

## ***Interactive comment on “Multiresolution analysis (MRA) classification of plurennial to multi-decadal climate drivers to streamflow in France using Wavelet Transform and Geostatistical Euclidean Distance Clustering” by Manuel Fossa et al.***

**Anonymous Referee #1**

Received and published: 18 January 2017

### **Overall Review and Recommendation**

This paper introduces a method of correlating multi-year streamflow anomalies to atmospheric anomalies and clustering stream gauge sites into groups with similar drivers. This is done using wavelet decomposition to extract 6, 10, and 21 year anomalies from streamflow and climate. Then, correlation fields between each site's flow anomaly and gridded atmospheric anomalies are determined. These correlation fields are ultimately used as input to a novel clustering algorithm that weights clusters based on distance.

C1

Although I believe there may be value in this approach, I recommend rejecting this paper in its current state. There are severe problems with the paper's organization and because of this, it is unclear what is actually being proposed. The method, which I believe is the main objective of the paper, is poorly described and therefore it would be impossible to replicate the approach elsewhere based on the description. The writing is uneven, with excessive details in sections that are less important (Background on wavelets and Discussion regarding cluster centroids) and a lack of detail in critical sections (Methods and Results).

Additionally, there is no clear problem statement, objective, or hypothesis that is being tested. If this paper is meant to present a new approach (i.e. the new clustering algorithm), it would be important to present the methods in great detail and to test how the results differ from more traditional approaches. If this study is instead meant to explore the drivers of streamflow in France, it must do this in much greater detail, comparing results to other studies and attempting to provide a rationale for these drivers.

### **Major Comments**

1. Please provide a more explicit problem statement or hypothesis. Without this, the purpose of the paper is unclear. Is there an important question specific to France about the drivers of 6-21 year discharge patterns? If this is the purpose, there should be more interpretation of the climate drivers. Is the purpose to highlight the new clustering method? If so, there could be a comparison with other clustering methods. Or, is the paper a test of whether, as said in Lines 65-66, there are differences in clustering due to TSV and climate drivers. It is possible to address several of these in a single paper, but they must be explicitly stated at the start and the paper must be framed around them.
2. There is an overall lack of citations.

C2

When referring to methods and terminology, there must be references so the reader can follow what is being done. For example, the AHC method (line 147) or silhouette width should include references. Also, I have not seen the terminology TSV used previously. I could understand the meaning from context, but if this term is used elsewhere, please provide a citation. Similarly, the term “correlation climate field” is referred to as “so-called”, meaning it is a common method used elsewhere and should therefore include a reference.

Several areas of background information are not properly cited. For example, on Lines 56-58, if something is “in the literature”, you must provide examples of this. Similarly, in Lines 125-127, if a concept is “well known” it must still be supported with citations.

3. The overview of the Methods is unclear. I was ultimately able to understand through several readings, but there are still gaps. For example, if Fig 2, sub-Fig 2 shows point-wise correlation between a single gauge and geopotential height, how is this different from Fig 3, which also shows correlation for each of the 152 gauges? This particular step is covered in only one Methods sentence (117-118) and it does not explain the process.

4. The Methods section on wavelets provides great detail (1.5 pages) on theory, but it does not provide clear detail on methodological decisions. For example, why are these specific time scales chosen (Lines 113-114)? I don't see any unifying pattern to choosing 3, 7, 12 months and 1.5, 4, 6, 10, 21 years. These decisions are important, whereas some of the wavelet material could be covered with citations, especially because this is not the main purpose of the paper.

5. Another unclear methodological issue is the choice to only consider winter months (Line 57). This is not described in the Methods section. At what point in the process is the winter subset created? Is wavelet decomposition applied to the full time series and then correlation calculated only for the winter? Or is the winter subset extracted before decomposition? I don't think the latter is possible, but please be clear about this.

### C3

6. Please describe why clustering was performed for each TSV and climate field separately (resulting in 12 separate clustering schemes). Is this done to test the inter- / intra-driver comparison? If so, this should be stated explicitly as an objective or hypothesis at the start of the study.

7. The Results section provides few details. Please provide some detail regarding clustering fit or decisions in Section 4.1. Could silhouette results be presented here? Are there any other metrics that could be used to measure the cluster fit? Perhaps correlation/covariance in streamflow among the clusters or some form of variance explained. Additionally, if this paper is focused on the meaning of clusters, Section 4.1 must provide enough detail that the results can be interpreted in the Discussion section.

8. Clusters in Fig 3 relate to the fields in Fig 4, but it is impossible to link the two because the clusters are not labeled in Fig. 3. You might say C1 is always black, C2 is always orange, etc. This is particularly confusing in sentences like Line 191, where it is unclear which clusters in Fig. 4 correspond to the southern stations.

9. The difference between clusters over the African continent is an interesting result (Line 195), but this and other findings from Section 4.2, are not discussed in the Discussion section. Has this correlation been noted in other studies? What atmospheric process could link geopotential height over north Africa to streamflow anomalies in France?

10. There is almost a full page (Lines 226-251) regarding how to calculate a representative centroid for each cluster. This is certainly an interesting topic, but it is never discussed in the paper before this point. This has little bearing on the method or the results and should be shortened (or turned into a paper of its own).

11. There is a very strong conclusion statement on Line 336 “geopotential are the source for most of the variation of streamflow in France ...”. This would be an interesting conclusion; however, this was not introduced as an objective for the paper, nor did

### C4

I find any results presented that support this conclusion.

### **Minor Comments**

1. There are extra right parentheses on citations throughout the paper.
2. Line 54: What does “when applicable” mean in this context? When is it not applicable?
3. How is the non-normality of monthly flows handled by the wavelet approach?
4. Are the sentences on lines 67 and 116 meant to be subsection headings?
5. The clustering method is referred to as GEDC (Line 117) and GeoEDC (Line 128). Please choose one and do a global replace.
6. In the line 135 equation, I assume sigma squared represents variance of the distance matrix. Please introduce this variable in Line 136.
7. Please consider referring to Fig. 4 as Fig. 4, 5, 6, etc.
8. Section 4.2: I struggled with the organization of this section. The paragraphs were organized by TSV, but the figures were organized by climate variable. I suggest you either (a) state this organization explicitly as an intro to Line 179, or (b) reorganize the paragraphs to match the figures or the figures to match the paragraphs.
9. Line 184: I see a negative correlation for zonal winds in the North Atlantic. Please verify this.
10. Line 281: You might consider using the word “anomaly/anomalies” to describe this.

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-395, 2016.