Authors' response

Editor

<u>Editor</u>: Thank you for your detailed and complete reply to the reviews. Both Reviewers have now provided their feedback on your new version of the manuscript. Referee#1 recommends publish as is since all the issues have been satisfactorily addressed; however, Referee#2 identifies some issues that, despite being addressed in your reply, cannot be fully found in the revised text. Could you please check whether all the changes were implemented in the revised version? Please, provide in your reply specific points to identify the modifications.

<u>Reply</u>: We appreciate the efforts of the editor and the reviewers in revising the manuscript another time. We checked and further modified the manuscript following the last document of "authors reply". We apologize for any inconvenience.

Editor: I have checked some of the new texts, and changes, you provide in your reply and couldn't find them in the revised manuscript. The Referee #2 is just wondering whether all were finally implemented or whether some not definite version was finally upload. I kindly ask you to check this, <u>and I suggest that you provide in the new reply each point in the text were the changes can now be found in the new document</u> so that the Referee is not concerned about this.

<u>Reply</u>: We followed the suggestions of the Editor and prepared a new document of reply.

Reviewer #1

<u>Reviewer</u>: I appreciate the authors' effort in improving the manuscript addressing all the reviewers' comments.

<u>Reply</u>: We thank the Reviewer #1 for his/her support.

Reviewer #2

<u>Reviewer</u>: I stumbled across various discrepancies between the file with the responses to the referee's comments and the actual changes in the manuscript. It should be checked whether really the correct files have been uploaded. The authors have responded to the first two of my comments (concerning the lengths of the records analysed and the criteria for inclusion in the database). The authors now describe more clearly how they selected the streamflow series for their analysis. They now clearly state that 3485 stations were used in the analysis over the period 1950–2013. The new Fig. 1 is unfortunately hard to read, since each of the 3913 stations is represented by an individual line. This could be better represented by a figure that shows the number of gauges over time.

Reply: This representation of the data availability is commonly used in scientific literature as reported in Durocher et al. 2019 (Hydrological Processes. 2019; 33:1143‑1151). The figure allows to summarize the length of dataset in functions of the number of gauged station. In particular, it clearly displays the maximum period when the data are available. We tried to improve the quality of figure as far as possible.

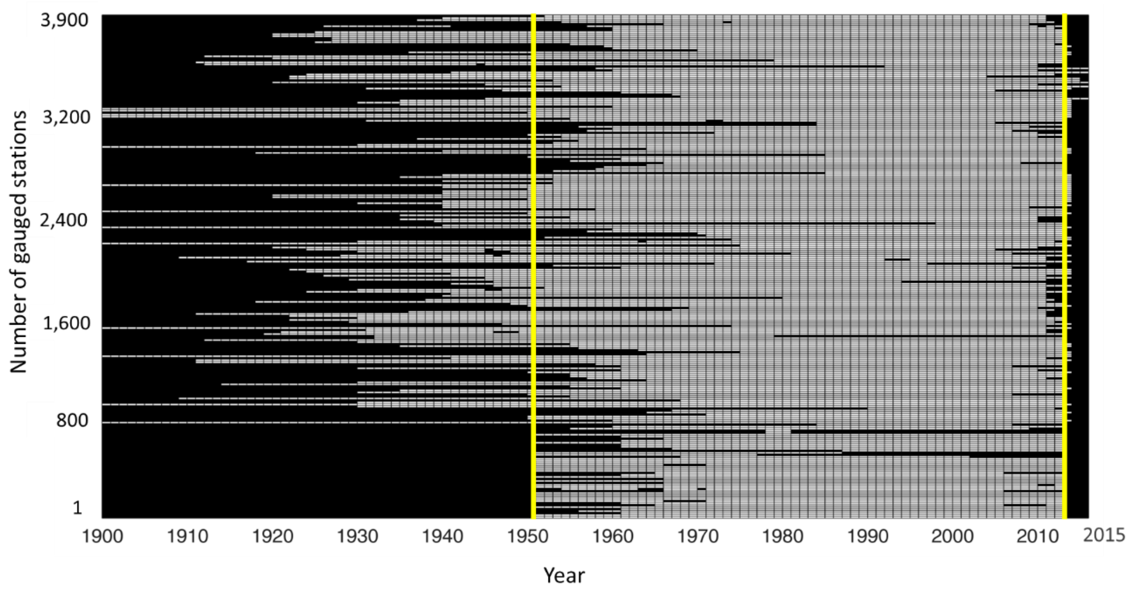

Figure 1. Data availability.

