SCRS/2019/179 Collect. Vol. Sci. Pap. ICCAT, 76(6): 881-891 (2020)

FROM FISHERMEN TO SCIENTIFIC TOOLS: PROGRESS ON THE RECOVERY AND STANDARDIZED PROCESSING OF INSTRUMENTED BUOYS DATA

M. Grande[1], M. Capello[2], Y. Baidai[2], J. Uranga[1], G. Boyra [1], I. Quincoces[3] , B. Orue[1], J. Ruiz[3], I. Zudaire[1],
H. Murua[1], M. Depetris[2], L. Floch[2], J. Santiago[3]

SUMMARY

In the tropical tuna purse seine fishery, the fishing efficiency and dynamics of the fleet are evolving rapidly due to the fast-technological development and the increasing use of FADs attached to echosounder buoys, which provide information on the accurate geo-location of FADs and estimation of fish biomass aggregated underneath along its trajectory. This evolution and the data gaps on FADs and buoys attached made it difficult to obtain tuna abundance indicators from purse fisheries fishing with FADs. Therefore, initiatives such as the EU funded RECOLAPE project was focused on developing of buoy data collection and pre-processing protocols to provide reliable indicators of operational buoys at sea to be used in CPUE standardization procedure. In addition, the recovery of echosounder-buoy derived data opens the possibility to developing alternative indicators of tuna biomass, to assess natural variations on target species abundance. This work presents the progress done in buoy data collection on buoys attached to FADs for filling data gaps and for the establishment of procedures for buoy data pre-processing for its use in support of stock assessment and tuna fisheries management.

RÉSUMÉ

Dans la pêcherie de senneurs ciblant les thonidés tropicaux, l'efficacité de la pêche et la dynamique de la flottille évoluent rapidement en raison du développement technologique rapide et de l'utilisation croissante des DCP avec des bouées échosondeur, qui fournissent des informations sur la géolocalisation précise des DCP et l'estimation de la biomasse des poissons regroupés sous les DCP tout au long de sa trajectoire. Cette évolution et les lacunes des données sur les DCP et les bouées attachées ont rendu difficile l'obtention d'indicateurs d'abondance des thonidés auprès des pêcheries de senneurs opérant sous DCP. Par conséquent, des initiatives telles que le projet RECOLAPE financé par l'UE étaient axées sur l'élaboration de protocoles de collecte et de prétraitement des données des bouées afin de fournir des indicateurs fiables des bouées opérationnelles en mer à utiliser dans la procédure de standardisation de la CPUE. En outre, la récupération des données obtenues des bouées-échosondeurs ouvre la possibilité de développer des indicateurs alternatifs de la biomasse des thonidés, afin d'évaluer les variations naturelles de l'abondance des espèces cibles. Ce travail présente les progrès réalisés dans la collecte des données des bouées fixées aux DCP pour combler les lacunes des données et pour l'établissement de procédures de prétraitement des données des bouées en vue de leur utilisation à l'appui de l'évaluation des stocks et de la gestion des pêcheries de thonidés.

RESUMEN

En la pesquería de cerco de túnidos tropicales, la eficiencia pesquera y la dinámica de la flota están evolucionando rápidamente debido al veloz desarrollo tecnológico y al creciente uso de DCP con boyas ecosonda, que proporcionan información sobre la precisa geolocalización de los DCP y la estimación de la biomasa de peces agregada debajo a lo largo de su trayectoria. Esta evolución y las lagunas en los datos de los DCP y las boyas colocadas en ellos hacen difícil obtener indicadores de la abundancia de atún de las pesquerías de cero que pescan en DCP. Por lo tanto, iniciativas como el proyecto RECOLAPE, financiado por la UE, se han centrado en desarrollar protocolos de recopilación y preprocesamiento de los datos de las boyas para proporcionar indicadores fiables de las boyas operativas en el mar que se utilizarán en el

---

[1] AZTI Herrera Kaia, Portualdea z/g, 20110 Pasaia, Gipuzkoa (Spain).
[2] MARBEC, Univ Montpellier, CNRS, Ifremer, IRD, Sète, France
[3] AZTI Txatxarramendi Ugartea z/g, 48395 Sukarrieta, Bizkaia (Spain).

procedimiento de estandarización de la CPUE. Además, la recuperación de datos derivados de boyas ecosonda abre la posibilidad de desarrollar indicadores alternativos de biomasa de túnidos para evaluar las variaciones naturales en la abundancia de las especies objetivo. Este trabajo presenta el progreso realizado en la recopilación de datos de boyas colocadas en DCP para llenar las lagunas existentes en los datos y para establecer procedimientos para el preprocesamiento de los datos de boyas con el fin de utilizarlos en apoyo de la evaluación de stock y la ordenación de las pesquerías de túnidos.

KEYWORDS

FADs, Instrumented buoys, purse seiner, tropical tuna

1. Introduction

The introduction of Fish Aggregating Devices (FADs) in conjunction with the satellite linked echo-sounder buoys is one of the most significant innovations introduced in the industrial tropical tuna purse seine fishery (Lopez *et al.*, 2014). These buoys provide information on the accurate geo-location of the FADs and estimation of fish biomass aggregated underneath the FAD along its trajectory, which increases the efficiency of the fishing operations. This technological development has broken the link between searching time and effective fishing effort (Torres-Irineo *et al.*, 2014) making difficult to obtain reliable catch per unit effort (CPUE) indices for tropical tunas from purse fisheries fishing with FADs (Torres-Irineo *et al.*, 2014). Given that abundance indices for tuna are derived from commercial CPUE, distinguishing between the impacts of technological innovation and natural variations in fish abundance is crucial for stock assessment (Torres-Irineo *et al.*, 2014; Katara *et al.*, 2018). In this scenario determining the operational buoys at sea and its evolution through time is essential to improve the standardization procedure and to provide alternative abundance indices (Baidai *et al.*, 2018; Santiago *et al.*, 2019).

