THIRD WORKSHOP ON APPROPRIATE SAMPLING SCHEMES FOR PROTECTED, ENDANGERED AND THREATENED SPECIES BYCATCH (WKPETSAMP3)

VOLUME 6 | ISSUE 1

ICES SCIENTIFIC REPORTS

RAPPORTS
SCIENTIFIQUES DU CIEM

ICES INTERNATIONAL COUNCIL FOR THE EXPLORATION OF THE SEA
CIEM CONSEIL INTERNATIONAL POUR L'EXPLORATION DE LA MER

International Council for the Exploration of the Sea
Conseil International pour l'Exploration de la Mer

H.C. Andersens Boulevard 44-46
DK-1553 Copenhagen V
Denmark
Telephone (+45) 33 38 67 00
Telefax (+45) 33 93 42 15
www.ices.dk
info@ices.dk

ISSN number: 2618-1371

ICES Scientific Reports

Volume 6 | Issue 1

THIRD WORKSHOP ON APPROPRIATE SAMPLING SCHEMES FOR PROTECTED, ENDANGERED AND THREATENED SPECIES BYCATCH (WKPETSAMP3)

Recommended format for purpose of citation:

ICES. 2024. Third Workshop on appropriate sampling schemes for Protected, Endangered and Threatened Species bycatch (WKPETSAMP3).
ICES Scientific Reports. 6:1. 96 pp. https://doi.org/10.17895/ices.pub.25061522

Editors

Sara Königson • Estanis Mugerza • Katja Ringdahl

Authors

Paolo Carpentieri • Tom Clegg • Bram Couperus • Ruth Fernandez • Gildas Glemarec
Paula Gutierrez Muñoz • Jani Helminen • Ailbhe Kavanagh • Allen Kingston • Sara Königson
David Lusseau • Kim Magnus Bærum • Caterina Maria Fortuna • Estanis Mugerza • Kay Panten
Henrik Pärn • Katja Ringdahl • Torbjörn Säterberg • Gudjon Sigurdsson • Margaret Siple
Rebeka Tetere • Didzis Ustups • Rita Vasconcelos

International Council for the Exploration of the Sea
Conseil International pour l'Exploration de la Mer

Contents

i Executive summary

The workshops WKPETSAMP2 and WKPETSAMP3 were convened following a special request from the European Union's Directorate-General for Environment (DG ENV) on appropriate sampling schemes for endangered, threatened and protected (ETP) species. In particular, these workshops were tasked with providing concrete inputs and results to inform ICES advice to DG ENV on 'appropriate bycatch monitoring systems at Member State level and on regional coordination'.

An aim of the PETSAMP workshops was to generate improved insights into how aspects of sampling design may impact the precision and accuracy of bycatch estimates and the detection probability of bycatch events. The workshop considered key issues such as: how sampling coverage (percentage of monitored fishing operations) impacts the precision of bycatch estimates and how this is dependent on the bycatch probability (how often a bycatch is encountered); if stratification improves precision and if this is dependent on bycatch probability; if it is better to sample few vessels but many trips (e.g. typical of reference fleets and Electronic Monitoring programmes) or many vessels but fewer trips (e.g. typical of at-sea observer programmes). To do this the WKPETSAMP2 extended the simulation framework (SCOTI) developed by WGBYC in 2022. This framework was used in WKPETSAMP3 and was parameterized with data from several case studies. The case studies are from ongoing or historical sampling programs and represent different waters across Europe and different fisheries.

To explore the basic relationship between sampling coverage, bycatch rate precision, and bycatch rate accuracy, a simulation framework was developed based on a hypothetical simplified fishery situation (i.e., homogeneous fishing activity where every fishing operation has the same bycatch probability) and a range of different bycatch rates ranging from relatively frequent (a bycatch event every 10 fishing operations) to very rare (a bycatch event every 10,000 fishing operations). This bycatch rate range reflects real-world situations where highly variable bycatch rates between species are often encountered within a single fishery. The rate at which accuracy and precision improves with increased monitoring as well as the inflection point where the improvement starts to slow down depends on the bycatch probability itself.

The results from the simulations parameterized with case study data show that there is not one monitoring design that is universally better to achieve a precise and accurate BPUE estimate. Overall, in many instances, increasing the number of vessels monitored tends to increase accuracy and precision, even if it means sampling fewer fishing operations per vessel. Stratification is not always a beneficial approach even when fishing characteristics that influence bycatch probability (differences among métiers/gear types) are known. Stratification by métier will yield more precise and accurate BPUE estimates for species with very low bycatch probability. However, improvements in estimation with stratification are more negligible as the bycatch probability increases. Importantly, it appears that stratification is particularly valuable if you have to distribute the total monitoring coverage over fewer vessels and therefore increase the number of fishing operations monitored for each vessel. A reference fleet, given that the reference fleet include the full range of vessel-specific variation that practitioners believe exists, might, compared to observer programmes, be an effective way to collect data on bycatches, especially in cases where refusal rates are high.

The collection of some key biological and ecological parameters is essential for developing models to assess the impact of fisheries on ETP species. The most commonly used biological and demographic parameters (or their proxies) are listed in WKPETSAMP3 report. Some parameters refer to information on the "nature" of bycaught animals (e.g. species, sex, etc.) that need to be

collected at vessel-level by e.g. observers or via electronic monitoring. Quality of the models using these parameters is dependent on the quality of the input data, which thereby carefully should be considered. Considering the difficulty for observers to identify all bycaught specimens to species with certainty, WKPETSAMP3 recommends that bycatch events are documented with photographic evidence. That would also provide additional information for some of the biological parameters.

WKPETSAMP3 also conducted a review of literature.

ii Expert group information

Expert group name	Workshop on appropriate sampling schemes for Protected Endangered and Threatened Species bycatch (WKPETSAMP3)
Expert group cycle	Annual
Year cycle started	2023
Reporting year in cycle	1/1
Chair(s)	Katja Ringdahl, Sweden
	Sara Königson, Sweden
	Estanis Mugerza, Spain
Meeting venue(s) and dates	February 2023 (by correspondence)
	6-10 March 2023, ICES Headquarters, Copenhagen Denmark (ToR B and ToR C)
	13 – 17 November 2023, ICES Headquarters, Copenhagen Denmark

1 Introduction and approach taken by the workshop

The workshops WKPETSAMP2 and WKPETSAMP3 were convened following a special request from the European Union's Directorate-General for Environment (DG ENV) on appropriate sampling schemes for endangered, threatened and protected (ETP) species. In particular, these workshops were asked to provide concrete inputs and results to inform ICES advice to DG ENV on 'appropriate bycatch monitoring systems at Member State level and on regional coordination'. According to the special request any system to monitor incidental bycatch that aims to fulfil legal requirements under the BHD (Birds and Habitats Directives) and the MSFD (Marine Strategy Framework Directive) must provide a satisfactory estimate of the level of incidental capture and killing of protected species, with a high degree of reliability. "It must cover the whole range of fisheries and métiers for which incidental bycatches are known or suspected to occur according to risk assessments, and it must cover a sufficiently large sample of vessels to produce results with sufficient statistical power to permit a reliable assessment of the impact on the conservation status of species (at the level of the local and whole population)."

It is challenging to design and implement data collection programmes that provide satisfactory (howsoever that is defined) estimates of bycatches with a high degree of reliability. The primary reason for this is that bycatch occurrence for many ETP species tends to be rare. This can be a result of a species simply not being particularly susceptible to getting bycaught because of behavioural or ecological reasons, or because a species may have very low abundance and/or a limited overlap with the distribution of a given fishery.

Assuming data collection protocols are appropriate, a single monitoring programme will generate data on a multitude of bycatch species. However, many of these species will have different core distribution areas, abundances, migration patterns and behavioural traits which will lead to high variability in the spatial and temporal patterns of bycatch rates between species. It will also affect the drivers of precision and accuracy in those bycatch rates. If prioritization between species does not exist, it is essentially impossible to optimize the design of a data collection programme to efficiently and cost effectively support management and conservation needs across the full range of potentially affected species without monitoring almost all fishing operations.

The aim of the PETSAMP workshops was to generate improved insights into how aspects of sampling design may impact the precision and accuracy of bycatch estimates and the detection probability of bycatch events. The workshop considered key issues such as: how sampling coverage (percentage of monitored fishing operations) impacts the precision of bycatch estimates and how this is dependent on the bycatch probability (how often a bycatch is encountered); if stratification improves precision and if this is dependent on bycatch probability; if it is better to sample few vessels but many trips (e.g. typical of reference fleets and Electronic Monitoring programmes) or many vessels but fewer trips (e.g. typical of at-sea observer programmes).

To do this the WKPETSAMP2 extended the simulation framework (SCOTI) developed by WGBYC in 2022. This framework was used in WKPETSAMP3 and was parameterized with data from several case studies.

Regarding the part of the DG ENV request on how to translate the simulation framework into concrete inputs to inform ICES advice to DG ENV on 'appropriate bycatch monitoring systems at Member State level and on regional coordination', WKPETSAMP3 discussed a number of aspects. It needs to be stressed that when designing, implementing and developing monitoring programmes it is important to fully understand the objectives of the programme; Will data be used for quantitative mortality assessments? Does that mean that certain demographic data about the bycaught animals should be collected to help improve knowledge of possible population level impacts? What are the expectations from the end-user on a "satisfactory" precision of estimates that will enable robust decision making? When is improved precision no longer important and so increased effort of bycatch monitoring?

To establish efficient and useful bycatch monitoring programmes, it is important to ensure that feedback loops between end-users of data (e.g. management organizations) and monitoring programmes (e.g. scientific institutions) are developed and maintained. Those are crucial to understand how uncertainties, precision and accuracy can be interpreted during the decision-making process in management organizations and how regulatory texts can be interpreted when deciding on precision and accuracy which needs to be tested by scientific institutions (Figure 1).

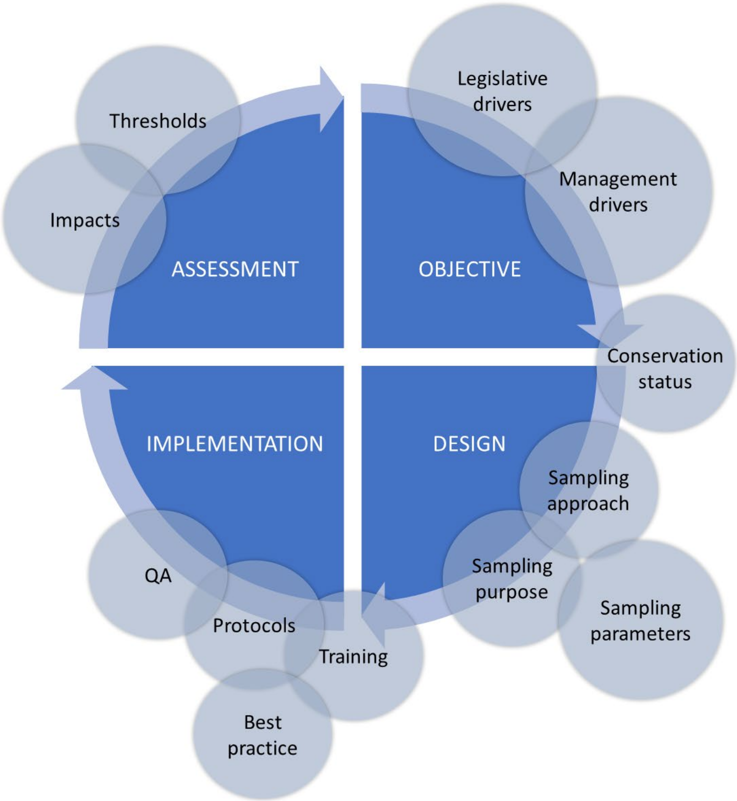


Figure 1 The inter-connected components to consider when designing a bycatch monitoring programme.

2 ToR A: Identify criteria and best practices for designing a multipurpose programme for sampling and estimating bycatch of PETS in order to assess population level impacts

2.1 Summary of Literature review

A literature review was conducted during the WKPETSAMP3 workshop. This is a summary of the literature review and the complete literature review including references are included is presented in Annex 1. The literature review was originally a term of reference for WKPETSAMP2 but was not completed during that workshop due to time restrictions. Subject matter experts, who were workshop participants, were invited to identify potentially relevant reports and publications for consideration. Thirty publications were identified, compiled, and reviewed. A selection of publications referenced in the identified publications and reports were also included in this review. The review process focused on criteria and data quality thresholds, and conclusions are presented in Annex 1.

Key findings from the literature review include recommendations and conclusions on monitoring methods, coverage levels depending on objectives, alternative bycatch risk assessments, and recommendations on how the sampling design could be optimized.

In relation to sampling methods, onboard observers and, to some extent, electronic monitoring are considered the most robust monitoring methods. Therefore, these methodologies should be prioritized for monitoring bycatch whenever possible.

A combination of other monitoring methods has also been used to produce bycatch estimates and collect relevant biological parameters, however with variable reliability. An example of this is the use of self-sampling and dockside collection of bycatch samples by port observers in the Norwegian Reference fleet.

When data on bycatch are lacking and it is not possible to generate quantitative estimates of bycatch mortality, alternative methodologies such as bycatch risk assessments can be used to evaluate the relative risk of bycatch spatially and temporally. Alternative methods to highlight spatial and temporal risk of bycatch include anecdotal information such as interview studies or data collected from strandings. These approaches are appropriate for designing onboard observer programs, and may not be sufficient for determining bycatch mortality.

The coverage needed to detect bycatch depends on the objective of the sampling program, as well as on factors such as bycatch occurrence and heterogeneity. Tools are available to assess observer coverage requirements to document and estimate bycatch with specified criteria and data quality thresholds (e.g. ObsCovgTools; Curtis and Caretta 2020). Studies have suggested that observer coverage levels of 5% may be adequate to collect information identifying bycatch risks and issues in some instances and for some species. However, this level of coverage is likely insufficient for effectively quantifying bycatch of species that are rarely bycaught. This concerns species that are either rare or less susceptible to a given fishery. Higher levels of coverage are likely to be required to obtain a specified level of precision when bycatch events are rare and exhibit highly variable statistical distributions. Observer coverage should aim to be representative, taking into consideration factors such as seasonality of fishing and between-vessel operational variability.

In the case of rare bycatch events/species, multi-annual approaches can be used by pooling a number of years of observation data to achieve more precise bycatch estimates (ICES, 2021, 2022). This assume though that the rate is the same for all pooled years, an assumption which may not always be reasonable. Approaches are developed in ICES WGBYC (2023) to formally assess when such assumptions can be made. Bycatch estimates may also be improved by monitoring a proportion of fishing sets from all fishing trips rather than full coverage from a proportion of all fishing trips (e.g. Peatman *et al*., 2023). Monitoring programs designed for data-poor species should acknowledge that certain species may never have enough data for quantitative assessments. In these cases, there is a trade-off between the costs associated with data collection, the economic value of the fishery and the conservation status of the ETP species/impact of mortality at population level. Adopting a sufficiently precautionary approach may be the only practical means of managing some low-value, data-poor species (Smith *et al*., 2009).

2.2 Overview of common methods to assess human-induced mortality

In most ETP bycatch monitoring programs, the objective is to gather sufficient information to estimate bycatch levels to ensure that they are below a level that will not negatively impact the productivity of the ETP population. There are several approaches for setting reference points for fishery-induced mortality or incidental bycatch (*e.g.* PBR, see Wade, 1998; CLA, see IWC 1991; PST, see Richard & Abraham 2013 and Roberts *et al.* 2019; RLA, see Hammond *et al.* 2019). All of these methods employ some version of population viability analysis (PVA), each with specific data requirements (summarized in Table 1). PVA refers to various theoretical (typically stochastic) demographic models used to understand extinction risk and to forecast future scenarios of population growth and decline for certain species, within a given time frame and under specific external conditions, including the effect of total natural and human-induced mortality on the population's long-term survival.

Reference points, are integral to Bycatch Risk Assessments (BRA). The latter can range from qualitative and semi-quantitative, to fully quantitative approaches. Examples include the ICES WGBYC BEAM (ICES, 2022; 2024), the ByRA (Verutes *et al.* 2020; ICES, 2023), the MSC Productivity Susceptibility Analysis (PSA) (Good *et al*., 2023), the New Zealand spatially-explicit fisheries risk assessment (SEFRA) (Richard & Abraham 2013 and Roberts *et al.* 2019). For all approaches, including PBR or similar, data on the age and sex structure of the population of bycaught animals are important for management advice. For example, if there is a skewed selectivity of reproductively mature females, a PBR-informed threshold may be insufficiently precautionary (Brandon et al., 2017). In addition, although PBR has been extensively tested for marine mammals, its usage for other taxonomic groups could be problematic (*e.g.* Marchowski *et al*., 2020). If PBR is to be applied *sensu stricto* to other taxonomic groups, careful sensitivity analyses based on demographic parameters will be necessary to adequately choose values of R_{max} and F_R. If taking the U.S. Potential Biological Removal (PBR) approach or a modified version of it (for example, with a different conservation objective) PBR is the reference point for minimum sustainable bycatch, and the minimum inputs that need to be collected in an on-board monitoring programme is the total number of bycaught animals per species to compare with PBR (in addition to an estimate of population size). The other components of the PBR formula do not explicitly require observer data: the maximum net productivity rate (R_{max}) can be derived from existing population models and literature (at least for some taxonomic groups, *e.g.* see Dillingham *et al*., 2016), and the recovery factor F_R is obtained from discussions with managers (or in the absence of manager input, different values can be simulation tested).

Table 2.1. Examples of some of the most common threshold-setting methods for ETP bycatch assessment. The conservation objective within each approach can be modified depending along with the recovery time set.

Approach	Example of conservation objective	Removal limit algorithm	Needed input data
Potential Biological Removal (PBR) – original US PBR (Wade 1998)	Maintain population size at or above the Optimum Sustainable Population (OSP) *[50% K 95% of cases]*	N_{min} x $0.5R_{max}$ x F_R	Current minimum population estimate (N_{min}). $N_{min} = \frac{\hat{N}}{exp\left(z\sqrt{ln(1+CV(\hat{N})^2)}\right)}$ Where N=abundance estimate for the population; z is the degree of compensation; CV=Coefficient of variation of the abundance estimate. *Other (even approximated) parameters*: assumed max population growth rate (R_{max}), usually set at 0.04 or at 0.02. Recovery factor related to conservation objectives (F_R) chosen between 0.1 and 1.0.
Catch Limit Algorithm (CLA)	72% K 50% of cases	α x R_{max} (D_T-β) x N_T	Time series of population estimates, including current population size (N_T) Time series of (by)catch data *Other parameters, estimated via simulations*: max population growth rate (R_{max}) current population status (D_T) conservation objectives tuning factors (⍰ and ⍰) RLA: *y*= species specific tuning factor
Removal Limit Algorithm (RLA)	80% K 50% of cases		
Population Sustainability Threshold (PST)	Variable depending on the taxon and management objectives	N x 0.5 x R_{max} x ϕ	Total population size ϕ= user-specified population-based management tuning factor. *Other needed parameters, roughly estimated*: spatially-explicit estimation of annual deaths (D) by SEFRA models (referred to as "Annual Potential Fatalities" or "APF" by previous SEFRA implementations, e.g. Abraham et al. 2017) The risk ratio (R) is then calculated as R=D/PST where R expresses anthropogenic threat-specific deaths (D) as a proportion of the threshold (PST) and is presented as a posterior probability distribution, propagating uncertainty in both D and the PST.
Population Viability Models	All options are open	Typically (but not exclusively) stochastic demographic models	As many demographic parameters as possible. Usually at least: Total population estimate Adult survival Survival of other age classes Sex ratio per age class Age at first reproduction Number of offspring Maximum age

	etc.

2.3 Design of bycatch sampling programmes

An effective design of an ETP monitoring program is dependent on e.g. bycatch probabilities (how often a bycatch appear), objectives, available funds and prior knowledge of fisheries and bycatches. There is not one monitoring design that is optimal in all given situations. This is also why it is important to have feedback loops between users of data and scientists designing and implementing monitoring programme (as well as organizations providing funds for monitoring). It is important to understand if a monitoring programme should meet multiple objectives (such as eg. bycatch rate for several ETP species and/or other species) or a single objective. It might be easier to optimize the monitoring in the latter case.

A designer of a monitoring programme need to consider how large sampling coverage that is needed to meet the objective(s) (if this/these are identified) and how to most effectively use the sampling resources. Implementation of monitoring programmes also often involve logistical constrains such as access to vessels (for different reasons) which also need to be considered. Efficient use of sampling resources involves questions such as choice of monitoring method (e.g. observers or electronic monitoring), choice of selection of PSU (primary sampling units, e.g. vessels) and need for stratification (splitting the PSUs into groups depending on e.g. time, space or métiers). It is however, if prior knowledge on bycatch in a fishery is limited, difficult to understand how much sampling coverage that is needed and how it best should be used. A flowchart with some key elements of design of bycatch monitoring programmes is shown in figure 2. Within WKPETSAMP3 we aim, through simulations parameterized by case studies and by depleting a data rich case study (down-sampling), to generate improved insights in how precision and bias are impacted by different aspects of sampling design (section 3). It is indicated in the figure where insight are achieved from these simulations.

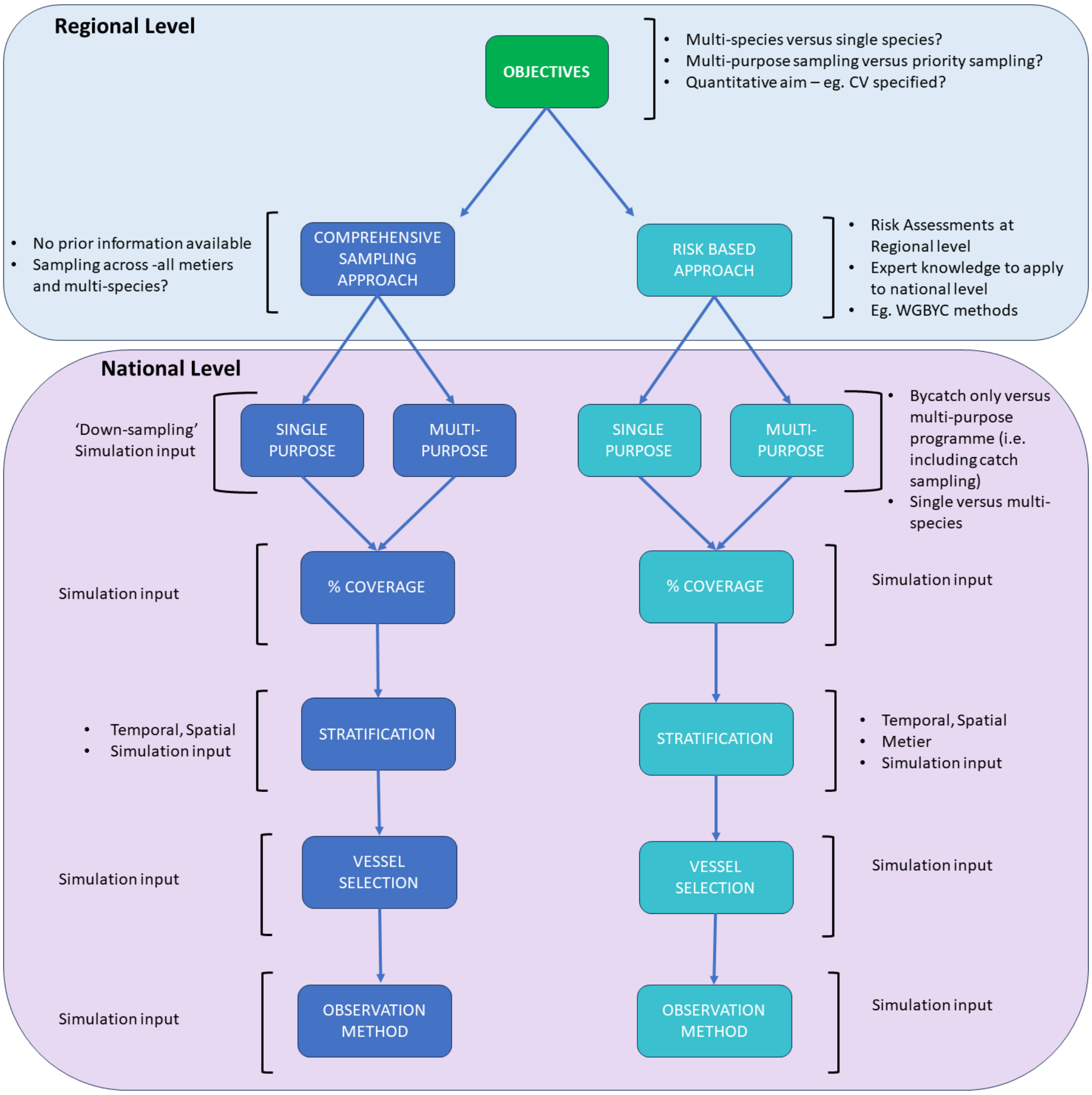


Figure 2 Flowchart for bycatch monitoring design. "Simulation input" and "Down-sampling" indicates that simulations were done by WKPETSAMP.

2.4 Description of monitoring methods

Different methods can be used to collect bycatch data. However, the choice of method used for collecting bycatch data in a particular fishery is based on several factors, including:

- **Completeness**: do the data cover the entire range (temporal, spatial, fleet segments etc.) of the fisheries that interact with the species of concern?
- **Cost**: Is the method cost-effective?
- **Timeliness**: How quickly are the data available to fisheries scientists, managers, industry and other end-users?
- **Safety**: how safe is the data collection method compared with other monitoring methods and what safeguards are in place to ensure the safety of the data collectors?
- **Logistics**: How easily is the programme implemented and maintained?
- **Planned use of data**: do the management goals require a level of detail, quality and timeliness that only certain methods can provide?

- **ETP species conservation status**: do the bycaught species require immediate mitigation actions rather than bycatch monitoring, given their conservation status and existing knowledge of bycatch levels? Is the precision of bycatch rates important given the species' conservation status?

Decisions on the selection of the methodologies to be used needs to be analysed at regional scale and on a case-by-case basis.

2.4.1 Monitoring methods in high-risk fisheries

Independent observations made by trained scientific observers are widely considered to be the most reliable and useful means of collecting data on bycatch. Many at-sea monitoring programmes are undertaken around the world where scientific observers are onboard vessels operating in different fisheries. The objectives of these programmes are diverse, as are the tasks covered by the observers. A specific at-sea scientific observer programme focused only on collecting bycatch data is the optimal way to collect the data required for quantitative bycatch assessments. These programmes can be relatively expensive to implement, especially in situations where incidental bycatch rarely occurs in a sampled haul or trip. Hence, it becomes important to decide when increasing monitoring is necessary or when alternative data/proxies can be sufficient to inform management actions and ensure the conservation of the ETP species. A cost-effective alternative might be well designed multi-purpose programmes with detailed and gear specific data collection protocols that ensure that bycatch events are noticed, recorded and reported and where the absence of bycatch in observed fishing operations is also properly documented, taking into account the best knowledge of ETP species distribution, abundance and conservation status.

Electronic monitoring (EM) is the other preferred method for high-risk fisheries monitoring. EM technology consists of multiple closed-circuit television cameras, a variety of sensors including Global Position System (GPS), winch rotation and hydraulic pressure, all connected to a video and data storage box. EM systems are designed to operate autonomously and continuously while a fishing vessel is at-sea. EM based fisheries monitoring has been carried out over the past decade in a wide range of geographical locations, fishing gears and fishing vessels. In terms of protected species bycatch monitoring, EM can be more cost effective than other monitoring methods, as significant coverage levels can be required for detection of these rare events. However, some issues related to accurate species identification, non-representative data and video review costs are widely recognized but are not insurmountable in the longer term. EM also provides a potential alternative for fisheries and fleets where observer access is limited.

EM programmes are more complex than observer programmes. Essential elements of an EM programme include equipment supply, responsive field services for keeping EM programmes operational on the fleet, data interpretation services for production of standard fishery data from sensor and image data sets, services for consolidating results from EM with other data sources, and an overall management structure to coordinate all elements of the programme. It is important to note that all these elements need to be accounted for when calculating the full costs of electronic monitoring programmes.

2.4.2 Monitoring methods in low-risk fisheries

Monitoring of low-risk fisheries and/or fleets that are difficult to monitor (e.g. small-scale fleets) is associated with a number of issues that need to be considered when selecting the methodology to be used to collect data.

Onboard observers and electronic monitoring programmes are relatively costly way to obtain data from some fisheries (e.g. small-scale fleets), and may still result in limited sampling effort.

In the case of observer programmes, space on-board the vessels (for work tasks or sleeping) is required, which may be challenging on small scale vessels or on large vessels that carry a full crew complement.

In some regions, there is little reliable data available on the actual impact of fisheries on protected species populations, particularly related to bycatch of seabirds and marine mammals in passive gears. This is mainly due to the very high numbers of vessels in these fisheries and comparatively rare and very variable bycatch events, and low sampling levels which make a statistically reliable extrapolation difficult. Vessels using passive gears, also tend to be relatively small. To increase knowledge of bycatch in these type of fisheries other data collection methodologies can be utilized, such as self-sampling, questionnaires/interviews and reference fleets.

Self-sampling

Self-sampling by industry through different reporting systems (specific logbooks, apps), might be considered. However, there is an obvious risk of negative bias as bycatches are usually considered as something detrimental to the industry or bycatch occurrences may simply be unobserved by crew, which usually leads to significant under-reporting. The information provided by this methodology should be validated with EM or observer data when possible or can be used as part of a preliminary screening phase.

Questionnaires/interviews

Individual fishers tend to have useful knowledge of where and when bycatch generally occurs in their own fishing operations, and thus collectively can possess a significant amount of information over a much larger spatial scale. Questionnaires and interviews are a way to access this knowledge. However, the resulting information is not suitable for use in quantitative assessments but can be part of a preliminary screening to highlight possible high-risk areas needing specific attention when designing programmes or can inform validation of standard sampling programmes.

Reference fleet

A reference fleet is a group of fishing vessels that provide detailed information about their fishing activities. The sampling and data collection methods are normally similar to those used by scientific observers but in a reference fleet the sampling is carried out by a trained crew member. Comprehensive training for the crew is important to ensure reliable data are obtained, and is also useful for building trust-based co-operation between fishers and scientists. However, as in any self-sampling scheme, it is important to validate the data by comparing them with alternative sources of information. Examples of a successful sampling scheme based on a reference fleet can be found in Norway.

2.5 Long-term data collection

Monitoring can be defined as an ongoing process of *collecting and analysing* data to track progress towards achieving and/or maintaining specific goals and objectives. In the context of European Union (EU) bycatch policy (which is focused primarily on ensuring the long-term viability of affected populations) this implies a long-term commitment to data collection activities from which fisheries impacts on the current and future status of relevant populations can be judged with reasonable certainty. ETP bycatch monitoring may be necessary to monitor bycatch rates (Bycatch Per Unit Effort (BPUE)) and/or the effectiveness of mitigation measures.

Long-term data collection activities are beneficial because they generate cumulative growth in knowledge that incrementally improves understanding of the finer details of and trends in what are often rare and unpredictable occurrences, and helps facilitate multi-annual analytical approaches that can be used in some situations to increase the data available for assessments of K-selected (large, long-lived, low fecundity) type species that are typical of many ETP species.

Multi-annual analytical approaches have been used for several years to produce bycatch estimates for some ETP species in the United States, Iceland and the United Kingdom, and more recently by the ICES Working Group on Bycatch of Protected Species (WGBYC). Multi-annual approaches can help improve the precision of bycatch estimates and are particularly suited to situations where bycatch rates do not exhibit high inter-annual variability. Long-term data collection combined with multi-annual analytical approaches can also be helpful in improving knowledge in situations where a species exhibits low bycatch rates because of very low abundance or very low susceptibility to capture in fishing gears, and so might not be recorded at all or at useable levels in shorter term studies or within annual analytical approaches.

Bycatch monitoring activities provide data on the frequency, magnitude and spatio-temporal distribution of bycatch events which are used to estimate bycatch rates. Rates can vary widely between species, gear types, seasons and geographical areas. Elucidating those underlying patterns improves the estimation of bycatch rates and the reliability of subsequent mortality assessments and is one of the main objectives of a broadscale long-term bycatch monitoring programme.

In addition to providing data fundamental to bycatch assessments, long-term and suitably designed monitoring programmes can also be used to:

- Explore the operational and/or environmental factors that might be consistently associated with higher/lower bycatch and which can inform the development of mitigation measures.
- Identify areas where bycatch rates or overall mortality is consistently higher and so might form suitable candidate areas for trialling or introducing mitigation measures.
- Inform risk assessment approaches where data availability does not permit reliable quantitative assessments.
- Provide biological data to improve knowledge of species demographics and may indicate species distribution shifts and/or population trajectories for species where abundance estimates are unavailable.
- Provide a useful conduit for knowledge exchange and collaboration between industry and the marine science community.

2.6 Description of case studies

Prior to the workshop it was agreed that some participants would bring data to parameterize simulation models with different case studies representative of European fisheries. Data was brought in a specified format but was not shared in the workshop. The data used followed the data profiling tool (DPT) developed by ICES. This DPT is designed for ICES experts groups as a checklist for those data flows and data products that primarily feed scientific and/or advice outputs through ICES working groups and workshops. The case studies are from ongoing or historical sampling programs and represent different regional waters across Europe and different fisheries. Some monitoring programs have focused on collecting data on rare species and some are multi-purpose programs. A description of the case studies used in the simulations is provided below. The level of detail provided in the description of each case study varies depending on the data policy concerns of the respective institutions that provided the data.

2.6.1 Case study A, Gillnets in Southern European waters

Case study A refers to a fishery/métier that is operated during a particular season of the year by a few vessels (< 5) with a large overall length (>15m). These vessels perform a small number of trips per year (<50 for all vessels combined), and each trip has several days at sea (<5), and several hauls per day at sea (<5) and per trip (<10). These vessels also operate in other métiers throughout the year.

The dataset used in the case study refers to a monitoring programme conducted during a short period (1 year). Data collection was carried out by onboard scientific observers, and there were no refusals from the fleet during the implementation of this sampling programme (refusal rate = 0). There was no stratification in the sampling design of the monitoring programme. Sample selection was performed with the vessels as primary sampling unit, trip as secondary sampling unit and haul as a tertiary sampling unit. The sampling protocol defined that all hauls were sampled during each sampled trip, and it was considered all bycatch events and individuals were detected, and that all individuals were correctly identified.

