Comment on os-2021-109

Anonymous Referee #1

Referee comment on "Clustering analysis of the *Sargassum* transport process: application to beaching prediction in the Lesser Antilles" by Didier Bernard et al., Ocean Sci. Discuss., https://doi.org/10.5194/os-2021-109-RC1, 2021
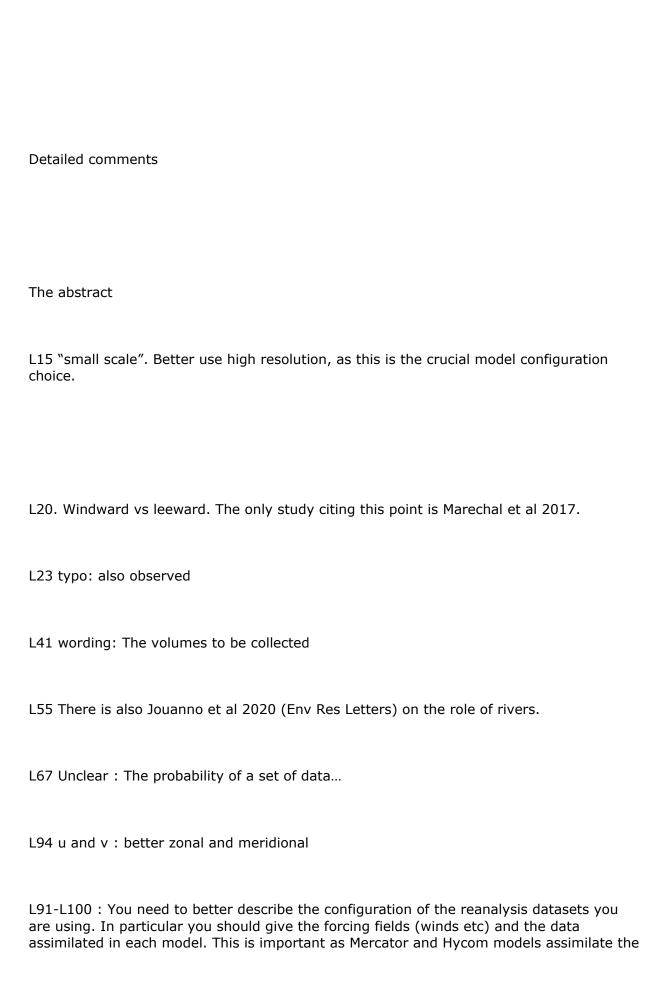

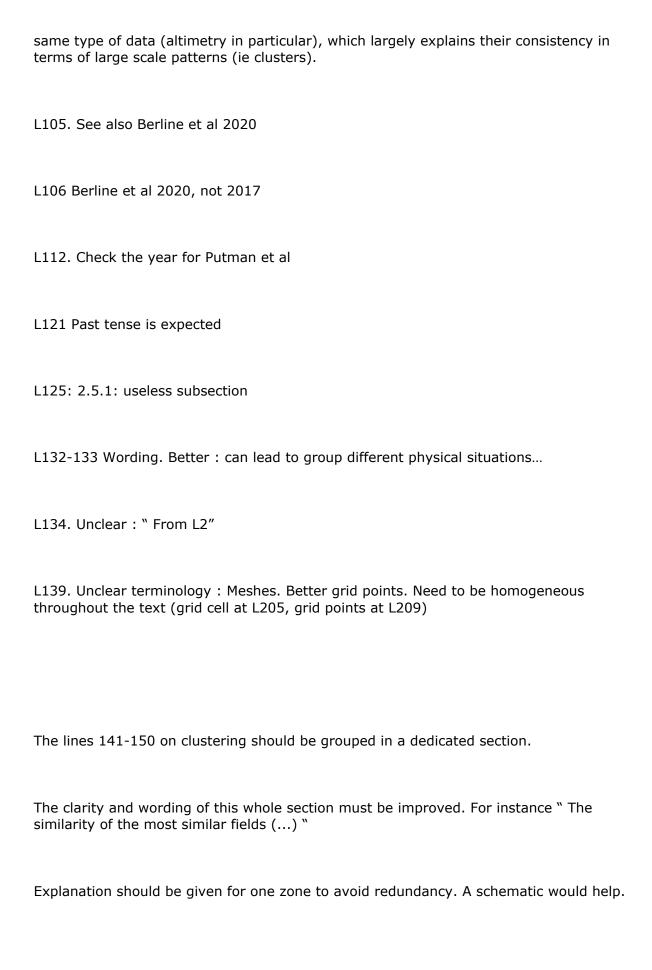Review of manuscript by Bernard D et al

This study deals with an analysis of the surface circulation in the vicinity of the Lesser Antilles in order to determine the situations favoring Sargassum stranding. It uses clustering of the surface current fields from two reanalysis datasets to identify recurring patterns. Based on these clusters and stranding observations, a decision tree classifier is built to forecast stranding probability. This is an original work addressing the drivers of stranding based on state of the art datasets, and dedicated methods. While the performance of the classifier is rather low, the topic and methodological approach is appealing and worth a publication.

The main remarks I have are

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).

2) Strandings occuring 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.

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.

Detailed comments

The abstract

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

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

L23 typo: also observed

L41 wording: The volumes to be collected

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

L67 Unclear : The probability of a set of data…

L94 u and v : better zonal and meridional

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).

L105. See also Berline et al 2020

L106 Berline et al 2020, not 2017

L112. Check the year for Putman et al

L121 Past tense is expected

L125: 2.5.1: useless subsection

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

L134. Unclear : " From L2"

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

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 ?

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?

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

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

L161 Wald's or Ward ?

L164 "At a given location" : which one ?

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

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

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

L191 Is it including windage ?

L191 "intensities ": better magnitude

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

L205 "At sea" : you mean offshore

L226 parangon

L241-242 should be in methods

L245 "most important" : better highest

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

L276-281 : Avoid redundancy.

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

L334 North Current: You mean NEC?

L344 "Out of sync"?

L345 remove "and"

L357 "independent variables": you mean explanatory

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

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)

Figures

Fig3 current magnitude

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

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

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.

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

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

Table 3. Use cluster names as in figures (MC1, HC1, ..)

Table 6. Add recall.