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'

# Evaluation of indicators for ecosystem structure and functioning against screening criteria

G.J. Piet, H.M Jansen & M.J. Rochet

Project Deliverable Numbers 19, 20 & 21

Dissemination Level: Public

Due date: October 2006

Submission date: October 2006

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

Lead name and organisation: Indrani Lutchman, 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 INTRODUCTION

The third and final deliverable of the INDECO project work packages 2-4 is based on an indicator evaluation, designed for this project. This report presents the results of the indicator evaluation, together with the outcome of the analyses and the discussions at a workshop held on 5-6 September in London.

The evaluation aims at selecting key ecosystem indicators that are applicable for all geographical areas/ecosystems. The evaluation is based on an evaluation framework developed by Rice & Rochet (2005) for selecting an appropriate suite of indicators to support an ecosystem approach to fisheries management. The evaluation framework is structured as a sequence of eight steps:

Step 1: Determining user needs

Step 2: Listing candidate indicators

Step 3: Determining screening criteria

Step 4: Scoring indicators against criteria

Step 5: Summarizing scoring results

Step 6: Deciding how many indicators are needed

Step 7: Final selection

Step 8: Reporting

Within this study we focus on steps 2, 3,4 and 5 of the evaluation framework.

As this study was intended as one of the first attempts to explore the Rice & Rochet (2005) framework for the evaluation of this suite of indicators we will discuss the methodology used and its advantages, disadvantages and possible sources of bias.

The objectives of Work Packages 2-4 were to identify key biological and ecosystem indicators, but during the course of the project, it became evident that this would not be entirely possible due to a number of factors. These included the lack of detailed management objectives for the CFP, the state of knowledge about different ecosystem components and the lack of engagement of stakeholders in the evaluation process.

# 2 MATERIALS AND METHODS

## 2.1 Candidate indicators

In step 2 of the framework candidate indicators are listed. Candidate indicators should measure the ecosystem status relative to the management objectives. The list of candidate indicators used within this study is developed within deliverables 10, 11 and 12 and was based on the input of all INDECO partners. The aim was to cover all relevant ecosystem features for the State indicators and matching Pressure indicators including less informative proxies in case the required information is not available. For this we introduce a hierarchy from very broad and general features (e.g. physical/chemical, fish or other ecosystem components) to more specific features (e.g. physical environment or abundance of commercial stocks) to the actual indicator (e.g. 'Proportion of commercial stocks that are within safe biological limits'). In cases where no specific indicator has been developed for a particular ecosystem feature we used a more general phrasing (e.g. Abundance index of selected marine mammal species).

**Physical/chemical**

- Physical environment:
  - Temperature
  - NAO
- Chemical environment
  - Salinity
  - Oxygen levels
  - N and P levels (Eutrophication)

**Plankton**

- Phytoplankton
  - Primary production
  - Water transparency
  - Chlorophyll a level
- Zooplankton
  - CPR derived plankton indicators
  - Zooplankton biomass

**Fish**

- Abundance of commercial stocks
  - Proportion of commercial stocks that are within safe biological limits
- Abundance of populations that are not regularly assessed
  - Abundance (numbers) index of selected species (e.g. elasmobranches)
  - Biomass index of selected species (e.g. elasmobranches)
  - Decline (threat) indicator
- Size/Age Structure of a fish species
  - Average length of selected species
  - Average weight of selected species
  - Average age of selected species
- Genetic composition of a fish species
  - Maturation norm
- Size structure of the fish community
  - Mean weight
  - Mean length
  - Proportion of large fish
- Species composition including biodiversity of the fish community
  - Mean maximum length
  - Biodiversity indicator (Hill's N0)
  - Biodiversity indicator (Hill's N1)
  - Biodiversity indicator (Hill's N2)
  - Proportion of target species
- Abundance of the fish community
  - Total numbers
  - Total biomass

**Other ecosystem components**

- Status of marine mammals
  - Abundance index of selected marine mammal species
- Status of Seabirds
  - Abundance index of selected seabird species

- Status of marine reptiles
    - Abundance index of selected species
- Status of benthos
    - Abundance index of sensitive benthic species
    - Epibenthos community indicator
    - Infauna community indicator
- Status of sensitive habitat
    - Area coverage of highly sensitive habitats

**Ecosystem**

- Ecosystem functioning including trophic level
    - Primary production required
    - Catch ratios
    - Mean transfer efficiency
    - Trophic level
    - Fishing in Balance index
    - Finn Cycling Index

**Fishing pressure**

- Fleet capacity
- Fishing effort per métier and its spatial and temporal distribution
    - Days-at-sea or hours fished per spatial unit (e.g. ICES rectangle) per time (e.g. year or month)
- Fishing impact including catch, by-catch and habitat destruction
    - Fishing-induced proportion mortality of commercial fish species
    - Fishing-induced proportion mortality of non-assessed fish species
    - Fishing-induced proportion mortality of benthic species
    - Fishing-induced proportion mortality of marine mammals
    - Fishing-induced proportion mortality of vulnerable and/or protected species
    - Proportion of catch discarded
    - Proportion of area or (sensitive) habitat impacted

## 2.2 Criteria and stakeholders

In step 3 of the framework the screening criteria are determined. A list of criteria on which indicators can be evaluated is published by Rice & Rochet (2005). Table 1.1 provides for each of the screening criteria, and constituent considerations (sub-criteria) in conducting the scoring (H, high; F, fair; M, moderate; L, low) for an indicator (IND). Stars on items labelled H and L indicate that, if the consideration (or method of evaluation) is relevant, scoring high there is of high importance, and scoring low is a nearly fatal flaw, respectively. INDECO Deliverables 11, 12 and 13 (Piet and Pranovi, 2006) provide some background information on each of the criteria. The information presented there is not comprehensive but should give some more information on criteria used in the evaluation.

Since different stakeholder groups use indicators in different ways, ranking of indicators will give different results between the stakeholder groups. In order to obtain an overall and comprehensive view on the candidate indicators, the indicator evaluation should include all involved stakeholder groups (scientists, managers, politicians, community or environmental groups, economic stakeholders, public). Within this study we have tried to incorporate the opinions of all those stakeholder groups. However, since return rates of the non-scientist groups was low we only present the results of the scientists.

Table 1.1 List of criteria and sub-criteria on which indicators were evaluated (Rice & Rochet, 2005)

**Concreteness**

- Concrete property of physical/biological world (H), or abstract concept (L)?
- Units measurable in the real world (H), or arbitrary scaling factor (L)?
- Direct observations (H), or interpretation through model (L)?

**Public awareness**

- Is it a property with a high (H) or low (L) public awareness outside the use as an IND?
- Does public understanding correspond well (H) or poorly (L) with technical meaning of IND?
- If awareness high, is public likely to demand action that is: (i) proportional to IND value as determined by experts (H); (ii) disproportionately severe (M); (iii) largely indifferent (L)
- Does the nature of what constitutes ''serious harm'' (used to define a reference point) depend on values that are widely shared (H) or vary widely across interest groups (L)?
- Internationally binding agreements, national or regional legislation require that a specific IND be reported at regular intervals (H), to agreements/legislation require environmental status reporting, but IND not specified (M) to no such requirements (L).

**Theoretical basis (number of competing theories to allow contrast is important)**

- Not contested among professionals (H); (ii) basis credible, but debated e can account for patterns in many data sets (H-F, depending on how other models fit the same data); (iii) credible, but competing theories have adherents and empirical support is mixed (M); (iv) adherents, but key components untested or not generally accepted (M-L)
- If IND derived from empirical observations: (i) concepts readily reconciled with established theory (H); (ii) concepts not inconsistent with, but not accounted for by, ecological theory (M); (iii) concepts difficult to reconcile with ecological theory (L)**;
- Theory allows calculation of reference point associated with serious harm (M)*

**Cost**

- Uses measurement tools that are widely available and inexpensive to use (H), to needs new, costly, dedicated, and complex instrumentation (L)

**Measurement**

- Can variance and bias of IND be estimated? Yes (H); No (L)
- If variance can be estimated, is variance low (H) to high (L)
- If bias can be estimated, is bias low (H) to high (L)?
- If IND biased, is direction usually towards overestimating risk (H), or towards underestimating risk (L)
- If both can be estimated, have variance and bias been consistent over time (H), or have they varied substantially (L)
- Probability that IND value exceeds reference point can be estimated with accuracy and precision (H), to coarsely or not at all (L)**

- IND measured using tools with known accuracy and precision (H), to unknown or poor/ inconsistent (L)
- Value obtained for indicator unaffected by sampling gear (H), to sampling methods can be calibrated (M), to calibration difficult or not done (L)
- Seasonal variation unlikely or highly systematic (H) to irregular (L)
- Geographic variation irrelevant or stable and well quantified (H), through random (M) to systematic on scales inconsistent with feasible sampling (L)**
- Taxonomic representivity: IND reflects status of all taxa sampled/modelled (High), through ecologically predictable subset of species (M), to only specific species with no identifiable pattern of representivity (L)

**Availability of historical data**

- Necessary data are available for: periods of several decades (H) to only relatively recent period (M), to opportunistic or none available (L)
- Necessary data are: from the full area of interest (H), to restricted but consistent sampling sites (Moderate), to opportunistic and inconsistent sources, or none (L)**
- Necessary data have high contrast, including periods of harm and recovery (H), to high contrast but without known periods of harm and recovery (M), to uninformative about range of variation expected (Low)
- The quality of the data and archiving is known and good (H), to data scattered with reliability but not systematically certified, and archives not maintained (L) MP (e.g. environmental IND);
- Data sets are freely available to research community (H), to private or commercial holdings (L)

**Sensitivity** (length of time-series used for testing important)

- IND responds to fishing in ways that are: (i) smooth, monotonic, and with high slope (H)**; (ii) smooth, monotonic, and with low slope (M); (iii) smooth, monotonic over a restricted range of effort characteristics (M-F); (iv) unreliable (M-F, depending on when it fails to inform about fishing effects); (v) insensitive or irregular. Magnitude of response does not depend on magnitude of signal in effort (L)

**Responsiveness** (length of time-series used for testing important)

- IND changes within 1-3 years of implementation of measures (H), to IND only reflects system responses to management on decadal scales or longer (L)

**Specificity** (contrast in data set used for testing important)

- Is impact of environmental forcing on IND known, and small (H) or strong (L)?
- If environmental forcing affects IND, effect systematic and known (H), to irregular or poorly understood (L)**
- Relative to other factors, IND: (i) known to be unresponsive (H); (ii) responds to specific factors in known ways (M); (iii) thought to be unresponsive (F); (iv) responds to many factors in only partly understood ways (L)**

### 2.3 Scoring indicators against criteria; the questionnaires

In step 4 of the framework the indicators are scored against the criteria. The INDECO project includes partners with expertise on four ecosystems concomitant with the RAC areas: Baltic Sea, North Sea, Bay of Biscay (representative for the SW waters), and the Mediterranean. Therefore the INDECO partners (n=24) were asked to do the scoring for at least one of the above ecosystems, preferably the one they were most familiar with.

