Hydrology and
Earth System
Sciences
Discussions

Interactive comment on "Characteristics and process controls of statistical flood moments in Europe – a data based analysis" *by* David Lun et al.

Anonymous Referee #2

This manuscript has the potential to serve as a strong reference for characterizing the spatial variability of annual peak-flow moments at sites without strong anthropogenic modifications, such as reservoirs, throughout Europe. The leave-one-out cross-validation of a multiple regression model predicting flood moments (mean, Cv, Cs) suggests that, with follow-up efforts, this work could be used to estimate flood moments at ungaged locations with reasonable accuracy in many locations. This work also documents large-scale spatial patterns in controls on flood moments throughout the continent, although the process controls revealed are not especially surprising to people with knowledge of European hydrology. However, numerous technical and presentation improvements detailed below are needed to make this manuscript publish-

able in HESS. In addition, a more compelling case for how this research could benefit both stationary and nonstationary flood-frequency analysis would be helpful. I have also attached a Tracked Changes Word document with more specific writing and presentation suggestions and some more minor technical inquiries.

SOME BASIC CHARACTERISTICS OF THE FLOOD TIME SERIES NEED TO BE CLARIFIED UPFRONT. The authors should state in their abstract whether their set of 2,370 flood series are from stations in anthropogenically impacted basins and whether the "maximum annual flows" they analyze are daily mean flows or instantaneous peak flows. This is important given the small drainage areas of some basins. The authors state that they used the version of the European Flood Database used in Blöschl et al. (2019), which excluded catchments with strong human modifications, such as reservoirs, but did not exclude basins subject to more local anthropogenic perturbations – given their focus on elucidating broad regional patterns. While this dataset contains both [instantaneous?] peak flows and maximum daily mean flows in each year, it seems like the authors might have strictly used peak flows based on descriptions at the beginning of Section 2: "This study uses the data set of European flood discharges of Blöschl, Hall et al. (2019), which . . . consists of 2370 annual maximum peak discharge series from 33 countries". Also, the authors only used 2,370 stations whereas Blöschl et al. (2019) used 3,783. This discrepancy should be explained briefly. Finally, the authors should clarify earlier in the manuscript whether they used calendar years or a designated water year when identifying annual peaks.

MOMENT ESTIMATION BIASES MUST BE ADDRESSED. The authors need to discuss the bias in their estimates of the Cv (coefficient of variation) and Cs (coefficient of skewness). First, with regards to the Cv, Ye et al. (2020) demonstrated the extent to which common Cv estimators can be biased when data are skewed or do not adhere to i.i.d. assumptions. While the degree of bias is not as pronounced as it is with daily flow data, quick calculations using the equations described in this paper demonstrate that Cv of annual peak flows can have a substantial bias. Numerous references have

also demonstrated the bias of skewness estimates from small samples, including their dependence on record length (Wallis et al., 1974; Bobee and Robitaille, 1975; Carney, 2016). In their discussion, the authors should also recognize the literature on regional skewness coefficients as well as the weighted skewness approaches combining at-site and regional information that the U.S. Geological Survey employs.

NONSTATIONARITY AND ITS POTENTIAL IMPACTS ON MOMENTS MUST BE CONSIDERED. Blöschl et al. (2019) reported regional-scale climate-driven trends in northwestern, southern, and eastern Europe (see Fig. 1). Is it worthwhile to describe the sample moments of sites without considering these changes? In my opinion, the authors should either develop a procedure to exclude sites subject to trends or provide a rationale for treating all sites as stationary given their research goals. In making this decision, the authors should consider the ongoing shift from nival to pluvial regimes in 3/5 regions in Europe makes this an important consideration. If they wish to distinguish trends from long-term persistence, an argument often used to refute nonstationary treatments of hydrologic records, the authors could test for trends of a given trajectory against null hypotheses of long-term persistence (see Matalas and Sankarasubramanian, 2003; Cohn et al., 2005). The authors should also note trends in both the mean and variance affect both Cv and Cs estimates [see Serago and Vogel (2018) for some initial guidance for making these adjustments. Hecht and Vogel (2020) offer one approach for modeling trends in variability and reference a handful of other moment-based ones, including Strupczewski et al. (2001).

