**Referee comment on the paper:**

**Machine learning-constrained projection of bivariate hydrological drought magnitudes and socioeconomic risks**

By Rutong Liu, Jiabo Yin, Louise Slater, Shengyu Kang, Yuanhang Yang, Pan Liu, Jiali Guo, Xihui Gu, Xiang Zhang, Aliaksandr Volchak

## General comments

This paper outlines the development of a method consisting in the execution a sequence of analytical and simulation techniques, such as deep learning – long short-term memory neural networks, hybrid modelling, bivariate Copula-based analysis, machine learning and multivariate analysis of variance. After performing multiple hydrological simulations, the most likely realization approach allows to project climate change effects on the duration and severity of hydrological droughts in 179 catchments of China, while assessing their uncertainties, and the interaction and potential impact on population and gross domestic product under three shared socioeconomic pathways.

I consider that the proposed methodology is well founded and has been applied with precision, based on very complete and good quality data and observations. In my opinion, the strategy of using the joint probability density of drought characteristics to optimize the most likely realization for the selection of joint design values was particularly interesting. Similarly, the finding that simulated streamflow performance is better in the southern region of China relative to the northern region, and that this may be attributed to a greater dependence of streamflow on rainfall in South China, which is governed by a humid climate pattern, is of utmost relevance in this study. The paper is very well written, with interesting and clear tables and figures. Notwithstanding the above, minor clarifications and additions from the authors could further facilitate the good explanation offered on the conceptualization and application of the chosen cascade modelling techniques, as well as the interpretation of the results and conclusions obtained.

## Specific comments

In **L.22** and **L.40** it is mentioned that climate change (or a warming world) accelerates the (global) water cycle, but this statement is not very precise. The hydrological cycle is a complex phenomenon, and much more elaboration would be required to generally state that, as such, it is accelerated by climate change. It seems to me that Allan et al. (2020), cited here, refer to the expectation of acceleration of global precipitation responses (as warming increases and aerosol forcing decreases), providing quotes on how the warming influence of continued rises in $CO_2$ concentration + declining aerosol cooling is expected to accelerate increases in global precipitation and its extremes as transient climate change progresses. They also point that nonlinear changes in streamflow over multidecadal timescales are expected in some regions as accelerated glacier melt is followed by declining glacier volume. However, these findings do not necessarily support the far-reaching argument for an accelerated global water cycle.

**L.51**: I would suggest reviewing the use of the adjective "uneven" to refer to the distribution of precipitation under the effects of climate change. In principle, precipitation is already an uneven phenomenon, and perhaps this term could be replaced by "rapidly changing", or another that is more precise and refers rather to some process of change than to a static characteristic.

**L.264-265**: It may be worth explaining further what is meant by selecting joint design values according to “the same frequency hypothesis” that has been applied in previous studies.

**L.480** and **L.536**: Although the word accuracy is commonly used to assess model performance in many publications, I would suggest double-checking whether it applies here. The performance of a model can be determined in terms of its efficiency (it has good predictive skills that can be tested by measures such as the Nash-Sutcliffe model efficiency coefficient), as well as by uncertainty analysis that allows the model to be characterized in terms of precision (how well each modelled value agrees with each other, i.e., width of confidence intervals constructed around modelled values) and accuracy (how well modelled values agree with “true” values, i.e., percentage of observed test values contained within certain confidence intervals built around modelled values).

**L.353-355**: Very interesting indeed is the finding that the severity of droughts measured by the TWS-DSI index is twice that of the hydrological drought, and perhaps in addition to the explanation that the TWS-DSI metric incorporates all vertical water fluxes, thus offering a comprehensive view of shifts in water scarcity, a conceptual discussion could be included around the concepts and differences between meteorological, hydrological and agricultural droughts. In this case, it should be noted that TWS-DSI does not involve aquifer recharge processes, which are fundamental to explain baseflow and, therefore, the hydrological drought in its entire extension, especially for catchments with aquifer recharge and storage capacity that exceeds several times the time step of the analysis.

**L.376-380**: How are these percentages calculated? Are the relationships of *SH*, *RH*, radiation, etc. established considering the entire tributary catchment area to the respective streamflow gauging site? Otherwise, point-to-point comparisons on the location of gauging sites would be meaningless since streamflow is a catchment-supported process. Also, what does a sensitivity rate >10% mean? Is it significant? Are the negative contributions of *RH* and shortwave radiation to streamflow significant? What do you mean by a “pronounced” negative sensitivity of shortwave radiation? If comparisons are made with catchment-support and not point-to-point, as indicated above, it does not seem to make sense that *RH* has an opposite effect on streamflow at 179 stations. Is this relationship statistically significant?

