Long Yang, Lachun Wang, Xiang Li, and Jie Gao, "On the Flood Peak Distributions over China", submitted to Hydrology and Earth System Sciences (hess-2019-322).

This paper presents an investigation of the characteristics of annual maximum streamflow using a comprehensive database of more than 1000 stream gauges across China. The authors also related flood occurrences and associated changes to catchment regulations and other climatic drivers with a focus on tropical storms. The authors' effort in collating and analysing streamflow data from different sources is a significant contribution to large-scale hydrology. I'm particularly impressed by the database being used, and found the paper has the potential to make significant contribution to the field, at both regional and global scales.

To make the contributions of this study really come through, the authors must address several major issues. Most of these issues belongs to the structure and content, but there are also some methodological issues. I outlined these major issues in "General Comments" section, followed by more specific comments to further clarify my concerns. I'm recommending a major revision to provide the authors sufficient time to thoroughly revise the manuscript.

NOT CONTENT RELATED

Perhaps this is an editorial decision but I found the bottom-left panel of Figure 1 (the marine region southern of China) irrelevant to the scientific content of this paper, as there was no gauge located in that area. Please also note that there is political tension around this region due to ongoing territorial disputes involved several countries (see https://en.wikipedia.org/wiki/Territorial_disputes_in_the_South_China_Sea). As political science is not the focus of this study, I strongly recommend omitting this feature from the map to avoid unnecessary debates.

GENERAL COMMENTS

1. Scientific writing issues

The nature of an analogous study

This paper shares a significant similarity to that of Villarini and Smith (2010) in terms of manuscript structure, narrative, methodology and content. As indicated at the beginning of my review, the findings are still significant contributions to the literature as this study used an unprecedented database for China and have improved the state-of-understanding of flood hazard at the regional and global scale. Nevertheless, I strongly recommends the authors to revise the manuscript substantially (i.e. making the manuscript more independent to its counterpart for the eastern of US) as the presented study has more potential than a simple incremental research.

I would also consider it appropriate to highlight the link between this analysis to similar works across the globe to place this contribution in a broader context of current assessments and datasets being used worldwide to assess flood peak characteristics. There are some studies for other regions mentioned in the introduction, but they have not been discussed systematically to highlight the significance of a China-wide investigation. Considering the quality/magnitude of the database and analyses presented, I consider the study's most significant contribution (which unfortunately was not adequately highlighted in the current manuscript) is to complement the limited understanding of flood characteristics at the global scale (e.g. changes in flood magnitude and timing). Specifically, recent observation-based findings on flood hazard magnitude and timing (Do et al., 2019;Do et al., 2017;Hodgkins et al., 2017;Burn and Whitfield, 2018;Mangini et al., 2018;Hall and Blöschl, 2018;Slater and Villarini, 2016) have not provided much information for Asia (note that China contributes a significant share to Asia land area). The key barrier is the insufficient streamflow records presented with the most comprehensive global datasets to-date such as GRDC or GSIM (Do et al., 2018a;Gudmundsson et al., 2018a), and thus large-scale hydrologists do not possess a representative sample of streamflow observation (especially for Asia). Current understanding of global flood hazards may therefore be biased toward Europe and North America, making studies similar to the present study very demanding.

If the authors prefer to stay on the current objective (i.e. conducting an analogous study of a previous study), I want to see this objective introduced more prominent (e.g. explain why a nearly identical study was conducted) as it is currently mentioned very vaguely in the introduction (page 2, lines 35-37). This potentially makes a notion that the authors did not give Villarini and Smith (2010) the merits it deserves. The authors should also focus on clarifying the technical aspects that were identically repeated and the ones that were done differently (beside different datasets and study areas) – the rationale of these choices should also be discussed. In addition, "distributions" in Villarini and Smith (2010) refer to several terminologies (e.g. spatial, temporal and statistical distributions) and the rationale of using this word was presented adequately in the introduction of that paper. The authors must rethink the key motivation of their study to clarify which "distributions" are the focus of their research and justify why a specific characteristic of flood peaks is chosen (similar to the link between Typhoon Nina and timing of flood peaks currently presented). The loose ties between the "distributions" and the objectives currently made the paper read fragmented and confused.

