

**ATLANTIC-WIDE RESEARCH PROGRAMME ON BLUEFIN TUNA
(ICCAT-GBYP – PHASE 5 - 2015)**

**POWER ANALYSIS AND COST-BENEFIT ANALYSIS FOR THE ICCAT
GBYP AERIAL SURVEY ON BLUEFIN TUNA SPAWNING
AGGREGATIONS**

Tender No. ICCAT-GBYP 08/2015, ITEM A

Final Draft Report

12 February 2016

Ana Cañadas & Abdelouahed Ben Mhamed

Alnilam Research and Conservation Ltd

Pradillos 29, 28491 Navacerrada, Madrid, Spain

Background

The objectives of the comprehensive ICCAT Atlantic-Wide Research Programme on Bluefin Tuna (GBYP) are to improve (a) the understanding of key biological and ecological processes, (b) current assessment methodology, (c) the management procedures, and (d) advice.

Key tasks are to reduce uncertainty in stock assessment and to provide robust management advice. This requires improved knowledge of key biological processes and parameters. However, currently almost all the data used in stock assessments are obtained from the fisheries-dependent data. It is therefore important to obtain data from alternative sources, i.e. aerial survey or tagging studies, in order to verify the assumptions made when conducting the assessments.

Important parts of the mayor research tasks under the ICCAT Atlantic-wide Bluefin Tuna Research Programme (GBYP) are an aerial survey on bluefin spawning aggregations in the Mediterranean Sea, a large, wide and intensive scientific tagging program to address several important biological and ecological topics regarding Atlantic bluefin tuna, and an Atlantic-wide sampling programme including a wide range of biological studies (ageing, genetics, micro.-chemistry and otolith shape analyses).

The comprehensive GBYP activities were initially assessed by a panel of external reviewers in 2013. The ICCAT GBYP Steering Committee recommended to carry out analysis on the main GBYP research activities before the end of Phase 5 (February 21, 2016), in order to have a more focused overview of the works carried out so far and have further details for adopting the best research strategy in Phase 6.

At the 2015 SCRS meeting it was recognized that if the GBYP continues to use the same methods without evaluating how the data and knowledge gained will improve the scientific advice framework, the programme may fail to meet management objectives. To avoid this potential risk it is essential to conduct a cost/benefit analysis to help design a programme which will meet the programme objectives in a cost-effective way. This also requires a clear definition of objectives and milestones to monitor progress.

SCRS/2015/146 detailed how to conduct a power analysis for one of the GBYP programmes, specifically for the aerial survey. The benefits of a particular aerial survey design also depend on the knowledge gained by other programmes of the GBYP, i.e. on population structure and behavior. Conducting a full cost-benefit analysis of the main research programmes of the GBYP, however, will not allow a decision on the aerial survey programme to be made before the start of the 2016 campaign.

Taking into account that the recent SCRS BFT species group meeting clarified that several fishery-related indices could not be updated and therefore used in the next assessment, the potential importance of any additional GBYP data may be increased for future assessments and for the MSE process.

Objectives

The primary objective is to evaluate whether the aerial survey can provide a reliable and robust index of abundance of the spawning stock with sufficient precision to be used in BFT stock assessment. The specific objectives are:

1. A comprehensive review of the ICCAT GBYP surveys conducted so far and an analysis of the power of the current design to detect changes in the stock for a range of population growth rates.
2. An analysis of the design and costs to detect a range of population growth rates in 3, 5 and 10 years with no reduction in the additional variance.
3. An analysis of the design and costs to detect a range of population growth rates in 3, 5 and 10 years with a reduction in additional variance. Discussion on how other programmes under the GBYP, e.g. tagging or a better understanding or habitat usage should be designed.
4. Provision of Recommendations.

I. Comprehensive review of the ICCAT GBYP surveys conducted so far and an analysis of the power of the current design to detect changes in the stock for a range of population growth rates

I.1 Review of how the different designs used in the surveys conducted in the first four years affected the results

The ICCAT GBYP is now known in marine science as a one of the most challenging and efficient research programme, currently the most comprehensive on one single marine species. The value of the programme is increased by the fact that some scientific activities, namely the aerial survey for spawning aggregations, were made possible only because they have been supported by an international organization and by the ICCAT CPCs concerned, all working for the common purpose of making feasible an otherwise impossible endeavor. The coordination team faced enormous difficulties, particularly when extended surveys have been planned. The aerial surveys carried out so far by ICCAT GBYP have been the first over a so extended area and in so many different FIRs belonging to a number of countries, having all different rules. Just the fact that these surveys have been carried out is a standing alone result for ICCAT. The scientific results of the survey can be obviously discussed and for sure they could improve under more favourable conditions and medium-term planning.

I.1.1 Data availability and analysis

The information available from previous reports for the surveys of 2010, 2011, 2013 and 2015 was used, including the parallel work being done as an in-depth analyses of the collected data to assess the reliability and consistency with which the survey protocols have been implemented within years among the different companies and airplanes.

I.1.2 Results and discussion

a) Spatial extent of the surveys

Most of the inside sub-areas have been slightly (A, C, G) or to a greater extent (E) modified across the years. Table 1 shows these variations, as well as the overlap areas defined in 2015; i.e. the areas common to all years and therefore comparable (see report Cañadas and Vazquez, 2015).

Table 1. Surface area (km²) of the inside sub-areas surveyed each year, and the surface area of the overlap areas defined in 2015

Year	Inside sub-areas				Total
	A	C	E	G	
2010	62,150	54,636	132,453	68,819	318,058
2011	62,150	54,636	104,366		221,151
2013	62,194	56,329	82,054	56,329	254,754
2015	62,150	64,610	117,718	68,013	312,491
Overlap	61,933	53,868	93,614	56,211	265,627

The CVs are generally lower when estimating abundance in the overlap areas is compared with the pre-overlap areas (see report Cañadas and Vazquez, 2015) but equally important is that by using every year the same area, comparisons are possible without the potential bias created by changing the extent of the areas. Furthermore, due to logistic constraints, some years there have been changes in the survey areas after the survey design was done, and some sectors remained unsurveyed (the problem

is mostly in E), compromising the equal coverage probability assumption of the line transect methodology. These changes and unsurveyed sectors should be avoided in future surveys. Of course, due to the many problems in the Mediterranean (security, political issues, sudden flight restrictions, etc.), these facts cannot be fully excluded.

The choice of the areas was originally done based on the available scientific knowledge of the spawning behaviour of BFT by the experts and of 3-year VMS data. We **recommend** that the selected overlap areas are maintained as stand-alone areas for the survey design, independently on potential extensions when it is considered necessary, so comparisons are possible in future surveys.

Another issue to consider is to survey or not the outside areas. The effort in the inside areas was very much reduced in 2013 and 2015 due to the allocation of 50% of the effort to the outside areas. The amount of observations in the outside areas have been minimum, not even allowing a robust abundance estimation in most cases. The reduction of effort allocated in the inside areas in 2013 and 2015 has yielded a reduced amount of observations of BFT in them and an increased CV in most cases (decrease of precision in the estimates), which is an undesirable effect, especially when there is little or no gain in information in the outside areas. This large reduction in effort in the inside areas makes also much more difficult to compare between years with so different survey effort. In this sense, we **recommend** that the amount of effort is maintained in the inside areas every year of survey. When an extended survey will be decided, then it would be necessary to allocate extra resources to survey the outside areas for checking any possible presence of bluefin tuna spawning aggregation in other areas, always maintaining the level of effort in the inside areas.

b) Timing of the survey

The timing of the survey was selected each year based on the knowledge of the biology and ecology of BFT during the spawning season, which occurs generally between the second part of May and the first part of July, with a peak in June. It is known the relationship between BFT spawning behaviour and some oceanographic features such as the thermocline and sea surface temperature (SCRS paper Di Natale *et al*, 2015).

The timing of the survey has varied slightly from year to year (Table 2) mainly due to logistic reasons and also trying to adapt to the suspected starting of the spawning season according to the recorded values of sea surface temperature in the Mediterranean in late spring and early summer. In 2011 surveys started a few days later than in 2010, 2013 and 2015 in sub-area A and C, while in E it was 2013 the year with latest start, due to permits problems. In general surveys start earlier in area A, and the latest in area G and E, where permit issues are usually more relevant. Date for ending the surveys are more variable with no clear pattern, probably more dependent on weather conditions to finish all or most of the tracks designed, and most finish around end of June or mid-July, except in 2010 when area E was prolonged to beginning of August.

Table 2. Dates of survey in 2010 – 2015. ‘n’ is the number of survey days within the survey period.

