Dear Francis Chiew,

Thank you very much for your time and for providing insightful feedback that has significantly improved our manuscript. We appreciate the opportunity to revise our work and have addressed all the points you raised. Below, we provide a point-by-point response to your comments and detail the corresponding changes made in the revised manuscript.

Comment

The paper presents a modelling analysis to quantify the uncertainty in runoff and hydrological drought projections arising from model calibration considerations (dry/wet and data length) and climate change projections (CMIP6 GCMs and different SSPs). The modelling was carried out using the SWAT hydrological model for four catchments in Korea.

This is an okay paper and is a useful addition to the literature. The paper is simplistically and nicely written, and whilst the study could have delved into nuances, the analysis here is probably sufficient for the interpretation and conclusions.

The results show that the uncertainty in the climate change (in particular rainfall, the study could specifically note this, as I am sure the range in the GCM rainfall projection is much higher than the range in the temperature or PET projection) projections is considerably higher than the differences in hydrological modelling considerations, confirming what have been reported in many studies. Nevertheless, whilst this is true when considering the sensitivity of runoff to changes in the climate inputs, the uncertainty in hydrological non-stationarity (changes in runoff-rainfall relationship, catchment response under higher temperature, PET and CO_2 not seen in the historical data, as models are extrapolated to predict the future using parameter values obtained calibration against historical data) which is not considered in these studies, could be high.

A couple of technical queries/comments below:

Response:

We sincerely thank the reviewer for their time, thoughtful evaluation, and constructive comments on our manuscript. We are grateful for the positive assessment that our paper is a "useful addition to the literature" and is "simplistically and nicely written."

The reviewer accurately summarizes the core objective of our study. Regarding the potential for delving into further nuances, our primary goal was to provide a clear and direct comparison of the major uncertainty sources (hydrological model calibration choices and climate change projections). We believe this focused approach provides a clear and valuable contribution, and we are pleased that the reviewer found the analysis sufficient for the interpretation and conclusions. We believe that by addressing these points and the specific technical queries that follow, the manuscript has been significantly improved. Our detailed point-by-point responses are provided below.

Comment 1: Some periods could be easier to model than others resulting in higher KGE values. How is this considered in the paper? through cross-sampling or cross-consideration of all possible combinations of calibration lengths in different periods?

Response:

We thank the reviewer for raising this important point, which touches upon a core strength of our experimental design. We acknowledge that model performance can indeed be sensitive to the specific characteristics of the validation period. To address this, we implemented a rigorous validation protocol that goes beyond a simple split-sample approach.

Instead of validating the model against a single, continuous block of remaining years, we performed a year-by-year validation. For example, for a model calibrated using data from years 1 to 5, we did not evaluate its performance on the entire 6-20 year period as a whole. Instead, we calculated 15 separate, single-year KGE values for year 6, year 7, and so on, up to year 20.

This meticulous approach ensures that the model's predictive skill is tested against a wide spectrum of individual annual hydrological conditions (including various dry, normal, and wet years), rather than being smoothed over a long-term average. By strictly separating each validation year from the calibration data, we obtain a more robust and unbiased assessment of how calibration period length and conditions affect the model's ability to predict outcomes in diverse, non-overlapping future scenarios. This methodology is central to our goal of quantifying the uncertainty that arises from these choices.

Changes made:

3.2 SWAT parameter calibration

The simulated runoff data were analyzed for performance using the Kling-Gupta Efficiency (KGE; Gupta et al., 2009). KGE was developed to overcome some limitations of the commonly used Nash-Sutcliffe Efficiency (NSE) in performance analysis (Gupta et al., 2009). The attributes of KGE include focusing on a few basic required properties of any model simulation: (i) bias in the mean, (ii) bias in the variability, and (iii) cross-correlation with the observational data (measuring differences in hydrograph shape and timing). The parameter optimization of SWAT was performed as shown in Fig. S. 2, considering the data length of the calibration period from 1 to 20 years. A rigorous validation scheme was adopted to prevent bias from specific period characteristics and to ensure a robust evaluation of predictive performance. For

any given calibration period, the validation was not performed on the entire remaining period as a single dataset. Instead, we conducted a year-by-year validation, calculating a separate KGE value for each individual year not included in the calibration set. For instance, if a model was calibrated on years 1-5 from a 20-year record, 15 distinct single-year KGE values were calculated for years 6 through 20. This approach strictly separates calibration and validation datasets and ensures that model performance is assessed across a diverse range of annual hydrological conditions, providing a robust foundation for the subsequent uncertainty analysis.

Following parameter optimization, KGE values as shown in Fig. 2 were found to be suitable for conducting the study, with all four dam basins achieving values above 0.60. The performance improvements are as follows: AD's KGE increased from 0.55 before calibration to 0.64 after calibration, CJ's from 0.68 to 0.75, HC's from 0.70 to 0.80, and SJ's from 0.50 to 0.73. This improvement in KGE after calibration underscores the robustness of the hydrological models used and their enhanced capability in projecting future runoff.

Comment 2: We know that models calibrated against dry period will simulate the dry period better than if calibrated against wet period and vice-versa. Could we speculate (or perhaps even extend this analysis) what parameters we should use then to mode/project the future (e.g., wetter versus drier future)? That said, the uncertainty quantification in the paper provides an indication of how much this would matter, at least for the modelling and catchments here.

Response:

The reviewer raises a fundamental and critical question in hydrological modeling for non-stationary futures. As our response to the previous comment highlights, our year-by-year validation protocol (detailed in Fig. 4) thoroughly assesses how parameters calibrated under specific conditions (e.g., Dry Flow) perform across a wide variety of individual years (dry, normal, and wet).

This detailed analysis reinforces the conclusion that no single parameter set can be deemed universally optimal for an uncertain future that may be wetter or drier. Therefore, rather than attempting to select a single "best" parameter set, the focus of our study was to embrace this very issue as a key source of uncertainty. Our primary goal was to quantify the magnitude of uncertainty stemming from hydrological modeling choices (such as calibration data length and hydrological conditions). Our findings indicate that while the choice of calibrated parameters is important, its contribution to the total uncertainty is secondary to that of the climate projections. This underscores the importance of an ensemble-based approach for future projections, which incorporates a range of plausible hydrological model parameterizations.

Changes made:

4. Discussion

This study quantified the cascade of uncertainties caused by various factors in the process of projecting future runoff and analyzing future hydrological drought. Previous studies (Chegwidden et al., 2019; Wang et al., 2020) have reported that climate data from GCMs and SSP scenarios are the primary sources of uncertainty in future hydrological analysis. The results of this study also identified GCMs as the major contributor to uncertainty in future hydrological analysis. However, recent research has begun to identify and quantify the cascade of uncertainties caused by factors beyond GCMs and SSP scenarios (Chen et al., 2022; Shi et al., 2022). This study focused on the uncertainties inherent in the calibration of hydrological models, which are essential for future water resource management. Rather than

seeking a single optimal parameter set, the central aim of this study was to quantify the uncertainty that arises from this very choice.

There have been limited studies that consider the uncertainties in runoff projection due to various calibrated parameter cases (Lee et al., 2021a). However, this study further subdivided the observation data used in the calibration period of hydrological model parameters by the amount of data and hydrological conditions to quantify the uncertainties more precisely. The results showed that hydrological conditions had a greater impact than the amount of calibration data period on the uncertainties in the calibration of hydrological model parameters.

This study went beyond merely projecting future runoff by also quantifying the cascade of uncertainties in the analysis of future hydrological drought using this runoff projection. Many studies on future drought prediction reported that hydrological drought becomes more complex and uncertain due to its association with human activities and the use of future climate data and hydrological models (Ashrafi et al., 2020; Satoh et al., 2022). Most existing studies on future hydrological drought analysis focused on the severity and frequency of droughts. However, this study quantified the cascade of uncertainties that arise in the process of future drought analysis. Although the contribution of hydrological model uncertainty to future hydrological drought may be lower compared to future runoff projections, the characteristics of uncertainty differ between drought and runoff projections, clearly indicating the necessity to separately analyze and consider these uncertainties in future hydrological analyses.

Comment 3: I suggest using blue (i.e., good) colour for Figure 4?

Response:

We thank the reviewer for the constructive suggestion. We agree that a more intuitive color scheme would improve the readability of Figure 4. Accordingly, the figure has been revised using a blue-to-red color scale to represent KGE performance more clearly, which enhances the visual interpretation of the results.

Changes made:

KGE

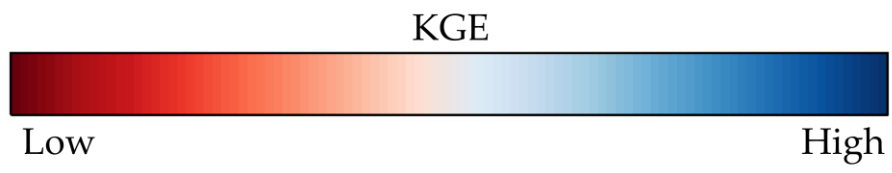Low   High
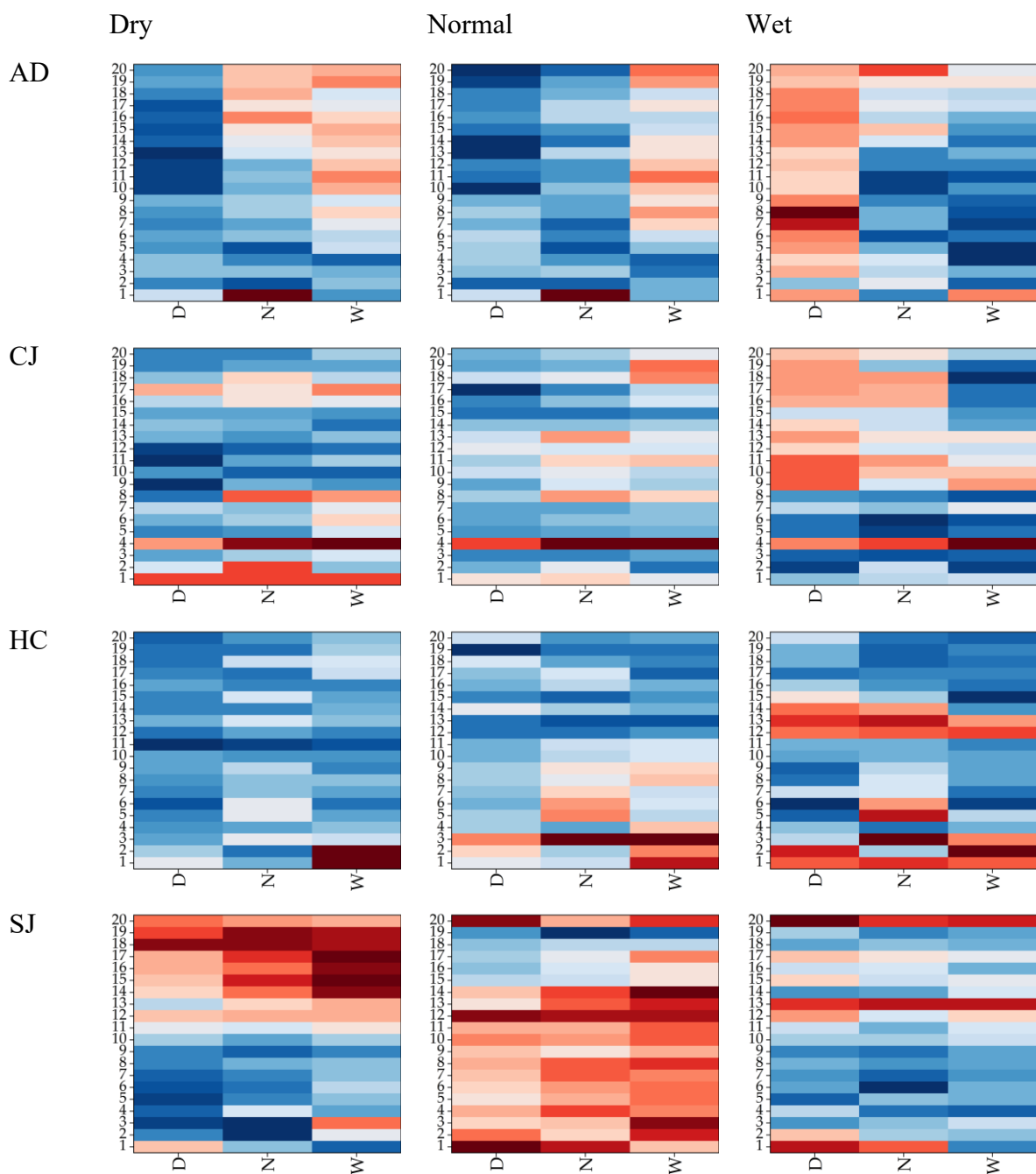
Figure. 4. KGEs classified by hydrological conditions for the calibration-validation period