If the authors decide to follow another approach, the study may excluded some analyses (e.g. GEV distribution) to focus more on an aspect of flood hazards (e.g. trends in flood magnitude and timing) and relate the paper findings to previous observation-based studies. As indicated above, with this unprecedented dataset, I believe there are several rooms for analyses and discussions beyond the current strategy (i.e. an analogous study of Villarini and Smith (2010)). There are opportunities for deeper and more critical discussions included in the "Specific comments" section that are applicable if the authors choose this direction.

Abstract:

This section should be revised significantly to provide a clear synthesis of the objectives, the methods being used and the main findings. The opening sentence reiterates the ambiguous terminology presented in the title (i.e. "flood peak distributions") without sufficient clarifications in the rest of the abstract. There are several statistical analyses presented with little "hints" about the role of each assessment as well as the links between the results and specific objectives. This shortcoming has led to the notion that the paper is a collection of unrelated analyses (which is not the case when reading further to the main text).

Introduction

After reading through the introduction twice, I'm still unclear about "what are the research questions being introduced?", and "what are the key contributions of this study?" The confusions partly come from the loose link between the ideas being presented and the choice of terminologies/wordings that sometimes overcomplicates the ideas. This section must be revised to provide a more synthesized literature review and simplify the ideas where possible. Please also refer to my previous comments for some thoughts about improving the introduction. The authors may find some of the "specific comments" useful as well.

Results and Discussions

This section was too “report focused” with very loose connection between individual sub-sections. This led to a notion that the “Results and Discussion” is a compilation of four separate studies. I strongly recommend the authors to remedy this issue and better explain the contribution of each sub-section to the overarching aims. One possible way is to have an “introduction sentence/paragraph” at the beginning of each sub-section to clarify how the subsequent discussion contributes to a better understanding of “flood peak distributions for China”.

Conclusions

This section reads more “Summary” than “Conclusion”. It would be more useful if the authors be more concise in summarizing the key findings and then focus on highlighting (1) what was the contribution of this study to the state-of-understanding for flood hazard (China-wide and global scale) and (2) the fact that collected data set is unprecedented and has much more potentials beyond this study (e.g. adding values to the literature of global scale hydrology). It would beneficial to also comments on the benefit, challenges and potential of making China streamflow data becomes more FAIR (Findable, Accessible, Interoperable and Reusable; see Wilkinson et al., 2016). Some potential options toward this ambitious goal is to publish metadata and indices (Do et al., 2018b;Gudmundsson et al., 2018b) and even include hydrological variable time series (Addor et al., 2017).

2. Methodological issues

There are some data aspects that should be clarified in Section 2 (or presented in supplementary) as this is the first study to use such dataset, which may be unfamiliar to other hydrologists. Some recommended clarifications:

- Data quality statements. The definition of “continuous records of at least 50 years” (e.g. non-missing data consecutively for 50 years) and “Strict quality control procedures are implemented to ensure consistency and accuracy of the records”. These aspects are very important on the credibility of detectable trends.
- Filtering criteria related to missing data. Please note that annual value of maximum are very sensitive to missing data. Previous studies usually applied a threshold of number of missing data points per year to assign “N/A” value for a specific year (e.g. station A has 30 missing data points in 1950 so N/A value was assigned for that particular year). If all gauges have zero missing data-point, this feature should be highlighted to show the exceptionally good quality of this dataset.
- A graph presents the number of stations with available data over time may be useful. This will also provide the an understanding of the uncertainty related to step-change analysis. Figure 2 of Do et al. (2017) is an example for this type of plot.
- Figure 1 should show the location of "discarded stations" prior to further analyses. It seems to me that the station density reduced significantly in other figures. If all stations showed in Figure 1 were used, please note in the caption. If there are stations removed, the number of stations used/removed for each subsequent analysis should also be mentioned.
- Additional maps should be added to Figure 1 to show (i) data length, (ii) the beginning and (iii) end of records for each station to complement the subsequent analysis of change-point. For instance, one hypothesized reason for change point detected mostly over 1980-2010 was most stations have streamflow data available for only 1970s onward.

