Aquat. Living Resour. 2022, 35, 13

https://doi.org/10.1051/alr/2022013

Aquatic Living Resources
Available online at:
www.alr-journal.org

Research Article

Open Access

A standard processing framework for the location data of satellite-linked buoys on drifting fish aggregating devices

Yannick Baidai[1,2,*], Jon Uranga[3], Maitane Grande[3], Hilario Murua[4], Josu Santiago[3], Iñaki Quincoces[3], Guillermo Boyra[3], Blanca Orue[5], Laurent Floch[1] and Manuela Capello[1]

[1] MARBEC, Univ Montpellier, CNRS, Ifremer, IRD, Sète, France
[2] Centre de Recherches Océanologiques (CRO), 29, rue des Pêcheurs BPV. 18, Abidjan Côte d'Ivoire, Republic of Côte d'Ivoire
[3] AZTI-Tecnalia, Herrera kaia portualdea z/g, 20110 Pasaia (Gipuzkoa), Spain
[4] International Seafood Sustainability Foundation (ISSF), Washington, DC, USA
[5] Collecte Localisation Satellites (CLS), Parc Technologique du Canal, 11 Rue Hermès, 31520 Ramonville-Saint-Agne, France

Received 9 January 2022 / Accepted 9 August 2022

Handling Editor: Verena Trenkel

Abstract – Satellite-linked buoys used by tropical tuna purse-seine vessels on drifting fish aggregating devices (DFADs) provide a continuous stream of information on both the ocean characteristics and the presence and size of fish aggregations associated with DFADs, enabling the study of pelagic communities. This unprecedented amount of data is characterized by ocean-scale coverage with high spatial and temporal resolutions, but also by different data formats and specifications depending on buoy model and brand, as well as on the type of data exchange agreements into play. Their use for scientific and management purposes is therefore critically dependent on the abilities of algorithms to process heterogeneous data formats and resolutions. This paper proposes a unified set of algorithms for processing the buoys location data used by the two major purse seine fleets operating in the Atlantic and Indian oceans. Three main issues that need to be addressed prior to the exploitation of the data are identified (structural errors, data records on land and on-board vessels) and five specific filtering criteria are proposed to improve the data cleaning process and, hence, quality. Different filtering procedures are also compared, and their advantages and limitations are discussed.

Keywords: Instrumented DFADs / satellite-linked buoys / purse seiners / tropical tunas / data processing

1 Introduction

Defined as man-made floating objects specifically designed to attract tunas and improve catches, Drifting Fish Aggregating Devices (DFADs), are a major fishing tool used in tropical tuna purse seine fisheries (Fonteneau et al., 2013). It has been estimated that approximately 65% of the global tropical tuna purse seine landings stem from catches made using DFADs (Scott and Lopez, 2014). The DFAD-based fishery relies on a behavioral trait exhibited by several pelagic marine species, including tropical tunas, which leads them to gather in mass around objects floating at sea. Since their introduction in the tropical tuna purse seine fisheries, DFAD technology has rapidly evolved from simple floating objects often equipped with radar reflectors to help fishermen locate them, to more complex raft designs equipped with electronic tracking devices, ranging from radio transmitters to GPS beacons (Dagorn et al., 2013; Lopez et al., 2014). Currently, all deployed DFADs are equipped with satellite-linked buoys that incorporate echosounder devices to estimate the biomass underneath DFADs (Moreno et al., 2019). These buoys remotely provide near real-time information on their position, drift and the presence and size of the fish aggregation associated to the DFADs. This telemetered information has resulted in a significant increase in the fishing efficiency of purse seine vessels (Fonteneau et al.,1999; 2013; Lopez et al., 2014; Wain et al., 2020). The major changes in fishing strategies resulting from the use of DFADs have also introduced substantial uncertainties in the catch per unit effort traditionally used to assess tuna populations from commercial purse seine data. This stems especially from the non-random nature of DFAD based fishery, which adds considerable complexity to the estimation of the purse-seiner fishing effort (Fonteneau et al., 1999; 2013; Torres-Irineo et al., 2014). Moreover, the role of DFADs in improving purse seine fishing efficiency and their intensive use have raised several questions related to their impacts on tuna stocks and their ecology, as

*Corresponding author: yannick.baidai@ird.fr

well as on marine ecosystems (Dagorn et al., 2013). As a result, a major current concern for tuna Regional Management Organisations (tRFMOs) surrounds the need for complementary data on DFADs. In particular, growing concerns about the impacts of DFAD use have led to a number of specific plans for their management by tRFMOs, incorporating, *inter alia*, the strengthening of reporting requirements on DFAD activities and densities (e.g. IOTC: Res. 19/08; ICCAT: Rec 19-02; IATTC: C-19-01; WCPFC: CMM 2018-01). The vast amount of position and biomass data collected by the instrumented buoys used to monitor DFADs constitutes a unique and extremely important asset for scientists. Because of their large number, wide spatial distribution and constant maintenance by fishermen, satellite-linked echosounder buoys allow effortless and cost-effective collection of various types of data likely to provide valuable insights into ocean dynamics (Imzilen et al., 2019), distribution and behaviour of fishes (Lopez et al., 2017; Orue et al., 2019a; Baidai et al., 2020a), or to be used to derive novel abundance indices for tuna populations (Santiago et al., 2016, 2020), as well as new methods to reduce bycatch in the tropical tuna purse seine fisheries (Mannocci et al., 2021). The instrumented DFADs represent an unprecedented observatory of marine pelagic communities (Brehmer et al., 2018; Moreno et al., 2016).

