

Author response

We thank the editor and referees for their careful reviewing and helpful comments. Our reply is given below. The page and line numbers correspond to the modifications done on the revised manuscript.

Editor

L54: lack of forecast performance → This has been corrected. (Page 2. Line 54)

L68: Numerous studies have been conducted → This has been corrected. (Page 2. Line 68)

L73: specific regions → This has been corrected. (Page 2. Line 73)

L82 wether to use SFFs, and when. → This has been corrected. (Page 2. Line 82)

L84: check spelling of reference for van Dijk → We have corrected this typo. (Page 2. Line 84)

L85: Mention is made of the coarse resolution of seasonal forecasts, suggesting these are 1x1 degrees. I would agree that this is the case for seasonal forecasts as made available through e.g. the CDS from which the forecasts have been download. It may be good to note that the original resolution of these forecasts is in fact on the order of 0.4x0.4 degrees, which have been resampled on the CDS. Please also see the documentation and references for the SEAS5 forecasts.

→ We thank you for this comment. In the Introduction section, we aim to present previous findings on the overall impact of bias correction. The details regarding the resolution of ECMWF's seasonal weather forecasts used in this study, are described in Section 2.1.2 (Data) and Section 4.2 (Limitations and directions for future research).

L104-L106: I appreciate that it is relevant to study the skill of seasonal forecasts in South Korea. However, I would consider the remark that precipitation is correlated to flow to be somewhat trivial as this applies to most catchments in the world.

→ We agree with your point, so this sentence has been removed.

L113: and also → This has been corrected. (Page 3. Line 113)

L123: rephrase this as: the catchments upstreamt of 12 multi-purpose reservoir

→ We agree with you and have rephrased the sentence as suggested. (Page 3. Line 122)

L125: the locations → This has been corrected. (Page 3. Line 125)

L138: please drop "changes in" as this is confusing - monthly precipitation and PET is sufficient.

→ We agree with the editor's comment, and we dropped 'changes in'. (Page 4. Line 137)

L141: vary → This has been corrected. (Page 4. Line 140)

L141: divide the year → This has been corrected. (Page 4. Line 140)

L150: simulate flow → This has been corrected. (Page 4. Line 149)

L151/L153: "produced" and "generated" are strange verbs to use in this context. Consider using "from"
→ We have replaced them to 'from'. (Page 4. Line 150-153)

L156: provide reference to FAO.

→ Thank you for this comment. We have added the reference. (Page 4. Line 155)

L160: I presume you mean hear the inflow to the reservoirs. May be good call this "daily inflow data" to make that clear. → We agree and have corrected it. (Page 4. Line 159)

L161: Instead of water supplies - I would describe this as releases from the reservoir - as that is what I presume you mean → We agree and have corrected it. (Page 4. Line 161)

L191: we analyse (note: given that this is presented in the current paper it should be present tense) → We agree and have corrected it. (Page 5. Line 190)

L228: original forecast datasets → This has been corrected. (Page 6. Line 227)

L229: each seasonal weather forecast → This has been corrected. (Page 6. Line 228)

L233: soil moisture structures → This has been corrected. (Page 6. Line 232)

L234: it is not clear what the comparison, so rather than "higher" just mention "good" → We agree and have corrected it. (Page 6. Line 233)

L281: $H(x)$ is often referred to as the Heaviside Function, rather than the Indicator function - hence the Heaviside

→ Thank you for this advice. We have replaced it to 'Heaviside function'. (Page 7. Line 280)

L282: members would exactly match the observations, and CRPS would equal 0.

→ This has been corrected. (Page 7. Line 281)

L283: a lower performance → This has been corrected. (Page 7. Line 282)

L287: I would say it compares to a benchmark forecast, as it is not a comparison to the method, but rather an application of the method → We agree and have corrected it. (Page 7. Line 286)

L293: has a lower performance than the benchmark → This has been corrected. (Page 8. Line 292)

L299: Low or no skill → This has been corrected. (Page 8. Line 297-298)

L338-L360: In this section and ensuing sections the terminology referring to the variables (Temperature, Precipitation and PET) that are used as forcing either from the observed dataset or from the forecast dataset. Various terms are used; weather forecasting forecast, weather forecats, weather forcing, etc. This makes the section confusing to read. This should be clarified and a consistent terminology used.

→ We thank you for this comment. To maintain consistency in the use of terminology throughout this paragraph, we have adjusted these expressions to 'weather forcing' (Page 9, Line 337-359).

L376: which leads → This has been corrected. (Page 10. Line 375)

L376: correct partitioning → This has been corrected. (Page 10. Line 375)

L392: should this read - to assist analysis? I am not sure who "analysts" refers to → We agree with your comment. We have replaced it to analysis. (Page 11. Line 391)

L446: should this read: wettest years and driest years? Or is this only the case in these two single years?
→ Thank you for this comment. Yes, it is the case in two single years, but they are the wettest and driest years in history in Chungju reservoir. To avoid confusion, we have added the specific year next to wettest and driest year. (Page 13. Line 445)

L475: has an equivalent level of performance → This has been corrected. (Page 14. Line 474-475)

L481: geographic locations → This has been corrected. (Page 14. Line 480)

L493-495: the sentence starting with "While" is somewhat confusing, please rephrase
→ We agree with you. We have modified the sentence. (Page 14. Line 492-494)

L500: the actual score → This has been corrected. (Page 15. Line 499)

L501: the theoretical score → This has been corrected. (Page 15. Line 500)

L563: the real world → This has been corrected. (Page 16. Line 563)

Referee 1

Line 145, is it “inter-annual” or “intra-annual” that is referred to?

