Deliverable 6.1

Applying Methods to integrate currently collected fisheries data as part of regular reporting processes to understand how to best develop a monitoring programme

Version 3

Work Package no.	6
Deliverable no.	6.1
Lead Beneficiary	DTU
Topic	Bycatch monitoring design
Grant Agreement No.	101114301
Dissemination Level	PU
Date	27 Oct 2025

Contact: [WP6, David Lusseau, davlu@dtu.dk]

Co-funded by the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them.

Contact us at:
CIBBRiNA@minlnv.nl

Executive summary

This document outlines the approach used to help guide the design of a bycatch monitoring programme developed as part of the CIBBRiNA project. It also provides results based on the application of this method to given fishing conditions. The approach is centred around the application of a simulation platform which we developed to appraise the performance of monitoring strategies depending on the characteristics of the fisheries and the sensitive species considered. These results are presented to help illustrate the capabilities of the simulation platform with some applications for case studies. In addition to monitoring design recommendations, simulation of realistic fishing and monitoring conditions can also help define the power – a priori as well as a posteriori – to detect given effect sizes of mitigation techniques during trials. Bycatch mitigation trials are challenging to design as a number of considerations require to minimise sample size while at the same time dealing with a response variable which is often difficult to represent using traditional statistical distributions.

Here we showcase the simulation platform by providing insights on three long-standing challenges in bycatch monitoring: i. How to deal with realistic bycatch rate statistical distribution, ii. Understand under which circumstances stratification is useful, iii. Understand the consequences of monitoring bycatch at a 'Days-At-Sea" scale compared to the effort unit at which bycatch occurs (fishing operation level). We also provide insight for one of CIBBRiNA's case studies to help design an efficient and robust monitoring strategy. Finally, we help guide the design of mitigation trials for one of CIBBRiNA's case study by estimating a posteriori the power to detect mitigation effects given that no bycatch was observed in the first phase of the trial and a priori to define sequentially what sample size will be required in the second phase to determine mitigation effects.

This platform is now released and has been used to produce monitoring advice outside the project. It continues to be available to case studies of CIBBRiNA as the project progresses to help case studies develop monitoring schemes and mitigation experimental design.

Background to CIBBRiNA

The Coordinated Development and Implementation of Best Practice in Bycatch Reduction in the North Atlantic, Baltic and Mediterranean Regions (CIBBRiNA) project aims to minimise the bycatch of Endangered, Threatened and Protected (ETP) species in the North-East Atlantic, Baltic, and Mediterranean seas, working collaboratively as fishers, authorities, scientists, and other relevant stakeholders to achieve this. The species that we focus on include a variety of mammals, birds, turtles, and elasmobranchs (sharks, skates, and rays).

Through cross-border and cross-sectoral collaboration involving stakeholders from 13 European countries, CIBBRiNA is establishing mitigation, monitoring, and assessment programmes in a selection of fisheries with a higher risk of bycatch. Within a proactively fostered "Safe Working Environment" characterised by mutual trust, safety, and cooperation, we will build on a review of current approaches and learning from our Case Study fisheries to deliver an innovative toolbox designed to be integrated into policy and best practice in European fisheries management.

CIBBRiNA is co-funded by the EU's LIFE programme and runs from 2023 to 2029.

Contents

1. Introduction

1.1. Bycatch estimation: the statistical challenge

We are unable to have a complete census of every single bycatch event occurring during all fishing activity at this time. It means that we cannot measure bycatch rates and total bycatch, and we need to estimate both the rate at which sensitive species are caught incidentally in different fishing operations and the total bycatch emerging from the product of the bycatch rate and the total fishing effort. We cannot measure bycatch rate and total bycatch, we can only estimate those. This process of estimation relies on taking samples of fishing operations during which the occurrence of bycatch events can be observed. We consider a population of fishing operations, a statistical term used to describe all possible fishing operations of interest, whether they are observed or not, we know that for an estimate emerging from sampling that population to be precise and accurate (Figure 1), samples need to be representative, independent of each other, and large enough. If that is the case, then the bycatch rate and total bycatch estimates from those samples will be a precise and accurate representation of the 'true' bycatch rate and total bycatch of all fishing operations in the 'population' of fishing operations on which we focused. We can be sure of this because we have a mathematical theorem, called the Central Limit Theorem, which demonstrates that it is true for all instances respecting the sampling propositions listed above.

We know that multiple factors affect, to varying degrees, the probability that a bycatch event occurs during a fishing operation. The same applies to the number of individuals that will be accidentally caught during that bycatch event. Statistically, the bycatch rate therefore emerges from two statistical processes: does a bycatch event occur or not? If so, how many individuals are bycaught during that event? Also, for both processes, this implies that fishing operations can come from different 'populations' of fishing operations. This may be for example that bottom otter trawl with mesh size of 100-119mm targeting demersal fish are a statistical population and bottom otter trawl with mesh size of 40-99mm targeting demersal fish are another. We therefore need to ensure that our samples are representative of these different populations.

The probability of bycatch and the mean number of individuals caught during a bycatch event will depend on several factors. These are not necessarily the same: what affects bycatch probability may not affect the typical number of individuals caught during a bycatch event; and vice-versa. We know some of those factors already (e.g. the duration of a fishing operation or the population density of the sensitive species). Those are known known factors. From experience, we have some understanding that we are more likely to have bycatch when we fish in particular ways or under certain oceanographic conditions, but it is hard for us to articulate precisely why that is. Those are known unknown factors. There are also factors that can change bycatch probability and/or the mean number of individuals we will catch during a bycatch event, but we are unaware of those. Those are unknown unknown factors.

The challenge becomes how to best distribute samples of fishing operations (i.e. monitoring coverage) across the different 'population' of fishing operations to have a precise and accurate estimate of total bycatch. A secondary objective is to estimate the bias (lack of accuracy) and precision of our estimates given a particular sampling design. In both cases, how can we be certain, or at least confident, that we have met the challenge of achieving a precision and accuracy sufficient to use the estimates meaningfully?

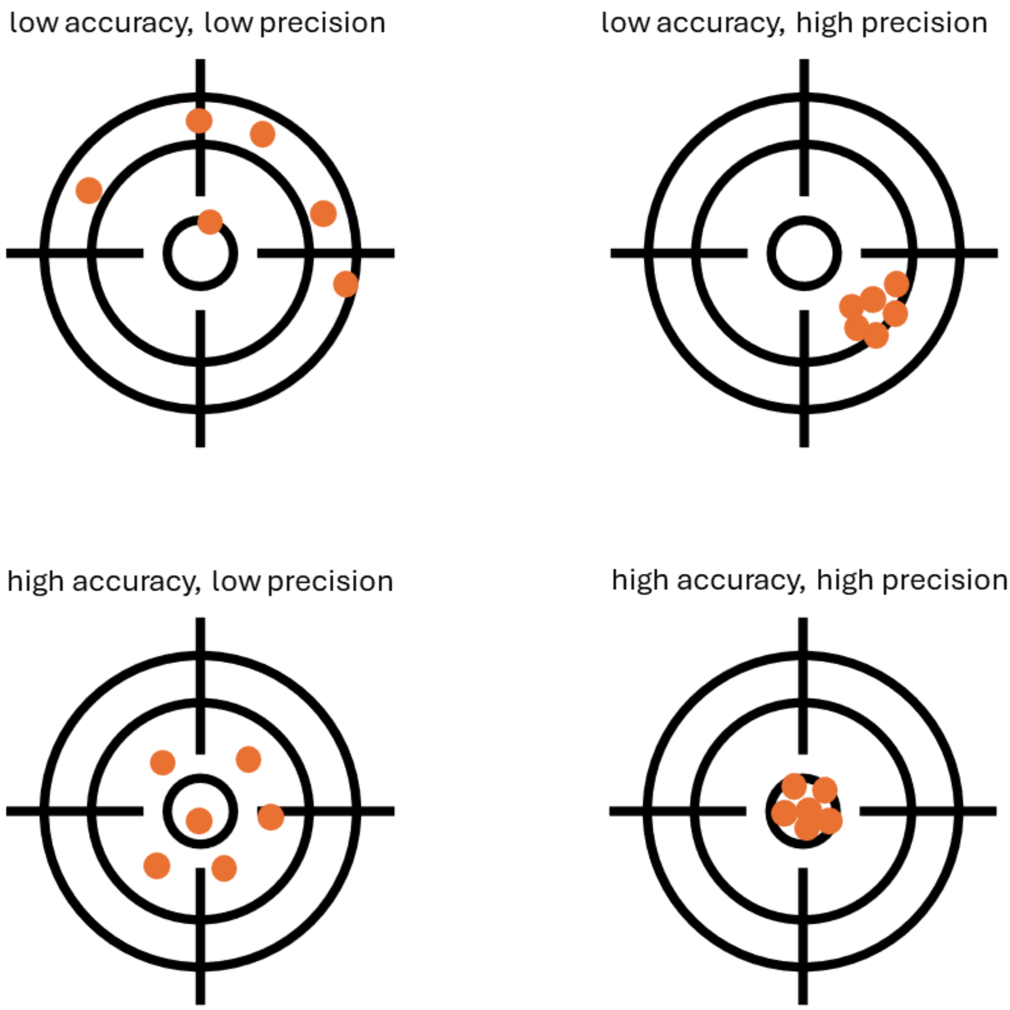


Figure 1. A visual representation of the statistical concepts of precision and accuracy.

1.2. Why simulations

A simulation is a way to describe and articulate a situation or "system" in a simple and constrained manner so that we can use computers to explore features, behaviours, and characteristics of the real, true system we are trying to grasp. We can replicate this simulation process many times and therefore gain insights from generalising across repeated outcomes.

We can rarely know with certainty whether the estimates of bycatch rate and total bycatch resulting from our monitoring design are accurate and precise representations of bycatch; given our understanding of the complexity of the circumstances leading to bycatch events. We have circumvented this shortcoming in the past by assuming that we can apply the Central Limit Theorem to the samples monitored and therefore use them as representing the population of fishing operations. However, the resulting estimates do not always match expectations based on empirical observations and on the raw data. Observers, scientists or fishers with extensive field experience on bycatch events perceive that the estimated bycatch rates seem to be at times over- estimating bycatch and at times under-estimating bycatch. Clearly, there are, at times, shortcomings to this approximative approach.

Here we try to understand how monitoring coverage design can affect precision and accuracy of bycatch rate and total bycatch given the processes described in the previous section. To do so we simulate a fishery using realistic characteristics. It can therefore either represent one population of fishing operations or be composed of multiple populations. We therefore know what the resulting ‘true’, simulated, bycatch is for that fishery. We then apply monitoring (sampling) to this fishery, using realistic characteristics of the ways bycatch monitoring can occur, and estimate bycatch rate and total bycatch based on the monitored observations. We can repeat the monitoring many times for the same simulated fishing operations, therefore obtaining many estimates for the same true (simulated) bycatch rate/total bycatch. We can then use these many estimates to measure, rather than estimate, the precision and accuracy of the monitoring programme.

1.3. Monitoring characteristics

Finally, it is worth spelling out a few additional constraints associated with monitoring logistics that increase the statistical challenge of bycatch rate estimation. Typically, to maintain the independence of samples, sampling design requires samples to be drawn at random from the population we aim to understand. In other words, we would like to be able to randomly draw the fishing operations which will be observed from all possible operations in the fisheries of interest. That is of course logistically challenging. Observers cannot shift from one vessel to another at each fishing occasion. Even when using electronic monitoring, we are in practice not able to cover all vessels and in most cases not even all a given fleet or area. So inherently, because we work at sea, bycatch monitoring schemes introduce some level of dependence between samples (for example, samples from a vessel are not independent and may also not be independent between some vessels). The effect of accuracy and precision is not systematically known and will depend on how much independence there is in bycatch itself between fishing operations within fishing trips.

2. SCOTI – Simulations for Characterising Optimal moniToring Implementations

This work benefited from contributions of partners outside the CIBBRiNA project. It was co-developed at three events organised by ICES: the annual meeting of the Working Group on Bycatch (WGBYC) in 2023, and the second and third workshops on Appropriate Sampling Schemes for Protected Endangered and Threatened Species Bycatch in 2023 and 2024. Further development and implementation took place between these proceedings and since.

The SCOTI development and application team is:

David Lusseau, Henrik Pärn, Paula Gutiérrez-Muñoz, Kim Magnus Bærum, Torbjörn Säterberg, Margaret Siple, Katja Ringdahl, Sara Königson, Estanis Mugerza, Allen Kingston, Gudjon Sigurdsson, Caterina Fortuna, Rita Vasconcelos, Gildas Gilmarec, Simon Northridge, Marjorie Lyssikatos.

The simulation is implemented in the R programming language. All code is available at https://www.github.com/dlusseau/SCOTI.

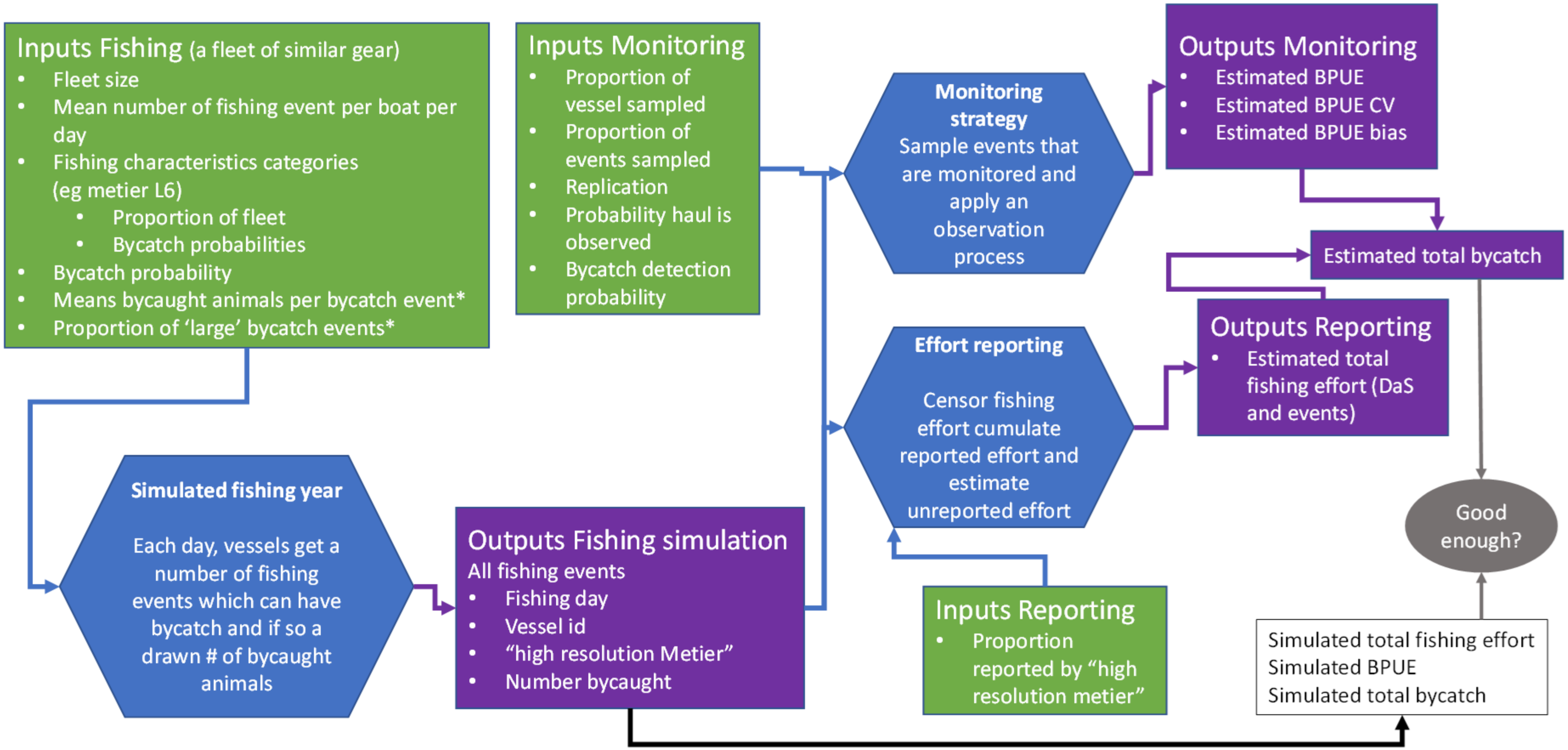
This work has contributed to the 2024 ICES Advice responding to the "EU request on appropriate bycatch monitoring systems at Member State level and on regional coordination" (https://doi.org/10.17895/ices.advice.25562220).

The simulation platform is versatile and provides mechanisms to consider aspects of fishing characteristics, species ecology, and monitoring schemes that can affect the precision and accuracy of bycatch rate (Figure 2). It is species independent. It can be applied to any fishery and sensitive species as long as we have data or expert opinion to understand how to parameterise any particular case of interest. Here we use Bycatch Per Unit Effort (BPUE) as a definition of bycatch rate and the effort unit considered is 'Days-at-Sea" (DaS). These were selected to account for existing reporting schemes. We know that DaS can be a complicated measure of effort in several fisheries and further work in CIBBRiNA WP6 will tackle the relation between DaS and higher resolution measures of fishing effort to better understand the consequences of the fishing effort unit selected.

The simulation process involves first simulating a fishing year (365 days) for a fishing segment (e.g. fishery, fishing métier/gear in a particular region). The characteristics of the fishing operations are tuned to a case study of interest using summary statistics for that particular case study. Once the *simulated fishing year* (Figure 2) is produced, then 'true' BPUE and total bycatch is measured (called here simulated BPUE and simulated total bycatch). Then, monitoring is applied to the simulated fishing operations. The monitoring strategy is either tuned based on an existing monitoring programme or developed based on scenarios that users want to explore. The same monitoring strategy is applied many times (at least 1000 times) and BPUE is estimated each time. This results in a distribution of BPUE estimates from which precision can be estimated. We then estimate bias by comparing estimated BPUEs based on monitoring strategies to the simulated BPUE.

The central objective of subsequent analyses is to understand how changes to monitoring strategies can meaningfully improve the precision and accuracy of BPUE and total bycatch estimates in a realistic manner. We only focus on BPUE at this stage in the examples provided.

In this section, we outline the details of each of those components as the simulation platform stands in February 2025. Extensions are possible if case studies require it.



*number of animals bycaught at a bycatch event is simulated as a mixture of two processes: 'large' bycatch events and 'typical' bycatch events

Figure 2. Workflow diagram of SCOTI identifying the main functions to characterise the simulated fishing 'year' and the simulated monitoring strategy. For convenience fishing characteristics may be categorised using métiers and their levels (https://vocab.ices.dk/?codetypeguid=edfc216f-b466-4c33-84f6-da2c998fe09f). Outputs are in purple. Outcome is a contrast between the 'true' (simulated) bycatch (white box) and the estimated bycatch based on monitoring.

The inputs described in the following sub-sections and summarised in Figure 2 are tuned parameters, i.e. they are informed by case study specialists. The information can be semi-quantitative or quantitative, i.e. respectively case study specialists can use fisheries data to estimate the parameters or expert opinion to qualify them. It may be that case study specialists would like to run "what-if" scenarios to explore bycatch patterns under hypothetical scenarios. All those approaches are possible with this simulation platform. The interpretation of results will then depend on the assumptions made during parameter setting and tuning.

A formal description of the simulation can be found in the report of the third workshop on Appropriate Sampling Schemes for Protected Endangered and Threatened Species Bycatch (https://doi.org/10.17895/ices.pub.25061522.v2).

2.1. Fishing

All code for this section is available in the function "make_fishing_year_metier.R". We first simulate a fishing fleet for one year. Note that this process can be replicated to investigate the effect of stochasticity emerging from the random processes introduced by drawing randomly from a statistical distribution (see below) in the simulation on the BPUE. All inputs are summarised in Table 1.

Fishing is simulated over 365 fishing days during which n vessels can fish (n_{vessel} is an input which is the size of the fleet considered for a particular case). On each day, each vessel is assigned a random number of fishing events (a unit of fishing operation such as a haul). The number of fishing events per vessel per day is drawn randomly from a statistical distribution that is informed by two inputs: λ_f the mean number of fishing events per vessel per day, and whether this mean differs between vessels in the fleet or is homogeneous among all vessels (stochastic). If the input is that the mean number of fishing events per day varies between vessels, then at the start of the simulation each vessel is assigned a mean number of fishing events per day (λ_{fi}) randomly drawn from a statistical distribution informed by λ_f.

The fleet is then simulated to fish for the year as described above. At each fishing event, a bycatch event can happen with a given probability $p_{bycatch}$ which is an input. This probability can vary by fishing characteristics and by sensitive species. If a bycatch event occurs, the number of individuals bycaught is randomly drawn from a bimodal statistical distribution. The inputs for that distribution are the probability that the bycatch event contains many individuals (p_{large}) and the mean number of expected bycaught individuals for both a regular bycatch event and a large bycatch event. This feature was introduced to capture realism from some bycatch situations where the observed bycatch numbers per bycatch event appear to have this feature where most events have a small number of bycaught individuals, and a few have a very large number of individuals (can be especially important for sea birds but also others). This intrinsic feature of the bycatch process can severely bias BPUE estimates and so it was important to introduce it to determine its effects. Note that p_{large} can be set to zero when no such feature is expected.

We can set fishing and bycatch characteristics by features of the vessels. This is implemented by an input métier which can be specified for each vessel or qualified as a proportion of the fleet belonging to each métier. Métier is a term adopted by ICES to capture fishing characteristics with varying level of details: métier level 3 captures fishing gear groups (https://vocab.ices.dk/?ref=1497), level 4 captures gear types in gear group (https://vocab.ices.dk/?ref=1498), level 5 captures target species with level 4 gear type (https://vocab.ices.dk/?ref=1499), and level 6 captures fishing activities to access target species (https://vocab.ices.dk/?ref=1647).

The simulation outputs a table containing all realised fishing events with their bycatch characteristics (Table 2). This forms the basis of the information which is then sampled when applying the monitoring strategy next.

Table 1. Input parameters for the fishing simulation.