## Technical corrections

Below I recommend technical and typographical corrections to this manuscript, and some typing suggestions.

Some acronyms or abbreviations in the document are not defined or appear for the first time without having been defined, such as: GCM (**L.27**), GDP (**L.37**), HMs (**L.117**), ML (**L.128/Fig.1**), POP (not defined), $T_{or}$ (**L.272 & L.275/Fig.3**), KGE (L.391).

Consider homogenizing/equating the use of terms such as streamflow/runoff, watershed/catchment, etc.

I would strongly advise including a table of abbreviations in the paper.

**L.98-99**: …GCMs outputs under the Coupled Model Intercomparison Project phase six (CMIP6)…

**L.100**: …to quantify the sensitivity of daily streamflow to different meteorological variables.~~ to daily streamflow.~~

**L.116-117**: I would suggest including some reference for the ERA5-Land 116 dataset.

**L.128/Fig.1**: I would advise including the MANOVA analysis as a process here.

**L.131**: Perhaps you could include an introductory paragraph to section 2.1, to explain why you are calculating 2-meter relative and specific humidities.

**L.132-133**: How is Eq.(1) deriving air temperature *T?*

**L.135**: Constants $T_0$, $e_0$, $L_0$ and $R_0$ could be further explained

**L.137-138 & L.140**: Does it imply (since it is not mentioned) that $T_{2m}$, $T_2$ and *ps* are available in ERA5?

**L.144**: … The RF model is used to calculate the sensitivity of runoff to different meteorological variables~~ for runoff~~…

**L.144**: I would suggest including some reference for the RF model.

**L.153**: But is there any modeling that has been done without all the meteorological variables?

**L.171**: I do not agree with the assessment that a model containing 21 parameters is simple and efficient.

**L.177-178**: I would not consider infiltration as a type of runoff.

**L.184-185**: Could you provide more references in addition to Hu et al. (2005), or further explanation, to support the statement: “To date, it is widely reported that the XAJ model usually shows the best accuracy in simulating hydrological conditions in China”?

**L.188**: We used the SCE-UA…

**L.189-190**: “The most complete 20-year observation period is selected to calibrate five models in each watershed.” At this point it might be convenient to specify the modelling time step. Is it a daily time step?

**L.211/Eq.(9)**: Aren’t the subindexes *oh* and *ox* inverted in $W_{oh}$ and $W_{ox}$, in relation to the orders employed in Eqs.(5),(6)&(7)?

**L.213-214**: Are $W_{\bullet f}$, $W_{\bullet i}$, $W_{\bullet \hat{c}}$ and $W_{\bullet o}$ from Eqs(5), (6), (7)&(9) also weights?

**L.215**: In "…are the cell state of the LSTM and the hidden unit at ~~the~~ time $t$, respectively; $c_{t-1}$ and $hs_{t-1}$ at the former…", could you please further explain the term "hidden unit"?

**L.219-220**: If in the following statement HMs stand for Hydrological Models: "…The hydrological outputs together with other climate variables are used as inputs to feed the LSTM model (i.e., the HMs are thus constrained by the LSTM)…", then I would say that the LSTM are the ones that are being constrained by the HMs, and not the other way around.

**L.237**: What is $\overline{TWSA_y}$ and where is it used?

**L.247**: Why do you calculate two drought indexes, if Table S1 only classifies drought according to DI? If so, what is the SRI for? It is not clearly stated, but it seems like sometimes you use TWS-DSI as DI, and other times you use SRI (also not explicitly stated). Could you clarify when TWS-DSI s used and when SRI? Also, could you please explain in more detail how the maps in Figs. 4 & 5 are calculated?

**L.263**: "which contains infinite combinations of values of these two multivariate arrays of variables."

**L.275-277**: I suggest including the terms $d_T$, $s_T$, $F_S$, $F_D$ and $T_{or}$ in the legend of Fig.3 and explaining what each of them is.

**L.279**: Could you clarify what does it mean the expression "the future period" here?

**L.280**: Could you please clarify why you consider this definition of socioeconomic exposure to be "static"? (At this point you had only mentioned that it varied from 0 to 100%).

**L.286**: Please note that the conditions $T_h - T\mathrm{f} > 0$ and $T_h - T_f > 0$ are not mutually exclusive!

**L.287**: "…*POP* and ~~(~~*GDP*~~)~~ denote~~s~~ the population and the gross domestic product (in USD)~~(GDP)~~ of a given catchment in the future climate, respectively…."

