Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Ocean & Coastal Management
Ocean & Coastal Management 46 (2003) 235–259
www.elsevier.com/locate/ocecoaman

Environmental health indicators

Jake Rice*

Department of Fisheries and Oceans, Coordinator—DFO Science Advisory Secretariat, 200 Kent Street, Ottawa, Ont., Canada K1A 0E6

Abstract

Ecologists have proposed hundreds of quantitative indicators of the status of ecosystems for evaluation of and reporting on the status of marine ecosystems. The talk applies a common approach to classifying indicators, summarises the main properties of each class of indicator, and provides some illustrations.

Indicators of ecosystem status have roles in both communication and decision support. For both roles, strengths and weaknesses of indicators are usually only partially known. Few have been tested systematically for sensitivity and robustness across a range of contexts… However, to determine best practices for selecting indicators for specific uses, one must have a fairly complete understanding of the information content of the various indicators. The paper explores some alternative approaches to documenting the information content of various indicators of ecosystem status.

As the Precautionary Approach becomes broadly used as the basis for management decision-making, the role of indicators of ecosystem status becomes central. The PA is made operational in decision-making through use of indicators and reference points. This means that it is necessary to identify values of an ecosystem indicator associated with harm to the environment that is serious or difficult to reverse. Methods for identifying and justifying reference points for ecosystem indicators are being developed and tested, but the task is turning out to be complex. Alternative strategies for identifying reference points are reviewed.

When choices must be made from a suite of candidate indicators, it is desirable make the selection on objective grounds. This requires explicit a priori criteria, on which there is not yet scientific consensus. Recent developments in this area are reviewed, and again a way forward proposed.

*Tel.: +1-613-990-0288; fax: 613-954-0807.
E-mail address: ricej@dfo-mpo.gc.ca (J. Rice).

0964-5691/03/$ - see front matter
doi:10.1016/S0964-5691(03)00006-1

1. Introduction

Throughout the paper, I usually refer to Indicators of Ecosystem Status, a more neutral phrasing than Indicators of Environmental Health. The more phenomenological phrase allows performance properties of indicators to be evaluated separately from potentially differing views on what comprises a healthy ecosystem. The paper also refers to unspecified "forcing factors", as well as specific perturbations such as pollution, resource harvesting, eutrophication, substrate removal, etc. Integrated coastal zone management requires indicators that are informative both about effects of specific perturbations, and about situations when the nature of the forcing factors is not known a priori.

The background document [1] gives some history of the development of ecological indicators. A review prepared three years ago [2] found over two-hundred different indicators of ecosystem status, without being exhaustive. The challenge is not to find an indicator of ecosystem status to use. It is to choose the set that will serve the users' needs best. The talk will describe briefly the general classes of indicators, and some of their key attributes, but will not review any particular set of indicators in detail. The focus will be on how to select indicators from the huge spectrum that is available. The talk presupposes that the coastal zone manager already understands:

- the management problem to which the indicators are to be applied,
- the management tools and strategies which the indicators are to support,
- what society (or experts) have determined are acceptable and unacceptable states for the ecosystem.

Even with those things known, choice of the correct indicators is not straightforward. It requires explicit criteria for performance of the candidate indicators, and objective tools for evaluating performance against the criteria.

The talk will develop these considerations and criteria more fully. It will not provide the definitive answer to the question of which indicator(s) to use in Integrated Coastal Zone Management. Rather, it will help users of Indicators of Ecosystem Status ask better questions when selecting the indicator(s) to use. If the question "Which Indicator(s) to use?" is asked in a better way, better answers may be found.

2. Categories of indicators of ecosystem status

Ecosystem indicators have been classified in many ways. Four classes have proved effective at organising discussion on ecosystem effects of fishing [2]. Indicator species are also included because the class is used frequently in some environmental quality studies, although in resource management applications single-species indicators, although common, are rarely thought of as indicators of larger ecosystem properties.

2.1. Indicator species

In management of exploited species, indicators of the status of the harvested species, particularly biomass of the mature population and mortality rate, are core tools [3]. However they are not considered indicators of the status of the larger ecosystem, but as only indicators of the species being exploited or specific species impacted as bycatch [4].

In studies of pollution and general environmental reporting, the use of indicator species is more widespread, and has a long history [5]. Whether the species is reliably informative about the status of "the environment" is determined in part by the way in which the indicator species is chosen and used. When the indicator species approach is used to evaluate the impact of a specific activity on an ecosystem, there may be enough information to select species that are likely to be both fully exposed to the activity and highly sensitive to it. Then the status of the single species indicator may be a reliable indicator of whether or not the activity is being conducted in a sustainable manner. For example, the fledging success of kittiwakes is considered a reliable indicator of the sustainability of harvests of sandeels in parts of the North Sea, because kittiwakes depend more exclusively on sandeels than do other marine predators, and the fledging period is the time when foraging demands are highest. If the sandeel stock is adequate for this particular ecosystem need, then it is considered highly likely to be adequate for other ecosystem requirements as well [6].

When the indicator species is simply chosen to be "representative", its value as an indicator of general ecosystem status depends completely on whether it truly is representative. Ease of measurement is a nice property for an indicator species. However, it does not assure that trends in such an indicator species reflect trends in many other ecosystem components. That can only be established when the covariance of trajectories for many species in the ecosystem is known [7]. When data are adequate for rigorous community analyses, species are identified as playing core roles in system structure and function—the keystone species [8], more often than some species are identified as "typical" of whole species groups [9]. There are many excellent reasons to monitor the status of species with particularly important ecosystem roles. These are certainly indicators of "ecosystem health" but ought not be considered indicators of overall state of the environment.

There are, of course, circumstances when reporting on the status of individual species is required to meet legal or policy commitments. Often these are species that are either designated as threatened or endangered by a recognised agency (in Canada, via COSEWIC, in the US under the ESA, in Europe through both national (UK Biodiversity Group 2000) and OSPAR (BCD 2000 [10]) initiatives, or of high iconic value. In such cases indicators of the status of individual species certainly will influence decision-making in coastal zones, and may be interpreted as a measure of the health of the coastal environment. However, without a great deal more information it would be dangerous to interpret indicators of the status of listed species as reflecting anything more than the status of the particular species of concern.

2.2. Diversity indices

Diversity is a joint construct of both how many species are present in a collection (richness), and how similar their abundances are (evenness). Of the dozens of diversity indices (and their close relatives, similarity indices) that have been proposed since Preston [11], many vary only in the relative weight given to richness vs. evenness [12]. Some indices additionally try to emphasise dominance; the role of the most important species in a community [13]. Recently, Warwick and Clarke [14] have proposed diversity indices that also include taxonomic similarity as well as information about the number and relative abundance of species in a collection.

In trying to collapse all information on species richness and relative abundance into a single number, any index of diversity can be misleading in (at least) two ways. Two communities with different numbers of species can have similar estimated ‘‘diversities’’, if the distribution of abundances across species differs in a reciprocal manner (i.e. the richer community is dominated by a few very common species; whereas species in the species-poor community are similar in abundance). Alternatively, two communities which are quite similar in all their common species can still have different estimated ‘‘diversities’’, if many rare species are recorded in one of the communities, whereas few are recorded in the other. Another possible source of concern is that diversity indices intrinsically treat all species as equally informative with regard to community structure and impacts.