The new figures 2 and 3 illustrate a data gap and a series with a step change but I think they are not necessary for the manuscript.

Reply: We accepted the suggestion of the Reviewer and removed both the figures.

In the responses to the referee's comments the authors agreed with many of the comments and suggested modifications. However, unfortunately, these modifications have apparently in the end not been implemented.

Reply: We apologize for the misunderstanding. We prepared a new document modifying the last response to reviewer comments and specified the changes in the manuscript. In particular, the comments of reviewers are reported in black, the first reply is in blue and the current replay (new reply) in green.

Previous review: Reviewer #2

The study analyses trends in annual streamflow over the period 1950-2015 in Europe. This is a relevant topic certainly within the scope of HESS. The study generally applies standard methods for trend analysis (Theil-Sen slope, Mann-Kendall test). The spatial patterns of the trends are compared to spatial patterns of air temperature and precipitation. The study extends previous work on observed streamflow trends in Europe by including a higher number of catchments, particularly in Portugal, Spain, France and Italy. This was possible through assembling the database of streamflow records from various sources. The results largely confirm previous studies with dominant positive trends in northern Europe and dominant negative trends in the Mediterranean region.

Reply: We thank Rev#2 for her/his comments and suggestions, which allowed to improve the quality of the manuscript in this revised version.

New Reply: No comments to add.

Main comments:

Reviewer: Since the study states that records with missing data for more than two years were excluded from the database (L 107), I initially assumed that the calculated trends all relate to the period 1950-2015, which, looking at Fig. 2a, is apparently not the case.

Reply: Fig.2a was deleted because it created misunderstanding both in Rev#1 and #2. The original dataset included 3913 stations and after the checks on reliability, consistency and uniformity of series of data, 428 stations were discarded. The 63-year study period (from 1950 to 2013) has been chosen as the optimal threshold between maximizing series length and avoiding missing data, as shown in the following plot.

New Reply: We prefer to maintain Fig.1. Please, see the above comments.

Reviewer: This has of course a strong influence on the results and needs to be clarified. If the series length vary between catchments it will probably be more useful to analyze trends for different periods with nearly complete records, as the trends of course depend on the period analyzed (as discussed in the introduction).

Reply: We agree with the Rev#2 about the influence of length of series of data on trend identification. Dixon et al. 2006 coped with this problem by splitting the dataset in periods of different length, with a different number of stations for each period (see also Birsan et al. 2005). Nevertheless, one of the added values of our work was to consider a continuous dataset as large as possible over the entire study domain in order to evaluate spatial trends over European basins with a consistent sample size. It was the same approach proposed in the recent work by Durocher et al. (2019) where stations with missing data were discarded and a single time frame for all study domain was considered.

New Reply: We are firmly convinced that one of the added values of our work is to consider a continuous dataset as large as possible over the entire study domain in order to evaluate spatial trends over European basins with a consistent sample size.

We added a sentence at the end of Section 1.2 as follows:

"In the present study, we decided to maintain the integrity of the dataset focusing on the same time frame for all the study domain without splitting it in periods of different lengths. This procedure was already proposed in the study of Durocher et al. (2019) where preferred to discard all those time-series with missing data over a threshold rather than considered different time windows." (lines 190-195)

Reviewer: The criteria for inclusion/exclusion from the database should be described very clearly. It is not so clear whether the study aimed at only including near natural catchments. How were gaps smaller than 2 years treated? The steps that were undertaken to exclude inhomogeneous series, or series strongly affected by human interventions need to be mentioned clearly. For example, did the authors try to get information from the data providers on human interventions such as changes in flow abstractions etc. It should be described clearly how the database was 'consolidated and validated'. Did you apply any automatic screening tests to systematically check the series for possible inhomogeneities?