Parameter	Parameter name in function	Data type	Description
$\lambda_{bycatch}$	mean.bycatch.event	Real $(0,\infty)$	Expected number of animal bycaught given that a "regular" 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.
$\lambda_{largebycatch}$	mean.bycatch.large.event	Real $(0,\infty)$	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. $\lambda_{largebycatch} > \lambda_{bycatch}$
$p_{bycatch=large\|bycatch_{i,d,h}=1}$	p.large.event	Real [0,1]	The probability that a large bycatch event occurs, given that a bycatch occurs. subscript i is for vessels, d for fishing day and h for bycatch event.
n_{vessel}	Nvessel	Integer ($n_{vessel} \geq 1$)	Number of vessels in the fleet.
λ_f	mean.fishing.event.vessel.day	Real $(0,\infty)$	Mean number of fishing operations conducted per vessel and day (when stochastic=FALSE).
$\lambda_{\bar{f}}$	mean.fishing.event.vessel.day	Real $(0,\infty)$	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 p_{metier} vector.
p_{metier}	p.metier	Simplex vector	Distribution of métiers in the fishery. This vector distributes metiers to vessels such that $\sum_{i=1}^{n} p_{metier,i} = 1$.
	Stochastic	Logical	Should the mean number of fishing operations per vessel and day be drawn randomly?

Table 2. Output of the fishing simulation is a table with the following information provided for each fishing event in the simulated fishing year (table dimensions: five columns and as many rows as there were fishing events in the simulated fishing year).

Parameter	Data type	Description
fishing.day	Integer {1,2,…,365}	Day of the year for the fishery simulation.
Vessel	Integer {0,1,…,nvessel}	ID of vessel.
Metiers	Integer {1,…,nmetier}	Métier category.
Bycatch	Integer {0, 1}	Indicator of whether bycatch occurred or not.
nbycatch	Integer {0,1,…, ∞}	Number of bycaught individuals.

2.2. Monitoring

All code for this section is available in the function "monitor_BPUE_metier.R". The monitoring functions aim to offer flexibility to design a number of monitoring strategies. This minimum set of monitoring options can be extended to fit additional considerations in which case studies may be interested (for example seasonality in bycatch patterns associated with sensitive species ecology).

Given the fishing simulation outputs (Table 2) we first define the sampling design we want to apply using inputs (Table 3):

vessel_samp=FALSE & bymetier=FALSE: random monitoring – this corresponds to a design when we have full control of which fishing day is monitored, regardless of the vessel which carried it out. There is no stratification of sampling.

vessel_samp=TRUE & bymetier=FALSE: sampling vessels – in this situation, a proportion of vessels (p_{vessel}) are sampled and for each of those vessels a proportion of their fishing days is sampled ($p_{monitor|vessel}$).

vessel_samp=TRUE & bymetier=TRUE: sampling is stratified by métier – in this situation the fishing fleet had multiple métiers, which may or may not have different bycatch probabilities, and sampling is stratified (distributed) between métiers.

Currently, the fishing simulation sets the métier at the vessel level rather than at the fishing event level. This can be extended, if need be, but it explains why the sampling design vessel_samp=FALSE and bymetier=TRUE is missing.

Finally, we introduce observation processes on the monitored fishing events which aims to capture some of the factors that may lead to lack of observations for planned monitoring events. This is based on inputs: the probability that a fishing event was observed by an observer or electronic monitoring ($p_{fishing\_event\_obs}$), the probability that when a bycatch event was observed an individual bycaught animal was detected out of all individuals bycaught in that event ($p_{detection}$), and the probability that a vessel selected for monitoring refused to engage ($p_{refusal}$).

The number of vessels and the number of fishing days that are to be monitored can be varied to understand how monitoring coverage under different sampling design affects BPUE precision and accuracy. Once a monitoring strategy is set, fishing days are randomly drawn from the simulated fishing table (Table 2) using inputs (Table 3) and the BPUE estimated as the mean bycatch per unit effort (here fishing event is the effort unit, but we can also estimate BPUE using DaS as effort unit).

To estimate the precision and accuracy of BPUE estimates (BPUE_est) we replicate the application of monitoring to fishing many times (n_{sample}, typically 1000 times). We can then estimate the variance of BPUE_est using its coefficient of variation (Table 4).

Table 3. Input parameters for the monitoring simulation

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

Table 4. Output parameters for the monitoring simulation

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 metiers 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 metiers are simulated in the fishery a CV estimate for each métier is provided.
effort_mon	Real (effort_mon ≥ 0) or a vector of reals	Mean effort for BPUE estimates. If multiple metiers are simulated in the fishery an effort value for each métier is provided.

2.3. Analyses

All code for this section is available in the function “estimate_bias_and_precision.R”. Following the simulation of fishing and the simulation of monitoring we end up with a known BPUE (Table 2) and a mean BPUE estimate and its CV based on our monitoring strategy for this known BPUE (Table 4). A measure of bias (lack of accuracy, *sensu* Figure 1) can also be estimated by contrasting $BPUE_{estimate}$ and the real BPUE.

The whole process can be replicated across multiple fishing years (year), typically 100 to 1000 replicated fishing years. The outcome is a table with year estimates of BPUE CV and BPUE

bias for each monitoring scenario. We therefore have estimates of precision and accuracy and we can assess how those change when we change the monitoring strategy, or the knowledge about the fishery.
In the next sections we apply SCOTI to generic questions about monitoring design. Multiple examples of SCOTI implementation on real world examples are available in the third workshop on Appropriate Sampling Schemes for Protected Endangered and Threatened Species Bycatch (https://doi.org/10.17895/ices.pub.25061522.v2). Here we provide two additional examples to show the potential for this simulation platform and stimulate ideas for case studies. In addition, a workshop was held with case studies during the 2025 CIBBRiNA Spring meeting in Vigo, Spain, to identify case-study relevant SCOTI implementation and build modelling scenario proposals. Two types of questions emerged from this process: i. detailed implementation of SCOTI for case-study specific monitoring design, ii. prospective analyses of the use of simulation for mitigation trial power analyses. The next sections present these four types of implementations.

3. The influence of bycatch patterns on bycatch rate estimates

Species density can influence the number of individuals caught during a bycatch event. Higher density than usual can happen for example simply because individuals are aggregated on a prey patch. This means that for some species the realised distribution of the number of individuals caught per bycatch event can be complicated (Figure 3) and this can affect the BPUE estimates. We demonstrate this with a simple set of simulations where p_{large} varies for a given $p_{bycatch}$.

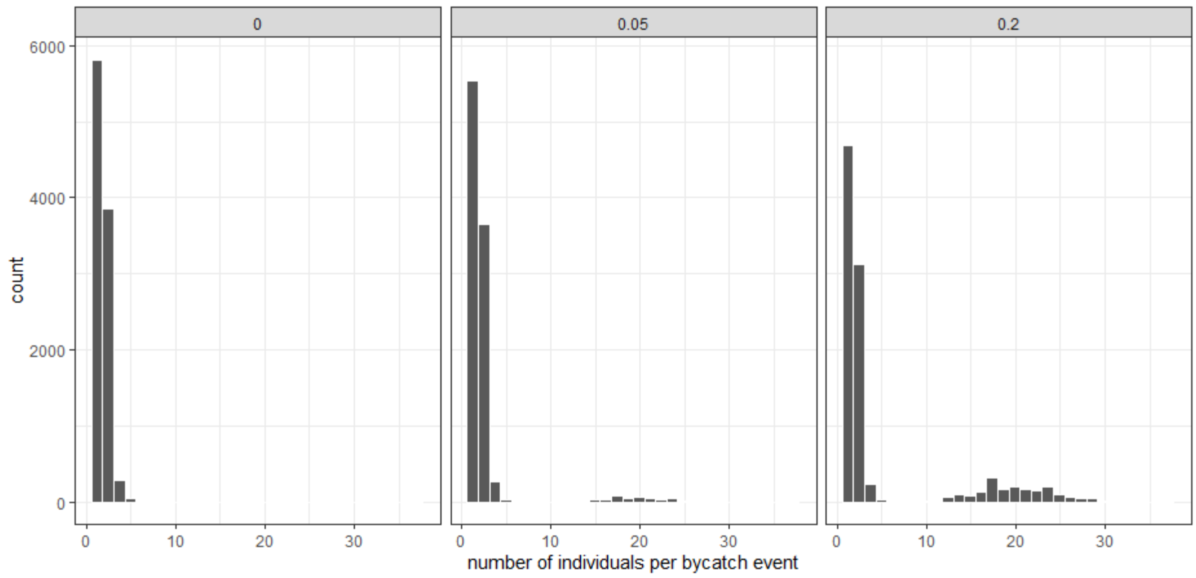


Figure 3. Illustrative example of the distribution of 10,000 simulated bycatch events (number of individuals caught). Panels show three different probabilities that a “large” event takes place (p_{large} = 0, p_{large} = 0.05, and p_{large} = 0.2).

Here how this simple example is parameterised:

We assume a fishing fleet of 30 vessels that carry out on average 2.5 fishing events/day/vessel. This fishing pattern is the same for all vessels (stochastic=FALSE). Therefore, on average a simulated fishing year will contain 27,375 fishing events.

When a ‘regular’ bycatch event occurs on average 2 individuals are bycaught and when a large bycatch event occurs on average 20 individuals are bycaught.
We simulate the probability of a bycatch event to be either 0.005, 0.01, or 0.1. For each of these probabilities of bycatch we simulate a probability of a large bycatch event to be either 0, 0.05, 0.1, or 0.2. We therefore have 12 sets of fishing conditions where only bycatch patterns change.

We then randomly monitor the simulated fishing (vessel_samp=FALSE & bymetier=FALSE). We assume there is no monitoring refusal, that the detection probability of bycaught individuals is 1 and the probability to observe a monitored fishing event is 1. We vary the proportion of fishing events observed to be either 0.01, 0.05, 0.1, 0.25, or 0.5. As a result, we have 5 monitoring strategies and therefore simulate 5 monitoring strategies applied to 12 sets of fishing conditions, which is 60 unique simulation sets. We replicate the monitoring 1000 times and replicate the fishing simulation 100 times. We therefore carry out 1200 fishing simulations and 6,000,000 monitoring simulations.

As an indication, as the code stands, it took 80 minutes on a Core i9-10920X CPU without parallelisation to carry out the full simulation implementation in this case including the production of BPUE mean, CV, and bias estimates for each replicated simulated fishing and monitoring conditions.

The resulting 6000 BPUE mean, CV, and bias estimates can then be used to estimate the association between bias and precision and the factors we varied in the simulation. We can model and predict BPUE bias (Figure 4) and BPUE CV (Figure 5).

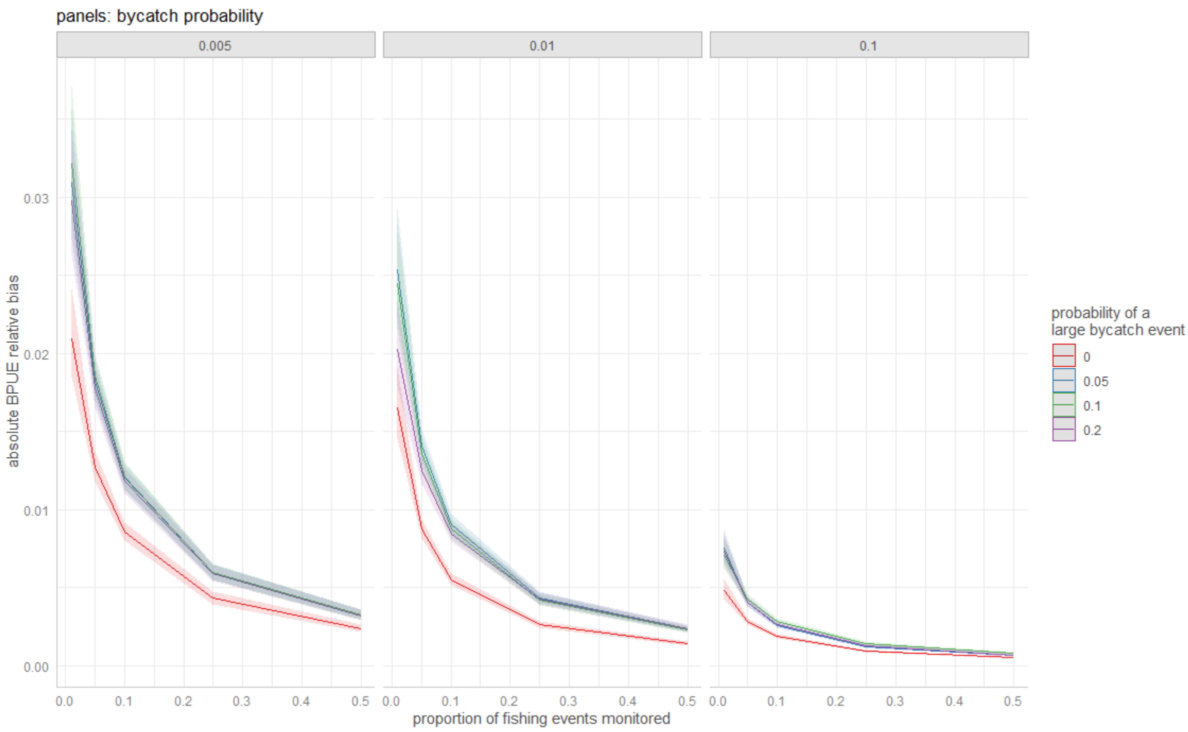


Figure 4. The predicted effect of bycatch probability (columns), probability of a large bycatch event (line colours), and the proportion of fishing events monitored (x-axis) on the **BPUE estimate bias** (absolute BPUE relative bias: 0.01 = 1% of BPUE). Predictions obtained from the best generalised linear model (Gamma residual distribution) including the 3-way interaction between the three factors (proportion of events monitored, bycatch probability and probability of a large event).

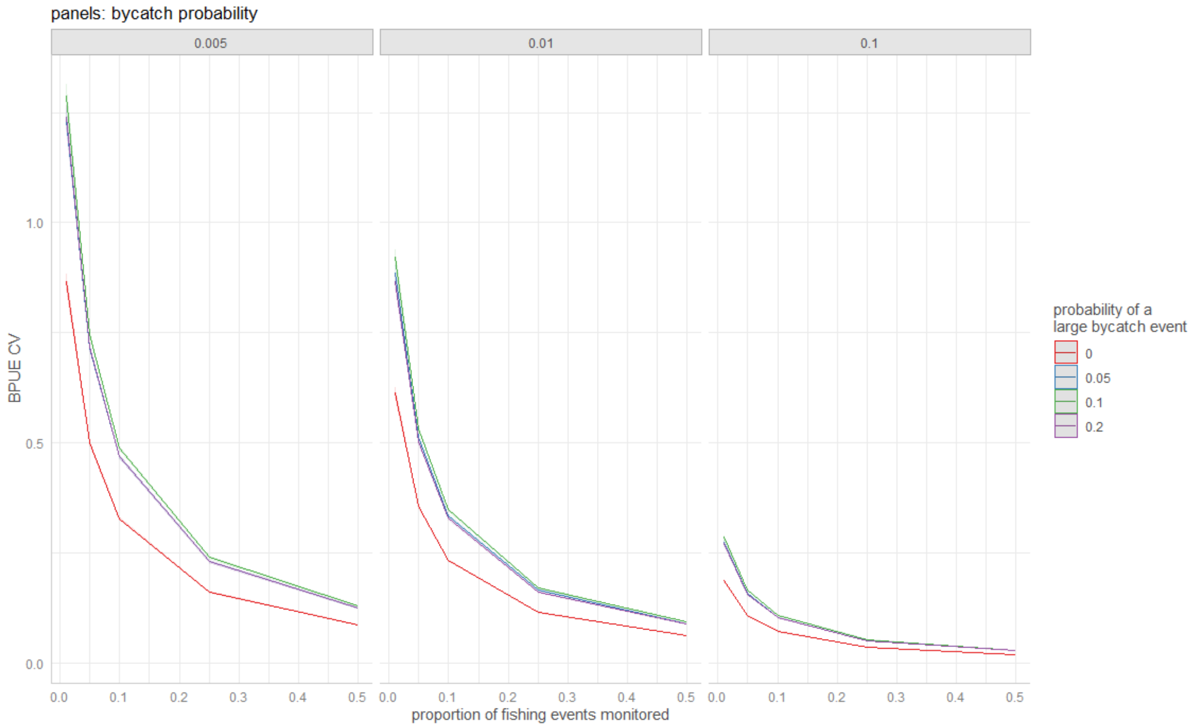


Figure 5. The predicted effect of bycatch probability (columns), probability of a large bycatch event (line colours), and the proportion of fishing events monitored (x-axis) on the **BPUE estimate coefficient of variation (CV)**. Predictions obtained from the best generalised linear model (Gamma residual distribution) including the 3-way interaction between the three factors (proportion of events monitored, bycatch probability and probability of a large event).

This simulation helps to show that beyond the effect of monitoring coverage (the more, the better), bycatch patterns also affect the bias and precision of BPUE estimates. When bycatch is rare (left columns), we need substantially more coverage to achieve similar levels of accuracy and precision as we can achieve with less coverage when bycatch is less rare (right columns). Any complexity in the patterns of the number of individuals bycaught during a bycatch event also decreases accuracy and precision. This can be seen as the difference in CV and bias is very small between all cases when a large bycatch event could occur (all line colours except red, $p_{large}>0$, Figures 4 and 5) compared to the difference between those cases and cases when there were no large bycatch events (red lines in Figures 4 and 5). When large bycatch events become more common, i.e. when p_{large} becomes larger, then the CV and bias decrease (get closer to the red line representing no large events). It is because by chance, we are more likely to be able to capture the complexity of the bycatch rate in our samples (large bycatch events are better represented). This decrease in bias and imprecision would continue as we increase p_{large}.

3.1. Implementing a case study

Here we walk through a synthesis of one of the examples used in the third workshop on Appropriate Sampling Schemes for Protected Endangered and Threatened Species Bycatch (https://doi.org/10.17895/ices.pub.25061522.v2) to showcase how a simulation can be tuned to ask a specific question.

In all monitoring programmes, the number of fishing events (operations) that can be monitored is limited, primarily due to budgetary constraints. A classical challenge is therefore to decide whether monitoring coverage should be limited to a few vessels in the fleet to which more monitoring can be dedicated or whether monitoring coverage should be spread to as many different vessels as possible even if it means a much lower monitoring coverage per vessel.

At the same time, bycatch rate may vary by fishing characteristics (e.g., métier level 6 within a particular métier level 4) and so it may be advantageous to stratify sampling (e.g. by métier level 6) to increase precision. Finally, it is not uncommon for a fishery to face multiple bycatch challenges, since several sensitive species can be bycaught in the fishery with different rates (i.e., the combination of bycatch probability and the typical number of individuals bycaught per event). In such a, rather typical, case we face multiple interacting challenges: i) we cannot monitor all fishing operations, ii) different segments of the fleet will have different bycatch probabilities, and iii) we need to find a consensus monitoring strategy across multiple species which have different bycatch probabilities.

What is then the best sampling design?

Here we simulated four monitoring strategies: i) fishing events (operations) randomly monitored, ii) monitoring coverage stratified by métier, iii) monitoring coverage focussed on a proportion of the fleet, iv) monitoring coverage stratified by métier and focussing on a proportion of vessels in the fleet from each métier.

In ICES WKPETSAMP3, bycatch rate is defined as: Very high – one individual taken as bycatch per 10 fishing operations, Medium/High – one individual per 100 fishing operations, Rare – between one individual per 100 and one individual per 1 000 fishing operations, Very rare – between one individual per 1 000 and one individual per 10 000 fishing operations, Extremely rare – one individual per > 10,000 fishing operations (https://doi.org/10.17895/ices.advice.25562220.v1).

In this example the fishing fleet bycatches three ETP species, one rarely (species 1), one commonly (species 2), and one often (species 3). All three species display complex patterns in the number of individuals caught per bycatch event which can be described by a 'typical' bycatch number and a smaller proportion of bycatch events having larger numbers of individuals. The probability of a large bycatch event is more than 0.1 for all. The fleet is composed of many tens of vessels and more than two types of fishing characteristics that have different bycatch rates for all three species (but somewhat within the same order of magnitude).

We have an outcome for each species (Figure 6-8). Each outcome is displayed in the same manner: we have four panels, one for each monitoring strategy.
The proportion of fishing operations monitored is comparable across the four panels. The legend in the bottom row of panels represents the proportion of fishing operations monitored for the whole fleet (so the blue line in the bottom row panels corresponds to a value of 0.05 on the x-axis of the top row panels). In the *top row of panels* this monitoring effort/coverage is assigned at random across all vessels. In the *bottom row* first vessels participating in monitoring are selected (a particular proportion of the fleet given on the x-axis of the bottom row panels) and then that monitoring coverage is distributed randomly on all fishing operations of those vessels. The *right column* represents outcomes when the sampling in both cases (operations selected at random or first vessels selected at random and then random operations of those monitored) is stratified by métier. The different métiers have different bycatch rates and so stratification is thought to assure that the métiers most used receive more monitoring to increase representativeness.
Note that the scale of the y-axis is the same for all figures and all panels to increase comparability.

For the same total monitoring coverage, the best monitoring design will be different for different species. For the rarer species, stratification improves precision (Figure 6, top row) for all levels

of monitoring coverage. Precision is degraded (CV larger) if we focus the monitoring on a few vessels (comparing top and bottom row of the right column in Figure 6).
For the other species it is possible to increase the BPUE estimate precision by focussing the monitoring to a subset of the fleet, but it takes a substantial proportion of the fleet to have a meaningful effect on precision (Figures 7-8). Given the monitoring coverage considered, this focus on monitoring specific fleet segments does not differ much from random sampling of all fishing operations. Further constraints on the sampling design (focussing on a subset of the fleet and stratifying by métiers) leads to detrimental effects on the BPUE estimate precision. Such a sampling design captures well how the monitored subset of the fleet operates, but that advantage means that it is worse at generalising bycatch patterns to the whole fleet. If the fleet was more homogeneous in bycatch patterns it would probably be less of an issue.

