

**IMPLEMENTACION DEL MUESTREO AEREO PARA LA
CALIBRACION Y VALIDACION (ICCAT GBYP 15/2019)
DENTRO DEL PROGRAMA “ATLANTIC-WIDE
RESEARCH PROGRAMME FOR BLUEFIN TUNA”
(ICCAT-GBYP Fase 9)**

INFORME FINAL



ELABORADO POR:

José Antonio Vázquez Bonales

30 julio 2019

1.- INTRODUCCION

La estima de los tamaños de grupos de especies de grandes peces y mamíferos marinos es un factor clave a la hora de obtener la abundancia de las especies objetivo y sus tendencias en el tiempo. En los muestreos en que se utiliza la metodología “distance sampling” se asume que los observadores que participan son experimentados y han realizado ejercicios previos de entrenamiento. Sin embargo, siempre pueden existir diferencias de mayor o menor magnitud entre ellos y por ese motivo en la mayoría de los análisis se incluye el factor observador como covariable, para testar si efectivamente afecta o no a la función de detección. Recientemente Lennert-Cody *et al.*, (2018), en una revisión sobre las metodologías “line transect” para estimar la abundancia de delfines en el Pacífico Tropical Oriental, sugirió, entre otras mejoras, obtener estimas más realistas mediante la inclusión de los tamaños de grupo y la utilización de factores de calibración para corregir esas estimas. Sin embargo, existen pocos trabajos que hayan abordado la cuestión de la calibración de las estimas de tamaños de grupo entre observadores que participan en muestreos de tñidos (Evenson *et al.*, 2018).

Los muestreos aéreos que ICCAT lleva promoviendo desde el año 2010 a través del programa GBYP (Atlantic-Wide Research Programme for Bluefin Tuna) tienen como último objetivo establecer un índice de abundancia independiente de datos pesqueros. A lo largo de las diferentes fases del programa se han identificado diversas fuentes potenciales de error que deben ser cuantificadas correctamente para conocer en qué grado afectan al índice y, por lo tanto, para evaluar la validez del mismo como herramienta de seguimiento de la abundancia de reproductores de atún rojo. La necesidad de realizar un ejercicio de calibración entre los observadores participantes en los muestreos es una cuestión que se viene planteando por parte del Comité de Dirección del GBYP desde hace años, tal como ha quedado reflejado en varios informes anuales. Siguiendo esa recomendación, dentro del programa de actividades del GBYP para el año 2019 se incluyeron varias partidas presupuestarias encaminadas a abordar esta cuestión de la calibración entre los observadores, una para el diseño teórico y determinación de requisitos prácticos mínimos para abordar el ejercicio de calibración y una segunda parte para su desarrollo en sí y el análisis de los datos que se registren durante el mismo.

Este informe corresponde al contrato relacionado con esta última partida, detallando las actividades realizadas durante el ejercicio de calibración y describiendo el proceso de análisis de los datos y los resultados obtenidos.

2.-METODOLOGIA

2.1. DISEÑO Y METODOLOGIA DE MUESTREO AÉREO

2.1.1.- OBSERVADORES

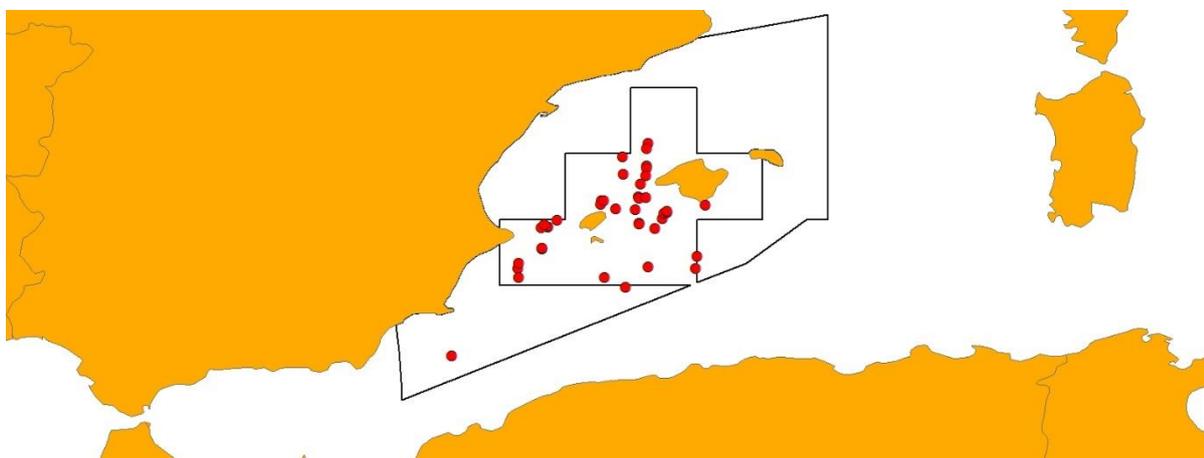
Debido a que este tipo de ejercicios conllevan un gasto económico considerable la elección de los observadores participantes se realizó de modo que se maximizara el potencial uso de los resultados para la mejora de la serie histórica. Con esta premisa, se hizo una revisión de la frecuencia de participación de cada observador en los muestreos aéreos desde su inicio en 2010 (18

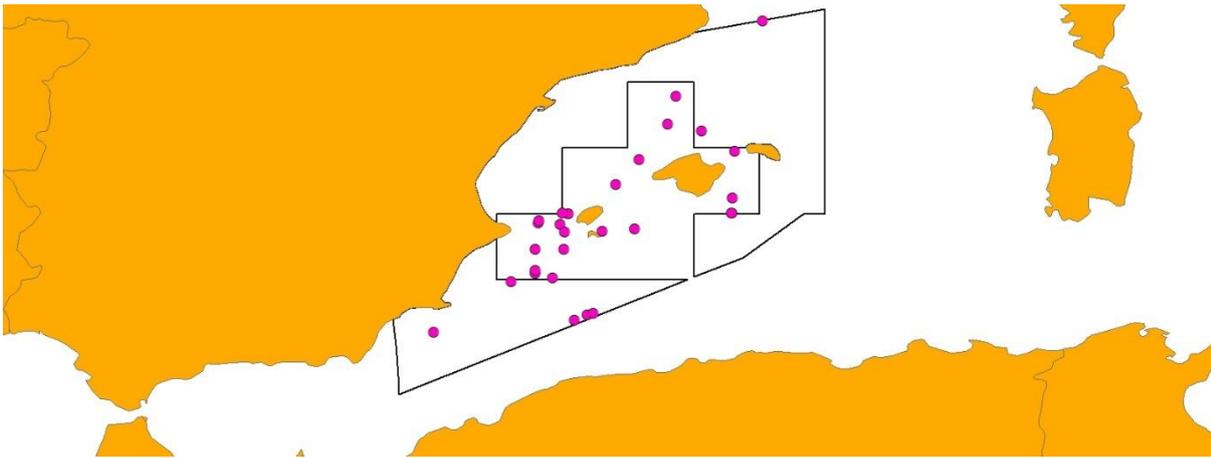
observadores profesionales y 41 observadores científicos). Por cuestiones logísticas, se decidió que lo más práctico era contar única y exclusivamente con los observadores profesionales (PS) ya que, en último término, son los valores de sus estimas los que se utilizan para obtener las abundancias finales en cada área de estudio. Teniendo en cuenta que hay un PS por área de estudio, el número total de PS tendría que ser mínimo de 4, que a su vez era el máximo que permitían los recursos disponibles y la capacidad de las avionetas adecuadas para llevar a cabo estas prospecciones. Atendiendo a su frecuencia de participación en la serie histórica de los muestreos aéreos, especialmente en los últimos tres años (incluyendo 2019), en los que se ha seguido una estrategia y metodología de muestro general más estandarizada, los observadores profesionales seleccionados fueron:

- ✓ AREA A: Luis Navarro (6 participaciones).
- ✓ AREA C: Khalifa Zariohi (4 participaciones).
- ✓ AREA E: Salvatore De Martino (5 participaciones).
- ✓ AREA G: Silvere Fontanet (2 participaciones).

2.1.2.- PERIODO Y AREA DE MUESTREO

Por motivos logísticos y de valores históricos de tasa de encuentro relativamente altos en comparación con el resto de áreas, se eligió la zona A, la de las Islas Baleares, como el área idónea para realizar el ejercicio de calibración. Se identificó un periodo de aproximadamente 10 días posterior a la finalización del muestreo aéreo anual como el óptimo para realizar el ejercicio de calibración, es decir entre el 29 de junio y 8 de julio. En base a estas premisas se revisó la distribución de avistamientos de BFT registrados en la zona de Baleares durante la última semana de junio y primera de julio de la serie histórica de los muestreos aéreos. Como se puede observar en el mapa 1, se identificaron dos zonas potencialmente aptas para centrar los esfuerzos de muestreo durante el ejercicio de calibración; una primera situada en la zona norte del canal de Mallorca y, una segunda situada al suroeste de la isla de Ibiza.





Mapa 1. Localizaciones de los avistamientos de BFT registrados durante el período de estudio (2010-2018) en el área A (Islas Baleares). En el mapa superior se muestran los avistamientos correspondientes a la última semana de junio y en el mapa inferior los correspondientes a la primera semana de julio.

2.1.3.- TIPO DE AVIONETA

La avioneta necesaria para la realización de este tipo de ejercicio de calibración tiene que cumplir una serie de requisitos indispensables:

- ✓ Disponer de todos los elementos de seguridad requeridos en la legislación vigente.
- ✓ Capacidad para alojar a 4 observadores profesionales y un coordinador científico de manera cómoda durante al menos 4 horas de vuelo.
- ✓ Disponer de ventanas adecuadas para la detección de grupos de atún rojo, así como para la estimación de los datos mínimos necesarios sobre esos cardúmenes (número de individuos, tamaño medio de los individuos y peso total del grupo).
- ✓ Disponer de cascos para todos los miembros de la tripulación y la capacidad de poder aislar las conversaciones entre el piloto y la torre de control del resto de los miembros de la tripulación.
- ✓ Disponer de salida de 200v para la conexión de un ordenador portátil.

2.1.4.- PROTOCOLO DE MUESTREO

Dado que el principal objetivo del ejercicio de calibración es encontrar el máximo número posible de grupos de BFT, se siguió una estrategia de muestreo similar al método usado durante el periodo en el que se permitía la pesca con avioneta. Los planes de vuelo se diseñaron y fueron llevados a cabo teniendo en cuenta las zonas históricas de tasas de encuentro altas, las

localizaciones de los avistamientos realizadas durante el mes de junio, las condiciones meteorológicas y las indicaciones del PS con más experiencia en el área de estudio.

Si bien no era estrictamente necesario para la intercalibración, la velocidad y altitud de vuelo se mantuvo siempre que las condiciones meteorológicas lo permitieron en 100 nudos y 300m, con el fin de realizar el ejercicio en las mismas condiciones que durante las campañas de prospección aérea habituales. Para la toma de datos se utilizó un ordenador portátil con el programa informático Logger 2010 (IFAW <http://www.marineconservationresearch.co.uk/downloads/logger-2000-rainbowclick-software-downloads/>) instalado y un dispositivo GPS conectado para la recogida automática tanto de los datos de esfuerzo como de los datos correspondientes a los avistamientos. El GPS utilizado fue en los dos casos un modelo GARMIN GPSMAP 78, con antena magnética GA25MCX y cable serie de conexión NMEA con adaptador puerto serie – USB. El formulario de avistamientos del programa Logger 2010 fue modificado de modo que se pudieron anotar las estimas de los 4 PS. A cada PS se le entregó una libreta pequeña para que apuntara las estimas de número, peso medio y peso total del grupo de animales, de manera independiente. Al término de cada avistamiento, el coordinador científico se aseguró de recibir las 4 hojas con las estimas de cada observador con todos los datos necesarios. Para la toma de las fotos de los cardúmenes durante los vuelos en círculos utilizó una cámara réflex NIKON modelo D5200 en modo S (prioridad a la velocidad) y disparo en ráfaga, con objetivo NIKON modelo 55-220 VR de enfoque automático y con estabilizador.

Por último, aprovechando la presencia del B/O Angeles Alvariño en la zona para el desarrollo de una campaña ictioplanctónica dirigida a la recogida de larvas de atún rojo, el coordinador científico del ejercicio de calibración se coordinó con la jefa de dicha campaña para que la avioneta pudiera sobrevolar, al menos en parte del tiempo cada vuelo, una zona próxima al buque y, en caso de detectar algún grupo de atún rojo cerca del mismo, darle indicaciones para orientarlo en la dirección correcta y así poder caracterizar el cardumen detectado mediante técnicas acústicas, concretamente el sónar científico MS70, con el fin último de intercalibrar ambos tipos de observaciones (aérea y acústica). La comunicación entre la avioneta y el barco se realizó mediante de uso de una radio portátil VHF icom ic-m72.

2.2.- EJERCICIO COMPLEMENTARIO CON FOTOS

De manera complementaria al ejercicio de calibración con avistamientos reales detectados durante los vuelos y en previsión de que una posible falta de un número suficiente de registros en condiciones de campo impidiera obtener resultados estadísticamente significativos, se preparó un ejercicio de gabinete, similar al realizado en los últimos tres años durante el curso de formación que se imparte en Madrid antes de la realización de las campañas de prospección aérea, basado en fotos de cardúmenes obtenidas de la base de datos GBYP. En esta ocasión se seleccionó un set de fotos más amplio del habitual, llegando a un total de 40 fotos, de las que 20 se clasificaron como de buena calidad y otras 20 como de calidad regular o mala. El propósito de utilizar fotos de diferente calidad era la de evaluar de una manera más amplia la capacidad de los observadores profesionales a la hora de estimar los diferentes valores necesarios para los análisis de abundancia.



Foto 1. Ejemplo de procesamiento de imágenes de cardúmenes de atún rojo para la obtención del número total de individuos presentes. Foto izquierda original, y foto derecha procesada.

