

Dear referee 2,

We thank you sincerely for your comments which helped us to improve the quality of the paper.

Firstly, we would like to draw your attention on some major changes we proposed to strengthen the evaluation of the decision tree classifier and to improve its recall scores. To strengthen the performance evaluation, the testing period was extended from the first four months of 2021 (i.e., from January 2021 to April 2021) to the full year of 2021 including seasonal variations of the offshore Sargassum abundance. To improve the recall score of the classifier, the module A producing the monthly probability of beaching was replaced by a new module based on satellite observations which produces the weekly probability to reach the maximum observed cumulative floating algae density in an area of 100 km radius offshore Guadeloupe. The performance evaluation of the classifier was also extended by adding three temporal uncertainty ranges around the decision day, respectively: ± 1 days, ± 2 days, ± 3 days. While the classifier may reproduce 61.5% of the observed beachings in 2021 with an accuracy lower than one day (this value reached 41.7% with the old module A and the limited testing period of four months), this recall score reaches 74.4% at ± 3 days accuracy.

Please find below our answers to your remarks (in bold). The proposed changes in the text are marked in red.

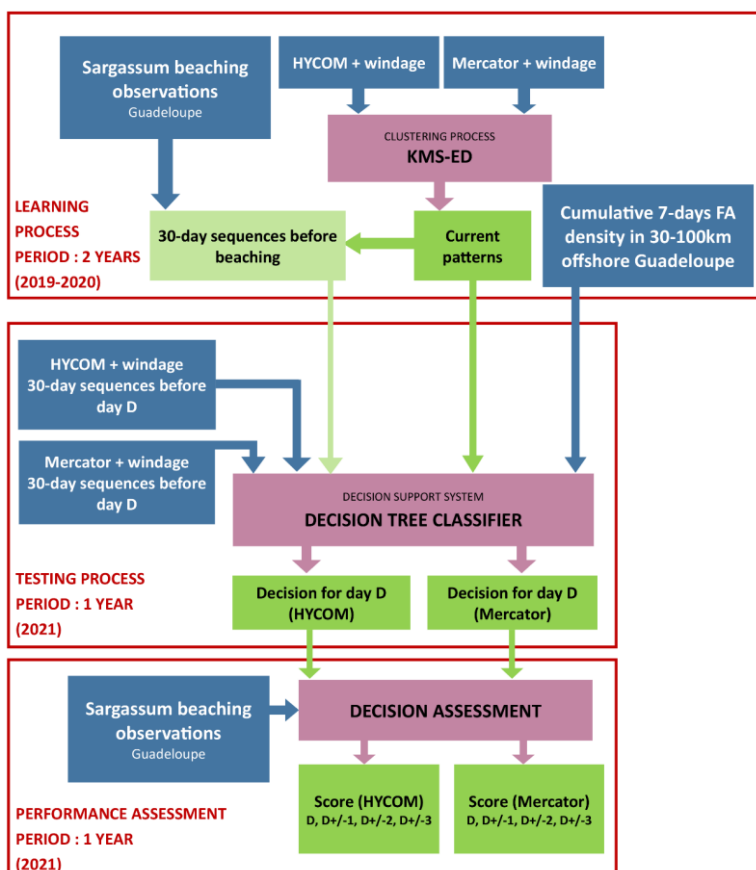
1. General comments

The authors present a very interesting framework and method to better understand the ocean dynamics behind the strandings of Sargassum in the Lesser Antilles and to estimate their occurrence. The methodology presented is quite complex as well.

A better explanation of the methodology is necessary, especially for the oceanographic audience of this journal to adequately follow and understand this interesting study.

Section 2 I believe can be improved by making it easier for the reader to follow, especially the non-experts in these clustering methods. The technical details necessary for the reader to follow the study should be clearly described and the other details can be added as a section in supplementary material.

Our answer: In the revised manuscript, we will try to clarify the text with language improvements and to better explain the overall approach. The following schematic will be added in the methods section to summarize the overall methodology.

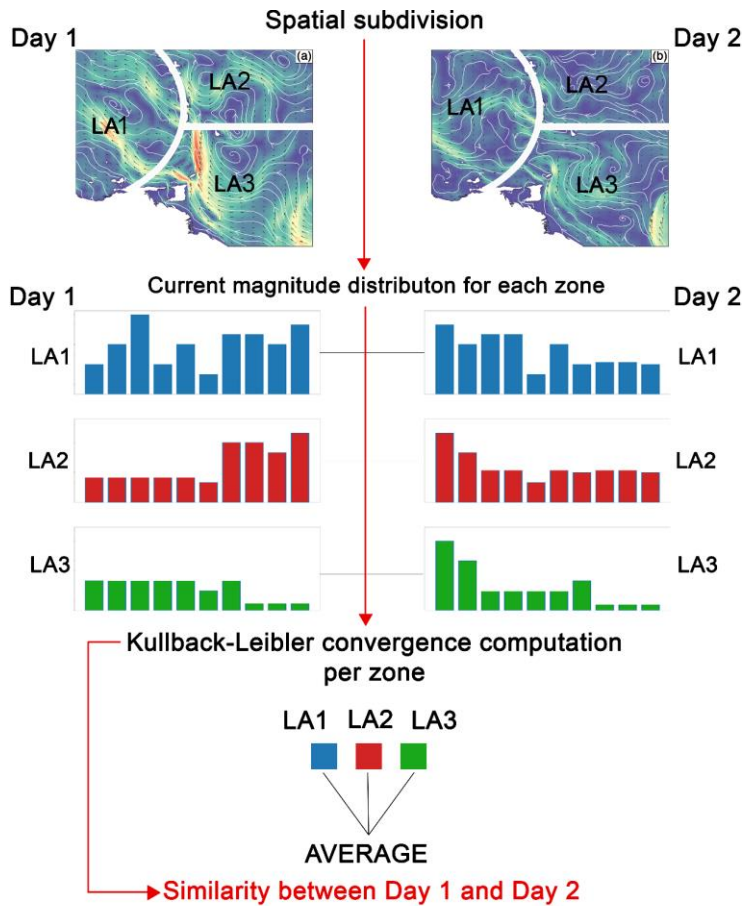


“Figure XX: The schematic of the adopted methodology.”

A schematic of the method is given in fig. 2 for Section 2.7, but maybe a schematic for sections 2.5 and 2.6 could help too.

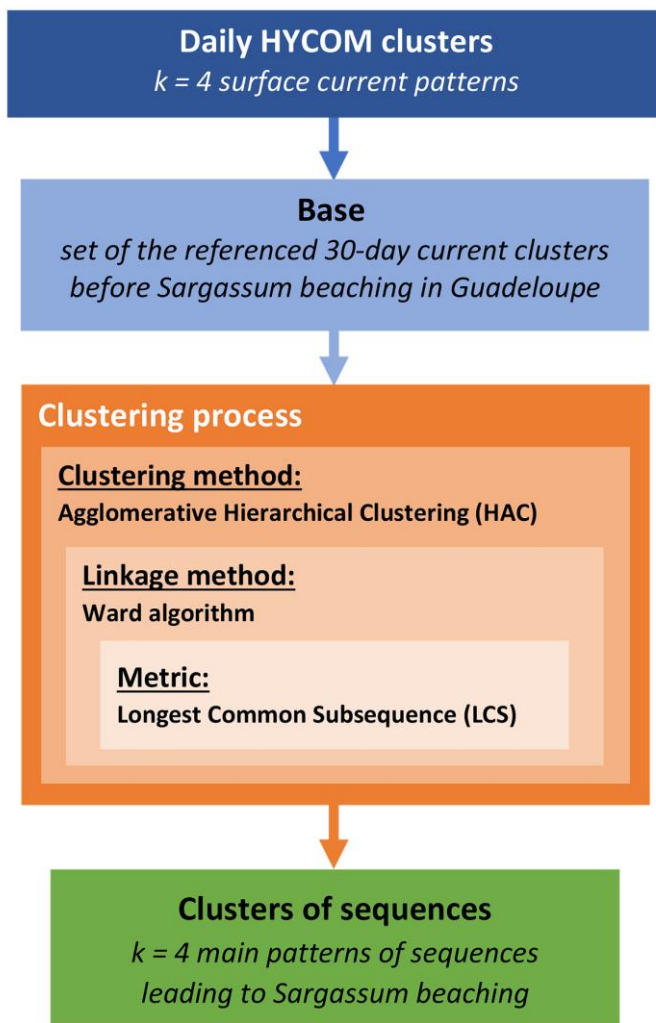
Our answer: Following your suggestion, the two schematics, below, will be added in sections 2.5 and 2.6, respectively.