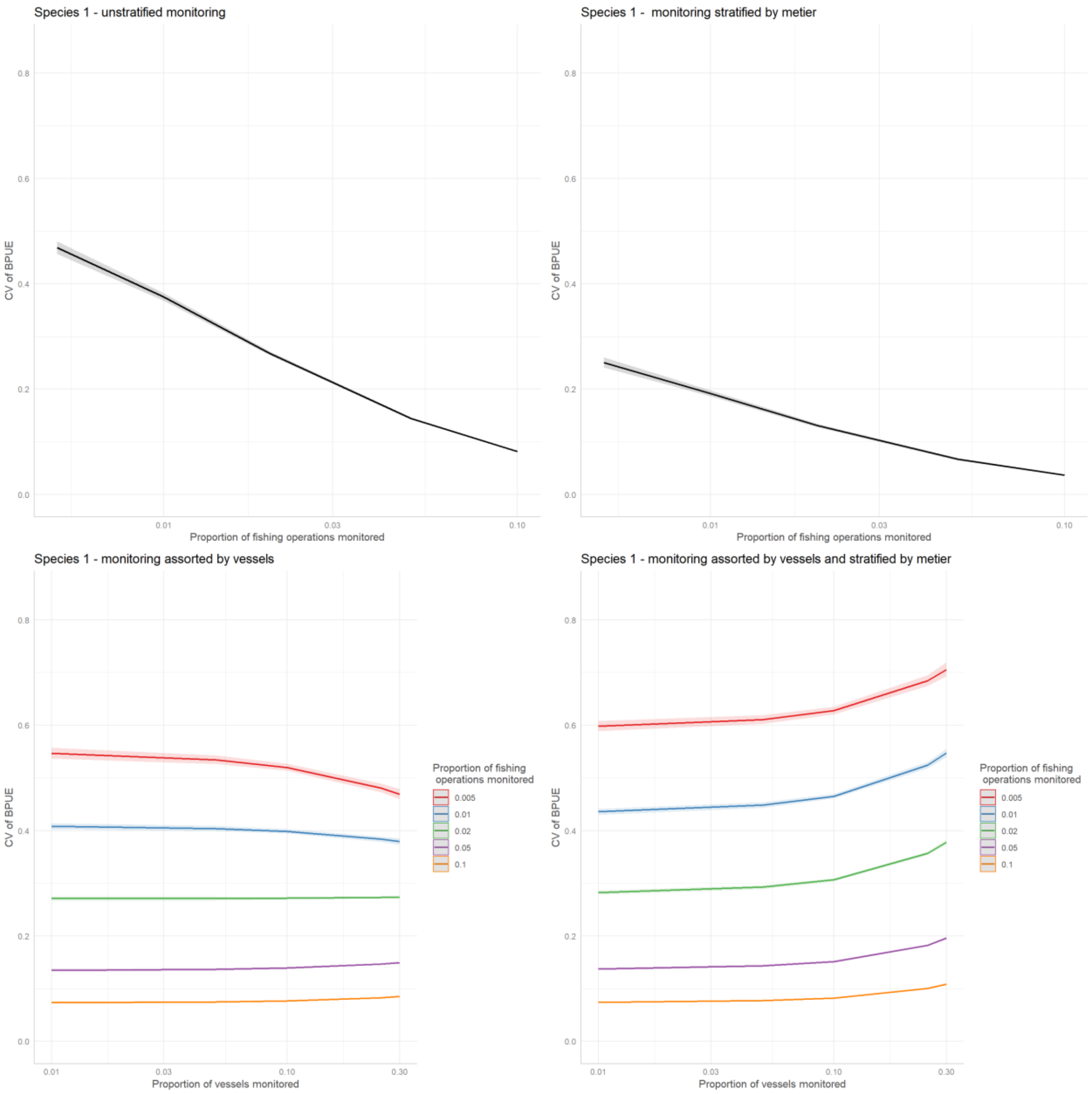


Figure 6. Predicted BPUE estimate CVs for species 1 (bycaught rarely) under the four monitoring strategies (four panels), with 1000 monitoring replicates per fishing year and 100 year replicates. Predictions from the best generalised linear model associating CV to the interaction between proportion of fishing operations monitored (x-axis in top row and line colours in bottom row) and the proportion of vessels monitored (x-axis in bottom row) explanatory variables assuming a Gamma residuals distribution.

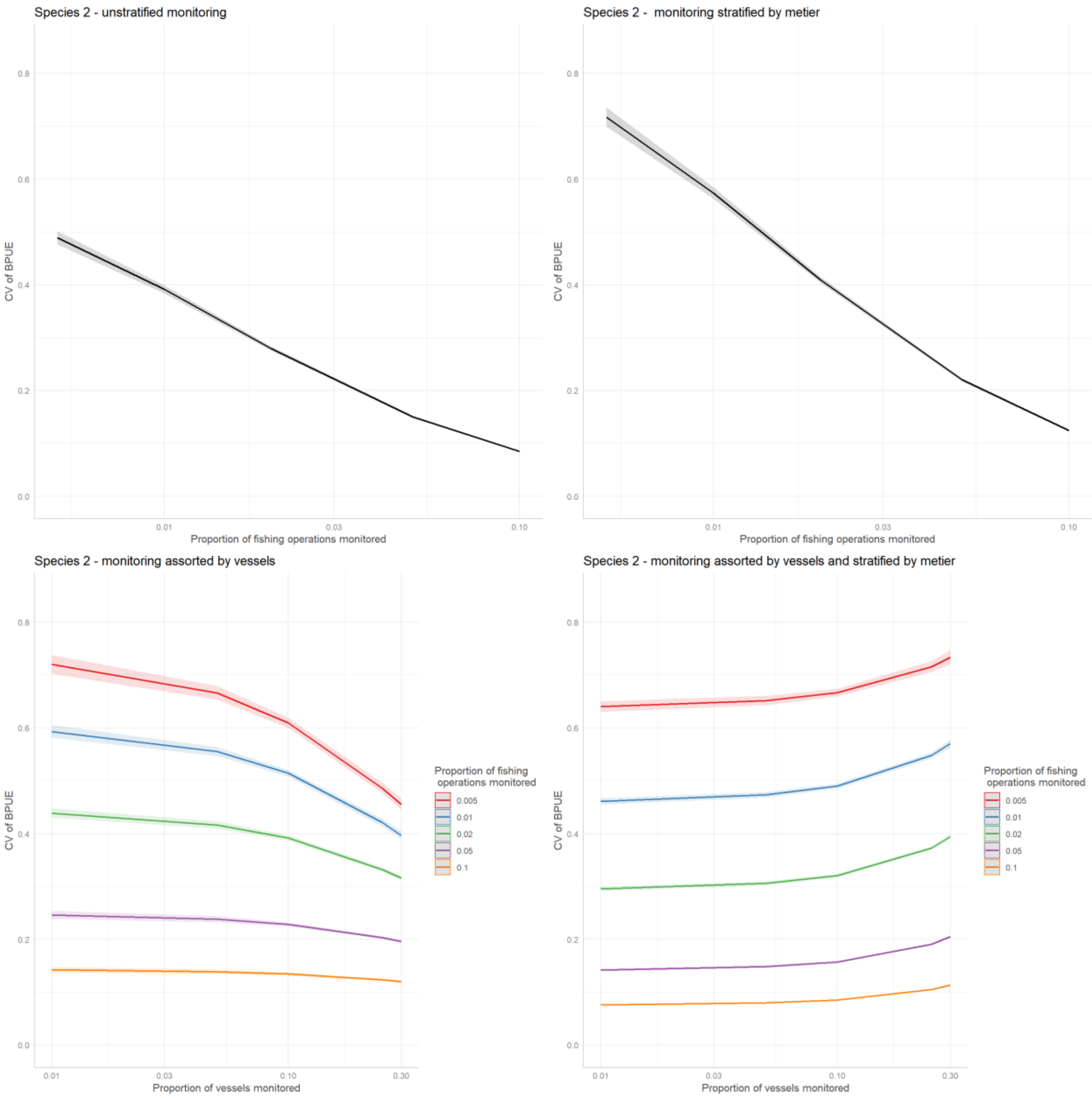


Figure 7. Predicted BPUE estimate CVs for species 2 (bycaught commonly) under the four monitoring strategies (four panels), with 1000 monitoring replicates per fishing year and 100 year replicates. Predictions from the best generalised linear model associating CV to the interaction between proportion of fishing operations monitored (x-axis in top row and line colours in bottom row) and the proportion of vessels monitored (x-axis in bottom row) explanatory variables assuming a Gamma residuals distribution.

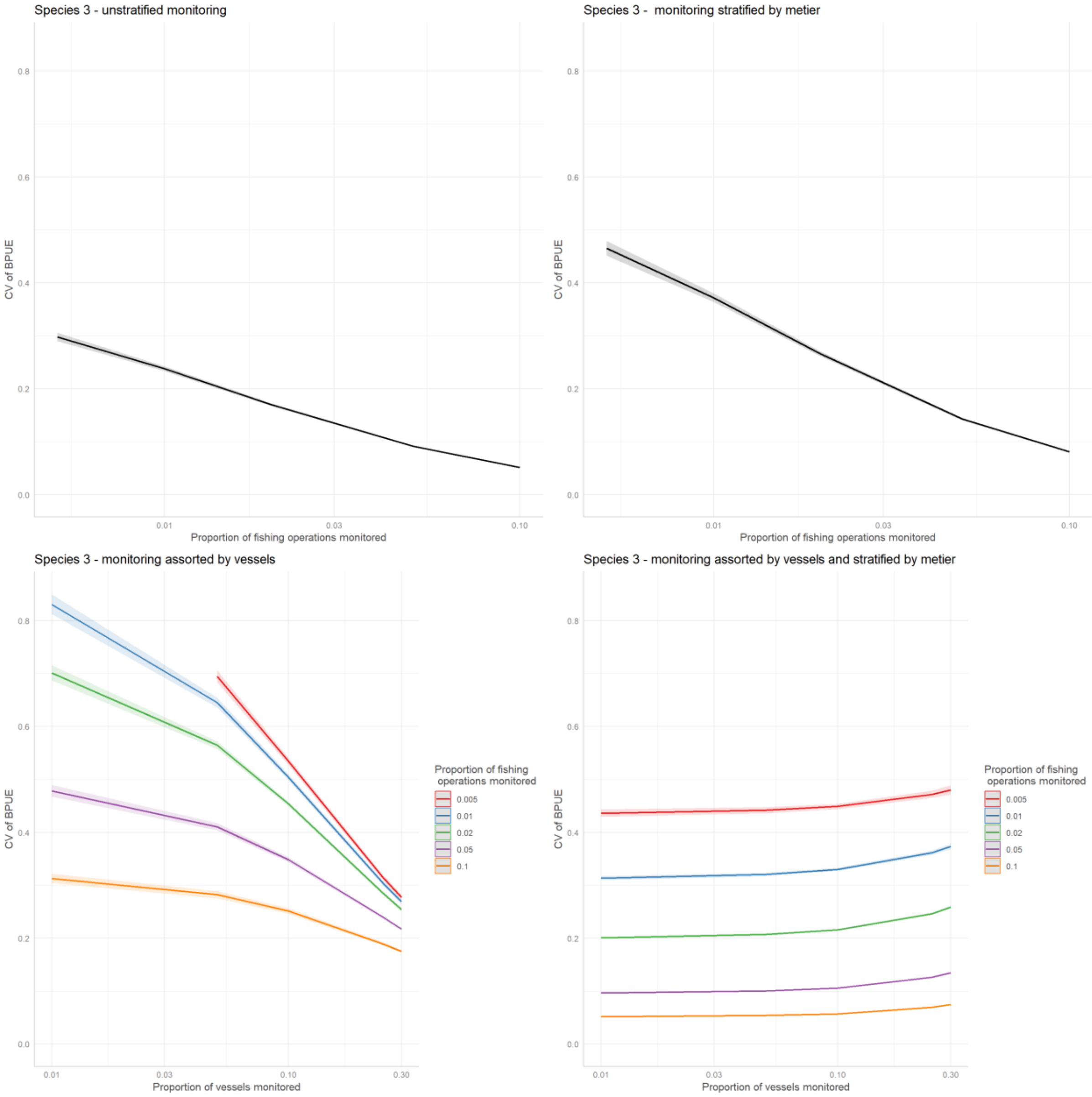


Figure 8. Predicted BPUE estimate CVs for species 3 (bycaught often) under the four monitoring strategies (four panels), with 1000 monitoring replicates per fishing year and 100 year replicates. Predictions from the best generalised linear model associating CV to the interaction between proportion of fishing operations monitored (x-axis in top row and line colours in bottom row) and the proportion of vessels monitored (x-axis in bottom row) explanatory variables assuming a Gamma residuals distribution.

4. The consequences of the scale at which bycatch rate is reported for the precision and accuracy of bycatch estimates

4.1. Introduction

Fishing effort in bycatch estimation is typically calculated at two main levels: 1) Operational level: This includes metrics like number of hooks deployed, net length, tow duration, or number of sets. This method provides a more precise measurement of effort directly related to fishing operations, or 2) Days-at-sea level: This is a broader temporal measure that accounts for the time vessels spend fishing but does not account for variations in fishing intensity within those days. Days-at-sea level reporting is more commonly available through vessel monitoring systems and logbook requirements, while operational-level effort data requires more detailed observer programs or electronic monitoring systems that have more limited coverage due to associated costs and labour. Here, we simulated if the choice between the monitoring effort reporting level, operational level or days-at-sea level, affects the accuracy and precision of total bycatch (TB) estimates.

The design of monitoring programs has also a direct effect on the reliability of total bycatch (TB) estimates. Monitoring assumes that data collected from a sample of fishing operations can be extrapolated to represent the entire fleet. However, this assumption is only valid when fishing activities, and the related bycatch probabilities, are relatively homogeneous across the fleet. When important differences exist, the design of the monitoring program determines whether those differences are captured or ignored.

In an unstratified monitoring design, all fishing operations are treated as the same. Monitoring effort is therefore distributed randomly across vessels or operations, without prior consideration of potential heterogeneity in the fleet. In contrast, a stratified monitoring design explicitly acknowledges such heterogeneity. The fleet is divided into segments based on characteristics known or expected to influence bycatch, and monitoring effort is allocated in a way that accounts for this variation.

To test the effect of monitoring effort level on the precision and accuracy of total bycatch estimates under different fishing and monitoring conditions, we used 3 different scenarios:1) homogenous fishing fleet, 2) heterogenous fishing fleet under unstratified monitoring design and 3) heterogenous fishing fleet under stratified monitoring design.

Simulations

Parameters used in the simulations included different bycatch probabilities (ranging from 0.0001 to 0.2), different proportions of fishing effort monitored (ranging from 5% to 30 %), mean number of fishing events per vessel per day (1, 2 or 5) and the effort level (operational or days-at-sea). In the second simulation, we changed the fishing fleet characteristic from homogenous fleet to a heterogenous one. This was done by introducing different bycatch probabilities to the two segments of the fishing fleet and by varying the proportion of the segments within the fleet (25 v 75 %, 50 v 50 % and 75 v 25 %).

The effect of the bycatch estimation effort level (days-at-sea or operational) on the precision (CV) and accuracy (calculated as relative bias (TB_estimated-TB_real)/TB_real) of bycatch estimates, were investigated using generalised linear models (GLM). Model selection was performed using a stepwise backwards model selection.

4.2. Results

Homogenous fishing fleet

The output of the simulation included estimated total bycatches at each level of the above parameters (bycatch probability, proportion of fishing effort and mean number of fishing events per vessel per day) and a coefficient of variation (CV) around these estimates. In addition, the monitoring effort as total number of operations monitored (for both options "operational" and "DaS") and as total number of days at sea monitored (if option "DaS") was calculated.

Precision in the total bycatch estimates

Under the homogenous fishing fleet scenario, the bycatch estimate precision was strongly influenced by pbycatch, NmeanFishingEvent, and pmonitor, with significant two- and three-way interactions among these factors (table 5). Although the level of monitoring effort (options) had no main effect, it appeared in significant interactions with pbycatch and in the four-way pbycatch:NmeanFishingEvent:pmonitor:options term.

Table 5. Analysis of Deviance table for the final model explaining bycatch estimate precision. Generalised linear model assuming a Gamma distribution of residuals.

term	χ^2	Df	p-value
pbycatch	474687	1	<0.00001
NmeanFishingEvent	152	1	<0.00001
pmonitor	36731	1	<0.00001
options	0.01	1	0.92
pbycatch:NmeanFishingEvent	8400	1	<0.00001
pbycatch:pmonitor	50256	1	<0.00001
NmeanFishingEvent:pmonitor	4.46	1	0.03
pbycatch:options	74	1	<0.00001
pbycatch:NmeanFishingEvent:pmonitor	2365	1	<0.00001
pbycatch:NmeanFishingEvent:pmonitor:options	45	1	<0.00001

The coefficient of variation (CV) of total bycatch estimates consistently decreases as the proportion of fishing operations monitored increases (x-axis), regardless of whether effort is measured in days-at-sea (red) or at the operational level (blue) (Figure 9). Differences between the two monitoring effort calculation methods are minimal, with nearly overlapping trends across all scenarios. Lower bycatch probabilities (top rows) and fewer fishing events per vessel per day (left columns) are associated with larger CVs, indicating greater uncertainty in bycatch estimates under these conditions. Conversely, at higher fishing frequencies and higher bycatch probabilities, the CVs are reduced more quickly with monitoring, suggesting more precise estimates with even modest levels of coverage. Overall, operational-level and days-at-sea monitoring provide nearly equivalent precision in estimating total bycatch.

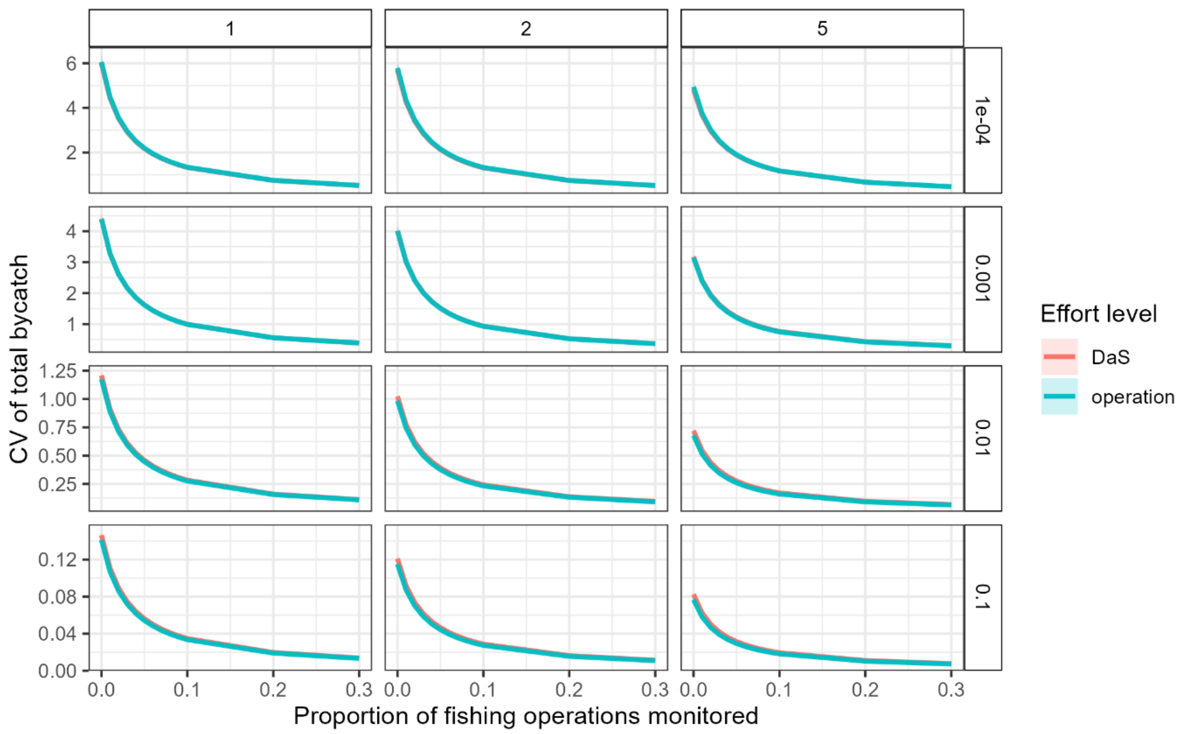


Figure 9. Predicted TB estimate CVs under the two different monitoring effort levels, days-at-sea (red) and operational level (blue) (1000 monitoring replicates per fishing year and 100 year replicates). Predictions from the best generalised linear model associating CV to the interaction between proportion of fishing operations monitored, bycatch probability (rows) and mean number of fishing events per vessel per day (columns) explanatory variables assuming a Gamma residuals distribution. Notice the logarithmic scale on the y-axis.

Accuracy in the total bycatch estimation

The analysis of deviance for the final model indicated that bycatch estimate accuracy was significantly affected by several predictors. Strong main effects were observed for pbycatch, NmeanFishingEvent, and $p_{monitor}$, with multiple significant two- and higher-order interactions contributing to model fit. As previously, level of monitoring effort calculation methods (options) did not show a significant main effect but appeared in one higher-order interaction as a significant factor which is why it was kept in the model.

Table 6. Analysis of Deviance table for the final model explaining bycatch estimate accuracy. Generalised linear model assuming a Gamma distribution of residuals.

term	χ^2	Df	p-value
pbycatch	166933	1	<0.00001
NmeanFishingEvent	297	1	<0.00001
pmonitor	10782	1	<0.00001
options	0.76	1	0.3826
pbycatch:NmeanFishingEvent	2954	1	<0.00001
pbycatch:pmonitor	17643	1	<0.00001
NmeanFishingEvent:pmonitor	56	1	<0.00001
pbycatch:options	26	1	<0.00001
NmeanFishingEvent:options	4.3	1	0.03
pbycatch:NmeanFishingEvent:pmonitor	758	1	<0.00001

The accuracy of total bycatch estimates, measured as bias, improves steadily as the proportion of fishing operations or days-at-sea monitored increases (x-axis, Figure 10). Both monitoring effort calculation methods—days-at-sea (red) and operational level (blue)—produce nearly identical trends, with very minor differences across all conditions. Accuracy is lowest (i.e., bias is highest) at low monitoring coverage (x-axis), particularly when bycatch probability is high. Overall, these results suggest that both monitoring effort calculation methods provide comparable and unbiased estimates of total bycatch, with accuracy dependent on monitoring coverage, as found for precision of total bycatch rate estimation.

