Reply To Referee#2 Comments

We thank referee for the valuable comments. Our replies to all comments are shown in blue and the original referee's comments are shown in black.

Summary: this manuscript attempts to show the value of seasonal flow forecasts for reservoir management by applying a simulation/optimization scheme and multicriteria decision making (MCDM) techniques to a reservoir system in South Korea. Seasonal Streamflow Forecasts (SSF) are used to run the model, and the performance is compared to that achieved by ESP and two deterministic scenarios consisting on the worst-case and 20yr observed droughts. The authors propose a metric to evaluate forecast value, consisting on the count value of times a forecast-simulation scenario outperforms the historical operation of the system. Then they evaluate the sensitivity of this forecast value to a series of methodological choices, including forecast lead time, MCDM technique, frequency of decision making, and type of streamflow forecast (deterministic or probabilistic) used.

General comments:

1. The authors tackle an important topic such as optimizing reservoir operations considering ever-increasing pressure on water resources and a less predictable hydrological regime due to climate change. The manuscript is well crafted, with attractive figures and an agile style which makes it enjoyable to read.

→ We sincerely appreciate your efforts in providing valuable feedback.

2. However, I am unconvinced that the material presented supports many of the claims and generalizations the manuscript makes. A fundamental concern is that of experimental design. The paper studies two reservoirs and three drought events, in what can hardly be considered a sample large enough to support claims about robustness.

→ We totally agree that our findings are limited to few reservoirs and drought events and thus we cannot draw robust general conclusions based on these limited cases. However, this is a limitation somehow 'intrinsic' to this type of simulation studies due to the rare occurrence of extreme drought events and the limited temporal cover of seasonal forecast data, which only became available since 1993. Since our results have shown the dependency of forecast value on reservoir systems and events, it is crucial to continue further testing to establish more general patterns in the skill-value relationship (if they exist), as well as to compare the performance between different forecast products. We hope that sharing our workflow and open-source code will stimulate and facilitate such further research.

In response to your comment, we will include these considerations as the very first point of sub-section 5.2 (Limitations and directions for future research) in the Discussion. We will also modify the title of the paper to 'Exploring the value of seasonal flow forecasts for drought management in South Korea' and clarify throughout the manuscript that such exploration is still based on a limited number of events and does not give a comprehensive answer yet regarding the skill-value relationship.

3. It appears that the events included are in all cases hydrological droughts with return periods of less than 20 years, because the 20-year drought always underestimates monthly flows. If this is actually the case, the experimental setup is predetermined to favor the ensemble over the deterministic methods, as the latter will underestimate flows and generate a poorer reservoir performance (evaluated ex-post).

→ Thanks for this comment, which made us realise that Figure 4 may be causing a misunderstanding regarding the magnitude of the 2014-2016 drought event. As shown in Figure R1 below, the cumulative reservoir inflows fell between the 20-year return period and the worst-case drought inflows for nearly two years, from June 2014 to May 2016 (it is only one out-of-season wet event in June 2016 that made the cumulative inflow catch up with and temporarily exceed the 20-year return period scenario). Given that the reservoirs in South Korea are designed to supply water for a year under a drought with a 20-year return period, this event is significant for practical reservoir operations. Since our simulation began in June 2014 and ran continuously until the end of 2016, it captures the impact of this severe drought event. To enhance the clarity regarding the magnitude of this drought event, we will include Figure R1 in the supplementary material of the revised manuscript.