Year	A inside			C inside			E inside			G inside			Total		
	n	Start	End	n	Start	End									
2010	13	01/06	02/07	8	05/06	29/06	23	06/06	03/08	14	05/06	30/06	58	01/06	03/08
2011	19	15/06	11/07	13	19/06	08/07	15	13/06	29/06				47	13/06	11/07
2013	17	06/06	06/07	5	18/06	28/06	11	22/06	12/07	9	20/06	15/07	42	06/06	15/07
2015	20	01/06	11/07	5	01/06	06/06	11	12/06	28/06	5	20/06	25/06	41	01/06	11/07
Total	69	01/06	11/07	31	05/06	06/06	60	06/06	28/06	28	05/06	25/06	188	01/06	11/07

It is important to “capture” within the survey time the main spawning season of BFT. Therefore, as long as there are no too large variations in the timing, we **recommend** that the survey should always cover the various areas around the usual peak of the spawning season. Even if organizing the surveys start time according to analysis, by the relevant experts, of the information on oceanographic conditions in the weeks/days prior to the usual spawning season, will possibly create important problems for the contracts, because it implies to pay for a longer availability of both aircrafts and crews, and it would possibly create additional problems for permits and additional costs.

c) Duration of the survey

Despite having an extended period of around one month to do the surveys each year, the actual days spent surveying are much less, going from 5 to 23 days (the latter in E 2010, when two months in total were allocated). In C, E and G the number of days for surveying the inside areas in 2013 and 2015 was reduced in comparison to 2010 and 2011; this was mostly induced by the time spent in surveying the outside areas. It is important to allocate enough time to the survey in order to cover the whole extent of the designed tracks, taking into account the stand-by days caused by unfavourable weather conditions. So far, GBYP coordination had provided a punctual overview of the percentage of unfavourable days in each area after each survey and this may help the prediction. Therefore, the duration itself of the survey does not seem to affect its efficiency, because all the primary designed tracks were covered (except in E in some cases) in all surveys and in some cases even extra designed tracks were done. Hence, it is important to maintain enough duration (minimum one month, as the previous years) to ensure that there is enough time to cover everything despite the days of bad weather.

d) Range of planes and spotters utilized

The use of different aircrafts and many different spotters, a survey procedure that is inevitable for covering so many areas at the same time within a very limited period of time, introduces undesirable variability when trying to compare estimates. It is practically impossible to conclude whether differences observed among areas or years are due to real differences in abundance, or to differences induced by different types of aircrafts or/and by the individual observers (different experiences, different skills, ways of dealing with the protocols, criteria to read the inclinometers, criteria to estimate school sizes and weight, etc.). Furthermore, the variables linked to the environmental conditions (in a broad comprehensive sense) and to the both fast-moving platform and target are factors very difficult to duly assess.

According to the analysis done within the contract “Elaboration of 2015 data from the aerial survey on spawning aggregations” (see report Cañadas and Vazquez, 2016), quite surprisingly Cessna has a positive effect on the encounter rate compared to Partenavia. As described in that report, after comparing the effect of the aircraft type on BFT detections and also on all species pooled together: “In the case of BFT it could be argued that as the different airplanes were used in different areas, these effect may be mainly due to the different densities of BFT in the different areas. But the fact that the same effect is observed for all species pooled together suggests that it is not only a matter of density but maybe the type of aircraft has indeed some real effect for some reason. On the other hand, it does not have any effect on the distance of detection when considering all species together, but it does when looking only at BFT. With Cessna there is a tendency for detection at closer distances in average, and much larger with Partenavia in average. Whether this is an effect of different searching behaviours by the different teams using each airplane or due to the configuration of the airplane itself, is unknown.”

In terms of spotters, the very large amount of different spotters even within a single team, precludes any reliable analysis, comparison or conclusion, but also any possible calibration. In terms of school sizes, in average, SS give larger estimates than PS, by 22% in school size and by 16% by weight. But these differences (which are quite logical, due to the different experience and skills) are not constant and are just an average, with differences ranging from negative to positive in both cases, and with no observable pattern.

The number of observations of BFT per observer is very variable and generally very low in most cases. Therefore an appropriate exploration of the searching patterns of individual observers would only be possible for very few of them. Hence, it is in practice impossible to make reliable comparisons among them. However, an exploration has been done to compare Professional Spotters as a whole (PS) with Scientific Spotters (SS). Results are shown in the report for “Elaboration of 2015 data from the aerial survey on spawning aggregations” (see report Cañadas and Vazquez 2016). The main conclusions are that (a) PS prove to be much more efficient in finding schools of BFT than SS (much larger encounter rate), presumably due to their large experience in this task; and (b) it is clear that in most cases PS tend to look further away than SS, despite having repeatedly insisted that the most important observations for the analysis are those at the shortest distances. This searching pattern by the PS needs to be improved for searching more at closer distances.

Furthermore, apart from the impossibility to make any analysis on individual spotters, the fact of having so many different ones, with different degree of experience, introduces huge variability and noise in the data. This is an issue that should be discussed before the next survey, but if the survey will be carried out on the four overlapping areas, even this variability will be smoothed; furthermore, it is **recommended** to possibly always use the same spotters in the same areas, because this approach will reduce biases due to changes in crew components, as it was discussed during the SCRS BFT Species Group meeting in 2015.

e) Methods for estimating perpendicular distances

Two methods have been used to estimate perpendicular distances: (a) with inclinometer (vertical angle transformed later in perpendicular distance); and (b) by circling over the animals and calculating *a posteriori* the distance between the track and the centre of the circle over the animals.

Both methods have their own associated problems. Measuring the angle with the inclinometer may be difficult at times, for example, due to lack of experience, not having enough time for the measurement before the school is missed, or when the school is spotted by the PS and the plane turns towards it before the SS with inclinometer takes a proper measurement; furthermore, any possible turbulence may increase imprecision. On the other hand, as demonstrated in the report by Cañadas and Vazquez (2016), circling is not always possible, as sometimes the school is missed before arriving to it; but even when circling is properly done, sometimes it is difficult to estimate the centre of such circles (theoretically the position of the school) when they are not concentric. This happens mostly when the target is moving, which is a usual condition at least for bluefin tuna.

When both measurements are taken, there is a huge variability in the difference between the perpendicular distance estimates derived from both, either in favour of one or the other method. Not a clear pattern can be extracted from these differences due to such extreme variability. It is very difficult also to assess which one is more precise, as no ground truth is available, although logic would lead to think that estimating a position *a posteriori* using GPS coordinates, when it is possible and circling is well done, is a safer method.

In the survey report by Cañadas and Vazquez (2016) there is an experiment to test what difference would it make using one estimate of perpendicular distance or the other, on the abundance estimates. The results are strikingly different in 2011 (much higher) and 2015 (much lower), although very similar in 2013. 2010 does not have data for GPS derived distances. One of the effects of changing the perpendicular distances is that the amount of observations inside the truncation distance may vary.

In this exercise, two observations from C in 2011 and two in A 2013 were removed, one observation in A 2015 was included. Additionally, the “corrected” distances derived from GPS are in some cases very different from the original ones derived from the angle. These differences can be of several thousand meters in some cases, which can produce an important effect on the detection function.

f) Use of bubble windows

All aerial surveys for marine mammals strongly encourage the use of bubble windows. On the same principles it is strongly encouraged for aerial surveys for other megafauna such as BFT. The lack of observations in the blind sector below the airplane when no bubble windows are available may produce a bias difficult to quantify, even if the observers in the front could potentially look directly on the track line. It is assumed that the probability of detection would be the same or similar to the closest distance available (next bin of frequencies), and therefore the curve of the detection function is extrapolated from the available information until distance zero, usually through the shoulder of a hazard-rate key function. But this assumption may or may not be true, as shown in the report of Cañadas and Vazquez (2016), and there is much variability among years/aircrafts in the searching pattern from below the aircraft towards further away.

It is clear that by not using bubble windows, much information is lost for the same resources, and more importantly, the most valuable information for a reliable detection function, which is the observations closer to the track lines. Therefore we strongly **recommend** that bubble windows continue to be used in future surveys.

I.2 Estimate of the additional variances associated with the current indices

I.2.1 Data availability and processing

Aerial surveys were conducted during spawning seasons in 2010, 2011, 2013 and 2015. The study areas differ from one year to another and a final four overlapped areas were identified, A-inside, C-inside, E-inside and G-inside. All study areas were surveyed each year except for G-inside in 2011, due to the lack of permit.

A primary analysis using the aerial survey data was conducted using the software DISTANCE 6.2 to estimate different parameters that will be used in the next analysis. The estimation of the additional variance and the assessment of population trends was conducted using the Bayesian framework given that it allows different facilities. Bayesian Hierarchical Modeling (BHM) is a framework that represents the underlying processes generating the data in a multi-level form and allows the estimation and inference on parameters using in addition to the observable data, the external knowledge or expert opinion (Clark & Gelfand, 2006). In the context of distance sampling, BHM provide numerous advantages concerning estimation and inference on abundance and related trends. Firstly, the abundance and trends estimation and inference are handled in the same model. In addition, this framework allows the separation of sources of variation (e.g. sampling and process error). Moreover it enables accommodating fixed and random effects implicitly (Moore & Barlow, 2011).

The proposed model consists of two components: a state process that describes the underlying biological dynamics of interest and an observation process that describes the relationship between the unobservable processes (state process) and the observed data. In our context the state model describes the dynamics of the population density (D_t), while the observation model characterizes the probability of observing n_t individuals during survey given D_t and the detection probability process specific to the distance sampling design.

1.2.2 Data analysis

The abundance estimate following (Buckland et al., 2001) is given by:

$$\hat{N}_{jt} = \hat{D}_{jt} \cdot A_{jt} \quad (\text{eqn. 1})$$

where N_{jt} is the population abundance, D_{jt} is population density, and A_{jt} is the study area of stratum j during year t . The density may be estimated as:

$$\hat{D}_{jt} = \frac{n_{jt} \cdot s_{jt} \cdot f_{jt}(0)}{2 \cdot L_{jt} \cdot g(0)} \quad (\text{eqn. 2})$$

where n_{jt} is the number of schools detected; s_{jt} is mean school size; $f_{jt}(0)$ is the evaluation at distance $y=0$ of the probability density (pdf) for detection probability; $g(0)$ is the probability of detection on the transect line if not assumed to be 1; and L_{jt} is the on-effort transect length. Table 3 shows some of the parameters estimated using DISTANCE.

Table 3. Parameters estimated with Distance.