Asimismo, se intentó que el set de fotos representara, en la medida de lo posible, diferentes tamaños de grupo, que se correspondieran con lo encontrado en la mar durante los muestreos anuales. En el caso de las fotos clasificadas como “buenas” y siempre que la calidad de la foto lo permitiese se contó el número total de los individuos presentes en cada foto con la ayuda del programa informático *ImageJ*, un programa de procesamiento de imagen digital de dominio público programado en Java, desarrollado en el National Institutes of Health de Estados Unidos de América, que es utilizado en ámbitos tan variados como, por ejemplo, el conteo de células en histología o el de estrellas en astronomía.

4.- RESULTADOS

4.1.- TRIPULACIÓN Y AVIONETA

Finalmente, las fechas del ejercicio de calibración fueron del 29 de junio al 8 de julio. Por motivos personales Khalifa Zariohi no pudo formar parte de la tripulación, y en su sustitución se seleccionó a Marc Bayle, quien se incorporó al grupo el día 1 de junio. Así, la tripulación quedó compuesta de la siguiente manera:

- ✓ Piloto (P): Jean-Jerome Houdaille (del 29 de junio al 8 de julio).
- ✓ Coordinador Científico (CL): José Antonio Vázquez (del 29 de junio al 8 de julio).
- ✓ Observador Profesional (PS): Luis Navarro (del 29 de junio al 8 de julio).
- ✓ Observador Profesional (PS): Salvatore De Martino (del 29 de junio al 8 de julio).
- ✓ Observador Profesional (PS): Silvere Fontanet (del 29 de junio al 8 de julio).
- ✓ Observador Profesional (PS): Marc Bayle (del 1 al 8 de julio).



Piloto (P):
Jean-Jerome Houdaille



Coordinador Científico (SS):
José Antonio Vázquez



Observador Profesional (PS):
Luis Navarro



Observador Profesional (PS):
Salvatore De Martino



Observador Profesional (PS):
Silvere Fontanet



Observador Profesional (PS):
Marc Bayle

La avioneta utilizada durante el ejercicio de calibración fue una BN2 Islander G-AWNT (Foto 2). La avioneta se posicionó en el aeropuerto de Ibiza el día 29 de junio. El día 30 de junio, se comprobó que los cristales de la zona media de la avioneta no eran aptos para el propósito del trabajo debido a que no eran lo suficientemente claros y presentaban rayas en diferentes zonas (Foto 3). El coordinador científico, tras consultas con el piloto y el coordinador del GBYP, decidió intentar buscar en Ibiza algún establecimiento para poder cambiar al menos una de las ventanas.



Foto 2. Avioneta BN2 Islander G-AWNT utilizada durante el ejercicio de calibración.



Foto 3. Aspecto de la ventana central derecha antes de su sustitución.

El día 1 de junio por la mañana se consiguió cambiar la ventana central derecha de la avioneta. En consecuencia, se decidió cambiar además la organización del posicionamiento de la tripulación durante el vuelo, priorizando que todos los observadores profesionales tuvieran el mejor acceso a la visualización de los cardúmenes y así facilitar sus estimas. Tras una reunión con todo el equipo el día 30 por la tarde, se decidió pasar de una configuración inicial donde el piloto y el coordinador científico se sitúan en la parte delantera y observadores profesionales en la parte central y trasera, con círculos alternos a derecha e izquierda en el momento de detectar grupos de atunes, a una configuración modificada en la que un observador profesional se situaría en el puesto del copiloto, otro en la parte central derecha, donde se había sustituido la ventana, y los otros dos en la parte trasera (Figura 2). El coordinador científico se situaría en la parte central izquierda, donde seguía habiendo una ventana en mal estado. Con esta nueva configuración los círculos alrededor del cardumen de atunes se realizaron siempre en sentido horario ya que en la parte trasera de la avioneta había espacio suficiente para que los observadores intercambiaran sus posiciones para así poder tener acceso visual directo a los animales.

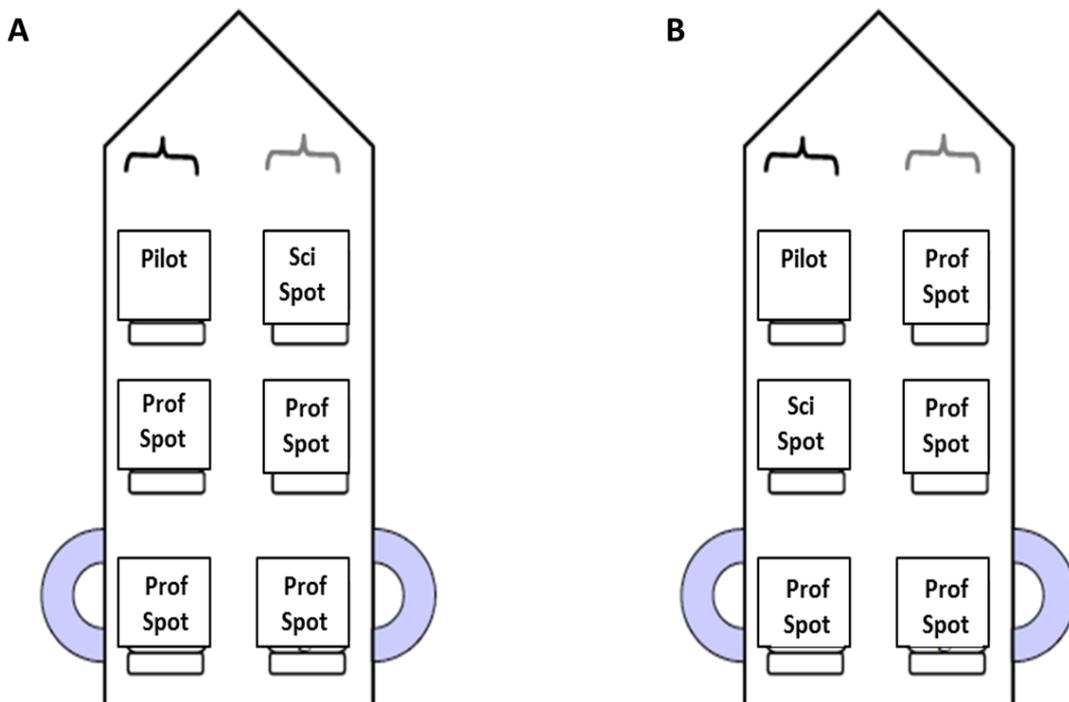
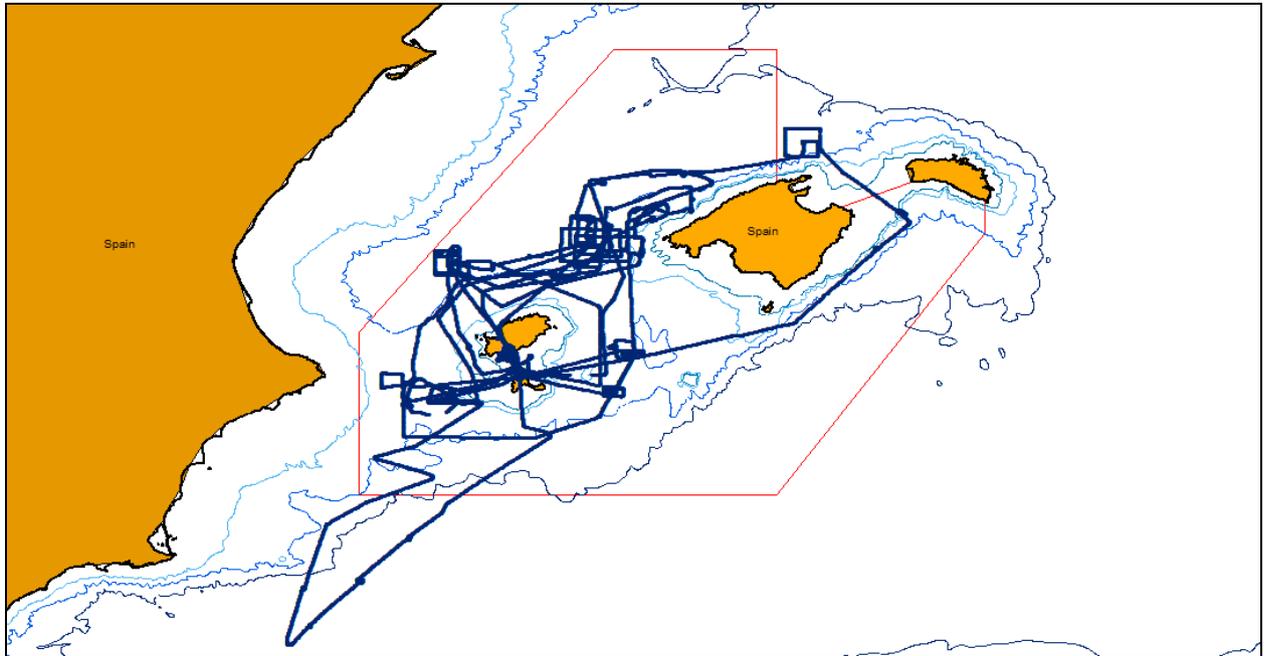


Figura 2. Configuración original A y modificada B de la tripulación.

4.2. – ESFUERZO Y AVISTAMIENTOS

En el mapa 1 se muestra el esfuerzo realizado en el área de Baleares durante el ejercicio de calibración. En total se han realizado 7 vuelos que han sumado 27 horas y 4936 km de esfuerzo de muestreo.



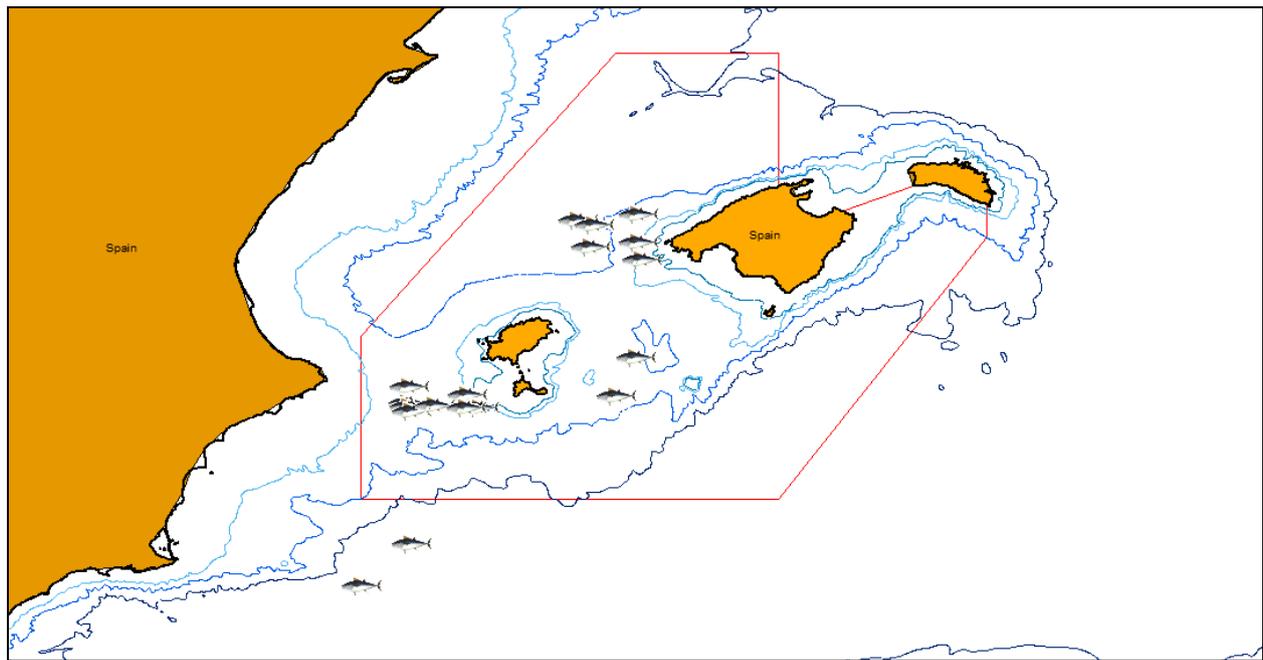
Mapa 1. Esfuerzo realizado durante el ejercicio de calibración.

En la Tabla 1 se muestra el resumen de los días volados y días no volados en el área de estudio. El único día que no se pudo volar fue debido a malas condiciones meteorológicas.

Tabla 1. Resumen días de vuelo y días no volados durante el ejercicio de calibración.

	Volado	No volado	Causa
29 Junio	Posicionamiento		
30 Junio			
1 Julio			
2 Julio			
3 Julio			
4 Julio			
5 Julio			
6 Julio			
7 Julio			Meteorología
8 Julio	Vuelta a base		

En el mapa 2 se muestra los avistamientos de BFT registrados durante el ejercicio de calibración.



Mapa 2. Avistamientos de BFT durante el ejercicio de calibración.

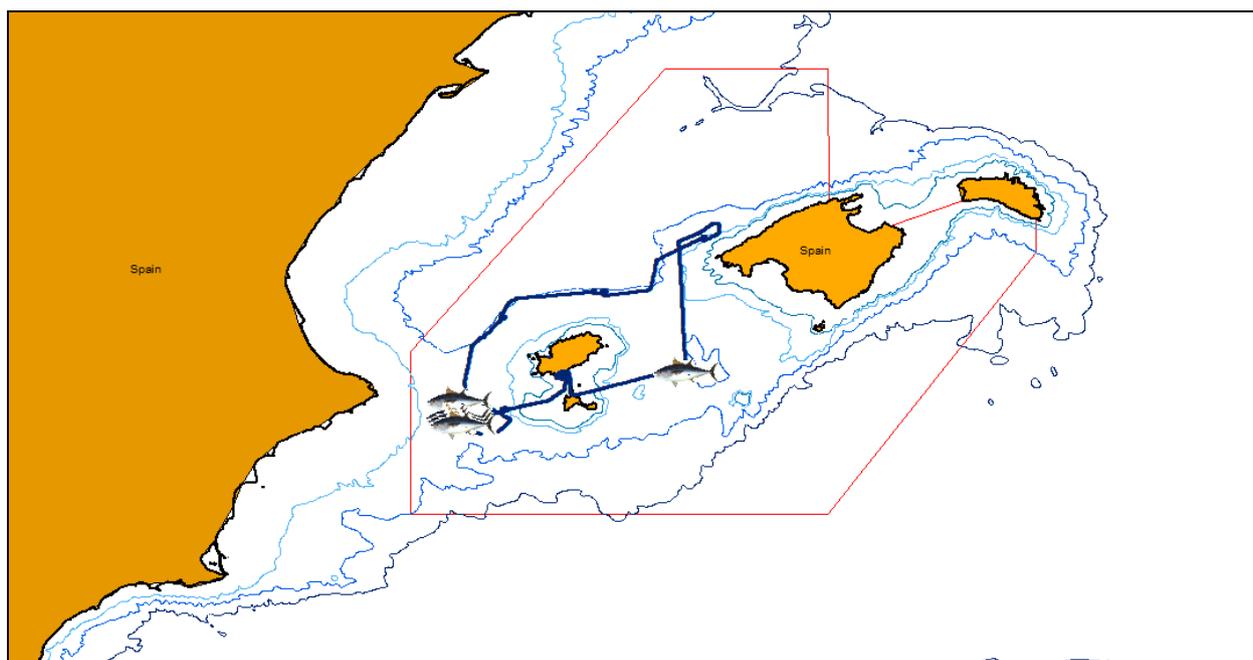
En la Tabla 2 se muestran las posiciones de cada uno de los 18 avistamientos de BFT registrados durante el ejercicio de calibración.

