Dear referee 1,

We thank you for the attention that you paid to this review and for your helpful comments and suggestions.

Firstly, in the introduction of your report, you mentioned the *"rather low performance of the classifier"*. Following this remark, we made some major changes 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) The discussion of the appropriate time and space scale must be introduced in the methods to justify the choices made (30 days sequences, areas, monthly probability).

Our answer:

We will justify our choices of time and space scale (30 days sequences, areas, monthly probability) in the methods section.

Areas are already described in the "2.5.2 Use of Expert Deviation in clustering algorithms" methods subsection, Lines 135-140:

The LA study area was separated into three parts (Fig. 1b) based on the Sargassum rafts transport centers of action reported in the literature (Franks et al., 2016; Berline et al., 2020). To the west of LA, the first zone, LA1, is centered on the Caribbean Sea. To the east, the Atlantic zone has been split into two areas towards 13.5°N, just above Barbados island. To the south-east is the LA3 zone under the influence of the North Equatorial Recirculation Region (NERR) and its retroflection rings, while to the north-east is the LA2 zone, more representative of the North Equatorial Current. 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).

Concerning the 30 days duration, the following sentences will be added in the methods, L154 (Section 2.6):

"The 30 days duration corresponds to the empirical transport time of a passive particle moving from the main entrance location of Sargassum rafts in the Lesser Antilles area (i.e., in LA3 zone, 8°N; -55°E) to Guadeloupe (i.e., LA2 zone). Based on the mean current magnitude of 0.2 m s^{-1} (average value over the LA zone, in HYCOM and in Mercator data) and the distance of 500 km between the main entrance location and the Guadeloupe coasts, 29 days are obtained for the transport. For simplicity, the duration of 30 days was selected instead of 29 days."

To improve the decision support system, the stranding monthly probability (i.e., Module A in the decision tree) will be replaced by a weekly probability of Sargassum presence in an area of 100 km radius offshore Guadeloupe. This probability is based on the cumulative 7-day Floating Algae density (Wang and Hu, 2016) estimated in this area during the two years 2019 and 2020.

The following section will be added in the "Datasets and method" section:

*"**2.5 Satellite-based offshore abundance of Sargassum***

Sargassum satellite observations were included in the present decision support system. To quantify the abundance of Sargassum in an area of 100 km radius offshore Guadeloupe, the 7-day Floating Algae (FA) density fields derived from the Alternative Floating Algae Index (Wang and Hu, 2016) were analyzed. As described by Trinanes

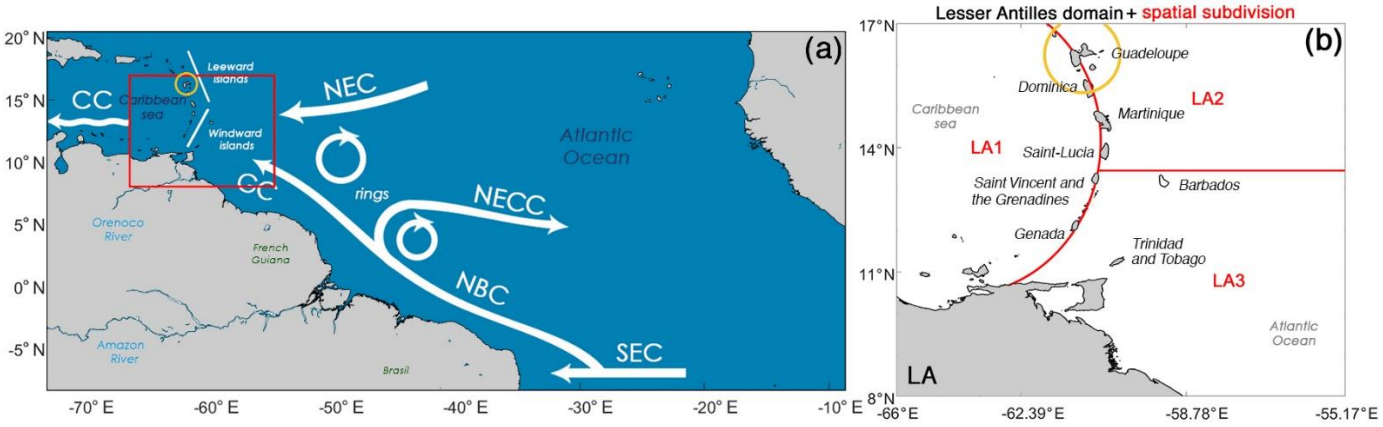
et al. (2021), the 7-day Floating Algae (FA) density fields are accumulated on 7 days and have a 0.1° resolution. Due to optical complexity in nearshore waters, the FA density fields are masked with missing values within 30 km from shoreline (Trinanes et al. 2021). The cumulative FA density values were summed in the area 30-100 km offshore Guadeloupe (Fig. 1) then weekly averaged during the two years 2019 and 2020."

In the "Decision support system" section, the following sentence (L.168):

"Module A takes as input the month of the selected day and returns the associated monthly probability (frequency) of stranding;"

Will be replaced by:

"Module A takes as input the week number of the selected day and returns the associated weekly probability to reach the maximum offshore abundance of Sargassum (based on observational FA density values during the two years 2019 and 2020)."



"Figure 1: (a) Main oceanic currents occurring and interacting in the central Atlantic and the Lesser Antilles regions; Caribbean Current (CC), North Equatorial current (NEC), North Brazil current (NBC), North equatorial Counter Current (NECC), South Equatorial current (SEC). Lesser Antilles domain (LA): the red rectangle corresponds to the study area (55-66° W, 8-17° N); (b) Spatial subdivision of the study area into three sub-areas: LA1 (i.e., Caribbean Sea), LA2 (i.e., North Tropical Atlantic above Barbados (13.2° N)) and LA3 (i.e., North Tropical Atlantic below 13.2° N). The yellow circle corresponds to the 100 km offshore Guadeloupe area in which the satellite-based Sargassum abundance is analysed."