Currently three major manufacturers dominate the DFAD buoy industry[1] in Atlantic and Indian Oceans, and each offer models that differ in terms of their hardware and software (Moreno et al., 2019). The output datasets can also vary greatly in both their nature and format depending on manufacturer and model. Furthermore, because the provision of DFAD buoy data to scientists is still conducted at regional or national level, the characteristics of the dataset can also depend on the details of the specific data exchange agreement between industry and the respective national authority. Although some processing protocols have already been proposed (Maufroy et al., 2015; Orue et al., 2019a), they have primarily been applied to single buoy types and uniform datasets. The application of these methods can quickly become limited when faced with datasets that mix several fishing fleets, brands and models of buoys. Just like the intensive work undertaken a few years ago on VMS data filtering procedures, (e.g. Gerritsen and Lordan 2011; Hintzen et al., 2012; Lambert et al., 2012; Lee et al., 2010), the design of a standardized framework to capture the heterogeneity that typifies DFAD-related data for their integration into research and management processes is now emerging as a key priority. Buoy location data, in particular, are critical to achieve some of the major objectives of DFAD management; including estimating and monitoring the actual number of DFADs at sea (Chassot et al., 2019; Escalle et al., 2021; Gershman et al., 2015), improving the conservation measures regarding DFAD limitations and their enforcement by fishermen (Goñi et al., 2017; Lennert-Cody et al., 2018), or defining strategies to mitigate their stranding in sensitive areas (Curnick et al., 2020; Davies et al., 2017; Escalle et al., 2021, 2019; Imzilen et al., 2022, 2021). Hence, a processing framework to transform this vast amount of industrial data into harmonized information, notably through standard procedures, independent of the characteristics of the databases, appears to be of primary importance.

In this study a set of processing algorithms were proposed and applied to the raw data provided by the buoys from the two major tropical tuna purse seine fleets (French and Spanish) in the Atlantic and Indian oceans. The outcomes of the filtering algorithms were compared between the various buoy models and heterogeneous buoys location data formats that make up the two databases.

2 Material and methods

2.1 Buoy data

Buoy data has been collected in the Atlantic and Indian Oceans, under specific data-exchange agreements signed between different research organizations (i.e. AZTI and IRD) and EU tuna purse seine associations (i.e. ORTHONGEL[2] for the French purse-seine fleet, Echebastar and Atunsa companies in ANABAC[3] and OPAGAC[4] for Spanish fleets), under the framework of the EU project RECOLAPE[5].

A dataset was created for each fleet and ocean. They consisted of information collected by a sample of 1000 buoys during a random month of the year 2016. Since the objective of this work was not to compare the characteristics of the databases between fleets, buoy manufacturers or research organizations, but rather to define a common processing protocol, the datasets examined for each fleet (resulting from specific data exchange-agreements) were referred to as D1 and D2. Details on the key features and differences between the two datasets are provided in the following section. Following the same principle, the different buoy models included in this study were also anonymized. The D1 datasets consisted of 62,902 and 61,194 rows for the Atlantic and Indian oceans respectively, whereas the D2 datasets were composed of 25,304 rows for the Atlantic Ocean, and 22,461 rows for the Indian Ocean.

2.2 Datasets characteristics

Table 1 provides the list of the different buoy brands, as well as a description of the raw data contained in each dataset. Common information includes the buoy identification code, hour, date, position (latitude and longitude) and buoy speed. The other data types are dependent upon the buoy brand and data-exchange agreements. In each of the two oceans, large differences between the composition of buoy models available in the different fleet datasets (D1 and D2) are observed (Fig. 1). The D1 datasets is largely dominated by a single model of buoy, especially in the Indian Ocean, but between two and four models were present, depending on the Ocean. The D2 datasets is more heterogeneous, with roughly twice as many buoy

[1] Marine Instruments (www.marineinstruments.es), Satlink (www.satlink.es) and Zunibal (www.zunibal.com).

[2] Organisation française des producteurs de thons congelés et surgelés.

[3] Asociación Nacional de Armadores de Buques Atuneros Congeladores.

[4] Organización de Productores Asociados de Grandes Atuneros Congeladores.

[5] MARE/2016/22 "Strengthening regional cooperation in the area of fisheries data collection".

Table 1. Description of the raw position and acoustic data received from the various buoy manufacturers and under the different data-exchange agreements (D1 and D2).

		Marine Instruments		Satlink	Zunibal
		D1	D2	D2	D2
Buoy Operation Data	Buoy identification code	☒	☒	☒	☒
	Owner vessel(s)	☒	☒	☒	☐
	Buoy activation date	☒	☐	☐	☐
	Buoy deactivation date	☒	☐	☐	☐
	Flash (notification on activation of the buoy flash)	☒	☐	☐	☐
	Buoy operating mode)	☒	☒	☐	☐
	Buoy battery level	☒	☒	☐	☐
	in/out water sensor	☐	☐	☐	☒
Location data and other data	Timestamp of GPS position data	☒	☒	☒	☒
	GPS position data	☒	☒	☒	☒
	Buoy speed	☒	☒	☒	☒
	Buoy drift angle	☒	☐	☐	☐
	Sea water temperature	☒	☒	☐	☐
Acoustic sampling data	Timestamp of acoustic data collection	☒	☒	☒	☐
	Sampling frequencies	☒	☒	☐	☐
	Echosounder detection range	☒	☒	☐	☐
	Gain of acoustic samplings	☒	☒	☐	☐
	Resolution (number of bits used in each layer)	☒	☒	☐	☐
	Number of depth layers	☒	☒	☐	☐
	Total biomass index (estimated tonnage from the echosounder)	☒	☐	☒	☒
	Maximum biomass estimated at any layer	☐	☐	☒	☐
Acoustic data format		Integers (from 0-7 or 0-15) representing the intensity of the acoustic signal detected		Biomass (tons) estimated per layer of 11.2 m over 115 m depth (10 layers), based on a buoy-integrated algorithm	Total biomass (tons)
Buoys models in the dataset		M3+, M4I, M3I, MSI		DL+, DSL+, ISD+, ISL+	T7+

models, and two to three dominant models in each ocean. The temporal resolution of the buoy location data is also characterized by significant variability depending on the datasets. The D2 datasets is limited to a single location per buoy per day, regardless of the ocean or buoy model. This is contrasted by a higher resolution in the data of the D1 datasets, with between 3 and 9 location data per day for both oceans, depending on the buoy model (Fig. 2).

