Fisheries and Oceans Canada / Pêches et Océans Canada
Ecosystems and Oceans Science / Sciences des écosystèmes et des océans

Canadian Science Advisory Secretariat (CSAS)

Research Document 2019/013

National Capital Region

Unified framework for the statistical assessment of fishery monitoring programs

Jacques Allard[1] and Hugues P. Benoît[2]

[1] Atlantic Statistical Analysis Inc.
Dieppe, N.B. E1A 4Z7

[2] Fisheries and Oceans Canada
Maurice Lamontagne Institute
Mont Joli, QC G5H 3Z4

March 2019

Foreword

This series documents the scientific basis for the evaluation of aquatic resources and ecosystems in Canada. As such, it addresses the issues of the day in the time frames required and the documents it contains are not intended as definitive statements on the subjects addressed but rather as progress reports on ongoing investigations.

Published by:

Fisheries and Oceans Canada
Canadian Science Advisory Secretariat
200 Kent Street
Ottawa ON K1A 0E6

http://www.dfo-mpo.gc.ca/csas-sccs/
csas-sccs@dfo-mpo.gc.ca

ISSN 1919-5044

Correct citation for this publication:

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.

Aussi disponible en français :

Allard, J, et Benoît, H.P. 2019. Cadre unifié pour l'évaluation statistique des programmes de surveillance des pêches. Secr. can. de consult. sci. du MPO. Doc. de rech. 2019/013. vi + 67 p.

TABLE OF CONTENTS

ABSTRACT

Many monitoring programs are employed across the numerous and diverse fisheries in Canada to estimate parameters that are used directly in fishery management or are used to produce scientific advice for the sustainable management of fisheries. These parameters include catch amounts or catch rates for targeted and incidentally captured species, catch composition (e.g., size composition, percentage of soft-shelled crustaceans) and fishing effort. Fisheries and Oceans Canada (DFO) is currently finalizing a national fishery monitoring policy to ensure that it has dependable, timely and accessible information on fisheries to manage them sustainably and to ensure a unified approach for setting the type and degree of monitoring employed across fisheries nationwide. This report presents an important component for the implementation of the new policy: a unified framework to evaluate the quality (how close a parameter estimate is to the true value) and dependability (ability of an estimation process to achieve its intended objectives) of DFO's fishery monitoring programs. Quality and dependability are affected by statistical characteristics of the monitoring program design and by operational characteristics which result in deliberate or unintentional differences between the implementation of the program and the program design, or which contribute data errors (e.g., measurement, data handling and modelling errors). The proposed assessment framework accounts for statistical and operational characteristics of a program and can be implemented using quantitative data and expert opinion. It is structured and founded on statistical concepts which are used heuristically to ensure consistent application and rigor. Furthermore, the framework is applicable to parameter estimations that are used directly (e.g., catch amounts for stock assessment purposes) or that are used to gauge compliance with respect to some limit (e.g., fulfillment of a quota or achievement of a bycatch limit). The goal for this report is to provide the details of and relevant justifications for the proposed assessment framework. Technical details are provided only where they are required; however, this report is not meant to be a user's guide and a separate document will therefore be required for the implementation of the proposed assessment framework.

1. INTRODUCTION

Fisheries and Oceans Canada (DFO) is currently drafting a national fishery monitoring policy to ensure that it has dependable, timely and accessible information for fisheries to manage them sustainably and to ensure a unified approach for setting the type and degree of monitoring employed across fisheries nationwide. As part of this project, a unified framework and a corresponding tool to evaluate the statistical quality and dependability of DFO's fishery monitoring programs are required. The purpose of the present report is to propose such a framework, from which an assessment tool will be concurrently developed and documented separately.

DFO's fishery catch monitoring programs include data reported by resource users such as fisher questionnaires, purchase slips and logbooks, and data reported by independent monitors such as dockside monitoring, at-sea observers and video monitoring systems. Beauchamp et al. (2019) provide a review of these methods, including their strengths and weaknesses in providing dependable catch data.

The number of catch monitoring programs for all captured species in all Canadian fisheries managed by DFO is very large. In many cases, detailed information will be very difficult to obtain, only qualitative in nature 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 was developed based on these considerations.

The proposed assessment methodology is structured: it requires that the impact of each factor contributing to the statistical quality of the monitoring program be assessed separately. The structured approach increases the ease and reliability of the assessment by, for example, allowing factors that do not impact a particular monitoring program to be ignored. It also facilitates consistency by ensuring that each evaluation of statistical quality considers all relevant identified factors.

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 assess and, in some cases, must be based on expert knowledge ("expert" referring here to a person with detailed knowledge about how the fishery operates and is monitored, including possible cheating by resource users). 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 would usually have to be assessed from expert knowledge.

The proposed assessment methodology is applicable to the various monitoring tools, either singly or in combination. For example, the estimate of total catch of a species targeted by several fisheries (fixed gears, mobile gears, for example), involves assessing several monitoring programs separately and then considering their combined statistical quality. Similarly, an evaluation of the statistical quality of catch estimators derived from separate monitoring of effort and catch-per-unit-effort (CPUE), requires an evaluation of each monitoring programs and their joint effect.

Finally, the proposed methodology requires thorough documentation for each relevant element of the assessment.

The methodology is inspired by statistical concepts which are used heuristically to guide the development of the assessment framework. Complete understanding of the underlying statistical concepts is not required to understand and apply the framework and technical details are presented in this report only where they are required. Furthermore, a case study is provided to demonstrate the use of this framework. The goal for this report is to provide the details of and relevant justifications for the proposed assessment framework. This report is not meant to be a user's guide and is likely to be too detailed for some users. A separate user's guide to facilitate the use of the framework and to further ensure consistency in its application is being produced.

2. PRESENTATION

The following text styles are used throughout the text:

- *Term* that is defined in the text and/or in the glossary (upon first use).
- An *EXAMPLE* follows.
- An *impact* on quality is described.
- ***This point is important.***
- Parameters that should be set before deployment of the framework (example: scoring scales).

Warnings, limitations, etc.

Furthermore, instances in which a measurement scale is described are explicitly labelled as such (e.g., 'Measurement: …').

3. FISHERY MONITORING PROGRAM AND PARAMETER ESTIMATION

The objective of a fishery monitoring program is to estimate one or several *parameters* for that fishery, including, for example, the total target species catch, total bycatch for each incidentally captured species, total effort and average catch per unit of effort (CPUE), and the composition of the catch with respect to sex, age, size or condition.

A monitoring program is generally oriented towards a single fishery, typically defined by target species, location, time period, gear and sector (commercial, recreational or food, social and ceremonial). In many cases, a fishery will target one or a few species but also capture other species or components of the target species (e.g., sizes, condition) incidentally, as 'bycatch'. A single monitoring program can involve several monitoring tools (e.g., at-sea observers on a random sample of trips, mandatory logbooks and dockside monitoring of all trips).

Similarly, the estimation of some parameters depends on several monitoring programs. For example, some species are harvested by several fisheries and the estimation of total catch may involve monitoring programs specific to each fishery. Also, the fishery parameter may be estimated from the results of two or more monitoring tools. For example, estimating the total recreational fishery catch of Atlantic salmon could require a telephone census of licenced fishers to estimate the total fishing effort and a creel sampling survey to estimate the CPUE. Understanding the quality of the estimator of the parameter (total salmon catch) requires evaluating the quality of the estimators of the components, total fishing effort and CPUE, and their joint effect on quality.

There are two principal uses for parameter estimations. First, as part of fishery management, the parameter estimates may be used for the application of established quotas, limits on the bycatch, etc. That is, to determine whether and when a limit is reached. Second, for stock management and for the protection of species of conservation concern (SCC), the parameter estimates may be used to estimate stock composition and size, which in turn can be used to determine appropriate quotas and other management measures.

Figure 1 illustrates the complexity of assessing DFO's monitoring programs. Species 1 may be the target species of the Fishery A. Species 2 may be a bycatch in Fishery A but a target species in Fishery B. Therefore, the results of several monitoring programs must be combined for stock management or for the protection of SCC. Consequently, an assessment of the quality of the data used for management purposes must integrate quality across monitoring programs.

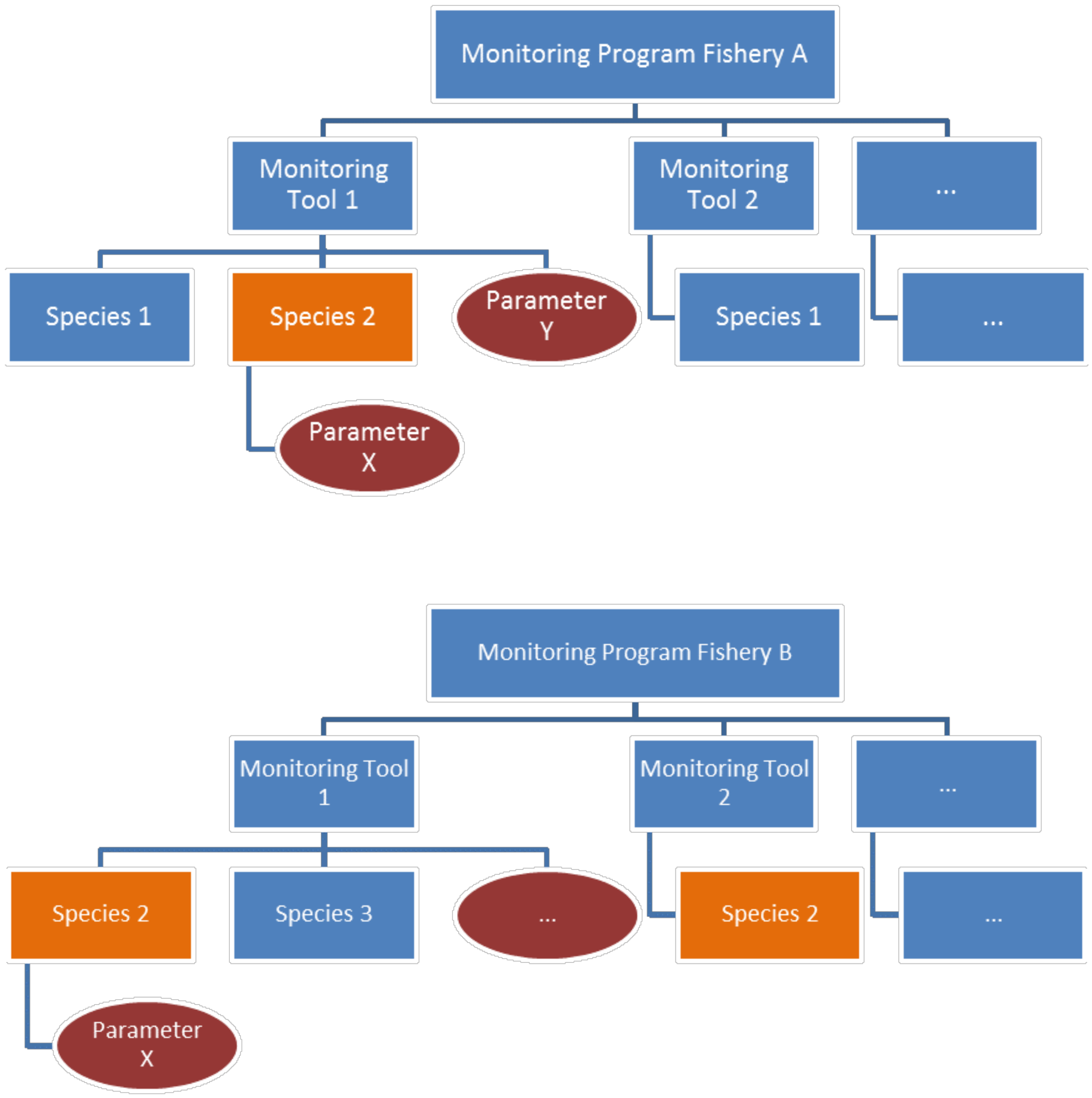


Figure 1. A single monitoring program can involve several species and therefore several parameters of interest (species 1 and 2 and parameters of interest X and Y). A species (species 2, in the figure) may be captured in several fisheries and observed in several monitoring programs. Parameter X could be total catch of Species 2 in each fishery and Parameter Y could be the ratio of bycatch to total catch in Fishery A.

The smallest unit of assessment will be a single parameter measured or estimated in a specific monitoring program from a single fishery (e.g., total catch of a given species estimated from data obtained by at-sea observers), which we call a *parameter estimation process* or, briefly, an *estimation process*. The estimation process includes all steps required to produce the estimate of the parameter, including the acquisition of the data, transcription, correction, and mathematical computations.

Each monitoring program must be assessed separately for each parameter estimated in that monitoring program. A monitoring program can be found dependable for a certain parameter (e.g. total target species catch) but not for another parameter (e.g. total bycatch and discard estimates of a specific SCC).

To understand the need to assess estimation processes separately, consider the following examples. An at-sea observer monitoring program is used to estimate the catch of a target species and the catch of a rare, incidentally captured SCC (bycatch). The observers may very easily identify fish from the target species, producing high quality data for the target species, but may have difficulty correctly and consistently identifying rare species, producing low quality data for those species. Similarly, consider a catch monitoring program with mandatory logbooks and dockside monitoring, used to measure total catch of target species and discards. The logbook entries for retained catch of the target species are likely to be accurate, given that they may also be verified at dock-side by an independent observer, while the entries for catches of discarded species may be much less reliable, given that they cannot be verified. Furthermore, logbook entries in such a case may be particularly biased if there are incentives for the fisher to under-report bycatch, such as bycatch quotas that result in fishery closures once they are reached.

Each monitoring program will consist of one or several estimation processes. Overall assessment of the monitoring program will be constructed from these elemental assessments.

For the application of fishery regulations in a single fishery, the assessment of a single monitoring tool for one or more parameters may be sufficient, i.e. the assessment of one or more (related) estimation processes. For parameters that require the results of several monitoring programs, some modifications to the assessment methodology are required. For example, when a species is caught by several fisheries, a joint assessment of several estimation processes must be carried out to assess the dependability of these programs with respect to the management of this species. This will be accomplished by combining the outcomes of the individual estimation processes.

4. APPROACH

A monitoring program targets a specific fishery. The target population consists of all units of interest. Relevant examples of target populations include the set of all fishing trips in a fishery, the set of all recreational fishers during a season, and the set of all possible daily effort counts in a fishery. A monitoring program can be based on a sample survey, where a sample of units is selected for observation, or on a census, where all units are to be observed. The term survey refers to either a sample survey or a census.

Note: The available list of units of interest is called the frame. It may differ from the list of all units of interest. The frame will be further defined below.

4.1. Quality

In assessing an estimation process, we use the term *quality* to describe the validity of the estimate of the parameter, i.e. how close to the true value it is likely to be.

The characteristics affecting the quality of an estimate based on the results of a sample survey or a census can be separated into two classes:

- “Statistical characteristics” describe the impact of the randomness in the random sampling protocol and the impact of the properties of the *estimator* (the mathematical computations) used to obtain the estimate. Statistical characteristics of an estimator are mathematical properties: they can be established through mathematical proofs and assessed using statistical computations. The random sampling protocol presumed to have been used (e.g., simple random sampling, stratified sampling) determines the statistical methods that are applied. Statistical characteristics, as defined above, are not present in censuses unless the non-response rate is high.

Assessment of the statistical characteristics affecting the quality of a sample survey must include an assessment of the standard error and of the estimator bias based on statistical analysis of actual or, possibly, similar or historical data, or based on theory.

- “Operational characteristics” are related to the implementation of the sampling protocol and properties of the estimator or of model-derived estimates. They include either deliberate or unintentional differences between the actual sampling protocol and the sampling protocol assumed in the statistical analysis, i.e., departures from the sampling protocol and other sources of errors due to the implementation of the program. They also include characteristics such as measurement errors and errors associated with calculating the estimates (e.g., calculating total catch weight from gutted-weight observations and a gutted-weight conversion factor). Operational characteristics can impact both sample surveys and censuses.

In the proposed assessment methodology, the assessment of the quality of the estimation process can, in many cases, be broken down into assessing the prevalence and the impact of each operational characteristic potentially affecting the quality of the estimates.

4.2. 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.

We separate the statistical objectives of fishery monitoring programs into two classes: measurement and compliance. Measurements are important for administrative purposes (e.g., reporting the total economic value of given fishery) or scientific purposes (e.g., stock assessment). 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 of the parameter is used to determine if the limit has been respected or not.

Measurement and compliance are inherently different types of objectives requiring different assessment approaches. The assessment of dependability for measurement objectives will be based on comparing the quality of the estimate with the scientific or administrative requirements. For example, this may involve requirements for the *precision* (defined below) of the estimate. The assessment of dependability for compliance objectives will be based on a hypothesis testing framework, e.g., 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.

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 boardings by fisheries officers). The present report does not address non-statistical objectives.

5. MAIN DEFINITIONS

We define the most important concepts used in this report. Other terms are defined in the Glossary.

5.1. Preliminary: what does "on average" mean?

Throughout this section of the report the expression "on average" is used several times and warrants some attention.

Suppose that one desires to estimate the total fishing effort (fisher days) for Miramichi River recreational salmon fishers using a telephone sample survey. At the end of the season, a random sample of recreational fishers holding a license is called and asked how many days they fished. The total effort will be estimated by multiplying the average effort reported by fishers in the telephone sample survey, by the total number of fishers holding a license.

Now, consider the following "thought experiment". Suppose that we can repeat this process a very large number of times in identical conditions. Due to difference between random samples, differences between cases of non-response, difference between momentary memory lapses, etc., we would get a different estimate each time. ***"On average" means taking an average of the estimates obtained from these impossible-to-carry-out, theoretical repetitions.*** Mathematical theorems give information about the impact of randomness on these estimates. However, it is not possible to say something specific about, for example, the error in a single estimate (e.g., how different it is from the population "true" value) because the true value is unknown.

Note that we use "on average" to refer to any kind of mean, including the quadratic mean used in the computation of the standard error. Furthermore, if only sampling randomness was considered, in statistics this would be termed the "expected value".

Technical note: The central limit theorem

In classical statistics, computation of the sampling error relies on mathematical theorems and, most often, on the central limit theorem proven by the French mathematician Pierre-Simon Laplace (Laplace, 1812).

The central limit theorem states that the means of random samples drawn from a population with mean μ and variance σ^2 will have an approximately normal distribution with a mean equal to μ and a variance equal to σ^2/n, if n, the sample size, is "large" and the population size is much larger than n. It is the central limit theorem that gives the usual formula for the confidence interval: $\bar{X} \pm z_{\alpha/2} \times s/n^{1/2}$ where $\bar{X}$ and s^2 are respectively the sample mean and variance used to estimate the population mean μ and variance σ^2.

Under restrictive assumptions, the central limit theorem can be used in other sampling situations (e.g. to estimate the parameters of a regression).

5.2. Estimator vs estimation process

Obtaining a parameter estimate from data involves mathematical computations. For example, given a simple random sample, the population mean is usually estimated by computing the usual ("arithmetic") mean of the observations. The term "estimator" will be used to refer to this mathematical step. Another estimator of the population mean is the truncated mean: it involves computing the mean of the observations remaining after removing, say, the 5% smallest and the 5% largest observations. This is a different estimator of the mean (in some situations, it is a better estimator than the usual mean).

The estimator refers only to the mathematical computations. The estimation process involves an estimator but also includes all the other steps leading to the estimate.

5.3. Quality of an estimation process

The quality of an estimation process describes the quality of the estimation, i.e. how close to the true value the estimate is expected to be. The quality of the estimation will depend on its accuracy (converse: inaccuracy or bias) and its precision (converse: imprecision or variability).

5.3.1. Accuracy/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 estimation process bias, or, in this document, simply, the bias, is the average of the differences between the estimated values and the true value, if the estimation process was repeated many times. The term bias will also be used in another context. “Estimator bias” is a statistical concept defined mathematically and is unrelated to factors such as measurement errors and unintended deviations from planned sampling protocols. The term estimator bias will be reserved for this specific case.