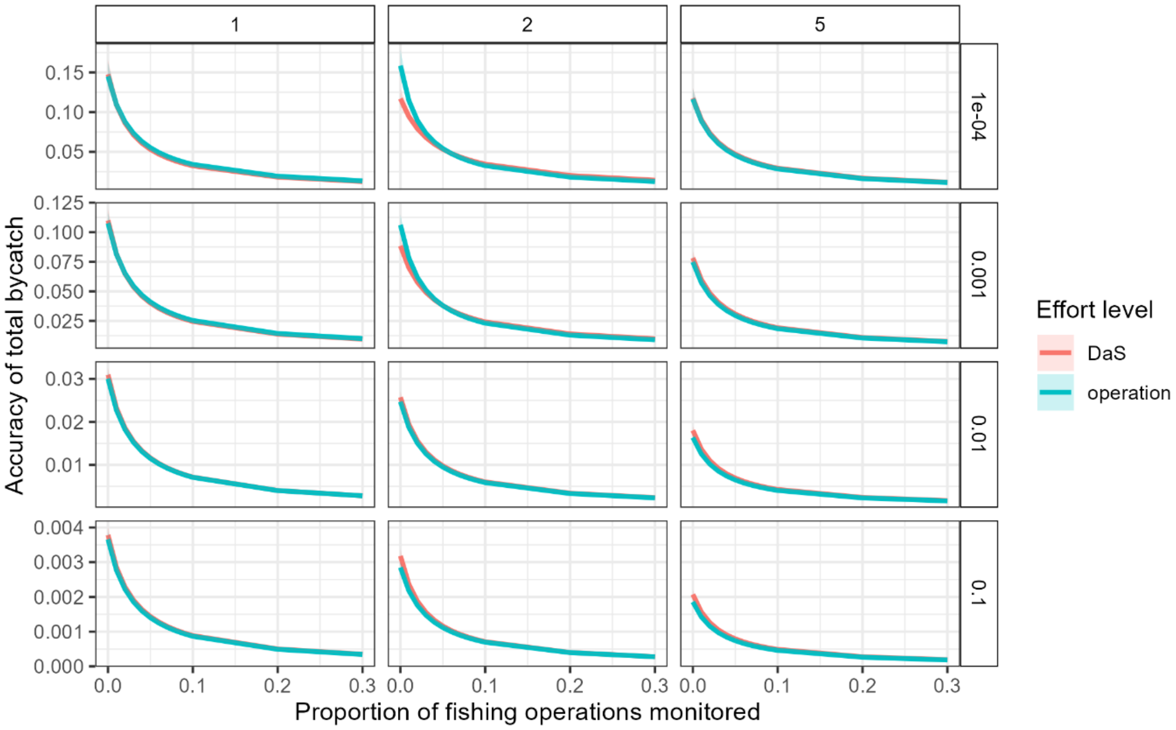


Figure 10. Predicted accuracy (measured as bias) in the TB estimates under the two different monitoring effort levels, days-at-sea (red) and operational level (blue) (1000 monitoring replicates per fishing year and 100 year replicates). Predictions from the best generalised linear model associating accuracy to the interaction between proportion of fishing operations monitored, bycatch probability (rows) and mean number of fishing events per vessel per day (columns) explanatory variables assuming a Gamma residuals distribution.

Heterogenous fishing fleet

Precision in the total bycatch estimates under unstratified monitoring

All main effects included in the model, apart from monitoring effort methods (options), were highly significant (χ^2 = 512396–137483, df = 1, $p < 0.00001$), indicating strong individual contributions to the model (Table 7). In contrast, monitoring effort calculation method (options) was not significant (χ^2 = 0.22, df = 1, p = 0.63) but was still kept in the model due to its significance in interactions (Table 7).

Table 7. Analysis of Deviance table for the final model explaining bycatch estimate precision under unstratified monitoring. Generalised linear model assuming a Gamma distribution of residuals.

term	χ^2	Df	p-value
pbycatch	512396	1	<0.00001
dbycatch	4523	1	<0.00001
pmetier1	2877	1	<0.00001
NmeanFishingEvent	11902	1	<0.00001
pmonitor	137483	1	<0.00001
options	0.22	1	0.63
pbycatch:dbycatch	3966	1	<0.00001
pbycatch:NmeanFishingEvent	6370	1	<0.00001
dbycatch:NmeanFishingEvent	399	1	<0.00001
pmetier1:NmeanFishingEvent	366	1	<0.00001
pbycatch:pmonitor	66955	1	<0.00001
dbycatch:pmonitor	1564	1	<0.00001
NmeanFishingEvent:pmonitor	3927	1	<0.00001
pbycatch:dbycatch:pmetier1	1241	1	<0.00001
pbycatch:dbycatch:NmeanFishingEvent	217	1	<0.00001
dbycatch:pmetier1:NmeanFishingEvent	217	1	<0.00001
pbycatch:dbycatch:pmonitor	1442	1	<0.00001
pbycatch:pmetier1:pmonitor	868	1	<0.00001
dbycatch:pmetier1:pmonitor	862	1	<0.00001
pbycatch:NmeanFishingEvent:pmonitor	2206	1	<0.00001
dbycatch:NmeanFishingEvent:pmonitor	208	1	<0.00001
pbycatch:NmeanFishingEvent:options	158	1	<0.00001
pbycatch:dbycatch:pmetier1:NmeanFishingEvent	121	1	<0.00001
pbycatch:dbycatch:pmetier1:pmonitor	431	1	<0.00001
pbycatch:dbycatch:NmeanFishingEvent:pmonitor	78	1	<0.00001
dbycatch:pmetier1:NmeanFishingEvent:pmonitor	170	1	<0.00001
pbycatch:NmeanFishingEvent:pmonitor:options	54	1	<0.00001
pbycatch:dbycatch:pmetier1:NmeanFishingEvent:pmonitor	50	1	<0.00001

The coefficient of variation (CV) of total bycatch estimates decreases steadily as the proportion of monitored days-at-sea or operations increases (x-axis), across all bycatch probabilities (Figure 11). Differences between the two monitoring effort calculation methods (days-at-sea in red, operations in blue) are minimal, with nearly overlapping results in all scenarios. Higher bycatch probabilities (lower panels) result in lower CVs overall, indicating greater precision in bycatch estimates under these conditions with operational level monitoring effort calculation yielding in slightly more precise estimates.

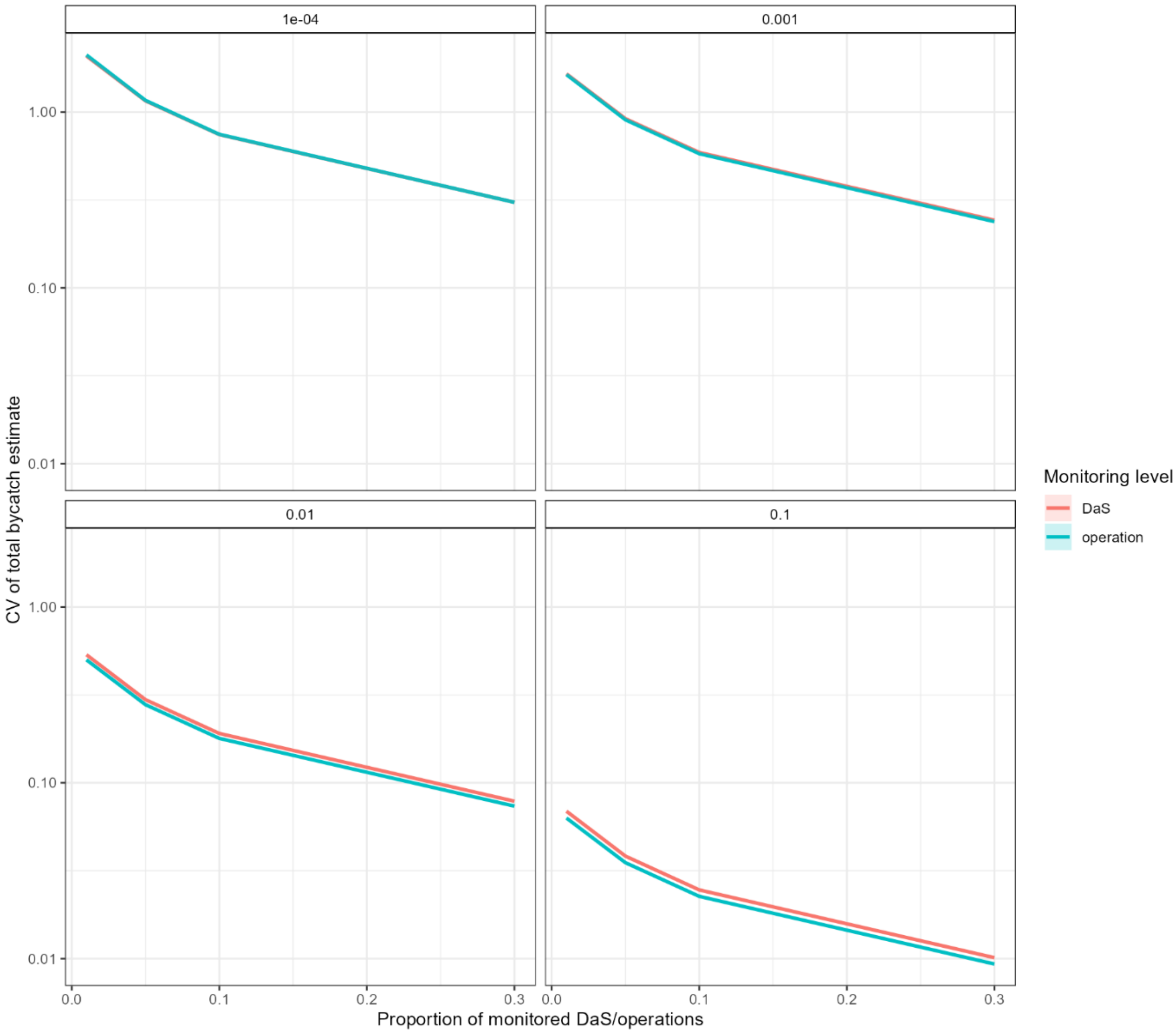


Figure 11. Predicted TB estimate CVs under the two different monitoring effort calculation method, days-at-sea (red) and operational level (blue) (1000 monitoring replicates per fishing year and 100 year replicates) under unstratified monitoring. Predictions from the best generalised linear model associating CV to the interaction between proportion of days-at-sea or operations monitored (x-axis) and bycatch probability (panels) explanatory variables assuming a Gamma residuals distribution. The plot shows the model outputs for scenarios when the mean number of fishing events per boat per day was five. Note the logarithmic scale of y-axis.

Accuracy in the total bycatch estimation under unstratified monitoring

Again, all main effects included in the model were highly significant ($\chi^2 = 715–127459$, df = 1, $p < 0.00001$), indicating strong to modest individual contributions to the model (Table 8). Monitoring effort level (options) was not significant ($\chi^2 = 1.29$, df = 1, $p = 0.25$) but as with estimating precision, was still kept in the model due to its significance in some of the interactions (Table 8).

Table 8. Analysis of Deviance table for the final model explaining bycatch estimate accuracy (bias) under unstratified monitoring. Generalised linear model assuming a Gamma distribution of residuals.

terms	χ^2	Df	p-value
pbycatch	127459	1	<0.00001
dbycatch	950	1	<0.00001
pmetier1	715	1	<0.00001
NmeanFishingEvent	2717	1	<0.00001
pmonitor	33995	1	<0.00001
options	1.29	1	0.25
pbycatch:dbycatch	1004	1	<0.00001
pbycatch:pmetier1	705	1	<0.00001
dbycatch:pmetier1	171	1	<0.00001
pbycatch:NmeanFishingEvent	1696	1	<0.00001
dbycatch:NmeanFishingEvent	115	1	<0.00001
pmetier1:NmeanFishingEvent	91	1	<0.00001
pbycatch:pmonitor	16699	1	<0.00001
dbycatch:pmonitor	418	1	<0.00001
pmetier1:pmonitor	259	1	<0.00001
NmeanFishingEvent:pmonitor	903	1	<0.00001
pbycatch:dbycatch:pmetier1	163	1	<0.00001
pbycatch:pmetier1:NmeanFishingEvent	42	1	<0.00001
dbycatch:pmetier1:NmeanFishingEvent	42	1	<0.00001
pbycatch:dbycatch:pmonitor	390	1	<0.00001
dbycatch:pmetier1:pmonitor	82	1	<0.00001
pbycatch:NmeanFishingEvent:pmonitor	619	1	<0.00001
dbycatch:NmeanFishingEvent:pmonitor	41	1	<0.00001
pmetier1:NmeanFishingEvent:pmonitor	36	1	<0.00001
pbycatch:pmonitor:options	20	1	<0.00001
pbycatch:dbycatch:pmetier1:pmonitor	324	1	<0.00001
pbycatch:dbycatch:NmeanFishingEvent:pmonitor	47	1	<0.00001

The bias in total bycatch estimates decreases steadily as the proportion of monitored days-at-sea or operations increases, across all bycatch probabilities (Figure 12). Differences between the two monitoring effort calculation methods (days-at-sea in red, operations in blue) are negligible, with almost identical model outputs in all scenarios. Higher bycatch probabilities (lower panels) result in lower bias in estimates but the differences in actual numbers are very small (Figure 12).

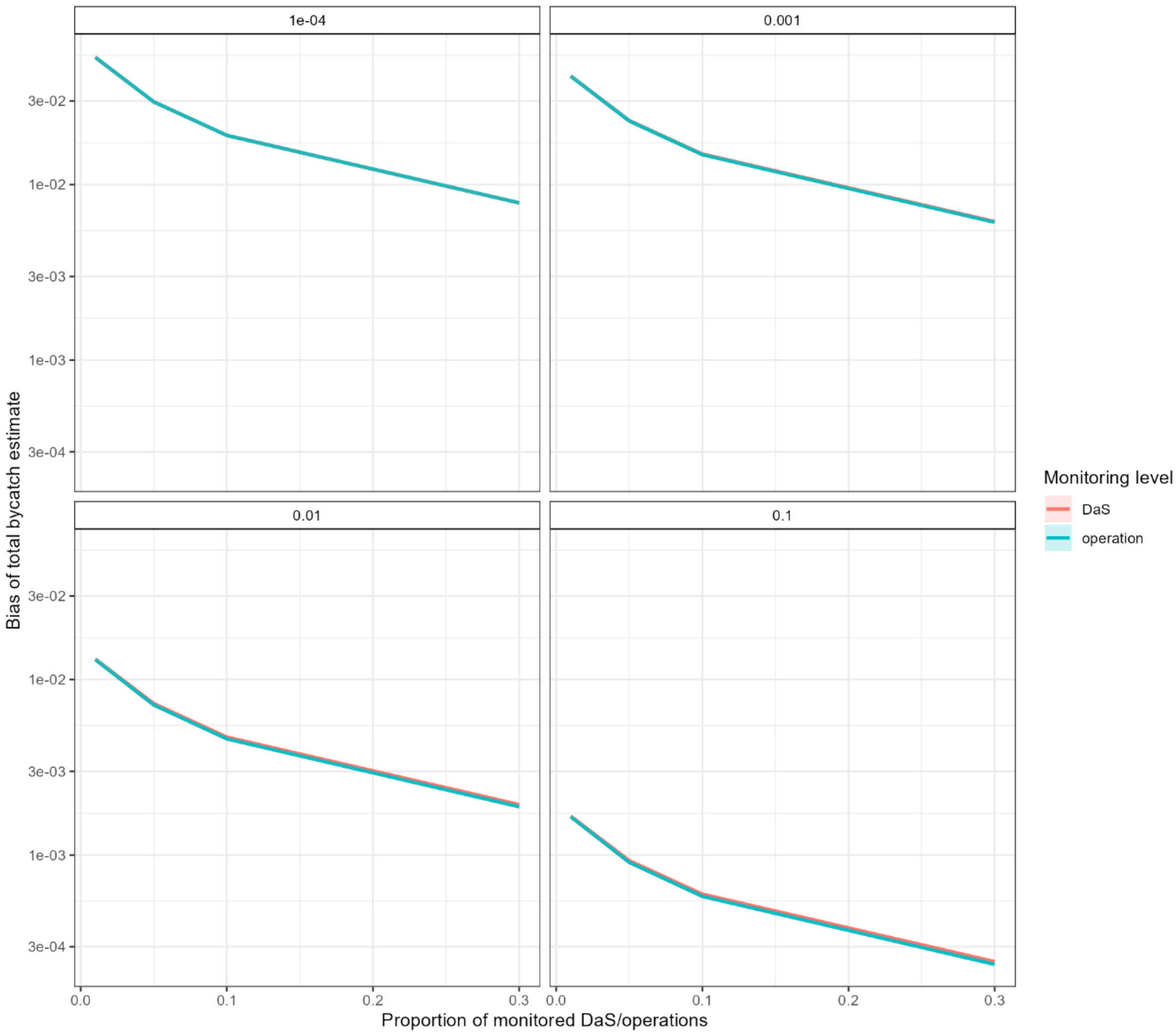


Figure 12. Predicted accuracy (measured as bias) in the TB estimates under the two different monitoring effort levels, days-at-sea (red) and operational level (blue) (1000 monitoring replicates per fishing year and 100 year replicates) under unstratified monitoring. Predictions from the best generalised linear model associating bias to the interaction between proportion of days-at-sea or operations monitored and bycatch probability (panels) explanatory variables assuming a Gamma residuals distribution. The plot shows the model outputs for scenarios when the mean number of fishing events per boat per day was five. Notice the logarithmic scale of y-axis.

Precision in the total bycatch estimates under stratified monitoring

As under the unstratified monitoring design, the main effects included in the model, apart from monitoring effort calculation method (options), were highly significant and explaining much of the variation in the data ($\chi^2 = 2952–511348$, df = 1, $p < 0.00001$) (Table 9). Also, as previously, monitoring effort calculation method (options) was not significant ($\chi^2 = 0.74$, df = 1, $p = 0.38$) but was still kept in the model due to its significance in interactions (Table 9).

Table 9. Analysis of Deviance table for the final model explaining bycatch estimate precision under stratified monitoring. Generalised linear model assuming a Gamma distribution of residuals.

Term	χ^2	Df	p-value
Pbycatch	511348	1	<0.00001
Dbycatch	4465	1	<0.00001
pmetier1	2952	1	<0.00001
NmeanFishingEvent	11818	1	<0.00001
Pmonitor	136279	1	<0.00001
Options	0.74	1	0.38
pbycatch:dbycatch	3821	1	<0.00001
pbycatch:pmetier1	2801	1	<0.00001
dbycatch:pmetier1	718	1	<0.00001
pbycatch:NmeanFishingEvent	6368	1	<0.00001
dbycatch:NmeanFishingEvent	410	1	<0.00001
pmetier1:NmeanFishingEvent	348	1	<0.00001
pbycatch:pmonitor	66879	1	<0.00001
dbycatch:pmonitor	1536	1	<0.00001
pmetier1:pmonitor	1102	1	<0.00001
NmeanFishingEvent:pmonitor	3889	1	<0.00001
pbycatch:dbycatch:pmetier1	724	1	<0.00001
pbycatch:pmetier1:NmeanFishingEvent	115	1	<0.00001
dbycatch:pmetier1:NmeanFishingEvent	109	1	<0.00001
pbycatch:dbycatch:pmonitor	1613	1	<0.00001
dbycatch:pmetier1:pmonitor	283	1	<0.00001
pbycatch:NmeanFishingEvent:pmonitor	2249	1	<0.00001
dbycatch:NmeanFishingEvent:pmonitor	144	1	<0.00001
pmetier1:NmeanFishingEvent:pmonitor	119	1	<0.00001
pbycatch:pmonitor:options	128	1	<0.00001
pbycatch:dbycatch:pmetier1:pmonitor	1298	1	<0.00001
pbycatch:dbycatch:NmeanFishingEvent:pmonitor	214	1	<0.00001
dbycatch:pmetier1:NmeanFishingEvent:pmonitor	41	1	<0.00001
pbycatch:NmeanFishingEvent:pmonitor:options	48	1	<0.00001
pbycatch:dbycatch:pmetier1:NmeanFishingEvent:pmonitor	55	1	<0.00001

The coefficient of variation (CV) of total bycatch estimates decreases steadily as the proportion of monitored days-at-sea or operations increases, across all bycatch probabilities (Figure 13). However, compared to the unstratified monitoring design, differences between the two monitoring effort levels (days-at-sea in red, operations in blue) were slightly more visual. Higher bycatch probabilities (lower panels) resulted in higher precision and during these conditions, the operational level monitoring effort was producing more precise estimates.

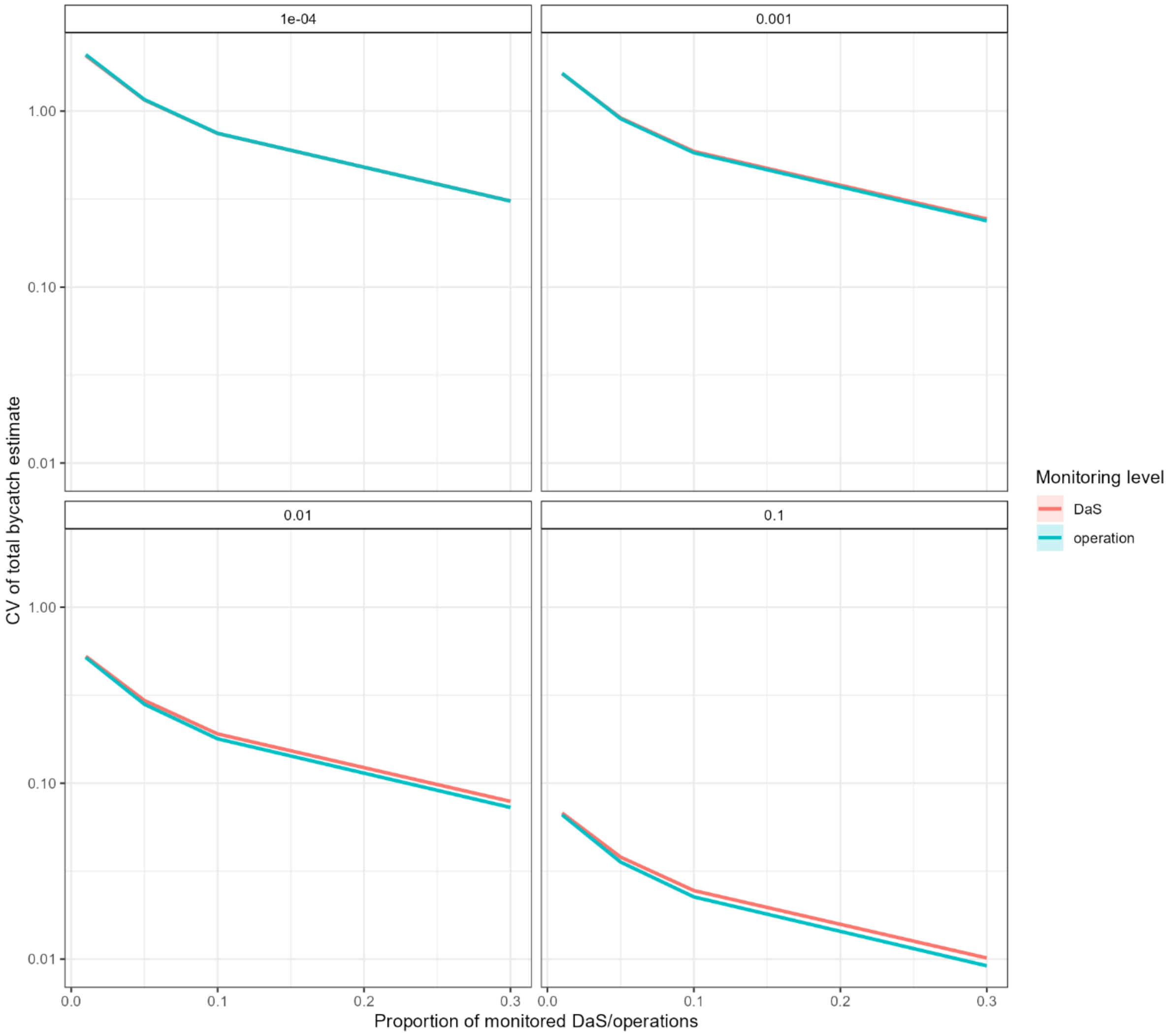


Figure 13. Predicted TB estimate CVs under the two different monitoring effort calculation method, days-at-sea (red) and operational level (blue) (1000 monitoring replicates per fishing year and 100 year replicates) under stratified monitoring. Predictions from the best generalised linear model associating CV to the interaction between proportion of days-at-sea or operations monitored (x-axis) and bycatch probability (panels) explanatory variables assuming a Gamma residuals distribution. The plot shows the model outputs for scenarios when the mean number of fishing events per boat per day was five. Notice the logarithmic scale of y-axis.

