Classification of synoptic circulation patterns with a two-stage clustering algorithm using the modified structural similarity index metric (SSIM)

Overview

This manuscript shows the development of a two-stage clustering algorithm that combines two classical clustering algorithms, hierarchical agglomerative clustering (HAC) and k-medoids while utilizing structural similarity index metric (SSIM) over traditional distance metrics by using the 500-hPa geopotential height from ERA-Interim reanalysis data. This approach seems to offer many benefits over traditional classification methods, and I can see it being useful for a wide range of applications within atmospheric and climate science. While the manuscript is very extensive in describing the developed algorithm and showing its robustness with reanalysis data, I have a couple of general comments and several specific comments that I believe should be addressed. I believe this paper fits well into the scope of Earth System Dynamics and I would approve this work for publication pending major revisions made to address the comments attached.

General Comments

1. While the manuscript is very detailed in explaining and testing the methodology that was developed to classify synoptic circulations the connection and application of the method to the main motivation for its development, "to extend the evaluation routine for climate simulations", is not given the same amount of detail and attention as it should. The manuscript as is should more clearly demonstrate how the method accomplishes this objective and how it adds value to the current evaluation of climate simulations that would warrant the effort required to implement it. One possible suggestion, given the length of the paper and detail provided to the actual methodology and its testing, could be to make the application of evaluating CMIP6 simulations with this algorithm as a separate manuscript where that specific application of the method can be discussed and demonstrated in a complete manner.

2. While there is a good discussion in the introduction with respects to building synoptic classes in pervious work there was no mention of works that used approaches such as Machine Learning and AI which is becoming more popular within Earth system science as well as other fields. For example, Gervais et al. (2016) uses Self-Organizing Maps to classify Artic Air Masses from CESM-LE. I think it would be important to discuss how approaches like SOMs, Random Forrest, etc. have been used in the classification of synoptic patterns and how this new approach compares to them.

Specific Comments

LINE 20 – Why not state what the alternative reanalysis is instead of keeping it vague by just saying "alternative reanalysis"?

LINE 175 – Would this method also work if considering more than one atmospheric variable mapped on the same domain, or can it only work with the use of a single variable?

LINE 175 – For this work, one time step a day was used, is the reason for this due to computational/time constraints or are there other issues that may arise using this method with more regular time steps, such as all timesteps in ERA-Interim or even if moving to the hourly timesteps in ERA-5. If there are restrictions associated with the method and temporal/spatial resolution of data that can be used it would be good to mention them at some point.

LINE 195 – Its not clear why NCEP1 was chosen as the alternative reanalysis compared to other available reanalysis datasets. Why would the assumption "*Assuming that the alternative reanalysis captures the synoptic circulation of the reference data ERA-Interim better than any unconstrained global circulation mode*" be made? Can more be said about this decision?

LINE 198 – I am assuming all datasets are normalized with EQ. 1? Is this correct?

LINE 375 – I'm not sure this is clear, is the "final cluster" what is used as the initialization clusters, or the final result of the entire method being presented in the manuscript?

LINE 444 – When stating "well separated …from the entire data set" does this mean the clusters should be well separated from the data that is not assigned to the given cluster?

LINE 451 – Are these "similarity diagrams" what is shown in Figure 10?

LINE 461 – If it has been established that using values such as Euclidean distance does not perform well when considering things such as synoptic patters what is the value in calculating Metric 2?

LINE 585 – To clarify, there are 183 "runs" but each run is done for varying data volumes from 1 to 40 years. So, is it correct to say the method is done 183 x 40 times? Or the output of each run is just saved after each year of data is added?

LINE 660 – It is difficult to see the dashed and grey lines in Figure 10.

LINE 805 – While I understand the reasoning for showing the 5 most frequent SP-classes one of the benefits mentioned was the ability for the algorithm to preserve less frequent patterns that are more likely to be associated with extremes. I think it is important to demonstrate this ability/benefit. I would suggest maybe showing a couple of these patterns instead of just focusing on the most frequent SP-classes.