THE ATLANTIC-WIDE RESEARCH PROGRAMME FOR BLUEFIN TUNA
(GBYP Phase 10)

Independent Peer Review of the ICCAT GBYP Aerial Survey Design, Implementation and Statistical Analyses

(ICCAT GBYP 12/2020)

Prepared by the Center for Independent Experts (CIE)

January 2021

This project is co-funded by
the European Union

Independent Peer Review by the Center for Independent Experts of the ICCAT GBYP Aerial Survey Design, Implementation and Statistical Analyses

Executive Summary

This report describes the process used by the Center for Independent Experts (CIE) to complete an independent peer review of the aerial surveys of the bluefin tuna (BFT) stocks of the Mediterranean Sea completed in each of seven years between 2010 and 2019. The review focused on a recent re-analysis of the surveys, which showed large differences in estimates of BFT spawning biomass from previous estimates and high interannual variation both within and between regions, possibly because spawning locations and spawning times vary across years. Two independent experts (reviewers), commissioned by the CIE, reviewed the survey design, field methods, and the methods of analysis. Sampling efforts for the surveys have focused on four subareas of the Mediterranean that are assumed to represent the main BFT spawning areas. There is strong evidence that a long-term monitoring program will require a survey design that covers much of the Mediterranean, including areas outside of the four subareas currently surveyed. Given cost limitations, one option to accomplish this is to continue annual spatial and temporal sampling coverage in the four main spawning areas at current levels, and to cover the remaining spawning area with less effort. Model-based methods could be used to combine data from the two survey components and complement the design-based methods used to date. Given the difficulties that observers face in recording reliable data for the line transect method, the use of high-resolution imagery should be considered as an alternative to observers, possibly in conjunction with long-distance drones. Video or still images taken from higher altitude provide a permanent record, allowing verifiability. Several inconsistencies found in the re-analysis results suggest errors in the computer code that need to be corrected.

Description of the Work Carried Out

In early September 2020, the CIE commissioned the services of Dr S.T. Buckland, an expert in line-transect methods for estimating animal abundance, and Dr. J.H. Vølstad, an expert in the application of statistics in fisheries stock assessment, to conduct the review of the BFT aerial transect surveys. The ICCAT Secretariat provided the CIE reviewers with the background documents that were the focus of the review (marked with asterisks in the *List of References and Literature Cited* section below). Reviewers asked questions of the assessment authors via email, and the correspondence was followed by a virtual meeting on October 12, 2020, hosted by the ICCAT project contact, and attended by the assessment authors, CIE reviewers, and CIE lead coordinator, to discuss various technical questions regarding the 2019 re-analysis by Cañadas and Vázquez (2020) of the aerial survey time series, the main focus of the review. The primary deliverable products of the review are this Summary Report and the reports by the two reviewers, included in this report as appendices. The contract between ICCAT and the CIE specified that the CIE would make a virtual presentation of the review findings to the Bluefin Tuna Working Group, but the parties agreed to drop that work product.

The reviewers were instructed to address a list of specific tasks in their review reports, listed below in abbreviated form. The complete Terms of Reference are included in the reviewers' reports (Appendices 1 and 2).

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.

ii. Survey design. Evaluate the historical protocols and analytical approaches used in this survey as well as the recommended changes to the design procedures.

iii. Evaluate statistical treatment and index calculation of the Mediterranean survey time series.

iv. Suitability of the GBYP aerial survey (e.g., whether known logistical / biological / unaccountable factors are adequately addressed).

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

Finalization of the reviewers' written reports conformed to the following process. Draft versions of the CIE reviewer reports were initially examined by members of the Bluefin Tuna Working Group to identify any factual inaccuracies to be corrected, of which there were none. The review reports were then evaluated by the CIE Steering Committee and CIE Technical Team to check the reports for formatting, identify any readability issues, and to verify that the reports addressed all the Terms of Reference.

Description of Final Results Achieved

The CIE Steering Committee deemed that both reviewers' reports satisfied the Terms of Reference for the review. Both reports provide detailed commentary, directed at the requested specific tasks, that should be useful to the Bluefin Tuna Working Group in their deliberations on how to proceed with possible modifications to the BFT aerial survey. The following bullet points provide some notable highlights.

