



**SCIENTIFIC COMMITTEE**  
**SEVENTEENTH REGULAR SESSION**

Online meeting  
11-19 August 2021

---

**Project 88 final report: Acoustic FAD analyses**

---

**WCPFC-SC17-2021/MI-IP-05**

**10 August 2021 Rev 1**

Lauriane Escalle<sup>1</sup>, Tiffany Vidal<sup>1</sup>, Beth Vanden Heuvel<sup>2</sup>, Ray Clarke<sup>3</sup>, Steven Hare<sup>1</sup>, Paul Hamer<sup>1</sup> and Graham Pilling<sup>1</sup>

<sup>1</sup> Oceanic Fisheries Programme, The Pacific Community (SPC), Noumea, New Caledonia

<sup>2</sup> Cape Fisheries;

<sup>3</sup> South Pacific Tuna Corporation (SPTC)

**Revision 1:**

Revision one of this paper includes a modification of the recommendation on page 5 and 34.

## Executive Summary

This report presents final results from the WCPFC SC Project 88 'Acoustic FAD analyses', funded by the European Union and WCPFC. This is a continuation of the work started in previous years, in collaboration with three fishing companies. The available data comprise 4.7 million acoustic transmissions from buoys deployed on drifting Fish Aggregating Devices (dFADs). This included data from three different satellite echosounder buoys: Satlink, Zunibal and Kato, which present different operational characteristics, such as biomass estimates, depth bins, or transmission frequency. Given the amount of data available and the specific characteristics of these buoy brands, analyses presented in the paper are based on Satlink data only (3.8 million transmissions), which best suited the specific objectives of this project.

Results from analyses associated with this project will be used to identify whether the following objectives might ultimately be addressed using acoustic data from satellite buoys deployed on dFADs:

- 1) Whether acoustic buoys on drifting dFADs can provide a novel and efficient source of fishery-independent data for stock assessments (e.g., indices of abundance or additional information for purse seine CPUE standardisation);
- 2) Whether limiting sets to only those dFADs that have a large estimated biomass beneath them could reduce the proportion of small bigeye and yellowfin caught.

Matching the trajectory of echosounder buoys attached to dFADs and dFAD activities from logsheet (associated fishing sets) and observer data (associated fishing sets and dFAD visits) was performed using date/time and position. This allowed estimated biomass levels and trends to be related to specific fishing events and associated catch records.

To evaluate the utility of the acoustic biomass estimates, initial analyses examined the patterns of biomass accumulation following deployment, and related biomass estimates just prior to setting to the resulting catch levels achieved. Following deployment, a clear increase was detected in the estimated biomass up to around 30–40 days post deployment, with a maximum around 60 days. Pattern of biomass accumulation before a fishing set showed an increase through time from 10 to 1 day before a set, with most buoys in the 3 days prior to a set having an estimated biomass between 25 and 60 t (up to 140 t). Examining the level of catch relative to the estimated biomass around the time of setting, larger catches generally corresponded to larger estimated biomasses over a period of 5-days before a fishing set. For very large catch events, the echosounder often underestimated the biomass or alternatively, the whole tuna school was not detected under the cone of the echosounder.

### **Towards an independent biomass index for stock assessment**

For this project objective, the focus was on skipjack tuna given the regional concerns over the reduction in the spatial extent and volume of the current pole and line abundance index for this stock. Two alternative methods were developed to classify echosounder transmissions as a relative index of tuna abundance or as presence/absence of skipjack tuna. The first method is a clustering analysis that classified echosounder transmissions in four different clusters based on biomass at depth, total biomass and biomass accumulation rate. One cluster (0) corresponded to no biomass detected. Two clusters (2 and 3), detecting biomass with differing depth profiles, corresponded to almost all transmissions a few days before a fishing set and therefore likely indicated the presence of tuna.

Finally, the last one (cluster 1), might be a mix of bycatch and tuna. Spatial distribution of each cluster was investigated and clusters 2 and 3 displayed a higher proportion in the main purse seine fishing area, while cluster 0, corresponding to transmissions with no biomass associated to the dFAD, had a low proportion. Higher proportions of cluster 0 were however found outside the main purse seine fishing areas, as expected. General additive models (GAMs) were used to investigate factors influencing total tuna catch. The limited number of skunk sets precluded the investigation of tuna presence/absence. Explanatory variables in the best model of tuna catch explained 25.9% of deviance. An increase in tuna catch was detected with increasing echosounder biomass, up to approximately 150 t, but for larger sets the biomass was underestimated by the echosounder. Other significant variables linked to higher tuna catch were time drifting, higher absolute values of biomass accumulation rates, full moon and higher biomass detected by the echosounder, latitude and longitude (higher tuna catch north of the equator and to the east of the WCPO).

The second method is a Random Forest analysis that classified echosounder transmissions into presence/absence of skipjack tuna, based on a learning dataset of matched observer set/acoustic transmissions and Vessel Monitoring Systems/acoustic transmissions; and buoys with no associated biomass. Skipjack presence generally showed longer drift times (median of 79) compared to skipjack absence (median of 21). Skipjack presence was also associated with higher absolute biomass accumulation rate values. Finally, it was also found that skipjack presence was found in areas with higher vessel density compared to absences. Interestingly, the match in results between the random forest classification and the clustering analysis was very good, with more than 97% of transmissions in clusters 1, 2 and 3 classified as skipjack presence; and 92.7% of transmission from cluster 0 classified as absence.

The classification of presence/absence of skipjack tuna was then used to explore an integrated standardization approach that combines CPUE time series from the purse seine fishery with presence/absence data from acoustic dFAD buoys drifting throughout the WCPO. Two models were compared, both included the WCPO-wide purse seine catch and effort data, while only one included the presence/absence data set. Both models predicted similar patterns, with notable differences in the magnitude of relative abundance; as well as lower predicted encounter probability for the combined data set, and higher predicted abundance during the dFAD closure period. The distribution of skipjack density was also notably different between the two models, mostly due to the information about skipjack encounters from the dFAD network. The positive catch rates showed similar patterns, but with a difference in magnitude. This analysis is a preliminary exploration of the methodology, with a relatively short time-series. These results are not intended to be indicative at this stage, but rather an opportunity to evaluate alternative approaches to obtain more reliable information on trends in skipjack abundance.

- While further analyses are required on a larger data set, this preliminary study suggests that acoustic biomass estimates from regional drifting FADs could assist in the development of an independent biomass estimate for skipjack tuna in the WCPO.

#### **Exploration of bigeye and yellowfin tuna proportion in fishing sets related to estimated biomass**

Bigeye and yellowfin tuna proportions derived from the catch composition were examined in relation to the levels of total biomass estimated by the echosounder buoys in the days preceding a set. The proportion of yellowfin tuna in the catch appeared relatively constant with increasing levels of

estimated biomass, while the proportion of bigeye tuna appeared to increase with higher levels of biomass estimated by the echosounder buoy. GAMs were then used to investigate factors influencing bigeye and yellowfin catch proportions. Explanatory variables in the GAMs explained 23.8 and 27.0% of deviance of the bigeye and yellowfin tuna proportions. Biomass estimated by the echosounder was not significant for bigeye tuna models, but lower yellowfin tuna proportions were found for lower biomass estimated. Other variables influencing bigeye tuna proportions were drift time, low biomass accumulation rates just before the set (2-days), time relative to sunrise, longitude and the interaction between longitude and latitude. Other variables influencing yellowfin tuna proportions were moon phase and longitude. The limited number of sets considered here should be noted, as well as the fact that the echosounder is calibrated to detect skipjack tuna. The presence of a swim bladder for yellowfin and bigeye tuna, contrary to skipjack tuna, might therefore influence the acoustic signal. Accessing data from multi-frequency echosounder buoys that are now more frequently being deployed in the WCPO may allow these analyses to be refined.

- While the current analysis suggests little relationship between biomass estimates and the catch proportion of bigeye tuna, statistical models showed a slight decrease in yellowfin proportion for larger estimated biomass. Other factors, including drift time and trends prior to setting (rather than at setting) showing more potential.

It should be noted that the acoustic data available for this analysis represent a small subset, available through partnership with three fishing companies, of the full dFAD network in the WCPO. The assistance of those companies is gratefully acknowledged. Therefore, results presented in this report should be viewed as preliminary and highlighting potential methods to be further explored with a more complete data set. Our intent is that WCPFC members and fishing companies see the utility of the analyses presented, in particular towards improving the monitoring of tropical tuna stocks and encourage or commit to making their acoustic dFAD data available for these endeavors.

**We invite WCPFC-SC17 to:**

- Note the results from Project 88 on acoustic data from echosounder buoys deployed on dFADs.
- Note that while the acoustically estimated biomass related to the proportion of bigeye in the catch, this proved non-significant within models, but statistical models showed a slight decrease in yellowfin proportion for larger estimated biomass and other identified variables may show some promise to pre-identify sets likely to lead to greater proportions of bigeye and yellowfin in the catch.
- Note the potential, over the longer-term, to use echosounder data as a source of fishery-independent data for stock assessments, either as an independent relative index of abundance or to provide additional information for purse seine CPUE standardization.
- Recommend the need for better identification of particular dFAD buoys (e.g., via the buoy identification numbers) by commercial vessel operators or via observer reports.
- Endorse the continued cooperative relationship with the fishing community to obtain commercially sensitive data for analysis for the purpose of scientific and other research, particularly with regard to dFADs, and the fishing strategies involved in their use. Highlight the need for additional data covering the whole WCPFC convention area, including that from now available multi-frequency echosounder buoys, and encourage other industry partners to become involved in the project.

## 1. Introduction

The deployment of satellite and echosounder buoys on drifting Fish Aggregating Devices (dFADs) by purse seine vessel operators has dramatically increased over the last two decades in all ocean basins. In the Western and Central Pacific Ocean (WCPO), the reported number of dFAD sets and the number of dFADs deployed have been relatively stable over the last decade, the latter varying between 30,000 and 40,000 each year (Escalle et al., 2021a). Harvest from dFAD sets accounts for approximately 40% of the WCPO purse seine tuna catch (Williams and Ruaia, 2021). Satellite and echosounder buoys are relatively new technological developments that allow fishers to track dFAD locations and monitor the relative biomass of tuna beneath the buoy, which has the potential to increase fishing efficiency (i.e., catch per unit effort - CPUE) dramatically (Lopez et al., 2014; Wain et al., 2021). DFADs also have the potential to be a useful source of information for scientific investigations (Moreno et al., 2016). In the WCPO, the majority of dFADs are now deployed with an echosounder (99% in 2020, Escalle et al., 2021b). This represents a rich and robust data source that has the potential to help inform mitigation approaches to reduce catches of small bigeye and yellowfin tuna, increase our understanding of fleet dynamics, and potentially provide a new source of fishery-independent data on tuna abundance for regional stock assessments.

In the WCPO, skipjack tuna *Katsuwonus pelamis* catches are the highest of all tuna species (approximately 1,769,202 tons in 2020), with 82% of the harvest coming from the purse seine sector (Williams and Ruaia, 2021). However, despite the substantial contribution of the purse seine fishery to the WCPO skipjack tuna harvest, CPUE time series from this fishery are largely absent from the assessment process (Vincent et al., 2019). The most recent assessment relied primarily on CPUE data from the pole-and-line fishery (a fishery which contributed about 6% of the total harvest in 2019, and has been spatially contracting through time) to inform the model on trends in abundance (Vincent et al., 2019). This absence of the purse seine CPUE data from skipjack stock assessments is in part due to the perceived hyperstability of purse seine catch rates (Hoyle et al., 2014; Hamilton et al., 2016) and lack of proportionality with the underlying stock, but the consequence is a growing disconnect between the dominant fishery sector and the data used to assess the underlying stocks.

Developing approaches to integrate data more fully from the purse seine fishery into tuna assessment and monitoring programs may be key to continued sustainability of these stocks. Approaches could either include fishery-independent relative abundance indices developed from the echosounder buoys deployed on dFADs throughout the WCPO (Diallo et al., 2019; Santiago et al., 2019; Uranga et al., 2021), or integration of dFAD acoustic data into the CPUE standardization approach (Maunder and Punt, 2004), by combining time series from the purse seine fishery with presence/absence data from echosounder buoys.

Preliminary analyses, as previously reported to SC, identified that the format of available data, i.e., from single frequency echosounders, allowed some investigations to be performed (Escalle et al., 2020b, 2019b). In particular, it was determined that acoustic dFAD data and logsheet/observer set data could be linked (although assumptions are required), and some signals in the acoustic data could be related to catch levels, with notable variability. A subset of the dataset was selected, for which trajectory and fishing activities corresponded, with high confidence, to the same buoy attached to a dFAD (Escalle et al., 2020b). Additional parameters were also compiled and included dFAD soak time, drift speed, trends in biomass accumulation leading to a fishing set, biomass moving averages, biomass

by depth, total catch or catch per species, moon phase, time of the set, and spatial areas. It allowed for additional investigation that increased knowledge of biomass accumulation dynamics and of the signal captured by echosounder buoys.

This report presents results from the WCPFC project 88 ‘FAD acoustics analyses’. This is a continuation of the work started in previous years (Escalle et al., 2020b, 2019b), in collaboration with two US-based private sector firms Cape Fisheries and South Pacific Tuna Corporation (located in San Diego, California) and one Solomon Islands-based firm National Fisheries Developments (NFD). The three firms own several large purse seine vessels that operate in the WCPO and exhibit broad spatial coverage of the purse seine fishing grounds, with operational characteristics that are representative of a large segment of the fishery. Results from analyses associated with this project will be used to identify whether the following objectives might ultimately be addressed using acoustic data from satellite buoys deployed on dFADs:

- 1) Whether acoustic buoys on drifting dFADs can provide a novel and efficient source of fishery-independent data for stock assessments (e.g., relative indices of abundance or additional information for purse seine CPUE standardization);
- 2) Whether limiting sets to only those dFADs that have a large estimated biomass beneath them could reduce the proportion of small bigeye (*Thunnus obesus*) and yellowfin (*Thunnus albacares*) tunas caught.

It should be noted that the acoustic data available for this analysis is a small subset, available through partnership with those three fishing companies, of the full dFAD network in the WCPO. Therefore, results presented in this report should be viewed as potential methods that could be further explored with a more complete data set. Our intent is that fishing companies will see the utility in improved monitoring of tropical tuna stocks and commit to making their acoustic dFAD data available for these endeavors.

## 2. General description of the data and processing methods

### 2.1 Available data

Biomass estimates from echosounder buoys of the three partner fishing companies, Cape Fisheries, NFD, and South Pacific Tuna Corporation, were independently provided to SPC. These data were obtained from three satellite echosounder buoy providers (Satlink, Zunibal and Kato), with each brand encompassing unique data characteristics (Table S1). The characteristics of each buoy brand and the corresponding datasets are detailed in Appendix 1. The data available for analysis comprised over 4.7 million acoustic transmissions from buoys deployed on dFADs in the WCPO in 2016–2018 (Table S1).

Generally, the analyses performed on acoustic data must be tailored to the specific characteristics of the data obtained from different buoy brands, as they are not directly comparable at this time. Therefore, analyses need to be performed separately for each dataset. In this paper, we focused exclusively on acoustic data received from Satlink buoys for several reasons. Specifically, most of the Zunibal buoys did not include a precise hourly position nor depth-disaggregated acoustic signals, which precluded analyses of interest. For Kato buoys, acoustic data only corresponded to a relative index per depth layer (between 0 (no biomass) to 15 (very high biomass)), which will need specific investigations. The Satlink buoys represent the majority of transmissions from the data set we have

received, with a wider spatial distribution than the other two brands. In addition, Satlink data provide precise location and time of the dFAD buoy transmissions as well as biomass estimates by depth layer below the dFAD. Therefore, we have only considered Satlink data for the analyses presented in this report.

## 2.2 Data processing

Consistent with the analysis of the Parties to the Nauru Agreement (PNA) dFAD tracking dataset (Escalle et al., 2020a), and analyses performed previously on similar dFAD echosounder datasets (Escalle et al., 2019b), the raw position and acoustic dataset included transmissions from active buoys drifting at-sea but also included data from some that were still on-board a vessel (before deployment, or following recovery). Therefore, data were processed by identifying at-sea and on-board positions following the approach of Maufroy et al. (2015). Random Forest models (Breiman, 2001) were calibrated using a learning dataset, then used to predict the class (at-sea or on-board) of positions in the echosounder dataset. Models were applied using the R package *randomForest* (Liaw and Wiener, 2002). Additional correction procedures were also performed to eliminate isolated or short at-sea or on-board sections surrounded by long on-board or at-sea positions (Escalle et al., 2019a). Each buoy trajectory with acoustic data then consisted of one or several drifting ('at-sea') segments, separated by 'on-board' positions. Deployment positions were identified as the first at-sea position.

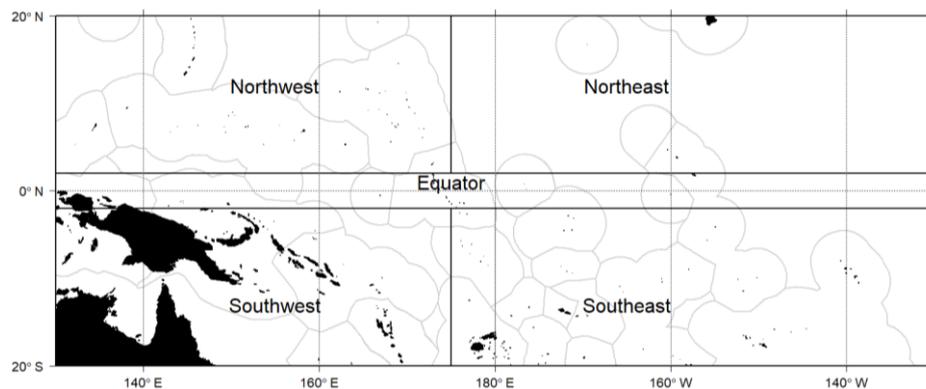
A total of 3,804,244 acoustic transmissions from Satlink buoys were available after the filtering process (Table S1). DFADs with no biomass estimates (only position, see Appendix 1) over the whole trajectory were removed. Only at-sea segments were selected, removing part of trajectories where dFADs were on-board a vessel. The trajectory data were then summarised to produce one record per dFAD buoy per day. For each dFAD buoy, we identified the transmission showing the maximum acoustic biomass for a given day (across all depth layers), and then retained additional summary statistics associated with that transmission including buoy ID number, dFAD segment ID (indicating each at-sea segment separated by a re-deployment), and the vessel(s) that received the transmission, date, time, and location. The echosounder recorded biomass at depth in ten depth layers of 11.2m, from 3m to 115m (Appendix 2). The time relative to sunrise was also calculated, as tunas tend to be more closely aggregated to dFADs around sunrise (Harley et al., 2009). After the aggregation of the data, there were 1,420,371 acoustic unique daily transmissions available for analyses.

In addition, the amount of time the buoys had been drifting since the last deployment was calculated (in days). However, given the common practice of exchanging buoys on a dFAD found drifting at-sea, some of the buoy deployments would not correspond to a dFAD deployment, but only deployed on a dFAD already drifting. To overcome this issue and isolate, when possible, the time that the dFAD had effectively been drifting at-sea, first deployments were identified when several deployments by the same vessel occurred in a row (see Appendix S1.3), hereafter referred to as "grouped deployments".

Finally, the relative biomass accumulation rate over the preceding period (5, 10, and 20 days) was calculated. In other oceans, it has been shown that biomass tends to increase after an initial colonization period (Baidai et al., 2020; Uranga et al., 2021) and, in an industry survey, fishers indicated that they monitor the accumulation of biomass at a dFAD with respect to deciding where and when to fish (Wichman and Vidal, 2021). The accumulation rate was calculated as the slope of the linear

biomass trend, over the preceding period, was predicted to be important when attempting to ascertain whether a dFAD was retaining tuna biomass or not.

Five main areas of the WCPO were also considered: i) equator, between 2°S and 2°N; ii) southwest, south of 2°S and west of 175°E; iii) southeast, south of 2°S and east of 175°E; iv) northwest and west of 175°E; and v) northeast, north of 2°S and east of 175°E (Figure 1). This was based on changes in dFAD drifting speed, especially the latitudinal variability (Figure S5). For the analyses focused on the development of relative abundance indices for skipjack, we restricted the spatial extent of the data to the assessment regions associated with purse seine fishing activity (i.e., regions 6–8) from the most recent assessment (Vincent et al., 2019; Figure S7).



**Figure 1.** Map of the delineated areas of the WCPO used in the study.

### 2.3 Matching between echosounder buoy trajectories and fishery data

In this study, matching between the trajectory of echosounder buoys attached to dFADs with dFAD activities from logsheet and observer data was performed using date, time and position. Associated fishing sets (sets in association with dFADs or logs) from all vessels in the WCPO from logsheet operational data were considered. For the observer data, associated fishing sets and dFAD visits that did not result in a fishing set, from all vessels in the WCPO, were extracted from the Pacific Islands Regional Fishery Observer Program (PIRFO). Observer coverage is mandated for 100% of observer trips, and the realized coverage averages around 84–98% (Panizza et al. (2021); excepting 2020, due to COVID-19). For both logsheet and observer data, information included position, date and time, and for fishing sets, the total catch and the catch per species (three tuna species, and bycatch per category for the observer data) and covered the 2016–2018 period in the WCPO to match the available time period of acoustic data. Note that the species composition is from operational logsheet and observer data, hence the regional species composition adjustment could not be performed at the set level.

A total of 42,529 and 31,669 associated sets from all vessels in the WCPO were available for matching from the logsheet and observer data, respectively. In addition, 156,489 visits (visiting, servicing, or retrieving dFADs or buoys from all vessels in the WCPO) recorded by observers were available for matching with trajectories. Only visits to a dFAD for which the vessel did not have an associated set the same day were selected. As mentioned above, the acoustic trajectory data used for matching included 3,804,244 acoustic transmissions in the 2016–2018 period (non-aggregated by day dataset).

Each acoustic transmission was matched to the nearest associated set or dFAD visit from logsheet and observer data using the “distHaversine” function of the *geosphere* package in R (Hijmans, 2019) and

the time difference between the two data points calculated, in hours. Sets and visits were identified as from a vessel's own dFAD, or the dFAD of the vessel's parent company, and the number of unique vessels receiving each acoustic transmission was calculated.