Areas	Year			
	2010	2011	2013	2015
A-inside				
n_{jt}	8	10	10	6
L_{jt}	6277	7975	6743	4119
s_{jt}		678.1	611	825
C-inside				
n_{jt}	6	10	10	3
L_{jt}	8168	8466	2682	2658
s_{jt}	733	291	1285	1533
E-inside				
n_{jt}	29	45	20	13
L_{jt}	12621	9806	3720	4484
s_{jt}	1015	1715	361	2030
G-inside				
n_{jt}	33		12	2
L_{jt}	2900		1716	785
s_{jt}			336	600

a) Process model

The process model describes the underlying dynamics of the density taking into account the inherent spatial and temporal variability. The model describes the density as a function of the mean stratum differences (fixed intercept), a yearly trend coefficients (fixed effect) and a stochastic component (random variable) $c(j,t)$. The full density model is:

$$D_{jt} = \exp(\beta_0 + \beta_k \cdot strata_k + \beta_{K+1} \cdot t + \gamma_{jt}) \text{ for } k=1, \dots, K \quad (\text{eqn. 3})$$

$$\gamma_{jt} \sim \text{Norm}(0, \sigma_d)$$

Where β_0 is an intercept for log-density; β_k for $k=1, 2, \dots, K$ are fixed effect for binary dummy variables for strata A-inside, C-inside and E-inside; and $\beta_{(K+1)}$ the year-specific trend coefficient considered as a fixed effect. γ_{jt} is a random effect with mean zero and variance σ_d that describes process variation in year-to-year density.

b) *Observation model*

The observation model describes the relationship between the unobservable state process and the observed data. Rearranging (eqn. 2) and treating observed counts as a Poisson random variable:

$$n_{jt} \sim \text{Pois}(E[n_{jt}])$$

$$E[n_{jt}] = \frac{2 \cdot L_{jt} \cdot g(0)}{s_{jt} \cdot f_{jt}(0)} \cdot D_{jt}$$

The main inputs of the model are the transect length L_{jt} which is considered measured without error, mean school sizes s_{jt} and $f_{jt}(0)$ which are estimated based on the perpendicular distances using the software Distance 6.2. The correction factor $g(0)$ was estimated independently based on the tagging data that records the percentage of time spent in a depth band (ICCAT report Cañadas and Vazquez 2016).

c) *Mean school size*

To account for the inter-annual variability in spatial and temporal distribution of spawning schools and school size, raw data representing the on-effort observations of schools and their sizes were used. The range of available data cover all survey years 2010, 2011, 2013 and 2015 and all overlapped areas A-inside, C-inside, E-inside and G-inside. The exception was for 2010 in areas A-inside and G-inside and in 2011 for G-inside.

The observed school sizes were assumed having a Negative-Binomial distribution with a time varying over-dispersion parameter. As in the model of density we consider covariates such as strata as fixed effects, and year as fixed trend effect. The likelihood of the observed school sizes ($S. obs$) is given by:

$$S. obs_{jt} \sim \text{negBin}(\lambda_{jt}, r_t)$$

Where r_t is the time-dependent random over-dispersion parameter with mean \bar{r} and variance σ_r^2 .

The process model of school size is given by:

$$\lambda_{jt} = \exp(\beta_{s0} + \beta_{sk} \cdot Strata_k + \beta_{s,K+1} \cdot t + \alpha_{jt})$$

$$\alpha_{jt} \sim \text{Norm}(0, \sigma_s)$$

where β_{s0} is an intercept, β_{sk} for $k=1, 2, 3$ are fixed effect coefficients for the dummy variables representing stratum A-inside, C-inside and E-inside, $\beta_{s,K+1}$ is a fixed effect trend coefficient and α_{jt} is a normal random variable for each stratum-year with mean 0 and variance σ_s^2 .

d) *Parameter estimation*

Parameter estimation was implemented using a Bayesian MCMC approach in JAGS 3.4.0 (Plummer, 2003, 2012) that was called from the statistical software R 3.1.1 using the package 'R2jags' 0.5-7. The prior distributions of all parameters were informative normal and uniform distributions except for $g(0)$ that was chosen so that it has a mean of 0.47 (CV=52%). The $g(0)$ distribution parameters were estimated based on the diving data collected from the electronic tags used by ICCAT. The parameters estimation was held using 2 chains each of length 100,000 with a thinning interval of 50.

30,000 iterations were run as a burning period to ensure the convergence of the Markov chains of all parameters.

e) $g(0)$ estimation

The Electronic tagging data provided by ICCAT contains the information about the percentage of time spent by BFT in depth bands that goes down from the surface to 10m maximum. The estimation of trackline detection probability, $g(0)$, was based on data collected from the overlapping areas A, E, C during 2011, 2013 and 2015 in day time. The filtered data kept 13 tagged fishes and 154 data points. For the purpose of this analysis the estimation of $g(0)$ was conducted independently of area and year. Table 4 shows a summary of the data used in the analyses.

Cañadas and Vazquez (2016) have estimated $g(0)$ by area and year. The estimation shows that it varies between a 0.3 and 0.7. Based on the data available, an overall estimate of $g(0)$ was 0.47 (CV=52%). A bounded beta prior between 0.3 and 0.7 was used in the analysis to avoid unrealistic estimates of abundance that could be obtained when $g(0)$ takes values near zero.

Table 4. Number of data points per area and year

NB - IDArgos		Overlapping areas			
Years	Month	A	C	E	Total
2012	June			10	10
	June	42	9	41	92
2013	July	4			4
2015	June	13	8	27	48
Total		59	17	78	154

f) Model selection

Model selection was conducted using the Deviance Information Criterion (DIC) (Spiegelhalter *et al.*, 2002). This criterion is defined as $DIC = \bar{D} + pD$ where \bar{D} is the posterior mean model deviance and pD is termed the effective number of parameters which account for the complexity of the model. The model selection was held without incorporating the process error because almost all models behave equally well when it is present.

1.2.3 Results

a) Mean school size model

Five school size models were analysed. Table 5 shows the DIC values. Based on the criteria it is clear that model *S1* that has the lowest DIC is the model that fits the data well. The convergence of the Markov chains to a stationary state was verified using the Gelman-Rubin diagnostic (Gelman & Rubin 1992). Parameter estimation for the school size model was held using the model *S1*. Table 6 shows the estimates of the mean school size using the model.

Table 5. DIC values for the different models of school size

ID	Model	\bar{D}	pD	DIC	ΔDIC
S1	$\lambda_{jt} = \exp(\beta_0 + \beta_t + \beta_k \cdot starat_k)$	2591.11	7.33	2598.44	0
S2	$\lambda_{jt} = \exp(\beta_0 + \beta_1 \cdot t + \beta_k \cdot starat_k)$	2593.47	8.41	2601.84	3.39
S3	$\lambda_{jt} = \exp(\beta_0 + \beta_1 \cdot \log(t) + \beta_k \cdot starat_k)$	2593.06	7.85	2600.91	2.47
S4	$\lambda_{jt} = \exp(\beta_0 + (\beta_1 + \beta_j) \cdot t + \beta_k \cdot starat_k)$	2593.29	8.57	2601.86	3.41
S5	$\lambda_{jt} = \exp(\beta_0 + (\beta_1 + \beta_j) \cdot \log(t) + \beta_k \cdot starat_k)$	2586.04	14.52	2600.56	2.12

Table 6. Mean school size estimates (CV between ()) based on model S1.

Year	A-inside	C-inside	E-inside	G-inside
2010	-	1077 (73 %)	1031 (35 %)	-
2011	937 (46 %)	861 (73 %)	1664 (27 %)	-
2013	703 (55 %)	1232 (49)	670 (48 %)	396 (53 %)
2015	960 (71 %)	1785 (125 %)	1905 (60 %)	784 (120 %)

b) Density model

The estimation of the additional variance is conducted using a Bayesian hierarchical model. The data used are estimated using the aerial survey data gathered by ICCAT. Table 3 shows the main input used in the analysis. The mean school sizes are corrected using the model in [1.2.2) c)]. Table 6 shows the estimated mean value of the mean school sizes used in the model.

Six different models were fitted to data and a final model was selected based on the deviance information criteria. The selected model incorporate stratum effect and a yearly linear trend. The strong relation between the fitted and the observed number of schools and the absence of correlation between the residuals and the fitted number of schools suggest a good fit to the data (Figure 1).

Table 7. DIC of the different models of density

ID	Model	\bar{D}	pD	DIC	ΔDIC
D1	$D_{jt} = \exp(\beta_0 + \beta_k \cdot starat_k)$	474.03	18.46	492.50	2.96
D2	$D_{jt} = \exp(\beta_0 + \beta_t + \beta_k \cdot starat_k)$	473.94	16.73	490.67	1.14
D3	$D_{jt} = \exp(\beta_0 + \beta_1 \cdot T + \beta_k \cdot starat_k)$	473.97	16.87	490.84	1.31
D4	$D_{jt} = \exp(\beta_0 + \beta_1 \cdot \log(T) + \beta_k \cdot starat_k)$	473.69	15.83	489.53	0
D5	$D_{jt} = \exp(\beta_0 + (\beta_1 + \beta_j) \cdot T + \beta_k \cdot starat_k)$	473.90	17.88	491.78	2.25
D6	$D_{jt} = \exp(\beta_0 + (\beta_1 + \beta_j) \cdot \log(T) + \beta_k \cdot starat_k)$	473.72	16.50	490.22	0.68

1.2.4 Discussion

The analysis followed above allows the quantification of the process error which could be used for the prediction of abundance. Another advantage of the hierarchical structure of the model is that it has shown flexible ability to accommodate fixed and random effects. The lack of calibration of the school size is a key question in the estimation of abundance. Such correction might be held using parallel external data like aerial photos or sonar images to calibrate the abundance estimates, but all available methodologies have currently limits and biases. The quantification of the track-line probability of detection induce an underestimation of the abundance given that the used $g(0)$ estimate account only for availability bias. The posterior mean of the additional variance estimated based on the density model is 0.8 (CV=28%). This high value of the process error could be the result of the estimation based on the pooled data of all areas. The estimation of the additional variance by area has been tested but it yields relatively low additional variance but with very high CV. For this reason it was decided to use the first one. Table 8 shows a comparison of the estimates of the total abundance by area and year and their CV using the model of density and DISTANCE. It is remarkable that the estimates using the model have a reduced CV compared to those of DISTANCE.

