Response to the review comments to the manuscript **hess-2020-600**

“Characteristics and process controls of statistical flood moments in Europe – a data based analysis”

by D. Lun, A. Viglione, M. Bertola, J. Komma, J. Parajka, P. Valent and G. Blöschl

We want to thank the editor and the referees for their useful and constructive comments. Here we reproduce the comments of the editor and of all referees in *italic characters*, followed by our answers. The line numbers of the referee comments refer to the line numbers of the revised manuscript, if not stated otherwise.

Thomas Kjeldsen (Editor)

The reviewers have considered the revised version of the manuscript. While they are generally impressed with the study they have both suggested further, relatively minor, revisions. In particular reviewer #2 have asked for more clarification on aspects raised in the first review. In particular, asked the authors to consider shortening the paper to below 10,000 words. Please consider the comments in detail and submit a revised manuscript for consideration.

> We thank Thomas Kjeldsen for the useful and constructive comments. We have addressed all comments of the referees below. The main text of the paper, excluding tables, figure captions and the appendix is currently 10,262 words. We have expanded the text following the suggestion of the referees regarding more clarification and it is now 10,800 words, which we believe is close to the target of 10,000.

Kolbjorn Engeland (Referee)

I think the paper is suitable for publication following some minor clarifications.

> We thank Kolbjorn Engeland for the valuable comments on the second revision of the manuscript that helped improve the quality of the manuscript. All his comments are reproduced and addressed in the following paragraphs, the line numbers refer to line numbers in the revised manuscript with tracked changes accepted.

1: Are peak floods or daily floods used in the analysis? The sentence on line 12 makes it unclear: 'Annual maximum discharges were derived from instantaneous peak flows and daily mean flows for each calendar year.'

> About 25% of the 2,370 flood peak series are instantaneous annual peaks and the rest are annual maxima of daily mean flows, depending on data availability. Given that the average catchment size is about 2,500km^2, we consider the effect of this inhomogeneity small relative to the spatial contrasts of floods in Europe. For example, Merz et al. (1999) found that the ratio of annual maxima of instantaneous peaks and daily flows for a catchment size of 2,500km^2 is on average 1.2, which is small relative to the spatial contrasts in Europe.

2: Figure 6: It might help to first show MAF for all sub-regions and then CV for all sub-regions. Then the visual interpretation of the heat map is easier.

We changed the ordering of the columns of Figure 6, as suggested.

3: Lines 665-670 I would be careful to make a direct link to the paper by Wang et al., 2017 since you in the current paper uses data from catchments that are only limited influenced by reservoir operations. What you analyse is the effect of natural lakes, whereas Wang et al (2017) consider the influence of reservoirs. Reservoirs introduce much more non-linearity than natural lakes (reservoirs often introduces thresholds in the system response when the dam is overtopped, whereas for lakes, there is a much more gradual transition). In addition, LUW does not account for the location of the lake in the catchment and does not directly tell how large part of the catchment runoff that has to flow through the lakes.

We fully agree with this assessment, and have therefore modified the sentence to emphasize that reservoirs and natural lakes tend to have different response characteristics.

“The former is consistent with retention effects while the relationship between CV and water body size may be non-linear (increasing CV up to a water body threshold and decreasing CV beyond as shown by Wang et al., 2017 for reservoir effects) which is not captured by Spearman correlation. However, in comparing natural lakes and reservoirs it should be noted that reservoirs tend to introduce more non-linearity in flood frequency behaviour because of a threshold effect when the spillway is activated.“

4: Lines 685 : You could add one or two sentences suggesting non-linear approaches that could be used, e.g. generalized additive models GAM (Rahman et al, 2018, Umlauf & Kneib, 2018) and Random forest (e.g. Desai eta al, 2021) that

We have added a sentence on possible non-linear modelling procedures, as suggested by the referee.

“While here we examined monotonic relationships and linear relationships, it would also be worth exploring non-monotonic relationships between flood moments and covariates (see e.g. Blöschl and Sivapalan, 1997; Smith, 1992; Pallard et al., 2009). Possible approaches for modelling non-monotonic relationships include generalized additive models (Rahman et al., 2018, Umlauf and Kneib, 2018) and Random forest regression (Desai et al., 2021).”

References

Blöschl, G. and M. Sivapalan (1997) Process controls on regional flood frequency: Coefficient of variation and basin scale. Water Resources Research, 33 (12), pp. 2967-2980.

Desai, S., Ouarda, T.B.M.J. (2021) Regional hydrological frequency analysis at ungauged sites with random forest regression, Journal of Hydrology, 594, https://doi.org/10.1016/j.jhydrol.2020.125861.

Merz, R., G. Blöschl und U. Piock-Ellena (1999) Zur Anwendbarkeit des Gradex-Verfahrens in Österreich (Applicability of the Gradex-Method in Austria). Österreichische Wasser- und Abfallwirtschaft, 51, (11/12), pp. 291-305.

Pallard, B., Castellarin, A., and Montanari, A.: A look at the links between drainage density and flood statistics, Hydrol. Earth Syst. Sci., 13, 1019–1029, https://doi.org/10.5194/hess-13-1019-2009, 2009.