Trends analysis: as a China-wide investigation is one of the key motivations of this study, I’m recommending the use of “reference period”. The authors may conduct the analysis for only one period (e.g. all stations were assessed for 1969-2019) or different period (e.g. 1980-2010, 1950-2019, and 1900-2019). This is particular important to remedy data limitation, which could dismiss the usefulness of the detected trends wherever too short records was used to assess trend in floods over a long period (e.g. using 50 years of data to infer changes in floods for a 150-year period).

For Mann-Kendall test: is there any criterion for data length to conduct the analysis (e.g. at least ten data points)? This is relevant for the test conducted for before-step-changes and after-step-changes time series (there may be insufficient data points). I noted that the figures 5b and 5c have different numbers of stations, potentially due to insufficient data points for a specific sub-group?

For flood timing, it is unclear the motivation for assessing trends in this flood index. In addition, it is more useful to use an analysis where the magnitude of changes is visible. The reason is even when changes are significant, it is not practically meaningful if flood timing shifts only a small time-window (e.g. one day over 100 years). Theil-Sen slope estimator is a useful statistical technique for this type of analysis (Blöschl et al., 2017).

SPECIFIC COMMENTS

Title: “Flood Peak Distributions over China” made the notion that “spatial distribution” is the key feature being analysed (which is not the case).

Line 2: “flood peak distributions across China” in the abstract also made the notion that “spatial distribution” is the key feature being analysed.

Line 16: the research objectives were “to provide improved understandings on the nature of upper tails of flood peaks and innovative methods for flood frequency analysis in a changing environment”. This statement has two issues: (1) it is unclear what is “the nature of upper tails of flood peaks”? (Any alternative for “upper tails of flood peaks”, which was used quite often across the manuscript?), and (2) “provide innovative methods for flood frequency analysis” indicates the development of a new method, which is not the focus of this paper. The goal should be rewritten.

Line 17: It is unclear how the four presented themes linked to the two objectives.

Line 26: “staionarity” should be “stationarity”.

Line 35: “highlight possible factors that induce the changes” should be clarified: some of potential factors and why did they are chosen?

Line 36: it is unclear to me what “dominant modes of violation for the stationarity assumption” is. Would there be a more simple way to explain it?

Line 37: the final sentence seems out of place. I was expecting a clarification of the “dominant modes...” mentioned it the previous sentence, or justification of why the authors followed Villarini and Smith (2010) rather than Hodgkins et al. (2019).

Line 46: it is unclear what “space-time rainfall organizations” means.

Line 48: I don’t think the sub-sequent sections discussed anything related about “the necessity of improved procedures for regional flood frequency analysis with spatial heterogeneity in flood hydro-climatology considered”. Please revise this statement or extend the discussion.

Lines 52: please note that “seasonality” may refer to more than the timing of floods (Villarini, 2016).

Lines 57-62: please focus more on introducing the mechanisms generating floods across China.

Line 63: it is unclear what “monsoon-related systems” means.

Lines 64-67: these two sentences read out of place.

Line 67-68: this section is repetitive. I’m recommending the authors to link this paragraph to the subsequent one (starting at line 70) as timing of tropical cyclones were then related to flood timing.

Lines 70-92: I found the detail introduction of Typhoon Nina is too disruptive – and a distraction. This is also a reason for confusion about the key research question of this research. If the key motivation was “to examine the impact of tropical cyclone on the upper tail properties of flood peak distribution over China” (line 85; and I think this is a great research question itself), the structure of the manuscript need to be revised to reflect this main research objective. The current manuscript presents the link between floods and tropical storms at the very end (section 4.4), after a very long discussions for other factors (e.g. step-changes, seasonality, GEV...) with almost no “reminder” for readers. At the time I got to section 4.4, I almost forgot the motivation to analyse “Tropical cyclones and upper tail properties” in this study.

Line 88: I’m not sure why “Results presented in this study can promote a predictive understanding of flood hazards associated with landfalling tropical cyclones”.

Line 121: please explain the rationale of applying Mann-Kendall test for two sub-groups and the contribution of this analysis to the main objective.

Line 124: “significance level of 5%” is for one-tail or two-tail?

Line 125: please justify why circular statistic, a common approach for flood timing investigations (Villarini, 2016;Blöschl et al., 2017;Hall and Blöschl, 2018), was not used in this study to prevent the issues where two flood peaks occurring on calendar day 1 and calendar 365 only have one day difference.