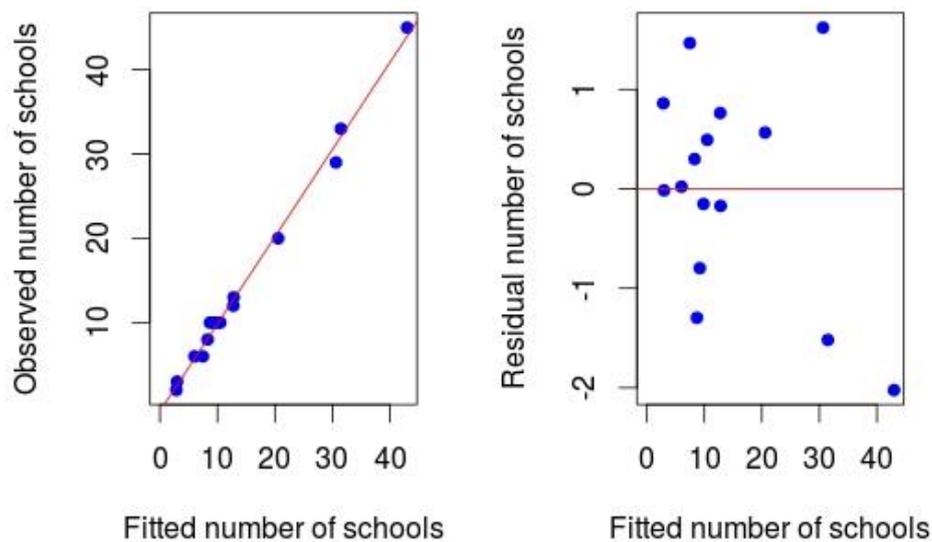


Figure 1. Observed vs fitted number of schools (left) and residual vs fitted number of schools (right)

Table 8. Abundance estimates by year and area using the Density model and DISTANCE.

Area	Year	N (Model)	CV % (Model)	N (DISTANCE)	CV % (DISTANCE)
A-inside	2010	31636	43		
	2011	52542	40	38720	45.5
	2013	24069	39	18717	44.5
	2015	30513	46	38248	44.7
C-inside	2010	19142	44	9797	59.2
	2011	42180	40	13614	45.3
	2013	76426	42	86114	38.3
	2015	38253	57	47900	65.5
E-inside	2010	84130	31	73676	37.8
	2011	535824	29	541634	32.2
	2013	123552	34	60614	75.4
	2015	180266	37	283100	64.1
G-inside	2010	233273	32		
	2011				
	2013	58870	37	44041	54.8
	2015	56432	59	44162	95.9

I.3 Estimate the use of electronic tagging or other data to evaluate spatial and vertical differences between spawning seasons and provide an estimate of additional variance independently of I.2.

I.3.1 Data availability and processing

The evaluation of the spatial and vertical differences between spawning seasons was assessed based on data from electronic tags of 36 fish including 320 observations (Table 9). The data filtered according to overlapping areas indicate the percentage of time, in periods of 6 hours in average, that the observed fish spent in different depth bands (0(0 m), 2 (0.1-2m) and 10 (2.1-10m)), during the observed time period.

Table 9. Number of observations by area and year

Years	NB - IDArgos					Total
	A	C	E	G	Overlapping areas	
2011	4					4
2012		42	14			14
2013	62	12	55			129
2015	26	16	52	79		173
Total	92	28	121	79		320

A graphical exploration was held to show differences of time spent within 0-10m depth band according to area, year and day time (Figure 2). Generally there is a large dispersion of the time spent within depth bands according to day and night. Figure 3 shows the histograms of time spent within depth bands. The data are skewed to the left with important number of values near zero. The exception is for depth (10m) where data tend to be symmetric.

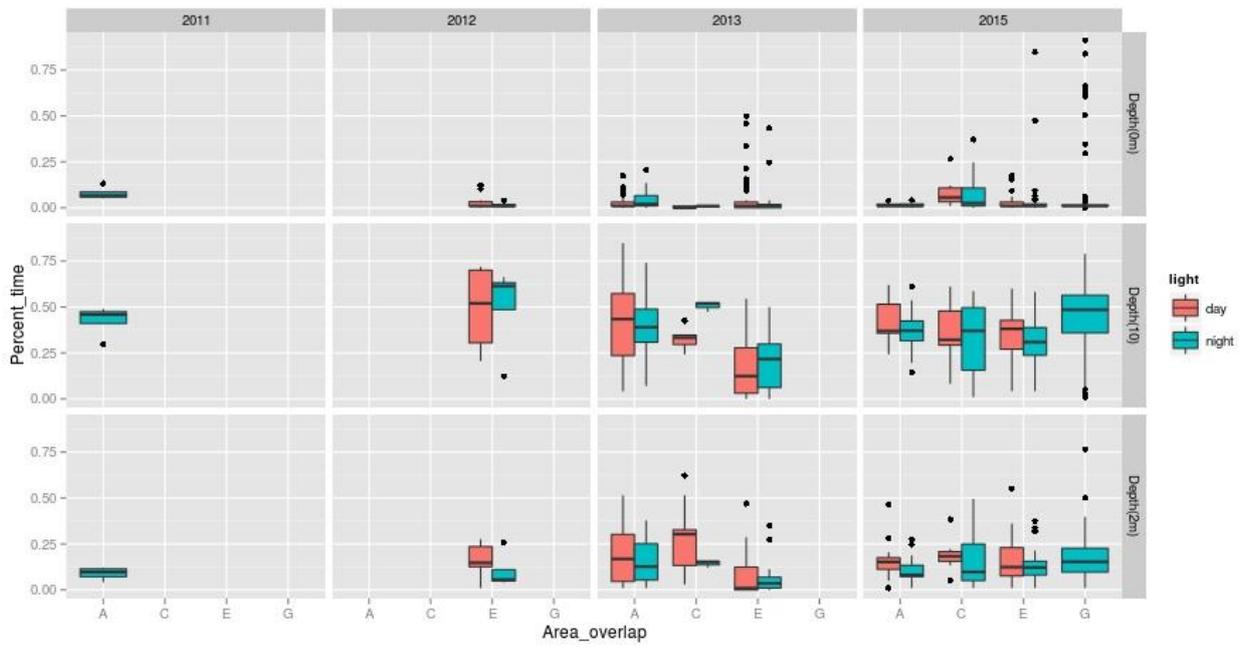


Figure 2. Time spent within each depth band by year area

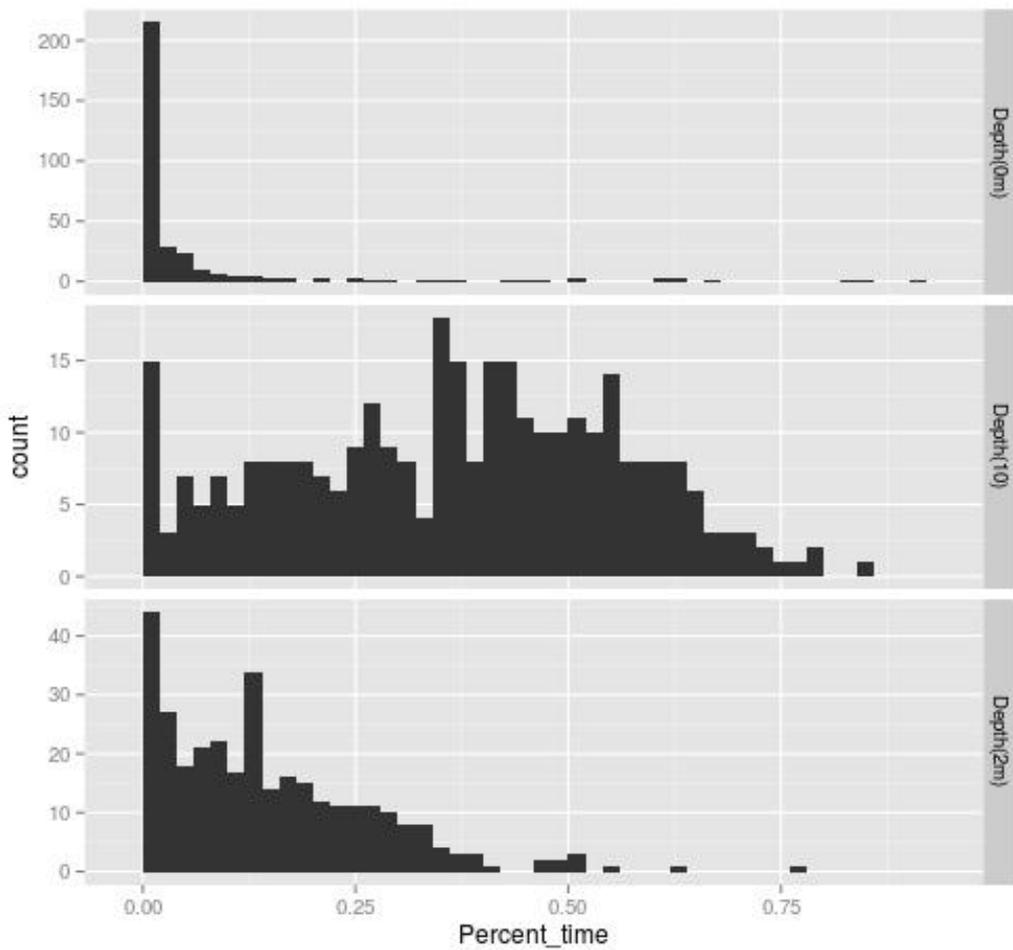


Figure 3. Histogram of the percent time spent within a depth band

1.3.2 Data analysis

The collected data represent the percentage of time spent by a fish in a depth band. The particularity of this continuous data is that it lays between zero and one. Such data could be modelled using a suitable distribution such as the beta distribution. This distribution is extremely flexible and one can model its parameters as functions of explanatory variables which is the purpose of the present exercise. However the data contains zero proportion of time spent within some depth bands. This could be critical for the beta distribution because it is defined on the open interval (0,1). Another alternative is to combine a mixture of models to allow for the zero-one inflation (Ospina and Ferrari, 2010, 2012)