Reply: We thank the reviewer for raising this issue. In the revised manuscript, we have added details about the pre-processing activity done to select the discharge time series used for the analysis. In particular, to ensure quality of discharge observations, the following steps were followed: 1) check on data availability; 2) check for outliers (i.e. five st.dev. higher or lower than the means; 3) check on the presence of in-homogeneities through automatic screening tests. In order to filter out catchments affected by human disturbance, each discharge time series was accurately scrutinized through visual hydrograph inspection to identify disturbed hydrographs due to e.g. the presence of dams/reservoirs. Discharge time series characterized by disturbed hydrographs were discarded from the analysis. It should be noted that most of the basins considered in the analysis are taken from the EWA database, i.e. a discharge data collection of near-natural streamflow records from small catchments (Stahl et al, 2010). Moreover, the Global Reservoir and Dam (GRanD, https://sedac.ciesin.columbia.edu/data/set/grand-v1-dams-rev01) has been used to identify if (how many) dams/reservoirs are actually present in the selected basins. At the end of this analysis we expect that no substantial differences will be found between the basins retained for the analysis and the basins for which a certain degree of disturbance can be tolerated. Only stations with low human impact (no presence of dams/reservoir in the analysis period or no appreciable dam impact in the hydrograph); with less than 20% of missing data, showing no inhomogeneities in the time series were retained in the compiled dataset. Gaps smaller than two years were retained as missing data; during trend calculations, missing data were discarded on a case-by-case basis.

New Reply: We added two bullets in the methodology regarding this time-series check:

"The quality control was conducted in succession on daily and aggregated time-series following the steps reported in Gudmundsson and Seneviratne (2016):

(i) a visual hydrograph inspection to identify evident malfunction, consistent gaps (Fig. 2) and hydrograph disturbs such as presence of dams or reservoirs;

(ii) excluding catchments with a drainage area larger than 100,000 km^2 to minimize the possibility that the human actives can significantly cause disturbances on the streamflow time-series (Piniewski et al. 2018);

(iii) remove values with negative daily streamflow values;

(iv) remove time-series with more than 2 years of missing data."

See lines 120-135

Reviewer: Some results are not very clear. The results section reports significant trends in 95% of the stations, which disagrees with results reported in Table 1. In the results section, it is not always clear whether results on trends also include non-significant trends.

Reply: The number of basins reported in tab 1 (tab. 2 in the revised version of the manuscript) were incorrectly transcribed by the authors. They referred to the total number of stations in each macro-region (i.e. 3,485). In the table, only significant positive or negative trends are shown. These were 95% of total gauged stations (i.e. 3,310 stations). In the revised manuscript, the number of stations in each macro-region has been corrected. The manuscript will also clearly state whether any summary result includes non-significant trends.

New Reply: We added into the manuscript:

We clarified the results modifying Tab. 2.

"Results found that in 95% of the European gauged stations (i.e. 3,310 stations) the MK test confirmed the presence of a trend in annual streamflow volumes."

In addition, we modified Table 2:

Tab. 2. Number of significant (i.e. 3,310 stations) positive and negative trends in annual streamflow volumes in the European macro-regions.

Region	Number of stations	Positive trend	Negative trend
Boreal	323	307	16
Continental	694	472	222
Atlantic	1191	846	345
Mediterranean	1102	88	1014
Total	3310	1713	1597

Reviewer: I disagree with the finding of an inversion point in 1985 for the average series in the Mediterranean region. I do not see a change in the trend direction or trend slope in 1985. The fact that streamflow is above average before and below average after 1985 is a rather arbitrary result that depends on the selected study period. Streamflow has been decreasing since about 1965, and if anything, the rate of decrease has rather slowed down since the late 1980s.

Reply: The reviewer is right. Figure 7 in the manuscript highlights that streamflow has been decreasing since about 1965 and the rate of decrease has rather slowed down since the late 1980s. In the revised manuscript the sentences related to Figure 7 has been modified accordingly and supported by new statistical trend analyses on the entire time period.

"Fig. 7 shows a change in the annual streamflow volume pattern between 1980 and 1985 moving from positive to negative availabilities with respect to the mean of annual streamflow volume observations. This finding is consistent with the results found by Hannaford et al. (2013) on the marked decreasing of low flow regimes in southern Europe in the last thirty years as well as with the conclusions of the International Panel

of Climate Change (IPCC) work on climate change prospective (IPPC 2007) which highlighted how in the Northern Hemisphere climate change effects in reducing water resource availability have increased notably from the post- 1980 period."