**L.302**: "…is quantified by the variance of each source ~~by~~ to the total variance…."

**L.311-312**: "…with at least 20 years of data ~~are~~ were selected…"

**L.319-320**: "…As these three mason solutions are produced at different spatial resolutions, we ~~produce~~ generated blended TWS data based on the…"

**L.324**: "…precipitation, temperature~~, and~~ air pressure, etc. The spatial resolution of the dataset is 9 km…"

**L.328**: I assume that here $T_{dew}$ is the dew-point temperature, which in section 2.1 you first called $T_d$. Please check consistency.

**L.331-332**: "…The climate outputs of five GCMs ~~under~~ of the historical scenario and three SSPs (i.e., SSP1-26, SSP3-70, 331 SSP5-85) under CMIP6 are used to represent different climate scenarios…"

**L.344**: "4.1 Observed changes in SRI and TWS-DSI based drought"

**L.346**: "…employed the TWS-DSI as a supplement…."

**L.348-349**: In relation to Figs. 4 & 5, which describe drought trends based on TWS-DSI and SRI, I did not find a clear explanation in the paper on how these maps were calculated, and then several questions arise. For example, is Fig. 4 somehow calculated using Eq.(12) and how, considering that this equation is month-specific? Furthermore, it is indicated that the maps in Fig.4 correspond to the periods 2002-2022, but Fig.5 does not provide any reference time period. Could you include further description about this in the methodology section?

**L.350**: Could you explain in more detail how you concluded that drought hazards have increased in recent decades? If I understood correctly, at this point in the paper you have only provided spatial trends, not temporal trends.

**L.350-353**: How did you estimate the percentage increase in TWS-DSI droughts? Again, maybe you produced monthly maps using Eq.(12) and analyzed temporal trends for specific locations on the map? In Figs. 4 & 5 I cannot distinguish any catchment. It looks like a normal grid-based GIS analysis, rather than a catchment-based analysis. Are you using streamflow measurements to draw conclusions about hydrological droughts?

**L.364/Fig.4**: It might be convenient to explain the unit of measurement "/10 years" for drought trends, since although it is a little more intuitive in the case of frequency (count of events in a 10-year period) and duration (number of months over a 10-year period), in the case of the TWSA-DSI index this might be less obvious.

**L.381**: When you say "This" do you mean increasing or decreasing *RH*-streamflow relationships?

**L.386**: Could you refer to the different maps of Fig.6 here?

**L.389/Fig. 6**: Why do you use a thin plate smoothing spline method to interpolate your data, rather than more data-driven techniques such as directional kriging? Also, when you say point-based station data, are you referring to your hydrological stations?

**L.391/Fig. 7**: This is a very interesting figure, but it needs better explanation. For example, the legend of Fig. 7a should indicate that the regions colored according to the best performing models are the study catchments. Fig. 7b should perhaps include labels for the different categories/models, since the use of only colors is a little ambiguous.

**L.418**: Higher carbon emissions and other climate forcing factors should be listed and explained (at least briefly) in the methodology section.

**L.434-436**: A better description could be provided for terms such as IPSL_CM6A_LR. Perhaps you could also better explain in the methodological section the geo-statistics behind terms like median relative change of severity.

**L.438**: Regarding the finding of substantial spatial heterogeneity of drought across China: Are the study catchments distributed homogeneously and equally throughout the country?

Considering geospatial sampling techniques, a homogeneous density of catchments may be necessary to reach such a conclusion, in a strict sense.

**L.440**: "…intensification as a result of global warming.…"

**L.441/Fig. 10**: Color legends seem to be missing for the five GCMs.

**L.445**: "…drought severity and duration, we used a Copula…"

**L.448**: "…The medians of the projected future JRP are 38.78 years, 14.52 years and 19.24 years under…"

**L.449**: "…under SSP3-70 and SSP-5-85…"

**L.455**: The use of bivariate drought analysis can "synthesize", or rather "amplify" the individual effects of two drought characteristics?

**L.457**: Are the figures contained in Figure 12 absolute or marginal? If they are absolute, you should also present the relative change with respect to the period of reference.

**L.479**: It is very surprising to me that HTM is the main source of uncertainty, since this analysis also includes SSPs, which, being products of socioeconomic studies and models, I would think involve much higher levels of uncertainty. Does this analysis quantify uncertainties, or rather the variance explained by each of the factors?

**L.505**: Is "interference" the right term? Perhaps "intervention" would be more appropriate. In addition to that, how about other high impacting factors such as political and economic crises, changes in culture and expectations of the populations, and wars and other conflicts?

**L.545**: In line with previous comments, it really seems that your uncertainty analysis is returning rather explained variance, and not induced uncertainty.