The fishery/métier was monitored with a low number of monitored trips (<20) but that represented a high coverage rate (50% of vessels, and about 40% of the hauls performed by those vessels in the year).

Monitoring data showed that in this fishery/métier there was bycatch of several species, with different frequency of occurrence in monitored hauls: species A (~50%), species B (~90%), species C (~10%), species D and E (each ~2%). It was considered that the typical number of individuals in a bycatch event was 1 except for species B where it was 2, and that large bycatch events were those bycatch events where the number of individuals were above 3 individuals (species A), 5 (species B) or 2 (species C, D or E) (as defined by the 75% quantile) but with no large bycatch events detected in species D and E.

2.6.2 Case study B

Case study B refers to a fishery/métier that is operated throughout the year but with higher effort during several different seasons of the year. It is performed by a large number of vessels (mean of circa 125 per year) with different lengths overall (including high representation of vessels below and above 15m, i.e., vessels for which electronic logbooks and VMS are mandatory). These vessels perform many trips per year (generally about 5000-15000 for all vessels combined), and each trip has few days at sea (<2) and few hauls per day at sea (mean <2) and per trip (mean <2).

The case study is based on the compilation of data sets from several monitoring programmes implemented over a long time period (15 years, several programmes just one year each, one programme circa 5 years and one programme 15 years). The programmes were based on data collected by scientific onboard observers and/or vessel crew, and there were refusals from the fleet during the implementation of this sampling programme (refusal rate considered as 0.25, but this was based on expert judgement as it was not consistently recorded throughout the study period and/or the several monitoring programmes).

There was no stratification in the data set used in the case study. Sample selection was performed with the vessels as primary sampling unit, trip as secondary sampling unit and haul as a tertiary sampling unit. The sampling protocol defined that all hauls were sampled during each sampled trip, and it was considered all bycatch events and individuals were detected and that all individuals were correctly identified.

In the data set used for the case study, monitoring effort was irregular among monitoring programmes and years. The number of trips monitored per year was circa 20 to 200 depending on the year, the number of vessels monitored was circa 5 to 25 per year which represented a coverage rate of vessels of 5 to 25% per year, and the number of hauls monitored was circa 5 to 150 per year, which represented a coverage rate of hauls of circa <5%.

Monitoring data showed that in this fishery/métier study there was bycatch of several species, but only one ETP species was considered for the present case study, with a frequency of occurrence in monitored hauls of 2%. It was considered that the typical number of individuals in a bycatch event was 1, and that events with more than 3 individuals could be considered as large bycatch events (as defined by the 75% quantile).

2.6.3 Case Study C, Gillnets in the North Atlantic, (Metiers/fleets C5, C6 and C7)

Case studies C, Gillnets in the North Atlantic, represent a fleet of set gillnets, that is composed by different métiers defined at Metier Level 6, named as C5, C6 and C7 with varying mesh sizes operating regularly throughout the year. Overall vessel length varies between 6m and 40m, with a mean length of 10m. Most of the vessels only operate one métier, but the distribution of métiers in terms of effort is not regular (ranging from 0.02 to 0.88 of the vessels using each métier), therefore producing differences in the species-specific bycatch patterns. The total number of vessels operating in the fleet corresponds to the boats registered in the official census in one year. In these case studies the fleet size has been set to 100 vessels. On average, each vessel carries out 2.5 fishing operations per day regardless of the métier they are using.

Bycatch data for each of the species groups is from an onboard monitoring programme, carried out by scientific observers, limited to vessels with a given length. The overall proportion of vessels monitored is around 7%, varying from 2% to 50% between métiers, which is correlated with the number of vessels in each group. Sampling design takes the trip as the primary sampling unit and the haul as the secondary sampling unit. Within a trip, almost all hauls are observed (99%). Sampling protocol was focused on the detection of ETP species bycatch, observing all the phases of the fishing operations.

Metier/fleet C5 aims to analyse the variation in the BPUE (and CV) of a group of cetacean species, with low bycatch probability of 9 bycatch events occurring every 1000 fishing operations (range = 0.003 - 0.026). When a bycatch event occurs, the typical number of animals per event is 1 (as defined by the 75% quantile) and the probability of it involving a large number of animals is low (0.16). The mean number of animals recorded for those large events is 2 individuals.

Metier/fleet C6 aims to analyse the variation in the BPUE (and CV) of a group of seabird species, with a low mean bycatch probability of 2 bycatch events occurring every 100 fishing operations (range = 0.005 - 0.038). Typical number of animals per event is 2 (as defined by the 75% quantile). The probability of a large bycatch event occurring, in terms of number of animals bycaught, is 0.3 and where it does occur, the mean number of animals recorded is 4.

Metier/fleet C7 aims to analyse the variation in the BPUE (and CV) of a group of protected shark species, with a relatively high bycatch probability of 1 bycatch event occurring every 10 fishing operations (range = 0.01 - 0.19). Typical number of animals per event is 6 (as defined by the 75% quantile). The probability of a large bycatch event occurring, in terms of number of animals bycaught, is 0.13 and where it does occur, the mean number of animals recorded is 16.

2.6.4 Case study D, Gillnets in Northern European waters

Case study D refers to a fishery that is operated throughout the year. Vessels can use different gillnet métiers and have variable effort depending on the target species and fishing season. The vessels used are mainly less than 12 meters. They normally go out over a single day and come back to the same harbour. The case study is based on two sampling methods (onboard observers and electronic monitoring) and data have been collected over several years (2017-2023). In total, 63 vessels have had either an onboard observer or cameras of their fishing activity. The objective of the sampling program has been to monitor ETP species.

The sampling program has been spatially stratified based on the abundance of the most protective specie. Different sampling effort was allocated (vessels have been allocated to the area were most of its fishing activity takes place) to five areas where the risk of bycatch of concerned species is assumed to be different. Therefore, the sampling coverage varies in the stratified areas, with the maximum coverage in an area as high as 24 %. Sampling coverage of the entire fishery was 7% in 2022.

Sample selection was performed with the vessels as primary sampling unit. The sampling protocol defined that all fishing operations were sampled during each sampled trip. All bycatch events and individuals were considered detected. The first three years on average 30 trips were monitored per year. The following 4 years on average 265 trips were monitored per year.

Since all species were collected there are bycatch rates for several rare species however for the simulation exercise they are grouped as cetaceans or seabirds These two taxa have different distributions and thereby different bycatch characteristics. The bycatch rate of a species defined to be a rare species (cetaceans) was from 0.00 to 0.011 individuals per trip. Seabirds exhibited higher bycatch rates of 0.00 to 0.073 individuals per trip. The bycatch rates were different depending on the area of stratification.

2.6.5 Case study E, Mid-water pair-trawlers in Adriatic waters

We used bycatch data collected by onboard observers on pair trawlers in the northern Adriatic Sea 2006-2013 as implementation of Regulation EC 812/2004 (Fortuna et al. 2010). The fishing operations in the area are carried out in shallow waters (20-40 m). In total, over 9,000 hauls on 21 different vessels were monitored. The monitoring coverage (proportion of hauls monitored) ranged between 2.4 and 10.4% among the study years, with a mean of about 5%.

Italian mid-water trawlers, called '*volanti*', operate in pairs (see Table 2 for details on this métier). Pair trawlers tow a net about 150m long, with a mouth opening of about 15-18m width and 6-10m height, to target small pelagic fish. The speed and relative distance of the boats, and the size and depth of the net mouth can be altered during the towing. Hauls generally last on average 30-45 min. However, given the relatively shallow waters (20-40 m) of the northern Adriatic Sea and the net dimensions, these midwater pair trawlers can catch species that are found in various parts of the water column from the seabed to the surface. This causes a multispecies (often unwanted) bycatch, which includes dolphins, loggerhead turtles, HD bony fish, sharks and rays (Fortuna et al. 2010; La Mesa et al. 2015, 2016; Bonanomi et al. 2018; Pulcinella et al. 2019). Because of the short duration of tows, the post-release survival might be not too bad for some of these taxa. Hence, this case study is an example of multispecies bycatch fishery, using relatively coastal and offshore areas but always operating in shallow waters. Four species were selected for various representativeness reasons linked to their observed bycatch rates.

The focal species in this simulation are: *Tursiops truncatus* (Common bottlenose dolphin) - a cetacean species (listed in Annex IV of the HD) relatively abundant in the area but rarely bycaught; *Caretta caretta* (Loggerhead turtle) - a marine reptile species (listed in Annex IV of the HD) abundant in the area and bycaught twenty times more often than *T. truncatus*; and *Myliobatis aquila* (Common eagle ray) - a pelagic ray species (considered vulnerable in the Mediterranean Sea and usually released when bycaught) of unknown density in the region - and *Squalus acanthias* (Spiny dogfish) - a shark (considered endangered in the Mediterranean, but subject to exploitation and listed in Appendix II of the CMS) - which are both bycaught about two hundred times more often than *T. truncatus*.

Table 2.2 Mid-water pair trawlers operating in northern Adriatic Sea (registered in Veneto, Emilia Romagna and Friuli-Venezia Giulia)

<table>
<tr><td rowspan="8">Fleet description</td><td colspan="2">Total number of pairs</td><td colspan="2">Annual average fishing days per pair</td></tr>
<tr><td colspan="2">84</td><td colspan="2">120-160 (depending on the year)</td></tr>
<tr><td colspan="2">Vessel length</td><td colspan="2">Nominal power</td></tr>
<tr><td colspan="2">>18m</td><td colspan="2">150 and 900kW</td></tr>
<tr><td colspan="2">Fleet average annual fishing days (pairs)</td><td colspan="2">Fleet annual hauls (pairs)</td></tr>
<tr><td colspan="2">5,500</td><td colspan="2">24,000</td></tr>
<tr><td colspan="2">Average hauls duration</td><td colspan="2">Target species</td></tr>
<tr><td colspan="2">30-45 min</td><td colspan="2">• Anchovy (Engraulis encrasicolus): about 70% of catch biomass and 80% of individuals.
• European sardine (Sardina pilchardus): totalling about 20% of biomass.</td></tr>
<tr><td rowspan="5">Bycaught species</td><td>Species</td><td>Average bycatch rate</td><td>Species relative density</td><td>Conservation / legal status</td></tr>
<tr><td>Tursiops truncatus</td><td>0.0012</td><td>0.057 (ind/km²)</td><td>LC, HD Ann IV</td></tr>
<tr><td>Caretta caretta</td><td>0.0245</td><td>0.405 (ind/km²)</td><td>LC, HD Ann IV</td></tr>
<tr><td>Myliobatis aquila</td><td>0.1969</td><td>NA</td><td>VU</td></tr>
<tr><td>Squalus acanthias</td><td>0.2219</td><td>NA</td><td>EN, CMS App II</td></tr>
</table>

2.6.6 Case study F, Longline fisheries in Norwegian waters (Norwegian Reference fleet)

The Norwegian case study uses sampling of bycatch of northern fulmar (*Fulmaris glacialis*) by the Norwegian Reference Fleet (NRF) sampling programme in the offshore demersal longline fishery between 2019 and 2021. The fishery consists of 31 vessels targeting cod (*Gadus morhua*), haddock (*Melanogrammus aeglefinus*), ling (*Molva molva*), and tusk (*Brosme brosme*), following seasonal trends in species distribution and quota allocations throughout Norwegian and Russian waters. Vessels use a range of bycatch mitigation devices to reduce incidental mortality of seabirds. This includes bird scaring lines, laser deterrents, turning off lights when setting at night. Some modern vessels fitted are fitted with moonpools through which lines are hauled to reduce exposure.

The NRF are a group of active fishing vessels (n=3) tasked with self-sampling catches and fishing activity. Vessels are selected using an expert judgement sample through an open tender process which aims to recruit ‘typical’ vessels in the priority fisheries. An agreement between fishers, scientists, and the Norwegian authorities ensures that data shall not be requested or used for control or enforcement purposes. This agreement allows the NRF to record bycatches without fear of prosecution.

Sampling follows a one-stage cluster sampling design in which vessels report total seabird bycatches for each fishing day. Seabird bycatches were almost exclusively northern fulmar, with negligible bycatches of seven other seabird species.

Over the three-year study period, the three NRF vessels fished a total of between 330 and 706 fishing days per year, which equates to 6-12% of total annual fishing days in the fishery.

Note: this reference fleet case study is also relevant to electronic monitoring programmes that have similar sampling design (i.e., repeated sampling of a small group of voluntarily participating vessels which aim to be representative of the wider fishery).

3 Using simulations (SCOTI) and case-studies to highlight aspects to consider when designing a bycatch monitoring programme

The simulations explored during WKPETSAMP3 extended the work carried out in WKPETSAMP2. The simulation platform has been described in detail in the WKPETSAMP2 report and we refer the readers to this previous report for details. We also provide in Annex 2 a description of how to use the open access code used to run these simulations (available at https://www.github.com/dlusseau/scoti). A function wrapper is provided to replicate the approaches used here and can be used by any interested party to ask similar questions and tune simulations to meet their own interests. The authors of SCOTI are available for queries on how to best approach these simulations while SCOTI continues to move through testing and release phases.

3.1 Methods

We first assessed how changes in monitoring coverage affects the precision of Bycatch Rate per Unit Effort (BPUE) estimates under generic fishing conditions. We then tuned fishing simulations to match conditions in existing fisheries (i.e., using the case studies described above) and assessed how varying monitoring coverage affected BPUE precision.

To complement these simulations, we also used a rich bycatch monitoring dataset (Case study E, Mid-water pair-trawlers in the Adriatic Sea) to assess how down-sampling of bycatch monitoring samples affected both the precision of BPUE for multiple species with varying bycatch rates and the ability to detect bycatch for a species which is caught very rarely.

Finally, we assessed how sampling stratification might affect BPUE estimates precision by carrying out a detailed assessment of two case studies. Data from Case study F (NRF) were used to assess how heterogeneity in spatial and temporal overlap can affect BPUE estimates emerging from monitoring schemes which may be unaware of these heterogeneities. Then heterogeneities in bycatch probability between multiple métiers from case study C (Gillnets in the North Atlantic), were used to determine whether stratification by métier provides more precise BPUE estimates for given monitoring coverage rates.

The first step in these simulations consisted of characterizing a series of parameters describing the fisheries in each case study. These parameters are inputs in the simulation study and are derived from the datasets presented by the case studies described above. The following parameters were considered:

- Fishing parameters: number of vessels in the fleet (nboat), mean number of fishing events (e.g. hauls) per day at sea (mean.fishing.event.boat.day), and if this is homogeneous or heterogeneous across vessels (stochastic), proportion of vessels belonging to each métier (i.e. here a set of fishing activities with similar catch and bycatch patterns) (p.métier);
- Bycatch parameters: probability of bycatch event per fishing event (e.g.haul) (p.bycatch), mean number of individuals in a regular bycatch event (mean.bycatch.event), probability of a given bycatch event being large (p.large.event), mean number of individuals in a large bycatch event (mean.bycatch.large.event);
- The number of simulations: number of fishing events (e.g. hauls) to be used in the simulations (nsample);

- Monitoring parameters: proportion of vessels monitored (p_monitor_boat), and proportion of fishing events (e.g. hauls) monitored for each monitored vessel (throughout the regular fishing activity of those vessels during a year) (pmonitor);
- Sampling design parameters: whether sampling is stratified by métier (bymétier), proportion of monitoring allocated to each métier (p_monitor_métier), (strat_vessel), and finally if there is direct selection of hauls for monitoring or if there is first a selection and then boat (boat_samp);
- Vessel/captain property parameters: probability that a vessel selected for sampling refuses to accept observers onboard (refusal_rate);
- Monitoring event property parameters: probability that a haul is observed during a monitored trip (p_haul_obs), probability of detection of each individual in a bycatch event (detect_prob), probability of misidentification of the bycaught species (misclassification).

The characterization of these parameters is explicitly necessary for the implementation of the simulation study.

In addition, it was shown that these key features/parameters affect the outcome in terms of the CV of BPUE (CV_{BPUE}) that can be obtained from a given sampling design (WGBYC 2022). Therefore, the characterization of the key features/parameters is very relevant for the definition of the sampling design and effort of a programme, regardless of the approach taken, either in a simulation approach such as the present one where the parameters are explicitly used; or a statistical approach (based on effort or variability); or in an approach based on fixed objectives. The characterization of some of these parameters requires different levels of knowledge of fishing activity of the case study considered and on bycatch. In the case studies when there is insufficient information to characterize these parameters, then a hypothetical range of values for a parameter could be hypothesized based on other cases with similar characteristics, or it may be relevant to acquire some information that allows characterization of the parameters. Therefore, this implies an exercise of careful consideration of specific characteristics of the fishery and of bycatch events (and in the case of several species this needs to be done separately). This results in better knowledge of the fisheries and characteristics of bycatch events, but also in increased awareness of the set of features that affect CV_{BPUE} and the accuracy of BPUE that should be taken into consideration when designing a sampling programme.

Simulation scenario outcomes

3.2 Effect of monitoring coverage on precision and accuracy of bycatch Rates

It is generally considered that historical and current monitoring levels in European fisheries are insufficient to provide statistically reliable information on bycatch rates for many ETP species and métier combinations (WGBYC 2022). To explore the basic relationship between sampling coverage, bycatch rate precision, and bycatch rate accuracy, a simulation framework was developed based on a hypothetical simplified fishery situation (i.e., homogeneous fishing activity where every fishing operation has the same bycatch probability) and a range of different bycatch rates ranging from relatively frequent (a bycatch event every 10 fishing operations) to very rare (a bycatch event every 10,000 fishing operations). This bycatch rate range reflects real-world situations where highly variable bycatch rates between species are often encountered within a single fishery.

If we use the analogy of looking for needles in a haystack, where the haystack is the fishery, and the needles are bycatch events, as the number of needles (frequency of bycatch) increases so does the probability of encountering them with different levels of "looking." This means that the bycatch probability interacts with the proportion of monitored fishing operations and this interaction affects both the precision and accuracy of resulting bycatch rate (BPUE) estimates. If we randomly sample the fishing operations, the accuracy and precision of the BPUE changes in a predictable way as we increase/decrease the proportion of monitored fishing events. However, the rate at which accuracy and precision improves with increased monitoring depends on the bycatch probability itself. This is illustrated in figures 4 and 5.

Raw outputs from the simulations are simulated bias and CV. The results shown in this section's figures (Figure 3 – Figure 16) are predictions from a generalized linear model in which the response variables (simulated bias and CV) are associated to the interaction between the explanatory variables (proportion of events monitored and the probability of bycatch), assuming a Gamma distribution of residuals

$$bias_i \sim \mu_{bias} + \varepsilon_i, \qquad \varepsilon_i \in \Gamma(k, \Theta)$$

$$\frac{1}{\mu_{bias}} \sim \beta_0 + \beta_1 p_{monitor_i} + \beta_2 p_{bycatch_i} + \beta_3 p_{monitor_i} : p_{bycatch_i}$$

Where i∈[1,100nm] and n is the number of $p_{monitor}$ values for which bias was simulated and m is the number of $p_{bycatch}$ values for which bias was simulated. The same modelling approach was applied to simulated CV.

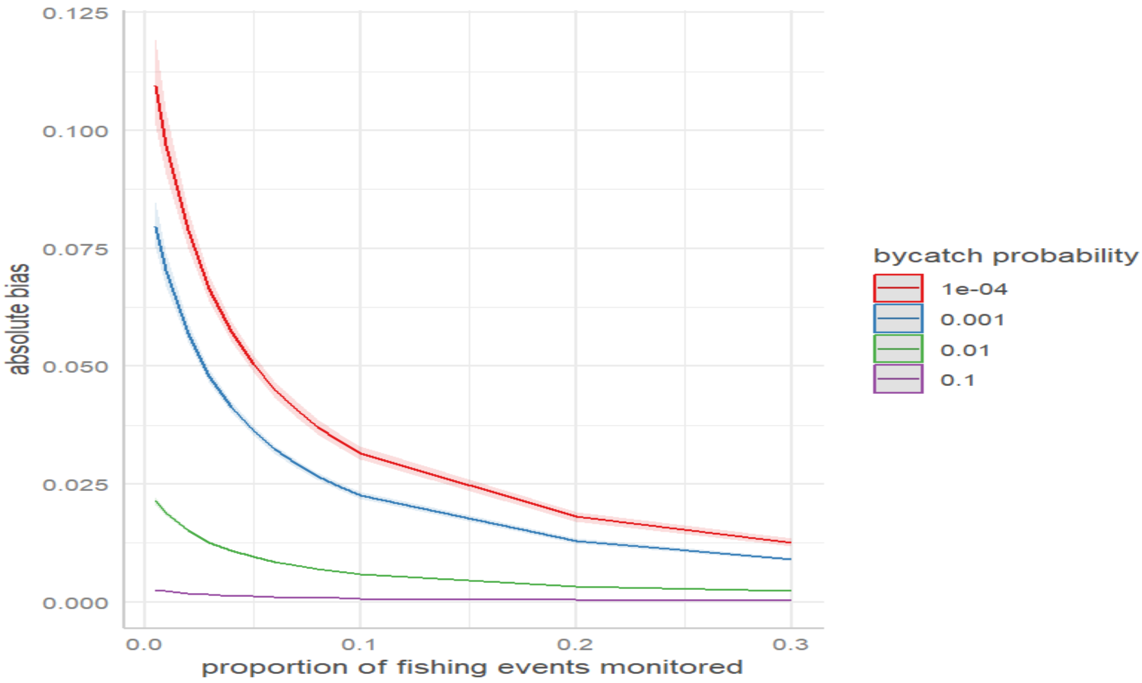


Figure 3 Predicted BPUE bias (absolute value) estimated by applying an unstratified randomized sampling design from 0.5% to 30% of fishing operations from simulated fisheries where bycatch probability is homogeneous across all fishing operations and ranges from 0.0001 to 0,1 (only 4 examples are shown). Shaded bands are 95% confidence intervals.

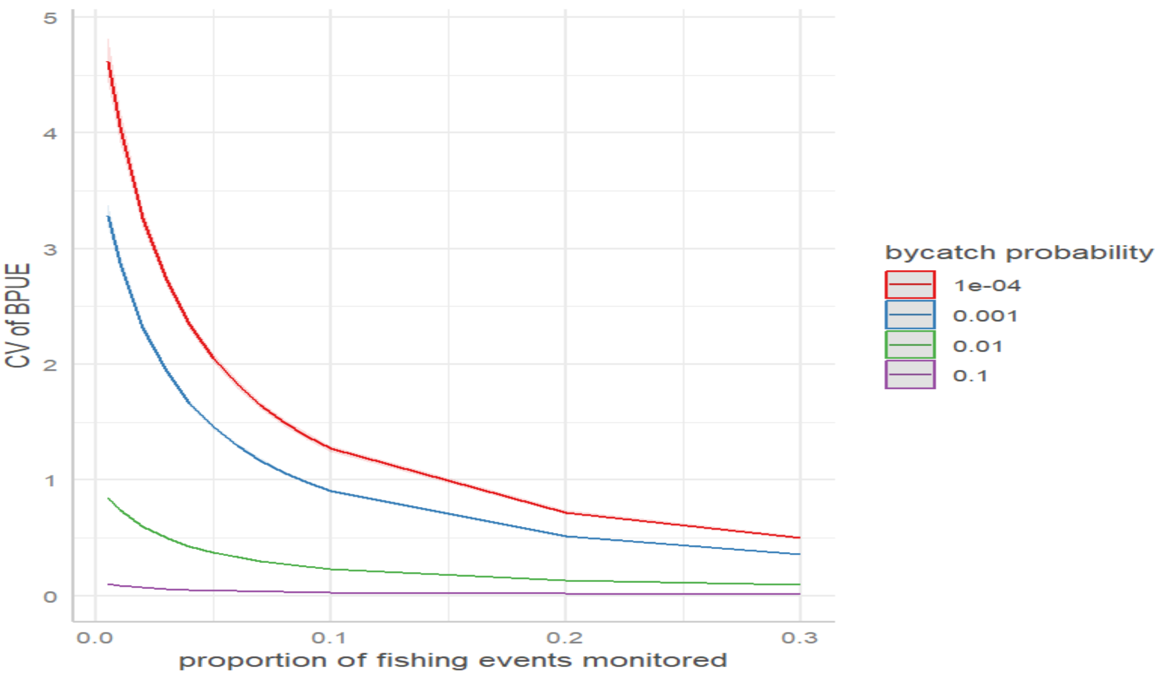


Figure 4 Predicted CV_{BPUE} estimated by applying an unstratified randomized sampling design from 0.5% to 30% of fishing operations from simulated fisheries where bycatch probability is homogeneous across all fishing operations and ranges from 0.0001 to 0.1 (4 examples are shown). Shaded bands are 95% confidence intervals.

In cases of very high bycatch frequency (1 bycaught individual per 10 fishing operations, purple line in Figure 3 and Figure 4) little statistical benefit (reduced bias and uncertainty) is obtained by increasing coverage levels beyond the simulation starting with coverage level of 0.5 %. As the bycatch frequency decreases, increasing coverage levels reduces bias and uncertainty quite rapidly but then arrives at an inflection point, which indicates a slowing, and then eventual levelling off in the statistical benefit associated with increased monitoring effort. Note though that the achieved CV and bias at those 'plateaus' increase as the bycatch probability decreases.

For medium/high bycatch frequency situations (1 bycaught individual per 100 fishing operations, green line in figure 4 and 5) the inflection point occurs at approximately 3% monitoring coverage, but approximately 7% coverage would be required to reduce CVs to previously proposed target levels (e.g. CV of 0.3 from EU Regulation 812/2004).

For very rare bycatch situations (1 bycaught individual per 1000 to 1 bycatch per 10,000 fishing operations, blue and red lines in figure 4 and 5) the inflection point occurs between 5% to 8% monitoring coverage but circa 30% coverage would be required to reduce CVs to previously proposed target levels.

In ultra-rare bycatch frequency situations (e.g. 1 bycatch event per 10,000+ fishing operations), which are not represented in Figure 3 and Figure 4 , it is likely that reliable (low bias, low uncertainty) bycatch rates would be unachievable with any level of monitoring effort simply because of the statistical properties of bias and uncertainty associated with extremely rare events. However, in these cases it would be more important to understand the causes of such low bycatch rate (i.e., extremely rare species/occasional visitors or extremely small and threatened population or abundant species not susceptible to bycatch?).

This simulation indicates that if monitoring coverage was implemented based only on achieving reliable rates for higher frequency situations, it is very likely that reliable bycatch rate estimates would not be achievable, at least in the short term, for the less frequently bycaught species. However, multi-annual analytical approaches may be appropriate for improving bycatch rate estimation for less frequently bycaught species in the medium to long term (see section 2.5). Conversely, if coverage levels were targeted at producing reliable rates for less frequently bycaught species it is likely that significant oversampling (i.e., additional effort with little statistical benefit) would be undertaken for the more frequently encountered species.

In the context of a multi-purpose multi-métier monitoring programme this simple simulation exercise has highlighted that important sampling design choices would need to be made on how monitoring efforts might best be allocated within and between métiers with different bycatch risk profiles. This is explored in more detail in the following detailed case study simulations.

3.2.1 General guidance for allocating monitoring coverage

Monitoring sampling designs are first faced with decisions on how a given desired monitoring coverage (proportion of fishing operations for a particular fishery) should be apportioned between vessels. Sampling designers can try to maximize the number of vessels sampled but sample fewer fishing operations per vessel or they might choose to focus on a few vessels (e.g. if they plan to use electronic monitoring) but sample more fishing operations per vessel.

In practice, sampling designers are often faced by constraints associated with non-responses or refusals to participate in monitoring which will constrain their sampling scheme. In such instances, it may be useful to know what the consequences of sampling fewer vessels than desired might be on the precision and accuracy of the BPUE estimates. To explore these consequences, we tuned SCOTI's fishing simulations to real fisheries case studies and applied monitoring to many (100 or more in some instances) replicates of each tuned fishing conditions. Tuning was based on observer data available for the case studies as well as expert knowledge of these fisheries. Total monitoring coverage was varied from 0.1% to 20% (except for case study A which focused on a small fishery only composed on 4 vessels and therefore coverage varied from 1% to 25%). Then conditions were set so that these proportions of monitored fishing operations were randomly distributed (evenly) across 1%, 5%, 10%, 25%, and 30% of vessels. This resulted in 3900 samples of BPUE precision and accuracy (estimated from applying the selected monitoring scheme 1000 times for each of the 100 replicate simulations).

We could then assess whether precision and accuracy varied with monitoring coverage and whether this effect of monitoring coverage depended on the proportion of vessels over which the monitoring effort was distributed. We tested this by fitting generalized linear models with an assumption of Gamma-distributed residuals (see Annex 2) and using model selection to assess the effects of which explanatory variables (proportion of fishing operations monitored, proportion of vessels monitored, and their interaction) was best supported by the simulation outcomes. We find that in all instances monitoring coverage effect on precision was large, as expected, but depended on its distribution over vessels and this interaction had varying effect size for the different case studies (Figure 5 – Figure 7). Overall, the precision of BPUE increases as more vessels are involved in the monitoring scheme (and therefore have fewer fishing operations monitored for each vessel) and this effect decreases as we increase the monitoring coverage (the proportion of fishing operations monitored).

There are, however, notable and important exceptions to this general outcome. For example, in Case Study A, which has few vessels, the gain over the proportion of vessels monitored (which varied from 25%, to 50%, to 75% to 100%) is negligible and indeed changes direction as the monitoring coverage increases. The shift in direction of the effect occurred between 1 and 2% coverage for species that are caught more regularly (higher bycatch probability) and around 5% for the other 3 species which are bycaught less often (Figure 7). So, in the context of case study A, it is not always advantageous to increase the number of vessels sampled for a given amount of monitoring effort available but it is beneficial to increase the number of fishing operations sampled over fewer vessels instead.

The case study B and case study C show the general pattern described above, with the former having a strong diminishing return of vessels assignment as monitoring coverage increases and the latter having the strongest effect of monitoring effort distribution among vessels (Figure 5).

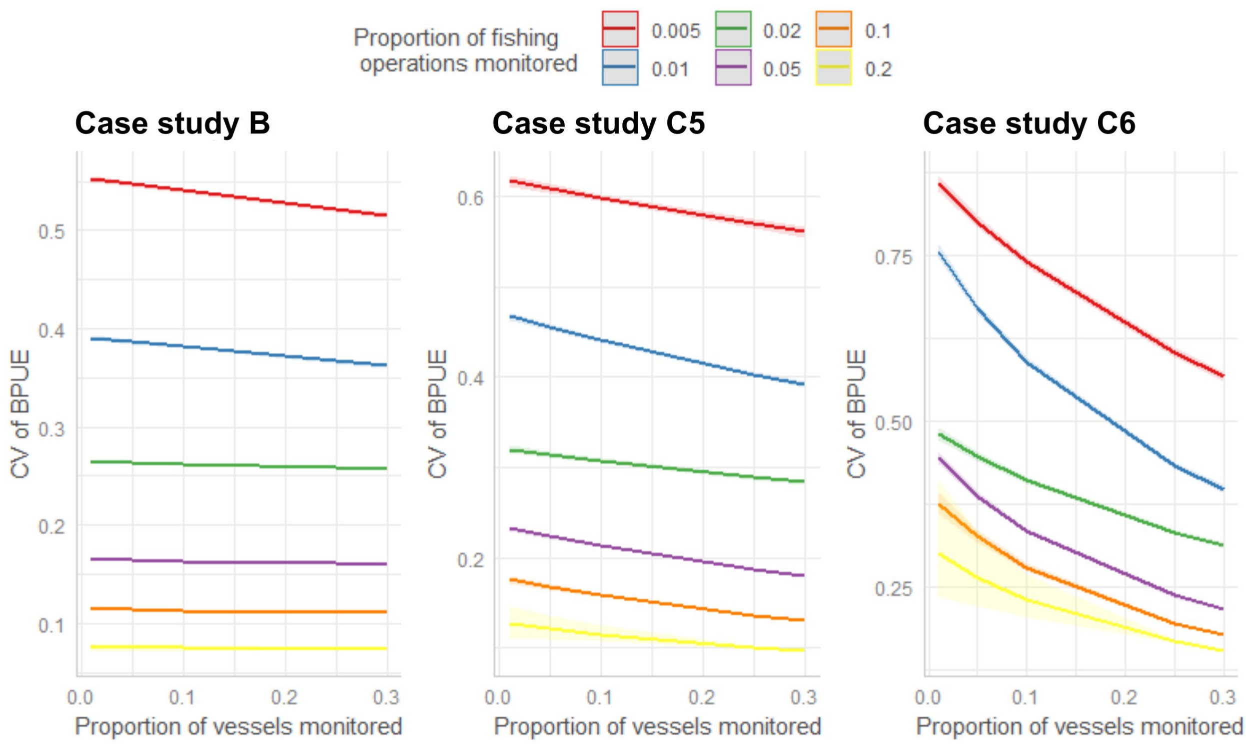


Figure 5 predicted CV_{BPUE} estimated by applying a randomized sampling design sampling from 0.5% to 20% of fishing events from simulated fisheries from Case Study B and Case Studie C (C5, C6), over proportion of vessels monitored. Lines

show predictions from a generalized linear model in which both bias and CV response variables are associated to the interaction between proportion of fishing operations monitored and the proportion of vessels monitored explanatory variables assuming a Gamma distribution of residuals. Shaded bands are 95% confidence intervals.

In case study D, Gillnets in North waters, we observed both variability in the effects between areas and between species (Figure 6), both indicated again an interaction with bycatch probability (or features affecting it).

