

Fishery improvement projects: Performance over the past decade

James Cannon^a, Pedro Sousa^{a,*}, Isidora Katara^{a,b}, Pedro Veiga^a, Braddock Spear^a, Douglas Beveridge^a, Tracy Van Holt^c

^a Sustainable Fisheries Partnership Foundation, 4348 Waialae Ave. #692, Honolulu, HI 96816, United States

^b Centre for Environment, Fisheries and Aquaculture Science, Pakefield Road, Lowestoft, Suffolk NR33 0HT, United Kingdom

^c Center for Sustainable Business, Stern School of Business, New York University, 40 West 4th Street, New York, NY 10012, United States

ARTICLE INFO

Keywords:

Fishery improvement project
FIP
FishSource
Management
Supply chain
Sustainability

ABSTRACT

Fishery improvement projects (FIPs) are multi-stakeholder platforms for engaging retailers, importers, processors, and others in seafood supply chains directly in the policy-making and management of fisheries. FIPs vary in design and aim, making their evaluation complex. Studies to date have highlighted successes but also raised concerns about the performance of FIPs in improving fisheries. Drawing on a comprehensive dataset of attributes on all public FIPs, combined with sustainability performance data on the management of the target fisheries, their fishing levels, and stock status, this paper evaluates the performance of FIPs worldwide on improving fisheries, using exploratory data analysis methods and regression-based statistical approaches. The results showed that FIPs improved critical problems in target fisheries in the range between 60% and 82%, depending on the sustainability criteria considered. Performance did not vary between artisanal and industrial FIPs or according to the economic development status of the country. The probability of achieving improvements in management and overfishing domains is higher for fisheries with FIPs compared to those without. Variability in performance was related to the specific characteristics and history of each FIP, based on which further steps in research were suggested.

1. Introduction

Fishery improvement projects (FIPs) rapidly expanded over the past decade, but academic research into their performance on addressing sustainability issues is still scant. Individual case studies have analyzed the contribution of FIPs in specific fisheries [1–4] or in a small number of similar fisheries [5,6]. A broader study of the FIP model and its performance has been carried out using relatively coarse measures of progress [7].

This paper evaluates the performance of all publicly reported FIPs globally in rebuilding biomass, reducing fishing mortality levels, reducing illegal fishing, aligning quotas set by managers with those advised by scientists, and introducing precautionary harvest control rules (HCRs) that mandate reductions in fishing mortality at low biomass levels. The questions investigated were:

- Did FIPs improve fisheries?
- Did FIP performance vary depending on whether the fishery was artisanal or industrial, or in countries at different levels of economic development?

- Did FIPs improve critical problems in the fisheries?
- How fast did FIPs improve fisheries with critical problems?
- Did fisheries with FIPs improve more than those without?

A comprehensive database on FIP attributes, progress, and sustainability performance was developed for all FIPs known to have been active at some point in the past decade, based on FIP public records and their respective sustainability indices from FishSource.com,¹ typically updated on an annual basis, which were used to analyze FIP progress. This is the first study on which specific measures of fisheries' sustainability are used to evaluate FIP performance at the global scale.

1.1. Background: the origins and diversity of FIPs

In 2002, fishery improvement partnerships were introduced as a multi-stakeholder platform for engaging retail and restaurant seafood buyers and their suppliers as partners directly in the policy-making and management of fisheries they sourced from [8]. These improvement partnerships focused on fisheries important to international supply chains, meaning they were often large and globally significant sources

* Corresponding author.

E-mail address: pedro.sousa@sustainablefish.org (P. Sousa).

¹ FishSource.com, a program by Sustainable Fisheries Partnership Foundation, is an online website with information on status of fisheries and fish stocks.

Table 1

FishSource score indices and fisheries status and management criteria, typically updated on an annual basis, with underlying principles and rules of measurement for each score. FishSource scores were used for measuring fisheries' sustainability performance.

Score	Question	Principle	Quantitative measure
Harvest Strategy	Is the management strategy precautionary?	Harvest rates (F, fishing mortality) should be reduced when biomass is very low.	$F_{\text{at low biomass}}/F_{\text{target}}$
Management	Do managers follow the scientific advice?	The catch limits (TAC) set by the managers should align with the catch limits advised by the stock assessment.	$TAC_{\text{set by managers}}/TAC_{\text{advised}}$
Compliance	Do fishers comply?	The actual catches should not be higher than the catch limits set by managers.	$\text{Catch}/TAC_{\text{set by managers}}$
Stock Health	Is the fish stock healthy?	Current stock biomass (B) should be higher than the target biomass.	$B_{\text{current}}/B_{\text{target}}$
Overfishing	Is the stock overfished? Will the fish stock be healthy in the future?	Current fisheries mortality should remain at or below the fishing mortality set as a target mortality by the stock assessors or managers.	$F_{\text{current}}/F_{\text{target}}$

of seafood but their future was at risk because of poor fisheries management [9].

Major seafood buyers supporting these early FIPs described the strategy as “fix the worst first,” meaning prioritize engaging the worst performing fisheries in their supply chains, and within those fisheries, focus improvement efforts on the worst problems (e.g., [10]). These FIPs typically focused on urgent issues (such as rebuilding depleted stocks) and postponed other needed improvements until adequate progress had been made on the top priority issues. These early FIPs typically focused on larger fisheries within existing supply chains that were prioritized for action by buyers based on their greater commercial importance, were almost all large in scale and sought to cover the entire biological stock and management unit (e.g., Russian pollock, Barents Sea cod FIPs).

As other organizations adopted and adapted the FIP concept, different models emerged that varied according to a range of factors. California Environmental Associates identified four key factors [11]: (1) structure (*basic*, i.e., focused on one or two serious problems, versus *comprehensive*, i.e., working on all problem areas); (2) main lead (either by industry or a non-governmental organization (NGO)); (3) fishery status (i.e., improving a fishery with significant problems or celebrating a relatively “good” fishery with the intent of helping it rapidly achieve certification (e.g., Marine Stewardship Council (MSC))); and (4) the presence or absence of international supply chain engagement. FIPs also varied significantly in their scale, from small FIPs run by individual companies on only a few vessels or a small geographical portion of a fishery, up to large FIPs involving all the main producers and supply chain companies, and covering the entire biological or management unit of the fishery.