Once the datasets were matched, some criteria were developed to identify the matches that were likely to be real. In theory, observers record the dFAD buoy ID number; but in practice this is not always possible (buoy ID number recorded in 20% of the sets and 50% of the deployments, from all WCPO vessels) nor is it consistently recorded properly (Escalle et al., 2018). In addition, the transmission timing for dFAD acoustics may not align perfectly with the timing of a set or a visit. Therefore, exploratory and sensitivity analyses were performed using observer data to identify the best distance and time difference to reliably associate a transmission with a set or visit (Appendix 1.4). Following the sensitivity analyses, only matches where sets or visits were made less than 5 km distant and within 8 hours of an acoustic transmission were considered. Only sets, or other dFAD-related activities, with the same buoy ID in observer and acoustic trajectory data were considered for analyses presented in this section.

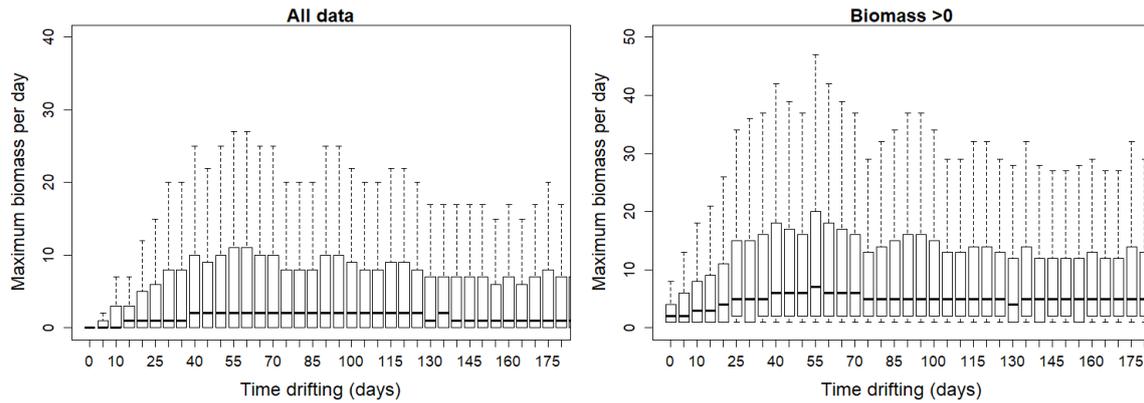
Finally, only matched acoustic records occurring within +/- 5 hours of sunrise were retained, as this is the time when tunas are known to aggregate near the surface. Signals outside of the time window were not predicted to reliably represent tuna presence. Selected matches included 638 logsheet sets, 481 observer sets and 179 visits from observer data.

### **3. Potential use of acoustic signal as index of tuna presence or abundance**

Although independent estimates of biomass are not available to ground-truth the estimates derived from the echosounder buoys, the relationship between estimated biomass and resulting catch can be used to provide some useful information.

#### **3.1 Colonization of biomass after dFAD deployment**

DFAD colonization processes after deployment were investigated for dFADs with no associated sets matched to the trajectory and classified as "grouped deployments" (Figure 2) or all dFADs with no associated sets matched to the trajectory (Figure S8). For dFADs with relatively certain drift time, a clear increase in the estimated biomass up to around 3–40 days post deployment is detected (Figure 2), with a maximum around 60 days. Similar pattern is detected for all dFADs (Figure S8), although the biomass is sometimes higher for very short drift times, highlighting the fact that some deployments detected in this study might be a buoy deployment on a dFAD found drifting at-sea.

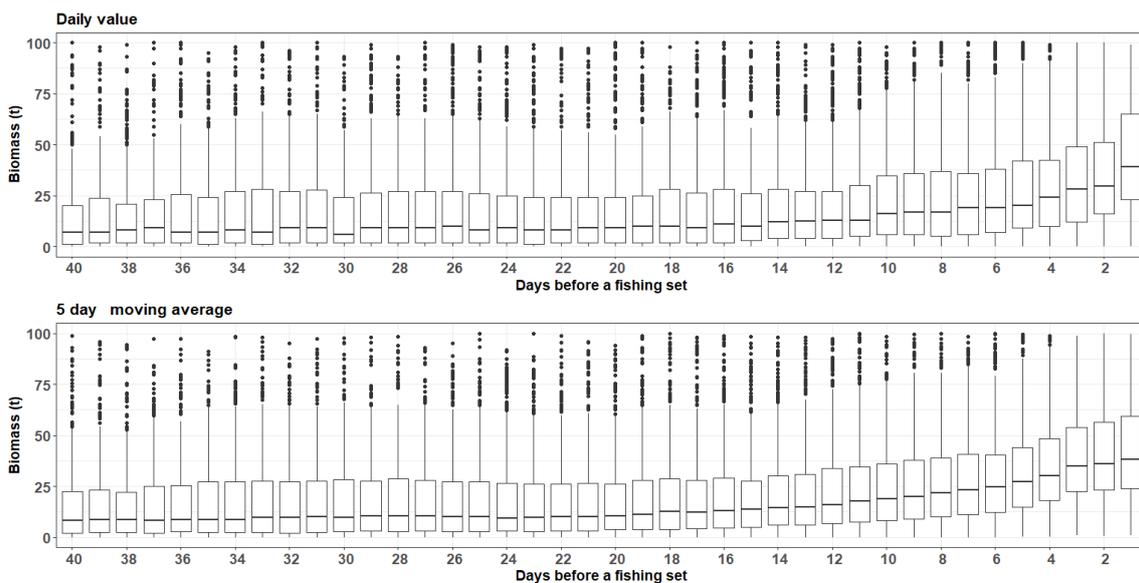


**Figure 2.** Evolution of the maximum biomass per day (t) estimated by the echosounder buoys (transmission within +/- 5h of sunrise) depending on time drifting, for dFADs deployed in “grouped deployments” and with no sets matched, with either all data considered (left) or only transmissions with estimated biomass >0 (right). Horizontal line corresponds to the median, boxes to the lower and upper quartiles.

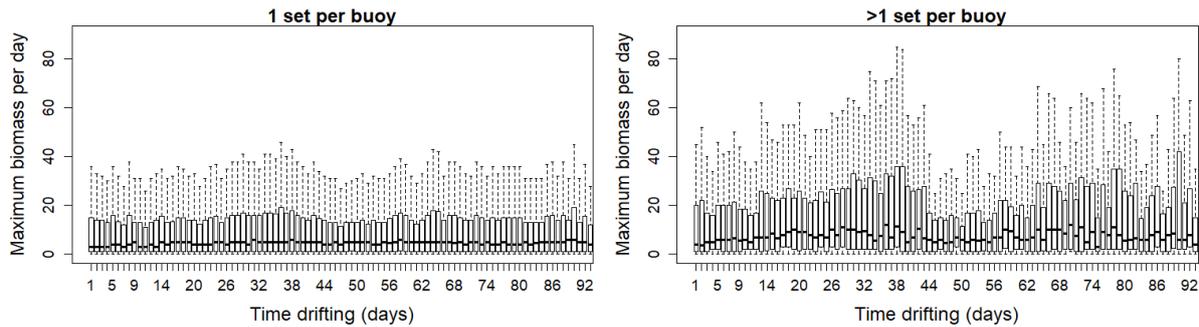
### 3.3 Biomass accumulation before a fishing set

The pattern of the maximum daily biomass before a fishing set was examined (Figure 3). The range of maximum biomass the day prior to a set was from 0 to 140t, with most estimates being between 0 and 30t. Biomass was generally highest from 10 to 1 days before a set. Within 3 days prior to a set, most buoys estimated between 25 and 60t (Figure 3). Similar patterns were found using the daily value and the 5-day moving average, although a more gradual increase leading to the fishing set was detected with moving averages (Figure 3).

This was compared to the estimated biomass after a fishing set, specifically in the case that one or more fishing set was performed 10 days or more after the first set (Figure 4). If no set was performed, the biomass remained low (below 20t), however an increase was detected, specifically after 30 days.

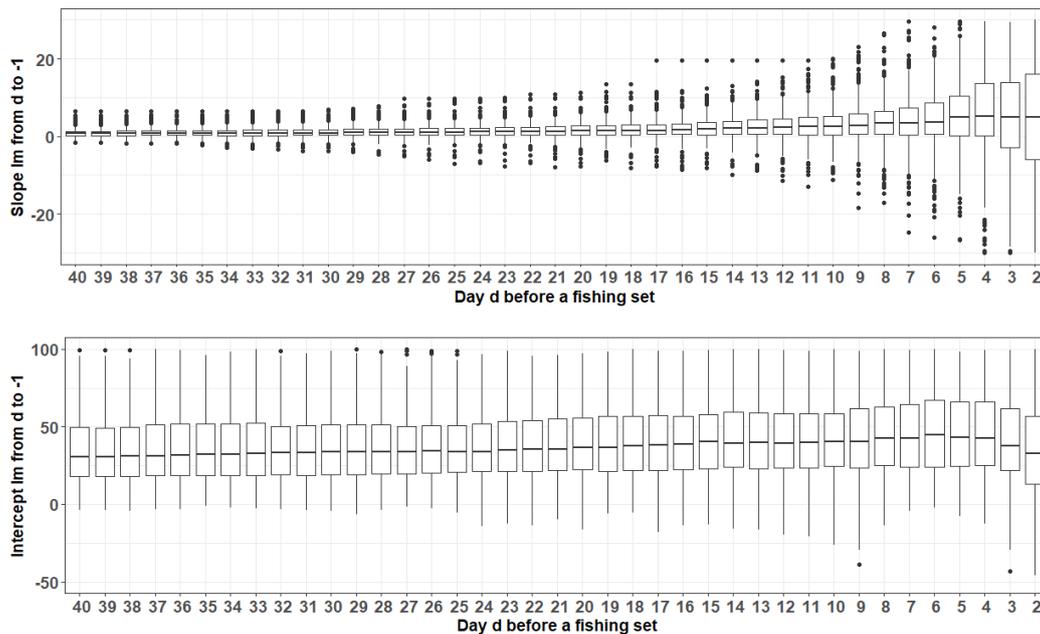


**Figure 3.** Evolution of the maximum daily biomass estimates (values >0.95 quantile not shown here, maximum of 340t) from Satlink echosounder buoys each day before a fishing set, using the daily value (top) or a 5-day moving average (bottom).



**Figure 4.** Evolution of the maximum daily biomass estimates from Satlink echosounder buoys each day after a fishing set, for buoys with only 1 set (top) or more than one set (bottom) with a 10day interval between first and second set.

Biomass accumulation rates before a fishing set were analysed using the slope and intercept from linear models of the estimated biomass per day over varying periods of time prior to the set. These were compiled (see linear regressions on Figures S9 and S10) over varying periods before a fishing set (40–2 days to the day before the set; Figure 4). In general, an increasing trend in biomass before a set was detected, except sometimes for periods just before the fishing set, when the slopes were sometimes negative. This might result from overall high biomass levels already reached and some day-to-day variability.



**Figure 5.** Slopes and intercepts from linear models of the maximum daily biomass estimates over varying periods prior to the set, depending on the period considered (40–2 days to the day before the set) from Satlink echosounder buoys.

### 3.3 Relative index of tuna abundance

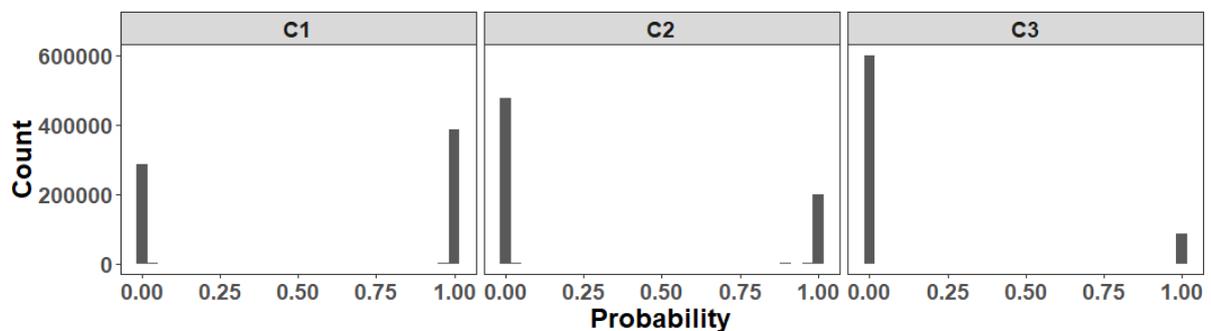
While previous investigations have shown that the link between total biomass and catch per set was difficult to discern (Escalte et al., 2020b), the type of signal detected in terms of biomass at depth and biomass accumulation are likely to be an indication of different levels of tuna abundance. Development of a relative index of tuna abundance was therefore investigated. Given the type of data

available, this could only be examined for all tuna species combined but could potentially be examined for individual tuna species if multi-frequency echosounder buoys data were made available.

A model-based clustering method (Fraley and Raftery, 2002) was used to identify groups of echosounder readings with similar data features in terms of biomass estimates and profile of biomass at depth and accumulation rates over previous days. This method was preferred compared to traditional clustering algorithms (Kaufman and Rousseeuw, 1990) because it assumes a data model and measures probability of cluster assignments. It can also be applied to very large datasets. The data in a model-based clustering are considered as coming from a distribution, in this case a normal distribution, that is mixture of two or more components or clusters. Each cluster is modeled by a Gaussian distribution characterized by some parameters: mean vector, covariance matrix, associated probability of belonging to each cluster.

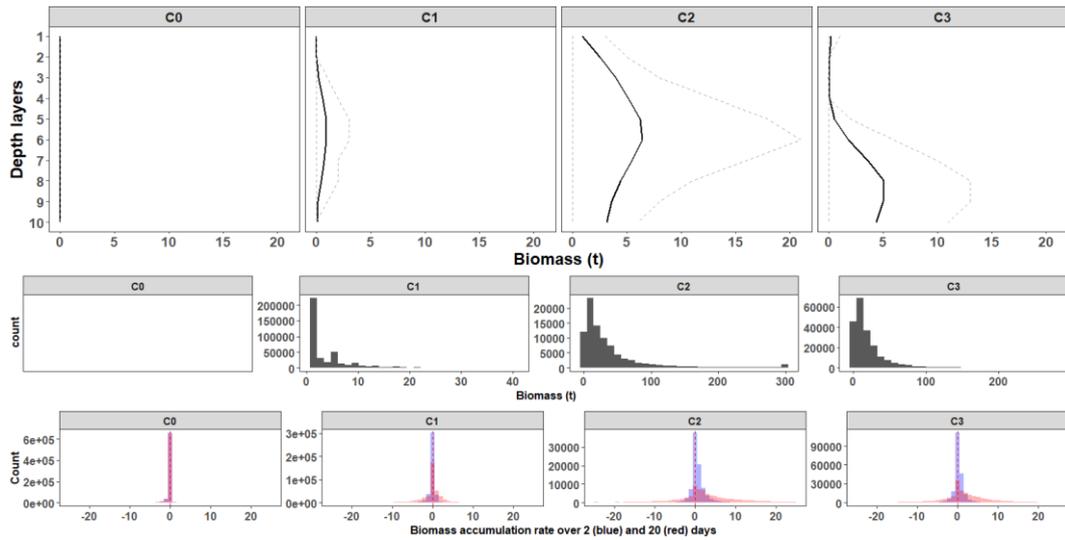
Clustering was performed using the *mclust* package (Fraley et al., 2012; Scrucca et al., 2016) applied to a dataset comprising, for each observation, the biomass estimated at each of the 10 depth layers (Appendix 2). Observations with total biomass across depth layers of 0, were separated before the clustering analyses and grouped as cluster 0.

Different models, based on maximum likelihood, are fitted with different covariance matrix parameterizations and a range of clusters (Fraley et al., 2012; Scrucca et al., 2016). The best model is selected using Bayesian Information Criterion (BIC), balancing model fit and parsimony, to determine an optimum number of clusters. The selected best model was the ellipsoidal, equal shape model, which identified three clusters (four in total, with the additional cluster 0). Probability of belonging to each cluster was very high for all transmissions classified, with very few observations showing intermediate probability of being in one cluster or another (Figure 6).



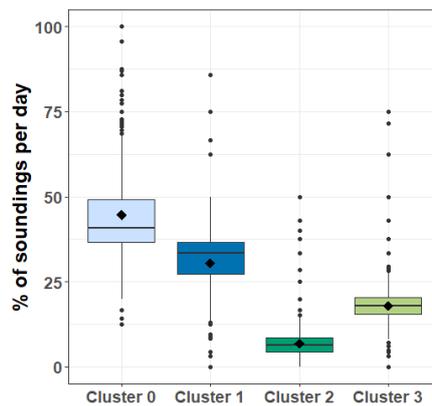
**Figure 6.** Probability of belonging to each cluster classified used the model-based clustering analyses for each maximum daily biomass transmission.

The four clusters (three identified by the model-based clustering and the additional cluster 0) showed very different profiles of biomass at depth, total biomass and biomass accumulation rate (Figure 7). Cluster 0 has no biomass detected and as expected, null or low decreasing biomass accumulation rates. Cluster 1 has relatively low biomass, mostly less than 10t and detected from 36 to 92m (layers 4 to 8). Clusters 2 and 3 have higher total biomass and a generally increasing biomass accumulation rate over 2 and 20 day periods. The profile of biomass at depth is however different, with mostly biomass detected in the deepest layers in cluster 3, from 81 to 115m (layers 8 to 10) and mostly biomass from 36 to 92m (layers 4 to 8, Figure 7).

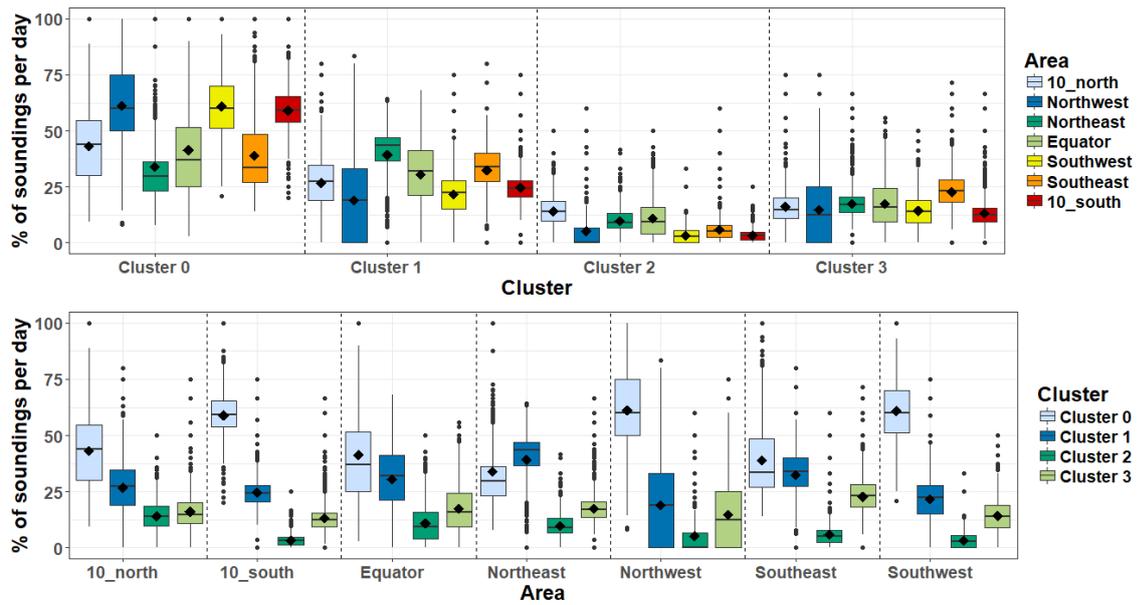


**Figure 7.** Characteristics of each cluster, in terms of distribution of biomass at depth for each transmission (top), distribution of total biomass (middle) and biomass accumulation rates over 2 (blue) and 20 (red) days (bottom).

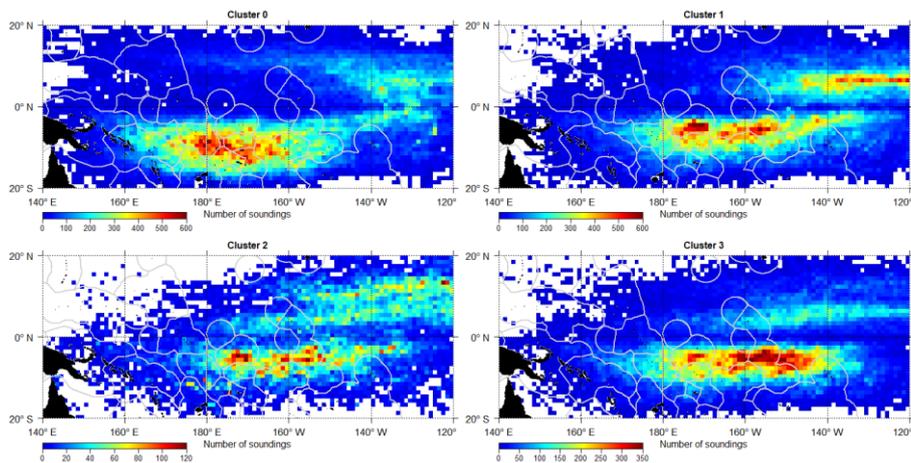
Most echosounder transmissions (around +/- 5 hours of sunrise) were classified in cluster 0 (average of 47% of daily transmissions), indicating that on most days, there is no fish biomass detected under the majority of dFADs (Figure 8). Cluster 1 was associated with the next highest group association (average of 27%), following by cluster 3 (15%) and finally cluster 2 (7%). Some variability in the percent associations with each cluster was however detected depending on the area (Figures 9, 10 11 and S11), but not relative to the dFAD closure period (Figure S12). Higher proportions of clusters 2 and 3 were detected in the main purse seine areas, from Kiribati Gilbert Islands, to Kiribati Line Island, between 10°N and 10°S (Figures 10 and 11). In this area, the proportion of cluster 0 was very low, but was very high in areas outside the purse seine fishing area, especially above 10°N and below 10°S in the west. In the northeast of the WCPO, high proportion of cluster 1, and to a lesser extent cluster 2, were detected (Figure 10 and 11).



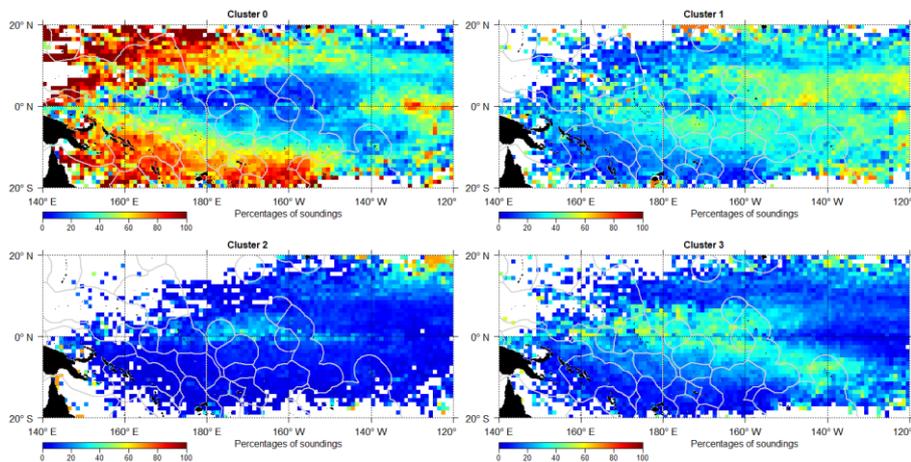
**Figure 8.** Percentage of echosounder transmissions (around +/- 5 hours of sunrise) per day that are within each of the clusters.



