Accurately estimating catch

AN ILLUSTRATION OF THE EFFECTS OF BIAS IN INDEPENDENT MONITORING OF FISHERIES

The Nature Conservancy

gema
Global Electronic Monitoring Accelerator

UNIVERSITY of TASMANIA
IMAS
Institute for Marine and Antarctic Studies

Contents

Accurately estimating catch

AN ILLUSTRATION OF THE EFFECTS OF BIAS IN INDEPENDENT MONITORING OF FISHERIES

AUTHORS:

Christopher J. Brown
Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, Tasmania (contact: c.j.brown@utas.edu.au)

Vienna R. Saccomanno
The Nature Conservancy (contact: v.r.saccomanno@tnc.org)

Ben Gilmer
The Nature Conservancy (contact: ben.gilmer@tnc.org)

INDEPENDENT PEER-REVIEW:

Kylie Scales
Associate Dean (Research), School of Science, Technology & Engineering University of the Sunshine Coast (contact: kscales@usc.edu.au)

Key findings

- Using a computer simulation of a hypothetical longline tuna fleet, this study demonstrates that allocating comprehensive independent monitoring coverage across a fleet and reviewing the associated monitoring data at random is the best option to ensure that catch data are accurate, adequate, consistent, and unbiased.
- When monitoring coverage was anything less than 100% across the simulated fleet, this invited opportunities for opt-in bias and behavioral changes, which resulted in underestimates of mean catch rates and higher annual variability in catch rate estimates.This translates into an increased likelihood of inaccurate estimates of market and bycatch species populations, risking the overall sustainability of marine wildlife populations.
- We recommend that Regional Fisheries Management Organizations (RFMOs) and sustainability certifications adopt clear guidelines requiring 100% independent monitoring across a fleet or fishery with random review of at least 20% of fishing activity, to ensure fisheries managers have credible catch estimates and better visibility of fishing practices to advance the long-term health of fish populations.

Introduction

Electronic monitoring (EM) uses video cameras and sensors to independently record fishing activity; the electronic records are later reviewed as a source of information that is independent of logbooks. As EM programs have expanded globally over the past decade, a variety of EM coverage and footage review strategies have emerged—shaped by fishery characteristics, monitoring goals, species prevalence, funding, and human capacity to perform data review.

EM has become an essential tool for meeting independent observation requirements set by sustainability certifications such as the Marine Stewardship Council (MSC). Likewise, since the adoption of EM standards across various RFMOs, RFMO Members are now exploring how EM can fulfill coverage and data submission requirements. To meet target monitoring rates and generate representative catch statistics, EM programs often rely on sub-sampling methods when reviewing EM footage. However, allocation of EM coverage and video review rate differ significantly from traditional human observer programs, prompting a need for clarity on how EM sampling procedures affect fisheries data quality and accuracy.

Scientifically robust sampling of EM footage looks at a subset of fishing activity that is representative of the characteristics of the entire fleet. If sampling is not representative, catch estimates may be biased and over- or underestimate catch events and result in missed observations of important interactions—including those with endangered, threatened, and protected (ETP) species. For instance, the MSC is proposing independent observation of at least 20% of fishing events per year

for high seas operations to track statistics like the catch rate of both market and bycatch species. In recent years, technical guidance on minimum standards for individual vessels with EM has been developed for key fisheries such as longline and purse seine tuna (e.g., Murua et al. 2025); however, there is limited guidance on how coverage should be allocated across an entire fishery. This opens the door for selective monitoring—where nations or companies might unintentionally (or intentionally) monitor only their "cleanest" 20% of vessels or trips, potentially masking problematic fishing practices or underrepresenting catch activity.

To explore this issue, we conducted a computer simulation using a hypothetical tuna longline fishery informed by the operational dynamics of real-world fishing (Brown et al. 2021) to assess whether different methods of selecting fishing activities for monitoring and review yield accurate estimates of true catch rates. We evaluated three distinct monitoring and review scenarios and analyzed their impact on the accuracy of estimated catch rates for both market species and bycatch.

Why this matters: If monitoring is biased toward vessels or trips within a fleet with cleaner fishing practices, regulators and the public may be misled into believing a fishery is performing better than it actually is. Conversely, if monitoring is biased toward vessels with poor fishing practices, the data will be skewed accordingly. This can result in misinformed management plans and allow harmful fishing practices to persist undetected, undermining sustainability goals and the credibility of fisheries management systems.

How this study was done

Simulating catch

We aimed to illustrate how different monitoring and review strategies of fishing activity affect the accuracy of catch rate estimates (see Box 1 for terminology details). We developed a computer simulation of a fictional longline tuna fleet composed of 50 vessels, informed by the operational dynamics of real-world fishing (e.g., Brown et al. 2021). In our fictitious world, a fishery manager must decide how to allocate monitoring and data review resources for the upcoming year, knowing only the number of vessels—not how much fishing effort each vessel will exert or how much they will catch.

Fishing activity across the fleet was simulated over a one-year period. Each vessel completed one or more trips, with each trip consisting of a randomized number of longline sets (averaging 26 sets per trip). The variability in fishing activities reflected realistic differences in trip duration and fishing effort. In this fictitious fishery, the catch composition varied among individual sets, across trips, and across vessels. This variation is representative of differences in fishing location, timing, hook number, bait type, skill level, and/or fisher behavior and/or knowledge.

The fishery had five types of catch events, and each event had a different average catch rate : (1) a commonly caught target market species (e.g., yellowfin tuna); (2) a commonly caught species with high variability in catch rates across vessels, representing differences in captain expertise; (3) a commonly caught market species with high variability in catch rates across trips, simulating variation in skill and flexible fishing strategies by captains to modify bycatch; (4) a bycatch species caught less consistently (e.g., blue shark); and (5) a rare and vulnerable bycatch species (e.g., green turtles). The rare event could also represent other infrequent but significant events such as transshipment.

These parameters were chosen to be representative of catch rates for a range of different types of industrial fishing activity, including longline and gillnet fisheries (see Appendix A for details).

BOX 1: important terminology

Effective fisheries management depends on high-quality data to inform management tools such as stock assessments and ensure regulatory compliance. Traditionally, managers have relied on surveys, paper logbooks, and human observers to track catch and discards. More recently, electronic monitoring has emerged as a cost-effective and comprehensive alternative, using video footage to document fishing activity for later analysis.

Because EM captures video for post-trip review and human observers record events in real time, these two methods entail different processes to yield final fishing activity data. Because of these differences, we use the following terms intentionally throughout this study:

» **Monitoring coverage** refers to the proportion of vessels, trips, or sets with either EM or human observers present continuously collecting fisheries data.

