Center for Independent Experts (CIE) independent peer review of a predictive model of discarded catch that leverages self-reporting and electronic monitoring on commercial fishing vessels

Dr Coby Needle

Aberdeen, Scotland

November 2021

Executive Summary

The method proposed is valid, sufficiently justified, and appropriate for the purpose of using electronic monitoring to correct bias in self-reported estimates of commercial discards. The statistical approach used is of high quality and makes full use of the best scientific information available. The report presented would benefit from improved treatment and explanation of the modelling approaches, but this does not detract from the quality of the analysis itself.

Background

The aim of this paper is to present a modelling framework that combines electronic monitoring (EM; i.e. human viewing of discard footage) with self-reporting to attempt to correct for biases in the latter, and thereby generate accurate estimates of discards. The main idea here is to use EM to essentially ground-truth and calibrate self-reporting estimates - the principal focus of the discard estimation would still be the self-reported estimates. The data used were from trips with both EM and logbook reports.

All fish recorded on camera were positioned singly on a measuring strip under a camera for subsequent review, and the EM estimates are assumed to be the truth against which logbook estimates are to be compared. Studies in Europe have shown that skipper estimates of discards are very often significant underestimates, so some form of correction would certainly be needed – particularly as the logbook entries are by NMFS statistical area, rather than by haul. The report also determines (through simulation modelling) the rate of review that would be required to generate estimates with specified levels of precision.

Modelling is by generalised linear mixed models (GLMMs), with several different versions presented: the gamma-hurdle (or delta) model; the zero-inflated Poisson model (which is analogous to the Poisson-link model in Thorsen (2018)); and a simplified Poisson model.

Description of the Individual Reviewer's Role in the Review Activities

To conduct this review, I have:

1. Attended an online pre-review meeting to raise any questions required to the report's lead author, and to discuss further both the modelling and the review process.
2. Carefully read four background documents.
3. Read and made detailed notes on the principal review document (Lindon 2021).
4. Written the review report.
5. Responded to Steering Committee comments on my first review report.

Summary of findings for each ToR

1. **Do the statistical analyses address the objectives of providing reasonably accurate and precise estimation of groundfish discards on EM vessels?**

1a. Are the statistical models adequately described?

I conclude that the statistical models are adequately described for an experienced reader, but probably not for the intended audience. The model notation assumes a lot of statistical modelling experience on the part of the reader, and in most cases is not sufficiently well explained. The model descriptions would only really make sense to an experienced reader, and in particular one who is familiar with the supplied background papers. As the intended audience is policy colleagues (and it will also be made available for the public), the paper is probably too difficult for many readers, and even an experienced reader may find themselves questioning exactly what notation is being used for what. If the models are to be presented, I would suggest that each term is defined as it is introduced (as would be standard in a published peer-reviewed paper). So, for example, it would be helpful to highlight that the first two equations on page 4 refer to "encounter probability", while the second two equations express "biomass given encounter". Alternatively, the equations and detailed model descriptions could be moved to an appendix, as the intended audience is unlikely to be interested in them. There also needs to be a table summarising all the notation used in the paper.

In Table 3, the existence of two different ways of calculating R^2 is introduced without any discussion. Having experience in this area, and having read the background papers, I know what the difference is, but the reader might not. I would suggest that a short note is included (with references) to indicate the rationale for this choice of goodness-of-fit metric.

1b. Do the statistical models adequately fit the data?

The goodness-of-fit of the three models presented are summarised in Table 3, as well as in Supplement 1, and cover both Bayesian and Leave-One-Out (LOO) R^2 estimates as well as error (with confidence intervals). This is a very comprehensive approach that allows the reader to determine which of the models performs the best. The more familiar gamma hurdle model is shown to have a relatively low explanatory power by both R^2 measures, and in addition shows a higher error. Both the ZIP (zero-inflated Poisson) and simpler Poisson models have higher explanatory power and lower error, although the author notes that the simpler Poisson model did not accommodate zero-inflation very well. Given that the main aim of using these models is to deal with many zero observations, one could argue against the use of the simple Poisson model (even though it does seem to fit better)? On balance, though, I think the decision to continue with the simpler Poisson model on the grounds of predictive power and simpler structure is the correct one. Parsimony and a good fit to the data are the key issues here, which outweigh a potential problem around zero inflation.