The term “fishery improvement project” (FIP) was adopted in 2008 to encompass this diversity of FIPs worldwide [12]. A formal definition was agreed in 2012 by the main NGOs engaging seafood buyers and supply chains [13]. FIPs work to improve fisheries that are themselves highly diverse. Fisheries vary according to a number of factors, often interlinked, such as their starting conditions in terms of management and sustainability (e.g., stock status and quality of existing science, monitoring, and enforcement), the magnitude of total annual catches, their importance to national policy-makers, heterogeneity of gears and fleets, the number of management jurisdictions, fisheries management budgets, ecological complexity, and social problems. Such factors can have a bearing on the success of fisheries management [14–17], and hence on the likelihood of success of a FIP, the speed with which improvements can be made, and the time it may take to raise the fishery to high levels of performance.

2. Data and methods

2.1. Data

A comprehensive global database of a total of 109 FIPs publicly reported as active at some point in the past decade (as per June 2016) was compiled with more than 60 different attributes per FIP (SI Tables

1, 2) including fisheries sustainability performance indices on management and stock status as well as external factors that could be influential upon FIP performance. The FIPs included in the database all publicly reported in conformance with Conservation Alliance for Sustainable Seafood guidelines and definition for a FIP [13]. Projects self-describing as FIPs, but not conforming with the Conservation Alliance for Sustainable Seafood definition and public reporting guidelines, were excluded from the database (e.g., development-agency-funded programs from the 1980s and 1990s that used the term “fishery improvement project” but did not include significant involvement of international supply chains, as described by [18–20]).

The FIPs' sustainability performance data was derived from FishSource, a global database of fisheries maintained by the Sustainable Fisheries Partnership Foundation that holds data on fisheries characteristics, related sustainability assessments, and associations (or lack thereof) to FIPs (SI Table 3). FishSource scores, which rate fisheries management and stock health (details in Table 1; [21]), are available as time series for most fisheries profiled on FishSource, not just fisheries with FIPs. In cases where a single FIP operated on multiple fisheries, the time series of scores were constructed from the lowest scores across all the fisheries within the scope of the FIP. The current analysis of fisheries performance was limited to the five FishSource scores. Even though these are good indicators of how fisheries are doing overall, their focus is on management quality and stock health. For fisheries where the main sustainability issues link to environmental (e.g., significant bycatch levels, impacts on vulnerable bottom habitats, etc.) and social impacts, the current analysis will not detect any changes in sustainability performance.

When quantitative measures cannot be derived, due to either a lack of publicly available data or an unusual assessment or management system, information may still be available to allow a qualitative response to each of the scores' underlying questions. Qualitative scores are obtained by using cut-off points: “< 6” refers to a high-risk condition, indicating a negative reply to the specific question being asked; “≥ 6” to a medium-risk condition, indicating that although not “high risk,” improvements are required on the specific matters being addressed by the question; and “≥ 8” to a low-risk condition, indicating an affirmative reply to the specific underlying question. Determining qualitative scores is always associated with some inherent subjectivity, as opposed to quantitative scores where calculations are fixed and based on unequivocal rules (for more information on FishSource scores, data pre-processing, and metadata, see SI).

External factors that could influence FIP performance and were considered in the analysis included: seafood sectors, because they dictate strategic goals for NGOs, the industry and the supply chain (SI Table 4); macro-geographical regions [22], because they can be indicative of differences in governance and culture and sustainability strategies and goals are in general organized around regions; international Human Development Indicators [23], since FIP performance has been argued to differ based on the human, economic and social development status of the country [7]; and fleet-type characteristics, because these fundamental differences may affect how fisheries

management can be implemented (artisanal, industrial). The starting status of the fishery was also considered, particularly whether the fishery was suffering from critical issues at FIP start, which would be highlighted by scores below 6 on FishSource (e.g. if catch surpassed the quota by 50%, a Compliance score of 2 is retrieved). FishSource scores range continuously from 0 to 10, with less than 6 denoting critical issues. The range between 6 and 8 (exclusively) represents the absence of critical issues but with room for improvement, while scores greater than or equal to 8 denote a fishery in good condition against the respective sustainability criterion measured by the score.

The data were collected and cross-validated, using publicly available FIP records and contacting FIP implementers to fill in data gaps and correct any errors in the history of the FIP (e.g., start dates, participants, etc.). The database allowed the evaluation of the performance of FIPs as a whole and the sustainability performance of groups of FIPs with similar characteristics, testing the hypothesis that specific FIP traits could affect the pace and magnitude of achieved improvements. The sample was limited to 57 FIPs (120 fisheries associated to FIPs) that had complete data for two or more years (i.e. “mature”), to achieve a balance between an appropriate sample size and time scales long enough for the FIP to achieve improvements. All FIP categories, as defined by the factors referred above, were represented in the data subset (SI Fig. 1).

FishSource provided data for a control group of fisheries not associated with FIPs, consisting of a sample of about 470 fisheries covering all seafood sectors and a wide range of macro-geographical regions. While these controls were not associated with FIPs, they were all sources for the international supply chain, and hence are not a random selection of the world's fisheries.

2.2. Methods

2.2.1. Did FIPs improve fisheries? Did FIP performance vary by fisheries subsets? Did FIPs improve critical problems?

FishSource scores at the beginning and end (or current date, for active FIPs) of the FIPs were compared using star plots [24] for all FIPs and for subsets of fisheries with FIPs: according to the fleet type (artisanal vs. industrial) and development (high or very high United Nations Development Programme – UNDP index vs. low or medium UNDP index). Star plots were also derived for those FIPs with starting FishSource scores below 6, where a FishSource score below 6 denotes a critical problem and a high-risk fishery. The differences between scores at the beginning and end of the FIPs (i.e., before and after the FIP) were tested using the two paired samples Wilcoxon signed-rank test (H_0 : median difference = 0; [25]) for all FIPs and for the group of FIPs with critical problems (FishSource scores < 6) at the start of the FIP. Statistical testing of differences was not carried out for other subsets given their small sample sizes.

2.2.2. How fast did FIPs improve fisheries, especially those with critical problems?

Linear regression lines were fitted to the time series of each FishSource score for each FIP and the slopes for each line were derived:

$$FSS_{ij} \sim a + b_{ij} \text{year} \quad (1)$$

where FSS stands for a FishSource score, i denotes one of the five FishSource scores and j one of the 57 FIPs. A positive slope (trend, b_{ij}) shows an increase of the score and thus implies achieving an improvement over time. Annual increments in slopes were used to derive estimates of improvements in the actual variables the scores are based on (see SI for calculations). Histograms of the slopes for all FIPs per FishSource score showed the frequency of the trends that are indicative of the number of FIPs that achieved improvements and of the rate of change for the majority of FIPs (high slopes denote a higher rate of change). The subset of slopes for FIPs with critical issues at the start of the FIP was superimposed to the histograms for comparison purposes

against the dataset with all FIPs.

