SCRS/2020/163 Collect. Vol. Sci. Pap. ICCAT, 77(2): 988-1005 (2021)

REVIEW OF THE REVISION OF GBYP AERIAL SURVEY DESIGN, IMPLEMENTATION AND STATISTICAL ANALYSES (ICCAT GBYP 12/2020) OF THE ATLANTIC-WIDE RESEARCH PROGRAMME FOR BLUEFIN TUNA (ICCAT GBYP Phase 10)

Jon Helge Vølstad[1]

SUMMARY

The focus of this review is the GBYP aerial survey design, field methods, and methods employed in the 2019 re-analysis of the whole time series by Cañadas and Vázquez (2020). Several inconsistencies were found in the re-analysis results, suggesting errors in the R-script that needs to be corrected. There is strong evidence that a long-term monitoring program will require a survey design that covers much of the Mediterranean. An option is to expand the survey coverage outside the four main spawning areas by rotating the sampling of remaining spawning areas annually over time. Model-based methods could be used to combine data from the two survey components. We suggest the use of high-resolution video or digital photography and development of automatic image analysis through machine learning as an alternative to observers for collecting abundance data from standardized strip transects. Such methods could ensure standardized counts of individual animals (and their lengths) within a defined narrow transect width and could reduce cost and eliminate many of the sources of errors that are identified for the current field data collections with observers.

RÉSUMÉ

Cette étude porte sur la conception de la prospection aérienne du GBYP, les méthodes de terrain et les méthodes employées dans la nouvelle analyse de l'ensemble de la série temporelle de Cañadas et Vázquez (2020) en 2019. Plusieurs incohérences ont été constatées dans les résultats de la nouvelle analyse, ce qui suggère des erreurs dans le script R qui doivent être corrigées. Il est évident qu'un programme de surveillance à long terme nécessitera une conception de prospections couvrant une grande partie de la Méditerranée. Une option consiste à étendre la couverture des prospections en dehors des quatre principales zones de frai en faisant une rotation annuelle de l'échantillonnage des autres zones de frai au fil du temps. Des méthodes basées sur le modèle pourraient être utilisées pour combiner les données des deux composantes de prospection. Nous suggérons l'utilisation de la vidéo haute résolution ou de la photographie numérique et le développement de l'analyse automatique des images par apprentissage automatique comme alternative aux observateurs pour la collecte de données d'abondance à partir de transects en bande standardisés. Ces méthodes pourraient garantir un comptage standardisé des animaux individuels (et de leur longueur) dans une largeur de transect étroite définie et pourraient réduire les coûts et éliminer de nombreuses sources d'erreurs qui sont identifiées pour les collectes actuelles de données sur le terrain par les observateurs.

RESUMEN

Esta revisión se centra en el diseño de la prospección aérea del GBYP, los métodos de campo y los métodos empleados en el reanálisis de 2019 de toda la serie temporal por Cañadas y Vázquez (2020). Se encontraron varias incoherencias en los resultados del reanálisis, lo que sugiere errores en el script R que deben corregirse. Hay pruebas sólidas de que un programa de seguimiento a largo plazo requerirá un diseño de prospección que cubra gran parte del Mediterráneo. Una opción es ampliar la cobertura del estudio fuera de las cuatro zonas principales de desove, rotando el muestreo de las zonas de desove restantes anualmente a lo largo del tiempo. Se podrían utilizar métodos basados en modelos para combinar los datos de los dos componentes de la prospección. Sugerimos el uso de vídeo de alta resolución o fotografía digital y el desarrollo del análisis automático de imágenes a través del aprendizaje automático como alternativa a los observadores para la recogida de datos de abundancia en transectos de franja estandarizados. Dichos métodos podrían garantizar recuentos estandarizados de animales individuales (y sus tallas) dentro de un estrecho ancho de transecto definido y podrían reducir el coste y eliminar muchas de las fuentes de error que se identifican en las actuales recopilaciones de datos de campo con observadores.

---

[1] Strangehagen 22, 5011 Bergen, Norway; jonhelge@hi.no Representing Center of Independent Experts

KEYWORDS

aerial surveys; bluefin tuna; line transect sampling

1. Background

The BFT aerial survey conducted in the Mediterranean is one of the major activities of the Atlantic Wide Research Programme for Bluefin Tuna (GBYP). The BFT aerial survey was launched in 2010 with the purpose of obtaining a relative abundance index of spawning biomass for the Mediterranean Sea. The survey has been conducted in 2010, 2011, 2013, 2015, 2017, 2018, and 2018. The aim is to provide reliable indices of spawning stock abundance that can track trends over time and provide key input to stock assessments. Due to the large extent of the potential spawning area, over multiple jurisdictions, the survey has faced numerous logistical challenges and has had to alter its design and data processing protocols multiple times. Cañadas and Vázquez (2020) provides updated estimates of abundance, and total weight of the BFT spawning stock spanning the period 2010-2019. The re-analysis included adjustments of the main areas surveyed to ensure that annual estimates are provided for fixed areas. The updated estimates exhibit large variability, which has raised the question of whether the surveys are able to provide reliable data to inform management of the stocks. The updated estimates of abundance indices exhibit substantial differences from prior time series and the index exhibits high interannual variability both within and between regions. The magnitude of the difference between prior time series and the high variability has raised concerns regarding the estimation procedures and the overall efficacy of the survey to reflect annual spawner abundance in the Mediterranean Sea. Given the need to evaluate the survey and to soon take decisions regarding the nature of its continuation, ICCAT has requested an independent desk review of the survey design, statistical treatments and analytical procedures and of its general capacity to achieve its objectives. The main purpose of this report is to provide an independent review of the Mediterranean Sea Bluefin tuna aerial survey design and statistical analysis used in the development of an index of spawning stock biomass, with an emphasis on the 2019 re-analysis of the time series (Cañadas and Vázquez 2020). The Terms of Reference are provided in the Appendix.

2. Description of Role in the Review Activities

This CIE desk review was conducted independently (where Dr. Steve Buckland and I served as the CIE independent peer reviewers), with focus on the BFT aerial survey design, statistical treatments and analytical procedures, and of its general capacity to achieve its objectives. We have collective expertise and long experience in aerial survey design, statistical time series evaluation, and a strong understanding of population modeling and stock assessment. Dr. Buckland is a world renown expert on distance sampling applied to the estimation of animal abundance and has published standard reference textbooks and papers on this subject. Dr. Buckland has intimate knowledge of the Distance project, and the “Distance” software (Thomas *et al.* 2010) for the design and analysis of distance sampling surveys of wildlife populations. It is my understanding that the Distance software for Windows has been used in the survey design and prior data analysis of BFT aerial surveys. The 2019 re-analysis were conducted using R-packages from the Distance Project in an R-script developed by the authors.

I bring international research and management experience in quantitative fisheries biology and ecological statistics, specializing in statistical survey methods. I have broad hands-on experience in the development and optimization of fisheries-dependent and fisheries- independent monitoring programs to support stock assessments and ecosystem-based fisheries management. My experience with the design and analysis of acoustic-trawl surveys, with transects as primary sampling units, and aerial surveys and roving creel surveys of recreational fisheries are relevant for this review. In this review I have focused especially on aspects related to the survey design of the BFT aerial surveys and provide some thoughts on possible future improvements. For in-depth review of the statistical methods applied in the 2019 re-analysis of BFT aerial surveys I defer to Dr. Buckland´s independent CIE review.

Dr. Manoj Shivlani (CIE) provided comprehensive background material, including all historic analysis reports, reports on survey design, fields protocols, and prior reviews for this desk-top peer review through google drive and links to ICCAT websites via email.

3. Review of Methods

3.1 Survey Design

Stock assessment of the Atlantic bluefin tuna (BFT) is conducted separately for (1) the western Atlantic and (2) eastern Atlantic and Mediterranean stocks (ICCAT 2017). The eastern Atlantic stock (EBFT) mainly spawns in the Mediterranean Sea, and the western Atlantic stock (WBFT mainly spawns in the Gulf of Mexico (Fromentin and Powers, 2005). Although stock mixing occurs, this seems to have small effects on the stock assessments of the eastern BFT (Morse *et al*. 2017). The Atlantic-Wide Research Programme for Bluefin Tuna (GBYP) use Aerial surveys and line-transect DISTANCE sampling to estimate the relative abundance of the eastern BFT spawning stock in the Mediterranean. Reliable indices of abundance provide key input to stock assessments, for example as tuning series in Virtual Population Analysis, or as input to Statistical Assessment Models (SAMs).