There do remain some points regarding the data which I think could be clarified, however, and which would help the reader understand the details of the model fits.

- Why did review rates decline so much in 2019? I don't think this is stated anywhere.
- The author mentions that white hake could be confused with other hake species on camera, and so were not included in the analysis. Does this mean that all hake estimates must be suspect? If so, then this should be noted.

1c. Is the complexity of the statistical models justified?

Three models are presented in the paper with different levels of complexity.

The gamma hurdle model uses a two-phase approach, with the likelihood of encounter modelled separately from the expected biomass given encounter. In itself this approach leads to a model which could be considered to be more complicated than necessary, in terms of analysis steps and the number of parameters to estimate, and also (as outlined by Thorson, 2018) makes parameter estimates difficult to interpret. The zero-inflated Poisson (ZIP) model is structurally similar, with a two-stage process, although in this case a different assumption about the probability structure is used. Given that it is a similar approach, my conclusions about the gamma hurdle model apply equally well to the ZIP model. Finally, the simpler Poisson model involves a single analysis step with fewer parameters. Parsimony suggests that a simpler model will intrinsically be preferable to a complicated model if it describes the data equally well or better, and the paper shows that this is the case here. Hence, I would conclude that the level of complexity is appropriate to interpret what is a detailed dataset.

This is maybe a naïve point, but is there any particular reason why Bayesian estimation is preferred here to frequentist approaches? There may be valid justifications, but they are not stated and a frequentist option does not seem to have been considered. My issues with Bayesian analysis are that: a) there is a danger that the posterior probability distribution of a particular parameter is dependent (in some way) on the chosen prior distribution, which may itself be *ad hoc*, and b) that parameter estimation can take a long time, which in some circumstances can prove impractical. The author notes that priors were deliberately vague for this study, which helps with a), but in this case the point of Bayesian analysis (which is to use prior knowledge where available) is rather lost. If the author is more comfortable with Bayesian analysis, then that is fine, but a short note explaining the choice of estimation approach would be helpful.

1d. Are there alternative models that would be preferable in terms of estimation performance and computational efficiency?

From the background papers and my own experience, I have concluded that the model presented in the paper is appropriate in terms of both estimation performance and computational efficiency, and (as I have mentioned above) I therefore do not think alternative models would be preferable – what is presented is good. However, I do have a few comments to make about alternative sources of data, or different ways to use existing data, rather than about the models *per se*.

In addition to EM and VTR logbooks, a third source of data would be onboard observers. Where available, were these considered for the analysis, and do you think they would give a different estimate to either EM or VTR logbooks? This aspect is mentioned briefly in the "Data" section, but an additional paragraph or two of explanation would be helpful.

Why weren't haul-by-haul logbook entries requested for this project? It may be that this would have led to more accurate skipper recording – there could be several hauls within an NMFS statistical area, making it harder for skippers to remember discards at each (and thus the total for the area).

2. Have the sources of uncertainty and caveats in the analyses been adequately described? Given the objectives and the performance of presented models, are there additional potential sources of uncertainty that can be quantified and should be incorporated?

In general, I did find that the sources of uncertainty and caveats in the analysis were well described. However, for me, there remain some elements of the models and of the data used that could still bear improved explanation.

The paper proposes the use of the simpler Poisson model as the most appropriate for analysing the given data. However, there remains a potential issue with this model, in that it may not account for zero-inflation very well. That model does fit the data better (in terms of both error and explanatory power) than the other two suggested models, so it may not be a significant issue, but I wonder if this remains a problem given that the main purpose of the modelling is to address the bias in the data towards zero observations. I concluded above that the choice was probably reasonable, but it might be worth further consideration in future analyses.