2.3 Data processing protocol

The data processing protocol comprises the definition of five specific filtering criteria (defined here from F1 to F5), structured into three main processing stages (Fig. 3).

2.3.1 Stage 1: Structural errors filtering

Structural errors, such as duplicate or irrelevant rows in the dataset, resulting from failures during data collection or transfer are defined at this stage. Three types of structural errors, principally related to failures during satellite communication, were flagged in the databases.

(i) Filter F1: Duplicate rows

Duplicate data refer to rows with identical buoy codes, timestamps and positions. Generally, all other information in the duplicate rows remain strictly identical, however in rare cases, missing values may occur for some lines. Duplicates were identified based on their identical buoy codes, timestamps and locations. In the cases of missing data, the row kept, and considered as the original, was the one with the most complete information.

(ii) Filter F2: Ubiquitous rows

Ubiquitous rows consisted of cases where two rows have identical buoy codes and timestamps, but different locations (Supplementary Fig. 1). Rows with these characteristics were identified and the distance between locations was calculated. When the two positions were separated by less than 1 km, a randomly selected row was retained while the other was considered "ubiquitous". Otherwise, the two rows were assigned as "ubiquitous".

(iii) Filter F3: Isolated positions

Positions separated from their nearest neighbors on a buoy track, by more than 48 h (considering the general time resolution in the data collection) or having an inconsistent speed (considering a threshold of 35 knots, a value far above speeds of both tuna purse seine vessels and ocean currents), were referred to as isolated positions (Supplementary Fig. 2). By addressing the buoy track (a collection of positions belonging to a single GPS buoy), this filtering step allowed the identification of a distinct series of consecutive positions (segments) separated by potential GPS failures, buoy relocations or buoy deactivation/reactivation events on a given buoy track.

2.3.2 Stage 2: Filtering of land positions (Filter F4)

Buoys located on land (due to beaching events or active buoys brought back to port) were detected using shoreline data from the GSHHG database (Global Self-consistent, Hierarchical, High-resolution Geography; Wessel and Smith, 1996). The influence of different shoreline resolutions on the filtering procedure was assessed through the comparisons of results from low and high-resolution shorelines (see details in Wessel and Smith, 1996) buffered with 0.05° (Supplementary Fig. 3).

2.3.3 Stage 3: Filtering of "on-board"/"at sea" buoy positions (Filter F5)

Echosounder buoys can be activated and transmit on-board vessels prior to their deployment. Similarly, buoys retrieved from the sea may continue to collect data on-board vessels for variable durations. In order to discriminate "on-board" from "at sea" positions, two different approaches were compared. The first approach was based on a rule-based algorithm using the buoy's speed as the main classification variable. This was referred to as the "kinetic classification algorithm". The second applied a random forest model (Breiman, 2001) trained using a learning dataset containing information from a single buoy type (Model 9, Tab. 1). The two algorithms classified the data into three classes: "on-board" (for buoys emitting while on-board a vessel), "at sea" (for buoys deployed in the water), and "undetermined" (a subset of positions that remained unclassified). Finally, comparisons of the classification results of the two algorithms were carried out through the calculation of simple matching coefficient estimated from confusion matrices derived from the outputs of the two approaches (Sokal, 1958). For this purpose, unclassified buoy positions were considered as emitting from water ("at sea").

(i) Kinetic classification algorithm (KiC)

The kinetic classification algorithm uses deterministic rules encoded in the form of if-then-else statements as a representation of knowledge (Baidai et al., 2017; Grande et al., 2020). The different rules were derived from knowledge of surface currents in tropical oceans and the behaviour of tuna purse seine vessels. The algorithm consisted of two iterative classification steps based on three main parameters (see Fig. 4A): (*i*) the *buoy speed* (calculated between consecutive positions), (*ii*) the *buoy speed history* (the maximum value of the speed recorded during a time window of three days before the current position), (*iii*) and the *change in buoy speed* (absolute value of the speed difference between two consecutive points). The three rules governing the first classification step were stated as follows:

- A position with a *buoy speed* higher or equal to 6 knots correspond to an on-board position ("on-board");
- A position with a *buoy speed history* of less than 6 knots corresponds to a buoy emitting from the water ("at sea");
- A position that does not meet either of the two rules has an undetermined status.

The selected cut-off value of 6 knots is largely higher than the theoretical maximum drift speed in the Atlantic and Indian Oceans. Since the average speed of tuna purse vessels is well above this value (from 9 to more than 12 knots), buoys with speeds exceeding this threshold are very likely to be on-board a vessel (Supplementary Fig. 4). The second rule relates to the fact that active purse seine vessels rarely maintain speeds below 6 knots for long durations, therefore buoys that display low speeds for a continuous period can be considered as actually emitting from the water. A number of segments (series of consecutive positions along a track) can be classified from this set of rules. From these

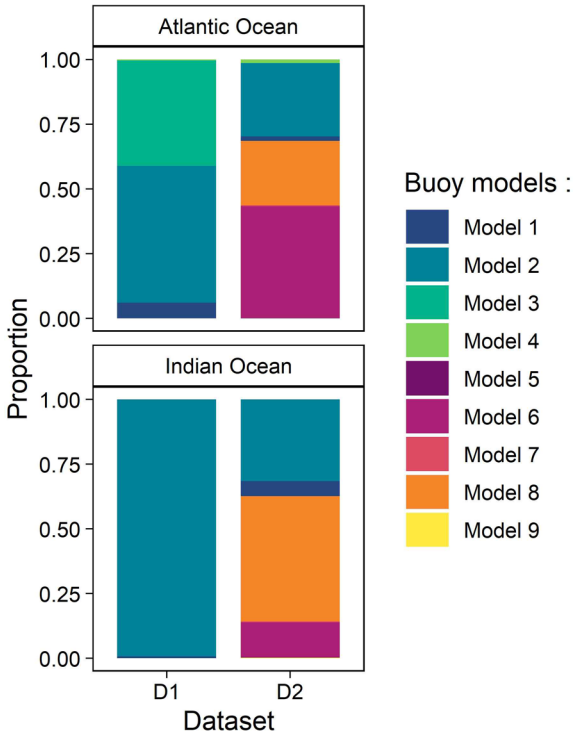

Fig. 1. Proportion of buoy model constituting the D1 and D2 datasets in the Atlantic and Indian Oceans.