Bias has a sign: it is either positive (a tendency to over-estimate the true value) or negative (a tendency to under estimate the true 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.

The converse of bias is accuracy. On occasion, we will use “inaccuracy” as a synonym of bias.

5.3.2. Precision/variability

If the same estimation process is carried out many times, the estimates may differ from each other due to the randomness of the sampling protocol or to some other characteristic. We use the term variability to describe this variation: it is a measure of how much estimates from an estimation process vary, on average. Variability does not have sign (it is a positive number): it only describes how much the estimates would differ from each other if one repeated the same estimation process many times.

The *standard error* of an estimator (see § 6.1.2) is a measure of its variability: it measures the variability due randomness of the sampling protocol.

We reserve the term *variance* for the variability due to a random sampling protocol as usually measured in statistics. In this case, “on average” refers to a quadratic mean.

Variability typically decreases as sample size increases. In a true census, there is no error due to randomness of the sampling, but variability may still exist due to other error sources such as measurement error, transcription error, and recall error.

The converse of variability is *precision*. On occasion, we will use “imprecision” as a synonym of variability.

5.3.3. Quality of an estimation process

The quality of an estimation process is determined by the combination of accuracy and precision for the process.

Figure 2 and Figure 3 illustrate the four possible combinations of accuracy and precision, the components of quality.

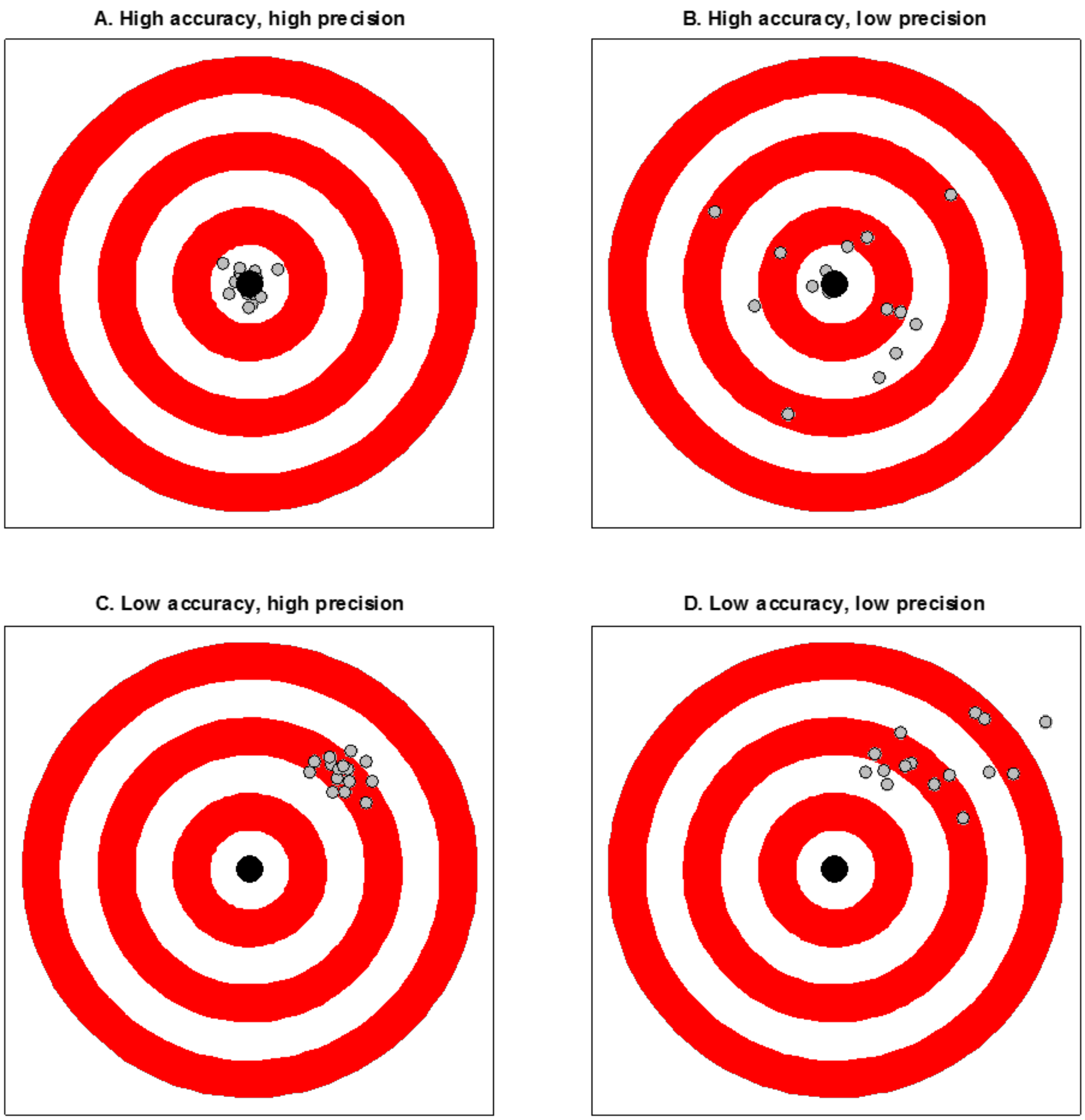


Figure 2. Classic illustration of accuracy (converse: bias) and precision (converse: variability), the components of quality. The middle of the bullseye represents the true value of the parameter and the dots represent examples of estimated values from (theoretical) repetitions of the estimation process.

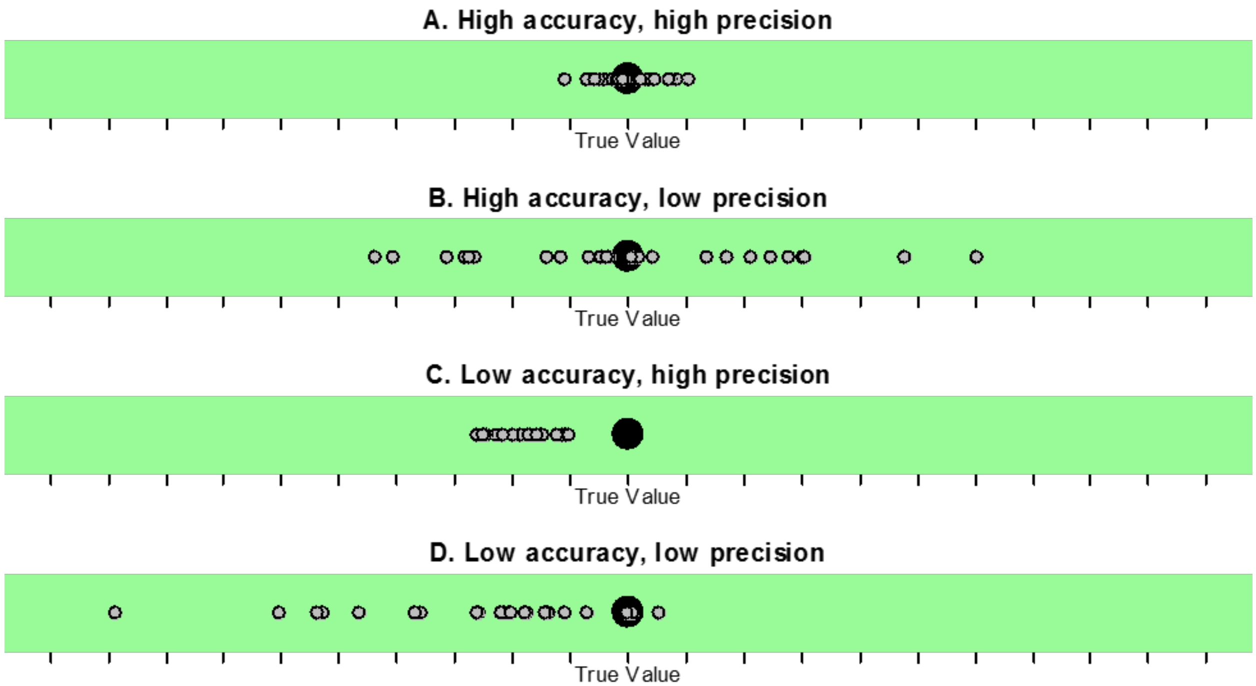


Figure 3. A one-dimensional illustration of accuracy (converse: bias) and precision (converse: variability), the two components of quality. The solid dot represents the true value of the parameter and the grey dots represent examples of estimated values from (theoretical) repetitions of the estimation process.

Technical note: Bias and variability are non-intersecting, "orthogonal" concepts

Bias and variability are non-intersecting, "orthogonal" concepts: they measure entirely different aspects of an estimation process. For example, consider a sample survey with a large sample size to estimate discards. Suppose that large discard values are systematically under-reported. The estimate would have small, correctly estimated, variability (statistically measured by the standard error) due to the large sample size but would have a large negative bias due to the under-reporting of large values. ***The assessment should reflect this situation: low variability and large negative bias.***

If the under-reporting was corrected, the variability of the estimation process, and therefore the standard error, would increase but the bias of the estimation process would decrease. The assessment should ***not*** suggest that the under-reporting contributes (negatively) to the variability of the estimation process. It should only indicate that it contributes to the bias of the estimation process.

In statistics, the root mean square error (RMSE) summarizes how well an estimator estimates a parameter. The equality $RMSE^2 = \sigma^2 + Bias^2$, essentially expressing Pythagoras' theorem, describes the orthogonality of the statistical concepts of variability and bias.

5.4. Characteristics influencing quality, with examples

Several characteristics of an estimation process influence its quality. These characteristics can influence the accuracy and the precision of the process in different ways. In this report, characteristics influencing quality are divided into two groups: statistical characteristics and operational characteristics (see § 6).

The impact of statistical and operational characteristics on bias and variability differs (Table 1). In sample surveys, statistical characteristics have a central impact on variability but, in most cases, they will have a small or correctible impact on bias. In censuses, they have no impact. Operational characteristics can have an important impact on bias in both sample surveys and censuses but will often have a small or negligible impact on variability.

For this report, the definition and evaluation of operational characteristics affecting sample surveys draws heavily on Groves et al. (2009). This book, while oriented toward human population sample surveys, gives an extensive overview of operational difficulties present in sample surveys of all types. For a discussion of operational sources of bias and variability in fishery monitoring programs, specifically at-sea observer programs, see Babcock and Pikitch (2003).

Table 1. Overview of the impact of the characteristics of an estimation process on its quality.

		Characteristic	
		Statistical	Operational
Quality	Bias	Usually quantifiable; correction sometimes possible; ***often, small impact***; in most cases, does not apply to censuses.	Often present, documented in the survey literature; difficult to identify and/or quantify in a specific sample survey or census; ***typically, important impact*** including in censuses.
	Variability	***Very important impact***. Always quantifiable; in sample surveys based on a probabilistic sampling protocol; this is the core of classical statistical inference; does not apply to censuses.	For some operational characteristics, ***a small impact***, including in censuses (example: measurement errors). For others, ***an important impact*** (example: differences between the actual sampling protocol and the protocol assumed in the computation).

Examples of estimation process characteristics and their potential impact on quality are provided to illustrate the concepts presented above, however specific details on the full suite of operational characteristics are only provided later.

5.4.1. Impact of statistical characteristics on bias/accuracy

Some estimators, under specific sampling protocols, are biased. This is a statistical characteristic of the sampling protocol and the estimator. In some cases, there are statistical methods to estimate this estimator bias (analytical methods, bootstrapping) and to correct the estimate for the estimator bias. Whether or not such methods were applied should be part of the monitoring program documentation.

The (mathematical) bias of the estimator used, if any, in an estimation process contributes to the bias of the estimation process if no correction has been applied.

In most cases, the impact of statistical characteristics on bias/accuracy of the estimation process (i.e. of the estimator bias) will be small.

EXAMPLE: Under simple random sampling, the sample mean as an estimator of the population mean is not biased: on average, it gives the population mean. The estimator bias is 0.

EXAMPLE: Suppose that, in a given fishery, effort is known for each trip and that catch is observed at sea for a simple random sample of the trips. Total catch can be estimated using a *ratio estimator*, i.e. multiplying the effort by the estimated catch per unit of effort for the observed trips. The ratio estimator is known to be biased and formulae have been developed to estimate and correct for this bias (Adbola and Oshungade, 2012).

Technical note: statistical bias and censuses

The statistical bias of most common estimators decreases as the sample size increases (example: the biased of the ratio estimator – see Lohr § 4.1.2). Exceptions may be mostly academic (example: using a trimmed mean to estimate the mean of a skewed distribution).

The bias due to operational characteristics does not decrease as the sample size increases. For example, an incorrect scale tare will lead to a biased estimate of the mean weight even in a census.

5.4.2. Impact of statistical characteristics on variability/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* (SE) of the estimator, the relative standard error *(*RSE; the SE divided by the estimate*)* or by the confidence interval (CI) of the estimate for a specified confidence level (typically 95%). The relative standard error (RSE) is the main measure of variability used in the remainder of the report.

The method of computation of the standard error (or of other measures of sampling error) depends 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. The sample size has a direct impact on standard error. However, the sampling proportion (i.e. the sample size relative to the total number units in the target population) can also be important if the proportion is large (e.g. 40%). For further details, see § 13.1.

In sample surveys, the impact of statistical characteristics on variability/precision should be at the core of an assessment.

EXAMPLE: Under simple random sampling, the standard error of the sample mean as an estimator of the population mean is approximately $s/n^{½}$ where s is the sample standard deviation and n is the sample size.

EXAMPLE: The population standard deviation can be "guesstimated" by (97.5th centile – 2.5th centile)/4. The SE of the mean can then be obtained by dividing the result by $n^{½}$. A similar heuristic approach can be applied to some more complex estimators.

5.4.3. Impact of operational characteristics on bias/accuracy

At implementation, there may be differences between the actual sampling protocol and that presumed in the statistical computation, and departures from planned protocol. Hereafter, we will use the terms unintended and unintentional when referring to these departures, which were neither planned nor assumed by the analyst during data processing (i.e., not accounted for). These unintended departures can cause biases in the estimation process. The presence of such biases has been documented in the general literature on surveys and in fishery-specific literature (e.g., observer effects: Benoît and Allard, 2009; Faunce and Barbeaux, 2011). However, it may be difficult to demonstrate and to quantify bias in a specific survey.

Here, the estimated contribution of operational characteristics to bias will be added arithmetically (taking into account the sign of the estimated impact) to the computed estimator bias to adjust the assessed quality of the estimation process.

The potential impact of operational characteristics on bias/accuracy should be considered important in the assessment of both sample surveys and censuses.

EXAMPLE: If the sample is taken in such a manner that it is not representative of the whole population, then bias can occur. For example, at-sea observers may be excluded from trips where high bycatches are likely, leading to a negative bias on bycatch estimates.

EXAMPLE: Misreporting in logbooks may follow a specific pattern. Most often, logbook misreporting will tend to under report incidental catches, in which cases the estimates of these catches will be negatively biased. However, in some situations, it may tend to over report catches and, therefore, be positively biased.

EXAMPLE: A continuous video monitoring program is intended as a census. However, inclement weather may interrupt data availability. If weather and the quantity measured (for example, catch per tow) are correlated, the impact of this operational characteristic can create a bias.

5.4.4. Impact of operational characteristics on variability/precision

The RSE obtained from statistical analysis reflects mostly the variability of the estimator due to the randomness of the sampling protocol. The RSE computation formulae depend on the sampling protocol that is assumed. Unintended departures from the sampling protocol may result in a mischaracterization of the estimation process variability, which would otherwise be estimated from the RSE.

The impact of departures from the sampling protocol on variability will be best described by a correction factor to be applied to the RSE to adjust the assessed quality of the estimation process. For example, a 20% increase will mean that the computed standard error must be multiplied by 1.2 to assess the true variability of the estimation process; similarly, a 20% decrease will mean that standard error must be multiplied by 0.8. Methods based on sampling theory to estimate the required correction factor are proposed in the text (see § 13.3, 13.4, 13.5).

The impact of various operational sources of random error will be best described as a quadratic addition to the RSE, i.e. taking the squared root of the sum of squares. For example, for the estimation a mean under the simple random sampling protocol, if the error of measurement tool has a standard deviation of σ_ε, then the corrected relative error of the estimation process will be $\sqrt{RSE^2 + (\sigma_\varepsilon/\mu)^2/n}$ where μ is the anticipated mean and n is the sample size (see § 13.6).

The following examples illustrate the impact of operational characteristics on variability.

EXAMPLE: In a census, unintentional (random) non-response will create variability like that of sample survey. The statistical analysis yields RSE = 0, since it is a census. However, the actual variability is positive. If the rate of unintentional non-response is very high (say, above 20%), the RSE should be computed using the actual number of observations. In the sampling case, the RSE will naturally be computed using actual number of responses, i.e. considering the number of non-responses.

EXAMPLE: Consider an observer monitoring program that assumes a simple random sampling protocol. Suppose that, for practical reasons, the sampling is clustered: an observer travels to a randomly chosen location, samples the fishery occurring at that location, moves to another randomly chosen location, samples, and so on. The sampling protocol actually applied is called "cluster sampling". Unwittingly analyzing the resulting data assuming simple random sampling could cause the RSE to incorrectly estimate the true sampling variability. If the locations are similar to each other relative to the parameter of interest (i.e. there is little variation between locations), the RSE computed assuming simple random sampling will correctly represent the true estimator variability. If the locations are not similar to each other, the RSE computed assuming simple random sampling will underestimate the true variability (see a more technical explanation in § 13.3).

EXAMPLE: More generally, in a sample survey based on random sampling, the actual sampling protocol may be different from that determined by the sampling protocol: a subset of the target population will be excluded from the sampling protocol or subjected to a sampling probability different than that determined by the sampling protocol. If this subset is like the entire population, the variability of the estimation will not be impacted. However, if this subset is different (e.g. for estimating a mean or a total, more homogeneous or more heterogeneous; for estimating a ratio, closer or farther from the regression line) from the entire population, the estimated variability (estimated RSE) will be different than the true variability of the estimator (see more technical explanations in § 13.2).

EXAMPLE: A continuous video monitoring program is a census. However, inclement weather may interrupt data availability. If weather and the parameter (for example, catch per tow) are uncorrelated, the impact of this operational characteristic will create variability: the planned census will have become a sample survey.

EXAMPLE: Variation between observers in their visual estimation of catch amounts may increase the variability of the catch estimation in an observer program. However, unless the variation is very large, variability will not be meaningfully impacted.

In sample surveys, we expect that the impact of many operational characteristics on variability/precision will tend to be small and negligible. However, the impact of irregular sample selection (for example: targeted sampling) that is unaccounted for may be important.

5.5. Dependability

We will use the term *dependability* to describe the overall adequacy of a monitoring tool relative to the objective of the monitoring program. Dependability is a function of the accuracy and the precision of the estimation process (jointly, the quality) and the objective for the estimate. As indicated above, the estimates obtained from monitoring programs address one of two types of objectives: measurement and/or compliance.

For scientific research, stock assessment, administrative purposes or environmental reporting and protection, the quality of the estimate (measurement) must be sufficient to accomplish the objective.

EXAMPLE: To monitor the removals from a population of a long-lived, at-risk marine mammal species, scientists may require that the annual estimate be within $\pm$10% of the true value (with a 95% probability to account for the sampling randomness). In contrast, to monitor removals from a large, lightly exploited and productive population, an annual estimate within $\pm$50% of the true value (again with 95% probability) might be sufficient.

For compliance application, the objective is to determine if the parameter satisfies or not some regulatory limit. In this situation, the quality of the estimate should be sufficient to ensure that the correct decision is reached with a given probability. If exceeding the limit results in a high risk to conservation of the population, then a high probability of making the correct decision would be required; conversely if the risk to conservation is low, a lower probability could be acceptable.

EXAMPLE: For a given probability of making a correct decision for a fishery, if the typical catch is close to the total allowed catch, a high-quality estimate will be required to reach the correct conclusion. If the typical catch is far from the limit, a much lower quality estimate will be sufficient.