Is there a potential bias involved in the self-selection of vessels for the trial? As the vessels were all volunteers, would there be something unusual about them that might make them non-representative of the fleet as a whole? Were there inducements offered to encourage participation? The potential issue with a "reference fleet", used to infer the behaviour and characteristics of the fleet as a whole, is that it may not be representative of general activity, and this either needs to be accommodated or at least acknowledged in the analysis if seen as a problem. A short analysis or note covering this issue would increase confidence in the robustness of the overall modelling approach, in my opinion.

The presentation of a single fish at a time to the video camera is an obvious limitation of the approach. In Scottish trials, we have seen that fish lying on discard belts often occlude each other, or can be bent, or may not fit into the frame, and these issues all limit to current utility of EM for evaluating discards (see Needle et al 2015, French et al 2019). Was this considered to be a problem for the long-term future of this type of monitoring?

Similarly, the use of human viewers is another limitation of the current approach. Various approaches to automated image analysis are being trialled around the world (French et al 2019) – were any of these systems considered for use in this project? It would be interesting to compare the bias and variance characteristics of human vs automated EM image analysis, and determine what proportion of video would need to be analysed using the latter for an equivalent precision, and I would recommend this as a suitable topic for a follow-up analysis.

Overall Conclusion

We have been asked to consider whether the reduced bias from the modelling is worth it, given the potential increase in uncertainty. I feel that it certainly is worth it, as skippers' logbook estimates of discard rates are notoriously prone to bias and correction of the type outlined in this paper is essential. I have noted some points that could be reconsidered above (along with some more minor editorial points below), but overall I feel that **the method proposed appears both valid and appropriate for the purpose**. In my view, the paper itself is lacking a lot of required detail and almost feels like an early draft, but that doesn't detract from the modelling. Having said that, there may be specificities around the vessels and species used for this analysis, and I think it would be helpful to consider further applications of this method to other fishery situations before concluding that the approach is appropriate for every circumstance.

Looking at the problem more widely, I do think that the models presented here are of the appropriate type, given that there will always be a bias towards zero in the encounter data due to the generally-patchy nature of fish distribution. The more basic question of whether using EM data to ground-truth and supplement skipper-based VTR data is an interesting one. I would argue that VTR estimates of discards are always likely to be flawed (and, specifically, probably underestimates), given the nature of discarding (no one really counts their garbage). The paper has shown that it is possible to use EM to correct for this bias, but the corrections are to a certain extent model dependent and it could be questioned whether they get us nearer to the actual truth. An alternative approach would be to consider EM estimates of discarding as the main source of data, either alone or in conjunction with scientific observer data. While there may be uncertainties around both of these, they should not be subject to the same biases as VTR data. There remain significant problems with EM, however, in terms of the resources required to implement and operate the system, but (as I mentioned above) ongoing developments with automated image analysis may help with this in the near future. Overall, I feel that the solution suggested in the paper is a valid one for the data available, but that a more holistic EM-and-observer approach may be more beneficial in the long run.

I should also note that I very much appreciated the inclusion of four background papers to assist with my review. While this extended the length of time required, they served as excellent revision to the subject and I feel my review is more relevant in consequence.

Minor points

I would also like to include some more minor points as part of my review. These do not fit easily into the ToR structure above, and some reflect a personal preference, but they might improve the final paper.

