Discard indices based in on-board observers data: the case of Spanish fresh trawlers targeting black hake in NW Africa

Maria SOTO, Javier REY, Ramón GARCÍA-CANCELA, María LÍEBANA and Lourdes FERNÁNDEZ-PERALTA

INSTITUTO ESPAÑOL DE OCEANOGRAFÍA-CSIC

EUROPEAN COMMISION propose quantification of Discard Per Unit Effort (DPUE) as a measure to manage the discarding of commercially fished organisms. DISCARDS of the Spanish fresh trawling fleet operating in North West Africa, targeting black hakes, *Merluccius polli* and *M. senegalensis* are variable and, in general, an important part of the total catch. ONBOARD OBSERVER DATA from commercial surveys from 2016 to 2018 of the Spanish fresh trawling fleet provide a detailed source of scientific information about catches, discards, effort and technical factors. We define DPUE for the whole discards as a unit of stock using as effort set duration. Similarity groups in discards hauls were explored through K-means clustering algorithm complemented with the predictions obtained by the Random Forest to validate the groups. We propose a model to describe an index for trends in discards for each group through GLMs with confidence intervals. The main variable determining discards hauls groups for the Spanish trawling fleet is the DEPTH, that characterizes two modes of fishing: targeting HAKE or targeting other MIXED species. The DPUE follows the same trend as the hake CPUE index.

Observer on-board collecting data during the OBSERVER PROGRAM

OBSERVER'S DATA

Data from 606 hauls along 31 surveys of one-week each were analyzed in Mauritanian and in Moroccan waters (Western Sahara). Hauls were performed at depths comprised between 93 and 815 m. Data covered the 16% of the activity of the whole fleet 2016-2018

SURVEY	GROUND	YEAR	MONTH	% discard in survey	Depth (m) min	Depth (m) max	number of hauls by depth MIXED (<300 m)	number of hauls by depth HAKE (>300m)	hauls sampled
1	MOROCCO	2016	11	52%	595	725		8	3
2		2016	11	40%	506	756		18	8
3		2016	11	45%	502	763		21	10
4		2017	12	44%	446	740		17	7
5		2017	12	47%	558	789		13	6
6		2018	6	43%	476	779		18	10
7	MAURITANIA	2016	1	32%	104	725	2	17	8
8		2016	2	35%	334	815		15	6
9		2016	3	51%	102	712	12	14	10
10		2016	4	50%	112	688	11	13	10
11		2016	6	38%	112	766	6	14	8
12		2016	6	45%	121	697	7	16	9
13		2016	10	40%	112	735	3	17	7
14		2017	2	48%	98	697	8	18	12
15		2017	2	49%	106	727	9	18	14
16		2017	4	37%	502	787		20	9
17		2017	4	35%	493	725		23	12
18		2017	6	43%	106	707	3	19	10
19		2017	6	38%	478	740		23	11
20		2017	9	33%	115	744	2	20	8
21		2017	9	31%	132	725	2	14	6
22		2017	11	25%	521	725		18	8
23		2018	1	41%	225	768		18	6
24		2018	2	34%	93	697	5	16	9
25		2018	3	29%	100	787	8	12	9
26		2018	4	38%	231	763		23	11
27		2018	5	45%	651	800		15	7
28		2018	7	25%	362	732		21	4
29		2018	8	38%	304	690		16	3
30		2018	9	19%	407	770		18	7
31		2018	10	13%	409	781		15	6
					TOTAL HAULS		78	528	254

