# Testing whether biodiversity indicators detect policy induced change in marine ecosystems

Emily Nicholson1\*, Alberto Barausse2, Steven Mackinson3, EJ Milner-Gulland, Kathryn Sullivan4,5, Reg Watson6, Ben Collen7

\* e.nicholson@deakin.edu.au

1 – Deakin University, Australia

2 –

3 –

4 – Division of Biology, Imperial College London, Silwood Park, Berkshire, SL5 7PY, United Kingdom

5 – Institute of Zoology, Zoological Society of London, Regent’s Park, London NW1 4RY United Kingdom

6 –

7 – Centre for Biodiversity & Environmental Research, Department of Genetics, Evolution & Environment, University College London, Gower Street, London WC1E 6BT, United Kingdom

### Problems 6 July2017

* Uncertainty in RL status at T1. Use median instead of mean, then we get a real simulation?
* JUST REALISED THAT I AM INCLUDING DD SPECIES IN SAMPLED RLI, WITH A RANDOM NOT MEDIAN RLSTATUS. SHOULD I LAVE THEM OUT ALTOGETHER? OR SHOULD I ASSIGN A STATUS WITH MEDIAN SIMULATION? BOTH?
* What portion of the fishing effort is reuced by stoping bottom trawling in each model?
* Bootstrapping
* Could add the Gulf of California data? Check Nth Sea with Synonyms? Benguela?
* Improve figures
* Should we do a global analysis, where we combine the model outputs?
* Run whole thing with only species named in the model
* Add sppid into tables. Run a check for old listings, e.g. LR/lc in Nth Sea species or manually check.
* Should I exclude extrapolated fish spp by trophic level from fishlife?

some errors in RL status and RLI – need to do a check with RL website. Did that but some species not coming up in search.:

* Molva dypterygia

I’ve now set up synomyms to work for LPI searches, so species names are based on RL. But I should also do RL searches on synonyms. I need to go back and do this for Nth Sea. I have done for Aluetians and Nth Adriatic.

BIG assumption we have to take clear – assume that all species in a Fg respond in the same way.

Also wold expect LPI to do better in model only as focus on more common species; RL focus on threatened (often rare) species, also not named in mode.

Need to include only-in-model species as an analysis (ie minimal species extrapolation).

### Chat with Alberto Barausse 26 June 2017

**1. MTI**

Only model-based tests seems ok (as opposed to species list from FAO). Doesn’t help us with species that are not included in the models, for example species for which there is little information thus not in the model. There are real data on fish that are landed that are not included in the model.

Uncertainty/sensitivity analysis? Bootstrapping of biomass in functional groups (but also trophic level). Alberto thought that having less biomass per landed FG was ok – tests under-sampling.

**2. Extrapolation**

It should be species specific to a functional group, and the parameters relating to growth rate, size, trophic level, diet, predators and fecundity. So not just distribution and broad taxonomic group. May bring too much bias into the model. Though it not uncommon to see this done, but could bring strong biases.

For discussion – even for a species within a functional group, we don’t know what the individual species are doing, and relative abundance of those species.

### 3. Sensitivity analyses:

try only including species named in models, so not extrapolating out???

### Abstract

Biodiversity indicators provide vital information on the changing state of nature, and are used to measure progress towards global biodiversity targets, guide policy, and inform management interventions. Of the suite of indicators designated for such a role, few have been subjected to scrutiny on their ability to represent trends of interest, particularly in guiding policy choices. Here we combine model predictions of the impact of two broad fisheries policies to assess the performance of three biodiversity indicators in evaluating changes in marine ecosystems. We find that…[etc.]. To better guide policy choices, we recommend indicators are designed to specifically inform environmental management [etc.]. Further work is required to modify general ecosystem models to respond appropriately and realistically to anthropogenic pressures. With these modifications, indicators have the potential to tell us how we can best conserve biodiversity, not simply that we are failing to do so.

## Introduction

The targets set for sustainability under the Sustainable Development Goals (SDGs) and to stem biodiversity loss under the Convention on Biological Diversity (CBD) are associated with a suite of indicators. These indicators are used to report on progress towards the targets ([e.g., Butchart *et al.* 2010](#_ENREF_6); Tittensor et al. 2014). There has been a recent proliferation of indicators under such targets focussed on different aspects of biodiversity, including state, trends, pressures and responses – a necessary response to trying to measure a variety of metrics of *biodiversity*, a complex and multifaceted term (Purvis & Hector 2000 Nature). While it has been noted that the performance of indicators should be evaluated to ascertain their suitability to the task of measuring biodiversity change (Collen & Nicholson 2014; Mace & Baillie 2007), there have been few tests of their ability to represent trends of interest (but see [Branch *et al.* 2010](#_ENREF_3); [Nicholson *et al.* 2012](#_ENREF_19)){Costelloe, 2015 #2762}, particularly as a suite of indicators.

To date, the vast majority of indicators have been used to track change in metrics related to biodiversity e.g. abundance, extinction risk, distribution, with a further few measuring pressures causing those changes e.g. forest loss (Hansen et al. 2013), Human appropriation of Net Primary Productivity (Krausmann et al. 2013 PNAS). Research has focussed on measuring and predicting decline. While certain indicators have been subjected to testing on their design and coverage (e.g. Collen et al. 2009; Baillie et al. 2008), relationship to ecological theory (e.g. McCarthy et al. 2014; Buckland et al. 200X), and reliability to predict changes (e.g. Branch et al. 2010; Fulton et al. 2005), such testing is not widespread. In a mid-term review, Tittensor *et al*. (2014) identified 160 indicators that might provide information on progress toward the CBD Aichi targets. Only 55 of those indicators passed a rudimentary inspection and selection process; only 2 have undergone any thorough testing. There has been relatively little work carried out on the performance of indicators, and in particular on whether they can be used more proactively to generate predictions about the effectiveness of policy interventions.