The beta zero-inflated (BEZI) distribution is given by :

$$\begin{aligned} bezi(y; \alpha, \mu, \phi) &= \alpha && \text{if } y = 0 \\ bezi(y; \alpha, \mu, \phi) &= (1 - \alpha)f(y; \mu, \phi) && \text{if } y \in (0,1) \\ bezi(y; \mu, \phi) &= \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1} && \text{if } y \in (0,1) \end{aligned}$$

where $\Gamma(\cdot)$ is the gamma function, $0 < \alpha < 1$, $0 < \mu < 1$ and $\phi > 0$

$$P(y=0)=\alpha, E(y) = (1 - \alpha)\mu, Var(y) = (1 - \alpha)\frac{V(\mu)}{\phi+1} + \alpha(1 - \alpha)\mu^2$$

The two components μ and α were modelled using covariates such as strata, day light and depth band as fixed effects and a random normal variable that accounts for inter-annual variability as random effect.

$$\text{logit}(\alpha) = \gamma_0 + \gamma_1 * \text{strata} + \gamma_2 * \text{depth} + \gamma_3 * \text{dayLight} \quad (\text{eqn4})$$

$$\text{logit}(\mu) = \beta_0 + \beta_1 * \text{strata} + \beta_2 * \text{depth} + \beta_3 * \text{dayLight} + \epsilon \quad (\text{eqn5})$$

where eqn4 is the process component for the discrete model that models the presence / absence of zero percent time spent in depth bands, and eqn5 is the model for the mean of the Beta distribution with the random effect ϵ representing the process error with mean 0 and variance σ .

The modelling process was held using the Bayesian framework with JAGS via the statistical software R using R2jags.

1.3.3 Results

Modelling the percentage of time spent within a depth band could be used as an estimation for availability of BFT for the aerial survey. With a representative sample of BFT and good geographic coverage it is possible to build an environmental model to predict the availability of BFT when aerial surveys were being conducted (Scott-Hayward *et al*, 2014). Table 10 and Table 11 show the estimates of the different parameters of the discrete model for zero data and the continuous model for the data in (0,1).

The process error is estimated to have a standard deviation of 0.282 (CV=57 %).

Table 10. Parameters for the discrete model

Parameter	Mean	SD
b.v.0.zero	-5.196	0.604
b.v.strata.zero[1]	0.651	0.644
b.v.strata.zero[2]	1.380	0.704
b.v.strata.zero[3]	2.256	0.587
b.v.light.zero	0.560	0.293
b.v.depth.zero[1]	1.837	0.338
b.v.depth.zero[2]	0.243	0.413

Table 11. Parameters for the continuous model

Parameter	Mean	SD
b.v.0	-0.310	0.292
b.v.strata[1]	-0.134	0.365
b.v.strata[2]	-0.067	0.373
b.v.strata[3]	-0.308	0.358
b.v.light	-1.024	0.063
b.v.depth[1]	-2.011	0.078
b.v.depth[2]	-1.115	0.067
Sig.v.jt	0.284	0.162
tau	6.513	0.326

I.4 Compare the additional variance estimates with the estimates of survey variance obtained using DISTANCE

The accuracy of estimates and the reduction of variability could always be attained by increasing enough the sample size. Taking into account the cost of gathering data, it is always desirable to determine the minimal sample size or the minimal sampling duration that is required to achieve a desired accuracy (Kruschke, 2011).

To conduct the question of comparing the additional variance estimates with the estimates of DISTANCE, multiple scenarios on the coverage percentage were analysed. The analysis of the effect of increasing or reducing the total transects length or the percentage of the surveyed area is measured by the improvement and the deterioration of the coefficients of variation of the estimates. For this exercise, six scenarios (Table 12) were analysed where the total range of areas coverage varies from 10% to 60% (as in Cañadas and Vazquez, 2016). For each scenario, the total length and the number of schools detected were calculated based on the available data according to each scenario.

Table 12. Transect length by area and year corresponding to six area coverage scenarios

Area	Year	Survey area (km ²)	Real coverage %	Real transect length	10 %	20 %	30 %	40 %	50 %	60 %
A inside	2010	61933	30	6277	2092	4185	6277	8369	10462	12554
C inside	2010	53868	44.9	8168	1819	3638	5457	7277	9096	10915
E inside	2010	93614	40.0	12621	3155	6311	9466	12621	15776	18932
G inside	2010	56211	15.3	2900	1895	3791	5686	7582	9477	11373
A inside	2011	61933	17.5	7975	4557	9114	13671	18229	22786	27343
C inside	2011	53868	21.4	8466	3956	7912	11868	15824	19780	23736
E inside	2011	93614	14.2	9806	6906	13811	20717	27623	34528	41434
G inside	2011	56211								
A inside	2013	61933	32.6	6743	2068	4137	6205	8274	10342	12410
C inside	2013	53868	14.9	2682	1800	3600	5400	7200	9000	10800
E inside	2013	93614	11.9	3720	3126	6252	9378	12504	15630	18756
G inside	2013	56211	9.2	1716	1865	3730	5596	7461	9326	11191
A inside	2015	61933	20.2	4119	2039	4078	6117	8156	10196	12235
C inside	2015	53868	15.0	2658	1772	3544	5316	7088	8860	10632
E inside	2015	93614	14.5	4484	3092	6185	9277	12370	15462	18554
G inside	2015	56211	4.2	785	1869	3738	5607	7476	9345	11214

Based on the data calculated according to each scenario, the model of density presented in I.2 was fitted to each of them and the abundances by area and year and their CVs were estimated. Table 13 shows the main results of the six scenarios. Figure 4 presents the trend of the CVs of abundance. It is clear that above a certain level of area coverage the gain in terms of variability reduction becomes unimportant compared to the cost of coverage. For example, when the total area coverage increases from 10 % to 20 % the average reduction of the abundances CVs is equals to 19.8 %. The average reduction of the CVs when increasing the total coverage from 20 % to 30 % in all areas in 9.66 %. Above 40 % total coverage of all areas, the CVs reduction decrease slowly from 5.35 % for 40 % coverage and 3.8 % for 60% coverage.

Table 13. Abundance estimates and their CVs by year and area.

Area	Year	Actual		N1	CV1 %	N2	CV2 %	N3	CV3 %	N4	CV4 %	N5	CV5 %	N6	CV6 %
		Actual N	CV %												
A-inside	2010			37294	57.3	30946	48.4	31777	43.1	32378	39.5	30733	37.5	31190	36.4
C-inside	2010	9797	59.2	26173	66.4	24599	52.3	20742	48.5	18702	46	19527	41.8	18224	40.5
E-inside	2010	73676	37.8	94742	42.3	91487	35	86472	32.3	84121	30.7	83046	30.1	83419	28.9
G-inside	2010			231058	34.9	233067	30.9	238697	29.6	239331	29.1	241146	28.6	239911	28
A-inside	2011	38720	45.5	52821	47.5	50754	39.6	53746	35	55502	32.8	56148	31.2	55064	30.1
C-inside	2011	13614	45.3	45366	48.3	41525	41	42592	35.9	43341	33.7	42250	32.2	42571	30.9
E-inside	2011	541634	32.3	531352	31.2	537071	27.9	546944	26.6	552979	26.5	553171	26.1	553053	25.6
G-inside	2011														
A-inside	2013	18717	44.5	25650	54.2	24153	44.6	23691	39.9	23460	37.3	23336	35.5	23093	33.6
C-inside	2013	86114	38.3	75554	47.4	76236	38	81234	33.9	83824	31.8	85752	30.5	84066	29.2
E-inside	2013	60614	75.4	125120	34.1	123331	30	120024	28.3	120401	27.7	120646	27.2	120280	26.5
G-inside	2013	44041	54.8	57813	36.4	56385	31.2	55878	29.4	55944	28.5	55795	27.9	55553	27.1
A-inside	2015	38248	44.7	30751	57.7	30501	46	30665	40.9	30698	37.7	30675	35.4	30691	34
C-inside	2015	47900	65.5	38634	65.6	38114	51.9	38069	45.9	38070	42.8	38396	39.5	38017	37.5
E-inside	2015	283100	64.1	178004	41.1	180242	34.2	181158	31.4	181494	30.1	182489	29	181443	28.3
G-inside	2015	44162	95.9	48667	46.7	45702	38.9	41675	35.6	41489	33.5	41767	32	41484	30.4