Scoring has two components: evaluation of quality criteria ('indicator scoring') and strength of evidence by which information is judged ('criteria weighting'). An ordinal scoring of 5 ranks for each indicator on each criterion is used. The strength of evidence of how the criteria are judged is also done by means of ordinal scoring of 5 ranks for each criterion. Judging of sub-criteria is done on a relative scale (sum of sub-criteria for each criterion is 1). Moreover, 'familiarity scoring' was included to investigate the relation between scoring and the familiarity of the evaluator with the indicators. Here an ordinal scoring of 3 ranks was used for each indicator.

Different types of questionnaires were provided (Table 2.2). The questionnaires were based on a hierarchical scale; both for the indicators (see section 2.1) and the criteria (see section 2.2) a hierarchical scale was introduced. In the quick version one should only score the indicators (general or specific) directly (i.e. not against specific criteria). The simple version scores general indicators against criteria, while the extended version is most comprehensive and all specific indicators were scored against all sub-criteria.

Table 2.2 Hierarchical scale of questionnaires

| | Indicators | | | Criteria | | |
|---|---|---|---|---|---|---|
| | General | Specific | None | Main | Sub-criteria | None |
| Indicator scoring - Quick version simple | X | | | | | X |
| Indicator scoring - Quick version extended | | X | | | | X |
| Indicator scoring - Simple version | X | | | X | | |
| Indicator scoring - Extended version | | X | | | X | |
| Familiarity scoring – Simple | X | | | | | X |
| Familiarity scoring – Extended | | X | | | | X |
| Criteria weighting | | | X | X | | |
| Sub-criteria weighting | | | X | | X | |

### 2.4 Analysis

For step 5 of the framework, 'Summarizing scoring results' we create a table where for each indicator (general or specific) a mean score is given. For the Quick versions this mean score is calculated as the mean across all responses, for the Simple and Extended versions the 'indicator scoring' per (sub)-criterion is weighted with the '(sub)criteria weighting' in order to derive one score per indicator per response and the mean is calculated across all responses. For the Simple and Extended versions a matrix with weights assigned to the (sub) criteria was available. For the simple and extended indicator scoring a second matrix with scores of each indicator on each criterion (separate scores for quality and weight) was available. In addition, weights assigned to each sub-indicator were available for the extended indicator scoring. Based on the available questionnaires (table 2.2) we distinguished different scenarios to derive a mean score per indicator.

For interpretation of the results we also had access to a scoring of the familiarity of the different respondents to the indicators that were scored. Together this resulted in 10 scenarios that provided scorings of indicators.

Scenario 1 – Quick version simple (average)

Scenario 2 – Quick version extended (average)

Scenario 3 – Simple version (average)

Scenario 4 – Simple version (median)

Scenario 5 – Extended version (average)

Scenario 6 – Extended version (median)

Scenario 7 – Extended to simple (average)

Scenario 8 – Extended to simple (best sub-indicator)

Scenario 9 – Familiarity scoring simple version

Scenario 10 – Familiarity scoring extended version

Rice and Rochet (2005) suggest graphic methods instead of the relatively simple methods as described above. Within such analysis we also make use of the same input: i) weights for criteria and ii) indicators against these criteria. First, figures display the distribution of scores and weights among scientists, or criteria. Second, we explore how Indicators and Criteria are judged relative to each other, across all scientists by a factor analysis of the scores (Mardia *et al.*, 1979), with missing data scored '0'. By positioning the indicators and criteria in a common space, the degree to which scientists differentiate these factors can be visualised.

# 3 RESULTS

## 3.1 Response level

As there were only very few responses from non-scientists (in total 4 non-scientist stakeholders completed the simple version) we excluded those in the analyses to avoid bias. The total number of responses from scientist according to each of the ecosystems is shown in Table 3.1.

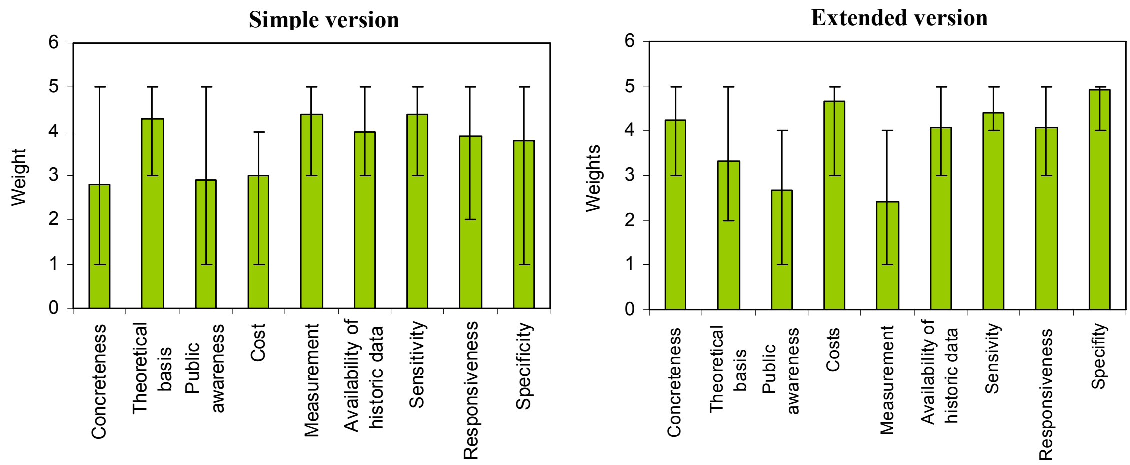
Table 3.1 Number of responses for each questionnaire per geographic region

| | Baltic | North Sea | Bay of Biscay | Mediterranean | **Total** |
|---|---|---|---|---|---|
| Indicator scoring - Quick version simple | 4 | 6 | 1 | 4 | **14** |

| | | | | | |
|---|---|---|---|---|---|
| Indicator scoring - Quick version extended | 4 | 6 | 1 | 4 | **14** |
| Indicator scoring - Simple version | 2 | 4 | 1 | 4 | **11** |
| Indicator scoring - Extended version | 2 | 3 | 2 | 5 | **12** |
| Familiarity scoring – Simple | 4 | 6 | 1 | 7 | **18** |
| Familiarity scoring – Extended | 4 | 6 | 1 | 7 | **18** |

## 3.2 Weights of criteria

The analysis of the weights of the criteria based on the simple version using only the results of the scientist respondents showed that overall, concreteness, public awareness, and cost got the lowest weights (Figure 3.1 left). For summarizing 'weighting results' by using the graphic methods see figure 3.3a. Results from the extended version showed a somewhat different pattern; public awareness weight was also low but here the weights of theoretical basis and




measurement were lowest (Figure 3.1 right).

Figure 3.1 Weights assigned to the criteria based on the Simple version (left) and Extended version (right)

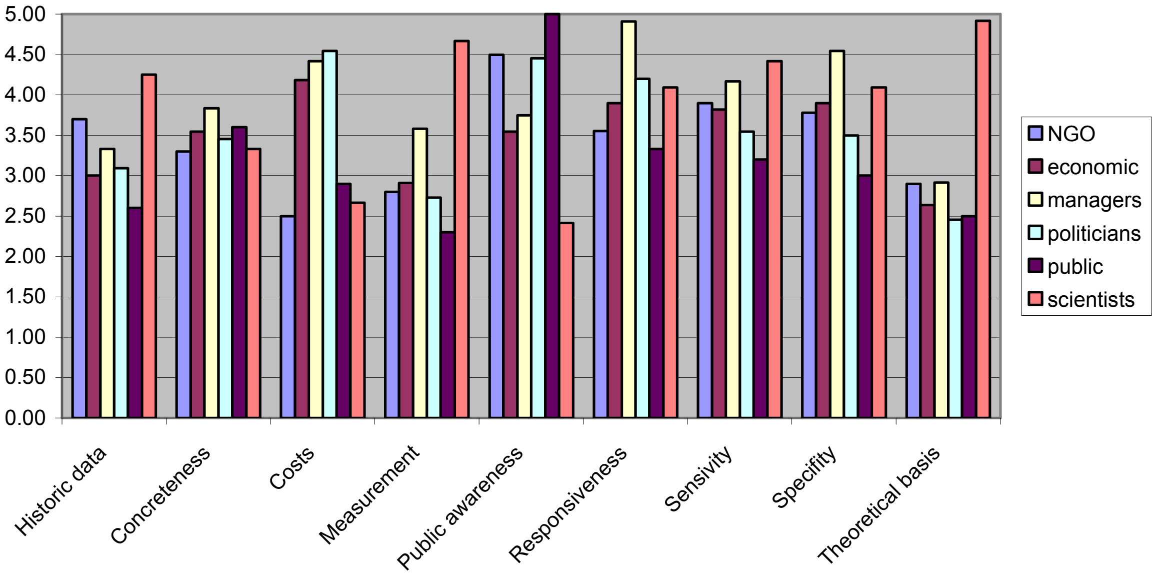



Figure 3.2 Weights that other stakeholder groups would assign to the criteria according to the scientists.

