Product Description Document (PDD)

Buoy dataset

Contributors: Ivan Manso-Narvarte, Lohitzune Solabarrieta, Ainhoa Caballero, Asier Anabitarte and Jose A. Fernandes-Salvador.

Version	Changes made by	Nature of changes
V_2023_08	I. Manso-Narvarte, L. Solabarrieta, A. Caballero, A. Anabitarte, J.A. Fernandes-Salvador	Document generation

1 Introduction

Position data of buoys used in fisheries represent an opportunity for obtaining current (zonal and meridional components) data, potentially complementing other observation platforms (e.g. scientific drifters). This document describes a dataset built from near-surface current velocity data along trajectories in the Indian Ocean from October 2010 to December 2020, generated in the framework of the SusTunTech project. Current velocity data were monthly averaged and provided into a 4.5º x 4.5º grid. The obtained velocities have been validated against scientific drifters (derived velocity data at -15 m). Therefore, the obtained currents are considered representative of that depth.

2 Data provenance

Buoys are attached to structures typically made of a bamboo raft, equipped with floats to ensure buoyancy, and a subsurface structure built of non-entangling material (e.g., ropes or panels). This submerged structure can go from a few meters (2-3 m) to 80 m depth and might have different forms, depending on the area/season and the ocean. The buoys transmit position and date data from every 15 minutes to once a day. The information is sent by GPS to the servers of the buoy provider and forwarded to fishing vessels. From position and time data, current velocities can be inferred, thus becoming a new invaluable source of information on near-surface currents.

3 Dataset description

The buoy dataset described herein covers the Indian Ocean for the period from 2010 to 2020. Positions were obtained by GPS usually every 24 hours and 99.3% of the time differences between consecutive positions were of 72 hours or less. Once the raw data of the positions were obtained, spurious data were removed to get rid of duplicated data, data on land, data on-board etc. To that end, data quality control (QC) filters were applied (see next section). Then, positions were regularly interpolated by the Kriging technique in 6 hours intervals (Hansen and Herman, 1989; Hansen and Poulain, 1996); and velocities were derived from finite differences of their fixed position using a 12-hours centred scheme. Finally, velocities were monthly averaged onto a 4.5º x 4.5º grid.

4 Data QC, processing and validation

The raw data were QC by filtering erroneous locations mainly related to failures in satellite communication and location data acquisition; identifying buoys on land positions; and identifying on-board positions. Therefore, several QC filters were applied based on Baidai et al. (2017, 2022), the Annexe 13 of the RECOLAPE project (Ruiz et al., 2019) and Hansen and Poulain (1996). The QC filters used are listed in Table 1 and were applied in a sequential way.

Table 1*. Sequentially applied filters for QC raw data.*

FILTER	Description
F1. Duplicated	Adjacent points with the same date and position for the same buoys are erroneous. When these points are found, the first datum is kept, and the following ones are removed.
F2. Ubiquitous	Multiple positions for the same date and buoys are also erroneous. In this case, the first point is kept if the second one is further than 1 km (as in Ruiz et al., 2019). If not, they are all removed.
F3. Single location	Trajectories of one point are removed since they do not provide any information of the ocean drift.
F4. Isolated	Points with aberrant velocities (>18 m/s, as in Baidai et al. (2017)) with respect to adjacent points are removed.
F5. Land	Buoys with continental positions are removed using the NOAA - Global Self-consistent, Hierarchical, High-resolution Shoreline (Wessel and Smith, 1996), with a buffering of 0,05 arc degrees around the shoreline (as in the RECOLAPE project).
F6. Bathymetry	Positions with depths shallower than 50 m are removed since the structure attached to the buoy usually reaches 40-60 m depth, and therefore the trajectory would be affected by the drag against the sea floor.
F7. Zero velocity	Adjacent points with the same date are supposed to be removed before this step; thus, points with zero velocity corresponds to adjacent points in time of zero distance, which is not realistic. Thus, these points are removed.
F8. On board	Buoys are many times recovered by the ships and transported to other areas and then deployed again. These paths do not represent the drift of the buoy in the ocean; therefore, 'on-board' points must be removed. For this purpose, a classification system is used based on velocities and acceleration values estimated between adjacent points. Herein, the basic ideas are presented (for more details see Baidai et al. (2017,2022)): • First, the points with velocity > 3m/s are flagged as 'on-board' since ocean currents in the Indian ocean are below this threshold value. • The rest of the points (≤3 m/s), if the velocities during the previous 3 days have also been ≤3 m/s, are classified as 'at-sea' since typically buoys are at sea at least for 3 days. The remaining points are flagged to 'on-board' or 'at-sea' by comparing their acceleration with the distributions of accelerations estimated for constant (sea-sea or on board-on board) and transition (on board-sea and vice versa) sequences. These comparisons are made using the t-test at a confidence level of 0.95. Once the kind of sequence of the points are known the state ('on-board' or 'at-sea') of each point can be deduced from initially known states.
F9. Velocity	Finally, a velocity filter is applied to remove the remaining spurious data. The method used was proposed by Hansen and Poulain (1996) and has been used to process several drifter datasets (e.g., AOML and CMEMS drifters). The filter is applied to each trajectory:

