Ecological Indicators 84 (2018) 546–552

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

The need for validation of ecological indices

P.E. Moriarty[a,*], E.E. Hodgson[a], H.E. Froehlich[b], S.M. Hennessey[c], K.N. Marshall[d], K.L. Oken[e,f], M.C. Siple[a], S. Ko[f], L.E. Koehn[a], B.D. Pierce[a], C.C. Stawitz[e]

[a] School of Aquatic and Fishery Sciences, University of Washington, Box 355020, Seattle, WA 98195, United States
[b] National Center for Ecological Analysis and Synthesis, University of California, 735 State St. Suite 300, Santa Barbara, CA 93101, United States
[c] Department of Integrative Biology, Oregon State University, 2701 SW Campus Way, Corvallis, OR 97331, United States
[d] Cascade Ecology, LLC, PO Box 25104, Seattle, WA 98165, United States
[e] Quantitative Ecology and Resource Management, University of Washington, Box 352182, Seattle, WA 98195, United States
[f] Department of Marine and Coastal Sciences, Rutgers University, 71 Dudley Rd., New Brunswick, NJ 08901, United States

ARTICLE INFO

Keywords:
validation
development
criteria

ABSTRACT

Increased recognition of the need for ecosystem-based management has resulted in a growing body of research on the use of indicators to represent and track ecosystem status, particularly in marine environments. While multiple frameworks have been developed for selecting and evaluating indicators, certain types of indicators require additional consideration and validation. In particular, an index, which we define as an aggregation of two or more indicators, may have unique properties and behaviors that can make interpretation difficult, particularly in a management context. We assert that more rigorous validation and testing is required for indices, particularly those used to inform management decisions. To support this point we demonstrate the need for validation and then explore current development and validation processes for ecosystem indices. We also compare how other disciplines (e.g., medicine, economics) validate indices. Validating indices (and indicators) is particularly challenging because they are often developed without an explicit objective in mind. We suggest that exploring the sensitivity of an index to the assumptions made during its development be a pre-requisite to employing such an index.

1. Introduction

Over the past several decades, recognition of the importance of ecosystem approaches to natural resources management has increased (Fulton et al., 2014; Pikitch et al., 2004). In many countries, these approaches have begun to be codified into law and implemented in management strategies. The Marine Strategy Framework Directive in Europe, which provides a legislative framework for protecting the European Union's marine waters using an ecosystem approach, is one example. Australia has also developed tools for ecosystem-based management, including ecosystem risk assessments and harvest strategy frameworks (Smith et al., 2007). Recent advances in the U.S. include Integrated Ecosystem Assessments (IEAs) (Levin et al., 2009) and fishery ecosystem plans (FEPs) (e.g., Pacific Fishery Management Council, 2013). Management at the ecosystem scale requires metrics that track ecosystem status and provide easily interpretable information about changes in that status.

Indicators are measurable properties that track changes in attributes of ecosystems that cannot be measured directly (Kerschner et al., 2001). Indicators provide a way to track progress towards management objectives. They are also used in decision analysis to evaluate the impacts of alternative management strategies (Fulton et al., 2005), or to define thresholds and goals for management (Gsell et al., 2016). Existing metrics range from relatively simple ones, such as the abundance of a single species, to complex, multifaceted indices (plural of 'index') that combine ecological attributes with economic and social factors, such as the Ocean Health Index (Halpern et al., 2012). Here, we distinguish between an indicator, which involves only one data stream that is directly observable (such as survey counts of a single species), and an index (plural: indices), which is a quantitative aggregation of two or more variables (Mayer, 2008). "Indicator" and "index" are often used interchangeably in the literature, though the two are different concepts and warrant explicit definitions.

Indices can be a useful management and communication tool, as they collapse multiple indicators into a single value. However, their simplicity may conceal important ecosystem complexity. Indices have been used to characterize system attributes such as physical drivers of ecosystem processes (e.g., the Pacific Decadal Oscillation (PDO),

Abbreviation: FEC, focal ecosystem component

* Corresponding author.
E-mail address: pmoriart@uw.edu (P.E. Moriarty).

http://dx.doi.org/10.1016/j.ecolind.2017.09.028
Received 13 June 2017; Received in revised form 13 September 2017; Accepted 14 September 2017
1470-160X/

Mantua and Hare, 2002), inherent properties of an ecosystem (e.g., Shannon biodiversity index), or ecosystem services (e.g., Ocean Health Index, Halpern et al., 2012). An index is easy for managers to understand, but aggregating multiple time series may dampen important features of the attribute the index is designed to represent (Figge, 2004).

We suggest that the aggregate nature of indices mandates additional validation. Validation is the process of establishing that an indicator or index meets performance criteria chosen for the specific circumstances (Rykiel, 1996). This is necessary to gauge how they track the desired attribute of system status over time. A body of literature provides guidance for developing and validating indicators (Samhouri et al., 2012; Samhouri et al., 2009), and indicator development is a key component of IEAs (Levin et al., 2009). However, these guidelines do not differentiate between indicators and indices, though indices require an extra level of scrutiny. Each indicator composing an index must be validated, but additionally, properties of indicators included in an index and the way that indicators are combined likely influence the ecological index value and its temporal dynamics. For example, the variance and amplitude of individual indicators will affect how they act when combined into a single index.