- *The feasibility and cost of conducting surveys using high-resolution video or still photos instead of observers, and from either piloted aircraft or long-distance drones, should be assessed. This should include a pilot study carried out say in Area A, with aircraft that have bubble windows and observers as well as high-resolution cameras, so that data from the two approaches can be compared*. (Recommendation 1 from Buckland)
- *Survey designs that offer improved spatial and temporal cover at acceptable cost should be developed*. (Recommendation 2 from Buckland)
- *Information on behaviour of spawning schools and on timing of spawning should be reviewed, to assess whether there might be time trends in the proportion of schools that are at the surface and hence detectable at a given time point during the surveys, and if so, whether an approach along the lines proposed by Quílez-Badía et al. (2016) is feasible*. (Recommendation 3 from Buckland)
- *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*. (Recommendation 1 from Vølstad)
- *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*. (Recommendation 2 from Vølstad)
- *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 3 from Vølstad)

The appendix reports should be consulted for details regarding the above recommendations and for additional observations, comments, and recommendations.

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.

List of References and Literature Cited

This list of references is a compilation of the reference lists provided in the reports by Dr. S.T. Buckland (Appendix 1) and by Dr. J.H. Vølstad (Appendix 2). Leading asterisks (*) indicate reference materials provided by the ICCAT Secretariat as focal or background references for the review.

*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. and Thomas, L. (2001). Introduction to distance sampling: estimating abundance of biological populations. Oxford University Press, Oxford.

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L. and Thomas, L. (eds) 2004. Advanced Distance Sampling. Oxford University Press, Oxford.

Buckland, S.T., Burt, M.L., Rexstad, E.A., Mellor, M., Williams, A.E. and Woodward, R. 2012. Aerial surveys of seabirds: the advent of digital methods. J. App. Ecol. 49: 960-967.

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

*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

*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.

*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.

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.

*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.

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.

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., 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

Miller, D.L., L-M- Burt, E.A. Rexstad, 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

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

*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 Natale, 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 surveys 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.

Appendices

Appendix 1. Review report by Dr. S.T. Buckland

INDEPENDENT PEER 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)

SUMMARY
Aerial surveys of the bluefin tuna stocks of the Mediterranean were carried out in each of seven years between 2010 and 2019 inclusive. The most recent time series of estimates shows large differences from previous estimates, and high interannual variation both within and between regions. I review the survey design, the field methods, and the methods of analysis. I conclude that spatial and temporal coverage of the survey may be insufficient to yield a reliable time series of estimates, especially if spawning locations and spawning times vary across years. Given the difficulties that observers face in recording reliable data for the line transect method, I suggest that the use of high-resolution imagery be explored, possibly in conjunction with long-distance drones. Video or still images taken from higher altitude provide a permanent record, allowing verifiability. I review the methods of analysis used to date, and suggest more advanced model-based methods to complement the design-based methods used to date. I also note the large inconsistencies in some estimates, which point to problems in the computer code.

1. Background

A key part of the Atlantic-wide Research Programme on Bluefin Tuna is aerial surveys of four areas of the Mediterranean Sea that are considered to be the main spawning areas. The latest estimates of abundance, spanning the period 2010-2019, exhibit large variability (Cañadas and Vázquez, 2020), which raises the question of whether the surveys are able to provide reliable data to inform management of the stocks.

This review was commissioned by the Center for Independent Experts (CIE) to provide an independent review of the 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). 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

e. 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.

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

2. Description of the individual reviewer's role in the review activities

My expertise is in distance sampling. The bluefin tuna surveys use line transect sampling, which is the most widely-used distance sampling method. My review therefore concentrates on survey design, field methods and analysis methods adopted for the bluefin line transect surveys. I note that Fonteneau et al. (2013) proposed that other approaches to estimating adult biomass be investigated. I do not have the expertise to advise on their suggestions, so I have not addressed them. Di Natale and Idrissi (2013) compared the strategy of surveying spawning adults with that of surveying juveniles, and concluded that the former was preferred. MRAG (2016) concluded that aerial surveys and close-kin mark-recapture methods are the leading contenders. Again, I don't have sufficient experience of close-kin mark-recapture to comment on that approach.