**Figure 9.** Percentage of echosounder transmissions (around +/- 5 hours of sunrise) per day that are within each of the clusters per area considered (see Figure 1), with cluster (top) on the x-axis or area (bottom).



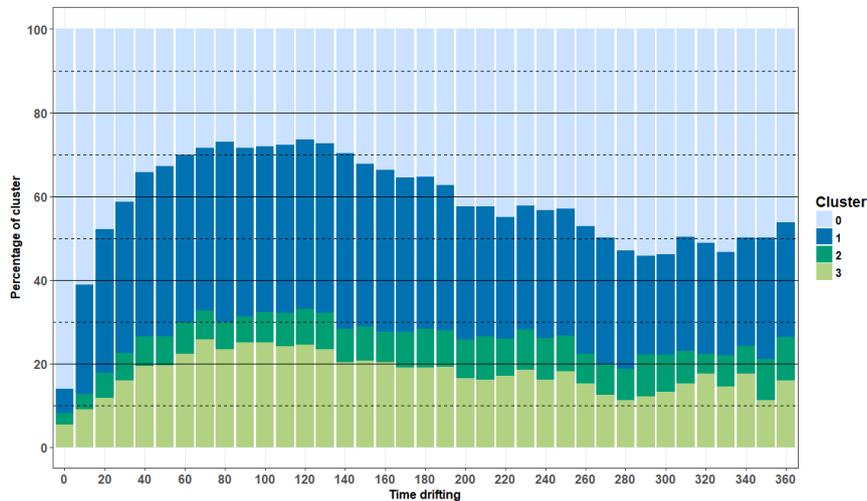
**Figure 10.** Number of echosounder transmissions per 1° cell classified in each cluster.



**Figure 11.** Percentage of echosounder transmissions per 1° cell classified in each cluster.

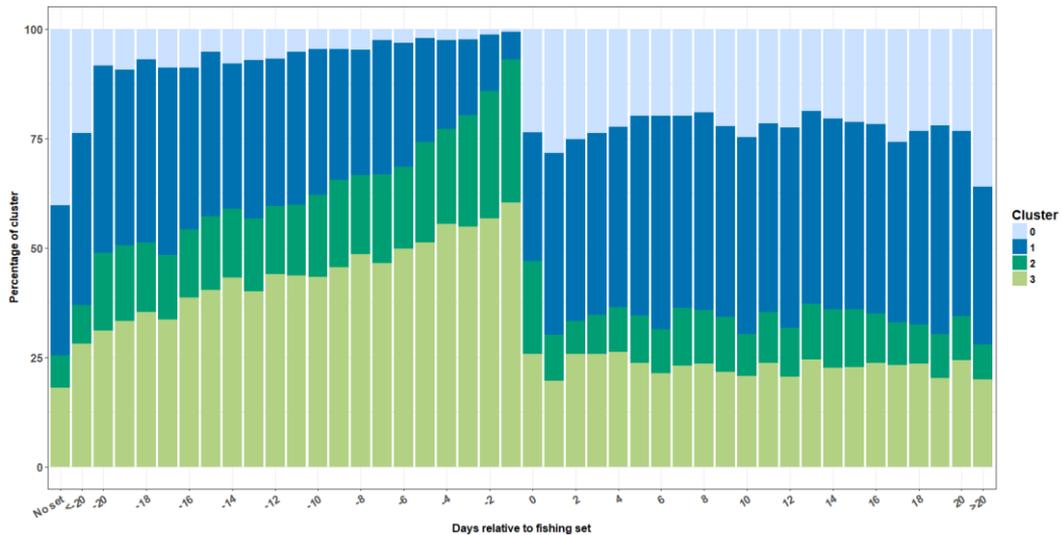
The percentage of transmissions classified within each cluster depending on dFAD drift time is also examined (Figure 12 and S13). For the first 10 days, the majority of transmissions were classified in

cluster 0, indicating no biomass aggregated, then the importance of the other clusters increases until 60 days. Between 2 months and 5 months, cluster 3 is around 25% of the transmissions, cluster 2 is around 8%, cluster 3 around 25% and cluster 1 around 35%. After 5 months, the percentage of cluster 0 transmissions increases again and the transmission of other clusters decreases, specifically clusters 1 and 3.



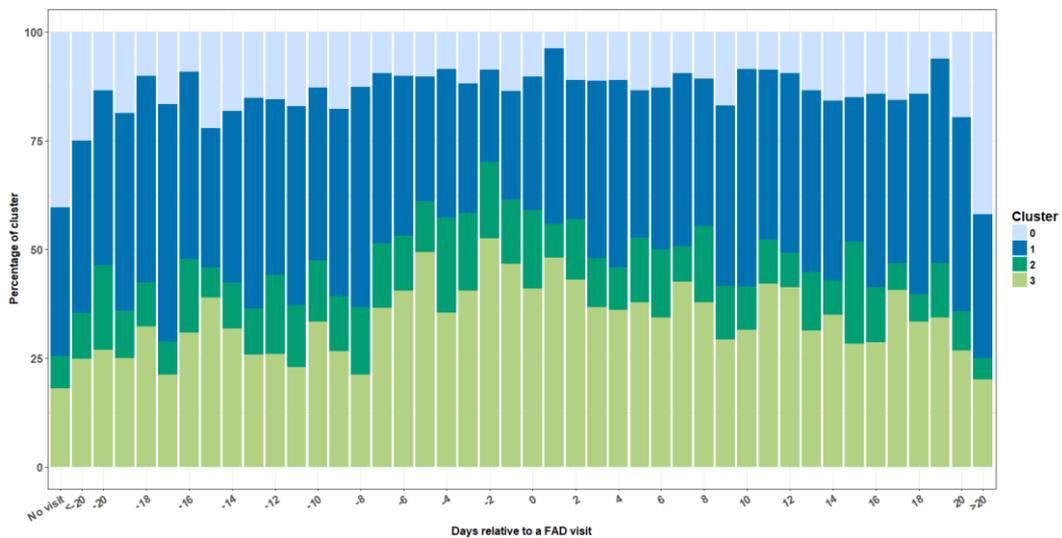
**Figure 12.** Percentage of echosounder transmissions (around +/- 5 hours of sunrise, and within 10°N and 10°S) per clusters depending on time drifting (only dFAD deployed in grouped deployments were considered).

An increase in the percentage of transmissions from clusters 2 and 3 was also detected in the days preceding a fishing set (Figures 13, S14 and S15). In particular, it can be noted that from 20 days before a fishing set, the percentage of cluster 0 in the transmissions is less than 15% (almost 0 the day preceding a set), very low compared to the percentage before 20 days (25%) or any days for dFAD with no set performed (55%). The percentage of cluster 1 also decreased substantially during the period preceding a fishing set, until less than 5% the day before a fishing set. To the contrary, the percentage of clusters 2 and 3 increases rapidly in the days preceding a set, to reach 55% for cluster 3 and 35% for cluster 2 one day before a fishing day. Hence, we can consider that clusters 2 and 3 may indicate tuna biomass, while cluster 1, a mix of species that would likely include bycatch. Sets that led to higher catch (>50t) showed a very high percentage (>80%) of both clusters 2 and 3 over the last 5 days before a set, while smaller sets showed a progressive increase in the dominancy of clusters 2 and 3 the few days before the set (Figure S15).



**Figure 13.** Percentage of echosounder transmissions (around +/- 5 hours of sunrise, and within 10°N and 10°S) per clusters on days relative to a fishing set (only sets with >5t of tuna catch were considered).

This can be compared to trends before a dFAD visit that did not lead to a fishing set (Figure 14). It can be noted that the percentage of transmissions from cluster 1 is high in the days preceding a visit, indicating that the dFAD echosounder buoy was estimating some tuna biomass, potentially bycatch, but when the vessel came in the vicinity of the dFAD, it considered that it was not worth making a set.



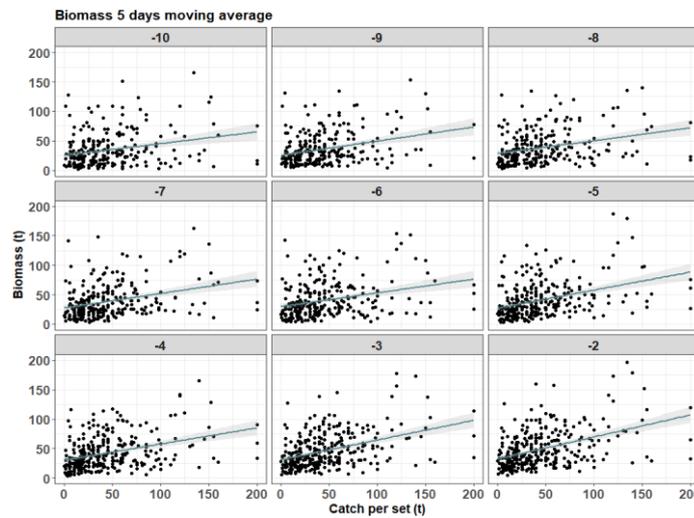
**Figure 14.** Percentage of echosounder transmissions (around +/- 5 hours of sunrise, and within 10°N and 10°S) per clusters on days relative to a fishing set (only sets with >5t of tuna catch were considered).

### 3.4 Relation between achieved catch and estimated biomass

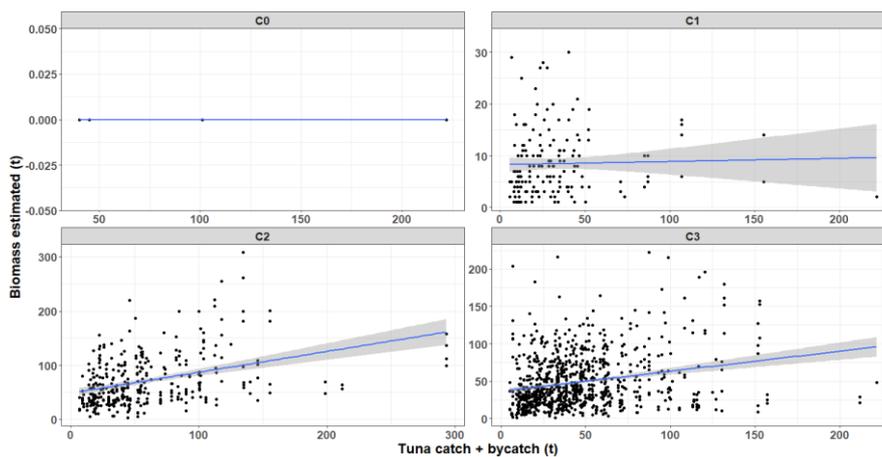
#### 3.4.1 Total catch

Linear relationships between catch and 5-day moving average of biomass from echosounder buoys were compiled but no clear pattern could be identified (Figure 15). Similarly, a linear relationship between total catch and bycatch, and the biomass estimated the five days before a fishing set was investigated and showed similar trends (Figure 16). In general, larger sets corresponded to larger estimated biomasses over a period of 5-days before a fishing set (Figure 17), although some variations are detected. It should be noted that, for very large sets, the echosounder often underestimated the

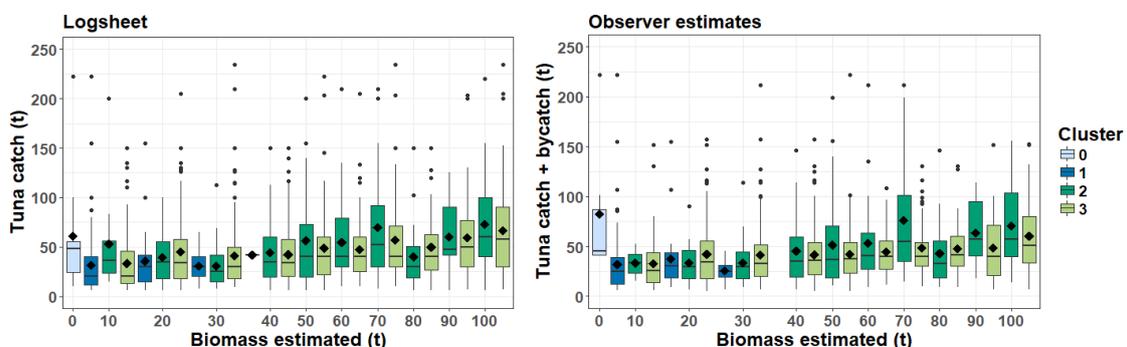
biomass or alternatively, it can be hypothesised that the whole tuna school was not detected under the cone of the echosounder. The low levels of bycatch compared to tuna catch, generally below 5 t, lead to very few differences when including the bycatch in the total catch of a set.



**Figure 15.** Relationship between catch per set (t) (from logsheet data) and the estimated biomass (t) from Satlink buoys over various periods preceding the set (note the low  $R^2$ , below 0.15 for all the regressions).

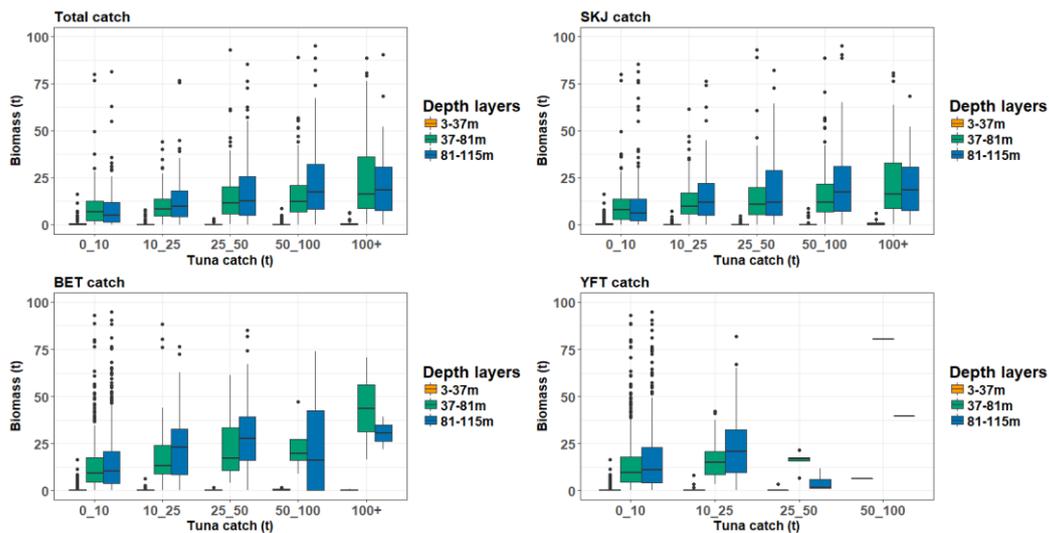


**Figure 16.** Relationship between total catch per set (t) (including targeted species and bycatch, from observer data) and the estimated biomass (t) from Satlink buoys over the 5-day period before a fishing set, for each cluster of echosounder transmission (note that the  $R^2$  was less than 0.15 for all regressions). Similar trends were found with tuna catch from logsheet data.



**Figure 17.** Tuna catch from logsheet data (left) and tuna catch and bycatch (from observer data) depending on levels of biomass estimated from the echosounder buoys over 5 days before a fishing set. Colors indicate the cluster of each biomass transmission.

Investigation of biomass estimates per depth layer and total catch or catch per species was also conducted (Figure 18). As mentioned previously, almost no biomass was detected for the shallower depths from 3 to 37m. For total catch and skipjack catch, higher biomass levels were detected for the largest sets, indicating that catch and estimated biomass could be related, at least for large aggregations. Biomass levels were higher in the 81–115m layer, except for very large sets where the 37–81m layer was of similar levels. For bigeye and yellowfin catch, higher biomass levels in the 91–115m layer were detected for intermediate catch (10 to 50t). No large yellowfin catches were available in the selected data set, while large bigeye catches corresponded to high biomass detected in the 37–81m layer (Figure 18).



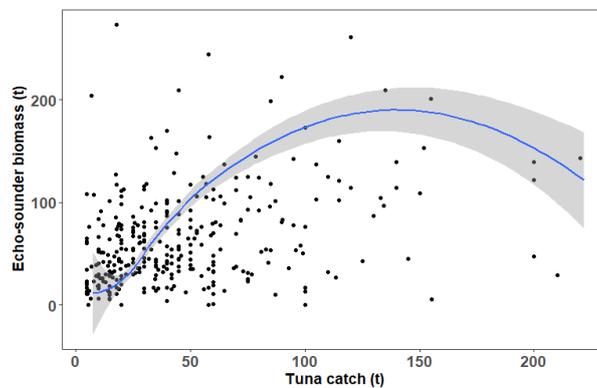
**Figure 18.** Daily biomass (t) estimated per depth layers over the 5 days prior to a fishing set function of total, skipjack, bigeye and yellowfin catch per set.

General additive models (GAMs, Hastie and Tibshirani, 1986) were used to investigate factors influencing total tuna catch. The available explanatory variables used were: the cluster or the biomass estimated by the echosounder the day preceding the set (average over 3 days and 5 days were also tested); biomass accumulation rates over 2 days and 20 days; the geographic area or the latitude and longitude; the time relative to sunrise; drift speed of the buoy; moon phase and time drifting. Note that only biomass transmission at +/- 5 hours of sunrise were considered. A delta model approach was used, first looking at variables influencing presence/absences of tuna in a binomial model, then when present, factors influencing tuna biomass caught were considered in a lognormal model. GAMs were run in R using the *mgcv* package (Wood, 2013).

A total of 314 fishing sets were considered for the models. However, the tuna presence/absence model could not be run because of the limited number of skunk sets (14 compared to 320 with tuna catch). The best models of tuna catch explained 25.9% of the deviance (Table 1). Tuna catch increased with time drifting, with high absolute values of biomass accumulation rates, around full moon, and for higher biomass detected by the echosounder (Figure 19 and S16). In particular, regarding tuna catch and biomass estimated by the echosounder, an increase was detected in tuna catch with increasing echosounder biomass, up to approximately 150 t; for larger sets the biomass was underestimated by the echosounder (Figure 19). Latitude and longitude were also significant, with higher tuna catch above the equator and to the East of the WCPO.

**Table 1.** GAM used to investigate the factors influencing tuna catch and bigeye tuna proportion in fishing sets, based on variables derived from echosounder buoy data.

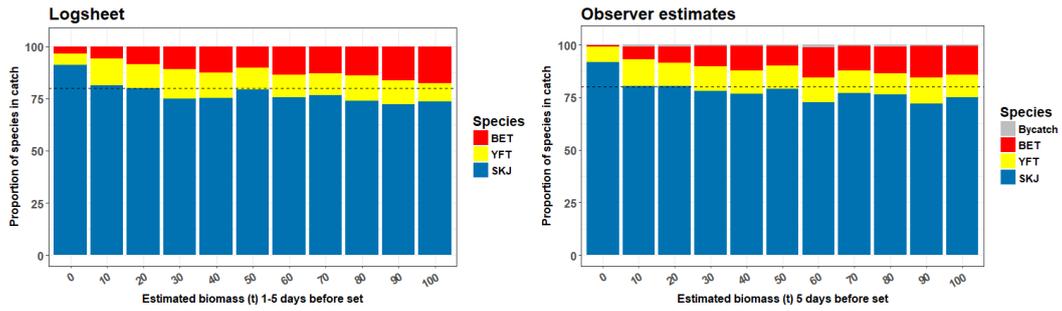
Model	Distribution	Explanatory variables	Deviance explained	R <sup>2</sup>
Tuna presence	Binomial	-		
Tuna catch	Log normal	Time drifting; moon phase; time relative to sunrise; 20-days biomass accumulation rate; 2-days biomass accumulation rate; biomass; latitude; ti(latitude, longitude)	25.9%	0.234
Proportion BET	Lognormal	Time drifting; time relative to sunrise; longitude, ti(latitude, longitude)	23.8%	0.197
Proportion YFT	Lognormal	Moon phase; biomass; longitude; ti(latitude, longitude)	27.0%	0.242



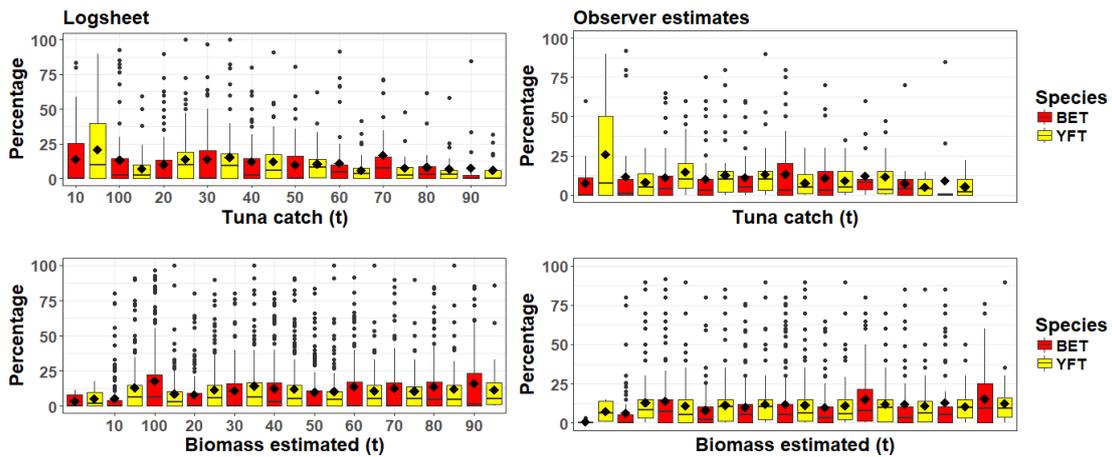
**Figure 19.** Dot plot of the relationship between tuna catch in a set and echosounder biomass estimated the day before (observed), the curve (loess) corresponds to the predicted data from the GAM.

