DTU Library

Workshop on Unavoidable Survey Effort Reduction 2 (WKUSER2)

Anderson, Sean; Bacheler, Nate; Barnett, Lewis; Berg, Casper; Blackheart, Kristian; Bolser, Derek; Börjesson, Patrik; Bryan, Meaghan; Cariou, Thibault; Chaves, Corina

Total number of authors:
46

Link to article, DOI:
10.17895/ices.pub.22086845

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

WORKSHOP ON UNAVOIDABLE SURVEY EFFORT REDUCTION 2 (WKUSER2)

VOLUME 5 | ISSUE 13

ICES SCIENTIFIC REPORTS

RAPPORTS SCIENTIFIQUES DU CIEM

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

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

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

ISSN number: 2618-1371

ICES Scientific Reports

Volume 5 | Issue 13

WORKSHOP ON UNAVOIDABLE SURVEY EFFORT REDUCTION 2 (WKUSER2)

Recommended format for purpose of citation:

ICES. 2023. Workshop on Unavoidable Survey Effort Reduction 2 (WKUSER2).
ICES Scientific Reports. 5:13. 115 pp. https://doi.org/10.17895/ices.pub.22086845

Editors

Hans Gerritsen • Stan Kotwicki • Kotaro Ono

Authors

Sean Anderson • Nate Bacheler • Lewis Barnett • Casper Berg • Kristan Blackhart • Derek Bolser
Patrik Börjesson • Meaghan Bryan • Thibault Cariou • Corina Chaves • Elizabeth Chilton • Jason Conner
Annica De Groote • Lukas DeFilippo • Paul Dolder • Catherine Foley • Hans Gerritsen • Madison Hall
Espen Johnsen • Eoghan Kelly • Stan Kotwicki • Sven Kupschus • Gwladys Lambert • Andy Lipsky
Emily Markowitz • Guillermo Martin • Angelia Miller • Coilin Minto • Bríd O'Connor • Kotaro Ono
Zack Oyafuso • Iosu Paradinas • Maria Grazia Pennino • Elizabeth Phillips • John Gabriel Ramirez
Yves Reecht • Paul Regular • Margaret Siple • Matt Siskey • David Stokes • Anna Stroh
James T Thorson • Ralf van Hal • Joel Vigneau • Daniel Vilas • Kai Wieland • Semra Yalcin

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

Contents

i Executive summary

The Workshop on Unavoidable Survey Effort Reduction 2 (WKUSER2) focused on best-available approaches that can minimize information loss and ensure continuity in survey time series when unavoidable changes to survey effort occur. WKUSER2 recognised that reductions, reallocations, or increases in survey effort present similar set of problems, and therefore concentrated on all aspects of survey effort changes. The workshop reviewed available research, current practices, and recommended future directions on four key topics: (i) key elements of flexibility of a survey, (ii) why and how to combine data from different sources (e.g. surveys, fishery sampling) and deal with survey gaps, (iii) how to configure estimation and simulation models, and (iv) review existing tools and technology to evaluate consequences of survey effort changes.

Road maps were developed for the key topic areas i, ii, iii, and iv, whenener possible, to assist scientists and survey managers in making decisions on how to evaluate and mitigate the impact of survey effort changes on data and advice quality. Many tools are available or are being developed for that purpose, but the group recognized two important needs during the workshop: i) defining clear objectives and priorities of a survey, which are essential to properly evaluate consequences of survey changes; and ii) making all tools accessible, reproducible, and transparent to benefit the whole community. This requires organisational and cultural shift to create support systems that ensure the development and sustainability of such tools in the future.

ii Expert group information

Expert group name	Workshop on Unavoidable Survey Effort Reduction 2 (WKUSER2)
Expert group cycle	Annual
Year cycle started	2022
Reporting year in cycle	1/1
Chair(s)	Stan Kotwicki, United States
	Hans Gerritsen, Ireland
	Kotaro Ono, Norway
Meeting venue(s) and dates	13-17 September 2022, Galway, Ireland

1 Introduction

1.1 Unavoidable Survey Effort Reduction

Fisheries-independent surveys of marine resources (hereafter referred to as surveys) are an important source of the information used in fish stock and ecosystem assessments, and fisheries research. Surveys are often the only source of information providing population abundance estimates, spatial distribution, and population structure data needed for sustainable management of marine resources. In addition, environmental data collected during surveys are used to develop ecosystem indicators for assessment of the ecosystem health as well as forecasting potential changes to marine resources in response to climate change.

Surveys follow standardized and repeatable scientific protocols that assure the continuity of both survey design and the data collected. Nevertheless, most survey programs have to make substantial changes to survey operations from time to time due to challenges such as reductions in staff and budgets, bad weather and vessel breakdowns / repairs / unavailability. Other challenges are associated with changes in the environment such as ocean warming and shifts in species distributions. Lastly, human activities such as commercial fisheries and wind energy developments can also affect surveys by restricting access to sampling locations. These challenges often result in effort reductions, which typically compromise the long-term objectives of survey series in terms of accuracy, precision, and consistency of population estimation.

Failure to adapt to changing needs is the biggest challenge to survey programmes. Not adapting will mean obsolescence: the value of the survey is gradually reduced to the point where it is not worth continuing to collect the data. Defining best practice beyond broad generalizations is challenging because of the differences in the survey objectives, severity of impacts on survey design in different areas for different resources. However, changes often need to be implemented in a short time-frame, leaving little time for planning and quantitative evaluation, so there is a need to develop methods and tools that provide a better understanding of the risks and allow evaluation of different mitigation options when changes to the survey effort are made.

1.2 Challenges in making changes to survey effort

The overarching challenge when dealing with survey effort reductions is the understanding of the consequences in terms of precision and bias associated with survey sampling processes, which includes operational, environmental & biological processes. The first ICES Workshop on Unavoidable Survey Effort Reduction (WKUSER, 2020) focused on the following key issues:

- How to choose the survey data products to focus on? (e.g. abundance, composition)
- How to reduce the effort? (e.g. sampling density, spatial coverage, survey frequency)
- How to measure consequences of change in sampling effort?
- How to assess value or loss of information?
- How to minimize loss of information and optimize survey effort?
- How to assure continuation of the standardized time-series?
- How to assure constant catchability or, how to deal with variable catchability?
- What is the minimum survey effort required to provide useful information?
- How to weigh environmental vs biological data when deciding on effort changes?
- How to propagate increased uncertainty from survey to stock assessment outputs?

The outcomes of WKUSER (2020) included: 1. Documentation of current processes and strategies for dealing with survey effort reduction and roadmaps (which are already guiding decisions of survey managers). 2. Better understanding of sampling and "total uncertainty" and a roadmap for research needed to obtain "total uncertainty". 3. Increased knowledge on how to plan for survey calibrations and how to communicate calibration and survey continuity needs to stakeholders. 4. Increased awareness of the tools to help decision making that are currently available (or in development) and their utility in helping the decision-making process and informing the quality of survey deliverables and advisory products.

1.3 The need for the WKUSER2

WKUSER 2020 and the ICES Working Group on Improving Use of Survey Data for Assessment and Advice (WGISDAA) identified issues that needed attention and additional work, which resulted in the proposal for WKUSER2 to be conducted in 2022. Changes to survey operations are a recurring theme in many monitoring agencies, and more coherent planning and a long-term response strategy is desirable. It is important to have a better understanding of the effects of these changes on survey time-series, in particular in relation to stock assessment advice. A clearer understanding is needed of the mitigation measures that can be implemented to minimise the impact of such changes. Because survey scientists / managers often have to make near instantaneous decisions, it is valuable to have a framework or a set of methods that can be applied to the specific problems. This can allow decisions to be made in the absence of data or the opportunity to evaluate options statistically.

WKUSER2 met in Galway, Ireland on 13-17 September 2022 to address the following terms of reference (TORs):

- **TOR 1: Survey design for flexibility.** Review and summarise desired attributes of survey design that allow for flexibility when dealing with unavoidable changes in survey effort and need to expand survey into new areas of species expansion due to changes in the ecosystem.
- **TOR 2: Combining surveys, dealing with data gaps.** Collate advice on methods to combine data from different sources, how to deal with data gaps and how to perform survey calibrations.
- **TOR 3: Modelling and simulations.** Further develop model-based estimation, model validation through simulations, use of auxiliary information to improve survey data products, **including appropriate propagation of uncertainty.**
- **TOR 4: Tools and technology development.** Describe the development of methods that aim to provide quantitative decision-making tools that describe the effects on the quality of the survey deliverables and ultimately advisory products.

1.4 How does WKUSER2 fit into the EOSG and ICES

The ICES Ecosystem Observation Steering Group (EOSG) is responsible for guiding and supporting fisheries-independent data collection, i.e. all at-sea scientific surveys collecting data on marine resources and the ecosystem using research oceanographic vessels. The mandate given to EOSG is to ensure the various data requests from assessment and ecosystem scientists are met with effectively coordinated, integrated, quality-assured products from cost-effective monitoring in the ICES region and beyond.

ICES has recently released three key documents: the Science, Strategic, and Advisory plans. From these documents the priorities for EOSG can be given as follows:

1. Assessing quality, reproducibility, and transparency;
2. Incorporating innovation, future of scientific survey means, and tools; and,
3. Identifying needs for the ecosystem approach, particularly focusing on regionalization.

In 2022, two key workshops are planned under EOS: WKUSER2 and WKPILOT-FIRMOG (Fisheries Independent Regional Monitoring Groups). WKUSER2 has a mandate to progress on: Assessing survey data quality, reproducibility, and transparency, and Incorporating innovation while preparing for the future of scientific survey methodology and tools while operating in reduction of effort. As such, WKUSER2 is an important part of the strategic reflection needed for EOSG and ICES with accompanying necessary innovation and adaptation to a changing environment (in both ecological and societal terms).

WKPILOT-FIRMOG is planned for November 2022 and is focused on making better use of the collective research survey data within a region. This workgroup will investigate the workability of a regional group on fisheries independent data. Since the topic is strongly related to optimization of the use of the surveys at the regional level, participation in this workshop was promoted to the WKUSER2 participants. WKPILOT-FIRMOG directly links to point 3 of ICES strategy, i.e. identifying needs for the ecosystem, approach, and focusing on regionalization.

1.5 Conduct of the workshop

WKUSER2 took place in Galway, Ireland on 13-17 September 2022 and was attended by 47 people from 12 countries (USA, Ireland, Norway, Canada, Sweden, Denmark, Spain, UK, Italy, France, Portugal and The Netherlands); 39 people attended in person and 8 remotely.

The agenda of the meeting is provided in Annex 3. The first day of the meeting consisted of six 30-minute keynote presentations as well as discussions on the challenges and priorities of the workshop. The second day consisted of twenty-one 15-minute presentations related to the ToRs of the workshop (section 4 collated all abstracts from the talks). The presentations were intended to provide a basis for the discussions in the subgroups. The remaining 2½ days consisted mainly of subgroup work: agreeing on outputs and report writing. Each of the ToRs was addressed by a separate subgroup and each subgroup had between seven and twelve members. Each subgroup had two or three subgroup leaders. Subgroups reported in plenary to the group on the fourth day of the workshop.

On the morning of the fourth day of the workshop, one of the participants tested positive for COVID-19. The workshop chairs decided to conduct the remainder of the meeting remotely. Some participants attended from their hotel rooms, others gathered in hotel lobbies, cafes and outdoor terraces. Unfortunately, at least 8 participants contracted COVID-19.

WKUSER2 taught us a few lessons (summarised in section 3.2) that future hybrid workshops (e.g. WKUSER3) could benefit from.

2 TOR Group Reports

The following summarises the results of each TOR group addressing each term of reference.

2.1 TOR I. "Survey design for flexibility": Review and summarize desired attributes of survey design that allow for flexibility when dealing with changes in survey effort and need to expand survey into new areas of species expansion due to changes in the ecosystem.

Subgroup leads: Lewis Barnett, Kai Wieland

Participants: Joel Vigneau, Ralf van Hal, Andy Lipsky, Catherine Foley, Dave Stokes, Jason Conner, Annica de Groote, Philip Politis (remote), Elizabeth Phillips (remote), Daniel Vilas (remote)

While the WKUSER workshop series is intended to address *unexpected* survey effort reduction, perhaps it would be better to consider it as a need to transform unexpected effort reduction into *expected* effort reduction and address it proactively. One solution that meets this need is the development of flexible survey sampling designs and estimators that enable the survey practitioner to react to short or long term reductions in sampling effort.

What do we mean by flexible surveys? We define flexible surveys as: approaches which facilitate multiple robust estimation options to retain the ability to acquire consistent and/or approximately unbiased estimates given change in survey resources, distribution of resources and monitoring access, and observation requirements.

WKUSER1 (ICES, 2020a) introduced the discussion on the need to anticipate unexpected survey effort reduction to cope with urgent or anticipated situations (extreme weather, vessel breakdown, non-accessibility of certain areas, increase of surveyed areas and/or parameters/variables to collect with the same effort, decreased budget, etc.). WKUSER1 then evaluated the consequences of these effort reductions on survey outputs, stock assessments and fisheries management and highlighted the need to consider different strategies for coping with survey changes in time. These strategies included spreading reductions as evenly as possible through the survey effort distribution and increasing flexibility in survey implementation.

WKUSER1 elaborated on the current processes used in dealing with unavoidable change in survey effort. These are:

- Improving survey efficiency, e.g. reducing tow time to certain extent, increase working hours, introducing new technology for easing the staff workload, improving catch processing mainly with more subsampling
- Reducing station density
- Reducing survey frequency

- Changing survey design
- Cancelling surveys as the ultimate solution when no funding is available

WKUSER2 focused on survey design in relation to flexibility. Indeed, it was deemed that survey flexibility is a constant need since the scientist in charge of implementing a survey is dealing, together with the vessel master, with the daily plan to adapt the survey design to *in situ* context. Over a time series of indices, it is very unlikely that all years were monitored exactly with identical design. A common example of consequences for the quality of the indices comes from surveys with stratified designs with a large number of strata and, hence, few fishing operations per strata. Year after year, the risk of having poorly or unsampled strata is high with direct consequences on the robustness of the time series of indices. A key desired attribute of a flexible survey design is one that avoids allocating a small number of samples to a large number of strata. More broadly, the consensus is that survey scientists should carefully consider their survey objectives, required sampling area and habitat, available resources, and statistical assumptions that can best be met by their survey capabilities when designing a survey for flexibility.

2.1.1 Survey flexibility, objectives, assumptions, and design

One of the primary concerns of sampling design is the error of the estimates thereof. Estimation error consists of two sources: bias (i.e. systematic error) and variance (i.e. random error). Bias, that is the degree to which an estimate deviates systematically from its study parameter, may exist for various reasons, two that are relevant for the fishery survey practitioner: 1) design-bias, where sampling violates the assumptions of independent and identically distributed (IID) samples; 2) selection-bias, where sampling diverges from a defined study population. Design-bias may occur in sampling designs such as sampling from a systematic grid, because even though samples are identically distributed, they are not independent. Selection-bias may occur when the population of interest is not confined to the sampling frame of the study (e.g. a study area limited by international borders), resulting in a sample population that may not derive from the same distribution as the study population.

If a survey estimate is design-unbiased, as in simple random sampling with replacement, wherein samples conform to the assumptions of IID, the error of this estimate is equal to its variance. However, it is often infeasible or undesirable to execute a design-unbiased survey, for reasons including: cost efficiency per sampling unit (e.g. vessel fuel cost); management requirements for spatially balanced observations; accessibility of sampling units within a study area (e.g. gear-inaccessible habitat, offshore windfarm or pipeline infrastructure, international borders). Sampling designs that may be biased include systematic sampling, where an estimate will be biased if there are patterns in the study population that covary with the sampling pattern (Gabler and Stenger 2012), and stratified sampling using an allocation algorithm (e.g. Neyman allocation) that prioritises factors other than randomness in samples.

The appropriate statistical design of a survey will depend on the objectives of the study. A number of survey designs can be considered inflexible. For example, simple random sampling is often used when any bias is unacceptable to the study objectives, but logistical inefficiencies make such a design inflexible (e.g., adding sampling units that are long distances away from the natural survey progression). Similarly, systematic sampling, which often enables the most precise point estimates at the risk of some unknown bias, is also inflexible because the full sampling plan must be completed to avoid bias associated with spatially unbalanced sampling (unless the resolution of the sampling universe can be coarsened to cover the entire remaining domain, which is not possible in a fixed-station design). In contrast, a cluster sampling design, which enables design-based evaluation of the variance of an estimate, is a flexible design that can yield more information on variance within and among cluster units, but resulting in fewer primary sampling units (see TOR II for definitions of sampling terminology) and larger variance estimates. By imposing stratification on a study area, it may be possible to improve precision while maintaining unbiased estimates. Stratification may also facilitate the combination of multiple sampling strategies in a hybrid design, however one must ensure that the appropriate estimators are available and employed. Below, we discuss general properties of designs that make them more or less flexible, and how to approach design decisions with statistical robustness.

2.1.1.1 If stratification is to be used for survey sampling designs, the objectives of stratification should be clearly identified.

From Cochran (1977) the potential reasons for stratifying are as follows:

1. If data for certain subdivisions are to be known with specific precision.
2. Certain administrative needs may dictate the use of a subdivision of the population. For example, the agency conducting the sampling may have fieldworkers, each of whom must oversee a portion of the population.
3. Sampling problems may differ for different parts of the population and sampling approach will be appropriate for each situation.
4. Stratification can produce a gain in precision in the estimates of the parameters of the entire population. It may be possible to divide a heterogeneous population into sub-populations, which are themselves more homogeneous. If each stratum is homogeneous, in the sense that the measures vary very little from unit to unit, a precise estimate of each stratum can be obtained from a small sample. These estimates are then combined to obtain an accurate estimate of the entire population.

The number of divisions of the population for sampling purpose will be directly dictated by the answers to the question of the objectives of the stratification. When it comes to stratifying for increasing the precision of the estimates, Cochran expresses that the ideal variable to stratify by is the value of y - the quantity to be measured in the study. If we could stratify by y-value, there will be no overlap between strata, and the within-stratum variance would be smaller than the total variance, especially if there are many strata. In practice, of course, one cannot stratify according to the values of the variable to be measured. But certain cases may approach this situation, and thus provide large gains in precision, if they satisfy the three following conditions:

1. The population is composed of sets that vary greatly in size
2. The main variables to be measured are related to the sizes of the sets
3. A good measure of set size is available to determine strata

These conditions apply perfectly to length sampling in auctions, where all fish are graded by commercial categories, i.e. ranges of individual weights which are directly linked to the length of the fish. In the survey design to estimate marine resources abundance and biomass, the use of trawlable areas linked to a range of bathymetry is often used for stratifying the population. These trawlable areas may obey the three conditions although point 2 may only apply to a subset of targeted species. It is important to note that Cochran (1977) inferred the optimal number of strata in a population and demonstrated that very small reduction of variance is expected over 6 strata.

The advantages of using fewer strata in fisheries independent surveys is commonly agreed (Gavaris and Smith, 1987, Folmer and Pennington, 2000).

2.1.1.2 Survey design should ensure even distribution of samples throughout the surveyed area

This principle, highlighted in WKUSER, is often a non-written objective of survey designs, which sometimes may lead to some aversion to the use of random draw of fishing operations in a large area or an incentive to create a maximum number of strata. We note that while strata are not always defined by spatial boundaries in geographical space, they are typically discussed in this context within this report. Three options to ensure maximum randomness and even distribution of samples were discussed, with more details given in the sections on strategic (2.1.4.1) and tactical (2.1.4.2) pathways to survey flexibility:

- **Systematic sampling**: Division of the surveyed area in N identical surfaces (cells), N being the planned number of samples so that one sample is taken in each cell of the grid.
    - Advantages: Easy to implement, given to be more precise than the simple random sampling, samples evenly distributed throughout the sampling areas.
    - Conditions: Randomness of the design is conditioned to a random draw of the first position in the cell. Then after there are different options but the most commonly used one is to repeat that position in every cell of the grid. It may be as simple as dividing every cell into two (e.g. North / South or East/ West), three or four areas, draw in which area should sampling occur and proceed.
    - Advantages in terms of flexibility: the size of the N′ surfaces to sample (N′ < N for any reason occurring during the survey) is totally flexible and can easily be adapted to the N′ remaining samples to be taken; Differentiated inclusion probabilities between the first samples (1/N) and the remaining (1/N′) would allow statistical inferences without bias.
      If needed to extend the geographical area with the same effort, adapt the size of the grid as necessary.
    - Disadvantages: Not easy to estimate the variance, with the possibility to use the variance of random draw as a proxy. This can only be done if the condition (point above) is respected.

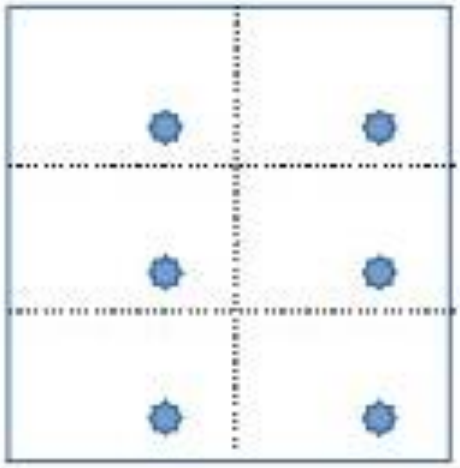


Figure 2.1-1: Systematic survey sampling design illustration

- **Random draw with an exclusion zone** (buffer): Defining a radius around the sample meant to determine a minimal distance between two samples
  - Advantages: Easy to implement, variance well defined, size of the radius ensuring the even distribution of the samples, the radius offers a bit of flexibility to the vessel master in positioning the gear and in tow direction.
  - Conditions: radius size to be determined objectively (e.g. using a predetermined threshold in mean decorrelation distance derived from previously collected data) and documented.
  - Advantages in terms of flexibility: since the samples have been drawn at random, there is the possibility to remove samples at random from the remaining number of primary sampling units at any moment in time, provided that redefined inclusion probabilities are attached to each sample.
    If need to extend the geographical area with the same effort, not much to do except maybe increase the radius around each sample.
    More research needs to be developed in this case on the flexibility potential and the need to avoid systematic effect of samples being collected with spatial ordination (i.e. vessel collecting samples between nearest locations first).
  - Disadvantages: Difficult to implement in narrow strata, e.g., near the continental shelf edge where depth changes significantly over a small area.

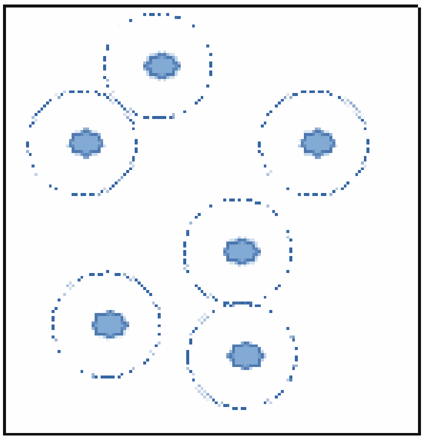


Figure 2.1-2 Random draw with exclusion zone sampling design illustration

- **Mixed systematic and random draw**: Allocate part of your total number of samples to a systematic grid and draw at random the remaining number of samples (see Zhao *et al.* 2018 and Zinger 1980).
    - Advantages: The systematic part comes as the priority samples to be covered; the same variance statistics apply to both (see point above on systematic sampling); when used in a stratified design, the systematic part represents the even part all over the surveyed area and the random draw can be done using optimised allocation (e.g. Neyman allocation)
    - Conditions: as for the systematic sampling, draw at random the first position in the cells and repeat the position for all the systematic parts.
    - Advantage in terms of flexibility: As for the random draw, the extra samples can be removed randomly as required with the reduction of effort.
      If needed to extend the geographical area with the same effort, adapt the size of the grid as necessary.
    - Disadvantages: Same as for both systematic and random sampling.

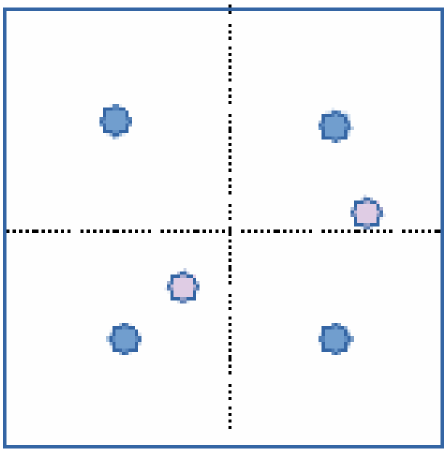


Figure 2.1-3: Mixed systematic and random draw sampling design illustration

In conclusion, the desired attributes of survey design that allows for flexibility is to i) stratify parsimoniously the total survey area and document clearly the reasons for stratifying and ii) use a sampling design with as much probabilistic nature as possible taking the three options detailed above as potential candidates. It is worth noting that some variants of these are said to provide better estimation accuracy and precision (e.g. Liu *et al.* 2009). Finally, documentation of the process for determining a stratification is critically important to making coherent changes to the survey in the future. For example, one common challenge is when habitat or range shifts require an extension of the survey domain, in which case it is necessary to understand and replicate the same design principles from the core survey to these extended areas.

2.1.2 Summary of literature review

A review of the limited literature regarding design and modification of flexible fishery-independent surveys, demonstrated that statistical guidance has generally been consistent with the conclusions of the discussions and presentations of TOR 1 (Madow 1953; Cochran 1977; Gavaris and Smith, 1987; Folmer and Pennington, 2000; Chen et al. 2004; Liu *et al.* 2009; Xu et al. 2015a,b; Zhao *et al.* 2018). However, this guidance is typically not well synthesized within a single text as it bridges the statistical methods with the specific logistics of application to fisheries and ecological monitoring. Perhaps partly as a result of this disconnect, or out of necessity due to the need for in situ decisions, there are many cases where there appears to be a divergence between idealized statistical guidance and practical implementation of survey modifications. We hope that the sections below (particularly 2.1.4) synthesize our knowledge in a framework that is more actionable for survey practitioners and the stock assessment enterprise. First, we highlight a few case studies where surveys were faced with challenges, the mitigation steps taken, and lessons learned.

2.1.2.1 Gulf of Alaska bottom trawl survey

The NOAA Fisheries, Alaska Fisheries Science Center Gulf of Alaska bottom trawl survey has been monitoring demersal living resources of the continental shelf and slope of this productive region since 1984. The survey has historically used a stratified random sampling design, where strata were largely defined by depth. Beginning in 2011 (Raring et al., 2016), survey resources were reduced from three to two chartered commercial vessels, requiring some decision regarding how the fewer vessels and total number of samples would be allocated across the region. Coincidentally, neither of the two vessels had enough wire on their winches to conduct survey hauls at extreme depths, so it was decided that the response would be to drop all sampling from the deepest strata, from 700 to 1000 meters. Eliminating an entire strata is suboptimal because it is simultaneously changing the boundary of the survey domain and the total effort, and limits the interpretation of the time series using only design-based indices. However, it still allows one to use model-based estimators, but those methods must use additional information to predict the response in an unsampled area, which is often difficult to do with high predictive skill. Upon revisiting this survey design to make it flexible to variation in survey effort, Oyafuso et al. (2022) proposed an alternative approach (Oyafuso et al. 2021) to improving efficiency and flexibility by optimizing stratification and sample allocation relative to a constraint on the precision of an estimate (in this case a suite of abundance index CVs across species of interest). They simulated fish densities across the complete spatial domain from a spatiotemporal GLMM conditioned on the full suite of historical data, then simulated sampling with the proposed and existing designs to compare their performance at multiple effort levels. The results showed that there was a trade-off between precision and bias, where most notably the prior design was expected to give more precise estimates but with less accurate estimates of the true precision than under the proposed design. Furthermore, for species that have a significant portion of their habitat on the continental slope, the truncation of sampling at 700m could lead to bias in the scale of design-based abundance estimates. Oyafuso et al. (2022) demonstrated that the design could be truncated at 700m and re-stratified while continuing to provide unbiased design-based estimates of abundance (with improved accuracy of associated uncertainty) with a similar total number of primary sampling units. This was accomplished by implementing a novel approach to sample allocation among strata, which can be reallocated as new data are collected or added flexibility. In summary, to the extent possible one should seek to avoid dropping entire strata to conserve inference to a given spatial footprint, and each survey group should consider preparing for unavoidable survey effort changes by employing a framework/tool to objectively redistribute primary sampling units among strata to support decision making when survey changes are needed.

2.1.2.2 Northeast Atlantic surveys

From about 1999 onwards, demersal trawl surveys in the NE Atlantic area began to coordinate under the International Bottom Trawl Surveys (IBTS) working group at ICES (ICES, 2017). Survey designs and gears were already established in the area by that time and difficulties in standardising these aspects in particular generated much discussion and review. Broadly speaking: the North Sea (NS) Area is a fixed station design with a common survey trawl and ostensibly similar target species and depth range throughout the area. With the exception of the west of Scotland (referred below as Western Area, WA), in the Atlantic area the surveys are stratified random designs. Trawls are often tailored to local, area specific target species, depth ranges and bottom type. Differences between WA survey series in catchability assumptions as well as minimal overlap between surveys meant that it has been problematic to address gaps in survey coverage. This made annual data collection quite inflexible compared to a NS standardised combined index, although evolving assessment methods have gone some way to addressing that.

For this reason, The International Program of Standardised Trawl Surveys (ICES, 2001) (IPROSTS) undertook a review of methods and design between the IBTS WA surveys for Scotland, France and Ireland. Integral was whether calibration coefficients would be required to produce standardised combined indices as these particular surveys utilised similar survey trawls already. The analytical approach was for a series of parallel fishing tows between vessels and it built on the work of (Fryer et al, 2003 and Milar and Fryer, 1999). Results indicated that, where numbers at length for a species appeared in the catch on both paired hauls, reasonable precision was seen above 15-20 paired hauls (Figure 2.1-4).

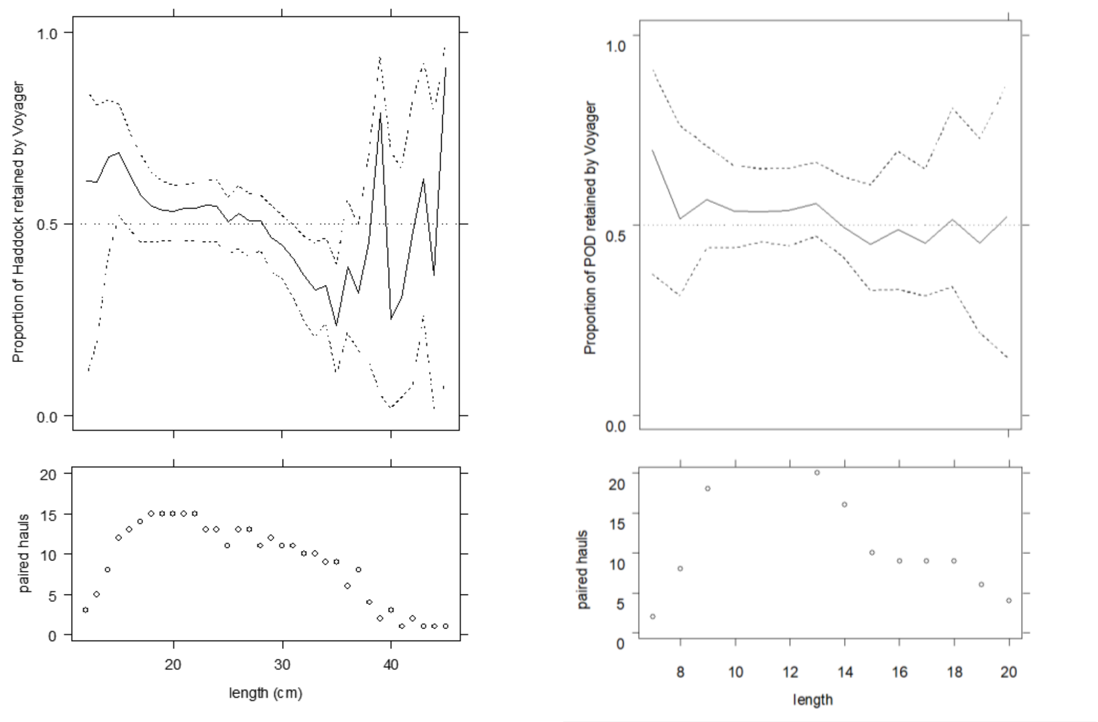


Figure 2.1-4: Results of IPROSTS comparative hauls length frequency analysis for haddock (left) and poor cod (right). Upper panel shows the weighted average back-transformed smoothing curve with 95% confidence intervals. Lower panel gives the number of paired tows available in the analysis. A ratio of 0.5 would indicate that both vessels caught equal amounts for that species at that length, i.e. overall catchability was equal.

One of the first recommendations towards ensuring flexible survey designs for WA surveys therefore came from the IPROST outputs suggesting even moderate datasets of paired haul data could provide two important inputs for data collection programs. One being that trends in one survey series could be useful to predict likely trends in missing data of another survey, should in-survey issues arise. The second use for collecting paired data on an ongoing basis would be to monitor relative changes such that unintentional drift in catchability didn't develop unnoticed over time in either survey. Changes to survey designs were implemented over the proceeding few years to ensure at least a minimal 'corridor' of spatio-temporal overlap between neighbouring surveys annually. Compared to a specific calibration exercise, this would be cost neutral and incrementally build an up to date data set if and when needed.

In planning ahead for the potential need to adjust a survey design 'on the fly', some largely pragmatic design adjustments can be seen in a few NE Atlantic surveys. The Irish Anglerfish & Megrin Survey (IAMS) for example is another random stratified survey coordinated under IBTS (see Kelly et al, 2021). Many random (potential) sampling positions are generated, incrementally the positions with the shortest distance to their nearest neighbour are removed. A rank is assigned to each sampling position in reverse order to the sequence in which they were removed. So, if 1000 potential locations are generated, the pair of stations with the shortest nearest-neighbour distance is identified and one of those stations is removed from the set and it will receive the rank 1000. The nearest neighbour algorithm is repeated, and the next station is removed and allocated the rank 999 etc. Before the survey, a target number of *n* stations is decided and the sampling positions with rank 1 to *n* are selected to be sampled. If early on in the survey it becomes clear that survey effort will be reduced, a lower number of stations can be selected. The point of allocating a ranking to the stations that is related to the distance to their nearest neighbour is that whichever value of *n* is chosen, the sampling positions will always be randomly but evenly distributed in space (Figure 2.1-5). This provides the chief scientist with a theoretically unbiased design for dropping stations where a significant issue arises, while still maintaining spatial coverage.

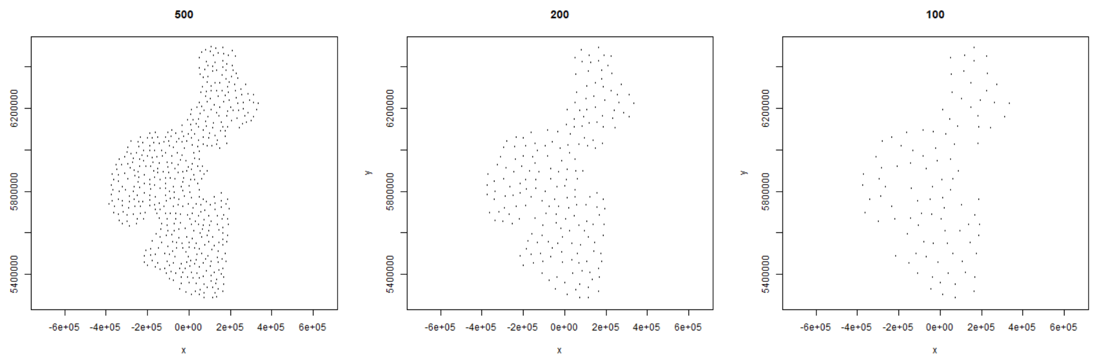


Figure 2.1-5: Distribution of sampling positions when n = 500 stations (left), 200 (middle) and 100 (right) are selected. Regardless of the choice of n stations, they will be evenly distributed in space. This allows changes to be made at short notice without repeating the station selection procedure.

In a recent IAMS survey for example, significant weather events led to the likelihood that large areas or even strata would not be sampled. Prioritising stations based on their rank, was combined with reducing the tow time to the minimum allowed for a valid haul. This ensured the number of 'valid' hauls per day were maintained as high as possible, while also re-allocating some resources from fishing back into transit between stations. This proved very successful in terms of achieving station targets. However, a few questions remain and are being investigated. Firstly, the impact of systematically introducing a ~25-30% reduction in tow time and equivalent increase in 'end effects' (Battaglia et al, 2006; Moriarty et al, 2018). Secondly, if dropping stations of approximately equal 'priority' as you move through the survey area is optimal? Are there gains to be made by optimising hauls in areas of higher variance? A few WA surveys now employ spatio-temporal models to produce relative or absolute indices and have a lot of model diagnostics to hand suggesting where spatial model fits are poorest. When working against the clock, and even against the weather, one can be faced with dropping the next few stations to the east or the west all being of mixed lower priority status. It is tempting then, where priorities over the whole day will not be so clear cut, to review the performance of the indices and surmise whether heading east vs west would bring the greater benefits of precision over the risk of bias.

An additional consideration with random stratified surveys specifically comes back to the catchability assumption. Even with an agile survey design, relative indices will assume relative stability in the distribution of the population being sampled. Invariably trawl efficiency will vary with depth, ground type and so forth (Godø and Engås, 1989; Von Szalay and Sommerton, 2005). One of the drivers for reviewing flexible survey designs derives from possible shifts in spatial distribution of the stock. Where this involves depth or movements into neighbouring surveys with different catchabilities it is also useful to monitor stability in which survey areas are contributing most to the index over time (e.g. discussion in Palma-Pedraza, 2020).

In terms of design-based surveys, the Mackerel and Horse Mackerel Egg Survey is coordinated under ICES across about eight countries and covers the area from the west coast of Norway down to the Gulf of Cadiz in the south. Sampling across the survey area runs from January to July every three years. Covering such a large area means an adaptive strategy has been used in allocating sampling levels along fixed transects. This is to ensure the outer bounds of the egg distribution can be identified and population therefore contained. Recent surveys have encountered some difficulties in ensuring that can still be done as egg distribution has stretched survey resources. Survey protocols have thus evolved to recommend a first pass using every second transect and sample the alternate transects on the return legs (ICES, 2019). The recommendation remains however to not drop two consecutive transects so as to maintain spatial coverage while also minimising data gaps. Again this approach can alleviate moderate reductions in survey effort, but struggles to address the impact of a complete vessel(s) breakdown or significant spatial/temporal shift in the population being sampled.