Terminology
The choice of terms (accuracy, bias, precision, variability, quality, dependability, etc.) used in this report is not universal. Statisticians, biologists, social scientists, medical researchers, etc. have adopted certain terms with discipline-specific definitions. For example, the term "validity" may be interpreted differently by different people. The choice of terms used here is not necessarily better then other choices but was felt to be reasonable and is largely consistent with the use of these terms in fisheries science.

6. CHARACTERISTICS OF AN ESTIMATION PROCESS IMPACTING ITS QUALITY

We now describe the main characteristics impacting the quality of a monitoring tool. ***Assessment of these characteristics will form the basis of the assessment of an estimation process***. The list is detailed but not exhaustive and other characteristics could be added to suit the needs of particular monitoring programs.

Following the description of each characteristic, we present the anticipated contribution to the bias and the variability of the estimation process and some examples.

Note that the following list of characteristics is appropriate for parameters referring to both common and rare events. However, the measurement methods and the anticipated impact will generally differ between the two. For example, intentional non-response or misidentification may have a very large impact on the estimate of the bycatch of a rare, possibly little-known, species.

6.1. Statistical characteristics

6.1.1. Estimator bias

By estimator bias, we mean the expected value of the difference between the estimate and the true value. For true censuses, the estimator bias is 0. The relative estimator bias is the quotient of estimator bias over the true value, typically expressed in percentage.

The estimator bias can be computed using statistical methods, theoretical considerations (e.g. central limit theorem) or numerically (e.g. by bootstrapping). It is sometimes possible to correct for estimator bias using established statistical methods. In the assessment, if such a correction is applied, only the estimator bias ***after correction*** is important.

Measurement*:* The estimator bias should be reported with the estimate unless a reliable bias correction has been applied, in which case it should be reported as 0.

Impact: In most cases, the impact of the estimator bias, after correction, on the bias of the estimation process, will be small.

EXAMPLE: Under simple random sampling, the estimator bias of the sample mean as an estimator of the population mean is 0 and the relative estimator bias is 0%.

EXAMPLE: Biases inherent in ratio estimators, such as used in the analysis of creel sample surveys, are well known and quantitative correction methods exist. Once applied, the relative estimator bias should be 0%.

6.1.2. Standard error

The standard error measures the variability of the estimates due to the randomness of the sampling. For censuses, the standard error is 0.

The standard error will reflect the impact of some operational characteristics (as described in § 6.2). To understand this impact, consider the following "thought experiment": suppose that, using a simple random sampling protocol with sample size n, we are estimating the mean of a variable that has a very small standard deviation σ. If the measurement error is negligible, the standard error of the sample mean, $\sigma/\sqrt{n}$, will be small. If measurement error is very large, the standard error of the sample mean will be large, not due to the randomness of the sampling but to the measurement errors. See §13.6 for a detailed explanation.

Measurement*:* The standard error of an estimator is a common measure of the variability of an estimate due to the randomness of the sampling. The relative standard error (RSE) is the quotient of the standard error over the true value, typically expressed in percentage. For assessment of estimation processes, we retain the RSE. The RSE is akin to the coefficient of variation for the population mean.

The RSE of an estimator can be computed analytically, based on theoretical considerations (e.g. central limit theorem), or numerically (e.g. bootstrap). It would normally be computed and shown with estimated values.

If the RSE has not been reported, it can be approximated, for the estimation process assessment, by borrowing from other similar sample surveys, by carrying out some simulations or by heuristic computations.

Impact: ***In sample surveys, the standard error will be the basic source of the variability of an estimation process.*** When a well-designed random sampling protocol is implemented nearly perfectly, the standard error will represent most of the variability.

EXAMPLE: Under simple random sampling, the standard error of the sample mean as an estimator of the population mean is approximately $\sqrt{1-n/N}\,s/\sqrt{n}$ where s is the sample standard deviation, n is the sample size and N is the population size. Therefore, the RSE can be estimated as $\left[\sqrt{1-n/N}\,s/\sqrt{n}\right]/\bar{X}$ where $\bar{X}$ is the sample mean. When the sample size is small relative to the population size, the factor $\sqrt{1-n/N}$ is close to 1 and can be ignored.

EXAMPLE: Often, the population standard deviation *s* can be approximated by $s_* = 97.5\text{th centile} - 2.5\text{th centile})/4$. Then, the RSE can be "guesstimated" as $\left[\sqrt{1-n/N}\,s_*/\sqrt{n}\right]/\bar{X}$. Similar heuristic approaches can be applied to more complex estimators.

Technical note: What is the bootstrap?

Mathematical theorems to estimate estimator bias and standard error exist only for a limited number of sampling situations and estimators. Bootstrapping allows for the estimation of estimator bias and standard error in many or, even, most situations (Efron, 1979).

Bootstrapping is "computer intensive"; it relies on resampling the observations thousands of times and computing the estimator on these new samples. In effect, bootstrapping assumes that the sample of observations reflects the population so that resampling simulates the original sampling protocol.

Bootstrapping is now routinely applied to estimate bias and standard error.

6.2. Operational characteristics

Statistical computations depend on the sampling protocol. In many cases, the implementation of a monitoring program will depart from the planned sampling protocol. When the actual sampling does not correspond to the planned sampling protocol, the bias and standard error of the estimator may not describe correctly the variability and the bias of the complete estimation process. The following is a list of most likely departures and their respective potential consequences on the quality (precision and accuracy) of the estimation process.

The impact of some of the following operational characteristics will already be reflected, at least partly, in the calculated RSE and bias. Measurement error is one such characteristic. These characteristics are nonetheless discussed below, and their impact should be quantified as they present opportunities for improving the quality of estimates. For example, using a motion-compensated balance to measure catch weight, rather than a visual estimate, will lead to a smaller RSE and reduce the estimation process variability. Unbiased response errors are another example of such a characteristic.

The following definitions apply hereafter.

"*Independent observer data*" refers to measurements collected by a person or a technology specifically tasked with observing and reporting on fishery activities and at arm's length from the fishing industry or community, such as at-sea and dockside observers and on-board cameras and vessel monitoring systems.

"*Resource user data*" refers to measurements made and/or reported by the fishing industry or community, including fishers, plant personnel and buyers. Logbook records, purchase slips and answers to recreational fisher surveys are examples of resource user data.

We separate operational characteristics related to these two cases for the following reasons. Independent observers are presumed not to be in conflict of interest and in many instances are certified as such. They can be required to undergo training to carry out specific measurements or observations, a potential remedy for, say, misidentification of species of conservation concern (SCC). In contrast, resource users can be in conflict of interest (for example, when reporting of discards if high discard levels lead to a fishery closure). Furthermore, it is difficult to require that resource users undergo training on reporting.

Thus, independent observer data are expected to be relatively unbiased under normal operating conditions, whereas resource user data may be subject to bias given conflicts of interest and incentives. Nonetheless, while independent observers are meant to be arm's length and without conflict of interest, voluntary or involuntary (e.g., resulting from intimidation) collusion with resource users would likely violate the assumption of unbiasedness.

6.2.1. Undercoverage

In this document, we use the term "*frame*" to mean the ***available*** list of units of the target population. This is the simplest case of "frame". The document points out situations where a more general definition of "frame" is required. In a census, the census frame is the list of all units to be observed. In a sample survey, the sampling frame is the list of all units from which the sample will be drawn. Ideally, the frame and the target population would be identical. The frame may be created before the survey (e.g. the holders of a recreational fishery license for the current season) or during the survey (e.g. the herring gillnet fishing trip in Herring Fishing Area 5).

Coverage refers to the relationship between the target population and the frame.

Undercoverage occurs when a subset of the target population is not included in the sample survey sampling frame or in the census frame. In a census, the excluded units will not be observed. In a sample survey, the excluded units cannot be part of the sample.

Contribution to estimation process bias: If the excluded subset is like the target population, the undercoverage will have only a small to nil impact on the bias. If the excluded subset is different from the target population, the impact will depend on the difference, as follows for the estimation of a mean or a total:

- Excluded subset associated with large values: the under-coverage will result in a negative bias (sample surveys and censuses);
- Excluded subset associated with small values: the under-coverage will result in a positive bias (sample surveys and censuses);
- Excluded subset associated equally with small and large values: no impact on bias;
- Excluded subset associated with middle values: no impact on bias.

Statistically speaking, the third case above is similar to using a truncated mean. For symmetrical distributions with heavy tails, this type of undercoverage may in fact represent a more efficient sampling scheme.

Contribution to estimation process variability: None anticipated.

EXAMPLE: Incomplete list of recreational fishers targeted for a post-season interview.

EXAMPLE: Vessels occasionally forget to hail-out prior to departure, causing trips to be randomly excluded from the on-board observer program.

EXAMPLE: Observers avoiding uncomfortable vessels that otherwise have fishing patterns like other vessels in the fleet, resulting in no impact on bias.

EXAMPLE: Suppose that vessels prone to illegally discarding avoid timely pre-departure hail-outs, causing trips with large bycatch to be excluded from an at-sea observer program. In a sample survey or a census, the estimate of the total bycatch will be negatively biased.

EXAMPLE: In a sport fishery, occasional, less skilled fishers who happen to have lower catch rates may be underrepresented, leading to a positive bias in estimated catch.

6.2.2. Overcoverage

See the definition of frame in § 6.2.1.

Overcoverage occurs when units outside the target population are incorrectly included in the sampling frame or census frame. In a census, the outside units will be incorrectly included in the observations. In a sample survey, the outside units may be selected as part of the sample.

Contribution to estimation process bias: The impact will be generally opposite to that of undercoverage (see 6.2.1 Undercoverage). For example, if the incorrectly included subset is associated with large values the over-coverage will result in a positive bias.

Contribution to estimation process variability: None anticipated.

EXAMPLE: Incorrect allocation of a trip to a target species.

EXAMPLE: Fishing trip outside the target fishing zone of the monitoring program incorrectly recorded within the monitoring program.

6.2.3. Unintended clustering of samples (sample surveys only)

In a sample survey, cluster sampling occurs when subsets are selected (stage 1) and units are selected from these subsets (stage 2). Cluster sampling is unintentional if the sampling protocol did not call for this sampling protocol.

Cluster sampling consists of partitioning the population into subsets called clusters, taking ***a random sample of clusters***, and then, taking a census from each selected cluster (one-stage cluster sampling) or a sample from each selected cluster (two-stage cluster sampling). Using appropriate computations, cluster sampling may yield a lower total cost than simple random sampling, especially if the cost of reaching each statistical individual or unit is high. A basic assumption of cluster sampling is that all clusters are similar to each other and to the whole population. A key distinction here is that the clusters to sample are chosen randomly. The case where some clusters have a sampling probability of zero is treated in 6.2.5 below.

If cluster sampling is occurring unintentionally while formulae for simple random sampling are used, and clusters are more homogeneous than the population, the computed standard error will be ***smaller*** than the true standard error, and the variability of the estimation process will be underestimated.

In practice, it can be very difficult to detect whether cluster sampling has occurred simply by examining the data because even apparent clustering in the data may have resulted from a broader random process. Understanding the sampling protocol, such as how observers are deployed, relative to how the fishery operates, is a more reliable manner of inferring whether cluster sampling is likely to have occurred. For example, if observers must travel long distances to reach certain ports, fishing trips from more proximate ports may be sampled preferentially and some very distant ports may only be sampled in some years.

Contribution to estimation process bias: None expected.

Contribution to estimation process variability: Variability will be underestimated if clusters are more homogeneous than the population.

EXAMPLE: Allocation of monitoring resources to wharfs, where vessels at different wharfs have different catch characteristics (e.g. each week, 3 wharfs are randomly selected for monitoring); allocation of monitoring to time periods, where the catch increases toward a peak season and then decreases (e.g. monitoring will occur in randomly selected weeks 2, 7, 9 and 12 of the fishing season).

Technical note: Computing the impact of unintended clustering on variability of the estimation process

The impact of unintended clustering on the variability, and therefore the correction factor that could be applied to the standard deviation, can be estimated using the following formula in the case where the number of units sampled within each cluster is close to the total number of units in the cluster, i.e. close to one-stage clustering:

N: anticipated number of clusters

n: anticipated number of clusters sampled

M: anticipated average number of units within each cluster

m: anticipated average number of units sampled within each cluster

$s^2_{between}$: estimated or anticipated variance between cluster means

s^2_{within}: estimated or anticipated variance within each cluster

s^2: estimated or anticipated population variance

$$\sqrt{\frac{\left(1-\frac{n}{N}\right)\frac{s^2_{between}}{n}+\frac{1}{N}\left(1-\frac{m}{M}\right)\frac{s^2_{within}}{m}}{\left(1-\frac{nm}{NM}\right)\frac{s^2}{nm}}}$$

The term $\frac{1}{N}\left(1-\frac{m}{M}\right)\frac{s^2_{within}}{m}$ is often negligible if N is large.

For the assessment of monitoring programs, the variances can be estimated from heuristic rules (e.g. $[0.25\times\text{range}]^2$), historical data, outside sources, etc.

In more complex cases of unintended clustering, more elaborate computational approaches are required.

6.2.4. Unintended sampling stratification (sample surveys only)

In a sample survey, stratified sampling occurs when the target population has been partitioned into subsets (strata) and a sample is drawn separately from ***each*** stratum.

Strata differ from clusters (6.2.3 above): in stratified sampling, a sample must be drawn from each stratum; in cluster sampling, a sample or a census is taken from a sample of clusters.

Stratified sampling is unintentional if the sampling protocol did not call for this sampling protocol. For example, unintended stratification could be temporal (e.g. equal number of observations each week of the season or a number of observations proportional to the weekly number of units) or spatial (e.g. one observer assigned to each dock).

Correctly planned and implemented stratified sampling can be efficient from a sampling perspective. Using appropriate computations, stratified sampling will yield lower standard error for a given total sample size if the strata are more homogeneous than the population.

If stratified sampling is occurring unintentionally while formulae for simple random sampling are used and strata are more homogeneous than the population, the computed standard error will be ***larger*** than the true standard error, and the variability of the estimation process will be overestimated.

Of course, the statistician could compute the true RSE of the estimator using the formulae for stratified sampling if he was aware of the stratification and if the stratum information had been included in the dataset. See §13.4.

Contribution to estimation process bias: Bias may result if strata are more homogeneous than the population and the sample allocation is not proportional to stratum size.

EXAMPLE: In a river salmon recreational fishery, separate allocation of the sampling effort to all pools, where the pools have different characteristics.

EXAMPLE: In an at-sea observer program where one observer is deployed to each port and some ports have more vessels which also have a greater propensity for large catches.

Contribution to estimation process variability: The variability of the estimation process will be smaller than that described by the standard error if strata are more homogeneous than the population.

6.2.5. Other irregular selection probabilities or exclusions

Besides unintended cluster or stratified sampling, several other situations may cause the sample selection probabilities to be incongruent with those determined by the sampling protocol.

Note: Irregular selection probabilities or exclusions refer to events ***after*** the frame has been established. Exclusions that occurred when the frame was established are addressed in 6.2.1

A. Some units may be unexpectedly excluded from the frame. In a census, they are not observed. In a sample survey, they are excluded from the sample.

EXAMPLE: The sole observer a certain dock is sick for 2 weeks and cannot be replaced. In a sample survey, the probability of any trips leaving during these two weeks being sampled becomes 0. In census, they are not included.

B. In a sample survey, some units not selected within the sampling protocol may be included in the sample for external reasons. Their probability of inclusion in the sample is 1.

EXAMPLE: Targeted sampling, where given units (a specific fishing trip, the activities a of specific fisher, etc.) are included in the sample independently from the sampling protocol, is an important example of such a situation. The probability of selection of the targeted samples is 1.

Targeted sampling of repeat offending fishers is common in enforcement and compliance contexts. Due to the private nature of these investigations, analysts will often not be able to distinguish which samples were targeted and which were selected based on the sampling protocol when analyzing the resulting catch data. It can therefore be very difficult to accurately account for targeted sampling in these contexts.

C. In a sample survey, the probability of a unit being included in the sample may be different from the probability determined by the sampling protocol.

EXAMPLE: At-sea observers preferring particular vessels because of comfort or attitude of the crew.

EXAMPLE: Weather events reducing at-sea observer coverage.

Contribution to estimation process bias: Likely to create bias.

Contribution to estimation process variability: Increase or decrease variability, depending on the details.

EXAMPLE: At-sea observers tend to avoid particular vessels due to the poor quality of the food or cleanliness. May not impact bias or variability if these vessel characteristics are unrelated to the characteristics of the catch.

EXAMPLE: At-sea observers tend to avoid specific vessels due to frequent harassment by the captain. May impact bias if the harassment behaviour is associated with an aversion to follow responsible fishing practices that have an impact on catch composition (e.g., voluntary fisher agreement to avoid fishing in certain areas).

EXAMPLE: Statistical analysis conducted assuming simple random sampling of fishing trips while actual sampling includes trips that were specifically targeted by at-sea observers for enforcement purposes. If the targeting due to repeated breach of catch regulation, the targeting is likely to contribute to the bias of the estimation process. The impact may be extremely difficult to assess since the targeted fishers are likely to change behaviour when an observer is on board. See observer effect in § 6.2.6.

Technical note: Computing the impact of targeted sampling

It may be possible to roughly estimate the impact of targeted sampling on the estimation of a mean or population total based on the following statistical considerations.

Let

N: population size

n_r: size of the random sample

n_t: size of the targeted sample

$n = n_r + n_t$: total sample size

$\bar{X}_r$: mean of the random sample

$\bar{X}_t$: mean of the targeted sample

s_r: standard deviation of the complete sample

s_r: standard deviation of the random sample

$s_t = 0$: standard deviation of the targeted sample

$\hat{T}$: Estimate of the population total T

$\hat{\mu}$: Estimate of the population mean μ

Then, the computation of the estimates and their standard error is as follows:

$$\hat{T} = (N - n_t)\bar{X}_r + n_t\bar{X}_t$$

$$\widehat{\sigma_{\hat{T}}} = \frac{s_r}{\sqrt{n_r}}(N - n_t)$$

$$\hat{\mu} = \hat{T}/N = \left(1 - \frac{n_t}{N}\right)\bar{X}_r + \frac{n_t}{N}\bar{X}_t$$

$$\widehat{\sigma_{\hat{\mu}}} = \frac{\widehat{\sigma_{\hat{T}}}}{N} = \frac{s_r}{\sqrt{n_r}}\left(1 - \frac{n_t}{N}\right)$$

Not considering the targeting, the $\widehat{\sigma_{\hat{\mu}}}$ would be estimated by $\frac{s}{\sqrt{n}}\left(1 - \frac{n}{N}\right)$. The quotient $\frac{s_r}{\sqrt{n_r}}\left(1 - \frac{n_t}{N}\right)/\frac{s}{\sqrt{n}}\left(1 - \frac{n}{N}\right)$ is the correction factor that must be applied to the RSE to obtain the variability of the estimation process.

For the assessment of the estimation process, the various values must be estimated from historical data or other sources.

6.2.6. Observer effect

An observer effect occurs when the presence or anticipated presence of a human observer or of a technological surveillance tool causes a change in the fishing activity. There can be no impact of an observer effect if the entire population is observed, meaning, for example, that at-sea observers are present on all trips and observe continuously each trip.

Contribution to estimation process bias: Likely to create a bias. The magnitude of bias will depend on the degree to which the alteration in fishing activity results in a change in the property being measured (e.g., catch amount) and on the prevalence of observer effects.

Contribution to estimation process variability: None.

EXAMPLE: Selection of a different fishing zone or a more general change in fishing patterns when an observer is on board.

EXAMPLE: Fishing behavior changed if the vessel knows that it will be monitored when it returns to port, such as the release of a non-retainable species that it would otherwise have illegally been kept.

EXAMPLE: Logbook entries may be accurate mainly when an observer or enforcement officer is aboard ensuring compliance. In this case, the observer effect is to eliminate biased reporting of resource user data only for the monitored trips.

EXAMPLE: An at-sea observer is present on all trips. However, he/she observes only 25% of the fishing activity on each trip. The vessel may change its discarding process when the fishing activity is not observed.