2.2.3. Did fisheries with FIPs improve more than those without?

To answer the question of whether FIPs make a difference in achieving improvements towards sustainability, a comparison was made between the progress of fisheries with FIPs and the progress of fisheries without (the control group). Progress was measured using two indicators: (i) the probability of regression slope (trend) of a FishSource score of a fishery being positive, where the slope is calculated using a regression model

$$FSS_{ik} = a + b_{ik} \text{year} \quad (2)$$

with FSS_{ik} being the FishSource score i (with $i = 1-5$) of fishery k ; and (ii) the probability of the difference of the FishSource score at the beginning and end of the time series to be positive, where the difference is calculated as

$$\text{diff. } FSS_{ik} = FSS_{ik/END} - FSS_{ik/START} \quad (3)$$

The two indicators were used as response variables in a linear regression model

$$P(\text{trend}_{ik} > 0) \sim a + d_0 * FIP + b_1 * ps, \text{ weights} = w_{ik} \quad (4)$$

$$P(\text{diff}_{ik} > 0) \sim a + d_0 * FIP + b_1 * ps, \text{ weights} = w_{ik} \quad (5)$$

$P(\text{trend}_i > 0)$ is the probability of the slope for the time series of FishSource score i to be positive, i.e., increasing trend of the score, $P(\text{diff}_i > 0)$ is the probability of the difference of the FishSource score i between the beginning and end of the time series being positive, FIP is 1 for fisheries with FIPs and 0 for the control. ps refers to the propensity score for the fishery obtained by fitting a binomial regression (see Eq. (6) below for details), w_{ik} is the length of the time series of FishSource score i for fishery k , and d_0 is the difference between FIP-associated fisheries and the baseline (control). The Bonferroni correction was applied to avoid erroneous inferences due to multiple comparisons (a total of 10 in this case).

Given that our study is observational and hence non-randomized [26], propensity scores were employed to decrease the bias induced by confounding factors (see [27] on propensity scores; [28] on how to adjust for the covariates; and [29–31] for examples in fisheries). The propensity score is the conditional probability of a fishery being associated to a FIP, given a vector of observed covariates (confounding factors, see below). Eq. (6) gives the binomial regression used to calculate the propensity scores:

$$P(\text{fishery} \in \text{FIP}) \sim f(\text{confounding factors}) \quad (6)$$

where the dependent variable is the probability of a fishery being associated to a FIP or not ($P(\text{fishery} \in \text{FIP})$) and $f()$ is a linear function of the *confounding factors*.

The confounding factors that were considered were:

- 1) Seafood sector: to account for imbalances in the number of fisheries belonging to each sector for fisheries with FIPs and for those without.
- 2) UN region: used as a grouping factor concerning conservation goals and policies. Adding this variable to the model increased the number of parameters to estimate and led to model convergence issues, so the variable was dropped.
- 3) UNDP index and UNDP index-based categories: to account for possible differences between developed and developing countries such as level of awareness of sustainability problems, capacity to solve environmental problems, differences in scientific output.
- 4) Being related to a Supply Chain Roundtable: because supply chain roundtables use their leverage to initiate FIPs for fisheries of their interest.
- 5) Having critical problems (i.e. one or more FishSource scores < 6) at the beginning of the time series: the variable was considered

because the framework of FIPs aims to deliver improvements in fisheries with serious sustainability issues, and often stakeholders prioritize these fisheries.

- 6) Latitude: This variable is often used as a proxy for a number of unobservable factors. Different levels of biodiversity [32,33] and ecosystem complexity relate to latitudinal gradients [34,35] and pose challenges to scientists and managers. The characteristics of fisheries also likely to vary with latitude.
- 7) Species catch/total catch of the fishing sector, at country level when possible or at the FAO area level when information at the country level was not available: This variable was used as an indicator of the relative importance of the fishery against the country's production. A relatively small fishery in the global landscape can be a key fishery for a given country or region, and thus very important for the country's government to manage well.
- 8) Global catch of the target species: as an indicator of the importance of the fishery at a global scale.

Model selection was based on the framework provided in Brookhart et al. [36] and on a stepwise regression, with the full model including 2-way interactions between confounding factors. The final model was selected using the Akaike and Bayesian Information Criteria [37] aiming for a realistic and parsimonious model. The model was fitted to derive the propensity score for each fishery. The propensity scores are then used for covariate adjustment (Eqs. (4) and (5)). All graphical and statistical analyses were conducted using the R language (version 3.3.0; [38]).

3. Results

3.1. Did FIPs improve fisheries?

The star plots of the average of the five scores from the beginning to the end of FIPs, for all FIPs, showed increase or stability for all sustainability criteria as measured by FishSource scores (Fig. 1a), but the Wilcoxon test showed that only the Harvest Strategy score had a statistically significant increase (p -value < 0.001; SI Table 5) when considering all FIPs.

3.2. Did FIP performance vary by fisheries subsets?

The star plots of FishSource scores suggested positive differences between the beginning and end of FIPs (or current state) for all subsets – artisanal and industrial fleets, Low or Medium UNDP index, and High or Very high UNDP index (Fig. 1b). The most difference was for FIPs for artisanal fleets, and for FIPs from developing countries (Low or Medium UNDP index) (Fig. 1b). The increasing pattern was observed for all scores except for Stock Health. Statistical testing was not rendered possible for FIPs subsets given small sample sizes. A closer look at the data regarding the score on Stock Health for artisanal FIPs showed that the overall slightly decreasing trend can be attributed to artisanal shrimp FIPs in the Gulf of California. Despite the decrease in the average, all scores for stock health remained above six for artisanal FIPs, with the lowest end-point observed being 6.6.

3.3. Did FIPs improve critical problems?

A clear pattern of improvement in all five FishSource scores was observed for scores from FIPs that were below six at the FIP start (critical issues) (Fig. 1c). The observed improving pattern of critical issues by score was as follows: Harvest Strategy – 62%, Management – 72%, Compliance – 82%, Stock Health – 64%, Overfishing – 60%. The Wilcoxon test showed that all score differences between start and end of FIP were statistically significant (p -value < 0.001) with the exception of the score on stock health (SI Table 6).

3.4. How fast did FIPs improve fisheries with critical problems?

Linear regression lines for each FIP and histograms of the respective regression slopes are presented in Fig. 2 (left and right columns, respectively) for all five FishSource scores (rows of Fig. 2). Regression lines are trimmed to the period of activity of FIPs. The distribution of slopes for FIPs on fisheries that started with critical problems (scored < 6 at the start of the FIP) is shown overlapped in red in histogram plots. Medians for all FIPs and for those with critical issues at the start are shown in histogram plots as vertical lines.

