REPLY TO REFEREE#1 COMMENTS

In this paper, the authors have conducted an evaluation of seasonal hydrological forecasts using Tank model driven by the ECMWF SEAS5 seasonal weather forecasts in various South Korean catchments. The topic is timely and brings insights to the field of hydrological forecasting. However, I have identified several areas that require attention to ensure the manuscript meets the standards of the journal.

Thank you for your valuable and insightful comments to our paper. We are committed to address them in our revision.

1. The manuscript currently needs a clearer explanation of the methodologies employed, which also relates to the second point of terminology use. Some parts of method description need to be expanded. For example, the calculation of contributions from each variable needs more details. When examining the contribution of forecasted precipitation by substituting the other two variables with observations, please clarify if this was done for all ensemble members, and is it skill that you are compared on or score?

   We thank you for this comment and agree with your point. We will provide more specific details of the methodologies used, particularly in explaining the contribution of weather variables to the performance of SFFs and clarify the skill/score issue (see next point).

2. Consistent Use of Terminology: The paper misuses the word “skill” in different ways which should be sometimes specified by performance or quality, and this inconsistent use lead to confusion unfortunately. There seems to be also some ambiguity in the use of “skill” and “score”. For instance, “theoretical skill” and “actual skill”, are they referred to CRPS or CRPSS? If they are skills, could you specify which benchmark is being used to calculate them? If they are scores, then I suggest not using skill in the names. I would suggest the authors to clearly distinguish these terms and maintain consistent usage throughout the manuscript to improve clarity.

   Thanks to this comment, we have recognized that our terminology might be causing confusion among readers. We will replace the term ‘skill’ with ‘score’ in sections 3.1 and 3.2, where the ‘CRPS’ is used. The term ‘skill’ is retained only in sections 3.3 and 3.4, where the overall skill (calculated using ‘CRPSS’) is discussed (and we will clarify that the benchmark is ESP).

3. Abbreviation Usage: The manuscript needs a thorough review to ensure that all abbreviations are properly introduced upon their first occurrence. Additionally, to avoid redundancy, each abbreviation should only be defined once. Moreover, certain abbreviations have been assigned multiple meanings within the manuscript (for instance, CRPS at Line 183 and Line 258). This presents a significant source of confusion.

   We agree with this comment. We have checked the overall usage of abbreviations and will correct them across the manuscript. Please refer to the details shown below (Line by line reply).

4. Actual Skill and Theoretical Skill: The authors raise an interesting point about the significance of using actual skill over theoretical skill to provide more insights for water resource management on whether to use SFFs and when. However, in the analysis, this is conducted by calculating CRPSS, with ESP as benchmark, thus the use of either actual or theoretical references does not play such a big role, as long as the benchmark is using the same reference as the forecasts. On the other hand, the information gained from theoretical skill in this paper is to validate the performance of the hydrological model by showing its proximity to the actual skill (or perhaps more appropriately,

the “actual score”). In this case, it didn’t really reflect the argument of providing significant information for the users.

We appreciate your comment. This misunderstanding is also caused by the misuse of the term ‘skill’. In this study, we use CRPSS to compute ‘overall skill’ by comparing the flow forecasts with real observations. In previous literature, this was often referred to as ‘actual skill’. As acknowledged in the introduction, we believe that ‘actual skill’ (compared with real observations) would be more informative for water managers than ‘theoretical skill’ (compared with pseudo-observations). Additionally, to validate the performance of the hydrological model (Section 3.1), we calculated the actual and theoretical ‘score’ using CRPS, therefore, no benchmark is used here. To clarify this, we have modified Figure 2 (see page 4 in this document) and proposed revisions to the manuscript in line-by-line reply below.

Line by Line comments (reply):

Line 23, “actual skill” here sounds ambiguous since there is no other information explaining this term, which might lead to misunderstanding.

→ Agree. We will replace the term ‘actual skill’ with ‘overall skill’ with additional information explaining this term.

Line 25, please add brief information on the methods that you use to get the conclusion that precipitation is the most important variable.

→ Agree, we will add explanations of the methods.

Line 57, this is the first time that ESP is mentioned (excluding abstract), therefore full explanation is needed here.

→ Agree, we will add the full explanation of ESP.

Line 77, to my knowledge, the reference Pechlivanidis et al., 2020 is not using ESP in the analysis, therefore cannot support the argument here.

