Machine learning for characterization of tuna aggregations under drifting FADs from commercial echo sounder buoys data

Yannick Baidai (1,2), Monin Justin Amande (2), Daniel Gaertner (1), Laurent Dagorn (1), Manuela Capello (1)

(1) MARBEC, Univ Montpellier, CNRS, Ifremer, IRD, Sète, France ; (2) Centre de Recherches Océanologiques (CRO), DRAV, 29, Rue des pêcheurs, BPV 18, Abidjan, Côte d'Ivoire. Main authors contact details: yannick.baidai@ird.fr, manuela.capello@ird.fr

Summary

The echosounder buoys that equip the drifting FADs (DFADs) used in tropical tuna purse-seine fisheries offer unique opportunities to observe pelagic communities and can potentially provide fishery-independent abundance indices for tropical tunas. Such data, however, differs considerably in nature depending on the software and hardware of the different brands and models, which is very often a limitation to their scientific use. This work proposes a new methodology based on machine learning, for characterizing fish aggregations under DFADs from the acoustic data collected by these devices. Our approach consists in specific processing of acoustic information, combined with random forest algorithm, to translate the raw data provided by the buoys into metrics of tuna presence and abundance. The classifications were built from a training dataset constituted from cross-referencing of acoustic and catch data recorded on the same schools, considered as tuna occurrences, and acoustic data recorded a few days after new DFAD deployments, or before DFAD visits without sets, considered as tuna absences. Our results evidenced that detection of tuna aggregations from echo sounder buoys was typically more effective during daytime periods and at ocean-specific depths. Our approach shows very good efficiency for pattern recognition of presence and absence of tuna aggregation under DFADs regardless of the ocean (75 and 85 % of correct predictions respectively in the Atlantic and Indian Ocean), but is less accurate for estimating the precise range of aggregation sizes. This work is one of the milestones towards of development of novel fishery-independent indices of abundance for tropical tuna based on acoustic data.

Introduction

Although still poorly explained, the natural aggregative behavior of tropical tuna under floating objects is widely exploited by fishermen around the world, to improve their catches (Kakuma 2001; Dempster and Taquet 2004). In the tropical tuna purse-seine fishery, the use of artificial drifting devices referred to as "drifting fish aggregating devices" or DFADs has significantly improved the probability of success of tuna seiners' fishing sets (Fonteneau *et al.* 2000) and, currently, over half of the yearly tuna catches worldwide originate from DFADs (Miyake *et al.* 2010; Fonteneau *et al.* 2013, Dagorn et al. 2013). It has been estimated that between 50 000 and 100 000 DFADs, generally instrumented with GPS and echosounder devices, are deployed every year in the world, by the purse-seine fishery in the three major oceans (Baske *et al.* 2012; Maufroy *et al.* 2017). The number and wide spatial distribution of DFADs across the different oceans, coupled with their constantly evolving technology (Lopez *et al.* 2014), grant to these fishing tools the status of a privileged platform for the observation of the pelagic communities that aggregate under floating objects (Moreno *et al.* 2016; Brehmer *et al.* 2018). For instance, the exploitation of data collected at FADs could provide fisheries-independent abundance indices for tropical tuna, as highlighted by the works of Capello et al. (2016) and Santiago et al. (2016). However, the characterization of fish aggregations under FADs through echosounder buoys remains severely limited by the reliability and variability of the information provided, which depends on the hardware and software characteristics of the buoys, which vary between manufacturers (Lopez *et al.* 2014; Moreno *et al.* 2016; Santiago *et al.* 2016). Currently echosounder buoys data are of heterogeneous types, natures and formats and few studies provided an assessment of their accuracies to the purpose of exploiting them for scientific purposes (Lopez *et al.* 2016; Baidai *et al.* 2017). We present here a dedicated approach using data collected by echosounder buoys to characterize tuna aggregations under DFADs, based on machine learning algorithms. The accuracy of the method is evaluated in the Atlantic and Indian Oceans.

