Reply to Editors' and Reviewers' Comments

Legend

Editors' and reviewers' comments

Authors' responses

Direct quotes from the revised manuscript

The Editor:

Dear authors,

Thank you for this article exploring important issues regarding operation of river infrastructure under change. Two experts have reviewed your manuscript, and while they were supportive overall, they suggested some significant changes to framing and discussion. Thus, we will reconsider the article after major revisions.

In addition to the author comments, I also have some comments on the article, which are included at the bottom of this message.

Both reviewers emphasise the lack of discussion and limitations. Please give due attention to this in your revision. It is good to see the start of one in your response to Reviewer 1, but please note my comments below. Note, if a reviewer has explicitly asked for discussion, please do so--e.g. your response to Review 2's comment #3 is not acceptable (i.e. yes, your method is consistent with other studies and yes, this dataset seems relatively robust, but it is still a limitation).

When you prepare your response document (for any journal), it needs to be clear in the response document which parts of the text have been changed. E.g. you provide large paragraphs in red text of "direct quotes from revised manuscript" but it is unclear which parts are new versus from the previous version. This is particularly important in this case because you are not yet at the stage of submitting your track changes manuscript.

Response: Thank you for your review and for the careful assessment of our manuscript. We also sincerely appreciate the time and effort invested by the two reviewers. We fully acknowledge the key points emphasized by you and the reviewers—particularly the need to clarify the framing in the Introduction and to substantially strengthen the Discussion, including a clear and explicit statement of limitations.

Accordingly, we have systematically revised the manuscript to better articulate the novelty and scope of applicability. We have also added dedicated text on uncertainty and limitations, explicitly addressing Reviewer 2's Comment #3 rather than only justifying our methodological choices.

We also acknowledge your guidance on preparing the response document. In the revised response, we provide point-by-point replies and clearly indicate where changes were made in the manuscript (with section/line references), so that reviewers can readily distinguish newly added text from content retained from the previous version.

Details of the revisions are provided below.

Specific comments

The framing of the article in the introduction is unclear--please improve it. Specifically, you say that existing efforts do not leverage "actual reservoir operation data"--please give more detail. Further, you say that "A state-of-the-art tool that can scientifically mine massive historical operating data" is needed--what do you mean by "massive", do you mean multi-site? Multi-site is not what you do in the paper. Overall, please try your best to better frame the novel aspects of this contribution, and pay close attention to the reviewers' requests in this regard.

Response: Thank you for this helpful comment. We agree that the framing in the original Introduction was not sufficiently clear, particularly regarding what we meant by "actual reservoir operation data" and "massive historical operating data." In the revised manuscript, we explicitly clarify that "actual reservoir operation data" refers to observed operating records that directly reflect operational decisions, i.e., inflow/outflow/storage time series, and we explain that many existing drought-projection studies rely on generic rule-based operating representations with empirically calibrated parameters, without explicitly using such observed records to constrain or evaluate the operating representation (Introduction, revised Lines 69–82).

We also revise the wording around "massive" to avoid any implication of a multi-site dataset: our focus is on systematically learning from long-term historical operating records, and we specify that the learning is conducted over periods with relatively stable objectives and constraints (Introduction, revised Lines 87–89).

Finally, we revised the surrounding paragraphs (especially the first sentence of the last paragraph in Introduction, revised Lines 120–123 as shown below) to better align the stated novelty and scope of the study with the reviewers' requests and the actual analysis performed.

Here, we aim to advance current reservoir-related drought assessment frameworks by (i) replacing traditional process-based hydrological models with a fully data-driven LSTM framework for hydrological drought quantification, and (ii) explicitly exploring the adaptive performance of optimal operating policies under future climate change.

The title indicates a focus on a reservoir-regulated case study, but the case study map does not show the regulated reach, nor does Section 2.1 characterise it. Please rectify in both cases. Also, please make it clear early (e.g. 2.1) the length of the reach downstream that is directly influenced by the reservoir operation.

Response: Thank you for pointing this out. We have revised both the case-study map and Section 2.1 to explicitly define and visualize the reservoir-regulated reach examined in this study. Specifically, in Figure 1 we now label the Ankang Reservoir and the downstream Ankang Hydrological Station (control section) and update the caption to state that the reservoir-regulated reach extends ~30 km downstream from the dam to the station (Figure 1, revised caption; Section 2.1, revised Lines 142-146). Because the dam–station distance is short relative to the basin scale, it is difficult to distinguish on a basin-wide map; we therefore added clear labels and stated the reach length explicitly in both Section 2.1 and the figure caption.

The abstract and conclusions have a strong emphasis on "reconcil[ing] the tradeoff" between drought and hydropower, but I find the discussion of this lacking. Rather than removing this language as you suggest to Reviewer 1, I would rather you use your existing results to explore this issue. In particular, I find it curious that you have not included the Pareto curve results in the main manuscript, since this curve is an important source of information about the strength of tradeoffs--or alternatively, whether there exists a solution that can meet the different criteria simultaneously without trading off the performance. Given the shape of your curve, what can you say in this case? Please revise accordingly (including abstract/conclusions if necessary).

Response: Thank you for this comment. We now include the Pareto curve in the main manuscript (Figure 10) and provide an expanded discussion in Section 4.5 (Lines 617–629) to evaluate the future hydropower and drought performance of the Pareto-optimal policies relative to the historical operating policy. Specifically, the reliability-oriented solution, namely the one with the highest power-generation guarantee rate, performs well across most future scenarios compared to the historical operating policy, except for drought frequency (as shown in Figure 11), whereas the hydropower-maximizing solution does not show consistent advantages. Moreover, hydropower generation is broadly similar across the Pareto set in future periods, suggesting limited differentiation in this metric. Together, these findings indicate that, for the Ankang system, the reliability-oriented solution is a more defensible candidate for implementation.

Also, regarding the Pareto plot:

- The black dot is marked as "Historical simulation from the observations" in the legend and "historical operation derived from observed records" in the caption. Which is it?

- Assuming it's the latter, please provide further discussion as to why the black dot is so far from the Pareto optimal solutions. Is it because the operators are not operating it properly, or are the other operational constraints not considered in your methodology, and/or has operation changed over time (sensu Reviewer 1's comment)? Depending on your answer, this could be a key point to discuss in your limitations.

Response: Thank you for noting this inconsistency. The black marker represents the historical operation derived from observed operating record. We have therefore revised the figure legend to use consistent wording with the caption and the manuscript text (Figure 10; revised legend).

