DEVELOPMENT STATUS OF THE NEW TROPICAL TUNAS TREATMENT (T3) SOFTWARE

Duparc A.[1], Depetris M.[1], Floch L[1]., Cauquil P.[2], Bach P.[1], Lebranchu J.[1]

Abstract

This paper is an information note on the progress of the development of the new version of T3 following the development of a new statistical model.

Introduction

The multi-species nature of tropical tuna surface fisheries gives rise to a series of difficulties when estimating both catch by species and catch by size statistics. The T3 processing was built about 30 years ago in order to correct biases in the species composition reported in logbooks and to provide more accurate estimates of catch by species for the European purse seine fleet. However, the evolution of fishing practices and fishing stocks according to climate change have challenged the T3 methodology on some part of its processing. The different biases in the existing T3 processing were already presented and discussed (Duparc et al., 2018). In order to take these different biases into account, we developed a new statistical methodology (Duparc et al., 2019). The aim of this paper is to present the status of the new T3 implementation, which began in 2019 with a complete redesign of the code and the upcoming milestones.

A redesign for the T3 code

The original methodology developed in the Fortran language was previously described (Pallarés and Petit, 1998). The lack of a dedicated detailed documentation of the treatments performed was one of the major drawbacks. Reverse engineering work was carried out at the beginning of 2010 with the aim of replacing the tool with a documented program in a modern language, Java. A new version was developed reproducing the same methodology based on large sample areas, relying on T3 on the documentation of reverse engineering (Cauquil et al., 2018). However, the version 2 of T3 could not succeed completely because it did not resolve all the biases of the first version.

Based on these facts, and in order to provide a functional version, we decided to completely review the program code.

The new version of T3 process is developed in R language and available through a package. One of the advantages of this language is that it is more understandable and manipulated by the fisheries' scientific community rather than Java language. Moreover, this development follows the principles of the open and reproducible science:

- Process transparency: the code is fully documented and the documentation of how works the program will be available.
- Open access and Open source: the code source is hosted on a github repository (https://github.com/OB7-IRD/t3) under a GPL version 3 license. In the end, it would be available on the Comprehensive R Archive Network (CRAN).

[1] MARBEC, IRD, Univ Montpellier, CNRS, Ifremer, Sète, France
[2] MARBEC, IRD, Univ Montpellier, CNRS, Ifremer, Abidjan, Côte d'Ivoire

- Referencing of the software : enable result reproducibility and versioning (Depetris et al., 2020) : version 0.9.0 - DOI: 10.5281/zenodo.3878125.

Changes in the new version

The overall schema for the processing remains unchanged from the last version of T3 (**fig. 1**). As previously described (Duparc et al., 2018), the first two steps are the standardization of the logbook reporting and the sampling data, and third step is the assessment of the species composition of the catches. Thus, the development of the T3 version aims at fixing, or reducing as much as possible, biases identified previously (Duparc et al. 2018). Upgrades of the process are summarized below.