As biodiversity indicators become increasingly integrated into environmental management, an integral part of that process must be to assess their ability to inform policy (Nicholson et al. 2012; Collen & Nicholson 2014). Effective environmental and conservation decisions require an explicit understanding of the links between desired outcomes of conservation, how those outcomes can be measured, and the proposed actions needed to achieve them (Sparks et al. 2011 Oryx; Collen & Nicholson 2014). More extensive stress testing of biodiversity indicators would enhance knowledge of how biodiversity is changing, show whether the existing indicators can adequately measure that change, and help identify the most appropriate policies to counteract biodiversity declines.

In this study, we sought to test 1) how the mathematical design of the indicators affects their behaviour, and 2) how sampling bias typical of these indicators affect their performance. We model the impacts of two broad fisheries policies (halting and halving bottom trawling effort) on three biodiversity indicators used by the Convention on Biological Diversity: the Red List Index (RLI, Butchart et al. 2004), the Living Planet Index (LPI, Loh et al. 2005) and the Marine Trophic Index (MTI), also referred to as mean trophic level ([Pauly & Watson 2005](#_ENREF_24)). For each indicator, we compare a modelled output (which we use to represent the ‘truth’) with the change illustrated by each of the indicators. Specifically, we asked (1) how well the current set of indicators reflected changes in hypothetical policies, (2) whether each indicator responded in the same way to the underlying ecosystem change, and (3) whether current data coverage of each indicator could adequately inform proactive attempts at policy comparison.

## Methods

### *Overall Approach*

We compiled ecosystem models from eight different ecosystems (Figure 1; Table 1) and simulated two broad hypothetical fisheries policies, which we imposed on the ecosystem models. From the model outputs, we sampled the projected changes in biomass, and from these data evaluated the response of three indicators: the IUCN Red list Index (RLI, Butchart et al. 2004), the Living Planet Index (LPI, Loh *et al*. 2005) and the Marine Trophic Index (MTI, [Pauly & Watson 2005](#_ENREF_24)). We selected these indicators as they represent important metrics for biodiversity conservation and global fisheries management: extinction risk, abundance trend, and trophic level, respectively.

### *Models and scenarios*

We used Ecopath with Ecosim models (Christensen & Walters 2004) implemented for eight Large Marine Ecosystems (Table 1). The Ecopath with Ecosim approach has been widely used as a tool for analysis in aquatic ecosystems, particularly for exploring fishing impacts and environmental change. Each of the models was run for 20 years with current fishing levels to stabilise before policy implementation in the 21st year.

After a 20 year run in, two policies were implemented: (i) a complete halt to trawling, where all bottom trawling fleets had their effort instantaneously changed to zero, and (ii) halving bottom trawling, where bottom trawl fleets instantaneously had their effort halved. The models were then run for 40 years under each scenario. Values for each of the three biodiversity indicators were estimated at annual intervals for 40 years. This time period allowed each of the models to stabilise again after the policy implementation (for certain models, particularly those with long-lived species, stabilisation took some time), while remaining within a feasible time period for projection. The models were focussed purely on the two hypothetical policies, and assumed stasis in a range of other factors such as change of effort in other gear types or fisher behaviour, environmental change such as climate change, and assumed no other management policies were implemented for the duration.

Halting trawling amounts to a reduction of between ¼ and 1/6 in the fishing fleet. This amounts to a reduction in total catch of

* Nth Adriatic: 42% reduction
* Nth Sea: 22%
* Aleutians: 27%
* Gulf California: 10%

### *Compiling species lists*

For each model, we compiled lists of species captured by the model functional groups. Lists were compiled between Nov 2015 and July 2017. Despite the systematic efforts, many species will have been omitted, mostly invertebrates, which have been less fully assessed.

First, we collated the names of all species named in the documentation for each model, e.g. for the North Sea, from the text and table 14.6 in {Mackinson, 2007 #2755}, or from supplementary material (Barausse) – these documents are listed in Table X. Some functional groups comprised a single species (or a stage of a single species), while many others were multiple species, some of which were named specifically (thus included), others were at the genus level (the IUCN RL was searched for those species). Only species will full species names, or readily identified by common name, were included from text and tables in the model documentation. Taxa identified to the genus level only, and not easily identified (particularly in larger functional groups) were excluded. We recorded whether species were listed in LPI, RL index or Fishbase, and collate all relevant data (e.g. IUCN status, generation time, link to Fishbase if applicable etc).

However, the named species per modelled functional group is not a comprehensive list of species the functional group might represent, but only the main species; several species may be missing from species lists, or captured by genus only, for the lower functional groups in particular (e.g. invertebrates). We sought to capture these species to make a fuller list, which meant extrapolating out, by making assumptions about unnamed species within each functional group, and whether they are captured effectively by the modelled output. In particular rare species, or odd ones that have little biomass, may not match well. The matching of species to functional groups was done by HG, checked by EN (and the modellers, e.g. Mackinson? We may then ask model authors to cross check these, which may then mean they should be co-authors). We extracted all RL, RLI and LPI species from the LMEs where models were situated (date). Each of the species not already linked with a functional group was allocated a functional group, where appropriate, or not allocated a group at all, if not adequately represented in the model.