Regarding why the observed historical operation lies far from the Pareto-optimal solutions, we emphasize that this likely reflects additional real-world objectives and constraints that are not explicitly represented in our two-objective optimization, such as flood-control requirements, water-supply reliability, ecological releases, and operational risk preferences. We have added this clarification in the revised manuscript (Lines 617–619). We also provide an objective comparison of the operational performance of the historical operation and the Pareto-optimal policies under plausible future scenarios, including both hydropower generation and drought-related metrics (Lines 614–633; Figures 10–11).

Lastly, regarding LSTM, your response to reviewer 1 clarifies that you have applied LSTM to only one catchment. This is not consistent with guidance in the literature such as https://doi.org/10.5194/hess-28-4187-2024. Please cite this article, discuss the issues it raises, and justify your modelling decisions.

Response: Thank you for this important comment. We agree with recent guidance for rainfall–runoff LSTM modeling, which cautions against training LSTMs on a single basin and generally recommends multi-basin training to improve robustness and generalization, particularly under rare and extreme conditions (Kratzert et al., 2024). We have now cited this HESS Opinions article and addressed the issues it raises in Section 4.5 (“Limitations, uncertainties, and implications”, Item II) of the revised manuscript.

Following Kratzert et al. (2024), we understand that multi-basin regional training can improve overall streamflow performance and, in particular, better represent extremes (e.g., flood peaks), largely by expanding the training envelope through increased sample diversity. Since our study focuses on drought-relevant low-flow behavior, we found that the single-basin LSTM achieves satisfactory NSE performance for our case study, which supports the subsequent analysis of relative changes in drought metrics.

Nevertheless, we acknowledge that single-basin training constrains the training envelope to the historical hydroclimatic and operational conditions of the basin, which may reduce extrapolation reliability. Our choice of single-basin training was primarily driven by limited access to harmonized reservoir-operation records across regulated basins and by our focus on diagnosing regulated-system behavior in the case-study basin. We have added this clarification in Lines 603–613 of the revised manuscript. As data availability permits, we will conduct multi-basin (or multi-reservoir-system) training followed by fine-tuning to the target basin, and explicitly evaluate how this training strategy affects the simulation of drought-relevant low flows and extreme drought events.

Reference

Kratzert, F., Gauch, M., Klotz, D., and Nearing, G.: HESS Opinions: Never train a Long Short-Term Memory (LSTM) network on a single basin, Hydrol Earth Syst Sc, 28, 4187-4201, 10.5194/hess-28-4187-2024, 2024.

Thank you for your contribution to HESS.

Thank you for your careful handling of our manuscript and for the constructive guidance provided by you and the reviewers. We have revised the manuscript accordingly and hope that this version adequately addresses all concerns raised. We would be happy to make further revisions if needed.

Kind regards

Keirnan Fowler

Reviewer #1:

General comments

The authors present a study projecting future hydrological drought in the Upper Hanjiang River Basin by coupling LSTM-based hydrological models with reservoir operational models. I like the integration of LSTM hydrology and reservoir operation (although have some questions around the need for LSTM-based hydrology in its' apparent implementation, see below). I have some concerns around the calibration methods and how the authors have interpreted some key results. I also think there needs to be more discussion of hydrological and operational non-stationarity in the context of 'mining' systems dynamics using a machine learning model.

The manuscript is in general well written, with clear communication and good quality figures. But I think some substantial revisions are required to address the comments below.

Response: We sincerely thank Referee #1 for the thorough and constructive general comments. We acknowledge the referee's main concerns regarding (i) the calibration and evaluation strategy of the hydrological and reservoir-operation models, (ii) the interpretation of key results, and (iii) the need for a deeper discussion of hydrological and operational non-stationarity when using machine-learning-based approaches to explore system dynamics. We have addressed these concerns by conducting additional analyses and substantially strengthening the Discussion section in the revised manuscript. The revisions are described below in a point-by-point manner. We hope these changes satisfactorily address the referee's concerns, and we would be happy to make further revisions if needed.

Specific comments

Line 44: I would add Oceania here as multiple significant, record-breaking droughts have impacted Australia over the past ~20 years, with climate change a contributing factor.

Response: Thank you for this helpful suggestion. We have revised the text in the Introduction to include Oceania among regions affected by increasingly frequent hydrological droughts (refer to Line 45 in the revised manuscript).

Line 47: I find it a bit odd to say: "the time series of land temperatures." You could just replace with "land temperatures are projected to…"

Response: Thank you for the suggestion. We have revised the wording by replacing “the time series of land temperatures” with the more standard phrasing “land temperatures are projected to continue to rise” (refer to Line 48 in the revised manuscript).

Lines 51 to 63: To me, it’s an interesting question of whether you can separate the impact of the dam itself on the flow regime from the opportunity it provides for water abstraction and water resources development. By itself, I agree with Wanders and Wada that dams can buffer against low flow impacts by releasing passing flows. But let’s not forget that it’s because of the dam being there that enables much more intense consumptive water use for various industries. The net impact may be that downstream users are more impacted because of the water abstractions and diversions below/from the dams.

Response: We thank the reviewer for this insightful comment. We agree that the impacts of reservoirs on hydrological droughts arise from two distinct but interacting mechanisms: (1) hydraulic flow regulation by the reservoir itself, which can buffer low-flow extremes, and (2) intensified consumptive water use enabled by reservoirs, such as irrigation expansion and other anthropogenic withdrawals, which may exacerbate downstream drought conditions.

Following the reviewer’s suggestion, we have revised the Introduction to explicitly distinguish between these two mechanisms. The revised text clarifies that the dampening effects reported in strongly regulated basins primarily reflect hydraulic regulation (Wanders and Wada, 2015), whereas the intensification of hydrological droughts reported by Wan et al. (2018) is largely attributed to enhanced water abstractions associated with irrigation reservoirs. This clarification helps reconcile seemingly contrasting findings in the literature and emphasizes the region-dependent nature of reservoir impacts on future hydrological droughts.

The corresponding revisions have been made in Section 1 (Introduction), Lines 52–68.