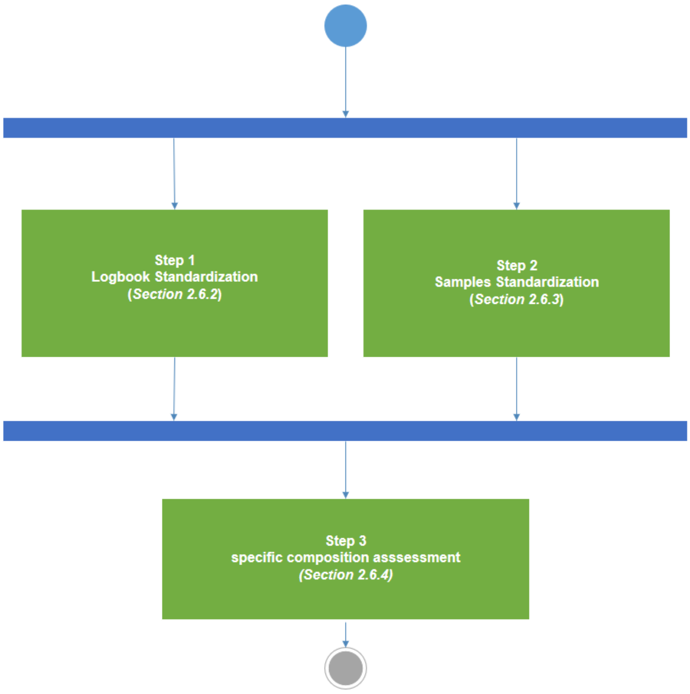

Fig. 1: Activities diagram of the main step in T3 processing (source: Duparc et al. 2018). Sections corresponds to the paragraph in Duparc et al. (2018) which details the step.

Step 1: Standardization of logbooks

- Code optimization: computing time reduced
- Improve fishing time estimation: compute fishing time according to daylight time related to the vessel location

Step 2: Standardization of samples

- Code optimization: computing time reduced

Step 3: Assessment of the species composition to non-sampled sets

- Strict selection of samples used in the model training according to several objective criteria on fishing sets and wells: school type, location, date, sample quality.
- Classification of unknown school type (no data available) by clustering according to major tuna species composition.
- Development of a statistical model (random forest model described in Duparc et al., 2019) accounting for:

   - Spatial and temporal structure of the fishery: location and date of the catch
   - Vessel specificity which impacts the species composition of the catch
   - Crew reporting information which enables more flexibility in the model when a marginal composition is caught
   - Several years of fishing in the training to increase the robustness of the estimates
- Model Checking to ensure reliability of the results (see example in **Appendix 1**)
- Computation of a confidence interval (bootstrap method) to quantify the accuracy of estimates of catch per species.

Planning

Initially, following the development of the new model, we established a six-month development program with the objective of presenting the new tool to the IRD partners, namely IEO[3] (Spain), CRODT[4] (Senegal) and SFA[5] (Seychelles) during the annual data preparatory meeting that takes place within the framework of the RCG LP[6] at the end of March 2020 (**Fig. 2**).

During this meeting, it was planned to test data from each partner in order to experiment the program and then to provide a first version in April to be presented at the next dedicated RFMOs meeting for the Atlantic Ocean (ICCAT) and the Indian Ocean (IOTC).

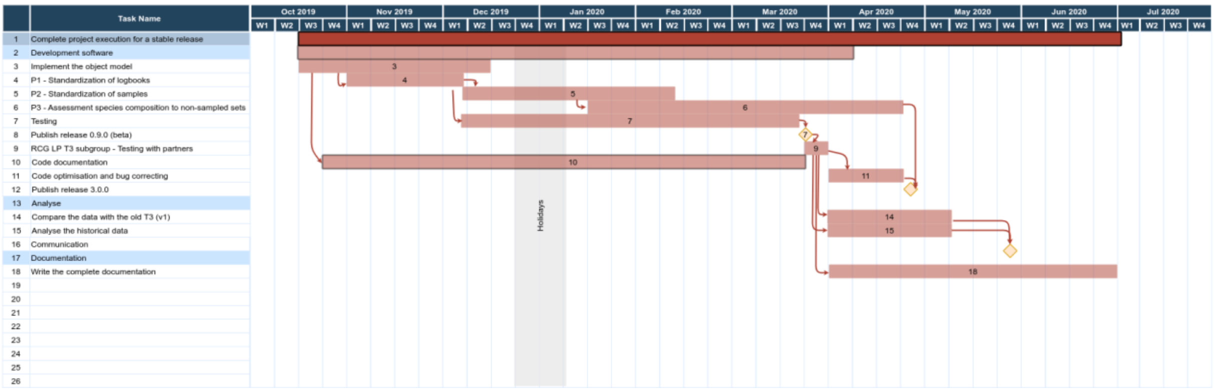

Fig. 2: Gantt diagram of the initial work schedule

However, the COVID-19 pandemic has led to a significant delay in implementation due to the containment from March 17 in France. Although working remotely, the time available for the project was therefore impacted and we had to review the development program for the remaining steps (**fig. 3**).