» **Review** refers to the analysis of fishing activity that had monitoring coverage—whether video footage or observer records—to generate final catch estimates. **Review rate** refers to the proportion of monitored fishing activity that is used to generate final catch estimates.

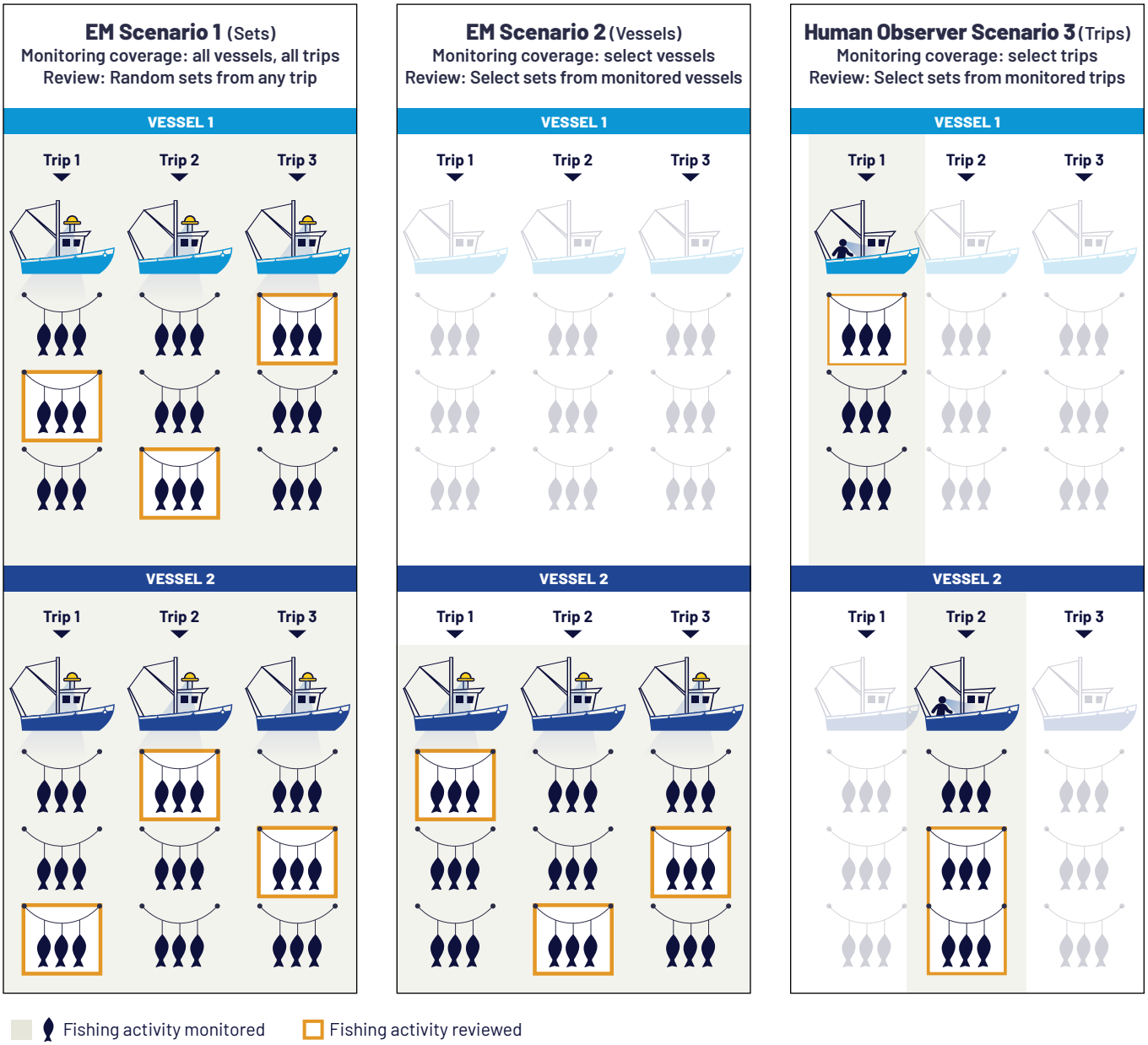
Simulating monitoring and review scenarios

We tested three monitoring and review scenarios that our fictional manager could choose between if given the guideline "20% coverage of independent monitoring". The first two scenarios focus on EM as the monitoring tool; the third scenario focuses on human observation (see Monitoring Scenarios schematic on page 5). For all scenarios, data review was of whole sets.

- **EM Scenario 1: monitor 100% of vessels and review 20% of sets at random.** EM was present on 100% of vessels in the fleet and 20% of sets were randomly selected for review. This reflects a fleet with full electronic monitoring coverage but limited capacity to review EM footage.

- **EM Scenario 2: monitor 20% of vessels and review 20% of sets.** EM was present on 20% of vessels and 20% of the sets that were monitored were reviewed, representing a fleet where EM systems are installed on a subset of vessels. The manager could select vessels for monitoring in two ways: 1) at random or 2) biased towards selecting vessels with lower-than-average catch rates, simulating strategic deployment of EM on vessels with cleaner fishing practices.
- **Human Observer Scenario 3: monitor and review 75% of sets on 20% of all trips.** Human observers were present on 20% of trips made by vessels in the fleet and 75% of the sets were analyzed, representing limitations of human observers to analyze complete fishing activity due to periods when they might not be on deck, such as sleeping, eating, etc. The manager could select trips for monitoring in two ways: 1) at random or 2) biased toward trips with lower-than-average catch rates, simulating behavioral changes that might occur under observation (e.g., gear adjustments to reduce bycatch, Benoît and Allard 2009).

Schematic: monitoring and review scenarios



Simulation methods

We explored whether the three fictional monitoring and review scenarios would obtain unbiased catch rates when applied to each of the five catch event categories. We simulated fishing activity for the fictional fleet across 1,000 replicates. These replicates capture the range of variation in catch rates that is possible to encounter in a fishing year. For each replicate we applied each of the three monitoring scenarios, and allowed the manager to allocate review in a random or biased way.

For each catch event category across the replicates we calculated **mean bias**. The percent bias was the percent difference between the monitored catch rate and the true catch rate. The mean bias was the mean of percent bias across the 1,000 replicates. This metric indicates how close each monitoring and review scenario comes to accurately estimating the true catch rate. A mean bias of 0% represents monitoring that is accurate on average across many years of fishing.

In addition to mean bias, we assessed the **variance in the bias statistic.** The variance represents how consistent the results will be across different years of fishing. Ideally, a monitoring and review scenario would yield both low bias and low variance, meaning it consistently produces accurate estimates year after year. However, a scenario with low average bias but high variance means that catch rates in any single year could be well above or well below the true catch rate. High variance is a problem for managers because it increases the chance of spurious estimates in catch rates across multiple years of fishing.