3. Review of methods

Cañadas and Vázquez (2020) report annual abundance estimates that vary widely. They also vary widely from one set of analyses to another. Even within a set of analyses, there seem to be large inconsistencies (see Appendix 1). It is worth considering possible sources of such large inconsistencies:

1. Survey design.
2. Field methods.
3. Methods of analysis.

I consider each of these in turn.

2.1 Survey design

Survey design varied across years, and the boundaries seem to have been rather arbitrarily drawn. Detections in all areas appear to occur right up to the boundary of areas of overlap across surveys (and where there was search effort, beyond the boundary), suggesting that these overlap areas are too small. Movements in spawning areas across time may therefore compromise the time series of abundance estimates. Areas could be expanded, or a 'buffer zone' created around each area, with a lower level of effort. Given that schools occur through much of the Mediterranean, it may be necessary to develop a stratified design, with good coverage in existing survey areas, and a low level of coverage through all or most of the rest of the Mediterranean. Survey effort might also be spread through the time over which schools might spawn, with a view to developing a spatio-temporal model of spawning school density, from which an annual index might be estimated (see Section 2.3).

The issue of spatial coverage of the survey was raised by the GBYP Steering Committee in 2012, and Cañadas and Vázquez (2013) subsequently developed a proposal that included survey effort with lower coverage through much of the Mediterranean outside the main survey areas. The only areas excluded were ones with no historical spawning, or ones for which flight permits could not be obtained. 'Outside' areas were surveyed in 2013 and 2015. Because effort in the main survey areas was substantially reduced to keep total cost within budget, Cañadas and Ben Mhamed (2016) recommended that in future, outside areas should not be sampled unless additional resources were allocated. ICCAT (2012) and MRAG (2016) both provide a useful summary of the advantages and disadvantages of surveying a wider area.

2.2 Field methods

Survey methods have varied across surveys, with bubble windows used in some surveys and not others, and with different companies doing the surveys. It is important to keep methods as comparable as possible.

The standard line transect sampling method assumes that a school located on the line is detected with certainty. It is questionable whether this is the case, even when a bubble window is used. If line transect surveys are to be continued, consideration should be given to the use of double-platform methods, so that this assumption can be relaxed. In addition, there is a question of whether all schools are available for detection at the time of the surveys. Double-platform methods can address the problem of schools at the surface and on the trackline that are not detected, but they do not address the problem of schools that are not at the surface (i.e. are unavailable for detection) when the aircraft passes overhead. Cañadas and Ben Mhamed (2016) estimated that only 47% of schools were available for detection at a given time, but presumably did not have adequate data to assess how this proportion varies across years, or through the spawning season within a year.

It is possible that observers in different years were searching in different ways. If one is not already in place, an observer training programme should be implemented that includes sufficient detail of the analysis methods that observers understand what is needed if the method is to work. In the absence of such training, observers tend to try to maximise the number of schools detected, which results in detections at large distances which are of little value in the line transect analysis, and potentially a failure to detect small schools near the line. The training should address how to search so that for schools at the surface, probability of detection is at or close to one near the line, and does not fall off rapidly with distance from the line.

Serious consideration should be given to using high-resolution imagery (video or stills) instead of observers. This allows survey aircraft to fly at higher altitude (reducing safety concerns), and strips will be surveyed without the problem of detectability falling off with distance. Also, the images provide a permanent record of what was seen, which can be verified and re-analysed as required. By contrast, the data recorded by the observers are limited by the fact that they are the observers' interpretation of fleeting images. Thus distance estimates and school sizes and weights cannot be verified, and there is no information on missed schools. Using high-resolution imagery also allows the use of long-distance drones rather than piloted aircraft (subject to any restrictions on use of drones in the areas to be surveyed). In the UK, two companies, HiDef (video) and APEM (stills) routinely use high-resolution imagery for marine surveys (see for example Buckland et al., 2012). Ideally, the method would be calibrated against current methods necessitating two surveys in at least one year, but given the lack of consistency in the methods used to date, such a calibration exercise would probably have limited value. (Di Natale, 2016, was similarly pessimistic about the potential for calibrating surveys against each other, to try to account for this lack of consistency.) Instead, each estimate might be assumed to be an unbiased estimate of abundance, in which case all estimates would be comparable – at least in principle.

