INDECO

DEVELOPMENT OF INDICATORS OF ENVIRONMENTAL PERFORMANCE OF THE COMMON FISHERIES POLICY

Project no. 513754

INDECO

Development of Indicators of Environmental Performance of the Common Fisheries Policy

Specific Targeted Research Project of the Sixth Research Framework Programme of the EU on 'Modernisation and sustainability of fisheries, including aquaculture-based production systems', under 'Sustainable Management of Europe's Natural Resources'

A review of modelling methods of indicators for the identification of fishing impacts on marine ecosystems

Dr. Murdoch McAllister and Dr. Pia Orr (Imperial College, London)

Project Deliverable Number 9

Dissemination Level: Public

Due date: April 2005

Submission date: February 2006

Start date of project: 1 December 2004 Duration: 24 months

Lead name and organisation: Indrani Lutchman and James Brown, Institute for European Environmental Policy (IEEP)

[Final]

The INDECO project

The purpose of this Co-ordination Action is to ensure a coherent approach to the development of indicators at EU level, in support of environmental integration within the CFP and in the context of international work on indicators. The principal objectives of INDECO are:

1. to identify quantitative indicators for the impact of fishing on the ecosystem state, functioning and dynamics, as well as indicators for socio-economic factors and for the effectiveness of different management measures;
2. to assess the applicability of such indicators; and
3. to develop operational models with a view to establishing the relationship between environmental conditions and fishing activities.

A consortium of 20 research organisations from 11 EU Member States is implementing INDECO. An Advisory User Group will provide a link between the researchers and policy makers, managers and stakeholders.

More information on INDECO can be found on the project's website:

http://www.ieep.org.uk/research/INDECO/INDECO_home.htm

This report has been carried out with the financial support of the Commission of the European Communities, under the specific RTD programme 'Specific support to policies, SSP-2004-513754 INDECO'. It does not necessarily reflect its views and in no way anticipates the Commission's future policies in this area. The information in this document is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

Table of Contents

1 INDICATORS FOR MANAGEMENT

Fishing activity can potentially impact on all components of a marine ecosystem resulting in increased mortality of both target and by-catch species in a marine community and changes to fish growth and reproductive patterns (Hall 1999, Jennings & Kaiser 1998). The removal of fish from a marine ecosystem also has significant effects on its predators, prey and competitors as well as indirect consequences for other species such as marine mammals and sea-birds.

With the current interest in the development of a broader ecosystem-based approach amongst fisheries scientists and managers, there has been an acknowledgement of the need to extend beyond a consideration of the impact of fishing at the single species level (ICES 2000, Link *et al.* 2002, Garcia *et al.* 2003). This is reflected in European fisheries management legislation, with the Common Fisheries Policy (CFP) reform process leading to 'the progressive implementation of an eco-system based approach to fisheries management' being an objective of the new 'basic' CFP Regulation (Article 2(1), Regulation 2371/2002). While ecosystem-based management (EBFM) is yet to be defined in a EU policy context, it involves a move away from the "hard predictability" and quantitative predictions of the classical fisheries management model to the "soft predictability" and qualitative predictions of indicators (Degnbol & Jarre 2004, Degnbol 2005).

Reliable Indicators of fishing impact should provide a means for assessing the relative effect of fishing on the marine ecosystem. However, the incorporation of fishing impact indicators into EBFM decision criteria is still in the developmental phase (Link *et al.* 2005). In a management situation, indicators are evaluated in relation to specific reference values, either targets or limits, which, if violated, are designed to trigger a management response (Sainsbury & Sumaila 2001). The challenge is to identify key indicators with reference points that reflect the scale of the management unit, are case-specific, cost-effective and reflect the effects of fishing.

2 INDICATORS FOR FISH COMMUNITIES

Many indicators have been proposed for the measurement of fishing impact on marine ecosystems at organisational levels ranging from populations to whole ecosystems (see FAO 1999, Rice 2000, 2003, Rochet & Trenkel 2003). The types of indictors that are available for use in an EBFM context and their associated requirements are described in a separate INDECO report "A review of the indictors for ecosystem structure and functioning" (2005).

Single stock indicators and their associated reference points have been widely used in fisheries management to define overfishing for several decades (Degnbol & Jarre 2004). While single stock indicators are not considered to be indicators of ecosystem properties, it may be appropriate to consider a suite of species or population indicators from a single study as part of a "holistic" approach to the evaluation of the effects of fishing at an ecosystem level (Caddy 2002, Dale & Beyeler 2001). However, the problems that have previously been encountered with single stock indicators also apply to ecosystem indicators, in particular the difficulty in distinguishing between anthropogenic effects (such as fishing) on indicators and other non- anthropogenic impacts (Trenkel & Rochet, 2003, Jennings & Blanchard 2004) such as environmental fluctuations which can also modify ecosystems over time (Walther *et al.* 2002).

There have been several recent reviews of indicators for the detection of fishing impacts (Vandermeulen 1998, Hall 1999, Murawski 2000, Rice 2000, Rochet & Trenkel 2003) and relatively few of these indicators have been subjected to testing (Link *et al.* 2000, Trenkel & Rochet 2003, Rochet & Trenkel 2003, Nicholson & Jennings 2004, Fulton *et al.* 2005, Piet & Jennings 2005, Rochet& Rice 2005). It is not the purpose of this review to provide an exhaustive analysis of the modelling of every indicator but rather to review those models that provide an objective evaluation of the behaviour of an indicator in response to the effects of fishing, with the intention of applying these results within a decision making framework for fisheries management. Since many studies have used trend analysis for evaluating changes exhibited by indicators with time, therefore this review will focus on time-series models of indicator trends.

2.1 Establishing a hypothesis for testing indicators

Direct comparison of indicator values with reference points is only possible in the unlikely situation that the whole system is sampled. More commonly, the system is sub-sampled and conclusions about the fish community as a whole are made within a hypothesis-testing framework which assumes a null hypothesis that the value of an indicator is equal to a specific reference point or exhibits no significant change. In this instance the alternative hypotheses are that the value of the indicator is either greater or less than a specific reference point or else shows a trend over time as a result of the impact of fishing and/or environmental changes over time (Rochet & Trenkel 2003, Trenkel & Rochet 2003). It is necessary to exercise caution in the interpretation of indicator values with respect to reference points or target values as both empirical and model-derived indicators are estimated quantities and therefore subject to inherent error and uncertainty (Punt *et al.* 2001).

2.2 Establishing a baseline as a reference point

One criticism frequently levelled at ecosystem indicators is the lack of appropriate reference points (Degnbol & Jarre 2004). Ideally, knowledge of the structure of the unexploited system is required in order to establish a baseline as a reference point which can then be compared to indicators which reflect subsequent levels of fishing impacts (Steele & Schumacher 2000, Jennings & Blanchard 2004, ICES 2001). However, fisheries exploitation usually starts long before the advent of any scientific programs designed to assess the impact of fishing on an ecosystem, making it difficult to establish reference points which reflect a system in a "pre-exploited" state. Furthermore, many existing community indicators of size are only sensitive to fishing impacts during the early phase of exploitation (Jennings & Kaiser 1998, Hall 1999). After the onset of exploitation, short-term indicator time series can be subject to "shifting baseline syndrome" (Pauly 1995) whereby indicators reflect a system which is subject to increasing levels of exploitation over time.

The recent development by Jennings & Blanchard (2004) of a size-based model based on macroecological theory for predicting total fish abundance and size structure in an unexploited ecosystem may prove to be a powerful new tool for setting ecosystem indicator reference levels and assessing the effect of fishing on community structure. Another approach proposed by Link *et al.* (2002) allows for the identification of reference limits through the use of principal components analysis (PCA) to track changes in system state. Both of these methods require further development and testing but show considerable promise. Presently, in nearly all situations, the status of the system in the absence of fishing remains unknown (Trenkel & Rochet, 2003).

2.3 Reference Directions for Indicators

The definition of appropriate reference points or limits is a major obstacle in the translation of indicators into decision criteria (Link 2005). The majority of indicators that can be assessed with reference points operate at a single species level and therefore provide little insight into processes and operations at an ecosystem level (Hall & Mainprize 2004). While there is little consensus on reference points, a range of studies have shown that the direction of trends in indicator time series can reflect the effect of fishing (Table 1). Considering the difficulties in establishing baseline reference levels and the lack of agreed reference points for the majority of indicators, it would appear that the only viable alternative for assessing the relative impact of fishing is by the examination of consistent changes or trends in indicators over time (Rochet & Trenkel 2003). Reference directions (RDs) may provide useful management targets as management response will be triggered by counter trends in RDs. The use of RDs within a management framework is described by Link *et al.* (2002) and Bellail *et al.* (2003).

A review of recent studies shows that reference directions are commonly used for assessing indicators of size, diversity and trophodynamics (Table 1). Therefore, these indicators will be of more immediate use to management than indicators without reference points. The following section describes the derivation and use of indicators from size, diversity and trophodynamic models.