Results: monitoring and review scenario findings

EM Scenario 1 → *Monitor 100% of vessels and review 20% of sets at random*

This scenario resulted in no bias on average across all five catch categories and generally low variability across replicates (Figure 1). For the market species catch category, the consistency of results means that the manager can be confident that the estimated catch rate in any given year will fall within ±19% of the true value. Rare bycatch events had higher variability (±34% of the true value), meaning that there is a higher likelihood that when these events occur they could be missed or over-estimated during the 20% EM footage review process. For example, if an average of 139 rare turtles were caught as bycatch and 97,500 market tuna were caught by the fleet in a given year, this monitoring scenario would have estimated between 70-215 turtles (95% CI) and 67,800-134,900 tuna (95% CI).

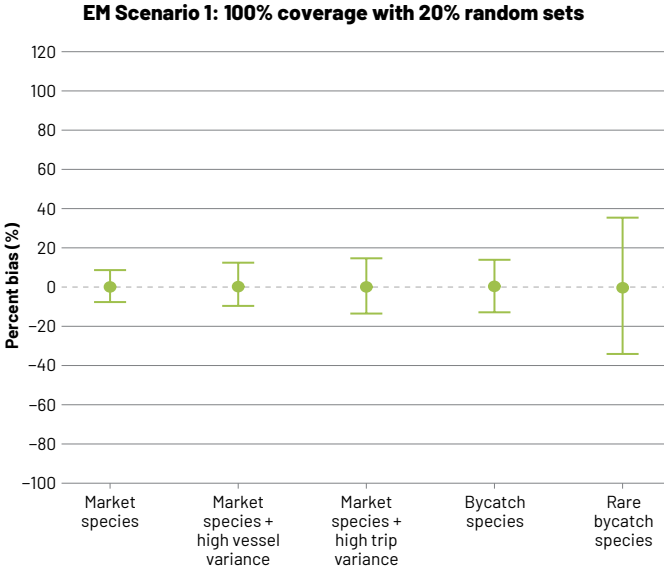


Figure 1. Percent bias for each catch event category across simulation replicates for Scenario 1. Points show mean bias and error bars show 95% quantiles for percent bias across replicate simulations (95% of simulations fill within the bounds). The mean bias is the amount of bias we would expect to see on average across many years of fishing. The variability shows the range of bias values a manager would be exposed to in data from any single year of fishing.

This page: © David Hills Photography; Opposite: © Jonne Roriz

EM Scenario 2 → *Monitor 20% of vessels and review 20% of sets*

When vessels were selected at random in this scenario, there was no bias on average but generally high variability in bias across replicates (dark green scenario). The high variability means the manager cannot be sure that the catch rate for any given year is accurate. For the market species catch category, the catch rate in a single year could be overestimated by as much as 70% or underestimated by as much as 50%. When vessel selection was biased toward those with lower catch rates, average catch rates were underestimated by up to 60% and variability was high (up to 100%) (Figure 2).

When vessels were not selected at random, mean bias systematically decreased below the true catch rate (orange scenario). Non-random selection skews the overall catch estimates because vessels who catch less than other vessels may opt-in and/or vessels with cleaner fishing practices are more willing to have EM installed. This introduces the risk of misrepresenting overall fishery performance. Furthermore, the mean bias was amplified in the catch event category with market catch and high variability in catch rates among vessels. For example, if an average of 97,500 market tuna were caught by a fleet with high variability in catch rates, the vessel based monitoring scenario could estimate as little as 45,000 tuna.

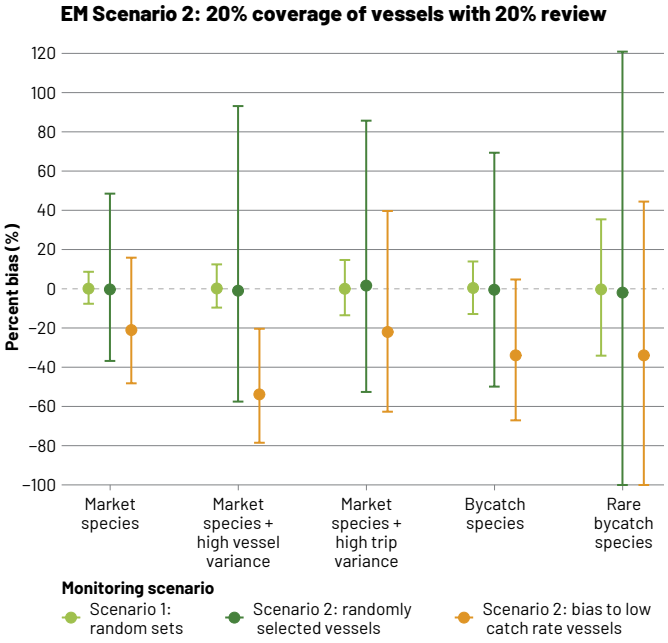


Figure 2. Percent bias for each catch event category across simulation replicates for Scenario 2. Points show mean bias and error bars show 95% quantiles for mean bias across replicate simulations (95% of simulations fill within the bounds). The mean bias is the amount of bias we would expect to see on average across many years of fishing. A bias of 100% means the estimate was double the true catch, a bias of -100% means we never see a species that occurred in the catch. The variability shows the range of bias values a manager would be exposed to in data from any single year of fishing.

Human Observer Scenario 3 → *Monitor and review 75% of sets on 20% of all trips*

When whole trips were selected randomly for human observation results showed greater year-to-year variability compared to Scenario 1, due to the clustering of reviewed data within trips. This meant the manager had lower confidence in the accuracy of annual catch rate estimates, especially in fisheries where average catches varied trip-to-trip (e.g. they may overestimate catch rates by as much as 50% or underestimate catch rates by as much as 40%).

When trip selection was biased toward those with lower catch rates—simulating behavioral changes by captains under observation—mean catch rates were significantly underestimated, by as much as 80% (Figure 3). This bias was most pronounced if it was assumed the fishery had high variability across trips - this assumption reflects a situation where captains alter fishing practices or locations when they know they have an observer on board. For example, if an average of 97,500 market tuna were caught by the fleet in a given year, this monitoring scenario would have estimated as little as 36,100 tuna (lower quantile for confidence).