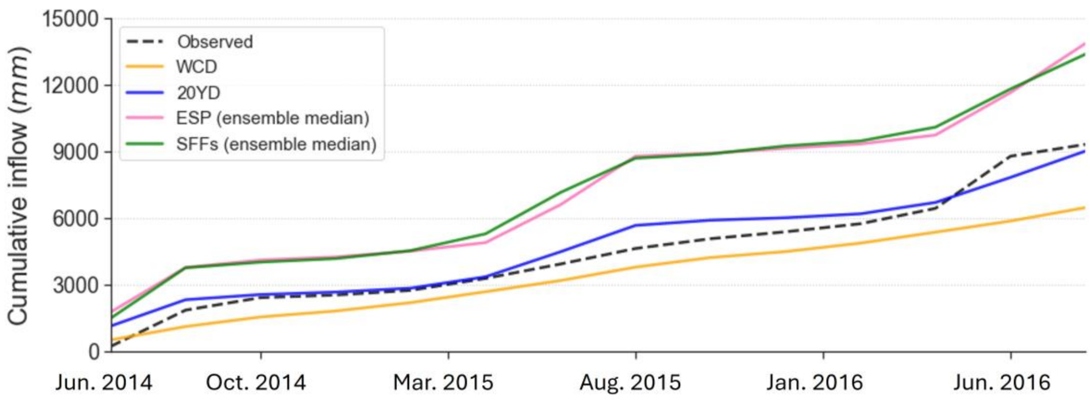

Figure R1: Cumulative flow observation (black dashed line) and forecasts (coloured solid lines) for Soyanggang-Chungju reservoir system for 2-month lead time. This figure corresponds to Figure 4(a) with 2-months lead time.

4. Concerning the historical operation of the system, from the example shown in figure 5 it appears that reservoir operators did not realize they were facing a drought event until Jun-Jul 2015. After that date they accumulate large supply deficits, which can be more moderate when some forecast information allows hedging at the beginning of the simulation period (late 2014).

→ We fully agree with your point. As you mentioned, reservoir operators did not recognise the impending drought until Jun-July 2015, which exacerbated the damage caused by the drought. This operational failure highlights the limitations of the current reservoir operations in South Korea and further motivate our study to enhance our reservoir operation methodology utilising seasonal forecasts.

5. The discussion section mostly indicates that the results presented here confirm or align well previous research by some of the coauthors or by other groups. It is not easy to determine what is the main, novel, contribution of this paper to the wider body of literature, aside from the methodology for evaluating forecast value based on the count of outperforming scenarios.

→ We agree that some of our key findings well align with (few) previous studies that explore the topic with a similar approach to ours (i.e. quantifying the value of seasonal forecasts by simulating their use in historical events). We believe we have provided some useful contributions though: a methodology to measure value (as recognised by the author); a workflow to assess the sensitivity of simulation results to key set-up choice (something that was never done before, to our knowledge, and is both a methodological contribution for others to use, and provide interesting insights on which of those choices are very important and which are not); an attempt at beginning to systematise simulation experiments by analysing multiple reservoirs in the same region and multiple drought events - an incomplete attempt, admittedly, but a start to overcome the "single case study" approach of previous papers.

6. I would suggest that the authors take advantage of the fact that they have a limited sample of cases and analysed them in more detail: what are the hydrological characteristics of the drought events under study? How is the interplay between the temporality of monthly flows and demand? Are these reservoirs operated under multi-year criteria? Can operating rules (hedging) be introduced in combination with streamflow forecasts? In this way, they could glean further insights from their experiments. In the current version, many of the claims made in the discussion seem speculative.

→ We appreciate your suggestions. We think that these suggestions such as hydrological characteristics of drought events could be useful to improve our paper and will include those results in the manuscript. For example, Figure R1 will be included in the revised manuscript, accompanied by an explanation of the reservoir operation designed to ensure a single year water supply during 20-year return period drought events.

Specific comments:

L91. How is the material presented here relevant to the wider community outside of South Korea? All insights that are scalable to other systems should be highlighted.

→ Thank you for this comment. We will highlight the insights of our study to the wider research community outside of South Korea in Section 6 (Conclusions) by underlining that 'This study explores the potential of SFFs in improving drought management, and our workflow, accompanied by open-source code, will facilitate the replication of this analysis in other regions around the world'.

L112. This contribution does not seem substantial enough to justify the publication of this paper in its current form.

→ We thank you for this comment. While proposing a new methodology to quantify forecast value (L112) is very important point of our study, the key contribution of this study lies in exploring the forecast value of seasonal forecasts for enhancing drought management and providing our workflow for further research across other regions. We will further clarify this in the manuscript.

L132. It is said that both reservoirs operate as one, but later in the results section there are shown separately. Why?