Rahman, A., Charron, C., Ouarda, T.B.M.J. et al. Development of regional flood frequency analysis techniques using generalized additive models for Australia. Stoch Environ Res Risk Assess 32, 123–139 (2018). https://doi-org.ezproxy.uio.no/10.1007/s00477-017-1384-1

Smith, J. A. (1992). Representation of basin scale in flood peak distributions. Water Resources Research, 28(11), 2993-2999.

Umlauf, N. and Kneib, T. (2018) A primer on Bayesian distributional regression, Statistical modelling, 18(3.4): 219-247

Wang, W., H.-Y. Li, L. R. Leung, W. Yigzaw, J. Zhao, H. Lu, Z. Deng, Y. Demisie and G. Blöschl (2017) Nonlinear filtering effects of reservoirs on flood frequency curves at the regional scale. Water Resources Research, 53, 8277–8292, doi: 10.1002/2017WR020871

Anonymous Referee #2

The authors have improved the article and incorporated some suggestions from reviewers. However, further improvements described below are needed before this article can be accepted for publication. I have also attached a Tracked Changes document to this round of revisions for writing suggestions and minor technical comments. I've incorporated my feedback on the authors' responses to my initial comments below, including some places where I agreed with their responses.

> We thank the anonymous referee for the valuable comments on the second revision of the manuscript. All writing suggestions have been adopted. The technical comments are summarised and addressed in a table at the end of this document.

Effects of nonstationarity on flood series moments ---

The authors described nonstationarity as being outside the scope of their paper in their response to my initial review. While I agree that it is not the goal of their paper, I still think a quick investigation as to whether trends might affect estimates of sample moments is imperative, especially given that snowmelt comprises a major control on flood generation in 3/5 regions they examine.

This could be limited to an analysis of the effects of trends in the MAF on estimates of the CV following the conditional moments framework of Serago and Vogel (2018). If a trend is not accounted for, the CV can be overestimated since the overall variance of the peak flows will also include the variance explained by the trend.

The effects of trends in the mean and variance on CS can be mathematically derived but given general estimation challenges with at-site skewness arising from sampling variability, it seems like a less essential endeavor for this study.

I strongly recommend that any choice to retain the assumption of stationarity for any region should be supported with at-site trend analyses of these site records as well as any prior literature, including studies that examine trends over periods of record of more than 50 years *(to avoid confounding apparent trends with artifacts of inter-decadal variability) and studies specifically focused on snow trends.*

Finally, I agree with the authors that the autocorrelation of the annual flood series is often weak and, consequently, that adjusting significance inferences for persistence is of second-order importance for their continental-scale investigation. However, it should be noted very briefly to avoid any inappropriate uptake of this work.

> We agree that the presence of trends is a relevant aspect in the analysis of flood moments. Trends of exactly the same data set have already been analysed and published previously (Blöschl, Hall et al., 2019).
>
> We prefer not to include a trend analysis and/or an analysis of the moments of the residuals to a trend, as this would completely change the focus of the paper. Our approach is pragmatic in the sense that we are interested in the statistical flood moments for the period 1960-2010. We use the standard product moment estimators to make the results comparable with a large body of literature.
>
> We now do acknowledge in the discussion section the existence of trends in the data and the possible effect on floods moments. We also state that flood moments could be investigated by using a framework including trends in some of the moments, such as the one suggested by Serago and Vogel (2018).
>
> "A possible extension of the analysis presented here could be the consideration of non-stationarities in flood moments, for example in the spirit of Serago of Vogel (2018). Blöschl, Hall et al. (2019) have found that significant trends do exist in the mean flood of the data set in 28.02% of the stations. Trends affect the estimation of flood moments. For example the detrended data tend to exhibit smaller CVs than the raw data, while the effect on the sample mean may be smaller."
>
> We believe we already refer to the issue regarding autocorrelation in line 194.

Sample moment estimation biases under stationarity --

While the authors point out some good reasons for not pursuing further efforts to correct common product moment estimators for bias, I think that the authors should pay a little more attention to this. In their revised manuscript, the authors write "while the estimation uncertainty of the mean is small, the uncertainty and bias of the estimators of CV and CS (equations 3 and 4) can be substantial. Ye et al. (2020) illustrate the uncertainty and bias in the estimation of CV". This does not provide readers with an idea of the magnitude of this bias nor a sense of how to determine it should they be concerned about it for a practical application.

The authors are correct in observing that the specific methods for bias correcting CV estimates that Ye et al. (2020) employ (recommended in prior round of review) assume distributions other than the GEV, a distribution whose prominence in many parts of Europe has been previously established. However, Ye et al. (2020) provide these methods as examples and make it clear that the bias of the common product common CV estimator is not specific to any theoretical probability distribution.

Ye et al. (2020) cite the following general relation between the bias of the product moment estimator of the CV and the population CV, population CS and record length from Breunig (2011):

*Bias(CV_est) = CV_true^(3/2)/N * [3*sqrt(CV_true) – 2*CS_true]*

Indeed, this equation is difficult to apply for CV bias correction without knowing the true value of the CV unless Monte Carlo experiments requiring distribution assumptions are simulated.

Yet, it is possible to use this equation to assess the general magnitude of CV estimation bias by examining ranges of plausible values of the true CV and true CS based on a priori knowledge of sites in a region.