The weights that different stakeholder groups would attribute to the various criteria could only be studied using the weights that scientists thought the different stakeholder groups would give. To what extent these weights correspond with reality could not be tested as there were too few non-scientist stakeholders that responded. The weights the scientists gave, however, are similar to those based on the simple version in spite of the fact that these were not the same respondents which suggests that there is consistency and scientists at least appear to know how (other) scientists would weight the criteria.

This exercise shows that according to the scientists we can expect marked differences between the stakeholder groups with, not surprisingly, NGOs giving the highest weights to public awareness and the lowest to costs, economists giving highest weights to cost, lowest to theoretical basis, the managers giving the highest weights to responsiveness, lowest to theoretical basis, and finally the politicians giving the highest weights to cost and public awareness, lowest to theoretical basis.

## 3.3 Overview of indicator scoring

### *3.3.1 Simple methods*

Based on the relevant scenarios the scorings and rank order of both general and specific indicators are shown in respectively tables 3.2 and 3.3.

Overall, the fishing pressure indicators scored highest in most scenarios followed by the status of commercial stocks. Ecosystem functioning indicators and plankton indicators scored lowest. For the Physical/chemical indicators there were large differences between scenarios with scores that ranked them among the best indicators (ranks 2-4) to the worst indicators (ranks 19-20).

Table 3.2 Results of (general) indicator scoring using different methods for analyzing the data. For each scenario the rank order of the indicators is given between brackets.

| Indicators | Scenario | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 3 | 4 | 7 | 8 | 9 |
| **Physical/Chemical** | | | | | | |
| Physical environment | 2.6 (20) | 3.6 (4) | 3.3 (6) | 3.3 (6) | 3.7 (2) | 2.1 (12) |
| Chemical environment | 2.7 (19) | 3.5 (7) | 3.3 (6) | 3.4 (3) | 3.5 (4) | 2.0 (14) |
| **Plankton** | | | | | | |
| Phytoplankton | 2.9 (16) | 3.0 (15) | 2.9 (18) | 3.1 (9) | 3.3 (9) | 1.6 (18) |
| Zooplankton | 3.1 (15) | 2.9 (17) | 2.7 (19) | 2.6 (19) | 2.7 (19) | 1.7 (17) |
| **Fish** | | | | | | |
| Abundance of commercial stocks | 4.8 (1) | 3.8 (3) | 3.7 (3) | 3.4 (5) | 3.4 (6) | 2.7 (2) |
| Abundance of populations that are not regularly assessed | 3.9 (8) | 3.1 (12) | 3.1 (11) | 3.2 (8) | 3.3 (7) | 2.4 (7) |
| Size/Age Structure of a fish species | 4.3 (4) | 3.6 (5) | 3.6 (4) | 3.4 (3) | 3.5 (5) | 2.6 (3) |
| Genetic composition a fish species | 2.9 (17) | 2.9 (17) | 2.9 (14) | 2.6 (17) | 2.6 (20) | 1.4 (19) |
| Size structure of the fish community | 3.9 (7) | 3.2 (8) | 3.3 (8) | 3.3 (7) | 3.3 (7) | 2.6 (3) |
| Species composition including biodiversity of the fish community | 3.9 (8) | 3.1 (13) | 2.9 (14) | 2.7 (16) | 3.2 (11) | 2.4 (7) |
| Abundance of the fish community | 3.6 (13) | 3.1 (9) | 3.2 (10) | 2.9 (12) | 2.9 (14) | 2.5 (6) |
| **Other ecosystem components** | | | | | | |
| Status of Marine mammals | 3.8 (10) | 3.1 (10) | 3.1 (12) | 3.1 (10) | 2.8 (15) | 1.9 (15) |
| Status of Seabirds | 3.5 (14) | 3.1 (11) | 3.3 (9) | 2.8 (13) | 3.1 (12) | 1.7 (16) |
| Status of Marine reptiles | 2.9 (18) | 2.9 (16) | 2.9 (16) | 2.8 (15) | 2.8 (18) | 1.4 (20) |
| Status of Benthos | 3.7 (12) | 2.8 (19) | 2.9 (16) | 2.8 (14) | 2.8 (17) | 2.3 (10) |
| Status of sensetive Habitat | 4.1 | 3.0 | 3.1 | 2.6 | 2.8 | 2.1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | (6) | (14) | (13) | (18) | (15) | (11) |
| **Ecosystem** | | | | | | |
| Ecosystem functioning including trophic level | 3.7 (11) | 2.6 (20) | 2.6 (20) | 2.4 (20) | 3.1 (13) | 2.1 (12) |
| **Fishing Pressure** | | | | | | |
| Fleet capacity | 4.1 (5) | 4.3 (1) | 4.3 (1) | 3.9 (1) | 3.9 (1) | 2.7 (1) |
| Fishing effort per métier and its spatial and temporal distribution | 4.7 (2) | 3.9 (2) | 4.0 (2) | 3.7 (2) | 3.7 (2) | 2.6 (3) |
| Fishing impact including catch, by-catch and habitat destruction | 4.5 (3) | 3.5 (6) | 3.5 (5) | 3.0 (11) | 3.2 (10) | 2.4 (7) |

Table 3.3 Results of specific indicator scoring using different methods for analyzing the data. For each scenario the rank order of the indicators is given between brackets.

| Indicators | | Scenario | | | |
|---|---|---|---|---|---|
| **General** | **Specific** | **2** | **5** | **6** | **10** |
| **Physical/Chemical** | | | | | |
| Physical environment | Temperature | 2.85(41) | 3.65(2) | 3.58(4) | 2.22(23) |
| | NAO | 2.6(46) | 3.02(25) | 2.98(29) | 1.72(45) |
| Chemical environment | Salinity | 2.73(44) | 3.52(4) | 3.47(7) | 2.17(25) |
| | Oxygen levels | 2.82(43) | 3.49(5) | 3.62(2) | 2.11(28) |
| | N and P levels (Eutrophication) | 2.91(40) | 3.21(18) | 3.24(17) | 1.78(40) |
| **Plankton** | | | | | |
| Phytoplankton | Primary production | 3.08(37) | 2.92(31) | 2.89(35) | 1.83(37) |
| | Water transparency | 2.08(51) | 3.25(14) | 3.22(18) | 1.78(40) |
| | Chlorophyll a | 2.58(47) | 3.09(21) | 3.19(19) | 1.83(37) |
| Zooplankton | CPR derived plankton indicators | 2.64(45) | 2.46(43) | 2.58(42) | 1.39(51) |
| | Zooplankton biomass | 3.15(36) | 2.65(39) | 2.63(41) | 1.78(40) |

| Fish | | | | | |
|---|---|---|---|---|---|
| Abundance of commercial stocks | Proportion of commercial stocks that are within safe biological limits | 4.5(3) | 3.4(8) | 3.51(5) | 2.5(10) |
| Abundance of populations that are not regularly assessed | Abundance (numbers) index of selected species (e.g. elasmobranchs) | 4.21(5) | 3.29(10) | 3.32(13) | 2.33(17) |
| | Biomass index of selected species (e.g. elasmobranchs) | 3.86(16) | 3.25(14) | 3.31(15) | 2.39(16) |
| | Decline (threat) indicator | 4.14(8) | 2.97(26) | 3.04(26) | 2.22(23) |
| Size/Age Structure of a fish species | Average length of selected species | 4.15(6) | 3.45(6) | 3.44(8) | 2.56(5) |
| | Average weight of selected species | 3.92(14) | 3.44(7) | 3.48(6) | 2.56(5) |
| | Average age of selected species | 3.69(21) | 3.34(9) | 3.43(9) | 2.5(10) |
| Genetic composition of a fish species | Maturation norm | 2.83(42) | 2.63(40) | 2.94(31) | 1.72(45) |
| Size structure of the fish community | Mean weight | 3.42(29) | 3.27(13) | 3.32(13) | 2.61(2) |
| | Mean length | 3.42(29) | 3.28(12) | 3.33(12) | 2.61(2) |
| | Proportion of large fish | 3.5(25) | 3.29(10) | 3.17(20) | 2.56(5) |
| Species composition including biodiversity of the fish community | Mean maximum length | 3.25(32) | 3.18(20) | 3.13(21) | 2.56(5) |
| | Biodiversity indicator (Hill's N0) | 2.4(50) | 2.4(45) | 2.5(46) | 1.83(37) |
| | Biodiversity indicator (Hill's N1) | 2.44(48) | 2.37(46) | 2.46(47) | 1.89(34) |

| | | | | | |
|---|---|---|---|---|---|
| | Biodiversity indicator (Hill's N2) | 2.44(48) | 2.37(46) | 2.46(47) | 1.89(34) |
| | Proportion of target species | 3.25(32) | 3.06(23) | 3.06(25) | 2.5(10) |
| Abundance of the fish community | Total numbers | 3.54(24) | 2.92(31) | 2.9(33) | 2.5(10) |
| | Total biomass | 3.46(27) | 2.94(29) | 2.97(30) | 2.5(10) |
| **Other ecosystem components** | | | | | |
| | Abundance index of selected marine mammal species | 3.86(16) | 2.84(33) | 2.91(32) | 1.78(40) |
| Status of Seabirds | Abundance index of selected seabirds species | 3.64(22) | 2.76(37) | 2.78(37) | 1.67(48) |
| Status of Marine reptiles | Abundance index of selected species | 3.08(37) | 2.8(35) | 2.89(35) | 1.5(49) |
| Status of Benthos | Abundance index of sensitive benthic species | 3.86(16) | 2.84(33) | 2.76(38) | 2.28(20) |
| | Epibenthos community indicator | 3.33(31) | 2.61(41) | 2.57(43) | 2.06(31) |
| | Infauna community indicator | 3(39) | 2.42(44) | 2.54(45) | 1.89(34) |
| Status of sensitive Habitat | Area coverage of highly sensitive habitats | 3.5(25) | 3.07(22) | 2.99(28) | 2.17(25) |
| **Ecosystem** | | | | | |
| Ecosystem functioning including trophic level | Ecosystem functioning including trophic level | 3.77(20) | 2.13(50) | 1.82(51) | 2.11(28) |
| | Primary production required | 3.55(23 | 2.57(42) | 2.57(43) | 2(32) |

