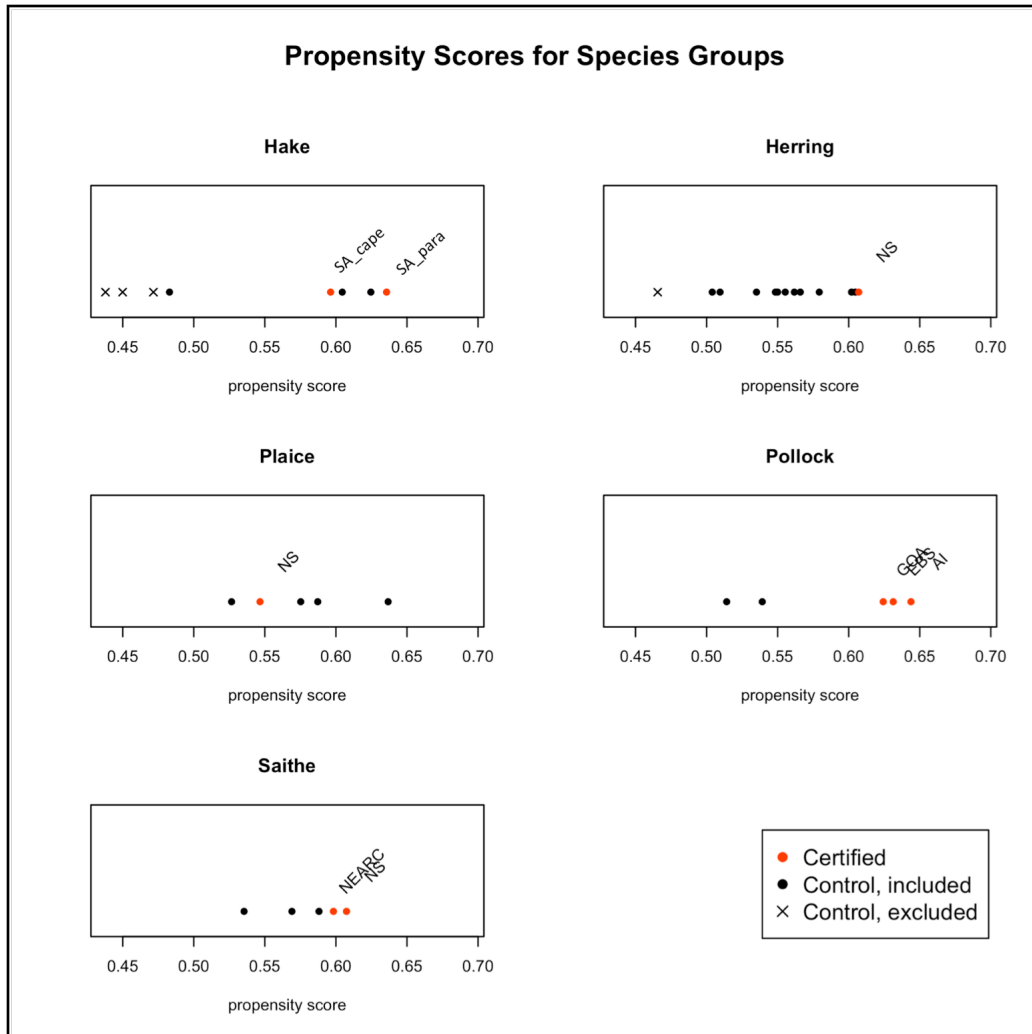
Table C1. Decision Makers and potential covariates for the propensity score model, along with grounds for exclusion, where applicable.

Decision Maker (Basis of Decision)	Covariate	Description	Excluded?
MSC (Health of Target Stock, Management, Ecosystem impact)	F_0	Standardized fishing mortality rate at the beginning of time series. Numeric.	No
	B_0	Standardized SSB at the beginning of the time series. Numeric.	Yes; excluded by stepwise algorithm.
	Stock Size: ln(SSB)	Natural log of SSB at beginning of time series (measure of stock size). Numeric.	Yes; excluded by stepwise algorithm.
	Rebuild	Stock's rebuilding status. Binary.	Yes; data unavailable or unreliable.
	Fixed Gear	Does the fishery use fixed gear (lines, traps, etc)? Initially, a mobile gear covariate was considered for inclusion, but every stock uses mobile gear of some sort. Binary.	Yes; excluded by stepwise algorithm.
Individual Fishery (Primary Market Location)	EUR	Europe and Western Russia, grouped together because fisheries rarely seem to sell in one market and not the other. Binary.	No
	SAMER	South America. Binary.	Yes; too few representative stocks
	NAMER	North America. Binary.	Yes; excluded by stepwise algorithm.
	ASIA	Asia, except Western Russia. Binary.	Yes; too few representative stocks

Table C2. Estimated coefficients of the propensity score model. P-values are provided in parentheses under the estimated coefficients.

Variable	Estimated Coefficient (p-value)
(intercept)	-0.027 (0.87)
F_0	-0.22 (0.004)
EUR	0.41 (0.04)
R ²	0.26 (0.006)

Figure C3. Diagram of stock propensity scores by taxonomic group.



APPENDIX D. DETAILED MODEL OUTPUT

Table D1 (A-F) . Estimated coefficients of the certification analysis by taxonomic group. Standard errors are provided in parentheses. Three asterisks (***) indicates that the coefficient is significant at the 1% level, two asterisks at the 5% level, and one asterisk at the 10% level. Greek characters refer to Eq. 1.

A. Herring			
Metric		ln(SSB)	Standardized F
Stock Intercepts ($\beta_{1,i}$)	North Sea Autumn Spawners	14.5	-0.32
	Gulf of Maine/George's Bank	12.7	-0.17
	West of Scotland	11.5	-0.13
	Irish Sea	9.23	0.37
	Celtic Seas	10.2	1.7
	Gulf of Riga	11.3	0.39
	Norwegian Spring Spawners	15.7	0.04
	Bothnian Sea (area 30)	13.0	-0.20
	Bothnian Sea (area 31)	9.26	1.5
	Baltic Spring Spawners	11.9	0.55
	Icelandic	13.1	0.0
Eastern Baltic	13.1	1.1	
time (β_2)		0.013 (0.0095)	-0.035 (0.0083)
time:certification (β_3)		-0.034 (0.033)	-0.020 (0.029)
B. Hake			
Metric		ln(SSB)	Standardized F
Stock Intercepts ($\beta_{1,i}$)	South African (<i>M. capensis</i>)	12.1	-0.740
	South African (<i>M. paradoxus</i>)	11.6	0.382
	Southern stock in NE Atlantic	9.5	2.21
	Pacific	14.6	-0.281
	Chilean	12.3	-0.302
time (β_2)		-0.064** (0.021)	0.028 (0.013)
time:certification (β_3)		-0.0065 (0.034)	-0.0049 (0.021)
Autocorrelation Coefficient (ϕ)		0.33	--

C. Saithe			
Metric		ln(SSB)	Standardized F
Stock Intercepts ($\beta_{1,i}$)	Northeast Arctic	12.9	0.549
	North Sea	12.5	-0.0060
	Faroe Islands	11.4	0.636
	Iceland	11.9	-0.0550
		-0.019	-0.0290
time (β_2)		(0.015)	(0.018)
time:certification (β_3)		-0.048** (0.022)	0.0073*** (0.026)
Autocorrelation Coefficient (ϕ)		0.57	0.098

D. Plaice			
Metric		ln(SSB)	Standardized F
Stock Intercepts ($\beta_{1,i}$)	North Sea	10.6	0.122
	Iceland	10.5	-0.0870
	Irish Sea	9.7	-0.931
	Southwest of Ireland	11.0	0.283
	Kattegat and Belts Sound	11.9	0.912
time (β_2)		0.014 (0.010)	-0.086*** (0.029)
time:certification (β_3)		0.11*** (0.023)	-0.044 (0.064)

E. Pollock			
Metric		ln(SSB)	Standardized F
Stock Intercepts ($\beta_{1,i}$)	Aleutian Islands	11.1	-0.964
	Eastern Bering Sea	14.7	-0.594
	Gulf of Alaska	11.9	-0.472
	Sea of Okhotsk	15.3	0.00800
	Japanese	12.4	-0.862
time (β_2)		0.020 (0.018)	-0.043*** (0.016)
time:certification (β_3)		-0.027 (0.023)	0.0392* (0.020)
Autocorrelation Coefficient (ϕ)		0.65	--

F. Rock Lobster			
Metric		ln(SSB)	Standardized F
Stock Intercepts ($\beta_{1,i}$)	S. Australian North	7.73	4.80E-05
	S. Australian South	8.20	1.03E-04
	W. Australian Abrolhos	5.63	-
	W. Australian Central	4.78	-
	W. Australian North	5.44	5.83E-05
	W. Australian South	4.98	1.00E-05
time (β_2)		-0.00450 (0.014)	-0.038*** (0.0062)
time:certification (β_3)		0.00510 (0.017)	0.048*** (0.0088)
Autocorrelation Coefficient (ϕ)		0.81	-

APPENDIX E. ALTERNATIVE LINEAR MODELS

Various statistical models were investigated to analyze the data in this project. For the sake of consistency, we presented results from the model that most commonly received the greatest support based on AIC. This was the generalized linear model fit by ordinary least squares, and in some cases, fitted by generalized least squares with autocorrelated residuals. This appendix presents results from successful trials of two alternative forms of our model: a linear mixed effects model with random intercepts, and a model with nonlinear terms.