Given the discussion above, there are risks with current survey designs with possible spatial-temporal shifts in fish populations along with escalating costs and carbon footprints with surveys. That being the case, it may be useful for some species to look at sampling just following natural aggregations such as spawning for example where the subjects themselves do some of the commuting. Clearly recruitment estimates become less precise, but stock definition in contrast is simplified.

2.1.3 Summary of workshop presentations.

The plenary presentations from this TOR touched on several common challenges and a few new suggestions for pathways to improving survey flexibility in general and specifically with application to certain ecosystems and surveys.

Lewis Barnett introduced a new multivariate optimization approach to improve survey efficiency and flexibility by reducing the number of strata in the Gulf of Alaska Bottom Trawl Survey and allocating primary sampling units to provide the best chance of meeting survey objectives. In this case, the primary objective was to provide abundance estimates with a minimum standard of precision across all assessed stocks. The research demonstrates how moving from univariate survey optimization tools (e.g., those based on Neyman optimization) to multivariate tools (e.g., via this application of the Bethel algorithm) can help improve survey flexibility by tuning a survey design to allocate effort in proportion to the data needs for management while accounting for differences in species distributions and relative importance to a fishery or ecosystem (Oyafuso et al, 2021; 2022).

Patrik Börjesson discussed challenges with reduced effort (due to rising fuel costs) and spatial preclusion (due to marine protected areas and wind farm development) of bottom trawl surveys in the Kattegat and Skagerrak. He and his team are in the midst of determining how to prioritise stations within existing fixed station designs and how to potentially combine data from surveys with different designs to make up for survey reductions. They are approaching station prioritisation through a sensitivity analysis to determine which stations have the greatest influence on key data products such as abundance indices. Finally, they are hoping to evaluate the consequences of prior surveys changing from fixed station to stratified random designs.

Annica de Groote presented proposed changes to the design of the Kattegat cod survey. The existing design consists of a station grid that is stratified and then sampled within strata with a mixture of random sampling and independent random groups depending on the stratum. Annica proposes using spatially balanced sampling using the local pivotal method, which creates a strong negative correlation between inclusion indicators of units that are close to each other, based on the values of some auxiliary variables. As a result, nearby units are unlikely to appear together in a sample, which also aids in providing representative dispersion of effort throughout the domain.

Jason Conner presented an analysis of the eastern Bering Sea bottom trawl survey, evaluating the expected impact of changes in survey effort, statistical design, and estimators. They used a simulation framework, conditioning a spatiotemporal GLMM on the historical systematic survey and simulating survey sampling. They found that systematic sampling provided the most precise abundance estimates across three species, in accordance with sampling theory, but only when the appropriate local estimators were used (those computed on small groupings of samples over space). Applying appropriate estimators could make up for losses in precision due to reductions in total effort.

Zack Oyafuso presented a proposed new bottom trawl survey design for the U.S. Chukchi Sea. He fitted a spatiotemporal GLMM to limited historical data from two gear types and many species to serve as an operating model over which to optimise stratification and sample allocation using a genetic algorithm and the Bethel algorithm as described in Oyafuso et al. (2021). He found that a stratified random sampling design provided the best expected results in terms of consistent, reliable and precise estimates of abundance. Such a design would also facilitate more flexibility in future changes to the design, as opposed to a systematic design. The systematic design also produced biased precision estimates and less consistent results across effort levels. Thus, a new stratified random design is proposed for if and when future monitoring is needed in the Chukchi Sea.

Andy Lipsky presented a summary of the challenges and opportunities arising from the massive expansion of offshore wind farms on the Pacific, Gulf of Mexico, and Atlantic ocean with large-scale construction set to begin in the northeast U.S.. Most importantly, he and collaborators are developing a strategy to account for the development of offshore wind farms by determining how preclusion of many sampling gears due to wind farms will influence the interpretation of existing long-term fisheries survey data. They are also attempting to determine the effects of this buildout on the coastal habitat and physical oceanography, and what impact that may have on fish distributions. Readers should look out for the final strategy for mitigating the impacts of offshore wind energy development on fisheries surveys, to be released in the Fall of 2022 by NOAA Fisheries and the Department of the Interior’s Bureau of Ocean Energy Management.

2.1.4 Recommendations for designing flexible surveys.

Flexibility against changes in survey effort should be incorporated into both existing and new survey designs; however, determining how to accomplish this is rarely straightforward. We propose the use of a roadmap or decision tree to provide guidance on designing flexible surveys (Figure 2.1-6).

In the initial design of flexible surveys, decision makers and survey practitioners must have a clear agreement on the objectives and priorities of a survey. This may include precision targets or thresholds for population estimates based on stock assessment data needs and the relative importance of different stocks to a fishery. Objectives may also include estimation of environmental variables that are important for ecosystem-based fisheries management (e.g., biodiversity indices, forage abundance, temperature, oxygen and salinity landscapes). Survey objectives are often not explicitly stated in such a way, making decisions difficult as there is no framework for evaluating consequences relative to the survey objectives. When there is no guidance as to the prioritisation of objectives, it should be explicitly stated that the aim is to equally weight all objectives to obtain the most robust estimates across all components of an ecosystem, community, or assemblage. With clear, stated objectives, survey scientists may then consider appropriate and available modes of observation (e.g. trawl, trap, video, etc.) to address the study’s objectives. Given these modes of observation, scientists may then develop an appropriate statistical design and determine how to allocate sampling effort within this design to best address the objectives.

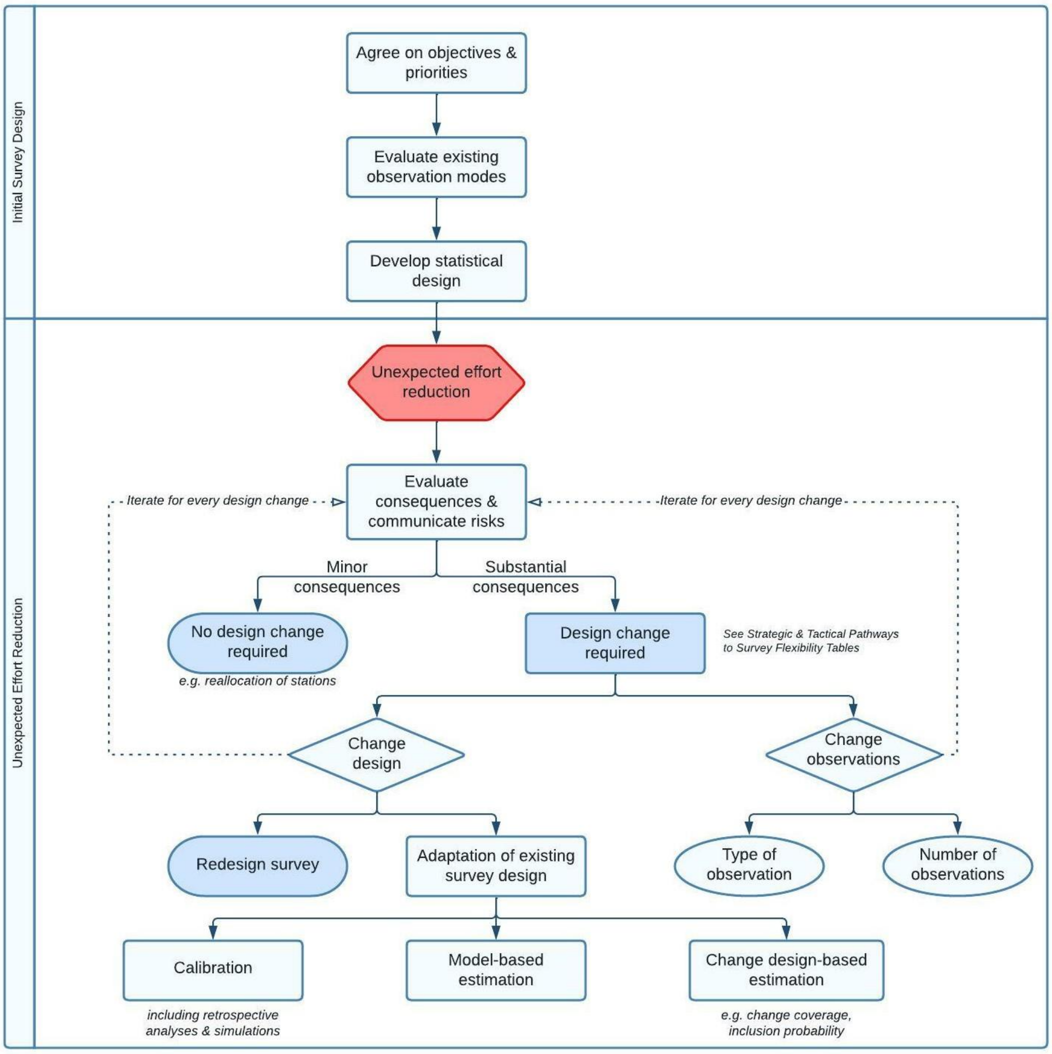


Figure 2.1-6: Road map providing guidance to increase the flexibility of existing survey designs or create new flexible designs.

Should a change in survey effort become necessary, it is essential for survey personnel to evaluate the consequences of various potential changes that can accommodate the shift in survey effort and communicate the risks of implementation to agency leadership (those tasked with ultimate oversight of the assessment enterprise). If it is determined that consequences of a design change are minimal, no change in survey design is necessary. Rather, implementation may be adjusted through reallocation of survey effort to accommodate the required effort reduction. If, however, potential consequences of effort reduction are substantial, adjustments to the survey are required (see Tables 2.1.1 and 2.1.2). Determining thresholds for what magnitude of change in estimates requires a change in the survey is not straightforward and will likely require simulation analyses specific to the focal system to determine how a given change in an estimate will affect management advice (e.g., by propagating the change through a simulated stock assessment model).

Changes to survey design may take the following forms:

1. Changes in design (which may also require changes in estimation)
2. Changes in observations amount and type

A change in survey design requires either a complete redesign of the survey (i.e. the creation of a new time series) or the alteration of existing survey methodologies. For example, an alteration of an existing survey may result in the expansion or contraction of the survey's spatial domain. The result of survey alteration may require calibration of new methods to historical survey products, perhaps incorporating retrospective analyses and simulation. Should calibration be logistically infeasible, alternative alteration approaches may include reliance solely on model-based estimation or adjusting inclusion probabilities and thus modifying design-based estimates.

Alternatively, survey practitioners may change observations associated with a survey design. This may include changing the amount or frequency of observations or the type of observation. For example, the inclusion of supplemental surveys or data may allow for a flexible solution to changes in survey effort.

Regardless of the type of change to a survey, it is critical that survey scientists document the potential consequences of inaction along with clear and objective reasoning for why and how modifications were or were not made. In the two sections below, we expand on strategic and tactical pathways to survey flexibility in the face of common challenges through design, observation, and estimation. We encourage survey scientists and stock assessment authors to carefully consider the consequences of inaction and the costs and benefits of each pathway to modifying surveys to address these consequences.

2.1.4.1 Strategic pathways to survey flexibility

Table 2.1-1 demonstrates long-term (one to many years ahead of cruise), strategic options for survey flexibility in the face of three categories of common changes that represent challenges to survey continuity. These changes include 1) <u>decreased survey resources</u> (time, money, vessels, staffing) relative to the scope of objectives (e.g., the number of stocks to monitor or the area of the sampling domain), which could encompass cases where the survey resources are simply not keeping pace with changes in the scope of objectives; 2) <u>habitat expansion or contraction</u>, which includes modifications to the amount of habitat available to the survey due to species range shifts, expansion or contraction; 3) <u>reduced sampling universe</u>, wherein some portion of the prior survey domain are no longer accessible to traditional sampling gear. For each change, we detail the consequences of inaction to emphasize the risk of failing to adjust the survey enterprise, and then provide options for actions that would account for such changes while ensuring survey continuity. In each subsection below, we elaborate on the options for accounting for survey changes, structured by the source of flexibility: sampling design, observation method, and estimation method.

2.1.4.1.1 Sampling design

Several different sampling designs from existing fisheries surveys were presented during WKUSER2, which included: systematic and random trawl surveys, systematic transect acoustic-trawl surveys, and fixed station trap and video surveys. Various flexibilities in these designs were discussed, which were generally caused by emergent decreases in sampling capability during survey execution, and how in many such situations these solutions were required to be created in-situ with little time to consider the full scope of consequences. The consensus of TOR I is that contingency designs should be made in advance and that the impacts of such contingencies should be well defined relative to the study objectives and priorities. Much of this contingency planning is covered in the subsequent section Tactical Pathways to Survey Flexibility, while here we emphasize that when substantial changes are necessary, survey planners should consider adoption of statistical designs that lend themselves to flexibility in the long term.

Classes of sampling designs that offer more flexibility are those that use probabilistic sampling and stratification, such that samples can be added or subtracted randomly while simply accounting for differences in inclusion probability in the estimation process. Likely the most common of such methods in fisheries surveys is the stratified random sampling, which facilitates design-unbiased estimation that is robust to most prudent survey modifications through re-stratification and/or sample reallocation (see Tactical Pathways to Survey Flexibility below for additional details). Cluster sampling can be another flexible survey design that maximizes the number of samples collected per research platform with unbiased estimators. One implementation of cluster sampling would treat each vessel sampling day as a cluster unit. Pre-selected random starting points would be allocated, and in one day, the vessel would acquire as many samples (e.g. trawls) as feasible, using a random direction and distance for each subsequent sample. Transit between random starting points could be done overnight to begin a new cluster the following day. Another example would be the adaptive cluster sampling that concentrates samples where a targeted species is likely to be abundant. Such design can be favored when monitoring highly patchily distributed species that are rare in marginal habitat (or as it is applied in many acoustic transect surveys to find the edges of the core range). Here, the selection of adjacent sampling units are made given that an observation meets a specific criteria, such as occurrence of a particular species or observation of population density above some threshold value.

Table 2.1-1: Strategic pathways to survey flexibility (long term ahead of survey)

		Source of flexibility to ensure comparability over time		
Change	Consequence	Sampling design	Observation	Estimation
Decreased survey resources (relative to scope of objectives)	Variance increase; potential bias; reduced information quality; timeliness of management advice	Reduce sampling density; Reduce domain area and maintain sampling density; Use randomized design with flexible allocation; Optimize design to prioritize objectives to reduce their scope; change statistical design; reduce survey frequency	Change to more cost-efficient observational methods (or combine with other traditional or non-traditional methods)	Correct for bias; correct for change in variance structure; use auxiliary information to shrink variance
Habitat expansion or contraction	Bias due to change in spatial or temporal availability; changes in variance structure, total uncertainty	Extend domain, potentially at cost of decreased sampling density (informed by, e.g., tagging research); use adaptive sampling; restratify; change statistical design	Combine alternative observational methods (time-efficient, habitat-appropriate combine surveys) and perform calibration	Evaluate change in spatial availability; stitch together years with different spatial distribution of observations; correct for change in catchability or selectivity
Reduced sampling universe	Bias; changes in variance structure, total uncertainty	Reduce scope of inference (spatial domain); increase resolution of design grid; change statistical design	Combine alternative observational methods or data sources and perform calibration; Integrate auxiliary information within inaccessible habitats	Develop and test for robust environment-biology relationships to predict in unobserved locations using modeling and simulations (given appropriate “before” data)

Nevertheless, one of the most flexible sampling designs might be a combination of systematic and random sampling. Zinger (1980) proposed unbiased estimators for the mean of a finite population and its variance. The systematic component of sampling would be allocated on a coarse grid so that the number of stations could be reliably sampled with some fraction of total expected sampling effort. Then a random, or equal probability stratified random, component would overlay the sampling frame, excluding the systematic sampling units. Thus, in situ reductions to random sampling could be accomplished using random elimination of remaining random samples if the need arises, without introducing bias to the mean or standard error.

2.1.4.1.2 Observation method

Changes due to decreased survey resources, changes in habitat domains, or reduced sampling universe may necessitate changes in our observation systems, including survey platforms and survey sampling methods. Such changes require an examination of the survey objectives and sampling priorities. Survey objectives are often broad and prioritization is often overlooked as the desire for and use of survey data are significant and heavily relied on for the management of many stocks. A clear set of survey objectives, potentially including priority or 'core' objectives and secondary objectives that can be dropped if needed, will aid in decision-making related to necessary changes in observations. These objectives may be related to representation (e.g., biodiversity monitoring, characterization of the environment) or to prioritization among species to obtain more precise/accurate abundance and compositional data for specific stocks while accepting less precise/accurate estimates for other stocks, based on importance to management. An example of observation system change in response to a decrease in survey resources may include the following: reducing overall survey effort (e.g., days at sea, or the number of primary sampling units), reducing biological sampling by prioritizing what sampling metrics will be executed per survey effort, e.g., reduction in biological subsample sizes and analysis such as prioritization of species for stomach content analysis, lengths, ageing structures, or maturity information. An example of observation system changes in response to changes in habitat domains include the following: adaptively expanding beyond original survey area bounds if surveyed species are detected near the boundaries and using alternative sampling methods. Examples of observation system changes in response to a reduction of the sampling universe may include the following: using smaller vessel platforms or unmanned systems to effectively and safely operate in wind energy areas where larger vessels may be excluded, using remote sampling techniques including acoustics, optical, or emerging eDNA sampling methods to accurately and precisely obtain samples in untrawlable habitats where areas of change are likely, integrating other data such as environmental covariate data from other sources of information that can be used to make necessary inferences.

2.1.4.1.3 Estimation method

If we adjust the sampling design in response to changes in the survey conditions (available resources, habitat or sampling universe), we also need to adjust the estimation, since sampling and estimation are intertwined (at least from a design-based perspective). For instance, assume that the sampling design is stratified sampling with simple random sampling within strata. If we respond to a budget cut or a habitat change by reducing the total sample size or changing the allocation of the sample over strata, this affects the inclusion probabilities of the sampling units and thus the estimation. A smaller sample size will also increase the variance of the parameter estimates. We can try to counteract this in the estimation by use of auxiliary information. A design-based model-assisted way of doing this is regression estimation, see e.g. Ch. 6-8 in Särndal et al (1992). If the auxiliary variables covary with the study variable, the regression estimator has smaller variance than the design-unbiased estimator. The regression estimator is however only approximately unbiased (for large samples). Another example of the relationship between sampling and estimation is if the sampling universe is reduced. Involuntary exclusion of part of the target population from sample selection is a type of cut-off sampling. In general, cut-off sampling leads to biased estimates of target population parameters. One way of dealing with this is to use a ratio adjustment for the cut-off part of the population (see e.g. p. 531-533 in Särndal et al (1992)). Another possibility is to estimate the cut-off through modelling.

2.1.4.2 Tactical pathways to survey flexibility

Table 2.1.2 demonstrates short-term (immediately prior to, or within cruise), tactical options for survey flexibility in the face of three categories of common changes that represent challenges to survey continuity (defined in section 2.1.4.1 above). The format is similar to Table 2.1.1, with the addition of a column distinguishing how solutions vary according to the type of survey design currently in place. In each subsection below, we elaborate on the options for accounting for survey changes, structured by the source of flexibility: sampling design, observation method, and estimation method.

2.1.4.2.1 Sampling design

On the tactical timescale, decisions about required survey changes must be heavily weighted by the existing statistical design of the survey. The sources of flexibility and the diversity of options varies substantially depending on the sampling design. One useful dichotomy here is that designs using probability sampling (those that assign an explicit, non-zero probability of sampling each primary sampling unit in the domain), particularly in combination with stratified sampling, generally allow for the most flexibility. In contrast, non-probability sampling designs (e.g., fixed station designs) are more limiting due to the fact that randomized reallocation of samples among areas is not always possible or straightforward.

Systematic sampling designs are inflexible because the predefined sampling extent must be completely sampled and no added efficiencies (such as minimizing distance travelled between primary sampling units selected for observation, Oyafuso et al. 2022) are available to survey practitioners. Phillips reported on several events in the time series of the coast-wide Pacific hake acoustic-trawl survey where sampling platforms became unavailable and the transect spacing was coarsened to allow observations covering the full spatial extent of this survey. A contingency plan for systematic sampling may consider ad hoc stratification of the sampling frame, where completed samples are pooled into 1 or more strata with the original sampling resolution and random starting point and remaining samples are collected into 1 or more new strata with a separate resolution and starting point. Then the design-based estimates for these strata may be combined using area-weighted stratification methods. The utility of this approach for point sampling methods such as trawl, trap, or drop-camera surveys can be limited by the distance between consecutive sampling locations and the time it takes for a vessel to transit this distance.

Table 2.1-2: Tactical pathways to survey flexibility (short term: just prior or within cruise)

			Source of flexibility to ensure comparability over time		
Change	Consequence	Type of survey	Design	Observation	Estimation
Decreased survey resources (relative to scope of objectives)	Potential bias; change in variance structure; change in spatial domain	Non-probability sampling (e.g., fixed stations)	Decrease resolution of sampling grid	Prioritize sampling (e.g., increase subsampling of biological data for some species)	Account for differences in inclusion probability, sample density; models to fill in data gaps
		Probability sampling (e.g., stratified random)	Reallocate samples with lower density in some strata given objective criteria; aggregate strata	Prioritize sampling (e.g., increase subsampling of biological data for some species)	Account for differences in inclusion probability; models to fill in data gaps
Habitat expansion or contraction	Bias due to change in spatial availability	Any	Add new strata to best partition the sampling universe; redesign	Evaluate whether conditions differ in portions of the sampling universe added or removed	Account for differences in inclusion probability; models to fill in data gaps (e.g., account for missing prior data from new strata)
Reduced sampling universe	Bias; changes in variance structure, total uncertainty	Non-probability sampling (e.g., fixed stations)	Reduce stations; increase station density	Gather alternative data (e.g., environmental only)	Account for differences in inclusion probability, sample density; models to fill in data gaps
		Probability sampling (e.g., stratified random)	Reallocate samples with lower density in some strata given objective criteria	Gather alternative data (e.g., environmental only)	Models to fill in data gaps

Simple random sampling designs are also inflexible due to the need to rigorously adhere to a defined sampling allocation. When non-random processes are imposed on random sampling (e.g. sampling at the end of a survey is truncated), bias is introduced to a design that is meant to be unbiased. As with the systematic sampling design, contingency plan – ad hoc stratification whereby the remaining sampling frame is constituted into a different stratum and inclusion probabilities for proportional sampling are calculated – may be useful.

On the other hand, stratified random sampling designs are flexible in that the partitions needed to respond to effort reallocation are established and based on biological or physical criteria. However, the presence of too many strata risks the inability to calculate reliable estimates of variance. Thus, adding strata to account for a habitat expansion is often the best response unless samples are already stretched too thin across many strata, in which case it may be necessary to aggregate strata via poststratification. If the strata are allocated samples proportional to their area, then altering sampling density in one or more strata requires using estimators for probabilities proportional to size (Cochran, 1977).

2.1.4.2.2 Observation

Probably the most common problems that affect surveys are short-term, arising from a variety of factors including weather, mechanical, or personnel issues. Such problems may also occur at more than one stage of a given survey, compounding the impact to that survey year. It is necessary to plan for this reduction in effort as this is a question of "when" rather than "if" and surveys must be prepared to adapt quickly.

Reductions in survey effort will have consequences and reductions to accuracy and precision for any species should be clearly communicated. Perhaps the most important factor to guide the decision-making process for how to deal with reduced effort is prioritization of objectives and desired level of precision from survey estimates. For example, a CV of 20% may be desired for some species, but a CV of 40% may be acceptable for others. All components of data collected by a survey should be prioritized, including the species of interest, survey region, biological sampling and auxiliary sampling (e.g. oceanographic and plankton). Some prioritization may require analyses, such as quantifying the acceptable level of subsampling or optimal number of samples per strata. Although priorities may change over the course of a long-running survey time-series, these priorities should remain consistent when dealing with short-term effort reductions to avoid confusion in the decision-making process.

Often, those executing surveys are not statisticians and require clear guidelines for making decisions to reduce sampling effort in order to limit any potential to introduce bias or violate the statistical survey design. It is important to avoid dropping entire strata and sample reductions should be spread throughout the entire survey region when possible. To do so, decisions must be made early with the full picture of the remaining survey region and available effort in mind. Methods, and tools to implement them, could be developed to allow those executing surveys to make decisions regarding which samples to remove. It may be desirable to remove the samples from the farthest, deepest or most difficult areas of a survey due to the increased gain in time; however, doing so routinely could bias estimates and alter the perception of distribution.

There are other viable options to avoid losing entire stratum by increasing the efficiency of the survey progress or by reducing the amount of effort at any one location. Increasing the transit efficiency where applicable could be an effective way to mitigate impacts due to lost survey time. Approaches such as the "traveling salesperson problem" may help to optimize survey routes and decrease the overall time required to complete a survey. Increasing the efficiency of survey execution may often be at odds with reducing costs. Completing a survey quicker might require faster transit speeds or increased personnel, which does not guarantee success and may not be feasible for many surveys due to vessel or other limitations. Another option may be to reduce the tow or transect duration or distance. For surveys with relatively long tow durations (~30 minutes), reducing the duration (to ~15 minutes) may increase the overall number of samples taken in a survey and may allow sampling in areas difficult to operate due to bad bottom or fixed gear by shortening the overall path to be sampled. Surveys should have clear and established tolerance ranges for the optimum and minimum sample distance and duration. Tools can be developed for use at sea to measure and verify this in real time to ensure standardization throughout a survey time-series. Calculated indices of abundance must properly account for the shortened tow duration. In addition to accounting for differences in effort when expanding estimates of local density to regional abundance, it is necessary to consider how such changes may influence catchability or detection probability, thus some additional research effort may be required to allow for more efficient sampling and observation near the boundaries of excluded zones.

2.1.4.2.3 Estimation

Regardless of the type of change that survey practitioners are faced within the short term (just prior to or within a cruise), there are typically two options for mitigating these changes in situ: 1) adjusting design-based estimators to account for modified sample density or inclusion probabilities (the probability that a given primary sampling unit would be sampled); 2) accounting for unbalanced sampling using model-based estimators that can provide an informed estimate for unsampled areas given spatial and/or temporal correlation or by relying on established relationships between the response and some covariate(s) that can be measured in areas without direct ecological observation. If the existing survey design is flexible enough to allow for design-based estimation with modified inclusion probabilities, this is the most straightforward and low-cost approach. However, one should be careful to ensure that such changes are not introducing bias. When approach 1 is not possible or does not meet all the survey objectives it is often necessary to use approach 2 with model-based estimators to fill in missing data or otherwise integrate multiple data sources with different temporal or spatial extents. Specifically, it is becoming increasingly common to use spatiotemporal models to knit together surveys or account for unbalanced data (Ono et al. 2018; Ianelli et al. 2019; O’leary et al. 2020; Thorson et al. 2021).

2.1.5 What are the current major challenges?

2.1.5.1 Disruption of Surveys due to Offshore Wind Energy Development

As of 2022, North America and Europe are pursuing aggressive climate mitigation strategies to decarbonize national energy systems. The use of offshore fixed and floating wind technologies are being advanced to achieve these goals. The U.S. has set a target of 30 gigawatts of offshore wind production by 2030 with a pathway to 110 gigawatts by 2050 (GWEC, 2022). In 2021, The European Commission set a 300 gigawatt goal for 2050 (GWEC, 2022) with new commitments by some European nations to increase and speed up these targets to address an energy crisis and the war in Europe. The amount of marine space needed to develop necessary transmission systems and turbine generators is substantial. In the U.S., reaching the 2030 targets will require approximately 3,411 turbine generators and approximately 10,000 miles of new submarine cabling over 2.37 million acres in the Northeast U.S. shelf ecosystem (BOEM, 2022a). As of September 2022, over 22.37 million acres of the U.S. northeast shelf ecosystem has been designated by the Bureau of Ocean Energy Management (BOEM) as offshore wind leases, wind energy areas, and wind planning areas (BOEM, 2022b, BOEM, 2022c). Applying similar technology requirements, the 110 gigawatts U.S. target will require over 10,000 turbines and 33,000 miles of submarine cable. Development at this pace, scale, and magnitude could have profound interactions on the marine ecosystem, NOAA trust resources, and the survey enterprises need to monitor populations, stocks, and ecosystem conditions (Saba et al. in press). This emerging new use requires existing regional fisheries independent survey programs to evaluate the space-use impacts associated with this development on survey missions/operational requirements. This presents both a massive challenge and opportunity to evaluate and propose potential changes in the design, observation systems, and estimation approaches to meet existing and new scientific requirements.

The U.S. has evaluated the effects and impacts of this proposed development and determined there would be major adverse impacts on NOAA's ability to continue to conduct existing core long term fisheries independent surveys that underpin the management of commercial and recreational fisheries and conservation and recovery of protected marine mammals, sea turtles, and fish species (BOEM, 2022c). This analysis determined that offshore wind development would have four primary impacts: 1.Preclude the ability for NOAA to continue to carry out its large vessel and aircraft survey missions in wind energy areas due to operational safety issues, 2. Impact existing statistical survey designs, 3. Result in habitat change that could result in changes in variance structure of stocks/populations inside and outside wind areas, and 4. Impact survey efficiency by increasing vessel transit time and reducing aircraft operations (Hare et al., 2022). To address these impacts NOAA Fisheries has collaborated with its sister agency, the Bureau of Ocean Energy Management, to develop a programmatic strategy and mitigation framework to address the challenges of implementing current scientific surveys in a new era of offshore wind energy development.

NOAA Fisheries *Survey Mitigation Implementation Strategy* is national in scope but focused on the Northeast U.S. region and describes goals, objectives, a strategy framework, and actions necessary to effectively mitigate the impacts of offshore wind. The document provides guidance and direction for survey program leads to develop mitigation plans for impacted surveys. In the Northeast region there are thirteen long-term surveys that will be impacted by proposed offshore wind development and require such plans. Survey mitigation plans would include a number of elements including: the objectives of the survey, specific stakeholders for the data collected, evaluation of impacts on the design and sampling methods of the survey through e.g. observation system simulation and modeling approaches; planned mitigation measures that would include solutions to the impacts on statistical design and sampling methods, necessary calibrations and/or data integration approaches, estimated costs, and proposed schedules for implementation. Plans would also include mechanisms for adaptive management, evaluation, and peer-review, communication strategies necessary to implement the plans, and database design, management, and data availability and accessibility requirements for new or altered data streams.

Early actions NOAA Fisheries is taking to mitigate impacts on the Northeast Fisheries Science Center's (NEFSC) fishery-independent, multi-species bottom trawl survey include development of a spatial model framework to simulate and evaluate the impacts of spatial overlap of wind energy development in the survey area and evaluate alternative sampling strategies to minimize impacts of wind energy development on estimates of abundance and distribution for multiple fish and invertebrate stocks currently sampled by the survey (ref Angelia Miller talk]. NOAA Fisheries is also advancing survey mitigation activities to address the impacts to the NEFSC Atlantic Scallop Survey enterprise which consists of a dredge survey, habcam survey, and various other cooperative partner monitoring efforts. The New England Fisheries Management Council established the Atlantic Sea Scallop Survey Working Group to develop recommendations to improve the Atlantic sea scallop survey system; including integration of existing scallop surveys, mechanisms to incorporate new approaches to address survey spatial coverage, sampling intensity and frequency; data standardization, storage and access, and ways to address potential impacts from offshore wind development (NEFMC, 2022). In addition, NOAA Fisheries is at an early phase of augmenting its HabCam observation system from a vessel towed system that will not be capable of sampling within wind energy areas to an uncrewed system capable of operating autonomously within wind energy areas.

2.1.5.2 Technological creep and calibration for vessel and gear changes

Regardless of the underlying survey design, standardized long-term monitoring to create time-series has also to be flexible enough to deal with changes in vessels, gears, and technological developments. The first consideration is often to deal with these changes by performing calibration experiments to provide conversion factors. However, carrying out special vessel/gear calibration experiments is almost impossible due to high effort need and such an approach is unlikely to provide conclusive conversion factors, as changes are expected to be relatively small and require large numbers of observations to statistically detect them (for more extensive discussion, see TOR II. Thus, understanding survey gear efficiency is critical for minimising the amount of physical calibration that is needed (Somerton et al. 2007).

Furthermore, technological creep is likely to occur as minor changes are introduced (e.g. newer gear materials, improved bridge software, etc.) over time. However, as time passes, the accumulation of limited/very limited changes might become more and more significant. This is especially the case in multi vessels/country monitoring situation which could lead to deviation in the standardized population indices.

Owing to the technological creep, it is unclear how new vessel/gear should be calibrated to especially, as these minor changes are often not well documented. Hence, WKUSER2 supports the conclusion from WKNSIMP (ICES 2019) favouring pragmatic approaches for combining time series as discussed more extensively in TOR II.

2.1.5.3 Survey costs

At sea surveys are generally expensive so most of the surveys are already designed to be cost-effective as budgets are often limited. Oversampling is for various reasons (budget, ecological impact, workload) unlikely to occur in a manner that makes reductions possible without losing precision and accuracy or increasing the likelihood of bias. Some reduction in effort might be possible in some cases but will lead to a higher risk of short-term impacts e.g. weather and mechanical malfunctioning might affect a survey season in such a way that the information gained from the survey no longer has any value because the signal to noise ratio becomes too low (e.g., the CV of an abundance index is so high that population trends cannot be identified).

Flexibility in respect to sampling additional habitats or by following a species distribution has a similar impact as reducing effort. It is possible when time spent in the original survey area is reduced to allow time in the new areas. Such a reduction in the original area might lead to a reduction in the precision and accuracy in the original area. This might be acceptable when the standards of the original objects are reduced. However, there is again the risk that the information will fall below the noise level. When the signal to noise ratio of a survey becomes too low, the total costs made for the survey are wasted.

2.1.5.4 Impact of increasing fuel costs

Reducing the fuel cost of a survey comes to reducing either or both the distance covered and the gear towing. The former is counter-intuitive when thinking of flexibility, since a solution to avoid any mid-trip surprise would be to make a first visit of the whole surveyed area before densifying the samples in a second part. Other solutions such as cluster sampling using a day or half a day as a primary sampling unit (was also discussed above), but this would deserve more research and testing.

There were discussions regarding increasing the efficiency of transit times which is a significant source of fuel consumption. There are fuel consumption trade-offs between increasing vessel transit speed and sailing fewer sea days *v.s.* reducing transit speed to conserve fuel over the duration of a survey. Another option is to utilize tools or methods to optimize the route on which a survey is executed or utilize varying ports between sampling legs to minimize transit distances.

Another solution could herald in internationally coordinated design when the whole area is covered by many nations and vessels, so that there may be room to reallocate the sampling using a minimum distance allocation. An example for this would be the NS-IBTS. Here, however, it must be ensured that there remains a minimum overlap between the different vessels in order to allow the estimation of potential vessel/country effects in model-based calculation of the abundance indices as outlined by WKINSIMP (ICES 2019). The west coast Pacific hake survey is conducted by two vessels, one from the US and one from Canada. This serves to, in some cases, minimize the total transit time either vessel needs to make by designing the survey based on the planned port location of the vessels prior to, during, and after the survey is completed.

The forthcoming workshop (WKPILOTNS-FIRMOG) will use the North Sea as a test case on providing an overview of all surveys providing data to ICES fisheries and ecosystem advice and this regional overview could provide the necessary information for considering such survey optimization. Reducing the towing force of a gear has a lot of potential, using additional means of collecting data together with the main gear (see section on UAVs and e-DNA), reducing the tow duration, working on gear rigging and/or using video. These are options developed in several ongoing research projects (e.g. SmartFish, Game of trawls, FishGenome), workshops, e.g. WKING (ICES, 2020b), monitored and discussed in ICES/WGFTFB under the supervision of ICES/DSTSG, which outputs and synthesis of work will need to be discussed in further WKUSER sessions.

2.1.6 Recommended future directions.

1. Develop and deploy tools to evaluate the costs and consequences of making survey changes, along with inaction, and communicate with transparency
2. If survey changes are needed, implement change incrementally (to allow attribution of effect)
3. Document and disseminate descriptions of the changes made
4. Improve collaboration between survey practitioners and stock assessors and within disciplines across national boundaries
5. Consider reducing effort made to acquire each replicate to limit lost observations, particularly when there is inadequate space or time to obtain an observation given conventional protocols (e.g., by taking smaller subsamples of biological data, or shorter tow or transect length or duration; however, note that these changes require an evaluation of their potential effect on catchability or detection, but may be a last resort if traditional collection methods are unsuitable)
6. Create survey-specific protocols for making objective in situ modifications when needed
7. Adopt probability sampling designs to the extent possible
8. Create objective criteria for how to make design changes such that observations can be removed or added with statistical rigor
9. Collaborate with other working groups facing similar challenges (e.g., WGOWDF, WGNAEO, WGIBTS, WGFTFB)

2.1.7 References

Battaglia, A., Trenkel, V.M., Rochet, M.-J., 2006. Estimating end effects in trawl catches. ICES Journal of Marine Science 63, 956–959. https://doi.org/10.1016/j.icesjms.2006.03.002

BOEM, 2022a. Renewable Energy GIS Data. Bureau of Ocean Energy Management: Maps and GIS Data. Accessed June 2022 https://www.boem.gov/renewable-energy/state-activities

BOEM, 2022b. State Activities. Bureau of Ocean Energy Management: Renewable Energy. Accessed June 2022 https://www.boem.gov/renewable-energy/state-activities

BOEM, 2022c. Ocean Wind Draft Environmental Impact Statement. Bureau of Ocean Energy Management. EIS/EA BOEM 2022-XX. Accessed June 2022. Retrieved from: https://www.boem.gov/renewable-energy/state-activities/ocean-wind-1.

Chen, J., Thompson, M.E., Wu, C., 2004. Estimation of Fish Abundance Indices Based on Scientific Research Trawl Surveys. Biometrics 60, 116–123. https://doi.org/10.1111/j.0006-341X.2004.00162.x

Cochran, W.G. 1977. Sampling Techniques, 3rd ed.; John Wiley and Sons: New York, USA; 900p

Folmer, O.; Pennington, M. 2000. A statistical evaluation of the design and precision of the shrimp trawl survey off West Greenland. Fish. Res. **2000**, 49, 165–178.

Fryer, R.J., Zuur, A.F., Graham, N., 2003. Using mixed models to combine smooth size-selection and catch-comparison curves over hauls. Can. J. Fish. Aquat. Sci. 60, 448–459.

Gabler, S., Stenger, H., 2012. Design effect of randomized systematic sampling. Statistics 46, 131–148. https://doi.org/10.1080/02331888.2010.491252

Gavaris, S.; Smith, S.J. 1987. Effect of allocation and stratification strategies on precision of survey abundance estimates for Atlantic Cod (*Gadus morhua*) on the eastern Scotian shelf. J. Northwest Atl. Fish. Sci. 1987, 7,137–144.