Tabla 2. Posición de los avistamientos de BFT registrados durante el ejercicio de calibración.					
AVISTAMIENTO	VUELO	FECHA	LATITUD	LONGITUD	
1	1	30/06/2019 08:24:38	38.87317	2.13092	
2	1	30/06/2019 10:16:40	38.69474	0.73576	
3	1	30/06/2019 10:27:00	38.59629	0.72854	
4	1	30/06/2019 10:30:00	38.57499	0.74168	
5	1	30/06/2019 10:39:23	38.54722	0.75625	
6	2	01/07/2019 15:59:12	38.6369	2.00756	
7	3	02/07/2019 15:22:34	39.75408	2.14529	
8	3	02/07/2019 16:07:30	39.71443	1.77393	
9	3	02/07/2019 16:13:15	39.6846	1.87274	
10	3	02/07/2019 16:19:15	39.55473	1.85462	
11	4	03/07/2019 09:16:30	39.47717	2.16827	
12	4	03/07/2019 09:24:40	39.58378	2.14411	
13	5	04/07/2019 09:09:31	38.64389	1.0962	
14	5	04/07/2019 09:23:40	38.58034	0.89632	
15	5	04/07/2019 09:36:11	38.57006	1.16805	
16	7	06/07/2019 11:43:27	38.56572	1.08774	
17	7	06/07/2019 13:32:23	37.47323	0.44658	
18	7	06/07/2019 13:52:59	37.73089	0.7498	

A continuación se resumen el tiempo, la distancia, los avistamientos registrados y las incidencias más relevantes acaecidas durante cada uno de los vuelos realizados.

30 JUNIO 2019

VUELO 1 - 30 JUNIO	
HORA INICIO	09:44:23
HORA FIN	13:37:43
TIEMPO TOTAL	03:53:20
DISTANCIA TOTAL	706km
AVISTAMIENTOS BFT	5



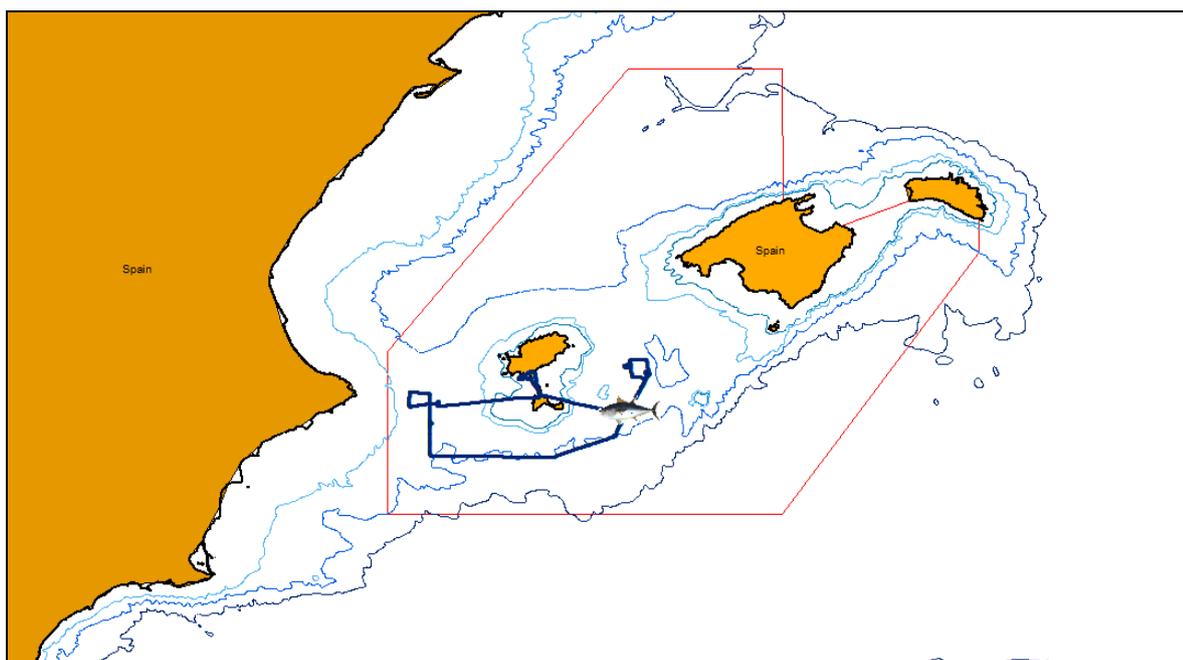
Mapa 3. Trayecto recorrido y Avistamientos de BFT

COMENTARIOS:

Al llegar al aparcamiento donde está la avioneta la tripulación comprueba que las ventanas situadas en la zona central no son aptas para el trabajo ya que los cristales no están totalmente transparentes. El coordinador científico se lo hace saber al piloto, al responsable de GAM y al coordinador de GBYP. Aun así, se decide hacer un vuelo de prueba para evaluar el alcance del problema.

1 JULIO 2019

VUELO 2 - 1 JULIO	
HORA INICIO	16:25:04
HORA FIN	19:25:55
TIEMPO TOTAL	03:00:51
DISTANCIA TOTAL	563km
AVISTAMIENTOS BFT	1



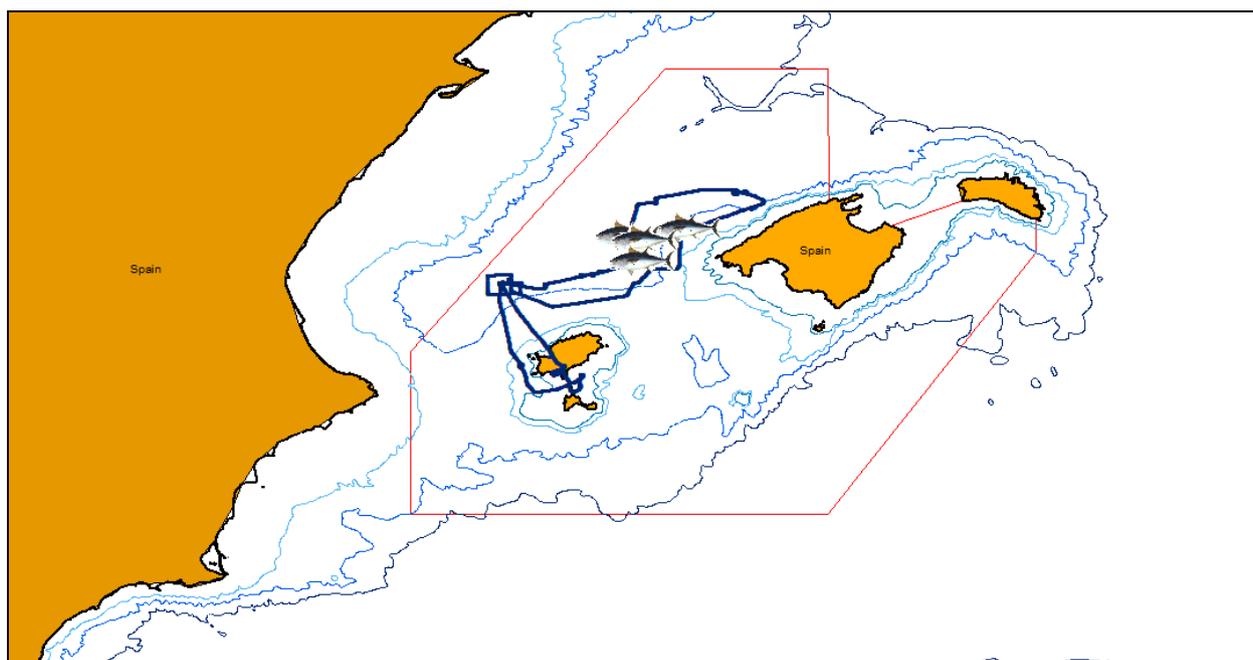
Mapa 4. Trayecto recorrido y Avistamientos de BFT

COMENTARIOS:

Tras realizar una labor ímproba de búsqueda el día 30 de junio conseguimos encontrar un establecimiento donde nos pueden hacer una nueva ventana. El día 1 de julio se instala la nueva ventana en la zona central derecha de la avioneta. La configuración inicial del equipo se modifica de modo que se garantice que los 4 PS puedan ver correctamente los grupos de atunes detectados. Se realiza el primer contacto por radio con el B/O Ángeles Alvariño con éxito.

2 JULIO 2019

VUELO 3 - 2 JULIO	
HORA INICIO	15:22:46
HORA FIN	19:22:19
TIEMPO TOTAL	03:59:33
DISTANCIA TOTAL	740km
AVISTAMIENTOS BFT	4



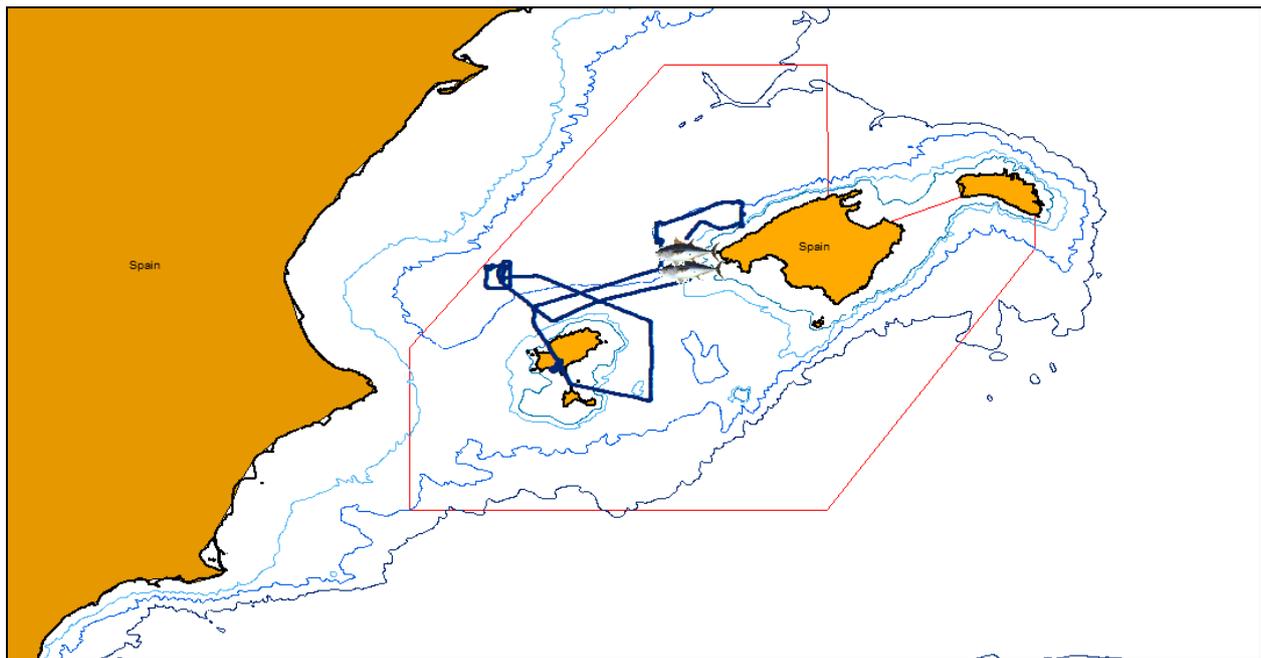
Mapa 5. Trayecto recorrido y Avistamientos de BFT

COMENTARIOS:

Se acuerda con el equipo del B/O Ángeles Alvariño que durante los vuelos se dedicará un periodo a volar 5 millas alrededor del buque para intentar encontrar algún grupo de atunes en sus proximidades, con el fin de poder dirigir al buque al mismo y poder así llevar a cabo ejercicios de intercalibración entre metodologías. Durante el vuelo se registran 4 avistamientos de BFT, pero ninguno de ellos cerca del buque. El primero de los avistamientos corresponde a una jaula que transporta el barco Salve Regina. Los observadores estiman el número, peso medio y peso total de los atunes que lleva la jaula para comparar posteriormente con los datos reales proporcionados por la empresa e utilizar este dato como referencia para una posible validación de la calibración.