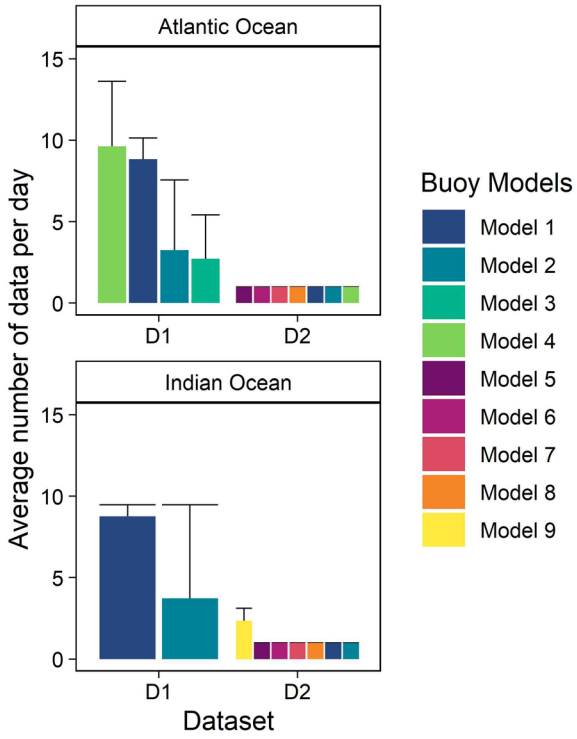

Fig. 2. Average number of location data provided per day by the different buoy models in the Atlantic and Indian Ocean datasets.

segments, (1) "constant sequences" defined as consecutive positions with the same predicted status (i.e. on-board – on-board, or at sea – at sea) and (2) "transition sequences", where the buoy shifts from one status to another, were defined.

The second classification step relies on the comparison of *changes in buoy speed* recorded for positions with undetermined status with those found for "constant" and "transition sequences" (Fig. 4B). The comparison starts from the first undetermined position that immediately follows a classified position (i.e. defined from the first classification step). The *change in buoy speed* between the undetermined and classified positions is estimated and compared to the *changes in buoy speed* found in the constant and transition sequences, respectively, using a Student *t*-test at confidence level of 95%. The status of the undetermined position is then assigned according to the result of the test of comparison. For example, an undetermined position following an "on board" position, and whose *change in buoy speed* is not significantly different from "constant sequences" will be classified as "on board". Conversely, if the comparison is not significantly different from "transition sequences", its status will be classified as "at sea". This classification step is first performed from neighbor to neighbor, moving along the buoy segment. The same procedure is then carried out backwards (from the end of the trajectory to the beginning), considering the remaining unclassified positions (Supplementary Fig. 5).

(ii) Random forest approach (RF)

The RF approach was derived from the procedure developed by Orue et al. (2019a). The learning dataset was built using data from a single buoy model (model 9, Tab. 1), which have a conductivity sensor to detect when the buoy is immersed in seawater. The RF classification model was constructed using (*i*) distance between two consecutive points,

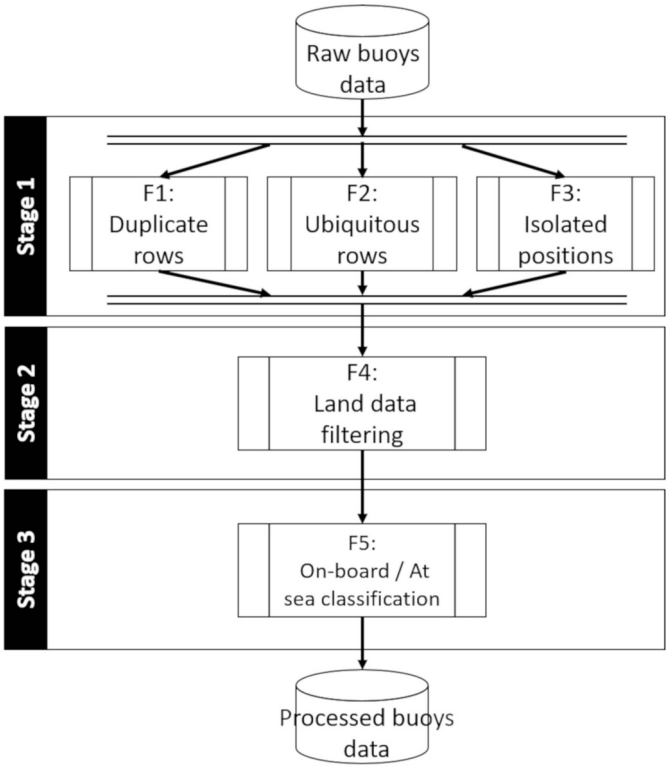

Fig. 3. Flowchart of the standard processing protocol for satellite-linked echosounder buoys used in tropical tuna purse seine fisheries.