None of these problems are insurmountable, because the relative sensitivity of the various indices to richness, evenness, and taxonomic relatedness are known [15]. This knowledge can be used to select the appropriate index for specific applications, although the selection offers a number of meta-problems. To choose wisely which diversity indices to use, one must know which of the major forcers perturbs the community most strongly. To illustrate, fishing rarely extirpates marine species [16], so if fishing is an important pressure, diversity indices which weight richness more strongly than evenness will be insensitive. However, pollution often causes loss of species intolerant to the chemical while allowing the few most tolerant species to thrive, so indices emphasising richness are often recommended [17].

In summary, if a researcher or manager is interested in one specific type of perturbation, and knows a substantial amount about the ecosystem of interest, it may be possible to choose a diversity index with the proper weighting of richness, evenness, and dominance. However, such a strategy allows one's expectations to influence strongly the analytical results to be obtained, which is rarely desirable. The strategy also requires that the perturbation of interest be the dominant one affecting the system, whereas the point of integrated coastal zone management is to address multiple interacting forcing factors.

2.3. Ordination methods

Ordination techniques are a class of multivariate techniques which place things in order. A complex ‘‘community matrix’’ (commonly i rows of sampling sites, j columns of species, cell entries as abundances of species j at site i) is reduced to a small number of dimensions whose axes are mathematically translatable into to the original community matrix. In this reduced space sites with similar species compositions are close together and sites with very different species compositions are far apart. Statistically independent gradients usually are represented as orthogonal ordination axes. This gives the scores for individual species or sites on the ordination axes some desirable statistical properties [18]. The computational steps of identifying the underlying gradients and allocating cases on the gradients are almost always automated. It is the ecological interpretation of the ordination axes that requires skill and judgements.

The various classes of ordination methods differ in the assumptions made about the underlying statistical distributions of the abundances of species and the functional relationship between the species' abundances and the underlying gradients. For example, the frequently encountered principal components analysis (PCA), assumes that the abundance data have normal error structure, zero abundances are rare, and gradient(s) are well sampled across their full range(s). As a consequence, PCA and its close relatives are only applicable when the ecological range sampled is narrow, with similar species present over the full range of sites. Correspondence analyses can be used when the gradient is longer and the abundance matrix is sparser [19]. However, correspondence analysis performs poorly when all species do not share a common, monotonic functional form for their distribution; a problem when dealing with a mixture of both widely distributed species and specialists, or which species that aggregate.

Non-metric scaling (MDS) provides an alternative to dealing with problematic underlying distributions of abundances. With MDS the similarities or dissimilarities among sites are estimated using the weaker assumption that the rank order of abundance of a species across sites is informative, but the actual quantitative estimates of abundance may not be [20]. MDS then represents the species by site matrix in a space defined by many fewer dimensions, but preserving as well as possible the (dis)similarities of cases in the smaller dimensional space. This weaker assumption of ordinal rather than interval information increases robustness in the face of irregular distributions of abundance and high sampling variance. As a result MDS has become a preferred technique for ecological ordinations of marine communities [21]. However, it is not without its problems. MDS methods require specifying a priori the number of axes which exist and a starting configuration for patterns in the data. These a priori requirements create substantial opportunity for the analyst's preconceptions to influence, if not dominate, the analytical results. MDS also requires selecting a measure of (dis)similarity, which brings with it all the complexities discussed under diversity indices.

Choosing among ordination methods can be done objectively, based on statistical properties of the site-abundance matrix. Interpreting results in applied contexts may not be so easy. The simplest interpretations are by narrative; the positions of cases on the extracted axes are examined, and ecological inferences are drawn from knowledge of the cases. Opportunity for expectation bias is high. More rigorously, the values of cases on the ordination axis can be regressed on some environmental variable, reflecting intensity of perturbation due the forcing factor. Unfortunately, unless the forcing factor is such a dominant signal that it describes a community gradient of its own, its effect may be masked altogether when extracting orthogonal axes.

2.4. Aggregated indicators of ecosystem status

Intermediate between full multivariate ordinations of biological communities and condensations of species' abundances into single indices lie some metrics which aggregate information on the occurrences of many species into a single functional or graphical relationship. The two most common aggregate metrics are number (or sometime biomass) spectra and species dominance curves. The former has begun to be used in investigating the effects of fishing on the exploited community [22] whereas the latter has become established as a metric of impacts of pollution on biological communities [23].

Size spectra first partition the sample of specimens into equivalent size classes, and then aggregate numbers (x) in each size class across all species in a collection. The size spectrum refers to the smoothly decreasing relationship of aggregate $\ln(x)$ to size interval across the fully sampled size range. The intercept of the size spectrum of a community is related to ecosystem productivity, whereas the slope of the size spectrum reflects how quickly animals die off. Decreased productivity makes the intercept lower; increased mortality makes the slope steeper [24]. Its theoretical and empirical value as a useful (but not universal) indicator of fishing pressure is well established [25]. However, although of substantial interest to ecological theoreticians, they have been used little in other applied contexts.

Dominance curves present the species in a community or collection ranked by their abundances. Like size spectra, it is the full shape of the curve (or parameters from which the shape may be reconstructed) which contains information about the community under study [26]. The logic behind using k-dominance curves to evaluate ecosystem effects of perturbations is similar to the logic behind using diversity indices: perturbations cause a subset of species which tolerate the perturbation to thrive, while many other intolerant species either disappear or become very rare. The changes would make k-dominance curves of perturbed communities lie above and to the left of the curves of unperturbed communities [27]. They have been primarily used in pollution studies, but appear to be informative about the effects of fishing, as well [28]. Both aggregate metrics seem to be quite informative about changes to communities, although neither will be sensitive to impacts experienced primarily by the uncommon species in a community.

2.5. Metrics of "emergent properties" of ecosystems

This class of indicators of community status move beyond aggregating or ordinating data, to reflect some hypothesised underlying properties of the community or ecosystem. These properties require the intervention of some form of ecological model, representing hypotheses about the trophic (food-web) interactions among taxa in the model. The question of how well a model-derived indicator reflects a core property of the ecosystem cannot be disassociated from the question of how well the model represents the ecosystem. It is impossible for food web models to represent all the species in a community and all their interactions. Therefore indicators from food-web models also reflect users' decisions about which taxa and relationships to include, and how to represent them.

Indicators from classic food web models [29] are things like resilience to species invasions or losses, or stability over perturbations to parameter values. These models and metrics must be used cautiously when evaluating the effects of forcers on food web properties and dynamics of marine ecosystems. Model representations of systems of even moderate complexity are necessarily indeterminate, so they cannot provide specific forecasts of consequences of forcing activities [30]. Also, when life-history linkages within species are added to food-web models they dominate over trophic linkages between species [31]. Subtle details about how life-history dynamics are parameterised then determine the values of the indicators, rather than the inter-species linkages and external forcers about which conclusions are being drawn. When food-web models are used to quantify effects of pollutants and nutrient enrichment users tend to ignore the synthetic model properties, and interpret dynamics of individual nodes as reflecting specific biological consequences of the perturbation [32].

Mass Balance models focus more directly on local parameterisation of biomasses and flow rates [33], with biomass and energy accounted for at all trophic levels. Mass-balance models attempt to avoid the indeterminacy problem by requiring enough biomasses and flows to be specified by the user that the remaining family of simultaneous equations has only one unique set of solutions. However, by setting such data constraints one allows only a single final configuration of the ecosystem to exist, and values of some very uncertain biomasses and flow rates may have to be assumed.