→ This is additional information regarding the inter-annual climatic characteristic in South Korea. We have modified and relocated the sentence to improve its clarity (Page 4. Line 144-147)

Referee 2

L55: “real-world applications” → This has been corrected. (Page 2. Line 55)

L69 and L81: I would suggest introducing specific terminologies chosen in this paper (“actual skill”, “theoretical skill”) in the Methodology section instead, because the terminologies are not clearly indicated as coming from the literature, and because it is easier to identify the term with the help of Figure 2. In addition, here you still refer to skill, and in particular “theoretical skill”, but you calculate ‘actual scores’, ‘theoretical scores’ and a skill which could be called “actual skill” but no “theoretical skill”.

→ Thank you for your comment. We believe that separating the two skills in the introduction would indeed help in clarifying their respective meanings. Additionally, the following literature reviews are subject to each skill.

L107: “of various sizes” → This has been corrected. (Page 2. Line 106)

L110: I would suggest “long-term probability” → This has been corrected. (Page 2. Line 109)

L110: Here as well, the notion of “overall skill” might be too specific for the introduction. You could simply refer to “skill” here and detail the terminologies (actual, theoretical, actual) in the Methodology section.

→ Thank you for this comment. To prevent confusion with the term 'skill' in following sections, it would be more appropriate to use 'overall skill' from the introduction.

L135: Just to be sure I understand correctly: for each month, you first take the minimum (or maximum), and then average the values obtained for a given month over all years. Is that correct?

→ Yes, it is correct.

L192: Consider adding “calculated using flow observations” after “actual score” to

→ We agree with you and have modified it. (Page 5. Line 191)

L194: Here you ignore all errors from the hydrological model, and not only the ones due to the structure. Consider expanding. → This has been corrected. (Page 5. Line 193)

L195: “influences” → This has been corrected. (Page 5. Line 194)

L249-250: In my opinion, this statement still requires some clarification. Results may be superior but surely for certain flow characteristics. This objective function should be good to fix the bias, but will, for instance, not focus on extreme values.

→ Based on the previous research (Kang et al., 2004), this objective function generally showed good performance, not for specific conditions. We have modified the sentence. (Page 7. Line 248)

L260-261: This is not a critical issue, but here you are considering two metrics that resemble bias (assessing overestimation and underestimation), and that are redundant. → Although they can both be used to assess overestimation and underestimation, they have distinct physical meanings.

L283: "indicates a lower performance" → This has been corrected. (Page 7. Line 282)

L299: "the presence of few extremely negative values" → This has been corrected. (Page 8. Line 298)

L320-323: Here, a factor that may better explain the loss from calibration and validation is rather a temporal one, such as a trend or inhomogeneity. Could human influences in these three catchments cause this loss in performance?

→ Thank you for this comment. In selecting target reservoirs, we ruled out all the reservoirs with human influence, such as artificial inflow. (Page 3. Line 122-124)

L329 "based on accumulated monthly mean flow": I am not sure I understand this. Based on the Methodology section it seemed you were considering monthly mean flow. Do I understand correctly here that you still take the monthly average and then accumulate it over lead months? If you refer to the variable shown in Figure 9, values do not seem to be averaged monthly. Regarding the decrease in score values with the lead time, this is expected due to the loss of information from initial conditions, both in the weather and hydrological systems.

→ In calculating monthly CRPS, we used accumulated flow for a given lead time. The value of CRPS for each reservoir, shown in Figure 4, represents the mean monthly CRPS for the reservoir. However, Figure 9 illustrates cumulated daily simulated flow for specific lead times, rather than scores or skills, so it is not averaged.

L335: "the driest" → This has been corrected. (Page 9. Line 334)

L347-350: Please consider rephrasing this sentence. The subject of "forced" is not fully clear.

→ We have rephrased the sentence. (Page 9. Line 346-348)

L419-420: This phrasing is confusing. I suggest rephrasing simply to "The pale black points represent the overall skill in each catchment." → This has been corrected. (Page 12. Line 487-419)

L421: This does not seem true for the dry season, since SFFs outperform ESP only (maybe) in the first lead month. → We agree with you and have modified it. (Page 12. Line 420-421)

L422: "SFFs are" → This has been corrected. (Page 12. Line 421)

L431: "in the Southern region" → This has been corrected. (Page 12. Line 430)

L444-445: This sentence now seems redundant with the one added. Please consider removing.

→ They have different meaning. While added sentence (Line 434-435) explains the general tendency of 'overall skill' across entire period, this sentence (Line 443-444) refers to the performance of flow forecast in specific year and season, explaining Figure 9.

L446: I suggest nuancing. For instance, for the wettest year, there is underestimation at the scale of the season but overestimation in the first lead month.

→ It is difficult to agree that 'there is overestimation in the first lead month'. But we slightly modified the sentence as you suggested. (Page 13. Line 445-446)

L448-449: It seems to depend on the horizon you look at. If ensembles were less sharp, they might have had a chance at including the observation in their range (Jul-Oct).

→ We agree with you that the term 'sharpness' might raise confusion; therefore, we have replaced it to performance. (Page 13. Line 447)