Reply to Reviewer 2

We would like to thank Thomas Over for the time dedicated to our paper and for his comments that contributed to improve the paper.

In the following, reviewer's comments are in Italic (R2), Authors' comments are in normal text (AC). Moreover, authors' changes in manuscript based on comments of all referees are summarized at the end of this document.

R2: This paper presents a method for estimating FDCs during an ungauged period at a "target" location that is gauged during another period. […].As such it is similar to a record extension application of the approach of Smakhtin and Masse (2000, Hydrological Processes, Vol. 14, pp. 1083-1100), except they estimated daily flow in ungauged basins, and to the work of Hughes and Smakhtin […] Ideally, however, the authors would investigate the distinction and show how in application one might make different choices regarding parameters or selection of the reference basin when estimating the FDC as opposed to daily streamflow.

Regarding the two papers recalled by the reviewer, Smakhtin and Masse [1] used the weather at a donor site, represented by the current precipitation index (CPI), to extend the daily hydrograph at a destination site through the monthly FDC of the destination site itself. The monthly FDC at the "destination" site is found using different methods such as (i) regionalization of FDCs based on available observed records from several adjacent gauges Smakhtin et al. [2] or (ii) conversion of FDCs calculated from monthly data into 1-day FDCs (Smakhtin, [3]). The procedure presented by Smakhtin and Masse [1] is an extension of a previous work proposed by Hughes and Smakhtin [4] to extend and/or filling in daily flow time series. The drawback of the procedures proposed by Hughes and Smakhtin [4] and Smakhtin and Masse [1] is the necessity of retrieving the monthly FDC of the target site with well-known literature methods before applying the methodology to extend the hydrograph. While the novelty of the approach we propose is the possibility to retrieve the FDCs at partially ungauged sites from weather recorded at a reference catchment. Thus, without the need of literature procedures such as the regionalization. We will recall the two papers by Hughes and Smakhtin [4] and Smakhtin and Masse [1] in the Introduction section, better highlighting the novelty of our work.

R2: (1) Why select another gauged basin and use only its API, not its streamflow?

For the German case study, we derived the streamflow and thus the FDC of each specific basin from the streamflow of another basin. We will better explain it in the text. We reported the evaluation metrics of the procedure to show the goodness of the method. We will better highlight the use of the streamflow as support variable.

p.C2: (2) Why not use the API at the target basin?

Because to retrieve the discharge at the target basin gauged during the target period, the discharge values only are necessary.

R2: 2.a. The idea of using the reference gauge discharge is raised and in the first presentation of the methodology on page 10, the FDC at the reference basin is computed, but it is never used.

AC: We will improve the presentation of the methodology and we will remove what is unnecessary to the development of the procedure.

R2: 2.b. The results for the US basins are presented quite differently than those of the German basins, including using different performance criteria.

AC: We decided to show results, in terms of flow duration curves, of the US catchments only to reduce the amount of possibly redundant data. In the revised version of the manuscript. we will use the same estimation metrics for both sites.

R2: 2.c. How the performance criteria could be applied for individual predictions was not clear to me.

AC: We will better explain it the text (Sect. 4.1) with the following text: "The **X** percentile is defined as the set containing all "**X**,…" numbers where the dots stand for the decimal points. For instance, the 1.09%, 1.36%, 1.63%, 1.91% belong with the 1rst percentile."

R2: 2.d. The choice of basins for the study seems rather arbitrary. For example, there are hundreds of basins in the MOPEX, including many others that do not have much snow.

AC: We decided to use these catchments also because the land use did not consistently change in the time window we used for analysis (see beginning of Sect. 3.2 (now Sect.3): "As resulted from the KS test, FDCs cannot be considered an invariant characteristic of a basin. The fact that FDCs are not invariant suggests that the weather is a driver of annual runoff variability. Indeed, the reason of the invariance must be the weather conditions as others (e.g. the catchment area, the land use) did not change.").

R2: 2.e. The consideration of energy versus water limitation as a measure of similarity is interesting but it is not clear that it is relevant when API is being used.

AC: In this paper the API is used to estimate the streamflow at a specific site. Annual streamflow variability is driven by the availability of water (i.e., provided by the precipitation) and energy (i.e., the evapotranspiration), [5]. Therefore, it is relevant to distinguish between water and energy limited catchments as their behavior is different.

