FRAMEWORK FOR THE QUALITATIVE ASSESSMENT OF THE DEPENDABILITY OF CATCH DATA FROM EXISTING FISHERIES MONITORING TOOLS

Figure 1. The administrative regions of Fisheries and Oceans Canada (DFO). The dashed line indicates Canada's Exclusive Economic Zone (EEZ).

Context

Fisheries and Oceans Canada (DFO) is currently developing a national fishery monitoring policy to ensure that DFO has dependable, up-to-date, and accessible fishery information to manage fisheries sustainably. Meeting this objective is essential to the long-term sustainable management of fisheries and is an important step in increasing the public's confidence in the government's stewardship of public resources. It is proposed that during the implementation of this new policy there will be a risk assessment of individual fisheries in Canada. Part of the proposed risk assessment will be to review existing catch monitoring tools and the dependability of inferences on catch (retained and non-retained catch) that can be derived from the data collected by each tool as implemented in different fisheries. For these risk assessments to be nationally consistent, National Fisheries Policy requested a framework be developed to provide guidance on how different catch monitoring tools could be qualitatively assessed to determine the extent to which they provide data from which catches can be accurately and precisely inferred.

This Canadian Science Advisory Secretariat (CSAS) national peer review did not assess individual fisheries, but instead developed a nationally consistent framework on how to qualitatively assess common tools used across Canada to monitor marine and anadromous fisheries. This framework is an *important first step in fulfilling the DFO National Fisheries Policy request. Future meetings may be held to apply the framework to Canadian fisheries to inform fishery-specific decisions about the requirements for fishery monitoring.*

This Science Advisory Report is from the June 20-23, 2017 meeting Framework for the qualitative assessment of the dependability of catch data from existing fisheries monitoring tools. Additional publications from this meeting will be posted on the [Fisheries and Oceans Canada (DFO) Science Advisory Schedule] as they become available.

SUMMARY

- DFO is developing a policy on fishery monitoring. It is proposed that the implementation of the policy will require DFO to assess the dependability of estimates from catch monitoring programs
- Dependability is the ability of each catch monitoring tool to achieve the objectives for which it was intended, such as compliance with a threshold or estimation with a desired level of quality.
- A review of industry-reported and independent-observer catch monitoring tools was completed to identify their strengths, weaknesses, and limitations in providing dependable catch estimations. Feasibility and cost effectiveness were not considered.
- The effects of a number of statistical and operational factors that affect dependability cannot always be explicitly quantified and therefore a framework that accepts both expert opinion and data-based assessment of the quality and dependability of estimates from fishery monitoring programs has been developed.
- This framework is applied individually to a measured property of a monitoring tool (e.g. target species discards). By its design, the framework allows for the assessment of the dependability of a monitoring program using multiple tools and for the dependability of catch estimations using multiple monitoring programs.
- A practitioner's guide is being developed and will help users apply the framework in a consistent manner.

BACKGROUND

Fisheries and Oceans Canada (DFO) is currently developing a national fishery monitoring policy to ensure that DFO has dependable, up-to-date, and accessible fishery information to manage fisheries sustainably. Meeting this objective is essential to the long-term sustainable management of fisheries and is an important step in increasing the public's confidence in the government's stewardship of public resources.

It is being proposed that the new policy will require risk assessments of individual fisheries in Canada. DFO National Fisheries Policy requested a framework be developed to ensure national consistency in these risk assessments. Part of these proposed risk assessments will involve reviewing existing catch monitoring tools and the dependability of inferences on catch (retained and non-retained catch) that can be derived from the data collected by each tool as implemented in each fishery. The framework will provide guidance on how different catch monitoring tools could be qualitatively and quantitatively assessed to determine the extent to which they provide data from which catches can be accurately and precisely inferred. In applying the policy, the framework will help guide decisions about the type and level of monitoring that is required in a fishery to meet the data needs of the Department.

This Canadian Science Advisory Secretariat (CSAS) national peer review did not assess individual fisheries. Instead it developed a nationally consistent framework for the qualitative assessment of the dependability of inference on fishery catches made using data obtained from common tools used across Canada to monitor marine and diadromous fisheries. This framework is an important first step in fulfilling the DFO National Fisheries Policy request. A future meeting may be held to apply the framework to Canadian fisheries to inform fishery-specific decisions about the requirements for fishery monitoring.

DEFINITION OF TERMS

Retained catch: portion of the catch that is retained for use. This includes landed catch and catch that is used in some way but not landed, such as catch that is used for bait (DFO 2013).

Non-retained catch: consists of any species or specimens that are not retained for use and that are returned to the water. The returned catch may be alive, injured or dead. This includes catch brought on board and thrown back, catch released from gear before it is brought on board (such as catch released from a purse seine before the seine is fully pursed), and catch that becomes visibly entangled in fishing gear, such as entangled whales, birds and sea turtles. This does not include catch that escaped the fishing gear, that was removed by predators and scavengers, or that dropped out dead from the gear (DFO, 2013).

Bycatch: a) retained catch that includes species and specimens of the target species, such as specimens of a particular sex, size or condition, that the harvester is not licensed to direct for but may or must retain; and, b) all non-retained catch, including catch released from gear and entanglements, whether alive, injured or dead, and whether of the target species or the non-target species (DFO, 2013).

Rare catch: A species or specimens of a species that are caught very infrequently and typically in low numbers. This category of catch is treated separately because the estimation of catch characteristics (e.g., total catch) is challenging. The sample sizes typically required to estimate these characteristics with reasonable precision are quite large and the data often do not conform to conventional statistical distributions.

Species at Risk Act (SARA) catch: species listed on the List of Wildlife Species at Risk as set out in Schedule 1. A species at risk means an extirpated, endangered, or threatened species or a species of special concern (*Species at Risk Act*).

Target catch: retained catch that consists of the species that the harvester is licensed to direct for, in other words, the target species of the fishery. In a multispecies fishery, this includes any species that the licence holder is licensed to direct for on a given fishing trip regardless of whether the licence holder did so or not (DFO, 2013).

Independent-observer data: data that are observed, measured, obtained, and/or recorded by a third party, Fisheries and Oceans Canada, or through an electronic format.

Data reported by resource users: data that are measured, recorded, or reported by the fisher conducting the activity or the buyer of commercial harvests.

ANALYSIS

Catch monitoring tools used in Canadian fisheries

Catch monitoring provides information on catch and other details related to fishing activity. This is performed by Fisheries and Oceans Canada (DFO) staff, resource users, or designated third-party individuals. Data from catch monitoring are used primarily to support fisheries management efforts and are critical inputs to resource assessments, but they can also be used for other purposes such as enforcement activities and directed scientific research. Resource managers use the collected data to make ongoing management decisions, for within-season management actions, longer-term fisheries planning, and national and international reporting on landings. Scientists use the collected data to feed into stock assessments that provide an evaluation of the stock status and the risks to conservation of different management options, and for directed research. Fishery officers use the collected data to carry out compliance and enforcement activities. Dependable data are necessary to support these efforts and contribute to the long-term sustainable management of fisheries.