To develop our long-list of species, we extracted species found within the modelled areas from the following databases:

* *Living Planet index:* Extract all species listed by the LPI that occur within the appropriate modelled area using the excel master spreadsheet (filtering data by location)
* *Red list index:* Advanced search for all species listed to occur in the appropriate FAO marine region (i.e. Eastern Central Pacific for the North Adriatic Sea - the narrowest search term available by location), by taxonomic group (e.g. seabirds, mammals, skates & rays, bony fish etc).
* *Lists of species in FAO marine areas (via locations):* Assess whether each species occurs within the appropriate LME (i.e. the North Sea etc) using range maps provided.
* *Fishbase***:**Identifying additional species that occur in each region that may not be listed by LPI or RL, through Advanced Search for all species listed to occur in the LME (where a model does not encompass an entire LME – e.g. North Adriatic Sea making up part of the Mediterranean LME – range maps are again used to determine if the species occurs within the area of interest).
* *The sea around us/Sealifebase:* Identifying additional invertebrate species that occur in each region that may not be listed by RL; Advanced search for all species listed to occur in LME.

We used only change in adult biomass when a species was found across multiple functional groups (e.g. Stellar Sea Iions in Aleutians, or some fish species e.g. Pollock in Aleutians). Only species names used, not to subspecies.

The result was a series of spreadsheets of species per model, with information on data sources, whether included in the LPI or RLI, and whether named in the model documentation, generation length, and current IUCN Red List status and criteria under which it is listed.

### *Calculation of indicators*

For each indicator, we calculated several versions, based on: 1) those species that typically make up the input data for the indicator; and 2) the longer compiled list of species available (which acted as ‘truth’, or at least, closer to the truth), which was between 50% to 100% longer that the typical indicator list (see Table X). In effect, differing subsets of the long list of species were sampled by each of the indicators in each scenario, also differentially sampling the functional groups. As an additional indicators we also compared outputs with trends in biomass in different functional groups or species. We included uncertainty in the indicator values, largely through bootstrapping (see details below), though other approaches could be applied {Soldaat, 2017 #3854}.

### *Marine trophic index (MTI)*

The MTI in a given year is a function of the landed biomass, *Y*, and the trophic level, *TL*, of landed groups or species *i* ([Pauly & Watson 2005](#_ENREF_24)):

Ideally this indicator should include all catch, i.e. landings and discards, rather than just landings ([Pauly & Watson 2005](#_ENREF_24)), to better represent the trophic level of the ecosystem. We calculated multiple MTIs based on model outputs; in all cases trophic levels (TLs) were those calculated within the Ecopath model for the given year.

In this analysis, we used only modelled outputs, and did not use any empirical data or lists of species currently used in the MTI (unlike the other two indicators). Using modelled biomass outputs, we generated four indices. We calculated two measures of MTI based on typical data:

(1) a landings-based measure (MTI-L), which included only species that were landed; and

(2) a catch-based measure (MTI-C), where discards are included. We generated two MTI indicators which represent a ‘true’ response with which to compare our policy impacts:

(3) Including all functional groups (and thus species) from the model above a given TL (see below), MTI-B; and

(4) Including all biomass of landed species, MTI-BL (as opposed to landed biomass).

Each of the four MTIs were calculated using two thresholds for trophic levels included:

1. Including functional groups with TL≥2, cited as the current CBD version ([Pauly & Watson 2005](#_ENREF_24));
2. Including functional groups with TL≥3.25, as recommended by Pauly & Watson ([2005](#_ENREF_24)).

In order to check our pre-policy implementation levels for each of the models, we compared the MTI value for each LME with those calculated by the Sea Around Us project (<http://www.seaaroundus.org/>; see appendix Table X).

We calculated proportional change in MTI from current levels after a 20-year run in time for each policy. Other studies of this nature (e.g. Branch et al. 2010) have compared MTI response standardised across the modelled systems by comparing with zero fishing. In the context of this study, the systems selected are mostly heavily fished, and have been in some cases for a considerable time. Therefore running the models removing fishing would take the models outside known state space, with potential spurious results.

To represent uncertainty around the estimated indices, we undertook bootstrapping, removing up to XX% the biomass of each group to replicate species being removed/excluded from the database.

### *Red List Index (RLI)*

The Red List Index is an aggregated measure of the response of biodiversity to the policy scenarios, measured by change in extinction risk class of a group of species over time. It is based on the proportion of species in each category on the IUCN Red List, and changes in this proportion over time resulting from genuine improvement or deterioration in the status of individual species (Butchart et al. 2004 PLoS Biology). We implemented the revised version of the Red List Index (Butchart et al. 2007 PLoS ONE), where the Red List Index (RLI) is defined as:

Where *W* is the weight assigned to the species *S* in each category *C* at time *t*, and *N* is the total number of species assessed at time *t*. Weightings for each category are: EX = 5, CR = 4, EN = 3, VU = 2, NT = 1, LC = 0.

From the model output for each of the two policies, we calculated RLIs for the species in each LME at 10-year intervals to compare to the RLI value from the year before implementation. At each decadal interval, each species was classified into a Red List category under criterion A (IUCN 2001; Mace et al. 2008), using the LME as if it were a global population. This is, in effect, a regional Red List assessment (Ref JP Rodrigues); RLIs have recently been calculated at the sub-global scale (e.g. Szabo et al. 2012 Biol Cons). This criterion relates to a measured rate of population reduction (criterion A), and in this study was based on an index of abundance (biomass). Due to the lack of distribution data, we were unable to implement criterion B. Species were assumed to not fall to the levels where criteria C or D were relevant.

