An econometric approach to the construction of ecological indices: A case study using the food provision goal of the Ocean Health Index

by

Caitlin Nicole Millar

B.Sc., University of British Columbia, 2011

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

Report No. 622

in the
School of Resource and Environmental Management
Faculty of Environment

SIMON FRASER UNIVERSITY

Summer 2015

Approval

Name:	**Caitlin Nicole Millar**
Degree:	**Master of Resource Management**
Project No.	**622**
Title:	***An econometric approach to the construction of ecological indices: A case study using the food provision goal of the Ocean Health Index***
Examining Committee:	**Chair:** Lindsay Gardner Master of Resource Management candidate

Dr. Andrew Cooper
Senior Supervisor
Associate Professor

Dr. Sean Cox
Supervisor
Associate Professor

Date Defended/Approved: April 20, 2015

Abstract

Ecological indices summarize large sets of complex data to improve performance monitoring, benchmarking, policy analysis, and public communication. Indices, such as the Ocean Health Index, are sensitive to the aggregation method and weighting scheme used in the construction of the index. This analysis investigates the differences in the mathematical properties and aggregate behaviour of eight aggregation methods and weighting schemes, and considers how information about desired index behaviour can be used to make a final choice of the mathematical form for the index. The mathematical properties of these aggregation methods and weighting schemes are compared using the axiomatic approach to index number theory, and a case study of the fisheries sub-goal of the Ocean Health Index. The results show that the exponentially weighted geometric mean with proportional weight, and the weighted arithmetic mean with proportional weight perform the best in terms of their underlying mathematical properties and aggregate behaviour. The results highlight the importance of setting achievable objectives that inform which mathematical properties and aggregate behaviour are most desirable; from this, one can select an appropriate aggregation method for constructing indices that rely on averages.

Keywords: Index number theory; aggregation; averages; ecological index

Acknowledgements

I would like to thank Andrew Cooper and Sean Cox for their contribution to my education and this project. Thank you to Benjamin Halpern, Catherine Longo and Kristin Kleisner for their helpful feedback on the manuscript. I am thankful to my colleagues at REM for their support over the years, especially Lindsay Gardner, Michelle Jones, Amanda Schroeder and the rest of the Fisheries Group.

Funding for this project was provided by a Canada Graduate Scholarship and Discovery Grant from the National Sciences and Engineering Research Council of Canada, and the Ocean Health Index.

Table of Contents

List of Tables

List of Figures

List of Acronyms

A	Weighted Arithmetic Mean
EEZ	Exclusive Economic Zone
EG	Exponentially Weighted Geometric Mean
ES	Expenditure Shares Weighting Scheme
FAO	Food and Agriculture Organization of the United Nations
G	Weighted Geometric Mean
HDI	Human Development Index
OHI	Ocean Health Index
P	Proportional Weight Weighting Scheme
PA	Proportionally Weighted Arithmetic Mean
PG	Proportionally Weighted Geometric Mean
SFU	Simon Fraser University

Data Accessibility

Data layers for the 2012, 2013 and 2014 analyses are publically available online at http://www.oceanhealthindex.org.

1 Introduction

In ecology, indices are widely used for performance monitoring, benchmarking, policy analysis, and public communication (Convention on Biological Diversity, 2010; Halpern et al., 2012; World Bank Group, 2014). For example, indices are used extensively to detect declines in abundance (Czech & Krausman, 1997; Kerr & Cihlar, 2004; Mace et al., 2008), to quantify impacts of human activities on ecosystems (Ban, Alidina, & Ardron, 2010; Halpern et al., 2008), and as a resource accounting tool to measure environmental impacts (Rees, 1992). Other fields use indices extensively as well, for example, as a measure of inflation in an economy (Organization for Economic Cooperation and Development, 2004), and in the monitoring of key dimensions of human development (United Nations Development Programme, 2014). An index is referred to in this paper as a metric formed by aggregating individual components into a single number, synonymous with composite and aggregate indices. Each of the examples mentioned above involve the simplification and summarization of large sets of separate measures into a single value, i.e., the index, that represents more manageable information that can be easily communicated to decision-makers, stakeholders and the public (Singh, Murty, Gupta, & Dikshit, 2009).

There is a vast literature on how to design indices, so that they may be most useful to managers or other audiences (Collen & Nicholson, 2014). This analysis focuses specifically on how the mathematical form of the aggregation index calculation influences the properties of the index. An ecological index, for example, could be based on a set of components (i.e., individual measures or a set of indices) describing the state of nature of a population, community, location, or ecosystem (e.g., IUCN indicator of declining abundance, Simpson's Diversity Index, Living Planet Index, Human Development Index) (Mace et al., 2008; Simpson, 1949; WWF, 2014). Common practice assigns a weight to each component and uses an aggregation method, such as averaging, to combine these components into an index. These weights can be derived from statistical or participatory methods (e.g., factor analysis or expert opinion) (Singh et al., 2009). The construction of indices can be complicated due to their multi-dimensionality and the heterogeneity of the individual components (Nardo et al., 2005). For example, the Human Development Index (HDI) produced by the United Nations Development Programme is composed of three components: (1) Health: measured by life expectancy in years, (2) Education: measured by literacy and gross enrolment ratios, and (3) Standard of Living: measured by per capita GDP in US dollars (United Nations Development Programme, 2014). The ability of the HDI to convey accurate information is contingent upon the underlying assumptions made about how the component parts of the index are weighted and summarized. For instance, if the three components are equally weighted, it means that they are assumed to all contribute equally to human development. Additionally, if the range of values is not standardized across the three components (e.g., using a rescaling transformation), the underlying assumption is that an increase by one point in Education represents an equivalent level of development as an increase by one US dollar unit in Standard of Living. These assumptions affect the index behaviour and thus the information communicated to interested parties (Singh et al., 2009).

Once the components of an index have been selected and calculated, one must choose appropriate standardization procedures and weighting schemes for the components to ensure that the units of the components are compatible. Furthermore, one must identify the most appropriate aggregation method to ensure that the aggregation method's properties are meaningful (Singh et al., 2009). Examples of common aggregation methods include the arithmetic mean, geometric mean, and the average in log space (Nardo et al., 2005). Most indices, such as the ecological footprint analysis and several biodiversity metrics (Buckland, Magurran, Green, & Fewster, 2005; Rees, 1996; Singh et al., 2009), use the arithmetic mean or summation to aggregate components. Other indices, such as the Living Planet Index and many biodiversity indices for wild birds (Gregory & Strien, 2010; Loh et al., 2005), have adopted the geometric mean as their aggregation method.

Despite the wide use of indices in ecology, sustainability science, and management, there has been very little consideration of the intrinsic behaviour of an aggregation method and its resulting index in these fields with a few exceptions in the biodiversity literature (e.g., Buckland, Studeny, Magurran, Illian, & Newson, 2011; van Strien, Soldaat, & Gregory, 2012). In contrast, economics has an entire field of research, called index number theory, devoted to comparing aggregation methods and weighting schemes. Index number theory examines how to summarize economic price and quantity data into a smaller set of numbers or a single number, such as the consumer price index (International Monetary Fund, 2004).

The axiomatic approach to index number theory examines the behaviour of indices by comparing how well aggregation methods satisfy a suite of conditions or properties that express intuitive notions about how indices should behave (Hill, 1988; Pfouts, 1966). Although an index will rarely satisfy all the performance criteria of the axiomatic approach, it may still be used to determine the approach that is closest to optimal. For example, when applied to a set of indices of biodiversity (e.g., Simpson's Index, Shannon Index, arithmetic mean of abundance), the axiomatic approach revealed that the index based on the geometric mean of relative abundance satisfies many of the properties and was the most relevant when considering factors such as statistical properties, practicality, and ease of communication (Buckland et al. 2011, Van Strien et al. 2012).