2.4 Size-based models

Fishing is a size-selective process which generally targets larger fish and exploitation results in an increase of smaller-sized organisms in an ecosystem (Jennings & Kaiser 1998, Hall 1999, Jennings *et al.* 1999, Nicholson & Jennings 2004). Individuals and species which generally occupy the top trophic levels in a fish community are usually the first to be targeted by exploitation (Ault *et al.* 1998, Gislason & Rice 1998, Pauly *et al.* 1998). Given that mean body size is strongly correlated with trophic level in the community (Jennings *et al.* 2001), and exploitation decreases the biomass of large-bodied fish, it is reasonable to expect that a size-based indicator will reflect the impact of fishing. Shin *et al.* (2005) present a detailed analysis of the theoretical basis for the use of size-based indicators for the detection of fishing impacts.

The collection of size data is a matter of routine for most surveys and commercial fisheries and the process of gathering together a time series of size-related information is usually simple and cost-effective. This information can be combined in different ways to produce size-based indicators for processes at different hierarchical levels within an ecosystem. In addition, size indicators (e.g. independent estimates of mean community length and mean community weight) often demonstrate similar patterns of residual variance making them good candidates for MANCOVA analysis (see section 4.5).

2.4.1 Sized based indicators at the population level

At the population level, changes in mean length, mean weight, mean maximum length and mean maximum weight all reflect the removal of larger fish (Trenkel & Rochet 2003, Ault *et al.* 2005). Mean population length has been shown to be unbiased with an assumption of constant recruitment to the exploited stock (Ehrhardt & Ault 1992, Quinn & Deriso 1999) and in some cases is also relatively insensitive to trends in recruitment (Ault *et al.* 2005). Trenkel and Rochet (2003) found that this indicator was very precise and had a high level of statistical power. Rochet (1998) found that population mean length-at-maturity of an exploited population increased with time, a finding consistent with the theoretical response of exploited stocks to density-dependent changes in growth and maturation. However, other studies have shown that the performance of this indicator is inconsistent (see Rochet 1998).

2.4.2 Size based indicators at the community level

Community indicators of mean size and weight provide a simple, robust and easily understood indicator of the impact of fishing on marine ecosystems (Link & Brodziak 2002, Link *et al.* 2002, Rochet & Trenkel 2003, Dulvy *et al.* 2004, Laë *et al.* 2004, Nicholson & Jennings 2004, Sala *et al.* 2004, Blanchard *et al.* 2005, Piet & Jennings 2005). Blanchard *et al.* (2005) found that temporal trends in maximum length and average weight of fish communities respond to the effects of fishing in variable environments therefore these size-based indicators are useful for detecting fishing impacts against a background of climatic variation. Community mean length-at-maturity is not considered to be any more informative than maximum community length (Shin *et al.* 2005).

Mean community size ranks highly in terms of acceptability by stakeholders (Degnbol & Jarre 2004). It is usually a relatively simple procedure to calculate the mean length of catch and provided that mean and median length are similar, a normal distribution can be assumed (Punt *et al.* 2001, Trenkel & Rochet 2003). Link (2005) proposes an annual change in mean size < 30% as a precautionary threshold with a limit reference point of a 50% decline, although most studies use RDs to detect the impact of fishing.

2.4.3 Fish size spectra models

Fish size spectra is a measure of the distribution of biomass across size classes in a fish community. As fishing selectively removes larger individuals the predation pressure on smaller fish is indirectly reduced and the slope of the fish size spectrum increases (Murawski & Idoine 1992, Greenstreet & Hall 1996, Rice & Gislason 1996, Bianchi et al. 2000, Benoît & Rochet, 2003, Nicholson & Jennings 2004). Link (2005) proposes a limit reference point of a slope increase of 10% per year.

Bianchi et al. (2000) found that the slope of the size spectra responded consistently to exploitation in high latitude regions while the decreased sensitivity of the slope to changes in fishing in tropical waters was attributed to the higher growth rate of tropical species. In simulations, exploitation affected the curvature and stability of the size spectrum rather than the slope causing Benoît & Rochet (2004) to conclude that the slope of the size spectra is not a consistent indicator of fishing impacts. The usefulness of size spectra models for management is limited as currently there is insufficient empirical and theoretical information available to allow meaningful interpretation of slopes and intercepts (Bianchi et al. 2000).

2.5 Sources of potential bias in size-based indicators

The samples from research surveys and commercial catches used to produce the data for estimating size-based indicators of fish communities are typically both size and species-selective. Gear selectivity can result in the introduction of systematic bias in the size distribution of catches. Inconsistencies in survey design, such as gradual changes to the sampling methodology over the duration of the research period, can cause interannual variability in the indicator which will confound the detection of temporal trends (Piet & Jennings 2005). Consistent spatial coverage of the sampling area is essential when collecting size data (Shin *et al.* 2005).

2.6 Performance of size-based indicators

Nicholson & Jennings (2004) tested a range of size-based indicators and found that the power of a trawl survey to detect trends over 10 years or less was too low for the provision of effective provision of management advice. Trenkel & Rochet (2003) tested many candidate indicators and found that the best performer over a short term in terms of precision and power was mean catch length.

Trends in size-based indicators are not exclusively due to the effect of fishing. Environmental changes and fluctuations in recruitment will also influence trends (Jennings & Dulvy 2005, therefore researchers must attempt to account for as many sources of variation as possible by considering complementary information from size-based and other indicators (Shin *et al.* 2005).

2.7 Diversity indices -Species richness, Species diversity & Species evenness

Fishing theoretically reduces diversity by selectively removing species (Jennings & Kaiser 1998, Gislason et al 2000) and over-long time periods various diversity indices have been shown to decrease in response to exploitation (Greenstreet & Hall 1996, Rijnsdorp *et al.* 1996, Greenstreet *et al.* 1999, Hall 1999). However, in other situations, diversity indices have increased in response to increasing evenness following the removal of more abundant stocks (Rice 2000). Other factors unrelated to the effect of fishing can also affect diversity indices leading to the conclusion that diversity indicators respond inconsistently to the effects of fishing.

2.8 Mean Trophic Level (TLm) and Fishing in Balance (FiB)

The mean trophic level (*TLm*) has been proposed as an indicator of the effect of fishing on food webs and the theoretical basis for the reference direction is undisputed (Pauly *et al.* 1998). Trends in the mean trophic level of catches are not exclusive to fishing impacts but are biased by economic and technological factors (Caddy *et al.* 1998, Caddy & Garibaldi 2000). Trends in the mean trophic level at the community level should theoretically provide a better indicator of fishing impacts (Rochet & Trenkel 2003). The performance of this indicator in a range of studies was inconsistent. For example, Jennings *et al.* (2002) found no consistent detectable trend in a long time series of the trophic level of the North Sea fish community whereas Milessi *et al.* (2005) accounted for economic and technological factors and found that mean trophic levels declined with increasing exploitation.

The Fishing in Balance *(FiB)* indicator is a measure of "trophic level balance" which is similar to mean trophic level. FiB incorporates measures of mean trophic level and assumes a known value for transfer efficiency which is constant across trophic levels (Pauly *et al.* 2000). The *FiB* indicator is suitable for comparison between different ecosystems but is based on strong data assumptions and a complex modelling approach that includes a high level of uncertainty. This indicator scored poorly among stakeholders (Degnbol & Jarre 2004). In addition, both *TLm* and *FiB* have high data requirements which restricts their application (Rochet & Trenkel 2003).

2.9 Mass Balance Models: Ecopath & Ecosim

Mass balance models are a well established group of ecosystem models (Rice 2000). The Ecopath suite of software (Walters *et al.* 1997, Christensen & Walters 2000, Pauly *et al.* 2000) provides an aggregate system-modelling package for the evaluation of the impact of fishing on the trophic flow in ecosystems. The Ecopath approach was first developed in the 1980s (Polovina & Own 1983, Polovina 1984,1985) and expanded by various researchers to include temporal and spatial ecological analysis (Walters *et al.* 1997, 1999, 2000) and policy optimisations (Walters *et al.* 2002). The Ecopath with Ecosim (EwE) software package produces indicators such as the level of primary production required to support the fishery, mean trophic level and the transfer efficiency between trophic levels that can be used to assess the ecosystem effects of fishing (Bundy 2001, Buchary *et al.* 2003, Christensen & Walters 2003, Rice 2003). In some cases, these trophic macrodescriptors have been shown to be more sensitive and therefore better indicators of fishing impacts than structural descriptors such as total fish biomass or diversity indices (Arias-Gonzalez *et al.* 2004).

As with all multi-species models, the application of EwE is limited by the quality and quantity of the available data (Aydin 2004). Plagányi & Butterworth (2004) advise that EwE provides a valuable complement to traditional single-species models but in its current form EwE does not adequately address uncertainties in either the data or model structure therefore it should be used with caution. These authors also advise caution when applying EwE to assess marine mammals and seabirds. The EwE model was specifically developed for teleost fish therefore applications that include marine mammals, seabirds and some shark species need to account for the different life history parameters of these species. The EwE model was used by Blanchard *et al.* (2002) to model marine mammal and fishery interactions in the Barents Sea who found the model was extremely sensitive to the vulnerability parameter, a common aspect of many EwE studies and widely recognised (Christensen & Walters 2004). The role of external forcing cannot be readily accommodated by mass balance models making it difficult to ascertain to what degree an indicator is responding to the impact of fishing in the presence of environmental fluctuations (Rice 2000)

3 INDICATOR TRENDS FOR OTHER POPULATIONS

3.1 Seabirds

Trends in population abundance, productivity, survival and other responses of even the most common seabird species may be useful indicators of stability and health in marine ecosystems (Cairns 1987, Batty 1989, Aebischer *et al.* 1990, Ainley 1994, Le Corre & Jaquemet 2005). The high variability in seabird populations generally makes it difficult to detect the effect of fishing on abundance (Anderson & Gress 1980, Diamond & Devlin 2003). However, in a study of Alaskan seabird populations Hatch (2003) showed that a relatively short time series of 5 years of survey data could detect declines of 20% per year with 90% power.