For instance, using the 75% values of the estimated CV (0.61) and CS (1.69) as true values, one obtains the following correction factor for a 50-year peak-flow series:

(0.61)^(3/2)/50[3*sqrt(0.61) – 2(1.69)] = -0.009 = -0.9%*

For the 25% estimated CS (0.62), this rises to just 1.1%. Unfortunately, I cannot compute the bias over the full range of estimated CV and CS values (since only 25%/50%/75% are reported in Table 1). The authors may also want to expand this range given that CS and CV informing this range are estimated values, not true ones.

However, it could end up that the bias in the CV is relatively minor for the range of CV and CS in the study. In this case, the authors could state that after testing plausible CV_true and CS_true values reflecting the range of sites in their study, the adjustments to the CV general did not exceed a low percentage (e.g. 10%), therefore making it reasonable to use common product moment estimators of the CV for the sake of comparing their work with the body of literature that uses this estimator.

It would also be nice to mention that future work should involve the generation of GEV-based bias correction factors using an approach similar to the one that Ye et al. (2020) undertook with the lognormal, kappa, and Wakeby distributions.

With regards to skewness estimation bias, the authors could explore using the GEV-based bias correction method from Carney (2016), although this is really a second-order issue given the pronounced effects that sampling variability can have on skewness coefficient estimates.

In order to inform readers about the magnitude of the bias in the estimation of CV in more detail, we have modified the text in section 2.3 in the following way, using the equations suggested by the referee:

"While the estimation uncertainty of the mean is small, the uncertainty and bias of the estimators of CV and CS (equations 3 and 4) can be substantial. Ye et al. (2020) illustrate the uncertainty and bias in the estimation of CV. The bias in the estimation of CV is relatively small for ranges of CV and CS as in this study (using their equation 2: the bias is at most 0.065 in absolute value, in the case of CV ranging from 0.25 to 0.97 and CS ranging from 0.09 to 3.18, which encompasses 90% of the observed values in this study) making it reasonable to use the common product moment estimator of the CV. "

Regarding skewness, we agree with the referee that the sampling variability is very pronounced for records as short as in the present study. We believe we inform the readers about the estimation uncertainty and bias in the estimation of CS and prefer not to pursue this issue in more detail, as it is not at the heart of the manuscript. Carney (2016) investigates bias in the estimation of parameters of the GEV-distribution via the method of L-moments, whereas this paper does not fit a GEV distribution and focusses on product moments. Extrapolating their results to a bias-correction for the estimation of skewness would require additional attention.

OLS regression model assumptions --

In their response to initial feedback, the authors are correct in stating that regression coefficient estimates are unbiased even when the assumptions of normality and heteroscedasticity are violated. I viewed the need for these assumptions to be evaluated as requisite for making hypothesis testing-based inferences using p-values and other standard *error-based criteria. However, if the goal is to understand the range of coefficient magnitudes without making hypothesis-oriented inferences, then ignoring these assumption evaluations is less critical. However, the authors should make an explicit statement if this is a scope limitation that they would like to establish. If the take this approach, it is yet another reason for them to apply an all subsets modeling strategy in lieu of the stepwise one that they reported, which presumably uses a statistical significance-based criterion in adding and removing variables from the multivariate regression models (see below).*

I appreciate the information in the appendices, and the importance of including the variance inflation factors to prevent excessive multicollinearity among explanatory variables as well.

> We have now added a sentence stating, that we do not look at significance tests for coefficient estimates:
>
> "The OLS-estimator still remains unbiased and consistent under these conditions (Hayashi, 2000), but no inferences such as significance tests of individual coefficients should be made from standard properties of the OLS-estimator. In Tables A.1 and A.2 we report the standard errors of the coefficient estimators, which should be interpreted with care and are thus not used for hypothesis tests. "
>
> The stepwise selection is based on an information criterion (Mallow's Cp) and aims at a good predictive performance respectively a good model fit, but does not rely on a significance-based criterion. We therefore choose to retain the stepwise procedure to derive meaningful covariates for the regional regression models. We also performed additional analyses comparing the results of the stepwise selection procedure and an all subsets modelling strategy, as suggested by the referee. Both procedures select the same variables for the regional regression models for MAF and CV for the data in the present study.

Model building and variable selection --

The authors used a stepwise [forward] selection process to build their multivariate regression models. The leaps R package has an all subsets routine that they could use to check if they missed any strongly performing models by using a stepwise selection process, which does not evaluate all possible combinations of explanatory/predictor variables.

> We now performed additional analysis using the R-package 'leaps', as suggested by the referee. The best performing models for MAF and CV were the same for the stepwise selection and the all subset selection using the leaps-package and Mallow's Cp as the selection criterion.

Temperature as a proxy for snowmelt ---

The authors include two temperature variables (min winter temp and min spring temp) as proxies for snowmelt impacts on annual peak flows. Negative coefficients on their relationship with peak flows assume that colder winters and springs lead to greater snowmelt contributions to flooding. However, how well correlated are winter/spring temperatures with both seasonal snowpack and the rate at which it melts?

At a minimum, the authors should describe some constraints regarding the collection of consistent snowpack depth and snow cover data in Europe as well as limiting assumptions of their use as proxies and efforts to circumvent them. Other examples of regional regression studies using temperature variables as snowmelt proxies would also be helpful.

We now include the following sentence in the discussion, pointing out the limited information content of temperatures as proxies for snowmelt.