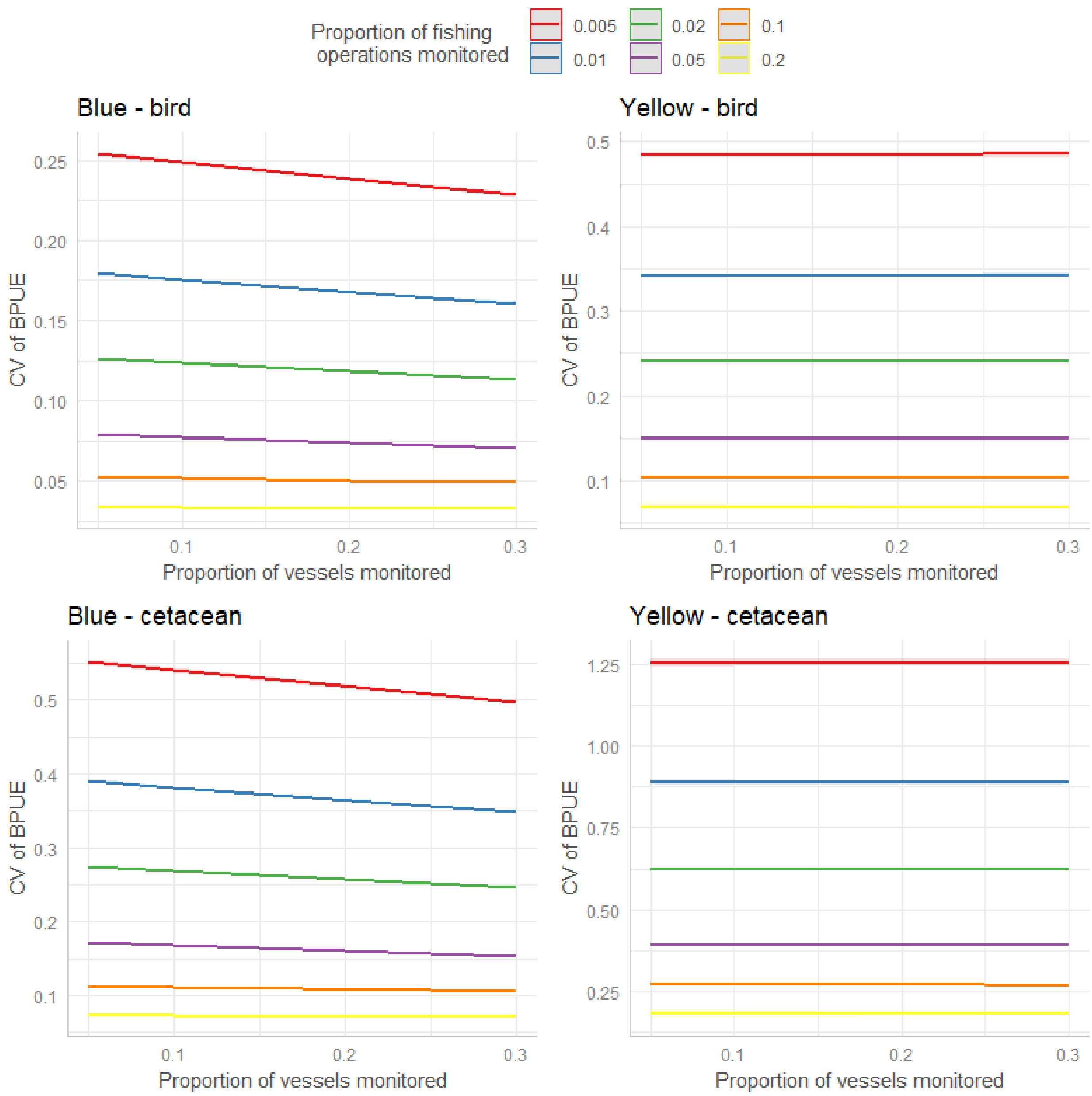


Figure 6 Predicted CVBPUE estimated by applying a randomized sampling design sampling from 0.5% to 20% of fishing events from simulated fisheries from the two regions (blue and yellow) in case study D distributed over varying proportion of vessels monitored. Prediction from a generalised linear model in which both response variables are associated to the interaction between proportion of fishing operations monitored and the proportion of vessels monitored explanatory variables assuming a Gamma distribution of residuals. Shaded bands are 95% confidence intervals.

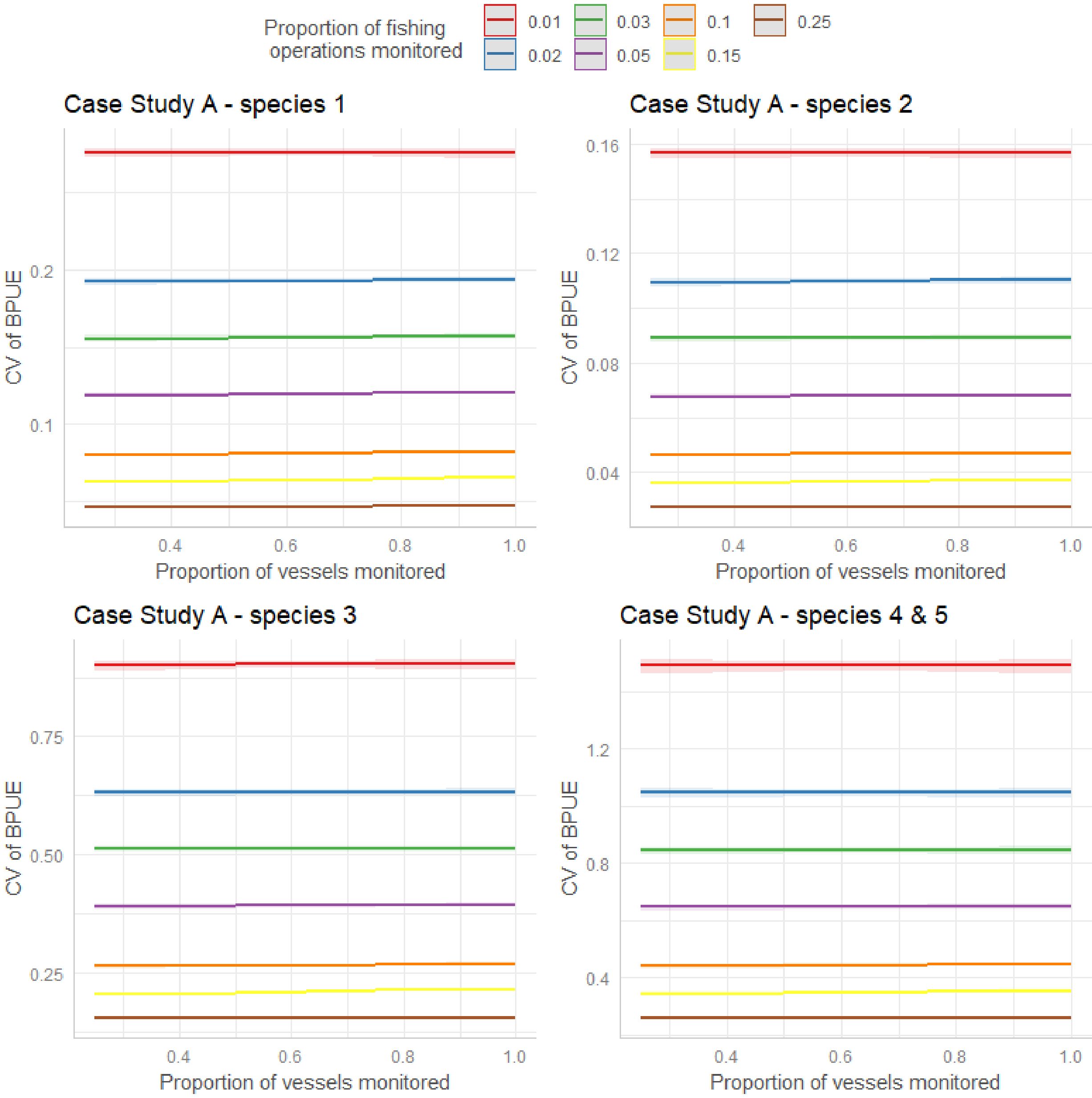


Figure 7 Predicted CVBPUE estimated by applying a randomized sampling design sampling from 0.5% to 20% of fishing events from simulated fisheries from Case Study A distributed over varying proportion of vessels monitored. Prediction from a generalised linear model in which both response variables are associated to the interaction between proportion of fishing operations monitored and the proportion of vessels monitored explanatory variables assuming a Gamma distribution of residuals. Shaded bands are 95% confidence intervals.

3.2.2 Stratification of monitoring by métier

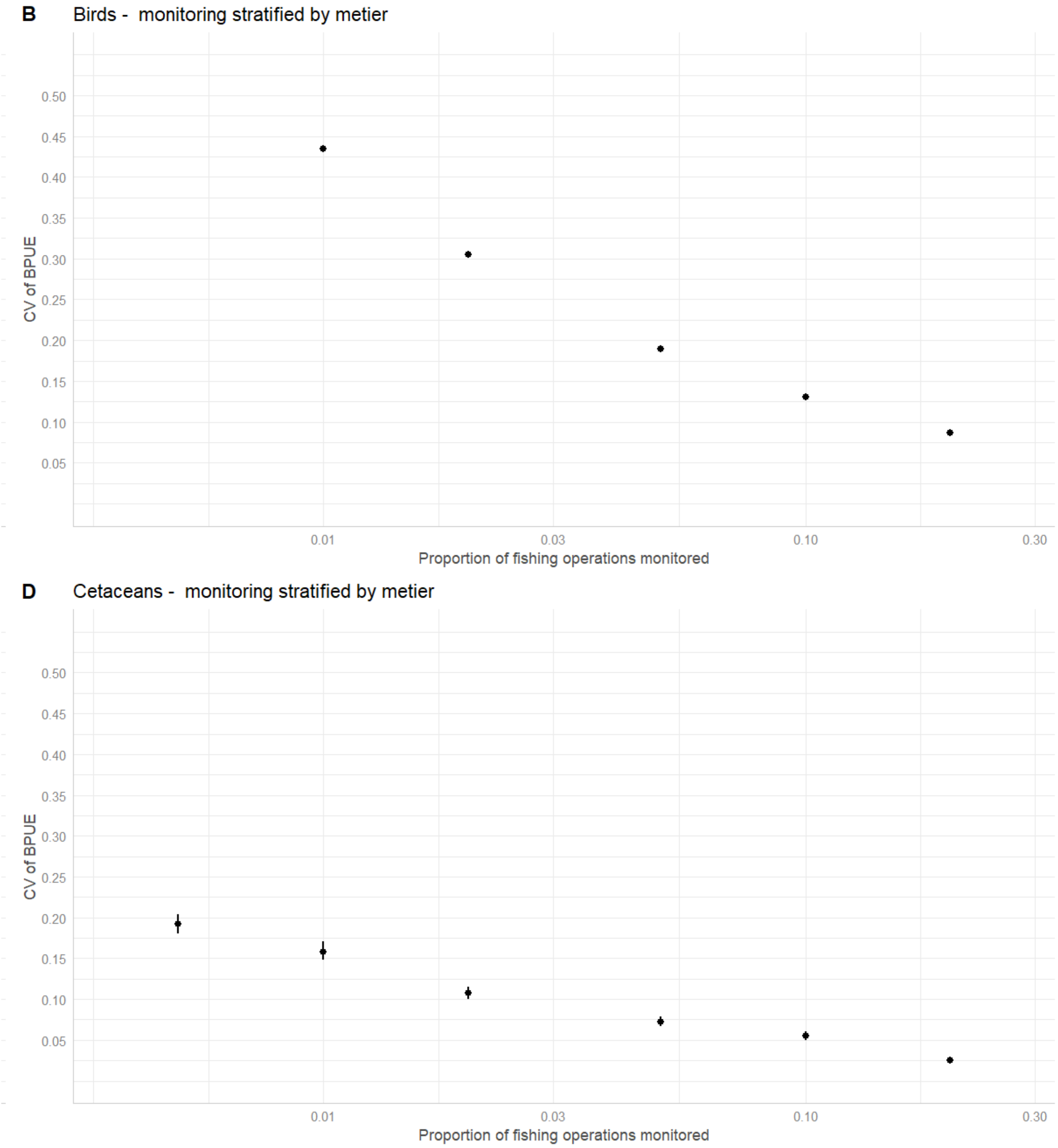
Calculation of total BPUE for a given gear or fleet, as well as its precision in terms of CV or bias, can be influenced by the stratification of the monitoring effort, among other factors, especially when the fleet is composed of several métiers with different behaviours and parameters (e.g. from soak time, time of day and depth to the probability of bycatch for each métier, for different species).

The purpose of this case study was to investigate the influence of monitoring stratification by métier (métier level 6) on the total BPUE and its precision (in terms of CV and absolute bias) for a fleet that consists of three different métiers with distinct characteristics, such as mesh size, target species and soak time. Total BPUE was estimated for three species group with differing probabilities of bycatch in the fleet. Detailed information about the fleet characteristics can be found in the description of the "Case Study C," since the same fleet parameters were used in these simulations. The SCOTI simulation framework, described in the previous section, was used to investigate these differences. The simulations were carried out with a) the métier-specific parameters and b) the whole fleet parameters. From these estimations, the total BPUE for the fleet and its precision and accuracy (CV and bias) were estimated for both métier scenarios (a and b in previous sentence). At this early stage of SCOTI development, each vessel was assumed to use only one métier because this is the most common scenario in the fleet and which simplified the simulations.

The results of the simulations, described in Annex 3 were subsequently modelled through GLMs and using as explanatory variables the possible interactions between the proportion of fishing operations monitored and the proportion of vessels monitored. The model results demonstrate the following for the present case study:

1. When a fleet is composed of several métiers with a wide variety of characteristics, monitoring stratified by métier reduces absolute bias and CV of total BPUE for the whole fleet for those species groups presenting very low bycatch probabilities (Figure 8C and D – mean bycatch probability = 0.009), but does not represent a remarkable improvement in groups of species with higher bycatch probabilities (Figure 8A and B – mean bycatch probability = 0.021).
2. Regardless of whether the monitoring is stratified by métier or not (in the present case study), for any of the simulated species' groups, the higher the number of fishing operations monitored, the lower the CV and absolute bias of total fleet BPUE (Figure 8).
3. Considering the characteristics of the simulated fleet, in cases where monitoring is stratified by métier, the CV and absolute bias of total fleet BPUE will increase as the number of monitored vessels increases, therefore there is statistical gain[*] from exhaustive monitoring of some vessels of each métier composing the fleet (Figure 9).

[*] Sentence edited base don reviewers' comments.


B
Birds - monitoring stratified by metier
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
CV of BPUE
0.01
0.03
0.10
0.30
Proportion of fishing operations monitored
D
Cetaceans - monitoring stratified by metier
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
CV of BPUE
0.01
0.03
0.10
0.30
Proportion of fishing operations monitored

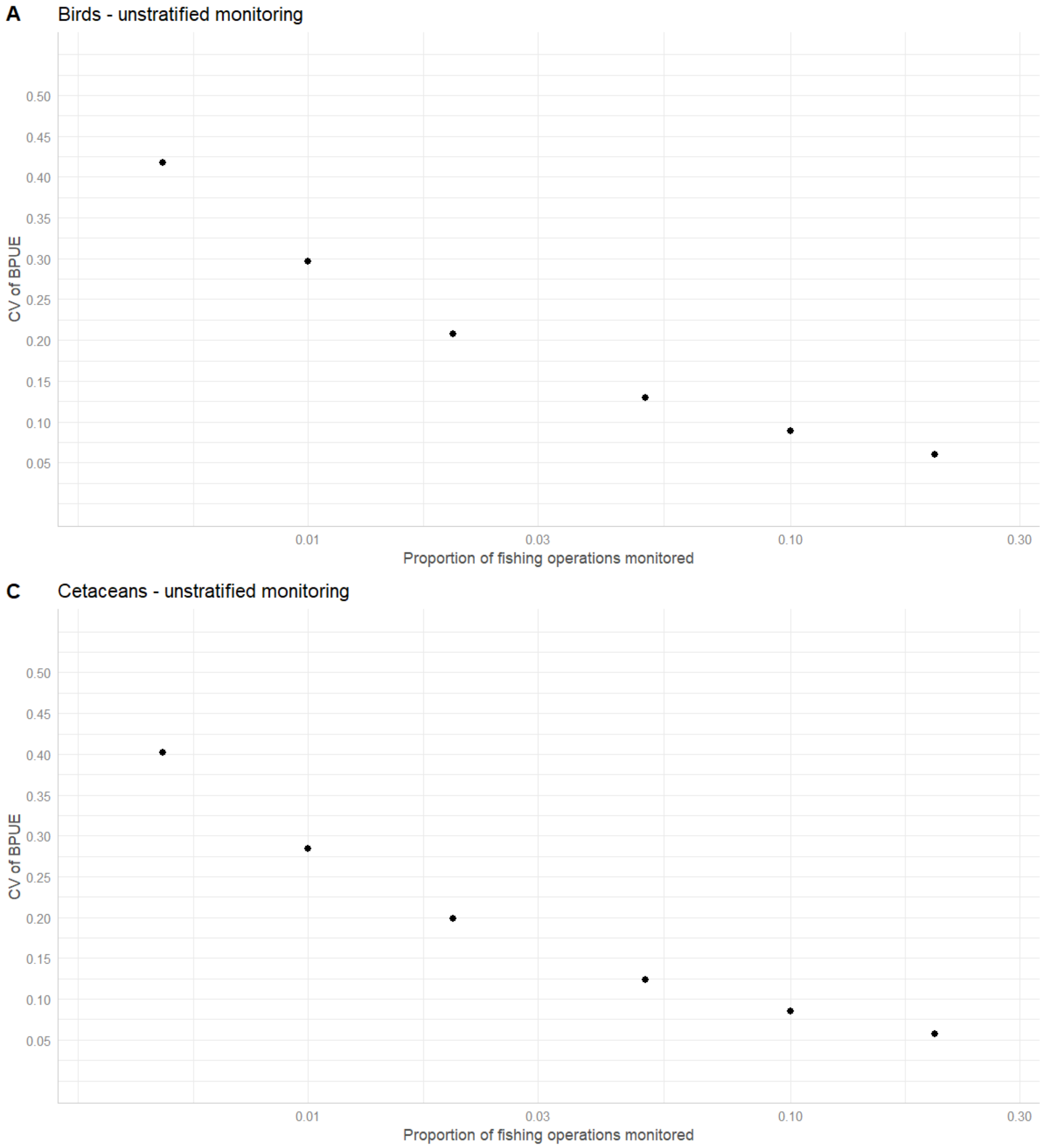


Figure 8 Predicted CV of total fleet BPUE of two groups of species (seabirds in graphs A and B; cetaceans in graphs C and D) for different proportions of fishing operations monitored when monitoring is unstratified (graphs A and C) or stratified by métier (graphs B and D). In the graphs, the y-axes have been equalized in the four graphs so that the changes in the CV can be better observed, if there is any, when stratification is performed by métier.

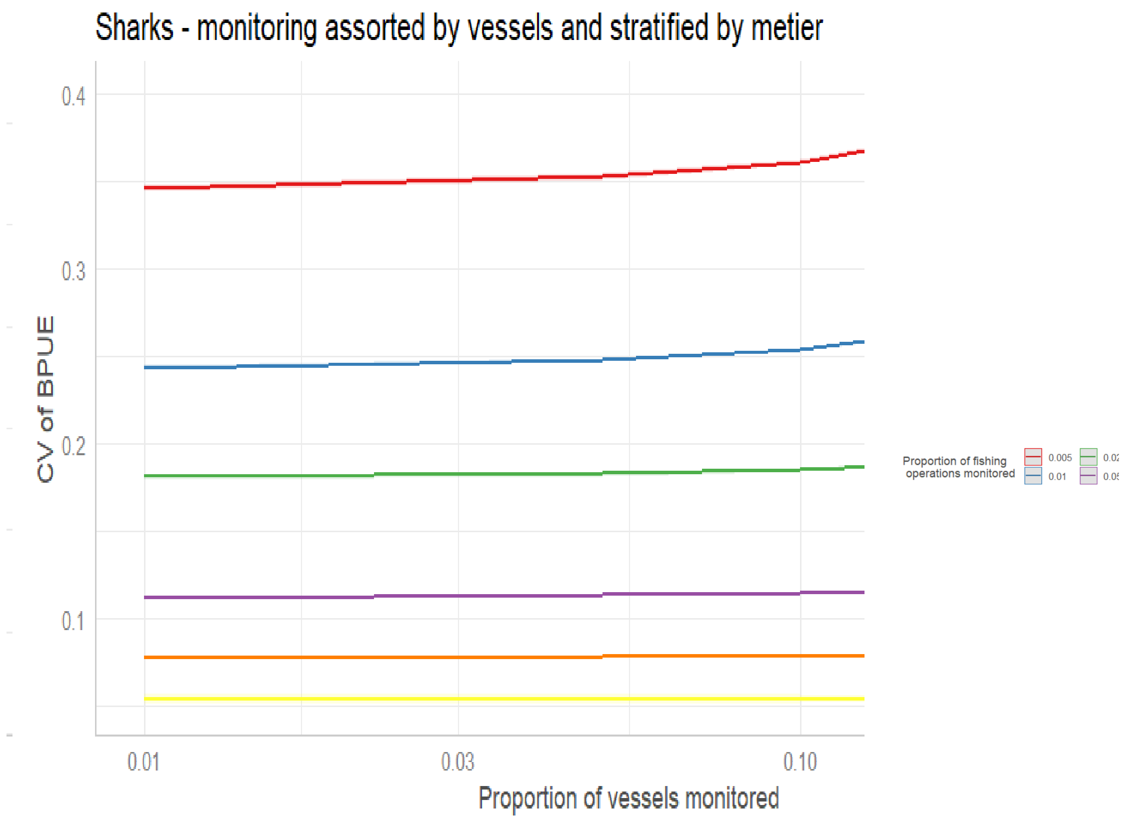


Figure 9 Predicted CV of total fleet BPUE of a shark species group for different proportion of vessels monitored when monitoring is stratified by métier. Coloured lines represent different proportions of fishing operations monitored.

3.2.3 Assessing the Sampling Design of a Reference Fleet

Parameterization for a reference fleet with seabird bycatch

We parameterized SCOTI simulations based on data and expert opinion from the Case study F, Reference fleet. The fishery has bycatch of Northern fulmars (*Fulmarus glacialis*), and three vessels in case study F report seabird bycatch observations. The parameterization of these simulations is described in more detail in the Annex to this section (Annex 4). Because this model is not a statistical model fitted to data from the fishery, we attempted to tune the simulation model to produce a similar bycatch per unit effort (BPUE) and CV_{BPUE} to existing values estimated from a linear model before running any simulations.

Currently, the BPUE for the Norwegian Reference Fleet is estimated using a generalized linear mixed model (GLMM) with vessel as a random effect. SCOTI currently assumes that all vessels share a common mean probability of a bycatch event and a common mean number of individuals caught if a bycatch event occurs (i.e., the same probability distributions describe bycatch for every vessel in the fleet). For the CV of the BPUE estimate from simulations to more closely match the model-based estimate from the GLMM, we modified the fishery simulation step in the SCOTI functions to include an added vessel effect. The effect is additive on bycatch numbers for hauls with positive bycatch.

We emphasize that the results below should not be interpreted absolutely as a representation of the Norwegian Reference Fleet or as prescriptive advice for the fishery. We have parameterized a set of simulations to try to generate a realistic degree of precision. The results below should be interpreted relative to each other and relative to the base case, not in absolute terms.

Incorporating a vessel effect in the model

To account for repeated sampling by vessels, we introduced additional variation in bycatches for each vessel after bycatch observations were generated. This was done by adding random noise to the final estimate using a normal distribution with $\mu = 0$ and σ defined as an additional parameter in the simulation (named `vessel.effect`). To ensure the additional variation didn't generate negative values, we applied it to log-transformed observations before back-transforming to the original scale.

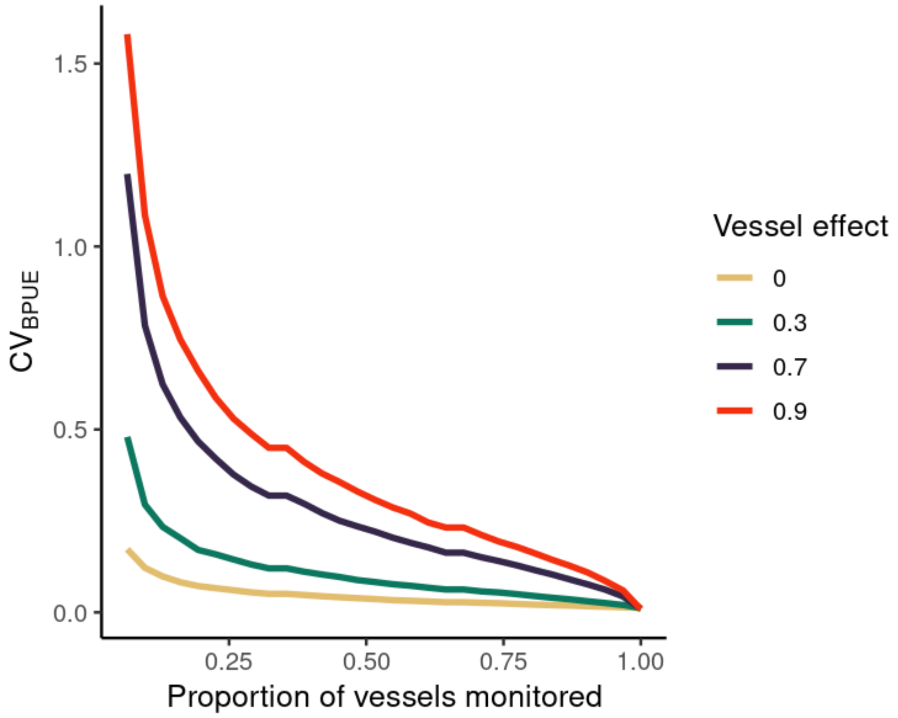


Figure 10 Increasing coverage in a reference fleet with different vessel effects. The vessel effect that generates a similar CVBPUE to the one estimated with a GLMM fitted to Norwegian Reference Fleet data.

The vessel effect that generates a CV like that estimated from the Norwegian Reference Fleet is σ = 0.7. When the vessel effect is stronger, the coverage (number of vessels monitored) may never be high enough to achieve a similar CV_{BPUE} (Figure 10). Vessel effects may be caused by spatial or temporal differences in effort between vessels, skipper experience, use of bird scaring devices, or other vessel-specific characteristics. In cases where these effects arise from skipper differences, vessel effects can be reduced through training programs. In situations where bycatch data come from an observer program instead of a reference fleet, it will be important to account for vessel effects in observer allocation as well, because this can lead to bias in fishery-wide estimates of BPUE.

Comparison of a Norwegian reference fleet to an observer program

We simulated a hypothetical observer program to evaluate the sampling effort needed to improve the performance of seabird bycatch estimation relative to current estimates that use the Case study F (Reference fleet) data. The values used to parameterize SCOTI for this example are provided in Annex 4.

An observer program would sample more vessels but fewer fishing operations per vessel. However, a refusal rate would be larger yet unknown, so we explored a wide range of values. The detection probability would increase given that the observer would have a single task. From the base case, we increased the `p_haul_obs` parameter and reduced `pmonitor` (Annex 4: table 1) to represent these characteristics of a hypothetical observer program.

In this simulation framework, a dedicated observer program implemented on the full fishery, where the refusal rate is around 60%, would need a target coverage of about 50% in the fishery to estimate a CV comparable to what one would get from estimating CV_{BPUE} from a Norwegian Reference Fleet (Figure 11).

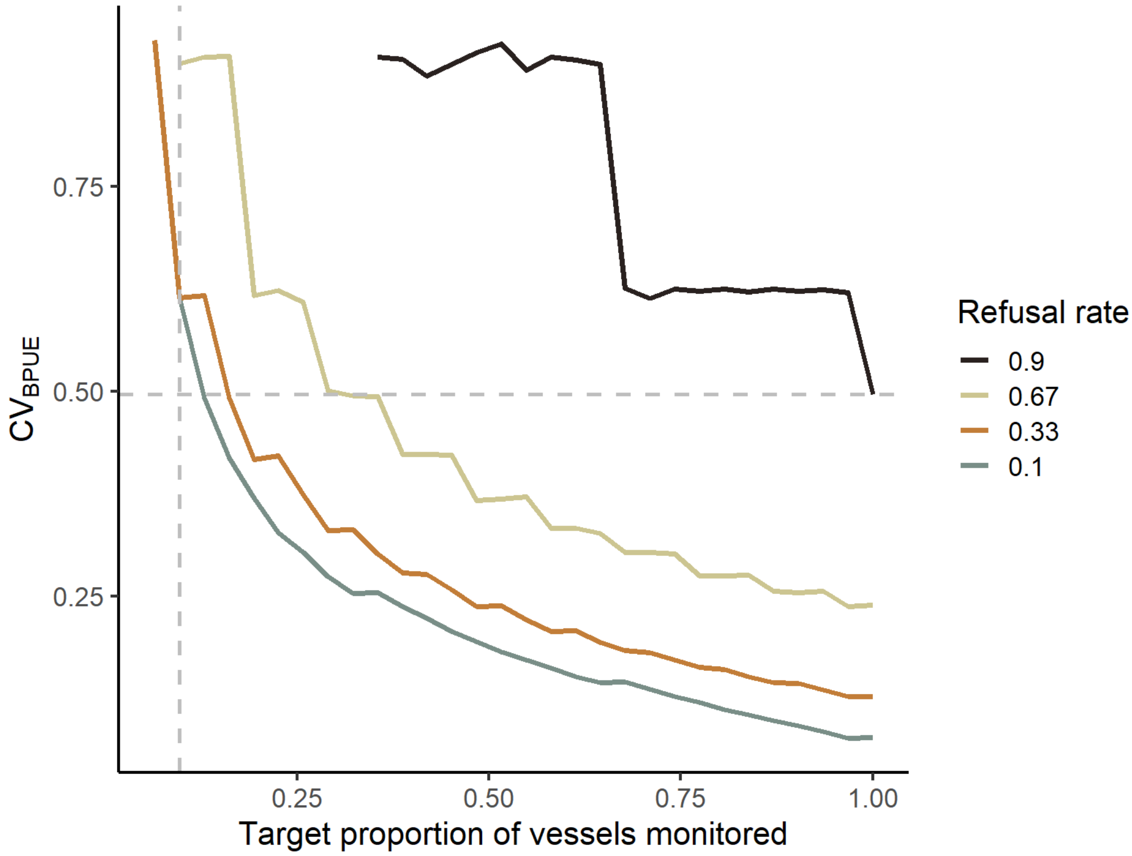


Figure 11 CV_{BPUE} estimated from a simulated observer program with different refusal rates (colored lines) compared to CV_{BPUE} estimated for the Norwegian Reference Fleet (horizontal dotted line). Current proportion of vessels monitored by the Norwegian Reference Fleet indicated by vertical dotted line. The target amount of coverage on the x-axis is the target coverage; the realized coverage is the product of the target coverage and the refusal rate.

The simulated observer programme exhibits a step-wise decrease in CV_{BPUE}, the magnitude of which increases with refusal rate (Figure 11). This occurs because the large refusal rate suppresses the realized vessel sampling: High refusal rates mean that even though target proportion increases, the realized proportion does not change. Therefore, the CV stays constant until the target proportion of vessels monitored overcomes the refusal rate enough to add one more vessel. The small size of the fishery in our example (31 vessels) magnifies the effect.

Spatial and temporal variation in the probability of a bycatch event

We built a framework for incorporating spatial and temporal hot spots (i.e., areas and time periods where the probability of bycatch events and number of bycaught individuals are higher than average), as ETP species can have strong spatio-temporal patterns in density and representative monitoring requires characterizing this variation. The extended framework also includes the possibility to change the fishery intensity in space and time. This allows analysts to test how well the monitoring captures true BPUE during different scenarios of overlap between species and fisheries.

We extended the original SCOTI framework to represent situations where there is spatial variation in the probability of a bycatch event occurring (i.e., bycatch hotspot(s)). The bycatch hotspot(s) in this function can be switched on and off during specific periods in a year. We modified the fishery-simulating function in SCOTI to include multiple areas, one or more of which can be identified as a hotspot(s) (e.g. an area where seabirds might be actively feeding) and a temporal trend (e.g. an area with a breeding colony where there is a temporal effect on the probability that a bycatch event will occur). The model assumes a collection of fishing areas, among which boats move with equal probability, unless the user specifies temporal and spatial trends in fishery effort across areas.

To emulate the Norwegian case study, we simulated a single hotspot for birds that was active during the summer months (aggregation due to breeding season) and simulated a fishery with a tendency to fish more intensively in the same hotspot area throughout the year (this type of situation has been observed with seabird bycatch in other fisheries as well, e.g. Fox et al. (2021)). The inclusion of a spatio-temporal bycatch hotspot thus increased the overall BPUE in the simulation, compared to a simulation without a hotspot area and with increased probability of bycatch. We did not have enough information in the real data to specifically parameterize how the BPUE is expected to change between hotspot areas and non-hotspot areas in the simulation.

As the mean BPUE in this case study is very small (close to zero), a small increase in the mean bycatch number (for example from 0.02 to 0.03 average birds per haul) will also change CV_{BPUE} substantially compared to changes in the standard deviation. This is thus more an artefact of how the CV is calculated (sensitive to small changes in the mean when the mean is close to zero) than an effect of variations in fisheries or hotspot areas for this case study (Figure 12). In general, the CV may be less informative about the precision of monitoring effort when the mean $BPUE$ in the fishery is close to zero, which can be the case of bycatch of ETP species. The hotspot framework requires further testing and parameter tuning before it should be used. In the section below, we also provide some recommendations about accounting for spatial and spatio-temporal variation in estimates of $BPUE$.