### 3.4.1 Proportion of bigeye and yellowfin tuna

The proportion of bigeye and yellowfin tuna derived from the catch composition was examined in relation to the levels of total biomass estimated by the echosounder buoys in the days preceding a set (Figures 20 and 21). The proportion of yellowfin tuna in the catch appears relatively constant with increasing levels of estimated biomass (Figure 20). The proportion of bigeye tuna, however, increases with higher levels of biomass estimated by the echosounder buoy (Figures 20 and 21). This may be due to the presence of a swim bladder for yellowfin and bigeye tuna, while the echosounder is calibrated to detect skipjack tuna, which does not have a swim bladder (Moreno et al., 2019). There was also a tendency for the echosounder to underestimate the size of the school for very large sets, which might overestimate proportions as well. Finally, the limited number of sets considered here should be noted. The proportion of bigeye tuna from the catch composition for the sets considered was therefore assessed as well and did not show any trend with the level of actual total tuna catch (Figure 21).

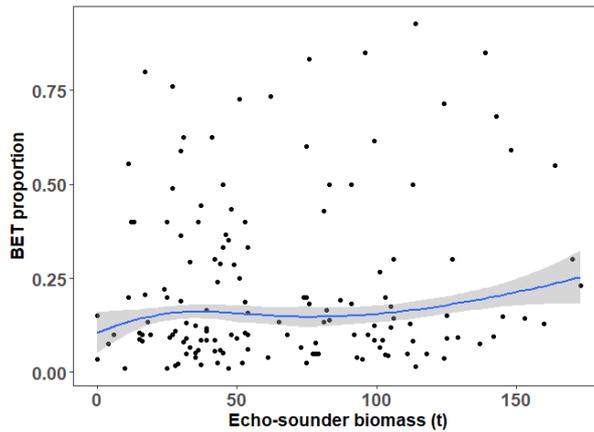


**Figure 20.** Average percentage of each tuna species and bycatch, depending on the level of biomass estimated over a 5-days period before a fishing set from logsheet (left) and observer data (right).

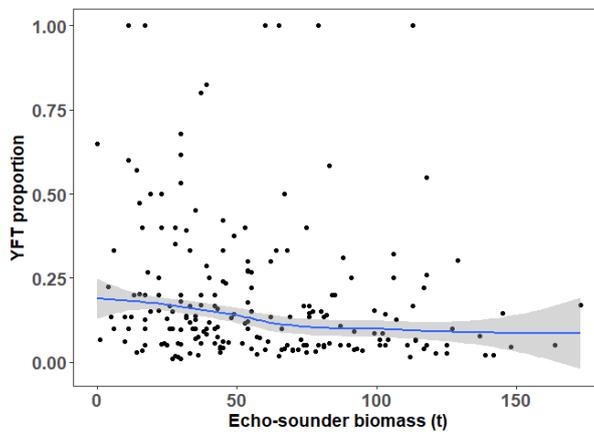


**Figure 21.** Percentage of bigeye and yellowfin tunas, depending on the level of catch (top) or biomass estimated over a 5-days period before a fishing set (bottom) from logsheet (left) and observer data (right).

GAMs were used to investigate factors influencing the proportion of bigeye and yellowfin tuna catch using the same data explanatory variables as for total catch, with a lognormal distribution. Models explained 23.8 and 27.0% of deviance for the bigeye tuna and yellowfin tuna models, respectively (Table 1). Higher bigeye tuna proportions were influenced by drift time around 40–60 days, decreasing biomass accumulation rates just before the set (2-days), time relative to sunrise (higher from 0 to 5h after sunrise) and area (higher in the east of the WCPO) (Figures 22 and S18). The biomass estimated by the echosounder, although showing a decreasing trend, was not a significant variable in the model. Higher yellowfin tuna proportions were found to be influenced by the moon (higher around new moon), longitude (higher in the west of the WCPO) and for lower biomass estimated from the echosounder (Figure 23 and S19). Note, however, the limited number of sets considered in these models (145 for bigeye and 204 for yellowfin).



**Figure 22.** Dot plot of the relationship between bigeye tuna proportion in a set and echosounder biomass estimated the day before (observed), the curve (loess) corresponds to the predicted data from the GAM.



**Figure 23.** Dot plot of the relationship between yellowfin tuna proportion in a set and echosounder biomass estimated the day before (observed), the curve (loess) corresponds to the predicted data from the GAM.

### 3.5 Identification of presence/absence of skipjack tuna

A combination of observer, dFAD acoustic transmissions, and Vessel Monitoring Systems (VMS) data were used to work towards developing reliable criteria to identify presence or absence of skipjack tuna, from the acoustic signals alone. This is an important research area to improve upon the estimation of relative abundance trends for skipjack tuna from the purse seine fishery. The process of assuming presence or absence of tuna species from acoustic dFAD buoy signals alone is not yet straightforward; therefore, several steps were taken to establish criteria for this determination.

#### 3.5.1 Identification of presence and absence

Using descriptive statistics and data visualization, we identified a subset of the data elements that showed contrast between skipjack presence and absence, based on the matched observer-dFAD acoustic and VMS-dFAD acoustic datasets. We developed a training data set associated with 'known' presence/absence of skipjack or that could be assumed with a high level of confidence. We then used a random forest classification (Breiman, 2001) to evaluate the classification accuracy using a suite of predictor variables from the acoustic data and the general fishing strategy.

Confidently identifying skipjack presence from the acoustic data alone is challenging. We first used the matched observer-dFAD acoustic sets. Matched sets that had less than 1 t of tuna were removed, as they were considered a failed set and unlikely to be representative of true abundance. We then

explored the catch composition, and if there was any amount of skipjack tuna caught, the set was classified as a presence, otherwise an absence, of skipjack. The main limitation to this approach for informing on presence/absence more broadly is that skipjack are caught in a very high proportion of purse seine sets, thus resulting in very few absences. To address this limitation, we used the VMS and acoustic signals to impute additional zeros, as described below.

Prediction of absences from the acoustic data is relatively straightforward because we can look to see which dFADs/buoys had no associated biomass. We can bolster this classification by evaluating biomass accumulation rates from the previous few days. We must recognize, however, that not all acoustic signals will be so clear, and to increase the utility of these data it is imperative to develop approaches to interpret noisier acoustic signals, relative to these research questions. For example, there may be an acoustic signal associated with non-target species only. This signal should be classified as an absence of tuna or skipjack more specifically, but with our simple classification criteria of no biomass, it would be left in the 'unknown' category.

To begin to address this uncertainty around acoustic signals, we incorporated VMS data to identify situations where we could assume that a dFAD was not holding skipjack (or tuna more generally) based on location, acoustic signal, and local fishing effort/vessel proximity. At this point, we have no way to verify these classifications, but the intention is to employ a conservative approach to further evaluate the potential of this methodology. Absence values were therefore imputed using the following criteria: a dFAD had low biomass (<5 t), a non-positive 5-day accumulation rate, and the vessel registered to the dFAD came within 5 km of its location and 5 hours from the associated transmission time, but did not set. All observations under consideration were between +/- 5 hours of sunrise. Based on these criteria, we feel reasonably confident that the dFAD was unlikely to be supporting a tuna school, and therefore, it was unlikely that there were skipjack at that location, on that day.

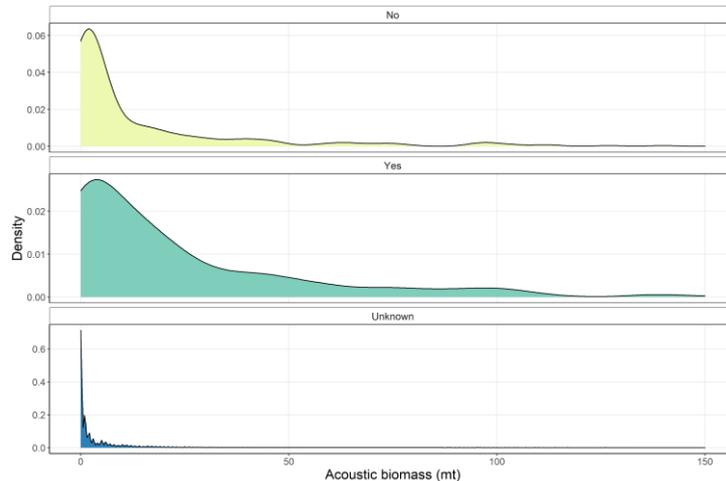
VMS data, from all vessels in the WCPO, were also used to estimate the local fishing effort surrounding a dFAD. For each buoy transmission, we estimated the number of purse seine vessels within 1, 5, 20, and 50km from the buoy location. This information was predicted to be potentially informative about relative productivity of fishing areas, using density of fishing vessels as a proxy; and therefore, potentially informative with respect to the predicted presence or absence of skipjack.

A random forest approach was then applied to the training dataset using the R package *randomForest* (Liaw and Wiener, 2002). The random forest model was evaluated using classification accuracy based on the percentage of observations recovered accurately and Cohen's  $\kappa$  statistic. Finally, the model was applied to the full acoustic data set to predict presence or absence of skipjack. Once the presence/absence data sets were predicted for the full data set, they were then combined in the standardization procedure with the purse seine catch-per-unit effort time series (see section 3.6).

### 3.5.2 Preliminary results

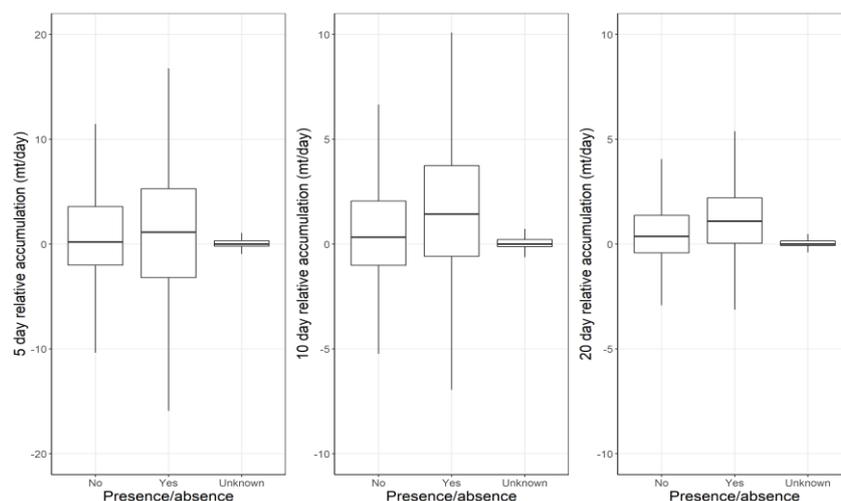
The combined matched and groomed data set (dFAD acoustic, observer, and VMS) included just over 1 million records and 525 matched observer sets. Of the matched sets retained in the filtered data set, 363 had skipjack catch while 7 reported no skipjack in the catch. The VMS data offered an additional 26 absences, which was insignificant given the size of the data set. The absence imputation from the acoustic signal, assuming that if the biomass estimate from the same day was zero and the previous day was either zero or unavailable (potentially suggesting a new deployment), then there was no skipjack there, added 377,787 absences to the training data set.

The main data elements explored for patterns related to skipjack presence included: sum of acoustic biomass, biomass accumulation rates at dFADs, dFAD drifting time, and local fishing effort. The buoys that were set on with positive skipjack catch rates generally showed a broader distribution over higher biomass estimates as compared to the observations that we were unable to match sets to (Figure 24). The sets without skipjack showed evidence of relatively high acoustic signals for some observations, although it is important to keep in mind that these groups have very few observations, and though there was no skipjack harvested, yellowfin and/or bigeye were harvested from those sets.



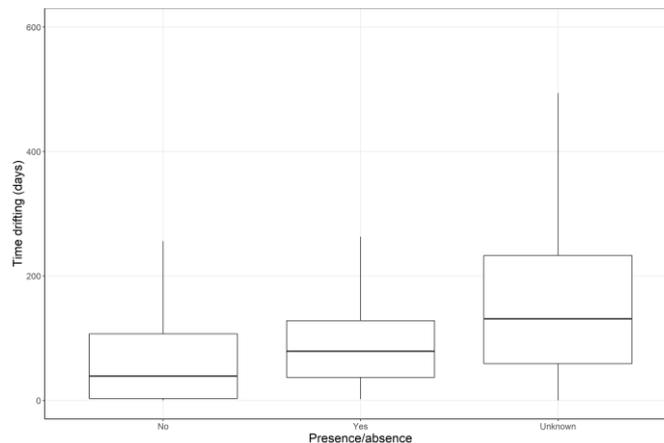
**Figure 24.** Distribution of acoustic biomass signal (aggregated across all depths), relative to whether skipjack was detected in observed sets (Yes) or not (No), or whether skipjack presence is unknown because a match to observer data was not found (Unknown). These data include only the observations from the dFAD-observer matched data set and does not include the imputed absences based on the acoustic signal.

Accumulation rates are considered as important for interpreting acoustic signals (Baidai et al., 2020). Here we see a strong signal in the biomass accumulation rates, indicating that when a set was made, the accumulation rates tended to be more variable than the dFADs with an unknown status. Of course, this is largely influenced by sample size, as many of the dFAD acoustic transmissions with an ‘unknown’ skipjack status will not be associated with any biomass (Figure 25). It is interesting to note that sets that caught skipjack were generally associated with more variable accumulation rates. Typically, we assume that biomass accumulation would be increasing prior to a set, but it may also be the case that a large decrease in biomass is an equally useful indicator of tuna or skipjack presence, as it suggests that a large school has been associated recently.



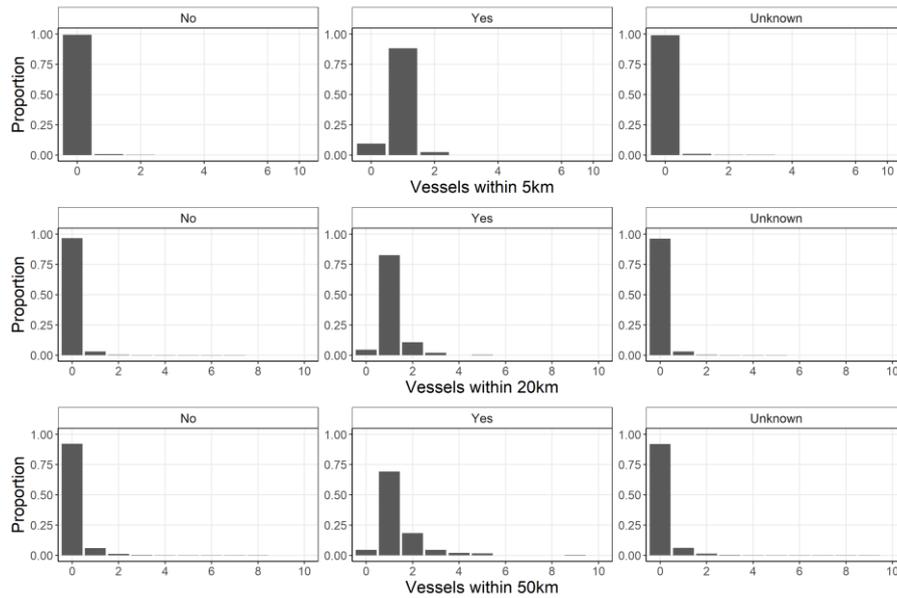
**Figure 25.** Distribution of relative biomass accumulation from acoustic dFAD signals over 5, 10, and 20 days, relative to whether skipjack was detected in observed sets (Yes) or not (No), or whether skipjack presence is unknown because a match to observer data was not found (Unknown). These data include only the observations from the dFAD-observer matched data set and does not include the imputed absences based on acoustic signal.

The amount of time a dFAD has been in the water has been used in other studies (Baidai et al., 2020; Santiago et al., 2019; Uranga et al., 2021) to filter out dFADs that may be unrepresentative as they have yet to be colonized (e.g., virgin segments). Here, we simply looked at the distribution of drifting time across the three categories of skipjack presence (Figure 26). For dFADs that were known to support skipjack, the drifting times ranged from 2 to 507 days, with a median of 79. For the dFADs holding tuna, but no skipjack, the drifting time ranged from 1 to 463 days, with a median of 21.



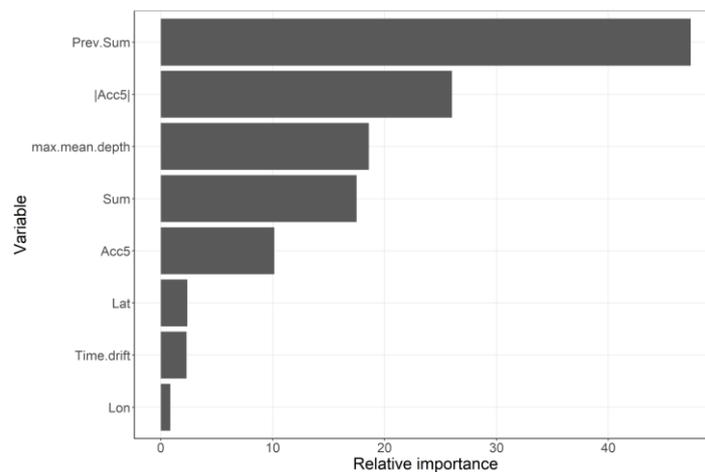
**Figure 26.** Distribution of dFAD drifting time (days), relative to whether skipjack was detected in observed sets (Yes) or not (No), or whether skipjack presence is unknown because a match to observer data was not found (Unknown). These data include only the observations from the dFAD-observer matched data set, and does not include the imputed absences based on acoustic signal.

We predicted vessel density to be potentially important with respect to skipjack presence as it may be an indication of local fish density. Fishers tend to aggregate in areas of relatively high density, and therefore, we may expect higher encounter rates in areas of high vessel density. Here, there is some indication of high vessel density in association with positive skipjack sets (Figure 27). The dFAD observations with no sets were generally in areas with lower vessel density, especially when looking at a 20 and 50km radius.



**Figure 27.** Distribution of proportion of vessel density (from the VMS database with all WCPO vessels) within 5, 20, and 50km from the dFAD (by row), relative to whether skipjack was detected in observed sets (Yes) or not (No), or whether skipjack presence is unknown because a match to observer data was not found (Unknown). These data include only the observations from the dFAD-observer matched data set and does not include the imputed absences based on the acoustic signal.

We randomly sampled 200 matched observations from activities where skipjack were classified as being present and 200 for which skipjack were classified as absent. Using a model with 150 trees, we evaluated the importance of each predictor variable included (Figure 28). We used a bootstrap approach ( $n=50$  iterations), to evaluate the sensitivity of the prediction accuracy to the random subsample of the data. All models produced a prediction accuracy of 99%. The  $\kappa$  statistic varied between 0.90 and 0.99, with a mean of 0.96. This level of accuracy is perhaps not surprising given the minimal variability in the acoustic signals associated with skipjack absence.



**Figure 28.** Mean variable importance identified by the random forest model, from the 50 bootstrap iterations.

Across the 50 model iterations, the acoustic biomass signal from the previous day was the most important predictor variable followed by the absolute value of the 5-day biomass accumulation rates, the mean depth of the maximum acoustic biomass signal, the acoustic biomass on the same calendar day, and the 5-day accumulation rate (Figure 28). The geographic location and dFAD drifting time offered little to the classification of presence/absence.

Given the relatively good performance of the classification models, we sampled the training data set at a higher rate to more fully capture all the information associated with confirmed presence events ( $n = 724$ ; 362 per classification category) to fit a random forest model and then used the model to predict presence/absence for the full acoustic data set ( $n \approx 1.04$  million). The model predicted approximately 411,000 records to be absent of skipjack while 622,000 were expected to have skipjack present. We then manually adjusted these assignments, based on conservative logic. Specifically, if the acoustic signal indicated no biomass on a given day as well as the previous day and the model estimated presence of skipjack, those were adjusted to be absences. Secondly, if the random forest model misclassified the presence for an observed set, we used the real observation instead. In total, we adjusted 92,828 records from presence to absence (about 11% of the data set).

This adjustment has benefits and limitations at this point. The first limitation being the distinction between tuna presence and skipjack presence. In our data set, most of the observer sets with tuna, but no skipjack, were classified as presence. We included the depth layer of the maximum acoustic signal in the random forest, but at this point those data are not very precise. Obtaining more detailed data at depth from dual or multi-frequency echosounders is expected to improve these classifications. Secondly, we have made some assumptions from the acoustic and VMS data about what an absence is, but uncertainty in those criteria may be complicating our classification as well. It should be noted that the multi-cluster approach described earlier in the paper to distinguish tuna from bycatch and no biomass at all, could be used to further refine the classification here, especially if we are able to distinguish yellowfin and bigeye signals (tunas with a swim bladder) from skipjack (a species without one).

### 3.5.3 Comparing with clusters of echosounder transmissions

Results from the random forest were compared to results from the clustering analysis for the same echosounder transmissions (Table 2). In the training dataset, almost all transmissions from cluster 2 and 3 were classified as skipjack presence (> 97%), and 83% of the transmission from cluster 1 as skipjack presence. 100% of transmissions from cluster 0, were classified as tuna absence. Results from the random forest classification on the whole echosounder dataset showed a very good match between both methods, more than 97% of transmissions in clusters 1, 2 and 3 were classified as skipjack presence. 92.7% of transmission from cluster 0 were classified as absence, a similar level as the training dataset. Finally, the random forest classification was adjusted to account for the fact that some tuna can be detected, but not necessarily skipjack. A more pronounced difference was then detected between results from the clustering and the random forest classification, in particular for clusters 0, 1 and 3 (Table 2).