In addition to having different uses, the various catch monitoring tools also provide information that differs in content, scope, resolution, and data quality. For example, some tools only report on retained catch, with varying degrees of accuracy, while others report on both retained and non-retained catch.

Catch monitoring tools used in Canadian fisheries include: commercial sales slips; fisher pre-departure and pre-arrival notifications (hails), fisher questionnaires, creel surveys, logbooks, dockside monitoring, at-sea observers, electronic monitoring systems (with video), vessel monitoring systems, and patrolling. Each of these tools has benefits and limitations that impact the quality of the data they provide (Beauchamp et al. 2019). Brief descriptions of these tools are provided below and characteristics of each tool and the data they provide are summarized Table 1. Assessing the cost, feasibility, and usefulness for enforcement activities was not part of the dependability assessment.

*Table 1. Canadian catch monitoring tools and their ability to collect data on different categories of catch, catch composition, fishing location data, and effort data. Each tool is also designated as recording independent-observer (O) or providing resource user (R) data. Reports = annual/seasonal reports, questionnaires and surveys that are either voluntary or required; Hails = harvester pre-arrival and pre-departure notifications; ASOP = at-sea observer programs; DMP = dockside monitoring programs; EM = electronic monitoring systems with video; VMS = vessel monitoring programs. P indicates the tool is primarily used to collect this type of data, while * indicates that this tool can collect this type of data but may be prone to strong biases or mis-reporting.[1]EM and VMS tools are fully independent and do not have an observer at time of recording, providing an achievable record of activities free from human biases.*

			Catch Monitoring Tools									
			Reports	Logbooks	Hails	Sales Slips	ASOP	DMP	EM (Video)	VMS	Patrolling	Creel
		Data (R or O)	R	R	R	R	O	O	O[1]	O[1]	O	O/R
Catch Monitoring	Retained	Target	P*	P*	P*	P*	P	P	P			P
		Bycatch	*	*	*	P*	P	P	P			*
		Rare	*	*	*	P*	P	P	P			*
		SARA	*	*	*	P*	P	P	P			*
	Non-retained	Target	*	*			P		P			*
		Bycatch	*	*			P		P			*
		Rare	*	*			P		P			*
		SARA	*	*			P		P			*
		Catch composition					P	P	P			
		Location	*	*	*	*	P	*	P	P	P	*
		Effort	*	*	*		P		P	P	P	

DATA REPORTED BY RESOURCE USERS

There are several catch monitoring tools where fishers or other members of the fishing industry (e.g., buyers) may be required (or volunteer) to provide data or to answer questions on catch. Tools involving self-reporting include: fisher questionnaires, summary reports, logbooks, and creel surveys. Much, if not all, of the information collected through self-reporting is fisher-dependent data and therefore can be confounded by responder bias and intentional mis-reporting for a variety of reasons. Each of these tools is described in more detail below.

Fisher pre-departure and pre-arrival notification (Hails, Hailing out/in)

Communication between a commercial fishing vessel and a third-party monitoring company, fishery managers, or enforcement officials, to indicate the departing and/or returning to port, or reporting the activities after a day of fishing. This is a fisher-dependent monitoring tool and the information reported can be limited by incorrect or imprecise reporting. Often the purpose of the tool is to assist in planning further monitoring. Communications indicating a vessel is departing port are typically provided within a mandated pre-departure timeframe and allow observer companies to plan the deployment of at-sea observers; this can significantly increase the randomness of observer deployments. Communications indicating a vessel is returning to port are typically provided to dockside monitoring companies to allow them to plan dockside monitoring activities. In some fisheries, daily hails are used to report daily catches on multi-day (extended) trips which allows for area or individual quota management. This tool is not usually used as a sole source of reporting catch, but rather in conjunction with other catch monitoring tools. Hail systems can be used to verify compliance with mandatory logbook reporting (for example cross referencing logbook entries with pre-departure communications on intended fishing locations).

Commercial sales slips

Commercial sales slips report on the amount of fish that are sold at the first point of sale. Sale slips are documents or information produced by official fish buyers and provided to DFO. They are a relatively common tool across Canada. However, catches that are retained for personal use, private sales (e.g., restaurant), or sold to other fishers (e.g., for bait) are often not accounted for through sales slips. Therefore, the information they provide could be underestimates. Additionally, there may be deliberate misreporting via commercial sales slips to hide catch or to inflate it. Factors that motivate hiding catch include underreporting revenue for tax purposes or to allow for continued fishing in quota managed fisheries. Factors that motivate inflating catch include anticipation of the imposition of catch shares based on historical individual catches. This tool does not provide any information on non-retained catch.

Fisher questionnaires

Fisher questionnaires may be conducted by mail, telephone, or electronically (e-mail or web/application based platform). They are often used in recreational fisheries to estimate catch. They are also used in commercial fisheries either during or post-season. For recreational licence holders, questionnaires are usually randomly distributed to a subset of the licensees to solicit their responses. In commercial fisheries, active licence holders that sold fish in the target season typically form the population from which participants are selected. Advantages of questionnaires include a lack of face-to-face bias (for example, tailoring responses based on the interviewer's reaction), more time for responses, and the ability to ask more complex questions. However, disadvantages include relying on recall and non-responses, both of which can lead to increased variability.

Furthermore, when recall is biased to unusual events such as large catches, or if non-responses are intentional, then bias could be introduced. It is also not possible to clarify questions in mail or electronic questionnaires, therefore clear-sounding questions may result in divergent answers due to misunderstanding of the questions' intent. All-in-all, the quality of data provided by harvester questionnaires can be negatively affected by the lack of clarity and specificity in the questions, the elapsed time between the questionnaire and the events that are meant to be recalled, and response bias and error.

Logbooks

Logbooks provide fisher accounts of catch and other fishing related details. They can vary greatly from fishery to fishery, with different fishery-related information being collected. Logbooks are used in most Canadian commercial fisheries but have limitations and can be prone to intentional and unintentional inaccurate reporting. Fishers can be required to submit logbooks at different times, such as when they complete a fishing trip or at the end of the season. Generally, the quality of the data is considered to be greater the earlier this information is submitted, reducing the need to recall information and increasing the ability of DFO to undertake corrective actions should logbooks not be completed or be completed incorrectly. Compared to records by at-sea observers (described below), reports in logbooks often underreport catch amounts, overreport the frequency of zero catches, and report a smaller diversity of species. The quality of the data reported in logbooks further depends on having clear, accurate, and specific instructions for completing each of the fields.