Material and Methods

This study focused on the M3I Marine Instruments echosounder buoy, one of the main buoy models which equips the DFADs deployed by the French fleet of tuna purse-seiners. This buoy internally processes the raw acoustic signals sampled by 3-meter depth layers into discrete indices (varying from 0 to 7). By cross referencing the echosounder buoys database hosted by Ob7/IRD[1] and the fishing activities reported in logbook and observer databases, we designed a training dataset for the Atlantic and Indian oceans. This dataset consisted of acoustic data associated to catch data, as well as to new DFADs deployments and DFAD visits without fishing sets (both considered as DFADs without tuna). The dimensions of acoustic data were reduced following two steps: (1) a temporal sub-sampling, selecting the best echo (namely, the highest value of the total acoustic energy) every 4 hours over 24 hours and (2) a spatial clustering, which combines the discrete indices recorded in the different depth layers into scores per group of homogenous layers. The dimensionality reduction transforms the 12x50 acoustic matrix collected over a full day of sampling (corresponding to 12 acoustic samples for the default mode for 50 depth layers) into a matrix of 6 rows (for the groups of layers) and 6 columns (for each time slot), referred to as "daily acoustic matrix", which summarizes the acoustic information on a daily scale. The final dataset consisted of **3381 acoustic matrices from catch data** (916 and 2465 respectively in Atlantic and Indian oceans) labelled as tuna presence, and **8336 matrices identified as tuna absence** (880 and 7456 for Atlantic and Indian oceans, respectively). Two types of classification models were built, relying on the random forest algorithm (Breiman 2001): (1) a **binary model describing the absence or presence of tuna,** and (2) **a multiclass classification model** considering different sizes of aggregations under DFAD (no tuna, less than 10 tons, between 10 and 25 tons, more than 25 tons). Model training and evaluation were performed through a two-fold cross-validation replicated 10 times (Dietterich 1998), and importance of the different predictors in the classification process, in each ocean, was assessed through the analysis of mean decrease accuracy (Breiman, 2001).

Results and discussion

The classification models exhibited a good performance in successfully discriminating tuna presence and absence, in both oceans, with an overall accuracy of **75 and 85 % in the Atlantic and Indian ocean, respectively**. In the Atlantic Ocean, the model is highly effective in detecting DFAD aggregations with tuna (**0.82 for sensitivity**). However, its outputs remain characterized by a noticeable level of false positive classifications (**0.67 for specificity**). An opposite trend is observed in the Indian Ocean. The evaluation metrics indicates a slightly lower ability than in Atlantic ocean for correctly classifying tuna presence (**0.78 for sensitivity**), contrasted with a much higher performance for recognition of acoustic patterns from non-tuna aggregations (**0.91 for specificity**). Multi-class classifications were considerably less efficient than binary classifications, as indicated by their **low overall accuracies (50 and 45 %) observed for Atlantic and Indian oceans** respectively. This poor performance could potentially result from biases induced by the multispecific composition of the tuna aggregations. Indeed, acoustic signature associated with a given aggregation class size can significantly vary depending on the specific composition of the aggregation (mix of species with different target strengths). In the Atlantic, the highest proportion of misclassifications is associated with tuna aggregations between 10 and 25 tons (**0.23 in precision**), whereas tuna schools below 10 tons, and above 25 tons, share relatively similar classification performances (**precision of 0.36 and 0.35 respectively**). In the Indian ocean, tuna schools over 25 tons constitute the best detected size class (precision of 0.43), while the classification performances of small and intermediate aggregation sizes (less than 10 tons, and between 10 and 25 tons) are less good (**precision of 0.37 and 0.31 respectively**). The importance of predictive factors is highly ocean-dependent. In the Atlantic Ocean, the characterization of fish schools under DFADs is mainly driven by a set of acoustic variables recorded from 6 to 45 meters and from 0 to 4 pm (8 am to 4 pm for multiclass classification), while in the Indian Ocean, the main predictors correspond to deeper layers (from 30 to 150 meters), over a wider period of time (4 am to 4 pm) (**Figures 1 and 2**). Although several reasons can be mentioned to explain these differences, one of the most likely is related to a possible ocean-specific vertical distribution of fish under DFADs. Further studies, comparing the presence-at-depth profiles of tuna species in the two oceans, should be conducted to confirm these findings.

Assessing the reliability of biomass estimates obtained from the echosounder buoy constitutes an essential step towards the exploitation of this data for scientific purposes. This work opens the way towards future studies aiming at deriving fisheries-independent indices for assessing the spatial distribution and abundance of tropical tuna in both oceans.

[1] Observatoire des Écosystèmes Pélagiques Tropicaux exploités (Institut de Recherche et de Développement)

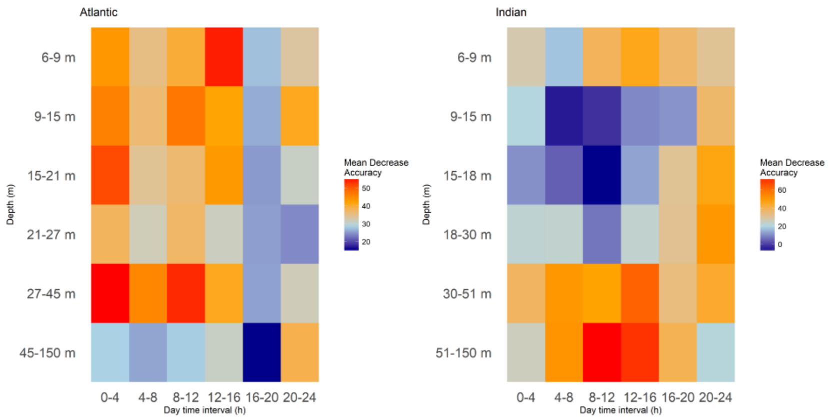

Figure 1 : Importance of depth layers and day period in presence/absence classification for the Atlantic and Indian oceans. Each cell represents a combination between a depth and a time period. Color indicates the relevance of the predictor in the classification