- p. 1, last para: "100% video surveillance of vessel fishing effort"– do you mean effort *per se* here, or actual discard activity?
- p.2, 2nd para: I find "leverage" an odd word to use here (and in the paper title) – I usually take it to mean "use (something) to maximum advantage", and that is not really the task here. Another word might be "supplement"?
- I might have preferred the figures to be floats within the document (I'm assuming the report has been done in LaTeX!), rather than all towards to end – there needs to be quite a lot of scrolling as it is.
- Figure 1 takes quite a lot of work on the part of the reader, and some of the lines overlap making interpretation difficult. The only thing lost if we didn't have Figure 1 would be the time-series element which I'm not sure is all that informative anyway? Also, I might prefer to see Figures 2 and 3 split out by fishing year – this would make for many more plots, but again it is currently difficult to tell one year from another.
- The delta-model is referenced (Thorsen 2018), but the other two are not. Without working through the notation, the other two models look analogous to the Poisson-link models also proposed by Thorsen (2018) – is this correct? I assume so given that this paper was included as background reading. If so, then this should be cited – if the source is something else, then this also should be cited (or if the approach is novel, this should be noted).
- In the caption for Figures 7-9, it would be helpful to note that "rate" refers to the subsampling rate of trips, as this is not immediately clear.
- The appendix is useful and should be retained, but I would suggest there needs to be some more explanatory introduction text. In most cases the variable names are obvious, but not in all, and it would also be helpful to include a notation table showing how a variable is notated in both the main paper and the appendix.

References

French, G., Mackiewicz, M., Fisher, M., Holah, H., Kilburn, R., Campbell, N. and Needle, C. (2019). Deep neural networks for analysis of fisheries surveillance video and automated monitoring of fish discards, *ICES Journal of Marine Science* **77**(4): 1340–1353.

Needle, C. L., Dinsdale, R., Buch, T. B., Catarino, R. M. D., Drewery, J. and Butler, N. (2015). Scottish science applications of Remote Electronic Monitoring, *ICES Journal of Marine Science* **72**(4): 1214–1229. DOI: 10.1093/icesjms/fsu225.

Appendix 1: Bibliography of materials provided for the review

Main review document

Linden, D.W. 2021. A predictive model of discarded catch that leverages self-reporting and electronic monitoring on commercial fishing vessels. Working Paper

Other background material

Gelman, Andrew, Ben Goodrich, Jonah Gabry, and Aki Vehtari. 2019. "R-Squared for Bayesian Regression Models." The American Statistician 73: 307–9.

New England Fishery Management Council. 2021. Northeast Multispecies Fishery Management Plan – Amendment 23. Including a Final Environmental Impact Statement and Initial Regulatory Flexibility Analysis. August 2021 Vols 1 and 2. https://www.nefmc.org/library/amendment-23

Thorson, J. T. (2018). Three problems with the conventional delta-model for biomass sampling data, and a computationally efficient alternative. Canadian Journal of Fisheries and Aquatic Sciences, 75(9), 1369-1382.

Vehtari, Aki, Andrew Gelman, and Jonah Gabry. 2017. "Practical Bayesian Model Evaluation Using Leave-One-Out Cross-Validation and WAIC." Statistics and Computing 27 (5): 1413–32

Venables, W. N., & Dichmont, C. M. (2004). GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research. Fisheries research, 70(2-3), 319-337

Appendix 2: CIE Performance Work Statement

Performance Work Statement

National Oceanic and Atmospheric Administration (NOAA)

National Marine Fisheries Service (NMFS)

Center for Independent Experts (CIE) Program

External Independent Peer Review

A predictive model of discarded catch that leverages self-reporting and electronic monitoring on commercial fishing vessels

Background

The National Marine Fisheries Service (NMFS) is mandated by the Magnuson-Stevens Fishery Conservation and Management Act, Endangered Species Act, and Marine Mammal Protection Act to conserve, protect, and manage our nation's marine living resources based upon the best scientific information available (BSIA). NMFS science products, including scientific advice, are often controversial and may require timely scientific peer reviews that are strictly independent of all outside influences. A formal external process for independent expert reviews of the agency's scientific products and programs ensures their credibility. Therefore, external scientific peer reviews have been and continue to be essential to strengthening scientific quality assurance for fishery conservation and management actions.

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards[1]. Further information on the Center for Independent Experts (CIE) program may be obtained from www.ciereviews.org.