Independent verification of logbooks can minimize many of the limitations and increase the quality of the data from this tool. Some examples include: communications (hails) and data from fishery enforcement overflights can be used to validate locational and effort data collected in logbooks; boarding by fisheries officers can be used to verify the accuracy of catch records in the logbooks; logbook catch records and catch records provided by third-party at-sea observers (or electronic monitoring systems with video) can be compared to evaluate logbook data quality. However, the presence of observers can produce an invigilation effect, increasing logbook compliance only when an observer is present. Therefore, concordance between observer and logbook records is not a reliable measure of the quality of logbook data.

Creel surveys

A creel survey involves conducting interviews of fishers at a landing port, water-body access point, or onsite (e.g., on river) during or after fishing. This tool is most often used in recreational fisheries. Creel surveys are often conducted over very large areas for an extended period of time. Generally, fishers (often recreational anglers) are surveyed after fishing trips and this information is used in combination with estimates of effort to estimate total catch. The data collected is a mix of resource user-reported and independently observed data.

Creel surveys reporting on non-retained catch rely on the fisher's recall and their ability to correctly identify fish species and may be subject to reporter bias (resource user-reported data); data collected on landed catch can be verified by the surveyor (independent-observer data).

Often many assumptions need to be made to estimate total catch, which decreases the quality of the data. For example, the assumption of a constant catch rate over the duration of a fishing trip may not be suitable for some fisheries, such as a gill net fishery where net saturation can result in a declining catch rate. Or, the assumption that fishing in one area is independent of that in another area. Data quality is affected if it cannot be extrapolated accurately to the fishery. While obtaining a dependable estimate of total catch from creel surveys is difficult, it is one of the few tools available to estimate catch from recreational fisheries.

INDEPENDENT OBSERVER TOOLS

Dockside monitoring

Dockside monitoring programs (DMP) are regularly used in commercial fisheries. Dockside monitoring companies, either private companies or not-for-profit corporations, and the dockside observers they employ, are designated by the Department to perform the duties related to the Dockside Monitoring Program, as indicated in the *Fishery (General) Regulations*. They are qualified according to the Canadian General Standards Board (CGSB) Program Manual which ensures all the dockside monitoring companies have quality management systems in place. This includes proper documentation and training, established procedures and reporting standards, quality control, accountability, and internal audit.

In many fisheries, dockside monitoring is the Department's primary source of verified landing information. Catches are weighed at the wharf providing a direct and typically accurate measurement. This is in contrast to catches recorded by at-sea observers which are often based on a visual estimation of catch weight (see below). However, adjustment errors may be an issue as the condition of landed fish can vary (dressed, split, whole, frozen, fresh) and adjustments are required for the catch statistics which are typically in whole-weight equivalents. As noted above, mandatory communications of fishers returning to port with catches (hails) can help in planning dockside monitoring activities and can be used to correctly stratify the data for producing estimates.

At-sea observers

At-sea observer programs (ASOP) place designated third-party observers aboard fishing vessels to monitor/verify fishing activities, collect scientific and fishing data, and monitor industry compliance with fishing regulations and licence conditions. While retained catch is quantified, observer reports are not typically used to estimate landings as DMP programs are more accurate. However, observer reports are one of the main sources of information used to quantify non-retained catch.

Similar to dockside monitoring programs, ASOP companies and the at-sea observers they employ are designated by the Department to perform the duties related to the At-Sea Observer Program, as indicated in the *Fishery (General) Regulations*, and are qualified according to the CGSB Program Manual and the At-Sea Observer Program Policy and Procedures. The CGSB qualification program aims to ensure that ASOP companies have adequate quality management systems and practices in place. This includes proper documentation and training, established procedures and reporting standards, quality control, accountability, and internal audit. Regular audits are carried out by CGSB to ensure that the companies conform to the recognized standard.

Required communications from fishers leaving port are often used to help plan trips on which to deploy an observer when targeted coverage of the fleet is below 100% (see communications above).

For in-season fishery management actions this is the only third-party tool that provides near instantaneous reporting on catch characteristics that may result in management actions. Examples of in-season management actions include limits on the incidental catch of non-target species, size limits for target catch and changes in sex ratios in some decapod crustacean fisheries.

This tool is particularly prone to the influence of observer effects. Harvesters have been shown to alter their fishing patterns to modify the amount and species composition of catches in the presence of an observer when there are incentives to do so, which there often are. It is sometimes possible to statistically test for an observer effect and to estimate its magnitude of effect on the reported retained catch; this is not possible for discarded catch. However, it may be possible to infer whether an observer effect on discards is likely. This can be achieved by testing for an effect on retained catch (amount and size composition) and by comparing the spatial location and effort of fishing activities associated with an observer and not, provided that statistical power of those tests was sufficiently high. Indeed the mandatory use of VMS systems may act as a deterrent for observer effects caused by changes in fishing patterns when an observer is present.

The uncertainty in catch and bycatch estimates changes as observer coverage increases. The challenges for these programs is to determine coverage levels that are optimal with respect to statistical precision of the estimators while balancing costs. Unfortunately, the impact of deployment (described below in unrepresentative sampling) and observer effects on the variability and bias of the estimates can be difficult to predict. Assessments that ignore deployment and observer effects are therefore likely to provide misleading results for planning observer coverage levels unless these effects are believed to be small because of the procedures that are in place (e.g., mandatory hailing) and lack of incentives (or presence of strong disincentives) to modify behaviours when an observer is present.

Electronic monitoring systems (with video)

Electronic monitoring systems using digital video-recording (EMSV) devices are used with global positioning systems (GPS) to record fishing operations and catch composition which can later be analyzed. These systems can provide independent electronic catch data and a comprehensive record of fishing activity that can be stored. This allows the data to be audited or referenced at a later date to verify accuracy or clarify discrepancies if desired.

Electronic monitoring systems collect data on both retained and non-retained catch; however, accurate and reliable catch data can be difficult or impossible to obtain in some instances. For example, data are difficult to capture in high volume fisheries where fish do not necessarily pass through restricted locations that are easy to record (e.g., conveyor belt) and in fisheries with species that are similar in form and colour. EMSV can provide data on the count and/or size (e.g., length) of fish but cannot directly provide data on the weight of catch, though this can be estimated from counts and sizes. Other issues that can affect data quality include changes to fishing behaviour to ensure catch is captured on video, image quality that affects species identification, image quality that varies as a function of sea conditions or weather, and inadequate camera coverage.

Data acquired from an EMSV system can require a large amount of time to process and review and therefore an audit approach is typically used, where a predefined subset of video data is reviewed. An advantage of EMSV is that it normally provides complete coverage of sampled trips. It is therefore possible to employ optimal sampling strategies when selecting part of the video for detailed analysis, thereby ensuring efficient and unbiased sampling of the available images. Furthermore, EMSV allows for optimization of sampling efforts to different parts of the sampling hierarchy (e.g., vessels within the fleet, trips within vessels, and fishing sets within trips), such as to maximize precision for a given subsampling effort. In contrast, such optimization is more problematic in at-sea observer surveys given constraints on the number of available observers and the fact that once an observer is deployed, they only sample at one level of this hierarchy (sets within trips).

