BLUEFIN TUNA LARVAL INDICES IN THE WESTERN MEDITERRANEAN, ECOLOGICAL AND ANALYTICAL SOURCES OF UNCERTAINTY

D. Alvarez-Berastegui[1], M.P. Tugores[1], D. Ottmann[2], M. Martín-Quetglas[2], P. Reglero

SUMMARY

The main objective of this study is to provide the knowledge to design adequate sensitivity analyses on the assessment models used for the Eastern stock of Bluefin tuna. We analyze how different configuration for the same environmental variable (temperature in the mixed layer depth) and different modeling approaches (nonlinear Delta-log,delta-gamma, tweedy and bayesian) affects to the variability of the larval indices of the Eastern bluefin tuna from data collected in the Balearic Sea (Western Mediterranean). We also investigate the effects on the index caused from having differences in the total sampled area among years. We used these results to interpolate larval index values in years with not standard larval surveys but with some ichthyoplankton surveys available, and to propose a "revised version" of the index providing parameters of uncertainty.

RÉSUMÉ

L'objectif principal de cette étude est de fournir les connaissances nécessaires pour concevoir des analyses de sensibilité adéquates sur les modèles d'évaluation utilisés pour le stock oriental de thon rouge. Nous analysons comment différentes configurations pour la même variable environnementale (température dans l'épaisseur de la couche de mélange) et différentes approches de modélisation (Delta-log non linéaire, delta-gamma, tweedy et bayésienne) affectent la variabilité des indices larvaires du thon rouge de l'Est à partir de données collectées dans la mer des Baléares (Méditerranée occidentale). Nous étudions également les effets sur l'indice causés par les différences dans la zone totale échantillonnée entre les années. Nous avons utilisé ces résultats pour interpoler les valeurs de l'indice larvaire pour les années où il n'y avait pas de prospections larvaires standard mais où certaines prospections d'ichthyoplancton étaient disponibles, et pour proposer une "version révisée" de l'indice fournissant des paramètres d'incertitude.

RESUMEN

El principal objetivo de este estudio es proporcionar conocimientos para diseñar análisis de sensibilidad adecuados en los modelos de evaluación utilizados para el stock oriental de atún rojo. Analizamos cómo afectan las diferentes configuraciones para la misma variable medioambiental (temperatura en la profundidad de la capa de mezcla) y los diferentes enfoques de modelación (Delta lognormal no lineal, delta-gamma, tweedy y bayesiano) a la variabilidad de los índices larvarios del atún rojo oriental a partir de los datos recogidos en el mar Balear (Mediterráneo occidental). También se investigan los efectos sobre el índice causados por las diferencias en la zona total muestreada entre los años. Se han usado estos resultados para interpolar los valores del índice larvario en los años sin prospecciones de larvas estándar, pero con algunas prospecciones de ictioplancton disponibles, y para proponer una «versión revisada» del índice aportando parámetros de incertidumbre.

KEYWORDS

Standardization of CPUE, environmental variability, Balearic Sea

[1] ICTS-SOCIB, Balearic Islands Coastal Observing and Forecasting System, Palma de Mallorca, SPAIN. E-07121.

[2] IEO, Instituto Español de Oceanografía, Muelle de Poniente s/n.Palma de Mallorca, SPAIN.E- 07015

1. Introduction

Larval indices are standardized means of larval abundances from ichthyoplankton surveys, normally expressed as capture per unit area (CPUA). In the framework of the SCRS bluefin tuna working group, larval indices are used as proxy of the spawning stock biomass. The first larval abundance index for the bluefin tuna Easter stock was presented to the SCRS bluefin tuna group in 2013 (Ingram *et al.,* 2013). This larval index covered the five years (2001- 2005) and was computed following standard methods already applied for the species Western stock, in the Gulf of Mexico (Ingram *et al.,* 2010). Since then, new surveys and methodological developments have been incorporated to improve the accuracy and precision of the index. Some of these advances are the integration of environmental covariables (Ingram *et al.,* 2017), the development of gear standardization models using experimental fishing (Alvarez-Berastegui *et al.,* 2018), the improvement in standard errors calculation after the logarithmic transformations applied in delta-lognormal models, and the integration of nonlinear effects of covariates.

In 2017 the Atlantic bluefin tuna larval index in the Western Mediterranean was incorporated in the Virtual Population Analysis (VPA) assessment as a proxy for temporal trends in the spawning stock biomass (Anon, 2017). This version of larval index was constructed with a delta-lognormal approach where compensation for factors were resolved using Least-Squared Means (Lenth, 2016). An updated version of the larval index was tested in 2018 using a delta-gamma approach but allowing non-linear responses of covariates and resolving the effect of unbalanced factors using a bootstrap approach. At that moment, it was relevant that both approaches resulted in indices with some differences in the interannual trends and standard errors. Exploring the advantages and disadvantages of different fisheries abundance standardization methods and the variability of the results they provide is a relevant question for advancing in the proposal for "best practices" in the field of fisheries assessment (Forrestal *et al.,* 2019). In order to improve the larval abundance indices (reduce uncertainty and increase accuracy) it is desirable to provide new updates following state-of-the-art biological knowledge and mathematical standardization methods. Nevertheless, the incorporation of an updated abundance index computed with a different method within a fisheries assessment process may force to readjust the complete fisheries assessment models, which is not always possible nor desirable.

To date, the Spanish Oceanographic Institute (IEO) and the Balearic Islands Coastal Observing and Forecasting System (SOCIB) provide two different versions of the larval index for the bluefin tuna in the Western Mediterranean the "VPA strict update index" and the "MSE strict update index". These are applied as input data for the Virtual population Analysis and the Management Strategy Evaluation (MSE), both calculated from same input data but using different explanatory covariates and modelling approaches. These versions are expected to be used for successive stock assessments updating. Additionally, we calculated an alternative larval index called "revised version", which is not constrained by the need of applying a specific standardization method, and uses the currently best available information and modelling technique. This "revised version" shows the most realistic trend of larval abundances and can be used to explore deviations from the previous index (currently used for the VPA and the MSE) and design a sensitivity analysis to compare uncertainty of these two fishery assessment approaches.

Regarding the possibility of advancing on the information obtained from the larval index in the Western Mediterranean there is also the open question of whether it is possible to retrieve abundance values of years missing standardized surveys (2006 to 2011), by using available information of other ichthyoplankton surveys.

Here we evaluate a number of questions related to how different aspects of the standardization process for the larval index affect the accuracy and precision of the final result. These questions are:

1- What are the effects of applying different statistical standardization methods?

2- What are the effects of changing the way that an environmental indicator is parameterized when included in the standardization process?

3- Can we reconstruct the larval abundance for those years lacking standard surveys directed to bluefin tuna spawning?

4- How relevant is the issue of changing total sampling area among years?

The results of these analyses allow to select a **revised a version of the larval index** and associated SE and CV to design sensitivity analyses for the VPA and MSE assessments of the Eastern stock of bluefin tuna.

2. Material & Methods

Biological data

Atlantic bluefin tuna (*Thunnus thynnus*) larvae were collected during fifteen ichthyoplankton surveys around the Balearic Islands during summer (**Table 1**). Eleven of these surveys were systematic ichthyoplankton surveys targeting the spawning peak of Atlantic bluefin tuna in the area. Systematic surveys were carried out in June–July in two periods, i.e. 2001–2005 and 2012-2017. Survey design consisted of a regular grid of 10 × 10 nautical miles covering the area between 37.858–40.358 N and 0.778–4.918 E, covering an area of 86,351 Km^2 (**Figure 1A**). Fishing operations were conducted at around 2 knots, during 8–10 minutes and covered a linear distance of about 600 m. The volume of water filtered was measured with flowmeters located at the centre of the net. During the period 2001-2005, tows were performed using a bongo net of 60 cm mouth diameter, stepped obliquely to a depth of 70 m (deep oblique), or from 5 m above the bottom at coastal stations, to the surface maintaining the vessel speed at two knots. For the second period, the sampling was conducted using bongo 90 nets fitted with 500 microns meshes towed down to 20 – 30 meters (mixed layer oblique), covering the whole mixed layer depth in this area and season (Torres *et al.*, 2014). Additionally, bluefin tuna larvae information from non-standard surveys was retrieved for four years, i.e. 2006, 2008, 2010 and 2011 (**Figure 1B**). These data were collected during ichthyoplankton surveys that did not cover the entire sampling area of the standard surveys or were displaced in the sampling time (see dates and % of area covered in **Table 1**). Hence, survey design, survey period and area coverage were not standard. Sampling was performed by means of a bongo 90 net, except for 2010 when a bongo 60 was used.