Accuracy in the total bycatch estimation under stratified monitoring

Again, all main effects, apart from monitoring effort calculation method (options), included in the model were highly significant (χ^2 = 767–127397, df = 1, p < 0.00001), indicating strong individual contributions to the model (Table 8). Monitoring effort level (options) was not significant (χ^2 = 0.008, df = 1, p = 0.93) but as with estimating precision, was still kept in the model due to its significance in some of the interactions (Table 10).

Table 10. Analysis of Deviance table for the final model explaining bycatch estimate accuracy (bias) under stratified monitoring. Generalised linear model assuming a Gamma distribution of residuals.

term	χ^2	**Df**	**p-value**
pbycatch	127397	1	<0.00001
dbycatch	1154	1	<0.00001
pmetier1	767	1	<0.00001
NmeanFishingEvent	3022	1	<0.00001
pmonitor	34011	1	<0.00001
options	0.008	1	0.93
pbycatch:dbycatch	912	1	<0.00001
pbycatch:pmetier1	651	1	<0.00001
dbycatch:pmetier1	175	1	<0.00001
pbycatch:NmeanFishingEvent	1609	1	<0.00001
dbycatch:NmeanFishingEvent	133	1	<0.00001
pmetier1:NmeanFishingEvent	77	1	<0.00001
pbycatch:pmonitor	16884	1	<0.00001
dbycatch:pmonitor	376	1	<0.00001
NmeanFishingEvent:pmonitor	1040	1	<0.00001
pmonitor:options	0.49	1	<0.00001
pbycatch:dbycatch:pmetier1	185	1	<0.00001
pbycatch:dbycatch:NmeanFishingEvent	55	1	<0.00001
pbycatch:dbycatch:pmonitor	427	1	<0.00001
dbycatch:pmetier1:pmonitor	309	1	<0.00001
pbycatch:NmeanFishingEvent:pmonitor	655	1	<0.00001
dbycatch:NmeanFishingEvent:pmonitor	29	1	<0.00001
pbycatch:dbycatch:pmetier1:NmeanFishingEvent	50	1	<0.00001
pbycatch:dbycatch:pmetier1:pmonitor	377	1	<0.00001
pbycatch:dbycatch:NmeanFishingEvent:pmonitor	33	1	<0.00001
dbycatch:pmetier1:NmeanFishingEvent:pmonitor	49	1	<0.00001
pbycatch:NmeanFishingEvent:pmonitor:options	71	1	<0.00001

The bias in total bycatch estimates, verified the tendency of operation level monitoring effort calculation method (blue) providing slightly more accurate total biomass estimates than days-at-sea method under stratified monitoring design (Figure 14). As with precision, the effect was present only at higher bycatch probabilities (lower panels) bearing in mind that the differences in actual numbers are very small.

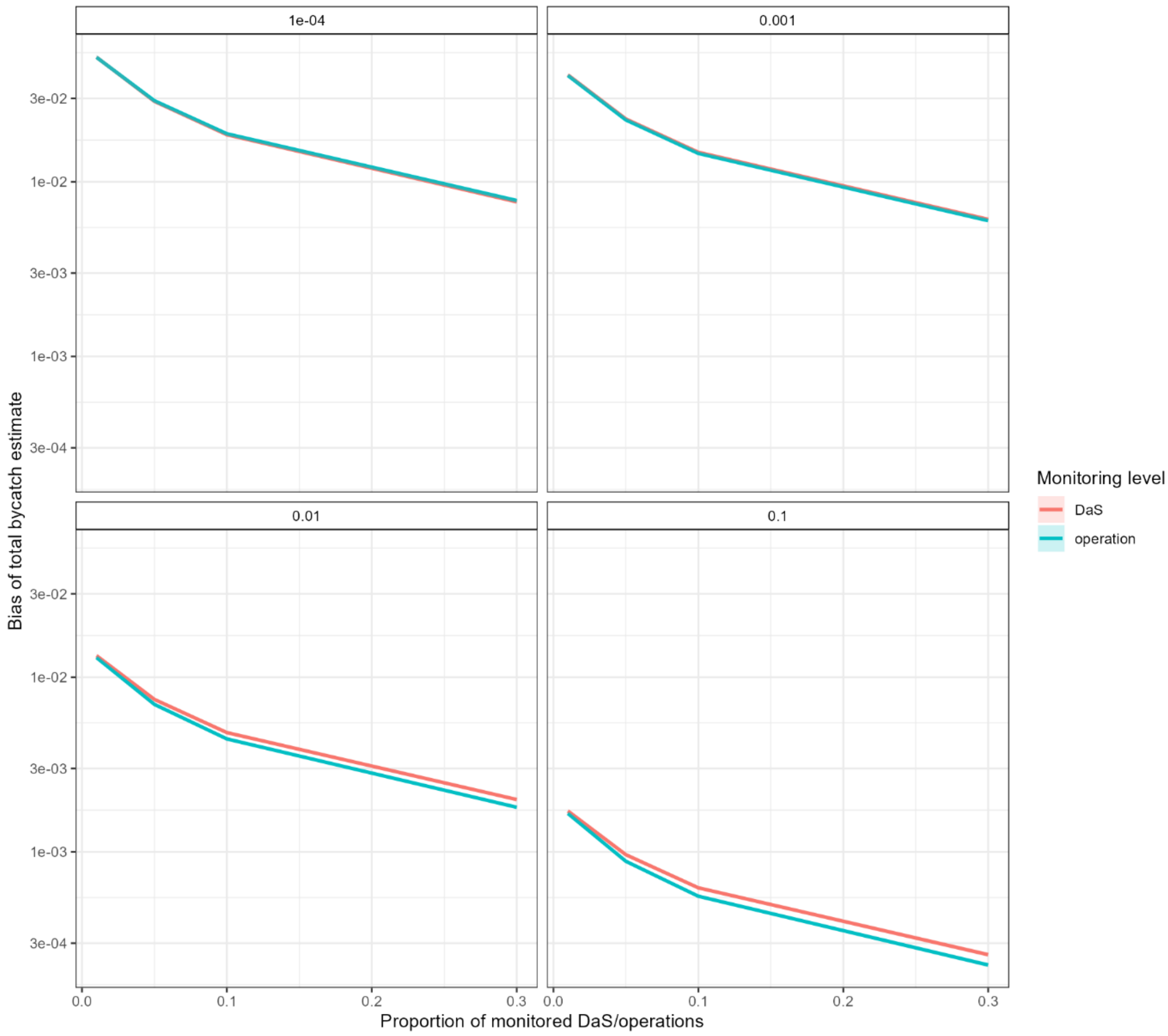


Figure 14. Predicted accuracy (measured as bias) in the TB estimates under the two different monitoring effort calculation method, days-at-sea (red) and operational level (blue) (1000 monitoring replicates per fishing year and 100 year replicates) under stratified monitoring. Predictions from the best generalised linear model associating bias to the interaction between proportion of days-at-sea or operations monitored (x-axis) and bycatch probability (panels) explanatory variables assuming a Gamma residuals distribution. The plot shows the model outputs for scenarios when the mean number of fishing events per boat per day was five. Notice the logarithmic scale of y-axis.

4.3. Discussion

The simulations show that both precision and accuracy of total bycatch (TB) estimates were primarily driven by the proportion of fishing effort monitored, the underlying bycatch probability, and fishing intensity (number of events per vessel per day). In contrast, the topic of this study, the choice of fishing effort and monitoring effort reporting level—operational-level versus days-at-sea (DaS)—had only marginal influence. Across all scenarios, increases in monitoring coverage consistently reduced uncertainty (CV) and bias in TB estimates, regardless of whether effort was calculated at the operational or DaS level.

For homogeneous fleets, operational and DaS monitoring produced nearly identical results. Precision and accuracy improved steadily with increasing coverage, with no systematic advantage of one effort level over the other. This suggests that for relatively uniform fisheries, the more readily available DaS data may provide estimates comparable in reliability to operational data, particularly when moderate to high coverage is achieved.

For heterogeneous fleets, the monitoring design played a more central role. Under unstratified monitoring, estimates remained unbiased but imprecise at low coverage, and the differences between days-at-sea and operational metrics were minimal. In contrast, under stratified monitoring, which explicitly accounted for fleet heterogeneity, operational-level monitoring produced slightly more precise and accurate TB estimates than days-at-sea-based monitoring, particularly at higher bycatch probabilities. Even if these differences were statistically detectable, their magnitude was small, indicating that both approaches remain broadly comparable in practice.

It is important to note, however, that in these simulations vessels were assumed to be able to fish during all reported days-at-sea. In reality, days-at-sea reporting often includes periods of steaming to and from fishing grounds, during which no fishing is possible. This introduces a systematic bias in estimates based on days-at-sea, as the effective effort (and thus the mean bycatch per unit effort, BPUE) will be lower than assumed. One solution is to apply a correction factor, for instance to calculate the ratio of fished to non-fished days-at-sea within trips.

In summary, monitoring coverage wields a much stronger influence on total bycatch estimation reliability than the choice of bycatch monitoring effort level. For instance, for homogeneous fishing fleets, days-at-sea level data appear sufficient. Under stratified monitoring of heterogeneous fleets, operational-level data can offer some advantage in precision and accuracy of bycatch estimation. In some contexts, days-at-sea level monitoring, however, can offer a practical and cost-effective alternative, provided any systematic bias introduced by non-fishing days is corrected using straightforward adjustments.

5. Informing deepwater longline fishery monitoring for mainland Portugal

5.1. Introduction

In CIBBRiNA Case study 4, on the studied fisheries is the deepwater longline fishery for scabbardfish off mainland Portugal, and it offers an interesting challenge for the design of bycatch monitoring schemes. The fishery is characterised by a small number of vessels operating, year-round, in non-overlapping areas. For the purpose of simulations, we assumed that the fishery was composed of 15 vessels, an upper limit to the range of vessel numbers that can be considered operating annually. These areas are well delineated and consistently used by the same vessel. They operate similar fishing gears and the number of hooks do not greatly differ within vessels. Most often deploy one haul per day. This haul can be composed of 5 to 12 segments, typically 8, each containing around 1000 hooks.
The fishery bycatches several species of deepwater sharks at very different bycatch rates (n=19 species in the time series between 2005-2019, some of which are occasional and need species confirmation). The number of individuals caught for each species of concern appear to follow a simple unimodal process. Onboard observers may not be able to observe all segments, they observe all segments in about 60% of cases, when they cannot observe all segments, they typically observe 60% of segments and extrapolate the total bycatch for the haul based on the proportion of observed segments. The inability to sample all segments is related to the length of the gear and the consequent long hauling duration.
The probability that a haul contains at least one bycaught individual from those species vary across several orders of magnitude. To represent them we will here simulate bycatch of five representative species with a bycatch probability of 0.025, 0.05, 0.1,0.5,0.75. The challenge is that bycatch probability can be expressed per haul. For these species and this fishery, the bycatch process occurs by individuals getting caught on hooks, same as for target species. Theoretically it would mean that the bycatch effort unit is not necessarily the same for all species. Practically, the number of individuals bycaught per haul means that hooks are unlikely to be ‘unavailable’ (saturated) to bycatch sharks. So practically we can move to a larger unit of effort for which we can assume similar bycatching processes across species. We will use haul segments as bycatch simulation unit.

Fishing simulations

The fishing simulation will therefore depart from our description in section 2. We assume that each vessel can go on a fishing trip every day. We assume that for each trip a vessel can have one haul and the number of segments per haul will vary stochastically for all vessels every day.
Fishing is less intensive in reality. Each vessel can go on a fishing trip every day, however vessels tend to go out three times per week. This does not alter the relevance of the simulation outcomes as the primary purpose of this stage of the simulations is to align the number of hauls possible per fishing days, and here we ensure that there is one haul possible per day (rather than multiple hauls). This means that the magnitude of the total bycatch will be less in reality compared to the outcomes of the simulations, but our interest lies with precision and bias which are not affected by total bycatch magnitude (only bycatch rates).
This deployment schedule is informed by summary information available from previous monitoring programmes. The number of segments can vary from 5 to 12, 8 segments are 3 times more likely that 5,6,7, or 9 segment, 6 times more likely than 10 segments and 12 times more likely than 11 or 12 segments (Figure 15). For the purpose of the simulation, we assume that each haul, species can be caught with five differing haul bycatch probability (0.025, 0.05, 0.1,0.5, and 0.75). We have not previously demonstrated if bycatch probability is related to the number of segments in the haul, but practically it can be, if it fishes at different areas or

depths (higher probabilities at deeper depths, e.g. Veiga et al., 2013). We will therefore assume that these haul bycatch probabilities are those resulting from the mean number of segments in the haul. We therefore divide for each simulated fishing year the haul bycatch probabilities by the mean number of segments deployed per haul in the fishing year to retrieve a bycatch probability per segment for each species. For example, if the mean number of segments, s, per haul in the year is 7.5, then the segment bycatch probabilities are:

$$p_{segment} = 1 - (1 - p_{haul})^{\frac{1}{7.5}}$$

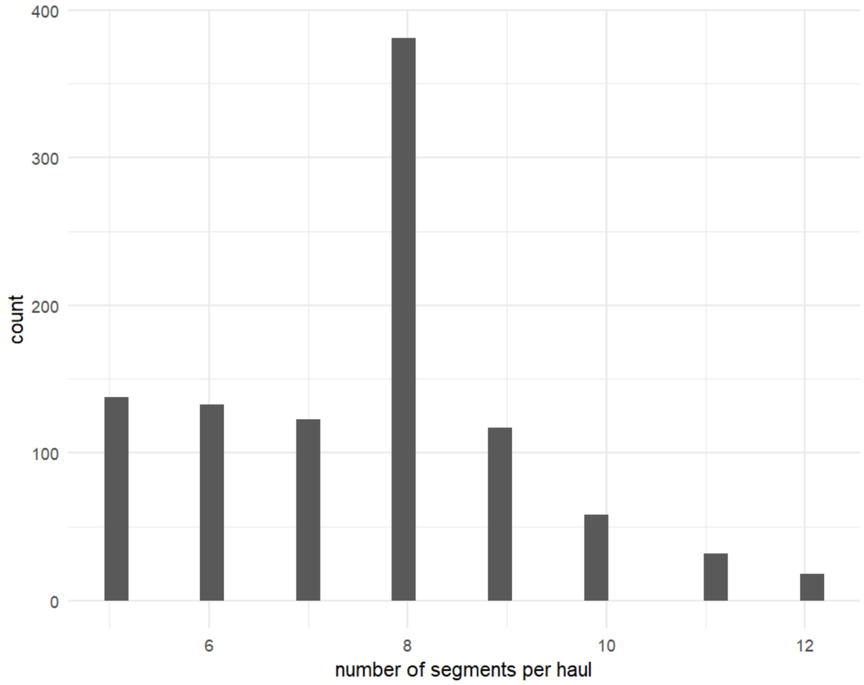


Figure 15. Frequency distribution the number of segments per haul for 1000 simulated hauls in the simulated longline fishery

The mean number of segments per haul, s, is a common to all vessels here.

Area-specific variance in bycatch probability

Vessels operate in mutually exclusive waters. It is therefore possible that bycatch rates vary between areas not because of fishing characteristics but for ecological reasons; the simplest being that the density of sharks differ between areas, as a function of depth and possibly other factors. The lack of overlap in fishing effort between vessels is an important characteristic of this fishery which need to be captured in simulation so it can be investigated for its potential to affect monitoring design. Here we simply add a random effect of vessel on the haul bycatch probabilities and varies it from a small effect (1% of p_{haul}) to a large effect (50% of p_{haul}).

Here is a pseudocode of how bycatch probability is defined for each segment:

$p_{bycatch}$ = [0.025, 0.05, 0.1,0.5, 0.75] one value per species
$vessel_{effect} \in (0,0.5)$
$p_{bycatch,vessel\_i}$=(1-Uniform(-$vessel_{effect}$,$vessel_{effect}$)) * $p_{bycatch}$
$p_{segment,vessel\_i}$ = 1-(1- $p_{bycatch,vessel\_i})^{1/s}$

where we assume that the mean number of bycaught individuals *per segment* is one when a bycatch occurs. This approximates observed and estimated number of individuals bycaught per haul.

These adaptations to the fishing simulation in section 2 allow us to then produce a fishing year resembling the studied deepwater longline fishery for scabbardfish off mainland Portugal in case study 4. During this fishing year for simplicity each vessel operates one haul per day every day, rather than 3 times per week as currently operated, and a haul is composed of different number of segments with a resulting bycatch rate which we can then estimate by monitoring some of the fishing.

Monitoring strategy

The monitoring question reconciles with those described in section 3: given that capacity is available to cover a small proportion of hauls, should we monitor a few vessels more or should we spread monitoring between vessels? We can therefore explore the effect of the fishing characteristics we highlight above on this consideration. We introduce here also the observer effect we described: in 40% of cases all segments of a monitored haul are observed, and in 60% of cases a proportion of those is observed with the mean proportion observed being 0.6 (random normal distribution mean= 0.6, sd=0.1, Figure 16).

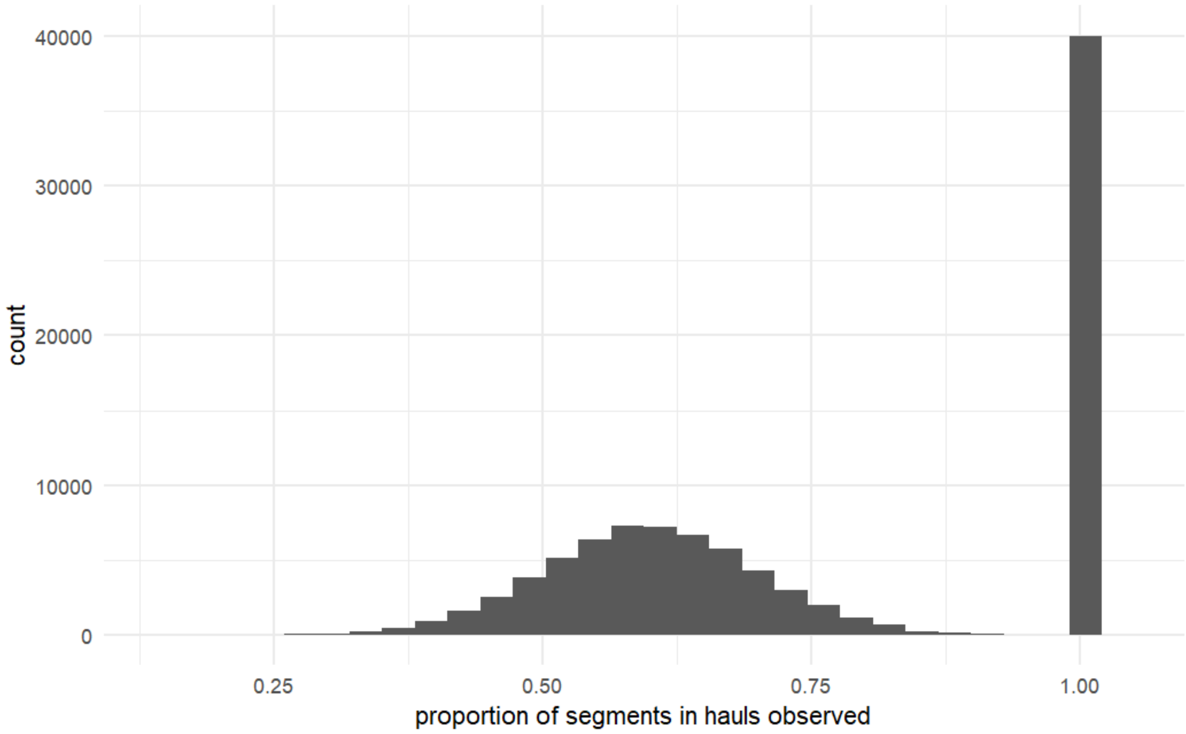


Figure 16. Frequency distribution of 100,000 simulated values for the proportion of segments in a haul that will be observed.

5.2. Results

Monitoring is spread across all vessels

Designs with low coverage can achieve sensible precision and accuracy when monitoring is spread between vessels.

Bias in bycatch estimate

There was no effect of “unknown” between-vessel variability in BPUE (Figures 17 and 18). The best model describing the bias in Total bycatch estimate assumed an interacting effect between $p_{monitor}$, $p_{bycatch}$ and $p_{segment}$ (AIC= -114742.6) compared to the full model where all components are interacting (AIC= -114727.7).