Table 1. Survey information

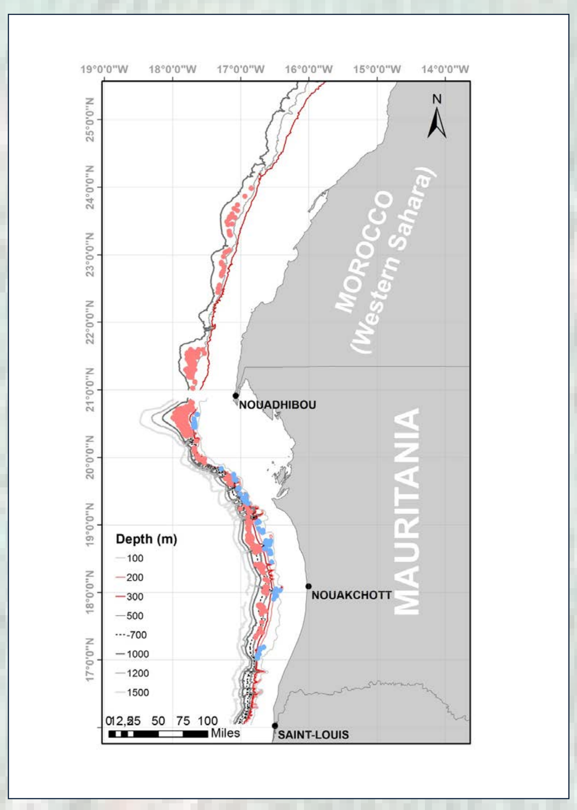


Figure 1. Hauls distribution map. Pink (>300 m), blue (<300 m)