| | | | | | |
|---|---|---|---|---|---|
| | | ) | | | |
| | Catch ratios | 3.85(19) | 3.05(24) | 3.04(26) | 2.33(17) |
| | Mean transfer efficiency | 3.2(34) | 2.16(49) | 2.24(49) | 1.78(40) |
| | Trophic level | 3.92(14) | 2.68(38) | 2.73(39) | 2.28(20) |
| | Fishing in Balance index | 3.2(34) | 2.34(48) | 2.66(40) | 1.72(45) |
| | Finn Cycling Index | 3.43(28) | 1.99(51) | 2.19(50) | 1.44(50) |
| **Fishing Pressure** | | | | | |
| Fleet capacity | Fleet capacity (number of vessels) | 4.15(6) | 3.86(1) | 4(1) | 2.78(1) |
| Fishing effort per métier and its spatial and temporal distribution | Days-at-sea or hours fished per spatial unit (e.g. ICES rectangle) per time | 4.54(2) | 3.65(2) | 3.61(3) | 2.61(2) |
| Fishing impact including catch, by-catch and habitat destruction | Fishing-induced proportion mortality of commercial fish species | 4.57(1) | 3.19(19) | 3.26(16) | 2.56(5) |
| | Fishing-induced proportion mortality of non-assessed fish species | 4.14(8) | 2.95(28) | 3.11(23) | 2.28(20) |
| | Fishing-induced proportion mortality of benthic species | 4.07(11) | 2.8(35) | 2.9(33) | 2.11(28) |
| | Fishing-induced proportion mortality of marine mammals | 4.07(11) | 2.96(27) | 3.08(24) | 1.94(33) |
| | Fishing-induced proportion mortality of vulnerable and/or protected species | 4.5(3) | 2.93(30) | 3.13(21) | 2.17(25) |
| | Proportion of catch discarded | 4.07(11 | 3.23(16) | 3.35(11) | 2.44(15) |

| | | | | | |
|---|---|---|---|---|---|
| | | ) | | | |
| | Proportion of area or (sensitive) habitat impacted | 4.14(8) | 3.23(16) | 3.41(10) | 2.33(17) |

For each of the tables we tested the similarity in the outcomes of the different scenarios using a correlation analysis. This shows that the scenarios are reasonably consistent in their evaluation of the indicators and that this is significantly affected by the familiarity with the indicators. For both the general and specific indicator scoring the scenario that differs most from the others is scenarios 1 and 2: the 'Quick version'.

Table 3.4 Correlation coefficients (upper right) significance (lower left) for different scenarios against each other

| | **Scenario_1** | **Scenario_3** | **Scenario_4** | **Scenario_7** | **Scenario_8** | **Scenario_9** |
|---|---|---|---|---|---|---|
| Scenario_1 | | 0.47 | 0.56 | 0.33 | 0.30 | 0.77 |
| Scenario_3 | 0.04 | | 0.97 | 0.91 | 0.84 | 0.63 |
| Scenario_4 | 0.01 | <0.01 | | 0.89 | 0.78 | 0.66 |
| Scenario_7 | 0.16 | <0.01 | <0.01 | | 0.87 | 0.58 |
| Scenario_8 | 0.20 | <0.01 | <0.01 | <0.01 | | 0.62 |
| Scenario_9 | <0.01 | <0.01 | <0.01 | <0.01 | <0.01 | |

Table 3.5 Correlation coefficients (upper right) significance (lower left) for different scenarios against each other

| | **Scenario_2** | **Scenario_5** | **Scenario_6** | **Scenario_10** |
|---|---|---|---|---|
| Scenario_2 | | 0.32 | 0.30 | 0.62 |
| Scenario_5 | 0.03 | | 0.97 | 0.67 |
| Scenario_6 | 0.05 | <0.01 | <0.01 | 0.61 |
| Scenario_10 | <0.01 | <0.01 | <0.01 | |

### *3.3.2 Graphic methods*

Different scientists gave different distributions of scores (Figure 3.3a and 3.4). As a consequence, average or even modal scores are difficult to interpret. This is the reason why in this section we describe the distribution of scores across scientists, rather than statistical summaries. For scientists overall, sensitivity, measurability and theoretical basis appeared to be more important than concreteness, public awareness, and cost (Figure 3.3b). Indicators with the highest scores across scientists included Abundance of commercial stocks, Fleet capacity, Size/Age Structure of a fish species, Fishing effort per métier and its spatial and temporal distribution, and Fishing impact including catch, by-catch and habitat destruction; those with the lowest scores included Status of Marine reptiles, Ecosystem functioning including trophic level, Zooplankton and Phytoplankton, and Genetic composition of a fish species (Figure 3.3c).

Abundance of commercial stocks, Size/Age Structure of a fish species, and Fishing impact including catch, by-catch and habitat destruction, were given the highest scores on Sensitivity, Responsiveness and Specificity, whereas Fleet capacity, Physical and Chemical environment, the Size/Age Structure of a fish species, and Plankton indicators were scored high on Concreteness, Measurability, Cost, Availability of historic data, and Public awareness (Figure 3.4 and 3.5). Status of Marine reptiles and Ecosystem functioning including trophic level were both given very low scores on these criteria. Generally indicators received a wide diversity of scores across the multiple criteria (Figure 3.4, bottom). As a result, many of them have intermediate average scores.

An extended evaluation of a more detailed list of indicators against a list of sub-criteria reveals a different ranking, with two physical indicators scoring highest. However, fleet capacity and size-based indicators of the fish community still have high scores with this extended method, whereas indicators of ecosystem functioning and plankton still have low scores (Figure 3.6). Generally, sub-criteria within each criterion were correlated, and this was particularly true for Measurability, Availability of historic data, and Concreteness (Figure 3.7).

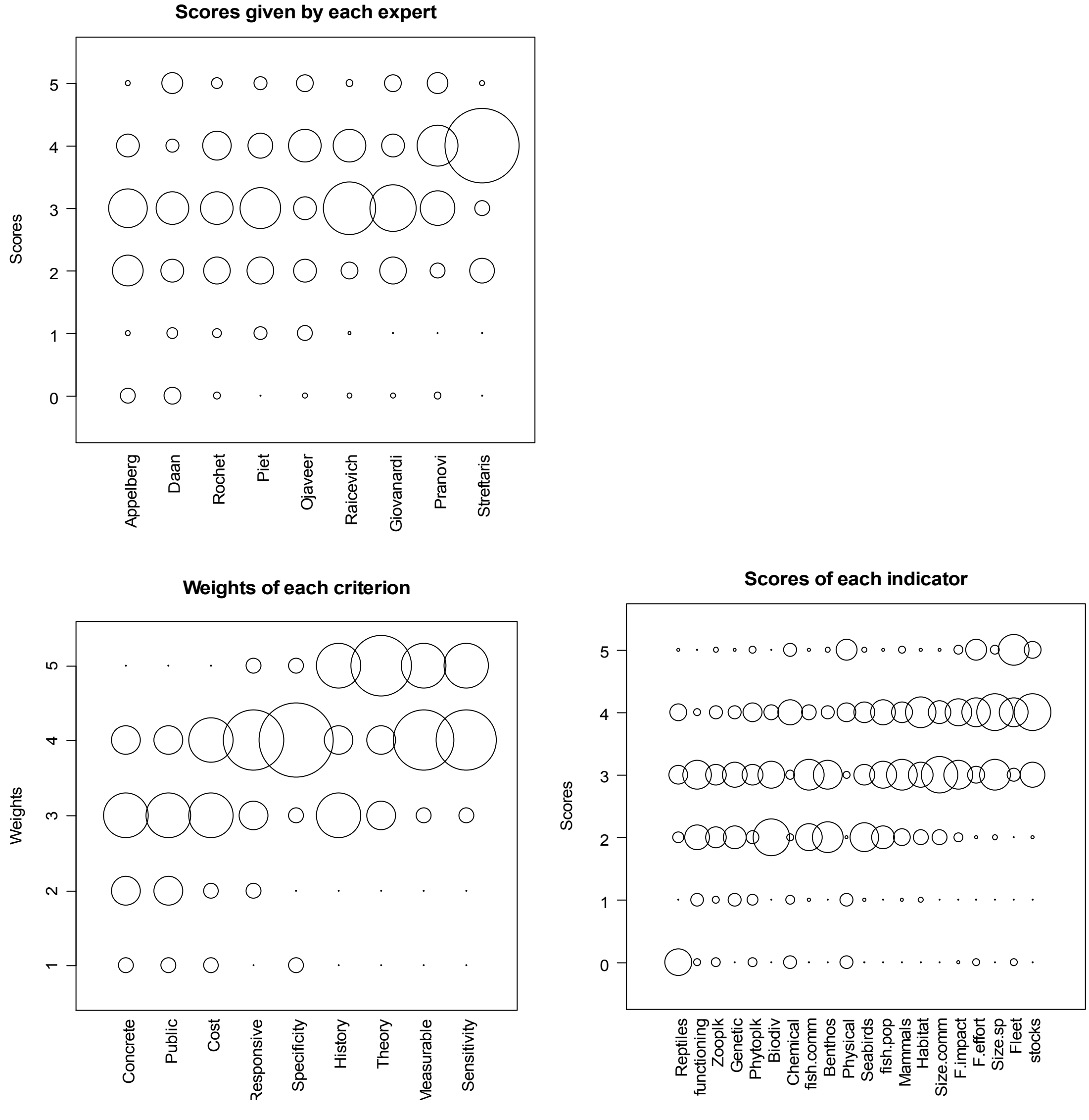



Figure 3.3 Distribution of a) scores given by scientists across indicators, b) weights given to criteria across scientists c) scores given to indicators across scientists. Within each plot, items are ordered from left to right by increasing sum of scores (or weights).