	• The velocity between adjacent points is first measured and a velocity threshold is set to determine if a point is bad or not. A point is considered bad when its velocity exceeds this threshold. This threshold is set as the mean·4·std of the velocity (as in CMEMS drifter datasets) of each trajectory. • When a bad point is detected, the velocity is calculated with respect the next point… and so on until a good velocity (which does not exceed the fixed threshold) is found. • Then, this process is repeated backwards in time (within the trajectory). • If a point is good/bad for both forward and backward directions in time, it is flagged as good/bad. • For the points which are good/bad for only one direction the direction that has more good points is taken as reference. The points with bad data are finally removed.

After the QC, 12.68% of the raw data were removed. Note that trajectories with on-board sequences (which were removed) were split. Also, trajectories with time differences bigger than 3 days were split for avoiding interpolating data between two points that are far from each other (time differences equal to or smaller than 3 days are 99.3% of the data). The interpolation was carried out for obtaining position data every 6 hours. This was achieved by the Kriging technique, which is based on the optimal weighting of the nearby observations of the point to be interpolated (Hansen and Herman, 1989; Hansen and Poulain, 1996). Latitude and longitude data were separately interpolated in time, considering the data of the previous and next 5 dates as in Hansen and Poulain (1996). Once the positions were interpolated, the velocities were estimated using a 12-hours centred scheme and then decomposed into zonal and meridional components. A few positions (0.04% of the data) still provided unrealistic velocities higher than 3 m/s for one of the current components (peak speeds of 2.6 m/s were observed in the Agulhas current (Lutjeharms, 2006)); thus, those positions were removed.

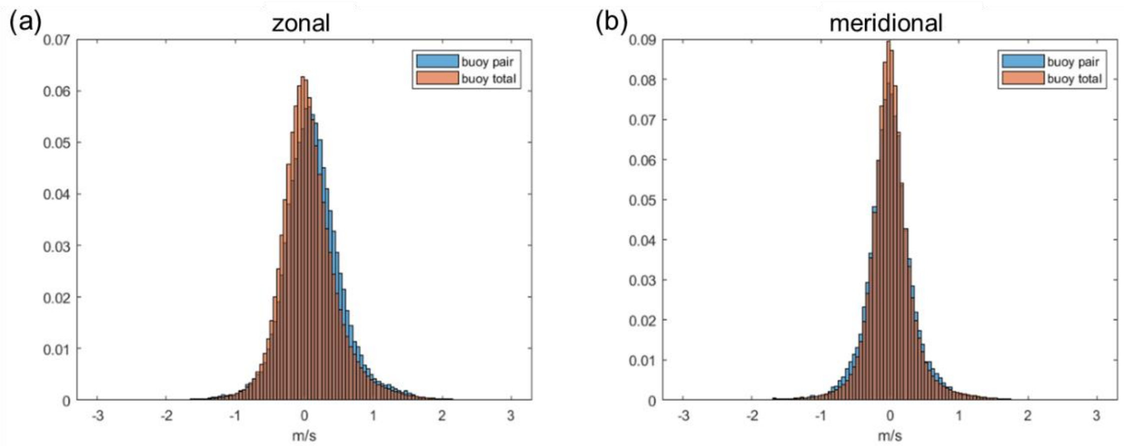

Figure 1. *Normalized histograms of the buoy data considered for the comparison against the total buoy data, compared to the total buoy dataset for zonal (a) and meridional (b) current components.*