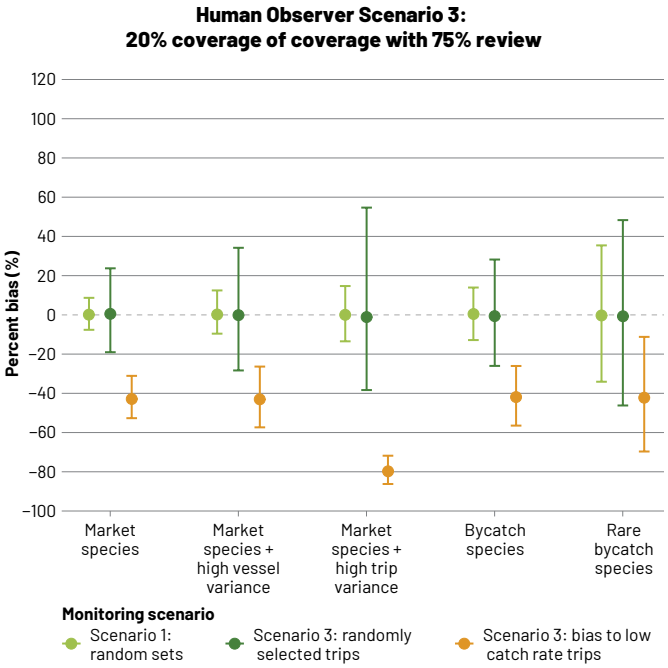


Figure 3. Percent bias as a function of catch event category across simulation replicates for Scenario 3, where review is allocated to whole trips. Points show mean bias and error bars show 95% quantiles for mean bias across replicate simulations (95% of simulations fill within the bounds). The mean bias is the amount of bias we would expect to see on average across many years of fishing. The variability shows the range of bias values a manager would be exposed to in data from any single year of fishing.

This page, from top: © Jonne Roriz; © Erin Feinblatt; © Alejandro Lopez-Serrano/TNC

Summary and discussion

Our simulated scenarios reflect realistic ways that monitoring and review of fishing activity could be allocated by a manager and the impacts said allocation could have on the accuracy and consistency of fishery data, which in turn impacts the manager's ability to achieve essential sustainability objectives.

We found that EM Scenario 1, comprehensive EM coverage across the fleet with random review (Box 2) across a subset of fishing sets, resulted in unbiased and generally consistent data on fishing activities. This scenario had the lowest mean variance across all catch categories, meaning it consistently produced reliable catch rate estimates year after year. Therefore, EM Scenario 1 is best suited to help the manager annually estimate bycatch and market species catch rates –a fundamental fishery management activity that is needed to ensure populations are healthy and to help inform accurate management plans, along with basic evidence for sustainability certifications like the MSC Fisheries Standard. For 100% coverage, we found that increasing the review rate above 20% improved accuracy (Appendix B).Comprehensive EM coverage across a fleet can also be used to incentivise higher quality logbook reporting, and subsequently fill gaps from partial EM footage review (Box 4).

When monitoring coverage was not comprehensive across the whole fleet in Scenarios 2 and 3, there were opportunities for bias and behavioral changes. When monitoring effort was allocated by selecting a subset of fishing trips or vessels to be monitored, the variability increased and the catch rate estimates were erroneous, with mean catch rates appearing much lower than they actually were across all five catch categories. The impacts of this bias are that the manager consistently underestimates annual average catch rates, resulting in potentially inaccurate management plans and assessments as to the health and productivity of market fish -and bycatch- populations. The compounding impacts of biased catch rate estimates over time could threaten the long-term sustainability of marine wildlife populations and possibly result in an undetected drop in fish numbers below sustainable limits. Beyond accurate catch accounting, these impacts of bias also prevent managers from having an accurate, fleet-wide picture of other monitoring objectives, such as safe handling of bycatch or adherence to gear practices.

BOX 2: an alternative to random review

An alternative to random selection is **stratified sampling**, where monitoring is allocated based on factors that are known to influence catch rates, such as vessel type, gear, or fishing location. Stratification could match or even outperform random sampling in precision, but it requires a strong understanding of the drivers of catch variability. If key factors are overlooked, monitoring based on stratification may still deliver biased catch estimates.

© Jonne Roriz

BOX 3: alternative monitoring and review rates

Our three scenarios simulate realistic ways that monitoring and review resources can be allocated. Due to this realism, the total sample size of sets reviewed varied across the scenarios. For example, 100% EM coverage with 20% review in EM Scenario 1 is a greater number of sets reviewed than 20% EM coverage with 20% review in EM Scenario 2. Therefore, we also considered the implications of keeping the number of sets reviewed the same for the three scenarios of allocating monitoring effort (by sets, trips, or vessels). Additionally, we compare different monitoring coverage rates of 20%, 30% and 100% (all with 20% review) across the three scenarios.

Even with equal sample sizes, Scenarios 2 (EM on a subset of vessels) and 3 (Observers on a subset of trips) had greater variance than Scenario 1 (random sets). Further, when monitoring coverage was less than 100%, Scenarios 2 and 3 were biased towards underestimating catch rates and variability increased across all scenarios, meaning the manager would have less certainty in annual catch rates (Figure 4).