At the same time, the rapid global expansion of reservoirs as a major manifestation of human intervention in river systems has introduced new challenges for assessing future hydrological droughts. Currently, more than 55,000 reservoirs have been registered by the International Commission on Large Dams, with a total storage capacity of 14,602 km3 (Eriyagama et al., 2020). Such an extensive storage capacity suggests that reservoirs can substantially affect hydrological drought characteristics by regulating the spatiotemporal distribution of river flows (Ho and Ehret, 2025; G. Ribeiro Neto et al., 2023). From the perspective of hydraulic regulation alone, reservoirs are often found to dampen low-flow extremes in strongly regulated river basins, particularly in Europe and North America, thereby alleviating drought severity during dry seasons (Wanders and Wada, 2015). However, reservoirs also enable intensified consumptive water use, including irrigation expansion and other anthropogenic withdrawals, which may counteract or even outweigh the buffering effects of flow regulation. For example, Wan et al. (2018) reported that irrigation reservoirs could increase the duration and intensity of global hydrological droughts by up to 50% during 2070–2099, largely due to enhanced water abstractions. Consequently, the net impact of reservoir operation on future hydrological droughts is highly region-dependent, reflecting the combined effects of hydraulic regulation, reservoir-enabled water use, and the heterogeneity of regional climate change.

Lines 64 to 78: You might like to discuss the approach by Culley et al. (2016, https://doi.org/10.1002/2015WR018253) in your literature review, where they assess the range of changes in climate that reservoir operational adjustments can adapt to. I am not an author on this paper. While it's less about only drought, and more about broader reservoir operating objectives, I think it's relevant to your study.

Response: We thank the reviewer for this valuable suggestion. After carefully reading Culley et al. (2016), we found that our study shares a similar conceptual basis with their work: we evaluate the operational adaptive capacity of reservoir systems by testing the performance of historically derived, real-world operating rules under future climate scenarios, and we further construct and examine optimized operating rules for comparison. Although Culley et al. (2016) is not solely focused on drought and instead emphasizes the broader adaptive range of multi-objective reservoir performance, their bottom-up perspective on assessing the limits of operational adaptability is highly relevant to our study. Accordingly, we have revised and expanded the related text in Lines 78–89 to highlight the importance of reproducing and characterizing realistic reservoir operational patterns when assessing reservoir responses to plausible future hydrological conditions.

These drought experiments demonstrated the feasibility of coupling hydrological and reservoir modules for such problems, but their conclusions may remain sensitive to empirical assumptions about reservoir releases when observed operating records (i.e., inflow/outflow/storage time series) are not explicitly used to constrain or evaluate the operating representation. As one of the most influential human-engineered interventions under a changing climate, reservoir systems warrant particular attention regarding the extent to which realistic operating patterns can sustain system performance under plausible future scenarios (Culley et al., 2016). Historical operating records contain rich decision-making information that reflects how operators have adapted release strategies to diverse inflow conditions (Zheng et al., 2022). Therefore, state-of-the-art tools that can systematically learn from long-term historical operating records during periods with relatively stable objectives and constraints are critical for capturing drought-relevant reservoir releases.

Line 70: What is the CSSPV2? Is it a hydrological model?

Response: Thank you for pointing this out. We have revised Lines 75-76 to clarify that CSSPV2 is a hydrological model and explicitly stated that it is coupled with a reservoir module.

Lines 73 to 74: I think this statement needs some additional evidence. What was the approach Ji et al. used to simulate reservoir operations? I think you just need one more line to demonstrate that it does not "consider… actual reservoir operation data."

Response: Thank you for this helpful comment. We have added one sentence in Lines 73–78 clarifying that Ji et al. (2023) represented reservoir operations using the generic rule-based scheme of Hanasaki et al. (2006) with empirically calibrated parameters, rather than being driven by observed reservoir operation records.

Reference:

Hanasaki, N., Kanae, S., and Oki, T.: A reservoir operation scheme for global river routing models, J Hydrol, 327, 22-41, 10.1016/j.jhydrol.2005.11.011, 2006.

Line 74 to 76: Yes I agree in principle, but also consider the possibility of reservoir operational procedures changing over time due to policy or infrastructure updates. If the change is significant, it may mean that older data is less relevant due to non-stationary operating conditions.

Response: We thank the reviewer for this important and thoughtful comment. We agree that reservoir operating procedures may evolve over time due to policy adjustments, infrastructure upgrades, or changes in management objectives, which could lead to non-stationary operating conditions and reduce the relevance of older records. To acknowledge this limitation, we have revised Lines 87–89 and 576-586 to clarify that the proposed data-driven approach is intended to mine historical reservoir operating data during periods with relatively stable objectives and constraints, where major structural or policy shifts are absent. This revision explicitly defines the applicability domain of the approach and highlights the need for caution when substantial non-stationarity in reservoir operations is present.

Lines 87 to 91: In these cases, what deficiencies in the traditional hydrological models are the AI models trying to address? Why would a "fully artificial intelligence-based simulation" provide new insights? I think this is a fairly generic statement and needs to be better linked to the research objectives or gaps here.

Response: Thank you for this insightful comment. We acknowledge that the original statement was overly generic and somewhat subjective in its wording. In response, we have revised the manuscript to remove the vague claim that a “fully artificial intelligence–based simulation” would inherently provide new insights. Instead, we now ground this discussion in the existing literature (Arsenault et al., 2023) and explicitly clarify the conceptual and practical differences between LSTM-based models and traditional process-based hydrological models. Specifically, we now state that traditional hydrological models (e.g., VIC and CSSPV2) typically rely on basin-by-basin calibration and require extensive physiographic inputs and parameterization, which can limit their transferability across regions. By contrast, LSTM-based models learn system behavior directly from large-sample historical data and have been shown to better capture nonlinear dynamics (Tran et al., 2025).

References:

Arsenault, R., Martel, J.-L., Brunet, F., Brissette, F., and Mai, J.: Continuous streamflow prediction in ungauged basins: long short-term memory neural networks clearly outperform traditional hydrological models, Hydrol Earth Syst Sc, 27, 139-157, 10.5194/hess-27-139-2023, 2023.

Tran, H., Zhou, T., Tan, Z., Fang, Y., and Ruby Leung, L.: Improving the prediction of daily reservoir releases over the CONUS using conditioned LSTM, J Hydrol, 661, 133750, 10.1016/j.jhydrol.2025.133750, 2025.

The corresponding paragraph is revised in Lines 90–105 of the manuscript and also shown as follows.

Against this background, machine learning (ML) offers a promising complementary approach to reproducing historical reservoir operation processes. A range of data-driven ML models, including artificial neural networks (ANN) (Özdoğan-Sarıkoç et al., 2023), nonlinear autoregressive models with exogenous input (NARX) (Yang et al., 2019), and long short-term memory networks (LSTM) (Tran et al., 2025), have been applied to simulate reservoir operations using large-sample historical records. Among them, LSTM-based models have demonstrated particularly favorable performance. Embedding physical mechanisms or operational constraints can further enhance their ability to represent operational behaviors under hydrological extremes, thereby allowing for a more accurate representation of high- and low-flow dynamics (Zheng et al., 2022). Building on this line of research, coupling an LSTM-based reservoir operation module with an LSTM-based hydrological process model can offer a pathway towards an integrated data-driven framework for more automated drought diagnosis. This direction is motivated by key limitations of traditional process-based hydrological models (e.g., VIC and CSSPV2), including their reliance on basin- specific calibration and substantial requirements for physiographic inputs and parameterization (e.g., topography, land use, and soil properties), which together constrain model transferability across regions (Arsenault et al., 2023).