Assessment of species under IUCN Red List criterion A requires knowledge of generation length of the species, as the population reduction is measured over the longer of 10 years or three generations. We used the following steps to estimate their generation length:

1. Generation lengths from existing Red List assessments, and the mammal estimates by {Pacifici, 2013 #3367}{Pacifici, 2013 #3807} (sharks? SRL of fish).
2. Estimated GL for over 10,300 spp from a NCEAS working group (Keith et al., 2015)
3. 1/adult mortality + age of first reproduction (option 2 in the RL guidelines {IUCN, 2016 #3808}. Adult mortality (M) is estimated per functional group from the model, and includes predation and other natural mortality (ie excludes fishing mortality. We scraped age of first reproduction from Fishbase using rfishbase v2.0 [ref]
4. Mean value for the functional group.
5. If no estimates were available for species in the FG, we used 1/(P/B), ie B/P from the ECOpath model for the functional group.

For the longer-lived species in our analysis, the assessment timeframe extended beyond the model timeframe, and so required estimates about past biomass/abundance trends. To generate estimates of trend previous to 2010, we used current Red List assessment to derive a best estimate of proportional change in abundance, and bounded (worst and best case) estimates of potential trends. The best and worst case estimates (Table 2) were used in a sensitivity analysis (see below); we assumed all trends were linear (Figure 2).

The ‘current’ or ‘sampled’ RL index is all species that have been assessed on RL in July 2017. The ‘true’ RLI is all species on the list we have assembled. For those that have not been Red Listed, we assume they are threatened at the same proportion as listed ones, randomly allocating a starting (year 1) RL status to each NE species; we run 1000 simulations of randomly allocated RL status. We did the same for DD species, although it is likely that many DD species are in fact more threatened than species that have not yet been assessed (Bland?). We then take the median RLI in year 1 as the main RLI for comparison.

Other options for allocating RL startus to non-redlisted species include:

* + - * + assume they are all fine
        + assume they are all threatened
        + assume a distribution from others in the functional group

Birds, mammals, amphibians, chondrichthians (sharks and rays), are all assessed. Most corals – all schleratinia group. Undersampled groups/groups with NE species will be some fish (though there is a sampled fish RL done – we can use that to estimate distribution of threat).

RL status at tome 1 is the current RL status; we used global status rather than regional.

We estimated two Red list indices: 1) only including species that are currently on the red List, ie have been assessed, and could potentially contribute to the Red List Index (noting that at present, only four taxonomic groups currently contribute to the Red list Index – corals, mammals, birds and ?); and 2) all species in the compiled list of species, with assumed generation length and Red List status (contributing to uncertainty around the estimates).

We then bootstrap the RLIs (‘true’ and sampled) to explore uncertainty due to species sampling. We (will as currently haven’t) taken the top and bottom 75th percentile of the simulated starting RL status and bootstrap those as well, to express uncertainty in both RL status and sampling.

### *Living Planet Index (LPI)*

The Living Planet Index tracks the relative change in abundance of a set of vertebrate species (Loh et al. 2005). We derived a simple formulation of the index assuming linear change (referred to as the chain method in Collen et al. 2009). This method simply calculates the log difference of biomass estimates for each species (*Nx*) in subsequent years (*t* and *t*-1) (assuming only one population of each species within each LME):

:

The rate of change per species x were averaged to give a mean inter-annual rate of change across species for each year *t*:

were aggregated into an index that is effectively a weighted geometric mean relative abundance:

The index was calculated by setting the base year (2015) to 1, and calculating subsequent changes in reference to the previous year.

We calculated LPIs for each of the two policies for each LME:

1. An LPI based on species currently in the LPI database;
2. an LPI with all species in the compiled list, which was always longer, and typically included additional function groups, as well as sampling more frequently from the functional groups currently represented in the LPI.

We undertook sensitivity analysis by bootstrapping… (see LPI papers).

**Analyses**

Put in details on…

* How do current indicators sample what is in the LME? (essentially a comparison of data in the RL and LPI datasets vs what is in the LME)
* Uncertainty
  + Generation length: range of Index values (sensitivity analysis)
  + Assumption about decline outside the simulation window (sensitivity analysis)
  + Maybe trophic level across FGs?
  + LPI: bootstrapping (the chance/influence of including a given species, or 10%)
  + Treat current index samples as ‘samples’. Try other subsets (e.g. random subsets of the same size, non-random across functional groups but random within functional groups.
  + How would we actively add to the current samples to redress the balance?

## Results

The concordance of response of the three indicators was relatively similar but variable among the three modelled systems. Each demonstrated that a complete reduction in fishing effort resulted in higher index values than halving effort. Etc…

How well they represented the ‘truth’ depending on evenness of representation across groups?? How to analyse that?

### *Marine trophic index*

Across the 4 modelled systems thus far, the MTI was highly variable. Generally the biomass MTI (‘truth’) increased slightly with the policies (with the exception of the Gulf of Califnornia – not sure what is happening there).

In short, the MTI of landings or catch were unpredictable, sometimes increasing, sometimes decreasing, sometimes with landings and catch going in the opposite directions.

Given that an increase in MTI suggests an improvement in biodiversity, the landings MTI is often giving the wrong picture. This is because a decline in the MTI may be the result of a range of factors, including reduction in effort, and type of effort (i.e. species targeted), which we would expect to result in certain shifts in landings MTI. A decline in MTI could simply mean that fisheries have changed target species (and this indeed is the case here – bottom trawl targets higher trophic level species) or spatial area targeted. Thus, a decline in MTI is supposed to be negative from an ecological point of view only if it takes place together with a certain, specific reduction in the amount of landings. This makes the interpretation of the index under policy scenarios complex. Note that Pauly and colleagues developed the “fishing in balance” (FIB) index to formalize this concept ([Pauly & Palomares 2005](#_ENREF_23); [Pauly & Watson 2005](#_ENREF_24)). [EXPAND ON REASONS IN DISCUSSION].