**Table 2.** Comparison between presence (yes)/absence (no) and skipjack and clusters of tuna biomass.

Cluster	Training dataset				Estimated by random forest		Adjusted	
	No	Yes	% No	% Yes	% No	% Yes	% No	% Yes
0	377,793	72	100.0	0.0	92.7	7.3	82.6	17.4
1	22	105	17.3	82.7	0.0	100.0	22.2	77.9
2	2	76	2.6	97.4	0.0	100.0	5.7	94.3
3	3	110	2.7	97.3	2.3	97.7	12.6	87.4

### 3.6 Potential inclusion to estimate skipjack relative abundance index

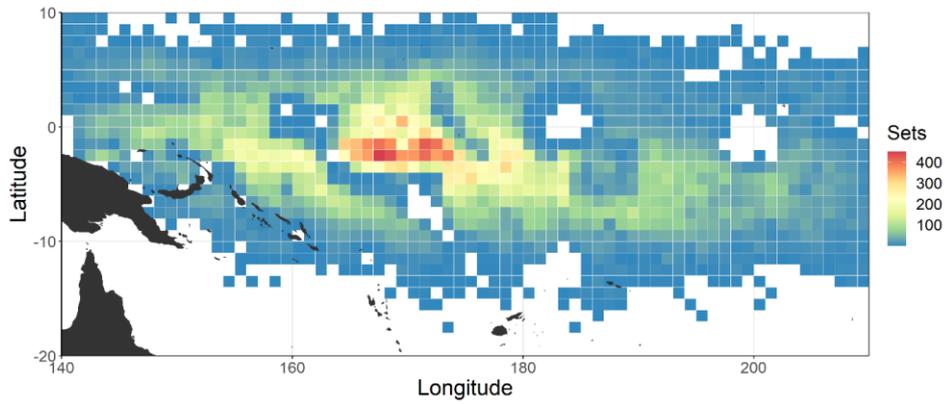
The previous section focused on identifying reliable signals in the dFAD acoustic data to detect presence or absence of skipjack tuna, using observer-reported and VMS data to develop these criteria. This information was then used to explore an integrated modeling approach that combined CPUE time series from the purse seine fishery with presence/absence data from acoustic dFAD buoys drifting throughout the WCPO. The idea behind combining these data sets is to provide a more realistic assessment of skipjack dynamics that are not entirely dependent on data from the fishery. Drifting dFADs provide a unique observation platform that could provide data throughout the spatial domain of interest, including spatial and temporal strata absent of fishing effort. Additionally, one of the challenges with purse seine CPUE time series, is that encounter probability for skipjack tuna is very high (excepting failed sets which are unlikely to be representative of an absence of tuna), largely because purse seiners do not set their nets unless a school is detected. As a result, predicted encounter probability from CPUE standardization models are likely to be inflated. Our hope is that by combining an additional source of information about skipjack presence and absence from fishery-independent data, we can develop more reliable estimates of relative abundance trends.

This approach is built on the work developed by Grüss and Thorson (2019), where they illustrated how combining data types (i.e., biomass, count, and encounter/non-encounter data) from different monitoring programs in the Gulf of Mexico improved the precision of reconstructed population trends and variables of habitat usage. For this analysis, the motivation for including presence/absence data was twofold: i) to better inform the model on ‘known’ absences, and ii) to incorporate realism in the encounter probability, across the spatial domain.

#### 3.5.1 Development of abundance indices

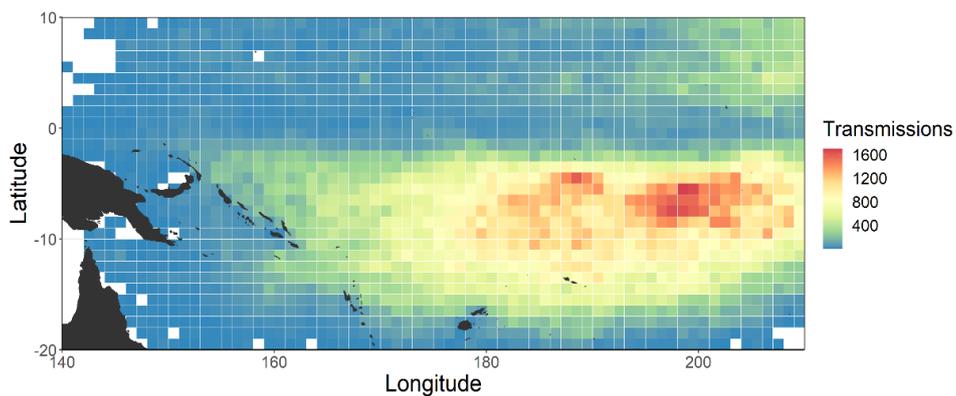
The spatiotemporal modeling approach is described in Appendix 4 and followed an approach developed by Grüss and Thorson (2019) by using a Poisson-link delta model (Thorson, 2018). The spatiotemporal model was fit using the *VASTR* package (Thorson et al., 2015). Two models were fitted: one that included the purse seine catch and effort data only, and a second that combined the CPUE data with the presence/absence data set. The results were compared. Although we do not know the true underlying biomass with certainty, we would not expect the overall index values from the two models to vary significantly; and if they do, it would likely represent a problem with the modeling framework and assumptions being made.

For the CPUE data we included only sets made around dFADs. Given that all acoustic data were obtained from dFADs, we wanted to ensure consistency in the analysis; however, it may be of interest to model both free-school and dFAD sets simultaneously in future model iterations. Figure 29 shows the distribution of dFAD fishing effort, by set type from 2016–2018. The distribution of skipjack catch over the time series was almost identical to the effort distribution.



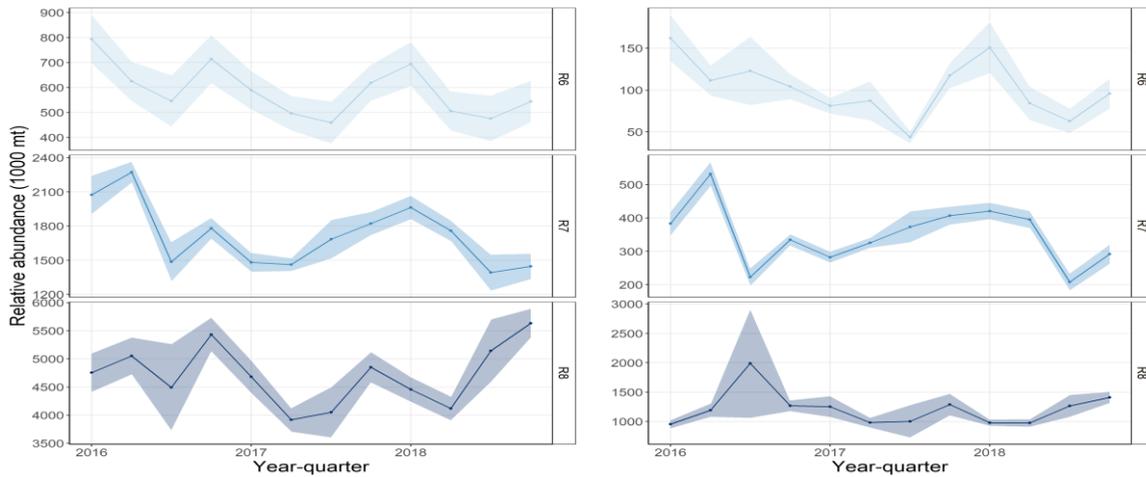
**Figure 29.** The spatial distribution of dFAD purse seine sets, from all vessels in the WCPO, used for the CPUE analysis, from 2016–2018.

The spatial distribution of the acoustic data is more comprehensive than the fishery data, within the core drifting zones (Figure 30). Throughout the equatorial region, there are few dFADs due to oceanic current patterns. DFADs are deployed and are drifting throughout the tropical waters north of the equator, but we have yet to develop data sharing agreements with the fleets and companies associated with those dFAD networks. Therefore, at this point in time, we do not have access to those data, creating an important data gap for these analyses.



**Figure 30.** Spatial distribution of acoustic transmissions used in the combined data analysis, from 2016–2018.

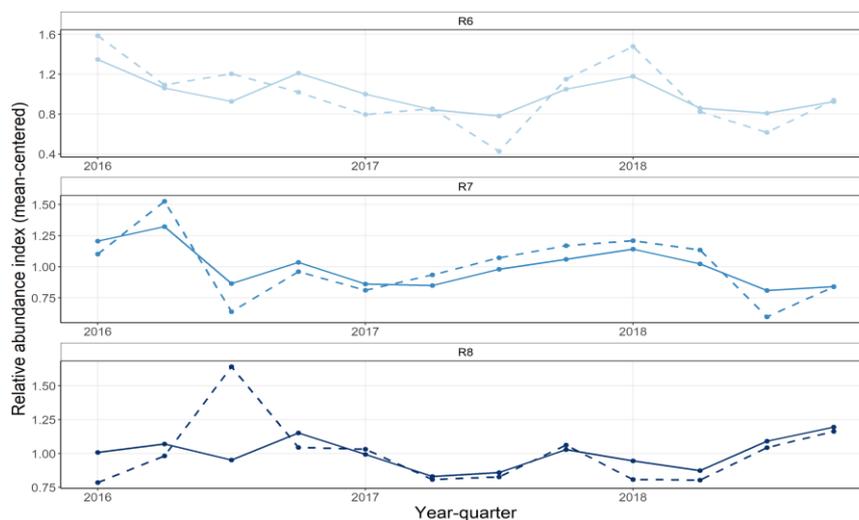
The CPUE-only model demonstrated relatively stable trends in relative abundance over time (Figure 31). It should be noted that there has been relatively little effort during the 3<sup>rd</sup> quarter of each year due to the dFAD closure period. There are, however, exemptions that allow dFAD fishing to occur, but those estimates are likely uncertain due to relatively small sample sizes. In regions 6 and 7 (Figure S7), we see a general decline in relative abundance, with a slight increase in Region 8 beginning around mid-year in 2017.



**Figure 31.** Area-weighted relative abundance indices for skipjack tuna from the CPUE-only model (left) and the combined data model (right), for the three regions of interest, from 2016–2018.

The combined data model predicted similar patterns across the three regions, with notable differences in the magnitude of relative abundance, based on the area-weighting of the indices (Figure 31). The predicted encounter probability was naturally much lower for the combined data set than predicted by the CPUE model alone. In addition, the impact of the additional data source is apparent during the dFAD closure period, during which the CPUE is not overly informative.

The mean-standardized indices of relative abundance enable a more straightforward comparison of the differences in the abundance trends estimated from the two different models (Figure 32). For much of the time series the two models estimate very similar trends, but in a few time steps, there are notable departures.



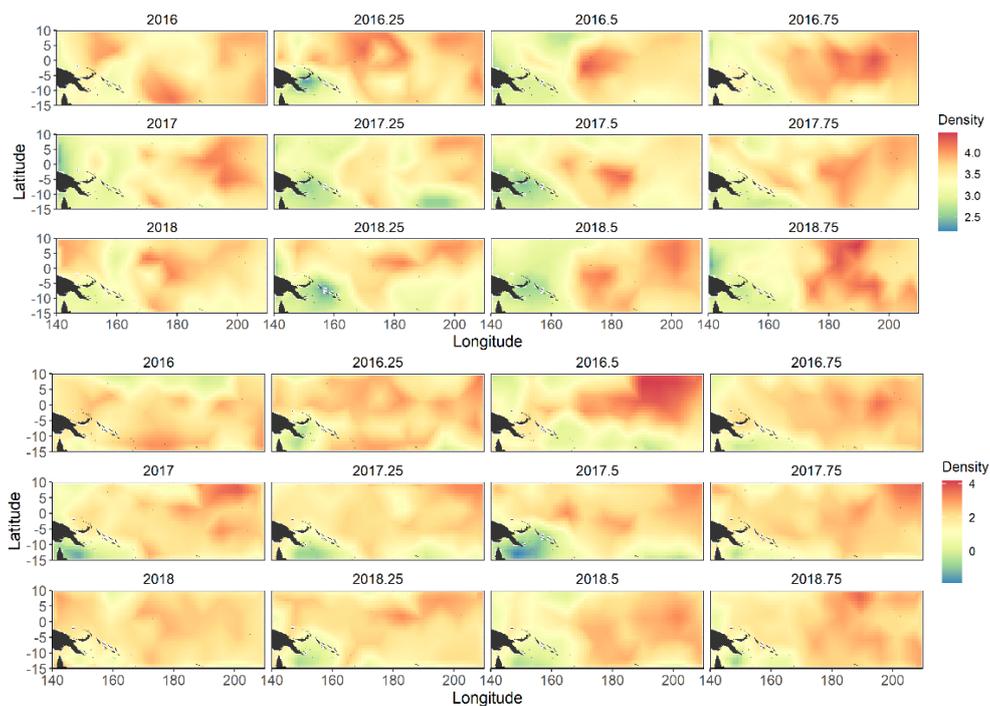
**Figure 32.** Mean-standardized index values from the CPUE only model (solid line) and the combined data (dashed line) model.

The estimates of spatial ( $\sigma_\epsilon$ ) and spatiotemporal variability ( $\sigma_\omega$ ) were comparable within each linear predictor; however, the spatial variation tended to be greater than the spatiotemporal variability, in both models (Table 3). These patterns are not necessarily unexpected given the relatively short time series we have modeled here. In addition, it should be noted that the variability for the first linear predictor, related to encounter probability, is much greater from the combined data model.

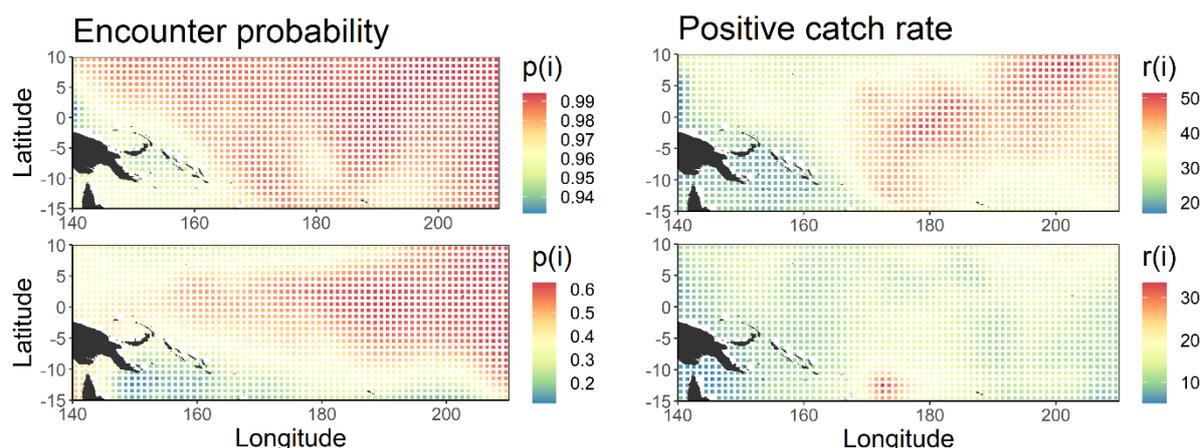
**Table 3.** Estimates of spatial and spatiotemporal variability from the two linear predictors for the CPUE-only and combined data models.

Model	Spatial variability		Spatiotemporal variability	
	1 <sup>st</sup> linear predictor	2 <sup>nd</sup> linear predictor	1 <sup>st</sup> linear predictor	2 <sup>nd</sup> linear predictor
<b>CPUE-only</b>	$\sigma_\varepsilon=1.20$	$\sigma_\varepsilon=0.84$	$\sigma_\omega=1.22$	$\sigma_\omega=0.73$
<b>Combined</b>	$\sigma_\varepsilon=2.50$	$\sigma_\varepsilon=0.58$	$\sigma_\omega=3.26$	$\sigma_\omega=0.58$

The distribution of skipjack density is also notably different between the two models, even though the trends in relative abundance are fairly similar (Figure 33). These differences arise largely from the information about skipjack presence from the dFAD network. When using catch and effort data alone, the estimated encounter probability is often greater than 0.9. This encounter probability is applied to the spatial domain, thereby assuming an almost guaranteed encounter of skipjack throughout the region. Skipjack is estimated to be a fairly abundant species, but we would not necessarily expect such a high encounter probability for a highly mobile pelagic fish species. When we combined the two data sources, we obtain what might be considered a more reasonable estimate of encounter probability, that maps more intuitively to the predicted spatial distribution of skipjack (Figure 34). The positive catch rates show similar patterns, but with a difference in magnitude.



**Figure 33.** Estimate (log) skipjack density, at the level of the extrapolation grid (i.e., 1°x1°) from 2016 through 2018, from the CPUE only model (top) and the combined data model (bottom).



**Figure 34.** Estimated encounter probability (left) and positive catch rates (right) from the CPUE only model (top) and the combined data model (bottom).

The exploration of a combined data approach, using the VAST framework, for the estimation of relative abundance indices is novel. This is a preliminary exploration of the methodology, with a relatively short-time series. These results are not intended to be informative at this stage, but rather an opportunity to evaluate alternative approaches to obtain more reliable information on trends in skipjack abundance. The skipjack assessment is hindered by what is often believed to be an uninformative purse seine CPUE index, and a lack of alternative time series to inform the model on abundance trends. Our hope is that by combining additional information on the presence and absence of skipjack, these models made prove more informative. At this stage, the information from the dFAD network is fairly limited relative to the number of dFADs in the WCPO, but also with respect to the spatial distribution. There remains much development and exploration into the utility of this combined data approach for estimating abundance indices, but these preliminary results show promise to further explore the utility for potential use in the upcoming skipjack assessment, it would be valuable to secure greater access to the data from the dFAD network drifting across the whole of the WCPO.

## 6. Discussion and Conclusion

This report presents analyses using data from over 3.8 million acoustic transmissions from Satlink echosounder buoys deployed on dFADs in the WCPO from 2016 to 2018, provided under a cooperative agreement with three commercial fishing companies. The three brands of echosounder buoys presented different characteristics, including the basic algorithm used to estimate total aggregated biomass, making inter-brand comparison difficult. Given the data available and their characteristics, analyses presented in this report focused on Satlink buoys. Colonization and recolonization (i.e., after a fishing set) patterns were studied using echosounder and catch data (from logsheet and observer data). Refined analyses to more precisely identify sets and other dFAD-related activities performed on buoys with available echosounder data allowed a dFAD's 'life history' to be followed, and to better characterise biomass levels in relation to catch per set. Biomass estimates in relation to total catch per set were investigated, as well as relative indices of tuna biomass and presence/absence of skipjack tuna.

It was noted that the ability to accurately track a dFADs 'life history' (not only the buoy) was challenging here and could have significant implications for integration of dFAD associated

information into regional stock assessments. In addition, many factors may influence both the echosounder estimated biomass and the catch per set. First, the whole school might not be detected by the echosounder, or not completely caught during a set. Second, a mix of species will influence the echosounder readings; the algorithm for estimating biomass is programmed to estimate biomass for schools of skipjack only. In particular, bycatch species will also be detected by the echosounder, or yellowfin and bigeye tuna might have a stronger influence on the acoustic signal, due to the presence of a swim bladder.

This study continues to increase our knowledge of biomass accumulation dynamics and of the signal captured by echosounder buoys. A big challenge within this project relates to identifying, with the most precision, fishing activities that are performed on the trajectories available. Although that information exists in the fishery data, identification of a precise record of buoy ID number remains a challenge. Accordingly, a subset of the dataset was selected, for which trajectory and fishing activities corresponded, with high confidence, to the same buoy attached to a dFAD. Several parameters were compiled here, that could be used in further analyses performed on the dataset. These parameters included: dFAD soak time, drift speed, trends in biomass accumulation leading to a fishing set, biomass moving averages, biomass by depth, total catch or catch per species, moon phase, time of the set, and spatial areas. Bycatch per set, as well as tuna discards, as recorded by observers were also added to better characterize the whole catch per set and compare it to the biomass estimated by the echosounder buoy.

Following the two main objectives of the project, the investigations focused on:

- 1) the potential use of acoustic buoys on dFADs to provide new fishery-independent data for stock assessments;
- 2) the link between high aggregated biomass and the level of small of bigeye and yellowfin in the catch.

Two alternative methods were developed to classify echosounder transmissions as a relative index of tuna abundance or as presence/absence of skipjack tuna. A very good match in results between the two methods was found. The first one distinguished three clusters of fish biomass indicating different profiles at depth and level of aggregated biomass, and a fourth cluster of tuna absence. This could be used as an indication of tuna biomass, however, it is not informative at this stage regarding species composition. The second method, however, has been tailored directly to detect skipjack tuna presence and absence, but the currently available biomass data could not be used to evaluate skipjack biomass. Previous analyses have explored presence/absence of tuna using a similar approach (Baidai et al., 2020) or the possibility to derive an index of abundance of skipjack tuna using a combination of echosounder buoy data and species composition and average size over large spatial areas. (Uranga et al., 2021). Accessing multi-frequency echosounder data might allow for species discrimination and investigation of estimated biomass per species (Diallo et al., 2019; Moreno et al., 2019).

The classification of presence/absence of skipjack tuna was then used to explore an integrated standardization approach that combines CPUE time series from the purse seine fishery with presence/absence data from acoustic dFAD buoys drifting throughout the WCPO. This was compared to a standardization model with purse seine catch and effort data only. This preliminary exploration of the methodology was based on a relatively short time-series and data from three fishing companies only with specific operational regions. Hence, results are not intended to be informative at this stage,

but rather an opportunity to evaluate alternative approaches to obtain more reliable information on trends in skipjack abundance. Larger and more comprehensive datasets are needed to fully explore the approach developed.

Analyses presented in this paper are based on a small subset of the full dFAD network in the WCPO, for three years only. An expansion of the current study, with a larger, longer and more comprehensive dataset from multiple fleets, should be promoted. In addition, accessing multi-frequency echosounder data would also allow additional investigation to be performed, in particular linked to species composition and proportion of bigeye and yellowfin tuna in aggregated tuna schools.

We invite WCPFC-SC17 to:

- Note the results from Project 88 on acoustic data from echosounder buoys deployed on dFADs.
- Note that while the acoustically estimated biomass related to the proportion of bigeye in the catch, this proved non-significant within models, but statistical models showed a slight decrease in yellowfin proportion for larger estimated biomass and other identified variables may show some promise to pre-identify sets likely to lead to greater proportions of bigeye and yellowfin in the catch.
- Note the potential, over the longer-term, to use echosounder data as a source of fishery-independent data for stock assessments, either as an independent relative index of abundance or to provide additional information for purse seine CPUE standardization.
- Recommend the need for better identification of particular dFAD buoys (e.g., via the buoy identification numbers) by commercial vessel operators or via observer reports.
- Endorse the continued cooperative relationship with the fishing community to obtain commercially sensitive data for analysis for the purpose of scientific and other research, particularly with regard to dFADs, and the fishing strategies involved in their use. Highlight the need for additional data covering the whole WCPFC convention area, including that from now available multi-frequency echosounder buoys, and encourage other industry partners to become involved in the project.

## Acknowledgments

The authors would like to thank Cape Fisheries, National Fisheries Developments (NFD) and South Pacific Tuna Corporation fishing companies for giving us access to their data for this analysis. We also acknowledge Satlink, Zunibal and Kato buoy companies for their collaboration and helpful support in data export and processing. WCPFC Project 88 ‘Acoustic FAD analyses’ is carried out with funds from the European Union, and the WCPFC. We thank Nan Yao for valuable comments on an earlier version of the paper.

## References

- Baidai, Y, Dagorn, L., Amandè, M.J., Gaertner, D., Capello, M., 2020. Tuna aggregation dynamics at Drifting Fish Aggregating Devices: a view through the eyes of commercial echosounder buoys. *ICES J. Mar. Sci.* 77, 2960–2970. <https://doi.org/10.1093/icesjms/fsaa178>
- Baidai, Yannick, Dagorn, L., Amandè, M.J., Gaertner, D., Capello, M., 2020. Tuna aggregation dynamics at Drifting Fish Aggregating Devices: a view through the eyes of commercial echosounder buoys.