In all surveys, plankton samples were preserved with 4% formalin buffered with borax. Tuna larvae were identified to the species level and measured in standard length. Once in the laboratory, the number of larvae were counted and standardised at each hauling station, following Álvarez-Berastegui *et al.,* (2017). First, the abundance of Atlantic bluefin tuna are standardised to larvae of 2 mm to avoid the exponential decay due to natural mortality, following the equation:

$$N_{2mm} = 1092944e^{-0.722\,L_i}, \qquad \text{(Eq. 1)}$$

where N_{2mm} is number of larvae at 2 millimetres and L_i total length of larvae, in mm. Second, the catch per unit area (CPUA, in N larvae / m^2) is computed following the equation:

$$CPUA = \frac{N\ larvae}{V_{filtered}}\ D_{tow} \quad , \qquad \text{(Eq. 2)}$$

where $V_{filtered}$ is the volume of water filtered by the net (in m^3) and D_{tow} is the towing depth (in m). Finally, CPUA is standardised or the two nets used following an exponential relationship between B60 deep oblique and B90 mixed layer oblique of the form:

$$CPUA_{B90} = 0.58\ CPUA_{B60}\ e^{0.00115\ CPUA_{B60}}, R^2 = 0.998 \qquad \text{(Eq. 3)}$$

Environmental data

In-situ environmental data regarding mean temperature and salinity of the water column down to the mixed layer depth (TMLD and SMLD respectively) were retrieved during the surveys using a CTD. Longitude, latitude, year, day of the year (DY) and time at which the tows were produced were recorded at each sampling station. Temperature anomaly (Tanom) and salinity anomaly (SALanom) were computed as the temperature and salinity at the mixed layer standardised to the annual mean temperature and salinity, respectively. For temperature, the residual temperature obtained after removing the temporal increasing trend was also computed (tempres2).

Larval index standardization models applied

The relationship between larval abundance and environmental data was inspected through generalised additive models (GAM). First, the we computed a model named (REFERENCE.MODEL) with same explanatory variables of those included in the previous version of the MSE-strict updated larval index, but resolved using emmeans (estimated marginal means) for the compensation of factors and calculation of the standard error (back transformed within the emmeans) and including additional data of non-standard surveys.

Two-stage models including lognormal and gamma models were constructed, applying both frequentist and Bayesian statistics in order to evaluate whether different inference approaches produce meaningfully different results. Two-stage models first analyse the presence-absence data using a binomial distribution and then evaluate the abundance data, given presence. GAMs were fitted following a stepwise forward method, starting from models with only one variable and subsequently adding significant variables by means of restricted maximum likelihood (REML; Wood, 2011). REML is more efficient than other available methods like general cross-validation (Marra and Wood, 2011). The degree of smoothness of each particular variable was limited in order to avoid overfitting, i.e. a maximum of 3 knots for single variable relationships and 9 knots for interactions between two variables. Bayesian models were fitted with uninformative priors to set unbiased expectations and maximize the influence of field data on the model outputs. Ten thousand post-warm-up samples distributed in four chains were drawn and thinned at 1/10 to prevent sample-autocorrelation. All models presented good mixing of chains and 99-100% of the Pareto k diagnostic values were <0.7, indicating moderate to good model performance (Vehtari *et al.*, 2017).

Additionally, a one-stage GAM model was constructed using a tweedie distribution. Tweedie distributions are based on probabilities and do not use any explicit analytic form of the density function. The tweedie distribution is a specific class of exponential dispersion model, defined by a power relationship between the mean (μ) and variance (V), of the form $V(\mu) = \mu p$ for some p (Tweedie, 1984; Dunn and Smyth, 2005).They exist for all values of p and the commonly used models like the normal distribution, the Poisson or the gamma are characterised by the value of p, i.e. 0, 1 and 2 respectively. Delta models have been widely used in CPUE standardisation of zero inflated datasets, frequently using generalised linear models (Forrestal *et al.*, 2019) and there are also examples of tweedie models being applied with the same aim (Shono, 2008), although generally using a frequentist modelling approach. Following Ingram *et al.*, (2015), a factor year variable was incorporated into each model to analyse the inter-annual variability of the larval index.

Integration of different forms of the explanatory variable related to sea temperature variability

Four alternative approaches were tested for each of all modelling strategies, including: i) only the variable year as a factor, ii) a linearly detrended temperature variable (tempres2), iii) the annual anomaly of temperature (Tanom) and iv) the absolute value of temperature (TMLD). This was performed to avoid correlated variables to be included in the same model but that still can have some ecological meaning. Model selection was adapted to each statistical approach and was not applied neither to the REFERENCE.MODEL larval index nor to the models including only year as a factor variable. Classical F-tests were applied to compare frequentist models, and variables were considered significant at 0.95 (p-value < 0.05). Variable selection of the Bayesian models was based on the leave-one-out cross-validation and widely applicable information criteria (Vehtari *et al.*, 2017). To compare model performance across all modelling approaches, the root mean squared error (RMSE) was computed for each model using the differences between fitted and observed values.

Comparison larval index model precession

The annual means for the larval index and confidence interval was calculated for each of all modelling approaches (20 in total) using marginal means, through the emmeans R package (Lenth R., 2020). Marginal means uses the mean of the continuous variables included in the model to predict mean estimates of the independent variable. Coefficient of variation of each annual mean estimate was also computed from the standard error estimate provided by the emmeans function. When the response variable was log-transformed, the back transformation of the standard error was processed within the emmeans.

Robustness of the larval index to changes in sampling intensity

The robustness of the larval index computed with the REFERENCE.MODEL to changes in sampling intensity was analysed to assess the effects of using years in the larval index with small area coverage and for design of future sampling. In particular, two tests were performed:

i) To assess the adequacy of using interpolated years for the larval index three Index values were calculated using the "REFERENCE.MODEL" model but removing a 25%, 50% and 75% percentage of the samples in year 2012 (always ensuring the high abundance area were included in the model, i.e. between 0.7-1.7º longitude and 38-39º latitude and ii), then this results were compared to the index in 2012 when all data was applied (**Figure 2**)

ii) To assess how a larval index would be if only the south of the archipelago would be sampled we subset the sampling area of all the years to an area where larvae of Atlantic bluefin tuna is recurrently present, i.e. 1 - 4.6º longitude and 38-39.5º latitude (**Figure 2**).

All modelling and data analysis was conducted using the R software (R Core Team, 2019), using the packages "mgcv", "stanarm" v2.18.2 and "brms" v2.9.0 for modelling (Wood, 2017; Becker *et al.*, 2018; Bürkner, 2017) and "maptools" for mapping (Bivand & Lewin-Kok, 2019).

Selection of the revised version of the larval index, SE and CV for sensitivity analyses

For selecting one model to be applied as revised version of the larval index we calculate the mean of all standardized model values for each year, and identify which of the 20 models is more similar to this mean. This analyses is performed with values after standardization of each model, due to the different scale of the index when logarithmic transformations are applied. For this "revised larval index" we provide the SE and CV computed from EMMEANS. We also provide the CV derived from comparing all models. We also applied the previous analyses of the effect of total area sampled to decide whether a particular year is going to be included in the time series.

3. Results

Model performance

The final selected models and their performance are shown in **Table 2**. In general, the inclusion of temperature and other environmental variables allow reducing RMSE in relation to models that only included the variable year as a factor (**Table 2**). However, the particular temperature variable (i.e. empres2, Tanom or TMLD) included does not make a great difference in the improvement in RMSE of the model outcomes.

Variability among standardization models and environmental indicators.

The outcomes of marginal mean estimates of Atlantic bluefin tuna estimated from the different models and approaches show a general increasing trend in abundance (**Figure 3**). In the period 2001-2005 low abundance is observe, with values generally ranging between 1.3 and 198.9 larvae/m^2 and generally around 25 larvae/m^2 (**Table 3**). This is followed by a period (2005-2010) of dramatically low abundances, sometimes with values close to larvae/m^2. Some models, however, predict a pick of abundance in the year 2008. In the year 2011, the abundance index showed signs of recovery to values observed in the first 5-year period. In the recent years, bluefin tuna larval indices show a remarkable increase. Almost all models detect the year 2017 as the one with the highest abundances. However, there is some degree of uncertainty in relation to the particular year in which other local maxima are observed, varying from one model to another. In general, models including tempres2 (approach 2) and TMLD (approach 4) tend to detect a partial maximum in the year 2013 while models including only the year as a factor variable (approach 1) or including Tanom (approach 3) usually identify the year 2012 as a local maximum and a second local pick in 2015 (**Figure 3**).