These whole-ecosystem models have been used to evaluate the impacts of fishing, pollution, and climate change on marine ecosystems. For fishing effects, indicators included primary production required to support the fishery itself, the mean trophic level at which the fishery operates, and the transfer efficiency between exploited trophic levels [34]. Ecosystem responses to climate forcing and pollution were measured with transfer efficiencies and distributions of biomass among trophic levels [35]. If the assumptions and formulation of the mass-balance approach are accepted, it is a powerful tool for providing quantitative indicators of ecosystem status and response to forcers. The assumptions and formulations have not been accepted by all marine ecosystem scientists.

3. Potential errors when using of indicators of environmental status

If there were no risk of committing errors when using indicators, then the choice of indicators would be a matter of taste and cost-effectiveness. On the other hand, if errors can be made, then their nature and the associated consequences should be known. This concern should be investigated for use of indicators both as communication tools and as support (formal or informal) for decision-making.

3.1. Potential errors in communications uses

When technical experts want to convey information about their field of knowledge, particularly to a non-technical audience, they have a few choices. Narrative description is always possible. However, when the field of knowledge is the status of the coastal zone environment, it would be a long narrative. The message would require recipients to decide which parts of the narrative were the important ones, to assimilate a large amount of possibly quite detailed information, and construct their conclusions about the status of the system. If they had expectations of what was a desirable or typical state, they would have to compare one narrative description to another. Although these are tasks human beings do daily, there are many opportunities for the message that the technical expert wants to convey to become distorted in the transfer. Differences in vocabulary, personal histories, preconceptions, and value systems all could contribute to distortions of the message, including selective use of a biased subsample of the narrative.

Good indicators should reduce the possibility for distortion of messages. To succeed, they must summarise a large body of technical information into a small number of values that can be interpreted unambiguously. What might cause that goal to not be realised?

One of the communications risks for indicators of environmental status is that the indicator may not carry the intended meaning to the audience. Non-specialists are likely to understand what is meant by the statement that 45,000 puffins breed on an island or 100% of the Canadian population of a small cetacean migrate through a restricted area. However, saying that an area has a Hill's N2 of 3.21 means little to most audiences. Likewise, merely presenting scores of sites on ordination axes communicates little. Even community ecologists need more information about the area to interpret the value of the indicator. There is a message but little meaning. Between those extremes, accuracy of communication can still be distorted by nuances of language. The properties reflected in changing values of a diversity index are not identical with the more colloquial concept of biodiversity, but the similarities are close enough that technical experts and the public can have a long dialogue without understanding each other fully.

The other communications risk is that the colloquial interpretation of the indicator might correspond generally to the technical meaning, but either include information that is not part of the colloquial interpretation, or not cover some facets that a non-specialist would assume to be addressed. For example, the values of indicators derived from ecosystem models, say abundance of top predators, might reflect pooling of many species and complex decisions about linkages among species. However, from the name of the indicator, recipients might infer they were being informed about the status of a few particularly charismatic species. On the other hand, using the first empirical orthogonal factor (a version of PCA population in physical oceanography) to describe the state of the coastal ocean environment tells only part of the story, even if to the technical expert that part of the story is explained very well.

These types of communications errors are not unique to indicators of ecosystem status, nor to indicators of any type. Neither is the protection against them unique to this context. Clarity and care in language is paramount, as is awareness of the target audience. Avoidance of value-laden terms is a virtue. Testing messages with focal groups is effective when there is a specific message that needs to be conveyed to a large audience. All these things are good common sense. Although communications problems will occur, vigilance and co-operation with appropriate experts in communications arts are the keys to keeping them infrequent and minimally damaging.

3.2. *Potential errors in decision support*

3.2.1. *How indicators are used in decision support*

When Indicators of Ecosystem Status are used to support decision making, they can be used as informally as simply one more piece of information in the debate. The values of the indicators are presented, along with any other information that various stakeholders bring forward, and discussion ensues. In such cases, all the concerns about errors of communication with indicators apply, but no new considerations come into play.

Such uses of indicators do not meet standards for best practices in risk management, however, nor do they meet the increasingly formal standards for the application of precaution in government decision-making. The management of risk can include a variety of governance approaches [36] but presupposes that risk has been quantified. It is here that indicators contribute to decision-making in two ways. First, on the continuum of the indicator at least two reference points must be identified. One is a target, associated with a state of the environment which society has identified as desirable and capable of providing satisfactory benefits. The other is a limit, beyond which the risk of harm that is serious or impossible to reverse is unacceptable [37]. Second, the probability that the current value of the indicator is at or above the target reference point, and at or below the limit reference point must be estimated with the maximum possible quantitative rigour. ("Above" and "below" assume that high values of the indicator reflect good conditions and low values, poor ones, as with population data. If the indicator varies in the other direction, such as for contaminant levels, "above" and "below" are switched, of course.)

Effective management must allow fully for uncertainty in estimating both the current value of the indicator and the correct value for the limit reference point. Clients also often find it difficult to operate in a management system that flip-flops from no restrictions to no opportunities. Both of these considerations can be addressed by the use of additional "precautionary" reference points. These are used in risk management contexts to trigger commencement of moderate regulatory restrictions before the risk of violating limits associated with severe conservation or human health issues exceeds the preagreed tolerances [37].

How these probabilities are used then becomes a function of the governance system. In areas with long management histories and strong regulatory capacities, there can be quite rigid control rules. In advance of crisis situations managers and stakeholders agree on specific management actions, tied to specific probabilities of violating limit reference points and/or achieving targets. The decision support consists of estimating the probability that the current state of the indicator violates the limit (or achieves the target), and taking the predetermined action corresponding to the estimated risk. This approach occurs in a few places in resource management [38], and with many human health issues, where, for example beaches may be closed to recreation or shellfish harvesting when coliform bacteria levels exceed pre-identified reference points. In resource management contexts, though, the management system often uses softer rules. The scientific advice still reports the probability of violating limits and achieving targets, and may even recommend actions consistent with a precautionary approach [39]. However, the decision-making uses this information as one of many factors, although often an influential one, in the final decision.

Regardless of the governance use of the information, estimation of two properties associated with indicators is fundamental to effective decision support. One is reference points that are reliably indicative of conservation or human health concerns, and of provision of desired benefits. The other is the probability that the current state of the environment complies with the reference points.

3.2.2. *How error rates would be assessed*

When indicators are used as a component of decision-making in integrated coastal zone management, errors are more than just "misunderstandings". Indicators could be simply uninformative; with no values of the indicator reliably associated with harm or benefits. More seriously, values of an indicator could be hypersensitive, sensitive to things other than those of relevance to the management decision, or even insensitive to factors that users believed were reflected in the indicator. Finally, reference points could be identified incorrectly. In all those cases probabilities derived from current estimates of the indicator relative to reference points would be unreliable guides to decisions. Basing management decisions on indicators that were uninformative about true ecosystem status would have managers chasing noise, and squandering both resources and credibility. Using indicators that were either hypersensitive, or sensitive to the wrong things could be the basis for management actions that perturb ecosystems more seriously, rather than retain them within usual ranges of variation. Using indicators that were insensitive could lead to failure to take necessary conservation actions. Finally, positioning reference points incorrectly could lead to failure to take necessary actions or to unjustified restriction on activities that would pose little risk of harm. Because any of these errors could have serious consequences, it is important to understand the way that indicators can increase or decrease the risk of committing them.

Signal detection theory [40], well-established in psychology and engineering, provides a useful framework for developing that understanding. Selecting informative indicators of ecosystem status differs little from selecting informative signals about anything else. Managers need to be warned when environmental events have occurred in the real world so that measures to minimise harm or maximise opportunity can be implemented promptly. Managers do not want frequent erroneous warnings, however, prompting them to act unnecessary, and possibly creating problems where none existed.