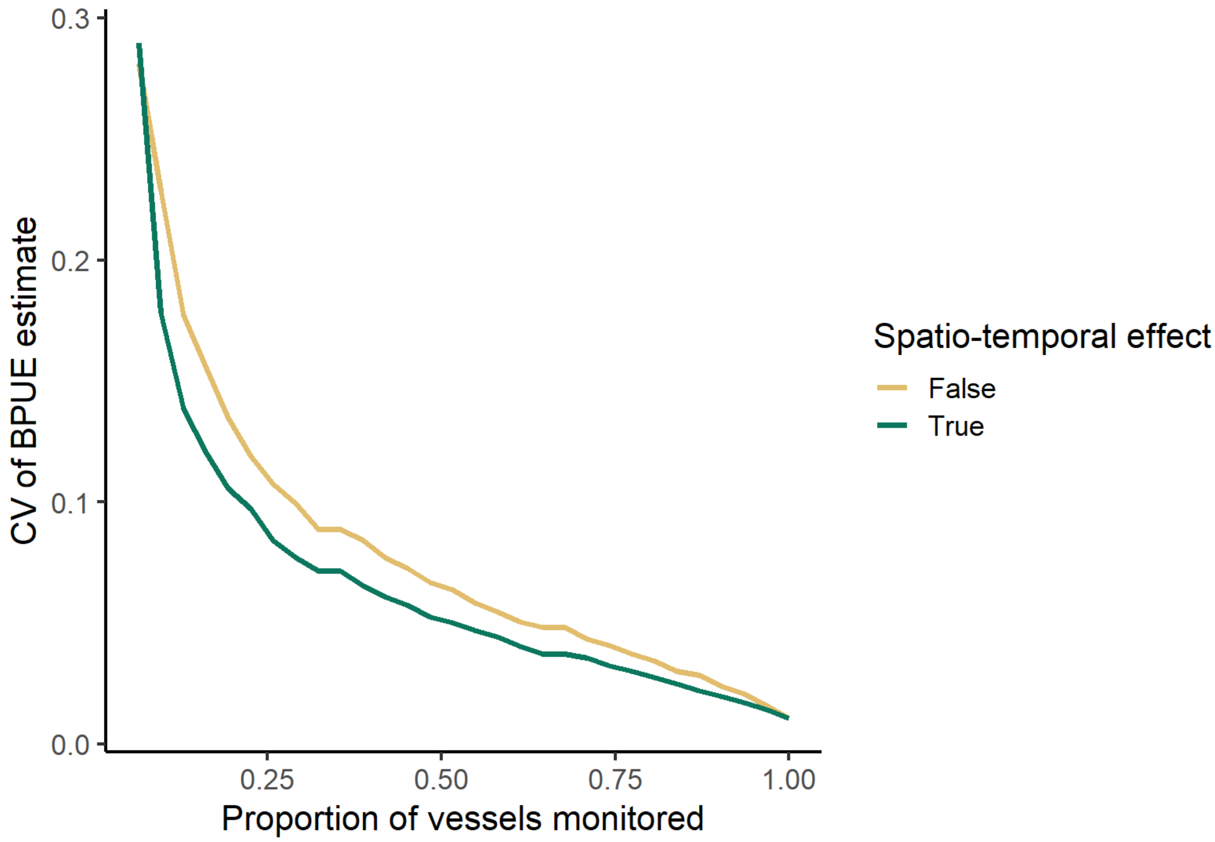


Figure 12 CV_{BPUE} estimated from a simulated observer program where there are spatio-temporal variations in seabird densities and fisheries, and for a situation without spatio-temporal variations.

Future spatial simulation work

We recommend that SCOTI itself be simulation tested and peer reviewed before it is used to guide specific management decisions on monitoring programmes. We provide some recommendations here as to how simulations and existing monitoring data can be used to provide scientific advice.

1. **Separate spatial, temporal, and spatio-temporal effects in simulation models.** Generalized simulations with spatial and spatio-temporal fishing and bycatch dynamics should separate spatial and temporal dynamics so observer programs can be designed to account for this variation. When there are sufficient data, simulations can be produced from spatio-temporal models that have been fitted to data (e.g. Arimitsu et al. (2023); Bi, Jiao, and Browder (2021)). These models could be simulation tested for their robustness to more data-poor situations.
2. **Take advantage of existing distribution and monitoring data to generate realistic simulations.** The spatio-temporal simulation extension in SCOTI needs input from both bird distribution data and fishery data in order to be parameterized to simulate realistic scenarios. Spatiotemporal species distribution models (e.g. VAST or sdmTMB; Anderson et al. (2022)) can be fitted with monitoring data and used to generate simulated distribution data that can be used in a simulation-estimation framework. Fishery data could be extracted from things like VMS data, to understand changes in fishery effort distribution throughout the year. In cases where spatial data may not necessarily be available, hierarchical models can still be fitted to data to provide useful simulations (e.g. Authier, Rouby, and Macleod (2021)).

3. **Generate spatial observation models to simulate spatial stratification in monitoring programs.** Monitoring programs should capture the existing variation in fleets to the best extent possible; useful advice on how to allocate bycatch monitoring effort spatially will require some information about spatial variation in bycatch. Spatiotemporal changes in the observation models is yet to be implemented in the SCOTI framework.

General conclusion of vessel effect and refusal rates

- The precision of bycatch per unit effort estimates (CV_{BPUE}) is sensitive to vessel effects, whether these are due to spatial, temporal, or spatio-temporal variation in effort. Practices for reducing or accounting for the size of the vessel effect may be more cost effective than increasing observer coverage for a given fishery.
- For a fishery to achieve a similar degree of precision in an observer program as in a reference fleet, it will require more coverage, especially in cases where refusal rates could be high. If refusal rates are high enough, an observer program will not be as useful as a reference fleet in reducing precision. Additionally, reference fleets sustain motivation more than observer programmes due to incentives such as payment and more involvement in the scientific process. As always, a reference fleet should include the full range of vessel-specific variation that practitioners believe exists in the full fleet.
- The precision of bycatch per unit effort estimates from the monitoring effort (CV_{BPUE}) is sensitive to the real $BPUE$ in the fishing fleet. This is an artifact of how CV is calculated, and that the mean $BPUE$ is often a number close to zero for ETPs. A small increase in mean $BPUE$ between simulations, without large changes in the standard deviation, will produce in general lower CV across most proportions of monitored effort as long as the monitoring is sufficient to include some extreme bycatch event in any hotspot areas.

3.2.4 Investigating the role of monitoring coverage and allocation of samples by downsampling bycatch data

In the previous sections, the relationship between properties of BPUE estimates (bias and CV) and the coverage of bycatch monitoring was investigated by applying various sampling schemes to simulated fisheries (generated virtual populations) and their associated bycatch. The simulated fishing year was created using parameter values on fishing and bycatch rates derived from different case studies. Here, an alternative approach is presented. Instead of simulating the fishing and bycatch data, the population of fishing events is downsampled from real field data from Case study E, Pair trawlers in Adriatic waters. To simulate a range of values of bycatch monitoring coverage, random subsamples of various proportions of the original data, with corresponding bycatch events, was generated. From this downsampled data, estimates of BPUE were then calculated.

Two different methods of downsampling were performed. In the first set of analyses or “simple random downsampling of hauls”, sampling of the original monitoring data was performed at the haul level. This allows us to evaluate the overall influence of monitoring coverage on different metrics of bycatch. In the second analyses, a “two-stage downsampling” was performed. In the first stage, vessels to be monitored were selected. In the second stage, among the selected vessels, hauls were sampled. In this two-stage downsampling, we also investigated the influence of allocation of monitored hauls among vessels and hauls. Two cases were considered: allocation of a given sample size (number of hauls to be monitored at a certain proportion of original data) among “few” vessels (3) with a larger number of hauls monitored on each vessels, or “many” vessels (10) with fewer hauls monitored on each vessel. Note that to be able to allocate a given number of hauls among a low number of vessels, we needed to consider a subset of vessels which had >=50 hauls / vessels in this set of analyses. The number of hauls to sample from each vessel was determined using proportional allocation.

The proportions of hauls sampled from the original data ranged from 0.001 to 0.5 for the simple random sampling of hauls. For the two-stage downsampling, the proportion of hauls ranged from 0.005 to 0.1; a lower maximum proportion than for the sampling of hauls was necessary to be able to allocate the hauls among a lower number of vessels. For each proportion, 10000 samples were drawn, each with bycatch counts of the four species. For each set of samples, corresponding to different proportion of hauls monitored, the bycatch per unit effort (BPUE; where the effort here is a haul) was estimated as the mean number of bycaught individuals across the random subsamples. The probability that a species was detected during monitoring was calculated as the mean number of samples with 1 or more individuals recorded. In addition, CV of the mean bycatch estimate was calculated (equal to standard deviation of the estimates / mean of the estimates).

3.2.5 Downsampling at the haul level

When downsampling of the original monitoring data was performed at the haul level, the results suggest that:

1. above a certain monitoring effort (circa >10%), the bias and CV on BPUE are low and does not change much with effort, especially so for more frequently bycaught species (species 1-3; Figure 13). That is, it seems to be a clear diminishing return of increasing sampling effort on the properties of the BPUE estimate in terms of bias and precision (CV).
2. Similarly, at low levels of monitoring effort, the bias and CV increases rapidly and, in some case, (e.g. 0.5% effort) would be unacceptable.
3. The detection probability of a species, i.e., the proportion of hauls where one or more individuals of a species were observed as bycatch, increases rapidly with increasing coverage, and less so for the rarer species. To be able to certainly detect a bycatch event when happens for species with “very high” or “high/moderate” or “very low” bycatch frequency the coverage needs to be higher than about 1% or 5% or 50%, respectively. For species with very low bycatch frequency a coverage of 7-10% is necessary to exit the field of detection causality/random detection.

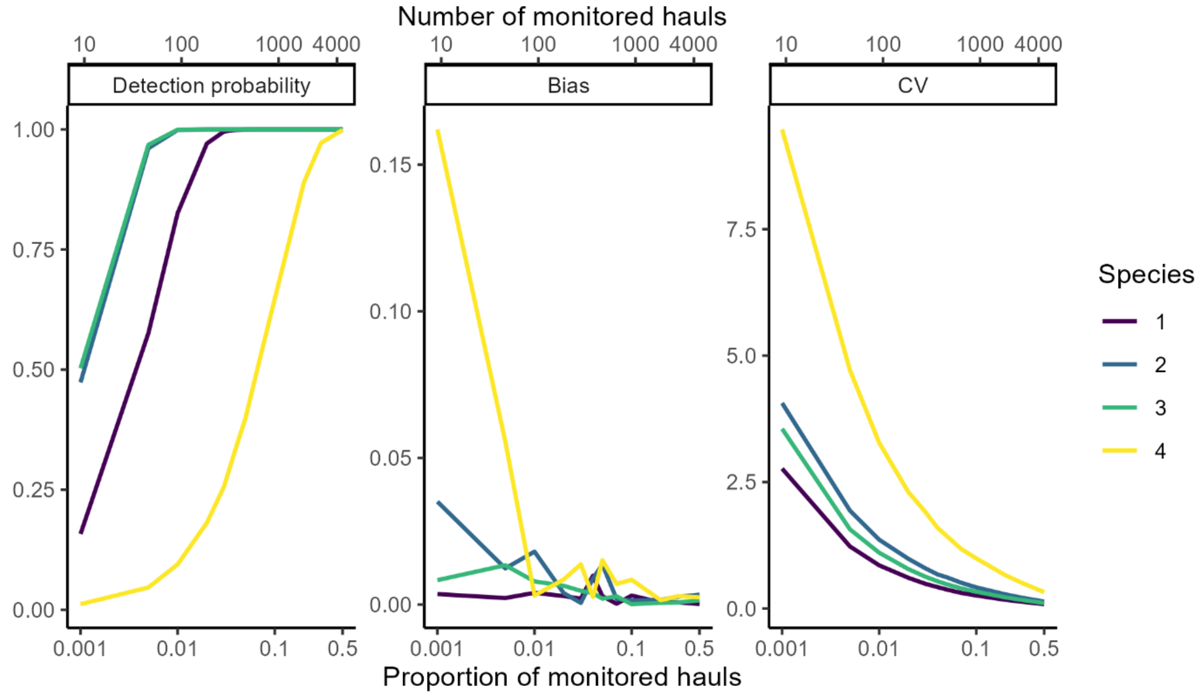


Figure 13 The influence of the proportion of monitored hauls that were subsampled from bycatch data on four different species from the northern Adriatic Sea, and detection probability (proportion of hauls with at least one individual observed), bias and CV in estimated bycatch rate (BPUE, number of individuals per haul). To facilitate the distinction of different species at lower proportions of monitoring, the x-axis has a base-10 logarithmic transformation.

3.2.6 Two-stage down-sampling

In the second set of analyses, downsampling of the original monitoring was performed in two steps: first vessels were sampled, either "few" (3) or "many" (10) vessels. The hauls to be observed are then selected by random subsampling among the randomly selected vessels. Proportional allocation was used to determine the number of hauls to sample from each of the selected vessels. Thus, for each level of monitoring, a fixed number of hauls is allocated among either a few or many vessels. Everything else being equal, this analysis may provide some insight into optimal allocation between primary sampling units (vessels) and secondary sampling units (hauls), and how different allocations may influence precision (CV) and the probability to observe a bycatch in a haul. The results from the two-stage downsampling suggest that:

1. The CV is consistently lower when sample size is allocated to a larger number of vessels (Figure 14, yellow lines), across all species and the entire range of monitoring effort, i.e. number of hauls to be sampled.
2. When monitoring effort is allocated to a larger number of vessels, the CV is less sensitive to a decrease in monitoring effort than when the same effort is allocated to a few, i.e. the slopes of the "many vessel" lines are less negative than the "few vessel" lines.
3. The probability to observe a bycatch species is higher when a given monitoring effort is allocated to a larger number of the primary sampling unit, i.e. vessels (Figure 15), especially so for a rarer bycatch species (panel 1).

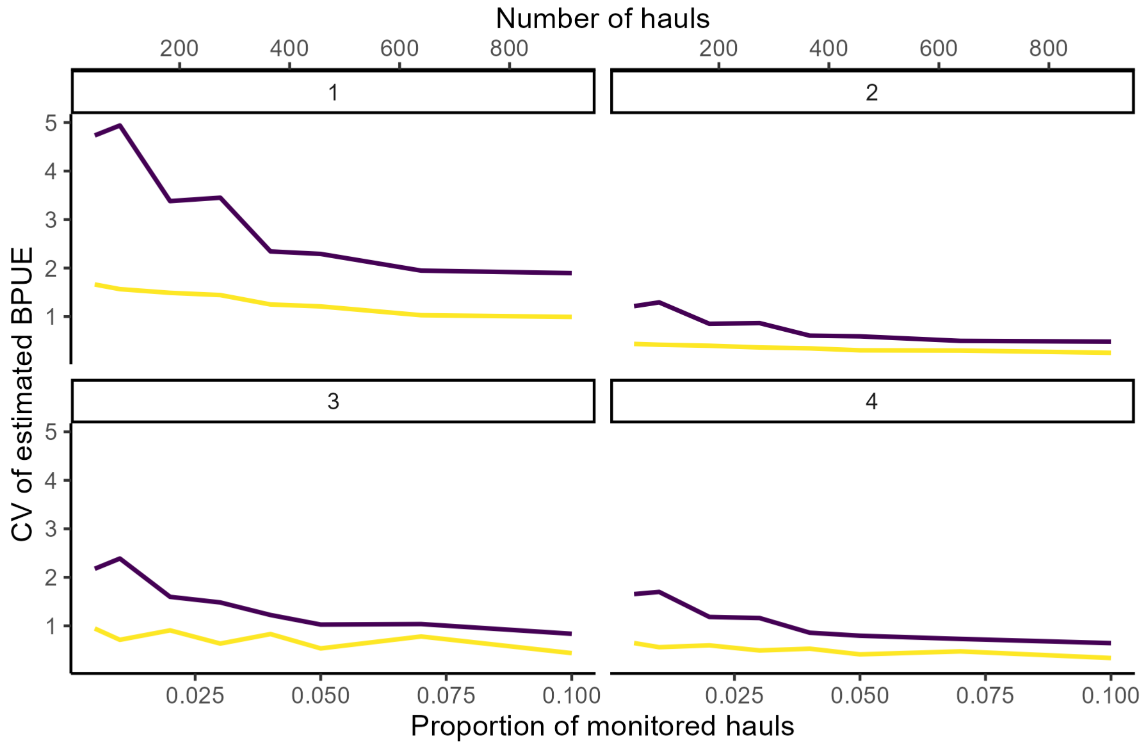


Figure 14 The relationship between proportion of monitored hauls (corresponding number of hauls are shown on upper x-axis) that were subsampled from bycatch data on four different species from the northern Adriatic Sea and CV of estimated number of bycatch per haul (BPUE). Two different sample size of vessels were used, indicated by the line colour (purple: 3 vessels; yellow: 10 vessels). The sample size of hauls was distributed among the vessels using proportional allocation. Panels (1-4) represent different species.

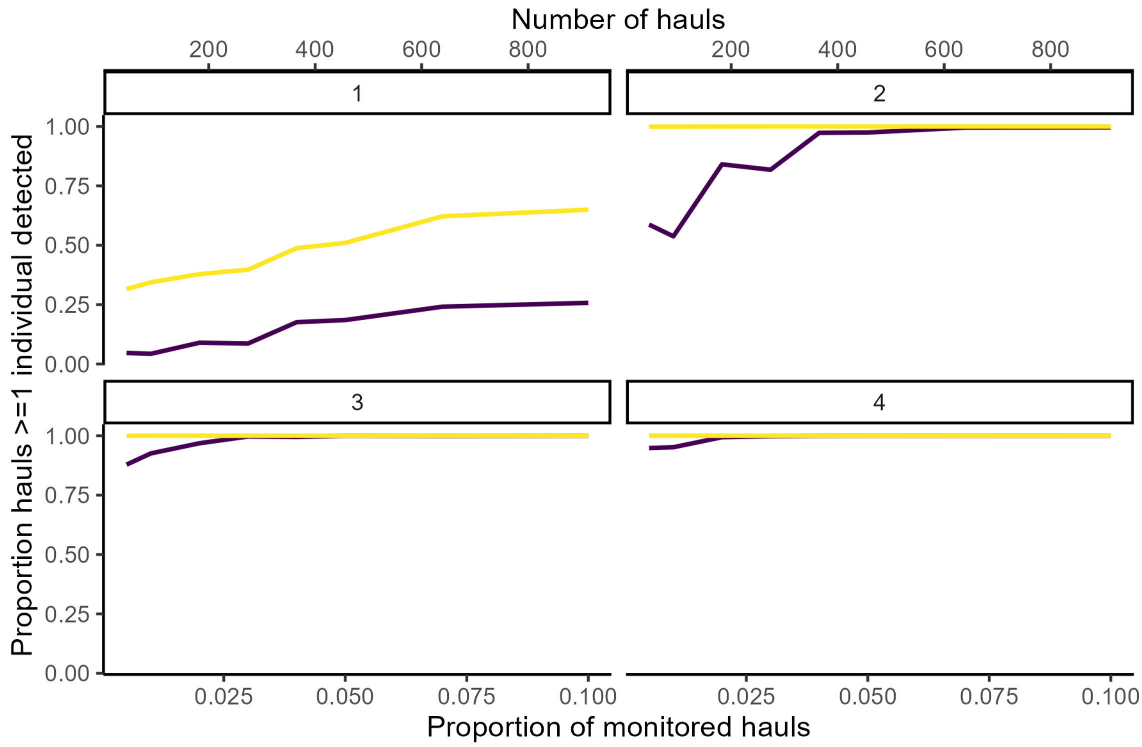


Figure 15 The relationship between proportion of monitored hauls (corresponding number of hauls are shown on upper x-axis) that were subsampled from bycatch data on four different species from the northern Adriatic Sea and proportion

of hauls where 1 or more individuals were observed as bycatch. Two different sample size of vessels were used (3, purple line; 10, yellow line). Panels (1-4) represent different species.

Above, downsampling of the original data on bycatch in the northern Adriatic Sea was used to create a virtual range of bycatch monitoring. This allowed us to examine the relationship between monitoring effort and probability to observe a bycatch species in a haul. This relationship was also investigated using annual variation of monitoring and bycatch events of a very rarely bycaught species (regularly present in the fishing area) in the original data (Figure 15). Although the number of observations is rather low and the entire range of monitoring coverage did not appear in this observational data, the pattern suggests that for this species with a monitoring effort of less than 4% of the hauls, a bycatch event (one or more individuals bycaught in a haul) is unlikely to be detected. On the other hand, with an effort above 6%, a bycatch event can be detected. This is confirmed by Figure 16 plotting the relationship between annual monitoring effort (proportion of hauls with onboard observers) and observed bycatch events of such very rarely bycaught species.

This is in line with what is shown in the first panel of Figure 13 where, in order to have a probability of detection higher than just "by chance" (50%) for our "rarely bycaught species", on-board monitoring effort should be closer to 10% than 5%.

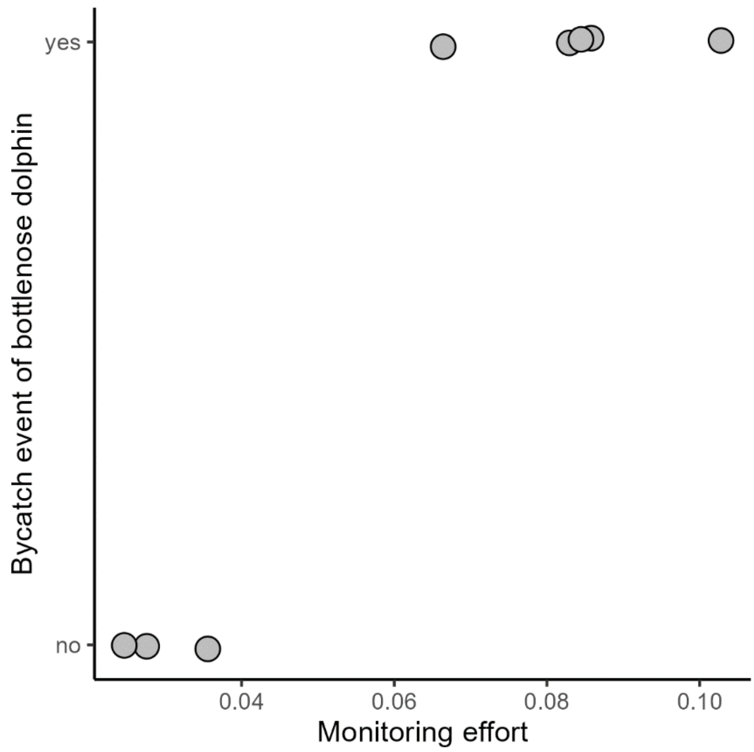


Figure 16 The relationship between annual monitoring effort (proportion of hauls with onboard observers) and observed bycatch events of a very rarely bycaught species using data collected 2006-2013 by onboard observers on pair trawlers in the Northern Adriatic.

4 Overview of common biological and demographic parameters to assess human-induced mortality

European Union Member States are required to collect "data to assess the impact of Union fisheries […] on bycatch of non-target species, in particular species protected under Union or international law" (Regulation (EU) 2017/1004). Data on ETP species incidental bycatch are collected routinely by most MSs, but often lack a level of detail sufficient to estimate the total number of animals incidentally captured in fisheries (ICES, 2024) or to inform mitigation measures. Although total bycatch and mortality estimates are essential to assess the effect of fisheries at species or population level, information on the biological characteristics of bycaught animals is key to refine any subsequent management advice.

In the past few years, several expert meetings discussed what type of fisheries-dependent information should be collected to inform bycatch management, mainly in the context of additional information that should be collected by observers under the DCF sampling programme and the design of the RDBES (e.g. WGCATCH 2019, TOR C). These discussions were mostly driven by the wish to collect any possible information in the DCF sampling scheme, taking into account that incidental bycatches are rare, and that more information should be recorded if possible.

We envision that data for bycatch assessments to be carried out by WGBYC will be provided by the RDBES in the future. These data already include information on the number of individuals bycaught but should be completed with information for key biological parameters that are required as a minimum to carry out such bycatch assessments. In this section we review and discuss the minimum requirements for on-board observer monitoring programmes regarding biological parameters required to assess the effect of fisheries bycatch on affected populations/species.

The collection of some **key biological and ecological parameters** is essential for developing models to assess the impact of fisheries impact on species (or populations). Such knowledge of the parameters associated with bycatch of ETP species is also often critical to identify and implement the best mitigation approaches. The most commonly used biological and demographic parameters (or their proxies) are listed in Table 4.1. Some key parameters refer to populations (or species) demographics and distribution (e.g. population abundance, population growth rate, etc.) and must be provided/collected outside fisheries monitoring programmes, while others refer to information on the "nature" of bycaught animals (e.g. species, sex, etc.) that need to be collected at vessel-level by e.g. observers or via electronic monitoring. In all cases, the quality of the models using these parameters is dependent on the quality of the input data, which should be carefully considered. Some of the parameters in Table 4.1 may be derived from other more basic measurements. For example, length and weight are often used for some taxonomic groups to derive age or sexual maturity, while for some taxa sexual dimorphism becomes evident at sexual maturity, which is connected to length. Demographic parameters are also key to build more elaborated and precautionary population models for different taxa. In this case, specific non fishery-dependent studies need to be carried out to obtain them. Demographic parameters can also be obtained from existing long-term monitoring programmes of populations (e.g. mark-recapture studies). Biological parameters are useful to understand the nature of the interactions and inform mitigation studies. This is especially true when interactions between fishing and ETP species are exacerbated by specific conditions (e.g. young inexperienced animals getting entangled, pregnant females, or sick individuals taking advantage of an energetically less expensive meal).

Table 4.1 Commonly used biological and demographic parameters. Examples of the type of information these parameters can be derived from is indicated. The table points at which data should be collected onboard fishing vessels in priority (Mandatory: the parameter is essential for bycatch assessments; Recommended: the parameter is important to improve bycatch assessment; NA = not applicable or collected outside onboard monitoring programmes).

Parameter	Priority for onboard monitoring	Derivable from
Species	MANDATORY	Onboard monitoring Direct identification if possible, or picture of the animal including key features for later identification
Dead/alive	RECOMMENDED	Onboard monitoring Usually directly noticeable, but may need to be further defined (e.g. WGCATCH 2019)
Sex	MANDATORY	Onboard monitoring Visual identification if sexual dimorphism; a picture of the genital area; a tissue sample (via DNA)
Length	MANDATORY	Onboard monitoring For the animals brought onboard: picture of the entire animal with a ruler or of the animal on a measuring board. If the animal is too big for handling, or if the handling could put the observer at risk of injuries, a picture including some fixed reference point (e.g. an apparatus onboard the fishing vessel, or a human “for scale”) to estimate the length of the animal is recommended instead
Age	RECOMMENDED	Onboard monitoring Age determination can be directly visible from external features for some species, or derived from proxies like length, weight, etc. sample taken onboard
Weight	RECOMMENDED	Onboard monitoring Where relevant weight is susceptible to provide information on the by-caught animals together with other parameters like length
Maturity	RECOMMENDED	Onboard monitoring Sexual maturity determination can be directly visible from external features for some species, or derived from proxies like length, weight, etc.
Gravidness	RECOMMENDED	Onboard monitoring Pregnancy determination can be directly visible from external features for some species, or derived from proxies like length, weight, etc.
Condition	RECOMMENDED	Onboard monitoring Usually directly noticeable, but may need to be further defined
State of decomposition	RECOMMENDED	Onboard monitoring Picture of animal
Total Abundance	NA	Population monitoring (survey)
Population growth rate	NA	Population monitoring (survey) Abundance estimates

Carrying capacity/ maximum population	NA	Population monitoring (survey)
Adult survival	NA	Dedicated studies Mark-recapture data
Average age at maturity	NA	Dedicated studies Age/length curves & direct observations (post-mortem examination, long-term individual monitoring)
Maximum age	NA	
Fecundity	NA	Dedicated studies

4.1 List of key biological parameters to collect for assessing fisheries effect on sensitive populations/species

Species

Determining the species and, when possible, the population a bycaught animal belongs to is an **essential first information** to assess bycatch impacts. Species identification for a wide range of taxa susceptible to bycatch is challenging, and organizations responsible for national data collection programmes should provide adequate training for observers, and supply identification guides and procedures. Considering the difficulty for observers to identify all bycaught specimens to species with certainty, **the subgroup advises that bycatch events are documented with photographic evidence** that would also provide additional information for some of the biological parameters described below. When the identification at species level is not possible with certainty, the closest certain taxonomic level should be entered instead.

Dead/alive

Out of the parameters outside of species, the information on whether the bycaught animal is dead or alive was discussed thoroughly in this subgroup. Such information would be a useful addition to the models and the subgroup agreed it should be collected in the programmes where possible. The information can be categorized *e.g.* as dead, injured, or unknown (see *e.g.* Carretta 2021), especially in certain types of fisheries that are suitable for such assessment. However, defining whether a released animal is dead or injured can be complicated onboard and defining an objective method for all species is difficult. Some potential issues include for instance survival rates after release, overall disturbance to the animal, and large differences between different species and fishing methods. More specifically, the question is how to estimate the survival after release: a released animal can be barely moving and it may look like it will not survive for very long but still be able to recover, depending on *e.g.* the released species and the habitat it is released in. It can also be released quickly with no visible harm but still not be able to recover or reproduce due to non-visible issues such as internal damage. Considering these limitations, there is a risk that observers use different criteria that are not comparable spatially nor for different species and can lead to wrong estimates of mortalities. It was discussed whether collection of such information is useful at all. If this is used as a mandatory field in the RDBES, it should at least have the option “NA”. However, based on previous experience, only a small amount of data could be received if the field is marked as non-mandatory.

In the case of ETP species protected under EU laws, MSs need to minimise both the capture and the mortality induced from capture. HD article 12.4 explains that MSs '*shall take further research or conservation measures as required to ensure that* ***incidental <u>capture</u> and killing does not have a significant negative impact*** *on the species concerned* [emphasis added]. Hence, in this sense the fate of an ETP species individual taken as bycatch is negligible when assessing the overall impact of bycatch, which includes mortality and unwanted capture. Nevertheless, the level of mortality caused by a given métier is an important aspect to inform more effective mitigation strategies. This can be estimated based on species-specific studies, expert opinions, or assumed as total when information on post-release mortality is unavailable. In many instances, the data collected by the DCF or other similar programs may be sufficient to assess the post-release mortality for certain species and areas. Therefore, there are recommendations for data recording and for the database format. In the field, the recommendation is to record the animal "*dead*" when it is certain (or almost certain) the animal is dead. Additionally, the observers should record "*injured*" if the animal is visibly injured. "*No visible injuries*" should be recorded when the observer can subjectively assess the animal behaving as it should be assumed to behave in such an instance.

Sex

The determination of the sex of bycaught animals was discussed by the subgroup. Recording the sex of all bycaught individuals is susceptible to provide important information on the potential impact of fisheries at population/species level. A sex ratio different in the bycatch records than in the originating population could negatively affect the reproductive capacity of that population. For instance, a disproportionate catch of mature females can significantly reduce this capacity. A disproportionate catch of males would give the opposite effect (Addink et al, 1997). Although sex determination is relatively straightforward for species with marked sexual dimorphism, *e.g.* elasmobranchs, or Anatidae (ducks), it may be difficult or even impossible for many other species. The subgroup recommends that, where possible, sex determination is carried out by observers directly onboard fishing vessels, but also advises that observers take pictures of every bycatch individual they record to help identify or confirm observations. Observers should be provided with detailed instructions on the species and body parts to be documented with photographic evidence. In a database, sex should be entered in a mandatory field with the option "NA".

Age-related parameters

Age, weight, length, maturity, and pregnancy of bycaught animals are all useful information for informing population models that should be collected when possible. Some of these parameters are correlated and can be used as proxies for others (*e.g.* length can be a proxy for age and maturity). There are differences whether these data can be collected, depending on the bycaught species and the fisheries considered. It is likely that at least some of this information can be collected or at least estimated for most bycaught specimens, and in many cases one parameter can be used to estimate another. A measure for the age is very useful information to construct the part of the population that is taken as bycatch. This is especially the case for long-lived animals, and therefore the case for many endangered species. Age is a difficult parameter to obtain onboard and for that reason will not often be available; however, when possible (*e.g.* for bony fish), an age sample can be taken and analysed afterwards. For all species groups except for birds, the body length of the animal can probably be used as a proxy for age. Maturity in itself gives an indication if the specimen took part in the reproduction process. This is in particular true for pregnant females. Pregnancy can be indicative for species with a low production - and a high survival rate.

Condition/fitness

Body condition is a parameter that could be used to give insight into the health of an animal's body. Quantifying energetic reserves shows how prepared they are to reproduce, migrate and survive predators in a changing environment. Changes in condition factor give insight into possible changes that influence physiological processes, parasitic interactions and food abundance at a certain time. This value varies depending on breeding. Many methods for calculating condition index exist and their use depends on required precision and availability to gather enough data as some methods require more and harder to obtain measurements (*e.g.* large shark body circumference; Irschick & Hammerschlag, 2014). Most basic ones require different forms of body length and width measurements combined with weight. Depending on taxonomic group specific characteristics may be used such as tarsus length for birds, shape of plastron for turtles and blubber thickness for cetaceans (Merilä et al. 2001, Jordan et al. 2009, Castrillon and Bengtson Nash 2020). Health and body condition of marine animals is largely influenced by the presence of parasites and disease. Parasites may use up energy reserves that could be otherwise used for reproduction, migration and dealing with predators as well as interfere with normal functioning of organs (Geraci and Aubin 1987). They weaken an animal's immune system making them more susceptible to bacterial disease. This group recommends that any externally visible parasites and diseases should be noted by the observers to give a more detailed insight into what has caused a potential change in body condition.

Decomposition

State of decomposition impacts the quality of other information gathered from the animal and allows approximating how long it has been deceased. Advanced states of decomposition may lead to the inability to perform certain required measurements or values may not reflect the animal correctly. If a bycaught animal is dead, observers should note carcass characteristics to determine the decomposition category (see e.g. Valverde et al. 2020). This group recommends using vague categories due to environmental factors such as temperature influencing the rate of decomposition. In some fisheries with large secondary bycatch, it is important to have clear onboard instructions on how to record a bycaught specimen. For instance, "fresh, very likely a bycaught specimen", "not sure if the animal was not already dead" and "given the state of decomposition, this specimen was for sure not alive when getting caught".

4.2 DNA sampling and eDNA

Exact species identification, stock (or population) affiliation, sex determination, etc. can be determined from collecting DNA samples of bycaught animals. Although DNA monitoring may be more difficult to put in place than traditional observers sampling, it can be the only realistic way to assess fisheries effects of bycatch for some mixed populations. Moreover, it is advisable that storage of DNA samples for subsequent analysis should be limited to freezing as the use of chemical fixatives is not permitted on fishing vessels. The use of environmental DNA (eDNA) is also a way of identifying rare species that come into contact or may come into contact with fisheries. This application requires a water sample from the fishing depth.

4.3 On board practicalities

The requirements for assessments of bycatch effects are closely connected to the training of the onboard observers and the protocols they use on board. These onboard best practices have been addressed in earlier reports and are summarized in this report (ICES 2020). In this section, we merely want to point out that it can be difficult to make a record of even the minimum required parameters – *i.e.*, species – as observers or EM analysts may lack the adequate training to identify all possible ETP species bycatches they may encounter. We recommend that fisheries observers document all bycatch events with photos that they can send out to identification specialists. As the saying goes, a picture is worth a thousand words. Depending on the species group, the information that can be inferred from photographs includes species, sex, length, condition, maturity. Protocols for onboard observers should contain detailed instruction by species group on how to take these pictures, but we provide here what we believe should be the minimal requirements:

1. Always put a reference object for size next to the specimen to be photographed, preferably a ruler.
2. The body parts that need to be visible for identification purposes differ per species group, but in general there should always be a picture of the complete animal from the side (lateral view).
3. Depending on the species or the vessel configuration, more than one picture per animal may be necessary.
4. Metadata associated with the picture should be conserved to ensure the correct allocation of the bycatch events to the corresponding fishing event.
5. A picture must not contain objects or parts which could make it possible to identify a vessel or a person. The context of the picture – an incidental bycatch by a fishery vessel – should not be recognizable.