Comparison of larval index model precission

The precision of the abundance estimates follows a very similar temporal trend for all modelling approaches (**Figure 4**). In the initial period (2001-2005), CV varies from 22.1 to 92.4 %, generally around 46% while higher precision (lower CV) is achieved in the recent period (2011-2017) when CV ranges between 15.7 and 65.0 (**Table 4**), generally around 30.8 %. Among the non-standard surveys data, the years 2006, 2008 and 2010 show a general low precision and great variability among models, with values ranging between 55.0 and 152.4 %, while in the year 2011 precision is pretty acceptable (32.8-65.0%). In the year 2010, the few available samples explains the low precision while in the years 2006 and 2008 the reason behind the lack of precision could also be a lack of mismatch between the survey design and the area of high abundance.

Robustness of the larval** index **to changes in sampling intensity

The analysis of robustness of the larval index computed with the REFERENCE.MODEL shows that, provided that the sampling in the high abundance area is retained, a reduction of a 25 % of the sampling effort in 2012 does not modify the larval index abundance (**Figure 5**). When the reduction of the sampling effort is increased up to 50 or 75%, a potential overestimate of the larval abundance is observed in 2012 with increase in the dispersion of the data, even if these differences are not statistically significant. On the other hand, when the sampling area of the whole time series is reduced to the core area of Atlantic bluefin tuna spawning the larval index increases in the recent years but time series trend kept the same (**Figure 5**).

Selection of the revised version of the larval index, SE and CV for sensitivity analyses

Model fitting of the revised version of the larval index is provided in Annex I. The mean value of all models and the associated CI are presented in **Figure 6**. **Figure 7** shows the mean time series and the model that is more similar (this is the Delta-log normal with the "tempres2" environmental variable), and the values of the larval index submitted in January 2020 to the bluefin tuna Management Strategy Evaluation group to be integrated in the Operational Models (OMs) for comparison. Values of the larval index "Revised Version" SE and CVs associated are presented in **Table 6**.

The year 2010 was excluded from the time series as the analyses on the effects of sampling area reduction show that the number of data this year (9.0 % of the samples in 2003) was too low to be considered. The results from 2006 were also excluded; as this year presented extremely low values that may be associated to no low values in abundances but a change in fishing operations (fishing tows did not reach the mixed layer depth). Therefore the years 2008 and 2011 were considered for incorporation in the time series.

4. Discussion

Here we summarize the most relevant results from this study for understanding the sources of variability in the larval index (associated to inclusion of environmental variability ad modelling approaches) and for the design of sensitivity analyses in the MSE and VPA for bluefin tuna.

In relation to the effect of different environmental indicators. Incorporating environmental variables affecting the species reproduction, such as temperature, in larval abundance index standardization process is very relevant. The design of the environmental indicator derive in differences among model results, especially when applying gamma and tweedy modelling families (See **Figure 3**)

In relation to the variability associated to the different modelling approaches. The different modelling approaches showed a similar overall temporal trend of bluefin tuna abundance, with a general increase in the recent years and a historical maximum in 2017 (**Figure 3**), but with some models showing a different maximum. Delta log normal models showed the lowest model RMSE. The gamma families were the ones presenting higher deviances associated to few out layers (for example in year 2013) (see **Figure 3.B**). The Bayesian approach with logarithmic link was the one providing more different results in relation to all other models, with index values in the 2001-2005 period similar to the 2012-2016 period, besides, these Bayesian models did not show differences when environmental variables where included or excluded (**Figure 3.D**), so the model may not be correcting for differences in sampled habitats or dates. Exploring the inclusion of priors in the Bayesian models to investigate responses would be recommendable.

In relation to the variability associated differences in sampled area among years and the interpolation of years with no standard sampling area coverage.

An adequate match between the spawning area and the survey design as well as a sufficient amount of samples, are important factors introducing variability in the larval index (**Figure 5**). Years with a number of samples higher than 50% of those in 2012 (around 60) can be considered to be included in the larval index, but it is also important to consider the location of these samples that should located in areas where larval habitats were adequate, around the salinity front each year.

5. Conclusion

In relation to the selection of the final "Revised larval index" and the design of sensitivity analyses.

The proposed "revised larval index" is the model that showed higher similarity with the mean of all 20 other models. The selected modelling approach was a non linear delta-log normal model accounting for unbalanced factors and associated standard errors in EMMEANS.

It is relevant to point that as main result, the revised version of the larval index is almost the same as the one provided to the BFT-MSE group (R^2=0.99), but the coefficients of variation are higher. This is due to the different methods applied in each version for the back transformation of the standard error from the logarithmic scale. We recommend using the CVs associated to the revised version, where the logarithmic back-transformation of the SE was calculated within the emmeans. We also propose to consider the variability obtained in the analyses of how different models vary among them (all information included in **Table 6**).

6. Acknowledgement

This article has been developed in the framework of the European research Project PANDORA - PAradigm for Novel Dynamic Oceanic Resource Assessments, European Union's Horizon 2020 research and innovation programme under Grant Agreement No 77371, and the project Bluefin Tuna from SOCIB and IEO.

References

Álvarez-Berastegui D., Ingram Jr G.W., Reglero P., Ferrà C., Alemany F. (2018). Changes of bluefin tuna (*Thunnus thynnus*) larvae fishing methods over time in the Western Mediterranean, calibration and larval indices updating. Collective Volume of Scientific Papers of the International Commission for the Conservation of Atlantic Tunas (ICCAT) 74(6): 2772-2783, www.iccat.int/Documents/CVSP/CV074_2017/colvol74.html.

Álvarez-Berastegui D., Saber S., Ingram W.G.Jr, Díaz-Barroso L., Reglero P., Macías D., García-Barcelona S., Ortiz de Urbina J., Tintoré J., Alemany F. (2018). Integrating reproductive ecology, early life dynamics and mesoscale oceanography to improve albacore tuna assessment in the Western Mediterranean. Fisheries Research, 208C (2018) pp. 329-338. https://doi.org/10.1016/j.fishres.2018.08.014

Álvarez-Berastegui, D., Ingram Jr, G.W., Reglero, P., Ferrà, C., Alemany, F. (2017). Changes of bluefin tuna (*Thunnus thynnus*) larvae fishing methods over time in the western Mediterranean, calibration and larval indices updating. SCRS/2017/033

Anon. ICCAT-SCRS, (2017), REPORT OF THE 2017 ICCAT BLUEFIN STOCK ASSESSMENT MEETING, SCRS/2017/010, Collect. Vol. Sci. Pap. ICCAT, 74(6): 2372-2535 (2018), https://www.iccat.int/Documents/CVSP/CV074_2017/n_6/CV074062372.pdf

Becker, R.A., Wilks, A.R., Browninrigg, R., Minka, T.P., Deckmyn, A. (2018). Maps: draw geographical maps. R package version 3.3.0.

Bivand, R., Lewin-Koh, N. (2019). maptools: Tools for Handling Spatial Objects. R package version 0.9-8. https://CRAN.R-project.org/package=maptools

Bürkner, P-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. Journal of Statistical Software, 80 (1): 10.18637/jss.v080.i01

Dunn, P.K. and G.K. Smyth (2005) Series evaluation of Tweedie exponential dispersion model densities. Statistics and Computing 15:267-280

Forrestal, F., Schirripa, M., Goodyear, C.P., Arrizabalaga, H., Babcock, E.A., Coelho, R., Ingram, W., Lauretta, M., Ortiz, M., Sharma, R., Walter, J. (2019). Testing robustness of CPUE standardization and inclusion of environmental variables with simulated longline catch datasets. Fisheries Research, 210: 1-13

Ingram G.W., Jr., Alemany F., Álvarez D., García A. (2013). Development of indices of larval bluefin tuna (*Thunnus thynnus*) in the western Mediterranean sea. Collective Volume of Scientific Papers of the International Commission for the Conservation of Atlantic Tunas (ICCAT), 69(2): 1057-1076. www.iccat.int/documents/cvsp/cv069_2013/n_2/cv069021057.pdf

Ingram Jr., G.W., Alvarez-Berastegui, D., Reglero, P., Balbín, R., García, A., Alemany, F. (2017). Incorporation of habitat information in the development of indices of larval bluefin tuna (*Thunnus thynnus*) in the Western Mediterranean Sea (2001–2005 and 2012–2013). Deep Sea Res. Part II Top. Stud. Oceanogr. 140, 203–211.