Lines 94 to 95: I don't see how reservoir operation optimisation is a nature-based solution? It's artificial water regulation infrastructure. I think you should just remove the reference to NBS here, since you don't refer to it again in the manuscript. Your point still stands that optimisation is a good option because it doesn't require additional capital.

Response: Thank you for this helpful comment. Following this suggestion, we have removed the reference to nature-based solutions (NBSs) from the manuscript.

Lines 96 to 99: Consider some of the work by Wenyan Wu and colleagues (https://scholar.google.com.au/citations?hl=en&user=7N-YnaQAAAAJ&view_op=list_works) which I think do include drought performance metrics in reservoir optimisation approaches.

Response: Thank you for this helpful comment. We have carefully reviewed the recent works by Wenyan Wu and her colleagues (Huang et al., 2025; Huang et al., 2026) and agree that several of their studies explicitly consider drought-related performance metrics in reservoir operation and optimization frameworks. In particular, these studies assess the impacts of optimal operating rules on water supply systems under dry conditions using water deficit-based performance measures.

In response to this comment, we have revised the manuscript to explicitly acknowledge and cite these contributions at the appropriate locations. At the same time, we clarify that such metrics primarily reflect the impacts of dry conditions on managed water supply systems, rather than explicitly quantifying hydrological drought states using drought indices. We further emphasize that, despite these important advances, it remains unclear whether and how integrating such optimal operating strategies into current water management regimes would ultimately improve or deteriorate basin-scale resilience to hydrological drought extremes under climate change. The relevant text has been revised accordingly in Lines 110–119.

References:

Huang, J., Wu, W., Maier, H. R., Hughes, J., Wang, Q. J., and Cao, Y.: Comprehensive framework for long-term reservoir management under deep uncertainty, Environ Modell Softw, 195, 106740, 10.1016/j.envsoft.2025.106740, 2026.

Huang, J., Sangiorgio, M., Wu, W., Maier, H. R., Wang, Q. J., Hughes, J., and Castelletti, A.: Solving the robustness puzzle: The joint impact of optimization approach, robustness metrics, and scenarios on water resources management under deep uncertainty, Journal of Environmental Management, 373, 123540, 10.1016/j.jenvman.2024.123540, 2025.

Lines 127 to 128: Is this because there is a history of disaster-related damages in the basin?

Response: Thank you for the comment. We removed this sentence in the revised manuscript to avoid any confusion or unintended implications.

Lines 220 to 222: Can you provide some additional details on the inputs to the LSTM hydrology model? Were basin characteristics also considered, or are the inputs just meteorological variables? Were more than pr and t used as inputs? If it's just pr and t as inputs, I'm not sure why you used an LSTM model rather than a simple conceptual rainfall-runoff model, given the additional effort required in model calibration and challenges inherent in extrapolating outside training data (see Maier et al. (2023) for discussion here https://doi.org/10.1016/j.envsoft.2023.105776).

Response: Thank you for this insightful comment. We have clarified and expanded the description of the LSTM inputs in Lines 247–251 of the revised manuscript.

Specifically, the inputs to the LSTM consist of multiple meteorological variables, including precipitation, daily maximum and minimum air temperature, relative humidity, and wind speed, together with selected lag times. The lag structure for each input variable is determined using cross-correlation analysis to account for delayed hydrological responses and catchment memory effects. The model output is the near-natural reservoir inflow at time *t*. The data sources used in this study are described in detail in Section 2.2, and the overall model architecture is illustrated in Figure 3b.

Basin physiographic characteristics were not explicitly included as input variables in this study because our analysis focuses on a single reservoir catchment, within which such characteristics are effectively time-invariant and thus provide limited additional explanatory power for the temporal variability of inflow. We acknowledge that physiographic attributes can play an important role in regional-scale or multi-reservoir applications, and their inclusion should be prioritized in future work, which is also mentioned in the revised Discussion section (Lines 603-613).

Regarding the model choice, although the input variables are meteorological, the LSTM is employed to capture nonlinear rainfall–runoff relationships and temporal dependencies that are difficult to represent using simple conceptual models. We acknowledge the challenges of extrapolation outside the training range highlighted by Maier et al. (2023). To reduce this risk, the model is trained using a relatively long observational period, and future impacts are evaluated in terms of relative changes with respect to a predefined reference period.

Reference

Maier, H. R., Galelli, S., Razavi, S., Castelletti, A., Rizzoli, A., Athanasiadis, I. N., Sànchez-Marrè, M., Acutis, M., Wu, W., and Humphrey, G. B.: Exploding the myths: An introduction to artificial neural networks for prediction and forecasting, Environ Modell Softw, 167, 105776, 10.1016/j.envsoft.2023.105776, 2023.

Line 233: What are the differences in the hydroclimate regime in the validation period compared to the calibration period? Normally the differential split sample method you are describing here is supposed to contrast calibration and validation performance between two different climatic regimes to demonstrate out-of-sample performance for the calibration parameter set.

Response: Thank you for pointing this out. We apologize for the confusion caused by the wording in the original manuscript. In our study, the calibration/validation partition was implemented as a temporal hold-out split to evaluate out-of-sample performance in time, rather than to enforce two clearly separated climatic regimes as in a differential split-sample design. We have revised the manuscript accordingly to clarify this point.

Revised text (Lines 261-264)

For historical simulations, meteorological data from 1992 to 2020 were used. The year 1992 was reserved as the model spin-up period to minimize the influence of initial conditions, while the remaining data were divided into a calibration period (1993–2014) and a validation period (2015–2020) to evaluate the out-of-sample performance of the trained LSTM model.

Section 3.1.1: I think this section is missing a little bit of detail on calibration approaches. What objective function was used for calibration? What optimisation algorithm was used? You include this in 3.1.3 but I think it's part of 3.1.1 methodology. Maybe you can just combine these two sections.

Response: Thank you for this helpful suggestion. In the revised manuscript, we have added a concise description of the calibration objective function and the optimization algorithm in Section 3.1.1, including that model calibration was performed by maximizing the Nash–Sutcliffe efficiency (NSE) using the Adam optimizer. Please refer to Linex 264-266 of the revised manuscript.

