Reply to the Reviewers' Comments

We would like to thank the Editor and the Referees for reviewing the revised version of our paper. We acknowledge that one reviewer suggests publication of our manuscript in the present form, while the second reviewer suggests rejection.

While we appreciate the constructive approach adopted by the Editor, and felt very comfortable with the first round of the review process, we would like to kindly point out that we do feel uncomfortable with the second review round. The reason is that one **new** review (the report by Reviewer #2 in the second round) is not open and therefore the review process is not transparent as it should be (according to the journal's policy).

The problem is originated by the fact that a **new** reviewer was involved in the second review round (who did not respond to the invitation to review the paper in the first round). Therefore, his/her report in the second round is actually a **first round review**, which should be open and therefore publicly available. In fact, according to our understanding of the journal's policy, the second review round aims to assess whether the criticism expressed in the public review was successfully addressed or not. New criticism by a new reviewer should not be expressed. In fact, the email we received after the publication of our paper in HESS-D reads as (cut and pasted text is reported between asterisks, with relevant text in red):

****
----- Mensaje reenviado de editorial@copernicus.org -----
Fecha: Tue, 3 Apr 2018 09:11:07 +0200 (CEST)
De: editorial@copernicus.org
Asunto: hess-2018-134 (author) - manuscript available for public review and discussion

You are receiving the following email copy due to your co-authorship of the manuscript hess-2018-134. The original message was sent to the contact author defined upon manuscript registration. Please contact us in case of any discrepancies with regard to the manuscript.

Dear Theano Iliopoulou,
We are pleased to inform you that your following manuscript has been posted as a discussion paper in HESSD, the scientific discussion forum of HESS:

OMISSIS

As soon as the open discussion phase is over, no more referee comments or short comments will be accepted. During the following final response phase, however, you will have the opportunity to post final author comments. Before submitting a revised version of your manuscript for publication in HESS, you are obliged to have answered all referee comments and relevant short comments in one or more author comments in the discussion forum of your paper.

OMISSIS
****

We decided to submit our paper to HESS because we appreciate the transparency of the open review process and we appreciate the opportunity of the public reply to the reviewer comments. If a **new reviewer**, and therefore a **new report**, is involved in the second round, the distinguishing feature of the open review system vanishes. This is particularly relevant in this case as we do not agree with the **new** concerns that were raised by the reviewer in the second review round and therefore we would like to have the opportunity to publicly reply. We are confident that the Editor will recognize that our reply below is providing interesting arguments that, therefore, deserve to be known by the community.

Furthermore, we would like to stress that the audience would never know the real reason why the paper was not published if the review is not made open and our paper is finally rejected. This would be in contrast with the essential feature of the open review, namely, transparency.

Finally, we would like to point out that the second-round policy that was adopted here may stimulate reviewers to skip the first review round to avoid open publication of their report, therefore annihilating the benefit of submitting papers to HESS.

Therefore, we kindly ask the Editor that the review report by Reviewer #2, alongside with the present document which reports our replies, is published in the open discussion of the first review round. We believe that publication of reviews is important to keep the editorial process of HESS fully transparent.

Here below we reply to the concerns of Reviewer #2.

Reply to Reviewer #2

In the following, the comments of the Reviewer are copied in italic.

We first reply to the **general comment** of the Reviewer that reads as:

> *In the author's own words, the results are often 'expected' and the discussion section mostly 'confirms' previous work and understanding of what is controlling catchment streamflow.*

First, we feel it is necessary to clarify that the results were mostly not "expected". In fact, we use the term "expected" several times in the paper to highlight the conjectures that led us to design our experiment. In fact (lines and text refer to the revised version of the paper from the first round reviews):

- at line 130 we write "We use the mean flow in the previous month as a robust proxy of 'storage' in the catchment that is expected to reflect the state of the catchment, i.e., wetter/drier than usual";
- at line 156 we write "...as lakes and glaciers are expected to increase catchment storage thus affecting persistence";
- at line 173 we write "Geological features are expected to be linked to persistence properties...";
- at line 188 we write "We expect the presence of multi-collinearity among the explaining variables and therefore Principal Component…."；
- at line 520 we write "The former result may be explained considering that increased evapotranspiration (higher temperature) is expected to dry out LFS flows….";
- at line 532 we write "However, in the glacier dominated regime of western Alpine and central Austrian catchments this is not expected to be [equivalent to "expected not to be"] a relevant driver of higher correlation".