This project expands on this approach by applying index number theory to ecology using a case study, and applies this methodology to aggregation methods and weighting schemes appropriate for the purposes generally needed by ecologists and managers. This analysis investigates the differences in the mathematical properties and aggregate behaviour of eight aggregation methods and weighting schemes, and considers how information about desired index behaviour can be used to make a final choice of the mathematical form for the index. The implications of this choice are discussed for the interpretation of the index when applied as a monitoring and assessment tool.

The Ocean Health Index (OHI) evaluates the status of the world's ocean ecosystems according to ten goals (i.e., food provision, artisanal fishing opportunities, natural products, carbon storage, coastal protection, coastal livelihoods and economies, tourism and recreation, sense of place, clean waters, and biodiversity). These goals are considered beneficial to people and important for the health of marine ecosystems (Halpern et al., 2012). Indices are created to represent each goal and then indices are combined across the 10 goals to form the OHI. Scientists, managers, policy makers and the public use the OHI as a communication and benchmarking tool to help assess ocean health. The ten goals can be considered individually or as an aggregate index number.

The fisheries sub-goal index, which is a sub-set of the food provisioning index of the OHI, measures the degree to which wild-capture fishery resources are fully and sustainably exploited within an exclusive economic zone (EEZ). This index provides a good test case because the number of EEZs, and the number of stocks aggregated for each EEZ, provide sufficient variability in data structures and a large enough sample size to draw meaningful conclusions. Additionally, the index itself measures attributes that are commonly used in conservation science and in management, e.g. population abundance; thus, the results of the exploration are likely to find broad applicability. For example, this analysis could be adapted to a wide range of other ecological applications, such as indicators of endangerment, that do not currently consider the mathematical properties and aggregate behaviour of their methodology.

In this paper, I compare five generic aggregation methods and two weighting schemes by evaluating whether the methods satisfy 11 key properties of the axiomatic approach to index number theory (Table 1). Performance of an aggregation method was determined by the degree to which the method satisfied all or most of these properties. The four best performing aggregation methods are applied to the fisheries sub-goal index. The functional behaviour of each version of the index is compared to the desired behaviour of the index. Finally, a sensitivity analysis is conducted to compare the index scores of the final two aggregation methods and weighting schemes. The results show that the aggregation methods and weighting schemes considered in this paper have different mathematical properties and aggregate behaviour, and result in significantly different index numbers. The choice in aggregation method and weighting scheme for the OHI is dependent on the perception of the associated mathematical properties, aggregate behaviour, sensitivity and relevance.

2 Methods

Five aggregation methods and two weighting schemes are used to compute an index that averages a set of components (Section 2.1). These indices are applied to a case study of the fisheries sub-goal of the OHI, which uses fisheries status as the component values, and fisheries landings as the weighting values (Section 2.2). The aggregation methods and weighting schemes are compared using the axiomatic approach to index number theory. Strong performance of an aggregation method and weighting scheme is determined based on the number of axioms that it satisfies (Section 2.3). The four best performing aggregation methods and weighting schemes are further analyzed for how sensitive they are to changes in fisheries status or landings (Section 2.4). Finally, a sensitivity analysis is conducted to compare the index score and rankings of two aggregation methods and weighting schemes (Section 2.5).

2.1 Aggregation Methods and Weighting Schemes

I derived eight indices by combining examples from index number theory with more conventional indices used in ecology. The aggregation methods and weighting schemes are listed below, where $x_{i,t}$ is the value of component *i* at time *t*, $w_{i,t}$ is the set of weighting values, *N* is the number of components in the data set, and $s_{i,t}$ is the weighting scheme.

Proportionally Weighted Arithmetic Mean (PA)

$$I_{PA} = \frac{1}{N}\sum_{i=1}^{N} x_{i,t} \cdot w_{i,t} \qquad \text{(eqn 1)}$$

Exponentially Weighted Geometric Mean (EG)

$$I_{EG} = \prod_{i=1}^{N} x_{i,t}^{\,s_{i,t}} \qquad \text{(eqn 2)}$$

Proportionally Weighted Geometric Mean (PG) $$I_{PG} = \prod_{i=1}^{N} (x_{i,t} \cdot w_{i,t})^{\frac{1}{N}}$$ (eqn 3)

Weighted Geometric Mean (G) $$I_{G} = \prod_{i=1}^{N} (x_{i,t} \cdot s_{i,t})^{\frac{1}{N}}$$ (eqn 4)

Weighted Arithmetic Mean (A) $$I_{A} = \sum_{i=1}^{N} x_{i,t} \cdot s_{i,t}$$ (eqn 5)

The weighting scheme can take on one of two forms: (1) $s_{i,t} = x_{i,t} \cdot w_{i,t} \Big/ \sum_{j=1}^{N} x_{j,t} \cdot w_{j,t}$, which is called the "expenditure shares" as in index number theory, or (2) $s_{i,t} = w_{i,t} \Big/ \sum_{j=1}^{N} w_{j,t}$, which is called the "proportional weight". The weighting scheme is only applied to the exponentially weighted geometric mean, weighted geometric mean, and the weighted arithmetic mean. Superscripts "ES" and "P" represent expenditure shares and proportional weight, respectively. The proportionally weighted arithmetic mean and geometric mean do not use a weighting scheme (e.g., $x \cdot s$), and are instead weighted directly using the weighting values (e.g., $x \cdot w$).

2.2 Case Study

The fisheries sub-goal index reflects the degree to which sustainable seafood is being provided from wild-capture fisheries in an EEZ. This analysis refers to a stock as a species fished within a Food and Agriculture Organization of the United Nations (FAO) major fishing area. The status of each fished 'stock' within an EEZ is assessed a score ranging from 0 to 1, where 1 represents an optimally harvested stock (i.e. to its maximum sustainable potential) and 0 represents a stock that is not harvested due either to over- or under-exploitation. The contribution of each of the stocks to total seafood being provided depends on how much of that species is being harvested. Therefore, the index requires averaging status across all individual stocks fished within an EEZ ($x_{i,t}$), weighted by their corresponding level of landings ($w_{i,t}$). See the Appendix for more details on the underlying data.

2.3 Desirable Properties

This analysis uses a set of 11 axioms originally developed for indices in index number theory (Auer, 2008). See Table 1 for description and examples. These axioms are applied to compare the performance of the aggregation methods introduced in Section 2.1 when calculated on an artificial dataset of ten randomly generated component and weighting values (Table 2). This artificial dataset is simulated to test whether the axiom's mathematical definition is violated. Strong performance of an aggregation method was determined based on the number of axioms that it satisfied.

The complete list of axioms from the axiomatic approach is included in this analysis, however not all axioms should necessarily contribute equally to one's understanding of performance of the aggregation method, and one should not be limited to the list of axioms provided in the axiomatic approach. In this paper, the eleven axioms are used as a starting point to compare a set of aggregation methods and weighting schemes. The behaviours of the aggregation methods and weighting schemes are further analyzed in Section 2.4 as they relate to the OHI case study.

Based on the results of the axiomatic analysis, the four best performing aggregation methods are applied to real data from the fisheries sub-goal index of the OHI case study to demonstrate the differences between these aggregation methods. In particular, this analysis calculated the index score over time for Australia and Canada EEZs that tend to have medium to high landings and divergent index patterns.

2.4 Aggregate Behaviour

The four best performing aggregation methods and weighting schemes from Section 2.3 are also used to produce an index from the simulated status and landings data of ten hypothetical stocks in a single year (Table 2). I analyzed how sensitive the aggregation methods, weighting schemes and their overall index scores are to changes in either status or landings of a single stock. For nine of the hypothetical stocks, the status and landings data are held constant throughout the exercise. The status and landings data for the tenth stock is varied under different scenarios:

(1) Status is varied between zero and one, and landings are held constant. This is done to analyze the effect of improving / declining status of a stock.