In case of a bycatch of a large animal it is not easy to arrange that a picture can be taken of the full length of an animal with a ruler or another reference measure. It is suggested to ask crews to paint a ruler along the side of the hold or another object on the vessel where large bycatches are likely to be moved after having been caught. As the examples below refer to onboard sampling, it also contains non-endangered species. Onboard sampling (for example under the DCF) should deal with any catch, especially if the species is not yet known.

4.3.1 Examples of bycatch photographs categorized by species group

Elasmobranchs

Sharks and rays may be difficult to identify with certainty onboard fishing vessels, notably because of the high number of species susceptible to be recorded as bycatch. In addition to a picture from the side with a ruler, pictures of elasmobranchs should be focus on the mouth with the animal’s teeth visible and on the fins; moreover, taking pictures of the ventral side and of the anal fins is useful for sex determination (Figure 17-Figure 19).

Figure 17 Example of a picture suitable for identification purposes for sharks and rays: full length with a ruler. In this case an additional picture of the mouth is required (picture by Bram Couperus; species *Centrophorus squamosus*).

Figure 18 Close of the mouth. The shape of the teeth of deep-sea sharks are often helpful for the identification of the species.

Figure 19 Example of picture taken of shark's genitals (male).

Bony fish

Many specimens of incidentally caught bony fish may be difficult to identify onboard. Fortunately, bony fish often can be collected or photographed easily. In addition to a picture from the side with a ruler, pictures may be taken of all fins and a close-up of the head (Figure 20-Figure 21).

Figure 20 This specimen, caught west of Ireland at Porcupine Bank, appeared to be a new species *(Microichthys grandis;* Fricke and Couperus, 2023).

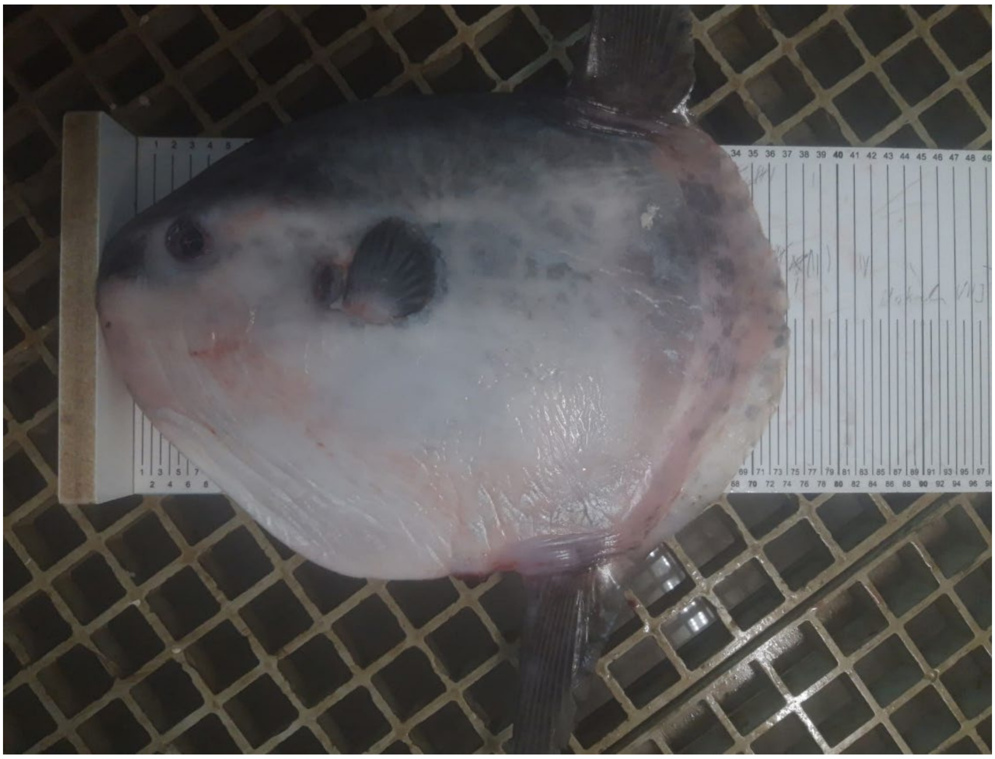


Figure 21 Large specimens of incidentally caught bony fish are often easy to identify. This sunfish (*Mola mola*) was light enough to be measured on the measuring board.

Turtles

In addition to a picture from the side, for turtles it is very helpful to have a dorsal picture, giving view on the plates of the head and the shell (Figure 22).

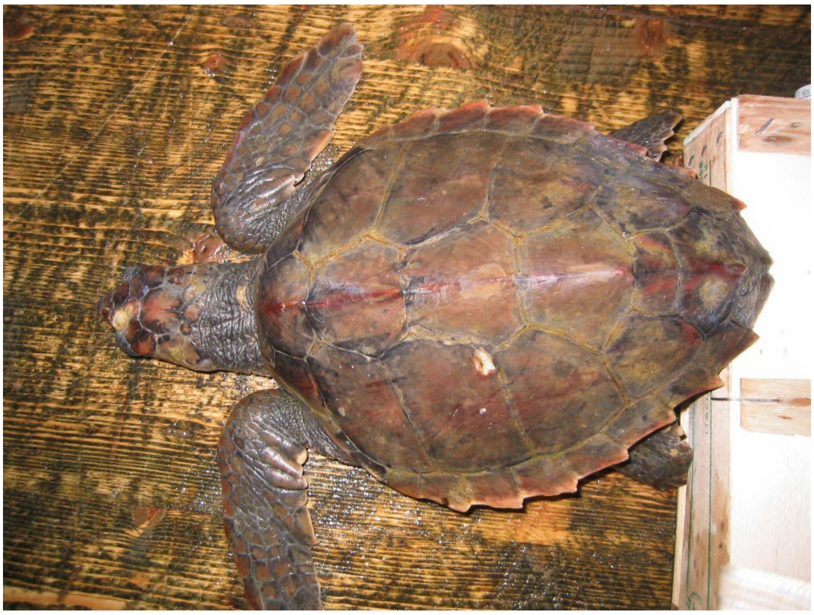


Figure 22 Loggerhead turtle (*Caretta caretta*)

Seals

In addition to a picture with a lateral view of the complete specimen. A close-up of the head makes distinction possible between several species, particularly between harbour seal and grey seal (Figure 23).

Figure 23 Full length seal with a ruler. For seals it is recommended to take a close-up picture of the head from the side to help with identification.

Cetaceans

A picture from the side with a reference measure is usually sufficient to identify the species. For the establishment of the gender, a picture should be taken of the anal region.

Birds

General shape and size of the animal, plumage colour, size and shape of the bill, or feet colour are key features for bird species identification. Photos of both sides of a stretched wing are very useful to determine age for many species (Figure 24, Figure 25).

Figure 24 Common eider male adult. Clear view of the plumage, interior of the wing (incl. the state of feathers), head, beak, and feet permit identification with 100% certainty. Source: DTU Aqua.

Figure 25 Multispecies bycatch (all were captured in the same haul). Specimen can easily be identified to species and sexed (for the ducks) from the picture, but a close-up of each individual with a stretched wing may help collect more additional information on the age of the bird. Source: DTU Aqua.

5 Make recommendations for improving monitoring systems for ETP bycatch at a Member State level and for regional level coordination. Amongst others, it should include proposals for adjusting DCF sampling to cover all PETS bycatch relevant fisheries.

5.1 Conclusions regarding design of bycatch monitoring programmes

There is not one monitoring design that is universally better to achieve a precise and accurate BPUE estimate. Overall, in many instances, increasing the number of vessels monitored tends to increase accuracy and precision, even if it means sampling fewer fishing operations per vessel because of limits on the total monitoring effort which can be achieved.

The bycatch probability of the sensitive species, for which the observed BPUE can be a useful indicator, play an important role in shaping the precision and accuracy which can be achieved. This may seem like a trivial result of this work, but it is an underappreciated limiting factor to what can be achieved with any monitoring programme. It is overall very challenging to obtain precise and accurate BPUE for species which are accidentally caught rarely. However, this might be not necessary in some case at the extremes of the 'species' conservation status' boundaries. Low BPUEs can emerge from two opposite scenarios. First, in instances when the density of the species is very low, be it because of a poor conservation status (very small populations). Second, when the species does not have a challenged conservation status, is relatively abundant in the region, but it is not prone to bycatch (at all or in certain fishing gears) or when this species is exposed at the margins of its range. In both cases, it may be useful to focus monitoring sampling design with an observation effort at sufficient level to detect whether bycatch occurs or not (e.g. around 10%) and produce a rough estimate, rather than aim to an unachievable estimate precision that would require, for example, over 50% (if not full) coverage. This rough estimate (even semi-quantitative) together with information on the conservation status of that species/population will suffice to identify the most suitable mitigation measure, if necessary. For example, in the first case (small population/poor conservation status), after the preliminary monitoring study, engaging directly with robust mitigation measures would seem a sensible trade-off. In the second case, mitigation measures could be milder or none.

It is important to consider the above extreme examples as the first step of a triage protocol for ETP species bycatch monitoring. More frequently monitoring programmes are expected to produce BPUE estimates for multiple species characterized by bycatch rates of different orders of magnitude. Relaxing the target for very rarely accidentally caught species can ensure that a consensual monitoring programme can be achieved where the coverage and its distribution can aim to retrieve the appropriate insights about bycatch for all the species of concern. This is important because different species may require different stratification strategies. The case studies we simulated in this work show that bycatch probability influences whether monitoring should be distributed over many vessels or be concentrated over fewer vessels. This is before we introduced realistic between-vessel variance in fishing behaviour that influence bycatch probability (the known unknown fishing characteristics influences on bycatch).

When we have 'known known' fishing characteristics that influence bycatch probability (differences among métiers/gear types), stratification is not always a beneficial approach. Stratification by métier will yield more precise and accurate BPUE estimates for species with very low bycatch probability. However, improvements in estimation with stratification are more negligible as the bycatch probability increases. Importantly, it appears that stratification is particularly valuable if you have to distribute the total monitoring coverage over fewer vessels (stratified by métier still) and therefore increase the number of fishing operations monitored for each vessel.

The precision of bycatch per unit effort estimates is sensitive to vessel effects, whether these are due to spatial, temporal, or spatio-temporal variation in effort.

A reference fleet, given that the reference fleet include the full range of vessel-specific variation that practitioners believe exists, might be an effective way to collect data on bycatches, especially in cases where refusal rates in observer programmes could be high.

Overall, engaging with this simulation exercise can be a useful way for monitoring programme managers to identify the key characteristics that need to be thought about when distributing monitoring effort, and even to know from first principle whether precise and accurate BPUE estimates can be achieved with the budget available to fund the total monitoring coverage.

Long-term data collection activities generate cumulative growth in knowledge, improving our understanding of the details and trends in what are often rare and unpredictable occurrences. These activities also facilitate multi-annual analytical approaches, which can be employed in some situations to augment the data available for bycatch assessments of large, long-lived species with low fecundity, such as many ETP species. In situations where bycatch rates lack high inter-annual variability, multi-annual approaches enhance the precision of bycatch estimates. The combination of long-term data collection and multi-annual analytical approaches is also beneficial for situations where a species exhibits low bycatch rates due to very low abundance or low susceptibility to capture in fishing gears, potentially going unrecorded or reaching unusable levels in shorter-term studies or annual analytical approaches.

WKPETSAMP3 also developed flowcharts assisting managers in setting up an ETP species bycatch monitoring programmes (single and multi-species) also taking into consideration the link between the species' bycatch probability and its conservation status (Figure 26). The latter flowchart was applied to two case studies (case study C and E) to show concrete examples on how it can work (Figure 27-Figure 28).

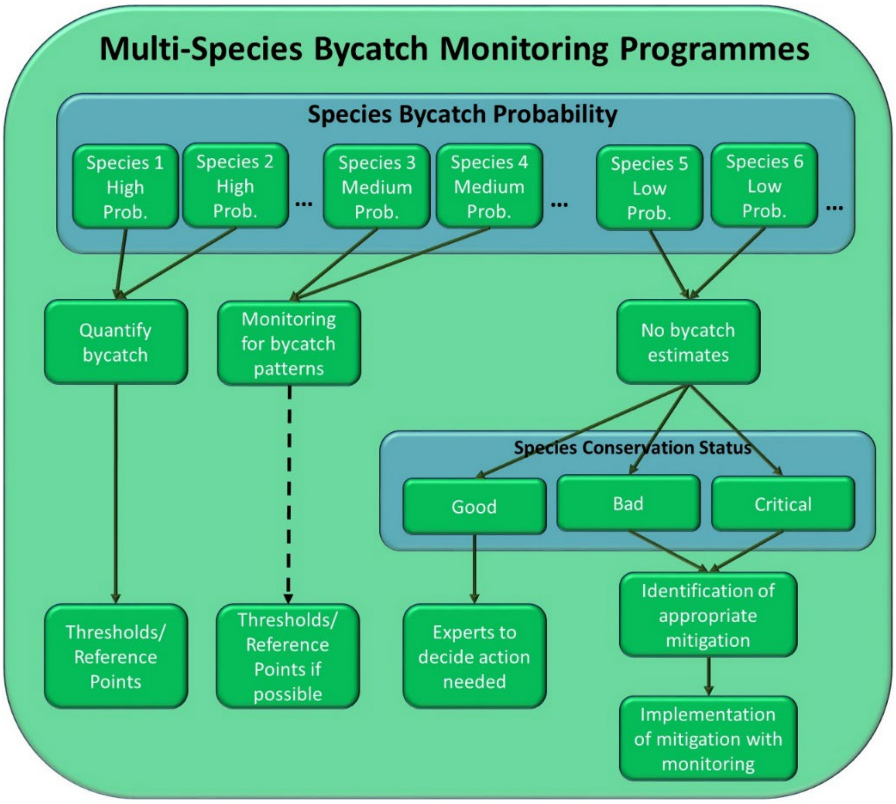


Figure 26 Decision-tree for setting up a multi-species bycatch monitoring programme according to the species bycatch probability and conservation status

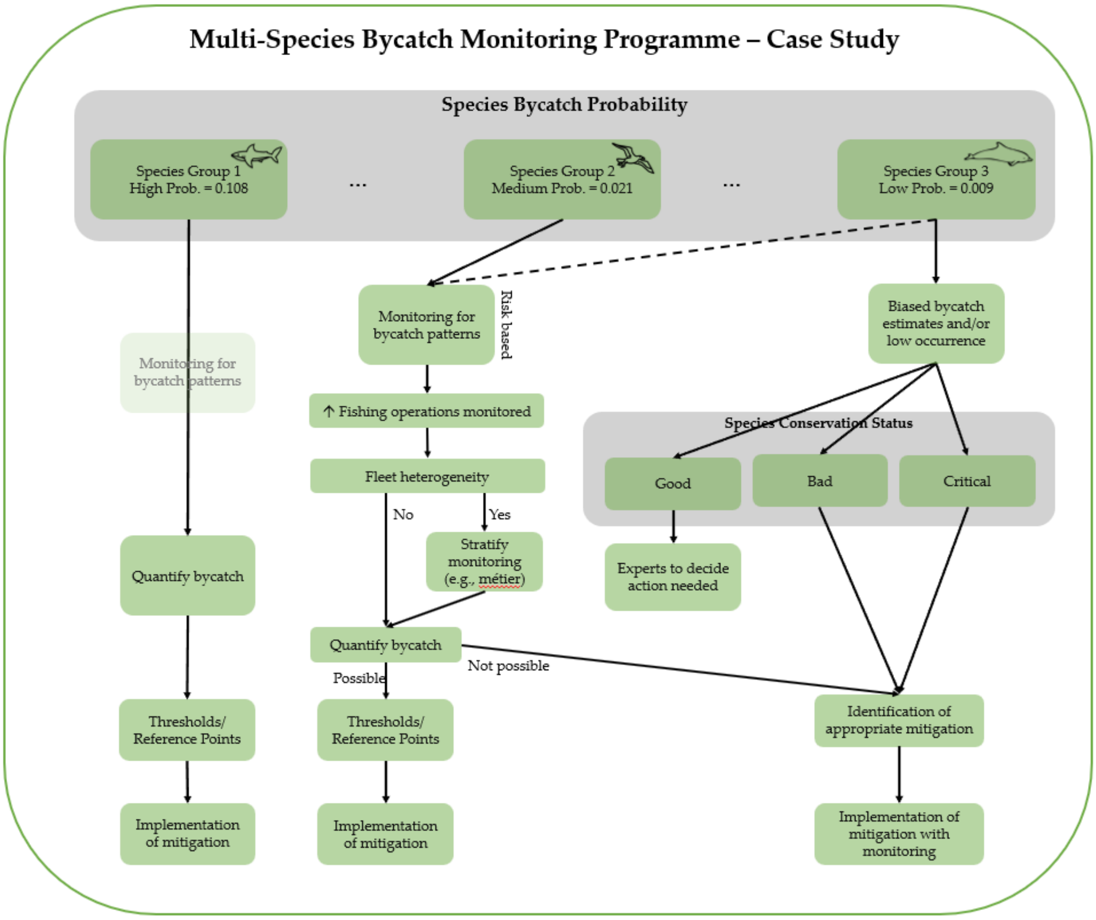


Figure 27 Flowchart based on Case study C, representing a multi-species bycatch monitoring programme. The differences in the bycatch probability of the different groups of species, which in turn can be influenced by the species occurrence (abundant, rare, etc.) will affect to make more informed decisions.

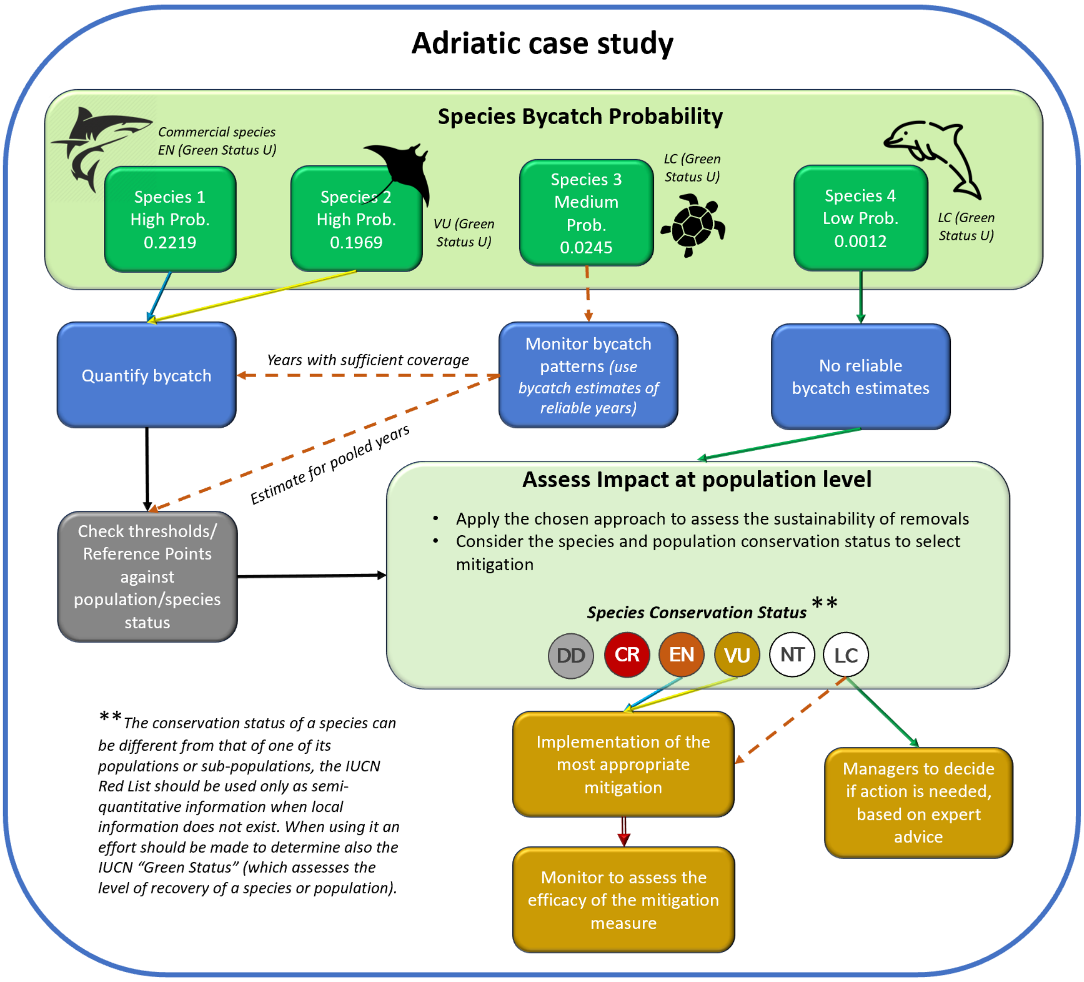


Figure 28 Flowchart based on Case study E, representing a multi-species bycatch monitoring programme. The differences in the bycatch probability of the different groups of species, which in turn can be influenced by the species occurrence (abundant, rare, etc.) will affect to make more informed decisions.

5.2 Recommendations

WGBYC is currently carrying out risk-based bycatch assessments of fisheries. WKPETSAMP3 recommends that ICES WGBYC continues to carry out bycatch risk assessments in all ICES and European union waters and rank which fisheries/métiers to be monitored at relevant spatial scale, taking into account the species ecological range. Where deemed necessary member states should carry out risk assessment for species, fisheries or areas at national level, if such scale it still appropriate for the concerned species.

WKPETSAMP3 recommend that the Regional Coordination Groups (RCGs) of the Data Collection Framework (DCF) ensure that fisheries/métiers identified as high risk, by for example by ICES WGBYC, are sufficiently covered by sampling.

WKPETSAMP3 recommend that DCF, RCGs have the responsibility to ensure that sampling plans and protocols, where appropriate, are coordinated regionally.

WKPETSAMP3 recommend sampling programmes to set up data quality procedures specifically for species identification and other key parameters collection (e.g. sex, age as informed by other proxies such as length or weight), based on documentation with pictures to be used for validation of sampling data.

WKPETSAMP3 recommend that expert groups (e.g. WGMME, JWGBIRD, WGEF) collate and maintain data on key biological and demographic parameters for bycatch-sensitive species to inform population model and threshold definition.

References

Anderson, Sean C., Eric J. Ward, Philina A. English, and Lewis A. K. Barnett. 2022. "sdmTMB: An R Package for Fast, Flexible, and User-Friendly Generalized Linear Mixed Effects Models with Spatial and Spatiotemporal Random Fields." Preprint. Ecology. https://doi.org/10.1101/2022.03.24.485545.

Arimitsu, Mayumi L., John F. Piatt, James T. Thorson, Katherine J. Kuletz, Gary S. Drew, Sarah K. Schoen, Daniel A. Cushing, Caitlin Kroeger, and William J. Sydeman. 2023. "Joint Spatiotemporal Models to Predict Seabird Densities at Sea." *Frontiers in Marine Science* 10. https://www.frontiersin.org/articles/10.3389/fmars.2023.1078042.

Authier, Matthieu, Etienne Rouby, and Kelly Macleod. 2021. "Estimating Cetacean Bycatch From Non-Representative Samples (I): A Simulation Study With Regularized Multilevel Regression and Post-Stratification." *Frontiers in Marine Science* 8 (October): 719956. https://doi.org/10.3389/fmars.2021.719956.

Bi, Rujia, Yan Jiao, and Joan A. Browder. 2021. "Climate Driven Spatiotemporal Variations in Seabird Bycatch Hotspots and Implications for Seabird Bycatch Mitigation." *Scientific Reports* 11 (1): 20704. https://doi.org/10.1038/s41598-021-00078-z.

Dillingham, P.W., Moore, J.E., Fletcher, D., Cortés, E., Curtis, K.A., James, K.C., Lewison, R.L., 2016. Improved estimation of intrinsic growth rmax for long-lived species: integrating matrix models and allometry. Ecological Applications 26, 322–333. https://doi.org/10.1890/14-1990

Fox, Ch, C Robertson, Pd O'Hara, R Tadey, and Kh Morgan. 2021. "Spatial Assessment of Albatrosses, Commercial Fisheries, and Bycatch Incidents on Canada's Pacific Coast." *Marine Ecology Progress Series* 672 (August): 205–22. https://doi.org/10.3354/meps13783.

Addink, M.; García Hartmann, M.; Couperus, B. 1997. A note on life-history parameters of the Atlantic white-sided dolphin (*Lagenorhynchus acutus*) from animals bycaught in the northeastern Atlantic. Rep. Int. Whal. Commn. 47, 637–639.

Carretta, J. V. 2021. Estimates of marine mammal, sea turtle, and seabird bycatch in the California large-mesh drift gillnet fishery: 1990-2019. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-SWFSC-654. https://doi.org/10.25923/7emj-za90

Castrillon, J, Bengtson Nash, S. Evaluating cetacean body condition; a review of traditional approaches and new developments. Ecol Evol. 2020; 10: 6144–6162. https://doi.org/10.1002/ece3.6301

Brandon, J. R., Punt, A. E., Moreno, P., Reeves, R. R. 2017. Toward a tier system approach for calculating limits on human-caused mortality of marine mammals. ICES Journal of Marine Science, 74(3): 877-887.

Dillingham, P.W.; Moore, J.E.; Fletcher, D.; Cortes, E.; Curtis, K.A.; James, K.C.; Lewison, R.L. 2016. Improved estimation of intrinsic growth rmax for long-lived species: integrating matrix models and allometry. Ecological Applications 26 (1), 322–333.

Fricke, Ronald, and Bram Couperus. 'Microichthys Grandis, a New Species of Deepwater Cardinalfish from off Ireland, Northeastern Atlantic Ocean (Teleostei: Epigonidae)'. *Ichthyological Research* 71, no. 1 (1 January 2024): 40–47. https://doi.org/10.1007/s10228-023-00909-1.

Geraci, J.R., St. Aubin, D.J. 1987. Effects of parasites on marine mammals. International Journal for Parasitology, Volume 17(2):407-414, ISSN 0020-7519. https://doi.org/10.1016/0020-7519(87)90116-0

Good, Sd, M Gummery, S McLennan, K Dewar, Sc Votier, and Ra Phillips. 'Evaluating the Appropriateness of Risk-Based Approaches to Assess the Sustainability of Fishery Impacts on Seabirds'. *Endangered Species Research* 51 (6 July 2023): 161–72. https://doi.org/10.3354/esr01251.

Hammond, P.S., Paradinas, I. & Smout, S.C. 2019. Development of a Removals Limit Algorithm (RLA) to set limits to anthropogenic mortality of small cetaceans to meet specified conservation objectives, with an example implementation for bycatch of harbour porpoise in the North Sea. JNCC Report No. 628, JNCC, Peterborough, ISSN 0963-8091

ICES 2020. Working Group on Commercial Catches (WGCATCH). International Council for the Exploration of the Sea (ICES). ICES Scientific Report Vol. 2 No. 66 https://doi.org/10.17895/ices.pub.7428

ICES 2022. Working Group on Bycatch of Protected Species (WGBYC). ICES Scientific Reports. Report. https://doi.org/10.17895/ices.pub.21602322.v1

ICES 2023. NEAFC request on seabird bycatch in the NEAFC regulatory areas (RAs). ICES Advice: Special Requests. Report. https://doi.org/10.17895/ices.advice.24204291.v1

ICES 2024. Working Group on Bycatch of Protected Species (WGBYC). ICES Scientific Reports. Available at: https://doi.org/10.17895/ICES.PUB.24659484

Irschick, D.J. and Hammerschlag, N. 2014. A new metric for measuring condition in large predatory sharks. J Fish Biol, 85: 917-926. https://doi.org/10.1111/jfb.12484

Jordan A. Thomson, Derek Burkholder, Michael R. Heithaus, Lawrence M. Dill; Validation of a Rapid Visual-Assessment Technique for Categorizing the Body Condition of Green Turtles (*Chelonia mydas*) in the Field. Copeia 1 June 2009; 2009 (2): 251–255. doi: https://doi.org/10.1643/CE-07-227

Marchowski, Dominik, Łukasz Jankowiak, Łukasz Ławicki, Dariusz Wysocki, and Przemysław Chylarecki. 'Fishery Bycatch Is among the Most Important Threats to the European Population of Greater Scaup Aythya Marila'. *Bird Conservation International* 30, no. 2 (June 2020): 176–93. https://doi.org/10.1017/S0959270919000492

Merilä, J., Kruuk, L.E.B. and Sheldon, B.C. (2001), Natural selection on the genetical component of variance in body condition in a wild bird population. Journal of Evolutionary Biology, 14: 918-929. https://doi.org/10.1046/j.1420-9101.2001.00353.x

Regulation (EU) 2017/1004 of the European Parliament and of the Council of 17 May 2017 on the establishment of a Union framework for the collection, management and use of data in the fisheries sector and support for scientific advice regarding the common fisheries policy and repealing Council Regulation (EC) No 199/2008, OJ L. Available at: http://data.europa.eu/eli/reg/2017/1004/oj/eng (Accessed: 24 February 2020)

Richard, Yvan, and Edward R. Abraham. *Application for Potential Biological Removal Methods to Seabird Populations*. Ministry for Primary Industries, 2013

Roberts, J.O., D.N. Webber, W.D. Roe, C.T.T. Edwards, and I.J. Doonan. 'Spatial Risk Assessment of Threats to Hector's and Māui Dolphins (Cephalorhynchus Hectori)'. New Zealand Aquatic Environment and Biodiversity Report. WELLINGTON: Fisheries New Zealand, 2019. https://fs.fish.govt.nz/Doc/24714/AEBR-2019-214-Spatial-risk-assessment-Maui-Hectors-dolphin.pdf.ashx

Valverde, I., Espín, S., María-Mojica, P. et al. 2020. Protocol to classify the stages of carcass decomposition and estimate the time of death in small-size raptors. Eur J Wildl Res 66, 93. https://doi.org/10.1007/s10344-020-01429-3

Verutes, Gregory M., Andrew F. Johnson, Marjolaine Caillat, Louisa S. Ponnampalam, Cindy Peter, Long Vu, Chalatip Junchompoo, Rebecca L. Lewison, and Ellen M. Hines. 'Using GIS and Stakeholder Involvement to Innovate Marine Mammal Bycatch Risk Assessment in Data-Limited Fisheries'. *PLOS ONE* 15, no. 8 (20 August 2020): e0237835. https://doi.org/10.1371/journal.pone.0237835

Wade, Paul R. 'Calculating Limits to the Allowable Human-Caused Mortality of Cetaceans and Pinnipeds'. *Marine Mammal Science* 14, no. 1 (1998): 1–37. https://doi.org/10/dxjhws

Annex 1: Literature review of criteria and quality thresholds applied to bycatch monitoring

Introduction

One of the tasks of WKPETSAMP2 was to "review relevant scientific literature where criteria and data quality thresholds have been applied to PETS bycatch monitoring data in order to derive bycatch mortality estimates". At the beginning of the WKPETSAMP2 and the PETSAMP3 workshops, members who are subject matter experts were asked to upload potentially relevant reports and publications for consideration. Thirty documents were submitted and reviewed. If the submitted publications or reports had referenced literature relevant to the task, a selecting of these were also been reviewed. The review process focused on criteria and data quality thresholds.

To determine population level consequences of bycatch mortality, reliable and precise estimates of bycatch mortality with reasonable confidence intervals are needed (ICES advice 2021). In order to achieve this, estimates of bycatch need to be improved, associated bias minimised, and precision estimated accurately. One way to achieve this is by providing guidance on implementation, including common sources of estimation bias and data analysis (*e.g.* ICES 2019; Murphy et al. 2022, Moore et al. 2021).

Generally, four types of report and publications were reviewed; one that specified criteria and threshold values required to achieve a predetermined objective (*e.g.* Babcock and Pikitch, 2003), those that provide tools to allow users to explore the criteria associated with estimating bycatch for their own monitoring programs (*e.g.* Curtis and Caretta 2020, those that have described best practices for achieving unbiased estimates of bycatch (*e.g.* Moore et al. 2021, Wade et al. 2021), and those that focus on case-specific approaches to estimating bycatch (*e.g.* Jannot et al. 2021).

When considering the following reviews, it should be noted that in cases where criteria and threshold values with associated coverage are case-specific, they will not directly apply to studies where data is collected under different circumstances (Babcock and Pikitch, 2003.

Sampling programme design

Bycatch sampling programme design includes the consideration of the criteria and data quality thresholds that are required to estimate reliable bycatch mortality estimates. Sampling bias can be related to sampling coverage and the methods that are used for selecting the primary sampling units such as trips and vessels, among other factors (*e.g.* ICES Advice 2021; Murphy et al. 2022). In addition to achieving proportional monitoring coverage, it is also crucial to know the proportion of the unique vessels and whether the monitored vessels were selected randomly. There might be a need to stratify sampling in order to achieve decreased bias. However, selecting high-risk métiers and areas for monitoring can result in positively biased estimates (*e.g.* ICES Advice 2021; Murphy et al. 2022).

Currently, dedicated onboard observers and EM devices are considered to provide the most robust and reliable bycatch estimates (e.g. Jannot et al. 2021; Moore et al. 2021; Wade et al. 2021). However, this presumes random sampling, and in practice this is not always achieved. For example, many observer programs allocate sampling effort opportunistically to vessels that volunteer to carry observers.

Other methods that can be used to collect data and identify if bycatch occurs include self-reporting by fishers, interviews of fishers, and stranding data (e.g. Moan *et al*, 2020; Wade *et al*., 2021, Moore *et al*, 2010; Pardalou and Tsikliras, 2018; Peltier *et al*, 2016, 2020). However, the reliability of self-reported data is considered limited due to issues with factors such as low reporting rates and misidentification of species (Moore *et al*., 2021) . Interviews with fishers conducted at the dockside can gather information regarding bycatch of protected species (Moore *et al*., 2021). Data collected during interviews, particularly those led by individuals recognized by the fishing community, can provide significant insights into bycatch of fisheries where no other information is available (*e.g.* Moore *et al*, 2010; Pardalou and Tsikliras, 2018). Additional data source include information collected from stranded animals which have been used to identify and in some cases quantify bycatch (*e.g.* Moore and Read, 2008; Williams *et al*, 2011; Carretta *et al*, 2016; Peltier *et al*, 2016, 2020). Finally, these methodologies can be combined; to improve bycatch estimates, for example, using fisher interviews and stranding data (*e.g.* Mustika *et al*. 2014), and bycatch estimates have been derived from reference fleets combining self-sampling with data collected on the dockside, as used as a sample program in Norway (Clegg and Williams, 2020; Moan *et al*, 2020).