Ingram JR., G.W., Richards, W.J., Lamkin, J.T., Muhling, B. (2010). Annual indices of Atlantic bluefin Tuna (*Thunnus thynnus*) larvae in the Gulf of Mexico developed using delta-lognormal and multivariate models. Aquat. Living Resour. 23, 35–47.

Ingram, G.W., Álvarez-Berastegui, D., García, A., Pollack, A.G., López-Jurado, J.L., Alemany, F. (2015). Development of indices of larval bluefin tuna (*Thunnus thynnus*) in the western Mediterranean Sea. ICCAT, Collect. Vol. Sci. Pap., 72 (3): 1279-1296

Marra, G., Wood, S.M. (2011). Practical variable selection for generalized additive models. Computational Statistics & Data Analysis, 55 (7): 2372-2387

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Russell Lenth (2020). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.4.5. https://CRAN.R-project.org/package=emmeans

Shono, H. (2008). Application of the Tweedie distribution to zero-catch data in CPUE analysis. Fisheries Research, 93: 154-162

Torres, A. P., Dos Santos, A., Balbín, R., Alemany, F., Massutí, E., Reglero, P. (2014). Decapod crustacean larval communities in the Balearic Sea (western Mediterranean): seasonal composition, horizontal and vertical distribution patterns. Journal of Marine Systems, 138: 112-126

Tweedie, M. C. K. (1984). An index which distinguishes between some important exponential families. Statistics: Applications and New Directions. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference (Eds. J. K. Ghosh and J. Roy), pp. 579-604. Calcutta: Indian Statistical Institute.

Vehtari, A., Gelman, A., Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross validation and WAIC. Statistics and Computing, 27: 1413-1432

Wood, S.N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society (B) 73(1):3-36

Wood, S.N. (2017). Generalized Additive Models: An Introduction with R (2nd edition). Chapman and Hall/CRC.

Table 1. Dataset used in the analysis of bluefin tuna larval index and characteristics of the surveys. Squared are years with the newly processed data; spatial coverage: is the percentage of area covered in relation to the year with wider sampled area (2003).

Year	Systematic grid	Gear	Nº samples	Survey dates	Spatial coverage (%)
2001	Y	B60	162	16/06 - 07/07	81.8
2002	Y	B60	171	07/06 - 28/06	86.3
2003	Y	B60	198	03/07 - 29/07	100.0
2004	Y	B60	166	18/06 - 08/07	83.8
2005	Y	B60	186	27/06 - 23/07	93.9
2006	Y	B90	51	17/06 - 14/07	25.7
2007	-	-	-	-	-
2008	N	B90	41	29/07 – 11/08	20.7
2009	-	-	-	-	-
2010	N	B60	18	18/06 – 19/06	9.0
2011	N	B90	84	14/05 – 17/07	42.4
2012	Y	B90	153	21/06 – 08/07	77.2
2013	Y	B90	124	20/06 – 10/07	62.6
2014	Y	B90	92	13/06 – 30/06	46.4
2015	Y	B90	94	23/06 – 09/07	47.4
2016	Y	B90	95	21/06 – 07/07	47.9
2017	Y	B90	92	26/06 – 12/07	46.4

Table 2. Model formulation and performance of the evaluated models, showing mean squared error (MSE) of the models. BN: binomial; Gau: Gaussian; tw: tweedie; Zt: zero truncated.

Type of model	Approach	Formula	Family	MSE
PREVIOUS-LIKE		lpres~as.factor(year)+s(lat,lon,k=9)+s(jd,k=3)+s(SMLD,k=3)+s(residualtemp,k=3) log(BFTab_gs)~as.factor(year)+s(lat,lon,k=9)+s(TMLD,k=3)+s(SMLD,k=3)+s(hournorm,bs="cc", k=7)	BN ("logit") Gaussian	85.9
GAM Delta-lognormal	1	lpres ~ as.factor(year) log(BFTab_gs) ~ as.factor(year)	BN ("logit") Gaussian	86.5
	2	lpres ~ as.factor(year)+s(jd, k=3)+s(SALanom, k=3)+ s(tempres2, k=3) log(BFTab_gs) ~ as.factor(year)+s(lon, lat, k=9)+s(SALanom, k =3)+s(tempres2, k=3)	BN ("logit") Gaussian	86.1
	3	lpres ~ as.factor(year)+s(lon, lat, k=9)+s(jd, k=3)+ s(SALanom, k=3)+s(Tanom, k=3) log(BFTab_gs) ~ as.factor(year)+s(lon, lat, k=9)+s(SALanom, k=3)	BN ("logit") Gaussian	86.1
	4	lpres ~ as.factor(year)+s(SALanom, k=3)+s(TMLD, k=3) log(BFTab_gs) ~ as.factor(year)+s(SALanom, k=3)+s(TMLD, k=3)	BN ("logit") Gaussian	86.1
GAM Delta-gamma	1	lpres ~ as.factor(year) BFTab_gs~ as.factor(year)	BN ("logit") Gamma ("log")	104.4
	2	lpres~as.factor(year)+s(jd, k=3)+s(SALanom, k=3)+s(tempres2, k=3) BFTab_gs ~ as.factor(year)+s(lon,lat, k=9)+s(SALanom, k=3)+ s(tempres2, k=3)	BN ("logit") Gamma ("log")	102.8
	3	lpres ~ as.factor(year)+s(lon,lat, k=9)+s(jd, k=3)+s(SALanom, k=3)+s(Tanom, k=3) BFTab_gs ~ as.factor(year)+s(lon,lat, k=3)+s(SALanom, k=3)	BN ("logit") Gamma ("log")	102.9
	4	lpres ~ as.factor(year)+s(TMLD, k=3)+s(SALanom, k=3) BFTab_gs ~ as.factor(year)+s(TMLD, k=3)+s(SALanom, k=3)	BN ("logit") Gamma ("log")	101.6
GAM-Tweedie	1	BFTab_gs ~ as.factor(year)	Tweedie ("log")	134.4
	2	BFTab_gs ~ as.factor(year)+s(lon, lat, k=9)+s(jd, k=3)+s(SALanom, k=3)+s(tempres2, k=3)+s(hournorm,	Tweedie ("log")	132.3
	3	BFTab_gs ~ as.factor(year)+s(lon, lat, k=9)+s(jd, k=3)+ s(SALanom, k=3)+s(Tanom, k=3)+s(hournorm, bs	Tweedie ("log")	134.8
	4	BFTab_gs ~ as.factor(year)+s(lon, lat, k=9)+s(TMLD, k=3)+s(hournorm, bs="cc", k=7)+s(SALanom, k=3)	Tweedie ("log")	126.9
Bayesian Delta-lognorm	1	lpres~as.factor(year) BFTab_gs~as.factor(year)	BN ("logit") Gaussian ("log")	130.3
	2	lpres ~ s(lon, lat,k=9)+as.factor(year)+s(jd,k=3)+s(tempres2,k=3)+s(SALanom,k=3) BFTab_gs ~ s(lon, lat,k=9)+as.factor(year)+s(tempres2,k=3)+s(SALanom,k=3)	BN ("logit") Gaussian ("log")	127.8
	3	lpres ~ s(lon, lat,k=9)+as.factor(year)+s(jd,k=3)+s(Tanom,k=3)+s(SALanom,k=3)	BN ("logit")	127.8

		BFTab_gs ~ s(lon, lat,k=9)+as.factor(year)+s(SALanom,k=3)	Gaussian ("log")	
	4	lpres ~ s(lon, lat,k=9)+as.factor(year)+s(jd,k=3)+s(TMLD,k=3)+s(SALanom,k=3)	BN ("logit")	126.6
		BFTab_gs ~ s(lon, lat,k=9)+as.factor(year)+s(jd,k=3)+s(TMLD,k=3)+s(SALanom,k=3)	Gaussian ("log")	
Bayesian Delta-gamma	1	lpres~as.factor(year)	BN ("logit")	137.0
		BFTab_gs~as.factor(year)	Gamma (“log”)	
	2	lpres ~ s(lon, lat,k=9)+as.factor(year)+s(jd,k=3)+s(tempres2,k=3)+s(SALanom,k=3)	BN ("logit")	137.6
		BFTab_gs ~ s(lon, lat,k=9)+as.factor(year)+s(hournorm, bs ="cc",k=3)+s(jd,k=3)+s(tempres2,k=3)	Gamma (“log”)	
	3	lpres ~ s(lon, lat,k=9)+as.factor(year)+s(jd,k=3)+s(Tanom,k=3)+s(SALanom,k=3)	BN ("logit")	138.0
		BFTab_gs ~ s(lon, lat,k=9)+as.factor(year)+s(jd,k=3)+s(Tanom,k=3)	Gamma (“log”)	
	4	lpres ~ s(lon, lat,k=9)+as.factor(year)+s(jd,k=3)+s(TMLD,k=3)+s(SALanom,k=3)	BN ("logit")	132.4
		BFTab_gs ~ s(lon, lat,k=9)+as.factor(year)+s(jd,k=3)+s(TMLD,k=3)+s(SALanom,k=3)	Gamma (“log”)	

Table 3. Marginal mean abundance estimates of bluefin tuna larvae (N larvae / m^2) computed by emmeans. In bold: local maxima of the time series.