→ Agree and will remove the reference from the paragraph.

Line 93, I’m a little bit suspicious on this sentence here that “only a few studies” have used SEAS5 for seasonal hydrological skill assessment. For example, the reference you mentioned before Pechlivanidis 2020 is actually using SEAS5 at higher spatial resolution.

→ Here, we wanted to emphasize that there are not many previous studies using ECMWF SEA5 and analysing the performance of SFFs compared to ESP. To clarify our intention, we will modify the sentence.

Line 139, what is the criteria of dividing the four seasons, are they based on precipitation or flow?

→ The criteria of dividing the four seasons is based on monthly precipitation. This is intended to maintain continuity with our previous research (Lee et al., 2023) and is consistent with the general seasonal classification in South Korea. We will clarify this point.

Line 142, the information of annual variability is not shown in the figure but only in the text, right?

→ Yes, the inter-annual variability is not shown in the figure. We will include additional explanation and a reference (Lee et al., 2023) to support this.

Line 143, typhoon and monsoon might not need to start with an uppercase character here.

→ Agree, we will change them to lowercase.

Line 149, the abbreviation of KMA should be noted in the previous sentence when it is firstly mentioned.

→ Agree, we will add the abbreviation of KMA.

Line 169, regarding SEAS5 data, here the period 1993-2020 is mentioned, but in the method part and in Figure 2, based on my understanding, the forecast period is 2011 to 2020. Please clearly specify this.

→ Thank you for this comment. Our analysis focuses on the period from 2011 to 2020. However, we also analysed SEAS5 data from 1993 to 2010 to compute the bias correction factors. A detailed explanation of this process will be provided, along with Figure R1 (shown below). This figure will be added in the supplementary material.

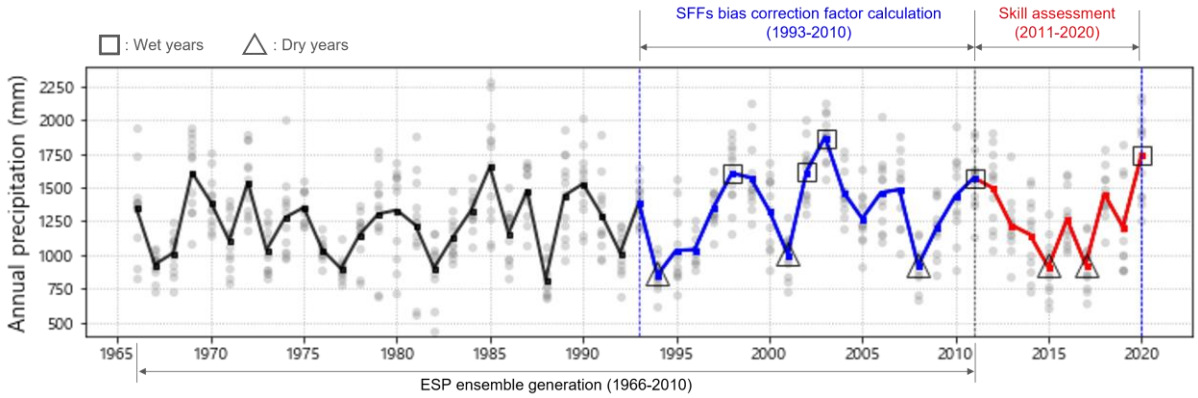


Figure R1: Observed annual precipitation (dots) from 1966 to 2020 in the 12 catchments feeding the reservoirs considered in this study. The solid line represents the mean annual precipitation over the 12 catchments. The red line represents the period for assessing the seasonal flow forecasts (2011–2020), and the blue line represents the period used to compute the bias correction factors (1993–2010). ESP ensembles are generated using observed data from 1966 to 2010.

Line 181, SFFs has been mentioned many times already.