Note I have only done the simulations for TL of 3.25, as recommended by Pauly et al, not 2, as per CBD. Comparing ‘true’ MTIs: We find there is little difference between biomass MTIs of only landed species (MTI-B) or all species (MTI-B) with TL cut-off of 2, with no difference when the TL cut-off is 3.25. Therefore the baseline ‘truth’ for comparison matters little, and we use the all biomass MTI.

### *Red List Index*

The RLI was relatively reliable (once accounting for uncertainty in status of species that had not been evaluated), but insensitive. Does particularly well in the Nth Adriatic because most of the species named in the model are included in the RL. The LPI doesn’t have such good coverage.

How closely the sampled RLI tracked the ‘truth’ depended on how well groups other than mammals and birds were sampled. Because they tended to be over-represented, the sampled RLI tended to be lower than the true RLI, as it as more threatened species represented, or groups that decline (e.g. birds).

IN the North Sea, the sampled RLI only increases slowly. This is because of the large proportion of seabirds in the RLI, and their negative response to halting trawling due to loss of a major food source – discards. In fact, seabirds were the only functional group that showed a negative trend in the RLI, and thus any RLI that included seabirds also decreased in the first decade (Figure 5). The recovery in the RLI, even for seabirds, is due to the time weighting: for shorter-lived species, the decline in biomass is ‘forgotten’ due to the shorter time periods for assessment of decline. Thus although bird populations do not recover, the RLI loses track of the baseline for many species because it extends beyond the assessment period.

### *Living Planet Index*

The LPI was relatively reliable, the sampled LPI tended to be similar in shape to the ‘true’ LPI, though sometimes higher or lower, depending on the system. Only in one case so far, the Gulf of California, does it go the wrong way, because there are very few species in the LPI database within that modelled system; those that are there are birds, which decline in this system due to decline of key food species (“other fish”), through trophic interactions.

### *Biomass indicators*

Various groups declined or increased with a halt to trawling, depending on whether they were targeted directly by trawling, or had direct competitors or prey that increased or decreased as a result of trawling (see figs below – not very informative, need to find a better way to express these results; also need to go through the model in greater detail to understand why groups go up or down). Broader aggregations tended to give a very small change, because increases and decreases in different groups were effectively averaged out.

## Discussion

Biodiversity indicators are at a crucial point in their development. Environmental targets such as those devised under the Sustainable Development Goals and Convention on Biological Diversity have encouraged a large number of new indicators to be developed reflecting change in the many important metrics relating to biodiversity (Walpole et al. 2009; Pereira et al. 2013 Science EBV paper). While notable success has been achieved in monitoring changes in biodiversity with these indicators (Butchart et al. 2010; Tittensor et al. 2014), the extent to which those same indicators are capable of informing about the impact of proposed policy implementation or management interventions has received comparatively little scrutiny. Such a role could be important. Evidence–based modelling such as that undertaken here, can help define the causal relationships between policy actions, biodiversity change, and indicators. Appropriately designed indicators can then start to measure the success and failures of management, informing how we can best conserve biodiversity, not simply that we are failing to do so.

*Plan for discussion*

1. Our results demonstrate that… there are clear trade offs among indicators/all indicators perform well/something in between.

2. Critique of the indicators – how well do they reflect the ‘truth’?

3. What is the data coverage like among the indicators?

4. The underlying model – mass balance, equilibrium, etc…

A potential criticism of our study is that the indicators that we tested are not necessarily fit for the purpose of comparing management options into the future. Nevertheless, it rather defeats the purpose of having an indicator if all it can do is tell us what we know

if can’t tell us what we hope, and instead we rely on ‘knowing’ that stopping bottom trawling will be better for biodiversity, rather than being able to see the benefits reflected in our key biodiversity indicators.

Major assumptions that hopefully don’t affect the conclusions about the indicators:

* Extrapolated species match FG
* That all species in a FG respond in the same way to change (when clearly they don’t’)
* Ignoring other threats/factors outside model (e.g. nest predation as threat for seabirds)

***Em’s notes:***

The three CBD indicators all use the biomass trends from the model (and in fact, all track biomass or abundance trends), sampling different groups and weighted in different ways.

* The **MTI** weights biomass by the trophic level of the group, placing greater emphasis on those species at higher trophic levels; the MTI is based on landings and therefore only focuses on commercially important species, generally of high trophic levels.
* Because it is unevenly sampled, the **LPI** weights change in biomass by data availability, with each population time series weighted equally; the LPI database is taxonomically biased, including only vertebrates and dominated by birds.
* The **RLI** scales changes in biomass by the timeframe over which the change occurs, where the timeframe is species-specific and relates to the biology (generation time) of the species. It is also taxonomically biased, with larger proportions of vertebrates, especially birds.
* Our biomass indices basically weight by proportion of biomass; therefore changes in species/groups that contribute little to the biomass of the system (whatever their contribution ecologically) will have little overall impact on the biomass-based indicator. It samples everything that is in the model, though some groups may be left off, depending on the model coverage, or aggregated, depending on the model objectives.

There are two problems that result in the indicators behaving the opposite to expected/hoped (i.e. showing a negative effect of stopping trawling): sampling bias (the groups they are sampling); and using an indicator designed for other purposes to reflect biodiversity trends. The RL were not designed with such an indicator as the RLI in mind, but to assess relative extinction risk ([Mace *et al.* 2008](#_ENREF_17)). The MTI was not necessarily designed to assess change in large policies relating to fishing type or effort, but to assess the hypothesis of fishing down the food web ([Pauly *et al.* 1998](#_ENREF_22)).