Another benefit of EMSV is its ability to function as a verification and incentivization system. Tying the degree of image subsampling required (a cost to the industry) inversely to the accuracy of logbooks creates an incentive to improve the quality of logbook reporting (Stanley et al. 2011, 2015). For example, in the British Columbia groundfish hook-and-line and trap fishery 100% at-sea monitoring is required, however EMSV coupled with an audit system is used to defray costs and to eliminate the need for an at-sea monitor on every vessel. In this example EMSV data are used to audit harvesters' logbook data on effort, catch, and catch composition. In these audits, 10% of the fishing events on these vessels are independently monitored at random. A low level of agreement between the logbooks and videos can lead to additional audits that are directly funded by the responsible fishers.

Framework for the assessment of the quality and dependability of monitoring tools

The approach described below provides a unified framework to evaluate the statistical quality (how close a parameter estimate is to the true value) and dependability (ability of an estimation process to achieve its intended objectives) of the data collected by DFO's fishery monitoring programs. Several considerations were required in devising the framework.

The number of monitoring programs for all captured species in all fishery sectors in Canadian fisheries is very large. In many cases, detailed information required to measure quality will be very difficult to obtain or unobtainable. Consequently, the assessment framework must be scalable, allow for incomplete information and real-life limitations, while aiming for consistency in application and rigor.

The proposed assessment methodology is structured: it requires that the impact of each factor contributing to the quality of the monitoring program be assessed separately. A structured approach ensures that the assessment is exhaustive and consistent between programs. A structured approach increases the reliability of the assessment preventing overlaps between factors and eliminating factors not applicable to a specific program.

The proposed assessment methodology is semi-quantitative: it accepts available information, whether it is obtained from data or from expert opinions. The impact of some factors on quality can be quantified, possibly based on some quality control procedures or studies of other, similar, monitoring programs. For example, errors resulting from observers visually estimating catch weight may have been studied in quality control experiments in a particular fishery and the results can be applied to assessments in other fisheries. The impact of other factors may be difficult to estimate and, in some cases, must be based on expert knowledge ("expert" referring here to a person with a deep knowledge of the fishery). For example, changes in fisher fishing patterns when an at-sea observer is aboard are difficult to observe and their impact on the amount and species composition of discards must mostly be estimated from expert knowledge.

The proposed assessment methodology is applicable to the various monitoring tools, either singly or in combination.

A single monitoring program may be used to estimate several parameters. For example, an at-sea observer monitoring program may estimate the following parameters: total target species catch, total catch of individual bycatch species, total effort, and catch per unit of effort (CPUE). Conversely, some parameter estimation processes may depend on more than one monitoring program. For example, the estimation of total catch may depend on a monitoring program to estimate CPUE and another one to estimate total effort or the estimate of total catch of a species targeted by several fisheries. In this example, it involves assessing several monitoring programs simultaneously.

A single parameter and the monitoring program or monitoring programs required to estimate it is termed a parameter estimation process or, succinctly, an estimation process. Therefore, a single monitoring program may be involved in one or several estimation processes and a single estimation process may involve several monitoring programs. The quality and dependability of each estimation process must be assessed separately.

Quality

The quality of an estimation process describes how close to the true value the estimate is likely to be. The quality of the estimation will depend on its accuracy (converse: bias) and its precision (converse: variability). Several characteristics of an estimation process influence its quality. These characteristics can influence the accuracy and the precision of the process in different ways. Characteristics influencing quality are divided into two groups: statistical characteristics and operational characteristics.

Statistical characteristics are related to the data and the model used to derive an estimate. Factors that influence statistical characteristics include sample sizes and the sampling design. The statistical characteristics of an estimate are generally known, based on theory or simulation. Operational characteristics are factors that result in unplanned and uncorrected deviations from the sampling design or from accurate data collection, such as unplanned non-random sampling and observer effect. The effects of operational characteristics on catch estimates will often not be directly measurable, thereby requiring the use of expert knowledge or assumptions about the data collection.

Accuracy

Accuracy is defined here as a measure of how close an estimate is to the true (unknown) value; the converse of accuracy is bias. When an estimation process tends, on average, to under- or over-estimate the true value of the parameter, it is said to be negatively or positively biased. The bias is the average of the differences between the estimated values and the true value, if the estimation process was repeated many times. Bias has a sign: it is either positive (a tendency to over-estimate the true value) or negative (a tendency to under estimate the value). Bias may not decrease as the sample size increases. For example, a bias due to underreporting of discards in logbooks will remain the same for any sampling proportion, including for a census.

Impact of statistical characteristics on accuracy

Some estimators, under specific sampling protocols, are biased. This is a statistical characteristic of the sampling protocol and the estimator. There are statistical methods to estimate this bias and, in some cases, to correct the estimate for the bias. In most cases the estimator’s bias should be zero.

Impact of operational characteristics on accuracy

Differences between the actual sampling process and that presumed in the statistical computation and departures from the planned protocol can cause biases in the estimation process. Such operational characteristics can affect the accuracy of both sample surveys and censuses, and has been documented in the general literature on surveys and in fishery-specific literature. Observer effects, misreporting in logbooks, and incorrect adjustments applied to catches to produce round-weight equivalents are examples of operational characteristics that can affect the accuracy of a catch estimate. For many operational characteristics it may be difficult to demonstrate and to quantify bias in a specific survey. Information from other fisheries, expert opinion, and simulations can be used to inform on the magnitude and direction of bias resulting from an operational characteristic. The contributions to bias of different operational characteristics is expected to be additive. Furthermore, in this framework the contributions are assumed to be independent such that the bias resulting from one operational characteristic (e.g., underestimation of catch due to an observer effect) is independent of biases resulting from other characteristics (e.g., catch-weight measurement bias).

Precision

Precision is a measure of the degree to which repeated estimations are close in value; the converse of precision is variability. Precision does not have a sign (it is a positive number). If the same estimation process is repeated many times, the estimates may differ from each other due to the randomness of the sampling process or to some other characteristic of the estimation process. The term variability is used to describe this variation; it is a measure of how much estimates from an estimation process vary, on average. It describes how much the estimates would differ from each other if one repeated the same estimation process many times.

Variability typically decreases as sample size increases. In a true census, there is no sampling randomness but variability may still exist due to certain operational characteristics such as measurement error, transcription error, variations in memory lapses (in human surveys), etc.