With the exception of Stock Health, all medians of slopes for critical issues are clearly greater than zero, demonstrating clear improvement patterns. Such an outcome is not observed when considering all FIPs combined (grey histogram bars and grey vertical lines standing for Medians: in that case, only the Harvest Strategy median is notably greater than zero).

Translating the patterns observed only for FIPs with critical issues at the start into effective improvements, a positive median of slopes of 0.4 for Harvest Strategy stands for a 10% annual reduction in the anticipated fishing mortality at low biomass levels compared to the target fishing mortality, a considerable improvement. Similarly, a median of slopes of 0.6 for Management indicates a general pattern of reduction of 3.75% of the set total allowable catch compared to advised levels. A median slope of 0.8 for the Compliance score indicates that actual catches reduced by 5% relative to the set TAC, a significant reduction in overshoot of the quota and potential illegal overfishing. A median slope of 0.25 annually for the Overfishing score means a reduction in fishing mortality of 6.25% compared to target (see SI for detailed calculations).

3.5. Did fisheries with FIPs improve more than those without?

The comparison between fisheries with and without FIPs showed those with FIPs have a statistically significantly higher probability of achieving improvements regarding Management and Overfishing compared to fisheries without FIPs (Table 2, p -values < 0.001). For the other FishSource scores the difference between the two groups was not statistically significant (p -values > 0.05).

Fisheries related to FIPs show a significantly higher probability of a positive difference between the scores on Overfishing at the beginning and end of the time series. For the score on Management, both comparisons (first using the FishSource score trend as the dependent variable and second using the difference between the score at the beginning and end of the score time series) showed a statistically significant difference. The propensity scores used to standardize the data were based on three confounding factors: the seafood sector (p -value < 0.001), the UNDP index-based categories (p -value < 0.001), and latitude (2nd order polynomial; p -value < 0.001).

4. Discussion and conclusions

4.1. The impact of FIPs on fisheries with critical problems

Our results showed that FIPs made statistically significant improvements in fisheries with critical problems in Management (Harvest Strategy, Management, and Compliance scores) and Overfishing score. This supports the intended principal goal of FIPs of improving problem areas rather than improving areas where fisheries are already performing well.

The absence of a statistically significant result for rebuilding depleted fish stocks is potentially caused by two factors: (i) target Stock Health was, in general, a less severe issue at the start of FIPs compared to Management or Compliance, and (ii) FIPs have not had enough time yet to rebuild stocks. Depleted stocks are typically the last thing to improve because for stocks to start to rebuild, fishing mortality must, in general, be reduced first (as measured by Overfishing score). For fishing mortality to be reduced, typically there needs to be a harvest control

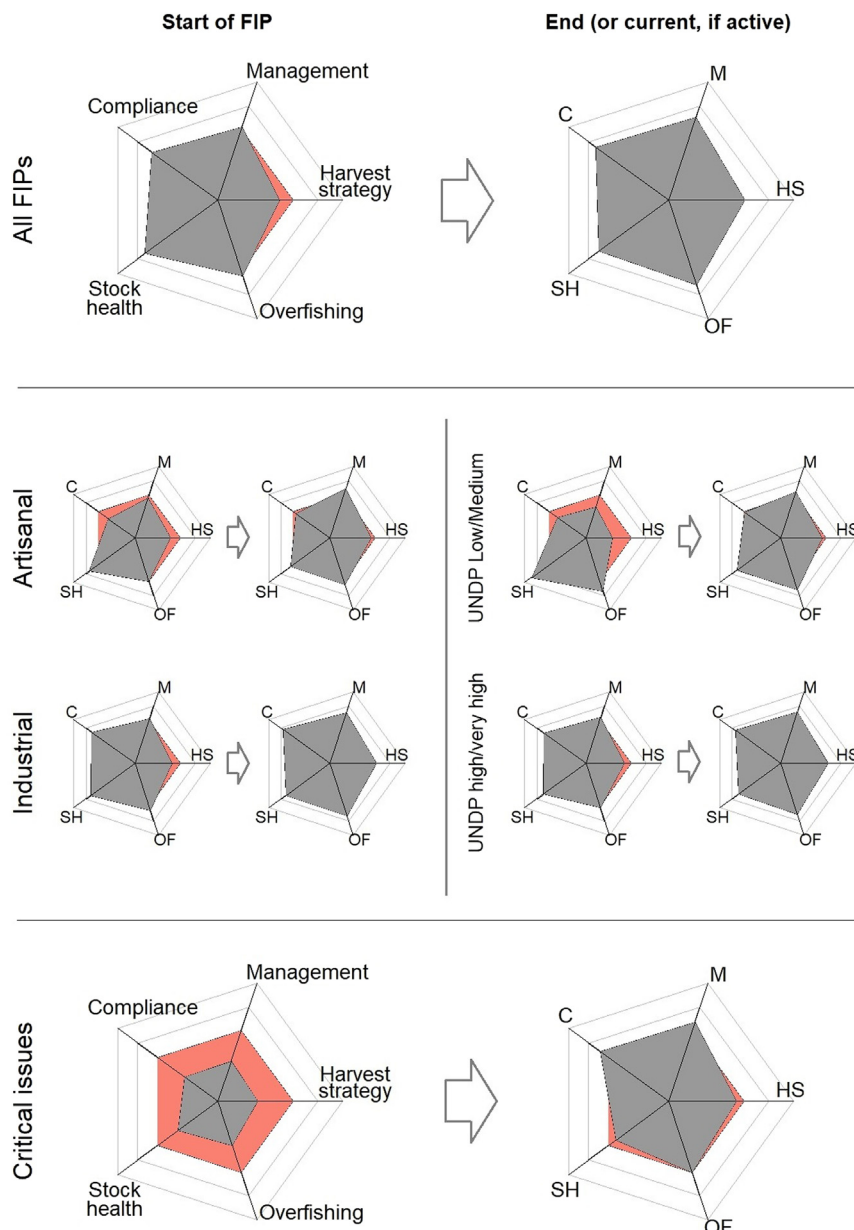


Fig. 1. Star plots showing average starting and current or end FishSource scores of FIPs. Top row: all FIPs analyzed; Middle row: FIP fishery categories artisanal, industrial, low or medium UNDP country index, high or very high UNDP country index; Bottom row: FIPs with critical issues (FishSource scores < 6) at the start of the FIP. Each vertex of the pentagon represents the average FishSource score. For the “start” star plot, the grey polygon is the average of all scores at the beginning of the FIPs. For the “end” star plot, the grey polygon is the average of all scores at the end of the FIPs for the same FIPs that are depicted in the left-hand star plot. The red polygons highlight the area corresponding to scores < 6 and are used in all plots to facilitate visual comparison. The radii to which polygons are superimposed correspond to guideposts for scores = 6, 8, and 10 (the outermost radii). The number of FIPs depicted for each score varies due to data availability and/or number of FIPs in each category (e.g., there were 26 FIPs with a score < 6 for Harvest Strategy at the start of FIPs, and 17 FIPs with a score < 6 for Compliance at the start of FIPs).