New Reply: This part has been already added in the manuscript.

Reviewer: The calculation of the Sen's slope from annual streamflow anomalies is described as innovative, but if I do not overlook something this should not affect trends (and has probably been done in many studies).

Reply: By using anomalies to detect trends, the absolute random error is minimized (Pandžić and Trninić, 1992), but the reviewer is right in that it does not affect the trend (i.e., regression slope against time). Also, it is routinely carried out in both hydrologic and climatologic research. The methods section has been amended accordingly.

New Reply: The term "innovative" was removed. We modified as follows:

"Theil-Sen's line, known as Theil-Sen's slope or Sen's slope, was calculated on the annual anomalies in streamflow volumes, an alternative modality with respect to the common application on direct streamflow data (Birsan et al. 2005). "

Reviewer: The introduction should be improved. The introduction should clearly convey what has been found previously on annual streamflow trends in Europe? What is the gap in the current literature? How is this approached by this study? Please also check the logic of individual sentences and the subdivision of the introduction into paragraphs.

Reply: We thank the reviewer for this suggestion also underlined in the short comment by Adriaan Teuling. In the revised version of the manuscript, the introduction includes a more complete summary of what has been found by recent studies on annual streamflow trends in Europe, what is missing in the current literature and in which way this study will fill the gap. The revised introduction also relies on a more logical paragraphing.

New Reply: We largely modified the introduction summarizing the shared results on the trend detection as follows:

"Most studies have identified two separate trends, both from recent observations and using model projections sensitive to climate change: reduced flows in Southern and Eastern Europe (e.g., Stahl et al. 2010, Caloiero et al. 2018), and increased flows in Central and Northern Europe (up to minus or plus 45% after 1962 according to Teuling et al. 2019; -10-30% and +10-40% respectively by year 2050, under SRES A1B, according to Milly et al. 2005). Lehner et al. (2006) indicated large critical regions in southern and southeastern Europe for which significant changes in river flow drought are expected, and Feyen and Dankers (2009) projected increases in streamflow drought severity and persistence in most parts of Europe.

Models have also highlighted a reduction through time of areas with increased runoff, and an expansion of those decreased runoff (e.g., a north-bound expansion of drying in the Mediterranean area) (Milly et al. 2005). Similar trends were also found when analyzing trends in zero flow days (Tremblay et al. 2020) and peak flows or flooding frequency, although with high sensitivity to catchment size (Bertola et al. 2020). Most of such change has been attributed to changes in precipitation, with a less important role for land use and evapotranspiration change (Teuling et al. 2019).

Seasonal flows were also found to experience significant change (Bard et al. 2015, Bormann et al. 2017). Positive trends were found in the winter months in most catchments. A marked shift towards negative trends was observed in April, gradually spreading across Europe to reach a maximum extent in August. Low flows have decreased in most regions where the lowest mean monthly flow occurs in summer, with some exceptions in catchments buffered by a large groundwater storage capacity (e.g. Fleig et al., 2010; Laizé et al., 2010). Bates et al. (2008) summarized European studies that have found generally similar but more spatially explicit patterns including, for example, decreasing future summer flow in Central and East Europe. Also, models sensitive to climate change project that the peak in discharge will occur approximately one month earlier due to increased temperatures and earlier snowmelt in the future, with changes that are much more pronounced and statistically significant for all months under RCP8.5 compared to RCP4.5 (Lobanova et al. 2018).

Most studies, however, are based on observations limited to the second half of the 20th century (Piniewski et al. 2018, Renard et al. 2008, Birsan et al. 2005, KLIWA 2003, Schmocker-Fackel and Naef 2010, Demeterova and Skoda 2005, 2009, Fiala 2008, Fiala et al. 2010, Teuling et al. 2019). In addition, several studies have highlighted the extreme sensitivity of river streamflow to data selection, method of trend detection, and time window for the analysis (Stahl et al. 2010). Kundzewicz et al. (2005, 2017) advocated particular caution in interpreting streamflow trend signals resulting from a restricted number of stations with a small recording period, as even small gaps in the data time series or missing values could alter the significance of the statistical tests. Finally, even though trends highlighted by the literature are broadly consistent with spatial patterns of evapotranspiration and precipitation change, the effect of climate change on hydrology at the river basin scale is complex. Large-scale climate or hydrological models can reproduce broad patterns, but are still unable to capture all relevant spatially-distributed characteristics of physical catchment structures and associated processes, particularly in regimes with storage and release of water across the seasons (Stahl et al. 2010). Also, extending the analysis to longer time series might reveal unexpected influences from long-term climate variability modes, such as the North Atlantic Oscillation (Hannaford et al. 2003 , Steirou et al. 2017) or expected changes in the Atlantic Meridional Overturning Circulation (Rousi et al. 2020), which might introduce spurious trends in analyses focusing on shorter time spans. Finally, noise can be introduced by human modification and appropriation of streamflow, which may also reverse forecasted changes in river flow (Forzieri et al. 2014)."