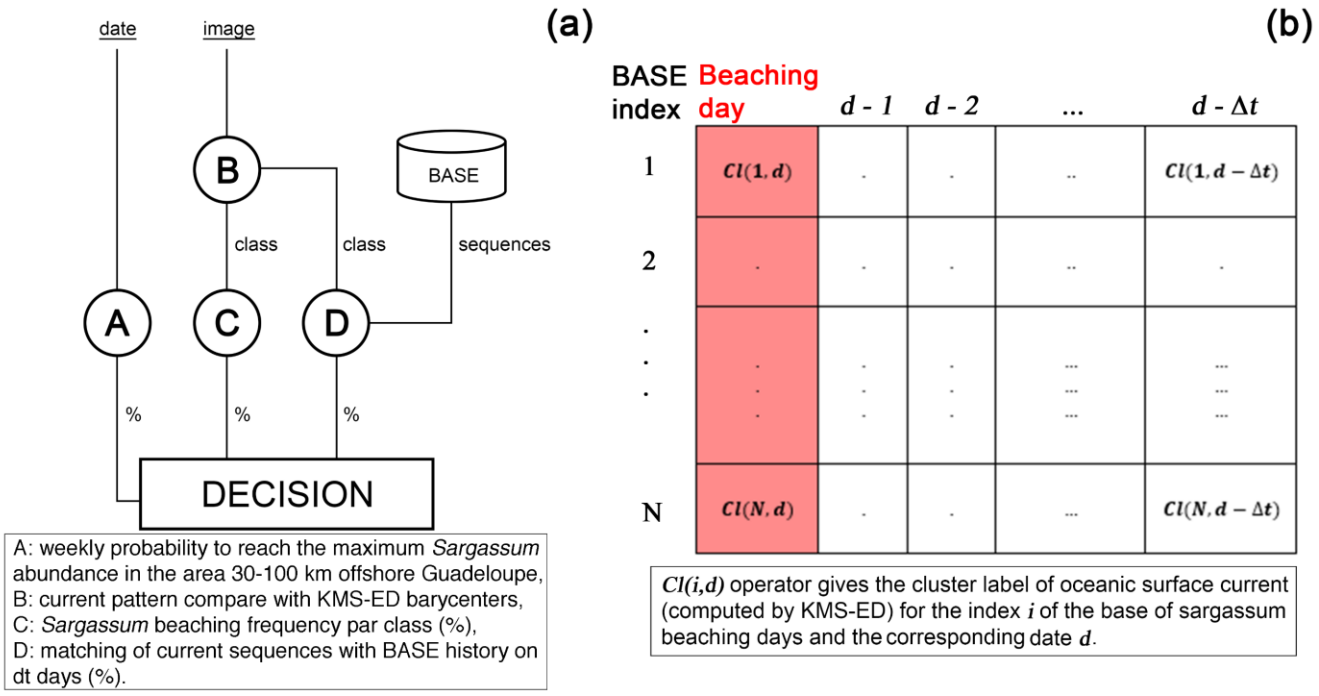


Figure 2: (a) Scheme of the decision tree classifier to predict Sargassum stranding probability. (b) Combination base of oceanic currents clusters labels obtained by KMS-ED from each stranding day to Δt days before.

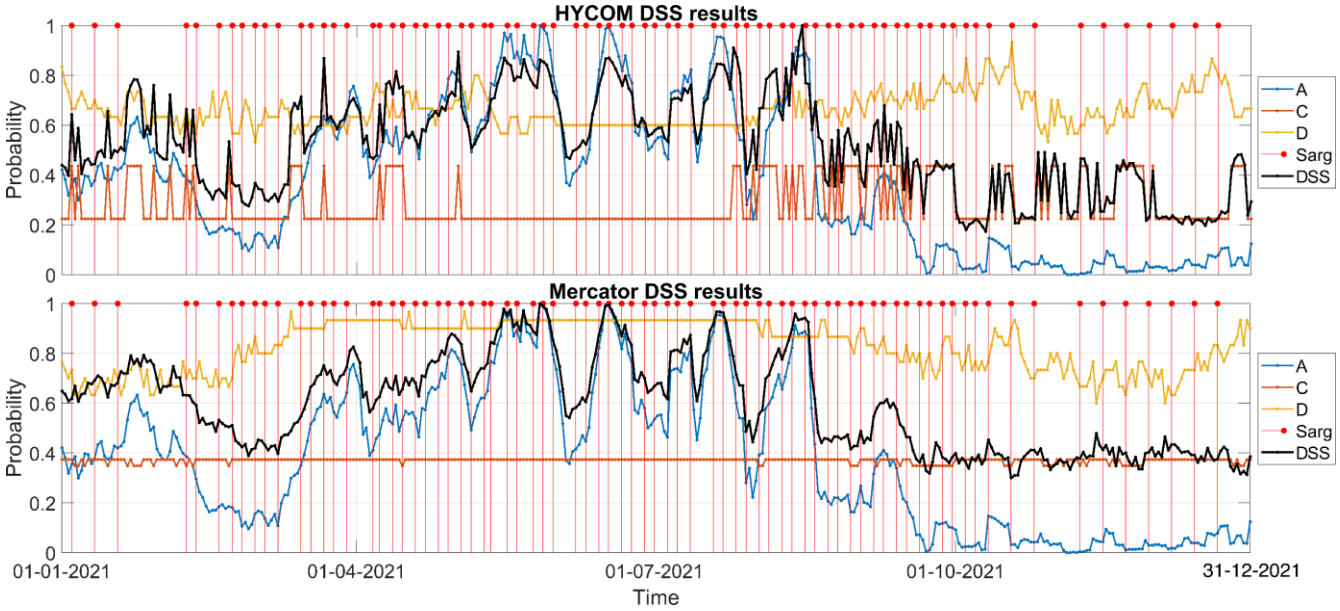


Figure 15: Decision Support System (DSS) results: probability of beaching obtained per module. Weekly probability to reach the maximum Sargassum abundance in the area 30-100 km offshore Guadeloupe for module A (blue line), stranding frequency per cluster for module C (orange line), match percentage for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

2) Strandings occurring in Guadeloupe are not affected by the dynamics of zones LA3 and LA1, but only by LA2. You should take it into account in the study.

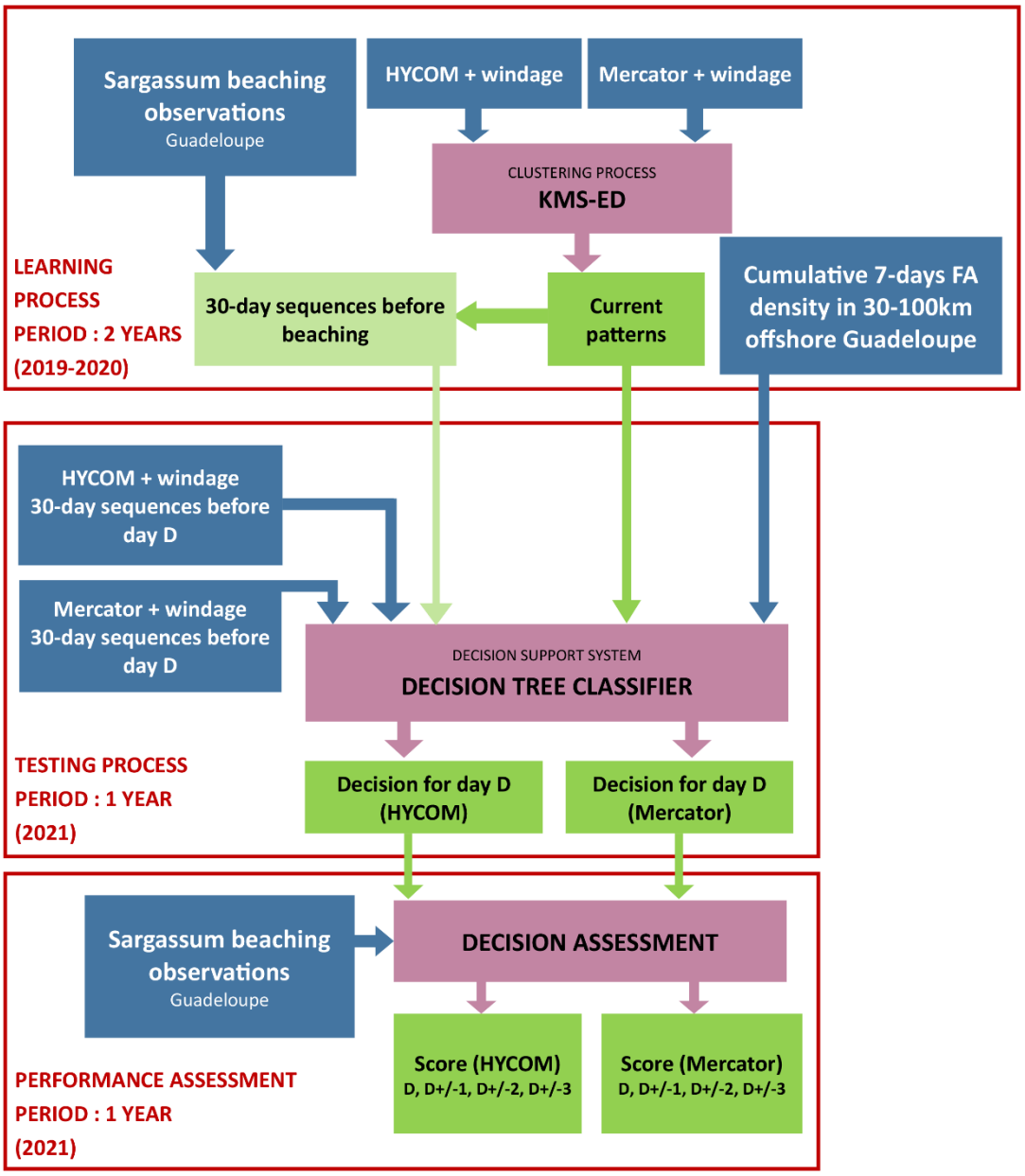
Our answer:

In the present study, the short-range transport from the LA2 zone is examined as well as the medium-range transport of Sargassum from the LA3 zone (i.e., the South of the Lesser Antilles Arc). Strandings occurring in Guadeloupe may be affected by dynamics of both zones LA2 and LA3.

Although interesting, in its present form the overall approach needs to be better explained and the text is quite difficult to read. The language can be significantly improved and small typos removed.

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.



"Figure XX: The schematic of the adopted methodology."

Detailed comments

The abstract

L15 "small scale". Better use high resolution, as this is the crucial model configuration choice.

Our answer:

Following your suggestion, *"small scale"* will be replaced by *"high resolution".*

L20. Windward vs leeward. The only study citing this point is Marechal et al 2017.

Our answer:

To correct this error, the following sentences L19-22:

"During the periods 2011-2012, then 2014-2019, massive Sargassum strandings impacted most coasts of the Lesser Antilles (LA), mainly those facing east and southeast. LA received large amounts of algae on the windward Atlantic coastline, while leeward Caribbean coastal areas remained slightly affected (Franks et al., 2012, Gower et al., 2013, Johnson. et al., 2014, Hu et al., 2016, Wang and Hu, 2016, Maréchal et al., 2017)."

will be changed to:

"During the periods 2011-2012, then 2014-2019, massive Sargassum strandings impacted most coasts of the Lesser Antilles (LA), mainly those facing east and southeast (Franks et al., 2012, Gower et al., 2013, Johnson. et al., 2014, Hu et al., 2016, Wang and Hu, 2016). LA received large amounts of algae on the windward Atlantic coastline, while leeward Caribbean coastal areas remained slightly affected (Maréchal et al., 2017)."

L23 typo: also observed

Following your suggestion, *"Strandings were also to be observed"* will be replaced by *"Strandings were also observed".*

L41 wording: The volumes to be collected

Following your suggestion, *"The volumes needed to be collected"* will be replaced by *"The volumes to be collected".*

L55 There is also Jouanno et al 2020 (Env Res Letters) on the role of rivers.

Our answer:

Following your suggestion, the reference "Jouanno et al. 2021" will be added here and in the references Section.

Jouanno, J., Moquet, J.-S., Berline, L., Radenac, M.-H., Santini, W., Changeux, T., Thibaut, T., Podlejski W., Ménard, F., Martinez, J.-M., Aumont, O., Sheinbaum, J., Filizola N. and Moukandi N'Kaya G. D.: Evolution of the riverine nutrient export to the Tropical Atlantic over the last 15 years: is there a link with Sargassum proliferation?, Environ. Res. Lett., 16, 8 pp, https://doi.org/10.1088/1748-9326/abe11a, 2021.

L67 Unclear : The probability of a set of data…

Our answer:

- Line 67: 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 replaced by:

"None of them used predictive modelling based on a decision tree including current patterns and probabilities related to Sargassum strandings. This Sargassum beaching predictive tool based on repeatable current patterns would be useful for risk prevention managers."

L94 u and v: better zonal and meridional

Following your suggestion, *"giving the u and v components"* will be replaced by *"giving the zonal (u) and meridional (v)".*

L91-L100 : You need to better describe the configuration of the reanalysis datasets you are using. In particular you should give the forcing fields (winds etc) and the data assimilated in each model. This is important as Mercator and Hycom models assimilate the same type of data (altimetry in particular), which largely explains their consistency in terms of large scale patterns (ie clusters).

Our answer:

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 HYCOM + NCODA 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.".

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.

L105. See also Berline et al 2020

Our answer:

Following your suggestion, we will add *"Berline et al. 2020"* here.

Following the CC1 community comment:

"Line 104: I am confused, what is the basis for assuming the "optimal factors of Cw = 0.01"? Surely this is not the case based on data from Johns et al. 2020, which showed no evidence that a windage factor of 1% was appropriate for *Sargassum*. They simply picked the "reasonable" value that has been used in the earlier publication Putman et al. 2018. The value of 1% was chosen by Putman et al. 2018 to test the sensitivity of model predictions to windage and did not claim that it was optimal (or even somewhat correct). Work since that point has been conducted which seems to suggest that the situation is somewhat more complicated, see Putman et al. 2020 (already cited elsewhere) and Johnson, D.R., Franks, J.S., Oxenford, H.A. and Cox, S.A.L., 2020. Pelagic Sargassum Prediction and Marine Connectivity in the Tropical Atlantic. Gulf and Caribbean Research, 31(1), pp.GCFI20-GCFI30. Whether the best windage value is 0, 0.5%, 1%, 3% or something else likely depends on the oceanographic region and the ocean circulation model and wind product used."

the sentence Line 104:

"*Surface wind influences the transport of floating seaweed rafts, with an optimal factor of Cw = 0.01, which corresponds to the drag coefficient or windage, following Johns et al. (2020). A first clustering (KMS-L2) on Mercator analysis without windage had been proposed by Bernard et al. (2019). Berline et al. (2017), Putman et al. (2018) and Johns et al. (2020) have shown that the windage improves the Lagrangian simulations of Sargassum rafts transport in the Caribbean region. The windage was included in the present surface current clustering.*"

will be replaced by:

"Surface wind influences the transport of floating seaweed rafts and a drag or windage coefficient must be added to the surface currents. The value of Cw = 0.01 was used by Putman et al. 2018, Johns et al. (2020) and Berline et al. (2020). The use of other windage values should be investigated in a further study."

L108 The sentence:

"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

"Surface wind data (at 1000 hPa) from the ERA-5 model were integrated with Mercator and HYCOM currents following this formula:"

L106 Berline et al 2020, not 2017

Our answer:

We will fix this error in the revised manuscript.

"Berline et al. (2017)" will be replaced by *"Berline et al. (2020)".*

L112. Check the year for Putman et al

Our answer:

We will fix this error in the revised manuscript.

"Putman et al. (2016)" will be replaced by *"Putman et al. (2018)".*

L121 Past tense is expected

Our answer:

We will fix this error in the revised manuscript.