6.2.7. Missing values due to unintentional factors, including unintentional non-response

In a sample survey, units are selected from the sampling frame in the sampling protocol; in a census, all units of the target population are selected. A missing value is data from any selected units that cannot be obtained.

An unintentional missing value occurs when an observation cannot be made due to event(s) outside the control of the people (fishers, plant personnel, observers) or technology involved in the monitoring.

Contribution to estimation process bias: If the missing values are random, no impact. If the missing values due to unintentional factors are larger or smaller, on average, than the typical population values they will impact the bias.

Contribution to estimation process variability: In sample surveys, increases variability since the actual sample size will decrease; this contribution will be included in the computed standard error. Therefore, a reduction of the number of missing values due to unintentional factors will lead to a smaller standard error and, therefore, a lower variability of the estimation process. In censuses, this may create variability that should be accounted for if the number of missing values is large, i.e., the census should be then analyzed as a sample survey.

EXAMPLE: In a post-season telephone sample survey, some selected recreational fishers are not reached within the indicated time frame; possibly unintentional, random non-responses.

EXAMPLE: In a in-season telephone sample survey of recreational fishers targeting effort, the most active recreational fishers are reached less often because they are often out fishing. This will contribute a negative bias to the estimate.

6.2.8. Missing values due to intentional factors, including intentional non-response

A missing value due to an intentional factor occurs when an observation is not made due to an intentional action, most often a refusal to provide information.

Contribution to estimation process bias: Missing values due to intentional action are likely to be biased and therefore to contribute to bias.

Contribution to estimation process variability: None anticipated.

EXAMPLE: In a pre-departure hail-out monitoring program used to plan at-sea observer activities, fishers intentionally do not send the hail-out or send it too late to allow the deployment of an at-sea observer.

EXAMPLE: In a log-book monitoring program, fishers intentionally do not supply some log-books.

Technical Note: Impact differences between missing values due to intentional and non-intentional factors.
Consider the following thought experiment. Suppose that we can repeat a post-season telephone survey of recreational salmon fishers many times in identical conditions. Consider missing values due to intentional factors, whereby some respondents have decided not to supply information. If one virtually repeats the survey, the same respondents will again refuse to supply the information and the missing values will always be the same. Therefore, these missing values will not contribute to the variability of the estimation process (the SE will correctly account for the variability of the responses of the other respondents) but may well contribute to its bias (the reasons for the refusal are likely related to the respondent's fishing activity). Consider missing values due to unintentional factors where some respondents are unavailable due to reason which are not their own and which are unrelated to the fishery. If one virtually repeats the survey, similar reasons may apply but to different respondents. Therefore, these missing values will contribute to the variability of the estimation process (the SE will correctly account for the variability of the responses and for the smaller sample size) but are unlikely to contribution to its bias.
Consider missing values may be due to unintentional factors related to the fishery. In the example above, the most active fishers may be more involved in the fishers' association and, therefore, more difficult to reach in the post-season survey. These missing values contribute to the variability of the estimation process and, possibly, also to its bias. This case also illustrates the issue of irregular selection probabilities. Finally, notice that the above thought experiment applies to a census.

6.2.9. Errors in data reported by resource users

In this section, we consider ***recurring*** errors related to the implementation of the program including, for example unintentional errors due lack of training, carelessness, etc. and intentional errors aiming to mislead fishery managers.

Errors are biased when their average is not 0, i.e. the reported values are more likely to be above and more likely to be below the true values. Errors will typically be biased when they are intentional, but biased unintentional error are also possible.

Errors will usually have some variability which will contribute to the variability of the estimation process. For non-census, this contribution will be partially accounted for in the SE. See § 13.6 for details.

Contribution to estimation process bias: Biased errors will to contribute to the bias of the estimation process.

Contribution to estimation process variability: Variability of the errors in resource user's data contributes to the variability of the estimation process. However, in sample surveys, the SE partly accounts for this contribution.

EXAMPLE: In an end-of-season mail-out survey of recreational harvesters for catch, poor recall leads to unintentional, unbiased error.

EXAMPLE: Recreational fishers under-report their effort due to forgetfulness leading to unintentional biased errors, in turn resulting in a negative bias.

EXAMPLE: Vessel logbooks underreport discards to avoid reaching or exceeding bycatch limits or to avoid stigma by environmental organizations or consumer groups, leading to intentional biased errors.

EXAMPLE: Commercial sales slips incorrectly reported to hide or to inflate catch, leading to intentional biased errors.

6.2.10. Errors in data reported by independent observers

In this section, we consider ***recurring*** errors related to the implementation of the program including, for example, unintentional errors due lack of training, carelessness, etc. and intentional errors aiming to mislead fishery managers.

Intentional errors by independent observers are anticipated to be less prevalent than those by resource users. However, they may occur in cases of collusion, harassment, etc.

Auditing of independent observers is advisable to assess the prevalence their errors.

Errors are biased when their average is not 0, i.e. the reported values are more likely to be above and more likely to be below the true values. Intentional errors will typically be biased.

Errors will usually have some variability which will contribute to the variability of the estimation process. For sample surveys, this contribution will be partially accounted for in the SE. See § 13.6 for details.

Contribution to estimation process bias: Biased errors will to contribute to the bias of the estimation process.

Contribution to estimation process variability: Variability of the errors in independent observer's data contributes to the variability of the estimation process. However, the SE partly accounts for this contribution, except for censuses.

EXAMPLE: At-sea observers misjudge quantities that are assessed visually, leading to unintentional, possibly unbiased, errors.

EXAMPLE: Systematic misidentification of a species in video monitoring, leading to unintentional, possibly biased, errors.

EXAMPLE: Recreational fishing effort measured using aerial counts may unintentionally misclassify larger pleasure vessels as recreational fishing vessels resulting in an overestimate of fishing effort for larger vessels.

EXAMPLE: Collusion between an at-sea observer and fishers may involve under-reporting of catches of protected species that may trigger a fishery closure, resulting in intentional bias in the bycatch estimates.

Note: Errors from resource users vs errors from independent observers
Errors in data reported by resource users and in data reported by independent observers have similar impacts on the estimation process. However, we separate them in the assessment because correcting these errors will depend on their sources. For example, DFO can easily impose further training to independent observers but not to resource users; independent observers committing intentional errors can be more easily penalized than resource users.

6.2.11. Equipment error

Error due to a measuring tool inaccuracy and/or imprecision.

Errors are biased when their average is not 0, i.e. the values reported by the equipment are more likely to be above and more likely to be below the true values.

Contribution to estimation process bias: Biased errors will to contribute to the bias of the estimation process.

Contribution to estimation process variability: In sample surveys, the contribution is included in the computed standard error. Reduction of the measuring tool error will lead to a smaller standard error and, therefore, a lower variability of the estimation process. In censuses, the contribution may create variability if the measurement errors are numerous and large.

EXAMPLE: Recurring weigh-scale calibration error leading to a bias.

EXAMPLE: Measurement errors because scales only measure to nearest kg results in an unbiased equipment error.

6.2.12. Data handling error

Data handling refers to all data manipulation steps occurring ***after*** the initial recording of the observations. A data handling error occurs when a data manipulation creates an error.

Auditing can be used to assess the prevalence of data handling errors.

Contribution to estimation process bias: Biased errors will to contribute to the bias of the estimation process.

Contribution to estimation process variability: In a sample survey, the contribution is included in the computed standard error. Reduction of the data handling errors will lead to a smaller standard error and, therefore, a lower variability of the estimation process. In censuses, it may create variability if the response errors are numerous and large.

EXAMPLE: Data entry error; data copy error.

6.2.13. Adjustment error

Adjustments are required when observations are obtained using different methods. Such adjustments may depend on models, experimental results, etc.

An adjustment error occurs when an adjustment yields incorrect values.

Contribution to estimation process bias: May lead to bias if the equation or routine used to adjust the data is not accurate.

Contribution to estimation process variability: None expected.

EXAMPLE: Incorrect weight adjustment for landings where catch weight is corrected due to presence of ice or for cleaned vs whole fish.

Technical note: Adjustment error and variability
Adjustments are anticipated to be have been carried using tables or formulae. The SE will typically be computed after the adjustment. A mistake in the tables or formulae will not contribute to further variability.

6.2.14. Imputation error

An imputation is the replacement of a missing value by a value obtained (imputed) from other information available, including the values for spatially and temporally adjacent or otherwise similar observed units. Imputation is a valuable tool in sample surveys and censuses but it depends on the availability of auxiliary data and on a statistical model to compute the imputed values. If a large number of values are imputed, in a sample survey, the computation of the SE will normally take into account the uncertainty due to the imputation; in a census, this uncertainty should be computed.

Imputation error occurs when an imputation yields a value different from the true value.

Contribution to estimation process bias: If imputed values are, on average, smaller/greater than the true values that would otherwise have been observed, the imputation will lead to a negative/positive bias.

Contribution to estimation process variability: The imputation itself may introduce errors in the estimation process. These errors may or may not have been integrated in the computation of the SE.

EXAMPLE: Using CPUE from neighboring rivers or from similar time periods to estimate catch from effort for a river for which CPUE values are not available.

EXAMPLE: In a census of landings, missing values are imputed by the mean of the observed landings.

EXAMPLE: In a census of landings where the durations of the trips are known, missing values are imputed by computing the CPUE of the observed landings and applying it to the other trips.

6.2.15. Modelling error

Some statistical estimators are based on a statistical model, which can fit the data more or less well.

Diagnostic tools are available to verify a model's goodness of fit (e.g., cross-validation). Such diagnostic tools should be applied each time the model is reused. If lack of fit is present, simulations can be carried out to measure the impact of the lack of fit on the estimation process.

A modelling error occurs when the assumed statistical model does not fit the data well.

Contribution to estimation process bias: Possible impact on bias.

Contribution to estimation process variability: Possible impact on variability.

EXAMPLE: Estimating the CPUE by the ratio estimator (i.e. [total sample catch]/[total sample effort]) assumes that the catch is proportional to effort (i.e. the model is catch = CPUE × effort + residual). If the CPUE depends on effort (e.g. if very active recreational fishers are more skilled and have higher individual CPUE), the ratio estimator will give a biased estimate due to an incorrect model.

EXAMPLE: (Technical) If a linear regression model is applied, the quotient of the sum of squared residuals of a locally weighted smoothing (LOESS) over the sum of squared residuals of the linear model can be used as measure of goodness of fit and of non-random departures from the model.

7. ASSESSING OVERALL QUALITY

The assessment of the overall combined impact of operational characteristics will be obtained in the following way.

7.1. Assessing the impact of operational characteristics on bias

Given that individual biases are additive, we propose to compute the impact of each characteristic on the estimation process bias by combining the prevalence (proportion of observations affected) and impact (magnitude of bias when an observation is affected) in the following way:

- Let p be the proportion of targeted observations impacted for an operational characteristic that affects bias (e.g. biased response, missing value, etc.).
- Let δ be the anticipated relative difference between the average of the values for the impacted observations and the average of the unimpacted values. For example, in the case of characteristics affecting under- or over- coverage it is the difference between the average of impacted observations and the average of the observed values.

Then, the impact on relative bias is $p \times (1 + \delta) + (1 - p) \times 1 - 1 = p\delta$.

In the case that there are individual proportions p_i associated with different anticipated biases δ_i the impact on relative bias for the characteristic will be $\sum_i p_i \delta_i$.

Knowledge of the proportion of targeted observations impacted is useful. However, if the overall impact has already been computed in an external study, the above formula applies with $p = 1$, i.e. the impact is δ.

7.2. Assessing the impact of operational characteristics on variability

The impact of operational characteristics on variability will be assessed using computations specific to the situation. The result is a measure of variability that reduces to the standard error if there are no impacts from operational characteristics. We sometimes refer to this measure as a “heuristic standard error”.

7.3. Estimation process bias

We sometimes refer to the estimation process bias as the “pseudo-bias”.

The estimator bias and the contribution of the 15 operational characteristics to relative bias will be added, taking into account the sign of the bias.

b_{ep} = estimator bias + sum of contribution to bias from the operational characteristics.

7.4. Estimation process variability

We sometimes refer to the estimation process variability as the “pseudo-SE”.

7.4.1. Sample surveys: Impacts on variability unaccounted for in the computed SE

Impacts of operational characteristics representing departures from the sampling protocol will typically be expressed as a proportion of the SE. The corresponding correction to the SE will be multiplicative. If no other contribution is present, for a single operational characteristic, the computation will be as follows:

$$s_{ep} = \text{SE} \times [\text{multiplicative contribution to variability}].$$

Computation of the multiplicative factors will depend on the departure.

Impacts of operational characteristics representing added variability in the data will be added quadratically to the SE. If no other contribution is present, for a single operational characteristic, the computation will be as follows:

$$s_{ep} = \sqrt{\text{SE}^2 + [\text{additive contribution to variability}]^2}.$$

However, care must be taken since the SE may already take into account some of the added variability (see § 7.4.2).

7.4.2. Sample surveys: Impacts on variability already accounted for in the computed SE

In general, these impacts will be due to random errors. The combined contribution of these characteristics to the overall variability will be the quadratic sum of each contribution:

$$\sqrt{\sum [relative\ contribution\ to\ the\ error]^2)}$$

See § 13.6 for further details.

7.4.3. Censuses: Impacts on variability unaccounted for in the computed SE

In general, these impacts will be due to random errors. The combined contribution of these characteristics to the overall variability will be the quadratic sum of each contribution:

$$\sqrt{\sum [relative\ contribution\ to\ the\ error]^2)}$$

See § 13.6 for further details.

7.5. Estimation process error

We sometimes refer to the Estimation process error as the “pseudo-RMSE”.

The estimation process error will be obtained by the following formula, corresponding to the relationship between the root mean square error and the estimation process variability and the estimation process bias:

$$e_{ep} = \sqrt{(estimation\ process\ bias)^2 + (estimation\ process\ variability)^2}$$

7.6. Anticipated true value

We use the following symbols:

θ : True value of the parameter

$\theta_{anticipated}$: Anticipated true value of the estimator

$\hat{\theta}$: An estimate of the parameter obtained from the monitoring program (most recent value or median observed catch over the last 3, 5 or 10 years)

The anticipated true value is obtained from an estimation value by subtracting the bias:

$$\theta_{anticipated} = \hat{\theta} - b_{ep}$$

7.7. Summaries of the quality of an estimation process

The relative estimation process bias, variability and error are obtained by taking the quotient by the anticipated true value:

$$rb_{ep} = \frac{b_{ep}}{\theta_{anticipated}}$$

$$rs_{ep} = \frac{s_{ep}}{\theta_{anticipated}}$$

$$re_{ep} = \frac{e_{ep}}{\theta_{anticipated}}$$

The relative values will usually be reported as percentages.

The last value summarizes the quality of an estimation process. However, given the different nature of bias and variability, all 3 values should be reported.

7.8. Assessment based on several monitoring programs

There are two general cases in which several monitoring programs may be involved in deriving a final parameter of interest.

When parameters from several estimation processes are summed to estimate a final parameter of interest such as total catch, effort, etc., the assessment of the estimation process for the parameter of interest is obtained as follows:

Assuming that the total is obtained as a weighted sum of the individual estimates:

- The bias of the overall estimation process for the total is the sum of individual estimation process biases.
- The variability of the overall estimation process for the total is the square root of the sum of the squared variability of the individual estimation process variability.

Weights may be required to reflect the relative contribution of each estimation process to estimating the parameter of interest. For example, if total catch is estimated from landings, which are estimated from dock-side observations, and discards, estimated from an at-sea observer survey, the proportion of total catch represented by each source could be a weight.

The values e_{ep}, rb_{ep}, rs_{ep}, and re_{ep} are then computed.

In contrast, when a final parameter of interest is a product of two or more estimation processes, such as the estimation of catch from separate monitoring programs that estimate the fishing effort and the CPUE, then the assessment of the estimation process for the parameter of interest can be obtained heuristically by applying the formulae for the bias and variance of a product of two independent random variables:

If $E(X) = \mu_X + Bias_X$ and $E(Y) = \mu_Y + Bias_Y$, then

$$E(XY) = E(Y)E(Y) = (\mu_X + Bias_X)(\mu_Y + Bias_Y)$$
$$= \mu_X\mu_Y + Bias_X\mu_Y + \mu_X Bias_Y + Bias_X Bias_Y)$$
$$Var(XY) = Var(X)Var(Y) + Var(X)E(Y)^2 + Var(Y)E(X)^2$$

8. ASSESSING DEPENDABILITY

8.1. Common events, rare events, special cases

By common events, we mean events that will occur on most 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. In these cases, we can presume that the central limit theorem applies to the estimation process. These are the most frequent cases. We describe a detailed approach to assessing dependability in these cases.

By rare events, we mean events that will occur only infrequently and that will typically involve small counts or amounts. In these cases, we cannot presume that the central limit theorem applies to the estimation process. We describe an approach to assessing dependability in rare-events cases where the objective is estimating a small count and a Poisson distribution is applicable (i.e. without large under- or over-dispersion).

Special cases include, for example, situations where the slope of a linear regression must be estimated. This would be the case if the compliance limit is a limit on the ratio of total catch weight of a particular bycatch species and the target total catch weight. While the principles presented for common events and rare events apply to these situations, the mathematical details will depend on each situation.

8.2. Assessment of an estimation process for estimation applications: common event case

For an estimation application, the user should determine a required quality level. We propose that the user set this quality level using the largest acceptable relative root mean square error of the estimate, which we denote $rrmse_{max}$. Then, the dependability of the estimation process will be the quotient:

$$\frac{rrmse_{max}}{re_{ep}}$$

A score ≥1.0 clearly indicates that a program meets its dependability requirements. However, in some applications, managers may choose to accept values that are below 1 if, for example, fishing activities pose little risk of harm to resources. In such instance it might be instructive to derive categories of quotient scores that might scale with risk tolerance, for example. The following scoring scale is selected in such a way that the square of the threshold is approximately halved at each step, i.e. the MSE as oppose to the RMSE approximately halved at each step.

Score	A	B	C	D	E
Quotient	≥ 1.000	1.00 to ≥ 0.700	0.700 to ≥ 0.500	0.500 to ≥ 0.350	Less than 0.350

8.3. Assessment of an estimation process for estimation applications: A rare event case

Consider an event (for example, catch of a large mammal) that is anticipated to occur randomly at a small rate p occurrence(s)/observation (for example, on average, one large mammal is caught every $1/p$ trips). The user requires an estimate $\hat{p}$ of p. In the rare event case, the sampling distribution of $\hat{p}$ may be highly asymmetrical and, therefore, the $RMSE$ is not a suitable description of the error.

In such a situation, we recommend that the user should determine a precision level by setting the largest acceptable relative bias rb_{max} and left-side and right-side length, $rlength_{left,max}$ and $rlength_{right,max}$, for the confidence interval for p with a confidence level 0.683. The choice of 0.683 insures compatibility with the common event case described in § 8.2 (see technical note below).

The left- and right- side largest acceptable estimation process error (pseudo-RMSE) is calculated as:8.2

$$rrmse_{max,left} = \sqrt{rlength_{left,max}{}^2 + rb_{max}{}^2} \text{ and } rrmse_{max,right} = \sqrt{rlength_{right,max}{}^2 + rb_{max}{}^2}$$

Supposing that the anticipated value of the parameter is $p_{anticipated}$ and that the sample size is n_{trips} trips, then, the total number x of events observed will follow a Poisson distribution with parameter $\lambda_{anticipated} = n_{trips} \times p_{anticipated}$. Let $[ci_{left}\ ;\ ci_{right}]$ be the confidence interval for λ with confidence level 0.683, if $\lambda_{anticipated}$ occurrences are observed. The relative left-side and right-side confidence interval lengths are $(\lambda_{anticipated} - ci_{left})/\lambda_{anticipated}$ and $(ci_{right} - \lambda_{anticipated})/\lambda_{anticipated}$, respectively. Let

$$rep_{left} = \sqrt{[(\lambda_{anticipated} - ci_{left})/\lambda_{anticipated}]^2 + [rb_{ep}]^2}$$

$$rep_{right} = \sqrt{[(\lambda_{anticipated} - ci_{right})/\lambda_{anticipated}]^2 + [rb_{ep}]^2}$$

Then, the dependability of the estimation process will be the smallest quotient:

$$min\left[\frac{rrmse_{max,left}}{rep_{left}}, \frac{rrmse_{max,right}}{rep_{right}}\right]$$

We propose the same scoring scale as in § 8.2.

