From fishermen´ to scientific tools: Progress on the recovery and standardized processing of echosounder buoys data

Maitane Grande, Manuela Capello, Jon Uranga, Yannick Baidai, Guillermo Boyra, Iñaki Quincoces, Blanca Orue, Jon Ruiz, Iker Zudaire, Hilario Murua, Mathieu Depetris, Laurent Floch, Josu Santiago

Objective

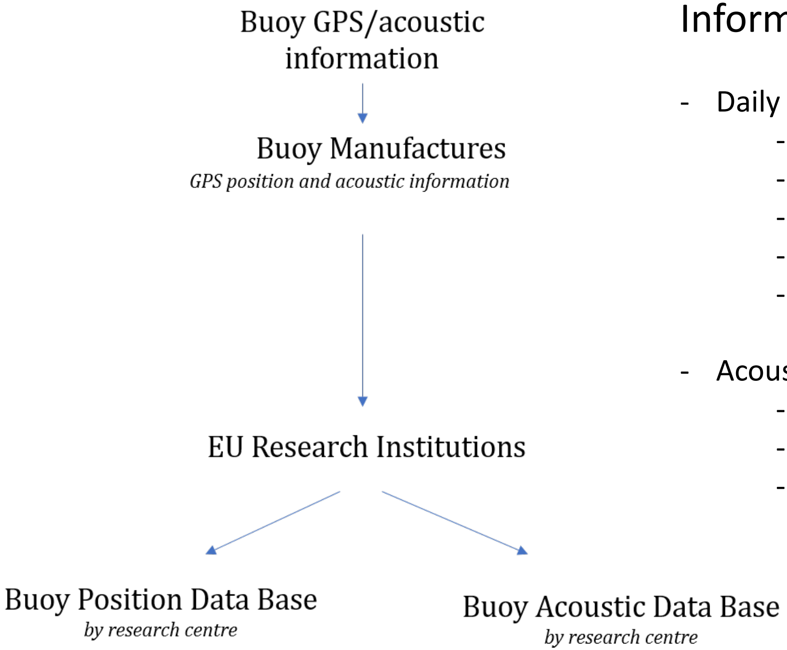
General Objective:

Develop a data collection strategy of information on buoys to be used for CPUE standardization, and in alternative abundance indices in tuna fisheries

Specific Objectives:

- Gathering information on buoys, and description of the buoys databases provided to research centers from the EU purse-seine fleet (ORTHONGEL, ANABAC and OPAGAC) in the Atlantic and Indian Ocean
- Definition of a common data pre-processing protocol for buoy information