R2: 1.a. Are there snow effects in the Upper Neckar basin? How addressed?

AC: Snow effects are considered by a simple snow accumulation and snowmelt model using a degree day approach. This allows to convert snow to a daily liquid water which is then used for the calculation of the API.

R2: 1.b. Do you take the karstic effects on Upper Neckar flows into account?

AC: For the karstic catchments a direct transfer seams not to be plausible. However the temporal stability of the API/Runoff can be considered as invariant. We'll check this assumption specifically.

R2: 2.a.i. Figure 3: Perhaps plot and check correlation with temperature of ET/P instead of just ET?

AC: The plot and the correlation of temperature and mean annual runoff (Q/P) already provide a similar information. It would be redundant.

R2: 2.a.ii. Figure 3: Need to consider uncertainty around correlation estimates: for the Peace R. it seems unlikely that rho = 0.027 is a significantly positive value; rather probably this basin is balanced between energy and water limitation by this criterion.

AC: We thank the reviewer for the suggestion, we will add the following consideration to the paper: "For instance, measurements at Peace River (LA) suggest that the catchment is balanced between energy and water limitation by the correlation criterion, Figure 3 upper panel."

R2: 2.b. Last sentence on p. 6: It is not possible…" It sounds plausible, but has this assertion been tested? The statement itself is very categorical; in fact there are degrees of water and energy limitation. How different do they need to be to make this true (if it is)? In particular, for the present application, the methodology may account for the water versus energy limitation; it may be that the timing of the weather is the most important thing to have in common.

AC: We tested the assertion. As a result it was not possible to estimate streamflow values of a water-limited catchment from the data of an energy-limited ones. See also reply to comment 2e.

R2: 3.a.i. Sect. 3.1. 2nd paragraph: It seems there should already be a well-established way of addressing autocorrelation effects on the K-S test.

AC: Weiss [6] proposed a methodology to account for modifying the K-S test for autocorrelated data. Later, Xu [7] suggested a method that can be applied to two sample test. However, our way to take into account for the autocorrelation is easier to implement and has a nice interpretation of equivalent sample size adjustment. More importantly, our method can be easily generalized to two samples test. We will introduce the following paragraph in the section: "[…] Since the streamflow data presents autocorrelation, the autocorrelation effects the KS test. Weiss (1978) proposed a methodology to account for modifying the K-S test for autocorrelated data. Later, Xu (2013) suggested a method that can be applied to two sample test. The information contained in the data is (usually) less than an i.i.d. sample with the same size. In other words, the number of equivalent independent observations is fewer than the sample size. In the following we explain how we took into account the equivalent sample size. It is easier to implement and more importantly, it can be easily generalized to two sample test. We can assume that the autocorrelation effect attenuates after three days [...]"

R2: 3.a.ii. a. Sect. 3.1. 2nd paragraph: The last two sentences of this paragraph seem to be referring to a test on a particular basin, but they are stated as if these relations are generally (i.e., mathematically) true. Which is it?

AC: This example is given for streamflow values at daily resolution recorded during a year, thus for a time series of 365 values. We will specify it with the following sentence in the paper: "For instance, let's take as an example the 1 year FDCs. If the samples were 3 times smaller and for instance their length would equal 122 …"

R2: 3.b. 3rd paragraph: This paragraph seems to include "Results", not "Methodology".

AC: In this paragraph we provide details about the methodology, but we also anticipate results regarding the KS test. This is done because those results explain the reasons why we applied the methodology presented in the following part of the paper to estimate the FDCs. For the sake of clarity, we decided to move the Sect. 3.1 in Sect. 2 and name it "2.4 Preliminary analysis".

R2: 4.a. Section 3.2: First paragraph: It is not always true that the non-weather properties (land use) do not change. Did you check that your study basins satisfy this assumption?

AC: We considered basins where the land use did not deeply change in time.

R2: 4.b. Section 3.2: Last sentence: Did you test different values of alpha other than 0.85, or just select that value for the reasons given?

AC: We tested also other values of alpha. However, we decided to proceed with alpha=0.85 because when α tends to 1, API represents the long memory of the basin as it includes the effect of precipitation occurred many days before. Moreover, this is in agreement with the value used by [8] for the Neckar Catchment. Therefore, we will add a line: "To capture this behaviour, in this study α is chosen equal to 0.85, this is in agreement with a previous study by Sugimoto (2014) who investigated the same case study area (i.e. Neckar catchment)."