3.2 Marine mammals

The estimation of trends in population abundance is a standard method for monitoring the impact of fishing on marine mammal populations. However, abundance is an unreliable indicator of fishing effects. The detection of changes in long-lived slow-growing species, such as seals, can be especially difficult as long time series are required to detect trends. Forcada (2000) found that a population survey of endangered monk seals had insufficiently low power to detect changes in abundance over a time period shorter than 8 years.

3.3 Benthos

Fishing can cause physical damage to the benthos resulting in habitat modification (Auster *et al.* 1996), changes to the sedimentation pattern (Churchill 1989), nutrient cycling (Mayer *et al.* 1991) and the attraction of predators and scavengers to exposed and damaged organisms (Kaiser & Spencer 1994). Fishing impacts can be monitored by trends in indicator species and rates of recovery of benthic communities following disturbance by fishing (Collie *et al.* 2000, Ellis *et al.* 2000, Kaiser 2003).

Table 1 The number of recently published studies that analysed the indicators listed for time series trends.

Indicator	***Reference Direction (RD)***	***Number of studies which reported that the indicator time series DID reflect the RD***	***Number of studies which reported that the indicator time series DID NOT reflect the RD***	***Number of studies with inconclusive results***	***Study***
Mean length in catch	↓	2			Trenkel & Rochet (2003); Laë *et al.* (2004)
Mean length in population	↓	2			Trenkel & Rochet (2003); Ault *et al.* (2005)
Mean length in community	↓	3	1		Trenkel & Rochet (2003); Dulvy *et al.* (2004); Laë *et al.* (2004); Nicholson & Jennings (2004)
Mean length of individual species	↓	1			Laë *et al.* (2004)
Mean maximum length in community	↓	4			Jennings *et al.* (1999); Nicholson & Jennings (2004); Blanchard *et al.* (2005); Piet & Jennings (2005)
Mean maximum length in catch	↓	1			Laë *et al.* (2004)
Mean weight in community	↓	6	1		Jennings *et al.* (2002); Trenkel & Rochet (2003); Nicholson & Jennings (2004); Blanchard *et al.* (2005): Jennings & Dulvy (2005); Mueter & Megrey (2005); Piet & Jennings (2005)
Mean maximum weight in community	↓	2			Jennings *et al.* (2002); Nicholson & Jennings (2004)
Mean age at maturity community	↓	1			Jennings *et al.* (1999)
Mean length at maturity community	↑	1			Jennings *et al.* (1999)

Indicator	*Reference Direction (RD)*	*Number of studies which reported that the indicator time series DID reflect the RD*	*Number of studies which reported that the indicator time series DID NOT reflect the RD*	*Number of studies with inconclusive results*	*Study*
Mean length at maturity population	↑	1			Rochet (1998)
Slope/intercept of size spectra	↑↓	10	3		Gislason & Rice (1998); Bianchi *et al.* (2000); Jennings *et al.* (2002); Benoît & Rochet (2003,2004); Trenkel & Rochet (2003); Nicholson & Jennings (2004); Blanchard *et al.* (2005); Jennings & Blanchard (2005); Jennings& Dulvy (2005); Stobberup *et al.* (2005); Piet & Jennings (2005);
Species richness	↓	1	2		Bianchi *et al.* (2000); Amand *et al.* (2004); Laë *et al.* (2004)
Species diversity	↓	1	2	1	Greenstreet & Hall (1996); Thrush *et al.* (1998); Bianchi *et al.* (2000); Laë *et al.* (2004)
Species evenness	↑		1		Laë *et al.* (2004)
Mean trophic level	↓	3	2	1	Jennings *et al.* (2002); Nicholson & Jennings (2004); Cury *et al.* (2005); Gascuel *et al.* (2005); Milessi *et al.* (2005); Piet & Jennings (2005)
Ecopath/Ecosim	↑↓	2		4	Christensen (1998); Shannon *et al.* (2000); Steven *et al.* (2000); Shannon & Cury (2003); Moloney *et al.* (2005); Pinnegar & Polunin (2004)
Fishing in Balance (FIB)		2			Cury *et al.* (2005); Milessi *et al.* (2005)

4 MODELLING TRENDS IN INDICATORS

4.1 Trend Analysis

When reference points, targets or limits are either unavailable or difficult to determine, reference directions such as the trends exhibited by an indicator over time can be used to examine the effect of fishing on ecosystems.

The analysis of temporal trends in indicators has been included in many studies (e.g. Pauly 1998, Philippart 1998, Jennings et al. 1999, Jennings et al. 2002, Link et al. 2002, Hall & Mainprize 2004, Blanchard et al. 2005) but few have considered an analysis of statistical power (Hayes & Steidl 1987, Peterman 1990). Statistical power analysis has been used in studies of time series trends using indicators such as catch-per-unit-effort (de la Mare 1984, Mueter & Megrey 2005), fish abundance (Peterman & Bradford 1987, Jennings et al. 1999, Ham & Pearsons 2000, Maxwell & Jennings 2005), fish size (Nicholson & Jennings 2004, Jennings & Dulvy 2005), marine mammal abundance (Gerrodette 1987, Holt et al. 1987, Forcada 2000, Forney 2000, Thompson et al. 2000), seabird abundance (Hatch 2003) and benthic condition (Jackson et al. 2000). For the purpose of this review, statistical power is the probability that a certain size of effect (e.g. the impact of fishing or slope in a linear time trend in an indicator) will be detected in a statistical hypothesis test if the effect actually exists.

Obviously, indicator studies that are designed with low power will not be able to provide appropriate information for management purposes. Low power studies run the risk of being unable to detect statistically significant effects or trends even when these effects are actually substantial. If the power of a monitoring program is low then researchers may continue to monitor indicator trends, oblivious to the fact that any real changes in indicator status will remain undetected. Peterman (1990) goes as far as to state that low power tests are worse than no tests at all.

4.2 The concept of statistical power

Studies of indicator trends are generally concerned with the direction and magnitude of the trend. The null hypothesis H_O is usually expressed as “there is no effect of fishing on an indicator trend” with an alternative hypothesis H_A of the form “the indicator trend is significantly affected by fishing”. Formal statistical tests are structured to decide whether or not to reject H_O and it is necessary to emphasize that the null hypothesis can never be proved rather it can only be disproved at a given level of significance (Popper 1962). There are four possible outcomes of statistical hypothesis tests and two types of errors (Table 2). A Type I error is made when a decision is made to reject a null hypothesis that is actually true. A Type II error is made when a decision is made to accept a null hypothesis that is actually false. The rate of acceptance of Type I errors (α) is conventionally set at 0.05 for most ecological studies. However, α has been set at values higher than 0.05, e.g., at 0.10, in environmental studies where the costs of a Type II error have been understood to be very high. Most studies involving hypothesis testing report an α level for the rejection of H_O but very few apply the concept of statistical power and extend the analysis to the determination of 1- β where β is the probability of a type II error, i.e., failing to reject H_O when H_O is actually false (Table 2). The power of a test is its ability to avoid Type II errors and is defined as the probability of correctly rejecting a null hypothesis that is false (Peterman 1990).

Table 2 The four possible outcomes for the statistical test of a null hypotheses given the actual state of nature. Probabilities of each outcome are shown in parentheses (adapted from Peterman 1990).

Actuality	**Decision**	
	Do not reject H_O	Reject H_O
H_O is true	Correct (1-α)	Type I error (α)
H_O is false	Type II error (β)	Correct (1- β) = Power

It is important to recognise that a failure to reject H_O is not equivalent to a conclusion that H_O is true. An analysis which does not reject H_O can only lead to a conclusion that H_O is true when the probability of a type II error (β) is sufficiently low. The term "sufficiently low" is generally assumed to refer to a β-value at least <0.2 which gives a power >0.8 (Peterman 1990). [Note that power analysis applies only to those situations where testing results in a failure to reject H_O. If testing indicates a rejection of H_O at a given level of α then the calculation of power is not warranted]. Type II errors are common in cases where large sampling variability or small sampling size prevent the detection of real effects that lead to the failure to reject H_O when H_O is actually false.

Nearly 15 years after Peterman (1990) reported that the majority of studies in fisheries aquatic sciences failed to consider the statistical power of hypothesis tests when null results were obtained, the importance of reporting statistical power is still largely ignored. This review surveyed 36 studies which analysed indicator trends for fishing impacts. Of those that did not reject a null hypothesis, only 17% reported the statistical power of the test and yet 93% of the studies with failed rejections of null hypotheses and either low power or no power calculated formulated conclusions and management recommendations on the basis that H_O was true.