In Section 2.5:



“Figure XX: The schematic of the Expert Distance process.”

In Section 2.6:



“Figure XX: The schematic of the clustering process on the current sequences leading to Sargassum beachings.”

In the discussion, I found that some comment on the impact (if any) of considering processes other than windage (e.g. presence of nutrients, sinking of Sargassum, waves?) could have on an even better understanding of the Sargassum strandings, was missing.

Our answer:

Following your suggestion we will add this sentence at the end of the discussion (L365, end of section 4.3):

“The present study does not take into account the effects of other factors (e.g., presence of nutrient, sinking of Sargassum and waves) which would allow a more realistic understanding of the Sargassum strandings.”

2. Specific comments

L23: “Strandings were also be observed in Africa (Széchy et al., 2012).” Why mention the occurrence of strandings in Africa? Any connection with the Caribbean strandings? Did the Sargassum strandings also cause natural hazards on the African coast?

Our answer:

This sentence will be removed.

L61: “MODIS AFAI satellite images”, please define/describe

Our answer:

“A combination of MODIS AFAI satellite images” will be replaced by:

“A combination of satellite-based Alternative Floating Algae Index (AFAI, Wang and Hu, 2016) fields”

L66-68: Could be useful to include some references of the methodology here.

Our answer:

The sentence *“None of them used predictive modelling, including classifiers, to determine the probability of a set of data belonging to another set in order to discover repeatable patterns, allowing to produce a decision for risk prevention managers.”* Will be modified as below:

“None of them used predictive modelling (Geisser, 1993; Kuhn and Johnson, 2013) including classifiers (Friedl and Brodley, 1997), to determine the probability of a set of data belonging to another set in order to discover repeatable patterns, allowing to produce a decision for risk prevention managers.”

These references will be included to improve the understanding.

Friedl, M. A., and Brodley, C. E.: Decision tree classification of land cover from remotely sensed data, Remote Sensing of Environment, vol. 61, Issue 3, pp. 399-409, ISSN:0034-4257, [https://doi.org/10.1016/S0034-4257\(97\)00049-7](https://doi.org/10.1016/S0034-4257(97)00049-7), 1997.

Geisser, S.: Predictive Inference: An Introduction, Chapman & Hall, ISBN: 978-0-412-03471-8, 1993.

Kuhn, M., and Johnson, K.: Applied predictive modeling. New York, Springer. isbn:978-1-4614-6848-6, doi:10.1007/978-1-4614-6849-3, 2013.

L69: A general definition of predictive modelling is missing in the introduction for the readers which do not know about this method and how it compares with a conventional forecast. For example could be included here (Line 69).

Our answer:

The following sentences will be added here L69:

“Predictive modelling refers to mathematical and computational methods for predicting future events on the analysis of the statistical patterns in the input dataset (Geisser, 1993; Friedl and Brodley, 1997; Kuhn and Johnson, 2013). Compared to other conventional forecast, predictive modelling methods requiring low computational costs are characterized by their flexibility, and their intuitive simplicity (Friedl and Brodley, 1997).”

L75-76: “To optimize the final partitioning, an additional metric based on the Kullback Leiber divergence (Kulback and Leibler, 1951, Biabiany et al., 2020) will be included” : quite specific on the methodology, for readers not familiarised with this method it could be hard to follow in this point in the introduction. More general details can be given, or this point can be moved to the methods section.

Our answer:

Following your remark, this sentence will be removed. The Expert Distance process with the Kullback Leiber divergence will be explained in the method section (2.5).

L82-83: “This ocean region corresponds to the CA and TA1 boxes in Johns (2020)”, maybe say approximately corresponds, as not exactly the same. The LA3 region goes further south and LA2 and LA3 go until -55°E, whales region TA1 till -50°E. Most importantly, why choose the study regions to correspond to CA and TA1 boxes from Johns (2020)?

Our answer:

The study regions were not choose to correspond to CA and TA1 boxes from Johns (2020). The sentence “*This ocean region corresponds to the CA and TA1 boxes in Johns (2020).*” will be removed.

L96-96: From what I understand this dataset was not used before to simulate Sargassum trajectories, but was it used in any other Lagrangian study? Any validation studies done on the velocity outputs of this dataset?

Our answer:

Firstly, there was a mistake about the HYCOM output resolution, we will correct it in the revised manuscript. The right version of HYCOM output used here is the HYCOM GLBy0.08 which has a grid resolution of 0.08 degree in longitude and 0.04 degree in latitude. To perform the present study, the native HYCOM fields have been preliminarily interpolated on the Mercator uniform lon/lat 0.08-degree grid with a bilinear method.

The sentence L95-96: “*These fine resolution current data were not used in previous studies dealing with Sargassum hazard (Putman et al., 2018; Johns et al., 2020).*” will be replaced by :

“Putman et al. (2018) and Johns et al. (2020) used a previous version of HYCOM model including uniform lon/lat 0.08° scale grid to successfully simulate Sargassum trajectories.”

Line 91-100 the two paragraphs,

2.1 HYCOM surface current dataset

“Fine scale surface current data from the 1/25-degree HYCOM + NCODA Gulf of Mexico analysis model (GOMu0.04/expt_90.1m000 version, Hogan et al, 2014; Helber et al., 2013; Cummings and Smedstad, 2013; Cummings, 2005) between 1st January 2019 (i.e., available data starting date) and 31 December 2020 were analyzed. Daily 12Z fields giving the u and v components of the current at 50 cm depth were used. These fine resolution current data were not used in previous studies dealing with Sargassum hazard (Putman et al., 2018; Johns et al., 2020).

2.2 Mercator surface current dataset

The daily 50-cm depth current components from the PSY4V3R1 Mercator 1/12-degree 3D analysis system including the version 3.1 of the NEMO ocean model (Lellouche et al., 2018; Gasparin et al., 2019) were also analyzed along the same period as HYCOM..”

will be replaced by:

“2.1 HYCOM surface current dataset

Daily 12Z surface current components from the 41-layer Hybrid Coordinate Ocean Model (HYCOM) global 1/12-degree analysis (HYCOM GLBy0.08 version), were examined. The HYCOM surface forcing including 10-m wind velocities are extracted from Climate Forecast System Version 2 (CFSv2). The Navy Coupled Ocean Data Assimilation (NCODA) system is used to assimilate available observational data: satellite altimeter sea surface height, satellite and in-situ sea surface temperature, temperature vertical profiles and salinity vertical profiles (Cummings, 2005; Cummings and Smedstad, 2013; Helber et al., 2013). The Bathymetry used is the GEBCO8 (Becker et al., 2009) with 30 arc second of resolution. The HYCOM GLBy0.08 grid resolution is 0.08 degree in longitude and 0.04 degree in latitude. To perform the present study, the native HYCOM fields have been preliminarily interpolated on the Mercator uniform lon/lat 0.08-degree grid with a bilinear method. Putman et al. (2018) and Johns et al. (2020) used a previous version of HYCOM model including uniform lon/lat 0.08° scale grid to successfully simulate Sargassum trajectories.”

2.2 Mercator surface current dataset

The daily 12Z surface current components from the 50-layer PSY4V3R1 Mercator 1/12-degree 3D analysis system (Lellouche et al., 2018; Gasparin et al., 2019) were also analyzed. The atmospheric surface forcing are extracted from the 3-hourly ECMWF (European Centre for Medium-Range Weather Forecasts) IFS (Integrated Forecast System). Assimilated observational data types are quite similar to HYCOM model. Unlike the HYCOM GLBy0.08 native grid including higher resolution in latitude (i.e., 0.04 degree), the Mercator native grid is uniform in longitude and latitude with 0.08-degree scale. This would suggest that HYCOM may better reproduce small scale patterns than Mercator. Moreover as described by Lellouche et al. (2018), the Mercator bathymetry includes GEBCO8 data in regions shallower than 200 m and the coarse 1 arc-minute ETOPO1

data (Amante and Eakins, 2009) in regions deeper than 300 m. The complex bathymetry of the Lesser Antilles Arc studied here could be less realistic in Mercator than in HYCOM.”

These additional references will be added to the Reference Section:

Amante, C. and Eakins, B. W.: ETOPO1 1 Arc-minute global relief model: procedures, data sources and analysis, NOAA Technical Memorandum NESDIS NGDC-24, Marine Geology and Geophysics Division, Boulder, Colorado, 25 pp., <https://doi.org/10.1594/PANGAEA.769615>, 2009.