Linear Mixed Effects Model

Model Description

Mixed effects models account for variation due to a predictable grouping factor (random effects, in this case stock membership) before accounting for variation due to the treatment level (fixed effect, in this case certification status). This linear mixed effects model (lme) included a random stock effect for both the intercept and the slope. Thus, it controls for the effect of stock membership on both intercept and slope before assessing certification's effect on them. Like the linear models already presented, the lme groups the stocks into control and certified groups. However, each stock is assigned its own intercept and slope. Within each treatment group (certified and uncertified) the stocks share a common distribution of intercepts and slopes. The mean slope of the certified stocks is then compared to the mean slope of the control stocks. This model, for the i^{th} stock is described as:

$$y_{i,t} = X_{i,t} + Z_{i,t} + \eta_{i,t}$$

$$X_{i,t} = \text{fixed effects} = (\beta_1 + \beta_2 * \text{time} + \beta_3 * \text{certification} + \beta_4 * \text{time} * \text{certification})$$

$$Z_{i,t} = \text{random effects} = (b_{1,i} + b_{2,i} * \text{time})$$

$$b_i \sim N(0, \Psi_i)$$

$$\eta_{i,t} = \phi \eta_{i,t-1} + \varepsilon_{i,t}$$

$$\varepsilon_{i,t} \sim N(0, \sigma_i^2)$$

Where Ψ is the variance for the random effects and σ^2 is the variance of errors in stock i . Note that the fixed effect coefficients (β 's) are the same across all stocks, while the random effect coefficients (b 's) are different for each stock. As with the linear model, a grouping structure was included in the fixed effects, random effects, and autocorrelation formula to sort the data by stock. If the autocorrelation coefficient of this model approached one (was greater than 0.9), it was fixed at a value estimated from a basic linear model (see Methods Section).

Results and Discussion

In several cases, the linear mixed effects model with random slopes and intercepts was inestimable. This observation contributed to our rejection of this model for the final analysis. The seven analyses for which the mixed effects model was estimable showed evidence of random intercepts and random slopes within groups (certified or uncertified). This is observable in Figure E1 (A-G), which shows that stocks within the same treatment group have different slopes. For example, while the Baltic spring spawning herring stock was fit with a negative slope in biomass, the Gulf of Maine/George's Bank stock was found to be increasing (Fig. E1.A). The random effects model captures this variation among control stocks, whereas the fixed effect model in the final analysis forced a single overall slope on all controls stocks. In both

models, however, a single overall slope is calculated for each group, which allows for a comparison of how certified stocks change relative to control uncertified stocks. In 5 of these 7 analyses, the certified group showed lower overall performance (had lower trends in SSB or higher trends in F) than the control groups (Table E1, Fig. E1). One exception to this observation was rock lobster (SSB), in which the certified group was increasing in biomass and the control group decreasing (Fig. E1.G). The other exception was Herring (F), whose certified stocks decreased in F more quickly than the control stocks (Fig. E1.B). Nonetheless, in all cases, the effect of certification on stock trends was insignificant.

The results from the lme model generally lead to the same conclusions as the fixed effects models in the final analysis. In every case, the fixed and mixed effects models agree on the direction of change (positive or negative) in stock status for both control and certified groups. For all but one case (saithe SSB, Fig. E1.E), the models agreed on the significance of the certification effect. In the final analysis, saithe stocks that were certified were found to decrease in SSB significantly faster than uncertified stocks. The mixed effects model, while it fit the certified stocks with an overall lower slope than the uncertified stocks, did not suggest that the difference in trends between these two groups was statistically significant. Note that while one control stock (Faroe Islands) decreases in SSB, the other control stock (Icelandic) increases. Thus the certified stocks might be decreasing in SSB significantly faster than the former stock, but not the latter. The random effects model accounts for this variation when it calculates the overall effect of certification. In summary, several cases considered in this study showed evidence of random effects on slope and intercept, which supports

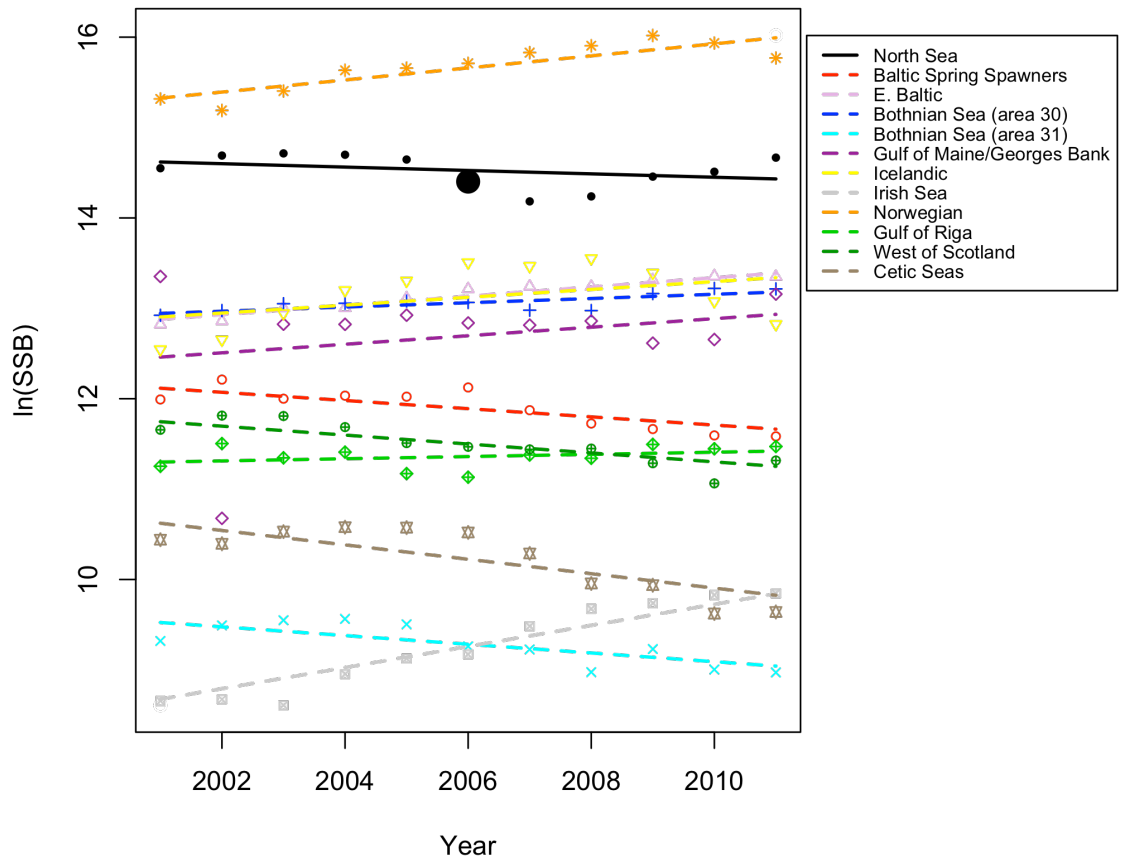
the use of a mixed effects model. However, this model was inestimable for many analyses, and if ultimately selected for the analysis, would have severely limited the scope of the study. Moreover, in all but one instance, the linear mixed effects model corroborates the conclusions reached by the fixed effects model, so the cost of using the mixed effects model outweighed the benefits.

