Manuscript by Doan et. al. presents a S k-means clustering framework, improving on standard k-means clustering, and demonstrate their application to several climate datasets.

Manuscript presents a methods focused study, which however lacks sufficient discussion to demonstrate the benefits of the proposed algorithmic improvements to standard k-means algorithm. Section "Results and Discussions" focus more on Results and less on Discussion, which is the critical weakness of the manuscript in its current form.

1. Manuscript is missing several key references from the reference list. Wang et. al. 2004, Wang and Bovik, 2009 Mo et al., 2014; Han and Szunyogh, 2018; Doan et al., 2021

2. One of the motivation for the proposed work, as discussed in introduction, is to mine the unique "structuredness" of temporal and spatial climate data (Line 67-81). However, rest of the manuscript focused on comparison of various clustering methods based on Silhouette scores, uncertainty degree etc. Proposed S k-means consistently shows better scores than the other methods, but if and how it better captures the "structuredness" of the data need to be discussed, since that's the key contribution of the study.

3. Structural similarity metric (Section 2.2) is the most important part of the study. However, several symbols/terms in equations 2, 3 and on lines 142-145 are not defined or explained. In particular the equations for luminance, contrast and structure. And the cited articles (Wang et. al. 2004, Wang and Bovik, 2009) that developed the similarity metrics are missing from the reference list. That makes it difficult to understand the similarity metric. Aside from describing equations for S-SIM, there are discussions, in methods section or later, as to how these structural metrics capture the spatial and temporal structuredness of climate data.
4. Discussion of clustering results in Section 5.2 is very high level. Question remains, aside from slightly higher scores what unique and new insights does the S k-means clustering enabled?

5. I am glad to see S k-means being compared with three other k-means variants. They were all run for a 11 different 'k' and with 10 random ensembles each, resulting in a total of 1320 clustering runs. BUT were all four k-means variants run with exactly the same random starting centroids for the purpose of comparison? It's important to do that for a fair comparison. Also, was a consistent convergence criteria used for all four methods? Convergence criteria was mentioned on Lines 128-129, but what criteria was used in the study never discussed.

6. Lines 364-365 "As the first study to address this issue, we believe that CUEF can constitute a new standard for addressing uncertainty issues when performing data clustering in (but not limited to) climate science." -- This is an overstatement. It's well know that clustering algorithms are local search methods that are sensitive to random start, however, there are number of approaches in published literature to identify good seeds and ensure that algorithms can converge to a consistent cluster set.

7. Lines 370-374: "This makes sense because different data have different topologies, which can make them unsuitable or even invalid for a clustering solution. The question of whether it is valid or meaningful to apply a clustering solution to a dataset is more important than how to find the best method of clustering. Although this issue is fundamentally important, to the authors' best knowledge, no studies have addressed this question or proposed a solution, at least among the climate sciences." -- this again is broad and biased inference based on the demonstrated applications and results.

8. Authors have termed their clustering framework to be novel, including in the title of the manuscript, which in my opinion is overstated and not justified. There are three key methodology elements in the paper + application to three select climate datasets.

Application component of study is weak and limited in scope. But author's acknowledge that application/interpretation was not the focus of their study, Lines 277-278 "We do not intend to physically interpret the specific clustering outcomes, although some phenomenal explanations are provided in the manuscript." So novelty is not in the three applications.

Three elements of methodology are adopted from published literature:

3. Clustering uncertainty degree and information theory (Vinh et al. (2009))
Building upon published literature is normal discourse of scientific research. But I suggest reconsidering the use of term "novel".