GWEC. 2022., "Global Wind Energy Council. Global Offshore Wind Report 2022. https://gwec.net/gwecs-global-offshore-wind-report/#:~:text=The%20offshore%20wind%20industry%20enjoyed,the%20global%20offshore%20wind%20industry. Retrieved: October 20, 2022, 13:25.".

Godø, O.R., Engås, A., 1989. Swept area variation with depth and its influence on abundance indices of groundfish from trawl surveys. J. Northw. Atl. Fish. Sci. 9, 133–139.

Hare, J. A., B. Blythe, K. Ford, B. Hooker, B. Jensen, A. Lipsky, C., Nachman, L. Pfeiffer, M. Rasser, and K. Renshaw. 2022. NOAA Fisheries and BOEM Federal Survey Mitigation Implementation Strategy - Northeast U.S. Region. in public review.

Kelly, E., Gerritsen, H., Stokes, D., and Moore, S.J. (2021). Cruise report: Irish Anglerfish & Megrim Survey 2021. FEAS Survey Series: IAMS 2021. Marine Institute

Ianelli, J., Kotwicki, S., Honkalehto, T., McCarthy, A., Stienessen, S., Holsman, K., … Fissel, B. (2019). Stock assessment and fishery evaluation report for the 2019 North Pacific groundfish fishery. 2019 Assessment of the walleye pollock stock in the eastern Bering Sea". North Pacific Fishery Management Council. https://archive.afsc.noaa.gov/refm/docs/2019/EBSPollock.pdf

ICES. 2017. SISP 15 - Manual of the IBTS North Eastern Atlantic Surveys. Series of ICES Survey Protocols, SISP 15 (v. 4). https://doi.org/10.17895/ices.pub.3519

ICES, 2019. SISP 6 - Manual for mackerel and horse mackerel egg surveys, sampling at sea (report). Series of ICES Survey Protocols (2012–2020). https://doi.org/10.17895/ices.pub.5140

ICES. 2019. Workshop on Impacts of planned changes in the North Sea IBTS (WKNSIMP). ICES Scientific Reports. 1:67. 25 pp. http://doi.org/10.17895/ices.pub.5609

ICES. 2020a. ICES Workshop on unavoidable survey effort reduction (WKUSER). ICES Scientific Reports. 2:72. 92pp. http://doi.org/10.17895/ices.pub.7453

ICES. 2020b. ICES Workshop on Innovative Fishing Gear (WKING). ICES Scientific Reports. 2:96. 130 pp. http://doi.org/10.17895/ices.pub.7528

ICES, 2001. International Program of Standardised Trawl Surveys (IPROSTS) (EU Study contract 98-057 No. Contract Reference DG XIV Study Contract 98-057). ICES, Copenhagen, Denmark.

Liu, Y., Chen, Y., and Cheng, J. 2009. A comparative study of optimization methods and conventional methods for sampling design in fishery-independent surveys. – ICES Journal of Marine Science, 66: 1873–1882. https://doi.org/10.1093/icesjms/fsp157

Madow, W.G., 1953. On the Theory of Systematic Sampling, III. Comparison of Centered and Random Start Systematic Sampling. The Annals of Mathematical Statistics 24, 101–106.

Millar, R.B., Fryer, R.J., 1999. Estimating the size-selection curves of towed gears, traps, nets and hooks. Rev. Fish Biol. Fish. 9, 89–116.

Moriarty, M., Sell, A.F., Trenkel, V.M., Lynam, C.P., Burns, F., Clarke, E.D., Greenstreet, S.P.R., McGonigle, C., 2018. Resolution of biodiversity and assemblage structure in demersal fisheries surveys: the role of tow duration. ICES J Mar Sci 75, 1672–1681. https://doi.org/10.1093/icesjms/fsy050

NEFMC (New England Fishery Management Council's Scallop Survey Working Group and O'Keefe, C). 2022. Atlantic Sea Scallop Survey Working Group Report and Recommendations. 62pp. https://s3.us-east-1.amazonaws.com/nefmc.org/4b.-SSWG-Report-September-2022.pdf (accessed 10/7/22).

O'Leary, CA, Thorson, JT, Ianelli, JN, Kotwicki, S. Adapting to climate-driven distribution shifts using model-based indices and age composition from multiple surveys in the walleye pollock (Gadus chalcogrammus) stock assessment. Fish Oceanogr. 2020; 29: 541– 557. https://doi.org/10.1111/fog.12494

Oyafuso, ZS, Barnett, LAK & Kotwicki S. 2021. Incorporating spatiotemporal variability in multispecies survey design optimization addresses trade-offs in uncertainty. ICES Journal of Marine Science 78(3):1288–1300. https://doi.org/10.1093/icesjms/fsab038

Oyafuso, ZS, Barnett, LAK, Siple, MC & Kotwicki S. 2022. A flexible approach to optimizing the Gulf of Alaska groundfish bottom trawl survey design for abundance estimation. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-434, 142 p. https://doi.org/10.25923/g5zd-be29

Palma-Pedraza, S., Sarrazin, V., Clarke, M., Stokes, D., 2020. Status of non-assessed fish species in Irish waters (Technical Report). Marine Institute.

Kotaro Ono, James N Ianelli, Carey R McGilliard, André E Punt, Integrating data from multiple surveys and accounting for spatio-temporal correlation to index the abundance of juvenile Pacific halibut in Alaska, ICES Journal of Marine Science, Volume 75, Issue 2, March-April 2018, Pages 572–584, https://doi.org/10.1093/icesjms/fsx174

Raring, N.W., E.A. Laman, P.G. von Szalay, and M.H. Martin. 2016. Data report: 2011 Gulf of Alaska bottom trawl survey. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-330, 231 p. doi:10.7289/V5/TM-AFSC-330

Särndal, C.-E., Swensson, B. and Wretman, J. 1992. Model Assisted Survey Sampling. Springer-Verlag.

Somerton, D.A., Munro, P.T. and Weinberg, K.L., 2007. Whole-gear efficiency of a benthic survey trawl for flatfish. Fishery Bulletin 105:278–291.

Thorson, J.T., Arimitsu, M.L., Barnett, L.A.K., Cheng, W., Eisner, L.B., Haynie, A.C., Hermann, A.J., Holsman, K., Kimmel, D.G., Lomas, M.W., Richar, J. and Siddon, E.C. (2021), Forecasting community reassembly using climate-linked spatio-temporal ecosystem models. Ecography, 44: 612-625. https://doi.org/10.1111/ecog.05471

Von Szalay, P.G., Somerton, D.A., 2005. The effect of net spread on the capture efficiency of a demersal survey trawl used in the eastern Bering Sea. Fish. Res. 74, 86–95.

Xu, B., Ren, Y., Chen, Y., Xue, Y., Zhang, C., Wan, R., 2015a. Optimization of stratification scheme for a fishery-independent survey with multiple objectives. Acta Oceanol. Sin. 34, 154–169. https://doi.org/10.1007/s13131-015-0739-z

Xu, B., Zhang, C., Xue, Y., Ren, Y., Chen, Y., 2015b. Optimization of sampling effort for a fishery-independent survey with multiple goals. Environ Monit Assess 187, 252. https://doi.org/10.1007/s10661-015-4483-9

Zhao, J.; Cao, J.; Tian, S.; Chen, Y.; Zhang, S. 2018. Evaluating Sampling Designs for Demersal Fish Communities. *Sustainability* **2018**, *10*, 2585. https://doi.org/10.3390/su10082585

Zinger, A., 1980. Variance Estimation in Partially Systematic Sampling. Journal of the American Statistical Association 75, 206–211. https://doi.org/10.1080/01621459.1980.10477454

2.2 TOR II. “Combining surveys, dealing with data gaps”: Collate advice on methods to combine data from different sources, how to deal with data gaps and how to perform survey calibrations

Sub-Group Leads: Jim Thorson, María Grazia Pennino

Participants: Iosu Paradinas, Madison Hall, Patrik Börjesson, Yves Reecht, Guillermo Martin, Bríd O'Connor, Corina Chaves (remote), John Gabriel Ramirez (remote)

2.2.1 Why it is often necessary to combine surveys

Fisheries-independent surveys provide primary information to understand the stock status for both commercial and non-target fish species, often providing the only data source available to estimate relative abundances for non-commercial species. These surveys tend to be discrete monitoring programs, operating at local scales usually associated with the exclusive economic zones of countries that manage the surveys. However, populations often extend beyond jurisdictional boundaries and, therefore, the integration across national jurisdictional boundaries and multiple surveys is often required (Pinsky *et al.*, 2018; Moriarty *et al.*, 2020; Maureaud *et al.*, 2021).

Similarly, surveys typically sample only a portion of the population existing at a given location. In some cases, this portion is (essentially) a random subsample of the population at that location. If this subsampling process is random and its rate is constant across space and time, then its effect can sometimes be approximated by estimating a catchability coefficient and/or selectivity function when subsequently using a survey-based data product (e.g., as index in a stock assessment model). However, in other cases the portion that is sampled may not be random. For example:

1. Bottom trawls primarily sample those individuals that are near the bottom, and acoustic/midwater trawls sample primarily those higher in the water column
2. Baited gears (hook-and-line or videotrap) will likely only sample individuals that respond quickly to bait
3. Surveys operating in the one season may miss the portion of the stock at that location that is available to a fishery operating in a different season

In these cases, each survey technology is applicable to a portion of a given population.

In probability sampling theory (Cochran, 2007), we often seek to identify a set of primary sampling units that are (1) non-overlapping, (2) cover the entire stock, and (3) can be associated with a known “inclusion probability”, where alternative sampling designs involve different protocols for assigning inclusion probabilities to sampling units (see Table 2.2.1 for definitions). This set of sampling units is then called the sampling frame. Even in cases when a sampling gear can only sample some subset of the sampling frame, we still find it useful to define this “stockwide sampling frame” that covers the entire stock so that we can make inference from the gear to the entire stock (see Fig. 2.2.1). We can then refer to a separate “survey-specific sampling frame” for each individual survey, representing sampling units from the stockwide frame that are included for sampling by that survey given the gear used and other logistical constraints. This allows us to apply design-based estimators for each survey-specific sampling frame individually, and also potentially to define a separate design-based estimator for the stockwide frame. For example, we might define a stockwide sampling frame that includes primary sampling units defined as all combinations of defined geographical areas and also defined vertical layers of the water column.

In this case, a bottom trawl gear might access only the bottom layers, and other surveys (and associated sampling methods) are necessary to sample higher in the water column. In other cases, these survey-specific sampling frames may overlap somewhat, such that auxiliary information is needed to combine samples in those different sampling frames.

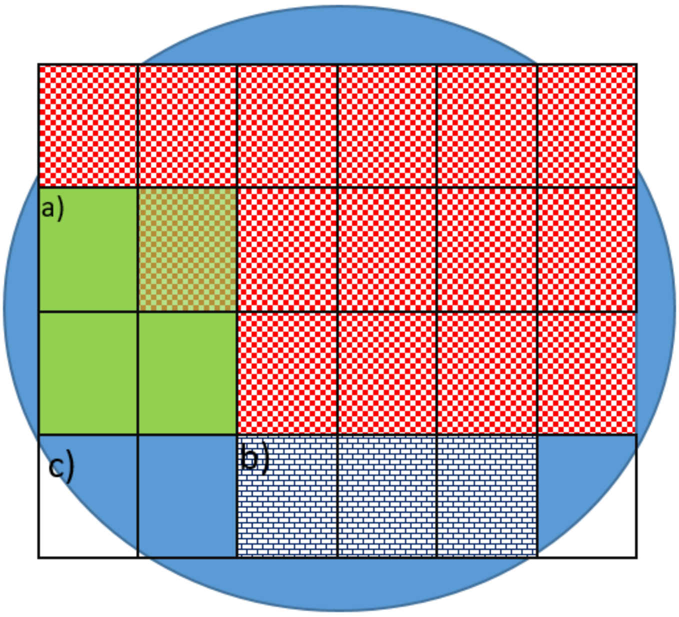


Figure 2.2-1: blue circle is the stock, the square-mesh is the stockwide sampling frame, with each grid cell representing a primary sampling unit, red and white represents the units sampled by the most extensive survey (15 units) and every grid cell that is not covered by this survey is considered a gap. Green (4 units) and blue-white (3 units) cells represent sampling units that are sampled by other "opportunistic" data. a), b) and c) refer to different "gap-filling" scenarios. a) cases where the opportunistically paired samples are available to intercalibrate the primary (red) and opportunistic (green) surveys, b) cases where surveys do not overlap while sampling the same stock, and c) cases where there are no samples at all.

Creating a sampling design involves specifying the sampling frame, but also assigning a sampling (a.k.a., inclusion) probability for each primary sampling unit. Design-based estimators typically assume that these sampling probabilities are assigned a priori and that they are followed exactly when implementing the design.

2.2.2 Gear calibration and fishing power ratios

So far, we have introduced the idea of a stockwide sampling frame, where individual surveys each sample a portion of that stockwide frame, and some primary units are sampled by multiple gears. However, we have not discussed how specifically information from these different surveys might be combined. Combining all these sources of information into a single estimator is difficult because of potential differences in catchability (the expected ratio of the sampled response occurring in the same primary sampling unit at the same time) and/or selectivity (the expected ratio of the response for a given age/length/sex). Accounting for different catchability/selectivity is called "survey calibration."

Survey calibration requires that two surveys are sampling the same underlying variable. In cases where this is violated, then the ratio of the estimated catchability or selectivity – referred to as fishing-power ratio hereon – will vary systematically over time and space (i.e., because the relationship between variables being sampled by each survey itself varies over time and space). In this case, the fishing-power ratio might not be useful to extrapolate in space or time, which greatly degrades the value of estimating a fishing-power ratio.

Survey calibration (almost) always requires paired sampling, and samples might be paired in an experimental or opportunistic way. The traditional approach to combine surveys is to quantify the gear efficiency through "experimentally paired sampling", in which samples are conducted with two vessels that fish in parallel, where paired-sampling stations are allocated following a probability design, and subsequently comparing catches to infer an expected fishing-power ratio that represents its value when averaging over the sampling frame as a whole. This fishing-power ratio can then be applied to data from one gear to predict what would have been observed if sampling had occurred with the other gear. This type of calibration approach is expensive and dependent on availability of two research vessels. Separately estimating a calibration ratio and then converting data using this ratio prior to analysis also fails to propagate uncertainty about the fishing-power ratio.

We introduce alternative vocabulary to understand cases when data are available but are not fully paired. Specifically, each survey sample has some attributes that can be measured from the time and place of sampling, including, but not limited to:

1. primary unit being sampled
2. location of sampling within a primary sampling unit
3. time of day
4. vertical position
5. features of seafloor bathymetry including depth, rugosity, and aspect
6. day of year ("seasonality")
7. sea state including tide and wave height

There are also controllable aspects of the deployment of gear, which affect the results of sampling given the attributes already listed. These include:

1. gear configuration and geometry (wing spread and bottom contact)
2. speed of towing for mobile gears
3. skipper attributes including knowledge (measured via econometrics)

These latter "controllable aspects of gear deployment" are sometimes called "catchability covariates", and these collectively measure "sampling effort". These aspects should be measured for every sample, so that we can calculate effort associated with each sample. Fully paired sampling then has the same sampling attributes except for those attributes that are under experimental control. In experimentally paired sampling, an analyst can then fit a statistical model that includes experimental treatments as a factor, and estimating density for each primary sampling unit (or replicated samples within each), and estimating fishing power ratio as the coefficient associated with experimental treatments.

Using more formal definitions (rather than fishing power-ratio), calibration experiments are designed to measure the expected ratio of the response for two or more gears in the same sampling unit (termed "catchability ratio"), or the ratio for a given portion of the population (termed "selectivity ratio"). Sampling units are typically defined as spatial locations within a domain, and occurring at a fixed time, so catchability/selectivity ratios are often described as a ratio of expected responses for samples at a given place and time. When measuring a ratio for a given sample, one of the other gears often measures a zero (no individuals present). In this case, it is not possible to calculate a ratio (no division by zeros). In these cases, the measured ratio can be transformed to and subsequently analyzed as a proportion of the total across both gears caught by each individual gear (Brooks et al., 2022). As an alternative, it is possible to define fishing power as a ratio of expected catches (which are typically never exactly zero) rather than observed catches (which often includes zeros). This alternative does not require converting ratios to proportions, and instead models the expectation of each gear while estimating their ratio (e.g., as a generalized linear model with a log-link, where the additive coefficient associated with experimental treatment is the estimated log-ratio). This alternative can be derived as a thinned and marked point process (Thorson et al., 2022).

In other cases, analysts might have "opportunistically paired sampling" that is similar in some sampling attributes but not others. In these cases, we obtain measurements of as many of these sampling attributes as possible, and include those as covariates to "control for" differences in sampling attributes that are not under experimental control. In some cases, opportunistically paired samples might not perfectly overlap. For example, we might have two surveys that operate in adjacent but non-overlapping areas. In this case, it is theoretically possible to still estimate the fishing power ratio between gears. This estimate requires assuming that the change in the target variable is smooth (i.e., that densities on both sides of a boundary separating surveys are similar). In this case, estimating a fishing power ratio for not-quite overlapping "opportunistically paired samples" is a type of "regression discontinuity design" (i.e. the rate of change between stations in adjacent strata is constrained by the assumed smoothness of the density estimated by the model).

In many cases, different gears sample different portions of a sampling frame corresponding to the unit stock under management or for assessment. For example, in the Four-Spot Megrim (*Lepidorhombus boscii)* in area ICES in Divisions VIIIc and IXa, separate gears are used for Porcupine Bank *vs.* Celtic Sea and Bay of Biscay areas. In this case, there are some "opportunistically paired" samples which occur at nearby locations, but which are not well paired with respect to other attributes (e.g., occuring in different seasons). However, it is also possible to do experimentally paired sampling, and this can help to validate estimates of catchability/selectivity ratio arising from analyzing opportunistically paired samples.

When designing calibration experiments, it is important to make a calibration-experimental design that allows inference to the whole sampling frame for which calibration will be used. For example, paired calibration samples might be randomized spatially over the entire spatial extent that is sampled by gears. It is also important to ensure that paired samples have the same value (or experimentally randomized values) for as many attributes as is feasible.

Different studies have analyzed "opportunistically paired sampling" in various ways (Table 2.2.2), and have confirmed that statistical models can offer the opportunity of overcoming challenges in combining data across surveys with varying gear efficiencies and resolutions.

Opportunistic data often also arise from a data-generating process with some unknown probability of samples in each primary sampling unit, where these sampling probabilities are clearly outside of experimental control. “Preferential sampling” arises when this true but unknown sampling probability is correlated with target densities (Conn et al., 2017; Pennino et al., 2019; Rufener et al., 2021; Alglave et al., 2022).

2.2.3 Characteristics of alternative sampling gears

We identify many types of sampling gears, which might require different designs to intercalibrate in different pairwise combinations. These gears include:

1. Bottom trawl, whether large mesh (typically otter trawl) or small mesh (typically beam trawls)
2. Visual surveys, whether scuba/towed/airplane-based counts, or using deployed video, including tows, drifted, stationary, and net-mounted cameras
3. Acoustical surveys, including opportunistic, fixed-station vs. mobile, single vs. multi-frequency for species/size identification
4. Traps, whether baited or behavioral, including chevron traps, pots, intercept traps, and gill nets
5. Environmental DNA (eDNA), including metabarcoding (i.e., proportional measurements) vs. calibrated qPCR (i.e., measurements of eDNA density that is calibrated against aquaculture individuals to estimate densities)
6. Pelagic trawl, whether surface, midwater, oblique tow, using various single or paired nets (Methot, Bongo, etc.), targeting adult, juvenile, ichthyo-, zoo- or phytoplankton
7. Shore-based sampling, e.g., purse seine, weirs
8. Hook-and-line, whether longline, angler surveys
9. Subsurface sampling gears, including van-Veen grabs, dredge

Some of these are in-situ and do not involve removing individuals from their habitat, while others involve collecting the individuals. Gears that involve removing individuals facilitate additional types of subsampling including otoliths (for ages), weight, gonadosomatic index, etc. Each pair of gears presents different logistical challenges in designing a calibration study, whether experimental or model-based.

These different gears then become important when designing a survey protocol that has samples from (almost) all primary sampling units. For example, a small mesh trawl is typically better at capturing small-bodied animals than a large-mesh trawl, and large-bodied animals often have lower numerical density. Therefore, the large-mesh trawl can cover a larger distance without requiring additional time to sort individual animals by species. To cover larger distances, large-mesh surveys often use a net with a wider opening, often pulled open using otter-trawl doors. However, these otter-trawls are not easily deployed in shallow water, so beam-trawls are instead typically used in shallow waters. For species that utilize both shallow and deep waters for a given life-stage, it might be necessary to define a stockwide sampling frame that includes sampling units at all depths, but then use large-mesh otter trawls offshore and small-mesh beam trawls inshore. However, new technologies may then be applicable to both habitats (e.g., environmental DNA), and these new sampling technologies present both challenges and opportunities for adapting a survey protocol.

In some cases, these gears may already be in use by commercial, recreational, and subsistence resource users, and we call records resulting from these activities "fishery-dependent data". Fishery-dependent data can be entirely outside of experimental control (i.e., fishery CPUE), or might arise cooperatively with scientists following protocols that are defined a priori (i.e., cooperative data). There is also increased emphasis on "citizen-science data" and these can be used for marine resource monitoring (Thorson *et al.*, 2014).

These fishery-dependent and citizen-science data typically differ from conventional surveys in many of the attributes listed previously, and these differences must be addressed during analysis. However, they also have many benefits including:

- reduced survey costs
- increased sample sizes overall
- wider spatial and seasonal coverage of sampling
- improved relationships with stakeholders
- better understanding of stakeholder concerns
- information about relevant fishery trendsd
- increased buy-in of survey results

We therefore note that analyzing opportunistic (fishery-dependent and citizen-science) data has many ancillary benefits beyond improvements in sampling efficiency (Kaplan and McCay 2004; Johnson and van Densen 2007).

2.2.4 How gaps in sampling coverage arise, and how to respond

We define a "sampling gap" as the set of primary sampling units that have a zero probability of inclusion by any gear. These sampling gaps can arise at multiple time scales:

1. Seasonal time-scales: gaps can arise unexpectedly for many reasons including inclement weather, vessel technology and maintenance issues, staff injuries and illnesses, and many other reasons. When these situations occur, survey coordinators face the decision to select sample stations that minimise its impact on the data quality.
2. Interannual-to-decadal time-scales: other types of spatial-temporal gaps could arise for systematic issues as for example areas that cannot be sampled due to many reasons (e.g., type of habitat, spatial closures due to marine protected areas or wind farm areas).
3. Indeterminate time-scales: Spatial gaps also could be due to the species habitat as for example for species that are in the middle part of the column of the water that is not usually sampled; species that live in shallow waters that are not sampled by the traditional trawl survey performed for demersal species, etc.

A survey coordinator might have the following questions when addressing sampling gaps:

1. How to select the new or dropped sampling stations?
2. Whether to weight information differently for different managed or assessed stocks?
3. How to weigh the importance of information for other species and ecosystem components?

In all these situations a possible solution will be to have tools that help to plan how to deal with these gaps on the time-scale allowed (discussed more in detail in section TOR IV). These tools may differ for species that have a persistent spatial-temporal distribution but not for species that have a changing distribution. Obtaining information ahead of time to prepare is typically easier for long than short time scales (e.g., before new spatial closures are created).

As stated previously, both experimentally and opportunistically paired sampling can be intercalibrated to develop a model that represents spatio-temporal variation in the target variable, as well as differences among sampling gears. This model can then be used to simulate new data, and subsequently used in a simulation experiment to predict the effect of sampling changes on management objectives.

In model-based approaches, it can also be helpful to compile covariates that can explain variation in sampling responses among primary sampling units. These covariates can then improve predictions in areas with sparse or absent data, but also require some consideration of relevant spatial scales when calculating the covariate values to use.

2.2.5 General recommendations regarding combining data

We recommend several principles for defining a model used to integrate data from multiple sampling programs, including surveys, fishery-dependent and citizen-science programs (see diagram "Combining Data"):

1. When multiple data sets are available but each does not individually capture the stock-wide sampling frame, we recommend exploration of combining data sets. This effort would then include: (A) identifying relevant experimentally or opportunistically paired data; (B) discussing with scientists and stakeholders whether the data sets are sampling the same underlying variable or have some known differences that prevent combining them; (C) fitting data sets individually and confirming that they do not disagree more than expected from sampling variability alone; (D) if either B or C suggest it, using catchability covariates to correct for known or observed differences (see Fig. 2).
2. Regardless of whether calibration (catchability/selectivity) is measured experimentally or opportunistically, we recommend that analysts estimate densities for each data set individually. In cases when density maps or index trends differ between data sets, then we recommend that analysts develop hypotheses to explain these differences. In simple cases, a difference may arise from low sample sizes (i.e., estimates are all within the standard errors of the others), such that combining data types will resolve these differences. How to estimate the statistical significance of differences is an active area of research, but (Rufener *et al.*, 2021) proposed using a chi-squared test for this purpose. In other cases, the estimated densities are largely consistent in which case it is straightforward to combine data in a single model while estimating catchability ratios (Grüss and Thorson, 2019).
3. We recommend designing the collection of experimentally paired sampling data, and evaluating whether it is worth allocating available resources to implement the design. If opportunistically paired samplings are already being used in a model-based framework to combine data, then these experimentally paired sampling data can then be used to validate the estimated calibration ratio. If two data sets cannot be combined using opportunistically paired sampling data, then, the new experimentally paired data provides an avenue to combine them in the future.
4. We recommend that densities measured in each data set be "area-weighted" when calculating a quantity intended to represent total abundance (or other variables typically calculated from densities). Area-weighting implies that the index variance is derived from the predictive variance (and covariance) at each primary sampling unit, such that information arising from a large portion of a unit stock (a high proportion of primary units in a sampling frame) is given larger importance than data arising from a small

portion of a stock. This recommendation precludes calculating separate indices and then combining them via an arithmetic average. It also precludes many commonly-used methods for combining indices post-hoc, including those that estimate a latent trend that is shared among indices produced for each data set individually (Conn, 2009; Peterson *et al.*, 2021).

5. To ensure that data sets are "area-weighted" we recommend that data be analyzed within a model that estimates density at each primary sampling unit, where this density is acknowledged to vary (i.e., as a set of fixed effects, or spatially smoothed random effects), and where catchability ratios are estimated as the ratio of expected observations at a given sampling unit. This catchability ratio may vary among sampling units either as a function of measured attributes (i.e., covariates), or based on unmeasured attributes (i.e., spatially varying catchability). Estimating variation in catchability is often necessary for both experimentally and opportunistically paired samples, to ensure that the differences among surveys are properly attributed to differences in survey attributes rather than differences in the spatial location of samples.
6. When combining data that are not paired experimentally, we recommend including variables to control for attributes that are not paired between gears. For example, if two gears differ in time of day, then we recommend estimating the response to time-of-day as a covariate, and then predicting densities at a standardized time of day. This treatment has been called a "catchability covariate" in index standardization modeling.
7. Where age/length/sex subsamples are available for multiple surveys, we recommend exploring age/length/sex-specific modelling to generate estimates of abundance or composition by category. We also include continued efforts to test performance using simulation, recognizing that model-based expansion of age-based indices or compositions using multiple data sets has received less research attention that biomass-aggregated models (although see Thorson and Haltuch, 2018);
8. As a special case of #6 that deserves special mention, data are often not well paired seasonally. In this case, then analysts must consider potential movement between those times. Analyzing the impact of movement occurring between the timing of two surveys is an active area of research but could be addressed using a variety of mechanistic and correlative approaches (Dormann *et al.*, 2012; Hanks *et al.*, 2015; Thorson *et al.*, 2021);
9. We recommend developing good-practices for diagnosing model issues when combining data sets. Such diagnostics include: (A) checking for confounding between estimated random effects and catchability covariates, which arises when a catchability covariate is highly correlated with spatial or spatio-temporal random effects; (B) checking for preferential sampling in individual data sets by comparing estimated random effects against the sampling/inclusion probability for each data set; and (C) retrospective testing, where models are fitted to temporal or spatial blocks of available data to check for stability when adding/excluding data;
10. We recommend ongoing efforts to communicate successes/failures and developing "good practice guidance" when applying models to intercalibrate multiple data sets for generating abundance indices, age/length/sex composition and other data inputs. This would be most helpful if including members from the US, Europe and other regions, and would be helpful occurring every 2-4 years to keep up with rapid developments.

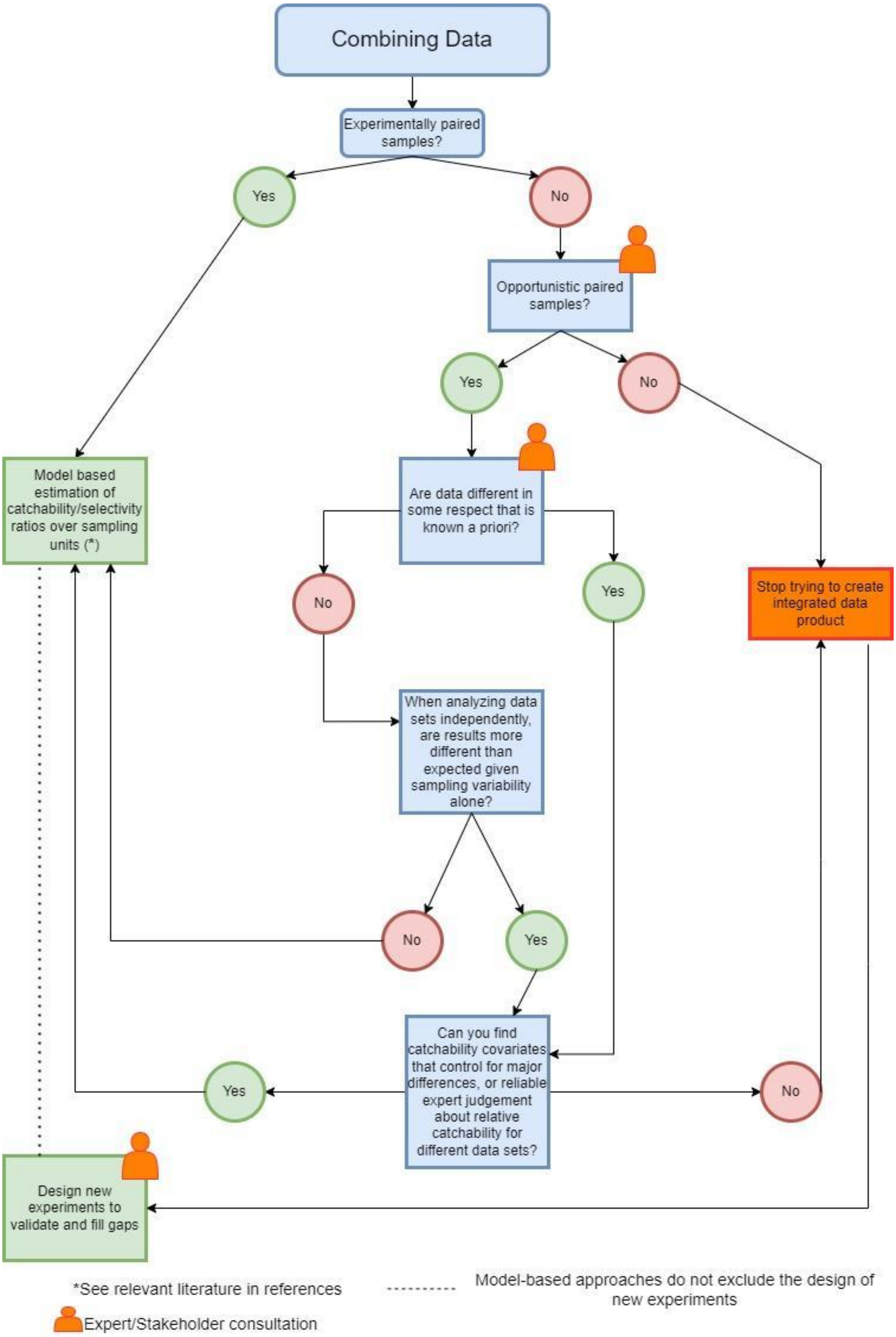


Figure 2.2-2: A proposed roadmap for TOR II

Table 2.2-1: Definitions used in TOR II

Term	Definition
Stock	A component of a population under management or being modeled, often defined geographically based upon jurisdictional boundaries
Sampling frame	Complete partition of a stock into distinct and non-overlapping “primary” sampling units, often defined based on geographic boundaries
Primary sampling unit	One unit of the sampling frame
Sampling design	Process for assigning a sampling probability to each primary sampling units within a sampling frame
Data gap	Primary sampling units that have zero probability of sampling within a given design, either due to logistical constraints (outside of national jurisdiction, too expensive, limited time), technical constraints (inability to use a gear within that unit), or disruption (vessel damage)
Paired sampling	Sampling where the underlying process/variable being sampled is the same (i.e., measuring the same species); often this is not known until after data are investigated to see if differences can be explained a priori or a posteriori
Experimentally paired sampling	Sampling with the same attributes in all respects except those under experimental study, i.e., occurring at the same place and time (i.e., sampling sampling unit), but using different gears
Sampling attributes	Aspects of sampling (either measured or latent), including those under experimental control (i.e., gear, location and season in experimentally paired sampling) or not (e.g., weather, crew, tide), where some attributes may be partially controlled (e.g., tow duration and performance in a bottom trawl)
Opportunistically paired sampling	Samples that are similar in at least one sampling attribute and different in others, e.g., occurring at nearby locations in space and/or time
Catchability covariates	Measurements of sampling attributes that are available to control for “sampling confounders” when attributing sampling response to underlying stock
Catchability ratio	Expected ratio (or proportion if data are pre-transformed) of response for two gears that would occur in experimentally paired samples. This ratio might be assumed to be constant across sampling units, or varying among units, and is often estimated in a statistical model that estimates variation in stock density among primary sampling units and also: for experimentally paired samples, typically includes experimental treatment as a catchability covariate; or for opportunistically paired samples, necessarily includes multiple confounder variables as catchability covariates.
Selectivity ratio or proportion	Same as catchability ratio, but where response is defined for a specific category i.e., age, length, sex, etc.

Table 2.2-2: List of research (not review or synthesis/perspectives) papers describing experimental or model-based synthesis and calibration of multiple sampling gears

Reference	Regions	Gears	Topics
(Moriarty *et al.*, 2020)	northeast Atlantic	Beam/Otter	spatio-temporally explicit relative abundance distribution maps from multiple surveys
(Conn, 2010)	US Atlantic coast	Gillnet, castnet, hook-and-line, recreational survey data, logbooks	Hierarchical framework for analyzing multiple, noisy indices with the goal of estimating a single time series of relative abundance
(Kotwicki *et al.*, 2017) (Kotwicki *et al.*, 2013)	Bering Sea	Bottom trawl and midwater-acoustic trawl	Combine bottom trawl and acoustic-midwater trawl to estimate acoustic dead-zone and vertical overlap
(Thompson *et al.*, 2022)	Gulf of Mexico	Videotrap surveys (no inter-calibration ended)	Generate a combined index more representative and of the relative abundances of commercial species
(Shelton *et al.*, 2014)	California Current	bottom trawl, autonomous underwater vehicle	Essential fish habitat
(Perretti and Thorson, 2019)	Northeast US	bottom trawl (otter and beam)	Combine nearshore and offshore bottom trawls
(O'Leary *et al.*, 2022)	Bering Sea	bottom trawl	Combine US and Russian bottom trawls
(Maureaud *et al.*, 2021)	global	bottom trawl	Inventory of bottom trawl surveys worldwide
(Monnahan *et al.*, 2021)	Bering Sea	Bottom trawl and midwater acoustic-trawl	Joint model for bottom and miwater-acoustic trawl
(Gwinn *et al.*, 2019)	Southeast US	Chevron trap and videotrap	Create a single integrated index
(Alglave *et al.*, 2022)	Bay of Biscay	Survey and commercial bottom trawl	Combine fishery and survey data for habitat
(Ono *et al.*, 2018)	Bering Sea, Gulf of Alaska, Aleutian Islands	Bottom trawl	Combine bottom trawl surveys across Alaska
(Adams *et al.*, 2021)	Northeast US	Bottom trawl	Combine Canada and US surveys for quota allocation
(Rufener *et al.*, 2021)	Western Baltic Sea	Survey and commercial bottom trawl	Account for preferential sampling
(Grüss and Thorson, 2019)	Gulf of Mexico	bottom trawl, longline, acoustic-midwater trawl	Combine biomass, counts, and presence/absence data
(Peterson *et al.*, 2021)	Simulation study	various indices	Dynamic Factor Analysis (DFA) to combine survey indices post-hoc
(Grüss *et al.*, 2021)	nearshore Alaska	beam trawl, beach seine	Nearshore fish atlas for nearshore habitat associations

(Thorson *et al.*, 2020)	Bering Sea, Northeast US	bottom trawl, midwater trawl	Seasonal dynamics when combining surveys
(Martin Gonzalez *et al.*, 2021)	North Sea	survey gear/ commercial	Combine independent and commercial data sources to predict spatial-temporal distributions
(Thygesen *et al.*, 2019)	Namibia	paired trawls	statistical intercalibration in paired fishing operations
(Cadigan and Dowden 2010)	Simulation study	data from previous paired trawl study (Cadigan et al 2003)	test calibration approaches, measure accuracy of statistical inferences
(Miller *et al.*, 2010)	Northeast US	paired bottom trawls	simulations, statistical behavior of calibration factor estimators
(Miller, 2013)	Northeast US	paired bottom trawls	hierarchical models, selectivity, catchability
(Rademeyer and Butterworth 2013)	South Africa	paired bottom trawls	catchability, calibration of new and old survey gear
(Webster *et al.*, 2019)	Eastern Bering Sea	setline/ bottom trawl	spatiotemporal modelling, density indices, distribution models
(Olsen *et al.*, 2021)	Red Sea	Baited traps and gillnets	Stratified GAMs, not spatially explicit. Combined CPUEs, but also functional diversity index.
(Zhu *et al.*, 2018)	East China Sea	Survey and commercial bottom trawl	Multivariate auto-regressive state-space model (MARSS); combination of CPUEs
(Fowler and Showell 2009)	Scotian shelf, Canada	bottom trawl	catchability calibration between survey vessels for abundance indices
(Delargy *et al.*, 2022)	Wales, UK	paired dredges, survey and commercial gear	utilizing commercial data for estimating scallop abundance indices

2.2.6 References

Brooks, M.E., Melli, V., Savina, E., Santos, J., Millar, R., O'Neill, F.G., Veiga-Malta, T., Krag, L.A. and Feekings, J.P., 2022. Introducing selfisher: open source software for statistical analyses of fishing gear selectivity. Canadian Journal of Fisheries and Aquatic Sciences, 99(999), pp.1-9.