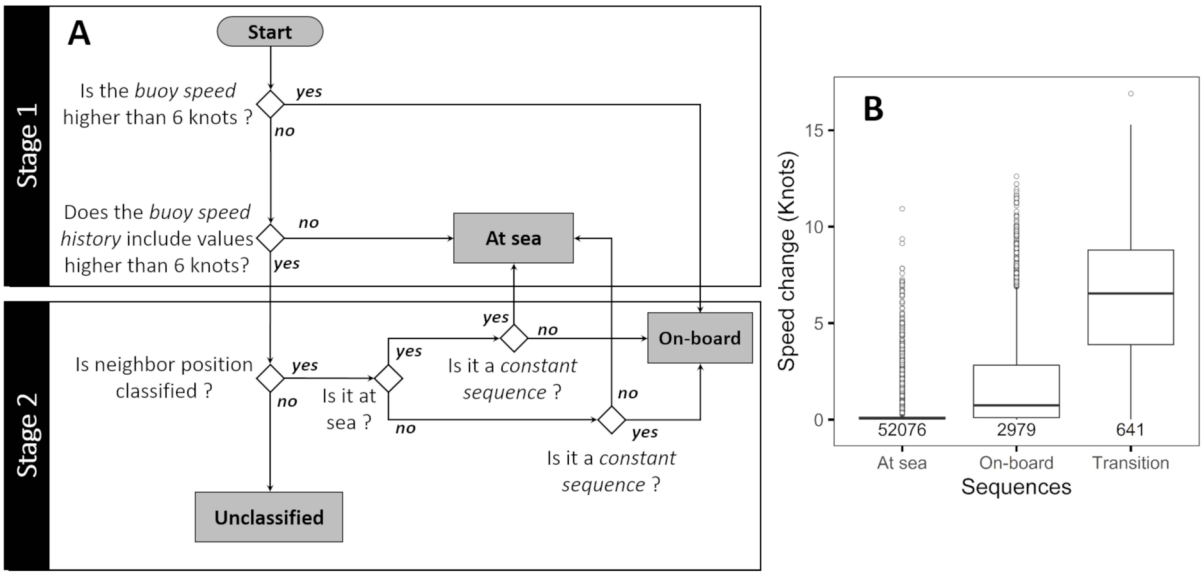

Fig. 4. Description of the kinetic classification algorithm (KiC). (A) KiC flowchart: buoy speed history corresponds to speeds recorded 3 days prior to the buoy position. (B) Example illustrating speed changes in constant ("on-board" – "on-board", "at sea" – "at sea"), and transition states ("on-board" – "at sea", "at sea" – "on-board") from the D1 dataset in the Atlantic Ocean. Values represent the number of data in each sequences.

(*ii*) buoy speed; (*iii*) change in speed; (*iv*) acceleration, (*v*) azimuth (degree), (*vi*) change in azimuth (degree) and (*vii*) time since the first and last observation of the corresponding buoy trajectory (days), as predictive variables. This RF model has previously performed well in discriminating "on-board"/"at sea" positions from a buoy model, (*kappa* = 0.87, further details regarding the model construction and evaluation are provided in Orue et al., 2019a)

3 Results

3.1 Structural errors filtering outputs

For all oceans and datasets, structural errors represented less than 1.5% of all data (Tab. 2). Duplicate rows only occurred in the D1 datasets, while the largest amount of isolated and ubiquitous rows was reported for the D2 dataset in the Indian Ocean.

3.2 Land filtering outputs

Overall, land positions comprised between 0.9% and 8.1% of the data, depending on the dataset and resolution, with a larger proportion observed in the Atlantic Ocean. The difference in the amount of data filtered when using the low and high-resolution shoreline data was minor and more noticeable in the Indian Ocean than in the Atlantic Ocean for both datasets (Tab. 3).

3.3 F5 outputs: on-board / at sea classifications

Cross-comparisons of classifications performed by random forest and kinetic algorithms resulted in high matching coefficients for the four datasets. The two approaches showed very strong agreements for the D2 dataset in the Indian Ocean (99%), and D1 datasets in both oceans (more than 96%). The weakest agreement (94%) was observed for the D2 dataset collected in the Indian Ocean (Tab. 4).

Less than 0.5% of positions from the D1 datasets remained unclassified after processing with the kinetic algorithm. This value was considerably higher for the D2 datasets. For the Indian Ocean, 10 times more unclassified data were observed than in the D1 dataset while for the Atlantic Ocean this figure rose to more than 30 times that of the D1 dataset (Tab. 5). More than 87% of the positions were classified as "at sea" by both approaches, while "on-board" positions varied from 0.1 to 5.5% depending on the algorithm, ocean and dataset.

4 Discussion

This study proposes a set of processing algorithms to be applied to data from satellite-linked buoys provided by the two major tropical tuna purse seine fleets operating in the Atlantic and Indian Oceans. To date, there is still a significant lack of information on the numbers and local densities of DFADs worldwide, although such information is crucial for addressing current issues related to their massive use in tropical tuna fisheries. As outlined by Dagorn et al. (2013), the route towards the sustainable use of DFADs requires a careful assessment of their impacts on tuna stocks and non-target species, as well as on habitats and ecosystems. The data provided by satellite-linked buoys could be a valuable source of information for ensuring that DFADs are monitored adequately, thus supporting the various DFAD management plans adopted by the tRFMOs in recent years.

The first step towards the exploitation of this data is to ensure an adequate level of quality in the data provided. Achieving this requires an appropriate protocol for processing this industry-based data, originally only intended for use at vessel or fleet level, into standardized data that can be utilized for research and management purposes. In this work, the proposed protocol

Table 2. Number and percentage (in brackets) of structural errors for the different datasets in the Atlantic Ocean (AO) and the Indian Ocean (IO).

Filters	D1		D2	
	AO	IO	AO	IO
F1. Duplicated	47 (0.07%)	94 (0.15%)	0 (0%)	0 (0%)
F2. Ubiquitous	11 (0.02%)	11 (0.02%)	0 (0%)	149 (0.66%)
F3. Isolated	38 (0.06%)	46 (0.07%)	91 (0.36%)	174 (0.77%)
Total	96 (0.15%)	151 (0.24%)	91 (0.36%)	323 (1.43%)

Table 3. Number and percentage (in brackets) of data recorded on land for the different datasets in the Atlantic Ocean (AO) and Indian Ocean (IO).