"we shall use" will be replaced by *"we used".*

L125: 2.5.1: useless subsection

Our answer:

Following your suggestion, the subsection "2.5.1 Clustering methods process" will be removed.

The paragraph *"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. The result is an automated analysis with its own expertise on the input data."

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)."*

L132-133 Wording. Better : can lead to group different physical situations…

Our answer:

We will fix this error in the revised manuscript.

"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."*

L134. Unclear : " From L2"

Our answer :

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)."

L139. Unclear terminology : Meshes. Better grid points. Need to be homogeneous throughout the text (grid cell at L205, grid points at L209)

Our answer:

We will fix these errors in the revised manuscript.

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

L205: *"grid cell"* will be replaced by *"grid point".*

The lines 141-150 on clustering should be grouped in a dedicated section.

The clarity and wording of this whole section must be improved. For instance " The similarity of the most similar fields (...) "Explanation should be given for one zone to avoid redundancy. A schematic would help. I do not clearly understand the algorithm for clustering. What is the role of the average divergence ?

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,

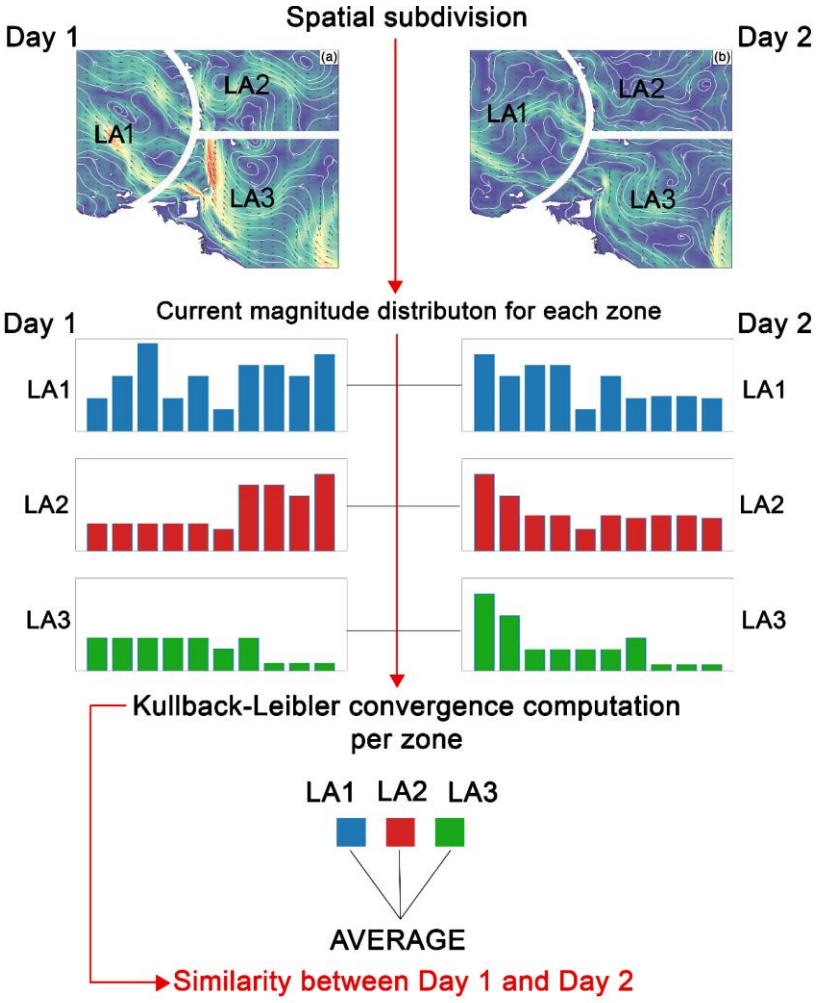
1951). This allows the entropy between two distributions to be expressed without having a priori reasoning concerning the probability distribution. The similarity between two histograms was quantified this way. The last step consisted in calculating the average of the divergence values for each zone. 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 point, the following schematic will be added in the methods section.



"Figure XX: The schematic of the Expert Distance process."

L152-162 (Section 2.6): wording and clarity should be improved. The word 'backward' is misleading here as there is no time integration in your analysis. You simply take the 30 days before one peculiar stranding event. What justifies the 30 days duration? Transport? Then is it consistent with the areas LA1, 2, 3?

Our answer:

We will try to improve wording and clarity. The words *"the past stranding 30-day current backward sequences"* will be replaced by *"the 30-day sequences before beaching"*.

The following sentences will be added in the methods, L154 (Section 2.6):

"The 30 days duration corresponds to the empirical transport time of a passive particle moving from the main entrance location of Sargassum rafts in the Lesser Antilles area (i.e., in LA3 zone, 8°N; -55°E) to Guadeloupe (i.e., LA2 zone). Based on the mean current magnitude of 0.2 m s^{-1} (average value over the LA zone, in HYCOM and in Mercator data) and the distance of 500 km between the main entrance location and the Guadeloupe coasts, 29 days are obtained for the transport. For simplicity, the duration of 30 days was selected instead of 29 days."

L158-162. Unclear. "optimal matching methods" : which one ?

You compute a distance metric between the sequences of cluster numbers from previous section?

Our answer:

The Longuest Common Subsequence (LCS) methods were used to compare the back-sequences. 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)."*

can be replaced by: *"Dissimilarities between these backward sequences were calculated before dividing the population into several groups using a hierarchical classification (Larmarange et al., 2015)."*

L161 Wald's or Ward ?

Our answer: *We will fix this error in the revised manuscript. The right term is Ward.*

L164 "At a given location" : which one ?

Our answer: *This group of words has been deleted.*

L168 : Why monthly? Are the stranding observations autocorrelated at this scale ?

To improve the decision support system, the stranding monthly probability (i.e., Module A in the decision tree) will be replaced by a weekly probability of Sargassum presence in an area of 100 km radius offshore Guadeloupe. This probability is based on the cumulative 7-day Floating Algae density (Wang and Hu, 2016) estimated in this area during the two years 2019 and 2020.

The following section will be added in the "Datasets and method" section:

*"**2.5 Satellite-based offshore abundance of Sargassum***

Sargassum satellite observations were included in the present decision support system. To quantify the abundance of Sargassum in an area of 100 km radius offshore Guadeloupe, the 7-day Floating Algae (FA) density fields derived from the Alternative Floating Algae Index (Wang and Hu, 2016) were analyzed. As described by Trinanes et al. (2021), the 7-day Floating Algae (FA) density fields are accumulated on 7 days and have a 0.1° resolution. Due to optical complexity in nearshore waters, the FA density fields are masked with missing values within 30 km

from shoreline (Trinanes et al. 2021). The cumulative FA density values were summed in the area 30-100 km offshore Guadeloupe (Fig. 1) then weekly averaged during the two years 2019 and 2020."

In the "Decision support system" section, the following sentence (L.168):

"Module A takes as input the month of the selected day and returns the associated monthly probability (frequency) of stranding;"

Will be replaced by:

"Module A takes as input the week number of the selected day and returns the associated weekly probability to reach the maximum offshore abundance of Sargassum (based on observational FA density values during the two years 2019 and 2020)."

L170 L172 L176. Wording : "which " can be removed

Our answer: We will fix this error in the revised manuscript. *"which"* will be removed.

L184 I understand you compute the average of P over an ensemble j pertaining to R. Use this notation then.