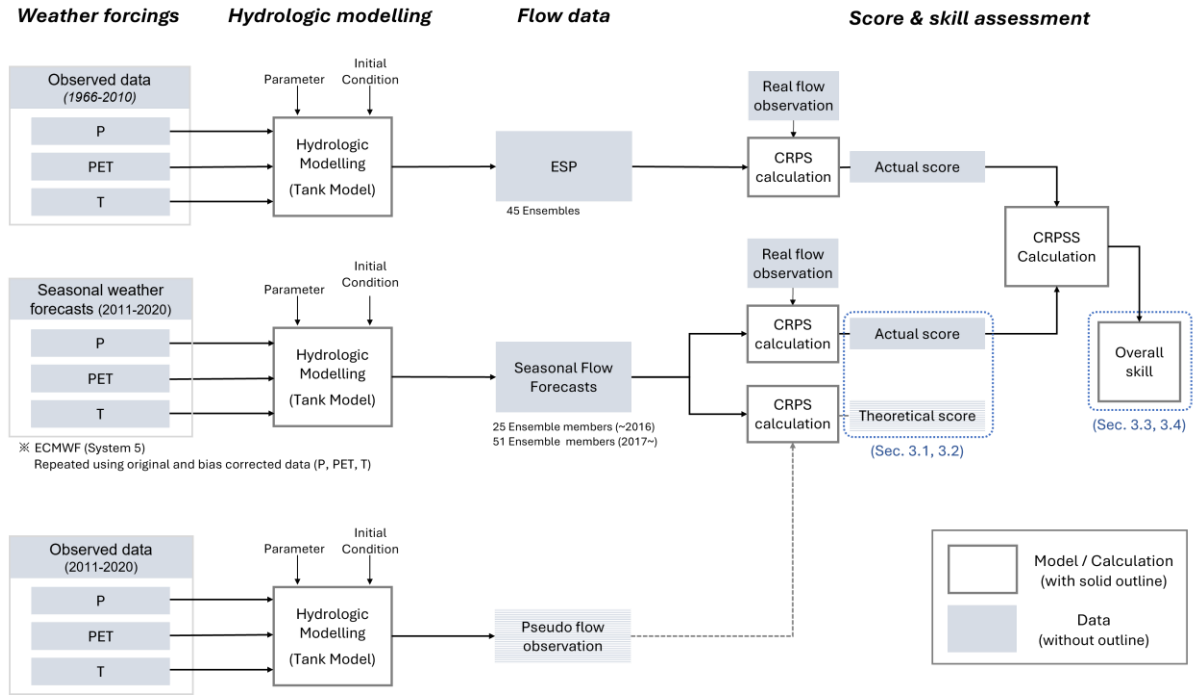
→ We will remove the full form of SFFs.

Line 183, here CRPS is referred to as skill but later it is referred to as score (Line 258).

→ We will use the term score consistently.

Line 188, the plot needs to be improved. To calculate CRPS needs the forecast (either ESP or SEAS5) and the reference (either real or pseudo-observation), therefore the arrows should lead from corresponding systems to the box of CRPS. However, this is not systematically shown in the plot.

→ Thank you for this suggestion. We have improved Figure 2 to clarify our methodology and the term. Please see the modified Figure 2 below.



Modified Figure 2: Schematic diagram illustrating analysis method of the study.

Line 190, to my knowledge there is SEAS5 forecasts with higher spatial resolution that is available.

→ We thank you for this comment. In this study, we have utilised the forecasts with 1×1° because we wanted to maintain consistency with our pervious study on the seasonal precipitation forecasts (Lee et al. 2023). In addition, the Copernicus Climate Data Store (CCDS) website officially states that their service has a horizontal resolution of 1×1° (https://cds.climate.copernicus.eu/cdsapp#!/dataset/seasonal-original-single-levels?tab=overview).

To clarify the relationship between the higher (36×36km, approximately 0.3×0.3°) and lower resolutions (1×1°) data, we directly contacted ECMWF. ECMWF indicated that they provide lower resolution data publicly through CCDS, and this data is generated from higher resolution data using an in-house interpolation method.

We have conducted a test to compare the forecasts with higher and lower resolution in three catchments with a large, medium, and small area, respectively. Figure R2 shown below represents the monthly P, T and PET obtained from higher and lower resolutions (before bias correction). The difference between both products is generally small, with a somewhat greater disparity in PET (orange points). However, overall, the differences were not substantial, therefore, increasing the spatial resolution may not significantly impact our results and conclusions for the catchments analysed in this study.

Additionally, we also have compared the mean P, T and PET for a single coarse grid cell (1×1°) and nine finer grid cells (36×36km) contained in the single course grid. This approach has been repeated for two different locations, see Figure R3. Again, the differences between the two datasets are small and consistent with the previous findings shown in Figure R2.

To clarify this point, we will include additional explanation in Section 4.2 (Limitations and directions for future research).