R2: 5.a Section 3.3. First paragraph, last sentence: Why do you assume "large scale precipitation"? What do you mean by that?

AC: Small-scale variability of rainfall can be assumed to vary in a range lower than 10–20 km [9]. Therefore instead of APIs calculated from point precipitation areal precipitation is considered. The wording may be inappropriate and will be changed.

R2: 5.b. Last complete paragraph on p. 10: It seems it would be better to interpolate between Pj and Pj+1 rather than taking the mean, but it may not make a lot of difference.

AC: Thank you for the suggestion, as you anticipated, the difference is not significant.

R2: 6.a. Section 4, p. 11, discussion of figures 5-8: a. Several statements regarding goodness of fit are made without being quantified. However the K-S technique has been presented and could be applied; indeed, it would be ideal to provide K-S test results to accompany the results in each panel of these plots.

AC: Thank you for the suggestion, the KS distance D* will be added to each panel.

R2: 7.a. Section 4, p. 13, figure 10 and discussion of it: a. Why present 30, 70, 90, and 99th percentiles? As one can see, 90th and 99th (though the lower right panel of figure 10 is labeled as the 95th percentile), are almost the same. The complementary percentiles, i.e., 70, 30, 10, and 1st percentiles (exceedance probabilities) would be more interesting, in my opinion.

AC. We chose these percentiles as they are flow percentiles usually investigated in literature. For instance, the approach by Franchini and Suppo [10] regionalises these streamflow quantiles.

The title of the plots will be checked and made consistent with the caption.

R2: 7.b. You say (lines 5-6 of p. 13): "it is not possible to estimate the flow quantiles using regression methods that do not take into account the weather characteristics." This may be an over-statement. You have demonstrated that if you want to transfer across time, weather fluctuations need to be considered. But for prediction at ungauged basins for a fixed period of time, that may not be true

AC: The moving average is computed to show that the between-year variability of the discharge of a specific percentile can be high. Therefore, this suggests that percentiles cannot be considered an invariant characteristic of the basin and thus they cannot be estimated using geographic and morphologic characteristic of the basin only.

This is true also for prediction at ungauged basins for a fixed period of time as we demonstrated applying the K-S test to streamflow values gauged during the same time window at two different sites (see Sect.3.1 lines 21-22).

R2: 1. Section 3.1, 3rd paragraph: This paragraph seems to include "Results", not "Methodology.

AC: In this paragraph we provide details about the methodology, but we also anticipated results regarding the KS test. This is done because the results justify the methodology we present in the paper to estimate the FDCs.

R2: 2. Section 3.3, in steps 2&3 of the "procedure to predict" (p. 10), the FDCs of the reference catchment A is computed, but it does not seem to be used in the procedure.

AC: We will remove it and rephrase the methodology section to make it clearer.

R2: 3. Section 3.3, step 8 of the "procedure to predict" (p. 10): Suggest "qBrj is taken to be the value of discharge that occurred…" rather than simply "qBrj is the value of discharge that occurred…".

AC: The Section will be rewritten to make it clearer.

R2: 4. Section 3.3, last paragraph (p. 11): It is stated here that in the paper both discharge and precipitation will be used as the support variable. But everything before indicates that only precipitation will be used. And I don't see any results using discharge as the support variable.

AC: For the German case study, the support variable is the discharge (please, see also lines 5-7 p.11). We will better explain it.

R2: 5.a. Section 4, figures 5-8: From what period is this FDCref_site that is plotted? As it does not seem to be used in the procedure, why is it plotted?

5.b. Section 4, figures 5-8: I think however you should add the FDC of the target site during the reference period to these plots so the reader can see how much the FDC has changed from reference period to target period.

AC: Since usually the FDC of a donor site is used to retrieve the FDC of a target site, the FDCref_site was plotted to show the difference between the FDC at the donor site and the FDC at the target site recorded during the same period of time. Each FDCref_site is recorded during the reference period reported either in the plot or in the caption.

R2: 6.a.i. Section 4, figure 9 and discussion of it: Discussion of figure 9 on p. 12, lines 9&10: "Results shown that the distance between the former pairs is bigger than the distance between the latter, Figure 9.":i. I don't think you ever defined the K-S distance. That needs to be done.

