Reviewer 1

General comments

In this paper, the authors develop a statistical method for assessing CMIP5 climate model simulations of upwelling in the Senegalo-Mauritanian upwelling region and briefly discuss future projections of upwelling from a subset of the best-performing models. The method for assessing the models appears sound and seems to produce acceptable results in evaluating the models. However, I found the description of the method and its application difficult to follow at some points, as detailed in the specific comments below. There are also a number of typographical, grammatical, and organizational issues which impede the reader's ability to interpret the writing at some points.

I have provided some specific comments on some corrections needed in the technical corrections section, but this is not an exhaustive list and the authors should carefully proofread the paper prior to submitted any revised version. Finally, there seems to be a mismatch in the wind data discussed in the Data section versus the wind data used to produce Figure 10, as detailed in the specific comments below. All of these issues should be corrected before any subsequent version of the paper can be evaluated.

We would like to thank the reviewer for his/her careful reading of the manuscript and his/her relevant comments. We answered them carefully, and we believe the manuscript has definitely improved. We apologize for the typos and grammatical issues underlined by the reviewer. We have carefully revised the whole manuscript to improve this point specifically. We detail below ll the modifications that have been implemented in the text.

Specific comments

Lines 22-23: Give a brief summary of the main findings on the future behavior of the Senegalo-Mauritanian upwelling in the abstract, rather than simply stating that the future behavior was assessed.

This sentence was changed into "The future reduction of the Senegalo-Mauritanian upwelling proposed in recent studies is then revisited using this multi-model selection."

Lines 106-107: More detail is needed on the ERSST_v3b data set. How is this dataset produced, and why was it chosen as the "observation field" for comparison with the Models?

Some details on the ERSST_v3b were added. We also explained that other data sets would have been available, in particular HadISST. A previous study has shown that differences in the Senegal-Mauritania upwelling are relatively weak. Yet, sensitivity of the method to the target field should definitely be addressed in the future, we thank the reviewer for this remark.

Lines 112-114: Similar to the ERSST_v3b data set, please provide some additional detail on the QUICKSCAT (sic) product. And is this product actually used in the paper?

In the caption for figure 10, the TropFlux data set is referenced rather than QuikSCAT in the discussion of the wind stress and Ekman transport (see comment on line 826), and there is no mention of QuickSCAT or TropFlux in the discussion of this figure in section 5.1. Additionally, “QuikSCAT” is the correct spelling of this satellite.
We apologize for the typo on the name of the satellite. We also apologize for the fact that QuickSCAT data is indeed not used in the final version of the paper. The wind stress was evaluated against the TropFlux reanalysis. Information about this product was added (end of section 2.1)

Line 118: Does “Sylla et al. (in rev.)” refer to the Sylla et al. 2019 Climate Dynamics paper, or another work? If this is another paper, it should be added to the reference List.
Sylla et al. (in rev.) and Sylla et al (2019) are indeed the same paper, accepted in the course of the preparation of the present manuscript. The citations have been homogenized, we apologize for this.

Section 3.2 (starting on line 168): It’s not clear to me why the SOM classification followed by the HAC clustering was necessary. What is the reason for performing both classifications rather than just using one method or the other?

The SOM model has been used to determine a vector quantization of the dataset: i.e. to determine referent vectors that are a representative summary of the learning dataset. The vector quantization compresses the total database into a quite small (with respect to the size of the database) number of referent vectors such as each data is not too different of its nearest referent according to a distance (The Euclidean distance in the present case). The exact number of referents, that is the number of neurons, does not really matter because this number will be reduced by the HAC. Doing so allows us to take the non-linearities of the dataset into account in the analysis. The exact number at the end of the SOM+HAC procedure is not known a priori but at the end of the study by looking at the HAC dendrogram, which suggests several possibilities for the number of classes to estimate. A compromise between the number of classes we can explain from a physical point of view and the number we need to include the information embedded in the dataset is made. This procedure has been used with success in several papers (Jouini et al, 2016, JGR; Farikou et al, 2015, JGR; Sawadogo et al, 2009, IEEE; Niang et al, 2003, RSE).

We have added an introduction at the beginning of section 3 to explain this point. Note that it was also raised by reviewer 2. We are grateful to both reviewers for requiring this classification.

Lines 197-198: What is a “standard statistic algorithm”? Is this referring to the calculation of the standard deviation?
It refers indeed to the calculation of the spread in each neuron. We have re-phrased this sentence as:
“for each region-cluster, we estimated the monthly mean of the SST seasonal cycles and the associated spread captured by the neurons constituting this region-cluster” . Thank you for this clarification.

Section 3.4 (lines 255-330): Perhaps it is just my own ignorance, but I find figure 4 and the accompanying discussion quite difficult to interpret. What, conceptually, do the x and y axes and the grouping of the points on the plot represent? The description says that "proximity between a model and a region-cluster leads us to affirm that this region-cluster is well represented by that model", but the observations and the "highest skill" models 7 and 25 are far away from any of the region clusters. . .? And I think I understand that model 7 is considered as having good skill because it lies close to the "obs" point on the plot, but why is model 25 considered to have better skill than models 24, 19, 8, and 40, which are located a similar distance from the "obs" point as model 7? Have any previous studies used this method to assess the skill of models?