Conn, P. B. (2010). Hierarchical analysis of multiple noisy abundance indices. Canadian Journal of Fisheries and Aquatic Sciences, 67(1), 108-120.

Delargy, A. J., Hold, N., Heney, C., Cann, R. P., Bhandari, K., Colvin, C. N., Moore, A. B. M., et al. 2022. Detection and quantification of differences in catch rates among research vessel gears and commercial vessels. Fisheries Research, 254: 106371.

Adams, C. F., Brooks, E. N., Legault, C. M., Barrett, M. A., and Chevrier, D. F. 2021. Quota allocation for stocks that span multiple management zones: analysis with a vector autoregressive spatiotemporal model. Fisheries Management and Ecology, n/a. http://onlinelibrary.wiley.com/doi/abs/10.1111/fme.12488 (Accessed 13 May 2021).

Alglave, B., Rivot, E., Etienne, M.-P., Woillez, M., Thorson, J. T., and Vermard, Y. 2022. Combining scientific survey and commercial catch data to map fish distribution. ICES Journal of Marine Science: fsac032.

Cochran, W. G. 2007. Sampling techniques. John Wiley & Sons.

Conn, P. B. 2009. Hierarchical analysis of multiple noisy abundance indices. Canadian Journal of Fisheries and Aquatic Sciences, 67: 108–120.

Dormann, C. F., Schymanski, S. J., Cabral, J., Chuine, I., Graham, C., Hartig, F., Kearney, M., et al. 2012. Correlation and process in species distribution models: bridging a dichotomy. Journal of Biogeography, 39: 2119–2131.

Grüss, A., and Thorson, J. T. 2019. Developing spatio-temporal models using multiple data types for evaluating population trends and habitat usage. ICES Journal of Marine Science, 76: 1748–1761.

Grüss, A., Pirtle, J. L., Thorson, J. T., Lindeberg, M. R., Neff, A. D., Lewis, S. G., and Essington, T. E. 2021. Modeling nearshore fish habitats using Alaska as a regional case study. Fisheries Research, 238: 105905.

Gwinn, D. C., Bacheler, N. M., and Shertzer, K. W. 2019. Integrating underwater video into traditional fisheries indices using a hierarchical formulation of a state-space model. Fisheries Research, 219: 105309.

Hanks, E. M., Hooten, M. B., and Alldredge, M. W. 2015. Continuous-time discrete-space models for animal movement. The Annals of Applied Statistics, 9: 145–165.

Kotwicki, S., De Robertis, A., Ianelli, J. N., Punt, A. E., and Horne, J. K. 2013. Combining bottom trawl and acoustic data to model acoustic dead zone correction and bottom trawl efficiency parameters for semipelagic species. Canadian Journal of Fisheries and Aquatic Sciences, 70: 208–219. NRC Research Press. https://cdnsciencepub.com/doi/10.1139/cjfas-2012-0321 (Accessed 9 September 2022).

Kotwicki, S., Ressler, P. H., Ianelli, J. N., Punt, A. E., and Horne, J. K. 2017. Combining data from bottom trawl and acoustic surveys to estimate an index of abundance for semipelagic species. Canadian Journal of Fisheries and Aquatic Sciences. http://www.nrcresearchpress.com/doi/abs/10.1139/cjfas-2016-0362.

Maureaud, A. A., Frelat, R., Pécuchet, L., Shackell, N., Mérigot, B., Pinsky, M. L., Amador, K., et al. 2021. Are we ready to track climate-driven shifts in marine species across international boundaries? - A global survey of scientific bottom trawl data. Global Change Biology, 27: 220–236.

Monnahan, C. C., Thorson, J. T., Kotwicki, S., Lauffenburger, N., Ianelli, J. N., and Punt, A. E. 2021. Incorporating vertical distribution in index standardization accounts for spatiotemporal availability to acoustic and bottom trawl gear for semi-pelagic species. ICES Journal of Marine Science.

Moriarty, M., Sethi, S. A., Pedreschi, D., Smeltz, T. S., McGonigle, C., Harris, B. P., Wolf, N., et al. 2020. Combining fisheries surveys to inform marine species distribution modelling. ICES Journal of Marine Science, 77: 539–552.

O'Leary, C. A., DeFilippo, L. B., Thorson, J. T., Kotwicki, S., Hoff, G. R., Kulik, V. V., Ianelli, J. N., et al. 2022. Understanding transboundary stocks' availability by combining multiple fisheries-independent surveys and oceanographic conditions in spatiotemporal models. ICES Journal of Marine Science, 79: 1063–1074.

Olsen, E., Axelsen, B. E., Moland, E., Utne-Palm, A. C., Elamin, E. M., Mukhtar, M. A., Saleh, A. M., et al. 2021. Distribution and diversity of fish species along the Sudanese Red Sea coast based on three combined trap and gillnet surveys. Fisheries Research, 242: 106032. Elsevier, Amsterdam. https://www.webofscience.com/api/gateway?GWVersion=2&SrcAuth=DynamicDOIArticle&SrcApp=WOS&KeyAID=10.1016%2Fj.fishres.2021.106032&DestApp=DOI&SrcAppSID=EUW1ED0EC3416aVEdHaJjdyqCjytS&SrcJTitle=FISHERIES+RESEARCH&DestDOIRegistrantName=Elsevier (Accessed 9 September 2022).

Ono, K., Ianelli, J. N., McGilliard, C. R., and Punt, A. E. 2018. Integrating data from multiple surveys and accounting for spatio-temporal correlation to index the abundance of juvenile Pacific halibut in Alaska. ICES Journal of Marine Science, 75: 572–584.

Pennino, M. G., Paradinas, I., Illian, J. B., Muñoz, F., Bellido, J. M., López-Quílez, A., and Conesa, D. 2019. Accounting for preferential sampling in species distribution models. Ecology and Evolution, 9(1), 653-663.

Perretti, C. T., and Thorson, J. T. 2019. Spatio-temporal dynamics of summer flounder (Paralichthys dentatus) on the Northeast US shelf. Fisheries Research, 215: 62–68.

Peterson, C. D., Wilberg, M. J., Cortés, E., and Latour, R. J. 2021. Dynamic factor analysis to reconcile conflicting survey indices of abundance. ICES Journal of Marine Science, 78: 1711–1729.

Pinsky, M. L., Reygondeau, G., Caddell, R., Palacios-Abrantes, J., Spijkers, J., and Cheung, W. W. L. 2018. Preparing ocean governance for species on the move. Science, 360: 1189–1191.

Rufener, M.-C., Kristensen, K., Nielsen, J. R., and Bastardie, F. 2021. Bridging the gap between commercial fisheries and survey data to model the spatiotemporal dynamics of marine species. Ecological Applications, 31: e02453.

Shelton, A. O., Thorson, J. T., Ward, E. J., and Feist, B. E. 2014. Spatial semiparametric models improve estimates of species abundance and distribution. Canadian Journal of Fisheries and Aquatic Sciences, 71: 1655–1666.

Thompson, K. A., Switzer, T. S., Christman, M. C., Keenan, S. F., Gardner, C. L., Overly, K. E., and Campbell, M. D. 2022. A novel habitat-based approach for combining indices of abundance from multiple fishery-independent video surveys. Fisheries Research, 247: 106178.

Thorson, Barbeaux, S. J., Goethel, D. R., Kearney, K. A., Laman, E. A., Nielsen, J. K., Siskey, M. R., et al. 2021. Estimating fine-scale movement rates and habitat preferences using multiple data sources. Fish and Fisheries, n/a. https://onlinelibrary.wiley.com/doi/abs/10.1111/faf.12592 (Accessed 18 October 2021).

Thorson, J. T., Scheuerell, M. D., Semmens, B. X., and Pattengill-Semmens, C. V. 2014. Demographic modeling of citizen science data informs habitat preferences and population dynamics of recovering fishes. Ecology, 95: 3251–3258.

Thorson, J. T., Adams, C. F., Brooks, E. N., Eisner, L. B., Kimmel, D. G., Legault, C. M., Rogers, L. A., et al. 2020. Seasonal and interannual variation in spatio-temporal models for index standardization and phenology studies. ICES Journal of Marine Science, 77: 1879–1892.

Thorson, J. T., Arimitsu, M. L., Levi, T., and Roffler, G. H. 2022. Diet analysis using generalized linear models derived from foraging processes using R package mvtweedie. Ecology, 103: e3637.

Zhu, M., Yamakawa, T., and Sakai, T. 2018. Combined use of trawl fishery and research vessel survey data in a multivariate autoregressive state-space (MARSS) model to improve the accuracy of abundance index estimates. Fisheries Science, 84: 437–451. https://doi.org/10.1007/s12562-018-1190-9 (Accessed 9 September 2022).

Miller TJ, Das C, Politis PJ, Miller AS, Lucey SM, Legault CM, Brown RW, Rago PJ (eds). 2010. Estimation of Albatross IV to Henry B. Bigelow calibration factors. Northeast Fish Sci Cent Ref Doc. 10-05; 233 p. Available from: National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, or online at http://www.nefsc. noaa.gov/nefsc/publications/

Cadigan, N. G., and Dowden, J. J. 2010. Statistical inference about the relative efficiency of a new survey protocol, based on paired-tow survey calibration data. Fishery Bulletin, 108: 15–29.

Christiansen, H. M., Switzer, T. S., Keenan, S. F., Tyler-Jedlund, A. J., and Winner, B. L. 2020. Assessing the Relative Selectivity of Multiple Sampling Gears for Managed Reef Fishes in the Eastern Gulf of Mexico. Marine and Coastal Fisheries, 12: 322–338.

Fowler, M., and Showell, Mark. 2009. Calibration of bottom trawl survey vessels: comparative fishing between the Alfred Needler and Teleost on the Scotian Shelf during the summer of 2005. Canadian Technical Report of Fisheries and Aquatic Sciences, 2824: iv–25.

Miller, T. J. 2013. A comparison of hierarchical models for relative catch efficiency based on paired-gear data for US Northwest Atlantic fish stocks. Canadian Journal of Fisheries and Aquatic Sciences, 70: 1306–1316.

Rademeyer, R. A., and Butterworth, D. S. (2013). An update of the catchability calibration factor between the Africana with the old and the new gear, with an attempt to estimate its length-dependence: 10.

Thygesen, U. H., Kristensen, K., Jansen, T., and Beyer, J. E. 2019. Intercalibration of survey methods using paired fishing operations and log-Gaussian Cox processes. ICES Journal of Marine Science, 76: 1189–1199.

Webster, R. A., Soderlund, E., Dykstra, C. L., and Stewart, I. J. 2020. Monitoring change in a dynamic environment: spatiotemporal modelling of calibrated data from different types of fisheries surveys of Pacific halibut. Canadian Journal of Fisheries and Aquatic Sciences, 77: 1421–1432.

Martin, G., Wiff, R., Marshall, C. T., and Cornulier, T. 2021. Estimating spatio-temporal distribution of fish and gear selectivity from pooled scientific survey and commercial fishing data. Fisheries Research, 243: 106054.

Johnson, T. R., and van Densen, W. L. T. 2007. Benefits and organization of cooperative research for fisheries management. ICES Journal of Marine Science, 64: 834–840.

Kaplan, I., and McCay, B. 2004. Cooperative research, co-management and the social dimension of fisheries science and management. Marine Policy, 28: 257–258.

2.3 TOR III. "Modeling and simulations": Further develop model performance evaluation through simulations, use of auxiliary information to improve survey data products, including appropriate propagation of uncertainty

Sub-Group Leads: Meaghan Bryan, Sean Anderson, Paul Regular

Participants: Margaret Siple, Coilin Minto, Gwladys Lambert, Angelia Miller, Semra Yalcin (remote), Anna Stroh

2.3.1 Introduction

Fisheries-independent surveys are usually carefully designed, accounting for species catchability and spatial distribution (Pennington and Strømme, 1998). Indices of abundance calculated from such surveys have long been derived from design-based theory, which does not require observations to follow a particular frequency distribution (Cochran, 1977; Smith, 1990). Under a design-based approach, samples are assumed to come from a finite population of sampling units (e.g. trawl sites) and, as such, the bias and precision of design-based mean and variance estimators are a function of the survey design (Cochran, 1977; Lohr, 2021). Robust estimates of stock size can therefore be derived from consistently implemented surveys and, thanks to foundational work by Cochran (1977) and other statisticians, such "model-free" analyses have been tractable since the 1970s. A key caveat is that this approach requires a consistently implemented survey, which is an ideal that is rarely possible in the real world.

Unfortunately, survey effort reductions are unavoidable and often force changes that violate the assumptions of design-based analyses (ICES, 2020a). For example, short-term changes can manifest as the inability to sample stations or the cancellation of surveys due to unforeseen events, budget cuts, or marine spatial planning campaigns, which in turn causes gaps in temporal and spatial data coverage. Moreover, sudden shifts in species distribution in response to climate-change, related ecosystem changes, and disturbance introduced by human activity (e.g., wind energy developments) may lead to total or periodic reallocation of survey effort or indefinitely restrict survey effort spatially, and ultimately lead to increasingly biased or uncertain design-based estimates. Though there is simplicity in the "model-free" nature of design-based estimators, this simplicity offers few options for compensating for flaws in the implementation of a survey. It is in such circumstances when predictions from model-based approaches, which infer finite population characteristics based on a model for the population (Lohr, 2021), may help compensate for survey inconsistencies.

The development of model-based approaches has allowed us to address some major challenges that can impact estimates from fishery-independent surveys. Generalised linear, additive, and associated linear and additive mixed models (GLMs, GAMs, GLMMs, GAMMS) have been used to correct for factors such as when and where survey sampling occurs that can impact fishing efficiency. In recent years, geostatistical models have been developed and are increasingly used to account for temporal, spatial, and spatiotemporal correlations to overcome problems with temporally inconsistent and spatially imbalanced survey designs, changing availability of fish populations to existing surveys, and shifting spatial distributions. Recently, employing model-based approaches has allowed us to use catchability and density covariates to improve estimates of abundance and uncertainty; leverage data from multiple surveys to develop single, comprehensive survey indices; and better predict species distributions (e.g., Shelton *et al.*, 2014; Grüss and Thorson, 2019; Thorson, 2019a; O'Leary *et al.*, 2020; Barnett *et al.*, 2021; Monnahan *et al.*, 2021). In the 'Design-based inference' and 'Model-based inference' sections below, we provide a review of design-based sampling theory; outline common characteristics of model-based approaches to index standardisation including decisions about response distributions, temporal structure, spatial effects, and spatiotemporal structure; and touch on the topic of model validation and selection.

Simulation modelling is an integral tool for testing new estimation methods, optimising survey designs and allocating sampling effort, and evaluating the impact of survey changes on abundance indices, stock assessment outputs, and management advice. Depending on the simulation objectives, two general approaches—resampling and model-based simulations—have been used. In the 'Simulation modelling' section we discuss the importance of simulation modelling, provide a review of how resampling and model-based simulation approaches have been used and when they are most appropriate, and discuss some of their limitations.

Overall, the aim is that the text for TOR III serves as a guide for practitioners seeking to develop a survey index or conduct simulation testing. The core ideas presented throughout this section are encapsulated in Figure 2.3.1. We conclude by providing a set of recommendations and suggestions for future research.

2.3.2 Contributed talks

There were six talks presented in TOR III at WKUSER2. These talks introduced important considerations for design- and model-based inference from survey data; emphasised the importance of diagnostics for determining model fit and suitability, the strengths and limitations of models for combining datasets, generating indices, and filling information gaps about the survey area; and highlighted the use of simulation to quantify the impact of survey changes on stock assessment outcomes. In terms of generating indices, there was consensus that practitioners should continue to implement well-designed surveys with design-based estimators where possible, as they are stable, intuitive, and fast/easy to calculate.

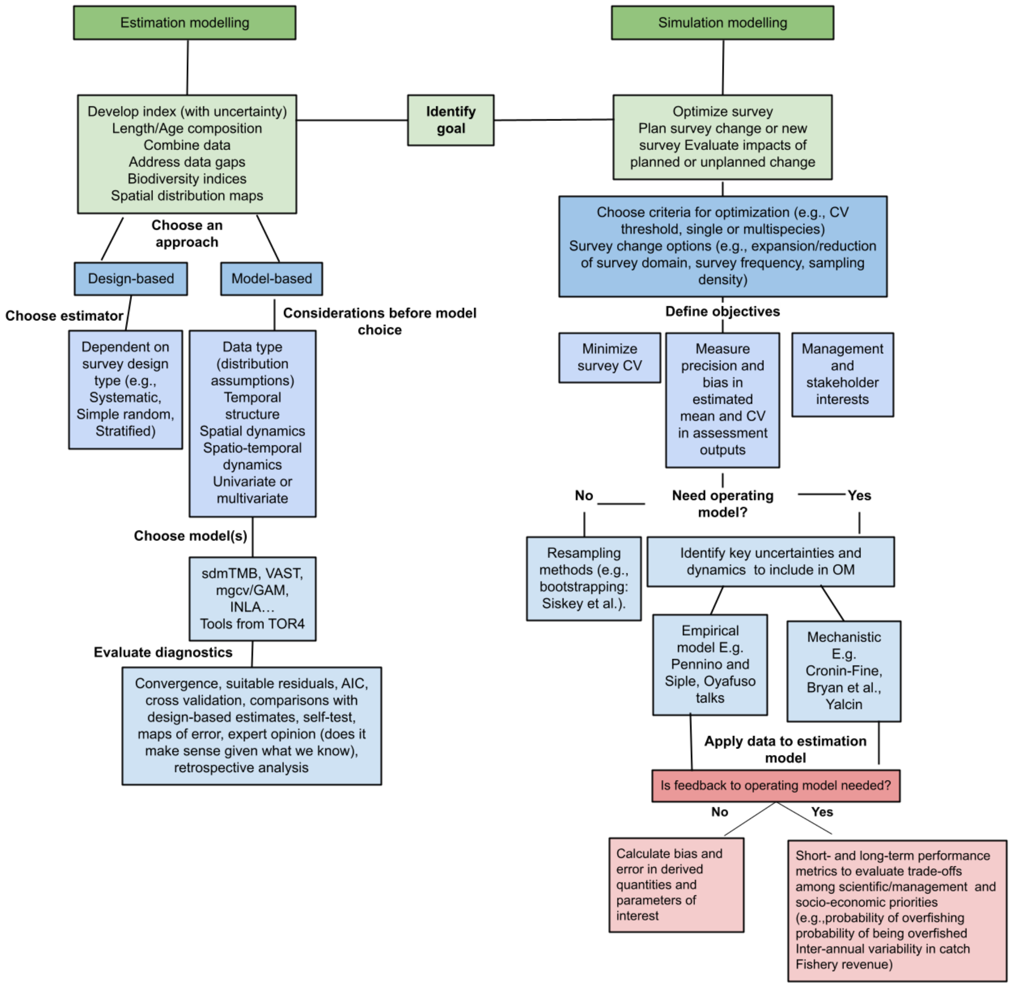


Figure 2.3-1: Roadmap for estimation and simulation modelling exercises.

James Thorson (NOAA AFSC) introduced spatio-temporal age- and length-composition expansion implemented in the Vectorized Auto-regressive Spatiotemporal (VAST) model (Thorson and Barnett, 2017; Thorson, 2019b). Most existing approaches use an age-length key that is assumed constant over space and time (but see Berg *et al.*, 2014; Babyn *et al.*, 2021). In this situation, sampling noise and error around age-length fit are not propagated through the age-length conversion, so errors in the estimates of age compositions made in these models do not reflect the true variability. Thorson presented new features of VAST that propagate this error through to the final age-composition estimates.

Maria Grazia Pennino (Instituto Espanol de Oceanografia) introduced a multispecies species distribution model implemented in INLA that uses covariance structure between surveyed species to fill in data gaps in abundance and distribution. This approach is similar to the “robin hood approach” to sharing data between poorly sampled and well sampled areas/species (Punt *et al.*, 2011). In a situation where there are multiple co-occurring species being surveyed, a model-based approach can help analysts share data between species to obtain density information.

Meaghan Bryan (NOAA AFSC), on behalf of Lee Cronin-Fine (University of Washington and NOAA AFSC Affiliate), presented a closed-loop simulation framework that includes 1) a spatially explicit operating model that includes a population dynamics accounting for movement among 8 areas on the Bering Sea shelf and slope, 2) an observation dynamic model with the same spatial structure generates fishery data and survey data (i.e., biomass and composition data) and used to model different survey strategies, 3) a spatially aggregated stock assessment model, and 4) a catch model that feeds back into the OM. The goal of the project was to examine the impact of survey changes (i.e., spatial extent and/or frequency) on assessment outcomes through the management system and the economic value of the fishery. Closed-loop simulations like this are helpful for evaluating the value of information, even if an economic model is not available and value is estimated on a relative scale. The closed-loop simulation approach allows practitioners to evaluate changes in management from multiple perspectives. The process of developing a closed-loop simulation is also a valuable communication tool that can be leveraged for improving survey science and identifying areas of future research.

Semra Yalcin (DFO) introduced a simulation approach with a mechanistic model with two life history types (cod- and yellowtail-like), each defined by their distribution patterns (depth preference and diffusion). This study aimed to determine whether models can accurately estimate the accuracy of design- and model-based indices of abundance and associated variance. Using SimSurvey (Regular *et al.*, 2020), Yalcin simulated a stratified random survey of two stocks (cod- and yellowtail-like) under four effort reduction scenarios (no change [base], random loss of effort, blocked loss of effort, and loss of strata). For each scenario, the accuracy of design-based estimators were compared to model-based estimates derived from sdmTMB (Anderson *et al.*, 2022). Yalcin found that the performance of correctly specified sdmTMB models was comparable to design-based estimators under the base case and random reduction scenarios. She also demonstrated that model-based approaches, especially models using depth as a covariate, outperformed design-based estimates when effort was reduced in a spatially blocked pattern. Finally, simulation results indicate that variance estimates from the best models were accurate as the confidence interval coverage (portion of cases when the true population available to the survey was within the 95% confidence intervals) was close to 95%, and AIC appears to be an indicator of the most accurate models. Overall, results indicate that correctly specified model-based approaches can reduce the impact of survey effort loss on accuracy.

Anna Stroh (ATU) presented a 4-year PhD project on improving survey abundance indices using spatio-temporal modelling. Four research objectives are defined: The first research objective aims at a comprehensive review of the causes and effects of survey (data) gaps and how they were dealt with. The second research objective aims at univariate approaches to study error sources (observational and process error), model diagnostics, and covariate structures for key species from the Irish Groundfish Survey. The third objective takes a multivariate approach and uses spatio-temporal modelling to improve the accuracy and precision of survey indices, and index standardisation, and aims to advance knowledge on the VAST approach. The fourth objective is using the MixFishSim package (Dolder et al., 2020) to create simulations for groundfish populations and test various survey designs (i.e. new gears, survey gaps caused by untrawlable areas, as well as to assess the power of simulations to detect true abundance metrics).

Margaret Siple (NOAA AFSC) presented an ongoing study comparing model-based indices of abundance from GAMs to GLMMs (implemented in VAST). Stocks without strong agreement between design- and model-based indices of abundance may benefit from additional model-based indices for comparison, especially in cases when there may be environmental drivers with nonlinear effects on the population. This study uses survey data from Alaska groundfish and crab stocks and fits models with similar structures to compare mean abundance index values and estimated CVs. Groundfish stocks from the Eastern Bering Sea (walleye pollock and yellowfin sole) with agreement between design- and model-based indices in terms of annual indices did not have similar CVs over time, and the design-based index performed differently for red king crab, a stock with a patchy distribution that may be less likely to be captured by the survey. Applying different model-based indices to existing stocks can provide a basis of comparison and/or consensus for a model-based index, especially in cases where the assumptions of a design-based index may be violated.

Talks from other TORs, though not summarised here (see abstract in section 4), provide insight into the relative strengths of design- and model-based inference for producing indices of abundance, age compositions, estimates of catchability, and other survey data products. Angelia Miller (UMass Dartmouth) presented a proposal for a management strategy evaluation that will evaluate the impacts of planned offshore wind developments on the east coast of the US. Miller summarised steps of the MSE process that have occurred so far, including a consultation with experts about anticipated changes in abundance and distribution of managed stocks. Throughout this section, we draw examples from talks contributed to other TORs as needed.

2.3.3 Design-based inference

Design-based inference—also called randomization inference (Lohr, 2021)—treats characteristics of interest (e.g., numbers of fish) as fixed and introduces randomness in how the sampling units are chosen from a finite population of sampling units (i.e., survey design) (Särndal *et al.*, 1978; Smith, 1990). Below, are two common examples:

A simple random sample occurs when every subset of n samples (e.g., hauls) has the same probability of being drawn from a finite population of size N. In simple random sampling without replacement (SRSWR) the probability associated with n samples from N population units is given by the inverse binomial coefficient (Cochran, 1977; Lohr, 2021). In simple random sampling, the sample mean is the design-unbiased estimate of the finite population mean. The variance is the finite-population-corrected sample variance (Cochran, 1977; Lohr, 2021).

Stratified sampling occurs when sampling units are grouped into strata (sampling subgroups; e.g., depth bins), samples are taken per strata, and those samples are then combined across strata. Stratified sampling typically produces more precise estimates as the variability within strata is often lower than the total variability (Lohr, 2021). Strata means are combined for the total sample mean by weighting them by the proportion of the total sample units in the strata (Cochran, 1977; Lohr, 2021). Similarly, the total sample variance uses the square of the strata weights (Cochran, 1977; Lohr, 2021). Various allocation methods are available to allocate observations to strata. Proportional allocation allocates observations in proportion to the size (number of sampling units) in the strata, whereas optimal allocation attempts to allocate observations to minimise the variance of the estimator for a given total cost (Lohr, 2021). The sample mean in stratified random sampling weights the strata sample means by the proportion of the observations in each strata (Lohr, 2021).

For all sampling designs though, it is important to keep in mind the recommendations provided in TOR I with respect to survey flexibility.

2.3.4 Model-based inference

Model-based inference assumes that observations can be represented as samples from some statistical or probability model (i.e., a set of mathematical relationships). In contrast to design-based inference where randomness enters only in which sample units are chosen, randomness enters model-based inference by assuming the characteristics of interest (e.g., numbers of fish) are random variables. Under the assumption of independent random variables with an expected population mean and variance, model-based estimators are the same as simple random sampling design-based estimators. However, model-based inference can deal with changes in methodology and are therefore an important tool in fisheries-independent survey data analysis—particularly when facing unavoidable changes to survey design where they may help limit impacts on survey indices (e.g., Yalcin et al. talk). Furthermore, under some conditions, model-based estimators can improve upon design-based estimators, particularly in terms of precision (Petitgas, 2001; Thorson *et al.*, 2015). Here we provide an overview in model-based inference for practitioners. Detailed methodologies and comparisons among models can be found in Thorson (2019b) and Anderson et al. (2022).

2.3.4.1 Distributional assumptions

For model-based inference, we need a model, which (in a parametric setting) starts with a distributional assumption for the data. This depends on the nature of the data (e.g., count, continuous measurements, categories). Here we briefly review common distributional assumptions.

For count data, the Poisson distribution is often assumed (McCullagh and Nelder, 1989) with a rate parameter (lambda) controlling the expected rate per unit of time or space or both. Ecological count data often contain more variability than that expected in the Poisson (termed overdispersion), which is reflected in the common use of the negative binomial, which allows for a gamma-mixture on the Poisson rate parameter (Hilbe, 2011). The expected number of zeros under either a Poisson or negative binomial distribution is sometimes lower than that observed in ecological data, which is termed zero-inflation. Zero-inflated Poisson and zero-inflated negative binomial distributions are designed to deal with the excess of zeros by mixing together a count distribution (Poisson or negative binomial) and an additional quantity of zeros (Lambert, 1992).

For continuous data, the distributional assumption depends on whether the observations are restricted to some space on the real line or not. Common positive continuous distributional assumptions include lognormal or gamma distributions. These distributions are often assumed for catch rate data where the response is in units of mass (Maunder and Punt, 2004). Similarly to count data, there often exists many zeros in continuous data (e.g., no catch of a given species in a haul). Delta models allow for a proportion of zeros in continuous positive data by combining a distribution such as the binomial with the gamma (delta-gamma) or lognormal (delta-lognormal) (Aitchison, 1955; Pennington, 1986; Smith, 1990; Stefánsson, 1996). The Tweedie distribution (Tweedie, 1984) can include discrete count distributions (Poisson) to purely continuous models (gamma)—the Tweedie is also termed the compound Poisson-gamma distribution. Use of the Tweedie in survey index estimation is typically focussed on cases with a mass at zero that are otherwise positive and continuous (e.g., catch weight or biomass density from survey hauls). The Tweedie is being increasingly adopted for index standardisation (e.g., Anderson *et al.*, 2019), in part because of the speed with which it can be fit in the R package TMB (Kristensen *et al.*, 2016). Thorson (2021) showed how distributional assumptions can influence the scale of the estimated abundance indices with delta-gamma and Tweedie performing relatively well compared to design-based estimates.

While much survey index modelling focusses on single species and age-classes, models such as VAST allow multivariate responses that can cover multiple compositional classes (Thorson and Haltuch, 2019) or species (Dolder *et al.*, 2018). VAST allows multiple categories to load on spatial factors explaining spatial or spatiotemporal variability (Thorson, 2019b).

2.3.4.2 Temporal structure

Traditionally, the main goal of fitting models to survey data has been the estimation of abundance indices. This is typically achieved by including independent fixed year effects in the model, which then form the basis of the estimated index (see Maunder and Punt, 2004 in a commercial context). While a wide variety of other structures are possible (e.g., linear or semi-parametric trends over time, random walks, or autoregressive structures), allowing for yearly fixed effects imparts no assumption on the inter-annual changes, leaving this to the stock assessment (Stefánsson, 1996). Other more constrained structures, such as a random walk (e.g., Monnahan et al., 2021), are also possible.

2.3.4.3 Spatial effects

Incorporating spatial effects into model-based survey indices can range from blocking areas (e.g., strata) and simple latitudinal trends to highly flexible random fields. Geostatistical models that assume correlation decreases with distance among observations have been shown to improve upon the precision of stratified/blocked models, as the between-observation variance is explained by spatial variability whereas it would otherwise contribute to the overall variation (Thorson *et al.*, 2015). Both VAST (Thorson, 2019b) and sdmTMB (Anderson *et al.*, 2022) allow for Gaussian random fields that follow a spatial correlation structure (how correlation between observations decays with distance, often modelled assuming a Matérn correlation function). The spatial domain is broken into cells or polygons comprising a mesh specified by a set of knots; and correlation structure, possible anisotropy (correlation decaying at different rates in different directions) are estimated using marginal likelihood with software such as TMB (Kristensen *et al.*, 2016) or (approximate) Bayesian inference with software such as INLA (Lindgren and Rue, 2015). Other common approaches to modelling spatial effects include two and three-dimensional smooths in additive models (e.g., Berg *et al.*, 2014; Wood, 2017).

2.3.4.4 Spatiotemporal effects

When combined with yearly fixed effects, an average spatial field shows how relative abundance or density is spread out in space across the entire modelled time period; to allow the spatial distribution to evolve in time requires spatiotemporal modelling. Predicted abundance at a point in space can vary over time in many ways. Values between years can be assumed to be conditionally independent given the rest of model structure (IID), that is, separate random fields are estimated each year. Alternatively, an autoregressive structure allows the correlation between years for the same location to decay at a given rate over time. An extreme example of this is a random walk where the value next year is the value last year plus some process error. Practically, these methods can allow the spatial field to evolve over time reflecting differences in distribution or spatially correlated effects from missing (latent) variables that evolve over time (e.g., temperature, dissolved oxygen, predator or prey effects).

2.3.4.5 Diagnostics and model selection

Many of the geostatistical models typically used to fit survey data for index standardisation, as described above, are complex statistical models and care needs to be taken that fitting routines have converged, that models appropriately represent the data, and that inference from the models makes sense.

Options for model comparison and selection include information criteria such as AIC (Akaike information criterion) (Akaike, 1973) or BIC (Bayesian information criterion) (Schwarz, 1978), likelihood ratio tests, and cross validation. We note the following:

- Information criteria such as AIC have additional complexities when applied to mixed-effects models and can be applied as marginal AIC, conditional AIC, or various "corrected" conditional AIC versions (Liang *et al.*, 2008; Greven and Kneib, 2010). Future work in geostatistical modelling for survey data could consider such corrected conditional AIC forms that may more appropriately penalise random effect complexity compared to current commonly applied approaches.
- Cross-validation represents a gold standard for model comparison, but can be time consuming for complex geostatistical models and introduces decisions about how to best split data into fitting and testing groups to assess the intended aspect of model performance (e.g., random, spatial blocking, time blocking, space and time blocking, block size) (e.g., Commander *et al.*, 2022). As an example, in the context of this workshop, prediction across missing areas due to survey restrictions (e.g., MPAs [marine protected areas] or wind farms), may be best tested by cross-validation with spatial blocking with the spatial blocks of approximation the same size as the restricted areas.
- As an alternative to or in addition to model selection, parameters can be penalised or shrunk towards zero through the use of priors (Bayesian inference) or penalties (likelihood inference) and various regularisation approaches such as ridge regression (Hoerl and Kennard, 1970). Penalised complexity priors are a recent area of research that may be particularly helpful in the context of penalising complexity in models with random fields (Fuglstad *et al.*, 2019). Penalised complexity priors (and traditional priors or penalties) are available in INLA (Lindgren and Rue, 2015) and sdmTMB (Anderson *et al.*, 2022) for example.

There are many approaches to assessing if models are consistent with the data and that inference from models makes sense. Options include residual inspection, self-simulation experiments, retrospective peels, comparing model-based indices to design-based indices, and consulting with species experts. We note the following:

- Residual diagnostics for geostatistical models, especially those involving spatial structured random effects estimated via the Laplace approximation, are a complex and ongoing topic of research. Current best-practices for such models include one-step-ahead prediction residuals (Thygesen *et al.*, 2017) and MCMC (Markov chain Monte Carlo) residuals, possibly with fixed-effects fixed at their MLEs (Rufener *et al.*, 2021), although both can be slow to calculate in practice. Simulation-based residuals are possible (e.g., facilitated with the DHARMa R package; Hartig, 2021), although it remains unclear the reliability of such residuals for these models.
- Self-simulation experiments can be used to evaluate the ability for models to be self-consistent and recover parameters if new data are simulated from a fitted model; cross-model simulation experiments can be used to assess the consequences of model misspecification or the robustness of certain modelling choices (e.g., Thorson *et al.*, 2021b). Retrospective peels—successively removing trailing years of data and refitting the model—can be used to assess self-consistency and the presence of retrospective model patterns.

- Finally, model-based inference from surveys can be checked by comparing it to design-based inference when appropriate—particularly for identifying issues with abundance or biomass scale (e.g., Thorson *et al.*, 2021b). Species experts can be consulted to ensure that predicted species distributions and covariate relationships (e.g., relationships with depth or environmental variables) appear reasonable given prior knowledge.

2.3.5 Simulation modelling

Planning and implementing fisheries-independent surveys is a non-trivial task as such surveys often span large areas, sample multiple species, and conduct sub-sampling for biological measures. It is not always possible to determine accurate and pragmatic sampling strategies using analytical methods and, as such, simulations have frequently been applied for planning and risk assessment purposes. In the past, the predominant use of simulation methods and models was to ensure appropriate allocation of effort for a given survey design and sampling domain (e.g., Schnute and Haigh, 2003). However, planned and unplanned changes present challenges to the continuity and quality of the important survey data products and can lead to greater bias and imprecision in their estimates. Another source of uncertainty is the rapidly changing spatial and temporal dynamics of fish populations. Although these complexities are acknowledged, the majority of simulation experiments conducted to date lack the spatial resolution to assess spatially blocked effort reductions or species distribution shifts. Simulation modelling that accounts for spatial dynamics is increasingly needed and will play an important role in the testing of emerging geostatistical modelling techniques for the standardisation of indices and composition data and combining data sources, planning optimised survey designs over space and time and among species, contingency planning for anticipated in-season obstacles, and evaluating the impact of survey changes on assessment outcomes through to the management system and reliant fisheries.

Developing a realistic operating model that can simulate data with similar dynamics properties can be a daunting task because of all the complexities behind species distributions and multi-level sampling techniques. Despite all of the nuances, practitioners frequently face scenarios where surveys are disrupted (e.g., exclusion from an MPA; Benoît *et al.*, 2020) or require improvements and the impacts of substantive changes need to be assessed. Practitioners can apply two general simulation approaches to address changes in survey design and operation and to quantify their impacts by either resampling existing observed data or through model-based simulations. Generally, model-based simulation allows analysts to simulate data based on the fitted model input under different conditions, producing simulation scenarios which can be compared with one another, usually involving a base scenario. For reduced survey effort in particular, simulation approaches are valuable in evaluating for example i) the effect of including unconventional data input (e.g., oceanographic satellite data or trophic interactions; Pennino *et al.*, 2016; Barber *et al.*, 2021), ii) the performance of various forms of combined data (ref TOR II), iii) survey design optimisation scenarios, iv) contingency planning for anticipated in-season obstacles, and v) the impact of survey changes on assessment outcomes through to the management system and reliant fisheries.

2.3.5.1 Resampling

Resampling methods (e.g, bootstrapping; sampling with replacement) can be used to compare the precision and bias of design- or model-based indices across a range of alternative sampling scenarios. With resampling, the analyst uses observed data to generate a large number of simulated samples. A major benefit of this simulation approach is that it implicitly produces realistic samples as they are based on real-world data. For instance, results from a recent resampling study focused on whiting in the Celtic Sea (see Stroh et al.) helped demonstrate the ability of spatiotemporal models to remain unbiased, albeit more uncertain, when survey effort is reduced, even if there is a non-random loss of a survey. Another resampling analysis of data collected in the Gulf of Alaska reiterated the impact of the number of hauls (primary sampling unit) on the variance of age-disaggregated indices; the age sub-sampling protocol had a relatively minor impact (Sikey et al., Siskey *et al.*, 2022). Jourdain *et al.* (2020) used similar bootstrapping methods to evaluate sampling protocols for Atlantic cod (*Gadus morhua*) in the North Sea and results were similar to Siskey *et al.* (2022). Other key benefits of resampling approaches is that they are relatively quick to implement and a complex operating model does not need to be defined. It is important to note that resampling approaches assume that previous samples were representative of the underlying population and sampling process.