Becker, J. J., Sandwell, D. T., Smith, W. H. F., Braud, J., Binder, B., Depner, J., Fabre, D., Factor, J., Ingalls, S., Kim, S. H., Ladner, R., Marks, K., Nelson, S., Pharaoh, A., Trimmer, R., Von Rosenberg, J., Wallace, G., and Weatherall, P.: Global Bathymetry and Elevation Data at 30 Arc Seconds Resolution: SRTM30_PLUS, Mar. Geod., 32, 355–371, 2009.

HYCOM GLBy0.08 version, <https://www.hycom.org/dataserver/gofs-3pt1/analysis>, last access: 17 January 2022.

Cummings, J. A.: Operational multivariate ocean data assimilation. Quart. J. Royal Met. Soc., Part C, 131(613), 3583–3604, <https://doi.org/10.1256/qj.05.105>, 2005.

Cummings, J. A., Smedstad O. M.: Variational Data Assimilation for the Global Ocean. In: Park S., Xu L. (eds) Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. II), Springer, Berlin, Heidelberg, https://doi.org/10.1007/978-3-642-35088-7_13, 2013.

Helber, R. W., Townsend, T. L., Barron, C. N., Dastugue, J. M., Carnes, M. R.: Validation Test Report for the Improved Synthetic Ocean Profile (ISOP) System, Part I: Synthetic Profile Methods and Algorithm, Report, Naval Research Laboratory, Mississippi, 127 pp, <https://www7320.nrlssc.navy.mil/pubs/2013/helber1-2013.pdf>, 2013.

L101: “Comparison between HYCOM and Mercator results” Do you mean the results from the Sargassum trajectories or a comparison of the velocity outputs of these datasets?

Our answer: Because of the previous changes in HYCOM and Mercator outputs description, the sentence *“Comparison between HYCOM and Mercator results would help to better understand the effects of spatial resolution on surface current patterns in the focused region”* will be removed in the revised manuscript.

Section 2.3: What is the spatial and temporal resolution of the ERA-5 wind dataset?

Our answer:

The ERA-5 wind dataset has a spatial resolution of 31 km and hourly fields are available.

Line 108, the part *“Surface wind data (at 1000 hPa) from the ERA-5 model for the time period 2019 to 2020 were integrated with Mercator currents following this formula:”*

will be replaced by

“The daily 12h (UTC) surface wind data (at 1000 hPa) from the 31-km scale ERA-5 model were integrated with Mercator and HYCOM currents following this formula:”

L128: “Ward's method for HAC” Please explain and add reference.

Our answer:

The sentence *“Besides the measures and the classes of distance between objects such as the Euclidean distance for K-means and the Ward's method for HAC, a new metric was also added (Biabiany et al. 2020)”*

Will be modified as below

“Besides the measures and the classes of distance between objects such as the Euclidean distance for K-means and the Ward's method which allows to identify homogeneous subsets of data (Ward, 1963), a new metric was also added (Biabiany et al. 2020)”

This reference will be added

J. H. Ward Jr. (1963) Hierarchical Grouping to Optimize an Objective Function, Journal of the American Statistical Association, 58:301, 236-244, DOI:10.1080/01621459.1963.10500845.

L129-130: “with its own expertise on the input data” What do you mean by these? Please provide further explanations. Also, the new method name is not specified at all in section 2.5.1, and it will help for the reader to better follow the methodology. This section is only 5 lines long, more details on the process of the clustering methods could be given.

Our answer:

The sentence *“The result is an automated analysis with its own expertise on the input data.”* Will be removed.

Moreover to help the reader better following the methodology the section “2.5.1 Clustering methods process” will be removed.

The paragraph (L126-129) *“Unsupervised learning methods such as Hierarchical Agglomerative Clustering (HAC) and K-means algorithms are used in the present study. Besides the measures and the classes of distance between objects such as the Euclidean distance for K-means and the Ward's method for HAC, a new metric was also added (Biabiany et al. 2020). This metric integrates a set of knowledge about the dynamics of the data to be partitioned as well as their spatio-temporal properties.”*

will be moved L125 after the sentence *“This method allowed significant improvement in clustering analysis dealing with climate data characterized by high spatio-temporal variability, such as precipitation (Biabiany et al., 2020).”*

L132: “L2 clustering methods...” Please explain L2 in this context.

The sentences: *“The ED metric, which seems more suitable for this study, was used. L2 clustering methods can lead, within the same cluster, to gatherings of different physical situations (Biabiany et al., 2020).”*

To remove these biases linked with L2 clustering, the first step of the method used here is to consider the spatial variability in the dynamics of the analyzed daily surface currents from L2.”

will be replaced by

“The ED metric, which seems more suitable for this study, was used. Clustering methods using euclidean distance (L2) can lead to group different physical situations within the same cluster (Biabiany et al., 2020).”

L133: “gatherings of different physical situations”. What do you mean by this? Maybe give an example of physical situations for this particular study scenario.

Our answer:

We will clarify this part in the revised manuscript.

The part *“can lead, within the same cluster, to gatherings of different physical situations”* will be replaced by *“can lead to group different physical situations within the same cluster.”*

You refer to this in the next phrase as “biases”. Is there then a tendency towards a specific physical situation?

Our answer:

There is no trend towards a specific physical situation. These side effects have been described in Biabiany et al. (2020).

As stated above, the sentence (L133) *~~“To remove these biases linked with L2 clustering, the first step of the method used here is to consider the spatial variability in the dynamics of the analyzed daily surface currents from L2.”~~* will be removed in the revised manuscript.

L134: “spatial variability” : At what scales?

Our answer: As stated above, the sentence (L133) *~~“To remove these biases linked with L2 clustering, the first step of the method used here is to consider the spatial variability in the dynamics of the analyzed daily surface currents from L2.”~~* will be removed in the revised manuscript.

L139-L140: “The analyzed daily fields include a total of 14 279 meshes (4 282 meshes in LA1, 3 407 meshes in LA2 and 4 536 meshes in LA3). The remainder corresponds to land areas.” What do you refer to here with meshes?

Our answer:

L139-140: “meshes” will be replaced by *“grid points”*

The land areas then correspond to Sargassum strandings? For clarity, these details could be described in a dataset section better, rather than in the middle of the methods description.

Our answer:

The land areas do not correspond to Sargassum strandings, they correspond to areas over land (e.g., islands).

To clarify this point, the sentence *“The remainder corresponds to land areas”* will be replaced by *“The remainder corresponds to areas over land (e.g., islands).”*

L141-142: “The second step was to group the information carried by the daily current velocity fields conditionally to the three given zones into histograms.” More details on histograms, for example binning, velocity data from HYCOM and Mercator?

Our answer:

Histogram bins are given in Table 1 and Fig. 4. The values correspond to deciles from HYCOM and Mercator datasets.

The sentence *“The second step was to group the information carried by the daily current velocity fields conditionally to the three given zones into histograms.”* Will be modified as follows

“The second step was to group the information carried by the daily current velocity fields conditionally to the three given zones into histograms (Table 1, Fig.4).”

L158: “optimal matching methods” Please explain and add some references.

Our answer:

The words “optimal matching methods” will be removed.

The sentence at L158 *“Dissimilarities between these backward sequences were calculated with optimal matching methods before dividing the population into several groups using a hierarchical classification (Larmarange et al., 2015).”*

will be replaced by:

“Dissimilarities between these backward sequences were calculated before dividing the sequences dataset into several groups using a hierarchical classification (Larmarange et al., 2015).”

L158: “dividing the population” what do you refer to exactly here by population? Population of strandings or backward sequences?

Our answer :

To clarify this point, the words “dividing the population” will be replaced by “dividing the sequences dataset”

L160-162: Please give further details (maybe as supplementary material?) and add more references.

Our answer :

“Wald’s algorithm” will be corrected to “Ward algorithm” already described (Line 128) as a “method which allows to identify homogeneous subsets of data (Ward, 1963)”.

Ward, J. H. Jr.: Hierarchical Grouping to Optimize an Objective Function, Journal of the American Statistical Association, 58:301, 236-244, DOI: 10.1080/01621459.1963.10500845, 1963.

L186-L187: “was experimented on the first 120 days...”. Was experimented to...? Recall aim of doing these tests. Also why 120 days and during this period of time? Could results vary a lot if done during the northern hemisphere Summer months instead?

Our answer:

To strengthen the performance evaluation, the testing period was extended from the first four months of 2021 (i.e., from January 2021 to April 2021) to the full year of 2021 including seasonal variations of the offshore Sargassum abundance.