2.3 Methods of analysis

Current estimates should be reviewed and revised, in light of the issues raised in the appendix. It would appear that the estimates should vary by far less than do some of the estimates tabulated in Cañadas and Vázquez (2020). The histograms of distance estimates given by Cañadas and Vázquez (2020) are badly spiked; such data are notoriously difficult to model reliably. It appears that many schools are only detected if they are very close to the trackline (within 300m), and there is very little benefit in searching out to 5km or beyond in this circumstance. This alone suggests that survey methods should be revised. This could for example involve bubble windows, with search concentrated within say 1km either side of the line, but a better option would be to use high-resolution imagery from higher altitude, to allow good quality data on a strip of width perhaps 500m or so. Given the extremely rapid fall-off in detectability with distance when using observers, there would be little loss and substantial gain from near-complete detection in a much narrower strip.

Left-truncation of the distance data may affect estimates for 2010-2013. At the start of Section 2.1 of Cañadas and Vázquez (2020), it is stated that *esw* is the reciprocal of the probability (which I will call P_a, and which is referred to as p in Cañadas and Vázquez, 2020) of detecting a school that is within the strip. In fact, $esw = w.P_a$ where w is the right-truncation distance. The simple (but not unique) way to handle left-truncation is to define $P_a = \int_u^w g(y)dy$ where $g(y)$ is the probability that a group at distance y from the line is detected, for $u \leq y \leq w$, so that u is the left-truncation distance, which is zero when there is no left-truncation. Thus given a fitted detection function $\hat{g}(y)$, an estimate of P_a and hence of esw can be obtained. (With covariates in the detection function, this estimation can be carried out for each individual detection, and averaged across detections.) If this is done, I would expect estimates of P_a to be smaller for 2010-2013, for which left-truncation was implemented, than for subsequent years, yet the estimates tabulated are larger. This may be because the effect of left-truncation was slight, but the calculations may need to be checked.

I note here a very large discrepancy of the histograms of perpendicular distances for 2010-2013 of Cañadas and Vázquez (2020) and Figure 1 of ICCAT (2012), showing the perpendicular distance distribution for 2010. For the histograms of Cañadas and Vázquez, left-truncation at 109m appears to eliminate the effect of not being able to detect schools under the aircraft that have no bubble window, there is a large spike of detections between this distance and 300m, and estimated probability of detection for a school 3km from the lines is estimated to be well below 0.2. By contrast, Figure 1 of ICCAT (2012) indicates that no schools were detected within 1km of the line, and the spike of detections occurs at 3km. This discrepancy may have already been investigated. If not, it should be.

Cañadas and Vázquez (2020) have used the MCDS engine of Distance, with probability of detection varying by school size. In this case, a Horvitz-Thompson-like estimator is used, where school size appears in the numerator inside the summation of the estimator. This only reduces to the first equation of Section 2.1 of Cañadas and Vázquez (2020) when probability of detection is assumed to be a function of distance from the line only. Using the mean school size as in Section 2.1 of Cañadas and Vázquez (2020) will introduce bias when probability of detection is a function of school size as here.

Estimation of animal abundance and total weight requires reliable estimates of school size and weight of school. Vázquez and Cañadas (2019) note that observers' estimates are often poor, and suggest that calibration is needed.

They also tested different cameras for obtaining images of schools, potentially allowing calibration to be carried out. Such calibration would seem to be prudent. Alternatively, surveys might switch from line transect surveys with observers to strip transect surveys with high-resolution imagery. Grup Air-Med (2019) successfully implemented a calibration trial, demonstrating the feasibility of calibrating observer estimates.

Consideration should be given to spatial modelling of the data, which may help inform estimation of the extent of spawning areas, and whether these areas move around across years. Density surface modelling (Hedley and Buckland, 2004) should be adequate for this purpose, although more sophisticated point-process models (Yuan et al., 2017) might be considered. Given possible variation in spawning times, it would be advisable to include date in the model, and so develop a spatio-temporal model of spawning density. An index could then be defined in several ways. For example, averaging over space within any given area would give spawning density as a function of date within any given year, and an index could be estimated by integrating over date.

Good progress towards developing a model-based approach was reported by Cañadas and Ben Mhamed (2016), and further development of that approach to show how density varies over space and time might prove useful.