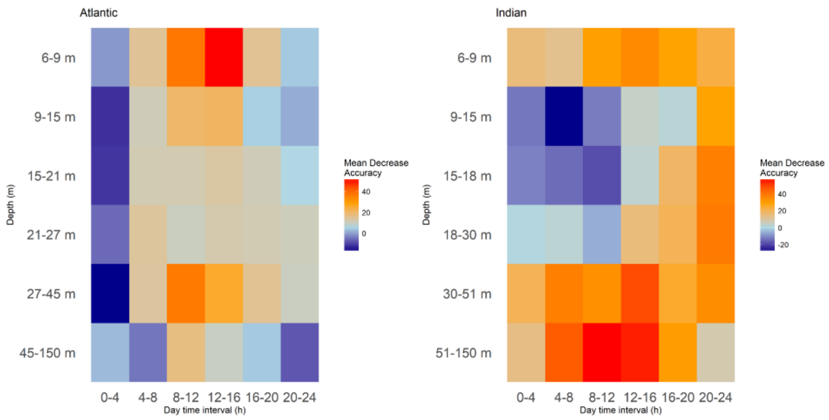

Figure 2 : Importance of depth layers and day period in multiclass classification for the Atlantic and Indian oceans. Each cell represents a combination between a depth and a time period. Color indicates the relevance of the predictor in the classification

References

Baidai, Y., Capello, M., Billet, N., Floch, L., Simier, M., Sabarros, P. and 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_Rev1. Mahé.

Baske, A., Gibbon, J., Benn, J. and Nickson, A. (2012) Estimating the use of drifting Fish Aggregation Devices (FADs) around the globe. *Pew Discussion Paper*, 8.

Brehmer, P., Sancho, G., Trygonis, V., et al. (2018) Towards an Autonomous Pelagic Observatory: Experiences from Monitoring Fish Communities around Drifting FADs. *Thalassas: An International Journal of Marine Sciences*.

Breiman, L. (2001) Random forests. *Machine Learning* **45**, 5–32.

Dempster, T. and Taquet, M. (2004) *Fish aggregation device (FAD) research: Gaps in current knowledge and future directions for ecological studies*, (Vol. 14).

Dietterich, T.G. (1998) Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. *Neural Computation* **10**, 1895–1923.

Fonteneau, A., Chassot, E. and 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. *Aquatic Living Resources* **26**, 37–48.

Fonteneau, A., Pallarés, P. and Pianet, R. (2000) A worldwide review of purse seine fisheries on FADs. In: *Pêche thonière et dispositifs de concentration de poissons, Caribbean-Martinique, 15-19 Oct 1999*. Mahé, pp 15–35.

Kakuma, S. (2001) Current, catch and weight composition of yellowfin tuna with FADs off Okinawa Island, Japan. *Actes de colloques-ifremer*, 492–501.

Lopez, J., Moreno, G., Boyra, G. and Dagorn, L. (2016) A model based on data from echosounder buoys to estimate biomass of fish species associated with fish aggregating devices. *Fishery Bulletin* **114**, 166–178.

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

Maufroy, A., Kaplan, D.M., Bez, N., De Molina, A.D., Murua, H., Floch, L. and Chassot, E. (2017) Massive increase in the use of drifting Fish Aggregating Devices (dFADs) by tropical tuna purse seine fisheries in the Atlantic and Indian oceans. *ICES Journal of Marine Science* **74**, 215–225.

Miyake, M.P., Guillotreau, P., Sun, C.-H. and Ishimura, G. (2010) *Recent developments in the tuna industry*.

Moreno, G., Dagorn, L., Capello, M., et al. (2016) Fish aggregating devices (FADs) as scientific platforms. *Fisheries Research* **178**, 122–129.

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

Acknowledgements

This project was co-funded by Ob7 and by the ANR project BLUEMED (ANR-14-ACHN-0002). The authors sincerely thank the contribution of the staff of the Ob7 “Observatoire des Ecosystèmes Pélagiques Tropicaux exploités” for providing the echosounder buoys and observers’ data. We are also deeply grateful to the buoy manufacturers for their useful advice and information on their echosounder buoys.