The sentence “The proposed tree in Fig. 2 was experimented on the first 120 days of the year 2021, from 1st January 2021 to 30 April 2021, i.e., 120 tests.” Will be replaced by

“The proposed tree in Fig. 2 was tested on the full year of 2021 (from 1st January to 31 December), i.e., 360 tests.”

The original Fig. 2 will be replaced by the following figure:

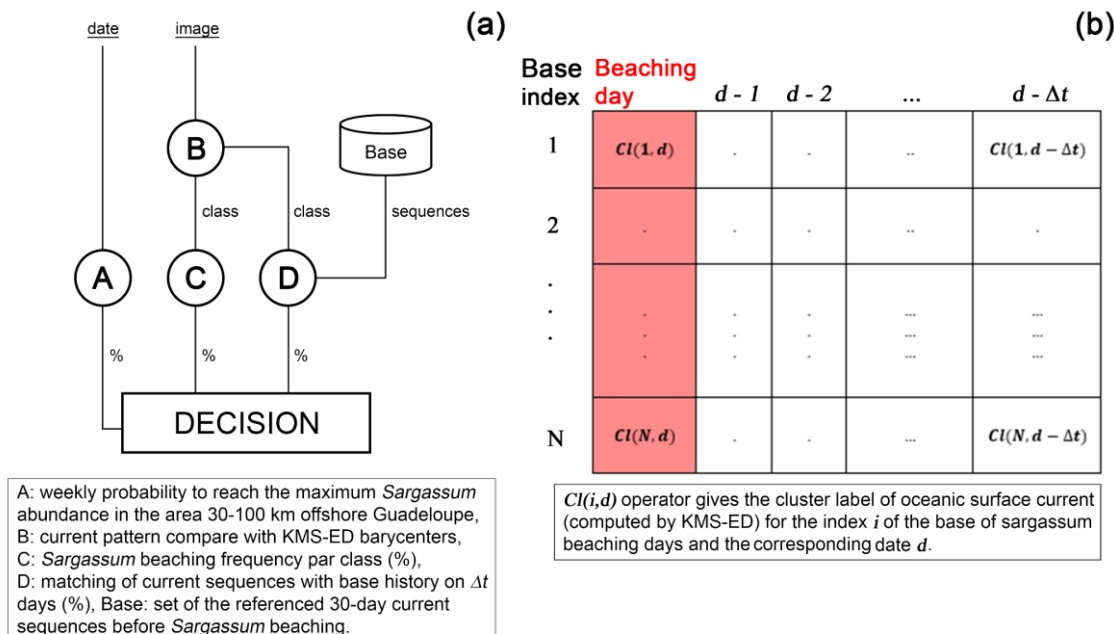


Figure 2: (a) Scheme of the decision tree classifier to predict Sargassum stranding probability. (b) Combination base of oceanic currents clusters labels obtained by KMS-ED from each stranding day to Δt days before.

The Decision support system results figure will be replaced by the following figure:

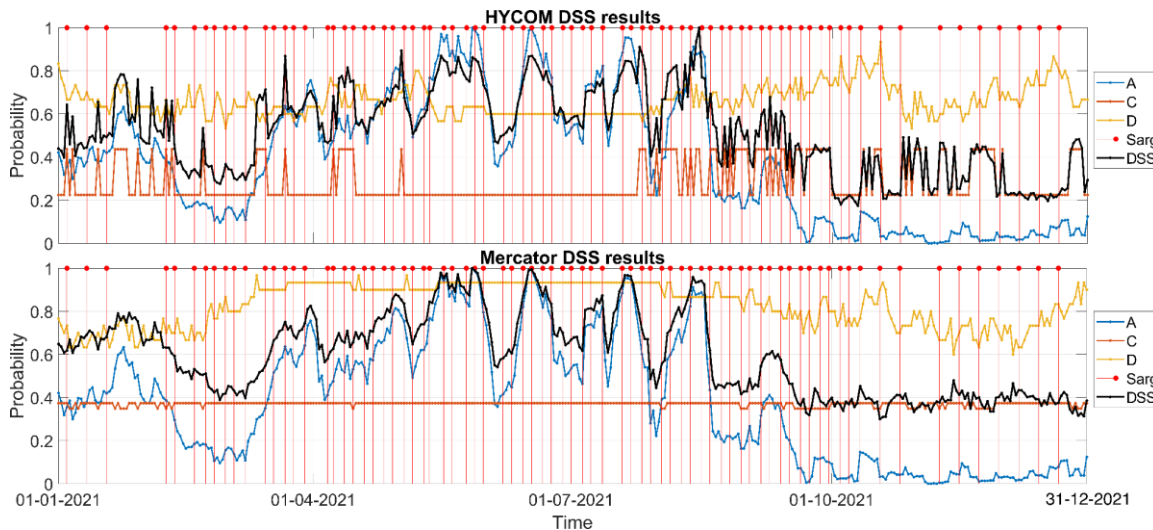


Figure 15: Decision Support System (DSS) results: probability of beaching obtained per module. Weekly probability to reach the maximum Sargassum abundance in the area 30-100 km offshore Guadeloupe for module A (blue line), stranding frequency per cluster for module C (orange line), match percentage for module D (yellow line), DSS Decision (black line). Day of observed beaching on Guadeloupe coasts (red dots): HYCOM (a) and Mercator (b).

L190: Can maybe start section 3.1 giving some context on why this analysis is done.

Our answer:

The following sentence will be added L190 at the beginning of the section 3.1:

“In view of the lack of study dealing with surface current patterns in the Lesser Antilles area, this preliminary analysis is presented here.”

L191: “90% of them remain below 0.65 m/s”. For both models exactly same?

Our answer :

The sentence “For both models HYCOM and Mercator, the velocity intensities do not exceed 2.57 m s⁻¹ and 90% of them remain below 0.65 m s⁻¹”

will be replaced by

“For both models HYCOM and Mercator, the velocity intensities do not exceed 2.57 m s⁻¹ and 90% of them remain below 0.65 m s⁻¹ (the respective 90th centile values are respectively 0.6515 m.s⁻¹ and 0.6458 m.s⁻¹ for HYCOM and Mercator).”

L193: Figure 3 distributions how are they calculated? With histograms? Kernel Density Estimator or something else applied to obtain this “smooth” distribution curves?

Our answer :

After the sentence (L193): “Figure 3 shows skewed distributions with skewness equal to 1.31 and 1.21.”

The following sentence will be added *“A normal kernel was used to obtain these distributions.”*

L194-L195: 5 times greater for both models?

Our answer :

The sentence “There are extreme values indicating surface current speeds with 195 deviations 5 times greater than the standard deviation.” Will be removed.

L207-208: what are the implications of these differences?

Our answer:

Firstly, the original Fig. 5 will be replaced by the following figure:

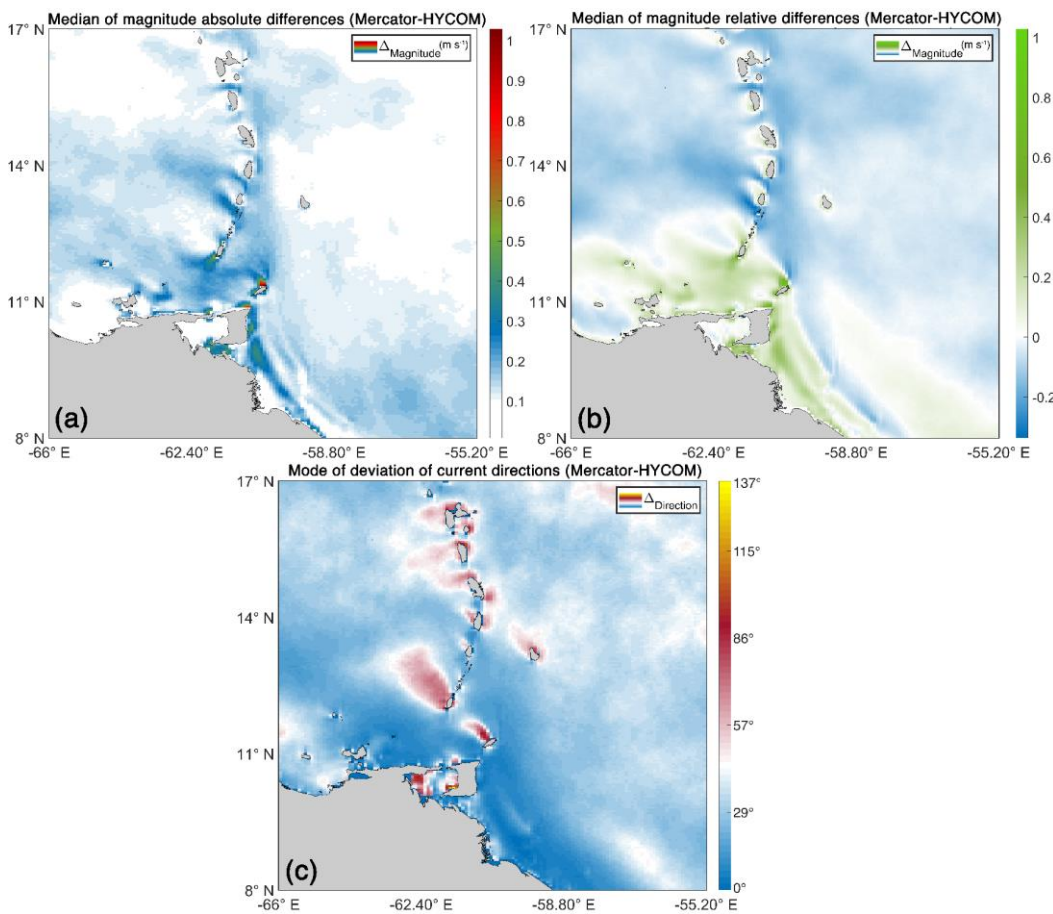


Figure 5: Comparison between Mercator and HYCOM surface currents from 2019 to 2020 on the same 0.08° grid: (a) median of magnitude absolute differences (Mercator-HYCOM) in m s^{-1} and (b) median of magnitude relative differences (Mercator-HYCOM) in m s^{-1} and (c) mode of current direction differences (Mercator-HYCOM) in degree.

The sentence “The largest differences, above 0.5 m s^{-1} are observed in the South part of the LA arc, around Trinidad and Tobago.” Will be replaced by the following sentence:

“In the South part of the LA arc, around Trinidad and Tobago, Mercator current magnitudes are globally higher than HYCOM current magnitudes. Thus, Mercator surface currents might induce higher Sargassum influx from the Western Central Atlantic to the Caribbean Sea in this area.”

L272-L273: “The monthly evolution of observed stranding days on the Guadeloupe coasts, the monthly evolution of Sargassum abundance over the Central Atlantic region (SaWS, <https://optics.marine.usf.edu/projects/SaWS.html>)” I imagine it should be: “Guadeloupe coasts and the monthly evolution...”, to make clear you talking about two datasets. The observed stranding dataset is mentioned in the dataset section (section 2.4), but not the Sargassum abundance over the Central Atlantic region.

Our answer:

Line 272, the sentences:

“The monthly evolution of observed stranding days on the Guadeloupe coasts, the monthly evolution of Sargassum abundance over the Central Atlantic region (SaWS, <https://optics.marine.usf.edu/projects/SaWS.html>) were also analyzed on the focused period 2019-2020 (Figs. 11 and 12). During these two years, the amount of Sargassum over the Central Atlantic region increased significantly from February to July, then decreased from July to November.”

will be replaced by:

“The monthly evolution of observed stranding days on the Guadeloupe coasts, the monthly evolution of Sargassum abundance in the area 30-100 km offshore Guadeloupe were also analyzed on the focused period 2019-2020 (Figs. 11 and 12). During these two years, the amount of Sargassum which may enhance the beaching risk in Guadeloupe increased significantly from February to May, then decreased from May to November.”

Figs. 11 and 12 will be replaced by the following figures:

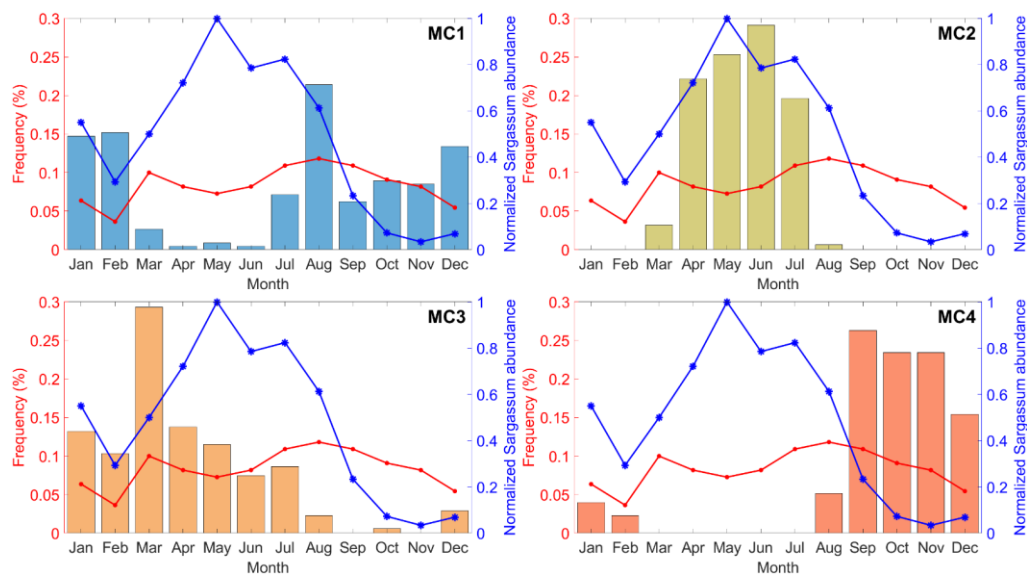


Figure 11: Monthly distribution of cluster occurrence from Mercator outputs, from 2019 to 2020, in the Lesser Antilles (55-66°W, 8-17°N): MC1 (a), MC2 (b), MC3 (c) and MC4 (d). The red line shows the monthly distribution of Sargassum strandings on the coasts of Guadeloupe during the same period. The blue line indicates the monthly evolution of Sargassum abundance in the area 30-100 km offshore Guadeloupe normalized on the maximum value.

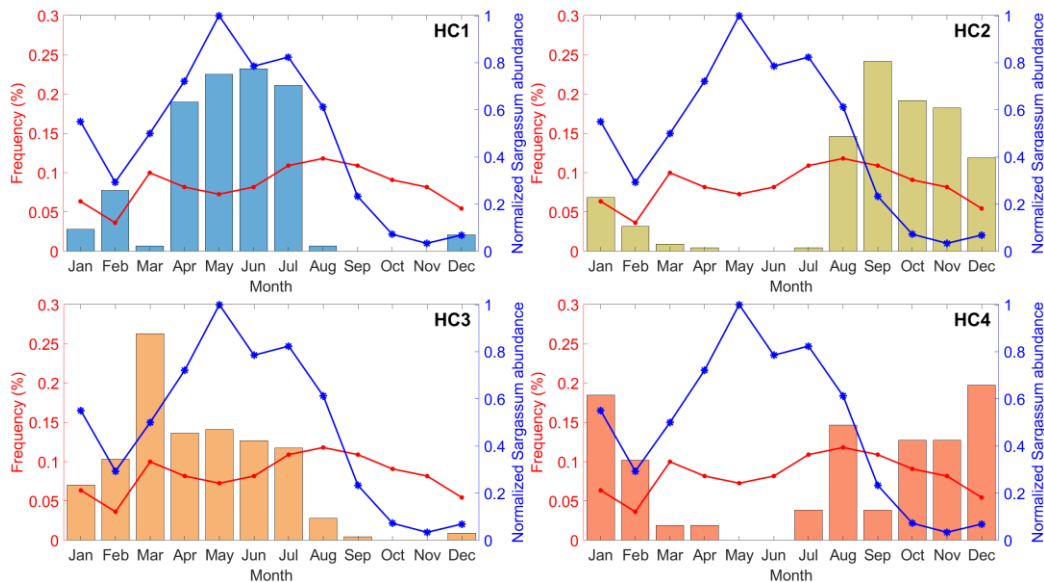


Figure 12: Monthly distribution of cluster occurrence from HYCOM outputs, from 2019 to 2020, in the Lesser Antilles (55-66°W, 8-17°N): HC1 (a), HC2 (b), HC3 (c) and HC4 (d). The red line shows the monthly distribution of Sargassum strandings on the coasts of Guadeloupe during the same period. The blue line indicates the monthly evolution of Sargassum abundance in the area 30-100 km offshore Guadeloupe normalized on the maximum value.

3. Technical corrections

Please write *Sargassum* in italics, like it is done in other studies like for example Johns et al., (2020), as you are writing its scientific name, and even if it is just the genus in this case.

Our answer: Following your suggestion, “*Sargassum*” will be italicized in the revised manuscript.