In other cases, we highlight that the results were **not** "expected". In fact:

- at line 338 we write "…indicates that it is not a key determinant of correlation";
- at line 347 we write "The impact of lake area (Fig. S1a) on correlation for LFS and HFS is not significant but positive…";

- at line 374 we write “Therefore, a spatially consistent pattern does not clearly emerge...”;
- at line 384 we write “Figure S2 in the Supplement shows that there is not a prevailing pattern in either case...”;
- at line 408 we write “Presence of lake, glaciers, karstic and Flysch areas do not appear significantly effective at a 5 % significance level.”;

Finally, only in some cases we indeed point out that the results confirmed our expectation and/or the outcome of previous studies. For instance:

- at line 464 we write “As expected from Eq. (3) and (4), the variance of the updated (conditioned) distribution decreases while the mean value increases.”;
- at line 479 we write “This result was expected since the LFS correlation refers to average flow while the HFS correlation is related to rapidly occurring events.” Please note that this sentence is relevant to our reply to comment #2 below.

Indeed, when we found a potentially interesting result we tried to provide physically based reasoning, and/or review of the previous literature, to give further support to our findings, namely, to provide evidence that they are not merely due to “noise”. This is what a rigorous scientific approach requires, rather than a sign of “conspiracy” (see the unfortunate wording that is used by the Reviewer in his/her comment #2 that is copied below).

Actually, we do see the fact that results confirm our previous conjectures as a positive outcome. When a deductive approach is used, the scientist first elaborates a conceptual reasoning to explain what is observed. In this case, we did observe that **peak flows in the high flow season** (HFS) are often preceded by high **mean flows in the previous month**. Therefore, in a previous work we decided to explore the correlation between the two random variables above (highlighted in bold) for two rivers only. The results confirmed our expectation. Therefore, the present contribution aims to (1) extend the analysis to the low flow in the LFS, (2) extend the analysis to several other rivers, and (3) explore the physical drivers of river memory.

About the latter issue, we of course needed to select physically based metrics to explain correlation. We conjectured what physical properties (metrics) may determine correlation and therefore elaborated an expectation. We therefore designed the experiment precisely with the aim to confirm our conjecture. The Reviewer seems to imply that confirmation of conjectures (expectations) makes the results meaningless. We regret to report that we disagree. Rather, confirmation of expectations means that the experiment is well designed.

1. *1. The following aspects of the methodology are unclear: For the HFS, the max daily discharge in the 3-month HFS is chosen. Is this value distinct from the max yearly discharge? In most cases, I suspect not. If the max discharge is in the second month of the HFS, does the lag-1 represent the correlation with the previous months mean discharge (also technically in the 3 month HFS), or with the last month before the onset of the HFS? If it is the latter, then the analysis is no longer technically a lag-1 analysis, and the study could be a big mix of lag-1, lag-2 and even lag-3 analyses that are all confused as representing a lag-1 value. Of course, it could be argued that you wish to be outside the HFS season for the correlation analysis, but what is the value of having inconsistent time periods in your lag analysis, especially given how sensitive the correlation will be to changing lag lengths? Moreover, what if a single catchment has max discharge always moving between the first and third month of the HFS over all the years of record? This will have a large impact on the ‘lag-1’ correlation even before any hydrological interpretations are involved. Some clarification on the mechanics of this analysis would really help.*

Strictly speaking, the Reviewer is right, but a mixed lag is not infrequent in hydrological analysis. As an example when we examine daily maximum discharges of consecutive years, we usually speak about the average time lag which is one year, but in fact this is mixed and varies between 1 day (if the max values were observed in 1 Jan of one year and 31 Dec of the previous year) to 730 days (if the max values were observed in 31 Dec of one year and 1 Jan of the previous year). In our view the important thing is to clarify the terminology and the methodology, and consistently define the related random variables. This does not necessarily require that data are sampled at regular time step or that the time distance of consecutive high flow events is constant.

In our case, we rigorously define in the paper the random variables which we consider. For instance, for the HFS they are:

- Peak flow in the high flow season (with arbitrary but rigorously identified length);
- Mean flow in the previous months.

We also clearly define that we denote with lag-1 the correlation between the peak flow in the HFS and the average flow in the previous month, before the onset of HFS. In the same way we define lag-2 correlation and so on. We regret to report that we do not agree with the criticism of the Reviewer and therefore did not make any major change to the manuscript in this respect. However, we have added this further clarification in the revised manuscript: “In the case of HFS, a correlation is sought between the maximum daily flow occurring in the HFS period and the mean flow in the previous months, <u>before the onset of HFS</u>.” (Line 128-129).

2. *The authors do not consider how the design of their study may have conspired to control the reported results before referring to a myriad of hydrological explanations. The core issue is one of signal vs noise. The LFS lag analysis uses a correlation between mean values that are by definition weighted by the central tendency of the data being considered, whereas the HFS uses a correlation between a max value and a mean, which is by design a far noisier signal, and hence displays little to no correlation with other variable throughout the study. Can the authors image a scenario where this would not be an expected result?*

We fully agree with the Reviewer that correlation between monthly data is expected to be higher with respect to correlation between local variables like peak flow. This is precisely the reason why the correlation that we found between peak flow in high flow season and average flow in previous month is a relevant (and not expected) result. It implies timely predictability of the probability distribution of peak flows, which is a relevant finding.