Aerial surveys have been conducted in 2010, 2011, 2013, 2015, 2017, 2018, 2019. Sampling effort (total length of transects) have focused on four subareas that are assumed to represent the main spawning areas. Surveys with extended spatial coverage were conducted in 2013 and 2015 to cover the majority of the potential spawning areas in the Mediterranean Sea. The recognized DISTANCE software has been used in designing the annual surveys. The annual surveys were generally conducted in multiple rounds (2 – 4 rounds) for each of the main spawning areas to cover the main spawning period, with flights along equally spaced parallel transect lines that were selected with a random starting location in each round. This results in equal coverage probability in space over each subarea. In the expanded surveys conducted in 2013 and 2015 the areas outside the main spawning grounds were only covered in one survey round. One disadvantage of evenly spaced parallel transects (compared to a zig-zag design) is that some flying time is spent in transit between transects. There is no unbiased estimator of the variance for systematic sampling. However, in practice, variance estimates based on the assumption of simple random sampling of transects is likely to overestimate the true variance and is likely to provide more reliable estimates than simple random spacing of transects.

Depending on how much flying time is spent between transects, the alternative zig-zag design may be considered. Harbitz (2019) developed a randomized zigzag design for straight line and curved transects that guarantees equal coverage probability. This method has been used since 2018 in acoustic surveys of Norwegian Spring Spawning Herring (ICES 2018), treating each straight line in the zigzag design as randomly selected primary sampling units (Simmonds and MacLennan 2008), which is a fairly strong assumption. See also Skaug *et al*. (2004) for an example where double-platform shipborne visual sighting surveys are used to estimate the abundance of minke whales in the NE Atlantic, and where transects were constructed as zig-zag tracks with a random starting point. The main advantage of the zig-zag design is that the costly ship-time between equally spaced transects has nearly been eliminated.

MRAG (2016) lists some key factors that affects reliability of abundance indices for monitoring changes in abundance based on the aerial surveys:

a) degree of inter-annual variability in timing of spawning relative to the timing of surveys,
b) spatial distribution of fish particularly, if the distribution changes as a function of population size, and
c) behavioral factors such as time fish spend near the surface where they can be seen from aircraft.”

These factors mainly affect bias in annual estimates of abundance indices. Of particular concern is variable biases, which generally are very difficult to quantify and correct for. In addition to the above factors there are many factors related to the execution of the field data collections from aerial surveys that affects precision and bias.

Precision of estimates will primarily be determined by the survey design, sampling effort, detectability, and choice of estimators.

MRAG (2016) also list other challenges that largely cannot be controlled in the BFT aerial surveys: “In addition, there are serious logistic challenges with aerial surveys. For example, surveying some areas originally included in the survey design is not feasible because of security concerns in areas near military conflicts. There have also been problems obtaining authorizations to survey within the airspaces of some Mediterranean CPCs, sometimes causing delays that adversely impact field programmes. In light of some or all of these concerns, the fifth aerial survey of the Mediterranean Sea in 2016 was cancelled.”

A discussion of main sources of bias follows. Sources of uncertainty related to field observations are discussed in section 3.2.

3.1.1 Timing of spawning relative to timing of survey (a)

A literature review by Piccinetti (2013) show that the main spawning by BFT in the Mediterranean Sea occurs from mid-May to mid-July, with a peak in June, with limited variability in timing depending on oceanographic and environmental conditions. Alemany *et al*. (2010) suggest that the spawning BFT have preferences for waters with salinities between 36.9 and 37.7, and that spawners prefer temperature in the range of 21.5–26.5 °C. Clearly the aerial BFT surveys must be planned well in advance, which leaves few options to adjust the timing based on real-time information on salinity and temperature, if available. The multiple survey rounds in the aerial surveys may reduce biases caused by annual variations in timing. Surveys were designed as equal spaced parallel lines and so that the whole sub-area could be surveyed in two days and then repeated multiple times. The number of 2-day surveys planned for each sub-area was based on the size of the sub-area. The annual data analysis reports generally do not specify the timing of the survey's rounds within annual surveys, so it is difficult to assess if the timing covers the spawning season representatively each year.

3.1.2 Spatial coverage of survey

The survey coverage of the spatial distribution of the spawning stock is a major concern in this aerial survey, especially if the distribution changes over time, which is likely when abundance increases over time. An increase in the overall abundance of eastern BFT has been documented in recent years, and BFT on feeding migrations during summer have been observed in increasing numbers since 2012 in Norwegian waters, after several decades of absence (Nøttestad *et al*. 2020).

The 2010 GBYP report on Data recovery plan (Cañadas, Hammond & Vázquezis 2010) stated that "To minimize natural variation in using survey estimates as indices of abundance over time, surveys in future years should ideally occur in the same areas at the same time of year." A survey design with fixed subareas that only cover a small portion of the Mediterranean would be cost-effective, and could provide relative abundance indices, if they representatively cover a fixed proportion of total spawning stock over time. However, if the 4 inner subareas that have been surveyed every year in the time series only partially cover the spawning stock, and the coverage varies from year to year, then the abundance indices may not reliably track trends over time.

The GBYP Steering Committee raised the issue of incomplete coverage of the BFT spawning distribution in 2012. Surveys that expanded the coverage to cover much of the Mediterranean were conducted in 2013 and 2015 (Cañadas and Vázquez, 2013; 2015). Only areas with no historical data on spawning, and areas with closed airspace were excluded. The four main survey areas (Figures 1, 2, labeled with yellow) were covered with two survey rounds and denser transects each year (to save costs), while a smaller survey effort was allocated to "outer areas", with one survey round and larger distance between transects.

Main spawning areas 1,2,3,4 in 2013 approximately overlap with inner subareas areas A, C, E, G in 2015. In the following we will refer to the main spawning areas as inner areas A, C, E, G.

The estimated abundance of BFT in the combined inside areas (A, C, E, G) accounted for 43% and 26% of the total abundance (Inside and Outside areas combined) in 2013 and 2015, respectively. The estimated total weight of BFT in the combined inside subareas (A, C, E, G) accounted for 75% and 25% of the total weight in the extended survey area (Inside and Outside areas combined) in 2013 and 2015, respectively (**Tables 4, 5**). The estimated total abundance of BFT in the combined inside subareas (A, C, E, G) accounted for 43% and 26% of the total weight in the extended survey area (Inside and Outside areas combined) in 2013 and 2015, respectively.

Also, the re-analysis of survey time series with adjusted (reduced boundaries) for the inside subareas (A, C, E, G) that overlaps for 2010-2018 show that there are significant number of BFT sightings at the edge of the adjusted overlapping areas (**Table 6**). Also, tagging studies suggest that there is little evidence to support that BFT home in on specific spawning areas over time, and multiple-spawning behavior can occur over 3-6 weeks, in multiple areas over the same spawning season (Carruthers *et al*. 2018). This suggest that a substantial and variable portion of the spawning stock may not be adequately covered by surveying only the inside areas A, C, E, and G (**Table 7**). Thus, surveys that only cover inside areas are like to provide annual abundance and weight estimates with highly variable bias. I recommend that areas of these four areas be modified slightly so that they include the annual variations in boundaries.

Spatial modelling of the BFT time series may be used to map the extent of spawning areas. Druon *et al*. (2011) derived daily mapping of potential BFT feeding and spawning habitats in the Mediterranean Sea based on satellite-derived sea surface temperature (SST). Their study suggests high year-to-year variations for the potential spawning habitat.

Guiardo *et al*. (2019) propose a large scale generalizable deep learning system for automatically counting whales from satellite and aerial images. They demonstrate proof of concept by applying the method to free Google Earth coastal imagery in 10 whale-watching hotspots. Possibly, similar methods could be used in the development of spawning habitat maps for BFT, and to study the timing of spawning.

In conclusion, a long-term monitoring program will require a survey design that covers much of the Mediterranean. One option could be to secure good annual coverage in the four main spawning areas (i.e., inner areas of A, C, E, G), and to cover the outer areas with less effort. Presumably it is prohibitively expensive to cover the entire outer area every year, even with large spacing between transects. An alternative method is to split the outer area into survey regions (blocks) that each is surveyed with synoptic coverage in a single year, achieving full coverage of all blocks over several years. Skaug *et al*. (2004) employed such methods in double-platform shipborne sighting surveys to quantify the abundance of minke whales in the Northeast Atlantic. Blocks with assumed uniform densities were defined, taking into account topographical and oceanographic features.

3.1.3 Vertical distribution of BFT

Based on tagging studies Bauer *et al*. (2017) show that BFT in the Mediterranean were more surface orientated during summer. However, a proportion of spawners may stay in the layers from 1-2 meter below the surface down to 10 meters ((Fromentin, *et al* 2003), thus hardly being detectable from the plane. Cañadas and Ben Mhamed (2016) estimated that only 47% of schools were available for detection at a given time. It is unclear how this proportion varies across years, or through the spawning season within a year. A key assumption when using the BFT abundance indices to track changes over time is that the diving behavior is relatively constant over time. This assumption can be monitored through acoustic methods and acoustic tagging studies.