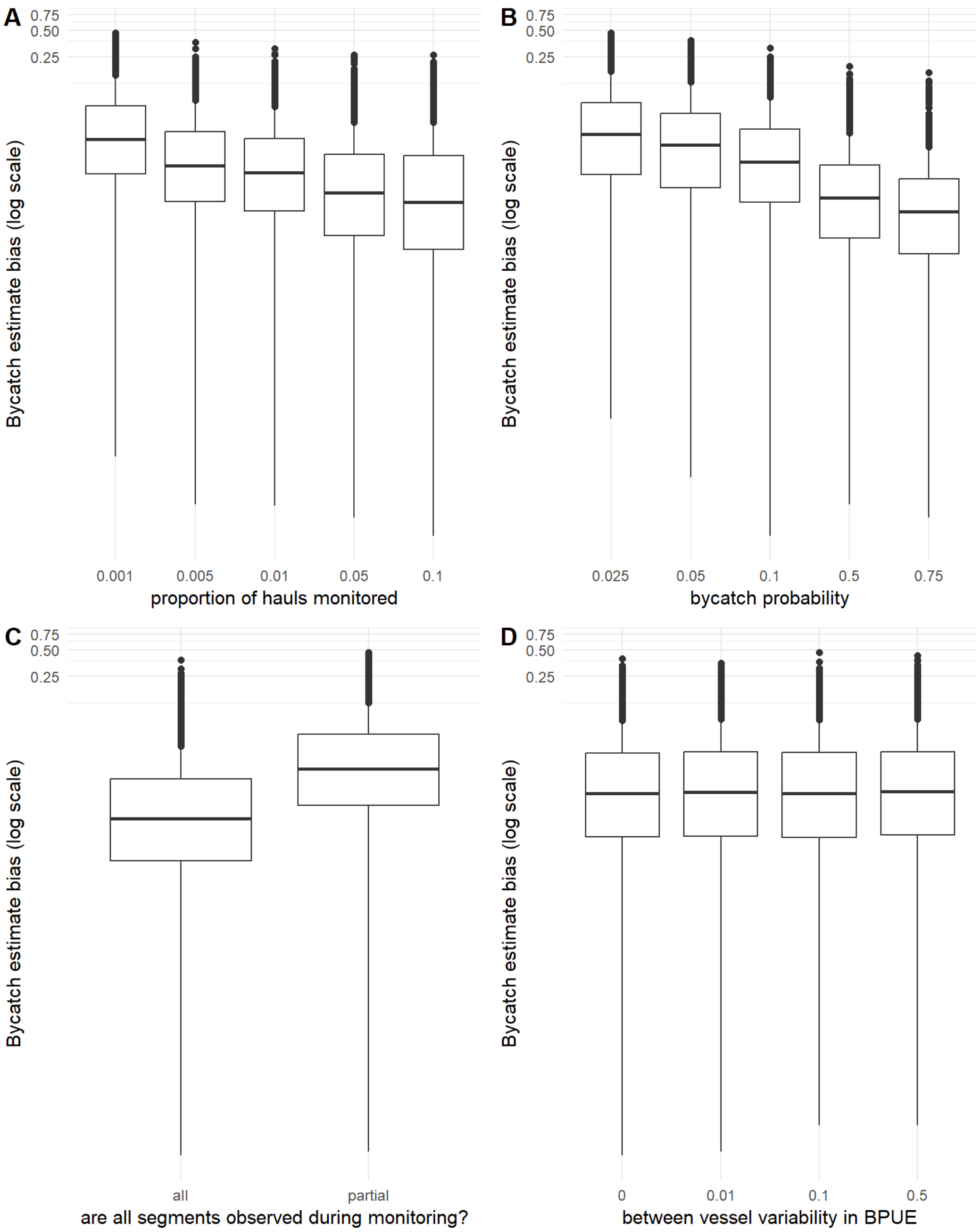


Figure 17. Summary statistics of the bias in total bycatch estimate observed in simulation run following the parameter settings for (A) pmonitoring, (B) pbycatch, (C) psegment, and (D) $vessel_{effect}$ described in the methods (n=20000 simulation sets)

The best model predicts that, as expected, the bycatch estimate bias decreases as the bycatch probability increases (line colours) and the proportion of hauls monitored increases (x-axis) (Figure 18). It shows that the current approach to extrapolate bycatch observations to the whole haul based on partial observations of segments introduces a non-trivial bias in the bycatch estimate, and this effect becomes more evident as the bycatch probability decreases (Figure 18).

Table 11. Analysis of Deviance table for the final model explaining bycatch estimate bias. Generalised linear model assuming a Gamma distribution of residuals.

terms	χ^2	df	p-value
pbycatch	4561	1	<0.00001
pmonitor	1251	1	<0.00001
psegment	3413	1	<0.00001
pbycatch:pmonitor	54.7	1	<0.00001
pbycatch:psegment	144.4	1	<0.00001
pmonitor:psegment	427.5	1	<0.00001
pbycatch:pmonitor:psegment	31.7	1	<0.00001

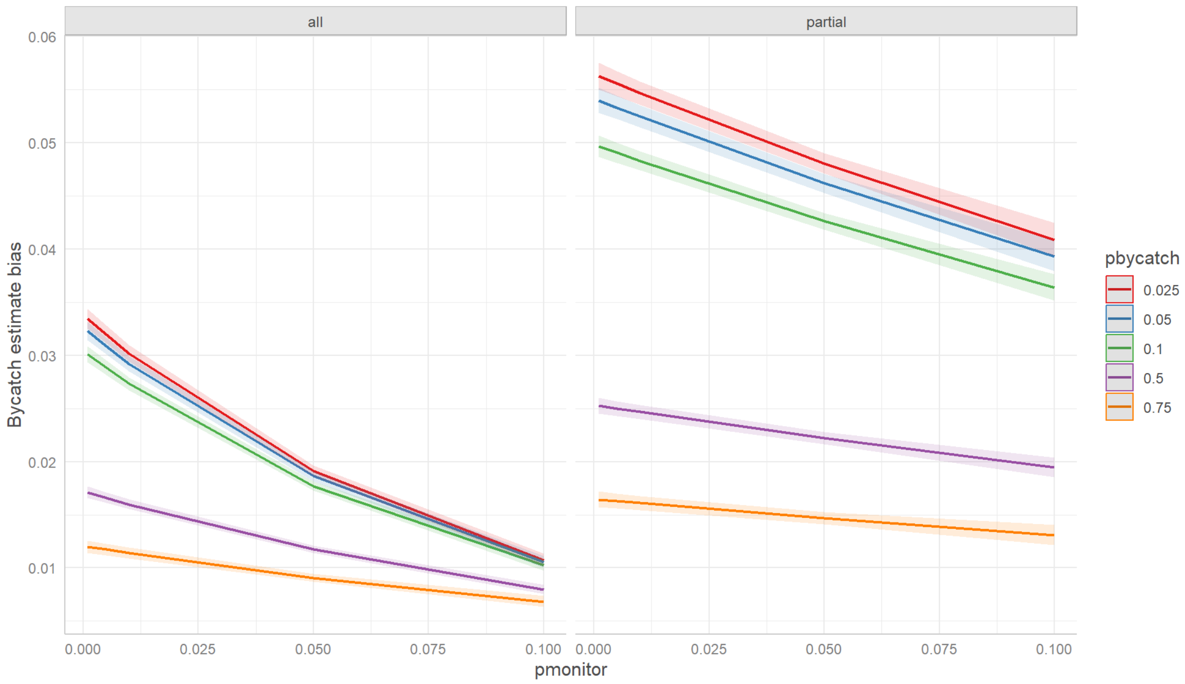


Figure 18. Predicted bycatch estimate bias based on the best model describing simulation sets (Table 11). Left panel all segments of monitored hauls are observed, right panel segments of monitored hauls are partially observed following the tuned observation prevalence (Figure 16).

Precision of bycatch estimate

Similar effects emerge for the precision of the bycatch estimate (Figure 20) in which between-vessel variability does not affect the precision of the estimate (CV).

There was no effect of "unknown" between-vessel variability in BPUE (Figures 19 and 20). The best model describing the CV of the total bycatch estimate assumed an interacting effect between $p_{monitor}$, $p_{bycatch}$ and $p_{segment}$.

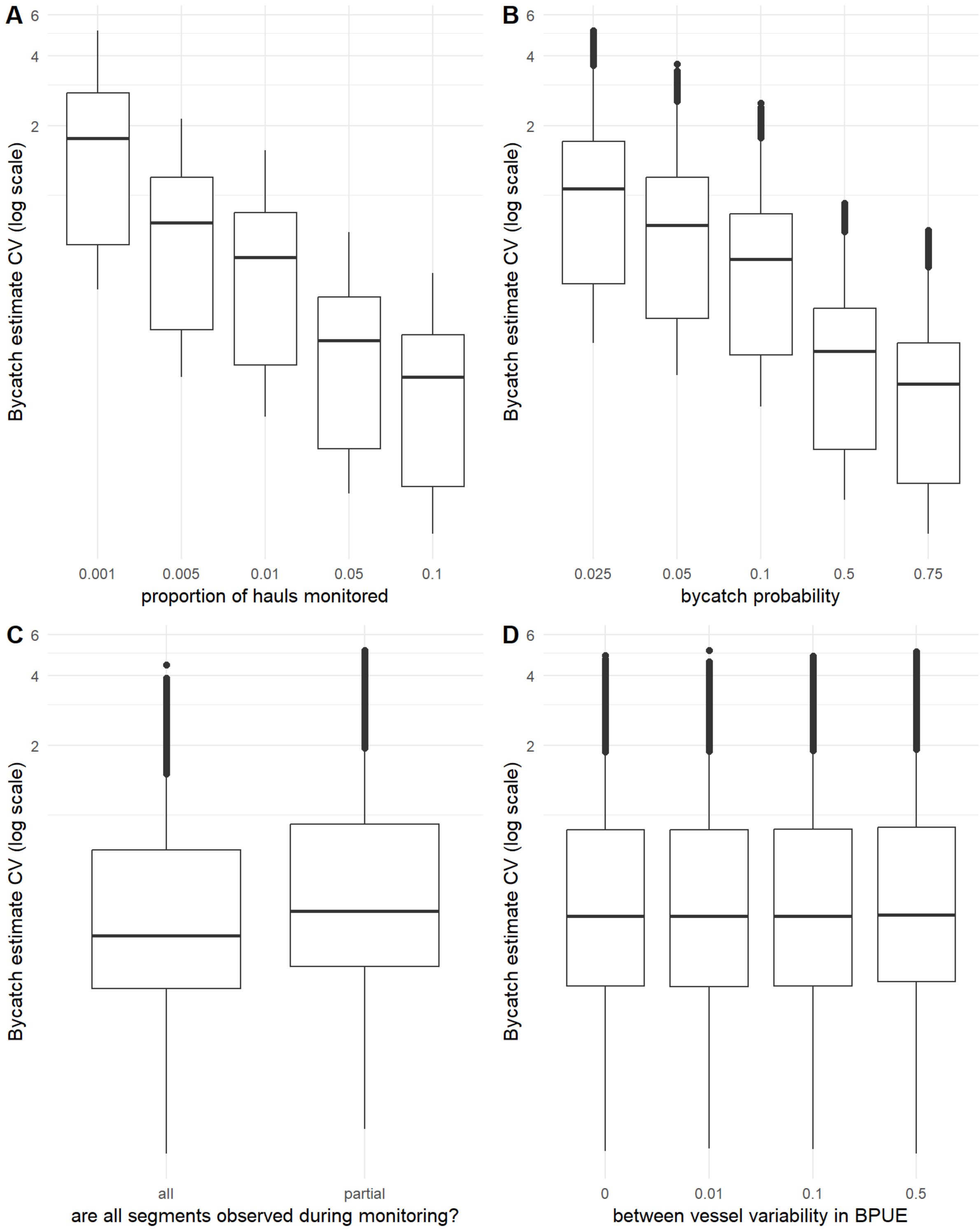


Figure 19. Summary statistics of the coefficient of variation of the total bycatch estimate observed in simulation run following the parameter settings for (A) $p_{monitoring}$, (B) $p_{bycatch}$, (C) $p_{segment}$, and (D) $vessel_{effect}$ described in the methods (n=20000 simulation sets)

The best model predicts that, as expected, the bycatch estimate CV decreases (precision increases) as the bycatch probability increases (line colours) and the proportion of hauls monitored increases (columns) (Figure 20). It shows that the proposed approach to extrapolate bycatch observations to the whole haul based on partial observations of segments introduces a non-trivial decrease in precision in the bycatch estimate, and this effect becomes more evident as the bycatch probability decreases (Figure 20). If whole monitored hauls are observed then the monitoring design can achieve a CV of 0.3 with 7.5% of hauls monitored for all bycatch probabilities (Figure 20), but it needs to raise to 9% to do so if monitored hauls are partially observed.

Table 12. Analysis of Deviance table for the final model explaining bycatch estimate CV. Generalised linear model assuming a Gamma distribution of residuals.

terms	χ^2	df	p-value
pbycatch	54701	1	<0.00001
pmonitor	69023	1	<0.00001
psegment	1946	1	<0.00001
pbycatch:pmonitor	24543	1	<0.00001
pbycatch:psegment	614.5	1	<0.00001
pmonitor:psegment	878.1	1	<0.00001
pbycatch:pmonitor:psegment	267.7	1	<0.00001

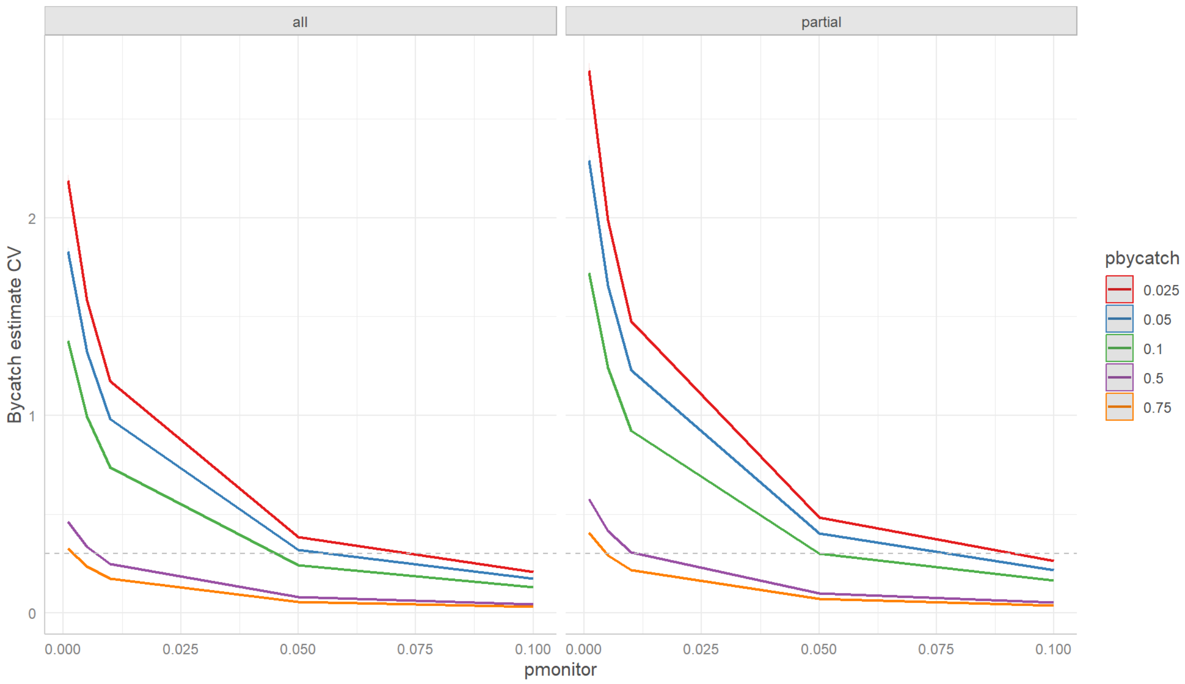


Figure 20. Predicted bycatch estimate precision (CV) based on the best model describing simulation sets (Table 12). Left panel all segments of monitored hauls are observed, right panel segments of monitored hauls are partially observed following the tuned observation prevalence (Figure 16).

Monitoring is concentrated in a few vessels

In this scenario we achieve the same number of hauls monitored as we did in the previous section, but instead of those monitored hauls being distributed across all vessels, they are concentrated in a few vessels. It is important to stress that here $p_{monitor}$ is applied to the total number of hauls realised in the fleet that year, so that the proportion of the hauls monitored for the vessels on which monitoring is concentrating is much higher. As an example, if we monitored 10% of hauls and the fleet carried out 360 hauls that year, it means we monitor 36 hauls. If we only sample vessel A and vessel B and each of them fished 18 hauls, it means we would have to sample 100% of the hauls those vessels did. This approach allows us to compare directly the precision and accuracy achieved when monitoring is concentrated in a few vessels rather than spread randomly between all vessels (results in the previous section). It therefore allows us to make inference about the usefulness of either monitoring designs.

We replicated the simulation set while concentrating monitoring in 2, 4, or 7 vessels (out of 15). The fewer vessels were monitored, the less accurate the bycatch estimate was (Figure 21) and the precision seems to be overall deteriorated compared to the previous monitoring design (Figures 17 & 19).

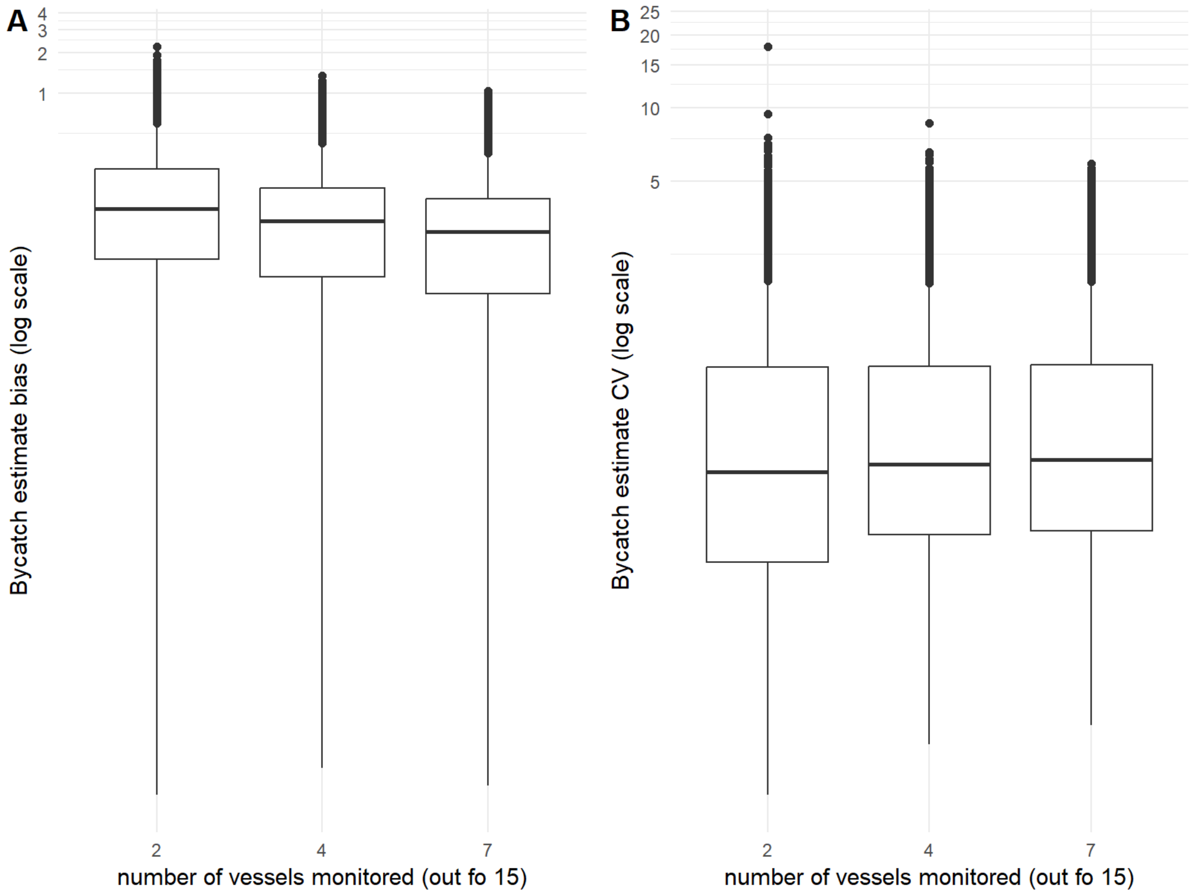


Figure 21. Summary statistics of the bias (A) and the coefficient of variation (B) of the total bycatch estimate observed in simulation run depending on the number of vessels monitored out of 15 (60000 simulation sets).

Bias in bycatch estimate

The best model retains all variables (AIC=-139293) yielding complex interactions (Table 13).

Table 13. Analysis of Deviance table for the final model explaining bycatch estimate bias. Generalised linear model assuming a Gamma distribution of residuals.

Terms	χ^2	df	p-value
pbycatch	1464	1	<0.0001
pmonitor	7.47	1	0.006
vesseleffectf	5641	3	<0.0001
sboat	3344	2	<0.0001
psegment	3501	1	<0.0001
pbycatch:pmonitor	1.76	1	0.18
pbycatch:vesseleffectf	1497	3	<0.0001
pmonitor:vesseleffectf	4.81	3	0.18
pbycatch:sboat	4.20	2	0.12
pmonitor:sboat	7.01	2	0.03
vesseleffectf:sboat	25.1	6	0.0003
pbycatch:psegment	626	1	<0.0001
pmonitor:psegment	9.21	1	0.002
vesseleffectf:psegment	2112	3	<0.0001
sboat:psegment	930	2	<0.0001
pbycatch:pmonitor:vesseleffectf	0.54	3	0.91
pbycatch:pmonitor:sboat	0.80	2	0.67
pbycatch:vesseleffectf:sboat	8.37	6	0.21
pmonitor:vesseleffectf:sboat	8.77	6	0.19
pbycatch:pmonitor:psegment	2.48	1	0.11
pbycatch:vesseleffectf:psegment	4739	3	<0.0001
pmonitor:vesseleffectf:psegment	4.50	3	0.21
pbycatch:sboat:psegment	133	2	<0.0001
pmonitor:sboat:psegment	16.1	2	0.0003
vesseleffectf:sboat:psegment	485	6	<0.0001
pbycatch:pmonitor:vesseleffectf:sboat	1.80	6	0.93
pbycatch:pmonitor:vesseleffectf:psegment	8.33	3	0.04
pbycatch:pmonitor:sboat:psegment	3.15	2	0.21
pbycatch:vesseleffectf:sboat:psegment	691	6	<0.0001
pmonitor:vesseleffectf:sboat:psegment	21.0	6	0.002
pbycatch:pmonitor:vesseleffectf:sboat:psegment	16.4	6	0.01

The variable $p_{segment}$ has the same effect as in the previous scenario and it does not vary with the number of vessels selected (Figure 22). Overall, this monitoring design yields less accurate estimates than the previous scenario (Figures 18). This further deteriorates if there is between-vessel heterogeneity (Figure 23, line colours) and increasing monitoring coverage does not significantly improve precision (Figure 23). Increasing monitoring coverage only increases accuracy when the number of vessels monitored increases (Figure 24 columns).