Comment 4: I assume that the paper used the QQM bias corrected GCM data as input into SWOT for both the historical and future periods. It may be worth having a look at the historical modelled versus observed runoff. I suspect that the modelling with bias-corrected GCM data will underestimate the observed runoff, as the GCM is likely to underestimate the serial correlation (or multi-day wet rainfall totals) (e.g., Charles et al. and Potter et al. 2020 HESS papers). This however may not (or may) matter when considering the relative differences in the runoff projections.

Response:

We appreciate the reviewer's insightful comment on the potential limitations of GCM data. To clarify, a critical distinction in our methodology is the data used for different stages of the analysis. The SWAT model calibration and validation for the historical period were conducted exclusively using observed meteorological data and observed dam inflow records, not GCM outputs. Our model's historical performance was thus validated against actual observations.

The bias-corrected GCM data were used solely for the projection of future runoff. We acknowledge that GCMs have inherent limitations, such as underestimating serial correlations in rainfall, which is an important factor contributing to uncertainty in future projections. In our study, this inherent uncertainty stemming from the GCM data itself is precisely what is captured and quantified by the 'GCM' factor in our ANOVA. To prevent any misunderstanding, we will explicitly clarify in the methodology section (Chapter 2) that observed data were used for model calibration/validation, while bias-corrected GCM data were used for future projections.

Changes made:

2.3 Soil and water assessment tool (SWAT)

The SWAT was used to calibrate hydrological processes in our study basin. The SWAT is particularly adept at simulating runoff and other hydrological variables under a wide range of environmental conditions and is a robust, physically based, semi-distributed model. Its efficiency in modelling hydrological cycles within basins relies on simple input variables to produce detailed hydrological outputs. The capability of this model has been effectively shown in various studies, including those in South Korea (Kim et al., 2022; Song et al., 2022).

The core of the SWAT model is the water balance equation, which integrates daily weather data with land surface parameters to calculate water storage changes over time:

$$SW_t = SW_0 + \sum_{i=0}^{t}\left(R_{day} - Q_{surf} - E_a - w_{seep} - Q_{gw}\right) \qquad (1)$$

where SW_0 is the initial soil moisture content (mm), SW_t is the total soil moisture per day (mm), R_{day} is precipitation (mm), Q_{surf} is surface runoff (mm), E_a is evapotranspiration (mm), W_{seed} is penetration, Q_{gw} is groundwater runoff (mm), and t is time (day).

For rainfall-runoff analysis, the SWAT model is structured into several sub-basins, each of which is further subdivided into Hydrologic Response Units (HRUs) based on different soil types, land use and topography. Each HRU independently simulates parts of the hydrological cycle, allowing a granular analysis of basin hydrology. This setup reflects the spatial heterogeneity within the basin and allows continuous simulation of hydrological processes over long time periods, enhancing the utility of the model for climate change studies. The model was calibrated and validated using R-SWAT for parameter optimization. R-SWAT incorporates the SUFI-2 algorithm, which is known for its rapid execution and precision in parameter optimization, ensuring accurate and reliable simulation results (Nguyen et al., 2022). In this study, the setup and evaluation of the SWAT model for the historical period were performed using observed data. The model was forced with observed meteorological data, and the parameters were calibrated and validated against historical daily dam inflow records for the period 1980-2023.

2.5 General Circulation Models (GCMs)

In this study, M1 to M20 GCMs from the CMIP6 suite that have been consistently used in studies for East Asia and Korea were selected for future runoff projection and hydrological drought analysis. The details of the development institutions, model names and resolutions of these 20 GCMs were presented in Table S2.

The climate data from the GCMs were evaluated using daily observed climate data provided by the Korea Meteorological Administration (KMA). The evaluation used observed data from the past period (1985-2014) to evaluate the future climate data from the GCMs, which were analyzed for two future periods: the near future (NF) and the distance future (DF). The future climate change scenarios used were SSP2-4.5, SSP3-7.0 and SSP5-8.5. The SSP scenarios are

divided into five pathways based on radiative forcing, reflecting different levels of future mitigation and adaptation efforts (O'Neill et al., 2016). The SSPs are numbered from SSP1 to SSP5, with SSP1 representing a sustainable green pathway and SSP5 representing fossil fuel driven development. The numbers 4.5 to 8.5 indicate the level of radiative forcing (4.5: 4.5 W m-2, 7.0: 7.0 W m-2 and 8.5: 8.5 W m-2). For the analysis of future changes, the calibrated SWAT model was then driven by bias-corrected future climate projection data from the 20 GCMs under the three SSP scenarios. This approach ensures that the model's baseline performance is grounded in observational data, while the future analysis specifically assesses the uncertainties propagated from the climate projections and hydrological modeling choices.

Comment 5: It is interesting that the uncertainty in the hydrological drought projection is lower than the runoff projection. Can the modelling (or a bit more analysis) shed some light? because of the lag/storage effect in runoff? because there is less uncertainty in the multi-year characteristics in the GCM simulation compared to the average rainfall?

Response:

This is a very interesting and accurate observation. The primary reason for the lower quantified uncertainty in hydrological drought projections lies in the fundamental difference between raw runoff and the Streamflow Drought Index (SDI).

Monthly runoff is a direct physical quantity (m³/s) with high variability. In contrast, the SDI is a standardized statistical index derived from accumulating runoff over several months. This calculation process inherently smooths out the high-frequency fluctuations present in the monthly runoff data. As a result, the numerical range and variance of the SDI values are naturally smaller than those of the raw runoff. In the ANOVA, this lower total variance in the drought index directly leads to smaller calculated uncertainty contributions. This explains not only the difference in the percentage contributions but also why the overall pattern of uncertainty differs from that of the direct runoff analysis.

Changes made:

3.9.3 Uncertainty contribution of future hydrological drought

The quantification of uncertainty in future hydrological drought was conducted using ANOVA. The uncertainty in future hydrological drought projections caused by SSP, GCM, and hydrological modelling parameters was clearly quantified by ANOVA. Fig S.10 shows the contribution of each factor to the total uncertainty. Among single-factor uncertainties, GCM contributed the most, averaging over 30%. The largest contributor to the total uncertainty, however, was the interaction between SSP and GCM, averaging over 50%.

Fig. 7 and Table 8 present the contribution of hydrological modelling parameters to the uncertainty in future drought projections. The uncertainty contribution from hydrological model parameter estimation in future hydrological drought analysis averaged 2.7%, which is lower than that observed for future runoff projections. The uncertainty contribution from hydrological model calibration for future drought conditions was highest in HC, followed by CJ, AD, and SJ, respectively. These results differ from those obtained in the runoff projections. The contribution of uncertainty in hydrological drought analysis decreased for

AD and SJ, where uncertainty in future runoff projection due to hydrological model calibration was relatively high. In contrast, HC showed high uncertainty contributions from hydrological model calibration in both runoff and drought analyses. Monthly runoff is a direct physical variable with high temporal volatility. In contrast, the SDI, used here to quantify hydrological drought, is a processed statistical indicator. It is calculated by accumulating and standardizing runoff over multi-month timescales. This integration process acts as a filter, effectively smoothing the high-frequency variability of the raw runoff series. Consequently, the absolute numerical fluctuation of the SDI is significantly smaller than that of the runoff itself. This reduced total variance in the drought index is the primary reason why the quantified uncertainty contributions appear lower and exhibit a different pattern compared to the runoff analysis. This highlights that while the underlying drivers of uncertainty are the same, their manifestation can differ depending on the temporal scale and the nature of the hydrological variable being analyzed. These findings confirm the necessity to separately analyze and consider uncertainties in future runoff projection and hydrological drought analysis.

We believe that these revisions have thoroughly addressed the reviewer's concerns and have substantially strengthened the manuscript. We look forward to your positive consideration of our revised work.

Sincerely,

Kim Jin Hyuck
on behalf of all authors

General Comments

The manuscript addresses relevant scientific questions within the scope of the journal. It presents novel ideas, as this exact combination of uncertainty drivers— GCM, SSP, calibration period length, hydrological conditions during calibration, and model parameter uncertainty—has not been investigated previously (to my knowledge). Interesting conclusions are reached: GCMs contribute most in general, while model uncertainty contributions differ for general future runoff and drought prediction (lower for drought prediction).

Answer)

We sincerely thank you for your thorough and constructive review. We are very grateful for your positive assessment, particularly your recognition that our work addresses relevant scientific questions, presents a novel combination of uncertainty drivers, and reaches interesting conclusions. Your insightful feedback has been invaluable in helping us to further strengthen the manuscript.

The methodology is valid but quite complex and not always straightforward. An overview plot or flowchart would help clarify how the different simulations are organized and how the combination of GCMs, SSPs, calibration periods, hydrological conditions, and parameter sets is applied. Figures and tables are informative but must be clarified (more extensive captions and axis labels). Overall, this is a strong and complex manuscript with interesting results. Minor clarifications regarding methodology, figure/table captions, and interpretation of results would further strengthen the clarity.

Answer)

Thank you for these excellent suggestions. We agree that the multi-step methodology, involving several interacting uncertainty drivers (GCM, SSP, HC, PL), can be complex to follow. We also agree that several captions and labels needed more detail to improve clarity.

To address the methodological complexity, we have added a new comprehensive concept as Figure 1 in the manuscript. We have also revised Section 2.1 (Procedure) to explicitly refer to this new figure, which now serves as a visual guide to the steps described.

Furthermore, following your general advice, we have reviewed and revised the captions for figures and tables throughout the manuscript to be more descriptive, self-explanatory, and include clear axis definitions where needed.

