Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2016-395-AC1, 2017

Hydrology and Earth System Sciences
Open Access
Discussions

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.

Manuel Fossa et al.

manuel.fossa@etu.univ-rouen.fr

Received and published: 19 January 2017

Reply to AR#1 published 18 January 2017

We first thank AR#1 for its detailed and very constructive review of the paper. Whatever the decision from the editor, we will correct the article taking into account the several remarks made by AR#1. We wanted to clarify some important points though as we think, by our own mistake, have led to some misunderstanding.

Reply to Overall Review and Recommendation

In the overall review and recommendation AR#1 underlines several major problems with the article:

- The paper's organization makes the actual added value of the paper unclear.

- The GeoEDC method, which AR#1 believes is the main objective of the paper is poorly described.

- There's uneven repartition of details along chapters

- AR#1 advices to focus either on comparison of the GeoEDC with other clustering methods or on a climate forcings analysis

-Overall the AR#1 thinks either parts (GeoEDC and Climate analysis) are not sufficently developped.

We think that AR#1first comment is due to unclear statement of the hypothesis. The paper doesn't make the hypothesis clear enough and it is not introduced with enough details.

The hypothesis is underlined in the last sentence of the introduction (line 36)

"The present study aims at finding out if a classification of climate fields that are correlated with the 6, 10 and 21 year TSV 36 streamflow over France is possible and does so by introducing a clustering method called GeoEDC"

The hypothesis is thus to find whether the 152 discharge gauging stations show spatial structures depending on their climate drivers. This study is done within the context of two other works by our lab: -Recent publication (not published at the time of the paper writing) by Nicolas Massei ("multi-time-scale hydroclimate dynamics of a regional watershed and links to large-scale atmospheric circulation: application to the seine river catchment, france") did a multiresolution analysis of climate drivers to streamflow of the Seine river in France. -An unpublished work was done by Marie Nicolle on whether a spatial structure exists for the same 152 gauging stations by considering their multiresolution streamflow variability only (not considering their link to climate fields).

The present study thus expends on those two studies by "bridging" the spatial structure analysis of the streamflow with the climate field drivers.

As explained from the introduction in the paper, this brings the necessity of classifying climate fields which are represented as 2 dimensional scalar functions (function of latitude and longitude). The methods used so far either work well for one climate variable but not for some other or their intra cluster variability is too big and scope for improvement was there, hence the GeoEDC method.

We partly agree with the second point of AR#1's comments in that the GeoEDC is not described enough. However and considering our previous point reply, we do not agree that the main objective of the article is presentation of the GeoEDC method.

We again partly agree with the third point. We totally agree on the uneven detail repartition for Methods and Results, however we do not agree on cluster centroid, which we believe is a major consideration relative to the hypothesis. The point is developed in our reply to the major comment #10

The fourth and fifth points are in our opinion the result of the unclear hypothesis. According the our hypothesis we do not aim at discussing the climate forcings neither comparing the method to other method available (other than the necessary explanation of where it is supposed to bring improvement, obviously). The former point is the most important and some of the Major Comments by #AR1 are the consequence of our misleading introduction and hypothesis presentation. Since the study is whether there's spatial structures, at different time scales, for the 152 discharge gauging stations, finding rationales for the climate drivers would be beyond the scope of this paper. It may appear as a waste but the result part shows already the complexity of describing the spatial structure of climate driver influence on discharge in France depending on the TSV and climate variable considered. Likewise, while the methodology presentation and its comparison with other methods would be definitely interesting, we

think it would difficulty fit within a not too lengthy article whose hypothesis is not the performance of the GeoEDC method.

Summing our argumentation for the Overall Review and Recommendation, we totally agree with AR#1 about the unclearness and several unbalances in detail. Whatever the editor decision, we will correct the article taking into account the remarks. We however feel that some of the remarks, including major ones, made by AR#1 are due to a misunderstanding of the hypothesis (by our own fault).

Reply to 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.

We agree on that point, especially the 6-21 years, that wasn't even justified (which is an oversight from us). The reason why we choose those time scales is that low frequency variabilities in both discharges and climate field is a major topic in the climate change context due to climate models difficulty to represent variabilities above the annual time scale. The 6 year time scale has been shown by several studies (Massei 2007, Ghill 2010, Palus 2014) to be a major component of local hydroclimatic variability. Several studies covered the decadal time scale. Multi-decadal variabilites are the holy grail as they have strong long memory behavior.

2/There is an overall lack of citations

We agree.

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.

On this point, we will try to be more precise but we feel that the workflow was correctly explained and that Figure 2 and 3 are sufficiently explained in the text so as not to confuse between the two. If AR#1 speaks about the left bottom sub figure in Figure 2 then it is just an example of the multiple cluster maps produced by GeoEDC method and presented in Figure 3.

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.

We agree on that point, it should have been precised that those scale are not chosen arbitrarily but a result of the Multiresolution that breaks the signal into a dyadic series of time scale (each time scale is a power of 2 of the time series length).

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

We agree this was unclear. The decomposition is applied on the full time series as well as the correlation analysis, then winter months are chosen.

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

Agreeing about the general unclearness of the hypothesis we will make it clearer. However, the very reason to test each TSV is clearly explained in the introduction with references to motivate that choice. As for the different variable investigated, we agree that this should be in the introduction.

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 ex- plained. 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 sec- tion

Presenting correlation/covariance in streamflow among the cluster would be an assessment of the correlation not the clustering and the latter is done only on the shape of their correlation cluster fields. We think any clustering fit analysis would have to be centered on shape variance in ccfs and possibily the number of clusters. The shape variance in ccfs is analyzed in the study via the study of transition at the interface of stations that are geographically close but belong to different clusters. If such two stations show similar patterns it means the transition between clusters is smooth and thus that the limit between clusters is down to a threshold in the clustering not from some inconsistency of the clustering method. Silhouette algorithm is explained (surely not enough) which motivates our decisions as to the number of clusters for each TSV and each variable. We thus agree more explanation would be necessary but we think clustering fit has been at least partially addressed in the study.

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.

We agree, adding a reference to cluster colors is necessary.

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?

Line 195 doesn't relate to Africa so we assume that AR#1 is talking about line 175-176. Recalling AR#1 that the focus of this study is to characterize the spatial structure of discharge gauging stations over France according to their climate driver, the explanation of all correlation climate fields is beyond the scope of this article. We provided a comparison with results from other studies in section 5.3 (especially from line 273 on) but we think trying to analyze physical processes behind every correlation climate field is beyond the scope of this article.

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

As explained in the Overall Review and Recommendation reply, we think it is actually an important point in the context of the hypothesis because it gives elements of response as to whether the clusters have representative shapes or if an extension of the clusters beyond the geographical coverage of the stations is needed.

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

Lines 250-253 explain the rationale: The geopotentials have always the lowest number of a clusters for each TSV. As the study goes from large scale climate fields to local discharge, it is likely that a process that is the most homogeneous over the area considered is the first in a chain of events that are all more local. This is supported by atmospheric physics that shows that especially at the 850 hPa level, winds are essentially driven by geopotentials. We agree this could be better explained.

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