Lines 136-140: this is repetitive (has been mentioned in introduction). The authors should clarify about how to examine the dependence of GEV parameters on drainage area.

Line 145: is there any reference for this choice (i.e. 500km and two weeks).

Lines 149-155: these information seems irrelevant as only the circulation centre location was used. In case the sub-sequent analyses will divide tropical storms into sub-categories (ET/TS), the authors should clarify this technical aspect.

Line 160: “The majority of stations tend to show smaller values”. Please clarify what is “the majority” and how “smaller” the values are (perhaps in %).

Line 169: “We are able to relate some of the changes in annual flood peaks series to intentional human activities”. Please clarify the procedure used to identify these relationships (e.g. metadata inspection?) and what is the magnitude of “some” (e.g. 5% of stations with significant step-change detected?).

Figure 3: Please also show the locations of gauges with insignificant results. The geographical location of individual stations analysed in Figure 4 should also be highlighted (e.g. starred symbol).

Line 197: this is somewhat expected as changes in climate variables occur quite gradually in general. Previous studies generally link abrupt changes to human interventions rather than natural climate drivers.

Line 203: please be careful with your conclusion that “Abrupt change rather than slowly varying trend is a common mode of the violation of the stationarity assumption for the annual flood peak series over China” as both modes of non-stationarity can present at the same station. For instance, “naturalized streamflow records” (e.g. the difference in means between two before/after step-change time series were removed) may reveals gradual change, which is more relevant to climate changes and variabilities. As a result, assessing linear trends over only stations that did not exhibit significant step-change is not sufficient to support this statement.

Figure 5b and Figure 5c: please clarify why these two figures have different number of data points. If it is due to statistical insignificance, please also show the location of stations that did not exhibit significant linear trends.

Line 227: Please note that the findings do not support the statement that “external climate factors (i.e., extreme rainfall), and changes in soil moisture on flood hydrology” leads to flood stationarity.

Lines 232-235: it is unclear what “state-of-art process-based approaches” and “statistical modelling approaches” are – please clarify. Please also make it clearer why these approaches important to “flood frequency analyses across China” (i.e. how could these approaches address the non-stationarities in flood frequency).

Section 4.2: maps of the average flood timing and associated concentration (Villarini, 2016) would be a nice addition.

Lines 239: “the first peaks”? Considering the distribution of the floods timing (Figure 6), I thought there is only one peak per group?

Lines 240-241: please provide reference.

Line 252: please clarify how “tropical cyclone floods” defined (this should be presented in methodology).

Line 260: forward-reference is not recommended.

Lines 265-280: without sufficient evidence of “how many days flood peaks have shifted”, it is hard to justify these statements.

Line 271: “Villarini (2016) found ...” sounds out of place.

Lines 273-278: these statements read contradicting to each other.

Line 287: please clarify why only these stations used (should explain in the methodology).

Line 312: “contrasting space-time organizations” is unclear. Please clarify.

Line 343: the presented results only show the impacts of tropical cyclone on flood occurrence rather than “flood peak distributions”.

Figure 12: please plot also the stations within 500-km distant and the annual maximum streamflow does not coincide with the occurrence of the selected tropical cyclones (i.e. flood timing is outside the two weeks threshold) and extend the discussions appropriately (e.g. the proportions of stations influenced by tropical storms).