	MSE Strict update	GAM: 2-stage lognormal				2-stage gamma				Tweedie				Bayesian: 2-stage lognormal				2-stage gamma			
Year		**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
2001	8.2	3.3	5.6	5.8	5.3	6.6	8.4	7.8	9.9	6.6	7.9	8.5	11.2	19.6	20.3	21.6	15.5	41.8	26.0	25.7	26.5
2002	**17.2**	2.1	9.6	2.6	11.3	4.2	19.7	1.3	41.5	4.2	11.0	2.9	36.4	20.3	15.5	17.2	14.3	42.6	26.5	8.0	56.1
2003	6.2	4.4	3.3	**12.5**	6.3	8.2	4.0	**50.5**	26.0	8.2	9.1	**54.3**	12.8	**35.7**	27.1	**34.7**	**39.2**	71.9	50.1	**198.9**	87.4
2004	**13.0**	**5.4**	**11.5**	7.1	**11.6**	**28.9**	**76.9**	32.2	**53.0**	**28.9**	**22.2**	14.7	**35.7**	35.4	**30.7**	27.9	21.9	**190.2**	**123.7**	74.7	**93.2**
2005	4.4	3.2	2.9	5.5	3.5	8.2	5.1	15.2	12.4	8.2	6.3	14.9	9.8	15.7	11.2	14.2	11.0	40.7	23.0	38.4	25.2
2006	0.6	0.4	0.3	0.4	0.5	0.7	0.4	0.6	1.0	0.7	0.5	0.8	0.9	1.7	1.1	1.4	1.2	3.0	2.3	3.3	2.4
2008	3.0	0.4	2.4	2.2	0.3	0.5	4.7	19.8	3.9	0.5	48.1	231.8	0.7	2.2	3.8	2.2	6.3	2.6	26.7	78.5	14.1
2010	5.7	0.8	2.6	1.2	4.2	0.9	4.0	0.7	10.1	0.9	5.4	1.5	7.7	2.9	2.6	1.9	2.7	4.1	4.3	1.1	4.4
2011	9.1	4.3	6.8	4.9	7.7	25.2	35.1	18.8	70.6	25.2	43.7	32.0	38.9	13.6	10.2	10.0	10.0	81.6	71.9	60.2	58.9
2012	45.9	**28.8**	30.7	**37.8**	**37.2**	**212.9**	142.6	**200.3**	243.4	**212.9**	158.6	**257.5**	249.9	**41.5**	30.3	**38.0**	24.6	**310.9**	146.3	**189.2**	120.5
2013	**48.8**	15.7	**35.0**	19.1	**37.2**	178.0	**560.1**	163.4	**368.1**	178.0	**369.2**	127.6	**494.1**	25.2	**32.0**	20.6	**30.4**	287.7	**427.4**	139.7	**284.3**
2014	28.6	12.7	22.2	20.6	24.3	85.2	111.7	62.6	165.2	85.2	172.4	120.7	182.3	23.8	20.1	21.7	15.1	164.9	104.8	69.6	87.0
2015	**56.5**	**39.9**	**39.2**	**50.6**	**49.6**	196.3	132.0	222.9	236.2	196.3	147.3	286.6	245.8	**49.8**	36.0	**45.4**	29.4	245.6	138.5	224.8	123.1
2016	42.8	16.2	32.8	24.6	29.4	185.4	346.4	185.2	279.2	185.4	488.4	276.0	465.1	26.0	33.7	28.5	27.4	302.3	497.8	295.7	293.6
2017	**118.9**	**76.1**	**85.3**	**108.5**	**104.8**	**489.6**	**348.3**	**669.7**	**590.1**	**489.6**	**448.3**	**902.6**	**644.7**	**98.4**	**86.1**	**100.9**	**74.7**	**644.0**	**512.3**	**870.2**	**438.4**

1: year; 2: tempres2; 3: Tanom; 4: TMLD

Table 4. Precision (Coefficient of Variation, in %) of the annual marginal mean abundances of bluefin tuna larvae computed by emmeans.

		GAM:												Bayesian:							
	MSE Strict	2-stage lognormal				2-stage gamma				Tweedie				2-stage lognormal				2-stage gamma			
Year	update	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
2001	46.1	39.7	40.7	41.0	40.4	61.4	58.7	59.5	55.1	28.4	36.5	34.9	34.0	35.9	37.0	37.8	39.6	32.8	36.9	37.9	37.6
2002	57.9	48.6	48.0	61.0	48.7	74.8	71.1	92.4	68.4	29.4	47.3	39.0	41.5	49.2	48.2	47.5	54.5	42.1	59.1	65.2	68.3
2003	72.9	42.3	52.5	47.1	59.5	65.1	67.5	78.5	75.7	25.0	48.1	39.3	50.1	40.0	44.3	40.8	61.4	34.5	48.4	55.7	54.1
2004	48.6	40.5	40.9	42.2	40.3	62.6	60.4	61.0	55.9	22.9	34.1	32.6	32.2	38.4	40.7	38.1	44.7	34.2	45.5	44.0	41.4
2005	47.1	33.5	37.6	34.5	37.8	51.9	54.3	51.3	51.0	25.7	38.6	34.8	35.9	31.0	35.1	30.8	38.9	26.9	32.4	33.8	33.3
2006	**67.6**	**58.3**	**60.1**	**59.5**	**61.5**	**90.4**	**87.3**	**88.8**	**84.8**	**69.5**	**73.7**	**72.9**	**75.0**	**62.3**	**59.8**	**58.8**	**62.6**	**55.3**	**57.3**	**55.0**	**56.6**
2008	**79.6**	**70.6**	**74.5**	**66.8**	**75.5**	**108.9**	**115.4**	**146.4**	**102.6**	**81.4**	**116.4**	**111.8**	**86.1**	**74.3**	**87.3**	**77.5**	**152.4**	**65.4**	**127.6**	**147.3**	**127.9**
2010	**93.7**	**86.6**	**84.9**	**85.8**	**86.4**	**134.0**	**132.3**	**134.8**	**128.3**	**111.7**	**111.4**	**109.8**	**113.2**	**96.2**	**106.1**	**102.3**	**110.3**	**99.4**	**95.8**	**106.3**	**109.2**
2011	46.0	39.1	38.5	41.1	40.1	61.1	58.0	65.0	57.1	32.8	40.0	40.2	37.7	37.1	39.2	37.3	43.2	33.0	37.7	44.6	39.9
2012	30.3	19.0	21.5	19.5	20.2	30.6	32.2	29.4	28.1	18.1	30.6	26.6	27.1	17.9	21.8	19.0	24.4	15.7	23.0	20.0	22.3
2013	38.1	22.3	30.3	22.5	26.2	35.8	53.9	34.4	40.7	20.6	36.4	27.8	32.2	21.8	36.2	21.2	38.8	18.5	37.0	25.3	34.0
2014	32.6	28.8	28.7	29.5	28.9	45.7	43.1	50.5	40.7	26.5	35.0	30.5	30.6	27.8	30.1	26.6	35.4	23.7	33.4	36.4	32.4
2015	28.6	22.1	24.7	22.3	23.2	35.9	38.1	34.6	33.2	23.4	31.3	27.3	28.6	21.3	25.9	21.5	27.4	19.2	25.7	22.9	24.5
2016	32.3	25.7	27.1	26.3	26.0	41.1	41.9	39.7	37.7	23.4	28.7	27.9	28.0	24.9	27.1	25.1	30.0	21.7	28.7	28.8	33.5
2017	34.4	23.0	25.4	23.9	25.3	37.2	37.0	38.0	35.2	20.8	29.4	26.5	28.4	22.9	25.1	22.6	27.4	19.1	25.9	29.1	26.6

1: year; 2: tempres2; 3: Tanom; 4: TMLD

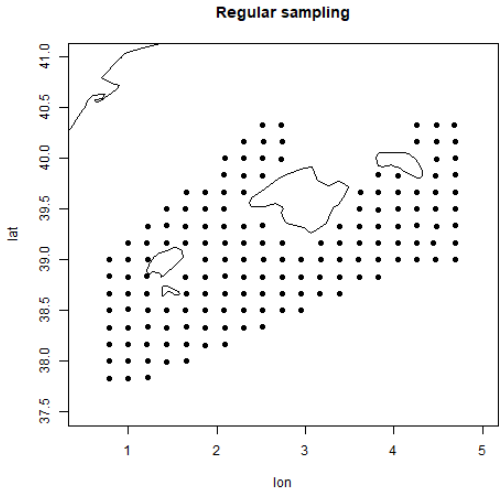
Table 5. RMS of each model (sample by sample prediction of the two-submodels against real data), by quartiles of the larval index.