Impact of statistical characteristics on precision

In random sampling, the variability of an estimate due to the randomness of the sampling is referred to as “sampling error”. This is a statistical characteristic of the sampling protocol and the estimator. The sampling error is often described by the standard error of the estimator or by the confidence interval of the estimate for a specified confidence level (for example 95% confidence intervals).

Methods to compute the standard error (or other measures of sampling error) depend on the sampling protocol presumed to have been used and the estimator. For example, computations required for simple random sampling and for stratified sampling are different.

The sampling error decreases as sample size increases and is 0 for a census with 100% response. In sample surveys, the standard error of the estimator is the basis of the assessment of the impact of statistical characteristics on variability/precision.

Impact of operational characteristics on precision

The standard error obtained from statistical analysis reflects the variability of the estimator due to the randomness of the sample selection process.

Some operational characteristics, such as measurement error due to imprecision in the measurement tool, will create variability in a census and will increase the calculated standard error in a sample survey (i.e., their effect will be reflected in the calculated standard error). The contributions of these errors are expected to be independent and additive. While the impact of these characteristics on precision is expected to be small and negligible in many cases, in other cases, they will present an opportunity to improve the precision of the estimate by improving the measurement tool or data collection process.

Other operational characteristics will cause, in sample surveys, unplanned departures from the sampling design. These characteristics will often make the estimated standard error appear smaller or larger than it really is. For example, unplanned cluster sampling, where sampling units are selected in groups instead of individually, and unplanned stratified sampling, where sampling units will be drawn separately from an exhaustive partition of the population, will lead, respectively, to underestimation and to overestimation of the standard error; similarly, targeted sampling may lead to either impact. Methods based on sampling theory can be used to estimate the required corrections which will usually be expressed as a multiplicative factor (e.g. the true standard error is 1.2 times that which is calculated).

We expect that the impact of these characteristics on precision will be important in many cases.

Quality's Statistical and Operational Characteristics

The combination of accuracy and precision determines the quality of an estimation process. Figure 2 illustrates the four possible combinations of accuracy and precision, the components of quality.

In statistics, quality is often summarized by the root mean-squared error (RMSE):

$$RMSE = \sqrt{standard\_error^2 + bias^2}$$

In this framework, this concept is extended heuristically by including the effect of both statistical and operational characteristics on the variability and the bias.

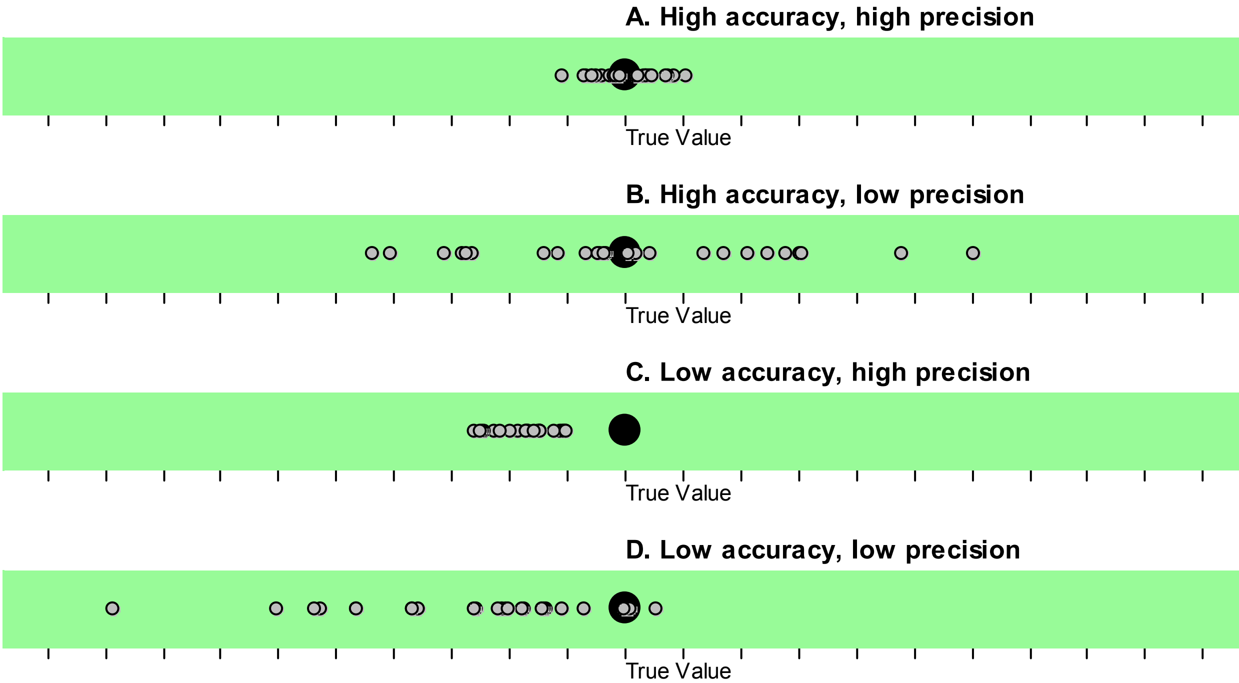

Figure 2. Illustration of accuracy (converse: bias) and precision (converse: variability), the two components of quality. The solid dot represents the true value of the parameter and the open dots represent examples of estimated values from theoretical repetitions of the estimation process. (A) Estimates are centered on the true value (high accuracy) and close to each other (high precision). (B) Estimates are centered on the true value but far from each other (low precision). (C) Estimates are generally lower (negative bias) than the true value but close to each other (high precision). (D) Estimates are generally lower than the true value and far from each other (low precision).

Assessment of the quality of an estimation process

The assessment of the quality of an estimation process involves the statistical bias and standard error from the statistical characteristics, resulting from the data and estimation method, as well as the contributions of operational characteristics to bias and variability.

In this framework, the bias of an estimation process (b_{ep}) is measured by adding the statistical bias of the estimator to the sum of any biases resulting from the 15 operational characteristics described below (also see Table 2).

In sample surveys, the variability of an estimation process, s_{ep}, will usually be measured by multiplying the standard error of the estimator by the factors describing the contribution to variability of the operational characteristics due to unplanned departures from sampling design (Table 2; Allard and Benoît 2019).

In censuses, the variability of an estimation process is calculated from the random errors due to operational characteristics (Table 2; Allard and Benoît 2019).

The quality is summarized by the estimation process error (e_{ep}), which is calculated by combining the bias, b_{ep}, and the variability, s_{ep}, into a RMSE equation ($e_{ep} = \sqrt{s_{ep}^2 + b_{ep}^2}$).

It is also possible and desirable under the forthcoming DFO fishery monitoring policy to be able to compare the quality of different monitoring programs, or of a single program over time. This can be done by comparing the b_{ep}, s_{ep} and e_{ep} to an anticipated true value of the parameter.