AC: The K-S distance will be defined as:

"Moreover, the test allows us to estimate the distance between couple of FDC:

$$\mathrm{D}^{*} = \max_{x}\left(\left|F_1(x) - F_2(x)\right|\right), \qquad (3)$$

where $F_1(x)$ is the proportion of x_1 values less than or equal to x and $F_2(x)$ is the proportion of x_2 values less than or equal to x. F_1 and F_2 are two FDCs."

R2: 6.a.ii. Section 4, figure 9 and discussion of it: a. Discussion of figure 9 on p. 12, lines 9&10. I am willing to believe this assertion is true, but it is hard to see just from the plot. Can you provide some summary results such as the mean and median difference between 9 (top) and 9 (bottom) to give evidence of the assertion.

AC: We will provide the following summary to evidence the findings: "Moreover, we wanted to know which is the distance between pairs of FDCs built at the same site in different periods of time. This distance is bigger than the distance between simulated and observed FDCs at the same period of time for the 73% of the cases, Figure 9."

R2: 6.a.iii. Section 4, figure 9 and discussion of it: a. Discussion of figure 9 on p. 12, lines 9&10. This assertion should be restated without the shorthand of "former" and "latter". It is hard to understand the way it is currently phrased, and it is a very important point.

AC: We will rephrase the sentence as explained in the comment above.

R2: 7. Section 4, pp. 11-14: It is not clear why the Results section starts by giving a lot of results for the U.S. catchments and none for the German ones.

AC: We decided to show results of U.S. in a comprehensive way (both FDCs and performance criteria are shown) to keep compact the manuscript avoiding redundant plots. For the German case study we shown the performance criteria which are much more representative than the FDCs. The performace criteria are shown in an extensive way as they are reported for both case studies.

R2: 8.a. Section 4.1, pp. 14-17, Definition of performance criteria: Are all these computed for Q in mm units? Even though those are units used throughout, it would be worth re-emphasizing that here.

AC: Yes, Q is in mm. We will better highlight it with this sentence "Discharge values are in mm."

R2: 8.b. BIAS: This is not a simple bias as it is normalized by Qsim; it is more like a relative bias or "relative mean error"; however usually one divides by Qobs. Actually, ME (defined later) is more like a simple bias.
c. Why apply different criteria for the German catchments?
d. "Ratio":
i. Can you give it a more meaningful name?
ii. This formula looks odd. If the square root were only on the numerator, it would be the standard error divided by the mean error (and the quantity would be non-dimensional). But why apply the square root to the mean error in the denominator?

AC: We thank you the reviewer for the suggestion, we will use the same metrics for both case study areas. We reviewed the estimation metrics, there was a typo in the BIAS formula, the correct form will be reported in the paper and below. Results were estimated with the following formula in agreement with Castellarin et al. [11]

$$\mathrm{BIAS} = \frac{1}{N} - \sum_{i=1}^{N} \left(\frac{Q_{sim,i} - Q_{obs,i}}{Q_{obs,i}} \right).$$

R2: 9. Section 4.1, figures 11-13: Many of the colors these figures are shifted so each box has more than one color, making them hard to interpret. This effect needs to be fixed.

AC: We are sorry for this issue, we will fix it.

R2: 10. Section 4.1: I don't see how the Performance criteria were applied to create the results shown in figures 11-13. As I understand, there is only one prediction of each quantile, for a fixed reference catchment and decade. Then how does one do the summations indicated in the performance criteria formulae? Following the definition of NSE on p. 15 it says: "N is the number of discharge values related to a specific percentile". How many of those are there? Is there ever more than one? If so, how? The situation with the correlation coefficient values presented in figure 14 seems to be the same: How does one compute correlation coefficients without multiple values? If there are multiple values, where are they coming from?

AC: We will better explain it the text with the following sentence: "The **X** percentile is defined as the set containing all "**X**,…" numbers where the dots stand for the decimal points. For instance, the 1.09%, 1.36%, 1.63%, 1.91% belong with the 1rst percentile."

R2: 11.a. Conclusions: p. 21, lines 2-5: "Here it is shown that two FDCs built for the same catchment, but with data corresponding to two different time windows, cannot be regarded as the same continuous distribution. The same results when two FDCs of two different catchments built for the same time window are analysed. Thus, it is not possible to infer a FDC using parameters retrieved from the distribution of another FDC without considering the weather." The first sentence supports the assertion in the third, but the second does not. If two different catchments experience possibly similar weather but produce a different streamflow, the cause is not the weather.