"Mean spring and winter temperature were used in the analysis to capture snow processes because of the better data availability. Chaoimh (1998) and Bednorz (2003) identified correlations between spring and winter temperature with snowpack-depth and days with snow-cover and more generally air temperature is often used as an indicator of snowmelt (Ohmura, 2001). Future work could enrich the analysis by using snow data directly, although remote sensing products may have some limitations related to the duration (see e.g. Parajka and Blöschl, 2012). "

Soil moisture data biases ---

The authors write that "Soil moisture (SM) was taken from the CPC Soil moisture database, which contains model-calculated soil moisture values. Fan and Van Den Dool (2004) discuss some biases of the soil moisture data set, which may distort some of the findings here." However, the authors do not comment further on any of these biases/distortions and their implications regarding inferences on process controls of annual peak flows.

The discussion of the soil moisture biases of Fan and Van Den Dool (2004) is rather vague. For example they state: "The results show that the Climate Prediction Center (CPC) global soil moisture data, in spite of its simplicity, simulates the seasonal to interannual variability of observed soil moisture reasonably well in many places." It is therefore difficult to be more explicit about the potential biases and their effects on the results in this paper.

Seasonality analysis not well integrated into manuscript narrative --------------------------------

The seasonality analysis is interesting its own right, but it could be better integrated into the manuscript. A phrase or sentence in the abstract should mention this purpose considering the amount of text devoted to this component of the study.

We have modified the following sentence in the abstract adding information about the the purpose of the seasonality analysis:

"The process controls on the flood moments in five predetermined hydroclimatic regions are identified through correlation and multiple linear regression analyses with a range of covariates and the interpretation is aided by a seasonality analysis."

Writing ---

The authors should aim to reduce the length of the article to roughly 10,000 words by combining sentences, getting rid of unnecessary phrases and wordy language and possibly reducing the discussion of CS given the challenges that sample variability poses to at-site skewness estimates. See the Tracked Changes document for additional writing suggestions.

We are especially grateful for the specific writing suggestions. They are all adopted in the revised manuscript. The main text of the paper, excluding tables, figure captions and the appendix is currently 10,262 words. We have expanded the text following the suggestion of the referees regarding more clarification and it is now 10,800 words, which we believe is close to the target of 10,000.

A minor comment on reproducibility ---

I still think it is valuable from a reproducibility perspective to identify the 22 catchments you omitted from your study due to insufficient covariate data. This does not have to be overemphasized, as you could add a quick list to your supplemental material or data repository.

> We have added a table to the supplementary material indicating the row number of the series in the data of the supplemental material from Blöschl, Hall et al. (2019), which were omitted from the regional regression models.

Table 1: Row-indices of omitted catchments for regional regression models. The row-indices refer to the supplementary material of Blöschl, Hall et al. (2019). Only catchments that were used in their Figure 1 are used in this paper.

221
320
415
416
566
601
630
632
969
2154
2232
3382
3456
3467
3468
3511
3570
3593
3610
3620
3640
3679

References

Bednorz, E. (2004). Snow cover in eastern Europe in relation to temperature, precipitation and circulation. International Journal of Climatology: A Journal of the Royal Meteorological Society, 24(5), 591-601.

Carney, M.C. (2016), Bias correction to GEV shape parameters used to predict precipitation extremes, doi: 10.1061/(ASCE)HE.1943-5584.0001416

Chaoimh, Ú. N. (1998). European snow cover and its influence on spring and summer temperatures. Geographical Journal, 41-54.

Hayashi, F., (2000), Econometrics, Princeton University Press.

Ohmura, A. (2001). Physical basis for the temperature-based melt-index method. Journal of applied Meteorology, 40(4), 753-761.

Parajka, J., & Blöschl, G. (2012). MODIS-based snow cover products, validation, and hydrologic applications. Multiscale Hydrologic Remote Sensing: Perspectives and Applications, edited by: Chang, N.-B. and Hong, Y.

Serago, J. M., & Vogel, R. M. (2018). Parsimonious nonstationary flood frequency analysis. Advances in Water Resources, 112, 1-16.

Ye, L., Gu, X., Wang, D., & Vogel, R. M. (2021). An unbiased estimator of coefficient of variation of streamflow. Journal of Hydrology, 594, 125954.

Technical comments

First column is the number of comment, second column is the line in the annotated word-file provided by the referee, third is the comment of the referee, fourth is the response, fifth are the text changes

		Review comments	**Response**	**Text change**
A1	11	Consider using introductory sentences like this to make the scope and contribution of your work clear from the get-go	Thank you for this suggestion.	Text adopted.
A2	35	Add something about how this can support future work?	Future work is supported by providing a baseline for local studies, as mentioned in the text.	
A3	35	Abstract doesn't say anything about seasonality analysis.	The abstract now includes the seasonality analysis.	The process controls on the flood moments in five predetermined hydroclimatic regions are identified through correlation and multiple linear regression analyses with a range of covariates and the interpretation is aided by a seasonality analysis.
A4	60	More studies than this.	We have added two references.	While USGS regional flood-frequency studies based on observed data have revealed non-climatic controls (Parrett et al., 2011, Paretti et al., 2014, England et al., 2019), most