→ This is incorrect. All the results shown in this manuscript pertain to the Soyanggang-Chungju and Andong-Imha reservoir systems as combined entities rather than as individual reservoirs. Although Soyanggang and Chungju are individual reservoirs, they are managed through conjunctive (combined) operation. Taking into account this, we have analysed them as Soyanggang-Chungju reservoir system. (This also applies to Andong and Imha reservoirs.)

L150: Here it is said that PET was computed, but the Penman-Monteith (FAO) metod computes real ET. This was confusing. Also, it is mentioned that evaporation from the reservoirs was neglected. This was surprising, because once you are running a simulation model, adding an empirical evaporation estimate shouldn't be too challenging. What was the reason to neglect this evaporation term?

→ We disagree with this comment. The Penman-Monteith method has been widely used to compute PET (e.g. Cai et al., 2007; Córdova et al., 2015). Additionally, we have neglected the evaporation from a reservoir due to the lack of relevant empirical equations in South Korea regarding the reservoir evaporation. This is discussed in 5.2 'Limitations and directions for future research' (L.531-538).

L304: replace "simply" with "simple".

→ We will replace this.

L315: It is difficult in the present manuscript to identify such insights.

→ This comment is also related to the robustness of the study. Please see the reply to the second comment in 'General comments' on page 1. We believe that this issue can also be addressed more by clearly articulating our objectives, contributions, and limitations.

L361: I don't think it is possible to say that "generally" something is true in this context, because the number of scenarios analyzed is very limited.

→ This is also related to the robustness issue. We will remove 'generally' here and clarify our objectives and contribution as the new attempt for assessing the value of seasonal forecasts for drought management.

L406. One problem of using the scenario count as forecast value is that it does not take into account the magnitude of the deficits generated in either of the scenarios tested.

→ Thank you for this comment and we agree with your comment. We also acknowledge that our methodology cannot measure the quality of the value. We think that incorporating the hypervolume concept, which defined as the space enclosed by a set of points in a multi-dimensional space (While et al., 2006; Sanchez-Gomez et al., 2019) could enhance our method for measuring the magnitude (or quality) of the value. Therefore, we will further discuss this in Section 5.2 'Limitations and directions for future research'.

L422-431. Here we see various results, but little analysis goes into gleaning the reasons why the MCDM method impacts differently the forecast value for either reservoir and drought. Before it was said that both reservoirs are operated jointly, but the results presented here suggest otherwise. Please clarify.

→ We thank you for this comment. We hypothesize that the value can be influenced by the MCDM method, as well as the characteristic of analysed drought events.

When the Pareto front is plotted with the options selected by the different MCDM methods, a distinct decision-making trend can be found (Figure R2). Except for the variable weighting method (Simple Selective, Multi Weighting), which applies different weights based on storage volume status, other methods demonstrate consistent decision-making trend and order as illustrated in Figure R2. The method emphasizes storage availability most significantly with the Storage-prioritized approach, followed by the Utopian point, TOPSIS, Balanced, Knee point, and, finally, the Supply-prioritized approach, as illustrated from right to left on the x-axis.