Our answer: Following your remark, the equation (4) L184 will be replaced by:

"DECISION (i) = P(i) > MEAN(P(j))

where j ∈ R, the set of past days (2019-2020),…"

L191 Is it including windage ?

Our answer: Yes it is including windage.

The caption of the Figure 3: "*Distributions of oceanic surface currents including windage for both models, HYCOM (blue) and Mercator (red) datasets.*"

will be replaced by:

"*Distributions of oceanic current magnitude including windage for both models, HYCOM (blue) and Mercator (red) datasets*"

L191 "intensities ": better magnitude

Our answer: Following your suggestion, *"intensities"* will be replaced by *"magnitude"*.

L205 Mercator and Hycom outputs are not given on the same grid. Then how do you compare them?

Should be explained in methods.

Our answer:

As added above, the description of current data will be corrected. Mercator and HYCOM fields examined in the present study are given on the same 0.08-degree grid.

In "HYCOM surface current dataset" section, the sentences:

"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."

will be added.

L205 : "At sea" : you mean offshore

Our answer: Following your remark, *"At sea"* will be replaced by *"offshore"*.

L226 parangon

Our answer: Following your remark, *"paragon"* has been replaced by *"parangon"*.

L241-242 should be in methods

Our answer:

Following your suggestion, this part including matching formula (5) will be moved in method section before the section 2.6.

L245 "most important": better highest

Our answer: Following your suggestion *"most important"* has been replaced by *"highest"*.

L273 Where is the central Atlantic region used to quantify offshore abundance? Add it on the map.

Our answer:

The offshore abundance from the satellite-based Sargassum Watch System (SaWS) will be replaced here by the monthly averaged offshore abundance of Sargassum (based on the floating algae density in the area 30-100 km offshore Guadeloupe).

Figs. 11 and 12 will be modified this way :

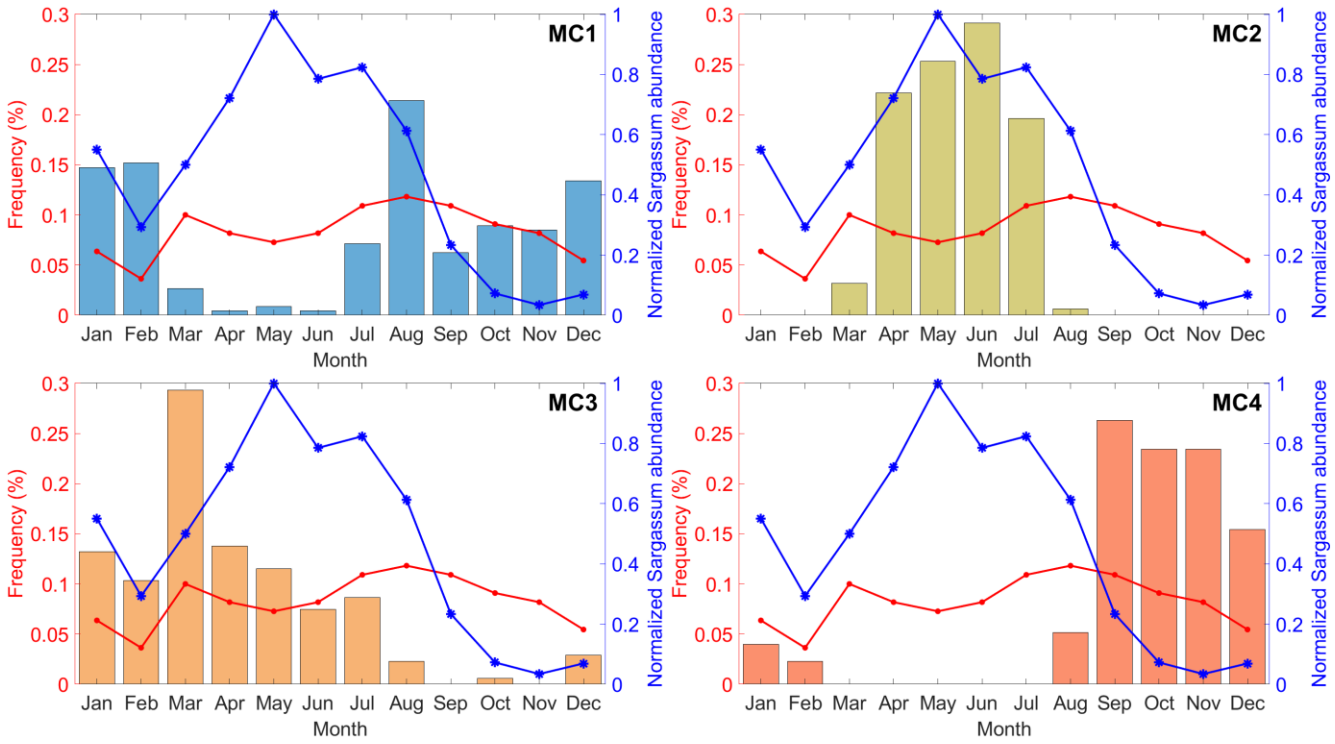


Figure 11: Monthly distribution of cluster occurrence from Mercator outputs, from 2019 to 2020, in the Lesser Antilles (55-66°W, 8-17°N): MC1 (a), MC2 (b), MC3 (c) and MC4 (d). The red line shows the monthly distribution of Sargassum strandings on the coasts of Guadeloupe during the same period. The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