Sampling design and coverage

A general recommendation for sampling design is that a fishery should be sampled randomly or systematically. This sampling should cover all components of the fishery, simultaneously allocating coverage levels that are high enough to adequately sample every stratum of the stratified sampling design (Babcock and Pikitch 2003). Importantly, levels of coverage and objectives should be defined by the management goals of the fishery being monitored, with the level of observation needed to achieve these goals depending on the fishery characteristics (NMFS 2004). Examples of objectives include, monitoring catch/effort for stock assessment, bycatch monitoring for management/stock assessment, ETP species monitoring, technical monitoring (performance of fishing gear or mitigation) and compliance monitoring (NMFS 2004). There are no universal or legally defined required levels for observer coverage by fishery or area, although there have been guidelines set out in various studies (*e.g.* REFS). For rare species, a coverage of above 50% has been suggested, unless managers can show that lover levels of coverage give sufficient precision and accuracy (Babcock and Pikitch, 2003; Pierre, 2009).

Tools are available for assessing the needed observer coverage to document and estimate rare event bycatch (*e.g.* Curtis and Carretta 2020). Such tools can be used as a standalone analysis, but it is recommended to combine them with a risk assessment approach to assess impacts on vulnerable species that may not be observed in bycatch but that may interact with the fishery and where there is very low (<10%) observer coverage (Curtis and Carretta 2020).

A literature review and simulation study conducted by Babcock and Pitkitch (2003) suggested that coverage levels of at least 20 percent for common species, and 50 percent for rare species, could give a reasonable estimates of total bycatch, if the observer samples are an unbiased sample of the fishery. In regions where minimal mortality rates pose a threat to the recovery of Endangered, Threatened, or Protected (ETP) species, Babcock and Pitkitch (2003) advocates 100% coverage. However, this coverage may be seasonal, such as during periods when for example endangered species such as right whales are calving.

In specific instances studies have estimated coverage needs for particular species/area combinations. A study on turtles in US longline fisheries suggested increasing coverage from 4 to 12% with the purpose of monitoring the effectiveness of measures and if they were being implemented (Dietrich *et al*. 2004). Similarly, the optimal level of coverage for ETP species in longline fisheries was evaluated by Debski *et al*. (2016). It was concluded that a reasonable precise estimate of seabird bycatch requires a coverage level of 20% of hooks observed, although levels of over 2.5 times that would be required to detect captures of rare species. Similarly, 20% coverage has been recommended in many other communications (*e.g.* ICCAT 2016; IATTC 2017; Anonymous 2015). If a seabird is occasionally captured but caught in large numbers when they are caught, then higher levels of coverage will be required to obtain a specified level of precision. Controversially, a species often captured but in low numbers per event will require less coverage for the same level of precision. Other factors that influence bycatch estimates are seabird abundance and behavior, vessel behavior and mitigation use among others.

Relevant results from a study on what factors affect the precision of the length of the catch, which is of importance when assessing for example fish, revealed that the catch standard error was reduced when the sampling occurs across a number of different vessels and the number of samples taken per trip (Pennington and Helle, 2011). The number of fish taken per sample had little effect (Pennington and Helle, 2011).

An increase in monitoring effort should ideally be distributed among fleets in a representative manner (Pierre, 2009). A study on bycatch of marine mammals, fish turtles and elasmobranchs in North Pacific albacore tuna (*Thunnus alalunga*) fishery using longlines or purse seines included simulations on how precision in catch rates is affected by varying observer level coverage. The simulations indicated estimates would be improved by monitoring a proportion of fishing sets from all fishing trips rather than full coverage from a proportion of all fishing trips. For example, precision would be improved by having a 10% coverage of sets from all trips, rather than covering all sets from 10% of trips. Moreover, for frequently caught species, coverage of 10% of sets from all trips obtained more precise estimates than full coverage of 20% of trips (Peatman *et al*. 2023).

Statistically modelling to overcome bycatch data challenges

Bycatch mortality estimates that are based only on the collected data are considered minimum estimates. For more complete estimates, the unobserved mortality should be estimated and included in mortality estimates (*e.g.* Carretta *et al*. 2004). There are several statistical modelling methodologies available that can account for the use of different types of data in different situations, to estimate unobserved and observed bycatch in fisheries that are partially monitored. For example, regression trees can result in more stable annual bycatch estimates with better precision than e.g. ratio estimators, because the models are informed using all available data (Carretta 2021). Multilevel modeling (Authier *et al*. 2021) is a flexible modelling approach that can accommodate a variety of different information such as spatial and random effects. This method is recommended when estimating bycatch rate and numbers (Authier *et al*. 2021). Similarly, bayesian models can be effective in estimating rare bycatch events as they can yield probabilities for unobserved entanglements (Zhou et al., 2019; Jannot et al. 2021; Pirotta et al. 2023 [†]). When data are rare and over-dispersed, more complex distributions need to be employed to estimate bycatch and uncertainty more accurately (Jannot *et al*. 2021).

Simulation models can also be very useful in informing sampling programme design and consideration of criteria and data quality thresholds to be used or that can be achieved. For example, when bycatch events are rare, Carretta and Moore (2014) recommended pooling a number of years of observation data to achieve more precise bycatch estimates. The number of years to pool varied depending on the management goals; in general, adding more years was recommended, particularly in the case of rare-event bycatch (Carretta and Moore 2014). However, prior to pooling data, the likelihood that characteristics of the fishery of concern have remained sufficiently constant need to be considered (Carretta and Moore 2014).

There are studies that have looked into the possibility of using assessments from data-rich species to inform assessments of data-poor species (the "Robin Hood" approach) (*e.g.* Smith *et al*.,

[†] References added based on reviewers comments.

2009; Punt *et al*., 2011). Smith *et al*., (2009) set up general recommendations using case studies from two multispecies trawl fisheries. The recommendations included the possibility of using the "Robin Hood approach" to assess data-poor species and that for some data-poor species, sufficient data may never be available to enable quantitative assessments to be conducted and adopting a sufficiently precautionary approach may be the only realistic way to manage some data-poor species.

Data limited situations

When data on bycatch are unavailable or limited, alternative methodologies are available to assess risk of bycatch both spatially and temporally to guide fisheries management and spatial planning toward species conservation. At a most basic level, bycatch risk assessments consider the spatio-temporal overlap between fishing effort and the distribution of species to identify high-risk gear (*e.g.* Hines et al., 2020; Wade *et al*. 2021). Other information, such as strandings, interviews, port-observers and landings can also be incorporated into these assessments (Wade *et al*. 2021), however, associated biases need to be considered when estimating bycatch using data recorded with these methodologies. By setting initial priorities, a rapid risk assessment can be done to identify if bycatch occurs (Wade *et al*. 2021). These approaches can determine bycatch risk for species, areas, and fishing métiers, and can then be correlated with current monitoring coverage and species abundance and distribution information, to identify the overall bycatch risk, highlighting sampling needs, and identifying gaps in monitoring programmes. Bycatch risk assessment methodologies are documented in a number of locations, including throughout the scientific literature (*e.g.* Hines *et al*. 2020), and in reports such as FishPi and those of the ICES Working Group on Bycatch (MASTS 2016; ICES 2020; ICES 2021; ICES 2022).

A method to broadly assess the bycatch risk was developed within the FishPi Project identifying fisheries and fishing areas where bycatch of PET species was considered a conservation threat. The results of the project provided relative indices of what species, areas, and métiers are most at risk of significant bycatch in areas across the NE Atlantic and for multiple species groups. Since 2020, ICES WGBYC have identified fishery métiers (métier Level 4 and ICES Division) that are relatively under-sampled with respect to PETs bycatch, as a way of informing coordinated sampling plans, and used métier specific risk index scores produced within the FishPi project, and data on fishing and monitoring effort from the WGBYC database to provide an overview of how sampling coverage is related to the FishPi relative risk scores.

The ByRa (Bycatch Risk Assessment) is a freely downloadable modelling framework which supports modelling of various input data into GIS layers for analysis and modelling of the physical environment, spatial and temporal patters of animal distribution and movement, fishing effort and gear distribution (Hines *et al*. 2020). The tools can also assess the risk in different seasons throughout the year (Hines *et al*. 2020).

Conclusions from the literature review

- Onboard observers and to some extent electronic monitoring are considered most robust monitoring methods, as such these methodologies should be prioritized to monitor bycatch where possible. However, in order to balance the scope and detail of the data that can be collected by observers a regionally agreed multi-purpose sampling protocols and manuals need to be developed and agreed.
- In some specific instances, a combination of monitoring methods, can be used to produce bycatch estimates and collect relevant biological parametres. An example of this

is the use of self-sampling and dockside collection of bycatch samples by port observers in the Norwegian Reference fleet.
- When data on bycatch are lacking and it is not possible to generate estimates of bycatch mortality, alternative methodologies such as bycatch risk assessments can be used to assess risk of bycatch both spatially and temporally to guide fisheries management and spatial planning toward species conservation.
- Alternative methods to highlight spatial and temporal risk of bycatch include anecdotal information such as interview studies or data collected from strandings.
- Tools are available to assess observer coverage requirements to document and estimate bycatch with specified criteria and data quality thresholds.
- Observer coverage levels of 5% may be adequate to collect information identifying bycatch risks and issues in some instances and for some species. However, this level of coverage is likely insufficient for effectively quantifying bycatch of rarer species such as seabirds. Higher levels of coverage is required to obtain a specified level of precision when species are bycaught seldom but in large numbers
- Observer coverage should aim to be representative, taking into consideration factors such as seasonality of fishing, between-vessel variation within a fishery, timing of sets, and location of fishing activities.
- In the case of rare bycatch events/species, research recommends pooling of a number of years of observation data to achieve more precise bycatch estimates (Carretta and Moore, 2014).
- Bycatch estimates could be improved by monitoring a proportion of fishing sets from all fishing trips rather than full coverage from a proportion of all fishing trips (Peatman *et al*. 2023).
- Control rules designed for data-poor species should acknowledge that certain species may never have enough data for quantitative assessments. In these cases, there exists a trade-off between the expenses associated with data collection and the economic value of the fishery. Adopting a sufficiently precautionary approach may be the only practical means of managing certain low-value, data-poor species (Smith *et al*,. 2009).

References

Anonymous. (2015). Tuna RFMO Expert Working Group: Harmonisation of Longline Data Collected by Tuna RFMOs. ISSF Technical Report 2015-08. International Seafood Sustainability Foundation, Washington, D.C., USA

Authier, M., Rouby, E., and Macleod, K. (2021). Estimating cetacean bycatch from non-representative samples (I): A simulation study with regularized multilevel regression and post-stratification. Frontiers in Marine Science, 1459.

Babcock E. A. and Pikitch E. K. (2003). How much observer coverage is enough to adequately estimate bycatch? Unpublished report, Pew Institute for Ocean Science and Oceana, 2003.

E.A. Babcock, E.K. Pikitch, C.G. Hudson. 2003. How Much Observer Coverage Is Enough to Adequately Estimate Bycatch? Page University of Miami (2003). Available from: http://oceana.org/sites/default/files/reports/BabcockPikitchGray2003FinalReport1.pdf,

Carretta, J. V., Price, T., Petersen, D., and Read, R. (2004). Estimates of marine mammal, sea turtle, and seabird mortality in the California drift gillnet fishery for swordfish and thresher shark, 1996–2002. Marine Fisheries Review, 66(2), 21-25.

Carretta, J. V. and Moore, J. E. (2014). Recommendations for Pooling Annual Bycatch Estimates When Events are Rare. NOAA Tech. Memo. NOAA-TM- NMFS-SWFSC-528. La Jolla, CA.

Carretta, J. V., Danil, K., Chivers, S. J., Weller, D. W., Janiger, D. S., BermanKowalewski, M., *et al*. (2016). Recovery rates of bottlenose dolphin (Tursiops truncatus) carcasses estimated from stranding and survival rate data. Mar.Mamm. Sci. 32, 349–362. doi: 10.1111/mms.12264

Carretta, J. V. (2021). Estimates of marine mammal, sea turtle, and seabird bycatch in the California large-mesh drift gillnet fishery: 1990-2019. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-SWFSC-654.

Clegg, T. and Williams, T. (2020). Monitoring Bycatches in Norwegian Fisheries. Species registered by the Norwegian Reference Fleet 2015-2018 (IMR) RAPPORT FRA HAVFORSKNINGEN NR. 2020-8. https://hdl.handle.net/11250/2685855

Curtis, K. A. And Carretta, J. V. (2020). ObsCovgTools: Assessing observer coverage needed to document and estimate rare event bycatch. Fisheries Research, 225, 105493.

Debski, I., Pierre, J., and Knowles, K. (2016) Observer coverage to monitor seabird captures in pelagic longline fisheries. WCPFC-SC12-2016/ EB-IP-07.

Dietrich, K. S., Cornish, V. R., Rivera, K. S., and Conant, T. A. (2007). Best practices for the collection of longline data to facilitate research and analysis to reduce bycatch of protected species: report of a workshop held at the International Fisheries Observer Conference, Sydney, Australia, November 8, 2004

Hines, E., Ponnampalam, L. S., Junchompoo, C., Peter, C., Vu, L., Huynh, T., ... and Verutes, G. M. (2020). Getting to the bottom of bycatch: a GIS-based toolbox to assess the risk of marine mammal bycatch. Endangered Species Research, 42, 37-57.

IATTC. (2017). Inter-American Tropical Tuna Commission 92nd meeting. Minutes of the Meeting. 24 – 28 July 2017. Mexico City, Mexico. Inter-American Tropical Tuna Commission.

ICCAT. (2016). Report of the Standing Committee on Research and Statistics (SCRS). Madrid, Spain. 3 – 7 October 2016. International Commission for the Conservation of Atlantic Tunas

ICES Advice. (2021). Bycatch of protected, endangered, and threatened species (PETS). ICES Advice. Ecoregions in the Northeast Atlantic and adjacent seas. Published 2 December 2021.

ICES. (2019). Joint WGBYC-WGCATCH Workshop on sampling of bycatch and PET species (WKPETSAMP), 24–26 April 2018, SLU Aqua, Lysekil, Sweden, ICES CM 2018/EOSG:35. 76 pp.

ICES. (2020). Working Group on Bycatch of Protected Species (WGBYC). ICES Scientific Reports. ICES Scientific Reports. Volume 2, Issue 81. 209 pp.

ICES. (2021). Working Group on Bycatch of Protected Species (WGBYC). ICES Scientific Reports. Volume 3, Issue 107. 175 pp.

ICES. (2022). Working Group on Bycatch of Protected Species (WGBYC). ICES Scientific Reports. Volume 4, issue 91. 265 pp.

Jannot, J. E., Ward, E. J., Somers, K. A., Feist, B. E., Good, T. P., Lawson, D., and Carretta, J. V. (2021). Using Bayesian models to estimate humpback whale entanglements in the United States west coast sablefish pot fishery. Frontiers in Marine Science, 8, 775187.

Lewin, W.-C., Freyhof, J., Huckstorf, V., Mehner, T., and Wolter, C. (2010). When no catches matter: coping with zeros in environmental assessments. Ecological Indicators 10, 572–583.

Martin, S. L., Stohs, S. M., and Moore, J. E. (2015). Bayesian inference and assessment for rare-event bycatch in marine fisheries: a drift gillnet fishery case study. Ecological Applications 25, 416–429. doi: 10.1890/14-0059.1

MASTS. (2016). fishPi - strengthening regional co-operation in fisheries data collection. Report of the Marine Alliance for Science and Technology for Scotland. Available from: https://www.masts.ac.uk/media/36266/fishpi-final-report.pdf

Moan, A., Skern-Mauritzen, M., Vølstad, J. H., and Bjørge, A. (2020) Assessing the impact of fisheries-related mortality of harbour porpoise (Phocoena phocoena) caused by incidental bycatch in the dynamic Norwegian gillnet fisheries. ICES Journal of Marine Science 77 (7-8), 3039-3049.

Moore, J. E., and Read, A. J. (2008). A Bayesian uncertainty analysis of cetacean demography and bycatch mortality using age-at-death data. Ecological Applications, 18(8), 1914-1931.

Moore, J. E., Cox, T. M., Lewison, R. L., Read, A. J., Bjorkland, R., McDonald, S. L., ... and Kiszka, J. (2010). An interview-based approach to assess marine mammal and sea turtle captures in artisanal fisheries. Biological Conservation, 143(3), 795-805.

Moore, J. E., Heinemann, D., Francis, T. B., Hammond, P. S., Long, K. J., Punt, A. E., ... and Zerbini, A. N. (2021). Estimating bycatch mortality for marine mammals: concepts and best practices. Frontiers in Marine Science, 8, 752356.

Murphy, S., Borges, L., and Tasker, M. (2022). External report on the review of monitoring PETS bycatch of mammals, birds, turtles and fish for ICES under the service of EC DG Environment. ICES Scientific Reports. 4:17. 69 pp.

Mustika, P. L. K., Purnomo, F. S., and Northridge, S. (2014). A pilot study to identify the extent of small cetacean bycatch in Indonesia using fisher interview and stranding data as proxies. Denpasar: Updated Report to the International Whaling Commission.

NMFS. (2004). NMFS Fisheries Observer Workshop – Defining a Basis, summary Report. July 29-31 2003, Alaska Fisheries Science Centre.

Pardalou, A., and Tsikliras, A. C. (2018). Anecdotal information on dolphin-fisheries interactions based on empirical knowledge of fishers in the northeastern Mediterranean Sea. Ethics in Science and Environmental Politics, 18, 1-8.

Peatman, T., Allain, V., Bell, L., Muller, B., Panizza, A., Phillip, N. B., Pilling, G., and Nicol, S. (2023). Estimating trends and magnitudes of bycatch in the tuna fisheries of the Western and Central Pacific Ocean. Fish and Fisheries, 24, 812–828.

Pirotta, Enrico , Schick, Robert S. , Hamilton, Philip K. , Harris, Catriona M. , Hewitt, Joshua , Knowlton, Amy R. , Kraus, Scott D. , Meyer-Gutbrod, Erin , Moore, Michael J. , Pettis, Heather M. , Photopoulou, Theoni , Rolland, Rosalind M. , Tyack, Peter L. , Thomas, Len. 2023. Estimating the effects of stressors on the health, survival and reproduction of a critically endangered, long-lived species. Oikos, vol. 2023, issue 5, id. e09801

H. Peltier, M. Authier, R. Deaville, W. Dabin, P.D. Jepson, O. van Canneyt, P. Daniel, V. Ridoux. 2016. Small cetacean bycatch as estimated from stranding schemes: the common dolphin case in the northeast Atlantic. Environ. Sci. Pol., 63 (2016), pp. 7-18, 10.1016/j.envsci.2016.05.004

H. Peltier, M. Authier, W. Dabin, Dars, C, Demaret, F, Doremus, G, Van Canneyt, O, Laran, S, Mendez-Fernandez, P, Spitz, J, Daniel, P, Ridoux, V. 2020. Can modelling the drift of bycaught dolphin stranded carcasses help identify involved fisheries? An exploratory study. Global Ecol. Conserv., 21 (2020), p. e00843, 10.1016/j.gecco.2019.e00843

Pennington, M., and Helle, K. (2011). Evaluation of the design and efficiency of the Norwegian self-sampling purse-seine reference fleet. ICES Journal of Marine Science, 68(8), 1764–1768.

Pierre, J. P. (2009). Observer coverage for monitoring bycatch of seabirds and other ETP species in pelagic longline fisheries. Ninth Meeting of the Seabird Bycatch Working Group. SBWG9 Inf 26 Rev 1. Agenda Item 13.1.

Punt A. E., Smith, D. C., and Smith A. D. M. (2011). Among-stock comparisons for improving stock assessments of data-poor stocks: the "Robin Hood" approach, ICES Journal of Marine Science, 68 (5), 972–981.

Punt, A. E., Siple, M., Francis, T. B., Hammond, P. S., Heinemann, D., Long, K. J., ... and Zerbini, A. N. (2020a). Robustness of potential biological removal to monitoring, environmental, and management uncertainties. ICES Journal of Marine Science, 77(7-8), 2491-2507.

Punt, A. E., Siple, M. C., Francis, T. B., Hammond, P. S., Heinemann, D., Long, K. J., ... and Zerbini, A. N. (2020b). Can we manage marine mammal bycatch effectively in low-data environments? Journal of Applied Ecology, 58(3), 596-607.

Siple, M. C., Punt, A. E., Francis, T. B., Hammond, P. S., Heinemann, D., Long, K. J., ... and Zerbini, A. N. (2022). mmrefpoints: Projecting long-term marine mammal abundance with bycatch. Journal of Open Source Software, 7(71), 3888.

Smith, D., Punt, A., Dowling, N., Smith, A., Tuck, G., and Knuckey, I. (2009). Reconciling Approaches to the Assessment and Management of Data-Poor Species and Fisheries with Australia's Harvest Strategy Policy. Marine and Coastal Fisheries, 1: 244-254.

Wade, P. R., Long, K. J., Francis, T. B., Punt, A. E., Hammond, P. S., Heinemann, D., ... and Zerbini, A. N. (2021). Best practices for assessing and managing bycatch of marine mammals. Frontiers in Marine Science, 8, 1566.

Williams, R., Gero, S., Bejder, L., Calambokidis, J., Kraus, S. D., Lusseau, D., *et al*. (2011). Underestimating the damage: interpreting cetacean carcass recoveries in the context of the Deepwater Horizon/BP incident. Conserv. Lett. 4, 228–233. doi: 10.1111/j.1755-263X.2011.00168.x

Zhou, C., Jiao, Y., and Browder, J. 2019. How much do we know about seabird bycatch in pelagic longline fisheries? A simulation study on the potential bias caused by the usually unobserved portion of seabird bycatch. PLoS ONE, 14: e0220797. https://doi.org/10.1371/journal.pone.0220797

Annex 2: General overview of the SCOTI software

The overall objective with the SCOTI software is to simulate a fishing fleet's bycatches in a representative manner and thereafter monitor the simulated fleet using different monitoring strategies. By varying fisheries characteristics (including its bycatch properties) and monitoring strategies, it is possible to ask questions as to how a specific bycatch monitoring strategy is appropriate or not for a given type of fishery.

The SCOTI software contains two main functions: a function for simulating fisheries and their associated bycatches (make_fishing_year_métier.R) and a monitoring function (monitor_BPUE_métier.R), which is used to monitor the fishery. The output of the fishery simulation function (make_fishing_year_métier) is a data.frame ("fishing") containing parameters describing the fishery and its general bycatch properties, along with simulated number of bycaught animals for each fishing event (number of bycaught individuals a given day of the year). By specifying a number of monitoring properties, such as the proportion of vessels to monitor and the proportions of fishing events to sample in a vessel, the monitor_BPUE_métier function estimates BPUE as well as the coefficient of variance of this estimate. These estimates are derived by randomly sampling (based on the monitoring properties being defined as input to the function) the fishing events from the fleet (the "fishing" data.frame) a predefined number of times and thereafter using the number of bycatches and sampling effort from these to derive mean BPUE and CV (BPUE) estimates.

The SCOTI simulation framework contains a script ("estimate_bias_and_precision.R") that can be used to test a large number of fishery(with associated bycatch properties)-monitoring combinations. In this script, a user first defines, in sub-lists of a list (fishery_params), unique parameter sets describing fisheries and their presumed bycatches properties. Second, a user defines, in another list with sub-lists (monitor_params), unique parameter sets describing the different ways in which the fisheries are to be monitored. Each monitoring strategy (defined in each parameter sub-list of monitor_params) is implement on each of the simulated fisheries (defined in each parameter sub-list of fishery_params). The script ("estimate_bias_and_precision.R") outputs a data.frame with columns describing the monitoring-fishery combinations as well as the true BPUE from the fishery simulation, and the estimated BPUE and CV of the BPUE estimate retrieved from the monitoring function.

R-code for running the SCOTI simulation tool is available in an open repository: https://github.com/dlusseau/scoti

<u>Detailed description of the main functions in the SCOTI software</u>

1.1 Fishing and bycatch simulation algorithm

In the SCOTI simulation tool, a fishery and its bycatches are simulated over 365 fishing days, that is, during one year. On each day, all boats in the fleet are assigned a random number of 'fishing events' (e.g. hauls). The mean number of fishing events conducted per boat and day can either be assumed different (stochastic=TRUE) or the same across vessels (stochastic=FALSE). If the mean number of fishing events per boat and day are assumed the same across vessels (stochastic=FALSE), the number of fishing events for a given boat and day is assigned randomly from:

$f_{i,d} \sim Pois(\lambda_f)$ where $d \in \{1,2,\dots,365\}$ and $i \in \{1,2,\dots,nboat\}$ [Eq. 1] where $f_{i,d}$ is the number of fishing events conducted day *d* by vessel *i*, *nboat* is the number of vessels in the fleet and λ_f is the mean number of fishing events per boat and day.

If the mean number of fishing events per boat and day is assumed to differ across boats (stochastic=TRUE), the number of fishing events per boat and day are assigned randomly from:

$f_{i,d} \sim Pois(\lambda_{fi})$ [Eq.2]

where

$\lambda_{fi} \sim ZTPois(\lambda_{\bar{f}})$ [Eq.3]

where $f_{i,d}$ is the number of fishing events conducted day *d* by vessel *i*, λ_{fi} is the mean number of fishing events per day conducted by boat *i*, $\lambda_{\bar{f}}$ is a rate parameter describing how the mean number of fishing events per boat and day is distributed among boats. $ZTPois$ refers to a zero-truncated poisson distribution.

For each fishing event, a bycatch event can happen with a given probability. Once a bycatch event occurs, the number of bycaught animals is randomly drawn from a mixture of random distributions of 'regular' bycatch events and rare 'large' events. Therefore, bycatch is introduced as a combination of bycatch probability and bycatch intensity (number of individuals caught during an incident). Such complex distributions appear to be noticed in many real-world conditions (bycatch incidents of many PET species often involve only a few individuals and in rare cases many).

For each fishing event, *h*, it is first assessed whether a bycatch occurs:

$bycatch_{i,d,h} \sim Bern(p_{bycatch})$ [Eq. 4]

where $p_{bycatch}$ is the probability that bycatch occurs.

If a bycatch occurs, it is thereafter assessed whether the bycatch is a 'large bycatch event':

$largebycatch_{i,d,h} \sim Bern(p_{bycatch=large|bycatch_{i,d,h}=1})$ [Eq. 5]

Where $p_{bycatch=large|bycatch_{i,d,h}=1}$ is the probability of a large bycatch event given that a bycatch occurred.

The number of individuals bycaught for each fishing event where a bycatch occurs is drawn from:

$n_{i,d,h|(bycatch_{i,d,h}=1\ \&\ largebycatch_{i,d,h}=0)} \sim Pois(\lambda_{bycatch})$

$n_{i,d,h|(bycatch_{i,d,h}=1\ \&\ largebycatch_{i,d,h}=1)} \sim Pois(\lambda_{largebycatch})$ [Eq. 6]

Where $\lambda_{bycatch}$ is the mean number of bycaught animals in the event of regular bycatch instances, and $\lambda_{largebycatch}$ is the mean number of bycaught animals in the event of large bycatch instances.

Simulations of fishery fleets and their associated bycatches can be extended with more complexities. One such complexity is the métier used by the vessels. In the SCOTI simulation tool, the distribution of métiers within the fleet is determined by a simplex vector, $\mathbf{p}_{\text{métier}}$ ($\sum_{i=1}^{n} p_{metier,i} = 1$), which is used to randomly assign a métier to each vessel in the fleet. Once a métier is assigned to a vessel a métier specific bycatch probability, $p_{bycatch,metier\ i}$, is assigned to the vessel. Hence, if multiple métiers are apparent in the fleet, bycatch probability is provided as a vector with unique bycatch probabilities defined for each métier.

Input parameters for the fishing and bycatch simulation algorithm (make_fishing_year_métier)

Table A1. Input parameters for the make_fishing_year_métier function.

Parameter	Parameter name in function	Data type	Description
$\lambda_{bycatch}$	mean.bycatch.event	Real (0,∞)	Expected number of animal bycaught given that a “normal” bycatch occurs. This parameter is the rate parameter of a zero truncated poisson distribution and the number of bycaught animals, given that a bycatch occurs, is distributed following eq.6.
$\lambda_{largebycatch}$	mean.bycatch.large.event	Real (0,∞)	Expected number of animal bycaught given that a “large” bycatch event occurs. This parameter is the rate parameter of a zero truncated poisson distribution and the number of bycaught animals, given that a large bycatch occurs, is distributed following eq. 6. $\lambda_{largebycatch} > \lambda_{bycatch}$
$p_{bycatch=large\|bycatch_{i,d,h}=1}$	p.large.event	Real [0,1]	The probability that a large bycatch event occurs, given that a bycatch occurs.
n_{boat}	Nboat	Integer (n_{boat} ≥1)	Number of boats in the fleet
λ_f	mean.fishing.event.boat.day	Real (0,∞)	Mean number of fishing operations conducted per vessel and day. (when stochastic=FALSE)
$\lambda_{\bar{f}}$	mean.fishing.event.boat.day	Real (0,∞)	Rate parameter of a zero truncated poisson distribution used for assigning mean fishing events per day to vessels (when stochastic=TRUE)
$p_{bycatch}$	p.bycatch	Real or vector of reals (0,1]	Probability of bycatch. If a vector with parameter values is provided, values correspond to bycatch probabilities for different métiers. The first value in this vector corresponds to first value in the $\mathbf{p}_{métier}$ vector.
$\mathbf{p}_{métier}$	p.métier	Simplex vector	Distribution of métiers in the fishery. This vector distributes métiers to vessels such that $\sum_{i=1}^{n} p_{metier,i} = 1$
	Stochastic	Logical	Should the mean number of fishing operations per boat and day be drawn randomly? If yes (TRUE), the number of fishing operations per vessel and day is assigned based on eq.2 and eq.3; if no (FALSE), the number of fishing operations per vessel and day is assigned based on eq. 1.

Output from the fishing and bycatch simulation algorithm (make_fishing_year_métier).

Table A2. Output from the make_fishing_year_métier function. Output is provided in a data frame (fishing) with the following columns. Rows in this data frame correspond to unique combinations of fishing day and vessel IDs.

Parameter	Data type	Description
fishing.day	Integer {1,2,…,365}	Day of the year for the fishery simulation
Boat	Integer {0,1,…,nboat}	ID of vessel
Métiers	Integer {1,…,nmétier}	Métier category

Bycatch	Integer {0, 1}	Indictor of whether bycatch occurred or not.
nbycatch	Integer {0,1,…, ∞}	Number of bycaught individuals

Monitoring algorithm

Vessels can be monitored in a number of different ways using the SCOTI simulation tool. The general approach is to first define a monitoring scheme (for example the proportion of vessels to sample and the probability of sampling a fishing events from vessels) that will be implemented on the simulated bycatch data (Table A2). This monitoring scheme is thereafter implemented a predefined number of times by randomizing the vessel selections, fishing event selections etc. according to the predefined monitoring scheme. For each randomization, a specific number of bycatches will be observed along with the actual monitoring effort that was applied. Hence, the randomization process makes it possible to investigate uncertainty in bycatch monitoring, and hence how precision and bias may be affected by the specific sampling scheme being implemented. For one randomization BPUE is calculated following:

$$BPUE_{estimated} = \frac{\sum_{k \in monitored\ vessels} nbycatch_k}{total\ number\ of\ monitored\ vessels} \quad \text{Eq. 7}$$

From a predefined number (n_{sample}) of BPUE estimates the monitoring algorithm outputs the mean and the coefficient of variation of these estimates (Table A4) as:

$$\overline{BPUE}_{estimated} = mean(BPUE_{estimated}) \quad \text{Eq. 8}$$

$$CV(BPUE_{estimated}) = \frac{\sigma_{BPUE_{estimated}}}{\overline{BPUE}_{estimated}}$$

Below follows a description of the specificities of the SCOTI monitoring algorithm. Specifically, we describe four general sampling schemes that can be implemented using SCOTI.

Completely random monitoring

In the SCOTI simulation tool, sampling effort can be distributed across the fleet based on a number of predefined input parameters. The simplest sampling scheme, which does not consider any sampling complexities such as stratification or vessel dependencies, is to sample fishing events from the fleet completely at random. Such a random sampling scheme can be set by defining two parameters in the monitoring function (boat_samp=FALSE, bymétier=FALSE). This setting infers that sampling should not be distributed among boats (boat_samp=FALSE) and that sampling should not be stratified by métiers (bymétier=FALSE). In this case the total coverage of the fleet (proportion fishing events sampled out of the total number of fishing events) is controlled by the parameter $p_{monitor}$, which, in this case, is the probability of monitoring a fishing event:

$$coverage = p_{monitor} \quad \text{Eq.9}$$

Monitoring stratification by métier and random sampling per fishing event

In this case (bymétier=TRUE, boat_samp=FALSE), the probability that a fishing event for a boat with a given métier, is sampled is:

$$p_{monitor|metier} = p_{metier} * p_{monitor|vessel} \quad \text{Eq. 11‡}$$

where $\sum_{i=1}^{n} p_{metier_i} = 1$ for n métiers in the fleet

it is possible to explore non-traditional stratification designs where sampling is unrelated to the proportion of fishing operations in each métier. We invite readers to inspect the code monitor_BPUE_métier.r at github.com/dlusseau/scoti for the algorithm implemented to do so.

Monitoring stratification by métier and random sampling of vessels

In this case (bymétier=TRUE, boat_samp=TRUE), the probability that a fishing event for a vessel with a given métier, is sampled is given by:

$$p_{monitor|metier} = p_{metier} * p_{vessel} * p_{monitor|vessel} \quad \text{Eq. 12‡}$$

Note that at this stage the simulations assume that one vessel is ascribed fishing operations from only one métier. This can be extended to represent cases where vessels can deploy multiple métiers during their operations.