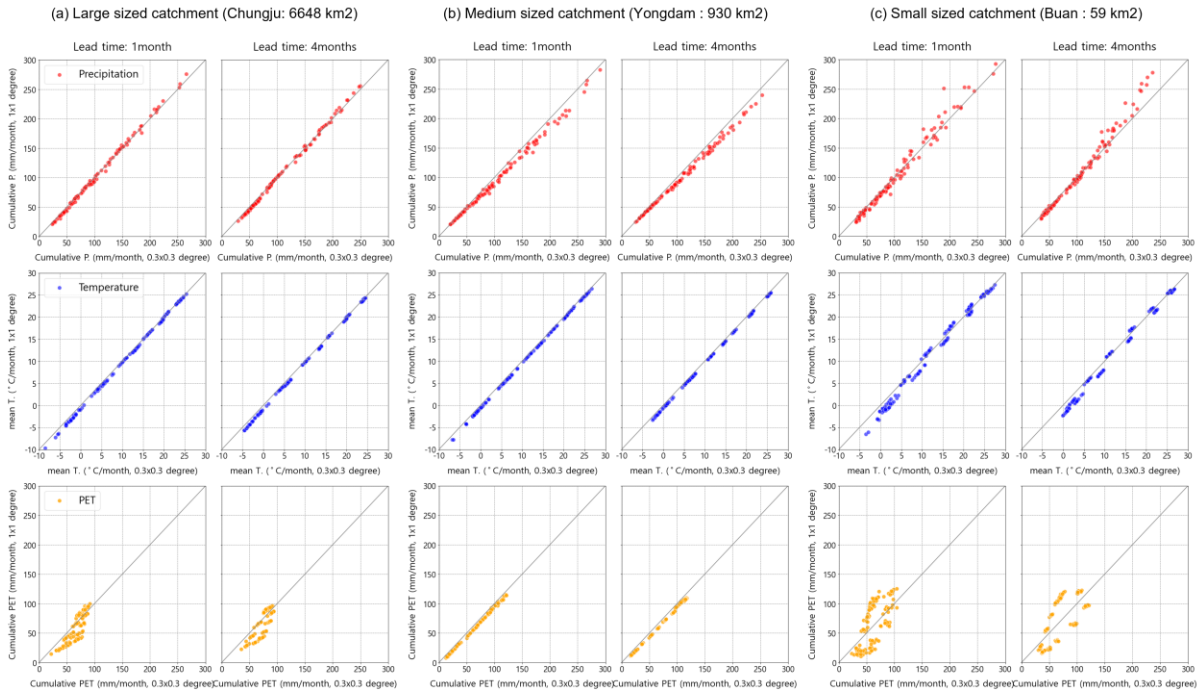


Figure R2. Comparison between mean monthly weather forecasts of the high resolution product (0.3×0.3° or 36×36km, on x-axis) and low resolution (1×1°, y-axis) (first row: precipitation, second row: temperature, third row: PET) at 1- and 4-month lead times from 2011 to 2020. Analysis is repeated in three catchments: (a) large-size (Chungju), (b) medium (Yongdam) and (c) small (Buan).