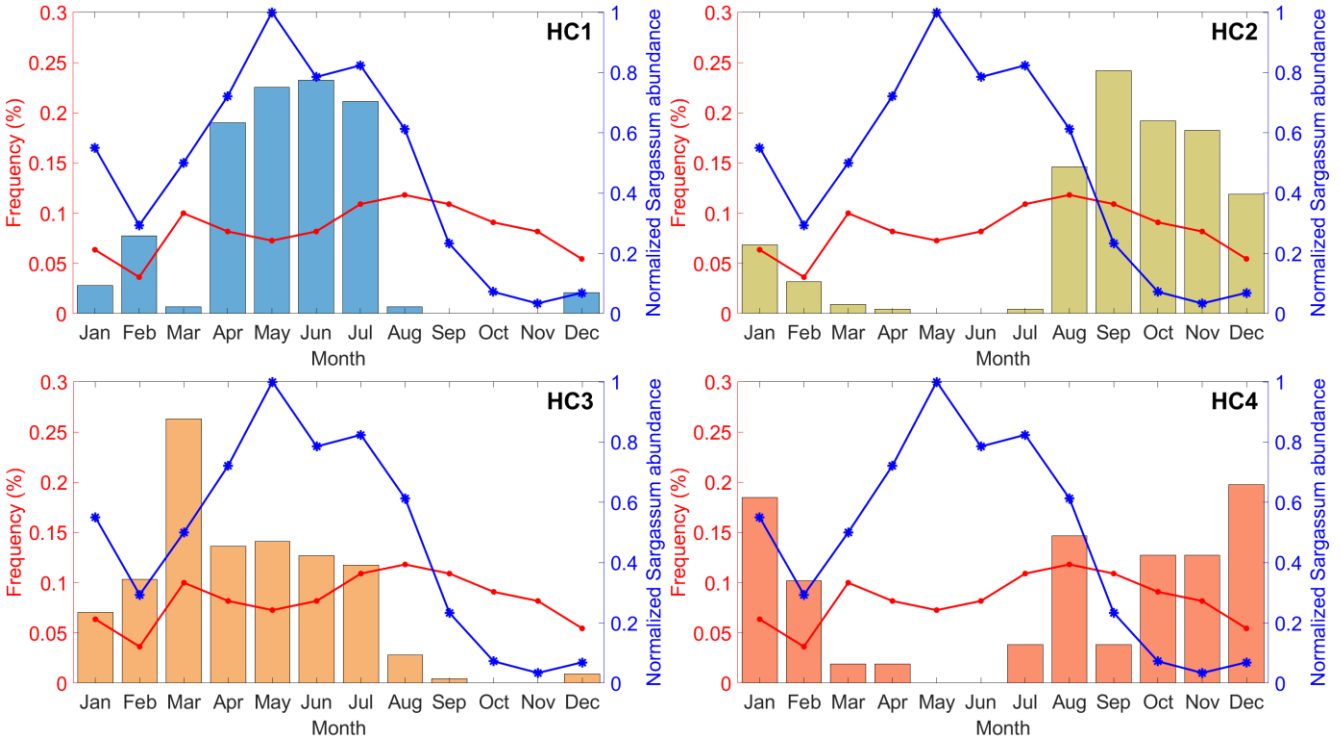


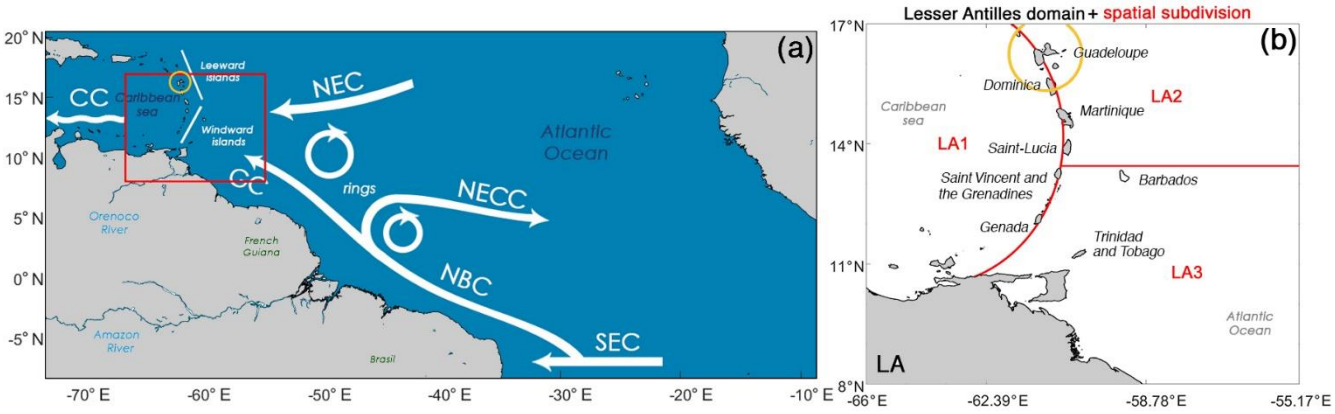
Figure 11: Monthly distribution of cluster occurrence from HYCOM outputs, from 2019 to 2020, in the Lesser Antilles (55-66°W, 8-17°N): HC1 (a), HC2 (b), HC3 (c) and HC4 (d). The red line shows the monthly distribution of Sargassum strandings on the coasts of Guadeloupe during the same period. The blue line indicates the monthly evolution of Sargassum abundance in the area 30-100 km offshore Guadeloupe normalized on the maximum value.

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."



"Figure 1: (a) Main oceanic currents occurring and interacting in the central Atlantic and the Lesser Antilles regions; Caribbean Current (CC), North Equatorial current (NEC), North Brazil current (NBC), North equatorial Counter Current (NECC), South Equatorial current (SEC). Lesser Antilles domain (LA): the red rectangle corresponds to the study area (55-66° W, 8-17° N); (b) Spatial subdivision of the study area into three sub-areas: LA1 (i.e., Caribbean Sea), LA2 (i.e., North Tropical Atlantic above Barbados (13.2° N)) and LA3 (i.e., North Tropical Atlantic below 13.2° N). The yellow circle corresponds to the 100 km offshore Guadeloupe area in which the satellite-based Sargassum abundance is analysed."

L276-281: Avoid redundancy.

Our answer:

To avoid redundancy the sentence Line 279 : *"The pairs (MC1, HC2) and (MC3, HC3) include the greatest number of observed stranding days in Guadeloupe (Table 4)."* will be deleted.

Discussion

L317 (Discussion) I suggest splitting this section. One for surface current and one for Sargassum stranding (L343-352).

Our answer: According to your recommendations, this section will be splitted. A subsection will be added for *Sargassum* stranding.

L334 North Current: You mean NEC?

Our answer: Yes we mean NEC.

The sentence L344: "*The last identified factor is related to surface currents present in the North Atlantic region due to the North Equatorial Current and the associated gyre circulation.*"

will be replaced by: *"The last identified factor is related to the North Atlantic Gyre and the associated the North Equatorial Current. As the seasons change from winter to summer, the gyre shifts South by a few degrees in latitude."*

L344 "Out of sync"?

Our answer:

The sentence *"The monthly distribution of clusters and the distribution of observed strandings in Guadeloupe are out of sync."* will be removed.

L345 remove "and

Our answer:

Following your suggestion *"and"* will be removed here.

L357 "independent variables": you mean explanatory

Our answer:

Yes we mean explanatory: *"independent variables"* will be replaced by *"explanatory variables".*

L372 "ocean current 3D models": better ocean current reanalysis.

Our answer:

We used analysis fields from HYCOM and Mercator models, not reanalysis. Following your suggestion, *"ocean current 3D models"* will be replaced by *"ocean current analysis".*

L404. This discussion of the appropriate time and space scale must be introduced earlier to justify the choices you made (30 days sequences, areas, monthly probability)

Our answer:

As mentioned at the beginning of the document, we will justify our choices of time and space scale (30 days sequences, areas, monthly probability) in the methods section.

Areas are already described in the "2.5.2 Use of Expert Deviation in clustering algorithms" methods subsection, Lines 135-140:

The LA study area was separated into three parts (Fig. 1b) based on the Sargassum rafts transport centers of action reported in the literature (Franks et al., 2016; Berline et al., 2020). To the west of LA, the first zone, LA1, is centered on the Caribbean Sea. To the east, the Atlantic zone has been split into two areas towards 13.5°N, just above Barbados island. To the south-east is the LA3 zone under the influence of the North Equatorial Recirculation Region (NERR) and its retroflection rings, while to the north-east is the LA2 zone, more representative of the North Equatorial Current. 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).

Concerning the 30 days duration, the following sentences will be added in the methods, L154 (Section 2.6):

"The 30 days duration corresponds to the empirical transport time of a passive particle moving from the main entrance location of Sargassum rafts in the Lesser Antilles area (i.e., in LA3 zone, 8°N; -55°E) to Guadeloupe (i.e., LA2 zone). Based on the mean current magnitude of 0.2 m s^{-1} (average value over the LA zone, in HYCOM and in Mercator data) and the distance of 500 km between the main entrance location and the Guadeloupe coasts, 29 days are obtained for the transport. For simplicity, the duration of 30 days was selected instead of 29 days."