Sub-section 3.4 is now section 4. In fact the MCA method used in this section is different from the SOM and quite new in geophysics; it therefore deserves a dedicated section in which we give more details on the functioning of the MCA. We have rewritten the presentation of the MCA analysis in order to get an easier understanding of the functioning of this analysis and to facilitate the understanding of Figure 4. The new writing of the beginning of the MCA presentation (section 4) is shown below. Our changes are in yellow in this new writing.

"In order to further progress in the selection of the models, the 47 climate models and the Observation field were then analyzed by using a Multiple Correspondence Analysis (MCA in the following). MCA is a multivariate statistical technique that is conceptually similar to principal component analysis (PCA in the following), but applies to categorical rather than continuous data. Similarly as PCA, it provides a way of displaying a set of data in a two-dimensional graphical form.

In the following, we applied a MCA analysis to the (47 , 7) matrix **R** = [Rm*i*] whose elements represent the skills of the clusters of the models shown in front of the color bars in Fig. 3: the rows m represent the 47 different models, the columns *i* the 7 region-clusters. The MCA, as the PCA does, projects the initial matrix in a new basis in such a way that the new axes are the matrix eigenvectors (PC), the inertia of each axe being the related eigenvalues. According to the theory, the MCA matrix analysis gives 6=(7-1) independent PCs. Each model is now associated with a 6-dimensional vector. The MCA uses for this analysis the khi 2 distance. In figure 4, we present the projection of the models and the "region clusters" in the plane formed by the two first axes (x=PC1 and y= PC2) of the MCA that represent 70 % of the total inertia. Each model is represented by a small circle and each Region-cluster by a purple square. Moreover, we projected the observation field (green diamond) on that plane as a supplementary individual. The proximities in figure 4 is represented by the khi2 distance. To have a more precise view, it should be necessary to consider the projection on the 5 other PCs which represent 30% of the inertia.

In the (PC1, PC2) plane, the shorter the distance between two models, the more similar the distribution of their region-cluster skills. The seven clusters of the observation field are represented by purple squares. Proximity between a model and a region-cluster leads us to affirm that this region-cluster is well represented by that model. ..;

…….

……;

In this above analysis, we must be aware that we are confronted to the intrinsic difficulty to represent multidimensional data in a plan. The representation of some data can be biased thank to the importance given to the other axes."

Line 826: What is the TropFlux data set? This needs to be described and its use justified in the Data section.
The TropFlux reanalysis combines the ERA-Interim reanalysis for turbulent and long-wave fluxes, and ISCCP (International Satellite Cloud Climatology Project) surface radiation data for shortwave fluxes. This wind stress product is described and evaluated in Praveen Kumar et al. (2011).
These lines were added in the Data section.

Technical corrections
The writing in this paper is frequently conversational in tone rather than technical, and there are many instances of imprecise filler words like "very", "pretty", and "nicely".
In line 250, "let us say. . ." is a conversational phrase that is not appropriate to use in a scientific manuscript in this context.
Also, the mention of "ongoing studies in our group" (line 527) is fine for a conference presentation but, in my opinion, is not appropriate to write in a scientific paper.
Please proofread the paper and correct these and other such instances of informal language.
The text was largely revised and improved in this respect. The two specific sentences cited above were corrected. We thank the reviewer for this remark that led us to significantly improve the language of the manuscript.

There are several excessively long paragraphs that are taxing on the reader. For example, the paragraph from lines 30-73 in the Introduction and the paragraph from lines 332-377 in section 4 are very difficult to read due to their length. Please break up and reorganize these and other long paragraphs.
These two paragraphs and several others have been cut. Thanks also for this remark that improves the readability of the paper.

There are a number of typographical and grammatical errors throughout the paper, which impede its interpretability to the reader in some places. I have given a few examples below, but this is not an exhaustive list, and the authors should check the entire paper carefully for such errors in any subsequent versions of the manuscript.

- Title: Extra space after the word "in"
Corrected

- I am not able to read the full "short summary" on the discussion paper web site, but the part I can see contains three misspelled words.

We are not sure why the short summary could not be read from the website, but the short summary was corrected

- Errors in capitalization of words (e.g. “Observation field” in line 108, seasonal “Cycle” in line 299)
These capitalizations intended to put emphasis on the elements of the mathematical procedure. Yet, we realize that this was not clear so that we removed all the capital letters in these expressions.

- Lines 16-18: Abrupt shift from first-person to third person (“We used a neural classifier. . .” to “One can then determine. . .”)
corrected

- Line 83: Typo (“lies at is the southern. . .”)
corrected

- Lines 109-110: Typo (“were been regridded”)
corrected

- Line 525: Typo (“costal” instead of “coastal”)
corrected

Line 16: Typically self-organizing maps are described as an “artificial neural network” rather than a “neural classifier”.
This was changed: In the text we now use a artificial neural network (Self-Organizing Maps),

Line 21: What is meant by the phrase “performing multi-model ensemble”?
This phrase was changed to “an efficient multi-model combination of 12 climate models”

Line 26: CMIP5 stands for the “Coupled Model Intercomparison Project, Phase 5” (not the “5th Climate Model Intercomparison Project”).
corrected

Line 383: Shouldn’t this say “(Fig. 8, left)”?
Corrected, apologizes for this mistake

Line 755: Figure 1 appears distorted and blurry in the PDF version of the paper. Please correct this figure to make it easier to read.
This figure was corrected and the .eps version attached to the revisions is not blurred.