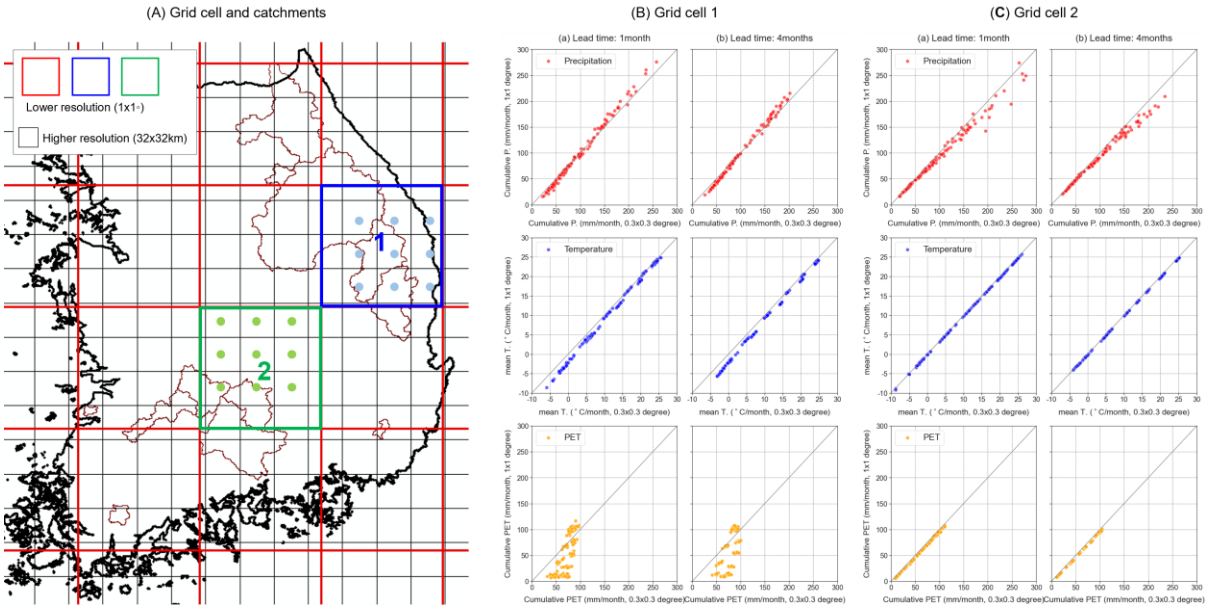


Figure R3. Comparison between the mean monthly weather forecasts for a grid cell of the lower resolution (1×1°) product (vertical axis) and the average of 9 grid cells of the higher resolution (36×36km) product (horizontal axis), at different lead time (a: 1 month, b: 4 months) from 2011 to 2020. Analysis is repeated for two regions.

Line 205, a potential problem for linear scaling on precipitation is, it might generate very large values. Have you had any solutions to avoid this?

→ Thank you for your comment. In our study, we could not find any problem generating very large values. Additionally, the linear scaling method has demonstrated its usefulness in literature (Azman et al., 2022; Crochemore et al., 2016; Shrestha et al., 2017) and South Korea (Lee et al., 2023).

Line 247, as defined in Eq.4?

→ We have corrected this typo.

Line 265, what does SPFs stand for? Or maybe you mean SFFs? Otherwise please add the full name for the abbreviation.

→ Thanks for spotting this typo. We will replace it with SFFs.

Line 270 and Line 258, redundant information.

→ Agree. We will remove this redundant information.

Line 275, Major does not need an uppercase here.

→ Agree. The words will be changed to lowercase.

Line 275, here the CRPS of ESP is calculated using real observation as reference, it is correct?

→ Yes, it is. Basically, we have computed the overall skill using real observations. To clarify this, we have modified Figure 2 (see page 4 in this document) and the sentence.

Line 285, here comes the explanation of SPFs, but it is already mentioned many times before this.

→ This has been amended in accordance with your previous comment related to Line 265.

Line 310, here I would strongly recommend to distinguish skill from score, since you have CRPSS later which are actually skills, but here these are scores.

→ Thanks for your advice. We will modify the term 'skill' to 'score' in those sentences where the CRPS is used.

Line 327, this part should be described in method session, and more details are needed for fully understanding.

→ We agree with you and have moved this part to method session and added descriptions.

Line 498, are these conclusions from Figure 8? Considering there are only two dry years and two wet years, the conclusion needs to be drawn carefully, otherwise it's not very scientifically valid.

→ We agree that having only two dry and two wet years means that we cannot draw definitive conclusions. We will revise our discussion to recognise this in our conclusions. Additionally, we have produced an additional Figure with the same analysis as Figure 8 but including data from the calibration period used to calculate bias correction factors. This extended the analysis to 5 dry years (1994, 2001, 2008, 2015, 2017) and 5 wet years (1998, 2001, 2002, 2011, 2020), respectively. (Please note that, due to the lack of observed data, here we can only use 7 catchments: Soyanaggang, Chungju, Andong, Imha, Hapcheon, Namgang, Sumjingang).