rule or equivalent in place that calls for reduction in fishing mortality at low biomass (Harvest Strategy score), and managers must follow the scientific advice and actually aim to bring fishing mortality down, e.g., by setting quotas in line with advised catch levels (Management score), and fishers must comply with the limits (Compliance score). It may take time for compliance to improve, and as stricter management policies are implemented, compliance scores may temporarily get worse. In other words, it is reasonable to expect to first detect improvements in the Management (Harvest Strategy and Management scores) and the Compliance scores, then improvements in reducing fishing mortality (Overfishing score), before improvements are likely in rebuilding depleted stocks (Stock Health score) [39].

This sequence means improvements in the Stock Health score in depleted fisheries may only be detected in FIPs that have been active for many years, and potentially have completed their work. Similarly, a reasonable period of time may be needed to improve data collection and overcome data deficiencies, or for fisheries to improve compliance, which may even get worse first in response to stricter regulations. However, only 17 FIPs have concluded and 70% of the FIPs in the database are still active (i.e. have not yet finished their work) and the

average lifetime of FIPs is only 6 years (with an overall range in age for all observed FIPs from 2 to 9 years).

Many FIPs, therefore, may not have been operational for a sufficient time period to have completed inherently slow processes such as changes in fishers' behavior or dealing with data deficiency. Future analyses will benefit from longer time series and be able to confirm whether FIPs ultimately improved compliance with stricter regulations and rebuilt depleted fish stocks.

4.2. Setting reasonable expectations and deadlines for FIPs to deliver

While more time may be needed to resolve the questions above, our results provide useful insight as to how fast improvements of critical problems typically occur. Fisheries that started with critically low scores related to management (Harvest Strategy, Management, and Compliance scores) showed the most rapid pace of improvement. This is perhaps to be expected since improvements in management can be done through changes by regulators or policy-makers to written rules (e.g., changing the F target at low biomass in a harvest control rule). In contrast, reductions in F are typically introduced slowly and

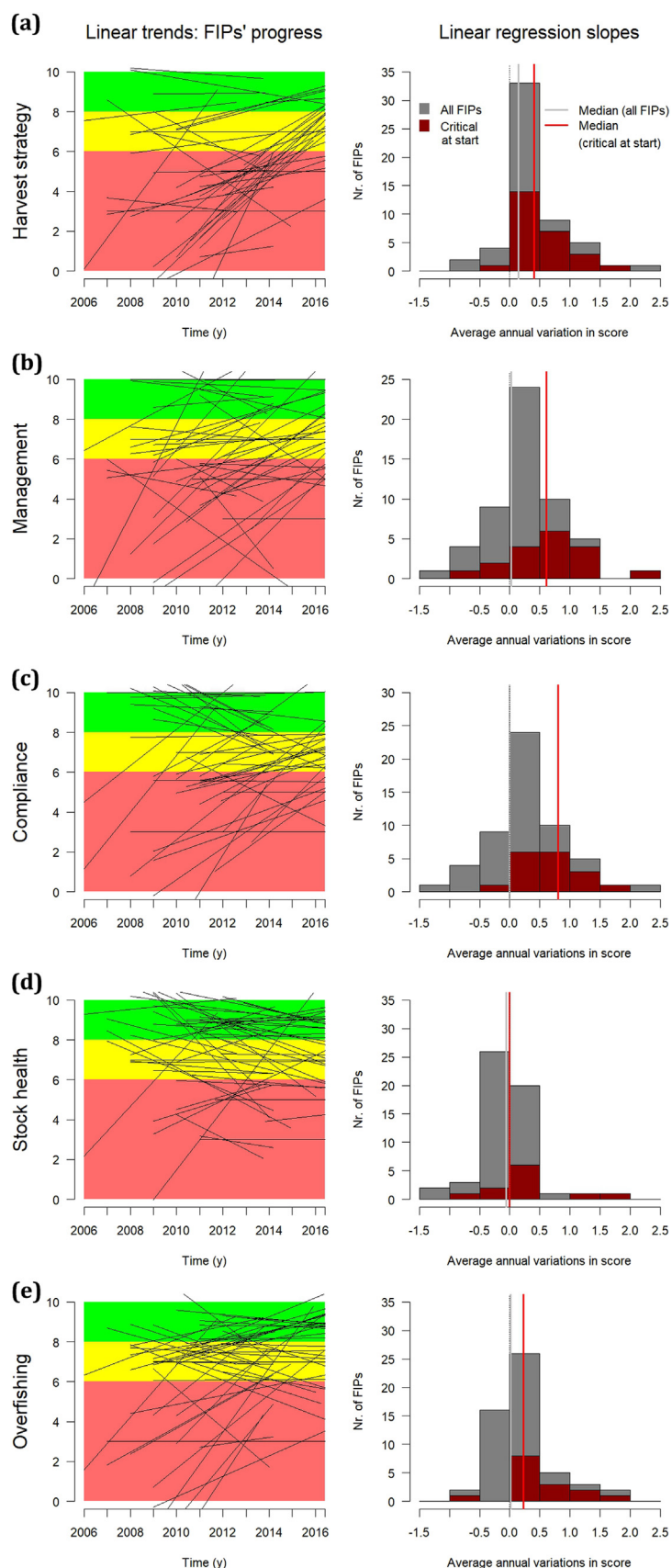


Fig. 2. Changes over time of the FishSource scores – linear trends (left column) and histograms of the linear regression slopes (right column): Harvest Strategy, Management, Compliance, Stock Health, Overfishing (rows). Predictions for some FIPs can extend beyond the 0–10 range because each regression line is fitted for the respective period of activity of the FIP. The subsets of slopes for FIPs with a score < 6 at the beginning of the FIP (denoting critical issues) are overlapped in the histogram plots (red). Medians for all FIPs (grey) and for only those FIPs with critical issues at the start (red) are displayed as vertical lines in the histogram plots.