Table E1. Output summary of linear mixed effects model with random slopes and intercepts. Analyses for which the model did not converge are excluded. Standard errors are provided in parentheses. An asterisk (*) indicates that the phi estimate was fixed at the given value.

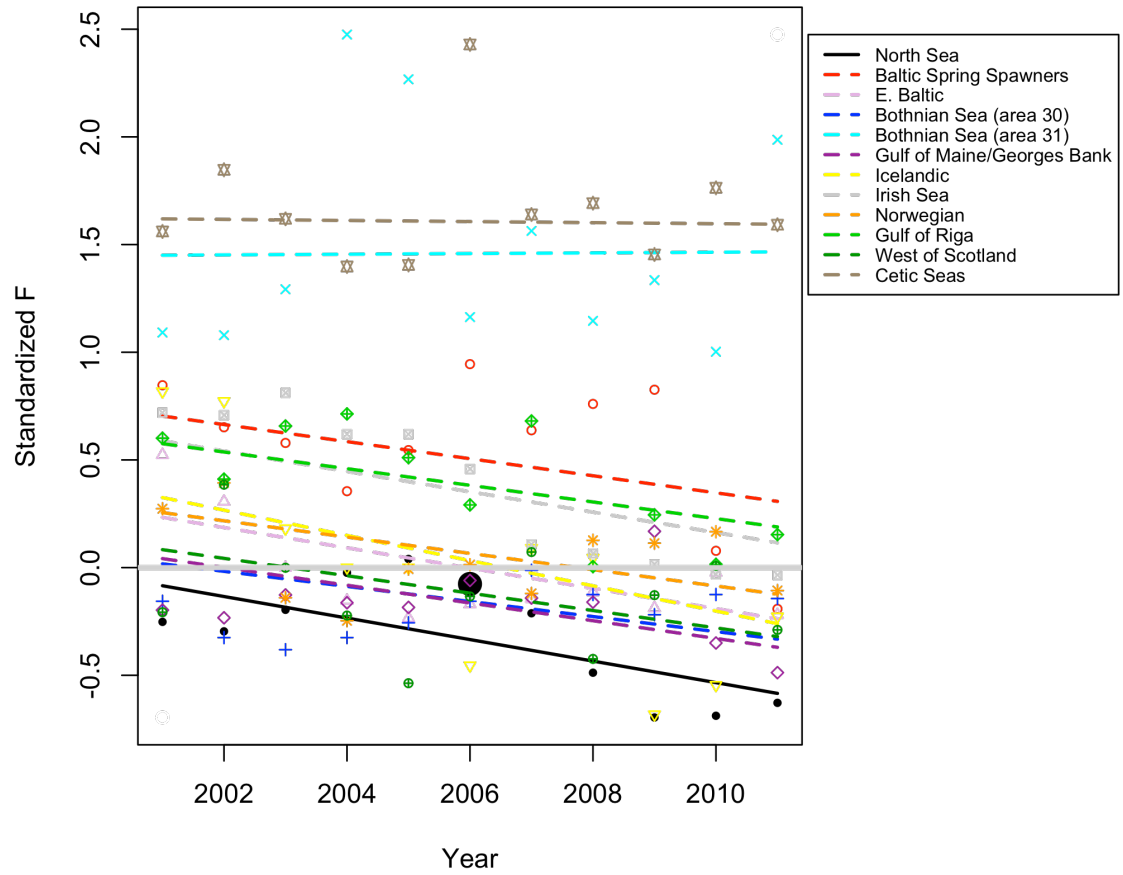
	Herring (SSB)	Herring (F)	Hake (SSB)	Hake (F)	Saithe (SSB)	Pollock (SSB)	Rock Lobster (SSB)
Observations	132	132	55	55	42	55	66
Mean Certified Intercept	14.5	-0.33	11.8	-0.26	12.6	12.7	5.21
Mean Control Intercept	11.9	0.36	12	0.53	11.6	13.9	7.97
<u>time</u>	0.013 (0.02)	-0.035 (0.01)	-0.049 (0.06)	0.02 (0.02)	-0.019 (0.03)	0.02 (0.04)	-0.0044 (0.03)
<u>time*certification</u>	-0.031 (0.07)	-0.015 (0.05)	-0.00036 (0.09)	-0.008 (0.04)	-0.048 (0.04)	-0.027 (0.05)	0.0049 (0.03)
Overall Certified Slope (y ⁻¹)	-0.019	-0.050	-0.049	0.012	-0.067	-0.0069	0.00048
Overall Control Slope (y ⁻¹)	0.013	-0.035	-0.049	0.020	-0.019	0.020	-0.0044
Phi	0.11	0.43	0.78*	0.76	0.57*	0.65*	0.80*

Figure E1. Observed data (points) and trends fitted by mixed effects model (lines) for certified (solid) and uncertified (dashed) stocks of each taxonomic groups.