2.3.5.2 Model-based simulations

A limitation of the resampling approach is that the samples must conform to the historical design. That is, resampling cannot be used to simulate samples from an alternate sampling design (e.g., systematic samples from a stratified-random sample). Another limitation is that the "truth" remains unknown and, as such, performance must be measured against the best available information, which may be insufficient. It is therefore important to consider whether previously collected samples are sufficiently representative of the population and sampling process and whether it is necessary to know the "truth". When doubts loom regarding the previous sampling, practitioners often turn to model-based simulations. When developing a simulation study, it is important to keep the operating model as simple as possible to meet the study objectives. Depending on the study, it may be useful to focus on mechanisms, empirical dynamics, or a combination of both.

Realistic population dynamics and species distributions can be generated using "mechanistic" models that explicitly define functions for processes such as cohort tracking or environmental associations (Kearney and Porter, 2009). If the objective is to test the performance of a species distribution model, it may be useful to simulate samples from an environmental suitability map derived from species-environment relationships (e.g., Leroy *et al.*, 2016; Regular *et al.*, 2020, Regular and Anderson talk, Yalcin *et al.* talk). The biggest hurdle with this approach lies with the process of defining response functions to various environmental variables (covariates). While it is conceptually appealing to define these functional relationships, the approach becomes difficult to apply in cases where functional relationships with the environment are not well understood. The same limitations apply if population dynamics need to be simulated as the functional relationships between spawning stock biomass and recruitment (Subbey *et al.*, 2014) or environmental variables and natural mortality (Johnson *et al.*, 2015) are rarely well defined. It is therefore often convenient and necessary to describe population and spatial dynamics using stochastic processes; that is, apply an empirical approach.

"Empirical" simulations implicitly capture many of the temporal and spatial dependencies apparent in real survey data by adding stochasticity using distributional assumptions. For example, noisy survey samples are frequently described using a mixture of the binomial and gamma distributions (Aitchison, 1955; Schnute and Haigh, 2003). There are a growing number of general tools (see TOR IV) that facilitate the fitting of empirical models to data and these same models can be used to simulate observations throughout the survey domain. This offers a relatively straightforward way to generate observations collected using different strategies. Using the geostatistical model behind VAST, Oyafuso et al. (2021) modelled the distributions of 14 species in the Gulf of Alaska and simulated samples from surveys with different stratifications and allocations. In doing so, they demonstrated that if survey priorities change, objective modifications could be made to the stratification and allocation of the survey to lower the survey CV of selected species, potentially at the cost of other species. Within species, one might focus on size-composition and assess the merits of alternate sampling protocol or estimators. Size-specific indices can be obtained from geostatistical models by supplying age or length disaggregated haul-level data. This requires the intervening application of design-based expansions of sub-sampled data (e.g., age length keys). A simulation self-test of this approach using VAST indicated that age-disaggregated model-based indices are comparable, and perhaps slightly more accurate, to design-based indices (Thorson talk; Thorson and Haltuch, 2019). If inefficiencies in sub-sampling strategies are suspected, simulations of alternate multi-level sampling of size-specific species distribution models may reveal unexpected flaws in the sampling design (see case study in Regular *et al.*, 2020).

2.3.5.3 Simulating management consequences

Multiple resampling and model-based simulation studies have demonstrated that changes in the design and implementation of surveys will either improve or degrade the accuracy (bias or precision) of survey data products. When changes to the performance are observed, a natural next question to ask is "Does it matter?" In many cases it may be obvious that the change is so small that it will have a negligible impact on the assessment of the stock; however, major changes to a survey may result in levels of bias and variance that are too big to ignore. This is where open or closed-loop simulations become a useful approach to apply with the former assessing the consequences to stock assessment estimates and the latter including feedback between the survey index, assessment model, and management strategy on the projected population allowing practitioners to evaluate the risks associated with a changing survey.

Open-loop simulations are a powerful tool to evaluate the performance of stock assessment models (Hilborn and Walters, 1987; Magnusson and Hilborn, 2007; Deroba *et al.*, 2015). Performance is often quantified as bias and imprecision in quantities of interest to fisheries managers including current stock size and management reference points. This simulation approach is done using an operating model to generate the true population and observation dynamics of a population for a specified period. The operating model can be conditioned on an existing stock assessment model or developed independently and include more complexity. Existing tools include ss3sim (Anderson, *et al.*, 2014) and FLR (Kell *et al.*, 2007) for example. Survey observations are generated from the observation dynamics model and used as inputs in the assessment model. The impact of survey information can be evaluated by simulating scenarios with different levels of observation error or bias in indices of abundance and composition data or removing entire time series. The quantities estimated by the assessment model are then compared to the true dynamics from the OM. Simulation-estimation experiments have been conducted to evaluate the impact of different data types and their availability on stock assessment outcomes. For example, Ono *et al.* (2015) found survey information was more informative in assessments when conducted less frequently over more years than more frequently over fewer years. Muradian *et al.* (2019) conducted a simulation experiment where scenarios dropped one of four surveys to evaluate the trade-off in the cost of conducting the individual surveys and the information provided to the assessment model and, as a result, were able to rank the value of each sampling program. Studies such as these can help to justify the continued support for the collection of survey data. In some cases, we want to understand how changes in survey design will impact the management decision making process, this is where closed-loop simulations are a preferred method.

Closed-loop simulations are central to the Management Strategy Evaluation (MSE) approach where scientists and resource managers collaboratively develop a series of operating models, estimation models, and candidate management strategies to test the robustness of various policies to sources of uncertainty (Smith, 1994; Punt *et al.*, 2016). While this approach is increasingly common, operating models have largely generated spatially aggregated indices at age. This has limited the analysts' ability to evaluate the management implications of survey changes (i.e., spatially explicit survey data of varying quality are rarely simulated). Within the spatially aggregated MSE context, the "shortcut approach" (ICES, 2020b) enables the evaluation of some change by simulating survey data with error using parameter estimates from a distribution (parametric), or resampling from the data (non-parametric). If the purpose of the simulation exercise is to evaluate a change in survey design (for example, moving from a systematic to stratified random survey design), resampling approaches are no longer appropriate for the observation model and the dynamics of the existing survey should be included in the simulation. In this case existing models like MixFishSim (Dolder *et al.*, 2020) or spatialSim (Nottingham and Millar, 2021) can simulate population, survey, and fishery dynamics mechanistically. Harford and Babcock (2016) describe one example where spatially explicit samples were simulated using an individual-based model within an MSE framework to evaluate management risks under different data sampling scenarios. Tools for implementing closed-loop simulation include FLR (Kell *et al.*, 2007), openMSE (formerly DLMtool; Carruthers and Hordyk, 2018), and SSMSE (Doering and Vaughan, 2022).

2.3.6 Recommendations and future work

Here we consider recommendations and future work based on the talks and discussion within this TOR. We first consider recommendations that apply primarily to simulations. Then, since models and simulations often go hand-in-hand, with simulations being used to test estimation models, and fitted estimation models being used to generate survey simulations, we consider recommendations that apply to both.

Recommendations and future work for simulation modelling:

1. **Establish clearly defined goals early in the process and let those goals guide model or simulation complexity.** For example, feedback in the simulation (i.e., closed-loop simulation) is needed when the goal is to evaluate the impacts of survey changes on socio-economic and management objectives but would introduce unnecessary complexity in other circumstances. Evaluations of tactical changes and their impacts on stock assessment outcomes can often be effectively addressed by resampling existing data and using open-loop simulation experiments.
2. **Involve those working in survey implementation, survey modelling, and stock assessment when designing survey simulations if possible.** Identifying key uncertainties is an important component in the development of simulation scenarios. This is where each representative's perspective will be impactful. For example, group discussions led to hypotheses about directional movement due to climate change (Cronin-Fine *et al.* talk) and wind energy development (Miller *et al.* talk) that represented important process-level uncertainties in the respective systems. We can expect an emergent property of the simulation will be uncertainty, bias, or both in survey products due to the interaction between process uncertainties and survey design.
3. **Use closed-loop simulation to quantify the value of information in current surveys and assess the degree to which reductions in survey effort are expected to change probabilities of achieving management and conservation objectives.** It is difficult to infer consequences of increased index uncertainty and bias on management and economic consequences without formal analyses. Value of information analyses (Harford and Babcock, 2016, Cronin-Fine talk) with closed-loop simulation can help justify survey programs, highlight critical parts of surveys to retain, and design reduced survey programs such that they have minimal impact on conservation and economic consequences.

4. **Consider all data streams resulting from a survey—not just indices of abundance for assessments.** For example, indices derived from large suites of species are important for ecosystem modelling, indices of predator or prey species abundance provide important context when modelling a species of interest (Pennino *et al.* talk), survey data are often used for marine spatial planning, and biological samples (length, weight, age, and maturity samples) are also critical data streams for stock assessment (Thorson talk).
5. **Maintain both resampling and model-based simulations.** Resampling and model-based approaches each have their strengths and weaknesses. Resampling approaches are fast to implement and realistic complexities are "baked in" for free; however, the underlying truth is not known and sampling schemes are restricted based on historical sampling. Model-based approaches know the underlying truth and can explore entirely new sampling scenarios or conditions but only incorporate the complexities in states of nature that are explicitly included.
6. **Improve conditioning of "mechanistic" simulation model elements through fitted empirical models.** Simulation models such as the model underlying the SimSurvey package (Regular *et al.*, 2020) incorporate processes and mechanisms that are more complicated than can typically be fit to data. On the other hand, related empirical models can be fit to characterise aspects of such simulation models. Therefore, there is much potential in "linking" empirical fitted models to more "mechanistic" or semi-mechanistic simulations to condition aspects of the models (Regular and Anderson talk). For example, abundance relationships with depth or random field properties can be characterised by related empirical models.

Recommendations and future work for estimation and simulation modelling:

1. **Continue development of models and simulations that incorporate movement.** Understanding changes in fish migration and distribution under climate change is a major challenge and movement is a primary cause for changes in availability and resulting survey catchability. Jim Thorson presented on the inclusion of fish movement in models of surveyed abundance via diffusion-advection-taxis (Thorson et al., 2021a). Such models have the potential to improve inference about movement processes and also make better predictions of spatial distribution—potentially after sections of the survey domain have been excluded from survey effort (e.g. for MPAs, Benoît et al., 2020). Movement can also be implicitly included in survey models through, for example, seasonal spatiotemporal random fields (Thorson talk).
2. **Consider possible improvements in realism and inference from multivariate, joint (dynamic) species distribution models (JSDMs).** In some instances, JSDMs can improve model predictions, particularly for more poorly sampled species (Thorson and Barnett, 2017; Pennino et al. talk), but a primary benefit may be improved ecological inference (e.g., Dolder et al., 2018; Thompson et al., 2022). In the context of unavoidable reductions in survey effort, JSDMs may have applicability in reducing uncertainty loss for more poorly sampled species and using correlations between species established in years before survey effort reductions to improve inference after effort reductions.
3. **Consider including spatially varying coefficients in SDMs of survey data to make better local predictions of missing survey observations.** Spatially varying coefficient models allow covariate effects on the response variable to vary smoothly in space (Hastie and Tibshirani, 1993; Gelfand et al., 2003, Thorson talk). In the context of fish surveys, this could, for example, let regional climate indices help predict the local distribution of a species (Thorson, 2019a) or allow for spatially varying trends in abundance (Barnett et al., 2021).

4. **Continue to develop approaches and make existing approaches more accessible for combining multiple data types, including incorporating data sources with preferential sampling such as commercial catches into survey modelling.** In WKUSER2, Iosu Paradinas (OSU) presented a joint modelling approach to combine trawl and trammel net surveys, and Nathan Bacheler (NOAA SEFSC) presented the challenges of combining paired camera and trap survey data in the Atlantic from North Carolina to Florida. The field of integrated SDM modelling is developing rapidly (Isaac et al., 2020) and we have begun to see applications within fisheries-independent survey modelling (e.g., Grüss and Thorson, 2019; Rufener et al., 2021). Integrated SDMs have the potential to help fill gaps incurred by unavoidable reductions in survey effort (see more discussion TOR II). In line with this point, we recommend using and further developing tools such as MixFishSim (Dolder et al., 2020) and spatialSim (Nottingham and Millar, 2021) that can simulate survey and fishery dynamics simultaneously to test integrated modelling approaches.
5. **Continue developing approaches for propagating uncertainty from survey data through to indexes of abundance and stock assessment.** TOR II of the first WKUSER meeting focused on total survey uncertainty, which is a value that incorporates error from biophysical processes, observation processes, and analytical considerations (ICES, 2020a, pp 18–30). TOR III from WKUSER2 largely focused on analytical considerations for survey index production, particularly geostatistical methods that attempt to estimate uncertainty and mitigate bias introduced by flaws in the implementation of a survey. While a number of simulation studies have shown that these methods produce confidence intervals with expected rates of coverage (e.g., Jardim and Ribeiro, 2007; Yalcin et al. talk), further research is required to ensure that error introduced by pre-processing survey data for length- or age-selectivity calibrations (e.g., Webster et al., 2020) or age-length key conversions (e.g., Thorson and Haltuch, 2019; Thorson talk) is accounted for and propagated through to model-based indices and assessments. Studies that simulate size-structured spatiotemporal populations may be useful for assessing confidence interval coverage of length- or age-disaggregated survey indices from emerging model-based methods.
6. **Develop a more rigorous and standard process for model diagnostics, validation, and selection.** There are many decisions a practitioner has to make in the process of fitting the complex SDMs that are frequently used to model survey data (e.g., GLMMs with random fields or GAMMs) and these decisions can have meaningful consequences on inference (e.g., Commander et al., 2022). Models with correlated random effects, and especially such models involving the Laplace approximation, can introduce several complexities with model diagnostics (Thygesen et al., 2017). It would help practitioners to have more guidance on assessing the fit of the geostatistical models often used for the purpose of modelling survey data.
7. **Maintain survey designs such that design-based estimators can still be applied if at all possible, given unavoidable survey effort reduction.** While model-based estimators can account for many inconsistencies in survey design or implementation, they are likely to perform best with data that also work for design-based estimators, and designed-based estimators have several advantages including being trusted, intuitive, easy to calculate, and easy to communicate.
8. **Develop open-source well-documented user-friendly tools for new models and simulation approaches.** This encourages rapid uptake of approaches among practitioners and allows the community to efficiently build on past work. This topic is covered in detail within TOR IV.

2.3.7 References

Aitchison, J. 1955. On the distribution of a positive random variable having a discrete probability mass at the origin. Journal of the American Statistical Association, 50: 901.

Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle. In Second International Symposium on Information Theory, pp. 267–281. Akademiai Kiado, Budapest.

Anderson, S. C., Monnahan, C. C., Johnson, K. F., Ono, K., and Valero, J. L. 2014. ss3sim: an R package for fisheries stock assessment simulation with Stock Synthesis. PLOS ONE, 9: e92725.

Anderson, S. C., Keppel, E. A., and Edwards, A. M. 2019. A reproducible data synopsis for over 100 species of British Columbia groundfish. DFO Can. Sci. Advis. Sec. Res. Doc., 2019/041: vii + 321 p.

Anderson, S. C., Ward, E. J., English, P. A., and Barnett, L. A. K. 2022, March 27. sdmTMB: an R package for fast, flexible, and user-friendly generalized linear mixed effects models with spatial and spatiotemporal random fields. bioRxiv. https://doi.org/10.1101/2022.03.24.485545.

Babyn, J., Varkey, D., Regular, P., Ings, D., and Mills Flemming, J. 2021. A gaussian field approach to generating spatial age length keys. Fisheries Research, 240: 105956.

Barber, X., Conesa, D., López-Quílez, A., Martínez-Minaya, J., Paradinas, I., and Pennino, M. G. 2021. Incorporating Biotic Information in Species Distribution Models: A Coregionalized Approach. Mathematics, 9: 417.

Barnett, L. A. K., Ward, E. J., and Anderson, S. C. 2021. Improving estimates of species distribution change by incorporating local trends. Ecography, 44: 427–439.

Benoît, H. P., Dunham, A., Macnab, P., Rideout, R., Wareham, V., Clark, D., Duprey, N., et al. 2020. Elements of a framework to support decisions on authorizing scientific surveys with bottom contacting gears in protected areas with defined benthic conservation objectives. DFO Can. Sci. Advis. Sec. Res. Doc., 2020/011.

Berg, C. W., Nielsen, A., and Kristensen, K. 2014. Evaluation of alternative age-based methods for estimating relative abundance from survey data in relation to assessment models. Fisheries Research, 151: 91–99.

Carruthers, T. R., and Hordyk, A. R. 2018. The Data-Limited Methods Toolkit (DLMtool): An R package for informing management of data-limited populations. Methods in Ecology and Evolution, 9: 2388–2395.

Cochran, W. G. 1977. Sampling Techniques. Wiley Series in Probability and Mathematical Statistics - Applied. John Wiley & Sons. N.Y. 428 pp.

Commander, C. J. C., Barnett, L. A. K., Ward, E. J., Anderson, S. C., and Essington, T. E. 2022. The shadow model: how and why small choices in spatially explicit species distribution models affect predictions. PeerJ, 10: e12783. PeerJ Inc.

Deroba, J. J., Butterworth, D. S., Methot, R. D., De Oliveira, J. A. A., Fernandez, C., Nielsen, A., Cadrin, S. X., et al. 2015. Simulation testing the robustness of stock assessment models to error: some results from the ICES strategic initiative on stock assessment methods. ICES Journal of Marine Science, 72: 19–30.

Doering, K., and Vaughan, N. 2022. Management Strategy Evaluation (MSE) using Stock Synthesis (SS). https://github.com/nmfs-fish-tools/SSMSE.

Dolder, P. J., Thorson, J. T., and Minto, C. 2018. Spatial separation of catches in highly mixed fisheries. Scientific Reports, 8: 13886.

Dolder, P. J., Minto, C., Guarini, J.-M., and Poos, J. J. 2020. Highly resolved spatiotemporal simulations for exploring mixed fishery dynamics. Ecological Modelling, 424: 109000.

Fuglstad, G.-A., Simpson, D., Lindgren, F., and Rue, H. 2019. Constructing Priors that Penalize the Complexity of Gaussian Random Fields. Journal of the American Statistical Association, 114: 445–452. Taylor & Francis.

Gelfand, A. E., Kim, H.-J., Sirmans, C. F., and Banerjee, S. 2003. Spatial modeling with spatially varying coefficient processes. Journal of the American Statistical Association, 98: 387–396. Taylor & Francis.

Greven, S., and Kneib, T. 2010. On the behaviour of marginal and conditional AIC in linear mixed models. Biometrika, 97: 773–789.

Grüss, A., and Thorson, J. T. 2019. Developing spatio-temporal models using multiple data types for evaluating population trends and habitat usage. ICES Journal of Marine Science, 76: 1748–1761.

Gunderson, D. M. 1993. Surveys of Fisheries Resources. Wiley, New York. https://www.wiley.com/en-us/Surveys+of+Fisheries+Resources-p-9780471547358 (Accessed 4 October 2022).

Harford, W. J., and Babcock, E. A. 2016. Aligning monitoring design with fishery decision-making: examples of management strategy evaluation for reef-associated fisheries. Aquatic Living Resources, 29: 205.

Hartig, F. 2021. DHARMa: Residual diagnostics for hierarchical (multi-level/mixed) regression models. manual. https://CRAN.R-project.org/package=DHARMa.

Hastie, T., and Tibshirani, R. 1993. Varying-Coefficient Models. Journal of the Royal Statistical Society. Series B (Methodological), 55: 757–796. [Royal Statistical Society, Wiley].

Hilbe, J. M. 2011. Negative Binomial Regression. Cambridge University Press.

Hilborn, R., and Walters, C. J. 1987. A General Mode for Simulation of Stock and Fleet Dynamics in Heterogeneous Fisheries. Canadian Journal of Fisheries and Aquatic Sciences, 44: 1366–1369.

Hoerl, A. E., and Kennard, R. W. 1970. Ridge regression: biased estimation for nonorthogonal problems. Technometrics, 12: 55–67. [Taylor & Francis, Ltd., American Statistical Association, American Society for Quality].

ICES. 2019. International Bottom Trawl Survey Working Group (IBTSWG). ICES. https://ices-library.figshare.com/articles/_/18617900 (Accessed 6 October 2022).

ICES. 2020a. ICES Workshop on unavoidable survey effort reduction (WKUSER). ICES Scientific Reports, 2: 72.

ICES. 2020b. The third Workshop on Guidelines for Management Strategy Evaluations (WGMSE3). ICES. https://ices-library.figshare.com/articles/_/18621194 (Accessed 17 September 2022).

ICES. 2022. International Bottom Trawl Survey Working Group (IBTSWG). ICES Scientific Reports. https://ices-library.figshare.com/articles/report/International_Bottom_Trawl_Survey_Working_Group_IBTSWG_/20502828 (Accessed 6 October 2022).

Isaac, N. J. B., Jarzyna, M. A., Keil, P., Dambly, L. I., Boersch-Supan, P. H., Browning, E., Freeman, S. N., et al. 2020. Data integration for large-scale models of species distributions. Trends in Ecology & Evolution, 35: 56–67.

Jardim, E., and Ribeiro, P. J. 2007. Geostatistical assessment of sampling designs for Portuguese bottom trawl surveys. Fisheries Research, 85: 239–247.

Johnson, K. F., Monnahan, C. C., McGilliard, C. R., Vert-pre, K. A., Anderson, S. C., Cunningham, C. J., Hurtado-Ferro, F., et al. 2015. Time-varying natural mortality in fisheries stock assessment models: identifying a default approach. ICES Journal of Marine Science, 72: 137–150.

Jourdain, N. O. A. S., Breivik, O., Fuglebakk, E., Aanes, S., and Vølstad, J. H. 2020. Evaluation of sampling strategies for age determination of cod (Gadus morhua) sampled at the North Sea International Bottom Trawl Survey. ICES Journal of Marine Science, 77: 859–869.

Kearney, M., and Porter, W. 2009. Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. Ecology Letters, 12: 334–350.

Kell, L. T., Mosqueira, I., Grosjean, P., Fromentin, J.-M., Garcia, D., Hillary, R., Jardim, E., et al. 2007. FLR: an open-source framework for the evaluation and development of management strategies. ICES Journal of Marine Science, 64: 640–646.

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., and Bell, B. 2016. TMB: Automatic Differentiation and Laplace Approximation. Journal of Statistical Software, 70. http://arxiv.org/abs/1509.00660 (Accessed 16 September 2022).

Lambert, D. 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics, 34: 1–14. [Taylor & Francis, Ltd., American Statistical Association, American Society for Quality].

Leroy, B., Meynard, C. N., Bellard, C., and Courchamp, F. 2016. virtualspecies, an R package to generate virtual species distributions. Ecography, 39: 599–607.

Liang, H., Wu, H., and Zou, G. 2008. A note on conditional AIC for linear mixed-effects models. Biometrika, 95: 773–778.

Lindgren, F., and Rue, H. 2015. Bayesian spatial modelling with R-INLA. Journal of Statistical Software, 63: 1–25.

Lohr, S. L. 2021. Sampling: Design and Analysis. Chapman and Hall/CRC, Boca Raton, FL.

Magnusson, A., and Hilborn, R. 2007. What makes fisheries data informative? F I S H and F I S H E R I E S: 22.

Maunder, M. N., and Punt, A. E. 2004. Standardizing catch and effort data: a review of recent approaches. Fisheries Research, 70: 141–159.

McCullagh, P., and Nelder, J. A. 1989. Generalized linear models. Chapman {\&} Hall, New York.

Monnahan, C. C., Thorson, J. T., Kotwicki, S., Lauffenburger, N., Ianelli, J. N., and Punt, A. E. 2021. Incorporating vertical distribution in index standardization accounts for spatiotemporal availability to acoustic and bottom trawl gear for semi-pelagic species. ICES Journal of Marine Science, 78: 1826–1839.

Muradian, M. L., Branch, T. A., and Punt, A. E. 2019. A framework for assessing which sampling programmes provide the best trade-off between accuracy and cost of data in stock assessments. ICES Journal of Marine Science, 76: 2102–2113.

Nottingham, C. D., and Millar, R. B. 2021, September 5. spatialSim: multi-species spatiotemporal size-structured operating model for management strategy evaluation. arXiv. http://arxiv.org/abs/2109.02228 (Accessed 25 September 2022).

O'Leary, C. A., Thorson, J. T., Ianelli, J. N., and Kotwicki, S. 2020. Adapting to climate-driven distribution shifts using model-based indices and age composition from multiple surveys in the walleye pollock (Gadus chalcogrammus) stock assessment. Fisheries Oceanography, 29: 541–557.

Ono, K., Licandeo, R., Muradian, M. L., Cunningham, C. J., Anderson, S. C., Hurtado-Ferro, F., Johnson, K. F., et al. 2015. The importance of length and age composition data in statistical age-structured models for marine species. ICES Journal of Marine Science, 72: 31–43.

Oyafuso, Z. S., Barnett, L. A. K., and Kotwicki, S. 2021. Incorporating spatiotemporal variability in multispecies survey design optimization addresses trade-offs in uncertainty. ICES Journal of Marine Science, 78: 1288–1300.

Pennington, M. 1986. Some statistical techniques for estimating abundance indices from trawl surveys. Fisheries Bulletin: 519–525.

Pennington, M., and Strømme, T. 1998. Surveys as a research tool for managing dynamic stocks. Fisheries Research, 37: 97–106.

Pennino, M. G., Conesa, D., López-Quílez, A., Muñoz, F., Fernández, A., and Bellido, J. M. 2016. Fishery-dependent and -independent data lead to consistent estimations of essential habitats. ICES Journal of Marine Science, 73: 2302–2310.

Petitgas, P. 2001. Geostatistics in fisheries survey design and stock assessment: models, variances and applications. Fish and Fisheries, 2: 231–249.

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

Punt, A. E., Butterworth, D. S., de Moor, C. L., De Oliveira, J. A. A., and Haddon, M. 2016. Management strategy evaluation: best practices. Fish and Fisheries, 17: 303–334.

Regular, P. M., Robertson, G. J., Lewis, K. P., Babyn, J., Healey, B., and Mowbray, F. 2020. SimSurvey: An R package for comparing the design and analysis of surveys by simulating spatially-correlated populations. PLOS ONE, 15: e0232822. Public Library of Science.

Rufener, M.-C., Kristensen, K., Nielsen, J. R., and Bastardie, F. 2021. Bridging the gap between commercial fisheries and survey data to model the spatiotemporal dynamics of marine species. Ecological Applications, In press: e02453.

Särndal, C.-E., Thomsen, I., Hoem, J. M., Lindley, D. V., Barndorff-Nielsen, O., and Dalenius, T. 1978. Design-based and model-based inference in survey sampling [with discussion and reply]. Scandinavian Journal of Statistics, 5: 27–52. [Board of the Foundation of the Scandinavian Journal of Statistics, Wiley].

Schnute, J. T., and Haigh, R. 2003. A simulation model for designing groundfish trawl surveys. Canadian Journal of Fisheries and Aquatic Sciences, 60: 640–656.

Schwarz, G. 1978. Estimating the dimension of a model. The Annals of Statistics, 6: 461–464. Institute of Mathematical Statistics.

Shelton, A. O., Thorson, J. T., Ward, E. J., and Feist, B. E. 2014. Spatial semiparametric models improve estimates of species abundance and distribution. Canadian Journal of Fisheries and Aquatic Sciences, 71: 1655--1666.

Siskey, M., Punt, A. E., Hulson, P.-J. F., Bryan, M. D., Ianelli, J. N., and Thorson, J. T. 2022. The estimated impact of changes to otolith field-sampling and ageing effort on stock assessment inputs, outputs, and catch advice. Canadian Journal of Fisheries and Aquatic Sciences. https://cdnsciencepub.com/doi/abs/10.1139/cjfas-2022-0050 (Accessed 5 October 2022).

Smith, A. D. M. 1994. Management strategy evaluation – the light on the hill. In Population Dynamics for Fisheries Management.

Smith, S. J. 1990. Use of statistical models for the estimation of abundance from groundfish trawl survey data. Canadian Journal of Fisheries and Aquatic Sciences, 47: 894–903. NRC Research Press.

Stefánsson, G. 1996. Analysis of groundfish survey abundance data: combining the GLM and delta approaches. ICES Journal of Marine Science, 53: 577–588.

Subbey, S., Devine, J. A., Schaarschmidt, U., and Nash, R. D. M. 2014. Modelling and forecasting stock–recruitment: current and future perspectives. ICES Journal of Marine Science, 71: 2307–2322.

Thompson, P. L., Anderson, S. C., Nephin, J., Haggarty, D. R., Peña, M. A., English, P. A., Gale, K. S. P., et al. 2022. Disentangling the impacts of environmental change and commercial fishing on demersal fish biodiversity in a northeast Pacific ecosystem. Marine Ecology Progress Series, 689: 137–154.

Thorson, J., Shelton, A., Ward, E., and Skaug, H. 2015. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. ICES Journal of Marine Science, 72.

Thorson, J. T., and Barnett, L. A. K. 2017. Comparing estimates of abundance trends and distribution shifts using single- and multispecies models of fishes and biogenic habitat. ICES Journal of Marine Science, 74: 1311–1321.

Thorson, J. T. 2019a. Measuring the impact of oceanographic indices on species distribution shifts: The spatially varying effect of cold-pool extent in the eastern Bering Sea. Limnology and Oceanography, 64: 2632–2645.

Thorson, J. T. 2019b. Guidance for decisions using the Vector Autoregressive Spatio-Temporal (VAST) package in stock, ecosystem, habitat and climate assessments. Fisheries Research, 210: 143–161.

Thorson, J. T., and Haltuch, M. A. 2019. Spatiotemporal analysis of compositional data: increased precision and improved workflow using model-based inputs to stock assessment. Canadian Journal of Fisheries and Aquatic Sciences, 76: 401–414. NRC Research Press.

Thorson, J. T., Barbeaux, S. J., Goethel, D. R., Kearney, K. A., Laman, E. A., Nielsen, J. K., Siskey, M. R., et al. 2021a. Estimating fine-scale movement rates and habitat preferences using multiple data sources. Fish and Fisheries, 22: 1359–1376.

Thorson, J. T., Cunningham, C. J., Jorgensen, E., Havron, A., Hulson, P.-J. F., Monnahan, C. C., and von Szalay, P. 2021b. The surprising sensitivity of index scale to delta-model assumptions: Recommendations for model-based index standardization. Fisheries Research, 233: 105745.

Thygesen, U. H., Albertsen, C. M., Berg, C. W., Kristensen, K., and Nielsen, A. 2017. Validation of ecological state space models using the Laplace approximation. Environmental and Ecological Statistics, 24: 317–339.

Tweedie, M. C. K. 1984. An index which distinguishes between some important exponential families. In Statistics: Applications and New Directions. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference, pp. 579–604. Ed. by J. K. Gosh and J. Roy. Indian Statistical Institute, Calcutta.

Webster, R. A., Soderlund, E., Dykstra, C. L., and Stewart, I. J. 2020. Monitoring change in a dynamic environment: spatiotemporal modelling of calibrated data from different types of fisheries surveys of Pacific halibut. Canadian Journal of Fisheries and Aquatic Sciences. https://cdnsciencepub.com/doi/abs/10.1139/cjfas-2019-0240 (Accessed 13 October 2020).

Wood, S. N. 2017. Generalized additive models: An introduction with R. Chapman and Hall/CRC.

2.4 TOR IV. "Tools and technology development": Describe technological and analytical tools (e.g. R packages, AI, video analysis, etc.) that can provide quantitative assessment of the effect of effort changes on the quality of survey deliverables and advisory products

Sub-Group Leads: Espen Johnsen, Zack Oyafuso

Participants: Emily Markowitz, Nathan Bacheler, Derek Bolser, Kristan Blackhart, Lukas DeFilippo, Matthew Siskey

The goal of this TOR is to review the technology and analytical tools that are available to assess the effect of survey changes on the quality of survey deliverables, stock assessments, and advisory products. During the plenary session of WKUSER2, a number of presentations discussed the application of available tools, techniques, and technologies that would help researchers achieve the goal of TOR IV. In addition, discussions that stemmed from them also highlighted possible gaps that are present in the current suite of tools (e.g., availability and accessibility of tools). We intend to provide advice in order to guide survey and stock assessment scientists in their analyses given their objectives and data availability. Below, we discuss the recommendations for survey tools and technology across three main themes:

- *2.4.1 Evaluation:* Effects of changes in survey designs/total effort on survey products and assessment/management outputs.
- *2.4.2 Mitigation:* How to mitigate the effects of survey changes identified by those evaluations.
- *2.4.3 Accessibility:* Approaches to increase the availability, transparency, and standardisation of survey tools and their outputs.

2.4.1 Tools for Evaluation of the Impacts of Survey Changes

Simulation testing is a principal tool used to evaluate the effects of anticipated changes in survey design and total effort on survey products and stock assessments that use them. Mechanistic simulation models make explicitly defined assumptions about the underlying relationships among ecological variables (e.g., demographic rates, habitat suitability, movement, species interactions). Correlative simulation models describe these underlying relationships empirically (i.e., fitting of observations via a specified model structure). The simulation tools presented in this workshop included both mechanistic and correlative models (Table 2.4-1), but a common workflow included 1) specifying an operating model from which surveys could be simulated, 2) calculating survey products resulting from this simulated truth (e.g., estimated abundance and associated precision), and 3) evaluating the effects of using these survey products on downstream stock assessments and management output.

Table 2.4-1: Software used to evaluate and/or mitigate impacts of survey changes on survey products

Software Name	Description and Use	Source/Workshop Examples
R-INLA	Approximate Bayesian inference for Latent Gaussian Models. R-INLA is a major component of other packages in this Table including VAST and sdmTMB. Used to simulate spatiotemporal population density for use in survey evaluation.	Lindgren and Rue (2015); Kotwicki and Ono (2019); Conner (WKUSER2); Paradinas (WKUSER2)
SamplingStrata	Optimisation software for the design of multivariable (i.e., multispecies) stratified random designs via the Genetic and Bethel algorithms. Used to mitigate effort reduction by optimising effort allocations across strata with lower total sampling effort.	Barcaroli 2014; Oyafuso(WKUSER2); Barnett (WKUSER2)
sdmTMB	Spatiotemporal generalised linear mixed effects model via template model builder, R-INLA, and Gaussian Markov random fields. Used in abundance index estimation, species distribution modelling, and spatiotemporal simulation of population density.	Anderson et al. 2022; Regular/Anderson (WKUSER2)
SimSurvey	Simulation framework to evaluate sampling protocol of dynamic populations that vary across ages, time and space. Used to simulate survey designs (e.g., simple and stratified sampling) along with hierarchical sub-sampling of lengths and ages.	Regular et al. 2020; Yacin (WKUSER2); Regular/Anderson (WKUSER2); Miller (WKUSER2)
spatioTemporalIndices	Latent Gaussian spatiotemporal model.. Used to estimate abundance indices by length-class with uneven survey coverage.	Breivik et al. 2021
StoX & Rstox	Calculates estimates of abundance and catch-at-age from acousitc trawl surveys and fishery-dependent data	Johnsen et al. 2019

SurveyIndex	GAM-based approach to index production for biomass and compositional data. Used in both estimation and simulation of spatiotemporal population densities.	Berg et al. 2014; Siple (WKUSER2); Markowitz (WKUSER2)
VAST	Spatiotemporal delta-generalized linear mixed model fitted via template model builder, R-INLA, and Gaussian Markov random fields. Used in both estimation of abundance indices and simulation of spatiotemporal population densities.	Thorson (2019); Thorson et al. (2015); Thorson and Haltuch (2018); Thorson and Barnett (2017); Oyafuso (WKUSER2); Barnett (WKUSER2); Thorson (WKUSER2s);

2.4.1.1 Tools for Evaluation of the Impacts of Survey Changes

Simulation testing is a principal tool used to evaluate the effects of anticipated changes in survey design and total effort on survey products and stock assessments that use them. Mechanistic simulation models make explicitly defined assumptions about the underlying relationships among ecological variables (e.g., demographic rates, habitat suitability, movement, species interactions). Correlative/empirical simulation models describe these underlying relationships empirically (i.e., fitting of observations via a specified model structure). The simulation tools presented in this workshop included both mechanistic and correlative models (Table 2.4-1), but a common workflow included 1) specifying an operating model from which surveys could be simulated, 2) calculating survey products resulting from this simulated truth (e.g., estimated abundance and associated precision), and 3) evaluating the effects of using these survey products on downstream stock assessments and management output.

2.4.1.2 Summary of Existing Approaches to Evaluate the Impacts of Survey Changes

Mechanistic and semi-mechanistic simulation models allow the user to develop operating models where functional relationships among ecological variables are explicitly specified by the user. The R package SimSurvey (Regular et al. 2020) is a recently developed tool for simulating age-structured and spatially explicit population dynamics. SimSurvey was designed to simulate sampling designs (e.g., simple random sampling, stratified random sampling, etc.) as well as hierarchical sampling protocols (i.e., subsampling length and length-stratified age data at the haul level). Miller's case study in this workshop proposed the use of this tool to evaluate changes in survey outputs due to reductions in survey area, highlighting the potential utility of this tool to support evaluations of anticipated effort changes. The specification of causal and explicit relationships in the simulation is a strength of these mechanistic simulation tools but these relationships can be difficult to parameterize and groundtruth. However, in their keynote presentation, Regular and Anderson proposed pairing output from geostatistical models to inform mechanistic simulation models.

Correlative models rely on empirically-driven specifications of ecological distributions and relationships conditional on observed data. These types of models were more common in the case studies presented by workshop participants relative to mechanistic models. The most common R packages used for simulation among the talks presented in this workshop were VAST (Thorson et al. 2016) and R-INLA (Lindgren and Rue 2015). Spatial and spatiotemporal variation in density are usually defined via latent variables (e.g., random fields), and fixed effects on density and catchability can be fitted via linear, polynomial, and nonlinear relationships (e.g., splines similar to those used in the mgcv R package). The R package sdmTMB (Anderson et al. 2022) is a recently developed package that also has the capability to simulate densities from a fitted spatiotemporal model similar to VAST. SurveyIndex (Berg et al. 2014) is a GAM-based approach to fit similar types of spatiotemporal models; workshop case study presentations (Siple and Markowitz) focused on its use in abundance index generation but simulation is also a function of the package. Although less tools are available for simulating acoustic trawl surveys, Holmin et al (2020) produced an observation model by combining the NORWegian ECOlogical Model system (NORWECOM; Skogen et al. 1995) and the StoX survey estimation tool (Johnsen et al. 2019) to test different in silico survey strategies.