3 JULIO 2019

VUELO 4 - 3 JULIO	
HORA INICIO	08:47:31
HORA FIN	12:52:36
TIEMPO TOTAL	04:05:05
DISTANCIA TOTAL	729km
AVISTAMIENTOS BFT	2



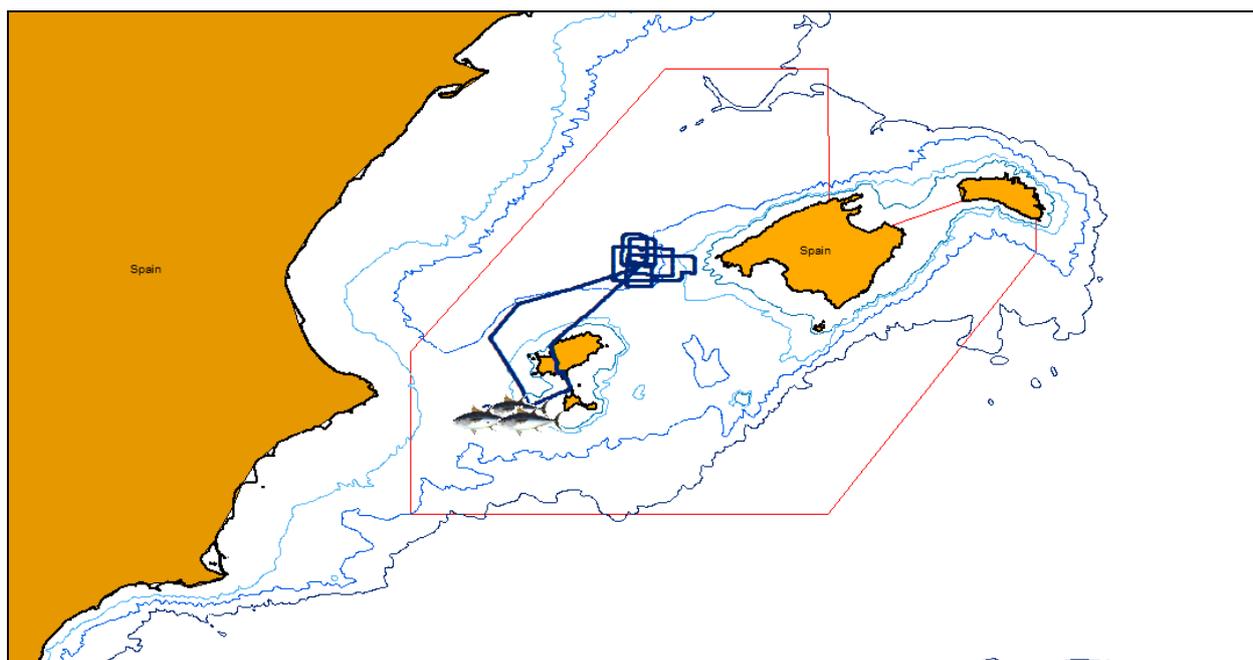
Mapa 6. Trayecto recorrido y Avistamientos de BFT

COMENTARIOS:

Durante el vuelo se registran 2 avistamientos de BFT, ninguno de ellos cerca del B/O Ángeles Alvariño.

4 JULIO 2019

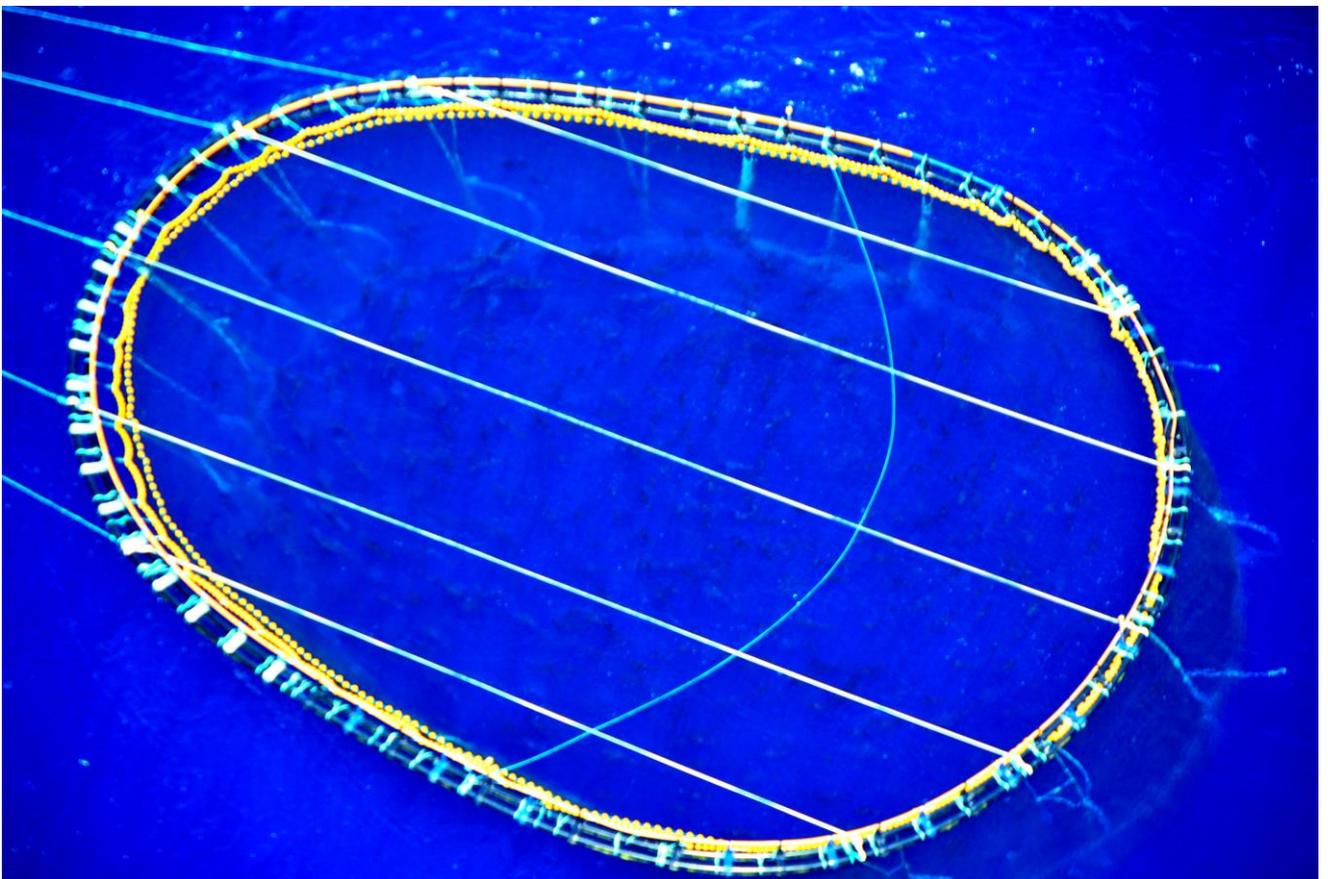
VUELO 5 - 4 JULIO	
HORA INICIO	10:48:04
HORA FIN	14:52:13
TIEMPO TOTAL	04:04:09
DISTANCIA TOTAL	740km
AVISTAMIENTOS BFT	3



Mapa 7. Trayecto recorrido y Avistamientos de BFT

COMENTARIOS:

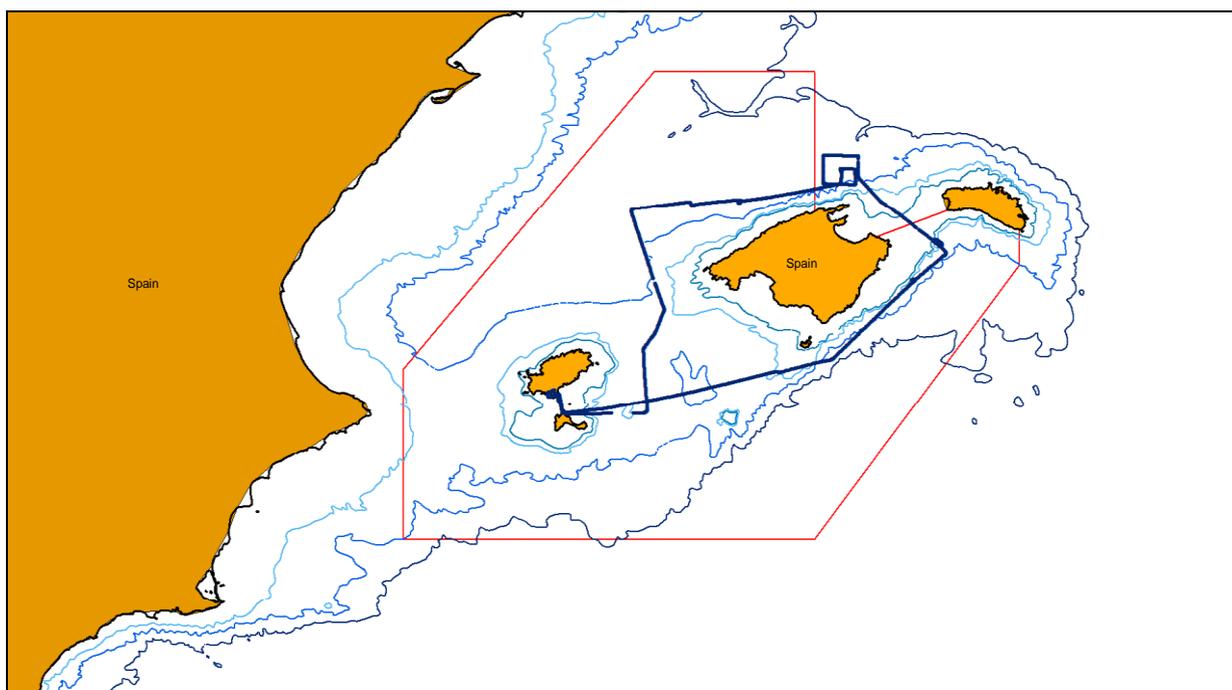
Se acuerda quedar con en B/O Ángeles Alvariño en la zona norte del canal de Mallorca después de que termine el ejercicio de sondeo de la jaula de atunes del buque Salve Regina, jaula que se sobrevoló ya el día 2 de julio para obtener estimas de tamaño y peso por los observadores. En el inicio del vuelo se registran 3 avistamientos de BFT, al suroeste de Ibiza, pero no en la zona norte donde se concentra el esfuerzo.



Fotos 4 y 5. B/O Ángeles Alvariño durante su aproximación a la embarcación Salve Regina que traslada una jaula con atún rojo.

5 JULIO 2019

VUELO 6 - 5 JULIO	
HORA INICIO	09:41:34
HORA FIN	13:39:01
TIEMPO TOTAL	03:57:27
DISTANCIA TOTAL	712km
AVISTAMIENTOS BFT	0



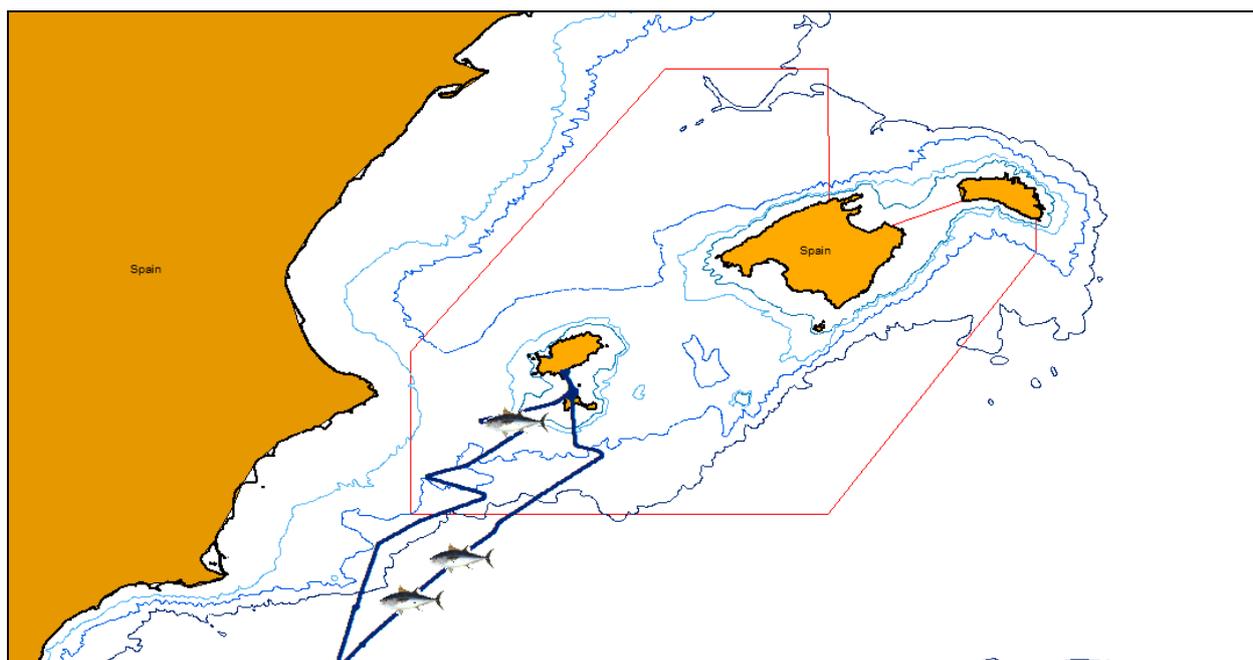
Mapa 8. Trayecto recorrido y Avistamientos de BFT

COMENTARIOS:

Dado que en este día el B/O Ángeles Alvariño se encuentra cerca de la zona noreste de Mallorca y considerando además las mejores condiciones meteorológicas en la misma, se decide hacer un vuelo alrededor de la isla de Mallorca pasando por la posición del B/O. Durante este vuelo no se detecta ningún avistamiento de BFT.

6 JULIO 2019

VUELO 7 - 6 JULIO	
HORA INICIO	13:10:08
HORA FIN	17:12:23
TIEMPO TOTAL	04:02:15
DISTANCIA TOTAL	746km
AVISTAMIENTOS BFT	3



Mapa 9. Trayecto recorrido y Avistamientos de BFT

COMENTARIOS:

El último día que vamos a poder volar decidimos ir al sur de Ibiza ya que el parte meteorológico es menos malo en esa zona y porque lo normal es que los atunes en esta época vayan trasladándose al sur en su viaje hacia el estrecho de Gibraltar. Durante el vuelo se registran 3 avistamientos de BFT.

4.3.- COMPARACION ESTIMAS DE NÚMERO, PESO MEDIO Y TOTAL DE LOS CARDUMENES

Como se muestra anteriormente en la tabla 2 en total se han realizado 18 avistamientos de BFT. Sin embargo, no en todos ellos ha sido posible registrar los datos correspondientes al número, peso medio y peso total. En concreto, en los dos últimos avistamientos del día 4 de julio se detectó el “temblor” que indicaba la presencia de los cardúmenes, pero al llegar al punto donde se encontraban los animales no se pudieron observar debido a que estaban demasiado profundos. En el caso del avistamiento registrado el día 1 de julio, solo un observador profesional pudo facilitar los valores de las diferentes estimas, ya que fue el único que los vio. Por otro lado, en los 6 primeros avistamientos solo se obtuvieron estimas de 3 de los 4 observadores profesionales, ya que el cuarto observador profesional, Marc Bayle, llegó con 1 día de retraso a Ibiza. Asimismo, en los 6 primeros avistamientos el coordinador científico detectó al llegar al hotel que el observador profesional Salvatore De Martino solo había facilitado el peso total de los cardúmenes. Tras insistir varias veces, este observador entregó los datos del peso medio de los 6 avistamientos, pero solo facilitó el número de ejemplares de los dos primeros. En los siguientes días el coordinador científico insistió a todos los observadores profesionales sobre la necesidad de aportar todos los datos requeridos y de la importancia de esta cuestión para la consecución del objetivo del ejercicio de calibración.