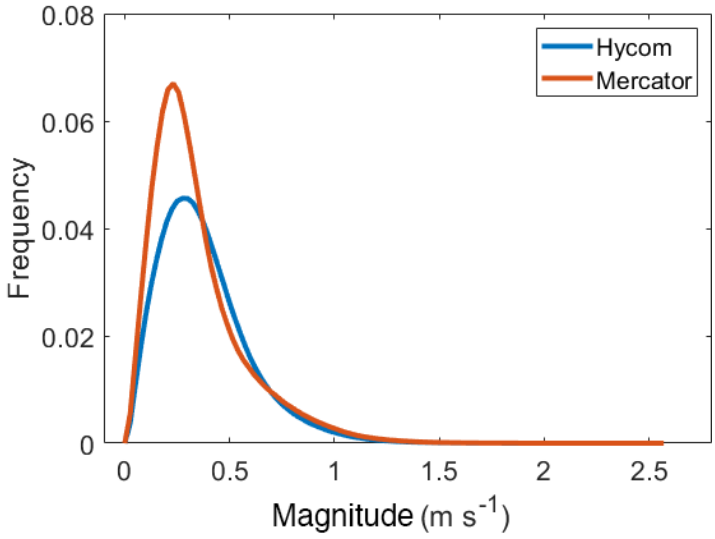
Concerning the monthly probability, the following sentences will be added in the methods, L170 (Section 2.7), in Module A explanation:

"The monthly period was selected to reduce possible biases associated with the uncertainties on the daily coastal observations. This sampling scale also allows comparisons with the monthly offshore Sargassum abundancy (SaWS, https://optics.marine.usf.edu/projects/SaWS.html) and with the monthly distribution of cluster occurrence."

Figures

Fig3 current magnitude

Our answer: This figure will be corrected, "velocity" will be replaced by "current magnitude" on the x-axis.



The caption of the Figure 3: *"Distributions of oceanic surface currents including windage for both models, HYCOM (blue) and Mercator (red) datasets"*

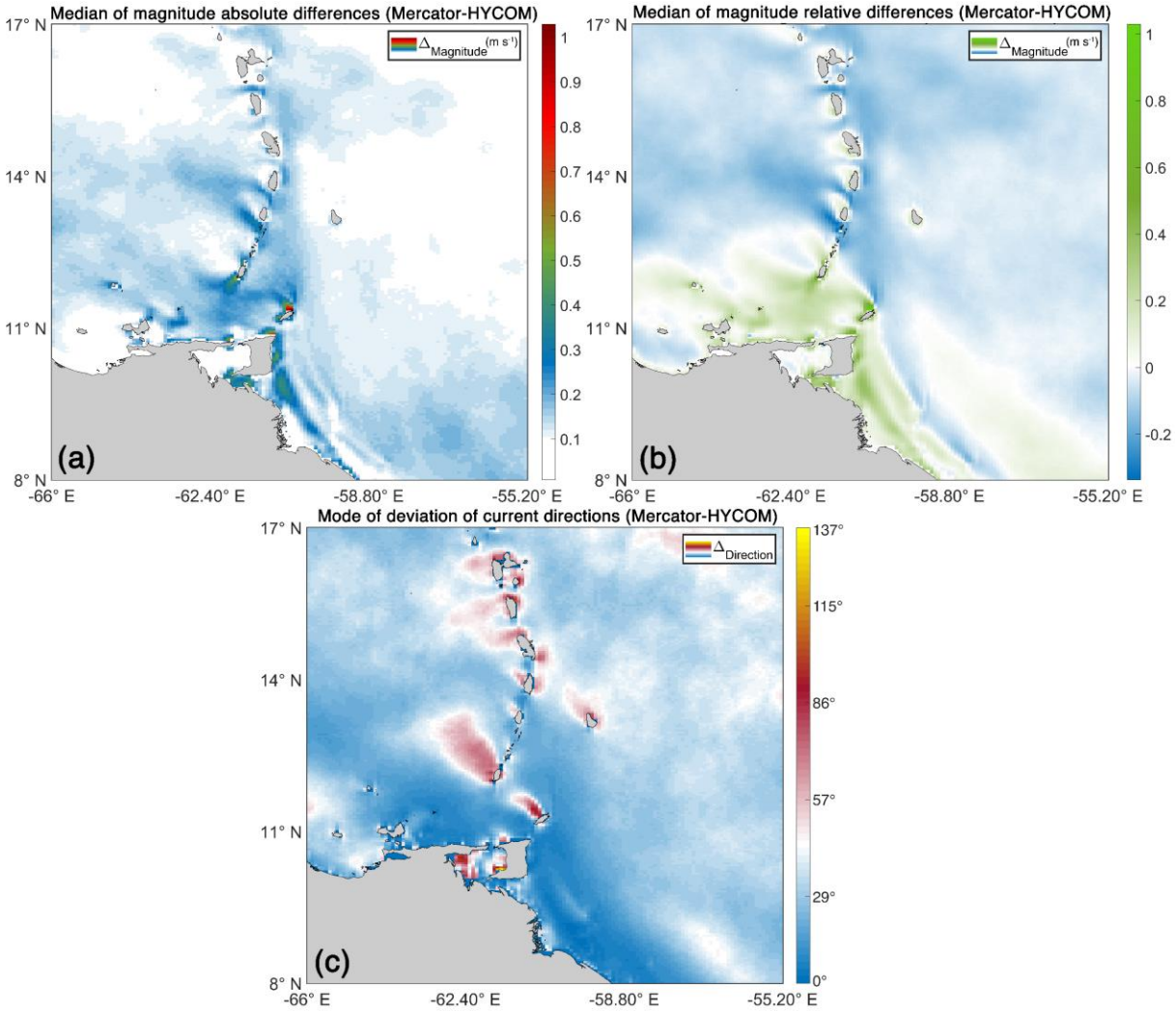
will be replaced by

"Distributions of oceanic surface current magnitudes including windage for both models, HYCOM (blue) and Mercator (red) datasets"

Fig5 Why not showing relative difference of magnitude, to see if Hycom is higher than Mercator for instance. What is the grid shown?

Our answer:

Firstly, the grid shown is the same 0.08° grid for Mercator and HYCOM. Following your remarks, the Figure 5 will be modified including three panels: (a) Median of magnitude absolute differences (Mercator-HYCOM) (b), Median of magnitude relative differences (Mercator-HYCOM) and (c) Mode of current direction differences (Mercator-HYCOM).



The caption of the Figure 5: *"Comparison between Mercator and HYCOM surface currents from 2019 to 2020: current direction median differences in degree (a), current intensity median differences in m s-1 (b)."*

Will be replaced by

"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."

Fig7 and 8. Mention Parangon as in the text. How is computed the stream function?

Our answer:

Paragon has also been replaced by parangon. The stream function is calculated from the u and v components on the lon lat grid.