In addition to simulation approaches, survey effort changes could also be evaluated using purely data-driven or empirical approaches that rely on resampling techniques or simply data removal. A couple of case study presentations during the workshop highlighted these approaches. DeFilippo's WKUSER2 presentation analyzed the impacts of dropping bottom trawl survey stations on survey CV and resulting stock assessment and management output. Another presented study used a bootstrap sampling approach to evaluate how data-weighting metrics for age composition (i.e., input sample size), stock assessment output (i.e., estimates of overfishing limit), and otolith sampling costs were affected by changes to otolith sampling rates (± 0%, 33%, 67%) and changes to sampling methods (i.e., whether changes to sampling were administered via a change in the number of otoliths collected per tow or the number of tows upon which otoliths were sampled) (Siskey et al., 2022). Applications of resampling techniques were fewer in the cases presented in WKUSER2 than WKUSER1, replaced by more simulation approaches. This is concurrent with the development of new software and packages to simulate populations (Table 2.4-1). Resampling techniques are more easily implementable relative to simulation methods but assume the data are representative of the population and limit evaluation of other sampling designs or sampling of new areas.

2.4.1.3 Limitations of Existing Approaches and Recommendations for Future Development

Existing methodology and software packages for evaluating the effects of survey design changes rely primarily on simulation analyses. While the design and characteristics of these simulations may vary, the metrics used to evaluate alternative sampling designs are typically the precision and accuracy of survey data products. We recommend that simulation approaches consider a broader range of metrics for evaluation (e.g., accuracy and precision of model-based indices and compositional estimates; Siskey et al., 2022). Indeed, as model-based indices become more common, optimising surveys with respect to design-based estimates alone may lead to incomplete or one-sided recommendations.

In addition to considering a wider range of survey data products, we recommend that evaluations of survey design changes also consider outcomes beyond the survey data products themselves. In particular, linking survey simulations with stock assessment models to explore the effects of design changes on assessment model estimates (e.g., overfishing limits) offers several advantages. Stock assessment models provide the guidance that is ultimately used in management decision-making, and typically incorporate multiple data sources (e.g. survey abundance indices, fisheries harvest data, and compositional data from both surveys and fisheries) which interact to affect model predictions via a joint likelihood (Maunder and Punt, 2013). As such, variation in the precision or accuracy of survey data products may not necessarily affect assessment model output in a predictable manner. Thus, optimising survey design with respect to the quality of survey data products alone may not necessarily be representative of the effects on assessment outcomes. We recommend building the capacity for existing survey simulation approaches and software packages to interface with stock assessment models (e.g., Stock Synthesis, ASAP, etc.) in order to propagate the simulated effects of survey design changes into assessment outputs. Similarly, building interfaces between survey simulation approaches and existing packages for management strategy evaluation (MSE; e.g., openMSE, mseR, SSMSE) would allow further investigation of the impacts of survey design and effort reduction on management and economic objectives (e.g., Cronin-Fine presentation).

An additional challenge for simulation approaches that evaluate survey designs is to consider the role of environmental change and non-stationarity. Existing methods for simulating and evaluating survey design are based on fixed functional relationships (e.g., mechanistic models) and/or conditioning on existing survey data (e.g., semi-empirical/empirical methods). However, climate change is likely to lead to non-stationarity in the biological processes underlying these functional relationships as well as spatial, spatiotemporal, and multispecies covariance structures. As such, inferences on survey design choices based on simulations conditioned on existing data/relationships may break down as future conditions change, leading to decisions that are optimal for present but not future conditions. Considering that changing ecosystems and species distributions are major challenges for fisheries-independent surveys, including predictions of future states (and associated uncertainty) in survey design simulations is an important consideration. Accounting for environmental change and uncertainty in survey simulation approaches may be achieved by linking existing survey simulation packages and regional climate models (e.g., Bering 10k ROMS). As such, we recommend development of improved interfaces between climate and ecosystem models and survey simulation tools as a priority in future developments.

Beyond improving the capabilities of existing simulation approaches to consider a broader range of objectives and inputs, we also recommend building the capacity for such simulations to be deployed in a tactical capacity. Often the initial effort to develop a reliable simulated operating model (be it mechanistic or empirical) can be substantial, making these approaches difficult to deploy for in-season decision making. While some effort reductions may be anticipated over longer time horizons (e.g., foreseen budget cuts, area closures), often such circumstances may be unanticipated until the survey is under way (e.g., area inaccessibility due to weather, unanticipated crew shortages, vessel breakdowns). As such, we recommend that survey groups build and maintain the capacity to rapidly evaluate effort reductions within the survey season. This can likely be accomplished using existing software and approaches by ensuring that models are adequately developed and conditioned prior to the survey season. Rather than being forced to confront these decisions with little or no empirical guidance, we recommend that survey programs plan for unanticipated effort reduction challenges by having flexible simulation tools conditioned and ready to evaluate situations that may arise. For instance, if hazardous weather is a consistent problem for a given survey area, making it difficult to complete the prescribed sampling plan, field staff may need to decide which stations/areas to drop to minimise impacts. Having a simulation model that can objectively evaluate a suite of possible options would provide valuable information for in-season decision-making.

Finally, there is little guidance on the suitability of alternative survey simulation approaches for different situations. While a range of approaches and software packages exists (e.g., mechanistic, semi-empirical, empirical; SimSurvey, VAST, sdmTMB), the strengths and limitations of each are seldom discussed in the literature, providing little information for a prospective analyst to choose among available options. Furthermore, available approaches and software packages are seldom evaluated side-by-side. If different approaches provide conflicting advice, this should be considered in the treatment of uncertainty in the results of the simulation model. We recommend further study comparing alternative survey simulation approaches and software packages, and better characterization of the differences between them in the literature to provide information for analysts.

2.4.2 Tools for Mitigating Impacts of Survey Change

Survey effort reductions occur for a variety of reasons (e.g., reductions in days at sea owing to a lack of staff, untrawlable conditions or habitat, reduced funding for processing collected samples, etc.) and are an unavoidable aspect of conducting fishery-independent surveys. Tools to mitigate the impacts of survey changes are available at the survey design and estimation stage, and when selecting observation methodologies.

2.4.2.1 Estimation Models

Design-based methods have been the most common approach for generating survey indices and age/length compositions. Based on sampling theory, these indices often use information on survey design (i.e., effort, stratum area) to expand information collected at the tow level (e.g., catch, length frequency, age frequency) up to the population level (i.e., population abundance, population-at-length, population-at-age). Advantages of design-based indices include relative ease of interpretability and a long history of implementation. It is highly advantageous for data collectors to understand the index calculation process so that the impact of decisions made can be fully understood, and the relative simplicity of design-based indices facilitates this. Case study presentations by Barnett and Oyafuso used the Bethel algorithm to perform multispecies optimizations of optimal effort allocation for stratified random designs in the Gulf of Alaska region across a range of total effort levels. While useful for planning, this tool can be used in the short-term for mitigating in-season reduced sampling. The advantage of this approach is that decisions of where to reduce sampling (i.e., across strata) are done objectively in a way to minimise the information lost (i.e., loss in precision) from reduced effort, however the choice of species weightings in the optimization need to be considered by survey planners.

Changes to survey area and effort that result in spatial and/or temporal imbalances in survey data may be overcome by using model-based approaches for generating survey data products. Model-based indices assume that the quantity of interest (e.g., biomass, age composition, fish condition) can be described by a statistical model with terms that explain spatio-temporal patterns in the distribution of a stock. This property makes model-based indices especially well-suited for dealing with unbalanced survey effort and shifting distributions of stocks (WKUSER2 presentation from Paradinas). Model-based indices also allow for gears to be combined in the creation of a unified index, and naturally facilitate testing, development, and inclusion of new data collection methods in survey programs. Many of the simulation tools developed for evaluation of survey change are geostatistical spatiotemporal models (Table 2.4-1). Studies that have compared model-based indices derived from these tools (e.g., Brodie et al. 2020; Breivik et al., 2021; Anderson et al. 2022) found that models fit with these tools can be configured in similar ways to produce similar outputs, but their specific properties make implementation, interpretation, and capabilities variable depending on the specific survey to which they are applied.

2.4.2.2 Platforms

In the event that fishery survey vessels cannot complete a full survey, there are several options for alternative survey platforms that can collect the missing data. One option is to employ a co-operative approach in which industry partners survey areas where the fishery survey vessel cannot. For unplanned, short-term reductions in survey coverage, industry partners are generally well-equipped to mobilise and provide data in a rapid manner, but issues arise in the context of legal permissions, funding to conduct survey effort, arranging contracts, and perhaps most importantly, standardisation between survey vessels and gear. For planned disruptions, differences in configurations and designs between the survey vessel and the industry vessel can make industry vessels better equipped to sample in challenging areas (e.g., wind farms, areas untrawlable to survey trawling gear). Another option that involves industry partners is direct use of fishery-dependent data without changes to fishing practices. This approach presents problems related to selectivity, catchability, and preferential sampling, but may be useful for both planned and unplanned disruptions as it provides data where they would otherwise not exist. It is most advantageous to use these data for specific needs (e.g., compositional data) but could also be used for generating abundance indices (with some challenges already mentioned in TOR II). Fishing vessels can also be equipped with devices that allow collection of additional data beyond their catch to provide data that would otherwise be lost in a design change or effort reduction (e.g., echosounders on ships of opportunity, cameras).

An additional option when survey effort is reduced or designs are changed is the employment of uncrewed systems (Chu et al., 2019; Thompson and Guihen, 2018; Sepp et al., 2022). For unplanned reductions, uncrewed systems can be used to maintain survey coverage when and where it is not possible to conduct crewed surveys (e.g., De Robertis et al. 2021). For planned reductions, uncrewed systems may be more nimble than crewed platforms, allowing sampling in challenging areas (e.g., wind farms). They may also be useful for expanding survey effort in both spatial and temporal domains in an efficient and cost-effective manner. However, there are significant issues associated with uncrewed systems surveys related to the lack of biological sampling associated with them. This is an especially important issue for uncrewed acoustic surveys, which need compositional data to scale acoustic backscatter and estimate the biomass of specific species and age classes. In the absence of fishery-independent compositional data, fishery-dependent data could be used with its limitations in mind, but further research is needed to validate this approach (see presentation by Bolser). If circumstances preclude a typical allocation of effort to survey vessels, further efforts to integrate and validate data collected by uncrewed systems could result in an expansion of survey effort and new streams of data that otherwise would not be possible in a given logistical or funding situation.

2.4.3 Future Directions

Specific observation methods (i.e., sensors, gears, combinations of gears) and data processing techniques show promise for minimising the impact of a loss of survey effort or specific type of data. However, significant advancements are necessary before implementation in regular fishery resource survey program operations. For example, environmental DNA (eDNA) sampling can be conducted with minimal effort and can yield insights about species composition, distribution, and abundance (Shelton et al. 2016; Lacoursière-Roussel et al. 2016; Rourke et al. 2022). Promising research in a fishery resource survey program suggests that eDNA-derived estimates of the distribution and abundance of Pacific hake (Merluccius productus) closely reflect trawl-derived estimates at a broad scale for the U.S. West Coast (Shelton et al. 2022). It will be necessary to work towards a more nuanced description of species compositions and conduct calibration experiments for eDNA sampling to fully compensate for a lack of data collected by other means, but at present eDNA sampling could be useful for partial compensation. Other genetics-based approaches (e.g., close kin mark recapture) show promise for generating estimates of population size (Bravington et al. 2016; Ruzzante et al. 2019; Trenkel et al. 2022) but have different burdens on sampling effort than conventional surveys (e.g., focus on tissue vs. biomass sampling), which might be advantageous in the event of unplanned or planned reductions in survey effort.

Broadband (i.e., frequency modulated) acoustic sampling can differentiate species more accurately than is possible with analysis of single frequency (i.e., continuous wave) data (Benoit-Bird et al. 2020; Boswell et al. 2020; Roa et al. 2022). However, at present there are obstacles to employing broadband acoustic data to provide species composition data that are equivalent to data collected by other gears. These include a lack of target strength information for common species and a lack of reproducible and scalable methods for classifying scattering sources. When methods for processing broadband acoustic data advance, they will facilitate independent uncrewed systems surveys that produce results that may be comparable to conventional acoustic surveys. Corroborating estimates of species composition derived through broadband methods with estimates of compositional data derived from spatio-temporal models fit to available (e.g., fishery-dependent) data has the potential to add confidence to results (presentation by Bolser). Artificial intelligence and machine learning approaches have the potential to facilitate scalable processing of broadband acoustic data (Malde et al. 2020), and will be highly beneficial to develop for processing acoustic and optical data in general (e.g., Richards et al. 2019; Kilfoil et al. 2020; Roa et al. 2022). If such methods are applied to data collected by automated systems deployed on fishing vessels (e.g., camera sampling for composition and length measurements), they can provide a large amount of data that could be useful for addressing gaps in survey data. For these examples of observation methods and data processing techniques, advancement in methodologies to address design changes and reduction in survey effort could not only meet this need, but also eventually result in the provision of richer data for assessments.

2.4.4 Accessibility of Tools to Analyse Changes to Surveys

Simulation models, packages, code routines for producing survey indices, and other critical tools for the survey enterprise produce vital inputs to contemporary fisheries management. Ease of use, predictable outputs, and clear documentation are crucial to maintaining stakeholder trust and ensuring defensible scientific results. As science agencies encounter increasing instances of unavoidable effort reduction in their scientific surveys due to budgetary constraints, climate impacts, and shifts in survey focus, it is more important than ever that technology and software tools meet community standards and needs.

Fisheries researchers often find themselves in a position of having to quickly develop bespoke solutions to immediate problems that are then adapted and distributed for broader use beyond their original intent. Software development standards and best practices are widely available, but are not broadly adopted by fisheries practitioners. Fisheries tools are often developed by individuals without professional software development training, made to address a specific problem, distributed to other potential users without adequate documentation, and lack a plan for long term maintenance of the tool despite their importance and utility. However, these issues can be addressed by taking organisational and cultural steps towards improvements in tool accessibility and usability, tool documentation and transparency, tool checking and code review, and tool testing and comparison. Additionally, improvements in data governance, while beyond the scope of WKUSER, are key to improving the transparency and accessibility of data necessary to plan and evaluate the impacts of survey effort reductions on fisheries.

We have compiled a list of recommendations for tool developers on three fronts: (1) Tool accessibility and usability, (2) documentation and transparency, and (3) checking and review.

2.4.4.1 Tool Accessibility and Usability

An issue hampering forward progress in addressing unavoidable survey effort reductions is not necessarily the lack of adequate tools to assess impacts to abundance indices and resulting catch advice of survey reductions, but the accessibility and usability of currently available tools. A current challenge that is relatively easy to address is the lack of available information on the full suite of tools available to support assessing changes to surveys. Table 2.4-1 provides an important contribution towards filling this gap, by compiling a list of common survey tools and their attributes.

1. Developers of survey tools should ensure their tools are made widely available by sharing their tools and resources to a central tool repository for easy access and discovery e.g.,

NOAA Fisheries Integrated Toolbox (https://noaa-fisheries-integrated-toolbox.github.io); alternative or additional toolbox locations with a more international focus may also be considered.

2. Useful information to include on the user interface of a tool include: the tool name, description, link to github repository or website (e.g., pkgdown page), capabilities and outputs from the tool, developer and contact information, maintenance information, platform and build environment information (e.g., the `sessionInfo()` function in the R software environment), run checks and level of testing, statistical and computational limitations, and a description of tool diagnostics.

2.4.4.2 Tool Documentation and Transparency

Documentation shows users how to appropriately apply tools to their situation or research question. Thorough tool documentation should include metadata (e.g., when the tool was developed, by who, when it was last updated); descriptions of function or tool intended use, guides, inputs, and outputs; minimal reproducible examples; and case studies and vignettes. Moving towards a standardised layout and documentation procedure for model and tool manuals would be helpful to both developers and users. This procedure will help developers understand what information should be included and help users to readily and efficiently find the information they need to appropriately assess and apply the tool. As noted above, documentation should include examples of the tool in action. Often an example is provided of successful use of the tool, but additional, more nuanced examples would help users better understand operational use and capacity of the tool. These nuanced examples could include instances where the tool was inappropriately applied, where another tool would have been a better choice and why, or what unsatisfactory results may look like.

1. Software tools to support surveys must have thorough and standardised documentation to ensure users beyond the original developer can successfully utilise the tool and understand results.
2. To the extent practicable, variable definitions, data and variable input formats, and data output formats should be standardised to improve comparability between tools and enterprise cross-tool collaboration opportunities.

It is imperative that tools developed and distributed to users outside of the development team meet contemporary software development standards and best practices. Often, the development of software is prioritised by applied research needs and less so on meeting community standards for wider use. To meet this goal, fisheries researchers must have the support of program managers and leadership to maintain tools. This support can be in the form of time, training, and resources for the tool developer or team to systematically improve the tool or to hire professional developers who can upkeep the tool. This is not often the case, whether development is contracted out to professional developers (with no plan in place for maintenance), or it is necessary to adopt the perspective that tools are effectively valueless without helpful and intuitive documentation and documentation is an integral part of the tool development process. Many of these best practices for tool and package development are outlined by institutions like the ROpenSci (www.ropensci.org/).

2.4.4.3 Tool Checking and Code Review

Creating standardised and comprehensive code checking routines is imperative, as many of the products being used for producing survey indices and assessing the impact of survey effort reduction need to be professional and defendable. These products often inform catch advice and funding priorities and thus have major legal and economic implications. Products produced for these efforts should represent the best available science and be predictably applicable by anyone who attempts to use them.

Each tool or model must be assessed through test routines to ensure the product is continuing to produce anticipated results. These tests should be implemented through known case studies and unit testing packages like the testthat R package (Wickham 2011) or similar. Rigorous peer review should follow guidelines from or similar to the rOpenSci code review standards, be clearly scheduled (e.g., each version release, annually), and should be conducted by different scientific team members and product developers. Scientific team members are critical for assessing if there are advances in science or methodology that need to be incorporated into the product. Professional product developers are trained resources that can ensure that changes in support packages, operating systems, and system preferences will not inadvertently cause the product to 'break' or produce incorrect results.

1. Each tool should have an accompanying unit testing script and follow guidelines on code review standards.
2. Each public facing tool undergoes regular, end-to-end, peer review to make sure the tool is performing consistently and correctly. Tools should be assessed by both researchers who can assess the science, methodology, and outputs, and by product developers who can ensure the durability of the product as technology and platforms change.

2.4.4.4 Tool Testing and Comparison

With the large and ever-growing body of research, development, and tools, it is essential that guidelines are formulated to help users choose the most appropriate tool or model for the application. The selection of which tool to use is largely a subjective process, skewed heavily by the comfort of the user with the tool being used and availability and simplicity of the model. However, little information is available on how to provide quantitative comparisons of survey tools and models. As noted above, some single-species comparison studies have been conducted (including case studies presented during this workshop) that conclude with inexplicable mismatches in results between methods and tools.

1. ICES should commission a broad cross-comparison study of groups of common survey tools to validate model outputs and assumptions.
2. ICES should develop simple decision trees to support user evaluation of common groups of survey tools.

In addition to providing additional assurance in the results and interchangeability of available tools, a cross-comparison study would also be invaluable towards developing guidelines for tool users for model and tool selection. Currently, no such generalised guidelines (e.g., decision trees or similar decision support tools) exist for the survey community.

2.4.5 References

Anderson, S.C., E.J. Ward, P.A. English, L.A.K. Barnett. 2022. sdmTMB: an R package for fast, flexible, and user-friendly generalized linear mixed effects models with spatial and spatiotemporal random fields. bioRxiv 2022.03.24.485545; doi: https://doi.org/10.1101/2022.03.24.485545

Barcaroli G (2014). "SamplingStrata: An R Package for the Optimization of Stratified Sampling." *Journal of Statistical Software*, **61**(4), 1–24. doi:10.18637/jss.v061.i04, https://www.jstatsoft.org/v61/i04/.

Benoit-Bird, K.J. and Waluk, C.M., 2020. Exploring the promise of broadband fisheries echosounders for species discrimination with quantitative assessment of data processing effects. The Journal of the Acoustical Society of America, 147(1), pp.411-427.

Berg, C. W., Nielsen, A., & Kristensen, K. (2014). Evaluation of alternative age-based methods for estimating relative abundance from survey data in relation to assessment models. *Fisheries research*, *151*, 91-99.

Boswell, K.M., Pedersen, G., Taylor, J.C., LaBua, S. and Patterson III, W.F., 2020. Examining the relationship between morphological variation and modeled broadband scattering responses of reef-associated fishes from the Southeast United States. Fisheries Research, 228, p.105590.

Bravington, M.V., Skaug, H.J. and Anderson, E.C., 2016. Close-kin mark-recapture. Statistical Science, 31(2), pp.259-274.

Breivik, O. N., Aanes, F., Søvik, G., Aglen, A., Mehl, S., & Johnsen, E. (2021). Predicting abundance indices in areas without coverage with a latent spatio-temporal Gaussian model. *ICES Journal of Marine Science*, *78*(6), 2031-2042.

Brodie, S.J., Thorson, J.T., Carroll, G., Hazen, E.L., Bograd, S., Haltuch, M.A., Holsman, K.K., Kotwicki, S., Samhouri, J.F., Willis-Norton, E. and Selden, R.L., 2020. Trade-offs in covariate selection for species distribution models: a methodological comparison. Ecography, 43(1), pp.11-24.

Chu, D., Parker-Stetter, S., Hufnagle, L. C., Thomas, R., Getsiv-Clemons, J., Gauthier, S., and Stanley, C. 2019. 2018 Unmanned Surface Vehicle (Saildrone) acoustic survey off the west coasts of the United States and Canada. In OCEANS 2019 MTS/IEEE SEATTLE, pp. 1–7.

De Robertis, A., Levine, M., Lauffenburger, N., Honkalehto, T., Ianelli, J., Monnahan, C. C., Towler, R., et al. 2021. Uncrewed surface vehicle (USV) survey of walleye pollock, Gadus chalcogrammus, in response to the cancellation of ship-based surveys. ICES Journal of Marine Science, 78: 2797–2808.

Holmin, A. J., Mousing, E. A., Hjøllo, S. S., Skogen, M. D., Huse, G., & Handegard, N. O. (2020). Evaluating acoustic-trawl survey strategies using an end-to-end ecosystem model. *ICES Journal of Marine Science*, *77*(7-8), 2590-2599.

Johnsen E. , Totland A. , Skålevik Å. , Holmin A. J. , Dingsør G. E. , Fuglebakk E. , Handegard N. O. 2019. StoX: an open source software for marine survey analyses. *Methods in Ecology and Evolution*, 10: 1523–1528.

Kilfoil, J.P., Rodriguez-Pinto, I., Kiszka, J.J., Heithaus, M.R., Zhang, Y., Roa, C.C., Ailloud, L.E., Campbell, M.D. and Wirsing, A.J., 2020. Using unmanned aerial vehicles and machine learning to improve sea cucumber density estimation in shallow habitats. ICES Journal of Marine Science, 77(7-8), pp.2882-2889.

Kotwicki S, Ono K. The effect of random and density-dependent variation in sampling efficiency on variance of abundance estimates from fishery surveys. Fish Fish. 2019;20:760–774. https://doi.org/10.1111/faf.12375

Lacoursière-Roussel, A., Côté, G., Leclerc, V. and Bernatchez, L., 2016. Quantifying relative fish abundance with eDNA: a promising tool for fisheries management. Journal of Applied Ecology, 53(4), pp.1148-1157.

Leroy B., Meynard C.N., Bellard C., Courchamp F. 2016. virtualspecies: an R package to generate virtual species distributions. *Ecography* 39:599-607

Lindgren, F., & Rue, H. (2015). Bayesian spatial and spatio-temporal mod- elling with R-INLA. Journal of Statistical Software, 63, 1–25. https:// doi.org/10.18637/jss.v063.i19

Malde, K., Handegard, N. O., Eikvil, L., & Salberg, A. B. (2020). Machine intelligence and the data-driven future of marine science. *ICES Journal of Marine Science*, *77*(4), 1274-1285.

Maunder, M. N., & Punt, A. E. (2013). A review of integrated analysis in fisheries stock assessment. *Fisheries Research*, *142*, 61-74.

Regular PM, Robertson GJ, Lewis KP, Babyn J, Healey B, et al. (2020) SimSurvey: An R package for comparing the design and analysis of surveys by simulating spatially-correlated populations. PLOS ONE 15(5): e0232822. https://doi.org/10.1371/journal.pone.0232822

Richards BL, Beijbom O, Campbell MD, Clarke ME, Cutter G, Dawkins M, Edgington D, Hart DR, Hill MC, Hoogs A, Kriegman D, Moreland EE, Oliver TA, Michaels WL, Piacentino M, Rollo AK, Thompson C, Wallace F, Williams ID, Williams K. 2019. Automated Analysis of Underwater Imagery: Accomplishments, Products, and Vision. NOAA Technical Memorandum NOAA-TM-NMFS-PIFSC-83. 59 p.

Roa, C., Pedersen, G., Bollinger, M., Taylor, C. and Boswell, K.M., 2022. Taxonomical classification of reef fish with broadband backscattering models and machine learning approaches. The Journal of the Acoustical Society of America, 152(2), pp.1020-1034.

Rourke, M.L., Fowler, A.M., Hughes, J.M., Broadhurst, M.K., DiBattista, J.D., Fielder, S., Wilkes Walburn, J. and Furlan, E.M., 2022. Environmental DNA (eDNA) as a tool for assessing fish biomass: A review of approaches and future considerations for resource surveys. Environmental DNA, 4(1), pp.9-33.

Ruzzante, D.E., McCracken, G.R., Førland, B., MacMillan, J., Notte, D., Buhariwalla, C., Mills Flemming, J. and Skaug, H., 2019. Validation of close-kin mark–recapture (CKMR) methods for estimating population abundance. Methods in Ecology and Evolution, 10(9), pp.1445-1453.

Sepp, E., Vetemaa, M., and Raid, T. 2022. Use of autonomous research vehicles in Baltic fisheries acoustic surveys: Potential benefits and pitfalls. In Trends in Maritime Technology and Engineering Volume 2. CRC Press. 4 pp.

Siskey, M.R., Punt, A.E., Hulson, P-J.F., Bryan, M.D., Ianelli, J.N., and Thorson, J.T. 2022. The estimated impact of changes to otolith field-sampling and ageing effort on stock assessment inputs, outputs, and catch advice. *Canadian Journal of Fisheries and Aquatic Sciences*, doi: 10.1139/cjfas-2022-0050.

Shelton, A.O., O'Donnell, J.L., Samhouri, J.F., Lowell, N., Williams, G.D. and Kelly, R.P., 2016. A framework for inferring biological communities from environmental DNA. Ecological Applications, 26(6), pp.1645-1659.

Shelton, A.O., Ramón-Laca, A., Wells, A., Clemons, J., Chu, D., Feist, B.E., Kelly, R.P., Parker-Stetter, S.L., Thomas, R., Nichols, K.M. and Park, L., 2022. Environmental DNA provides quantitative estimates of Pacific hake abundance and distribution in the open ocean. Proceedings of the Royal Society B, 289(1971), p.20212613.

Skogen M. D. , Svendsen E. , Berntsen J. , Aksnes D. , Ulvestad K. B. 1995. Modelling the primary production in the North Sea using a coupled three-dimensional physical-chemical-biological ocean model. *Estuarine, Coastal and Shelf Science*, 41: 545–565.

Thompson, F., and Guihen, D. 2019. Review of mission planning for autonomous marine vehicle fleets. Journal of Field Robotics, 36: 333–354.

Thorson, J.T., 2019. Guidance for decisions using the Vector Autoregressive Spatio-Temporal (VAST) package in stock, ecosystem, habitat and climate assessments. Fish. Res. 210, 143–161. https://doi.org/10.1016/j.fishres.2018.10.013

Thorson, J.T., Barnett, L.A.K., 2017. Comparing estimates of abundance trends and distribution shifts using single- and multispecies models of fishes and biogenic habitat. ICES J. Mar. Sci. 74, 1311–1321. https://doi.org/10.1093/icesjms/fsw193

Thorson, J. T., and Haltuch, M. A. 2018. Spatio-temporal analysis of compositional data: increased precision and improved workflow using model-based inputs to stock assessment. Canadian Journal of Fisheries and Aquatic Sciences. doi:10.1139/cjfas-2018-0015. URL: http://www.nrcresearchpress.com/doi/abs/10.1139/cjfas-2018-0015#.W0oloTpKiUk

Thorson, J.T., Cunningham, C.J., Jorgensen, E., Havron, A., Hulson, P.J.F., Monnahan, C.C., von Szalay, P., 2021. The surprising sensitivity of index scale to delta-model assumptions: recommendations for model-based index standardization. Fisheries Research, 233, 105745. https://doi.org/10.1016/j.fishres.2020.105745

Trenkel, V.M., Charrier, G., Lorance, P. and Bravington, M.V., 2022. Close-kin mark–recapture abundance estimation: practical insights and lessons learned. ICES Journal of Marine Science, 79(2), pp.413-422.

Wickham H (2011). "testthat: Get Started with Testing." *The R Journal*, **3**, 5–10. https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf.

3 Synthesis of the Workshop

3.1 Summary of each TOR group

3.1.1 TOR I. "Survey design for flexibility"

Why is it important to have a flexible survey? This was the main question that established the need for WKUSER2 workshop in Galway, Ireland in September 2022. Simply put, fish distribution may change unpredictably in response to changing environmental and anthropogenic forcings (e.g. range expansion, contraction, etc), monitoring resources may fluctuate due to budget reduction, vessel breakdown, and the sampling frame can also be reduced due to establishment of MPAs or wind farms for example. Not adapting to these changes mean obsolescence: the value of the survey is gradually reduced to the point where it is not worth continuing to collect the data.

TOR I therefore started with creating a definition of a flexible survey: Approaches which facilitate multiple robust estimation options to retain the ability to acquire consistent and/or approximately unbiased estimates given change in survey resources, distribution of resources and monitoring access, and observation requirements.

The group then provided short-term (just prior or during cruise, named as "tactical" – Table 2.2-1) as well as long-term (in the planning stage, long time ahead of survey, named "strategic" – Table 2.2-2) recommendations for making flexible adjustments to the survey design to preserve the continuity/comparability of the survey time series over time when facing unavoidable survey effort changes due to i) decreased survey resources, ii) habitat expansion or contraction, or iii) reduced sampling universe. The flexibility was tackled from three axes i.e. combining/changing gear, considering alternate/auxiliary observation method(s), and improving/developing estimation method(s).

The main take home messages from the group were as follow:

1. Tools (that are accessible, transparent, and reproducible) are needed to evaluate the cost/consequences of "reduced survey effort". But in order to evaluate, objective criteria and prioritizations are needed.
2. These objectives and priorities should be generated together with stock assessment scientists, biologists, managers, and stakeholders.
3. If changes are deemed necessary, it is best to evaluate the proposed design/sampling changes a priori, if possible. These proposed changes (as suggested in Table 2.2-1 or Table 2.2-2) should also be incremental, documented, with "quality control" i.e. assess whether the objectives were achieved and the associated cost.
4. The above evaluation should also consider contingency plans in case of sudden unforeseeable problems.

3.1.2 TOR II: "Combining surveys, dealing with data gaps"

One pathway to creating a flexible survey was to develop an estimation method that can combine various data sources – the term "data sources" is used here because data is not limited to the survey but could also be coming from the fishery or other sources. TOR II investigated best practices in doing so. Combining data from different surveys should start with understanding each survey and their sampling frame & design i.e. what "portion" of the stock each survey covers?

If two (or more) data sources have something in common i.e. the samples overlap in space, time, or biological attributes (such as depth, SST, etc), then, data could potentially be combined. However, doing so requires estimating the difference in scale between the data sources (i.e. catchability, or selectivity ratio). These differences can be estimated via a "paired sampling experiment" (where two vessels fish in parallel thus overlap in most attributes except for the treatment factor(s). Such experiments are costly but provide a straightforward way to scale the two data sources). Alternatively, these differences in scale can be estimated via "opportunistically paired sampling" (when data sources share only part of attributes). This method often estimates the scaling factor within the model itself but requires including covariate(s) that control for the difference in attributes between the data sources i.e. the "catchability covariate". The group also recommended that data sources should always be analysed separately first to find and understand any discrepancies. If differences are explainable and reasonable, then, user need to think about how to weight the datasets (e.g. by area coverage of the stock) or data products (i.e. index of abundance). Finally, one should always perform diagnostics to examine whether the combined model exhibits potential issues such as: preferential sampling, parameters confounding, or retrospective pattern.

However, there are situations where data sources do not have enough (or any) overlap. In that case, one should stop trying to combine data. Instead, they should analyse the data independently and/or investigate experiment or changes in sampling design that may increase overlap between the data sources.

3.1.3 TOR III. "Modelling and simulations"

In this TOR, participants reviewed and identified gaps in the current best practices for using/developing simulations and modelling tools to answer questions related to the changes in survey effort. They did so from a practitioner point of view to avoid getting lost in the detail.

The group first reviewed the modelling best practices (section 2.3.4). These practices can be largely grouped into two important considerations: model development (e.g. choice between single species *vs.* multi-species model, available and important covariates, spatial and temporal scale of the covariates, etc.) and model validation and diagnostics (e.g. model convergence, parameter estimates, trustworthiness of covariance matrix, residual analysis).

The group next reviewed current best practices in developing simulation models (section 2.3.4). Two major pathways of simulation were identified: resample-based (which captures all the intricate structure in the data but lacks the "truth" to compare the estimates against) and model-based (where the truth is specified but is more challenging to capture all the important factors and interactions affecting the population of interest and data generation process). Model-based simulation can further be divided into two sub-types: mechanistic model (i.e. where user explicitly simulate/specify the underlying mechanisms that generated the observed survey data – thus is very flexible but may not reflect the true data generation process) and correlative/empirical model (i.e. that is based on a model fit to actual data hence can better reflex the observed pattern but is constrained by model specifications and data availability. Such models may be limited in use for evaluating future conditions). The group also highlighted the importance of extending the simulation to evaluate the risk of survey effort reductions to achieving management and/or conservation objectives i.e. conducting management strategy evaluations.

While the progress in simulations and estimation methods has been apparent (e.g. through development of software packages; Table 2.4-1), there are still many remaining challenges. For example: clear objectives of the survey need to be defined early in the process (see also TOR I), tools that can simulate and test all data streams resulting from a survey (not just indices of abundance but also biological data) are needed. In addition, tools are needed for integration of different data types (e.g. acoustic, eDNA that are increasingly being discussed/used to mitigate survey effort reduction). These new tools should also be open-source, user-friendly, and well-documented to increase its usefulness (a discussion point that is tackled in more detail in TOR IV).

3.1.4 TOR IV. "Tools and technology development"

In this TOR, participants investigated tools, technologies, and concepts that are currently available and important when assessing the effect of effort reduction on the quality of survey deliverables and advisory products.

First, they focused on the existing tools to evaluate sampling design changes and created a table with R packages and other software programs that are currently available and their capabilities/limitations (summarized in Table 2.4-1). Many of the currently available solutions are focused on the development of computational/statistical estimation methods capable of handling various "data issues". Other tools can help with adding/changing survey "platforms" e.g. instead of focusing only on survey data, they can fill in the gaps or complement the data with fishery dependent data, or use new technologies e.g. saildrones to collect data where it was impossible before. Moreover, novel types of data and data processing methods are increasingly being investigated/available (e.g. video, still images, broadband acoustic, artificial intelligence for data processing) and these can potentially be used to address the problem of "unavoidable survey effort" and more.

All these new methods and tools can address many of the issues described here but there is a great need for standardization and making these tools accessible, reproducible, and transparent. We need to move away from a black-box situation as new users need to be able to fully understand and control what they do. Furthermore, increasing transparency also requires that code documentation, testing, and examples are accessible and easily understandable to the public.

Some institutions have started moving in this direction (e.g. NOAA fisheries integrated toolbox, or the TAF framework in ICES) but this is something that WKUSER2 would like to push forward to the national and international communities and make it the standard for the future. This also means that institutions/country/funding sources need to act and implement measures to develop a framework that can support the maintenance and updates to the available tools. Many tool developers might not have adequate experience with software development training to ensure that their tool is accessible, reproducible and transparent.

3.2 Lessons learned on workshop conduct

- A large number of presentations were planned for the second day of the workshop. Participants were asked to leave time for discussion but most presentations took the full 15 minutes allocated, leaving no time for discussion. Time for discussion was only possible because a number of presentations were retracted. Stronger guidance on time limits may have allowed for more discussion time. Alternative ways to ask/collate questions may also have been beneficial (e.g. online chat, flipchart etc) these questions could help the subgroups and report writing (even if they could not be discussed in plenary).
- While all contributed presentations were informative, a more balanced representation would have been beneficial; there was a strong focus on NOAA work and on modelling and less on sampling design and operational issues. A better balance could have been achieved by targeting / encouraging presentations on certain topics.
- Each of the subgroups was productive and efficient. This was largely due to the direction provided by the subgroup leaders. They were chosen (well before the meeting) for their

expertise on the subject but also for their skills in leadership and facilitating discussions. This approach contributed significantly to the success of the workshop.

- It is difficult to facilitate meaningful remote participation. Ensuring that remote participants can see presentations, hear discussions, and ideally see the presenter and audience is challenging. Allocating this task to a dedicated person would have been helpful but it is a full-time task. In absence of a strong focus on remote participation it would have been good to communicate more clearly that video links were provided mainly to allow people to listen in (and learn), rather than actively contribute.
- COVID-19 guidelines vary by country, venue and over time. It is important to plan for changes in guidelines and what to do if participants contract covid before, during or even after the meeting: what actions should be taken; who should be informed and how; alternative venues; remote access; support for people who are self-isolating; etc. It is also important to inform participants of these plans and to make sure they can be contacted at short notice (e.g. through a whatsapp group). Not all of these facets were in place and some time was lost during the unplanned transition from an in-person to a fully remote meeting.

3.3 Scope for further development – topics for future research

WKUSER 2 continued on the work initiated during WKUSER 1(ICES 2020) on improving the understanding of uncertainty associated with survey sampling processes (operational, environmental & biological). WKUSER 2 provided advice on designing surveys for flexibility, methods for combining data and dealing with data gaps, and synthesizes best-practices as well as limitations of existing frameworks to test the consequences of survey effort reductions. WKUSER 2 also identified remaining challenges and research topics that should be considered in the future in the short-, mid-, and long-term.

3.3.1 Major challenges to survey practitioners

The WKUSER workshops identified major challenges that survey practitioners face in addressing the needs for maximizing survey efficiency and adapting surveys to new environments. The challenges include limited resources, administrative and governmental requirements specific to different countries, and many kinds of boundaries that prevent seamless sampling of the ecosystem.