[3] Instituto Español de Oceanografía
[4] Centre de Recherches Océanographiques de Dakar-Thiaroye
[5] Seychelles Fishing Authority
[6] Regional Coordination Group of Large Pelagic

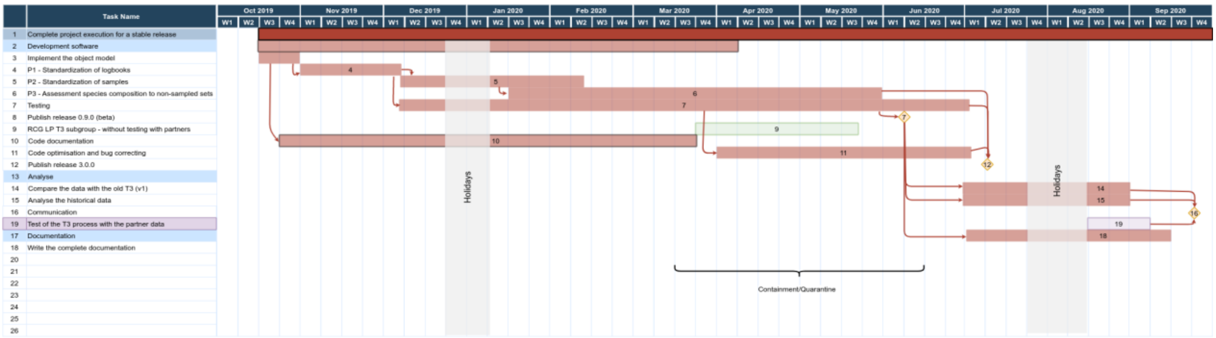

Fig. 3: Gantt diagram of revised work schedule

The main milestones up to October 2020 are:

- Make available on GitHub a release candidate of the package
- Write the complete documentation (User guide and Dev guide)
- Analyse the historical data and publish the results

At this time, the beta version (0.9.0) is available (Depetris et al., 2020) but the implementation of the new version is not fully achieved. Several tests remained to perform to check for the data traceability and to optimize the model parametrization. Then the last two decades could be revised and compared to the T3 v1 process. Historical data, prior to the 2000s, will require adaptation of the new model considering the partial availability of the information and the decrease of the data quantity.

Acknowledgments

We are grateful for all technical staff that collected data used in T3 processing and used in this paper. The data used were collected through the Data Collection Framework (Reg 2017/1004 and 2016/1251) funded by both IRD and the European Union.

References

Cauquil, P., Chassot, E., Billet, N., 2018. T3+ - Documentation des traitements - *internal report*. IRD.

Depetris, M., Duparc, A., Floch, L., Cauquil, P., Lebranchu, J., 2020. OB7-IRD/t3: Beta version of T3 software. Zenodo. https://doi.org/10.5281/zenodo.3878125

Duparc, A., Aragno, V., Depetris, M., Floch, L., Cauquil, P., Lebranchu, J., Gaertner, D., Marsac, F., Bach, P., 2019. Assessment of the species composition of major tropical tunas in purse seine catches: a new modeling approach for the Tropical Tuna Treatment processing, in: Report of the 21st Session of the IOTC Working Party on Tropical Tunas, IOTC-2019-WPTT20-10. Presented at the 21st session of the IOTC Working Party on Tropical Tunas, IOTC, San Sabastián, Spain, pp. 1–35.

Duparc, A., Cauquil, P., Depetris, M., Dewals, P., Gaertner, D., Hervé, A., Lebranchu, J., Marsac, F., Bach, P., 2018. Assessment of accuracy in processing purse seine tropical tuna catches with the T3 methodology using French fleet data. Case of the French fleet in the Indian Ocean, in: Report of the 20th Session of the IOTC Working Party on Tropical Tunas, IOTC-2018-WPTT20-16. Presented at the 20th session of the IOTC Working Party on Tropical Tunas, IOTC, Victoria, Seychelles, pp. 1–19.