(2) Status is held constant at a low value (status = 0.05), and landings are varied between zero and three hundred, where the other fisheries landings range from 35 to 198. This is done to analyze the effect of increased / decreased landings of a stock that is in poor condition.

(3) Status is held constant at a high value (status = 0.95), and landings are varied between zero and three hundred, where the other fisheries landings range from 35 to 198. This is done to analyze the effect of increased / decreased landings of a stock that is in good condition.

The choice of scenarios allows me to evaluate the effects of the weighting scheme and the aggregation method separately, further analyze the axioms from the axiomatic approach, and detect if there is an asymmetric behaviour between the overall index score and the status or landings variable (i.e., stocks with high status and low status have different ways of influencing the score). The results allow me to narrow down the selection of aggregation method and weighting scheme to two by deciding that the monotonicity axiom is an essential axiom.

2.5 Sensitivity Analysis

Two aggregation methods are chosen from Section 2.4 and a graphical sensitivity analysis is conducted that compares the two aggregation methods' fisheries sub-goal index scores and rankings for all EEZs in 2011. Finally, the two aggregation methods are applied to the fisheries sub-goal index over time (1980-2011) for a subset of randomly selected EEZs.

3 Results

3.1 Desirable Properties

The axiomatic analysis found that no one index met all 11 of the axioms, but the exponentially weighted geometric mean with proportional weight (I_{EG}^{P}), the exponentially weighted geometric mean with expenditure shares (I_{EG}^{ES}), the weighted arithmetic mean with proportional weight (I_{A}^{P}), and the weighted arithmetic mean with expenditure shares (I_{A}^{ES}) satisfied the greatest number of axioms (Table 3). In particular, I_{EG}^{P} and I_{A}^{P} fulfilled nine axioms, while I_{A}^{ES} and I_{EG}^{ES} fulfilled eight axioms.

The four best-performing aggregation methods are applied to the fisheries sub-goal index of the OHI case study using Canada and Australia to examine the differences between the aggregation methods. The value and trend in Australia's overall index score depends on the aggregation method (Figure 1A) because Australia has increased the proportion of stocks being caught from both low and high status fisheries over time (Figure 2A). When stocks were aggregated using I_{A}^{ES}, the overall index number was relatively high with a slight increase over time. When the stocks are aggregated using I_{EG}^{ES}, the overall score is relatively high and stable over time. When the stocks are aggregated using either I_{A}^{P} or I_{EG}^{P}, the overall score is relatively low and slightly decreasing over time. In contrast to Australia, the value and trend in Canada's overall index score is more robust to the choice of aggregation method with each giving similar values and trends (Figure 1B). Canada has much more diversity in the types of stocks that are caught at different status levels (Figure 2B).

3.2 Aggregate Behaviour

The four best performing aggregation methods used simulated data of ten stocks' landings and status scores to produce four indices, and analyze the sensitivity of these indices to changes in landings or status scores of a single stock. I_A^P results in a linear relationship between changes in a stocks' status and changes in the overall index score (Figure 3A). When using the expenditure shares weighting scheme (i.e., I_A^{ES} and I_{EG}^{ES}) there is a parabolic-shaped relationship between the status of an individual stock, and overall index score. I_{EG}^P results in decreasing marginal returns of overall index score. Furthermore, the overall index score responds non-linearly to increases in stock landings, which correspond to an increase in the weight of that stock's status, when using proportional weighting (Figures 3B and 3C). Although more difficult to see, this is also true for the expenditure shares weighting scheme.

I_{EG}^P rewards the overall index score more when landings of a low status stock are reduced than when landings of a high status stock are increased. Based on the slopes of the lines, increasing landings of a stock that has a high status (Figure 3C) does not increase overall index score as fast as if one were to decrease the landings of a stock with a low status (Figure 3B). In contrast, the expenditure shares weighting scheme rewards increasing landings of a stock with a high status (Figure 3C) more than reducing landings of a stock with a low status (Figure 3B). With the expenditure shares weighting scheme, it is very difficult to significantly improve one's overall index score by reducing landings of a stock with a low status (Figure 3B), but it is much easier to increase one's overall index score by increasing landings of a stock with a high status (Figure 3C); whereas I_A^P equally rewards increased landings of stock with a high-status and decreased landings of stock with a low status.

3.3 Sensitivity Analysis

I_A^P and I_{EG}^P are compared in the sensitivity analysis. The analysis resulted in different overall index scores and produced different overall rankings for intermediately ranked EEZs (Figure 4A and 4B). The arithmetic mean tended to produce higher index scores than the geometric mean (Figure 4A). The best and worst rankings did not change noticeably between I_A^P and I_{EG}^P (lines are relatively parallel), but did cause considerable reordering of the intermediate ranking EEZs (Figure 4B).

4 Discussion

This project applies index number theory to the fisheries sub-goal of the OHI by applying this methodology to aggregation methods and weighting schemes appropriate for the purposes of the OHI. This analysis investigates the differences in the mathematical properties and aggregate behaviour of these aggregation methods and weighting schemes, and considers how information about desired index behaviour can be used to make a final choice of the mathematical form for the index. The results show that I_A^P and I_{EG}^P perform the best in terms of their underlying mathematical properties and aggregate behaviour.

4.1 Desired Properties

The axiomatic analysis found that none of the indices analysed satisfied all eleven axioms, however I_{EG}^P, I_{EG}^{ES}, I_A^P, and I_A^{ES} satisfied the greatest number of axioms. None of these aggregation methods satisfied the weak and strict commensurability axioms (axioms A10 and A11), whereas I_{PA} and I_{PG} did. This suggests that a change in the component values by a common factor cannot be offset by a change in the weighting values by the reciprocal of that factor. Auer (2008) showed that the single observation axiom cannot be satisfied at the same time as the weak and strict commensurability axioms; thus, one must consider the trade-offs between these two axioms, and decide which of the axioms is most important to the aggregation method.

I_{EG}^P, I_{EG}^{ES}, I_A^P, and I_A^{ES} were applied to the fisheries sub-goal of the OHI for Australia and Canada. The four indices resulted in significantly different overall index scores, especially in the case of Australia. The overall index score for Australia varies significantly between aggregation method and weighting scheme, whereas the overall index score for Canada varies much less.

Australia has increased the proportion of fisheries with both high and low status. In 1980, 54.6 percent of Australia's landings were from stocks with an individual status of less than 0.5; this percentage increased to 79.4 by 2011. In contrast, in 1980, only 5.2 percent of Australia's landings were from stocks with an individual status of greater than 0.9; this percentage increased to 18.6 by 2011. Part of this change is attributed to the increase in landings of historically low status stocks, such as the southern bluefin tuna (*Thunnus maccoyii)*. There has been a widespread decline of stock status without a reduction in fishing activity, such as the great barracuda (*Sphyraena barracuda)* stock. For some fisheries in Australia, the low status is the result of poor reporting of taxonomic scale (e.g., pelagic fishes, marine animals). There has been an increase in landings and status over time of many stocks in Australia, including the yellowfin tuna (*Thunnus albacares* stock).

In contrast, Canada has a greater diversity in scores than Australia. Over time, Canada has increased its proportion of catch taken from stocks with high status, and marginally decreased its proportion of catch taken from stocks with moderate to low status. This pattern is the result of changes to the status of species being fished, and changes in the proportion of landings of species being fished. For example, between 1980 and 2011, there was a decrease in the status of some species in Canada, such as albacore (*Thunnus alalunga*) and sablefish (*Anoplopoma fimbria)*, and an increase in the status of other species, such as Alaska pollock (*Theragra chalcogramma)* and arrowtooth flounder (*Atheresthes stomias)*. In some of these cases, the decrease or increase in status of the stock is reflected in the proportion of landings of that stock. Canada also experienced a decline in low status fisheries in the early 1990s. This is caused by the significant reduction in landings of Atlantic cod (*Gadus morhua*) from 22 percent of total landings in 1988 to 1 percent by 1995. The fisheries sub-goal index score for Canada increased over this time.