Here, we illustrate the challenge of validating ecological indices and suggest improvements to the practice of validation. We begin with a simulated example to show how indicator properties influence index performance. We then evaluate several widely used ecosystem indices against selection criteria developed for indicators with a focus on validation criteria. Finally, we describe how other disciplines validate indices and compare their methods to those used in marine ecology. We close with suggestions to avoid potential pitfalls in the validation and use of indices to inform management decisions.

2. Why is index validation necessary?

Combining multiple indicators into an index requires the developer to make decisions that may ultimately affect the behavior, performance, and reliability of the index. To explore the consequences of these decisions and provide a rationale for the importance of validation, we use a simulation study. The simulation is intended to be a generic example of how indicator characteristics and methods of combining indicators influence the resulting index. To concretize this example, we present the simulation in the context of a hypothetical scenario.

Consider, for example, a management body that wants to track food web status of a marine system (the attribute) and chooses to monitor benthic fish community composition (the index). The management body chooses the status of benthic fish species because they are an important component of marine food webs (Kershner et al., 2011; Levin and Schwing, 2011). Biomass estimates for four species, based on trawl survey samples, are chosen as indicators to represent species status. The index, benthic fish community composition, is then a sum of survey data on all four species. The abundances are averaged to create the index.

For this demonstration, we generated a 50-year time series of a focal ecosystem component (FEC) (e.g., ecosystem status, details of simulations in Supplement 1). The time series of the FEC was used to generate four simulated time series that were correlated with the FEC (e.g., individual species' biomasses, Fig. 1).

We modified the strength of correlation and the variance of indicator time series and explored how those factors and the method of combining the indicators influenced the ability of the index to represent the FEC.

We considered how the indicator quality and responsiveness affected the performance of the resulting index. Quality and responsiveness are two aspects of indicators commonly considered during indicator development and selection (Kennedy and Jacoby, 1999; Kershner et al., 2011; Niemeijer and de Groot, 2008). We define *quality* as the ability of the indicator to track the FEC of interest (e.g., individual species biomass ability to track food web status), which we defined here as the indicator's correlation to the FEC. High quality indicators have high correlation (correlation = 0.7) to the FEC, where low quality indicators have low correlation (correlation = 0.5; Fig. 2a,b). We define *responsiveness* as how abruptly the indicator can change in response to external conditions, which may or may not be related to the focal ecosystem component. Indicators with high responsiveness are more independent from external conditions (autocorrelation = 0.3), whereas indicators with poor responsiveness are more dependent on external conditions (autocorrelation = 0.7; Fig. 2c,d).

We used three performance metrics to evaluate the ability of the index to track the FEC: 1) the average Pearson's correlation between the index and FEC (from here on referred to as 'correlation'), 2) the proportion of simulations for which a comparison of the directions of the trends (positive or negative) of the index and the FEC, calculated over the final 5 years ('5 year trend') were in agreement, and 3) the proportion of simulations for which the mean of the last 5 years was more than 1 standard deviation above or below the long-term mean for both the index and the FEC ('1 SD'). Metrics (2) and (3) are coarser than correlation, and capture the status and trends of ecological indicators represented in several ecosystem status reports developed for U.S. Fishery Management Councils (e.g., Pacific, North Pacific, and Gulf of Mexico) (Pacific Fishery Management Council, 2013).

We used two scenarios to investigate the influence of quality and responsiveness on index performance. The 'quality scenario' included indices composed of indicators with different qualities (Kershner et al., 2011; Niemeijer and de Groot, 2008), holding responsiveness constant: 1) four low quality indicators, 2) four high quality indicators, and 3) two low quality and two high quality indicators. Similarly, in the 'responsiveness scenario', indices had combinations of indicators with different levels of responsiveness (Kershner et al., 2011; Niemeijer and de Groot, 2008), holding quality constant: 1) four poorly responsive indicators, 2) four highly responsive indicators, and 3) two poorly and two highly responsive indicators.

We also investigated how different weighting schemes for combining indicators affects index performance, as weighting scheme influences index values, and potentially conclusions (Halpern and Fujita, 2013). We summed two high quality indicators and two low quality indicators, while holding responsiveness constant. We then weighted the indicators using three schemes. The first scheme used the default approach of additive weighting (Halpern et al., 2009), where the index value at each time step is an average of component indicators. Second, we used expert judgment weighting (Halpern et al., 2012). Since all time series were randomly generated, we optimistically assumed experts knew the relative quality (correlation to true system state) of the four indicators and weighted them accordingly. Third, we assigned random weights to the four indicators and calculated their weighted sum to simulate failed expert weighing. This may occur if expert judgement was used, but experts judged relative indicator quality poorly.

We found that the coarser metrics (the 5-year trend and standard deviation) were robust to changes in indicator properties (Fig. 3, 4). When using coarser evaluation metrics, indicator quality had no effect and responsiveness had only a small effect on index performance (Fig. 3, 4). These metrics are very robust to changes in the indicators and will only detect changes in the corresponding index if there is a large fluctuation. This may be beneficial if managers are interested in wide fluctuations in the focal ecosystem component, such as regime shifts, or may be cause for concern if smaller fluctuations are important.

When using the fine scale metric, 'correlation', properties of the indicators affected index performance (Fig. 3a). Summing four low quality indicators resulted in a low index quality, with a mean correlation to the true system state of 0.21 (SD = 0.19) and summing four high quality indicators resulted in high index quality, with a mean correlation of 0.93 (SD = 0.02) (Fig. 3a).

When an index included two low quality indicators and two high