ICES J. Mar. Sci. <https://doi.org/10.1093/icesjms/fsaa178>

- Breiman, L., 2001. Random Forests. *Mach. Learn.* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Diallo, A., Baidai, Y., Mannocci, L., Capello, M., 2019. Towards the derivation of fisheries-independent abundance indices for tropical tuna: Report on biomass estimates obtained from a multi-frequency echosounder buoy model (M3I+). IOTC Tech. Rep. IOTC-2019-WPTT21-54.
- Escalle, L., Hare, S.R., Vidal, T., Brownjohn, M., Hamer, P., Pilling, G., 2021a. Quantifying drifting Fish Aggregating Device use by the world's largest tuna fishery. *ICES J. Mar. Sci.* <https://doi.org/10.1093/icesjms/fsab116>
- Escalle, L., Muller, B., Brouwer, S., Pilling, G., 2018. Report on analyses of the 2016/2018 PNA FAD tracking programme. WCPFC Sci. Comm. WCPFC-SC14-2018/MI-WP-09.
- Escalle, L., Muller, B., Hare, S., Hamer, P., Pilling, G., PNAO, 2020a. Report on analyses of the 2016/2020 PNA FAD tracking programme. WCPFC Sci. Comm. WCPFC-SC16-2020/MI-IP-14.
- Escalle, L., Muller, B., Hare, S., Hamer, P., PNAO, 2021b. Report on analyses of the 2016/2021 PNA FAD tracking programme. WCPFC Sci. Comm. WCPFC-SC17-2021/MI-IP-04.
- Escalle, L., Muller, B., Scutt Phillips, J., Brouwer, S., Pilling, G., PNAO, 2019a. Report on analyses of the 2016/2019 PNA FAD tracking programme. WCPFC Sci. Comm. WCPFC-SC15-2019/MI-WP-12.
- Escalle, L., Vanden Heuvel, B., Clarke, R., Brouwer, S., Pilling, G., 2019b. Report on preliminary analyses of FAD acoustic data. WCPFC Sci. Comm. WCPFC-SC15-2019/MI-WP-13.
- Escalle, L., Vanden Heuvel, B., Clarke, R., Dunham, R., Pilling, G., 2020b. Updates on Project 88: FAD acoustics analyses. WCPFC Sci. Comm. WCPFC-SC16-2020/MI-IP-20.
- Fraley, C., Raftery, A.E., 2002. Model-Based Clustering, Discriminant Analysis, and Density Estimation. *J. Am. Stat. Assoc.* 97, 611–631. <https://doi.org/10.1198/016214502760047131>
- Fraley, C., Raftery, A.E., Murphy, T.B., Scrucca, L., 2012. Mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation. Tech. Rep. No. 597, Dep. Stat. Univ. Washington. <https://www.stat.washington.edu/research/reports/2012/tr597.pdf>.
- Grüss, A., Thorson, J.T., 2019. Developing spatio-temporal models using multiple data types for evaluating population trends and habitat usage. *ICES J. Mar. Sci.* 76, 1748–1761. <https://doi.org/10.1093/icesjms/fsz075>
- Hamilton, R.J., Almany, G.R., Stevens, D., Bode, M., Pita, J., Peterson, N.A., Choat, J.H., 2016. Hyperstability masks declines in bumphead parrotfish (*Bolbometopon muricatum*) populations. *Coral Reefs* 35, 751–763. <https://doi.org/10.1007/s00338-016-1441-0>
- Harley, S., Williams, P., Hampton, J., 2009. Analysis of purse seine set times for different school associations: a further tool to assist in compliance with FAD closures? *West. Cent. Pacific Fish. Comm. 5th Regul. Sess. WCPFC-SC5-2009/ST-WP-07*, Port Vila, Vanuatu.
- Hastie, T., Tibshirani, R., 1986. Generalized Additive Models. *Stat. Sci.* 1, 297–310.
- Hijmans, R.J., 2019. R package “geosphere”. Spherical trigonometry.
- Hoyle, S.D., Langley, A.D., Campbell, R.A., 2014. Recommended approaches for standardizing CPUE

- data from pelagic fisheries. West. Cent. Pacific Fish. Comm. 10th Regul. Sess. WCPFC-SC10-2014/SA-IP-10.
- Kaufman, L., Rousseeuw, P.J., 1990. Finding Groups in Data: An Introduction to Cluster Analysis, Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, USA. <https://doi.org/10.1002/9780470316801>
- Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. R News 2, 18–22.
- Lopez, J., Moreno, G., Boyra, G., Dagorn, L., 2016. A model based on data from echosounder buoys to estimate biomass of fish species associated with fish aggregating devices. Fish. Bull. 114, 166–178. <https://doi.org/10.7755/FB.114.2.4>
- Lopez, J., Moreno, G., Sancristobal, I., 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. Fish. Res. 155, 127–137. <https://doi.org/10.1016/j.fishres.2014.02.033>
- Maufroy, A., Chassot, E., Joo, R., Kaplan, D.M., 2015. Large-scale examination of spatio-temporal patterns of drifting fish aggregating devices (dFADs) from tropical tuna fisheries of the Indian and Atlantic Oceans. PLoS One 10, 1–21. <https://doi.org/10.1371/journal.pone.0128023>
- Maunder, M.N., Punt, A.E., 2004. Standardizing catch and effort data: a review of recent approaches. Fish. Res. 70, 141–159. <https://doi.org/10.1016/J.FISHRES.2004.08.002>
- Moreno, G., Boyra, G., Sancristobal, I., Itano, D., Restrepo, V., 2019. Towards acoustic discrimination of tropical tuna associated with Fish Aggregating Devices. PLoS One 14, e0216353. <https://doi.org/10.1371/journal.pone.0216353>
- Moreno, G., Dagorn, L., Capello, M., Lopez, J., Filmalter, J., Forget, F., Sancristobal, I., Holland, K., 2016. Fish aggregating devices (FADs) as scientific platforms. Fish. Res. 178, 122–129. <https://doi.org/10.1016/J.FISHRES.2015.09.021>
- Panizza, A., Williams, P., Falasi, C., Loganimoce, E., Schneiter, E., 2021. Status of observer data management. WCPFC Sci. Comm. WCPFC-SC17-2021/ST-IP-02.
- Santiago, J., Uranga, J., Quincoces, I., Orue, B., Grande, M., Murua, H., Merino, G., Urtizbera, A., Pascual, P., Boyra, G., 2019. A novel index of abundance of Juvenile yellowfin tuna in the Indian Ocean derived from echosounder buoys. IOTC Tech. Rep. IOTC-2019-WPTT19-45.
- Scrucca, L., Fop, M., Murphy, T.B., Raftery, A.E., 2016. mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. R J. 8, 289–317.
- Thorson, J.T., 2018. Three problems with the conventional delta-model for biomass sampling data, and a computationally efficient alternative. Can. J. Fish. Aquat. Sci. 75, 1369–1382. <https://doi.org/10.1139/cjfas-2017-0266>
- Thorson, J.T., Cunningham, C.J., Jorgensen, E., Havron, A., Hulson, P.-J.F., Monnahan, C.C., von Szalay, P., 2021. The surprising sensitivity of index scale to delta-model assumptions: Recommendations for model-based index standardization. Fish. Res. 233, 105745. <https://doi.org/10.1016/J.FISHRES.2020.105745>
- Thorson, J.T., Shelton, A.O., Ward, E.J., Skaug, H.J., 2015. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. ICES J. Mar. Sci. 72, 1297–1310. <https://doi.org/10.1093/icesjms/fsu243>
- Uranga, J., Lopez, J., Grande, M., Lennert-Cody, C.E., Quincoces, I., Granado, I., Maunder, M.N., Aires-

- da-Silva, A., Merino, G., Murua, H., Santiago, J., 2021. Tropical tuna biomass indicators from echosounder buoys in the Eastern Pacific Ocean. Inter-American Trop. Tuna Comm. Ad-hoc Perm. Work. Gr. FADs. Fifth Meet. FAD-05-INF-E.
- Vincent, M., Pilling, G., Hampton, J., 2019. Stock assessment of skipjack tuna in the western and central Pacific Ocean. West. Cent. Pacific Fish. Comm. 15th Regul. Sess. WCPFC-SC15-2019/ SA-WP-05.
- Wain, G., Guéry, L., Kaplan, D.M., Gaertner, D., 2021. Quantifying the increase in fishing efficiency due to the use of drifting FADs equipped with echosounders in tropical tuna purse seine fisheries. ICES J. Mar. Sci. 78, 235–245. <https://doi.org/10.1093/icesjms/fsaa216>
- Wichman, M.-O.-T.-A., Vidal, T., 2021. Incorporating industry knowledge to understand purse seine effort creep and evolution of fishing strategies in the WCPO. WCPFC Sci. Comm. SC17-2021/MI-WP-07.
- Williams, P., Ruaia, T., 2021. Overview of tuna fisheries in the western and central Pacific Ocean, including economic conditions - 2020. WCPFC Sci. Comm. SC17-2021/GN-IP-01.
- Wood, S.N., 2013. Package “mgcv” - CRAN. Mixed GAM Computation Vehicle with GCV/AIC/REML Smoothness Estimation. Version 1.7-27. <http://cran.r-project.org/web/packages/mgcv>.

## Appendix 1. General description of the data and processing method

### S1.1 Available acoustic datasets

Acoustic and position data from three different satellite echosounder buoy providers, Satlink, Zunibal and Kato, were available.

Satlink buoys transmitted acoustic (echosounder readings) and position information separately, with generally two transmissions of position data and three acoustic readings (typically around sunrise) per day (Table S1). In order to access the position related to each acoustic transmission, we linearly interpolated the position dataset at those times. Hence, for each echosounder reading, we had access to estimated position, date/time, processed total biomass estimates (t) and biomass estimates at 11.2m depth intervals or bins from 3 to 115m.

Zunibal buoys transmitted every hour with data for both the position of the buoy and echosounder readings (Table S1). Transmissions included position, date/time and total estimated biomass (t). However, data received included fixed position per day (same latitude and longitude throughout the whole day, while normally one position per hour should be available), which led to very high and unrealistic drift speeds between days. In order to process the data (see following section), we therefore retained one transmission per day, corresponding to the highest biomass reading. For some of the data (provided by one of the fishing companies), transmissions included raw echosounder data at 1.6m depth bins from 1.6 to 120m. Finally, Zunibal buoys includes a sensor archiving in/out water position, allowing access to buoy deployment position, which was also available for a subset of the Zunibal data only (Table S1).

**Table S1.** Summary of data parameters per echosounder buoy by commercial brand.

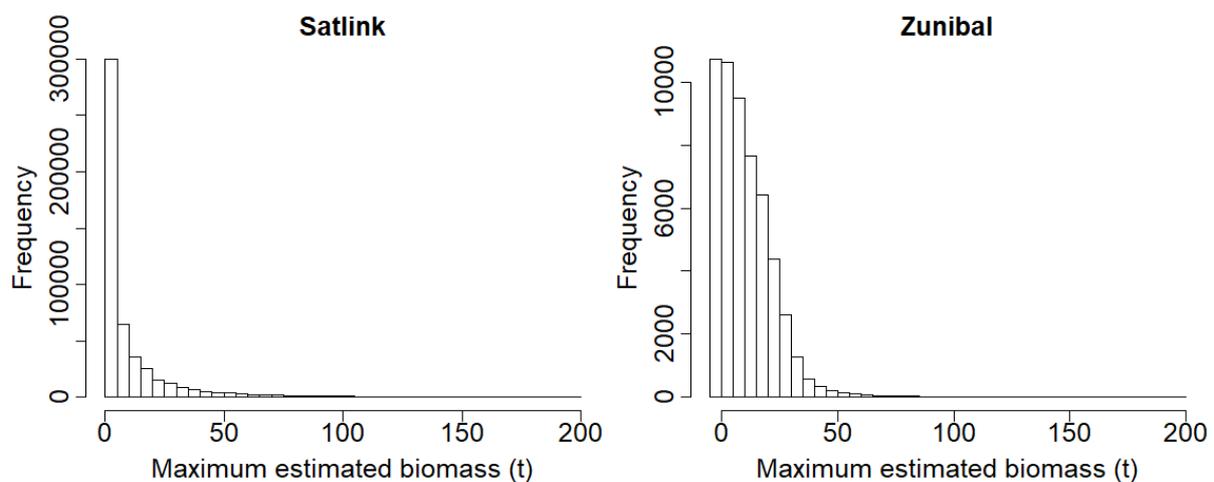
Brand	Satlink	Zunibal	Kato
<b>Year</b>	2016–2018	2016–2018	2017–2018
<b>Frequency of echosounder readings</b>	Generally 3 per day (range 0-4)	Generally 1 per hour with associated position	1 per day with associated position
<b>Frequency of position transmission</b>	Generally 2 per day	Generally 1 per hour with associated acoustic data	1 per day with associated acoustic data
<b>Number of echosounder transmissions</b>	3,813,009	811,006	61,859
<b>Biomass estimates</b>	Total estimates in tons	Total estimates in tons	No
<b>Biomass estimates per depth bin</b>	Yes 11.2m; from 3 to 115m	No	Raw data (strength of echosounder 6–100%)
<b>Echosounder raw data</b>	No	Occasionally by depth bin of 1.6m; from 1.6 to 120m	Yes, by depth bin of 10m; from 0 to 150m
<b>Buoy deployment position</b>	No	Occasionally	No

Kato buoys transmitted once per day with data for both the position of the buoy and echosounder readings at the same time (Table S1). Transmissions included position, date/time and a signal

corresponding to the echosounder reading. This corresponds to the acoustic signal strength, with a relative index between 0 and 7 given for each 10m depth bin between 0 and 150m.

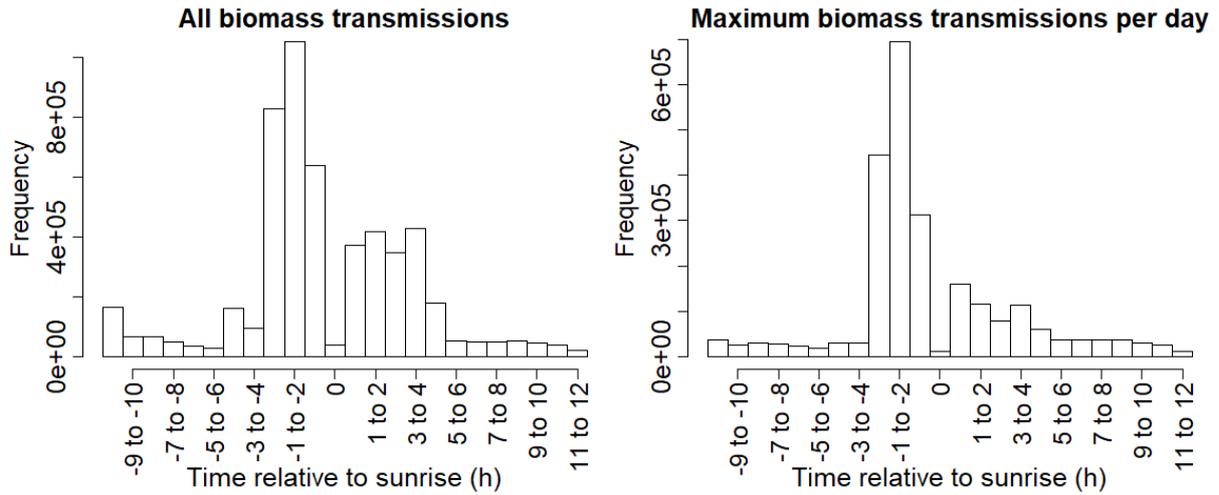
## S1.2 Maximum estimated biomass and influence of the time of the day

The biomass estimated by Satlink and Zunibal echosounder buoys (no estimates in tons for Kato buoys) ranges from 0 to 350t (Figure S1). However, the profile of the distribution of maximum daily biomass estimates was different between Satlink and Zunibal buoys. For the majority of days, the maximum estimated value from Satlink buoys was between 1 and 5t. Note that when the echosounder estimated a biomass of less than 1t, no acoustic signal was sent, only a transmission of the position was received. The distribution of maximum daily biomass estimates decreases gradually from 5 to 100t. Zunibal buoys presented maximum daily estimated biomass mostly between 0 and 25t, then a gradual decrease in the distribution of maximum daily biomass from 30 to 100t was seen (Figure S1).

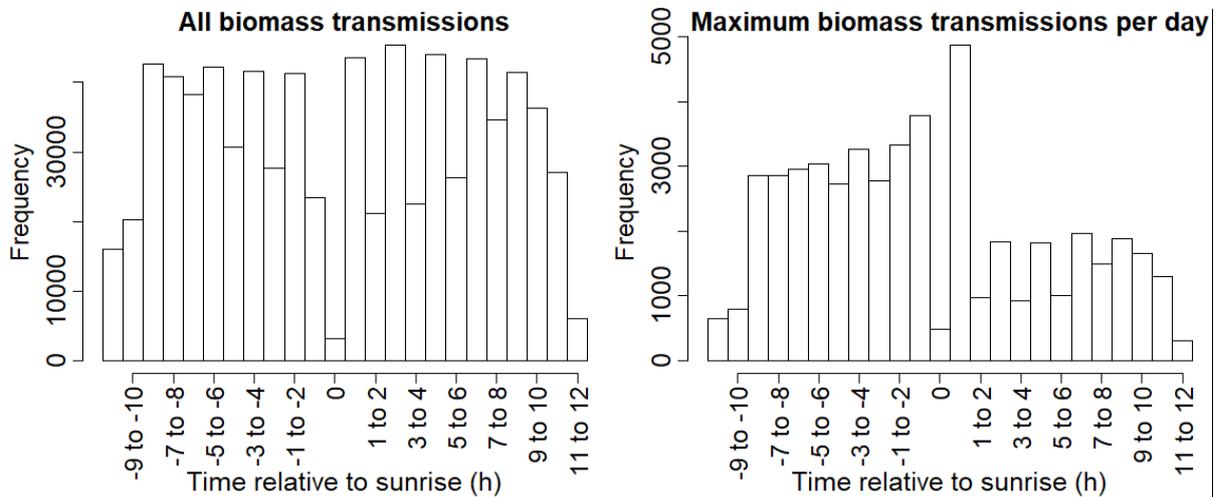


**Figure S1.** Maximum estimated biomass during the day for the Satlink and Zunibal echosounder buoys. Values above 200 t (0.4% of all values) were removed from the histogram to increase interpretability.

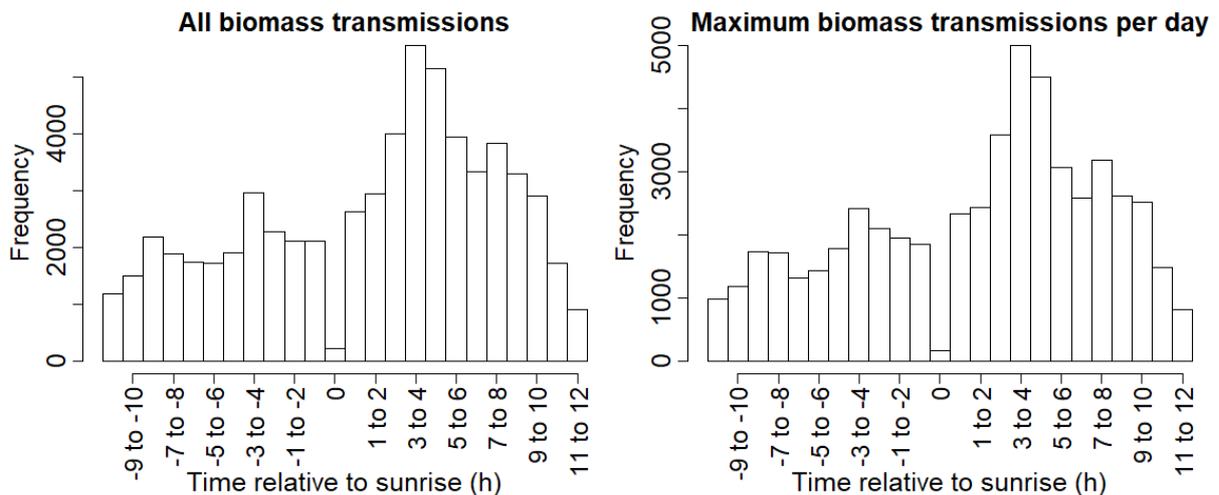
Higher numbers of biomass readings were found during the three hours before sunrise for the Satlink buoys (Figure S2). Similarly, the maximum daily biomass also most commonly corresponded to this period before sunrise (Figure S2). Zunibal and Kato buoys showed different patterns. Zunibal transmitted biomass regularly throughout the day, with a maximum daily biomass being typically just after sunrise (Figure S3). Kato buoys transmitted one biomass reading per day, hence explaining the similarity in the histograms between the biomass transmissions and the maximum daily biomass transmission (Figure S4). The maximum daily biomass was found from 3 to 5 hours after sunrise (Figure S4). However, it should be noted that for Kato buoys, no general estimate across the whole echosounder detection cone is given, but only some index by depth bin. Adding those indices might not therefore be relevant, and additional exploratory analyses are needed.



**Figure S2.** Time relative to sunrise of acoustic data transmission (left) and of the maximum biomass estimated during the day (right) for Satlink buoys.



**Figure S3.** Time relative to sunrise of acoustic data transmission (left) and of the maximum biomass estimated during the day (right) for Zunibal buoys.



**Figure S4.** Time relative to sunrise of acoustic data transmission (left) and of the maximum biomass estimated during the day (right) for Kato buoys.

## S1.2 Drift speed

Drift speed of echosounder buoys showed some variability with latitude and longitude (Figure S5), which may influence the acoustic signal, with higher, potentially overestimated, biomass detected at higher drifting speed.