Whether the aggregation methods show consistent patterns will depend on the specific fisheries in that EEZ. For example, I_{EG}^{P} is especially sensitive to low values in fisheries status. For EEZs landing fish predominantly from groups with status in the mid range (e.g., Canada), the aggregation methods show similar relative but different absolute overall index scores. For EEZs landing a significant proportion of fish from groups in the low status range (e.g., Australia), the geometric mean will penalize overall index score in a non-linear way; for instance, when an EEZ increases fishing of a low status stock. As a consequence of this, the aggregation methods may not show similar relative patterns, and will not show similar absolute patterns. This non-linear aggregate behaviour resulting from harvesting low status stocks will be discussed more in the following section.

4.2 Aggregate Behaviour

For increasing stock status, I_{EG}^{P} has decreasing marginal returns in overall index score. This suggests that one can improve the overall index score more by improving the status of a stock that has low status (e.g., x = 0.05), than a stock that has high status (e.g., x = 0.95), but improving the status of the stock always increases the overall index score. Depleted stocks are harder to rebuild than overexploited stocks, and fishing less of a stock in good health is more easily remedied than fishing too much of a stock in poor health. Furthermore, a stock may have a low status as a result of overexploitation and fishery collapse, and may not be fished anymore due to low abundances (e.g., Atlantic cod). Small landings from a low status fishery may not be reflected in the index number as much as it should be. For this reason, it seems appropriate to penalize EEZs more for increasing landings of a low status stock than reducing landings of a high status stock. However, this property results in the aggregation method being very sensitive to low values, which can become problematic when the status of a single stock is zero. The geometric mean offers a more pessimistic view based on the absolute value of overall index score by providing a lower score than other aggregation methods.

The expenditure shares weighting scheme (i.e., I_{A}^{ES} and I_{EG}^{ES}) results in a parabolic-shaped relationship between an individual stock status and overall index score. This suggests that when fisheries status is low, a one-unit increase in fisheries status can result in a decrease in the overall fisheries sub-goal index score. This pattern was detected using the axiomatic approach, which found that the expenditure shares weighting scheme failed to fulfill the monotonicity axiom; this axiom states that if one component increases with everything else remaining constant, the index number should also increase. The expenditure shares weighting scheme used in this analysis was borrowed from index number theory, which commonly uses this weighting scheme for indices that scale to a baseline point in time. For indices that don't scale to a baseline, such as the fisheries sub-goal, the expenditure shares weighting scheme results in a parabolic shape as shown in Figure 3A. The equation for this relationship can be easily derived; for example, for the I_A^{ES}, the relationship can be derived as approximately $I_A^{ES} = (x^2 + 445.51)/(x + 606.43)$, where *x* is fisheries status, and *I* is the overall index score.

Some subjective judgment is required to choose which aggregation method and weighting scheme is most suitable for the fisheries sub-goal index of the OHI case study. This depends on what aggregate behaviour and sensitivities are considered most suitable for the aggregation method; this will depend on the case study being considered. This decision should be informed by the information considered in the analysis above. For example, one needs to decide whether it is good behaviour to reward the fisheries sub-goal index score more when landings of a stock with a low status are reduced than when landings of a stock with a high status are increased as with I_{EG}^P. For the OHI, this analysis assumes that monotonicity is an essential property – increasing the status of a single stock should always increase the index score. With monotonicity as a prerequisite, the set of aggregation methods and weighting schemes are narrowed down to two: I_A^P and I_{EG}^P.

4.3 Sensitivity Analysis

I_A^P and I_{EG}^P resulted in significantly different overall index scores and rankings. This emphasizes that our choice in aggregation method matters and the importance of considering the mathematical properties and aggregate behaviour a priori.

The dendrogram performed moderately well in identifying the related groupings of countries (Figures 5A and 5B). The dendrogram may identify clusters of countries with similar geographic location, as this may reflect similar stocks being fished and similar or shared fisheries management, or with similar economic factors. It is worth noting here that while the stock status is estimated at the scale of the statistical area defined by the FAO, which may include multiple EEZs, landings data used for the weights are obtained for each country's EEZ. Thus, country-specific factors that affect the rate of fishing at any given time would emerge from the dendrogram. Lithuania, Russia (Baltic Sea), Poland and Latvia were successfully identified as being related for both aggregation methods. Other less-related countries, such as Canada and Morocco, which do not seem to have any relation in reality, were identified as sharing similar patterns in overall index score. The aggregation methods resulted in some different groupings in the dendrogram; for example, I_{EG}^{P} identified Estonia as being related to Lithuania, Russia (Baltic Sea), Poland and Latvia, whereas the I_{A}^{P} did not.

5 Conclusions

Examining the axiomatic properties beforehand allows one to understand the consequences of choosing one aggregation method over another. The results of the sensitivity analysis highlight the need to look at this issue a priori. The choice of aggregation method not only affected the overall index score, but also the relative rankings of EEZs. This analysis showed that the choice in aggregation method required a number of tradeoffs and value judgments. The axiomatic approach was valuable in narrowing the field of aggregation methods, but the final choice was dependent on the perception of the associated mathematical properties, aggregate behaviour, sensitivity and relevance. The perception of the appropriateness of these factors will differ on a case-by-case basis, since interested parties may value the changes in a component value differently.

In practice, ecological indices are more likely to be effective and applicable if they acknowledge the interconnected nature of science and policy (Gieryn, 1983). Science is used in the collection of data and the construction of ecological indices; however, ecological indices are also shaped by social preferences and considerations. Therefore, indices should be constructed to ensure that they are relevant to both ecology and policy (Turnhout, Hisschemoller, & Eijsackers, 2007). An ecological index that does not consider these perspectives will likely be less successful at contributing to meaningful policymaking and public communication, and may even lead to incorrect policy determinations. The construction of ecological indices does not fall clearly in the science or policy domain, but rather somewhere in the middle, where science and policy overlap. This requires the transfer of knowledge from one domain to the other (Turnhout et al., 2007). Scientists must communicate with non-scientists and frame their work within a social and political context. Thus, in addition to the scientific reasoning behind the chosen aggregation method, one must also consider the social and political landscape that defines the goals and purpose of the index. For the index to be effective, the aggregation method may need to consider social values and policy needs. For example, whether an index should be more sensitive to changes in low values than high values is as much a policy judgment as a scientific question, because the appropriateness of such a behaviour is defined by the political, social and ecological contexts.

When producing an index of any form, a procedure should be followed to pick the most appropriate aggregation method. One can follow a similar procedure to what was used in this analysis (Figure 6). The axiomatic approach is a good starting point to narrow down the number of aggregation methods being considered. One can evaluate which mathematical properties are most important by setting achievable objectives by having the axiom list reflect these value judgments. The axiomatic approach is flexible to these value judgments and the diverse range of ecological applications that we work with. One can also consider how each aggregation method performs under different scenarios specific to the application. I recommend working alongside interested parties to decide which aggregate behaviour is appropriate and which is not.

Future work could consider the use of bilateral indices in ecological applications. Bilateral indices, such as the consumer price index, are used to track relative changes over time or space and could be a useful tool to track the change in an ecological process (e.g., to track the change in sea level rise). Bilateral indices are more commonly used in index number theory because they have more favourable mathematical properties than indices that track a single point in time (Auer, 2008). Bilateral indices were not used in this analysis because bilateral indices lose the original units in which they were measured and are instead scaled to a baseline point in time. Here a baseline is defined as the state of a system in one point in time. For example, an index in year 2012 could be scaled to the index baseline in year 2000; this would result in an adimensional index number, independent of time. In order to be effectively communicated to stakeholders and policy makers, the fisheries sub-goal of the OHI needs to be meaningful in its original units. Bilateral indices were deemed inappropriate for the goals of the OHI. Further research should also consider the robustness of indices to reasonable departures from assumptions. For example, in managed systems, attaining perfect information about the system is unlikely; however, management decisions need to be made regardless of process and observation uncertainty. A power analysis could be performed to observe the sensitivity of aggregation methods to various levels and types of uncertainty. This would provide information on the probability of detecting an effect with a given level of confidence (e.g., false positive and negative rates).