Score	A	B	C	D	E
Quotient	≥ 1.000	1.00 to ≥ 0.700	0.700 to ≥ 0.500	0.500 to ≥ 0.350	Less than 0.350

Example

Suppose that scientists require that the rate of occurrence of an event be estimated using the largest acceptable absolute relative bias $rb_{max} = 0.1$ and left-side and right-side length, $rlength_{left,max} = 0.5$ and $rlength_{right,max} = 0.5$, for the confidence interval for p with a confidence level $1 - \alpha = 0.683$. Then, $rrmse_{max,left} = rrmse_{max,right} = 0.510$

Suppose that $p_{anticipated} = 0.2$, $b_{ep} = -0.05$ (i.e. 25% of the occurrences are anticipated to be unreported). Suppose also that a sample size of $n_{trips} = 15$ is planned.

Then the anticipated number of occurrences is $\lambda_{anticipated} = 3$ and the confidence interval for λ with confidence level 0.683, if $\lambda_{anticipated} = 3$ occurrences are observed, is $[1.37\ ;\ 5.92]$ and we obtain $rep_{left} = 0.60$ and $rep_{right} = 1.00$.

Then, the dependability of the estimation process is $min\left(\frac{0.51}{0.60}, \frac{0.51}{1.00}\right) = min(0.85\ , 0.51) = 0.51$ and the dependability score is “B”.

In this example, a sample size of $n_{trips} = 48$ would be required to obtain a dependability score of “A”.

Technical note
Confidence intervals for the parameter of a Poisson distribution can only be computed for integer observations. Therefore, $[ci_{left}\ ;\ ci_{right}]$ must be obtained by interpolation if the anticipated value is not an integer. The user may prefer to set the non-relative $length_{left,max}$ and $length_{right,max}$ first. Setting these two values equal ensures that the procedures for this rare event case converges to the procedure for common events described in § 8.2, when $p_{anticipated}$ and/or n_{trips} become large. Because the sum of independent Poisson random variables is also a Poisson random variable, the above procedure applies to more complex cases including, for example, a situation where several fleets with different values of $p_{anticipated}$ are involved.

Technical note: the choice of 0.683
The following explains the choice of the 0.683 confidence level. In the common event case (§ 8.2), if we neglect bias, the dependability of an estimation process for estimation application is measured by the quotient $s_{ep,max}/s_{ep}$ of the required error over the assessed estimation process error (using the relative values gives the same quotient). For a parameter θ where the normal distribution is an acceptable model (for example, the estimation of the mean with 30 observations from a symmetrical distribution), the confidence interval with confidence level 0.683 is given by $\hat{\theta} \pm z_{\alpha/2}\sigma_{\hat{\theta}} = \hat{\theta} \pm \sigma_{\hat{\theta}}$, since $z_{0.683/2} = 1$. Therefore, the left-side or the right-side length of the confidence interval are each equal to $\sigma_{\hat{\theta}}$. This shows that the procedure above generalizes that of § 8.2 if we use the confidence interval with confidence level 0.683. ***Example*** (continued) In the example above, $\sigma_{\hat{\lambda}} = 1.73$, suggesting incorrectly that the variability of the estimator spreads equally below and above the true value $\lambda = 3$, while the confidence interval if $\hat{\lambda} = 3$ is observed, $[1.37\ ;\ 5.92]$ is approximately twice larger above λ than below.

8.4. Assessment of an estimation process used to verify compliance with a limit: common event case

We propose to assess the dependability of an estimation process for compliance applications by examining the ability of the estimation process to prove that an upper limit has not been reached or breached. ***(As indicated previously, the inequalities in the equations that follow would have to be reversed in the case of a lower limit)***. An objective of this approach is that fisheries for which the catch is expected to be close to the total allowable catch will require better quality estimates (e.g. larger sample sizes) than fisheries for which the catch is expected to be far below this limit.

Figure 4 illustrates several combinations of accuracy, precision and relationships between the true value and the compliance limit. Since monitoring programs must be assessed for their value over time, when assessing an estimation process, the “true value” will be replaced by the “anticipated” value of the parameter.

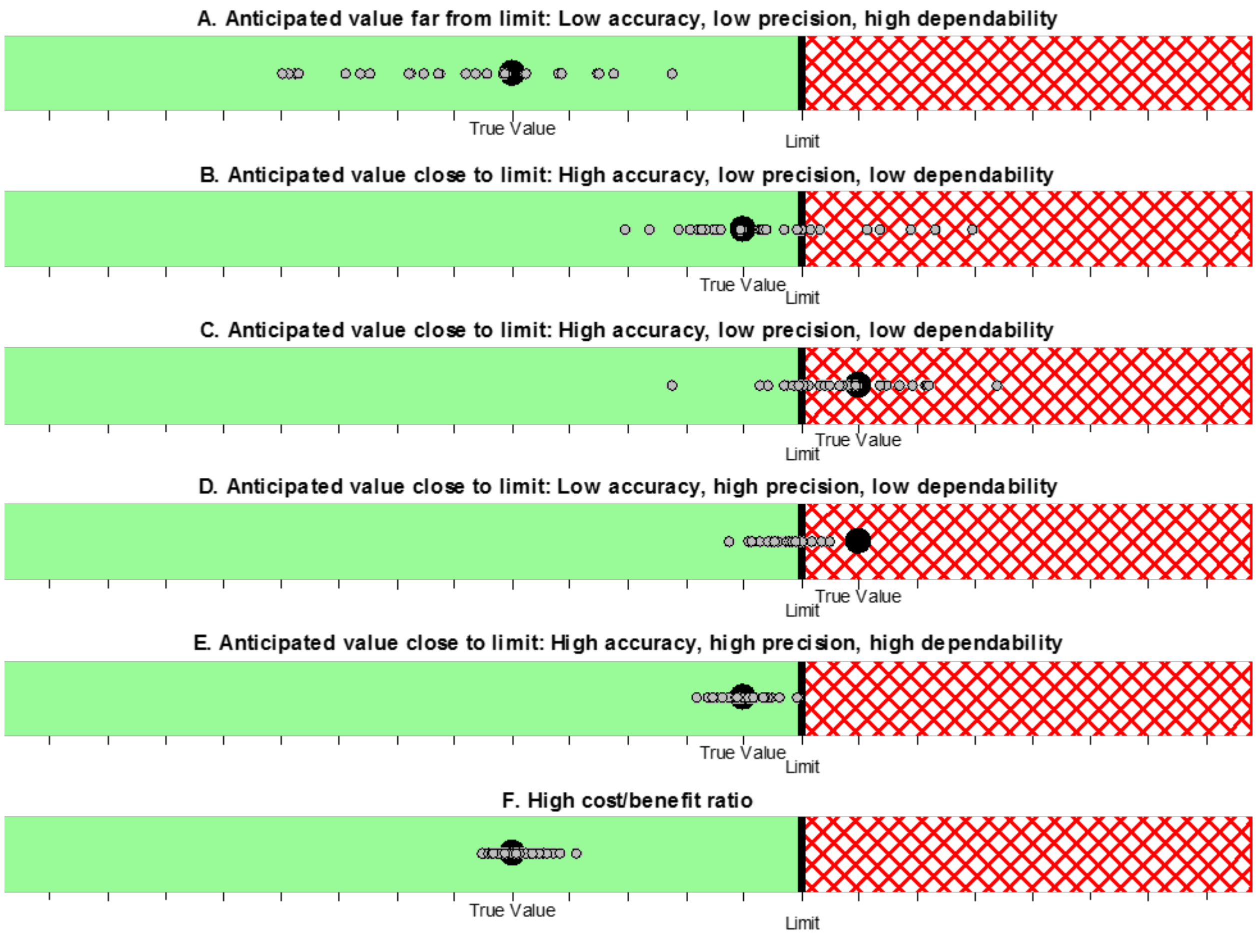


Figure 4. Dependability of an estimation process depends on its accuracy and precision and on how close the true value of the parameter (solid dot) is to the limit (vertical line). The open dots represent examples of estimated values from theoretical repetitions of the estimation process. The graphics illustrates an upper compliance limit, e.g. a fishery's Total Allowable Catch (TAC). If the true value is far below the limit, a low accuracy, low precision (A) estimation process is dependable. If the true value is close to the TAC, a low precision (B, C) or low accuracy (D) estimation process is not dependable but a high accuracy, high precision (E) estimation process is dependable. If the true value is far from the compliance limit, a high quality (F) estimation process suggests that cost savings are possible.

In the following, we suppose that the limit is an upper limit on a quantity (e.g. a total allowable catch). The case for a limit based on a proportion is similar for limits that are away from 0% or 100%. Different formulae are required for limits on proportions close to 0% or 100% or for limits on rare events. See § 8.5 for a computation on a case of rare events.

Heuristically, this approach is based on a statistical test to ***reject the hypothesis that the limit has been reached or breached***, assuming that the error on the estimate follows normal distributions. The concepts of statistical tests of significance are only used as a guide. The development does depend on verifying the assumption required to apply the significance tests (e.g. normality of the estimator). Special cases for other distributions, such as the Poisson, could be equivalently elaborated, as discussed below for rare event cases.

Note: Terminology
On occasion, we use the following vocabulary based on detection of non-compliance: A "false positive" refers to concluding that the limit has been exceed when in fact it was not. A "false negative" refers to concluding that the limit has been respected (not exceeded) when in fact it was exceeded. The terminology is similar to the terminology used in medical diagnostics.

We use the following notation:

L : Upper limit (e.g. TAC)

θ : True value of the parameter

$\theta_{anticipated}$: Anticipated true value of the estimator (defined in § 7.6).

$\hat{\theta}$: Estimate of the parameter, obtained from the estimation process

$\sigma_{\hat{\theta}}$: The standard error of the estimator

$RSE = RSE_{\hat{\theta}} = \sigma_{\hat{\theta}}/\theta$: Relative standard error of the estimator

φ : Cumulative distribution function of the standard normal distribution

$0 < \alpha < 1$: A pre-selected value to be thought of as a significance level

$z_\alpha = \varphi^{-1}(\alpha)$

Consider the one-sided statistical hypothesis test H_0: $\theta \geq L$ vs H_1: $\theta < L$ with a predetermined significance level α. The limit will be considered satisfied if the null hypothesis H_0 is ***rejected***, i.e. if $\hat{\theta} < L + z_\alpha \sigma_{\hat{\theta}}$. Under normality assumption, the power of the test at θ is $\varphi\left(\frac{L-\theta}{\sigma_{\hat{\theta}}} + z_\alpha\right) = \varphi\left(\frac{L/\theta - 1}{\sigma_{\hat{\theta}}/\theta} + z_\alpha\right)$.

The predetermined significance level of the test, α, is the probability of concluding, incorrectly, that the limit was not reached when it was reached exactly. If $\alpha = 0.50$, $z_\alpha = 0$ and the conclusion is based on the simple comparison between $\hat{\theta}$ and L, the usual approach to monitoring for compliance. Choosing a small value for α (e.g. $\alpha = 0.05$) corresponds to a precautionary or risk-averse approach.

Technical notes
1. The "false negative" and "false positive" defined above are opposite to the same notions typically associated to the test H_0: $\theta \geq L$. "False negative" and "false positive" for the non-compliance application corresponds to a Type I error and Type II error, respectively, for the statistical test H_0: $\theta \geq L$. Because the heuristic approach is based on a test of the null hypothesis of ***non***-compliance, the association is reversed. 2. The equality is included in the null hypothesis H_0: $\theta \geq L$ for technical reasons related to the theory of statistical test. The mathematical probability that the limit L be reached exactly is zero. Therefore, the fact that $\theta = L$ is mathematically considered unacceptable is immaterial in practice.

We propose the following measure of dependability of an estimation process for compliance applications:

The probability that the statistical test H_0: $\theta \geq L$ yields the correct conclusion, if $\theta = \theta_{anticipated} = \hat{\theta} - b_{ep}$ and $\sigma_{\hat{\theta}} = s_{ep}$.

If $\theta_{anticipated} < L$, the measure of dependability equals

$$[the\ power\ of\ the\ test\ at\ \theta = \ \theta_{anticipated}] = \varphi\left(\frac{L-\theta}{\sigma_{\hat{\theta}}} + z_\alpha\right).$$

If $\theta_{anticipated} \geq L$, the measure of dependability equals

$$1 - \left[the\ power\ of\ the\ test\ at\ \theta = \ \theta_{anticipated}\right] = 1 - \ \varphi\left(\frac{L-\theta}{\sigma_{\hat{\theta}}} + z_\alpha\right).$$

The probability will usually be reported as a percentage.

At this point in time, $\alpha = 0.50$ and, therefore $z_\alpha = 0$, the current practice, should be retained. If a more precautionary approach is desired, a smaller value of α should be adopted.

The behaviour of the proposed dependability measure relative to the true value of the parameter is illustrated in Figure 5. In all cases, the estimation process is considered dependable if the true value is far from the limit. If the bias of the estimation process is 0, the quality of the estimator varies symmetrically for true values on either side of the limit. A lower precision corresponds to a wider band of low dependability around the limit. If the bias of the estimation process is negative, the dependability of the estimation process will be low for true values just above the limit: the negative bias will hide the breach of the compliance limit.

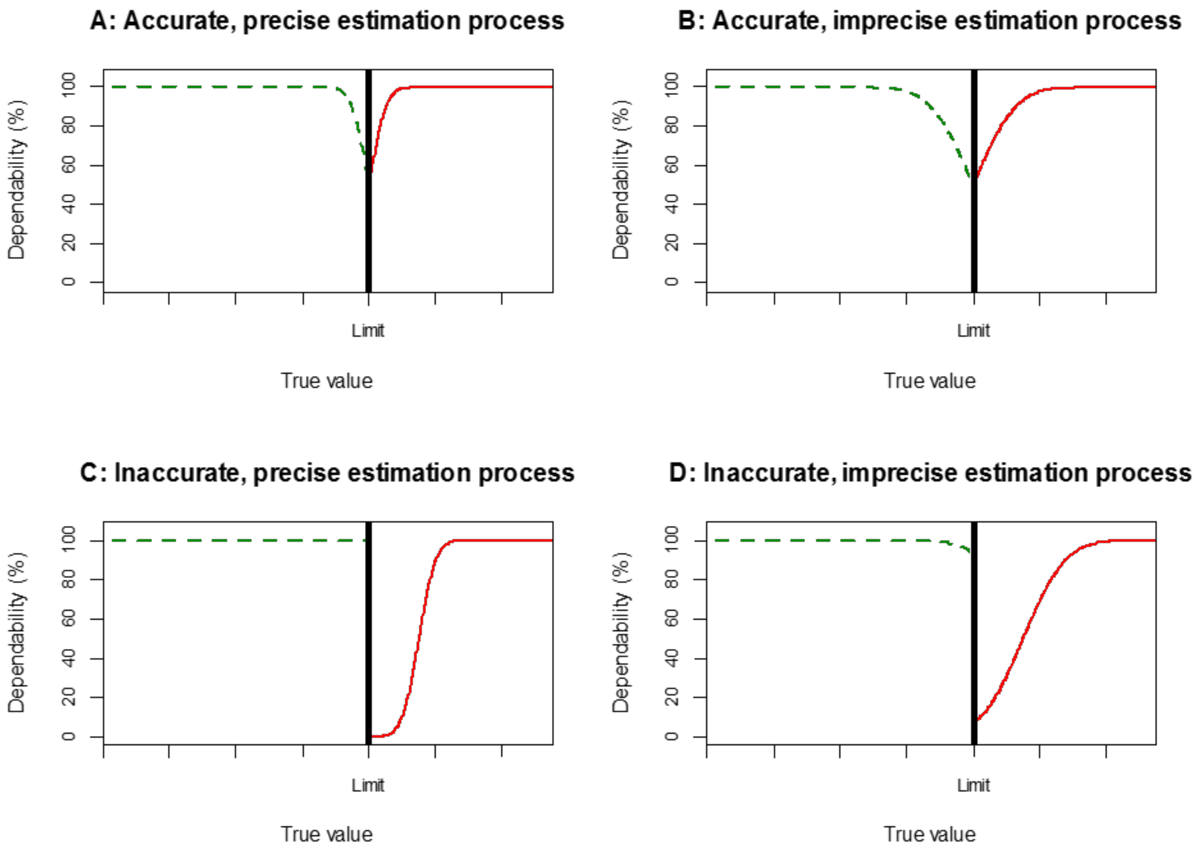


Figure 5. Dependence of the dependability score on the true value of a parameter, the limit, the accuracy and the precision of the estimation process for α= 0.50, the choice corresponding to the usual practice. The thick vertical line shows the limit. The thin lines show the probability of drawing the correct conclusion: that the limit has been respected (green dashed) or not (red solid). (A) Without bias and with low variability, the probability of a correct conclusion is high except if the true value is very close to the limit. (B) Without bias but with high variability, the probability of a correct conclusion is low further away from the limit. (C) With a negative bias and low variability, there is an extremely low probability of drawing an incorrect conclusion for true values just above the limit. (D) A negative bias combined with high variability worsens the situation.

To be consistent with the use of precaution when managing fishery removals, we discuss briefly the impact of using a smaller value of α, illustrated in Figure 6, which is directly comparable to Figure 5. In the context presented, α represents the probability of incorrectly concluding that the limit is respected when it is not. Therefore, choosing a smaller value for α reduces this probability. However, it also increases the probability of incorrectly concluding that the limit is not respected when in fact it is.

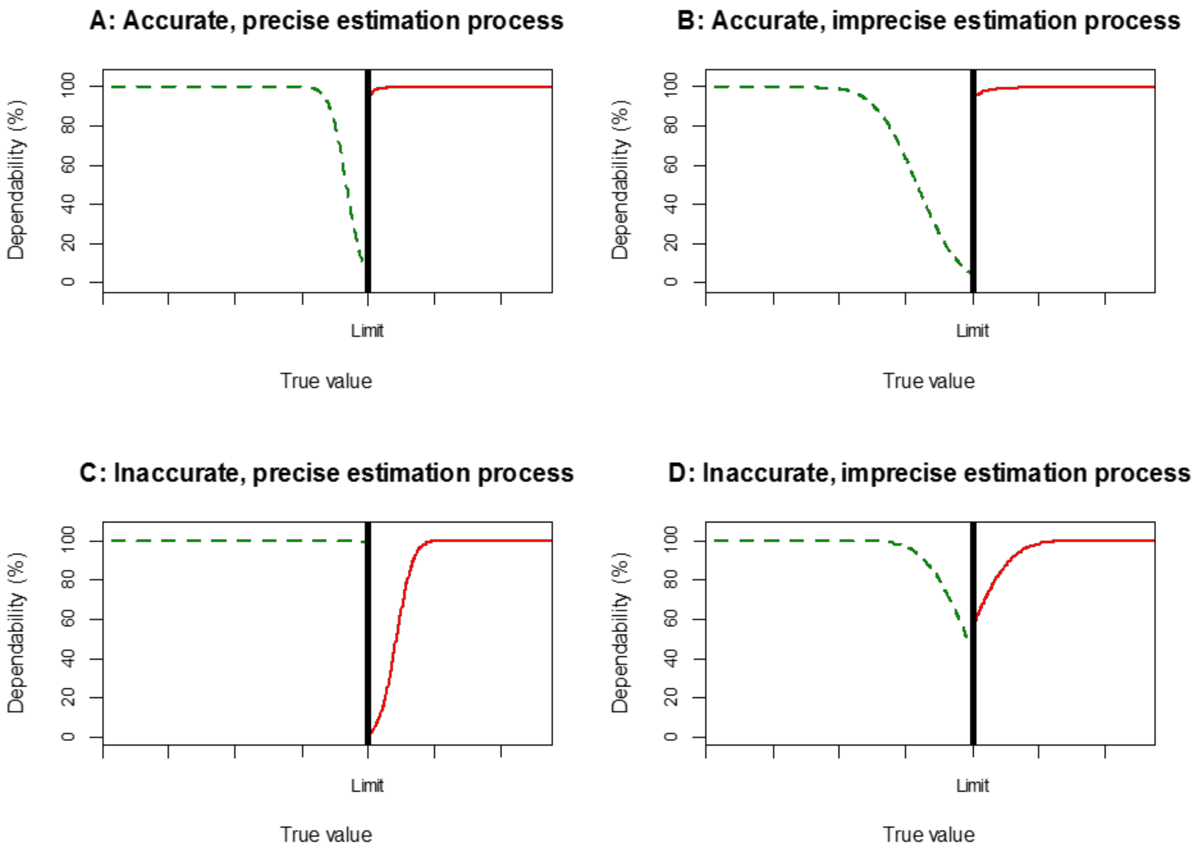


Figure 6. Dependence of the dependability score on the true value of a parameter, the limit, the accuracy and the precision of the estimation process for α= 0.05, a precautionary choice. Comparing with the previous figure (Figure 5), a smaller value for α reduces this probability of incorrectly concluding that the limit is respected if it is not in all cases (unless it was already 100%). Correspondingly, it increases the probability of incorrectly concluding that the limit is not respected when it is.