Data gathering on buoys: Buoy data collection protocol

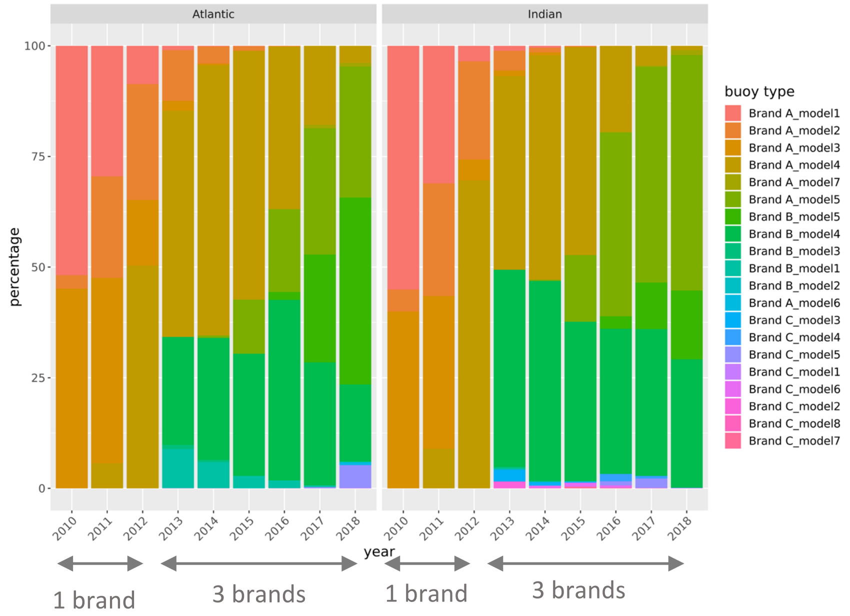


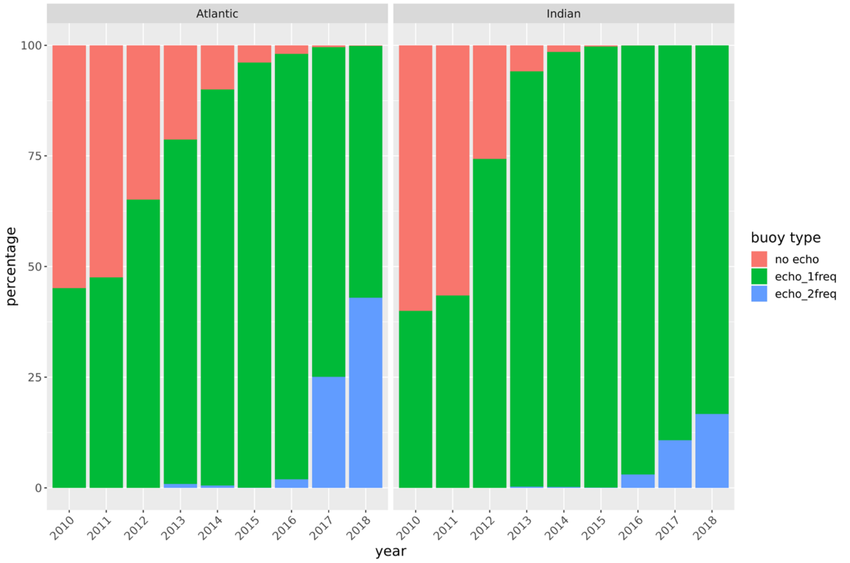
Information collected:

- Daily positions – AZTI (one per day), IRD (all available per day)
  - Lat and long
  - Hour for each position (GMT)
  - Velocity for each position
  - Temperature (IRD)
  - Vessels receiving the information (IRD)

- Acoustic record of each sounding
  - Brand A: Biomass by layer (11 layers) – max per layer
  - Brand B: Intensity of the acoustic signal by layer (50 layers) – max per layer
  - Brand C: Total biomass

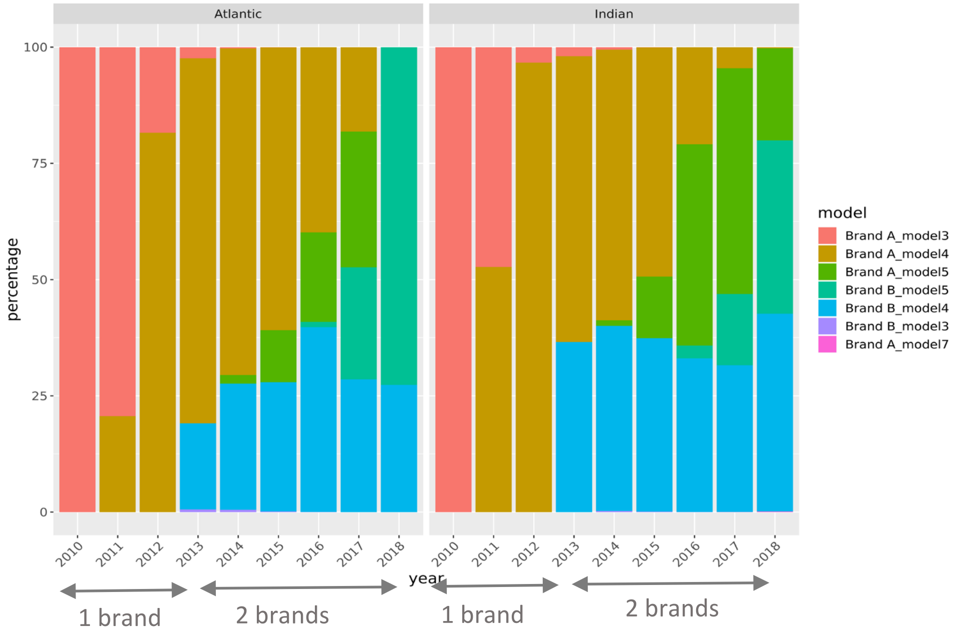
Data gathering on buoys: **Buoy position DataBase** in AZTI - EU Spain





- Information on Brand A (2010-2018), Brand B and Brand C (2013-2018)
- From 2012 the echosounder buoys replaced progressively the non-echosounder buoys (in 2016 almost all buoys with echosounder).
- One frequency buoys are mainly used, although from 2016 the used of the two-frequency buoy start increasing.