The OHI faces a number of challenges that cannot be solved through aggregation methodology alone. First, landings data is obtained at the EEZ scale and is sensitive to the level of catch reporting of that EEZ. In contrast, status data is obtained at the FAO region scale and may include status information from multiple EEZs; thus, poor fisheries management by one EEZ may skew the status values of good fisheries management of another EEZ if both EEZs are within the same FAO region. These factors may favour some EEZs more than others. Furthermore, these factors will change over time depending on the discrepancy between status values within an FAO region, and social, political and economic considerations within an EEZ. Second, when a low status fishery collapses, reducing landings of that fishery (e.g., the Atlantic cod fishery in Canada), the fisheries sub-goal index score will improve. This behaviour can be minimized by using an aggregation method that is sensitive to low values.

This paper showed that collaboration between science and policy is necessary to produce an effective index. The behaviour of ecological indices needs to be tied to the goals and purpose of the index, to change in the appropriate manner, and to be communicable to the intended audience. All of these issues require a thorough examination of the intrinsic properties of the index before they are implemented.

Tables

Table 1. Axiomatic approach axioms, their descriptions, and an example of each from the Ocean Health Index.

	Axiom	Formal Definition	Simplified Definition	OHI Example
A1	Anonymity Axiom	$I(x^t, w^t)$ is exclusively a function of *x* and *w*.	The index number should only consider the set of components being summarized (x) and the set of weighting values (w) as inputs into the index formula.	The index number should not be affected by the number of fisheries (N) being aggregated.
A2	Invariance to Re-Ordering Axiom	$I(x^t, w^t) = I(\tilde{x}^t, \tilde{w}^t)$ where $\tilde{x}^t$ and $\tilde{w}^t$ are uniform permutations of the vectors x^t and w^t	If the set of components being summarized and the set of weighting values are reordered in the same way, then the index number remains unchanged.	If there is information from three fisheries being aggregated, it should not matter which order the fisheries are aggregated in.
A3	Single Observation Axiom	$I(x^t, w^t) = x_1$ when N=1	If the set of components being summarized is composed of only one observation, then the index number should take the value of this single observation.	If there is one fishery being aggregated, and this fishery has a status equal to 0.1, then the index number should equal 0.1 regardless landings.
A4	Uniformity Axiom	When $x_i = x$ for $i = 1,2, \dots, N$ $I(x^t, w^t) = x$	If the set of components being summarized is composed of a number of observations of the same value, then the index number should also take on that value.	If all of the fisheries being aggregated have the same status, then the index number should equal that same status regardless of landings.
A5	Mean Value Axiom	$\min_i \{x_i\} \le I(x^t, w^t) \le \max_i \{x_i\}$	The index number should take a value between the largest and smallest components.	If fisheries status ranges between 0.1 and 0.5, then the index number should also range between 0.1 and 0.5 regardless of landings.

	Axiom	Formal Definition	Simplified Definition	OHI Example
A6	Positivity Axiom	$I(x^t, w^t) > 0$ if $x^t >> 0$ and $w^t >> 0$	The index number is positive so long as the set of components being summarized and the set of weighting values are positive.	If there are no fisheries with a status of zero, then the index number should be positive.
A7	Linear Homogeneity Axiom	$I(\lambda \cdot x^t, w^t) = \lambda \cdot I(x^t, w^t)$ $\forall \lambda > 0$	If each observation in the set of components being summarized is changed by the same factor and the set of weighting values remains the same, then the index number should also change by the same factor.	If the set of fisheries status values doubles, then so should the index number.
A8	Quantity Proportionality Axiom	$I(x^t, \lambda \cdot w^t) = I(x^t, w^t)$ $\forall \lambda > 0$	If each observation in the set of weighting values changes by the same factor and the set of components being summarized remains the same, then the index number should remain unchanged.	If the set of landings values doubles, then there should be no change to the index number.
A9	Monotonicity Axiom	$I(x^t, w^t) > I(x^{*t}, w^t)$ when $x^{*t} \geq x^t$, $\forall x^{*t}$ and $\forall x^t$ and for at least one element the strict relation holds.	If the set of components being summarized is greater than or equal to another set of components with at least one element being strictly greater than the first index number should be strictly greater than the second.	If the set of fisheries status values for one EEZ is greater than another EEZ, then the first EEZ should have a greater index number.
A10	Weak Commensurability Axiom	$I(\lambda \cdot x^t, w^t / \lambda) = I(x^t, w^t)$ $\forall \lambda > 0$	If the set of components being summarized changes by a factor and the set of weighting values changes by the reciprocal of this factor, the index number should remain unchanged.	If the set of fisheries status values double in size and the set of landings values are reduced by half, then the index number should remain unchanged.
A11	Strict Commensurability Axiom	$I(\Lambda \cdot x^t \cdot \Lambda, w^t \cdot \Lambda^{-1}) = I(x^t, w^t)$ Λ is an arbitrary N by N diagonal matrix with positive elements λ_i	If each observation in the set of components being summarized changes by different factors within a set of factors and each observation in the set of weighting values changes by the reciprocal of these same factors, then the index number remains unchanged.	If each fisheries status value changes by a different factor, and the corresponding landings values change by the reciprocal of those factors, then the index number should remain unchanged.

Table 2. **Example dataset with 10 simulated component and weighting values. This simulated numerical example is used to evaluate the axiomatic approach in Section 2.3, and the aggregate behaviour of different aggregation methods in Section 2.4.**

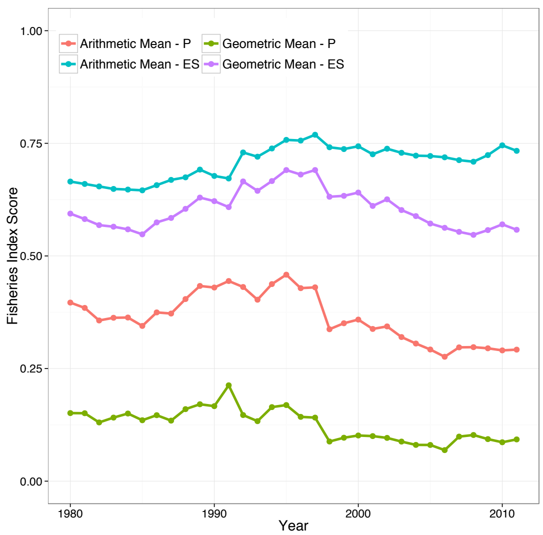
Stock No.	Status (Component Value)	Landings (Weighting Value)
1	0.2655	41.1949
2	0.3782	35.3114
3	0.5729	137.4046
4	0.9082	76.8207
5	0.2017	153.9683
6	0.8984	99.5399
7	0.9447	143.5237
8	0.6608	198.3812
9	0.6291	76.0070
10	0.0618	155.4890

Table 3. Results of the axiomatic approach. The '✔' signifies when an aggregation method satisfies the axiom requirements.

		I_{PA}	I_{EG}^{ES}	I_{EG}^{P}	I_{PG}	I_{G}^{ES}	I_{G}^{P}	I_{A}^{ES}	I_{A}^{P}
(A1)	Anonymity		✔	✔				✔	✔
(A2)	Invariance to Re-Ordering	✔	✔	✔	✔	✔	✔	✔	✔
(A3)	Single Observation		✔	✔				✔	✔
(A4)	Uniformity		✔	✔				✔	✔
(A5)	Mean Value		✔	✔				✔	✔
(A6)	Positivity	✔	✔	✔	✔	✔	✔	✔	✔
(A7)	Linear Homogeneity	✔	✔	✔	✔	✔	✔	✔	✔
(A8)	Quantity Proportionality		✔	✔		✔	✔	✔	✔
(A9)	Monotonicity	✔		✔	✔		✔		✔
(A10)	Weak Commensurability	✔			✔				
(A11)	Strict Commensurability								
	Total Number of Properties Satisfied	5	8	9	5	4	5	8	9

Figures

A.