To assess the goodness of the velocity data obtained, they were compared against drifter data. In a previous study, buoy-derived and drogued (at -15 m) drifter-derived velocities showed quite a good agreement in the Indian Ocean (Imzilen et al., 2019); therefore, a similar comparison was carried out for this dataset. The scientific drifter dataset used was the Global Drifter Program - 15 m drogued data (https://www.aoml.noaa.gov/phod/gdp/interpolated/data/all.php), which have been processed following Hansen and Poulain (1996) and are a widely used and reliable ocean current data source. The period covered by the drifter dataset is the same as reported by our buoy datasets. To compare buoy-derived against drifter-derived velocities, the same date and a maximum distance of 10 nm (as in Imzilen et al., 2019) were found and compared. Before carrying out these comparisons, the representativity of the buoy data, considered within those pairs, was analysed. To this end, the normalized histograms of the total buoy data and the buoy data considered within the pairs were compared for each current component (Figure 1), showing the percentage of the number of data (Y-axis) corresponding to each velocity value (X-axis). These comparisons provide an idea of the similarity in the data distribution for each dataset and showed a high degree of similarity with a higher percentage of negative (small) values for buoy zonal (meridional) velocities.

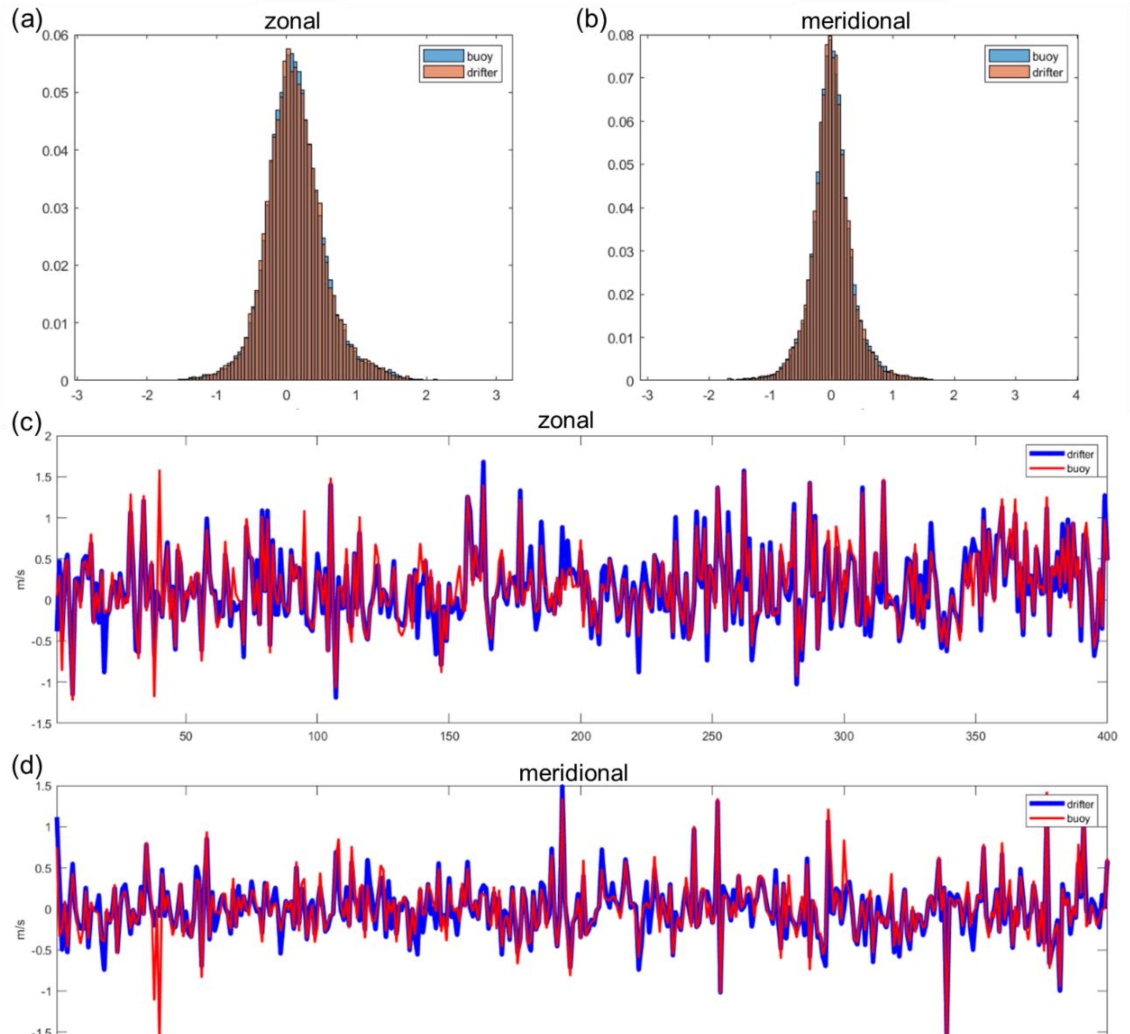

Figure 2. *Normalized histograms of the buoy and drifter data considered for the comparison for zonal (a) and meridional (b) current components. Subsamples of 400 points of the data series of zonal (d) and meridional (d) components of the drifter (blue) and buoy (red)-derived velocities (in m/s).*

After observing that the buoy data used for the comparisons could represent the entire buoy dataset, drifter and buoy datasets were compared. The normalized histograms of each dataset show almost the same distributions of the velocities for each current component (Figure 2ab). This agreement is also observable in the time series, as shown by a subsample of the series (Figure 2cd). In addition to the good agreement, the results show a low bias between drifter and buoy velocities.