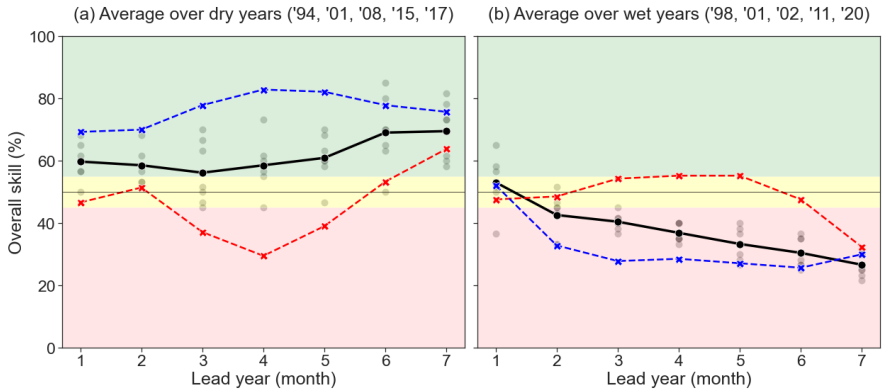


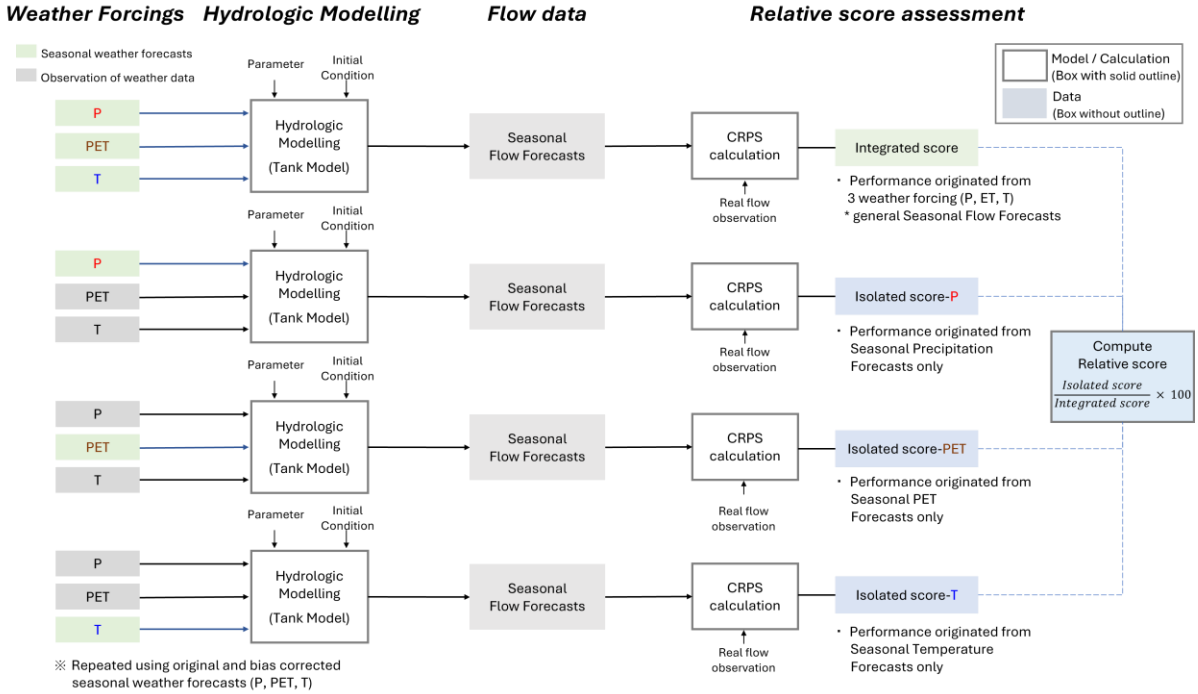
Figure R4. Overall skill of bias corrected SFFs over 7 catchments averaged over (a) dry years (mean annual P < 900mm) and (b) wet years (mean annual P > 1500mm) during all seasons (black lines), dry seasons (red dashed lines) and wet seasons (blue dashed lines). Here, mean annual precipitation is averaged across the catchments and years.

As shown in Figure R4, the results are generally consistent with Figure 8 (b, c), which is encouraging and, considering the available seasonal forecasts dataset (1993-), likely the broadest analysis that we can conduct. We will bring this discussion in the revised manuscript and acknowledged this limitation of our study.

Figure S2, please explain which benchmark is used here to calculate from CRPS to skill (skill-P, skill-T).

→ No benchmark was used; thus, we have modified Figure S2 as shown below. This figure illustrates that how we computed the contribution of each weather variable to the 'score' of SFFs (not skill).

.



Modified Figure S2: Schematic diagram of calculating the relative scores.