Peterman (1990) identified four interrelated factors which affect statistical power.

1. The probability of a type I error (α) and the level of the P-value below which H_O is rejected. Power and α are directly related, thus an increase in α from the conventional 0.05 to 0.10 results in a substantial increase in statistical power and may be considered when the costs of type II errors are large relative to the cost of type I errors (Peterman 1990).
2. The size of the true effect being tested. The greater the hypothesized effect size, the higher the power and the higher the probability of obtaining a statistically significant outcome (Ortiz 2002).
3. The sample size (*N*). Power can always be improved by increasing the sample size (Cohen 1988)
4. The sample variance (s^2). Increasing precision of measurement and decreasing variability in the sampled parameter will increase power.

4.3 A Priori Analysis

A priori power analysis is ideally suited for determining whether a proposed monitoring scheme is likely to detect postulated effects (Steidl *et al.* 1997). A priori analysis conducted before the start of an experiment or management program can be used to

1. To calculate the sample size required to produce a desired level of power for an assumed effect size or,
2. To calculate the detectable effect size necessary to produce a given level of power.

Monte Carlo simulation can also be used to explore how a wide range of assumptions regarding the sample size, expected effect size and sampling variability can affect statistical power (De la Mare 1984, Scandol & Forrest 2001).

4.4 A Posteriori Analysis

When statistical testing fails to reject the null hypothesis, *a posteriori* analysis is useful for estimating the effect size or sample size that would have been necessary for a study to achieve an acceptable level of power. The majority of studies of indicator time series that incorporate *a posteriori* analyses of power have concluded that statistical power was insufficient to detect declines in indicator trends within a realistic management time frame (De la Mare 1984, Nicholson & Fryer 1992, Fryer & Nicholson 1993, Ham & Pearsons 2000, Punt *et al.* 2001, Hatch 2003, Maxwell & Jennings 2005, Piet & Jennings 2005).

Few resources are available for dedicated studies of indicator trends and most researchers use data from international research surveys instead (Maxwell & Jennings 2005). In general, these large-scale surveys do not include the design requirements of indicator trend studies as a major priority thus the power of many indicator studies to date has been low. It is essential that researchers include a retrospective analysis of power of indicator trends before making recommendations to management (Peterman 1990).

Retrospective power analysis of an indicator time series is a relatively straightforward procedure. A comprehensive review of statistical analysis software is provided by Thomas & Krebs (1997). The first step is to fit a model (e.g. linear, quadratic, exponential) to the time series data that best explains the systematic variation in indicator annual mean values. One procedure would be to fit a lower-order polynomial or robustly locally weighted smoothed curve (see Fryer & Nicholson 1993). An examination of the residuals will assist in deciding the model that provides the best description of the data. There are various methods for describing the variance structure of the model. For a standard linear model the residual variance is equal to the variance of the deviations between dependent observations and the linear model (adjusted for two estimated parameters), given that the model is correct (Trenkel & Rochet 2003).

Power can be increased by increasing the number of samples per year or by extending the length of the indicator time series. Obviously it is unreasonable to expect the detection of a trend in an indicator series over a short time period (e.g. < 3 years) however it is equally unrealistic to extend a time series to 20 or even 30 years in order to achieve a desired level of power. An alternative approach to increasing the power of an indicator trend study is to use MANCOVA (Multivariate Analysis of Covariance) which is available through software packages such as S-Plus and SAS. Many types of indicators (e.g. mean length and mean weight) are based on similar community characteristics and thus exhibit high levels of correlation. MANCOVA provides a simple and inexpensive way of improving power simply by combined analysis of the covariance of a set of indicators which are not necessarily independent (e.g., all obtained from the same survey) and can be expected to exhibit similar trends. For example, the annual mean weight of a set of fish species captured in a survey could be expected to be positively correlated with, and is not independent of, the annual mean length of the same set of fish species in the same survey. MANOVA and MANCOVA are specially formulated to analyse the variance and covariance of sets of inter-related dependent variables. So instead of this being a univariate analysis of variance where there is one dependent variable, there is more than one dependent variable and these are related to a set of factors or covariates that might help to explain the variation in the dependent variables. It is of benefit to evaluate the set of inter-related dependent variables in a single statistical analysis rather than one statistical analysis per dependent variable because the structure of the error variance is more accurately estimated and taken into account in a MANOVA and MANCOVA when there is covariance among dependent variables in the error deviations from model predictions.

4.5 Consequences of Type I and Type II errors

By selecting a value for α, the researcher is also making an (often unrecognized) assumption regarding the relative importance of Type I and Type II errors. If α is decreased then β will increase with the result that power may be reduced to an unacceptably low level. A more conservative management approach would result from increasing α (e.g. from the conventional 0.05 to 0.10) (Peterman & M'Gonigle 1992). The concept behind this approach is to allow quicker implementation of a management response in the event that adverse trends are identified (Hatch 2003). The costs associated with Type II errors in environmental studies are typically very high (Peterman 1990) and the risks and consequences of Type II errors may far outweigh those associated with Type I errors (Toft & Shea 1983).

4.6 Reversing the burden of proof

In hypothesis testing of indicator trends, the burden of proof is typically to demonstrate that *fishing is having an effect* by rejection of a null hypothesis of no effect of fishing. Inherent in this approach is the willingness to accept the consequences of Type II errors over those of Type I errors (Steidl *et al.* 1997). Instead, it may be appropriate to shift the burden of proof and require proof that *fishing is not having an effect* by a <u>failure</u> to reject the null hypothesis that fishing is not having an adverse effect with acceptably high power (Peterman 1990, Mangel 1993, Taylor *et al.* 2000, Sanderson & Petersen 2001, Hatch 2003). Previous studies have shown that the concept of reversing the burden of proof is justified in the case of environmental management due to the overwhelming costs of Type II errors and a track record that shows that a history of failure to reject a null hypothesis of no effect of fishing has ultimately led to the collapse of many pelagic stocks (Saetersdal 1980). A reversal of the burden of proof is a shift towards a more precautionary management strategy (Peterman & M'Gonigle 1992).

The reversal of the burden of proof is a major departure from current practice but is a standard method for testing pesticides, food and drugs (Peterman 1990). Another precautionary approach is to set the null hypothesis equal to some detectable effect size (which may be a target or reference value e.g. H_O=2) rather than to set a null hypothesis of no effect (e.g. H_O=0). This would result in the rejection of the null hypothesis only in the event that the direction of the indicator trend was towards an absence of adverse effects of fishing.

4.7 Reducing model uncertainty

Trends in abundance indices, catch-per-unit-effort, size-related indicators and diversity indicators have all been used to detect the impact of fishing. However the detection and interpretation of these trends is often confounded by environmental variability. This variability causes a decrease in statistical power and introduces uncertainty as to whether the observed trends are due to true changes in the indicator or to environmental fluctuations.

One possible approach to account for environmental variability is to model abundance as a function of a suite of environmental variables using generalised additive models (GAMs) (Hastie & Tibshirani 1990) or generalised linear models (GLMs) (McCullagh & Nelder 1989). GAMs and GLMs share many common statistical properties and both can use linear, logarithmic or polynomial models to describe relationships between predictor and response variables. However, unlike GLMs, GAMs can also include non-linear relationships to describe patterns in the data and therefore they are of particular interest to researchers in ecology (Cury *et al.* 1995, Swartzman 1995). Statistical optimisation criteria such as maximum likelihood or model choice criteria such as Akaike's information criterion (AIC) are useful tools in the determination of the appropriate number of effects for inclusion in the model.

While it is unlikely that GAMs and GLMs will capture the complexities of all environmental variables contributing to model variability, with the careful choice of appropriate environmental variables, GAMs and GLMs can potentially be used to reduce unexplained variation and improve trend analyses (Forney 2000).

GLMs are typically used in fisheries management studies for the standardisation of trends in catch per unit effort. In a similar way, GLMs could also be used to account for factors such as vessel size and catchability which particularly affect the analysis of trends in size-based indicators (Daan *et al.* 2003, Piet & Jennings 2005).

4.8 Recommendations for statistical power analysis of candidate indicators

In addition to Peterman's (1990) four factors important to statistical power analysis in applied scientific work, which are pertinent also for power analysis of candidate indicators, we offer a few additional factors also important to power calculations for candidate ecosystem indicators.

4.8.1 One versus two tailed tests to evaluate trends in indicators

The selection of a two-tailed test when a one-tailed test is more appropriate lowers power unnecessarily. We argue that in most instances where indicators may be utilized in offering management advice, a one tailed test is appropriate. This is because most candidate indicators, e.g., those listed in Table 2, have a recognized reference direction. For example, if a statistically detected decline in an indicator over some pre-specified time period is to trigger a set of remedial management actions, the null hypothesis should thus be of the form that the slope is greater than or equal to zero rather than equal to zero. Obviously statistical power analyses of candidate indicators should be devised accordingly.