To keep the methodology section well organized and avoid redundancy across modules, we retained Section 3.1.3 as a separate subsection to formally define the calibration objective function and evaluation metric (i.e., the average NSE based on inflow and outflow; Equations (9)–(10)). This is because the calibration objective involves both near-natural inflow simulation (Section 3.1.1) and human-regulated outflow simulation (Section 3.1.2), and Section 3.1.3 provides a unified formulation applicable to both.

Line 267: A note here that while NSE is commonly adopted in hydrological studies, its' squared error formulation means it will place far more emphasis on high flows compared to low flow performance. This is fairly well established in hydrology literature with various other objective functions or transformations used to overcome the issue when low flow performance is important. Can you offer some commentary on why high flows are more important for your study, considering your focus on hydrological drought?

Response: Thank you for your insightful comments. We agree that *NSE* tends to place more weight on high flows due to its squared-error formulation. In this study, *NSE* was adopted as the calibration objective primarily to ensure a reasonable overall reproduction of both reservoir inflow and outflow across the full flow range. In the Results section, when comparing the historical operating rules learned by the physics-guided LSTM model with the optimized operating rules, we assessed performance from multiple perspectives, including hydrological drought characteristics, hydropower generation, and power generation guarantee rate; notably, hydropower generation is calculated based on simulations over the entire flow regime. In addition, our analysis emphasizes relative changes in drought characteristics (future conditions relative to the historical reference period), rather than absolute values. We also acknowledge that the choice of calibration metric may, to some extent, influence low-flow and drought simulation performance; this limitation has been briefly discussed in the revised Discussion section (Lines 592-602).

Line 355: I disagree that such an NSE threshold is 'widely accepted.' There has been a lot of criticism in the literature over such arbitrary threshold-based approaches to model evaluation (see Knoben et al. (2019) https://doi.org/10.5194/hess-23-4323-2019), and recommendations to move towards purpose-dependent evaluation. I am not disputing that your model performance is good overall, but I would remove the reference to Moriasi et al. and this sentence. I would suggest in your case, you include some specific graphical or quantitative evaluation of low flow performance (such as flow duration frequency curves) because of your specific focus on hydrological drought. Figure 5 alone is insufficient as it's very difficult to separate performance across the flow regime. Based on a cursory inspection of Figure 5, it appears as though your simulations are biased high in the periods of lowest flow.

Response: Thank you for this insightful comment. After reading see Knoben et al. (2019), we realized that describing an NSE threshold (e.g., NSE > 0.5) as "widely accepted" is inappropriate. We therefore removed the reference to Moriasi et al. (2007) and deleted the corresponding statement from the original manuscript, and we will keep this important point in mind in future work.

Reference

Knoben, W. J. M., Freer, J. E., and Woods, R. A.: Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores, Hydrol Earth Syst Sc, 23, 4323-4331, 10.5194/hess-23-4323-2019, 2019.

Given our focus on hydrological drought, we followed your suggestion and added an explicit evaluation using flow duration curves (FDCs). Figure 5 has been revised to include FDCs for both inflow and outflow (Figure 5(b) and Figure 5(d)), enabling performance assessment across the full flow regime rather than relying solely on hydrographs. The FDCs show that the simulations generally reproduce the overall flow distribution reasonably well; however, as you noted, the lowest flows tend to be overestimated at high exceedance probabilities. We now explicitly acknowledge this low-flow bias as a limitation and will discuss its implications in the Discussion section, including directions for future improvement.

Finally, we clarify that the subsequent drought analyses are conducted within a relative framework (i.e., future changes relative to the ISIMIP3b_ref baseline), which reduces the sensitivity of our results to systematic biases in the absolute magnitude of low flows. We therefore consider the model suitable for hydrological drought analysis. The corresponding revisions have been made in Section 4.1 (Lines 388-393 and Lines 407-412), and Figure 5 has been updated accordingly.

Lines 388-393

As shown in Figure 5(a), the LSTM model reproduced the near-natural reservoir inflow well at the monthly scale, with NSE values of 0.95 and 0.93 for the calibration and validation periods, respectively. Figure 5(b) further evaluates the model performance across the full flow regime using flow duration curves (FDCs), showing that the simulated flow distribution generally follows the observed pattern across a wide range of exceedance probabilities.

Lines 407-412

FDCs in Figure 5(b) and 5(d) further indicate that simulated low flows tend to be overestimated at high exceedance probabilities, which may affect the absolute magnitude of simulated low-flow conditions. Nevertheless, since our subsequent analyses focus on changes relative to the ISIMIP3b_ref baseline, the influence of this systematic bias is likely to be reduced; therefore, the simulations are used for the subsequent hydrological drought analysis.