				knowledge on these effects comes from process-based simulation studies.
A5	62	Do you mean process-based simulation based?	We refer to process-based simulation studies.	Sentence adopted.
A6	66	Including nonstationary ones	Suggestion added.	The role of these variables, can to some extent be inferred from their use as covariates in flood frequency regionalization models (see, e.g. Zaman et al., 2012; Rosbjerg et al., 2013; Miller and Brewer, 2018), including nonstationary ones.
A7	74	This co-evolution doesn’t always favor flood generation though. For instance, a wet year could increase vegetation, which could increase transpiration during the following year, reducing runoff generation.	Thank you for pointing this out, we agree and modified the sentence.	Moreover, climate, vegetation, soils and land forms may co-evolve with MAP, thus exerting a longer-term influence which may increase or decrease floods (Gaál et al., 2012, Perdigão and Blöschl, 2014).
A8	75	Of annual peak flows?	Yes, clarification added.	Farquharson et al. (1992) found CV of annual peak flows (variability between years) to increase with the Aridity Index (the ratio of potential evaporation and MAP).
A9	79	Clarify this earlier	Clarification moved to	

			previous sentence of paragraph.	
A10	83	Pallard drainage area studies: https://hess.copernicus.org/articles/13/1019/2009/	We are discussing here a slightly different point, i.e. CV as a function of area rather than drainage density.	
A11	86	What about the infiltration excess mechanism causes this decrease with area? What about the increase in CV with area in basins where saturation excess overflow dominates? Adding a sentence or so to explain these mechanisms would be helpful.	The authors are not clear about the physical process controlling the scaling of CV. We have therefore chosen to remove this sentence	
A12	120	Check to see if this is referenced later on in paper.	This is not referenced later in the paper.	
A13	120	See General Comment	We address this in the general comment regarding temperature as a proxy for snowmelt.	
A14	123	How biased is it? Discuss this in the discussion?	See our answer to the comment regarding Soil moisture data biases.	
A15	150	So mixed rainfall and snowmelt?	Yes, added clarification.	The Central-Eastern region has a continental climate with cold winters and warm summers and floods mainly occur in spring with snow-melt contributions (resulting in a mixture of rainfall and snowmelt).

A16	Table1	Consider creating three columns for these values	We adopted the suggestion and modified the table accordingly.	
A17	Table1	Evapotranspiration?	Yes, thank you.	Changed to evapotranspiration
A18	Table1	Note this article in your discussion: https://www.researchgate.net/publication/352322291_A_note_on_some_uncertainties_associated_with_Thornthwaite's_aridity_index_introduced_by_using_different_potential_evapotranspiration_methods	While this is interesting in its own right, we feel this would go beyond the scope of the paper.	
A19	161	See general comment about adjustments for nonstationarity	We address this in the general comment regarding nonstationarity.	
A20	174	What do they find? How big of an issue is it?	This is also addressed in the response to the referee's comment regarding the sample moment estimation biases under stationarity. A More detailed calculation is now presented in the manuscript.	While the estimation uncertainty of the mean is small, the uncertainty and bias of the estimators of CV and CS (equations 3 and 4) can be substantial. Ye et al. (2020) illustrate the uncertainty and bias in the estimation of CV. The bias in the estimation of CV is relatively small for ranges of CV and CS as in this study (using their equation 2: the bias is at most 0.065 in absolute value, in the case of CV ranging from 0.25 to 0.97 and CS ranging from 0.09 to 3.18, which corresponds to roughly 90% of

				observed values in this study).
A21	175	Please describe.	We have clarified the text.	Based on a simulation study.
A22	183	See General Comment. Consider stating this at the beginning of the paragraph before getting into estimation biases.	We have a slight preference for leaving this statement where it is.	
A23	191	Any model performance stats to report briefly?	The model performance stats are reported in Table A.6.	
A24	192	Consider enumerating each of these steps	We are now enumerating the steps of our analysis.	
A25	210	Interesting analysis in its own right, but you should describe how you used these results to inform your interpretations of process controls	We added a sentence clarifying the use of the results from the seasonality analysis in the paper.	In the spirit of Blöschl et al. (2017) we used the seasonality of floods to identify dominant flood-generating mechanisms, e.g. spring snowmelt vs winter storms which to some extent explain variations in the flood moments (Merz and Blöschl, 2003).
A26	223	What about all possible subsets approach?	We performed additional analyses and the all possible subsets approach yields exactly the same selection of covariates for all regional regression models for MAF and CV.	
A27	247	Did you also use Cs as a response variable?	No, this is stated at the beginning of section 3.5.	

A28	248	Does this tie in with your analysis of process controls? It seems like it could be a second paper if you validate your ordinary kriging model.	No process controls are used for the ordinary kriging model. It merely serves as comparison for predictions of the regional regression models.	
A29	254	Worth showing other descriptive stats of these values in a table? For instance, sd, min, max?	We believe that the quantiles (25%/50%/75%) give a comprehensive picture of the characteristics of the distribution and the moments and the extremes are perhaps not needed.	
A30	261	How large is the MAF?	We added the relevant information.	On the other hand, the Northeastern region has the smallest average CV and CS and below average MAF (0.39, 0.82 and 0.13 respectively).
A31	285	Consider the inherent relationship between these two here: CS = (X-M)^3/S^3 while CV = S/M If S goes up, |CS| goes down and CV goes up Can you discuss why CS and CV are positively correlated across sites despite the expected relationship between CS and CV at a single site described above?	Yes, a detailed answer is given below the table.	
A32	Fig 4	Consider moving to supplement	We have a total of 10 Figures which we do not consider an excessive number, so	