Problems with fish moving across survey and country boundaries create logistical challenges because not all surveys can be easily modified due to jurisdictional issues. In these situations, it is important to develop international agreements to assure efficient collaboration between survey groups. Even within national borders, issues emerge when survey effort must shift into the jurisdiction of another organization or group. Coordination between groups is imperative in these situations, as difficult decisions must be made about the allocation of resources (e.g., funds, human capital, ship time) and standardization between surveys. The expansion into new areas can present a number of new logistical challenges beyond jurisdiction due to: increase or change of survey footprint, the need to obtain new resources, development of sampling strategies for new areas, collection of all required permits, etc. Developing good surveys in new areas also requires knowledge about the area, but since the area is new, the information about habitat and the distribution of species may be limited. In cases of expansion of survey into new areas, it may be desirable to perform initial preliminary survey(s) with the goal to inform future survey design for assessment purposes.

Many survey programs are also struggling because of a limited number of staff, which make it difficult to complete surveys and work on important issues necessary to modify the survey in response to new developments in survey science and changes in environment. Many organizations within the ICES community are set up in a way that separates data collections from assessment programs. This separation is seriously hindering progress of modernizing surveys. Organizations need to create mechanisms for the survey and stock assessment scientists to have productive communication and collaboration. Survey programs that depend on chartered commercial fishing vessels often face problems such as limited number of boats available from an aging fleet that are difficult to modify to meet survey needs. The number of research vessels is also limited and building research vessels takes a long time and is very costly. In addition, the crew on the research vessels are usually not as experienced in some of the survey methods (e.g. bottom trawling) as members of the fishing vessels, which can make the survey difficult or even impossible to conduct on a research vessel. Further, when the skill level of crew changes between survey legs or years there could be unquantified changes in catchability.

At the heart of the issues discussed in WKUSER2 is the decision whether to break time series and incorporate new technologies, sampling strategies, and change survey area or continue sampling the same way and account for new dynamics by changing estimation methods alone. It is difficult to make generalized recommendations as all, some, or none of these options may be prudent in different situations, and this decision should be made through careful deliberation between surveyors, assessors, managers, and stakeholders with the aid of simulation studies and field tests. It is important to develop frameworks for making these types of decisions, as when ecosystem dynamics, habitat types, species distributions, and survey resources change, inaction can have similar consequences to action on survey products and assessments.

3.3.2 Research topics in the short-term

- Decision trees for survey managers for different types of surveys and different issues.
- Development of universal survey analysis tools to process survey data. Assure that tools can be used across different survey types and databases (e.g. R packages, augment existing code from researchers and make it universal for the processing survey data and/or estimating reductions impacts). Universal survey tools will make analytical studies easier and continue to streamline the process of QC of the survey data.
- Development of methods to evaluate importance criteria for survey locations. These criteria can be determined using environmental data and information on animal distribution and abundance from the past. For example, the information on density, variance, or spatio-temporal covariance could be used to identify consistent patterns to allow for lower sampling in the future. The information from the past surveys can be used to allocate effort relatively to the expected information contents. Adaptive sampling based on previous data and expected distribution can help with efficient effort allocation.
- Develop methods for dealing with elimination of entire survey areas (e.g. due to area closures to trawling, wind farms, high commercial fishing densities).
- Consider technological developments to improve the ability to collect more data with higher accuracy and less effort (e.g. by using automated weight, length, age, and species classification methods).
- Increase ecosystem data collection during survey for use in model-based estimation and process studies to inform spatio-temporal models and decrease uncertainty in model-based estimates of survey data products.
- Use simulation studies to explore causes of additional variability estimated within assessment models.

- Perform calibration exercises during existing surveys to test different designs using the same gear.
- Continue exploration of model-based estimation and make comparisons with design-based estimates.

3.3.3 Research topics in the mid-term

- Develop methods for multispecies/multi-objective optimization, addressing tradeoffs between data types and different approaches for estimating survey data products. Develop alternative metrics as basis for optimization, provide advice on how to agree on optimization metrics. Develop strategies for communication with stakeholders on how to balance survey objectives (e.g., focus particularly on less abundant species, environmental information, etc.) and to agree on weights for different species and different data types (e.g. consider developing multivariate matrix to weigh between environmental and biological objectives).
- Prepare for the ecosystem change. Conduct ecosystem process studies to inform catchability and spatial dynamics of species on seasonal level and in response to the climate change.
- Continue working on methods to design surveys that allow for flexibility in survey effort allocation between years and between areas to maximize information that can be obtained from surveys.
- Develop methods for optimal stratification and sample allocation to achieve reduced uncertainty from a given effort. For example, consider changing sampling density within strata of existing surveys in response to expected changes in distributions.
- Provide advice on how to expand surveys into the new areas where species are moving to.
- Research topics in the long-term
- Develop and test technological methods to obtain absolute or relative estimates of biomass or abundance to: calibrate existing surveys, obtain estimates of survey catchability and variation in catchability, and/or to improve outputs from existing surveys.
- Develop methods for incorporating new technologies to reduce data processing requirements. For example, use video surveys at some sampling locations.
- Perform studies on use of AI or other multivariate approaches to discover predictable relationships between available ecosystem data and data products needed for assessment and advice.

3.4 What's next?

WKUSER participants agreed that there is a need to continue WKUSER work to assure progress in modernization of the survey enterprise. The challenges and issues resolved and remaining to date are complex and require continuous work. The last two WKUSER meetings showed a very good participation rate (~ 50 people, including leading researchers from all around the world) and productive outputs. Therefore, WKUSER provides the necessary venue for continued research coordination and cooperation to assure progress in the critical research areas as defined by WKUSER TORs which are also in line with the WGISDAA and EOSG missions.

Between WKUSER 1 and 2, the focus of the workshop expanded from unavoidable survey effort reduction to adapting surveys to periodic effort reductions due to logistical issues and implementing changes to surveys that are necessary due to many different biological and ecological issues (presented above in this report). Because of this generalization of the focus of WKUSER, an additional recommendation is that the WKUSER should change its name in the future to reflect this expansion in focus.

During the October 2022 meeting, WGISDAA confirmed their support for the next workshop and discussed priorities for TORs for consideration for the next edition of the workshop. It was recommended that future work includes continued advice on how to conduct necessary changes to survey, increase in understanding of the level precision (bias) of survey data products that is required for survey data to be useful for assessments, and expanding work into data products, other than indices of abundance. WGISDAA also pointed out the importance of incorporating new technologies into existing surveys and in starting new surveys.

4 Abstracts from Presentations

Simulation tools for survey data products – Keynote presentations

Simulation tools for survey data products present and future

Sean Anderson (remote), Paul Regular, DFO (Canada)

Abstract: The adage that change is the only constant presents a major challenge to standardized surveys of marine resources. Changing technologies, weather, budgets, and priorities frequently disrupt the continuity of standardized indices and complicate the provision of science advice. Simulation studies are frequently employed to assess the impacts of such changes. Approaches can be grouped into two categories: (1) mechanistic, which aim to generate a system based on underlying theoretical relationships, and (2) empirical, which simulate from a model that describes observed data via fitted approximations. Advances in computational tools have enabled the development of species distribution models that form the basis of many contemporary simulation studies. Here, we review several such tools in the context of simulating survey data products. For example, 'virtualspecies' can simulate mechanistic species distributions based on environmental suitability; however, it lacks several complexities common in fish surveys and stock assessment. 'SimSurvey' is a semi-mechanistic tool that combines an age-structured model and Gaussian random fields to simulate populations that vary in space and time. A crux of such mechanistic models is how to condition them to reflect realistic population dynamics. One approach is through fitting related empirical models such as 'sdmTMB', 'VAST', or 'INLA', which can also be used to simulate population dynamics themselves, but lack built-in multi-stage sampling tools. We review and touch on strengths and limitations of selected tools that can evaluate stock assessment and management consequences of simulated changes to surveys including 'ss3sim', 'SSMSE', 'openMSE', and the FLR toolset. Finally, we consider future directions for simulation tools including better tools for linking empirical models to mechanistic and semi-mechanistic models, sampling tools for existing empirical simulation models, incorporating financial cost into simulation tools, improved multispecies mechanistic survey simulation tools, and tools for simulating survey data from species on the move under climate change.

The importance of multivariate and seasonal dynamics in simulation designs, and tools to address it

James Thorson, NOAA (USA)

Abstract: Fisheries scientists use simulation experiments throughout their enterprise, but reliable inference depends upon how closely the simulation model represents real-world processes. As a result, conditioning simulation models upon system observations improves inference in many simulation experiments. I therefore discuss circumstances where multivariate spatio-temporal models are uniquely suited to evaluate alternative sampling designs and estimators. I start by highlighting trait-based multispecies models, which use trait information to specify shared responses to known or latent environmental drivers, and where species complexes are represented by covariance in spatial dynamics. These have been used as simulation models to optimize

survey designs under alternative stratification and frequency constraints, but (to my knowledge) have seen limited use as operating model when evaluating closed-loop management performance.

Despite their importance, spatial simulation models typically have little capacity to represent within-year dynamics. This lack is important when evaluating new sampling designs and estimators, given that these frequently involve samples occurring at different times of year. To elaborate, I classify seasonal spatio-temporal models as being either *implicit* (where seasonality is represented as affecting catchability) or *explicit* (where density is tracked at every date, year, and location), and either *mechanistic* (where seasonal dynamics occurs via advective-diffusive-taxis movement) or *descriptive* (where seasonal patterns arise from spatially correlated variables). I highlight a few practical implementations of each, and conclude by recommending greater emphasis on seasonal spatio-temporal simulations that condition upon available data.

TOR I. Survey design for flexibility: Review and summarise desired attributes of survey design that allow for flexibility when dealing with changes in survey effort and need to expand survey into new areas of species expansion due to changes in the ecosystem.

A Flexible Approach to Optimizing the Gulf of Alaska Groundfish Bottom Trawl Survey Design for Abundance Estimation

Lewis Barnett, NOAA (USA).

Abstract: Flexible and efficient survey designs are needed to mitigate problems associated with unavoidable fluctuations in sampling effort over time. We use simulation to evaluate the performance of stratified random survey designs for a multispecies survey, the Gulf of Alaska (USA) bottom trawl groundfish survey, under alternative stratifications and sampling effort allocations to achieve targets for abundance index precision across species. In the new approach, we combined a genetic algorithm to optimize the placement of stratum boundaries over the simulated data with a multivariate Bethel optimal allocation algorithm. This method minimized total survey sample size subject to target precision constraints on abundance indices for a suite of species of high commercial or ecological relevance. Given the proposed and status quo survey designs, performance metrics of bias, precision, and uncertainty in precision estimates were computed across repeated simulations using independent draws with observation error. To determine how the spatial scale of optimization may produce the most precise and accurate abundance estimates at the scale required for informing management decisions, we conducted the optimization at two spatial scales: a regional scale and the finer scale of management area. In general, newly optimized survey designs at both spatial scales produced abundance estimates with similar precision to the status quo survey, yet also increased the accuracy of abundance estimates and both precision and accuracy of their associated variances by reducing biases present for some species. Overall, we found that precision targets across species could be met under the proposed optimal effort allocation, regardless of which species were prioritized for greater precision. We determined the proposed optimized survey is practically feasible by solving the Traveling Salesperson Problem to determine the optimal station order, which indicated that the distance between stations and total expected cruise duration is similar to the status quo. We conclude that the proposed design is expected to improve the accuracy of abundance indices and their variances for

many species with no change in average effort, while also providing improved continuity in the face of interannual changes in effort and management priorities.

Optimizing survey effort in ICES division 27.3.a, the Kattegat and Skagerrak

Patrik Börjesson[1], Annica de Groote[2]

[1]Department of Aquatic Resources, Swedish University of Agricultural Sciences.

[2]Department of Energy and Technology, Swedish University of Agricultural Sciences.

Abstract: Abundance and biomass of demersal species in the Kattegat and Skagerrak are monitored through several bottom trawl surveys, some of which have been carried out for decades. The expansion of offshore wind power and implementation of nature reserves and Natura 2000 areas may entail limitations in which areas be monitored with bottom trawl fishing in the future. In addition, increased ship costs including fuel costs and budget restrictions already limit what can be carried out in some of these surveys. Within the project "Optimizing survey effort in ICES division 27.3.a", we will evaluate the effectiveness of current and alternative survey designs, and look at options for combining different surveys. Initially, we will focus on the Swedish parts of the ongoing investigations, but we hope to expand the work to a collaboration with Norwegian and Danish colleagues.

At present, seven bottom trawls surveys operate in the area on an annual basis. These include three internationally coordinated surveys; the International bottom trawl survey (NS-IBTS), the Baltic international trawl survey (BITS) and the Kattegat cod survey (CODS), and four national surveys; the Norwegian shrimp survey (NO-SH), the Danish sole survey (DK-SO), the Swedish coastal survey (COASTS) and the Swedish Skagerrak survey (SKAGS). Together these surveys generate data from more than 400 hauls per year, approximately corresponding to 0.01 haul per km2.

The Kattegat and Skagerrak Seas were included in the NS-IBTS quarter 1 survey in the 1970s and the quarter 3 survey was initiated in 1991. The area is mainly fished by Sweden and in quarter 1 a list of fixed stations are used, currently 27 hauls in Skagerrak and 19 hauls in Kattegat. The Swedish quarter 3 survey was following the same protocol until 2005 when the design was changed to a semi-random depth stratified design, but the changes has so far not been evaluated.

In this study, we compare the depth-stratified design with the standard NS-IBTS design based on two hauls per ICES rectangle. We also evaluate the fixed station design to determine how much the different stations contribute to trends in target species abundance and indices used for evaluation of MSFD indicators. Ranking different stations will enable an informed reduction in effort if needed, and may free up ship resources that could be used to increase overlap with other surveys, or for one-off initiatives.

Use of balanced sampling for a cod survey in Kattegat

Annica de Groote[1], Patrik Börjesson[2]

[1]Department of Energy and Technology, Swedish University of Agricultural Sciences. [2]Department of Aquatic Resources, Swedish University of Agricultural Sciences.

Abstract: The Kattegat cod survey is a Swedish-Danish survey conducted annually since 2008 (except 2012) in Kattegat. It is managed jointly by SLU Aqua (Department of Aquatic Resources, Swedish Agricultural University) and DTU Aqua (National Institute of Aquatic Resources, Technical University of Denmark). The main goal of the survey is to provide fisheries independent data for monitoring trends in abundance, biomass, recruitment and distribution of cod. The results are used to strengthen scientific advice on the cod stock in Kattegat. The survey design has remained largely unchanged since 2008, but some weaknesses with the current design are emerging and could justify a change. In this presentation, we consider an alternative to the existing sampling design.

First, let's have a brief look at the main features of the survey. The main target species is cod. Due to lack of a sampling frame for cod, the target population is defined as the sea area of Kattegat and the number (or weight) of cod is treated as an attribute of this population. Sea area is converted into a finite population by use of a grid of 5×5 nmi cells placed over the area, and the list of all squares is used as the sampling frame. The main target variables are catch of cod by age class, and the parameters of interest are mainly the population totals of these variables. The population of squares is partitioned into five strata. In some strata, squares are selected by simple random sampling (SRS), and in others by independent random groups (IRG). Chartered Swedish and Danish commercial trawlers with scientific staff onboard collect the data. The vessels are free to choose a starting position and tow direction for one haul within each selected square. From collected data, population totals are estimated with the usual (Horvitz-Thompson, HT) estimator for stratified SRS.

The present design has many strengths: i) Conversion of sea area into a finite population of squares makes it possible to use well-known finite population sampling and estimation methods; ii) the squares also ensure a spatial distribution of the hauls; iii) the use of probability sampling to select squares provides some protection against selection bias and guarantees an "objective" sample; iv) the stratification is likely to improve the precision of the estimators; v) the use of IRGs in some strata facilitates investigation of potential "vessel effects" (i.e. whether reported catches vary depending on vessel); and vi) the estimation formulae are simple, at least for stratum and population totals. The design also has some weaknesses: i) it does not support estimation of parameters for geographical domains that cross the borders of squares, which is problematic since there is increasing demand for domain estimates (linked to the introduction of various fishing regulations); ii) unequal sample inclusion probabilities complicate statistical data analyses, such as regression and time series analyses; iii) the non-probability sampling of haul locations within large squares carries a risk of selection bias; and iv) use of IRG complicates planning and execution of the fieldwork and deviations from the plan are common.

Based on the above, the main requirements on a sampling design for the cod survey are the following: The design should be probability-based and make use of auxiliary variables to improve precision and ensure spatial distribution of hauls; good precision should be pursued for cod estimates and also for estimates for other species; the design should be flexible in terms of estimation for different domains of study; and the design should allow simple estimation of population parameters, as well as statistical analyses of collected data.

One alternative to the present sampling design is spatially balanced sampling. A sample is said to be "balanced" on some auxiliary variables if the HT estimates of the population totals of the auxiliary variables equal the known true values. In surveys of the environment, sampling units are likely to be spatially correlated. For instance, in a trawl survey, two hauls close to each probably produce more similar catches than two hauls far apart. Methods to distribute a sample geographically, and at the same time balance it on some auxiliary variables, are called spatially balanced sampling. One way of selecting a spatially balanced sample is the local pivotal method (LPM), which creates a strong negative correlation between inclusion indicators of units that are close to each other, based on the values of some auxiliary variables. As a result, nearby units are unlikely to appear together in a sample. Spatially balanced sampling with LPM is used e.g. in the Swedish National Forest Inventory (NFI) and National Inventories of Landscapes in Sweden (NILS).

As an illustration of the method, let us look at the NILS sampling design. The sampling frame used in that survey is a map of all land and freshwater in Sweden, plus a buffer zone. The area of the frame is denoted a_F and the sampling units are "tracts", where a tract consists of 196 circular plots ordered in a regular grid within a 1 ×1 km square. A sample of tracts is selected in two phases. First, a very large sample of tracts is selected from the frame by systematic sampling of one tract per 100 hectares, corresponding to a sampling intensity of c=1/100. Selected tracts that end up in the buffer zone are included in the sample if some of their plots are in the population. In the second phase, a final sample of tracts of size n is selected from the phase one sample, using LPM and inclusion probabilities c-1(n/aF). It follows that the sampling intensity for the final sample is constant (n/aF).

A spatially balanced design, similar to that in NILS, appears to match the sampling design requirements for the cod survey. However, the method needs to be tailored to the specific conditions and needs of the cod survey, e.g. appropriate sampling units must be defined (such as circular plots with 2.5 km radius), and auxiliary variables must be selected. In addition, the systematic sampling in phase one must be specified, since it can be conducted in different ways.

Designing for change: the impact of altering sampling design and density on survey indices

Jason Conner[1], Stan Kotwicki[1], Kotaro Ono[2] and Lewis Barnett[1]

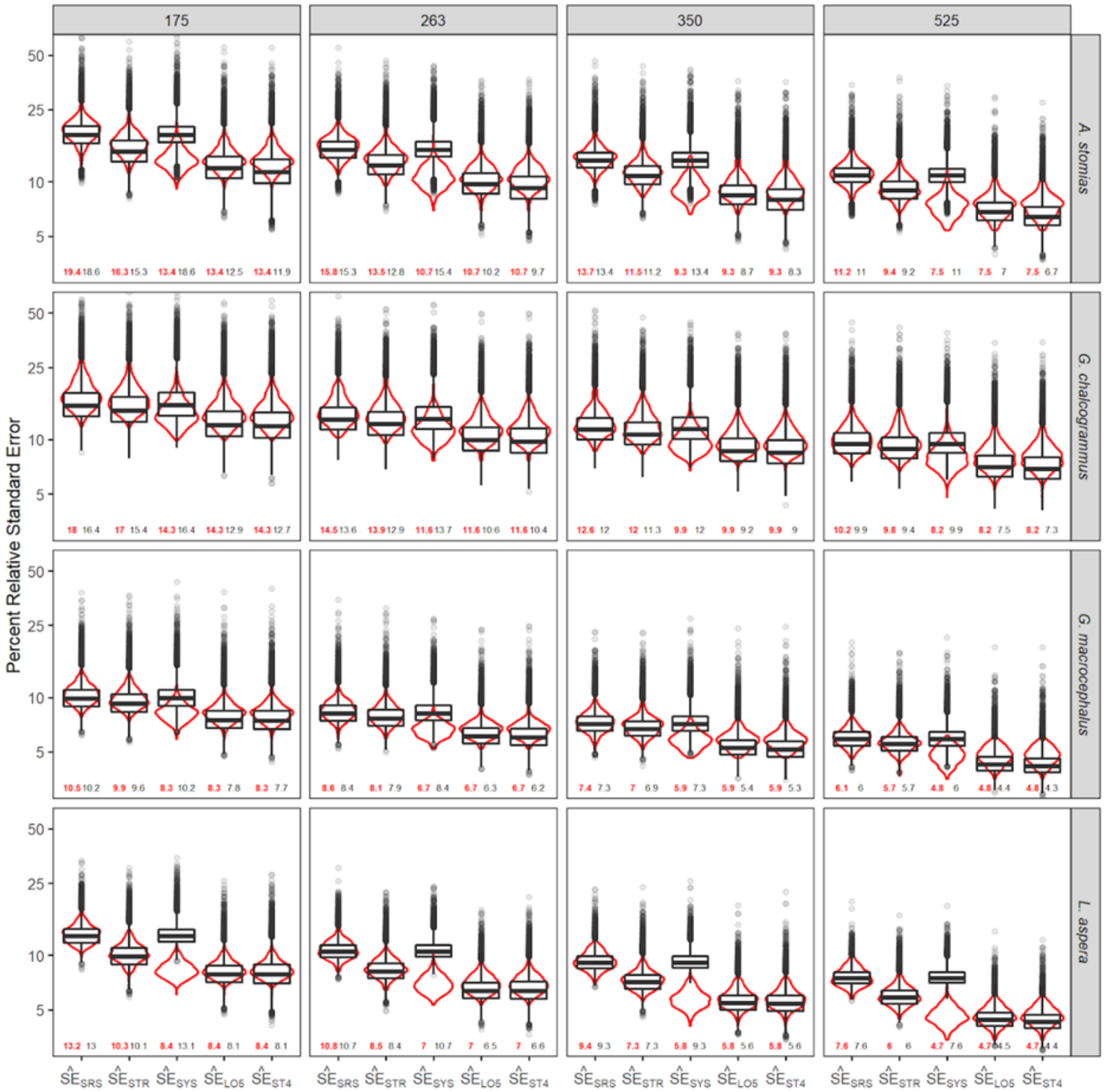
[1]NOAA (USA)

[2]IMR (Norway)

Abstract: Fisheries-independent bottom trawl surveys are employed throughout the world to provide fisheries managers with abundance indices (AIs). Disruptions to AI time-series may occur for a myriad of reasons, including: survey logistics (e.g. vessel availability or breakdowns), distribution shifts (e.g. changes in the vulnerability of a fish stock to a survey), funding shortfalls (e.g. inability to procure vessels or staffing) and conflicts with infrastructure (e.g. active fisheries affecting survey sampling). Understanding the impacts of changes to sampling design and density on the accuracy and precision of AI estimates is therefore pivotal to the sustainable management of fisheries. We present a case-study of an AI time-series of 35 years in the eastern Bering Sea for *Atheresthes stomias* (arrowtooth flounder), *Gadus chalcogrammus* (walleye pollock), *Gadus macrocephalus* (Pacific cod) and *Limanda aspera* (yellowfin sole). Using a spatio-temporal model to simulate distributions for each species, we evaluated 3 sampling designs (simple random, stratified random and systematic grid) and 4 sampling densities. Additionally, we assessed the accuracy and precision of 2 alternative estimators for estimating the standard error of the mean for systematic sampling.

Our findings support the sampling theory principle that systematic sampling results in higher precision estimates than simple random or stratified random sampling at each sampling density for each species. The increase in precision for systematic sampling designs, however, is sacrificed when borrowing the standard error estimator for random sampling, resulting in mean relative biases from 24% to 63%. These relative biases may be ameliorated by using either of the 2

alternative standard error estimators based on localized groupings of observations we evaluated. Our study establishes a benchmark for addressing the consequences of changes to survey sampling on AIs used for fisheries stock assessments.



Distribution of the percent relative standard error (*PRSE*), often referred to as the coefficient of variation, of simulated mean CPUE by sampling density (panel columns), grouped by SE estimator (X-axis) for each of 4 species (panel rows). Survey sampling and estimators consist of simple random sampling (SRS), stratified random sampling (STR) and systematic sampling (SYS). Boxplots represent the estimates PRSE (N = 35,000, means labeled in black), whereas the red violin plots represent the distribution of the "true" PRSE for each simulation (N=350; means labeled in red). Note that the Y-axis is log-scale and that the distribution of "true" PRSE is the same for SESYS, SELO5 and SEST4 estimators because these are based on the same SYS surveys.

Survey design evaluation of a new multispecies bottom trawl survey in the US Chukchi Sea

Zack Oyafuso, NOAA (USA)

Abstract: The US Chukchi Sea is an extremely dynamic area with climate change and offshore fishing affecting the marine ecosystem as well as the Arctic coastal communities that rely on healthy ecosystems. In anticipation for more frequent ecosystem monitoring in the region, it is urgent for there to be robust planning and research on the best means to flexibly allocate limited survey resources. This analysis focused on the types of bottom trawl surveys (otter and beam trawl) standardized by the NOAA-NMFS-AFSC and three types of survey designs: simple random, stratified random, and systematic. First, spatiotemporal distributions for 20+ representative demersal fish and invertebrate taxa were fitted using the VAST R package. We then simulated taxon densities from the spatiotemporal distributions to evaluate design-based estimates of abundance and precision from the three survey designs across a range of sampling effort. Modest increases in precision were gained from stratifying the design when compared to a simple random design with either a similar or decreasing level of uncertainty and bias of the precision estimates. There were often strong tradeoffs between the precision and bias of the systematic estimates of precision across species. We did not find inconsistencies in the bias of the estimated abundance indices across designs but sample precisions were slightly negatively biased, with the systematic designs being the most biased. The stratified random design provided the most consistent, reliable, and precise estimates of abundance indices and is likely to be the most robust to changes in the survey design.

Addressing interactions between offshore wind energy development and fisheries independent surveys in the United States: Development of a NOAA Fisheries and BOEM federal survey mitigation implementation strategy for the Northeast U.S region

Andy Lipsky, Jon Hare, Kathryn Ford, Brad Blythe, Brian Hooker, Brandon Jensen,, Lisa Pfeiffer

NOAA, USA

Abstract: The U.S. Department of Commerce's National Oceanic and Atmospheric Administration (NOAA) Fisheries and the U.S. The Department of the Interior's Bureau of Ocean Energy Management share a commitment to develop offshore wind energy, while protecting biodiversity and promoting ocean co-use. There are many elements to achieve these goals, including mitigation of the impact of offshore wind energy development on NOAA Fisheries surveys. Nationally, NOAA Fisheries assesses the status of approximately 450 fishery stocks, 200 marine mammal stocks, and 165 threatened and endangered species. These assessments rely on more than 50 long-term, standardized surveys, many of which have been ongoing for more than 30 years. Each survey uses different methods, platforms, and designs, with the goal of providing information on a subset of species to support sustainable management. For example, bottom trawl surveys provide information on bottom fishes, plankton surveys provide information on the early life stages of fishery species as well as ocean production (phytoplankton and zooplankton), and aircraft and vessel visual surveys provide information on the abundance and distribution of whales, dolphins, and seals. Owing to the precautionary approach, increased uncertainty in the data originating from these surveys typically results in more restrictive management. As a result, NOAA Fisheries has made extensive efforts to maintain consistency in surveys over time to reduce uncertainty and increase accuracy and precision. Sustaining surveys with consistent sampling designs/methods is an essential feature of their value, allowing understanding of the status and trends of managed species consistently through time. These surveys are essential for sustainably managing our nation's fisheries, promoting the protection and recovery of marine mammals and endangered and threatened species, conserving coastal and marine habitats; and are

also critical to understanding the impacts of climate change on marine resources and marine ecosystems. During the environmental review of the first offshore wind energy project in U.S. federal waters, four impacts to NOAA Fisheries surveys were identified: 1.Preclusion of NOAA Fisheries sampling platforms from the wind development area due to operational and safety limitations; 2.Impacts on the random-stratified statistical design; 3.Alteration of benthic and pelagic habitats, and airspace, requiring new designs and methods to sample new habitats; and 4. Reduced sampling productivity through navigation impacts of wind energy infrastructure on aerial and vessel surveys. For this reason, in 2021 NOAA Fisheries and BOEM committed to developing a federal survey mitigation implementation strategy to mitigate the impacts of offshore wind energy on NOAA Fisheries surveys. In Fall 2022, the final implementation strategy will be released. This presentation will provide an overview of the process of developing the strategy and outline key elements of the approach for developing, implementing, and adapting NOAA fisheries survey mitigation activities in response to the impacts of offshore wind development.

TOR II. Combining surveys, dealing with data gaps: Collate advice on methods to combine data from different sources, how to deal with data gaps and how to perform survey calibrations.

Using spatio-temporal models to provide compositional data for acoustic surveys: facilitating autonomous vehicle sampling and inferences on non-target species in a fishery resource survey program

Derek Bolser[1*], Aaron Berger[2], Dezhang Chu[2], Jim Hastie[2], Julia Clemons[2], Lorenzo Ciannelli[1]

[1]Cooperative Institute for Marine Resources Studies, Oregon State University

[2]Northwest Fisheries Science Center, NOAA Fisheries

*Present affiliation: Office of Science and Technology, NOAA Fisheries

Abstract: Pairing compositional (i.e., size, age) data with acoustic data is required to estimate fish biomass-at-age from an acoustic survey. Accordingly, considerable effort is expended to collect biological samples of the species or species complex of interest in most fishery resource survey programs. The need for biological samples limits the use of acoustic data collected by alternative platforms (e.g., autonomous vehicles) or surveys of non-target species for generating biomass-at-age indices for the stock assessment of a target species, even if the target species can reliably be identified in the acoustic data. However, compositional data from sources independent of the acoustic survey could be fit to spatio-temporal species distribution models and replace contemporaneously collected biological samples when their collection is not possible. We evaluated the validity of this procedure by examining a case study with Pacific Hake (Merluccius productus; 'hake') on the U.S. West Coast. Specifically, we generated estimates of compositional data with a vector-autoregressive spatio-temporal (VAST) model fit to a combination of fishery-dependent and fishery-independent data that were independent of the hake acoustic trawl (AT) survey. The performance of the VAST model was assessed with simulation testing and comparisons between VAST estimates of age composition and those from midwater trawls in the hake AT survey. The challenges we encountered when fitting the VAST model to a relatively rich dataset (e.g., data coverage, age class resolution, model stability in simulation testing) indicated

that this approach may not be suitable in all situations, but our model produced estimates of age composition that were reasonably comparable to midwater trawls (+/- 10% over the entire domain, +/- 2% in data-rich regions). Our approach allows us to use acoustic data collected in an autonomous vehicle (Saildrone) survey and a non-target (coastal pelagic species) survey to estimate hake biomass-at-age, which dramatically increases the amount of data used to understand hake biomass distribution. Ultimately, the ability to differentiate species in the acoustic data and the potential for differences between survey platforms remain major hindrances to using estimates derived from this approach in a stock assessment, but ongoing research is addressing these challenges (e.g., development of machine learning algorithms, broadband acoustics research, inter-vessel comparisons). In an increasingly challenging funding environment, using spatio-temporal models to provide compositional data for acoustic surveys could allow survey programs to maintain historical coverage or leverage acoustic data from other survey programs, ships of opportunity, and autonomous vehicles to expand survey coverage.

Integrating different fishery surveys through joint modelling

Iosu Paradinas, AZTI (Scotland)

Abstract: In this study we focused on the use of Integrated Species Distribution Models (ISDMs). ISDMs are able to formally accommodate different types of data and scale proportional gear efficiencies. ISDMs use joint modeling to integrate information from different surveys to fit shared environmental, temporal and spatial effects. Conventional ISDMs assign equal weights to every data point, thus borrowing more information from the larger dataset. This may not always be desirable, so we investigate the potential effect of different data weightings. We illustrate this method first using a simulated example and a case study that combines data coming from a fishery independent trawl survey and a fishery dependent trammel net survey on {\it Solea solea}. We find that ISDMs produce better results and improve predictions, while we urge that ISDMs require proportional gear efficiencies across surveys. Lastly, we explore ensemble modelling to combine population trends, and obtain similar results given when the response variable and link functions are the same.

Combining trap and video data from the US Southeast Reef Fish Survey to index reef fish abundance

Nathan M. Bacheler[1], Kyle W. Shertzer[1], and Walter J. Bubley[2]

[1]Southeast Fisheries Science Center, US National Marine Fisheries Service, 101 Pivers Island Drive, Beaufort, NC 28516, USA

[2]South Carolina Department of Natural Resources, 217 Fort Johnson Road, Charleston, SC 29412, USA

Abstract: Numerous economically important fish species associate with reef habitats on the continental shelf of the southeast United States Atlantic coast. These fish have been surveyed annually since 1990 using baited chevron fish traps by the Southeast Reef Fish Survey (SERFS), and video cameras were added to traps region-wide in 2011 to provide relative abundance information for trap-shy species. For many reef-associated fish species, relatively precise trap- and video-based indices of abundance (CVs < 0.20) are now available, but their inclusion in stock assessments is complicated because these indices are not independent. We review the various

approaches that have attempted to combine SERFS trap and video indices of abundance for use in stock assessments. One approach, developed by Gwinn et al. (2019), used a Bayesian state-space model that incorporated paired gear observations to estimate gear-specific catchabilities and ultimately combine trap and video data into a single index, but this approach makes the strong assumption of equal selectivity between gears. A second approach was to treat the indices as independent and include both into the stock assessment, and then have the weights of those indices sum to 1.0 in the assessment. A strength of this approach is that selectivity does not have to be equal between gears. The last approach, developed by Conn (2010), was to use a hierarchical model formulation where each index provides a sample of relative abundance, but the indices are subject to index-specific sampling and process errors. This approach was robust to differences in selectivity. The Gwinn et al. (2019) and Conn (2010) approaches are both Bayesian state-space models, but they differ in that the Conn (2010) approach uses annual index values (and can incorporate any indices, related or not), while the Gwinn et al. (2019) approach uses raw data and was specifically designed for paired observations. When SERFS trap and video indices have been combined for use in stock assessments in the southeast United States, the most common approach has been Conn (2010) with the selectivities of the two gears implicitly assumed to be equal. Some recent work has investigated ways to maintain separate selectivities within the assessment model even when fitting to the single, combined index produced by the Conn (2010) approach.

Attributes of the US-Canada Integrated Ecosystem and Acoustic Trawl Survey for Pacific hake that contribute to successful fisheries management, ecosystem monitoring, and expanded collaborations

Elizabeth Phillips[1], Rebecca Thomas[1], Julia Clemons[1], Chelsea Stanley[2], and Stéphane Gauthier[2,3]

[1]NOAA Fisheries Northwest Fisheries Science Center, Fishery Resource Analysis and Monitoring Division, Seattle, WA, USA

[2]Institute of Ocean Sciences, Fisheries and Oceans Canada, Sidney, BC, Canada

[3]Department of Biology, University of Victoria, Victoria, BC V8P 5C2, Canada

Abstract: The coast-wide Pacific hake (Merluccius productus) groundfish stock on the west coast of the U.S. and Canada has been monitored jointly by both countries since 1995. The Pacific Hake/Whiting Treaty was signed in 2003, and the joint U.S.-Canada Integrated Ecosystem and Acoustic Trawl Survey for Pacific hake has been led since then by scientists from Canada's Department of Fisheries and Oceans (DFO) and NOAA Fisheries Northwest Fisheries Science Center's (NMFS) newly formed Fisheries Engineering and Acoustic Technologies team. The coast-wide stock assessment survey is conducted biennially in odd years, and biomass estimates are provided to stock assessors working under the joint treaty. An additional stock assessment survey was also conducted in 2012 due to high uncertainty in the 2009 and 2011 assessments. During even years, research cruises are conducted to improve and expand survey capabilities, refine acoustic species discrimination and classification, and address Scientific Review Group (SRG) requests aimed at improving the stock assessment.

The success of this joint survey and management approach is due in large part to cooperation between scientists and managers from NMFS and DFO, which allows for a flexible survey design and greater ability to combine surveys and deal with data gaps. This includes coordinated survey and research planning throughout the year, open and transparent communication between

NMFS and DFO survey and stock assessment scientists, and built-in modularities that ensure complete surveys of the entire coastal Pacific hake stock.

A key attribute of the joint U.S.-Canada Integrated Ecosystem and Acoustic Trawl Survey for Pacific hake is its two-vessel design, which allows for the entire area from southern California to Dixon Entrance, Alaska to be surveyed in approximately 3 months between June and September. The survey consists of 10-nmi spaced transects along most of the coast, covering ~221,000 km2, and the two-vessel design overcomes limitations in days-at-sea allocations for each survey team, which typically are not enough to cover the full geographic area in the required time window. By splitting the survey effort between two vessels, one led by NMFS and one led by DFO, the spatiotemporal resolution of the sampling, including transect spacing, is less likely to be sacrificed. Furthermore, the use of two vessels allows for coordination of geographic coverage by vessel, including interleaving transects to increase efficiency, and flexibility in the northern and southern survey extent.

The two-vessel design also provides flexibility in the event that either vessel becomes unavailable. Because both countries are committed to covering the entire geographic area without prioritizing work in their country's waters over the other country, scientists can effectively collaborate on operations to ensure survey goals are met when ship time is limited or a ship becomes unavailable. For example, in 2021 survey scientists from NMFS and DFO coordinated efforts to maximize transect resolution to ensure complete geographic coverage when DFO's vessel became unavailable, resuming the original survey design after another DFO vessel was identified. Regular communication with the Joint Technical Committee ensured that changes in survey design and sampling resolution were acceptable for stock assessment purposes. Similarly, in both 2005 and 2007, DFO's vessel experienced mechanical issues, and the NMFS vessel was able to take over the full survey.