A la hora de comparar los datos entre los diferentes observadores se han tenido en cuenta varias cuestiones, siendo la primera el valor de referencia a utilizar. En la bibliografía se describen diferentes aproximaciones. Gerrodete *et al.*, (2002; 2019) utilizan las estimas del número de delfines en fotos tomadas desde un helicóptero como referencia para corregir las de observadores que están en el barco, mientras que Hammond *et al.*, (2016) utilizan las estimas del número de delfines o ballenas de los observadores situados en una plataforma superior del barco para corregir las estimas de los observadores situados en una plataforma inferior. Por otro lado, Evenson *et al.*, (2018) consideran las estimas de peso y número de atunes rojos del observador más experimentado como los valores de referencia con el que comparar las estimas del resto de observadores que han participado en el muestreo a lo largo del periodo de estudio. La segunda cuestión tiene que ver con el método usado para calcular el factor de corrección. Mientras que Gerrodete *et al.*, (2002;2019) utilizan una combinación de modelos de regresión lineales y no lineales con modelos bayesianos para calcular los diferentes factores de corrección de las estimas del número de animales para cada observador incorporando los efectos del estado de la mar y de los diferentes años, Hammond *et al.*, (2016), calculan los factores de corrección para cada especie como la división entre la suma de las estimas de los observadores de la plataforma superior e inferior. En el caso de Evenson *et al.*, (2018) se utilizan modelos lineales generalizados para ajustar las diferencias dos a dos entre el observador de referencia y el resto de observadores. Hay que señalar que todos estos estudios de referencia, a diferencia del nuestro, se basan en una considerable cantidad de datos recogidos durante varios años.

En este ejercicio tenemos varias limitaciones. El número de muestras del que se dispone es muy bajo (n=16) y no en todas ellas se tienen los datos correspondientes a los 4 observadores. Por otro lado, debido a las características propias de los avistamientos de BFT y de la disposición de los miembros del equipo dentro de la avioneta, no se cuenta con fotografías claras en las que este todo el cardumen representado, permitiéndose así poder contar con un número total de individuos

mediante el recuento directo con el programa ImageJ. Por último, a pesar de tener estimas sobre un cardumen que se encontraba en una jaula de transporte, hasta la fecha todavía no se dispone de los valores reales de número de individuos, peso medio y peso total. Únicamente se dispone del valor aproximado del peso total (177 toneladas).

A la hora de aplicar los análisis estadísticos es necesario conocer en primer lugar si las muestras se ajustan a una distribución normal o no. En la tabla 1 se muestran los resultados del test de normalidad Shapiro-Wilk, para las tres variables; número de individuos, peso medio y peso total. En el caso del número de individuos los datos correspondientes de Luis y Silvere no se ajustan a una distribución normal mientras que los datos correspondientes a Marc y Salvatore si se ajustan a una distribución normal. Para la variable peso medio el único observador cuyos datos no se ajustan a una distribución normal son los de Salvatore y para la variable peso total el único observador cuyos datos no se ajustan a una distribución normal son los de Luis. Dado que los test de normalidad de Shapiro-Wilk no muestran resultados concluyentes, considerando al bajo número de muestras, y además debido al hecho de que desde el punto de vista conceptual no es esperable que las variables observadas presenten en la población distribuciones normales unimodales, a la hora de comparar las estimas entre los observadores se ha decidido utilizar test no paramétricos, en concreto se han utilizado el test de suma de rangos de Kruskal-Wallis que permite evaluar si las hay diferencias entre las diferentes muestras y el test de Wilcoxon o también denominada prueba de los rangos con signo de Wilcoxon que permite evaluar si hay diferencias entre muestras mediante la comparación pareada de los datos. Este último test también se utiliza en estudios de calibración de observadores que estimas las edades de los peces a través de la lectura de los otolitos (Eltnik, 2000). A pesar de que, como se ha comentado anteriormente, los datos disponibles son escasos, se ha explorado el uso de las regresiones lineales para obtener una primera idea de cuál es el comportamiento de cada observador con respecto al resto. Debido a que los datos disponibles no apoyan la candidatura de ninguno de los observadores como valor de referencia, se ha decidido tomar el valor medio de las estimas de cada una de las variables.

Tabla 1. Resultados del test de normalidad Shapiro-Wilk para las tres variables medidas por los observadores; número de individuos, peso medio y peso total (* = $p < 0.05$).

	LUIS		SILVERE		MARC		SALVATORE	
	Valor <i>w</i>	<i>p</i>						
Numero individuos	0.69547	0.00015*	0.87847	0.04502*	0.92585	0.40830	0.94515	0.56750
Peso medio	0.90146	0.08481	0.92271	0.21180	0.90861	0.27160	0.82137	0.00700*
Peso total	0.66842	0.00007*	0.79744	0.00464	0.95582	0.73730	0.95296	0.57220

En cuanto al **número total de individuos**, el coeficiente de variación para las estimas de los 4 observadores varía entre 0% y 108%, con un valor medio del 33.8% (desviación estándar 26.8%) que indica que las muestras presentan un nivel de homogeneidad relativamente alto, siendo los valores de CV para las estimas menores de 1000 individuos (35.7%) ligeramente mayores que los valores de CV para las estimas mayores de 1000 individuos (31.4%).

Tabla 2. Resultados del test de Wilcoxon para la variable número de individuos comparando parejas de observadores.

	valor v	lwr	upr	p
MARC-LUIS	6	-1795	100	0.0929
SALVATORE-LUIS	28	-1310	415	0.9592
SILVERE-LUIS	46	-525	326	0.6829
SALVATORE-MARC	35	-75	530	0.1386
SILVERE-MARC	32.5	-249	700	0.6101
SILVERE-SALVATORE	36	-475	545	0.8139

El resultado del test de suma de rangos de Kruskal-Wallis indica que no existen diferencias estadísticamente significativas ($p=0.932$) entre las estimaciones de los 4 observadores. El resultado del test de Wilcoxon es coincidente con el test anterior e indica que no existen diferencias estadísticamente significativas entre ninguna de las comparaciones dos a dos (Tabla 2).

En la figura 3 se muestra la relación entre las estimas del número de individuos correspondientes a cada observador y el valor promedio de las mismas, así como las líneas de las regresiones lineales en cada caso.

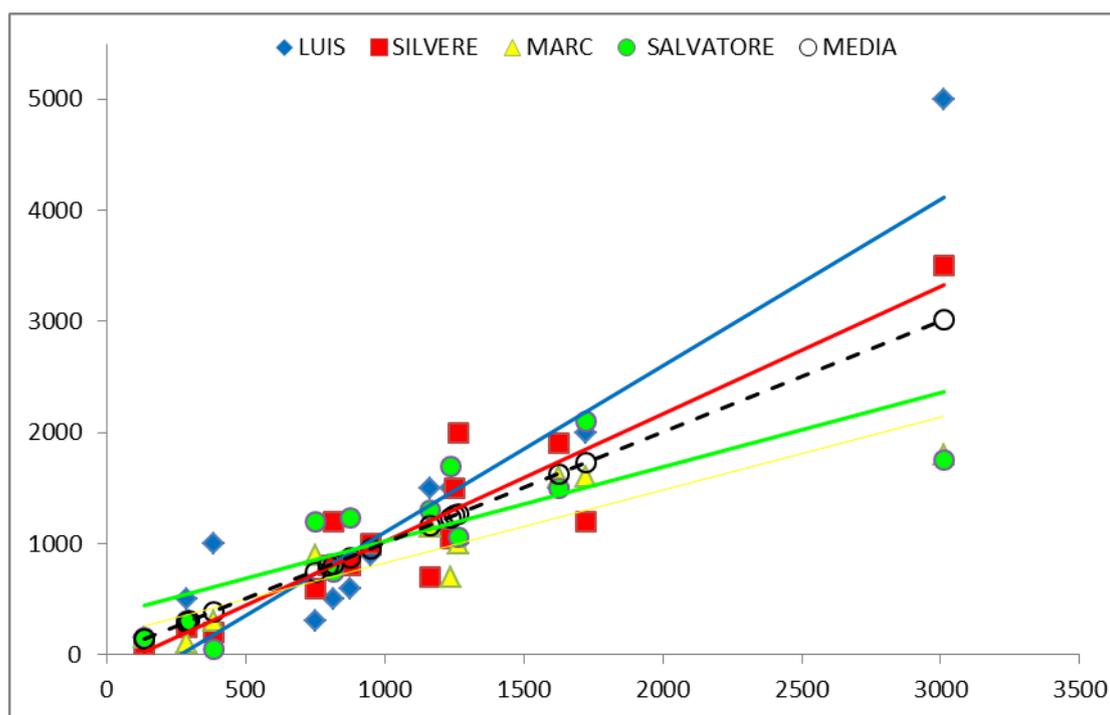


Figura 3. Relación entre las estimas del número de individuos correspondientes a cada observador y el valor promedio de las mismas. Las líneas son el resultado del análisis de regresión lineal para cada observador; azul=Luis; rojo=Silvere; amarillo=Marc y verde= Salvatore. La línea negra indica los valores de referencia.

En todos los casos los resultados de las regresiones han sido estadísticamente significativos, con valores de R^2 de 0.84, 0.86, 0.82 y 0.63 en el caso de Luis, Silvere, Marc y Salvatore respectivamente (Tabla 3).

Tabla 3. Resultados del análisis de regresión múltiple entre las estimaciones del número de individuos de los observadores y el valor medio.

	Intercepto	Pendiente	R2	p	
LUIS	-394.8	1.499	0.845	4.88E-07	***
SILVERE	-126.4	1.148	0.858	7.04E-07	***
MARC	169.7	0.657	0.821	3.02E-04	***
SALVATORE	346.7	0.672	0.635	1.92E-03	***

En base a los valores obtenidos de las pendientes de las rectas, se puede decir que Silvere, con un valor de 1.15 es el observador que más se aproxima a la recta de referencia, y que Luis, con un valor de 1.50 es el observador que más se aleja de la recta de referencia. Generalizando se podría concluir que por debajo de un número de individuos cercano a 1000 Marc y Salvatore tienden a sobrestimar ligeramente, mientras que Luis y Silvere tienden a subestimar ligeramente. Por encima de este valor arbitrario de 1000 individuos las tendencias cambian y, mientras que Marc y Salvatore tienden a subestimar, Luis y Silvere tienden a sobreestimar.

En cuanto al **peso medio**, el coeficiente de variación para las estimas de los 4 observadores varía entre 8.3% y 46.8%, con un valor medio del 27% (desviación estándar 40.8%) que indica que las muestras presentan un nivel de homogeneidad ligeramente mayor que el obtenido para la variable número total de individuos.

Tabla 4. Resultados del test de Wilcoxon para la variable peso medio comparando parejas de observadores.

	valor v	lwr	upr	p
MARC-LUIS	12	-61	5	0.1137
SALVATORE-LUIS	0	-83	-52	0.0009*
SILVERE-LUIS	67.5	-18	25	0.6696
SALVATORE-MARC	10.5	-80	4	0.0829
SILVERE-MARC	45	10	65	0.0075*
SILVERE-SALVATORE	104	43	97	0.0012*

El resultado del test de suma de rangos de Kruskal-Wallis indica que existen diferencias estadísticamente significativas ($p=0.00044$) entre las estimaciones de los 4 observadores. Los resultados del test de Wilcoxon son diferentes en función de la pareja considerada. Mientras que en el caso de las parejas Marc-Luis, Silvere-Luis y Salvatore-Marc las diferencias no son estadísticamente significativas, en el resto, Salvatore-Luis, Silvere-Marc y Silvere-Salvatore, las diferencias son estadísticamente significativas (Tabla 4).

En la figura 4 se muestra la relación entre las estimas del peso medio correspondientes a cada observador y el valor promedio de las mismas, así como las líneas de las regresiones lineales en cada caso.

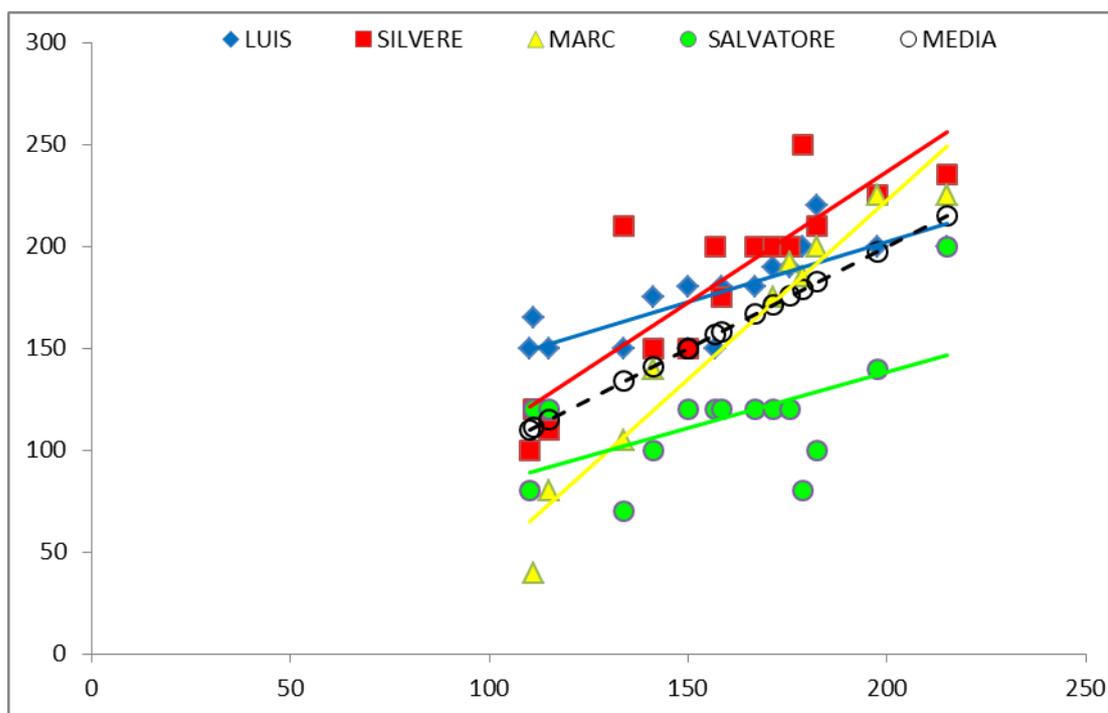


Figura 4. Relación entre las estimas de peso medio correspondientes a cada observador y el valor promedio de las mismas. Las líneas son el resultado del análisis de regresión lineal para cada observador; azul=Luis; rojo=Silvere; amarillo=Marc y verde= Salvatore. La línea negra indica los valores de referencia.