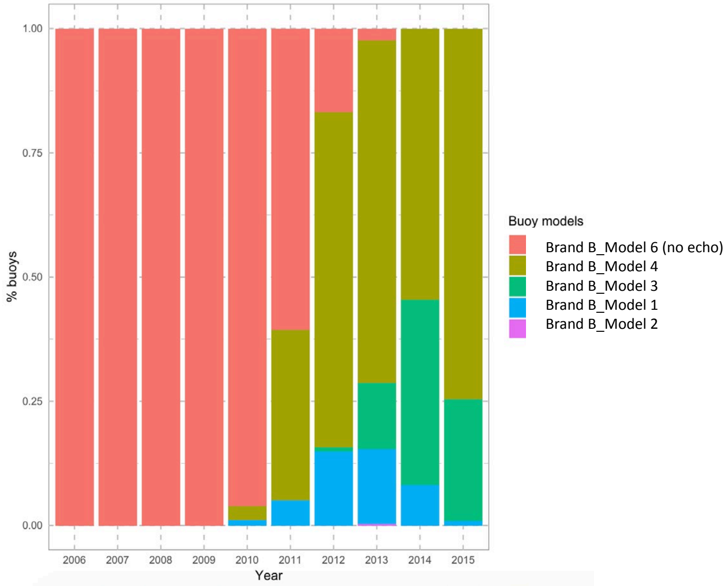
Data gathering on buoys: **Acoustic DataBase** in AZTI – EU Spain



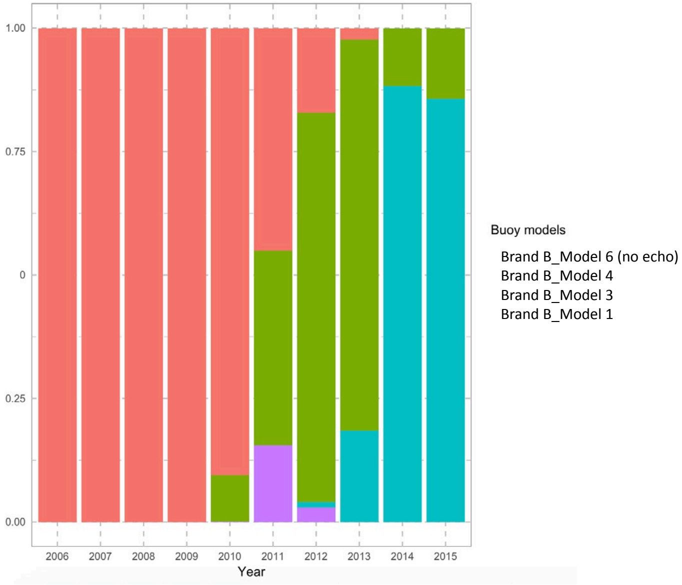
- Information on Brand A (2010-2018), Brand B (2013-2018)
- The gathering process of 2018 is on-going.
- Heterogenity in the acoustic information, given the characteristics of different buoy brands and models

