MARINE BIODIVERSITY EXPLOITATION & CONSERVATION

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

Joint Tuna RFMOs Meeting (8-10 May 2019 – San Diego)

Y. Baidai, M.J. Amandé, D. Gaertner, L. Dagorn and M. Capello

Sponsors PhD Grant

Goal of the project

- To develop fisheries-independent indices of abundance for tropical tunas, using their associative behavior to floating objects

Objective of the study

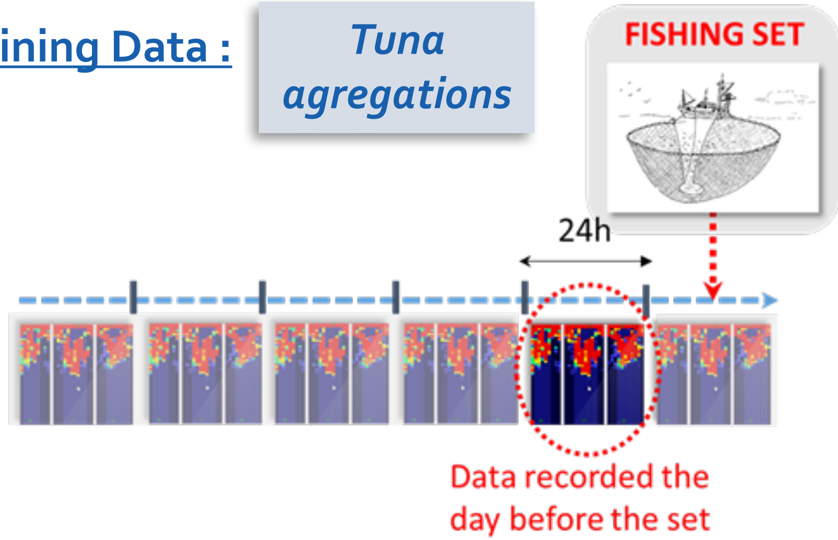
- To have reliable biomass estimations from echo-sounder buoys
  - Presence/absence of tuna
  - Size of tuna aggregations

?? Tons of tunas

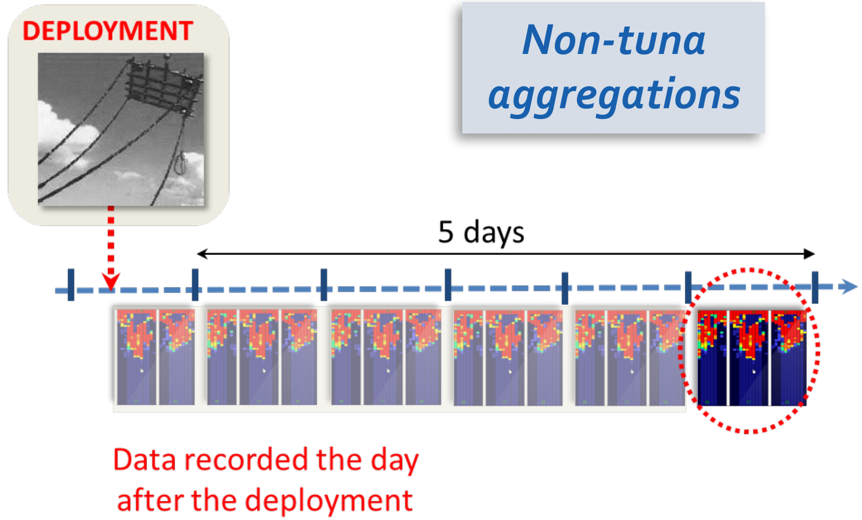
- Preliminary analyses evidenced that the biomass index computed by the internal buoy algorithm has very poor reliability (Baidai et al., 2017).
- M3i Buoy data and catch data in the AO & IO from the French fleet (2010-2017): 20 millions echosounder buoy data

Characterizing DFADs aggregation from Marine Instruments echosounder technology

Training Data : ***Tuna agregations***



Non-tuna aggregations



Acoustic samplings **recorded the day before a fishing set.**

*Set of **3381** acoustic data recorded on fished aggregations, with their corresponding catch size and composition*