B.

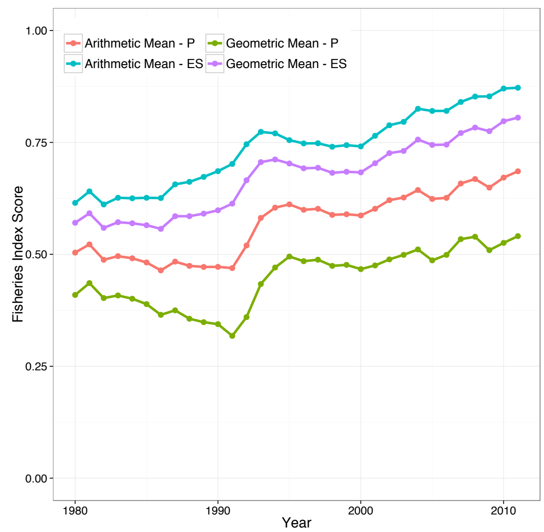
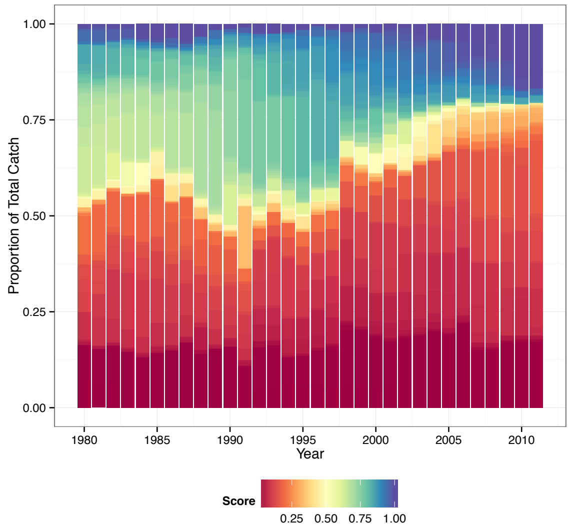


Figure 1. **Fisheries sub-goal index scores for Australia (Figure 1A) and Canada (Figure 1B) using the arithmetic and geometric mean with proportional weight (P) and expenditure shares (ES) weighting schemes.**

A.



B.

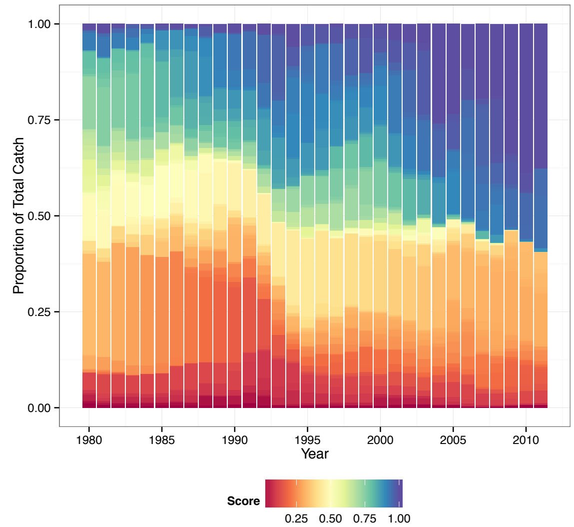


Figure 2. **The proportion of catch at each fisheries status level for Australia (Figure 2A) and Canada (Figure 2B), from 1980 through 2011.**

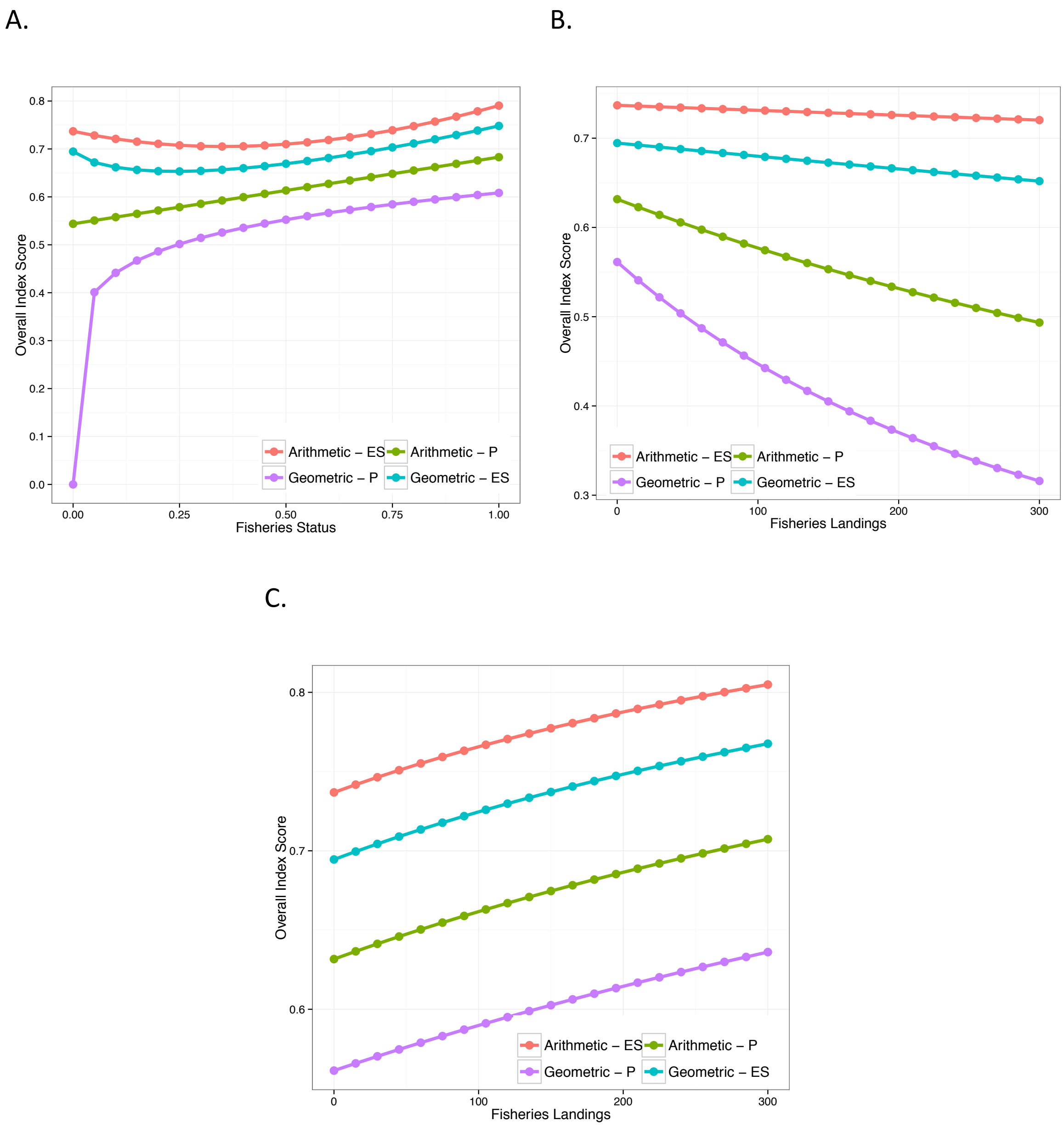


Figure 3. **The overall index score as a function of the status of one stock (Figure 3A), landings from one stock when status of that stock is 0.05 (Figure 3B), and landings from one stock when status of that stock is 0.95 (Figure 3C). All other data remains constant.**