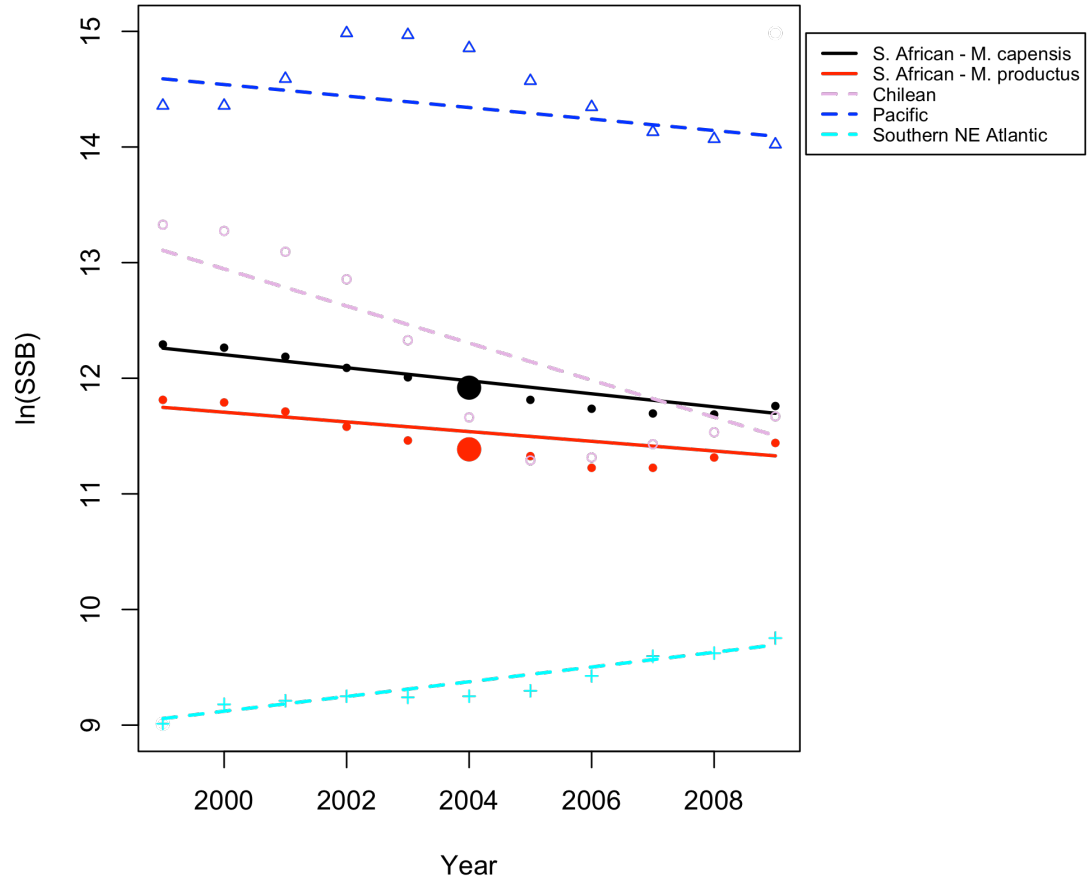
A) HERRING: Spawning Stock Biomass



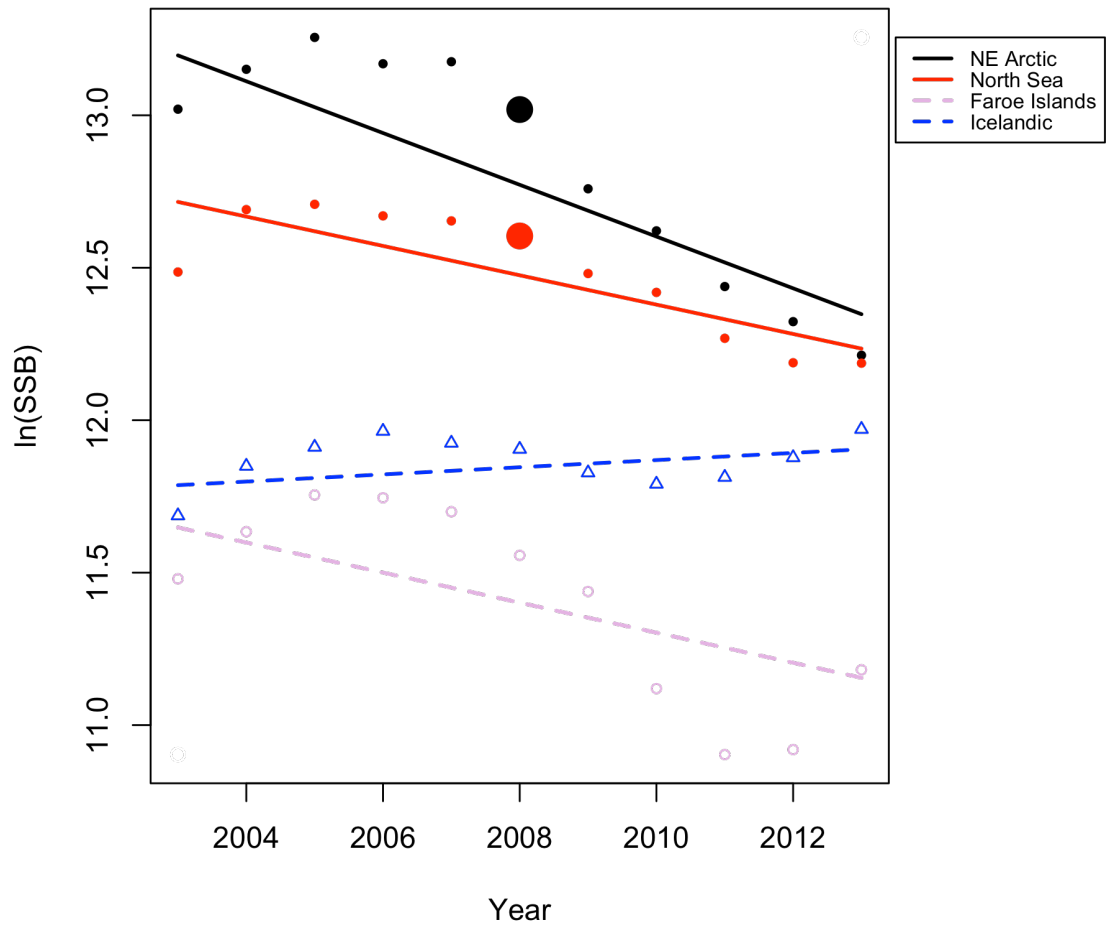
B) HERRING: Fishing Mortality Rate



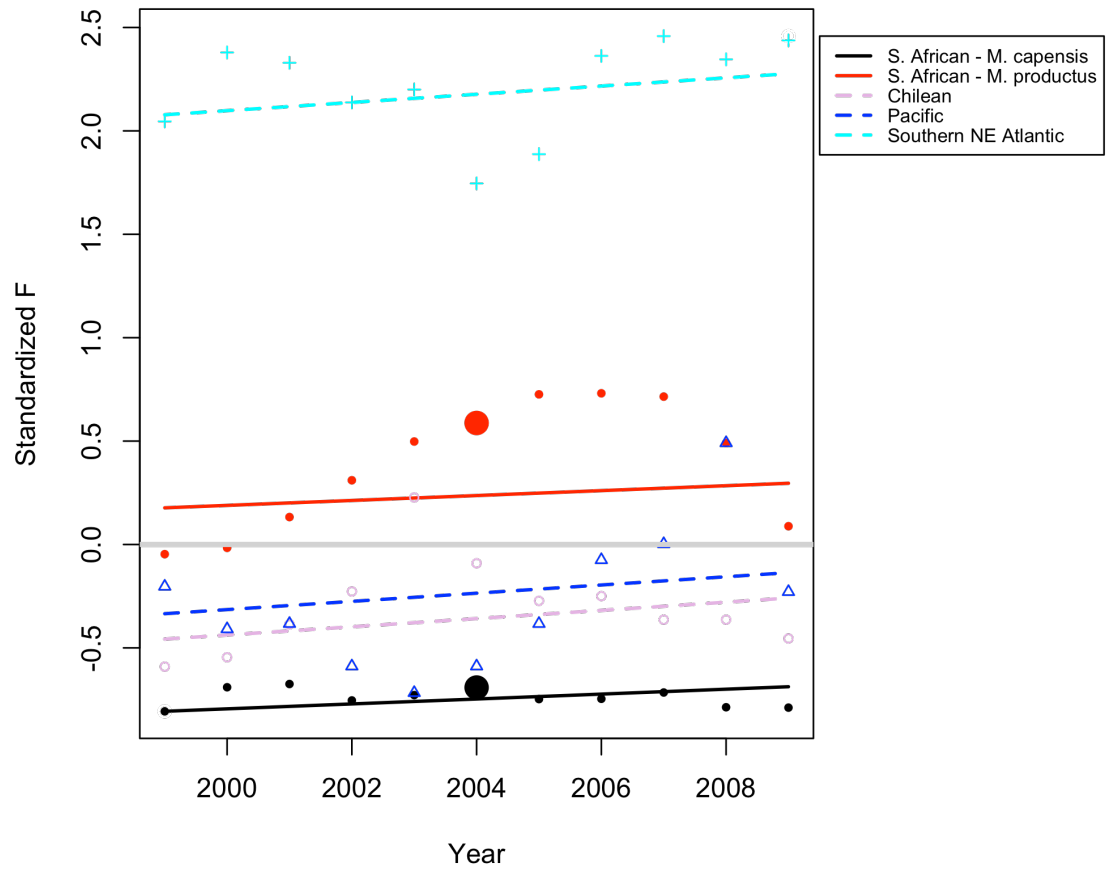
C) HAKE: Spawning Stock Biomass



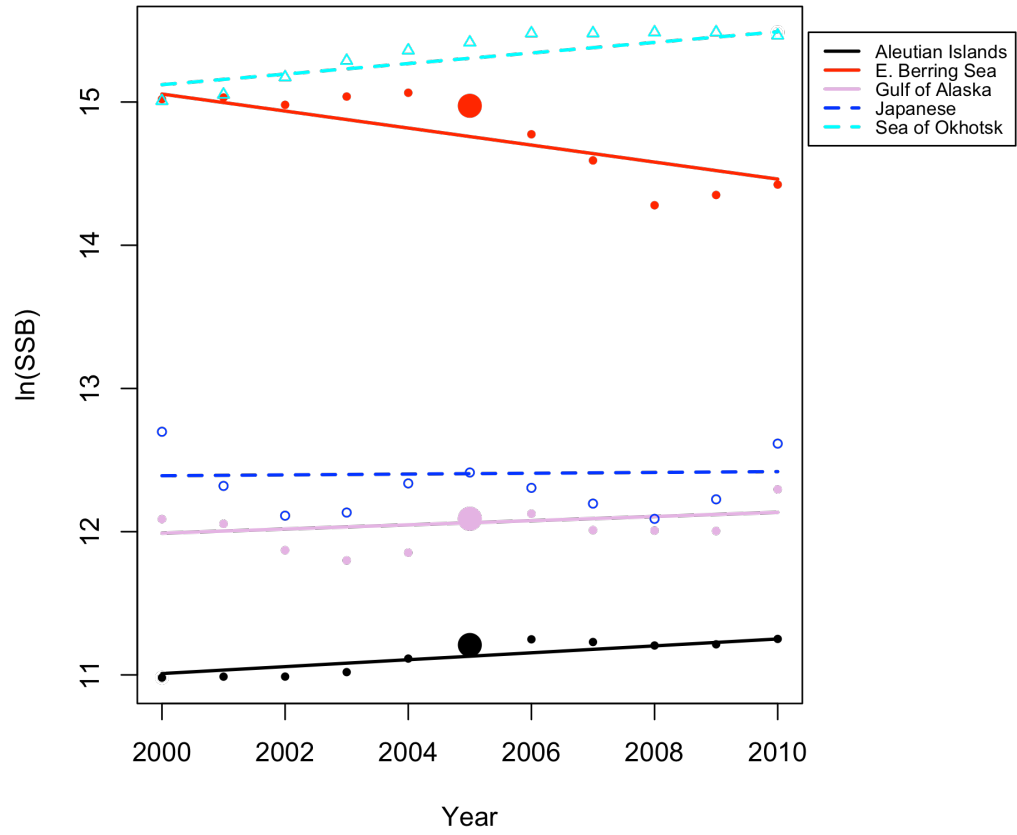
D) SAITHE: Spawning Stock Biomass



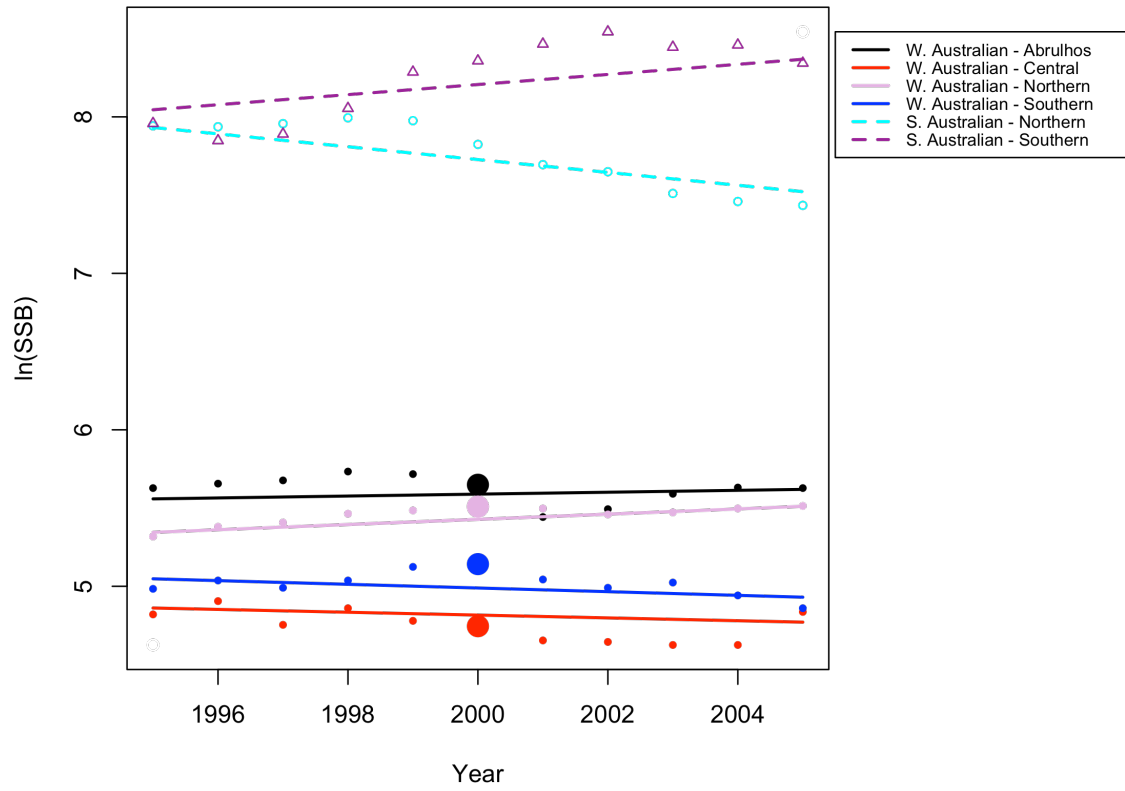
E) HAKE: Fishing Mortality Rate



F) POLLOCK: Spawning Stock Biomass



G) ROCK LOBSTER: Spawning Stock Biomass



Model with non-linear terms

Model Description

The data were also analyzed with a model containing non-linear terms for time (t^2 and t^3):

$$X_{i,t} = \beta_0 + \beta_{1,i} * \text{stock} + \beta_2 * t + \beta_3 * t^2 + \beta_4 * t^3 + \beta_5 * t * C + \beta_6 * t^2 * C + \beta_7 * t^3 * C + \varepsilon_{i,t}$$

where t is time (years), C is certification status (1 or 0), and $\varepsilon_t \sim N(0, \sigma^2)$. For both metrics (SSB and F) in each taxonomic group, a stepwise algorithm was used to eliminate terms that did not contribute significantly to the model fit. In most cases, the stepwise algorithm excluded the term describing the interaction of certification with time ($\beta_5 * t * C$). Given that this term addresses the question posed in this study, these trials were not of much use. However, the interaction term was retained in the 4 cases described below (Plaice SSB, Saithe F , Pollock SSB, and Rock Lobster F).

Time was retained as a covariate in each model, and was significant in all of them except for Pollock (SSB) (Table E2). The t^2 term was retained in the Plaice (SSB) and Saithe (F) trials, indicating that these models exhibit non-linear tendencies (Fig. E2.A-B). The cubic (t^3) function describes consecutive rises and falls in the dependent variable, and was retained in the saithe (F) and rock lobster (F) analyses (Fig. E2.B and E2.D). In the pollock analysis, all terms were excluded except for intercepts, t , and $t * C$, resulting in a linear model similar to the one used in the final analysis (Figure E2.C). For rock lobster, the stock term was excluded, suggesting the stocks did not have significantly different values of standardized F in the year of certification (Figure E2.C). In two trials (plaice SSB and rock lobster F), the term

describing the interaction of certification with t^2 or t^3 was retained (Table E2). These terms, because their estimates are less than zero, suggest that uncertified stocks are less linear than the certified stocks, and are therefore better described by quadratic or cubic functions.

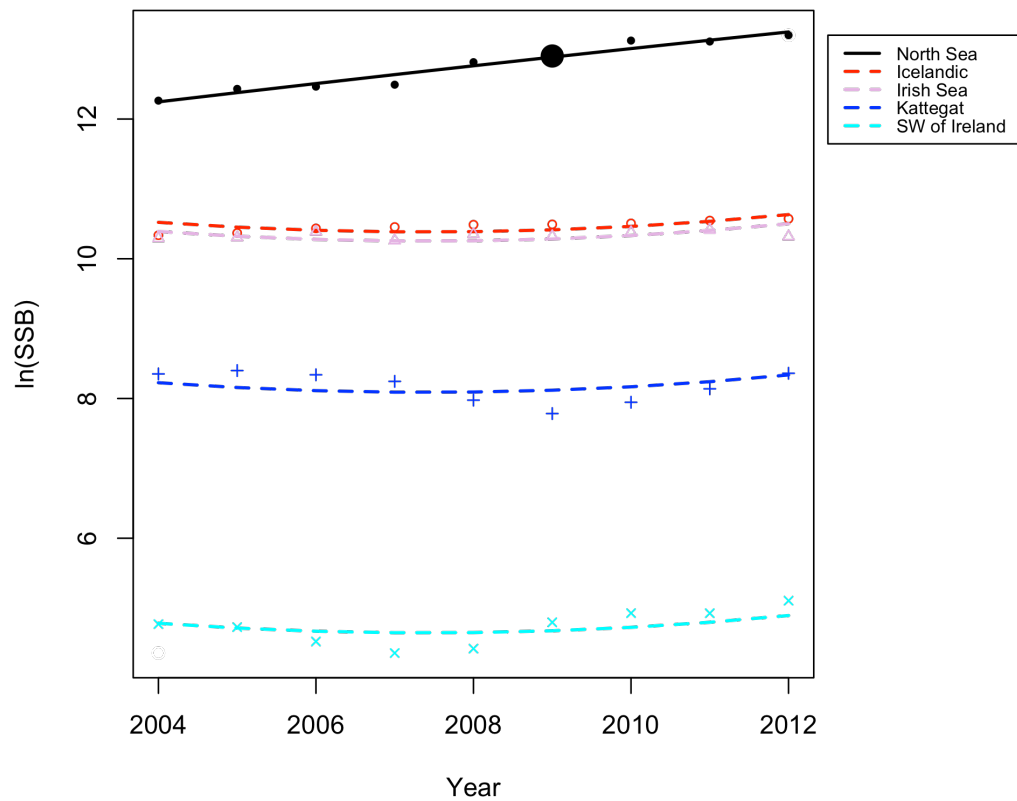
In each of these analyses, the effect of certification on stock trends was significant at the 10% level, and for plaice SSB, saithe F, and rock lobster F, at the 1% level. For plaice, the effect of certification was positive; certified stocks increased in SSB significantly faster than uncertified stocks. For the other three taxonomic groups, uncertified stocks showed better performance than certified ones, increasing faster in F (saithe and rock lobster), or decreasing faster in SSB (pollock). These conclusions agree strongly with those reached by the final analysis. Note that, though the mixed effects model described above did not indicate a significant certification effect for saithe (F), the final analysis and the present analysis with non-linear terms are in agreement.