The scatterplots of the time series (Figure 3) show that the clouds of points are relatively well adjusted to the 1:1 isolines of each current component, which represents the perfect agreement between drifter and buoy-derived velocities. In fact, the linear adjustments of the clouds have slope values close to 1 (which corresponds to the above-mentioned ideal isoline), again showing similar distributions and strong linear relationships between both datasets. The correlations between the data pairs are 90% and 89% for the zonal and meridional components, respectively, once again showing a high agreement. A slight overestimation of the buoy-derived velocities is observed compared to drifter-derived velocities.

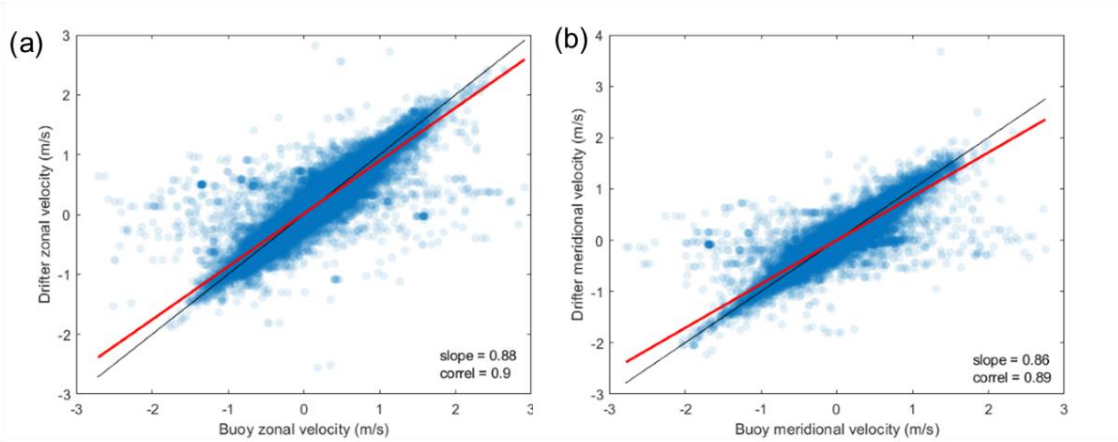

Figure 3. *Velocity comparisons between the buoy and drifter pairs for the meridional (left) and zonal (right) components. The red line indicates the major axis regression model, and the black line indicates the 1:1 isoline.*

Concerning the performance evaluation (Table 2), the mean of the absolute value of the difference between both datasets is around 10 cm/s and the root mean square difference (RMSD) values of around 18 cm/s. All these comparisons were made with data pairs that did not correspond to the same position (maximum distance of 10 nm). This means that these error ranges would be lower for pairs that were closer to each other and consequently, for values at the same position that would be the case of a perfect comparison strategy (which is not realistic since it is extremely unlikely to have the same position in the ocean). The RMSD relative (RRMSD) to the root mean square (RMS) of the drifter dataset ranges between 0.42 and 0.48. The depth of the structure attached to the buoys varies, going from 3 to 80 meters, being in the Indian Ocean normally shallower than 50 m depth. Therefore, the velocities derived from buoy positions might be representative of different depths. Nevertheless, given the good agreement between the drifter-derived and buoy-derived currents, the buoy dataset can be considered a robust dataset and the velocities contained in this dataset will be considered representative of the same depth as the drifters', that is -15 m. Finally, current velocities inside a 4.5º x 4.5º grid cell were monthly averaged and data were provided at the centre of each cell.