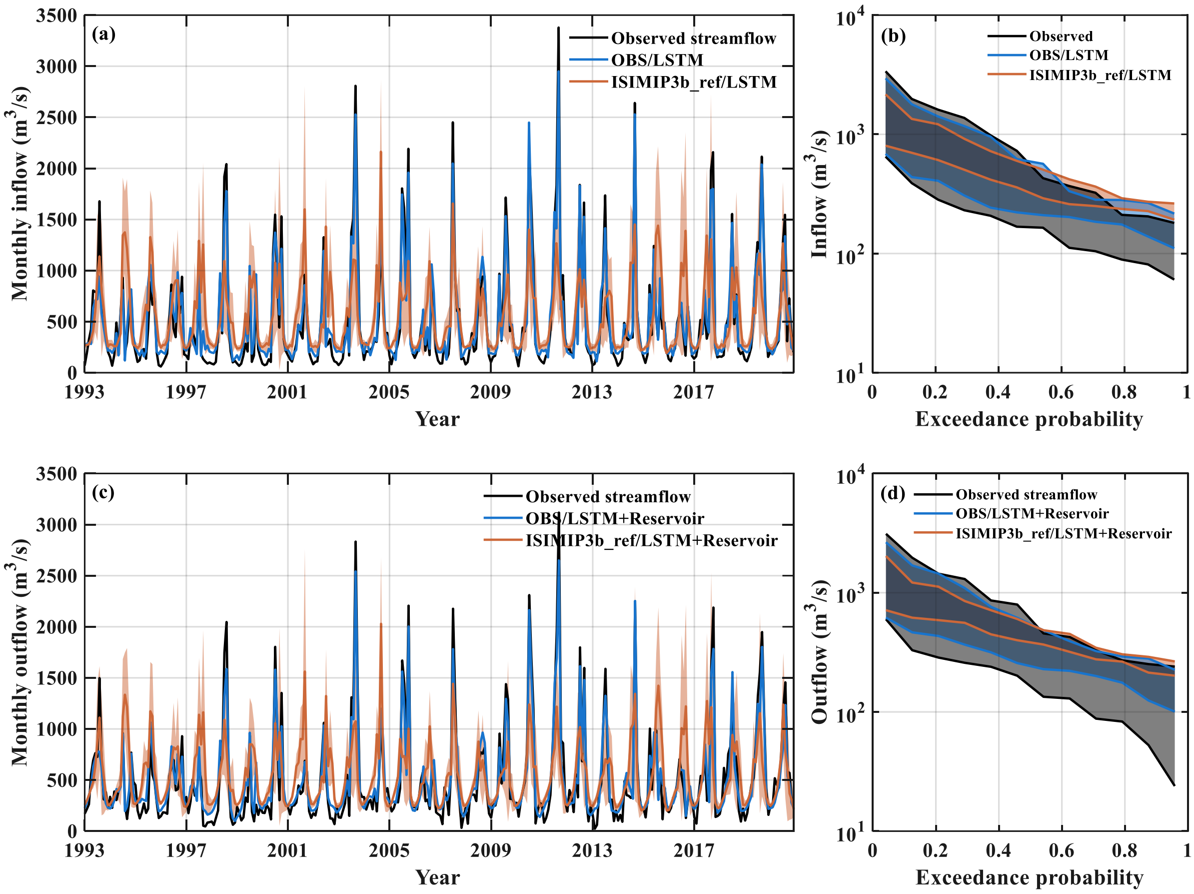

Figure 5. Evaluation of monthly reservoir inflow and release simulations. (a, c) Hydrographs of reservoir inflow and release; (b, d) Corresponding flow duration curves (FDCs). Simulations driven by meteorological observations (OBS/LSTM and OBS/LSTM+Reservoir) are marked as blue lines, while simulations driven by ISIMIP3b_ref forcings (ISIMIP3b_ref/LSTM and ISIMIP3b_ref/LSTM+Reservoir) are marked as orange lines. Shaded bands in (a, c) indicate the ensemble mean ±1 standard deviation of simulations driven by ISIMIP3b GCM forcings, while those in (b, d) denote the interannual range (min–max envelope) of annual FDCs.

Lines 548 to 551: I am not sure I agree with these points here. You say that your methods "can effectively reconcile the trade-offs between hydrological drought and hydropower benefits…" I interpret this as finding operational strategies that achieve better hydropower outcomes and reduce drought risk (which you mention as future research). Reconciliation means some compatibility or trade-off, and your methods only focus on hydropower rather than alleviating drought risk because optimisation only uses hydropower indicators. I think you just need to change the language here to something else to summarise your results. But I really don't think the optimal policy reconciles anything, it rather increases drought risk to pursue hydropower gains, which is very clear from your text and Figure 10.

Response: Thank you for this constructive comment. To address this concern, we have revised the corresponding statement in the Conclusions to avoid an inappropriate "reconciliation" claim. The revised text in lines 657-660 (also shown below) now emphasizes the existence of clear trade-offs between hydropower benefits and hydrological drought risk under the optimal operating rules.

3. Optimal reservoir operating policies at Ankang Reservoir, designed to maximize hydropower generation and power generation guarantee rates, highlight clear trade-offs between hydrological drought risk and hydropower benefits, especially in the near-future period (2031-2060).

Discussion section: This article is missing some key discussion on limitations of the adopted approach. There needs to be some acknowledgement of the limitations inherent in (these are examples and the authors should reflect on the key limitations and uncertainties) 1) the possibility of a non-stationary reservoir operating environment over the calibration period; 2) training the LSTM hydrology models on observed meteorology data and then using ISIMIP data for projections (rather than training on ISIMIP data); 3) the selected objective function for calibration which may bias model simulations towards high flows; 4) using the historic operating regime with far future climatic inputs (some recognition that operational strategies will co-evolve with climate and further anthropogenic development); etc.

Response: Thank you for this insightful comment. We agree that the original Discussion section did not sufficiently acknowledge key limitations and uncertainties associated with our framework. In response, we have substantially revised the Discussion by adding a new subsection (Section 4.5: "Limitations, uncertainties, and implications"), where we explicitly discuss the major limitations inherent in the adopted approach and clarify how these may affect the interpretation of our findings. Specifically, Section 4.5 now addresses: (i) the potential non-stationarity of reservoir operating conditions during the calibration period and limited climate-model sampling; (ii) the sensitivity of drought inferences to calibration choices and single-basin LSTM training, noting that NSE-based calibration may emphasize high flows at the expense of low-flow fidelity; and single-basin training can limit generalization, particularly for extreme events, and (iii) limitations in the design of our optimized operating policies, as the optimization was restricted to maximizing hydropower generation and power generation guarantee rates based on historical streamflow observations, and the resulting policy set was directly applied to future scenarios. Future developments could incorporate drought-event characteristics as explicit objectives or constraints and account for co-evolving human interventions. These additions improve the transparency of our assumptions and better constrain the scope and robustness of the conclusions drawn from this study.

The revised section 4.5 is also presented as follows.

4.5 Limitations, uncertainties, and implications

While the proposed framework provides a data-driven way to represent reservoir-regulated hydrological droughts under future climate scenarios, several limitations and sources of uncertainty should be explicitly acknowledged. These aspects are important for interpreting the robustness, scope, and broader applicability of our results.

I) Possible non-stationarity of the operating environment and limited climate-model sampling. In our framework, the physics-guided LSTM module learns reservoir operating behavior from historical conditions and is subsequently applied to the reference and future periods. This implicitly assumes that the governing operating objectives and constraints remain broadly stable, and that the learned decision logic is transferable across time. Our previous analyses indicate that, for the case investigated here, the reservoir-regulated hydrological response exhibits relatively stable patterns over multi-decadal timescales, supporting the feasibility of using a surrogate model to represent the whole-period operational behavior (He et al., 2023). However, such consistency is case-specific and may not hold in other basins where human interventions are stronger; therefore, caution is warranted when applying the proposed approach beyond the study basin. In addition, future projections are driven by a limited set of climate forcings (five ISIMIP3b GCMs), which may not fully span plausible hydroclimatic trajectories and extremes, thereby constraining the uncertainty range of the simulated drought responses. Key extensions include stress-testing surrogate transferability under plausible operational changes and quantifying climate-forcing uncertainty using a larger GCM ensemble.