Table E2. Output summary for model including non-linear covariates. Standard errors are provided in parentheses. Three asterisks indicates that the coefficient is significant at the 1% level, two asterisks at the 5% level, and one asterisk at the 10% level.

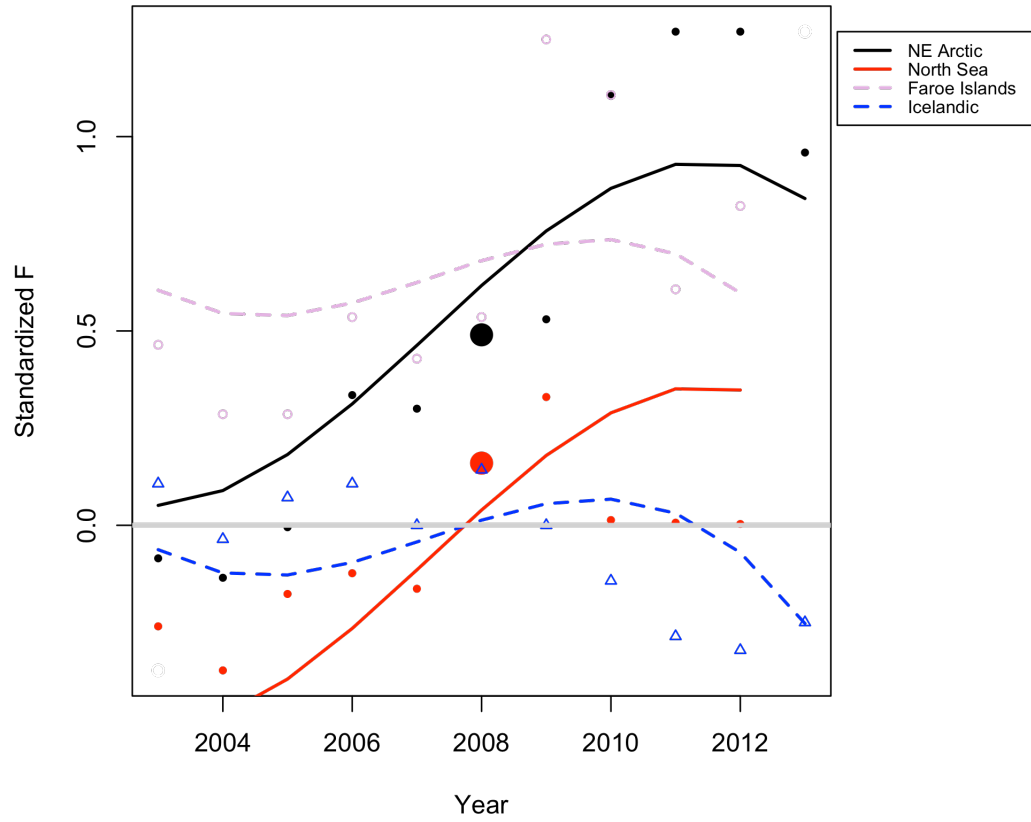
	Plaice (SSB)		Saithe (F)		Pollock (SSB)	Rock Lobster (F)
Stock	Irish Sea	10.3	NE Arctic	0.62	Aleutian Islands	11.1
Intercepts	Kattegat	8.11	North Sea	0.04	E. Bering Sea	14.7
	N. Sea	12.9	Faroe Islands	0.68	Gulf of Alaska	12.0
	SW of Ireland	4.70	Iceland	0.010	Japanese	12.3
	Iceland	10.4			Sea of Okhotsk	15.3
t_1	0.037 *** (0.01)		0.052 *** (0.03)		0.021 (0.01)	-0.080 *** (0.01)
t_2	0.012* (0.04)		-0.0068 (0.005)			
t_3			-0.0028 (0.02)			0.0023 *** (0.01)
t^* certification	0.086 *** (0.03)		0.098 *** (0.02))		-0.032 * (0.02)	0.091 *** (0.0007)
t^{2*} certification	-0.013 (0.009)					
t^{3*} certification						-0.0024** (0.001)

Figure E2. Observed data (points) and trends fitted by the model with non-linear terms (lines) for certified (solid) and uncertified (dashed) stocks of each taxonomic groups.