Signal detection theory in built on the 2×2 matrix of real-world events and indicator status, using the four possible combinations, labelled:

- hits—event occurred and the signal says something happened,
- misses—event occurred but the signal is indistinguishable from background noise,
- false alarms—nothing actually happened but the signal says something did,
- true negatives—nothing happened, and there was only background noise in the signal.

A perfect signal produces only hits and true negatives, but in a noisy environment, a perfect signal is impossible. The theoretical framework was developed to optimise performance of imperfect signal detection systems. In this optimisation signal detection theorists focused on several factors that can guide in the selection of indicators.

(1) The costs of the two types of errors may not be symmetrical, and the asymmetry usually can be specified in advance. By "cost" in our context, we mean the consequences of managers not acting when needed vs. acting when the actions are unnecessary.
(2) Using knowledge of the *relative* cost of the two types of errors, indicators can be selected whose error patterns best match the cost ratio.
(3) Knowing the *absolute* cost of errors guides investments in various monitoring and signal detection systems. Do the benefits of fewer errors from a more reliable indicator justify the cost of improved monitoring?

The parallel between developing reliable signals and selecting informative environmental indicators is direct. Informative environmental indicators tell managers about changes in ecosystem status. There will always be background variation (noise) in the indicator, from which signals requiring action must be differentiated. The differentiation can be done probabilistically in a risk management framework, or as a binary decision of a reference point either being exceeded or not. Science advisors, coastal managers, and stakeholders can explore in advance what the possible consequences of misses and false alarms are, both to the ecosystem and to those using the coastal zone. Misses are likely to have very different costs than false alarms, with the costs of false alarms often borne by society, and costs of misses borne by the ecosystem (and, of course. later by society). The interactions between error tolerance rates and costs to improve indicator performance can be discussed among all parties with a role in governance.

Thus, we have a framework for objective screening of candidate indicators of environmental status. Select case histories where, with all the benefits of hindsight, we know when some events occurred of magnitudes that coastal zone managers should have reacted. Calculate the values of the suite of candidate indicators for the areas, and see when the risk quantified for each indicator relative to its reference point would be large enough that technical advisors would have recommended action. Which ones have error rates within our tolerances? Which ones have miss and false alarm ratios that match our cost tolerances? Are the best error rates achievable good enough to provide the desired likelihood of achieving coastal zone management objectives?

Using a signal detection framework to select environmental indicators does not guarantee that we will find an effective basis for science advice on coastal zone management. Informative metrics with acceptable error rates and distributions may not exist, or be prohibitively expensive. It may be impossible to get consensus among diverse, legitimate stakeholders in what are acceptable error rates and distributions. These situations would still be successes for the approach, because it would have informed managers and stakeholders what to expect from the indicators being used in making decisions about coastal zone management, that they assume are science-based and risk averse.

4. Likely performance of the classes of environmental indicators

A signal detection approach for evaluating suites of candidate environmental indicators, would be more demanding than choosing indicators through a consensus-building exercise. It would also be more objective, particularly when done relative to risk of violating reference points, and provide indicators with known reliability. Despite these virtues to my knowledge no one has tried to apply it in selecting indicators for environmental risk management. Nonetheless, there have been enough trials with both field applications and simulations to allow conjectures about likely error tendencies of the classes of indicators presented above. The issue of determining sound reference points for each type of indicator will be addressed subsequently.

4.1. Indicator species as reliable signals

There is cause for cautious optimism here. When an indicator species was chosen specifically because of its relationship to a particular forcer, one would expect a priori that fluctuations in the indicator species relative to its reference point would be responding to changes in the forcer, and be informative to managers about appropriate actions. This should be true when the indicator species itself is the ecosystem property of interest, which is the case with species on protected or endangered lists. It is less likely to be met when the species was chosen primarily for convenience or to be a focus for stakeholder engagement. When the presence of a species is diagnostic of particular conditions, such as a pollution-tolerant species, false alarms may be rare, but miss rates might be high if the species has additional ecological restrictions or slow dispersal.

The experience with fisheries is informative here. Fisheries are a particularly strong forcer, affecting the target species directly (in addition to all their other effects [41]).

Many jurisdictions make large investments in monitoring several indicators of stocks status, identifying reference points, and synthesising results into integrated assessments of stock status. The individual indicators often demonstrate high error rates. Fishery independent surveys often produce false alarms, with single-year values turning out to be high or low "outliers" [42] whereas fishery-based catch per unit effort indicators often produce misses, due to fishing strategies that can maintain catch rates as stocks decline [43]. These patterns suggest that where indicator species are to be used in coastal zone monitoring programs, there is a potential high risk of each type of error, depending on the nature of the monitoring program.

4.2. Diversiy indices as reliable signals

Diversity and similarity indices blend changes in abundance or occurrence of many species into a single indicator. In marine ecosystems, big changes in indices emphasising richness would occur only when many species typical of an area were found to be missing or many additional species were encountered in a survey. Such changes might occur in response to atypical oceanographic conditions, when large changes in species composition can occur [44]. Such changes would not result in a false alarm, as the assemblage really had changed, and managers should know about it, even if there was little they could do to return to the assemblage to its previous composition except wait. *False alarms* could occur if sampling methods changed, affecting detectability of species. Science advisors ought to know if sampling methods changed, and frame their advice using the indices appropriately. However vigilance, ability, and enthusiasm of the samplers for identifying uncommon species can change in a time series without being recorded, causing large *false alarms* using richness-based indices [45]. Richness-based indices can also *miss* changes in assemblages that may be important to managers, because as long as the total number of species does not change greatly, there can be major changes in species composition without any signal in the indicator.

As noted earlier, evenness-based indicators can be very good at picking up effects of pollutants where species have highly differential tolerances. They can give misleading *false alarms*, however, when diversity values fall due to the recruitment of strong year-classes to a species in the assemblage [46]. More seriously, they can *miss* forcers which increase mortality rates of an assemblage, as long as the mortality is not highly species selective. In fact, if the more common species are experiencing somewhat greater increases in mortality, evenness-based diveristy indices will indicate diversity is increasing. This may be erroneously interpreted as a good thing, possibly encouraging managers to continue practices that detrimental to the entire assemblage.

4.3. Ordination scores as reliable indicators

Ordination methods are vulnerable to both misses and false alarms. Management applications ordinate the multispecies samples from a block of years, using the resultant species and site scores on major axes as a reference framework. Scores of monitoring sites are then tracked through ordination space over time. When sites make big excursions, big ecosystem changes at those sites are inferred. Because ordination methods are variance-structuring tools, the initial ordination is dominated by the species that were most variable in the reference years. If these are species whose abundances are inherently highly variable in space and time or there is a lot of sampling error, ordinations will give many *false alarms*. Alternatively, species which were initially either uncommon everywhere, or abundant and widespread but without clearly defined optima will get little weight in the initial ordination axes. Hence major increases in rare species, or declines in common but eruptive species, will not show up as big excursions of sites in ordination space, and be *misses*.

It is, of course, possible to ordinate and interpret a full time series. If there were long-term trends, either driven by environmental forcing or incremental effects of anthropogenic perturbations, they could be apparent in the systematic ordering of samples over time along latent axes. *Misses* would be rare if the forcing factor of interest contributed a meaningful and systematic portion of the total variance, so it was an independent axis itself. *Misses* would be common if different parts of the system responded to the forcing factor at different rates. Not all extracted axes would be informative about the effects of the forcing factor. *False alarms* could be common or rare, depending on the ability to interpret the information content of the axes in non-circular ways (that is, in ways that used information independent of the ordination itself to evaluate which axes were reflected effects of forces to which managers should respond).