THE RESIDUAL BEHAVIOR OF THE REGRESSION MODELS MUST BE EVALUATED. The authors do not report the normality, heteroscedasticity, autocorrelation of their residuals. They also do not report the variance inflation factor or alternatives measures of multicollinearity for their multiple regression equations. This is especially important if one is making process-based inferences using covariate matrix-derived statistics from regression models. The authors should consider using a Supporting Information (SI) section to display the residual behavior of their models. Also, the authors

report the tendency for large MAFs to be underestimated and small ones to be overestimated. This suggests that residuals might not be homoscedastic and that another covariate may be needed to produce a multiple regression model that meets the homoscedasticity requirement for making inferences from standard error-based metrics (Helsel et al., 2020).

THE CHOICE BETWEEN LOG-SPACE VS. REAL-SPACE MOMENTS SHOULD BE RECOGNIZED. The authors should also recognize in their manuscript that moments of log-transformed floods are often used in FFA and clarify that real-space moments are used upfront.

MIXED POPULATIONS SHOULD BE CONSIDERED IN THEIR INTERPRETATION OF RESULTS. While the authors somewhat recognize mixed populations (e.g. description of Alps and Norwegian coast flood-generating processes), they compute moments assuming floods at each site belong to homogenous populations. While statistically evaluating the presence of mixed populations at individual sites lies beyond the scope of this paper, it is important to consider mixed populations explicitly when interpreting results and to caution readers about problems associated with choices to neglect them at individual sites. While the authors use an analysis of flood timing to help identify drivers of floods, they do not specifically check for the presence of multi-modal peaks suggesting mixed distributions in them. This type of quicker analysis could support some of the good observations that authors make about mixed distributions in specific regions. Finally, the authors should communicate an awareness of this 'mixed populations' literature in their discussion of mixed populations.

MORE DETAILS ABOUT THE DRAINAGE AREA-NORMALIZED EQUATION(S) ARE NEEDED. I like the authors' idea of normalizing their analysis to a given drainage area (100 km2) since drainage area is still an important descriptor of flood-generating processes even when specific discharge values are used to express peak flows. However, it would be nice to report goodness-of-fit measures for this model and show the fit graphically, the latter which can be done in the SI section if space constraints remain.

The authors also describe the creation of equations that establish values of the Cv and Cs for 100-km2 drainage areas, but it is unclear if these DA-adjusted values are ever evaluated as response variables in the multiple regression models.

NON-MONOTONIC RELATIONSHIPS WITH COVARIATES SHOULD BE CHCKED - AT LEAST IN AN EXPLORATORY DATA ANALYSIS. The authors raise the possibility of non-monotonic relationships between moments and catchment descriptors in discussions of prior findings, but they only examine monotonic relationships in their linear regression models. In particular, they cite Smith et al. (1992), who found that floods in the Appalachian mountains in the eastern US demonstrated an increase in the CV with catchment area for catchments up to 100 km2 and then exhibited a decrease with catchment area in larger basins. They also recognize that Wang et al. (2017) found a non-monotonic relation between water body size and the Cv. In addition, Pallard et al. (2009) also found that Cv decreases with drainage density in catchments with sparse drainage networks but then increase after a reaching a minimum. I think that if the authors can claim that exploratory data analyses did not demonstrate any non-monotonic trajectories like these, they don't need to formally test hypotheses of non-monotonic change with statistical models, but they should briefly demonstrate that they performed exploratory data analysis (EDA) justifying the monotonic relationships they modeled.

DOES ARIDITY CAUSE FLOOD VARIABILITY? The authors make an important association between the aridity index (AI) and the Cv of annual floods. However, it is important to recognize more succinctly that arid regions tend to have greater interannual precipitation variability, and, for that reason, arid basins tend to have larger Cv's. This is important when considering the implications of these findings under climate change. If a region becomes drier, it's interannual precipitation variability will not necessarily increase. A discussion about the implications of these cross-sectional findings for projecting flood responses of environmental changes at a given location over time would enrich the paper.

ORDINARY KRIGING. Ordinary kriging visualizes broad regional patterns but may be

limited for applications in ungaged basins. The kriging results look visually pleasing and achieve the goal of illustrating broad regional patterns in flood moments. However, what if nearby basins have greatly different drainage areas (since this is stated to be a map of MAF and not MAF[alpha]) or pronounced differences in other catchment characteristics that can change abruptly? In the future, the authors could consider kriging in attribute space instead of geographic space. If the authors retain these kriged maps to display broad regional patterns, they should note the limitations of using these interpolations for characterizing flood regimes at ungaged sites. To me, it seems like the regression equations should work reasonably well for estimating moments at many sites. And if they choose to argue that kriging can be used to estimate moments in ungaged basins, then a more formal cross-validation analysis and more detailed reporting of model performance is necessary. Alternatively, they could turn this kriging exercise into a separate paper.