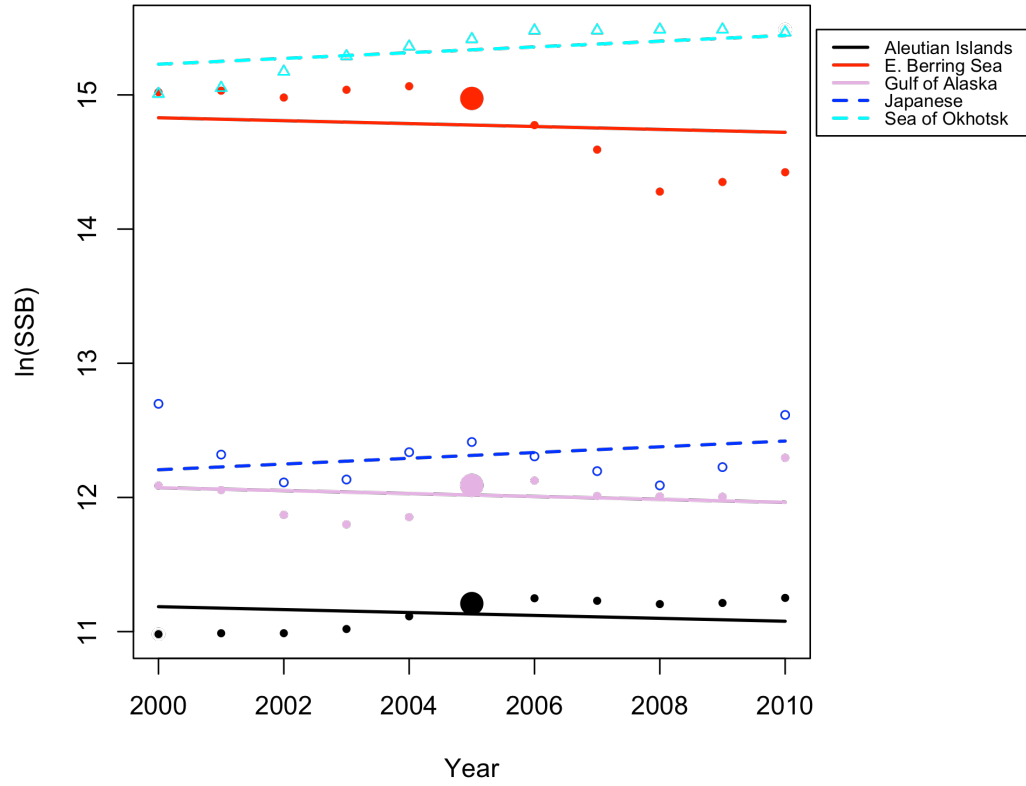
A) PLAICE: Spawning Stock Biomass



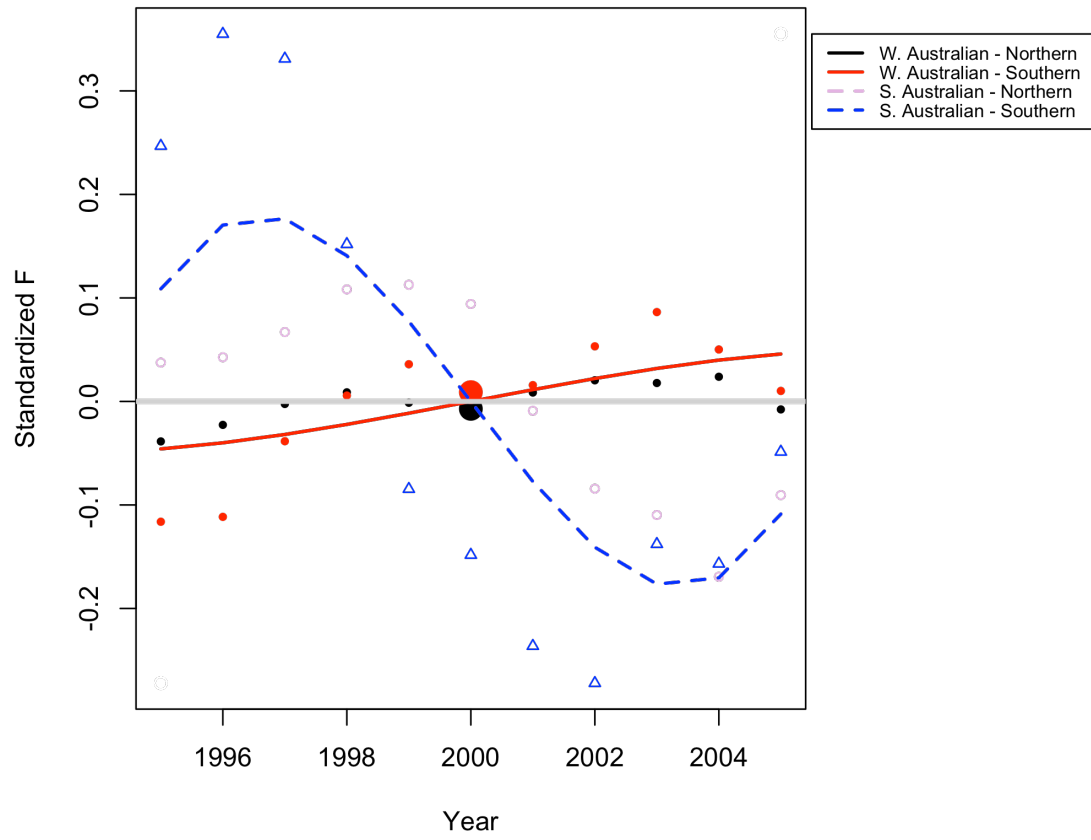
B) SAITHE: Fishing Mortality Rate



C) POLLOCK: Spawning Stock Biomass



D) ROCK LOBSTER: Fishing Mortality Rate



BIBLIOGRAPHY

- Agnew, D., Grieve, C., Orr, P., Parkes, G., and Barker, N. 2006. *Environmental benefits resulting from certification against MSC's Principles and Criteria for Sustainable Fishing*. Marine Resource Assessment Group Ltd. UK. and Marine Stewardship Council. London, UK.
- Agnew, D. J., Gutiérrez, N. L., Stern-Pirilot, A., Smith, A. D. M., Zimmermann, C., and Sainsbury, K. 2012. Rebuttal to Froese and Proelss "Evaluation and legal assessment of certified seafood". *Marine Policy* **38**: 551-553.
- Angrist, J. D. 1998. Estimating the labor market impact of voluntary military service using social security data on military applicants. *Econometrica* **66**: 249-288.
- Barbeaux, S., Ianelli, J., and Palsson, W. 2012. Assessment of the Pollock stock in the Aleutian Islands. In *Stock assessment and fishery evaluation report for the groundfish resources of the Bering Sea/Aleutian Islands Region*. North Pacific Management Council.
- Beddington, J.R., Agnew, D.J., and Clark, C.W. 2007. *Current problems in the management of marine fisheries*. *Science* **316**: 1713-1716.
- Bezzi, S.; Cañete, G.; Pérez, M.; Renzi, M. and Lassen, H., 1994. *Report of the INIDEP working group on assessment of hake (Merluccius hubbsi) North of 48° S (Southwest Atlantic Ocean)*. Instituto Nacional de Investigación y Desarrollo Pesquero Scientific Document No. 3.
- Christian, C., Ainley, D., Bailey, M., Dayton, P., Hocevar, J., LeVine, M., Nikoloyuk, J., Nouvian, C., Velarde, E., Werner, R.W., and Jacquet, J. 2013. A review of

- formal objections to Marine Stewardship Council fisheries certifications. *Biological Conservation* **161**: 10-17.
- Cooper, H., Hedges, L. V. 1994. Fixed effects models. *The handbook of research synthesis*: 285-299.
- Costello, C., Gaines, S. D., and Lynham, J. 2008. Can catch shares prevent fisheries collapse?. *Science* **321**: 1678-1681.
- DerSimonian, R., Laird, N. 1986. Meta-analysis in clinical trials. *Controlled clinical trials* **7**:177-188.
- Dorn, M., Aydin, K., Barbeaux, S., Darin, J., and Palsson, W., 2012. Assessment of the Walleye Pollock Stock in the Gulf of Alaska. In *Stock assessment and fishery evaluation report for the groundfish resources of the Gulf of Alaska*. North Pacific Management Council.
- Ferraro, P. J., McIntosh, C., & Ospina, M. 2007. The effectiveness of the US endangered species act: An econometric analysis using matching methods. *Journal of Environmental Economics and Management* **54**: 245-261.
- Fisheries Research Agency of Japan (FRAJ). 2012. *Summary Resource Assessment of Alaska Pollock in the Pacific Systems Group*. Hokkaido National Fisheries Research Institute.
- Fish Source. <http://www.fishsource.com>. (accessed 12 August 2014).
- Froese, R., and Proelß, A. 2012. Evaluation and legal assessment of certified seafood. *Marine Policy* **36**: 1284-1289.
- Gerrodette, T. (1987). A power analysis for detecting trends. *Ecology*. 1364-1372.

- Goyert, W., Sagarin, R., and Annala, J. 2010. The promise and pitfalls of Marine Stewardship Council certification: Maine lobster as a case study. *Marine Policy* **34**: 1103-1109.
- Gulbrandsen, L. H. (2009). The emergence and effectiveness of the Marine Stewardship Council. *Marine Policy* **33**: 654-660.
- Gutiérrez, N. L., Valencia, S. R., Branch, T. A., Agnew, D. J., Baum, J. K., Bianchi, P. L., ...and Williams, N. E. 2012. Eco-label conveys reliable information on fish stock health to seafood consumers. *Public Library of Science One* **7**: e43765.
- Hall, S.J., and Mainprize, B.M. 2005. Managing by-catch and discards: how much progress are we making and how can we do better?. *Fish and Fisheries* **6**: 134-155.
- Ianelli, J.N., Honkalehto, T., Barbeaux, S., Kotwicki, S., Aydin, K., and Williamson, N. 2012. Assessment of the walleye pollock stock in the Eastern Bering Sea. In *Stock assessment and fishery evaluation report for the groundfish resources of the Bering Sea/Aleutian Islands Regions*. North Pacific Management Council
- ICES. 2012. Report of the Baltic Fisheries Assessment Working Group (WGBFAS). *International Council for the Exploration of the Sea Council and Meeting Document* 2012/ACOM:10.
- ICES. 2013. Report of the Working Group on Widely Distributed Stocks (WGWIDE). *International Council for the Exploration of the Sea Council and Meeting Document* 2013/ACOM:15.

- ICES. 2012. Report of the North-Western Working Group. *International Council for the Exploration of the Sea Council and Meeting Document 2012/ACOM:07*.
- ICES. 2012. Report of the ICES Advisory Committee: Herring in Division IIIa and Subdivisions 22-24 (western Baltic spring spawners). *International Council for the Exploration of the Sea Advice*. Book 6.
- ICES. 2012. Report of the Working Group on the Assessment of Southern Shelf Stocks of Hake, Monk and Megrin (WGHMM). *International Council for the Exploration of the Sea Council and Meeting Document 2012/ACOM:11*.
- ICES. 2012. Report of the Herring Assessment Working Group for the Area South of 62 N (HAWG). *International Council for the Exploration of the Sea Council and Meeting Document 2012/ACOM:06*.
- ICES. 2013. Plaice in Subarea IV (North Sea). *International Council for the Exploration of the Sea Advice 2013*. Book 6, 6.4.18.
- ICES. 2013. Report of the Working Group for Celtic Seas Ecoregion (WGCSE). *International Council for the Exploration of the Sea Council and Meeting Document 2013/ACOM*.
- ICES. 2013. Report of the Arctic Fisheries Working Group (AFWG). *International Council for the Exploration of the Sea Council and Meeting Document 2013/ACOM:05*.
- ICES, 2012. Report of the North-Western Working Group (NWWG). *International Council for the Exploration of the Sea Council and Meeting Document 2012/ACOM:07*. 1425 pp.