To date, there has been significant data gap on FAD use worldwide for a science-based FAD fisheries management, while it is crucial for effort assessment and evaluation of associated impacts (Dagorn *et al.*, 2012). In this context, filling the data gaps on FADs has become a priority for RFMOs and other stakeholders which work on defining standards and procedures for FAD related data collection (Ramos *et al.*, 2017; Lopez *et al.*, 2018; Grande *et al.*, 2018). In this context, initiatives such as the EU funded RECOLAPE[4] project, aim at developing data collection and analysis procedures on instrumented buoys to provide indicators of operational buoys at sea in support of the CPUE standardization and the development of alternative abundance indices in tuna fisheries. The collaborative work conducted by the EU fishing industry, buoy providers and research institutions has allowed recovering historical information on buoy positions and acoustic data to be used by scientists for developing novel indicators for evaluating and managing tropical tuna stocks. This work presents the progress done in buoy derived data collection and presents a specific exercise developed for the establishment of procedures for buoy data pre-processing (i.e. data filtering protocol).

2. Material and Methods

2.1. Data Collection protocols and filtering criteria

Under specific data-exchange agreement signed between research organisms (i.e. AZTI and IRD) and EU tuna purse seiner associations (i.e. ORTHONGEL[5], Echebastar and Atunsa companies in ANABAC[6] and OPAGAC[7]) historical buoy positions and acoustic data has been gathered for both the Atlantic and Indian Ocean. The data has been sent from 3 buoy providers (i.e. Brand 1, Brand 2 and Brand 3) to the research institutions. In addition, buoy position data for the period 2006-2015 was directly provided to IRD by the fishing companies, under a first data-exchange agreement signed between IRD and ORTHONGEL. The position data contained details of buoy ID, information on geolocation and buoy movement such as date, hour, latitude, longitude, velocity, course (only for Brand 2 buoys and French association); information relative to the environment such as SST (for Brand 2 buoys); information relative to the functioning of the device such as the "status" (for Brand 2 buoys); battery level (for

---

[4] MARE/2016/22 "Strengthening regional cooperation in the area of fisheries data collection" Annex III "Biological data collection for fisheries on highly migratory species"

[5] Organisation française des producteurs de thon congelé et surgelé

[6] Asociación Nacional de Armadores de Buques Atuneros Congeladores

[7] Organización de Productores Asociados **de** Grandes Atuneros Congeladores

Brand 2 buoys), activation date (for Brand 2 buoys and French associations) and deactivation date; details of the ownership (in case of the Spanish associations), or vessels receiving the information (in case of Brand 2 and French associations). The acoustic data contained information on gain, frequency, resolution, intensity of the acoustic signal by layer (i.e. 50 layers) in the case of the Brand 2 buoys, biomass estimate by layer (i.e. 10 layers) in the case of Brand 1 buoys and biomass estimate in case of Brand 3 buoys. Note that the indicators of tuna biomass are different among buoy providers.

To develop common indicators of the number of operational buoys at sea and biomass from acoustic signals, the raw data need to be pre-processed for filtering erroneous location, data related to failures in satellite communication and location data acquisition; identifying buoys on land positions; and identifying buoys data recording on-board positions. In order to compare the performance of different methods used in AZTI and IRD and agree on a common method for data pre-processing, a common EU database was created and shared, integrating the position data recorded by 2000 buoys during 1 month in the Atlantic and Indian Ocean (i.e.,1,000 buoys from the Spanish and 1,000 buoys from the French fleet for each ocean). Based on the work of Baidai *et al.* (2017), the filters described in the **Table 1** were defined and applied in the common data base.

The F1, F2 and F4 filters were applied using the same data processing protocol detailed in **Table 1** for both organisms. Land position (filter F3) were assigned considering a low-resolution shoreline from GSHHG[8] buffered with 0.05° shapefile (IRD) or a high-resolution shoreline from GSHHG[4] buffered with 0.05° shapefile (AZTI). In order to filter the data on-board (F6) IRD applied the kinetic algorithm described in Baidai *et al.* (2017), which is based on the analysis of buoys speed and acceleration along the buoy trajectory, see **Figure 1** and Baidai et al 2017 for further details.

On the other hand, AZTI applied a random forest classification approach to classify the buoys at sea/onboard using information from the Brand 3 buoys, which have the capability to identify true positions at sea through a conductivity sensor. The sensor measures the ionic content between two electrodes and determines, through a simple algorithm, whether the buoy is in the water. The predictors variables used in the RF analysis were: distance between two points (km), velocity (km/h), change in velocity (km/h), acceleration (km/h2), azimuth (degree), change in azimuth (degree) and time since the first and last observation of the corresponding buoy trajectory (days) (**Fig. 2**, see Orue *et al.*, 2019 for further details).