Trenkel and Rochet (2003) applied one tailed tests for some indicators and two tailed tests for others. The rationale for this allocation was not made explicit and in at least one instance, the test applied was inconsistent with the reference direction assumed. For example, for the intrinsic rate of increase (r) a two tailed test was applied. Yet, a one tailed test might have been more appropriate because a single reference direction (i.e., decrease in r) and a single effect size (i.e., r = -0.10) was used in the statistical power calculation.

Nicholson and Jennings (2004) applied two-tailed tests for all statistical power calculations when according to existing literature (Table 2) all of the indicators investigated have only a single reference direction and a one-tailed test would have instead been more appropriate. Based on using a two-tailed test rather than a one-tailed test their calculations on statistical power of the candidate indicators investigated might have been biased too low. For example, when attempting to replicate their power calculations for the average mean maximum weight of fish in the North Sea International Beam Trawl Survey, a two-tailed test for a 10 year decline of 40 grams per year with alpha set at 0.05 gave power of 0.62 (compared to their result of 0.70). Making it a one tailed test increased power from 0.62 to 0.77. Power was also low from using a value for alpha that was perhaps too low. Increasing alpha to 0.10 and using a one-tailed test increased power to 0.87. We thus suggest that when there is a single reference direction for an indicator that a one-tailed hypothesis test rather than a two-tailed hypothesis test should be carried out.

4.8.2 Uncertainty in the error variance used in power calculations

The error variance required as an input to power calculations for environmental hypothesis testing is often imprecisely estimated due to low sample size or high variability in sample deviations and considerable uncertainty often exists over an appropriate model with which to estimate error variance (Pennington 1985). For example, if it is of interest to evaluate the power of a test whether a linear decline in some indicator has occurred, the variance in the deviates of such a model must be estimated from historic observations. If the time series of historic observations do not conform to a linear decline, then some other non-linear model might be fitted to the data to estimate the error deviates that might occur from a linear decline should one occur.

Moreover, it has recently emerged that different quality control and screening protocols for survey data can lead to extraordinarily different patterns in time series for the same indicators. For example, Nicholson and Jennings (2004) and Piet and Jennings (2005) produce time series of the same indicators from the IBT survey database. While showing similar trends, the interannual variability in the time series over the same time period is considerably higher in Piet and Jennings (2005) than that in the same time series from Nicholson and Jennings (2004). Clearly, the statistical power for the same indicator series to detect the same trend would be considerably lower based upon Piet and Jennings' (2005) time series. This highlights the urgent need for there to be carefully derived and commonly agreed standardized protocols for screening and quality control of survey databases from which indicators are to be derived.

If a nonlinear model is to be fitted to a historic time series to estimate error variances, then some appropriate nonlinear model must be chosen and there may be considerable uncertainty in the form of an appropriate model. Statistical techniques for dealing with model choice would thus need to be considered, e.g., likelihood ratio or F-tests for nested models, AIC, or Bayes' factors (Hiborn and Mangel 1997). Nicholson and Jennings (2004) utilized a simple difference-based variance estimator to estimate the error variance (Gasser *et al.*. 1986) for power calculations that has been found to provide a good balance between bias and precision when time series are relatively short. Trenkel and Rochett (2003) in contrast used among other things residual variance from fits of models to indicator time series values. However, the use of a single estimate of error variance in power calculations ignores uncertainty in the estimate of the error variance and masks the sensitivity of statistical power results to alternative plausible values for error variance. We therefore recommend that in future power analyses of indicators that the sensitivity of power to alternative methods of computing input variances be shown.

4.8.3 Statistical analysis of sets of non-independent co-varying indicators

Sets of indicators derived from the same survey may be expected to exhibit the same reference direction, and similar patterns in error variability if they happen to be analogous measures of some common ecological property such as the mean body size of fishes in the ecosystem. For example, some proposed indicators include mean length, mean weight, mean maximum length and mean maximum length and as these are all related to body size they might all expected to covary. Furthermore, as pointed out by Trenkel and Rochet (2003), these indicators are not statistically independent since they come from the same set of trawl survey hauls from which the indicators are derived. Thus, while it is valid to conduct hypothesis tests (providing that multiple tests correction factors are applied) and compute the statistical power separately for each separate indicator, it may be equally valid to include all related indicators from a single survey in a single statistical evaluation such as MANOVA or MANCOVA that is specifically designed to statistically analyse sets of dependent variables that are not statistically independent (see above). MANOVA and MANCOVA typically yield higher power tests than their univariate counterparts when there is covariance in the error structure of the dependent variables. However, the null hypotheses that may be tested using MANCOVA in conventional software may not necessarily be the ones of interest. For example, SPSS permits only the null hypothesis that the slope for the covariate is not different between the set of dependent variables. In contrast, it may be instead of interest to test the null hypothesis that the mean value for the slope for the covariate across all of the indicators in the MANCOVA is greater than or equal to zero. We thus recommend that such multivariate methods of analysis be considered for statistically analysing sets of indicators that are non-independent and have been found to exhibit positive covariance.

4.8.4 Formulating effect sizes of interest for power calculations for indicators

Effect size is a key input to statistical power analysis yet there appears to be a lack of consistency in the manner in which input values for effect size are derived for statistical power calculations. The effect size for statistical power calculations of candidate indicators should reflect the effect size, e.g., decline, in the indicator that is of importance in a management decision making context (Peterman 1990). In some instances, power is calculated over a range of values for potential effect sizes (Trenkel and Rochet 2003). In this latter paper, however, criteria were not specified for the choice of the particular ranges of effects sizes utilized. For example, a value of $r=-0.1$ was utilized in Trenkel and Rochet (2003) for power calculations but no justification for why this effect size was of concern was provided.

In other instances, the effect size used for power calculations is taken to be estimates from statistical analysis of historic data (e.g., also in Trenkel and Rochet 2003). Nicholson and Jennings (2004) fitted Loess smoothers to the historic candidate indicators and estimated the maximum observed historic change in the indicators based on the Loess smoothed trend. This maximum observed historic change was taken to be the maximum effect size for the statistical power analysis of the ability to detect future changes in the candidate indicator.

Without also identifying where effort intensification occurred or the periods in which fishing effort was considered to be most excessive, it may be the case that maximum observed effect sizes in available historic data may come from periods when there have been no changes of concern to the ecosystem and thus the observed effect sizes would be much smaller than those that could be expected to occur based on undesirable fishing induced changes to the ecosystem. Thus, rather than utilizing only observed historic values for possible effect sizes of concern for management decision making, it may be more prudent to link these analyses with evaluations of historic fishing effort as in Piet and Jennings (2005) and try to identify minimum effect sizes that would in principle be of concern based on population dynamics theory or ecological theory and possibly population dynamics and ecosystem modelling (Fulton *et al.*. 2005)

If it were known that large increases in fishing effort occurred at some point in the time series, this could help to identify possible expected undesirable changes in the candidate indicators based on observed changes in the indicators following the fishing effort intensification (Piet and Jennings 2005). Piet and Jennings (2005) analysis that attempted to identify responses in candidate indicators to fishing effort variation, appeared to use indicator time series with anomalously high variance compared to the same ones in Nicholson and Jennings (2004) and possibly because of this may have had unduly low power in their tests for such linkages.

4.8.5 Are linear models always the most appropriate choice for trend analysis?

When it is of interest to do statistical power calculations for the detection of time trends in an indicator, it may be more appropriate to consider an exponential model rather than a simple linear model. Many evaluations of abundance and ecosystem indicators presume a simple linear model and base statistical power calculations on the linear model (de la Mare 1984; Nicholson and Jennings 2004). Such power analyses also evaluate the sensitivity of power to the length of a linear trend and in some instances attempt to the number of years it would take to obtain sufficiently high power. The use of a linear model in such instances may be inappropriate because it ignores the fact that there may be practical limits to the extent of decline possible in the characteristic measured by the indicator (e.g., the mean body size of adult fish). A linear model may also be conceptually inappropriate since it implies that should the assumed trend continue for long enough, then the value for the indicator could take on a negative value. Exponential decline models may provide an appropriate alternative because they model a constant proportional decline per year and do not allow the modelled variable to drop below zero. The "effects" to evaluate in exponential models are also easier to interpret than the annual total rates of decline in linear models since in exponential models it is possible to evaluate given percentage annual rates of decline (Link 2005).

4.8.6 Utilizing indicators directly vs. fitted time series model values for advice

It has been suggested that for management purposes it may be of interest to detect changes in the values of indicators over relatively short time periods, e.g., 3-5 years. Nicholson and Jennings (2004) suggested that when a linear model is assumed, the statistical power to detect changes over such short time periods was unacceptably low and that most indicators available should not be used to try to detect changes in indicators over such short time periods. However, if it is still critical to be able to respond to apparent declines in indicators in the absence of other information on ecosystem state, it may be acceptable to apply time series model-based estimated declines in indicators to provide management advice (Pennington 1985; 1986). Management advice could thus be based on declines in the model predicted values for the indicator where some time series model has been fitted to the available data and a statistically rigorous model selection process has been undertaken to identify the most appropriate time series model. This is analogous to stock assessment where the decision rules adopted are model-based rather than model-free (McAllister *et al.* 1999). Where there is a set of related indicators that could a priori be expected to exhibit similar patterns in temporal variability in response to fishing and other environmental perturbations, it may be appropriate to undertake the model selection process with a single form of a time series model being chosen for the set of related indicators, rather than a separate form of model being selected for each of the related indicators. Decision rules could be formulated on joint observed changes in the modelled indicators, taking into account the reference direction for each particular indicator (Fulton *et al.* 2005).