			perhaps moving Figure 4 to the supplement is not needed.	
A33	382	Check for repetition	We checked for repetitions of this statement and they have been removed in the previous version of the document.	
A34	540	Steeper slopes? Smaller watersheds?	Suggestion adopted.	Further inland, various mountain ranges (Pyrenees, Massif Central, Alps Apennines, Ore mountains, Carpathians, Balkan mountains) stand out with higher MAF than the surrounding areas (mostly above 0.3 $m^3/s/km^2$) and summer as the dominant flood season due to their effects of enhancing rainfall and probably shallower soils as well as steeper slopes and smaller watersheds.
A35	541	Slovenia is close to the Mediterranean Sea (Adriatic Sea) while the Ore mountains are on the German-Czech border. How do highly variable Mediterranean storm tracks drive this pattern?	This is explained in more detail in Hofstätter et al. (2018). Flood-generating storms, such as Vb events tend to have preferred pathways (see e.g. Figure 5 and 6 in Hofstätter et al., 2018), which	

			both affect Slovenia and the Ore mountains. These tracks often extend over more than 1,000 km.	
A36	552	Are you just assuming that they have more nonlinear runoff generation processes because they lie in a region with less mean annual and extreme precip? To support this speculation, could you please add a citation about this general tendency in Europe? Or better, could you cite any work demonstrating nonlinear runoff generation processes in this specific region?	We now cite papers discussing the non-linearity of runoff generation in arid and semi-arid regions of Europe, including Ukraine and Hungary.	Some of the continental regions of Europe (Hungary, Poland, Ukraine) are particularly sheltered by mountain chains, resulting in low precipitation, both at the annual scale and for extreme events, which translates into low MAF and mostly high CV due to the more non-linear runoff generation as compared to wetter regions (Nováaky, 1991, Didovets et al., 2017, Ries et al., 2017).
A37	578	Evidence?	Blöschl and Sivapalan (1997) discuss the process controls on CV. The non-linearity of runoff generation is generally related to the soil moisture status and the reasoning is that during snowmelt floods the soil tends to be wetter than for other types	They

			of floods (Grillakis et al., 2016). The reference has been added to the paper.	
A38	583	How spatially transferrable is this finding? Should recognize need for more research on this given that this study was conducted in just one location.	We have included the referee’s suggestion in the manuscript.	This may be related to possible non-monotonous relationships between CV and area as suggested by Smith (1992), and more complex aggregation effects (Blöschl and Sivapalan 1997), although more research is needed on the transferability of this finding.
A39	607	Can you support this sentence with some of your findings?	We now support this sentence by referring to Figure 6.	While regional studies have suggested that MAP is a better predictor of MAF than other precipitation variables (Mimikou and Gordios, 1989; Merz and Blöschl, 2009) this does not seem to be the case at the European scale (see e.g. Figure 6).
A40	609	Since MAP is a better indicator of soil moisture than P95 and Pmax?	Yes, we added the referee’s sentence for clarification.	On the other hand, CV is always a better correlated with MAP than with P95 and Pmax, reflecting the decreasing degree with which antecedent soil moisture is captured as one moves from MAP

				to P95 and Pmax, since MAP better captures soil moisture conditions.
A41	613	Explain better	We further clarified our statement.	This effect is also represented in the negative correlations between CS and MAP (r=-0.35) and CS and P95 (r=-0.34) (Table A.1.5) in the Mediterranean, indicating a decrease in skewness for comparatively wetter catchments, which is related to a particularly large potential for this contrast in initial conditions.
A42	615	Why?	Given that AI also contains MAP, we decided on P95 as a representative variable for precipitation characteristics to avoid this overlap.	
A43	616	See this note: https://www.researchgate.net/publication/352322291_ A_note_on_some_uncertainties_associated_with_ Thornthwaite's_aridity_index_introduced_by_ using_different_potential_evapotranspiration_ methods	We feel a discussion of the estimation procedures of the aridity index would distract from the results of the paper, in particular given that we are not including a similar discussion for the other	

			variables. Within the levels of correlations obtained and the accuracy of the discharge measurements the effect of different approaches to estimating AI is probably minor.	
A44	623	Both winter and spring air temperatures?	We added more information.	As would be expected, the Spearman correlations between temperature and MAF_{α} and CV are comparatively high in the Northeastern and Central-Eastern region (higher for spring temperature), where snow-processes are important for floods.
A45	628	What about the spatial scale of snowmelt vs. rain events?	The spatial scale of snowmelt floods is usually larger than that of rain-floods, but according to our opinion this aspect does not tie in closely with the argument.	
A46	635	Whereas?	Suggestion adopted.	
A47	638	General comment: variables that explain spatial variability might not be best for explaining temporal variability	While we fully agree with this comment, there might be some symmetry which is worth looking into (Perdigão	

			and Blöschl, 2014).	
A48	706	Didn't you use pre-determined regions from another study?	The regions are also used in Lun et al. (2020).	
A49	725	Citation regarding the ability to capture these processes?	This is discussed in the cited literature i.e. section 3.4 in Boorman et al. (1995), Lilly et al. (1998) and section 3 in Maréchal and Holman (2005).	
A50	733	Your study also excluded sites with pronounced anthropogenic impacts, including these	We excluded sites with heavy urbanization, but not those with deforestation/ afforestation. We have added a comment to explain.	On the other hand: flood changes of small local streams may be much more controlled by land use changes, such as urban development and deforestation (Rogger et al., 2017), only a few of which are included in this study (average catchment size of 2,480 km^2).