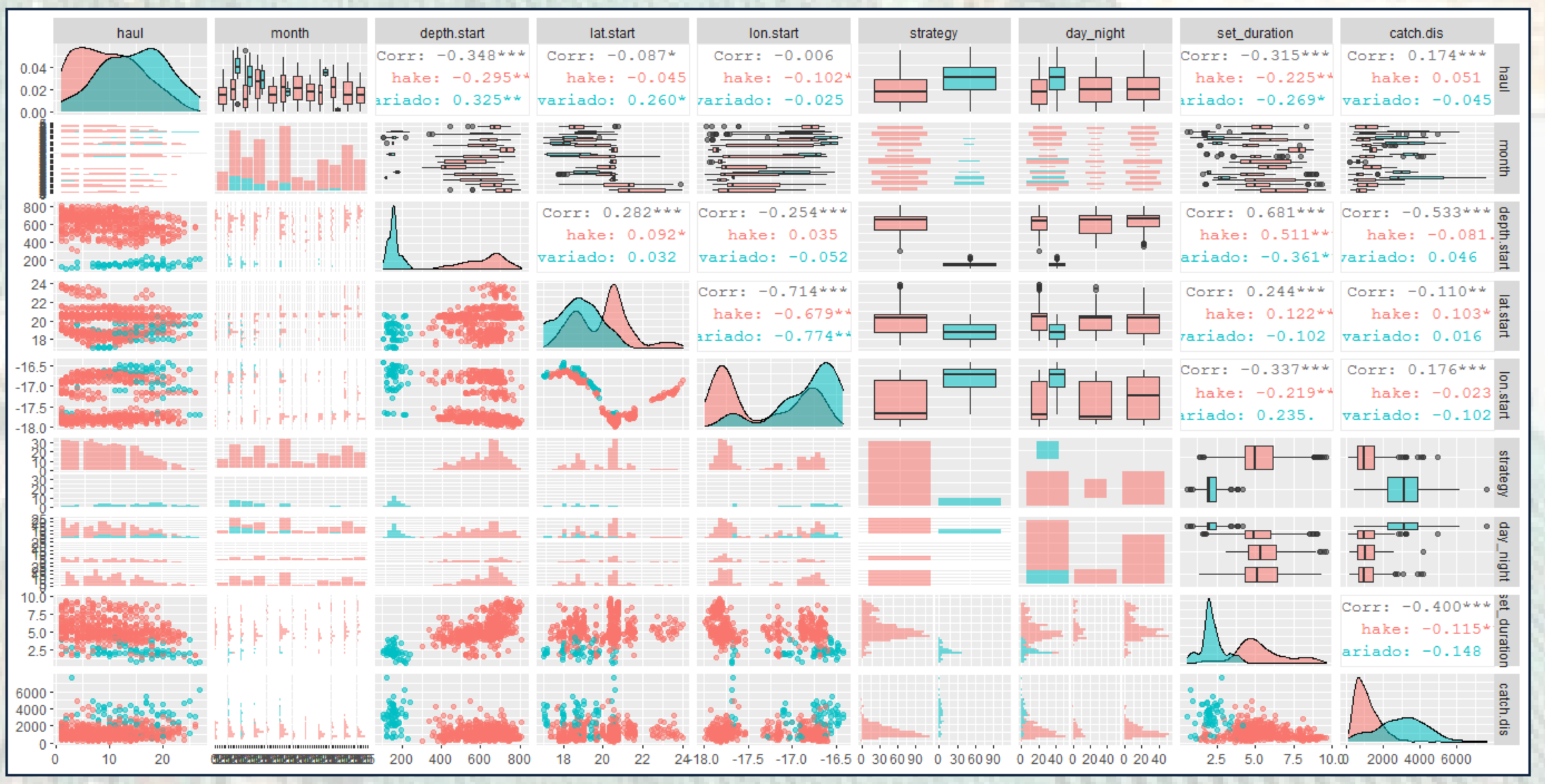


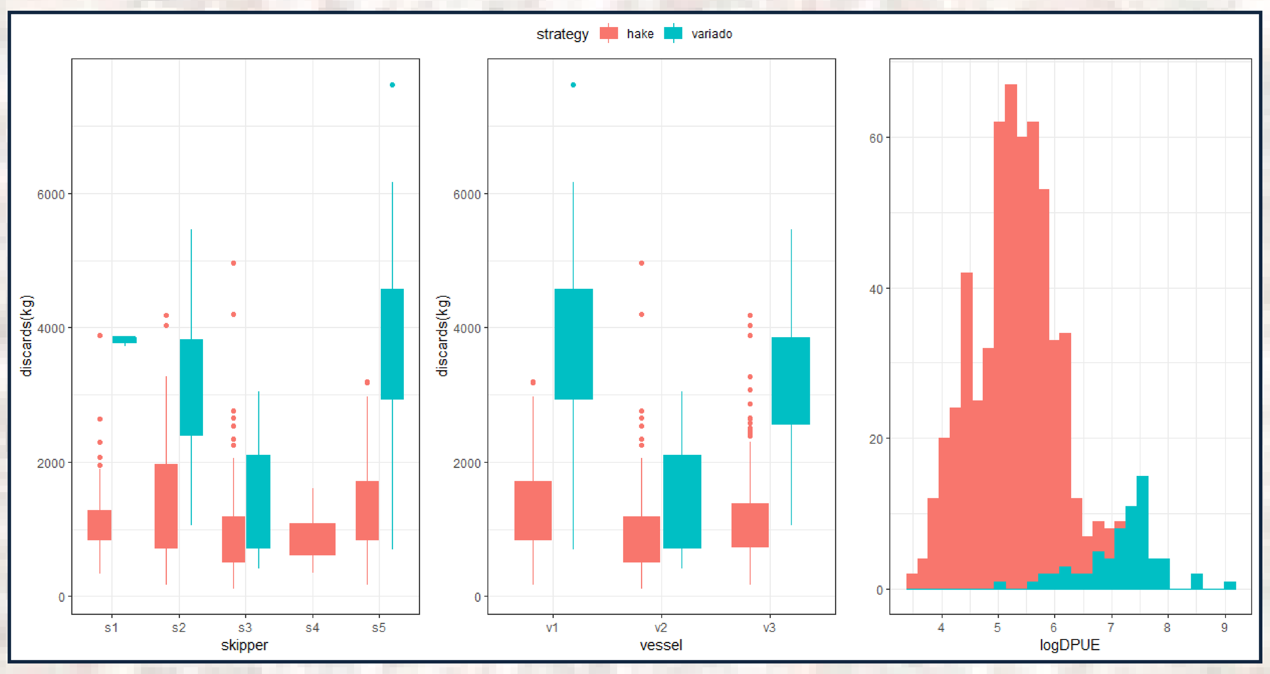
Figure 2. Observer data relationships for each haul sampled during 2016-2018. Variables considered were: order of the haul along the trip, month, year, depth, longitude and latitude at the beginning of the fishing operation, strategy of the skipper (hake or mixed), time of the day for each haul (day-night), duration of the haul and discarded catch of each haul.