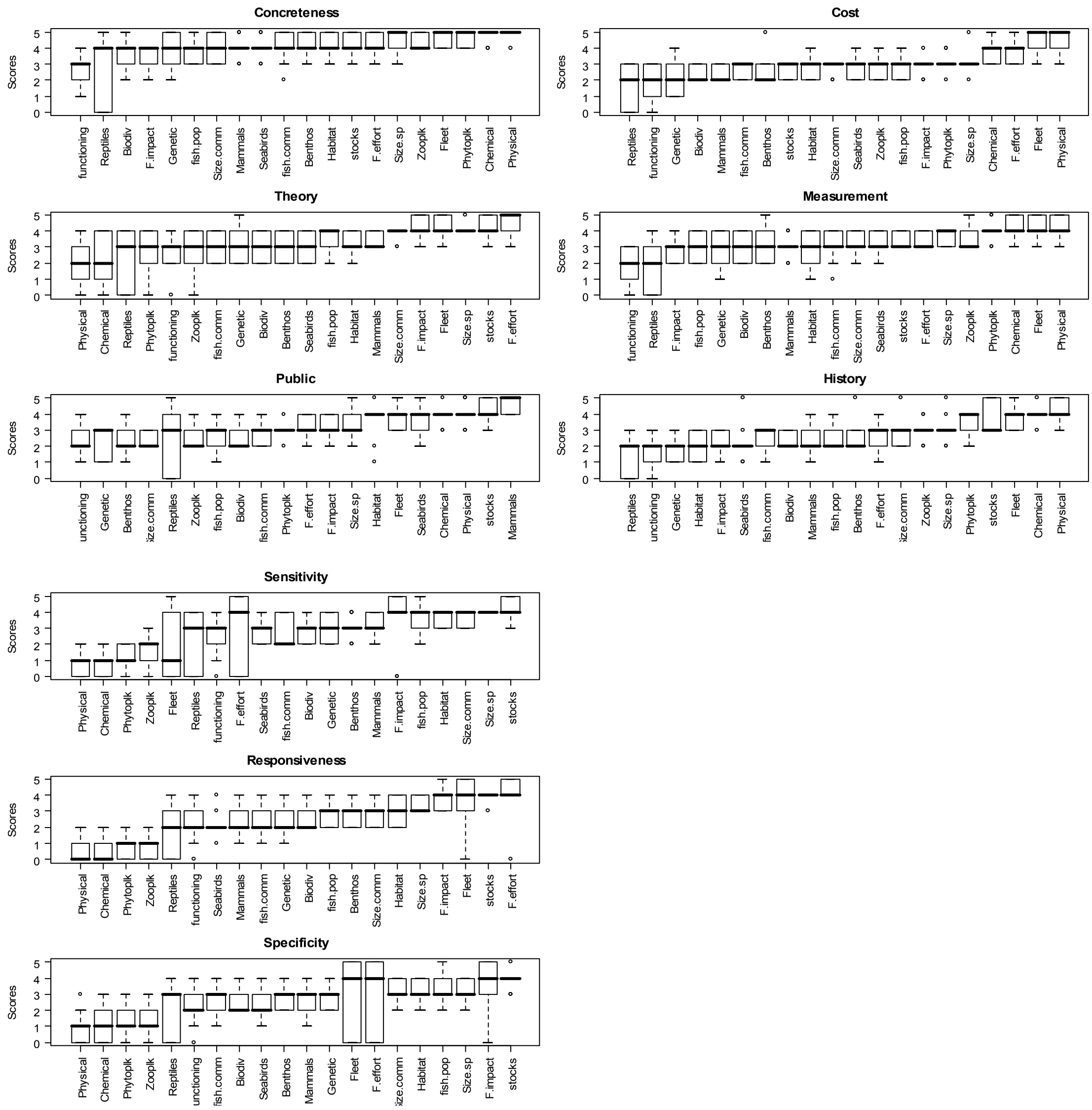



Figure 3.4 Distribution of scores given by 9 scientists to each indicator against each criterion, ordered by increasing sum of scores.

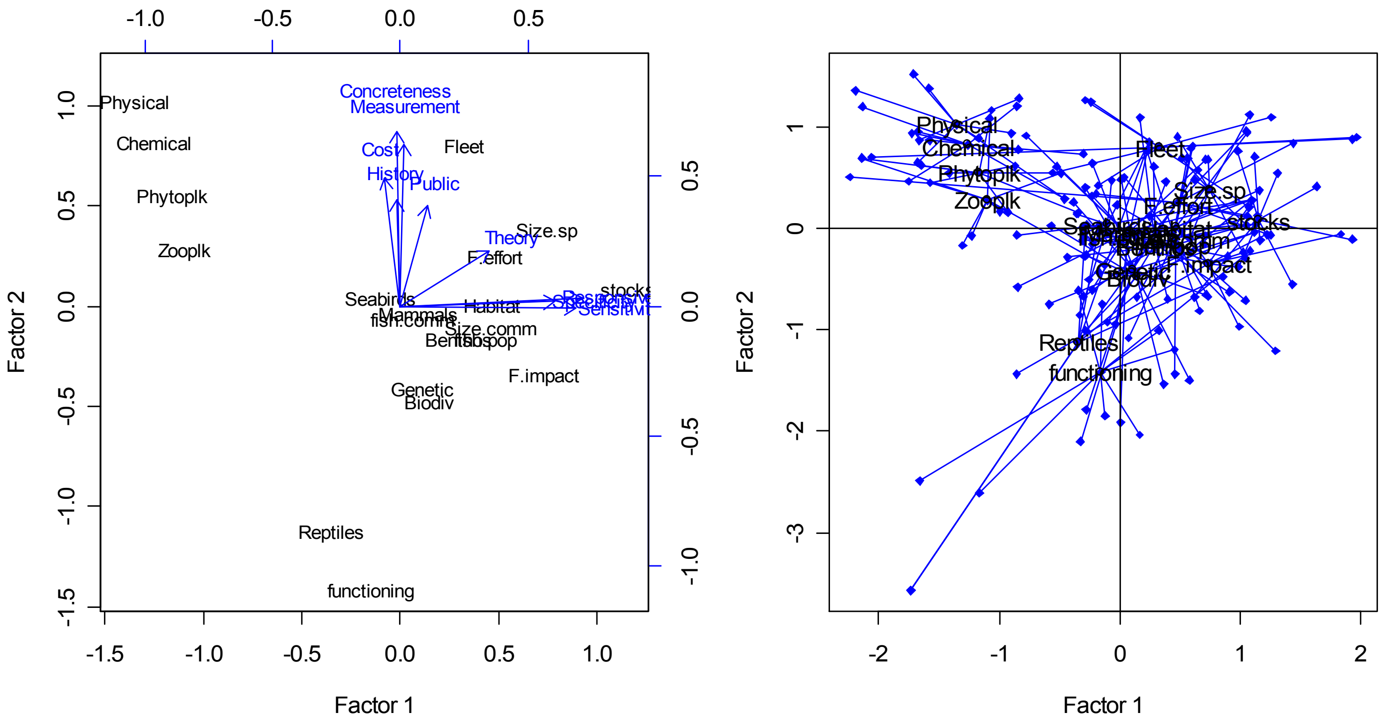



Figure 3.5 Factor analysis of indicators scores by 9 scientists. Average scores of indicators and criteria loadings on the factors 1 (26% of variance) and 2 (25%) (left). Dispersion of scores given to each indicator: each individual projection is connected to the average scores of indicators, along the same factors as top (right).

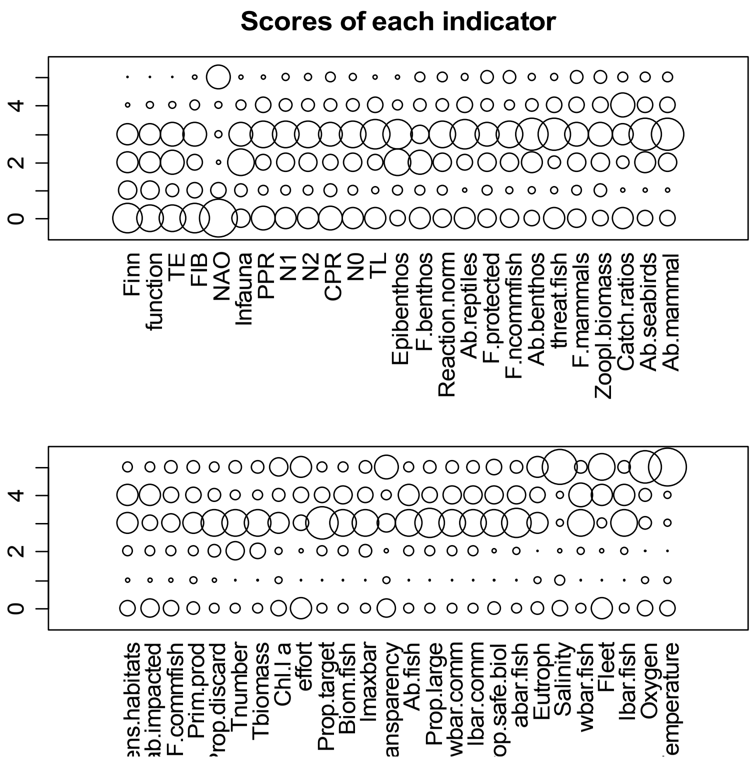



Figure 3.6 Distribution of scores given by scientists across indicators in the extended evaluation. Indicators are ordered from left to right by increasing sum of scores.