2.1 Procedure

The procedure of the study is as follows. The overall workflow, illustrating the main phases of data processing, model setup, and analysis, is visualized in Fig. 1. First, topographic data for four dam basins in South Korea were established, taking into account the overall hydrological characteristics of the region, and observed dam inflow data were utilized to consider the length and hydrological conditions of the hydrological model calibration data. The length of the calibration data considered ranged from 1 to 20 years, and hydrological conditions were categorized using the Streamflow Drought Index (SDI). Subsequently, validation performance analysis was conducted, with calculations varying according to the length of calibration data and hydrological conditions (Dry, Normal, and Wet). For the study, future climate data from 20

Coupled Model Intercomparison Project Phase 6 (CMIP6) GCMs and three SSP scenarios (SSP2-4.5, SSP3-7.0, and SSP5-8.5) were bias-corrected. Future runoff projection and hydrological drought were then analyzed using calibrated hydrological model parameters under different conditions along with the future climate data. Finally, the uncertainties in the future hydrological analysis were quantified using the Analysis of Variance (ANOVA).
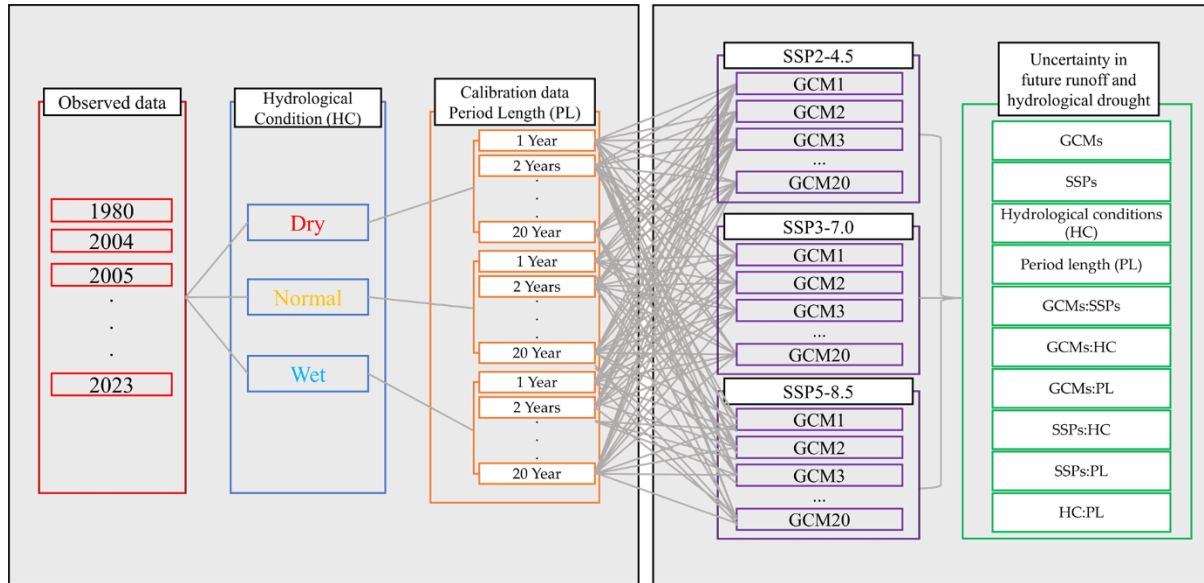

Figure. 1. Uncertainty concept in this study

Specific Comments

Methodology and Simulation Setup

Lines 284–286 mention 120 simulations, but the description of 20 GCMs, 3 SSPs, 3 hydrological conditions, and 20 calibration period lengths (not multiplying to 120, but 3600, though probably I misunderstood something) is confusing, especially when also using three different durations (3, 6, and 12 months) to determine the hydrologic conditions (HC).

Answer)

Thank you for highlighting this major point of confusion and our sincere apologies for this misleading error. You are absolutely correct, and we are grateful for your meticulous review.

The 120 unique combinations mentioned in our manuscript was a significant error in the text. Your calculation is correct.

The actual analysis setup, as correctly inferred by you and detailed in our new flowchart (Figure 2), consists of 60 climate scenarios (20 GCMs × 3 SSPs) combined with 60 distinct hydrological model parameterization types (3 Hydrological Conditions × 20 Period Lengths). This results in a full set of 3,600 combinations (60 × 60) for each basin, and the ANOVA was applied to this complete dataset.

We have completely rewritten this paragraph in Section 2.7 to remove the incorrect "120 combinations" reference and to clearly describe the full 3,600-combination set that forms the basis of our ANOVA.

2.7 Quantifying uncertainty

The ANOVA used in this study is an effective statistical method that decomposes the total sum of squares (SST) into contributions from different sources and their interactions. This method would be particularly useful in the study framework, as it allows us to assess not only the individual effects of each source of uncertainty but also the combined effects of these sources interacting with each other (Bosshard et al., 2013; Lee et al., 2021a).

For this analysis, the primary sources of uncertainty considered are General Circulation Models (GCMs), Shared Socioeconomic Pathway (SSP) scenarios, hydrological conditions (HC)

during the calibration period, and period length (PL). Each of these sources could have a significant impact on the projections of hydrological models; therefore, their comprehensive evaluation is crucial (Morim et al., 2019; Yip et al., 2011). Higher-order interactions (e.g., three-way) were excluded as they are often difficult to interpret physically and can introduce noise into the model.

$$SST = SS_{GCMs} + SS_{SSPs} + SS_{HC} + SS_{PL} + SS_{Interactions(2-way)} + SS_{Residuals} \qquad (7)$$

where each term (SS) indicates the sum of squares attributed to each factor or interaction. Here, SS_{GCMs}, SS_{SSPs}, SS_{HC}, and SS_{PL} represent the sum of squares due to GCMs, SSPs, HC, and PL, respectively, known as the main effects. The remaining terms represent the sum of squares due to the interactions among GCMs, SSPs, hydrological conditions, period length, their two-way interactions, and the residual error.

The model setup for ANOVA was designed to analyze the set of projections. As detailed in the flowchart (Fig. 1), this set was generated by combining 60 climate data (20 GCMs × 3 SSPs) with 60 distinct hydrological model parameterization (3 HC × 20 PL). This resulted in a total of 3,600 combinations for each basin analyzed. Initially, the SST, representing the total variation within the data, was calculated. Subsequently, the sum of squares attributable to each source of uncertainty was computed. To quantify the relative impact of each source, its contribution was calculated as the proportion of its Sum of Squares relative to the Total Sum of Squares. This provides a clear measure of the percentage of total uncertainty explained by each factor and interaction.

The statistical robustness and validity of the ANOVA models were rigorously evaluated. First, the overall goodness-of-fit for each model was assessed using the Adjusted R-squared (R^2_{adj}), defined as Eq. (8).

$$R^2_{adj} = 1 - \frac{(1-R^2)(n-1)}{n-k-1} \qquad (8)$$

Where, R^2 is the coefficient of determination, n is the number of observations, and k is the number of predictions. This metric is preferred over the Standard R-squared as it adjusts for the number of predictors in the model, providing a more accurate measure of model fit.

Second, a residual analysis was conducted to verify that the core assumptions of ANOVA were met. The normality of residuals was a primary focus of this validation, examined both statistically with the Shapiro-Wilk test and visually using Quantile-Quantile (Q-Q) plots. The Shapiro-Wilk test evaluates the null hypothesis that the residuals are normally distributed. However, given the large sample size in this study, which can lead to statistically significant results even for minor deviations from normality, greater emphasis was placed on the visual inspection of Q-Q plots to assess practical adherence to the normality assumption. The assumption of homoscedasticity (constant variance of residuals) was also inspected using Residuals vs. Fitted values plots. These validation steps ensure that the results of the uncertainty partitioning are statistically sound and reliable. All statistical analyses were performed using the R software environment.

Figure S.2 helps, but the explanation remains unclear

Answer)

Thank you for this specific feedback. You are correct that the original caption for Figure S.2. Description of calibration period data lengths in this study was too brief and uninformative. We agree that this figure is important for understanding our experimental design for defining the Period Length (PL) uncertainty factor.
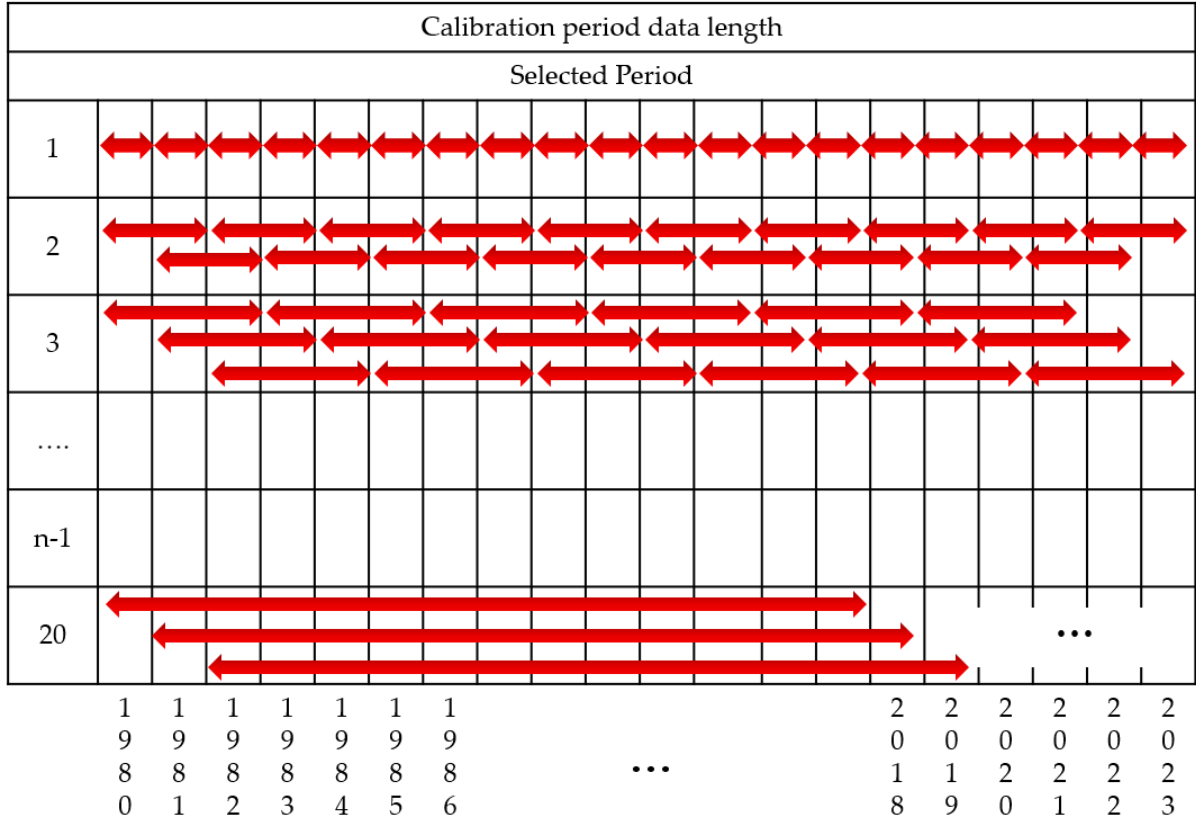

Figure. S. 2. Schematic illustration of the calibration period selection strategy. The horizontal axis represents the full historical data period (1980-2023). The vertical axis represents the calibration period length (PL) from 1 to 20 years. The red arrows illustrate the specific years of data used for calibration corresponding to each data length.

Results and Interpretation

An overview of the four basins' properties (land use, precipitation, slope, etc.) and how they differ would help interpret differences in uncertainty contributions, as far as I could see this was not done apart from differentiating between the catchment size. But since there are quite large differences between the catchments, this should be discussed more in depth.

Answer)

This is an excellent point. The reviewer is correct that we selected the basins based on their natural state, but we failed to use their physical and climatic characteristics to interpret the differences in our results.

We apologize if this was not clear, but the detailed characteristics (Area, Mean Temp, Mean Precip, Land Use Ratios) were already provided in Table S1 in the Supplementary Information.