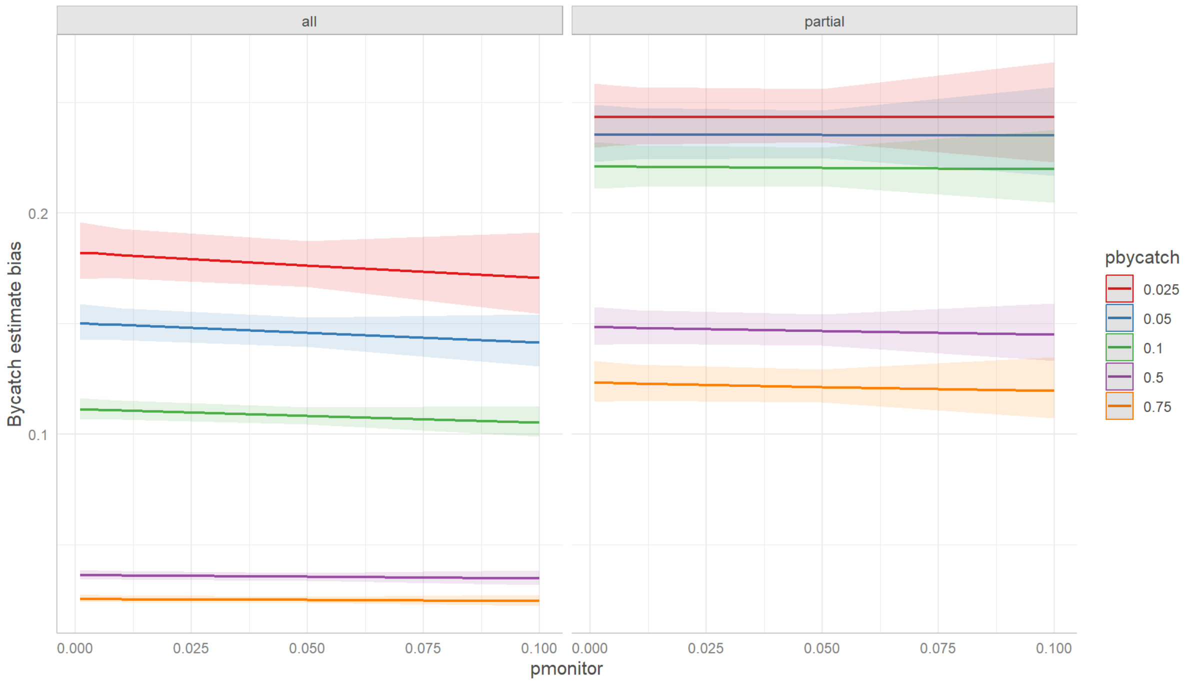


Figure 22. Predicted bycatch estimate accuracy (bias) based on the best model describing simulation sets (Table 13). Here we focus on the partial effect the number of segments observed (panels), the proportion of hauls fished that year which were monitored ($p_{monitor}$ in x-axis) and the bycatch probability ($p_{bycatch}$ in line colours).

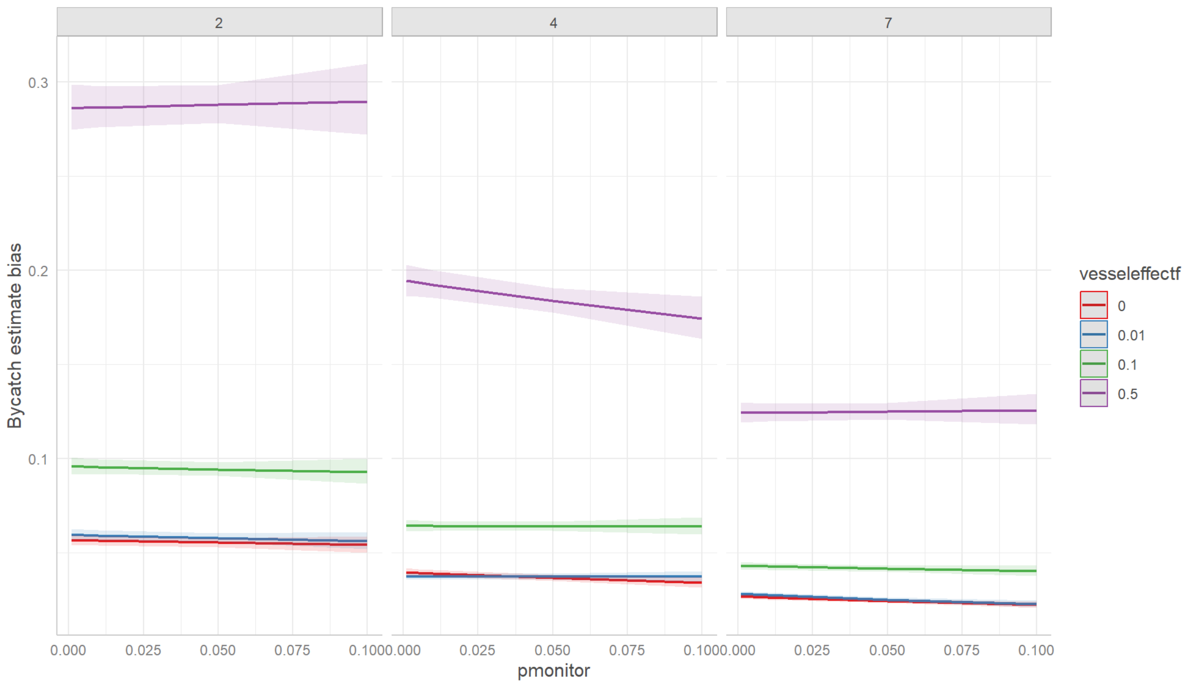


Figure 23. Predicted bycatch estimate accuracy (bias) based on the best model describing simulation sets (Table 13). Here we focus on the partial effect the number of vessels monitored (panels), the proportion of hauls fished that year which were monitored ($p_{monitor}$ in x-axis) and the between-vessel variability in bycatch ($vessel_{effect}$ in line colours).

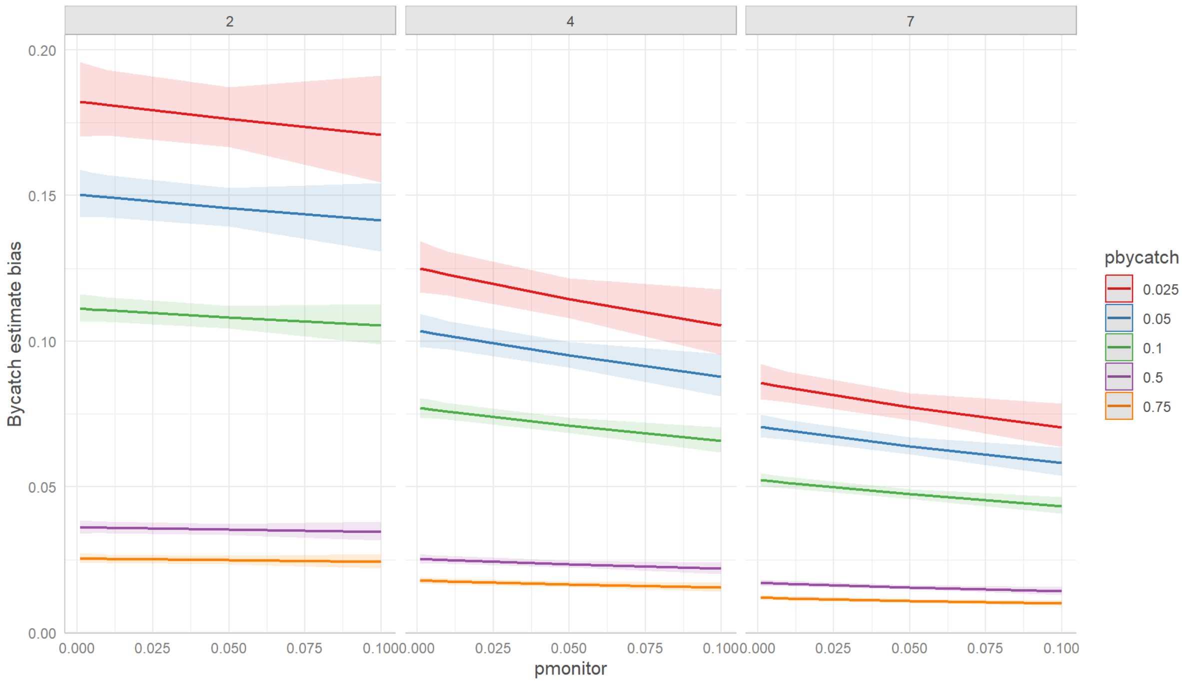


Figure 24. Predicted bycatch estimate accuracy (bias) based on the best model describing simulation sets (Table 13). Here we focus on the partial effect the number of vessels monitored (panels), the proportion of hauls fished that year which were monitored ($p_{monitor}$) and the bycatch probability ($p_{bycatch}$). Here the $vessel_{effect}$ is set to be 0.

Precision in bycatch estimate

The best model does not retain the five-way interaction, nor most four-way interactions (AIC= -125105, Table 14). It points to expected effect: precision increases (CV decreases) as $p_{bycatch}$ increases (x-axis) and $p_{monitor}$ increases (line colours) (Figure 27).

The between-vessel variability ($vessel_{effect}$) did not have much effect on precision (line colours Figure 25) and a potential target CV of 0.3 (EU Regulation 812/2004) can be achieved with low coverage (Figures 26, 27). It is important to recognise the difference in monitoring units compared to the previous analyses. Here the monitoring coverage corresponds to the coverage of the fleet. So, 1% of the monitoring coverage (**fleet coverage**) in a fleet that contains 15 vessels when monitoring is distributed between two vessels, and all vessels are fishing with the same intensity as it is the case here, corresponds to monitoring 7.5% **of the hauls** of the monitored vessels. We therefore retrieve similar monitoring coverage to those we estimated in the previous section. As before not being able to monitor all segments of the hauls ($p_{segment}$) deteriorates the bycatch estimate precision as well (Figure 26).

Table 14. Analysis of Deviance for the final model explaining bycatch estimate CV. Generalised linear model assuming a Gamma distribution of residuals.

Terms	χ^2	**df**	**p-value**
pbycatch	173887	1	<0.0001
pmonitor	265818	1	<0.0001
vesseleffectf	151.2	3	<0.0001
sboat	33.2	2	<0.0001
psegment	7024	1	<0.0001
pbycatch:pmonitor	90860	1	<0.0001
pbycatch:vesseleffectf	267.4	3	<0.0001
pmonitor:vesseleffectf	79.9	3	<0.0001
pbycatch:sboat	29.6	2	<0.0001
pmonitor:sboat	7629	2	<0.0001
vesseleffectf:sboat	11.8	6	0.07
pbycatch:psegment	3806	1	<0.0001
pmonitor:psegment	3701	1	<0.0001
vesseleffectf:psegment	2.59	3	0.45
sboat:psegment	2.09	2	0.35
pbycatch:pmonitor:vesseleffectf	156.3	3	<0.0001
pbycatch:pmonitor:sboat	2764	2	<0.0001
pmonitor:vesseleffectf:sboat	13.58	6	0.03
pbycatch:pmonitor:psegment	1993	1	<0.0001
pbycatch:vesseleffectf:psegment	26.29	3	<0.0001
pmonitor:sboat:psegment	100.0	2	<0.0001
pbycatch:pmonitor:vesseleffectf:psegment	9.74	3	0.02
pbycatch:pmonitor:sboat:psegment	69.8	2	<0.0001

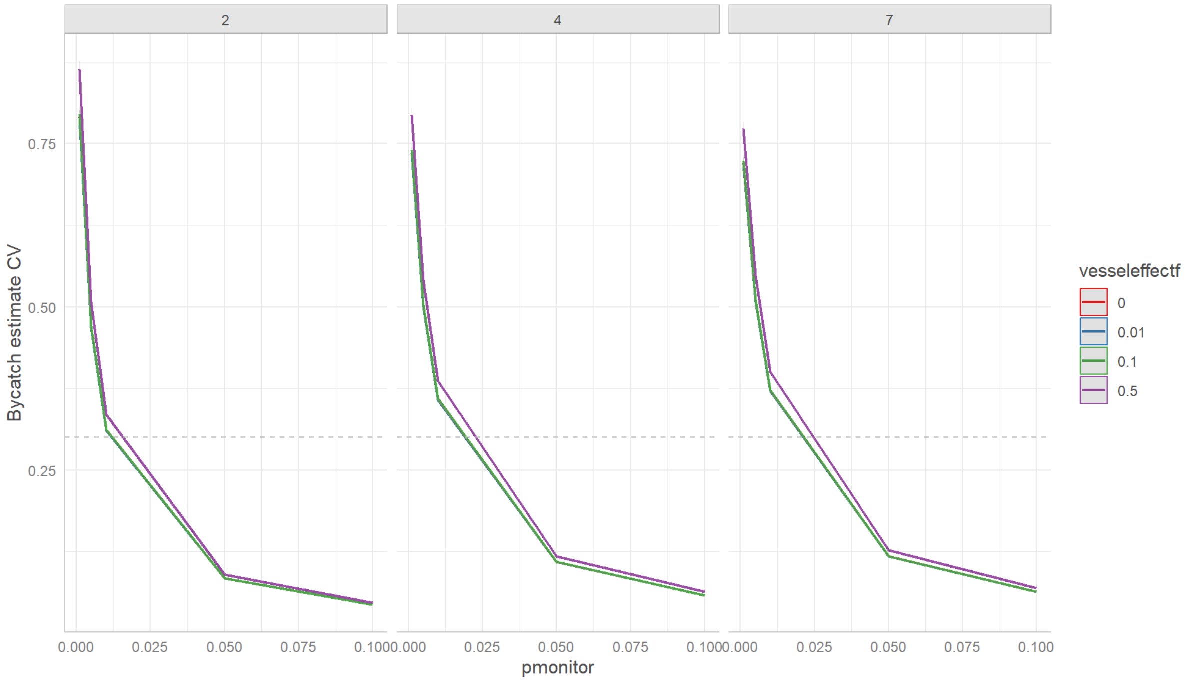


Figure 25. Predicted bycatch estimate precision (CV) based on the best model describing simulation sets (Table 14). Here we focus on the partial effect the number of vessels monitored (panels), the proportion of hauls fished that year which were monitored ($p_{monitor}$) and the between-vessel variability in bycatch ($vessel_{effect}$). potential target of CV = 0.3 is highlighted with a dashed line.

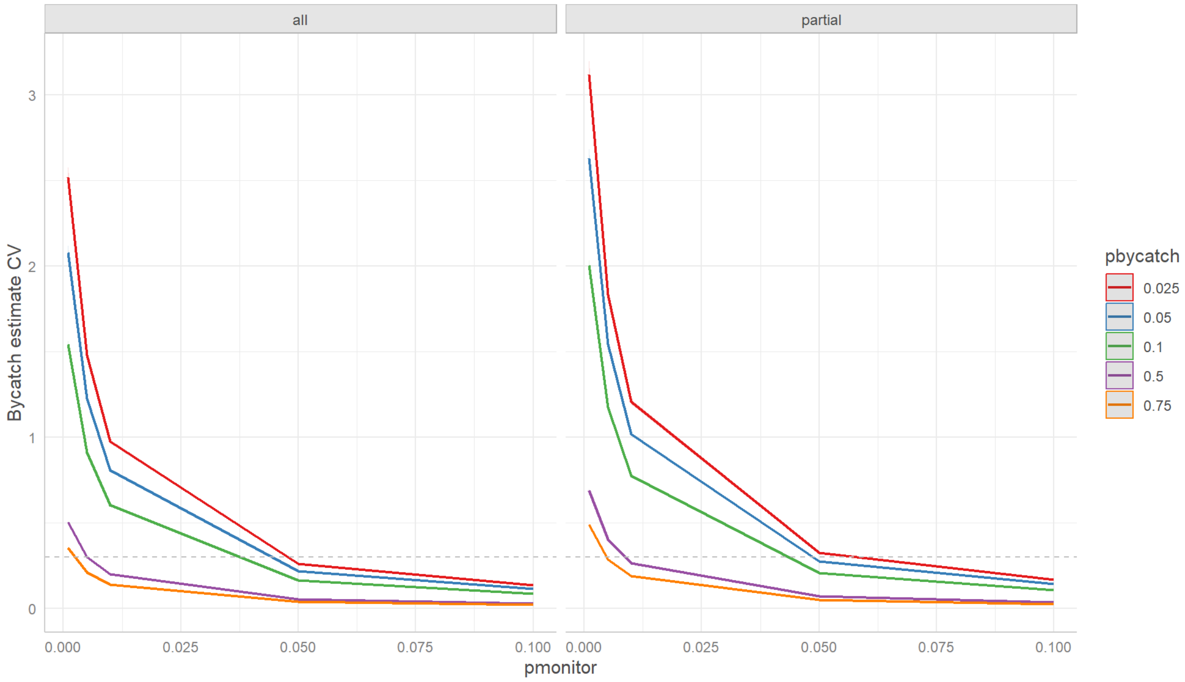


Figure 26. Predicted bycatch estimate precision (CV) based on the best model describing simulation sets (Table 14). Here we focus on the partial effect the number of segments observed (panels), the proportion of hauls fished that year which were monitored ($p_{monitor}$ in x-axis) and the bycatch probability ($p_{bycatch}$ in line colours).

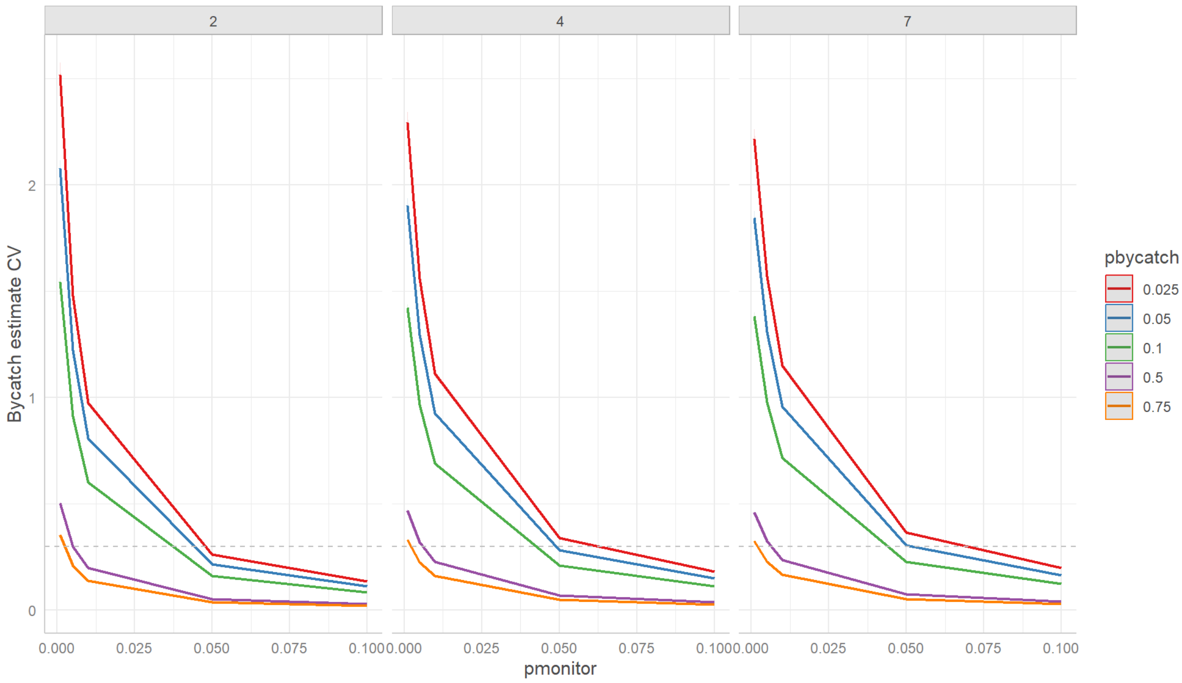


Figure 27. Predicted bycatch estimate precision (CV) based on the best model describing simulation sets (Table 14). Here we focus on the partial effect the number of vessels monitored (panels), the proportion of hauls fished that year which were monitored ($p_{monitor}$ in x-axis) and the bycatch probability ($p_{bycatch}$ in line colours). Here the $vessel_{effect}$ is set to be 0.

5.3. Discussion

Previous trials to estimate the level of bycatch in the deepwater longline fishery for scabbardfish off mainland Portugal (fishery considered in Case Study 4) concluded that the sampling effort achieved, by year, was insufficient to provide information on the abundance or biomass on deep-water shark species (Figueiredo and Moura, 2019; ICES, 2024). This was related to confounding effect between the fishing vessel and its fishing grounds, the potential aggregation of species in particular areas (e.g. Moura et al., 2014) and the effect of depth in the probability of capture (e.g. Veiga et al., 2013; Veiga et al., 2015).

Based on the current preliminary observed bycatch probabilities for the species of concern in Case Study 4, a bias of 2% can be achieved for all species with a monitoring coverage of 5% and a potential target CV (0.3, EU Regulation 812/2004) can be achieved with a monitoring coverage of 7.5%. However, this is only possible if all segments of monitored hauls are observed. If those are partially observed and haul bycatch is estimated by extrapolating the observed bycatch to the unobserved segments in the monitored hauls, then the process introduces bias and decreases precision. Observing whole monitored hauls can be logistically challenging at sea. One solution could be to estimate bycatch at the segment level instead and estimate segment level BPUE which is then raised to the number of segments fished. Regardless of the number of sampled hauls, this small methodological change would improve precision and accuracy.

Distributing monitored hauls across all vessels can counteract the effect of potential between-vessel variability in BPUE which is likely to arise from “unknown” area-specific variance in bycatch probability. Such design can control for such variance to retrieve a precise and accurate fleet level bycatch estimate.