A critical component of combining data from two platforms is an inter-vessel calibration (IVC) to ensure that the data collected by each vessel are comparable and not biased. Each time a new vessel is used by either country, an IVC must be performed. In 2019 the CCGS Sir John Franklin became the primary survey platform for DFO, although mechanical issues and the COVID pandemic prevented use of this vessel until 2022, when a 7-day IVC was conducted to combine data with NOAA Ship Bell M. Shimada. The IVC involved both vessels operating in close proximity, side-by-side, collecting simultaneous acoustic data from the same area to compare hake Nautical Area Scattering Coefficient (NASC) across multiple hake aggregation sizes and water depths. This involved constant communication between the vessels, and close coordination of transect and fishing efforts. Data sharing during the IVC and all survey effort is another important component of the joint survey, and applies to echo sounder calibration, data acquisition at-sea, expert scrutinized and judged echogram data exports, and the hake biomass estimate that is provided to stock assessment scientists. NMFS and DFO scientists coordinate acoustic and trawl data sharing throughout and after each survey, making data processing and biomass estimation efficient and collaborative. Furthermore, scientists from both groups have laid out clear priorities in terms of data requirements, making it easier to prioritize decisions related to survey effort and design that may impact data acquisition and quality. Open discussions about maximizing at-sea data collections have also led to broader ecological sampling and monitoring, including oceanographic data collections during nighttime operations when trawling operations are not occurring. This coordinated approach also facilitates expanded collaborations within NMFS and DFO programs, and with external partners that are expanding our ability to provide data to multiple stakeholders.

Cooperative rockfish research as supplemental survey effort: an early example from the Gulf of Alaska

Madison Hall, Mark Zimmermann, Pete Hulson, Brad Harris, Suresh Sethi, Curry Cunningham, John Gauvin, Julie Bonney, Stan Kotwicki

NOAA (USA).

Abstract: In Alaska, groundfish bottom trawl surveys conducted by the National Marine Fisheries Service (NMFS) use standardized fishing gear that is unable to sample in hard, rough, and rocky areas. Over the lifetime of the Gulf of Alaska (GOA) bottom trawl survey, many rough and rocky habitat stations have been marked as "untrawlable" and eliminated from sampling plans. For fish species that prefer rocky habitats, biomass estimates from the standardized surveys may be imprecise and/or negatively biased due to the exclusion of these areas. Even though NMFS is unable to sample in "untrawlable" habitats, several species that inhabit rocky habitat types are successfully targeted by commercial fisheries. Specifically, Pacific Ocean perch (*Sebastes alutus*), northern (*S. polyspinis*) and dusky (*S. variabilis*) rockfish are commercially important to the Amendment 80 catcher-processor fleet and the Kodiak-based catcher vessel trawl fleet in the Gulf of Alaska. Our project, the Science-Industry Rockfish Research Collaboration in Alaska (SIRRCA), has been working with the GOA fishing industry to build a cooperative survey that uses industry vessels and gear to sample in areas that are "untrawlable" to the NMFS bottom trawl survey. During the pilot phase of our project we have collected data on two catcher/processor vessels and one catcher vessel, successfully sampling 34 NMFS survey stations. SIRRCA project tows focus on either calibrating catch per unit effort (CPUE) between vessels, or gathering rockfish biomass information at “untrawlable” survey stations. We plan to use SIRRCA data as a supplemental source of population abundance data to inform, and perhaps improve, official GOA rockfish stock assessments. Additionally, we are using simulations to explore the impacts of omitting important habitats on rockfish biomass estimate bias and CV. We believe SIRRCA’s cooperative science model holds great promise for fisheries science, and we hope that sharing our cooperative model will help the development of other cooperative fishery projects both nationally and globally.

TOR III. Modelling and simulations: Further develop model performance evaluation through simulations, use of auxiliary information to improve survey data products, including appropriate propagation of uncertainty.

Model-based expansion of age, length, and prey subsampling while estimating equivalent multinomial sample size

James T. Thorson, NOAA (USA)

Abstract: Multivariate spatio-temporal models can be fitted to point-count data for multiple species, ages, sizes, or other relevant categories. The approach can standardize spatially unbalanced survey data, and estimated age-compositions are used annually as data in major stock assessments using bottom-trawl surveys in the eastern and northern Bering Seas. Here, I briefly review how the estimated precision for model-based abundance-at-age can be converted to an approximate multinomial sample size, for use as an upper bound on model-based age-composition weighting. I also outline how this approach has been extended to estimate total predator consumption based on independent surveys of predator density and stomach contents. I use these examples to argue that the approximate multinomial sample size is a natural and simple metric for measuring the impact of effort reductions on the precision of survey products. I conclude by recommending further research regarding model performance, either empirically testing how subsampling survey effort (either age and length subsampling, or total survey stations) affects either estimated precision, or simulation-based evaluation of bias and confidence interval coverage.

Multivariate geostatistical models for correcting sampling bias

Maria Grazia Pennino[1], Xavier Barber[2], David Conesa[3], Iosu Paradinas[4] Joaquín Martínez-Minaya[5], Antonio López-Quílez[3]

[1]Centro Oceanográfico de Vigo, Instituto Español de Oceanografía (IEO-CSIC), Subida a Radio Faro, 50-52, 36390 Vigo, Spain.

[2]Center of Operations Research (CIO), Universidad Miguel Hernández, 03202 Elche, Spain.

[3]Department of Statistics and Operations Research, University of Valencia, 46100 Valencia, Spain.

[4]Scottish Ocean's Institute, University of St Andrews, St Andrews KY16 9AJ, UK.

[5]Department of Applied Statistics and Operational Research, and Quality, Universitat Politecnica de Valencia, Valencia, 46022, Spain.

Abstract: Information about the distribution of marine species can essentially be derived from two main sources, namely, fishery-independent data (scientific surveys at sea) and fishery-dependent data (collection and sampling by observers in commercial vessels). Commonly scientific survey data are considered to be of higher quality because sampling and collection are

scientifically designed and standardized to ensure that estimates, as species occurrence and abundance, are unbiased.

However, under certain circumstances, surveys may produce imprecise estimates of species occurrence and abundance, particularly for species with preferential habitats that are in strata only partially included in the survey sampling design or for catchability issues. Indeed, catchability is influenced by diverse sources of variation such as size, time, density, and location. The spatial variation of catchability reflects some aspects of resource behaviour like population aggregation and movement, and habitat preference by developmental stage. In these cases, the spatial species variation is not adequately captured. This unequal survey coverage of a species distribution is often referred as sampling bias, sample selection bias or survey bias.

In these cases, the quality of the estimated species distribution can be strongly affected if entire parts of the environmental space suitable to a species are absent or poorly represented in the survey dataset. However, during scientific-surveys multiple species are collected at each spatial location which imply not only a dependence between species at each location but also association between species across locations. As a result, including as many as possible of these dependence relationships in species distribution models (SDMs) would help us to achieve unbiased estimations of occurrence and abundance. Multivariate geostatistics provides a good solution to study the behaviour of spatially correlated species by including data observed for each species as a multivariate response variable. In other words, multivariate geostatistics expands the idea of single-species SDMs to multi-species SDMs by implementing a Joint Species Distribution Model (JSDM from now on) as a model containing one spatial regression model for each species.

As predicting in multivariate geostatistical models can be complicated, we propose the use of Bayesian coregionalised models for multivariate spatial data, that have proved to work well in applied contexts. In the case of two species, a coregionalised model basically consists of a spatial regression model that describes the occurrence (or the abundance) of one of the species, and another spatial regression model that describes the occurrence (or the abundance) of the second species, but conditional on the latent geographical pattern of the first one. In our fishery context, this would imply that species well represented in scientific surveys could be used to infer the spatial behaviour of those species correlated but poorly represented.

In line with this context, here we discuss a computational efficient tool to perform inference and prediction in JSDMs using the Integrated Nested Laplace Approximation (INLA) and associated software. We illustrate the performance of the coregionalized model in species interaction scenarios using both simulated and real data. The simulation demonstrates the better predictive performance of the coregionalized model with respect to the univariate models. The case study focuses on the spatial distribution of a prey species, the European anchovy (Engraulis encrasicolus), and one of its predator species, the European hake (Merluccius merluccius). Results presented demonstrate how multivariate geostatistical models can show underlying patterns of similarities and differences in distributions between species and correct possible sampling bias.

Quantifying the Scientific and Economic Value of Surveys to Fisheries Management

Lee Cronin-Fine[1], Meaghan Bryan[2] (presenter), Stephen Kasperski[2], Lewis Barnett[2], and André E. Punt[1]

[1]University of Washington, School of Aquatic and Fishery Sciences, Seattle, WA

[2]Alaska Fisheries Science Center, NOAA Fisheries, Seattle, WA

Abstract: Stock assessments are used in fisheries management to help sustain fisheries as a valuable economic resource. They use multiple data sources such as catch, discard and surveys. Surveys are an important source of information since they are independent of fisheries and collect data in a more standard scientific fashion. Climate change can impact fisheries in multiple such as range shifts. Studies show that climate change induced range shifts have been occurring to multiple species in the Bering Sea of Alaska. Surveys are important information sources that help track these range shifts in populations. Unfortunately, surveys are very costly to run and could be modified in the future. Therefore, the goal of this project is to evaluate the scientific and economic benefits of different survey strategies through a management strategy evaluation (MSE) for fisheries in the Bering Sea of Alaska. This talk will discuss assumptions within the MSE such as how the impact of climate change is modeled and how to evaluate the economic impact of different survey strategies.

Geostatistical models limit impact of reduced survey coverage on indices of abundance

Semra Yalcin[1], Paul Regular[1], Sean C. Anderson[2]

[1]Northwest Atlantic Fisheries Center, Fisheries and Oceans Canada, NL, Canada

[2]Pacific Biological Station, Fisheries and Oceans Canada, Nanaimo, BC, Canada

Abstract: Changes in survey effort, such as reduced survey set density, reduced survey coverage, or exclusion from closed areas, are common challenges that may introduce bias and noise to indices of population size. We evaluated the performance of different methods for computing indices through simulations under a range of survey conditions. We simulated the population abundance and distribution of two fish species across space and time with SimSurvey: cod-like (size-specific spatial clustering) and flatfish-like (diffuse distribution). We then applied a stratified random sampling procedure for each population as a base condition. For each species, we reduced survey effort by decreasing set density in spatially random (e.g., budget reductions), spatially blocked (e.g., introduction of marine protected areas), or blocked patterns that eliminated entire strata. We then calculated abundance indices using design-based and model-based (spatiotemporal models fit with sdmTMB) approaches under several model parameterizations. For each simulation scenario and analytical approach, we calculated bias, accuracy, and coverage to assess the estimators' ability to recover known indices. We found that correctly specified spatiotemporal models had bias, accuracy, and coverage comparable to design-based estimators in base conditions or under spatially random survey effort reductions. However, model-based approaches with a depth covariate (especially for the more diffuse flatfish-like species with a strong depth preference) outperformed design-based estimators in all measures when survey effort was reduced in spatially blocked patterns. This effect was most pronounced when entire strata were lost; in this case model-based approaches—regardless of model—outperformed the design-based approach. Overall, model-based approaches can reduce the impact of blocked survey effort loss and therefore minimize effects on population assessment and the provision of scientific advice.

PhD project: Improving fishing survey indices though the use of spatio-temporal models

Anna Stroh, ATU, Ireland

Abstract: In the context of understanding the dynamics of fish populations, a 4-year PhD research project is introduced, focused on Improving fishing survey indices through the use of spatio-temporal models. Four main research objectives are defined: (1) Researching survey index estimation with a focus on survey gaps, (2) Influence of environmental covariates on univariate spatio-temporal models and derived indices, (3) Multivariate spatio-temporal modelling for survey index standardization, and (4) Spatio-temporal modelling and survey design. The spatio-temporal modelling in this project will be applied to case studies in the Celtic Sea i.e. whiting.

GAM-based methods for abundance indices in the Bering Sea

Siple, Margaret C., Markowitz, Emily H. Kotwicki, Stan., Thorson, James., Barnett, Lewis A. K.

NOAA (USA)

Abstract: Model-based indices of abundance have the potential to improve our understanding of the state of fisheries stocks compared to "design-based" indices of abundance based on stratum-weighted catch per unit effort (CPUE). However, such design-based and model-based estimates may differ among approaches and methods in terms of estimated scale (i.e., the average of the estimates over time) or trends (i.e., rates of change over time). The relative performance of spatial generalized additive models (GAMs) and spatial generalized linear mixed models (GLMMs, e.g., VAST) has been evaluated using simulated data. Here we use empirical observations from the NOAA Alaska Fisheries Science Center RACE bottom trawl surveys to compare them to one another and the traditionally used design-based indices. These models specify different spatial basis functions, estimate the penalty for basis-function coefficients differently , and differ in how (and whether) they can fit potentially complex, nonlinear relationships in fixed effects. The latter will be particularly important when there are nonlinearities in how the environment drives abundance, as some models more easily fit complex, nonlinear relationships in fixed effects. At this working group meeting, we will present preliminary results comparing a design-based approach, VAST, and GAMs, using data from pollock, yellowfin sole, and red king crab from the Eastern Bering Sea Bottom Trawl Survey. Specifically, we will compare predictions using cross-validation and the covariate effects estimated by each model-based index of abundance.

We will discuss appropriate methods for evaluating and comparing design- and model-based indices of abundance, including blockwise cross-validation and RMSE for the existing models, and the potential for developing a model comparison toolbox for evaluating design-based indices and those derived from species distribution models. This project will ultimately compare model-based indices of abundance created using GAMs with those created using GLMMs and with "design-based" indices of abundance (stratum-weighted mean CPUE) for three survey regions in Alaska (the Gulf of Alaska, Aleutian Islands, and the Bering Sea).

TOR IV. Tools and technology development: Describe technological and analytical tools (e.g. R packages, AI, video analysis, etc.) that can provide quantitative assessment of the effect of effort changes on the quality of survey deliverables and advisory products

Survey Simulation Experimentation and Evaluation Project (SSEEP): evaluating the impacts of offshore wind to the Northeast bottom trawl surveys

Angelia Miller, Gavin Fay, Philip Politis, Andrew Lipsky, Catherine Foley, Kathryn Ford, Catalina Roman, Madeleine Guyant, Catherine O'Keefe

University of Massachusetts Dartmouth (USA)

Abstract: When considering multiple-use management effects on fisheries (e.g., interactions due to infrastructure siting for other marine use sectors), there is a need to understand how spatial impacts to monitoring programs propagate through to scientific uncertainty in fisheries governance. Planned (and in-progress) wind development areas in the Northeast US and Mid-Atlantic overlap extensively with the footprints of the Northeast Fisheries Science Center (NEFSC) Multispecies Bottom Trawl Surveys, which have been conducted biannually since the 1960s and form the basis of fishery-independent data used in stock assessments for commercial and recreationally important fisheries. Current survey operations will be unable to take place near wind sites. Without changes to statistical design, methods of sampling, or supplemental monitoring these impacts to survey operations may diminish the accuracy and representativeness of data the surveys produce and affect the performance of stock assessment models and stock status determinations.

The Survey Simulation Experimentation & Evaluation Project (SSEEP) will apply a spatially-explicit observation simulation model for the NEFSC Bottom Trawl Surveys to evaluate the efficacy and statistical properties of changes to survey design as a result of preclusion to offshore wind areas, and also assess the performance of alternative methods for monitoring groundfish distribution, abundance, and trends. A general overview of the modeling framework is provided in Figure 1. The model will be able to emulate environmental drivers of fish population dynamics and spatial distribution, fish resource distribution, and alternative sampling strategies and design, and from this produce survey data products. Changes to the accuracy and precision of derived data products will then be used to compare performance of survey design alternatives given the various scenarios.

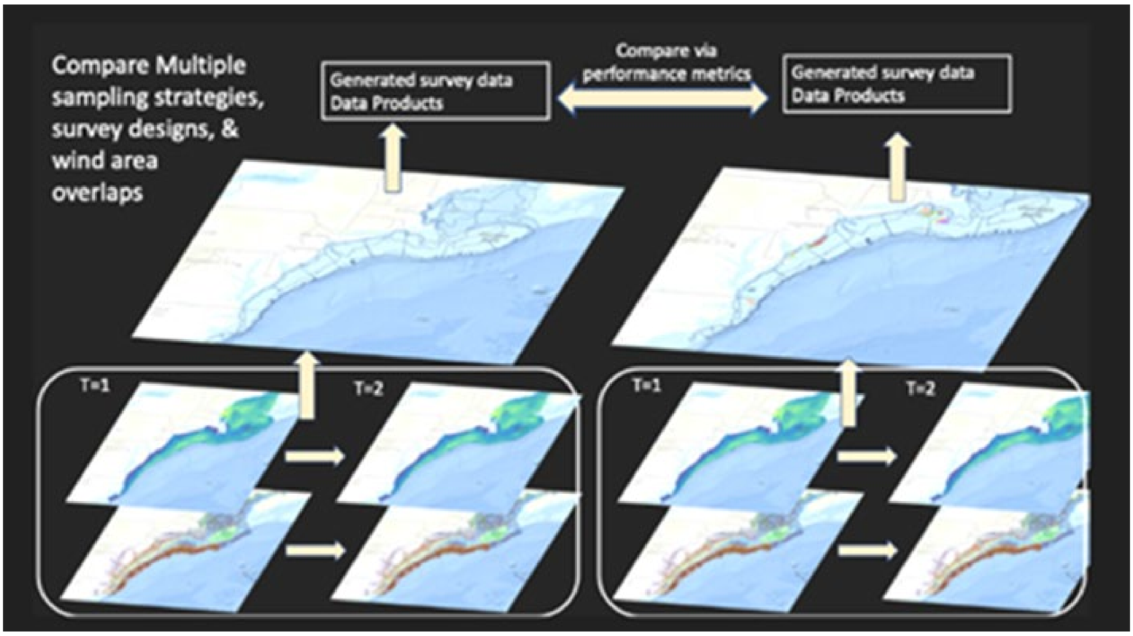


Figure 1. A simple representation of the SSEEP modeling framework.

Simulation scenarios for comparison include species interactions with proposed wind areas and the survey based on a set of species characteristics; changes to species' productivities and distributions in the presence of wind development and/or other drivers of change such as climate change effects; implementation timelines of wind installations and its impacts on the survey; and alternative sampling designs and supplemental sampling strategies to maintain the efficacy of the bottom trawl survey in the presence of wind development. To simulate a species distribution, its productivity and its availability to sampling, SSEEP will create a spatial grid that mirrors the NEFSC bottom trawl survey and sample simulated age-structured populations by employing and extending the SimSurvey R package (Regular et al. 2020). A simplified example of such a grid along with an area closed to sampling due to wind as represented by the blue rectangle is illustrated in Figure 2. Spatial allocation of survey tows may include or exclude areas depending on the degree to which parts of the survey area are precluded by wind.

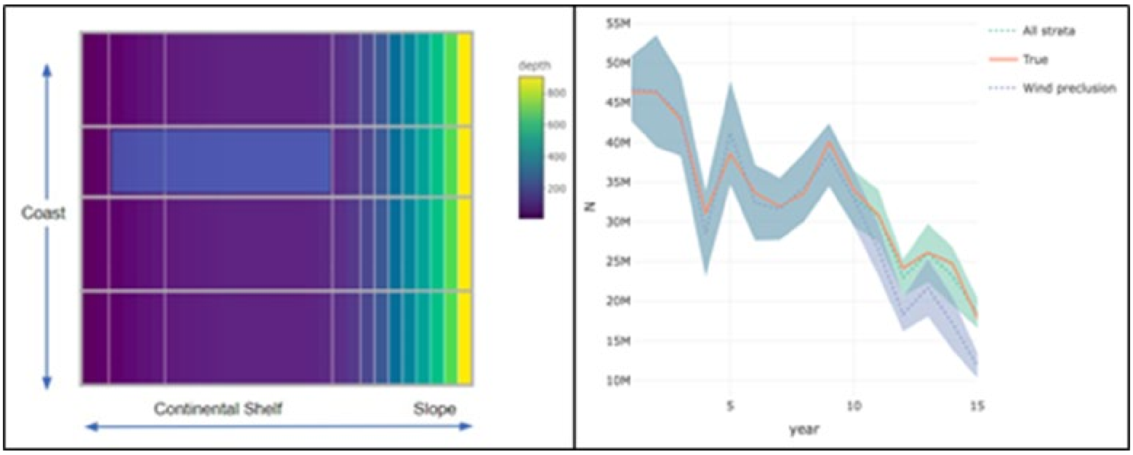


Figure 2. A spatial grid (left) representing the coast, continental shelf, and slope with a wind area as the blue rectangle representing areas unavailable to the trawl survey, which will be used to distribute and sample a population (right).

Overall, SSEEP is aimed at understanding the effects of changes in scientific monitoring as a result of increases in ocean use on the data used to support management advice and knowledge of ecosystem status. The project will also provide a framework for testing how scientific monitoring can adapt to meet current and future anticipated needs as part of coordinated, integrative approaches to ecosystem-based management of marine and fishery systems while continuing to support best available science.

Reference: Regular PM, Robertson GJ, Lewis KP, Babyn J, Healey B, Mowbray F (2020) SimSurvey: An R package for comparing the design and analysis of surveys by simulating spatially-correlated populations. PLoS ONE 15(5): e0232822. https://doi.org/10.1371/journal.pone.0232822

The estimated impact of changes to otolith field-sampling and ageing effort on stock assessment inputs, outputs, and catch advice

Matthew R. Siskey[1,5], André E. Punt[1], Peter-John F. Hulson[2], Meaghan D. Bryan[3], James N. Ianelli[3], James. T. Thorson[4]

[1]School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA, USA

[2]Auke Bay Laboratories, Alaska Fisheries Science Center, NMFS, NOAA, Seattle, WA, USA

[3]Resource Ecology and Fisheries Management Division, Alaska Fisheries Science Center, NMFS, NOAA, Seattle, WA, USA

[4]Habitat and Ecological Processes Research Program, Alaska Fisheries Science Center, NMFS, NOAA, Seattle, WA, USA

[5]Current affiliation: Washington Department of Fish & Wildlife, Olympia, WA, USA

Abstract: Generating accurate data for stock assessments is resource-demanding, necessitating periodic evaluation of survey sampling designs and potential impacts on stock assessments. We developed a framework for bootstrapped resampling of survey age data and calculation of input sample sizes as a function of among-bootstrap variance in age compositions. Data from this bootstrap estimator were then used to evaluate the influence of alternative sampling rates and methods on uncertainty in estimates of overfishing limit (OFL) calculated using stock assessment models. For dusky rockfish and Pacific ocean perch, a 10% decrease in the number of tows sampled upon led to a predicted 5-6% increase in the CV of OFL (log-log slope = -0.576 to -0.486), which was greater than the 0-2% increase from a 10% decrease in otoliths-per-tow (log-log slope = -0.238 to -0.029). Application of this approach across all stocks monitored in the survey of interest is required to identify which stocks (i) benefit the most from increased sampling of ageing structures, or (ii) cost the least in terms of OFL uncertainty owing to reduced sampling.

Evaluating the effects of survey effort reduction for the NOAA eastern Bering Sea bottom trawl survey

Lukas DeFilippo, NOAA (USA).

Abstract: Fisheries-independent surveys provide critical data products used to estimate stock status and inform management decisions. However, changes to sampling designs can complicate the interpretation of survey data, as it may be difficult to disentangle the effects of such changes from demographic variation in the surveyed population. Nonetheless, changes in the density

and allocation of sampling effort may be advantageous or unavoidable due to budgetary and staffing constraints, hazardous weather, habitat becoming inaccessible, and shifts in species distributions and abundance over time. While the ability to modify and redistribute sampling effort to address changing monitoring needs can improve survey flexibility, it is important to consider the consequences of such changes before their implementation. Here we present a framework for evaluating reductions in sampling effort that relies on existing survey time-series and simulation analyses. We apply this approach to evaluate the potential removal of high density sampling areas surrounding St. Matthew and the Pribilof Islands from the NOAA Alaska Fisheries Science Center's eastern Bering Sea (EBS) survey grid. These high density sampling areas were implemented to improve monitoring of king crab stocks (Paralithodes sp.) in these areas, which historically supported commercial fisheries but have since declined and are seldom eligible for harvest. Our approach considers the effect of removing these high density sampling areas on the focal blue (P. platypus) and red king (P. camtschaticus) crab stocks of these areas, as well as other crab and groundfish species monitored by the EBS survey. Using a combination of empirical and simulation-driven approaches, we estimate the effects of removing these high density sampling areas on the precision and accuracy of survey data products, as well as stock assessment outcomes. Our analyses provides a generic framework for evaluating tactical reductions in survey effort that can be readily applied to other species and regions.

GAM for Abundance Index Standardization of Alaska Bottom Trawl Survey Data

Markowitz, Emily H., Siple, Margaret C., Kotwicki, Stan., Thorson, James.,Barnett, Lewis A. K. Berg, Casper W.

NOAA (USA)

Abstract: Indices of abundance estimates that incorporate spatiotemporal environmental covariates can improve our past, present, and future understanding of the state of fishery stocks. Environmental covariates like depth, temperature, and the Bering sea cold pool extent have been shown to drive the distribution and abundance of marine organisms in Alaska's subarctic seas. These covariates may have complex, nonlinear effects on abundance that may not be well approximated by linear models often used to produce abundance indices. Generalized additive models (GAMs) utilize bivariate smoothing functions to estimate nonlinear responses and may make it easier to avoid overfitting and better account for structural uncertainty in response mechanisms. This presentation focuses on 1) preliminary applications of GAMs to produce indices of abundance, produced with assistance from the {surveyIndex} R package, for two groundfish and one crab species caught in the eastern Bering sea (EBS) bottom trawl survey conducted by the NOAA Alaska Fisheries Science Center (AFSC) Resource Assessment and Conservation Engineering (RACE) Division; and 2) efforts to adapt this approach for use for other regions, surveys, and species. Our example GAMs incorporate environmental covariates currently used in VAST-based abundance indices produced at AFSC so that we can ultimately compare how environmental covariates and parameterization are handled between these two model-based abundance indices and a design-based index. We also review secondary model testing and simulation analysis tools that can be harnessed from the {surveyIndex} package to compare and interpret indices, such as cross-validation and comparing covariate effects. We propose further extensions of this tool to accommodate changes in survey effort, implementation and adaptation to data and surveys in other regions, as well as the addition of standardized outputs that will have broad utility across stocks and environmental drivers.

Annex 1: List of participants

First Name	Last Name	Affiliation	E mail
Sean	Anderson	DFO, Canada	Sean.Anderson@dfo-mpo.gc.ca
Nate	Bacheler	NOAA, USA	Nate.Bacheler@noaa.gov
Lewis	Barnett	NOAA, USA	lewis.barnett@noaa.gov
Casper	Berg	DTU-Aqua, Denmark	cbe@aqua.dtu.dk
Kristan	Blackhart	NOAA, USA	Kristan.Blackhart@noaa.gov
Derek	Bolser	NOAA, USA	derek.bolser@noaa.gov
Patrik	Börjesson	SLU Aqua, Sweden	Patrik.Borjesson@slu.se
Meaghan	Bryan	NOAA, USA	meaghan.bryan@noaa.gov
Corina	Chaves	IPMA, Portugal	corina@ipma.pt
Elizabeth	Chilton	NOAA, USA	Elizabeth.Chilton@noaa.gov
Jason	Conner	NOAA, USA	jason.conner@noaa.gov
Annica	De Groote	SLU, Sweden	annica.isaksson.de.groote@slu.se
Lukas	DeFilippo	NOAA, USA	Lukas.DeFilippo@noaa.gov
Paul	Dolder	CEFAS, UK	paul.dolder@cefas.co.uk
Catherine	Foley	NOAA, USA	Catherine.Foley@noaa.gov
Hans	Gerritsen	MI, Ireland	hans.gerritsen@marine.ie
Madison	Hall	NOAA / Alaska Pacific University, USA	madison.hall@noaa.gov
Espen	Johnsen	IMR, Norway	espen.johnsen@hi.no
Eoghan	Kelly	MI, Ireland	eoghan.kelly@marine.ie
Stan	Kotwicki	NOAA, USA	stan.kotwicki@noaa.gov
Sven	Kupschus	JRC, Europe	sven.kupschus@ec.europa.eu
Gwladys	Lambert	CEFAS, UK	gwladys.lambert@cefas.co.uk
Andy	Lipsky	NOAA, USA	andrew.lipsky@noaa.gov
Emily	Markowitz	NOAA, USA	emily.markowitz@noaa.gov
Guillermo	Martin	MI, Ireland	guillermo.martin@marine.ie
Angelia	Miller	UMASSD, USA	amiller7@umassd.edu
Coilin	minto	ATU, Ireland	coilin.minto@atu.ie

Bríd	O'Connor	MI, Ireland	b.oconnor92@gmail.com
Thibault	Cariou	MI, Ireland	Thibault.Cariou@Marine.ie
Kotaro	Ono	IMR, Norway	kotaro.ono@hi.no
Zack	Oyafuso	NOAA, USA	Zack.Oyafuso@noaa.gov
Iosu	Paradinas	AZTI, Spain	jparadinas@azti.es
Maria Grazia	Pennino	IEO-CSIC, Spain	grazia.pennino@ieo.csic.es
Elizabeth	Phillips	NOAA, USA	elizabeth.phillips@noaa.gov
John Gabriel	Ramirez	GFCM, Italy	johngabrielrt@gmail.com
Yves	Reecht	IMR, Norway	yves.reecht@hi.no
Paul	Regular	DFO, Canada	Paul.Regular@dfo-mpo.gc.ca
Margaret	Siple	NOAA, USA	Margaret.Siple@noaa.gov
Matt	Siskey	W-DFW, USA	Matthew.Siskey@dfw.wa.gov
David	Stokes	MI, Ireland	david.stokes@marine.ie
Anna	Stroh	ATU, Ireland	anna.stroh@research.atu.ie
James T	Thorson	NOAA, USA	james.thorson@noaa.gov
Ralf	van Hal	WMR, Netherlands	ralf.vanhal@wur.nl
Joel	Vigneau	IFREMER, France	joel.vigneau@ifremer.fr
Daniel	Vilas		danielvilasgonzalez@gmail.com
Kai	Wieland	DTU-Aqua, Denmark	kw@aqua.dtu.dk
Semra	Yalcin	DFO, Canada	Semra.Yalcin@dfo-mpo.gc.ca

Annex 2: Resolution

2021/WK/EOSG03 The **Workshop on unavoidable survey effort reduction (WKUSER2)**, chaired by Stan Kotwicki, US, Hans Gerritsen*, Ireland Kotaro Ono*, Norway will meet in Galway, Ireland on 13-17 September 2022 to:

a) **Survey design for flexibility.** The workshop will review and summarise desired attributes of survey design that allow for flexibility when dealing with unavoidable reductions and increases in survey effort and need to expand survey into new areas of species expansion due to changes in the ecosystem.

b) **Combining surveys, dealing with data gaps.** Collate advice on methods to estimate combine data from different sources, how to deal with data gaps and how to perform survey calibrations.

c) **Modelling and simulations.** Further develop model-based estimation, model validation through simulations, use of auxiliary information to improve survey data proucts, including appropriate propagation of uncertainty.

d) **Tools and technology development.** Describe the development of methods that aim to provide quantitative decision-making tools that describe the effects on the quality of the survey deliverables and ultimately advisory products.

WKUSER will report to by 22nd October for the attention of ACOM/SCICOM trough EOSG and DSTSG.

Supporting information:

Priority	Marine surveys are expensive and under recent budgetary, poli tical, and pandemic associated presussures a number of decisions on survey implementation have had to be made at very short notice and with little opportunity to evaluate different options for effort reductions the effects of which will only become apparent in the next few years. The previous workshop WKUSER (2020) identified that such changes are recurring theme in many monitoring agencies, and more coherent planning and a long-term response strategy is desirable. It is in the interest of national governments making the decisions and ICES using such information for their advice to have a better understanding of the effects on stock assessment advice and a clearer understanding of the mitigation measures that can be implemented to minimse the impact of such events.

Scientific justification	Most survey programs are at one time or another asked to make substantial short term changes in survey effort due to budgetary constrains or need for more information. Usually these requests leave little time for planning and evaluation. There is a real need to develop methods that provide a better understanding of the different implementation options, and investigation of methods that can help to optimise available resources to maximise information obtained from surveys. Often survey scientist / managers are having to make near instantaneous decisions, the advisory consequences of which are poorly understood by the decision makers. Having a framework or a set of methods that can be applied to the specific problem is highly valuable together with summarisations of findings for general cases, which allow survey scientist to make decisions in the absence of data or the opportunity to evaluate options statistically.
Resource requirements	Many different approaches to evaluate effects and survey options have been developed independently at different times in response to specific cases. A large part of this work is to research programmes which provide the main input to this group are already underway, and resources are already committed. The additional resource required to undertake additional activities in the framework of this group is negligible.
Participants	Expected attendance 20–30 survey and assessment scientists along with monitoring program managers.
Secretariat facilities	None.
Financial	No financial implications.
Linkages to advisory committees	There is a direct link with the advisory committee as they require knowledge on the sensitivity of the advice to changes in surveys in order to provide precautionary advice when survey information is compromised.
Linkages to other committees or groups	The wporkshop should link closely back to WGISDAA which will maintain the tools / methods and broaden the approach over time. Work with stock assessment WG is thought to be essential.
Linkages to other organizations	The work of this group is closely aligned with similar work in FAO and in the Census of Marine Life Programme.

Annex 3: Workshop agenda

Tuesday, September 13th – Auditorium

Registration and introduction	
09:00	Bus from galway city centre
09:30	Registration and coffee
10:30	**Stan Kotwicki**, NOAA (USA). Opening Remarks, Terms of Reference
10:45	**Hans Gerritsen**, Marine Institute (Ireland). Orientation to the Facility
11:00	Participant introductions: What do you do? What do you hope to contribute/gain from WKUSER ?
WKUSER2 – Keynote presentations (20 min talk + 10 min for questions)	
11:30	**Joel Vigneau**, IFREMER (France). How does WKUSER fit into EOSG and ICES?
12:00	**Elizabeth Chilton**, NOAA (USA). US perspective on challenges (ecosystem and climate change, fish movements, wind farms, budgets, borders, etc)
12:30	**Sven Kupschus** (remote), JRC (EU). Future of EU fisheries Surveys; which challenges exist, which will develop and what WKUSER2 needs to consider
13:00	**Stan Kotwicki**, NOAA (USA). Lessons learned from WKUSER1 and looking forward to WKUSER2.
13:30	Lunch
Challenges and priorities	
14:30	Discussion on challenges and priorities (including examples of experiences and outcomes, defining extend of the problem across ICES countries)

14:45 Discussion on specific topics to work on during workshop:

- Why do we need flexible surveys?
- General approaches to combining surveys Why combine surveys?
- Advantages of simulations and modelling
- How to capture true uncertainty in survey data products and stock assessment?
- What tools do we need to improve survey data products and advice?
- Think about which breakout session you want to work in to tackle specific issues.

15:45 **Session chairs**: short summary of each theme session (5 min each)

16:00 Coffee

Simulation tools for survey data products – Keynote presentations

16:30 **Sean Anderson** (remote)**, Paul Regular**, DFO (Canada). Simulation tools for survey data products present and future

17:00 **James Thorson**, NOAA (USA). The importance of multivariate and seasonal dynamics in simulation designs, and tools to address it

17:30 Discussion on simulation tools

17:45 **Session chairs** and **WK chairs**. Preparation for the next days: practical issues etc.

18:00 Bus to Galway city centre

SOCIAL EVENT

19:30 Dinner at the Brasserie on the Corner, 25 Eglinton Street, Galway

Wednesday September 14th Contributed talks – Auditorium

9:00 Bus from Galway city centre

TOR I. Survey design for flexibility: Review and summarise desired attributes of survey design that allow for flexibility when dealing with changes in survey effort and need to expand survey into new areas of species expansion due to changes in the ecosystem.

9:30 **Lewis Barnett**, NOAA (USA). Spatiotemporal models and optimization approaches to the design of multispecies surveys.

9:45 **Patrik Börjesson**, SLU (Sweden). Optimising survey effort in 3a

10:00 **Annica De Groote**, SLU (Sweden). Improving the sampling design of a groundfish survey: a case study from ICES division 27.3.a

10:15 **Jason Conner**, NOAA (USA). Improving design-based error estimates for systematic surveys and implications for model-based estimates

10:30 **Zack Oyafuso**, NOAA (USA). Survey design evaluation of a new multispecies bottom trawl survey in the US Chukchi Sea.

10:45 **Andy Lipsky**, NOAA (USA). Addressing the scientific challenges of offshore wind development in the US.

11:00 Coffee

11:30 ~~**Philip Politis,** NWFSC. NOAA (USA). Multispecies bottom trawl survey: overview and impacts of offshore wind development.~~

TOR II. Combining surveys, dealing with data gaps: Collate advice on methods to combine data from different sources, how to deal with data gaps and how to perform survey calibrations.

11:45 **Derek Bolser**, OSU (USA). Using spatio-temporal models to provide compositional data for acoustic surveys: facilitating autonomous vehicle sampling and inferences on non-target species in a fishery resource survey program

12:00 **Iosu Paradinas**, AZTI (Scotland). Integrating different fishery surveys through joint modelling

12:15 **Nathan Bacheler**, NOAA (USA). Combining trap and video data from the US Southeast Reef Fish Survey to index reef fish abundance

12:30 **Elisabeth Phillips** (remote), NOAA (USA). Attributes of the US-Canada Integrated Ecosystem and Acoustic Trawl Survey for Pacific hake that contribute to successful fisheries management, ecosystem monitoring, and expanded collaborations

12:45	**Madison Hall**, NOAA (USA). Cooperative rockfish research as supplemental survey effort: an early example from the Gulf of Alaska
	~~**Julia Clemons** (remote), NOAA (USA). Efforts to integrate transboundary stock assessment surveys in the California Current Ecosystem by 2025~~

Thursday September 15th – plenary then subgroup work in break-out rooms

9:00	Bus from galway city centre
9:30	Discussion, formulating specific topics for sub-working groups. Dividing into sub-groups
10:00	Work in subgroups
11:00	Coffee
11:30	Work in subgroups
13:30	Lunch
14:30	Work in subgroups
16:00	Coffee
16:30	Work in subgroups
18:00	Bus to Galway city centre

Friday September 16th - subgroup work, presentations, report writing

9:00	Bus from galway city centre
9:30	Subgroups presentations, discussion
11:00	Coffee

11:30	Work in subgroups, subgroup report drafting
13:30	Lunch
14:30	Subgroup report drafting
16:00	Coffee
18:00	Bus to Galway city centre

Saturday September 17th - synopsis & report writing

9:00	Bus from galway city centre
9:30	**Kotaro Ono.** Plenary: Synopsis: What did we learn and where do we go from here?
10:15	**Stan Kotwicki.** Challenges and priorities remaining. What's next, practical matters?
11:00	Coffee
11:30	Closing remarks
12:00	Chairs meeting and writing report
13:00	Bus to Galway city centre