4. Summary of findings for each TOR

In this section, I list each TOR in turn, and summarise my findings relevant to it.

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?

I found the survey documentation and supporting material to lack detail in places, so that I was unsure of how analyses had been conducted.

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 survey areas do not fully cover spawning areas, and there is no allowance made for potential changes over time in where schools spawn, or in dates of the main spawning period. While design-based analysis methods potentially yield robust estimates of density within survey areas at the time of the surveys, they offer no insights into how density varies within areas, and whether high density regions extend beyond area boundaries, nor do they indicate whether the surveys captured the main spawning period. Model-based methods applied to past data might help inform design changes, in terms of both spatial extent of survey areas and temporal extent of surveys within a year.

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

If field methods are considered part of survey design, then the introduction of bubble windows has been helpful, although it compromises to some extent the comparability of estimates across years. Improved training of observers would aid comparability. Ideally, the same company and same teams of observers would be used in each survey year. However, now may be the time to consider replacing observer line transect surveys by surveys using high-resolution imagery.

    c. Are there further unaccounted for factors?

Changes in aircraft and observer teams are not fully accounted for. Cañadas and Vázquez (2020) use left-truncation for those surveys conducted wholly or partially without bubble windows, which should increase comparability between those estimates and estimates for later years.

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?

Model-based analysis methods may help to assess this, and possibly provide improved and more precise estimates 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?

Detection of the same school from different transects, either because the school is visible from more than one transect or because the school moves from one strip to another, does not systematically bias estimates. See Buckland et al. (2001:36) for discussion of the first issue, and Buckland et al. (2001:32) for discussion of the second issue. Provided a Horvitz-Thompson-like estimator (Buckland et al., 2004:38-45) is used, modelling probability of detection as a function of school size does not bias estimation of abundance or biomass. However, if density is estimated simply by multiplying school density by mean school size, as stated in the first equation of Section 2.1 of Cañadas and Vázquez (2020), then bias does occur; abundance or biomass would be over-estimated, because mean school size in the sample overestimated mean school size in the population, as larger schools have a higher probability of detection than smaller schools.

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

Model-based methods may prove more effective. See recommendations.

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?

There are problems with the estimates as presented. See Appendix for examples. If these issues can be resolved, a useful index should be achievable.

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

I believe that a model-based approach, developing a spatio-temporal model, might give a more informative index. See recommendations.

iv. Suitability of GBYP aerial survey

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

Partial only. Spatial coverage is not sufficient, and temporal coverage within a year may also be problematic.

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

See recommendations.

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?

I believe a model-based approach may help address these issues. See recommendations.

d. Provide general recommendations for potential improvements

See recommendations.

e. 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.

See recommendations.

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

See recommendations.

5. Discussion and recommendations

Surveys by aircraft or long-distance drones at relatively high altitude and using high-resolution imagery offer several advantages over line transect surveys using observers. First, the raw data are images, which provide a permanent record of the survey and for which estimates can be updated if analysis methods are improved. By contrast, for line transect surveys, only the observers' interpretations of what was seen are available to the analyst. Second, detectability of schools should not be a function of distance from the transect line, and so the difficulty of modelling the rapid fall-off in probability of detection with distance from the line in the line transect surveys is avoided, as is the difficulty in detecting schools directly below the aircraft. Third, estimates of school size and weight can be estimated objectively from the images, rather than subjectively by observers. Fourth, once suitable image analysis software has been acquired or developed, analysis should be relatively straightforward and rapid.

Recommendation 1. The feasibility and cost of conducting surveys using high-resolution video or stills instead of observers, and from either piloted aircraft or long-distance drones, should be assessed. This should include a pilot study carried out say in Area A, with aircraft that have bubble windows and observers as well as high-resolution cameras, so that data from the two approaches can be compared.

Many spawning schools occur outside of the areas of overlap across years, and outside the dates over which the surveys are conducted.
Recommendation 2. Survey designs that offer improved spatial and temporal cover at acceptable cost should be developed.

The surveys can detect only spawning schools that are at the surface when the aircraft passes over. Quílez-Badía et al. (2016) proposed a combined aerial survey and tagging programme to allow estimation of the proportion of schools available to be detected in the survey.
Recommendation 3. Information on behaviour of spawning schools and on timing of spawning should be reviewed, to assess whether there might be time trends in the proportion of schools that are at the surface and hence detectable at a given time point during the surveys, and if so, whether an approach along the lines proposed by Quílez-Badía et al. (2016) is feasible.