We propose the following scoring scale.

Score	A	B	C	D	E
Dependability	≥ 90%	75% to < 90%	60% to < 75%	50% to < 60%	< 50%

Figure 7 illustrates the scoring scale for various combination of the true value of a parameter, the limit, the accuracy and the precision of the estimation process for α = 0.50, the current practice.

Figure 7. Dependability score for several combinations of the true value of the parameter (solid dot), limit (vertical line), accuracy and precision of the estimation process. The open dots represent examples of estimated values from theoretical repetitions of the estimation process, illustrating the accuracy and the precision of the estimation process. The graphics illustrates an upper limit, e.g. a fishery's Total Allowable Catch (TAC). If the true value is very far from the limit, a low accuracy, low precision (A) estimation process is dependable (dependability score A). If the true value is close to the TAC, a high precision (B) estimation process is dependable (score A). Low precision (C, D) or low accuracy (E) estimation processes are not dependable if the true value is close to the limit (scores C, D and E, respectively).Assessment of an estimation process used to verify compliance with a limit: A rare event case

8.5. Assessment of an estimation process used to verify compliance with a limit: A rare event case

Consider a compliance limit stating that an event (e.g. catch of a large mammal) should occur at no more than a certain rate L_p (e.g. the event can occur no more than once every x trips).

The relevant statistical hypotheses are now $H_0: p \geq L_p$ vs $H_1: p < L_p$.

Supposing that the events occur randomly, the total number of events observed over n_{trips} trips would follow a Poisson distribution with parameter $\lambda = n_{trips} \times p$. The test above becomes the test $H_0: \lambda \geq \lambda_0 = n_{trips} \times L_p$ vs $H_0: \lambda < \lambda_0 = n_{trips} \times L_p$ on the parameter λ of the Poisson distribution. For small values of $\lambda = n_{trips} \times L_p$ and a fixed α, a Poisson exact test exists and can be used to determine the rejection limit $R_x = R_{x,\alpha}$.

Suppose that the typical number of trips is n_{trips} and the anticipated rate $p_{anticipated} = \hat{p} - b_{ep}$. Then, the anticipated number of occurrences is $n_{trips} \times p_{anticipated}$ and the power of the test is $\Pr(X \leq R_x | X \sim Poisson(n_{trips} \times p_{anticipated}))$. Then, the dependability of the estimation process is given by:

If $p_{anticipated} < L_p$, the measure of dependability equals $[the\ power\ of\ the\ test\ at\ p = p_{anticipated}]$

If $p_{anticipated} \geq L_p$, the measure of dependability equals $[1 - the\ power\ of\ the\ test\ at\ p = p_{anticipated}]$.

Since a sum of Poisson distribution is again a Poisson distribution, the proposed method applies to a situation where the limit is proportional to effort, for example.

EXAMPLE

Consider a compliance limit stated as "at most 1 occurrence every 5 trips", i.e. $L_p = 1/5 = 0.2$

Suppose that:

The anticipated sample size is 50 trips

25% of the occurrences are anticipated to be unreported (i.e. $b_{ep} = -0.05$)

$\alpha = 0.25$, a mildly precautionary value has been retained

The compliance limit applied to the sample trips is $50 \times 0.2 = 10$.

The dependability of the estimation process for selected anticipated true values is shown in Figure 8.

The low dependability for anticipated values immediately above the compliance limit is due to the small sample size, the negative bias and the precautionary choice of α.

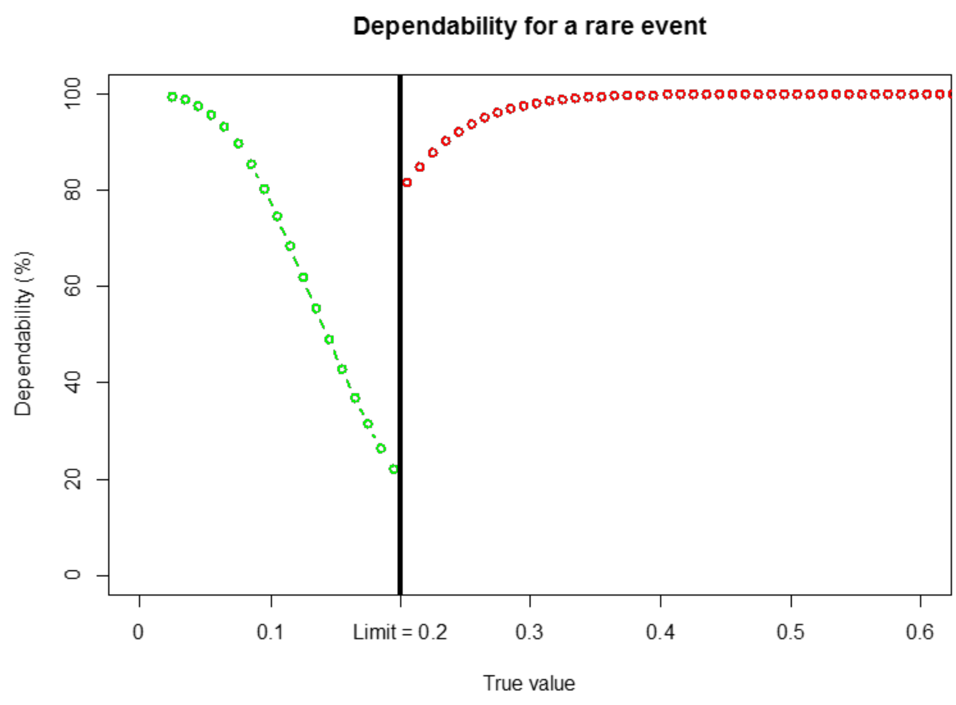


Figure 8. Dependability for a rare event where the estimation process is negatively biased and a mildly precautionary value of α has been selected. The compliance limit states that "at most 1 occurrence every 5 trips" are allowed.

8.6. Assessment of an estimation process used to verify compliance with a limit: uncertainty on the limit

Uncertainty around a compliance limit can come from several sources. For example, if a catch upper limit is a function of stock biomass (e.g. the $F_{0.1}$ rule), the uncertainty on the biomass estimate will create uncertainty on the catch upper limit. When several models or only expert opinions are available to set a limit, meta-analysis techniques can be applied to quantify uncertainty on the limit.

While compliance limits should be fixed values for enforcement practicalities, in environmental or biological risk management, the uncertainty on the limits will have an impact on decision making.

We note that the methodology proposed to assess the dependability of an estimation process can be easily modified to take into account uncertainty about a limit.

It suffices to consider heuristically the one-sided statistical hypothesis test H_0: $\theta \geq L$ vs H_1: $\theta < L$ under the assumption that the estimator $\hat{\theta}$ of θ has standard error $\sigma_{\hat{\theta}} \approx s_{\hat{\theta}}$ and that the estimator $\hat{L}$ of L has standard error $\sigma_{\hat{L}} \approx s_{\hat{L}}$, where the standard error on $\hat{L}$ may be obtained, for example, from a combining of opinion method. Then, the standard error $\sigma_{\hat{\theta}-\hat{L}}$ of $\hat{\theta} - \hat{L}$ can be estimated as $s_{\hat{\theta}-\hat{L}} = \sqrt{{s_{ep}}^2 + \ {s_{\hat{L}}}^2}$.

The definition of dependability presented in § 8.4 holds after replacing $\varphi\left(\frac{L-\theta}{\sigma_{\hat{\theta}}} + z_\alpha\right)$ by $\varphi\left(\frac{L-\theta}{\sigma_{\hat{\theta}-\hat{L}}} + z_\alpha\right)$.

Figure 9 illustrates the impact of uncertainty on the limit. In this example, the relative standard error on the parameter is 4% (precise case) or 10% (imprecise case) if the parameter equals the limit and the relative error on the limit is 25%. Comparing Figure 8 to Figure 5, we observe that the dependability is lower when there is uncertainty on the limit, for most true values of the parameter.

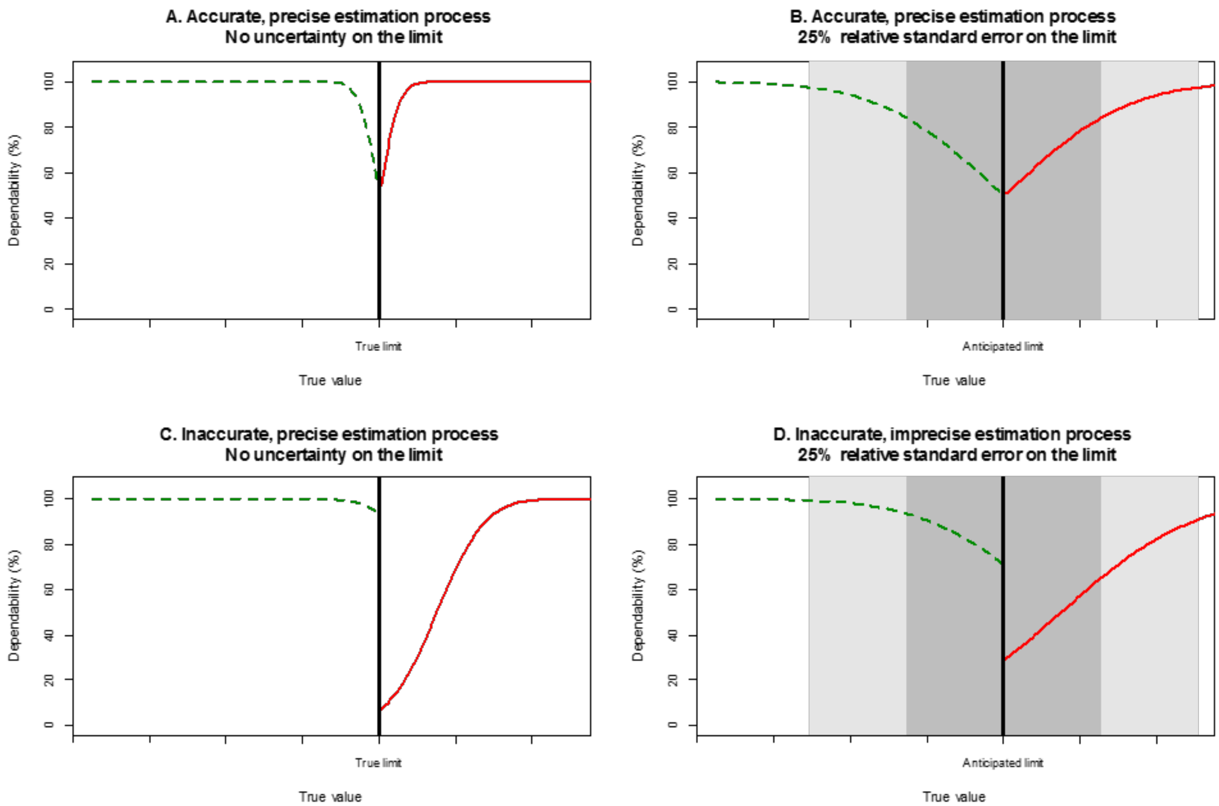


Figure 9. Dependence of the dependability score on the true value of a parameter, the limit, the accuracy and the precision of the estimation process for α= 0.50, the choice corresponding to the usual practice, without (left hand side) and with (right hand side) uncertainty on the limit. The grey bands show the uncertainty (68% and 95% C.I.) on the limit. See text for details.

9. CASE STUDY: SKATE DISCARDS IN THE SOUTHERN GULF OF ST. LAWRENCE

9.1. Description

Benoît (2013) estimated the landings and discards of three skate species in southern Gulf of St. Lawrence fisheries for the period 1991-2011. Here we consider only the total discards of skate over that period.

Total annual skate discards were estimated using a ratio estimator (Benoît 2013). First, the mean catch ratio of skates (kg skate/kg targeted species) was estimated for each fishery, defined by the target species and the gear class used (fixed or mobile gear), and year using the data collected by at-sea observers. Second, the catch ratios were multiplied by the total target species landings in each fishery and year, which were obtained from dockside monitoring for the majority of cases and from purchase slips for the remainder.

Therefore, assessing the quality and dependability of the annual skate discard estimates required consideration of operational characteristics affecting the at-sea observer program and those affecting the dockside monitoring program for the southern Gulf of St. Lawrence. (For simplicity, the contribution of characteristics affecting purchase slips was not considered).

Observations of skate discards are only available from the at-sea observer program which covers the groundfish and shrimp fisheries in the area. While target observer coverage levels differed among fisheries and ranged from 5% to 25% of trips annually, actual coverage levels were lower, ranging from 1% to 8% (Benoît and Allard, 2009).

Previous analyses of the observer records indicated that deployments of observers to vessels were not random (Benoît and Allard, 2009). The reasons underlying the irregular selection probabilities were not known but likely stem from two sources. First, observers may have avoided vessels that were perceived not to meet safety standards or were considered unfriendly. Second, there is known to have been targeted coverage of some vessels for enforcement purposes, but the targeted trips were not divulged to the analyst for privacy concerns. Evidence for observer effects was also found in the observer data, notably lower landings of commercially important species in the presence of observers (Benoît and Allard, 2009). Information was not available to directly consider whether the observer effects also resulted in changes in the ratio of skate discards to retained commercial catches. However, indirect evidence suggests that observer effects on skate discard ratios were small if present at all. Specifically, Benoît (2013) estimated skate landings based on observer records of retained catch (skates and commercial species) and compared these to estimates of landings based on dockside monitoring which were assumed to not be subjected to observer effects or other major sources of bias. The two compared favorably suggesting no observer effect on bycatch ratios for retained skate catch, and by inference on bycatch ratios for discarded skate catches.

The dockside monitoring program to measure target species landings is a census with close to 100% actual coverage, mandatory pre-arrival hails and mandatory dockside catch weighing. Therefore, it is not subject to impact from the statistical characteristics and it is subject to the impact of fewer operational characteristics.

The contributions of statistical and operational characteristics are described below, followed by a summary of the assessment for the case study. For simplicity, the contributions of monitoring programs for the different fisheries are considered jointly, rather than evaluating them separately and then summing the contribution.

The assessment covers the period from 1991-01-01 to 2011-12-31. The anticipated value for kg of discarded skates per kg of commercial species was taken as the average of annual estimates in Benoît (2013) and set at 0.041.

Acceptable errors were not set by fisheries managers; therefore, we used an acceptable relative estimation process error (pseudo-RMSE) of 30%, following the recommendations of 20% to 30% CV by National Marine Fisheries Service (NMFS) of the US National Oceanic and Atmospheric Administration in NMFS (2004). We note that the NMFS recommendation concerns only the variability (SE or pseudo-SE) while our selection is a combined limit on bias and variability.

9.2. At-sea monitoring skate discard ratio estimation process

9.2.1. Statistical characteristics

The anticipated statistically computed bias was assumed to be zero and the standard error was computed based on the average CV in Benoît (2013) and set at 0.010.

9.2.2. Operational characteristics

The assessment of operational characteristics is presented as follows:

Prevalence: Relative number of units included in the sampling or census frame impacted by the operational characteristic

Contribution to the estimation process bias (pseudo-bias): In most case, average difference between the values observed for these units and the true values due to the characteristic.

Contribution to the estimation process variability (pseudo-SE) expressed either as a (quadratically) additive term or as a multiplicative factor.

1 - Undercoverage

The sampling frame for the at-sea observer program is considered to include all population units (trips) for the fisheries, though there may be a very small number excluded. Therefore, very little impact from undercoverage expected.

Prevalence: 0% to 5%

Contribution to the estimation process bias: ~0%

Contribution to estimation process variability: ~0%

2 - Overcoverage

The sampling frame for the at-sea observer program does not extend beyond the fisheries it targets. Therefore, no impact from overcoverage.

Prevalence: 0%

Contribution to the estimation process bias: ~0%

Contribution to estimation process variability: ~0%

3 - Unintended clustering of samples

Due to the impossibility of separating the effect of unintended clustering, unintended stratification and other irregular selection probabilities, these effects are reported under “Other irregular selection probabilities”.

4 - Unintended sampling stratification

Due to the impossibility to separate the effect of unintended clustering, unintended stratification and other irregular selection probabilities, these effects are reported under “Other irregular selection probabilities”.

5 - Other irregular selection probabilities

Modified selection probabilities: In two of the main fisheries that catch skates (Greenland halibut fixed gear and offshore flatfish [plaice and witch flounder]) there is evidence of deployment effects: non-randomness among landing port districts, and vessels that tended to have higher average landings were disproportionately sampled (Benoît and Allard, 2009).

Based on the results of Benoît and Allard (2009), some vessels appear to be avoided perhaps for perceived safety reasons or perhaps unfriendly crew.

Deployment appeared random in the other major fishery, coastal flatfish mobile gear.

Forced inclusions**:** Furthermore, an unknown, variable, but likely small proportion of observer trips (5% to 20%), are targeted for enforcement or deterrence purposes. Targeted and randomly sampled trips are not distinguished in the at-sea observation database.

From simulations in Benoît and Allard (2009), we expect a relative bias of -15 to -10% overall due to the irregular sampling probabilities. However, landings of skate predicted using at-sea observation of skate discards match observed landings reasonably well, suggesting that bias is small if present (Benoît 2013).

From simulations in Benoît and Allard (2009) we expect that relative error will be underestimated requiring a correction of about 1.1 in fixed gear fisheries (which capture ~50% of skates) and ~1.6 in mobile gear fisheries (~50%), leading to a mean correction of ~1.35

Prevalence: 100% — The following contributions were assessed in Benoît and Allard (2009) for the whole population.

Contribution to the estimation process bias: −30% to −10%

Contribution to estimation process variability: +35%

6 - Observer effect

The results of Benoît and Allard (2009) suggest that an observer effect is likely for targeted catch and probably involves many vessels. It is not known if the observer effect extends to discarded catch; however, landings of skates predicted using at-sea observation of skate discards match observed landings reasonably well, suggesting that bias is small if present (Benoît 2013).

Anticipated prevalence, approximately 30-50% of observations; Average contribution to bias assumed to be approximately −10%. No contribution to variability expected.

Prevalence: 30% to 50%

Contribution to the estimation process bias: −10%

Contribution to estimation process variability: 0%

7 - Missing values due to unintentional factors

None observed or anticipated. Therefore, no impact anticipated. However, in some instances the observer may be imputing values for other sets during the trip, which is accounted in imputation error.

Prevalence: 0%

Contribution to the estimation process bias: ~0%

Contribution to estimation process variability: ~0%

8 - Missing values dues to intentional action

Independent at-sea observers are not expected to produce missing values intentionally, except perhaps in cases of strong coercion, which we assumed would be dealt with by fishery officers.

Prevalence: 0%

Contribution to the estimation process bias: ~0%

Contribution to estimation process variability: ~0%

9 - Errors in data reported by resource users

Independent observer program: not applicable.