3.2 Field Methods

The detection of BFT schools along transects will be affected transect width, observer skills, cluster size, the type of aircraft used (particularly if the aircraft has bubble windows or not), sea state, other weather conditions, time of day and more. In Distance it is assumed that these covariates affect detection only via the scale of the detection function, and do not affect the shape (Miller *et al*. 2019).

Although observer estimates may be better standardized and quality-checked through calibration experiments (Grup Air-Med 2019), the history of the aerial BFT demonstrate the logistical challenges to maintain standardized procedures. There have been substantial variations in the field data collections in the aerial survey that clearly introduce variable biases in the time series of abundance indices. I can understand that it is very difficult to standardize procedures in such a large and complex survey. Different companies have been contracted (presumably it is mandated to put out contracts for tender), so it is clearly important to have specifications for the aircrafts that minimize the effects on the counts of schools and animals. The most serious problem seems to relate to some aircrafts having bubble windows, and others not. Aircrafts with bubble windows will presumably improve the detection of animals right under the plane (center of the transect being searched).

For aerial surveys with observers, it is strongly recommended that double-platform methods (independent observer - independent counts) be used, if feasible. This would facilitate bias-corrections for counts caused by variable detection of schools, and for missing counts directly under the aircraft.

It is important that training of observers includes sufficient and accessible information on the principles of the methods. It is especially important that counts be restricted to fairly narrow distances from the transect line that can be searched consistently across observers. Clearly, expert spotters recruited from the industry have long experience in detecting schools, but their focus in their past have been to maximize profit. The best spotters may be able to spot schools at large distances, especially large schools. Such data are opportunistic, within unknown selection probability, and therefore difficult to incorporate in estimates of abundance indices.

Reliable counts of individual BFT within an accurately defined strip-width for each transect would be ideal. Aerial surveys using high-resolution video or digital photography and machine learning now provides an alternative for collecting abundance data from standardized strip transects. Such methods could be standardized so the data collections are largely independent of platform. High resolution images would allow the accurate counting of individual fish, and estimation of the length of individuals. This would eliminate the need for estimating school size and weight subjectively by observers. Also, the use of video or digital photography in the aerial surveys is likely to save time and money in the long run. It is likely that such methods could be operationalized within a couple of years. It would be particularly useful to employ such methods in parallel with observers for a period of time. In this period, independent counts would be collected by the observers and high-resolution video or camera. Also, it would be effective to involve the expert spotters in the interpretations of video and still photos, as part of the training of machine learning techniques.

Planes or drones can be outfitted with gyro-stabilized digital video, or cameras linked to computers to capture thousands of high-resolution digital photographs along the transects. For example, US Geological Surveys, Western Ecological Research Center (WERC), are now using machine learning techniques to automate the detection and counts of seabirds and marine mammals from imagery collected in photographic aerial surveys. Institute of Marine Research (IMR), Norway is using drones to conduct photographic surveys for the abundance estimation of ice breeding seals (harp and hooded) and coastal seals (grey and harbour seals). The images have been analyzed manually by trained experts. This is time consuming and costly, and also involves subjective human interpretation. IMR in collaboration with the Norwegian Computing Center is now developing methodology for automatic processing of aerial images (https://www.nr.no/en/projects/uavseal). Marine Scotland have contracted HiDef Aerial Surveying Ltd to conduct survey flights using high resolution video cameras in their strategic surveys of marine mammals and seabirds in Scotland. Even small objects can be detected from aircraft flying at 2000 feet.

Schofield *et al*. (2019) provides a review of methods that use drones to study marine vertebrates. Koen *et al*. (2019) conducted an experiment where they compared abundance estimates of narwal from aerial transect surveys based on counts by observers with counts from digital camera images. Their experiment involved fields methods that are similar to the ones used in the BFR surveys. The observer data in Koen *et al*. (2019) were collected in a double-observer experiment, with the two front observers (Observer pair 1) recording data independently of the two rear observers (Observer pair 2). Images were collected by two autonomously operated digital single lens reflex still cameras. Comparable numbers of individuals were detected by both platforms.

By combining digital aerial survey data (which has good spatial coverage) with moored acoustics observation systems (which provide good temporal coverage) reliable counts of animals at the surface may also be bias adjusted for changes in the timing of spawning and in the diving behavior over time.

3.3 Statistical Analysis

The analysis in the 2010-2019 re-analysis report Cañadas and Vázquez (2020) were conducted in R, using a script that was developed by the authors, based on methods in the R Distance Software (Miller 2019). A simple design-based estimator for the density of animals provided in the report is

$$\widehat{D_a} = {}^{(n\,\bar{s})}\!/_{(2\,esw\,L)}$$

where n is the number of separate schools observed, $\bar{s}$ is the mean number of animals per school, $2esw$ is the (estimated) effective search width of the transects (in km), and L is the total length of transects searched (in km). Hence, an estimator for the density of schools (number of schools per square km) should be

$$\widehat{D_s} = {}^{n}\!/_{(2\,esw\,L)}.$$

Total abundance (number of animals) is then estimated by scaling density estimates up to total survey area, A. The same principle should apply for estimating total number of schools. It should be noted that an estimate of $\bar{s}$ simply taken as the mean group size across observations would be biased in the case that detection of schools depends on school-size.

Cañadas and Vázquez (2020) do not provide estimators for the variance of their estimates of abundance or weight. Based on Thomas *et al*. (2010), variance estimation based on transects as primary sampling units (PSU) seems to be a good option. In a model-assisted approach, bootstrapping would be an option to incorporate estimation of the effective strip half-width based on fitting detection functions, and any “size-bias” adjustments for the case that detection of schools depends on school-size, in the variance estimates. Miller *et al*. (2019) provide estimators of abundance and an analytical estimator for the associated variances for line transect distance sampling. The Horwitz-Thompson type estimators in Miller *et al*. (2019) also accounts for variance in the estimated detection function related to schools’ size and other factors. It is strongly advised that future data analysis reports include a detailed description of the estimators.

There are several inconsistencies in the tabulated abundance of schools in Tables 4.1.x in Cañadas and Vázquez (2020) versus estimates from the estimator above, marked by * in the **Table 1**.

Also, the estimated total weights (T) annually in area A (Table 4.1.1) differ greatly from standard estimates based on T*= mean density (weight per km^2) x Area (61837km^2), while estimates of total abundance based on density of animals are consistent with standard estimates, as shown in the **Table 2**.

Due to many inconsistencies in the estimates from the 2019 re-analysis (**Table 3**), the estimators and the R-script needs to be reviewed and revised. In an updated report with corrected and quality-assured estimates the estimators should also be provided. It is recommended that an analyst with strong expertise in statistical survey methods and R-programming be contracted to assist in the re-analysis and documentation. The documentation of methods, along with the R-script used in the re-analysis would allow peer review. Because of all the complexities and the data, it may also be useful to test the script using simulated data where "the truth" is known.

4. Summary of Findings BY TOR

i. Review all relevant information (provided by the ICCAT Secretariat - GBYP) on the survey's design, implementation, and statistical approach for the development of the BFT index of abundance. If deemed necessary, discussion over a webinar between CIE reviewers and BFT aerial survey team. Is survey documentation and supporting material adequate to conduct this review?

The documentation of survey methods and analysis was insufficient in the annual data analysis reports, and in the 2019 re-analysis. With respect to survey design, it was not clearly specified how many survey rounds were conducted in each subarea and year, and the timing of the rounds were poorly documented. It is possible that such information is provided elsewhere, but details of the survey design should be part of main reporting. The reports did not include information about the randomization of the starting point for systematic surveys in each round, but this was confirmed in a conference call with the lead authors. Rather surprising, there was not documentation of the variance estimators employed in the data analysis reports. In the discussion via webinar it became apparent that the main authors had used R analysis software without knowledge about the variance estimation methods. Miller *et al*. (2019) provide suitable methods. Bootstrapping (with transects as PSUs) is a simple alternative to the analytical variance estimators in Miller *et al*. (2019).

ii. Survey design. Evaluate the historical protocols and analytical approaches used in this survey as well as the recommended changes to the design procedures.
    a. Is the current survey design and changes implemented over its history consistent with state-of the art aerial survey design and adequately accounted for in data or statistical treatment?

The annual surveys, with exception for 2013 and 2015, do not fully cover the spawning area. Also, the boundaries of some of the focus subareas (particularly inner subarea A) have changed over time. Biases due to variable coverage are difficult to quantify and correct for and has not been fully accounted for.

The definition of boundaries for inner sub-areas that encompass all variations in boundaries in the time series, combined with imputation based on spatial modelling to fill data-gaps in parts of subareas, may be useful.

    b. Have logistical issues that precluded full attainment of the design been adequately addressed?