4.4. Integrated indicators as reliable signals

Compared to diversity indices calculated from the same data, size spectrum methods should provide few *misses* or *false alarms* of changes in overall productivity, transfer efficiency, or mortality rate. *False alarms* should be uncommon because ln(number) spectra reduce distortions due to occasional sampling anomalies or strong year classes of one or a few species. If strong recruitments occur for several species or years, this *is* an increase in system productivity, and will affect the metric's value. The same is true, although possibly to a lesser extent, for dominance curves. They may become more markedly humped with sampling or recruitment anomalies, but unless productivity and survivorship of the rest of the species being sampled is affected at the same time, the dominance curve is unlikely to change substantially.

Both summary indicators will *miss* some types of ecosystem change, such as high turnover rates of uncommon species. However, few indicators will have a high hit rate for such types of community changes, where meta-populations of many uncommon species are being lost and re-colonised over time. Moreover, system changes due to several uncommon species becoming even rarer will be captured well by dominance curves, and if they are larger individuals, by size spectra as well. These types of changes might be of particular concern to coastal zone managers.

4.5. "Emergent property" indicators as reliable signals

Because of the diversity of ecosystem models, it is impossible to make general statements about the *hit, miss,* and *false alarm* rates of model properties, relative to true changes in the ecosystems being modelled. In fact, it may be inappropriate to view model-based ecosystem-model based indicators within a signal detection framework at all. Rather, ecosystem models that produce emergent properties, such as resilience, or mean trophic level, are posing hypotheses about the structure and dynamics of the ecosystems they are intended to represent. It is more conceptually consistent to view their performance as exploring scenarios about system responses. That is, the questions are of the nature "If the ecosystem functions as the model does, and if it is perturbed in such-and-such a way, this is how various properties will respond". Those can be important questions for a coastal zone manager, and the model may provide useful information. However, for changes in model outputs to be interpreted as indicators of the changing status of real ecosystems, it would be necessary to re-parameterise the model for each new value of the indicator. Otherwise the trajectory itself becomes a single indicator of the status under one set of conditions, and other trajectories, from other parameterisations or initialisations, become other values for the indicator.

It would be an interesting question, to my knowledge not yet explored, to determine how best to compare such indicators over time or space, and evaluate their error rates. It is worth noting though that the Ecosystem Effects of Fishing Working Group of ICES has addressed a related question twice and came to the same conclusion. If management reference points for ecosystem components other than biomass and fishing mortality of the target species were needed to ensure conservation of the ecosystem, what should be added to the suite of reference points? WGECO agreed that indicators for biomass and mortality of species taken as bycatch, for genetic diversity within species, for status of ecologically dependent predators, and for status of species preyed on by scavengers on fishery discards and offal, would all be essential. However, these are all single species indicators; merely for a longer list of species than those targeted by the fishery. WGECO concluded that at the time of the reviews (1999 and 2001) there was no evidence that more "integrated" ecosystem properties would be at risk from fishing, if compliance was high with the longer list of single-species indicators.

5. Selecting reference points for classes of indicator

The discussion of indicators within a signal detection theory context presupposes that it is possible to know when a signal has said that an event has occurred, whether the statement is a hit or a false alarm. When reference points are used in a risk management framework (including a precautionary approach) this knowledge is direct; the current value of the indicator exceeds the pre-identified risk tolerance of the reference point. How readily can the reference points be set for the various classes of indicators?

5.1. Reference points for indicator species

There is a large literature and extensive experience with setting reference points for indicator species. For exploited species, many science advisory and management agencies use limit and precautionary or buffer reference points as core tools (ICES NAFO, NASCO). Different jurisdictions may use somewhat different criteria for positioning the reference points on abundance and mortality rate indicators, some using more aggressive strategies than the F_{msy} and B_{msy} suggested by UNFA. What matters though, is that once the criteria are made explicit, estimating both the reference point and the probability that a current value violates the reference point, is just one more task in the assessment of stock status. If the data for the stock make a traditional assessment straightforward, it will be straightforward to estimate the reference point and risk as well. If the assessment is difficult and unstable, so will be the task of estimating reference points and risks.

The threatened or endangered species community also has extensive experience with using quantitative reference points. The IUCN Red List of 1996 [47] is a benchmark in using quantitative reference points for determining the appropriate category for species being evaluated. Although the IUCN SSG work refers to guidelines and criteria, in practice they are using reference points in combination with soft decision rules. Great effort when into determining what the criteria should be, and the values have generated substantial controversy [48]. The experience illustrates again, though, that estimation of reference points for indicator species is practical, as long as historic data are available, and are informative about what states comprise harm that is serious or difficult to reverse.

There is another message for coastal zone management in the experience of reference points used with fisheries and endangered species. These are the simplest cases, but they have been neither easy nor free of controversy. Stock assessments have not inspired uniform belief outside (or even within) the fisheries community. Application of IUCN criteria or their derivatives has been hotly contests for some types of species, and even the experts can argue at length about the proper category for a species, given the quantitative criteria, reference points, and the same data available to all. We should expect the task to be only harder for more complex classes of indicators.

5.2. Reference points for diversity indices

Reference points for diversity or similarity indicators are problematic. Modern simulation and resampling tools allow estimation of the uncertainty of current values of diversity indices, so risk of diversity or its components being below pre-identified values can be quantified. The difficulty is providing justification for any pre-identified reference point. When historic data are available, values can be calculated for previous periods. However, only if some of those time periods were periods when the status of the environment was unacceptable would the historic values be informative about defensible reference points to use as a basis for management decision-making. There is limited ability to borrow experience from other areas, because the components of diversity, particularly richness, vary with many factors; latitude, productivity, substrate, exposure, depth, etc. This makes it hard to justify taking a value that is well justified as a reference point for one area, and using it as a *de facto* reference point for other areas.

5.3. Reference points for ordination scores

Ordination scores give slightly more opportunity for objectively setting reference points than do diversity indices. When the only information available is the data used in the ordination then the situation is identical for the two types of indicators. Historic data can be used to provide estimates of ordination scores characteristic of previous states of the ecosystem. However, only if some of those states were unacceptable are they informative about where in ordination space reference limits (points on one axis, boundaries in spaces of more than one axis) should be located. Resampling methods can provide estimates of uncertainty of specific values of ordination scores, and when all the assumptions of the ordination are met, uncertainties can be estimated parametrically as well, so risk quantification should usually be possible.

The potential gain with ordination scores comes when there are data available for even a subset of sites, regarding independent variables that may be directly indicative of the forcer of interest; for example contaminant levels, nutrient loading, etc. In such cases, methods such as direct gradient analysis or canonical correspondence analysis [49] can be used to relate trends in the forcer to trends in the ordination scores. As long as one can justify extrapolation beyond the parameterisation data, the results can then be used to identify the location(s) in ordination space associated with unacceptable levels of the forcer and/or state of the ecosystem, even if the conditions have not yet been observed. These then can be used as objectively determined reference points.