Data gathering on buoys: **Buoy Position DataBase** IRD –EU France

FR: Indian Ocean



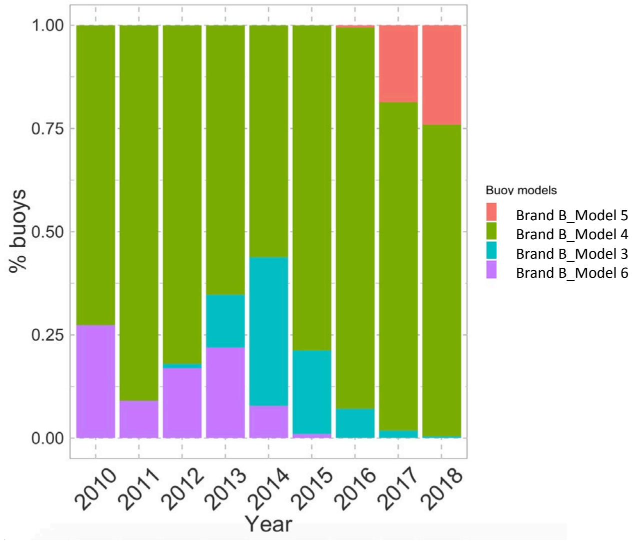
FR: Atlantic Ocean



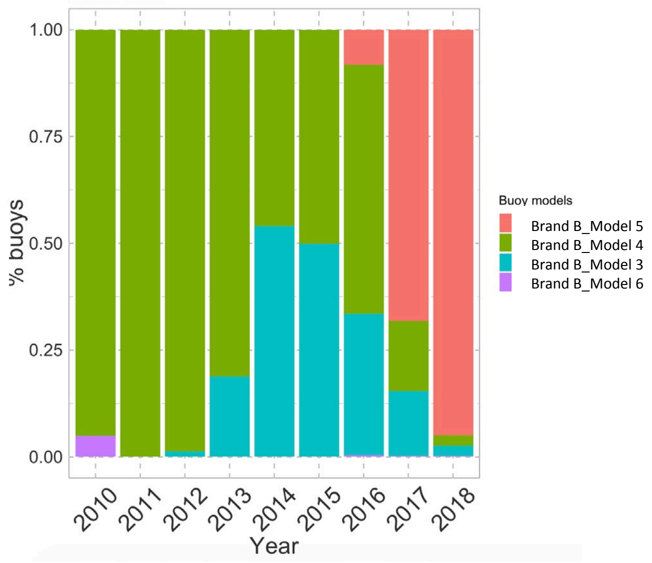
- All buoys are from Brand B
- From 2014, all buoys have echosounder.
- For years after 2015, the positions of all the echosounder buoys deployed by the French fleet are included in the acoustic database (see next slide)

Data gathering on buoys: **Acoustic Raw DataBase** in IRD - EU France

FR: Indian Ocean

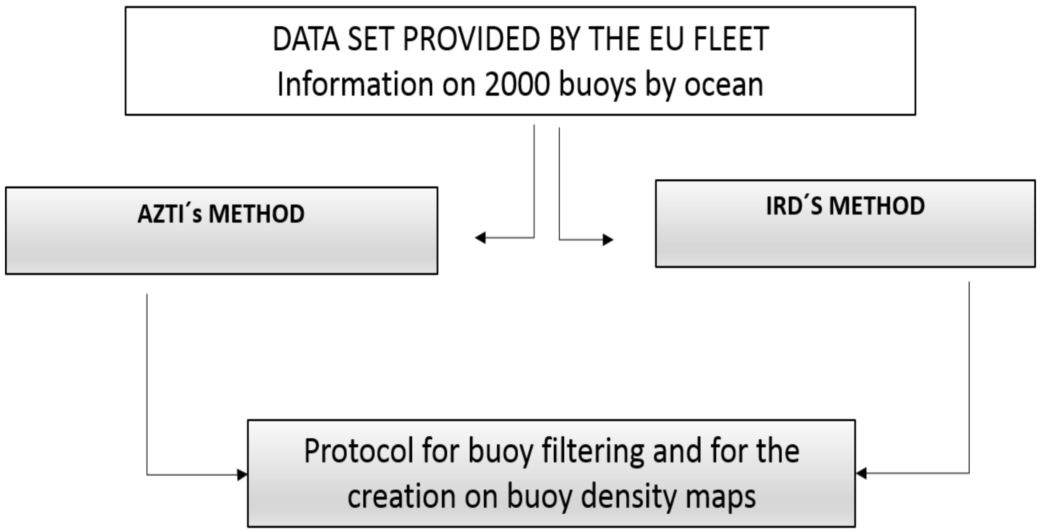


FR: Atlantic Ocean



- All buoys from brand B.
- Data on 2010-2018 for both buoy positions and echosounder of Brand B.
- Heterogenity in the acoustic information, given the characteristics of different buoy brands and models.