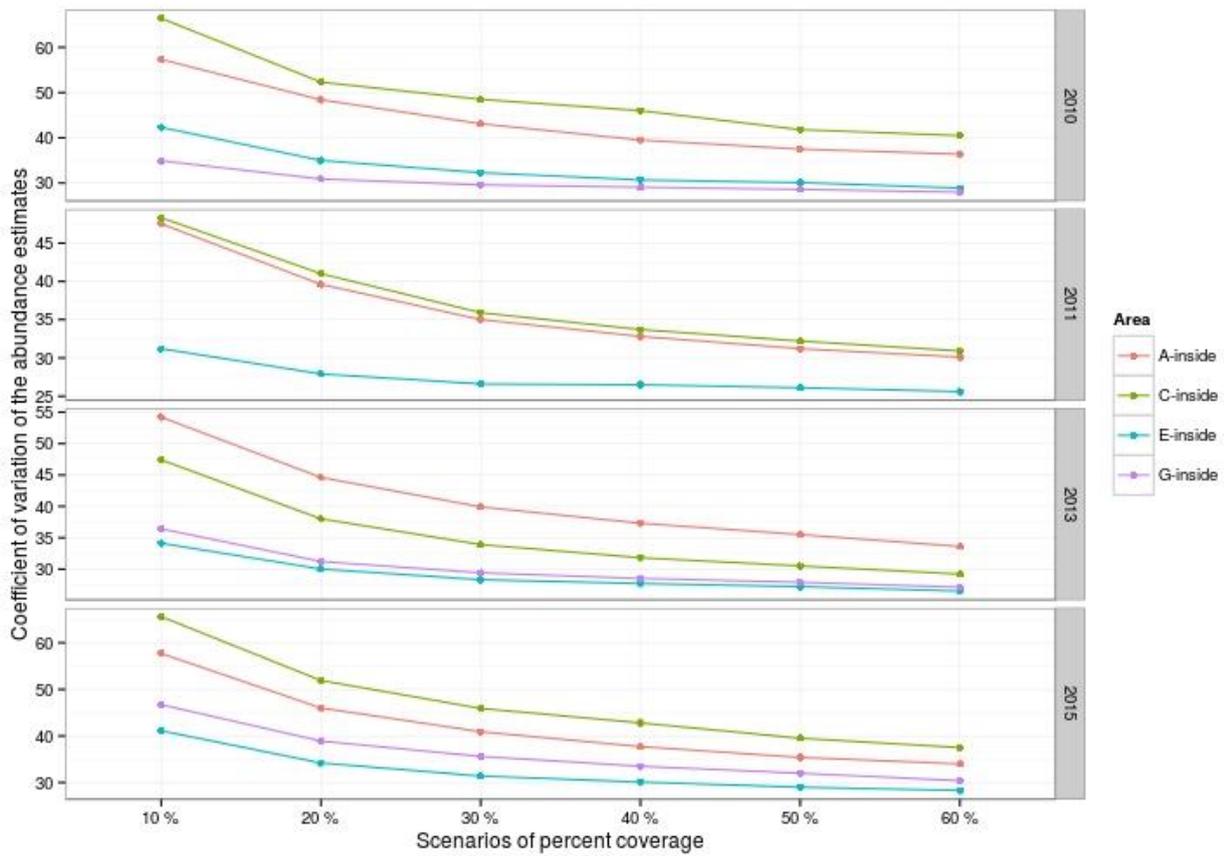


Figure 4. Trend of the CVs of abundance under various scenarios of area coverage.

Based on the coverage scenarios in table 12 and table 13, a comparison of the different strategies of surveying the overlapping areas was conducted. Figure 5 shows the comparison of average CVs of abundance of the four areas together for each coverage scenario based on 2010-2011, where the survey was held only in the overlapped areas, and 2013-2015, where 50 % of the survey was held outside the overlapped areas. Also the average based on the whole 2010-2015 period is shown. The figure shows that concentrating the survey in the overlapped areas (2010-2011) allows a gain in terms of reduction of abundance CVs within these overlap areas, when compared with years when extended survey is done (2013-2015).

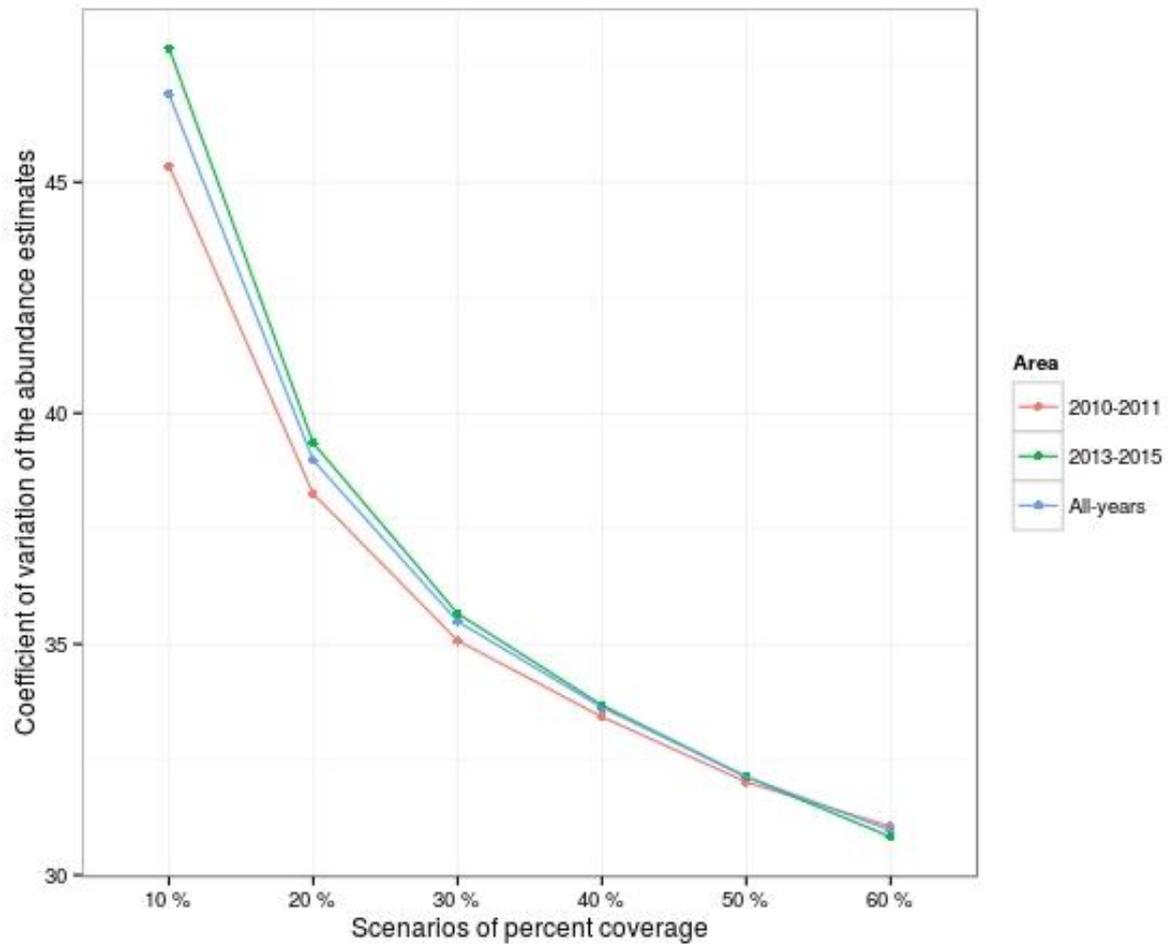


Figure 5. Trend of the CVs of abundance under various scenarios of area coverage in the overlap areas, based on the CVs of 2010-2011 (surveying only the inside areas), 2013-2015 (surveying both inside and outside areas), and the whole period.

II. Evaluation whether the aerial survey can provide a reliable and robust index of abundance for use in stock assessments

All current methodologies for obtaining any index of abundance for large pelagic species provide estimates which cannot be checked against a real reference number or ground truth. Both fishery-dependent indices (i.e. CPUE by gear and area) and fishery-independent indices (i.e. aerial surveys, larval index, sonar surveys) are always not precise. The advance of an aerial survey is that it produces estimates which are always prudential, because the number of sightings can be just a partial image of the reality, but it is almost impossible that they can be higher than the reality. As a matter of fact, this index is always conservative and prudential.

II.1 Analysis of the design and costs to detect a range of population growth rates in 3, 5 and 10 years with no reduction in the additional variance

The assessment of the reliability of the aerial survey index of abundance to reflect the underlying dynamics is conducted using a power analysis. Based on the model selected in I.2) above, Monte Carlo simulation were conducted to evaluate its ability to detect a known trend in different ranges of years. All parameters were given the posterior mean values estimated before, except for the trend parameter which is fixed according to each scenario. This parameter was given different values that represents increase and decrease in the density of BFT with magnitude of 20% and 10%. The annual changes were tested over 3, 5 and 10 years.

To test each combination of annual change and period, 100 random data sets were generated according to the following scheme:

$$n_{jt} \sim \text{Pois}(E[n_{jt}])$$
$$E[n_{jt}] = \frac{2 \cdot L_{jt} \cdot g(0)}{s_{jt} \cdot f_{jt}(0)} \cdot D_{jt}$$

$$D_{jt} = \exp(\beta_0 + \beta_1 \cdot t + \beta_k \cdot \text{starat}_k)$$

To estimate the power of the model to detect trends, each generated data set was analysed with the model selected in I.2. The trend is considered detected if the 95% confidence interval around the posterior mean value of the trend parameter fall completely in a region of practical equivalence (ROPE) of the fixed trend. It is obvious that the wider the ROPE the more the power will be improved. The ROPE tested were intervals around the known trend that have a width calculated based on a percentage of the trend value. For example for a trend θ , two ROPE will be tested that are $\theta \pm 0.1 \cdot \theta$, and $\theta \pm 0.2 \cdot \theta$.