A. B.

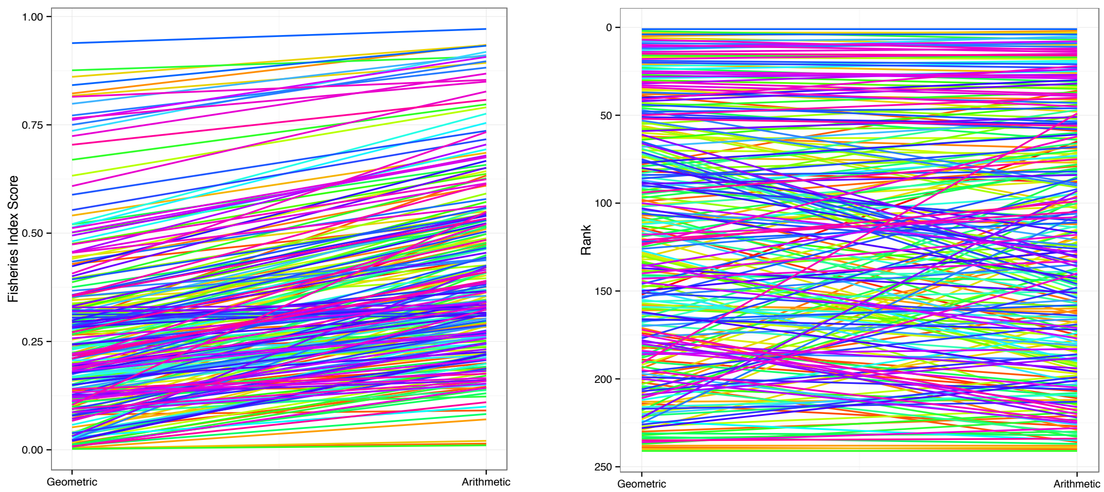


Figure 4. **Comparison of the arithmetic and geometric mean with proportional weight fisheries sub-goal index scores (Figure 4A) and rankings (Figure 4B) for all EEZs in 2011. A line segment is drawn to represent the difference between the two aggregation methods. If the line segment is perfectly horizontal, this suggests that the two aggregation methods do not result in different overall index scores or rankings. If the line segment is slanted one way or the other, the two aggregation methods result in different overall index scores or rankings.**

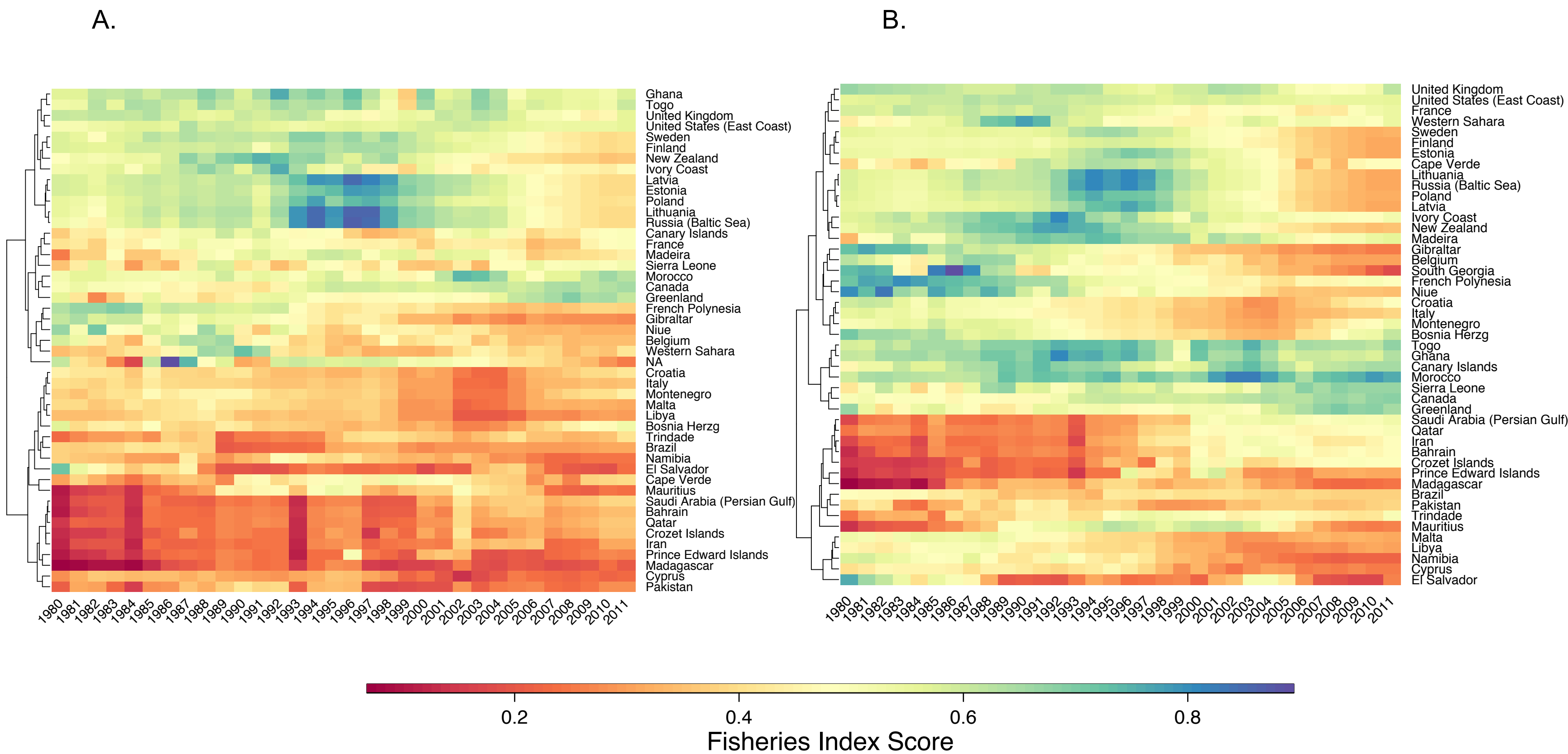


Figure 5. **The fisheries sub-goal of the Ocean Health Index for a subset of 48 randomly selected countries. The exponentially weighted geometric mean with proportional weight is shown in Figure 5A and the weighted arithmetic mean with proportional weight is shown in Figure 5B. The dendrogram on each plot clusters countries based on their correlation of scores.**

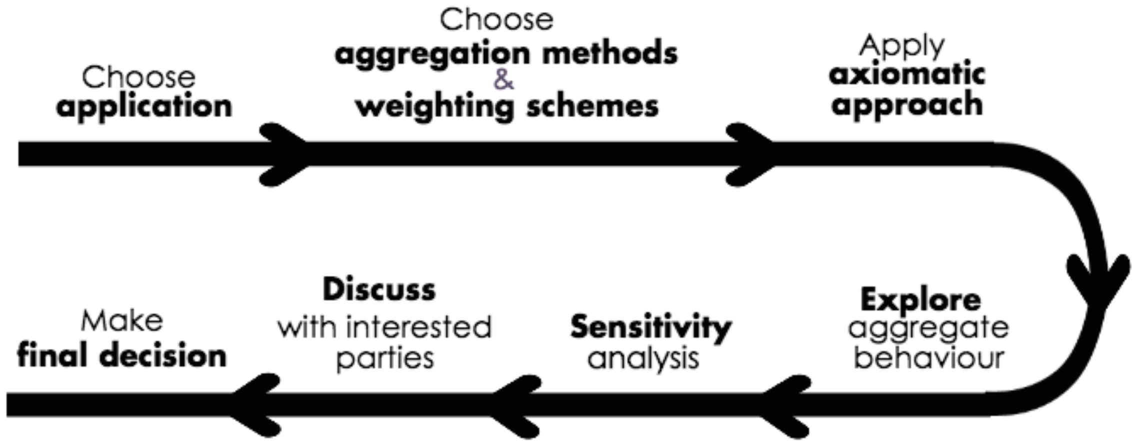


Figure 6. The general procedure for producing an index number.