II) Sensitivity of drought inferences to calibration choices and single-basin LSTM training. Simulated hydrological drought characteristics can depend on the objective function adopted during model calibration (Knoben et al., 2019). In this study, calibration primarily relied on NSE, which tends to emphasize high-flow conditions at the expense of low-flow fidelity. This may affect drought assessments because, at the low-flow end of the flow regime (i.e., high exceedance probabilities), streamflow deficits play a dominant role in shaping drought onset, persistence, and severity. Consistent with this concern, the FDCs in Figures 5(b) and 5(d) indicate a systematic high bias in low flows, implying that the absolute magnitude of drought intensity derived from simulated streamflow may be underestimated. Although our subsequent analyses focus on relative changes with respect to the ISIMIP3b_ref baseline, this sensitivity should be acknowledged. A more drought-facing calibration setup, e.g., low-flow-oriented objectives (log-transformed NSE), provides a direct path to reduce this sensitivity.

Beyond calibration, the physics-guided LSTM surrogate in our framework is trained using data from a single reservoir-regulated basin, which constrains the training envelope to the historical range of hydroclimatic and operational conditions represented in that system.

This design choice was primarily motivated by limited access to harmonized reservoir operation records across regulated basins and by our focus on the target reservoir-regulated basin. Recent guidance for rainfall–runoff LSTM modeling highlights that single-basin training can limit generalization, particularly for extreme events, whereas multi-basin training on hydrologically diverse data is often more robust (Kratzert et al., 2024). As data availability permits, a natural extension is to conduct multi-basin (or multi-reservoir-system) training followed by fine-tuning to the target basin, and to explicitly evaluate how such training affects the simulation of drought-relevant low flows and extreme drought events.

III) Limitations in the design of our optimized operating policy. Many reservoir optimization studies remain at a preliminary stage, where operating rules are optimized using historical observations and then deemed superior based solely on comparisons with historical performance. Although the observed historical operation, which is shaped by complex real-world objectives and constraints, is Pareto-dominated in the objective space of annual average hydropower generation and power-generation guarantee rate (Figure 10), it is not necessarily outperformed by most Pareto-optimal policies under future scenarios, particularly in terms of guarantee rate and drought frequency (Figure 11). In addition, a common challenge with Pareto-optimal solution sets is selecting a single implementable policy. By benchmarking the future performance of candidate Pareto solutions against the historical operating policy, we found that the reliability-oriented solution, that is, the one with the highest guarantee rate, performs better than the historical operating policy across most future scenarios, except for drought frequency, whereas the hydropower-optimal solution does not show consistent advantages. Moreover, hydropower generation is broadly similar across the Pareto set in future periods, suggesting limited differentiation in this metric. Together, these findings suggest that, for the Ankang system, the reliability-oriented solution is a more defensible candidate for implementation. Nevertheless, our drought-focused analysis optimized policies primarily for hydropower generation and guarantee rate and then directly applied them to future scenarios to examine trade-offs with hydrological drought mitigation. This indicates scope for further improvements in drought mitigation performance. Future work could incorporate drought-event characteristics as explicit objectives or constraints, and consider other human interventions (e.g., inter-basin water transfers and urbanization) to better bracket plausible drought-mitigation pathways under changing conditions (Wu et al., 2023; Firoz et al., 2018).

Technical corrections

Line 35 “pow” should be power

Response: Thank you for pointing this out. We have corrected the word in the revised manuscript.

Line 46: Ipcc should be IPCC

Response: Corrected. “Ipcc” has been revised to “IPCC”.

Lines 132: Extra parenthesis can be deleted

Response: Sorry for the oversight. The extra parenthesis has been deleted.

Line 186: You can remove the acronym BPTT because you don’t use it anywhere else in the manuscript

Response: Thank you for the suggestion. The acronym “BPTT” has been removed.

Line 307: d1 should be a subscript

Response: Thank you for pointing this out. “d1” has been formatted as a subscript (d_1).

Reviewer #2:

The manuscript investigates how climate change and reservoir operation jointly shape future hydrological droughts in a heavily regulated basin, using an Ankang Reservoir in the upper Hanjiang River Basin in China as an example. A hybrid framework was developed by coupling a LSTM-based hydrological model with a physics-guided LSTM reservoir operation module, which was then driven by ISIMIP3b CMIP6 projections (five GCMs, three SSPs) to simulate regulated streamflow in the near- (2031-2060) and far-future (2071-2100) periods. Hydrological drought characteristics are quantified using SSI-1, SSI-3, and SSI-12, and a multi-objective optimization (NSGA-II with RBF parameterization) is used to explore whether optimal operating policies can balance hydropower benefits and drought risks.

Overall, the paper addresses an important and timely topic, and the methodology is sound and clearly implemented. I believe the manuscript is suitable for publication after minor revisions to clarify the novelty, add some methodological details, and strengthen the discussion of limitations and applicability.

Response: Thank you for the positive and constructive assessment. We appreciate your recognition that the topic is timely and that the proposed methodology is sound and clearly implemented. In response to your suggestions, we have revised the manuscript accordingly by clarifying the novelty, adding key methodological details, and strengthening the Discussion on limitations and applicability (including a new Section 4.5). We believe these changes improve the clarity and robustness of the study.

Major comments

1. Clarify the paper’s novelty

The Introduction already reviews several related works (e.g., VIC-Reservoir and CSSPV2+reservoir frameworks for future drought projections, and recent studies coupling reservoir models with CMIP6). However, the specific advances of this study relative to those works could be articulated more explicitly.

Please more clearly state what is new in this manuscript: e.g., (i) replacing traditional hydrological models with a fully data-driven LSTM for inflow and outflow, and (ii) explicitly combining this hybrid model with multi-objective optimization of operating policies to examine drought–hydropower trade-offs under future climate scenarios.

It would help if the end of the Introduction contained a short, itemized list of the main contributions to distinguish this work from previous hybrid and reservoir–drought studies.

Response: Thank you for this constructive and insightful comment. We have substantially revised the final part of the Introduction to clearly articulate the novelty of this manuscript. Specifically, we now explicitly state that the main contributions of this study are:

(i) replacing traditional process-based hydrological models with a fully data-driven LSTM framework for hydrological drought quantification; and, (ii) explicitly exploring the adaptive performance of optimal operating policies under future climate change scenarios.

In addition, we have restructured the final paragraph of the Introduction to present these contributions in a concise and itemized manner, thereby clearly distinguishing this work from previous hybrid and reservoir–drought studies based on process-based models or conceptual operating rules. These revisions have been incorporated in Lines 120–135 of the revised manuscript.