10 - Errors in data reported by independent observers

Catch amounts are estimated visually by observers and therefore expect a small amount of unbiased error in nearly all observations. Observers can distinguish skates from other taxa therefore misidentification is assumed nil.

At-sea observers are certified and use standardized methodology. No biased observer error has been documented or is anticipated.

Prevalence: 100%

Contribution to the estimation process bias: ~0%

Contribution to estimation process variability: 5% to 15%

11 - Measuring tool error

Spring scales are used in conjunction with visual assessment to determine skate catch amounts. Moderate errors are anticipated, but no bias.

Prevalence: 100%

Contribution to the estimation process bias: ~0%

Contribution to estimation process variability: 5% to 15%

12 - Data handling error

Processing error may occur when observations are transcribed to data sheets and when data sheets are entered into the respective databases. There are some quality assurance and quality control procedures in place, particularly in the electronic data capture phase. The error associated with data handling is anticipated to be small to moderate.

Prevalence: 0% to 5%

Contribution to the estimation process bias: 0%

Contribution to estimation process variability: 5% to 15%

13 - Adjustment error

No adjustment applied. Therefore, no impact anticipated.

14 - Imputation error

In some fisheries in certain years there are too few observations to derive a valid estimate of bycatch rate for skates. In these cases (about 15% to 20% of fishery-year combinations), estimates are imputed using an average of data from adjoining years to minimize differences in relative catch rates resulting from changes in relative abundance and in fishing pattern (Benoît 2013). The expected bias is therefore small or nil. The error associated with imputation is also expected to be small.

In addition, observers are likely imputing some values for sets that are not (fully) observed. Some variability may be generated.

Prevalence: 75% to 100%

Contribution to the estimation process bias: 0%

Contribution to estimation process variability: 15% to 30%

Note
This imputation method is a smoothing procedure. As a rule, the imputed values will suffer from a larger bias if the smoothing "window" (here, the number of adjoining years) is larger or a larger variance if the smoothing window is narrow. This is sometimes called the "bias-variance trade-off". The direction of the bias may vary with each imputation. While it may be possible to assess statistically the bias and the variance, it may not be sufficiently important to warrant the effort.

15 - Modelling error

No modelling involved for this component specifically. However, modelling is involved when using landings to pro-rate discard amounts recorded by observers. No bias is expected, however variability is certainly generated, though it is at least partly included in the computed SE via a bootstrap (Benoît 2013).

Prevalence: 100%

Contribution to the estimation process bias: 0%

Contribution to estimation process variability: 15% to 30%

9.2.3. Overall assessment

After computations, we obtain the following summaries for the estimation process:

Estimation process bias (pseudo-bias): –26.3%

Estimation process variability (pseudo-SE): 33.1% (Statistical SE: 24.5%)

Estimation process error (pseudo-RMSE): 42.2%

The most important operational impacts are from “Irregular selection probability” (OC05) and “Observer effect” (OC06).

The impact of applicable operational characteristics OC09 to OC15 on variability is small because the sample size is large.

Potentially reducible error due to operational characteristics accounted for in the computed standard error: 0.4%.

We note that the variability of the estimation process is higher than that predicted by the statistical standard error and the bias is negative and important.

9.3. Dockside monitoring target species landing estimation process

9.3.1. Statistical characteristics

This is a census with actual coverage very near 100%: Bias and standard error are 0.

9.3.2. Operational characteristics

Prevalence: Relative number of units included in the sampling or census frame impacted by the operational characteristic

Contribution to the estimation process bias (pseudo-bias): In most case, average difference between the values observed for these units and the true values due to the characteristic.

Contribution to the estimation process variability (pseudo-SE) expressed either as a (quadratically) additive term or as a multiplicative factor.

1 - Undercoverage

Due to mandatory dockside monitoring of target species landing, the sampling frame is considered to include all population units (trips) for most years included in the study. Therefore, no impact from undercoverage.

Prevalence: 0%

Contribution to the estimation process bias: 0%

Contribution to estimation process variability: 0%

2 - Overcoverage

The sampling frame for the dockside monitoring of target species landing program does not extend beyond the fisheries it targets. Therefore, no impact from overcoverage.

Prevalence: 0%

Contribution to the estimation process bias: 0%

Contribution to estimation process variability: 0%

3 - Unintended clustering of samples

Not applicable (census).

4 - Unintended sampling stratification

Not applicable (census).

5 - Other irregular selection probabilities

Not applicable (census).

6 - Observer effect

Not applicable (census).

7 - Missing values due to unintentional factors

Given QA/QC procedures in place there should be very few missing values. Very little impact anticipated.

Prevalence: 0% to 5%

Contribution to the estimation process bias: ~ 0%

Contribution to estimation process variability: ~ 0%

8 - Missing values dues to intentional action

Given the checks and balances in place there should be very few instances and little impact.

Prevalence: 0% to 5%

Contribution to the estimation process bias: ~ 0%

Contribution to estimation process variability: ~ 0%

9 - Errors in data reported by resource users

Independent observer program: not applicable.

10 - Errors in data reported by independent observers

Retained catch amounts are almost all obtained from weigh-scales. Unbiased error is expected to be small.

Dockside observers are certified and use standardized methodology. No biased observer error has been documented or is anticipated.

Retained catch amounts are almost all obtained from weigh-scales. Dockside observers are certified and use standardized methodology. Transcription errors are possible, though there are no quantitative estimates of their prevalence or magnitude, though they should be small

Prevalence: 100%

Contribution to the estimation process bias: ~ 0%

Contribution to estimation process variability: 5% to 15%

11 - Measuring tool error

The data are obtained using weigh-scales. These are presumably calibrated regularly (so no bias); however, the authors don`t know the frequency and whether this is independently verified by DFO. The variability introduced is not known but expected to be small.

Prevalence: 100%

Contribution to the estimation process bias: 0%

Contribution to estimation process variability: 0% to 5%

12 - Data handling error

Processing error may occur when observations are transcribed to data sheets and when data sheets are entered into the respective databases. There are some quality assurance and quality control procedures in place, particularly in the electronic data capture phase. The error associated with data handling is anticipated to be small to moderate.

Prevalence: 0% to 5%

Contribution to the estimation process bias: 0%

Contribution to estimation process variability: 5% to 15%

13 - Adjustment error

Adjustments are made for the landing condition of fish. Corrections for dressing and icing of fish to whole fresh weight equivalents are made using empirical formulae. These adjustments are not anticipated to produce a bias. The error associated with adjustments is anticipated to be small to moderate.

Prevalence: 100%

Contribution to the estimation process bias: 0%

Contribution to estimation process variability: 0% to 5%

14 - Imputation error

Little imputation applied. Therefore, little impact anticipated.

Prevalence: 0% to 5%

Contribution to the estimation process bias: 0%

Contribution to estimation process variability: 5% to 15%

15 - Modelling error

Not applicable (census).

9.3.3. Overall assessment

After computations, we obtain the following summaries for the estimation process:

Estimation process bias (pseudo-bias): 0.0%

Estimation process variability (pseudo-SE): 0.0%

Estimation process error (pseudo-RMSE): 0.0%

Under- or over-coverage was not reported. Therefore, these operational characteristics do not impact bias.

While some operational characteristics create some variability, their impact on the estimation process variability is extremely small due to the large population size in this census situation.

9.4. Overall assessment of the total skate discard estimation process

The total skate discards are estimated using a ratio estimator based on the at-sea observation of skate discard ratios and the dockside monitoring of target species landings: the skate/target species catch weight ratio obtained from the at-sea observations is multiplied by total catch obtained from the dockside monitoring of the target species landings.

The result of the computation gives the following accuracy and precision for the total skate landings.

Estimation process bias (pseudo-bias): –26.3%

Estimation process variability (pseudo-SE): 33.1%

Estimation process error (pseudo-RMSE): 42.2%

Since no scientific or administrative quality requirements and no compliance limits have been set for this parameter, computation of the overall quality and dependability is presently not possible.

The using the NMFS-inspired 30% target relative RMSE, the dependability score is 0.71. On the proposed A-to-E assessment scale, the letter score is “B”.

While the estimation process variability is somewhat large, the errors due to the variability will vary between negative and positive values year-to-year. For example, if a time trend exists, it can be discovered by smoothing the time series (the relative residuals from the smoothing are anticipated to have a standard deviation around 33.1%). The estimation process variability is also useful to establish the year-to-year variations that are not due to the estimation process variability. For example, on a single-year basis, given a 33.1% pseudo-SE, only a change in the estimated skate discards of at least 50% can be considered meaningful.

On the contrary, the errors due to the estimation process bias will remain constantly negative. There is no statistical process similar to the above smoothing that allows the user to recognize or confirm the bias from the time series.

10. DISCUSSION

The framework presented here was designed to provide a thorough, reproducible and ideally consistent manner of evaluating the reliability of catch monitoring programs and the quality of the data they produce. While we have striven for completeness in the assessment framework, experience gained in applying it may further identify operational characteristics that should be considered. Likewise, some operational characteristics may turn out to be trivial or to contribute equally to all assessments, in which case they could be dropped to streamline 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. A user's guide for the framework will be completed concurrently with the publication of this report and will further help to ensure consistency.

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 base a score for 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. This report presents some heuristics than can be used to inform expert opinion. Similarly, Beauchamp et al. (2019) present examples of the effects of certain operational characteristics from the scientific literature, which can further inform decisions. With the accumulation of experience in applying the framework, further guidance will become available. This experience will also likely flag areas requiring targeted research aimed at understanding the consequences of operational factors affecting monitoring programs.

Completing an assessment effectively 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 variable selection probabilities 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 target 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 structure in sampling. It will therefore be critical that assessment teams be cross-sectorial and multi-disciplinary to ensure high quality evaluations.

11. ACKNOWLEDGEMENTS

We wish to acknowledge the help received from several Fisheries and Oceans Canada staff during the June peer review meeting and notably C. Faunce (US NOAA-NMFS), R. Houtman, H. Stone, N. Duprey, B. Beauchamp, M. Clemens and N. Kettenis who provided constructive comments on previous drafts of this document.

12. LITERATURE CITED

Adbola, F.B. and Oshungade, I.O. 2012. On a new method of bias reduction: alternative to approximately unbiased ratio estimators. Mathematical Theory and Modeling. 2: 80-89.

Babcock, E.A. and Pikitch, E.K. 2003. *How Much Observer Coverage is Enough to Adequately Estimate Bycatch?* Pew Institute for Ocean Science, Rosenstiel School of Marine and Atmospheric Science (verified on 2017-03-31).

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.

Benoît, H.P. 2013. Twenty years of annual landed and discarded catches of three southern Gulf of St. Lawrence skate species estimated under multiple sources of uncertainty. ICES J. Mar. Sci. 70: 554-563.

Benoît, H.P. and Allard, J. 2009. Can the data from at-sea observer surveys be used to make general inferences about catch composition and discards? Can. J. Fish. Aquat. Sci. 66:2025-2039.

Efron, B. 1979. "Bootstrap methods: Another look at the jackknife". The Annals of Statistics. 7: 1–26.

Groves, R.M., Fowler, F.J.Jr., Couper, M.P., Lepkowski, J.M., Singer, E., and Tourangeau, R. 2009. Survey Methodology 2nd ed. Wiley. ISBN-13: 978-0470465462.

Laplace, P.S. 1812. Théorie analytique des probabilités. Œuvres complètes, tome 7, ii-cliii+1-645.

Lohr, S. 2010. Sampling: Design and Analysis, 2nd edition. Brooks/Cole CENGAGE Learning, Boston, MA.

NMFS (National Marine Fisheries Service). 2004. Evaluating bycatch: a national approach to standardized bycatch monitoring programs. U.S. Dep. Commer., NOAA Tech. Memo. NMFSF/SPO-66, 108 p. On-line version, (view 2018-11-18)

13. ANNEX I - TECHNICAL NOTES

13.1. Sample sizes

Sample sizes are often described as a proportion of the target population. For example, an observer program will typically specify the fraction of trips that are required to carry an observer. In this case, the sampling unit is a trip and the population is the ensemble of trips.

Statistically, for simple random sampling, the relation between the standard error of the estimator of the population mean is given by the following formula:

$$\sqrt{1-\frac{n}{N}}\frac{\sigma}{\sqrt{n}}$$

where n is the sample size, N is the population size and σ is the population standard deviation.

The factor $\sqrt{1-\frac{n}{N}}$, called the "finite population correction factor", is negligible as soon as the sampling ratio or sampling portion n/N is smaller than 20% or, even, 40% (Figure 10).

Consequently, for small sampling fractions, the ***absolute sample size*** is the appropriate measure of sample size when discussing the variability of an estimator. The sampling fraction, i.e. the relative sample size, is only important when the sampling fraction is large. In a census, n=N, and SE is zero as can be seen in the equation above.

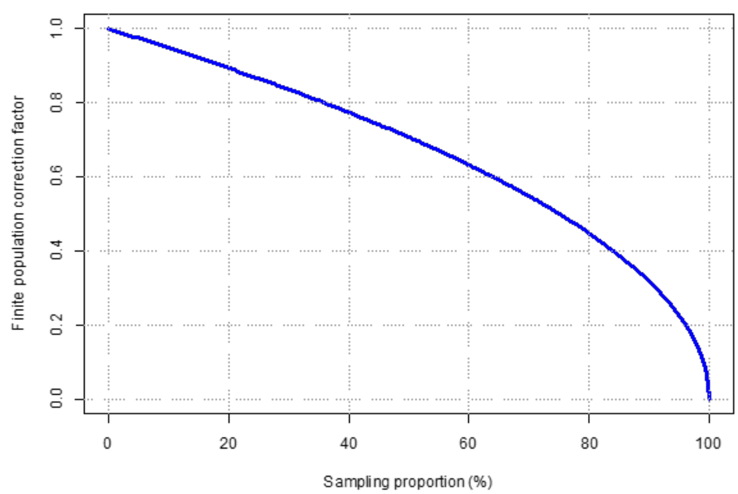


Figure 10. Relationship between the finite population correction factor and the sampling proportion.

13.2. Understanding the impact of operational characteristics on variability

We expect the impacts of operational characteristics on bias to be more important than those on variability. Furthermore, they are easier to understand.

The impact of operational characteristics on variability are less obvious. By variability, we mean the differences, on average, between estimates if we were to repeat the same estimation process many times. The variability must be first estimated using the relative standard error of the estimate (RSE), which represents the variability due to the randomness of the sampling process but may also include variability due to operational characteristics such as measurement errors. By “impact of operational characteristics on variability”, we mean effects that make true variability of the estimate smaller or (more often) larger than that reported by the RSE.

The following gives further details for censuses and sample surveys.

13.2.1. For censuses

For a census, the statistical variability is null: since the protocol requires that all units of the target population be observed, under ideal circumstances, the estimated value will be the true value.

In practice, some units of the population will not be observed. For example, in a telephone census of recreational fishers, a random subset of these fishers will not have been reached after the pre-set number of repeated calls due to unplanned circumstances (shopping habits, work schedules, etc. – as opposed to intentional avoidance). If the census was repeated many times in a thought experiment, different fishers may not be reached, creating variability in the estimate

Other operational characteristics such as measurement errors and data entry errors will also create some variability in the estimate.

13.2.2. For sample surveys

For sample surveys, the contribution of operational characteristics to variability may be difficult to assess and is likely to be case specific.

For example, consider the case of a short fishing season during which the CPUE increases and then decreases. Suppose also that this fishing season usually includes a vacation period during which the lone observer is not available. In this situation, the standard error of the estimator will be underestimated since it will not take into account the extra randomness of the vacation period relative to the CPUE temporal variation.

13.3. Impact of unintended cluster sampling

Cluster sampling and stratified sampling are two well established sampling methods. When implemented correctly and specialized statistical formulae are applied, they can give either a smaller standard error of the estimate for a fixed cost or a desired standard error for a lower cost.

When sampling is clustered and/or stratified unintentionally and statistical formulae appropriate for simpler sampling schemes (e.g., simple random sampling) are applied, for example because the analyst is unaware of the actual sampling scheme, the estimator can be biased and the standard error can be assessed incorrectly. In this section, we illustrate the impact of unintentional cluster sampling on standard error computations.

Clustering occurs when observations are taken from specific subsets of the population. For example, in a salmon recreational fishery, an observer can observe the fishers at several pools, possibly chosen at random. The pools present natural cluster of fishers. Another example concerns estimation of catch per unit of effort (CPUE) in seine fishing, with hauls as the unit of effort. If the sampling is carried out over several trips, each trip provides a cluster of hauls.

To illustrate the potential impact of unintentional clustering, suppose that a population of 1,000 is subdivided into 100 potential clusters of 10 individuals.

Suppose that the sample size is 100, taken from 20 of 100 clusters, with 5 individuals observed in each cluster.

We carried out simulation for 3 such populations, denoted A, B and C, illustrated in Figure 11. From A to C, the clusters were simulated to be progressively less similar to the whole population, as shown by adjusted R-square coefficients of 0.13, 0.30 and 0.70, respectively.

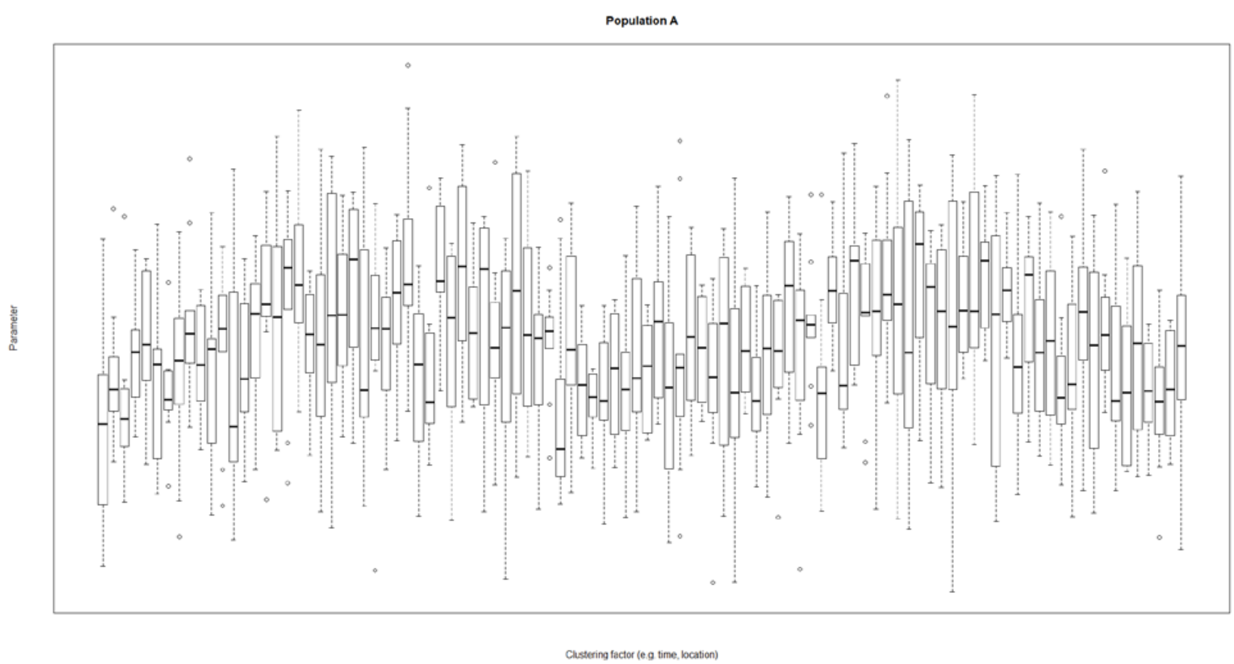
Correspondingly, the relative standard error of the estimate, computed on the incorrect assumption of simple random sampling, underestimates the true relative standard error by 10%, 25% and 47%, respectively.