The use of aircraft with bubble windows has reduced bias in school counts, and it would be a significant improvement if this could be standard in aerial flights with observers. Variable biases due to changes in observers are difficult to address. A discussed above, transition to aerial surveys using digital imagining systems could eliminate many of the biases related to field operations.

    c. Are there further unaccounted for factors?

iii. Evaluate Statistical treatment and index calculation of the Mediterranean survey time series.
    a. Are data treatments (spatial stratification, etc.) appropriate and adequate to account for known factors affecting detection and quantification of spawning biomass.?

The adjusted (reduced boundaries) for the inside subareas (A, C, E, G) that overlaps for 2010-2018 made sure that annual estimates could be provided for the same subareas. However, this further reduced the spatial coverage of the spawners, as evident by the many BFT sightings at the edge of the adjusted overlapping areas. An alternative analysis would be to define standard fixed boundaries for inside subareas (A, C, E, G) and expand the mean density to those areas based on imputation techniques.

b. For issues not addressed by (ii) above, does statistical treatment adequately account for issues affecting detectability, specifically does use of ‘school size’ in the detection function bias the detection estimates and does the method potentially double count schools detected multiple times?

If detection of BFT tuna schools is dependent on school size, then the estimator of abundance provided in the reanalysis report will be biased. Distance sampling in R (Miller *et al*. 2019) provides estimators for abundance and variances that appropriately accounts for varying detection related to school size.

c. Does the most recent (2019) index construction represent the most effective treatment?

The incomplete documentation of methods, and errors in the analysis, makes it difficult to answer this question. Also, as pointed out elsewhere, the more narrowly defined spawning areas may not have reduced annual biases.

d. Does the high inter-treatment variability of the index due to poorly estimated or highly variable detection functions render the index unreliable as a time series?

I believe that the time series will provide useful indices of abundance provided that the estimates are corrected. For a long-term time-series, it will be important to provide some level of coverage in the outside areas, especially since there are strong signs that the abundance of eastern BFT is increasing. Also, variable detection of schools could largely be eliminated using digital video or cameras. Since BFT is a long-lived species, time series analysis could likely improve the ability to detect trends in abundance if a long time series can be attained.

e. Are better statistical (spatial/temporal) treatments possible?

Cañadas and Vázquez (2020) have gone to great length to delineate areas that have been consistently covered in all years of the time series. However, even though these fixed areas have been surveyed throughout the time series, it is not possible based on current data to account for variable biases due to incomplete coverage of the stock. This further restriction of survey strata can even add to the variable bias in annual estimates, since clearly many schools were detected at the boundaries, or outside the revised boundaries (mainly in inside subarea A). I would recommend that an additional analysis be conducted where the inside areas A, C, G, E are redrawn to include all high-density transects. This would help bracket results under two different assumptions: (1) that the re-designed restricted strata contain a constant fraction of the total spawning stock every year of the times series, or (2) that the density of animals and schools within each original stratum is representative for the expanded stratum.

iv. Suitability of GBYP aerial survey

a. Does it achieve full objective (all Mediterranean spawning grounds) or partially (on specific spawning areas)

In general, the survey provides reasonable spatial coverage of four subareas that define important spawning grounds. However, it is currently unknown what portion of the entire spawning population is covered. Hence, it is important that the survey provide some level of coverage in the outside areas over time. The timing of the survey rounds within a year should also be carefully planned to ensure coverage of the main spawning period.

b. Are known logistical/biological/unaccountable factors adequately addressed?

c. Are unknown factors (availability of fish, timing of spawning, behavioral changes) too substantial, rendering the survey unable to achieve its full or partial goals?

d. Provide general recommendations for potential improvements

v. Determine if the current approach meets the established criteria for an index of abundance. If not provide an explanation of why and whether or not the data can be re-evaluated to meet these criteria.

vi. Provide recommendations on the future of this survey, as well as potential design modifications, standardization and/or research to improve the survey

I am impressed that you have been able to conduct annual aerial surveys that cover multiple jurisdictions! I believe that aerial surveys (using aircrafts, and possibly drones) provide the best basis for providing a time-series of relative abundance of BFT as input to stock assessment. I also believe that a long-term monitoring program will require a survey design that covers much of the Mediterranean. One option could be to secure good annual coverage in four main spawning areas (i.e., inner areas of A, C, E, G) with fixed boundaries, and to cover the remaining outer areas with less effort. As described earlier, one option is to split the outer area into survey regions (blocks) that each is surveyed with synoptic coverage in a single year, achieving full coverage of all blocks over several years. In terms of standardization, I believe the biggest gains can be achieve by evolving the aerial survey into a video or camera-based monitoring system, with automated analysis to provide counts of individual BFT and lengths. The lengths can be used to estimate weights, and also to better define the mature BFT. Such automated systems are likely to be cost-effective, reliable, and may allow the allocation of funds to improve temporal and spatial coverage.

5. Conclusions and Recommendations in accordance with the TORs

I believe that aerial surveys based on transect sampling is the best option for achieving a reliable time-series of abundance indices for BFT in the Mediterranean. Tagging studies are important for obtaining knowledge on tuna spawning behavior and migrations, but it is exceptionally challenging to operate tagging studies for BFT that can support abundance estimates based on mark-recapture methods.

Several improvements will be necessary to ensure reliable estimates over time. Most of these have already been pointed out in the many reports that were made available for this review.

Recommendation 1.

Inconsistencies in key estimates provided in Cañadas and Vázquez (2020) clearly need to be addressed. The R-distance software (Miller *et al*. 2019) seems to provide the necessary methods to provide key estimates and associated variances. It is recommended that an analyst with strong expertise in survey sampling statistics and strong experience with the R distance package be contracted to assist in the analysis of the time series of data. The key is to specify the appropriate estimators and to check that they are properly implemented in the R-script. This can hopefully be achieved through a modification of the R-code currently used. Also, it is recommended that analysis data be made available on the ICCAT website, and that an R-script be available for example via GitHub.

Because of the complexity in the data, it may also be necessary to test the methods using simulated data.

Recommendation 2.

An evolution towards using high-resolution video- or camera in counts of BFT has great potential to improve the reliability of abundance indices. In particular, such methods could ensure standardized counts of individual animals (and their lengths) within an accurately defined narrow transect width. This could eliminate the need to estimate school size, and weight, and also the need to estimate detection probabilities. It would be very advantageous to use video-camera systems in parallel with observers for a period. This would provide more insights on the observer's ability to estimate schools size, and size of animals. This will be important for assessing the reliability of current methods. Also, the observers, particularly the expert spotters, could provide valuable assistance in the quality assurance of image analysis. If high-resolution video- or still-cameras are used to count animals, then the alternative zig-zag design developed by Harbitz (2019), instead of parallel transects, may be considered for the aerial surveys. With the current parallel transect design, the time flying between transects provides a break for observers. If digital images are used, instead of observers, then the zig-zag design may improve cost-efficiency.

Recommendation 3.

Redesign the survey to include coverage of inside areas annually, and outside areas over multiple years. It is quite possible that a switch to video-camera observations from aircrafts or drones could free resources that allows an expansion of the survey coverage in space.

Recommendation 4.

Develop spawning habitat models that over time will allow better definition of the main spawning areas.

Recommendation 5.

There are many studies that provide information on the spawning behavior of BFT in the Mediterranean. It would be useful to systematize this information based on a thorough review. In particular, it is important to assess if timing of spawning changes over time, and if the vertical distribution of spawning BFT near the surface changes over time. It is of course essential that the multiple survey rounds within a year covers the main spawning period.

Acknowledgements

This work has been carried out under the ICCAT Atlantic-Wide Research Programme for Bluefin Tuna (GBYP), which is funded by the European Union, several ICCAT CPCs, the ICCAT Secretariat, and other entities (see https://www.iccat.int/gbyp/en/overview.asp). The content of this paper does not necessarily reflect ICCAT's point of view or that of any of the other sponsors, who carry no responsibility. In addition, it does not indicate the Commission's future policy in this area.

References

Alemany. F., Quintanilla, L., Velez-Belchíc, P., García, A., Cortés, D., Rodríguez, J.M., Fernández de Puelles, M.L., González-Pola, C., and J.L. López-Jurado. 2010. Characterization of the spawning habitat of Atlantic bluefin tuna and related species in the Balearic Sea (western Mediterranean).

Bauer A.K, J.M. Fromentin, H. Demarcq, S. Bonhommeau. 2017. Habitat use, vertical and horizontal behaviour of Atlantic bluefin tuna (*Thunnus thynnus*) in the Northwestern Mediterranean Sea in relation to oceanographic conditions.

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L. & Thomas, L. (2001). Introduction to distance sampling: estimating abundance of biological populations. Oxford University Press, Oxford.

Buckland, S.T., Oedekoven, C.S. & Borchers, D.L. Model-Based Distance Sampling. JABES 21, 58–75 (2016). https://doi.org/10.1007/s13253-015-0220-7

Cañadas, A. and Vázquez, J.A. 2020. 2010-2019 aerial survey data re-analysis. Atlantic-wide Research Programme on Bluefin Tuna Phase 9.