Acoustic samplings recorded **5 days after a virgin FAD deployment.**

Acoustic samplings recorded **the day before DFADs visits without sets.**

*Set of **8336** acoustic data recorded under newly deployed DFADs and on DFADs visited but on which no fishing sets were performed*

Characterizing DFADs aggregation from Marine Instruments echosounder technology

Classification models :

- Supervised learning based on **random forest algorithm** considering each ocean separately (Atlantic and Indian Oceans)

- Two types of classification models per ocean :

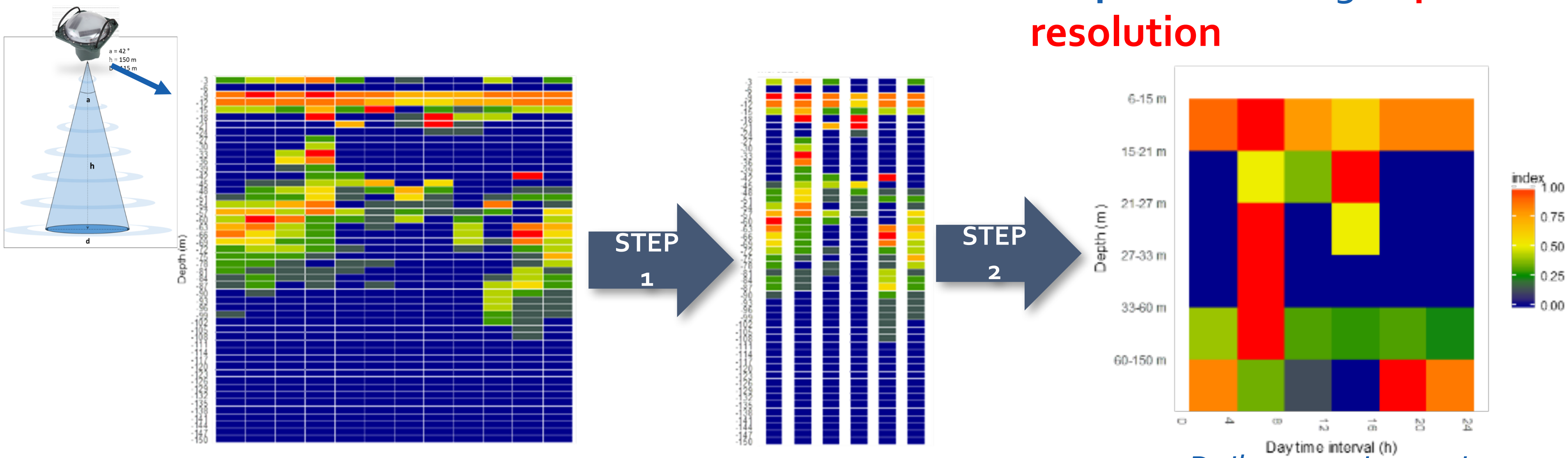
  - **Binary** : classification of presence or absence of tuna

  - **Multiclass** : classification of size classes of tuna aggregations:
    - no tuna
    - less than 10 tons
    - between 10 and 25 tons
    - more than 25 tons

Characterizing DFADs aggregation from Marine Instruments echosounder technology

Acoustic data processing: The daily acoustic matrix

- ✓ **Step 1 :** Reducing **time resolution**
- ✓ **Step 2** : Reducing **depth resolution**



Acoustic data recorded over a full sampling day

Low time resolution data

Daily acoustic matrix

Aggregation of the ***layers into groups of homogeneous layers*** *(6 groups of layers) identified from* ***clustering analysis*** *carried out separately for Atlantic and Indian ocean.*

Presence/absence of tuna

	Atlantic	Indian
Accuracy (proportion of correctly predicted)	75%	85%

- Ocean-specific classification skills :
  - In Atlantic ocean :
    - Highly effective in detecting aggregations with tuna
  - In Indian ocean :
    - Higher performance for recognition of acoustic patterns from non-tuna aggregations.