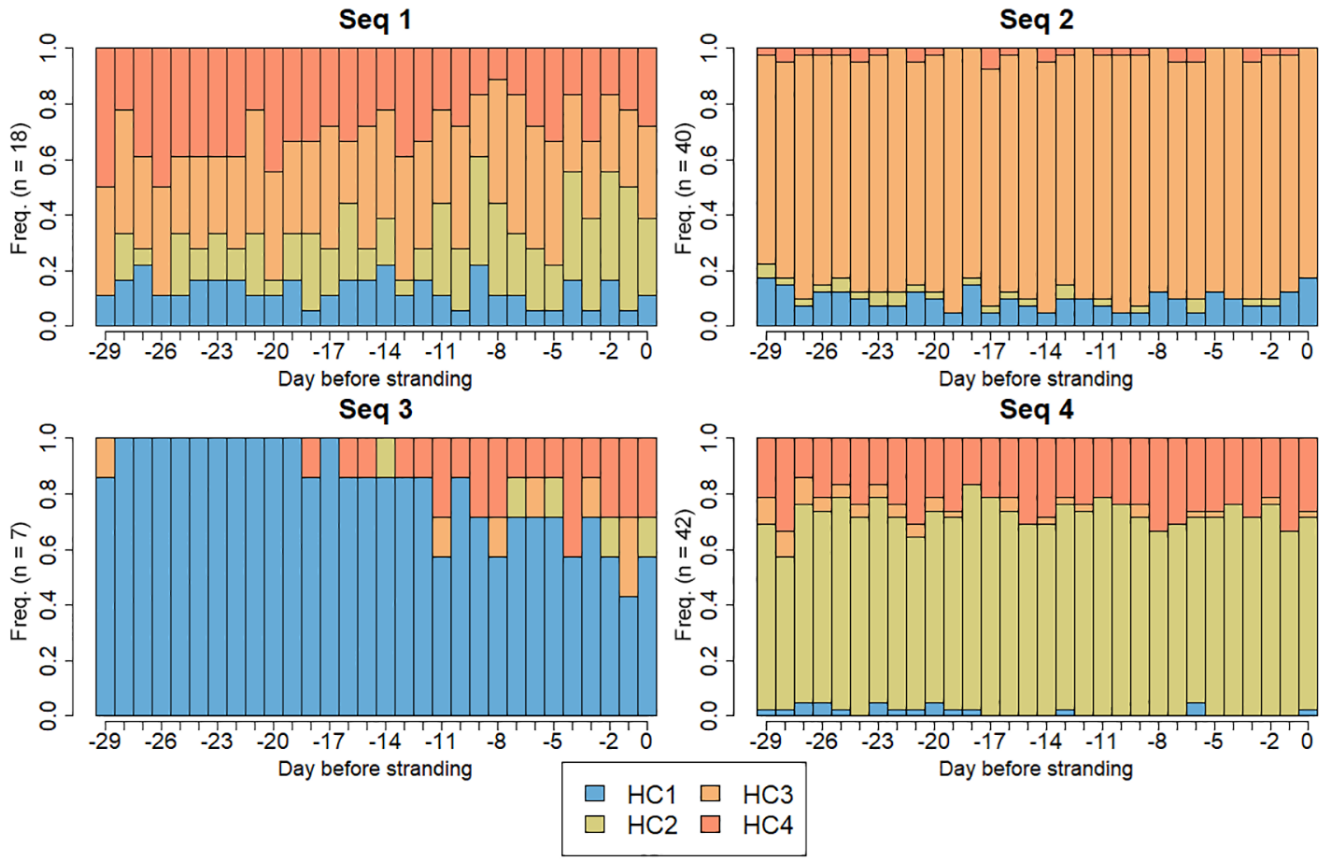
For all figures showing clusters, in the tables and text: for clarity, you should rename the clusters from Mercator and Hycom to make similar patterns match. As HC1 is consistent with MC2, rename MC2 into MC1, etc.

This similarity of patterns is expected given the similarity of data assimilated into the two models.

Our answer:

For greater clarity, we prefer keep this notation.

Fig13 These are clusters of sequences. Mention it. Use same color as in figs 11-12



Our answer:

The figure will be modified with same colors as figs 11-12.

The caption *"Figure 13: Distribution of current regimes over the 30-day stranding backward sequences: HC1 in green, HC2 in purple, HC3 in orange, and HC4 in yellow."*

Will be replaced by: *"Distribution of HYCOM current regime clusters (i.e., HC1 (in blue), HC2 (in green), HC3 (in orange), HC4 (in red)) in the 30-day sequence types (i.e., Seq1, Seq2, Seq3, Seq4)."*

Fig 15. Time index: What is the corresponding date?

Our answer: The corresponding dates will be added to the x-axis. The original figure will be modified with the new module A and the new testing period (i.e., all the year of 2021).

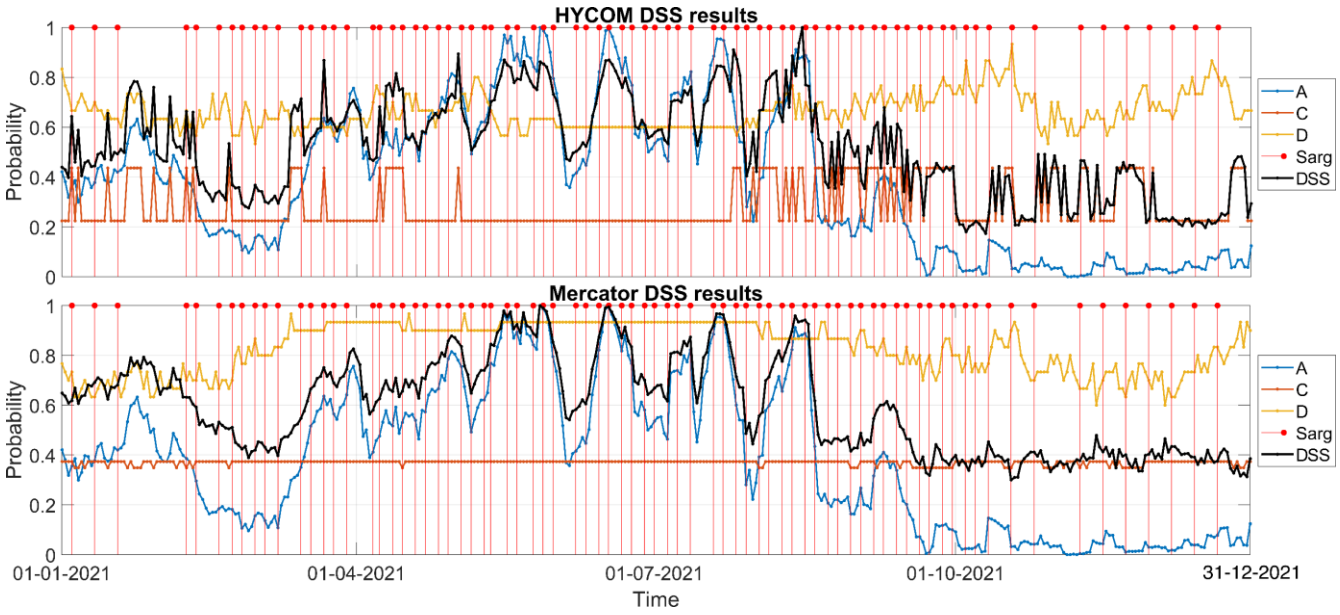


Figure 15: Decision Support System (DSS) results: probability of beaching obtained per module. Weekly probability to reach the maximum Sargassum abundance in the area 30-100 km offshore Guadeloupe for module A (blue line), stranding frequency per cluster for module C (orange line), match percentage for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

Table 6. Add recall.

Our answer: The old table will be replaced by the following table including recall and time uncertainty ranges (i.e., +/-1 days, +/-2 days,…).

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).