L10: “including windage effect”: gives the impression the HYCOM and Mercator datasets already include the windage effect, when you actually added separately. Please improve phrasing.

Our answer:

To clarify this point, the sentence:

“The input surface currents including windage effect were derived from the Mercator model and the Hybrid Coordinate Ocean Model (HYCOM).”

will be replaced by

“The input surface currents were derived from the Mercator model and the Hybrid Coordinate Ocean Model (HYCOM) outputs in which we integrated the windage effect.”

L20: “LA received...” to “The LA received...”

Our answer: “LA received” will be replaced by “The LA received”

L23: “...were also be observed...” to “...were also observed...”

Our answer: “were also be observed” will be replaced by “were also observed”

L46: Improve sentence, e.g. “... multi-year reanalysis of wind and current, and numerical models, both the role of subsurface nutrient supply and surface current transport were estimated.”

Our answer:

As you suggested the part:

“...multi-year reanalysis of wind and current, numerical models estimated both the role of subsurface nutrient supply and surface current transport.”

will be replaced by

“...multi-year reanalysis of wind and current, and numerical models, both the role of subsurface nutrient supply and surface current transport were estimated.”

L50: “Sargassum Watch System SaWS” to “Sargassum Watch System (SaWS)”

Our answer: “Sargassum Watch System SaWS” will be replaced by *“Sargassum Watch System (SaWS)”*

L83: “in Johns (2020)” et al. missing.

Our answer: “in Johns (2020)” will be replaced by *“in Johns et al. (2020)”*

L92: Please define the abbreviations HYCOM and NCODA (HYCOM defined in abstract but not in the main text)

Our answer: The pa

2.1 HYCOM surface current dataset

“Fine scale surface current data from the 1/25-degree HYCOM + NCODA Gulf of Mexico analysis model (GOMu0.04/expt_90.1m000 version, Hogan et al, 2014; Helber et al., 2013; Cummings and Smedstad, 2013; Cummings, 2005) between 1st January 2019 (i.e., available data starting date) and 31 December 2020 were analyzed. Daily 12Z fields giving the u and v components of the current at 50 cm depth were used. These fine resolution current data were not used in previous studies dealing with Sargassum hazard (Putman et al., 2018; Johns et al., 2020).

will be replaced by:

“2.1 HYCOM surface current dataset

Daily 12Z surface current components from the 41-layer Hybrid Coordinate Ocean Model (HYCOM) global 1/12-degree analysis (HYCOM GLBy0.08 version), were examined. The HYCOM surface forcing including 10-m wind velocities are extracted from Climate Forecast System Version 2 (CFSv2). The Navy Coupled Ocean Data Assimilation (NCODA) system ...

L94: Please define 12Z fields.

Our answer:

“12Z” will be replaced by *“12 UTC (i.e., Coordinated Universal Time) ”*

L94-95: “u and v components” to “zonal (u) and meridional (v) velocity components”

Our answer: “u and v components” will be replaced by *“zonal (u) and meridional (v) velocity components”*

L101-102: “Comparison...in the focused region” to “A comparison.. in the study region.”

Our answer:

This sentence “*Comparison between HYCOM and Mercator results would help to better understand the effects of spatial resolution on surface current patterns in the focused region.*” Will be removed in the revised manuscript.

L107: “Sargassum raft transport”, maybe trajectories instead of transport is more appropriate?

Our answer:

Following your suggestion the part “*Lagrangian simulations of Sargassum rafts transport in the Caribbean region.*”

Will be replaced by

“*Lagrangian simulations of Sargassum raft **trajectories** in the Caribbean region.*”

L112-113: “The region analyzed in the present work corresponds to the CA - TA1 region defined in Johns et al. (2020)” already mentioned in L82-83, is it necessary to repeat here?

Our answer:

This sentence will be removed.

L116-117: “This period includes 730 observational days with 110 days of observed strandings.” , phrasing not clear do you mean that out of the total 730 days of data, only 110 days included observations of Sargassum strandings?

Our answer:

During the two years 2019-2020, only 110 days of observed beachings in Guadeloupe have been recorded.

To clarify this point,

The following sentence:

This period includes 730 observational days with 110 days of observed strandings.”

Will be replaced by

“During this period of 730 days, only 110 days of observed beaching have been recorded (i.e., 30 days in 2019 and 80 days in 2020). During the year of 2021, 78 days of beaching were observed in Guadeloupe.”

L137: “above Barbados island” to “above the island of Barbados”

Our answer: “*above Barbados island*” will be replaced by “***above the island of Barbados***”

L142: “The similarity of the most similar fields is estimated per pair..” Improve phrasing. What do you refer to exactly? Per pair of Sargassum meshes?

Our answer:

Line 141-150 the two paragraphs,

“The second step was to group the information carried by the daily current velocity fields conditionally to the three given zones into histograms. The similarity of the most similar fields is estimated per pair and per zone based on the symmetrized Kullback-Leibler (KL) divergence computed from the histograms (Kullback and Leibler, 1951). This allows the entropy between two distributions to be expressed without having a priori reasoning concerning the probability distribution. The similarity between two histograms was quantified this way. The last step consisted in calculating the average of the divergence values for each zone. This allows to have a single value, named Expert Distance (ED) quantifying the similarity between the individuals of the database during clustering.

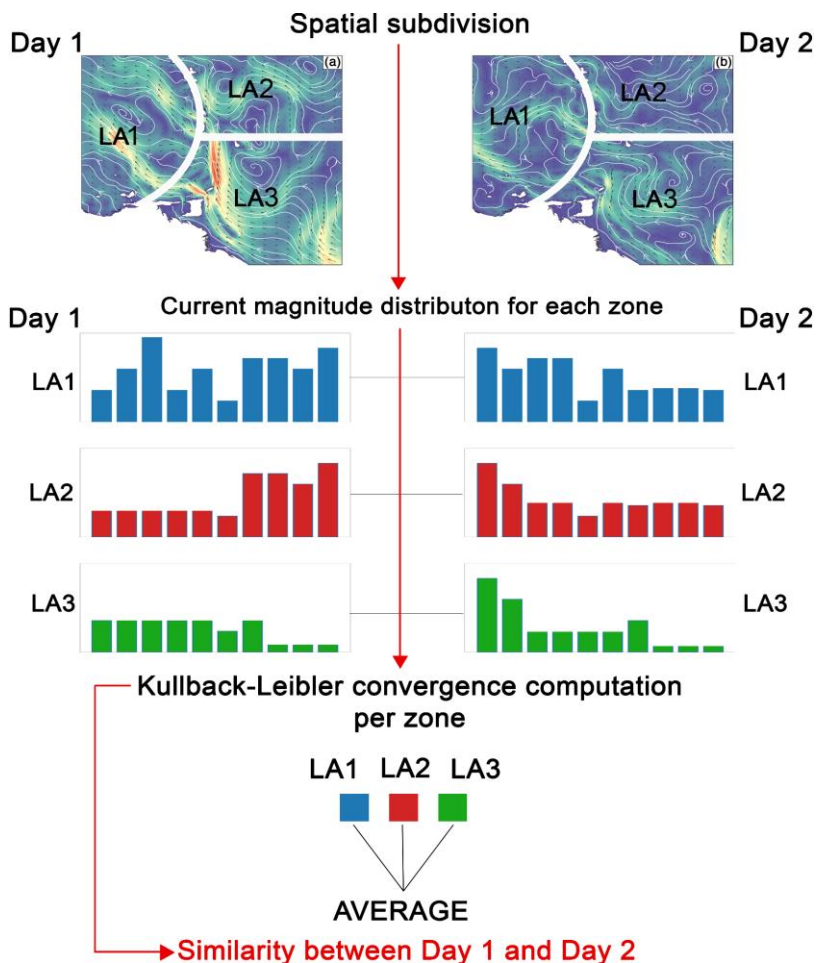
The clustering results have been evaluated using the Silhouette Index (Rousseeuw, 1987). The SaMk index defined in Biabiany et al. (2020) was used. This allows to express the quality of a clustering, by the average of

the quality of each cluster, which is itself the average of the silhouette indices $s(i)$ over the cluster elements. This index is defined as follows:”

will be replaced by:

“Clustering methods used in the present study were K-Means (KMS) and Agglomerative Hierarchical Clustering (HAC). Euclidean distance (L2) is usually computed to compare data in these algorithms. However, Biabiany et al. (2020) showed in a recent work that clustering results can be improved by using an Expert Distance (ED) to compare data. This ED is based on an empirical spatial subdivision and the use of Kullback-Leibler divergence, in order to quantify the similarity between two fields.”