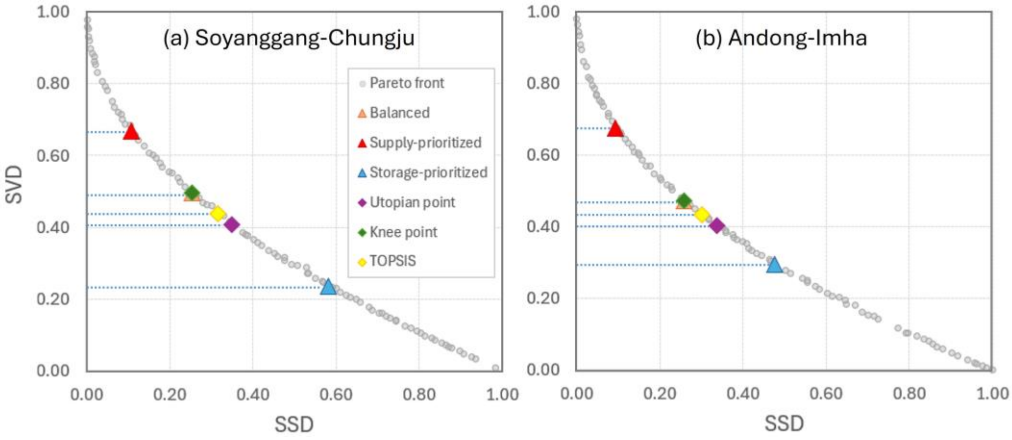

Figure R2. Examples of the Pareto front and decision-making results based on different MCDM methods for Soyanggang-Chungju (a) and Andong-Imha (b) in June 2014.

Taking into account this decision-making characteristic of the MCDM method, we reordered the MCDM methods (x-axis) from Storage-prioritized (Sto), Utopian point(UP), TOPSIS(TOP), Balanced(Bal), Knee point(KP) and Supply-prioritized (Sup), and isolated two Variable Weighting methods (SS, MW) as shown in Figure R3.

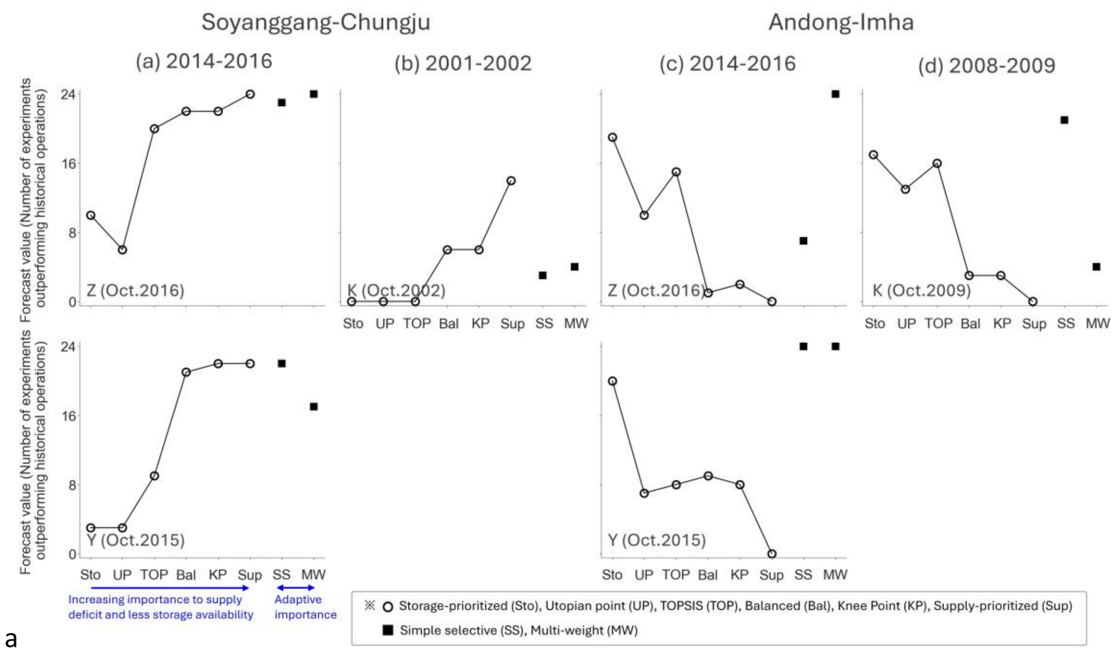

Figure R3: (First row) Forecast value (y-axis) at the end of different drought events (points at Z, K, Z, M in Figure 1) plotted against MCDM methods. The methods are ordered from left to right with increasing importance to supply deficit (hollow circles), along with two variable weighting methods (SS and MW, black squares). (Second row) Same as first row in the middle of drought event (points at Y in Figure 1). Here, the lines are not intended to imply continuity; they are included solely to clarify the direction for visualization purposes.

Figure R3 shows that, in the Soyanggang-Chungju reservoir system, the forecast value increases as the MCDM method prioritizes supply. In contrast, the opposite trend is observed in Andong-Imha. This discrepancy is closely linked to the characteristics of the drought events, as illustrated in Figure R4 below.