Quartile	MSE Strict update	Delta - lognormal				Delta - gamma				Tweedie				Bayesian hurdle-lognormal				Bayesian hurdle-gamma			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Q1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	6.3	6.0	6.0	6.1	20.9	20.3	19.2	20.1	7.2	7.5	7.6	10.4
Q2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	14.5	8.6	8.5	9.8	22.7	18.5	18.8	18.4	16.2	12.8	13.1	13.2
Q3	6.8	7.0	6.7	6.8	6.9	55.7	52.2	52.5	51.4	155.3	137.5	136.7	126.8	126.0	114.0	116.0	113.0	161.0	157.0	158.0	141.0
Q4	336.1	338.6	337.0	337.2	336.9	361.3	358.4	358.6	354.4	361.3	377.0	376.0	364.5	352.0	359.0	358.0	356.0	364.0	373.0	374.0	366.0
Global	85.9	86.5	86.1	86.1	86.1	104.4	102.8	102.9	101.6	134.4	132.3	134.8	126.9	130.3	127.8	127.8	126.6	137.0	137.6	138.0	132.4

Table 6. Values, SE and CV of the revised version of the index, CV associated to variability among the 20 models, and values of the MSE and VPA strict update indices provided in January 2020.

YEAR	Gear	Revised index (R.I)	SE R.I	CV R.I	CV inter-models	MSE.strickt update	VPA.strickt update
2001	B60	5,63	2,29	40,74	6,19	3,33	3,48
2002	B60	9,57	4,59	47,99	5,25	6,07	3,12
2003	B60	3,25	1,71	52,45	5,95	5,03	2,38
2004	B60	11,47	4,69	40,90	4,26	5,40	5,80
2005	B60	2,86	1,07	37,56	5,18	2,43	2,32
2006	-	-	-	-	-	-	-
2007	-	-	-	-	-	-	-
2008	B90	2,36	1,76	74,54	7,77	-	-
2009	-	-	-	-	-	-	-
2010	-	-	-	-	-	-	-
2011	B90	6,82	2,63	38,49	2,42	-	-
2012	B90	30,65	6,58	21,48	1,19	22,66	29,62
2013	B90	34,98	10,60	30,30	2,49	24,76	16,29
2014	B90	22,24	6,38	28,68	1,47	13,34	14,80
2015	B90	39,21	9,70	24,74	1,47	26,99	40,20
2016	B90	32,83	8,90	27,12	1,95	20,07	16,95
2017	B90	85,29	21,64	25,37	0,81	58,84	74,05

(A)



(B)

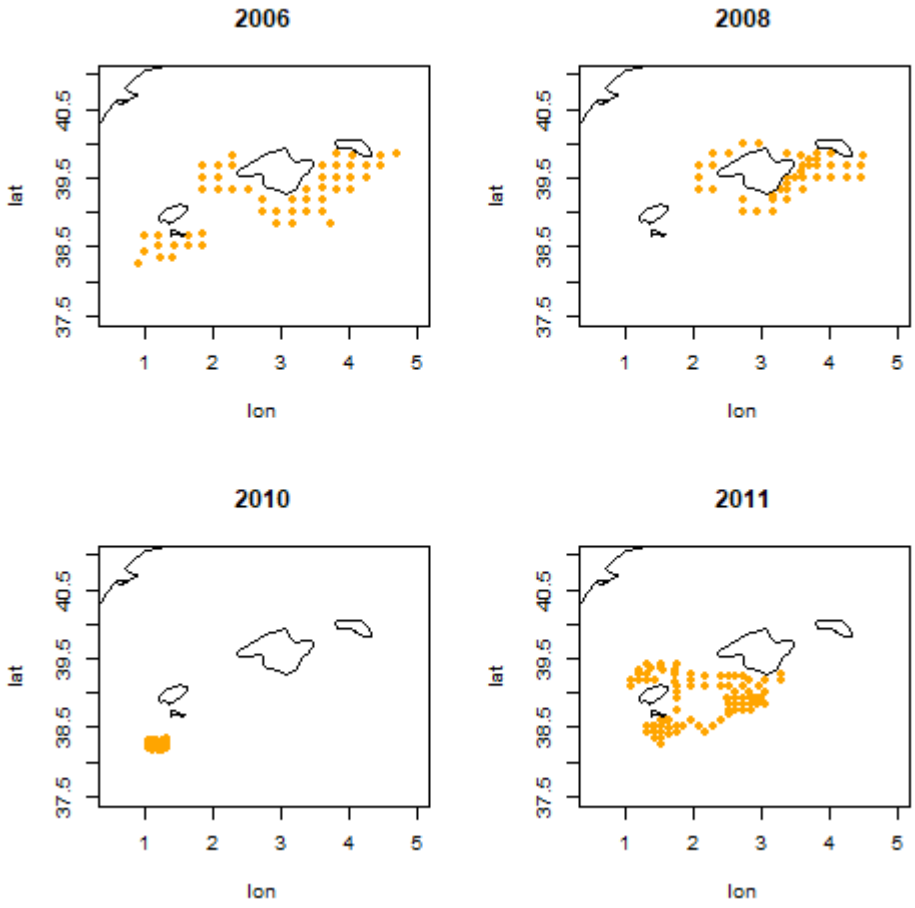


Figure 1. Sampling schemes for the: (A) standard and (B) non-standard surveys.

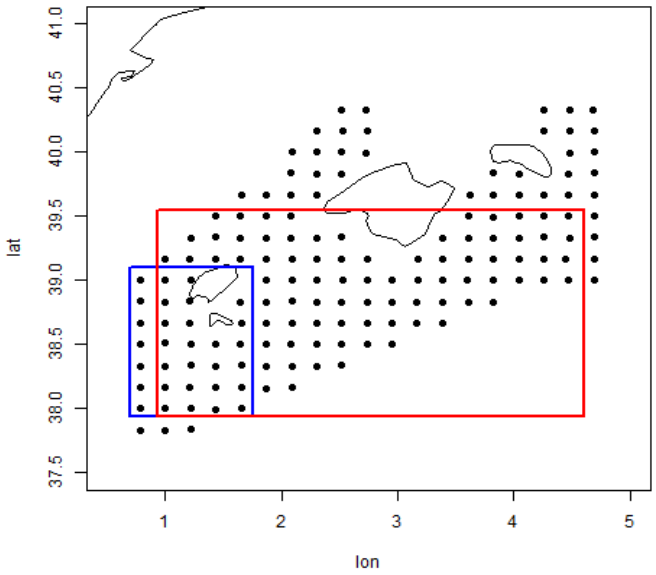


Figure 2. Areas identified for the analysis of the robustness of larval index to sampling intensity. In blue, area for the test on reducing the data in the year 2012 and in red, area for the reduction of the sampling area to bluefin tuna spawning area.