To clarify this part, the following schematic will be added in the methods section.



“Figure XX: The schematic of the Expert Distance process.”

L148: “The SaMk index” to “The Silhouette (SaMk) index”

Our answer: “The SaMk index” will be replaced by “The Silhouette (SaMk) index”

L151: Define all variables of equation 2!

Our answer:

This sentence will be added below the equation (2) :

where k is the number of clusters, C_j the set of days from the cluster j , i a day form C_j and $s(i)$ the silhouette index (Rousseeuw, 1987) value of day i .

L153-154: Improve phrasing.

The sentence: *“To better understand current regime dynamics which may lead to Sargassum strandings on the coasts of Guadeloupe, the past stranding 30-day current backward sequences were analyzed.”*

Will be replaced by

“To better understand current dynamics which may lead to Sargassum beaching in Guadeloupe, we analyzed the 30-day current sequences before beaching.”

L156: “January 2020” to “January 2019”

Our answer: “January 2020” will be replaced by *“January 2019”*

L165-L166: “ surface currents with windage effects (Mercator, HYCOM and ERA-5)” to “ surface currents (Mercator and HYCOM) with windage effects (ERA-5)”

Our answer: “ surface currents with windage effects (Mercator, HYCOM and ERA-5)” will be replaced by *“ surface currents (Mercator and HYCOM) with windage effects (ERA-5)”*

L186-L187: “The proposed tree in Fig. 2...”. Move to new line, to separate it from the phrase explaining the terms in equation (4)

Our answer: The sentence starting with *“The proposed tree in Fig. 2...”* will be moved to new line.

L191: “do not exceed 2.57 m/s”. Maybe better to say the maximum is 2.57 m/s, if not it sounds like 2.57 m/s is a key velocity value that should not be exceeded for some reason.

Our answer: *“ For both models HYCOM and Mercator, the velocity magnitudes do not exceed 2.57 m s⁻¹ and 90% of them remain below 0.65 m s⁻¹.”*

Will be replaced by

“For both models HYCOM and Mercator, the maximum surface velocity is 2.57 m s⁻¹ and 90% of them remain below 0.65 m s⁻¹.”

L193-L194: add at end to which model it each value corresponds to e.g. “.. for HYCOM and Mercator, respectively.”

Our answer: *“Figure 3 shows skewed distributions with skewness equal to 1.31 and 1.21”*

Will be replaced by

“Figure 3 shows skewed distributions with skewness equal to 1.31 and 1.21 for HYCOM and Mercator, respectively.”

L205: “Globally, at sea, the current..” Is it necessary to specify at sea? What do you exactly mean with at sea here, open ocean?

Our answer: “at sea” will be replaced by *“at open ocean”*

L210: “into three magnitude groups of 45°” to “into three magnitude groups of 45° intervals”?

Our answer: “into three magnitude groups of 45°” will be replaced by *“into three magnitude groups of 45° intervals”*

L215: Improve phrasing, gives the impression you used equation (1) to perform the clustering.

Our answer: the part “*according to equation (1)*” will be removed.

L244: “Table 3 shows results” to “Table 3 shows the results”

Our answer: “*Table 3 shows results*” will be replaced by “*Table 3 shows **the** results*”

L297: “remain with probabilities” add probabilities of... Help the reader follow better your study, recalling details

Our answer:

This part will be modified with new results produced by our improved version of Decision Support System. Please find below the modified Table 6 which includes recalling details.

Time range around D (day)	Datasets	TP (recall %)	TN (recall %)	FP (ratio %)	FN (ratio%)	Accuracy (ratio %)
0	HYCOM	48 (61.5%)	152 (53.1%)	134 (36.8%)	30 (8.2%)	200 (54.9%)
	Mercator	44 (56.4%)	141 (49.3%)	145 (39.8%)	34 (9.3%)	185 (50.8%)
+/- 1	HYCOM	53 (67.9%)	170 (59.4%)	(-)	(-)	(-)
	Mercator	47 (60.3%)	142 (49.6%)	(-)	(-)	(-)
+/- 2	HYCOM	54 (69.2%)	184 (64.3%)	(-)	(-)	(-)
	Mercator	47 (60.3%)	146 (51%)	(-)	(-)	(-)
+/- 3	HYCOM	58 (74.4%)	193 (67.5%)	(-)	(-)	(-)
	Mercator	47 (60.3%)	150 (52.4%)	(-)	(-)	(-)

Table 6: Decision tree performance (with TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative, (-): same as above).

L317: Improve wording of Section 4.2 title, for example can simply remove “hazard”

Our answer: “Hazard” will be removed. The new Section 4.2 title will be: “*4.2. Surface current analysis applied to Sargassum*”

L320: “retroflexion” to “retroflexion”

Our answer: “retroflexion” will be replaced by “*retroflexion*”

L345 “The first peak of strandings, in March and seems..” to “The first peak of strandings, in March, seems..”

Our answer: “*The first peak of strandings, in March and seems..*” will be replaced by “*The first peak of strandings, in March, seems..*”

L373: Write as K-Means, and also in L217, write method in the same way.

Our answer: “*k-mean*” will be replaced by “*K-Means*” (L217 and L373)

4. Figures and tables

Figure 2: Describe BASE abbreviation as in L175.

Our answer:

“*BASE*” is not an abbreviation, to avoid misunderstanding, the term will be written in the normal case (i.e., Base or base). Base will be explained both in the Fig. 2a and in the caption.

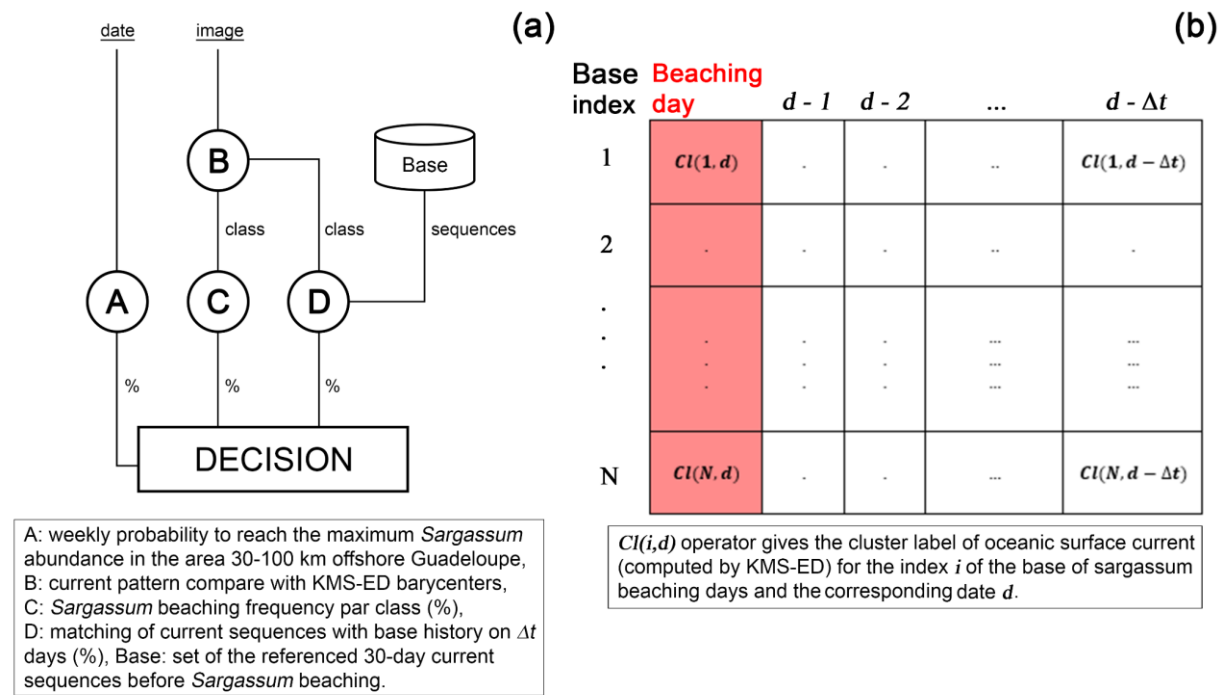


Figure 2: (a) Scheme of the decision tree classifier to predict *Sargassum* stranding probability. (b) Combination base of oceanic currents clusters labels obtained by KMS-ED from each stranding day to Δt days before. *Base* is the set of 30-day current sequences before *Sargassum* beaching.

Figures 4, 9 and 10: x-axis tick labels not clear, please improve.

Our answer: The x-axis tick labels will be improved as below.

Figures 4, 9 and 10 will be replaced by the following figures.