The anticipated true value should be taken as being equal to an estimate of the parameter obtained from the monitoring tool(s) minus the anticipated bias of the estimation process. In a stable context the estimate of the parameter could be a median of the values obtained in the last 5 or 10 years. If there is a temporal trend, the median of the most recent values or a projected value would be more appropriate. In some cases (e.g. a new monitoring program), the estimate of the parameter may be inferred from other programs and/or expert opinions.

Operational characteristics impacting quality

This framework identifies 15 operational characteristics that can impact accuracy and/or precision (Table 2). The quality assessment of the estimation process is based on first separately assessing the contribution of each of these 15 operational characteristics to the bias and variability of the estimation process. The assessment of each contribution can be based on research that provides a single value or on expert opinion that provides a range of values.

Some operational characteristics occur only in sample surveys (e.g. characteristics related to the sampling design). Some characteristics only or mostly impact bias or variability, not both. Finally, some characteristics impact the computed standard error; their influence is already reflected in the standard error computed using the original data and standard statistical methods. Unbiased measurement error is an example. In most cases, there is no requirement to explicitly adjust the assessed quality for such characteristics, however they do present an opportunity to reduce the standard error and, therefore, the error of the estimation process.

A distinction is made for some operational characteristics between values collected by DFO or independent (third party) observers and values reported by resource users.

Table 2. Operational characteristics contributing to the quality of an estimation process with their applicability to censuses (all apply to sampling surveys) and their potential for impact on the process bias, variability, and standard error (SE). "Observer" refers to an independent person or a technology (electronic monitoring with video) specifically tasked with observing and recording fishery activities; at-sea and at-dock observers are the main examples. "Resource user" refers to fishers, plant personnel, buyers, etc. that report information to DFO; logbooks and sales slips are examples.

	Operational characteristics	Applicable to censuses	May contribute to		
			process bias	process variability	SE
Q01	Undercoverage: The observed or sampled population (sampling frame) excludes a subset of the target population.	Yes	X	–	–
Q02	Overcoverage: The observed or sampled population (sampling frame) includes units outside the target population.	Yes	X	–	–
Q03	Unintended sample clustering: Cluster sampling is occurring unintentionally. If clusters are more homogeneous than the population, the computed standard error will be smaller than the true standard error and variability will be underestimated.	No	–	X	–
Q04	Unintended sampling stratification: Stratified sampling is occurring unintentionally. If strata are more homogeneous than the population, the computed standard error will be larger than the true standard error and variability will be overestimated. Bias may result if strata are more homogeneous than the population and the sample allocation is not proportional to stratum size	No	X	X	–
Q05	Other irregular selection probabilities or exclusions: including targeted sampling or deliberate avoidance of certain sampling units.	No	X	X	–
Q06	Observer effect: Occurs when the presence or expected presence of an independent observer (human or technological) causes a change in the fishing activity.	Yes	X	–	X
Q07	Missing values due to unintentional factors, including unintended non-response: an	Yes	X	–	X

	Operational characteristics	Applicable to censuses	May contribute to		
			process bias	process variability	SE
	observation was not obtained due to uncontrolled circumstances.				
Q08	Missing values due to intentional factors, including intentional non-response: any information not obtained due to a deliberate action.	Yes	X	–	–
Q09	Errors in data from resource users: recurring errors related to the implementation of the program including, e.g., unintentional errors due lack of training, carelessness, etc. and intentional errors aiming to mislead fishery managers. Intentional errors are likely to introduce bias	Yes	X	X	X
Q10	Error in data from independent observers: recurring errors related to the implementation of the program. Will generally be unintended (and unbiased) unless there is collusion with fishers or harassment.	Yes	– / X	X	X
Q11	Equipment error, including measurement error: error due to measuring tool inaccuracy and/or imprecision.	Yes	X	–	X
Q12	Data handling errors: errors introduce during the data manipulations.	Yes	X	–	X
Q13	Adjustment error: error in adjustments to the data that are required to obtain an estimate (e.g., conversion to round-weight equivalents).	Yes	X	– / X	– / X
Q14	Imputation error: error occurring when a missing value is replaced by a value obtained (imputed) from other information available	Yes	X	X	–
Q15	Modelling error: error due to using an inappropriate statistical model when computing the bias and the standard error of the estimator.	No	X	X	–

Table 3. The potential of Canadian catch monitoring tools to be impacted by the 15 operational characteristics that can influence parameter estimation quality and dependability. Reports = annual/seasonal reports that are either voluntary or required; Hails = harvester pre-departure or pre-arrival notifications; ASOP = at-sea observer programs; DMP = dockside monitoring programs; EM = electronic monitoring (with video); VMS = vessel monitoring programs.

	Operational characteristic	Catch Monitoring Tools									
		Fish slips	Hails	Self reporting	Logbooks	ASOP	DMP	EM (Videos)	VMS	Patrolling	CREEL
Q01	Undercoverage	X	X	X	X	X	X	X	X	X	X
Q02	Overcoverage	X	X	X	X	X	X	X	X	X	X
Q03	Unintended sample clustering	X	X	X	X	X	X	X	X	X	X
Q04	Unintended sampling stratification	X	X	X	X	X	X	X	X	X	X
Q05	Other irregular sample selection	X	X	X	X	X	X	X	X	X	X
Q06	Observer effect	–	–	–	–	X	–	X	–	–	–
Q07	Missing values	X	X	X	X	X	X	–	–	X	X
Q08	Missing values	X	X	X	X	–	–	–	–	–	X
Q09	Errors in data from resource users	X	X	X	X	–	–	–	–	–	X
Q10	Errors in data from independent observers	–	–	–	–	X	X	X	–	X	X
Q11	Equipment error	X	–	–	X	X	X	X	–	–	X
Q12	Data handling error	X	–	X	X	X	X	–	–	X	X
Q13	Adjustment error	X	–	–	–	X	X	X	–	–	X
Q14	Imputation error	X	–	X	X	X	X	X	–	–	X
Q15	Modeling error	X	–	X	X	X	X	X	–	–	X

Dependability

In assessing an estimation process, we use the term dependability to describe the ability of the estimation process to help reach the objectives for which it is to be used.

The objectives of fishery monitoring programs fall into two general classes: measurement and compliance. Measurements are important for administrative purposes (e.g., reporting the total economic value of a given fishery) or scientific purposes (e.g., stock assessment). The objective is to have an estimate that is of sufficient quality to be dependable. Compliance is important when some limit has been set (e.g., total allowable catch, total allowable bycatch as a function of the target species catch) and the estimate is used to determine if the limit has been respected or not.

There are other, non-statistical, objectives of fishery monitoring. These include deterrence (e.g., by placing at-sea observers on vessels that are likely to violate regulations or conditions of licence) and regulatory enforcement (e.g., at-sea vessel boarding by fisheries officers). These non-statistical objectives are not within the scope of the present framework.