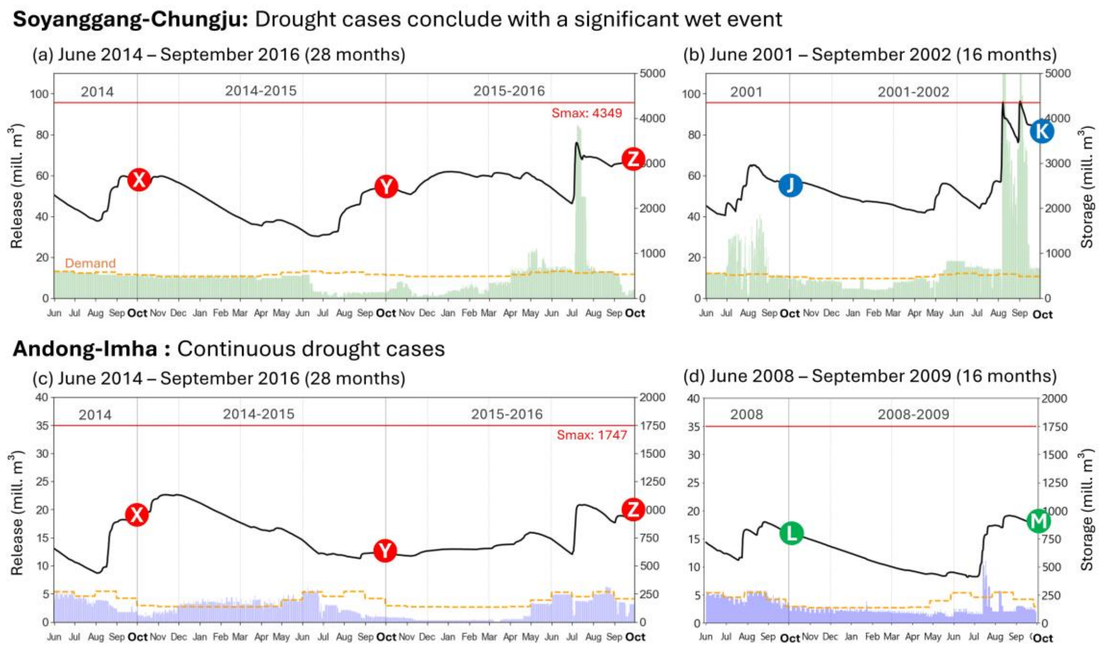

Figure R4: Daily reservoir operation records for the studied drought events (K-water, 2023). Points X, Y, Z, J, K, L, M represent the ends of the hydrological years (September 30th) which will be used as points for forecast value assessment.

For instance, the more emphasis is placed on supply deficit, the higher the value is achieved when the drought event concludes with a significant wet event (Soyanggang-Chungju, Figure R4(a, b)). This is because the effect of weighting storage availability is counteracted by the natural replenishment of storage from a wet event that occurs at the end of the simulation period. In contrast, MCDM methods that emphasize storage availability tend to achieve higher values when the drought event continues (Andong-Imha, Figure R4(c, d)).

We will address this in Section 5 'Discussion' and add the figures (R4-6) in the manuscript and supplementary material.

※ It is not true that our results show reservoirs separately. Please see reply to L132.

Figure 9: please clarify how is the "general expectation" curve obtained.

→ This curve conceptually illustrates the common understanding that higher forecast accuracy may lead to improved operational value. We will clarify this in the manuscript.

L461-462: please provide examples of the release scheduling policies derived by the different methods, in order to substantiate the idea of cautious operation. Also, please clarify the idea behind the concept "adverse events not seen during the optimization".

→ We can find the example of cautious operation before the wet event occurred in July 2016. Figure R5 shows that the flow forecast (a) and release decision (b) based on SFFs and deterministic scenarios (Worst Case Drought and 20 Year return period Drought scenarios). Considering flow uncertainty for the next two months, SFFs pertains higher releases compared to deterministic scenarios which predict continued dry conditions. Here, 'adverse events not seen during the optimization' can be the wet event occurred in July. We will add more details in the manuscript and include Figure R5 in the supplementary material.