Data processing protocol: Description of the working scheme



Data processing protocol: Filters

Common protocol

Different resolutions in the shape file

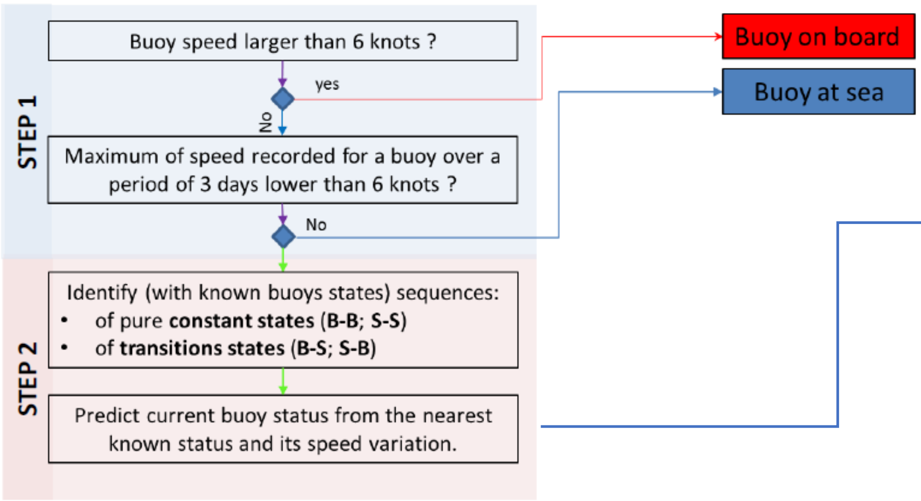
Common protocol

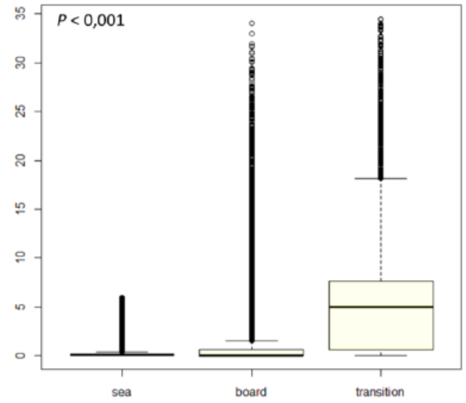
Two algorithms tested
IRD-Kinetic Algorithms
AZTI-Kinetic Random Forest

FILTER	Description
F1. Isolated	Isolated Position (>48 hours from another position or estimated speed above > 35 knots relative to next/previous position)
F2. Duplicated	Duplicated data (all fields are the same)
F3. Land and stationary	Data on land with speed <0.01 knots
F4. Land	Data on land with speed >0.01 knots
F5. Ubiquity	Data entry having from the same date/time different positions
F6. Not classified	Position not in the land and not classified by the at sea/on board algorithm
F7. Onboard	Buoys on board
F8. Water	Buoys at sea. Operational buoys: Active buoy that is transmitting a signal and is drifting in the sea (definition from RECOLAPE)

Data processing protocol: Filters – On board filtering

- Kinetic method of IRD (Baidai et al. 2017)





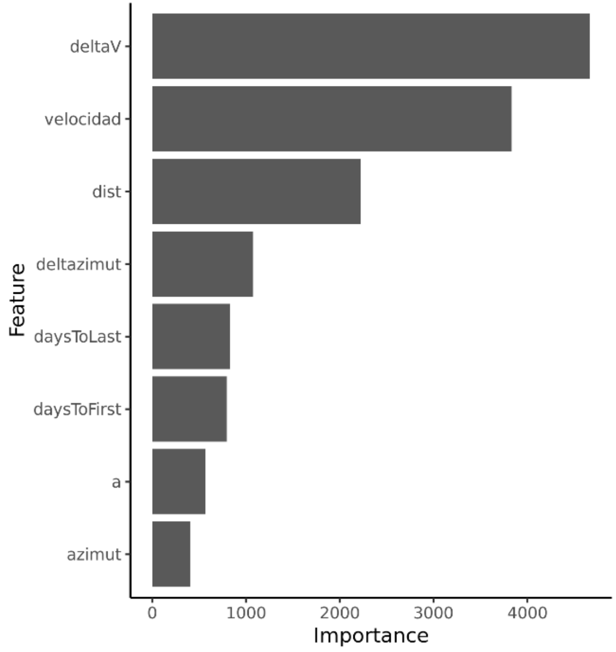
Boxplots of the mean acceleration (absolute values) for constant (sea and board) and transition states with p-value at ANOVA tests