Scope

Estimation of total catch in a commercial fishery is important for stock assessment and effective management decision-making, as dictated by the Magnuson-Stevens Act. A primary challenge of catch estimation lies in the largely unknown portion of catch that is discarded at sea rather than retained. Catch can be discarded for multiple reasons related to regulatory compliance (e.g., species and/or size retention prohibited) and non-compliance (e.g., illegal actions incentivized by quota limitations). While the retained portion is recorded by multiple information streams, discard estimation typically relies on a sampling of trips that carry a human at-sea observer. Low sampling rates may still achieve adequate precision (e.g., coefficients of variation < 30%) but rely on a random selection to ensure a

[1] https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2005/m05-03.pdf

representative sample, as with any survey design. In the absence of representative samples, estimators of discarded catch will be biased.

Amendment 23 of the Northeast Multispecies (i.e., groundfish) Fishery Management Plan (FMP) seeks to improve the reliability and accountability of catch reporting in the commercial groundfish industry. Justification for the amendment came from evidence that current monitoring approaches fail to adequately ensure compliance with regulations and may incentivize behavior that degrades the accuracy of catch estimation; evidence included differences in vessel fishing activity related to the presence/absence of a human observer on a given trip (i.e., observer bias). The proposed solutions involve increased and/or enhanced monitoring that maximizes the value of collected information while minimizing costs to the industry and NMFS. Electronic monitoring (EM) systems provide opportunities for unbiased estimation of catch by combining 100% video surveillance of vessel fishing effort with a random selection of video reviewing, removing the opportunity for observer bias. Pilot programs in the Northeast have suggested that consistent catch handling and self-reporting by vessels can allow for low (~20%) video review rates, enabling a framework of monitoring that achieves the intended goals and objectives of Amendment 23.

Here, we illustrate a modeling approach that leverages Vessel Trip Reports (VTRs) with EM video review to generate accurate estimation of groundfish discards on EM vessels. We fit and compared multiple generalized linear mixed and delta/hurdle models to estimate the relationship between EM-reviewed discards and VTR-reported discards, E(EM) ~ f(VTR + …). The models accommodated nuances of the data (e.g., zero inflation) and variation according to groundfish species, vessel/trip attributes, and changes in data availability during the fishing season. We used data from 2017–2021 on EM trips that were fully reviewed including 31 vessels, >1,100 sub-trips, and >3,500 hauls. The data were subset to explore review rates ranging towards 1% and to identify thresholds of reasonable precision, which depended on the model structure used. This modeling framework will be used for groundfish EM vessels during FY2021 and beyond.

Given the implications of this new monitoring approach, it is important that the methods represent the best available science and are statistically sound. Therefore, the CIE reviewers will conduct a peer review of the statistical modeling based on the Terms of Reference (TORs) referenced below. Given the public interest, it will be important for NMFS to have a transparent and independent review process of the model used in this assessment.

The specified format and contents of the individual peer review reports are found in Annex 1. The Terms of Reference (TORs) of the peer review are listed in Annex 2.

Requirements

NMFS requires three reviewers to conduct an impartial and independent peer review in accordance with this Performance Work Statement (PWS), OMB Guidelines, and the ToRs below. The reviewers shall have working knowledge and recent experience in **statistical modeling, with applications to fisheries management and/or quantitative ecology**. Each CIE reviewer's duties shall not exceed a maximum of 10 days to complete all work tasks of the peer review described herein.

Tasks for reviewers

Each CIE reviewers shall complete the following tasks in accordance with the PWS and Schedule of Milestones and Deliverables herein.

1. <u>Pre-review Background Documents:</u> Review the following background materials and reports prior to the review:

Thorson, J. T. (2018). Three problems with the conventional delta-model for biomass sampling data, and a computationally efficient alternative. *Canadian Journal of Fisheries and Aquatic Sciences*, 75(9), 1369-1382.

Venables, W. N., & Dichmont, C. M. (2004). GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research. *Fisheries research*, 70(2-3), 319-337.