For each combination of trend magnitude, number of year sampling and ROPE width, a fraction of the simulations do not detect the trend. This fraction represent the β - error (Type II error). The fraction of simulations that yields a confidence interval around the mean falling completely in the ROPE, represents the power.

a) Scenario 1: Data simulation without process error

In this scenario the data sets simulated to detect trends were held without using the process error. So the density was calculated deterministically as a function of the estimated parameters and only the number of schools generated using 'rpois' that adds some randomness to data. Table 14 shows the results of the trend detection. For the largest trend 20%, there is a probability above 0.6 to be detected within the range [0.16, 0.24].

Table 14. Power to detect a trend

No. of years	% annual change	Power to detect a trend	
		$\beta \pm 0.10\beta$	$\beta \pm 0.2\beta$
3	+20	0.10	0.29
5	+20	0.21	0.43
10	+20	0.36	0.68
3	-20	0.06	0.2
5	-20	0.17	0.33
10	-20	0.29	0.63
3	+10	0.02	0.09
5	+10	0.07	0.2
10	+10	0.12	0.37
3	-10	0.05	0.05
5	-10	0.08	0.12
10	-10	0.1	0.36

b) Scenario 2: Data simulation with process error

This scenario has the data generated by adding a normal random variable with mean zero and standard deviation equal to the additional variance. In this scenario we have tested large annual change equal to 25% and a wider ROPE with a half width equal the 20% of the trend. Results has shown that in the presence of the additional variance it becomes difficult to detect even a large annual change. Table 15 presents the results of the power to detect a 25% annual change.

Table 15. Power to detect a 25% annual change

No. of years	% annual change	Power to detect a trend
		$\beta \pm 0.2\beta$
10	+25	0.223
5	+25	0.087
3	+25	0.051
10	-25	0.18
5	-25	0.042
3	-25	0.032

c) Cost analysis:

The analysis of the cost of the aerial surveys designed to detect a range of growth rates in different range of years was based on the cost of one survey. The estimation of the total costs of the aerial surveys necessary to detect a trend could be calculated as the product of the cost of an actual survey and the number of years necessary to detect such a trend.

See cost analysis in II.2.

Discussion

The number of years to detect trends in the BFT population using the described methodology depends mainly on two parameters: 1) the magnitude of annual change to be detected and 2) the width of the ROPE which could be interpreted as the accuracy with which we want to detect the trend. A survey stability (meaning a survey conducted with the same methodology, always in the same areas, in the same periods and with the same spotters) will certainly and substantially improve the current estimates, which are necessarily based on the data from different survey approaches. Detecting an annual trend of 50% will be more likely than a 25% annual change. Similarly, detecting a trend within an interval of [20%, 30%] will be more probable than detecting a trend within [23%, 27%]. If a method requires a large increase or decline to be detected in large number of years, then the followed methodology will be of limited use (Forney *et al.*, 1991). The large estimate of the additional variance (80% (CV=28%)) is the main source that induced the weak power to detect the trends. The analysis confirms that a 25% annual change takes 10 years to be detected with a weak probability equal 0.2.

II.2 Analysis of the design and costs to detect a range of population growth rates in 3, 5 and 10 years with a reduction in additional variance

The reduction of the additional variance and the improvement of the power to detect a trend could be handled by increasing the sample size. In the next exercise multiple data sets were created and analyzed based on the model in I.2.

To increase the sample size and therefore reduce the additional variance, three scenarios specifying the total coverage of each overlapping area were analyzed. Table 16 shows the total transect length corresponding to 20%, 40% and 60% total coverage of each area. Three cases are shown, as the length of transects for a given coverage depends on the effective strip width (*esw*) considered for the detection function. The three cases considered of twice the *esw* (as the *esw* only counts one side of the track, so it needs to be multiplied by two to take into account both sides of the track) were:

- Esw x2 = 1.94 (being the *esw* of the last year, 2015)
- Esw x2 = 2.30 (being the average *esw* for the four years)
- Esw x2 = 2.96 (being the highest *esw* of the four years)

Table 17 shows the results of the power estimations. It is clear that by increasing the total length of transect the power increases also.

Table 16. Total transect according to each scenario of total coverage by area.

Area	esw x2=1.94			esw x2=2.30			esw x2=2.96		
	20 %	40 %	60 %	20 %	40 %	60 %	20 %	40 %	60 %
A-inside	6,385	12,770	19,155	5,385	10,771	16,156	4,185	8,369	12,554
C-inside	5,553	11,107	16,660	4,684	9,368	14,053	3,640	7,279	10,919
G-inside	9,651	19,302	28,953	8,140	16,281	24,421	6,325	12,651	18,976
E-inside	5,795	11,590	17,385	4,888	9,776	14,664	3,798	7,596	11,394
Total	27,384	54,768	82,152	23,098	46,196	69,294	17,948	35,895	53,843

Table 17. Power to detect a trend with increased transect length (considering here $esw \times 2 = 2.96$)

No. of years	% annual change	$\beta + 20\% \beta$		
		20 %	40 %	60 %
10	+20	0.13	0.145	0.20
5	+20	0.09	0.12	0.137
3	+20	0.02	0.06	0.07
10	-20	0.17	0.18	0.19
5	-20	0.06	0.09	0.11
3	-20	0.0	0.04	0.05

The results of the variance reduction shows that even with a large coverage of the different areas, the probability to detect a 20% annual change equal 0.2, is relatively weak. The major reason of this inability to detect a trend with sufficient confidence is due mainly to the process error which in our case equals 80% (CV=28%).

Cost analysis:

The available cost data provided by ICCAT cover the four aerial surveys conducted in the four phases 1, 2, 4 and 5. Table 18 presents the total transect length and Table 19 shows the total costs by year and the average cost by km transect on effort. Costs have been differentiated into field work cost and other costs (design, training and meetings, and analysis). Field costs depend on the transect length and the logistics, while ‘other costs’ are mostly independent of the coverage.

Table 18. Total transect length (km) in each year of survey

Year	Inside	Outside	Total
2010	30,879	-	30,879
2011	28,177	-	28,177
2013	15,669	13,278	28,947
2015	14,413	11,079	25,492
Total	89,138	24,357	113,495

Table 19. Total cost by phase and average cost by km of transect

Year	Cost (€)			Transect length on effort (km)			Cost per km		
	Field work	Other	Total	Outside	Inside	Total	Field work	Other	Total
2010	300,000	13,266	313,266		30,879	30,879	9.72	0.43	10.14
2011	282,097	34,236	316,333		28,177	28,177	10.01	1.22	11.23
2013	491,718	26,708	518,426	13,278	15,669	28,947	16.99	0.92	17.91
2015	431,425	48,002	479,427	11,079	14,413	25,492	16.92	1.88	18.81

As the costs have varied from year to year, a “typical” cost was chosen for the cost analysis. For ‘other costs’ the average cost for the four years of survey was taken. For field work, we considered only 2010-2011, as in 2013-2015 the outside areas were also surveyed increasing very much the costs associated with field work logistics. Therefore, the average of the field work cost of 2010 and 2011 was calculated, and prudentially a 10% was added to account for potential increases in costs of aircraft renting, personnel, fuel, and other associated costs. The resulting “typical” cost used for the cost analysis is shown in Table 20.

Table 20. “Typical” cost to be used in the cost analysis.

	Cost (€)		
	Field work	Other	Total
Total cost	320,153	30,553	350,706
Cost per km	10.85	1.11	11.96

Table 21 shows the cost analysis for five scenarios of coverage (20% to 60%) and the three cases of *esw*, as the amount of transect length is what would define the cost. The “typical” cost was applied. The reduction in CVs of abundance and with additional variance 1 were taken for the “Average Scenario” (see Cañadas and Vazquez, 2016).

Table 21. Cost analysis for one survey over the four overlap areas together (265,626 km²)

esw x2 (km)	% cov.	Transect length on effort (km)	Cost (€)			CV abundance				CV Final 1 (with additional variance from tagging)			
			Field work	Other	Total	A	C	E	G	A	C	E	G
1.94	20	27,384	297,113	30,467	327,580	0.56	0.58	0.49	0.60	0.63	0.64	0.57	0.66
	30	41,076	445,669	45,701	491,370	0.48	0.50	0.47	0.58	0.55	0.57	0.55	0.64
	40	54,768	594,226	60,934	655,160	0.43	0.46	0.46	0.57	0.52	0.54	0.54	0.63
	50	68,460	742,782	76,168	818,950	0.40	0.43	0.45	0.56	0.49	0.51	0.53	0.63
	60	82,152	891,339	91,401	982,740	0.38	0.41	0.45	0.56	0.47	0.50	0.53	0.62
2.30	20	23,098	250,608	25,698	276,307	0.56	0.58	0.49	0.60	0.63	0.64	0.57	0.66
	30	34,647	375,912	38,547	414,460	0.48	0.50	0.47	0.58	0.55	0.57	0.55	0.64
	40	46,196	501,217	51,397	552,613	0.43	0.46	0.46	0.57	0.52	0.54	0.54	0.63
	50	57,745	626,521	64,246	690,766	0.40	0.43	0.45	0.56	0.49	0.51	0.53	0.63
	60	69,294	751,825	77,095	828,920	0.38	0.41	0.45	0.56	0.47	0.50	0.53	0.62
2.96	20	17,948	194,729	19,968	214,698	0.56	0.58	0.49	0.60	0.63	0.64	0.57	0.66
	30	26,922	292,094	29,952	322,047	0.48	0.50	0.47	0.58	0.55	0.57	0.55	0.64
	40	35,895	389,459	39,937	429,395	0.43	0.46	0.46	0.57	0.52	0.54	0.54	0.63
	50	44,869	486,824	49,921	536,744	0.40	0.43	0.45	0.56	0.49	0.51	0.53	0.63
	60	53,843	584,188	59,905	644,093	0.38	0.41	0.45	0.56	0.47	0.50	0.53	0.62

When comparing the estimated “typical” cost for an ICCAT aerial survey for bluefin tuna, with other aerial surveys, it comes out cheaper. Table 22 shows a comparison with the costs of an aerial survey for marine life in the Adriatic Sea in 2013 (C. Fortuna, pers. comm.) with one aircraft for a total of 7,900 km on effort; with the budget for the planned aerial survey SCANS-III in 2016 (P. Hammond, pers. comm) in the NE Atlantic with 5 aircrafts and a coverage of 4.6% (total effort of 46,250 km): and with the budget for the next aerial survey in Italian waters and Pelagos in 2016 (S. Panigada, pers. comm) for 7,400 km on effort planned.

