

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 CO₂ 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.

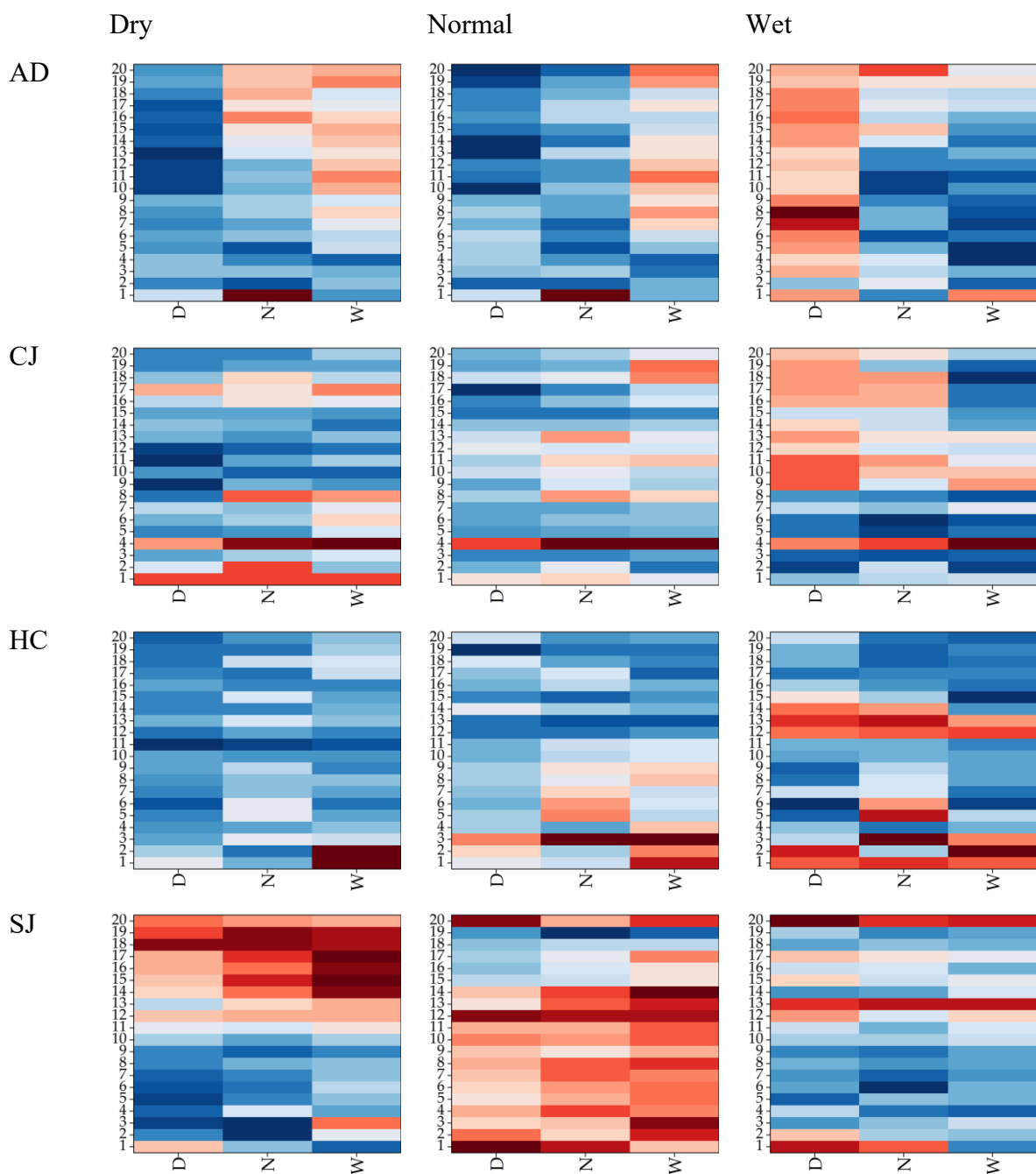
Comment 3: I suggest using blue (i.e., good) colour for Figure 4?

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:

Basins Hydrological conditions for validation period



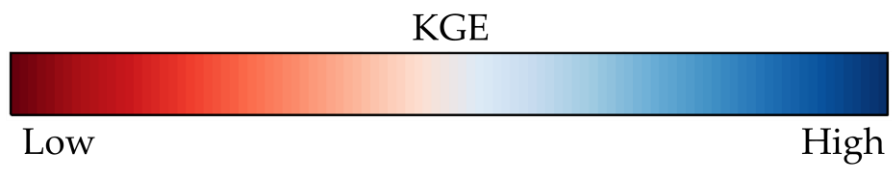


Figure. 4. KGEs classified by hydrological conditions for the calibration-validation period

Comment 4: I assume that the paper used the QQM bias corrected GCM data as input into SWAT 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:

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

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_{seep} 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 (m^3/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