5 CONCLUSIONS

Many studies have shown that size-based metrics are generally consistent indicators for the detection of fishing impacts on marine communities. The paucity of defined reference points for these indicators has resulted in a situation whereby the only viable approach for assessing the relative impact of fishing on an indicator is by the monitoring of trends in indicators over a period of time.

The power of many indicator monitoring programs has been shown by a number of published studies to be too low for effective management, however there are a number of alternatives available for improving power. Researchers within the INDECO project and beyond should consider

1. The relative costs of Type I and Type II errors and set a level of α accordingly and not accepting with out question the standard value for alpha of 0.05.
2. Using multivariate statistical methods such as MANCOVA to analyse the covariance among dependent variables in the error deviations from model predictions for a set of indicators.
3. Reversing the burden of proof.
4. Choosing 1-tailed rather than 2-tailed tests when there is an appropriate reference direction.
5. Evaluating a variety of approaches to computing error variance.
6. Applying both empirical methods that link indicators with historic fishing effort (Piet and Jennings 2005) and theoretical Population Dynamic Models and Ecological Modelling approaches (e.g., Fulton *et al.*. 2005) to identifying the effect sizes on percentage changes in the indicator that would be of interest to detect.

7. Applying agreed and carefully formulated standardized protocols with which to prepare survey data to construct candidate fishery indicators.
8. A time series longer than just 3 years in attempts to statistically detect trends in a set of indicators for use in fisheries management.
9. Using exponential models rather than linear models when evaluating the statistical power to detect annual changes in indicators.
10. Using time series modelling approaches to estimate trends in sets of analogous indicators and base decision rules on the modelled trend rather than the observed values.

These alternative approaches have the potential to make indicator trend analysis a more powerful tool for the detection of fishing impacts on marine communities.

6 LITERATURE REVIEWED

Aebischer, N.J., Coulson, J.C. & Colebrook, J.M. (1990). Parallel long-term trends across four marine trophic levels and weather. Nature 347, 753-755.

Ainley, D.G. (1980). Birds as marine organisms : A review. CalCOFI Rep., Vol. XXI, pp 48-53

Ainley, D.G., Sydeman, W.J., Hatch, S.A. & Wilson, U.W. (1994). Seabird population trends along the west coast of North America: causes and the extent of regional concordance. Studies in Avian Biology 15, 119-133.

Amand, M., Pelletier, D., Ferraris, J & Kulbicki, M. (2004). A step towards the definition of ecological indicators of the impact of fishing on the fish assemblage of the Abore reef reserve (New Caledonia). Aquatic Living Resources 17, 139-149.

Anderson, D.W. & Gress, F. (1980). Brown pelicans as anchovy stock indicators and their relationships to commercial fishing. CalCOFI Report 21, 54-61.

Arias-González, J.E., Nuňez-Lara, E. & González-Salas, C. & Galzin, R. (2004). Trophic models for investigation of fishing effect on coral reef ecosystems Ecological Modelling 172, 197-212

Ault , J.S., Bohnsack, J.A. & Meester, G.A. (1998). A retrospective (1979-1996) multispecies assessment of coral reef fish stocks in the Florida Keys. Fishery Bulletin 96(3), 395-414.

Ault, J.S., Smith, S.G. & Bohnsack (2005). Evaluation of average length as an estimator of exploitation status for the Florida coral-reef fish community. ICES Journal of Marine Science 62, 417-423.

Auster, P.J., Malatesta, R.J. Langton, R.W., Watling, L., Valentine, P.C., Donaldson, C.L.S., Langton, E.W., Shepard, A.N. & Babb, I.G. (1996). The impact of mobile fishing gear on seafloor habitats in the Gulf of Maine (Northwest Atlantic): Implications for conservation of fish populations. Reviews in Fisheries Science 4, 185-202.

Babcock, E.A., Pikitch, E.K., McAllister, M.K., Apostolaki, P.& Santora, C. (2005). A perspective on the use of spatialized indicators for ecosystem-based fishery management through spatial zoning. ICES Journal of Marine Science 62, 469-476.

Batty, L. (1989). Birds as monitors of marine environments. Biologist 136, 151-154.

Bellail, R., Bertrand, J., Le Pape,)., Mahé, J.C., Morin, J., Poulard, J.C., Rochet, M-J., Schlaich, I., Souplet, A., and Trenkel, V. (2003). A multi-species dynamic indicator-based approach to the assessment of the impact of fishing on fish communities. ICES Document, CM 2003/V:02, 12 pp.

Benoît, E. & Rochet, M-J. (2003). The meaning of fish size spectra, the effects of fishing on them and the usefulness of their slope as indicator of fishing impacts. ICES CM 2003/N:05

Benoît, E. & Rochet, M-J. (2004). A continuous model of biomass size spectra governed by predation and the effects of fishing on them. Journal of Theoretical Biology 226, 9-21.

Bianchi, G., Gislason, H., Graham., K., Hill, L., Jin, X., Korateng, K., Manichchand-Heilman, S., Payá, I., Sainsbury, K., Sanchez., F., Zwanenburg, K. (2000). Impact of fishing on size composition and diversity of denersal fish communities. ICES Journal of Marine Science 57, 558-571.

Bundy, A. (2001). Fishing on ecosystems:the interplay of fishing and predation in Newfoundland-Labrador. Can. J. Fish. Aquat. Sci 58, 1153-1167.

Caddy, J.F. (2002). Limit reference points, traffic lights, and holistic approaches to fisheries management with minimal stock assessment input. Fishery Research 56, 133-137.

Caddy, J.F. & Garibaldi, L. (2000). Apparent changes in the trophic composition of world marine harvests: the perspective from the FAO capture database. Open Coastal Management 43, 615-655.

Caddy, J.F., Csirke, J., Garcia, S.M. & Grainger, R.J.R. (1998). How pervasive is fishing down marine food webs? Science 282, 1383-1385.

Cairns, D.K. (1987). Seabirds as indicators of marine food supplies. Biological Oceanography 5, 261-271.

Christensen, V. (1998). Fishery-induced changes in a marine ecosystem: insights for models of the Gulf of Thailand. Journal of Fish Biology 53(Suppl. A), 128-142.

Christensen, V. & Walters, C.J. (2004). Ecopath with Ecosim : methods, capabilities and limitations. Ecological Modelling 172, 109-139

Churchill, J.H. (1989). The effect of commercial trawling on sediment resuspension and transport over the Middle Atlantic Bight Continental Shelf. Continental Shelf Research 9, 841-864.

Cohen, J. (1988). Statistical power analysis for the behavioural sciences. Second ed. Lawrence Erlbaum Assoc. N.J. 567 pp.

Collie, J.S., Hall, S.J., Kaiser, M.J. & Poiner, I.R. (2000). A quantitative analysis of fishing impacts on shelf-sea benthos. Journal of Animal Ecology 69, 785-798.

Crawford R.J.M. & Shelton P.A. (1978). Pelagic fish and seabird interrelationships off the coast of the Southwest and South Africa. Biol. Cons. 14, 85-109

Cury, P., Roy, R., Mendelssohn, R., Bakun, A., Husby, D.M., Parrish, R.H. (1995). Moderate is better: exploring nonlinear climatic effects on the Californian northern anchovy (*Engraulis mordax*). Canadian Special Publication of Fisheries and Aquatic Sciences 121, 417-424.

Cury, P.M., Shannon, L.J., Roux, J-P., Daskalov, G.M., Jarre, A., Moloney, C.L. & Pauly, D. (2005). Trophodynamic indicators for an ecosystem approach to fisheries. ICES Journal of Marine Science 62, 430-442.

Daan, N., Gislason, H., Pope, J. & Rice, J. (2003). Changes in the North Sea community structure: Evidence of indirect effects of fishing? ICES CM:10.

Dale, V.H. & Beyeler, S.C. (2001). Challenges in the development and use of ecological indicators. Ecological Indicators 1, 3-10.

De la Mare, W.K. (1984). On the power of catch per unit effort series to detect declines in whale stocks. Report to the International Whaling Commission 34, 655-662.

Degnbol, P. (2005). Indicators as a means of communicating knowledge. ICES Journal of Marine Science 62, 606-611.

Degnbol, P. & Jarre, A. (2004). Review of indicators in Fisheries Management – A development perspective. African Journal of Marine Science 26, 303-326.

Diamond, A.W. & Devlin, C.M .(2003) Seabirds as indicators of changes in marine ecosystems : ecological monitoring on Machias Seal Island. Environmental Monitoring and Assessment 88, 153-175.

Dulvy, N.K., Polunin, N.V.C., Mill, A.C., Graham, N.A.J. (2004). Size structural change in lightly exploited coral reef fish communities: evidence for weak indirect efeects. Canadian Journal of Fisheries and Aquatic Sciences 61, 466-475.