Table 2. *The mean of the absolute value of the difference between buoy and drifter-derived velocities and its standard deviation (STD), RMSD between both datasets, the RMSD relative RRMSD to the RMS of the drifter dataset and the slope and correlation values shown in Figure 3 for zonal (U) and meridional (V) current components.*

<table>
<tr><th></th><th><|drifter-buoy|> (STD) (cm/s)</th><th>RMSD (cm/s)</th><th>RRMSD (=RMSD/RMS_drifter)</th><th>Correlation</th><th>Slope</th></tr>
<tr><td>U</td><td>10.37 (15.64)</td><td>18.76</td><td>0.42</td><td>0.90</td><td>0.88</td></tr>
<tr><td>V</td><td>9.22 (13.74)</td><td>16.55</td><td>0.48</td><td>0.89</td><td>0.86</td></tr>
</table>

5 Product description

The gridded subsurface current velocity dataset is created as an individual NetCDF file. Note that the dataset published in data repositories, due to the standards of these repositories, may contain minor differences in the global attributes that do not affect the data. The general characteristics of these datasets are described in Table 3.

Table 3. *General description of the dataset.*

Spatial Coverage	Indian Ocean
Spatial Resolution	Gridded
Temporal Coverage	From October 2010 to December 2020
Temporal Resolution	Monthly
Variables	Zonal and meridional current velocities; and Metadata
File format	NetCDF 1.4
Dataset location	
Access to the dataset	Open access

Table 4 shows the variables that are embedded in each dataset, which compose the bulk of the datasets.

Table 4. *Variables contained in the dataset.*

Variable name	Description	Units
TIME	Date of the data	Days since 1950-01-01T00:00:00z
DEPH	Depth of the measurement	m
LATITUDE	Latitude of the data position	Degrees north
LONGITUDE	Longitude of the data position	Degrees east
EWCT	West-East current component	m/s
NSCT	South-North current component	m/s
EWCS	West-East current component standard deviation	m/s
NSCS	South-North current component standard deviation	m/s

Table 5 provides much more information and details about the dataset and enables a deep understanding of them, contributing to making the datasets FAIR (Findable, Accessible, Interoperable, Reusable). The proposed main and global attribute metadata followed the structures and standard names of the 'NetCDF CF Metadata Convention Standard Name Table Version 1.6' (https://cfconventions.org/cf-conventions/v1.6.0/cf-conventions.html). They also followed the ISOs 19115 and 19139, and several non-standard attributes. The SeaDataNet common vocabularies were also used. The use of common standard metadata and vocabularies made the dataset more FAIR.

Table 5. *Global attributes contained in the metadata of the dataset.*

Global attribute	Value
acknowledgment	'These data were collected by ships, processed by AZTI and made freely available by Marine Instruments through SusTunTech project and the programs that contribute to it'
area	'Indian Ocean'
cdm_data_type	'Point'
citation	'These data were collected by ships, processed by AZTI and made freely available by Marine Instruments through SusTunTech project and the programs that contribute to it'
comment	'The velocities are derived from buoy positions and dates: First raw data were quality controlled, then interpolated to obtain trajectory points every 6 hours, and finally velocities were derived from those positions. Given the good agreement with drifter-derived velocities at -15 m, the velocities of this dataset are considered representative of that depth. The final product is a set of monthly mean current velocities in a grid of cells of 4.5 degrees x 4.5 degrees'
contact_email	'sustuntech_WP4@azti.es'
Conventions	' '
creator_email	'sustuntech_WP4@azti.es'
creator_name	'AZTI'
creator_type	'Institution'
creator_url	'https://www.azti.es/'
data_assembly_center	'AZTI'
data_language	'eng'
data_mode	'D'
data_type	'gridded currents'
date_update	2020-12-01 00:00
distribution_statement	'These data are public and free of charge. User assumes all risk for use of data. User must display citation in any publication or product using data. User must contact PI prior to any commercial use of data'
DOI	
format_version	'1.4'
geospatial_lat_max	'11.25'
geospatial_lat_min	'2.25'
geospatial_lat_units	'degrees_north'
geospatial_lon_max	'65.25'
geospatial_lon_min	'56.25'
geospatial_lon_units	'degrees_east'
geospatial_vertical_max	'-15'
geospatial_vertical_min	'-15'
geospatial_vertical_units	'm'
history	2021-03-01 10:30 - 2023-03-09 23:30 data collected. 2023-06-08 12:34 netCDF file created using Matlab software
infoUrl	'https://www.sustuntech.eu/communicationmaterials/deliverables/'
institution	'AZTI (Spain)'
institution_edmo_code	'1623'
institution_references	'AZTI'
keywords	'OCEAN CURRENTS, FISHERIES'
keywords_vocabulary	'GCMD Science Keywords'
last_update	2023-06-08 12:34
license	'buoy-derived current velocity dataset is licensed under a Creative Commons Attribution 4.0 International License. You should have received a copy of the