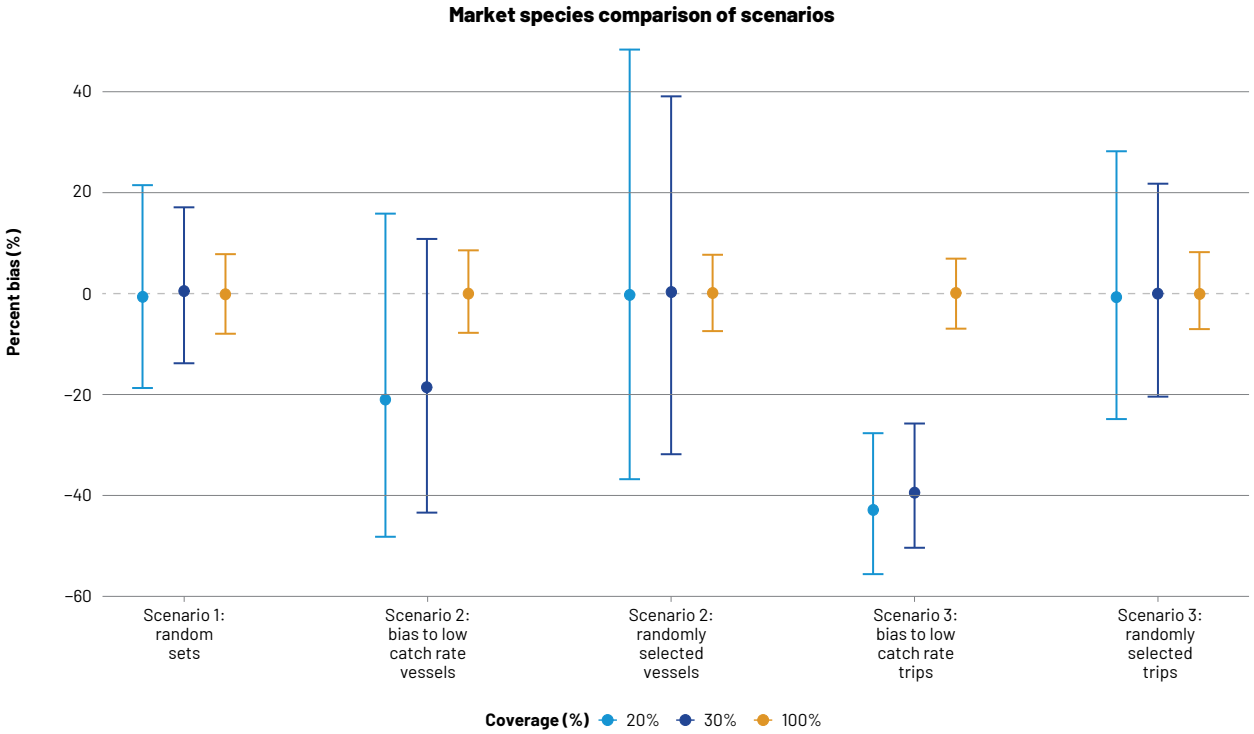


Figure 4. Percent bias across the three monitoring and review scenarios for monitoring coverage rates of 20%, 30% and 100% (all with 20% review) for the market species catch category. Points and bars of the same colour have the same total number of sets reviewed.

Conclusions

Independent monitoring and unbiased catch data are essential to the sustainable management of market catch and bycatch species. Management agencies and certification groups alike should adopt clear guidance on how monitoring coverage and review should be allocated to ensure management and sustainability objectives are achieved. **This study suggests that allocating comprehensive monitoring coverage across a fleet and reviewing the fishing activity that had monitoring coverage at random is the best option to ensure that catch data are accurate and unbiased.**

We recommend that RFMOs and sustainability certifications require 100% independent monitoring coverage across a fleet or fishery with a random review of at least 20% of fishing activity to ensure fisheries managers have credible catch estimates and better visibility of fishing operations.

BOX 4: implications of monitoring scenarios for bias in logbook reporting

Captains who complete logbooks can misreport catch. Especially common is underreporting of bycatch or ETP species, compared to observer data (e.g. Emery et al. 2019; Brown et al. 2021). Further, logbook records are often incomplete, and these data gaps may be biased towards some components of a fishing fleet (e.g. Bellanger, Macher, and Guyader 2016). Independent monitoring can be used to validate logbook records, estimate the rate of underreporting and create an incentive for more accurate logbook reporting.

Observers or EM can increase the accuracy of logbooks when fishers know they are being monitored and there are consequences for inaccurate logbooks (Emery et al. 2019; Bremner et al. 2009). For example, in a long-line fishery in Australia logbook records of turtle interactions went up by about 10 times when electronic monitoring was implemented (Emery et al. 2019). Monitoring with 100% coverage and partial review of randomly selected fishing events creates an incentive for more accurate logbook reporting. The scientific credibility of logbook data is weaker when there is selective application of independent monitoring, especially if it is biased towards trips or vessels with low bycatch rates.

Biased reporting of logbook data also has implications for stock assessments and quota allocations. Stock assessments can be biased to find the current stock status is either too conservative, or not conservative enough when catch data are underreported (Van Beveren et al. 2017; Rudd and Branch 2017). Increased variability in catch estimates across years also creates potential for spurious trends that could bias stock assessments. The effects of catch underreporting on stock assessments are complex, because assessments typically use complex models with many interacting factors. Key findings are that estimates of reference points will be unreliable if catch is under-reported and that underreporting is not accounted for (Van Beveren et al. 2017). The biggest impact will also occur when the level of bias changes over years. In particular, if the level of bias increases, stock assessments will tend towards being less conservative (i.e. not recognizing overfishing), whereas if the level of bias decreases then stock assessments will tend towards being more conservative (Rudd and Branch 2017).

Logbook reporting of catch share management systems provides an example of potential underreporting. When a fishery is restructured, catch shares are most often allocated on the basis of historical catch data (Lynham 2014). This can create an incentive to over-report in logbooks, if impending management changes are known about by fisheries. In the long-term, biased logbook data may also result in inequitable distribution of catch shares to fishers (Lynham 2014).

There is also an opportunity to deploy a risk-based logbook audit model that, when successfully deployed, can help lower EM review rates over time as logbook reporting becomes more accurate and reliable. This model can lower EM program costs and further empower fishers as the primary self-reporting actor through this type of "trust but verify" approach.

References

Bellanger, Manuel, Claire Macher, and Olivier Guyader. 2016. "A New Approach to Determine the Distributional Effects of Quota Management in Fisheries." Fisheries Research 181 (September): 116–26. https://doi.org/10.1016/j.fishres.2016.04.002.

Benoît Hughes P, and Allard, Jacques 2009. "Can the data from at-sea observer surveys be used to make general inferences about catch composition and discards?" Canadian Journal of Fisheries and Aquatic Sciences. 66: 2025-2039. https://doi.org/10.1139/F09-116.

Bremner, Graeme, Peter Johnstone, Tracy Bateson, and Philip Clarke. 2009. "Unreported Bycatch in the New Zealand West Coast South Island Hoki Fishery." Marine Policy 33 (3): 504–12. https://doi.org/10.1016/j.marpol.2008.11.006.

Brown, Christopher J, Amelia Desbiens, Max D Campbell, Edward T Game, Eric Gilman, Richard J Hamilton, Craig Heberer, David Itano, and Kydd Pollock. 2021. "Electronic Monitoring for Improved Accountability in Western Pacific Tuna Longline Fisheries." Marine Policy 132: 104664.

Emery, Timothy J., Rocio Noriega, Ashley J. Williams, and James Larcombe. 2019. "Changes in Logbook Reporting by Commercial Fishers Following the Implementation of Electronic Monitoring in Australian Commonwealth Fisheries." Marine Policy 104: 135–45. https://doi.org/https://doi.org/10.1016/j.marpol.2019.01.018.

Lynham, John. 2014. "How Have Catch Shares Been Allocated?" Marine Policy 44 (February): 42–48. https://doi.org/10.1016/j.marpol.2013.08.007.

Murua H., Ruiz J., Justel-Rubio A., and Restrepo V. 2025. Minimum Standards for Electronic Monitoring Systems in Tropical Tuna Purse Seine and Longline Fisheries. ISSF Technical Report 2022-09Rev. International Seafood Sustainability Foundation, Pittsburgh PA, USA

Pierre, J.P., Dunn, A., Snedeker, A. *et al.* Optimising the review of electronic monitoring information for management of commercial fisheries. *Rev Fish Biol Fisheries* 34, 1707–1732 (2024). https://doi.org/10.1007/s11160-024-09895-7.