- Additional algorithms available - Random Forest of IRD (Maufroy et al. 2015)

Data processing protocol: Filters – On board filtering

❑ Random Forest method of AZTI

- Algorithm based also in a random forest classification approach.
- Data source: Zunibal 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.
- Predictor 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).
- The most important variables in the RF model are velocity and variation in velocity.



Data processing protocol: Comparation of outputs (kinetic-IRD vs. Random Forest-AZTI)

Water (water+not classified)
vs.
Not water (duplicated + ubiquous + isolated + land + land and stationary + onboard)

Spanish buoys ATL Ocean
(Simple matching coefficient = 0.99)

	IRD	
AZTI	water	not water
water	24764	13
not water	213	314

French buoys ATL Ocean
(Simple matching coefficient = 0.96)

	IRD	
AZTI	water	not water
water	53735	1457
not water	1061	6649

Spanish buoys IND Ocean
(Simple matching coefficient = 0.94)

	IRD	
AZTI	water	not water
water	20892	25
not water	1245	299

French buoys IND Ocean
(Simple matching coefficient = 0.97)

	IRD	
AZTI	water	not water
water	57843	347
not water	1233	1771

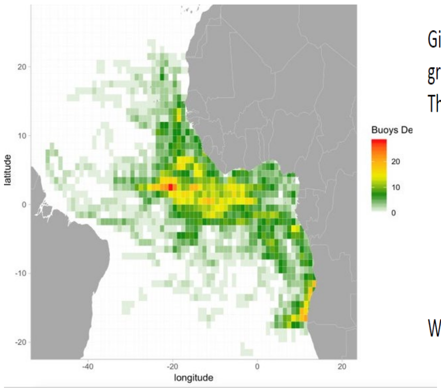
Data processing protocol: Filters – Acoustic filtering

❑ Description of the data processing protocol for acoustic information: Filters

- Filters defined for buoy position
- Bathymetry (British Oceanographic Data Centre, UK, www.gebco.net with resolution of 15 arc-second intervals), acoustic records from buoys located in areas with a bottom depth shallower than:
  - 150 m (in case of IRD working with MI buoys)
  - 200 m (in case of AZTI working with MI and Satlink buoys)
- Battery voltage on MI buoys: a voltage of 11.5 V
- Vertical boundaries:
  - Blind area of 3 m in Satlink and 6 m in MI
  - 25m depth to separate non-tuna from tuna (it can oscillate by season and area)
- Time of the day:
  - Samples obtained around sunrise, between 4 a.m. and 8 a.m., are supposed to capture the echosounder biomass signals that better represents the presence and abundance of tuna under the FADs (it can oscillate by season and area)

Applications

Information buoy density



Given b(i,r) representing the number of different bouys counted during day *i* for 1° grid cell r.
The 1°/month density will be estimated as:

$$D(r) = \frac{\sum_{i=1}^{N} b(i,r)}{N}$$

Where N is the total number of days for a given month.