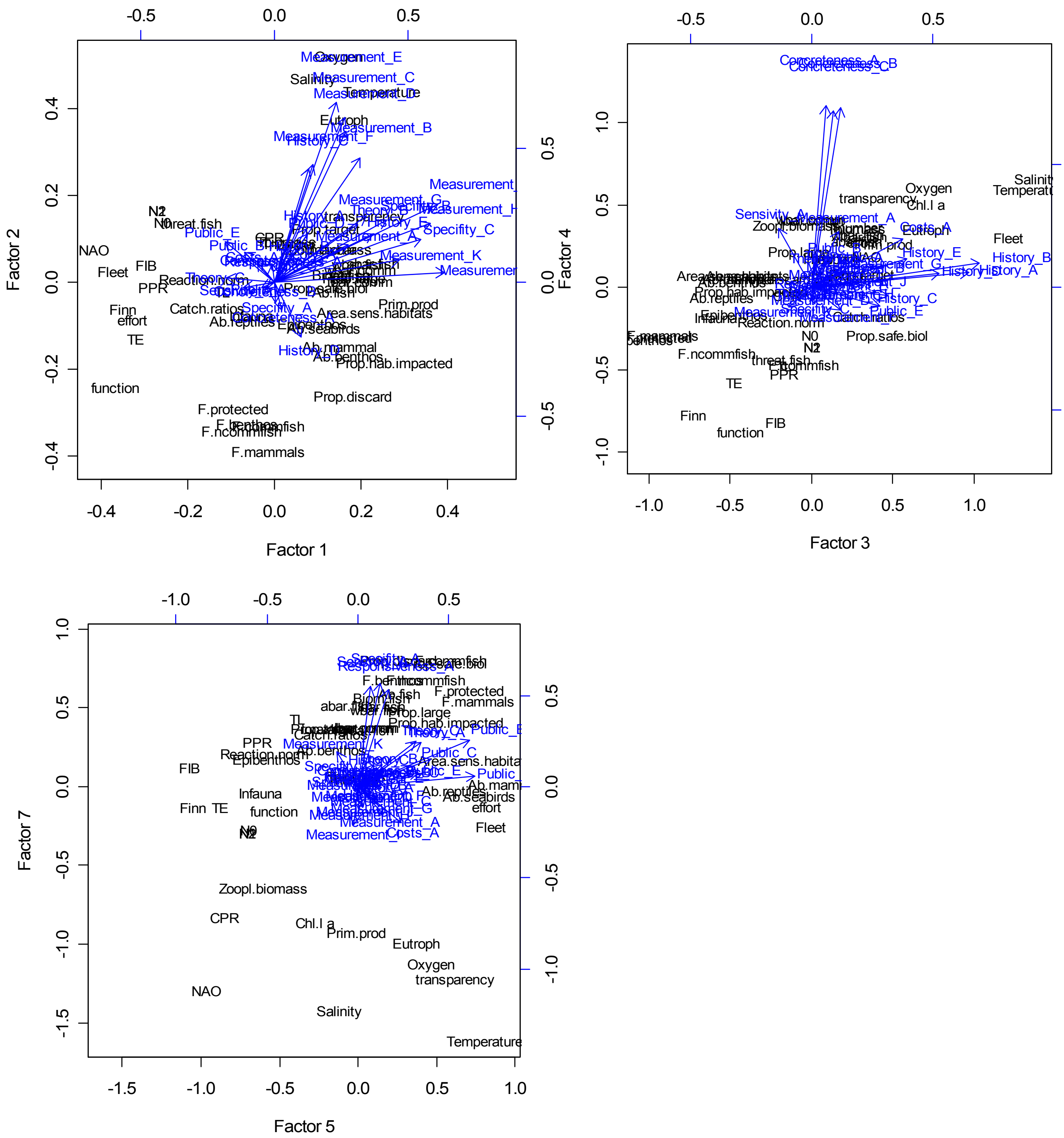



Figure 3.7 Factor analysis of extended evaluation of indicators by 1 scientists. Average scores of indicators and criteria loadings on the factors 1 to 5, and 7 (respectively: 13, 11, 9, 9, 7, and 7% of variance).

### 3.4 Differences between regions

Table 3.6 shows an overview of indicator scoring between the four regions. No clear differences in scoring patterns between the regions were observed.

Table 3.6 Results of (general) indicator scoring by area using scenario-3 for analyzing the data

| Indicators | Area | | | |
|---|---|---|---|---|
| | Medi-terranean | North Sea | Baltic | Bay of Biscay |
| **Physical/Chemical** | | | | |
| Physical environment | 3.3 (6) | 4.5 (2) | 3.4 (5) | 3.7 (7) |
| Chemical environment | 3.3 (5) | 3.8 (3) | 3.3 (7) | 3.4 (9) |
| **Plankton** | | | | |
| Phytoplankton | 2.9 (14) | 2.6 (17) | 3.0 (12) | 2.7 (16) |
| Zooplankton | 2.9 (15) | 2.6 (18) | 2.9 (19) | 2.6 (18) |
| **Fish** | | | | |
| Abundance of commercial stocks | 3.7 (3) | 3.2 (9) | 3.6 (3) | 4.4 (1) |
| Abundance of populations that are not regularly assessed | 2.8 (17) | 2.9 (13) | 3.1 (9) | 3.4 (8) |
| Size/Age Structure of a fish species | 3.7 (4) | 3.6 (4) | 3.3 (6) | 3.8 (4) |
| Genetic composition a fish species | 2.4 (19) | 2.3 (19) | 3.1 (10) | 2.8 (15) |
| Size structure of the fish community | 3.2 (9) | 3.3 (8) | 2.9 (14) | 3.7 (5) |
| Species composition including biodiversity of the fish community | 2.9 (13) | 2.9 (16) | 3.0 (11) | 3.2 (11) |
| Abundance of the fish community | 2.9 (7) | 3.3 (7) | 3.0 (13) | 3.1 (12) |
| **Other ecosystem components** | | | | |
| Status of Marine mammals | 3.2 (8) | 3.0 (10) | 2.9 (15) | 3.3 (10) |
| Status of Seabirds | 3.1 (16) | 3.0 (11) | 3.3 (8) | 2.8 (14) |
| Status of Marine reptiles | 2.9 (12) | 2.9 (14) | 2.9 (16) | 2.5 (19) |
| Status of Benthos | 1.9 (20) | 2.9 (15) | 2.9 (17) | 2.7 (17) |
| Status of sensitive Habitat | 3.1 (10) | 2.9 (12) | 2.9 (18) | 2.9 (13) |
| **Ecosystem** | | | | |
| Ecosystem functioning including trophic level | 2.5 (18) | 2.1 (20) | 2.7 (20) | 2.5 (20) |
| **Fishing Pressure** | | | | |
| Fleet capacity | 4.0 (1) | 4.5 (1) | 4.1 (1) | 4.3 (2) |

| | | | | |
|---|---|---|---|---|
| Fishing effort per métier and its spatial and temporal distribution | 3.8 (2) | 3.6 (5) | 3.9 (2) | 4.1 (3) |
| Fishing impact including catch, by-catch and habitat destruction | 3.3 (11) | 3.5 (6) | 3.5 (4) | 3.7 (6) |

### 3.5 Familiarity scoring

To what extent the scoring of the indicators is affected by the level of familiarity with an indicator was tested by asking everyone to indicate the level of familiarity with an indicator for both general (table 3.7) and specific indicators (table 3.8). It appears that the quick scores are more affected by the level of familiarity than the scores based on an evaluation against criteria (Figure 3.8). However, correlation analyses show that scoring of all questionnaires (quick, simple, extended) is significantly influenced by the familiarity with the indicators (see section 3.3.1).

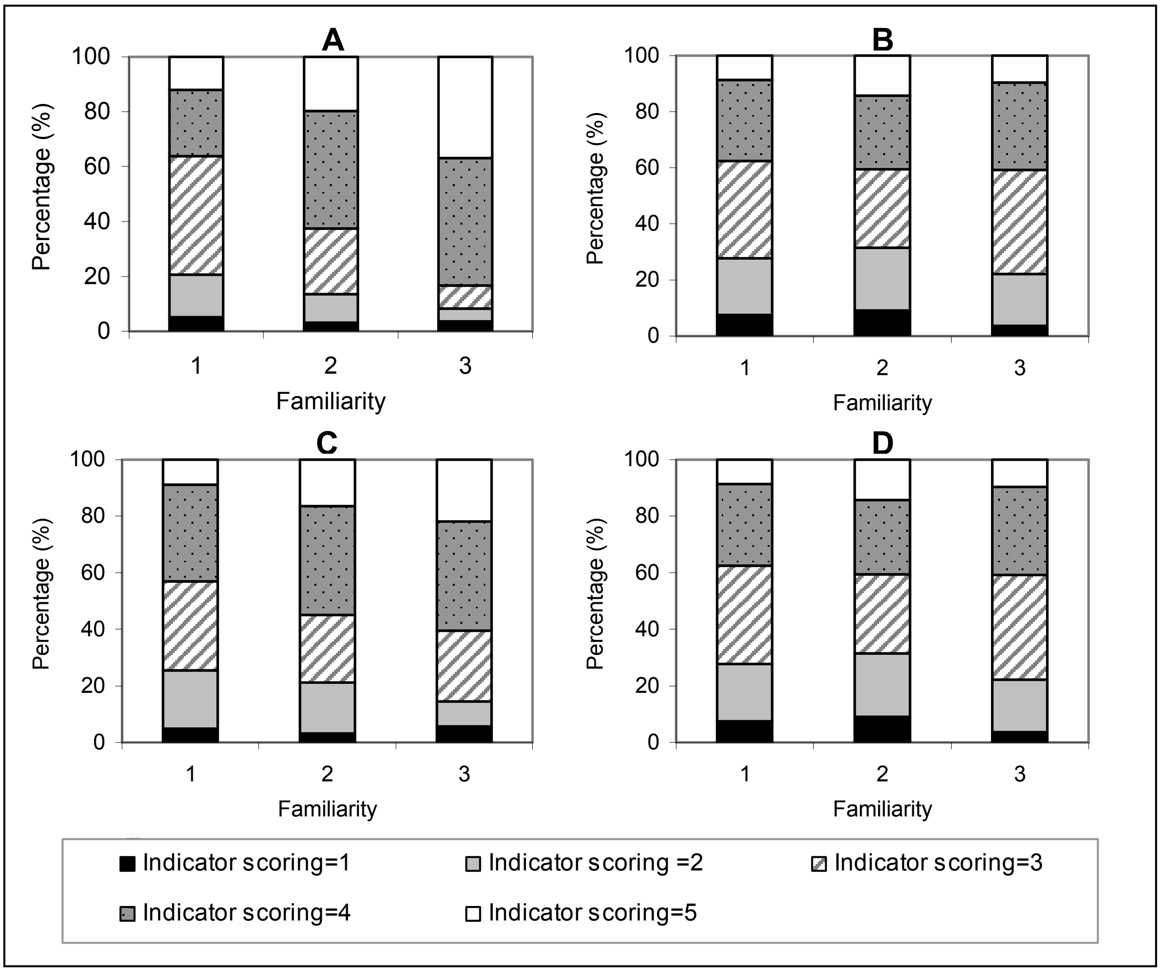



Figure 3.8 Relation indicator scoring versus familiarity

(A: based on indicator scoring quick version simple; B: based on indicator scoring simple version;