IMPORTANCE FOR FFA IN PRACTICE. This paper successfully elucidates broad regional patterns in flood moments across Europe. Their leave-one-out cross-validation suggests that flood moments can be reasonably estimated in many regions at sites whose covariate values are known. The implications of these errors for design flood estimates could be made stronger by computing the design floods with a GEV quantile function (noting issues with this distribution in specific regions from prior studies, such as Salinas et al. (2014)) using moments estimated from observations and from the multiple regression models. The authors should also address practical concerns regarding nonstationarity described above. In addition, the authors should note the contribution that their study makes to improve upon other recent prediction in ungaged basins efforts in Europe.

OVERALL PRESENTATION. The paper reads a bit like a lab report in places and generally has the potential to be shortened considerably without losing much content. In some places, starting paragraphs with more topic sentences could help orient the reader better and curtail the 'rambling' nature of some sections, such as the bivari-

ate correlation results. The correlation analysis is important for interpreting regression model results and many of the insights on multicollinearity in the data are good, but the presentation of it should be a bit more focused on supporting the multiple regression model analysis and not a comprehensive review of the entire correlation matrix. The submission also requires more editing for fluidity/conciseness and proper punctuation. While I made some writing and grammar suggestions in the Track Changes document, I did not perform a comprehensive check for these issues and suggest that the authors find someone else who can do that.

REFERENCES. This list contains references not already cited in the submitted manuscript:

Bobee, B., R. Robiataille (1975) Correction of bias in the estimation of the coefficient of skewness, Water Resour. Res., doi: 10.1029/WR011i006p00851

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

Cohn, T.A., H.F. Lins (2005), Nature's style: naturally trendy, Geophys. Res. Lett., 32, L23402, doi:1029/2005GL024476

Hecht, J.S., R.M. Vogel (2020), Updating urban design floods for changes in central tendency and variability using regression, doi:10.1016/j.advwatres.2019.103484

Helsel, D.R., Hirsch, R.M., Ryberg, K.R., Archfield, S.A., and Gilroy, E.J., 2020, Statistical methods in water resources: U.S. Geological Survey Techniques and Methods, book 4, chapter A3, 458 p., https://doi.org/10.3133/tm4a3. [Supersedes USGS Techniques of Water-Resources Investigations, book 4, chapter A3, version 1.1.]

Matalas, N.C., A. Sankarasubramanian (2003), Effect of persistence on trend detection via regression, Water Resour. Res., doi:10.1029/2003WR002292.

Pallard, B., A. Castellarin, and A. Montanari (2009), A look at the links between drainage density and flood statistics, Hydrol. Earth Syst. Sci., 13, 1019-1029,

doi:10.5194/hess-13-1019-2009

Salinas, J.L., A. Castellarin, A. Viglione, S. Kohnova, and T.R. Kjeldsen (2014), Regional parent flood frequency distributions in Europe – part 1: is the Gev model suitable as a pan-European parent? Hydrol. Earth Syst. Sci., 18, 4381-4389, doi: 10.5194/hess-18-4381-2014

Serago, J.M., R.M. Vogel (2018), Parsimonious nonstationary flood frequency analysis, Adv. Water Resour., 112, 1-16, doi: 10.1016/j.advwatres.2017.11.026

Strupczewski, W., Z. Kaczmarek (2001), Non-stationary approach to at-site flood frequency modelling II. Weighted least squares estimation, J. Hydrol., 248(1-4), 143-151, doi: 10.1016/S0022-1694(01)00398-5

Wallis, J.R., N.C. Matalas, J.R. Slack (1974) Just a moment! Water Resour. Res., doi: 10.1029/WR010i002p00211

Ye, L., X. Gu, D. Wang, and R.M. Vogel. An unbiased estimator of coefficient of variation of streamflow. J. Hydrol., 594, doi: 10.1016/j.jhydrol.2021.125954

I wish the authors good luck with their revision and am happy to discuss any of these issues!

Please also note the supplement to this comment:
https://hess.copernicus.org/preprints/hess-2020-600/hess-2020-600-RC2-supplement.pdf