Review of Latto and Romanou: The "Ocean Carbon States" Database: a proof-of-concept application of cluster analysis in the ocean carbon cycle:

Summary: This study presents a methodology for the analysis of model output against observations of CO_2 fluxes. The methodology is based on a "data-mining" technique, where both observations and model output are subjected to a cluster analysis, which identifies the main seasonal regimes of variability of sea surface temperature (SST) and surface water pCO_2 in the North Atlantic, and Southern Ocean. The study is focused on the evaluation of uncertainties related to the ocean component driving the CO_2 flux in the model.

Specifically, the analysis is initially carried out on a monthly climatology of sea surface temperature (SST) and surface water pCO_2 observations, in order to identify the number of clusters (K) that best represent variations in the data. This analysis is repeated with model output. Subsequently, model and observations maps of CO_2 flux for each regimen/cluster (K) are compared in order to discern regions of disagreement. Finally, the authors identify sources of bias in the model by evaluating the physicochemical factors that contribute the most to the disagreement between model and observations.

Major scientific comments: This study test an interesting way of assessing uncertainties in climate models. A key step in the analyses is the determination of the number of clusters K, which defines the main regimes within observations and model output. My main concern here is the apparent subjectivity is this determination, described in "**Sensitivity to predefined number of clusters"**. The optimal number of clusters is determined *"…by identifying the K with the highest score and no significant change of the score thereafter."* (page 7, line 21). This is based on a visual inspection of Fig.2b, 10b, S3b, and S10b. Regarding the case for the North Atlantic, the authors mention that the optimal K number is 3 for both observations and model output. Based alone on the visual assessment of Fig. S3b, I am inclined to say that 5 is a more appropriate number of K for model outputs. How would this discrepancy affect your results? Furthermore, it seems convenient that in both cases, North Atlantic and Southern Ocean, the optimal number of K in model and observations is the same, how would you compare model and observations if they had a different number of K?

Overall, these data-mining technique should provide a more objective method for model assessment. However, the subjectivity involved in the determination of the number of K, seems to defeat this purpose.

General comments:

- Abstract: This is a problematic section. There is a lot of emphasis on the utility of data-mining techniques and little or none mention of the biogeochemical findings regarding model biases in the North Atlantic or the Southern Ocean. It is not clear what are the main findings/conclusions of the study.

- Methods, Section 3.1.1: It took me several reading attempts to understand the way in which the K-means clustering works. It may be useful to include a diagram or include equations describing the iterative clustering process. As mentioned above, the method used to determine the final number of K clusters needs to be improved, or at least, better justified.

- Conclusions: *"A method is provided for an objective way to accurately determine the optimal number of clusters for the cluster analysis."* I highly disagree with this statement based on my **Major scientific comments**, above.

- The discussion of the attribution of model errors is well written. However, the conclusions are compromised by the subjective determination of the number of K.

- Figures. The distribution of figures between the main manuscript and the supplementary material is confusing. I recommend merging the frequency histograms in figures S1 and S8 in one figure, and including it in the main manuscript. In general figures are duplicated for the North Atlantic and Southern Ocean. With the exception of the 2D histograms, I recommend merging some of the similar figures for different regions (e.g., Fig 2 and Fig 10.), and also include in the main manuscript the analysis of the Southern Ocean (Fig S9, S10, S11, and S13). Fig S2 is not clear. It seems that some of the lines in the label are not included in the figure (e.g., K=2 and K=9 are both dark blue, but there is only one black blue line in the figure, and they would be impossible to distinguish). Fig 2a, 10a, 3a, and 10a: For more clarity, it would help to label the month corresponding to each bar.

Specific comments:

Page 10, Line 21: *"As shown in Fig. S5, we identify a subpolar region…"* What is the basis for the determination of these sub-regions in the North Atlantic and Southern Ocean? If they are arbitrarily chosen, "*we define*", would be a more appropriate wording. The criteria used to define these regions is important in the final attribution of processes driving the errors between models and observations.

Page 11, Line 25: "*Here, nitrate biases are probably due to misrepresentation of nitrogen fixation in the GISS climate model…*" In oligotrophic regions, relative errors in nitrate may be higher as the absolute concentration is overall lower, with much less seasonal amplitude when compared with higher latitudes. I agree that nitrogen fixation is likely a problem, but it might be useful to consider this aspect as well (see Arteaga et al, 2015, GRL, doi:10.1002/2014GL062937).

Page 12, Line 29: "*For almost all regimes and regions, biases in nitrate are large partly because of lack of a closed, state-of-the art nitrogen cycle representation in the climate model. However, observations are too scarce in the region, due to inclement weather and biases to specific seasons, so the model skill would be more adequately assessed as more in situ measurements are made (e.g. from the SOCCOM experiment; Johnson et al., 2017).*" This is also problematic regarding observations of CO_2 flux, particularly in winter. How do you account for a possible bias towards summer fluxes due to a lower amount of observations during winter months?

Final remarks: Overall, I find this study interesting and important for the community. However, further clarity is needed in the explanation and justification of the methods involved. My impression is that this study could be considered for publication after major revisions.