C: based on indicator scoring quick version extended; D: based on indicator scoring extended version)

Table 3.7 Results of familiarity scoring (simple). Distinction is made between familiarity scores of all respondents and respondents who filled in both familiarity scoring and the simple version of the indicator scoring.

| **Indicators** | **Familiarity scoring** | |
|---|---|---|
| | All respondents familiarity scoring (simple) (n=18) | Respondents familiarity scoring (simple) and simple version indicator scoring (n=14) |
| **Physical/Chemical** | | |
| Physical environment | 2.06 | 2.00 |
| Chemical environment | 2.00 | 1.93 |
| **Plankton** | | |
| Phytoplankton | 1.61 | 1.57 |
| Zooplankton | 1.67 | 1.57 |
| **Fish** | | |
| Abundance of commercial stocks | 2.67 | 2.64 |
| Abundance of populations that are not regularly assessed | 2.39 | 2.36 |
| Size/Age Structure of a fish species | 2.56 | 2.50 |
| Genetic composition a fish species | 1.39 | 1.43 |
| Size structure of the fish community | 2.56 | 2.57 |
| Species composition including biodiversity of the fish community | 2.39 | 2.36 |
| Abundance of the fish community | 2.50 | 2.50 |
| **Other ecosystem components** | | |
| Status of Marine mammals | 1.89 | 1.86 |
| Status of Seabirds | 1.72 | 1.79 |
| Status of Marine reptiles | 1.35 | 1.46 |
| Status of Benthos | 2.28 | 2.29 |
| Status of sensetive Habitat | 2.11 | 2.07 |
| **Ecosystem** | | |
| Ecosystem functioning including trophic level | 2.06 | 1.93 |
| **Fishing Pressure** | | |
| Fleet capacity | 2.72 | 2.71 |

| | | |
|---|---|---|
| Fishing effort per métier and its spatial and temporal distribution | 2.56 | 2.64 |
| Fishing impact including catch, by-catch and habitat destruction | 2.39 | 2.50 |

Table 3.8 Results of familiarity scoring (extended). Distinction is made between familiarity scores of all respondents and respondents who filled in both familiarity scoring and the extended version of the indicator scoring.

| **Indicators** | | **Familiarity scoring** | |
|---|---|---|---|
| | | All respondents familiarity scoring (extended) (n=18) | Respondents familiarity scoring (extended) and simple version indicator scoring (n=10) |
| **Physical/Chemical** | | | |
| Physical environment | Temperature | 2.22 | 2.30 |
| | NAO | 1.72 | 2.00 |
| Chemical environment | Salinity | 2.17 | 2.20 |
| | Oxygen levels | 2.11 | 2.20 |
| | N and P levels (Eutrophication) | 1.78 | 1.80 |
| **Plankton** | | | |
| Phytoplankton | Primary production | 1.83 | 1.90 |
| | Water transparency | 1.78 | 1.90 |
| | Chlorophyll a | 1.83 | 1.90 |
| Zooplankton | CPR derived plankton indicators | 1.39 | 1.50 |
| | Zooplankton biomass | 1.78 | 1.60 |
| **Fish** | | | |
| Abundance of commercial stocks | Proportion of commercial stocks that are within safe biological limits | 2.50 | 2.60 |
| Abundance of populations that are not regularly assessed | Abundance (numbers) index of selected species (e.g. elasmobranchs) | 2.33 | 2.50 |
| | Biomass index of selected species (e.g. elasmobranchs) | 2.39 | 2.60 |

| | | | |
|---|---|---|---|
| | Decline (threat) indicator | 2.22 | 2.30 |
| Size/Age Structure of a fish species | Average length of selected species | 2.56 | 2.70 |
| | Average weight of selected species | 2.56 | 2.70 |
| | Average age of selected species | 2.50 | 2.60 |
| Genetic composition a fish species | Maturation norm | 1.72 | 1.80 |
| Size structure of the fish community | Mean weight | 2.61 | 2.60 |
| | Mean length | 2.61 | 2.60 |
| | Proportion of large fish | 2.56 | 2.50 |
| Species composition including biodiversity of the fish community | Mean maximum length | 2.56 | 2.70 |
| | Biodiversity indicator (Hill's N0) | 1.83 | 2.20 |
| | Biodiversity indicator (Hill's N1) | 1.89 | 2.20 |
| | Biodiversity indicator (Hill's N2) | 1.89 | 2.20 |
| | Proportion of target species | 2.50 | 2.60 |
| Abundance of the fish community | Total numbers | 2.50 | 2.50 |
| | Total biomass | 2.50 | 2.50 |
| **Other ecosystem components** | | | |
| Status of Marine mammals | Abundance index of selected marine mammal species | 1.78 | 1.80 |
| Status of Seabirds | Abundance index of selected seabirds species | 1.67 | 1.60 |
| Status of Marine reptiles | Abundance index of selected species | 1.50 | 1.50 |
| Status of Benthos | Abundance index of sensitive benthic species | 2.28 | 2.40 |
| | Epibenthos community indicator | 2.06 | 2.10 |
| | Infauna community indicator | 1.89 | 2.00 |
| Status of sensetive Habitat | Area coverage of highly sensitive habitats | 2.17 | 2.10 |

| | | | |
|---|---|---|---|
| **Ecosystem** | | | |
| Ecosystem functioning including trophic level | Ecosystem functioning including trophic level | 2.11 | 2.50 |
| | Primary production required | 2.00 | 2.40 |
| | Catch ratios | 2.33 | 2.30 |
| | Mean transfer efficiency | 1.78 | 2.20 |
| | Trophic level | 2.28 | 2.70 |
| | Fishing in Balance index | 1.72 | 2.10 |
| | Finn Cycling Index | 1.44 | 1.70 |
| **Fishing Pressure** | | | |
| Fleet capacity | Fleet capacity (number of vessels) | 2.78 | 2.80 |
| Fishing effort per métier and its spatial and temporal distribution | Days-at-sea or hours fished per spatial unit (e.g. ICES rectangle) per time | 2.61 | 2.70 |
| Fishing impact including catch, by-catch and habitat destruction | Fishing-induced proportion mortality of commercial fish species | 2.56 | 2.70 |
| | Fishing-induced proportion mortality of non-assessed fish species | 2.28 | 2.50 |
| | Fishing-induced proportion mortality of benthic species | 2.11 | 2.40 |
| | Fishing-induced proportion mortality of marine mammals | 1.94 | 2.20 |
| | Fishing-induced proportion mortality of vulnerable and/or protected species | 2.17 | 2.50 |
| | Proportion of catch discarded | 2.44 | 2.50 |
| | Proportion of area or (sensitive) habitat impacted | 2.33 | 2.40 |

# 4 DISCUSSION

For the process of summarizing the results we explored several different scenarios that would result in different conclusions pertaining to the suitability of the various indicators. As this was intended as a first attempt to explore the Rice & Rochet (2005) framework for the evaluation of this suite of indicators we will discuss here the methodology used and its advantages, disadvantages and possible sources of bias. Although the However, this exercise was limited to input from one interest group, the scientists and therefore a full evaluation of the indicators is still required. This exercise was intended to deliver a final evaluation of the indicators. However, due to the limited participation of stakeholders this was not possible. Despite this limitation, the rankings area intended as a guide to compare the outcomes of the different scenarios and although not intended to suggest that one indicator with a higher ranking is necessarily a better indicator.

This exercise and the subsequent discussion of the outcome raised a number of issues that affect the methodology, potentially causing bias and therefore needing to be resolved before a final evaluation of the indicators can be done.