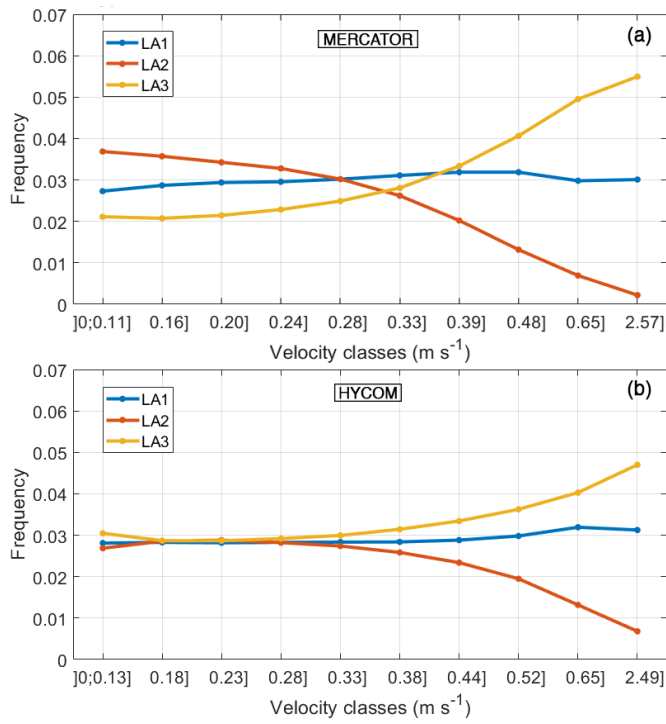


Figure 4: Relative frequency distribution of current speeds for the three offshore sub-regions around the Lesser Antilles (2019-2020), LA1 (blue), LA2 (red), LA3 (yellow). (a) Mercator with ERA-5 windage and (b) HYCOM with ERA-5 windage.

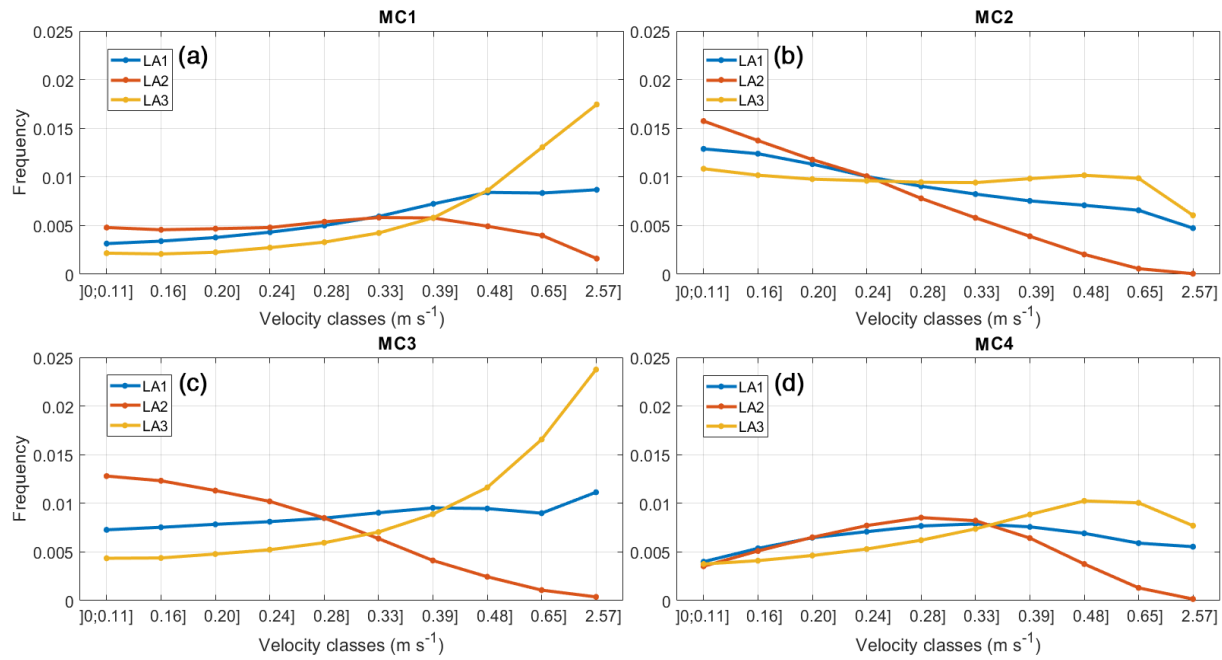


Figure 9: Relative frequency distribution of current speeds for the three offshore sub-regions: MC1 (a), MC2 (b), MC3 (c) and MC4 (d). The representative elements were obtained after KMS-ED clustering for Mercator.

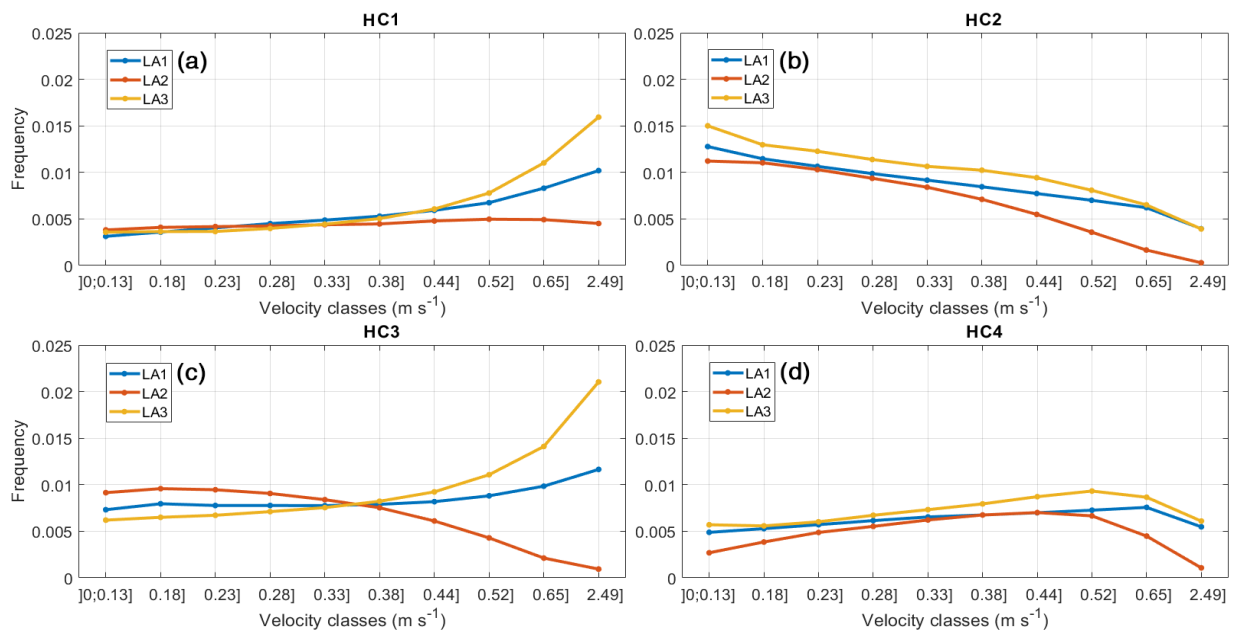


Figure 10: Relative frequency distribution of current speeds for the three offshore sub-regions: **HC1 (a)**, **HC2 (b)**, **HC3 (c)** and **HC4 (d)**. The representative elements were obtained after KMS-ED clustering for HYCOM.

Table 1: Header mean to Mean

Our answer: “mean” will be replaced by “*Mean*”

Deciles (D _i)	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Max	Mean	Sigma
Mercator (m s ⁻¹)	0.11	0.16	0.20	0.24	0.28	0.32	0.39	0.48	0.65	2.57	0.33	0.22
HYCOM (m s ⁻¹)	0.13	0.18	0.23	0.28	0.32	0.38	0.44	0.52	0.65	2.49	0.36	0.21

Table 1: Boundaries of the histogram classes used to quantify surface currents velocity data with Sigma as Standard deviation.

Table 5: Caption mention what n and % refer to exactly.

Our answer:

Table 5 will be modified as shown below:

30-day current sequence before beaching (HYCOM)	Seq1	Seq2	Seq3	Seq4
Number (n)	18	40	7	42
Ratio (%)	16.8	37.4	6.5	39.3

The caption will be modified as below:

Table 5: Distribution of backward sequence clusters, with (n) corresponding to the number of sequences in each cluster and (%) corresponding to the ratio of the number of sequences on the total.