Dependability for measurement applications

The assessment of dependability for measurement objectives is based on comparing the quality of the estimate with the quality required for the scientific or administrative objectives.

Dependability for compliance applications

The assessment of dependability for compliance objectives must depend on the quality of the estimation process and on how far the typical true value is from the compliance limit. A property of this approach is that fisheries for which the catch is typically close to the total allowable catch will require better quality estimates (e.g. based on larger sample sizes) than fisheries for which the catch is far from this limit.

The underlying principle applies to all cases, though the mathematical details may vary for example between cases for common events, covering most DFO programs and cases for rare events.

Figure 3 illustrates several combinations of accuracy, precision and relationships between the true value and the compliance limit (e.g. total allowable catch). If the true value is far from the limit, a low accuracy, low precision (Figure 3A) estimation process is dependable. If the true value is close to the limit, a low precision (Figure 3B, Figure 3C) or low accuracy (Figure 3D) estimation process is not dependable, but a high accuracy, high precision (Figure 3E) estimation process is dependable. If the true value is far from the limit, a high quality (Figure 3F) estimation process may not be needed, depending on other departmental needs and fishery logistics.

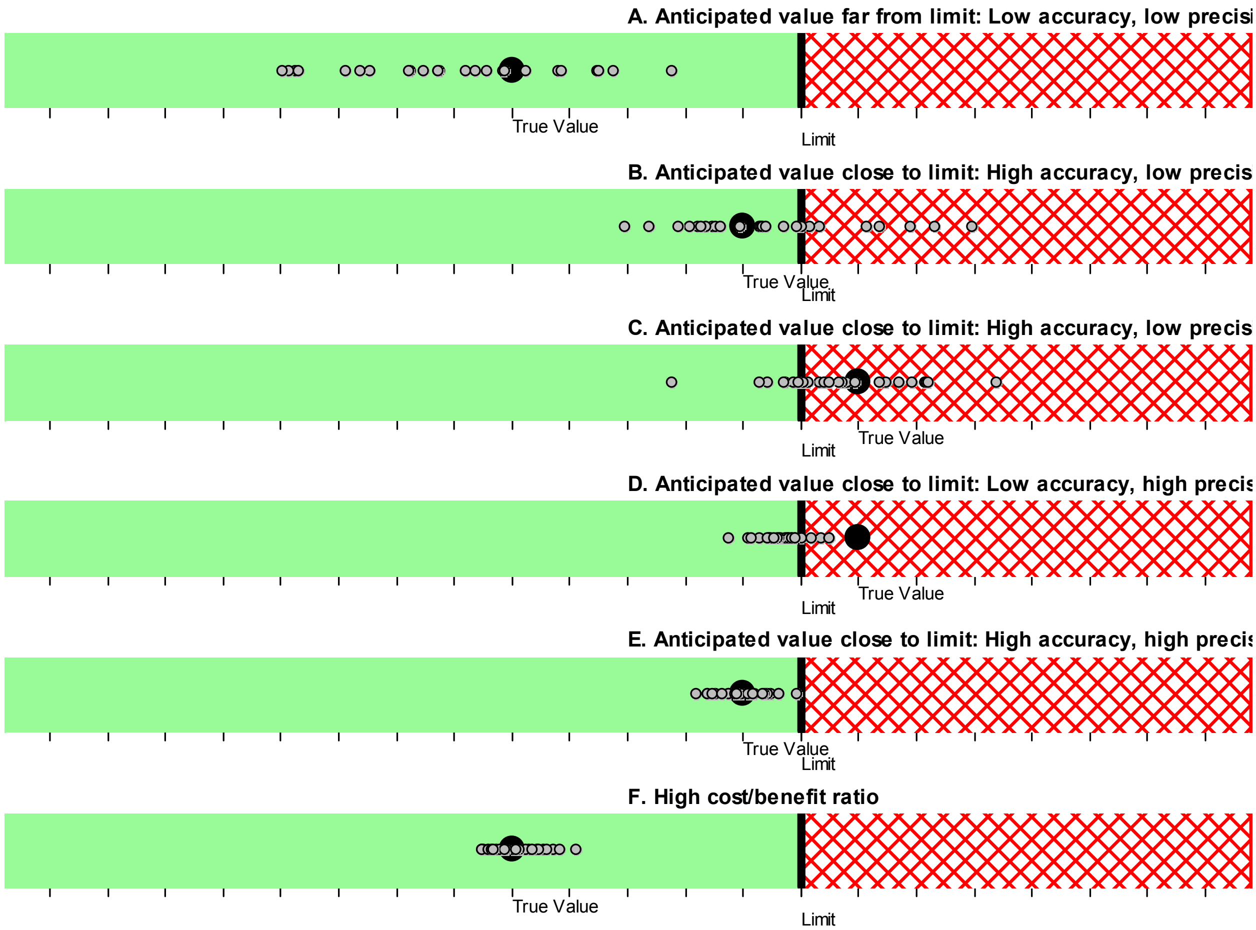

Figure 3. The dependability of an estimation process depends on its quality (accuracy and precision) and on how close the true value of the parameter (solid dot) is to the limit (vertical line). The small gray dots represent examples of estimated values from theoretical repetitions of the estimation process. The graphic illustrates an upper limit, e.g. a fishery's Total Allowable Catch (TAC).

Assessment of the dependability of an estimation process

Common events are taken here to mean events that will occur on many or all sampling occasions. This includes weights or counts of target catch, common bycatch, etc., where the parameter of interest is the total weight or the total number of units. The central limit theorem, from statistical probability theory, can be assumed to apply to the estimation process in these cases. Statistical methods to assess dependability can therefore be used in a heuristic manner (i.e., not entirely correct, but sufficient for this purpose). These will be the most frequent applications.

Rare events are defined as events that occur only infrequently and that typically involve small counts or amounts. In these cases, we cannot presume that the central limit theorem applies. Other distributions would need to be considered for approaches to assessing dependability in rare-events (Allard and Benoît 2019).

Measurement applications

For common events, where the estimation processes is heuristic and can be based on the central limit theorem, dependability is assessed using the quotient of the maximum error acceptable for the scientific or administrative objectives, as measured by the RMSE over the estimation process error. A value of 1 (100%) or above indicates a level of quality that is sufficient or better.

For other cases, the central limit theorem may not apply. For example, the sampling distribution of the estimate of the parameter obtained from the monitoring tool may be highly asymmetrical and, therefore, the RMSE is not a suitable description of the error. In cases where the Central Limit theorem does not apply other approaches can be used to determine the dependability relative to the scientific or administrative objectives (Allard and Benoît 2019).

Compliance applications

The assessment of dependability for compliance objectives is based on a hypothesis testing heuristic for statistical power: the probability the estimation process leads to a correct conclusion that the limit has been respected or not respected. This approach allows for the establishment of a uniform measure of dependability and for a risk-based assessment (details in Allard and Benoît 2019). Approaches for common and rare events would be different (Allard and Benoît 2019).