F4. Land	D1		D2	
	AO	IO	AO	IO
Low Res.	5099 (8.1%)	1708 (2.8%)	317 (1.3%)	205 (0.9%)
High Res.	4915 (7.8%)	2352 (3.8 %)	325 (1.3%)	333 (1.5%)

Table 4. Simple matching coefficients (percentage of agreement) between the random forest and the kinetic algorithm classifications for the different datasets in the Atlantic and Indian Oceans.

	Atlantic ocean	Indian ocean
D1	96%	97%
D2	99%	94%

Table 5. Number and percentage (in brackets) of "at sea", "on-board" and unclassified positions from kinetic classification (KiC) and random forest (RF) algorithm in the different datasets for Atlantic (AO) and Indian (IO) Oceans.

F5. Water/Board		D1		D2	
		AO	IO	AO	IO
On-board	RF	2746 (4.4%)	595 (1.0%)	122 (0.5%)	971 (4.3%)
	KiC	3469 (5.5%)	492 (0.8%)	22 (0.1%)	170 (0.7%)
At sea	RF	55,135 (84.5%)	56,020 (91.5%)	22,853 (90.3%)	18,976 (84.5%)
	KiC	54,136 (86.1%)	58,679 (95.9%)	22,897 (90.5%)	21,307 (94.9%)
Unclassified	RF	2010 (3.2%)	2076 (3.4%)	1924 (7.6%)	1941 (8.7%)
	KiC	102 (0.2%)	164 (0.3%)	1977 (7.8%)	726 (3.2%)

focused on the location data provided by the buoys. It used a set of filters, which were applied to different datasets with highly varied structure and characteristics, resulting from the various data-exchange agreements between national research institutes and fleets, as well as the buoy specificities.

The first set of data processing filters (i.e. structural filters) targeted possible satellite communication errors. Despite isolated and ubiquitous positions being rare (<1%), they were present in all datasets. Conversely, duplicated data, which had to be filtered to avoid being counted twice, were only detected in the D1 datasets (both in the Atlantic and in Indian Oceans).

The percentage of data recorded on land showed only minor changes with respect to the resolution of the shoreline data. Lower resolutions could potentially lead to a slight underestimation of location data collected on land, particularly in the Indian Ocean, due to the possible removal of small reefs and islands. However, due to the vast amount of data to be handled (e.g. the "Marine Instruments" buoys operated by the French purse seine fleet, represents, a raw data volume of around 150 million entries for the 2010–2018 period), and the subsequent computational costs required for their processing, the use of low-resolution data should not be excluded, should the study allow it. For example, studies related to DFAD beaching events and their impacts on sensitive habitats (Davies et al., 2017; Escalle et al., 2019; Maufroy et al., 2015) could require full or high resolutions spatial data, whereas more

global indicators on the number of buoys used for catch per unit of effort (CPUE) standardization purposes (Katara et al., 2016, 2017) could be obtained using low-resolution data.

Furthermore, a major issue inherent in buoy data is the discrimination between buoys emitting on-board a vessel from those actually deployed at sea. Although, some buoy models include built-in sensors that identify when the buoy is immersed in seawater, this information is not available for the vast majority of models. Maufroy et al. (2015) were the first authors to describe a processing protocol for buoy data in their analysis of the spatio-temporal patterns in the use of DFADs. They proposed an automatic classification of "at sea" and "on-board" positions, based on a random forest approach trained with a subset of manually pre-classified data, and followed by a post-processing step to improve classification performance. The random forest approach from Orue et al. (2019a), on the other hand, benefited from ground-truthed information provided by sensors integrated in some models of buoys, indicating whether the buoy is in the water or not. The comparisons of the results of this approach with the kinetic algorithm classification proposed in this work, revealed very high agreement rates for the classified positions. This is not surprising, since the analysis of the importance of the predictors in the random forest model revealed that the most relevant variables for discriminating "on-board" from "at sea" positions (i.e. buoy speed and its variation, see Orue et al., 2019a), are also the main parameters on which the KiC algorithm is based. Nevertheless, the random forest algorithm produces a higher number of unclassified positions than the KiC algorithm, as the first or the last positions of the analyzed buoy segments are systematically unclassified due to the impossibility of calculating the predictive parameters associated with them. For the KiC algorithm, unclassified positions are due to both the presence of very short trajectories and the low temporal resolution of data. Since the KiC algorithm analyzes the buoy segment, searching for characteristic changes in speed between neighboring positions, its performance depends on the temporal resolution of the data. Consequently, the larger number of unclassified positions produced by the KiC algorithm in datasets D2, compared to D1, results from the lower temporal resolution of these datasets (only one position per day compared to the higher temporal resolution of datasets D1). Globally, the use of high-resolution data (all the positions recorded in a day), if available, is recommended to reduce the number of unclassified positions.

In addition, despite this study is not including a framework to process the acoustic data collected from the different echosounder buoy models, the filtering approaches presented here remains a requirement before any acoustic data analysis could be conducted. This is because the buoys continuously sample the acoustic energy beneath them, even when they are located on land or on-board a fishing vessel. As such, the processing framework presented here aims to achieve a standardized protocol for preliminary data processing, prior to the implementation of specific algorithms working on the acoustic data that the buoys provide (Lopez et al., 2016; Orue et al., 2019b; Baidai et al., 2020b).

Finally, the main objective of this paper was to propose a standard protocol for processing buoy location data, particularly with respect to datasets that include a mix of fishing fleets and buoy brands and models. Although the provided protocol was designed from the most commonly used buoy models in tropical tuna fisheries, this work could be further improved by extending it to all existing satellite buoy data, including other brands and models for which data availability remains limited to date. However, through its main objective, as well as the definition of a standard data format for buoy location data (Supplementary Tab. 1), as a starting point for the implementation of a general standard format suitable for all the data collected by buoys, this work provides the basis for the development of a dedicated software environment for the processing, analysis, visualization, and possibly the estimation of fish biomass from echosounder buoy data, in order to simplify their use for both scientific and management purposes.