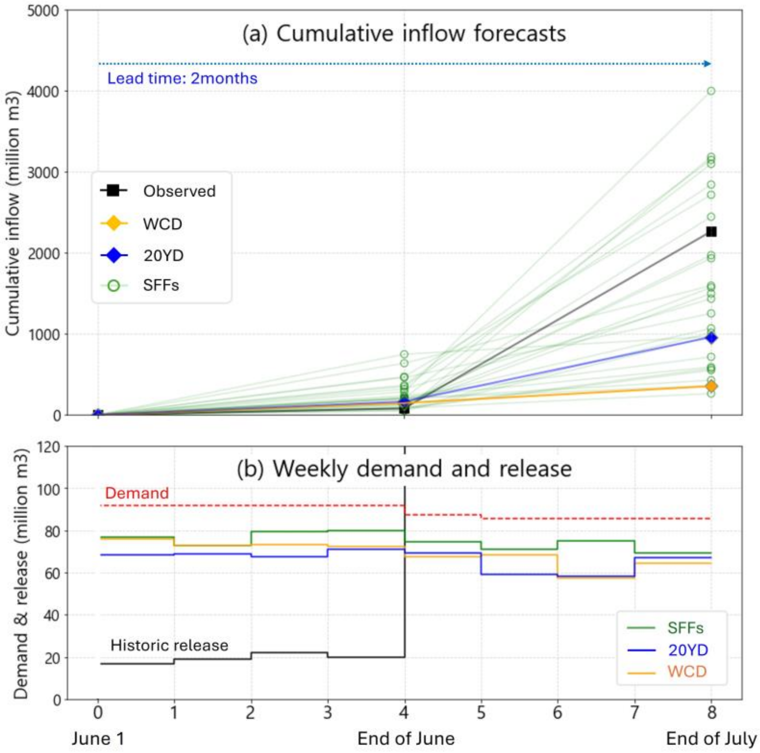

Figure R5. (a) Cumulative flow observation (black square) and forecasts for Soyanggang-Chungju from June to July 2016, using WCD (orange diamond), 20YD (blue diamond), and SFFs (ensemble: hollow green circle, median: red circle). The black square represents the observed cumulative inflow during the same period. (b) Weekly demand and release, following the same colour coding and time period as in (a).

L490. The logic behind this sentence is not evident.

→ Here, we aimed to highlight the importance of considering the uncertainty in flow forecasts. Specifically, while Figure 8(b) shows no clear relationship between the MCDM method and forecast value, Figure 8(d) indicates that the use of ensemble forecasts (i.e. considering the uncertainty in flow forecast) consistently provides higher value. We will revise this sentence to ensure a stronger connection to Figure 8, enhancing readers understanding.

L496-502. This is mostly speculation, which I suggest avoiding.

→ We believe that these sentences are grounded in practical considerations rather than speculation. For example, using the supply-prioritized method must be a good choice for a reservoir operation with ample storage capacity but lower demand. These sentences are significant as they offer recommendations for potential users in selecting an appropriate MCDM method, taking into account the primary purpose and operational characteristics of the reservoir.

L504-505. Why is this?

→ Even with perfect forecast of future inflows, limited lead times restrict our ability to consider subsequent conditions. As illustrated in Figure R6, if we use a lead time of 2-month and already recognize very dry conditions during this period, decisions will be made based solely on that dry condition. However, it is crucial to note that even a perfect forecast cannot account for wet events occurring beyond the 2-month horizon. We will revise the manuscript to explain this in more detail.