References

Auer, von, L. (2008). *Axiomatic Analysis of Unilateral Price Indices* (pp. 1–8). Germany: Universitat Trier.

Ban, N. C., Alidina, H. M., & Ardron, J. A. (2010). Cumulative impact mapping: Advances, relevance and limitations to marine management and conservation, using Canada's Pacific waters as a case study. *Marine Policy*, *34*(5), 876–886. doi:10.1016/j.marpol.2010.01.010

Buckland, S. T., Magurran, A. E., Green, R. E., & Fewster, R. M. (2005). Monitoring change in biodiversity through composite indices. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1454), 243–254. doi:10.1098/rstb.2004.1589

Buckland, S. T., Studeny, A. C., Magurran, A. E., Illian, J. B., & Newson, S. E. (2011). The geometric mean of relative abundance indices: a biodiversity measure with a difference. *Ecosphere*, *2*(9), 1–15.

Collen, B., & Nicholson, E. (2014). Conservation. Taking the measure of change. *Science (New York, N.Y.)*, *346*(6206), 166–167. doi:10.1126/science.1255772

Convention on Biological Diversity. (2010). Strategic Plan for Biodiversity 2011–2020, including Aichi biodiversity targets

. *Http://Www.Cbd.Int/Doc/Strategic-Plan/2011-2020/Aichi-Targets-en.Pdf*. Retrieved March 13, 2015, from

Czech, B., & Krausman, P. R. (1997). Distribution and causation of species endangerment in the United States. *Science*, *277*(5329), 1116–1117. doi:10.1126/science.277.5329.1116

Gieryn, T. F. (1983). Boundary-Work and the Demarcation of Science from Non-Science: Strains and Interests in Professional Ideologies of Scientists. *American Sociological Review*, *48*(6), 781–795. doi:10.2307/2095325

Gregory, R. D., & Strien, A. V. (2010). Wild Bird Indicators: Using Composite Population Trends of Birds as Measures of Environmental Health. *Ornithological Science*, *9*(1), 3–22. doi:10.2326/osj.9.3

Halpern, B. S., Longo, C., Hardy, D., McLeod, K. L., Samhouri, J. F., Katona, S. K., et al. (2012). An index to assess the health and benefits of the global ocean. *Nature*, *488*(7413), 615–620. doi:10.1038/nature11397

Halpern, B. S., Walbridge, S., Selkoe, K. A., Kappel, C. V., Micheli, F., D'Agrosa, C., & Bruno, J. F. (2008). A Global Map of Human Impact on Marine Ecosystems. *Science*, *319*, 948–952.

Hill, T. P. (1988). Recent Developments in Index Number Theory and Practice. *OECD Economic Studies*, *10*, 123–148.

International Monetary Fund. (2004). Basic Index Number Theory. In *Producer Price Index Manual Theory and Practice* (pp. 370–402). Washington, D.C.

Kerr, J. T., & Cihlar, J. (2004). Patterns and causes of species endangerment in Canada. *Ecological Applications : a Publication of the Ecological Society of America*, *14*(3), 743–753. doi:10.1890/02-5117

Loh, J., Green, R. E., Ricketts, T., Lamoreux, J., Jenkins, M., Kapos, V., & Randers, J. (2005). The Living Planet Index: using species population time series to track trends in biodiversity. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*(1454), 289–295. doi:10.1098/rstb.2004.1584

Mace, G. M., Collar, N. J., Gaston, K. J., Hilton-Taylor, C., Akcakaya, H. R., Leader-Williams, N., et al. (2008). Quantification of Extinction Risk: IUCN's System for Classifying Threatened Species. *Conservation Biology*, *22*(6), 1424–1442. doi:10.1111/j.1523-1739.2008.01044.x

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., & Giovannini, E. (2005). Handbook on Constructing Composite Indicators: Methodology and Use Guide, *4*.

Organization for Economic Cooperation and Development. (2004). *Consumer price index manual - Theory and practice*. Geneva.

Pfouts, R. W. (1966). An Axiomatic Approach to Index Numbers. *Review of the International Statistical Institute*, *34*, 174–185.

Rees, W. E. (1992). Ecological footprints and appropriated carrying capacity: what urban economics leaves out. *Environment and Urbanization*, *4*(2), 121–130. doi:10.1177/095624789200400212

Rees, W. E. (1996). Revisiting carrying capacity: Area-based indicators of sustainability. *Population and Environment*, *17*(3), 195–215. doi:10.1007/BF02208489

Simpson, E. H. (1949). Measurement of diversity. *Nature*, *163*(4148), 688–688. doi:10.1038/163688a0

Singh, R. K., Murty, H. R., Gupta, S. K., & Dikshit, A. K. (2009). An overview of sustainability assessment methodologies. *Ecological Indicators*, *9*(2), 189–212. doi:10.1016/j.ecolind.2008.05.011

Turnhout, E., Hisschemoller, M., & Eijsackers, H. (2007). Ecological indicators: Between the two fires of science and policy. *Ecological Indicators*, *7*(2), 215–228. doi:10.1016/j.ecolind.2005.12.003

United Nations Development Programme. (2014). *Human Development Report 2014* (pp. 1–239). New York.

van Strien, A. J., Soldaat, L. L., & Gregory, R. D. (2012). Desirable mathematical properties of indicators for biodiversity change. *Ecological Indicators*, *14*(1), 202–208. doi:10.1016/j.ecolind.2011.07.007

World Bank Group. (2014). The Worldwide Governance Indicators (WGI) project. *Http://Info.Worldbank.org/Governance/Wgi/Index.Aspx#Home*. Retrieved March 13, 2015, from

WWF. (2014). *Living Planet Report 2014: species and spaces, people and places*. (R. McLellan, L. Iyengar, B. Jeffries, & N. Oerlemans) (pp. 1–180). Gland, Switzerland: WWF.

Appendix

Data for the fisheries sub-goal of the Ocean Health Index

The objective of the fisheries sub-goal of the Ocean Health Index is to produce an index number that reflects the state of fisheries in an exclusive economic zone (EEZ). This requires an averaging of fisheries status of all fisheries in an EEZ, weighted by the corresponding level of landings for each fishery. Fisheries status is determined by whether or not an EEZ is under-utilizing (i.e., biomass greater than B/Bmsy) or over-utilizing (i.e., biomass less than B/Bmsy) a fishery, such that:

$$status = \begin{cases} B/B_{msy} \;\; if \;\; B/B_{msy} < 0.95 \\ 1 \;\; if \;\; 0.95 < B/B_{msy} < 1.05 \\ \max\left\{1 - \alpha(B/B_{msy} - 1.05), \beta\right\} \;\; if \;\; B/B_{msy} > 1.05 \end{cases}$$

where B is biomass, B_{msy} is biomass at maximum sustainable yield, α is the rate at which status declines as the population moves away from its optimal size, and β represents the lowest status possible for an under-utilized population. This analysis takes $\alpha = 0.5$ and $\beta = 0.25$. Fisheries status was obtained at the Food and Agriculture Organization of the United Nations (FAO) region level, and landings data was obtained at the EEZ specific level; therefore, status of a fishery in an EEZ was assumed to be the same for all EEZs within an FAO region, whereas the landings of a fishery in an EEZ was specific to the EEZ. For landings in a given year from an EEZ that are reported at coarser taxonomic levels than species, the landings' status will be based on the minimum of the B/B_{msy} estimates across all species in that FAO region in that year. Scoring it this way makes it such that landings reported at taxonomic levels coarser than species can never get a higher status than the species with the lowest status in the region. See Halpern et al. (2015) for more details on the underlying data.