Both of these classification algorithms can leave a subset of positions unclassified (e.g., for short trajectories). It was agreed that the unclassified positions should not be eliminated from the dataset and accounted as buoys “at water”, to avoid underestimating the buoys at sea. The final comparisons of the performance of the algorithms for classifying the buoys at water were carried out through the calculation of simple matching coefficient (Sokal and Michener, 1958), estimated from confusion matrices derived from the outputs of the two classification methods.

3. Results

3.1. Historical information on buoys

Information on three buoy brands has been gathered in the Atlantic and Indian Ocean covering the period from 2006 to 2018 in the case of buoys used by ORTHONGEL fleet and 2010 to 2018 in the case of buoys used by ANABAC (i.e. Atunsa and Echebastar companies) and OPAGAC fleet.

Regarding to the information coming from Spanish associations two data bases has been created, one relative to individual buoy GPS positions (one position per day has been stored referred to the last position of the day) and the other relative to the acoustic information including all sounding per day (in case of Brand 2, soundings also include details of transmission position). The qualitative description of the position and acoustic databases is included in the **Figure 3** and **Figure 4**, respectively. For each buoy brand various models are used which has been grouped considering acoustic specifications: absence or presence of the echosounder and the echosounder frequency (i.e. 1 frequency or 2 frequencies). Vessels within Spanish Associations work with 3 buoy Brands. In the 2010-2012 period the information of individual positions for Brand 1 could be obtained, but information on individual buoy positions of Brand 2 could not be exported and integrated in the database, due to a technical limitation in data exportation process. From 2013 to 2018 information on all buoys Brands could be integrated in the database: Brand 1 (including 6 different models), Brand 2 (including 5 models) and Brand 3 (including 5

---

[8] Wessel, P., and W. H. F. Smith (1996), A global, self-consistent, hierarchical, high-resolution shoreline database, J. Geophys. Res., 101(B4), 8741–8743, doi:10.1029/96JB00104.

models). At the beginning of the series about 50% of the buoys were working with echosounder. This percentage has increased gradually and nowadays all buoys deployed by the fleet have echosounder. Nowadays, buoys used by the Spanish fleet in both oceans work mainly with a unique frequency, but since 2016 the two-frequency echosounder buoys are also used.

On the other hand, French tuna purse seiner association ORTHONGEL mainly works with Brand 2 buoys. In this case high resolution information is received, i.e. all positions available per day. The information on buoys is gathered in two data bases (i.e. position data base and acoustic data base) at IRD. The raw position database contains information on the buoys deployed by the French fleet between 2006 and 2015 (**Fig. 5**). For the following years, the information on the buoy position data was part of the acoustic database (**Fig. 6**). Since 2010, echosounder buoys replaced non-echosounder buoys gradually to 2015 in which all buoys attached to FADs were equipped with echosounder (**Fig. 5** and **Fig. 6**). In recent years (2016-2018), in the Atlantic Ocean, echosounder buoys working with 1 frequency are being replaced by buoys working with two frequencies. On the other hand, in the Indian Ocean, the French fleet continued working principally with echosounder buoys working with 1 frequency (**Fig. 6**).

3.2. Buoy Filtering outputs on the common data set

The above-described data filtering procedures were applied on the common dataset composed by 2,000 buoy tracks from the EU fleet by ocean. Results for the Atlantic and Indian ocean are presented in **Table 2** and **Table 3**, respectively.

The inspection of the outputs of the filtering algorithm run by IRD and AZTI on the common database demonstrated a high rate of agreement between the two algorithms. The main differences occurred in the land classification. The shapefile resolution could impact the filtering of land positions. In addition, minor differences among the two methods occurred in the number of buoy classified as on-board. These differences were higher for the Spanish dataset, since the performances of the algorithms are affected by the resolution of the databases (i.e. one position per day in the Spanish database relative to high-resolution position data in the French database).

The on-board classification algorithms leave a subset of positions unclassified. In the case of the algorithms developed by AZTI it refers to the first position of the track of the buoy. In the case of the kinetic algorithm of IRD unclassified positions are both due to the presence of short trajectories and, for the Spanish dataset, to the low resolution of the data. For these classification algorithms that leave a subset of positions unclassified, it was agreed that the unclassified position should not be eliminated from the dataset and included as buoys at water (i.e., buoys with unclassified positions will be still considered as buoys “at water”).

The final comparisons of the performance of the algorithms for classifying the buoys at water were carried out through the calculation of simple matching coefficient (Sokal and Michener, 1958), estimated from confusion matrices derived from the outputs of the two classification methods. Results are included in **Table 4, Table 5, Table 6**, and **Table 7**. Overall, the two methods show high matching coefficients (>94%) in all oceans and datasets. In the Atlantic Ocean, the performances of the classification protocol by IRD and AZTI to classify the buoys at water are >96%. The weaker agreement (94%) is observed in the Indian Ocean in the Spanish data set, possibly due to the characteristics of this data set with shorter tracks and lower temporal resolution (i.e. a position per day).

4. Conclusions and Recommendations

The collaborative work conducted by the fishing industry, buoy providers and research institutions has allowed recovering historical information on buoy positions and acoustic data to be used by scientists for developing novel indicators for evaluating and managing tropical tuna stocks. The access to the data has been obtainned thanks to specific agreements with the data owners. IRD has recovered and integrated in the database the data related to the Brand 2 buoys used by the ORTHONGEL fleet since 2006 to 2018. In the case of AZTI information of ANABAC (i.e. Echebastar and Atunsa) and OPAGAC fleeet covering 2010-2018 period has been obtained and integrated in the database. Some buoy providers faced difficulties when exporting historical data, therefore, in the future in order to progress with the recovery of information on buoys, periodical deliveries would be a potential solution.