Table 22. Cost of other aerial surveys.

Survey	Cost per km (€)		
	Field work	Other	Total
“Typical” cost ICCAT	10.85	1.11	11.96
SCANS-III (2016)	11.35	2.05	13.41
Italy (2016)			15.05
Adriatic Sea (2013)	12.28	0.91	13.19

Additionally, a comparison has been made with other ICCAT techniques carried out during the past years. Even if the different techniques yield very different and complementary information, it at least puts into perspective the overall cost of the aerial surveys within the research of ICCAT. Table 23 shows a comparison for the 5 phases. The costs per year for the aerial surveys are overall more or less similar than those for the biosampling; being the survey much cheaper when done only within the inside areas (2011) and more expensive when surveying the extended areas (when the costs increases due to operational and logistical issues), than the biosampling. The tagging, however, is much more expensive (up to 2 or 3 times) every year than the aerial surveys.

Table 23. Total cost (€) by phase for the different techniques.

Year	Aerial survey			Tagging			Biosampling		
	Field work	Other	Total	Field work	Other	Total	Field work	Other	Total
2010	300,000	13,266	313,266	0	36,604	36,604			
2011	282,097	34,236	316,333	568,376	350,569	918,945	440,406	10,082	450,488
2012				1,080,000	40,157	1,120,157	396,000	0	396,000
2013	491,718	26,708	518,426	943,963	172,710	1,116,673	350,000	0	350,000
2015	431,425	48,002	479,427	245,701	394,877	640,578	342,496	0	342,496
Total	1,505,240	122,212	1,627,452	2,838,040	994,916	3,832,957	1,528,902	10,082	1,538,984

II.3 Previous power analysis

Previous power analyses (2010 and 2013) were carried out under a different situation, at first because these analyses were based on the four main spawning areas only, the strategy at that time. The results of these two previous power analyses were very much promising in terms of opportunities for getting a SSB trend within the life span of GBYP. The situation changed after the second analysis, because the strategy

requested by the GBYP Steering Committee was to carry out extensive surveys in all available Mediterranean areas where Bluefin tuna spawning may potentially occur.

The problem was that this strategy was enforced without a parallel budget increasing for ensuring the same previous coverage in the four main spawning areas selected at the beginning of the aerial survey. Furthermore, in the period 2010-2015, the aerial survey was not carried out in 2012 and 2014, for budget shortage in one case and for a decision of the Steering Committee in the other case.

The combination of these changes and suspensions obviously resulted in a general increase of CV for the whole surveys and mostly in the last two. As a consequence, the results of this power analysis are not much comparable with the two previous analyses. In a possible simulation, one additional aerial survey in 2016, providing a coverage able to ensure a similar number of replicates and transect length in the four main areas (defined in the analysis as “overlapped areas”), could provide major elements for better defining a more reliable power analysis under a stable survey strategy.

III. Recommendations

Reduction of coefficients of variation

The main recommendation from this report is that if a reduction of coefficients of variations, at several levels (encounter rates, school size, detection function and additional variances), is required to be able to detect trends in population growth, increased coverage in terms of kilometers of tracks on effort should be done.

This reduction could be gained by a more stable survey strategy, focused on the same areas over the years and by using always the same type of aircraft and the same team of spotters in each area, because this approach will reduce most of the most relevant variables. The **main recommendation** is, thus, to concentrate the survey effort in the inside overlap areas for future surveys. This would reduce considerably the CVs, but also the cost per km surveyed, and the operational and logistical problems. Surveying the extended areas, on the other hand, do increase CVs, costs and operational and logistical problems, without increasing or improving the scientific knowledge (apart from verifying, once more, tht the chosen survey areas are the right ones).

Tables of different cost analysis and power analysis have been provided throughout the report so that ICCAT can evaluate the level of power (and therefore coverage) they want to achieve with the available resources.

It has also been shown that additional variance, when considering spatial and temporal variability, is very high. However, if additional variance wants to be applied for each area in particular, so that trends can be detected in each of them, then only temporal variability needs to be considered in the aera-specific additional variance. These could not be estimated with the available data due to the small sample size. It is possible that one more year of survey will provide enough data to estimate the area-specific additional variances with a higher confidence. But the requirement for area-specific or global (or any combination of areas) additional variations would ultimately depend on the consideration about the population structure (i.e. whether the whole Mediterranean is considered a single population, in which case all areas could be pooled together for abundance estimates, or 2 or more coexist and different areas can be assigned to different populations, in which case not all of them should be pooled together).

The additional variance from tagging data is much smaller, but possibly it could be reduced even further if more precise data from the tags will be available and the sample size gets increased. Also information on population structure could help in this sense, as data on tagging may need to be stratified, if possible, according to populations if more than one is present in the Mediterranean.

Long-term planning for the GBYP

ICCAT GBYP in general is increasing in a substantial way the scientific knowledge on bluefin tuna and this is essential for both assisting the recovery of this species and its sustainable management. The variety of bluefin tuna behaviours and its capacity to immediately respond to any environmental change are the great challenge of the bluefin tuna research and the aerial surveys are surely a part of this challenge.

The aerial surveys are still one of the very few available methodologies for providing fishery-independent data; no-one of the fishery independent methodologies is perfect and without biases but, considering the cost of the ICCAT GBYP aerial survey, this approach is not among the most expensive.

Several data have been collected in the first four surveys and surely the stop-and-go strategy, induced also by budget unavailability in some years, played a severe role in partly affecting the quality of the data, resulting in a substantial CV. The difficulties in keeping active the same teams in each area were another limit. These problems can be smoothed and possibly reduced to a minimum in a medium- long-term strategy, building on the knowledge already achieved with many efforts.

The Steering Committee and the SCRS had already planned a longer GBYP activity in 2014, extending it up to 2021, and this proposal was endorsed by the Commission. The current assessment of the aerial survey activity is that it is a clear operational success so far and that the scientific results need more years and efforts for providing the necessary trends to be used for scientific and management purposes. This was

already clearly stated in previous power analyses, because any trend needs several years to be duly detected and assessed, considering any possible improvement included in this report. The necessary budget should be provided in the following years to ensure that the aerial surveys will continue following a more stable strategy.

IV. References

- Buckland et al., 2001: Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L. & Thomas, L. , Introduction to distance sampling, 2001
- Cañadas, A. and Vazquez, J.A. 2016. Elaboration of 2015 data from the aerial survey on spawning aggregations. Atlantic-wide research programme on bluefin tuna (ICCAT-GBYP – PHASE 5 - 2015).
- Clark & Gelfand, 2006. James S. Clark and Alan E. Gelfand, Hierarchical Modeling for the Environmental Sciences, 2006
- Forney, K.A., Hanan, D.A. and Barlow, J. 1991. Detecting Trends in Harbor Porpoise Abundance from Aerial Surveys Using Analysis of Covariance. Fishery Bulletin, U.S. 89:367-377.
- Gelman, Andrew, and Donald B. Rubin. 1992. "Inference from iterative simulation using multiple sequences." *Statistical Science* (1992): 457-472.
- Kruschke J. K., 2011, Doing Bayesian Data Analysis : A tutorial with R, JAGS and Stan.
- Moore, J. E. and Barlow, J. 2011. Bayesian state-space model of fin whale abundance trends from a 1991-2008 time series of line-transect surveys in the California Current
- Ospina, R., & Ferrari, S. L. 2010. Inflated beta distributions. *Statistical Papers*, 51(1), 111-126.
- Ospina, R., & Ferrari, S. L. 2012. A general class of zero-or-one inflated beta regression models. *Computational Statistics & Data Analysis*, 56(6), 1609-1623
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & van der Linde, A. 2002 Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B*, 64, 583–639.
- Scott-Hayward, L.A.S., D.L. Borchers, M.L. Burt, S. Barco, H.L.Hass, C.R. Sasso and R.J. Smolowitz. 2014. Use of Zero and One-Inflated Beta Regression to Model Availability of Loggerhead Turtles off the East Coast of the United States. Final Report. Prepared for U.S. Fleet Forces Command. Submitted to Naval Facilities Engineering Command (NAVFAC) Atlantic, Norfolk, Virginia, under Contract No. N62470-10-D-3011, Task Order 40, issued to HDR Inc., Norfolk, Virginia. Prepared by CREEM, University of St. Andrews, St. Andrews, Scotland. July 2014.