Ehrhardt, N.M. & Ault, J.S. (1992). Analysis of two length-based mortality models applied to bounded catch length frequencies. Transactions of the American Fisheries Society 121, 115-122.

Ellis, J.I., Norkko, A., Thrush, S.F. (2000). Broad-scale disturbance of intertidal and shallow sublittoral soft-sediment habitats; effects on the benthic macrofauna. Journal of Aquatic Ecosystem Stress and Recovery 7, 57-74.

FAO (1999). Indicators for sustainable development of marine capture fisheries. FAO Technical Guideline for Responsible Fisheries 8.

Forcada, J. (2000). Can population surveys show if the Mediterranean monk seal colony at Cap Blanc is declining in abundance? Journal of Applied Ecology 37, 171-181.

Fox, D.R. (2001). Environmental power analysis – a new perspective. Environmetrics 12, 437-449.

Fulton, E.A., Smith, A.D. & Punt, A.E. (2005). Which ecological indicators can robustly detect effects of fishing? ICES Journal of Marine Science 62, 540-551.

Garcia, S.M., Zerbi, A., Aliaume, C., Do Chi, T. & Laserre, G. (2003). The ecosystem approach to fisheries : issues, terminology, principles, institutional foundations, implementation and out-look. FAO Technical Paper 443. 71 pp

Gascuel, D., Bozec, Y-M., Chassot, E., Colomb, A., Laurans, M. (2005). The trophic spectrum: theory and application as an indicator. ICES Journal of Marine Science 62, 443-452

Gasser, T., Sroka, L., & Jennen-Steinemetz. C. (1986). Residual variance and residual pattern in non-linear regression. Biometrika 73: 625-633.

Gerrodette, T. (1987) A power analysis for detecting trends. Ecology 68, 1364-1372.

Gerrodette, T. (1991). Models for detecting trends : A reply to Link and Hatfield. Ecology 72 (5), 1889-1892.

Gislason, H. & Rice, J. (1998). Modelling the response of size and diversity spectra of fish assemblages to changes in exploitation. ICES Journal of Marine Science 55, 362-370

Greenstreet, S.P.R. & Hall S.J. (1996). Fishing and the ground-fish assemblage structure in the north-western North Sea: an analysis of long-term and spatial trends. Journal of Animal Ecology 65, 577-598.

Hall, S.J. (1999). The effect of fishing on marine ecosystems and communities. Blackwell Science, London, 274 pp.

Hall, S.J. & Mainprize, B. (2004). Towards ecosystem-based fisheries management. Fish and Fisheries 5, 1-20.

Ham, K.D. & Pearsons, T.N. (2000). Can reduced salmon population abundance be detected in time to limit management impacts? Canadian Journal of Fisheries and Aquatic Sciences 57, 17-24.

Harvey, E., Fletcher, D., Shortis, M. (2000). Improving the statistical power of length estimates of reef fish : A comparison of estimates determined visually by divers with estimates produced by a stereo-video system. Fishery Bulletin 99, 72-80.

Hastie, T.J. & Tibshirani, R.J. (1990). Generalized additive models. Monographs on Statistics and Applied Probability 43.

Hatch, S.A. (2003). Statistical power for detecting trends with applications to seabird monitoring. Biological Conservation 111, 317-329.

Hayes J.P. & Steidl R.J. (1997). Statistical power analysis and amphibian population trends. Conservation Biology 11, 273-275

Hilborn, R., and Mangel, M. (1997). The Ecological Detective. Princeton University Press, 330p.

Holt, R.S., Gerrodett, T. & Cologne, J.B. (1987). Research vessel survey design for monitoring dolphin abundance in the eastern tropical Pacific. Fisheries Bulletin U.S. 85, 435-446.

ICES (2000). Ecosystem effects of fishing. Proceedings of an ICES/SCOR Symposium held in Montpellier, France, 16-19 March 1999. ICES Journal of Marine Science 57, 465-792.

ICES (2001). Report of the Working Group on Ecosystem Effects of Fishing Activities. International Council for the Exploration of the Sea, Committee Meeting, 2001/ACME : 09.

Jackson, L.E., Kurtz, J.C. & Fisher, W.S. [eds.] (2000). Evaluation guidelines for ecological indicators. EPA/620/R-99/005. U.S. Environmental Protection Agency, Office of Research and Development, N.C., 107 pp.

Jenning, S., Greenstreet, S.P.R & Reynolds, J.D. (1999). Structural change in an exploited fish community : A consequence of differential fishing effects on species with contrasting life histories. The Journal of Animal Ecology 68(3), 617-627.

Jennings, S. & Blanchard, J.L. (2004). Fish abundance with no fishing: predictions based on macroecological theory. Journal of Animal Ecology 73, 632-642.

Jennings, S. & Kaiser, M.J. (1998). The effects of fishing on marine ecosystems. Advances in Marine Biology 34, 201-351.

Jennings, S., Pinnegar, J.K., Polunin, N.V.C. & boon, T.W. (2001). Weak cross-species relationships between body size and trophic level belie powerful size-based trophic structuring in fish communities. The Journal of Animal Ecology 70(6), 934-944.

Jennings, S., Greenstreet, S.P.R., Hill, L., Piet, G.J., Pinnegar, J.K. & Warr, K.J. (2002). Long-term trends in the trophic structure of the North Sea fish community : evidence from stable-isotope analysis, size-spectra and community metrics. Marine Biology 141, 1085-1097.

Jennings, S.J. & Dulvy, N.K. (2005). Reference points and reference directions for size-based indicators of community structure. ICES Journal of Marine Science 62, 397-404.

Kaiser, M.J. & Spencer, B.E. (1994). Fish scavenging behaviour in recently trawled areas. Marine Ecology Progress Series 112, 41-49.

Kaiser, M.J. (2003). Detecting the effects of fishing on seabed community diversity : Importance of scale and sample size. Conservation Biology 17(2), 512-520.

Laë, R., Ecoutin, J-M. & Kantoussan, J. (2004). The use of biological indicators for monitoring fisheries exploitation : Application to man-made reservoirs in Mali. Aquatic Living Resources 17, 95-105.

Le Corre, M. & Jaquemet, S. (2004). Assessment of the seabird community of the Mozambique Channel and its potential use as an indicator of tuna abumdance. Estuarine, Coastal and Shelf Science 63, 421-428.

Link, J.S. & Brodziak, J. (eds.) (2002). Report on the status of the NE US Continental Shelf Ecosystem. NEFSC Ecosystem Status Working Group. Northeast Fisheries Science Center Reference Document, 02-11. 245 pp.

Link, J.S. (2005). Translating ecosystem indicators into decision criteria. ICES Journal of Marine Science, 62, 569-576.

Link, J.S., Brodziak, J.K.T., Edwards, S.F., Overholtz, W.J., Mountain, D., Jossi, J.W., Smith, T.D. & Fogarty, M.J. (2002). Marine ecosystem assessment in a fisheries management context. Canadian Journal of Fisheries and Aquatic Sciences 59, 1429-1440.

Link, W.A. & Hatfield, J.S. (1990). Power calculations and trend analysis: a comment. Ecology 71, 1217-1220.

Mangel, M. (1993). Effects of high-seas driftnet fisheries on the northern right whale dolphin *Lissodelphis borealis*. Ecological Applications 3(2), 221-229.

Maxwell, D & Jennings, S. (2005). Power of monitoring programmes to detect decline and recovery of rare and valuable fish. Journal of Applied Ecology 42, 25-37.

Mayer,L.M., Schick, D.F., Findlay, R.H. & Rice, D.L. (1991). Effects of commercial dragging on sedimentary organic matter. Marine Environmental Research 31, 249-261.

McCullagh, P. & Nelder, J.A. (1989). Generalized linear models. 2nd edition. Monographs on Statistics and Applied Probability 37.

Milessi, A.C., Arancibia, H., Neira, S. & Defeo , O. (2005). The mean trophic level of Uruguayan landings during the period 1990-2001. Fisheries Research 74, 223-231.

Moloney, C.L., Jarre, A., Arancibia, H., Bozec, Y-M., Neira, S., Roux, J-P. & Shannon, L.J. (2005). Comparing the Benguela and Humboldt marine upwelling ecosystems with indicators derived from inter-calibrated models. ICES Journal of Marine Science 62, 493-502.

Mueter, F.J. & Megrey, B.A. (2005). Distribution of population-based indicators across multiple taxa to assess the status of Gulf of Alaska and Bering Sea groundfish communities. ICES Journal of Marine Science 62, 344-352.

Murawski, S.A. (2000). Definitions of overfishing from an ecosystem perspective. ICES Journal of Marine Science 57, 649-658.

Murawski, S.A., & Idoine, J.S. (1992). Multispecies size composition : a conservative property of exploited fishery systems? Journal of Northwest Atlantic Fishery Science 14, 79-85.

National Academy of Sciences (2000). Ecological Indicators for the Nation, National Academy Press, Washington D.C., 133 pp.

Nicholson, M.D. & Fryer, R.J. (1992). The statistical power of monitoring programmes. Marine Pollution Bulletin 24(3), 146-149.