In this specific exercise, for the analysis of data filtering protocols and the agreement of a common protocol for buoy data pre-processing, a wide set of filters has been defined and tested using a common database. Filters run in each research institute were identical except the shapefile for land and onboard filtering, for which a specific algorithm was developed by each institute. The outputs of the filtering algorithms run by each research center show high rates of agreement (>94% agreement on buoys labeled as "in water"), validating both methods for data pre-processing. Minor differences occur on land and on-board positions, for which a specific algorithm has been developed by each research center. These differences were higher for the Spanish dataset in the Indian Ocean, since the performances of the algorithms are affected by the characteristics of the databases (i.e. lower performance on shorter tracks and lower temporal resolution). In this sense, in order to minimize the misclassification, the use of high-resolution data is recommended if available. In addition, some factors were identified to be valuable to further improve the filtering and to evaluate the number of buoys followed by each vessel:

- Water temperature
- IMO of the vessels receiving the buoy information
- Activation and deactivation date
- Mode of the buoy
- High-resolution data (all the positions in a day)

Moreover, the addition of an on-board/at sea sensor to the buoys would be a technical improvement that could improve the quality of the data.

Nowadays this information is beign shared voluntarily by the EU purse seiner industry under specific data use agreements. Considering the potential of this data for effort assessment, as well as for an alternative indicator of target species abundance (Baidai *et al.*, 2018; Santiago *et al.*, 2019) and for the evaluation of indicators defined in the frame of ecosystem report card (Zudaire *et al.*, 2019), buoy derived information should be made available for scientific use.

Acknowledgements

We acknowledge the RECOLAPE project (MARE/2016/22 "Strengthening regional cooperation in the area of fisheries data collection" Annex III "Biological data collection for fisheries on highly migratory species"), the fishing companies (ORTHONGEL, Echebastar and Atunsa companies in ANABAC and OPAGAC) and buoy providers (Satlink, Marine Instruments and Zunibal) that shared the buoy data.

References

Baidai, Y., Amande M.J., Gaertner D., Dagorn L. Capello M. 2018. Recent advances on the use of supervised learning algorithms for detecting tuna aggregations under FADs from echosounder buoys data. IOTC-2018-WPTT20-25_Rev1

Baidai Y, Capello M, Amandé J, Billet N, Floch L, Simier M, Sabarros, P and Dagorn L. Towards the derivation of fisheries-independent abundance indices for tropical tunas: progress in the echosounders buoys data analysis, 2017. 19th IOTC WPTT, Victoria, Seychelles, 17-22 Oct 2017. IOTC-2017-WPTT19-22

Dagorn. L., Holland, K.N., Restrepo, V., Moreno, G., 2012. Is it good or bad to fish with FADs? What are the real impacts of the use of drifting FADs on pelagic marine ecosystems? Fish and Fisheries 14:391-415.

Grande M, Baez J.C., Ramos M.L., Ruiz J., Zudaire I., Murua H., Santiago J., Pascual P., Abascal F., Gaertner D., Cauquil P., Floch L., Maufroy A., Muniategi A., and Herrera M., 2018. Best standards for data collection and reporting requirements on FOBs: towards a science-based FOB fishery management Collect. Vol. Sci. Pap. ICCAT, 75(7): 2259-2282 (2019).

Katara I., Gaertner D., Marsac F., Grande M., Kaplan D., Urtizberea A., Guéry L., Depetris M., Duparc, A., Floch L., Lopez L. and Abascal, F., 2019. Standardisation of yellowfin tuna CPUE for the EU purse seine fleet operating in the Indian Ocean. IOTC–2018–WPTT20–36_Rev1

Lopez, J., Moreno, G., Sancristobal, I., and Murua, J. 2014. Evolution and current state of the technology of echo-sounder buoys used by Spanish tropical tuna purse seiners in the Atlantic, Indian and Pacific Oceans. Fisheries Research, 155: 127–137.

Lopez, J., Altamirano, E., Lennert-Cody, C., Maunder, M., Hall, M., 2018. Review of IATTC resolutions C-16-01 and C-17-02: available information, data gaps, and potential improvements for monitoring the fad fishery. 3rd Meeting of the Ad Hoc Working Group on FADs La Jolla, California USA, 11-12 May 2018

Orue, B., Lopez, J., Moreno, G., Santiago, J., Boyra, G., Uranga, J., Murua, H., 2019. From fisheries to scientific data: A protocol to process information from fishers' echo-sounder buoys. Fisheries Research 215 (2019) 38–43

Ramos, Mª. L., Báez, J.C., Grande, M., Herrera, M.A., López, J., Justel, A., Pascual, P.J., Soto, M., Murua, H., Muniategi, A., Abascal, F.J., 2017. Spanish FADs logbook: solving past issues, responding to new global requirements. Joint t-RFMO FAD Working Group meeting Doc. No. j-FAD_11

Santiago, J., Uranga J., Quincoces I., Orue B., Grande M.., Murua H., Merino M., Urtizberea A., Pascual P., Boyra G., 2019. A novel index of abundance of juvenile yellowfin tuna in the indian ocean derived from echosounder buoys.