As for the low flows, we demonstrated that the correlation that we found is higher than the correlation computed for the whole set of monthly data. This means that focusing on the specific correlation of the monthly flow for the LSF season and the monthly flows of the previous months again allows us to improve predictability of low flows. Again, this is not an expected result. In both cases, we found that there is a specific signal that emerges above other signals and noise. Please note: it’s not just a question of signal versus noise, which highlights an oversimplified view of the inherent processes. It’s a matter of recognizing a specific signal – namely, correlation between previous monthly flows and LFS low flows and HFS peak flow – over other signals (monthly correlation, for instance, for the LFS) and random components.

Turning to the physical explanation that we sought, we do not see the reason why the fact that the results were expected would downgrade their value (please see our reply to the Reviewer’s general comment above). Therefore we rebut the statement that we designed the experiment by “conspiring” (a very unfortunate term, as we already remarked) in order to obtain expected results. Again, the experiment was based on our preliminary conjectures that are in turn based on conceptual and physical reasoning. The fact that the results confirm expectation is a confirmation that the experiment was well designed. For what reason should we investigate possible physical explanations that are not expected to be sound?

Still, we would like to point out once again that many of the explanatory metrics we investigated turned out to be not effective on the correlation, such as, for instance, the presence of lakes and glaciers for the HFS, catchment elevation, flysch areas and so on. Therefore, not all of our results were expected.

To mitigate the concern of the Reviewer, we changed the wording throughout the manuscript to avoid many repetitions of the term "expected". We also made changes in the Discussion section to better highlight the purpose of the analysis and underline more some of the most important and less expected results. The relevant sentences of the revised manuscript (copied at the end of the present report) read:

- At line 482: "We also aim to investigate physical drivers for correlation <u>and quantify their relative impact on correlation magnitude</u>."
- At line 486: "We found that increasing basin area and baseflow index are associated with increasing seasonal streamflow correlation, <u>yet the latter has a stronger impact</u>."
- At line 492: "Our results additionally point out that catchment storage induces mild positive correlation, not only for low discharges which are directly governed by base flow, but also for high flows, <u>which is less anticipated</u>."
- At line 509: "<u>In fact, our finding that increased wetness has a negative impact on seasonal memory of both high and low flows, extends the above results to the seasonal scale and interestingly, to both types of extremes.</u>"
- At line 513: "We also confirm the role of lakes in determining higher catchment storage and therefore positive correlations for the LFS, which has <u>only</u> been reported for annual persistence in a few sites (Zhang et al., 2012)."

3. *Related to point 2, the authors use a suite of metrics, many of which (P, SR, BFI) have a natural correlation with HFS and LFS since they are either derived from the same data or help generate it. The HFS analysis produces such a noisy signal that no result can be found, and this is hardly a surprising result (as mentioned above). LFS is not as noisy, and so displays better correlations. The heart of the paper is then to say that the correlations are better with hydrological processes that will also natural reduced the noise, e.g. higher groundwater flow subsidies and snowmelt, and worse correlations with processes and drivers that have increased noise. Again, can the authors think of a situation where this would not be an expected result?*

First, we believe there is a misunderstanding here. We did find that correlation for the HFS season is relevant and helpful to improve predictability. Please see Section 4.2 and 7. Therefore there is indeed a signal that we discovered over what the Reviewer terms "noise". Furthermore, we demonstrated that such correlation is explained by catchment area, precipitation and catchment storage in general. Therefore we regret to report that we cannot agree with the statement that "*The heart of the paper is then to say that the correlations are better with hydrological processes that will also natural reduced the noise"* (sic). The heart of our paper is stated in the last sentence of the abstract: "Our findings suggest that there is a traceable physical basis for river memory which in turn can be statistically assimilated into high- and low-flow frequency estimation to reduce uncertainty and improve predictions for technical purposes."

Furthermore, we do not understand the criticism by the Reviewer "*natural correlation with HFS and LFS since they are either derived from the same data or help generate it".* For instance, we analyzed the correlation between rainfall and river flow. Would the fact that rainfall generates river flow make the analysis of their correlation meaningless? We regret to say that we cannot agree.