Discards: {Fondo, 2015 #3565}

Criticisms of MTI: see refs in ([Villasante *et al.* 2012](#_ENREF_28)): ([Essington *et al.* 2006](#_ENREF_9); [de Mutsert *et al.* 2008](#_ENREF_8); [Branch *et al.* 2010](#_ENREF_3); [Sethi *et al.* 2010](#_ENREF_25))

Changes in MTI reflect change in effort. Total landings and total catch are reduced (as is B in Nth Sea), and effort is shifted onto species with lower TL (check?), thus MTI goes down, although the MTL of the biomass goes up. How do we design MTI that accounts for changes in effort/targets/policy sensibly? (other than adding the FIB index?).

One key problem, especially given the emphasis on birds in both the LPI and RLI, is that they are modelled within the Ecopath models as a single group, seabirds, with uniform responses, thus ignoring all other threats or reasons for decline (e.g. loss of breeding habitat, other sources of mortality etc). Again, this is because we are borrowing models and using them for purposes other than their original construction and design. Ideally, the models would be updated to separate either different seabird species, or groups of birds based on ecological function or common threats.

### Acknowledgements

This paper was conceived in a workshop funded by the NERC Centre for Population Biology, which also supported KS. EN was supported by a Marie Curie Fellowship, a Centenary Research Fellowship, Inspiring women Fellowship from the Victorian Government and Veski, and the Australian Research Council (DP170100609). BC is grateful for a travel grant from the Centre of Excellence for Environmental Decisions.

### Cited literature

Arreguin-Sanchez F., Arcos E. & Chavez E.A. (2002). Flows of biomass and structure in an exploited benthic ecosystem in the gulf of California, Mexico. *Ecol Modell*, 156, 167-183.

Barausse A., Duci A., Mazzoldi C., Artioli Y. & Palmeri L. (2009). Trophic network model of the northern Adriatic Sea: analysis of an exploited and eutrophic ecosystem. *Estuarine Coastal and Shelf Science*, 83, 577-590.

Branch T.A., Watson R., Fulton E.A., Jennings S., McGilliard C.R., Pablico G.T., Ricard D. & Tracey S.R. (2010). The trophic fingerprint of marine fisheries. *Nature*, 468, 431–435.

Butchart S.H.M., Akçakaya H.R., Chanson J., Baillie J.E.M., Collen B., Quader S., Turner W.R., Amin R., Stuart S.N. & Hilton-Taylor C. (2007). Improvements to the Red List Index. *PLoS ONE*, 2, e140.

Butchart S.H.M., Stattersfield A.J., Bennun L.A., Shutes S.M., Akçakaya H.R., Baillie J.E.M., Stuart S.N., Hilton-Taylor C. & Mace G.M. (2004). Measuring Global Trends in the Status of Biodiversity: Red List Indices for Birds. *PLoS Biology*, 2, e383.

Butchart S.H.M., Walpole M., Collen B., van Strien A., Scharlemann J.P.W., Almond R.E.A., Baillie J.E.M., Bomhard B., Brown C., Bruno J., Carpenter K.E., Carr G.M., Chanson J., Chenery A.M., Csirke J., Davidson N.C., Dentener F., Foster M., Galli A., Galloway J.N., Genovesi P., Gregory R.D., Hockings M., Kapos V., Lamarque J.-F., Leverington F., Loh J., McGeoch M.A., McRae L., Minasyan A., Morcillo M.H., Oldfield T.E.E., Pauly D., Quader S., Revenga C., Sauer J.R., Skolnik B., Spear D., Stanwell-Smith D., Stuart S.N., Symes A., Tierney M., Tyrrell T.D., Vié J.-C. & Watson R. (2010). Global biodiversity: indicators of recent declines. *Science*, 328, 1164-1168.

Collen B., Loh J., Whitmee S., McRae L., Amin R. & Baillie J.E.M. (2009). Monitoring Change in Vertebrate Abundance: the Living Planet Index. *Cons Biol*, 23, 317-327.

de Mutsert K., Cowan J.H., Essington T.E. & Hilborn R. (2008). Reanalyses of Gulf of Mexico fisheries data: Landings can be misleading in assessments of fisheries and fisheries ecosystems. *Proceedings of the National Academy of Sciences*, 105, 2740-2744.

Essington T.E., Beaudreau A.H. & Wiedenmann J. (2006). Fishing through marine food webs. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 3171-3175.

Fay G., Large S.I., Link J.S. & Gamble R.J. (2013). Testing systemic fishing responses with ecosystem indicators. *Ecol Modell*, 265, 45-55.

Gribble N.A. (2005). Ecosystem Modelling Of The Great Barrier Reef: A Balanced Trophic Biomass Approach. In: *MODSIM 2005 International Congress on Modelling and Simulation* (eds. Zerger A & Argent RM). Modelling and Simulation Society of Australia and New Zealand, pp. 2561-2567.

Heymans S.J.J. (2005). Ecosystem models of the Western and Central Aleutian Islands in 1963, 1979 and 1991. *University of British Columbia Fisheries Centre Research Reports*, 13, 8-82.

Hornborg S., Belgrano A., Bartolino V., Valentinsson D. & Ziegler F. (2013). Trophic indicators in fisheries: a call for re-evaluation. *Biol Lett*, 9.

Houle J.E., Farnsworth K.D. & Reid A.G.R.G. (2012). Assessing the sensitivity and specificity of fish community indicators to management action. *Can. J. Fish. Aquat. Sci.*, 69, 1065-1079.

Jiang H., Cheng H.-Q., Xub H.-G., Arreguín-Sánchez F., Zetina-Rejón M.J., Del Monte Luna P. & Le Quesne W.J.F. (2008). Trophic controls of jellyfish blooms and links with fisheries in the East China Sea. *Ecol Modell*, 212, 492-503.

