

## RC2 round 2

Please find below, in green, the answers to round 2 of revisions. If no answer is provided, it implies agreement with your response.

### 1. General comments

Thank you very much for adding the new schematics to better understand the processing you used in this study.

### 2. Specific comments

**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).**

The following sentences were added here L76:

“Predictive modelling refers to mathematical and computational methods of predicting future events based on the analysis of the repeatable 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).”

” Compared to other conventional forecast...” to “Compared to other conventional forecasts...”

“predictive modelling methods requiring low computational costs are characterized by their flexibility” to “predictive modelling methods require low computational costs **and?** are characterized by their flexibility” add “and” or only the ones with low computational cost?

**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?**

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).” was replaced L117 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..”

were replaced by:

### 2.1 HYCOM surface current dataset

Daily 12 UTC (i.e. Coordinated Universal Time) 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 were 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 12 UTC 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.

Section 2.1 suggested modification (in bold the specific changes made):

## 2.1 HYCOM surface current dataset

Daily (**12 UTC, i.e. Coordinated Universal Time**) surface current components from the 41-layer Hybrid Coordinate Ocean Model (HYCOM) **at 1/12-degree, global 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 were **first** 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.

Section 2.2 suggested modification (in bold the specific changes made, and underlined phrases to be revised):

## 2.2 Mercator surface current dataset

The daily (**12 UTC**) 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 **is** 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 **horizontal grid resolution**. 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 the Mercator than in the HYCOM fields**.

“Assimilated observational data types are quite similar to HYCOM model.” This phrase is not very clear, with the assimilated observational data types are you referring to the Mercator dataset? If so, please specify whether the HYCOM and Mercator datasets include data assimilation.

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

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:” was replaced (L133) by:

“The daily 12 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:”

The above phrase and section 2.3 suggested to be re-written for clarity in the following way:

“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  $C_w = 0.01$  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. The daily 12 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:

$$us(x, t) = um(x, t) + C_wuw(x, t) \quad (1)$$

where  $us$  represents the oceanic surface currents with windage,  $um$  the oceanic surface currents velocity,  $C_w$  the windage and  $uw$  the surface winds velocity. This approach is consistent with Putman et al. (2018) and Johns et al. (2020) studies.”

To:

“Surface wind influences the transport of floating seaweed rafts and a drag or windage coefficient must be added to the surface currents. Daily (12 UTC) from the **31-km horizontal resolution** ERA-5 model was used. The wind data was integrated with Mercator and HYCOM **ocean currents data** following this formula:

$$us(x, t) = um(x, t) + C_wuw(x, t) \quad (1)$$

where  $us$  represents the oceanic surface currents with windage,  $um$  the oceanic surface currents velocity,  $C_w$  the windage and  $uw$  the surface winds velocity. This approach is consistent with Putman et al. (2018) and Johns et al. (2020) studies. The value of  $C_w = 0.01$  was used, **following** 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.”

Information on the ERA5 data still missing: is it the reanalysis dataset? Any references of the dataset? Is the temporal resolution daily or hourly and just the 12 UTC fields used?

Lastly, please add here or in the Data availability section where this data was downloaded or obtained from. Also, this information is missing for the HYCOM and Mercator datasets (sections 2.1 and 2.2).

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

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)”

was modified like below (L154)

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

Please clarify phrasing, maybe split sentence in 2 or 3?

**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?**

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 Sargassum offshore 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.” was replaced by (L218)

The proposed tree in Fig. 5 was tested on the full year of 2021 except 31 December 2021 including missing data, i.e. in total 364 tests.

Phrase not clear, do you mean the 31<sup>st</sup> of December was not included because of missing data?

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

The following sentence was added L223 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”

“the lack of study dealing” to “the lack of studies dealing”

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

The sentence “For both models HYCOM and Mercator, the velocity intensities do not exceed 2.57 m s<sup>-1</sup> and 90% of them remain below 0.65 m s<sup>-1</sup>” was replaced by (L224)

“For both models HYCOM and Mercator, the maximum surface velocity is 2.57 m s<sup>-1</sup> and 90% of them remain below 0.65 m s<sup>-1</sup> (the respective 90th centile values are respectively 0.6515 m s<sup>-1</sup> and 0.6458 m s<sup>-1</sup> for HYCOM and Mercator).”

“90% of them remain” to “90% of the velocity values remain”

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

After the sentence “Figure 3 shows skewed distributions with skewness equal to 1.31 and 1.21.” The following sentence was added (L229): “A normal kernel was used to obtain these distributions.”

Thanks for the clarification, maybe to clarify further modify “A normal kernel was used to obtain these distributions.” to “A **Gaussian** kernel was **applied** to obtain these distributions.”

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

Firstly, the original figure was replaced by the following figure:

Figure 8: 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<sup>-1</sup> and (b) median of magnitude relative differences (Mercator-HYCOM) in m s<sup>-1</sup> and (c) mode of current direction differences (Mercator-HYCOM) in degree.

For subplot c), colormap a bit confusing as the white regions suggest not difference in direction, but those regions are actually different. Maybe the same colormap as subplot a) would be more appropriate?

**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.**

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.”

were replaced L304 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.”

I think my remark here was not clarified. Please re-check.

### **3. Technical corrections**

Some *Sargassum* in the Conclusions section forgot to put in italics.

**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."**

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."

Was 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."

L47-48: "current, and numerical models, both the roles of both subsurface nutrient supply and surface current transport were estimated." to "current, and numerical models; the roles of both subsurface nutrient supply and surface current transport were estimated."

**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?**

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."

Was replaced by (L140)

During this period of 730 days, only 110 days of observed beaching were 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.

"During this period of 730 days, only 110 days of observed beaching were recorded" to "During this period of 730 days, only 110 days of *Sargassum* beaching were recorded"

**L151: Define all variables of equation 2!**

This sentence was 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 from  $C_j$  and  $s(i)$  the silhouette index (Rousseeuw, 1987) value of day.

"...value of day." to "...value of day  $i$ ".

**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.**

“ For both models HYCOM and Mercator, the velocity magnitudes do not exceed 2.57 m s<sup>-1</sup> and 90% of them remain below 0.65 m s<sup>-1</sup>.”  
was replaced L224 by

“For both models HYCOM and Mercator, the maximum surface velocity is 2.57 m s<sup>-1</sup> and 90% of them remain below 0.65 m s<sup>-1</sup>”

For the maximum velocity of 2.57 m/s for both models, is it exactly the same too? If so, why does in figure 7 the x-axis go up to 2.57 m/s for Mercator, but only 2.49 m/s for HYCOM?

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

This part was modified with new results produced by our improved version of Decision Support System.

Please find below the modified Table 6 which includes recalling details.

**Table 6: Decision tree performance scores.**

Please add details to the table caption. True/ negative positive of... and recall % referring to... Accuracy of... and ratio between... .

Moreover, in section 3.5 you first introduce table 6, but further details are still missing. I assume that with true positive you mean beaching and true negative non-beaching days? Please specify this. Moreover, the values in table 6 for true positive is the number of days beaching is observed and the percentage, the percentage of days with beaching observed? The corresponding for non-beaching and true negative? Also, how is the accuracy calculated? And the ratio?

#### **4. Figures and tables**

Figure 2: “The schematic of the overall methodology.” to “A schematic showing the overall methodology.”

Figure 4: “The schematic of the clustering process on the current sequences leading to beachings.” to “The schematic of the clustering process **used** on the **ocean** current sequences leading to beachings.” See whether best to say used on, applied on, or other.