CPUE Standardization

Information of presence of tuna and biomass

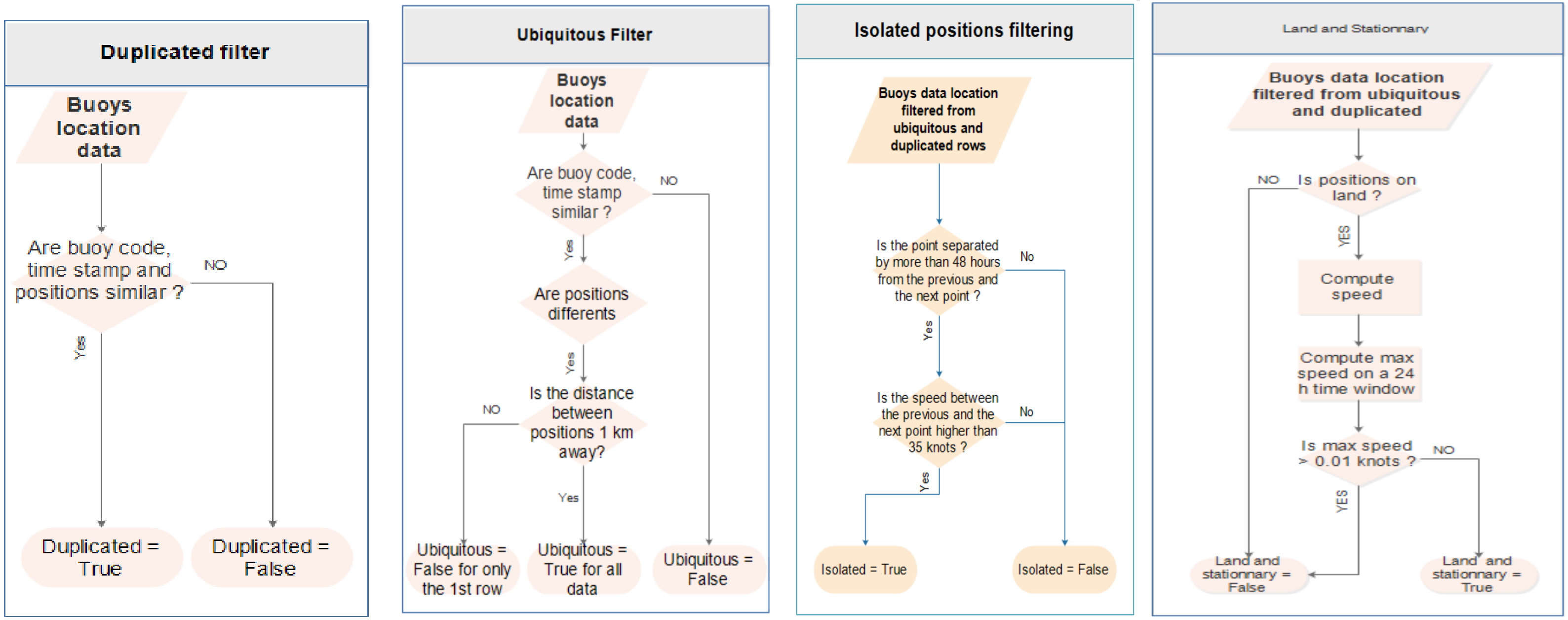
- IRD work with MI buoys and has developed algorithms to predict the presence/absence of tuna with good performance. The catch size can be also predicted, but with lower accuracy.
- AZTI is working with Satlink and MI instrument buoys in the improvement of tuna biomass estimation, applying new TS.

Alternative abundance indices

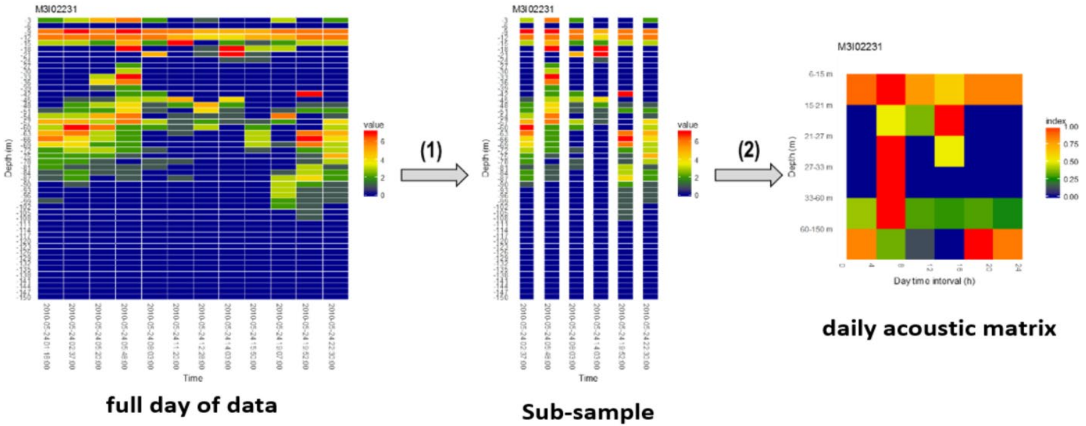
Concluding Remarks, Difficulties and Recommendations