5.4. Reference points for integrated indicators

Aggregate indicators like size spectra and dominance curves have not been used in formal decision support contexts, although k-dominance curves have been considered in evaluating pollution effects. The strong theoretical basis for size spectra, and the corresponding empirical and modelling support for systematic behaviour of the slope of the spectrum as a function of fishing mortality both suggest that in principle it should be possible to establish a slope for a spectrum that would correspond to an unsustainable mortality rate for the community as a whole. This would require some initialisation data, from sampling done when the community was considered healthy. The mortality rate associated with the observed slope of that community could be the basis for modelling to establish the mortality rate that would be unsustainable. The slope associated with that mortality would then be the reference point. Typical regression methods can estimate the uncertainty in the estimate of the slope from annual samples, so the risk quantification relative to the reference slope would be possible. Reference points for the intercept of the size spectrum of a community, reflecting changed productivity rates, would be harder to derive from a base sample and calculations with a sound theoretical basis. However, if the concern is with changed rates of productivity for the system, it would be wisest to simply measure productivity directly rather then infer it indirectly from size spectra.

Similar analyses probably would be possible for dominance curves. However, the abundances of individual species are preserved as dominance curves (but not the size spectra) are assembled. Therefore the approach would deal poorly with size-dependent mortality, which is pervasive in marine ecosystems [50]. Thus although resampling statistics have been developed for dominance curves, it is hard to see how a reference point could be justified theoretically.

5.5. Reference points for emergent properties of models

If one believes in a model strongly enough to use emergent model properties as indicators of ecosystem status, one should believe in the model enough to use it to estimate the boundary conditions beyond which the modelled system has suffered serious or irreversible harm. Hence it should always be possible to estimate reference points for model properties, as long as users understood the model being used. Quantifying the risk of violating the reference point, once set, is not as simple. Most ecosystem models complex enough to produce emergent properties require a lot of initialisation data. It is unlikely that all the initialisation data can be collected uniquely each year (or each time the model is run to advise managers). It is common, in fact, for much of the initialisation data to be drawn from different sources—different time periods of sampling, different levels of sampling error, etc, and only a subset of the inputs are updated with each run. In such cases the value of the indicator each year will be a mixture of influences of data values that have changed with other values that have not (whereas all the components will have changed in the system that is being managed). Depending on functional relationships in the model, particularly if some are non-linear, each year's value of the emergent property may reflect a different balance between inertia and buffering by data carried over, and responsiveness to inputs that were updated. This makes actual risk estimation, relative to reference points for the emergent property, unreliable. The model may have many ways to produce estimates of the uncertainty associated with various model properties. However, interpreting these uncertainties as quantified risk in the system represented by the model is making a profound statement of faith in the model. I have seen few models in which I have that much faith.

6. Criteria for selecting indicators

In Sections 4–6 I have laid out a rigorous framework for objectively selecting indicators whose performance standards would be known. All the steps in applying the framework are practical, but it reflects a noteworthy workload. Groups like SCOR Working Group 119 (Ecosystem Indicators of Fisheries Effects), and a large research team in Australia, are pursuing initiatives this ambitious. In the medium term it is realistic to expect the tools to be developed that would make operating with the necessary rigour widely possible. However, in the short term, it is likely that indicators will continue to be chosen on the basis of expert opinion and group consensus, rather than objectively determined performance standards. Towards that end a Task Group within SCOR WG 119 (Marie-Joelle Rochet, Philippe Cury, and I) have been trying to codify practical screening criteria for other Task Groups to use in selecting among indicators of ecosystem status for various system properties. This work extended work reported by the ICES Ecosystem Effects of Fishing Working Group, in their recommendations for scientific advice on Ecosystem Qualities and Ecosystem Quality Objectives to OSPAR and the North Sea Council of Ministers. Those developing selection criteria would be relevant for coastal zone managers, faced with the need to choose among a large suite of possible indicators.

- *Meaning*: Is the indicator readily understood by decision-makers and stake-holders, so values, when reported are interpreted meaningfully? As discussed in the section on Communications Errors, a particularly undesirable situation occurs when the public has a high awareness of the factor, but their interpretation of it is different than the meaning to technical experts. Meaning should be formally evaluated with the targeted audiences.
- *Measurement*: What needs to be done to provide the data for regular updates of the value of the indicator? Desirable situations are when the indicator can be calculated from data that already being provided by existing monitoring programs, and are in the public domain. It is also desirable that the tools needed to take the measurements are widely available, inexpensive to use, and of known and consistent accuracy and precision. All these considerations can be evaluated directly.
- *Accuracy/precision*: How well do the values of the indicator reflect the actual state of the environment? To be useful as anything more than narratives, the uncertainty and bias of indicators must be measurable and known. Indicators are particularly problematic when variance or bias is not stable over time or space, so the actual information content of the indicator is unstable. When there is potential for bias, it is particularly important to know the direction of the bias, so users know whether the indicator may underestimate the seriousness of environmental threats or overestimate the effectiveness of management actions. The more directly accuracy and precision can be measured the better, although it is common for simulations rather than field experiments to be the only practical way to measure these properties of an indicator.
- *Representativeness*: Can the dynamics of the indicator be taken to reflect more than the dynamics of the specific times and places where the data were collected? Seasonal and geographic variation in the properties being monitored are crucial here, as is the selectivity of the sampling tools. The interpretability of indicators degrades quickly if seasonal variation cannot be calibrated effectively or if sampling tools are not intercalibrated. Beyond those concerns, indicators become more useful as the legitimacy of generalising to larger geographic areas or more

taxonomic groups increases. Representativeness can be explored directly with experiments or post-stratification of samples, and indirectly using simulations.

- *Availability of historic data*: Is there a frame of reference in which to build an understanding of the performance of an indicator? The reasons for this are developed fully in Sections 3 and 4. Availability of historic data can be evaluated directly for any indicator.
- *Specificity*: If an indicator changes value, how indicative is the change of the impact of a particular forcing factor? Sensitivity to environmental variation may be desirable in a general indicator of ecosystem status. However if an indicator is used as a tool in management of human activities, such sensitivity can pose major interpretational problems, unless the linkage between the environmental forcing and the indicator is quantified well. These linkages have proven difficult to quantify, even for indicators of effects of fishing on target species [51]. Similarly, if an indicator is to be used in management of a single activity in the coastal zone, sensitivity to many other human activities creates wide scope for debate about the need for and effectiveness of specific restrictive measures. Sensitivity is best established through controlled experiments, but is much more commonly done through correlative studies of parallel time series, with all the dangers inherent in such approaches.
- *Ability to set reference points*: This was discussed fully in Section 5. Reference points associated with serious or irreversible harm can only be established with time series that have high contrast, or indicators supported by extremely robust theories.
- *Sensitivity*: Can the indicator be trusted to inform users fully if the environment really is undergoing change in properties of concern? The need for sensitive indicators was discussed in Section 3. The best indicators have smooth, monotonic relationships of high slope with changes to the ecosystem that they represent. Sensitivity can be established with controlled experiments, through careful analysis of historic time series, and in some cases, through simulation.
- *Responsiveness*: Will the indicator give feedback on ecosystem change on time scales that are useful in management decision-making? Indicators that do not inform stakeholder or managers of ecosystem changes until several years after the changes have occurred, are of very limited value. Responsiveness can be evaluated in the same ways as sensitivity.
- *Legal considerations:* Is the indicator something for which a legal requirement to report exists? The requirement could arise from international agreements to community bylaws. In most cases however, environmental scientists should consult legal professionals, rather than trusting their ability to interpret the intent and meaning of legal texts.
- *Theoretical basis*: Is there a sound basis in ecological theory for why the indicator should be reliably informative about effects of particular forcers? This is not a reason to reject empirically based indicators out of hand. However, given the tendency of correlation-based relationships in marine ecosystems to fall apart over time [52], consistency with uncontested ecological theory is a desirable additional property. On the other hand, indicators derived from ecological

theories that are hotly contested among professionals may pose a high risk of mis-informing users, if parts of the contested theory must be revised in future. In evaluating this property for an indicator, one must evaluate not only the consistency of an indicator with ecological theory, but also the degree to which the diversity of professional views all accept the theoretical arguments.