A31: Consider the inherent relationship between these two here:

CS = (X-M)^3/S^3 while CV = S/M

If S goes up, |CS| goes down and CV goes up

Can you discuss why CS and CV are positively correlated across sites despite the expected relationship between CS and CV at a single site described above?

The referee is correct in pointing out that in the above equations a ceteris paribus increase in S would result in a decrease of |CS| and an increase of CV. However, from this fact we cannot infer the correlation between these estimators. For small samples exact formulas for the correlation between these two estimators are hard to derive. Instead we use a limit theorem and check its validity for a sample size of 50, which is representative of our data.

For the asymptotic correlation between these two estimators, we use a multivariate central limit theorem for the estimators of the first three non-centralized moments. Using

$$\mu_n' = \mathbb{E}[X^n];\ \mu_n = \mathbb{E}[(X-\mu_1')^n];\ \tilde{\mu}_n = \frac{\mathbb{E}[(X-\mu_1')^n]}{(\mathbb{E}[(X-\mu_1')^2])^{n/2}};CV = \frac{{\mu_2}^{1/2}}{\mu_1'}$$

$$m_n' = \frac{1}{n}\sum_{i=1}^{n} X_i^n\,;\ m_n = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^n\,;\ cv = \frac{\sqrt{m_2}}{m_1'};\ g = \frac{m_3}{m_2^{3/2}}$$

According to a multivariate CLT (assuming iid-observations and 6-th order moments exist, see e.g. example 2.18 in Vaart, 2000)

$$\sqrt{n}\begin{pmatrix} m_1' - \mu_1' \\ m_2' - \mu_2' \\ m_3' - \mu_3' \end{pmatrix} \overset{d}{\rightsquigarrow} N(0,\Sigma)$$

$$\Sigma = \begin{pmatrix} \mu_2' - {\mu_1'}^2 & \mu_3' - \mu_1'\mu_2' & \mu_4' - \mu_1'\mu_3' \\ \mu_3' - \mu_1'\mu_2' & \mu_4' - {\mu_2'}^2 & \mu_5' - \mu_2'\mu_3' \\ \mu_4' - \mu_1'\mu_3' & \mu_5' - \mu_2'\mu_3' & \mu_6' - {\mu_3'}^2 \end{pmatrix}$$

Here $\overset{d}{\rightsquigarrow}$ refers to convergence in distribution. By using the delta method (e.g. Theorem 3.1 in Vaart, 2000) we can obtain the following limit theorem (we omit the calculation steps)

$$\sqrt{n}\begin{pmatrix} m_1' - \mu_1' \\ cv - {\mu_2}^{1/2}/\mu_1' \\ g - \mu_3/{\mu_2}^{3/2} \end{pmatrix} \rightarrow N(0,\Sigma^*)$$

$$\Sigma^* = \begin{pmatrix} \mu_2 & \frac{\mu_2}{\mu_1'}\left[\frac{\tilde{\mu}_3}{2} - CV\right] & {\mu_2}^{1/2}\left[\tilde{\mu}_4 - \frac{3}{2}{\tilde{\mu}_3}^2 - 3\right] \\ \frac{\mu_2}{\mu_1}\left[\frac{\tilde{\mu}_3}{2} - CV\right] & CV^2\left[\frac{\tilde{\mu}_4 - 1}{4} + CV^2 - \tilde{\mu}_3 CV\right] & CV^2\left[-\frac{5\tilde{\mu}_3}{4\,CV} - \frac{3\tilde{\mu}_3\tilde{\mu}_4}{4\,CV} - (\tilde{\mu}_4 - 3) + \left(\frac{3}{2}\right){\tilde{\mu}_3}^2 + (1/2)\frac{\tilde{\mu}_5}{CV}\right] \\ {\mu_2}^{1/2}\left[\tilde{\mu}_4 - \frac{3}{2}{\tilde{\mu}_3}^2 - 3\right] & CV^2\left[-\frac{5\tilde{\mu}_3}{4\,CV} - \frac{3\tilde{\mu}_3\tilde{\mu}_4}{4\,CV} - (\tilde{\mu}_4 - 3) + \left(\frac{3}{2}\right){\tilde{\mu}_3}^2 + (1/2)\frac{\tilde{\mu}_5}{CV}\right] & 9 - 6\tilde{\mu}_4 + \frac{9}{4}\tilde{\mu}_4{\tilde{\mu}_3}^2 - 3\tilde{\mu}_3\tilde{\mu}_5 + \frac{35}{4}{\tilde{\mu}_3}^2 + \tilde{\mu}_6 \end{pmatrix}$$

The correlation between the estimators of CV and CS depends on higher-order moments of the underlying distribution. $\tilde{\mu}_n$ refers to standardized moments (skewness, kurtosis, etc.). From this result, which is also documented in Bobee (1973), we can calculate the correlation between the estimators of CV and CS from the asymptotic distribution (simply as $\frac{\Sigma_{2,3}^*}{\sqrt{\Sigma_{2,2}^* \cdot \Sigma_{3,3}^*}}$), which we can use as a proxy for small-sample results.