Rudd, Merrill B, and Trevor A Branch. 2017. "Does Unreported Catch Lead to Overfishing?" Fish and Fisheries 18 (2): 313–23. https://doi.org/10.1111/faf.12181.

Van Beveren, Elisabeth, Daniel Duplisea, Martin Castonguay, Thomas Doniol-Valcroze, Stéphane Plourde, and Noel Cadigan. 2017. "How Catch Underreporting Can Bias Stock Assessment of and Advice for Northwest Atlantic Mackerel and a Possible Resolution Using Censored Catch." Fisheries Research 194 (October): 146–54. https://doi.org/10.1016/j.fishres.2017.05.015.

Wallace, B.P., Lewison, R.L., McDonald, S.L., McDonald, R.K., Kot, C.Y., Kelez, S., Bjorkland, R.K., Finkbeiner, E.M., Helmbrecht, S. and Crowder, L.B. (2010), Global patterns of marine turtle bycatch. Conservation Letters, 3: 131-142. https://doi.org/10.1111/j.1755-263X.2010.00105.x.

Appendices

© Jason Houston/TNC

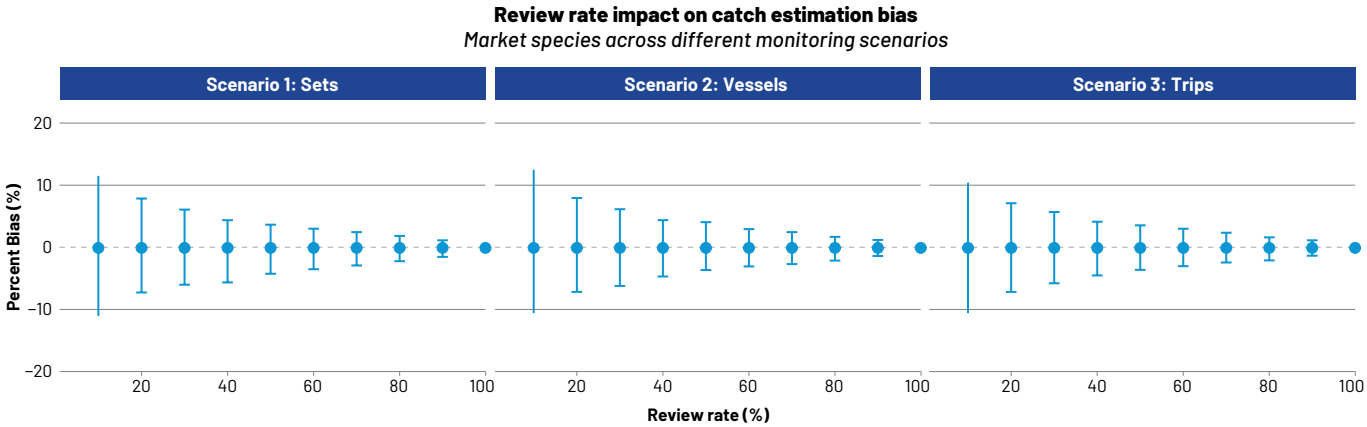
APPENDIX A: representativeness of the data used in the case study

The simulations were designed to be realistic to longline fishery catch of different species in the Pacific tuna longline fisheries, but also representative of a broad range of different fishery types. The simulation model we used was not specific to longline fishing and could be applied to any fishery with vessels, trips, and sets (e.g. vessels making daily trips to set gillnets, vessels making multi-month trips to purse seine, vessels making <1 week trips to set pots for crabs, etc.).

The fishery we simulated with 50 vessels setting around 20,000 sets per year is moderately large on a global scale. Small fleets (e.g. <20 vessels) could have much higher variability (i.e., broader error bars) for the bias statistics because the sample size of vessels and the amount of fishing activity is smaller.

The catch rates we used for the market, bycatch, and rare bycatch species are representative of catch rates in a broad range of gear types operating in different regions of the world. They amounted to an average of: 5.5 animals per set for the market species, 1 animal per 13 sets for the bycatch species and 1 animal per 147 sets for the rare bycatch species. As a comparison, Australian longline fisheries catch an average of 50 animals per set summed across five different market species (figures from Emery et al. 2019, assuming an average of 2500 hooks per set) and Australian gillnet fisheries catch an average of 14.6 of their target species per net (assuming 1000m long nets, Emery et al. 2019). These figures would be similar to our market species if they were apportioned by individual species. Bycatch per set of mammals, sharks, or turtles ranges from 1 animal per 40 sets to 1 animal per 330 sets (summing all species within each of those taxa). Our rates also align with global turtle bycatch rates. Turtle bycatch in fisheries where it is reported ranges from <1 turtle per 500 sets to almost 20 turtles per set, for gillnet, longline, and trawl fisheries globally (Wallace et al. 2010). The higher rates of turtle bycatch in Wallace et al. (2010) would be closer to our analysis of the market species. Lower rates of bycatch than our rare species will have greater uncertainty, increasing the importance of having high coverage and high data review to estimate catch rates of rarely caught species (Pierre et al. 2024).

APPENDIX B: different review rates with 100% monitoring coverage



Appendix B, Figure 1 Percent bias for market species catch for a range of review rates (assuming 100% monitoring coverage) and where monitoring is allocated at random to sets, trips, or vessels. Points show mean bias and error bars show 95% quantiles for mean bias across replicate simulations (95% of simulations fill within the bounds). Mean bias is unaffected by review rate, provided that monitoring coverage is 100%. The variability in bias reduces as review rate increases, because greater review rates mean more data are collected, resulting in greater certainty that the estimated catch rate is close to the true catch rate.

APPENDIX C: advanced methods and simulation code

Transparency statement
Simulation code is available at https://github.com/cbrown5/msc-review-rates

Advanced methods
We developed a general model for (1) simulating catch from a year's worth of fishing activity and (2) simulating monitoring and review of that fishing catch. The model was split into a catch event module that modelled catch per set (Table S1) and a monitoring module that modelled monitoring and review of a sub-sample of all sets (Table S2). Simulation of catch and allocation of monitoring effort were both stochastic. This allowed us to explore how different monitoring scenarios impacted bias in estimated catch rates, with realistic levels of uncertainty. Specifically, our models accounted for uncertainty in: how much fishing activity happens in the coming year, how much catch is taken and by which vessels and on which trips, how monitoring coverage is allocated to fishing activities, what parts of the monitoring data are reviewed (e.g. what video is turned into data sheets that we can estimate catch from) and bias in how monitoring coverage is allocated. These different sources of uncertainty were important because a plan for monitoring needs to be made before the fishing activity happens. With our model, we then simulated replicate years of fishing to obtain catch bias estimates along with ranges of variability.