	Atlantic	Indian
Sensitivity (Ability to correctly identify positive instances, e.g. tuna presence)	82%	78%
Specificity (Ability to correctly identify negative instances, e.g. tuna absence)	67%	91%

Sizes of aggregations of tuna

	Atlantic	Indian
Accuracy (proportion of correctly predicted)	50%	45%

- ➢ Considerably less efficient than binary classifications (AO 75% - IO 85%)

Sizes of aggregations of tuna

ATLANTIC	No tuna	<10 tons	[10 , 25 tons]	> 25 tons
Sensitivity	0.66 (0.04)	0.40 (0.05)	0.28 (0.07)	0.37 (0.06)
Specificity	0.84 (0.02)	0.79 (0.02)	0.83 (0.02)	0.87 (0.03)
Precision	0.80 (0.03)	**0.33** (0.03)	**0.23** (0.05)	**0.33** (0.05)

- In the Atlantic :
  - The **highest proportion of misclassifications** are associated with tuna aggregations **between 10 and 25 tons,**
  - **Similar** recognition **performance** for tuna aggregations **below 10 tons**, and **above 25 tons.**

INDIAN	No tuna	<10 tons	[10 , 25 tons]	> 25 tons
Sensitivity	0.88 (0.03)	0.19 (0.02)	0.23 (0.03)	0.51 (0.03)
Specificity	0.78 (0.01)	0.89 (0.01)	0.83 (0.01)	0.77 (0.01)
Precision	0.57 (0.01)	**0.37** (0.04)	**0.31** (0.03)	**0.43** (0.02)

- In the Indian :
  - **Intermediate aggregation size classes** (< 10 tons, and between 10 and 25 tons) represent the **poorly classified classes.**

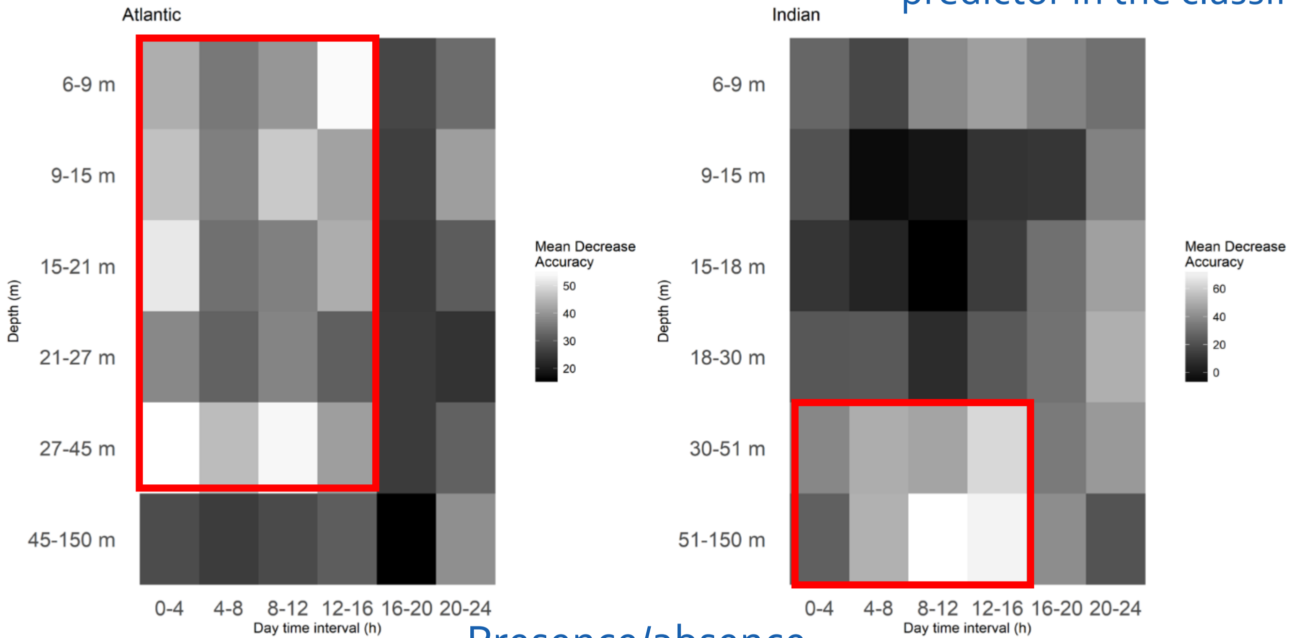
Precision: Proportion of tuna presence correctly identified among presence predictions

cro

marbec MARINE BIODIVERSITY EXPLOITATION & CONSERVATION

Predictors importance in classification

- Cells represent combinations between depth layers and time periods. Shade indicates the relevance of the predictor in the classification.