References

Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology for large-sample studies, Hydrol. Earth Syst. Sci. Discuss., 2017, 1-31, 10.5194/hess-2017-169, 2017.
Blöschl, G., Hall, J., Parajka, J., Perdigão, R. A. P., Merz, B., Arheimer, B., Aronica, G. T., Bilibashi, A., Bonacci, O., Borga, M., Čanjevac, I., Castellarin, A., Chirico, G. B., Claps, P., Fiala, K., Frolova, N., Gorbachova, L., Gül, A., Hannaford, J., Harrigan, S., Kireeva, M., Kiss, A., Kjeldsen, T. R., Kohnová, S., Koskela, J. J., Ledvinka, O., Macdonald, N., Mavrova-Guirguinova, M., Mediero, L., Merz, R., Molnar, P., Montanari, A., Murphy, C., Osuch, M., Ovcharuk, V., Radevski, I., Rogger, M., Salinas, J. L., Sauquet, E., Šraj, M., Szolgay, J., Viglione, A., Volpi, E., Wilson, D., Zaimi, K., and Živković, N.: Changing climate shifts timing of European floods, Science, 357, 588, 2017.
Burn, D. H., and Whitfield, P. H.: Changes in flood events inferred from centennial length streamflow data records, Advances in Water Resources, 121, 333-349, https://doi.org/10.1016/j.advwatres.2018.08.017, 2018.
Do, H. X., Westra, S., and Michael, L.: A global-scale investigation of trends in annual maximum streamflow, Journal of Hydrology, 10.1016/j.jhydrol.2017.06.015, 2017.
Do, H. X., Gudmundsson, L., Leonard, M., and Westra, S.: The Global Streamflow Indices and Metadata Archive (GSIM) – Part 1: The production of a daily streamflow archive and metadata, Earth Syst. Sci. Data, 10, 765-785, 10.5194/essd-10-765-2018, 2018a.
Do, H. X., Gudmundsson, L., Leonard, M., and Westra, S.: The Global Streamflow Indices and Metadata Archive - Part 1: Station catalog and Catchment boundary, in, PANGAEA, 2018b.
Do, H. X., Zhao, F., Westra, S., Leonard, M., Gudmundsson, L., Chang, J., Ciais, P., Gerten, D., Gosling, S. N., Schmied, H. M., Stacke, T., Stanislas, B. J. E., and Wada, Y.: Historical and future changes in global flood magnitude – evidence from a model-observation investigation, Hydrol. Earth Syst. Sci. Discuss., 2019, 1-31, 10.5194/hess-2019-388, 2019.
Gudmundsson, L., Do, H. X., Leonard, M., and Westra, S.: The Global Streamflow Indices and Metadata Archive (GSIM) – Part 2: Quality control, time-series indices and homogeneity assessment, Earth Syst. Sci. Data, 10, 787-804, 10.5194/essd-10-787-2018, 2018a.
Gudmundsson, L., Do, H. X., Leonard, M., and Westra, S.: The Global Streamflow Indices and Metadata Archive (GSIM) - Part 2: Time Series Indices and Homogeneity Assessment, in, PANGAEA, 2018b.
Hall, J., and Blöschl, G.: Spatial patterns and characteristics of flood seasonality in Europe, Hydrol. Earth Syst. Sci., 22, 3883-3901, 10.5194/hess-22-3883-2018, 2018.
Hodgkins, G. A., Whitfield, P. H., Burn, D. H., Hannaford, J., Renard, B., Stahl, K., Fleig, A. K., Madsen, H., Mediero, L., Korhonen, J., Murphy, C., and Wilson, D.: Climate-driven variability in the occurrence of major floods across North America and Europe, Journal of Hydrology, 552, 704-717, http://dx.doi.org/10.1016/j.jhydrol.2017.07.027, 2017.
Hodgkins, G. A., Dudley, R. W., Archfield, S. A., and Renard, B.: Effects of climate, regulation, and urbanization on historical flood trends in the United States, Journal of Hydrology, 573, 697-709, https://doi.org/10.1016/j.jhydrol.2019.03.102, 2019.
Mangini, W., Viglione, A., Hall, J., Hundecha, Y., Ceola, S., Montanari, A., Rogger, M., Salinas, J. L., Borzì, I., and Parajka, J.: Detection of trends in magnitude and frequency of flood peaks across Europe, Hydrological Sciences Journal, 63, 493-512, 10.1080/02626667.2018.1444766, 2018.
Slater, L. J., and Villarini, G.: Recent trends in U.S. flood risk, Geophysical Research Letters, 43, 12,428-412,436, 10.1002/2016GL071199, 2016.
Villarini, G., and Smith, J. A.: Flood peak distributions for the eastern United States, 46, 10.1029/2009wr008395, 2010.
Villarini, G.: On the seasonality of flooding across the continental United States, Advances in Water Resources, 87, 80-91, https://doi.org/10.1016/j.advwatres.2015.11.009, 2016.
Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M.,

Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., ’t Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data, 3, 160018, 10.1038/sdata.2016.18, 2016.