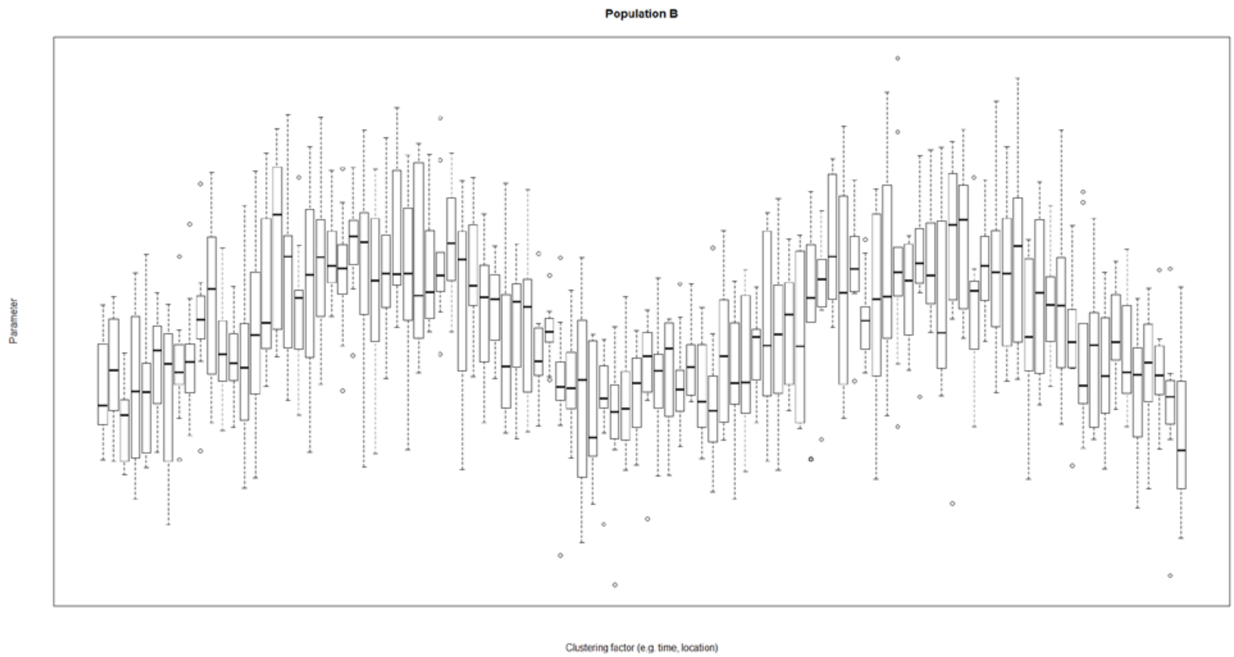
In general, unintentional clustering will lead to a larger underestimation of the true relative standard error when the clusters are less similar to the whole population (and to each other).

Population	Adjusted R-square (R_a^2)	Impact factor on the SE	SE Correction factor
A	0.13	0.90	1.11
B	0.30	0.75	1.33
C	0.70	0.53	1.89

The result of these simulations is consistent with theoretical results relating to cluster sampling (see Lohr, 2010)

Note: In the simulations, the estimators were found to be unbiased, a result consistent with properties of cluster sampling.





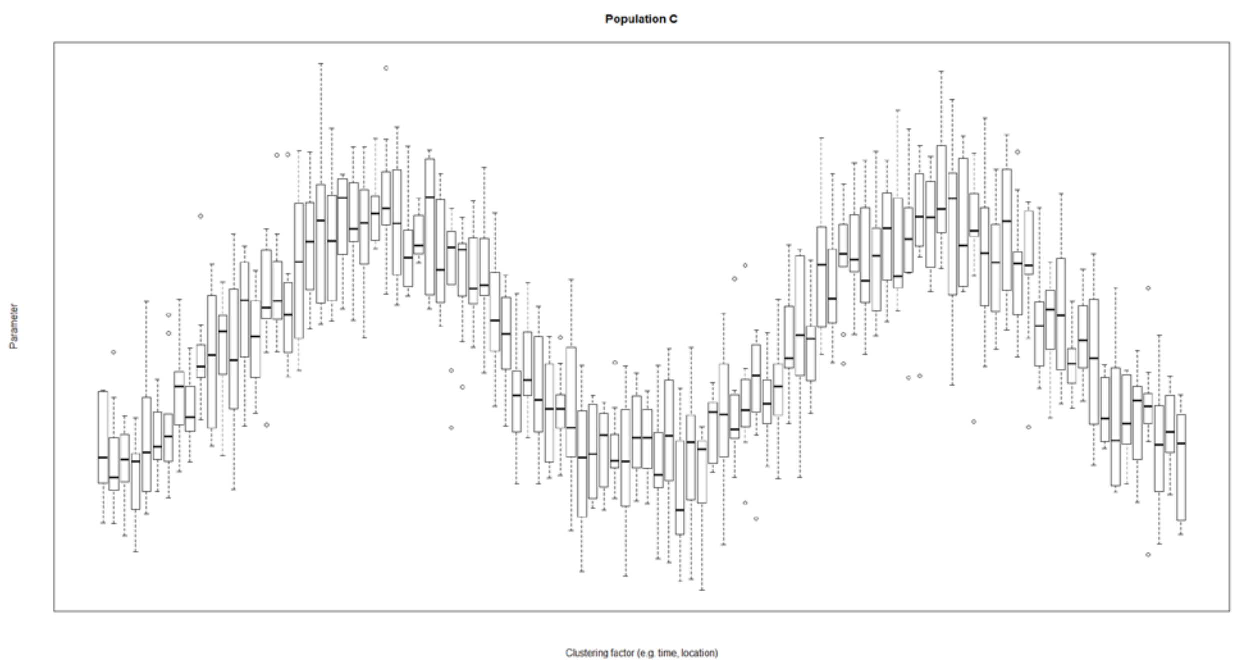


Figure 11. Box and whiskers plot of the values for simulated populations A, B and C, each of size N = 1,000, grouped into 100 unintentional subpopulations. The subpopulations (i.e. the potential clusters) are ordered in such a way as to illustrate the differences among them. The adjusted R-squared coefficient is 0.13, 0.30 and 0.70, respectively.

13.4.Impact of unintentional stratification on the RSE

Unaccounted for stratification can impact the RSE when the strata are very different from each other. To illustrate the impact of unintentional stratification we considered a simulation with 10 strata such that the adjusted R-square (R_a^2) of the ANOVA of the variable on strata is 0.93, i.e.

the strata are well separated (Figure 12). In this example, the RSE estimated by the simple random sampling formula is 3.1 times the true RSE.

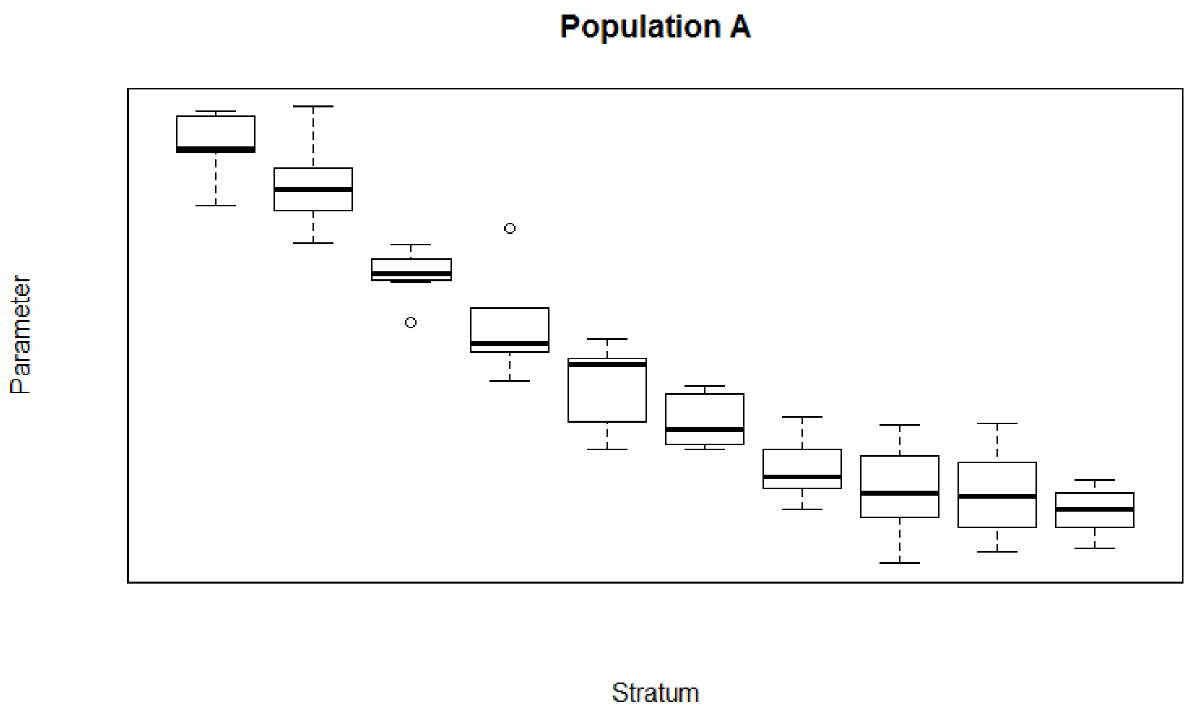


Figure 12. Population well stratified relative to the parameter of interest (R_a^2=0.93).

13.5.Impact of targeted sampling on the bias and the RSE

The following table shows the impact of targeting on the bias and RSE. The impact is largest when the targeted observations are at the extremes of the observations and smallest when they are around the mean of the observations. The impact is largest when the proportion of the sample targeted is highest.

This example is based on a population with a normal distribution. Results would be different if the population had an asymmetric distribution. For the RSE, the impact factors must be inversed to obtain the correction required: for example, if the impact factor is 0.80, the true RSE will be 1.25 times the computed RSE.

		Approximate location of the targets within the population (centile)			
		0th	10th	25th	50th
		Impact on the bias (%)			
Proportion of the sample targeted (%)	0	0	0	0	0
	5	-2	-1	-1	0
	10	-4	-2	-1	0
	20	-7	-5	-2	0
	30	-10	-7	-4	0
	40	-13	-9	-5	0
	50	-16	-12	-6	0

		Approximate location of the targets within the population (centile)			
		0th	10th	25th	50th
		Impact Factor on the RSE (%)			
Proportion of the sample targeted (%)	0	1	1	1	1
	5	0.87	0.97	0.97	1
	10	0.80	0.91	0.96	1
	20	0.72	0.85	0.95	1
	30	0.64	0.8	0.98	1
	40	0.59	0.73	0.95	1
	50	0.55	0.70	0.92	1

13.6.Impact of data errors on the estimation process

In the following, we illustrate the impact of data errors on the bias and variability of the estimation process in the case of the estimation of a population mean.

Definitions:

N: Number of units in the population

n: Number of units in the sample

x: True values

μ_x: Population mean

σ_x: Standard deviation of the population

ε: Observation error

σ_ε: Standard deviation of the errors

μ_ε: Expected value of the errors

y: Observed values

$y = x + \varepsilon$

s: Standard deviation of the observed sample

$\bar{Y}$: Mean of the observations in the sample

$\bar{X}$: Mean of the true values of observations in the sample

$\bar{\varepsilon}$: Mean of the observation errors

We assume that the data errors and the error due to the randomness of the sampling are independent.

For the bias, we have:

$$E(\bar{Y}) = E(\bar{X} + \bar{\varepsilon}) = E(\bar{X}) + E(\bar{\varepsilon})$$

$$Bias(\bar{Y}) = E(\bar{Y}) - \mu_x = E(\bar{X}) - \mu_x + E(\bar{\varepsilon}) = E(\bar{\varepsilon})$$

given $E(\bar{X}) - \mu_x = 0$ since $\bar{X}$ is an unbiased estimator of μ_x.

We conclude that biased errors contribute additively to the bias of the estimation process. The contribution is independent of the sample and the population size.

For the variability, we have:

Given the assumption of independence, we can write the statistical standard deviation as:

$$\sigma_{\bar{Y}}^2 = \sigma_{\bar{X}+\bar{\varepsilon}}^2 = \sigma_{\bar{X}}^2 + \sigma_{\bar{\varepsilon}}^2 = \left(1 - \frac{n}{N}\right)\frac{\sigma_x^2}{n} + \frac{\sigma_\varepsilon^2}{n}$$

Therefore

$$\widehat{\sigma_{\bar{Y}}^2} = \left(1 - \frac{n}{N}\right)\frac{s_x^2}{n} + \frac{\widehat{\sigma_\varepsilon^2}}{n}$$

Given that

$$s_{\bar{Y}}^2 = \left(1 - \frac{n}{N}\right)\frac{s^2}{n} = \frac{s_x^2 + s_\varepsilon^2}{n} - \frac{s_x^2 + s_\varepsilon^2}{N} = \left(1 - \frac{n}{N}\right)\frac{s_x^2}{n} + \frac{s_\varepsilon^2}{n} - \frac{s_\varepsilon^2}{N}$$

We obtain

$$\widehat{\sigma_{\bar{Y}}^2} = \left(1 - \frac{n}{N}\right)\frac{s_x^2}{n} + \frac{\widehat{\sigma_\varepsilon^2}}{n} = s_{\bar{Y}}^2 + \frac{\widehat{\sigma_\varepsilon^2}}{N}$$

We conclude that the contribution of the variability of the errors to the variability of the estimation process is partially accounted for in the statistical standard deviation as shown by the term $\widehat{\sigma_\varepsilon^2}/n$ in the equation for $\widehat{\sigma_{\bar{Y}}^2}$. This part of the contribution will be negligible if the sample size, n, is large.

The part not accounted-for is the term $\widehat{\sigma_\varepsilon^2}/N$. For a census, $n = N$ and $s_{\bar{Y}}^2 = 0$. Therefore, $\widehat{\sigma_{\bar{Y}}^2} = \widehat{\sigma_\varepsilon^2}/N$. This part of the contribution will be negligible if the population size, N, is large.

The accounted for contribution is dependent on the sample size. The unaccounted-for contribution is dependant on the population size.

In the assessment, when necessary, σ_ε^2 must be assessed from studies of the specific errors anticipated. For example, visual estimation of catch weight can be compared to scale measurements in a quality control exercise.

14. ANNEX II - GLOSSARY

Accuracy refers to the absence of a systematic error, i.e. an error which has either a positive or a negative expected value. Accuracy is the converse of *bias*.

Anticipated (*cf expected*): We reserve the word "expected" for statistical application as in "expected value". We use "anticipated" to refer to the act of foreseeing or predicting a value or an event.

Bias (operational): A systematic difference between the expected value of an estimator and the true value of the parameter estimated due to operational characteristics of the protocol. Applicable to random samples and to censuses.

Bias (statistical): The difference between the expected value of an estimator and the true value of the parameter estimated due to mathematical properties of the estimator. Bias is a signed value: a negative bias corresponds to underestimation, a positive bias to overestimation. Applicable only to random samples.

Bootstrap: When an estimate is obtained from a random sample, the probability distribution of the estimate often cannot be obtained from mathematical theorems. In most cases, the probability distribution of the estimate can be obtained, at least approximately, through a computer intensive method called the "bootstrap" in which the data are resampled as if their ensemble constituted the population. Once the probability distribution of an estimate is obtained, one can obtain its standard error and its bias.

Census: A survey where the complete population is targeted for observation.

Central Limit Theorem: When an estimate is obtained from a random sample, the probability distribution of an estimator can sometimes be obtained from mathematical theorems. In the simplest case, the mean of a simple random sample is used to estimate the population mean: in this case, the central limit theorem states that the probability distribution of the sample mean is a normal (also called Gaussian) distribution, at least if the sample size is large. It has been proven mathematically that the probability distribution of many other estimates is a normal distribution. Once the probability distribution of an estimator is obtained, one can compute its standard error and its bias.

Coverage: In sampling, the subset of the population that is included in the sample selection process. In a census, the subset of the population for which an observation is made or measurement taken. Fishery management *EXAMPLE:* Random samples or censuses of sport fishers may include only licensed fishers, excluding non-licensed fishers.

Dependability: Term used here to express the quality of an estimation process as a tool to fulfill fisheries management's objectives. Quality of the estimators is an important determinant of dependability.

Error (operational): Anticipated difference between an estimate and the value estimated in either a positive or a negative direction. *EXAMPLE*: Error due to mislabeling of a catch.

Error (statistical): Expected difference between an estimate and the value estimated in either a positive or a negative direction due to the randomness of a sampling protocol, most often described by the standard error. Applies only to random sampling.

Estimate: A quantity derived from a sample or a census used as an approximation of the parameter. Fishery management *EXAMPLE:* Using the product [CPUE (from sampling)] times [effort (from log books)] to estimate total catch. Typical notation: $\hat{\theta}$ for the parameter θ.

Estimation: Process of obtaining an estimate.

Estimator: A method to obtain an estimate from a sample based on data from a sampling protocol or a census. *EXAMPLES*: The sample mean of a sample obtained by simple random sampling is an estimator of the population mean. A 5%-truncated mean (whereby the smallest and largest 5% of the observations are removed before computing the mean) is also an estimator of the population mean.

Expected (*cf anticipated*): Used here in the statistical sense as in “expected value”.

Frame: In a census, list of all units to be observed. In a sample survey, the list of all units from which the sample will be drawn (*sampling frame*). In this document, we use the term “frame” to mean the available list of units of the target population. The frame may or may not coincide with the list of all units in the population. This is the simplest case of “frame”. The document points out situations where a more general definition of “frame” is required. Ideally, the frame and the *target population* would be identical. The frame may be created before the survey (e.g. the holders of a recreational fishery license for the current season) or during the survey (e.g. the herring gillnet fishing trip in Herring Fishing Area 5.)

Imputation: A process by which a missing value is replaced by a value assumed to be close to the missing value. There are several statistical methods to carry out imputation. Fisheries *EXAMPLE:* Using CPUE from neighboring rivers to estimate catch from effort for a river for which CPUE values are not available.

Independent observer data: Measurements collected by a person or a technology specifically tasked with observing and reporting on fishery activities and at arm’s length from the fishing industry or community, such as at-sea and dockside observers and on-board cameras and vessel monitoring systems.

Parameter: The quantity to be estimated. Fishery management examples: total catch of a target species, proportion of white crab in the catch, Lingcod catch per unit of effort. Typical notation: θ.

Parameter estimation process or *estimation process*: A complete process starting with the selection of units to be observed, the observations, the data entry, various computations, etc., and ending with the computation of the estimate.

Post-processing: Adjustments to data to compensate for variation in collection. Examples: gutted vs not-gutted catch weight; pound to kilogram conversion.

Precision refers to the repeatability/reproducibility of a measurement, i.e. an error which has an expected value of zero. Precision is the converse of *variability* or, in statistics, *standard error* and *variance*.

Quality: The measure of how close an estimate is to the true value: it summarizes *accuracy* and *precision*.

Resource user data: Measurements made and/or reported by the fishing industry or community, including fishers, plant personnel and buyers. Logbook records, purchase slips and answers to recreational fisher surveys are examples of resource user data.

Sample survey: A survey where only a *sample* from a population is observed. Opposed to census, where information is gathered from *all* individuals from a population.

Sampling protocol: The process used to obtain a sample. Fishery management example: The procedure used to decide when an at-sea observer should be deployed and which characteristics of the fishing trip they are meant to monitor and record.

Sampling variation (of an estimator): The variation of an estimator due to randomness of a sampling protocol.

Standard error (of an estimator): The most popular description of the variation of an estimator due to randomness of a sampling protocol. Typical notation: SE, $\sigma_{\hat{\theta}}$.

Standard error (SE): The standard of an estimator is a measure of the error due only to the randomness of the sampling protocol.

Target population: The set of all individual units that should be observed (census) or sampled (sample survey). Ideally, the *frame* and the target population would be identical. If they are not identical then there is either undercoverage (frame is smaller than the target population) or overcoverage (frame is larger than the target population). Examples: All fishing trips of a given fishery, all Atlantic salmon recreational fishers.

Variance (of an estimator): The square of the standard error. Statisticians prefer computing with the variance because some formulae become simpler and because "analysis of variance" allows a partitioning of the variance by source (but not of the standard error).