4. *Given these factors, it is unclear what processes or understanding can be revealed by such an analysis, since the study is producing most of the results by design, rather than by hydrological insight. In this sense the analysis in this manuscript obscures the actual hydrology, for example if you just plot actual baseflow on the maps in Figures 7 and 8 a clearer pattern of the hydrological controls on low flows would be revealed (or indeed baseflow against elevation, as documented by a lot of previous work). Surely, the LFS lag analysis only obscures these key hydrological drivers rather than making them clearer or easier to understand? I think this is also clearly shown in the discussion section, which is highly speculative about general processes and mostly confirms the results of previous workers rather than adding new understanding.*

We are glad that the Reviewer recognizes the value of previous studies that analyzed the correlation between baseflow and low flow, even if both baseflow and low flows are "*derived from the same data"*. We believe our contribution provides relevant new findings such as:

- We confirmed **by referring for the first time to a large set of basins** that the peak flow probability distribution and the low flow probability distribution can be usefully updated in real time one or more months in advance through data assimilation.
- The physical drivers of predictability of low flows and high flows are **quantitatively identified for the first time for the chosen variables** (please note that graphical depictions may provide a more immediately clear representation, as we all know, but do not allow a quantitative assessment unless a quantitative relationship is provided, as we did).

5. I don't see the value or utility of section 7, it is incredibly short and not at all mentioned in the discussion section, therefore its completely unclear what we have learnt from this exercise, or in what context it's results should be considered. This asymmetry is considerable given it has more length devoted to describing the methodology than anything else in the paper (section 2.3). However, after reading it a couple of times I found this to be the most interesting part of the paper, since it asks an interesting question about how you would expect HFS or LFS to change based on obtaining the new average discharge for the previous month (an update). However, this seems to have already been published and discussed in detail by Aguilar et al (2017), so what is the utility of the very brief repetition of the same work on a single river in this study? Given the results and methodology are far closer to Aguilar et al. (2017) than the rest of the submitted manuscript, it seems entirely out of place and only confirms their previous work.

We agree with the Reviewer that the application presented in Section 7 (which arguably is not "*incredibly short"*) whose theoretical basis is presented in section 2.3, is similar to what is presented by Aguilar et al. However, we refer here to a different river which has a higher memory with respect to the case studies previously analyzed and we also present a LFS application for the same river. Therefore we believe the case studied here is technically interesting. Sections 2.3 and 7 are titled "Technical experiment: Real-time updating of the frequency distribution of high and low flows" and "Real-time updating of the frequency distribution of high and low flows for the Oise River". They are meant to be a technical example. They do not present a scientific advance in the strict sense, but we believe they are an interesting addition to the paper. However, we may easily remove section 7 (and therefore section 2.3) if the Editor feels that they are redundant.

Our replies to the minor comments of the Reviewer follows here below.

> *Figure 10 c, no colour scale provided*

We do not understand the comment as in our vision there is colour scale.

> *131: I don't understand the basis of correlating LFS with the mean flow of the previous month on the expectation this is a robust proxy for storage. If you define the LFS as the month with the lowest flow, then by definition the previous months will have higher flow, so how will this be a robust proxy for storage? In fact, you will be correlating against months that could also be included in the definition of HFS, which we would not suggest are a good indicator of storage.*

Perhaps we missed the exact meaning of the comment, but in any case mass balance and energy balance apply to fluid mechanics and therefore river flow formation. Mass balance suggests that storage is related to river flow. The Reviewer, may refer to a simple conceptual model like the bucket model, where higher storage implies higher discharge and the river flow is clearly a proxy for storage. Besides, the Reviewer may feel free to use better proxies in his/her studies.

> *154: "SR (m3 s–1 km–2) is computed as the mean daily flow of the river standardized by the size of its basin area. It may be an important physical driver as it is an indicator of the catchment's wetness" – so this basically says that runoff can be considered as an indicator of how wet a catchment is. This is like saying rainfall can be considered an indicator of how much water is falling from the sky, hopefully the authors can see the silliness of such a statement without further explanation.*

We are negatively surprised by the offensive tone used by the anonymous Reviewer. We do not see the reason why specific runoff should not be related to catchment wetness or aridity.

> *479: "This result was expected since the correlation refers to average flow while the HFS correlation is related to rapidly occurring events" See major points 2 -4, the design of the study is a major control on the results reported here rather than actual hydrological processes.*

We regret to confirm that we fully disagree with the idea that an experiment should not be designed according to physical basis and scientific reasoning.

We respectfully submit a revised version of our paper. We regret to report that we do not agree with the criticism of the Reviewer and therefore did not make any major change to the manuscript in this respect, but only small clarifications (discussed above). We rely on the Editor assessment, in particular for the opportunity of keeping (or not) Section 7 and 2.3.

With our best regards,
Theano Iliopoulou, Cristina Aguilar, Berit Arheimer, María Bermúdez, Nejc Bezak, Andrea Ficchì, Demetris Koutsoyiannis, Juraj Parajka, María José Polo, Guillaume Thirel and Alberto Montanari