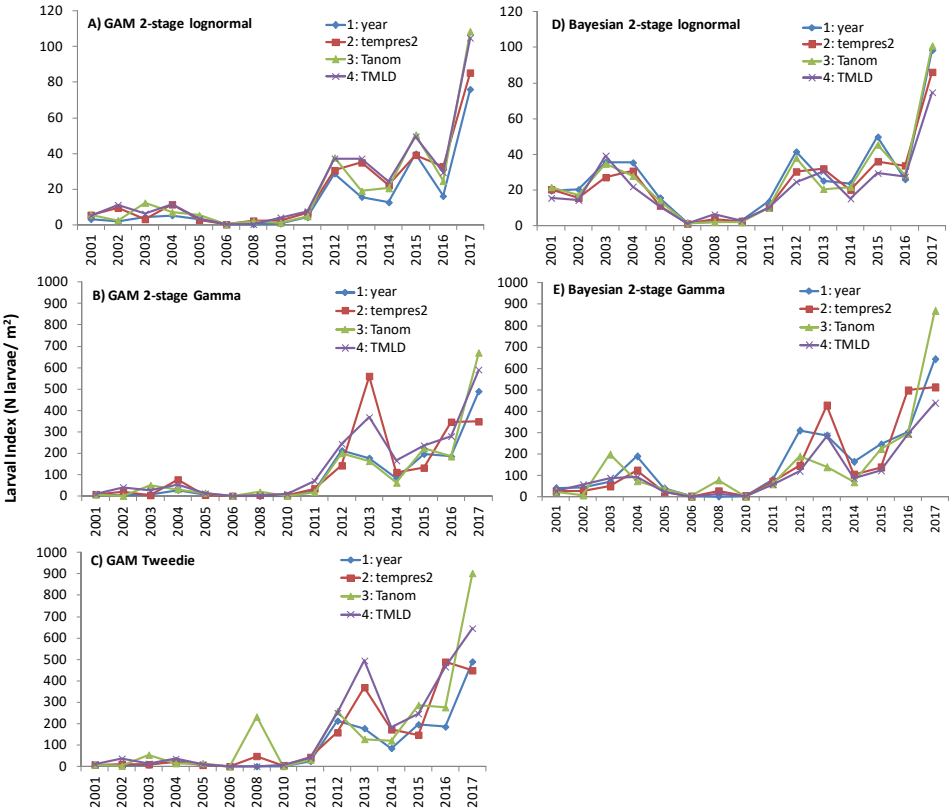


Figure 3. Bluefin tuna larval index for each modelling strategy and approach. Note that each y-axis has different ranges.

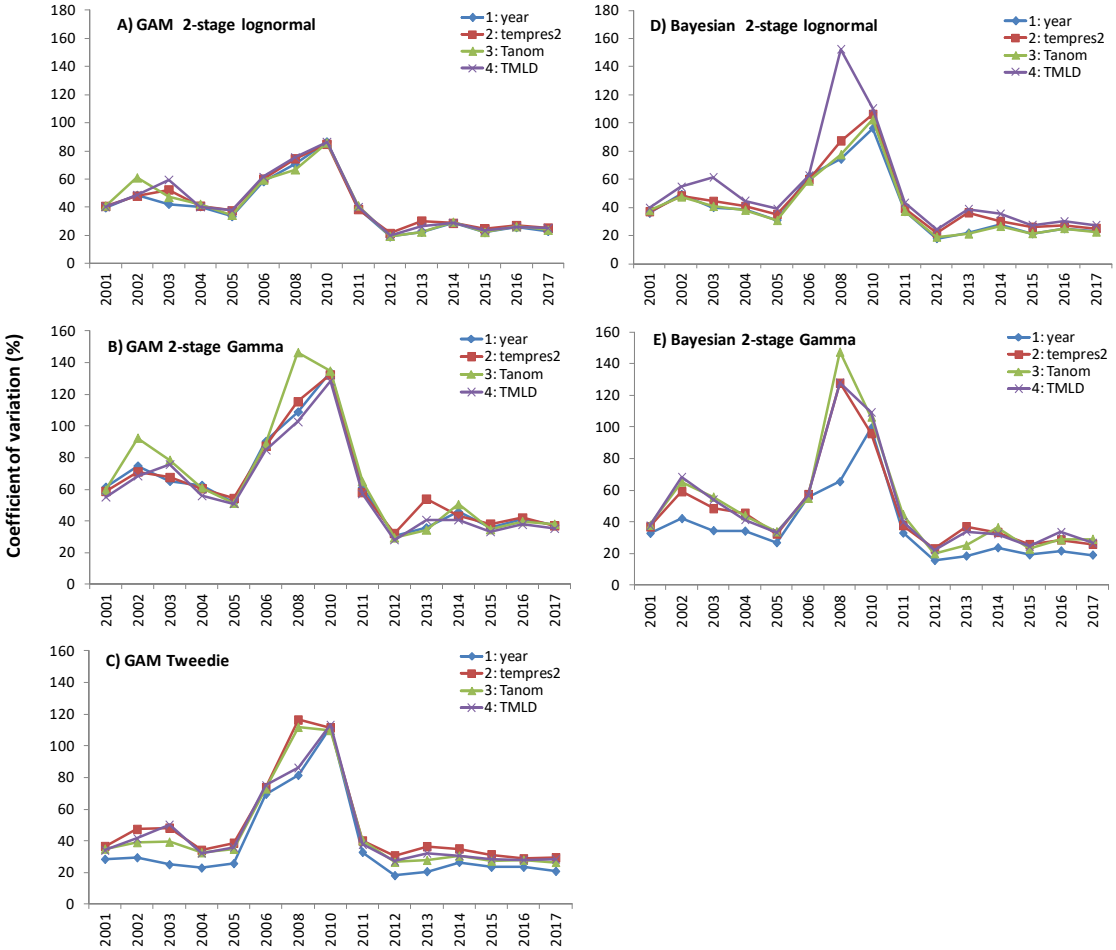
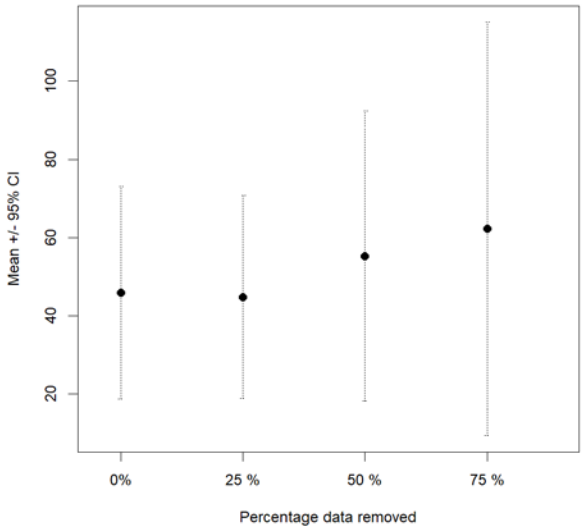


Figure 4. Precision of the bluefin tuna abundance indices for each modelling strategy and approach. Note that each y-axis has different ranges.

(A)



(B)

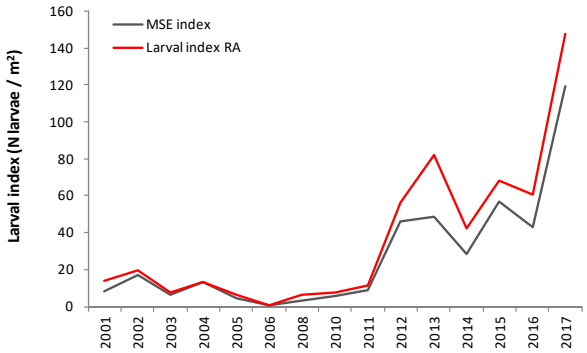


Figure 5. Robustness of the larval index to varying sampling intensity: (A) index and confidence limits of the year 2012, for the REF.MODEL where no samples are removed(0%), and test cases when 25%, 50% and 75% of the samples are removed in that year, (B) comparison of the REF.MODEL and the index calculated when subletting the sampling area for all the years to an area where bluefin tuna larvae is recurrently encountered (between 1 - 4.6º longitude and 38-39.5º latitude, see red area in Figure. 2) (Larval index RA).

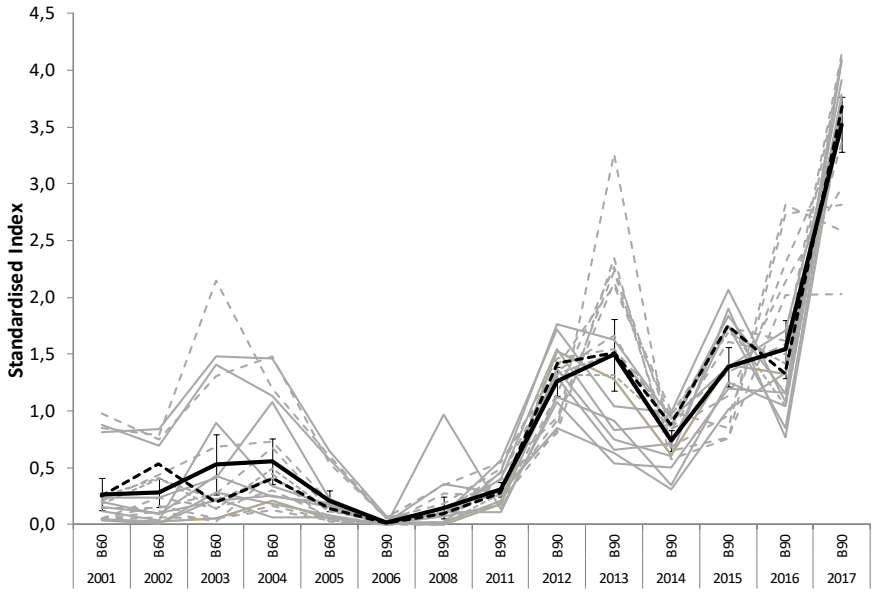


Figure 6. Mean of standardized values for all models and CI (95%), dotted line is the result of a delta-lognormal model with same variables as the MSE strict updated but resolved with EMMEANS.