Nicholson, M.D. & Fryer, R.J. (1993). The power of a contaminant monitoring programme to detect linear trends and incidents. ICES Journal of Marine Science 50, 161-168.

Nicholson, M.D. & Jennings, S. (2004). Testing candidate indicators to support ecosystem-based management:the power of monitoring surveys to detect temporal trends in fish community metrics. ICES Journal of Marine Science 61, 35-42.

Ortiz, M. (2002). Optimum sample size to detect perturbation effects : The importance of statistical power analysis – a critique. Marine Ecology 23(1), 1-9.

Pauly, D. (1995). Anecdotes and the shifting baseline syndrome of fisheries. Trends in Ecology and Evolution 10, 34.

Pauly, D., Christensen, V. & Walters, C. (2002). Ecopath, Ecosim and Ecospace as tools for evaluating ecosystem impact of fisheries. IICES Journal of Marine Science 57, 697-706.

Pauly, D., Christensen, V., Dalsgaard J., Froese, R., & Torres Jr., F. (1998). Fishing down marine food webs. Science 279, 860-863

Pauly, D., Christensen, V., Guénette, S., Pitcher, T.J., Sumalia, R.U., Walters, J.C., Watson, R.,& Zeller, D. (2002). Towards sustainability in world fisheries. Nature 418, 689-695.

Pennington, M.R. 1985. Estimating the relative abundance of fish from a series of trawl surveys. Biometrics 41: 197-202.

Pennington, M.R. 1986. Some statistical techniques for estimating abundance indices from trawl surveys. Fishery Bulletin 84:519-525.

Peterman, R.M. & Bradford, M.J. (1987). Statistical power of trends in fish abundance. Canadian Journal of Fisheries and Aquatic Sciences 44, 1879-1889.

Peterman, R.M. & M'Gonigle, M. (1992). Statistical Power Analysis and the Precautionary Principle. Marine Pollution Bulletin 24(5), 231-234

Peterman, R.M. (1990). Statistical power analysis can improve fisheries research and management. Canadian Journal of Fisheries and Aquatic Sciences 47, 2-15.

Philippart, C.M. (1998). Long-term impact of bottom fisheries on several by-catch species of demersal fish and benthic invertebrates in the south-eastern North Sea. ICES Journal of Marine Science 55, 342-352.

Piet, G.J. & Jennings, S. (2005). Response of potential community indicators to fishing. ICES Journal of Marine Science 62, 214-225.

Pinnegar, J.K. & Polunin, N.V.C. (2004). Predicting indirect of effects of fishing in Mediterranean rocky littoral communities using a dynamic simulation model. Ecological Modelling 172, 249-267.

Plagányi, É.E. & Butterworth, D.S. (2004). A critical look at the potential of Ecopath with Ecosim to assist in practical fisheries management. African Journal of Marine Science 26, 261-287.

Polovina J.J. & Own, M.D. (1983). Ecopath: a user's manual and program listings. NMFS/NOAA, Honululu Admin. Rep[. H-83-23, 46 pp.

Polovina J.J. (1985). An approach to estimating and ecosystem box model. US Fish. Bull. 83, 457-560

Polovina, J.J. (1984). Model of a coral reef system: I. The ECOPATH model and its application to French Frigate Shoals. Coral Reefs 3. 1-11.

Popper, K.R. (1962). Conjectures and refutations. Basic Books, New York, N.Y. 412pp.

Punt, A.E., Campbell, R.A. & Smith, A.D.M. (2001). Evaluating empirical indicators and reference points for fisheries management:application to the broadbill swordfish fishery off eastern Australia. Marine and Freshwater Research 52, 819-832.

Quinn, T.J. & Deriso, R.B. (1999). Quantitative Fish Dynamics. Oxford University Press. 542 pp.

Rice, J. & Gislason, H. (1996). 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 53, 1214-1225.

Rice, J.C. (2000). Evaluating fishery impacts using metrics of community structure. ICES Journal of Marine Science 57, 682-688.

Rochet, M.J. & Trenkel, V.M. (2003). Which community indicators can measure the impact of fishing? A review and proposals. Canadian Journal of Fisheries and Aquatic Sciences 60, 86-99.

Saetersdal, G. (1980). A review of past management of some pelagic fish stocks and its effectiveness. Rapp. P.-V. Reun. Cons. Int. Explor. Mer 177, 505-512.

Sainsbury, K. & Sumaila, U.R. (2001). Incorporating ecosystem objectives into management of sustainable fisheries, including "best practice" reference points and use of marine protected areas". In Summary of the Reykjavik Conference on Responsible Fisheries in the Marine Ecosystem, 1-4 October 2001. 19 pp.

Sala, E., Aburto-Oropza, O., Reza, M., Paredes, G. & Lopez-Lemus, L. (2004). Fishing down coastal food webs in the Gulf of California. Fisheries 29(3), 19-25.

Sanderson, H. & Petersen, S. (2001). Power analysis as a reflexive scientific tool for interpretation and implementation of the precautionary principle in the European Union. Environmental Science & Pollution Research 8, 1-6

Scandol, J.P. & Forrest, R.E. (2001) Commercial catches as an indicator of stock status in NSW estuarine fisheries: points, uncertainty and interpretation. Proceedings of the Australian Society for Fish Biology workshop "Towards sustainability of data-limited multi-sector fisheries". pp 77-97.

Shannon, L.J. & Cury, P.M. (2003). Indicators quantifying small pelagic fish interactions: apllication using a trophic model of the southern Benguela ecosystem. Ecological Indicators 3, 305-321.

Shannon, L.J., Cury, P.M. & Jarre, A. (2000). Modeeling effects of fishing in the Southern Benguela ecosystem. ICES Journal of Marine Science 57, 720-722.

Shin, Y-J. & Cury, P. (2004). Using an individual-based model of fish assemblages to study the response of size spectra to changes in fishing. Canadian Journal of Fisheries and Aquatic Sciences 61, 414-431.

Shin, Y-J., Rochet, M-J., Jennings, S., Field, J.G. & Gislason, H. (2005). Using size-based indicators to evaluate the ecosystem effects of fishing. ICES Journal of Marine Science 62, 384-396.

Steele, J.H. & Schumacher, M. (2000). Ecosystem structure before fishing. Fisheries Research 44, 201-210.

Steidl, R.J., Hayes, J.P. & Schauber, E. (1997). Statistical power analysis in wildlife research. Journal of Wildlife Management 61(2), 270-271.

Stevens, J.D., Bonfil, R., Dulvy, N.K. & Walker, P.A. (2000). ICES Journal of Marine Science 57, 476-494.

Stobberup, K.A., Inejih, C.A.O., Traoré, S., Monteiro, C., Amorim, P. & Erzini, K. (2005). Analysis of size spectra off northwest Africa: a useful indicator in tropical areas? ICES Journal of Marine Science 62, 424-429.

Swartzman, G.E., Silverman, E. & Williamson, N. (1995). Relating trends in walleye Pollock (*Theragra chalcogramma*) abundance in the Bering Sea to environmental factors. Canadian Journal of Fisheries and Aquatic Sciences 52, 369-380.

Taylor, B.L., Wade P.R., de Master, D.P. & Barlow, J. (2000). Incorporating uncertainty into management models for marine mammals. Conservation Biology 14(5), 1243-1252.

Thomas, L. & Krebs, C.J. (1997). A review of statistical power analysis software. Bulletin of the Ecological Society of America 78 (2), 126-139.

Thompson, P.M., Wilson, B., Grellier, K. & Hammond, P.S. (2000). Combining power analysis and populatin viability analysis to compare traditional and precautionary approaches to conservation of commercial cetaceans. Conservation Biology 14(5), 1253-1263.

Thrush, S.F., Hewitt, J.E., Cummings, V.J., Dayton, P.K., Cryer, M., Turner, S.J., Funnell, G.A., Budd, R.G., Milburn, C.J., Wilkinson, M.R. (1998).

Disturbance of the marine benthic habitat by commercial fishing : Impacts at the scale of the fishery. Ecological Applications 8(3), 866-879.

Toft, C.A. & Shea, P.J. (1983). Detecting community-wide patterns:estimating power strengthens statistical inference. American Naturalist 122, 618-625.

Trenkel, V.M. & M.J. Rochet (2003). Performance of indicators derived from abundance estimates for detecting the impact of fishing on a fish community. Canadian Journal of Fisheries and Aquatic Sciences 60, 67-85.

Vandermeulen, H. (1998). The development of marine indicators for coastal zone management. Ocean and Coastal Zone Management 39, 63-71.

Walters, C., Pauly, D. & Christensen, V. & Kitchell, J.F. (2000). Representing density dependent consequences of life history starategies in aquatic ecosystems: EcoSim II. Ecosystems 3, 70-83.

Walters, C., Pauly, D. & Christensen, V. (1999). Ecospace: prediction of mesoscale patterns in trophic relationships of exploited ecosystems, with emphasis on the impacts of marine protected areas. Ecosystems 2, 539-554.

Walther, G.R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T.J.C., Fromentin, J.C., Hoegh-Guldberg, O. & Bairlein, F. (2002). Ecological responses to recent climate change. Nature 416, 389-395.