Standard line transect methods assume that all schools on the line are detected. 'Double-platform' surveys allow estimation of the probability of detection, without assuming that detection of surface schools on the line is certain. Having front observers as one 'platform' and rear observers as the other is unlikely to be effective if only one platform has bubble windows. A better option is to have the observers as one platform and a high-resolution camera (or multiple cameras, to increase field of view) as the second. Another option of having two aircraft flying in tandem might be effective, but would double field costs.
Recommendation 4. If line transect surveys are continued, the feasibility of carrying out a 'double-platform' survey should be assessed, to allow the estimation of a detection function without having to assume that detection of schools at the surface and on the line is certain.

Observers must seek to detect all or nearly all schools near the line, ensure that probability of detection does not fall sharply with distance from the line, estimate the size and weight of detected schools, and generate comparable data over time. (The need to train observers on how to search was recognised by ICCAT, 2012.)
Recommendation 5. If line transect surveys are continued, and if such a programme does not already exist, an observer training programme should be developed, to ensure that observers understand what is required of them, and to gain experience in detecting schools, and estimating their size and weight.

Trials suggest that observers' estimates of school size and weight are very variable, and some observers' estimates may have large bias.
Recommendation 6. If line transect surveys are continued, it would seem essential to calibrate observers' school size and weight estimates, using photographic images of at least a sample of detected schools.

While design-based methods generally yield robust estimates of density, model-based methods offer several advantages: spatial models will indicate whether high densities potentially occur outside of surveyed areas; spatio-temporal models will indicate whether temporal coverage of the surveys is adequate; a degree of extrapolation to unsurveyed areas may be possible; precision may be improved.
Recommendation 7. A spatial (e.g. density surface) model should be fitted to the data for each survey year. This should be extended to a spatio-temporal model, with date within a year being added to the spatial model, to explore the options for estimating an annual index from the fitted model, which will allow for spatial and temporal variation in spawning densities across years.

Estimates of Cañadas and Vázquez (2020) show some inconsistencies that are probably a result of problems in the computer code.
Recommendation 8. If these inconsistencies cannot be resolved, an experienced user of the R distance software should be hired, either to review the code used to generate the estimates, or to write independent code to do the same analyses, allowing differences between the new estimates and the estimates of Cañadas and Vázquez (2020) to be identified and resolved.

The methods of analysis used by Cañadas and Vázquez (2020) are not fully described, and there appear to be inaccuracies in the description. A full description of the methods would allow verification of estimates.
Recommendation 9. The methods of analysis should be described in sufficient detail that an independent analyst could repeat the analysis, and the computer code used should be available.

References

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L. and Thomas, L. 2001. *Introduction to Distance Sampling.* Oxford University Press, Oxford.

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L. and Thomas, L. (eds) 2004. *Advanced Distance Sampling.* Oxford University Press, Oxford.

Buckland, S.T., Burt, M.L., Rexstad, E.A., Mellor, M., Williams, A.E. and Woodward, R. 2012. Aerial surveys of seabirds: the advent of digital methods. J. App. Ecol. 49: 960-967.

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

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.

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.

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.

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

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.

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

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.

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

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.

Appendix 1: Details on estimates' inconsistencies

Here, I give details of some of the inconsistencies in the estimates of Cañadas and Vázquez (2020).

1. Table 3.2.1. Estimates of 'average *p*' are consistently higher for 2010-13 than for later years, but as noted in the report, as there was left-truncation for these early years, I would expect them to be lower. However, the amount of left-truncation was relatively modest at 109m, so this may be just a chance effect.
2. At the start of section 2.1, the authors provide an equation for estimating density. This equation includes mean school size, which will give a biased estimate of density because probability of detection of a school was found to be a function of school size. However, if mean school size is dropped from this equation, it should give an estimate of school density. Applying this equation to the estimates of Table 4.1.1, school density is estimated as number of schools detected on effort divided by (twice the esw times the transect length). This gives the estimated densities of row 2 below, while the authors' estimates are in row 3:

2010	2011	2013	2015	2017	2018	2019
0.00049	0.00048	0.00078	0.00093	0.00221	0.00229	0.00179
0.00016	0.00022	0.00035	0.00058	0.00523	0.00402	0.00363

   There is clearly no correspondence in these estimates. While the discrepancies for 2010-13 may have something to do with left-truncation of the data, this is not the case for later years. Many of the other estimates in this table show the same discrepancies.
3. Expected cluster size in Table 4.1.1 for 2013 and 2015 are similar (415 vs 445), yet expected weight for 2013 is 0.505 and for 2015 is 79.361, indicating that average fish weight was more than 70 times greater in 2015 than in 2013. I do not have the data to check this.
4. As noted by the authors, total is estimated as density multiplied by area. Thus for area A in Table 4.1.1, total weight should be density of weight multiplied by 61,837. Assuming weight densities are correct, my estimates of total weight are in row 2 below, while the authors' estimates are in row 3.

2010	2011	2013	2015	2017	2018	2019
30300	8657	12367	2832135	9402934	10865998	6663555
2119	963	1946	2832	9403	10866	6664

   The authors do not specify their units, but for 2015-19, presumably the difference is kg vs tonnes. However, the estimates for 2010-13 bear no correspondence at all, and the density estimates of Table 4.1.1 appear to be orders of magnitude out.

Similar issues appear to be present in other tables, but I haven't checked these.

Appendix 2: 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 20200.

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).

Appendix 2. Review report by Dr. Jon Helge Vølstad

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)

SUMMARY

Aerial surveys with observers have been conducted in the Mediterranean in 2010, 2011, 2013, 2015, 2017, 2018, 2019 to provide indices of the abundance of the spawning population of the eastern stock of bluefin tuna (BFT). The recognized DISTANCE software has been used in the selection of random-systematic transects within subareas annually. Sampling efforts have focused on four subareas that are assumed to represent the main spawning areas. The spatial coverage was extended in 2013 and 2015 to cover the majority of the potential spawning areas in the Mediterranean Sea. According to the Terms of Reference, the focus of this review is the 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. Based on a review of extensive background material provided through the Center of Independent Experts there is strong evidence that a long-term monitoring program will require a survey design that covers much of the Mediterranean. Recognizing cost-limitations, an option is to continue annual spatial and temporal sampling coverage in the four main spawning areas at current levels, and to cover the remaining spawning area with less effort. This area outside the main spawning grounds could for example be split into smaller survey regions (blocks) that each can be surveyed with synoptic coverage in a single year, achieving full coverage of all blocks over several years. 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. In particular, such methods could ensure standardized counts of individual animals (and their lengths) within an accurately defined narrow transect width. Such methods could reduce cost and eliminate many of the sources of errors that are identified for the current field data collections with observers.

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 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.

(a) 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.

(b) 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.