Reviewer: The explanation of streamflow trends by trends in air temperature and precipitation remains a bit vague and overlooks areas where it is probably not possible to explain streamflow trends with trends in air temperature or precipitation (such as positive streamflow trends in northern Spain). Some arguments need to be clarified, e.g., it is not clear to me how groundwater or snowmelt effects would affect annual (and not only seasonal or monthly) streamflow.

Reply: The discussion on groundwater and snowmelt roles has been improved, also specifying that it will rely on speculation and literature and not direct measure or testing of such variables. The cases, in which the observed discrepancies between river discharge and weather series, could be explained by based on logical and science-supported hypotheses using likely drivers, will be highlighted with their most relevant examples (eg Northern Spain).

New Reply: We improve the discussion as follow:

"Concerning air temperature changes, the works of Staggle et al. (2017), Vicente-Serrano et al. (2014), Spinoni et al. (2015), Zeng et al. (2012), Willems (2013) and Madsen et al. (2014) confirm a global increase of mean temperatures with a marked trend in Mediterranean areas, where air temperature is expected to increase up to 0.3 °C/decade (as found in this study). The increase of air temperature directly impact glacierized and snow dominated basins where it can be responsible of the increase of runoff volume during the last sixty years due to the loss of ice masses (Sommer et al. 2020), however, depending on the basin elevation and trend in precipitation, some glaciers might have lost some sensitivity to an increased runoff production as a consequence of higher temperatures since there has not been more ice to melt and because, at high elevations, temperature might be not warm enough to counter balance the precipitation trend. In summary, for glacierized basins (or that use to be) there might be a causal effect of temperature on increased runoff volume (although this effect might have lost in time for some of them as explained below) while, for the others, precipitation seems again the main driver of runoff trend as it can be seen over the Alps by the contrasting trend found between the Italian side (negative) and continental side (positive) which reflects the trend in precipitation. On the other hand, temperature increase can impact negatively runoff over energy-limited environments by increasing evapotranspiration (Teuling et al. 2013, Avanzi et al. 2020) so some catchments might have experienced reduced runoff trend as a consequence of warming. This might explain the negative runoff trend found for some basins at high latitudes." (lines 305-340)

Detailed comments

Reviewer: P1, L28-30: The logic of the sentence is not clear. There is no contrast between a lot of research and not finding uniform streamflow trends in Europe. When mentioning a lot of research that aimed at investigating streamflow trends in Europe, this should be backed up by some references and their main findings (e.g. Stahl et al., 2010, Stahl et al., 2012).

Reply: The introduction, and in particular the review of past studies and their findings, has been deeply improved in the revised version of the manuscript. References has been added, including those suggested by the reviewer.

New Reply: We included additional references clarifying their outcome and findings as follows:

"Although the hydrological scientific community undertook a great effort, few research robustly demonstrates an ubiquitous and uniform trend in European annual streamflow volumes (e.g., Mediero et al., 2015; Alfieri et al., 2015; Hodgkins et al. 2017; Blöschl et al. 2019). …"

Then, we largely modified the introduction as shown in one of the previous point.

Reviewer: P2, L33-34: Did these studies also analyze changes in annual streamflow volume? What were the main findings? How did seasonal streamflow change?

Reply: As for the previous comment, the review of past studies and their findings has been deeply improved in the revised version of the manuscript.

New Reply: We largely modified the introduction as shown in one of the previous point.

Reviewer: P2 L40-47: The section on potential drivers of the streamflow trends remains a bit vague. Are changes in river cross-sections or boat tourism relevant for annual streamflow volumes?

