(ver. 1.1; March 2021)

Maciej Adamowicz, Ireneusz Wójcik
Department of Logistics & Monitoring
National Marine Fisheries Research Institute
Gdynia, Poland

Quality assurance of biological data collected from pelagic fisheries in SPRFMO waters by NMFRI scientific observers.

Introduction

On the basis of a multilateral agreement between Germany, Lithuania, Netherlands and Poland, since 2017 Poland has been coordinating a joint sampling programme for the collection of biological data from pelagic fisheries in SPRFMO waters. Sampling is carried out on EU vessels which conduct pelagic fisheries in the South Pacific according to licences and allocated fishing quotas.

NMFRI scientific observers collect data from catch sampling, length distribution and biological parameters of fish in accordance to the requirements of SPRFMO, which are specified in the documents listed below:

a) SPRFMO Conservation and Management Measure for Trachurus murphyi,
b) SPRFMO Data Standards,

which can be found on the official SPRFMO website (https://www.sprfmo.int/). Moreover, the website contains data on fishing vessels authorized to conduct fisheries in SPRFMO waters.

Data quality checks

Data processing and quality control are split into different stages.

Stage 1

Data are checked for common errors which could have been made while registering the data in a temporary database in Excel. This applies primarily to the information on date, time and coordinates of fishing operations. Two data sources are used in this step of data validation. The first one is the data provided by an operator of a fishing vessel. The second source of information is the data collected by scientific observers, registered in a dedicated form developed by SPRFMO, which is available on the official website (https://www.sprfmo.int/data/data-submission/). The compliance of both datasets is checked by a script developed in *Power Query*. A list of all found data issues is automatically generated in an Excel spreadsheet which is then forwarded for clarification.

Additionally, data on coordinates and catch weight are visualised on a map to verify that the fishing location is within the expected boundaries. This type of data check is implemented with application of the "*Map 3-W*" module.

Technical parameters of a vessel registered by scientific observers in the "Observer template" are compared with the ones that are available on the SPRFMO official website (Records of Vessels). In case of any inconsistencies, the data are corrected.

Stage 2

The next step of data validation checks that the data format complies with the specification that can be found in the document "SPRFMO Data Standards". A script in *Power Query* was developed which automatically detects incorrect data formats and corrects them.

Stage 3

This stage of data validation focuses on biological data collected by scientific observers. In order to check for outliers in biological parameters of fish, a scatter plot of weight-length and age-length is produced, which allows for a visual detection of invalid observations. If outliers are detected in the data, then the information is forwarded to an observer or NMFRI expert for additional checks. Outliers may be caused by typos while registering the data in a database. However, there are cases when an outlier is connected with natural reasons, e.g. diseases, parasites, poor condition. Therefore, expert knowledge is necessary for the unequivocal identification of outliers.

Another data validation check implemented in *Power Query* compares sample weight with a sum of individual weights of fish taken for biological analysis. This check assumes that sample weight should be greater or equal to the sum of individual weights.