Cañadas, A. and Vázquez, J.A. 2015a. Short-term contract for the aerial survey design of the Atlantic-wide Research Programme on Bluefin Tuna (ICCAT-GBYP Phase 5-2015). Final Report.

Cañadas, A. and Vázquez, J.A. 2015b. Atlantic-wide Research Programme on Bluefin Tuna (ICCAT-GBYP Phase 5-2015). Elaboration of 2015 data from the aerial survey on spawning aggregations. Final Report.

Cañadas, A. and Ben Mhamed, A. 2016. Power analysis and cost-benefit analysis for the ICCAT GBYP aerial survey on bluefin tuna spawning aggregations. Atlantic-wide Research Programme on Bluefin Tuna Phase 5.

Cañadas, A. and Vázquez, J.A. 2013. Short-term contract for assessing the feasibility of a large-scale survey on bluefin tuna spawning aggregations in all the Mediterranean Sea for obtaining useful data for operating modeling purposes. Atlantic-wide Research Programme on Bluefin Tuna Phase 3.

Cañadas, A., P. Hammond, and Vázquez, J.A. 2010. Atlantic-wide Research Programme on Bluefin Tuna (GBYP - 2010). Data Recovery Plan – Elaboration of 2010 Data from the Aerial Survey on Spawning Aggregations. Final Report.

Carruthers, T., A. Di Natale , M. Lauretta , A. Pagá García , S. Tensek. 2018. Migratory behaviour of Atlantic bluefin tuna entering the Mediterranean Sea. Collect. Vol. Sci. Pap. ICCAT, 74(6): 3082-3099.

Morse, M., Cadrin, S., Kerr, L., Secor, D., Siskey, M., Arrizabalaga,H., Hanke, A. and C.Porch. 2017. An updated analysis of bluefing tuna stock mixing. Vol. Sci. Pap. ICCAT, 74(6): 3486-3509(2018).

Di Natale, A. and Idrissi, M'H. 2013. ICCAT GBYP aerial survey: spawners vs. juveniles. A SWOT analysis of both perspectives. Collect. Vol. Sci. Pap. ICCAT, 69(2): 803-815.

Di Natale, A. 2016. Tentative SWOT analysis for the calibration of ICCAT GBYP aerial survey for bluefin tuna spawning aggregations. Collect. Vol. Sci. Pap. ICCAT, 72(6): 1463-1476.

Druon, J.N., Fromentin, Aulanier, F. and J. Heikkonen. 2011. Potential feeding and spawning habitats of Atlantic bluefin tuna in the Mediterranean Sea. Mar. Ecol. Prog. Ser. 39: 223–240. doi: 10.3354/meps09321

Fonteneau, A., Suzuki, Z. and Payne, A.I.L. 2013. Mid-term review of the ICCAT Atlantic-wide Research Programme on Bluefin Tuna Phase 4.

Fromentin Jean-Marc, Farrugio Henri, Deflorio Michèle, De Metrio Gregorio (2003). Preliminary results of aerial surveys of bluefin tuna in the western Mediterranean sea. Collective Volume of Scientific Papers, 55(3), 1019-1027. Open Access version : https://archimer.ifremer.fr/doc/00194/30531/

Fromentin, J.M. and J.E. Powers. 2005. Atlantic bluefin tuna: population dynamics, ecology, fisheries and management Fish. Fish., 6 (2005), pp. 281-306.

Grup Air-Med. 2019. Implementacion del muestreo aereo para la calibracion y validacion. Atlantic-wide Research Programme on Bluefin Tuna Phase 9.

Guirado, E., Tabik, S., Rivas, M.L. *et al*. Whale counting in satellite and aerial images with deep learning. Sci Rep 9, 14259 (2019). https://doi.org/10.1038/s41598-019-50795-9.

Harbitz, A. 2019. A zigzag survey design for continuous transect sampling with guaranteed equal coverage probability. Fisheries Research 213: 151-159.

Hedley, S.L. and Buckland, S.T. 2004. Spatial models for line transect sampling. Journal of Agricultural, Biological and Environmental Statistics 9: 181-199.

ICCAT. 2012. 2011 GBYP workshops on aerial surveys, and operational meetings on biological sampling and on tagging of bluefin tuna. Collect. Vol. Sci. Pap. ICCAT, 68(1): 1-65.

ICES.2018. Report of the Working Group on Widely Distributed Stocks (WGWIDE), 28 August-3 September 2018, Torshavn, Faroe Islands. ICES CM 2018/ACOM: 23. 488pp.

ICCAT 2017. Report of the 2017 ICCAT bluefin stock assessment meeting. Madrid, Spain.

ICCAT 2019. Report for Biennial Period, 2018-2019, Part I – Vol. 2. Standing Committee on Research and Statistics (SCRS). Madrid, Spain (2019), p. 450.

Koen C. A. B., R.G. Hansen, K. E. Leonard, W. R. Koski, M.P. Heide-Jørgensen. 2019. A comparison of image and observer based aerial surveys of narwhal. Marine Mammal Science 35(4): 1253–1279. https://doi.org/10.1111/mms.12586

Miller, D.L., L-M- Burt, E.A. Rexstd, and L. Thomas. 2013. Spatial models for distance sampling data: recent developments and future directions. Methods in Ecology and Evolution 4: 1001-1010. https://doi.org/10.1111/2041-210X.12105 Miller, D.L., E. Rexstad, L. Thomas. L. Marshall, and J.L. Laake. 2019. Distance sampling in R. J. of Sta. Softw. 89 (1). doi: 10.18637/jss.v089.i01

MRAG. 2016. Second review of the ICCAT Atlantic-wide Research Programme on Bluefin Tuna Phase 6.

Nøttestad, L., E. Boge, and K. Ferter. 2020. The comeback of Atlantic bluefin tuna (*Thunnus thynnus*) to Norwegian waters. Fish. Res. 231. https://doi.org/10.1016/j.fishres.2020.105689

Picenetti, C., A. Di Natala, and P. Arena. 2013. Eastern bluefin tuna (*Thunnus Thynnus*, L.) reproduction and reproductive areas and season. Collect. Vol. Sci. Pap. ICCAT, 69(2): 891-912.

Quílez-Badía, G, Tensek, S., Di Natale, A., Pagá García, A., Cañadas, A., Kitakado, T. and Kell, L.T. 2016. A consideration of additional variance in the Mediterranean aerial survey based on electronic satellite tags. Collect. Vol. Sci. Pap. ICCAT, 72(6): 1544-1552.

Santamaria, Nohely & Bello, Giambattista & Corriero, A. & Deflorio, M. & Vassallo Agius, Robert & Deniz, Tomris & Metrio, G.. (2009). Age and growth of Atlantic bluefin tuna, Thunnus thynnus (Osteichthyes: Thunnidae), in the Mediterranean Sea. Journal of Applied Ichthyology. 25. 38 - 45. 10.1111/j.1439-0426.2009.01191.x.

Schofield, Gail, Esteban, Nicole, Katselidis, Kostas A. and Hays, Graeme C. 2019, Drones for research on sea turtles and other marine vertebrates – A review, Biological Conservation, vol. 238, pp. 1-10, doi: 10.1016/j.biocon.2019.108214.

Simmonds, J, and David N. MacLennan.2005. Fisheries acoustics: theory and practice. John Wiley & Sons, 2008.

Skaug, H.J., N. Øien, T. Schweder, and G. Bøthun. 2004. Abundance of minke whales (*Balaenoptera acutorostrata*) in the Northeast Atlantic:variability in time and space. Can. J. Fish. Aquat. Sci.61: 870–886

Thomas, L; Buckland, ST; Rexstad, EA; Laake, JL; Strindberg, S; Hedley, SL; Bishop, JRB; Marques, TA; Burnham, KP. 2010. Distance software: design and analysis of distance sampling surveys for estimating population size. Journal of Applied Ecology 47(1): 5-14. DOI: 10.1111/j.1365-2664.2009.01737.x

Vázquez, J.A. and Cañadas, A. 2019. Design of an aerial survey calibration exercise and an updated aerial survey sighting protocol. Atlantic-wide Research Programme on Bluefin Tuna Phase 8.

Walli A, Teo SLH, Boustany A, Farwell CJ, Williams T, *et al*. (2009) Seasonal Movements, Aggregations and Diving Behavior of Atlantic Bluefin Tuna (*Thunnus thynnus*) Revealed with Archival Tags. PLoS ONE 4(7): e6151. doi:10.1371/journal.pone.0006151

Yuan, Y., Bachl, F.E., Lindgren, F., Borchers, D.L., Illian, J.B., Buckland, S.T., Rue, H. and Gerrodette, T. 2017. Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales. Annals of Applied Statistics 11: 2270-2297.