- First collaborative project on FAD buoys data recovery at EU level where the database characteristics and the data-processing protocols on EU are described.
- The collaborative work of the industry, buoy provides and research institutions has allowed recovering historical information on buoy positions and acoustics.
- IRD has record information from 2006 to 2018 from EU France and AZTI from 2010 to 2018 EU Spain in the Atlantic and Indian Ocean.
- A filtering protocol has been established, with common steps and different algorithm for the detection of buoys on board (i.e. kinetic method vs RF method)
- High rate of agreement is observed in the filtering performance.
- For improving the performance high-resolution position data is desirable if available (more than 1 position/day).
- Some problems in the exportation of historic data has been faced (2010-2012 period for one brand and fleet). Therefore, in order to improve the data flow, periodic deliveries are recommended.
- Additional data has been identified as potential variables to be collected :
    - Temperature to identify buoys at sea-on board (already received by IRD).
    - IMO of vessel receiving the information for effort assessment (already received by IRD).

Additional information: Data processing protocol- Filters



Additional information: Accuracy of algorithms for biomass estimates: Random Forest (Baidai et al 2018/ IRD)



Good accuracy for assessing presence/absence of tuna
Less accurate to evaluate the size of the aggregation

Evaluation Metrics	Atlantic	Indian
Accuracy	0.75 (0.02)	0.85 (0.01)
Kappa	0.50 (0.04)	0.69 (0.01)
Sensitivity	0.82 (0.03)	0.78 (0.01)
Specificity	0.67 (0.03)	0.91 (0.01)
Precision	0.72 (0.03)	0.89 (0.01)

Additional information: Accuracy of algorithms for biomass estimates (Orue et al. 2019, AZTI)

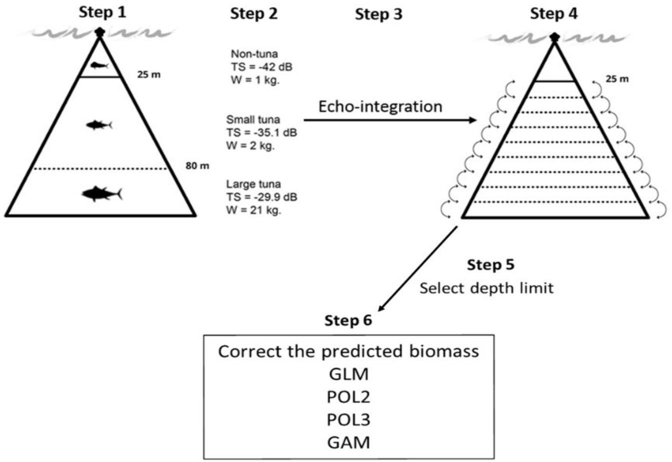


Table X. Coefficients of determination (r2) between catch and biomass estimated (manufacturer biomass, Manuf.; predicted biomass, Before correction; and corrected biomass obtained after different model corrections (GLM=generalized linear model; POL2=polynomial of order 2; POL3=polynomial of order 3; GAM=generalized additive model)) for all sets and each region.

Zone	Manuf.	Before correction	GLM	POL2	POL3	GAM
All sets	0.022	0.021	0.021	0.028	0.03	0.027
Somalia	0.025	0.025	0.025	0.026	0.029	0.025
Seychelles NW	0.047	0.050	0.050	0.158	0.158	0.159
Seychelles SE	0.065	0.073	0.073	0.093	0.093	0.073
Mozambique Channel	0.011	0.012	0.012	0.012	0.084	0.012