2. Assumption of stationary operation policy when projecting future droughts

The study assumes that the historical operation policy learned by the physics-guided LSTM (1992–2020) is directly applicable to the reference and future climate periods (ISIMIP3b_ref and ISIMIP3b_fut experiments). This is a reasonable and often necessary assumption, but it should be discussed more explicitly as a limitation: operation rules in reality may adapt to changing demands, policies, or infrastructure.

Response: Thank you for this insightful comment. We agree that our future projections rely on the assumption that the historical reservoir operating policy learned by the physics-guided LSTM module remains applicable to the reference and future periods. Following your suggestion, we have now explicitly acknowledged this as a key limitation in the Discussion (Section 4.5, Lines 576–586). Specifically, we clarify that the projected drought responses under reservoir regulation should be interpreted as conditional on a “stationary operating policy” hypothesis, rather than representing fully adaptive future management outcomes. The revised text is also shown below.

I) Possible non-stationarity of the operating environment and limited climate-model sampling. In our framework, the physics-guided LSTM module learns reservoir operating behavior from historical conditions and is subsequently applied to the reference and future periods. This implicitly assumes that the governing operating objectives and constraints remain broadly stable, and that the learned decision logic is transferable across time. Our previous analyses indicate that, for the case investigated here, the reservoir-regulated hydrological response exhibits relatively stable patterns over multi-decadal timescales, supporting the feasibility of using a surrogate model to represent the whole-period operational behavior (He et al., 2023). However, such consistency is case-specific and may not hold in other basins where human interventions are stronger; therefore, caution is warranted when applying the proposed approach beyond the study basin.

3. Uncertainty and generality of the results

The study uses five ISIMIP3b GCMs and a single basin–reservoir case. While this is already a substantial effort, readers would benefit from a more explicit reflection on the scope and limitations. Please expand the discussion of uncertainties related to (i) the limited number of climate models, (ii) the single-case setting (UHRB, Ankang) and whether the conclusions are transferable to other types of reservoirs or climate regimes, (iii) A short paragraph in the Discussion or Conclusions explicitly addressing "limitations and future work" would strengthen the paper and guide follow-up studies.

Response: Thank you for this suggestion. We agree that the scope and limitations should be stated more explicitly. In the revised manuscript, we have expanded Section 4.5 ("Limitations, uncertainties, and implications") to address these points.

(i) Limited number of climate models. We now explicitly acknowledge that the analysis is based on five ISIMIP3b GCMs and discuss the associated limitation in capturing the full range of climate-model uncertainty (Lines 586–590).

In addition, future projections are driven by a limited set of climate forcings (five ISIMIP3b GCMs), which may not fully span plausible hydroclimatic trajectories and extremes, thereby constraining the uncertainty range of the simulated drought responses. Key extensions include stress-testing surrogate transferability under plausible operational changes and quantifying climate-forcing uncertainty using a larger GCM ensemble.

(ii) Single-case setting and transferability. We have strengthened the discussion on the single-basin, reservoir-regulated case study (UHRB, Ankang) and clarified that the inferred transferability is conditional on the assumed stability of hydroclimatic and operational conditions. We note that the learned behavior may not hold in other basins where human interventions are stronger; therefore, caution is warranted when applying the proposed approach beyond the study basin (Lines 583–586).

However, such consistency is case-specific and may not hold in other basins where human interventions are stronger; therefore, caution is warranted when applying the proposed approach beyond the study basin.

(iii) Limitations and future work. We also added an explicit “future work” extension, noting that, as data availability permits, a natural next step is multi-basin (or multi-reservoir-system) training followed by fine-tuning to the target basin, and an evaluation of how this affects drought-relevant low flows and extreme drought events (Lines 603–613).

Minor comments

1. In the Abstract, there seems to be a small typo in “pow generation guarantee rate”; please correct to “power generation guarantee rate”.

Response: Thank you for pointing this out. We have corrected the typo in the Abstract.

2. Consider defining “LSTM+Reservoir” more explicitly at its first appearance in the Abstract or Introduction (e.g., “a hybrid LSTM-based hydrological and physics-guided reservoir operation model”).

Response: Thank you for this helpful suggestion. To avoid potential misunderstanding at an early stage of the manuscript, we have removed the term “LSTM+Reservoir” from both the Abstract and the Introduction, and instead provide its full definition and description in the experimental design section (Section 3.3), where the modeling framework is introduced in detail. The revised Abstract has been updated in Lines 26–29 (also shown below).

A long and short-term memory (LSTM)-based hydrological model, coupled with a physics-informed LSTM reservoir model, is developed and driven by bias-corrected climate outputs from five global climate models to project future drought conditions under three scenarios (SSP126, SSP370, and SSP585).

3. Notation and acronyms. Please ensure that all acronyms are defined at first use (e.g., SSI, NBS, NSGA-II, RBF). Some are introduced in the text or caption but could be clarified earlier for readability.

4. It may help to add a short list of symbols for key variables (e.g., V, I, O, THP, PGR, D, S) either in the main text or Supplement.

Response: Thank you for these helpful suggestions. We have revised the manuscript to define all acronyms at their first occurrence. We have also added a list of symbols for the key variables in the Supplementary Material to improve clarity. It is also shown below.

Acronyms and Notation

For ease of reading, all important and notations in the main text are summarized below.

Acronyms

UHRB	Upper Hanjiang River Basin
SSI	Standard Streamflow Index
LSTM	Long and Short-Term Memory
GCM	Global Climate Model
SSP	Shared Socioeconomic Pathway
ML	Machine Learning
CMA	China Meteorological Administration
ISIMIP3b	Inter-sectoral Impact Model Intercomparison Project 3b
NSGA-III	Nondominated sorting genetic algorithm version III
FDC	Flow Duration Curve
NSE	Nash–Sutcliffe efficiency

Notation for key variables

f	Operating objective vector
π_θ^*	Optimal operating policies
w^H	Historical climate conditions
V	Reservoir storage
I	Reservoir inflow
O	Reservoir release
D	Duration of hydrological drought
S	Severity of hydrological drought

5. For the parallel coordinate plots in Figure 10, you might consider adding an arrow in the left to clarify which direction is "better" for each metric (e.g., upward is optimal for all axes) to help non-expert readers interpret the trade-offs.

Response: Thank you for your suggestion. We note that all axes are oriented in the same "better upward" direction, so an explicit arrow is not strictly necessary. Instead, to improve readability for non-expert readers, we have clarified the direction of improvement in the caption of Figure 11 by adding the statement: "Each axis represents an objective, with the optimal direction oriented upwards."