Table 1. Extracts from Tables 4.1.x) in Cañadas and Vázquez (2020)

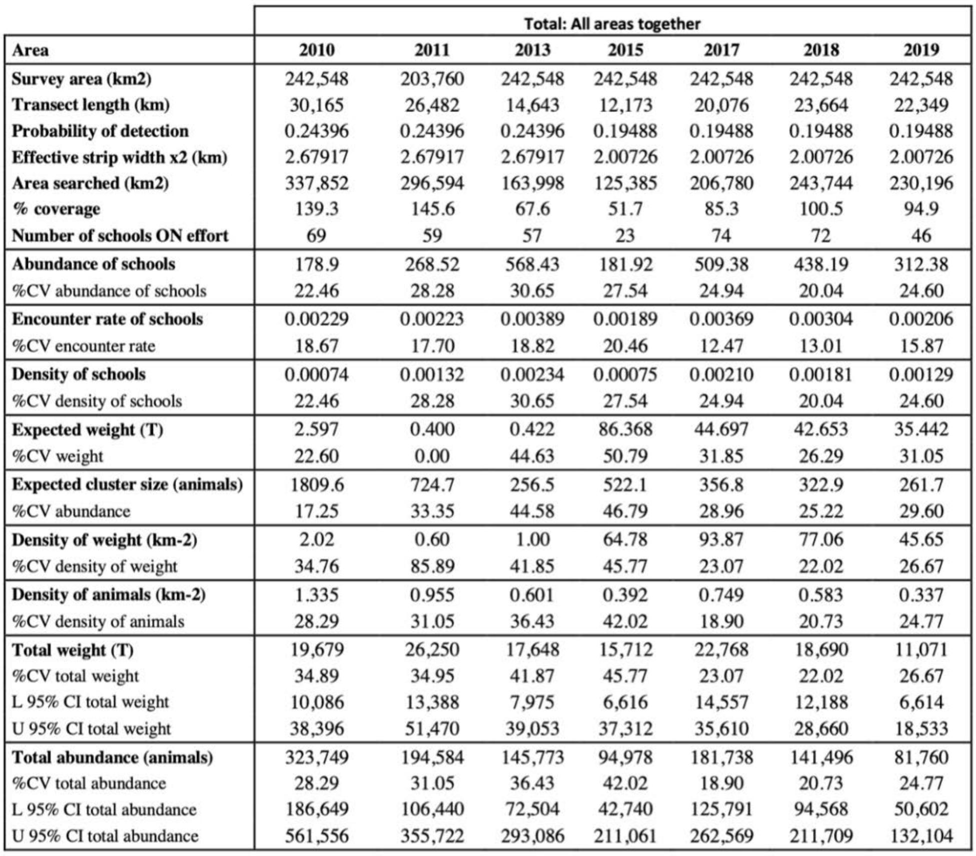
Strata	Area	Strip_length	Strip-width	Schools_N	Schools_Est*	Schools_tab
A	61837	6093	2,67917	8	30,30	10
C	51821	8354	2,67917	6	13,89	7,75
E	90102	12852	2,67917	30	78,50	46,83
G	38788	2866	2,67917	25	126,29	114,43
All	242540	30165	2,67917	69	207,08	178,90

Table 2. Extracts from Table 4.1.1 in Cañadas and Vázquez (2020)

	2010	2011	2013	2015	2017	2018	2019
Dens_A	0,299	0,114	0,147	0,257	0,86	1,079	0,67
Abun (animals)	18502	7028	9064	15894	53180	66713	41422
Abun(animals)*	18489	7049	9090	15892	53180	66722	41431
Dens_weight	0,49	0,14	0,2	45,8	152,06	175,72	107,72
Tot (weight) T	2119	963	1946	2832	9403	10866	6664
Tot (weight) T*	30300	8657	12367	2832135	9402934	10865998	6661082

Table 3. Re-analysis overlapping areas (Cañadas and Vázquez 2020)

Table 4.1.5. All areas together: Results of the re-analysis for all sizes of BFT, using the overlap areas between 2010-2018.



	Total: All areas together						
Area	**2010**	**2011**	**2013**	**2015**	**2017**	**2018**	**2019**
Survey area (km2)	242,548	203,760	242,548	242,548	242,548	242,548	242,548
Transect length (km)	30,165	26,482	14,643	12,173	20,076	23,664	22,349
Probability of detection	0.24396	0.24396	0.24396	0.19488	0.19488	0.19488	0.19488
Effective strip width x2 (km)	2.67917	2.67917	2.67917	2.00726	2.00726	2.00726	2.00726
Area searched (km2)	337,852	296,594	163,998	125,385	206,780	243,744	230,196
% coverage	139.3	145.6	67.6	51.7	85.3	100.5	94.9
Number of schools ON effort	69	59	57	23	74	72	46
Abundance of schools	178.9	268.52	568.43	181.92	509.38	438.19	312.38
%CV abundance of schools	22.46	28.28	30.65	27.54	24.94	20.04	24.60
Encounter rate of schools	0.00229	0.00223	0.00389	0.00189	0.00369	0.00304	0.00206
%CV encounter rate	18.67	17.70	18.82	20.46	12.47	13.01	15.87
Density of schools	0.00074	0.00132	0.00234	0.00075	0.00210	0.00181	0.00129
%CV density of schools	22.46	28.28	30.65	27.54	24.94	20.04	24.60
Expected weight (T)	2.597	0.400	0.422	86.368	44.697	42.653	35.442
%CV weight	22.60	0.00	44.63	50.79	31.85	26.29	31.05
Expected cluster size (animals)	1809.6	724.7	256.5	522.1	356.8	322.9	261.7
%CV abundance	17.25	33.35	44.58	46.79	28.96	25.22	29.60
Density of weight (km-2)	2.02	0.60	1.00	64.78	93.87	77.06	45.65
%CV density of weight	34.76	85.89	41.85	45.77	23.07	22.02	26.67
Density of animals (km-2)	1.335	0.955	0.601	0.392	0.749	0.583	0.337
%CV density of animals	28.29	31.05	36.43	42.02	18.90	20.73	24.77
Total weight (T)	19,679	26,250	17,648	15,712	22,768	18,690	11,071
%CV total weight	34.89	34.95	41.87	45.77	23.07	22.02	26.67
L 95% CI total weight	10,086	13,388	7,975	6,616	14,557	12,188	6,614
U 95% CI total weight	38,396	51,470	39,053	37,312	35,610	28,660	18,533
Total abundance (animals)	323,749	194,584	145,773	94,978	181,738	141,496	81,760
%CV total abundance	28.29	31.05	36.43	42.02	18.90	20.73	24.77
L 95% CI total abundance	186,649	106,440	72,504	42,740	125,791	94,568	50,602
U 95% CI total abundance	561,556	355,722	293,086	211,061	262,569	211,709	132,104

Table 4. Inside-outside comparison 2015 (Cañadas and Vázquez 2015b)