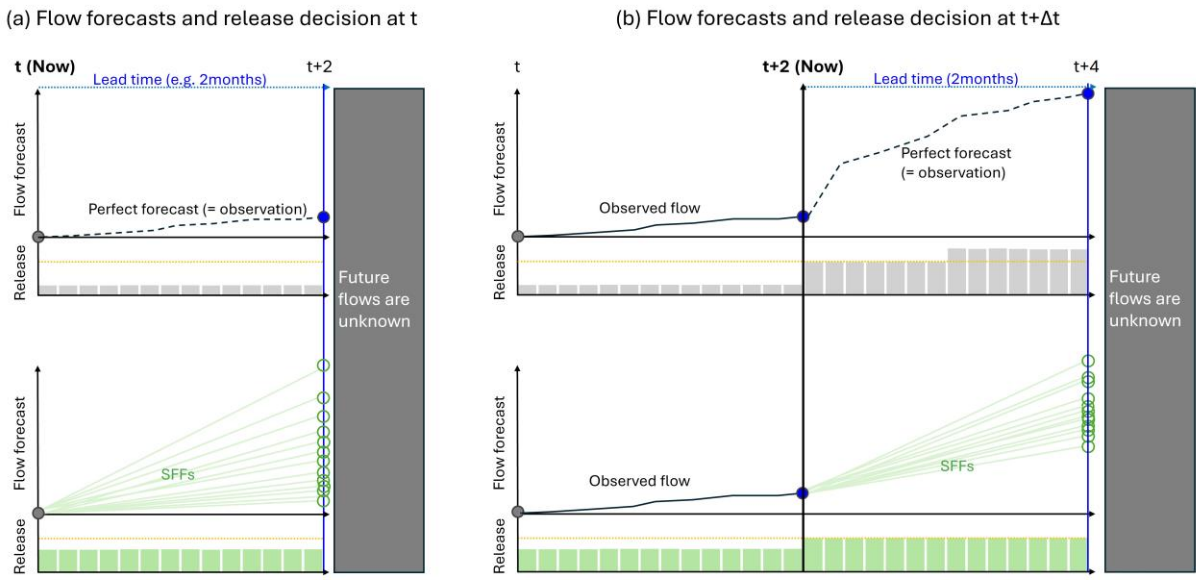

Figure R6. Conceptual illustration of the flow forecasts and release decision-making using a perfect forecast (upper row) and SFFs (lower row) at time t (a) and t + Δt (b).

L527. I dispute the idea that the results presented here have demonstrated anything, because the number of experiments is limited, and because no actual explanation has been provided for the mechanics of the obtained results.

→ We think this comment can be addressed by clarifying the main objectives of the paper, key contributions and the limitations of this study. Here, we will modify this paragraph saying that:

"Firstly, while we assessed forecast value across two reservoir systems and historical drought events, these case studies may not be sufficient to draw comprehensive conclusions. The primary reasons for the lack of study cases are the infrequent occurrence of extreme drought events and the limited availability of seasonal forecast data, which only became available in 1993. Since our results have shown the dependency of forecast value on reservoir systems and events, it is crucial to continue further application efforts to establish more general patterns in the relationship between accuracy and value, as well as to compare the performance between different forecast products. While our study faces challenges related to robustness, we expect that our open-source code developed through this study will enable potential users to easily replicate our experiments and validate our provisional results across other regions around the world."

Reference:

Cai, J., Liu, Y., Lei, T. and Pereira, L.S.: Estimating reference evapotranspiration with the FAO Penman–Monteith equation using daily weather forecast messages, Agricultural and Forest Meteorology, 145(1-2), 22–35, https://doi.org/10.1016/j.agrformet.2007.04.012, 2007.

Córdova, M., Carrillo-Rojas, G., Crespo, P., Wilcox, B. and Célleri, R.: Evaluation of the Penman-Monteith (FAO 56 PM) Method for Calculating Reference Evapotranspiration Using Limited Data, Mountain Research and Development, 35(3), 230. https://doi.org/10.1659/mrd-journal-d-14-0024.1, 2015.

Sanchez-Gomez, J., Vega-Rodríguez, M.A. and Pérez, C.: Comparison of automatic methods for reducing the Pareto front to a single solution applied to multi-document text summarization, Knowledge-Based Systems journal, 174, 123–136, https://doi.org/10.1016/j.knosys.2019.03.002, 2019.

While, L., Hingston, P., Barone, L. and Huband, S.: A faster algorithm for calculating hypervolume, IEEE Transactions on Evolutionary Computation, 10(1), 29–38, https://doi.org/10.1109/tevc.2005.851275, 2006.