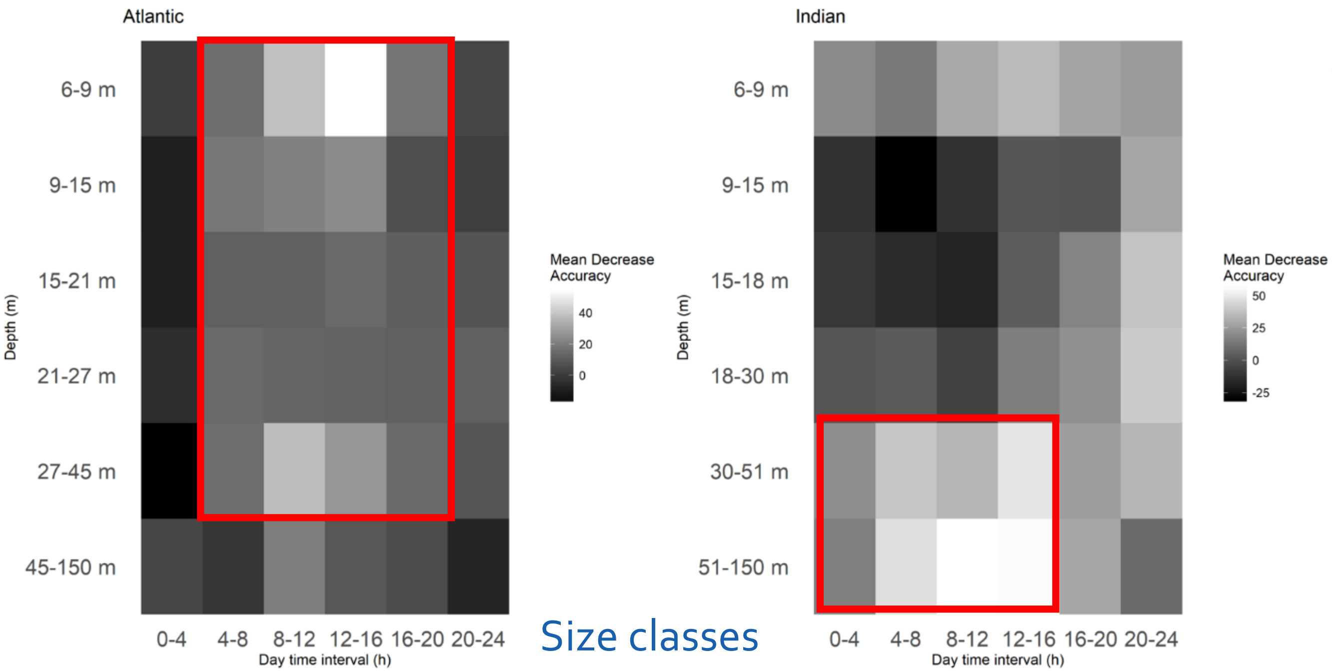


Presence/absence

Importance of depth layers and day period in presence/absence classification for the Atlantic (left) and Indian (right) oceans

Role of vertical behavior of tuna (different by ocean)

Daytime period = most relevant to characterize aggregations under DFADs



Size classes

Importance of depth layers and day period in multiclass classification for the Atlantic (left) and Indian (right) oceans

Conclusions and perspective

- ✓ **Accurate** results for the assessment of **presence/absence of tuna**
- ✓ Easily **adaptable** and **transferable** to other buoy model
- ✓ IOTC-funded study to work with M3i+ buoys (multifrequency)
- ✓ Milestones towards the use of echosounder buoy data to develop **fisheries-independent indices of abundance** for tropical tunas

Application example of presence/absence models : Monthly average of proportion of DFADs with tuna per day and 5x5° squares (2014).