Sources of Uncertainty

This framework has developed a method for assessing the dependability of data collected in fishery monitoring tools. Some of the input data is uncertain and the tool has been developed to incorporate qualitative evaluations of these estimation processes when quantitative data is not available. There is inherent uncertainty when using qualitative data; often this is based of expert knowledge. Results based off these qualitative assessments will have more uncertainty than results based off properly sampled, quality, quantitative data.

CONCLUSIONS AND ADVICE

The framework was developed based on statistical principles and an understanding of the operational factors affecting catch monitoring data. It was designed to provide a thorough, reproducible and consistent method of evaluating the reliability of catch monitoring programs and the quality of the data they produce, even in the absence of fishery-specific data. Experience gained in applying the framework to evaluate fishery monitoring programs may identify required additions and approaches for streamlining the process. A review of this framework after a few years of application is therefore recommended.

Documenting the basis for the scores or values used when completing the framework will be key to ensuring reproducibility and the defensibility of decisions made as a result of an assessment. Documentation will, in time, also contribute to enhancing the consistency of application of the framework to other monitoring programs and fisheries, serving as a reference base from which future assessments can draw information for completing the framework.

The assessment framework was designed to accommodate inputs from both quantitative measurements and expert opinion. While the use of quantitative inputs is highly desirable, in many instances the use of expert opinion will be unavoidable because the data will not be available for a particular case or the calculation for a given operational characteristic will simply not be possible (e.g., the quantification of an observer effect on bycatch quantities). The use of expert opinion runs the risk of biasing the assessment depending on how practitioners qualify cases where there is little or no information on which to assess the impact of a given operational characteristic. An overly cautionary approach may lead to an unduly pessimistic assessment, while a neutral response may fail to flag potential problem areas. Some guidance is provided in the research documents that were prepared for the review. Furthermore, the establishment of consistent review teams tasked with undertaking the assessments should help ensure that expert opinion is used in a consistent and reasonable manner. The experience gained by these review teams will also likely flag areas requiring targeted research aimed at understanding the consequences of operational factors affecting monitoring programs.

Completing an assessment in an effective manner will require bringing together information sources that may not be considered jointly on a regular basis. For example, information from DFO Conservation and Protection surveillance flights can, within the confines of information privacy rules, provide information on operational characteristics such as observer effects and undercoverage when combined with other sources of information (e.g., at-sea observer data, vessel monitoring, hails). Similarly, information from DFO licensing can aid in defining the statistical populations for assessment and potential structure in these populations (e.g., home ports and vessel classes) that can constitute clusters or strata in fishery monitoring. When combined with other monitoring, this can help inform the effects of operational characteristics related to coverage and unintentional structuring in sampling. It will therefore be critical that assessment teams be cross-sectorial and multi-disciplinary to ensure high quality evaluations.

OTHER CONSIDERATIONS

This advisory report presents an overview of the framework for assessing the dependability of data collected through catch monitoring programs. Additional justifications for the choice of framework as well as proposed approaches and methods for estimating the contributions of operational characteristics to precision and accuracy are provided in Allard and Benoît (2019). Furthermore, a practitioner’s guide is being developed and will help users apply the framework in a consistent manner.

LIST OF MEETING PARTICIPANTS

Last name	First name	Affiliation
Allard	Jacques	University of Moncton, New Brunswick
Ball	David	Fisheries Management – NL region
Barton	Leslie	Science – Pacific region
Beauchamp	Brittany	Science – National Capital region
Benoît	Hugues	Science – Gulf region
Bourdages	Hugo	Science – Quebec region
Bussell	Ann	Conservation and Protection – Pacific

Last name	First name	Affiliation
Clemens	Marc	National Fishery Policy – National Capital Region
Comeau	Peter	Science – Maritimes Region
Croft	Greg	Fisheries Management – Maritimes Region
Davis	Ben	Science – Newfoundland and Labrador Region
Duprey	Nicholas	Science – National Capital region
Faunce	Craig	National Oceanic and Atmospheric Administration, USA
Gauthier	Johanne	Science – Quebec region
Hedges	Kevin	Science – Central and Arctic region
Houtman	Rob	Science – Pacific region
Kettenis	Niels	National Fishery Policy – National Capital region
Lanteigne	Marc	Chair
Lester	Brian	Fisheries Management – National Capital region
Osborne	Derek	Science – Newfoundland and Labrador region
Stone	Heath	Science – Maritimes region
Tadey	Rob	Fisheries Management – Pacific region
Treble	Margaret	Science – Central and Arctic region

SOURCES OF INFORMATION

This Science Advisory Report is from the June 20-23, 2017 meeting Framework for the qualitative assessment of the dependability of catch data from existing fisheries monitoring tools. Additional publications from this meeting will be posted on the Fisheries and Oceans Canada (DFO) Science Advisory Schedule as they become available.

Allard, J, and Benoît, H.P. 2019. Unified framework for the statistical assessment of fishery monitoring programs. DFO Can. Sci. Advis. Sec. Res. Doc. 2019/013. vi + 63 p

Beauchamp, B., Benoît, H., and Duprey, N. 2019. Review of catch monitoring tools used in Canadian fisheries. DFO Can. Sci. Advis. Sec. Res. Doc. 2019/010. vi + 49 p.

DFO. 2013. Guidance on Implementation of the Policy on Managing Bycatch.

Stanley, R.D. McElderry, H., Mawani, T., and Koolman, J. 2011. The advantages of an audit over a census approach to the review of video imagery in fishery monitoring. ICES J. Mar. Sci. **68**:1621-1627.

Stanley, R.D., Karim, T., Koolman, J., and McElderry, H. 2015. Design and implementation of electronic monitoring in the British Columbia groundfish hook and line fishery: a retrospective view of the ingredients of success. ICES J. Mar. Sci. **72**: 1230-1236.

THIS REPORT IS AVAILABLE FROM THE:

Canadian Science Advisory Secretariat (CSAS)
National Capital Region
Fisheries and Oceans Canada
200 Kent St. Ottawa, Ontario K1A 0E6

Telephone: 613-990-0293
E-Mail: csas-sccs@dfo-mpo.gc.ca
Internet address: www.dfo-mpo.gc.ca/csas-sccs/

ISSN 1919-5087

Correct Citation for this Publication:

DFO. 2019. Framework for the Qualitative Assessment of the Dependability of Catch Data from Existing Fisheries Monitoring Tools. DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 2019/004.

Aussi disponible en français :

MPO. 2019. Cadre d'évaluation qualitative de la fiabilité des données sur les prises dérivées des outils actuels de surveillance des pêches. Secr. can. de consult. sci. du MPO, Avis sci. 2019/004.