En el caso de la variable peso medio (tabla 5), a pesar de que los resultados de las regresiones han sido estadísticamente significativos para los 4 observadores, el nivel de explicación de la variación de las estimas de cada observador en relación al valor medio es muy diferente para cada observador, siendo el valor mínimo de R2 el obtenido para Salvatore (0.314) y el máximo para Marc (0.944).

Tabla 5. Resultados del análisis de regresión múltiple entre las estimaciones del peso medio de los observadores y el valor medio.

	Intercepto	Pendiente	R2	p	
LUIS	84.0094	0.591	0.631	2.36E-04	***
SILVERE	-20.033	1.284	0.751	2.89E-05	***
MARC	-127.9895	1.755	0.944	2.86E-06	***
SALVATORE	29.2753	0.546	0.314	2.98E-02	*

En base a los valores obtenidos de las pendientes de las rectas, se puede decir que, de nuevo, Silvere, con un valor de 1.28 es el observador que más se aproxima a la recta de referencia. El resto de observadores se alejan del valor de referencia, con pendientes menores en el caso de Salvatore (0.54) y Luis (0.59), y pendientes mayores en el caso de Marc (1.75). Considerando todo el rango de valores de peso medio disponibles en la muestra, Silvere tiende a sobreestimar, Salvatore tiende a subestimar, mientras que Luis y Marc no presentan una tendencia clara.

Por último, en cuanto al **peso total**, el coeficiente de variación para las estimas de los 4 observadores varía entre 10.4% y 135.1%, con un valor medio del 46.1% (desviación estándar 34.8%), que indica que esta variable es la que menos nivel de homogeneidad presenta entre todos los observadores

Tabla 6. Resultados del test de Wilcoxon para la variable peso total comparando parejas de observadores.

	valor v	lwr	upr	p
MARC-LUIS	8	-313	-2	0.04685
SALVATORE-LUIS	19.5	-126	-11	0.02138*
SILVERE-LUIS	36.5	-93	30	0.31510
SALVATORE-MARC	18	-62	21	0.33290
SILVERE-MARC	44	-24	188	0.09260
SILVERE-SALVATORE	84	2	130	0.04793*

El resultado del test de suma de rangos de Kruskal-Wallis indica que no existen diferencias estadísticamente significativas ($p=0.4739$) entre las estimaciones de los 4 observadores. En cuanto a los resultados del test de Wilcoxon, salvo en el caso de las comparaciones Salvatore-Luis y Silvere-Salvatore donde las diferencias son estadísticamente significativas, con valores de p de 0.02138 y 0.04793 respectivamente, en el resto de parejas las diferencias no son estadísticamente significativas (tabla 6).

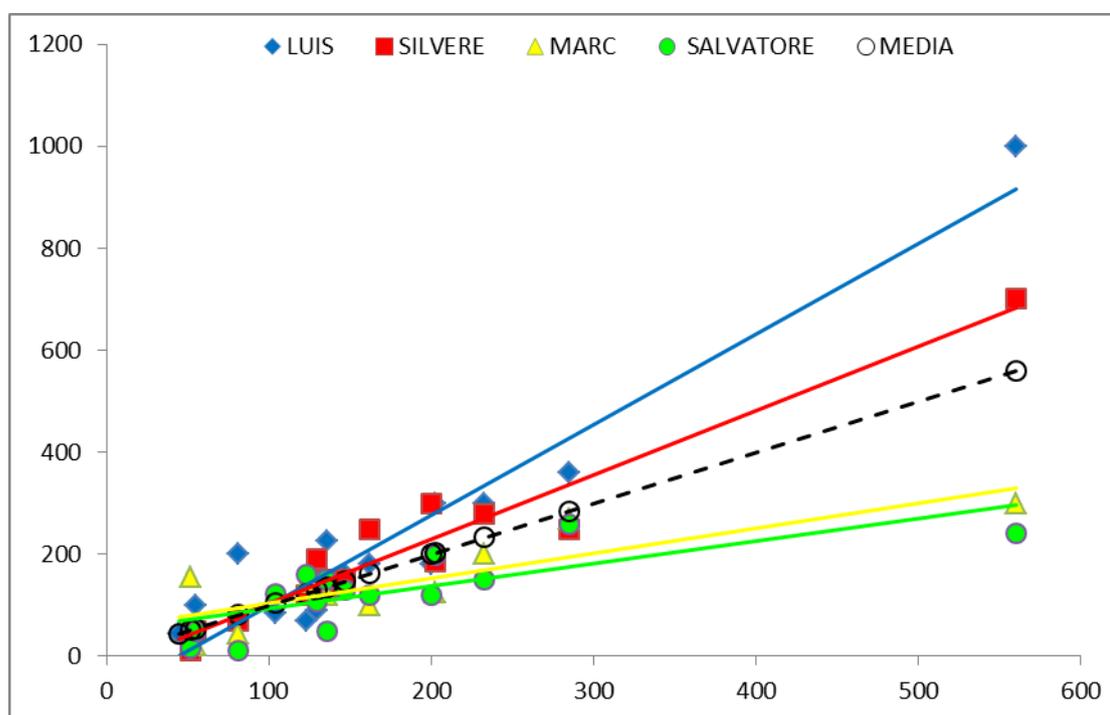


Figura 5. Relación entre las estimas de peso total correspondientes a cada observador y el valor promedio de las mismas. Las líneas son el resultado del análisis de regresión lineal para cada observador; azul=Luis; rojo=Silvere; amarillo=Marc y verde= Salvatore. La línea negra indica los valores de referencia.

En la figura 5 se muestra la relación entre las estimas del peso total correspondientes a cada observador y el valor promedio de las mismas, así como las líneas de las regresiones lineales en cada caso. En el caso de la variable peso total (tabla 7), los resultados de las regresiones han sido estadísticamente significativos para los 4 observadores. Una vez más, Silvere es el observador que presenta un valor de la pendiente más cercano a 1. Para Marc y Silvere los valores de la pendiente son muy similares, 0.488 y 0.446 respectivamente, e inferiores a 1, y para Luis el valor de la pendiente es superior (1.776) a la pendiente de referencia.

Tabla 7. Resultados del análisis de regresión múltiple entre las estimaciones del peso total de los observadores y el valor medio.

	Intercepto	Pendiente	R2	p	
LUIS	-79.1658	1.776	0.919	5.20E-09	***
SILVERE	-22.582	1.263	0.935	1.83E-08	***
MARC	55.7797	0.488	0.686	3.07E-03	***
SALVATORE	48.0828	0.446	0.565	1.23E-03	***

Generalizando, se puede decir que, mientras que Luis y Silvere tienden a sobreestimar, siendo Luis el que se sitúa en valores superiores, Marc y Salvatore tienden a subestimar a niveles muy parecidos entre ambos.

4.4.- COMPARACION DE ESTIMAS EN EL EJERCICIO CON FOTOGRAFIAS

De manera complementaria al ejercicio de comparación de estimas con datos recogidos durante los vuelos, se preparó un ejercicio práctico de comparación de estimas a partir de la visualización de fotos de cardúmenes de diferente calidad. El ejercicio de comparación de fotos se realizó el día 6 de julio a las 19:00 en las instalaciones del Ajuntament de Sant Josep de sa Talaia (foto 6). Los observadores se situaron de manera separada en la sala y tuvieron todo el tiempo que consideraron necesario para dar los valores del número de individuos, peso medio y peso total.



Foto 6. Sala del Ajuntament de Sant Josep de sa Talaia donde se realizó el ejercicio de calibración con fotografías de cardúmenes de atún rojo.

Antes de analizar los resultados de las comparaciones de las estimas entre los observadores, es necesario resaltar que en el caso de que los observadores consideraran que una foto no era lo suficientemente buena como para dar una estima realista se permitió que el observador pudiera dejar la casilla del formulario en blanco. En concreto, en relación al número total de individuos Luis Navarro dejó 10 de las 40 casillas en blanco, Silvere Fontanet 3 de 40, Marc Bayle 6 de 40 y Salvatore De Martino 2 de 40. El tratamiento estadístico que se ha seguido en el análisis de estos datos ha sido el mismo que para el caso de las estimas obtenidas durante los vuelos, es decir, en primer lugar se ha aplicado el test de normalidad de Shapiro-Wilk. La conclusión que se obtiene de los resultados de este test (tabla 7) es que, mientras que las variables número de individuos y peso total no se ajustan de manera significativa a una distribución normal para ninguno de los observadores, en el caso de la variable peso medio los resultados son dispares; ajuste significativo a distribución normal en el caso de Silvere y Marc, y ajuste no significativo en el caso de Luis y Salvatore. Por consiguiente, a la hora de comparar las estimas entre los observadores se ha decidido utilizar de nuevo los test no paramétricos; test de rangos de Kruskal-Wallis y el test de Wilcoxon.

Tabla 7. Resultados del test de normalidad Shapiro-Wilk para las tres variables estimadas por los observadores en base a las fotográficas; número de individuos, peso medio y peso total (* =p<0.05).

	LUIS		SILVERE		MARC		SALVATORE	
	Valor w	p	Valor w	p	Valor w	p	Valor w	p
Numero individuos	0.80731	0.00009*	0.84885	0.00015*	0.55958	6.178E-09*	0.83936	0.00007*
Peso medio	0.91348	0.01591*	0.94825	0.08472	0.95921	0.23060	0.87377	0.00050*
Peso total	0.6721	4.619e-07*	0.89403	0.00202*	0.83413	0.00010*	0.90864	0.00451*

Del mismo modo que se ha procedido para los datos de los avistamientos registrados durante los vuelos y, por los mismos motivos, se ha decidido usar como valor de referencia con el que comparar el valor promedio de las estimas. De las 40 fotografías que se han usado en este ejercicio práctico 20 fueron clasificadas como de buena calidad, es decir, con un enfoque, brillo, y distribución de los atunes adecuados para poder obtener un dato verosímil del número real mediante el programa ImageJ. Así que, en el caso del número de individuos, además de realizar la comparativa tomando como referencia el valor promedio de las estimas, se realizó un segundo análisis tomando como referencia el valor obtenido mediante el programa ImageJ.

En cuanto al **número total de individuos**, el coeficiente de variación para las estimas de los 4 observadores teniendo en cuenta los resultados de las 40 fotografías utilizadas, varía entre 6.1% y 137.5%, con un valor medio del 45.5% (desviación estándar 24.6%) que indica que las muestras presentan un nivel de homogeneidad relativamente alto.

Tabla 8. Resultados del test de Wilcoxon para la variable número de individuos comparando parejas de observadores.

	valor v	lwr	upr	p
MARC-LUIS	203.5	-90	85	0.9909
SALVATORE-LUIS	140.5	-200	25	0.1544
SILVERE-LUIS	270.5	35	290	0.0157*
SALVATORE-MARC	146	-230	55	0.3014
SILVERE-MARC	370	0	290	0.0474*
SILVERE-SALVATORE	466.5	100	410	0.0038*

El resultado del test de suma de rangos de Kruskal-Wallis indica que no existen diferencias estadísticamente significativas ($p=0.224$) entre las estimaciones de los 4 observadores. Los resultados del test de Wilcoxon son diferentes en función de la pareja considerada. Mientras que en el caso de las parejas Marc-Luis, Salvatore-Luis y Salvatore-Marc las diferencias no son estadísticamente significativas, en el resto, Silvere-Luis, Silvere-Marc y Silvere-Salvatore, las diferencias son estadísticamente significativas (tabla 8).

En la figura 6 se muestra la relación entre las estimas del número de individuos correspondientes a cada observador y el valor promedio de las mismas, así como las líneas de las regresiones lineales en cada caso.

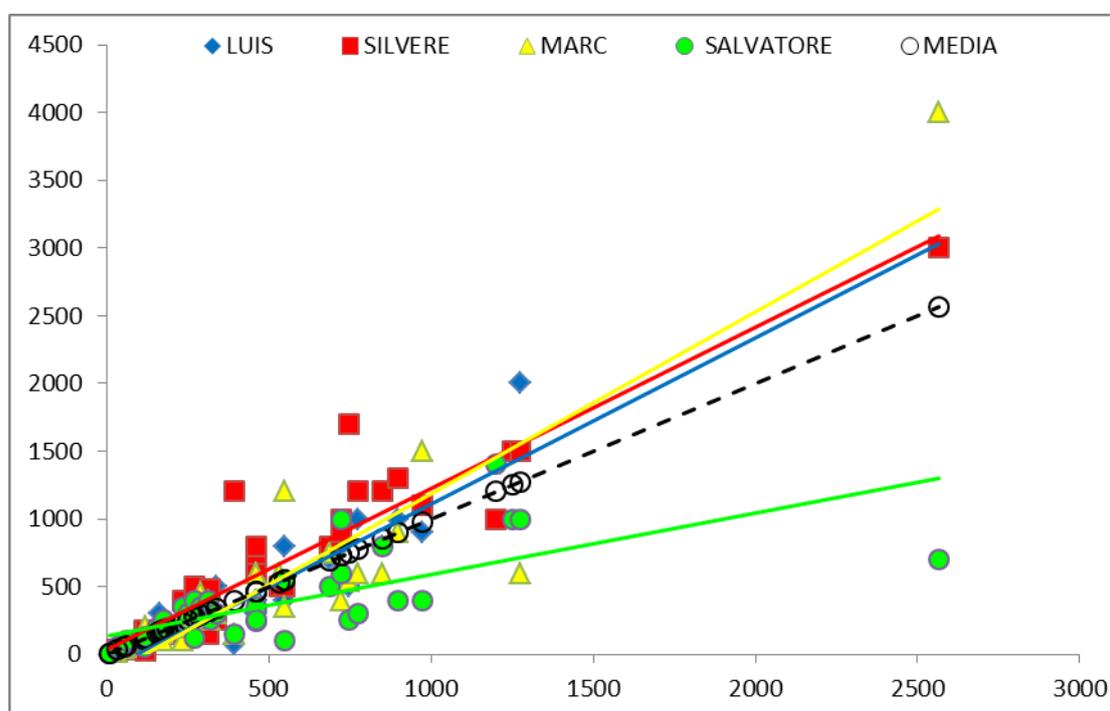


Figura 6. Relación entre las estimas del número de individuos correspondientes a cada observador y el valor promedio de las mismas. Las líneas son el resultado del análisis de regresión lineal para cada observador; azul=Luis; rojo=Silvere; amarillo=Marc y verde= Salvatore. La línea negra indica los valores de referencia.