MODELLING DPUE

Total discard is considered as a unit of stock and, hence, we define from haul detailed data from observers the Discard per unit of Effort:

$$DPUE_i = total\ discards_i / set\ duration_i$$

i=haul, total discard = catches of all species discarded in kg, set duration=time of trawling



SKIPPER VESSEL ESTRATEGY

From observer data we have detailed information about the skipper decisions and vessel characteristics. Information is not balanced, because not all the skippers target both hake and mixed species. This is an issue to model the standardized DPUE. All the vessels target both strategies and are comparable, so vessel is easily implementable in DPUE modeling. Nevertheless, from observer information we know that some skippers makes improvements to implement better practices to mitigate discards more than others. As the discard distribution varied depending on the strategy used, we modeled DPUE separately for HAKE and MIXED hauls.

Figure 6. Boxplots of discard by skipper and vessel and histograms of the log transformed DPUE for each fishing strategy

MODELING DPUE INDEX: We separate discards by strategy (hake or mixed) and estimate an index of DPUE for each through glm. Covariates year, month, depth.start, haul and vessel were initially used to model DPUE in lognormal scale with Generalised Linear Models (GLMs) (McCullagh and Nelder, 1989) in R. Variable selection for each index was done based on deviance explained and residual diagnostics.

HAKE

```
            Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                          546     248.60
year         2   56.870       544     191.73 < 2.2e-16 ***
month       11   20.410       533     171.32 6.063e-10 ***
depth.start  1    7.609       532     163.71 6.604e-07 ***
```

MIXED

```
            Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                           66     32.738
year         2  13.4441        64     19.294 3.967e-11 ***
depth.start  1   1.6122        63     17.682   0.01654 *
```

Table 2. Selected variables for each strategy model and partial residual plots

The index was calculated as the weighted average of the year LSMeans (Lenth, R. 2016) obtained for hake and mixed:

$$I_y = p_{y,hake} \cdot LSM_{y,hake} + p_{y,mixed} \cdot LSM_{y,mixed}$$

where p_y are the proportions of hauls targeting hake or mixed each year *y*=2016, 2017, 2018.

Standard error for the index was approximated by the Delta method:

$$SE_y = \sqrt{p_{y,hake}{}^2 \cdot SE_{y,mixed}^2 + p_{y,mixed}{}^2 \cdot SE_{y,hake}^2}$$

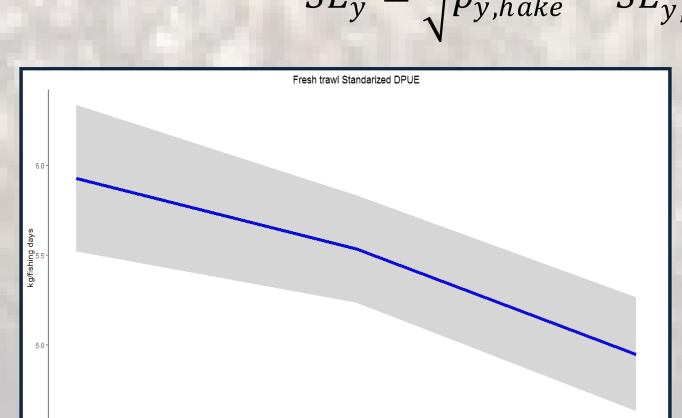


Figure 7. Standarised DPUE index for the Spanish trawling fleet in NW Africa with confidence bounds defined by DPUE∓1.96·SE. Effort is measured in set duration hours

The final DPUE standardised index presented a clear decreasing trend in DPUE abundance in the short period considered. Despite the effort units aredifferent for Black Hake CPUE and DPUE, the trend in 2016-2018 is very similar and more research is needed to determine the causes of this decline: discards declining, better practices implemented or discards driven by hake abundance