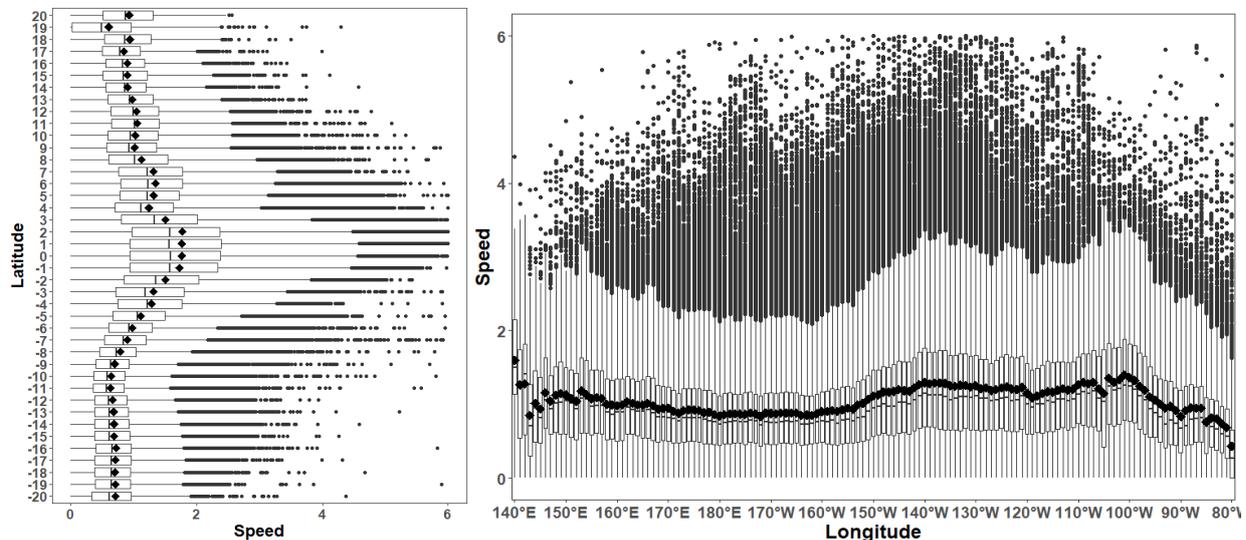


Figure S5. Variability in dFAD drift speed with latitude and longitude.

## S1.3 Identification of grouped deployments and precise dFAD drifting duration

Buoy deployment positions were identified using the Random Forest algorithm previously described and defined as the first at-sea position of a trajectory or of an at-sea segment (if more than one deployment per buoy was detected). This allowed the compilation of buoy drift time. However, more importantly for the study of tuna aggregation processes or re-aggregation processes after a fishing set, the actual initial deployment of the dFAD itself should also be identified. dFAD and buoy initial deployments and hence drift time, might be different as buoys can be deployed or re-deployed on dFADs found at-sea.

In the acoustic dataset received, there is no information regarding deployments and time drifting. Therefore, dFAD deployments could be identified using records made by observers and matched to acoustic time series using buoy ID numbers. However, these identifiers remain rarely recorded. Another method to identify dFAD deployments is to look at multiple buoy deployments around the same time and location ('grouped deployments'). Deployment of new dFADs (compared to buoys deployed on a dFAD found at-sea) were therefore identified with certainty when several buoys were deployed in a row by a vessel. Grouped deployments were identified using time difference (2h) and distance (30km, considering maximum cruising speed of 15 knots) between different deployments of the same vessel.

## S1.4 Sensitivity analyses in matching between echosounder buoy trajectories and fishery data

In order to access the sets and related catch made on the buoy attached to dFADs considered in this study, a match between the trajectory and the date and position of associated sets from logsheet and observer data and dFAD visits from observer data was performed using data from all vessels in the

WCPO. Given that buoys rarely transmitted a position every hour, and only one position per day for some buoys, a match was made on the same calendar day (UTC), with a distance varying between 0 and 10 km. Sensitivity analyses were then performed using observer data and recorded information about buoy owner (i.e., the vessel that initially deployed the dFAD and is paying the buoy satellite fees) to make sure the buoy considered in the acoustic data was the same as the one attached to the dFAD set on. Hence, distances and time difference between set information (observer) and acoustic transmission were compared for buoys: i) owned by a vessel of the same company as the vessel setting on the dFAD; ii) owned by the vessel actually performing the set; or iii) with the same buoy ID as the one recorded by the observer. The latter corresponded to the most confident matching but resulted in a limited number as the buoy ID number remains rarely recorded by observers.

Most matching of buoys from the same company as the vessel making the set or of the same vessel were at 1–5km and 1–8 hours difference, and matching on the exact same buoy at 1–2 km and 1–3 hours difference (Table S2). Maximum distance of 5km and time difference of 8 hours was therefore considered for all matches between acoustic trajectories and logsheet or observer data (Table S3). This might underestimate all possible matches, but will allow selection of a subset of set and dFAD-related activities from observer data for which we are almost certain that the buoy from the acoustic dataset corresponds to the one from the dFAD set upon.

**Table S2.** Investigation of best distance and time difference between buoy trajectory and fishing set to correctly identify the same dFAD on both the acoustic and observer datasets.

		All	Company		Vessel		dFAD ID	
			Same	Different	Same	Different	Same	Different
<b>Distance (km)</b>	Quantile 75%	7.06	<b>5.49</b>	8.10	<b>5.32</b>	7.50	1.93	7.90
	Quantile 90%	8.88	8.10	9.27	8.05	9.08	<b>5.18</b>	9.22
	Quantile 95%	9.46	9.14	9.64	9.09	9.54	7.29	9.58
<b>Time difference (h)</b>	Quantile 75%	10.60	<b>7.84</b>	12.18	<b>7.65</b>	11.35	2.47	9.87
	Quantile 90%	16.34	16.08	16.59	15.17	16.69	<b>7.46</b>	16.21
	Quantile 95%	18.94	19.09	18.74	17.29	19.25	18.27	19.31

A total of 4,342 sets from the logsheet dataset and 2,769 sets from the observer dataset were matched with a buoy in the acoustic dataset (Table S3). If multiple buoys were matched to the same set, the closest one or the one with the same buoy ID (observer data only) were kept. If we also consider matching at <5km and +/- 8h, a total of 2,224 sets from the logsheet dataset and 1,339 sets from the observer dataset were selected.

Other dFAD-related activities (deployments, visits, recoveries) from the observer data were also matched with acoustic trajectories (Table S4). In order to refine again the selection of sets and other activities matched with a buoy trajectory, we only selected those data where the observer's recorded buoy ID and that ID for the trajectory matched. All the other activities (e.g., fishing sets) recorded at the same time were considered to be made on the same dFAD. This resulted in 661 sets in the logsheet data, 657 sets in observer data, 1,198 visits, 211 dFAD deployments, 194 dFAD retrievals, 462 buoy deployments and 458 buoy retrievals considered for further analyses (Tables S3 and S4).

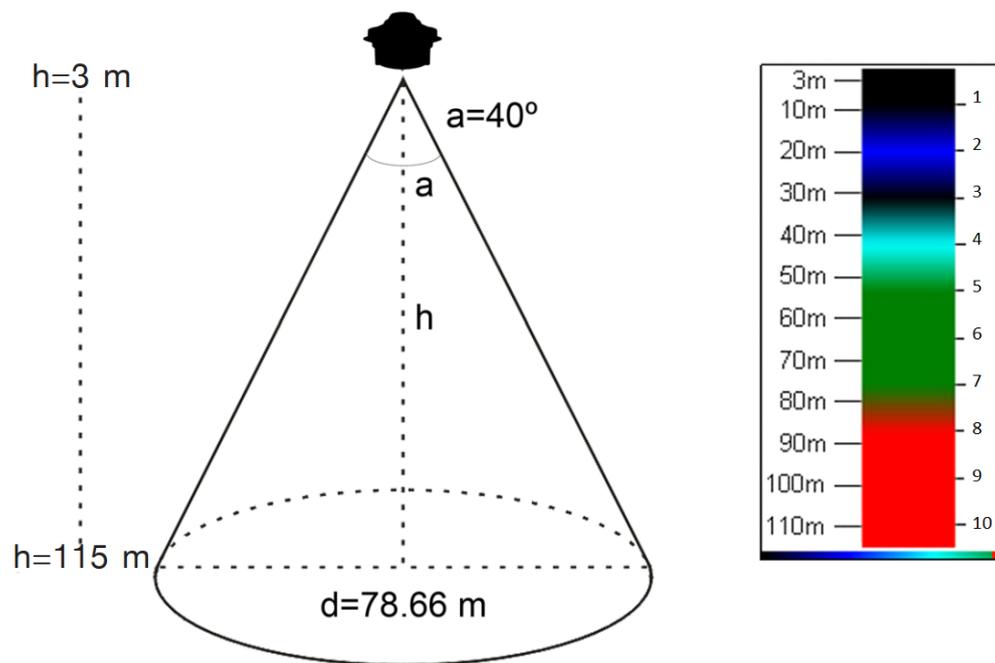
**Table S3.** Matches between buoys in the acoustic dataset and sets in logsheets, or set and other dFAD related activities in the observer data from all vessels in the WCPO.

	Logsheets				Observer			
	All	%	<5 km +/- 8h	%	All	%	<5 km +/- 8h	%
Total number of sets (all WCPO vessels)			42,529				31,669	
Number of matching	4,342		2,378		2,769		1,403	
Unique set matched	3,935	7.3	2,224	4.2	2,531	6.4	1,339	3.4
Unique buoy trajectory with matching	2,344	34.5	1,596	23.5	1,674	24.6	1,008	14.8
Set with only 1 buoy matched	3,617	83.3	2,106	88.6	2,349	84.8	1,290	91.9
Set with >1 buoy matched	318		118	5.0	181	6.5	48	3.4
Average number of sets per buoy	2.8		2.5		2.5		1.8	
Number of buoys with grouped deployments	1006		489		604		271	
<b>Selected matching</b>			661				657	

**Table S4.** Other matchings made between trajectories and the observer database, at less than 5km and 8h difference, from all vessels in the WCPO.

	Visits	dFAD deployment	dFAD retrieval	Buoy deployment	Buoy retrieval
Matching events	2,952	554	344	1,040	1,290
Unique buoy	1,586	458	299	976	755
<b>Selected matching events</b>	1,198	211	194	462	458

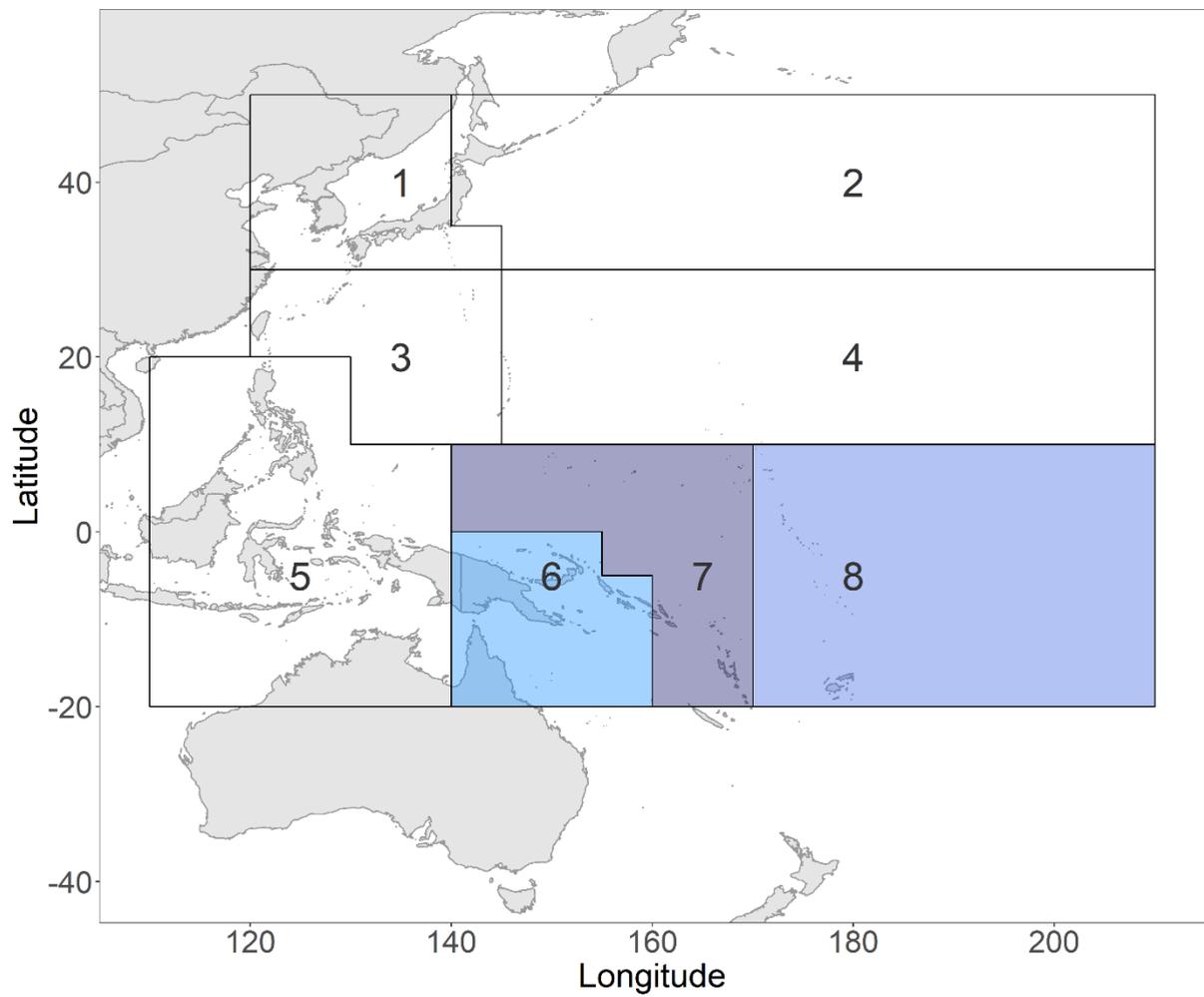
## Appendix 2. Biomass by depth layer



**Figure S6.** Depth layers and characteristics of the echosounder cone for Satlink echosounder buoys. Beam width (or angle) (a), depth range (h), and diameter (d) at 115 m. Derived from Lopez et al. (2016)

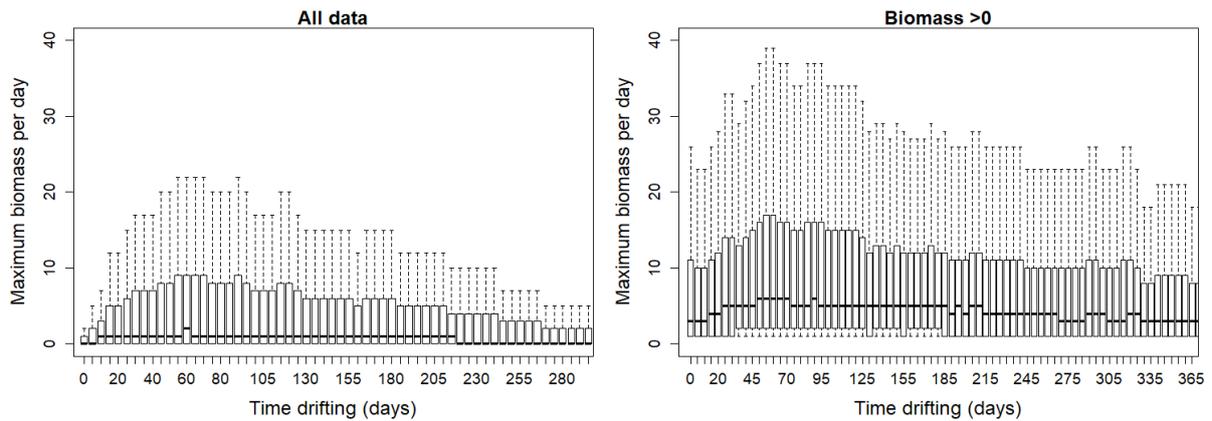
## Appendix 3. Additional figures

### S3.1 Skipjack tuna stock assessment regions



**Figure S7.** Spatial structure from the most recent skipjack tuna stock assessment (Vincent et al., 2019). Regions primarily associated with purse seine fishing activity (i.e., Regions 6–8) are colored.

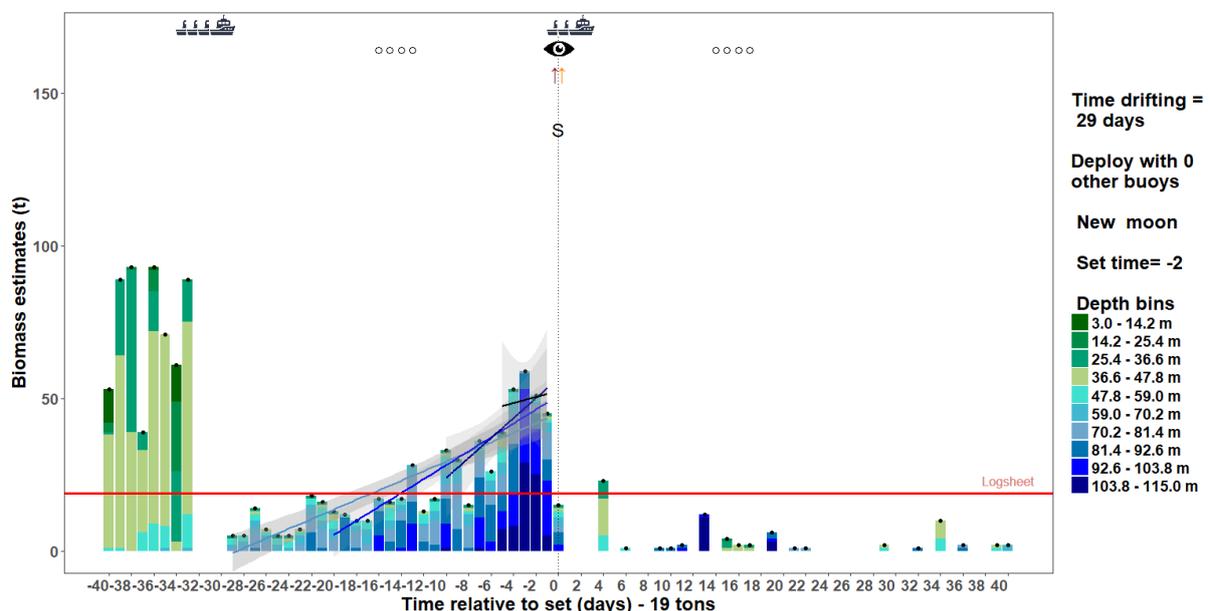
### S3.2 Colonization of biomass after dFAD deployment



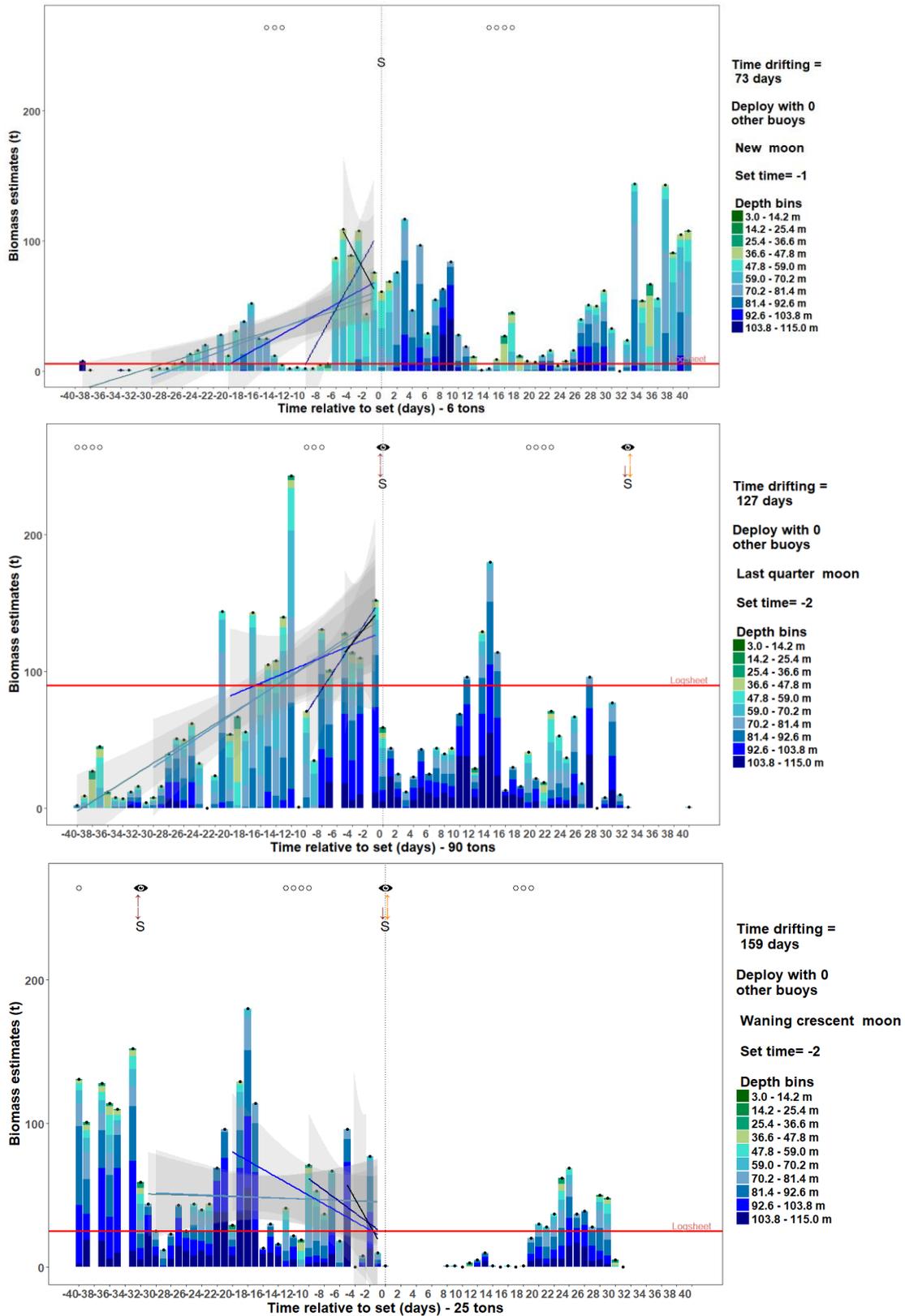
**Figure S8.** Evolution of the maximum biomass per day (t) estimated by the echosounder buoys (transmission within +/- 5h of sunrise) depending on time drifting, for all dFADs with no set matched.

### S3.3 DFAD life history patterns and dFAD accumulation processes before a set

Patterns of biomass estimates before and after a fishing set were investigated for some example dFADs to better understand the variability between catch and biomass (Figures S11 and S12). Figures were generated to follow the ‘life-history’ of buoys, with the different fishing sets, as well as any other activities recorded by observers, and periods when that same buoy was on-board a vessel.