Code availability

All the scripts used for the analysis can be found in the following GitHub repository: https://github.com/yannickBaidai/ProcessingBuoysLocationData.

Supplementary Material

Figure 1: Example of ubiquitous buoy positions. The two red points correspond to two distinct positions provided by the same buoy at the same timestamp (October 10, 2016 at 00:33).

Figure 2: Example of an isolated buoy position. The red points correspond to the position of a buoy separated from its closest neighbors by an inconsistent distance (the speed required to achieve this distance is far greater than the speeds of both tuna purse seine vessels and ocean currents).

Figure 3: Land positions. The white points corresponds to sample of buoy positions detected on land using a 0.05° buffered shoreline data from the GSHHG database (Global Self-consistent, Hierarchical, High-resolution Geography; Wessel and Smith, 1996). The red dashed line represents the buffer zone around the shoreline.

Figure 4: Boxplots of on-board and at-sea buoy speeds (in knots) from the training data used to build the random forest classification algorithm. The training data consisted of location data provided by buoys equipped with sensors that automatically detect their immersion in seawater.

Figure 5: Schematic description of the kinetic classification algorithm (KiC). (A) The green points represent the different positions with undetermined status, recorded along a buoy trajectory. The length of the black arrows roughly reflects the value of the speed associated with the position. (B) The red points correspond to buoy positions classified as "on board" after the first step of the KiC algorithm, given their buoy speed above 6 knots. (C) The step 2 of the KiC algorithm relies on the comparison of changes in buoy speed with those found for "constant" and "transition sequences. Here, the value of the speed change between the first undetermined position following a classified position is consistent with a transition sequence. The undetermined position is therefore classified as "at sea". (D) The

operation is performed along the buoy segment, classifying positions from neighbor to neighbor. (E) The same procedure is then carried out backwards (from the end of the trajectory to the beginning), considering the remaining unclassified positions.

Table 1: Standard data format for buoy location data.

The Supplementary Material is available at https://www.alr.org/10.1051/alr/2022013/olm.

Acknowledgments. This project was funded by the RECOLAPE project (MARE/2016/22 "Strengthening regional cooperation in the area of fisheries data collection" Annex III "Biological data collection for fisheries on highly migratory species"). IRD scientists were also supported by the BLUEMED project ANR project BLUEMED (ANR-14-ACHN-0002). We would like to thank the fishing companies (ORTHONGEL, Echebastar and Atunsa companies in ANABAC and OPAGAC) that shared the buoy data and buoy providers (Marine Instruments, Satlink and Zunibal) for the useful exchanges on the buoys technical characteristics.

References

Baidai Y, Capello M, Billet N, Floch L, Simier M, Sabarros P, Dagorn L. 2017. Towards the derivation of fisheries-independent abundance indices for tropical tunas: Progress in the echosounders buoys data analysis. IOTC-2017-WPTT19-22 Rev 1.

Baidai Y, Dagorn L, Amandè MJ, Gaertner D, Capello M. 2020a. Tuna aggregation dynamics at Drifting Fish Aggregating Devices: a view through the eyes of commercial echosounder buoys. *ICES J Mar Sci* 77: 2960–2970.

Baidai Y, Dagorn L, Amande MJ, Gaertner D, Capello M. 2020b. Machine learning for characterizing tropical tuna aggregations under Drifting Fish Aggregating Devices (DFADs) from commercial echosounder buoys data. *Fish Res* 229: 105613.

Brehmer P, Sancho G, Trygonis V, Itano D, Dalen J, Fuchs A, Faraj A, Taquet M. 2018. Towards an autonomous pelagic observatory: experiences from monitoring fish communities around drifting FADs. *Thalass An Int J Mar Sci* 35: 1–3.

Breiman L. 2001. Random forests. *Mach Learn* 45: 5–32.

Chassot E, Santiago J, Lucas V. 2019. Major reduction in the number of FADs used in the Seychelles purse seine fishery following IOTC limitations. IOTC-2019-WPDCS15-21_Rev1.

Curnick DJ, Feary DA, Cavalcante GH. 2020. Risks to large marine protected areas posed by drifting fish aggregation devices. *Conserv Biol*.

Dagorn L, Holland KN, Restrepo V, Moreno G. 2013. Is it good or bad to fish with FADs? What are the real impacts of the use of drifting FADs on pelagic marine ecosystems? *Fish Fish* 14: 391–415.

Davies T, Curnick D, Barde J, Chassot E. 2017. Potential environmental impacts caused by beaching of drifting fish aggregating devices and identification of management solutions and uncertainties. IOTC-2017-WGFAD01-08 Rev_1.

Escalle L, Hare SR, Vidal T, Brownjohn M, Hamer P, Pilling G. 2021. Quantifying drifting Fish Aggregating Device use by the world's largest tuna fishery. *ICES J Mar Sci* 78: 2432–2447.

Escalle L, Scutt PJ, Pilling G. 2019. Beaching of drifting FADs in the WCPO: recent science, management advice and in-country data collection programmes. *SPC Fish Newsl* 9–14.

Fonteneau A, Chassot E, Bodin N. 2013. Global spatio-temporal patterns in tropical tuna purse seine fisheries on drifting fish aggregating devices (DFADs): TAKING a historical perspective to inform current challenges. *Aquat Living Resour* 26: 37–48.

Fonteneau A, Gaertner D, Nordstrom V. 1999, An overview of problems in the CPUE-abundance relationship for the tropical purse seine fisheries. *Collect Vol Sci Pap ICCAT* 49: 259–276.