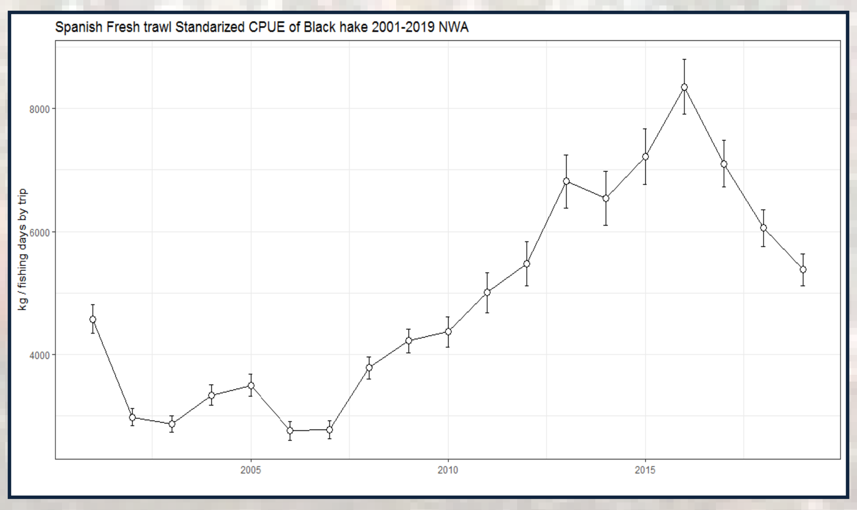
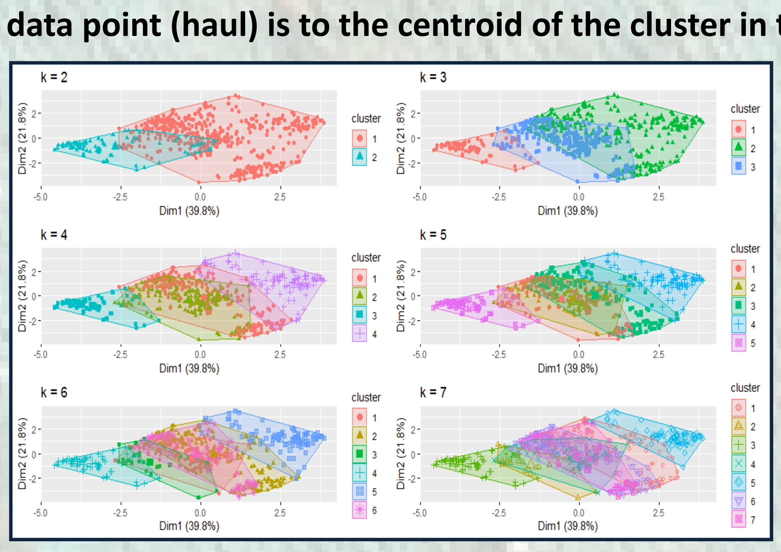


Figure 8. Standarized CPUE of Black Hake for the Spanish fresh trawling fleet in NW Africa. Effort is measured by fishing days by trip

EXPLORING OBSERVER'S DATA GROUPS

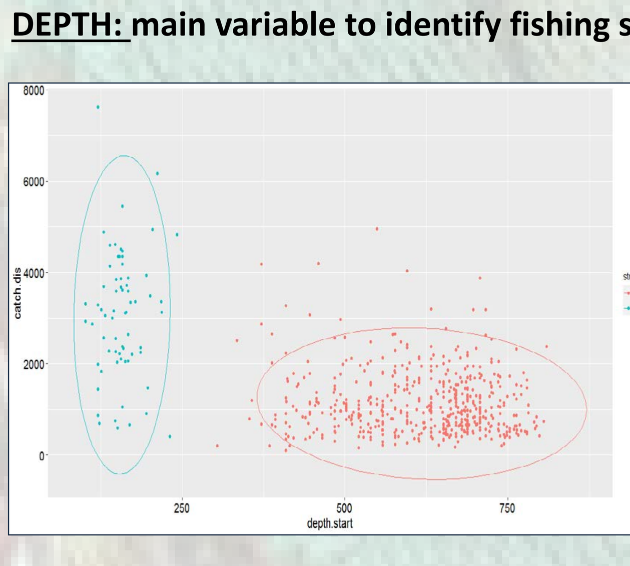
K-MEANS clustering algorithm (Giordani, 2020): the notion of similarity is derived by how close a data point (haul) is to the centroid of the cluster in the Euclidean distance measure.