However, we completely agree with the core of your comment: we did not discuss these differences in depth in the main text. To address this significant omission:

We have revised Section 2.2 (Study area and datasets) to more clearly summarize the diversity of the basins (e.g., precipitation ranges from 1,045 mm to 1,330 mm) and to more strongly signpost the reader to Table S1 for details.

More importantly, we have added a new discussion point to Section 4 (Discussion) that explicitly links these basin characteristics (e.g., differences in precipitation and area) to the observed differences in uncertainty contributions, addressing why basins like HCH and SJ show such different sensitivities.

2.2 Study area and datasets

The study areas selected in this study are the Andong (AD), Chungju (CJ), Habcheon (HCH), and Seomjingang (SJ) dam basins located in Korea as shown in Fig. 2. To achieve stable calibration and validation results for a hydrological model, it is imperative to choose catchments with extensive hydrological data records. This enables the accurate estimation of appropriate calibration data lengths through various testing periods of the hydrological model. Furthermore, incorporating a variety of basins is crucial to ensure that the findings of this study

are not biased by specific hydrological conditions. These four basins, which have the longest hydrological records in Korea, are situated in major river basins. Detailed basin characteristics are provided in Table S1. While all four basins are located in temperate climate zones and are predominantly forested (Forest ratio > 75%, except for CJ at 61.7%), they represent a diverse range of hydrological and climatic conditions. While all four basins are located in temperate climate zones and are predominantly forested (Forest ratio > 75%, except for CJ at 61.7%), they represent a diverse range of hydrological and climatic conditions. Area varies significantly from 763 km² (SJ) to 6,648 km² (CJ). Mean annual precipitation also ranges from 1,045.7 mm (AD) to 1,329.8 mm (SJ). These regions are devoid of artificial structures (Urban ratio < 5.3% for all basins), ensuring that runoff remains natural and unaltered. Located in different regions of Korea, these basins have a range of hydrological conditions and runoff characteristics, providing a representative cross-section of the country's hydrological characteristics.

Table S1. Information of study area

Basins	Area (km²)	Aver. Temp (°C)	Aver. Prec. (mm)	Urban ratio (%)	Forest ratio (%)
AD	1,584	12.2	1,045.7	0.7	78.8
CJ	6,648	11.7	1,214.3	5.3	61.7
HCH	925	13.2	1,289.9	2.0	78.3
SJ	763	13.3	1,329.8	1.5	75.4

4. Discussion

This study quantified the cascade of uncertainties caused by various factors in the process of projecting future runoff and analyzing future hydrological drought. Previous studies (Chegwidden et al., 2019; Wang et al., 2020) have reported that climate data from GCMs and SSP scenarios are the primary sources of uncertainty in future hydrological analysis. The results of this study also identified GCMs as the major contributor to uncertainty in future runoff analysis. This aligns with findings such as Her et al. (2019), who demonstrated that GCM uncertainty is dominant for rapid hydrological components, whereas parameter uncertainty becomes more significant for slower. However, recent research has begun to identify and quantify the cascade of uncertainties caused by factors beyond GCMs and SSP scenarios (Chen et al., 2022; Shi et al., 2022). This study focused on the uncertainties inherent in the calibration of hydrological models, which are essential for future water resource management. Rather than

seeking a single optimal parameter set, the central aim of this study was to quantify the uncertainty that arises from this very choice.

There have been limited studies that consider the uncertainties in runoff projection due to various calibrated parameter cases (Lee et al., 2021a). However, this study further subdivided the observation data used in the calibration period of hydrological model parameters by the amount of data and hydrological conditions to quantify the uncertainties more precisely. The results showed that hydrological conditions had a greater impact than the amount of calibration data period on the uncertainties in the calibration of hydrological model parameters.

This study went beyond merely projecting future runoff by also quantifying the cascade of uncertainties in the analysis of future hydrological drought using this runoff projection. Many studies on future drought prediction reported that hydrological drought becomes more complex and uncertain due to its association with human activities and the use of future climate data and hydrological models (Ashrafi et al., 2020; Satoh et al., 2022). For example, Gao et al. (2020), also using an ANOVA approach, found that for low flows, GCM and RCP uncertainty became increasingly pronounced. Most existing studies on future hydrological drought analysis focused on the severity and frequency of droughts. However, this study quantified the cascade of uncertainties that arise in the process of future drought analysis. Although the contribution of hydrological model uncertainty to future hydrological drought may be lower compared to future runoff projections, the characteristics of uncertainty differ between drought and runoff projections, clearly indicating the necessity to separately analyze and consider these uncertainties in future hydrological analyses.

Furthermore, the basin-specific characteristics presented in Table S1 help interpret the differing uncertainty contributions seen in the results. For example, in the hydrological drought analysis (Fig. 8), the uncertainty from model calibration was highest in HCH (5.56%) but lowest in SJ (0.26%), despite their similar areas (925 km² vs 763 km²). A key difference is that the SJ basin receives significantly higher mean annual precipitation (1,329.8 mm) compared to HCH (1,289.9 mm) and especially AD (1,045.7 mm). This suggests that basins with lower precipitation (like HCH and AD) may be more hydrologically sensitive to calibration data selection, leading to higher model-driven uncertainty, whereas the wetter conditions in SJ create a more robust (less sensitive) hydrological response regardless of calibration choice.

It would also be interesting to have an explanation of why some catchments yield a rather moderate KGE of 0.64.

Answer)

This is a fair question. The reviewer is correct to note that the KGE for the Andong (AD) basin (0.64) is 'moderate' relative to the high performance achieved in the HCH basin (0.80).

While a KGE of 0.64 is still considered 'Good' performance according to standard hydrological literature (Zhang et al., 2025), we agree this difference warrants explanation. This lower performance is not a flaw in the calibration methodology, but rather reflects the inherent, well-known hydrological complexities of the AD basin itself.

The primary reason is that the AD basin's historical period includes severe, record-breaking drought events (e.g., 2014-2015). As noted in our own manuscript (Section 3.1, citing Karunakalage et al., 2024), these extreme, non-linear outlier events are inherently difficult for hydrological models to capture perfectly, which lowers the overall KGE score.

Furthermore, other recent studies modeling the AD basin have also reported validation statistics in the 'Good' but not 'Very Good' range. For example, Han et al. (2019) reported validation NSE values of 0.52–0.69 for the Andong Dam basin. Similarly, Lee et al. (2020), in a study of the Nakdong River basin, reported NSE values as low as 0.59 for dam inflows including Andong Dam.

Therefore, we interpret the 0.64 KGE as a realistic and acceptable performance for this specific and challenging-to-model catchment.

Zhang, J., Kong, D., Li, J., Qiu, J., Zhang, Y., Gu, X., & Guo, M. (2025). Comparison and integration of hydrological models and machine learning models in global monthly streamflow simulation. Journal of Hydrology, 650, 132549.

Lee, J., Lee, Y., Woo, S., Kim, W., & Kim, S. (2020). Evaluation of water quality interaction by dam and weir operation using SWAT in the Nakdong River Basin of South Korea. Sustainability, 12(17), 6845.

Han, J., Lee, D., Lee, S., Chung, S. W., Kim, S. J., Park, M., Lim, K. J & Kim, J. (2019). Evaluation of the effect of channel geometry on streamflow and water quality modeling and modification of channel geometry module in SWAT: A case study of the Andong Dam Watershed. Water, 11(4), 718.

Line 607–609: clarify what is meant by the statement that the parameter set calibrated with dry periods shows higher performance—higher than the set calibrated for normal or wet conditions?

Answer)

Thank you for pointing out this ambiguity. You are correct that the original sentence was unclear by not specifying the basis for the comparison. The 11.4% and 6.1% figures represent the average improvement when comparing parameters calibrated in a dry period against parameters calibrated in a wet period. We have revised this key finding in the Conclusion to make this comparison explicit.

5. Conclusion

This study aimed to quantify the uncertainties in future runoff projections and hydrological drought analysis, considering various climate change scenarios and hydrological model calibrations. SWAT was used, and hydrological conditions were classified using the SDI. Additionally, 20 GCMs and three SSP scenarios were applied. The calibration data length ranged from 1 to 20 years, considering different hydrological conditions (Dry, Normal, Wet).

The main findings are as follows:

First, the validation performance of the calibrated hydrological model parameters depended significantly on the hydrological conditions of the calibration data. For instance, when compared against parameters calibrated using wet period data, hydrological model parameters calibrated with dry period data showed an average of 11.4% higher performance when validated under dry conditions and 6.1% higher performance when validated under normal conditions.

Second, the contribution of hydrological model uncertainty to future runoff projections ranged from 3.9% to 9.8%, with this uncertainty being more pronounced during low runoff periods. ANOVA results clearly indicated that GCMs contributed the most uncertainty, consistently accounting for over 60% on average, highlighting GCMs as the dominant source. In contrast, the contributions of SSP scenarios and hydrological model parameters were relatively smaller.

Third, the contribution of hydrological model uncertainty in future hydrological drought

analysis was on average 2.7%, lower than that observed for future runoff projections. The uncertainty contributions varied by basin, showing different patterns from runoff projections, thus confirming the necessity for separate analyses of future runoff and hydrological drought uncertainties.

The significance of this study lies in emphasizing the quantification of uncertainty from various sources, including hydrological conditions and calibration data length, in addition to climate model scenarios. The systematic approach using ANOVA provided insights into the dominant sources and interactions of uncertainties, offering important guidance for improving hydrological modeling practices and water resources planning under future climate scenarios. However, there remains a need to apply this methodology to other regions to generalize these findings further.

Conclusion: uncertainty in future Streamflow Drought Index (SDI) due to model parameter uncertainty (HC and PL) was on average 2.7%, whereas uncertainty for general runoff prediction was (more?) seasonally and catchment dependent, and generally higher. The implications of these findings could be made clearer, as I assume it means that predictions of low-flow periods are less sensitive to hydrological conditions in the calibration period than overall runoff predictions.

Answer)

This is a very interesting and accurate observation. We agree that the implications of this finding are important and require a clear explanation.

The primary reason for the lower quantified uncertainty in hydrological drought (SDI) lies in the fundamental difference between raw monthly runoff and the Streamflow Drought Index (SDI).

Monthly runoff is a direct physical variable with high temporal volatility. In contrast, the SDI, used here to quantify hydrological drought, is a processed statistical indicator. It is calculated by accumulating and standardizing runoff over multi-month timescales (e.g., 3-month SDI). This integration process acts as a filter, effectively smoothing the high-frequency variability of the raw runoff series.

Consequently, the absolute numerical fluctuation (and total variance) of the SDI is significantly smaller than that of the runoff itself. In our ANOVA, this reduced total variance in the drought index is the primary reason why the quantified uncertainty contributions from the model parameters (HC and PL) appear lower and exhibit a different pattern compared to the runoff analysis.

Realizing this was a key point needing clarification, we have added a detailed explanation to the revised manuscript in Section 3.9.3 to clarify this exact point.

3.9.3 Uncertainty contribution of future hydrological drought

The quantification of uncertainty in future hydrological drought was conducted using ANOVA. The uncertainty in future hydrological drought projections caused by SSP, GCM, and hydrological modelling parameters was clearly quantified by ANOVA. Fig S.10 shows the contribution of each factor to the total uncertainty. Among single-factor uncertainties, GCM contributed the most, averaging over 30%. The largest contributor to the total uncertainty,

however, was the interaction between SSP and GCM, averaging over 50%.