Table 2

Regression coefficients showing the difference between fisheries with FIPs and those without, with corresponding p-values. Each comparison is based on a measure of the change of a FishSource score over time. All five scores and two indicators of change (trends and the difference between start and end of time series) are considered.

Score	Change indicator	Regression coefficient	P-value
Harvest Strategy	slope	0.17	> 0.1
Management	slope	1.36	< 0.001
Compliance	slope	0.19	> 0.1
Stock Health	slope	0.22	> 0.1
Overfishing	slope	0.69	> 0.05
Harvest Strategy	difference	0.09	> 0.1
Management	difference	1.14	< 0.001
Compliance	difference	0.07	> 0.1
Stock Health	difference	0.16	> 0.1
Overfishing	difference	0.7	< 0.001

incrementally (Overfishing score) and, as noted above, typically require improvements in Management scores first. Increases in biomass (Stock Health score) depend on environmental and other factors, and in depleted fisheries increases in biomass typically require prior reductions in fishing mortality.

One of the concerns raised about FIPs is whether they are making rapid enough progress, and over how many years it is reasonable to expect improvements to take place [13]. Our results indicate an average increase of 0.6 score units per year for Management-related scores (Harvest Strategy, Management, and Compliance scores) as a preliminary estimate. Further analysis is required to determine realistic rates of improvement in fisheries, and hence to set deadlines or timetables for specific fisheries to achieve target levels of performance. For all FishSource scores, the most rapid increases in scores were up to four times the median. Further research to understand why some FIPs made such rapid progress may indicate ways to improve the performance of all FIPs.

4.3. The performance of FIPs in general, and for artisanal and developing country fisheries

Very few FIPs allowed critical problems to emerge, as indicated by scores falling below 6 during the FIP. However, some high scores (in the range of 8–10) did show declines (e.g., into the 6–8 range). Of particular concern are the declines in Compliance scores, though these may be temporary and potentially related to the introduction of stricter management regulations, as noted above.

For all FIPs combined, a slight decrease of Stock Health was observed, but the respective observed score average remained ≥ 6 , which may be regarded as a medium risk. Natural fluctuations over time are expected in biomass, but the declines observed for some tuna and South American stocks were substantial and lowered the average of the entire FIP sample. Additional research is required to understand the factors behind these specific declines.

The method and results indicate the crucial importance of measuring FIP performance against the impacts they need to achieve: improving critical problems. However, even when considering all FIPs working on all aspects of the fishery, including aspects where the fishery was already performing well, the vast majority of FIPs (93%), at the very least, maintained levels of performance, or delivered improvements in Harvest Strategy.

This pattern was observed for artisanal FIPs and FIPs from developing countries. Although small sample sizes did not allow for statistical inference for the subsets, average increases in all five scores were observed between the start and end of FIPs for both those subsets. Hence no evidence was found to support the conclusion of Sampson et al. in [7] that overall FIPs perform worse in developing countries. The disagreement between the results of this study and those of

Sampson et al. [7] are likely due to differences in measures of performance, and differences in how improvements in a single component of a multi-fishery FIP are weighted. In a FIP with n different fisheries, Sampson et al. [7] weighted improvements in a single component as $1/n$. In this study, however, an improvement in a single component was weighted as 1. Regarding measures of performance, this study used fisheries status and Management scores from FishSource.org, whereas Sampson et al. [7] used FIP “stages.” FIP stages were developed by Sustainable Fisheries Partnership (SFP) in 2009 to help seafood buyers distinguish between FIPs that had already made improvements from those that had not, based on self-reporting and public evidence. FIP stages do not incorporate information on the specific type of improvement made, the magnitude of improvement, or how recently or regularly improvements are being delivered (see [40] for more information on FIP stages). As such, they are a coarser measure of FIP progress compared to FishSource scores.

4.4. Fisheries with FIPs improve more than those without

Statistically testing FIPs against an appropriate control group was complicated by small sample sizes. Modeling was rendered inapplicable for testing only FIPs working on critical problems against a suitable control, and hence the sample size was expanded by considering all FIPs and how they performed on all scores. Hence, the impact of all FIPs on all scores was considerably lower, and thus less likely to out-perform a control group. The final sample, with all observations, regardless of the condition of the fishery at the start of the FIP, was balanced, with a similar number of observations in both groups (fisheries with FIPs and those without – SI Fig. 3).

The control group used—non-FIP fisheries on FishSource—is heavily focused on fisheries supplying international markets. These fisheries, despite not being in FIPs, have nonetheless been exposed to demands for improvement from international customers and may have responded even without formal publicly reporting FIPs [41]. This would further decrease the expected difference in performance between fisheries with FIPs versus those without. Nonetheless, our results showed fisheries with FIPs showed a significantly higher probability of improving Management and reducing Overfishing than those fisheries without FIPs.

The difference in improving precautionary management rules was not statistically significant. Adopting precautionary harvest strategies is a logical way for improvement efforts to start, and it may hence be the case that those took place to a large extent within the control group similarly as observed for the FIP group, thus rendering any comparison between the two improving patterns not statistically significant.

This study has not delved into which initiatives other than FIPs may have been playing an important role in improving fisheries (e.g., managers’ own led initiatives in improving management and fisheries science, or other non-FIP-related improvement initiatives). The ideal control group would be entirely unrelated to any improvement initiative, however, the information was lacking and such a selection of the control group was not possible. Despite this limitation of the data, based on the analysis of comparison of FIPs with the control group, our results showed that the FIP model is an efficient tool to address management issues and to mitigate the risk of overfishing.

4.5. Limitations of methodology and next steps in research

The evaluation of FIP progress was based on FishSource scores. The following characteristics of the FishSource scores should be accounted for when interpreting the results of this study. First, many FIPs were formed to deal with environmental problems (e.g., high bycatch levels, impacts on habitats), rather than with management, stock health or overfishing issues, which are the principles currently captured by FishSource scores. Extending the database to include environmental scores is needed to evaluate FIPs comprehensively.

Second, qualitative FishSource scores are inherently subjective, as opposed to quantitative scores where there are fixed and strict rules in the calculations. Also, they include three cut-off points in terms of fisheries performance (< 6 , $6-8$, ≥ 8), and are thus not able to capture finer-resolution changes.

The quantitative approach in this study was constrained to the coverage of the FishSource database. Despite being one of the largest fisheries databases globally, FishSource is far from covering all fisheries in the world. For some UN regions, the FishSource coverage is very limited, mainly driven by the absence of public information on those fisheries in the first place. Thus, the results from the current study may suffer from a bias toward more data-rich fisheries. As the FIP framework expands, more information on the status of the other fisheries will become available, and it is likely that any coverage-related bias will be mitigated in future studies.