Reply: Yes, if the shape of the river section is altered, or if flow itself is altered with recreational basin or locks for navigation. These sentences will be however moved to the Discussions to streamline the logical flow of the introduction. A missing reference to Vag et al. will be added in the References.

New Reply: We delete this part in the new version of the manuscript

Reviewer: P4, L97: I would suggest to first clearly list the criteria for selecting catchments and then mention the final number of selected catchments at the end.

Reply: The methods has bene amended accordingly – filtering criteria has been described in the methods, while the resulting number of catchments retained for analysis are reported in the Results.

New Reply: We largely modified the materials and methods, as follows:

"For assessing the reliability of streamflow daily values of each gauged station of the original dataset, a quality control and a homogeneity assessment were performed according the methodologies described in Buishand (1984), Chu et al. (2014), Ghiggi et al. (2019) and Kundzewicz (2015).

The quality control was conducted in succession on daily and aggregated time-series following the steps reported in Gudmundsson and Seneviratne (2016):

(v) a visual hydrograph inspection to identify evident malfunction, consistent gaps (Fig. 2) and hydrograph disturbs such as presence of dams or reservoirs;

(vi) excluding catchments with a drainage area larger than 100,000 km^2 to minimize the possibility that the human actives can significantly cause disturbances on the streamflow time-series (Piniewski et al. 2018);

(vii) remove values with negative daily streamflow values;

(viii) remove time-series with more than 2 years of missing data.

The homogeneity detection of data series (Fig. 3) was performed combining four different tests (Gudmundsson et al. 2018): (i) the standard normal homogeneity test of Alexandersson (1986); (ii) the Buishand range test (Buishand, 1982); (iii) the Pettitt test (Pettitt, 1979) and (iv) the Von Neumann ratio test (von Neumann, 1941). Homogeneity tests were carried out using the "iki.dataclim" statistical package for R (Orlowsky, 2014). The streamflow time series were considered as consistent when the null hypothesis at the 1% level was accepted at least in 3 of 4 tests (ECA&D) (Gudmundsson and Seneviratne, 2016; Merino et al., 2016). Despite potential levels of human-induced alterations of river flow regime could be still present in time-series data after the application of the aforementioned controls, a certain degree of disturbance can be tolerated (Murphy et al. 2013). In order to further reduce the disturbance, high flow conditions were not investigated and we focused the analysis on annual streamflow volumes."

Reviewer: P4, L101-102: You may use this in the introduction in order to emphasize your contribution in comparison to previous studies.

Reply: Suggestion accepted, the sentence will be integrated in the introduction.

New Reply: We added a key point into the objective as suggested:

"and (iv) to discuss the outcomes of the present study with previous investigations."

Reviewer: P4, L103-109: The description of the criteria for inclusion/exclusion from the database should be very clear. It is not very clear whether you aimed at including only near natural catchments. Did you check information from the data providers on human interventions such as changes in flow abstractions etc. (that would directly influence the trends)? Your database contains 3900 series of 65-years data. It is a lot of work to visually scan daily data of all these series. Could you provide some detail on how this was achieved? Did you apply any automatic screening tests? How were inhomogeneities identified?

Reply: Accepted - see reply to R2 comment 2 above.

New Reply: We largely modified the materials and methods (please see par 2.1 in the current version of the manuscript)

Reviewer: P5, L123ff: Why would it make any difference in terms of trend slope whether you calculate it on the original data or on the anomalies?

Reply: Accepted - see reply to R2 comment 5 above.

New Reply: We modified in the current version of the manuscript (see par. 2.1).

Reviewer: P5, L128: Delete "To homogenize the annual streamflow series", since dividing by catchment area cannot homogenize a time series.

Reply: Deleted.

New Reply: Done.

Reviewer: P5, L132ff: Have you checked the streamflow series for autocorrelation? How did you deal with series that contain significant autocorrelation?

Reply: The streamflow series of data were checked with lag-1 autocorrelation coefficient as proposed by Khaliq et al. (2009). The autocorrelation levels are reported in the picture in response to comment 4 of Rev#1. No series was significantly autocorrelated.