Sokal, R.R. and Michener, C.D. I958. A statistical method for evaluating systematic relationships. University of Kansas Science Bulletin, 38: 2.

Torres-Irineo, E., Gaertner, D., Chassot, E., Dreyfus-León, M., 2014. Changes in fishing power and fishing strategies driven by new technologies: The case of tropical tuna purse seiners in the eastern Atlantic Ocean. Fisheries Research, 155: 10-19.

Zudaire I., Grande M., Murua H., Ruiz I., Juan-Jordá, M.J., 2019. In support of the IOTC ecosystem report card: indicators for marine debris. IOTC-2019-WPEB15-50

Table 1. Filters defined for pre-processing raw position data.

FILTER	Description
F1. Isolated	Isolated Position (>48 hours from another position or estimated spee above > 35 knots relative to next/previous position)
F2. Duplicated	Duplicated data (all fields are the same)
F3. Land	Data on land
F4. Ubiquity	Data entry having from the same date/time different positions
F5. Not classified	Position not in the land and not classified by the at sea/on board algorithm
F6. Onboard	Buoys on board
F7. Water	Buoys at sea. Operational buoys: *Active buoy that is transmitting a signal and is drifting in the sea (definition from RECOLAPE)*

Table 2. Results from the application of the filtering protocol by AZTI and IRD in the atlantic ocean common data base. The table includes the number (n) and percentage (%) of records filtered in buoys coming from French associations (FR) and Spanish Associations (ESP), for the filtering made by IRD and AZTI.

Filters	n FR_ATL (AZTI)	% FR_ATL (AZTI)	n FR_ATL (IRD)	% FR_ATL (IRD)	n ESP_ATL (AZTI)	% ESP_ATL (AZTI)	n ESP_ATL (IRD)	% ESP_ATL (IRD)
F1.Duplicated	47	0.1	47	0.1	0	0.0	0	0.0
F2. Isolated	38	0.1	38	0.1	80	0.3	91	0.4
F3. LAND	4,915	7.8	4,595	7.3	325	1.3	232	0.9
F4. Ubiquity	11	0.0	11	0.0	0	0.0	0	0.0
F6. Onboard	2,746	4.4	3,473	5.5	122	0.5	4	0.0
F7 and F5 Water (and unclassified)	55,145	87.7	54,738	87.0	24,777	97.9	24,977	98.7
TOTAL	62,902	100.0	62,902	100.0	25,304	100.0	25,304	100.0

Table 3. Results from the application of the filtering protocol by AZTI and IRD in the indian ocean common data base. The table includes the number (n) and percentage (%) of records filtered in buoys coming from French associations (FR) and Spanish Associations (ESP), for the filtering made by IRD and AZTI.

Filters	n FR_ATL (AZTI)	% FR_ATL (AZTI)	n FR_ATL (IRD)	% FR_ATL (IRD)	n ESP_ATL (AZTI)	% ESP_ATL (AZTI)	n ESP_ATL (IRD)	% ESP_ATL (IRD)
F1.Duplicated	94	0.154	94	0.154	0	0	0	0
F2. Isolated	46	0.075	48	0.078	154	0.69	174	0.775
F3. LAND	2,352	3.844	1,574	2.572	333	1.48	138	0.614
F4. Ubiquity	11	0.018	11	0.018	86	0.38	149	0.663
F6. Onboard	595	0.972	496	0.811	971	4.32	14	0.062
F7 and F5 Water (and unclassified)	58,096	94.9	58,971	96.4	20,917	93.1	21,986	97.9
TOTAL	61,194	100.0	61,194	100.0	22,461	100.0	22,461	100.0

Table 4. Confusion matrix on AZTI´s filtering and IRD filtering on the Spanish buoys in Atlantic Ocean. Simple matching coefficient = 0.991;

	IRD	
AZTI	water	not water
water	24,764	13
not water	213	314

Table 5. Confusion matrix on AZTI´s filtering and IRD filtering on the French buoys in Atlantic Ocean. Simple matching coefficient= 0.96

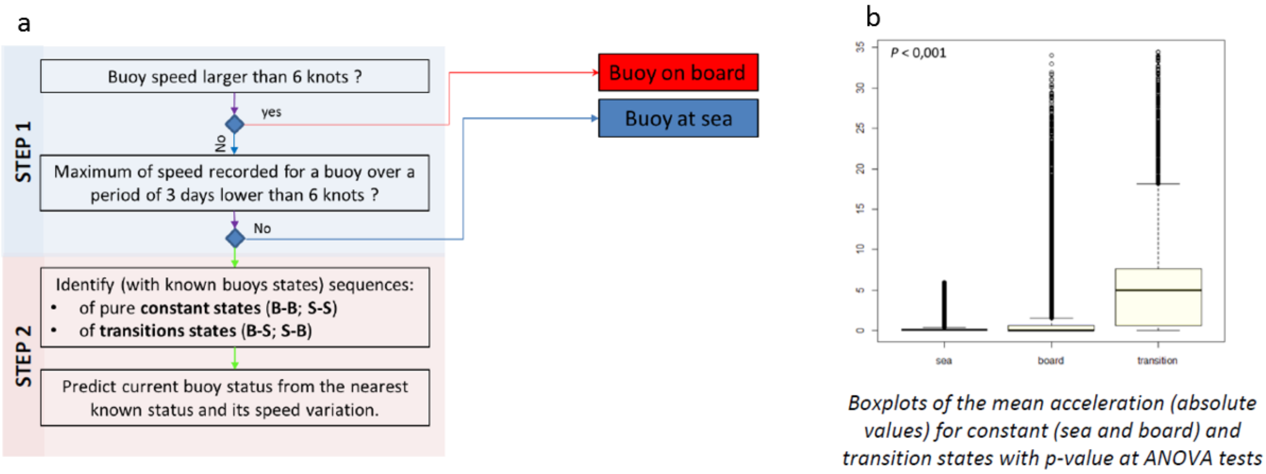
	IRD	
AZTI	water	not water
water	53,735	1,457
not water	1,061	6,649