Fig. 8 and Table 8 present the percentage contribution of hydrological modelling parameters to the total uncertainty of the future 3-month SDI value. The uncertainty contribution from hydrological model parameter estimation in future hydrological drought analysis averaged 2.7%, which is lower than that observed for future runoff projections. The uncertainty contribution from hydrological model calibration for future drought conditions was highest in HCH, followed by CJ, AD, and SJ, respectively. These results differ from those obtained in the runoff projections. The contribution of uncertainty in hydrological drought analysis decreased for AD and SJ, where uncertainty in future runoff projection due to hydrological model calibration was relatively high. In contrast, HCH showed high uncertainty contributions from hydrological model calibration in both runoff and drought analyses. Monthly runoff is a direct physical variable with high temporal volatility. In contrast, the SDI, used here to quantify hydrological drought, is a processed statistical indicator. It is calculated by accumulating and standardizing runoff over multi-month timescales. This integration process acts as a filter, effectively smoothing the high-frequency variability of the raw runoff series. Consequently, the absolute numerical fluctuation of the SDI is significantly smaller than that of the runoff itself. This reduced total variance in the drought index is the primary reason why the quantified uncertainty contributions appear lower and exhibit a different pattern compared to the runoff analysis. This highlights that while the underlying drivers of uncertainty are the same, their manifestation can differ depending on the temporal scale and the nature of the hydrological variable being analyzed. These findings confirm the necessity to separately analyze and consider uncertainties in future runoff projection and hydrological drought analysis.

Figures and Tables

Figure 2: Boxplots over all model chains before and after calibration—clarify whether the x-axis representing "number of years for the calibration period" refers simply to the length of the modeling period for pre-calibration conditions. The meaning of the different structures in the boxplots could perhaps be simplified or clarified.

Answer)

Thank you for this very specific and important question. You have identified a key point of confusion in this figure (which is now renumbered to Figure 3 in our revised manuscript). You are absolutely correct that the x-axis (1-20) in the 'Before' (pre-calibration) panel is confusing, and the original caption was too brief.

To clarify the figure's structure: The x-axis (1-20) defines the specific before calibration/after calibration data split. For any given x-axis value (e.g., '5'):

The 'After-5' boxplot shows the distribution of KGE values for the model calibrated on 5 years, evaluated on its corresponding calibration years.

The 'Before-5' boxplot shows the distribution of KGE values for the default, uncalibrated model, evaluated on the exact same validation years as the 'After-5' model.

This structure allows for a direct, fair comparison, demonstrating that the calibration process ('After') consistently outperforms the default model ('Before') when tested on the same data.

(a) AD

(b) CJ

(c) HCH

(d) SJ

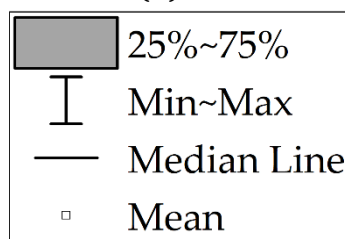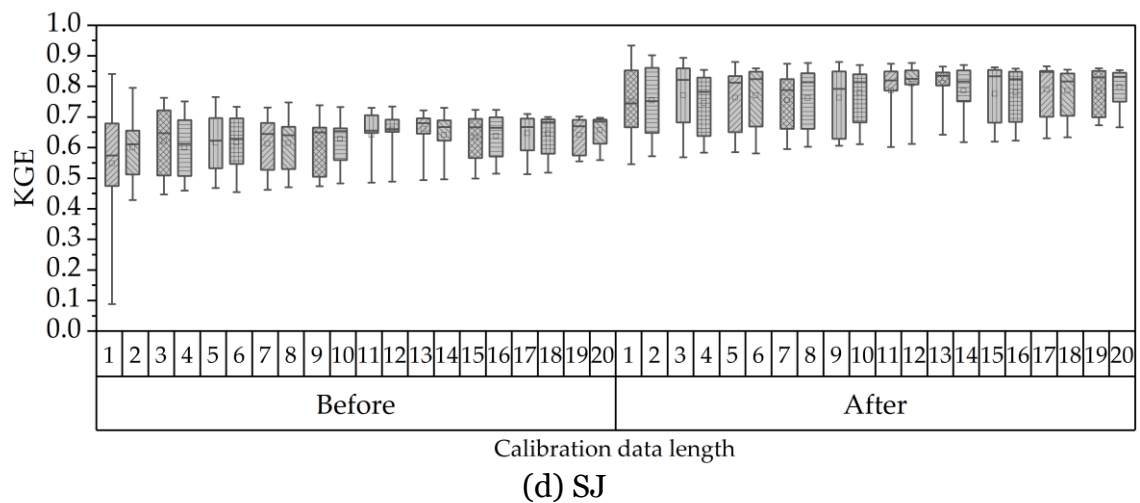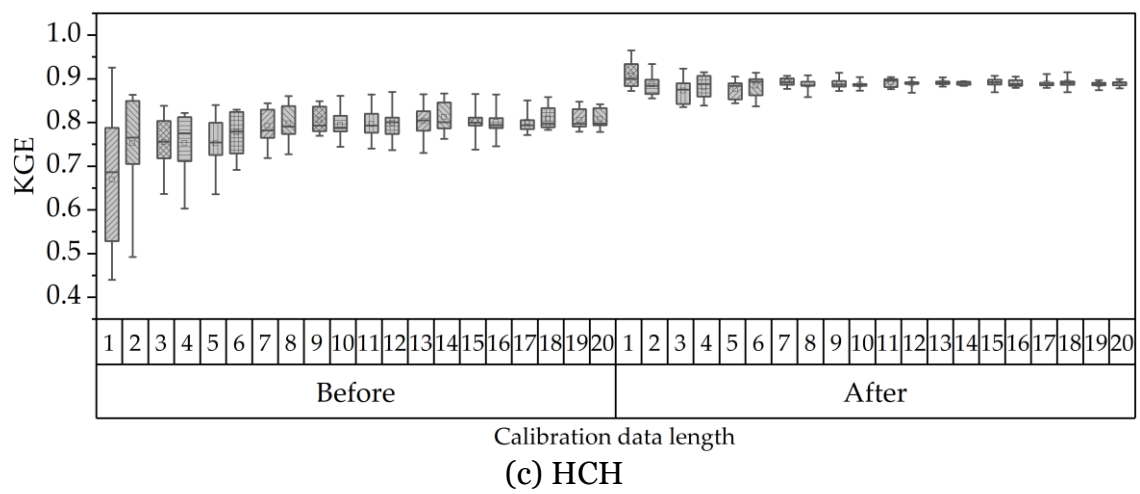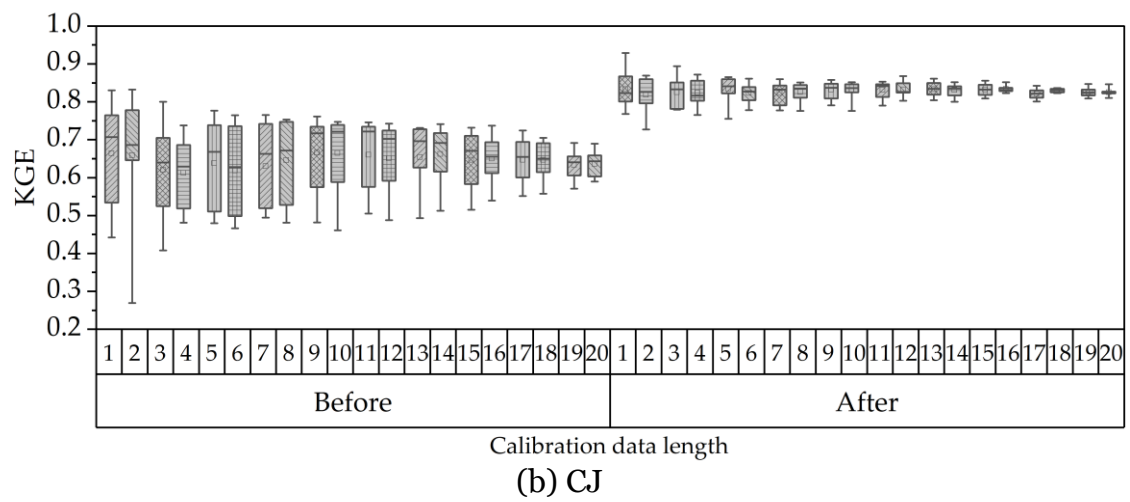| | 25%~75% |
	Min~Max
	Median Line
	Mean
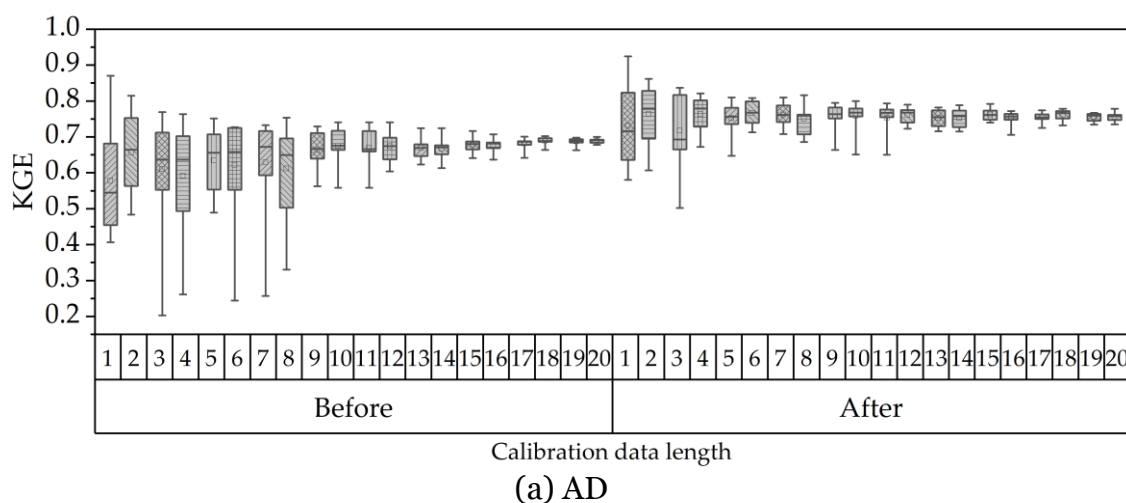
Figure. 3. Comparison of KGE values for the four basins using default parameters (Before)

and calibrated parameters (After). The x-axis (1-20) represents the calibration data length, which defines the before calibration/after calibration data split.

Figure 4: clearly indicate which wet (w), dry (d), and normal (n) conditions correspond to calibration and validation periods.

Answer)

Thank you for this critical feedback. You are correct that the original caption for this complex figure (now numbered Figure 4) was insufficient and did not explain the figure's hierarchical structure, making it difficult to interpret.

To improve clarity, we have made two key changes in the revised manuscript:

We had already updated the color scheme (based on previous feedback) from the original to a more intuitive scale where Blue indicates high KGE (good performance) and Red indicates low KGE (poor performance).