- ICES. 2012. Report of the Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak (WGNSSK). *International Council for the Exploration of the Sea Council and Meeting Document 2012/ACOM:13*.
- Instituto De Fomento Pesquero (IFOP). *Estatus y posibilidades de explotación biologicamente sostenables de los principales recursos pesqueros nacionales año 2012: Merluza Común, 2012*. Subsecretaría de Pesca.
- Irusta, C.G., and D'Atri, L.L. 2012. *Evaluacion del estado del efectivo norte de 41° S de la Merluza (Merluccius hibbsi) y estimación de la captura biológicamente aceptable para el año 2013*. Instituto Nacional de Investigación Desarrollo Pesquero (INIDEP).
- Jackson, J. B., Kirby, M. X., Berger, W. H., Bjorndal, K. A., Botsford, L. W., Bourque, B. J., ... and Warner, R. R. 2001. Historical overfishing and the recent collapse of coastal ecosystems. *Science* **293**: 629-637.
- Jacquet, J., Hocevar, J., Lai, S., Majluf, P., Pelletier, N., Pitcher, T., ... and Pauly, D. 2009. Conserving wild fish in a sea of market-based efforts. *Oryx* **44**: 45-56.
- Jacquet, J., Pauly, D., Ainley, D., Holt, S., Dayton, P., and Jackson, J. 2010. Seafood stewardship in crisis. *Nature* **467**: 28-29.
- Linnane, A., McGarvey, R., Feenstra, J. and Hoare, M. 2013. *Northern Zone Rock Lobster (Jasus edwardsii) Fishery Status Report 2012/13. Status Report to PIRSA Fisheries and Aquaculture*. South Australian Research and Development Institute (Aquatic Sciences), Adelaide. No. 750.
- Linnane, A., McGarvey, R., Feenstra, J. and Hawthorne, P. 2013. *Southern Zone Rock Lobster (Jasus edwardsii) Fishery Status Report 2012/13. Status Report*

- to *PIRSA Fisheries and Aquaculture*. South Australian Research and Development Institute (Aquatic Sciences), Adelaide. No. 749.
- Lotze, H. K., Lenihan, H.S., Bourque, B.J., Bradbury, R.H., Cooke, R.G., Kay, M.C., Kidwell, S.M., Kirby, M.X., Peterson, C.H., and Jackson, J.B.C. 2006. Depletion, degradation, and recovery potential of estuaries and coastal seas. *Science* **312**: 1806-1809.
- Martell, S. J. D. 2010. *Assessment and management advice for Pacific hake in U.S. and Canadian waters in 2010*. Pacific Fishery Management Council. Portland, Oregon.
- Martin, S. M., Cambridge, T. A., Grieve, C., Nimmo, F. M., and Agnew, D. J. 2012. An evaluation of environmental changes within fisheries involved in the Marine Stewardship Council certification scheme. *Reviews in Fisheries Science* **20**: 61-69.
- Marine Research Institute. 2012. *State of stocks 2011/2012, Prospects 2012/2013*.
- Marine Stewardship Council. 2010. *Marine Stewardship Council Fisheries Assessment Methodology and Guidance to Certification Bodies*.
- Marine Stewardship Council. 2010. *MSC Theory of Change*. Marine Stewardship Council, London. (last accessed December 2012). http://www.msc.org/documents/msc-brochures/msc-theory-ofchange/at_download/file
- Marine Stewardship Council. 2013. *Global Impacts Report*.
- Marine Stewardship Council. 2011. *Harnessing market forces for positive environmental change: The MSC theory of change*.

- "MSC around the world". Marine Stewardship Council. Web. (Accessed 7 June 2014).
<<http://www.msc.org/business-support/global-markets/europe>>
- Northeast Fisheries Science Center. 2012. Stock Assessment of Atlantic Herring - Gulf of Maine/George's Bank. 54th Northeast Regional Stock Assessment Workshop (54th SAW) Assessment Summary Report. *US Department of Commerce, Northeast Fisheries Science Center Reference Document*.
- Northeast Fisheries Science Center. 2010. *50th Northeast Regional Stock Assessment Workshop (50th SAW) Assessment Report*. US Department of Commerce, Northeast Fisheries Science Center Reference Document 10-17.
- O'Boyle, R., Japp, D., Payne, A., and Devitt, S. 2013. *Russian Sea of Okhotsk Mid-water Trawl Walleye Pollock (Theragra chalcogramma) Public Certification Report*. Intertek Moody Marine.
- Osmena, Peter. 2010. *Statistical power analysis using SAS and R: A senior project presented to the statistics department at California Polytechnic State University, San Luis Obispo*.
- Parkes, G., Young, J.A., Walmsley, S.F., Abel, R., Harman, J., Horvat, P., Lem, A., MacFarlane, A., Mens, M., and Nolan, C. 2010. Behind the signs—a global review of fish sustainability information schemes. *Reviews in Fisheries Science* **18**: 344-356.
- Peterman, R. M. 1990. Statistical power analysis can improve fisheries research and management. *Canadian Journal of Fisheries and Aquatic Sciences* **47**: 2-15.
- Pauly, D., Alder, J., Bennett, E., Christensen, V., Tyedmers, P., and Watson, R. 2003. The future for fisheries. *Science* **302**: 1359-1361.

- Rademeyer, Rebeccah. Marine Resource Assessment and Management Group, Department of Mathematics and Applied Mathematics, University of Cape Town, Rondebosch 7701, South Africa. Personal Communication.
- Roheim, Cathy A. 2009. An evaluation of sustainable seafood guides: implications for environmental groups and the seafood industry. *Marine Resource Economics* **24**: 301-10.
- Roheim, C.A., Asche, F., and Santos, J.I. 2011. The elusive price premium for ecolabelled products: evidence from seafood in the UK market. *Journal of Agricultural Economics* **62**: 655-668.
- Rosenbaum, P.R., and Rubin, D. B. 1997. Estimating causal effects from large data sets using propensity scores. *Annals of internal medicine* **127**: 757-763.
- Rotenberry, J. T., & Wiens, J. A. (1985). Statistical power analysis and community-wide patterns. *American Naturalist*, 164-168.
- Rubin, D. B. 2008. For objective causal inference, design trumps analysis. *The Annals of Applied Statistics*: 808-840.
- Sekhon, J. S. 2008. Multivariate and propensity score matching software with automated balance optimization: the matching package for R. *Journal of Statistical Software, Forthcoming*.
- Silvert, William. 1997. Ecological impact classification with fuzzy sets. *Ecological Modeling* **96**: 1-10.
- Subsecretaría de Pesca (Subpesca). 2011. *Cuota global anual de captura de merluza común (Merluccius gayi gayi), año 2012*. Informaciona Technica (R.Pesq.) N° 117/2011

- Trusty, Michael F. 2010. Environmental improvement of seafood through certification and ecolabelling: theory and analysis. *Fish and Fisheries* **13**: 1-13.
- UNCLOS (1982) United Nations Convention on the Law of the Sea. 1833: 1–186. (Accessed July 2014). <http://treaties.un.org/doc/Publication/UNTS/Volume%201833/volume-1833-A-31363-English.pdf>.
- Villarino, M.F., Santos, B.A., and Renzi, M.A. 2012. *Evaluacion del estado del efectivo sur de 41° S de la Merluza (Merluccius hubbsi) y estimación de la captura biológicamente aceptable para 2013*. Instituto Nacional de Investigación Desarrollo Pesquero (INIDEP).
- Ward, T. J. (2008). Barriers to biodiversity conservation in marine fishery certification. *Fish and Fisheries* **9**: 169-177.
- Wessells, Cathy Roheim, ed. Product certification and ecolabelling for fisheries sustainability. Vol. 422. *Fish and Agriculture Association*. 2001.
- West Coast Rock Lobster Harvest Strategy and Control Rules 2015 - 2019. 2013. Government of Western Australia Department of Fisheries. *Fisheries Management Paper* no. 263.
- Worm, B., Hilborn, R., Baum, J. K., Branch, T. A., Collie, J. S., Costello, C., ... and Zeller, D. 2009. Rebuilding global fisheries. *Science* **325**: 578-585.
- Zimmermann, H.J. 2001. Fuzzy set theory-and its applications. *Kluwer Academic*, Boston.