Our method to evaluate the speed of improvement in scores used simple linear regression models. Neither the variability of scatters of FishSource scores over time nor the statistical leverage of individual points upon the trend was explicitly considered in hypothesis testing. Especially the scores on Compliance and Management show high inter-annual fluctuations that the estimated trends cannot capture. The estimated trends can be “smooth” representations of reality that may under-represent score declines or increases over time or misrepresent real trends due to influential points.

The FIP model is a fairly new concept, and there are still a number of research questions that persist and pertain mainly to the factors that ensure FIPs’ success in improving fisheries. In particular, further research is needed to understand which factors may be preventing fisheries worldwide that have critical conditions from entering into FIPs. The FIP framework is very broad, leading to high levels of complexity, and more work is needed to decipher patterns behind FIPs dynamics. *Is the diversity of FIPs responsible for the varying performance we observe?* The current study makes no distinction between FIPs based on their quality or how they were set up. A complementary approach to the one followed in this study could focus on the most impactful and data-rich FIPs and examine the factors that contributed to their success. *Which FIP practices were most effective in creating impact? In what specific context (e.g., regional, political situation, type of fishery)? What was the level of supply chain involvement?* As FIPs become a preferred framework towards sustainability and reporting is improved, new datasets will allow scientists to identify the key factors and conditions that render a FIP successful. Future research should also include data on the fisheries production by value and global trade patterns. These factors can potentially be relevant to model the probability of a fishery being associated to a FIP, as well as an indication of market leverage as an integral part of the FIP model [42] and an important factor for the efficiency or success of a FIP. Another important layer of information is the inclusion of environmental and socioeconomic impact indicators, which would help assess FIPs that prioritize problems with habitat degradation and bycatch or FIPs with interlinked socioeconomic and sustainability issues.

Acknowledgements

We thank: Merul Patel, Priyanka Pathak and Ananta Mambo (Sustainable Fisheries Partnership – SFP) for assistance on data handling and database management; Ian Heath (SFP) for management assistance of the SFP FIP Research Program; Jessica Beels and Amy Sweeting (SFP) for proof reading and editorial review of the paper; Christiane Schmidt, Indrani Lutchman and Marina Mendes (SFP) for data collation and validation; Alexia Morgan, Indrani Lutchman, Jose Parajua, Mariana Bock Megan Westmeyer, and Patrícia Amorim (SFP) for providing valuable feedback on earlier versions of the manuscript. We are thankful to Professor Ray Hilborn and to Dr. Mike Melnychuk (School of Aquatic and Fishery Sciences, University of Washington) and to Dr. Beatrice Crona (Academy Program on Global Economic Dynamics

and the Biosphere, The Royal Swedish Academy of Sciences) for their valuable suggestions on the research leading to the final version of the paper.

Funding

This research was funded in part by a collaboration established by the Gordon and Betty Moore Foundation (005107-2.3FIP.2&3-SFP) and has also received funding support from the Walton Family Foundation (2016-685) and The David and Lucile Packard Foundation (2016-63634 and 2017-65567). Funding for Tracy Van Holt has been provided in part by The Swedish Research Council FORMAS Grant no. 2016-00375; Erling-Persson Family Foundation through the Global Economic Dynamics and the Biosphere Program of the Swedish Royal Academy of Sciences.

Declarations of interest

None.

Supplementary information

The Supplementary Information provided as a separate source contains all materials denoted by “SI” in the text.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.marpol.2018.06.007>.

References

- [1] M. Doddema, On Fisheries Improvement Projects. An Analysis of the Suitability of Fisheries Improvement Projects for Achieving Improvements in Fisheries. A Case Study of a Fisheries Improvement Project for the Yellowfin Tuna Fishery in Lagonoy Gulf, the Philippines (M.Sc. Thesis ENP 80436), Wageningen University, Environmental Policy Group (ENP), Department of Social Sciences, Wageningen, The Netherlands, 2012.
- [2] L.K. Deighan, L.D. Jenkins, Fishing for recognition: understanding the use of NGO guidelines in fishery improvement projects, *Mar. Policy* 51 (2015) 476–485, <http://dx.doi.org/10.1016/j.marpol.2014.10.009>.
- [3] D.E. Duggan, M. Kochen, Small in scale but big in potential: opportunities and challenges for fisheries certification of Indonesian small-scale tuna fisheries, *Mar. Policy* 67 (2016) 30–39, <http://dx.doi.org/10.1016/j.marpol.2016.01.008>.
- [4] G. Parkes, J.H. Swasey, F.M. Underwood, T.P. Fitzgerald, K. Strauss, D.J. Agnew, The effects of catch share management on MSC certification scores, *Spec. Issue Fish. Certif. Eco-Label. Benefits Chall. Solut.* 182 (2016) 18–27, <http://dx.doi.org/10.1016/j.fishres.2015.10.003>.
- [5] F. Tolentino-Zondervan, P. Berentsen, S. Bush, J. Idemne, R. Babaran, A.O. Lansink, Comparison of private incentive mechanisms for improving sustainability of Filipino tuna fisheries, *World Dev.* 83 (2016) 264–279, <http://dx.doi.org/10.1016/j.worlddev.2016.01.011>.
- [6] S.R. Bush, M. Bailey, P. van Zwieten, M. Kochen, B. Wiryawan, A. Doddema, S.C. Mangunsong, Private provision of public information in tuna fisheries, *Mar. Policy* 77 (2017) 130–135, <http://dx.doi.org/10.1016/j.marpol.2016.12.019>.
- [7] G.S. Sampson, J.N. Sanchirico, C.A. Roheim, S.R. Bush, J.E. Taylor, E.H. Allison, J.L. Anderson, N.C. Ban, R. Fujita, S. Jupiter, Secure sustainable seafood from developing countries, *Science* 348 (2015) 504–506.
- [8] Conservation International, Interview with Jim Cannon, *Frontlines*, 2003, 8–9.
- [9] McDonald’s Corporation, Interview with Klaus Neilsen and Helgi Eiriksson, 2004.
- [10] Sobeys, Press Release October 13th, 2010. <http://corporate.sobeys.com/wp-content/uploads/2014/08/Sobeys-Inc-Commits-to-Seafood-Sustainability.pdf>.
- [11] CEA, Summary Findings from the Global Landscape Review of Fishery Improvement Projects (FIPs), California Environmental Associates, San Francisco, CA, 2015.
- [12] WWF, FIP Handbook Guidelines for Developing Fishery Improvement Projects, WWF Oceans Program, Washington, DC, United States, 2015 (<http://seafood.sustainability.org/wp-content/uploads/2015/10/WWF-FIP-handbook-6-3-15.pdf>).
- [13] Conservation Alliance for Seafood Solutions, Guidelines for Supporting Fishery Improvement Projects, 2012, 22 p. <http://solutionsforseafood.org/wp-content/uploads/2015/03/Alliance-FIP-Guidelines-3.7.15.pdf>.
- [14] N.L. Gutiérrez, R. Hilborn, O. Defeo, Leadership, social capital and incentives promote successful fisheries, *Nature* 470 (2011) 386–389.
- [15] C.J. Lundquist, E.F. Granek, Strategies for successful marine conservation: integrating socioeconomic, political, and scientific factors, *Conserv. Biol.* 19 (2005) 1771–1778.