En todos los casos los resultados de las regresiones han sido estadísticamente significativos, con valores de R^2 de 0.82, 0.87, 0.82 y 0.49 en el caso de Luis, Silvere, Marc y Salvatore respectivamente (Tabla 9).

Tabla 9. Resultados del análisis de regresión múltiple entre las estimaciones del número de individuos de los observadores y el valor medio.

	Intercepto	Pendiente	R2	p	
LUIS	-114.3934	1.224	0.816	8.16E-12	***
SILVERE	36.43424	1.191	0.867	2.20E-16	***
MARC	-156.3789	1.343	0.824	1.25E-13	***
SALVATORE	139.14727	0.451	0.491	9.77E-07	***

En base a los valores obtenidos de las pendientes de las rectas, se puede decir que 3 de los observadores, presentan rectas muy similares entre sí y con valores de las pendientes ligeramente superiores a 1 (Silvere: 1.19, Luis: 1.22 y Marc: 1.34), y Salvatore que presenta un valor de pendiente de 0.45. Generalizando se podría concluir que mientras que Luis, Silvere y Marc tienden a sobreestimar los valores de número de individuos, Salvatore tiende a subestimar.

Si tenemos en cuenta únicamente los resultados de las 20 fotografías de calidad y utilizamos los valores resultantes del programa ImageJ, el coeficiente de variación varía entre 6.1% y 67.4%, con un valor medio del 37.3% (desviación estándar 16.9%), lo que comparativamente indica un nivel de homogenización mayor que cuando se tienen en cuenta las 40 fotografías.

Tabla 10. Resultados del test de Wilcoxon para la variable número de individuos comparando con los valores obtenidos mediante el programa ImageJ.

	valor v	lwr	upr	p
LUIS-IMAGEj	8	-348	-107	0.0012*
SILVERE-IMAGEj	52	-226	89	0.4080
MARC-IMAGEj	13	-330	-46	0.0026*
SALVATORE-IMAGEj	31	-340	-14	0.0313*

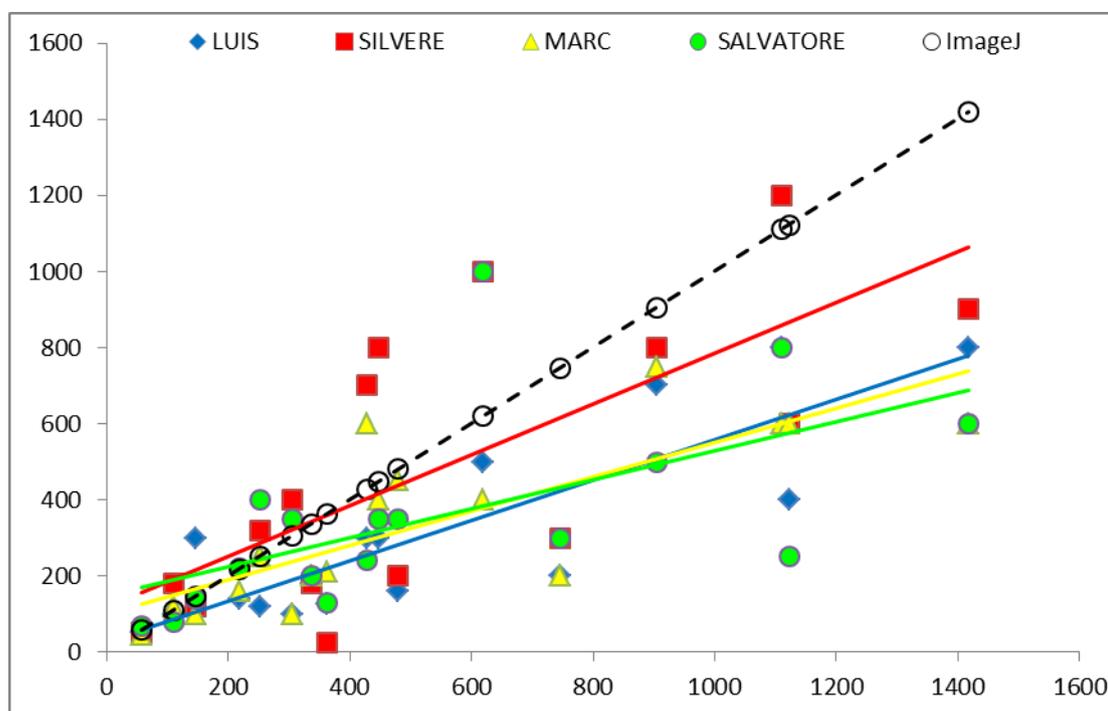


Figura 7. Relación entre las estimas del número de individuos correspondientes a cada observador y el valor obtenido mediante el programa ImageJ. Las líneas son el resultado del análisis de regresión lineal para cada observador; azul=Luis; rojo=Silvere; amarillo=Marc y verde= Salvatore. La línea negra indica los valores de referencia.

El resultado del test de suma de rangos de Kruskal-Wallis indica que no existen diferencias estadísticamente significativas ($p=0.297$) entre las estimaciones de los 4 observadores. Los resultados del test de Wilcoxon indican que, con la excepción de Silvere, existen diferencias estadísticamente significativas entre los valores estimados por Luis, Marc y Salvatore, y los valores obtenidos mediante el programa ImageJ (tabla 10).

En la figura 7 se muestra la relación entre las estimas del número de individuos correspondientes a cada observador y el valor promedio de las mismas, así como las líneas de las regresiones lineales en cada caso. En todos los casos los resultados de las regresiones han sido estadísticamente significativos, con valores de R^2 de 0.73, 0.52, 0.63 y 0.35 en el caso de Luis, Silvere, Marc y Salvatore respectivamente (Tabla 11).

Tabla 11. Resultados del análisis de regresión múltiple entre las estimaciones del número de individuos de los observadores y el valor obtenido mediante el programa ImageJ.

	Intercepto	Pendiente	R2	p	
LUIS	26.549	0.532	0.728	1.34E-05	***
SILVERE	116.8717	0.667	0.519	1.64E-03	***
MARC	100.67418	0.449	0.633	1.33E-04	***
SALVATORE	149.8446	0.379	0.355	1.16E-02	***

En base a los valores obtenidos de las pendientes de las rectas, se puede decir que cuando se comparan los valores resultantes del conteo de animales a partir de fotografías de calidad con el programa ImageJ con las estimas de los observadores, todos los observadores tienden a dar valores inferiores. Mientras que Luis, Marc y Salvatore, en general estiman de manera similar, Silvere se diferencia ligeramente de ellos y tiende a dar valores más próximos a los estimados con el programa ImageJ.

En cuanto al **peso medio**, el coeficiente de variación para las estimas de los 4 observadores varía entre 6.4% y 111.6%, con un valor medio del 34.2% (desviación estándar 24.5%) que indica que las muestras presentan un nivel de homogeneidad relativamente alto.

Tabla 12. Resultados del test de Wilcoxon para la variable peso medio comparando parejas de observadores.

	valor v	lwr	upr	p
MARC-LUIS	242.5	-2	40	0.0884
SALVATORE-LUIS	153.5	-35	17	0.3931
SILVERE-LUIS	183.5	-25	18	0.4616
SALVATORE-MARC	172	-40	2	0.0849
SILVERE-MARC	181.5	-37	2	0.0767
SILVERE-SALVATORE	305	-15	22	0.6613

El resultado del test de suma de rangos de Kruskal-Wallis indica que no existen diferencias estadísticamente significativas ($p=0.1325$) entre las estimaciones de los 4 observadores. El resultado

del test de Wilcoxon es coincidente con el test anterior e indica que no existen diferencias estadísticamente significativas entre ninguna de las comparaciones dos a dos (Tabla 12).

En la figura 8 se muestra la relación entre las estimas del peso medio correspondientes a cada observador y el valor promedio de las mismas, así como las líneas de las regresiones lineales en cada caso.

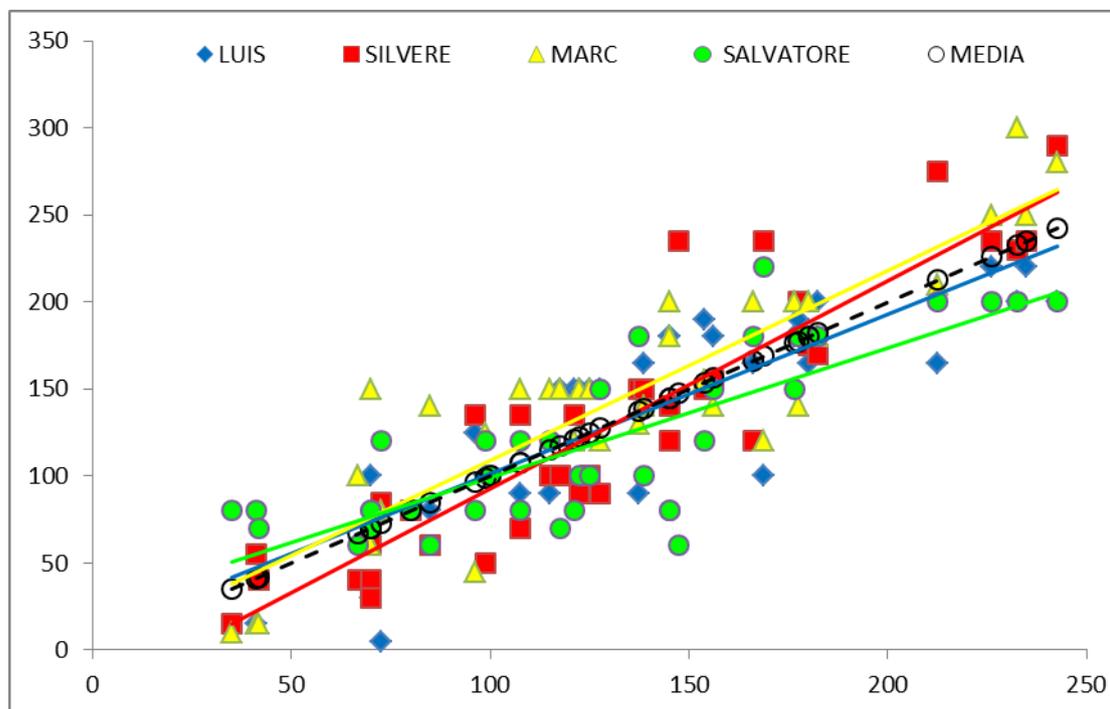


Figura 8. Relación entre las estimas de peso medio correspondientes a cada observador y el valor promedio de las mismas. Las líneas son el resultado del análisis de regresión lineal para cada observador; azul=Luis; rojo=Silvere; amarillo=Marc y verde= Salvatore. La línea negra indica los valores de referencia.

En el caso de la variable peso medio (tabla 13), a pesar de que los resultados de las regresiones han sido estadísticamente significativos para los 4 observadores. Con la excepción de Salvatore con un valor de pendiente de 0.75, en el resto de observadores los valores de las pendientes son muy cercanos a 1; Marc: 1.09, Luis: 0.92 y Silvere: 1.20.

Tabla 13. Resultados del análisis de regresión múltiple entre las estimaciones del peso medio de los observadores y el valor medio.

	Intercepto	Pendiente	R2	p	
LUIS	9.6111	0.917	0.694	6.15E-09	***
SILVERE	-26.66731	1.196	0.835	3.03E-15	***
MARC	0.08819	1.090	0.796	1.37E-12	***
SALVATORE	24.5432	0.746	0.638	1.85E-09	***

En base a los valores obtenidos de las pendientes de las rectas, se puede decir que, con la excepción de Salvatore, el resto de observadores tienen unas estimas muy similares del peso medio.

Por último, en cuanto al **peso total**, el coeficiente de variación para las estimas de los 4 observadores varía entre 8.3% y 157.6%, con un valor medio del 59.3% (desviación estándar 38.9%), que indica que esta variable es la que menos nivel de homogeneidad presenta entre todos los observadores

Tabla 14. Resultados del test de Wilcoxon para la variable peso total comparando parejas de observadores.

	valor v	lwr	upr	p
MARC-LUIS	278	-15	31	0.3493
SALVATORE-LUIS	153.5	-43	4	0.1041
SILVERE-LUIS	279	-2	33	0.0833
SALVATORE-MARC	152.5	-31	-4	0.0132*
SILVERE-MARC	312.5	-10	29	0.3643
SILVERE-SALVATORE	510.5	7	46	0.0053*

El resultado del test de suma de rangos de Kruskal-Wallis indica que no existen diferencias estadísticamente significativas ($p=0.0862$) entre las estimaciones de los 4 observadores. En cuanto a los resultados del test de Wilcoxon, salvo en el caso de las comparaciones Salvatore-Marc y Silvere-Salvatore donde las diferencias son estadísticamente significativas, con valores de p de 0.0132 y 0.0053 respectivamente, en el resto de parejas las diferencias no son estadísticamente significativas (tabla 14).