Table 6. Confusion matrix on AZTI´s filtering and IRD filtering on the Spanish buoys in Indian Ocean. Simple matching coefficient= 0.9435

	IRD	
AZTI	water	not water
water	20,892	25
not water	1,245	299

Table 7. Confusion matrix on AZTI´s filtering and IRD filtering on the French buoys in Indian Ocean. Simple matching coefficient= 0.9742

	IRD	
AZTI	water	not water
water	57,843	347
not water	1,233	1,771



Boxplots of the mean acceleration (absolute values) for constant (sea and board) and transition states with p-value at ANOVA tests

Figure 1. Description of the kinetic classification algorithm (a) and differences in mean acceleration of constant sequences and transition sequences (b).

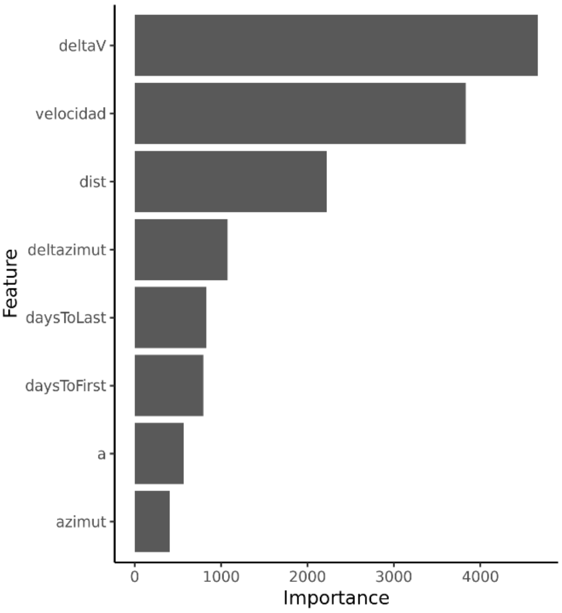


Figure 2. Variable Importance of the Random Forest Model. Name "deltaV" is the change in velocity, "velocidad" is the velocity, "dist" is the spatial distance between two points, "deltaazimut"is the change in azimuth, "daysToLast" is the time since the last observation, "daysToFirst" is the time since the first observation, "a" is the acceleration and "azimuth" is the azimuth. The average validation indices for sensitivity (i.e. 0.99), specificity (i.e. 0.89), Kappa (i.e, 0.87) and Area Under the Curve (AUC) (i.e, 0.94) were estimated to evaluate the performance effectiveness and efficiency of the RF classification.