7. Conclusions

- There are hundreds to thousands of potential indicators of ecosystem status that can be used in coastal zone management applications. They range in complexity from single-species indicators, through single-value indicators of diversity or similarity, to data ordinations, integrated (often graphical) indicators, and emergent properties of ecosystem models.
- Managers and science advisors have gained substantial experience with some of the possible indicators in some types of applications, but there have been few systematic evaluations of how well the different types of indicators match the requirements of the different types of applications of the indicators.
- Indicators can communicate incorrect information when the information content of the indicator has not been established rigorously by the technical experts, but also when the expectations of the recipients do not match the information content of the indicators.
- Indicators can contribute to two types of errors when used in decision-making contexts; the indicator could fail to inform about events that have occurred in the real world, or can provide false alarms about events that did not happen. The framework of signal detection theory can be used to evaluate the performance of indicators relative to these error rates, their respective costs, and society's risk tolerances.
- When used in decision-making contexts, indicators can guide risk management, but effective risk management requires that reference points can be set for the indicator, and risk of current status of the indicator relative to the reference point can be estimated.
- All the categories of indicators may perform well in the formal signal detection and risk management settings, but performance of even the relatively simple single-species indicators can be poor. It may not even be possible to establish the performance rates for complex indicators, particularly emergent properties of models, except through trial and error.
- It is unlikely that the formal evaluation of large numbers indicators required for objective choices among them will occur in the near future. In the interim, more consensus-based approaches to selecting indicators of ecosystem status should be based on 11 properties: Meaning, Measurement, Accuracy/precision, Representativeness, Availability of historic data, Specificity, Ability to set reference points, Sensitivity, Responsiveness, Legal requirements, Theoretical basis.

References

[1] Belfiore S, Balgos M, Galofre J, McLean B, Blaydes M, Tesch D. Background Paper; International Workshop on the Role of Indicators in Integrated Coastal Management. IOC document, 2002.

[2] Rice JC. Evaluating fishery impacts using metrics of community structure. ICES Journal of Marine Science 2000;57:682–8.

[3] Jennings S, Kaiser MJ. The effects of fishing on marine ecosystems. Advances in Marine Biology 1998;34:201–352.

[4] ICES. Report of the ICES advisory committee on fisheries management, 2001; ICES Cooperative Research Report 246(3 volumes);
Philippert CJM. Long-term impact of bottom fisheries on several bycatch species of demersal fish and benthic invertebrates in the south-eastern North Sea. ICES Journal of Marine Science 1998;55:342–52.

[5] Blacker RW. Benthic animals as indicators of hydrographic conditions and climatic change in Svalbard waters. Fishery Investigations (London) Ser. II 1957;20:1–49.

[6] ICES. Report of the ICES advisory committee on the marine environment, 1999. ICES Cooperative Research Report No. 239, 1999. 277pp.

[7] Spellerberg IF. Monitoring ecological change. Cambridge, UK: Cambridge University Press, 1991. 334pp.

[8] Menge BA. Indirect effects in marine rocky intertidal interaction webs: patterns and importance. Ecological Monographs 1995;65:21–74;
Zacharias MA, Roff JC. Use of focal species in marine conservation and management: a review and critique. Aquatic Conservation: Marine and Freshwater Ecosystems 2001;11:59–76.

[9] ICES. Report of the working group on ecosystem effects of fishing. ICES CM 1996/Assess/Env:1, 1996.

[10] OSPAR BDC. Testing of the Faial criteria for the selection of species and habitats which need to be protected with the deep-water and habitat-forming coral reef species *Lophelia pertusa*. Paper 00/6/5 presented by Norway to OSPAR Biodiversity Committee, 2000.

[11] Preston FW. On the commonness and rarity of species. Ecology 1948;29:254–83.

[12] Hill MO. Diversity and evenness: a unifying notation and its consequences. Ecology 1973;54:427–32;
Magurran AE. Ecological diversity and its measurement. London: Chapman & Hall, 1988.

[13] McNaughton J. Relationship among functional properties of California grassland. Nature 1967;216:168–9.

[14] Warwick RM, Clarke KR. New "biodiversity" measures reveal a decrease in taxonomic distinctness with increasing stress. Marine Ecology Progress Series 1995;129:301–5.

[15] Ludwig JA, Reynolds JF. Statistical ecology. New York: Wiley, 1988. 337pp;
Magurran AE. Ecological diversity and its measurement. London: Chapman & Hall, 1988.

[16] Powles H, Bradford MJ, Bradford RG, Doubleday WG, Innes S, Levings CD. Assessing and protecting endangered marine species. ICES Journal of Marine Science 2000;57:669–76;
FAO. An appraisal of the suitability of the CITES criteria for listing commercially exploited aquatic species. FAO Fisheries Circular, vol. 954, 2000. 66pp.

[17] Spellerberg IF. Monitoring ecological change. Cambridge, UK: Cambridge University Press, 1991. 334pp;
Diez I, Secilla A, Santolaria A, Gorostiagia JM. Phytobenthic intertidal community structure along an environmental pollution gradient. Marine Pollution Bulletin 1999;38:463–72.

[18] Gauch HG. Multivariate analysis in community ecology. Cambridge UK: Cambridge University Press, 1982. 298pp;
Jongman RHG, ter Braak CJF, van Tongeren OFR. Data analysis in community and landscape ecology. Pudoc-Wageningen, Den Haag, 1987. 299pp.

[19] Hill MO. Correspondence analysis: a neglected multivariate method. Applied Statistics 1974;23: 340–54.

[20] Kruskal JB. Multi-dimensional scaling by optimizing goodness of fit to a non-parametric hypothesis. Psychometrica 1964;29:1–27;

McRae G, Camp DK, Lyons WG, Dix TL. Relating benthic infauna community structure to environmental variables in estuaries using non-metric multi-dimensional scaling and similarity indices. Environmental Monitoring and Assessment 1998;51:233–46.

[21] Clarke KR, Ainsworth M. A method of linking multivariate community structure to environmental variables. Marine Ecology Progress Series 1993;92:205–19;
Cao Y, Bark AW, Williams WP. Measuring the response of macroinvertebrate communities to water pollution: a comparison of multivariate approaches, biotic and diversity indices. Hydrobiologia 1996;341:1–19.

[22] Rice JC, Gislason H. Patterns of change in the size spectra of numbers and diversity of the North Sea fish assemblage, as reflected in surveys and models. ICES Journal of Marine Science 1996;53:1214–25;
Bianchi G, Gislason H, Graham K, Hill L, Jin X, Koranteng K, Manckchand-Heileman S, Paya I, Sainsbury K, Sanchez F, Zwanenburd K. Impact of fishing on sinze composition and diversity of demersal fish communities. ICES Journal of Marine Science 2000;57:558–71.

[23] Clarke KR. Comparisons of dominance curves. Journal of Experimental Marine Biology and Ecology 1990;138:143–57;
Spellerberg IF. Monitoring ecological change. Cambridge, UK: Cambridge University Press, 1991. 334pp.