For the catch model we simulated the number of trips per vessel and number of sets per trip, as well as catch per set. Trips per vessel and sets per trip were modelled as random numbers, to reflect real world variation in their numbers. For simplicity we did not consider correlations in the number of trips or sets per vessel (e.g. if some vessels consistently make fewer longer trips). This would require more detailed fisheries data for analysis than the summary statistics in published work. The number of vessels was fixed at 50. Catch per set was modelled as a random number with additive components for vessel identity and trip identity. This simulation of catches emulates the statistical analyses of catch, which often uses mixed effects models (e.g. Roberson et al. 2025; Gilman et al. 2012). Therefore, catch events included covariation caused by vessel identity and trip identity.

The monitoring scenarios were modelled by randomly sampling sets, trips or vessels for monitoring. A proportion of sets (20%) was then selected within this sub-sample for data review. For the trip and vessel scenarios, a bias parameter was included such that monitoring coverage was biased towards trips or vessels with lower than average catch rates. We considered three monitoring scenarios, as described in the main text. Note that the total sample size in terms of number of sets varied across scenarios. This choice was made so we could represent coverage and review rates that are realistic to different real world interpretations of '20%' monitoring.

We also ran analyses where all of these scenarios the review rate as a proportion of the total sets would be on average 6% if our coverage is 30% and review rate is 20%.

This division between coverage (e.g. what you collect video data of) and review (what you actually watch to get datasheets) is reflective of how real world fisheries operate. The total review quantity was also an average across multiple replicate simulations, because in an individual simulation the actual review rate depends on the fishing activity model, which had stochastic elements. Therefore, our model is realistic to a situation a manager faces where they can allocate observers and cameras to vessels or trips, but they do not know ahead of time how much fishing will happen on those trips and vessels.

The random sampling of fishing activities and catch events was repeated to create 1000 datasets of annual fishing activities. Each monitoring scenario was then applied to each of the 1000 fishery datasets. The bias statistics were calculated as the difference between the true catch rate for that dataset and catch rate estimated by the monitoring scenario. The bias statistics therefore represent the difference between the estimated and true catch rates, conditional on each sample of catches. This conditional sampling is reflective of the situation managers face in the real world, where the observed catch rate differs from the true catch rate by an unknown amount. The confidence intervals across the 1000 simulations therefore represent our uncertainty about the accuracy of catch estimation in the coming year, given we know the number of vessels, but we don't yet know the number of trips, sets or catches those vessels will take.

For a given monitoring scenario M (Table S2) and catch data y, the estimated catch rate per set can be calculated as the average catch per set that was monitored and reviewed:

$$\hat{C} = \frac{\sum (y_{v,t,s} * M_{v,t,s})}{\sum M_{v,t,s}}$$

Where $y_{v,t,s}$ is a matrix of catches and $M_{v,t,s}$ is a matrix of 0/1 that indicates whether a given set was monitored and reviewed.

The true catch rate in each simulation, *C* was simply average catch per set.

We calculated two bias statistics, the absolute bias:

$$B = C - \hat{C}$$

and the bias as a percentage of the catch rate:

$$100 * B/C$$

The percent bias is what we present above, because it puts all scenarios on the same scale. The absolute bias was much smaller for rarer species and therefore hard to visualize in comparisons.

All parameters were chosen to be representative of the Western Pacific longline tuna fleets, data given in Brown et al. (2021), parameters are in Tables S3-s5. The analysis was performed with the R program (R Core Team 2024). The code is available online at https://github.com/cbrown5/msc-review-rates.

TABLE S1: model Equations and parameters of the catch event module

Equation	Description	Parameters
$T_v \sim dnegbin(\mu^{trips}, \theta^{trips})$	Number of trips per vessel per year	μ^{trips}: mean number of trips per year, θ^{trips}: dispersion in trips per year
$x_v \sim dnorm(0, \sigma^x)$	Vessel-level random effect for catches per set	σ^x: vessel effect standard deviation
$S_{v,t} \sim dnegbin(\mu^{sets}, \theta^{sets})$	Number of sets per trip	μ^{sets}: mean sets per trip, θ^{sets}: dispersion in sets per trip
$z_{v,t} \sim dnorm(0, \sigma^z)$	Trip-level random effect for catches	σ^z: trip effect standard deviation
$\mu_{v,t} = exp(\beta_0 + z_{v,t} + x_v)$	Expected catch rate per set for a given trip	β_0: baseline catch rate per set
$y_{v,t,s} \sim dnegbin(\mu_{v,t}, \theta^{catch})$	Catch per set	θ^{catch}: catch dispersion

Notes:

V: Number of vessels

T_v: Number of trips for vessel v

$S_{v,t}$: Number of sets for trip t of vessel v

x_v: Vessel random effect

$z_{v,t}$: Trip random effect

$\mu_{v,t}$: Expected catch rate for trip t of vessel v

$y_{v,t,s}$: Catch for set on trip t of vessel v

TABLE S2: model Equations and parameters of the monitoring module

Equation	Description	Parameters
$\phi = \text{logit}(p_{review})$	Logit of review probability per set	p: proportion of sets reviewed (e.g., 0.2). Applied to a sub-sample of sets that are monitored.
$\phi_v = \text{logit}(p_{cover})$	Logit of base monitoring probability per vessel. Likewise for trips.	p_{cover}: Proportion of vessels/trips that are covered by monitoring.
$M_{v,t,s} \sim \text{Bernoulli}(p_{review})$	Monitoring indicator for random sampling across sets within the vessels/ trips/sets with monitoring coverage	$M_{v,t,s}$: 1 if set monitored, 0 otherwise
$\phi_v = \text{logit}(p_{cover}) + \text{bias}_{vessels} \cdot x_v$	Logit of monitoring probability for vessel-based sampling, with bias	$\text{bias}_{vessels}$: vessel bias factor, x_v: vessel random effect
$V_v \sim \text{Bernoulli}(\text{logit}^{-1}(\phi_v))$	Monitoring indicator for vessel-based sampling	V_v: 1 if vessel monitored, 0 otherwise
$\phi_t = \text{logit}(p_{cover}) + \text{bias}_{trip} \cdot z_{v,t}$	Logit of monitoring probability for trip-based sampling, with bias	bias_{trip}: trip bias factor, $z_{v,t}$: trip random effect
$T_t \sim \text{Bernoulli}(\text{logit}^{-1}(\phi_t))$	Monitoring indicator for trip-based sampling	T_t: 1 if trip monitored, 0 otherwise

Notes:

$M_{v,t,s}$: Monitoring indicator for set s on trip t of vessel v (1 = monitored, 0 = not monitored)

p_{cover}: Proportion of sets to be monitoring coverage

p_{review}: Proportion of sets to be reviewed for data.