New Reply: We inserted in the manuscript the following section:

"About 90% of stations belongs to catchments with size less than 1,000 km2 of which more than 50% ranging from 1 to 200 km^2. Temporal autocorrelation level of the selected near-natural daily streamflow series was verified calculating lag-1 serial autocorrelation coefficient with a 95% of confidence bounds as suggested by Khaliq et al. (2009), Kulkarni and von Storch (1995) and von Storch (1995). All autocorrelation coefficients were found included in the confidence bounds, as shown in Fig. 5, and, therefore, they can be considered ready for the trend identification."

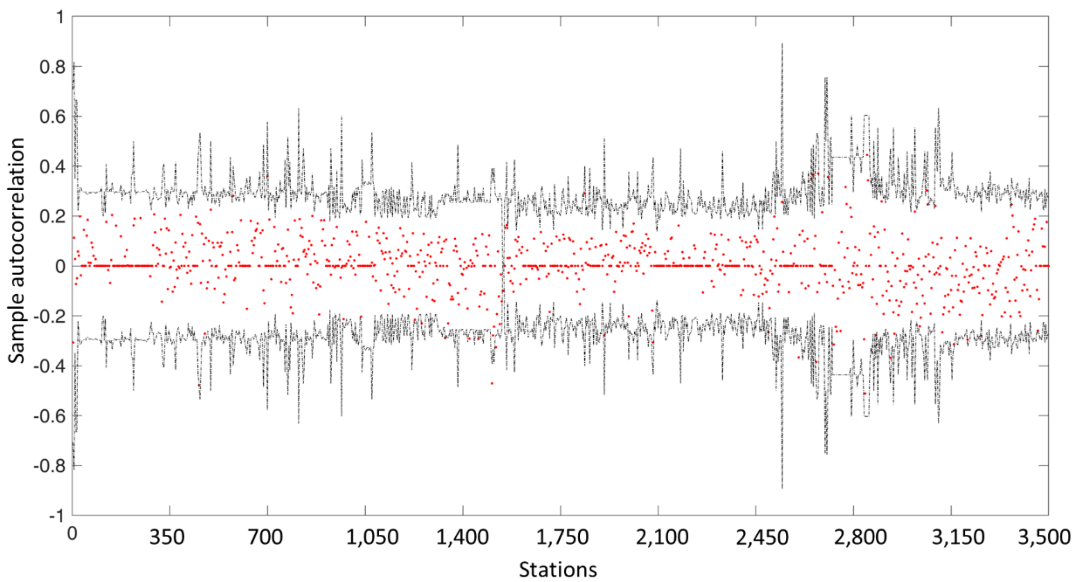

Fig.5. Samples autocorrelations. Red points are the value of lag-1 autocorrelation coefficient, whereas black dotted lines represent the 95% confidence bounds.

Reviewer: P6, L138: Since the streamflow volumes were divided by area, runoff depths would be more appropriate (instead of streamflow volume), no? (adjust throughout the paper)

Reply: Suggestion rejected – the reviewer is right, but streamflow volume is a widespread measure, which is readily understandable by managers and citizens. We decided to keep it that way.

New Reply: We decided to keep it that way.

Reviewer: P6, L145 and 146: This seems not correct, Table 1 shows positive trends in 7% and negative trends in 5% of the catchments?

Reply: The overall figures were corrected – 52% of positive trends and 48%, consistent with Table 2.

New Reply: We modified Tab. 2.

Reviewer: P6 Fig. 3: These figures are not necessary in my opinion.

Reply: The figure has been left in the revised manuscript just as an example of trend calculation.

New Reply: We decided to keep it that way.

1. **Reviewer**: P6, L151: The unit of annual streamflow per area is length/time (e.g. $m^3/(km^2$ year), or mm y^{-1}). Therefore the change in runoff over a certain period is length/$time^2$ (e.g. $m^3/(km^2$ $year^2$)).

Reply: Accepted – the values and units will be updated to reflect yearly change expressed in m3/(km2 year2).

New Reply: Done.

Reviewer: P7, L170; legend and caption of Fig. 5: replace rainfall by precipitation (assuming that snow is included).

Reply: Snow is included. Suggestion accepted.

New Reply: Done. We modified the caption as follows:

"Fig. 8. Comparison between annual streamflow volume trends and daily mean temperature (a) and rainfall (included snow-to-liquid equivalent) (b) trends over the European continent. Only significant trend are shown."