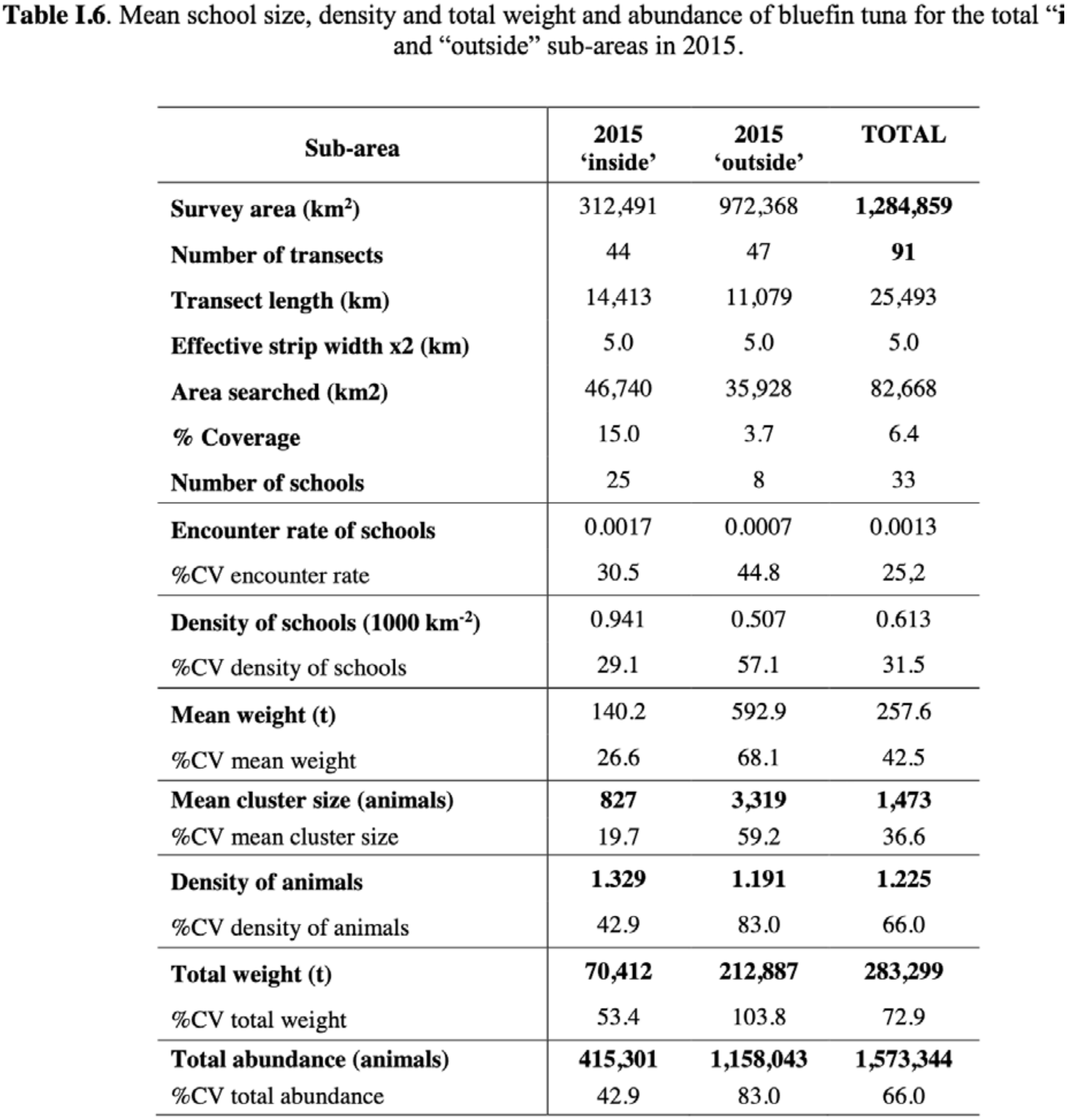


Table I.6. Mean school size, density and total weight and abundance of bluefin tuna for the total “i
and “outside” sub-areas in 2015.

Sub-area	**2015 ‘inside’**	**2015 ‘outside’**	**TOTAL**
Survey area (km^2)	312,491	972,368	**1,284,859**
Number of transects	44	47	**91**
Transect length (km)	14,413	11,079	25,493
Effective strip width x2 (km)	5.0	5.0	5.0
Area searched (km2)	46,740	35,928	82,668
% Coverage	15.0	3.7	6.4
Number of schools	25	8	33
Encounter rate of schools	0.0017	0.0007	0.0013
%CV encounter rate	30.5	44.8	25,2
Density of schools (1000 km^{-2})	0.941	0.507	0.613
%CV density of schools	29.1	57.1	31.5
Mean weight (t)	140.2	592.9	257.6
%CV mean weight	26.6	68.1	42.5
Mean cluster size (animals)	**827**	**3,319**	**1,473**
%CV mean cluster size	19.7	59.2	36.6
Density of animals	**1.329**	**1.191**	**1.225**
%CV density of animals	42.9	83.0	66.0
Total weight (t)	**70,412**	**212,887**	**283,299**
%CV total weight	53.4	103.8	72.9
Total abundance (animals)	**415,301**	**1,158,043**	**1,573,344**
%CV total abundance	42.9	83.0	66.0

Table 5. Inside-outside comparison 2013 (Table 7, Cañadas and Vázquez 2013)

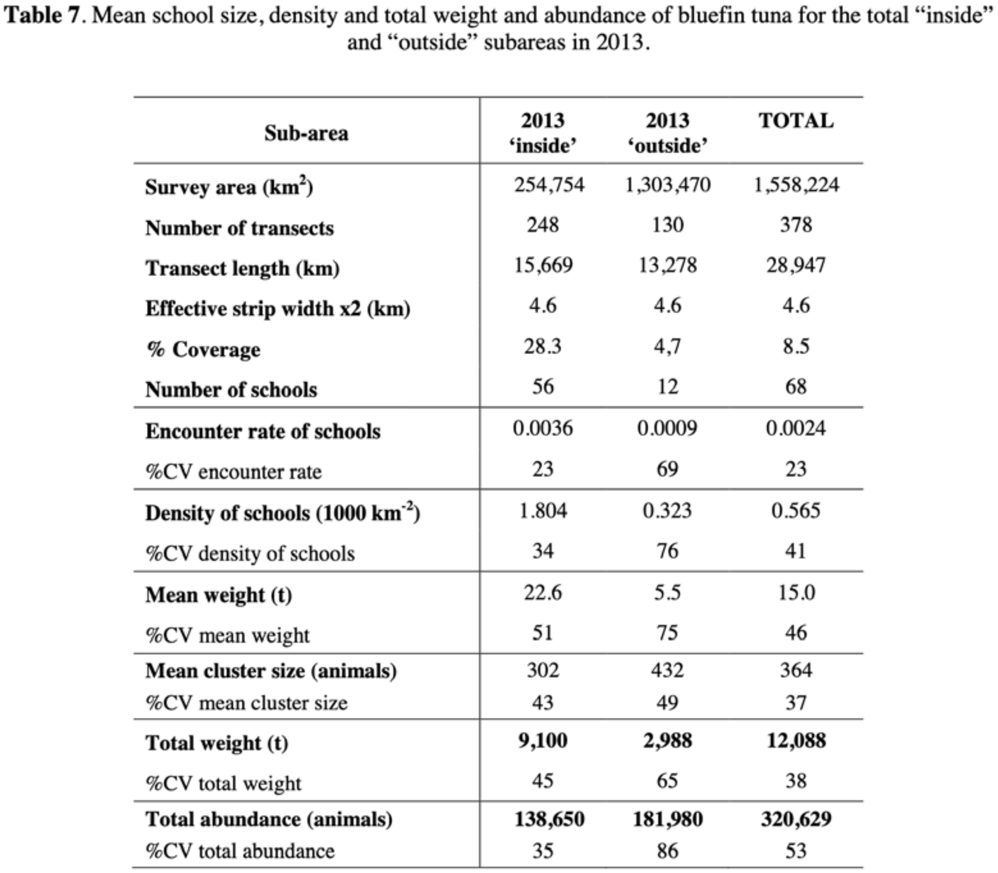


Table 7. Mean school size, density and total weight and abundance of bluefin tuna for the total “inside” and “outside” subareas in 2013.

Sub-area	2013 ‘inside’	2013 ‘outside’	TOTAL
Survey area (km^2)	254,754	1,303,470	1,558,224
Number of transects	248	130	378
Transect length (km)	15,669	13,278	28,947
Effective strip width x2 (km)	4.6	4.6	4.6
% Coverage	28.3	4,7	8.5
Number of schools	56	12	68
Encounter rate of schools	0.0036	0.0009	0.0024
%CV encounter rate	23	69	23
Density of schools (1000 km^{-2})	1.804	0.323	0.565
%CV density of schools	34	76	41
Mean weight (t)	22.6	5.5	15.0
%CV mean weight	51	75	46
Mean cluster size (animals)	302	432	364
%CV mean cluster size	43	49	37
Total weight (t)	**9,100**	**2,988**	**12,088**
%CV total weight	45	65	38
Total abundance (animals)	**138,650**	**181,980**	**320,629**
%CV total abundance	35	86	53

Table 6. Comparing re-analysis Cañadas and Vázquez (2020) to original estimates

	Original (Inside areas)	Re-analysis 2020 (Inside, overlap)	Original (Inside areas)	Re-analysis 2020 (Inside, overlap)
Year	2013	2013	2015	2015
Survey area	254754	242548	312491	242548
Transect length	15669	14643	14413	12173
Transect width	4,6	2,67917	5	2,00726
Number of schools ON effort	56	59	25	23
Abundance of schools	567,56	568,43	181,9	181,92
Abundance of schools*	197,93	364,77	108,41	228,31
Total weight	9100	17648	70412	15712
Total abundance (animals)	138650	145773	415301	94978

Table 7. Comparing Inside areas to total survey area

Estimate	Year	Inside	Inside+Outside	Ratio
Abundance	2013	138650	320629	0,43
Abundance	2015	415301	1573344	0,26
Weight	2013	9100	12088	0,75
Weight	2015	70412	283299	0,25