Keith D.A., Mahony M., Hines H., Elith J., Regan T.J., Baumgartner J.B., Hunter D., Heard G.W., Mitchell N.J., Parris K.M., Penman T., Scheele B.E.N., Simpson C.C., Tingley R., Tracy C.R., West M. & AkÇAkaya H.R. (2014). Detecting Extinction Risk from Climate Change by IUCN Red List Criteria. *Cons Biol*, n/a-n/a.

Mace G.M., Collar N.J., Gaston K.J., Hilton-Taylor C., Akçakaya H.R., Leader-Williams N., Milner-Gulland E.J. & Stuart S.N. (2008). Quantification of extinction risk: IUCN’s system for classifying threatened species. *Cons Biol*, 22, 1424-1442.

Mackinson S. & Daskalov G. (2007). An ecosystem model of the North Sea to support an ecosystem approach to fisheries management: description and parameterisation. In: *Science Series Technical Report*. Cefas Lowestoft.

Nicholson E., Collen B., Barausse A., Blanchard J.L., Costelloe B.T., Sullivan K.M.E., Underwood F.M., Burn R.W., Fritz S., Jones J.P.G., McRae L., Possingham H.P. & Milner-Gulland E.J. (2012). Making robust policy decisions using global biodiversity indicators. *PLoS One*, 7, e41128.

Okey T.A. & Mahmoudi B. (2002). An Ecosystem Model of the West Florida Shelf for use in Fisheries Management and Ecological Research: Volume II. Model Construction. In. Florida Marine Research Institute St. Petersburg, Florida USA.

Okey T.A., Vargo G.A., Mackinson S., Vasconcellos M., Mahmoudi B. & Meyer C.A. (2004). Simulating community effects of sea floor shading by plankton blooms over the West Florida Shelf. *Ecol Modell*, 172, 339-359.

Pauly D., Christensen V., Dalsgaard J., Froese R. & Torres F. (1998). Fishing Down Marine Food Webs. *Science*, 279, 860-863.

Pauly D. & Palomares M.-L. (2005). Fishing Down Marine Food Web: It is Far More Pervasive Than We Thought. *Bulletin Of Marine Science*, 76, 197-212.

Pauly D. & Watson R. (2005). Background and interpretation of the ‘Marine Trophic Index’ as a measure of biodiversity. *Philisophical Transcations of the Royal Society B*, 360, 415–423.

Sethi S.A., Branch T.A. & Watson R. (2010). Global fishery development patterns are driven by profit but not trophic level. *Proceedings of the National Academy of Sciences*, 107, 12163-12167.

Shannon L.J., Moloney C.L., Jarre A. & Field J.G. (2003). Trophic flows in the southern Benguela during the 1980s and 1990s. *Journal of Marine Systems*, 39, 83-116.

Shin Y.-J., Shannon L.J., Bundy A., Coll M., Aydin K., Bez N., Blanchard J.L., Borges M.d.F., Diallo I., Diaz E., Heymans J.J., Hill L., Johannesen E., Jouffre D., Kifani S., Labrosse P., Link J.S., Mackinson S., Masski H., Möllmann C., Neira S., Ojaveer H., Abdallahi K.o.M., Perry I., Thiao D., Yemane D. & Cury P.M. (2010). Using indicators for evaluating, comparing, and communicating the ecological status of exploited marine ecosystems. 2. Setting the scene. *ICES Journal of Marine Science*, 67, 692-716.

Villasante S., Rodríguez D., Antelo M., Quaas M. & Österblom H. (2012). The Global Seafood Market Performance Index: A theoretical proposal and potential empirical applications. *Marine Policy*, 36, 142-152.

# Figures and Tables

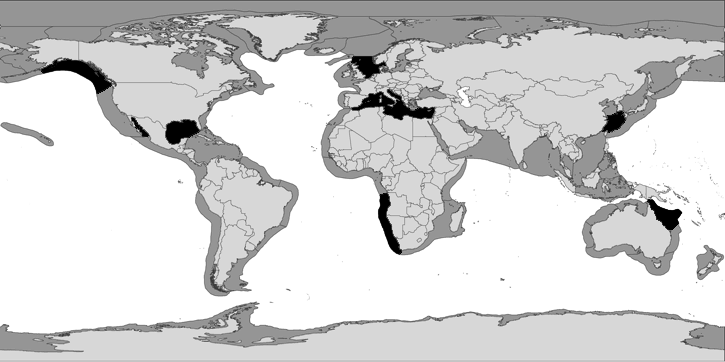


Figure 1. The Large Marine Ecosystem (black) modelled ….

Figure 3 – preliminary results (I will make it prettier!) **BUT see actual results in attached**.

Bue catch

Red landings

Black truth

|  |  |  |
| --- | --- | --- |
| LPI | RLI | MTI (relative change) |
| North Sea |  |  |
|  |  |  |
| Aleutians |  |  |
|  |  |  |
| Nth Adriatic |  |  |
|  |  |  |
| Gulf California |  |  |
|  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  | Nth Sea | Aleutians | Nth Adriatic |
|  | Gulf California |  |  |

**Figure 3.** Change in indicator values for the two scenarios in the North Sea, Aleutians and Northern Adriatic, using (a) Living Planet Index; and (b) Red List Index (will include Error bars show the upper and lower bounds of the estimates of Red list status based on two alternative assumptions about decline in the pre-model period); (c) Absolute values for MTI of species with TL greater than 3.5 (as recommended by Pauly & Watson 2005); (d) relative change in MTI; (e) shows the number of species within each broad taxonomic group for the RLI and LPI, compared with the overall species list (black).