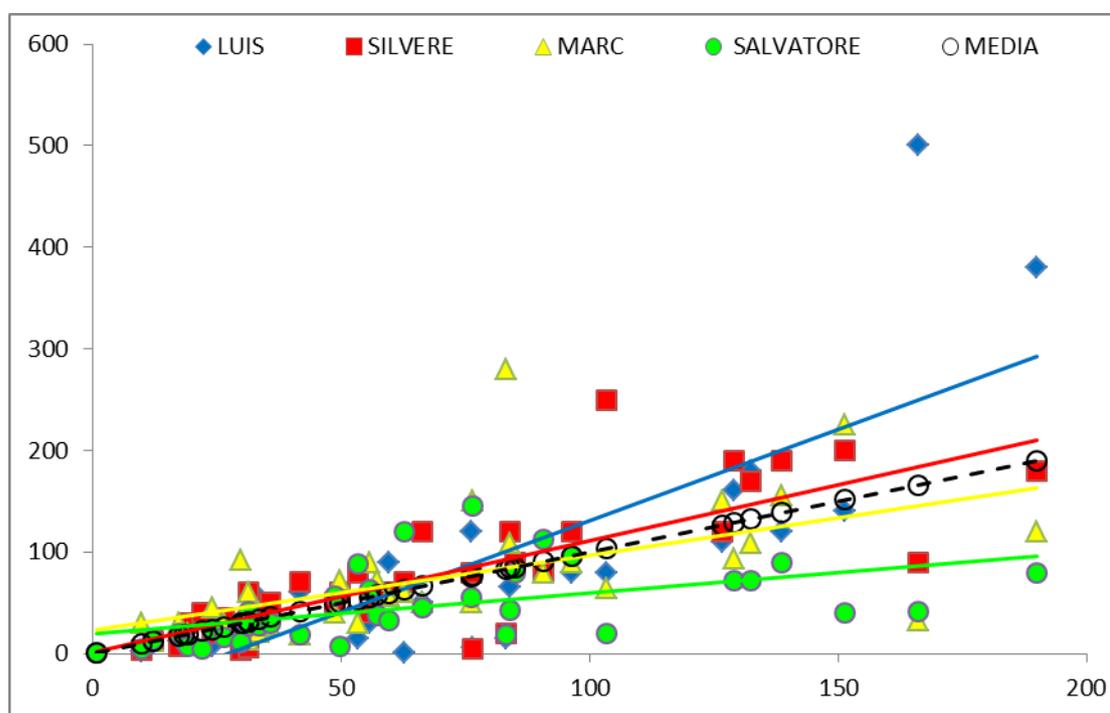


Figura 9. Relación entre las estimas de peso total correspondientes a cada observador y el valor promedio de las mismas. Las líneas son el resultado del análisis de regresión lineal para cada observador; azul=Luis; rojo=Silvere; amarillo=Marc y verde= Salvatore. La línea negra indica los valores de referencia.

En la figura 9 se muestra la relación entre las estimas del peso total correspondientes a cada observador y el valor promedio de las mismas, así como las líneas de las regresiones lineales en cada caso. En el caso de la variable peso total (tabla 15), los resultados de las regresiones han sido estadísticamente significativos para los 4 observadores. Los valores de las pendientes obtenidas para el peso total a partir de las fotografías son muy dispares, y varían desde 0.40 en el caso de Salvatore y 1.78 en el caso de Luis.

Tabla 15. Resultados del análisis de regresión múltiple entre las estimaciones del peso total de los observadores y el valor medio.

	Intercepto	Pendiente	R2	p	
LUIS	-49.2672	1.797	0.643	5.86E-08	***
SILVERE	1.4566	1.097	0.641	2.70E-09	***
MARC	23.0722	0.737	0.336	2.66E-04	***
SALVATORE	18.9605	0.404	0.278	6.73E-04	***

Generalizando, se puede decir que, Silvere es el observador que más se aproxima a los valores promedio de referencia y que mientras Marc y Silvere tienden a subestimar, Luis tiende a sobreestimar.

5.- DISCUSION

5.1.- ASPECTOS LOGÍSTICOS DEL EJERCICIO DE CALIBRACION

A pesar de la importancia que tiene el conocer las posibles diferencias que existen entre los observadores de un muestreo como el del ICCAT, y de haberse puesto de manifiesto la importancia de realizar ejercicios de calibración en los diferentes informes anuales GBYP, este es el primer intento de abordar este tema en el marco del programa “ATLANTIC-WIDE RESEARCH PROGRAMME FOR BLUEFIN TUNA” (ICCAT-GBYP Fase 9).

La primera cuestión que se debería mejorar de cara a futuros ejercicios de calibración es que estos se realicen en lo posible entre un mismo grupo de observadores expertos, con participación en un alto número de campañas a lo largo de la serie historia. Por ello, es fundamental que por parte de GBYP se haga todo lo posible para asegurar la presencia de los mismos observadores en los muestreos futuros. En este sentido hay que resaltar que no se pudo contar con Khalifa Zariohi quien contaba con 4 participaciones y en su lugar se tuvo que incorporar Marc Bayle quien, por motivos de solapamiento con el muestreo anual de junio, llegó dos días después que el resto de los observadores. Por otro lado, en este ejercicio de calibración se constató que alguno de los observadores ya tiene una edad avanzada (74 años) y que, por este motivo, puede que muy brevemente no pueda participar más en futuros muestreos.

Otra de las cuestiones que se tiene que mejorar para futuros ejercicios de calibración es la de comprobar que la avioneta destinada para realizar los vuelos es la adecuada, por este motivo, se

recomienda que en el caso de que se repita el coordinador científico viaje dos días antes a la base donde está la avioneta para asegurarse de que cumple con todos los requisitos necesarios.

Tal y como se ha planteado este ejercicio de calibración, no se cuenta con los valores reales de número de individuos, peso medio y peso total de cada uno de los cardúmenes detectados, por lo que no es posible abordar un análisis de validación. Una alternativa para contar con datos suficientes para ello sería diseñar futuros ejercicios durante el mismo tiempo en el que faena la flota cerquera y se realiza la prospección aérea científica habitual. La idea fundamental de este diseño sería la de validar los resultados de cada observador profesional en el área donde trabaja. Para llevarlo a cabo sería necesaria una coordinación entre los diferentes barcos cerqueros y el observador científico de cada avioneta para que se pudiera sobrevolar el cardumen antes de que el barco comience la maniobra de pesca. Este tipo de diseño permitiría comparar las estimas de los observadores con los datos reales a un coste relativamente bajo.

5.2.- COMPARACION DE LAS VARIABLES

Los resultados obtenidos para la variable **número total** de individuos a partir de las estimas de los avistamientos de atún rojo detectados durante los vuelos del ejercicio de calibración indican que las estimas de los observadores son relativamente homogéneas entre si y no son estadísticamente diferentes. Cuando se comparan con el valor promedio, en todos los casos las relaciones resultan significativas siendo Silvere el observador que más se aproxima a los valores de referencia. Cuando se utilizan las estimas obtenidas durante el ejercicio con fotografías, los resultados indican que los valores son más heterogéneos y que existen diferencias estadísticamente significativas entre Silvere y el resto de observadores. Precisamente es Silvere el observador que presenta valores más próximos en comparación con los valores promedio de referencia. Por último, cuando se utilizan únicamente las fotografías de calidad, el nivel de homogeneidad de las muestras se incrementa ligeramente y el único caso en el que no existen diferencias estadísticamente significativas con las estimas del programa ImageJ se produce con Silvere. En todos los casos los observadores tienden a subestimar el número de atunes presentes en las fotografías de calidad.

En resumen, se podría concluir que, independientemente de si se usan datos de avistamientos reales o datos de fotografías y de si se utiliza el valor medio de las estimas de los observadores, con el más que seguro efecto de correlación que ello implica, o los valores de las estimas obtenidas de manera independiente mediante el programa ImageJ, a falta de disponer de datos para una validación consistente, Silvere parece ser el observador más consistente y potencialmente candidato a ser tomado como referencia para la variable número de individuos en futuros ejercicios de calibración.

En cuanto al **peso medio** de los individuos, los resultados obtenidos son más dispares. A pesar de que en términos de coeficiente de variación los valores de las estimas entre los observadores son homogéneos, los test no paramétricos indican que existen diferencias significativas. En este caso, las estimas de Silvere no son estadísticamente significativas con las de Luis, pero sí lo son con respecto a las de Marc y Salvatore. Sin embargo, cuando se comparan los valores entre las estimas de los observadores y los valores promedio tomados como referencia, mientras que Silvere presenta una relación más cercana Luis se desmarca y presenta en general

valores más bajos que los de referencia. Por otro lado, cuando se utilizan las estimas obtenidas durante el ejercicio con fotografías, las diferencias entre los observadores no son significativas en ningún caso. Con la excepción de Salvatore, las relaciones entre las estimas de los observadores y los valores de referencias son muy similares y próximos entre sí.

Como conclusión, se podría decir que a los observadores les resulta más fácil coincidir en la estima del peso medio de los atunes cuando visualizan fotografías que cuando están sobrevolando un cardumen, lo cual tiene su lógica si tenemos en cuenta las dificultades para una buena visión del cardumen desde la avioneta y que los observadores tuvieron todo el tiempo necesario para estimar el peso medio durante el ejercicio con las fotografías. Es importante comentar que existen diferencias a la hora de definir los diferentes componentes del cardumen. Mientras que Silvere y Marc tienden a hacer una descripción más detallada del cardumen, Salvatore y sobre todo Luis, tienden a homogenizar y proporcionar el peso medio del componente más abundante. Estas diferencias a la hora de estimar el peso medio del cardumen hace que los valores que finalmente se han utilizado para la comparación en el caso de los observadores más detallistas provengan de la media del peso medio de todos los componentes. En cualquier caso, una vez más, y a tenor de los resultados expuestos, Silvere parece ser el observador más consistente y potencialmente candidato a ser tomado como referencia para la variable peso medio en futuros ejercicios de calibración.

Por último, en cuanto al **peso total** de los individuos, es necesario comentar que no todos los observadores obtienen este valor de la misma manera, es decir, mientras que algunos estiman el peso total de manera independiente al número de individuos y peso medio, otros utilizar la calculadora para obtener el valor final total. Luis Navarro presenta en 9 de 16 avistamientos diferencias del 2.22% de promedio entre el valor estimado y el valor calculado, Silvere Fontanet presenta en 11 de 16 avistamientos diferencias del -1.42% de promedio, Marc Bayle presenta en 4 de 10 avistamientos diferencias del -7.18% de promedio, y Salvatore De Martino presenta en 1 de 12 avistamientos diferencias del 7.04% de. Es difícil reconocer en que porcentaje estas diferencias se deben a errores en el uso de la calculadora. De las tres variables utilizadas, el peso total es la que presenta unos valores más heterogéneos y, a pesar de que cuando se analizan todas las muestras de manera conjunta no existen diferencias significativas, cuando se analizan por pares de observadores, existen diferencias significativas al menos entre Luis y Marc, Luis y Salvatore, y entre Silvere y Salvatore. Una vez más, Silvere es el observador que presenta una relación más próxima con los valores promedio de referencia. Al igual que en el caso de los datos de avistamientos reales, cuando se utilizan las estimas obtenidas durante el ejercicio con fotografías, los datos relativos al peso total son los más heterogéneos de las tres variables. En esta ocasión las diferencias significativas se producen entre Salvatore y Marc, y entre Salvatore y Silvere. De nuevo Silvere es el observador que presenta una relación más próxima con los valores promedio de referencia.

Resumiendo, el peso total de los atunes que forman un cardumen es la variable que más divergencias presenta, bien debido a las dificultades intrínsecas que lleva, al método que cada observador realiza para el cálculo, o bien una combinación de las dos. En cualquier caso, una vez más, y a tenor de los resultados expuestos, Silvere parece ser el observador más consistente y potencialmente candidato a ser tomado como referencia para la variable peso total en futuros ejercicios de calibración.

Como conclusión final, a pesar de que se han obtenido pocos datos de avistamientos y que las relaciones entre las estimas de las diferentes variables medidas comparadas para los diferentes observadores son de diferente signo y magnitud según el caso, este ejercicio de calibración aporta información relevante que ha permitido constatar la posibilidad de la realización de ejercicios de calibración entre observadores de atún rojo, y que permitirá en el futuro diseñar procedimientos más precisos para intentar obtener factores de corrección para los diferentes observadores profesionales que participan en los muestreos aéreos anuales del ICCAT-GBYP.

6.- BIBLIOGRAFIA

Eltink, A.T.G.W. 2000. Age reading comparisons. (MS Excel workbook version 1.0 October 2000) Internet: <http://www.efan.no>

Eveson JP., MV. Bravington and JH. Farley. 2018. Accounting for environmental and observer effects in estimating abundance of southern Bluefin tuna from aerial survey data. PLoS ONE 13(11): e0207790. <https://doi.org/10.1371/journal.pone.0207790>

Hammond, P. S., C. Lacey, A. Gilles, et al. 2017. Estimates of cetacean abundance in European Atlantic waters in summer 2016 from the SCANS-III aerial and shipboard surveys. Available at [https://synergy.st-andrews.ac.uk/scans3/files/2017/04/SCANS-IIIdesign-based-estimates-2017-04-28-final.pdf](https://synergy.st-andrews.ac.uk/scans3/files/2017/04/SCANS-III%20design-based-estimates-2017-04-28-final.pdf).

Gerrodette, T., W. Perryman and J. Barlow. 2012. Calibrating group size estimates of dolphins in the eastern tropical pacific ocean. ADMINISTRATIVE REPORT LJ-02-08.

Gerrodette, T., W. Perryman and C. Oedekoven. 2018. Accuracy and precision of dolphin school size estimates. Marine Mammal Science. Published ahead of print, 14 May 2018. <https://doi.org/10.1111/mms.12506>.

Lennert-Cody, C. E., S. T. Buckland, T. Gerrodette, A. Webb, J. Barlow, P. T. Fretwell, M. N. Maunder, T. Kitakado, J. E. Moore, M. D. Scott, H. J. Skaug. 2018. Review of potential line-transect methodologies for estimating abundance of dolphin stocks in the eastern tropical Pacific. CETACEAN RES. MANAGE. 19: 9–21, 2018.