Approximately, two weeks before the peer review, the NMFS Project Contacts will send by electronic mail or make available at an FTP site to the CIE reviewer all necessary background information and reports for the peer review. In the case where the documents need to be mailed, the NMFS Project Contacts will consult with the CIE on where to send documents. The CIE reviewer shall read all documents in preparation for the peer review.

2. Webinar: Additionally, approximately two weeks prior to the peer review, the CIE reviewers will participate in a webinar with the NMFS Project Contacts and other staff to address any clarifications that the reviewers may have regarding the ToRs or the review process. The NMFS Project Contacts will provide the information for the arrangements for this webinar.

3. Desk Review: Each CIE reviewer shall conduct the independent peer review in accordance with the PWS and ToRs, and shall not serve in any other role unless specified herein. Modifications to the PWS and ToRs cannot be made during the peer review, and any PWS or ToRs modifications prior to the peer review shall be approved by the Contracting Officer's Representative (COR) and the CIE contractor.

4. Contract Deliverables: Independent CIE Peer Review Reports: Each CIE reviewer shall complete an independent peer review report in accordance with the PWS. Each CIE reviewer shall complete the independent peer review according to required format and content as described in Annex 1. Each CIE reviewer shall complete the independent peer review addressing each ToR as described in Annex 2.

Place of Performance

Each CIE reviewer shall conduct an independent peer review as a desk review; therefore, no travel is required.

Period of Performance

The period of performance shall be from the time of award through October 2021. Each reviewer's duties shall not exceed 10 days to complete all required tasks.

Schedule of Milestones and Deliverables

The contractor shall complete the tasks and deliverables in accordance with the following schedule.

Within two weeks of award	Contractor selects and confirms reviewers
Two weeks prior to the review	Contractor provides the pre-review documents to the reviewers.
Within two weeks prior to the review	Reviewers participate in Webinar.

August 2021	Each reviewer conducts an independent peer review as a desk review
Within two weeks after review	Contractor receives draft reports
Within two weeks of receiving draft reports	Contractor submits final reports to the Government

Applicable Performance Standards

The acceptance of the contract deliverables shall be based on three performance standards: (1) The reports shall be completed in accordance with the required formatting and content (2) The reports shall address each ToR as specified (3) The reports shall be delivered as specified in the schedule of milestones and deliverables.

Travel

Since this is a desk review travel is neither required nor authorized for this contract.

Restricted or Limited Use of Data

The contractors may be required to sign and adhere to a non-disclosure agreement.

Project Contacts

Dr. Daniel Linden (lead contact)
NOAA/NMFS/GARFO, Analysis and Program Support Division
55 Great Republic Road
Gloucester, MA 01930
daniel.linden@noaa.gov
978-281-9265

Dr. J. Michael Lanning
NOAA/NMFS/GARFO, Analysis and Program Support Division
55 Great Republic Road
Gloucester, MA 01930
j.michael.lanning@noaa.gov
978-281-9308

Annex 1: Peer Review Report Requirements

1. The report must be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether or not the science reviewed is the best scientific information available.

2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.

3. The reviewer report shall include the following appendices:

a. Appendix 1: Bibliography of materials provided for review

b. Appendix 2: A copy of the CIE Performance Work Statement

Annex 2: Terms of Reference for the Peer Review

The reviewers will provide a scientific peer-review of the following document:

Linden, D.W. 2021. A predictive model of discarded catch that leverages self-reporting and electronic monitoring on commercial fishing vessels. *Working Paper*

The reviewers will provide input on the following questions:

1. Do the statistical analyses address the objectives of providing reasonably accurate and precise estimation of groundfish discards on EM vessels? a. Are the statistical models adequately described?

b. Do the statistical models adequately fit the data?

c. Is the complexity of the statistical models justified?

d. Are there alternative models that would be preferable in terms of estimation performance and computational efficiency?

2. Have the sources of uncertainty and caveats in the analyses been adequately described? Given the objectives and the performance of presented models, are there additional potential sources of uncertainty that can be quantified and should be incorporated?