Black lines show the ‘truth’ (full sample) for each indicator; red lines show impact of biased sampling (species currently red-listed in the RLI, species in the LPI database, and landings in the MTI); blue shows the catch-based MTI calculated with landings and discards; grey shows all landed species in the MTI, not just landed biomass; solid lines show a halt to bottom trawling, and dotted lines a halving of trawling effort at year 2.

What should NE species be assumed to be:

1. Assume all LC
2. Assume threatened in the same proportion as assessed (randomly assign a threat status, based on total average across all taxa, or across that functional group, or across functional with similar trophic level given that some FG have very few species)
3. Assume all threatened (similarly assigned by proportions across all species)

### Table 1. The eight Ecopath models used to simulate the policies of ending and halving bottom trawling, the stated objectives in the studies in which the models are described, the Large Marine Ecosystem (LME) in which the model system is situated, the number of functional or taxonomic groups (FG) each model contained, the number of these functional groups represented in the LPI, RLI and MTI, and the fraction of fishing fleets that were bottom trawl-based and thus affected by the policies. \*Models used in ([Branch *et al.* 2010](#_ENREF_3)); ^The model used for Southern Benguela is an updated version of the published model, provided by Lynne Shannon.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| System | Total number species | Total functional groups | Number species in LPI | Number functional groups in LPI | Number species in RLI | Number functional groups in RLI | Number functional groups in ‘true’ MTI | Number functional groups in landed MTI (50%) | Number functional groups in catch MTI (50%) |
| North Sea | 383 | 68 | 152 | 40 | 235 | 37 | 52 | 37 (39) | 42 (44) |
| Aleutians | 294 | 40 | 124 | 26 | 165 | 20 | 27 | 10(12) | 14(16) |
| NthAdriatic | 356 | 34 | 121 | 16 | 303 | 23 | 15 | 13 (13) | 14 (14) |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Model Region and reference*** | ***LME*** | ***Model objective*** | ***Model FG*** | ***FG in LPI*** | ***FG in RLI*** | ***Spp in RLI*** | ***FG in MTI*** | ***Area of model*** | ***Bottom trawl fleets/ total fleets*** |
| Central Gulf of California ([Arreguin-Sanchez *et al.* 2002](#_ENREF_1)) | Gulf of California  Whole LME | To characterize the trophic relationships and biomass flow paths; to learn the role of some functional groups, particularly of discards, in the ecosystem | 27 |  |  |  |  |  | 1/4 |
| East China Sea ([Jiang *et al.* 2008](#_ENREF_15)) | East China Sea  Whole LME | To examine possible mechanisms leading to jellyfish blooms and the impact of these blooms on fishery resources | 45 |  |  |  |  |  | 1/6 |
| \*Western and Central Aleutians ([Heymans 2005](#_ENREF_12)) | Gulf Alaska  This is its own LME - Aleutians | To examine the decline in the western stock of Steller sea lions, *Eumetopias jubatus* | 40 |  |  |  |  |  | 1/6 |
| \*North Sea ([Mackinson & Daskalov 2007](#_ENREF_18)) | North Sea | To quantitatively describe the ecological and spatial structure of species assemblages of the North Sea ecosystem; and to calibrate the dynamic responses of the modelled system by comparison with observed historical changes | 68 |  |  |  | 43 before policy, 41 after | 570,000 km2 | 4/12 |
| Northern Adriatic Sea ([Barausse *et al.* 2009](#_ENREF_2)) | Mediterranean Sea | To analyse the trophic structure of the system, identify the key trophic groups, and assess anthropogenic impacts on the ecosystem | 34 |  |  |  |  | 32,000 km2 | 2/6 |
| \*Great Barrier Reef ([Gribble 2005](#_ENREF_11)) | Northeast Australian shelf | To identify the effects of the major fisheries in each of the component systems, and the possible confounding effects of independently developed fisheries management plans | 32 |  |  |  |  |  | 1/3 |
| \*Southern Benguela  ([Shannon *et al.* 2003](#_ENREF_26)) ^ | Benguela Current (same model listed for Agulhas Current) | To identify data gaps and imbalances that result from inconsistencies between various stock assessments; …  to assess how observed differences or similarities in abundance, catches and dietary composition could affect overall trophic functioning, focusing on the pelagic part of the southern Benguela ecosystem | 27 |  |  |  |  |  | 1/6 |
| \*West Florida Shelf ([Okey & Mahmoudi 2002](#_ENREF_20); [Okey *et al.* 2004](#_ENREF_21)) | Gulf of Mexico | “to evaluate the potential effects of shading by phytoplankton blooms on community organization”  “The general questions addressed in this study were: (1) Are there multiyear trends in water transparency over the West Florida Shelf? (2) What proportion of the overall primary production on the West Florida Shelf is made up by microphytobenthos? (3) What broad community effects might result from nutrient enrichment and phytoplankton blooms?” | 59 |  |  |  |  |  | 1/11 |

**Table 2.** Rates of change used to infer a back-cast population change for species, prior to policy implementation, based on their 2014 IUCN Red List status.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Red List category** | **Criterion A rate of decline** | **Best estimate** | **Lower bound** | **Upper bound** |
| CR | 80-100% | -90% | -80% | -100% |
| EN | 50-80% | -65% | -50% | -79% |
| VU | 30-50% | -40% | -30% | -49% |
| NT | Inferred as 20-30% | -25% | -20% | -29% |
| LC | <20% up to any rate of increase | 0 | -19% | +20% |

# Supplementary material

Figure 2



**Figure 2.** Method used to infer declines in biomass prior to the modelled timeframe for longer-lived species, exemplified using (A) Northern Gannet (*Morus bassanus*), and (B) flounder (*Platichthys flesus*), in the North Sea.