[24] Boudreau PA, Dickie LM. Biomass spectra of aquatic ecosystems in relation to fisheries yield. Canadian Journal of Fisheries and Aquatic Science 1992;49:1528–38;
Thibaux ML, Dickey LM. Structure of the body-size spectrum of the biomass in aquatic ecosystems: a consequence of allometry in predator–prey interactions. Canadian Journal of Fisheries and Aquatic Science. 1993;50:1308–17.

[25] Gislason H, Rice JC. Modelling the response of size and diversity spectra of fish assemblages to changes in exploitation. ICES Journal of Marine Science 1998;55:362–70;
Gislason H, Lassen H. On the linear relationship between fishing effort and the slope of the size spectrum. ICES CM 1997/DD:05, 1997;
Bianchi G, Gislason H, Graham K, Hill L, Jin X, Koranteng K, Manckchand-Heileman S, Paya I, Sainsbury K, Sanchez F, Zwanenburd K. Impact of fishing on sinze composition and diversity of demersal fish communities. ICES Journal of Marine Science 2000;57:558–71.

[26] Beukema JJ. An evaluation of the ABC method (abundance biomass comparison) as applied to macrozoobenthic communities living on tidal flats in the Dutch Wadden Sea. Marine Biology 1988;99:425–33;
Clarke KR. Comparisons of dominance curves. Journal of Experimental Marine Biology and Ecology 1990;138:143–57.

[27] Lambshead PJD, Platt HM, Shaw KM. The detection of differences among assemblages of marine benthic species based on an assessment of dominance and diversity. Journal of Natural History 1983;17:859–74;
Clarke KR. Comparisons of dominance curves. Journal of Experimental Marine Biology and Ecology 1990;138:143–57.

[28] ICES. Report of the working group on ecosystem effects of fishing. ICES CM 1998/ACFM/ACME:1;
ICES 2001. Report of the working group on ecosystem effects of fishing. ICES CM ACME:01, 1998.

[29] Pimm SL. Food webs. New York: Chapman & Hall, 1982;
Cohen JE, Briand F, Newman CM. Community food webs: data and theory. Berlin: Springer, 1990. 308pp.

[30] Yodzis P. Introduction to theoretical ecology. New York: Harper & Row, 1989;
Yodzis P. Predator-prey theory and management of multispecies fisheries. Ecological Applications 1994;4:51–8.

[31] Pimm SL, Rice JC. The dynamics of multi-species, multi-life stage models of aquatic food webs. Theoretical Population Biology 1987;32:303–25.

[32] Breitburg DL, Sanders JG, Gilmore CC, Hatfield CA, Osmun RW, Riedel GF, Seitzinger SP, Sellner KG. Variability in responses to nutrients and trace elements, and transmission of stressor effects through an estuarine food web. Limnology and Oceanography 1999;44:837–63.

[33] Pauly D, Christensen V, Walters C. Ecopath, ecosim, and ecospace as tools for evaluating ecosystem impact of fishing. ICES Journal of Marine Science 2000;57:697–706;
Walters C, Christensen V, Pauly D. Structuring dynamic models of exploited ecosystems from trophic mass-balance assessments. Reviews in Fisheries and Fish Biology 1997;7:139–72.

[34] Jarre-Teichmann A. Seasonal mass-balance models of carbon flow in the central Baltic Sea with emphasis on upper trophic levels. ICES CM 1995/T:6, 1995. 26pp;
Jarre-Teichmann A. The potential use of mass-balance models for the management of upwelling ecosystems. Ecolological Applications 1998;8: 93–103;
Pauly D, Christensen V, Walters C. Ecopath, Ecosim, and Ecospace as tools for evaluating ecosystem impact of fishing. ICES Journal of Marine Science 2000;57:697–706.

[35] Dalsgaard J, Jarre-Teichmann A, Walters C, Pauly D. An approach to the modelling of persistent pollutants in marine ecosystems. ICES CM/1998, vol. 10, 1998;
Dalsgaard J, Wallace SS, Salas S, Preikshot D. Mass-balance model reconstruction of the Strait of Georgia: the present, one hundred, and five hundred years ago. UBC Fisheries Centre Research Reports 1998;6:72–91.

[36] Townsley P. Social issues in fisheries. FAO Fisheries Technical Report 375, 1998. 93pp.

[37] Garcia SM. The precautionary approach to fisheries, its implications for fishery research, technology, and management: an updated review. FAO Fisheries Technical Paper No 350, Part 2, 1996. pp. 1–76.

[38] Bell ED, Stefansson G. Performance of some harvest control rules. NAFO Scientific Council Research Document 98/7, 1998. 19pp.

[39] ICES. Report of the ICES advisory committee on fisheries management, 2001; ICES Cooperative Research Report 246(3 volumes), 2001.

[40] Helstrom CW. Statistical theory of signal detection. Oxford: Pergamon Press, 1968. 470pp.

[41] Hall SJ. The effects of fishing on marine ecosystems and communities. Oxford: Blackwell Science, 1999.

[42] Rice JC. Changes to the large marine ecosystem of the Newfoundland—Labrador shelf. In: Sherman KS, Skoldal HR, editors. Large marine ecosystems of the North Altantic. Oxford: Blackwell Scientific Press, in press.

[43] Rose GA, Kulka DW. Hyperaggregation of fish and fisheries: how catch-per-unit-effort increased as the northern cod (*Gadus morhua*) declined. Canadian Journal Fisheries and Aquatic Sciences. 1999;56(Su):118–27.

[44] Anderson PJ, Piatt JF. Community reorganization in the Gulf of Alaska following ocean climate regime shift. Marine Ecology Progress Series 1999;189:117–23.

[45] Daan N. A spatial, temporal diversity index taking into account species rarity, with an application to the North Sea fish community. ICES CM 2001/T:04, 2001. 6pp.

[46] Rice JC, Gislason H. Patterns of change in the size spectra of numbers and diversity of the North Sea fish assemblage, as reflected in surveys and models. ICES Journal of Marine Science 1996;53:1214–25;
ICES. Report of the working group on ecosystem effects of fishing. ICES CM 1996/Assess/Env:1, 1996.

[47] Baillie J, Groomsbridge B. 1996 IUCN red list of threatened animals. Cambridge UK: IUCN Publishing, 1996.

[48] Mace GM, Collar N, Cook J, Gaston K, Ginsberg J, Leader-Williams N, Maunder M, Miller-Gulland EJ. The development of the new criteria for listing species on the IUCN Redlist. Species 1992;19:16–22;
Musick JA. Endangered marine fishes: criteria and identification of North American stocks at risk. "Fisheries" (Bethesda) 1998;23(2):28–30;
FAO. An appraisal of the suitability of the CITES criteria for listing commercially exploited aquatic species. FAO Fisheries Circular, vol. 954, 2000. 66pp.

[49] Whitaker RH. Gradient analysis of vegetation. Biological Reviews 1967;49:207–64;
Gauch HG. Multivariate analysis in community ecology. Cambridge, UK: Cambridge University Press, 1982. 298pp;
Clarke KR, Ainsworth M. A method of linking multivariate community structure to environmental variables. Marine Ecology Progress Series 1993;92:205–19.

[50] Dickie LM, Kerr SR, Boudreau PR. Size-dependent processes underlying regularities in ecosystem structure. Ecological Monographs 1987;57:233–50.
[51] Daan N, Richardson K. Changes in the North Sea Ecosystem and their causes: Aarhus 1875 revisited. ICES Journal of Marine Science 1996;53:875–1225.
[52] Drinkwater KF, Myers RA. Testing predictions of marine fish and shellfish landings from environmental variables. Canadian Journal of Fisheries and Aquatic Science. 1987;44:1568–73.