Based on your specific suggestion, we have completely rewritten the caption for Figure 4. The new caption now explicitly details the structure: the main rows (Basins), the main columns (Validation Conditions), and, within each heatmap, the y-axis (Calibration Length) and x-axis (Calibration Conditions).
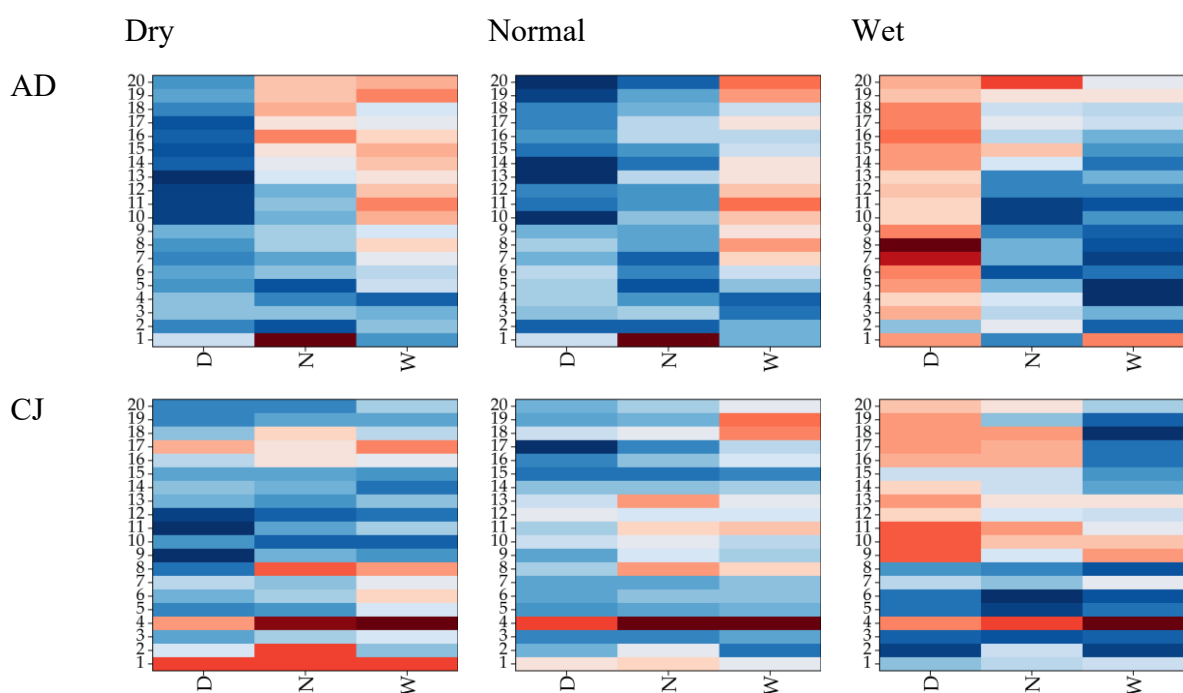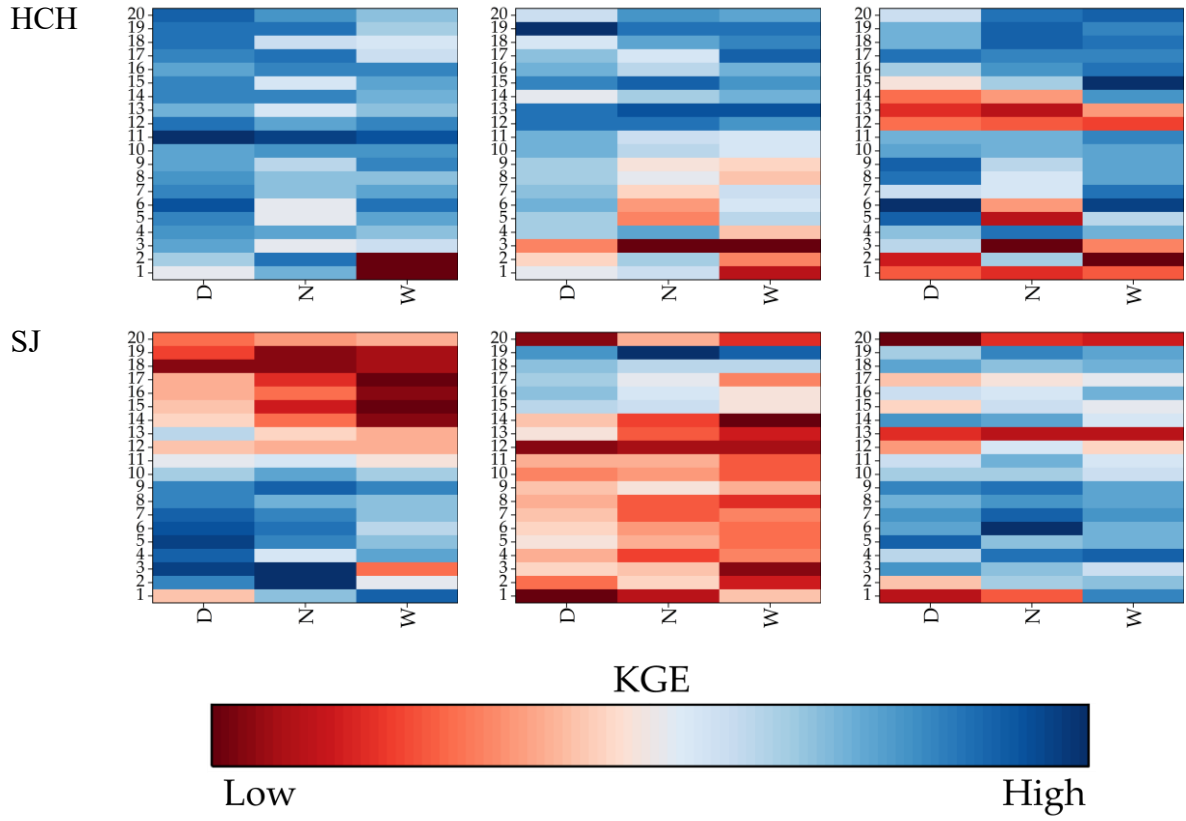
Figure. 5. Heatmap matrix of KGE performance by calibration and validation conditions. The four main rows represent the basins (AD, CJ, HC, SJ). The three main columns (labeled 'Dry', 'Normal', 'Wet') represent the hydrological conditions of the validation period. Within each individual heatmap, the y-axis represents the calibration data length (1-20 years), and the x-axis (labeled D, N, W) represents the hydrological conditions of the calibration period. Blue indicates high KGE (good performance) and red indicates low KGE (poor performance).

Figure 6: discussion on what causes the differences in contribution of different drivers would be useful. While the two drivers selected in the figure highlight differences, Figure S.9 seems more comprehensive; it may belong in the main text instead of the appendix.

Answer)

Thank you for this valuable suggestion regarding the presentation of the uncertainty results. We completely agree with you that Figure S.9 is a crucial figure that provides a comprehensive overview of all uncertainty drivers (GCM, SSP, HC, PL, and interactions).

Our rationale for placing Figure 7 (the new number) in the main text was to specifically highlight the novel findings of this study. While many studies have already confirmed that GCMs and SSPs are the dominant sources (which Fig. S.9 also shows), the core focus of our paper is to quantify the specific, often overlooked, uncertainty contribution stemming from the hydrological model calibration (HC and PL). To better integrate your excellent point, we have revised the text in Section 3.8.2. The revised text now explicitly introduces Figure S.9 first as the comprehensive overview (confirming the dominance of GCMs), and then introduces Figure 7 as the figure that specifically isolates and details the hydrological model uncertainty, which is the central theme of our paper.

3.8.2 Contribution of uncertainty using the ANOVA

A comprehensive overview of the relative contributions from all factors to the uncertainties in future runoff projections for each basin is provided in Fig. S. 9. As confirmed in Fig. S. 9, the differences in future climate data from the GCMs were found to be the largest source of uncertainty, consistently contributing over 60%. This contribution is more significant during NF than DF, as discussed in Section 3.6. Fig. 6 specifically highlights the uncertainty contributions attributed to hydrological models.

The uncertainty contributions from hydrological models were most significant during the spring (Mar-May) and winter (Dec-Feb) periods, as shown in Table S5. The results of the analysis for each basin were as follows: For AD, the hydrological model uncertainty was most significant in spring (NF: 7.54%, and DF: 5.86%), with a maximum of 9.76% in June for NF and 7.54% in April for DF. In CJ, the highest uncertainties were also found for NF in winter (3.9%) and for DF in spring (3.96%). HCH showed the highest uncertainty in winter (NF:

6.09%, and DF: 5.5%), with a maximum in November (NF: 9.76%, and DF: 8.92%). For SJ, the most significant contributions were found in spring (NF: 5.58%, and DF: 3.88%). In the end, hydrological model uncertainties were more significant in months with lower runoff.

Figure 7: clarify what is being shown—number of drought events?

Answer)

Thank you for this critical question, which identifies that the figure and its description were unclear. You are correct to question it; this figure (now renumbered to Figure 8) does not show the number of drought events.

Instead, it shows the percentage contribution (%) of the hydrological model factors (HC, PL, and their interaction) to the total uncertainty of the future 3-month SDI value (which is the standardized metric we use for hydrological drought analysis).

We have revised the text in Section 3.9.3 (specifically the sentence introducing Figure 8) to explicitly state that this figure shows the percentage contribution to the uncertainty of the 3-month SDI.
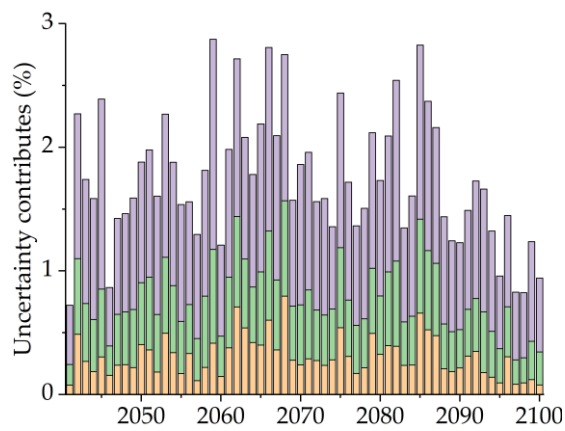
We have also rewritten the caption for Figure 8 to make this distinction clear.

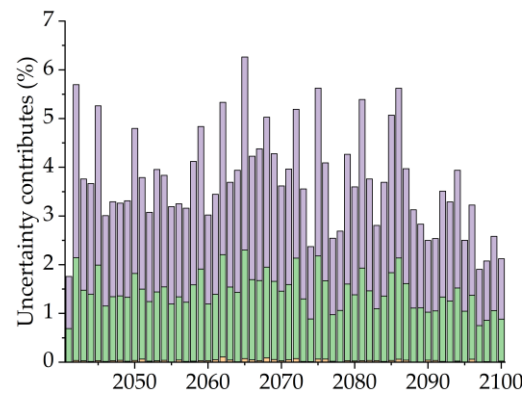3.9.3 Uncertainty contribution of future hydrological drought

The quantification of uncertainty in future hydrological drought was conducted using ANOVA. The uncertainty in future hydrological drought projections caused by SSP, GCM, and hydrological modelling parameters was clearly quantified by ANOVA. Fig S.10 shows the contribution of each factor to the total uncertainty. Among single-factor uncertainties, GCM contributed the most, averaging over 30%. The largest contributor to the total uncertainty, however, was the interaction between SSP and GCM, averaging over 50%.