- The operational objectives need to be known before a proper scoring of the indicators can be conducted
- The relevance of indicators and thus their scoring depends on the management framework. Initially only indicators for the effects of fishing were supposed to be considered. However in many EU waters eutrophication or non-anthropogenic factors like climate change may often be more important than fishing and only if these factors are known is it possible to assess and manage the effects of fishing. Therefore indicators like those for Physical/Chemical were also included. If it is decided that e.g. Physical/Chemical indicators should become part of the framework it needs to be specified how they will be used within this framework. For example: one possible response to the EU being told that temperature has gone up might be to say ‘we will take account of that in our decision taking’ (without specifying how); another response might be that they formally agree in advance that for every 0.1 degree rise in mean annual temperature, they would reduce the possible TAC by e.g. 5%; and yet another might be to ask RACs for their opinion. At present much of this is still unclear.
- The level of expertise of the respondents. Among scientists working in the field of indicators there are differences in the level of expertise and familiarity with certain indicators and it was shown that this determines their scoring. From the feedback we got from non-scientist stakeholders it became clear that most non-scientists do not consider themselves well-enough informed to score the indicators against criteria. It also appeared that among scientist the sharing and making available of information could cause respondents to change their scoring which would often have resulted in a convergence of the scores.
- The scale of application of the indicator needs to be considered when scoring the indicators, e.g. ‘whole RAC area’ is completely impractical for the application of a majority of indicators.
- The choice of and combination of scores against criteria and sub-criteria into a final score is difficult as the criteria and/or sub-criteria are not necessarily independent, nor additive.
- Unknown or irrelevant cells should be indicated beforehand (e.g. by marking them as 0's) as this is often open to interpretation

- There is an inherent difference between the scoring of generic indicators (simple version) and specific indicators (extended version). Some examples:
    - for the generic indicator 'physical environment', there are two specific indicators: 'temperature' and 'NAO'. In terms of concreteness an appropriate score for temperature would be 5, while for the NAO, which is a much more abstract concept, and appropriate score could be 2. Likewise for the criterion 'Public awareness'; the public have heard about temperature, but it is doubtful whether they are familiar with NAO.
    - the generic indicator 'Abundance of Commercial Stocks' sounds very concrete (score 4 or 5), something that the public is aware of (score 4 or 5) and likely to be tightly linked to fishing (score 4 or 5). But the specific indicator 'Proportion of commercial stocks that are within safe biological limits' is based on the outcomes of VPA models (concreteness score 1 or 2). Public awareness might be lower (what is safe biological limits?), and something that may be linked less tightly with fishing activity (score 2 or 3).
    - Within the generic indicator 'Status of marine mammals' there are considerable differences for the scoring against the criterion 'historical data' or 'measurement' if the specific indicator is 'seal population in the Wadden Sea' which would result in a high score while the specific indicator 'North Sea porpoise population' would give a much lower score. So how to score the generic indicator when being aware of these differences between potential specific indicators?
- Different approaches for summarizing the data are shown within this report. We created tables with mean scores and used graphic methods. The graphic methods are informative for the understanding of how scoring processes take place but they can not easily be used for the actual selection of indicators. In principle, this can be done based on the tables providing there is agreement on how the mean scores are determined. Note also that a possible disadvantage of these summarizing methods is that they tend to give similar scores to indicators with similar properties, fostering selection of redundant rather than complementary indicators (Rice & Rochet, 2005).

Some results of the analyses were discussed in more detail as we felt that these revealed some important issues for the process of selecting indicators.

Notably the factor analysis (Figure 3.5) but also the mean scores and their ranks (Tables 3.2 and 3.3) showed that two groups of generic indicators could be clearly separated from the bulk of the indicators:

1. Physical environment, Chemical environment, Phytoplankton and Zooplankton with very low scores on criteria theoretical basis, sensitivity, responsiveness and specificity
2. Status of Marine reptiles and Ecosystem functioning with very low scores on concreteness, public awareness, historical data, measurement and cost.

For the first group of indicators this reveals an inherent problem with these indicators: they are not directly affected by fishing and should therefore arguably not be part of a suite of indicators for the CFP. The reason they were included in the set that was evaluated was because they may be indicative of factors other than fishing (e.g. climate, eutrophication) that may have an impact on the ecosystem features that we are trying to conserve. Thus, they may be helpful in interpreting some of the patterns in the selected indicators and could therefore be useful for management of the fishery. The usefulness of these indicators probably differs between the various EU waters. The objectives should guide the decision on whether or not such indicators should be included.

The scorings of the second group of indicators also demonstrate some of the problems with these indicators. Ecosystem functioning indicators are model-based and hence should have low scores on concreteness. Also the development of most of these ecosystem functioning indicators lagged behind that of most other indicators which translates into a lack of historical data and because of this probably the low scores on measurement and cost. The lack of familiarity (see table 3.7 and 3.8) explains the low score on public awareness. Marine reptiles do not occur in most of the EU waters and even in those waters where they do occur there are no monitoring programs that collect data on them and indicators have not been developed. Therefore this indicator performs similar to the ecosystem functioning indicator. However, surprisingly the marine reptiles indicator also scores very low on concreteness which should not be the case as this describes a very concrete property. The latter may simply be the consequence of the respondents not being very familiar with the indicator.

Following up on the issue of how familiarity may affect the scoring: One detailed analysis of the scorings of a group of related indicators, i.e. the Pressure indicators, reveals some interesting patterns. For each criterion the scorings show similar patterns which are in line with familiarity (table 3.7 and 3.8): fishing capacity>effort>impact. This in spite of the fact that based on a recent study that was not known to most of the respondents (Piet et al. in press) at least for the criteria: theoretical basis, sensitivity and responsiveness the score should be the opposite. This stresses the point that if information is lacking the level of familiarity may largely determine the outcome of the scoring.

## 5 CONCLUSIONS

The results of this exercise and concerns that were put forward in the following discussion can be considered against the background of the evaluation framework developed by Rice & Rochet (2005). We will describe the consequences for each of the eight steps that need to be addressed before a final evaluation can be done:

- Step 1, Determining user needs: In Piet and Pranovi (2006) we found that at present there are no detailed operational objectives within the CFP for anything other than some commercial stocks subject to recovery or management plans. The lack of such operational objectives prevents a proper identification and evaluation of the indicators.
- Step 2, Listing candidate indicators: Too many indicators will aggravate the evaluation process. It is therefore advised to start with a limited suite of indicators. Another point is that for some ecosystem features exist several concrete indicators while for others none. We addressed this by distinguishing two levels of indicators: one generic, the other specific. While this was intended to resolve the discrepancy between the types of indicators available, the feedback of (notably non-scientific) respondents showed that for an evaluation by different stakeholders it is probably better to have them evaluate the generic ecosystem features as the specific indicators are often meaningless and only obfuscates the evaluation. The evaluation and selection of specific indicators within a generic ecosystem feature can be done by scientists who are sufficiently familiar with the merits of each of the different candidates. Even then it is advisable to provide these scientists all the information available to guide the scoring prior to the actual scoring as we observed that the sharing and making available of information could cause respondents to change their scoring which would often result in a convergence of the scores.
- Step 3, Determining screening criteria: The criteria used in this exercise appear appropriate. The scoring against the sub-criteria did not appear to affect or improve the scoring. The assumption is that the number and level of detail in the criteria should balance the level of expertise of the respondents. Adding more (sub)criteria among

which the differences become increasingly more subtle, will hamper the scoring process as soon as these criteria require more expertise than is present within the respondent group.

- Step 4, Scoring indicators against criteria: Using criteria and sub-criteria makes the scoring process more transparent. A direct scoring of the indicators will be affected by differences between respondents who each scored against their own implicit set of criteria. This highlights the issue of expertise mentioned in the previous steps. We observed marked differences in scoring between the respondents which we assume are largely determined by the differences in the level of expertise. Making the relevant information available prior to the scoring and allowing an exchange of viewpoints would considerably reduce the variation and bias in the scoring. If more information is available on the indicators and known and discussed within the group of respondents then more criteria can be used and the scoring exercise is more likely to deliver the best results. With the current level of expertise and information available only a few indicators good be distinguished reliably from a large body of indicators. The assumption is that with increasing level of expertise and information available it should also be possible to further differentiate within this large body of indicators.
- Step 5, Summarizing scoring results: For this another issue is most relevant, i.e. the weighting of the scores of each of the (sub)criteria in order to derive a final score. Allowing a different weighting for the criteria is useful as this allows the respondents to make explicit how they think the criteria should contribute to their score. For all respondents the weightings were markedly different and between respondents the respective weightings also differed considerably. Moreover, based on how the scientists thought the different stakeholder groups would weight the criteria, there may be systematic differences between stakeholder groups that may turn out to be helpful in understanding why different stakeholder groups differ in their preferences for certain indicators.
- Step 6, Deciding how many indicators are needed: Several considerations determine this choice. The first is that we need indicators for both state and pressure. A minimum requirement for the ecosystem state indicators would be that for all relevant ecosystem features represented by a generic indicator at least one specific indicator is selected. Here the relevancy is determined by whether or not that particular ecosystem component occurs in a region (e.g. marine reptiles may not be relevant in all EU waters), to what extent different features of the same ecosystem component are complementary or redundant and if this feature is likely to be affected by fisheries. This minimum selection may be expanded by also including ecosystem features that may affect the core ecosystem features but are not necessarily affected by the fishery. Finally, there is the choice to have more than one specific indicator for one or more of the generic indicators. Again, this should be determined by how much additional information this new specific indicator provides. The considerations in the previous step could be easily translated into suggested approaches to combine indicators such as the ‘hierarchical’ approach or the ‘headline’ indicator approach (Jennings, 2005). In the end, however, the number of indicators that are selected and how they are combined will not only be determined on scientific grounds but also by the requirements of the manager who needs to work with them.
- Step 7, Final selection: For the final selection of indicators the scoring of indicators against screening criteria can offer the information needed to guide the final selection of indicators provided that the shortfalls mentioned previously are resolved. A possible refinement of the approach could be to conduct this in two stages: a first stage where generic indicators are scored against (a subset of) the criteria by different stakeholders

and a second stage where for each generic indicator one or more specific indicators are evaluated against (a more detailed or extended set of) screening criteria by specialists including biologists, ecologists, social scientists and economists.

- Step 8, Reporting: This exercise has not provided any relevant information for the reporting process but this is an important aspect which needs to be addressed.

**REFERENCES**


Piet G and Pranovi F 2006**.** Performance of indicators for ecosystem structure and functioning against screening criteria. INDECO Project Deliverable Numbers 11, 12 and 13. August 2005.

Jennings, S. (2005) Indicators to support an ecosystem approach to fisheries. *Fish and Fisheries*, **6**, 212-32.

Rochet M.J. & Rice J.C, 2005 Do explicit criteria help in selecting indicators for ecosystem-based fisheries management? *ICES Journal of Marine Science*, 62:528-539

Rice J.C. & Rochet M.J., 2005. A framework for selecting a suite of indicators for fisheries management. *ICES Journal of Marine Science*, 62: 516-527.