- [16] R.S. Pomeroy, B.M. Katon, I. Harkes, Conditions affecting the success of fisheries co-management: lessons from Asia, *Mar. Policy* 25 (2001) 197–208.
- [17] J.R. Beddington, D.J. Agnew, C.W. Clark, Current problems in the management of marine fisheries, *Science* 316 (2007) 1713–1716.
- [18] W.R. Allison, The discovery bay fisheries improvement project: a status report, in: *Proceedings of the 42nd Ocho Rios Jam*, 1989, 331–337.
- [19] S. Vatcher, The discovery bay fisheries improvement project: biological and socio-economic findings, in: *Proceedings of the 43rd Gulf Caribb. Fish. Inst. Miami Fla*, 1990, 43.
- [20] D.N. Brown, R.S. Pomeroy, Co-management of Caribbean community (CARICOM) fisheries, *Mar. Policy* 23 (1999) 549–570.
- [21] J. Cannon, FishSource scores, how they're calculated and what they represent, v.1, *Sustain. Fish. Partnersh.* (2006) (20 pp), <http://www.fishsource.org/indices_overview.pdf>.
- [22] UNSD, Composition of macro geographical (continental) regions, geographical sub-regions, and selected economic and other groupings, 2013. <<http://unstats.un.org/unsd/methods/m49/m49regin.htm>>.
- [23] UNDP, International Human Development Indicators, 2016. <<http://hdr.undp.org/en/countries>>.
- [24] J.M. Chambers, W.S. Cleveland, B. Kleiner, P.A. Tukey, et al., *Graphical Methods for Data Analysis*, Wadsworth, Belmont, CA, 1983.
- [25] W.J. Conover, W.J. Conover, *Practical Nonparametric Statistics*, (1980).
- [26] R.J. Glynn, S. Schneeweiss, T. Stürmer, Indications for propensity scores and review of their use in pharmacoepidemiology, *Basic Clin. Pharmacol. Toxicol.* 98 (2006) 253–259, <http://dx.doi.org/10.1111/j.1742-7843.2006.pto.293.x>.
- [27] P.R. Rosenbaum, D.B. Rubin, The central role of the propensity score in observational studies for causal effects, *Biometrika* 70 (1983) 41–55.
- [28] P.C. Austin, An introduction to propensity score methods for reducing the effects of confounding in observational studies, *Multivar. Behav. Res.* 46 (2011) 399–424, <http://dx.doi.org/10.1080/00273171.2011.568786>.
- [29] C. Costello, S.D. Gaines, J. Lynham, Can catch shares prevent fisheries collapse? *Science* 321 (2008) 1678–1681.
- [30] M.C. Melnychuk, T.E. Essington, T.A. Branch, S.S. Heppell, O.P. Jensen, J.S. Link, S.J. Martell, A.M. Parma, J.G. Pope, A.D. Smith, Can catch share fisheries better track management targets? *Fish. Fish.* 13 (2012) 267–290.
- [31] J.T. Thorson, J.M. Cope, K.M. Kleisner, J.F. Samhouri, A.O. Shelton, E.J. Ward, Giants' shoulders 15 years later: lessons, challenges and guidelines in fisheries meta-analysis, *Fish. Fish.* 16 (2015) 342–361.
- [32] K.J. Gaston, Global patterns in biodiversity, *Nature* 405 (2000) 220–227, <http://dx.doi.org/10.1038/35012228>.
- [33] D.P. Tittensor, C. Mora, W. Jetz, H.K. Lotze, D. Ricard, E.V. Berghe, B. Worm, Global patterns and predictors of marine biodiversity across taxa, *Nature* 466 (2010) 1098–1101, <http://dx.doi.org/10.1038/nature09329>.
- [34] A.R. Longhurst, *Ecological geography of the Sea*, Academic Press, Amsterdam; Boston, MA, United States, 2007 2nd. Ed..
- [35] R.T. Paine, Food web complexity and species diversity, *Am. Nat.* (1966) 65–75.
- [36] M.A. Brookhart, S. Schneeweiss, K.J. Rothman, R.J. Glynn, J. Avorn, T. Stürmer, Variable selection for propensity score models, *Am. J. Epidemiol.* 163 (2006) 1149–1156.
- [37] J.K. Lindsey, *Applying Generalized Linear Models*, Springer Science & Business Media, New York, United States, 2000.
- [38] R. Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2014 <<http://www.R-project.org/>>.
- [39] B. Worm, R. Hilborn, J.K. Baum, T.A. Branch, J.S. Collie, C. Costello, M.J. Fogarty, E.A. Fulton, J.A. Hutchings, S. Jennings, O.P. Jensen, H.K. Lotze, P.M. Mace, T.R. McClanahan, C. Minto, S.R. Palumbi, A.M. Parma, D. Ricard, A.A. Rosenberg, R. Watson, D. Zeller, Rebuilding global fisheries, *Science* 325 (2009) 578–585, <http://dx.doi.org/10.1126/science.1173146>.
- [40] SFP, FIP evaluation: a standardized tool for measuring FIP progress, *Sustainable Fisheries Partnership (SFP)*, 2017.
- [41] Marine Stewardship Council, The MSC at 20: Wild. Certified. Sustainable, (2017) <<https://www.msc.org/global-impacts/msc-annual-report>>.
- [42] S.R. Bush, H. Toonen, P. Oosterveer, A.P.J. Mol, The “devils triangle” of MSC certification: balancing credibility, accessibility and continuous improvement, *Soc. Cult. Impacts Mar. Fish* 37 (2013) 288–293, <http://dx.doi.org/10.1016/j.marpol.2012.05.011>.