Fig. 8 and Table 8 present the percentage contribution of hydrological modelling parameters to the total uncertainty of the future 3-month SDI value. The uncertainty contribution from hydrological model parameter estimation in future hydrological drought analysis averaged 2.7%, which is lower than that observed for future runoff projections. The uncertainty contribution from hydrological model calibration for future drought conditions was highest in HCH, followed by CJ, AD, and SJ, respectively. These results differ from those obtained in the runoff projections. The contribution of uncertainty in hydrological drought analysis decreased
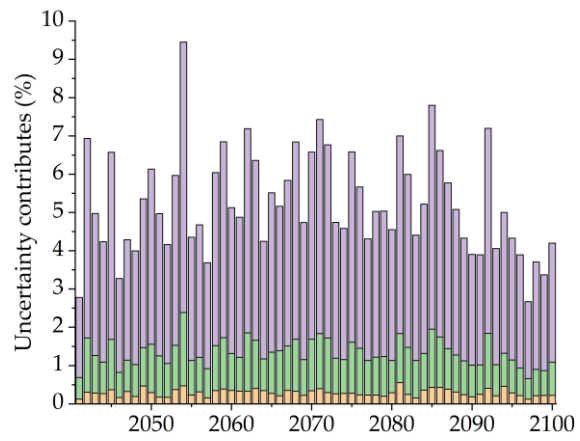
for AD and SJ, where uncertainty in future runoff projection due to hydrological model calibration was relatively high. In contrast, HCH showed high uncertainty contributions from hydrological model calibration in both runoff and drought analyses. Monthly runoff is a direct physical variable with high temporal volatility. In contrast, the SDI, used here to quantify hydrological drought, is a processed statistical indicator. It is calculated by accumulating and standardizing runoff over multi-month timescales. This integration process acts as a filter, effectively smoothing the high-frequency variability of the raw runoff series. Consequently, the absolute numerical fluctuation of the SDI is significantly smaller than that of the runoff itself. This reduced total variance in the drought index is the primary reason why the quantified uncertainty contributions appear lower and exhibit a different pattern compared to the runoff analysis. This highlights that while the underlying drivers of uncertainty are the same, their manifestation can differ depending on the temporal scale and the nature of the hydrological variable being analyzed. These findings confirm the necessity to separately analyze and consider uncertainties in future runoff projection and hydrological drought analysis.
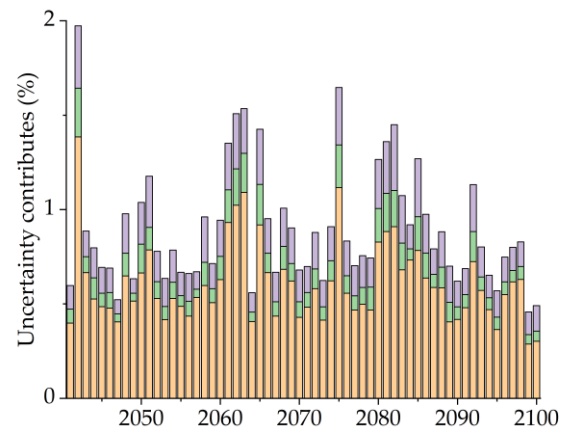

(a) AD   (b) CJ
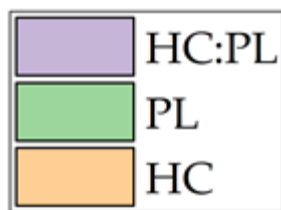
(c) HCH   (d) SJ

Figure. 8. Contribution of hydrological model parameters to the total uncertainty in the future 3-month SDI.

Table 3: provide a clearer definition, including what Q75 difference represents (difference in long-term discharge at the 75th quantile over different model parameterizations) and how the ratio relates to mean runoff. A large ratio meaning should be clarified.

Answer)

Thank you, this is a very important point. The original caption was too brief and failed to define the key metrics in the table, leading to valid confusion. As this table is critical for understanding the implications of calibration on drought-related flows, we have completely rewritten the caption.

Q75 (75% exceedance flow) is used here as our indicator for low-flow conditions.

The Q75 Differ (m³/s) column represents the physical difference in flow (range, max-min) of the projected Q75 values, when comparing results from parameter sets calibrated under different hydrological conditions (Dry, Normal, Wet).

The Ratio (%) column then expresses this physical flow difference ('Q75 Differ') as a percentage of the mean projected Q75 flow for that scenario.

Therefore, as you correctly inferred, a 'large ratio' signifies that the absolute difference in projected low flow (in m³/s) is large relative to the mean, indicating high sensitivity to the calibration conditions.

We have rewritten the caption for Table 3 to include these precise definitions, ensuring the table is now self-explanatory.

Table 3. Differences in projected low-flow (Q75) based on HC. Q75 Differ (m³/s) is the difference (range, max-min) in the magnitude of projected Q75 (75% exceedance flow) values when comparing results from models calibrated under different hydrological conditions (Dry, Normal, and Wet).

(unit: m³/s)

Basins	SSPs	NF		DF	
		Q_{75} Differ	Ratio (%)	Q_{75} Differ	Ratio (%)
AD	SSP2-4.5	7.24	10.28	7.00	10.42

	SSP3-7.0	7.04	9.58	7.71	9.56
	SSP5-8.5	7.43	9.32	7.88	9.94
CJ	SSP2-4.5	48.93	5.60	49.00	5.35
	SSP3-7.0	48.80	4.60	52.35	5.53
	SSP5-8.5	39.02	5.70	38.09	6.11
HCH	SSP2-4.5	5.84	12.67	5.86	13.93
	SSP3-7.0	5.55	13.86	5.95	12.86
	SSP5-8.5	6.03	12.86	6.44	14.62
SJ	SSP2-4.5	4.61	9.84	4.51	9.61
	SSP3-7.0	4.23	11.24	4.64	9.76
	SSP5-8.5	4.64	9.37	4.97	9.12

Figure S1: consider adding horizontal lines for the defined thresholds to indicate which condition each year would fall into.

Answer)

This is a very helpful suggestion for improving the figure's readability. We agree completely. We have updated Figure S1 by adding two horizontal lines at the defined thresholds (SDI = 0.5 and SDI = -0.5) to clearly visualize the 'Dry', 'Normal', and 'Wet' categories, as recommended.
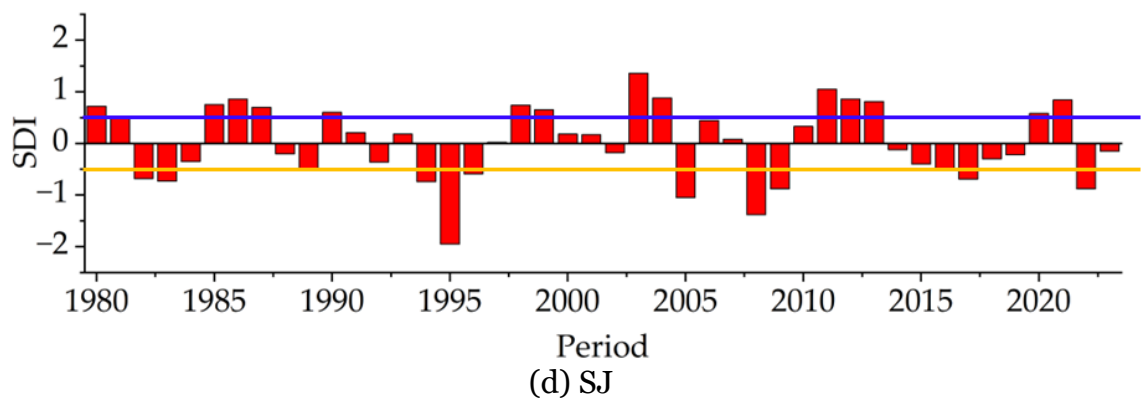

(a) AD

(b) CJ

(c) HCH

(d) SJ
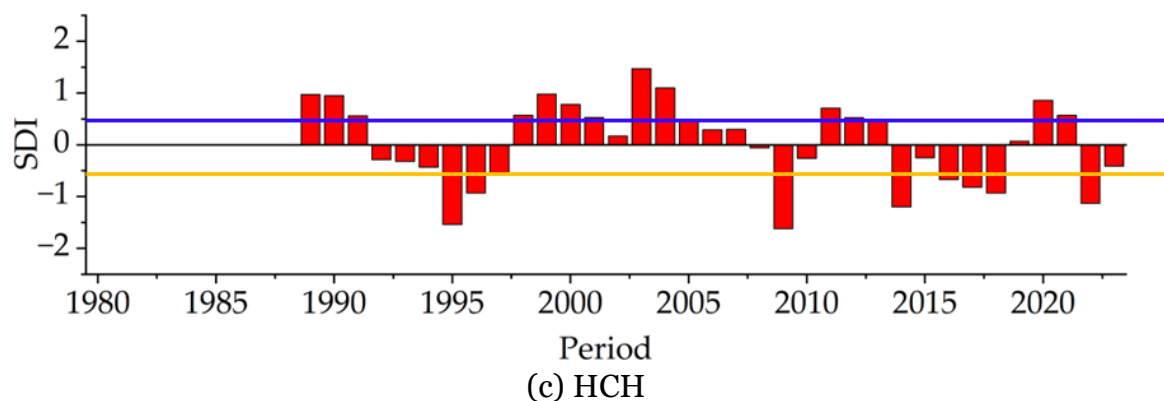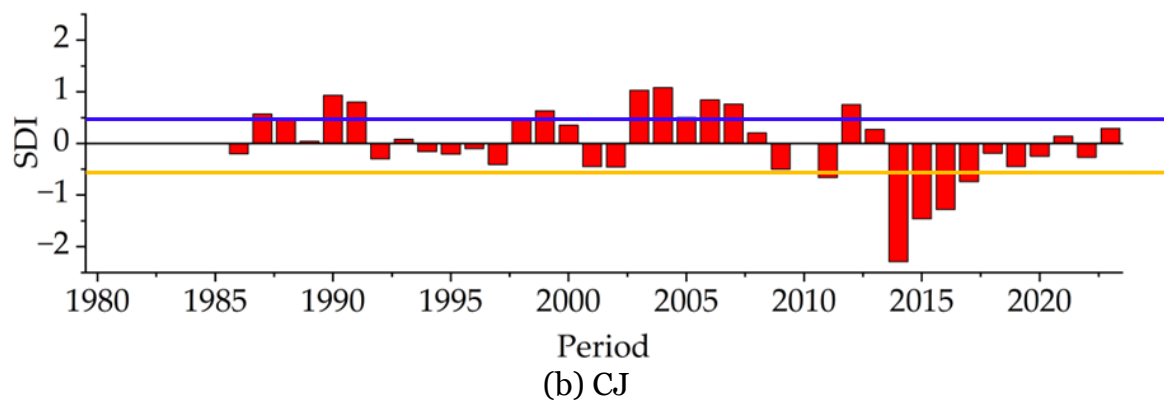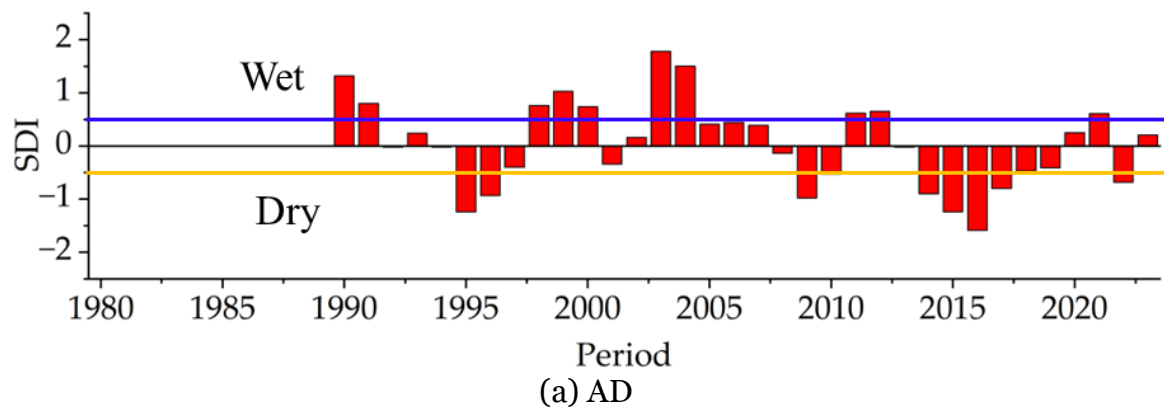
Figure S. 1. Hydrological conditions defined by Streamflow Drought Index for the historical period

Figures and Tables: axis titles are often missing; captions should be more extensive to improve interpretability.