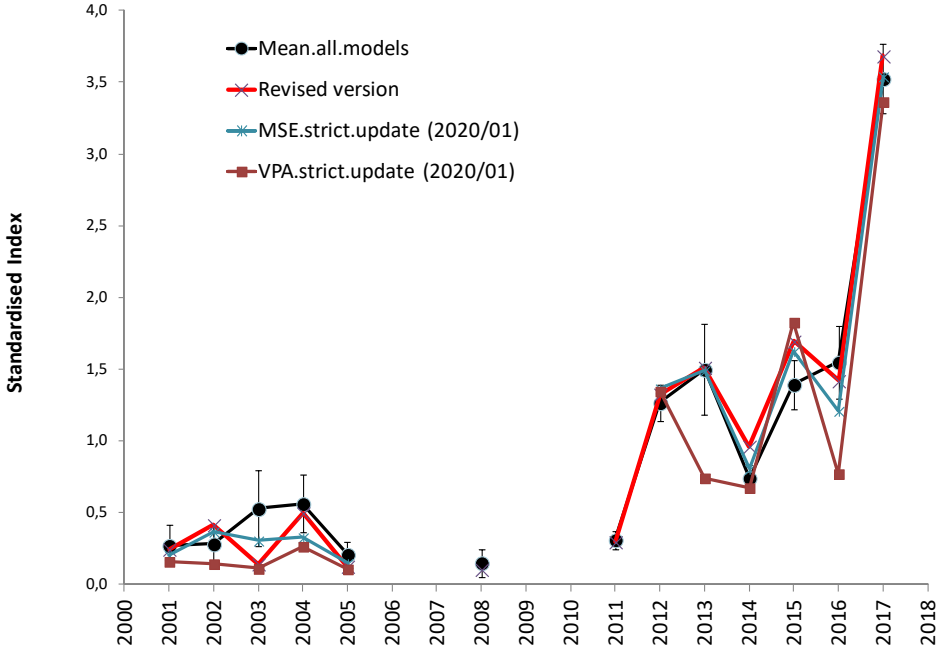


Figure 7. Black: Mean values from all models and confidence limits associated to the variability among models ; **Green:** "Revised version index" ; **Red:** Current larval index submitted to the Management Strategy Evaluation in January 2020 (NOTE: for the index submitted to MSE no values are available for years 2007, 2009, 2010)

Annex I. Model fitting of the larval index "revised version" model

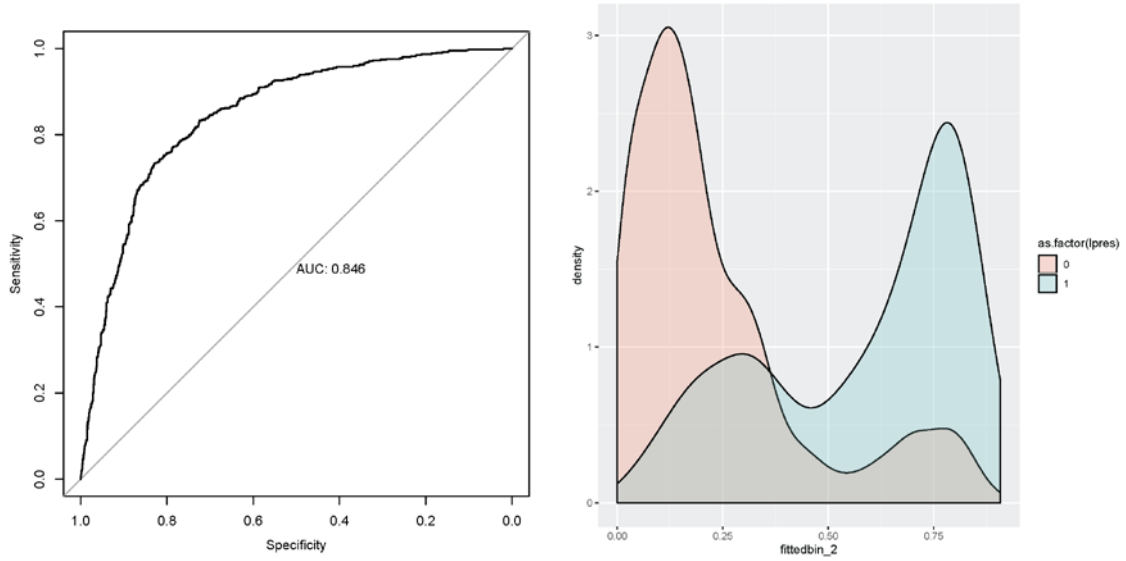


Figure AI.1. Model fit of the binomial stage of the "revised version" of bluefin tuna larval index model: plot of receiver operating characteristic, with Area Under the Curve (AUC) value (left) and density plot of real positive and negative estimates (right).

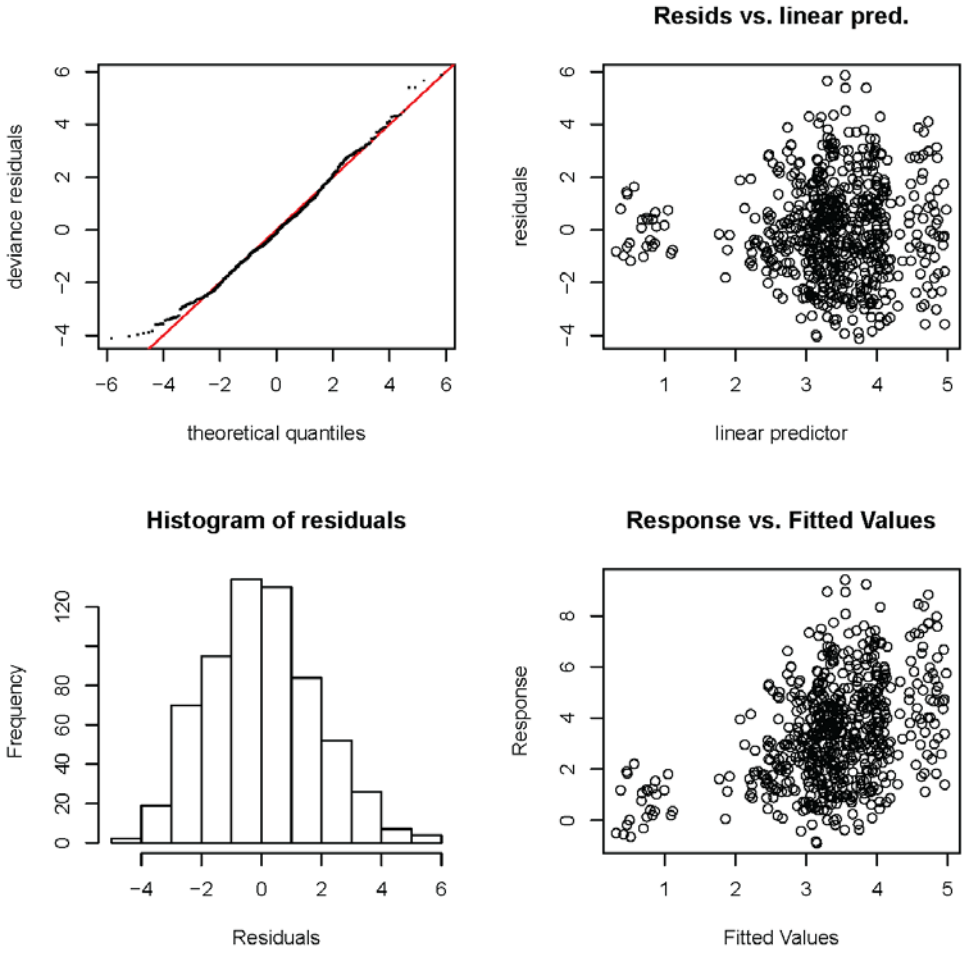


Figure AI.2. Model fit of the abundance part of the "revised version" of bluefin tuna larval index model: deviance of the residual versus theoretical quantiles (top left), residuals versus linear predictor (top right), histogram of the residuals (bottom left), response versus fitted values (bottom right).