AC: We thank the reviewer for the suggestion. The sentences will be: "Here it is shown that two FDCs built for the same catchment, but with data corresponding to two different time windows, cannot be regarded as the same continuous distribution. It is not possible to infer a FDC using parameters retrieved from the distribution of another FDC without considering the weather.

R2: 11.b. Conclusions: p. 21, lines 13-14: "Since precipitation data series are characterized by a high number of zeros, here we used the Antecedent Precipitation Index (API)." This statement misses the more important fact that the API combines in a streamflow-like way the history of the precipitation. (A similar statement is made at the beginning of section 3.2 near the bottom of p. 8.)

AC: We thank the reviewer for the suggestion. We will rephrase as:

"Since precipitation data series are characterized by a high number of zeros, here we used the Antecedent Precipitation Index (API). The API is used as it represents in a streamflow-like way the precipitation of the basin. It represents the memory of a basin as it provides the amount of precipitation released by the soil throughout the time."

R2: 11.c. p. 22, lines 26-27: Qualitative statement about similarity in shape from beginning of Section 4 is repeated. This assertion needs to be quantified somehow.

AC: A quantitative assessment of the goodness of the methodology is performed through the performance criteria. To describe why there can be a difference between the two FDCs, we will rephrase the sentence as:

"The distance between the simulated and observed FDCs can be due to the different temperature values characterizing the two catchments."

Author's changes in manuscript

Some hints regarding authors' changes in the manuscript have been already given in the comments section. In the following, a summary of authors' changes in manuscript based on comments of all referees is given:

1. In the Introduction, we will clarify the novelty and the contribution of the paper.
2. The Methodology section will be reorganized and improved to provide a clearer description of the method. For the sake of clarity, a figure will be added.
3. We will provide a clear statement of the hypothesis.
4. The sections will be organized on the base of the Introduction-Methods-Results-and-Discussion (IMRAD) format.
5. For both case studies we will use the same performance criteria and results will be discussed in-depth.

References

[1] V.Y. Smakhtin, B. Masse, Continuous daily hydrograph simulation using duration curves of a precipitation index, Hydrol. Process. 14 (2000) 1083–1100. doi:10.1002/(SICI)1099-1085(20000430)14:6<1083::AID-HYP998>3.0.CO;2-2.

[2] V.Y. Smakhtin, D.A. Hughes, E. Creuse-Naudin, Regionalization of daily flow characteristics in part of the Eastern Cape, South Africa, Hydrol. Sci. J. 42 (1997).

[3] V.Y. Smakhtin, Generation of natural daily flow time-series in regulated rivers using a non-linear spatial interpolation technique, Regul. Rivers Res. Manag. 15 (1999) 311–323.

[4] D.A. Hughes, V. Smakhtin, Daily flow time series patching or extension: a spatial interpolation approach based on flow duration curves, Hydrol. Sci. J. 41 (1996) 851–871. doi:10.1080/02626669609491555.

[5] G. Blöschl, M. Sivapalan, W. Thorsten, A. Viglione, H. Savenije, Runoff prediction in ungauged basins: Synthesis across Processes, Places and Scales, Cambridge University Press, 2013.

[6] M.S. Weiss, Modification of the Kolmogorov-Smirnov statistic for use with correlated data, J. Am. Stat. Assoc. 73 (1978) 872–875. doi:10.1080/01621459.1978.10480116.

[7] X. Xu, Methods in Hypothesis Testing, Markov Chain Monte Carlo and Neuroimaging Data Analysis, 2014. https://dash.harvard.edu/handle/1/11108711.

[8] T. Sugimoto, Copula based Stochastic Analysis of Discharge Time Series, 2014.

[9] D.K. Nykanen, E. Foufoula-Georgiou, W.M. Lapenta, Impact of Small-Scale Rainfall Variability on Larger-Scale Spatial Organization of Land–Atmosphere Fluxes, J. Hydrometeorol. 2 (2001) 105–121.

[10] M. Franchini, M.Suppo, Regional analysis of flow duration curves for a limestone region, Water Resour. Manag. 10 (1996) 199–218.

[11] A. Castellarin, D. Burn, A. Brath, Assessing the effectiveness of hydrological similarity measures for flood frequency analysis, J. Hydrol. 241 (2001) 270–285. doi:10.1016/S0022-

1694(00)00383-8.