Clustering is a subjective task and we already know the number of clusters expected, which is k=2 corresponding to the fishing mode: targeting hake and targeting mixed.

Figure 3. Results of k-means clustering algorithm for grouping observers data. Optimal number of clusters vary between 2-4. If we assume k=2, results are coherent with the discriminant variable depth obtained through random forest.

DEPTH: main variable to identify fishing strategies and groups of discards.



Hauls in shallower waters (<300m) generate higher and more dispersed amount of discards than deeper waters (>300m). This distinction is coherent with the definition of two métiers for each strategy.

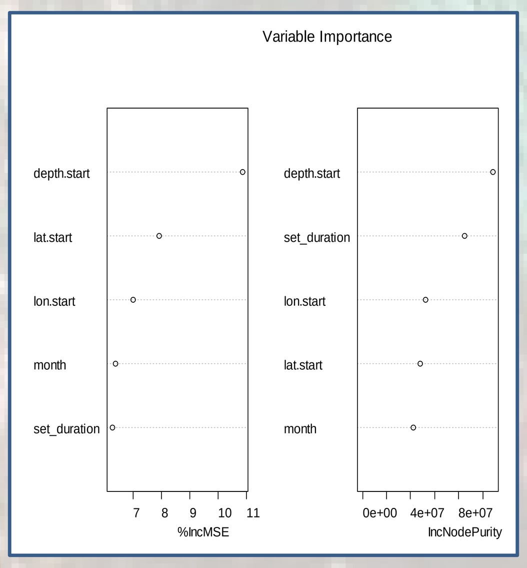
MIXED depth < 300m

HAKE depth > 300m

Figure 4. Relationship between the depth of each haul and the amount of discards in kg obtained by each fishing mode.

RANDOM FOREST have been used to rank the importance of variables in a regression classification problem: (Genuer, R. 2020)

catch ~ depth + lat + lon + day_night +haul + month + set_duration



The variables analyzed to discriminate were explored to create a balanced subset of training and test sets. To measure the rank of variable importance in the classification we use: the IncNodePurity is the total decrease in node impurities, measured by the Gini Index from splitting on the variable, averaged over all trees and the %IncMSE, which is the increase in MSE of predictions(estimated with out-of-bag-CV) as a result of variable j being permuted(values randomly shuffled). Both criteria showed that DEPTH is the best variable to classify discards hauls.

Figure 5. Results of variable importance rank in the random forest

CONCLUSSIONS

1. Unlike hake catches, discards were higher and more dispersed in shallower waters. 2. We identified two separate métiers determined by depth. 3. We treated total discards as a stock unit susceptible of being monitored, managed and assessed. 4. Vessel characteristics and strategy of the skipper are important effects on discards, but more contrast is needed with more trips in all vessels and several skkipers. 5. This study shows the importance of observer data and identifies recommendations for the improvement in the scientific usefulness of logbook information.

REFERENCES

- McCullagh P and JA Nelder, 1989. Generalized Linear Models, 2nd edn. London: Chapman & Hall.
- Giordani P, Ferraro MB and F Martella, 2020. An introduction to clustering in R. Springer.
- Genuer, R and Poggi, JM, 2020. Random forest in R. Springer
- Lenth, R (2016). Least-Squares Means: The R package lsmeans. Journal of statistical software, 69(1), pp. 1-33.