

High-level Code of practice for Scientific Data Verification

(Agreed at the 17th meeting of the Scientific Committee and adopted at CCSBT 19)

Introduction

The code of practice is intended to function as both a target and a guide to Members and Cooperating Non-Members (CNMs) on the procedures that should be in place for the verification of data. It is not intended to specify the types of data collection and monitoring systems that should be in place; instead, it provides information on the type of data management systems, and the types of verification/cross checking that are expected.

REPORTING ON SCIENTIFIC DATA VERIFICATION

To provide greatest understanding of the data, together with transparency and confidence in the data, all Members and CNMs are encouraged to report annually to the Extended Scientific Committee on the data verification conducted in accordance with this code of practice, together with the results of comparisons and the outcomes of any investigations into the data.

DATA MANAGEMENT SYSTEMS

It is expected that all scientific datasets maintained for CCSBT purposes would be managed using a robust database management system (e.g. SQL Server, Oracle), and preferably using a relational model. The database(s) should be professionally designed and implemented, and be accompanied by up-to-date documentation. Where a Member's datasets reside on different physical databases, systems should be in place to allow easy cross-checking and verification between the physically separated datasets.

Automatic checking should be conducted at the time of data entry/loading to prevent erroneous data being stored on the database(s). Automatic checks should include:

- **Validity checks:** These are checks or constraints on individual fields to ensure that the data is valid. They include checks on the format of the data (e.g. that a valid date is provided); the validity of codes (e.g. that a valid species code or statistical area is provided); the magnitude of a value (e.g. that a weight is within an acceptable range, and a date is not in the future etc.). A variety of validity checks should exist on nearly all fields within the database.
- **Plausibility checks:** These are checks to identify items that are unlikely, but not impossible. These checks will often be range checks such as: very small or large weights/numbers/hooks; small or large average weights for a species etc. As a minimum, when these checks reveal an unlikely item, the operator should check the data to ensure that a data entry error has not been made. Checks of this nature should be implemented for all relevant fields (most numeric and date fields) and the checks should be finetuned for the specific data (e.g. the actual species and gear) involved.

Checks of these types, when implemented at the database level as specified, significantly reduce the risk of erroneous data being stored.

CROSS-VERIFICATION OF DATASETS

The main data that Members and CNMs currently provide to the CCSBT for scientific purposes comprise: Total SBT catches, Catch and Effort, and Catch at Size/Age data. The CCSBT has also adopted a Catch Documentation Scheme to confirm catches of Members and CNMs, as well as Scientific Observer Program Standards with a target observer coverage of 10% for catch and effort monitoring for each fishery. Furthermore, different Members/CNMs have additional programs (such as real-time monitoring, landing inspections, and quota monitoring systems) in place to monitor and manage their catches.

An important component of this code of practice is that each scientific dataset be cross-verified against other, independent data sets wherever possible and that this cross-verification be conducted for each scientific dataset on an annual basis. The cross-checking recommended for each scientific dataset is as follows:

Total Annual SBT Catches

Members/CNMs report total annual SBT catches to the CCSBT as part of the “Total Catch by Fleet” data provided for the annual Scientific Data Exchange, in national reports to the Extended Scientific Committee and Extended Commission, and as part of the Final Catch by Vessel/Client reporting requirements. All these reports should be cross checked to ensure that the figures are the same. In addition, the following verification(s) should be conducted:

- The nationally reported annual SBT catches should be compared on a gear by gear basis with the annual catch estimated from CCSBT CDS documents for the same years¹. It is expected that the nationally reported catch should closely match the CDS figures. Discrepancies of greater than +/- 5% should be explained. If a clear explanation is not readily available, discrepancies of greater than +/- 5% should be investigated².
- The nationally reported annual SBT catches should be verified against any other independent nationally available total catch data sets such as quota monitoring system or landing inspections.

Commercial Catch and Effort data

Catch and Effort data is provided to the CCSBT Secretariat annually as a part of the Scientific Data Exchange. These figures should be verified where possible in the following ways:

- Commercial Catch and Effort data for observed trips should be crossed-checked against the observer’s data for the same parts of the same trips. Any discrepancies should be investigated.
- Commercial Catch and Effort data for non-observed shots should be compared with data for observed shots. Any substantial inconsistencies in the temporal or spatial CPUE

¹ The Secretariat can provide Members with CDS figures for their documents on request. However, there is a time lag in provision of CDS data such that figures for the most recent year may not always be complete.

² A technical working group at CCSBT 12 recommended that the principles for a CCSBT CDS should include a performance measure that the CDS be capable of accounting for at least 95% of all sources of fishing mortality of southern bluefin tuna (paragraph 90 of the CCSBT 12 report)

estimates or trends, or relative proportions of bycatch for the two datasets should be investigated.

- The weights of SBT from the unraised³ Catch and Effort data should be compared with:
 - Total Annual SBT Catches: Any substantial discrepancies⁴ (including differing trends in total catches between the two data sources between years) should be investigated.
 - CDS harvest data, stratified by statistical area and month¹: Again, any substantial discrepancies should be investigated.

Catch at Size data

Catch at Size data are provided to the CCSBT Secretariat annually as a part of the Scientific Data Exchange. For those Members/CNMs whose Catch at Size data is collected independently of CDS Catch Tagging Forms, a spatio-temporally stratified comparison should be made of the catch at size distributions of the two data sets⁵. Any substantial inconsistencies should be investigated.

³ Some Members raise their catch and effort data to match that of the total catch before providing that data to the CCSBT.

⁴ After adjusting for the Catch Effort reporting rate (e.g. log books not being provided for a certain percentage of fishing).

⁵ This is not possible for farmed product as the CDS length data is for grown out SBT.