Gerritsen H, Lordan C. 2011. Integrating vessel monitoring systems (VMS) data with daily catch data from logbooks to explore the spatial distribution of catch and effort at high resolution. *ICES J Mar Sci* 68: 245–252.

Gershman D, Nickson A, O'Toole M. 2015. Estimating The Use of FADS Around the World, *The Pew Chari table Trusts*.

Goñi N, Santiago J, Murua H, Fraile I, López J, Krug I, Ruiz J, Pascual P. 2017. Verification of the limitation of the number of FADs and best practices to reduce their impact on by-catch fauna. *Collect Vol Sci Pap ICCAT* 73: 988–1004.

Grande M, Capello M, Baidai Y, Uranga J, Boyra G, Quincoces I, Orue B, Ruiz J, Zudaire I, Murua H, Depetris M, Floch L, Santiago J. 2020. From fishermen to scientific tools: Progress on the recovery and standardized processing of instrumented buoys data. *Collect Vol Sci Pap ICCAT* 76: 881–891.

Hintzen NT, Bastardie F, Beare D, Piet GJ, Ulrich C, Deporte N, Egekvist J, Degel H. 2012. VMStools: Open-source software for the processing, analysis and visualisation of fisheries logbook and VMS data. *Fish Res* 115–116: 31–43.

Imzilen T, Chassot E, Barde J, Demarcq H, Maufroy A, Roa-Pascuali L, Ternon J-F, Lett C. 2019. Fish aggregating devices drift like oceanographic drifters in the near-surface currents of the Atlantic and Indian Oceans. *Prog Oceanogr* 171: 108–127.

Imzilen T, Lett C, Chassot E, Kaplan DM. 2021. Spatial management can significantly reduce dFAD beachings in Indian and Atlantic Ocean tropical tuna purse seine fisheries. *Biol Conserv* 254: 108939.

Imzilen T, Lett C, Chassot E, Maufroy A, Goujon M, Kaplan DM. 2022. Recovery at sea of abandoned, lost or discarded drifting fish aggregating devices. *Nat Sustain* 34: 731–754.

Katara I, Gaertner D, Billet N, Lopez J, Fonteneau A, Murua H, Baez JC. 2017. Standardisation of skipjack tuna CPUE for the EU purse seine fleet operating in the Indian Ocean. IOTC-2017-WPTT19-38.

Katara I, Gaertner D, Maufroy A, Chassot E. 2016. Standardization of catch rates for the eastern tropical atlantic bigeye tuna caught by the French purse seine DfAD fishery. *Collect Vol Sci Pap ICCAT* 72: 406–414.

Lambert GI, Jennings S, Hiddink JG, Hintzen NT, Hinz H, Kaiser MJ, Murray LG. 2012. Implications of using alternative methods of vessel monitoring system (VMS) data analysis to describe fishing activities and impacts. *ICES J Mar Sci* 69: 682–693.

Lee J, South AB, Jennings S. 2010. Developing reliable, repeatable, and accessible methods to provide high-resolution estimates of fishing-effort distributions from vessel monitoring system (VMS) data. *ICES J Mar Sci* 67: 1260–1271.

Lennert-Cody CE, Moreno G, Restrepo V, Román MH, Maunder MN. 2018, Recent purse-seine FAD fishing strategies in the eastern Pacific Ocean: what is the appropriate number of FADs at sea? *ICES J Mar Sci* 75: 1748–1757.

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.

Lopez J, Moreno G, Ibaibarriaga L, Dagorn L. 2017. Diel behaviour of tuna and non-tuna species at drifting fish aggregating devices (DFADs) in the Western Indian Ocean, determined by fishers' echo-sounder buoys. *Mar Biol* 164: 44.

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.

Mannocci L, Baidai Y, Forget F, Tolotti MT, Dagorn L, Capello M. 2021. Machine learning to detect bycatch risk: novel application to echosounder buoys data in tuna purse seine fisheries. *Biol Conserv* 255: 109004.

Maufroy A, Chassot E, Joo R, Kaplan DM. 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.

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.

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.

Orue B, Lopez J, Moreno G, Santiago J, Boyra G, Soto M, Murua H. 2019a. Using fishers' echo-sounder buoys to estimate biomass of fish species associated with drifting fish aggregating devices in the Indian Ocean. *Rev Investig Mar AZTI* 26: 1–13.

Orue B, Lopez J, Moreno G, Santiago J, Boyra G, Uranga J, Murua H. 2019b. From fisheries to scientific data: a protocol to process information from fishers' echo-sounder buoys. *Fish Res* 215: 38–43.

Santiago J, Lopez J, Moreno G, Quincoces I, Soto M, Murua H. 2016. Towards a Tropical Tuna Buoy-derived Abundance Index (TT-BAI). *Collect Vol Sci Pap ICCAT* 72: 714–724.

Santiago J, Uranga J, Quincoces I, Grande M, Murua H. 2020. A Novel Index of Abundance of Skipjack in the Indian Ocean Derived From Echosounder Buoys. *Collect Vol Sci Pap ICCAT* 76: 321–343.

Scott GP, Lopez J. 2014. The use of FADs in tuna fisheries, European Union. Directorate General for Internal Policies.

Sokal RR. 1958. A statistical method for evaluating systematic relationships. *Univ Kansas Sci Bull* 38: 1409–1438.

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

Wain G, Guéry L, Kaplan DM, Gaertner D. 2020. 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*.

Wessel P, Smith WHFF. 1996. A global, self-consistent, hierarchical, high-resolution shoreline database. *J Geophys Res Solid Earth* 101: 8741–8743.

Cite this article as: Baidai Y, Uranga J, Grande M, Murua H, Santiago J, Quincoces I, Boyra G, Orue B, Floch L, Capello M. 2022. A standard processing framework for the location data of satellite-linked buoys on drifting fish aggregating devices. *Aquat. Living Resour.* 35: 13