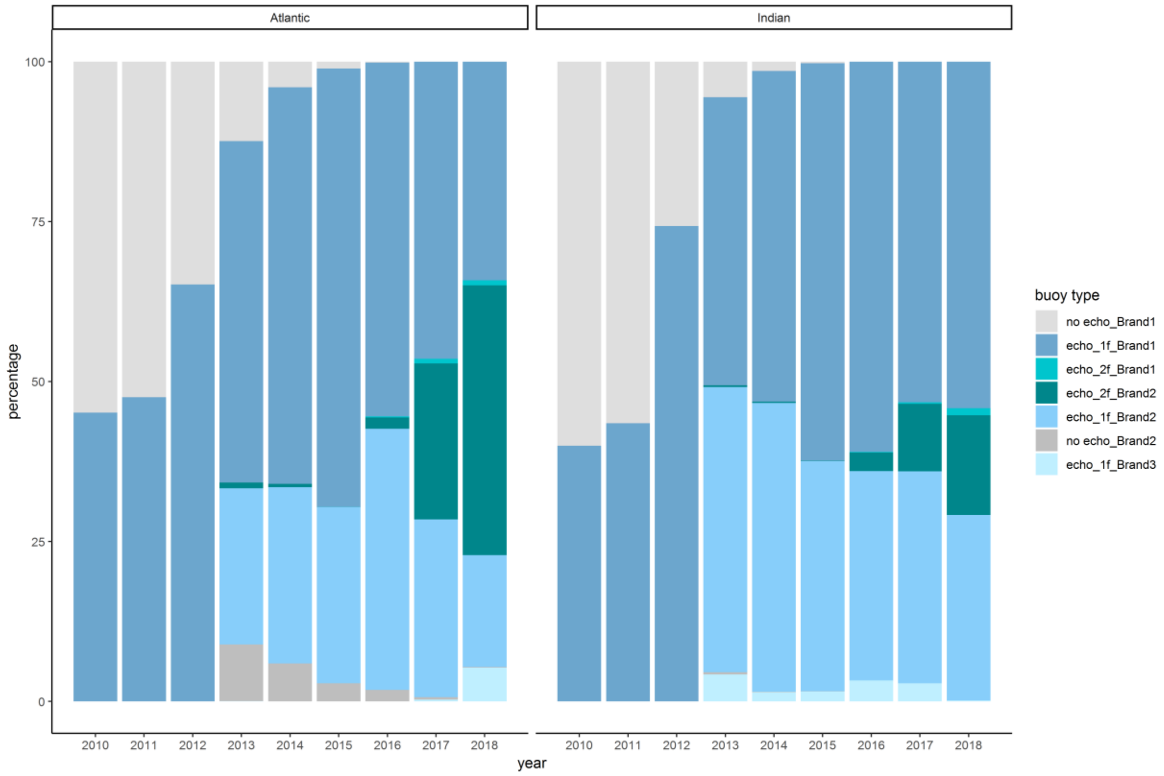


Figure 3. Percentage of buoy type and year constituting the raw Spanish Tuna Associations´ position database for the Atlantic and Indian Ocean from 2010-2018. Its of the buoy type categories is constituted by various buoy models Note that for the period 2010 to 2012 Brand 2 individual buoy positions could not be obtained.

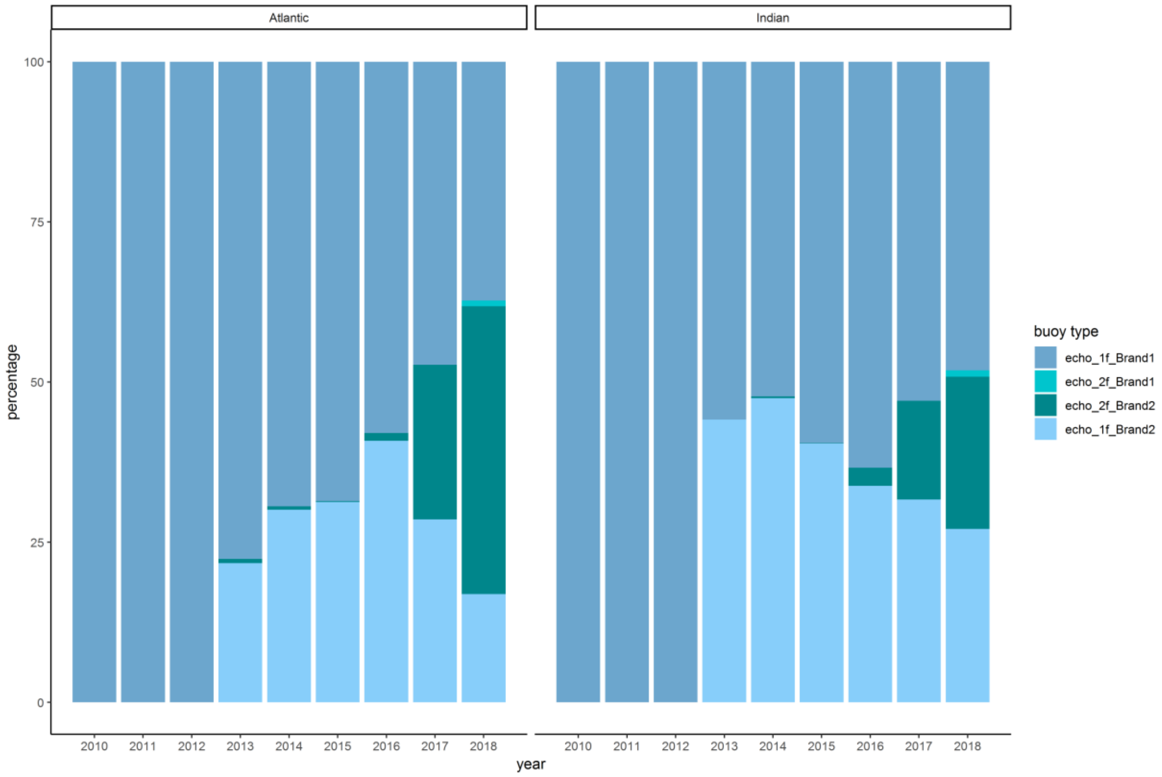


Figure 4. Percentage of buoy by type and year constituting the raw Spanish Tuna Associations´ acoustic database for the Atlantic and Indian Ocean from 2010-2018. Its of the buoy type categories is constituted by various buoy models. Note that for the period 2010 to 2012 acoustic information on Brand 2 could not be obtained.