	license along with this work. If not, see http://creativecommons.org/licenses/by/4.0/'
metadata_language	'eng'
naming_authority	'Marine Instruments, AZTI
NetCDF_format	'netcdf4_classic'
NetCDF_version	'netCDF-4 classic model'
pi_name	' '
platform_code	' '
platform_name	' '
publisher_email	'sustuntech@globalmarine.es'
publisher_name	'Marine Instruments'
publisher_type	'Institution'
publisher_url	'https://www.marineinstruments.es/es/'
qc_manual	'QC filters were applied based on Baidai et al. (2017, 2022), the Annexe 13 of the RECOLAPE project (Ruiz et al., 2019) and Hansen and Poulain (1996). '
references	'https://www.sustuntech.eu/, https://www.sustuntech.eu/communicationmaterials/deliverables/'
Standard_name_vocabulary	'NetCDF Climate and Forecast Metadata Convention Standard Name Table Version 1.6; and https://www.seadatanet.org/Standards/Common-Vocabularies'
summary	'The velocities are derived from buoy positions and dates: First raw data were quality controlled, then interpolated to obtain trajectory points every 6 hours, and finally velocities were derived from those positions. Given the good agreement with drifter-derived velocities at -15 m, the velocities of this dataset are considered representative of that depth. The final product is a set of monthly mean current velocities in a grid of cells of 4.5 degrees x 4.5 degrees'
time_coverage_end	2020-12-01 00:00
time_coverage_resolution	'monthly'
time_coverage_start	2010-11-01 00:0
title	'Buoy-derived monthly currents in the Indian Ocean by SusTunTech project'
update_interval	'void'
wmo_inst_type	' '
wmo_platform_code	' '

6 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. 12pp.

Baidai, Y., Uranga, J., Grande, M., Murua, H., Santiago, J., Quincoces, I., ... & Capello, M. (2022). A standard processing framework for the location data of satellite-linked buoys on drifting fish aggregating devices. Aquatic Living Resources, 35, 13.

Hansen, D. V., & Poulain, P. M. (1996). Quality control and interpolations of WOCE-TOGA drifter data. Journal of Atmospheric and Oceanic Technology, 13(4), 900-909.

Hansen, D. V., & Herman, A. (1989). Temporal sampling requirements for surface drifting buoys in the tropical Pacific. Journal of Atmospheric and Oceanic Technology, 6(4), 599-607.

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. Progress in oceanography, 171, 108-127.

Lutjeharms, J. R. (2006). The Agulhas current retroflection (pp. 151-207). Springer Berlin Heidelberg.

Ruiz, J., Depetris, M., Grande, M., Tserpes, G., Carbonara, P., Bach, P., Duparc, A., Cauquil, P., Krug, I., Spedicato, M. T., Capello, M., Gaertner, D., Mugerza, E., Thasitis, I., Garibaldei, F., Mariani, A., Santiago, J., Murua, H., Pascual, P., Baez, JC., Abascal, F., Uranga, J. & Baidai, Y. (2019). Strengthening regional cooperation in the area of large pelagic fishery data collection (acronym: RECOLAPE). Annex 13: Developing dedicated algorithms to provide the total number of active beaconed FADs.