Answer)

Thank you for this general, but very important, advice regarding the overall clarity of our figures and tables. We agree completely. Manuscript: In addition to addressing the specific items you pointed out, we have taken your advice to heart. We have conducted a thorough review of all figures and tables (including those in the Supplementary Information) to ensure every caption is comprehensive and self-explanatory, and that all axis titles are present and clear. We believe this has significantly improved the overall readability and quality of the manuscript, and we appreciate your constructive feedback.

Language and Terminology

Line 70: minor rephrasing could improve clarity.

Use "SWAT was used" instead of "the SWAT."

Avoid using "HC" for both the basin and an uncertainty driver.

Asnwer)

We are very grateful for these specific and helpful corrections to our language and terminology. We agree with all three points.

Line 70: You were correct, the original sentence ("...analyses... is...") was grammatically incorrect. We have revised this sentence for clarity and grammatical accuracy.

'the SWAT': Thank you. We have run a find-and-replace and corrected this term to 'SWAT' throughout the entire manuscript.

'HC' Abbreviation: This was an excellent point and a significant potential source of confusion. Thank you for catching this. We have changed the abbreviation for the Habcheon basin to HCH in all text, figures, and tables to avoid any overlap with 'HC' (Hydrological Conditions).

The future hydrological analysis considering uncertainty is essential for effective water management. These projections are largely based on General Circulation Models (GCMs) and hydrological models, which are critical tools for modelling the hydrological impacts of climate change. However, GCMs introduce significant uncertainty in future runoff prediction due to their inherent structural complexity and variability in scenario-based inputs (Broderick et al., 2016). This uncertainty has a direct impact on the accuracy of runoff predictions and poses a significant challenge to water resource management. The selection and use of GCMs have a crucial role in shaping these uncertainties, making the consideration of a variety of GCMs and shared socioeconomic pathways (SSP) scenarios essential for managing uncertainties and improving projections (Vetter et al., 2015; Chae et al., 2024a). Indeed, Shi et al. (2022) had shown how different evapotranspiration models embedded in GCMs affect runoff prediction, highlighting GCMs and Representative Concentration Pathways (RCPs) as major factors affecting uncertainty. Similarly, Lee et al. (2021a) had shown how the choice of GCMs

significantly affects prediction of water storage in wetlands under future climate scenarios. To understand these uncertainties, Wang et al. (2020) suggested the use of a broad ensemble of at least 10 GCMs, which allowed for a more comprehensive assessment of hydrological impacts and helped to reduce the inherent uncertainties associated with climate change. Thus, the use of a wide range of GCMs is an essential strategy for maximizing the effectiveness of water resource management under global climate change conditions.

References and Context

The authors properly credit related work and clearly separate their own contributions. Including Gao et al. (2020, DOI: 10.5194/hess-24-3251-2020) and Her et al. (2019, DOI: 10.1038/s41598-019-41334-7) in the discussion could further strengthen the context, as these studies seem quite relevant.

Answer)

Thank you for these excellent and highly relevant references. We have reviewed both papers and agree that they significantly strengthen the context and discussion of our findings.

We have integrated both references into our revised Section 4 (Discussion):

We have cited Her et al. (2019) in the first paragraph of the Discussion. Their finding—that GCM uncertainty is dominant for rapid components like runoff, while parameter uncertainty is dominant for slower components like groundwater—provides strong support for our results showing GCMs are the major uncertainty source for our runoff projections.

We have cited Gao et al. (2020) in the third paragraph of the Discussion. Their work, which also uses ANOVA to assess uncertainty in low flows (droughts), provides valuable context and corroboration for our own findings on the uncertainty drivers in hydrological drought analysis (Section 3.9.3).

4. Discussion

This study quantified the cascade of uncertainties caused by various factors in the process of projecting future runoff and analyzing future hydrological drought. Previous studies (Chegwidden et al., 2019; Wang et al., 2020) have reported that climate data from GCMs and SSP scenarios are the primary sources of uncertainty in future hydrological analysis. The results of this study also identified GCMs as the major contributor to uncertainty in future runoff analysis. This aligns with findings such as Her et al. (2019), who demonstrated that GCM uncertainty is dominant for rapid hydrological components, whereas parameter uncertainty becomes more significant for slower. However, recent research has begun to identify and quantify the cascade of uncertainties caused by factors beyond GCMs and SSP scenarios (Chen

et al., 2022; Shi et al., 2022). This study focused on the uncertainties inherent in the calibration of hydrological models, which are essential for future water resource management. Rather than seeking a single optimal parameter set, the central aim of this study was to quantify the uncertainty that arises from this very choice.

There have been limited studies that consider the uncertainties in runoff projection due to various calibrated parameter cases (Lee et al., 2021a). However, this study further subdivided the observation data used in the calibration period of hydrological model parameters by the amount of data and hydrological conditions to quantify the uncertainties more precisely. The results showed that hydrological conditions had a greater impact than the amount of calibration data period on the uncertainties in the calibration of hydrological model parameters.

This study went beyond merely projecting future runoff by also quantifying the cascade of uncertainties in the analysis of future hydrological drought using this runoff projection. Many studies on future drought prediction reported that hydrological drought becomes more complex and uncertain due to its association with human activities and the use of future climate data and hydrological models (Ashrafi et al., 2020; Satoh et al., 2022). For example, Gao et al. (2020), also using an ANOVA approach, found that for low flows, GCM and RCP uncertainty became increasingly pronounced. Most existing studies on future hydrological drought analysis focused on the severity and frequency of droughts. However, this study quantified the cascade of uncertainties that arise in the process of future drought analysis. Although the contribution of hydrological model uncertainty to future hydrological drought may be lower compared to future runoff projections, the characteristics of uncertainty differ between drought and runoff projections, clearly indicating the necessity to separately analyze and consider these uncertainties in future hydrological analyses.

Furthermore, the basin-specific characteristics presented in Table S1 help interpret the differing uncertainty contributions seen in the results. For example, in the hydrological drought analysis (Fig. 8), the uncertainty from model calibration was highest in HCH (5.56%) but lowest in SJ (0.26%), despite their similar areas (925 km² vs 763 km²). A key difference is that the SJ basin receives significantly higher mean annual precipitation (1,329.8 mm) compared to HCH (1,289.9 mm) and especially AD (1,045.7 mm). This suggests that basins with lower precipitation (like HCH and AD) may be more hydrologically sensitive to calibration data selection, leading to higher model-driven uncertainty, whereas the wetter conditions in SJ create a more robust (less sensitive) hydrological response regardless of calibration choice.

The title, "Insights into uncertainties in future drought analysis using hydrological simulation model," is appropriate.

The abstract is concise and complete, although it could mention the large contributions from GCM and SSP.

Original Comment (Eng): "The abstract is concise and complete, although it could mention the large contributions from GCM and SSP."

Answer)

This is a very valuable suggestion. We agree that mentioning the dominant contribution from GCMs provides crucial context for our novel findings on the hydrological model's uncertainty. We have revised the Abstract. We added a clause that, while our ANOVA results confirm that GCMs are the dominant source of total uncertainty, the specific focus of our study was to quantify the contribution from the hydrological model calibration process itself.

Abstract

Hydrological analysis utilizing a hydrological model requires a parameter calibration process, which is largely influenced by the length of calibration data period and prevailing hydrological conditions. This study aimed to quantify these uncertainties in future runoff projection and hydrological drought based on future climate data and the calibration data of the hydrological model. Future climate data were sourced from three Shared Socioeconomic Pathway (SSP) scenarios (SSP2-4.5, SSP3-7.0, and SSP5-8.5) of 20 general circulation models (GCMs). The Soil and Water Assessment Tool (SWAT) was employed as the hydrological model, and hydrological conditions were determined using the Streamflow Drought Index (SDI), with calibration data lengths ranging from 1 to 20 years considered. Subsequently, the uncertainty was quantified using Analysis of Variance (ANOVA). After calibrating SWAT parameters, the validation performance was found to be influenced by the hydrological conditions of the calibration data. Hydrological model parameters calibrated using a dry period simulated runoff with 11.4% higher performance in dry conditions and 6.1% higher performance in normal conditions, while hydrological model parameters calibrated using a wet period simulated runoff with 5.1% higher performance in wet conditions. While the ANOVA results confirmed that GCMs are the dominant source of total uncertainty, the uncertainty contribution from the

hydrological model calibration in estimating future runoff was analyzed to be 3.9~9.8%, particularly significant in the low runoff period. The uncertainty contribution in future hydrological drought analysis resulting from the calibration of hydrological model parameters was analyzed to be 2.7% on average, which is lower than that of future runoff projection.

Presentation and Complexity

The overall presentation is well structured, but due to the complexity of quantifying multiple uncertainty drivers, some sections are hard to follow. The ratio of results to discussion could be slightly adjusted, as some discussion points are already presented within the results section (but this is also a matter of taste).

The manuscript's novelty and strength lie in interpreting all the uncertainty drivers collectively, this is stated. But in the Abstract and Conclusion, the focus lies on the contributions of model uncertainty, the reasoning behind that could be made more clear. Also, I was expecting something like Figure S.9. within the manuscript, as this gives a great overview of all drivers' contributions in my opinion.

Answer)
We would like to once again express our sincere gratitude for your final overarching comments on the manuscript's Presentation and Complexity. We agree with your assessment entirely.

As you correctly pointed out, the methodology is complex, and our original presentation did not sufficiently guide the reader through our analytical framework or the narrative of our findings. Your feedback was crucial in helping us improve this.

Based on your suggestions, we have made the following key revisions, which are detailed in the point-by-point responses above:

To address the methodological complexity: We have added a new Flowchart as Figure 1 and revised Section 2.1 to refer to it. We also clarified the 3,600 simulation combinations (Sec. 2.7) and performed a thorough review of all figure and table captions (e.g., Fig. 3, 4, 8, S1, Table 3) to make them self-explanatory and clear, as you recommended.

To clarify the manuscript's narrative and focus: You made a crucial point about the apparent disconnect between our focus on "model uncertainty" (in the Abstract/Conclusion) and the comprehensive results (like Fig. S.9) showing GCMs are dominant. This was a key insight.

We have revised the Abstract and Conclusion (Sec. 5) to first acknowledge that GCMs are indeed the dominant source, and then clarify that the specific novelty and focus of this

study is the quantification of the often-overlooked hydrological model calibration uncertainty.

This directly addresses your excellent point about Figure S.9. As detailed in our response (and modified in Sec. 3.8.2), we now explicitly introduce Fig. S.9 in the text as the comprehensive overview of all drivers, while justifying that the main text figure (now Fig. 7) is presented to specifically detail our novel findings.

We are confident that these revisions, guided by your detailed and insightful review, have significantly improved the clarity, focus, and overall strength of the manuscript. We thank you again for your valuable time and constructive feedback, which has been invaluable in enhancing our paper.

Sincerely,

Kim Jin Hyuck

on behalf of all authors