Appendix 1: Model validity checking in the T3 process - example with the skipjack proportions in catches on school type associated with floating object

The model used was a random forest model with proportion of skipjack from landing samples as response variable and as explanatory variables: latitude, longitude, year, month, vessel id and skipjack proportion reported in the logbook (period: 2014 - 2017, N_{sample} = 782). These graphs (**Fig. A1-1**) will be a part of the outputs of the T3 v3 which enable the evaluation of the model validity by users.

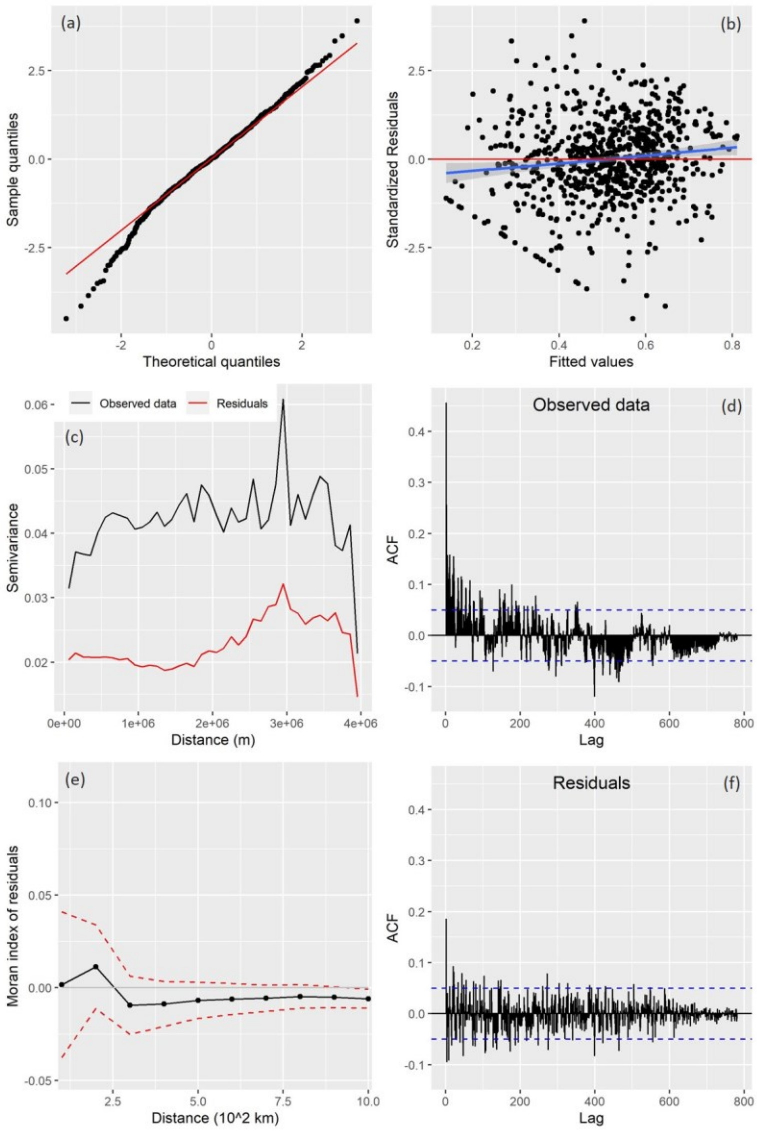

Fig. A1-1: (a) qqplot of the residuals. (b) Fitted values against standardized residuals of the model. (c) Semivariograms for the observed data and the residuals of the model. (e) Spatial autocorrelation of the residuals (Moran index mean and IC 95%). (d) and (f) represent the temporal autocorrelation (ACF) for observed data and residuals respectively.