$\text{bias}_{vessels}$, bias_{trip}: Bias factors for vessel/trip selection (higher values = stronger bias towards lower catch rates)

x_v: Vessel-level random effect (from catch model)

$z_{v,t}$: Trip-level random effect (from catch model)

TABLE S3: fixed parameter values. Fleet specific parameters was based on values from Brown et al. (2021).

Parameter	Value	Description
V	50	Number of vessels in the simulated fleet, representative of medium-sized longline fleet
μ^{trips}	10	Mean number of trips per vessel per year, based on typical longline operations
θ^{trips}	1.13	Dispersion parameter for trips per vessel, allows realistic variation in fishing effort
μ^{sets}	26	Mean number of sets per trip, typical for longline tuna fishing
θ^{sets}	1.8	Dispersion parameter for sets per trip, reflects operational variation
θ^{catch}	0.42	Catch dispersion parameter, controls overdispersion in catch counts
p_{cover}	0.2/0.3	Proportion of sets covered by monitoring (e.g. 20%, 30%), as specified by MSC requirements
p_{review}	0.2/1.0	Proportion of sets that are reviewed for data within those that are monitored.

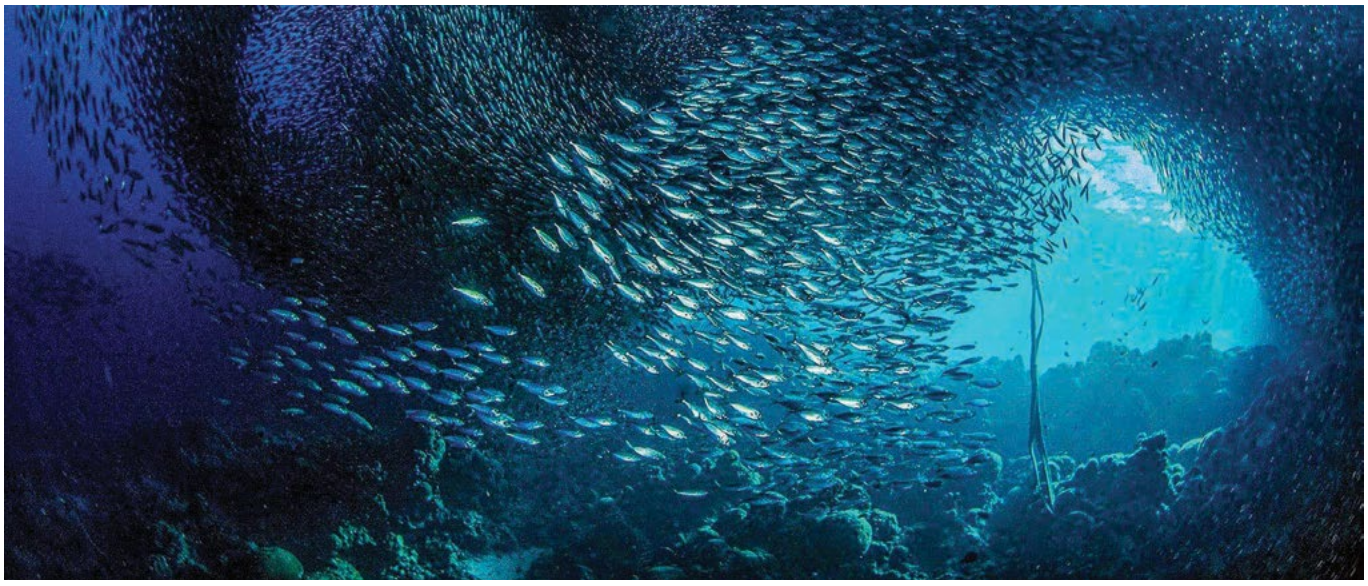


TABLE S4: parameters for monitoring scenarios.

Monitoring Set	Description	Coverage rate (p_monitor)	Vessel Bias (bias_v)	Trip bias (bias_factor)	Sets selection rate (p_sets_select)
1	Baseline	1.0 (100%)	0	0	0.2 (20%)
2	Vessel bias	1.0 (100%)	-2	0	0.2 (20%)
3	Trip bias	1.0 (100%)	0	-2	0.2 (20%)
4	Baseline 20%	0.2 (20%)	0	0	0.2 (20%)
5	Vessel bias 20%	0.2 (20%)	-2	0	0.2 (20%)
6	Trip bias 20%	0.2 (20%)	0	-2	0.2 (20%)
7	Baseline 75%	0.2 (20%)	0	0	0.75 (75%)
8	Vessel bias 75%	0.2 (20%)	-2	0	0.75 (75%)
9	Trip bias 75%	0.2 (20%)	0	-2	0.75 (75%)

TABLE S5: species-specific parameter values for catch simulations.

Species Set	Species Type	β_0	σ^x	σ^z	θ^{catch}
1	Market species	1.7	0.41	0.67	0.42
2	Market species + high vessel variance	1.7	0.82	0.67	0.42
3	Market species + high trip variance	1.7	0.41	1.3	0.42
4	Bycatch species	-2.52	0.55	0.65	0.42
5	Rare bycatch species	-4.9	0.55	0.65	0.42

Advanced methods references

Brown, Christopher J, Amelia Desbiens, Max D Campbell, Edward T Game, Eric Gilman, Richard J Hamilton, Craig Heberer, David Itano, and Kydd Pollock. 2021. "Electronic Monitoring for Improved Accountability in Western Pacific Tuna Longline Fisheries." *Marine Policy* 132: 104664.

Gilman, Eric, Milani Chaloupka, Andrew Read, Paul Dalzell, Jörg Holetschek, and Corrie Curtice. 2012. "Hawaii Longline Tuna Fishery Temporal Trends in Standardized Catch Rates and Length Distributions and Effects on Pelagic and Seamount Ecosystems." *Aquatic Conservation: Marine and Freshwater Ecosystems* 22 (4): 446–88. https://doi.org/https://doi.org/10.1002/aqc.2237.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Roberson, Leslie A, Christopher J Brown, Carissa J Klein, Edward T Game, and Chris Wilcox. 2025. "Opportunity to Leverage Tactics Used by Skilled Fishers to Address Persistent Bycatch Challenges." *Fish and Fisheries* 26 (2): 193–202.

The Nature Conservancy

gema
Global Electronic Monitoring Accelerator

UNIVERSITY of TASMANIA
IMAS
Institute for Marine and Antarctic Studies