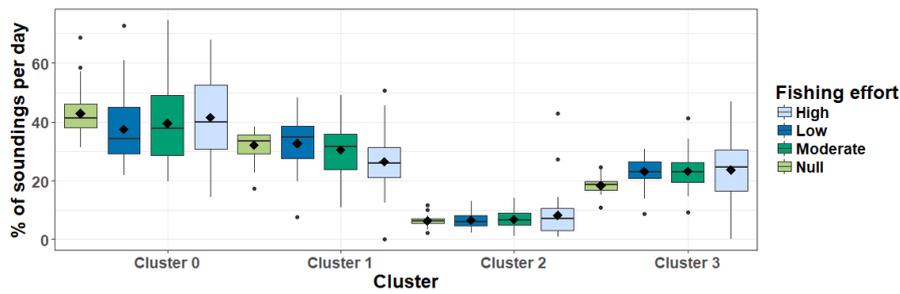


**Figure S9.** Example of biomass accumulation (t), from a Satlink buoy, before and after a fishing set. Color bars indicate the maximum biomass per day (40 days prior to and after a fishing set) and per depth bin, linear regression for 30-day, 20-day, 10-day and 5-day periods are shown as blue lines. On the right panel, time drifting, grouped deployments, moon phase (also indicated with circles on the graph for the full moon), time of the beginning of the set relative to sunrise and depth layers. Red horizontal line = total recorded catch of the set from logsheet data; S = day of a fishing set; top brown arrow = recovery of a dFAD; top orange arrow = recovery of a buoy, vessel = buoy on-board; eye = visit of the dFAD.

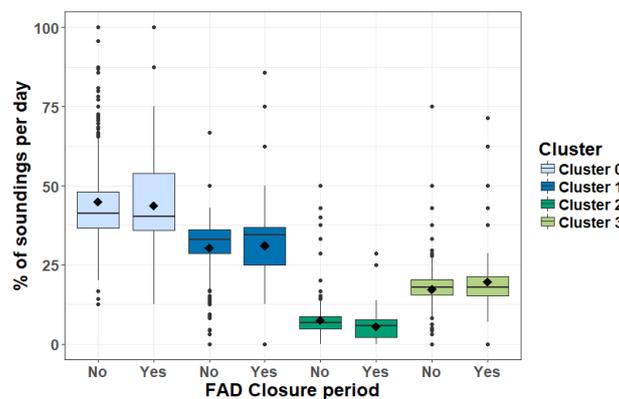


**Figure S10.** Examples of biomass accumulation (t), from a Satlink buoy, before and after a fishing set. Color bars indicate the maximum biomass per day (40 days prior to and after a fishing set) and per depth bin, linear regression for 30-day, 20-day, 10-day and 5-day periods are shown as blue lines. On the right panel, time drifting, grouped deployments, moon phase (also indicated with circles on the graph for the full moon), time of the beginning of the set relative to sunrise and depth layers. Red horizontal line = total recorded catch of the set from logsheet data; S = day of a fishing set; top/bottom brown arrows = recovery/deployment of a dFAD; top/bottom orange arrow = recovery/deployment of a buoy, vessel = buoy on-board; eye = visit of the dFAD.

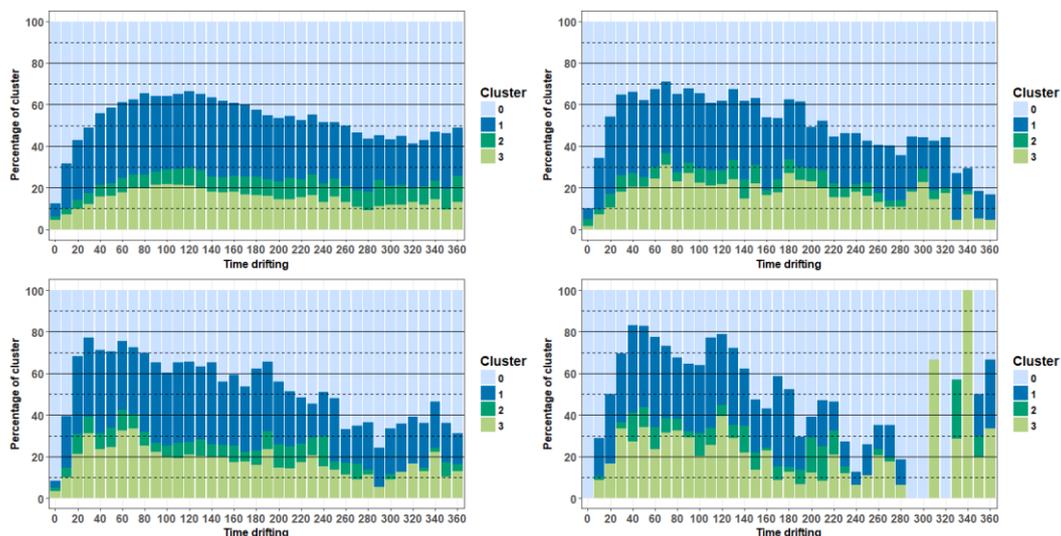
### S3.4 Clustering analyses of echosounder transmissions



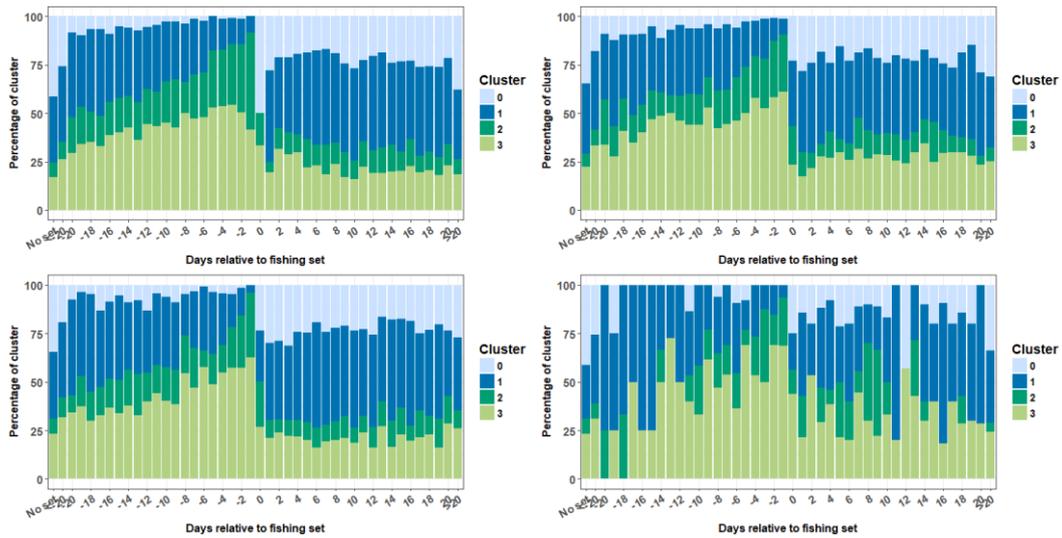
**Figure S11.** Percentage of echosounder transmissions (around +/- 5 hours of sunrise) per day in areas of varying fishing effort (Null = 0 fishing sets per 1° cell and month, Low = 1 or 2 fishing sets per 1° cell and month; Moderate = 3–15 fishing sets per 1° cell and month; High = >15 fishing sets per 1° cell and month).



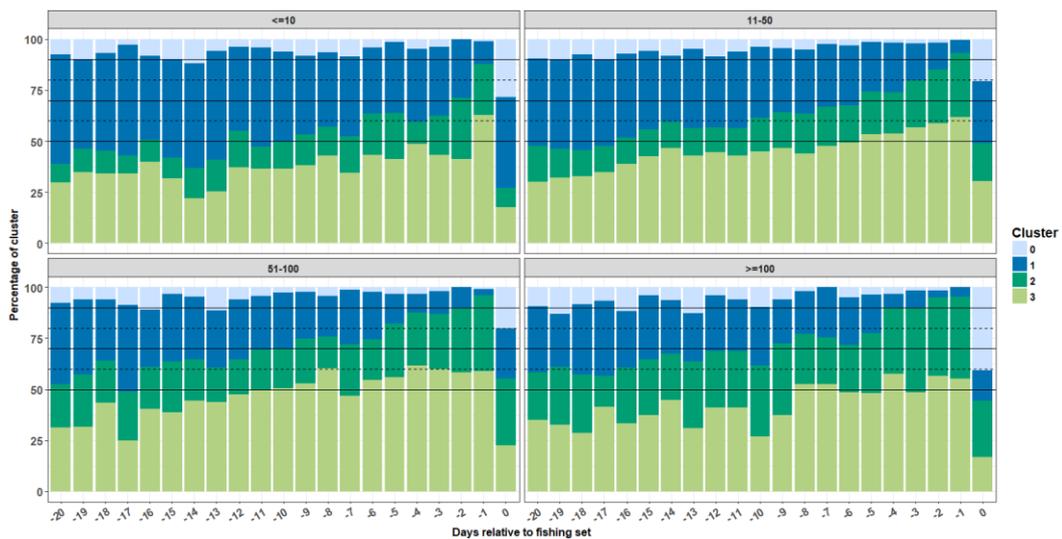
**Figure S12.** Percentage of echosounder transmissions (around +/- 5 hours of sunrise) per day during or outside the dFAD closure period.



**Figure S13.** Percentage of echosounder transmissions (around +/- 5 hours of sunrise, and within 10°N and 10°S) per clusters depending on time drifting (only dFAD deployed in grouped deployments were considered) and fishing effort (Null = 0 fishing sets per 1° cell and month (top left), Low = 1 or 2 fishing sets per 1° cell and month (top right); Moderate = 3–15 fishing sets per 1° cell and month (bottom left); High = >15 fishing sets per 1° cell and month (bottom right)).

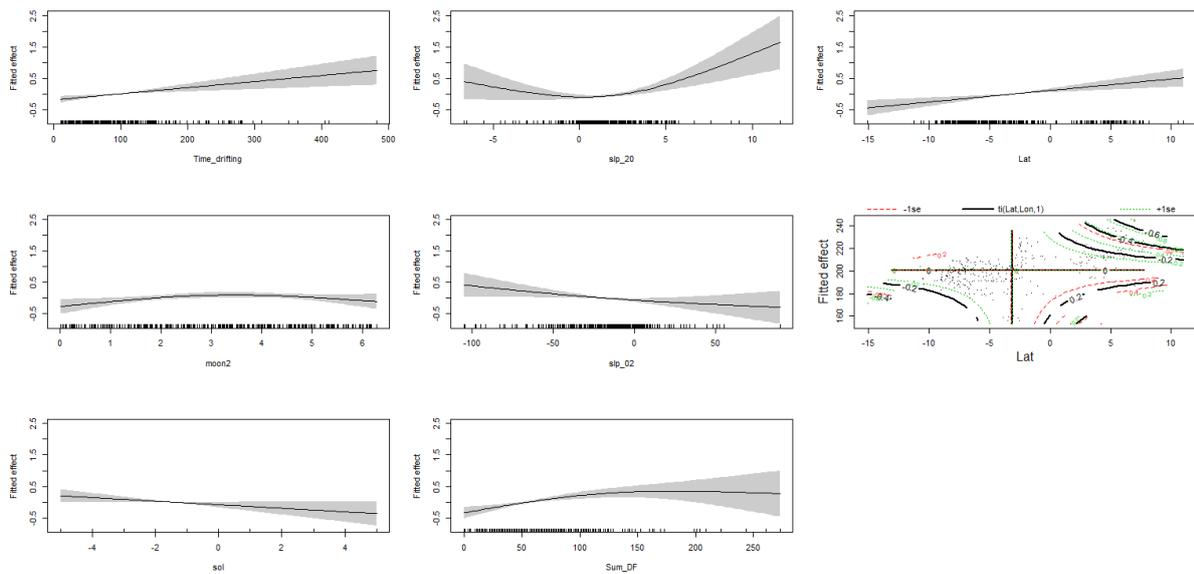


**Figure S14.** Percentage of echosounder transmissions (around +/- 5 hours of sunrise, and within 10°N and 10°S) per clusters on days relative to a fishing set (only sets with >5t of tuna catch were considered) and fishing effort (Null = 0 fishing sets per 1° cell and month (top left), Low = 1 or 2 fishing sets per 1° cell and month (top right); Moderate = 3–15 fishing sets per 1° cell and month (bottom left); High = >15 fishing sets per 1° cell and month (bottom right)).

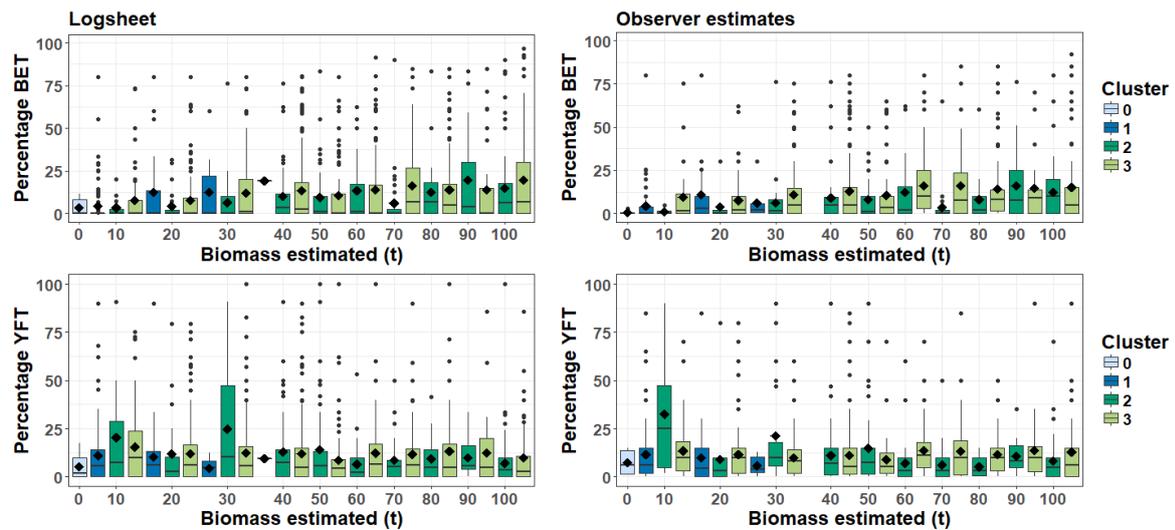


**Figure S15.** Percentage of echosounder transmissions (around +/- 5 hours of sunrise, and within 10°N and 10°S) per clusters on days relative to a fishing set (only sets with >5t of tuna catch were considered) and level of tuna catch achieved (<=10t (top left); 11-50 t (top right); 51-100 t (bottom left); >=100t (bottom right)).

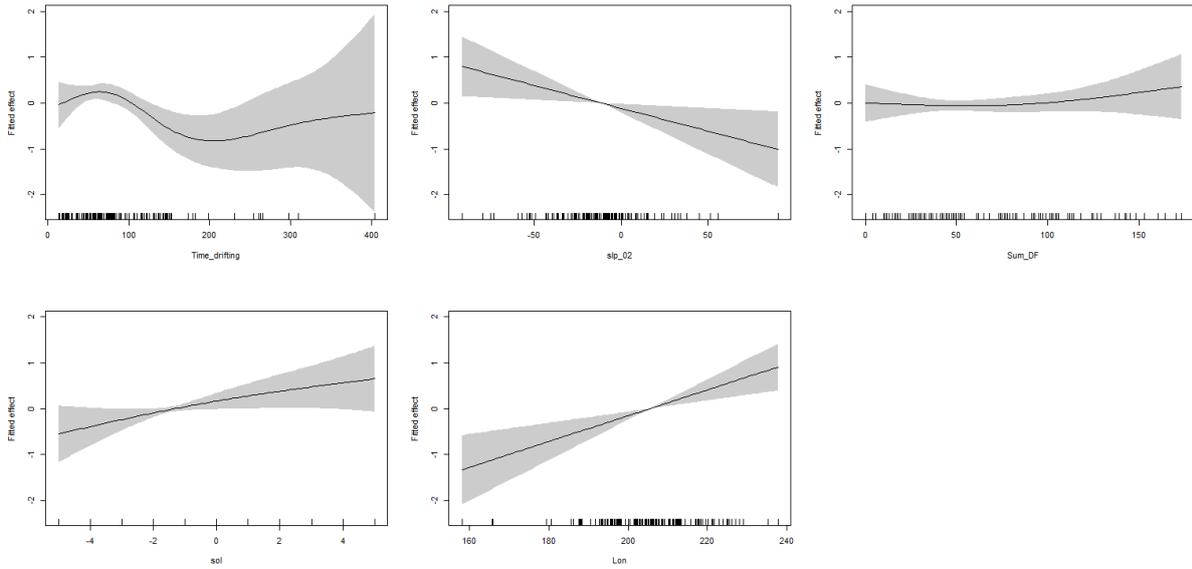
### S3.3 Models of tuna catch and bigeye and yellowfin proportions



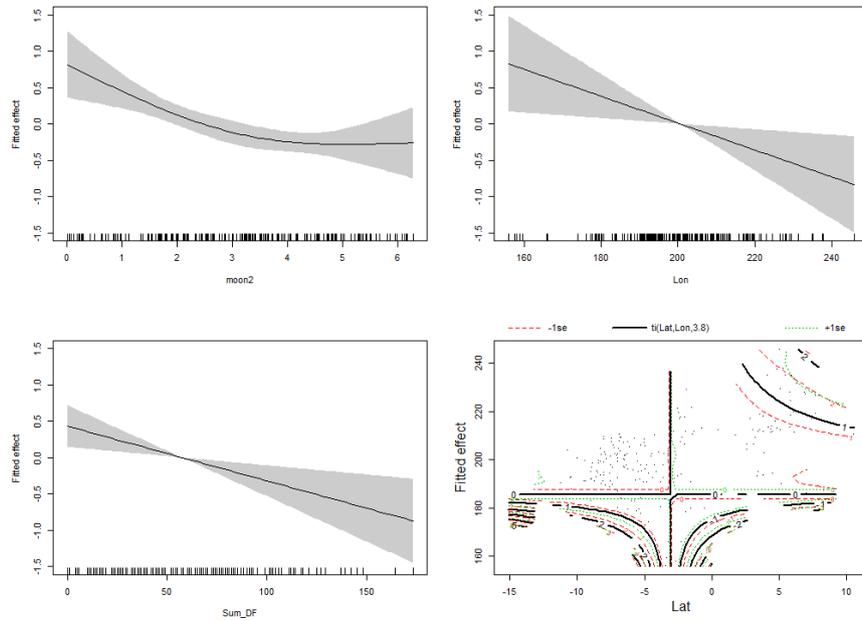
**Figure S16.** Smoothed fits of covariates modelling the total tuna catch in associated fishing sets using lognormal GAM.



**Figure S17.** Percentage of bigeye (top) and yellowfin (bottom) tunas, depending on the level of biomass estimated over a 5-days period before a fishing set from logsheet (left) and observer data (right).



**Figure S18.** Smoothed fits of covariates modelling the bigeye tuna proportion in associated fishing sets using lognormal GAM.



**Figure S19.** Smoothed fits of covariates modelling the yellowfin tuna proportion in associated fishing sets using lognormal GAM.

## Appendix 4. Spatiotemporal modeling approach

To develop the combined data model, we used a Poisson-link delta model (Thorson, 2018) and a model structure similar to that developed by Grüss and Thorson (2019). The echosounder data provided presence/absence data  $P$  which took on the value of 1 when a dFAD was classified as having skipjack present and 0 when no skipjack were detected (or predicted), and were modeled assuming a Bernoulli distribution

$$P \sim \text{Bernoulli}(p(i))$$

where the probability of encounter ( $p(i)$ ) follows a Poisson distribution with intensity equal to the local numbers density  $n(s_i, t_i)$ .

$$p(i) = 1 - \exp(-n(s_i, t_i))$$

We can think of this in the sense that the true variable of interest, for which we are using encounter probability to represent, is a Poisson process (i.e., number of individual fish). By using the estimated numbers density as the Poisson intensity to estimate the probability of encounter, we ensure that as local density increases, the expected probability of encounter also increases. The CPUE sampling data  $C$  represents skipjack catch (in tonnes) per set  $c(s, t)$  at each location  $s$  and time step  $t$  (year-quarter), and could take on any positive value or zero if no skipjack were encountered. Using a Poisson-link delta model, biomass density  $d(s_i, t_i)$  was predicted as the product of the encounter probability  $p(i)$  and the positive catch rates  $r(i)$ .

$$r(i) = \frac{n(s_i, t_i)}{p(i)} \cdot w_i \quad (1)$$

Assuming a gamma error distribution for the positive catch rate component (Thorson et al., 2021), the delta model for the CPUE data is defined as:

$$\Pr(c(i) = C) = \begin{cases} 1 - p(i) & \text{if } C = 0 \\ p(i) \cdot g(C|r(i), \sigma_{obs}^2) & \text{if } C > 0, \end{cases}$$

where  $g(C|r(i); \sigma_{obs}^2)$  is the gamma error distribution for unexplained variation in the positive catch rates and  $\sigma_{obs}^2$  is residual variability. As Grüss and Thorson (2019) demonstrate, this combined modeling approach is such that the likelihood components for each of the data types have shared parameters ( $n(s_i, t_i)$ ), making the spatiotemporal modeling process more straightforward. In their example, they also modeled count data, but here, we have a simpler application that combined biomass and presence/absence data. Briefly, the linear models used to estimate encounter probability  $p$  and magnitude of positive catch rates  $r$  were constructed as

$$\begin{aligned} \log(n(s_i, t_i)) &= \beta_n(t_i) + \omega_n(s_i) + \varepsilon_n(s_i, t_i) + \sum_{m=1}^{n_m} \gamma_m G(i, m) \\ \log(w_i(s_i, t_i)) &= \beta_w(t_i) + \omega_w(s_i) + \varepsilon_w(s_i, t_i) \end{aligned}$$

where  $\log(n)$  is defined as local numbers density and is used to estimate the encounter probability, and the biomass per individual is modeled as  $\log(w)$ . The  $\beta$  terms are the intercepts (time indices), the  $\omega$  parameters represent the spatial variation and the  $\varepsilon$  terms represent the spatiotemporal variation. Both  $\omega$  and  $\varepsilon$  are modeled as Gaussian Markov random fields. Lastly,  $\sum_{m=1}^{n_m} \gamma_m G(i, m)$  is the effect of sampling program  $m$  on the expected number of individuals sampled, where  $G(i, m) = 1$  for sampling program  $m$  that collected the samples and 0 otherwise. This reformulation of the traditional delta-model is possible because of the relationship defined in Equation 1, where numbers density  $n(s_i,$

$t_j$ ) is converted to encounter probability  $p(i)$  and biomass per group  $w(s_i, t_j)$  is converted to positive catch rates  $r(i)$  (see Thorson, (2018) for additional details). The spatiotemporal model was fit using the VAST R package (Thorson et al., 2015) with 50 spatial knots to create a uniform spatial surface for the estimation. To compare the results from the combined data model, we also fit a model to the CPUE data alone.