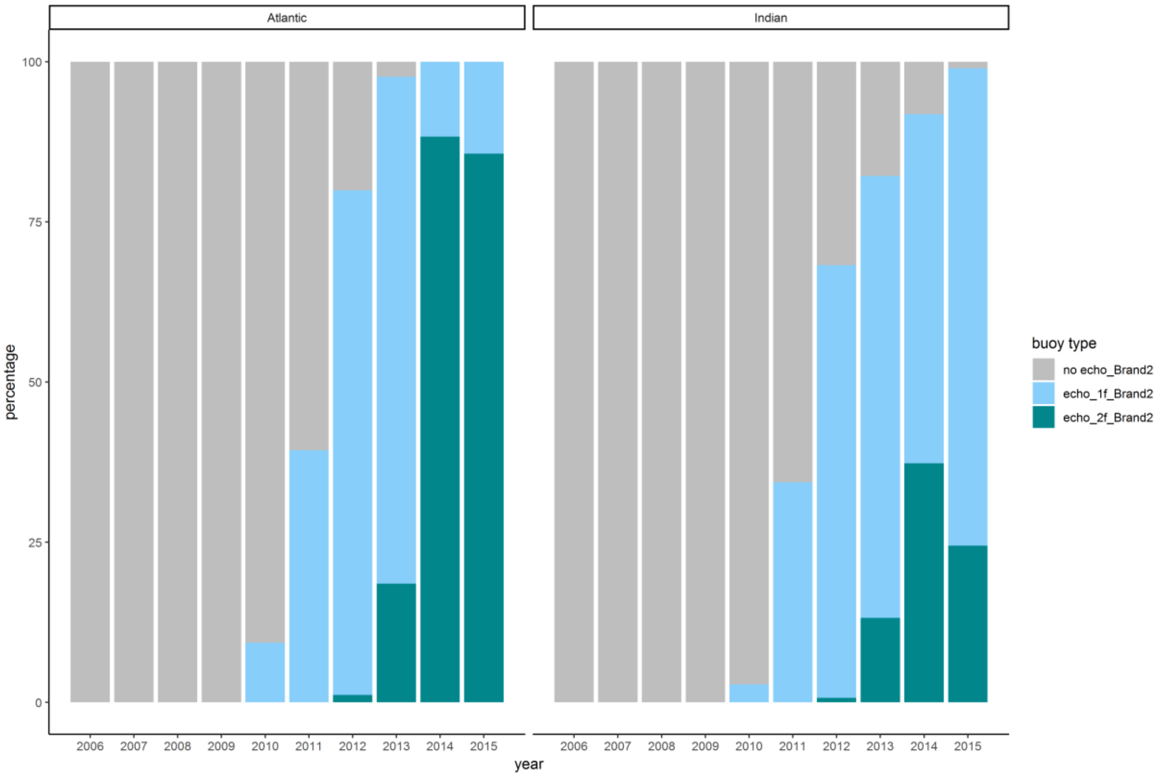


Figure 5. Percentage of buoys by type and year constituting the raw French Tuna Associations´ position database for the Atlantic Ocean (left panel) and Indian Ocean (right panel). Its of the buoy type categories is constituted by various buoy models.

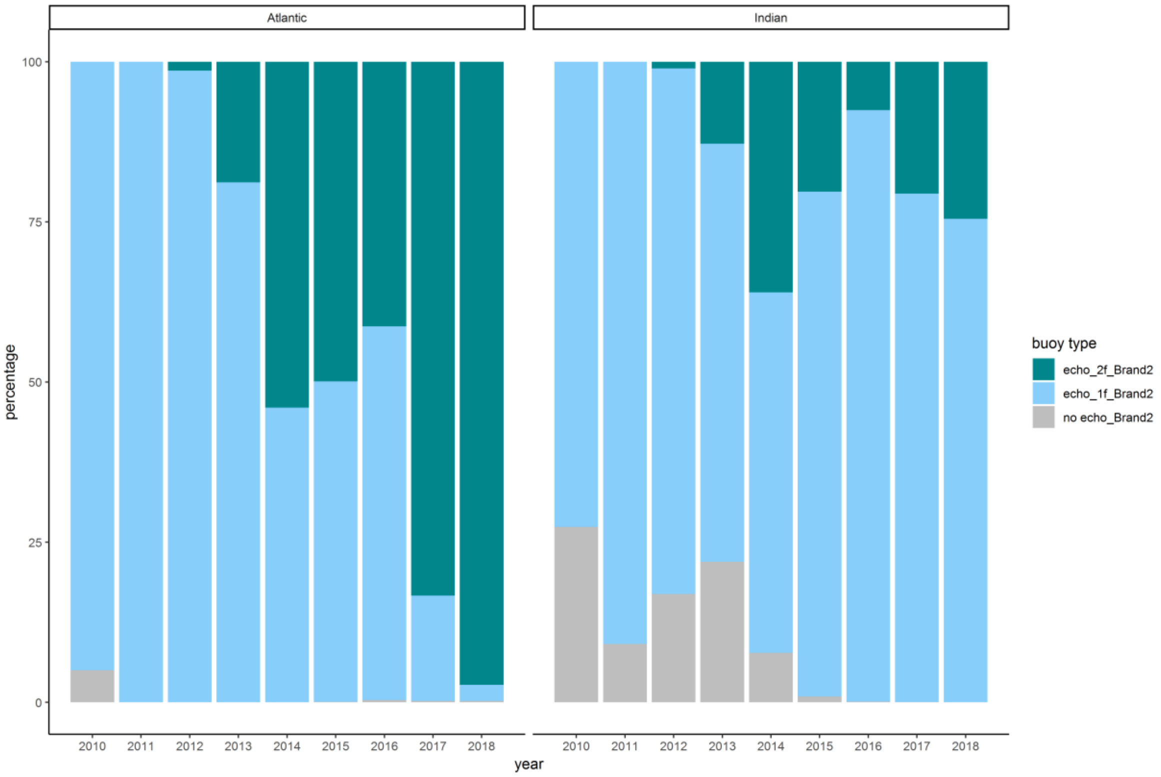


Figure 6. Percentage of buoys by buoy type and year constituting the raw French Tuna Associations´ acoustic database for the Atlantic Ocean (left panel) and Indian Ocean (right panel). Its of the buoy type categories is constituted by various buoy models.