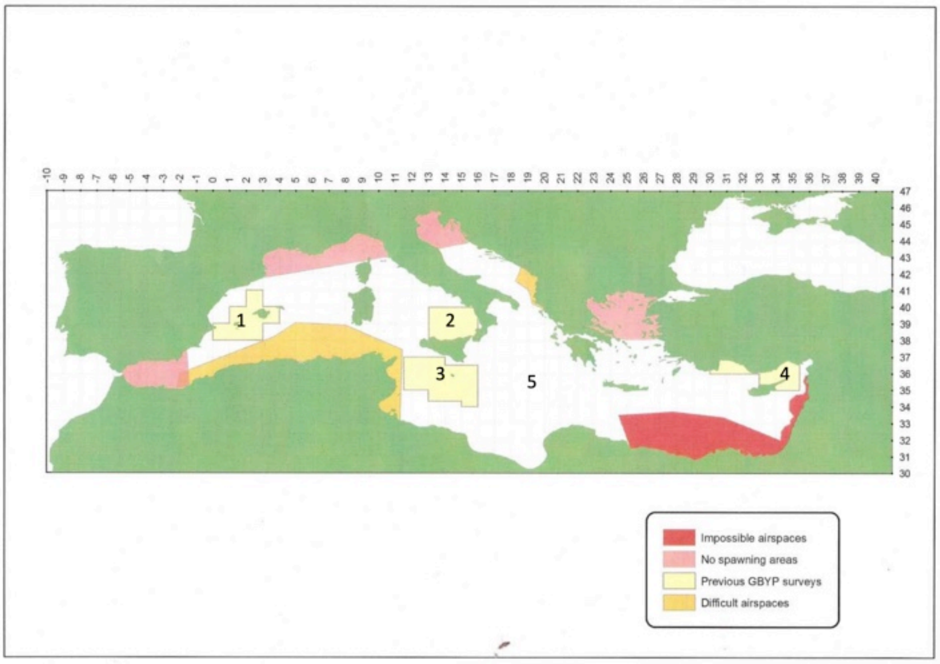


Figure 1. Blocks considered for this work

Figure 1. Survey blocks (strata) covered in 2013 BFT aerial survey (Cañadas and Vázquez 2013)

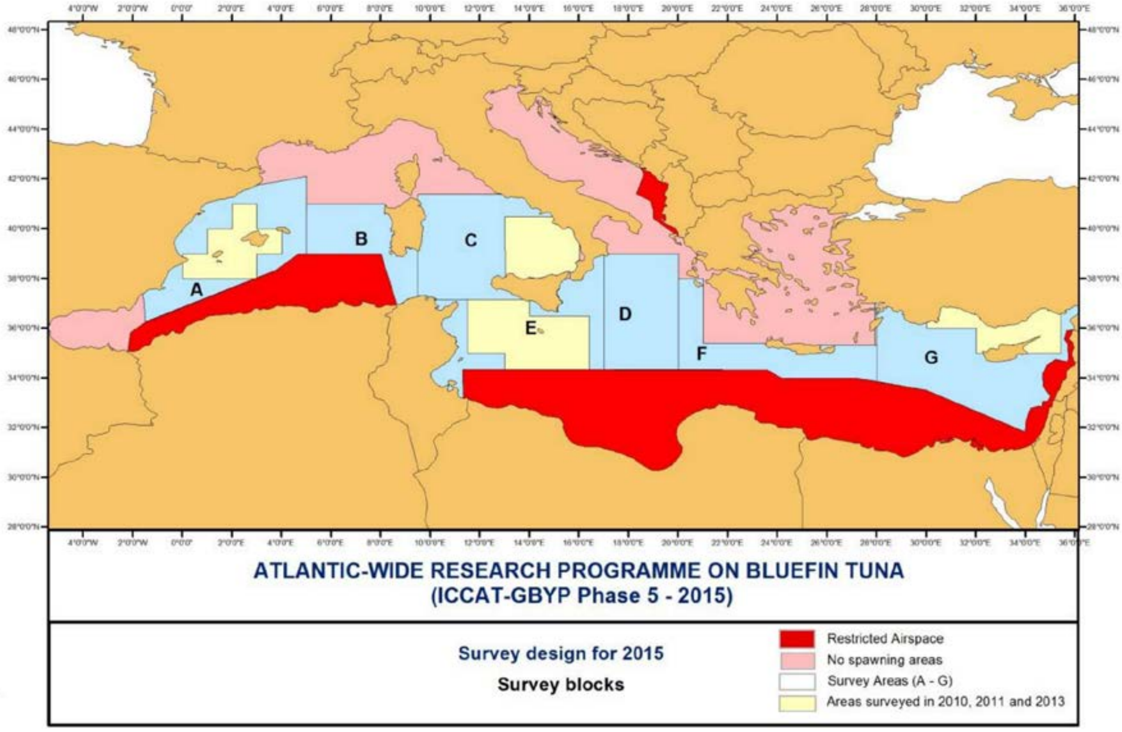


Figure 1. Survey blocks for 2015.

Figure 2. Survey blocks (strata) for 2015 extended coverage (Cañadas and Vázquez 2015a)

Appendix

Statement of Work

1. Background and Objectives

The BFT aerial survey is one of the major activities of the Atlantic Wide Research Programme for Bluefin Tuna (GBYP). It was launched in 2010 with the purpose of obtaining a relative abundance index of spawning biomass for the Mediterranean Sea. The index is obtained from aerial transects conducted during June in the four main spawning areas using a combination of scientific and professional spotters deployed on airplanes. Since its start, the survey has faced numerous logistical challenges and has had to alter its design and data processing protocols multiple times.

Currently, the most recent (2019) iteration of the index exhibits substantial differences from prior time series and the index exhibits high interannual variability both within and between regions. The magnitude of the difference between prior time series and the high variability has raised concerns regarding the estimation procedures and the overall efficacy of the survey to reflect annual spawner abundance in the Mediterranean Sea. Given the need to evaluate the survey and to soon take decisions regarding the nature of its continuation, ICCAT requests an independent desk review of the survey design, statistical treatments and analytical procedures and of its general capacity to achieve its objectives.

Expertise required to conduct this review will include two independent and highly qualified experts with a combined background and experience in aerial survey design, statistical time series evaluation, and a strong understanding of population modeling and stock assessment. Reviewers will have no financial or perceived conflicts of interest related to the subject matter to be reviewed. Finally, reviewers are to be approved by ICCAT upon selection but only as approval related to reviewer expertise to conduct the review and/or any conflicts of interest not discovered over the reviewer identification and selection process. The CIE will however make the final decision on the eligibility and effectiveness of all selections in such cases.

2. Reviewer Tasks

To provide an independent review of the Mediterranean Sea Bluefin tuna aerial survey design and statistical analysis used in the development of an index of spawning stock biomass, with an emphasis on the 2019 re-analysis of the time series. Specific tasks will include, but not be limited to, the following Terms of Reference (ToR):

i. Review all relevant information (to be provided by the ICCAT Secretariat - GBYP) on the survey's design, implementation, and statistical approach for the development of the BFT index of abundance. If deemed necessary, discussion over a webinar between CIE reviewers and BFT aerial survey team. Is survey documentation and supporting material adequate to conduct this review?

ii. Survey design. Evaluate the historical protocols and analytical approaches used in this survey as well as the recommended changes to the design procedures.
   a. Is the current survey design and changes implemented over its history consistent with state-of the art aerial survey design and adequately accounted for in data or statistical treatment?
   b. Have logistical issues that precluded full attainment of the design been adequately addressed?
   c. Are there further unaccounted for factors?

iii. Evaluate Statistical treatment and index calculation of the Mediterranean survey time series.
   a. Are data treatments (spatial stratification, etc.) appropriate and adequate to account for known factors affecting detection and quantification of spawning biomass.?
   b. For issues not addressed by (ii) above, does statistical treatment adequately account for issues affecting detectability, specifically does use of 'school size' in the detection function bias the detection estimates and does the method potentially double count schools detected multiple times?
   c. Does the most recent (2019) index construction represent the most effective treatment?
   d. Does the high inter-treatment variability of the index due to poorly estimated or highly variable detection functions render the index unreliable as a time series?
   e. Are better statistical (spatial/temporal) treatments possible?

iv. Suitability of GBYP aerial survey
    a. Does it achieve full objective (all Mediterranean spawning grounds) or partially (on specific spawning areas)
    b. Are known logistical/biological/unaccountable factors adequately addressed?
    c. Are unknown factors (availability of fish, timing of spawning, behavioural changes) too substantial, rendering the survey unable to achieve its full or partial goals?
    d. Provide general recommendations for potential improvements

v. Determine if the current approach meets the established criteria for an index of abundance. If not provide an explanation of why and whether or not the data can be re-evaluated to meet these criteria.

vi. Provide recommendations on the future of this survey, as well as potential design modifications, standardization and/or research to improve the survey

3. Deliverables

Deliverable #1- CIE reviewer shall submit a draft review report (formatted as an SCRS document) providing complete documentation of the review and recommendations (late September-early October 2020).

Deliverable #2 – CIE reviewer will present the draft review report findings to the Bluefin Tuna Working Group (BFTWG) at its next available meeting (early October 2020) (virtual presentation).

Deliverable #3- CIE reviewer will submit a final review report (formatted as an SCRS document), revised as based on comments provided by the BFTWG (first week in November 2020).