Observation processes

Apart from specifying overall monitoring parameters, such as the proportion of vessels monitored and the proportion of fishing events monitored, it is also possible to implement observer processes in the monitoring function. To this end, SCOTI includes three parameters describing potential error sources in monitoring. These are all implemented after boats and/or fishing events are selected for monitoring. First, after fishing events are selected for monitoring it is assumed that only a fraction ($p_{fishing\_event\_obs}$) of those events are actually observed. Second, it is assumed that the probability that a bycaught individual is detected is given by $p_{detection}$. Third, the vessels may refuse to engage in monitoring with a probability $p_{refusal}$.

Input parameters for the monitoring algorithm (monitor_BPUE_métier)

Table A3. Input parameters for the monitor_BPUE_métier function.

Parameter	Parameter name in function	Data type	Description
$p_{monitor}$	pmonitor	Real [0,1]	The probability of sampling a fishing event (boat_samp=FALSE).
$p_{monitor\|vessel}$	pmonitor	Real [0,1]	The probability of sampling a fishing event given boat selection (boat_samp=TRUE).
	boat_samp	Logical	Input parameter saying whether a predifed proportion of vessels (set by p should be sampled.

‡ Equation added based on reviewer comments.

p_{vessel}	p.monitor.boat	Real [0,1]	Proportion of vessels to sample in the fleet.
	fishing	Data frame	Data frame with simulated fishery and bycatch data (from make_fishing_year_métier function)
	nsample	Integer	Number of times to randomize the monitoring scheme.
$p_{fishing\_event\_obs}$	p_haul_obs	Real [0,1]	the probability that a fishing event (hauls) was observed by an observer
$p_{detection}$	detect_prob	Real [0,1]	The detection probability of each individual in the bycatch event
$p_{refusal}$	refusal_rate	Real [0,1]	The probability that a vessel selected for monitoring refuses to engage

Output from the monitoring algorithm (monitor_BPUE_métier)

The output from monitor_BPUE_métier is provided in a list. If there are more than one métier used in the monitoring simulation BPUE, CV and effort_mon are calculated and provided at the métier level.

Table A4. Output from the monitor_BPUE_métier function.

Parameter	Data type	Description
BPUE_est	Real (BPUE_est ≥ 0) or a vector of reals.	Mean bycatch per unit effort across all randomizations of the fishing and bycatch simulation algorithm. If multiple métiers are simulated in the fishery BPUE estimates for each métier is provided.
CV	Real (CV ≥ 0) or a vector of reals	Coefficient of variation of BPUE estimate. If multiple métiers are simulated in the fishery CV estimates for each métier is provided.
effort_mon	Real (effort_mon ≥ 0) or a vector of reals	Mean effort for BPUE estimates. If multiple métiers are simulated in the fishery effort values for each métier is provided.

Estimate bias and precision across monitoring strategies

The SCOTI simulation framework contains a script ("estimate_bias_and_precision.R") that can be used to test a large number of fishery (with associated bycatch properties)-monitoring combinations. In this script, a user first defines, in a list with sub-lists (fishery_params), unique parameter sets describing fisheries and their presumed bycatches properties. Second, a user defines, in another list with sub-lists (monitor_params), unique parameter sets describing the different ways in which the fisheries are to be monitored. Each monitoring strategy (defined in each parameter sub-list of monitor_params) is implement on each of the simulated fisheries (defined in each parameter sub-list of fishery_params). The script ("estimate_bias_and_precision.R") outputs a data.frame with columns describing the monitoring-fishery combinations as well as the true BPUE from the fishery simulation, and the estimated BPUE and CV of the BPUE estimate retrieved from the monitoring function.

Input to estimate_bias_and_precision R-script

Table A5. Input to the "estimate_bias_and_precision.R" R-script.

Input	Description
fishery_params	A list with sublists containing the parameters defined in Table A1. Each sublist, i.e. parameter set-up, represents one fishery/bycatch scenario.
monitor_params	A list with sublists containing the parameters defined in Table A3. Each sublist, i.e. parameter set-up, represents one monitoring scheme to be implemented.

-

Output from estimate_bias_and_precision R-script

The estimate_bias_and_precision R-script outputs the true value of BPUE from the fisheries simulation, the estimated BPUE from monitoring, the coefficient of variation of the BPUE estimate and effort from monitoring in a data frame with additional columns describing the specific monitoring and fisheries scenarios. If monitoring is stratified the script outputs additional columns, such as bycatch probabilities and BPUE estimates, for the different métiers.

Table A6. Output from the "estimate_bias_and_precision.R" R-script.

Parameter	Parameter name in function	Data type	Description
year	year	Integer {1,2,…,100}	Sampling year
$p_{bycatch}$	p_bycatch	Real or multiple reals [0,1]	The probability of bycatch per métier. If there are multiple métiers in the simulated fisher, bycatch probabilities for each métier are provided in separate columns (with name p_bycatch_i for métier *i*).
p_{metier}	p_métier		Parameters saying how large proportion of the fishery are associated with specific métier types. Proportions for each métier are provided in unique columns (with name p_métier_i for métier *i*).
$p_{monitor}$	pmonitor	Real [0,1]	Depending on the settings of boat_samp and bymétier it is defined as in eq. 9 and eq.10 (see section 2.2 Monitoring algorithm)
	p_monitor_boat	Real [0,1]	Probability of selecting vessels for monitoring.
	boat_samp	Logical	Indicator saying whether vessels were randomly assigned for monitoring (TRUE) or not (FALSE).
	bymétier	Logical	Indicator saying whether monitoring was stratified by métier (TRUE) or not(FALSE).
	BPUE_real	Real [0 ,∞)	True bycatch per unit effort for the simulated fishery. If multiple métiers are apparent in the fishery, BPUEs per métier are provided in separate columns (with name BPUE_real_i for métier *i*).
	BPUE_est	Real [0 ,∞)	Estimated bycatch per unit effort for the specific monitoring scheme. If multiple métiers are apparent in the fishery, BPUEs per métier are

			provided in separate columns (with name BPUE_est_i for métier *i*).
	BPUE_est_CV	Real [0 ,∞)	Coefficient of variance for the estimated BPUE estimate. If multiple métiers are apparent in the fishery, CVs per métier are provided in separate columns (with name CV_i for métier *i*).
	effort_real	Real [0 ,∞)	Total effort for simulated fishery. If multiple métiers are apparent in the fishery, effort is provided at the métier level in separate columns (with name effort_real_i for métier *i*).

SCOTI considerations and warnings

It is important to note that the SCOTI simulation tool is not a final product. The code is currently still ongoing beta testing and future extensions will be implemented. Hence, results derived using the SCOTI simulation tool should be interpreted with care. Here we document some important fact to consider when using the SCOTI simulation platform:

- It is important to consider the total monitoring effort being implemented for a single monitoring scenario. For example, imaging that an investigator wants to explore whether a limited amount of resources (total monitoring effort) should be distributed among a specific number of vessels or be completely randomly assigned across fishing events. Then it is important to make sure that the total effort being considered is the same for the two cases (see paragraph Monitoring algorithm; equations. 9 & 10).
- If monitoring is stratified based on métiers it is important to carefully consider how the distribution of métiers in the fishery relates to the proportion of métiers being monitored. To this end, it is also important to keep track of the total monitoring effort. For example, it is important to make sure that a result is not dependent on total monitoring effort rather than sampling stratification.

Future extensions

- Work is currently in progress on making the code spatially implicit, that is, that certain vessels (or areas) could be considered bycatch hotspots.
- Heterogeneity in vessel bycatch rate is currently being implemented.

Annex 3: Métier overlap" (Influence of monitoring stratification on BPUE precision and CV)

Presentation of the case study

The current case study was developed on the basis of data simulated from the Case study C, Gillnets in the North Atlantic, case study (Case Study C)., Case study C5, C6 and C7. Detailed information about the fleet description, characterization of bycatch parameters, and monitoring can be found in section "2.6.3 Description of case studies, case study C".

Methods

The methodology used is the same as that described in the general methodology section (cross-reference to section "Methods"), with slight modifications focused on testing the variations that métier stratification may produce in the precision or bias of the BPUE for three species groups. For this case study, each of the métiers was assigned a different probability of being monitored and each of the species groups presented a different bycatch probability (0.009 – very low probability, 0.021 - medium probability and 0.108 - high probability). The rest of the parameters are the same as described for case study"Gillnets in the North Atlantic" above. Once all the parameters are selected, simulations are run and the real BPUE, the estimated BPUE, the absolute bias in the BPUE and the CV of the BPUE for the whole fleet are calculated. Generalised linear models are fitted for the bias and CV of the BPUE by species group, using the different possible interactions between the proportion of fishing operations monitored and the proportion of vessels monitored as explanatory variables. After selecting the best model, based on their performance, predictions were made for different proportions of monitoring of fishing operations or of vessels, depending on the variables selected by the models, varying from 0 to 30% in variable steps.

Simulation results

Bias of BPUE

For the three groups of species simulated, absolute biases in the BPUE are always reduced when the proportion of fishing operations monitored is higher. Monitoring stratified by métier appears to reduce biases in BPUE for groups of species with low bycatch probabilities (0.009, Figure A.2.3B), but not significantly for species with higher bycatch probabilities (0.021 - 0.108, Figure A.2.2A and A.2.4C). A low proportion of vessels in the fleet should be monitored when monitoring is stratified by métier and vessel, as increasing this proportion leads to an increase in BPUE bias (Figure A.2.1D).

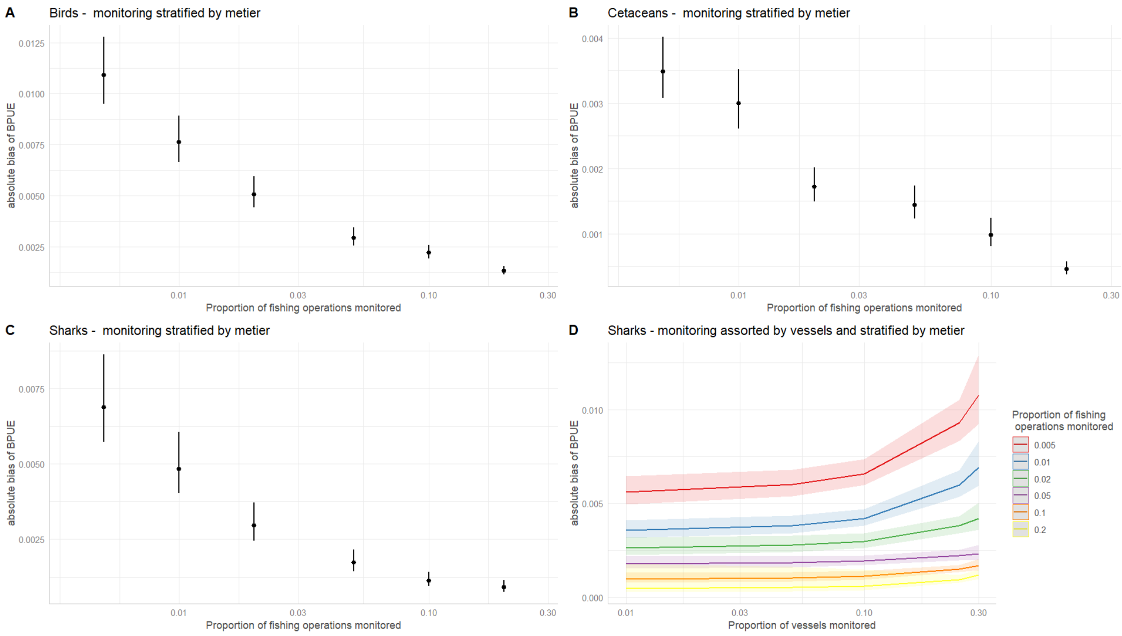


Figure A.2.1. Predicted absolute bias of BPUE of different species groups exemplifying a summary of the results of the simulations performed in the stratification of monitoring by métier and/or vessel. Shaded bands (transparent colours in graph D) represent 95% confidence intervals. Note that the y-axes have different scales and that the x-axis of graph D represents the proportion of vessels monitored and not the proportion of the fishing operations monitored, as the rest of the graphs in the panel.

The scenario outcomes detailed by species group are presented below.

Seabird species group

Absolute bias in BPUE is barely altered, slightly increased, when monitoring is stratified by métier or by vessels. Bias decreases as the proportion of fishing operations monitored increases, when monitoring is unstratified or stratified by métier. When monitoring is stratified by vessels, there is also a decrease in absolute bias as the number of fishing operations monitored and the proportion of vessels monitored increases. When monitoring is stratified by vessel and métier, the bias increases as the proportion of vessels monitored increases.

Summary: Regarding bias of BPUE, for a group of seabird species with a mean probability of bycatch of 0.021, stratification is not relevant apparently. There is a decrease in bias as we increase the number of fishing operations monitored, but keeping a low proportion of the vessels monitored (Figure A.2.2).

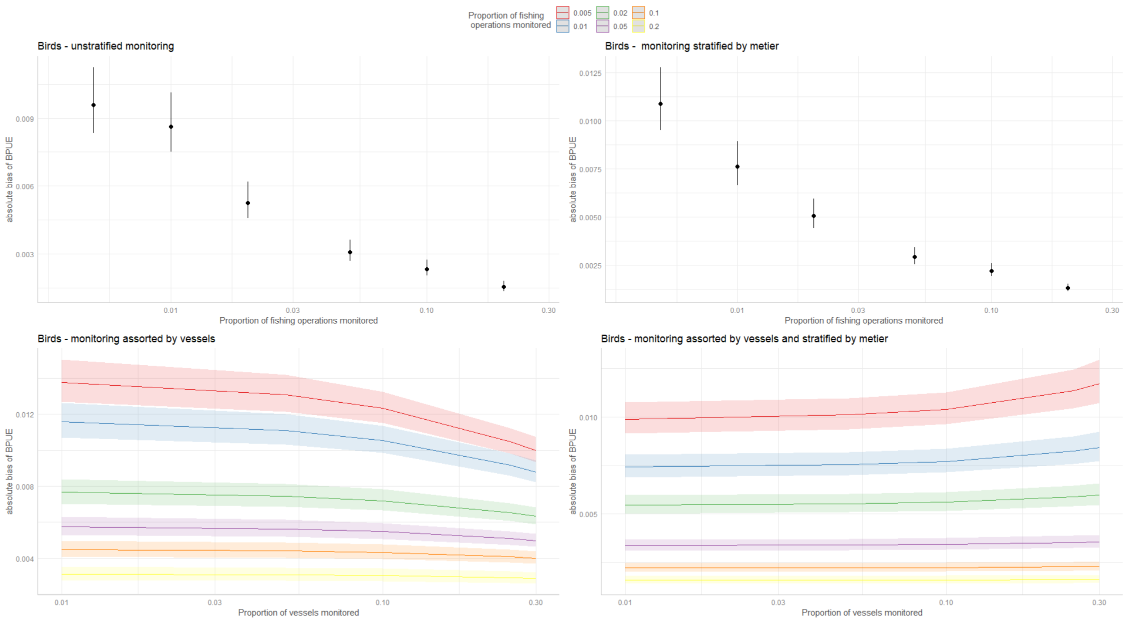


Figure A.2.2. Predicted absolute bias of BPUE of a simulated seabird species group obtained from generalised linear models in which the response variable (absolute bias of BPUE) is associated to different interactions between the proportion of fishing operations monitored and the proportion of vessels monitored, assuming a Gamma distribution of the residuals. Shaded bands (transparent colours in the lower panels) represent 95% confidence intervals. The graphs presented correspond to the prediction obtained from the best fitting models.

Cetacean species group

Biases are lower when monitoring effort stratification is performed by métier, leading to a decrease in absolute bias especially as the number of fishing operations monitored increases. When monitoring is stratified by vessel and métier, the decrease in absolute bias is smaller but bias also decreases as the number of fishing operations monitored increases. If monitoring is only stratified by vessels, the biases are larger but decrease slightly as the number of vessels monitored increases.

Summary: To reduce bias in BPUE of a group of cetacean species with a mean bycatch probability of 0.009, it seems to be more appropriate to stratify by métier, monitoring as many fishing operations as possible (Figure A.2.3).

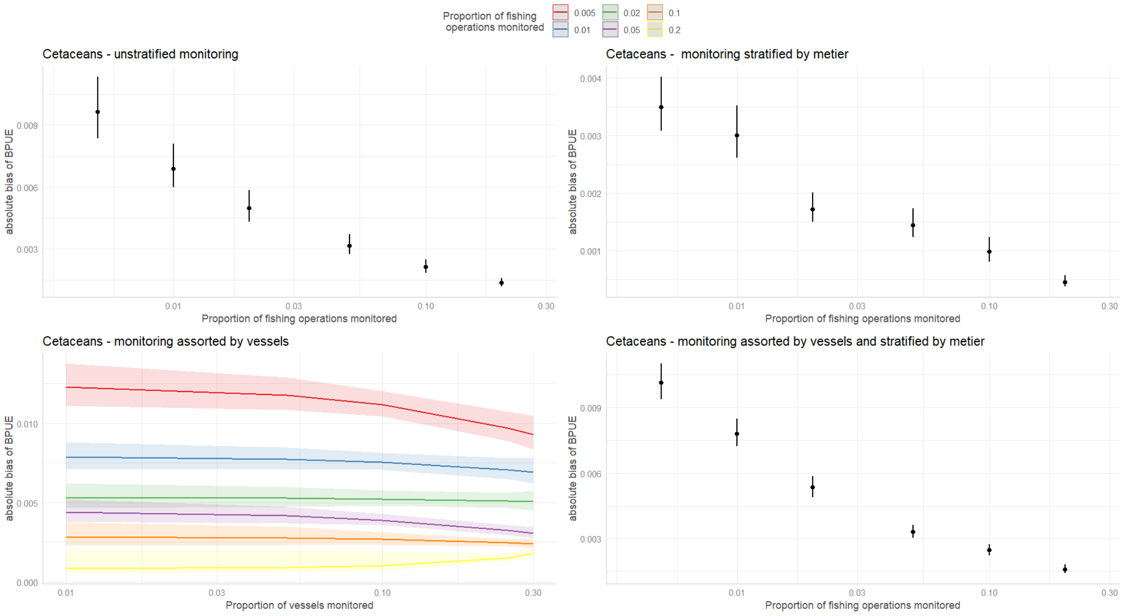


Figure A.2.3. Predicted absolute bias of BPUE of a simulated cetacean species group obtained from generalised linear models in which the response variable (absolute bias of BPUE) is associated to different interactions between the proportion of fishing operations monitored and the proportion of vessels monitored, assuming a Gamma distribution of the residuals. Shaded bands (transparent colours in lower left panel) represent 95% confidence intervals. The graphs presented correspond to the predictions obtained from the best fitting models.

Protected shark species group

Absolute bias in BPUE of a protected shark species group is hardly improved if monitoring is performed stratified by métier, although a decrease in bias is observed when we increase the proportion of fishing operations monitored in both cases (unstratifiedvs.métier-stratified).

When monitoring is stratified only by vessels, a considerable increase in absolute bias is observed, but as the proportion of vessels monitored is increased, the biases are reduced to levels similar to those obtained without stratification.

The opposite behaviour in bias is observed if the stratification of monitoring is performed by vessels and métier. In that case, the biases are on the scale of the biases when no stratification is performed but, in this scenario, as we increase the proportion of vessels monitored, the biases of the BPUE increase.

Summary: Stratification of monitoring by métier, vessel or by vessel and métier does not lead to a reduction in BPUE absolute biases for a protected shark species group, with a mean bycatch probability of 0.108. There is a reduction in bias as the number of fishing operations monitored increases, and if stratified by vessels and métier, monitoring should be concentrated on a small proportion of the vessels in the fleet (Figure A.2.4).

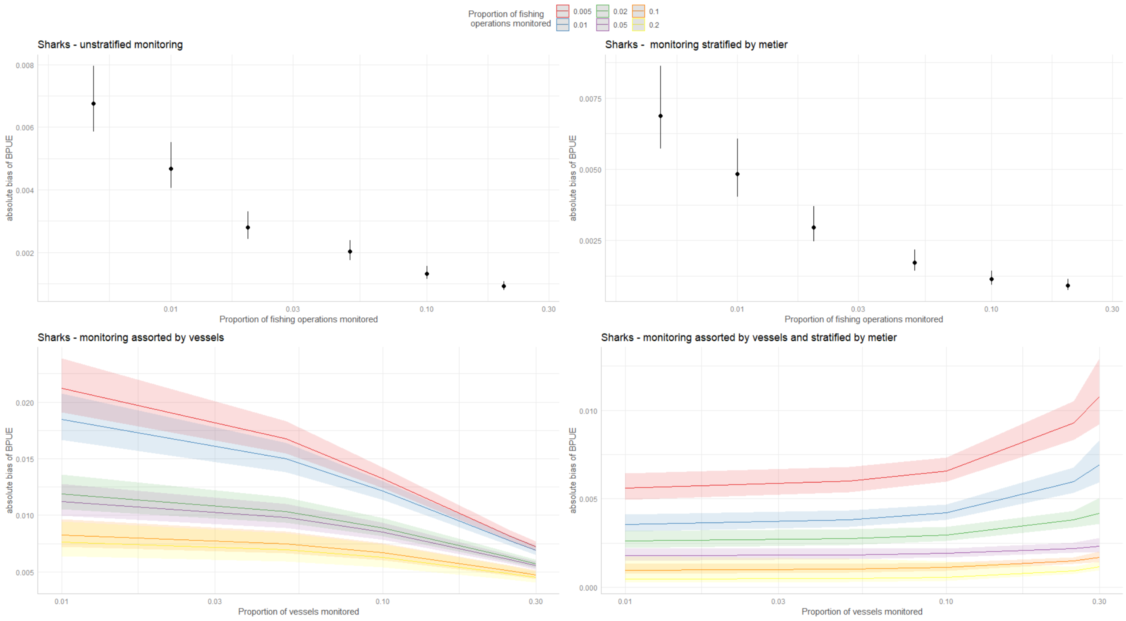


Figure A.2.4. Predicted absolute bias of BPUE of a simulated protected shark species group obtained from generalised linear models in which the response variable (absolute bias of BPUE) is associated to different interactions between the proportion of fishing operations monitored and the proportion of vessels monitored, assuming a Gamma distribution of the residuals. Shaded bands (transparent colours in lower panels) represent 95% confidence intervals. The graphs presented correspond to the predictions obtained from the best fitting models.

CV of BPUE

CVBPUE of species groups with very low bycatch probability (0.009) could be reduced through stratified monitoring by métier, while increasing the number of fishing operations monitored as much as possible (Figure A.2.5A). If there are species with a higher bycatch probability (0.02 - 0.10), it is better to conduct unstratified monitoring as well as increasing the number of fishing operations monitored in order to reduce CV (Figure A.2.5B and A.2.5C).

Across all the species groups analysed, stratification of monitoring by métier and vessel results in an increase in the baseline CVBPUE. In addition, with this stratification scheme, an increase in the proportion of vessels monitored will also result in an increase in the BPUE CV (Figure A.2.5D).

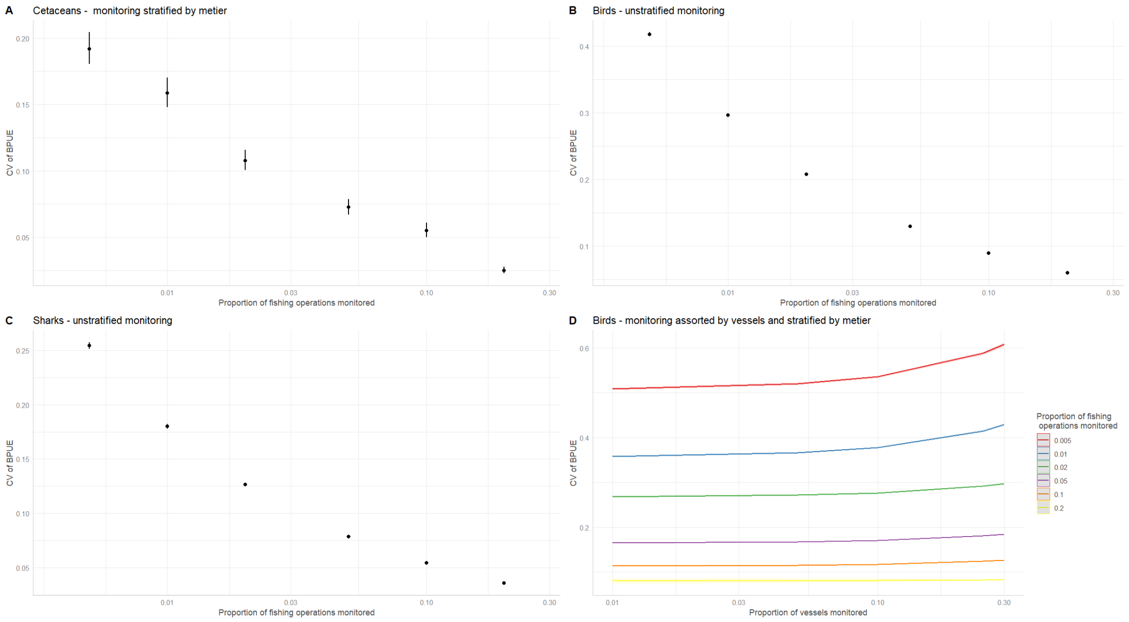


Figure A.2.5. Predicted CVBPUE of different species groups exemplifying a summary of the results of the simulations performed in the stratification of monitoring by métier and/or vessel. Shaded bands (transparent colours in graph D) represent 95% confidence intervals. Note that the y-axes have different scales and that the x-axis of graph D represents the proportion of vessels monitored and not the proportion of the fishing operations monitored, as the rest of the graphs in the panel.

<u>The scenario outcomes detailed by species group are presented below.</u>

Seabird species group

In seabirds, the CVBPUE increases (from 0.4 to 0.6) when monitoring is stratified by métier. If monitoring is unstratified or stratified by métier, any increase in the proportion of fishing operations monitored will reduce the CV. If monitoring is stratified by vessels, the CV will be reduced as the proportion of vessels monitored increases. On the contrary, if monitoring is stratified by vessels and métier the CV increases as the proportion of vessels monitored increases.

Summary: For a group of seabird species with a mean probability of bycatch of 0.021, the best option to reduce CVBPUE is to perform a non-stratified monitoring, increasing as much as possible the proportion of fishing operations monitored (Figure A.2.6).

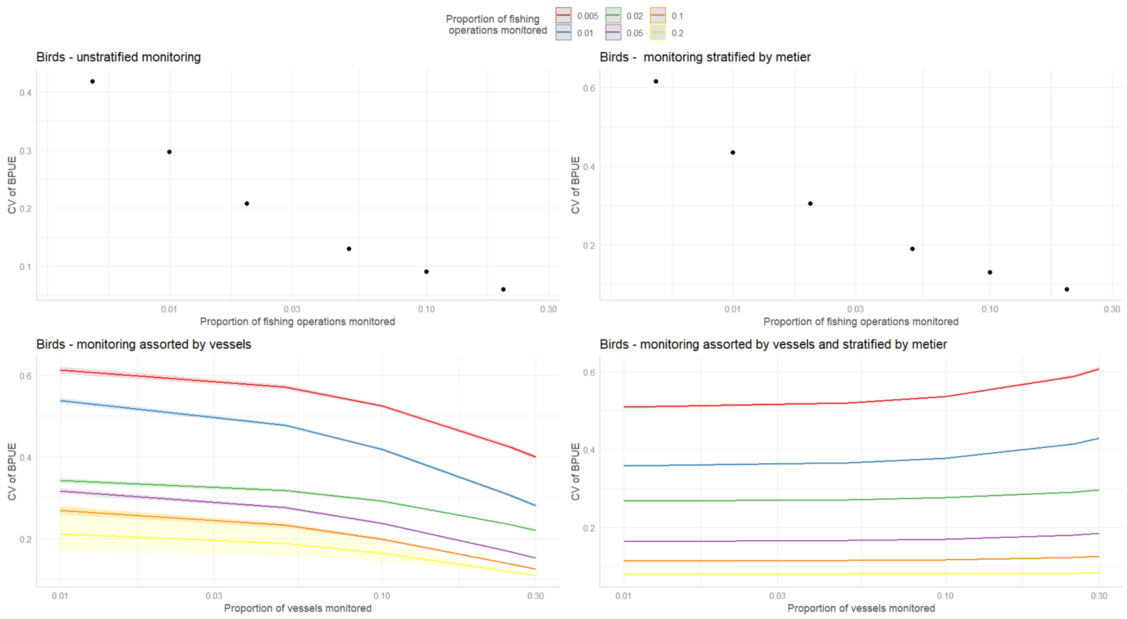


Figure A.2.6. Predicted CVBPUE of a simulated seabird species group obtained from generalised linear models in which the response variable (CVBPUE) is associated to different interactions between the proportion of fishing operations monitored and the proportion of vessels monitored, assuming a Gamma distribution of the residuals. Shaded bands (transparent colours in lower panels) represent 95% confidence intervals. The graphs presented correspond to the predictions obtained from the best fitting models.

Cetacean species group

In cetaceans, when monitoring is stratified by métier, the CVBPUE is lower than when monitoring is unstratified or it is stratified by vessel. If the monitoring is unstratified or stratified by métier, there is a decrease of the CV when the proportion of fishing operations monitored increases. If the monitoring is stratified by vessels, the CV decreases when the proportion of vessels monitored is increased. But when monitoring is stratified by vessels and métier, an increase in the proportion of vessels monitored will produce an increase in the CV.

Summary: For a group of cetacean species with a mean bycatch probability of 0.009, the best option to reduce the CVBPUE is to perform a monitoring stratified by métier, increasing as much as possible the proportion of fishing operations monitored (Figure A.2.7).

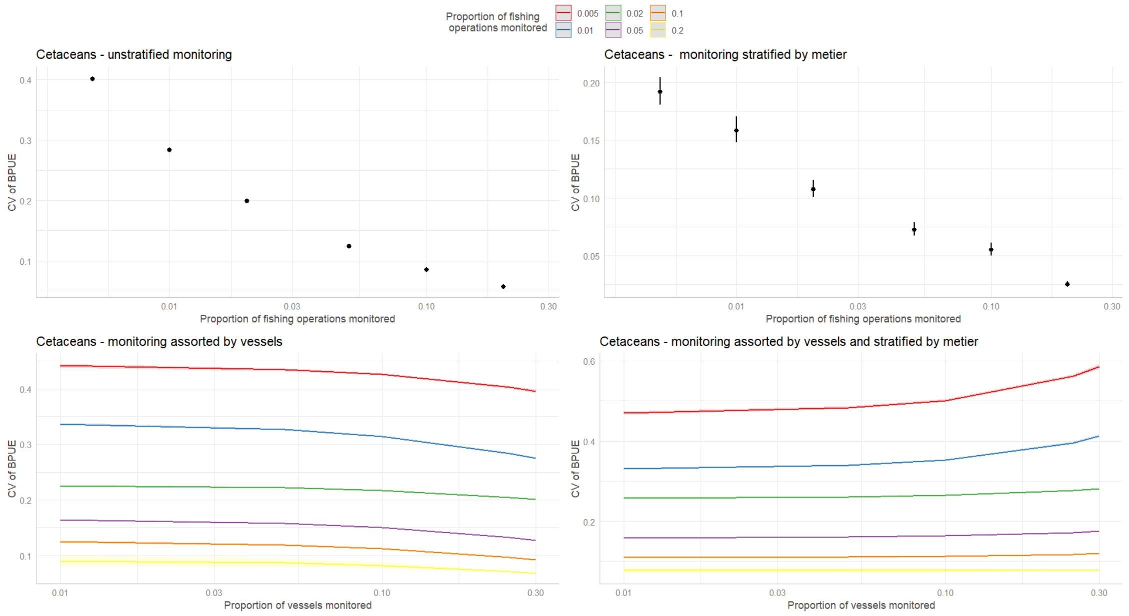


Figure A.2.7. Predicted CVBPUE of a simulated cetacean species group obtained from generalised linear models in which the response variable (CVBPUE) is associated to different interactions between the proportion of fishing operations monitored and the proportion of vessels monitored, assuming a Gamma distribution of the residuals. Shaded bands (transparent colours in lower panels) represent 95% confidence intervals. The graphs presented correspond to the predictions obtained from the best fitting models.

Protected shark species group

For protected shark species monitoring, there is an increase in the CV of the BPUE when we stratify sampling by métier (from 0.25 to 0.4), but in both cases the CV decreases as we increase the proportion of fishing operations monitored (reaching a CV=0.2 when the proportion of fishing operations monitored is around 0.01 and 0.015).

There is a very steep increase of the CV if we stratify monitoring by vessel, but CV is reduced to considerably low levels (similar to the CV obtained without any stratification = 0.2) as we increase the proportion of vessels monitored (especially when the proportion of monitored vessels is higher than 0.15).

When stratified by vessel and métier, the CV of the BPUE increases as the proportion of vessels and fishing operations monitored increases.

Summary: The best option to reduce the CV of the BPUE in a protected shark species group, with a mean bycatch probability of 0.108 is to perform an unstratified monitoring, and increase as much as possible the proportion of fishing operations monitored (Figure A.2.8).

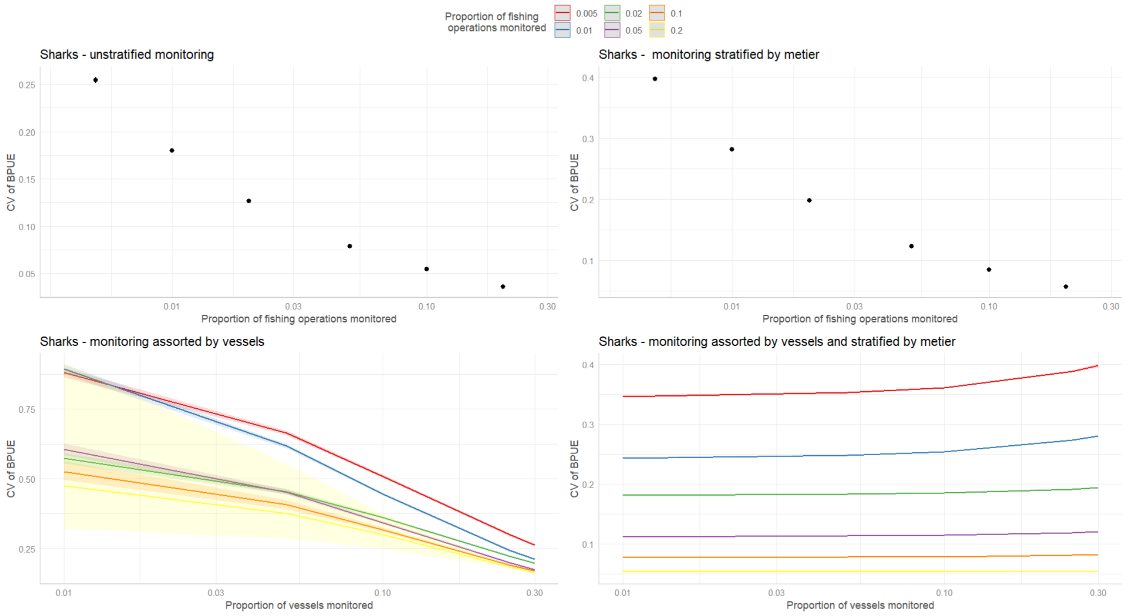


Figure A.2.8. Predicted CVBPUE of a simulated protected shark species group obtained from generalised linear models in which the response variable (CVBPUE) is associated to different interactions between the proportion of fishing operations monitored and the proportion of vessels monitored, assuming a Gamma distribution of the residuals. Shaded bands (transparent colours in lower panels) represent 95% confidence intervals. The graphs presented correspond to the predictions obtained from the best fitting models.