A quick calculation (higher-order moments of the GEV can be calculated with the formulas from Muraleedharan et al., 2011) reveals that these results predict a correlation of about 0.64 for the average parameter configuration of the study data, assuming a GEV-distribution as the data-generating process. Here we use the parametrization of the GEV as in chapter 18 of Maidment (1993). Using the average moments of the data as the population moments ($\mu_1' = 0.17, CV = 0.52, \tilde{\mu}_3 = 1.28$) the parameters of the GEV correspond to $\xi = 0.13, \alpha = 0.07, \kappa = -0.02$. The higher-order moments correspond to $\tilde{\mu}_4 = 6.13,\ \tilde{\mu}_5 = 24.11,\ \tilde{\mu}_6 = 132.61$. All numbers here are rounded to two digits.

A quick simulation study (100,000 times generating 50 observations from a GEV with the parameters as specified above and calculating the correlation between the estimates of CV and CS) results in an empirical correlation that is very close to the value predicted above (around 0.66 instead of 0.64 in the simulation). This result indicates that the asymptotic result is suitable for the sample size and parameter configuration considered here.

Considering this, the positive correlation between estimates of CV and CS across sites, assuming a GEV as the data-generating process and considering the average empirical moments of the study data, does seem plausible.

References

Bobee, B. (1973). Sample error of T-year events commuted by fitting a Pearson type 3 distribution. Water Resources Research, 9(5), 1264-1270.

Boorman, D. B., Hollis, J. M., & Lilly, A. (1995). Hydrology of soil types: a hydrologically-based classification of the soils of United Kingdom. Institute of Hydrology, Wallingford, UK.

Didovets, I., Lobanova, A., Bronstert, A., Snizhko, S., Maule, C. F., & Krysanova, V. (2017). Assessment of climate change impacts on water resources in three representative Ukrainian catchments using eco-hydrological modelling. Water, 9(3), 204.

England, J.F., Jr., Cohn, T.A., Faber, B.A., Stedinger, J.R., Thomas, W.O., Jr., Veilleux, A.G., Kiang, J.E., and Mason, R.R., Jr., 2019, Guidelines for determining flood flow frequency—Bulletin 17C (ver. 1.1, May 2019): U.S. Geological Survey Techniques and Methods, book 4, chap. B5, 148 p., https://doi.org/10.3133/tm4B5.

Grillakis, M. G., Koutroulis, A. G., Komma, J., Tsanis, I. K., Wagner, W., & Blöschl, G. (2016). Initial soil moisture effects on flash flood generation–A comparison between basins of contrasting hydro-climatic conditions. Journal of Hydrology, 541, 206-217.

Hofstätter M., Lexer A., Homan M. and G. Blöschl (2018) Large-scale heavy precipitation over central Europe and the role of atmospheric cyclone track types. International Journal of Climatology, 38, pp. e497– e517, https://doi.org/10.1002/joc.5386

Lilly, A., Boorman, D. B., & Hollis, J. M. (1998). The development of a hydrological classification of UK soils and the inherent scale changes. In Soil and Water Quality at Different Scales (pp. 299-302). Springer, Dordrecht.

Lun, D., Fischer, S., Viglione, A., & Blöschl, G. (2020). Detecting flood-rich and flood-poor periods in annual peak discharges across Europe. Water Resources Research, 56(7), e2019WR026575.

Maidment, D. R. (1993). Handbook of hydrology (No. 631.587). McGraw-Hill,.

Maréchal, D., & Holman, I. P. (2005). Development and application of a soil classification-based conceptual catchment-scale hydrological model. Journal of Hydrology, 312(1-4), 277-293.

Muraleedharan, G., Soares, C. G., & Lucas, C. (2011). Characteristic and moment generating functions of generalised extreme value distribution (GEV). In Sea Level Rise, Coastal Engineering, Shorelines and Tides (pp. 269-276). Nova.

Nováaky, B. (1991). Climatic effects on the runoff conditions in Hungary. Earth surface processes and landforms, 16(7), 593-599.

Parrett, C., Veilleux, A., Stedinger, J. R., Barth, N. A., Knifong, D. L., & Ferris, J. C. (2011). Regional skew for California, and flood frequency for selected sites in the Sacramento-San Joaquin River Basin, based on data through water year 2006. U. S. Geological Survey.

Paretti, N. V., Kennedy, J. R., Turney, L. A., & Veilleux, A. G. (2014). Methods for estimating magnitude and frequency of floods in Arizona, developed with unregulated and rural peak-flow data through water year 2010 (No. 2014-5211). US Geological Survey.

Perdigão, R. A. P., and G. Blöschl (2014) Spatiotemporal flood sensitivity to annual precipitation: Evidence for landscape-climate coevolution, Water Resour. Res., 50, 5492-5509, doi:10.1002/ 2014WR015365.

Ries, F., Schmidt, S., Sauter, M., & Lange, J. (2017). Controls on runoff generation along a steep climatic gradient in the Eastern Mediterranean. Journal of Hydrology: Regional Studies, 9, 18-33.

Smith, J. A. (1992). Representation of basin scale in flood peak distributions. Water Resources Research, 28(11), 2993-2999.

Van der Vaart, A. W. (2000). Asymptotic statistics (Vol. 3). Cambridge university press.