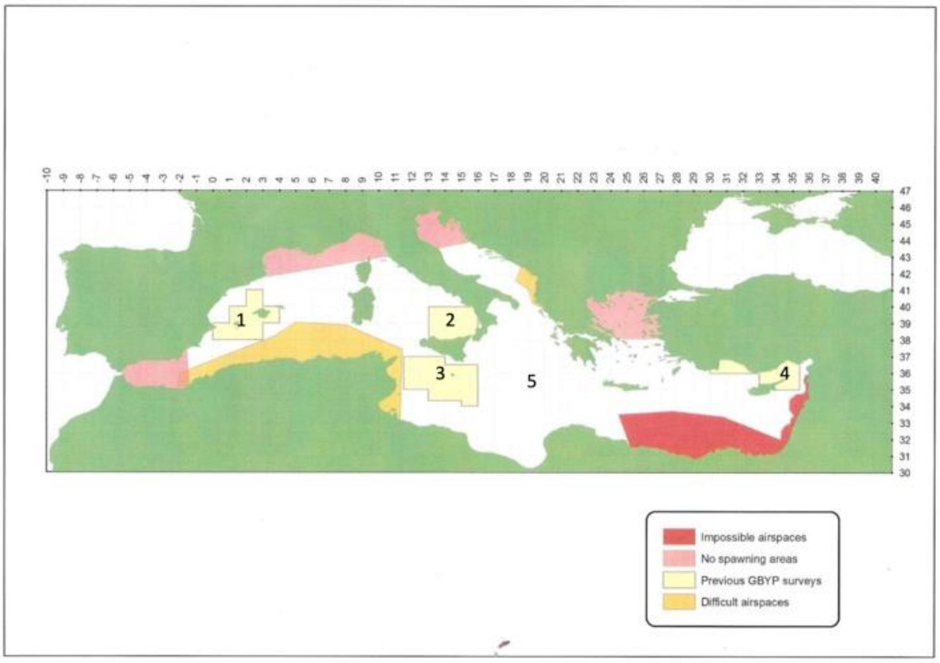

Figure 1. Blocks considered for this work

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

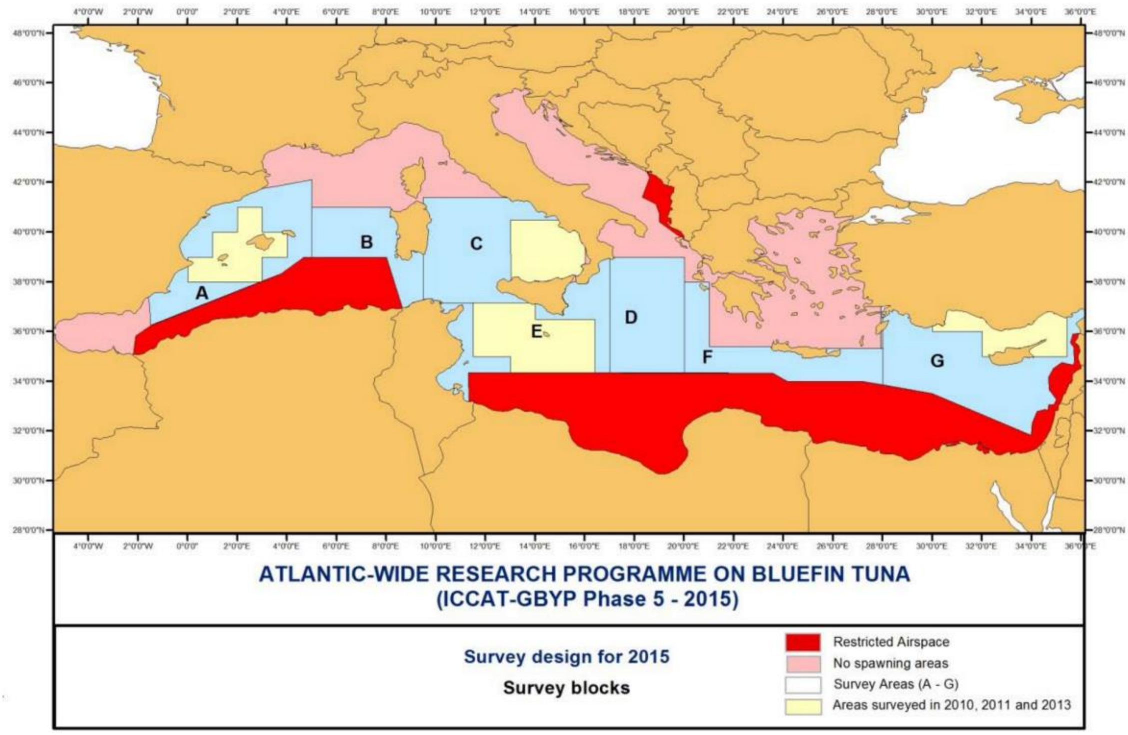

Figure 1. Survey blocks for 2015.

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

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. 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. 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. 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.

(c) 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

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

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 1.x in Cañadas and Vázquez (2020) versus estimates from the estimator above, marked by * in the example below.

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

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 ($61837 km^2$), while estimates of total abundance based on density of animals are consistent with standard estimates.

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

Due to many inconsistencies in the estimates from the 2019 re-analysis, 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

vi. 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).

vii. 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?

viii. 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.

ix. 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

x. 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.

xi. 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.

6. References of material provided for the review and other references used in support of the review.

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 (Balaenopteraacutorostrata) 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.

Tables and Figures

Table 1. 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 2- 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 (km²)	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 3 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 4. 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 5. 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

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):

vii. 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?

viii. 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?

ix. 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?

x. 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

xi. 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.

xii. 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).