Annex 4: Norwegian Reference Fleet case study – Parametrization for a seabird sample

The current example is based on data from the Norwegian Reference Fleet in Norway. More information on The Norwegian Reference Fleet can be found in Clegg and Williams, 2020. In this particular case we consider bycatch of Northern fulmars (*Fulmarus glacialis*) in a sample of the offshore longline fishery where we have high confidence in that the reported numbers are close to actual numbers of bycatch.

It is worth keeping in mind that the observed probabilities of bycatch events and number of individuals bycaught per event could be biased values (e.g. already including an observation error that might shape those parameters). Parameter values used in the 'base case' and their descriptions are in Table A.4.1.

***Table** A.4.1: Parameter values used in simulations.*

Simulation component	Parameter name	Description	Value in base case	Value in dedicated observer program
fishing	nboat	Size of the active fishing fleet	28	28
fishing	mean.fishing.event.boat.day	Mean number of fishing events per boat day	1	1
fishing	stochastic	Logical; should a vessel-specific mean number of events per vessel be used, or a single value for the entire fleet?	FALSE	FALSE
fishing	p.métier	Probability (proportion) of vessels belonging each métier	1	1
bycatch process	p.bycatch	Probability of bycatch event by haul and métier	0.12	0.12
bycatch process	p.large.event	Given that a bycatch event occurs, the probability of a large bycatch event	0.007	0.007

Simulation component	Parameter name	Description	Value in base case	Value in dedicated observer program
bycatch process	mean.bycatch.event	Given that a bycatch event occurs and that event is not large, the mean number of individuals caught at a 'normal' bycatch event	2	2
bycatch process	mean.bycatch.large.event	Given that a bycatch event occurs and that event is large, the mean number of individuals	17.6	17.6
simulation parameter	nsample	Number of samples to be taken from the fishing data	1000	1000
observation	p_monitor_boat	Proportion of vessels monitored	0.096774194	1
observation	pmonitor	Proportion of hauls monitored for each vessel	1	0.5 (Observer works same as crew shifts: on/off over 24 hours)
observation	p_monitor_métier	Proportion of monitoring allocated to (each?) métier	1	2:31/31
observation	bymétier	Logical; whether sampling is stratified by métier?	FALSE	FALSE
observation	boat_samp	If FALSE: sampling occurs at the fishing event level (haul); if TRUE: first sample vessels to be monitored, then sample hauls	TRUE	TRUE
observation	refusal_rate	Probability that a sampled vessel is rejected due to refusal to allow observer onboard	0	seq(0.1, 0.9, length.out = 4) (This value is sensitive to too many factors

Simulation component	Parameter name	Description	Value in base case	Value in dedicated observer program
				so is varied in the simulations)
properties of an average monitoring event	p_haul_obs	Probability that a haul is observed	1	0.95 (dedicated programme with trained observers has lower chance of missing observations)
properties of an average monitoring event	detect_prob	Probability of detection of each individual in a bycatch event	0.7	1
properties of an average monitoring event	misclassification	Probability of mis-identification of the bycaught species	0	0

{SCOTI} functions used

This code uses all simulation functions from the original SCOTI scripts, except for the `make_fishing_métier()` function, which has been modified to include a "vessel effect" (see below).

```
source("https://raw.githubusercontent.com/dlusseau/scoti/main/estimate_fishi
ng_effort_métier.R")
source("https://raw.githubusercontent.com/dlusseau/scoti/main/heterogeneity_
stats_distributions.r")
source("https://raw.githubusercontent.com/dlusseau/scoti/main/make_fishing_y
ear_métier.r")
source("https://raw.githubusercontent.com/dlusseau/scoti/main/monitor_BPUE_m
étier.R")
```

Simulating fishery data

We use the `make_fishing_year_métier()` function to simulate the fishery in the case study.

The next parameter controls the mean number of fishing events per boat day. In this case study, vessel is the primary sampling unit, and we assume that one day at sea equals one haul (i.e., mean fishing event per boat day is one). The data represent one métier (L4, LLS_DEF_0_0_0), so the probability of recording at one métier is set to one.

The next couple of parameters concerns the probability of a bycatch event: The mean number of fulmars taken given an event, as well as the potential of more extreme events (probability of a large bycatch event as well as the mean number of individuals taken in those events). In the data there are signs of bycatch following a bimodal distribution, in the sense that we have some rare events with a much higher number ($maxN \sim 36$) than the estimated mean bycatch ($\sim$ 2 individuals per haul). The probability of a large event, as well as the number of birds captured in these events are however rather uncertain as they are indeed rare events in the data. So, these parameters could be considered a mixture between data driven and more qualitatively driven (based on informal communication with fishers). In the data, there are unexplained variation between vessels in reported bycatch numbers, i.e. a vessel effect of bycatch which might be attributed to a number of variables in reality (training of observers, mitigation measures in place, spatio-temporal variation in fishing etc).

```
# Simulate the true state of the fishery
nboat <- 28
mean.fishing.event.boat.day <- 1
p.métier <- 1
p.bycatch <- 0.12  # range is from 0.02 to 0.12
p.large.event <- 0.007
mean.bycatch.event <- 2  # (rounded down from 2.4)
mean.bycatch.large.event <- 36  # max in dataset is 36
stochastic <- FALSE
BPUE_real <- sum(fishing$nbycatch)/dim(fishing)[1]
vessel.effect <- 0.7
```

Simulating monitoring

We don't have any actual data to calculate detection probability of individual birds in the catch, but we use a fixed probability 0.7 based on similar fisheries in the literature. The 'refusal rate' is assumed to be zero, as these values represent a reference fleet.

```
# Simulate monitoring
pmonitor <- 1
nsample <- 100  # How many times to run - this is independent of the fishery
, it's just how many draws you want to do.
```

```
p_monitor_boat <- 3/31  #  n boats that were willing to report seabirds
boat_samp <- TRUE
p_haul_obs <- 1  # should be fixed at 1 per David
detect_prob <- 0.7
refusal_rate <- 0  # ignoring refusal for now
misclassification <- 0
bymétier <- FALSE
p_monitor_métier <- 1
```

Incorporating a vessel effect

To account for non-independence of repeated sampling by vessels, we modified the fishery simulation step in the SCOTI functions to include an added vessel effect. This was done by introducing additional variation in bycatches for each vessels after bycatch observations were generated by adding random noise to the final estimate with $\mu = 0$ and σ defined as an additional parameter in the simulation (named `vessel.effect`). To ensure the additional variation didn't generate negative values, we applied it to log-transformed observations before back-transforming to the original scale.

Including a vessel effect in the simulation model shifted the CV of the BPUE estimate from simulations to more closely match the model-based estimate from the GLMM.

Spatial and temporal overlap extension

We extended the original SCOTI framework to represent a situation where there is spatial variation in the probability of a bycatch event occurring (i.e., bycatch hot spot(s)). The bycatch hot spot(s) in this function can be switched on and off during specific periods in a year. We modified the fishery-simulating function in SCOTI to include multiple areas, one of which can be identified as a hotspot (e.g. an area where seabirds might be actively feeding) and a temporal trend (e.g. an area with a breeding colony where there is a temporal effect on the probability of bycatch occurring).

The extended framework (in the `make_fishing_year_métier_space()` function) allows the analyst to control the amount of fishing effort across a simplified two-dimensional landscape. Because fisheries vary in intensity across areas during a year, the spatial extension we developed during the workshop controls fishing effort by area using a beta distribution to control skewness (i.e., how strong the hotspot is).

Bycatch hotspots were parameterized as one or more areas (`hotspot.area`) of a total number of areas (`narea`), which then will act as a bycatch hotspot with increased probability of bycatch during the specified time period during a year (`time.periods.bycatch`). This option can be switched on or off with the parameter `spatio.temporal.bycatch.trend`. The trend in fishing effort is similarly switched on or off with the parameter `spatio.temporal.fishery.trend`. This allows for control over which areas the vessels in the fleet should target their fishing effort. To do this, two sets of shape parameters controlling the shape of two different beta distributions should be defined. First, a general pattern of fishing activity across the total number of areas should be defined with the parameter `spatial.effort.skewness.general` (for example if the general pattern of the vessel is a uniform distribution of vessels across areas we use $\alpha = \beta = 1$). Then one could shift this distribution using other values of α and β in the parameter `spatial.effort.skewness.special`, if there are specific time periods during the year (defined in `time.periods.fishery`) where the fishery is more focused in some areas.

```
# Tuning of bycatch variability
spatio.temporal.bycatch.trend <- TRUE  #switch that turn on (TRUE) or off (F
ALSE) bycatch variability
narea <- 10  # Defines the total number of areas in the focal fishery, integ
r
```

```
hotspot.area = 10  #Defines a one or more areas between 1 and narea to
time.periods.bycatch <- 32:60  #Time periods of the year, in days, hotspot s
hould be active. Can be multiple time periods

# Tuning of fishery

spatio.temporal.fishery.trend <- TRUE  # this turns on or off the other spat
ial/temporal
spatial.effort.skewness.general <- c(1, 1)
spatial.effort.skewness.special <- c(1.7, 0.3)
time.periods.fishery <- 32:60
```

In its current state, the spatial extension produces patterns in the expected way but requires more testing in order to be useful for providing insight into how fisheries around hotspots should be monitored.

Code availability

The scripts used to generate these simulations are saved in a fork of the original SCOTI repository on GitHub (https://github.com/tomlclegg/scoti).

Reference

Clegg, T. and Williams, T. (2020). Monitoring Bycatches in Norwegian Fisheries. Species registered by the Norwegian Reference Fleet 2015-2018 (IMR) RAPPORT FRA HAVFORSKNINGEN NR. 2020-8. https://hdl.handle.net/11250/2685855

Annex 5: List of participants

Name	Dept/Institute	Email
Estanis Mugerza (chair)	AZTI Sukarrieta	emugerza@azti.es
Katja Ringdahl (chair)	SLU Department of Aquatic Resources	katja.ringdahl@slu.se
Sara Königson (chair)	SLU Department of Aquatic Resources Institute of Marine Research	sara.konigson@slu.se
Ruth Fernandez	International Council for the Exploration of the Sea	ruth.fernandez@ices.dk
Ailbhe Kavanagh	Marine Institute	Ailbhe.Kavanagh@Marine.ie
Allen Kingston	University of St Andrews	ark10@st-andrews.ac.uk
Anna Cheilari	European Commission Directorate-General for Environment A4	Anna.CHEILARI@ec.europa.eu
Bram Couperus	Wageningen Marine Research	bram.couperus@wur.nl
Caterina Maria Fortuna	Institute for Environmental Protection and Research	caterina.fortuna@isprambiente.it
David Lusseau	DTU Aqua, National Institute of Aquatic Resources	davlu@aqua.dtu.dk
Didzis Ustups	Institute of Food Safety, Animal Health and Environment	didzis.ustups@bior.lv
Gildas Glemarec	DTU Aqua, National Institute of Aquatic Resources	ggle@aqua.dtu.dk

Gudjon Sigurdsson	Marine and Freshwater Research Institute	gudjon.mar.sigurdsson@hafogvatn.is
Henn Ojaveer	International Council for the Exploration of the Sea	henn.ojaveer@ices.dk
Henrik Pärn	SLU Department of Aquatic Resources Institute of Marine Research	henrik.parn@slu.se
Jani Helminen	Natural Resources Institute Finland	jani.helminen@luke.fi
Kay Panten	Thünen-Institute of Sea Fisheries	kay.panten@thuenen.de

Annex 6: Terms of Reference

2022/WK/DSTSG07 The **Workshop on appropriate sampling schemes for Protected Endangered and Threatened Species bycatch (WKPETSAMP3)** chaired by Katja Ringdhal* (Sweden), Sara Königson* (Sweden), and Estanis Mugerza* (Spain), will meet in Copenhagen, Denmark on 13-17 November 2023 to:

a) Identify criteria and best practices for designing a multipurpose programme for sampling and estimating bycatch of PETS in order to assess population level impacts (Science Plan Codes: 3.2 and 3.3);
b) Make recommendations for improving monitoring systems for PETS bycatch at a Member State level and for regional level coordination. Amongst others, it should include proposals for adjusting DCF sampling to cover all PETS bycatch relevant fisheries. (Science Plan Codes: 6.4).

WKPETSAMP3 will report by 15 December 2023 to the attention of the HAPISG, ACOM and SCICOM.

Supporting information

Priority	The workshop is directly linked to a special request for advice from DGEnvironment on 'appropriate bycatch monitoring systems at Member State level and on regional coordination.'
Scientific justification	WKPETSAMP2 and WKPETSAMP3 will contribute to enhance data availability and improve data quality for bycatch estimates of protected species. Both workshops will support objective 4.2. of The Roadmap for ICES bycatch advice on protected, endangered and threatened species; propose options to improve the data availability and quality. The workshops will address two of the types of information needed to assess the conservation threat posed by fishery bycatch to a particular species: (i) the susceptibility of that population to bycatch in particular fisheries (based on monitoring effort); (ii) the scale of the fisheries concerned (based on total fishing effort by fishing gear for all relevant fleet segments and with effort given in meaningful metrics). Relevant outcomes from the Workshop on Estimation of Rare Events (WKRARE, 2021) will be considered. In addition, conclusions from the recent review of monitoring of bycatch of protected, endangered, and threatened species of mammals, birds, turtles and fish[4] will be taken into account The criteria mentioned in WKPETSAMP2 and WKPETSAMP3 ToR a may include: • Adequate temporal resolution (e.g. quarter, month, year) for the different taxa (mammals, birds, turtles);

4 ICES. 2022. EU request on the review of monitoring bycatch of protected, endangered, and threatened species of mammals, birds, turtles and fish under the service of EC DG ENVIRONMENT. In Report of the ICES Advisory Committee, 2022. ICES Advice 2022, sr.2022.04, https://doi.org/10.17895/ices.advice.10096

	• Adequate "primary sampling units" (e.g. haul level, trip level, other aggregation levels) for the different taxa (mammals, birds, turtles); • Use of standardized effort calculation methodologies and relevant total effort units (e.g. Fishing days vs. soak time) for different métiers; • Impact of the use of different effort units (e.g. Fishing days, hauls, km/hr) in bycatch rate calculations for a given métier; • Data quality of total effort data from different sources; • Identification of key geographic areas to be monitored; • Identification of key métiers to be monitored; • Identification of adequate monitoring methodologies (e.g. REM, dedicated observers) for the different métiers; • Adequate temporal frequency of the sampling. The case studies mentioned in WKPETSAMP1 ToR b will include data recorded through remote electronic monitoring, dedicated observer programs, crew observers from reference fleets.
Resource requirements	None beyond the funding for the workshops to be provided by DGEnvironment
Participants	The workshops will be attended by approximately 15 experts.
Secretariat facilities	SharePoint access and Secretariat support including assistance from the ICES Data Centre.
Financial	Financed through specific budget linked to a special request for ICES advice.
Linkages to advisory committees	ACOM
Linkages to other committees or groups	DSTSG, HAPISG, WGCATCH, WGBYC
Linkages to other organizations	OSPAR, HELCOM

Annex 7: Review of ICES WKPETSAMP2 and WKPETSAMP3 reports

Introduction

The European Commission, DG Environment has requested a special advice from ICES. To reply to the request, ICES organized two workshops on appropriate sampling schemes for Protected, Endangered, and Threatened Species bycatch (WKPETSAMP2 and WKPETSAMP3). The objectives of the workshops were to provide the scientific basis for advice on '*appropriate bycatch monitoring systems at Member State level and on regional coordination*'.

From reading Annex 2 to the WKPETSAMP2 report, it seems that the workshop was charged with:

1. Reviewing scientific literature where criteria and data quality thresholds have been applied to PETS bycatch monitoring data to derive bycatch mortality estimates.
2. Applying the relevant criteria and data quality considerations from the literature review to case studies using different sampling and monitoring approaches with an evaluation of the quality and certainty of bycatch assessments.
3. Applying a simulation framework to selected case studies and compare results to results from the aforementioned case studies.

From reading Annex 6 to the WKPETSAMP3 report, it seems that the workshop was charged with:

(1) Identifying criteria and best practices for designing a multipurpose program for sampling and estimating bycatch of PETS to assess population-level impacts.
(2) Making recommendations for improving monitoring systems for PETS bycatch at a Member State level and for regional level coordination.

ICES review guidelines ask reviewers to evaluate the following:

- Is the analysis technically correct?
- Are the scope and depth of the science appropriate for the request?
- Does the analysis contain the knowledge to answer the request for advice?

This report details Lisa Borges, Lee Banaka and José Castro comments and observations as external reviewers.

Lisa Borges

WKPETSAMP2

The report is clearly written and well organized, but section 5 on simulation model scenarios would benefit from an English language review. Several sentences are unclear due to for example using unspecific terms such as "fisheries characteristics" and "management programmes". Fisheries characteristics includes fishing behavior or only technical specifications? I'm not sure what a management programme is… a sampling programme detailed in a regulation? These terms should be defined or changed.

On section 3.2 Sampling Design the issue of the vicious circle of a fishery that is wrongly perceived to have low bycatch could be mentioned. A fishery in this situation may not have sampling attributed, and low or none sampling levels may not detect a rare bycatch event, which in turn will "prove" the fishery has low bycatch, perpetuating the initial wrong assumption of low bycatch. In this context, sampling programmes should be reassessed periodically and this fact should be highlighted, particularly considering the impact of climate change and consequent changes in species spatial distributions.

Section 5 – The simulations presented are very interesting and have produced several ways to move forward, but I notice that one of the conclusions of section 5.4 is that bias is not improved with stratification. I think this is an interesting result, and is discussed in WKPETSAM3, but should be discussed further in this report. There are also unspecific quantities that need addressing throughout this section text. For example, page 29 "*enough* data", "in cases where *enough* information" "*more* years of data" "*sufficient* spatial data are available." all these are extremely vague and not compared to a particular value so the reader wonders if 5 years of data are sufficient? or is it more 10 years? or maybe just 2 years?

Small remarks: Page 7 missing). Page 19 management systems replaced by reporting systems? Page 20 perhaps it is worth to refer to ongoing research projects like CIBBRiNA that are due to review efforts estimations in relation to PET bycatch estimations. Page 21 on sampling implementation, I understand the division between retained, discarded catch and bycatch of sensitive species, but I think its important to note that many discarded species are in fact sensitive species, and particularly when we referring to sharks and rays. Page 28 what are tier-based aproaches? Page 29 phrase "… are enough data to make hypotheses about sources of bias and uncertainty, there are sufficient data to make hypotheses about sources of bias or uncertainty." repeated? Page 29, 4th paragraph what framework?, what are closed-loop simulations? Page 30 "…specific context". specific objective?, what's an "index production"? page 31 I don't really understand Figure 5.1 scheme: for example why are diff types of uncertainty connected? Page 31 there is a good description of what métier levels are and I suggest this description be added in the beginning of the report when métiers are first mentioned. Finally page 50, "…waters outside Portugal", Portuguese EEZ? Offshore waters?

WKPETSAMP3

My main concern refers to the conclusion regarding reference fleets and observers programmes, and in my opinion regarding the misleading statement that reference fleets can be more effective. Namely, the following statement "*A reference fleet, given that the reference fleet include the full range of vessel-specific variation that practitioners believe exists, might be an effective way to collect data on bycatches, especially in cases where refusal rates in observer programmes could be high*." This first part of the sentence refers to the fact that the reference fleet should include vessels from all fisheries, but does it really? and at what proportion, e.g. 1 vessel per fishery? Also, has the impact of not fulfilling this assumption been tested in the effectiveness of the reference fleet to estimate bycatch? Because this assumption is applicable to observers programmes and EM programmes alike, i.e. that the vessels sampled should be representative of the fishery of interest, but this statement seems to imply that the reference fleet does not have a problem with this issue when the choice of participating in the reference fleet is voluntary. The other problem is related with the second part of the statement when it refers to refusal rates. Have actual real refusal rates been investigated and how much are they? Because this sentence refers to more than 50% refusal rates but have these ever been achieved in DCF observers programmes? If this is not the case or has not been investigated, then this is a misleading sentence. To a minimum the concepts "*high*" and "*full range*" in the sentence quoted should be clarified and specified, while caveats should be included regarding the limitations to the statement.

In addition, I think it is extremely important to highlight that EM differs considerably from a reference fleet, as the latest is based on self-reported data (namely on page 15). The reference fleet example operates in an industry with a high culture of compliance and somewhat less species diverse fisheries, which enhances participation and data quality, but that is in stark contrast to the reality in most EU MS fisheries. EM can also increase significantly the amount of fishing operations sampled but this is not a pre-requirement, as EM allows for only a subset of hauls be reviewed, so the cost and complexity of EM programmes can be much lower than observers programmes, and indeed likely of a reference fleet. EM also has the potential to store all data to be reviewed later in time if information or knowledge regarding the fishery and/or species changes, which is not the case with observers or reference fleet programmes.

On section 2.4.2 on monitoring small scale fisheries, perhaps there is a need to clarify that Electronic Monitoring is a part of a myriad of Electronic Technologies available that can be used to monitor fisheries activities, many of them simple and cost effective. For example, one can put a camera on board a vessel (ex. attached to a crew sorting in a small gillnetter) and does not need necessarily to be within an Electronic Monitoring system. However, issues common to all sampling programmes, for example the presence of "observer effect" would still apply.

On another aspect, I'm surprised that for rare bycaught species a very low percentage of sampling coverage (<10%) one may reach the BPUE accuracy and precision objectives. I was expecting much higher levels of sampling required, and I wonder why this is? My experience with estimating discards proportions is that their CVs can be much higher and variable, and I wonder if the CVs base simulations are not unrealistic lower?

Finally, I think the report should be more precise and succinct in the last section regarding making recommendations on monitoring programmes. From my reading, and based on the very good executive summary, to carry out a bycatch monitoring programme, one should do:

1. A pilot study to inform the probability or an estimated guess of likely bycatch probability (frequent - 1 bycatch event every 10 fishing operations, medium - 1 in 100, to very rare – 1 in 1,000 or 10,000 fishing operations);
2. Sample more vessels preferentially
3. Consider stratification if different métiers within a fishery, particularly with rarely bycaught species.
4. Use any sampling programme type, considering each has advantages and disadvantages, but should always be representative of the fishery
5. Reach at least 10% vessel sampling coverage to be able to detect rarely bycaught species, and to estimate BPUE with CVs of 30% for frequently caught species, although levels can be higher for fisheries with marked seasonal bycatch patterns (see last reviewer section, Castro et al., 2023).

I think clear bullet points as the ones suggested above could be included in the recommendations, while the schemes presented (figure 26-28) should have detailed explanation to make them clearer. Note the schemes are not just representing a monitoring scheme but part of a management strategy that includes monitoring, and therefore are incorrectly labelled.

Small comments: page iv – "WKPETSAMP3_that bycatch events" missing recommends, page 4 "In most ETP bycatch monitoring programs, the objective is to gather sufficient information to estimate bycatch levels and ensure that they are below a level that will not negatively impact the productivity of the ETP population" should be replaced by to, page 8 "EM programmes are more complex than observer programmes" should really be as complex, as observers need different types of training, safety & electronic equipment, databases and data checks, programme managers, etc., page 9 please use gender neutral word fishers, page 13 section 2.6.5 & table 2.2.2 should a country be identified?, page 30 "In general, the CV is may be less informative about the precision of monitoring" remove is, page 37 "between a fishing and ETP species are" remove a, page 40 "protected under EU laws" add national laws, page 50 section 5.2 "ensure that fisheries/métiers identified as high risk, by for example by ICES WGBYC, are sufficiently covered by sampling." please define what is sufficiently, section 4 and 5 title should refer to ToRs B & C.

Lee Benaka

It seems that responsive material for the above requests is presented for the most part in the WKPETSMP3 report. That is, the scientific literature review constitutes Annex 1 of the WKPETSAMP3 report. The case studies, and the comparison of the simulation framework data to the case study data, makes up Sections 2 and 3 of WKPETSAMP3. Criteria and best practices appear in Section 2, as well as Section 4, of WKPETSAMP3, and recommendations for improving monitoring systems can be found in Section 5 of WKPETSAMP3. WKPETSAMP2 sets the stage for the responsive sections of WKPETSAMP3 by providing an overview of selected monitoring programs and lessons learned, examples of simulations being used to guide monitoring programs, and an introductory simulation model.

I think the analysis used in these reports seems to have technical merit. Although I do not have a strong background in statistics, I was able to generally follow the simulations and understand the outputs of the models. I think the scope and depth of the science is appropriate for the request in terms of providing a helpful literature review, overview of various monitoring programs, and presentation of different case studies for meaningfully different modeling exercises. I think a possible shortcoming of these reports would be their responsiveness to the request to make recommendations for improving monitoring systems at a Member State level and for regional coordination.

Specifically, it would have been helpful for Section 5 of the WKPETSAMP3 report to circle back to the national monitoring programs described in Sections 2.1 to 2.7 of WKPETSAMP2. Section 5 includes some useful flowcharts tailored to the fishery case studies, with the idea that these could be adapted for particular Member State fisheries, but it could have been interesting to see such a flowchart adapted for a State fishery with a PET bycatch challenge. Another step that might be helpful would be to present concrete options for allocating a specific amount of resources to monitor a Member State's PET bycatch challenges, that is, can an approach be developed to identify the "true" monitoring priorities when there is not enough money to go around?

The recommendations in Section 5.2 of WKPETSAMP3 especially seemed in need of additional elaboration, for example, what Member States have not yet carried out robust risk assessments, and how can those States be incentivized or supported to carry those out? What fisheries/métiers have already been identified as high risk, and what level of sampling is deemed as sufficient coverage in light of limited resources? Can a specific framework for regional coordination be suggested, and how might that be effective?

My additional specific comments include the following:

- Each of the two reports should include an introductory list of acronyms. Some of the acronyms in the reports are not spelled out initially, which can be confusing for readers, for example, HELCOM and OSPAR on page 4 of WKPETSAMP2.
- All figures and tables should be cited within the text, and the cited figures and tables should generally appear following the paragraph that cites them. For example, Figure 2.2 in WKPETSAMP2 is not cited in the text, and Figure 5.2 in the same report is cited on page 31 but does not appear until page 36.
- For Section 5.2 of WKPETSAMP2, would it be helpful to create a matrix summarizing which types of models are best for different situations, and the relative characteristics of those models?

- In Tables 5.4 and 5.5 of WKPETSAMP2, the column headings (LR, df, p) should be explained or defined.
- On page 39 of WKPETSAMP2, should "Fig. 4 left panel" actually be Figure 5.4?
- Similarly, on page 42 of WKPETSAMP2, in the second full paragraph, there seems to be some confusion in the identification of tables being discussed. That is, should "table 3" be Table 5.3, and what about "Table1" and "figure 4a-c" and "figures 5a-c"? Please double-check that the tables are identified correctly.
- On page 21 of WKPETSAMP3, in the last sentence of the first full paragraph, should the phrase "Overall, the precision of BPUE increases as more vessels are involved in the monitoring scheme…" actually read, "Overall, the CV of the BPUE increases as more vessels are involved…"?
- On page 24 of WKPETSAMP3, in the last sentence of item #3, is the word "benefitting" the correct word to use? It reads as if increased bias is a benefit. I'm not sure how Figure 9 demonstrates what is being said there.
- For Figure 8 of WKPETSAMP3, I would rather see panels A and B on the same page and panels C and D on the same page to facilitate comparison.
- Figure 12 of WKPETSAMP3 should be cited in the text somewhere on page 30.
- In Table 4.1 of WKPETSAMP3, might Dead/alive and Weight be considered MANDATORY in some cases or for some types of PETS?
- In Figure 28 of WKPETSAMP3, please define the Species Conservation Status abbreviations; some are obvious, some (DD, CR, LC) are not.
- On page 57 of WKPETSAMP3, there seems to be a reference missing ("REF") in the third full paragraph on the page.
- On page 67 of WKPETSAMP3, are some equations (Eq. 11 and Eq. 12) missing?
- On page 81 of WKPETSAMP3, I don't understand why the refusalrate parameter is sensitive to so many factors. Isn't the strength of a Reference Fleet the cooperative aspect and presumed lack of refusals?

José Castro

WKPETSAMP2

Last paragraph of Section 2.4 (Spanish ongoing bycatch monitoring programmes) in page 10: "Using the area-specific correction factor, both parameters total and sampled fishing days are corrected in the same proportion so the estimated BPUE is equivalent to that calculated with the raw data". This conclusion is wrong. BPUE is not equivalent, in fact it can be doubled. What does not change is the estimate of total bycatch, since the total fishing days are recalculated in the same proportion as the sampled fishing days applying the same correction factor (compare Table 5 and Table 7 in Castro et al, 2023[5]).

Small remarks:

- Executive summary (page ii): 5th paragraph. 3rd sentence concluded with "explored". The period at the end of the sentence is missing.
- Section 2.2 (page 7): 2nd paragraph. The parenthesis needs to be closed in reference "Kindt-Larsen et al., 2023".
- Section 5.2 (page 29): 3rd paragraph. Sentence "there are sufficient data to make hypotheses about sources of bias or uncertainty" is repeated twice.
- Section 5.3.1 (page 31): "ICES WGBYG 2022" must be replaced by ICES WGBYC 2022.

[5] https://doi.org/10.1093/icesjms/fsad197

- Section 5.3.1 (page 35): last sentence "We pooled those estimates using weighted average (by monitor effort). To estimate bias and pooled CV". This sentence seems to be wrongly subdivided into two.
- Table caption of table 5.6 (page 42) is repeated twice.

WKPETSAMP3

I don't agree with concluding that reference fleets can be more effective than probability sampling programs with high refusal rate. This seems to be a particular case of the Norwegian sampling program. They define reference fleet (page 9) as "a group of fishing vessels that provide detailed information about their fishing activities, where sampling is carried out by a trained crew member". This does not guarantee the representativeness of the selected fleet. In the sampling programs of other countries, the reference fleet is understood to be the ad hoc sampling carried out on collaborating vessels, so the terminology can be confusing. For instance, Pan et al. (2022[6]) compares two periods of the same sampling program, one with a reference fleet (ad hoc) and other SRS with annotation of refusals, and only the first one showed to be biased. On the other hand, in the case of detecting bias in a random sampling program, the annotation of refusals can be an effective way to identify the source of bias (Castro et al., 2023).

Figures 3 & 4: "for cases of very high bycatch frequency (1 bycaught individual per 10 fishing operations) little statistical benefit (reduced bias and uncertainty) is obtained by increasing coverage levels beyond the simulation starting with coverage level of 0.5 %". As I mentioned above, this can be compromised in fleets with markedly seasonal bycatch.

Last paragraph in Page 30 underlines the CV's sensitivity to small changes in the mean when the mean is close to zero, which is quite common for some PET species. This idea is only commented on secondarily, but it could compromise the results of some simulations. Especially when species with low, medium and high incidental bycatch are compared under the same CV criteria.

The 2nd conclusion of "vessel effect and refusal rates" (page 32) indicates that "If refusal rates are high enough, an observer program will not be as useful as a reference fleet in reducing precision". Obviously, a high refusal rate can compromise the estimates derived from a sampling program. However, in sampling programs with a probabilistic design, the identification of refusals in the logbooks facilitates the subsequent exploration of both populations (accepted trips and rejected trips), helping to identify the source of bias.

Page 48: "A reference fleet, given that the reference fleet include the full range of vessel-specific variation that practitioners believe exists, might be an effective way to collect data on bycatches, especially in cases where refusal rates in observer programmes could be high". Once again, the authors make a characterization of what they call the "reference fleet" that seems very particular to their own sampling program.

Small remarks:

- 5th paragraph in "Executive summary" (page iii): word "impact" seems to be repeated.
- Section 3.1 (page 17), bullet 5th: sentence "or if there is first a selection and then boat (boat_samp)" should be clarified.

Conclusions & Recommendations

All three reviewers agreed that both workshops have worked hard and produce clear reports that detailed how the meetings reached the ToRs proposed. However, it was also acknowledged that WKPETSMAP3, although providing general methodological recommendations, should be

[6] https://doi.org/10.17895/ices.pub.21666584

nonetheless provide more specific recommendations to address the last ToR namely on providing "*recommendations for improving monitoring systems for PETS bycatch at a Member State level and for regional level coordination*".

In addition, all three reviewers disagreed with the conclusions reached that "*A reference fleet, given that the reference fleet include the full range of vessel-specific variation that practitioners believe exists, might be an effective way to collect data on bycatches, especially in cases where refusal rates in observer programmes could be high.*". The analysis detailed in the report does not support the conclusion reached. **We therefore strongly recommend that the conclusion regarding the reference fleet be redrafted and that the context, limitations and caveats of such conclusion be made clear.** Some of the caveats identified are: reference fleet is based on a voluntary scheme that may not be representative, operates in a culture of compliance that does not exist in many MSs, while self-reporting data of sensitive bycatch, and concepts "high" and "full range" of the sentence need to be defined. We also recommend that the conclusions of Castro et al. 2023 paper are taken into consideration, namely of prioritizing probabilistic sampling designs with annotation of refusals.

Smaller comments include that the minimum 10% coverage of fleet activity, proposed in the WKPETSAMP3 report, could be strongly compromised depending on the fishery bycatch pattern, and that both reports should include a list of acronyms.