Switching to a monitoring design which focusses sampling on a few vessels deteriorates the accuracy of the bycatch estimates. Importantly, in such a design, increasing the monitoring coverage of the vessels monitored does not increase the accuracy of the bycatch estimate for the fleet. As the proportion of monitored hauls are focussed on a smaller proportion of the fleet, increasing the proportion of monitored hauls increased the monitoring coverage for this segment of the fleet. In other words, we know more about this segment, but this segment is not necessarily representative of the fleet. We therefore retrieve more precise bycatch estimates, but these can be biased if used to represent the whole fleet, depending on how similar we can expect bycatch probabilities to be between vessels ($vessel_{effect}$).
Here these simulations provide a way to explore trade-offs between sacrifices of accuracy and precision. It may be easier to implement a monitoring design focussing on some vessels only, especially because of insufficient crew capacity of some vessel to accept observers onboard or due to other types of refusals.

In this case study, knowing a little bit about the bycatch patterns of all vessels is substantially more beneficial than knowing a lot about a few vessels. If only some vessels can be monitored, maximising the proportion of the fleet monitored will disproportionally improve accuracy: adding one vessel adds more accuracy than the accuracy gain that can be achieved by adding more monitoring to already monitored vessels. Precise bycatch estimates can be achieved with low coverage of the monitored vessels, because bycatch is not rare, but it is important to remember to interpret this precision in the context of the monitoring design: we can be precise about the segment of the fleet monitored, not the whole fleet. The level of between-vessel variability can be appraised by observing between-vessel variability in BPUE and therefore we can refer to simulation outcomes here to get a perception of how large the between-vessel variability is a problem to infer bycatch at the fleet level from the vessels monitored. We would suggest making sure to account for this between-vessel variability when estimating bycatch from monitoring data (e.g. random effect in generalised linear mixed effects model). This will ensure to retrieve a less biased bycatch estimate for the segment of the fleet monitored, but also an estimate of whether the between-vessel variability is impairing the ability to make inferences at the fleet level from this monitored segment estimate. As a final note, if a vessel-focussed monitoring design is chosen, then any monitoring of hauls performed by typically unmonitored vessels will provide a disproportionate benefit for the ability to make inferences for the whole fleet.

Limitations

One key limitation is the assumption that all vessels have the same hauling capacity: all fishing trips last one day and on each fishing trip, a vessel only deploys one fishing haul. This does not represent the current situation as a small proportion of vessels (c. 15-20%) are larger and can deploy two hauls per fishing trip. However, those hauls are smaller, in terms of the number of segments per haul, and we are somewhat capturing this effect with the difference in the number of segments between hauls (Figure 15). We will extend simulations to explore the implication of this fishing heterogeneity further by specifying that the segment/haul variable varies between vessels rather than between fishing trips to provide further guidance for case study 4.

5.4. References

ICES. 2024. Report of the Working Group on Elasmobranch Fishes (WGEF). ICES Scientific Reports. 06:75. 994 pp. https://doi.org/10.17895/ices.pub.26935504

Figueiredo, I., Moura, T. 2019. Onboard sampling information for deep-water sharks in Portuguese mainland waters (2009-2018). Working Document presented in the ICES Working Group on Elasmobranch Fishes (WGEF), Lisbon, 18-27th June. 21 pp.

Moura, T., Jones, E., Clarke, M.W., Cotton, C.F., Crozier, P., Daley, R.K., Diez, G., Dobby, H., Dyb, J.E., Fossen, I., Irvine, S.B., Jakobsdottir, K., López-Abellán, L.J., Lorance, P., Pascual-Alayón, P., Severino, R.B., and Figueiredo, I., (2014). Large- scale distribution of three deep-water squaloid sharks: integrating data on sex, maturity and environment. Fisheries Research, 157, 47–61.

Veiga, N., Moura, T., Figueiredo, I., (2013). Spatial overlap between the leafscale gulper shark and the black scabbardfish off Portugal. Aquatic Living Resources 26, 343–353.

Veiga, N., Moura, T., Figueiredo, I. (2015). Spatial overlap between Portuguese dogfish and the black scabbardfish off Portugal. Working Document for the ICES Working Group on Elasmobranch Fishes, Lisbon, 17th -23rd June 2015. 18 pp.

6. Mitigation trial power analyses

The Northern Gillnet case study (CS1) trialled pearl nets[1] as a gear modification to reduce or eliminate harbour porpoise bycatch. The first trial in Iceland was designed following existing guidelines to try and ensure sufficient replication to maximise effect detection and minimise bycatch[2]. The first trial resulted in no bycatch in control or modified nets. It is therefore crucial to understand what effect of the mitigation technique could have resulted in such an outcome given the experimental design.

We designed a simulation to replicate the experimental design and assessed the likelihood of the observed outcome (zero bycatch) under differing conditions of bycatch probability and mitigation effect sizes.

6.1. Experimental design

Here we assume that we deploy 4 control nets, 2 pearl nets, and 2 nets in which pingers are attached every day for 15 days. The location of the nets is such that we assume independence of bycatch probability between nets. That is, for example, if pingers displace porpoises from the vicinity of nets with pingers this process does not increase the bycatch probability in control nets. We then simulate the bycatch probability in control nets (p_c), the effect of pearl nets ($pearl_{effect}$), the effect of pingers ($pinger_{effect}$).

6.2. simulation design

We explored values of p_c ranging from 0.01 to 0.9 every 0.05 and values of 0.0001 and 0.001. We explored values of $pearl_{effect}$ and $pinger_{effect}$ ranging from 0.05 to 1 every 0.05. For each combination of values we iterate 1000 times the following simulation in order to get a good understanding of precision[3]. For each iteration, we simulate 15 days of mitigation trials following the experimental design. For each control net we randomly draw bycatch using a binomial probability density function with a mean of pc. For each modified net we do the same using $p_c(1-pearl_{effect})$ and $p_c(1-pinger_{effect})$ respectively as means for pearl nets and nets with pingers.

We can then ask first how often, for each parameter combination, did we simulate no bycatch observations. This will help us identify how likely the trial outcome was for different ranges of mitigation effect size and baseline bycatch probability. Secondly, if bycatch was observed in the simulation, we estimated the effect size of gear modification. We use a generalised linear model, assuming a binomial distribution of residuals, where bycatch is assumed to depend on net type. We did not assume any within-day covariance of net deployment in the simulations, we therefore did not include a random effect of day (day 1 to day 15), but this could be introduced if between-day variability was informed in the simulations in the future.
This is done for each iteration so that we can estimate the probability to detect an effect of the gear modification with a alpha level of 0.05 and also determine the accuracy of the effect size estimate.
Finally, we use the power analysis prospectively to estimate the change in power to detect an effect if the duration of the trials were reduced from 15 days to 9 days (exploring all durations

[1] Lotte Kindt-Larsen et al., 'Pearls Are Not Just for Girls: Plastic Spheres Do Not Interfere with Target Catches in a Set Net Fishery', *Fisheries Research* 276 (August 2024): 107032, https://doi.org/10.1016/j.fishres.2024.107032.
[2] Lotte Kindt-Larsen et al., 'Harbor Porpoise (Phocoena Phocoena) Reactions to Pingers', *Marine Mammal Science* 35, no. 2 (2019): 552–73, https://doi.org/10.1111/mms.12552.
[3] B F J Manly, *Randomization, Bootstrap, and Monte Carlo Methods in Biology* (Chapman & Hall, 1997).

from 9 to 15 days). To do so we replicate the simulation and analytical approach described above for experiment duration of 9 to 15 days with 1 day increase (seven simulation sets).

6.3. Results

In the results section we focus only on $pearl_{effect}$ contrasting control nets and pearl nets. The probability to observe zero bycatch over 15 days is dependent on p_c (Figure 28). This is probable if the control bycatch probability (p_c) is less than 0.05 and most likely if it is below 0.001.

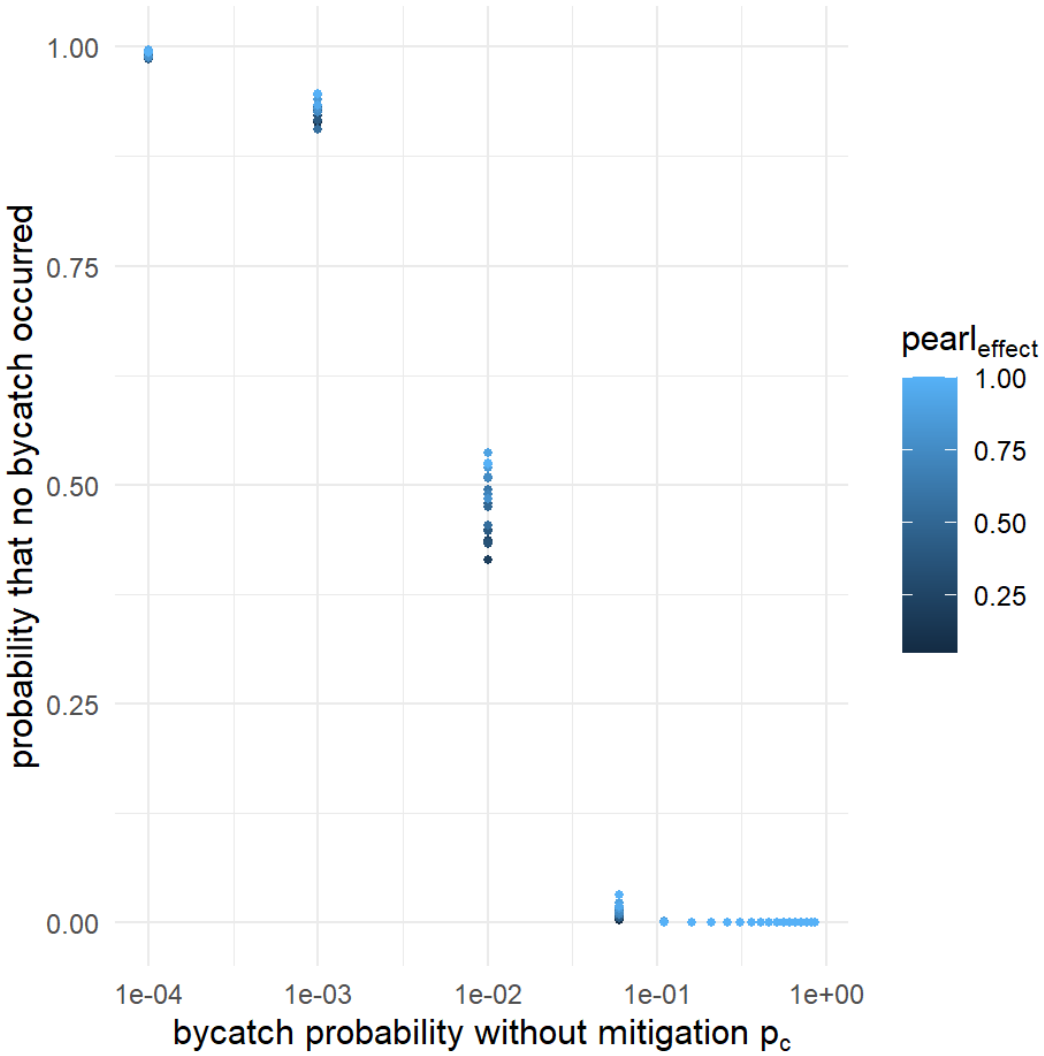


Figure 28. Probability that no bycatch occurred for 15 days of deploying the experimental design.

The experimental design will have power to detect large mitigation effects (more than 60% reduction in bycatch) when the control bycatch probability is also large (more than 0.3) (Figure 29).

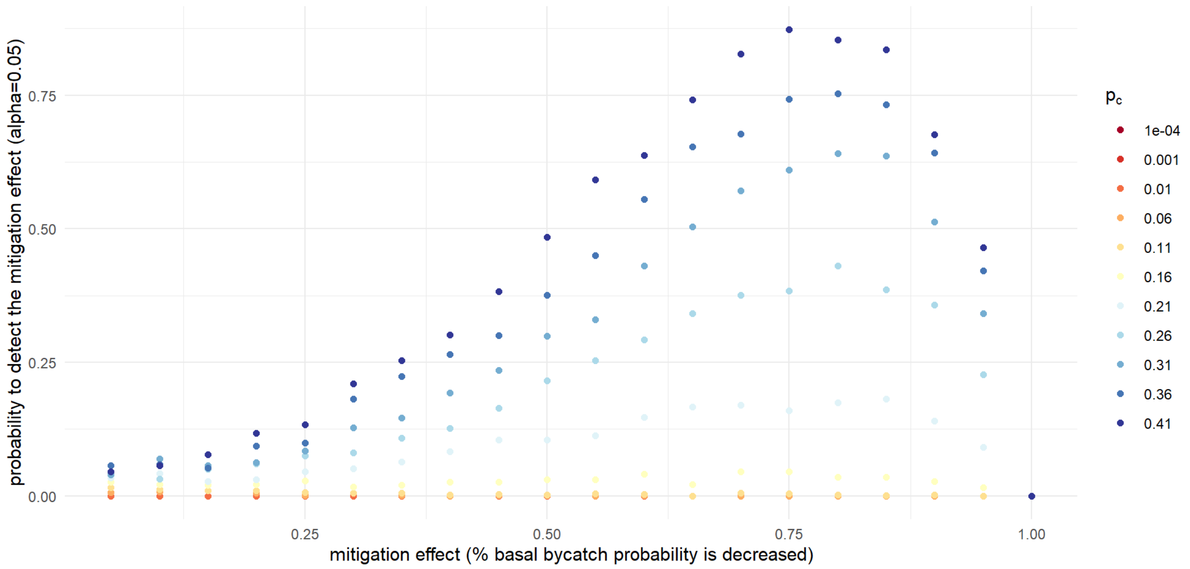


Figure 29. Probability to detect an effect of gear modification with the experimental design deployed for 15 days depending on the control bycatch probability (p_c) and the effect of gear modification ($pearl_{effect}$). Probability is estimated from the number of times effect detection was significant out of 1000 bootstrap replicates.

We can finally estimate the accuracy of the effect size estimate for those instances in which an effect can be detected. Here we retained only simulation sets for which the probability to detect a mitigation effect reasonable (Figure xx) and estimated the accuracy of the estimate of the mitigation effect as:

$$accuracy = \frac{\left|\widehat{pearl}_{effect} - pearl_{effect}\right|}{pearl_{effect}}$$

We find that when an effect is likely to be detected, it will be accurate; broadly falling within 5% of the real effect size ($pearl_{effect}$) (Figure 30). Of course, variability associated with between-day changes in non-measured variables (porpoise density, oceanographic conditions etc) will introduce additional variance which may decrease accuracy. However, models are available to account of this unobserved variance (e.g. generalised linear model with random effect of day).

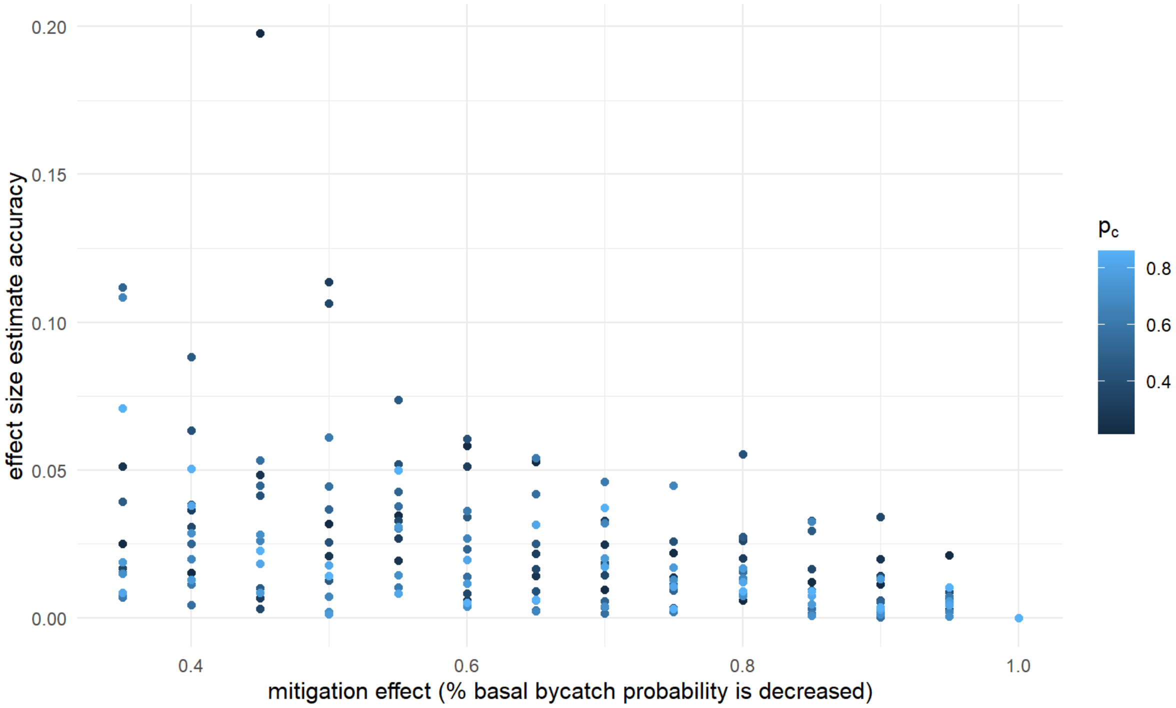


Figure 30. accuracy of the estimated pearl$_{effect}$ ($\widehat{pearl_{effect}}$) for simulations with medium and large effect size (>0.3) and a large control bycatch probability (>0.2).

6.4. Conclusions

Mitigation trials must consider a difficult trade-off between minimising replication to minimise bycatch itself, and maximising replication to be able to detect and estimate an effect when it is present. Overall, experiments designed with this trade-off in mind will be most likely to detect mostly large bycatch probability reduction. To do so, trials are best conducted in location where bycatch probability is large itself in the first instance. Of course, no one can control inter-annual variability in bycatch probability; therefore, the best laid plans still rely on knowledge prior to deployment of the experiment and therefore cannot control changes in control bycatch probability. The power analysis for this case study helps us identify this. In conclusion, mitigation trials are best conducted in location where bycatch rate is large and for gear modification techniques for which we can anticipate large bycatch reductions from first principles or a priori information collected in controlled environments.

7. What can be achieved with monitoring programmes when bycatch is rare

Simulations can help provide advice at all stages of monitoring design including a posteriori and a priori power analyses. Here we provide a general platform (SCOTI). It can be used to explore general guidance about monitoring design for broad-scale characteristics of fisheries x species interactions. We show it can also be tuned to specific case studies to provide guidance on monitoring design considering real-world access challenges monitoring programmes often face.

Overall, we see that when bycatch is rare (< 1 bycatch per 1000 fishing operations) it is difficult to obtain a precise and accurate BPUE estimate. The question becomes then how to best use monitoring. We are thus left with two options. First, move to a census approach where all fishing operations are fully monitored. For example, the use of EM may potentially cover all operations. However, deployment of EM at this scale is financially and logistically challenging. It would also require a large workforce to post-process the observations even with the aid of automation in bycatch detection (so-called “AI”). This will likely be associated with other bycatch detection and identification challenges and consequent bias. Second, an additional approach would be to shift the bycatch monitoring objective. At such low BPUE, it may be advantageous to focus on detecting whether bycatch occurs or not in the fishery, rather than estimate the rate at which it occurs. Bycatch can be rare for three reasons: i) the bycatch probability is not small, but the species is rare (e.g. bycatch exposure for a critically endangered species), ii) the species is not rare, but the bycatch probability is very small (e.g. bycatch from a relatively safe fishing method for an abundant species), and iii) the bycatch probability is small and the species is rare. In the first instance, any bycatch may represent a conservation risk for the population. Monitoring to guide management should then focus on detecting whether bycatch occurs, e.g. by concentrating monitoring effort on the segment of the fishery most likely to catch the sensitive species if it is known, rather than distributing the monitoring coverage to attempt to accurately estimate BPUE for the whole fishery. The same applies for the third instance. In the second instance, bycatch does not necessarily pose a threat to the population conservation status. Again, it is more informative to understand whether the probability of at least one bycatch occurring changes through time, rather than to estimate BPUE accurately, to monitor whether this seemingly safer situation changes and warrants further scrutiny. In short, not being able to estimate precisely and accurately BPUE is not necessarily paralysing bycatch management, monitoring is still crucial to inform whether bycatch occurs or not.

8. Document information

EU project reference	LIFE22-NAT-NL-LIFE-CIBBRiNA/101114301
Project name	Coordinated Development and Implementation of Best Practice in Bycatch Reduction in the North Atlantic, Baltic and Mediterranean Regions
Project acronym	CIBBRiNA
Project website	https://cibbrina.eu/

Milestone	**No.**	13	**Title**	Estimate robust monitoring programme
Work Package	**No.**	6	**Title**	Bycatch assessment toolkit
Work Package Leader	David Lusseau & Ailbhe Kavanagh			
Lead Beneficiary	DTU			
Authors	David Lusseau and Outi Tervo led the authorship of this deliverable. Early version of sections 1-3 benefited from comments and revisions by Martin Pastoors, Ailbhe Kavanagh, and Ruth Fernandez. Section 4 benefited from significant insights input from Al Kingston, Estanis Mugerza, and Gildas Glemarec. Inês Farias, Teresa Moura, Ivone Figueiredo, Rita Vasconcelos, Ana Cláudia Fernandes contributed as co-authors to section 5. Section 5 was co-created with those case study 4 members who not only contributed to the text of section 5 but also contributed parameterisation of the simulations. Lotte Kindt-Larsen contributed to the parameterisation of the power analyses in section 6.			
Reviewer(s)	Mindfully Wired, Inês Farias, Teresa Moura, Ivone Figueiredo, Rita Vasconcelos, Ana Cláudia Fernandes			
Due Date	30 Aug 2025			
Submission Date	30 Oct 2025			
Dissemination Level	PU			

Version Log			
Issue Date	**Version**	**Author**	**Change**
30 Sep 2025	0.0	David Lusseau and Outi Tervo	First version
10 Oct 2025	1.0	David Lusseau and Outi Tervo	Revisions after review by Mindfully Wired
25 Oct 2025	2.0	David Lusseau and Outi Tervo	Revisions after review by Inês Farias, Teresa Moura, Ivone Figueiredo, Rita Vasconcelos, Ana Cláudia Fernandes
28 Oct 2025	3.0	David Lusseau, Outi Tervo, Inês Farias, Teresa Moura, Ivone Figueiredo, Rita Vasconcelos and Ana Cláudia Fernandes	